University of California
Santa Barbara

# Supersymmetry and the S-matrix

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Physics

by

Timothy James Trott

Committee in charge:

Professor Mark Srednicki, Chair
Professor Steve Giddings
Professor Claudio Campagnari

March 2023

The Dissertation of Timothy James Trott is approved.

---

Professor Steve Giddings

---

Professor Claudio Campagnari

---

Professor Mark Srednicki, Committee Chair

December 2022

## Acknowledgements

The work presented in Chapters 1 and 2 were performed in collaboration with Aidan Herderschee and Seth Koren. The author's contribution is subdominant in significant parts of sections 1.4.1 and 1.4.2, as well as 2.4.4 and 2.$B$.

The work presented in Chapter 4 was performed in collaboration with Nathaniel Craig and Seth Koren. The author is not responsible for the introductory section, most of the conclusion and section 4.5.

Each chapter has previously been published in the Journal of High Energy Physics (JHEP) and may be reproduced here under the Creative Commons license (bibliographic information may be found below in Curriculum Vitae).

# Curriculum Vitæ

Timothy James Trott

## Education

| | |
|---|---|
| 2023 | Ph.D. in Physics (Expected), University of California, Santa Barbara, , supervised by Professor Mark Srednicki. |
| 2013 | M.Sc. in Physics, University of Melbourne, supervised by Professor Tony Gherghetta. |
| 2011 | B.Sc. In Physics, University of Melbourne. |
| 2011 | Dip. Math. Sc., The University of Melbourne |

## Publications


Trott, T., Causality, Unitarity and Effective Field Theory, J. High Energ. Phys. 07 143 (2021).

Herderschee, A., Koren, S. & Trott, T., Constructing $\mathcal{N} = 4$ Coulomb Branch Superamplitudes, J. High Energ. Phys. 08 107 (2019).

Herderschee, A., Koren, S. & Trott, T., Massive On-Shell Supersymmetric Scattering Amplitudes, J. High Energ. Phys. 10 092 (2019).

Craig, N., Koren, S & Trott, T., Cosmological Signals of a Mirror Twin Higgs, J. High Energ. Phys. 05 038 (2017).

Gherghetta, T., von Harling, B., Medina, A. D., Schmidt, M. A. & Trott, T., SUSY Implications from WIMP Annihilation into Scalars at the Galactic Centre, Phys. Rev. D 91, 105004 (2015).

MSc Thesis: Naturalness and Vacuum Stability in Non-Minimal Supersymmetry.

## Abstract

Supersymmetry and the S-matrix

by

Timothy James Trott

The functional form of the $S$-matrix is heavily constrained by both causality and unitarity to the extent to which it may be substantially reconstructed without computational recourse to a field-theoretic action, thereby avoiding complications arising from gauge fixing and field-variable redundancy. In this context, simplifications provided by supersymmetric theories have been fruitful in guiding these methods and enabling perturbatively deeper computations.

In this thesis, on-shell, unitarity-based $S$-matrix methods are generalised to supersymmetric theories of massive particles. An on-shell superspace for massive supermutiplets is developed and used to classify all three-particle interactions consistent with supersymmetry in gauge theories in four dimensions. In $\mathcal{N} = 4$ super-Yang-Mills theory, this is used to compute scattering amplitudes for BPS particles using on-shell recursion.

The impact of causality on the contact couplings parameterising effective theories are then considered. Dispersion relations are used to constrain the space of these couplings for several simple interactions relevant to extensions of the Standard Model of particle physics. It is shown how these bounds unify as components of the same constraint under supersymmetry.

Finally, as the observable world lacks supersymmetry, a possible intermediary extension of the Standard Model called the Twin Higgs, bridging a minor hierarchy of scales between the Higgs potential and a UV completion, is considered. In particular, the impact of numerous new, light, invisible particles on the spectrum of temperature

anisotropies of the Cosmic Microwave Background is calculated as a function of the twin electroweak scale. The observable impact of a possible dilution of the twin particles relative to their Standard Model counterparts by the decay of a particle reheating the universe is then projected.

# Contents

# Chapter 1

# Massive On-Shell Supersymmetric Scattering Amplitudes

We introduce a manifestly little group covariant on-shell superspace for massive particles in four dimensions using the massive spinor helicity formalism. This enables us to construct massive on-shell superfields and fully utilize on-shell symmetry considerations to derive all possible $\mathcal{N} = 1$ three-particle amplitudes for particles of spin as high as one, as well as some simple amplitudes for particles of any spin. Throughout, the conceptual and computational simplicity of this approach is exhibited.

## 1.1 Introduction

The spinor helicity formalism has been a key ingredient in developing a purely on-shell formulation of $S$-matrix computations in four dimensions. This is because helicity spinors may be used as a complete description of the data of external scattering states (that is, their momentum and spin polarisation) without recourse to the unnecessary non-linear gauge redundancy of polarisations used in the Feynman rules. This can be coupled with

on-shell methods, such as recursion and generalised unitarity, to perturbatively build the internal $S$-matrix structure out of on-shell units that bypass the need for the fictitious degrees of freedom that frequently arise in standard field-theoretic methods. However, the dream of a fully on-shell formulation of particle physics is far from realised.

Helicity spinors have been adapted to describe the kinematics of massive particles previously in (1–5). This generally involved decomposing time-like momenta into two null vectors and then proceeding with massless helicity spinors to describe each of these in some way. However, the element of arbitrariness in this decomposition often convoluted the method, as sought-after patterns could easily be obscured by an inappropriate choice. Furthermore, this choice of direction often involved a spurious breaking of Lorentz invariance in the amplitudes.

An advance on this formalism was made in (6), where the spinors of the null vectors were organised into representations of the little group for massive momenta, $SU(2)$. This symmetry represents the redundancy in the spinor description of momentum, analogous to the $U(1)$ redundancy in the massless case. However, it may also be utilised to describe the polarisations of the external massive states, or better, to directly use symmetries to build amplitudes that have the required transformation properties of the external states under their individual little group rotations. While there was never a gauge ambiguity in the polarisations of massive particles in the Feynman rules, non-existent time-like components still source tension in a symmetric treatment of a 4-vector description of these fundamentally 3-vector objects (recent use in effective field theories was given in (7)). Massive and massless particles may thus be treated on equal footing within the spinor helicity formalism.

Supersymmetry (SUSY) offers an idealisation that, in theories of massless particles, has enabled the utility of on-shell methods to be drastically extended. It is thus natural to look to supersymmetry as a testing grounds for on-shell methods for massive particles. We

therefore here amalgamate the little group-covariant helicity spinors for massive particles with the formulation of an on-shell superspace in which external scattering states are grouped into supermultiplets without reference to an external spin direction. This makes the relations between the amplitudes imposed by the supersymmetric Ward identities (SWIs) transparent while simultaneously preserving the polarisation structures. See (8) for a review of on-shell superspace for massless particles. An on-shell superspace for massive particles was first constructed in (9) and we will rediscover their results along the way, albeit re-expressed in the covariant formalism. This helps to organise the amplitudes into Lorentz-covariant terms that are simpler to interpret, identify and construct.

After laying the foundation by writing the superalgebra in a little group covariant form and constructing covariant supermultiplets, we turn to $\mathcal{N} = 1$ theories to exhibit the usage and utility of this formalism. We construct from first principles all possible three particle amplitudes, the most primitive on-shell scattering data, that are consistent with these symmetries and involve particle spins no greater than one. We also make some comments on how SUSY generally constrains interactions with higher-spin states. The on-shell supersymmetry allows us to simply catalogue the most general possible interactions given only the spectrum of a theory. It is also easy to further specialize by incorporating additional on-shell data such as the presence of a parity symmetry relating some component amplitudes to each other, or the absence of self-interactions for a vector in an Abelian theory. By studying the high energy or massless limits of superamplitudes, we may obtain the necessary dependence of the couplings on the masses of the external legs if the states are to be identified as elementary superfields.

This paper is structured as follows. In Section 2 we present the little group covariant on-shell SUSY algebra. This allows us in Section 3 to construct massive on-shell supermultiplets as coherent states of the supercharges in any reference frame. In Section 4 we discuss general features of superamplitudes and strategies for their construction,

including the implementation of parity symmetry. We exhibit all of this technology in Section 5 to construct elementary three particle amplitudes for flat space $\mathcal{N} = 1$ theories. We then conclude.

## 1.2 Little Group Covariant Superalgebra for Massive Particles

The general super-Poincaré algebra extends the Poincaré algebra to a graded Lie algebra in 4 dimensions through the introduction of $\mathcal{N}$ fermionic generators $Q_{\alpha A}$, $Q^{\dagger}_{\dot{\beta}B}$, where $\alpha, \dot{\beta}$ are $SL(2,\mathbb{C})$ indices and $A, B = 1 \ldots \mathcal{N}$ count the number of left-handed spinor supersymmetry generators. The Lie brackets of the generators of supersymmetry - or supercharges - with the generators of translations ($P_\mu$) and rotations/boosts ($M_{\mu\nu}$) are (following (10))

$$\begin{aligned}
[Q_{\alpha A}, P^\mu] &= 0, \\
[Q_{\alpha A}, M^{\mu\nu}] &= \frac{i}{4}\epsilon^{\dot{\alpha}\dot{\beta}}(\sigma^\mu_{\alpha\dot{\alpha}}\sigma^\nu_{\beta\dot{\beta}} - \sigma^\nu_{\alpha\dot{\alpha}}\sigma^\mu_{\beta\dot{\beta}})Q^\beta_A, \\
\{Q_{\alpha A}, Q^{\dagger B}_{\dot{\beta}}\} &= -2\delta^B_A(\sigma^\mu_{\alpha\dot{\beta}})P_\mu, \\
\{Q_{\alpha A}, Q_{\beta B}\} &= Z_{AB}\epsilon_{\alpha\beta}, \\
\{Q^{\dagger A}_{\dot{\alpha}}, Q^{\dagger B}_{\dot{\beta}}\} &= -Z^{AB}\epsilon_{\dot{\alpha}\dot{\beta}}.
\end{aligned} \tag{1.1}$$

The automorphism group of the supercharges preserving the anticommutation relations is the $R$-symmetry group and will be discussed further in what follows. The 'central charge' $Z_{AB} = -Z_{BA} = -(Z^{AB})^*$ is allowed for $\mathcal{N} > 1$ and typically breaks the $R$-symmetry to a subgroup.

We will be interested in the construction of superamplitudes, which package together scattering data for entire representations of the super-Poincaré algebra. Before discussing

this, we will first rewrite the superalgebra using the massive spinor helicity language. This allows the spinor indices to be stripped out of the supercharges, leaving an elegant, frame-independent formulation of the algebra from which massive representations can be simply constructed. This provides an aesthetic improvement over previous treatments (11) in addition to setting up our discussion of superamplitudes.

In Appendix 1.A.1, we present a lightning review of massive spinor helicity in which we also develop our conventions and provide relevant and useful identities. The reader can find there further introduction to the subject and the elementary mechanics which will not be remarked upon in the main text.

The external particles in a scattering amplitude are acted upon by the super-Poincaré generators as separate tensor factors of the scattering state. Each symmetry generator may therefore be represented on a scattering amplitude as the sum of its action on each external scattered particle. This will allow us to study symmetry generators and transformations on each leg separately. We will use spinor helicity variables to represent these generators, because these encapsulate the on-shell kinematic data for each leg. For massive particles, this means exhibiting the $SU(2)$ little group symmetry by expressing the symmetry generators acting on each particle $i$ in an appropriately covariant fashion. By construction, the momentum eigenvalue of particle $i$ is $p_i^{\dot{\alpha}\beta} = p_i^\mu \sigma_\mu^{\dot{\alpha}\beta} = \left|i^I\right\rangle^{\dot{\alpha}} \left[i_I\right|^\beta$. We can likewise define on-shell, little group covariant supersymmetry generators for each leg by projecting the supercharges onto the spinors of a given particle

$$q_{i,A}^I = \frac{-1}{\sqrt{2}m_i}\left[i^I Q_{i,A}\right], \qquad q_{i,I}^{\dagger A} = \frac{1}{\sqrt{2}m_i}\left\langle i_I Q_i^{\dagger A}\right\rangle, \tag{1.2}$$

where the factor of mass $m_i$ of the particle makes them dimensionless. Note that we are defining these operators as being restricted to single-particle momentum eigenspaces.

For convenience, the inverse relations are given by

$$Q_{i,\alpha A} = -\sqrt{2}\,|i_I]_\alpha\, q^I_{i,A} \qquad Q^{\dagger A}_{i,\dot\beta} = \sqrt{2} q^{\dagger A}_{i,I}\,\langle i^I|_{\dot\beta}\,. \tag{1.3}$$

The factor of $1/\sqrt{2}$ is just a normalisation convention and has been chosen here so that the little group covariant supercharges satisfy the anticommutation relations

$$\left\{q^I_{i,A}, q^{\dagger J,B}_i\right\} = -\epsilon^{IJ}\delta^B_A, \qquad \left\{q^I_{i,A}, q^J_{i,B}\right\} = -\epsilon^{IJ}\frac{Z_{i,AB}}{2m_i}, \qquad \left\{q^{\dagger I,A}_i, q^{\dagger J,B}_i\right\} = \epsilon^{IJ}\frac{Z_i^{AB}}{2m_i}. \tag{1.4}$$

These hold only on a particular single-particle momentum eigenspace, the labeling of which we leave implicitly subsumed in the particle label $i$. Here $Z_i$ is the particle's central charge. Also of note is that, as a result of the way the massive spinors transform under conjugation, $(q_{I,A})^\dagger = -q^{\dagger I,A}$ and $(q^I_A)^\dagger = q^{\dagger A}_I$. As usual, the $SU(2)$ little group indices may be raised and lowered using the Levi-Civita symbol. When the external legs are massless, the supercharges (1.3) become

$$Q_{i,\alpha A} = -\sqrt{2}\,|i]_\alpha\, q_{i,A} \qquad Q^{\dagger A}_{i,\dot\beta} = -\sqrt{2} q^{\dagger A}_i\,\langle i|_{\dot\beta}\,. \tag{1.5}$$

The little group covariant supercharges satisfy the algebra

$$\left\{q_{i,A}, q^{\dagger B}_i\right\} = \delta^B_A \tag{1.6}$$

and the other anticommutators are zero.

The stripping of the helicity spinor effectively exchanges manifest chirality for manifest spin polarisation (of which helicity is often a natural and useful example). For massless states, these are identical and each chiral spinor supercharge can only either raise or lower a state's helicity. However, for massive states, the supercharges in the form

of chiral spinors will do a superposition of both, for the usual reason that chirality and helicity/polarisation are no longer identical. The little group here describes the freedom in choosing a spin direction as a state label, which determines how the chiral spinor supercharges are decomposed into supercharges characterised by polarisation.

In the simple case in which all legs carry a single, electric central charge, $Z_{i,AB} = Z_i\,\Omega_{AB}$, where $Z_i \in \mathbb{R}$ while $\Omega_{AB} = -\Omega_{BA}$ is a symplectic 2-form:

$$\Omega_{AB} = \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix}, \tag{1.7}$$

where 0 is the $\frac{\mathcal{N}}{2} \times \frac{\mathcal{N}}{2}$ zero matrix, while $I$ is the identity of the same size. Specifically for $\mathcal{N} = 2$, $\Omega_{AB} = \epsilon_{AB}$, $Z_i$ may be complex (corresponding to two central charges) and this central extension is general. The supercharge labels $A, B$ give a manifest representation of a symmetry group that acts on $q^I_A$ (and on $q^{\dagger IA}$ in the conjugate representation) while preserving the algebra (2.3). If $Z_{AB} = 0$, this would be $SU(\mathcal{N})$ (or $U(\mathcal{N})$), while for the central charge considered above, this would be broken to $USp(\mathcal{N})$. The symplectic 2-form $\Omega_{AB}$ may then be used to convert $USp(\mathcal{N})$ tensor representations (such as the supercharges) into conjugate representations (i.e. raise and lower the explicit $R$-indices) in the way that the Levi-Civita tensor does for $SU(2)$.

For $|Z_i| < 2m_i$, the relations (2.3) may be simplified. Unlike for massless particle representations, the generators $q_{I,A}$ and their conjugates $q^{\dagger I,A}$ may mix because their index heights may be changed by $\epsilon_{IJ}$ and $Z_{AB}$. This allows for a rotation into a basis that canonicalises the anticommutators. This basis is given by

$$\bar{q}^I_{i,A} = \frac{1}{\sqrt{D}} \left( q^I_{i,A} + \left(\frac{2m_i}{|Z_i|}\right)^2 \left( 1 - \sqrt{1 - \left(\frac{|Z_i|}{2m_i}\right)^2} \right) \frac{Z_{i,AB}}{2m_i} q_i^{\dagger B,I} \right). \tag{1.8}$$

where $D = 2\Big(\Big(\frac{2m_i}{|Z_i|}\Big)^2 - 1\Big)\Big(1 - \sqrt{1 - \Big(\frac{|Z_i|}{2m_i}\Big)^2}\Big)$. The $\bar{q}_{I,A}$ and their conjugates then satisfy the anticommutation relations without a central charge:

$$\left\{\bar{q}^I_{i,A}, \bar{q}^{\dagger J,B}_i\right\} = -\epsilon^{IJ}\delta^B_A, \qquad \left\{\bar{q}^I_{i,A}, \bar{q}^J_{i,B}\right\} = 0, \qquad \left\{\bar{q}^{\dagger I,A}_i, \bar{q}^{\dagger J,B}_i\right\} = 0. \tag{1.9}$$

In such cases, representations of the supersymmetry algebra may be constructed with a structure identical to that of the case with $Z_i = 0$, although such multiplets still carry central charge (and this would still appear in relating $\bar{q}^I_{i,A}$ and $\bar{q}^{\dagger I,A}_i$ to $Q_{\alpha A}$ and $Q^\dagger_{\dot{\beta}B}$ for these states). Henceforth, this redefinition of the particles' supercharges will be implicit in subsequent discussions of SUSY representations with central charges satisfying $|Z_i| < 2m_i$ and the bars on the diagonalised supercharges will be omitted.

The relations (1.9) illustrate the symplectic $R$-symmetry of the massive representation. While (2.3) has a manifest $USp(\mathcal{N})$ $R$-symmetry, because the supercharges can mix with their conjugates while preserving the $SU(2)$ little group symmetry, the full $R$-symmetry group is actually determined by all of the automorphisms that preserve the anticommutation relations (1.9). Grouping the supercharges into a $2\mathcal{N}$ length vector $\mathbf{q}^I_{i,a} = (\bar{q}^I_{i,A}, \bar{q}^{\dagger I,B}_i)$, where $a = A$ for $a \leq \mathcal{N}$ and $a = B + \mathcal{N}$ for $a > \mathcal{N}$, (1.9) may be combined into the relation

$$\left\{\mathbf{q}^I_{i,a}, \mathbf{q}^J_{i,b}\right\} = \epsilon^{IJ}\Omega_{ab}. \tag{1.10}$$

Here $\Omega_{ab}$ is a $2\mathcal{N} \times 2\mathcal{N}$ symplectic 2-form. Thus the anticommutator is effectively itself a symplectic 2-form and the $R$-symmetry is enhanced to $USp(2\mathcal{N})$ (12). However, it is often broken by interactions. The enlarged $R$-symmetry does not occur for massless representations of the SUSY algebra because the non-zero supercharges have definite opposite helicity and cannot mix.

The case $|Z_i| = 2m_i$ is the special BPS limit. This typically occurs for elementary

particles which obtain mass through Higgsing of a vector multiplet (13). The redefinition of supercharges that give the canonical anticommutation relations described above fails in the BPS limit. This is because, for these representations, half of the number of supercharges are eliminated through the reality constraint

$$q_{i,IA} = \frac{-1}{2m_i} Z_{i,AB} q_{i,I}^{\dagger B}. \tag{1.11}$$

The phase of $Z_i$ may be absorbed into a redefinition of the supercharge. Calling this time $\bar{q}_{i,IA} = q_{i,IA} e^{-i(\arg Z)/2}$, the BPS condition reduces to

$$\bar{q}_i^{IA} = -\bar{q}_i^{\dagger IA} \qquad \bar{q}_{i,IA} = -\bar{q}_{i,IA}^{\dagger}. \tag{1.12}$$

This condition again preserves the supersymmetry algebra. Clearly, BPS states are annihilated by the combination $\bar{q}_i^{IA} + \bar{q}_i^{\dagger IA}$. For the central charge considered above with $Z_{AB} \propto \Omega_{AB}$, the multiplet is 1/2-BPS as it is annihilated by half of the supercharges. Configurations with multiple central charges are also possible in which some smaller fraction of supercharges annihilate the state.

The explicit $SU(\mathcal{N})$ symmetry of the SUSY algebra, which is broken to $USp(\mathcal{N})$ by the central charge of these massive single particle states, is therefore the massive $R$-symmetry group expected for a theory with half of the number of supersymmetries. A 1/2-BPS state in $\mathcal{N}$-SUSY may be represented as a massive non-BPS state of $\mathcal{N}/2$-SUSY. For example, for the simplest spontaneous symmetry breaking pattern in $\mathcal{N} = 4$ SYM, the massless $SU(4)$ $R$-symmetry is broken to $USp(4)$ when the central charge is generated. As the former is unbroken by dynamics and imposes stringent selection rules on scattering amplitudes at the origin of the moduli space, the latter should also be respected by the dynamics and organise the transition matrix structure away from the

origin. See (14) for further discussion. Further elaboration upon the representation of BPS states in scattering amplitudes has been recently made in (15).

More generally, with more complicated configurations of active central charges than the simple case discussed above, for each 1/2-BPS leg there is nevertheless an $SU(\mathcal{N})$ $R$-basis in which the central charge can be rotated into the form $Z_{i,AB} \propto \Omega_{AB}$. In such a basis, the representation of the leg's supercharges is just as described. However, as this basis is different for each leg, the linear combinations of supercharges that annihilate each state may differ by a $SU(\mathcal{N})$ rotation matrix, which must be accounted for when adding together the total supercharges. The $R$-symmetry group will also be broken further beyond $USp(\mathcal{N})$, but this will still be a symmetry restricted to the algebra of a single leg's supercharges.

Finally, the BPS bound itself, $|Z_i| \leq 2m_i$, may be derived for these scattering states from the fact that the operator,

$$\left(q_{i,IA} + \frac{1}{2m_i} Z_{i,AB} q_{i,I}^{\dagger B}\right) \left(q_{i,IA} + \frac{1}{2m_i} Z_{i,AC} q_{i,I}^{\dagger C}\right)^{\dagger} \tag{1.13}$$

being a sum of squares, must have non-negative spectrum. Using the algebra (2.3), (1.13) simplifies to $q_{i,IA} \left(q_{i,IA}\right)^{\dagger} \left(1 - \frac{|Z_i|^2}{(2m_i)^2}\right)$. The BPS bound follows by simply requiring that this be non-negative.

## 1.3 On-Shell Supermultiplets

We seek here to construct scattering amplitudes for supersymmetric theories, so need to understand the structure of supersymmetric scattering states. Scattering data is simplified considerably by the grouping of component states into coherent states of the supersymmetry algebra, known as 'on-shell superfields'. For these scattering states we

may describe their collective $S$-matrix entries using superamplitudes, which manifest both the supersymmetric Ward identities in a simple manner.

For massless theories, on-shell superfields have been established as a convenient organisation of the representations (16). The on-shell superspace was first conceived of in (17) for $\mathcal{N} = 4$ super Yang-Mills (SYM) and was employed later in (18) to formulate the supertwistor space representation of tree-level scattering amplitudes in these theories (it is worth noting that an off-shell superspace formulation of $\mathcal{N} = 4$ SYM does not yet exist). In particular, for $\mathcal{N} = 4$ SYM, it makes transparent the classification of the amplitudes into sectors of a fixed order of helicity violation, which close under both supersymmetry and $R$-symmetry (19–21). This enabled the formulation of the super-BCFW shift (19; 21) and the subsequent construction of all tree amplitudes and loop-level integrands in the limit of large gauge group dimension (22–24), as well as the elucidation of the dual superconformal symmetry and dual twistor representations of amplitudes on complex projective space (20). See (8) for a review of these topics. Amplitudes in theories with fewer supersymmetries have also been formulated in an on-shell superspace in (16), where on-shell superfields for $\mathcal{N} < 4$ massless theories were constructed. We refer the reader to these papers and the review for details of the construction of superamplitudes for massless theories, and now turn to the construction of massive supermultiplets.

General on-shell superspaces for massive states have been previously developed in (9). However, the manifestation of the massive little group for the external legs will allow us to improve upon the presentation of this exposition, as well as providing flexibility to choose a spinor basis best suited for the study of particular phenomena, such as high energy limits or complex momentum shifts. Much of the subsequent discussion here will parallel that of (9), with the improved organisation offered by the little group. This has been utilised more recently for $\mathcal{N} = 4$ super-Yang-Mills in (25) and will be elaborated upon in this context in (14).

From (1.9), the massive supersymmetry algebra is that of $N$ fermionic oscillators, where $N = 2\mathcal{N}$ if the representation is not BPS, but can be reduced by up to a factor of $1/2$ if shortened. Supermultiplets may be represented as coherent states which are eigenstates of $N$ 'lowering operators'. To build these states we introduce Grassmann variables which transform in the little group of each particle $\eta_{i,I}^{A}$, as well as their conjugates $\eta_{i,A}^{\dagger I}$. The labels here match those of the supercharges $q_{i,A}^{I}$. In this section, we will restrict our attention to multiplets which are not BPS, in which case the oscillator index may be identified with a supercharge $R$-index ('$A$' in the symbols just introduced). Following the conventions of (8), all particles will be represented as outgoing scattering states. We will reverse the heights of $R$-indices relative to this reference.

To ensure little group and $R$-covariance, either all of the $q_{i,IA}$ or all of the $q_{i,I}^{\dagger A}$ will be chosen as the lowering operators. These will have some Clifford vacuum states, $\langle\Omega|$ and $\langle\overline{\Omega}|$, which are annihilated by either set. Generally, any linear combination of $q_{i,IA}$ and $q_{i,I}^{\dagger A}$ for each such pair can be chosen as annihilation operators, the choice of which corresponds to the selection of a particular state in the multiplet as the Clifford vacuum in the superfield representation, but a choice that yields the most manifest symmetries is arguably desirable.

An entire supermultiplet may be encoded as a coherent state

$$\langle\eta_i| = \langle\Omega|\, e^{q_{i,A}^{I}\eta_{i,I}^{A}} \tag{1.14}$$

where $\eta_{i,I}^{A}$ are anticommuting Grassmann algebra generators. As is clear in this definition and will be made manifest below, the entire superfield transforms coherently under little group transformations with the same little group weight as the Clifford vacuum. The action of the supercharges on the states generalizes the action for massless particles described in (8), where little group and $R$-indices of the supercharge must be tensored

together and then decomposed. These are eigenstates of the annihilation operators, satisfying $\langle \eta_i | \, q_I^{\dagger A} = \langle \eta_i | \, (-\eta_{i,I}^A)$. The Grassmann Fourier transform may be used to define a basis of conjugate states. It is defined with its inverse respectively as:

$$\left\langle \eta^\dagger \right| = \int \mathrm{d}^{2\mathcal{N}}\eta \; e^{\eta_I^A \eta_A^{\dagger I}} \langle \eta | \qquad \langle \eta | = \int \mathrm{d}^{2\mathcal{N}}\eta^\dagger \; e^{\eta_A^{\dagger I} \eta_I^A} \left\langle \eta^\dagger \right| \tag{1.15}$$

The fact that both the two different $\eta$ and $\eta^\dagger$ representations for the same supermultiplet exist and are related by the Fourier transform will be useful in constraining the form of superamplitudes.

In the $\eta$ basis, the supercharges act as (assuming for simplicity the absence of central charges)

$$\langle \eta_i | \left\langle \theta_A Q^{\dagger A} \right\rangle = -\sqrt{2} \left\langle \theta_A i^I \right\rangle \eta_I^A \langle \eta | \qquad \langle \eta_i | \left[ \theta^A Q_A \right] = \sqrt{2} \left[ \theta^A i_I \right] \frac{\partial}{\partial \eta_I^A} \langle \eta | \tag{1.16}$$

where small $|\theta_A\rangle$ and $\left|\theta^A\right]$ parameterise a linearised supersymmetry transformation. The supercharges may therefore be represented as linear operators on the superamplitudes

$$Q^{\dagger A} = -\sqrt{2} \sum_i \left| i^I \right\rangle \eta_{i,I}^A \qquad Q_A = \sqrt{2} \sum_i |i_I] \frac{\partial}{\partial \eta_{i,I}^A} \tag{1.17}$$

or on individual legs:

$$q_{i,I}^{\dagger A} = -\eta_{i,I}^A \qquad q_{i,A}^I = -\frac{\partial}{\partial \eta_{i,I}^A}. \tag{1.18}$$

Supersymmetry transformations of both types act simply on these coherent states:

$$\langle \eta | \, e^{i\xi_A^{\dagger I} q_{i,I}^{\dagger A}} = e^{-i\xi_A^{\dagger I} \eta_I^A} \langle \eta | \, , \qquad \langle \eta | \, e^{-i\xi_I^A q_{i,A}^I} = \langle \eta + i\xi | \, . \tag{1.19}$$

Here, $\xi_I^A = \left[\theta^A i_I\right]$ and $\xi_A^{\dagger I} = \left\langle \theta_A i^I \right\rangle$ parameterise the supersymmetry transformation projected onto the spinors of leg $i$ of the appropriate chirality. The action of the supercharges encoded in (2.8) give the supersymmetric Ward identities relating the components.

As established in Appendix 1.A.1, massless limits are most naturally taken in the helicity basis for the massive little group (in which momentum is chosen as the quantisation axis). We will adopt the convention that this frame is chosen unless specified otherwise, so that little group indices always denote helicity by default.

By construction, the massive on-shell superfields do not depend upon a preferred frame of reference. However, as a result, the difference between massless and massive representations of the supersymmetry algebra is firmly ingrained in the formalism, as the massive (non-BPS) states non-trivially represent a larger algebra. In the massless limit, following the rules established in (1.A.1), the form of the spacetime supercharges (1.17) requires that the massive Grassmann variables are mapped onto the massless ones as $\eta_{i,-} \to \eta_i$. Here $\eta_i$ is the massless Grassmann variable used to construct the massless on shell superfields (as in e.g. (16)). For reference, massless coherent states are defined here as

$$\langle \eta_i | = \langle \Omega | \, e^{q_{i,A}\eta_i^A} \tag{1.20}$$

$$\left\langle \eta_i^\dagger \right| = \int d^{\mathcal{N}}\eta e^{\eta^A \eta_A^\dagger} \langle \eta | \, . \tag{1.21}$$

Massless analogues of the previous formulae may be obtained similarly.

However, for non-BPS states, the massless limit of the spacetime supercharges (1.17) reduces the number of supercharges represented on the multiplet in half, leaving the definitions of the spinor-stripped supercharges $q_{i,+}^{\dagger A}$ and $q_{i,A}^{+}$ ambiguous. As a consequence, expressions obtained upon taking the massless limit directly on coherent states or their matrix entries will involve a residual Grassmann variables denoted here as $\eta_{i,+} \to \hat{\eta}_i$.

This does not represent the action of a supercharge, but does delineate a division of the massive superfield into separate massless representations.

### 1.3.1 Superfields

The coherent state construction generically gives component fields in reducible representations of the little group and $R$-symmetries, which need to be disentangled to locate the field content. The structure of these vary significantly with the number of supersymmetries.

We consider first the simple case of $\mathcal{N} = 1$. The states in the multiplet are generated by acting $q^I$ on the Clifford vacuum and then decomposing the resulting little group tensors into irreducible representations, which will be further constrained by the needed fermionic antisymmetry of the supercharges. Choosing the Clifford vacuum to be a scalar $\phi = \langle\Omega|$, the resulting states are then $\chi^I = -\langle\Omega| q^I$ and $\tilde{\phi}^{IJ} = -\langle\Omega| q^I q^J = \epsilon^{IJ} \langle\Omega| \frac{-1}{2} q_K q^K$. Because the tensors are antisymmetric, the state $\tilde{\phi}^{IJ} = -\epsilon^{IJ}\tilde{\phi}$ is decomposed into a single scalar degree of freedom. The states of the chiral supermultiplet may therefore be arranged into coherent state

$$\Phi = \phi + \eta_I \chi^I - \frac{1}{2}\eta_I \eta^I \tilde{\phi} \tag{1.22}$$

All states in the multiplet must have identical internal quantum numbers (except for possible $U(1)_R$ charges). If the multiplet is self-conjugate, then the fermion is Majorana and the scalars are permitted to have opposite $R$-charges. Otherwise (as is necessary if the field is in a complex representation, like a quark in superQCD), an anti-superfield is required with conjugate internal quantum numbers which may be constructed similarly.

Component fields can be extracted from the full superfield via Grassmann derivations.

In this simple case we have the mapping

$$\phi = \Phi\big|_{\eta_I=0}, \qquad \chi^I = \frac{\partial}{\partial\eta_I}\Phi\big|_{\eta_I=0}, \qquad \tilde{\phi} = \frac{1}{2}\frac{\partial}{\partial\eta_I}\frac{\partial}{\partial\eta^I}\Phi\big|_{\eta_I=0} \tag{1.23}$$

which generalizes straightforwardly to other theories. By the equivalence of Grassmann differentiation and integration, the derivatives may be replaced by integration. The Grassmann differential operators above can be used to extract component amplitudes in the usual way as for massless superamplitudes.

In the massless limit, the superfield decomposes into components that may be described by opposite helicities:

$$\Phi \to \Phi^+\hat{\eta} + \Phi^- \tag{1.24}$$

with

$$\Phi^+ = \chi^+ + \eta\tilde{\phi}, \qquad \Phi^- = \phi + \eta\chi^-. \tag{1.25}$$

The limit is taken by simply replacing $\eta_- \to \eta$ and $\eta_+ \to \hat{\eta}$ in (1.22). Here $\eta$ is the Grassmann number that would represent the supercharge that acts non-trivially on the massless multiplet, while $\hat{\eta}$ is the variable corresponding to the trivially-acting component.

Similarly to the extraction of component states above, each resulting massless superfield may be extracted by either setting $\hat{\eta} = 0$ ($\Phi^-$ in this example) or differentiating with $\frac{\partial}{\partial\hat{\eta}}$ ($-\Phi^+$ here). This likewise allows for the extraction of massless superamplitudes from limits of massive ones.

We can next construct a vector superfield by starting with a fermionic Clifford vacuum. Because the two spin components of the fermion belong to different supermultiplets

(that is, the vector multiplet does not contain its $CPT$-conjugate states), a little group covariant representation necessitates that two multiplets be combined to create an on-shell superfield that itself transforms in a non-trivial representation of the $SU(2)$ little group, where each multiplet contains states of opposite spin projections. Here, this amounts to combining two Clifford vacua into an $SU(2)$ fundamental representation to describe the two polarisation states of a fermion's degrees of freedom. The superfield is

$$\mathcal{W}^I = \lambda^I + \eta^I H + \eta_J W^{(IJ)} - \frac{1}{2}\eta_J \eta^J \tilde{\lambda}^I, \tag{1.26}$$

where the components have already been decomposed to give the spin-1/2 fermion highest-level state $\tilde{\lambda}$, while we have both a real scalar $H$ and a massive vector $W^{(IJ)}$ at the first level. We can extract the different irreducible representations of the little group via $\frac{1}{2}\frac{\partial}{\partial\eta^I}\mathcal{W}^I = H$, and $\frac{1}{2}\left(\frac{\partial}{\partial\eta_J}\mathcal{W}^I + \frac{\partial}{\partial\eta_I}\mathcal{W}^J\right) = W^{(IJ)}$.

Taking the massless limit again, the massive vector supermultiplet decomposes into the two helicity components of a massless vector superfield and those of a massless chiral superfield as

$$\begin{aligned}
&\mathcal{W}^+ \to G^+\hat{\eta} + \Phi^+ && \mathcal{W}^- \to G^- + \Phi^-\hat{\eta}, \\
&G^+ = g^+ + \eta\tilde{\lambda}^+ && G^- = \lambda^- + \eta g^-, \\
&\Phi^+ = \lambda^+ + \eta\Big(\frac{1}{\sqrt{2}}W^L + H\Big) && \Phi^- = \Big(\frac{1}{\sqrt{2}}W^L - H\Big) + \eta\tilde{\lambda}^-
\end{aligned} \tag{1.27}$$

The longitudinally polarised vector, $W^L = \sqrt{2}W^{(+-)}$, combines with the scalar $H$ to give the two real scalar degrees of freedom in the massless chiral superfields.

For $\mathcal{N} = 2$ without a central extension, we essentially just have two copies of the $\mathcal{N} = 1$ superalgebra. There is only one supermultiplet to construct here, namely that which starts with a scalar Clifford vacuum, as any other choice takes us into supergravity.

The other familiar $\mathcal{N} = 2$ supermultiplets are short multiplets. Expanding the coherent state and keeping the $R$-indices gives the superfield

$$\Omega = \phi + \eta_I^A \psi_A^I - \frac{1}{2}\eta_I^A \eta_J^B (\epsilon^{IJ}\phi_{(AB)} + \epsilon_{AB} W^{(IJ)}) + \frac{1}{3}\eta_I^B \eta_{JB} \eta^{JA} \tilde{\psi}_A^I + \eta_1^1 \eta_1^2 \eta_2^1 \eta_2^2 \tilde{\phi}. \quad (1.28)$$

Each term has been decomposed into irreducible little group and $R$ components (remembering that $\epsilon_{AB}$ may be used to raise and lower the $SU(2)$ $R$-indices). The Grassmann order 3 and 4 terms respectively represent a pair of chiral fermions related by $R$-symmetry and a scalar. The fermion $\tilde{\psi}_A^I$ may be extracted by the action of the Grassmann derivatives for each of the Grassmann variables except $\eta_I^A$. Fermion statistics of the Grassmann generators implies that the Grassmann order 2 terms must be symmetric in either little group or $R$ indices, hence the little group triplet vector and $R$-triplet scalars. The scalars $\phi_{(AB)}$ may be extracted by the action of $\frac{1}{2}\frac{\partial}{\partial \eta_I^A}\frac{\partial}{\partial \eta^{IB}}$, while the vectors $W^{(IJ)}$ are extracted by $\frac{1}{2}\frac{\partial}{\partial \eta_I^A}\frac{\partial}{\partial \eta_{JA}}$. This superfield will be discussed further in (14) as a short multiplet in $\mathcal{N} = 4$ super-Yang-Mills.

Of course, higher-spin representations - either fundamental supergravity multiplets or composite superfields - may be constructed using the same methods. For example, a general massive $\mathcal{N} = 1$ superfield $S$ of spin $s$ has the form

$$S^{(I_1 \ldots I_{2s})} = \phi^{(I_1 \ldots I_{2s})} + \eta^{(I_1} \psi^{I_2 \ldots I_{2s})} + \eta_J \Psi^{(J I_1 \ldots I_{2s})} - \frac{1}{2}\eta_J \eta^J \tilde{\phi}^{(I_1 \ldots I_{2s})}, \quad (1.29)$$

where $\psi$, $\phi$, $\tilde{\phi}$ and $\Psi$ are its component states in order of increasing spin. In the massless limit, this decomposes into pairs of separate superfields each containing either one $\phi$ state (with helicity between $-s$ and $s$) or one $\tilde{\phi}$ state (the superfield having helicity between $-s + \frac{1}{2}$ and $s + \frac{1}{2}$). For reference, a massless higher spin superfield with Clifford vacuum

of helicity $h$ is

$$\Sigma^h = \varphi^h + \eta\xi^{h-\frac{1}{2}}, \tag{1.30}$$

where $\varphi$ and $\xi$ are the component states.

## 1.4 Constructing and Constraining On-Shell Superamplitudes

We wish to write down superamplitudes which package together the scattering data for full representations of the super-Poincaré algebra and allow for amplitudes of component states to be projected out in a simple manner. In this form the SUSY Ward identities will be simply represented. We first discuss general features of superamplitudes in 1.4.1 with a focus on three legs, and then lay out useful strategies for building them in 1.4.2, with a focus on $\mathcal{N} = 1$. In 1.4.3 we discuss the imposition of parity symmetry at the level of the superamplitude. We assume in this section the absence of central charges.

### 1.4.1 SUSY Invariants and the $\eta, \eta^\dagger$ Bases

Invariance under supersymmetry implies that each $n$-leg superamplitude, $\mathcal{A}_n$, must be annihilated by the supercharges. In the $\eta$ basis defined in Section 1.3, the multiplicative action of $Q^\dagger$ implies that $Q^\dagger\mathcal{A}_n = 0$ is solved if and only if $\mathcal{A}_n$ is proportional to the delta function

$$\delta^{(2\mathcal{N})}(Q^\dagger) = \prod_{A=1}^{\mathcal{N}} \left( \sum_{i<j}^{n} \left\langle i^I j^J \right\rangle \eta_{iI}^A \eta_{jJ}^A + \frac{1}{2} \sum_{i}^{n} m_i \eta_{iI}^A \eta_i^{IA} \right). \tag{1.31}$$

A straightforward calculation using momentum conservation shows that this delta function is also invariant under supersymmetry transformations by $Q_{A\alpha}$. However, as these

transformations are not multiplicative, this does not exhaust the constraints from $Q$ transformations.

If we had instead put all of our external states in the $\eta^\dagger$ representation, the $Q$ supercharges would act multiplicatively and the Ward identity $Q\mathcal{A}_n = 0$ would instead imply that the amplitude is proportional to the delta function

$$\delta^{(2\mathcal{N})}(Q) = \prod_{A=1}^{\mathcal{N}} \left( \sum_{i<j}^{n} [i_I j_J]\, \eta_{iA}^{\dagger I} \eta_{jA}^{\dagger J} + \frac{1}{2} \sum_{i}^{n} m_i \eta_{iA}^{\dagger I} \eta_{iIA}^{\dagger} \right). \tag{1.32}$$

The Fourier transform of this delta function, $\widetilde{\delta^{(2\mathcal{N})}(Q)} = \int \prod_{i=1}^{n} \mathrm{d}^{2\mathcal{N}} \eta_i^\dagger e^{-\eta_{iI}^A \eta_{iA}^{\dagger I}} \delta^{(2\mathcal{N})}(Q)$, is also a supersymmetric invariant in the $\eta$ basis, as can be seen by commuting $Q, Q^\dagger$ through the exponential. For amplitudes with massive particles, including three-leg amplitudes, this Fourier transformed delta function is always of degree at least as large as $\delta^{(2\mathcal{N})}(Q^\dagger)$.

Exceptions do exist in situations involving three-particle superamplitudes between BPS states in theories with extended supersymmetry. This will be elaborated upon further in (14), but we will merely comment here that, in these cases, some subset of the supercharges degenerate. The supersymmetric invariant in will instead be the product of all of the distinct supercharges. This is similar to the case of three massless particles, where, for example, special kinematics can imply that if the square brackets are nonvanishing, then $\left\langle Q^{\dagger A} Q^{\dagger A} \right\rangle = 0$. The supersymmetric invariant must instead be taken as $\prod_{A=1}^{\mathcal{N}} \frac{[23]}{\langle q1 \rangle} \left\langle q Q^{\dagger A} \right\rangle = \prod_{A=1}^{\mathcal{N}} \left( [23]\, \eta_1^A + [31]\, \eta_2^A + [12]\, \eta_3^A \right)$, with $|q\rangle$ a reference spinor satisfying $\langle qi \rangle \neq 0$ for all $|i\rangle$, which matches $\widetilde{\delta^{(2\mathcal{N})}(Q)}$.

The existence of the different $\eta$ or $\eta^\dagger$ bases for the same superamplitude yields a restriction on its maximum Grassmann degree from knowledge that the delta functions are the lowest Grassmann degree invariants. This restriction is especially important

for the construction and classification of three-leg superamplitudes. For any number of massive external particles, we can always write a three-leg superamplitude in either basis as

$$\begin{aligned} \mathcal{A}_3|_\eta &= \delta^{(2\mathcal{N})}(Q^\dagger)F(\eta_I) \\ \mathcal{A}_3|_{\eta^\dagger} &= \delta^{(2\mathcal{N})}(Q)\bar{F}(\eta^{\dagger I}) \end{aligned} \tag{1.33}$$

where $F, \bar{F}$ are so far undetermined and are also functions of momentum spinors. Naïvely, these functions could have maximum Grassmann degree $\mathcal{N}(M+3)$, where $M$ is the number of massive legs, since this is the number of independent Grassmann variables we have.

However, from above we know that the Grassmann Fourier transform relates fields in the $\eta$ basis to those in the $\eta^\dagger$ basis and thus such a transform of all legs relates the superamplitude in the two bases. That is, $\widetilde{\mathcal{A}_3|_\eta} = \mathcal{A}_3|_{\eta^\dagger}$. The Grassmann Fourier transform roughly returns the set complement of the $\eta^\dagger$s in the original expression from the total number of $\eta$s (a full discussion of the Grassmann Fourier transform may be found in Appendix 1.A.2). So we end up with $\left[\widetilde{\mathcal{A}_3|_\eta}\right]_{\eta^\dagger} = \mathcal{N}(M+3) - 2\mathcal{N} - [F]_\eta = \mathcal{N}(M+1) - [F]_\eta$, denoting by $[X]_\eta$ the Grassmann degree in $\eta$ of some polynomial $X$. However, the Grassmann degree of $\mathcal{A}_3|_{\eta^\dagger}$ is at least $2\mathcal{N}$, because this is the minimal Grassmann degree for the SUSY invariant to which it must be proportional. Hence we have the inequality

$$[F]_\eta \leq \mathcal{N}(M-1) \tag{1.34}$$

Of course, the same reasoning holds with $F$ replaced with $\bar{F}$. [1]

[1] As remarked previously, the situation is modified in the case of three massless particles because there is a SUSY invariant with Grassmann degree $\mathcal{N}$. In this case, SUSY directly implies that the only possible Grassmann structures are $\delta^{(2\mathcal{N})}(Q^\dagger)$ and $\widetilde{\delta^{(2\mathcal{N})}(Q)}$. The case in which the particles are BPS is also exceptional and will be explained in (14).

This simplifies our task of constructing general three-leg superamplitudes as we need only understand the structure of appropriately invariant functions of Grassmann degree $2\mathcal{N}$ at most.

### 1.4.2 Strategies for Enumerating Amplitudes without Central Charges

The main goal will be to construct three-leg superamplitudes in all simple supersymmetric theories with spins $\leq 1$. We presently discuss the procedure in brief and outline a number of simplifications.

Now that we only have a small number of Grassmann orders to worry about at most, our task will be to construct the function $F$ which multiplies the SUSY invariant delta function for various theories. This function is constrained by the little group covariance of the amplitude, which is set by the external legs as in the $\mathcal{N} = 0$ case. Supersymmetrically, it is constrained by the Ward identities, since half of the supercharges act derivatively. An important benefit of our representation of the supercharges (1.17) is that they are of uniform degree in $\eta$ and consequently these constraints do not mix up different Grassmann orders. This simplifies the procedure so that we may construct the amplitude order by order in $\eta$.

At each order, the $F$ factor consists of a sum of monomials in Grassmann variables. As the delta function is little group invariant, each of these terms must carry the little group representations of the superfield legs. The Grassmann variables themselves transform in non-trivial little group representations, so must be combined with coefficients built out of spinors in such a way as to give the necessary representation of the superamplitude. The possible combinations of spinors that satisfy this then correspond to the possible terms that are permitted by supersymmetry and Lorentz invariance. For example, a

superamplitude with three massive spin-1/2 legs will have an $F$ factor with a single spin index for each leg ($F^{I_1 J_2 K_3}$). An example of a candidate term with Grassmann degree 1 is then $c^{I_1 J_2 K_3 M_1}\eta_{1M_1}$, where the Grassmann variable from leg 1 contains a little group index for that leg, while the coefficient's tensor structure is then determined by that of $F^{IJK}$ and $\eta_{1M}$.

Each Grassmann variable carries either a fundamental $SU(2)$ index for a massive leg or a helicity weight of magnitude 1/2 for a massless leg. The rank and helicity weight of the representations of the possible coefficients are determined by the possible combinations of Grassmann monomials with the required little group structure. We define the 'total little group weight' $\mathbf{h}$ of a superamplitude to be

$$\mathbf{h} = \sum_{\text{massive legs}} 2s_i + \sum_{\text{massless legs}} 2|h_i|, \tag{1.35}$$

for spin $s_i$ (helicity $h_i$) of the massive (massless) leg $i$.

Each coefficient of the Grassmann monomials must involve an even number of contracted spinors (as the superamplitude is a Lorentz scalar). This implies that terms with an even number of Grassmann variables cannot arise if $\mathbf{h}$ is odd for the amplitude. Likewise, if the amplitude as a little group tensor has even $\mathbf{h}$, only even Grassmann degree terms are consistent.

The possible tensor coefficients of the Grassmann monomials may be constructed similarly to the way in which $S$-matrix amplitudes are constructed in (6). The coefficients like $c^{I_1 J_2 K_3 M_1}$ above may be expanded in a tensor basis spanned by a massive spinor of our choice for each of the required little group indices. We are then left to construct, for each monomial, an $SL(2,\mathbb{C})$ Lorentz tensor with the correct little group weight for the massless legs and massless Grassmann variables, which we may do by identifying a tensor basis and enumerating the possibilities as done in (6).

A similar procedure to that used in (26) may be used to determine $F$. As $F$ is ambiguous up to the addition of terms $\propto Q^\dagger$ (as these are annihilated by $\delta(Q^\dagger)$), arbitrary linear combinations of this supercharge may be added to simplify the superamplitude. The two components of each supercharge can be used to eliminate two Grassmann variables entirely from $F$ (for example, the two little group components of a particular massive leg). We may then apply the supersymmetry constraint $QF = 0$ to relate the spinor coefficients of different Grassmann monomials to each other.

An exceptional feature appears in the special case of a three-leg amplitude for two massive, equal-mass particles and one massless particle. In this special kinematic configuration, one finds that there is an additional object that can carry the little group weight of the massless particle. Following (6), this is

$$x \equiv \frac{1}{m}\frac{[q|\, p_2\, |3\rangle}{[q3]}, \tag{1.36}$$

where 3 is the massless leg, $m$ is the mass of legs 1 and 2, and $|q]$ is an arbitrary reference spinor defined so that $[q3] \neq 0$. In this unique case, the special kinematics of the legs implies that $p_1 \cdot p_3 = -\langle 3|\, p_1\, |3] = 0$ and so $p_1\, |3] \propto |3\rangle$. The constant of proportionality is $x$, which, as a $SL(2,\mathbb{C})$ scalar, nevertheless carries helicity weight 1 of leg 3. In no other kinematic configuration of massive legs in a 3-particle amplitude does such an alignment of massless spinors occur in which the relative orientation is described by a single scalar. This is the reason that (2.23) is independent of the reference spinor, despite its necessary appearance in inverting $\frac{1}{m} p_1\, |3] = x\, |3\rangle$, and also the reason that such a helicity-weight carrying scalar object doesn't exist in other kinematic configurations.

With this general method established, we turn to the construction of elementary amplitudes in simple SUSY theories, after first digressing to discuss parity.

### 1.4.3 Parity

While not obligatory, many theories exhibit parity ($P$) symmetry. We here explain how this acts upon on-shell superfields, from which relations between superamplitudes in a parity-invariant theory may be deduced. Details about the construction and spin quantisation of spinors can be found in Appendix C of (27).

As for general chiral spinors, parity acts on the super-Poincare group as (28)

$$PP_\mu P^\dagger = \mathcal{P}_\mu^\nu P_\nu \qquad PQ_\alpha P^\dagger = iQ^{\dagger\dot{\alpha}} \qquad PQ^\dagger_{\dot{\alpha}} P^\dagger = -iQ^\alpha. \tag{1.37}$$

where $\mathcal{P}_\mu^\nu = \text{diag}(1, -1, -1, -1)$.

The action of parity on the coherent states may be determined by its action on the Clifford vacuum and on the spinor-stripped supercharges $q$, defined in (2.2). It is important to remember here that these have been implicitly defined with restriction to a particular momentum eigenspace. The operators $q_i$ and $q_i^\dagger$, through their particle labels, implicitly also carry momentum labels. Under the action of parity, they are mapped to their representations on different momentum eigenspaces.

For massless legs, noting that

$$\begin{aligned} |\mathcal{P}p\rangle &= -e^{i\varphi}\,|p] \qquad |\mathcal{P}p] = e^{-i\varphi}\,|p\rangle \\ \langle\mathcal{P}p| &= e^{i\varphi}\,[p| \qquad [\mathcal{P}p| = -e^{-i\varphi}\,\langle p| \end{aligned} \tag{1.38}$$

for a phase $\varphi$, the action of $P$ on the supercharges $q_i$ and $q_i^\dagger$ is derived from (1.37) to be

$$Pq_iP^\dagger = -ie^{i\varphi}q^\dagger_{\mathcal{P}i} \qquad Pq_i^\dagger P^\dagger = ie^{-i\varphi}q_{\mathcal{P}i}. \tag{1.39}$$

Here, $\mathcal{P}i$ denotes leg $i$ with inverted 3-momentum. Note that helicity spinors are defined

up to a convention-dependent, arbitrary overall phase, which must be implicitly made in the definition of the spinor-stripped supercharge. This effectively determines an arbitrary phase multiplying the (complex) Grassmann variables $\eta^A$ in the coherent states, which, as will be shown below, can be defined to absorb these factors in the parity-conjugate superfield.

The existence and action of $P$ is a model-dependent property. Depending upon the theory, supermultiplets may be self-conjugate or mapped to distinct supermultiplets. Massless spinning particles must be mapped to states with opposite helicity, which are usually part of a distinct supermultiplet. However, because of (1.37), massless scalars and massive spinning particles, at least when selected as Clifford vacua, must also be mapped to states of distinct weight (in the same or possibly different multiplets) for consistency with SUSY. For theories with this property, the action of $P$ on a massless coherent state may be determined as follows. Taking for example the left-handed chiral multiplet $\Phi^-$ in (1.25) and explicitly labeling its 3-momentum $\vec{p}$,

$$\begin{aligned}\Phi^-_{\vec{p}} P^\dagger &= \langle \phi_{\vec{p}}| \, P^\dagger P e^{q_{\vec{p}}\eta_{\vec{p}}} P^\dagger \\ &= \zeta_\phi \left\langle \tilde{\phi}_{-\vec{p}} \right| e^{\eta^\dagger_{-\vec{p}} q^\dagger_{-\vec{p}}} = \zeta_\phi \widetilde{\Phi^+_{-\vec{p}}},\end{aligned} \tag{1.40}$$

calling the Grassmann variable of the parity conjugate coherent state $\eta^\dagger_{-\vec{p}} = ie^{i\varphi}\eta_{\vec{p}}$, absorbing the phase from the transformation of the supercharge. Here, the $\widetilde{\Phi^+}$ denotes Grassmann Fourier transform of the chiral superfield $\Phi^+$ in (1.25) (which, in general, need not have any other necessary relationship with $\Phi^-$). Finally, $\zeta_\phi$ is an intrinsic parity assigned to the scalars (note that the Clifford vacuum is not a parity eigenstate).

Similarly,

$$\Phi^+_{\vec{p}} P^\dagger = \zeta_{\lambda^+} \widetilde{\Phi^-_{-\vec{p}}} \tag{1.41}$$

$$G^+_{\vec{p}} P^\dagger = \zeta_{g^+} \widetilde{G^-_{-\vec{p}}} \tag{1.42}$$

$$G^-_{\vec{p}} P^\dagger = \zeta_{\chi^-} \widetilde{G^+_{-\vec{p}}} \tag{1.43}$$

where the Clifford vacuum for $\Phi^+$ maps under parity as $\left\langle \lambda^+_{\vec{p}} \right| P^\dagger = \zeta_{\lambda^+} \left\langle \lambda^-_{-\vec{p}} \right|$ (and analogously for the other coherent states). The factors of $\zeta_X$ are possible phases associated with intrinsic parity of the Clifford vacuum. For example, in SUSY QED, the action of parity on the photon's multiplet would introduce a factor of $\zeta_{g^+} = -1$ in (1.42) because of the intrinsic parity of the photon.

For massive legs, the null vectors that constitute the little group decomposition of the massive momenta transform in the same way as (1.38) under 3-momentum inversion:

$$\begin{aligned} \left|\mathcal{P}p^I\right] &= \left|p^I\right\rangle \qquad \left|\mathcal{P}p^I\right\rangle = \left|p^I\right] \\ \left[\mathcal{P}p^I\right| &= -\left\langle p^I\right| \qquad \left\langle \mathcal{P}p^I\right| = -\left[p^I\right|. \end{aligned} \tag{1.44}$$

Helicity reverses under parity because, while spin is invariant, the quantisation axis (defined as the 3-momentum) reverses. The massive little group components are expressed with respect to some external quantisation axis, rather than the 3-momentum, so should not change under parity. This is the reason that the phases that accompanied the transformation of the massless helicity spinors (and subsequently the supercharges) do not arise here. The massive supercharges therefore transform as

$$P q^I_i P^\dagger = i q^{\dagger I}_{\mathcal{P}i} \qquad P q^\dagger_{i,I} P^\dagger = i q_{\mathcal{P}i,I}. \tag{1.45}$$

Calling $\eta^{\dagger}_{-\vec{p},I} = i\eta_{\vec{p},I}$, the action of $P$ on a massive chiral superfield is

$$\Phi_{\vec{p}}P^{\dagger} = \zeta_{\phi}\left\langle \tilde{\phi}'_{-\vec{p}}\right| e^{q^{\dagger I}_{-\vec{p}}\eta^{\dagger}_{-\vec{p}I}} = \zeta_{\phi}\widetilde{\Phi'_{-\vec{p}}}, \tag{1.46}$$

where, depending upon other quantum numbers, $\Phi'$ may or may not be equal to $\Phi$. The scalar Clifford vacuum is importantly mapped to the scalar of opposite weight in the other superfield: $\phi \to \tilde{\phi}'$. For a massive vector, the transformation is similar but with fermionic Clifford vacuum mapped to the other fermionic degree of freedom in the multiplet with the same polarisation

$$\mathcal{W}^{I}_{\vec{p}}P^{\dagger} = \zeta_{\chi^I}\widetilde{\mathcal{W}'_{-\vec{p}}}^{\,I}, \tag{1.47}$$

where again $\mathcal{W}'$ may or may not be distinct from $\mathcal{W}$.

Parity invariance of a theory implies equality of the superamplitudes of a set of particles with that of their parity conjugates. Given the results above, this may be stated as

$$\mathcal{A}_n(X_{p_1}, X_{p_2}, \dots X_{p_n}) = \left(\prod_{i=1}^{n}\zeta_{X_i}\int \mathrm{d}^2\eta_{\mathcal{P}i}\; e^{\eta_{\mathcal{P}i,I}\eta^{\dagger I}_{\mathcal{P}i}}\right)\mathcal{A}_n(X^P_{\mathcal{P}p_1}, X^P_{\mathcal{P}p_2}, \dots X^P_{\mathcal{P}p_n}), \tag{1.48}$$

where $X^P$ is the parity conjugate superfield of $X$ (while we have written the Fourier transforms in (1.48) in the form specific for massive coherent states, they should be reinterpreted as their massless analogues for each massless leg). In other words, to relate couplings between superamplitudes in a parity symmetric theory, any superamplitude $\mathcal{A}_n(X_{p_1}, X_{p_2}, \dots X_{p_n})$ must be equal to that obtained by taking the superamplitude of the parity conjugate multiplets, Fourier transforming and then reversing the 3-momenta using (1.38), (1.39), (1.44), (1.45) and the relations between Grassmann variables $\eta^{\dagger}_{-\vec{p},I} = i\eta_{\vec{p},I}$ and $\eta^{\dagger}_{-\vec{p}} = ie^{i\varphi}\eta_{\vec{p}}$. Kinematic-dependent phases appearing in (1.48) from the use of (1.38)

and (1.44) may be dropped, representing arbitrary phases in the polarisations of the external legs.

## 1.5 $\mathcal{N} = 1$ Three-Particle Superamplitudes

In this section we systematically construct the possible three-particle superamplitudes for scattering of $\mathcal{N} = 1$ chiral and vector superfields and identify the types of theories to which they would belong. We furthermore discuss the dependence of the couplings on the masses of the different legs, how they behave in different limits and how they may appear in "tree-unitary" theories (29). We also present some simple results for higher spin multiplets. In Appendix 1.B, we additionally present some well-known results for higher leg amplitudes recast in the little group invariant helicity spinor language.

### 1.5.1 Three Chiral Supermultiplets

We begin with the case of three massive chiral supermultiplets, and will then find the cases with massless chiral supermultiplets via appropriate limits. Our representation of the massive chiral superfield is given in (1.22). This three-point superamplitude has the general form

$$\mathcal{A}(\Phi_1, \Phi_2, \Phi_3) = \delta^{(2)}(Q^\dagger) F(\eta_i^I), \tag{1.49}$$

where, from Section 1.4.1, $F(\eta_i^I)$ is at most a second degree polynomial and, since it has total little group weight $\mathbf{h} = 0$, contains only even orders in $\eta$. Since the Ward identities do not mix different Grassmann orders, we may construct each $i$-th order Grassmann term $F^{(i)}$ separately.

We will illustrate the general procedure by explicitly deriving the superamplitude from first principles using the method described in Section 1.4.2. From little group scaling,

$F^{(0)}$ is fixed to be a constant which we call $\lambda$. The second-order function can be simplified by using the supercharge conservation constraint imposed by the delta function, $Q^\dagger = 0$. We can use this to eliminate any dependence on $\eta_{3I}$, which then leaves us with

$$F^{(2)} = b\left[1^I 2^J\right]\eta_{1I}\eta_{2J} + c\left\langle 1^I 2^J\right\rangle \eta_{1I}\eta_{2J} + d_1\eta_{1I}\eta_1^I + d_2\eta_{2I}\eta_2^I. \tag{1.50}$$

The Ward identity $QF^{(2)} = 0$ is a first-order Grassmann equation and results in two independent spinor equations $(bm_2 - 2d_1)\left|1^I\right] + cp_2\left|1^I\right\rangle = 0$ and similarly with $1 \leftrightarrow 2$. The independent constraints may be found by contracting with $[1_I|$ and $\langle 1_I|\, p_2$, which allows one to solve for $d_1$ in terms of $b$ and find $c = 0$. Along with the same procedure for the other equation, this yields

$$F^{(2)} = b\left(\left[1^I 2^J\right]\eta_{1I}\eta_{2J} + \frac{1}{2}\left(m_2\eta_{1I}\eta_1^I + m_1\eta_{2I}\eta_2^I\right)\right). \tag{1.51}$$

The full superamplitude is then

$$\mathcal{A}(\Phi_1,\Phi_2,\Phi_3) = \delta^{(2)}(Q^\dagger)\left[\lambda + b\left(\left[1^I 2^J\right]\eta_{1I}\eta_{2J} + \frac{1}{2}\left(m_2\eta_{1I}\eta_1^I + m_1\eta_{2I}\eta_2^I\right)\right)\right]. \tag{1.52}$$

When all of the legs are identical the superamplitude can be written in the manifestly exchange symmetric form

$$\mathcal{A}(\Phi_1,\Phi_2,\Phi_3) = \delta^{(2)}(Q^\dagger)\left[\lambda + \frac{b'}{3m}\left(\sum_{i<j}\left[i^I j^J\right]\eta_{iI}\eta_{jJ} + m\sum_i \eta_{iI}\eta_i^I\right)\right]. \tag{1.53}$$

We have here redefined the coupling $b$ to make it dimensionless.

There are three special cases to consider corresponding to the number of different massless legs. Firstly, the massless limit $m_1 \to 0$ may be taken directly on (1.52) to

produce the most general expected superamplitudes

$$\mathcal{A}(\Phi_1^-, \Phi_2, \Phi_3) = \lambda\delta^{(2)}(Q^\dagger) \tag{1.54}$$

$$\mathcal{A}(\Phi_1^+, \Phi_2, \Phi_3) = -b\delta^{(2)}(Q^\dagger)\left(\left[12^J\right]\eta_{2J} + m_2\eta_1\right). \tag{1.55}$$

These expressions are independent of whether $m_2 = m_3$. We have assumed that the coupling $b$ is unaffected by the limit, which is self-consistent.

Similarly, taking the subsequent limit that $m_2 \to 0$ results in the superamplitudes for two massless legs:

$$\mathcal{A}(\Phi_1^-, \Phi_2^-, \Phi_3) = \lambda\delta^{(2)}(Q^\dagger) \qquad \mathcal{A}(\Phi_1^+, \Phi_2^+, \Phi_3) = b\delta^{(2)}(Q^\dagger)\left[12\right], \tag{1.56}$$

while $\mathcal{A}(\Phi_1^\pm, \Phi_2^\mp, \Phi_3) = 0$. It is again being assumed that the couplings present no obstruction to this, which is clearly self-consistent.

In the high energy limit, the superamplitude (1.53) does not diverge with inverse powers of a mass scale because of the special 3-particle kinematics. Note that $[i^+j^+] \sim \mathcal{O}(m^2/E)$ or $\langle i^-j^-\rangle \sim \mathcal{O}(m^2/E)$ as $m \to 0$ for some (complex) energy, depending upon the kinematic configuration that is converged to (individual spinor mass limits can be found in (1.115)). The superamplitude converges to (at leading order in energy)

$$\mathcal{A}(\Phi_1, \Phi_2, \Phi_3) \to \mathcal{A}(\Phi_1^-, \Phi_2^-, \Phi_3^-) - \mathcal{A}(\Phi_1^+, \Phi_2^+, \Phi_3^+)\hat{\eta}_1\hat{\eta}_2\hat{\eta}_3 \tag{1.57}$$

$$\mathcal{A}(\Phi_1^-, \Phi_2^-, \Phi_3^-) = \lambda\delta^{(2)}(Q^\dagger) \tag{1.58}$$

$$\mathcal{A}(\Phi_1^+, \Phi_2^+, \Phi_3^+) = -b'\widetilde{\delta}^{(1)}(Q), \tag{1.59}$$

where $\Phi^\pm$ are the massless superfields in the notation of (1.25). For the latter kinematic

configuration, the delta function is

$$\widetilde{\delta}^{(1)}(Q) = [23]\,\eta_1 + [31]\,\eta_2 + [12]\,\eta_3, \tag{1.60}$$

which is a Grassmann order 1 supersymmetry invariant that is the Fourier transform of $\delta^{(2)}(Q)$ in the $\eta^\dagger$ basis. In the first term of (1.57), $[ij] \to 0$, while in the second, $\langle ij \rangle \to 0$. The $(-)$ sign accompanying the second term arises because the Grassmann variables $\hat{\eta}_i$ must anticommute past the fermionic $\Phi_i^+$ states.

Note that if the limit that all particles are sent massless at the same rate is instead taken, then (1.57) is exact, rather than merely leading. The fully massive superamplitude (1.53) contains helicity violating couplings that, in the high energy limit, scale as mass-dependent constants and cannot be expressed as a massless superamplitude.

This massless limit is to be expected from field theory, where the three scalar component amplitudes contained in the two surviving superamplitudes are expected to vanish in the massless limit according to the superpotential. Also of note is that the massive superamplitudes (1.52) are totally determined by two parity conjugate sets of couplings. That there are no others is not completely obvious from a Lagrangian derivation, where the possibility of spontaneous supersymmetry breaking has to be explicitly checked for a given holomorphic superpotential. Here, constraints from unbroken supersymmetry are more directly applied. It automatically follows that candidate holomorphic superpotential terms, such as tadpoles and quartics that would naively give interactions that do not conform to the structures derived here, must induce spontaneous supersymmetry breaking.

The only remaining massless superamplitudes are those of superfields with mixed

helicity. These are determined by symmetries to be (up to a coupling constant)

$$\mathcal{A}(\Phi_1^+,\Phi_2^+,\Phi_3^-)=\delta^{(2)}(Q^\dagger)\frac{1}{\langle 12\rangle}\qquad \mathcal{A}(\Phi_1^+,\Phi_2^-,\Phi_3^-)=\widetilde{\delta}^{(1)}(Q)\frac{1}{[23]}. \tag{1.61}$$

However, these superamplitudes have peculiar locality properties. While non-divergent in the real momentum limit, they are also non-zero, being unsuppressed by helicity conservation (30). These are the supersymmetrisations of the helicity conserving scalar-fermion-fermion 3-leg amplitude found in (31). Consistent factorisation properties of 4-leg amplitudes were used to rule this out. Notably, while consistent with symmetries, they do not appear in the massless or high energy limit of the massive superamplitudes.

The theory of a single chiral supermultiplet has an accidental parity symmetry. This is a model-dependent and needn't be a general property of this three particle superamplitude. However, we take the opportunity to comment that parity may be imposed as described in Section 1.4.3 to relate the two otherwise independent couplings. Ignoring the possible non-trivial intrinsic parity phases, this gives $b'=\lambda$, in agreement with the massless and massive cases. The Wess-Zumino three-leg superamplitude is then

$$\mathcal{A}(\Phi_1,\Phi_2,\Phi_3)=\lambda\delta^{(2)}(Q^\dagger)\left[1-\frac{1}{3m}\left(\sum_{i<j}\left[i^I j^J\right]\eta_{iI}\eta_{jJ}+m\sum_i\eta_{iI}\eta_i^I\right)\right]. \tag{1.62}$$

It would be interesting to find an on-shell condition from which the accidental parity is derived as an outcome. One would have to study higher leg amplitudes in this theory with only a single chiral supermultiplet in order to derive this feature. In this regard, it would also be interesting to find how holomorphy of the superpotential is represented in the $S$-matrix. In the case where all particles are massless, each holomorphic composite operator in the superpotential contributes a contact interaction inducing a (super)amplitude that is holomorphic in helicity. The rest of the $S$-matrix is presumably

then generated by consistent factorisation involving these. Mass mixes states of different helicities, so produces a violation of this pattern of helicities induced by the holomorphic contact interactions. The on-shell superspace significantly clarifies the pattern, the foundations of which were described in (1) at the level of "seed" $MHV$ component amplitudes with the fewest legs.

### 1.5.2 One Massless Vector

We next turn to the case of two chiral supermultiplets interacting with a massless vector multiplet. This includes matter interactions in supersymmetric gauge theories (like superQCD). Because of this, in this section we refer to the chiral supermultiplets as quarks and the vector fields as gluons. Specifically in superQCD, the states of the quark supermultiplets arrange into the following on-shell superfields:

$$\begin{aligned}\mathcal{Q} &= \widetilde{Q}_L + \eta_I Q^I - \frac{1}{2}\eta_I\eta^I\widetilde{Q}_R \\ \overline{\mathcal{Q}} &= \overline{\widetilde{Q}}_L + \eta_I \overline{Q}^I - \frac{1}{2}\eta_I\eta^I\overline{\widetilde{Q}}_R,\end{aligned} \tag{1.63}$$

where $\mathcal{Q}$ are the quark and $\overline{\mathcal{Q}}$ are the antiquark states. The $L$ and $R$ subscripts identify each of the squarks. The arrangement of the states is to be contrasted with the field-theoretic off-shell superfields. However, while we will use the symbols $\mathcal{Q}$ and $\overline{\mathcal{Q}}$ to denote the chiral superfields in what follows, we will not be committing to identifying them with any particular theory beyond what we will find to be possible to construct.

It is easily shown using the methods of Section 1.4.2 that a three-leg superamplitude between two massive chiral multiplets and a massless vector multiplet is impossible unless the chiral multiplets have equal mass. This case is distinguished by the existence of $x$, which will allow for expressions with the required little group scaling to be constructed.

The $G^+$ and $G^-$ superamplitudes have total little group weights $\mathbf{h} = 2$ and $\mathbf{h} = 1$

respectively. The superamplitude for the positive helicity gluon superfield is simplest to construct as little group scaling immediately gives the unique form

$$\mathcal{A}(\overline{\mathcal{Q}}_1, G_2^+, \mathcal{Q}_3) = \delta^{(2)}(Q^\dagger)\frac{g}{x}, \tag{1.64}$$

where $x$ is defined in (2.23) and $g$ is the coupling constant (which may have suppressed dependence upon possible internal quantum numbers of the states). For the negative helicity superamplitude, little group scaling, supersymmetry invariance and the Grassmann counting rule of Section 1.4 determine the superamplitude up to a single coupling constant $b$:

$$\mathcal{A}(\overline{\mathcal{Q}}_1, G_2^-, \mathcal{Q}_3) = \delta^{(2)}(Q^\dagger)bx\left(\eta_2 + \frac{1}{2m}\left(\left[21^I\right]\eta_{1I} + \left[23^I\right]\eta_{3I}\right)\right). \tag{1.65}$$

That the superamplitudes are determined here by a single coupling constant is a reflection of the fact that the anomalous magnetic dipole moment of matter fermions in $\mathcal{N} = 1$ gauge theories is exactly zero (32). Supersymmetry determines the fermionic coupling to the gauge bosons from the scalar coupling, which has only one possible Lorentz structure. As a consequence, the supersymmetry implies that the matter-photon interaction is entirely characterised by the electric charge monopole.

Thus far we have not actually assumed anything beyond particles 1 and 3 having equal mass. However, these superamplitudes are antisymmetric under exchange of the two matter fields $1 \leftrightarrow 3$ (through $x$). As the superfields $\mathcal{Q}$ and $\overline{\mathcal{Q}}$ are bosonic, this implies that they must be distinct (the same argument applies to couplings of matter to massless vectors without supersymmetry as well).

Parity may additionally be imposed. Assuming that the $\mathcal{Q}$ and $\overline{\mathcal{Q}}$ multiplets are both self-conjugate under $P$ (the minimal assumption), this implies that $b = g$. Parity

invariance was an assumption used in the derivation of the Lie algebra structure of the matter couplings from consistent factorisation (6; 33), which is unaffected by the quark masses (with massless matter, $CP$ also suffices, which justifies it for chiral gauge theories). It would be interesting to clarify the role of the discrete symmetry needed to relate the amplitudes on each side of the factorisation channel. In Yang-Mills field theory, this symmetry is accidental. In the examples below, we always find parity emerge in the high energy limit of massive amplitudes, as well as the massless limits of individual legs, in the terms that match onto sensible amplitudes of massless vectors.

It is interesting to note that the $USp(2)$ massive $R$-symmetry of the SUSY algebra is broken in this theory because the gaugino couplings distinguish between the two squark states. The identification of the squarks as $L$ and $R$ is determined by the helicity of the gaugino that couples to them (the squarks are then oppositely charged under the residual unbroken massless $U(1)_R$). This is ultimately a reflection of the breaking of the $USp(2)_R$ by parity symmetry, which distinguishes between the two squarks. This would be restored in an $\mathcal{N} = 2$ gauge theory, where the gauginos are Dirac fermions.

The coupling of higher spin multiplets to photons follows a similar pattern. The superamplitude for the case of the positive helicity massless vector is

$$\mathcal{A}(\overline{S}_1^{(I_1\ldots I_{2s})}, G_2^+, S_3^{(J_1\ldots J_{2s})}) = \delta^{(2)}(Q^\dagger) T^{(I_1\ldots I_{2s})(J_1\ldots J_{2s})}. \tag{1.66}$$

SUSY places no further constraints upon $T^{(I_1\ldots I_{2s})(J_1\ldots J_{2s})}$, which can be constructed as in (6) just as a general amplitude for a photon coupled to a massive spin $s$ state. This implies that the coupling to photons of the spin $s+\frac{1}{2}$ states in the multiplet is determined by that of the spin $s$ state. For a massive particle of spin $s$, there are $2s+1$ such multipoles representing each possible independent Lorentz structure in the coupling. Following (6),

these are

$$T^{(I_1\ldots I_{2s})(J_1\ldots J_{2s})} = \frac{1}{x}\Big(c_0\prod_{i,j=1}^{2s}\big[1^{(I_i}3^{(J_j)}\big] + \frac{c_1}{m}x\prod_{i,j=1}^{2s-1}\big[1^{(I_i}3^{(J_j}\big]\,\big[1^{I_{2s})}2\big]\,\big[3^{J_{2s})}2\big] + \frac{c_2}{m^2}x^2\prod_{i,j=1}^{2s-2}\big[1^{(I_i}3^{(J_j}\big]\prod_{i,j=2s-1}^{2s}\big[1^{I_i)}2\big]\,\big[3^{J_j)}2\big] + \ldots\Big), \quad (1.67)$$

for coupling constants $c_i$. The additional multipole moment for the coupling of the $s+\frac{1}{2}$ state is therefore determined here entirely from the lower multipoles by SUSY. This is the generalisation of the protection of the magnetic dipole moment for supersymmetric matter fermions to higher spin states. We will see another explicit example of this in Section 1.5.5, where the electric quadrupole moment of the massive vector within the spin-half vector superfield is determined by the lower multipoles.

### 1.5.3 One Massive Vector

We next consider the three-leg superamplitude of two massive chiral multiplets and a massive vector multiplet, as may occur in a Higgsed gauge theory. Repeating the procedure as in previous sections, we can reduce the amplitude to

$$\begin{aligned}\mathcal{A}(\mathcal{Q}_1,\mathcal{W}_2^I,\mathcal{Q}_3) &= \delta^{(2)}(Q^\dagger)\Big(\frac{d_1}{m_2}\big\langle 2^I1^J\big\rangle\eta_{1J} - \frac{m_3d_2+m_2d_1}{m_1m_2}\big[2^I1^J\big]\eta_{1J} \\ &\qquad + \frac{d_1}{m_2}\big\langle 2^I3^K\big\rangle\eta_{3K} + \frac{d_2}{m_2}\big[2^I3^K\big]\eta_{3K}\Big) \\ &= \delta^{(2)}(Q^\dagger)\Big(-\frac{d_2}{m_3}\big\langle 2^I1^J\big\rangle\eta_{1J} + \left(\frac{(m_1^2-m_3^2)d_2}{m_1m_3m_2} - \frac{d_1}{m_1}\right)\big[2^I1^J\big]\eta_{1J} \\ &\qquad - d_1\eta_2^I + \frac{d_2}{m_2m_3}\big[2^I\big|\,p_3\,\big|2^J\big\rangle\eta_{2J}\Big). \quad (1.68)\end{aligned}$$

This leaves two undetermined couplings $d_1$ and $d_2$ after imposing supersymmetry invariance. The two forms stated are useful for taking massless limits $m_3 \to 0$ and $m_2 \to 0$

respectively.

Taking the vector massless differs depending on whether the chiral multiplets have equal mass. In the case $m_1 \neq m_3$, one recovers solely the three-chiral superamplitudes (1.54) with $b = d_2'(m_3^2 - m_1^2)/(m_1 m_3) - d_1/m_1$ and $\lambda = d_1$ (this mass scaling has been anticipated in the definition of $d_1$, as well as the assumption that it is non-zero and finite in this limit), where $d_2' = d_2/m_2$ must be finite (and hence must be suppressed by some other mass scale). This is consistent with our finding above that there was no consistent three-leg superamplitude for a massless gluon and two unequal mass chiral multiplets.

More interestingly, if $m_1 \to m_3$ at a rate $|m_1 - m_3| \sim \mathcal{O}(m_2)$ as $m_2 \to 0$, then non-zero superamplitudes involving massless vector multiplets may be recovered if $d_2$ remains a dimensionless constant. The reference spinors that appear in the factors of $x$ do so through the spinor limits in (1.116). This leaves the parity-symmetric terms in the superQCD amplitudes with $b = g = d_2$, as well as the three-chiral superamplitudes mentioned above. As alluded to above, parity in the vector coupling emerges in this special limit.

If we instead take the third leg massless, we find smoothly

$$\mathcal{A}(\mathcal{Q}_1, \mathcal{W}_2^I, \mathcal{Q}_3) \to \mathcal{A}(\mathcal{Q}_1, \mathcal{W}_2^I, \Phi_3^-) + \mathcal{A}(\mathcal{Q}_1, \mathcal{W}_2^I, \Phi_3^+)\hat{\eta}_3 \tag{1.69}$$

$$\mathcal{A}(\mathcal{Q}_1, \mathcal{W}_2^I, \Phi_3^-) = \frac{d_1}{m_1 m_2}\delta^{(2)}(Q^\dagger)\left\langle 32^I\right\rangle\left(\left[31^J\right]\eta_{1J} + m_1\eta_3\right) \tag{1.70}$$

$$\mathcal{A}(\mathcal{Q}_1, \mathcal{W}_2^I, \Phi_3^+) = \frac{d_2}{m_2}\delta^{(2)}(Q^\dagger)\left[2^I 3\right], \tag{1.71}$$

which are alternatively determined purely from symmetries. These expressions hold regardless of whether $m_1 = m_2$ or not. It is being assumed here that $d_1$ and $d_2$ do not vanish or diverge in this limit, which is self-consistent (they may still differ from their counterparts in (1.68) by terms of $\mathcal{O}(m_2)$).

Taking the further $m_1 \to 0$ limit of these superamplitudes requires $d_1 \sim m_1$ in (1.70)

and yields

$$\mathcal{A}(\Phi_1^+, \mathcal{W}_2^I, \Phi_3^-) = \delta^{(2)}(Q^\dagger)\frac{1}{m_2}\left[2^I 1\right], \tag{1.72}$$

where we have omitted the coupling and provided the dependence on mass necessary to realize the final massless limit smoothly (so $d_1 \sim 1/m_2$ and $d_2$ constant). The case in which the chiral multiplets have the same helicity is forbidden by symmetries, so does not appear in the limit. Taking finally $m_2 \to 0$, only the transverse polarisations interact non-trivially (see comments about the superamplitudes of mixed helicity chiral supermultiplets in Section 1.5.1). It is easily verified that $\mathcal{A}(\Phi_1^+, \mathcal{W}_2^+, \Phi_3^-) \to -\mathcal{A}(\Phi_1^+, G_2^+, \Phi_3^-)\hat{\eta}_1$ and $\mathcal{A}(\Phi_1^+, \mathcal{W}_2^-, \Phi_3^-) \to \mathcal{A}(\Phi_1^+, G_2^-, \Phi_3^-)$. This is expected from the Higgs mechanism if the massive vector is coupled to massless matter.

In the high energy limit (taking all masses small simultaneously at the same rate), then it can be verified that

$$\begin{aligned}
\mathcal{A}(\mathcal{Q}_1, \mathcal{W}^+, \mathcal{Q}_3) &\to \mathcal{A}(\Phi^+, G^+, \Phi^-)\hat{\eta}_1\hat{\eta}_2 + \mathcal{A}(\Phi^+, \Phi^+, \Phi^+)\hat{\eta}_1\hat{\eta}_3 - \mathcal{A}(\Phi^-, G^+, \Phi^+)\hat{\eta}_2\hat{\eta}_3 \\
\mathcal{A}(\mathcal{Q}_1, \mathcal{W}^-, \mathcal{Q}_3) &\to -\mathcal{A}(\Phi^+, G^-, \Phi^-)\hat{\eta}_1 + \mathcal{A}(\Phi^-, \Phi^-, \Phi^-)\hat{\eta}_2 + \mathcal{A}(\Phi^-, G^-, \Phi^+)\hat{\eta}_3.
\end{aligned} \tag{1.73}$$

In the $\mathcal{W}^+$ limit, the coupling $d_2$ is the cubic coupling among chiral multiplets, while $d_1$ is the parity conjugate coupling. The couplings of the chiral multiplets to the massless vectors are dermined by linear combinations of these weighted by combinations of the masses.

The possibility of distinct couplings $d_1$ and $d_2$ allows for parity violation in the massive superamplitudes and accounts for the way in which the massive amplitudes can combine together states of different helicities that would otherwise be described as having different interactions. Despite the observation that chiral multiplets coupling to massless vectors must have the same mass, there is not inconsistency with the massive multiplets having

different masses in the high energy limit.

### 1.5.4 Two Vector Superfields

We next turn to three-leg superamplitudes with two vector superfields and one chiral superfield. Starting with the case of one massive leg, we first look at a massive chiral superfield decaying into two massless vectors. The case where the massless decay products are instead both matter fields was addressed in Section 1.5.1, while no consistent superamplitude may be constructed if the massless multiplets are chiral and vector. Only the superamplitudes with massless vector multiplets of the same helicity are non-zero, following from the rules of Section 1.4.2. These are (calling $m$ the nonzero mass)

$$\mathcal{A}(G_1^-, \Phi_2, G_3^-) = \delta^{(2)}(Q^\dagger) a \langle 13 \rangle \qquad \mathcal{A}(G_1^+, \Phi_2, G_3^+) = \delta^{(2)}(Q^\dagger) \frac{b}{m} [31]^2 . \tag{1.74}$$

These superamplitudes would arise, for example, in a theory involving the quantum field couplings $[\Phi \mathcal{W}_A \mathcal{W}_B]_F$, for (off-shell) chiral superfield $\Phi$ and super-Yang-Mills curvatures $\mathcal{W}_{A,B}$ for some Abelian gauge groups (in other words, a massive supersymmetric axion or dilaton-like coupling). Demanding $P$ invariance would imply that $a = b$ (if the massive chiral multiplets in each superamplitude are antiparticles, then the couplings may be related by $CP$ instead). The couplings $a$ and $b$ have the expected inverse mass dimension of an irrelevant interaction. Assuming that, as defined in (1.74), they have no further dependence on the mass of the heavy chiral multiplet, then the massless limit may be taken while holding them fixed (if they instead scale as e.g. $\propto 1/m$, then this would obstruct the limit). This gives

$$\mathcal{A}(G_1^-, \Phi_2^-, G_3^-) = \delta^{(2)}(Q^\dagger) a \langle 13 \rangle \qquad \mathcal{A}(G_1^+, \Phi_2^+, G_3^+) = \widetilde{\delta}^{(1)}(Q) b [13] . \tag{1.75}$$

and the other components are zero.

The superamplitudes for a massive vector multiplet decaying into massless vector and chiral fields may be found similarly. Those that are permitted by the symmetries are (up to coupling constants)

$$\mathcal{A}(\mathcal{W}_1^I, G_2^+, \Phi_3^+) = \delta^{(2)}(Q^\dagger)\frac{1}{m}\left[1^I 2\right][23] \tag{1.76}$$

$$\mathcal{A}(\mathcal{W}_1^I, G_2^-, \Phi_3^-) = \delta^{(2)}(Q^\dagger)\left\langle 1^I 2\right\rangle \tag{1.77}$$

All other helicity combinations are zero. The other allowed decay channel for a massive vector multiplet was found above in (1.72).

The massless limits of the superamplitudes (1.76) and (1.77) converge to the superamplitudes (1.75). Both of these massive and massless superamplitudes may have a common origin, for example in the axionic coupling suggested above, where one of the vectors may become massive through the Higgs mechanism. As in previous cases, the coupling constants for (1.76) and (1.77) may be related by parity.

Finally, we note that it is not possible to find a superamplitude describing the decay of a massive vector multiplet into two massless vector multiplets, which is an expression of the Landau-Yang theorem.

Continuing to the two-massive-leg case, one may construct superamplitudes for massive chiral and vector supermultiplets with a massless vector, which are independent of whether the massive multiplets have the same mass or not:

$$\mathcal{A}(\Phi_1, \mathcal{W}_2^I, G_3^-) = a\delta^{(2)}(Q^\dagger)\left\langle 2^I 3\right\rangle \tag{1.78}$$

$$\mathcal{A}(\Phi_1, \mathcal{W}_2^I, G_3^+) = b\delta^{(2)}(Q^\dagger)\left[2^I 3\right]\left(\left[31^J\right]\eta_{1J} - \frac{m_1}{m_2}\left[32^K\right]\eta_{2K}\right) \tag{1.79}$$

Taking individual legs massless, one recovers solely those amplitudes already remarked

on above. The high energy behaviour of these superamplitudes is poor, scaling inversely with some mass scale contained within the couplings $a$ and $b$.

Next, the superamplitudes for two massive vector multiplets and one massless chiral multiplet may be similarly determined. Using the definitions of the massless chiral supermultiplets in (1.25), the usual arguments determine the three-particle superamplitudes to be

$$\mathcal{A}(\mathcal{W}_1^I, \Phi_2^+, \mathcal{W}'^J_3) = \delta^{(2)}(Q^\dagger) F_{1+}^{IJ} \left[ \frac{1}{m_1} \left[ 21^K \right] \eta_{1K} + \eta_2 \right] \tag{1.80}$$

$$\mathcal{A}(\mathcal{W}_1^I, \Phi_2^-, \mathcal{W}'^J_3) = \delta^{(2)}(Q^\dagger) F_{1-}^{IJ}, \tag{1.81}$$

where

$$F_{1\pm}^{IJ} = d_1^{(\pm)} \left\langle 1^I 3^J \right\rangle + d_2^{(\pm)} \left[ 1^I 3^J \right]. \tag{1.82}$$

These are again independent of whether the massive legs have equal mass or not. Taking the massless limit of the first leg, the coefficients in the $\Phi^-$ superamplitude, $d_i^-$, should have no mass dependence in order to smoothly match onto amplitudes (1.77) and (1.72). For the $\Phi^+$ superamplitude, both coefficients must scale as $d_i^{(+)} \sim m_1$ to return to (1.76) and (1.72). The couplings in both cases must be suppressed by a higher mass scale. Taking the third leg massless instead, the expected limits are obtained only if $d_i^{(+)} \sim 1/m_3$, so altogether $d_i^{(+)} \sim m_1/m_3$ to leading order in $m_1$ and $m_3$ if the limits are to be both non-trivial and unobstructed. However, in either of these cases, the resulting superamplitudes must be suppressed by other mass scales and, in this sense, are "effective". In contrast, taking both legs massless simultaneously is possible without introducing new mass scales. In this respect, these superamplitudes are merely a special example of the case in which the chiral multiplet is also massive, which will be explained

next.

Finally, the all-massive superamplitude for two vectors and a chiral multiplet is determined to be

$$\mathcal{A}(\mathcal{W}_1^I, \Phi_2, \mathcal{W}_3'^K) = \delta^{(2)}(Q^\dagger)\left(F_{(0)}^{IK} + F_{(2)}^{IK}\right), \tag{1.83}$$

where

$$F_{(0)}^{IK} = a\left\langle 1^I 3^K\right\rangle + a'\left[1^I 3^K\right] \tag{1.84}$$

$$F_{(2)}^{IK} = \left(b'\left\langle 1^I 3^K\right\rangle + b\left[1^I 3^K\right]\right)\left[\left[1^M 2^J\right]\eta_{1M}\eta_{2J} + \frac{1}{2}\left(m_2\eta_{1L}\eta_1^L + m_1\eta_{2N}\eta_2^N\right)\right]. \tag{1.85}$$

In the limit that the chiral multiplet becomes massless, the coefficients match on to those of (1.80) and (1.81) as $b' \to d_1^{(+)}/m_1$, $b \to d_2^{(+)}/m_1$, $a \to d_1^{(-)}$ and $a' \to d_2^{(-)}$. Making the matter massive does not really affect the structure of the interactions beyond their collection into the single superamplitude. The results from taking a single vector massless instead are similar as for (1.80) and (1.81) and will not be elaborated upon further.

More interesting instead are the high energy limits. The superamplitude (1.83) consists of two parity conjugate pairs of couplings. The couplings $a$ and $b$ represent "effective" couplings (like those of the axion/dilaton mentioned above or loop induced interactions in a perturbative field theory) that must be suppressed by some additional mass scale (and similarly $d_1^-$ and $d_2^+$ in the case with a massless chiral multiplet). On the other hand, $a'$ and $b'$ (or $d_1^+$ and $d_2^-$) correspond to couplings of a Higgs boson to massive vectors, where the Higgs belongs to a chiral multiplet (and is not part of the multiplet eaten by the vectors with the Goldstone boson). This happens when the quartic coupling of the scalar potential originates from the superpotential ("$F$-term").

To illustrate how the superamplitude (1.83) scales in the UV limit, assume that $a' = \bar{a}/v$ and $b' = \bar{b}/v^2$ for some mass scale $v$ of order the leg masses and call constants $c_i = m_i/v$ for leg masses $m_i$. The leading terms in the limit are then

$$\begin{aligned}
&\mathcal{A}(\mathcal{W}_1^+, \Phi_2, \mathcal{W}_3'^+) \to \mathcal{A}(\Phi_1^+, \Phi_2^-, G_3^+)\hat{\eta}_3 - \mathcal{A}(G_1^+, \Phi_2^-, \Phi_1^+)\hat{\eta}_1 - \mathcal{A}(\Phi_1^+, \Phi_2^+, \Phi_1^+)\hat{\eta}_2 \\
&\mathcal{A}(\mathcal{W}_1^+, \Phi_2, \mathcal{W}_3'^-) \to -\mathcal{A}(G_1^+, \Phi_1^+, \Phi_3^-)\hat{\eta}_1\hat{\eta}_2\hat{\eta}_3 + \mathcal{A}(\Phi_1^+, \Phi_2^-, G_3^+)\hat{\eta}_3
\end{aligned} \tag{1.86}$$

and similarly for parity conjugate states. All terms in the first line depend upon the coupling $\bar{a}$ and each term proportional to $\hat{\eta}_i$ is accompanied by a factor of $c_i$. In the second line, the first term depends upon $\bar{b}c_1c_3$, while the second has coupling constant $\bar{a}c_3$. Again, this pattern of couplings reverses for the parity conjugate limits.

However, there are also subleading terms that vanish in the massless limit that cannot be placed into massless superamplitudes. These represent the effective Goldstone boson couplings to the Higgs.

A supersymmetrised version of the argument used in (6) to demonstrate the Higgs mechanism may presumably be made from constructing a four-leg vector superamplitude from demanding consistent factorisation into 3-leg superamplitudes (1.83) on each factorisation channel. Notably, an exceptional case occurs when the Higgs couples to a massive and massless vector boson in a three-particle superamplitude, as in (1.78) and (1.79), which will induce unitarity-violating superamplitudes in the high energy limit in tree-level processes.

### 1.5.5 Massive and Massless Vector Multiplet Interactions

Let us begin with amplitudes with two massive vector superfields and one massless vector superfield, which has two distinct cases of interest. The first is $\mathcal{A}(\mathcal{W}^I, G, \overline{\mathcal{W}}^J)$, where the two massive vector superfields $\mathcal{W}$ and $\overline{\mathcal{W}}$ have the same mass. This arises

in many examples, such as the adjoint Higgsing of a simple gauge theory by a single vacuum expectation value (vev), which does not feature any 3-leg amplitudes entirely of massive vectors. In this case, the vectors are conjugates, which is the reason for our choice of notation, although we do not need to assume this at this point. The second is $\mathcal{A}(\mathcal{W}^I, G, \mathcal{W}'^J)$, where the two massive states are distinct and of different mass. These can occur, for example, in field theories with generalised Chern-Simons terms (34–36), where at least one of the vectors is Abelian and has a Stuckelberg mass, whereas another of the vectors may be separately Higgsed.

As in the superQCD case above, the positive helicity gluon superfield amplitudes are determined very simply. In these cases one finds

$$\mathcal{A}(\mathcal{W}_1^I, G_2^+, \overline{\mathcal{W}}_3^J) = \delta^{(2)}(Q^\dagger) \left[1^I\right|^\alpha \left[3^J\right|^\beta \left(\frac{g}{mx}\epsilon_{\alpha\beta} + \frac{g-h}{m^2}\left|2\right]_\alpha \left|2\right]_\beta\right) \tag{1.87}$$

$$\mathcal{A}(\mathcal{W}_1^I, G_2^+, \mathcal{W}_3'^J) = \delta^{(2)}(Q^\dagger)\, a \left[1^I 2\right] \left[3^J 2\right], \tag{1.88}$$

where, in both cases, the number of free parameters matches that in the non-supersymmetric amplitude for two massive fermions and one massless vector (6). As in all previous examples, we have here neglected to show that the coupling constants $g$ and $h$ may have internal quantum number structure. In the first case (1.87), the combination of terms with coupling $g$ corresponds to a massive vector 'minimally coupled' to the massless vector. As has been foreseen in the definition of dimensionless couplings in (1.87), in the limit that $m \to 0$ or, equivalently here, at energies $\gg m$, these terms converge to their expected massless counterparts.

The term proportional to $h$ would have the perturbative interpretation of an anomalous magnetic dipole moment for the massive vector (or electric dipole moment if it has a complex phase). This term has poor behavior in the UV limit for certain helicity configurations, which is the reason for the tree-level universality of the magnetic dipole

moment $h = 0$ for elementary particles (37). Note that supersymmetry has set fixed the possible quadrupole structure of the massive vector boson amplitudes that may otherwise exist as a further independent Lorentz structure in the vector boson component amplitude (6; 38). This derivation makes obvious the way that supersymmetry determines the vector amplitudes from their fermionic counterparts.

Finally, (1.87) appears to be symmetric under exchange of particles 1 and 3 ($x \mapsto -x$ under this exchange - see (2.23)). However, because the superfields are fermionic, the vector multiplets must be distinct.

In the second example (1.88), the coupling $a$ has mass dimension $-2$. However, unlike for the minimal coupling terms in the case above, the kinematic factors of the component superamplitudes corresponding to the + helicity states (such as $A(W^+G^+W'^+)$) contain terms that merely scale as $\sim \mathcal{O}(m_i)$ in the massless limit (see equations (1.115) in Appendix 1.A.1 for massless limits of spinors). The amplitude must therefore diverge in the high energy $E$ limit as $E/M$ for some mass scale $M$. Correspondingly, the examples of field theories cited above that feature these amplitudes are only effective up to a UV cut-off.

We can likewise find the negative helicity superamplitude purely from little group covariance and supersymmetry. From the same arguments as in the SQCD case, the Grassmann polynomial must only contain an order-one term.

$$\mathcal{A}(\mathcal{W}_1^I, G_2^-, \mathcal{W}_3^J) = \delta^{(2)}(Q^\dagger) \left[1^I\right|^\alpha \left[3^J\right|^\beta F_{2\alpha\beta} \left(-\frac{1}{m_1}\eta_{1K} \left[1^K 2\right] + \eta_2\right). \tag{1.89}$$

The tensor $F_{2\alpha\beta}$ is then determined from the little group representations of the legs. In

the equal mass case, this gives a superamplitude with two free parameters:

$$\mathcal{A}(\mathcal{W}_1^I, G_2^-, \overline{\mathcal{W}}_3^J) = \delta^{(2)}(Q^\dagger)\left[1^I\right|^\alpha \left[3^J\right|^\beta \left(\frac{g'}{m}x\epsilon_{\alpha\beta} + \frac{h'}{m^2}x^2\,|2]_\alpha\,|2]_\beta\right)\left(-\frac{1}{m}\eta_{1K}\left[1^K 2\right] + \eta_2\right). \tag{1.90}$$

Exchange (anti-)symmetry between $\mathcal{W}$ and $\overline{\mathcal{W}}$ may be manifested by adding terms proportional to $Q^\dagger$ to give

$$\begin{aligned}\mathcal{A}(\mathcal{W}_1^I, G_2^-, \overline{\mathcal{W}}_3^J) = \delta^{(2)}(Q^\dagger)\left[1^I\right|^\alpha \left[3^J\right|^\beta &\left(\frac{g'}{m}x\epsilon_{\alpha\beta} + \frac{h'}{m^2}x^2\,|2]_\alpha\,|2]_\beta\right)\\ &\times\left(-\frac{1}{2m}\eta_{1K}\left[1^K 2\right] + \eta_2 - \frac{1}{2m}\eta_{3K}\left[3^K 2\right]\right).\end{aligned} \tag{1.91}$$

If parity is a symmetry of the theory under consideration, then this relates the superamplitudes of $\mathcal{A}(\mathcal{W}^I, G^-, \overline{\mathcal{W}}^J)$ as discussed in Section 1.4.3. Assuming that the vector multiplets are self-conjugate, this requires that $g' = g$ and $h' = h$.

For the case where $m_1 \neq m_3$, the only option which has the correct scaling is $F_{2\alpha\beta} = b(p_3\,|2\rangle)_\alpha(p_3\,|2\rangle)_\beta$. Our amplitude in this case is

$$\begin{aligned}\mathcal{A}(\mathcal{W}_1^I, G_2^-, \mathcal{W}_3'^J) &= \delta^{(2)}(Q^\dagger)b'\left[1^I\right|p_3\,|2\rangle\left[3^J\right|p_3\,|2\rangle\left(-\frac{1}{m_1}\eta_{1K}\left[1^K 2\right] + \eta_2\right)\\ &= \delta^{(2)}(Q^\dagger)b\left\langle 1^I 2\right\rangle\left\langle 3^J 2\right\rangle\left(-\frac{1}{m_1}\eta_{1K}\left[1^K 2\right] + \eta_2\right)\end{aligned} \tag{1.92}$$

where the coupling $b$ has been redefined in the second line to absorb some factors of mass. If parity is a symmetry of this theory, then one finds $b = am_1/m_3$.

In the massless limit, the superfields are expected to decompose as shown in (1.27). In anticipation of the superamplitudes of the massless components being matched onto by the massless limit of the massive superamplitude, we first determine these directly from symmetries. The constraints of complex three-particle special kinematics, little group scaling and 'locality', in the sense that the three-particle amplitudes do not scale

as negative powers of momentum, determine that the superamplitudes of the massless supermultiplets are (neglecting coupling constants):

$$\mathcal{A}(G_1^+, G_2^+, G_3^-) = \widetilde{\delta}^{(1)}(Q)\frac{[12]^2}{[13]\,[23]} \tag{1.93}$$

$$\mathcal{A}(G_1^-, G_2^+, G_3^-) = \delta^{(2)}(Q^\dagger)\frac{\langle 13\rangle^2}{\langle 12\rangle\,\langle 23\rangle} \tag{1.94}$$

$$\mathcal{A}(\Phi_1^-, G_2^+, \Phi_3^+) = \widetilde{\delta}^{(1)}(Q)\frac{[23]}{[13]} \tag{1.95}$$

$$\mathcal{A}(\Phi_1^+, G_2^+, \Phi_3^-) = \widetilde{\delta}^{(1)}(Q)\frac{[21]}{[31]}. \tag{1.96}$$

Other superamplitudes between other possible combinations of massless superfields are also possible, but do not arise in taking the massless limit of (1.87).

Choosing a particular helicity configuration in (1.87), the massless limit may be taken using the limits presented in Appendix 1.A.1 and identified with the superamplitudes above. The limits may be calculated explicitly to be

$$\mathcal{A}(\mathcal{W}_1^+, G_2^+, \overline{\mathcal{W}}_3^+) \to 0, \qquad \mathcal{A}(\mathcal{W}_1^-, G_2^+, \overline{\mathcal{W}}_3^-) \to \mathcal{A}(G_1^-, G_2^+, G_3^-) \tag{1.97}$$

$$\mathcal{A}(\mathcal{W}_1^+, G_2^+, \overline{\mathcal{W}}_3^-) \to -\mathcal{A}(G_1^+, G_2^+, G_3^-)\hat{\eta}_1 + \mathcal{A}(\Phi_1^-, G_2^+, \Phi_3^+)\hat{\eta}_3, \tag{1.98}$$

and similarly for $\mathcal{A}(\mathcal{W}_1^-, G_2^+, \overline{\mathcal{W}}_3^+)$. Similar results may be shown for the limits of (1.90).

This demonstrates how the supersymmetrisation of the Higgs mechanism operates by combining well-defined UV amplitudes of massless chiral and vector multiplets into single superamplitudes of massive vector multiplets in the IR.

### 1.5.6 Self-interacting Massive Vector Supermultiplets

A similar analysis may be performed to determine the possible structure of three-leg superamplitudes of massive vector superfields. A general expression will include several

special cases, such as when the vectors have equal mass and belong to the same species, as well as the case in which there is only one type of superfield, which must be constrained so that there are no vector self-interactions.

Just as for the cases considered previously, supersymmetry implies that the amplitude has the form

$$\mathcal{A}(\mathcal{W}_1^I, \mathcal{W}_2^J, \mathcal{W}_3^K) = \delta^{(2)}(Q^\dagger) F_1^{IJKM} \Big( \eta_{1,M} + \frac{1}{m_2} \left[1_M 2^N\right] \eta_{2,N} \Big). \tag{1.99}$$

This is the extent to which supersymmetry determines the amplitude. The next step is to determine the number of independent Lorentz structures that can appear in $F_1^{IJKM}$. Altogether, there are 6 such independent terms (up to others related by the Schouten identity and kinematic relations):

$$\begin{aligned} F_1^{IJKM} = c_1 \left\langle 1^I 3^K \right\rangle \left[2^J 1^M\right] + c_2 \left[1^I 3^K\right] \left\langle 2^J 1^M \right\rangle + c_3 \left[1^I 3^K\right] \left[2^J 1^M\right] \\ + c_4 \left\langle 1^I 3^K \right\rangle \left\langle 2^J 1^M \right\rangle + c_5 \left\langle 2^J 3^K \right\rangle \epsilon^{IM} + c_6 \left[2^J 3^K\right] \epsilon^{IM}. \end{aligned} \tag{1.100}$$

One of the independent terms in this superamplitude represents a Higgs coupling, where the Higgs has a "$D$-term" quartic and is part of the chiral multiplet eaten with the Goldstone boson. In the Abelian Higgs theory, this is the only structure in the superamplitude. This may be identified by extracting the component amplitude of three vectors and setting it to zero. The component amplitude is

$$A(W_1^{I_1 I_2}, W_2^{J_1 J_2}, W_3^{K_1 K_2}) = F_1^{I_1 J_1 K_1 M} \left( \delta_M^{I_2} \left\langle 3^{K_2} 2^{J_2} \right\rangle - \frac{1}{m_2} \left[1_M 2^{J_2}\right] \left\langle 3^{K_2} 1^{I_2} \right\rangle \right), \tag{1.101}$$

where external spin indices are implicitly symmetrised over. After simplification this reduces to five independent spin structures. Demanding that these vanish implies that $c_2 = c_3 = c_4 = 0$, $c_6 = 0$ and $c_5 = -m_2 c_1$, thereby reducing the number of independent

couplings to one. The corresponding term in the superamplitude is then

$$\mathcal{A}(\mathcal{W}_1^I, \mathcal{W}_2^J, \mathcal{W}_3^K) = c_1\delta(Q^\dagger)\Big( \left(-m_2 \left\langle 2^J 3^K \right\rangle \epsilon^{IM} + \left\langle 1^I 3^K \right\rangle \left[2^J 1^M\right]\right) \eta_{1M} + \left(m_1 \left\langle 1^I 3^K \right\rangle \epsilon^{JM} - \left\langle 2^J 3^K \right\rangle \left[1^I 2^M\right]\right) \eta_{2M}\Big), \tag{1.102}$$

which is manifestly antisymmetric under the exchange $1 \leftrightarrow 2$. This constitutes one of the six independent contributions to the superamplitude (1.99) and is itself the three-particle superamplitude for the Abelian Higgs theory.

This contains component amplitudes of the form that would be expected in Abelian Higgs theories. For example, calling $H_i$ the scalar components of the supermultiplets, then

$$\begin{aligned} A_3(W_1^{I_1 I_2}, H_2, W_3^{K_1 K_2}) &= \frac{\partial}{\partial\eta_{1I_2}}\Big(\frac{1}{2}\epsilon_{J_1 J_2}\frac{\partial}{\partial\eta_{2J_2}}\Big)\frac{\partial}{\partial\eta_{3K_2}}\mathcal{A}(\mathcal{W}_1^{I_1}, \mathcal{W}_2^{J_1}, \mathcal{W}_3^{K_1}) \\ &= -c_1 m_3 \left[1^I 3^K\right] \left\langle 1^I 3^K \right\rangle . \end{aligned} \tag{1.103}$$

Completion of the identification of this with a Higgs amplitude would require that $c_1$ be inversely proportional to some mass scale and that $c_1 \sim 1/m_3^2$ as $m_3 \to 0$ (and likewise for the other masses, repeating this argument with the identities of particles $1, 2$ and $3$ permuted). These are the component amplitudes expected in the Abelian Higgs theory and, given the assumption that there are no vector self-interactions, $\mathcal{N} = 1$ supersymmetry implies that there is only a single Lorentz structure and coupling consistent with this.

The remaining five couplings each describe superamplitudes with vector boson self-interactions. The triple gauge coupling vertex of three massive vectors has been studied extensively in the past in the context of the electroweak bosons of the Standard Model. An effective Lagrangian describing the independent Lorentz structures has been given

in (39). The superamplitude (1.99) represents the supersymmetrisation of this. Supersymmetry restricts the seven independent couplings of (39) to five. The two prohibited terms are those originating from $F^3$ terms (for Yang-Mills curvatures $F$ of the vectors), just as for massless amplitudes.

Of the five remaining structures, one can be attributed to the Yang-Mills (tree) coupling. Just as for the Higgs couplings, the expected Yang-Mills vector self-interaction term may be identified by matching the component amplitude (1.101) to the expected expression. Doing so imposes $c_3 = c_4 = 0,\ c_6 = m_2 c_2$ and identifies the gauge coupling as $c_2 = -2g/(m_1 m_3)$. The Higgs coupling $c_5 = -m_2 c_1$ remains free. This structure, in addition to the Higgs coupling above, are distinguished as having UV limits that converge to massless three particle superamplitudes at leading order.

The remaining four couplings have poor UV scaling and correspond to field theoretic operators upon which gauge invariance is not linearly realised. Two of these (that are $CP$-odd) may be identified with the generalised Chern-Simons terms mentioned earlier (or are generated at loop-level by anomalies), while the other two correspond to the remaining two types of operators that may be constructed from vector multiplets with a single derivative. Of these, one corresponds to the anomalous magnetic dipole moment in the massless limit of one leg in (1.87) and (1.90). Its $CP$-odd counterpart, in the massless limit, provides the same Lorentz structure, but with a different phase in the coupling. The other two couplings vanish in the limit of a massless leg on-shell.

Further conditions may be used to constrain or interpret the couplings, such as requiring good UV limits and properties of higher leg amplitudes. Two simple examples are provided by demanding that this amplitude matches onto either of the two cases discussed in Section 1.5.5 in the limit that $m_1 \to 0$.

For the case where we leave $m_2 \neq m_3$, we demand that (1.99) converge to (1.92) and (1.88) for each helicity configuration of the massless vector multiplet. As long as the

couplings do not scale as $\sim 1/m_1$, this requires that $c_3 = -a$ and $c_4 = -b$, up to terms $\propto m_1$. One of the terms with couplings $c_1$ and $c_2$ vanishes (which depends upon the helicity choice for index $I$) while the other degenerates with the $c_5$ and $c_6$ terms and so cannot be independently determined. Finally, the couplings $c_5$ and $c_6$ (up to inclusion of possible contributions from $c_1$ and $c_2$ as just described) match onto the terms in (1.80) and (1.81) and may be identified with the couplings $d_1$ and $d_2$.

In the case where the two remaining masses approach equality, $m_2, m_3 \to m$ as $m_1 \to 0$, we can demand that the coefficients of (1.99) approach (1.87) and (1.90). This determines the coefficients to be $c_1 = c_2 = -g'/(m_1 m)$, $c_3 = h/m^2$ and $c_4 = h'/m^2$, while it is required that $g' = g$ in the massless amplitudes (so parity must be an accidental symmetry if $h = h' = 0$). These may be easily checked using the spinor limits provided in Appendix 1.A.1. Again, matching onto the superamplitudes with massless vectors, the remaining couplings must be $c_5 = -c_6 = g'/m_1$, but may additionally have extra terms that would be determined by matching onto the amplitudes with massless matter (1.80) and (1.81). Unlike the previous case, these limits ensure that the mass scale of the couplings is given by $m$ and $m_1$, so that, if $h = h' = 0$ (as is true at tree-level in perturbative gauge theories), the amplitudes would have the good UV limits arranged by the Higgs mechanism (6). Note that, as expected from (1.27), the superamplitudes have limits $\mathcal{A}(\mathcal{W}_1^+, \mathcal{W}_2^J, \mathcal{W}_3^K) \to \mathcal{A}(G_1^+, \mathcal{W}_2^J, \mathcal{W}_3^K)\hat{\eta}_1 + \mathcal{A}(\Phi_1^+, \mathcal{W}_2^J, \mathcal{W}_3^K)$ and $\mathcal{A}(\mathcal{W}_1^-, \mathcal{W}_2^J, \mathcal{W}_3^K) \to \mathcal{A}(G_1^-, \mathcal{W}_2^J, \mathcal{W}_3^K) + \mathcal{A}(\Phi_1^-, \mathcal{W}_2^J, \mathcal{W}_3^K)\hat{\eta}_1$ and involve terms that pick up the extra Grassmann variable $\hat{\eta}_1$ for the massless superfield. A similar analysis can be performed by instead $m_2 \to 0$ or $m_2 \to 0$ in order to find further consistency conditions on the couplings to match onto the superamplitudes in the previous sections, but we refrain from providing the results here. These are consistent with the identifications of the couplings made above - that is, $c_3$ and $c_4$ are associated with the couplings that determined the anomalous magnetic and electric dipole moments in the limits of a massless

leg, while linear combinations of $c_1$, $c_2$, $c_5$, $c_6$ correspond to the tree-level ("$D$-term") Higgs and Yang-Mills couplings, while $c_5$ and $c_6$ also contain the other non-Yang-Mills contact interactions, such as those induced from Stuckelberg axions and anomalies.

### 1.5.7 Higher Spin Amplitudes

While the number of possible Lorentz structures in three-particle amplitudes typically grows significantly with the spin of the interacting particles, the case of a heavy particle decaying into two massless products is especially simple. As described in (6), the amplitude for a spin $s$ massive particle $\bar{\phi}$ to decay into two massless particles $\varphi_1$ and $\varphi_2$ with respective helicities $h_1$ and $h_2$ is uniquely

$$A(\varphi_1^{h_1}, \varphi_2^{h_2}, \bar{\phi}_3^{(I_1 \ldots I_{2s})}) = G\,[12]^{s+h_1+h_2} \prod_{i=1}^{s+h_2-h_1} \left[3^{(I_i}1\right] \prod_{j=s+h_2-h_1+1}^{2s} \left[3^{I_j)}2\right], \tag{1.104}$$

where $G$ is some coupling constant of mass dimension $[G] = -(2s + h_1 + h_2 - 1)$. The notation is intended to indicate that all of the spin indices for the massive field are symmetrised over. It is being assumed that angular momentum selection rules permit this process to exist.

The supersymmetrisation of this is just as simple. Promoting $\varphi^{h_i}$ to massless supermultiplets (1.30) with Clifford vacua of helicities $h_i$ and likewise $\phi$ to the corresponding massive multiplet (1.29), then the three-particle superamplitude is also fixed as

$$\begin{aligned} \mathcal{A}(\Sigma_1^{h_1}, \Sigma_2^{h_2}, S_3^{(I_1 \ldots I_{2s})}) &= \frac{1}{m_S} \delta^{(2)}(Q^\dagger) A(\varphi_1^{h_1}, \varphi_2^{h_2}, \bar{\phi}_3^{(I_1 \ldots I_{2s})}) \\ &= \frac{1}{\langle 12 \rangle} \delta^{(2)}(Q^\dagger) A(\xi_1^{h_1 - \frac{1}{2}}, \xi_2^{h_2 - \frac{1}{2}}, \phi_3^{(I_1 \ldots I_{2s})}), \end{aligned} \tag{1.105}$$

where $m_S$ is the mass of the heavy multiplet.

The superamplitude for scattering of four massless particles by exchange of a massive

spinning particle may be constructed analogously to the non-supersymmetric case (6). Supersymmetry fixes the superamplitude to have the form

$$\mathcal{A}(\Sigma_1^{h_1}, \Sigma_2^{h_2}, \Sigma_3^{h_3}, \Sigma_4^{h_4}) = \frac{1}{\langle 34 \rangle} \delta^{(2)}(Q^\dagger) A(\varphi_1^{h_1}, \varphi_2^{h_2}, \xi_3^{h_3 - \frac{1}{2}}, \xi_4^{h_4 - \frac{1}{2}}), \tag{1.106}$$

where the component amplitude $A(\varphi_1^{h_1}, \varphi_2^{h_2}, \xi_3^{h_3-\frac{1}{2}}, \xi_4^{h_4-\frac{1}{2}})$ may be constructed out of the spinning Gegenbauer polynomials corresponding to the exchange of higher spin resonances, just as for the non-supersymmetric case (6).

On a massive resonance, the superamplitude respects a supersymmetric factorisation into three-particle superamplitudes. For example, in the $s$-channel,

$$\mathcal{A}(\Sigma_1^{h_1}, \Sigma_2^{h_2}, \Sigma_3^{h_3}, \Sigma_4^{h_4}) \to \int d^2\eta_P \mathcal{A}_L(\Sigma_1^{h_1}, \Sigma_2^{h_2}, S_P^{(I_1 \ldots I_{2s})}) \frac{1}{(p_1+p_2)^2} \mathcal{A}_R(\overline{S}_{-P(I_1 \ldots I_{2s})}, \Sigma_3^{h_3}, \Sigma_4^{h_4}), \tag{1.107}$$

where the intermediate superfield has Grassmann variables $\eta_P^I$ and the Grassmann integral accounts for the sum over all states in the multiplet of the intermediate resonance. It has been chosen to represent the massive multiplet as outgoing in $\mathcal{A}_L$ and incoming in the other factor. The incoming superfield is then represented as the analytic continuation of an outgoing multiplet. Crossing relations imply that this must be the antimultiplet, hence the bar and the opposite height spin indices. The component antiparticles occupy opposite levels in the superfield.

The factorisation of the superamplitude (1.107) is easily demonstrated as consistent with expectations from (1.106). Because of the simplicity of the three-particle superamplitudes, the Grassmann integral may be trivially evaluated using $\int d^2\eta_P \delta^{(2)}(Q_L^\dagger)\delta^{(2)}(Q_R^\dagger) = m_S \delta^{(2)}(Q^\dagger)$, where $Q_L^\dagger$ and $Q_R^\dagger$ are the supercharges associated with each respective sub-superamplitude above in (1.107). This requires use of the analytic continuation rules for

spinors and Grassmann variables given in (14), which here imply that $Q^\dagger_{i,-P} = -Q^\dagger_{i,P}$ for state $i$ of momentum $P$. The two representations of the three-particle superamplitude (1.105) can then be substituted to confirm that (1.107) is given simply by (1.106) with the exhibited component amplitude factorised into the component three-particle amplitudes shown in (1.105).

## 1.6 Conclusion

We have here initiated the study of the on-shell properties of supersymmetric theories by developing the on-shell superspace formalism in which states are described in a supermultiplet by their asymptotic quantum numbers - momentum, total spin and polarisation - without the need to commit to a frame of reference. This was used to construct massive supermultiplets and represent these in scattering amplitudes of supersymmetric theories, concentrating here on $\mathcal{N} = 1$ theories. Purely from the foundational principles of quantum mechanics, special relativity and supersymmetry, we constructed all of the possible elementary on-shell three-point amplitudes for multiplets of spin no greater than 1.

A more exhaustive study into the extent to which $S$-matrix postulates constrain supermultiplets and their interactions at weak coupling is warranted. Further constraints upon theories from assumptions about IR properties, such as factorisation or behaviour in the high energy limit, remain to be investigated.

It would be interesting to more broadly catalogue theories characterised by their spectra and interactions from conditions on IR properties and see whether they conspire to imply emergent symmetries or uniqueness properties (6; 31; 40). For example, consequences of supersymmetry on emergent properties of theories constructible from soft limits were recently investigated in (41). We do not foresee difficulties in extending our

analysis to scattering states of higher-spin composite superfields or including multiplets of supergravity or Kaluza-Klein modes (see recently (15; 42; 43) for a possible application to black holes).

To progress beyond single particle representations and 3-leg amplitudes, some guidance for systematically constructing higher order (loop and leg) amplitudes from infrared (on-shell) properties would be desirable, such as on-shell recursion. However, because the validity of massless recursion is often sensitive to the helicity of the shifted states, the effective combining of massless states of definite helicity into massive particle representations of the (super-)Poincare group poses a potential obstruction. Prospects for overcoming this are most promising in $\mathcal{N} = 4$ SYM where, for massless amplitudes, a myraid of constructibility properties have been discovered. Vestiges of these may remain present on the Coulomb branch, in particular the dual (super)conformal symmetry. In (14) we formulate a massive super-BCFW shift and prove its validity for the construction of all Coulomb branch tree superamplitudes. The constructibility of Coulomb branch superamplitudes seems to arise from a surprising 'nonlocality' present in the three-particle superamplitudes. This remains an interesting avenue for future work.

# Acknowledgments

We thank Tim Cohen, Nathaniel Craig, Henriette Elvang, and Callum Jones for comments on a draft of this work, Nathaniel Craig for discussions and support during the completion of this work, and Nima Arkani-Hamed and Yu-tin Huang for discussions on (6). AH and SK are grateful for the support of a Worster Fellowship. This work is supported in part by the US Department of Energy under the grant DE-SC0014129.

# Appendices

# 1.A Conventions and Useful Identities

## 1.A.1 Spinor Helicity for Massive Particles

We here summarise helicity spinors for massive particles and its consequences, taking the opportunity to establish the conventions and notation that we adopt throughout this article and also to present useful identities. The reader is referred to (8) for review of the spinor helicity method for scattering processes of massless particles, the conventions of which, in addition to (10), we (mostly) adhere to and will not restate.

Introducing helicity spinors with $SU(2)$ little group structure has consequences for the description of the internal and external structure of scattering amplitudes. Internally, as mentioned above, the starting point is that massive momenta (as representations of the spin group $SL(2,\mathbb{C})$: $p = p^\mu \sigma_\mu$) may be decomposed into two null momenta as

$$p^{\dot{\alpha}\beta} = -\sum_I |p_I\rangle^{\dot{\alpha}} \left[p^I\right|^\beta . \tag{1.108}$$

The two pairs of left- and right-handed spinors indexed by $I$, $|p_I]$ and $\langle p^I|$, transparently respect an $SU(2)$ symmetry that may be identified with the momentum's little group. These $SU(2)$ indices may be raised and lowered in the usual way to convert between representations and their conjugates:

$$\left\langle p^I\right|_{\dot{\alpha}} = \varepsilon^{IJ} \langle p_J|_{\dot{\alpha}} \qquad |p_I]_\alpha = \varepsilon_{IJ} \left|p^J\right]_\alpha . \tag{1.109}$$

Under conjugation, the spinors transform as

$$(\left[p^I\right|)^\dagger = |p_I\rangle \qquad (\left\langle p^I\right|)^\dagger = -|p_I] . \tag{1.110}$$

Fundamental tensor representations have lowered indices. We take all scattering states here to be outgoing, so naturally have raised internal indices (including little group) corresponding to the polarisations of the conjugated states.

As usual, $\det(p) = -p^2 = m^2$ for mass $m$. As the spinors in (1.108) are conjugates, $\det(p) = \det(|p]) \det(\langle p|) = |\det(|p])|^2$. The choice of the phase of $\det(|p])$ is free, so $\det(|p]) = m$ may be chosen without loss of generality (although see (44) for interpretation of the mass and its complex phase as the extra components of a 6d momentum and its consequences for dual conformal symmetry). The spinors then have bilinear products with themselves

$$\left\langle p^I p^J \right\rangle = m\varepsilon^{IJ} \qquad \left[p^I p^J\right] = -m\varepsilon^{IJ}, \tag{1.111}$$

obey the Weyl equations:

$$\begin{aligned} p\left|p^I\right] &= -m\left|p^I\right\rangle & p\left|p^I\right\rangle &= -m\left|p^I\right] \\ \left[p^I\right| p &= m\left\langle p^I\right| & \left\langle p^I\right| p &= m\left[p^I\right| \end{aligned} \tag{1.112}$$

and the spin sums:

$$\begin{aligned} |p_I]_\alpha \left[p^I\right|^\beta &= m\delta_\alpha^\beta & |p_I]_\alpha \left\langle p^I\right|_{\dot\beta} &= p_{\alpha\dot\beta} \\ |p_I\rangle^{\dot\alpha} \left\langle p^I\right|_{\dot\beta} &= -m\delta_{\dot\beta}^{\dot\alpha} & |p_I\rangle^{\dot\alpha} \left[p^I\right|^\beta &= -p^{\dot\alpha\beta}. \end{aligned} \tag{1.113}$$

The little group index effectively labels the two possible solutions to each of the Weyl equations, which may be rotated into each other by a Wigner rotation.

Externally, the $S$-matrix transforms as a tensor under the little group of each of its external particle legs, being an array of transition matrix entries between states of

different spin configurations. An external state of spin $s$ has polarisation wavefunction that can be described by a rank $2s$ symmetric tensor of the little group $SU(2)$. States of a particular polarisation $m_s$ may be extracted from this by choosing the symmetrised component of the tensor with $m_s + s$ indices aligned with the spin direction and $s - m_s$ indices opposite and then normalising. For example, a massive vector particle is described by symmetric polarisation tensor $T^{(I_1 I_2)}$, with $m_s = -1, 0, 1$ states respectively given by $T^{--}$, $\frac{1}{\sqrt{2}}\Big(T^{+-} + T^{-+}\Big)$ and $T^{++}$. See (45) for tensor methods to describe spin. We (mostly) restrict to particles of spin $\leq 1$ in this work, although a significant part of the versatility of this formalism is its ability to elegantly describe amplitudes of massive states of any spin.

The possible structures that may appear in the $S$-matrix and are consistent with Lorentz invariance are determined by the number of independent combinations of external state polarisations that can be made. The systematic construction of these was described in (6). Rather than build external polarisations directly from the tensor products of massive spinors (e.g. $T^{(I_1 I_2)} = \left|p^{(I_1}\right] \left|p^{I_2)}\right]$), a direct on-shell construction of elementary amplitudes can be performed instead by using the massive spinors to construct a tensor basis with respect to which the $S$-matrix may be decomposed. Spinors of either chirality (or both) may be used to do this. The coefficients of these basis tensors then represent polarisation-stripped Lorentz tensor amplitudes, in which the possible independent terms may be built out of external momenta and massless spinors. The helicities of massless legs then determine the amplitude's $U(1)$ little group scaling for each massless particle. As a simple example, the $S$-matrix entry for the decay of a massive vector $V_1$ into two massless right-handed fermions $\psi_2$ and $\psi_3$ is determined uniquely by symmetry to be

$$A(V, \psi_2, \psi_3) = g\Big( \left[1^{(I_1}\right|^{\alpha_1} \left[1^{I_2)}\right|^{\alpha_2} \Big) \times |2]_{\alpha_1} |3]_{\alpha_2} = g \left[1^{(I_1} 2\right] \left[1^{I_2)} 3\right], \tag{1.114}$$

for some coupling constant $g$. This method of deducing little group structures built out of spinors is used repeatedly throughout this work in constructing superamplitudes.

Part of the utility of this formalism is that the little group indices are an internal degree of freedom and allow for the polarisation to be projected onto any external spin frame or axis. The procedure for doing this is discussed in (6). In practice, we find that it is clearest to abuse notation and, once such an external frame is specified, simply reinterpret the little group indices as referring to components along this direction. In particular, as it is often most useful, especially in taking massless or high energy limits, to choose spin frames for each particle aligned with their momenta (so that the little group indices simply become helicity indices), we will leave this choice implicit unless stated otherwise.

In this case, the spinors have massless limits

$$\begin{aligned} \left|p^+\right] &\to |p] \qquad \left|p^-\right] \to 0 \\ \left|p^+\right\rangle &\to 0 \qquad \left|p^-\right\rangle \to -|p\rangle \end{aligned} \tag{1.115}$$

where the spinors without little group indices are the usual spinors for massless momentum $p$. More precisely, the spinors that vanish do so $\mathcal{O}(m)$. The limits may be expressed as

$$\lim_{m\to 0} \frac{1}{m} \left|p^-\right] = \frac{|q]}{[qp]} \qquad \lim_{m\to 0} \frac{-1}{m} \left\langle p^+\right| = \frac{\langle q|}{\langle qp\rangle}, \tag{1.116}$$

where the remaining spinors $|q]$ and $|q\rangle$ become the reference spinors and are ambiguous in the massless amplitude, as their direction arrived in taking the limit is arbitrary (up to requiring $[qp], \langle qp\rangle \neq 0$). In practice, it is often possible to take the limit while avoiding the introduction of the reference by using momentum conservation and other identities.

For 3-leg amplitudes involving the factor $x$, the following identities are useful:

$$x = \frac{1}{m}\frac{[q|\, p_2 \,|3\rangle}{[q3]} = \frac{m\,\langle q3\rangle}{\langle q|\, p_2 \,|3]} \tag{1.117}$$

and

$$x\left[32^I\right] = \left\langle 32^I\right\rangle, \quad x\left[31^I\right] = -\left\langle 31^I\right\rangle, \tag{1.118}$$

$$\frac{\left[1^I 2^J\right]}{x} = \frac{\left\langle 1^I 2^J\right\rangle}{x} + \frac{\left[1^I 3\right]\left[2^J 3\right]}{m}, \tag{1.119}$$

where $p_3$ is the massless leg and $p_1$ and $p_2$ are the massive legs, while $|q]$ and $|q\rangle$ are arbitrary reference spinors, not necessarily related, that satisfy $[q3] \neq 0$ and $\langle q3\rangle \neq 0$.

### 1.A.2 Grassmann Calculus

The Grassmann variables may be imbued with $SU(2)$ little group indices $\eta_I$. In this case, Grassmann differentiation may be defined in the usual way: $\frac{\partial}{\partial\eta_I}\eta_J = \frac{\partial}{\partial\eta^J}\eta^I = \delta^I_J$. However, this requires that the index height on the derivative be raised or lowered with an extra $(-)$ sign: $\frac{\partial}{\partial\eta^J} = -\epsilon_{JI}\frac{\partial}{\partial\eta_I}$. We note for convenience the identities

$$\frac{1}{2}\frac{\partial}{\partial\eta_I}\frac{\partial}{\partial\eta^I}\left(\frac{1}{2}\eta_J\eta^J\right) = -1, \tag{1.120}$$

$$\eta_I\eta_J = -\frac{1}{2}\epsilon_{IJ}\eta_K\eta^K. \tag{1.121}$$

The little group invariant Grassmann integration measures are defined here as

$$\mathrm{d}^2\eta = \frac{1}{2}\epsilon^{IJ}d\eta_I\mathrm{d}\eta_J \qquad \mathrm{d}^2\eta^\dagger = -\frac{1}{2}\epsilon^{IJ}d\eta^\dagger_I\mathrm{d}\eta^\dagger_J \tag{1.122}$$

where $\int \mathrm{d}\eta_I\eta^J = \delta^J_I$ and $\int \mathrm{d}\eta^\dagger_I\eta^{\dagger J} = \delta^J_I$. The index placement on the differential is a

property of the differential and not of the variable being integrated - that is, $d\eta_I = d(\eta^I)$. The strange positioning of the index is needed for this operation to be the same as differentiation and is an occurrence of the general topsy-turvyness of Grassmann numbers. Also, as for the derivative, $d\eta^I = -\epsilon^{IJ} d\eta_J$ (and likewise for the conjugate).

The Grassmann Fourier transform of some function $f(\eta)$ of a Grassmann variable $\eta$ is defined as $\tilde{f}$ and these are related by

$$\tilde{f}(\eta^\dagger) = \int \mathrm{d}\eta e^{\eta\eta^\dagger} f(\eta) \qquad f(\eta) = \int \mathrm{d}\eta^\dagger e^{-\eta\eta^\dagger} \tilde{f}(\eta^\dagger). \tag{1.123}$$

The Fourier transform from the $\eta^\dagger$ basis to the $\eta$ basis in $\mathcal{N} = 1$ is effected by the replacements

$$1 \to -\frac{1}{2}\eta_I\eta^I, \qquad \eta^{\dagger I} \to \eta^I, \qquad -\frac{1}{2}\eta^{\dagger I}\eta^\dagger_I \to 1. \tag{1.124}$$

For multiplets without a central charge, the Grassmann variables have massless limits in the helicity basis

$$\eta_- \to \eta, \qquad \eta_+ \to \hat{\eta}. \tag{1.125}$$

Here, $\hat{\eta}$ represents the redundant variable left-over from the division of the massive multiplets into smaller massless multiplets that each represent the smaller massless SUSY algebra. For the exceptional case of BPS multiplets, $\hat{\eta} = \eta^\dagger$, while for anti-BPS multiplets the limit picks up an extra negative sign.

## 1.B Comments on Higher-leg Amplitudes in SQCD

We here make some comments on various massive quark and multigluon amplitudes and rederive them in the little group covariant notation. Again, the arguments presented here are parallel to (9) and (19). Amplitudes stated here will be colour-stripped par-

tial amplitudes, following the usual rules for Yang-Mills theories, as prescribed in e.g. (46). This discussion is supplementary to further comments made in (14) concerning the relation between (S)QCD amplitudes and Coulomb branch amplitudes of $\mathcal{N} = 4$ SYM.

Firstly, the supersymmetric Ward identities provide relations and constraints between component amplitudes that can be exploited. Supersymmetry transformations can be found that set a Grassmann generator for a particular leg to 0. In particular, under the action of $-[\theta Q\rangle$, $\eta_{j,I}$ is translated to $\eta_{j,I} - i\,[\theta j_I\rangle$ for each leg $j$ (if the leg is massless, just omit the little group index). This can be used to set $\eta_{i,I} = 0$, for some single leg $i$ with polarisation in some direction given by $I$ in some little group frame, by choosing $[\theta| = \frac{-i}{m}\eta_{i,I}\left[i^I\right| + C\,[i_I|$ (no sum over $I$ is implied). Here, $C$ represents the remaining unused degree of freedom in the supersymmetry parameter. Component amplitudes that are obtained by integrating the superamplitude in the Grassmann parameters that are translated are unaffected by this transformation, because the integration variable can be likewise translated. After changing variables to absorb the supertranslation, the resulting integrand is completely independent of $\eta_{i,I}$, so integrating over it will give 0. The component amplitudes obtained by such projections must therefore be 0 by supersymmetry.

Simple illustrative examples of this are the squark-antisquark and $n$-gaugino amplitude and the squark-antisquark $n$-gluon amplitude.

$$A[\overline{\widetilde{Q}}_R\lambda^+\cdots\lambda^+\widetilde{Q}_L] = \int \mathrm{d}^2\eta_1 \prod_i \int \mathrm{d}\eta_i \int \mathrm{d}^2\eta_{n+2}\mathcal{A}[\overline{\mathcal{Q}}, G^+\ldots G^+, \mathcal{Q}] = 0 \tag{1.126}$$

$$A[\overline{\widetilde{Q}}_R g^-\cdots g^-\widetilde{Q}_L] = \int \mathrm{d}^2\eta_1 \prod_i \int \mathrm{d}\eta_i \int \mathrm{d}^2\eta_{n+2}\mathcal{A}[\overline{\mathcal{Q}}, G^-\ldots G^-, \mathcal{Q}] = 0 \tag{1.127}$$

Identical arguments in the $\eta^\dagger$ basis may be used to show that the $CP$-conjugate amplitudes are also 0. Identical arguments also demonstrate that amplitudes with additional

squark and antisquark pairs ($\overline{\widetilde{Q}}_R, \widetilde{Q}_L$) are 0.

Vanishing amplitudes of massive quarks, states of non-trivial polarisation, may also be obtained similarly. If all of the little group axes of the quarks are aligned, then the transformation $\eta_{j,J} - i\,[\theta j_J]$ does not affect the Grassmann numbers with opposite spin components to $J$ - which is now the same direction for each massive field. Thus the superamplitude integrated over only these components will be independent of the Grassmann variable that is being eliminated, so must vanish. This derives the fact that amplitudes with quarks and antiquarks all of identical polarisation are 0, as well as those that include some number of gluons or gauginos of identical helicity. These include amplitudes that are inherited by pure QCD at tree-level. The argument can be easily combined with that used in the previous paragraph to extend these vanishing amplitudes to those involving squarks.

The extra degree of freedom in the supersymmetry parameter can be further utilised to derive the vanishing of a further class of amplitudes. Choosing $C = \frac{-i}{m}\frac{[i^I j_K]}{[i_I j_K]}\eta_{i,I}$ (if $j$ is massless, omit its little group index in this expression), the Grassmann variable for leg $j$ does not shift under the supersymmetry transformation performed in the examples above. This means that the variables $\eta_{j,K}$ need not be integrated in order to obtain a vanishing amplitude. Thus one extra particle in any spin state may be added to any of the amplitudes above and the result will still be 0.

Tree-level amplitudes involving a quark-antiquark pair and any number of gluons of the same helicity have been previously determined in (3; 47) and little group covariantised in (48). A compact expression exists that may be derived inductively using BCFW recursion by shifting the massless legs in the usual way (49). The superamplitudes to which these amplitudes belong have the interesting property that they are fully determined by a single component amplitude, which we show in Appendix B of (14) by projecting these superamplitudes out from the massive $\mathcal{N} = 4$ theory.

# Bibliography


[1] T. Cohen, H. Elvang, and M. Kiermaier, *On-shell constructibility of tree amplitudes in general field theories*, *JHEP* **04** (2011) 053, [arXiv:1010.0257].

[2] E. Conde and A. Marzolla, *Lorentz Constraints on Massive Three-Point Amplitudes*, *JHEP* **09** (2016) 041, [arXiv:1601.08113].

[3] C. Schwinn and S. Weinzierl, *SUSY ward identities for multi-gluon helicity amplitudes with massive quarks*, *JHEP* **03** (2006) 030, [hep-th/0602012].

[4] R. Kleiss and W. J. Stirling, *Spinor Techniques for Calculating p anti-p —¿ W+- / Z0 + Jets*, *Nucl. Phys.* **B262** (1985) 235–262.

[5] S. Dittmaier, *Weyl-van der Waerden formalism for helicity amplitudes of massive particles*, *Phys. Rev.* **D59** (1998) 016007, [hep-ph/9805445].

[6] N. Arkani-Hamed, T.-C. Huang, and Y.-t. Huang, *Scattering Amplitudes For All Masses and Spins*, arXiv:1709.04891.

[7] Y. Shadmi and Y. Weiss, *Effective Field Theory Amplitudes the On-Shell Way: Scalar and Vector Couplings to Gluons*, arXiv:1809.09644.

[8] H. Elvang and Y. Huang, *Scattering Amplitudes in Gauge Theory and Gravity*. Cambridge University Press, 2015, arXiv:1308.1697.

[9] R. H. Boels and C. Schwinn, *On-shell supersymmetry for massive multiplets*, *Phys. Rev.* **D84** (2011) 065006, [arXiv:1104.2280].

[10] M. Srednicki, *Quantum field theory*. Cambridge University Press, 2007.

[11] J. Wess and J. Bagger, *Supersymmetry and supergravity*. Princeton, USA: Univ. Pr. (1992) 259 p.

[12] S. Ferrara, C. A. Savoy, and B. Zumino, *General Massive Multiplets in Extended Supersymmetry*, *Phys. Lett.* **100B** (1981) 393–398.

[13] P. Fayet, *Spontaneous Generation of Massive Multiplets and Central Charges in Extended Supersymmetric Theories*, *Nucl. Phys.* **B149** (1979) 137.

[14] A. Herderschee, S. Koren, and T. Trott, *Constructing $\mathcal{N} = 4$ Coulomb Branch Superamplitudes*, arXiv:1902.07205.

[15] S. Caron-Huot and Z. Zahraee, *Integrability of Black Hole Orbits in Maximal Supergravity*, arXiv:1810.04694.

[16] H. Elvang, Y.-t. Huang, and C. Peng, *On-shell superamplitudes in N¡4 SYM*, *JHEP* **09** (2011) 031, [arXiv:1102.4843].

[17] V. P. Nair, *A Current Algebra for Some Gauge Theory Amplitudes*, *Phys. Lett.* **B214** (1988) 215–218.

[18] E. Witten, *Perturbative gauge theory as a string theory in twistor space*, *Commun. Math. Phys.* **252** (2004) 189–258, [hep-th/0312171].

[19] N. Arkani-Hamed, F. Cachazo, and J. Kaplan, *What is the Simplest Quantum Field Theory?*, *JHEP* **09** (2010) 016, [arXiv:0808.1446].

[20] J. M. Drummond, J. Henn, G. P. Korchemsky, and E. Sokatchev, *Dual superconformal symmetry of scattering amplitudes in N=4 super-Yang-Mills theory*, *Nucl. Phys.* **B828** (2010) 317–374, [arXiv:0807.1095].

[21] A. Brandhuber, P. Heslop, and G. Travaglini, *A Note on dual superconformal symmetry of the N=4 super Yang-Mills S-matrix*, *Phys. Rev.* **D78** (2008) 125005, [arXiv:0807.4097].

[22] J. M. Drummond and J. M. Henn, *All tree-level amplitudes in N=4 SYM*, *JHEP* **04** (2009) 018, [arXiv:0808.2475].

[23] S. He and T. McLoughlin, *On All-loop Integrands of Scattering Amplitudes in Planar N=4 SYM*, *JHEP* **02** (2011) 116, [arXiv:1010.6256].

[24] N. Arkani-Hamed, J. L. Bourjaily, F. Cachazo, S. Caron-Huot, and J. Trnka, *The All-Loop Integrand For Scattering Amplitudes in Planar N=4 SYM*, *JHEP* **01** (2011) 041, [arXiv:1008.2958].

[25] F. Cachazo, A. Guevara, M. Heydeman, S. Mizera, J. H. Schwarz, and C. Wen, *The S Matrix of 6D Super Yang–Mills and Maximal Supergravity from Rational Maps*, arXiv:1805.11111.

[26] H. Elvang, D. Z. Freedman, and M. Kiermaier, *Solution to the Ward Identities for Superamplitudes*, *JHEP* **10** (2010) 103, [arXiv:0911.3169].

[27] H. K. Dreiner, H. E. Haber, and S. P. Martin, *Two-component spinor techniques and Feynman rules for quantum field theory and supersymmetry*, *Phys. Rept.* **494** (2010) 1–196, [arXiv:0812.1594].

[28] S. Weinberg, *The quantum theory of fields. Vol. 3: Supersymmetry*. Cambridge University Press, 2013.

[29] J. M. Cornwall, D. N. Levin, and G. Tiktopoulos, *Derivation of Gauge Invariance from High-Energy Unitarity Bounds on the s Matrix*, *Phys. Rev.* **D10** (1974) 1145. [Erratum: Phys. Rev.D11,972(1975)].

[30] S. L. Adler, *Collinearity constraints for on-shell massless particle three-point functions, and implications for allowed-forbidden $n+1$-point functions*, *Phys. Rev.* **D93** (2016), no. 6 065028, [arXiv:1602.05060].

[31] D. A. McGady and L. Rodina, *Higher-spin massless S-matrices in four-dimensions*, *Phys. Rev.* **D90** (2014), no. 8 084048, [arXiv:1311.2938].

[32] S. Ferrara and E. Remiddi, *Absence of the Anomalous Magnetic Moment in a Supersymmetric Abelian Gauge Theory*, *Phys. Lett.* **53B** (1974) 347–350.

[33] P. C. Schuster and N. Toro, *Constructing the Tree-Level Yang-Mills S-Matrix Using Complex Factorization*, *JHEP* **06** (2009) 079, [arXiv:0811.3207].

[34] L. Andrianopoli, S. Ferrara, and M. A. Lledo, *Axion gauge symmetries and generalized Chern-Simons terms in N = 1 supersymmetric theories*, *JHEP* **04** (2004) 005, [hep-th/0402142].

[35] P. Anastasopoulos, M. Bianchi, E. Dudas, and E. Kiritsis, *Anomalies, anomalous U(1)'s and generalized Chern-Simons terms*, *JHEP* **11** (2006) 057, [hep-th/0605225].

[36] P. Anastasopoulos, F. Fucito, A. Lionetto, G. Pradisi, A. Racioppi, and Y. S. Stanev, *Minimal Anomalous U(1)-prime Extension of the MSSM*, *Phys. Rev.* **D78** (2008) 085014, [arXiv:0804.1156].

[37] S. Ferrara, M. Porrati, and V. L. Telegdi, *g = 2 as the natural value of the tree level gyromagnetic ratio of elementary particles*, *Phys. Rev.* **D46** (1992) 3529–3537.

[38] A. Guevara, *Holomorphic Classical Limit for Spin Effects in Gravitational and Electromagnetic Scattering*, arXiv:1706.02314.

[39] K. Hagiwara, R. D. Peccei, D. Zeppenfeld, and K. Hikasa, *Probing the Weak Boson Sector in e+ e- —¿ W+ W-*, *Nucl. Phys.* **B282** (1987) 253–307.

[40] P. Benincasa and F. Cachazo, *Consistency Conditions on the S-Matrix of Massless Particles*, arXiv:0705.4305.

[41] H. Elvang, M. Hadjiantonis, C. R. T. Jones, and S. Paranjape, *Soft Bootstrap and Supersymmetry*, arXiv:1806.06079.

[42] M.-Z. Chung, Y.-t. Huang, J.-W. Kim, and S. Lee, *The simplest massive S-matrix: from minimal coupling to Black Holes*, arXiv:1812.08752.

[43] A. Guevara, A. Ochirov, and J. Vines, *Scattering of Spinning Black Holes from Exponentiated Soft Factors*, arXiv:1812.06895.

[44] J. Plefka, T. Schuster, and V. Verschinin, *From Six to Four and More: Massless and Massive Maximal Super Yang-Mills Amplitudes in 6d and 4d and their Hidden Symmetries*, *JHEP* **01** (2015) 098, [arXiv:1405.7248].

[45] H. Georgi, *Lie algebras in particle physics*, *Front. Phys.* **54** (1999) 1–320.

[46] M. L. Mangano and S. J. Parke, *Multiparton amplitudes in gauge theories*, *Phys. Rept.* **200** (1991) 301–367, [hep-th/0509223].

[47] C. Schwinn and S. Weinzierl, *On-shell recursion relations for all Born QCD amplitudes*, *JHEP* **04** (2007) 072, [hep-ph/0703021].

[48] A. Ochirov, *Helicity amplitudes for QCD with massive quarks*, *JHEP* **04** (2018) 089, [arXiv:1802.06730].

[49] S. D. Badger, E. W. N. Glover, V. V. Khoze, and P. Svrcek, *Recursion relations for gauge theory amplitudes with massive particles*, *JHEP* **07** (2005) 025, [hep-th/0504159].

# Chapter 2

# Constructing $\mathcal{N} = 4$ Coulomb Branch Superamplitudes

We study scattering amplitudes of massive BPS states on the Coulomb branch of $4d$ $\mathcal{N} = 4$ super-Yang-Mills, utilising a little group covariant on-shell superspace for massive particles. Super-BCFW recursion for massive amplitudes is constructed and its validity is proven for all Coulomb branch superamplitudes. We then determine the exact three-particle superamplitudes for massive states. These ingredients allow us to explicitly compute the four- and five-particle superamplitudes, which is the first non-trivial usage of BCFW recursion for amplitudes with entirely massive external states. The manifest little group covariance helps clarify both the role of special kinematic properties of BPS states and the organizational structures of the superamplitudes.

## 2.1 Introduction

The most powerful on-shell properties are to be found with maximal $\mathcal{N} = 4$ super-symmetry (at least for non-gravitational theories). Although a highly idealised model

of QCD, numerous hidden structures beyond the maximal, rigid supersymmetry have been uncovered and their role in nature remains to be ascertained. Some particular highlights include the computation of tree amplitudes at strong coupling by holography (1), the duality of planar (large number of colours) amplitudes with Wilson loops (2–6), the discovery of dual (super)conformal symmetry (in addition to regular spacetime superconformal symmetry) (7; 8), Yangian symmetry and integrable structure (9), constructibility of tree (10) amplitudes by BCFW recursion (11; 12), loop integrands by on-shell diagrams and full constructibility from leading singularities (13; 14) and the interpretation of amplitudes as volumes of polytopes (15; 16). Most of this work has focused on the origin of the moduli space, where the states are all massless and the theory is conformal.

The structure of amplitudes of massive particles with $\mathcal{N} = 4$ supersymmetry has received comparatively little attention. These nevertheless provide a further testing ground of the special symmetries and properties listed above and the extent to which they are deformed but not destroyed by Higgsing. Previous studies of massive amplitudes on the Coulomb branch have been made in (17–20), where a gamut of methods including soft limits, supersymmetric on-shell recursion and solutions to the supersymmetric Ward identities (SWIs) were proposed and used to compute some simple examples. Subsequently, some $4d$ tree-level amplitudes and loop integrands have been obtained by dimensional reduction from superamplitudes of the $6d$ $\mathcal{N} = (1, 1)$ SYM theory, for which dual conformal invariance has been established, despite the absence of conformality (21–25). However, a general procedure for explicitly constructing amplitudes beyond the fewest leg examples was not developed. More recently, a CHY (26) formula for all $6d$ $\mathcal{N} = (1, 1)$ massless amplitudes was found and reduced to give a general formula for all $4d$ massive $\mathcal{N} = 4$ tree amplitudes (27), from which a few examples were extracted (a new proposal was recently made in (28)). Partial use of the massive spinor helicity formalism discussed here was made to extract some simple examples of amplitudes contained within the gen-

eral formula. Nevertheless, much of the structure of these amplitudes thus far remains unexplored. We will review this subject more thoroughly in Section 2.5.

To proceed onto the Coulomb branch, we first discuss an on-shell superspace for massive BPS vector multiplets. Purely through the use of on-shell properties and maximal rigid supersymmetry, we construct the unique elementary three particle superamplitudes of the theory. These superamplitudes of massive legs have 'nonlocal' kinematic denominators analogous to that present in massless (S)YM, despite this feature not being present in any of the component amplitudes. This arises as a result of the special complex kinematics of the BPS states and suggests that the massive amplitudes share in the special constructibility properties of massless gauge theory. We confirm this by formulating a massive super-BCFW shift and proving the constructibility of all Coulomb branch tree amplitudes under it. Using this to fuse the four particle superamplitude from a single factorization channel between on-shell three-leg superamplitudes, we are able to explicitly locate the second pole of the four-point superamplitude as coming from the singular overlap of the two special kinematic configurations on either side of the factorization channel.

The establishment of super-BCFW for massive legs allows for the systematic computation of relatively compact expressions for massive superamplitudes. To illustrate this, we explicitly write down the five particle superamplitude for all-massive legs, which is the first non-trivial usage of on-shell recursion to construct an amplitude of fully massive external states. The way in which the massless sectors of helicity violation combine together when the states are massive is also shown.

This work is partnered with a companion paper (29) that discusses the on-shell properties of supersymmetric theories with massive particles (mostly with $\mathcal{N} = 1$ supersymmetry). This makes use of the adaptation of helicity spinors to describe the kinematics of massive particles made in (30) with manifest little group covariance.

This paper takes the following steps toward elucidating the structure of massive amplitudes in $\mathcal{N} = 4$ SYM. We firstly review, in Section 2.2, the representation theory of massive particles pertinent to the Coulomb branch of $\mathcal{N} = 4$. In Section 2.3, we introduce the 'non-chiral' superspace in which the superamplitudes are naturally formulated and explain the representation of BPS states (here massive elementary vector multiplets) in on-shell superspace. In order to construct higher-leg amplitudes, we implement BCFW recursion for massive superamplitudes in Section 2.4 and establish that all Coulomb branch amplitudes are constructible in this manner. In Section 2.5, we commence the calculation of massive scattering amplitudes. We find the three-particle superamplitudes in subsection 2.5.1, which features a 'special kinematics' of BPS states resembling that of massless particles with complex momenta, as well as a surprising 'nonlocality' in their superamplitudes. This enables us to recursively construct the four-leg superamplitude in subsection 2.5.2 (with some computational details shunted to Appendix 2.A). In subsection 2.5.3, after a discussion of the supersymmetric 'band structure', we are able to use the same technique to find the five particle superamplitude for all-massive states. We then conclude. In Appendix 2.B we make some comments about projecting Coulomb branch superamplitudes down to Yang-Mills theories with massive particles with fewer supersymmetries.

## 2.2 On-shell superfields for massive particles

In (29) we construct on-shell superspaces for massive supermultiplets that are covariant in the $SU(2)$ little group, recently introduced into helicity spinors in (30). We here briefly summarize the important results and refer the reader to (29) for further details, especially its appendix of conventions and identities.

For $\mathcal{N}$-extended SUSY, the supercharges carried by leg $i$, $Q_{i,\alpha A}$ and $Q^{\dagger A}_{i,\dot{\alpha}}$, satisfy the

commutation relations

$$\{Q_{i,\alpha A}, Q_{i,\beta B}\} = Z_{i,AB}\epsilon_{\alpha\beta} \qquad \{Q^{\dagger A}_{i,\dot\alpha}, Q^{\dagger B}_{i,\dot\beta}\} = -Z_i^{AB}\epsilon_{\dot\alpha\dot\beta}$$
$$\{Q_{i,\alpha A}, Q^{\dagger B}_{i,\dot\beta}\} = -2\delta^B_A(\sigma^\mu_{\alpha\dot\beta})P_{i,\mu} \tag{2.1}$$

where $P_{i,\mu}$ is the momentum and $Z_{i,AB}$ is the central charge, satisfying $Z_{i,AB} = -Z_{i,BA} = -(Z_i^{AB})^*$. The labels $A$ and $B$ are $R$-indices. On-shell, little group covariant supersymmetry generators are defined for each leg by projecting the supercharges onto the spinors of a given particle

$$q^I_{i,A} = \frac{-1}{\sqrt{2}m_i}\left[i^I Q_{i,A}\right], \qquad q^{\dagger A}_{i,I} = \frac{1}{\sqrt{2}m_i}\left\langle i_I Q^{\dagger A}_i\right\rangle, \tag{2.2}$$

which satisfy the anticommutation relations

$$\left\{q^I_{i,A}, q^{\dagger J,B}_i\right\} = -\epsilon^{IJ}\delta^B_A, \qquad \left\{q^I_{i,A}, q^J_{i,B}\right\} = -\epsilon^{IJ}\frac{Z_{i,AB}}{2m_i}, \qquad \left\{q^{\dagger I,A}_i, q^{\dagger J,B}_i\right\} = \epsilon^{IJ}\frac{Z_i^{AB}}{2m_i}. \tag{2.3}$$

The index $I$ denotes massive $SU(2)$ little group component while $m_i$ is the mass of the leg. For the simplest case, which will be considered here, $Z_{i,AB} = Z_i\,\Omega_{AB}$ for all $i$, where $Z_i \in \mathbb{R}$ while $\Omega_{AB} = -\Omega_{BA}$ is a symplectic 2-form

$$\Omega_{AB} = \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix}. \tag{2.4}$$

The case $|Z_i| = 2m_i$ is the special BPS limit and will be relevant for states on the Coulomb branch. For these representations, half of the supercharges are eliminated through the reality constraint

$$q_{i,IA} = \frac{-1}{2m_i}Z_{i,AB}q^{\dagger B}_{i,I}. \tag{2.5}$$

The phase of $Z$ may be absorbed into a redefinition of the supercharges $q_i$ and $q_i^\dagger$. This condition again preserves the supersymmetry algebra. BPS states are annihilated by the combination $q_i^{IA} \pm q_i^{\dagger IA}$ (the sign is determined by the sign of $Z_i$). For non-BPS representations with a central charge, linear combinations of supercharges may be found that will satisfy the algebra (2.3) with $Z_{i,AB} = 0$. The representation theory of these states is therefore unaffected by the existence of a central charge.

The explicit $SU(\mathcal{N})$ automorphism symmetry of the SUSY algebra is broken to $USp(\mathcal{N})$ by the central charge of these massive single particle states, which is exactly the massive $R$-symmetry group expected for a theory with half of the number of supersymmetries. A BPS state in $\mathcal{N}$-SUSY may be represented as a massive non-BPS state of $\mathcal{N}/2$-SUSY. For the simplest symmetry breaking pattern of the $\mathcal{N} = 4$ SYM Coloumb branch, the massless $SU(4)$ $R$-symmetry is broken to $USp(4)$ when the central charge is generated.

From (2.3), the massive supersymmetry algebra is that of $N$ fermionic oscillators, where $N = 2\mathcal{N}$ if the representation is not BPS, but can be reduced by up to a factor of 1/2 if shortened. Supermultiplets may be represented as coherent states which are eigenstates of $N$ 'lowering operators'. To build these states we introduce Grassmann variables which transform as fundamental spinors of the little group of each particle $\eta_{i,I}^A$, as well as their conjugates $\eta_{i,A}^{\dagger I}$. The $R$-index on the Grassmann variables is truncated for 1/2-BPS states to denote some subset of the $\mathcal{N}/2$ supersymmetries that do not leave the state invariant. We will use the fact that BPS states of $\mathcal{N} = 4$ obey the same algebra as the non-BPS state of $\mathcal{N} = 2$, which simplifies its construction.

To ensure little group covariance, we choose all of the $q_{i,I}^{\dagger A}$ as the lowering operators. An entire supermultiplet may be encoded as a coherent state

$$\langle \eta_i | = \langle \Omega | \, e^{q_{i,A}^I \eta_{i,I}^A} \tag{2.6}$$

where $\eta_{i,I}^A$ are anticommuting Grassmann algebra generators and $\langle\Omega|$ is the Clifford vacuum annihilated by $q_{i,I}^{\dagger A}$. These are eigenstates of the annihilation operators, satisfying $\langle\eta_i|\, q_{i,I}^\dagger = \langle\eta_i|\,(-\eta_{i,I})$. The action of the supercharges on the coherent states may be represented as

$$q_{i,I}^{\dagger A} = -\eta_{i,I}^A \qquad q_{i,A}^I = -\frac{\partial}{\partial\eta_{i,I}^A}. \tag{2.7}$$

Supersymmetry transformations generated by $q$ and $q^\dagger$ act simply on these coherent states:

$$\langle\eta|\, e^{i\xi_A^{\dagger I} q_{i,I}^{\dagger A}} = e^{-i\xi_A^{\dagger I}\eta_I^A}\,\langle\eta|\,, \qquad \langle\eta|\, e^{-i\xi_I^A q_{i,A}^I} = \langle\eta + i\xi|\,. \tag{2.8}$$

Here, $\xi_I^A = \left[\theta^A i_I\right]$ and $\xi_A^{\dagger I} = \left\langle\theta_A i^I\right\rangle$ parameterise the supersymmetry transformation projected onto the spinors of leg $i$ of the appropriate chirality, for some Grassmann spinors $\left[\theta^A\right|$ and $\left\langle\theta^A\right|$. The action of the supercharges encoded in (2.8) give the supersymmetric Ward identities (SWIs) relating the components.

Only elementary massive vector multiplets will be of interest to us in our investigation of scattering amplitudes on the Coulomb branch of $\mathcal{N}=4$. These are half-BPS, which are equivalent to long $\mathcal{N}=2$ vector multiplets. Expanding the $\mathcal{N}=2$ coherent state gives the superfield

$$\mathcal{W} = \phi + \eta_I^a\psi_a^I - \frac{1}{2}\eta_I^a\eta_J^b(\epsilon^{IJ}\phi_{(ab)} + \epsilon_{ab}W^{(IJ)}) + \frac{1}{3}\epsilon_{bc}\eta_I^b\eta_J^c\eta^{Ja}\tilde{\psi}_a^I + \eta_1^1\eta_1^2\eta_2^1\eta_2^2\tilde{\phi}, \tag{2.9}$$

See (29) for details. The $R$-indices $a, b, c$ are those of the $SU(2)_R$ of the $\mathcal{N}=2$ SUSY algebra. The states $\phi$, $\tilde{\phi}$ and $\phi_{(ab)}$ represent 5 scalar quanta, $\psi_a^I$ and $\tilde{\psi}_a^I$ represent the degrees of freedom of two Dirac fermions, while $W^{(IJ)}$ represents the spin triplet of massive vector states. This superfield and its massless limit will be discussed further in Section 2.3.2.

## 2.3 On-Shell Superspace for $\mathcal{N}=4$ Coulomb Branch

### 2.3.1 Non-Chiral Superspace

The massless supermultiplet of $\mathcal{N}=4$ at the origin of moduli space is commonly constructed in the 'chiral superspace' in which it is represented as a coherent state of $\eta^A$ for $SU(4)$ index $A$ (e.g. see for review (31)). These carry massless $U(1)$ helicity weights. This leads to a superfield[1]

$$G^+ = g^+ + \eta^A \lambda_A^+ - \frac{1}{2}\eta^A\eta^B S_{AB} - \frac{1}{6}\eta^A\eta^B\eta^C\lambda^-_{ABC} + \eta^1\eta^2\eta^3\eta^4 g^-, \tag{2.10}$$

where the superscript on the superfield labels the helicity of the supermultiplet. This contains the gluon $g^{\pm}$, four chiral gauginos with positive and negative helicities $\lambda_A^+$ and $\lambda^-_{ABC}$ respectively (the latter is totally antisymmetric in its $R$-indices and has only four independent components) and three complex scalars $S_{AB}$ satisfying self-duality $S_{AB} = \frac{1}{2}\epsilon_{ABCD}S^{*CD}$.

However, we will find in what follows that for the supercharges to be represented as homogeneously multiplicative or derivative on the superfields in the presence of massive BPS states, we are led to construct the massless multiplets in the 'non-chiral superspace', introduced in (24). To find the non-chiral superspace representation of the massless multiplet, we may perform a 'half-Fourier transform' from $\eta^3, \eta^4$ to $\eta_3^\dagger, \eta_4^\dagger$. This construction is natural from the perspective of the dimensional reduction of 6d $\mathcal{N}=(1,1)$ SYM to 4d $\mathcal{N}=4$ SYM, as used in (24; 25; 27) (also see (32) for developments of on-shell superspaces for similar $4d$ and $6d$ theories on brane world-volumes). The massless superfield will now be a coherent state expanded in $\eta^a$, for $a=1,2$, and $\eta_m^\dagger$, for $m=3,4$. The manifest massless $R$-symmetry is thus reduced to $SU(2)\times SU(2)$, although the multiplet

[1] We express this in the form of (31), defining the phases of the states to be those necessary to produce this from the action of $q_{i,A}$ on the Clifford vaccuum.

remains $SU(2,2)$ invariant. This form was used for the non-chiral superspace of (25).

However, this $SU(2) \times SU(2)$ is not a subgroup of the unbroken $R$-symmetry group $USp(4)$ (or $USp(2,2)$ after the half-Fourier transform), so will be broken in the superamplitudes on the Coulomb branch. Instead, as will become clearer in our discussion of BPS multiplets below, we will find it more useful to manifest a representation of a $U(2) \leq USp(4)$, under which the fundamental $USp(4)$ vector decomposes as $\mathbf{4} \to \mathbf{2} \oplus \bar{\mathbf{2}}$. Then $\eta^a$ and $\tilde{\eta}^{\dagger a} = \eta^\dagger_{a+2}$ both transform in the $\mathbf{2}$ representation of this $U(2)$ subgroup. In this notation, heights of the $R$-indices on the states in (2.10) are reversed for $A = 3, 4$ to show explicit $U(2)$ invariance of the superfield. The supermultiplet in the non-chiral superspace, first in the form of (25) with the manifest broken $SU(2) \times SU(2)$ and second in the form with the (partially) manifest $U(2)$, is

$$\begin{aligned} G &= -\frac{1}{2}S^m_{\ m} + \eta^\dagger_m \lambda^{+m} + \frac{1}{2}\eta^a \lambda^{-\ m}_{am} - \frac{1}{2}\eta_a \eta^a g^- + \frac{1}{2}\eta^\dagger_m \eta^{\dagger m} g^+ \\ &\quad + \eta^a \eta^\dagger_m S^m_{\ a} + \frac{1}{2}\eta^\dagger_m \eta^{\dagger m} \eta^a \lambda^+_a + \frac{1}{4}\eta_a \eta^a \eta^\dagger_m \lambda^{-mb}{}_b - \frac{1}{4}\eta^\dagger_m \eta^{\dagger m} \eta_a \eta^a S_b{}^b \\ &= S_{34} + (\tilde{\eta}^{\dagger 1}\lambda^+_4 - \tilde{\eta}^{\dagger 2}\lambda^+_3) + \eta^a \lambda^-_{a34} - \eta^1 \eta^2 g^- + \tilde{\eta}^{\dagger 1}\tilde{\eta}^{\dagger 2} g^+ \\ &\quad + \eta^a(\tilde{\eta}^{\dagger 1} S_{4\,a} - \tilde{\eta}^{\dagger 2} S_{3\,a}) + \tilde{\eta}^{\dagger 1}\tilde{\eta}^{\dagger 2}\eta^a \lambda^+_a + \frac{1}{2}\eta^a \eta^b (\tilde{\eta}^{\dagger 2}\lambda^-_{ab3} - \tilde{\eta}^{\dagger 1}\lambda^-_{ab4}) + \tilde{\eta}^{\dagger 2}\tilde{\eta}^{\dagger 1}\eta^1 \eta^2 S_{12}. \end{aligned} \tag{2.11}$$

The latter form will be henceforth assumed, although this will not actually be very important in what follows. In the former expression, index heights in each $SU(2)$ sector may be raised and lowered with the Levi-Civita symbol as usual. However, in the latter form, $G$ is charged under a $U(1) \leq U(2)$ subgroup. Each Grassmann variable carries a unit charge under a $U(1)$ subgroup, while the states are also charged such that each term above has an overall charge of $+2$ units. While possible to adjust the notation to make the $SU(2) \leq U(2)$ invariance manifest, we find that, in practice, the above form is clearest (these expressions are mostly useful for identifying extraction functions to find

component amplitudes).

The superfield (2.11) is the massless counterpart to the massive superfield in (2.9). The correspondance between the massless and massive on-shell superspace variables will be elaborated upon below.

The (complexified) $R$-symmetry generators for the $USp(2,2)$ on the non-chiral superspace are

$$\mathbf{m}_b^a = \sum_i \left( \tilde{\eta}_i^{\dagger a} \frac{\partial}{\partial \tilde{\eta}_i^{\dagger b}} + \eta_i^a \frac{\partial}{\partial \eta_i^b} - 2\delta_b^a \right) \qquad \mathbf{k}_{ab} = \sum_i \frac{\partial}{\partial \tilde{\eta}_i^{\dagger (a}} \frac{\partial}{\partial \eta_i^{b)}} \qquad \mathbf{p}^{ab} = \sum_i \tilde{\eta}_i^{\dagger (a} \eta_i^{b)}. \tag{2.12}$$

The symbols have been chosen to reflect the resemblance to the conformal group. All massless legs $i$ are summed over. The reader is referred to (25) for a larger catalogue of symmetry generator representations for the massless superfields in the non-chiral superspace.

Pure $\mathcal{N}=4$ super-Yang-Mills theory has a supersymmetry-preserving moduli space of vacua upon which the scalar components of the vector supermultiplets acquire a vev and spontaneously break the gauge theory to some smaller rank unbroken subgroup. We will generally consider the possibility of multiple breakings of the gauge group to factors of $\prod_k U(N_k)$. For simplicity, we will assume that the scalars' vevs are of the form $\langle S_{AB} \rangle = \oplus_k v_k \delta_{i_k}^{j_k} \Omega_{AB}$ for some $v_k \in \mathbb{R}$. Here, $i_k$ and $j_k$ are gauge indices of an unbroken $U(N_k)$ subgroup. This breaking pattern induces a central charge $Z_{AB} \propto \Omega_{AB}$ and modifies the SUSY algebra to the form discussed above. The $R$-symmetry in this case is broken to $USp(4)$, which corresponds to the simplest case in which there is only a single central charge. The vector superfields that become massive through this Higgsing are

BPS states and are bifundamentals of two of the unbroken gauge group factors. Calling these $U(N_{k_a}) \times U(N_{k_b})$, then their masses are $g|v_{k_a} - v_{k_b}|$, where $v_{k_a}$ and $v_{k_b}$ are the vevs that break the generators corresponding the the vector superfields. Conservation of the central charge then implies that, in any scattering process, the sum of the masses of the particles (states of positive central charge) must be equal to the sum of the masses of the antiparticles (states of negative central charge). This selection rule places an extra kinematic constraint upon the amplitudes.

### 2.3.2 BPS States

In the BPS case, the supersymmetry generators satisfy the reality condition

$$P_i^{\dot{\alpha}\alpha} Q_{i,\alpha,A} = \frac{1}{2} Z_{i,AB} Q_i^{\dagger,B\dot{\alpha}}, \tag{2.13}$$

which implies (2.5) when the little group symmetry is made manifest. This reduces the effective number of left-handed fermionic generators from $\mathcal{N}$ to $\mathcal{N}/2$. We use these remaining $\mathcal{N}/2$ generators to construct 'short' BPS supermultiplets that are equivalent to the 'long' massive supermultiplets of unextended $\mathcal{N}/2$ supersymmetry. For the Coulomb branch of $\mathcal{N} = 4$ SYM, the massive multiplets will all be short multiplets, which are equivalent to the $\mathcal{N} = 2$ multiplet given in (2.9).

There is a choice in how to represent the BPS SUSY algebra, which corresponds to a choice of raising and lowering operators for our supermultiplets. This affects the organization both of states and of superamplitudes in theories with BPS multiplets, such as $\mathcal{N} = 4$ on the Coulomb branch. Given our intent, it seems natural that we should make the choice which preserves manifest little group covariance of our BPS states. The formulation of an $\mathcal{N} = 2$ theory with BPS multiplets may be understood analogously, so we focus predominantly here on what happens for $\mathcal{N} = 4$. A similar on-shell superspace

for $\mathcal{N}=8$ supergravity incorporating half-BPS black holes was recently constructed in (33).

Firstly, on the BPS states, the supercharges satisfy $q_i^{\dagger I,A} = -q_i^{I,A}$. It is at this point that the breaking of the $R$-structure of the supercharges into the non-chiral form discussed above for massless representations becomes natural for describing the BPS states. For $\mathcal{N}=4$, after decomposing the supercharges into two separate pairs independently transforming under the $U(2)$ $R$-subgroup, the BPS condition equates supercharges of one doublet with the conjugates of the other, which are in the same $U(2)$ representation. The massive BPS on-shell superfield may then be expanded in two little group pairs of Grassmann variables $\eta_I^a$ (for $a \in \{1,2\}$), just as for the massive $\mathcal{N}=2$ superfield derived above. The supercharges are then represented on these as $q_{i,I}^{\dagger a} = -q_{i,I,a+2} = -\eta_{i,I}^a$ and $q_{i,a}^I = q_i^{\dagger I,a+2} = -\frac{\partial}{\partial \eta_{i,I}^a}$, for $a \in \{1,2\}$.

The anti-BPS superfields consist of the $CP$ conjugate states of the BPS superfields. For the anti-BPS states, the same coherent state basis may be selected, although, as the central charge has the opposite sign, the anti-BPS condition involves a relative negative sign $q_i^{\dagger Ia} = q_i^{Ia}$. This leaves a relative negative sign in the representations of the supercharges on the superspace compared to the BPS states. States at level $n$ in the BPS superfield are conjugate to states at level $\mathcal{N}-n$ in the anti-BPS superfield.

While only a $U(2)$ subgroup of the $R$-symmetry is manifest on the BPS multiplets, the full $USp(4)$ is still respected by the superamplitudes. The (complexified) $USp(4)$ $R$-symmetry generators (or, more precisely, $USp(2,2)$) represented on massive superfields are

$$\mathbf{m}_b^a = \sum_i \left(\eta_{i,I}^a \frac{\partial}{\partial \eta_{i,I}^b} - 2\delta_b^a\right) \qquad \mathbf{k}_{ab} = \frac{1}{2}\sum_i \pm \frac{\partial}{\partial \eta_{i,I}^{(a}} \frac{\partial}{\partial \eta_i^{b),I}} \qquad \mathbf{p}^{ab} = \frac{1}{2}\sum_i \pm \eta_{i,I}^{(a} \eta_i^{b),I}. \tag{2.14}$$

See discussion of the representation theory of the symplectic groups in (34). The (+) in the **k** and **p** generators is for BPS legs and the (−) is for anti-BPS. Note that the little group index on the Grassmann derivative is raised and lowered by $-\epsilon$ rather than $\epsilon$, so e.g. $\left|i^I\right] \frac{\partial}{\partial \eta^a_{i,I}} = |i_I] \frac{\partial}{\partial \eta_i^{I,a}}$. The expressions for the generators on the massless legs in (2.14) should be combined with those stated above to obtain the representation of the full superamplitude.

The massless limit of the BPS superfield in the form (2.9) produces the non-chiral representation of the massless superfield (2.11), which makes clearer why this representation is natural when formulating Coulomb branch superamplitudes. In this limit, the two supermultiplets are related as

| | | | | | | |
|---|---|---|---|---|---|---|
| Massive | $\phi$ | $\psi^I_a$ | $W^{IJ}$ | $\phi_{ab}$ | $\tilde{\psi}^I_a$ | $\tilde{\phi}$ |
| Massless | $S_{34}$ | $\lambda^-_a, \lambda^+_{a+2}$ | $g^\pm, S_{13}+S_{24}$ | $S_{14}, S_{13}-S_{24}, S_{23}$ | $\lambda^+_a, \lambda^-_{a+2}$ | $S_{12}$ |

The massless limit of the $\mathcal{N}=4$ BPS superfield therefore amounts to breaking up the little group indices, as we are familiar with in the non-supersymmetric case. We here send $\eta^A_- \to \eta^a$, $\eta^A_+ \to \tilde{\eta}^{\dagger a}$ (where $A$ here is the $\mathcal{N}=2$ $R$-index used in (2.9)). For the anti-BPS states, as a consequence of our definition of the massive superspace variables given above, the massless limit is modified to $\eta^A_+ \to -\tilde{\eta}^{\dagger a}$, as is required from the inverse relations implied by (2.2). The $R$-symmetry generators (2.14) clearly match onto (2.12). The fact that our covariant representation of the BPS state reduces to a mixed representation of the massless coherent state with a scalar Clifford vacuum suggests that this mixed (or non-chiral) representation may be useful for representing amplitudes on the Coulomb branch of $\mathcal{N}=4$. Previous works have instead (18; 19) implicitly worked with a massive representation that manifested an $SU(2)\times SU(2)$ subgroup[2] of the $USp(4)$ $R$-symmetry in which the massive little group was obscured. This representation led to massive

[2]This is a distinct subgroup from the broken $SU(2)\times SU(2)$ mentioned in the discussion preceding (2.11).

coherent states that appear similar to the first expression in (2.11), but with $R$-indices broken into pairs $(\eta^1,\eta^3)$ and $(\eta^2,\eta^4)$ and the vector's longitudinal mode replacing a single scalar. Similar tension in manifesting $R$-symmetries and little group symmetries in on-shell superspaces arises in $6d$ (22). Here we note that the BPS states $\mathcal{W}$ on the Coulomb branch are not self-conjugate (being eigenstates of the central charge) and so their massless limits are likewise complex.

The choice of non-chiral coherent state for the massless fields combines with the coherent state bases for the BPS states to ensure that the total supercharges $Q^{\dagger a}$ and $Q_{a+2}$ act multiplicatively on the superamplitudes (while their conjugates act on each leg homogeneously as derivatives). The full supercharges are therefore represented as

$$\begin{aligned}
\frac{1}{\sqrt{2}}Q_a &= |i_I]\,\frac{\partial}{\partial\eta^a_{i,I}} + |j_I]\,\frac{\partial}{\partial\eta^a_{j,I}} + |k]\,\frac{\partial}{\partial\eta^a_k}, & \frac{1}{\sqrt{2}}Q^{\dagger a} &= -\left|i^I\right\rangle\eta^a_{i,I} - \left|j^I\right\rangle\eta^a_{j,I} + |k\rangle\,\eta^a_k,\\
\frac{1}{\sqrt{2}}Q_{a+2} &= \left|i^I\right]\eta^a_{i,I} - \left|j^I\right]\eta^a_{j,I} + |k]\,\tilde{\eta}^{\dagger a}_k, & \frac{1}{\sqrt{2}}Q^{\dagger a+2} &= |i_I\rangle\,\frac{\partial}{\partial\eta^a_{i,I}} - |j_I\rangle\,\frac{\partial}{\partial\eta^a_{j,I}} + |k\rangle\,\frac{\partial}{\partial\tilde{\eta}^{\dagger a}_k},
\end{aligned} \tag{2.15}$$

where legs labeled $i$, $j$ and $k$ respectively enumerate $\mathcal{W}$, $\overline{\mathcal{W}}$ and $G$ legs and are implicitly summed over here. With the supercharges in this homogeneous form, the SWIs should be simplified.

While not considered here for simplicity, it is also possible to consider further breakings of the $R$-symmetry on the $\mathcal{N}=4$ Coulomb branch, by moving the vevs of the other scalar components away from the origin of the moduli space. As operators acting on external legs of elementary vector supermultiplets, the central charge eigenvalues $Z_{i,AB}$ may always be $SU(4)$ $R$-rotated into a form $Z_{i,AB}=z_i\Omega_{AB}$ and the BPS condition is unchanged. See (35–37) for discussion. However, if the $R$-symmetry is broken beyond $USp(4)$, this rotation is leg-dependent and the form of the supercharges represented on

the full superamplitude (and hence the SWIs) will be more complicated. On-shell representations of BPS states with more complicated configurations of central charges were recently discussed in (33) in the context of $\mathcal{N} = 8$ SUGRA for BPS black holes.

### 2.3.3 Superamplitude preliminaries

Our ultimate ambition is to construct an arbitrary $n$-point amplitude with both massless and massive external states, $\mathcal{A}_n(\mathcal{W}_1, \mathcal{W}_2 \ldots, \overline{\mathcal{W}}_j, \overline{\mathcal{W}}_{j+1} \ldots, G_k, G_{k+1}, \ldots)$. As is conventional in discussions of scattering amplitudes in gauge theories, we will be henceforth implicitly describing colour-stripped partial amplitudes $\mathcal{A}_n[\mathcal{W}_1, G_2, G_3 \ldots \overline{\mathcal{W}}_j, G_{j+1} \ldots]$, in which the ordering of the external legs is fixed. The full tree-level superamplitude is then obtained in the usual way by summing over all non-cyclic permutations of external legs and multiplying each partial amplitude with a single colour trace over the gauge group generators corresponding to each external leg in the order that they appear. See e.g. (38). In the case of interest here, some simple structure to the non-zero colour-traces can be used to identify possible orderings of the massive and massless vector multiplets.

As discussed in (18), because of the bifundamental nature of the massive vector multiplets with respect to the unbroken gauge group factors, partial amplitudes must be of the form $\mathcal{A}_n[\mathcal{W}_{mi}, G_i, G_i, \ldots \overline{\mathcal{W}}_{ij}, G_j, G_j, \ldots \mathcal{W}_{jn}, \ldots]$. Here, $G_i$ and $G_j$ are massless vectors of different unbroken gauge subgroups $SU(N_i)$ and $SU(N_j)$ respectively, while e.g. $\overline{\mathcal{W}}_{ij}$ has one fundamental $SU(N_i)$ index and one antifundamental $SU(N_j)$ index, so must be ordered to the left of a string of $G_j$ fields and to the right of a string of $G_i$ fields. The strings of massless vectors (of possibly zero length) can only terminate at a massive vector field with opposite index structure. Note that the overbar on the massive vectors merely distinguishes those with negative central charge ("anti-BPS") from those with positive central charge. BPS and anti-BPS vectors need not alternately appear in

the colour-ordered partial amplitudes for a general breaking pattern of the gauge group, but both must be present. In the subsequent discussion, we will not bother to distinguish between the vector multiplets belonging to different gauge subgroups, but will leave this implicit and fully encapsulated in the stripped colour trace.

Having established the colour-structure of the superamplitudes, we are now able to focus our attention on the more interesting kinematic structure of the superamplitudes with massive multiplets. The first feature to note is that all Coulomb branch superamplitudes $\mathcal{A}_n$ will be of homogeneous Grassmann degree $2n$ in our representation. This is a consequence of the $U(1)$ factor of the explicit $U(2) \leq USp(4)$ represented on the massive on-shell superspace. This subgroup is generated by the trace of the $\mathbf{m}_b^a$ generators in (2.12) and (2.14). As the vector bosons are $R$-invariant and the Grassmann variables carry a unit of charge under this generator, the massive superfield (2.9) must carry 2 units of this $R$-charge. As the component amplitudes must conserve this charge, the 2 units per leg in the superamplitude must be instead carried by accompanying Grassmann variables.

In this non-chiral superspace, the helicity-violating sectors of the massless superamplitudes appear as terms with Grassmann variables divided differently between $\eta^a$ and $\tilde{\eta}^{\dagger a}$ factors. This is clear from the contributions to the supercharges from the massless legs in (2.15), where $Q^{\dagger a+2}$ and $Q_a$ will not mix the sectors of definite helicity violation. However, both types of supercharges act on the massive Grassmann variables, so this structure is not respected by the massive legs. This is to be expected, because helicity is no longer a frame-independent property for massive particles. We will discuss how mass affects the sectors further below once we begin to compute higher leg amplitudes.

Finally, we adopt the convention that all particles are outgoing and that incoming states may be obtained by crossing outgoing legs. Under crossing, an outgoing leg of momentum $p$ is analytically continued to an incoming leg of momentum $-p$ and opposite

central charge. The mass of the leg is unchanged, but a negative sign now accompanies its appearance in the Weyl equation and the spin sums (see (29) for relevant identities in the conventions employed here). This is commented upon further below.

## 2.4 Massive Super-BCFW Recursion

### 2.4.1 Massless Super-BCFW

We will demonstrate below that supersymmetry fully determines the superamplitude with three external states. With more legs, supersymmetry is not enough and further properties of the $S$-matrix are required. To make progress in constructing higher-leg superamplitudes we will make use of BCFW recursion at tree level (11; 12). A BCFW shift on legs $i$ and $j$ consists, at the level of momenta, of finding a (complex) vector $r^\mu$ such that $p_i \cdot r = p_j \cdot r = r \cdot r = 0$, and shifting the two momenta to $p_i^\mu \to \hat{p}_i^\mu = p_i^\mu + z r^\mu$, $p_j^\mu \to \hat{p}_j^\mu = p_j^\mu - z r^\mu$, with $z$ a complex parameter. Note that this also necessitates shifting the polarisations of $i$ and $j$ as well, to maintain transversity. For massless legs, both of these deformations may be formulated simply at the level of spinors. An $[i, j\rangle$-shift is realised on the spinors as $|\hat{i}] = |i] + z\,|j]$ and $|\hat{j}\rangle = |j\rangle - z\,|i\rangle$ (so the shift vector $r = -\,|j]\,\langle i|$). A shift is called valid if the amplitude vanishes as $z \to \infty$. Cauchy's theorem then relates the value of the unshifted amplitude to a sum over complex poles of the shifted amplitude, which by tree-level unitarity occurs on on-shell factorization channels.

The supersymmetric extension of on-shell recursion, known as super-BCFW (10; 39; 40), allows us to construct full superamplitudes recursively. It has been shown that any amplitude of pure Yang-Mills and matter containing a negative helicity gluon is on-shell constructible under a BCFW shift (41). For $\mathcal{N} = 4$ at the origin of moduli space, the fact

that the other states are related supersymmetrically to the negative helicity gluon suffices to show that all superamplitudes are constructible using a supersymmetric extension of BCFW (40). These arguments do not rely on the masslessness of the other legs of the superamplitude and consequently this shows that any Coulomb branch superamplitude which has two massless legs is on-shell constructible under a super-BCFW shift.

The supersymmetrised BCFW-shift involves the standard BCFW shift described above supplemented with a shift in Grassmann variables to preserve the supercharge. For an $[i, j\rangle$-shift, the Grassmann variables are also shifted to $\hat{\eta}_i^A = \eta_i^A + z\eta_j^A$ in the chiral superspace. This may be derived by deducing the necessary shift in the supercharge $Q_i^{\dagger A} = \sqrt{2}\,|i\rangle\,\eta_i^A$ carried by leg $i$ resulting from demanding both that the total supercharge be conserved and that the SUSY algebra (2.1) be preserved (note that the derivatively represented $Q_{i,A}$ must also shift).

The standard super-BCFW shift may be converted into a form where it may be used in the non-chiral superspace. This can be obtained by half-Fourier transforming the shifted superamplitude in the chiral superspace. To implement a $[i, j\rangle$-supershift, the momentum shift is unchanged from that described above, while the Grassmann variables shift as $\hat{\eta}_i^a \to \eta_i^a + z\eta_j^a$ and $\hat{\tilde{\eta}}_j^{\dagger a} \to \tilde{\eta}_j^{\dagger a} - z\tilde{\eta}_i^{\dagger a}$. Constructibility continues to hold in this superspace, as the half-Fourier transform from the chiral superspace does not affect the large $z$ scaling of the superamplitude with shifted momentum.

### 2.4.2 Massive BCFW

While this standard super-BCFW shift is a powerful tool for constructing higher-leg Coulomb branch superamplitudes, it leaves open the question of constructing fully massive Coulomb branch superamplitudes. One path toward the on-shell construction of such superamplitudes is to formulate a supershift on massive legs. BCFW recursion for

massive legs has been introduced in (42) and (43). As in the massless case, the momenta shift as

$$\hat{p}_i^\mu = p_i^\mu + z r^\mu, \qquad \hat{p}_j^\mu = p_j^\mu - z r^\mu, \tag{2.16}$$

where $r$ has the same orthogonality properties as in the massless case. To construct the shift vector $r$, we find a little group frame for each particle where we can write $p_i$ and $p_j$ as linear combinations of the same two null vectors. Geometrically, these correspond to the two null vectors being coplanar with both massive momenta. Finding this little group frame requires solving

$$-\left|i^2\right]\left\langle i^1\right| = \frac{\alpha_i}{m_j^2}\left|j^1\right]\left\langle j^2\right|, \qquad -\left|j^2\right]\left\langle j^1\right| = \frac{\alpha_j}{m_i^2}\left|i^1\right]\left\langle i^2\right| \tag{2.17}$$

to find $\alpha_i = \alpha_j \equiv \alpha$, where

$$\alpha = -p_i \cdot p_j + \sqrt{(p_i \cdot p_j)^2 - m_i^2 m_j^2} \tag{2.18}$$

$$p_i = \left|i^1\right]\left\langle i^2\right| + \frac{\alpha}{m_j^2}\left|j^1\right]\left\langle j^2\right|, \qquad p_j = \left|j^1\right]\left\langle j^2\right| + \frac{\alpha}{m_i^2}\left|i^1\right]\left\langle i^2\right|. \tag{2.19}$$

Up to a single ambiguous phase, the spinors of each leg may be related in this special frame by

$$\begin{aligned} \left|i^1\right] &= \frac{m_i}{\sqrt{\alpha}}\left|j^2\right] \qquad & \left|i^2\right] &= -\frac{\sqrt{\alpha}}{m_j}\left|j^1\right] \\ \left|i^2\right\rangle &= -\frac{m_i}{\sqrt{\alpha}}\left|j^1\right\rangle \qquad & \left|i^1\right\rangle &= \frac{\sqrt{\alpha}}{m_j}\left|j^2\right\rangle. \end{aligned} \tag{2.20}$$

In this special little group frame, it is clear that we may take

$$r = \left|i^1\right]\left\langle j^2\right| \text{ or } r = \left|j^1\right]\left\langle i^2\right| \tag{2.21}$$

and satisfy the orthogonality requirements $p_i \cdot r = p_j \cdot r = r \cdot r = 0$ (17). It is clear that $r$ cannot be regarded merely as a function of the massive momenta $p_1$ and $p_2$, as it is determined by only a single helicity spinor associated to each. Its selection explicitly breaks little group invariance of the legs, as its existence relies on this preferred null vector decomposition.

The massive BCFW recursion may be illustrated on a simple example. Bhabha scattering in scalar QED is a constructible example, provided that, in the Lagrangian picture, there is a quartic scalar interaction with $-\frac{1}{2}e^2(\phi^*\phi)^2$ for electric charge $e$ (calling $\phi$ the scalar field) (31). The validity of the shift may be verified by derivation from the Feynman rules, from which it can be shown that the shifted amplitude $A(\phi, \phi^*, \phi, \phi^*) \to 0$ as $z \to \infty$. This is not unexpected, as this amplitude is well-known to be constructible by BCFW recursion when the scalars are massless, provided that the shifted particles have the same charge. Unlike for spinning particles, massive scalars do not carry more degrees of freedom than massless scalars. When the massive legs are spinning, the validity of recursion is expected to be less general. The validity of massive BCFW for QCD amplitudes with massive quarks was discussed in (43), which was spin-dependent. However, the case of massive scalars here does not introduce any substantial change.

Shifting the scalar $\phi$ legs 1 and 3, the amplitude is determined as a sum over two factorisation channels:

$$\begin{aligned} A(\phi_1, \phi_2^*, \phi_3, \phi_4^*) = &\sum_{h=+,-} \hat{A}(\hat{\phi}_1, \phi_2^*, \gamma^h_{\hat{P}_{12}}) \frac{-1}{s} \hat{A}(\hat{\phi}_3, \phi_4^*, \gamma^{-h}_{-\hat{P}_{12}})\Big|_{z_*^{(1)}} \\ &+ \sum_{h=+,-} \hat{A}(\hat{\phi}_1, \phi_4^*, \gamma^h_{\hat{P}_{14}}) \frac{-1}{u} \hat{A}(\hat{\phi}_3, \phi_2^*, \gamma^{-h}_{-\hat{P}_{14}})\Big|_{z_*^{(2)}}, \end{aligned} \tag{2.22}$$

where $\gamma$ is a photon and $\hat{P}_{12} = -\hat{p}_1 - p_2$ and $\hat{P}_{14} = -\hat{p}_1 - p_4$ are its (complex) momenta in each factorisation channel. The intermediate photon's helicity $h$ is summed over. The

unshifted Mandelstam variables are $s=-(p_1+p_2)^2$, $u=-(p_1+p_4)^2$ and $t=4m^2-s-u$, for scalar mass $m$. The poles $z_*^{(i)}$ are determined by finding the values of the shift parameter $z$ on which the shifted momenta are aligned on a factorisation channel, but their identity will not be necessary here.

At this point, we review an exceptional feature which appears in the special case of three-leg amplitudes with two massive, equal-mass particles and one massless particle, such as $\hat{A}(\phi_1,\phi_2^*,\gamma_3^{\pm})$. Introduced in (30), an additional object that carries helicity weight of the massless particle exists that may be used as an amplitude building block:

$$x \equiv \frac{1}{m}\frac{[q|\,p_2\,|3\rangle}{[q3]}, \tag{2.23}$$

where 3 is the massless leg, $m$ is the mass of legs 1 and 2, and $|q]$ is an arbitrary reference spinor defined so that $[q3]\neq 0$. This special case arises because $p_2\cdot p_3=-\langle 3|\,p_2\,|3]=0$, implying that $p_2\,|3]\propto|3\rangle$. The constant of proportionality is $x$ and carries helicity weight 1 of leg 3. It is independent of the reference spinor present in (2.23). See (29) for further details, conventions and identities.

The on-shell three-particle amplitudes in (2.22) are

$$A(\phi_1,\phi_2^*,\gamma_3^+)=\frac{em}{x} \tag{2.24}$$

$$A(\phi_1,\phi_2^*,\gamma_3^-)=emx. \tag{2.25}$$

Parity has been imposed. Denoting by $\hat{x}_{ij}$ the value of the $x$-factor at the shifted momentum in the three-leg amplitude with massive scalars $i$ and $j$, then for the purposes

here

$$\hat{x}_{12} = \frac{[q|\, p_2 \left|\hat{P}_{12}\right\rangle}{m\left[q\hat{P}_{12}\right]} = \frac{m\left\langle\rho\hat{P}_{12}\right\rangle}{\langle\rho|\, p_2 \left|\hat{P}_{12}\right]} \qquad \hat{x}_{34} = \frac{m\left\langle\rho(-\hat{P}_{12})\right\rangle}{\langle\rho|\, p_4 \left|(-\hat{P}_{12})\right]} = \frac{[q|\, p_4 \left|(-\hat{P}_{12})\right\rangle}{m\left[q(-\hat{P}_{12})\right]} \tag{2.26}$$

and similarly for $\hat{x}_{14}$ and $\hat{x}_{32}$. We leave implicit that these factors in (2.26) are to be evaluated on the pole $z = z_*^{(1)}$ while the others are on the $z = z_*^{(2)}$ pole. Here $|q]$ and $|\rho\rangle$ are reference spinors not aligned with the spinors of the internal momentum $\hat{P}_{12}$. The $x$-factors are independent of the reference spinors. With these expressions, the Bhabha scattering amplitude is then

$$\begin{aligned} A(\phi_1, \phi_2^*, \phi_3, \phi_4^*) &= \frac{-e^2m^2}{s}\left(\frac{\hat{x}_{12}}{\hat{x}_{34}} + \frac{\hat{x}_{34}}{\hat{x}_{12}}\right) + \frac{-e^2m^2}{u}\left(\frac{\hat{x}_{14}}{\hat{x}_{32}} + \frac{\hat{x}_{32}}{\hat{x}_{14}}\right) \\ &= \frac{-e^2}{s}\left(\frac{[q|\, p_2 \left|\hat{P}_{12}\right\rangle \left[(-\hat{P}_{12})\right| p_4\, |\rho\rangle}{\left[q\hat{P}_{12}\right]\left\langle\rho(-\hat{P}_{12})\right\rangle} + \frac{[q|\, p_4 \left|(-\hat{P}_{12})\right\rangle \left[\hat{P}_{12}\right| p_2\, |\rho\rangle}{\left[q(-\hat{P}_{12})\right]\left\langle\rho\hat{P}_{12}\right\rangle}\right) \\ &\quad + (2 \leftrightarrow 4) \\ &= e^2(2p_2\cdot p_4)\left(\frac{1}{s}+\frac{1}{u}\right) = e^2(2m^2 - t)\left(\frac{1}{s}+\frac{1}{u}\right). \end{aligned} \tag{2.27}$$

See below in (2.68) for spinor analytic continuation rules for negative momentum. Here, $\hat{P}_{ij}\cdot p_2 = \hat{P}_{ij}\cdot p_4 = 0$ on either complex pole ($\{(i,j) = (1,2),(1,4)\}$) imply that these momenta anticommute as bispinors, while the Clifford algebra has been used in the step in which the reference spinors cancel out when the two terms for each channel are added together. This calculation is almost identical to the gluing argument of (30).

Unlike in (super)-Yang-Mills, BCFW here merely automates the construction of the amplitude from its two possible factorisation channels. However, unlike massless gauge theories, the second factorisation channel of the amplitude does not automatically emerge from the first. While the on-shell three-particle amplitudes contain "non-local" kinematic

factors, these cancel in the sum over internal photon helicities, as explained in (30), along with the poles $z_*^{(i)}$. This happens regardless of the mass of the scalar legs. Foretelling further results below, at no point was the identity of the shift vector necessary in this computation. As the only source of little group violation, it cancelled-out in the end, being eliminated within each term in the BCFW expansion as part of the cancellation of the kinematic denominators upon each residue.

### 2.4.3 Massive Super-BCFW

Massless super-BCFW recursion has been established in 6 dimensional (44), (21) (and higher (45)) super-Yang-Mills. In $6d$, the extra dimensions allow for extra directions in which the shift vector can point. As a result, the possible shift vectors are parameterised by an arbitrary variable in the massless $6d$ little group $SU(2) \times SU(2)$ (as it is effectively like a polarisation vector of one of the states).

The Coulomb branch of $4d$ SYM is equivalent to the low energy limit of the $6d$ theory after dimensional reduction on a torus (with fluxes providing the masses (46)). The masses of the BPS states can be identified with the momenta in the compactified directions. The form of the supershift constructed here corresponds to the dimensional reduction of the $6d$ supershift defined in (21), having made the choice to align the six-dimensional shift vector along the four non-compact dimensions so that the $4d$ shift vector remains null. This reduces the possible $6d$ shifts to the two possibilities in $4d$ discussed above. It is presumably also possible to construct a super-shift for the Coulomb branch in which includes shifts to the masses. In the following, we will construct massive super-BCFW in $4d$ purely from consistency with the symmetry algebra and the non-supersymmetric shift constructed above.

In order to make the momentum shift supersymmetric, the supercharges of each leg

must be deformed in order to preserve both the SUSY algebra (2.1) and the BPS constraint (2.3.2). Demanding that the total supercharge still be conserved, the supercharges of the shifted legs become

$$\begin{aligned}\frac{1}{\sqrt{2}}\hat{Q}_{i,a+2} &= \frac{1}{\sqrt{2}}Q_{i,a+2} + \frac{z}{2}\Delta Q_{a+2} \qquad & \frac{1}{\sqrt{2}}\hat{Q}_{j,a+2} &= \frac{1}{\sqrt{2}}Q_{j,a+2} - \frac{z}{2}\Delta Q_{a+2} \\ \frac{1}{\sqrt{2}}\hat{Q}_{i}^{\dagger a} &= \frac{1}{\sqrt{2}}Q_{i}^{\dagger a} + \frac{z}{2}\Delta Q^{\dagger a} \qquad & \frac{1}{\sqrt{2}}\hat{Q}_{j}^{\dagger a} &= \frac{1}{\sqrt{2}}Q_{j}^{\dagger a} - \frac{z}{2}\Delta Q^{\dagger a}. \end{aligned} \tag{2.28}$$

The derivatively represented supercharges in (2.15) also shift.

The shift spinors above may be expanded in a basis of Grassmann variables (or their derivatives) and spinors. The commutation relations and the BPS constraints may then be imposed in order to determine the coefficients. We will give the supercharge shift assuming that leg $i$ is BPS and leg $j$ is anti-BPS. All other particle/anti-particle configurations are also possible, but conservation of central charge implies that this configuration will at least always be available in any superamplitude. Explicitly choosing the special little group frame selected by the momentum shift and considering only $r = |i^1]\langle j^2|$ for simplicity, the supercharges can be determined to shift as

$$\Delta Q_{a+2} = -\frac{2m_i m_j}{\alpha + m_i m_j}\left|i^1\right]\left(\eta_{j1}^a + \frac{\sqrt{\alpha}}{m_i}\eta_{i2}^a\right) \tag{2.29}$$

$$\Delta Q^{\dagger a} = -\frac{2m_i m_j}{\alpha + m_i m_j}\left|j^2\right\rangle\left(\eta_{i2}^a - \frac{\sqrt{\alpha}}{m_j}\eta_{j1}^a\right). \tag{2.30}$$

The supercharges shift in the spinor directions singled-out by the momentum shift vector. Note that these expressions may be converted into a form consisting of $r$ multiplying a little group invariant spinor expression. All little group violation may be contained to the shift vector $r$.

In contrast to the massless case, the BCFW shift implemented at the level of spinors and Grassmann variables has an ambiguity. This is because, while the shifted spinors of

each leg are related through (2.20), there is no analogue for the Grassmann variables. It is therefore possible to shift these by the Grassmann variables of the same leg, in addition to those of the other. This affects the numerical prefactor multiplying the spinor shift. Choosing the Grassmann variables to shift only by terms proportional to those of the opposite shifted leg, the supershift may be represented as:

$$\left|\hat{i}^2\right] = \left|i^2\right] - z\frac{m_j\sqrt{\alpha}}{\alpha + m_i m_j}\left|i^1\right] \qquad \left\langle\hat{i}^2\right| = \left\langle i^2\right| + z\left\langle j^2\right|\frac{m_i m_j}{\alpha + m_i m_j} \tag{2.31}$$

$$\left|\hat{j}^1\right] = \left|j^1\right] - z\frac{m_i m_j}{\alpha + m_i m_j}\left|i^1\right] \qquad \left\langle\hat{j}^1\right| = \left\langle j^1\right| + z\left\langle j^2\right|\frac{m_i\sqrt{\alpha}}{\alpha + m_i m_j} \tag{2.32}$$

$$\hat{\eta}^a_{i,1} = \eta^a_{i,1} - z\frac{m_i m_j}{\alpha + m_i m_j}\eta^a_{j,1} \qquad \hat{\eta}^a_{j,2} = \eta^a_{j,2} - z\frac{m_i m_j}{\alpha + m_i m_j}\eta^a_{i,2} \tag{2.33}$$

and the other components are unaffected. We are again only showing here the case for the momentum shift $r = |i^1]\langle j^2|$.

The spinor-level shift of the massive legs may be re-expressed in a way that relates the little group violation directly to the momentum shift vector:

$$\begin{aligned}
|\hat{i}_I] &= |i_I] + \frac{z}{2m_i}\rho|i_I\rangle - \frac{z}{2m_i m_j}p_j\rho|i_I] \\
\langle\hat{i}^I| &= \langle i^I| + \frac{z}{2m_i}[i^I|\rho + \frac{z}{2m_i m_j}\langle i^I|\rho p_j \\
|\hat{j}_I] &= |j_I] + \frac{z}{2m_j}\rho|j_I\rangle - \frac{z}{2m_i m_j}p_i\rho|j_I] \\
\langle\hat{j}^I| &= \langle j^I| + \frac{z}{2m_j}[j^I|\rho + \frac{z}{2m_i m_j}\langle j^I|\rho p_i.
\end{aligned} \tag{2.34}$$

Here, $\rho \equiv \pm\left(m_i m_j/\sqrt{(p_i\cdot p_j)^2 - m_i^2 m_j^2}\right) r$ ((+) for $r = |i^1]\langle j^2|$, (−) for $r = |j^1]\langle i^2|$), or equivalently $r_{\alpha\dot{\beta}} = \frac{-1}{2m_i m_j}\rho_{\alpha\dot{\alpha}}(p_i p_j - p_j p_i)^{\dot{\alpha}}{}_{\dot{\beta}}$. The corresponding shifts of the Grassmann

variables are

$$\begin{aligned}\hat{\eta}^a_{i,I} &= \eta^a_{i,I} - \frac{z}{2m_i m_j}\Big([i_I|\rho|j^J\rangle - \langle i_I|\rho|j^J]\Big)\eta^a_{j,J} \\ \hat{\eta}^a_{j,J} &= \eta^a_{j,J} - \frac{z}{2m_i m_j}\Big([j_J|\rho|i^I\rangle - \langle j_J|\rho|i^I]\Big)\eta^a_{i,I},\end{aligned} \tag{2.35}$$

where we have here again assumed that leg $i$ is BPS and leg $j$ is anti-BPS, although shifts with both legs of the same type are also possible and differ only in changes of signs both in (2.34) and (2.35). In the ensuing calculations, we will not actually need any of these results beyond the existence of the momentum and supercharge shifts and their abstract properties. Rather, we merely state them here for completeness.

All Grassmann dependence of the superamplitudes arise in the form of the supercharges of each leg. Since the superfield legs are scalars, the supershift may be regarded entirely as a shift in momentum and supercharge by the null vector $r$ and chiral spinors presented above in (2.29) and (2.30). From this point of view, it is clear that the supershift vector and spinors do not obstruct the freedom in choosing little group decompositions of the momenta and supercharges of each unshifted leg. However, they provide a prefered null direction which singles out the little group frames in which both the shift vector and spinors have the especially simple forms (2.21), (2.29) and (2.30), leading to the apparent breaking of covariance in the spinor (2.34) and Grassmann level shifts (2.35). The shift spuriously breaks the little groups of the shifted legs by providing a special direction in which the massive momenta may be decomposed. Use of super-BCFW will therefore preserve little group invariance of the recursed superamplitudes up to explicit appearances of the shift vector $r$. However, as this is the only source of the breaking and the superamplitude itself must be invariant, all appearances of the shift vector must ultimately cancel to leave a manifestly invariant expression. This is similar to the $6d$ perspective, where the cancellation of the shift also inevitably follows from the arbitrary

and spurious choice of direction that must be made in choosing it.

This issue does not appear for $4d$ massless superamplitudes, where the bispinor form of the shift vector appears to manifestly break the $U(1)$ little group invariance. Because the residue on the complex pole scales as $z_\star^{(i)} \propto 1/r$ on each factorization channel $i$, the combination $z_\star^{(i)} r$ is a little group invariant function when $r$ is constructed out of massless helicity spinors.

As mentioned in (30), for general massive amplitudes, the combination of different helicity states in the little group covariant formalism can obstruct on-shell constructibility, as not all helicity components have the correct large-$z$ behavior. However, as will be discussed in Section 2.5.1, supersymmetry forces the Coulomb branch three-leg superamplitudes to contain the precise "nonlocality" needed for them to combine to give the pole structure of the four-leg superamplitude. This first hint of simple factorization properties remarkably extends to all Coulomb branch superamplitudes, as it turns out that all such superamplitudes are on-shell constructible via massive super-BCFW.

### 2.4.4 Validity

While the underlying origin is likely a vestige of dual (super)conformal invariance remaining on the Coulomb branch, we here leave an exploration of this to future work and instead prove the shift validity by using soft limits to extend the known behavior at the origin of moduli space. The idea that Coulomb branch component amplitudes may be found from soft limits of massless amplitudes with scalar insertions was proposed in (18), expanded upon in (19) and proven in (20). The precise map is explained clearly around (4.3) of (19), but the details will not be necessary for us. All we rely on is the fact that the Coulomb branch component amplitudes may be written as a sum over amplitudes at the origin of moduli space.

We may utilize this relation to show that a massive super-BCFW shifted Coulomb branch superamplitude $\mathcal{A}\left[\{\hat{\lambda}_{1I},\hat{\eta}^a_{1I}\},\{\hat{\lambda}_{2I},\hat{\eta}^a_{2I}\},\dots\right]$ has the correct large $z$ scaling for a valid shift (where we are borrowing the notation of (40) to highlight both the momentum spinors and Grassmann variables of each leg). The first step is to perform a $z$-independent supertranslation which sets $\hat{\eta}^a_{1I},\hat{\eta}^a_{2I}\to 0$:

$$\begin{aligned}\mathcal{A}\left[\{\hat{\lambda}_{1I},\hat{\eta}^a_{1I}\},\{\hat{\lambda}_{2I},\hat{\eta}^a_{2I}\},\{\lambda_{3I},\eta^a_{3I}\},\dots\right]&\\ =\mathcal{A}\left[\{\hat{\lambda}_{1I},0\},\{\hat{\lambda}_{2I},0\},\{\lambda_{3I},\eta^a_{3I}\mp\langle 3_I\zeta\rangle-[3_I\zeta]\}\dots\right]&\quad(2.36)\\ |\zeta\rangle=\frac{1}{s_{12}}\left(p_2\left|1^I\right]\eta^a_{1I}-p_1\left|2^I\right]\eta^a_{2I}-m_2\left|1^I\right\rangle\eta^a_{1I}+m_1\left|2^I\right\rangle\eta^a_{2I}\right)&\quad(2.37)\\ |\zeta]=\frac{1}{s_{12}}\left(-p_2\left|1^I\right\rangle\eta^a_{1I}-p_1\left|2^I\right\rangle\eta^a_{2I}+m_2\left|1^I\right]\eta^a_{1I}+m_1\left|2^I\right]\eta^a_{2I}\right),&\quad(2.38)\end{aligned}$$

where we have assumed that leg 1 is BPS and leg 2 is anti-BPS in our explicit solutions for $|\zeta\rangle, |\zeta]$, but an analogous procedure may be done for any two shifted legs with either sign central charge. This is a supersymmetry transformation which relates all component amplitudes to those with two of the lowest-weight states, which are here the scalars $\phi$. The existence of such a transformation that sets the shifted Grassmann variables to zero while not reintroducing $z$ into the other Grassmann variables was first pointed out in the massless case in (40).

Each massive component of (2.36) is given by (19) as a soft limit of a sum of massless component amplitudes with scalar insertions. Importantly, all the components of the translated superamplitude (2.36) have two shifted lowest-weight scalars $\phi$, so for any massive component amplitude the sum will be over massless amplitudes with two shifted lowest-weight scalars $S_{34}$. Schematically, for any component of (2.36) we have

$$A\left[\hat{\phi}_1,\hat{\phi}_2,\dots\right]\sim\lim\sum A\left[\widehat{S}_{34},\varphi_{\rm vev},\dots,\varphi_{\rm vev},\widehat{S}_{34},\varphi_{\rm vev},\dots\right],\qquad(2.39)$$

where the left side is a Coulomb branch amplitude and the right side is a sum over amplitudes at the origin of moduli space with insertions of scalars $\varphi_{\text{vev}} = -\frac{1}{2}(S_{13} - S_{24}) = -\Re(S_{13})$, which are the massless scalar degrees of freedom which gain a vev on the Coulomb branch. These fields are taken soft by the limit. Each massless amplitude on the right side of (2.39) is obviously a component of some massless superamplitude

$$\mathcal{A}\left[\{\hat{\lambda}_1, 0\}, \{\lambda_{\text{vev}}, \eta_{\text{vev}}, \tilde{\eta}^{\dagger}_{\text{vev}}\}, \ldots, \{\lambda_{\text{vev}}, \eta_{\text{vev}}, \tilde{\eta}^{\dagger}_{\text{vev}}\}, \{\hat{\lambda}_2, 0\}, \{\lambda_{\text{vev}}, \eta_{\text{vev}}, \tilde{\eta}^{\dagger}_{\text{vev}}\} \ldots\right] \tag{2.40}$$

in non-chiral superspace, where the Grassmann variables of the two shifted lines have been set to zero.

It was shown in (40) that all massless $\mathcal{N}=4$ superamplitudes scale as $1/z$ in chiral superspace and, as mentioned above, the half-Fourier transform to non-chiral superspace does not modify the scaling. For any such superamplitude we may then perform the massless version of the supertranslation above to rid the superamplitude of the shifted Grassmann variables and bring it to the form of (2.40), where the shifted legs are lowest-weight scalars. This removes any factors of $z$ from the Grassmann monomials, so the superamplitude scaling immediately implies that the individual components of (2.40) must vanish as $1/z$. Then, from (2.39), the massive components, as sums of amplitudes scaling as $1/z$, must also scale as $1/z$. The translated massive superamplitude in (2.36) also has no $z$ dependence in its Grassmann variables, and so we may argue in reverse and upgrade the $1/z$ scaling of the component amplitudes to that of the full superamplitude. Thus the massive super-BCFW shifted superamplitude vanishes at infinity and therefore this is a valid shift.

This proves the validity of massive super-BCFW shifts of Coulomb branch superamplitudes. In concert with the aforementioned validity of super-BCFW when massless legs are shifted, this shows that all Coulomb branch superamplitudes are super-BCFW

constructible.

## 2.5 Scattering Amplitudes on the $\mathcal{N} = 4$ Coulomb Branch

The study of amplitudes at the origin of moduli space has revealed surprising structures and remarkable simplicity. The question of how much of this survives with massive states is not only of intrinsic interest, but also has use in understanding the loop-level properties of the massless theory. A first attempt to construct massive amplitudes and trace the way that the massless amplitudes are deformed by Higgsing was made in (18). They used a superspace representation analogous to that traditionally used at the origin of the moduli space, in which the full $R$-symmetry is manifest (although the little group is not). Of particular note for the discussion here is that they were able to deduce that the superamplitudes could be decomposed into distinct 'band' structures interrelated by SWIs, analogous to the usual sectors classified by degree of helicity violation, as well as find explicit expressions for the simplest cases of these. The use of soft limits discussed above was also proposed, which was then expanded upon in (19) to reconstruct tree-level amplitudes as a series expansion in mass.

After hints arising in loop computations (see e.g. (47; 48)), dual conformal symmetry was discovered in massless gluon amplitudes at strong coupling through holographic computations in (1), where it was shown that this was the conformal symmetry associated with Wilson loops $T$-dual to the amplitude. As mentioned above, this symmetry has also been discovered in the planar amplitudes at weak coupling, thereby suggesting some non-perturbative property of the theory that may be accessible through analytic techniques. At leading order, dual conformal symmetry is enhanced to superconformal and combines

with spacetime superconformality into a Yangian symmetry (see e.g. (49) for review of integrability in $\mathcal{N} = 4$ SYM). The breaking of dual conformality at loop level by IR divergences is understood from the Wilson loop duality (50) and has been used to fully determine the amplitudes with fewer than six legs (51). However, the extent of the usefulness and survival of the enhancements in scattering amplitudes at loop-level is still under investigation (52–56), although some of the progress has made use of this at the level of the loop integrands where infrared divergences can be sidestepped.

Loop-level investigations into dual conformal symmetry led to the suggestion of using Higgsing as a way of regulating IR divergences in loop amplitudes between massless particles (57). This was subsequently used in (58; 59) to constrain the form of loop integrands (an amusing application of this to computing the hydrogen spectrum in $\mathcal{N} = 4$ SYM was shown in (60)). The resulting prediction of this symmetry that 1-loop amplitudes do not involve triangle integrals (like their massless counterparts (61)) was verified in (17), where a massive on-shell superspace was also set-up.

Following (24), it was attempted in (25) to obtain massive amplitudes in $\mathcal{N} = 4$ SYM in $4d$ by dimensional reduction from amplitudes in $6d$, $\mathcal{N} = (1, 1)$ SYM. The $6d$ SYM amplitudes also feature dual conformal symmetry at tree-level (as well as at loop-level integrands), despite not being conformal themselves (22) (this was also observed in $10d$ $\mathcal{N} = 1$ SYM, which also reduces to this $6d$ theory (45)). The realisation of this symmetry and the way that it is inherited by the massive $4d$ amplitudes, its possible relation to the Yangian and its usefulness in providing a guiding structure for determining the superamplitudes were discussed in (22) and (25). The former used reduction of the $6d$ dual conformal symmetry to establish that the massive $\mathcal{N} = 4$ tree amplitudes and loop integrands were dual conformal invariant, while some progress was made in the latter using this, as well as $6d$ super-BCFW recursion (21), to build some superamplitudes at low numbers of legs with manifest dual conformal symmetry (tests at loop-level were made

in (23)). The symmetry algebra has been more recently discussed in (62). However, a general procedure for efficiently computing higher leg amplitudes is still left outstanding. As stated in the introduction, (27) (and, more recently, (28)) introduce CHY formulae for all $6d$ $\mathcal{N} = (1, 1)$ massless amplitudes which may reduced to give a general formula for all $4d$ massive $\mathcal{N} = 4$ tree amplitudes in a CHY form. It remains to be seen exactly how these special symmetries affect the structure of the massive amplitudes and can be used to explicitly construct them. The first step in such an investigation is to calculate and dissect some tree amplitudes in a presentable way. Once the patterns are identified, they can be used to guide the development of systematic computational techniques. We do this in the hope that it will ultimately help in grappling with the way in which the aspects of spin, supersymmetry and dual conformal symmetry interplay.

The first steps toward elucidating the special symmetries of the Coulomb branch amplitudes is to compute the simplest examples and search for the patterns. It is this goal that we initiate in the remainder of this section.

### 2.5.1 Special Massive Kinematics and Three Particle Superamplitudes

#### Special BPS Kinematics

Similarly to their massless counterparts, on-shell three-particle amplitudes of massive BPS vector multiplets exhibit special kinematical properties. Without loss of generality, we will consider the superamplitude $\mathcal{A}_3[\mathcal{W}, \overline{\mathcal{W}}, \mathcal{W}]$ with two BPS and one anti-BPS states. Conservation of the central charge implies that $m_2 = m_1 + m_3$. This configuration of masses yields precisely a massive analogue of the special 3-particle kinematics of massless 3-leg amplitudes, because restricting the momenta to be real implies that they are parallel.

The special kinematic features of the $6d$ three-particle amplitudes have been described

in (44). The $4d$ BPS particles have analogous properties. We will introduce these following the presentation in (44) before giving a more geometric account further below.

It is simple to show that $\left[i^I j^J\right] \pm \left\langle i^I j^J\right\rangle$ has vanishing determinant as a matrix in little group indices for any pair of legs $i$ and $j$, where the $(-)$ is to be chosen if $i$ and $j$ have central charges of opposite sign and $(+)$ is chosen if they are the same. This implies the factorisation

$$\left[i^I j^J\right] \pm \left\langle i^I j^J\right\rangle = u_i^I v_j^J \tag{2.41}$$

for some pure (complexified) $SU(2)$ spinors $u_i$ and $v_i$. It follows from these equations and the spin sums that $u_{i,I}\left[i^I\right| \propto v_{i,I}\left[i^I\right| \propto u_{j,I}\left[j^I\right|$ for each $i$ and $j$ and likewise $u_{i,I}\left\langle i^I\right| \propto v_{i,I}\left\langle i^I\right| \propto u_{j,I}\left\langle j^I\right|$ and that $u_i^I \propto v_i^I$ for each leg - it is only distinct $GL(1)$ rescaling redundancies that distinguishes $u_i$ from $v_i$. These $GL(1)$ rescaling freedoms represent the complexification of the $U(1)$ "tiny groups" of each pair of particles (63). This is the subgroup of Lorentz transformations that stabilises a pair of massive momenta. However, we find that this doubling is practically unnecessary and fix the scales so that $v_i = u_i$ for each $i$. The following identities then hold:

$$\begin{aligned}
\left[1^I 2^J\right] - \left\langle 1^I 2^J\right\rangle &= u_1^I u_2^J \\
\left[2^J 3^K\right] - \left\langle 2^J 3^K\right\rangle &= u_2^J u_3^K \\
\left[3^K 1^I\right] + \left\langle 3^K 1^I\right\rangle &= u_3^K u_1^I
\end{aligned} \tag{2.42}$$

and they imply

$$\begin{aligned}
u_{1,I}\left\langle 1^I\right| = u_{2,J}\left\langle 2^J\right| &= u_{3,K}\left\langle 3^K\right| \equiv \langle u| \\
u_{1,I}\left[1^I\right| = -u_{2,J}\left[2^J\right| &= u_{3,K}\left[3^K\right| \equiv [u|\,.
\end{aligned} \tag{2.43}$$

The spinors in a general little group frame may therefore be decomposed into components in this special frame in which the null vector decomposition "aligns" in this complexified way. The little group spinors may be decomposed into a magnitude and direction as $u_i^I = |u_i|\hat{u}_i^I$, where $\hat{u}_i^I$ is a unit $SU(2)$ spinor. To construct a little group basis including $u_{i,I}$, define

$$\hat{w}_{i,I} = \hat{u}_{i,I}^\dagger + \omega_i \hat{u}_{i,I}, \tag{2.44}$$

as linearly independent spinor, where $\omega_i \in \mathbb{C}$ is free (so $\hat{w}_i$ need not be a unit spinor). A little group spinor basis may be completed with

$$w_{i,I} = \frac{1}{|u_i|}\hat{w}_i. \tag{2.45}$$

This is effectively a dual spinor and satisfies $w_i^I u_{i,I} = 1$. This condition is necessary for the momenta to be on-shell (and the overall sign is fixed by requiring that the momenta be future-pointing in the real limit). The momenta may then be decomposed as

$$p_i = w_{i,I} \left|i^I\right] \langle u| \mp |u] \left\langle i^J\right| w_{i,J} \tag{2.46}$$

(the $(+)$ is for BPS, $(-)$ for anti-BPS). The real momentum limit corresponds to $\omega_i \to 0$, but for complex momenta, $\omega_i$ is an undetermined residual redundancy. In this latter case, the spinors in each term in the decomposition are not complex conjugates and the little group is complexified from $SU(2)$ to $SL(2, \mathbb{C})$.

The resemblance of (2.46) with the 3-particle massless special kinematics is clear. As will be shown further below, in the high energy limit, the null vectors in one of these sets all shrink to zero, recovering the usual special 3-particle kinematics for massless particles.

Calling the spinors $\hat{w}_{i,I}\left|i^I\right\rangle = |i^w\rangle$ (and similarly for the left-handed spinors), then

$$\left[i^I j^J\right] \pm \left\langle i^I j^J\right\rangle = \hat{u}_i^I \hat{u}_j^J \left([i^w j^w] \pm \langle i^w j^w\rangle\right), \tag{2.47}$$

The $\hat{u}_i$ give the direction in which the little group matrix on the left hand side of (2.47) has its only non-zero entry given by the accompanying factor. The combination

$$|u_i||u_j| = [i^w j^w] \pm \langle i^w j^w\rangle \tag{2.48}$$

is here the massive analogue of the spinor bilinears of massless particles like $\langle ij\rangle$ and $[ij]$, only one of which is non-zero, as determined by the configuration of special 3-particle massless kinematics. We will see below that the special massive kinematics will imply that the amplitudes will be functions of this combination of bilinears, along with accompanying little group tensor factors that encode polarisation information. However, first note that these relations may be inverted to give

$$|u_1| = \sqrt{\frac{\left([1^w 2^w] - \langle 1^w 2^w\rangle\right)\left([3^w 1^w] + \langle 3^w 1^w\rangle\right)}{\left([2^w 3^w] - \langle 2^w 3^w\rangle\right)}} \tag{2.49}$$

and similarly for the others. Note that because the $u_i$ spinors carry the scale in (2.49), they contain more information than merely a preferred little group decomposition in which the spinors of each leg align. This discussion is also entirely independent of the choice of $\omega_i$ variables in the definition of $w_i$ frame spinors (2.44).

As is clear in (2.46), for a single leg, the little group basis choice $\{u_i, w_i\}$ still has a remaining $GL(1,\mathbb{C})$ little group freedom under which $u_i$ and $w_i$ rescale oppositely. The choice of scales given by the equations (2.43) reduces this to a single $GL(1,\mathbb{C})$ for all three legs. In the special case of real momenta, (2.43) also leaves the overall direction free. However, this is fixed by the complex deformation.

Geometrically, a massive momentum vector may be decomposed into a sum of two future-directed, light-like vectors. The $SU(2)$ little group invariance represents the manifold of all such decompositions. Each null vector rotates about the massive vector under the Wigner rotations, but their sum remains unchanged. When the three leg momenta are real and parallel, each can be decomposed into a linear combination of the same two real null vectors. However, when complexified, the massive leg momenta need no longer be proportional and the requirement that their null vector decompositions coincide becomes a more stringent constraint. Here, "coincide" means that they have vanishing dot products (as is clear for e.g. the first terms in (2.46) for each $i$). However, in this complexified context, null vectors with vanishing dot product need no longer be linearly dependent (proportional). This is the complexification of the notion of massive momenta being parallel. The spinors constituting each null vector are also no longer complex conjugates. As a result, the little group is also complexified from $SU(2)$ to $SL(2, \mathbb{C})$.

The existence of the preferred spinor direction is analogous to that provided by the massless particle in the general 3-leg amplitude with one massless leg and two massive legs of equal mass, as classified in (30). It is therefore analogously possible to construct general 3-particle amplitudes between massive particles obeying this mass constraint by expanding the polarisation-stripped amplitude tensor in a basis spanned by tensor products of $|u]_\alpha$ and $\epsilon_{\alpha\beta}$. However, by Lorentz invariance, the amplitude tensor must be of even total rank, so factors of $|u]_\alpha$ may always be paired and eliminated using (2.42) and (2.43). Doing so will always leave (after applying the spin sums) terms proportional to $\epsilon_{\alpha\beta}$ and $(p_1 p_2)_{\alpha\beta}$, which are precisely the building-blocks used for the general 3-leg amplitude of three massive particles proposed in (30). Thus these special kinematics provide no new constraints on possible Lorentz structures in amplitudes nor any new features beyond the general case.

However, the special case in which one leg is massless and the other two have equal

mass can be regarded as a limiting case. Taking $m_3 \to 0$ and $m_2 \to m_1 = m$, then in the helicity basis for the little group frame,

$$u_{3+} \to \pm\sqrt{mx} \qquad u_{3-} \to \pm\sqrt{m/x} \tag{2.50}$$

(the sign choice in each limit is to be the same). These components of the frame spinor produce the helicity-weight-carrying scalar units $x = u_{3+}/u_{3-}$, introduced in (30). The remaining $u_1$ and $u_2$ spinors can still be used as building-blocks, but can be related to $x$ and $|3\rangle$ and $|3]$ through (2.42). Explicitly,

$$u_1^I = \mp\sqrt{\frac{x}{m}}\left[1^I 3\right] \qquad u_2^J = \pm\sqrt{\frac{x}{m}}\left[2^J 3\right]. \tag{2.51}$$

Nevertheless, the case of a massless leg with two massive legs of equal mass is distinguished from other cases obeying the mass selection rule in that it does not have a non-trivial, real momentum, collinear limit. As a result, $x$ is a purely complex momentum object with no analogue in amplitudes of other mass configurations. The significance of this was observed recently by (33), where it was shown that amplitudes of magnetic monopoles factorise differently on each of the two possible complex momentum configurations corresponding to the same factorisation channel, which was interpreted as signifying the presence of a Dirac string. In contrast, the Bhabha scattering calculation presented in Section 2.4.2 illustrates the simplest way in which the $x$ factors across a factorisation channel can be combined that does not depend upon the complex momentum configuration chosen, see discussion in (33).

### Three-Particle Superamplitude

To begin with, we present the 3-particle superamplitude for massless legs in non-chiral superspace, which is

$$\begin{aligned}\mathcal{A}_3[G_1,G_2,G_3] =& \frac{1}{\langle 12\rangle\langle 23\rangle\langle 31\rangle}\delta^{(4)}(Q^\dagger)\prod_a(\langle 12\rangle\,\tilde{\eta}_3^{\dagger a}+\langle 23\rangle\,\tilde{\eta}_1^{\dagger a}+\langle 31\rangle\,\tilde{\eta}_2^{\dagger a})\\ &+\frac{1}{[12][23][31]}\delta^{(4)}(Q)\prod_a([12]\,\eta_3^a+[23]\,\eta_1^a+[31]\,\eta_2^a).\end{aligned}\tag{2.52}$$

The first term is the MHV sector and the second is the anti-MHV ($\overline{\text{MHV}}$) sector. Each term is only non-zero for distinct special massless kinematical configurations. This may be obtained from the well-known chiral form by the half-Fourier transform. We henceforth choose to absorb the annoying factor of $\sqrt{2}$ in the supercharges in (2.15) into the definition of the coupling so that it is implicitly to be omitted in all appearances of the delta functions $\delta^{(4)}(Q)$ and $\delta^{(4)}(Q^\dagger)$.

We next turn to deriving the superamplitude for massive legs. Usually, supersymmetry invariance of an amplitude immediately implies that $\mathcal{A}_n \propto \delta^{(4)}(Q^{\dagger a})\delta^{(4)}(Q_{a+2})$. However, as will be shown below, the special kinematics here implies that 2 pairs of supercharges of each chirality degenerate, leaving only 6 independent (if the momenta were restricted to be real, then all 4 pairs would be related). This occurs as a result of the special spinor direction given by the $u_i^I$. It is simple to show that $\langle uQ^{\dagger a}\rangle = -[uQ_{a+2}] = \sum_i m_i u_i^I \eta_{iI}^a$. In this case, the superamplitude may be deduced from the little group scaling of the external legs (which are invariant in this coherent state basis) and invariance under the independent supersymmetries. Building a supersymmetry invariant involves introducing a new reference spinor $|q\rangle \not\propto |u\rangle$, effectively to decompose the supercharges into the shared components that are parallel to $|u\rangle$ and the remaining independent components. Projected onto $|u\rangle$ and $|q\rangle$, the delta functions may be factorised as $\delta^{(4)}(Q^{\dagger a}) =$

$\frac{1}{\langle qu\rangle^2}\delta^{(2)}(\langle qQ^{\dagger a}\rangle)\delta^{(2)}(\langle uQ^{\dagger a}\rangle)$ and $\delta^{(4)}(Q_{a+2}) = \frac{1}{m_1^4\langle qu\rangle^2}\delta^{(2)}(\langle q|\, p_1\, |Q_{a+2}])\delta^{(2)}(\langle u|\, p_1\, |Q_{a+2}])$, where both expressions are independent of $|q\rangle$. Up to a multiplicative prefactor, the supersymmetry invariant may be obtained by dropping the repeated factor in both $\delta^{(4)}(Q^{\dagger a})$ and $\delta^{(4)}(Q_{a+2})$. This is easily verified as being annihilated by all of the supercharges. To determine the numerical prefactor, we demand that the result match onto (2.52) in the limit of massless legs. The superamplitude is thus determined to be

$$\begin{aligned}\mathcal{A}_3[\mathcal{W}_1,\overline{\mathcal{W}}_2,\mathcal{W}_3] &= \frac{1}{m_1^2\,\langle q|\, p_1p_3\, |q\rangle}\delta^{(4)}\left(Q^{\dagger a}\right)\delta^{(2)}\left(\langle q|\, p_1\, |Q_{a+2}]\right)\\ &= \frac{1}{\langle q|\, p_1p_3\, |q\rangle}\delta^{(4)}\left(Q_{a+2}\right)\delta^{(2)}\left(\langle qQ^{\dagger a}\rangle\right). \end{aligned} \tag{2.53}$$

The superamplitude has been expressed in a form in which the auxiliary spinors $u_i$ do not appear explicitly, although they still constrain the reference spinor $|q\rangle$ to satisfy $\langle uq\rangle \neq 0$. While $\delta^{(4)}(Q^{\dagger a}) = \frac{1}{\langle qu\rangle^2}\delta^{(2)}(\langle uQ^{\dagger a}\rangle)\delta^{(2)}(\langle qQ^{\dagger a}\rangle)$ is clearly independent of the reference spinor, this remains true of $\frac{1}{\langle qu\rangle^2}\delta^{(2)}(\langle q|\, p_1\, |Q_{a+2}])$ up to terms that vanish when multiplied by the other delta functions. It is therefore justified to the drop of the factor of $\delta^{(2)}(\langle u|\, p_1\, |Q_{a+2}])$ in $\delta^{(4)}(Q)$ to obtain the SUSY invariant in the first form in (2.53) (and a similar argument applies to the second). The reference spinor itself is unnecessary for the component amplitudes and may be eliminated after these are extracted. However, it is needed to squash them all into the superamplitude in this way. A similar representation of the massless three particle superamplitude in $6d$ was found in (63), which presumably reduces to the expression above upon dimensional reduction.

Also of note is that this superamplitude combines terms that belong to distinct supersymmetric sectors (MHV and $\overline{\text{MHV}}$ in the massless limit) into a single Grassmann polynomial. We will return to this point and see the combination of sectors even more explicitly in Section 2.5.3.

The most remarkable feature of the massive 3-leg superamplitude (2.53) is the kine-

matic factor in the denominator. This vanishes in the collinear limit - the one situation in which the momenta can be both real and on-shell. This factor is reminiscent of the Parke-Taylor factors of the exact, massless Yang-Mills 3-leg amplitude e.g. $\sim \frac{\langle 12\rangle^4}{\langle 12\rangle\langle 23\rangle\langle 31\rangle}$, as well as its supersymmetrised counterpart $\frac{\delta(Q)}{\langle 12\rangle\langle 23\rangle\langle 31\rangle}$. For these theories, when glued into a 4-leg amplitude on a factorisation channel, the factors in the denominator combine to produce the pole representing the other factorisation channel of the amplitude, as arranged for automatically by BCFW recursion (64), (30). However, in the massive case here, the kinematic factor is neither present nor necessary in any of the component amplitudes. Instead, its appearance is orchestrated as a consequence of the maximal supersymmetry. Its presence likewise suggests that the Coulomb branch superamplitudes share in the special constructibility properties of their massless counterparts, as confirmed by the existence of super-BCFW. This will be explored further below.

An alternative representation of the three particle superamplitude also exists that more directly utilises the special kinematical properties of the BPS states. In the special frame selected by $\{u_i, w_i\}$, the multiplicative supercharges may be decomposed as

$$Q = |u] \sum_i \eta_{iw} - \sum_i \pm \frac{1}{|u_i|} |i^w] \, \eta_{iu} \qquad Q^\dagger = -|u\rangle \sum_i \eta_{iw} + \sum_i \frac{1}{|u_i|} |i^w\rangle \, \eta_{iu}, \tag{2.54}$$

calling Grassmann variables $\eta_{iu} = u_i^I \eta_{i,I}$ and $\eta_{iw} = w_i^I \eta_{i,I}$ (not $\hat{w}_i$ as used in the definition of $|i^w\rangle$). Now, partially solving the supercharge conservation constraints $[uQ] = 0$ and $\left\langle uQ^\dagger \right\rangle = 0$ implies that $\eta_{iu} = \pm \eta_{ju}$ for all legs $i$ and $j$ (where the $(+)$ applies if the central charges of $i$ and $j$ are the same and $(-)$ if they are opposite). This consequently implies that, on the support of this solution, the supercharges are parallel to the special frame spinor directions e.g. $Q \sim |u] \left(\sum_i \eta_{iw} - C\eta_{1u}\right)$. The constant $C$ may be determined by introducing the reference spinor $|q]$ satisfying $[qu] \neq 0$ (any of the $|i^w]$ would be possible

choices):

$$C = \frac{1}{[qu]} \sum_i \frac{1}{(u_i)} [qi^w] \,. \tag{2.55}$$

An alternative representation of the supersymmetric delta function may therefore be deduced by combining each of the three distinct Grassmann terms in (2.54) above into a single product

$$\mathcal{A}_3[\mathcal{W}_1, \overline{\mathcal{W}}_2, \mathcal{W}_3] = \prod_a \left( \left( \sum_i \eta^a_{iw} \right) (\eta^a_{1u}\eta^a_{2u} + \eta^a_{2u}\eta^a_{3u} - \eta^a_{3u}\eta^a_{1u}) - C\eta^a_{1u}\eta^a_{2u}\eta^a_{3u} \right) . \tag{2.56}$$

Note that, thus far, every expression involving a decomposition into this special little group frame is independent of the choice of $\omega_i$ in (2.45). These parameters remain free. Further simplification may be achieved by partially fixing the $\omega_i$ parameters to set $C = 0$, or equivalently

$$\sum_i \frac{1}{|u_i|} |i^w] = 0 \qquad \sum_i \pm \frac{1}{|u_i|} |i^w\rangle = 0 \tag{2.57}$$

(these two equations are equivalent). On the support of each other's delta functions, the supercharges then reduce to the first terms in (2.54). The superamplitude simplifies to

$$\mathcal{A}_3[\mathcal{W}_1, \overline{\mathcal{W}}_2, \mathcal{W}_3] = \prod_a \left( \sum_i \eta^a_{iw} \right) (\eta^a_{1u}\eta^a_{2u} + \eta^a_{2u}\eta^a_{3u} - \eta^a_{3u}\eta^a_{1u}) \,. \tag{2.58}$$

This is analogous to the form commonly presented in $6d$ (22).

The massive amplitudes are built out of these combinations of bilinears in (2.48). In (2.58), these are split apart into their 'square roots' $|u_i|$. In extracting a component amplitude, four factors of $u_i$ and two of their duals $w_i = \frac{1}{|u_i|}\hat{w}_i$ are produced. These combine into spinor bilinears through (2.48). This demonstrates how the frame spinors

$u_i^I$ can be used as alternative building blocks with which to construct the three particle amplitudes.

To illustrate this more explicitly, the three massive vector component amplitude may be extracted from (2.58) to give

$$A[W_1^{I_1 I_2}, \overline{W}_2^{J_1 J_2}, W_3^{K_1 K_2}] = \prod_{i=1}^{2} \left( \frac{|u_2||u_3|}{|u_1|} \hat{w}_1^{I_i} \hat{u}_2^{J_i} \hat{u}_3^{K_i} - \frac{|u_3||u_1|}{|u_2|} \hat{w}_2^{J_i} \hat{u}_3^{K_i} \hat{u}_1^{I_i} + \frac{|u_1||u_2|}{|u_3|} \hat{w}_3^{K_i} \hat{u}_1^{I_i} \hat{u}_2^{J_i} \right). \tag{2.59}$$

The little group indices are implicitly to be symmetrised over (we will assume this in all subsequent expressions where they arise as indexing polarisation states of external legs). The diagonal terms in the product have the form e.g.

$$\prod_i \left( \frac{|u_2||u_3|}{|u_1|} \hat{w}_1^{I_i} \hat{u}_2^{J_i} \hat{u}_3^{K_i} \right) = \frac{\left([2^w 3^w] - \langle 2^w 3^w \rangle\right)^3}{\left([1^w 2^w] - \langle 1^w 2^w \rangle\right) \left([3^w 1^w] + \langle 3^w 1^w \rangle\right)} \prod_i \hat{w}_1^{I_i} \hat{u}_2^{J_i} \hat{u}_3^{K_i} \tag{2.60}$$

It is clear that the prefactor multiplying the spinors is the massive upgrade of the Parke-Taylor factor. The remaining factor accounts for the spin components with respect to a given quantisation axis. Likewise, the cross terms are of the form

$$\left( \frac{|u_2||u_3|}{|u_1|} \frac{|u_3||u_1|}{|u_2|} \hat{w}_1^{I_1} \hat{u}_1^{I_2} \hat{u}_2^{J_1} \hat{w}_2^{J_2} \hat{u}_3^{K_1} \hat{u}_3^{K_2} \right) = \frac{\left([2^w 3^w] - \langle 2^w 3^w \rangle\right) \left([3^w 1^w] + \langle 3^w 1^w \rangle\right)}{[1^w 2^w] - \langle 1^w 2^w \rangle} \hat{w}_1^{I_1} \hat{u}_1^{I_2} \hat{u}_2^{J_1} \hat{w}_2^{J_2} \hat{u}_3^{K_1} \hat{u}_3^{K_2}. \tag{2.61}$$

The prefactor here suggestively resembles the massless amplitude for photon/gluon emission by a scalar.

In all expressions prior to (2.58), all occurrences of the $\omega_i$ parameters cancelled-out and could be set to zero without loss of generality (effectively setting $\hat{w}_i = \hat{u}_i^\dagger$).

While yielding the pleasing expressions above, the cost of the frame choice that sets $C=0$ is that further complication in the general expression has been transferred into the $\hat{w}_i$, which cannot be identified as unit spinors determined by the $\hat{u}_i^\dagger$ alone. Alternative expressions with this interpretation could be extracted directly from (2.56) at the expense of additional manifest complication.

In the limit that all legs become massless, (2.48) implies that the frame spinors all converge to a particular helicity, which corresponds to the configuration of massless special 3-particle kinematics. Either $\hat{u}_{i+} \to 0$ for each $i$ and the right-handed massless spinors align or $\hat{u}_{i-} \to 0$ and the left-handed spinors align. The combinations of bilinears (2.48) appearing in the superamplitude behave as $([i^w j^w] \pm \langle i^w j^w \rangle) \to \hat{u}_{i+}^\dagger \hat{u}_{j+}^\dagger [ij]$ or $([i^w j^w] \pm \langle i^w j^w \rangle) \to \pm \hat{u}_{i-}^\dagger \hat{u}_{j-}^\dagger \langle ij \rangle$ for each $i, j$. The surviving factors of $u_{iI}$ then become "square-roots" of the massless bilinears e.g. $u_{3+} \to \sqrt{\frac{[23][31]}{[12]}}$ or $u_{3+} \to \sqrt{\frac{\langle 23\rangle\langle 31\rangle}{\langle 12\rangle}}$.

Furthermore, in either massless complex kinematical configuration, $C \to 0$ and only the terms retained in the $C=0$ frame remain in the massless limit. The factors of $\hat{w}_i$ may be identified with as $\hat{u}_i^\dagger$ and consequently the massless limit may be read-off from the expressions (2.58) and (2.59). For example, in the case where all left-handed spinors become proportional, the second factor in (2.58) converges to $\delta^{(4)}(Q)$, while the first becomes the remaining Grassmann quadratic (including the Parke-Taylor factor) in the $\overline{\text{MHV}}$ term in (2.52). The diagonal term in the all vector component amplitude above (2.60) clearly converges to the Parke-Taylor three vector amplitude $A[g_1^\mp, g_2^\pm, g_3^\pm]$ for the relevant helicity and massless special kinematics choices, otherwise it converges to zero. Likewise, the cross-terms like (2.61) converge to amplitudes expected for a Goldstone boson emitting a gluon, as expected from the Higgs mechanism. The remaining factors of the unit frame spinors $\hat{u}_{i\pm}^{(\dagger)}$ ultimately cancel-out.

In practice, although carrying the redundant reference spinor, the form (2.53) is relatively easy to use in practical calculations. We will choose to continue to use the

spacetime spinor formulation of (2.53) in the remainder of this paper. Following a similar argument to that presented above for (2.53), a simple representation for the 2-massive-leg superamplitude may instead be derived by finding the SUSY invariant $\delta^{(4)}(Q^{\dagger a})\delta^{(4)}(Q_{a+2})$ and dropping one of the repeated factors of the degenerate supercharges. The overall coefficient is then fixed by the little group scaling of the legs. In this case, the special kinematics implies that $\langle 3|\, p_1\, |Q_{a+2}] = -m\left\langle 3Q^{\dagger a}\right\rangle$. The 3-leg superamplitude is then determined to be

$$\begin{aligned}\mathcal{A}_3[\mathcal{W}_1,\overline{\mathcal{W}}_2,G_3] &= \frac{-x}{m^3\left\langle q3\right\rangle^2}\delta^{(4)}\left(Q^{\dagger a}\right)\delta^{(2)}\left(\langle q|\, p_1\, |Q_{a+2}]\right)\\ &= \frac{-x}{m^3\left\langle q3\right\rangle^4}\delta^{(2)}\left(\left\langle 3Q^{\dagger a}\right\rangle\right)\delta^{(2)}\left(\left\langle qQ^{\dagger a}\right\rangle\right)\delta^{(2)}\left(\langle q|\, p_1\, |Q_{a+2}]\right)\\ &= \frac{-x}{m\left\langle q3\right\rangle^2}\delta^{(4)}\left(Q_{a+2}\right)\delta^{(2)}\left(\left\langle qQ^{\dagger a}\right\rangle\right).\end{aligned} \tag{2.62}$$

The reference spinor $|q\rangle$ must satisfy $\langle q3\rangle \neq 0$. Because this superamplitude must be invariant under the little group scaling of its legs, the helicity-carrying factor $x$ has been re-introduced. The presence of $x$ is expected because of its appearance in the component amplitudes like $A_3[W^{I_1I_2},\overline{W}^{J_1J_2},g^+]$ and it emerges in taking the massless limit $m_3\to 0$ of $u_{3,K}\left|3^K\right]/(\left\langle q3^K\right\rangle u_{3,K})$ in (2.53), as explained previously above.

Explicitly expanding the delta functions gives

$$\begin{aligned}\mathcal{A}_3[\mathcal{W}_1,\overline{\mathcal{W}}_2,G_3] = \frac{-1}{m}x\prod_a\Big(&-\left[32^I\right]\eta^a_{1M}\eta_1^{Ma}\eta^a_{2I}+\left[31^I\right]\eta^a_{1I}\eta^a_{2M}\eta_2^{aM}+\left[1^I2^J\right]\eta^a_{1I}\eta^a_{2J}\eta^a_3\\ &-x^{-1}\left\langle 1^I2^J\right\rangle\eta^a_{1I}\eta^a_{2J}\tilde\eta_3^{\dagger a}+\frac{1}{2}m\eta^a_{1M}\eta_1^{Ma}\eta^a_3+\frac{1}{2}m\eta^a_{2M}\eta_2^{Ma}\eta^a_3+\frac{1}{2}\frac{m}{x}\eta^a_{1M}\eta_1^{Ma}\tilde\eta_3^{\dagger a}\\ &+\frac{1}{2}\frac{m}{x}\eta^a_{2M}\eta_2^{Ma}\tilde\eta_3^{\dagger a}+\left[1^I3\right]\eta^a_{1I}\eta^a_3\tilde\eta_3^{\dagger a}-\left[2^I3\right]\eta^a_{2I}\eta^a_3\tilde\eta_3^{\dagger a}\Big),\end{aligned} \tag{2.63}$$

which allows the components to be efficiently read off. Notably, the reference spinor

introduced in the delta functions has completely disappeared and does not affect the components.

The form of the two-equal-mass superamplitude makes clear that the interactions of BPS states with massless gauge bosons are monomials in $x$. In the above case, this has the physical interpretation of the BPS states having gyromagnetic ratio $g = 2$ exactly (and likewise no anomalous electric quadrupole moment, as seen in the $\mathcal{N} = 1$ case in (29)). The different Lorentz structures of the couplings are fully protected by supersymmetry. Explicitly, we may extract the collection of such component amplitudes as

$$\mathcal{A}[\mathcal{W}, \overline{\mathcal{W}}, g^-] = \frac{x}{m} \prod_a \left( [1^I 2^J] \eta^a_{1I} \eta^a_{2J} + \frac{1}{2} m \eta^a_{1M} \eta^{Ma}_1 + \frac{1}{2} m \eta^a_{2M} \eta^{Ma}_2 \right) \tag{2.64}$$

$$\mathcal{A}[\mathcal{W}, \overline{\mathcal{W}}, g^+] = \frac{1}{mx} \prod_a \left( \langle 1^I 2^J \rangle \eta^a_{1I} \eta^a_{2J} - \frac{1}{2} m \eta^a_{1M} \eta^{Ma}_1 - \frac{1}{2} m \eta^a_{2M} \eta^{Ma}_2 \right). \tag{2.65}$$

### 2.5.2 Four Particle Superamplitudes

Using massive super-BCFW, we next present a derivation of the general 4-leg superamplitude for legs of arbitrary mass. In this case, the Grassmann dependence is entirely determined by the factor $\delta^{(4)}(Q^{\dagger,a})\delta^{(4)}(Q_{a+2})$. Thus, only the coefficient of the delta function need be calculated and this is fixed by any single component amplitude. While the expected form of the superamplitude is obvious and follows from supersymmetry, factorisation and the (trivial) spin of the external superfields, the following derivation will illustrate how these emerge from combining the on-shell 3-leg amplitudes (2.53). It will also provide a simple demonstration of the mechanics and use of massive super-BCFW.

We will calculate $\mathcal{A}_4[\mathcal{W}_1, \overline{\mathcal{W}}_2, \mathcal{W}_3, \overline{\mathcal{W}}_4]$. This colour-ordering implies that the masses obey the constraint $m_1 + m_3 = m_2 + m_4$. The cases with different combinations of particles and anti-particles may be obtained by obvious modification. For any such superamplitude that respects colour neutrality of the broken gauge group, there will

always be two consistent factorisation channels in which the on-shell, internal particle has mass given by the sum of the masses of the other legs on each subamplitude (weighted by the sign of their central charges).

As noted above, $4d$ massive super-BCFW may be obtained by dimensionally reducing that of massless $6d$ $\mathcal{N} = (1,1)$ SYM. An analogous calculation of the 4-leg superamplitude in $6d$ was performed in (21), supersymmetrising the computation in pure YM in (44). In $4d$, the special case of two massless legs have been previously calculated by (18) and (27), for the simple case of an $U(N+M) \to U(N) \times U(M)$ breaking pattern (where there are two possible structures with consistent colour-ordering). The former used non-supersymmetric BCFW recursion applied to the component amplitude $A_4[W, \overline{W}, g^+, g^+]$ to determine the kinematical coefficient of the delta functions, while (27) used the general CHY-like formula. These are a special case of our result.

First define generalized Mandelstam variables $s_{ij} = -(p_i + p_j)^2 - (m_i \pm m_j)^2$, where the masses are added if the lines have the same sign central charge and subtracted if opposite. For a general amplitude with any number of legs, these satisfy the useful identities $\sum_j s_{ij} = 0$ and $\sum_{j \neq k} s_{ij} = \sum_{j \neq i} s_{kj}$, by conservation of momentum and the mass constraint. Other relations may be similarly derived.

Applying the super-shift to legs 1 and 2, the superamplitude is determined from a single factorisation channel:

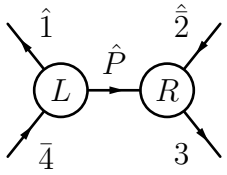



Figure 2.1: The single BCFW diagram for four-point recursion.

$$\mathcal{A}_4[\mathcal{W}_1, \overline{\mathcal{W}}_2, \mathcal{W}_3, \overline{\mathcal{W}}_4] = \int d^4\eta_{\hat{P}} \hat{\mathcal{A}}_L[\overline{\mathcal{W}}_4, \widehat{\mathcal{W}}_1, \mathcal{W}_{\hat{P}}] \frac{-1}{s_{14}} \hat{\mathcal{A}}_R[\overline{\mathcal{W}}_{-\hat{P}}, \widehat{\overline{\mathcal{W}}}_2, \mathcal{W}_3], \tag{2.66}$$

where $\hat{P}$ is the momentum of the internal line, taken as outgoing from the left subamplitude in Figure 2.1 and incoming into the right subamplitude. Hats denote shifted legs, to be evaluated on the residue determined by $s_{4\hat{1}} = 0$, although it will be unnecessary in this example to determine either the residue or the shift vector. Assuming that $m_1 < m_4$, then the internal on-shell particle has mass $m_P = m_4 - m_1$ and is BPS in the left superamplitude. In the right amplitude, it is an incoming BPS state, which can be regarded by crossing symmetry as an outgoing anti-BPS state with momentum $-P$.

Analytically continuing spinors and Grassmann variables from negative to positive energies requires the rules

$$\left|-P^I\right] = i\left|P^I\right] \qquad \left|-P^I\right\rangle = i\left|P^I\right\rangle \qquad \eta^I_{-P} = i\eta^I_P, \tag{2.67}$$

which are little group covariant (and consistent with (65)). These rules imply that the sign of the mass as it appears in the Weyl equation or the spin sums effectively reverses so that e.g. $p\left|-p^I\right] = -m\left|-p^I\right\rangle$ for a leg of mass $m$ and momentum $p$. See Appendix A of (29) for spinor conventions and identities. As a result, while the BPS condition for an analytically continued leg (2.13) is unchanged (noting that both the momentum and central charges reverse under crossing), a relative negative sign appears in the spinor-stripped counterparts (2.5). As a result, under the conventions employed here, the left-handed multiplicative supercharges of the crossed legs pick up an extra negative sign relative to that of the other outgoing legs.

Likewise, the corresponding massless variables in the conventions employed here must

also all acquire a factor of $i$ upon analytic continuation

$$|-p] = i\,|p] \qquad |-p\rangle = i\,|p\rangle \qquad \eta_{-p} = i\eta_p \qquad \eta^\dagger_{-p} = i\eta^\dagger_p. \tag{2.68}$$

The remaining calculation involves combining the delta functions and simplifying. It is through combining the delta functions that the extra pole is generated, effectively as a Jacobian factor arising from the mismatch between the aligned frame spinors $u^I_{\hat{P}}$ on the left and right on-shell amplitudes. This overlap was also the source of the additional pole in $6d$ YM (21). We give details of this in Appendix 2.A, but the result is that

$$\begin{aligned}\hat{\mathcal{A}}_L[\overline{\mathcal{W}}_4, \widehat{\mathcal{W}}_1, \mathcal{W}_{\hat{P}}]\hat{\mathcal{A}}_R[\overline{\mathcal{W}}_{-\hat{P}}, \widehat{\overline{\mathcal{W}}}_2, \mathcal{W}_3] = \frac{1}{\left(u^{(L)}_{\hat{P}M}u^{(R)M}_{\hat{P}}\right)^2}\frac{1}{\left(\langle q|\, p_4\hat{p}_1\,|q\rangle\right)^2} \\ \times\,\delta^{(4)}\left(Q\right)\delta^{(4)}\left(Q^\dagger\right)\delta^{(2)}\left(\left\langle q\hat{Q}^\dagger_R\right\rangle\right)\delta^{(2)}\left(\langle q|\,\hat{p}_1\left|\hat{Q}_R\right]\right).\end{aligned} \tag{2.69}$$

Here the $L$ and $R$ subscripts index parameters originating in the factorised on-shell amplitudes, the hats indicate that they are shifted and the $U(2)$ $R$-indices have been omitted for brevity. The supercharges without subscripts represent those for the full 4-leg superamplitude. The multiplicative factor arises from a succession of basis changes and invocations of constraints from the other delta functions. This critically provides the factor that will become the pole representing the other factorisation channel.

After factoring out the total supersymmetric delta function, the remaining Grassmann integral is simple to perform, giving

$$\int d^4\eta_{\hat{P}}\delta^{(2)}\left(\left\langle q\hat{Q}^\dagger_R\right\rangle\right)\delta^{(2)}\left(\langle q|\,\hat{p}_1\left|\hat{Q}_R\right]\right) = \left(\langle q|\,\hat{p}_1\hat{P}\,|q\rangle\right)^2. \tag{2.70}$$

Combining all of the factors and using $\hat{P} = -\hat{p}_1 - p_4$, the superamplitude reduces to

$$\mathcal{A}_4[\mathcal{W}_1, \overline{\mathcal{W}}_2, \mathcal{W}_3, \overline{\mathcal{W}}_4] = \frac{\delta^{(4)}(Q)\delta^{(4)}(Q^\dagger)}{s_{41}} \frac{-1}{\left(u^{(L)}_{\hat{P}M} u^{(R)M}_{\hat{P}}\right)^2}. \tag{2.71}$$

All of the kinematical factors cancel out with the exception of the internal propagator for this factorisation channel and another factor given by the overlap of the frame spinors for the internal line. This form was also reached in the analogous $6d$ calculation (21; 44) and the demonstration that this is the pole of the other factorisation channel is similar. We repeat the argument from the $4d$ perspective in Appendix 2.A, the result of which is that

$$\left(u^{(L)}_{\hat{P}M} u^{(R)M}_{\hat{P}}\right)^2 = -s_{12}. \tag{2.72}$$

As explained in Subsection 2.5.1, the $u_i$ spinors select a preferred decomposition of massive momenta into a sum of two parallel null vectors. This new pole occurs when the frame spinors for the internal line in the BCFW diagram align. This equivalently means that the two sets of parallel null vectors that span the massive momenta on each side of the factorisation channel align. This is just the complexification of the alignment of the external massive momenta on opposite sides of the the factorisation in the BCFW diagram, which is exactly the condition required for the alternative factorisation channel.

The 4-leg superamplitude is therefore

$$\mathcal{A}_4[\mathcal{W}_1, \overline{\mathcal{W}}_2, \mathcal{W}_3, \overline{\mathcal{W}}_4] = \frac{\delta^{(4)}(Q^{\dagger a})\delta^{(4)}(Q_{a+2})}{s_{12}s_{41}}. \tag{2.73}$$

The residue and the momentum shift ultimately cancelled out in this calculation and did not have to be solved for. This amplitude closely resembles its counterparts in unbroken

Yang-Mills, where both manifestly feature poles in both $s$ and $t$ factorisation channels in a single term. The superamplitude in which any of the legs is massless may be obtained as an obvious limiting case.

Just as in massless (super-)Yang-Mills, only one factorisation channel (or BCFW diagram) was sufficient to determine the 4-leg amplitude from the elementary 3-leg amplitudes. The pole providing the other factorisation channel originates from the kinematic "singularity" in the 3-leg superamplitude (2.53). The numerator (specified here by supersymmetry) determines the polarisation structure. It is a simple task to extract component amplitudes of massive states. As an example, the four massive vector boson amplitude may be found as

$$\begin{aligned} A_4[W_1^{I_1I_2},\overline{W}_2^{J_1J_2},W_3^{K_1K_2},\overline{W}_4^{L_1L_2}] = \frac{1}{s_{12}s_{41}} \\ \prod_{i=1}^{2}\Big( \left[1^{I_i}2^{J_i}\right]\left\langle 3^{K_i}4^{L_i}\right\rangle + \left\langle 1^{I_i}2^{J_i}\right\rangle\left[3^{K_i}4^{L_i}\right] + \left[1^{I_i}3^{K_i}\right]\left\langle 2^{J_i}4^{L_i}\right\rangle \\ + \left\langle 1^{I_i}3^{K_i}\right\rangle\left[2^{J_i}4^{L_i}\right] + \left[1^{I_i}4^{L_i}\right]\left\langle 2^{J_i}3^{K_i}\right\rangle + \left\langle 1^{I_i}4^{L_i}\right\rangle\left[2^{J_i}3^{K_i}\right]\Big) \end{aligned} \quad (2.74)$$

The massive little group indices are implicitly symmetrised over in the above expressions as usual. From the perspective of $6d$ Yang-Mills amplitudes dimensionally reduced to $4d$, each factor in the numerator is the reduction of the '4-bracket' of $6d$ spinors (44). It is clear that the expected helicity selection rules emerge in the massless limit (see Appendix A of (29)), where the amplitudes without split helicities are mass suppressed, most severely when all helicities are the same.

### 2.5.3 Five Particle Superamplitudes and Band Structure

#### Bands

Away from the origin of moduli space, the $R$-symmetry is broken to $USp(4)$ from $SU(4)$ and the sectors of distinct levels of helicity violation partially merge. This occurs because processes forbidden by helicity selection rules may now proceed at mass-suppressed rates. Instead, (18), who work in a chiral superspace in which a $SU(2) \times SU(2) \leq USp(4)$ is manifest, are able to classify the residual supersymmetric invariant sectors by their Grassmann orders under each $SU(2)$ factor. Each of these sectors, in their formulation, is an inhomogeneous polynomial that spans several overlapping even Grassmann orders, which were described as 'bands'. The polynomial of $(K+1)$th lowest degree was called the $\mathrm{N}^K\mathrm{MHV}$ sector, in analogy with the massless superamplitudes. Each invariant term in the superamplitude is then classified under this product structure as a $\mathrm{N}^{k_1}\mathrm{MHV} \times \mathrm{N}^{k_2}\mathrm{MHV}$ band.

In non-chiral superspace the superamplitudes are instead homogeneous of degree $2n$ in Grassmann variables and a distinct $U(2)$ $R$-subgroup is realized explicitly, which foretell a different organization of the bands here (we retain the term 'band' for supersymmetrically closed sector, as well as the $\mathrm{N}^{k_1}\mathrm{MHV} \times \mathrm{N}^{k_2}\mathrm{MHV}$ notation). The simplest non-trivial example of a superamplitude with independent (albeit simple) bands is at five legs and our exploration here will provide insight into the band structure for general Coulomb branch superamplitudes. The three and four leg superamplitudes discussed above are special cases.

As discussed in (18), the three leg massive superamplitude is non-trivial and actually contains three such independent terms. In the little group violating chiral superspace used by the authors, these appear as MHV and $\overline{\mathrm{MHV}}$ superamplitudes with a form almost identical to their massless counterparts, as well as a new $\mathrm{MHV} \times \overline{\mathrm{MHV}}$ term that vanishes

in the massless limit. None of these are manifestly visible in our expression (2.53), because they are represented by sectors of specific helicities, all of which are combined here into a massive little group invariant. That the three-leg superamplitude combines each helicity-violating band into a single, little group and supersymmetric invariant means that super-BCFW recursion cannot be automatically applied sector-by-sector as it is in the massless case. This weakening of the massless helicity selection rules may potentially complicate calculations if little group invariance is to be preserved.

We here illustrate the decomposition into bands of the 3-particle superamplitude, choosing the special case (2.62) for simplicity. To reveal the separate supersymmetric invariant sectors, we explicitly strip off a massive spinor from one of the supercharges. We define

$$\zeta_{1I}^{a} \equiv \frac{1}{m}\left\langle 1_I Q^{\dagger a}\right\rangle \tag{2.75}$$

such that we may write the degenerate component of the delta functions as

$$\delta^{(2)}\left(\left\langle 3Q^{\dagger a}\right\rangle\right) = \delta^{(2)}\left(\left\langle 31^I\right\rangle \zeta_{1I}^{a}\right). \tag{2.76}$$

The distinct bands closed under supersymmetry now correspond to the components of this sum in the helicity basis, so we may exhibit the band structure as

$$\begin{aligned}\mathcal{A}_3[\mathcal{W}_1,\overline{\mathcal{W}}_2,G_3] = {}& \frac{-x}{m^3\left\langle q3\right\rangle^4}\delta^{(2)}\left(\left\langle qQ^{\dagger a}\right\rangle\right)\delta^{(2)}\left(\left\langle q\right| p_1 \left|Q_{a+2}\right]\right) \\ & \times \frac{1}{2}\epsilon_{ab}\left[\left\langle 31^+\right\rangle^2 \zeta_{1+}^{a}\zeta_{1+}^{b} + 2\left\langle 31^+\right\rangle\left\langle 31^-\right\rangle \zeta_{1+}^{a}\zeta_{1-}^{b} + \left\langle 31^-\right\rangle^2 \zeta_{1-}^{a}\zeta_{1-}^{b}\right],\end{aligned} \tag{2.77}$$

where the first term corresponds to the $\overline{\text{MHV}}$ band, the last to the MHV band, and the middle to the MHV $\times$ $\overline{\text{MHV}}$ band, which vanishes in the massless limit. In other little group frames these bands will be scrambled, though still exist as separate supersymmetric

invariants. This decomposition into bands makes it clear the way in which the separate sectors of helicity violation are combined in the massive case.

The four leg superamplitude has only one distinct supersymmetric structure. Just as for the massless case, there is only the MHV sector, which is identical to its parity conjugate $\overline{\text{MHV}}$ sector.

Beyond 4 legs, the bands may be identified by solving the SWIs directly. At five legs, supersymmetry implies that $\mathcal{A}_5 = \delta^{(4)}(Q^{\dagger,a})\delta^{(4)}(Q_{a+2})F$, where $F$ is some function quadratic in Grassmann variables. Proceeding as in the general strategy laid out in (29), the appearance in $F$ of Grassmann variables for two of the lines may be eliminated here using the constraints imposed by the supersymmetric delta functions. Then supersymmetry requires that $Q_a F = 0$ and $Q^{\dagger a+2}F = 0$. These Grassmann PDEs may be solved by finding 'Grassmann characteristics' - combinations of Grassmann variables upon which $F$ cannot depend. Then $F$ is a function of the other independent Grassmann variables that 'label' the characteristics (this resembles the method used in (66) to solve the SWIs). In this manner one may construct linear combinations of Grassmann variables, which we term 'triads', that are annihilated by $Q_a$ and $Q^{\dagger a+2}$ and which include the $\eta$s of only three of the legs.

Choosing a BPS line $i$ and an anti-BPS line $j$, we define Grassmann triads 'anchored' at massless legs $k$ and massive legs $\ell$ as

$$\xi^a_{k,ij} \equiv \eta^a_k + \left(m_j \eta^a_{iI} \left\langle i^I\right| + m_i \eta^a_{jJ} \left\langle j^J\right|\right) \pi_{ij} \left|k\right] / \pi^2_{ij} \tag{2.78}$$

$$\tilde{\xi}^{\dagger a}_{k,ij} \equiv \tilde{\eta}^{\dagger a}_k + \left(m_i \eta^a_{jJ} \left[j^J\right| - m_j \eta^a_{iI} \left[i^I\right|\right) \pi_{ij} \left|k\right\rangle / \pi^2_{ij} \tag{2.79}$$

$$\xi^a_{\ell,ijL} \equiv \eta^a_{\ell L} + \left(m_j \eta^a_{iI} \left\langle i^I\right| + m_i \eta^a_{jJ} \left\langle j^J\right|\right) \pi_{ij} \left|\ell_L\right] / \pi^2_{ij} \pm \left(m_i \eta^a_{jJ} \left[j^J\right| - m_j \eta^a_{iI} \left[i^I\right|\right) \pi_{ij} \left|\ell_L\right\rangle / \pi^2_{ij} \tag{2.80}$$

where the upper sign in the last line is for BPS states and the lower sign for anti-BPS

states, and we have defined $\pi_{ij} \equiv m_i p_j + m_j p_i$ for ease of reference. Since $i, j$ differ in the signs of their central charges, $\pi_{ij}^2 = m_i m_j s_{ij}$, where $s_{ij}$ are the generalized Mandelstam variables. The triads have the massless limits

$$\xi^a_{k,ij} \to \eta^a_k + \frac{[jk]}{[ij]}\eta^a_i + \frac{[ki]}{[ij]}\eta^a_j \equiv \frac{m^a_{ijk}}{[ij]} \tag{2.81}$$

$$\tilde{\xi}^{\dagger a}_{k,ij} \to \tilde{\eta}^{\dagger a}_k + \frac{\langle jk\rangle}{\langle ij\rangle}\tilde{\eta}^{\dagger a}_i + \frac{\langle ki\rangle}{\langle ij\rangle}\tilde{\eta}^{\dagger a}_j \equiv \frac{\tilde{m}^{\dagger a}_{ijk}}{\langle ij\rangle} \tag{2.82}$$

$$\left(\xi^a_{\ell,ij+}, \xi^a_{\ell,ij-}\right) \to \left(\pm\frac{\tilde{m}^{\dagger a}_{ij\ell}}{\langle ij\rangle}, \frac{m^a_{ij\ell}}{[ij]}\right), \tag{2.83}$$

where the massless $m^a_{ijk} = [ij]\,\eta^a_k + [jk]\,\eta^a_i + [ki]\,\eta^a_j$ variables were recognized in (66) as useful for solving the SWIs in the chiral superspace at the origin of moduli space. It is straightforward to take limits where only line $i$ or $j$ becomes massless. In the following we will use the same symbols for triads regardless of the masses of lines $i, j$, and rely on these limits to provide their definitions.

We may now write any superamplitude as a sum of a large-enough set of products of these triads with undetermined coefficients, and then project onto various component amplitudes to fix them. For a 5-leg superamplitude with up to four massive legs, we may characterize the band structure using the triads of a single massless leg as

$$\mathcal{A}_5\left[G_1, V_2, V_3, V_4, V_5\right] = \frac{\delta^{(4)}(Q^{\dagger,a})\delta^{(4)}(Q_{a+2})}{2s^2_{45}}\epsilon_{ab} \times \Big[A_5[g^-, V_-, V_-, V_+, V_+]\xi^a_{1,23}\xi^b_{1,23} + 2A_5[S_{42}+S_{31}, V_-, V_-, V_+, V_+]\xi^a_{1,23}\tilde{\xi}^{\dagger b}_{1,23} + A_5[g^+, V_-, V_-, V_+, V_+]\tilde{\xi}^{\dagger a}_{1,23}\tilde{\xi}^{\dagger b}_{1,23}\Big]. \tag{2.84}$$

Here $V$ is either a massless $G$ or a massive $\mathcal{W}$ or $\overline{\mathcal{W}}$, while $V_\pm$ is the highest- or lowest-weight state in the multiplet, which are respectively $S_{12}, S_{34}$ and $\tilde{\phi}, \phi$ for the massless and massive vectors. We note that the denominator merely cancels out the kinematic factors in the delta function and is not a pole, as the kinematic poles are contained within the

component amplitudes which are here left undetermined.

It is clear in this form that each of the terms is closed under supersymmetry. In the language of (18), the first term in (2.84) is the MHV $\times$ MHV band, the third is its parity conjugate and the second is the MHV $\times$ $\overline{\text{MHV}}$ band (and its conjugate). Notably, this characterization of the bands respects little group covariance, but is determined by the massless multiplet's helicity states.

However, we may alternatively characterize the band structure using the triads of a single massive leg, which identifies the bands with the polarizations of the massive $W$. The 5-leg superamplitude with at least one massive leg may be written as

$$\mathcal{A}_5\left[\mathcal{W}_1, V_2, V_3, V_4, V_5\right] = \frac{\delta^{(4)}(Q^{\dagger,a})\delta^{(4)}(Q_{a+2})}{2s_{45}^2} A_5[W^{(IJ)}, V_-, V_-, V_+, V_+]\epsilon_{ab}\,\xi^a_{1,23I}\xi^b_{1,23J}. \tag{2.85}$$

The comparison of (2.84) and (2.85) thus reflects clearly how the introduction of masses combines amplitudes of different helicity components and how this in turn combines the different bands of the superamplitude.

As is evident in these formulae, the 5-leg superamplitudes have the special property that the bands are each fixed by a single component amplitude, so they may be fully determined once these are known. For the case of two massive legs, (18) used BCFW recursion to derive the partial amplitudes for a massive vector boson, its antiparticle and any number of massless gluons, which, after conversion to the little group covariant notation, may be written as

$$\begin{aligned} &A_n[W_1^{I_1 I_2}, \overline{W}_2^{J_1 J_2}, g_3^+, \dots g_n^+] \\ &\quad = \frac{-\left\langle 1^{I_1} 2^{J_1}\right\rangle \left\langle 1^{I_2} 2^{J_2}\right\rangle [3|\prod_{i=4}^{n-1}(m^2 - (p_i + \dots + p_n + p_1)(p_2 + \dots + p_i))\,|5]}{\langle 34\rangle \langle 45\rangle \dots \langle n-1 n\rangle \prod_{i=4}^{n}((p_2 + \dots + p_{i-1})^2 + m^2)}, \end{aligned} \tag{2.86}$$

where $n-2$ is the number of gluon legs. Likewise, partial amplitudes with any number

of massless scalars $S_{24}$ were derived as

$$A_n[W^{I_1I_2},\overline{W}^{J_1J_2},S_{24},\dots S_{24}]=\frac{m^{n-4}\left[1^{I_1}2^{J_1}\right]\left\langle 1^{I_2}2^{J_2}\right\rangle}{\prod_{i=4}^{n}((p_2+\cdots+p_{i-1})^2+m^2)}, \tag{2.87}$$

where $n-2$ is the number of scalar legs.

For 5-legs, these component amplitudes may be combined into the superamplitude

$$\begin{aligned}\mathcal{A}_5[\mathcal{W}_1,\overline{\mathcal{W}}_2,G_3,G_4,G_5]=&\frac{\delta^{(4)}(Q^{\dagger,a})\delta^{(4)}(Q_{a+2})}{s_{51}s_{23}s_{45}}\epsilon_{ab}\times\\&\left(\frac{\langle 3|\,p_2p_1-m^2\,|5\rangle}{2\,[34]\,\langle 45\rangle}\xi^a_{3,12}\xi^b_{3,12}+m\xi^a_{3,12}\tilde{\xi}^{\dagger b}_{3,12}+\frac{[3|\,p_2p_1-m^2\,|5]}{2\,\langle 34\rangle\,[45]}\tilde{\xi}^{\dagger a}_{3,12}\tilde{\xi}^{\dagger b}_{3,12}\right),\end{aligned} \tag{2.88}$$

where

$$\begin{aligned}\xi^a_{3,12}&=\frac{-1}{s_{12}}\left([3|\,p_1+p_2\left|1^I\right\rangle\eta^a_{1I}+[3|\,p_1+p_2\left|2^J\right\rangle\eta^a_{2J}+s_{12}\eta^a_3\right)\\\tilde{\xi}^{\dagger a}_{3,12}&=\frac{-1}{s_{12}}\left(\langle 3|\,p_1+p_2\left|1^I\right]\eta^a_{1I}-\langle 3|\,p_1+p_2\left|2^J\right]\eta^a_{2J}+s_{12}\tilde{\eta}^{\dagger a}_3\right).\end{aligned} \tag{2.89}$$

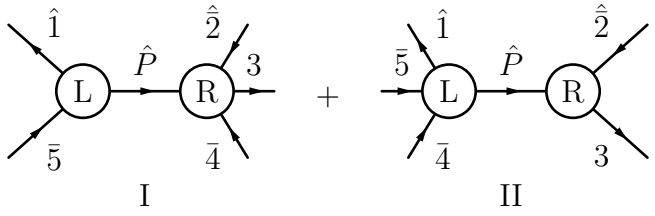
Note that the denominator of the superamplitude is somewhat different from (2.84) as the component amplitudes that have been matched onto are different, but the band structure is still clearly visible in terms of orders in helicity violation. As anticipated, the $\xi^a_{3,12}\tilde{\xi}^{\dagger b}_{3,12}$ term, which represents the MHV $\times$ $\overline{\text{MHV}}$ band, clearly vanishes in the massless limit, leaving the usual MHV sector and its parity conjugate $\overline{\text{MHV}}$.

With more legs, each band can consist of multiple combinations of triads and they are also no longer fixed by single component amplitudes. The exceptions to this, most clearly illustrated if there are enough massless legs for the superamplitude to be described entirely with massless triads, are always the MHV $\times$ MHV band, which corresponds to the only term that is purely holomorphic in triads anchored at massless legs (and analogously for the $\overline{\text{MHV}}\times\overline{\text{MHV}}$ band), and the MHV $\times$ $\overline{\text{MHV}}$ band, which involves a

single term with an equal number of triads and conjugate triads, each of a different type. For example, the 6-leg superamplitude $\mathcal{A}[\mathcal{W}, \overline{\mathcal{W}}, G, G, G, G]$ has bands described by $\xi^a_{3,45}$, $\xi^a_{4,56}$ and their conjugates. The MHV $\times$ MHV band is given by the single holomorphic term $\epsilon_{ab}(\xi^a_{3,45}\xi^b_{3,45})\epsilon_{cd}(\xi^c_{4,56}\xi^d_{4,56})$, the terms in the NMHV $\times$ MHV and MHV $\times$ NMHV bands are of the form $\sim \xi^3\tilde{\xi}$, while the terms in the NMHV $\times$ NMHV band are of the form $\sim \xi^2\tilde{\xi}^2$. However, when most of the legs are massive, there will not be a form in which all of the Grassmann triads are anchored to massless legs and the little group will combine the bands into components of an $SU(2)$ tensor, similar to that observed in (2.85).

In addition to having more available Grassmann structures, terms within each band are related by the massive $R$-symmetry generators (2.14) that are not part of the $U(2)$ linearly represented on the on-shell superspace. A similar analysis to (66) could be performed to determine the Grassmann structure for higher leg superamplitudes. We will instead return our attention toward super-BCFW recursion, which has the capacity to generate complete expressions instead.

**Five Particle Superamplitudes**




I II

Figure 2.2: The two BCFW diagrams for five-point recursion.

Using the insight provided above into the helicity structure of the the 5-leg superamplitude, we proceed to use massive super-BCFW to compute it in full generality. This gives a first non-trivial application of BCFW recursion to computing amplitudes in which every leg is massive.

Much of the ensuing calculation resembles that performed in $6d$ in (21) and (44). However, utilising the interpretation of the bands above, we are able to take short-cuts, despite the calculation presumably being attainable through dimensional reduction and not yet adapted with variables likely accommodating of dual conformal symmetry, as used in (22) and (25).

We will choose to compute the superamplitude $\mathcal{A}[\mathcal{W}_1, \overline{\mathcal{W}}_2, \mathcal{W}_3, \overline{\mathcal{W}}_4, \overline{\mathcal{W}}_5]$ for $m_1 < m_5$. Results for other choices of central charges and masses are obtained by trivial modification. Applying the massive super-BCFW shift to the first and second legs, the superamplitude recurses to the two factorisation channels depicted in Figure 2.2. The resulting superamplitude is

$$\begin{aligned}\mathcal{A}[\mathcal{W}_1, \overline{\mathcal{W}}_2, \mathcal{W}_3, \overline{\mathcal{W}}_4, \overline{\mathcal{W}}_5] = &\int d^4\eta_{\hat{P}} \hat{\mathcal{A}}_L[\overline{\mathcal{W}}_5, \widehat{\mathcal{W}}_1, \mathcal{W}_{\hat{P}}] \frac{-1}{s_{15}} \hat{\mathcal{A}}_R[\overline{\mathcal{W}}_{-\hat{P}}, \widehat{\overline{\mathcal{W}}}_2, \mathcal{W}_3, \overline{\mathcal{W}}_4]\Big|_{z_*^{(1)}} \\ &+ \int d^4\eta_{\hat{P}} \hat{\mathcal{A}}_L[\overline{\mathcal{W}}_4, \overline{\mathcal{W}}_5, \widehat{\mathcal{W}}_1, \mathcal{W}_{\hat{P}}] \frac{-1}{s_{23}} \hat{\mathcal{A}}_R[\overline{\mathcal{W}}_{-\hat{P}}, \widehat{\overline{\mathcal{W}}}_2, \mathcal{W}_3]\Big|_{z_*^{(2)}}. \quad (2.90)\end{aligned}$$

Each term is to be evaluated upon a different pole, respectively determined to be at

$$\begin{aligned} s_{\hat{1}5} = 0 \Rightarrow z_*^{(1)} &= \frac{s_{15}}{2r \cdot p_5} \\ s_{\hat{2}3} = 0 \Rightarrow z_*^{(2)} &= \frac{-s_{23}}{2r \cdot p_3}. \end{aligned} \quad (2.91)$$

In each term, combining the delta functions to produce the full 5-leg supersymmetric delta function is easy, as, on the support of the 3-leg superamplitude's delta function, the 4-leg superamplitude's delta function is equivalent to the overall delta function for the full superamplitude. This leaves the 3-leg delta function to be integrated in the state sum.

The calculation may be continued by substituting the shifted momenta and Grassmann variables into the two terms, adding them together and simplifying. Outside the

total supercharge conserving delta functions, only $\eta_1$, $\eta_2$, $\eta_3$ and $\eta_5$ manifestly appear in the two terms above. However, we know that these must able to be arranged (after use of the constraints imposed by the delta functions) into the triad structure discussed above, such as that succinctly presented in (2.85). Choosing to represent the superamplitude using triads $\xi^a_{3,12}$, we only need to identify the coefficient of the terms containing a factor of $\epsilon_{ab}\eta^a_{3K_1}\eta^b_{3K_2}$ to bootstrap the entire superamplitude. This is because, in this case, the massive bands are each determined by a polarisation component of a single component amplitude, $A[\phi_1,\phi_2,W_3^{K_1K_2},\tilde{\phi}_4,\tilde{\phi}_5]$. This allows us to henceforth discard all terms in the calculation that do not have a factor of $\epsilon_{ab}\eta^a_{3K_1}\eta^b_{3K_2}$, but only after first using the delta function constraints to eliminate $\eta_4$ and $\eta_5$, the latter of which appears in the first term.

Inverting the delta function constraints $Q=0$ and $Q^\dagger=0$ for the supercharges implies that, in expressing $\left|5^M\right\rangle\eta_{5M}$ as a linear combination of $\eta_1$, $\eta_2$ and $\eta_3$, the latter term is

$$\frac{m_4m_5}{s_{45}}\left(1-\frac{p_5p_4}{m_4m_5}\right)\left(\frac{p_4}{m_4}+\frac{p_3}{m_3}\right)\left|3^K\right]\eta_{3K}=\frac{m_4m_5}{s_{45}}\left|A^K\right\rangle\eta_{3K}, \tag{2.92}$$

where we define the spinor $\left|A^K\right\rangle$ above to condense notation (we do not bother here to present the terms proportional to other Grassmann variables, as these do not contribute to the $\eta_3^2$ term in the superamplitude).

Resuming our calculation of the BCFW diagrams in Figure 2.2, the first diagram contributes

$$\begin{aligned}\int d^4\eta_{\hat{P}}\hat{\mathcal{A}}_L[\overline{\mathcal{W}}_5,\widehat{\mathcal{W}}_1,\mathcal{W}_{\hat{P}}]\frac{-1}{s_{15}}\hat{\mathcal{A}}_R[\overline{\mathcal{W}}_{-\hat{P}},\widehat{\overline{\mathcal{W}}}_2,\mathcal{W}_3,\overline{\mathcal{W}}_4]\Big|_{z_*^{(1)}}\\ =\frac{\delta^{(4)}(Q)\delta^{(4)}(Q^\dagger)}{s_{15}s_{\hat{2}3}s_{34}}\frac{1}{\langle q|\,p_5\hat{p}_1\,|q\rangle}\prod_a\left(m_1\left\langle q5^M\right\rangle-\langle q|\,\hat{p}_1\left|5^M\right]\right)\eta^a_{5M}\Big|_{z_*^{(1)}},\end{aligned} \tag{2.93}$$

where we have performed the Grassmann intergral and retained only the $\eta_5^2$ terms.

The exchange symmetry of the little group indices of the factors contracted against the $\eta_5$ variables implies that, by fermion statistics, only the component of the product of $\eta_5$ variables that is antisymmetric in $R$-indices provides a non-zero contribution, so $\eta^1_{5M_1}\eta^2_{5M_2} \sim \frac{-1}{2}\epsilon_{ab}\eta^a_{5M_1}\eta^b_{5M_2}$. Then applying the Schouten identity, (2.93) can be simplified to

$$\frac{\delta^{(4)}(Q)\delta^{(4)}(Q^\dagger)}{s_{15}s_{\hat{2}3}s_{34}}\frac{-1}{m_5^2}\left\langle 5^{M_1}\right|\hat{p}_1 p_5\left|5^{M_2}\right\rangle\left(\frac{-1}{2}\epsilon_{ab}\eta^a_{5M_1}\eta^b_{5M_2}\right) \tag{2.94}$$

(leaving implicit evaluation on the first residue). Substituting in (2.92) gives the $\eta_3^2$ contribution from the first BCFW diagram

$$\frac{\delta^{(4)}(Q)\delta^{(4)}(Q^\dagger)}{s_{15}s_{\hat{2}3}s_{34}s_{45}^2}(-m_4^2)\left\langle A^{K_1}\right|\hat{p}_1 p_5\left|A^{K_2}\right\rangle\left(\frac{-1}{2}\epsilon_{ab}\eta^a_{3K_1}\eta^b_{3K_2}\right). \tag{2.95}$$

The spinor bilinear in the term above may be simplified to

$$\begin{aligned}\left\langle A^{K_1}\right|\hat{p}_1 p_5\left|A^{K_2}\right\rangle = \frac{1}{m_3 m_4^2}\Bigg(&\left\langle 3^{K_1}\right|p_4\left|3^{K_2}\right]\left(s_{45}s_{\hat{1}3}-s_{\hat{1}4}s_{35}\right)\\ &-\left\langle 3^{K_1}\right|\hat{p}_1\left|3^{K_2}\right]s_{45}s_{34}+\left\langle 3^{K_1}\right|p_5\left|3^{K_2}\right]s_{\hat{1}4}s_{34}\Bigg).\end{aligned} \tag{2.96}$$

The second BCFW diagram in Figure 2.2 may be evaluated almost identically to the first. In this case no factors of $\eta_4$ or $\eta_5$ appear, so only the coefficient of the $\eta_3^2$ term needs to be retained. This contributes

$$\frac{\delta^{(4)}(Q)\delta^{(4)}(Q^\dagger)}{s_{45}s_{23}s_{\hat{1}5}}\frac{-1}{m_3}\left\langle 3^{K_1}\right|\hat{p}_2\left|3^{K_2}\right]\left(\frac{-1}{2}\epsilon_{ab}\eta^a_{3K_1}\eta^b_{3K_2}\right)\Big|_{z_*^{(2)}}. \tag{2.97}$$

The next step is to add the two BCFW terms together and combine them into a simplified expression. Explicitly evaluated on the residues (2.91), the shifted Mandelstam

invariants appearing in each term may be expressed as

$$s_{\hat{23}}\Big|_{z_*^{(1)}} = \frac{1}{r\cdot p_5}\left(s_{23}(r\cdot p_5)+s_{15}(r\cdot p_3)\right), \qquad s_{\hat{15}}\Big|_{z_*^{(2)}} = \frac{1}{r\cdot p_3}\left(s_{23}(r\cdot p_5)+s_{15}(r\cdot p_3)\right). \tag{2.98}$$

Following (21) by calling $\phi = s_{23}r\cdot p_5 + s_{15}r\cdot p_3$, the two BCFW terms can be combined to give

$$-\frac{\delta^{(4)}(Q)\delta^{(4)}(Q^\dagger)}{m_3 s_{23}s_{34}s_{45}s_{51}\phi}\Big(\frac{s_{23}}{s_{45}}(r\cdot p_5)\Big(\left\langle 3^{K_1}\right| p_4 \left|3^{K_2}\right]\left(s_{45}s_{\hat{13}}-s_{\hat{14}}s_{35}\right) - \left\langle 3^{K_1}\right| \hat{p}_1 \left|3^{K_2}\right] s_{45}s_{34} \\ + \left\langle 3^{K_1}\right| p_5 \left|3^{K_2}\right] s_{\hat{14}}s_{34}\Big)\Big|_{z_*^{(1)}} + s_{34}s_{15}(r\cdot p_3)\left\langle 3^{K_1}\right| \hat{p}_2 \left|3^{K_2}\right]\Big|_{z_*^{(2)}}\Big). \tag{2.99}$$

The terms proportional to $\left\langle 3^{K_1}\right| r \left|3^{K_2}\right]$ arising from the shifted momenta cancel after substituting the residues. In order to progress further, the ambiguity from momentum conservation and the mass selection rule can be fixed to help combine terms. Choosing to do this by eliminating $p_4$ and $m_4$ from the expression, the special identities between Mandelstam invariants introduced just prior to the computation of the 4-leg superamplitude in Section 2.5.2 can be used for simplification. In doing so, all remaining dependence on the shift vector $r$ in the numerator of (2.99) factorises into a factor of $\phi$ and thus cancels against that in the denominator. This leaves an expression for the $\eta_3^2$ term in the 5-leg superamplitude that is independent of the shift vector and is little group covariant. The full superamplitude may then be obtained by the replacement $\eta_3 \mapsto \xi_{3,12}$. The result is

$$\mathcal{A}[\mathcal{W}_1,\overline{\mathcal{W}}_2,\mathcal{W}_3,\overline{\mathcal{W}}_4,\overline{\mathcal{W}}_5] = -\frac{\delta^{(4)}(Q)\delta^{(4)}(Q^\dagger)}{2m_3 s_{51}s_{23}s_{34}s_{45}^2}\Big(s_{23}(s_{15}+s_{25})\left\langle 3^{K_1}\right| p_1 \left|3^{K_2}\right] \\ - s_{23}s_{12}\left\langle 3^{K_1}\right| p_5 \left|3^{K_2}\right] + (s_{35}s_{12}-s_{13}(s_{25}+s_{15}))\left\langle 3^{K_1}\right| p_2 \left|3^{K_2}\right]\Big)\epsilon_{ab}\xi^a_{3,12K_1}\xi^b_{3,12K_2}, \tag{2.100}$$

which agrees precisely with (2.88) in the appropriate limit.

Remarkably, at no point in the calculation is the identity of the shift vector $r$ actually needed - it cancels out in the end. However, just as for the cancellation of spurious poles observed in massless recursion at higher legs, this only occurs after contributions from distinct factorisation channels (BCFW diagrams) are added together. This means that, despite manifestly breaking little group covariance, recursion nevertheless delivers a little group invariant expression. While invariance is not manifest term-by-term, it is broken in a controlled way. The shift vector seems only to be the needle threading the factorisation channels into a complete superamplitude. This clearly invites a search for an alternative picture of how the factorisation channels are being combined. Especially important to be investigated is the significance of the little group breaking in the BCFW representation of the superamplitude for dual (super)conformal invariance.

Although the 5-leg superamplitude does have non-trivial distinct bands, it is nevertheless an especially simple example in which each band is determined by a single supersymmetry invariant, and hence component amplitude, so that only the $U(2)$ $R$-symmetry subgroup provides non-trivial, independent constraints. This may be anticipated from its massless counterpart, which consists only of MHV and a distinct, yet parity-conjugate, $\overline{\text{MHV}}$ sector. These are especially simple to derive using massless super-BCFW recursion. In the massive case considered here, the three different bands, most clearly visible in (2.88) when some of the legs are massless, may be directly attributed to those of the three-leg superamplitude that are fused in super-BCFW recursion along the factorisation channel. We leave to be explored exactly how a massive manifestation of dual conformal invariance, which, for massless superamplitudes, is provided through super-BCFW, may interplay with both the little group and the band structure. Many simplifying features at five legs will not be present at six legs, which will provide a more acute test of the symmetries, their constraining power and the usefulness and meaning of recursion.

It was proposed by (24) that the $6d$ SYM superamplitudes (or equivalently, the $4d$ massive superamplitudes) could be entirely determined (or "uplifted") by their restrictions to $4d$ massless states. This made use of the expressions using dual variables, in which both dual conformal and permutation symmetries can be made manifest with relatively simple expressions for the superamplitudes. The uplift was demonstrated up to 5 legs, where the compact structure made it obvious by eye, once compact $4d$ and $6d$ building blocks manifesting the dual symmetries were identified. However, complications were encountered in (25) at six legs, where the form of the $4d$ superamplitude produced by BCFW recursion in non-chiral superspace was not automatically amenable to the uplift. Again, the difference arises because of the new, independent bands and their additional structures.

Because we are not fully manifesting the symmetries, in particular parity (through our use of chiral spinors) and the $6d$ Lorentz invariance, the uplift from (2.88) (or its fully massless limit) to (2.100) is not obvious, although there is a clear resemblance in the structures, especially in the way that the bands are combined (the converse operation, the massless limit, is easily seen and verified and relates the terms in the two expressions). It is suggested in (25) that the MHV sector by itself in the massless theory would be sufficient to determine the entire massive superamplitude, if the uplift were correct. This is plausible for the five leg case here, where the MHV sector corresponds to a single little group combination of the triads in (2.100). However, the continuation at six legs is the real test. The extent to which the embedded massless theory controls the structure of the massive theory remains an open question.

## 2.6 Conclusion

The spinor helicity formalism has provided a set of variables with respect to which a broad set of phenomena can be formulated and uncovered on-shell purely through recourse to fundamental principles of quantum mechanics and relativity, without introducing quantum fields and path integrals and their associated unphysical redundancies. The little group has provide an organisation of these variables enabling them to be adapted to insightfully describe the kinematics of massive particles. Treating both massive and massless states on the same footing, it may then be determined precisely to what extent features of quantum field theories are emergent from assumptions about infrared properties. Supersymmetry provides an idealisation that is already known to enhance many of the on-shell properties of unbroken Yang-Mills amplitudes.

The power of on-shell methods for massive theories may be strongest on the Coulomb branch of $\mathcal{N} = 4$ SYM, the maximally supersymmetric theory of massive particles, just as they are at the origin of moduli space for massless states. As a first step toward fully determining the on-shell properties of the theory, we have determined the elementary three-leg superamplitudes. These superamplitudes surprisingly have kinematic factors in their denominators akin to those of massless (super-)Yang-Mills, despite this not being a feature of their component amplitudes. Using super-BCFW recursion for amplitudes of massive particles, we have shown how, by combining on-shell 3-particle superamplitudes across a factorisation channel, a new pole emerges that completes the 4-leg massive superamplitude. This pole arises from combining supersymmetry invariants across the factorisation channel, an operation that simultaneously ensures that the arbitrary reference spinors in the 3-particle superamplitudes are cancelled. This property is not a feature of the non-supersymmetric Higgsed Yang-Mills counterpart. We have then provided the first non-trivial use of BCFW recursion to compute a scattering amplitude entirely in-

volving massive particles, doing so to determine the general 5-leg superamplitude on the Coulomb branch.

The next objective is to compute higher leg superamplitudes. We have shown here that massive super-BCFW recursion offers an avenue for doing this. However, guidance is still necessary for interpreting the expressions that it leaves. Just as for the massless superamplitudes, such a beacon may be provided by dual conformal symmetry. Super-BCFW for massless amplitudes was crucial in deriving a representation in which the dual superconformal symmetry could be deduced (as a sum over $R$-invariants that, in momentum twistor space, makes dual conformal symmetry manifest). However, its full consequences for the massive amplitudes and relationship with the little group has yet to be fully elucidated. Also, while we have demonstrated that super-BCFW is indeed valid, we expect that, just as at the origin of moduli space, this is more directly a consequence (or maybe expression of) dual conformal symmetry or a deeper structure. The hypothesized Grassmannian formulation of the $\mathcal{N} = 4$ SYM amplitudes on the Coulomb branch - the 'symplectic Grassmannian' (61), may make this more explicit. The $6d$ point of view may also provide the framework within which these structures can be seen (27), (28).

Having established the on-shell properties of this idealised theory, the extent to which they descend to theories with less supersymmetry remains to be explored. In the massless theory, the constructibility of the superamplitudes descend to those of pQCD. We have given a brief discussion of how certain tree superamplitudes may be projected down to theories of less supersymmetry in Appendix 2.B. Further progress would require a strategy for projecting out effectively closed subsectors or finding an adaptation of massive (super)-BCFW recursion to these theories.

# Acknowledgments

We thank Tim Cohen, Nathaniel Craig and Henriette Elvang for comments on a draft of this work, Nathaniel Craig for discussions and support during the completion of this work, and Nima Arkani-Hamed and Yu-tin Huang for discussions on (30). AH and SK are grateful for the support of a Worster Fellowship. This work is supported in part by the US Department of Energy under the grant DE-SC0014129.

# Appendices

## 2.A Four Particle Superamplitude Details

To begin combining the delta functions in the 3-leg superamplitudes, we express both left and right superamplitudes in the form in the first line of (2.53). Clearly $\delta^{(4)}(\hat{Q}_L)\delta^{(4)}(\hat{Q}_R) = \delta^{(4)}(Q)\delta^{(4)}(\hat{Q}_R)$. By construction, $\hat{Q}_L + \hat{Q}_R = Q$ is unshifted (and similarly for the conjugate supercharges). Representing the right delta function as the second line in (2.53) and using $\delta^{(4)}(\hat{Q}_R) = \frac{1}{m_1^2 \langle q|p_4\hat{p}_1|q\rangle}\delta^{(2)}\left(\langle q|\,\hat{p}_1\left|\hat{Q}_R\right]\right)\delta^{(2)}\left(u_{4L}\left\langle 4^L\right|\hat{p}_1\left|\hat{Q}_R\right]\right)$ gives $\delta^{(2)}\left(\langle q|\,\hat{p}_1\left|\hat{Q}_L\right]\right)\delta^{(2)}\left(\langle q|\,\hat{p}_1\left|\hat{Q}_R\right]\right) = \delta^{(2)}\left(\langle q|\,\hat{p}_1\,|Q]\right)\delta^{(2)}\left(\langle q|\,\hat{p}_1\left|\hat{Q}_R\right]\right)$. Note that the same reference spinor $|q\rangle$ may be used for both left and right factors. Such a spinor always exists that is parallel to neither $u_{4L}\left\langle 4^L\right|$ nor $u_{3K}\left\langle 3^K\right|$. This leaves the remaining delta functions $\delta^{(2)}\left(\left\langle q\hat{Q}_R^\dagger\right\rangle\right)\delta^{(2)}\left(\langle q|\,\hat{p}_1\left|\hat{Q}_R\right]\right)\delta^{(2)}\left(u_{4L}\left\langle 4^L\right|\hat{p}_1\left|\hat{Q}_R\right]\right)$ from which to extract the final factor required for the full 4-leg delta function.

On the combined support of the other delta functions, $\delta^{(2)}\left(u_{4L}\left\langle 4^L\right|\hat{p}_1\left|\hat{Q}_R\right]\right)$ $= C^2\delta^{(2)}\left(u_{4L}\left\langle 4^L\right|\hat{p}_1\,|Q]\right)$, where the constant $C^2 = \frac{m_1^2}{\left(u^{(L)}_{\hat{P}M}u^{(R)M}_{\hat{P}}\right)^2}\frac{\langle q|p_3\hat{p}_2|q\rangle}{\langle q|p_4\hat{p}_1|q\rangle}$. This follows from the relations

$$u^{(R)}_{\hat{P}M}\left\langle\hat{P}^M\right| = \alpha\, u^{(L)}_{\hat{P}M}\left\langle\hat{P}^M\right| + \beta\,\langle q| \tag{2.101}$$

$$u^{(R)}_{\hat{P}M}\left[\hat{P}^M\right| = \alpha\, u_{4L}\left\langle 4^L\right|\frac{\hat{p}_1}{m_1} + \beta\,\langle q|\,\frac{\hat{P}}{m_P}, \tag{2.102}$$

where

$$\alpha = \frac{u_{3K}\left\langle 3^K q\right\rangle}{u_{4L}\left\langle 4^L q\right\rangle} \qquad \beta = \frac{u_{3K}\left\langle 3^K 4^L\right\rangle u_{4L}}{\langle q4^L\rangle\, u_{4L}}. \tag{2.103}$$

The scalar coefficients of the spinors in the first line above have been obtained by use of (2.43). The second line may be obtained by (2.43) and the Weyl equations. On the support of $\delta^{(2)}\left(\left\langle q\hat{Q}_R^\dagger\right\rangle\right)$ and $\delta^{(4)}(Q^\dagger) \propto \delta^{(4)}\left(\left\langle q\hat{Q}_R^\dagger\right\rangle + \left\langle q\hat{Q}_L^\dagger\right\rangle\right)$, then $\left\langle q\hat{Q}_R^\dagger\right\rangle, \left\langle q\hat{Q}_L^\dagger\right\rangle \sim$

0 and terms proportional to these in the other delta functions may be dropped. Thus

$$\begin{aligned}
\delta^{(4)}(Q^\dagger) &\propto \delta^{(2)}(u^{(R)}_{\hat{P}M}\left\langle \hat{P}^M Q^\dagger \right\rangle) \\
&= \delta^{(2)}(u^{(R)}_{\hat{P}M}\left\langle \hat{P}^M \hat{Q}^\dagger_R \right\rangle + \alpha u^{(L)}_{\hat{P}M}\left\langle \hat{P}^M \hat{Q}^\dagger_L \right\rangle) \\
&= \delta^{(2)}(u^{(R)}_{\hat{P}M}\left[\hat{P}^M \hat{Q}_R\right] + \alpha u^{(L)}_{\hat{P}M}\left[\hat{P}^M \hat{Q}_L\right]) \\
&= \delta^{(2)}\left(\alpha u_{4L}\left\langle 4^L\right|\frac{\hat{p}_1}{m_1}\left(\left|\hat{Q}_R\right] + \left|\hat{Q}_L\right]\right) + \beta u_{4L}\left\langle 4^L\right|\frac{\hat{P}}{m_P}\left|\hat{Q}_R\right]\right) \\
&= \delta^{(2)}\left((\alpha+\beta\gamma) u_{4L}\left\langle 4^L\right|\frac{\hat{p}_1}{m_1}\left|\hat{Q}_R\right] + \alpha u_{4L}\left\langle 4^L\right|\frac{\hat{p}_1}{m_1}\left|\hat{Q}_L\right]\right) \\
\Rightarrow u_{4L}\left\langle 4^L\right|\hat{p}_1\left|\hat{Q}_R\right] &\sim \frac{-\alpha}{\alpha+\beta\gamma} u_{4L}\left\langle 4^L\right|\hat{p}_1\left|\hat{Q}_L\right].
\end{aligned} \tag{2.104}$$

In the penultimate line, it has been used that

$$\langle q|\frac{\hat{P}}{m_P} = \gamma u_{4L}\left\langle 4^L\right|\frac{\hat{p}_1}{m_1} + \lambda\langle q|\frac{\hat{p}_1}{m_1}, \tag{2.105}$$

where the identity of the scalar $\lambda$ is unimportant and

$$\gamma = -\frac{\langle q|\hat{p}_1 p_4|q\rangle}{m_P m_1\langle q4^L\rangle u_{4L}}. \tag{2.106}$$

Then

$$\begin{aligned}
\delta^{(2)}\left(u_{4L}\left\langle 4^L\right|\hat{p}_1\left|\hat{Q}_R\right]\right) &= \delta^{(2)}\left(\frac{\alpha}{\beta\gamma}u_{4L}\left\langle 4^L\right|\hat{p}_1\left|\hat{Q}_R\right] + \frac{\alpha}{\beta\gamma}u_{4L}\left\langle 4^L\right|\hat{p}_1\left|\hat{Q}_R\right]\right) \\
&= \delta^{(2)}\left(\left(\frac{\alpha}{\beta\gamma}\right)u_{4L}\left\langle 4^L\right|\hat{p}_1|Q]\right)
\end{aligned} \tag{2.107}$$

Thus $C^2 = \left(\frac{\alpha}{\beta\gamma}\right)^2$ and the expression stated above may be obtained upon simplification through use of (2.43), and the spin sums and the Weyl equations laid out in Appendix A of (29).

The delta function $\delta^{(2)}\left(u_{4L}\left\langle 4^L\right|\hat{p}_1\left|Q\right]\right)$ may be amalgamated with the factor of $\frac{1}{m_1^2\langle q|p_4\hat{p}_1|q\rangle}\delta^{(2)}\left(\langle q|\,\hat{p}_1\,|Q]\right)$ derived above to give $\delta^{(4)}(Q)$.

Use of the alternative representation of the 3-leg SUSY delta function in (2.58) may possibly be yield a simpler computation in this case once the Grassmann integral is performed (see (21; 25) for how this is done similarly in $6d$).

To derive (2.72), using the spin sums and special massive kinematics (2.42) (and that $s_{\hat{1}\hat{2}} = s_{12}$ is unshifted),

$$\begin{aligned}
-u_{\hat{1}I}u_{\hat{1}L}s_{12} &= u_{\hat{1}I}u_{\hat{1}N}\epsilon_{ML}\left(\left\langle\hat{1}^N\hat{2}_J\right\rangle\left\langle\hat{2}^J\hat{1}^M\right\rangle + \left[\hat{1}^N\hat{2}_J\right]\left[\hat{2}^J\hat{1}^M\right] + \left\langle\hat{1}^N\right|\hat{p}_2\left|\hat{1}^M\right] - \left[\hat{1}^N\right|\hat{p}_2\left|\hat{1}^M\right\rangle\right) \\
&= u_{\hat{1}I}u^{(L)}_{\hat{P}N}\epsilon_{ML}\Big(\left\langle\hat{P}^N\hat{2}_J\right\rangle\left\langle\hat{2}^J\hat{1}^M\right\rangle + \left[\hat{P}^N\hat{2}_J\right]\left[\hat{2}^J\hat{1}^M\right] \\
&\qquad\qquad + \left\langle\hat{P}^N\right|\hat{p}_2\left|\hat{1}^M\right] - \left[\hat{P}^N\right|\hat{p}_2\left|\hat{1}^M\right\rangle\Big) \\
&= u_{\hat{1}I}u^{(L)}_{\hat{P}N}u^{(R)N}_{\hat{P}}u_{2J}\epsilon_{ML}\left(\left[\hat{2}^J\hat{1}^M\right] - \left\langle\hat{2}^J\hat{1}^M\right\rangle\right) \\
&= u^{(L)}_{\hat{P}N}u^{(R)N}_{\hat{P}}\epsilon_{ML}u_{\hat{1}I}u^{(R)}_{\hat{P}J}\left(\left[\hat{P}^J\hat{1}^M\right] - \left\langle\hat{P}^J\hat{1}^M\right\rangle\right) \\
&= u^{(L)}_{\hat{P}N}u^{(R)N}_{\hat{P}}u_{\hat{1}I}u^{(R)}_{\hat{P}J}u_{\hat{1}L}u^{(R)J}_{\hat{P}} = u_{\hat{1}I}u_{\hat{1}L}\left(u^{(L)}_{\hat{P}N}u^{(R)N}_{\hat{P}}\right)^2 \\
&\Rightarrow \left(u^{(L)}_{\hat{P}N}u^{(R)N}_{\hat{P}}\right)^2 = -s_{12}.
\end{aligned} \tag{2.108}$$

## 2.B $\mathcal{N} < 4$ SYM Superamplitudes from $\mathcal{N} = 4$ SYM

In this appendix we investigate how tree-level superamplitudes in Yang-Mills theories with less-than-maximal supersymmetry may be constructed from $\mathcal{N} = 4$ SYM to determine the extent to which the valid BCFW shift may be exploited.

### 2.B.1 Massless $\mathcal{N} < 4$ SYM

It was observed in (67) that one may extract $\mathcal{N} = 0, 1, 2$ submultiplets from the $\mathcal{N} = 4$ massless vector multiplet via derivation or deletion of Grassmann variables.

After defining extraction operators on states, one may then act with these operators on on-shell superamplitudes to find subamplitudes which describe the interactions of the submultiplets inside the $\mathcal{N} = 4$ states.

Subsequent to extraction, it is of interest to investigate whether the spectrum may be truncated in order to obtain superamplitudes of theories with fewer supersymmetries. In particular, we would like to know when it is possible for the states that have been removed from the external legs of the superamplitude by the extraction process to be omitted from the theory altogether while still retaining a meaningful superamplitude. Calling $S$ a set of extracted superfields closed under some $\mathcal{N} < 4$ supersymmetries, we say that $S$ forms a 'closed subsector' of the $\mathcal{N} = 4$ tree-level theory if, for any tree-level subamplitude with external states only in $S$, it contains no contributions from off-shell states not in $S$. Then after extracting the subamplitudes of states in $S$, we may truncate the spectrum by ignoring the other states and we find the tree-level theory of states in $S$ enjoying some $\mathcal{N} < 4$ supersymmetry. We denote the steps of extraction and truncation together as 'projection' and say that this procedure projects from $\mathcal{N} = 4$ SYM to the lower $\mathcal{N}$ theory.

The case discussed in (67) is projection to pure (S)YM with $\mathcal{N} < 4$. That is, one sets $S = \{\mathcal{N} < 4 \,\text{vector multiplet}\}$. The truncation is valid here because in any $\mathcal{N}$ (S)YM, all other states only couple to the vector multiplets in pairs, so lines of these particles in Feynman diagrams cannot be produced internally without closing in loops. So we may extract any tree-level amplitude in, for example, pure $\mathcal{N} = 1$ SYM, $\mathcal{A}_n[G^\pm, G^\pm, \ldots, G^\pm]_{\mathcal{N}=1}$ (where $G^\pm$ are massless $\mathcal{N} = 1$ gluon superfields), from the tree-level amplitude $\mathcal{A}_n[G, G, \ldots, G]_{\mathcal{N}=4}$ in a manner we will make precise momentarily.

The procedure for finding the extraction operators begins by first choosing which subset of supersymmetries our residual coherent states will be built out of. We then find the different ways one may take derivatives with respect to or delete the Grassmann

variables corresponding to supersymmetries which will disappear. We end up manifestly with coherent states of the remaining supersymmetries.

Consider first the extraction of $\mathcal{N}=2$ submultiplets which are coherent states of $Q_a$, $a=1,2$ (thus reducing to a chiral superspace). For the massless multiplet, we may see on-shell the familiar statement that it consists of one $\mathcal{N}=2$ vector multiplet $G^{\pm}$ and one hypermultiplet $K$. With this choice of remaining supersymmetries, we may isolate these submultiplets from the massless $\mathcal{N}=4$ multiplet of (2.11) via

$$G^{+}_{\mathcal{N}=2} = \frac{1}{2}\frac{\partial}{\partial\eta^{\dagger m}}\frac{\partial}{\partial\eta^{\dagger}_{m}}G, \qquad G^{-}_{\mathcal{N}=2} = G\big|_{\eta^{\dagger}_{m}\to 0} \tag{2.109}$$

$$K^{+}_{\mathcal{N}=2} = \frac{\partial}{\partial\eta^{\dagger}_{3}}G\big|_{\eta^{\dagger}_{4}\to 0}, \qquad \bar{K}^{+}_{\mathcal{N}=2} = \frac{\partial}{\partial\eta^{\dagger}_{4}}G\big|_{\eta^{\dagger}_{3}\to 0}, \tag{2.110}$$

where the superscript refers to the helicity of the on-shell supermultiplet, and $K$ and $\bar{K}$ are the two $CP$-conjugate $\mathcal{N}=1$ chiral multiplets into which the $\mathcal{N}=2$ hypermultiplet may be decomposed. Here and throughout this appendix we use the $SU(2)\times SU(2)$ bronken $R$-symmetry notation for the superspace as in the first equation of (2.11).

If we were to instead extract the $\mathcal{N}=2$ submultiplets which are closed under $Q_1$ and $Q_3$, we would naturally end up in the non-chiral superspace for the massless multiplets

$$G^{+}_{\mathcal{N}=2} = \frac{\partial}{\partial\eta^{\dagger}_{4}}G\big|_{\eta^{2}\to 0}, \qquad G^{-}_{\mathcal{N}=2} = \frac{\partial}{\partial\eta^{2}}G\big|_{\eta^{\dagger}_{4}\to 0} \tag{2.111}$$

$$K_{\mathcal{N}=2} = G\big|_{\eta^{2},\eta^{\dagger}_{4}\to 0}, \qquad \bar{K}_{\mathcal{N}=2} = \frac{\partial}{\partial\eta^{\dagger}_{4}}\frac{\partial}{\partial\eta^{2}}G. \tag{2.112}$$

We may also go further and extract the $\mathcal{N}=1$ submultiplets from $\mathcal{N}=4$, where we see the massless supermultiplet decompose into a vector multiplet (and $CP$ conjugate)

$G^{\pm}$ and three chiral multiplets (and conjugate pairs) $\chi^m$ (for $m = 2, 3, 4$) as

$$G^{+}_{\mathcal{N}=1} = \frac{1}{2}\frac{\partial}{\partial\eta^{\dagger m}}\frac{\partial}{\partial\eta^{\dagger}_{m}}G\big|_{\eta^2\to 0}, \qquad G^{-}_{\mathcal{N}=1} = \frac{\partial}{\partial\eta^2}G\big|_{\eta^{\dagger}_{m}\to 0} \tag{2.113}$$

$$\chi^{2+}_{\mathcal{N}=1} = \frac{1}{2}\frac{\partial}{\partial\eta^{\dagger m}}\frac{\partial}{\partial\eta^{\dagger}_{m}}\frac{\partial}{\partial\eta^2}G, \qquad \chi^{2-}_{\mathcal{N}=1} = G\big|_{\eta^{\dagger}_{3},\eta^{\dagger}_{4},\eta^2\to 0}, \tag{2.114}$$

$$\chi^{m+}_{\mathcal{N}=1} = \frac{\partial}{\partial\eta^{\dagger}_{m}}G\big|_{\eta^{\dagger m},\eta^2\to 0}, \qquad \chi^{m-}_{\mathcal{N}=1} = \frac{\partial}{\partial\eta^{\dagger m}}\frac{\partial}{\partial\eta^2}G\big|_{\eta^{\dagger}_{m}\to 0} \tag{2.115}$$

where $m$ indexes which $R$-index we took a derivative with respect to, which is merely a more compact notation than we used for the massless hypermultiplet above.

Since the $\mathcal{N} = 2$ vector multiplet forms a closed subsector in pure SYM on its own, it can be split up into $\mathcal{N} = 1$ submultiplets $G^{\pm}$, $\chi^{4\pm}$. As discussed in (68), one may find amplitudes for fundamental quarks in ($\mathcal{N} = 0$) QCD from color-ordered amplitudes of adjoint gluinos merely by using different color factors when summing over color-orderings. Likewise, one may study $\mathcal{N} = 1$ SQCD with one flavor of massless fundamental quark chiral superfield at tree-level using this construction.

We can then proceed even further and go to $\mathcal{N} = 0$ non-supersymmetric Yang-Mills by simply considering each component field separately. As above, one may find closed subsectors from $\mathcal{N} = 1, 2$ supersymmetry which include, in addition to the gluons, massless fermions (from $\mathcal{N} = 1$) or both massless fermions and scalars (from $\mathcal{N} = 2$). Special combinations of amplitudes in the projected $\mathcal{N} = 2$ SYM theory were used by (68) in order to compute tree QCD amplitudes with multiple quark flavours while avoiding internal off-shell interactions with their scalar partners.

### 2.B.2 Massive $\mathcal{N} < 4$ SYM

The next question is whether any of this structure survives on the $\mathcal{N} = 4$ Coulomb branch now that we have another type of multiplet. We will not, in fact, find closed sub-

sectors which include massive states, for the reason that any massive submultiplet couples at tree-level to all of the massless submultiplets, which will be apparent after we give the extraction operators for massive states. However, we will be able to find effectively closed subsectors by restricting our attention to certain subsets of (super)amplitudes, in which only states in that subsector appear internally. This will allow us to deduce some interesting features of various $\mathcal{N} < 4$ theories at tree-level.

If we extract $\mathcal{N} = 2$ coherent states of the $Q_1$ and $Q_2$ supersymmetries, then for the massive multiplet we are left with the long vector multiplet of $\mathcal{N} = 2$. This has exactly the same field content as the short $\mathcal{N} = 4$ multiplet with which we've been working. It's clear from the form of the central charge that restricting our attention to the supercharges which anticommute leaves a massive multiplet without a central charge. We can then extract tree-level superamplitudes for $\mathcal{N} = 2$ by simply extracting the $\mathcal{N} = 2$ massless submultiplets in the chiral superspace as in (2.109). One finds nonzero tree-level three-leg amplitudes of the massive $\mathcal{N} = 2$ vector (denoted as $\Omega$ here) with both the massless vector and the massless hypermultiplet, for example, through

$$\begin{aligned} \mathcal{A}_{\mathcal{N}=2}[\Omega, G^+, \overline{\Omega}] &= \frac{1}{2}\frac{\partial}{\partial\eta_2^{\dagger m}}\frac{\partial}{\partial\eta_{2m}^{\dagger}}\mathcal{A}[\mathcal{W}, G, \overline{\mathcal{W}}] \\ \mathcal{A}_{\mathcal{N}=2}[\Omega, K^+, \overline{\Omega}] &= \frac{\partial}{\partial\eta_2^{\dagger 3}}\mathcal{A}[\mathcal{W}, G, \overline{\mathcal{W}}]\big|_{\eta_2^{\dagger 4}\to 0}. \end{aligned} \tag{2.116}$$

We thus cannot truncate the spectrum by deleting the hypermultiplets, as these appear in factorization channels of higher-leg subamplitudes containing only external massless and massive vectors.

Our other option to obtain $\mathcal{N} = 2$ submultiplets is to extract coherent states of a pair of supersymmetries whose supercharges have nonzero anticommutator, for example $Q_1, Q_3$. For the massless multiplets, this puts us in the non-chiral superspace representation of (2.111). For the massive multiplets, this extracts the BPS multiplets of $\mathcal{N} = 2$,

which are simply the massive $\mathcal{N}=1$ supermultiplets (the extraction of which will be demonstrated next). The massive states may be described solely as coherent states of $Q_1$ in both $\mathcal{N}=1$ and short $\mathcal{N}=2$ cases, so the differences between subamplitudes with either BPS $\mathcal{N}=2$ submultiplets or massive $\mathcal{N}=1$ submultiplets are attributable to the massless states present (2.113).

The massive multiplets decompose into one $\mathcal{N}=1$ vector multiplet and two $\mathcal{N}=1$ chiral multiplets as

$$\mathcal{Q}_{\mathcal{N}=1} = \frac{1}{2}\frac{\partial}{\partial\eta_I^2}\frac{\partial}{\partial\eta^{2I}}\mathcal{W} \qquad \mathcal{W}^I_{\mathcal{N}=1} = \frac{\partial}{\partial\eta_I^2}\mathcal{W}\big|_{\eta_J^2\to 0} \qquad \mathcal{Q}'_{\mathcal{N}=1} = \mathcal{W}\big|_{\eta_I^2\to 0} \tag{2.117}$$

$$\overline{\mathcal{Q}}_{\mathcal{N}=1} = \overline{\mathcal{W}}\big|_{\eta_I^2\to 0} \qquad \overline{\mathcal{W}}^I_{\mathcal{N}=1} = \frac{\partial}{\partial\eta_I^2}\overline{\mathcal{W}}\big|_{\eta_J^2\to 0} \qquad \overline{\mathcal{Q}'}_{\mathcal{N}=1} = \frac{1}{2}\frac{\partial}{\partial\eta_I^2}\frac{\partial}{\partial\eta^{2I}}\overline{\mathcal{W}}. \tag{2.118}$$

The massive matter states may be alternatively grouped into massive $\mathcal{N}=2$ hypermultiplets, just as for the massless case above. To reiterate, we interpret these either as $\mathcal{N}=1$ or $\mathcal{N}=2$ submultiplets depending upon which massless states are in the amplitude, which, at this point in the discussion, is simply a collection of components of a $\mathcal{N}=4$ superamplitude. We could of course go further and extract the $\mathcal{N}=0$ components easily.

Now that we have all of the extraction operators, we may ask which tree-level amplitudes may be obtained by truncating the spectrum. While we cannot project from the Coulomb branch to an entire theory of massive $\mathcal{N}<4$ SYM, we may still be able to project onto particular amplitudes in $\mathcal{N}<4$ SYM theories. The simplest examples are the three-leg amplitudes of any minimally-coupled matter with Yang-Mills theory. The extraction of the $\mathcal{N}=1$ three-leg amplitude for two equal mass vector superfields and a

positive-helicity massless vector gives

$$\begin{aligned}\mathcal{A}[\mathcal{W}^I, \overline{\mathcal{W}}^J, G^+] &= \frac{\partial}{\partial\eta_{1I}^2}\frac{\partial}{\partial\eta_{2J}^2}\frac{1}{2}\frac{\partial}{\partial\eta_3^{\dagger m}}\frac{\partial}{\partial\eta_{3m}^{\dagger}}\mathcal{A}[\mathcal{W}, \overline{\mathcal{W}}, G]\Big|_{\eta_{1K}^2,\eta_{2K}^2,\eta_3^2\to 0} \\ &= \delta^{(2)}(Q^\dagger) g \left[1^I\right|^\alpha \left[2^J\right|^\beta \left(\frac{1}{mx}\epsilon_{\alpha\beta} - \frac{1}{m^2}\left|3\right]_\alpha \left|3\right]_\beta\right). \end{aligned} \tag{2.119}$$

By comparison with the discussion in (29), we see that at tree-level the anomalous magnetic dipole moment of the massive vector superfield has been set to zero.

We may next look for higher leg tree-level amplitudes that are not affected by the absence of truncated particles. The key is that the massive states couple in pairs to the massless states in $\mathcal{N} = 4$, so this property is inherited in each projected theory and, as above, the other massless multiplets also couple to the reduced supersymmetry massless vector multiplet in pairs. This allows us to argue, for example, that the 2 massive leg, $n-2$ massless vector superamplitudes $\mathcal{A}_n[M, \overline{M}, G^\pm, G^\pm, \ldots, G^\pm]$ (gluon helicities arbitrary) may be found via an appropriate projection, where $M$ may here be any of the massive multiplets of $\mathcal{N} < 4$. From the above, no other states may appear internally. Of course this can also be taken one step further down to $\mathcal{N} = 0$, which allows us to find the tree-level amplitudes for any number of gluons and two massive particles of any spin $\leq 1$.

Furthermore, this projection allows us to see an interesting feature for the $\mathcal{N} = 1$ all-plus-helicity amplitudes. The $\mathcal{N} = 4$ Coulomb branch amplitudes have Grassmann degree $2n$, while the extraction operators for such an $\mathcal{N} = 1$ amplitude involve $2(n-2)+2$ derivatives, so that these subamplitudes will have Grassmann degree 2. This means that the Grassmann delta function saturates the Grassmann dependence, so these tree-level superamplitudes may be entirely characterized once one component amplitude is known,

for example

$$\mathcal{A}_n[\mathcal{Q}, \overline{\mathcal{Q}}, G^+, G^+, \ldots, G^+] = \frac{-1}{m}\delta^{(2)}(Q^\dagger)A_n[\widetilde{Q}_R, \overline{\widetilde{Q}}_R, g^+, g^+, \ldots, g^+], \tag{2.120}$$

in the notation of Section 5.2 of (29). This means that we may perform massless $\mathcal{N} = 0$ BCFW recursion to find a single component amplitude and get the rest for free, rather than needing to perform the recursion in $\mathcal{N} = 4$ and then project down. In particular, we may upgrade already-known results for all-$n$ amplitudes in QCD (42; 43; 69; 70) to full $\mathcal{N} = 1$ SQCD superamplitudes.

Some simple examples of tree-level amplitudes that may be obtained by projection to $\mathcal{N} = 1$ SYM with massive vectors are

$$\mathcal{A}[\mathcal{W}^I, \overline{\mathcal{W}}^J, G^+, G^+] = \frac{\delta^{(2)}(Q^\dagger)\left\langle 1^I 2^J\right\rangle [34]^2}{(p_1+p_2)^2\left((p_2+p_3)^2+m^2\right)}, \tag{2.121}$$

$$\mathcal{A}[\mathcal{W}^I, \overline{\mathcal{W}}^J, G^+, G^-] = \frac{\delta^{(2)}(Q^\dagger)\left(\left\langle 1^I 4\right\rangle \left[2^J 3\right] + \left\langle 2^J 4\right\rangle \left[1^I 3\right]\right)\left(\left[1^K 3\right]\eta_{1K} - \left[2^L 3\right]\eta_{2L}\right)}{(p_1+p_2)^2\left((p_2+p_3)^2+m^2\right)}. \tag{2.122}$$

The above is merely an initial exploration into what the Coulomb branch can tell us about massive amplitudes in Yang-Mills theories with fewer supersymmetries.

# Bibliography


[1] L. F. Alday and J. M. Maldacena, *Gluon scattering amplitudes at strong coupling*, *JHEP* **06** (2007) 064, [arXiv:0705.0303].

[2] J. M. Drummond, G. P. Korchemsky, and E. Sokatchev, *Conformal properties of four-gluon planar amplitudes and Wilson loops*, *Nucl. Phys.* **B795** (2008) 385–408, [arXiv:0707.0243].

[3] J. M. Henn, *Duality between Wilson loops and gluon amplitudes*, *Fortsch. Phys.* **57** (2009) 729–822, [arXiv:0903.0522].

[4] S. Caron-Huot, *Notes on the scattering amplitude / Wilson loop duality*, *JHEP* **07** (2011) 058, [arXiv:1010.1167].

[5] L. J. Mason and D. Skinner, *The Complete Planar S-matrix of N=4 SYM as a Wilson Loop in Twistor Space*, *JHEP* **12** (2010) 018, [arXiv:1009.2225].

[6] T. Adamo, M. Bullimore, L. Mason, and D. Skinner, *A Proof of the Supersymmetric Correlation Function / Wilson Loop Correspondence*, *JHEP* **08** (2011) 076, [arXiv:1103.4119].

[7] J. M. Drummond, J. Henn, G. P. Korchemsky, and E. Sokatchev, *Dual superconformal symmetry of scattering amplitudes in N=4 super-Yang-Mills theory*, *Nucl. Phys.* **B828** (2010) 317–374, [arXiv:0807.1095].

[8] N. Berkovits and J. Maldacena, *Fermionic T-Duality, Dual Superconformal Symmetry, and the Amplitude/Wilson Loop Connection*, *JHEP* **09** (2008) 062, [arXiv:0807.3196].

[9] J. M. Drummond, J. M. Henn, and J. Plefka, *Yangian symmetry of scattering amplitudes in N=4 super Yang-Mills theory*, *JHEP* **05** (2009) 046, [arXiv:0902.2987].

[10] J. M. Drummond and J. M. Henn, *All tree-level amplitudes in N=4 SYM*, *JHEP* **04** (2009) 018, [arXiv:0808.2475].

[11] R. Britto, F. Cachazo, and B. Feng, *New recursion relations for tree amplitudes of gluons*, *Nucl. Phys.* **B715** (2005) 499–522, [hep-th/0412308].

[12] R. Britto, F. Cachazo, B. Feng, and E. Witten, *Direct proof of tree-level recursion relation in Yang-Mills theory*, *Phys. Rev. Lett.* **94** (2005) 181602, [hep-th/0501052].

[13] N. Arkani-Hamed, J. L. Bourjaily, F. Cachazo, and J. Trnka, *Local Integrals for Planar Scattering Amplitudes*, *JHEP* **06** (2012) 125, [arXiv:1012.6032].

[14] N. Arkani-Hamed, J. L. Bourjaily, F. Cachazo, S. Caron-Huot, and J. Trnka, *The All-Loop Integrand For Scattering Amplitudes in Planar N=4 SYM*, *JHEP* **01** (2011) 041, [arXiv:1008.2958].

[15] A. Hodges, *Eliminating spurious poles from gauge-theoretic amplitudes*, *JHEP* **05** (2013) 135, [arXiv:0905.1473].

[16] N. Arkani-Hamed and J. Trnka, *The Amplituhedron*, *JHEP* **10** (2014) 030, [arXiv:1312.2007].

[17] R. H. Boels, *No triangles on the moduli space of maximally supersymmetric gauge theory*, *JHEP* **05** (2010) 046, [arXiv:1003.2989].

[18] N. Craig, H. Elvang, M. Kiermaier, and T. Slatyer, *Massive amplitudes on the Coulomb branch of N=4 SYM*, *JHEP* **12** (2011) 097, [arXiv:1104.2050].

[19] M. Kiermaier, *The Coulomb-branch S-matrix from massless amplitudes*, arXiv:1105.5385.

[20] H. Elvang, D. Z. Freedman, and M. Kiermaier, *Integrands for QCD rational terms and N=4 SYM from massive CSW rules*, *JHEP* **06** (2012) 015, [arXiv:1111.0635].

[21] T. Dennen, Y.-t. Huang, and W. Siegel, *Supertwistor space for 6D maximal super Yang-Mills*, *JHEP* **04** (2010) 127, [arXiv:0910.2688].

[22] T. Dennen and Y.-t. Huang, *Dual Conformal Properties of Six-Dimensional Maximal Super Yang-Mills Amplitudes*, *JHEP* **01** (2011) 140, [arXiv:1010.5874].

[23] Z. Bern, J. J. Carrasco, T. Dennen, Y.-t. Huang, and H. Ita, *Generalized Unitarity and Six-Dimensional Helicity*, *Phys. Rev.* **D83** (2011) 085022, [arXiv:1010.0494].

[24] Y.-t. Huang, *Non-Chiral S-Matrix of N=4 Super Yang-Mills*, arXiv:1104.2021.

[25] J. Plefka, T. Schuster, and V. Verschinin, *From Six to Four and More: Massless and Massive Maximal Super Yang-Mills Amplitudes in 6d and 4d and their Hidden Symmetries*, *JHEP* **01** (2015) 098, [arXiv:1405.7248].

[26] F. Cachazo, S. He, and E. Y. Yuan, *Scattering equations and Kawai-Lewellen-Tye orthogonality*, *Phys. Rev.* **D90** (2014), no. 6 065001, [arXiv:1306.6575].

[27] F. Cachazo, A. Guevara, M. Heydeman, S. Mizera, J. H. Schwarz, and C. Wen, *The S Matrix of 6D Super Yang–Mills and Maximal Supergravity from Rational Maps*, arXiv:1805.11111.

[28] Y. Geyer and L. Mason, *The polarized scattering equations for 6d superamplitudes*, arXiv:1812.05548.

[29] A. Herderschee, S. Koren, and T. Trott, *Constructing $\mathcal{N} = 4$ Coulomb Branch Superamplitudes*, arXiv:1902.07205.

[30] N. Arkani-Hamed, T.-C. Huang, and Y.-t. Huang, *Scattering Amplitudes For All Masses and Spins*, arXiv:1709.04891.

[31] H. Elvang and Y. Huang, *Scattering Amplitudes in Gauge Theory and Gravity*. Cambridge University Press, 2015, arXiv:1308.1697.

[32] M. Heydeman, J. H. Schwarz, and C. Wen, *M5-Brane and D-Brane Scattering Amplitudes*, *JHEP* **12** (2017) 003, [arXiv:1710.02170].

[33] S. Caron-Huot and Z. Zahraee, *Integrability of Black Hole Orbits in Maximal Supergravity*, arXiv:1810.04694.

[34] M. Gunaydin and R. J. Scalise, *Unitary Lowest Weight Representations of the Noncompact Supergroup Osp(2m*/2n)*, *J. Math. Phys.* **32** (1991) 599–606.

[35] P. Fayet, *Spontaneous Generation of Massive Multiplets and Central Charges in Extended Supersymmetric Theories*, *Nucl. Phys.* **B149** (1979) 137.

[36] H. Osborn, *Topological Charges for N=4 Supersymmetric Gauge Theories and Monopoles of Spin 1*, *Phys. Lett.* **83B** (1979) 321–326.

[37] C. Fraser and T. J. Hollowood, *Semiclassical quantization in N=4 supersymmetric Yang-Mills theory and duality*, *Phys. Lett.* **B402** (1997) 106–112, [hep-th/9704011].

[38] M. L. Mangano and S. J. Parke, *Multiparton amplitudes in gauge theories*, *Phys. Rept.* **200** (1991) 301–367, [hep-th/0509223].

[39] A. Brandhuber, P. Heslop, and G. Travaglini, *A Note on dual superconformal symmetry of the N=4 super Yang-Mills S-matrix*, *Phys. Rev.* **D78** (2008) 125005, [arXiv:0807.4097].

[40] N. Arkani-Hamed, F. Cachazo, and J. Kaplan, *What is the Simplest Quantum Field Theory?*, *JHEP* **09** (2010) 016, [arXiv:0808.1446].

[41] C. Cheung, *On-Shell Recursion Relations for Generic Theories*, *JHEP* **03** (2010) 098, [arXiv:0808.0504].

[42] S. D. Badger, E. W. N. Glover, V. V. Khoze, and P. Svrcek, *Recursion relations for gauge theory amplitudes with massive particles*, *JHEP* **07** (2005) 025, [hep-th/0504159].

[43] C. Schwinn and S. Weinzierl, *On-shell recursion relations for all Born QCD amplitudes*, *JHEP* **04** (2007) 072, [hep-ph/0703021].

[44] C. Cheung and D. O'Connell, *Amplitudes and Spinor-Helicity in Six Dimensions*, *JHEP* **07** (2009) 075, [arXiv:0902.0981].

[45] S. Caron-Huot and D. O'Connell, *Spinor Helicity and Dual Conformal Symmetry in Ten Dimensions*, *JHEP* **08** (2011) 014, [arXiv:1010.5487].

[46] J. Scherk and J. H. Schwarz, *Spontaneous Breaking of Supersymmetry Through Dimensional Reduction*, *Phys. Lett.* **82B** (1979) 60–64.

[47] Z. Bern, M. Czakon, L. J. Dixon, D. A. Kosower, and V. A. Smirnov, *The Four-Loop Planar Amplitude and Cusp Anomalous Dimension in Maximally Supersymmetric Yang-Mills Theory*, *Phys. Rev.* **D75** (2007) 085010, [hep-th/0610248].

[48] J. M. Drummond, J. Henn, V. A. Smirnov, and E. Sokatchev, *Magic identities for conformal four-point integrals*, *JHEP* **01** (2007) 064, [hep-th/0607160].

[49] N. Beisert et al., *Review of AdS/CFT Integrability: An Overview*, *Lett. Math. Phys.* **99** (2012) 3–32, [arXiv:1012.3982].

[50] J. M. Drummond, J. Henn, G. P. Korchemsky, and E. Sokatchev, *Conformal Ward identities for Wilson loops and a test of the duality with gluon amplitudes*, *Nucl. Phys.* **B826** (2010) 337–364, [arXiv:0712.1223].

[51] Z. Bern, L. J. Dixon, and V. A. Smirnov, *Iteration of planar amplitudes in maximally supersymmetric Yang-Mills theory at three loops and beyond*, *Phys. Rev.* **D72** (2005) 085001, [hep-th/0505205].

[52] G. P. Korchemsky and E. Sokatchev, *Symmetries and analytic properties of scattering amplitudes in N=4 SYM theory*, *Nucl. Phys.* **B832** (2010) 1–51, [arXiv:0906.1737].

[53] T. Bargheer, N. Beisert, W. Galleas, F. Loebbert, and T. McLoughlin, *Exacting N=4 Superconformal Symmetry*, *JHEP* **11** (2009) 056, [arXiv:0905.3738].

[54] A. Sever and P. Vieira, *Symmetries of the N=4 SYM S-matrix*, arXiv:0908.2437.

[55] S. Caron-Huot and S. He, *Jumpstarting the All-Loop S-Matrix of Planar N=4 Super Yang-Mills*, *JHEP* **07** (2012) 174, [arXiv:1112.1060].

[56] N. Kanning and M. Staudacher, *Graßmannian Integrals in Minkowski Signature, Amplitudes, and Integrability*, arXiv:1811.04949.

[57] L. F. Alday, J. M. Henn, J. Plefka, and T. Schuster, *Scattering into the fifth dimension of N=4 super Yang-Mills*, *JHEP* **01** (2010) 077, [arXiv:0908.0684].

[58] J. M. Henn, S. G. Naculich, H. J. Schnitzer, and M. Spradlin, *Higgs-regularized three-loop four-gluon amplitude in N=4 SYM: exponentiation and Regge limits*, *JHEP* **04** (2010) 038, [arXiv:1001.1358].

[59] J. M. Henn, S. G. Naculich, H. J. Schnitzer, and M. Spradlin, *More loops and legs in Higgs-regulated N=4 SYM amplitudes*, *JHEP* **08** (2010) 002, [arXiv:1004.5381].

[60] S. Caron-Huot and J. M. Henn, *Solvable Relativistic Hydrogenlike System in Supersymmetric Yang-Mills Theory*, *Phys. Rev. Lett.* **113** (2014), no. 16 161601, [arXiv:1408.0296].

[61] N. Arkani-Hamed, J. L. Bourjaily, F. Cachazo, A. B. Goncharov, A. Postnikov, and J. Trnka, *Grassmannian Geometry of Scattering Amplitudes*. Cambridge University Press, 2016.

[62] K. Bering and M. Pazderka, *6D dual superconformal algebra*, arXiv:1810.12674.

[63] R. H. Boels and D. O'Connell, *Simple superamplitudes in higher dimensions*, *JHEP* **06** (2012) 163, [arXiv:1201.2653].

[64] P. Benincasa and F. Cachazo, *Consistency Conditions on the S-Matrix of Massless Particles*, arXiv:0705.4305.

[65] M. Srednicki, *Quantum field theory*. Cambridge University Press, 2007.

[66] H. Elvang, D. Z. Freedman, and M. Kiermaier, *Solution to the Ward Identities for Superamplitudes*, *JHEP* **10** (2010) 103, [arXiv:0911.3169].

[67] H. Elvang, Y.-t. Huang, and C. Peng, *On-shell superamplitudes in N¡4 SYM*, *JHEP* **09** (2011) 031, [arXiv:1102.4843].

[68] L. J. Dixon, J. M. Henn, J. Plefka, and T. Schuster, *All tree-level amplitudes in massless QCD*, *JHEP* **01** (2011) 035, [arXiv:1010.3991].

[69] C. Schwinn and S. Weinzierl, *SUSY ward identities for multi-gluon helicity amplitudes with massive quarks*, *JHEP* **03** (2006) 030, [hep-th/0602012].

[70] A. Ochirov, *Helicity amplitudes for QCD with massive quarks*, *JHEP* **04** (2018) 089, [arXiv:1802.06730].

# Chapter 3

# Causality, Unitarity and Symmetry in Effective Field Theory

Sum rules in effective field theories, predicated upon causality, place restrictions on scattering amplitudes mediated by effective contact interactions. Through unitarity of the $S$-matrix, these imply that the size of higher dimensional corrections to transition amplitudes between different states is bounded by the strength of their contributions to elastic forward scattering processes. This places fundamental limits on the extent to which hypothetical symmetries can be broken by effective interactions. All analysis is for dimension 8 operators in the forward limit. Included is a thorough derivation of all positivity bounds for a chiral fermion in $SU(2)$ and $SU(3)$ global symmetry representations resembling those of the Standard Model, general bounds on flavour violation, new bounds for interactions between particles of different spin, inclusion of loops of dimension 6 operators and illustration of the resulting strengthening of positivity bounds over tree-level expectations, a catalogue of supersymmetric effective interactions up to mass dimension 8 and 4 legs and the demonstration that supersymmetry unifies the positivity theorems as well as the new bounds.

## 3.1 Introduction

Effective field theory (EFT) is a method to mock-up the long distance effects of high energy states with simplified and fake microphysics that nevertheless successfully approximates departures from otherwise universal behaviour. This is predicated on the approximate locality of the interactions involving heavy states (see e.g. (1) for review). The heavier the state, the shorter its range. If the scales over which these fields can propagate is too small to be resolved, then the effects can be instead approximated by a series of local contact interactions. These are usually restricted by the symmetries, from which the possible interactions can be systematically identified and used to parameterise the impacts of the microphysics on long distance observables without knowing what it actually is. This provides a general strategy for accounting for the effects of unknown short-distance effects and identifying a classification scheme for the possible interactions that low energy particles may be involved with.

However, demanding local contact interactions alone is not sufficient for consistency with causality (see e.g. (2)). The couplings parameterising the strength and phase of these interactions are restricted so that they cannot conspire to mediate macroscopic superluminal signal transmission. The focus of this work will be on exploring these consistency constraints on the space of effective interactions. The UV completion of these interactions will be assumed to be a conventional quantum field theory, obeying micro-causality (analyticity of the $S$-matrix) and polynomially bounded energy dependence. For discussions of the relevance of this to various ideas of theories where these conditions could be modified (usually in relation to quantum gravity), see e.g. (3), (4), (5), (6), (7). For applications of causality constraints to CFTs and holography, see e.g. (8), (9), (10), (11), (12), (13), (14), while for application to EFTs, the subject of discussion here, see (2), (15), (16), (17), (18), (19), (20), (21), (22) for a sample of past studies (among many

others).

Unitarity in quantum mechanics is the statement of the consistency of time evolution with the probability interpretation of quantum mechanics - that is, the square magnitude of a transition amplitude between different states at different times have the interpretation of a probability. Of interest here will be asymptotic scattering states. Unitarity of the $S$-matrix is expressed through the optical theorem and the Cutting rules. For (near) forward scattering, this relates the residues or discontinuities over the singularities of the $S$-matrix to on-shell particle production. Together with its analytic causal properties, this enables the construction of dispersion relations that constrain the $S$-matrix for general (complex) momentum. For low energy scattering states, the $S$-matrix is calculable from an EFT. For high enough order interactions, the dispersion relations form a sum rule that determines the EFT couplings entirely from the (usually unknown) on-shell production rates of states in the UV. The (by now standard) dispersion relation between IR EFT-calculable processes and high energy production rates is reviewed below in Section 3.2, but see again (2) for more background.

For scattering processes in which the identity of the particles do not change, the sum rule implies that the IR contact interactions are equated with a positive sum of production rates in the UV. Because this is necessarily a positive number, the resulting constraints on the corresponding low energy Wilson coefficients have been called "positivity theorems". However, unitarity implies that the dispersion relation contains more information than simply positivity.

As will be shown in Section 3.2.1, "inelastic processes" (processes in which the identity of the particles change) are bounded above by elastic ones according to the general

constraint:

$$|M^{ijkl}| + |M^{ilkj}| \leq \sqrt{M^{ijij}M^{klkl}} + \sqrt{M^{ilil}M^{kjkj}}. \tag{3.1}$$

Here, $M^{ijkl} = \frac{d^2}{ds^2}A^{ijkl}(s)\Big|_{s=0}$, where $A^{ijkl}(s)$ is the forward amplitude describing the scattering process $i, j \to k, l$, where $i, j, k, l$ are any species of particle. Consequently, the extent to which inelastic effective interactions can violate hypothetical symmetries is limited.

The term "inelastic" will be consistently (mis)used here to broadly refer to transitions in which the identity of the scattered particles change, rather than simply their masses. Here "identity" will usually reference quantum numbers with respect to a complete set of commuting observables, although this is, of course, a basis-dependent statement. In particular, (massless) spinning particles will usually be identified by helicity eigenstates (I will usually also misuse the word "helicity" to mean only the magnitude).

The remainder of the paper focuses both on applications of this result and its consequences, as well as general exploration of the structure of the causality constraints. This is mostly with an eye toward the Standard Model Effective Theory (SMEFT) (23), a general EFT parameterisation of the imprints of new, high energy microphysics on the Standard Model of particle physics (SM). The scope of these bounds are theoretically more far-reaching than the positivity constraints that have been largely the focus of previous attention. For example, elastic amplitudes that vanish in the forward limit can often be crossed into inelastic amplitudes that do not, thus failing to escape from the sum rule. See (24), (25), (26), (27), (28) for previous applications of causality bounds to the SMEFT.

### 3.1.1 Overview of results

Section 3.2.1 reviews the standard derivation of the dispersion relation between the twice differentiated forward scattering amplitude and the transition rates into on-shell states in the UV. This will be focused on forward scattering at dimension 8 level. It is then shown that unitarity implies the general constraint (3.1) on inelastic processes. Bounds of this form have been previously identified in (28) for the specific case of parity-symmetric weak boson interactions, although, to the author's knowledge, the general statement above is new. Section 3.2.2 gives some general discussion about the inclusion of loops of lower dimensional operators in the dimension 8 order contribution to the amplitude. In particular, unitarity implies the general expectation that the elastic dim-8 Wilson coefficients decrease with energy scale, which would therefore strengthen the positivity bounds that would otherwise be inferred at tree-level. The general results are illustrated in Section 3.3 by a simple example: a complex scalar field. It is shown that processes that would violate a (hypothetical) $U(1)$ charge must necessarily be bounded above by charge conserving processes.

In Section 3.4, causality constraints for EFTs with preserved internal symmetries are examined. Following (29), the causality sum rule can be equivalently characterised as a convex cone in which UV completable IR amplitudes must lie. In simple examples where there are no degenerate states unrelated by symmetry, the cone is polyhedral and elementary convex geometry allows for a complete set of positivity bounds to be extracted, including those inaccessible from scattering of factorised states. Section 3.2.3 reviews the convex geometry interpretation of (29) (similar ideas were suggested in (15)), for which much of this work was inspired at understanding. Using this picture, I compute the complete set of bounds for a single fermion species in the symmetry representations of the SM (fundamentals of $SU(2)$ and $SU(3)$) in Section 3.4.2, possibly also including exact

flavour symmetry. Some of these bounds are altogether new. It is also illustrated how, with increasingly more symmetry representations, the structure of the positivity bounds becomes increasingly intricate - the convex cone describing the space of allowed amplitudes becomes increasingly multifaceted. In Section 3.4.3, I give the general bounds on flavour violation admissible under the general bound on inelastic amplitudes, elaborating upon the observations of limits on flavour violation made by (25). The results discussed in this section have direct application to the SMEFT.

Section 3.5 discusses rotational symmetry and the treatment of spin and parity ($P$). Spin is discussed in generality in Section 3.5.1, which is then illustrated with the simple (and known) example of four photon scattering in Section 3.5.2. It is in particular shown how $P$ and helicity violation (electric-magnetic duality) are necessarily limited, explaining the observations made by (24) for vector boson scattering. The results are then generalised to two distinct species with the same helicity in Section 3.5.4.

Constraints on EFTs with multiple particles of different helicities will be derived in Section 3.5.5. These add to bounds derived previously for elastic processes involving states of various spin in (15), (24), (30). In Section 3.6, I catalogue all possible supersymmetric effective contact interactions with mass dimension up to eight and at most four particles, identifying which types of interactions are embeddable into supersymmetric EFTs. This culminates in the demonstration that the standard simple positivity theorems of (2), (16) and the new additions derived here unify under supersymmetry. Less supersymmetric inelastic interactions appear as amplitudes that must necessarily raise the lower bounds on the more supersymmetric elastic operators. In this sense supersymmetry must at least partially emerge from the positivity bounds.

An appendix gives a list of elementary results for projectors for spin indices represented in $SO(2)$ form, as well as a presentation of the sum rules for the toy two fermion and multispin theories.

Note added: As this work was being completed, the work of (20) appeared, which involves partial overlap with the general discussion of loops provided here. In particular, the general observation made here about strengthening of the constraints with RG flow from dim 6 loops is another instance of the result for Goldstone bosons described in (20), albeit applied in different examples.

More notes added: Subsequent to release of this preprint, the following relevant works appeared: (31) make some pertinent and interesting comments on positivity constraints with loops, problems with the forward limit and bounds on multiparticle theories, in addition to its main thesis of deriving new constraints on higher dimension operators away from the forward limit. (32) incorporates supersymmetry in constraints of the form of (31) on Yang-Mills operators, although their discussion has little overlap with that presented here. In (33), positivity bounds were derived for three flavours of SM quarks under the assumption of minimal flavour violation but only from consideration of scattering of pure states. Finally, (34) makes advances on the issue of constraining theories with multiple species in which the space of couplings is a non-polyhedral cone. This includes some more thorough derivations of bounds for theories with more than two distinct species unrelated by symmetries that, in some instances, improves over bounds presented here (such as on the flavour-violating operators of right-handed electrons).

## 3.2 Sum Rule and Unitarity

This section reviews the derivation of the standard dispersion relation between the IR scattering amplitude in an EFT and the full transition rates in the UV, along with the various caveats and assumptions that will be implicit throughout the remainder of the paper. This derivation can be found in numerous previous works e.g. (2), (15). The discussion presented below is closest to (28), from which much of this thinking is inspired.

The inelastic scattering constraint (3.1) will then be derived. The affects of loops will also be discussed.

The subsequent discussion will then turn to the process of extracting constraints on the EFT out of the sum rules. To do this, I will use the picture presented in (29) and (28) of the space of UV completions as delineating a convex cone to which the space of all forward effective transition amplitudes in the IR must belong.

All discussion in this paper will specifically focus on constraints at mass dimension 8 order in standard power counting arising from scattering in the forward limit.

### 3.2.1 Causality sum rule and structure

The standard assumptions of analyticity of the scattering amplitude as a function of energy will be made. From the point of view of QFT, this is an expected consequence of microcausality (35), (36), (37). See e.g. (38) for review. This has been somewhat established for correlators in massive theories and those obeying the Wightman axioms, like CFTs. If the theory has particles, these presumably extend to the $S$-matrix. I will assume that these results also hold for the simple massless theories described here by assuming that a massless limit can be taken that commutes with the ensuing derivations. Similarly, it will be assumed that the only singularities of the two-to-two $S$-matrix in the forward limit are poles and branch cuts close to the real energy axis directly associated with on-shell particle production in the $s$ or $u$-channels.

Call $A^{ijkl}(s)$ the forward scattering amplitude ($t = 0$) for the process $i, j \to k, l$. Then, for some complex-valued energy $\sigma$, define $\overline{M}^{ijkl}(\sigma^2) = \frac{d^2}{ds^2} A^{ijkl}(s)|_{s=\sigma^2} = \frac{1}{\pi i} \oint \frac{A^{ijkl}(s)}{(s-\sigma^2)^3} ds$. The contour is over a small loop enclosing $\sigma^2$ and no other singularities. The small loop may then be deformed in the standard way into a contour enclosing and railing along the poles and branch cuts on the real axis closed-off by a semi-circular arc of some large

enough radius. The Froissart bound (39) implies that the integrand decays fast enough along the arc that this part of the integral may be ignored. The supplemental locality assumption of polynomially bounded energy growth is invoked here (see e.g. (5), (4), (6), (7) for commentary about the necessity of this assumption and possible consequences of its modification, usually in the context of gravity).

Crossing can be used to relate the forward amplitude along the negative real axis to the $u$-channel amplitude $A^{ijkl}(s) = A^{i\bar{l}k\bar{j}}(4m^2 - s)$ (where $\bar{X}$ denotes the antiparticle of $X$). I will assume for simplicity that each particle has mass $m$, although this will be ignored for most of what follows (assumed to be small compared to the characteristic energies of the observed scattering processes). Crossing relations with spinning particles have been explained in (16) for the forward limit specifically for the present context. Note that the amplitude $A(X, Y \to Z, W)$ is obtained through LSZ reduction from a correlator with fields ordered as $\langle 0|WZXY|0\rangle$, which will be of relevance in Section 3.6. The remaining contour integral over the branch cuts is

$$\overline{M}^{ijkl}(\sigma^2) = \frac{1}{\pi i}\int_{4m^2}^{\infty}\left(\frac{\mathrm{Disc}A^{ijkl}(s)}{(s-\sigma^2)^3} + \frac{\mathrm{Disc}A^{i\bar{l}k\bar{j}}(s)}{(s+\sigma^2-4m^2)^3}\right)ds + \text{residues at poles.} \quad (3.2)$$

The amplitudes obey the reality condition following from analyticity and unitarity (40)

$$\left(A^{ijkl}(s)\right)^* = A^{klij}(s^*). \quad (3.3)$$

This is the $S$-matrix statement of requiring the Hamiltonian to be Hermitian. This relation expresses the fact that the discontinuity over the branch cut is related to the

intermediate state on-shell production rate through unitarity of the $S$-matrix:

$$\text{Disc}A^{ijkl}(s) = A^{ijkl}(s+i\epsilon) - \left(A^{klij}(s+i\epsilon)\right)^* = i\sum_X \mathcal{M}^{ij\to X}(s+i\epsilon)\left(\mathcal{M}^{kl\to X}(s+i\epsilon)\right)^*, \tag{3.4}$$

where $\mathcal{M}^{ab\to X}$ is the amplitude for particles $a$ and $b$ to transition into intermediate state $X$ and $\epsilon \to 0^+$ is implicit.

Mostly for simplicity, the theories discussed here will all be massless. While the results above have been derived under the assumption of a mass gap, I will disregard this and assume that the particle masses in the sum rule can be freely taken to zero without consequence at this stage. In particular, this will assume that the singularity structure of the $S$-matrix in the forward limit is not affected. See (16) for a list of some other possible issues. The validity of this remains an open question for investigation and becomes increasingly less certain with increasing spin. These questions have received particular recent attention in the context of gravity, see e.g. (41), (42), (43), although I am satisfied here with restricting to flat spacetime QFT with particle helicities $\leq 1$.

For massive theories, it is natural to make a real insertion $\sigma^2$ below the mass threshold where the amplitude is analytic. This would represent a region of energies in which RG evolution switches off and the energy-dependence of the amplitude is relatively simple. For massless theories, the branch cuts cleave the entire complex $s$ space. It will be assumed that the dispersion relation for these theories can be reached by taking a mass-deformed theory, analytically continuing the insertion point $\sigma^2$ above a branch cut to some $\sigma^2 + i\delta$ with $\delta \to 0^+$ (keeping $\sigma^2$ real) and then taking the massless limit (so that the cuts extend to the origin). Note that this procedure does not require the masses to be sent to zero exactly, but only that they be much smaller than the insertion point. The insertion point itself can then be taken small in the IR for simplicity: $\sigma^2/\Lambda^2 \to 0$ for UV

cut-off $\Lambda$. One advantage of doing this is that the sum rule becomes symmetric in $s$ and $u$-channel cuts.

So taking the massless limit and invoking unitarity, the sum rule becomes

$$\overline{M}^{ijkl}(\sigma^2) = \frac{1}{\pi}\int_0^\infty \frac{1}{(s-\sigma^2-i\delta)^3}\sum_X \mathcal{M}^{ij\to X}\left(\mathcal{M}^{kl\to X}\right)^* ds + \frac{1}{\pi}\int_0^\infty \frac{1}{(s+\sigma^2+i\delta)^3}\sum_X \mathcal{M}^{i\bar{l}\to X}\left(\mathcal{M}^{k\bar{j}\to X}\right)^* ds. \tag{3.5}$$

The limit $\delta \to 0^+$ is implicit. The insertion point $\sigma \ll \Lambda$ is chosen to be a characteristic IR energy scale so that the LHS can be evaluated in the EFT. The sum is over all possible intermediate state $X$, which may be infinite and continuous. There may also be poles on the real axis - these will be implicitly included in the integral over the cuts. In the IR, these may be explicitly calculated anyway.

The integral on the RHS of (3.5) is over both the known IR and the unknown UV. As a relation between the IR and the UV, the calculable IR part of the integral really belongs on the LHS and contains significant information. Defining this new combined left-hand side by $M^{ijkl}(\sigma^2)$, the sum rule therefore becomes

$$\begin{aligned} M^{ijkl}(\sigma^2) &= \overline{M}^{ijkl}(\sigma^2) - \frac{1}{\pi}\int_0^{\lambda^2}\Big(\frac{1}{(s-\sigma^2-i\delta)^3}\sum_{X\in IR}\mathcal{M}^{ij\to X}\left(\mathcal{M}^{kl\to X}\right)^* \\ &\qquad + \frac{1}{(s+\sigma^2)^3}\sum_{X\in IR}\mathcal{M}^{i\bar{l}\to X}\left(\mathcal{M}^{k\bar{j}\to X}\right)^*\Big)ds \\ &= \frac{1}{\pi}\int_{\lambda^2}^\infty\Big(\frac{1}{(s-\sigma^2-i\delta)^3}\sum_{X\in UV}\mathcal{M}^{ij\to X}\left(\mathcal{M}^{kl\to X}\right)^* \\ &\qquad + \frac{1}{(s+\sigma^2)^3}\sum_{X\in UV}\mathcal{M}^{i\bar{l}\to X}\left(\mathcal{M}^{k\bar{j}\to X}\right)^*\Big)ds. \end{aligned} \tag{3.6}$$

Here $\lambda$ is some high energy scale up to which the EFT is still reliable, which may be taken up to the cut-off $\Lambda$. If loops can be ignored in some approximation, then the IR integral

over the branch cut can be ignored and there is no problem with choosing $\sigma \approx 0$, where RG evolution (by assumption) has ceased. Otherwise the amplitudes are to be evaluated above mass thresholds, as would be necessary for (approximately) massless particles, and $\sigma$ is chosen to lie over a branch cut. The IR integral is non-trivial and can be computed to the required level of accuracy in the energy expansion underpinning the EFT. Because of the pole in the integrand, if $\sigma$ is to be identified with a real energy scale, then a non-zero $\delta$ is required that must be sent to 0 (it has been assumed that $\sigma^2$ is positive and lies above the $s$-channel cut in (3.6)). When evaluating the dispersion integral along the branch cut, this leaves behind a finite imaginary part that cancels against the imaginary part in the loop amplitude $\overline{M}^{ijkl}(\sigma^2)$. More generally, the loop amplitudes also include logarithmic RG-evolution from the renormalisation scale to the insertion point $\sigma$, while the dispersion integral accounts for further evolution from $\sigma$ to the scale up to which the integral is being evaluated.

Taking the insertion point to the origin $\sigma^2 \to 0$, the dispersion relation becomes

$$M^{ijkl}(0) = \frac{1}{\pi}\int_{\lambda^2}^{\infty} \frac{1}{s^3} \sum_X \left( \mathcal{M}^{ij\to X} \left(\mathcal{M}^{kl\to X}\right)^* + \mathcal{M}^{i\bar{l}\to X} \left(\mathcal{M}^{k\bar{j}\to X}\right)^* \right) ds. \tag{3.7}$$

The amplitudes $\mathcal{M}^{ij\to X}(s)$ are vectors in a complex inner product space, where both the energy $s$ and the couplings to each intermediate state $X$ in the UV completion are the (infinite and continuous) components. To emphasize this, I will rewrite this full complex vector as $\boldsymbol{m}^{ij}$. The sum over states $X$ and the dispersion integral define an inner product in these variables (the accompanying multiplicative factors in the integrand are positive, so enable this interpretation). The sum rule can then be expressed as

$$M^{ijkl} = \boldsymbol{m}^{kl} \cdot \boldsymbol{m}^{ij} + \boldsymbol{m}^{k\bar{j}} \cdot \boldsymbol{m}^{i\bar{l}}, \tag{3.8}$$

omitting the specification that $\sigma = 0$ from the notation for convenience. The sum rule (3.8) is the centerpiece of this work. All references to "the sum rule" refer to this equation, while the terms "LHS" and "RHS" will be used to refer to the left-hand side and right-hand side of this equation without qualification throughout.

Note that it is not essential to evaluate the sum rule with $\sigma = 0$ exactly. In this case, the $s$ and $u$-channel terms in (3.8) do not have identical coefficients, but differ only by subleading factors in $\sigma^2/\lambda^2$. Taking $\lambda$ close to $\Lambda$ and $\sigma \ll \Lambda$, these terms are already consistent with the truncation error of the low energy expansion.

The space of couplings of the states in the UV to those in the EFT are parameterised by the vectors $\boldsymbol{m}^{ij}$. Organised in this way, it is possible to draw many immediate conclusions from the sum rule about the EFT directly from the combination of vectors in this expression. This will be illustrated in numerous examples below. The traditional positivity theorems for elastic scattering following from the optical theorem are obvious from (3.8) when $k = i$ and $l = j$, as each term is the norm of a complex vector, which must be positive if non-zero. The elastic amplitudes all have the form

$$M^{ijij} = |\boldsymbol{m}^{ij}|^2 + |\boldsymbol{m}^{i\bar{j}}|^2. \tag{3.9}$$

However, inelastic processes are necessarily bounded from above by elastic processes as well. The Schwarz and triangle inequalities give upper bounds on the inelastic amplitudes:

$$|M^{ijkl}| = \left|\boldsymbol{m}^{kl} \cdot \boldsymbol{m}^{ij} + \boldsymbol{m}^{k\bar{j}} \cdot \boldsymbol{m}^{i\bar{l}}\right| \leq |\boldsymbol{m}^{kl}||\boldsymbol{m}^{ij}| + |\boldsymbol{m}^{k\bar{j}}||\boldsymbol{m}^{i\bar{l}}|. \tag{3.10}$$

A general upper bound can then be obtained as

$$|M^{ijkl}| + |M^{ilkj}| \leq \sqrt{M^{ijij}M^{klkl}} + \sqrt{M^{ilil}M^{kjkj}}. \tag{3.11}$$

This demonstrates the schematic pattern and is an entirely general result. When there are symmetries relating the states, a subset of the vectors $\boldsymbol{m}^{ij}$ are related and there are fewer independent vectors of UV couplings. In this case, stronger bounds may be possible after the states are classified into symmetry irreps, as will be discussed in Section 3.4. In general, the bounds of the form (3.11) are necessary but not sufficient, but are being highlighted here both because they are simple and that they directly demonstrate the way in which unitarity fundamentally limits the size of inelastic transitions. Improvements remain an open problem, such as those explored in (28).

Recently, (29), (28) offered the interpretation of the space of points of the form $\{\boldsymbol{m}^{kl} \cdot \boldsymbol{m}^{ij} + \boldsymbol{m}^{k\bar{j}} \cdot \boldsymbol{m}^{i\bar{l}}\}$ as a convex cone - a convex hull generated by positive linear combinations of a subset of vectors. Note that it is not necessary that the vectors themselves be real-valued, nor that the states be self-conjugate. See (29), (28) for further details. Convex cones can be described in two equivalent ways: by a set of inequalities delineating hyperplanes (or facets) that bound the cone, or by a set of extremal rays (ERs) that determine the edges of the cone. Extremal rays are 1d subspaces of single vectors that cannot be decomposed into a positive linear combination of any other set of linearly independent vectors in the cone. I will use the term ER ambiguously to mean either the subspace or a member vector. Any point in the cone can be expressed as a linear combination of extremal rays with positive coefficients, so the ERs generate the cone.

The inequality representation is a manifest statement of the constraints on the space of forward amplitudes or, equivalently, the bounds on the space of Wilson coefficients allowed in the EFT. The problem at hand is to extract from the sum rule a complete set of such bounds. The ER representation provides an intermediate alternative that is straight-forward to determine directly from the sum rule. If the cone is polyhedral, as expected for theories in which all transitions between states are rigidly fixed by symmetries,

then standard results from convex geometry may be applied to derive the inequalities describing the facets. This insight was applied to some simple examples by (29) to derive new constraints through rudimentary convex geometry that were inaccessible from considering only scattering amplitudes of factorised states. This will be applied to some more examples below in Section 3.4 with similar results.

It is worth discussing here the action of the discrete symmetries that the forward amplitudes. The kinematics of forward scattering preserves rotational invariance about the "beam direction" in the center of mass frame. The angular momentum of each state projected in this direction is a conserved charge that the external states are labelled by. This will be the subject of Section 3.5. Besides this however, there remains one further action of rotational invariance on the $S$-matrix. Rotations by $\pi$ perpendicular to the beam axis effectively interchanges (in the centre-of-mass frame) both particles one with two and three with four. This equates, up to a possible little group phase for inelastic amplitudes, the forward amplitudes $M^{ijkl}$ and $M^{jilk}$. This discrete rotation will be referred to as "$Y$". This symmetry is in addition to crossing, with which it can combine to produce $CPT$. In particular, $Y$-symmetry also acts on the vectors of UV couplings to imply that, in general, $|\boldsymbol{m}^{ij}| = |\boldsymbol{m}^{ji}|$. In many cases, when transitions between the $Y$-rotated pairs of states are prohibited (such as when there is angular momentum about the beam axis), the vectors themselves lie in orthogonal subspaces that may be directly identified through $Y$. Crossing, the Hermitian analyticity condition (3.3) and $Y$ symmetries (as well as the emergent $CPT$) can generally act to simplify the structure of the sum rule. In particular, $\boldsymbol{m}^{ij} \cdot \boldsymbol{m}^{kl} = \boldsymbol{m}^{\bar{k}\bar{l}} \cdot \boldsymbol{m}^{\bar{i}\bar{j}}$. Other discrete symmetries may exist for a particular theory, such as parity, time-reversal and identical particle exchange symmetries. Examples of these will be given throughout this work, but their existence is theory-dependent.

### 3.2.2 Loops

It is appropriate here to emphasise that, with no further assumptions beyond standard power counting, the dimension 8 order scattering amplitudes receive contributions not only from terms with single insertions of dimension 8 operators, but also from terms with multiple insertions of lower dimension operators that altogether give energy scaling of the same order. Double insertions of dimension 6 operators are particularly common. Much of previous work on these constraints has neglected the latter terms and are naively restricted in applicability to UV completions that generate small lower dimension Wilson coefficients (usually justified by appealing to a weak coupling expansion). It is not obvious how these bounds would apply to theories saturating naive dimensional analysis (44), which is characteristic of strongly coupled UV completions. One such example is chiral perturbation theory in the real world - see (45), (46), (47) for the results of applications of causality constraints to this. Double insertions of dim-6 cubic vector operators were, however, considered in (27) and (28), where it was interestingly observed that they enhanced the positivity constraints on the quartic vector operators.

A significant general statement about loop corrections from 4-point dimension 6 operators can be likewise made. A loop of two dim-6 insertions produce UV divergent bubble integrals, which are proportional to logarithm of the Mandelstam variable corresponding to the partitioning of the legs on either side of the loop. The coefficient of the logarithm is determined by the unitarity cut across the appropriate channel. For elastic amplitudes, both the $s$ and $u$-channel cuts are positive in the forward limit by the optical theorem. The $t$-channel cuts vanish in the forward limit by conservation angular momentum. This is because the dim-6 effective interactions can only mediate scattering in, at most, the $j = 1$ partial wave, implying that these terms cannot be proportional to more than one power of the crossed channel Mandelstam variable, here $s$ or $u$ (or combinations of spinor

bilinears that effectively behave as square roots of these). There must therefore be an overall factor that vanishes in the forward limit. The positivity of the $s$ and $u$-channel cuts implies that the coefficient of the UV log generated by the loop must be positive. This means that the dim-8 contact coefficients for elastic processes are always decreasing with increasing energy scale under renormalisation group (RG) evolution.

As explained above, the IR segment of the dispersion integral effectively accounts for RG evolution of the coupling from renormalisation scale $\mu$ to cut-off $\Lambda$. For elastic scattering, the higher $\Lambda$ is pushed, the more negative the loop correction appearing on the LHS becomes. It is for this reason optimal to integrate the IR dispersion integral as far as the cut-off. It is therefore the smaller, RG-evolved high-scale coupling that is constrained to be positive. The naive tree-level bounds on the low energy coupling are therefore strengthened by these dim-6 loops. The low-energy contact interaction is therefore subject to a stronger lower bound that depends both on the size of the dim-6 operators and the size of the energy hierarchy.

For bubble loop corrections to inelastic processes, another Schwarz-like bound can be placed on the cuts by noting that the sum over intermediate states (including phase space integration) is itself an inner product, so that:

$$\left|\sum_X \mathcal{M}^{ij\to X}\left(\mathcal{M}^{kl\to X}\right)^*\right| \leq \sqrt{\left(\sum_X \mathcal{M}^{ij\to X}\left(\mathcal{M}^{ij\to X}\right)^*\right)\left(\sum_X \mathcal{M}^{kl\to X}\left(\mathcal{M}^{kl\to X}\right)^*\right)}. \tag{3.12}$$

This bounds the size of the log coefficients for inelastic processes by those of elastic processes. In other words, the RG evolution of the corresponding tree operators is restricted by the elastic ones. This implies that the UV logs from loop corrections on the inelastic side of the constraints must be smaller than those on the elastic side. The RG evolution of the elastic amplitudes is therefore typically larger and determines how the constraints

tighten with scale.

These effects will be illustrated below in Section 3.3.1 in a simple, concrete example. All calculations performed here will be with the $\overline{MS}$ renormalisation scheme. Lower-dimension operators can also contribute to the amplitudes in more ways than simple UV renormalisations. A more thorough examination of these effects will be left for the future.

It will be likewise assumed that all marginal or renormalisable couplings are perturbative and that the loop corrections that they mediate are subdominant at leading order in the energy expansion of the EFT. These corrections may nevertheless be included in a similar way to the loops discussed above. If the energy hierarchy is large enough, the logarithms associated with these corrections become large and this is no longer justified. It would be interesting to also investigate how the RG flow would interact with the simplified conclusion derived above. Note that a perturbative treatment would apply under these conditions to the relevant couplings in the SM with the exception of the strong gauge coupling, which is non-perturbative at energies below $\sim$ 1 GeV. Extrapolating amplitudes in perturbative QCD to low energies is therefore unclear. A possible way of dodging the problem may be to modify the integration contour to cut-off the dispersion integral in the IR at some energy $s = r$ and then integrate over a semi-circular arc to the opposite branch cut. As long as $r \gg \Lambda_{QCD}$, then this should be computable within the perturbative theory, as long as the analytic continuation is still valid. If $r \ll \sigma^2 \ll \Lambda^2$, then this will also have only a small effect on the results derived under the assumptions above. Of course, this different contour choice does nothing to address the question of the validity of the foundational arguments underpinning the sum rule to Yang-Mills gauge theory, where the perturbative $S$-matrix must be presumably matched onto an inclusive IR observable.

### 3.2.3 Constraints

The sum rule in the form (3.8) and the convex cone picture of (29) yields a program for systematically extracting information in the sum rule into constraints on Wilson coefficients in an effective action that follows three stages:

1. Write down the sum rule and find the (potential) ERs.

2. Convert into inequalities among amplitudes.

3. Convert into inequalities among Wilson coefficients.

For simple enough theories, inequalities may be deduced directly from inspection of the sum rule without recourse to the convex cone picture. Invocation of convex geometry is most useful when the number of ERs is greater than the dimension of the space of independent amplitudes. As will be elaborated upon much more in Section 3.4, this typically occurs when many of the states are related by symmetries. While it is simple enough to outline the strategy above, much of the challenge lies in step 2 which itself thus far lacks a general procedure, although direct application of (3.11) is often substantial.

#### Sum rule and extremal rays

If sufficiently simple, constraints on the EFT can be deduced from the sum rule directly from inspection by expressing it in the form of (3.8), similarly to the way that the standard postivity results from elastic scattering and the inelastic bound (3.11) were derived. Examples of this will be given in the sections below. However, this is not always so simple when symmetries are present that impose further structure over the $S$-matrix.

Candidate ERs can be constructed by finding the ERs of the cone generated by only the $s$-channel term in (3.8), that is, the cone of positive semi-definite (PSD) matrices. Following (29), these will be referred to as potential ERs (PERs). Once the $u$-channel

term is added, all ERs must be PERs of the $s$-channel cone, but the converse does not necessarily hold and some PERs may be redundant (lie within the interior).

In the simple case where the symmetries of the theory are stringent enough to restrict $S$-matrix transitions to unique (irreps of) initial and final state species, then there is only a single independent vector in (3.8) that parameterises each transition and the magnitude squared of a single component of this vector, by itself, represents a full PSD matrix and defines a PER. Again, see Section 3.4.1 below for further details and explanation. This will be the situation discussed further below in Section 3.4.2. However, in the presence of multiple "degenerate" states (irreps) between which "inelastic" transitions are permitted, each PSD matrix consists of multiple parameters. Each parameter is a complex number that can be interpreted as a coupling of the IR states to a particular UV state with a specific set of quantum numbers. As rays, these are only of interest up to an overall scale. For example, for a theory with two degenerate irreps of scattering states under some symmetry (or, equivalently, distinct states with transition amplitudes permitted by symmetries), a PER has the form

$$\begin{pmatrix} 1 & r \\ r^* & |r|^2 \end{pmatrix} \tag{3.13}$$

for some unknown $r \in \mathbb{C}$. The undetermined components effectively parameterise a continuous family of (P)ERs that generate curved facets. A simple example of this will be illustrated in Section 3.5.2. Again see (28) for more details and discussion of application to SM electroweak bosons.

**From rays to amplitudes**

While the cone is fully determined as the convex hull of the rays, it is still required to convert the description into a set of inequalities on the amplitudes. For many of the simple examples described here, this step is relatively easy given the structure of the sum rule. However, when the shape of the cone becomes more intricate, as typically happens when the number of edges is larger than the dimension of the ambient space of amplitudes, then there is not a simple correspondence between coordinates/amplitudes and ERs. If the cone is polyhedral, the algorithm of vertex enumeration from convex geometry may be applied. This will be illustrated in the examples in Section 3.4.2 below. However, whenever degenerate states exist (two-particle states with the same quantum numbers, which are typically pervasive amongst theories), the cone is non-polyhedral. A systematic method for determining the shape of the cone, and hence the causality bounds, remains a problem for further work.

**From amplitudes to effective operators**

This step is well-known and not new. Given the effective action for the EFT, the standard Dyson series expansion can be performed to obtain the relevant scattering amplitudes at the relevant order of precision - here $\sim s^2/\Lambda^4$ for typical centre of mass energy scale $\sqrt{s}$. Their derivatives $\frac{d^2}{ds^2}A(s)$ that appear in the sum rule are then functions of the Wilson coefficients in the action.

The present work will include some exploration of lower dimension operators in the sum rule, mostly focused on dimension 6, and discuss how they modify constraints previously limited to dimension 8 Wilson coefficients. It is at this third stage in the program where this issue becomes relevant. However, an interpretation of the constraints directly on the structure of the $S$-matrix is unaffected.

It is interesting to wonder whether this step can be made altogether redundant. In such a formulation, the effective action would be redundant and the $S$-matrix may be perturbatively constructed directly from its singularity structure out of a set of contact interactions consistent with Lorentz invariance. The strength of the contact interactions would be an equivalent parameterisation to the Wilson coefficients. This program would require both a systematic understanding of the all-order singularity structure of the $S$-matrix (i.e. causality and locality) and a systematic method for actually performing this reconstructing in order to be a complete replacement, although for simple enough theories at low enough order and few enough legs (which cover all applications considered here), this is currently feasible.

## 3.3 Bounds on Inelastic Transitions

This section presents a simple example to concretely illustrate the general discussion presented above. However, the constraints presented here are also new and demonstrate the general way in which these bounds fundamentally limit the extent of symmetry violation by effective interactions.

### 3.3.1 Multiple scalars

For a theory of a single scalar field $\phi$, the positivity of the coefficient of the $(\partial\phi)^4$ operator is well known. This would be the leading irrelevant operator if the scalar was a Goldstone boson. Now consider a more general EFT of a complex scalar with effective interaction Lagrangian density

$$\mathcal{L}_{EFT_6} = \frac{c_6}{\Lambda^2}\phi\phi\left(\partial\phi^\dagger \cdot \partial\phi^\dagger\right) \tag{3.14}$$

and

$$\mathcal{L}_{EFT_8} = \frac{c_8}{\Lambda^4}\left(\partial\phi\cdot\partial\phi\right)\left(\partial\phi^\dagger\cdot\partial\phi^\dagger\right) + \frac{\tilde{c}_8}{\Lambda^4}\left(\partial\phi\cdot\partial\phi^\dagger\right)\left(\partial\phi\cdot\partial\phi^\dagger\right) + \frac{d_8}{2\Lambda^4}\left(\partial\phi\cdot\partial\phi\right)^2 + \frac{2\tilde{d}_8}{\Lambda^4}\left(\partial\phi\cdot\partial\phi\right)\left(\partial\phi\cdot\partial\phi^\dagger\right) + \text{conj.}. \tag{3.15}$$

Here $c_6$, $c_8$ and $\tilde{c}_8$ are real while $d_8$ and $\tilde{d}_8$ are complex. Complex scalars are usually associated with $U(1)$ symmetries, but this will not be assumed here. Only the existence of charge conjugation $C$ will be assumed to relate the two real scalar states. Nevertheless, at dim-6 level, the only possible 4-point operator is charge conserving (this would not be true if there were more species). See (48) for an analogous recent analysis of a two scalar system (axion and dilaton) in which both degrees of freedom are totally unrelated by symmetry (the bound is exactly that expected from (3.11)).

Note that marginal and relevant operators could also be included - it will be assumed that these are small perturbations such that they can be neglected from the leading contributions at each order the EFT expansion. Operators composed of more than four scalars $\phi$, such as $\phi^6$, have also been neglected for simplicity, but would ordinarily be considered at the same order. This would all be justified if the scalar was a Goldstone boson, in addition to ruling-out all possible dimension 6 operators and higher point dimension 8 operators so that $\mathcal{L}_{EFT_8}$ would give a complete leading order description of the interactions. However, I choose to include $\mathcal{L}_{EFT_6}$ here to provide a simple illustration of the inclusion of loops.

Firstly to analyse the structure of the constraints on the $S$-matrix entries. The sum rule can be organised into a matrix of incoming and outgoing states. This is given in Table 3.1. Here, there are four complex vectors with components corresponding to the amplitudes $m_i^{\phi\phi} = \mathcal{M}^{\phi\phi\to X_i}(s_i)$, $m_i^{\phi\bar{\phi}} = \mathcal{M}^{\phi\bar{\phi}\to X_i}(s_i)$, $m_i^{\bar{\phi}\phi} = \mathcal{M}^{\bar{\phi}\phi\to X_i}(s_i)$ and $m_i^{\bar{\phi}\bar{\phi}} = \mathcal{M}^{\bar{\phi}\bar{\phi}\to X_i}(s_i)$, where each entry in the vector corresponds to a particular state $i$ in the UV

| | $\phi\phi$ | $\phi\overline{\phi}$ | $\overline{\phi}\phi$ | $\overline{\phi\phi}$ |
|---|---|---|---|---|
| $\phi\phi$ | $\lvert\boldsymbol{m}^{\phi\phi}\rvert^2+\lvert\boldsymbol{m}^{\phi\bar{\phi}}\rvert^2$ | $2\boldsymbol{m}^{\phi\bar{\phi}}\cdot\boldsymbol{m}^{\phi\phi}$ | $\boldsymbol{m}^{\phi\phi}\cdot\boldsymbol{m}^{\phi\phi}+\boldsymbol{m}^{\phi\bar{\phi}}\cdot\boldsymbol{m}^{\phi\bar{\phi}}$ | $2\boldsymbol{m}^{\phi\bar{\phi}}\cdot\boldsymbol{m}^{\phi\phi}$ |
| $\phi\overline{\phi}$ | $2\boldsymbol{m}^{\phi\phi}\cdot\boldsymbol{m}^{\phi\bar{\phi}}$ | $\lvert\boldsymbol{m}^{\phi\bar{\phi}}\rvert^2+\lvert\boldsymbol{m}^{\phi\phi}\rvert^2$ | $2\boldsymbol{m}^{\bar{\phi}\phi}\cdot\boldsymbol{m}^{\phi\bar{\phi}}$ | $\boldsymbol{m}^{\bar{\phi}\bar{\phi}}\cdot\boldsymbol{m}^{\phi\bar{\phi}}+\boldsymbol{m}^{\bar{\phi}\phi}\cdot\boldsymbol{m}^{\phi\phi}$ |
| $\overline{\phi}\phi$ | $\boldsymbol{m}^{\phi\phi}\cdot\boldsymbol{m}^{\bar{\phi}\phi}+\boldsymbol{m}^{\phi\bar{\phi}}\cdot\boldsymbol{m}^{\bar{\phi}\bar{\phi}}$ | $2\boldsymbol{m}^{\phi\bar{\phi}}\cdot\boldsymbol{m}^{\bar{\phi}\phi}$ | $\lvert\boldsymbol{m}^{\bar{\phi}\phi}\rvert^2+\lvert\boldsymbol{m}^{\bar{\phi}\bar{\phi}}\rvert^2$ | $2\boldsymbol{m}^{\bar{\phi}\bar{\phi}}\cdot\boldsymbol{m}^{\bar{\phi}\phi}$ |
| $\overline{\phi\phi}$ | $2\boldsymbol{m}^{\phi\phi}\cdot\boldsymbol{m}^{\bar{\phi}\bar{\phi}}$ | $\boldsymbol{m}^{\phi\bar{\phi}}\cdot\boldsymbol{m}^{\bar{\phi}\bar{\phi}}+\boldsymbol{m}^{\phi\phi}\cdot\boldsymbol{m}^{\bar{\phi}\phi}$ | $2\boldsymbol{m}^{\bar{\phi}\phi}\cdot\boldsymbol{m}^{\bar{\phi}\bar{\phi}}$ | $\lvert\boldsymbol{m}^{\bar{\phi}\bar{\phi}}\rvert^2+\lvert\boldsymbol{m}^{\bar{\phi}\phi}\rvert^2$ |

Table 3.1: Sum rule for complex scalar theory.

completion up to an unimportant overall positive scalar coefficient. $CPT$ implies that all of the elastic amplitudes are equal so that $|\boldsymbol{m}^{\phi\phi}|^2+|\boldsymbol{m}^{\phi\bar{\phi}}|^2=|\boldsymbol{m}^{\bar{\phi}\phi}|^2+|\boldsymbol{m}^{\bar{\phi}\bar{\phi}}|^2$ and, up to an irrelevant phase, $\boldsymbol{m}^{\phi\bar{\phi}}\cdot\boldsymbol{m}^{\phi\phi}=\boldsymbol{m}^{\bar{\phi}\bar{\phi}}\cdot\boldsymbol{m}^{\bar{\phi}\phi}$. Then $Y$ symmetry by itself also implies that $|\boldsymbol{m}^{\phi\bar{\phi}}|=|\boldsymbol{m}^{\bar{\phi}\phi}|$ and $|\boldsymbol{m}^{\phi\phi}|=|\boldsymbol{m}^{\bar{\phi}\bar{\phi}}|$, while $\boldsymbol{m}^{\phi\bar{\phi}}\cdot\boldsymbol{m}^{\phi\phi}+\boldsymbol{m}^{\bar{\phi}\bar{\phi}}\cdot\boldsymbol{m}^{\phi\bar{\phi}}=2\boldsymbol{m}^{\phi\bar{\phi}}\cdot\boldsymbol{m}^{\phi\phi}$. This simplifies the matrix, in particular equating each single-charge violating amplitude in the upper triangle.

The standard positivity theorems on elastic forward scattering (2) are immediately clear from the diagonal entries in this table. However, there is clearly more information. These constraints can be extracted by applying the Schwarz and triangle inequalities. For example, $|2\boldsymbol{m}^{\phi\phi}\cdot\boldsymbol{m}^{\phi\bar{\phi}}|\leq 2|\boldsymbol{m}^{\phi\phi}||\boldsymbol{m}^{\phi\bar{\phi}}|\leq|\boldsymbol{m}^{\phi\phi}|^2+|\boldsymbol{m}^{\phi\bar{\phi}}|^2$, which is the statement on the LHS that

$$|M^{\phi\phi\phi\bar{\phi}}|\leq M^{\phi\bar{\phi}\phi\bar{\phi}}, \tag{3.16}$$

that is, the single-charge violating amplitude must be smaller than the charge conserving one. Likewise,

$$\begin{aligned}2|\boldsymbol{m}^{\phi\phi}\cdot\boldsymbol{m}^{\bar{\phi}\bar{\phi}}|+2|\boldsymbol{m}^{\phi\bar{\phi}}\cdot\boldsymbol{m}^{\bar{\phi}\bar{\phi}}|&\leq 2|\boldsymbol{m}^{\phi\phi}||\boldsymbol{m}^{\bar{\phi}\bar{\phi}}|+2|\boldsymbol{m}^{\phi\bar{\phi}}||\boldsymbol{m}^{\bar{\phi}\bar{\phi}}|\\&\leq 2\sqrt{\left(|\boldsymbol{m}^{\phi\phi}|^2+|\boldsymbol{m}^{\phi\bar{\phi}}|^2\right)\left(|\boldsymbol{m}^{\bar{\phi}\bar{\phi}}|^2+|\boldsymbol{m}^{\bar{\phi}\bar{\phi}}|^2\right)}\end{aligned} \tag{3.17}$$

implies that

$$|M^{\phi\phi\overline{\phi}\overline{\phi}}| + |M^{\phi\overline{\phi}\phi\overline{\phi}}| \leq 2M^{\phi\bar{\phi}\phi\bar{\phi}}. \tag{3.18}$$

These statements can then be converted into constraints on the Wilson coefficients. At tree-level, these bounds may be directly translated into the statement that the dim-8 charge conserving Wilson coefficients must be larger than the charge violating ones. This is, however, also an appropriate place to illustrate the inclusion of a loop process in the sum rule so that the affect of the dim-6 operators can be accounted for.

The relevant amplitudes at dim-8 order are

$$\begin{aligned} A\left(\phi,\overline{\phi}\to\phi,\overline{\phi}\right) = &\frac{c_8}{\Lambda^4}u^2 + \frac{\tilde{c}_8}{2\Lambda^4}\left(s^2+t^2\right) \\ &+ \frac{4}{(4\pi)^2}\frac{c_6^2}{\Lambda^4}\left(\frac{2}{3}u^2 + \int_0^1 x(1-x)\left(3u^2\log\left(\frac{\mu^2}{-x(1-x)u}\right)\right.\right. \\ &\left.\left. + t(t-u)\log\left(\frac{\mu^2}{-x(1-x)t}\right) + s(s-u)\log\left(\frac{\mu^2}{-x(1-x)s}\right)\right)dx\right) \end{aligned} \tag{3.19}$$

$$A\left(\phi,\phi\to\phi,\overline{\phi}\right) = \frac{\tilde{d}_8}{\Lambda^4}\left(s^2+t^2+u^2\right) \tag{3.20}$$

$$A\left(\phi,\phi\to\overline{\phi},\overline{\phi}\right) = \frac{d_8}{\Lambda^4}\left(s^2+t^2+u^2\right) \tag{3.21}$$

$$A\left(\phi,\overline{\phi}\to\overline{\phi},\phi\right) = A\left(\phi,\overline{\phi}\to\phi,\overline{\phi}\right)|_{u\mapsto t,t\mapsto s,s\mapsto u}. \tag{3.22}$$

As usual, $\mu$ is the renormalisation scale. The couplings are implicitly functions of this.

Taking the forward limit and differentiating give the entries for the LHS of the sum rule. Because of the singularities, I will take the insertion at $s = \sigma^2 + i\delta$ for some $\sigma^2 > 0$

and $\delta \to 0^+$, as explained in Section 3.2.1 above:

$$\overline{M}(\phi, \overline{\phi} \to \phi, \overline{\phi}) = \frac{2c_8}{\Lambda^4} + \frac{\tilde{c}_8}{\Lambda^4} + \frac{4}{(4\pi)^2}\frac{c_6^2}{\Lambda^4}\left(\frac{29}{18} + \frac{2\pi i}{3} + \frac{5}{3}\log\left(\frac{\mu^2}{\sigma^2}\right)\right) \tag{3.23}$$

$$\overline{M}(\phi, \phi \to \phi, \overline{\phi}) = \frac{4\tilde{d}_8}{\Lambda^4} \tag{3.24}$$

$$\overline{M}(\phi, \phi \to \overline{\phi}, \overline{\phi}) = \frac{4d_8}{\Lambda^4} \tag{3.25}$$

$$\overline{M}(\phi, \overline{\phi} \to \overline{\phi}, \phi) = \frac{2\tilde{c}_8}{\Lambda^4} + \frac{4}{(4\pi)^2}\frac{c_6^2}{\Lambda^4}\left(\frac{1}{9} + \frac{\pi i}{3} + \frac{2}{3}\log\left(\frac{\mu^2}{\sigma^2}\right)\right). \tag{3.26}$$

The IR part of the dispersion integral needs to be added to this to obtain the full LHS. This cancels the imaginary part of the amplitudes above (corresponding to above threshold production of the light states in the EFT), as well as the logarithmic dependence on $\sigma^2$, which can be taken arbitrarily soft, leaving behind a scheme-dependent correction to the coupling representing its RG evolution from $\mu$ to the cut-off. The terms from the dispersion integral relevant for each loop amplitude are:

$$\frac{2}{\pi}\int_0^{\Lambda^2} s\left(\frac{\sigma^{(6)}\left(\phi, \overline{\phi} \to \phi, \overline{\phi}\right)}{\left(s - \sigma^2 - i0^+\right)^3} + \frac{\sigma^{(6)}\left(\phi, \phi \to \phi, \phi\right)}{\left(s + \sigma^2\right)^3}\right) ds$$
$$\approx \frac{5}{12\pi^2}\frac{c_6^2}{\Lambda^4}\left(\frac{-3}{2} + \frac{2\pi i}{5} + \log\left(\frac{\Lambda^2}{\sigma^2}\right)\right) \tag{3.27}$$

$$\frac{1}{\pi}\int_0^{\Lambda^2}\left(\frac{1}{\left(s - \sigma^2 - i0^+\right)^3} + \frac{1}{\left(s + \sigma^2\right)^3}\right)\int A^{(6)}\left(\phi, \overline{\phi} \to \phi, \overline{\phi}\right)\left(A^{(6)}\left(\overline{\phi}, \phi \to \phi, \overline{\phi}\right)\right)^* d\Pi ds$$
$$\approx \frac{1}{6\pi^2}\frac{c_6^2}{\Lambda^4}\left(\frac{-3}{2} + \frac{\pi i}{2} + \log\left(\frac{\Lambda^2}{\sigma^2}\right)\right) \tag{3.28}$$

where $\sigma^{(6)}$ denotes cross section determined from the dim-6 tree amplitudes $A^{(6)}$ and $\Pi$ is the Lorentz-invariant phase space of the intermediate states being integrates over. The conventional optical theorem has been invoked in the statement of (3.27) because the relevant processes are elastic. The constraints (3.16) and (3.18) therefore translate into

bounds on Wilson coefficients:

$$4|\tilde{d}_8| \leq 2c_8 + \tilde{c}_8 + \frac{c_6^2}{(4\pi)^2}\left(\frac{148}{9} + \frac{20}{3}\log\left(\frac{\mu^2}{\Lambda^2}\right)\right) \tag{3.29}$$

$$2|d_8| + \left|\tilde{c}_8 + \frac{c_6^2}{(4\pi)^2}\left(\frac{20}{9} + \frac{4}{3}\log\left(\frac{\mu^2}{\Lambda^2}\right)\right)\right| \leq 2c_8 + \tilde{c}_8 + \frac{c_6^2}{(4\pi)^2}\left(\frac{148}{9} + \frac{20}{3}\log\left(\frac{\mu^2}{\Lambda^2}\right)\right). \tag{3.30}$$

The above example also makes explicit the issues of RG-scale dependence described in Section 3.2.1. The terms proportional to $\log\left(\frac{\mu^2}{\Lambda^2}\right)$ represent RG-evolution of the dim-8 couplings from the renormalisation scale $\mu$ to the cut-off $\Lambda$. The upper limit of the IR segment of the dispersion integral could have instead been chosen to be some $\lambda < \Lambda$. In this case, the above calculation would be mostly unchanged, but with $\Lambda \mapsto \lambda$ and the addition of terms $\mathcal{O}\left(\frac{\sigma^2}{\lambda^2}\right)$ on the RHS of (3.27) and (3.28). These latter terms were dropped with $\lambda = \Lambda$ because they are higher order in the energy expansion organising the EFT, but must be retained for smaller $\lambda$. They are nevertheless eliminated by taking $\sigma \to 0$. As a result, the constraints differ only in the replacement of the cut-off $\Lambda$ by the lower energy $\lambda$ in the logarithms, representing RG-evolution to the scale $\lambda$ instead. While the bounds must hold for all $\lambda$ and therefore represent constraints on the entire flow, they are typically optimised by taking $\lambda \to \Lambda$ because of the positive sign of the $c_6^2 \log$ contribution. This is the reason for integrating all the way to the cut-off. Interestingly, if the sign of the coefficient of the logarithm were instead negative, then taking $\lambda$ arbitrarily small would place arbitrarily strong lower bounds on the dim-8 coefficients mediating elastic scattering, effectively ruling them out. That this cannot happen is consequence of unitarity.

However, it is also of note that the logarithmic term is increasingly negative with higher cut-off. The division between the dim-8 contact coefficients and the rational terms

proportional to $c_6^2$ is renormalisation scheme-dependent and the sum of both terms should be the object of comparison with the pure tree-level bounds. The dim-6 loop contribution should be entirely attributed to the logarithm and this contributes negatively to the LHS of the sum rule, strengthening the bound relative to the tree-level expectation. Stated equivalently, the contact coefficients decrease with increasing RG scale to the extent that their IR values can be (at least partially) cancelled. For a given coupling at scale $\mu$, the constraints fundamentally limit the extent that the cut-off can be extended, or alternatively, for a given cut-off, improve the positivity bounds on the dim-8 contact interaction strength in the IR. The mere presence of dim-6 operators therefore enhances the lower bound on the size of dim-8 operators mediating elastic scattering.

The organisation of the sum rule presented here can be further applied to more complicated theories with multiple species, of which many examples will follow.

## 3.4 Bounds with Internal Symmetries - Flavour and Colour

The presentation of the sum rule illustrated in the previous sections demonstrates the nature of the new bounds for inelastic processes between distinct particles. The next level of sophistication to discuss is for theories with multiple states related by symmetries. Positivity bounds in theories with global symmetries have been discussed previously in (15). The present discussion will further examine the new positivity bounds suggested in (29) for theories of particles belonging to non-trivial representations of multiple symmetry groups. While the action of the symmetries on the states factorise, because all sets of indices are crossed simultaneously between the $s$-channel and $u$-channel terms in (3.8), they become effectively entangled across different entries in the full matrix of sum rules.

This raises the possibility of new constraints that cannot be accessed by considering only elastic scattering of factorised superpositions of states.

The bounds are determined from the ERs in the convex cone picture. Each ER is itself identified with a particular irrep under the global symmetry that a pair of external scattering states can couple to. Thus if a particular representation $R$ lies in the Clebsch-Gordan decomposition of the incoming states in either the $s$-channel or the $u$-channel, it will yield an ER in the RHS of the sum rule. Stage one of the procedure enumerated in Section 3.2.3 thus reduces to a decomposition of the theory's amplitudes into a set of partial amplitudes describing symmetry preserving transitions. This will be elaborated upon more precisely below.

The goal of this section is to analyse theories of global symmetries with the features just described and discover new bounds. This will provide a further educational illustration of the structure of the sum rule constraints, the way in which symmetries are managed and the convex cone of UV completions. More importantly however, the cases considered here will directly apply to the fermionic operators with the global symmetries of the SM. The important special case of rotational symmetry will be deferred to the next section in order to avoid distraction from the goal. However, for the theories of (hyper)charged chiral fermions considered here, such a treatment for spin is not necessary and each left-handed particle and right-handed antiparticle may be treated as independent states that have amplitudes related by only $CPT$, similarly to the scalars in the previous section (although hypercharge conservation will be assumed here). Note that the following method for the accounting of symmetries and the algorithm used to derive the positivity bounds are not intended to represent an application of the simple inelastic bounds derived earlier in (3.11) and could include more information.

### 3.4.1 Background group theory

A $2 \to 2$ scattering amplitude can be decomposed into partial amplitudes corresponding to symmetry preserving transitions between particular irreps. Assume that the incoming and outgoing particles transform under representations of some symmetry group. Generally, the initial states may be decomposed into irreps with Clebsch-Gordan coefficients defined here as

$$C^{ab}_{R_{\xi\iota}} = \langle R_\xi; \iota | \left( |r1; a\rangle |r2; b\rangle \right) \tag{3.31}$$

to give

$$|r1; a\rangle |r2; b\rangle = \sum_{R_\xi, \iota} C^{ab}_{R_{\xi\iota}} |R_\xi; \iota\rangle, \tag{3.32}$$

where $r_1$ and $r_2$ label the representations of the individual particles, $a$ and $b$ their components, while $R$ and $\iota$ index the product irreps and components. The index $\xi$ counts degenerate representations that may arise.

Projection tensors can be defined as

$$P^{abcd}_{R_{\xi\xi'}} = \sum_\iota C^{ab}_{R_{\xi\iota}} \left( C^{cd}_{R_{\xi'\iota}} \right)^* . \tag{3.33}$$

These obey orthogonality conditions

$$\frac{1}{\dim R} \sum_{a,b,c,d} \left( P^{abcd}_{R_{\xi\xi'}} \right)^* \left( P^{abcd}_{R'_{\eta\eta'}} \right) = \delta_{RR'} \delta_{\xi\eta} \delta_{\xi'\eta'}. \tag{3.34}$$

The final states may be likewise decomposed. The Wigner-Eckart theorem then implies that the resulting transition amplitude is diagonal in representation and compo-

nents, although transitions between distinct but degenerate representations are permitted. This will be especially important when spin is discussed later. The full amplitude decomposes as

$$\begin{aligned} A^{abcd} &= {}_{\text{out}}\left(\langle r4;d|\langle r3;c|\right)\left(|r1;a\rangle|r2;b\rangle\right)_{\text{in}} \\ &= \sum_{R,\xi,\xi'} P^{abcd}_{R_{\xi\xi'}} A_{R_{\xi\xi'}}, \end{aligned} \tag{3.35}$$

where the partial amplitudes are defined as

$$A_{R_{\xi\xi'}} = {}_{\text{out}}\langle R_{\xi'};\iota|R_{\xi};\iota\rangle_{\text{in}} \tag{3.36}$$

and may be extracted from the full amplitude by the action of projectors (note that the RHS of (3.36) is independent of the choice of the component $\iota$ and no sum is implied).

The projection operators thus encode all of the symmetry relations between amplitudes of different states. The irreps are the states that block-diagonalise the $S$-matrix. Each term in the $s$-channel of the sum rule (the first term in (3.8)) can be decomposed into irreps into the form

$$M_s = \sum_{R_s,\xi,\xi'} P^{abcd}_{R_{s\xi\xi'}}\, \boldsymbol{m}_{R_{s\xi'}} \cdot \boldsymbol{m}_{R_{s\xi}}, \tag{3.37}$$

where the $s$ subscript on the irrep label $R$ has been used to emphasise applicability to the $s$-channel decomposition. The u-channel term in (3.8) can likewise be decomposed with particles $b$ and $d$ exchanged with each others' antiparticles. The irreps in this case may be entirely different. However, by the Wigner-Eckart theorem, it must be possible for the $u$-channel projectors to be decomposed as linear combinations of the $s$-channel ones so

that the amplitude takes the form (3.35) given entirely in terms of $s$-channel projectors.

$$P^{a\bar{d}c\bar{b}}_{R_{u\rho\rho'}} = \sum_{R_s,\xi,\xi'} c_{R_{u\rho\rho'}R_{s\xi\xi'}} P^{abcd}_{R_{s\xi\xi'}}. \tag{3.38}$$

The numerical constants $c_{R_{u\rho\rho'}R_{s\xi\xi'}}$ of this decomposition are entirely determined by the Clebsch-Gordan coefficients of the group. These will be presented below for various examples relevant to the SM. As a result of this decomposition, the sum rule (3.8) in the presence of global symmetries may be expressed as

$$M^{abcd} = P^{abcd}_{R_{\xi\xi'}} \boldsymbol{m}_{R_{\xi'}} \cdot \boldsymbol{m}_{R_\xi} + \sum_{R_s,\xi,\xi'} c_{R_{u\rho\rho'}R_{s\xi\xi'}} P^{abcd}_{R_{\xi\xi'}} \boldsymbol{m}_{R_{u\rho'}} \cdot \boldsymbol{m}_{R_{u\rho}} \tag{3.39}$$

The $s$ label on the irreps of the $s$-channel has now been dropped. If there are no degeneracies, then each term in (3.39) is of the form $|\boldsymbol{m}_R|^2$ and can be identified with a PER.

The special case of $SU(3)$ will be used in examples below. Projectors for the irreps that arise in combining fundamental and antifundamental representations will be necessary. The Clebsch-Gordan coefficients may be easily inferred from the exchange symmetry structure of the representations in tensor form. The projectors onto each irrep appearing in the products $\mathbf{3} \otimes \mathbf{3}$ are then determined as

$$P^{ab}_{\mathbf{\bar{3}}\ cd} = \frac{1}{2}\left(\delta^a_c\delta^b_d - \delta^a_d\delta^b_c\right) \tag{3.40}$$

$$P^{ab}_{\mathbf{6}\ cd} = \frac{1}{2}\left(\delta^a_c\delta^b_d + \delta^a_d\delta^b_c\right). \tag{3.41}$$

For $\mathbf{3} \otimes \overline{\mathbf{3}}$ transitioning into $\mathbf{3} \otimes \overline{\mathbf{3}}$, they are

$$P^{a\ \ d}_{\mathbf{1}\,bc} = \frac{1}{3}\delta^a_b\delta^d_c \tag{3.42}$$

$$P^{a\ \ d}_{\mathbf{8}\,bc} = \delta^a_c\delta^d_b - \frac{1}{3}\delta^a_b\delta^d_c, \tag{3.43}$$

while for $\mathbf{3} \otimes \overline{\mathbf{3}}$ transitioning into $\overline{\mathbf{3}} \otimes \mathbf{3}$, they are

$$P^{a\ c}_{\mathbf{1}\,b\ d} = \frac{1}{3}\delta^a_b\delta^c_d \tag{3.44}$$

$$P^{a\ c}_{\mathbf{8}\,b\ d} = \delta^a_d\delta^c_b - \frac{1}{3}\delta^a_b\delta^c_d. \tag{3.45}$$

Raised indices indicate fundamental, lowered are antifundamental. The cases where the representations are conjugate are identical, with index heights reversed.

It will also be necessary to decompose the projectors in the $u$-channel into projectors for the $s$-channel, which is traditionally referred to as "finding the crossing matrix". For the projectors above, these are

$$P^{a\ \ b}_{\mathbf{1}\,dc} = \frac{1}{3}\left(P^{ab}_{\mathbf{6}\ cd} - P^{ab}_{\overline{\mathbf{3}}\ cd}\right) \tag{3.46}$$

$$P^{a\ \ b}_{\mathbf{8}\,dc} = \frac{2}{3}\left(P^{ab}_{\mathbf{6}\ cd} + 2P^{ab}_{\overline{\mathbf{3}}\ cd}\right) \tag{3.47}$$

$$P^{a\ c}_{\mathbf{1}\,d\ b} = \frac{1}{3}\left(P^{a\ c}_{\mathbf{1}\,b\ d} + P^{a\ c}_{\mathbf{8}\,b\ d}\right) \tag{3.48}$$

$$P^{a\ c}_{\mathbf{8}\,d\ b} = \frac{1}{3}\left(8P^{a\ c}_{\mathbf{1}\,b\ d} - P^{a\ c}_{\mathbf{8}\,b\ d}\right) \tag{3.49}$$

$$P^{ad}_{\overline{\mathbf{3}}\ cb} = \frac{1}{2}\left(P^{a\ \ d}_{\mathbf{8}\,bc} - 2P^{a\ \ d}_{\mathbf{1}\,bc}\right) \tag{3.50}$$

$$P^{ad}_{\mathbf{6}\ cb} = \frac{1}{2}\left(P^{a\ \ d}_{\mathbf{8}\,bc} + 4P^{a\ \ d}_{\mathbf{1}\,bc}\right). \tag{3.51}$$

Also of use will be the projectors for $SU(2)$. For the product $\mathbf{2} \otimes \mathbf{2}$, they are

$$P_{\mathbf{1}}^{ab}{}_{cd} = -\frac{1}{2}\epsilon^{ab}\epsilon_{cd} \tag{3.52}$$

$$P_{\mathbf{3}}^{ab}{}_{cd} = \frac{1}{2}\left(\delta^a_c\delta^b_d + \delta^a_d\delta^b_c\right). \tag{3.53}$$

The indices can be raised and lowered by $\epsilon$ tensors in order to relate these to the projectors appearing in amplitudes involving the conjugate representations. An additional factor of $-1$ must be included for each index either raised or lowered in this way (because, for a state $\psi^a$, defining $\psi_a = \epsilon_{ab}\psi^b$ and $\psi^{\dagger a} = \epsilon^{ab}\psi^\dagger_b$, then $(\psi_a)^\dagger = -\psi^{\dagger a}$). The $u$-channel projectors decompose as

$$P_{\mathbf{1}}^{a}{}_{dc}^{\ b} = -\frac{1}{2}\left(P_{\mathbf{1}}^{ab}{}_{cd} - P_{\mathbf{3}}^{ab}{}_{cd}\right) \tag{3.54}$$

$$P_{\mathbf{3}}^{a}{}_{dc}^{\ b} = \frac{1}{2}\left(3P_{\mathbf{1}}^{ab}{}_{cd} + P_{\mathbf{3}}^{ab}{}_{cd}\right). \tag{3.55}$$

For the product $\mathbf{3} \otimes \mathbf{3}$, they are

$$P_{\mathbf{1}}^{abcd} = \frac{1}{2}\delta^{ab}\delta^{cd} \tag{3.56}$$

$$P_{\mathbf{3}}^{abcd} = \frac{1}{2}\left(\delta^{ac}\delta^{bd} - \delta^{ad}\delta^{bc}\right) \tag{3.57}$$

$$P_{\mathbf{5}}^{abcd} = \frac{1}{2}\left(\delta^{ac}\delta^{bd} + \delta^{ad}\delta^{bc} - \delta^{ab}\delta^{cd}\right). \tag{3.58}$$

The indices here label components of the **3** representation, rather than fundamental. The

$u$-channel projectors decompose as

$$P_{\mathbf{1}}^{adcb} = \frac{1}{3}\left(P_{\mathbf{5}}^{abcd} - P_{\mathbf{3}}^{abcd} + P_{\mathbf{1}}^{abcd}\right) \tag{3.59}$$

$$P_{\mathbf{3}}^{adcb} = \frac{1}{2}\left(P_{\mathbf{5}}^{abcd} + P_{\mathbf{3}}^{abcd} - 2P_{\mathbf{1}}^{abcd}\right) \tag{3.60}$$

$$P_{\mathbf{5}}^{adcb} = \frac{1}{6}\left(P_{\mathbf{5}}^{abcd} + 5P_{\mathbf{3}}^{abcd} + 10P_{\mathbf{1}}^{abcd}\right). \tag{3.61}$$

### 3.4.2 Standard Model fermions

As the SM is a theory of chiral fermions, EFTs of these states will be the focus of this section. The isospin, colour and flavour representations of the SM fermions will be systematically considered, with hypercharge conservation imposed. Helicity and hypercharge are not independent quantum numbers, so both sets of representations are equivalent. However, in the product representations of two such states, the non-zero charged irreps correspond to the rotational singlets, while the charge singlets constitute the non-trivial angular momentum irreps (see Section 3.5.1 below for more explanation). This ensures that there are no transitions between distinct degenerate irreps and that hypercharge conservation is otherwise sufficient to account for both of these symmetries.

I begin with a theory of hypercharged chiral fermions in the fundamental representation of $SU(3)$. These results would apply to a single flavour of right-handed down or up quarks, with the $SU(3)$ symmetry being interpreted as colour, or to right-handed leptons in which the full $SU(3)$ flavour symmetry is preserved in the UV. I will use notation describing the former. The terms in the sum rule are determined by finding projectors for the irreps of the external legs, beginning with the $s$-channel and then crossing to the $u$-channel, in a similar way to that illustrated in the previous section. Again parameterising the couplings to UV states by complex vectors, the relevant amplitudes are of the

form

$$\begin{aligned} M\left(q_R, q_R \to q_R, q_R\right) &= M_{\overline{\mathbf{3}}} P_{\overline{\mathbf{3}}} + M_{\mathbf{6}} P_{\mathbf{6}} \\ &= \left(|\boldsymbol{m}_{\overline{\mathbf{3}}}|^2 - \frac{1}{3}|\boldsymbol{m}_{\mathbf{1}}|^2 + \frac{4}{3}|\boldsymbol{m}_{\mathbf{8}}|^2\right) P_{\overline{\mathbf{3}}} \\ &\quad + \left(|\boldsymbol{m}_{\mathbf{6}}|^2 + \frac{1}{3}|\boldsymbol{m}_{\mathbf{1}}|^2 + \frac{2}{3}|\boldsymbol{m}_{\mathbf{8}}|^2\right) P_{\mathbf{6}} \end{aligned} \tag{3.62}$$

$$\begin{aligned} M\left(q_R, \overline{q}_L \to q_R, \overline{q}_L\right) &= M_{\mathbf{1}} P_{\mathbf{1}} + M_{\mathbf{8}} P_{\mathbf{8}} \\ &= \left(|\boldsymbol{m}_{\mathbf{1}}|^2 - |\boldsymbol{m}_{\overline{\mathbf{3}}}|^2 + 2|\boldsymbol{m}_{\mathbf{6}}|^2\right) P_{\mathbf{1}} \\ &\quad + \left(|\boldsymbol{m}_{\mathbf{8}}|^2 + \frac{1}{2}|\boldsymbol{m}_{\overline{\mathbf{3}}}|^2 + \frac{1}{2}|\boldsymbol{m}_{\mathbf{6}}|^2\right) P_{\mathbf{8}}. \end{aligned} \tag{3.63}$$

The others are either related by $CPT$ or are zero. The partial amplitudes for only one of these transitions are independent - the other channel is determined by crossing. It is easy to see that the parameters $\boldsymbol{m}_{\mathbf{6}}$ and $\boldsymbol{m}_{\mathbf{8}}$ are redundant. The remaining terms clearly span a $2d$ cone and can be converted into inequalities

$$M_{\overline{\mathbf{3}}} + M_{\mathbf{6}} > 0 \tag{3.64}$$

$$M_{\mathbf{6}} > 0. \tag{3.65}$$

These bounds correspond to those found in (25) and (unsurprisingly) contain no new information.

Now to advance to fermions in the fundamental representation of $SU(2) \otimes SU(3)$, for example, left-handed leptons with flavour symmetry or a single flavour of left-handed

quark. Adopting the latter interpretation, the independent, non-zero amplitudes are

$$\begin{aligned}
&M\left(Q_L, Q_L \to Q_L, Q_L\right) \\
&\qquad = \left(|\boldsymbol{m}_{(\mathbf{1},\overline{\mathbf{3}})}|^2 + \frac{1}{6}|\boldsymbol{m}_{(\mathbf{1},\mathbf{1})}|^2 - \frac{2}{3}|\boldsymbol{m}_{(\mathbf{1},\mathbf{8})}|^2 - \frac{1}{2}|\boldsymbol{m}_{(\mathbf{3},\mathbf{1})}|^2 + 2|\boldsymbol{m}_{(\mathbf{3},\mathbf{8})}|^2\right) P_{\mathbf{1}} P_{\overline{\mathbf{3}}} \\
&\qquad + \left(|\boldsymbol{m}_{(\mathbf{1},\mathbf{6})}|^2 - \frac{1}{6}|\boldsymbol{m}_{(\mathbf{1},\mathbf{1})}|^2 - \frac{1}{3}|\boldsymbol{m}_{(\mathbf{1},\mathbf{8})}|^2 + \frac{1}{2}|\boldsymbol{m}_{(\mathbf{3},\mathbf{1})}|^2 + |\boldsymbol{m}_{(\mathbf{3},\mathbf{8})}|^2\right) P_{\mathbf{1}} P_{\mathbf{6}} \\
&\qquad + \left(|\boldsymbol{m}_{(\mathbf{3},\overline{\mathbf{3}})}|^2 - \frac{1}{6}|\boldsymbol{m}_{(\mathbf{1},\mathbf{1})}|^2 + \frac{2}{3}|\boldsymbol{m}_{(\mathbf{1},\mathbf{8})}|^2 - \frac{1}{6}|\boldsymbol{m}_{(\mathbf{3},\mathbf{1})}|^2 + \frac{2}{3}|\boldsymbol{m}_{(\mathbf{3},\mathbf{8})}|^2\right) P_{\mathbf{3}} P_{\overline{\mathbf{3}}} \\
&\qquad + \left(|\boldsymbol{m}_{(\mathbf{3},\mathbf{6})}|^2 + \frac{1}{6}|\boldsymbol{m}_{(\mathbf{1},\mathbf{1})}|^2 + \frac{1}{3}|\boldsymbol{m}_{(\mathbf{1},\mathbf{8})}|^2 + \frac{1}{6}|\boldsymbol{m}_{(\mathbf{3},\mathbf{1})}|^2 + \frac{1}{3}|\boldsymbol{m}_{(\mathbf{3},\mathbf{8})}|^2\right) P_{\mathbf{3}} P_{\mathbf{6}}
\end{aligned} \tag{3.66}$$

$$\begin{aligned}
&M\left(Q_L, \overline{Q}_R \to Q_L, \overline{Q}_R\right) \\
&\qquad = \left(|\boldsymbol{m}_{(\mathbf{1},\mathbf{1})}|^2 + \frac{1}{2}|\boldsymbol{m}_{(\mathbf{1},\overline{\mathbf{3}})}|^2 - |\boldsymbol{m}_{(\mathbf{1},\mathbf{6})}|^2 - \frac{3}{2}|\boldsymbol{m}_{(\mathbf{3},\overline{\mathbf{3}})}|^2 + 3|\boldsymbol{m}_{(\mathbf{3},\mathbf{6})}|^2\right) P_{\mathbf{1}} P_{\mathbf{1}} \\
&\qquad + \left(|\boldsymbol{m}_{(\mathbf{1},\mathbf{8})}|^2 - \frac{1}{4}|\boldsymbol{m}_{(\mathbf{1},\overline{\mathbf{3}})}|^2 - \frac{1}{4}|\boldsymbol{m}_{(\mathbf{1},\mathbf{6})}|^2 + \frac{3}{4}|\boldsymbol{m}_{(\mathbf{3},\overline{\mathbf{3}})}|^2 + \frac{3}{4}|\boldsymbol{m}_{(\mathbf{3},\mathbf{6})}|^2\right) P_{\mathbf{1}} P_{\mathbf{8}} \\
&\qquad + \left(|\boldsymbol{m}_{(\mathbf{3},\mathbf{1})}|^2 - \frac{1}{2}|\boldsymbol{m}_{(\mathbf{1},\overline{\mathbf{3}})}|^2 + |\boldsymbol{m}_{(\mathbf{1},\mathbf{6})}|^2 - \frac{1}{2}|\boldsymbol{m}_{(\mathbf{3},\overline{\mathbf{3}})}|^2 + |\boldsymbol{m}_{(\mathbf{3},\mathbf{6})}|^2\right) P_{\mathbf{3}} P_{\mathbf{1}} \\
&\qquad + \left(|\boldsymbol{m}_{(\mathbf{3},\mathbf{8})}|^2 + \frac{1}{4}|\boldsymbol{m}_{(\mathbf{1},\overline{\mathbf{3}})}|^2 + \frac{1}{4}|\boldsymbol{m}_{(\mathbf{1},\mathbf{6})}|^2 + \frac{1}{4}|\boldsymbol{m}_{(\mathbf{3},\overline{\mathbf{3}})}|^2 + \frac{1}{4}|\boldsymbol{m}_{(\mathbf{3},\mathbf{6})}|^2\right) P_{\mathbf{3}} P_{\mathbf{8}}.
\end{aligned} \tag{3.67}$$

Choosing the coordinates to be the partial amplitudes $(M_{(\mathbf{1},\overline{\mathbf{3}})}, M_{(\mathbf{1},\mathbf{6})}, M_{(\mathbf{3},\overline{\mathbf{3}})}, M_{(\mathbf{3},\mathbf{6})})$, the extremal rays can be read-off the sum rule and are (up to an arbitrary scale):

$$\{(1,0,0,0),(0,1,0,0),(0,0,1,0),(1,-1,-1,1),(-2,-1,2,1),(-3,3,-1,1)\}. \tag{3.68}$$

Note that the terms parameterised by both the vectors $\boldsymbol{m}_{(\mathbf{3},\mathbf{6})}$ and $\boldsymbol{m}_{(\mathbf{3},\mathbf{8})}$ are redundant, so have been excluded. The standard techniques of vertex enumeration may be directly applied to this system (see e.g. (49)) in order to convert the extremal rays into linear

inequalities among coordinates. The present example is readily computed by hand. However, the following examples rapidly grow in complexity. I use lrs (50) as a cross-check here and to compute the more complicated examples to follow. The resulting constraints are

$$M_{(\mathbf{1},\mathbf{6})} + M_{(\mathbf{3},\mathbf{6})} > 0 \tag{3.69}$$

$$M_{(\mathbf{3},\mathbf{6})} > 0 \tag{3.70}$$

$$M_{(\mathbf{1},\overline{\mathbf{3}})} + M_{(\mathbf{1},\mathbf{6})} + M_{(\mathbf{3},\overline{\mathbf{3}})} + M_{(\mathbf{3},\mathbf{6})} > 0 \tag{3.71}$$

$$M_{(\mathbf{3},\overline{\mathbf{3}})} + M_{(\mathbf{3},\mathbf{6})} > 0 \tag{3.72}$$

$$4M_{(\mathbf{1},\overline{\mathbf{3}})} + M_{(\mathbf{1},\mathbf{6})} + 9M_{(\mathbf{3},\mathbf{6})} > 0 \tag{3.73}$$

$$M_{(\mathbf{1},\overline{\mathbf{3}})} + 3M_{(\mathbf{3},\mathbf{6})} > 0. \tag{3.74}$$

These bounds can be compared to those given in (25) derived specifically from elastic forward scattering. At tree-level, each partial amplitude corresponds to a particular dim-8 operator of the form $(\psi D\psi) \cdot \left(\psi^\dagger D\psi^\dagger\right)$ for chiral fermionic operator $\psi$. The correspondence is determined by decomposing the bilinears $\psi D\psi$ into the irreps of $SU(2)$ and $SU(3)$ given above - each different operator corresponds to one of the four such representations and thus corresponds to one of the four such partial amplitudes. The irrep into which $\psi D\psi$ is decomposed can only be contracted into a singlet with its conjugate, thus fully determining the operator. From the correspondence between operators and partial amplitudes it is sufficient to see that this set of six irreducible bounds contains more information than the four linear inequalities stated in (25). In particular, the last two are new. As argued in (29), this is because the crossing between $s$ and $u$-channel terms in (3.8) involves simultaneously exchanging all degrees of freedom associated to the states, effectively entangling them.

This example illustrates the insight of the convex cone picture in cases where the space of allowed couplings is bounded by more faces than the dimension of the ambient space. In simple examples like the pure $SU(3)$ case described above, the number of linear positivity constraints are comparable to the dimension of the space being bounded, so bounds can be derived almost by direct inspection of the sum rule and some simple geometry. However, when the number of vectors parameterising the RHS of the sum rule exceeds the number of independent partial amplitudes, the space of allowed couplings becomes a multi-faceted polyhedral cone and the tools of convex geometry must be invoked.

Next are fermions in the fundamental representation of $SU(3) \otimes SU(3)$, such as right-handed quarks with flavour symmetry. This is similar to the above example. The forward elastic amplitude $M\left(q_R, q_R \to q_R, q_R\right)$ can be decomposed into four independent partial amplitudes $\{M_{(\overline{\mathbf{3}},\overline{\mathbf{3}})}, M_{(\overline{\mathbf{3}},\mathbf{6})}, M_{(\mathbf{6},\overline{\mathbf{3}})}, M_{(\mathbf{6},\mathbf{6})}\}$. The positivity constraints can be deduced by the same procedure to be

$$M_{(\overline{\mathbf{3}},\mathbf{6})} + M_{(\mathbf{6},\mathbf{6})} > 0 \tag{3.75}$$

$$M_{(\mathbf{6},\mathbf{6})} > 0 \tag{3.76}$$

$$M_{(\overline{\mathbf{3}},\overline{\mathbf{3}})} + M_{(\overline{\mathbf{3}},\mathbf{6})} + M_{(\mathbf{6},\overline{\mathbf{3}})} + M_{(\mathbf{6},\mathbf{6})} > 0 \tag{3.77}$$

$$M_{(\mathbf{6},\overline{\mathbf{3}})} + M_{(\mathbf{6},\mathbf{6})} > 0 \tag{3.78}$$

$$M_{(\overline{\mathbf{3}},\overline{\mathbf{3}})} + 2M_{(\mathbf{6},\mathbf{6})} > 0. \tag{3.79}$$

As the $SU(2)$ and $SU(3)$ projectors for these fundamental representations have the same tensor form, the first four bounds are analogous to those derived from forward scattering in the $Q_L$ case above. The last bound again corresponds to positivity of an entangled amplitude and differs from the example above because of the different crossing relations.

The final case considered here will be fermions in the fundamental representation of $SU(2) \otimes SU(3) \otimes SU(3)$, corresponding to states with isospin, colour and flavour, such as left-handed quarks. This is substantially more complicated than the previous two examples. The independent non-zero partial amplitudes are those of $M\left(Q_L, Q_L \to Q_L, Q_L\right) = \sum_{I,a,i} M_{(\boldsymbol{I},\boldsymbol{a},\boldsymbol{i})} P_{(\boldsymbol{I},\boldsymbol{a},\boldsymbol{i})}$, where $(\boldsymbol{I}, \boldsymbol{a}, \boldsymbol{i})$ indexes $SU(2) \otimes SU(3) \otimes SU(3)$ representations. This is an 8-dimensional space, with a general vector of partial amplitudes denoted by $\boldsymbol{M} = (M_{(\mathbf{1},\overline{\mathbf{3}},\overline{\mathbf{3}})}, M_{(\mathbf{3},\overline{\mathbf{3}},\overline{\mathbf{3}})}, M_{(\mathbf{1},\mathbf{6},\overline{\mathbf{3}})}, M_{(\mathbf{3},\mathbf{6},\overline{\mathbf{3}})}, M_{(\mathbf{1},\overline{\mathbf{3}},\mathbf{6})}, M_{(\mathbf{3},\overline{\mathbf{3}},\mathbf{6})}, M_{(\mathbf{1},\mathbf{6},\mathbf{6})}, M_{(\mathbf{3},\mathbf{6},\mathbf{6})})$. Repeating the procedure as above, there are 16 PERs (for each partial amplitude in this and the crossed channel), of which two are redundant, leaving 14 ERs. These may be converted

into 44 positivity bounds:

$$
\begin{array}{cccccccc}
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 3 & 2 & 0 & 2 & 0 & 0 & 6 \\
5 & 9 & 8 & 0 & 8 & 0 & 2 & 18 \\
11 & 15 & 16 & 0 & 16 & 0 & 0 & 48 \\
1 & 0 & 1 & 0 & 0 & 3 & 0 & 3 \\
4 & 0 & 4 & 0 & 1 & 9 & 1 & 9 \\
8 & 0 & 8 & 0 & 3 & 15 & 0 & 24 \\
1 & 1 & 0 & 0 & 0 & 0 & 2 & 2 \\
8 & 0 & 4 & 0 & 0 & 12 & 5 & 21 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 6 \\
7 & 0 & 2 & 0 & 2 & 0 & 0 & 36 \\
2 & 0 & 0 & 0 & 0 & 0 & 1 & 9 \\
4 & 0 & 0 & 0 & 0 & 4 & 5 & 13 \\
8 & 0 & 0 & 12 & 4 & 0 & 5 & 21 \\
3 & 0 & 0 & 5 & 0 & 5 & 0 & 13 \\
8 & 0 & 0 & 4 & 0 & 4 & 7 & 23 \\
4 & 0 & 0 & 4 & 0 & 0 & 5 & 13 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 2 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 3 & 1 & 0 & 2 & 1 & 0 & 7
\end{array}
\qquad
\begin{array}{cccccccc}
1 & 0 & 0 & 3 & 1 & 0 & 0 & 3 \\
4 & 0 & 1 & 9 & 4 & 0 & 1 & 9 \\
8 & 0 & 3 & 15 & 8 & 0 & 0 & 24 \\
0 & 0 & 0 & 0 & 4 & 0 & 1 & 9 \\
4 & 0 & 0 & 8 & 4 & 0 & 3 & 11 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 3 \\
1 & 0 & 1 & 0 & 6 & 0 & 0 & 18 \\
0 & 3 & 1 & 0 & 5 & 0 & 0 & 15 \\
0 & 1 & 0 & 1 & 3 & 0 & 0 & 9 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
4 & 0 & 4 & 0 & 0 & 8 & 3 & 11 \\
0 & 0 & 4 & 0 & 0 & 0 & 1 & 9 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 3 \\
1 & 0 & 6 & 0 & 1 & 0 & 0 & 18 \\
0 & 1 & 3 & 0 & 0 & 1 & 0 & 9 \\
0 & 3 & 5 & 0 & 1 & 0 & 0 & 15 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 4 & 8 & 0 & 8 & 0 & 7 & 23 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 3 & 2 & 1 & 1 & 0 & 0 & 7 \\
0 & 5 & 3 & 0 & 3 & 0 & 0 & 13
\end{array}
\tag{3.80}
$$

where each row $A_i$ corresponds to an inequality $A_i \cdot \boldsymbol{M} > 0$.

Operators mixing the different types of fermions together can also be considered, but these will be deferred to a more systematic analysis for now. The cases just described are especially simple because the UV cone is polyhedral. It is in this sense that the bounds derived here are well-described as generalisations of "positivity" constraints - they correspond to identifying the positive combinations of partial amplitudes implied by the optical theorem, which geometrises into the problem of finding the cone generated by a finite set of ERs. Examples of non-polyhedral cones, in which a section of the cone is described by smooth curved surface, will be discussed below. The parameters of the section correspond to the possible $S$-matrix transitions between distinct states that are permitted by the symmetries. No such transitions exist for the simple cases just discussed, but introducing new particles will usually spoil this.

To close this section, I emphasise that, for the restricted theories of a single species of fermion in the representations considered here, the listed constraints are complete. There are no further implications of the sum rule as stated in (3.8) for the structure of the EFT.

### 3.4.3 Flavour violation

The inelastic bounds can also be adapted to multiple flavours of particles unrelated by symmetries, of particular relevance to the SM. These were discussed in (25), where it was observed that flavour-violating fermion operators were bounded above by the flavour-conserving ones. I here show that this is a consequence of general inelastic unitarity bounds and derive general statements for the simplest cases.

Begin with right-handed leptons, the simplest states without non-Abelian symmetries, and allow for any number of flavours. As explained in Section 3.4.2, both angular momentum and hypercharge conservation ensure that, even with multiple flavours, am-

plitudes of the form $e_{Ri}, e_{Rj} \to e_{Rk}, e_{Rl}$ contain all possible independent interactions (other processes being related by $CPT$ or crossing). Labelling the UV coupling vectors as $\mathcal{M}^{ij\to X} = \boldsymbol{\alpha}^{ij}$ and $\mathcal{M}^{i\bar{j}\to X} = \boldsymbol{\beta}^{ij}$, then all elastic and inelastic amplitudes have the respective forms

$$M^{ijij} = |\boldsymbol{\alpha}^{ij}|^2 + |\boldsymbol{\beta}^{ij}|^2 \tag{3.81}$$

$$M^{ijkl} = \boldsymbol{\alpha}^{kl} \cdot \boldsymbol{\alpha}^{ij} + \boldsymbol{\beta}^{kl} \cdot \boldsymbol{\beta}^{il}. \tag{3.82}$$

This implies the existence of general bounds on flavour-violating transitions in any particular flavour basis

$$|M^{ijkl}| + |M^{ilkj}| \leq \sqrt{M^{ijij} M^{klkl}} + \sqrt{M^{ilil} M^{kjkj}}. \tag{3.83}$$

This result similarly holds for elastic scattering of right-handed leptons off any other species of fermion in the SM, as well as left-handed leptons off right-handed quarks, where the decompositions of the non-Abelian symmetries are trivial.

With left-handed leptons, the amplitudes must be decomposed into $SU(2)$ irreps. The partial amplitudes for the process $L_{Li}, L_{Lj} \to L_{Lk}, L_{Ll}$ are, using the projectors from Section 3.4.1,

$$M_{\mathbf{1}}^{ijkl} = \boldsymbol{m}_{\mathbf{1}}^{kl} \cdot \boldsymbol{m}_{\mathbf{1}}^{ij} - \frac{1}{2} \boldsymbol{m}_{\mathbf{1u}}^{kj} \cdot \boldsymbol{m}_{\mathbf{1u}}^{il} + \frac{3}{2} \boldsymbol{m}_{\mathbf{3u}}^{kj} \cdot \boldsymbol{m}_{\mathbf{3u}}^{il} \tag{3.84}$$

$$M_{\mathbf{3}}^{ijkl} = \boldsymbol{m}_{\mathbf{3}}^{kl} \cdot \boldsymbol{m}_{\mathbf{3}}^{ij} + \frac{1}{2} \boldsymbol{m}_{\mathbf{1u}}^{kj} \cdot \boldsymbol{m}_{\mathbf{1u}}^{il} + \frac{1}{2} \boldsymbol{m}_{\mathbf{3u}}^{kj} \cdot \boldsymbol{m}_{\mathbf{3u}}^{il}. \tag{3.85}$$

Here, the subscript $u$ has been introduced to distinguish the UV coupling vectors in the

$u$-channel term from the $s$-channel term. The bounds become

$$|M_{\mathbf{3}}^{ijkl}| + |M_{\mathbf{3}}^{ilkj}| \leq \sqrt{M_{\mathbf{3}}^{ijij} M_{\mathbf{3}}^{klkl}} + \sqrt{M_{\mathbf{3}}^{ilil} M_{\mathbf{3}}^{kjkj}} \tag{3.86}$$

$$\begin{aligned} |M_{\mathbf{3}}^{ijkl} + \frac{1}{3} M_{\mathbf{1}}^{ijkl}| + |M_{\mathbf{3}}^{ilkj} + \frac{1}{3} M_{\mathbf{1}}^{ilkj}| &\leq \sqrt{\left(M_{\mathbf{3}}^{ijij} + M_{\mathbf{1}}^{ijij}\right)\left(M_{\mathbf{3}}^{klkl} + M_{\mathbf{1}}^{klkl}\right)} \\ &+ \sqrt{\left(M_{\mathbf{3}}^{ilil} + M_{\mathbf{1}}^{ilil}\right)\left(M_{\mathbf{3}}^{kjkj} + M_{\mathbf{1}}^{kjkj}\right)}. \end{aligned} \tag{3.87}$$

The factors in the square roots on right-hand sides of these inequalities are positive, by analogous derivations to the $SU(3)$ case given at the beginning of Section 3.4.2. Because they have the same non-trivial non-Abelian symmetry structure, bounds for elastic scattering of left-handed leptons off left-handed quarks have the same form.

For right-handed quarks, analogous results can be derived but with $SU(3)$ partial amplitudes instead of $SU(2)$. In this case, the partial amplitudes have the form:

$$M_{\mathbf{6}}^{ijkl} = \boldsymbol{m}_{\mathbf{6}}^{kl} \cdot \boldsymbol{m}_{\mathbf{6}}^{ij} + \frac{1}{3} \boldsymbol{m}_{\mathbf{1}}^{kj} \cdot \boldsymbol{m}_{\mathbf{1}}^{il} + \frac{2}{3} \boldsymbol{m}_{\mathbf{8}}^{kj} \cdot \boldsymbol{m}_{\mathbf{8}}^{il} \tag{3.88}$$

$$M_{\bar{\mathbf{3}}}^{ijkl} = \boldsymbol{m}_{\bar{\mathbf{3}}}^{kl} \cdot \boldsymbol{m}_{\bar{\mathbf{3}}}^{ij} - \frac{1}{3} \boldsymbol{m}_{\mathbf{1}}^{kj} \cdot \boldsymbol{m}_{\mathbf{1}}^{il} + \frac{4}{3} \boldsymbol{m}_{\mathbf{8}}^{kj} \cdot \boldsymbol{m}_{\mathbf{8}}^{il}. \tag{3.89}$$

The bounds are

$$|M_{\mathbf{6}}^{ijkl}| + |M_{\mathbf{6}}^{ilkj}| \leq \sqrt{M_{\mathbf{6}}^{ijij} M_{\mathbf{6}}^{klkl}} + \sqrt{M_{\mathbf{6}}^{ilil} M_{\mathbf{6}}^{kjkj}} \tag{3.90}$$

$$\begin{aligned} |M_{\mathbf{6}}^{ijkl} + M_{\bar{\mathbf{3}}}^{ijkl}| + |M_{\mathbf{6}}^{ilkj} + M_{\bar{\mathbf{3}}}^{ilkj}| &\leq \sqrt{\left(M_{\mathbf{6}}^{ijij} + M_{\bar{\mathbf{3}}}^{ijij}\right)\left(M_{\mathbf{6}}^{klkl} + M_{\bar{\mathbf{3}}}^{klkl}\right)} \\ &+ \sqrt{\left(M_{\mathbf{6}}^{ilil} + M_{\bar{\mathbf{3}}}^{ilil}\right)\left(M_{\mathbf{6}}^{kj} + M_{\bar{\mathbf{3}}}^{kjkj}\right)}. \end{aligned} \tag{3.91}$$

This applies regardless of which species of right-handed quarks are identified with the pairs $i, k$ and $j, l$ (all that is important is that the amplitudes are elastic). Similarly, these results also apply to right-handed quarks scattering off left-handed quarks.

The bounds for left-handed quarks are more intricate because the convex cone de-

scribing the purely elastic, flavour-conserving amplitudes has a non-trivial (polyhedral) shape (in other words, more extremal rays than dimension). This issue also arises and is a general problem when there are multiple degenerate irreps of states in non-trivial symmetry representations. This was discussed in (28) in the case of the hypercharge boson coupling to $W$ bosons with parity symmetry respected, where restriction to the latter of is described by a such a non-trivial cone. This will be elaborated upon further below, but will here be left as an open problem. It is nevertheless clear that (3.11) can be directly applied to provide necessary upper bounds.

The general pattern described in the examples here is clear and would also apply to flavour-changing processes in which a fermion scatters off a boson. There are, of course, numerous other inelastic processes that can involve flavour violation that would likewise be bounded in more complicated ways (just as the underlying processes with flavour ignored). It should be again emphasised that (3.83) is, by itself, also not complete, and further constraints on the general three-flavour systems remain to be precisely determined.

## 3.5 Bounds with Helicity

In this section, the residual rotational invariance about the beam axis will be treated as a global symmetry in a similar way to the internal symmetries described above. This will allow bounds to be placed on theories with spinning particles in which there is a transfer of angular momentum.

### 3.5.1 Rotational symmetry

In the limit of exactly forward scattering, the rotational symmetry about the beam axis is preserved. This is an additional symmetry that can be managed just as for the

internal symmetries discussed above. In the center-of-mass frame, call the direction of particle 1 the $z$-direction, with respect to which all spin projections will be quantised. It is however natural to label the external states by helicity, or stated equivalently, by their little group symmetry of rotations about their momenta. There are two equivalent options for describing this: states of definite helicity (as was done in Section 3.4.2) and $SO(2)$ vectors, which were employed in (29), (28). Helicity eigenstates will be predominantly used in the following examples, in which case the angular momentum along the beam axis is treated as a $U(1)$ charge in a similar way to the fermion and scalar examples previously. However, as part of the simple illustrative example below in Section 3.5.2 of identical spinning particles, I will compare this with an analysis of the sum rule in $SO(2)$ form. This subsection will be devoted to deriving the relevant projectors and crossing relations required specifically for this case and addressing issues related to parity-violation. The more general results necessary for implementing the rotational symmetry in $SO(2)$ form will be given in the Appendix. The sum rules for the examples in Sections 3.5.4 and 3.5.5 will also be given there in $SO(2)$ form for comparison.

Call $h_i$ the magnitude of the helicity of particle $i$. The polarisation of particle $i$ may be represented equivalently by $\mathbf{2}_{\pm h_i}$ vectors, where a $\mathbf{2}_z$ vector responds to a spatial rotation of angle $\phi$ about the beam axis by a rotation by angle $z\phi$. Because the helicity quantisation axis of each particle is opposite, the rotational symmetry should act oppositely on each particle's little group indices, so it is natural to represent the polarisation of particle 1 by a $\mathbf{2}_{h_1}$ vector and particle 2 by a $\mathbf{2}_{-h_2}$ vector (and likewise for the outgoing states). The general relationship between the states in tensors of this form and helicity eigenstates $\{|h\rangle, |-h\rangle\}$ is

$$\mathbf{2}_{\pm h} \sim \frac{1}{\sqrt{2}} \begin{bmatrix} |h\rangle + |-h\rangle \\ \mp i\left(|h\rangle - |-h\rangle\right) \end{bmatrix}. \tag{3.92}$$

While there are two components to a $SO(2)$ vector, the Wigner-Eckart theorem does not require transitions between the two component states to be related by a symmetry transformation. Instead, only the $U(1)$ charges (in this case, spin projection) need be conserved. That the positive and negative charge irreps can have different partial amplitudes is an expression of the possibility of charge conjugation or parity violation. Separate projectors $P_\pm$ are therefore needed for each distinct charge configuration. It is possible to define projectors $P_P = P_+ + P_-$ and $P_{\not P} = P_+ - P_-$ corresponding to $P$ symmetric and violating transitions (these are normalised so that $(P_P)^* P_P = (P_{\not P})^* P_{\not P} = 2$). In many simple examples, such as those already discussed in Section 3.4.2, $CPT$ is sufficient to accidentally rule-out $(C)P$-violating transitions.

Now, consider the product of two states of helicities $h_1$ and $h_2$. The Clebsch-Gordan coefficients are

$$C^{ij}_{h_1+h_2} = \frac{1}{2}P^{ij} - \frac{i}{2}S^{ij} \qquad C^{ij}_{h_1-h_2} = \frac{1}{2}\delta^{ij} + \frac{i}{2}\epsilon^{ij} \tag{3.93}$$

and $C^{ij}_{-(h_1 \pm h_2)} = \left(C^{ij}_{h_1 \pm h_2}\right)^*$. The subscripts here denote $J_z$ eigenstate, while the superscript indices are $SO(2)$ $\mathbf{2}_{h_1}$ and $\mathbf{2}_{-h_2}$ components respectively. The symbol $P^{ij}$ is defined as having values $P^{11} = -P^{22} = 1$ and $P^{12} = P^{21} = 0$, while the symbol $S^{ij}$ is defined as having components $S^{12} = S^{21} = 1$ and $S^{11} = S^{22} = 0$. Note that here and throughout, as the four particle's little group indices are in altogether different representations, use of $\delta$ and $\epsilon$ is purely symbolic - these are not to be interpreted as invariant tensors. The helicities identified with the $i$ and $j$ indices are also necessary to uniquely specify the Clebsch-Gordan coefficients, but have been omitted from the notation here to avoid clutter, although there are several examples in the appendix where they must be kept track of.

For the special, yet prevalent case in which the particles have equal (non-zero) helicity

$h_1 = h_2 = h$, the $\mathbf{2}_{h_1-h_2}$ representation instead decomposes into two degenerate singlets. A basis for these will be chosen here to be labelled $A$ and $B$, where

$$|0\rangle_A = \frac{1}{\sqrt{2}}\left(|h\rangle|h\rangle + |-h\rangle|-h\rangle\right)$$
$$|0\rangle_B = \frac{1}{\sqrt{2}}\left(|h\rangle|h\rangle - |-h\rangle|-h\rangle\right). \tag{3.94}$$

These correspond to the standard $SO(2)$ components of the $\mathbf{2}_{h_1-h_2}$ vectors above. Note that the helicity labels denote spin numbers along opposite quantisation axes. Likewise, these states have Clebsch-Gordan coefficients

$$C_A^{ij} = \frac{1}{\sqrt{2}}\delta^{ij} \qquad C_B^{ij} = -\frac{i}{\sqrt{2}}\epsilon^{ij}. \tag{3.95}$$

Note that implicit in this discussion has been a particular phase convention in which eigenstates of helicity of particle 2 are directly equated with eigenstates of $J_z$, the rotation generator about the beam axis. In the present context, this has the further simplifying implication that the polarisations can be all chosen to be real (or, more precisely, their spinorial representations) and that the action of $P$ on the states does not produce a momentum-dependent phase. See Appendices C and I of (51) for more details, as well as (52).

For identical particles, states $A$ and $B$ are therefore $P$ eigenstates with opposite eigenvalues, once (anti-)symmetrisation is accounted for. Because fermion pairs have an intrinsic odd P phase, state $A$ is $P$ even for both bosons and fermions and state $B$ is $P$ odd. Transitions between these states in the $S$-matrix are prohibited if $P$ is conserved.

The next step is to find the projectors into which each spinning amplitude decomposes.

When $h_1 = h_2 = h_3 = h_4 = h \neq 0$, the projectors for $m_z = \pm 2h$ are

$$P^{ijkl}_{\not{P}2h} = \frac{i}{2}\left(\delta^{ik}\epsilon^{jl} + \delta^{jl}\epsilon^{ik}\right) \tag{3.96}$$

$$P^{ijkl}_{P\,2h} = \frac{1}{2}\left(\delta^{ik}\delta^{jl} + \delta^{il}\delta^{jk} - \delta^{ij}\delta^{kl}\right), \tag{3.97}$$

while for singlet states they are

$$P^{ijkl}_{AA} = \frac{1}{2}\delta^{ij}\delta^{kl} \tag{3.98}$$

$$P^{ijkl}_{BB} = \frac{1}{2}\left(\delta^{ik}\delta^{jl} - \delta^{il}\delta^{jk}\right) \tag{3.99}$$

$$P^{ijkl}_{AB} = \frac{i}{2}\delta^{ij}\epsilon^{kl} \tag{3.100}$$

$$P^{ijkl}_{BA} = \frac{-i}{2}\epsilon^{ij}\delta^{kl}. \tag{3.101}$$

The projectors in the $u$-channel may be decomposed as

$$P^{ilkj}_{P2h} = P^{ijkl}_{AA} + P^{ijkl}_{BB} \tag{3.102}$$

$$P^{ilkj}_{AA} = \frac{1}{2}\left(P^{ijkl}_{P2h} + P^{ijkl}_{AA} - P^{ijkl}_{BB}\right) \tag{3.103}$$

$$P^{ilkj}_{BB} = \frac{1}{2}\left(P^{ijkl}_{P2h} - P^{ijkl}_{AA} + P^{ijkl}_{BB}\right) \tag{3.104}$$

$$P^{ilkj}_{\not{P}2h} = -P^{ijkl}_{AB} - P^{ijkl}_{BA} \tag{3.105}$$

$$P^{ilkj}_{AB} = \frac{1}{2}\left(P^{ijkl}_{AB} - P^{ijkl}_{BA} - P^{ijkl}_{\not{P}2h}\right) \tag{3.106}$$

$$P^{ilkj}_{BA} = \frac{1}{2}\left(-P^{ijkl}_{AB} + P^{ijkl}_{BA} - P^{ijkl}_{\not{P}2h}\right). \tag{3.107}$$

### 3.5.2 Simple example: identical spinning particles

The complete set of dimension 8 positivity theorems for pure photon operators can be easily derived from requiring positivity of forward scattering of linearly polarised photons,

with polarisations inclined by some relative angle tuned to give an optimal constraint that is a function of the Wilson-coefficients (24). Here, I will present an alternative derivation directly from inspection of the sum rule. This is of particular educational value, as it provides a simple illustrates of several intricacies that can arise in the organisation of the symmetry structure of the sum rule. A discussion of these issues will also be both useful and necessary for further applications to theories of spinning particles. I will then present another derivation of the same results using the convex cone picture. While more complicated, this will again provide a simple archetypal example of a non-polyhedral cone. The arguments presented here applies generally for interactions of four identical particles of any non-zero helicity. While dimension 8 level is assumed here (as everywhere else), a near identical argument applies to any mass dimension $4n$ as well, which would be applicable for analogous results for gravitons assuming that the results of Section 3.2.1 continue to be valid.

Most of the argument is already complete given the projectors above. The four-particle amplitude can be decomposed into terms given by the projectors. The helicity violating partial amplitude is, expressed in terms of amplitudes between helicity eigenstates, $A_{\not{P}2h} = \frac{1}{2}(P^{ijkl}_{\not{P}2h})^* A^{ijkl} = \frac{1}{2}\left(A^{+-+-} - A^{-+-+}\right)$. In this form, it is clear that $CPT$ implies that this vanishes, as $CPT$ equates the two forward helicity amplitudes (as commented above, possible phases that may arise away from the forward limit are conventional and can be eliminated). The helicity-conserving amplitude is therefore time-reversal and parity-conserving. However, the crossing relations (3.106) and (3.107) would appear to generate this partial amplitude from crossing $A_{AB}$ and $A_{BA}$. This is avoided if the singlet partial amplitudes obey

$$A_{BA}(s) = -A_{AB}(s) = (A_{AB}(s))^*, \tag{3.108}$$

where the second equality uses (3.3) and assumes that $s$ is real (that the amplitude is evaluated away from the threshold singularities). In other words, $A_{AB}$, and hence $M_{AB}$, is purely imaginary. This constraint also saves the parity-violating singlet amplitudes from the Wigner-Eckart theorem. Without $P^{ijkl}_{\not{2}h}$, the tensors $P^{ilkj}_{AB}$ and $P^{ilkj}_{BA}$ cannot be individually decomposed into the $s$-channel projectors. However,

$$P^{ilkj}_{AB} - P^{ilkj}_{BA} = P^{ijkl}_{AB} - P^{ijkl}_{BA}, \tag{3.109}$$

as a result of the identity (3.179). The relation (3.108) thus ensures that it is only this crossing-consistent combination of parity-violating projectors that appear in the amplitude.

The $s$-channel term in the sum rule is

$$M^{ijkl}_s = |\boldsymbol{m}_2|^2 P^{ijkl}_2 + |\boldsymbol{m}_A|^2 P^{ijkl}_{AA} + |\boldsymbol{m}_B|^2 P^{ijkl}_{BB} + \boldsymbol{m}_A \cdot \boldsymbol{m}_B P^{ijkl}_{AB} + \boldsymbol{m}_B \cdot \boldsymbol{m}_A P^{ijkl}_{BA}. \tag{3.110}$$

Adding the $u$-channel crossed term gives

$$\begin{aligned} M^{ijkl} = &\left( |\boldsymbol{m}_2|^2 + \frac{1}{2}\left(|\boldsymbol{m}_A|^2 + |\boldsymbol{m}_C|^2\right)\right) P^{ijkl}_2 + \frac{1}{2}\left(2|\boldsymbol{m}_2|^2 + 3|\boldsymbol{m}_A|^2 - |\boldsymbol{m}_B|^2\right) P^{ijkl}_{AA} \\ &+ \frac{1}{2}\left(2|\boldsymbol{m}_2|^2 - |\boldsymbol{m}_A|^2 + 3|\boldsymbol{m}_B|^2\right) P^{ijkl}_{BB} + 2\boldsymbol{m}_A \cdot \boldsymbol{m}_B \left(P^{ijkl}_{AB} - P^{ijkl}_{BA}\right), \end{aligned} \tag{3.111}$$

where the crossing relation (3.108) implies that $\boldsymbol{m}_A \cdot \boldsymbol{m}_B$ is imaginary. This is to be expected because this transition amplitude is $P$-violating. The term parameterised by the vector $\boldsymbol{m}_2$ is redundant, so can be ignored.

Clearly $M_2 \geq 0$ and implies that the coefficient of the helicity-conserving amplitudes is positive. A second inequality can then be determined by finding a relation between

partial amplitudes in (3.111). The optimal constraint follows from $|\boldsymbol{m}_A \cdot \boldsymbol{m}_B| \leq |\boldsymbol{m}_A||\boldsymbol{m}_B|$ which, on the LHS, implies that

$$|M_{AB}| \leq \frac{1}{2}\sqrt{16M_2^2 - |M_{AA} - M_{BB}|^2}, \tag{3.112}$$

where $M_R$ denotes twice differentiated forward partial amplitude corresponding to representation $R$. This represents an upper bound on the size of the $P$-violating transitions determined from the $P$-conserving ones.

The partial amplitudes can be converted into amplitudes between helicity eigenstates. These are

$$M_2 = M(+,- \to +,-) \tag{3.113}$$

$$M_{AA,BB} = M(+,+ \to +,+) \pm \frac{1}{2}\left((M(+,+ \to -,-) + M(-,- \to +,+))\right) \tag{3.114}$$

$$M_{AB} = -\frac{1}{2}\left(M(+,+ \to -,-) - M(-,- \to +,+)\right), \tag{3.115}$$

where the $(+)$ corresponds to $AA$ and $(-)$ to $BB$. $CPT$ has been invoked for simplification. After some rearrangement, the constraints can be re-expressed as

$$M^{+-+-} > \frac{1}{4}\sqrt{|M^{++--} - M^{--++}|^2 + |M^{++--} + M^{--++}|^2}. \tag{3.116}$$

As mentioned in Section 3.5.1, with the phase conventions chosen here, $A(+,+ \to -,-) = A(-,- \to +,+)$ if $P$ is conserved. The $M_{AB}$ partial amplitudes (in the context of $W$-bosons) were not included in (28), where $P$-symmetry was assumed. In this case, the weaker bound on the $P$-conserving helicity violating interactions alone can be derived directly from positivity of the combinations of diagonal partial amplitudes $\frac{1}{2}\left(M_{AA} + M_{BB}\right) \pm \frac{1}{4}\left(M_{AA} - M_{BB}\right)$.

| | $+-$ | $++$ | $--$ | $-+$ |
|---|---|---|---|---|
| $+-$ | $\lvert\boldsymbol{m}^{+-}\rvert^2+\lvert\boldsymbol{m}^{++}\rvert^2$ | 0 | 0 | 0 |
| $++$ | 0 | $\lvert\boldsymbol{m}^{++}\rvert^2+\lvert\boldsymbol{m}^{+-}\rvert^2$ | $2\boldsymbol{m}^{--}\cdot\boldsymbol{m}^{++}$ | 0 |
| $--$ | 0 | $2\boldsymbol{m}^{++}\cdot\boldsymbol{m}^{--}$ | $\lvert\boldsymbol{m}^{--}\rvert^2+\lvert\boldsymbol{m}^{+-}\rvert^2$ | 0 |
| $-+$ | 0 | 0 | 0 | $\lvert\boldsymbol{m}^{+-}\rvert^2+\lvert\boldsymbol{m}^{--}\rvert^2$ |

Table 3.2: Sum rule for photons.

It is possible to perform the above analysis more directly with helicity eigenstates instead so that both helicity-violating interactions are treated symmetrically, as would be expected from the structure of the bound. In this case, the sum rule would have the same structure as the complex scalar example in Section 3.3.1, but with the single charge-violating amplitudes prohibited. This is given in Table 3.2. The helicity-violating amplitudes are immediately manifest and the bound in the form of (3.116) follows directly from bounding the off-diagonal entry, the real and imaginary parts of which correspond to $P$-conserving and violating interactions respectively.

Specialising now to photons, the general effective action up to dim-8 is

$$\mathcal{L}_{EFT_8} = \frac{c}{16\Lambda^4}\left(\left(F^2\right)^2+(F\tilde{F})^2\right)+\frac{d}{32\Lambda^4}\left(\left(F^2\right)^2-(F\tilde{F})^2\right)+\frac{e}{16\Lambda^4}F^2\left(F\tilde{F}\right). \quad (3.117)$$

The operator basis has been selected to match onto specific tree-level 4-leg amplitudes between helicity eigenstates. The first operator is helicity preserving, the others are helicity-violating, with the coefficient $e$ being $P$ and $CP$ violating and providing the imaginary part of the coupling in the corresponding amplitudes. Evaluating the LHS entries at tree-level, the constraints reduce to

$$c > \frac{1}{2}\sqrt{d^2+e^2}. \quad (3.118)$$

However, loops of scalar particles mediated by dim-6 operators of the form $F^2\phi^2$

also contribute to the dim-8 order four photon amplitudes. If the scalar is complex, the possible operators are

$$\begin{aligned}\mathcal{L}_{EFT_6} = &\frac{a}{\Lambda^2}F^2\phi^2 + \frac{b}{\Lambda^2}F\tilde{F}\phi^2 + \text{conj} \\ &+ \frac{\tilde{a}}{\Lambda^2}F^2\phi\phi^\dagger + \frac{\tilde{b}}{\Lambda^2}F\tilde{F}\phi\phi^\dagger. \end{aligned} \tag{3.119}$$

where $a, b \in \mathbb{C}$ and $\tilde{a}, \tilde{b} \in \mathbb{R}$. In a supersymmetric theory, discussed more below, $\tilde{a}$ and $\tilde{b}$ are prohibited if the scalars are the same, while $b = ia$. On-shell, this is the statement that the only permitted contact interactions induce amplitudes $\mathcal{A}(\gamma^+, \gamma^+, \overline{\phi}, \overline{\phi})$ and $\mathcal{A}(\gamma^-, \gamma^-, \phi, \phi)$. In this case, helicity charge of the photons may be extended into a conserved charge also carried by the scalars, which is a statement of electric-magnetic duality (53). Alternatively, $b = -ia$ is also compatible if the identification of particle and antiparticle is reversed. Including these contributions to the amplitudes, the constraints become:

$$\begin{aligned}2c + \frac{8}{(4\pi)^2}\left(\tilde{a}^2 + \tilde{b}^2 + 2|a-ib|^2 + 2|a+ib|^2\right)\left(2 + \log\left(\frac{\mu^2}{\Lambda^2}\right)\right) \\ > \left|d + ie + \frac{8}{(4\pi)^2}\left((\tilde{a}-i\tilde{b})^2 + 4(a-ib)(a^*-ib^*)\right)\left(2 + \log\left(\frac{\mu^2}{\Lambda^2}\right)\right)\right|. \end{aligned} \tag{3.120}$$

Just as in the scalar example discussed in Section 3.3.1, the dim-6 operators strengthen the lower bound on $c$ (absorbing into it the rational part of the loop correction), at least assuming that there is little change in the size of the expression on the RHS of the inequality.

It is possibly enlightening to consider the effect of each term in isolation, with all others set to zero. In order to maintain consistency with positivity, $c$ will remain active so that the dim-6 corrections can be consistently negative, while all $d$ and $e$ will be chosen to cancel the rational terms generated from the dim-6 operators. Activating only $\tilde{a}$ and

$\tilde{b}$, then both terms on the left and right side of the inequality are equal in magnitude. If both terms were positive, then these would simply account for each other on each side of the sum rule and saturate it to give no information. However, as the logarithm is negative, these terms actually reinforce each other and strengthen the lower bound on $c$ beyond 0, the degree to which depending on the size of the hierarchy between IR and UV scales. Next, if only $a$ and $b$ are active, then the terms on the LHS of the inequality are always greater in magnitude than those on the RHS although, as before, both strengthen the tree-level bound. Interestingly, for a give coupling $a$, the weakest contribution to the constraint is made for the symmetry-enhanced choice $b = \pm ia$. This is because the on-shell amplitudes into which the cut scalar loop factorise would be prohibited for the helicity-violating configurations, as well as for one of the two possible contributions to the helicity-conserving case. This is typical of the suppression of RG evolution caused by the enhancement of symmetries.

There are also other potential contributions to the dim-8 order four photon amplitudes, such as from a fermion box of three-particle dim-5 operators, that would appear on the LHS. A more thorough analysis of the way that lower-dim operators affect the constraints will be left for another work.

The bound (3.112) is therefore the statement that the helicity-violating amplitudes must be smaller than the helicity-conserving ones. For photons, this is a leading-order statement of the hierarchy in coupling strengths of symmetry preserving and symmetry violating interactions. Helicity conservation corresponds to electric-magnetic duality and is also selected by supersymmetry, as will be explored further below. This point of view also "explains" the observation of (24) of the consequential suppression of the $P$ and $T$ violation in the vector boson EFT. These discrete symmetries can only be violated by operators that mediate helicity-violating interactions, or, more generally, off-diagonal, inelastic $S$-matrix entries. Because the sum rule curbs the size of these interactions,

it consequently also places fundamental restrictions on the size of $T$-violation. Note that, while it is possible to perform a field redefinition (or "duality rotation") in the effective action (3.117) to remove the $T$-violating term (and transfer it into the coupling of the photon to sources), the combination $d^2+e^2$ remains invariant and the constraint is unchanged. This is the reason that the two couplings must necessarily appear added in quadrature. On-shell, this is reflected in the overall phase ambiguity of the amplitudes. This similarly applies to the complex $a$ coefficient in the case that $b = \pm ia$ and $\tilde{a} = \tilde{b} = 0$.

The bound (3.112) can be alternatively derived from the convex cone picture. Beginning with the sum rule (3.111) (with $\boldsymbol{m}_2 = 0$), the first step is to find the PERs. These may be determined as the independent contributions with the factorised form $\boldsymbol{m} \cdot \boldsymbol{m}$ in the $s$-channel before the crossed $u$-channel is added. Choose as independent coordinates the partial amplitudes $\boldsymbol{x} = (M_{AA}, M_{BB}, \Im(M_{AB}))$, where $M_{AB}$ is purely imaginary, so represents only one real dimension. Each PER corresponds to the contribution from a single UV state, so only a single component of the complex vectors in the sum rule need be chosen. These will be labelled as $\beta$ and $\gamma$. Because of the crossed-amplitude, there is only a single ray structure that combines all terms generated by both parameters. As the ray is only defined modulo positive real factors, the coordinates may be rescaled by a factor of $2/|\beta|^2$ to give $\boldsymbol{e}(r) = (3 - r^2, -1 + 3r^2, 4r)$, where $r = -i\gamma/\beta \in \mathbb{R}$. It can be verified that this is extremal for all $r$. This family of ERs is effectively a $2d$ surface in a $3d$ space parameterised by 2 real parameters (including a positive real parameter rescaling the ray). The cone itself has parabolic sections. The face of the cone is defined by the normal $\boldsymbol{n}$ to the surface, which (up to an arbitrary scale) has components $n_i = \epsilon_{ijk} e^j(r) \frac{\partial e^k}{\partial r} = (-1 - 3r^2, -3 - r^2, 4r)$. The EFT must induce amplitudes that lie inside the cone, so this implies that $\boldsymbol{x} \cdot \boldsymbol{n}(r) < 0$ for all $r \in \mathbb{R}$. Imposing this latter condition implies that $3M_{AA} + M_{BB} > 0$, $M_{AA} + 3M_{BB} > 0$ and $|M_{AB}| < \frac{1}{2}\sqrt{(3M_{AA} + M_{BB})(M_{AA} + 3M_{BB})}$, which is just a restatement of (3.112) given

the relation $M_2 = \frac{1}{2}(M_{AA} + M_{BB})$.

### 3.5.3 Non-Abelian vector bosons

I here make some comments about extending the above analysis to scattering of (massless) $W$-bosons in which $CP$-violation is permitted.

The $W$-bosons have helicity and adjoint $SU(2)$ indices. The projectors that span the amplitude are given by products of adjoint isospin and helicity projectors, of which there are 15. Denoting by $\boldsymbol{\alpha}_{I,m}$ the vector of UV couplings in the RHS of the sum rule for isospin irrep $I$ and $J_z$ irrep $m$, then the sum rule can be expressed as

$$M^{ai,bj,ck,dl} = \sum_{I,m,n} \boldsymbol{\alpha}_{I,m} \cdot \boldsymbol{\alpha}_{I,n} \left( P_I^{abcd} P_{mn}^{ijkl} + P_I^{adcb} P_{mn}^{ilkj} \right). \tag{3.121}$$

The crossed projectors are then decomposed using the relations stated in the previous sections to derive the partial amplitudes as a function of the complex vectors $\boldsymbol{\alpha}_{I,m}$. Similarly to photons, $CPT$ implies that the helicity-conserving amplitudes are parity symmetric, so the relations (3.108) hold for each isospin partial amplitude such that $\boldsymbol{\alpha}_{I,A} \cdot \boldsymbol{\alpha}_{I,B} = -\boldsymbol{\alpha}_{I,B} \cdot \boldsymbol{\alpha}_{I,A}$. The $\boldsymbol{\alpha}_{I,m} \cdot \boldsymbol{\alpha}_{I,n}$ coefficients determine the space of partial (forward) amplitudes.

In the simple examples above, the vectors parameterising the sum rule were loosely in correspondance with the elastic partial amplitudes. An expression of the form $|\boldsymbol{\alpha}|^2$ could be simply translated into a partial amplitude in order to determine the bounds. However, in the $W$ theory (and any other with sufficiently many degrees of freedom), these vectors exceed the number of independent partial amplitudes. If only diagonal $S$-matrix transitions were permitted, this would correspond to a non-trivial polyhedral cone with more facets than dimension. Each point in the cone may admit multiple decompositions into positive sums of extremal rays and different subsets of rays span

different regions of the cone, complicating the simple inspection arguments used to derive bounds in the examples above. Geometrically, the problem becomes that of performing vertex enumeration for non-polyhedral cones, or more simply finding the curved facets bounding the cone. Some ideas for addressing this were described in (28), but for now this will be left for future work. The solution to this problem represents the next step toward bootstrapping constraints on realistic EFTs such as Standard Model EFT.

### 3.5.4 Two chiral fermions

A simple expansion of the previous example in Section 3.5.2 is given by introducing a second distinct particle with the same helicity. I will commit to assuming that both particles are chiral fermions, $\psi$ and $\lambda$, because this will be of interest later. However, the conclusions are more general. It will be additionally assumed for simplicity that each is charged under its own $\mathbb{Z}_2$ symmetry so that they can only be destroyed or created in pairs.

The elastic amplitudes are affected by the same constraints derived above in Section 3.5.2. The sum rule for these are reproduced here in Table 3.3. Entries below the main diagonal have been omitted as they are simply related by Hermiticity of the matrix. Some of the vectors in the last block are $Y$-rotated versions of counterparts in the third block and satisfy $|\boldsymbol{m}^{\psi^+\lambda^+}| = |\boldsymbol{m}^{\lambda^+\psi^+}|$, $|\boldsymbol{m}^{\psi^+\lambda^-}| = |\boldsymbol{m}^{\lambda^-\psi^+}|$.

| | $\lambda^+\lambda^-$ | $\lambda^+\lambda^+$ | $\lambda^-\lambda^-$ | $\lambda^-\lambda^+$ |
|---|---|---|---|---|
| $\lambda^+\lambda^-$ | $\lvert\boldsymbol{m}^{\lambda^+\lambda^-}\rvert^2+\lvert\boldsymbol{m}^{\lambda^+\lambda^+}\rvert^2$ | $0$ | $0$ | $0$ |
| $\lambda^+\lambda^+$ | . | $\lvert\boldsymbol{m}^{\lambda^+\lambda^-}\rvert^2+\lvert\boldsymbol{m}^{\lambda^+\lambda^+}\rvert^2$ | $2\boldsymbol{m}^{\lambda^-\lambda^-}\cdot\boldsymbol{m}^{\lambda^+\lambda^+}$ | $0$ |
| $\lambda^-\lambda^-$ | . | . | $\lvert\boldsymbol{m}^{\lambda^-\lambda^-}\rvert^2+\lvert\boldsymbol{m}^{\lambda^+\lambda^-}\rvert^2$ | $0$ |
| $\lambda^-\lambda^+$ | . | . | . | $\lvert\boldsymbol{m}^{\lambda^-\lambda^-}\rvert^2+\lvert\boldsymbol{m}^{\lambda^+\lambda^-}\rvert^2$ |
| | $\psi^+\psi^-$ | $\psi^+\psi^+$ | $\psi^-\psi^-$ | $\psi^-\psi^+$ |
| $\psi^+\psi^-$ | $\lvert\boldsymbol{m}^{\psi^+\psi^-}\rvert^2+\lvert\boldsymbol{m}^{\psi^+\psi^+}\rvert^2$ | $0$ | $0$ | $0$ |
| $\psi^+\psi^+$ | . | $\lvert\boldsymbol{m}^{\psi^+\psi^-}\rvert^2+\lvert\boldsymbol{m}^{\psi^+\psi^+}\rvert^2$ | $2\boldsymbol{m}^{\psi^-\psi^-}\cdot\boldsymbol{m}^{\psi^+\psi^+}$ | $0$ |
| $\psi^-\psi^-$ | . | . | $\lvert\boldsymbol{m}^{\psi^-\psi^-}\rvert^2+\lvert\boldsymbol{m}^{\psi^+\psi^-}\rvert^2$ | $0$ |
| $\psi^-\psi^+$ | . | . | . | $\lvert\boldsymbol{m}^{\psi^-\psi^-}\rvert^2+\lvert\boldsymbol{m}^{\psi^-\psi^-}\rvert^2$ |
| | $\lambda^+\psi^-$ | $\lambda^+\psi^+$ | $\lambda^-\psi^-$ | $\lambda^-\psi^+$ |
| $\lambda^+\psi^-$ | $\lvert\boldsymbol{m}^{\lambda^+\psi^-}\rvert^2+\lvert\boldsymbol{m}^{\lambda^+\psi^+}\rvert^2$ | $0$ | $0$ | $0$ |
| $\lambda^+\psi^+$ | . | $\lvert\boldsymbol{m}^{\lambda^+\psi^-}\rvert^2+\lvert\boldsymbol{m}^{\lambda^+\psi^+}\rvert^2$ | $2\boldsymbol{m}^{\lambda^-\psi^-}\cdot\boldsymbol{m}^{\lambda^+\psi^+}$ | $0$ |
| $\lambda^-\psi^-$ | . | . | $\lvert\boldsymbol{m}^{\lambda^-\psi^-}\rvert^2+\lvert\boldsymbol{m}^{\lambda^-\psi^+}\rvert^2$ | $0$ |
| $\lambda^-\psi^+$ | . | . | . | $\lvert\boldsymbol{m}^{\lambda^-\psi^-}\rvert^2+\lvert\boldsymbol{m}^{\lambda^-\psi^+}\rvert^2$ |
| | $\psi^+\psi^-$ | $\psi^+\psi^+$ | $\psi^-\psi^-$ | $\psi^-\psi^+$ |
| $\lambda^+\lambda^-$ | $\boldsymbol{m}^{\psi^+\psi^-}\cdot\boldsymbol{m}^{\lambda^+\lambda^-}+\boldsymbol{m}^{\psi^+\lambda^+}\cdot\boldsymbol{m}^{\lambda^+\psi^+}$ | $0$ | $0$ | $0$ |
| $\lambda^+\lambda^+$ | $0$ | $\boldsymbol{m}^{\psi^+\psi^+}\cdot\boldsymbol{m}^{\lambda^+\lambda^+}+\boldsymbol{m}^{\psi^+\lambda^-}\cdot\boldsymbol{m}^{\lambda^+\psi^-}$ | $\boldsymbol{m}^{\psi^-\psi^-}\cdot\boldsymbol{m}^{\lambda^+\lambda^+}+\boldsymbol{m}^{\lambda^-\psi^-}\cdot\boldsymbol{m}^{\psi^+\lambda^+}$ | $0$ |
| $\lambda^-\lambda^-$ | $0$ | $\boldsymbol{m}^{\psi^+\psi^+}\cdot\boldsymbol{m}^{\lambda^-\lambda^-}+\boldsymbol{m}^{\psi^+\lambda^+}\cdot\boldsymbol{m}^{\lambda^-\psi^-}$ | $\boldsymbol{m}^{\psi^-\psi^-}\cdot\boldsymbol{m}^{\lambda^-\lambda^-}+\boldsymbol{m}^{\lambda^+\psi^-}\cdot\boldsymbol{m}^{\psi^+\lambda^-}$ | $0$ |
| $\lambda^-\lambda^+$ | $0$ | $0$ | $0$ | $\boldsymbol{m}^{\psi^+\psi^-}\cdot\boldsymbol{m}^{\lambda^+\lambda^-}+\boldsymbol{m}^{\lambda^+\psi^+}\cdot\boldsymbol{m}^{\psi^+\lambda^+}$ |

Table 3.3: Sum rule for two fermion theory.

Both $CPT$ and $Y$ can be invoked to reduce the number of independent vector magnitudes $|\boldsymbol{m}|$ to six. After applying the Schwarz and triangle inequalities to the off-diagonal amplitudes, these all have upper bounds of the form of either $|\boldsymbol{m}^{\lambda+\lambda-}||\boldsymbol{m}^{\psi+\psi-}|+|\boldsymbol{m}^{\lambda+\psi+}|^2$ or $|\boldsymbol{m}^{\psi^+\psi^+}||\boldsymbol{m}^{\lambda^+\lambda^+}|+|\boldsymbol{m}^{\lambda^+\psi^-}|^2$. Analogous bounds of the form of (3.11) are then given by

$$M_1 + M_2 \leq \sqrt{M^{\lambda^+\lambda^-\lambda^+\lambda^-}M^{\psi^+\psi^-\psi^+\psi^-}} + M^{\lambda^+\psi^-\lambda^+\psi^-}, \tag{3.122}$$

where

$$M_1 \in \{|M^{\lambda^+\lambda^+\psi^+\psi^+}|, |M^{\lambda^-\lambda^-\psi^-\psi^-}|, |M^{\lambda^+\lambda^+\psi^-\psi^-}|, |M^{\lambda^-\lambda^-\psi^+\psi^+}|\} \tag{3.123}$$

$$M_2 \in \{|M^{\lambda^+\lambda^-\psi^+\psi^-}|, |M^{\lambda^-\lambda^+\psi^-\psi^+}|\} \tag{3.124}$$

are, respectively, any of the $m_z = 0$ and $m_z = \pm 1$ amplitudes.

For reference, the general effective action for this theory has terms

$$\begin{aligned}\mathcal{L}_{EFT_6} = &\frac{f}{\Lambda^2}\psi\psi\psi^\dagger\psi^\dagger + \frac{g}{\Lambda^2}\lambda\lambda\lambda^\dagger\lambda^\dagger + \frac{h}{\Lambda^2}\psi\lambda\psi^\dagger\lambda^\dagger \\ &+ \frac{k}{\Lambda^2}\psi\psi\lambda^\dagger\lambda^\dagger + \frac{\tilde{k}}{\Lambda^2}\psi\psi\lambda\lambda + \text{conj.}\end{aligned} \tag{3.125}$$

and

$$\begin{aligned}\mathcal{L}_{EFT_8} =&\frac{a}{2\Lambda^4}\psi\psi\partial^2\left(\psi^\dagger\psi^\dagger\right) + \frac{b}{2\Lambda^4}\lambda\lambda\partial^2\left(\lambda^\dagger\lambda^\dagger\right) + \frac{c}{\Lambda^4}\psi\lambda\partial^2\left(\psi^\dagger\lambda^\dagger\right) + \frac{d}{\Lambda^4}\partial\psi\lambda\cdot\partial\psi^\dagger\lambda^\dagger \\ &+ \Big(\frac{\tilde{a}}{2\Lambda^4}\psi\psi\partial^2\left(\psi\psi\right) + \frac{\tilde{b}}{2\Lambda^4}\lambda\lambda\partial^2\left(\lambda\lambda\right) + \frac{\tilde{c}}{\Lambda^4}\psi\lambda\partial^2\left(\psi\lambda\right) \\ &+ \frac{\tilde{d}}{\Lambda^4}\psi\psi\partial^2\left(\lambda\lambda\right) + \frac{e}{\Lambda^4}\psi\psi\partial^2\left(\lambda^\dagger\lambda^\dagger\right) + \text{conj}\Big).\end{aligned} \tag{3.126}$$

The couplings $f, g, h, a, b, c, d \in \mathbb{R}$ and $k, \tilde{k}, \tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}, e \in \mathbb{C}$. The operators with Wilson

coefficients denoted with a tilde mediate helicity-violating interactions. Consistency with supersymmetry will be elaborated on further below (see Section 3.6). The bounds on amplitudes are readily converted into bounds on dim-8 Wilson coefficients at tree-level after identifying them with the particular transitions between helicity eigenstates and channels.

### 3.5.5 Bounds on simple theories of spinning particles

Similar analyses can be performed on systems with particles of different helicities. Here, simple cases of mixed scalar, fermion and vector amplitudes will be discussed in order to derive new constraints.

To begin with, the system will be restricted to consisting of a real scalar, a photon and a chiral fermion. The full table has dimensions $25 \times 25$ when external helicity eigenstates are chosen as a basis, but most of the entries vanish either altogether by conservation of angular momentum or, specifically at dimension 8, by incompatibility of Lorentz invariance and dimensional analysis (see discussion below in Section 3.6.2). For this reason, only the (few) important entries containing new information will be quoted here. The full sum rule in $SO(2)$ form is given in the Appendix.

Restricting entirely to the bosons to begin with, the entries relevant for photons are given in Table 3.2 in the section above, while the others of relevance are:

$$M^{\phi\phi\phi\phi} = 2|\boldsymbol{m}^{\phi\phi}|^2 \tag{3.127}$$

$$M^{\gamma^+\phi\gamma^+\phi} = 2|\boldsymbol{m}^{\gamma^+\phi}|^2 \tag{3.128}$$

$$M^{\gamma^-\phi\gamma^-\phi} = 2|\boldsymbol{m}^{\gamma^-\phi}|^2 \tag{3.129}$$

$$M^{\phi\phi\gamma^+\gamma^+} = \boldsymbol{m}^{\gamma^+\gamma^+} \cdot \boldsymbol{m}^{\phi\phi} + \boldsymbol{m}^{\gamma^+\phi} \cdot \boldsymbol{m}^{\phi\gamma^-} \tag{3.130}$$

$$M^{\phi\phi\gamma^-\gamma^-} = \boldsymbol{m}^{\gamma^-\gamma^-} \cdot \boldsymbol{m}^{\phi\phi} + \boldsymbol{m}^{\phi\gamma^-} \cdot \boldsymbol{m}^{\gamma^+\phi} \tag{3.131}$$

(and others related by crossing). Simplification with $CPT$ and $Y$ (which also equate $|\boldsymbol{m}^{\gamma^+\phi}| = |\boldsymbol{m}^{\phi\gamma^-}|$ has been invoked. The bounds are identified as:

$$|M^{\phi\phi\gamma^+\gamma^+}|, |M^{\phi\phi\gamma^-\gamma^-}| \leq \frac{1}{2}M^{\phi\gamma^+\phi\gamma^+} + \sqrt{\frac{1}{2}M^{\phi\phi\phi\phi}M^{\gamma^+\gamma^-\gamma^+\gamma^-}}. \tag{3.132}$$

Unlike the helicity-violating four-vector operators, these bounds are not accessible by considering elastic forward scattering of a scalar with a linearly polarised vector. Superpositions of vectors and scalars are instead necessary. The bounds are also stronger by various factors of 2 compared to what would be anticipated from direct application of (3.11). This is because the identity of the scalars is crossing symmetric, which simplifies the sum rule. The analogous bounds with complex scalars, used below in Section 3.6.3, are weaker.

The other constraint arises for a mixed spin amplitude. The relevant entries are

$$M^{\phi\psi^+\phi\psi^+} = |\boldsymbol{m}^{\phi\psi^+}|^2 + |\boldsymbol{m}^{\phi\psi^-}|^2 \tag{3.133}$$

$$M^{\gamma^+\psi^+\gamma^+\psi^+} = |\boldsymbol{m}^{\gamma^+\psi^+}|^2 + |\boldsymbol{m}^{\gamma^+\psi^-}|^2 \tag{3.134}$$

$$M^{\phi\psi^-\gamma^+\psi^+} = 2\boldsymbol{m}^{\gamma^+\psi^+} \cdot \boldsymbol{m}^{\phi\psi^-} \tag{3.135}$$

$$M^{\phi\psi^+\gamma^-\psi^-} = 2\boldsymbol{m}^{\gamma^-\psi^-} \cdot \boldsymbol{m}^{\phi\psi^+}, \tag{3.136}$$

where $|\boldsymbol{m}^{\phi\psi^+}| = |\boldsymbol{m}^{\phi\psi^-}|$. The resulting inelastic constraint is

$$|M^{\phi\psi^-\gamma^+\psi^+}|, |M^{\phi\psi^+\gamma^-\psi^-}| \leq \sqrt{2M^{\phi\psi^+\phi\psi^+}M^{\psi^+\gamma^+\psi^+\gamma^+}}. \tag{3.137}$$

Along with the postivity constraints given in (16), this completes the causality bounds for the simple minimal, toy theories of spinning particles. Similar arguments can be used to adapt these to more complicated theories with more states. This will be partly done

in the supersymetric case below.

## 3.6 Supersymmetry

Having addressed the management of spin in the sum rule, it is natural to now extend this to supermultiplets. Supersymmetry unifies states of different spin and likewise their interactions. Supersymmetry is a consistent extension of the spacetime symmetry algebra, so should not affect conclusions drawn from the (rigid) causal structure of background flat Minkowski space. It is therefore expected that causality constraints on scattering of a particular set of component states should be shared by the other interactions related by supersymmetry. Precisely these connections will be explored in this section. For simplicity, attention will be restricted to EFTs with minimal particle content.

### 3.6.1 Superamplitudes

Superspaces at the level of the effective action are generally arduous and cumbersome to work with. As is very well appreciated, on-shell scattering amplitudes cut-through the off-shell baggage of the effective action, not only making computations substantially easier, but also clarifying the presence and action of symmetries that are either not manifest or are convoluted in the Lagrangian field theory. The "on-shell superspace", to be employed here, makes the super-Ward identities (SWIs) manifest as relations between scattering amplitudes - see e.g. (54) for review. Amplitudes between individual states in a multiplet are unified into superamplitudes. This makes transparent the relation between the unified effective interactions and their component operators without recourse to an off-shell superspace. The especially simple case of $2 \to 2$ scattering amplitudes, under discussion here, are highly constrained by fundamental principles. See (55) for numerous examples in supergravity.

The chiral superspace of (56), later used by (57), will be employed here, where the highest helicity state is selected as the Clifford vacuum for the representation. This determines the little group representation of the entire "superfield". The massless multiplets for $\mathcal{N} = 1$ theories are

$$\Phi^+ = \psi^+ + \eta\overline{\phi} \qquad \Phi^- = \phi + \eta\psi^-$$
$$V^+ = v^+ + \eta\lambda^+ \qquad V^- = \lambda^- + \eta v^-.$$

For $\mathcal{N} = 2$, they are

$$K = \chi^+ + \eta^A\phi_A - \frac{1}{2}\epsilon_{AB}\eta^A\eta^B\chi^- \qquad \overline{K} = \overline{\chi}^+ + \eta^A\overline{\phi}_A - \frac{1}{2}\epsilon_{AB}\eta^A\eta^B\overline{\chi}^-$$
$$V^+ = v^+ + \eta^A\lambda_A^+ - \frac{1}{2}\epsilon_{AB}\eta^A\eta^B\overline{\phi} \qquad V^- = \phi + \eta^A\lambda_A^- - \frac{1}{2}\epsilon_{AB}\eta^A\eta^B v^-.$$

See (57) for general explanation of notation. The multiplet $K$ is a half-hypermultiplet and is usually paired with a conjugate multiplet of antiparticles, $\overline{K}$. However, as both multiplets have identical helicity structure, it will not be important here to continue to distinguish between the two. The $\mathcal{N} = 4$ vector is defined in (54) and will not be reproduced here.

All external states defining the superamplitudes will be taken to be outgoing. This is the convention adopted in (54). However, it will be necessary to cross two of the states to be incoming. Crossing has been discussed in the present context in (16) and will be performed here on the component amplitudes. As mentioned in at the end of Section 3.2.1, the ordering of the superfields in the correlator from which the amplitude is derived is $\langle 0|4312|0\rangle$. This gives the order of the states and the Grassmann variables in the superamplitude, which determines the order in which the Grassmann derivatives should be applied to extract the components. Note that the Feynman rules for external antifermion legs include a factor of $-1$ that is frequently dropped (58), but is required

here. This implies that a single fermion leg must be accompanied by a factor of $-1$ when crossed, in addition to the usual rules of reversing the momentum and replacing the external polarisation.

### 3.6.2 Effective operators

It is of general interest to classify the effective contact interactions combined together under various degrees of supersymmetry. Here these will be systematically classified from dimension 5 to dimension 8 for operators inducing contact interactions between three or four particles. Again, the discussion will be restricted to helicities $h \leq 1$ (so no (super)gravity). Different species of multiplets with the same superspin will not be distinguished in order to emphasise the purely kinematical structure of the allowed interactions, but no assumptions will be made about permutation symmetries and internal quantum numbers (unrelated to the supersymmetry algebra).

Most of the interactions considered here will be four-particle contact interactions. These are severely constrained by consistency with dimensional analysis, little group representation, Lorentz invariance, locality and supersymmetry. The last condition is the requirement that the superamplitude depend on the Grassmann variables exactly through $\delta^{(2)}(Q^\dagger)$, while the former conditions demand that the amplitudes be polynomials in spinor bilinears with the required mass dimension and total helicity charge for each leg.

While supersymmetry unifies interactions, it can also prohibit them. A common reason for this is that the spectrum of effective interactions for higher spin particles is sparser than for lower spin particles, so not all lower spin interactions can be uniquely paired with a higher spin interaction.

**Dimension** 5

An anomalous magnetic dipole moment (MDM)-like operator for matter fermions is prohibited by supersymmetry (this is already prohibited by exchange antisymmetry if the fermions are identical). However, the axion/dilaton coupling is promoted to

$$\mathcal{A}(\Phi^+, V^+, V^+) \propto \frac{1}{\Lambda}\tilde{\delta}^{(1)}(Q)\,[23]\,, \tag{3.138}$$

which also contains a mixed MDM-like interaction between the matter fermion and the gaugino. This interaction can be further promoted to $\mathcal{N} = 2$ in the superamplitude

$$\mathcal{A}(V^+, V^+, V^+) \propto \frac{1}{\Lambda}\tilde{\delta}^{(2)}(Q), \tag{3.139}$$

which contains no further interactions. There are likewise conjugate superamplitudes between the corresponding anti-multiplets. Notably, the axion/dilaton cannot belong to a hypermultiplet.

The Weinberg operator (uniquely) supersymmetrises into itself:

$$\mathcal{A}(\Phi^+, \Phi^+, \Phi^+, \Phi^+) \propto \frac{1}{\Lambda}\delta^{(2)}(Q^\dagger)\frac{[12]}{\langle 34\rangle}. \tag{3.140}$$

The conjugate amplitude $\mathcal{A}(\Phi^-\Phi^-\Phi^-\Phi^-)$ is similar. Despite appearances, all spinor prefactors are equivalent. The operator is altogether incompatible with $\mathcal{N} \geq 2$.

**Dimension** 6

Cubic vector interactions are altogether prohibited by supersymmetry, so there are no 3-particle operators to consider.

Permissible $\mathcal{N}=1$ superamplitudes are

$$\mathcal{A}(\Phi^+,\Phi^+,\Phi^-,\Phi^-) \propto \frac{1}{\Lambda^2}\delta^{(2)}(Q^\dagger)\,[12] \tag{3.141}$$

$$\mathcal{A}(V^+,V^+,\Phi^+,\Phi^+) \propto \frac{1}{\Lambda^2}\delta^{(2)}(Q^\dagger)\frac{[12]^2}{\langle 34\rangle} \tag{3.142}$$

and analogous conjugates. The first superamplitude combines dimension-6 four fermion operators with scalar and mixed fermion-scalar operators. The superamplitude must be helicity conserving, so helicity-violating matter interactions are forbidden (such as those induced by the operators of the form $\psi\psi\partial^2\psi\psi$). For the scalar interactions, "helicity preserving" becomes charge preserving (each $\phi$ must be paired with a $\phi^\dagger$ in the operator). The second superamplitude does allow for helicity violation, provided that it involves a gaugino and a matter fermion. It relates this to the bosonic operators of the form $F^2\phi^2$, where the scalar must be charge-violating, as well as the MDM-like operator in (3.138), dressed with an additional scalar.

Each of these superamplitudes may be respectively further enhanced to $\mathcal{N}=2$,

$$\mathcal{A}(K,K,K,K) \propto \frac{1}{\Lambda^2}\delta^{(4)}(Q^\dagger)\frac{[12]}{\langle 34\rangle} \tag{3.143}$$

$$\mathcal{A}(V^+,V^+,V^+,V^+) \propto \frac{1}{\Lambda^2}\delta^{(4)}(Q^\dagger)\frac{[12]^2}{\langle 34\rangle^2} \tag{3.144}$$

(and conjugates), neither of which contains any new types of component interactions. The chiral multiplets of (3.142) must descend from the vector multiplets in (3.144), while the $\mathcal{N}=1$ matter interactions can only be promoted to $\mathcal{N}=2$ matter interactions. Interestingly, the $\mathcal{N}=2$ vectors still cannot couple to the hypermultiplets at this order.

**Dimension 7**

The $\mathcal{N}=1$ possibilities are

$$\mathcal{A}(\Phi^+,\Phi^+,\Phi^+,\Phi^+) \propto \frac{1}{\Lambda^3}\delta^{(2)}(Q^\dagger)\{[12]\,[34]\,,[13]\,[24]\} \tag{3.145}$$

$$\mathcal{A}(V^+,V^+,\Phi^-,\Phi^-) \propto \frac{1}{\Lambda^3}\delta^{(2)}(Q^\dagger)\,[12]^2 \tag{3.146}$$

The first superamplitude is the first example that admits multiple possible independent terms, a particular basis for which is given inside the brackets. These correspond to helicity-violating operators of the schematic form $\psi\psi\partial^2\phi\phi$, where each term corresponds to a particular distribution of the derivatives. Like the Weinberg operator, these supersymmetrise into themselves. The second superamplitude describes operators of the form $F^2\psi^2$ and its superpartners: the gaugino-scalar coupling, similar to that in (3.145), but restricted to the term proportional to the Mandelstam invariant of both scalars' momenta, and a coupling of the schematic form $F\lambda\psi^\dagger\not{\partial}\phi^\dagger$.

The only $\mathcal{N}=2$ possibility is

$$\mathcal{A}(V^+,V^+,K,K) \propto \frac{1}{\Lambda^3}\delta^{(4)}(Q^\dagger)\frac{[12]^2}{\langle 34\rangle}. \tag{3.147}$$

This superamplitude unifies both of the $\mathcal{N}=1$ superamplitudes listed above (although selecting-out only one of the terms (3.145) determined by which chiral multiplets are embedded in the $\mathcal{N}=2$ vectors).

**Dimension** 8

$\mathcal{N} = 4$ compatible interactions become admissible at dimension 8. The only possible superamplitude consistent with dimensional analysis is

$$\mathcal{A}(V, V, V, V) \propto \frac{1}{\Lambda^4}\delta^{(8)}(Q^\dagger)\frac{[12]^2}{\langle 34\rangle^2}. \tag{3.148}$$

This is the supersymmetrisation of the helicity-preserving $F^4$ operator and is (kinematically) unique.

For $\mathcal{N} < 4$, more possibilities arise than for lower dimension, as dimensional analysis permits more derivatives and therefore more ways that they can be distributed, as well as new Lorentz-invariant combinations of fermion chirality. However, supersymmetry still places stringent constraints on the possible component interactions.

For $\mathcal{N} = 1$, the possible superamplitudes are

$$\mathcal{A}(\Phi^+, \Phi^+, \Phi^-, \Phi^-) = \frac{1}{\Lambda^4}\delta^{(2)}(Q^\dagger)\,[12]\,\{c_{\Phi^4 s}s, c_{\Phi^4 t}t\} \tag{3.149}$$

$$\mathcal{A}(V^+, V^-, \Phi^+, \Phi^-) = \frac{c_{V^2\Phi^2}}{\Lambda^4}\delta^{(2)}(Q^\dagger)\,[13]\,[14]\,\langle 24\rangle \tag{3.150}$$

$$\mathcal{A}(V^+, V^+, \Phi^+, \Phi^+) = \frac{1}{\Lambda^4}\delta^{(2)}(Q^\dagger)\,[12]\,\{d_{V^2\Phi^2 s}\,[12]\,[34]\,, d_{V^2\Phi^2 t}\,[31]\,[24]\} \tag{3.151}$$

$$\mathcal{A}(V^+, V^+, V^-, V^-) = \frac{c_{V^4}}{\Lambda^4}\delta^{(2)}(Q^\dagger)\,[12]^2\,\langle 34\rangle\,. \tag{3.152}$$

The numerical Wilson coefficients are retained in these expressions to match them with the operators to be given below.

Extending to $\mathcal{N}=2$, the permitted superamplitudes are

$$\mathcal{A}(K,K,K,K) \propto \frac{1}{\Lambda^4}\delta^{(4)}(Q^\dagger)\{[12]\,[34]\,,[13]\,[24]\} \tag{3.153}$$

$$\mathcal{A}(V^+,V^-,K,K,) \propto \frac{1}{\Lambda^4}\delta^{(4)}(Q^\dagger)\,[13]\,[14] \tag{3.154}$$

$$\mathcal{A}(V^+,V^+,V^+,V^+) \propto \frac{1}{\Lambda^4}\delta^{(4)}(Q^\dagger)\frac{[12]}{\langle 34\rangle}\{[12]\,[34]\,,[13]\,[24]\} \tag{3.155}$$

$$\mathcal{A}(V^+,V^+,V^-,V^-) \propto \frac{1}{\Lambda^4}\delta^{(4)}(Q^\dagger)\,[12]^2\,. \tag{3.156}$$

These mostly just describe promotions of the respective $\mathcal{N}=1$ superamplitudes into $\mathcal{N}=2$. The superamplitudes (3.154) and (3.156) also decompose into $\mathcal{N}=1$ components that include the $c_{\Phi^4 s}$ term in (3.149).

Note that the $\mathcal{N}=1,2$ helicity-violating interactions (which have been singled-out in (3.151) by having coupling labelled as $d$ rather than $c$), which are the only examples of inelastic superamplitudes listed above, are also the only type that do not appear when that $\mathcal{N}=4$ superamplitude is decomposed into lower $\mathcal{N}$ components and are therefore not $\mathcal{N}=4$ compatible.

For minimal field theories, it is interesting to consider the terms in an effective action that would generate the superamplitudes listed above and identify the Wilson coefficients united by supersymmetry. For a $\mathcal{N}=1$ chiral multiplet, the dimension 8 operators are

$$\mathcal{L}_{EFT_8} \propto \frac{c_{\Phi_s^4}}{\Lambda^4}\left(\frac{1}{4}\phi\phi(\partial^2)^2(\phi^*\phi^*)+\frac{1}{2}\psi\psi\partial^2(\psi^\dagger\psi^\dagger)+2i\partial_\mu\phi^\dagger\partial^\nu\phi\partial_\nu\psi\sigma^\mu\psi^\dagger\right) \tag{3.157}$$

(identical particle exchange symmetry rules-out the other possible operator displayed above in (3.149)). For the vector multiplet,

$$\mathcal{L}_{EFT_8} \propto \frac{c_{V^4}}{\Lambda^4}\left(\frac{1}{16}\left((F^2)^2+(F\tilde{F})^2\right)+\frac{1}{2}\lambda\lambda\partial^2(\lambda^\dagger\lambda^\dagger)+2i\lambda F_L\sigma^\mu F_R\partial_\mu\lambda^\dagger\right). \tag{3.158}$$

Here, $F_L = F_{\mu\nu}S_L^{\mu\nu}$ and $F_R = F_{\mu\nu}S_R^{\mu\nu}$. The mixed interactions are:

$$\begin{aligned}\mathcal{L}_{EFT_8} \propto \frac{c_{V^2\Phi^2}}{\Lambda^4}\Big(&2\mathrm{tr}\left(F_L\sigma^\mu F_R\overline{\sigma}^\nu\right)\partial_\mu\phi^\dagger\partial_\nu\phi - \partial_\mu\psi\lambda\psi^\dagger\partial^\mu\lambda^\dagger \\ &- 2i\lambda^\dagger\overline{\sigma}^\mu\partial_\nu\lambda\partial_\mu\phi^\dagger\partial^\nu\phi + 2i\psi F_L\sigma^\mu F_R\partial_\mu\psi^\dagger \\ &- \sqrt{2}i\partial_\mu\psi\lambda F^{\mu\nu}\partial_\nu\phi^\dagger + conj.\Big) \qquad (3.159)\\ + \frac{d_{V^2\Phi^2 s}}{\Lambda^4}&\left(\frac{1}{8}\left(F^2 + iF\tilde{F}\right)\partial^2\phi^2 + \frac{1}{4}\psi\psi\partial^2\left(\lambda\lambda\right) + 2\sqrt{2}i\partial_\mu\phi\lambda F_L\partial^\mu\psi\right) + \text{conj.}\end{aligned}$$

(3.160)

The terms in the effective action above, or equivalently, the amplitudes that they correspond to, can be compared to those listed in Section 3.5.4 for the two fermion system. The notable differences are that supersymmetry forbids the helicity-violating interactions involving only a single species $\psi$ or $\lambda$ (i.e. $\psi^2\partial^2\psi^2$, $\lambda^2\partial^2\lambda^2$ and their conjugates). For interactions between two different species of fermion, helicity-violating interactions are permitted. However, only the two interactions listed in (3.159) and (3.160) are allowed. The others listed in (3.126) are forbidden by supersymmetry. Note that $\partial\psi\lambda\cdot\partial\psi^\dagger\lambda^\dagger + \frac{1}{2}\psi\lambda\partial^2\left(\psi^\dagger\lambda^\dagger\right) = -\partial_\mu\psi\lambda\psi^\dagger\partial^\mu\lambda^\dagger$, so supersymmetry only permits this single combination of mixed helicity conserving interaction.

Before advancing on to the superymmetrised postivity constraints, I first digress to make a comment on the application to supersymmetry breaking given in (59). The general low-energy EFT of a goldstino and $R$-axion was constructed in (59) to describe the breaking of $\mathcal{N} = 1$ supersymmetry and its $R$-symmetry. This action included an interaction of the form given by the mixed interaction in (3.157) (but with a real scalar), as well as dimension 8 helicity preserving and violating pure goldstino operators. By demanding positivity of the mixed interaction, an upper bound on the vev of the superpotential was derived from the product of the $R$-axion and goldstino decay constants (all parameters determining the low energy constants in the effective action). This same bound could be

equivalently obtained by instead applying the conclusions of Section 3.5.2 directly to the goldstino interactions.

### 3.6.3 Unity of the positivity theorems and new bounds

To begin with, the positivity constraints on the scalar (2) and the fermion (16) in (3.157) are unified when the fields are combined in a supermultiplet. Note that these interactions are consistent with both a Goldstone shift-symmetry for the scalar and a goldstino non-linear supersymmetry for the fermion. These interactions are expected for a Goldstone multiplet in a theory with extended SUSY spontaneously broken to $\mathcal{N} = 1$. Likewise, the interactions in (3.158) must each have positive coefficients (16), which is consistent with their unification under supersymmetry.

The mixed interactions are similar to those discussed in the previous Section 3.5.5. The first four terms of (3.159) induce elastic scattering of different species off each other, so the positivity of the coefficient $c_{V^2\Phi^2} > 0$ is again expected. Interestingly however, the inelastic partner operators of the form $\partial_\mu\psi\lambda F^{\mu\nu}\partial_\nu\phi^\dagger$ seem to inherit this condition. It is unclear how these operators would be constrained in the absence of supersymmetry. Their on-shell contact amplitudes that vanish in the forward limit in all channels, so a departure away from the forward scattering would seem necessary to access them. This operator will remain a puzzle here.

Promoting to $\mathcal{N} = 2$, the $V^4$ and $V^2\Phi^2$ (super)-operators unify further and the positivity of their Wilson coefficients is combined into that of the ($\mathcal{N} = 2$) $V^4$ operator. Similarly, positivity of the $\mathcal{N} = 1$ chiral multiplet interactions (3.157) becomes positivity of the analogous $\mathcal{N} = 2$ hypermultiplet interactions. Promoting further to $\mathcal{N} = 4$, all of these are unified into the single positivity constraint on the dim-8 vector multiplet interaction.

Finally, the new inelastic constraints derived in the previous section also unify. With supersymmetry, the fermion constraints from Section 3.5.4 simplify, as the amplitudes

$$\begin{aligned} &A\left(\psi^{\pm}\psi^{\pm}\to\psi^{\mp}\psi^{\mp}\right)=A\left(\lambda^{\pm}\lambda^{\pm}\to\lambda^{\mp}\lambda^{\mp}\right)\\ &=A\left(\lambda^{\pm}\lambda^{\mp}\to\psi^{\pm}\psi^{\mp}\right)=A\left(\lambda^{\pm}\lambda^{\pm}\to\psi^{\pm}\psi^{\pm}\right)=0, \end{aligned} \tag{3.161}$$

while specifically for identical particles, $M\left(\psi^{\pm}\lambda^{\pm}\to\psi^{\mp}\lambda^{\mp}\right)=0$. This leaves only one type of inelastic amplitude and its parity conjugate and the constraints can be stated as

$$\frac{1}{2}|M^{\lambda^-\lambda^-\psi^+\psi^+}\pm M^{\lambda^+\lambda^+\psi^-\psi^-}|<M^{\psi^+\lambda^-\psi^+\lambda^-}+\sqrt{M^{\psi^+\psi^-\psi^+\psi^-}M^{\lambda^+\lambda^-\lambda^+\lambda^-}}. \tag{3.162}$$

The combinations appearing on the LHS correspond to the $P$ conserving and violating interactions in the inelastic transitions (corresponding to, at tree-level, the real and imaginary parts of the coupling $d_{V^2\Phi^2 s}$ above). Notably, the amplitudes forbidden by the SWIs would all appear as additional contributions to the left hand side of (3.162), strengthening the lower bound on the elastic amplitudes.

The analysis of Section 3.5.5 can be easily extended to a complex scalar and two fermion species. With the scalar complex, the bound (3.132) generalises to

$$\frac{1}{2}|M^{\phi\phi\gamma^+\gamma^+}\pm M^{\overline{\phi}\overline{\phi}\gamma^-\gamma^-}|<M^{\phi\gamma^+\phi\gamma^+}+\sqrt{M^{\phi\phi\phi\phi}M^{\gamma^+\gamma^-\gamma^+\gamma^-}} \tag{3.163}$$

(after simplifying with $CPT$ and $Y$), which has the expected form resembling (3.162). The last partner relation, involving the mixed fermion-boson amplitudes, can also be found to be

$$\frac{1}{2}|M^{\phi\psi^-\gamma^+\lambda^+}\pm M^{\overline{\phi}\psi^+\gamma^-\lambda^-}|<\sqrt{M^{\phi\lambda^+\phi\lambda^+}M^{\psi^+\gamma^-\psi^+\gamma^-}}+\sqrt{M^{\phi\psi^+\phi\psi^+}M^{\lambda^+\gamma^-\lambda^+\gamma^-}}. \tag{3.164}$$

SWIs imply that $M^{\phi\lambda^+\phi\lambda^+} = M^{\psi^+\gamma^-\psi^+\gamma^-}$, so this bound has identical structure to the previous two, completing the full super-positivity constraint. Again, the inelastic components of (3.159) do not appear in any of these bounds and seem only to be dragged into participation by supersymmetry. For Wilson coefficients at tree-level, these bounds are encapsulated by

$$|\Re(d_{V^2\Phi^2{}_s})|, |\Im(d_{V^2\Phi^2{}_s})| < c_{V^2\Phi^2} + \sqrt{c_{\Phi^4{}_s} c_{V^4}}. \tag{3.165}$$

In contrast to the case from Section 3.5.2, the space of consistent $P$-violating inelastic couplings is a square rather than a disc.

As mentioned above, the inelastic amplitudes are also the only type not consistent with $\mathcal{N} = 4$ supersymmetry. This indicates that the lower bounds in these inequalities are minimised by requiring increasingly more supersymmetry (where simple $\mathcal{N} = 1$ is sufficient to rule-out many possible inelastic amplitudes that may potentially appear, such as the other fermionic helicity configurations in Section 3.5.4). This is the (expected) consequence of symmetries imposing selection rules that prohibit inelastic processes, but also illustrates how extreme symmetry breaking can be prohibited by the positivity bounds.

## 3.7 Conclusion

The general implication of unitarity for the causality sum rule (3.8) is to bound the size of inelastic scattering amplitudes by the size of the elastic ones (3.11). This places fundamental limits on the extent to which hypothetical symmetries, which manifest themselves in the $S$-matrix as selection rules, can be broken by effective interactions. These constraints appear naively invisible in the construction of a general effective action of local contact interactions.

Employing the convex cone picture of (29), the general set of positivity bounds for fundamental $SU(2)$ and $SU(3)$ fermions were derived, including some that cannot be obtained from considering scattering of factorised states. Separately, general constraints on flavour violation were also derived for the cases in which the fermions are only non-singlets under one symmetry group. Simple inelastic bounds for spinning particles were also derived, in particular for two scalars and two vectors and the mixed case of two fermions, a vector and a scalar, where each particle has the same-sign helicity in the all outgoing convention. It was then shown that all of the standard bounds for particles of different spin unify under supersymmetry.

In the examples discussed in 3.4.2, all states under consideration are related by symmetries. When the particles transform under multiple symmetry groups, the convex cone picture is needed for a complete characterisation of the information in the sum rule. However, if transitions between multiple distinct states are permitted by the symmetries, then the cone is non-polyhedral and the standard results for polyhedral cones cannot be so simply applied. This was analysed in (28) for parity-symmetric weak boson operators, where a set of necessary constraints were derived using analytic and numerical methods. It is easy to apply (3.11) directly to obtain necessary conditions on the couplings. However, finding the complete set of sufficient bounds remains the most immediate open problem in applying the constraints from the sum rule to EFTs of multiple species, in particular the SMEFT. Subsequent to the release of this work, (34) were able to reformulate the the sum rule (3.8) as a positive semi-definite statement in a dual space to the space of external scattering states. Once appropriately crossing-symmetrised, a tensor in external particle labels may be contracted with (3.8) to obtain an expression of positive-definiteness. This dual space of positive-definite matrices was identified as a "spectrahedron", the geometry of which has been studied in (60), and, as a space of positive-definite matrices, enabled methods from semi-definite programming to be applied

(see (61) for introduction of recent applications of these ideas to solving the CFT bootstrap equations). See also (62) for recent progress in applying the $S$-matrix bootstrap to constraining Wilson coefficients in EFTs directly from the full crossing constraints. This leverages positive semi-definiteness of the $S$-matrix to utilise semi-definite programming.

Generalisation to higher mass dimension, departure from the forward limit and massive particles are obvious future directions (17), (20), (22), (21). See e.g. (63), (31) for recent discussions in string theory.

The entire discussion of this work has been concentrated at the level of dimension-8 order effective interactions. These have the minimum energy scaling to ensure that the integral over the contour deformed to infinity converges to zero and does not introduce an additional unknown UV ingredient into the sum rule with an unknown impact on the structure. The results described here at dimension 8 readily extends to higher dimension $4n$ for $n > 2$. However, no statement about lower dimensional amplitudes has been made, so it would seem that the remarks about symmetry violation would not extend to them. However, lower dimensional interactions will typically contribute to dimension-8 level interactions through multiple insertions and still appear in the sum rule. The impact of loops in higher order constraints was recently discussed in (20). In agreement with the observations here, these strengthen the positivity bounds that would naively apply at tree-level. It would be of interest to further investigate the implications of these constraints for RG flow, both in the context for the SMEFT and more generally. As mentioned in (20), this would require a treatment of IR divergences and inclusive observables. Furthermore, as recent works have shown, even operators appearing at higher dimension $4n + 2$, for which the standard sum rule does not imply positivity, are highly constrained away from the forward limit. It would be interesting to establish precise points of distinction between the rigid set of constraints applicable at dimension 8 and above and freedom for lower dimensional (including renormalisable) operators. Exami-

nation of UV completions may provide insight into this. Some discussion of dimension 6 operators has recently been given in (64), (65), (15), (66), although conclusions drawn for the IR interactions have been predicated on assumptions about the amplitudes having extra-friendly high-energy scaling.

While most of the applications presented here have had an eye toward the SMEFT, no analysis of the impacts of these constraints on tests of the SM has been attempted here. See (30), (67), (26), (27), (68) for some recent discussion of this. The bounds discussed here activate at dimension 8 order, which is expected to be typically sub-leading to the (many) dimension 6 interactions that pervade the SMEFT. However, it may still be possible to access the affected dimension 8 interactions through non-interference effects and angular distributions (69), (70). While amplitudes vanishing in the forward limit (possibly because of angular momentum conservation) appear naively unaffected, this is not necessarily true in a crossed, inelastic channel described by the same amplitude, and as result, such a process would not escape constraint.

This entire work has relied upon the $S$-matrix formulation of causality in order to derive constraints on effective interactions. However, it would also be interesting to construct background solutions (such as was done in (2)) in order to see precisely how such a breakdown arises if the constraints are violated, especially for the fermionic and mixed spin interactions. The validity and scope of the $S$-matrix formulation is also dependent upon the analytic structure being established and the standard derivation of the dispersion relations in Section 3.2.1 relied upon this. These have been established for Wightman theories (at least for the analogous correlation functions) and theories loosely satisfying the requirements of LSZ reduction. However, this does not include applicability to perturbative gauge theories. Clarification over this issue would be informative. Some possibly related foundational questions that have practical implications are the interpretation and use of the sum rule in the presence of IR divergences, as well as possible

extension away from the forward limit where some understanding of the singularity structure may become necessary. The assumption that all particles can be given a small mass and that the results will apply to the exactly massless theory also needs to be validated (in particular, the assumption that the forward and massless limits commute), although no effort has been made in this direction here. This is least clear for possible applications to gravitational systems, where other foundational assumptions about locality in the UV completion are also uncertain.

It would also be interesting to find applications of these bounds to model building. As mentioned above, the fact that dimension 8 operators usually contribute subleading effects naively poses an obstruction to the constraints having widespread, leading-order consequences. Continuing the analysis of (30) for constraining massive higher spin particles and their hypothetical coupling to the SM is another possible application with direct consequence for the constraining the space of possible particle models of dark matter or other hypothetically fields associated with other cosmological mysteries. See e.g. (71) for discussion of coupling of massive gravitons to (regular) matter, or (18) for various other theories related to modified gravity. Alternatively, it would be of interest if these results can be used to make general statements about the nature and scope of symmetry breaking (such as of time-reversal) that is possible at low energies.

# Acknowledgments

Thank you to Grant Remmen for discussion, Isabel Garcia Garcia for comments on a draft, Nathaniel Craig for assistance with funding and Cen Zhang for some questions and comments. This work is supported by the US Department of Energy under the grant DE-SC0011702.

# Appendices

## 3.A Sum Rules for Spinning Particles with $SO(2)$

### 3.A.1 Angular momentum projectors in $SO(2)$ form

The Clebsch-Gordan coefficients for general spinning particles were given in (3.93). If one of the helicities is zero, then the Clebsch-Gordan coefficients are

$$C_h^{i0} = C_h^{0i} = \frac{1}{\sqrt{2}} \begin{cases} 1, & i = 1 \\ -i, & i = 2 \end{cases} . \tag{3.166}$$

The 0 superscript denotes the scalar state. The parity-conjugate coefficients are $C_{-h}^{i0} = C_{-h}^{0i} = (C_h^{i0})^*$.

The next step is to find the projectors. There are several special cases. For the case in which one of the incoming and outgoing particles are scalars, the projectors are simply

$$P_{Ph}^{i0k0} = \delta^{ik} \qquad P_{Ph}^{0j0l} = \delta^{jl} \qquad P_{Ph}^{i00l} = \delta^{il} \qquad P_{Ph}^{0jk0} = \delta^{jk} \tag{3.167}$$

$$P_{\not{P}h}^{i0k0} = i\epsilon^{ik} \qquad P_{\not{P}h}^{0j0l} = i\epsilon^{jl} \qquad P_{\not{P}h}^{i00l} = i\epsilon^{il} \qquad P_{\not{P}h}^{0jk0} = i\epsilon^{jk}. \tag{3.168}$$

More generally, if none of the particles are scalars and $h_1 \neq h_2$, $h_3 \neq h_4$ (so that

singlets do not appear in the product irreps), then the possible projectors are

$$P^{ijkl}_{\not P\, h_1+h_2} = \frac{i}{2}\left(\delta^{ik}\epsilon^{jl} + \delta^{jl}\epsilon^{ik}\right) \tag{3.169}$$

$$P^{ijkl}_{P\, h_1+h_2} = \frac{1}{2}\left(\delta^{ik}\delta^{jl} + \delta^{il}\delta^{jk} - \delta^{ij}\delta^{kl}\right), \qquad \text{if}\, h_1+h_2 = h_3+h_4, \tag{3.170}$$

$$P^{ijkl}_{\not P\, |h_1-h_2|} = \frac{i}{2}\left(\epsilon^{ij}\delta^{kl} - \delta^{ij}\epsilon^{kl}\right) \tag{3.171}$$

$$P^{ijkl}_{P\, |h_1-h_2|} = \frac{1}{2}\left(\delta^{ij}\delta^{kl} + \delta^{ik}\delta^{jl} - \delta^{il}\delta^{jk}\right), \qquad \text{if}\, (h_1-h_2) = (h_3-h_4), \tag{3.172}$$

$$P^{ijkl}_{\not P\, |h_1-h_2|} = \frac{i}{2}\left(\delta^{ij}\epsilon^{kl} + \epsilon^{ij}\delta^{kl}\right) \tag{3.173}$$

$$P^{ijkl}_{P\, |h_1-h_2|} = \frac{1}{2}\left(\delta^{ij}\delta^{kl} - \delta^{ik}\delta^{jl} + \delta^{il}\delta^{jk}\right), \qquad \text{if}\, (h_1-h_2) = -(h_3-h_4), \tag{3.174}$$

$$P^{ijkl}_{\not P\, h_1+h_2} = \frac{-i}{2}\left(S^{ij}\delta^{kl} \pm P^{ij}\epsilon^{kl}\right) \tag{3.175}$$

$$P^{ijkl}_{P\, h_1+h_2} = \frac{1}{2}\left(P^{ij}\delta^{kl} \mp S^{ij}\epsilon^{kl}\right), \qquad \text{if}\, h_1+h_2 = \pm(h_3-h_4), \tag{3.176}$$

$$P^{ijkl}_{\not P\, h_3+h_4} = \frac{i}{2}\left(\delta^{ij}S^{kl} \pm \epsilon^{ij}P^{kl}\right) \tag{3.177}$$

$$P^{ijkl}_{P\, h_3+h_4} = \frac{1}{2}\left(\delta^{ij}P^{kl} \pm \epsilon^{ij}S^{kl}\right), \qquad \text{if}\, \pm(h_1-h_2) = h_3+h_4. \tag{3.178}$$

The $u$-channel projectors all agree with the required $s$-channel projectors that they should be related to under crossing. For example, if $h_1+h_2 = h_3+h_4$, it is easily verified that $P^{ilkj}_{h_1-h_4} = P^{ijkl}_{h_1+h_2}$ (in either parity-symmetric or violating cases). The identity

$$\delta^{ij}\epsilon^{kl} + \delta^{kl}\epsilon^{ij} = \delta^{il}\epsilon^{kj} + \delta^{kl}\epsilon^{il} \tag{3.179}$$

is useful in handling the parity-violating projectors.

For the special case in which exactly one of the particles is a scalar, the projectors

are

$$P^{0jkl}_{\not{P}h_3+h_4} = \frac{i}{\sqrt{2}}\left(\delta^{j1}S^{kl} - \delta^{j2}P^{kl}\right) \tag{3.180}$$

$$P^{0jkl}_{P\,h_3+h_4} = \frac{1}{\sqrt{2}}\left(\delta^{j1}P^{kl} + \delta^{j2}S^{kl}\right) \tag{3.181}$$

$$P^{0jkl}_{\not{P}h_3-h_4} = \frac{-i}{\sqrt{2}}\left(\delta^{j1}\epsilon^{kl} + \delta^{j2}\delta^{kl}\right) \tag{3.182}$$

$$P^{0jkl}_{P\,h_3-h_4} = \frac{1}{\sqrt{2}}\left(\delta^{j1}\delta^{kl} - \delta^{j2}\epsilon^{kl}\right) \tag{3.183}$$

$$P^{0jkl}_{\not{P}-h_3+h_4} = \frac{i}{\sqrt{2}}\left(\delta^{j1}\epsilon^{kl} - \delta^{j2}\delta^{kl}\right) \tag{3.184}$$

$$P^{0jkl}_{P\,-h_3+h_4} = \frac{1}{\sqrt{2}}\left(\delta^{j1}\delta^{kl} + \delta^{j2}\epsilon^{kl}\right). \tag{3.185}$$

Then $P^{i0kl} = P^{0ikl}$ and $P^{ij0l} = P^{ijl0} = (P^{l0ij})^*$ for each helicity configuration. These expressions encapsulate each of the possible relations between the helicities of the scattered particles. Again, it can be easily checked that these are consistent with crossing e.g. for the case $h_3 = 0$ and $h_1 + h_2 = h_4$, $P^{il0j}_{P\,h_4-h_1} = P^{ij0l}_{P\,h_1+h_2}$, while the parity-violating projectors pick-up a negative sign $P^{il0j}_{\not{P}h_4-h_1} = -P^{ij0l}_{\not{P}h_1+h_2}$.

For the special case in which the outgoing states are both scalars, the projectors are simply the Clebsch-Gordan coefficients (or if the scalars are incoming, their conjugates)

$$P_A^{ij00} = \frac{1}{\sqrt{2}}\delta^{ij} \qquad P_A^{00kl} = \frac{1}{\sqrt{2}}\delta^{kl} \qquad P_B^{ij00} = \frac{-i}{\sqrt{2}}\epsilon^{ij} \qquad P_B^{00kl} = \frac{i}{\sqrt{2}}\epsilon^{kl}. \tag{3.186}$$

These have crossing relations

$$P^{0lk0}_{Ph} = \sqrt{2}P_A^{00kl} \qquad\qquad P^{0lk0}_{\not{P}h} = -\sqrt{2}P_B^{00kl}. \tag{3.187}$$

When the particles are not scalars, then if $h_1 = h_4$ and $h_2 = h_3$, the $u$-channel

projectors decompose as

$$P^{ilkj}_{\not P\, h_1+h_2} = -P^{ijkl}_{AB} - P^{ijkl}_{BA} \tag{3.188}$$

$$P^{ilkj}_{P\, h_1+h_2} = P^{ijkl}_{AA} + P^{ijkl}_{BB} \tag{3.189}$$

$$P^{ilkj}_{\not P\, h_1-h_2} = P^{ijkl}_{AB} - P^{ijkl}_{BA} \tag{3.190}$$

$$P^{ilkj}_{P\, h_1-h_2} = P^{ijkl}_{AA} - P^{ijkl}_{BB}\,, \tag{3.191}$$

while in the opposite channel,

$$P^{ilkj}_{AA} = \frac{1}{2}\left(P^{ijkl}_{P\, h_1+h_2} + P^{ijkl}_{P\, h_1-h_2}\right) \tag{3.192}$$

$$P^{ilkj}_{BB} = \frac{1}{2}\left(P^{ijkl}_{P\, h_1+h_2} - P^{ijkl}_{P\, h_1-h_2}\right) \tag{3.193}$$

$$P^{ilkj}_{AB} = \frac{1}{2}\left(P^{ijkl}_{\not P\, h_1-h_2} - P^{ijkl}_{\not P\, h_1+h_2}\right) \tag{3.194}$$

$$P^{ilkj}_{BA} = -\frac{1}{2}\left(P^{ijkl}_{\not P\, h_1-h_2} + P^{ijkl}_{\not P\, h_1+h_2}\right). \tag{3.195}$$

The results given in Section 3.5.1 also apply to the case $h_1 = h_2 = h \neq h_3 = h_4$, except that only the transitions between singlet irreps are possible.

### 3.A.2 Sum rules for two fermion and multispin theories

The sum rule entries for scattering in the two fermion theory of Section 3.5.4 are, in $SO(2)$ form:

$$
\begin{aligned}
M\left(\lambda,\lambda\to\lambda,\lambda\right) = &\left(|\boldsymbol{m}_1^{\lambda\lambda}|^2+\frac{1}{2}\left(|\boldsymbol{m}_A^{\lambda\lambda}|^2+|\boldsymbol{m}_B^{\lambda\lambda}|^2\right)\right)P_{P1}^{ijkl}\\
&+\frac{1}{2}\left(2|\boldsymbol{m}_1^{\lambda\lambda}|^2+3|\boldsymbol{m}_A^{\lambda\lambda}|^2-|\boldsymbol{m}_B^{\lambda\lambda}|^2\right)P_{AA}^{ijkl}\\
&+\frac{1}{2}\left(2|\boldsymbol{m}_1^{\lambda\lambda}|^2-|\boldsymbol{m}_A^{\lambda\lambda}|^2+3|\boldsymbol{m}_B^{\lambda\lambda}|^2\right)P_{BB}^{ijkl}+2\boldsymbol{m}_A^{\lambda\lambda}\cdot\boldsymbol{m}_B^{\lambda\lambda}\left(P_{AB}^{ijkl}-P_{BA}^{ijkl}\right)
\end{aligned}
\tag{3.196}
$$

$$
\begin{aligned}
M\left(\psi,\psi\to\psi,\psi\right) = &\left(|\boldsymbol{m}_1^{\psi\psi}|^2+\frac{1}{2}\left(|\boldsymbol{m}_A^{\psi\psi}|^2+|\boldsymbol{m}_B^{\psi\psi}|^2\right)\right)P_{P1}^{ijkl}\\
&+\frac{1}{2}\left(2|\boldsymbol{m}_1^{\psi\psi}|^2+3|\boldsymbol{m}_A^{\psi\psi}|^2-|\boldsymbol{m}_B^{\psi\psi}|^2\right)P_{AA}^{ijkl}\\
&+\frac{1}{2}\left(2|\boldsymbol{m}_1^{\psi\psi}|^2-|\boldsymbol{m}_A^{\psi\psi}|^2+3|\boldsymbol{m}_B^{\psi\psi}|^2\right)P_{BB}^{ijkl}\\
&+2\boldsymbol{m}_A^{\psi\psi}\cdot\boldsymbol{m}_B^{\psi\psi}\left(P_{AB}^{ijkl}-P_{BA}^{ijkl}\right)
\end{aligned}
\tag{3.197}
$$

$$
\begin{aligned}
M\left(\psi,\lambda\to\psi,\lambda\right) = &\left(|\boldsymbol{m}_1^{\psi\lambda}|^2+\frac{1}{2}\left(|\boldsymbol{m}_A^{\psi\lambda}|^2+|\boldsymbol{m}_B^{\psi\lambda}|^2\right)\right)P_{P1}^{ijkl}\\
&+\frac{1}{2}\left(2|\boldsymbol{m}_1^{\psi\lambda}|^2+3|\boldsymbol{m}_B^{\psi\lambda}|^2-|\boldsymbol{m}_A^{\psi\lambda}|^2\right)P_{AA}^{ijkl}\\
&+\frac{1}{2}\left(2|\boldsymbol{m}_1^{\psi\lambda}|^2-|\boldsymbol{m}_B^{\psi\lambda}|^2+3|\boldsymbol{m}_A^{\psi\lambda}|^2\right)P_{BB}^{ijkl}\\
&+2\boldsymbol{m}_A^{\psi\lambda}\cdot\boldsymbol{m}_B^{\psi\lambda}\left(P_{AB}^{ijkl}-P_{BA}^{ijkl}\right)
\end{aligned}
\tag{3.198}
$$

$$
\begin{aligned}
M\left(\lambda,\psi\to\lambda,\psi\right) = &\left(|\boldsymbol{m}_1^{\lambda\psi}|^2+\frac{1}{2}\left(|\boldsymbol{m}_A^{\lambda\psi}|^2+|\boldsymbol{m}_B^{\lambda\psi}|^2\right)\right)P_{P1}^{ijkl}\\
&+\frac{1}{2}\left(2|\boldsymbol{m}_1^{\lambda\psi}|^2+3|\boldsymbol{m}_B^{\lambda\psi}|^2-|\boldsymbol{m}_A^{\lambda\psi}|^2\right)P_{AA}^{ijkl}\\
&+\frac{1}{2}\left(2|\boldsymbol{m}_1^{\lambda\psi}|^2-|\boldsymbol{m}_B^{\lambda\psi}|^2+3|\boldsymbol{m}_A^{\lambda\psi}|^2\right)P_{BB}^{ijkl}\\
&+2\boldsymbol{m}_A^{\lambda\psi}\cdot\boldsymbol{m}_B^{\lambda\psi}\left(P_{AB}^{ijkl}-P_{BA}^{ijkl}\right)
\end{aligned}
\tag{3.199}
$$

$$
\begin{aligned}
M\left(\lambda,\lambda\to\psi,\psi\right) = \frac{1}{2}\Big(\boldsymbol{m}_1^{\psi\psi}\cdot\boldsymbol{m}_1^{\lambda\lambda} + \boldsymbol{m}_{-1}^{\psi\psi}\cdot\boldsymbol{m}_{-1}^{\lambda\lambda} + \boldsymbol{m}_B^{\psi\lambda}\cdot\boldsymbol{m}_B^{\lambda\psi} + \boldsymbol{m}_A^{\psi\lambda}\cdot\boldsymbol{m}_A^{\lambda\psi}\Big)P_{P1}^{ijkl} \\
+ \frac{1}{2}\Big(2\boldsymbol{m}_A^{\psi\psi}\cdot\boldsymbol{m}_A^{\lambda\lambda} + \boldsymbol{m}_1^{\psi\lambda}\cdot\boldsymbol{m}_1^{\lambda\psi} + \boldsymbol{m}_{-1}^{\psi\lambda}\cdot\boldsymbol{m}_{-1}^{\lambda\psi} \\
+ \boldsymbol{m}_B^{\psi\lambda}\cdot\boldsymbol{m}_B^{\lambda\psi} - \boldsymbol{m}_A^{\psi\lambda}\cdot\boldsymbol{m}_A^{\lambda\psi}\Big)P_{AA}^{ijkl} \\
+ \frac{1}{2}\Big(2\boldsymbol{m}_B^{\psi\psi}\cdot\boldsymbol{m}_B^{\lambda\lambda} + \boldsymbol{m}_1^{\psi\lambda}\cdot\boldsymbol{m}_1^{\lambda\psi} + \boldsymbol{m}_{-1}^{\psi\lambda}\cdot\boldsymbol{m}_{-1}^{\lambda\psi} \\
- \boldsymbol{m}_B^{\psi\lambda}\cdot\boldsymbol{m}_B^{\lambda\psi} + \boldsymbol{m}_A^{\psi\lambda}\cdot\boldsymbol{m}_A^{\lambda\psi}\Big)P_{BB}^{ijkl} \\
+ \frac{1}{2}\Big(2\boldsymbol{m}_A^{\psi\psi}\cdot\boldsymbol{m}_B^{\lambda\lambda} - \boldsymbol{m}_1^{\psi\lambda}\cdot\boldsymbol{m}_1^{\lambda\psi} + \boldsymbol{m}_{-1}^{\psi\lambda}\cdot\boldsymbol{m}_{-1}^{\lambda\psi} \\
+ \boldsymbol{m}_A^{\psi\lambda}\cdot\boldsymbol{m}_B^{\lambda\psi} - \boldsymbol{m}_B^{\psi\lambda}\cdot\boldsymbol{m}_A^{\lambda\psi}\Big)P_{AB}^{ijkl} \\
+ \frac{1}{2}\Big(2\boldsymbol{m}_B^{\psi\psi}\cdot\boldsymbol{m}_A^{\lambda\lambda} - \boldsymbol{m}_1^{\psi\lambda}\cdot\boldsymbol{m}_1^{\lambda\psi} + \boldsymbol{m}_{-1}^{\psi\lambda}\cdot\boldsymbol{m}_{-1}^{\lambda\psi} \\
- \boldsymbol{m}_A^{\psi\lambda}\cdot\boldsymbol{m}_B^{\lambda\psi} + \boldsymbol{m}_B^{\psi\lambda}\cdot\boldsymbol{m}_A^{\lambda\psi}\Big)P_{BA}^{ijkl}\,. \qquad (3.200)
\end{aligned}
$$

Here $\boldsymbol{m}_{-1}^{ff'}$ and $\boldsymbol{m}_1^{ff'}$ for example represent the UV couplings of the $f^+f'^-$ and $f^-f'^+$ helicity configurations respectively for any fermions $f$ and $f'$. For brevity, the amplitude $M\left(\lambda,\psi\to\psi,\lambda\right)$ has not been stated as it is entirely determined from crossing $M\left(\lambda,\lambda\to\psi,\psi\right)$. The $Y$-symmetry implies that $|\boldsymbol{m}_{-1}^{\lambda\lambda}| = |\boldsymbol{m}_1^{\lambda\lambda}|$, $|\boldsymbol{m}_{-1}^{\psi\psi}| = |\boldsymbol{m}_1^{\psi\psi}|$, $|\boldsymbol{m}_B^{\lambda\psi}| = |\boldsymbol{m}_B^{\psi\lambda}|$, $|\boldsymbol{m}_A^{\lambda\psi}| = |\boldsymbol{m}_A^{\psi\lambda}|$, $|\boldsymbol{m}_1^{\psi\lambda}| = |\boldsymbol{m}_{-1}^{\psi\lambda}| = |\boldsymbol{m}_1^{\lambda\psi}| = |\boldsymbol{m}_{-1}^{\lambda\psi}|$, which has been used to (partially) simplify the elastic amplitudes. As for the example of Section 3.5.2, fundamental principles rule-out the existence of the parity-violating, spinning component amplitudes in all configurations above (although it is present in the omitted $M\left(\lambda,\psi\to\psi,\lambda\right)$).

The sum rule entries for the simple multispin theory of Section 3.5.5 in $SO(2)$ form are given next. The terms can be classified by spin projection $m_z = 0, \frac{1}{2}, 1, \frac{3}{2}, 2$. They are given in Table 3.4.

| $m_z = 0$ | $\phi\phi$ | $\psi\psi$ | $\gamma\gamma$ |
|---|---|---|---|
| $\phi\phi$ | $2\lvert\boldsymbol{m}^{\phi\phi}\rvert^2$ | $0$ | $\left(\boldsymbol{m}_A^{\gamma\gamma}\cdot\boldsymbol{m}^{\phi\phi}+\sqrt{2}\boldsymbol{m}_1^{\phi\gamma}\cdot\boldsymbol{m}_1^{\gamma\phi}+\sqrt{2}\boldsymbol{m}_1^{\gamma\phi}\cdot\boldsymbol{m}_1^{\phi\gamma}\right)P_A+$ $\left(\boldsymbol{m}_B^{\gamma\gamma}\cdot\boldsymbol{m}^{\phi\phi}-\sqrt{2}\boldsymbol{m}_1^{\phi\gamma}\cdot\boldsymbol{m}_1^{\gamma\phi}+\sqrt{2}\boldsymbol{m}_1^{\gamma\phi}\cdot\boldsymbol{m}_1^{\phi\gamma}\right)P_B$ |
| $\psi\psi$ | . | $\frac{1}{2}\left(2\lvert\boldsymbol{m}_1^{\psi\psi}\rvert^2+3\lvert\boldsymbol{m}_A^{\psi\psi}\rvert^2-\lvert\boldsymbol{m}_B^{\psi\psi}\rvert^2\right)P_{AA}+$ $\frac{1}{2}\left(2\lvert\boldsymbol{m}_1^{\psi\psi}\rvert^2-\lvert\boldsymbol{m}_A^{\psi\psi}\rvert^2+3\lvert\boldsymbol{m}_B^{\psi\psi}\rvert^2\right)P_{BB}+$ $2\boldsymbol{m}_A^{\psi\psi}\cdot\boldsymbol{m}_B^{\psi\psi}\left(P_{AB}-P_{BA}\right)$ | $0$ |
| $\gamma\gamma$ | . | . | $\frac{1}{2}\left(2\lvert\boldsymbol{m}_2^{\gamma\gamma}\rvert^2+3\lvert\boldsymbol{m}_A^{\gamma\gamma}\rvert^2-\lvert\boldsymbol{m}_B^{\gamma\gamma}\rvert^2\right)P_{AA}+$ $\frac{1}{2}\left(2\lvert\boldsymbol{m}_2^{\gamma\gamma}\rvert^2-\lvert\boldsymbol{m}_A^{\gamma\gamma}\rvert^2+3\lvert\boldsymbol{m}_B^{\gamma\gamma}\rvert^2\right)P_{BB}+2\boldsymbol{m}_A^{\gamma\gamma}\cdot\boldsymbol{m}_B^{\gamma\gamma}\left(P_{AB}-P_{BA}\right)$ |

| $m_z = \frac{1}{2}$ | $\phi\psi$ | $\psi\phi$ | $\psi\gamma$ | $\gamma\psi$ |
|---|---|---|---|---|
| $\phi\psi$ | $\lvert\boldsymbol{m}_{\frac{1}{2}}^{\phi\psi}\rvert^2P_{P\frac{1}{2}}$ | $0$ | $0$ | $\left(\boldsymbol{m}_{\frac{1}{2}}^{\psi\gamma}\cdot\boldsymbol{m}_{\frac{1}{2}}^{\phi\psi}+\boldsymbol{m}_{\frac{1}{2}}^{\phi\psi}\cdot\boldsymbol{m}_{\frac{1}{2}}^{\psi\gamma}\right)P_{P\frac{1}{2}}+$ $\left(\boldsymbol{m}_{\frac{1}{2}}^{\psi\gamma}\cdot\boldsymbol{m}_{\frac{1}{2}}^{\phi\psi}-\boldsymbol{m}_{\frac{1}{2}}^{\phi\psi}\cdot\boldsymbol{m}_{\frac{1}{2}}^{\psi\gamma}\right)P_{\not{P}\frac{1}{2}}$ |
| $\psi\phi$ | . | $\lvert\boldsymbol{m}_{\frac{1}{2}}^{\phi\psi}\rvert^2P_{P\frac{1}{2}}$ | $\left(\boldsymbol{m}_{\frac{1}{2}}^{\psi\gamma}\cdot\boldsymbol{m}_{\frac{1}{2}}^{\phi\psi}+\boldsymbol{m}_{\frac{1}{2}}^{\phi\psi}\cdot\boldsymbol{m}_{\frac{1}{2}}^{\psi\gamma}\right)P_{P\frac{1}{2}}-$ $\left(\boldsymbol{m}_{\frac{1}{2}}^{\psi\gamma}\cdot\boldsymbol{m}_{\frac{1}{2}}^{\phi\psi}-\boldsymbol{m}_{\frac{1}{2}}^{\phi\psi}\cdot\boldsymbol{m}_{\frac{1}{2}}^{\psi\gamma}\right)P_{\not{P}\frac{1}{2}}$ | $0$ |
| $\psi\gamma$ | . | . | $\left(\lvert\boldsymbol{m}_{\frac{3}{2}}^{\psi\gamma}\rvert^2+\lvert\boldsymbol{m}_{\frac{1}{2}}^{\psi\gamma}\rvert^2\right)P_{P\frac{1}{2}}$ | $0$ |
| $\gamma\psi$ | . | . | . | $\left(\lvert\boldsymbol{m}_{\frac{3}{2}}^{\psi\gamma}\rvert^2+\lvert\boldsymbol{m}_{\frac{1}{2}}^{\psi\gamma}\rvert^2\right)P_{P\frac{1}{2}}$ |

| $m_z = 1$ | $\psi\psi$ | $\phi\gamma$ | $\gamma\phi$ |
|---|---|---|---|
| $\psi\psi$ | $\left(\lvert\boldsymbol{m}_1^{\psi\psi}\rvert^2+\frac{1}{2}\left(\lvert\boldsymbol{m}_A^{\psi\psi}\rvert^2+\lvert\boldsymbol{m}_B^{\psi\psi}\rvert^2\right)\right)P_{P1}$ | $0$ | $0$ |
| $\phi\gamma$ | . | $\lvert\boldsymbol{m}_1^{\phi\gamma}\rvert^2P_{P1}$ | $\left(\boldsymbol{m}_1^{\gamma\phi}\cdot\boldsymbol{m}_1^{\phi\gamma}+\boldsymbol{m}_1^{\phi\gamma}\cdot\boldsymbol{m}_1^{\gamma\phi}+\frac{1}{\sqrt{2}}\boldsymbol{m}_A^{\gamma\gamma}\cdot\boldsymbol{m}^{\phi\phi}\right)P_{P1}+$ $\left(\boldsymbol{m}_1^{\gamma\phi}\cdot\boldsymbol{m}_1^{\phi\gamma}-\boldsymbol{m}_1^{\phi\gamma}\cdot\boldsymbol{m}_1^{\gamma\phi}-\frac{1}{\sqrt{2}}\boldsymbol{m}_B^{\gamma\gamma}\cdot\boldsymbol{m}^{\phi\phi}\right)P_{\not{P}1}$ |
| $\gamma\phi$ | . | . | $\lvert\boldsymbol{m}_1^{\gamma\phi}\rvert^2P_{P1}$ |

| $m_z = \frac{3}{2}$ | $\psi\gamma$ | $\gamma\psi$ |
|---|---|---|
| $\psi\gamma$ | $\left(\lvert\boldsymbol{m}^{\psi\gamma}_{\frac{3}{2}}\rvert^2 + \lvert\boldsymbol{m}^{\psi\gamma}_{\frac{1}{2}}\rvert^2\right) P_{P\frac{3}{2}}$ | $0$ |
| $\gamma\psi$ | . | $\left(\lvert\boldsymbol{m}^{\psi\gamma}_{\frac{3}{2}}\rvert^2 + \lvert\boldsymbol{m}^{\psi\gamma}_{\frac{1}{2}}\rvert^2\right) P_{P\frac{3}{2}}$ |
| $m_z = 2$ | $\gamma\gamma$ | |
| $\gamma\gamma$ | $\left(\lvert\boldsymbol{m}^{\gamma\gamma}_{2}\rvert^2 + \frac{1}{2}\left(\lvert\boldsymbol{m}^{\gamma\gamma}_{A}\rvert^2 + \lvert\boldsymbol{m}^{\gamma\gamma}_{B}\rvert^2\right)\right) P_{P2}$ | |

Table 3.4: Sum rule for theory of spinning particles.

The vanishing entries correspond to amplitudes that do not have dimension 8 order contributions. Amplitudes with the required mass dimension cannot be constructed consistently respecting Lorentz invariance and possessing the required little group scaling. Many of the processes, most notably the elastic ones, are also accidentally parity symmetric as a consequence of $CPT$ (and the fact that the particles are assumed to be self-conjugate in this theory). Both $Y$ and $CPT$ can be invoked to simplify the entries. The $Y$ symmetry further relates $|\boldsymbol{m}_1^{\phi\gamma}| = |\boldsymbol{m}_1^{\gamma\phi}|$. The couplings $\boldsymbol{m}_2^{\gamma\gamma}$, $\boldsymbol{m}_1^{\psi\psi}$ and $\boldsymbol{m}_{\frac{3}{2}}^{\psi\gamma}$ are redundant.

# Bibliography


[1] C. Burgess, *Introduction to Effective Field Theory*, *Ann. Rev. Nucl. Part. Sci.* **57** (2007) 329–362, [hep-th/0701053].

[2] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, *Causality, analyticity and an IR obstruction to UV completion*, *JHEP* **10** (2006) 014, [hep-th/0602178].

[3] G. Dvali, A. Franca, and C. Gomez, *Road Signs for UV-Completion*, arXiv:1204.6388.

[4] P. Cooper, S. Dubovsky, and A. Mohsen, *Ultraviolet complete Lorentz-invariant theory with superluminal signal propagation*, *Phys. Rev. D* **89** (2014), no. 8 084044, [arXiv:1312.2021].

[5] S. B. Giddings and R. A. Porto, *The Gravitational S-matrix*, *Phys. Rev. D* **81** (2010) 025002, [arXiv:0908.0004].

[6] L. Keltner and A. J. Tolley, *UV properties of Galileons: Spectral Densities*, arXiv:1502.05706.

[7] J. Tokuda, *Extension of positivity bounds to non-local theories: IR obstructions to Lorentz invariant UV completions*, *JHEP* **05** (2019) 216, [arXiv:1902.10039].

[8] T. Hartman, S. Jain, and S. Kundu, *Causality Constraints in Conformal Field Theory*, *JHEP* **05** (2016) 099, [arXiv:1509.00014].

[9] Z. Komargodski, M. Kulaxizi, A. Parnachev, and A. Zhiboedov, *Conformal Field Theories and Deep Inelastic Scattering*, *Phys. Rev. D* **95** (2017), no. 6 065011, [arXiv:1601.05453].

[10] X. O. Camanho, J. D. Edelstein, J. Maldacena, and A. Zhiboedov, *Causality Constraints on Corrections to the Graviton Three-Point Coupling*, *JHEP* **02** (2016) 020, [arXiv:1407.5597].

[11] S. Caron-Huot, *Analyticity in Spin in Conformal Theories*, *JHEP* **09** (2017) 078, [arXiv:1703.00278].

[12] N. Afkhami-Jeddi, S. Kundu, and A. Tajdini, *A Bound on Massive Higher Spin Particles*, *JHEP* **04** (2019) 056, [arXiv:1811.01952].

[13] M. Kologlu, P. Kravchuk, D. Simmons-Duffin, and A. Zhiboedov, *Shocks, Superconvergence, and a Stringy Equivalence Principle*, arXiv:1904.05905.

[14] S. Caron-Huot, D. Mazac, L. Rastelli, and D. Simmons-Duffin, *Dispersive CFT Sum Rules*, arXiv:2008.04931.

[15] B. Bellazzini, L. Martucci, and R. Torre, *Symmetries, Sum Rules and Constraints on Effective Field Theories*, *JHEP* **09** (2014) 100, [arXiv:1405.2960].

[16] B. Bellazzini, *Softness and amplitudes' positivity for spinning particles*, *JHEP* **02** (2017) 034, [arXiv:1605.06111].

[17] C. de Rham, S. Melville, A. J. Tolley, and S.-Y. Zhou, *UV complete me: Positivity Bounds for Particles with Spin*, *JHEP* **03** (2018) 011, [arXiv:1706.02712].

[18] C. de Rham, S. Melville, A. J. Tolley, and S.-Y. Zhou, *Positivity Bounds for Massive Spin-1 and Spin-2 Fields*, *JHEP* **03** (2019) 182, [arXiv:1804.10624].

[19] Z. Komargodski and A. Schwimmer, *On Renormalization Group Flows in Four Dimensions*, *JHEP* **12** (2011) 099, [arXiv:1107.3987].

[20] B. Bellazzini, J. Elias Miró, R. Rattazzi, M. Riembau, and F. Riva, *Positive Moments for Scattering Amplitudes*, arXiv:2011.00037.

[21] S. Caron-Huot and V. Van Duong, *Extremal Effective Field Theories*, arXiv:2011.02957.

[22] A. J. Tolley, Z.-Y. Wang, and S.-Y. Zhou, *New positivity bounds from full crossing symmetry*, arXiv:2011.02400.

[23] I. Brivio and M. Trott, *The Standard Model as an Effective Field Theory*, *Phys. Rept.* **793** (2019) 1–98, [arXiv:1706.08945].

[24] G. N. Remmen and N. L. Rodd, *Consistency of the Standard Model Effective Field Theory*, *JHEP* **12** (2019) 032, [arXiv:1908.09845].

[25] G. N. Remmen and N. L. Rodd, *Flavor Constraints from Unitarity and Analyticity*, *Phys. Rev. Lett.* **125** (2020), no. 8 081601, [arXiv:2004.02885].

[26] Q. Bi, C. Zhang, and S.-Y. Zhou, *Positivity constraints on aQGC: carving out the physical parameter space*, *JHEP* **06** (2019) 137, [arXiv:1902.08977].

[27] C. Zhang and S.-Y. Zhou, *Positivity bounds on vector boson scattering at the LHC*, *Phys. Rev. D* **100** (2019), no. 9 095003, [arXiv:1808.00010].

[28] K. Yamashita, C. Zhang, and S.-Y. Zhou, *Elastic positivity vs extremal positivity bounds in SMEFT: a case study in transversal electroweak gauge-boson scatterings*, arXiv:2009.04490.

[29] C. Zhang and S.-Y. Zhou, *A convex geometry perspective to the (SM)EFT space*, arXiv:2005.03047.

[30] B. Bellazzini and F. Riva, *New phenomenological and theoretical perspective on anomalous ZZ and Zγ processes*, *Phys. Rev. D* **98** (2018), no. 9 095021, [arXiv:1806.09640].

[31] Y.-t. Huang, J.-Y. Liu, L. Rodina, and Y. Wang, *Carving out the Space of Open-String S-matrix*, arXiv:2008.02293.

[32] J.-Y. Liu and Z.-M. You, *The supersymmetric spinning polynomial*, arXiv:2011.11299.

[33] Q. Bonnefoy, E. Gendy, and C. Grojean, *Positivity bounds on Minimal Flavor Violation*, *JHEP* **04** (2021) 115, [arXiv:2011.12855].

[34] X. Li, C. Yang, H. Xu, C. Zhang, and S.-Y. Zhou, *Positivity in Multi-Field EFTs*, arXiv:2101.01191.

[35] J. Bros, H. Epstein, and V. J. Glaser, *Some rigorous analyticity properties of the four-point function in momentum space*, *Nuovo Cim.* **31** (1964) 1265–1302.

[36] M. Gell-Mann, M. Goldberger, and W. E. Thirring, *Use of causality conditions in quantum theory*, *Phys. Rev.* **95** (1954) 1612–1627.

[37] A. Martin, *Extension of the axiomatic analyticity domain of scattering amplitudes by unitarity. 1.*, *Nuovo Cim. A* **42** (1965) 930–953.

[38] D. Simmons-Duffin, *CFT in Lorentzian Signature*, . https://physicslearning.colorado.edu/tasi/tasi_2019/tasi_2019.html.

[39] M. Froissart, *Asymptotic behavior and subtractions in the Mandelstam representation*, *Phys. Rev.* **123** (1961) 1053–1057.

[40] D. Olive, *Unitarity and evaluation of discontinuities*, *Nuovo Cim. A* **26** (1962) 3905.

[41] C. Cheung and G. N. Remmen, *Infrared Consistency and the Weak Gravity Conjecture*, *JHEP* **12** (2014) 087, [arXiv:1407.7865].

[42] J. Tokuda, K. Aoki, and S. Hirano, *Gravitational positivity bounds*, arXiv:2007.15009.

[43] L. Alberte, C. de Rham, S. Jaitly, and A. J. Tolley, *Positivity Bounds and the Massless Spin-2 Pole*, arXiv:2007.12667.

[44] B. Gavela, E. Jenkins, A. Manohar, and L. Merlo, *Analysis of General Power Counting Rules in Effective Field Theory*, *Eur. Phys. J. C* **76** (2016), no. 9 485, [arXiv:1601.07551].

[45] A. V. Manohar and V. Mateu, *Dispersion Relation Bounds for pi pi Scattering*, *Phys. Rev. D* **77** (2008) 094019, [arXiv:0801.3222].

[46] V. Mateu, *Universal Bounds for SU(3) Low Energy Constants*, *Phys. Rev. D* **77** (2008) 094020, [arXiv:0801.3627].

[47] Y.-J. Wang, F.-K. Guo, C. Zhang, and S.-Y. Zhou, *Generalized positivity bounds on chiral perturbation theory*, *JHEP* **07** (2020) 214, [arXiv:2004.03992].

[48] S. Andriolo, T.-C. Huang, T. Noumi, H. Ooguri, and G. Shiu, *Duality and axionic weak gravity*, *Phys. Rev. D* **102** (2020), no. 4 046008, [arXiv:2004.13721].

[49] K. Fukuda, *Lecture: Polyhedral computation.* 2016. http://www-oldurls.inf.ethz.ch/personal/fukudak/lect/pclect/notes2016/PolyComp2016.pdf.

[50] D. Avis, *lrs: A Revised Implementation of the Reverse Search Vertex Enumeration Algorithm. In: Kalai G., Ziegler G.M. (eds) Polytopes — Combinatorics and Computation*, *DMV Seminar* **29** (2000), no. 8 177.

[51] H. K. Dreiner, H. E. Haber, and S. P. Martin, *Two-component spinor techniques and Feynman rules for quantum field theory and supersymmetry*, *Phys. Rept.* **494** (2010) 1–196, [arXiv:0812.1594].

[52] S. Weinberg, *Feynman Rules for Any Spin. 2. Massless Particles*, *Phys. Rev.* **134** (1964) B882–B896.

[53] M. K. Gaillard and B. Zumino, *Duality Rotations for Interacting Fields*, *Nucl. Phys. B* **193** (1981) 221–244.

[54] H. Elvang and Y.-t. Huang, *Scattering Amplitudes*, arXiv:1308.1697.

[55] H. Elvang, D. Z. Freedman, and M. Kiermaier, *A simple approach to counterterms in N=8 supergravity*, *JHEP* **11** (2010) 016, [arXiv:1003.5018].

[56] S. Lal and S. Raju, *The Next-to-Simplest Quantum Field Theories*, *Phys. Rev. D* **81** (2010) 105002, [arXiv:0910.0930].

[57] H. Elvang, Y.-t. Huang, and C. Peng, *On-shell superamplitudes in N<4 SYM*, *JHEP* **09** (2011) 031, [arXiv:1102.4843].

[58] M. Srednicki, *Quantum field theory*. Cambridge University Press, 2007.

[59] M. Dine, G. Festuccia, and Z. Komargodski, *A Bound on the Superpotential*, *JHEP* **03** (2010) 011, [arXiv:0910.2527].

[60] M. Ramana and A. J. Goldman, *Some geometric results in semidefinite programming*, *Journal of Global Optimization* **7** (1995) 33–50.

[61] D. Simmons-Duffin, *The Conformal Bootstrap*, in *Theoretical Advanced Study Institute in Elementary Particle Physics: New Frontiers in Fields and Strings*, 2, 2016. arXiv:1602.07982.

[62] A. Hebbar, D. Karateev, and J. Penedones, *Spinning S-matrix Bootstrap in 4d*, arXiv:2011.11708.

[63] M. B. Green and C. Wen, *Superstring amplitudes, unitarily, and Hankel determinants of multiple zeta values*, *JHEP* **11** (2019) 079, [arXiv:1908.08426].

[64] G. N. Remmen and N. L. Rodd, *Signs, Spin, SMEFT: Positivity at Dimension Six*, arXiv:2010.04723.

[65] J. Gu and L.-T. Wang, *Sum Rules in the Standard Model Effective Field Theory from Helicity Amplitudes*, arXiv:2008.07551.

[66] A. Falkowski, S. Rychkov, and A. Urbano, *What if the Higgs couplings to W and Z bosons are larger than in the Standard Model?*, *JHEP* **04** (2012) 073, [arXiv:1202.1532].

[67] J. Gu, L.-T. Wang, and C. Zhang, *An unambiguous test of positivity at lepton colliders*, arXiv:2011.03055.

[68] B. Fuks, Y. Liu, C. Zhang, and S.-Y. Zhou, *Positivity in electron-positron scattering: testing the axiomatic quantum field theory principles and probing the existence of UV states*, arXiv:2009.02212.

[69] S. Alioli, R. Boughezal, E. Mereghetti, and F. Petriello, *Novel angular dependence in Drell-Yan lepton production via dimension-8 operators*, *Phys. Lett. B* **809** (2020) 135703, [arXiv:2003.11615].

[70] A. Azatov, R. Contino, C. S. Machado, and F. Riva, *Helicity selection rules and noninterference for BSM amplitudes*, *Phys. Rev. D* **95** (2017), no. 6 065014, [arXiv:1607.05236].

[71] A. Falkowski and G. Isabella, *Matter coupling in massive gravity*, *JHEP* **04** (2020) 014, [arXiv:2001.06800].

# Chapter 4

# Cosmological signals of a mirror twin Higgs

We investigate the cosmology of the minimal model of neutral naturalness, the mirror Twin Higgs. The softly-broken mirror symmetry relating the Standard Model to its twin counterpart leads to significant dark radiation in tension with BBN and CMB observations. We quantify this tension and illustrate how it can be mitigated in several simple scenarios that alter the relative energy densities of the two sectors while respecting the softly-broken mirror symmetry. In particular, we consider both the out-of-equilibrium decay of a new scalar as well as reheating in a toy model of twinned inflation, Twinflation. In both cases the dilution of energy density in the twin sector does not merely reconcile the existence of a mirror Twin Higgs with cosmological constraints, but predicts contributions to cosmological observables that may be probed in current and future CMB experiments. This raises the prospect of discovering evidence of neutral naturalness through cosmology rather than colliders.

## 4.1 Introduction

The electroweak hierarchy problem is one of the primary motivators for accessible physics beyond the Standard Model and has led to an expansive set of searches at the LHC and beyond. Recent null results in searches for conventional approaches to the hierarchy problem motivate the exploration of alternative solutions. "Neutral naturalness" provides one such promising alternative, in which the lightest states responsible for protecting the weak scale are partly or wholly neutral under the Standard Model (SM). In these theories, discrete symmetries enforce cancellations between finite threshold corrections to the Higgs mass. The discrete symmetries may be approximate or exact, although solutions with approximate symmetries typically require a plethora of new particles near the TeV scale.

Perhaps the simplest avatar of neutral naturalness is the "mirror" Twin Higgs (1), in which the new physics near the weak scale consists of an identical copy of the Standard Model related by an exact $\mathbb{Z}_2$ exchange symmetry. Higgs portal-type couplings between the Higgs doublets of the Standard Model and the twin sector lead to accidental global symmetries that protect the Higgs mass. The lightest partner particles are entirely neutral under the Standard Model, subject only to indirect bounds from precision Higgs coupling measurements. In conjunction with supersymmetry or compositeness at 5-10 TeV, this provides a complete solution to the "little" and "big" hierarchy problems consistent with current LHC limits. In this respect, the Twin Higgs naturally reconciles the observation of a light Higgs with the absence of evidence for new physics thus far at the LHC.

The primary challenge to the mirror Twin Higgs comes not from LHC data, but from cosmology. An exact $\mathbb{Z}_2$ exchange symmetry predicts mirror copies of light Standard Model states, which contribute to the energy density of the early universe. In particular, twin neutrinos and a twin photon provide a new source of dark radiation that is strongly

constrained by CMB and BBN measurements (2; 3). While these constraints could be avoided if the two sectors were at radically different temperatures, the Higgs portal couplings required by naturalness keep the two sectors in thermal equilibrium down to relatively low temperatures. Constraints on dark radiation in the mirror Twin Higgs have motivated models in which the $\mathbb{Z}_2$ symmetry is approximate (such as the orbifold (4; 5), holographic (6–8), fraternal (9), and vector-like (10) Twin Higgs), in which case the dark radiation component can be made naturally small. This problem was examined recently in (11), where the $\mathbb{Z}_2$ symmetry in the fermion Yukawa couplings was broken in order to find an arrangement that would reduce the residual dark radiation from the twin particles.[1] However, such cosmological fixes come at the cost of minimality, as models with approximate $\mathbb{Z}_2$ symmetries require a considerable amount of additional structure near the TeV scale.

In this work we take an alternative approach and investigate ways in which early universe cosmology can reconcile the mirror Twin Higgs with current CMB and BBN observations. In doing so, we find compelling scenarios that transfer the signatures of electroweak naturalness from high-energy colliders to cosmology. We consider several possibilities in which the energy density of the light particles in the twin sector is diluted by the out-of-equilibrium decay of a new particle after the two sectors have thermally decoupled. Crucially, the new physics in the early universe respects the exact (albeit spontaneously broken) $\mathbb{Z}_2$ exchange symmetry of the mirror Twin Higgs. This symmetry may be used to classify representations of the particle responsible for this dilution. We concentrate on two minimal cases: In the first, the long-lived particle is $\mathbb{Z}_2$-even and the asymmetry is naturally induced by kinematics. In the second, there is a pair of particles which are exchanged by the $\mathbb{Z}_2$ symmetry and which may be responsible for inflation.[2]

[1] For recent related work on the cosmology and cosmological signatures of non-minimal Twin Higgs scenarios, see e.g. (12–18).

[2] A third case exists, in which the particle is $\mathbb{Z}_2$-odd. This may additionally be related to the spon-

Moreover, in these cases the new physics does not merely reconcile the existence of a mirror twin sector with cosmological constraints, but predicts contributions to cosmological observables that may be probed in current and future CMB experiments. This raises the prospect of discovering evidence of electroweak naturalness first through cosmology, rather than colliders, and provides natural targets for future cosmological constraints on minimal realizations of neutral naturalness.

This paper is organized as follows: We begin in Section 4.2 by reviewing the salient features of the mirror Twin Higgs. In Section 4.3 we discuss the thermal history of the mirror Twin Higgs, with a particular attention to the interactions keeping the Standard Model and twin sector in thermal equilibrium and the cosmological constraints on light degrees of freedom. In Section 4.4 we present a simple model where the out-of-equilibrium decay of a particle with symmetric couplings to the Standard Model and twin sector leads to a temperature difference between the two sectors after they decouple. We turn to inflation in Section 4.5, constructing a model of "twinflation" in which the softly broken $\mathbb{Z}_2$-symmetry extends to the inflationary sector and leads to two periods of inflation. The first primarily reheats the twin sector, while the second primarily reheats the Standard Model sector. We conclude in Section 4.6.

## 4.2 The Mirror Twin Higgs

We begin by briefly reviewing the salient details of the mirror Twin Higgs. The reader is referred to any of the references listed in the previous section for further details. The theory consists of the Standard Model and an identical copy, related by a $\mathbb{Z}_2$ exchange symmetry at a scale $\Lambda \gg v$. The two sectors are connected only by Higgs portal-type

taneous $\mathbb{Z}_2$-breaking in the Higgs potential, although we find that a realisation of such a scenario is dependent upon the UV completion of the model.

interactions between the two $SU(2)$ doublet scalars.[3] Subject to conditions on the quartic coupling, the Higgs sector enjoys an approximate $SU(4)$ global symmetry.[4]

The Higgs potential is best organized in terms of the accidental $SU(4)$ symmetry involving the $SU(2)$ Higgs doublets of the SM and twin sectors, $H_A$ and $H_B$. The general tree-level Twin Higgs potential is given by (see e.g. (9))

$$V(H_A, H_B) = \lambda(|H_A|^2 + |H_B|^2 - f^2/2)^2 + \kappa(|H_A|^4 + |H_B|^4) + \sigma f^2 |H_A|^2 \tag{4.1}$$

The first term respects the accidental $SU(4)$ global symmetry, the second breaks $SU(4)$ but preserves the $\mathbb{Z}_2$ and the final term softly breaks the $\mathbb{Z}_2$. Clearly, $\kappa, \sigma \ll \lambda$ are required for the $SU(4)$ to be a good symmetry of the potential. The coupling $\kappa$ should naturally be of order the expected $SU(4)$-breaking radiative corrections to the potential induced by Yukawa interactions with the top/twin top, $\kappa \sim 3y_t^4/(8\pi^2)\log(\Lambda/m_t) \sim 0.1$ for a cut-off $\Lambda \sim 10$ TeV ($y_t$ being the top quark Yukawa coupling and $m_t$ its mass). Requiring $\lambda \gg \kappa$ therefore implies $\lambda \gtrsim 1$. As the SM and twin isospin gauge groups are disjoint subgroups of the $SU(4)$, the spontaneous breaking of the $SU(4)$ coincides with the SM and twin electroweak symmetry breaking. Three Goldstone bosons are eaten by the broken gauge bosons in each sector, leaving one Goldstone remaining. This will acquire mass through the breaking of the $SU(4)$ that is naturally smaller than the twin scale $f$. For future reference, it is convenient to define the real scalar degrees of freedom in the gauge basis as $h_A = \frac{1}{\sqrt{2}}\Re(H_A^0) - v_A$ and $h_B = \frac{1}{\sqrt{2}}\Re(H_B^0) - v_B$, where $\langle H_A^0 \rangle = v_A$ and $\langle H_B^0 \rangle = v_B$.

The surviving Goldstone boson should be dominantly composed of the $h_A$ gauge

[3]Here and in what follows we neglect possible kinetic mixing between the two $U(1)_Y$ gauge bosons; such mixing is not generated in the low-energy theory at three loops (1), and may be forbidden in UV completions where the mirror symmetry relates sectors with unified gauge groups.

[4]Properly speaking, the model must contain an $SO(8)$ global symmetry in order to enjoy a residual custodial symmetry (7; 8), but in linear realizations the $SU(4)$ is sufficient provided that higher-dimensional operators violating the custodial symmetry are adequately suppressed.

eigenstate in order to be SM-like. The soft $\mathbb{Z}_2$-breaking coupling $\sigma$ is required to tune the potential so that the vacuum expectation values (vevs) are asymmetric and that the Goldstone is mostly aligned with the $h_A$ field direction. The (unique) minimum of the Twin Higgs potential (4.1) occurs at $v_A \approx \frac{f}{2}\sqrt{\frac{\lambda(\kappa-\sigma)-\kappa\sigma}{\lambda\kappa}}$ and $v_B \approx \frac{f}{2}\sqrt{\frac{\sigma+\kappa}{\kappa}}$. The required alignment of the vacuum in the $H_B$ direction occurs if $\sigma \approx \kappa$. The consequences of this are that $v_A \approx v/\sqrt{2}$ and $v_B \approx f/\sqrt{2} \gg v$ (where $v$ is the vev of the SM Higgs, although $v_A \approx$ 174 GeV is the vev that determines the SM particle masses and electroweak properties), so that the SM-like Higgs $h$ is identified with the Goldstone mode and is naturally lighter than the other remaining real scalar, a radial mode $H$ whose mass is set by the scale $f$. The component of $h$ in the $h_B$ gauge eigenstate is $\delta_{hB} \approx v/f$ (to lowest order in $v/f$). Measurements of the Higgs couplings restrict $f \gtrsim 3v$ (9), and the naive tuning of the weak scale associated with this inequality is of order $f^2/2v^2$.

The spectrum of states in the broken phase consists of a SM-like pseudo-Goldstone Higgs $h$ of mass $m_h^2 \sim 8\kappa v^2$, a radial twin Higgs mode $H$ of mass $m_H^2 \sim 2\lambda f^2$, a conventional Standard Model sector of gauge bosons and fermions and a corresponding mirror sector. The current masses of quarks, gauge bosons, and charged leptons in the twin sector are larger than their Standard Model counterparts by $\sim f/v$, while the twin QCD scale is larger by a factor $\sim (1 + \log(f/v))$ due to the impact of the higher mass scale of heavy twin quarks on the renormalisation group (RG) evolution of the twin strong coupling. The relative mass of twin neutrinos depends on the origin of neutrino masses, some possibilities being $\sim f/v$ for Dirac masses and $\sim f^2/v^2$ for Majorana masses from the Weinberg operator. Mixing in the scalar sector implies that the SM-like Higgs couples to twin sector matter with an $\mathcal{O}(v/f)$ mixing angle, as does the radial twin Higgs mode to Standard Model matter. These mixings provide the primary portal between the Standard Model and twin sectors.

The Goldstone Higgs is protected from radiative corrections from $\mathbb{Z}_2$-symmetric physics

above the scale $f$. While the mirror Twin Higgs addresses the little hierarchy problem, it does not address the big hierarchy problem, as nothing stabilizes the scale $f$ against radiative corrections. However, the scale $f$ can be stabilized by supersymmetry, compositeness, or perhaps additional copies of the twin mechanism without requiring new states beneath the TeV scale. Minimal supersymmetric UV completions can furthermore remain perturbative up to the GUT scale (19), (20).

## 4.3 Thermal History of the Mirror Twin

The primary challenge to the mirror Twin Higgs comes from cosmology, rather than collider physics. The mirror Twin contains not only states responsible for protecting the Higgs against radiative corrections (such as the twin top), but also a plethora of extra states due to the $\mathbb{Z}_2$ symmetry that are irrelevant to naturalness. The lightest of these, namely the twin photon and twin neutrinos, contribute significantly to the energy density of the early universe around the era of matter-radiation equality, since they have a temperature comparable to that of the Standard Model plasma at all times. This is because the same Higgs portal coupling that makes the Higgs natural also keeps the two sectors in thermal equilibrium down to $\mathcal{O}(\text{GeV})$ temperatures. Then the identical particle content in the twin and Standard Model sectors guarantees that they remain at comparable temperatures even after they decouple - for every massive Standard Model species that becomes non-relativistic and transfers its entropy to the rest of the plasma, its twin counterpart does the same within a factor of $f/v$ in temperature.

In this section we undertake a detailed study of the decoupling between the Standard Model and twin sectors as well as the constraints from precision cosmology.

### 4.3.1 Twin Degrees of Freedom

In thermal equilibrium, each relativistic degree of freedom has roughly the same energy density. In general, we express the energy density of the universe $\rho$ during the radiation-dominated era as $\rho \equiv g_\star \frac{\pi^2}{30} T^4$, where we define $g_\star$ through this relation as the effective number of relativistic degrees of freedom and $T$ the temperature of the SM photons. This then determines the evolution of the scale factor through the first Friedmann equation

$$H = \frac{1}{M_{\rm pl}} \left[ \frac{\pi^2}{90} g_\star T^4 \right]^{1/2} \tag{4.2}$$

(assuming spatial flatness), where $M_{\rm pl}$ is the reduced Planck mass. In general, the energy density of a particular species $i$ may be computed from $\rho_i = g_i \int \frac{d^3p}{(2\pi)^3} f_i(p, T_i) E(p)$, where $g_i$ are the number of internal degrees of freedom, $E(p)$ is the energy as a function of momentum $p$, while $f_i(p, T_i)$ is the phase-space number density and is a Bose-Einstein or Fermi-Dirac distribution if the species is in equilibrium at temperature $T_i$. The number of effective relativistic degrees of freedom may then be defined for each sector separately as $g_\star^{\rm SM}(T)$ and $g_\star^{\rm t}(T)$ satisfying $\rho_{\rm SM}(T) = \frac{\pi^2}{30} g_\star^{\rm SM}(T) T^4$ and $\rho_{\rm t}(T) = \frac{\pi^2}{30} g_\star^{\rm t}(T) T^4$, respectively, where $\rho_{\rm SM}(T)$ and $\rho_{\rm t}(T)$ are the total energy densities of SM and twin particles. The values of $g_\star(T)$ for the SM and twin sectors are shown in Figure 4.1, where all species within each sector are in thermal equilibrium. These can then be used to calculate the total number $g_\star$ as a function of temperature, by weighting twin sector energy density by its temperature: $g_\star(T) = g_\star^{\rm SM}(T) + g_\star^{\rm t}(\hat{T})(\hat{T}/T)^4$, where $\hat{T}$ is the twin sector photon temperature when the SM photon temperature is $T$.

Likewise, entropy densities for each sector $i$ are defined as $s_i(T) = \frac{2\pi^2}{45} g_\star^{\rm i}(T) T^3$. We neglect the small differences between the number of relativistic degrees of freedom defined from energy and entropy densities, which are not significant over the range of tempera-

tures of interest here.

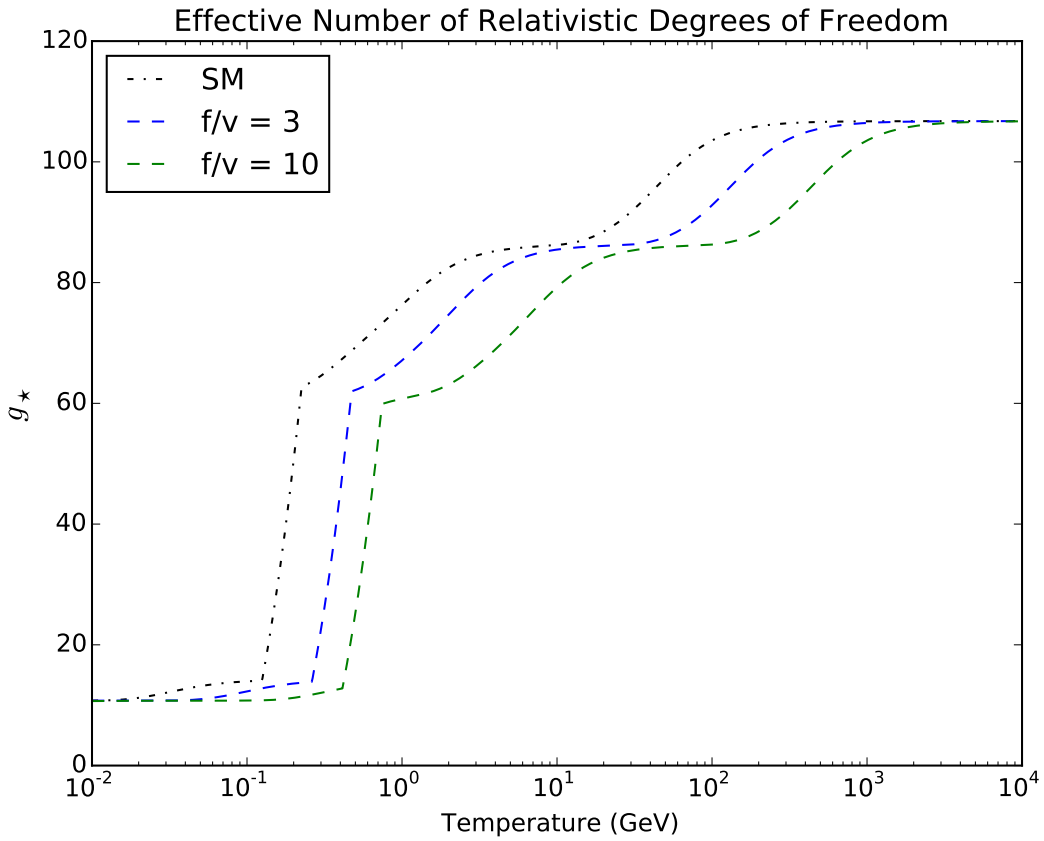



Figure 4.1: The effective number of relativistic degrees of freedom for mirror Twin Higgs models for different values of $f/v$. The dash-dotted line is the for the Standard Model $g_\star^{\rm SM}(T)$ and the dashed lines are the twin sector degrees of freedom $g_\star^{\rm t}(T)$. The evolution of $g_\star$ during the QCD phase transition (QCDPT) is not well-understood, so we assign the SM QCDPT a central value of 175 MeV and a width of 50 MeV and interpolate linearly between the values of $g_\star$ at 225 MeV for free partons and at 125 MeV for pions. Further discussion may be found in (21). For the twin sector we use a central value and width which are $(1 + \log(\frac{f}{v}))$ times larger than the SM values. Note that new mass thresholds, expected to appear at energies $\sim$ 10 TeV in UV completions of the twin Higgs, have not been included.

### 4.3.2 Decoupling

In the early universe, the two sectors are thermally linked by interactions mediated by the Higgs, which, through mixing with both $h_A$ and $h_B$ components, allows for

SM fermions and weak bosons to scatter off or annihilate into their twin counterparts. However, once the temperature drops sufficiently for this Higgs-mediated interaction to become rare on the expansion time-scale, the sectors decouple and thereafter thermally evolve independently. More precisely, thermal decoupling will occur once the rate at which energy can be exchanged between SM and twin particles (through the Higgs) falls below the Hubble rate.

Thermal decoupling is traditionally formulated from the Boltzmann equations describing the evolution of single-particle phase space number densities, wherein collisions induce instantaneous changes to the shape of these distributions. When the collisions occur faster than the expansion rate, the phase space probability density functions of the interacting species are expected to relax to an equilibrium distribution (Boltzmann, neglecting quantum statistics, will be applicable to our case). However, once the rate of collisions falls below the expansion rate, collisions become rare on cosmological time scales and the phase space distributions depart from equilibrium. The decoupling temperature is determined as that at which the scattering rate of a participating particle, $\Gamma$, drops below the Hubble rate, assuming that this occurs instantaneously across the entire phase space where the number density is significant. This formulation can be used to determine the time at which a particular species of particle will cease to scatter off twin particles on cosmological time scales.

In the case of interest here, however, both sectors of particles remain thermalised within themselves while the interactions between sectors freeze-out. This implies that the phase space number densities are still Boltzmann distributions throughout decoupling, with a different temperature for each sector. As it is the twin sector temperature that ultimately determines the impact of the light twin degrees of freedom on the cosmological observables (discussed below in Section 4.3.3), we wish to describe the thermal evolution of the two sectors by that of their entire energy or entropy content and the bulk heat

flows between them. They may then be identified as thermally decoupled once the rate at which they exchange energy falls below the expansion rate.

If the SM and twin sector plasmas have temperatures $T$ and $\hat{T}$ respectively, then calling $q$ the net heat flow density from the SM to the twin sector, the rate at which the twin entropy densities $s_{\rm t}$ and $s_{\rm SM}$ evolve is determined by

$$\frac{ds_{\rm t}}{dt} + 3Hs_{\rm t} = \frac{1}{\hat{T}}\frac{dq}{dt} = \frac{1}{\hat{T}}\Big(\frac{dq_{\rm in}}{dt} - \frac{dq_{\rm out}}{dt}\Big) \tag{4.3}$$

$$\frac{ds_{\rm SM}}{dt} + 3Hs_{\rm SM} = \frac{-1}{T}\frac{dq}{dt} = -\frac{1}{T}\Big(\frac{dq_{\rm in}}{dt} - \frac{dq_{\rm out}}{dt}\Big). \tag{4.4}$$

Here, $H$ is the Hubble rate. The heat flow rate has been decomposed into the sum of the energy transferred into and out of the twin sector by collisions in the second equality in each line, where $\frac{dq_{\rm in}}{dt}$ and $\frac{dq_{\rm out}}{dt}$ are both positive.

The rate of heat flow $q$ may be calculated by performing a phase space average of the rate that energy is transferred from the SM to the twin sector through particle interactions. Since the decay rates of top quarks or weak bosons are fast compared to their scattering rate and the Hubble rate, energy transferred to them is instantaneously transferred to the rest of the plasma. Similarly, the scattering rate of lighter fermions off other particles of the same sector (such as photons or gluons) is much faster than their interaction rate with twin fermions. Energy transferred to the lighter fermions therefore quickly diffuses throughout their respective plasmas. The rate of heat flow between sectors may therefore be well approximated by the rate at which energy is transferred from SM particles to twin particles in Higgs mediated interactions. This may occur through elastic scattering of SM particles off twin particles or annihilations of SM particle/antiparticle pairs into twin particles (or the reverse). The energy density

transferred to twin particle $i$ from SM particle $j$ in scattering is given by

$$\frac{dq_{ij\to ij}}{dt} = \frac{g_i g_j}{(2\pi)^6}\int\int \frac{d^3k}{2E_i(k)}\frac{d^3h}{2E_j(h)} f_i(k,\hat{T}) f_j(h,T) \\ \Big(4E_i(k)E_j(h)\int v_{rel}(E_i(p)-E_i(k))\frac{d\sigma_{ij\to ij}}{d\Omega}d\Omega\Big), \tag{4.5}$$

where $p$ is the outgoing 4-momentum of particle $i$. In the cosmic comoving frame, the phase space number densities $f_i$ and $f_j$ are just Boltzmann factors, although evaluated at the different temperatures of each sector. The factor $g_i$ is the number of internal degrees of freedom of particle $i$, which here includes colour (the cross section should not be colour averaged, as each colour of quark is present in the plasma in equal abundances and each mediates the exchange of energy, so have their contributions summed). Finally, $E_i(k)$ is the on-shell energy of particle $i$ with momentum $k$, while $\frac{d\sigma_{ij\to ij}}{d\Omega}$ is the differential scattering cross section for species $i$ scattering off $j$ per solid angle $\Omega$ and $v_{rel}$ is the usual relative speed of the incoming particles. As described in (22), the factor in the integrand giving the energy transferred per reaction is simply a component of a 4-vector,

$$X = 4E_i(k)E_j(h)\int (p-k)v_{rel}\frac{d\sigma_{ij\to ij}}{d\Omega}d\Omega. \tag{4.6}$$

This may be calculated in the centre-of-mass frame and then boosted back into the cosmic comoving frame where the integrals in (4.5) can be evaluated, similarly to the thermal averaging procedure described in (23).

The integral (4.5) may be decomposed into two terms giving the positive and negative energy changes of the twin particle, which respectively contribute to $\frac{dq_{\rm in}}{dt}$ and $\frac{dq_{\rm out}}{dt}$. When evaluated in the centre-of-mass frame, these terms correspond to the cases where the scattering angle of the twin particle is respectively less than and greater than the angle between its initial momentum and the total momentum of the system. However, when

$T \neq \hat{T}$, we find the integrals involved in this decomposition substantially more arduous than when they are evaluated together.

Energy transferred through annihilations may be similarly calculated as

$$\begin{aligned}\frac{dq_{j\bar{j}\to i\bar{i}}}{dt} = \frac{g_j^2}{(2\pi)^6}\int\int \frac{d^3k}{2E_j(k)}\frac{d^3h}{2E_j(h)} f_j(k) f_j(h) \\ \times \Big(4E_j(k)E_j(h)\int v_{rel}(E_j(h)+E_j(k))\frac{d\sigma_{j\bar{j}\to i\bar{i}}}{d\Omega}d\Omega\Big) \\ -\frac{g_i^2}{(2\pi)^6}\int\int \frac{d^3k}{2E_i(k)}\frac{d^3h}{2E_i(h)} f_i(k) f_i(h) \\ \times \Big(4E_i(k)E_i(h)\int v_{rel}(E_i(h)+E_i(k))\frac{d\sigma_{i\bar{i}\to j\bar{j}}}{d\Omega}d\Omega\Big),\end{aligned} \tag{4.7}$$

where $\frac{d\sigma_{j\bar{j}\to i\bar{i}}}{d\Omega}$ is now the differential annihilation cross section. This rate may be evaluated as described above and is more directly amenable to the factorisation of the integrals observed in (23). See also (24) for further details of similar calculations. The first term of (4.7) is the energy transferred from the SM to the twin sector and contributes to $\frac{dq_{\rm in}}{dt}$ in (4.3), while the second term is the energy transferred from the twin sector to the SM and contributes to $\frac{dq_{\rm out}}{dt}$.

In thermal equilibrium, the rate of energy transferred through collisions into one sector will be balanced by that of energy transferred out of it so that there is negligible net heat flow. This state will be rapidly attained (compared to the age of the universe) if $\frac{dq_{\rm in,out}}{dt} \gg 3H\hat{T}s_{\rm t}$. However, as the universe expands and the plasma cools, the energy transfer rates fall faster than the Hubble rate. This is demonstrated in the Figure 4.2 below. Once they drop below the Hubble rate, energy exchange ceases on cosmological time scales and the sectors thermally decouple, thereafter thermodynamically evolving independently.

To determine the decoupling temperature of the sectors, we calculate the rates of positive energy exchange for the twin particles interacting with the SM particles. The cross

sections are calculated using a tree-level effective fermion-twin fermion contact interaction that, in the full twin Higgs model, would be UV completed by a SM Higgs exchange (the heavier mass of the radial mode would make its exchange subdominant). The interaction strength is determined by the masses of the fermions through their Yukawa couplings, as well as the mixing angle of the SM-like mass state $h$ with the gauge eigenstate $h_B$, giving a 4-fermion coupling of strength $\frac{m_f m_{\hat{f}}}{m_h^2 f^2}$ (here $m_f$ and $m_{\hat{f}}$ are the masses of fermions $f$ and $\hat{f}$). See (19), (11) for a more detailed discussion of the cross sections. This effective interaction is appropriate for the temperatures of interest here and helps to simplify the integrals of (4.5). In order to further simplify the integrations of (4.5) when it is to be decomposed into terms in which the energy exchange is positive and negative, we calculate $\frac{dq_{\rm in}}{dt}$ under the assumption that the sectors have the same temperature (this ensures that the rate $\frac{dq_{\rm out}}{dt}$ is identical). This is then combined with the rate of energy transferred from annihilation. A similar calculation of these rates was recently performed in (11), for cases where the Yukawa couplings do not respect the $\mathbb{Z}_2$ twin symmetry.

In Figure 4.2 we compare the energy transfer rate to the Hubble rate in order to determine when decoupling occurs. As long as the energy exchange rate exceeds the expansion rate, the sectors will be thermalised and have the same temperature. Decoupling then occurs once this rate drops below the Hubble rate. From Figure 4.2, this occurs at a temperature $\sim 2$ GeV. However, even after the energy exchange rate drops below the Hubble rate, the sectors will remain at the same temperature unless some event that either injects or redistributes entropy occurs within a sector (such as the temperature dropping below a mass threshold). As the heavy quark masses roughly coincide with the decoupling temperature, these do cause the twin sector to be mildly reheated with respect to the SM below decoupling. However, the resulting temperature difference is small and the energy exchange rates are expected to continue to be well-approximated by the rates presented in Figure 4.2 beyond decoupling.

The lower plot of Figure 4.2 illustrates the decomposition of the energy exchange rates into contributions from interactions involving different SM quarks. The interaction cross sections are proportional to the Yukawa couplings of the interacting fermions. The greatest heat exchange is therefore expected to be mediated by the most massive particles, provided that their abundances are not too Boltzmann suppressed. As expected, at temperatures $\sim 1$ GeV, the bottom quark is the best conduit of thermal equilibration, followed by the charm quark and then the $\tau$ (with colour factors enhancing the former two with respect to the latter). The rate of heat flow that the top quarks and weak bosons can mediate at these temperatures (or below) is negligible because of Boltzmann suppression. The bend in the curves at temperatures $\sim 5$ GeV in the lower plot corresponds to a transition from temperatures where the dominant energy exchange rate is through scatterings to those where it occurs through annihilations, as can be seen in the upper plot. The annihilation rate into twin bottom quarks is the dominant component at high enough energies (again because of the larger Yukawa coupling), but this becomes rapidly threshold suppressed as the temperature drops. As can also be inferred in the upper plot, the energy exchange rate through annihilations involving the twin charmed quarks and tau leptons overtakes that of twin bottom quarks at similar temperature, but are still subdominant to scatterings.

The decoupling temperature depends upon $f/v$, which sets both the mass scale of the twin sector and the strength of the Higgs-mediated coupling. As $f/v$ is increased, decoupling occurs earlier because of the greater Boltzmann suppression, although this is only a relatively small effect that, for $f/v = 10$, increases the decoupling temperature by only 4 GeV.

When the twin sector is colder than the SM (which will be important for much of what follows) the heat flow is typically dominated by annihilations of SM into twin particles. However, the energy exchange from elastic scattering can be comparable to

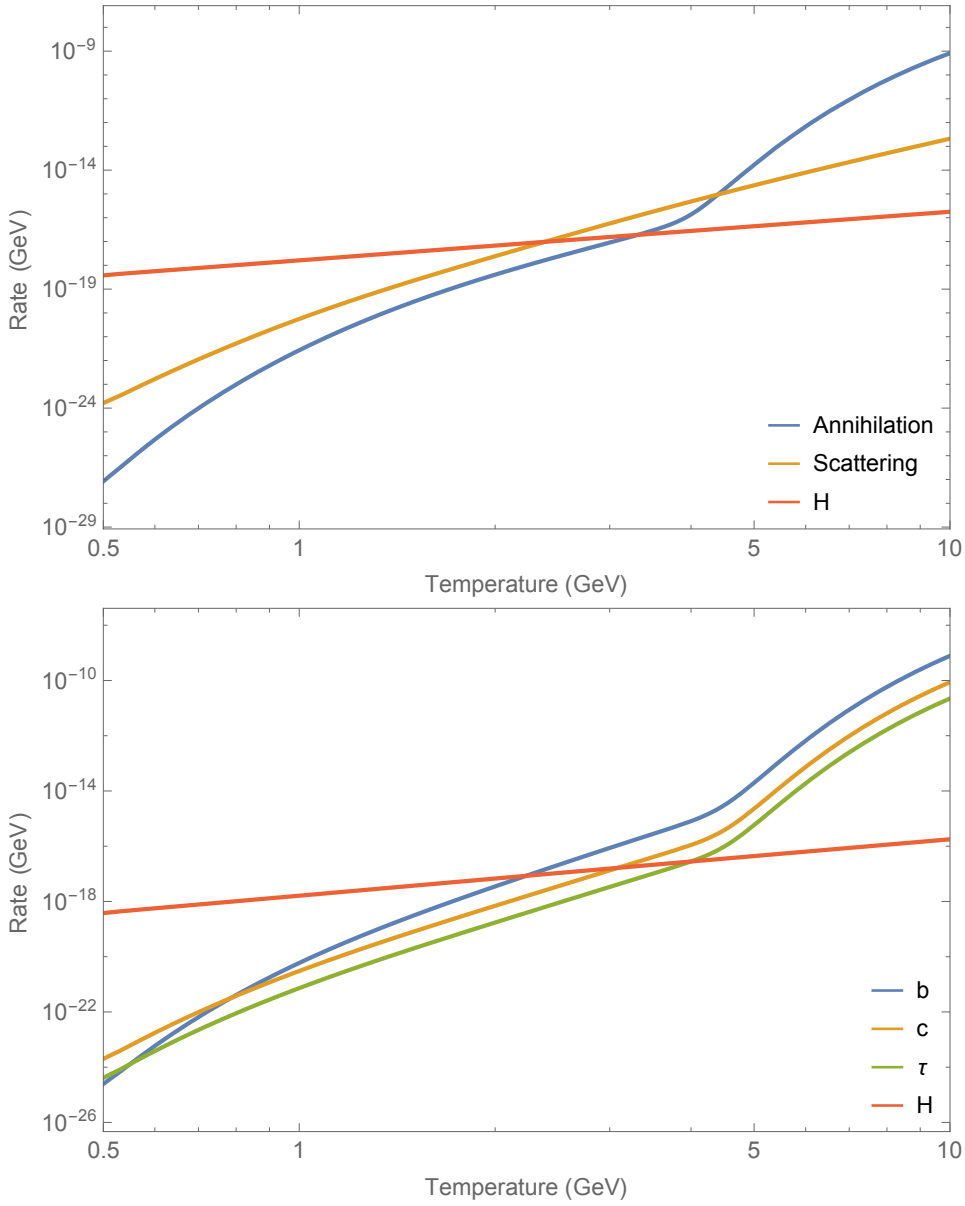



Figure 4.2: Rates of energy density exchange per twin entropy density ($\frac{1}{3s_t\tilde{T}}\frac{dq_{\rm in}}{dt}$) decomposed into contributions from scattering and annihilation (top) and for interactions involving different species of SM fermions (bottom), along with the Hubble parameter, for $f/v = 4$. The decoupling temperature is that where the sum of the energy exchange rates equals the Hubble rate, which occurs at $T_{\rm decoup} \approx 2$ GeV.

that from annihilations, as illustrated in Figure 4.2. Although the energy exchange in an annihilation will generally exceed that of a scattering because all of the energy involved in the process must be transferred, the annihilation rate also becomes more Boltzmann

or threshold suppressed when the temperature drops below the mass of the heavier twin particles. It is therefore not always clear that energy transfer through annihilations dominates.

Decoupling is not exactly instantaneous and there is some range of temperatures over which the rate of heat flow freezes-out. The net heat flow rate $\frac{dq}{dt}$ is greater for larger temperature differences between sectors. The generation of a potentially large temperature difference within this brief epoch of sector decoupling, such as those discussed below in Section 4.4, may be cut off when the heat flow rate becomes comparable to the Hubble rate. For a given SM temperature $T$, the minimum twin-sector temperature $\hat{T}_{\rm min}$ during the decoupling period may be roughly estimated as that which satisfies

$$H \sim \frac{1}{3s_{\rm t}\hat{T}}\frac{dq}{dt}\Big|_{\hat{T}=\hat{T}_{\rm min}}. \tag{4.8}$$

Twin temperatures colder than $\hat{T}_{\rm min}$ will partially thermalise back to this value. As the participating fermions are not non-relativistic, instantaneous decoupling is not as accurate an approximation as it is, for example, for chemical decoupling of a WIMP, although it is still reliable.

In Figure 4.3, we show the minimum temperature that the twin sector may have as a function of SM temperature for heat flow to freeze out, estimated using (4.8). Only annihilations have been included in the determination of the minimum temperature, although we have verified that, for these temperatures, the scatterings contribute only $\lesssim 10\%$ to the heat flow. Note that while the energy exchange rate, such as $\frac{1}{\hat{T}}\frac{dq_{\rm in}}{dt}$ in (4.3), in scattering processes may be faster, the net energy flow rate, or heat flow ($\frac{1}{\hat{T}}\frac{dq}{dt}$ in (4.3)), which is the difference between energy exchange rates into and out of the sector, is actually dominated by annihilations. Generally, we find that decoupling begins at temperatures $\sim 4$ GeV. The temperature difference can reach an order of magnitude

without relaxing once the SM temperature drops to $\sim 1$ GeV.

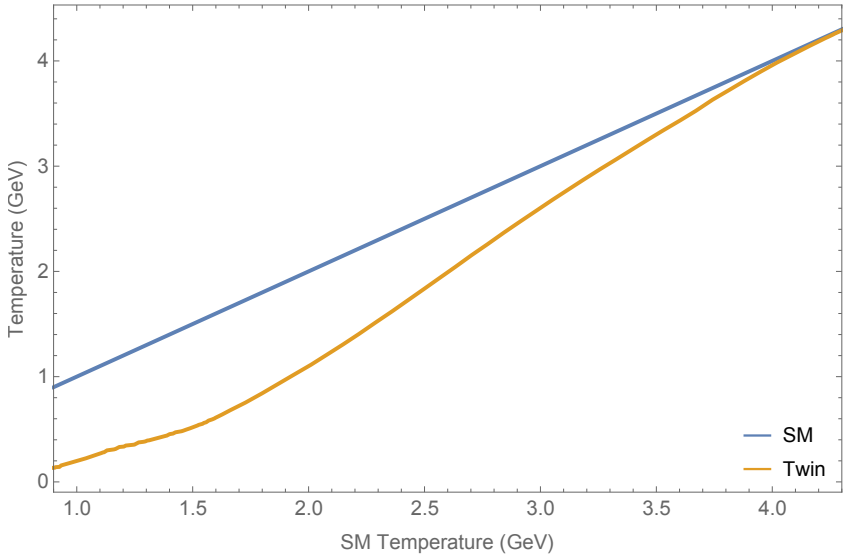



Figure 4.3: Minimum temperature of the twin sector that will not be heated by interactions with a hotter SM plasma, as a function of SM temperature, for $f/v = 4$. Also shown is the SM temperature, for reference.

While the extent of thermal decoupling is temperature dependent, the maximum temperature difference that will not relax grows quickly as the SM temperature drops. Then we may describe the two sectors as being decoupled if, in a given cosmology, all events that raise the temperature of one sector relative to the other (such as the crossing of a mass threshold and the resulting entropy redistribution, the most significant of which is the confinement of colour) induce temperature differences that are too small to partially relax.

At energies $\lesssim 1$ GeV in Figure 4.2, the reliability of the calculation of the heat flow rate diminishes because of the strengthening of the strong coupling and the eventual confinement of colour. Fortunately, for a cooler twin sector, which will be of interest in subsequent sections, annihilations from the SM dominate other processes over most of the parameter space. These are the least sensitive to higher order corrections and non-perturbative effects because of their higher temperature, and hence energy, compared

to the potentially cooler twin sector. The range of temperatures illustrated in Figures 4.2 and 4.3 have been selected to roughly illustrate the duration of decoupling, but may extend below the range where the perturbative calculation of the heat flow rate is valid. For example, at temperatures below the twin sector QCDPT, which occurs at $\sim \left(1 + \log(\frac{f}{v})\right)$ higher temperatures than in the SM, the partonic calculation of twin quark/anti-quark pair production must be replaced by a hadronic one. Furthermore, the growth of the twin strong coupling necessitates that the quark-Higgs Yukawa couplings be RG evolved to the scale of the energy exchanged, which can induce an $\mathcal{O}(1)$ change to the cross section, although this has only a relatively small effect on the decoupling temperature. It is nevertheless clear that decoupling is mostly complete by then and that these uncertainties are not large enough to affect this conclusion.

In the standard mirror Twin Higgs cosmology, knowing the decoupling temperature tells us how the temperatures of the two sectors will be related at subsequent times. The sectors separately evolve adiabatically after decoupling, though they redshift in the same way and differences in temperature only arise from events that redistribute entropy. Non-minimal cosmological events that could potentially cause the temperatures of each sector to diverge can therefore only be effective if they leave each sector colder than this approximate decoupling temperature.

### 4.3.3 Cosmological Constraints

Given that the twin and Standard Model sectors remain in thermal equilibrium to $\mathcal{O}(\text{GeV})$ temperatures, the simplest mirror Twin Higgs scenario is cosmologically inviable due to the presence of light twin species (photons and neutrinos) with abundances comparable to those of the SM. The cosmological observables through which evidence of light species may be inferred are typically represented by $N_{\text{eff}}$, the "effective number of

neutrino species" in the early universe; their individual masses, which determine their free-streaming distances; and the "effective mass" $m_\nu^{\rm eff}$, which parameterises their contribution to the present-day energy density of non-relativistic matter. These observables are probed by both the CMB and large scale structure (LSS).

**Effective number of neutrinos**

The parameter $N_{\rm eff}$ describes the amount of radiation-like energy density during the evolution of the CMB anisotropies before photon decoupling. It is defined as the effective number of massless neutrinos with temperature as predicted in the standard cosmology that would give equivalent energy density in radiation:

$$\rho_r = \rho_\gamma + \frac{7}{8}\left(\frac{4}{11}\right)^{4/3} N_{\rm eff}\rho_\gamma, \tag{4.9}$$

where $\rho_r$ is the energy density of radiation and $\rho_\gamma$ is the energy density of photons (the factor of $\left(\frac{4}{11}\right)^{4/3}$ arises from the relative reheating of the photons from electron/positron annihilation, which occurs after most of the neutrinos have decoupled, and the factor of 7/8 is from the opposite spin statistics). A deviation from the Standard Model prediction of 3.046 (25) is denoted by $\Delta N_{\rm eff} = N_{\rm eff} - 3.046$. This definition of radiation, or equivalently, relativistic degrees of freedom, becomes less clear if the new fields have a non-negligible mass, as we discuss further below.

We here review the CMB physics of dark radiation, summarising the discussion in (26). See also (2) for further review. The angular size and scale of the first acoustic peak is well-measured and this approximately fixes the scale factor at matter-radiation equality $a_{eq}$. If we imagine fixing all other ΛCDM parameters, extra radiation would delay the epoch of matter-radiation equality. This would have a pronounced effect on the power spectrum in the vicinity of the first acoustic peak through the early Integrated

Sachs-Wolfe (eISW) effect. The modes corresponding to this feature entered the horizon close to matter-radiation equality and the evolution of their potentials is highly sensitive to the radiation energy density. However, the impact of a $\Delta N_{\rm eff} \sim \mathcal{O}(1)$ deviation on the peak height can be simultaneously balanced by increasing the amount of non-relativistic matter, to the extent to which other observations providing independent constraints upon $\Omega_c$ permit (for $\Lambda$CDM+$N_{\rm eff}$, a variation of $\sim 10\%$ in $\Omega_c h^2$ is consistent with present CMB+BAO measurements (2), although these variations must be consistent with other observables). This degeneracy is not expected to be broken by CMB-S4 (27).

Given that $a_{eq}$ is approximately fixed, the utility of $N_{\rm eff}$ arises because, in simple extensions of the $\Lambda$CDM model, it approximately corresponds to the suppression of power in the small scale CMB anisotropies that arises from Silk damping. The reason for this is roughly that, although the greater expansion rate induced by the extra radiation reduces the time that CMB photons have to diffuse before decoupling, it also reduces the sound horizon size more severely. As the angular size of the sound horizon is determined by the location of the acoustic peaks and is also well measured, the reduction in the sound horizon must be compensated for by a reduction in the angular diameter distance to the CMB. This effectively raises the angular distance over which photon diffusion proceeds and results in a prediction of smoother temperature anisotropies at small scales. This correspondence with the Silk damping allows $N_{\rm eff}$ to be approximately factorised from other parameters and constrained independently, providing a direct observational avenue for detecting the presence of new, massless fields (26) (see (28) for further implications for model building). This relationship arises because the fixing of $a_{eq}$ implies that $N_{\rm eff}$ effectively determines the energy density of the universe, and hence the Hubble rate, during CMB decoupling. Note, however, that further extensions of $\Lambda$CDM may complicate this correspondence, in particular deviations from the standard Big Bang Nucleosynthesis prediction of the primordial helium abundance.

The contribution to $N_{\rm eff}$ (or $\Delta N_{\rm eff}$) in the mirror Twin Higgs arises from two sources: the twin photons, which can be treated as massless dark radiation with an appropriate twin temperature $T^{\rm t}_{\rm eq}$ at the time of matter-radiation equality, and the twin neutrinos, whose non-zero masses may need to be accounted for. For the twin photons, the contribution to $N_{\rm eff}$ is simple; their equation of state is always $w = 1/3$ and their energy density is given by $g\frac{\pi^2}{30}\left(T^{\rm t}_{\rm eq}\right)^4$, where $g = 2$. The twin temperature at matter-radiation equality is found from the SM temperature using comoving entropy conservation,

$$\frac{T^{\rm t}_{\rm eq}}{T^{\rm SM}_{\rm eq}} = \left(\frac{g^{\rm t}_\star(T_{\rm decoup})}{g^{\rm SM}_\star(T_{\rm decoup})}\right)^{1/3}\left(\frac{g^{\rm SM}_\star(T^{\rm SM}_{\rm eq})}{g^{\rm t}_\star(T^{\rm t}_{\rm eq})}\right)^{1/3}, \tag{4.10}$$

where the two sectors have the same number of thermalized degrees of freedom by this time. Here, $T^{\rm SM}_{\rm eq}$ is the SM photon temperature at matter-radiation equality and $T_{\rm decoup}$ is the sector decoupling temperature.

Since neutrinos are massive, their behavior is more complicated. Their equation of state parameter takes on a scale factor dependence which is controlled by their mass. In the Standard Model, this sensitivity is negligible because present CMB bounds imply that neutrinos are ultra-relativistic at $a_{\rm eq}$ to good approximation (2). However, the factor by which the twin neutrino masses are enhanced may raise them to order $T^t_{\rm eq}$ or greater (see Section 4.2 for discussion of the scaling of the masses with $f/v$).

To better describe the impact of the extra twin (semi-)relativistic degrees of freedom on the CMB, we choose to define $N_{\rm eff}$ through the effects of neutrinos at matter-radiation equality, when the impact on the expansion rate of the universe for most of the period relevant for the evolution of the CMB is greatest. Note that, in their presentation of joint exclusion bounds on $N_{\rm eff}$ and $\sum m_\nu$ (the sum of SM neutrino masses) or $m^{\rm eff}_\nu$ (effective mass contributing to the present-day non-relativistic matter density of an extra sterile neutrino), the Planck collaboration define $N_{\rm eff}$ as the value in (4.9) at temperatures

sufficiently high that the neutrinos are fully relativistic. Our values cannot be directly compared with their analysis, although we consider ours to be a reasonable rough estimate that is more representative of the CMB constraints. The ensuing correction from the finite neutrino masses is, in the cases considered in this work, a small effect anyway.

To determine this correction and provide a definition of $N_{\rm eff}$ that better describes the impact of quasi-relativistic particles on the CMB, we first define the epoch of matter-radiation equality as the time at which the average equation of state parameter of the universe is $\bar{w} = 1/6$ (the equation of state is defined as $\rho = \bar{w}P$, where $\rho$ is energy density and $P$ is pressure). We can express this condition as

$$\left.\frac{d\ln H}{d\ln a}\right|_{a_{eq}} = -\frac{7}{4}, \tag{4.11}$$

as in (29). Call the quasi-relativistic neutrino energy density $\tilde{\rho}(a)$ with time-evolving equation of state parameter $w(a)$, which is to be balanced against some extra non-relativistic energy density $\Delta\rho_{CDM}(a) \propto a^{-3}$ to keep $a_{eq}$ the same. This amount of non-relativistic energy density $\Delta\rho_{CDM}$ is

$$\Delta\rho_{CDM}(a_{eq}) = \rho_r(a_{eq}) - \rho_m(a_{eq}) - 2a_{eq}\left.\frac{d\tilde{\rho}}{da}\right|_{a_{eq}} - 7\tilde{\rho}(a_{eq}), \tag{4.12}$$

where $\rho_r$ and $\rho_m$ are the energy densities of the radiation and non-relativistic matter. For a perfect fluid, $\frac{d\tilde{\rho}}{da} = -3(1+w(a))\tilde{\rho}/a$ (neglecting the anisotropic stress that is expected only to contribute to a weak phase shift in the CMB (30)), this results in a Hubble parameter of

$$H^2(a_{eq}) = \frac{2}{3M_{\rm pl}^2}\left[\rho_r(a_{eq}) + 3w(a_{eq})\tilde{\rho}(a_{eq})\right]. \tag{4.13}$$

This suggests a definition of the effective number of neutrinos, $N_{\rm eff}$, via

$$H^2(a_{eq}) = \frac{2}{3M_{\rm pl}^2}\left(\rho_\gamma + N_{\rm eff}\rho^{th}_{\nu,m=0}\right)\Big|_{a_{eq}} \tag{4.14}$$

$$N_{\rm eff} \equiv \sum_i \frac{w_i}{1/3}\frac{\rho_i}{\rho^{th}_{\nu,m=0}}, \tag{4.15}$$

where $\rho_i$ is the contribution to the energy density from some species $i$ with equation of state parameter $w_i$ and $\rho^{th}_{\nu,m=0}$ is the energy density of a massless neutrino with a thermal distribution in the standard cosmology. Then $3w$ gives the 'relativistic fraction' of the energy density. Note that this is simply a ratio of the pressure exerted by the new fields to that of a massless neutrino. The effectiveness of this approximation was discussed in (31) in the context of thermal axions (while effective at keeping $a_{eq}$ fixed, changes to odd peak heights subsequent to the first are imperfectly cancelled and require further changes to $H_0$ to compensate - see Section 4.3.3 below).

Calling $T^i_\nu$ the temperature at which the neutrinos in sector $i$ freeze-out and $a^i_\nu$ the corresponding scale factor, then assuming instantaneous decoupling, the phase space number density for scale factor $a$ is given by a redshifted Fermi-Dirac distribution (32)

$$f^i_\alpha(p) \approx \left[1 + e^{pa/\left(a^i_\nu T^i_\nu\right)}\right]^{-1} \tag{4.16}$$

for the $\alpha$ neutrino mass eigenstate in the $i$ sector ($m^i_\alpha \ll T^i_\nu$, so has been dropped). The energy density and pressure are

$$\rho^i_{\nu_\alpha} = \frac{g_\alpha}{2\pi^2}\int_0^\infty dp\; p^2\sqrt{p^2 + \left(m^i_\alpha\right)^2} f^i_\alpha(p) \tag{4.17}$$

$$P^i_{\nu_\alpha} = \frac{g_\alpha}{2\pi^2}\int_0^\infty dp\; \frac{p^4}{3\sqrt{p^2 + \left(m^i_\alpha\right)^2}} f^i_\alpha(p), \tag{4.18}$$

where $g_\alpha = 2$ is the number of degrees of freedom for a neutrino species.

Since the neutrino decoupling temperature depends on the strength of the weak interaction as $T_\nu \propto G_F^{-2/3}$, while $G_F \propto v^2$, then the twin neutrino decoupling temperature $T_\nu^{\rm t}$ is related to the SM neutrino decoupling temperature $T_\nu^{\rm SM}$ by

$$T_\nu^{\rm t} = (f/v)^{4/3} T_\nu^{\rm SM}. \tag{4.19}$$

We can then simply use (4.17) and (4.18) at matter-radiation equality to find $\Delta N_{\rm eff}$ (assuming instantaneous decoupling). We thus obtain

$$H^2(a_{eq}) = \frac{2}{3M_{\rm pl}^2}\left(\rho_\gamma^{\rm SM} + 3.046\rho_{\nu,m=0}^{th} + \rho_\gamma^{\rm t} + \sum_\alpha 3w_{\nu_\alpha}\rho_{\nu_\alpha}^{\rm t}\right)\Bigg|_{a_{eq}} \tag{4.20}$$

and

$$\Delta N_{\rm eff} = \left(\frac{11}{4}\right)^{4/3} \frac{120}{7\pi^2\left(T^{\rm SM}\right)^4}\left(\rho_\gamma^{\rm t} + \sum_\alpha 3w_{\nu_\alpha}^t\rho_{\nu_\alpha}^{\rm t}\right), \tag{4.21}$$

where we now have equation of state parameters $w_{\nu_\alpha}$ for each neutrino, while $\rho_\gamma^{\rm SM}$ and $\rho_\gamma^{\rm t}$ are the SM and twin photon energy densities, $\rho_{\nu,m=0}^{th}$ and $\rho_{\nu_\alpha}^{\rm t}$ are the neutrino energy densities.

**Neutrino masses**

Because they are so weakly interacting, the neutrinos have a long free-streaming scale given by the distance travelled in a Hubble time $v_\nu/H$, with $v_\nu \propto m_\nu^{-1}$ the speed of the neutrino once it becomes non-relativistic. This defines a free-streaming momentum scale $k_{fs} = \sqrt{\frac{3}{2}}\frac{aH}{v_\nu} \propto m_\nu$, above which neutrinos do not cluster. Below this scale, perturbations in the matter density consist coherently of neutrinos and other matter, but well above it only non-neutrino matter contributes to density perturbations. This

results in a suppression of the matter power spectrum on large scales which is proportional to the fraction of energy density in the free-streaming matter. Since this occurs at late times when neutrinos are non-relativistic, the energy density is simply $\rho_{\nu_\alpha} = n_{\nu_\alpha} m_{\nu_\alpha}$ for each neutrino species $\alpha$, where $n_{\nu_\alpha}$ is the number density. Constraints on the sum of neutrino masses then come from the observations of power on small scales, which is suppressed relative to that expected for massless neutrinos by a factor $1 - 8f_\nu$, where $f_\nu = \Omega_\nu/\Omega_m$ is the fraction of non-relativistic energy in neutrinos at late times (33).

More generally, inferences of the matter power spectrum constrain the present-day energy density fraction of free-streaming species that do not cluster on small scales and have since become non-relativisitic, $\Omega_\nu = (\sum m_\nu + m_\nu^{\rm eff})/(94.1\,{\rm eV})$, where $\sum m_\nu$ is the sum of SM neutrino masses and $m_\nu^{\rm eff}$ is the sum of twin neutrino masses weighted by their number density

$$m_\nu^{\rm eff} = \frac{n_\nu^{\rm t}}{n_\nu^{\rm SM}} \sum_\alpha m_{\nu_\alpha}^{\rm t}. \tag{4.22}$$

Here $n_\nu^{\rm t}$ is the number density of a relic twin neutrino flavour and $n_\nu^{\rm SM}$ is that for a SM neutrino. It is assumed that the neutrinos have been thermally produced as hot relics.

The relic abundance of a neutrino species is given by its number density when it decoupled, diluted by the factor by which the universe has since expanded. The scale factors at which neutrino decoupling occurs in the two sectors, $a_\nu^{\rm SM}$ and $a_\nu^{\rm t}$ can be determined from (4.19), the relative temperatures in the two sectors and comoving entropy conservation, to obtain

$$a_\nu^{\rm t} \;=\; a_\nu^{\rm SM} \left(\frac{v}{f}\right)^{4/3} \left(\frac{g_\star^{\rm t}\left(T_{\rm decoup}\right)}{g_\star^{\rm SM}\left(T_{\rm decoup}\right)}\right)^{1/3} \tag{4.23}$$

where the same mass thresholds have been assumed in each sector below their neutrino decoupling temperatures, so that $g_\star^{\rm SM}\left(T_\nu^{\rm SM}\right) = g_\star^{\rm t}\left(T_\nu^{\rm t}\right)$. The neutrino number densities

are then

$$\frac{n_\nu^{\rm t}}{n_\nu^{\rm SM}} = \left(\frac{T_\nu^{\rm t} a_\nu^{\rm t}}{T_\nu^{\rm SM} a_\nu^{\rm SM}}\right)^3 = \frac{g_\star^{\rm t}\left(T_{\rm decoup}\right)}{g_\star^{\rm SM}\left(T_{\rm decoup}\right)}. \tag{4.24}$$

For $f/v$ from 3 to 10 and using $T_{\rm decoup} \sim 2 - 6$ GeV from Section 4.3.2, we find $g_\star^{\rm t}\left(T_{\rm decoup}\right) / \, g_\star^{\rm SM}\left(T_{\rm decoup}\right) \sim 0.8$ and thus arrive at

$$m_\nu^{\rm eff} \approx 0.8 \left(\frac{f}{v}\right)^n \sum_\alpha m_{\nu_\alpha}^{\rm SM}, \tag{4.25}$$

where $n = 1$ for Dirac masses and $n = 2$ for Majorana masses.

If they are sufficiently light and hot, the twin neutrinos only affect the CMB as dark radiation and their masses may then only be inferred from tests of the matter power spectrum. However, if heavier and colder, they are better described as a hot dark matter component. Their impact on the CMB is discussed in (34), where the shape of the power spectrum can depend upon the individual neutrino kinetic energies through their characteristic free-streaming lengths. The early Integrated Sachs-Wolfe effect (eISW) is also sensitive to the masses if the neutrinos become non-relativistic during decoupling (thereby affecting the radiation energy density and the growth of inhomogeneities) (33).

There is a significant degeneracy in cosmological fits to the CMB between $\Omega_m$ and $H_0$ (the Hubble constant) (35), where raising the non-relativistic matter fraction, such as with nonrelativistic neutrinos, can be accommodated by a decrease in $H_0$ (or equivalently, the dark energy density), which keeps the angular diameter distance to the CMB approximately fixed. This degeneracy can be broken by measurements of the baryon acoustic oscillations (BAOs), which are sensitive to the expansion rate of the late universe and provide an independent measurement of $\Omega_m$ and $H_0$. It is through combination with these results that bounds from Planck on neutrino masses are strongest (2).

### Bounds

The authors are unaware of any specialised analysis of the present and projected future cosmological constraints on scenarios with both massless dark radiation and additional light, semi-relativistic sterile neutrinos. In the absence of this, we use bounds from (2) as a rough indication of the present level of sensitivity to these parameters, which we nevertheless expect to be a reliable indication of the (in)viability of this model. The 95% confidence limits on these parameters are $N_{\rm eff} = 3.2 \pm 0.5$ and $\sum m_\nu < 0.32$ eV when each are constrained separately with the other fixed. This, of course, overlooks correlations between the impacts of masses and $\Delta N_{\rm eff}$ on the CMB and LSS. Bounds on an additional sterile neutrino as the only source of dark radiation are also presented with number density, or equivalently, contribution to $\Delta N_{\rm eff}$, left to float. These are similar to the limit on $\sum m_\nu$. It was found in (36) that, allowing $\sum m_\nu$ and $m_\nu^{\rm eff}$ to float independently for a single extra sterile neutrino, the bound mildly relaxes to $m_\nu^{\rm eff} \lesssim 1$ eV, although the bound may be stronger depending on the combination of data sets chosen (the lensing power spectrum presently prefers higher neutrino masses and raises the combined bounds if included). Other bounds from LSS on $\sum m_\nu$ exist and are potentially stronger than those placed from the CMB, possibly as low as $m_\nu^{\rm eff} \lesssim 0.05$ eV, again depending on data sets combined (see (37), (38)), although these are subject to greater uncertainties in the inference of the power spectra of dark matter halos from galaxies surveys and the Ly$\alpha$ forest.

It must also be noted that the shape of the CMB temperature anisotropies depends upon both the mass of individual neutrino components (through their free-streaming distance) and their contribution to the energy density of the nonrelativistic matter that does not cluster on small scales. However, it is not expected that improvements in bounds on the former will be made from improved measurements of the primary CMB itself,

but rather from weak lensing of the CMB, in conjunction with future measurements from DESI of the BAOs to break degeneracy with $\Omega_m$. The lensing spectrum, like inferences of the matter power spectrum made in galaxy surveys, is expected to measure the suppression of small scale power and therefore to strengthen constraints upon $m_\nu^{\rm eff}$, rather than the individual neutrino masses. One of the goals of CMB-S4 will be the detection of neutrino masses, given the present lower bound $\sum m_\nu \gtrsim 0.06$ eV from oscillations. Projected bounds are as low as $\sim 0.02$ eV (27), although this assumes no extra dark radiation or sterile neutrinos. A projection of the joint bound on $N_{\rm eff}$ (from extra massless dark radiation) and $m_\nu^{\rm eff}$ combining improved measurements CMB temperature measurements, lensing and BAOs indicates a limit of $m_\nu^{\rm eff} \lesssim 0.1$ eV at $1\sigma$ (27). Any contribution from additional states to $m_\nu^{\rm eff}$ may therefore be testable and bounded by the excess of the neutrino mass inference over the minimum neutrino mass, although laboratory measurements or measurements of $\Delta N_{\rm eff}$ will be required to further ascertain the contribution from the new particles.

Constraints on $\Delta N_{\rm eff}$ from improved measurements of the damping tail as part of CMB-S4 are projected to be $\sim 0.02 - 0.05$ at $1\sigma$ (27). In the following sections, we use an optimistic estimate of 0.02 for its reach in order to identify as much of the potentially testable parameter space as possible.

To estimate the impact of current and projected CMB limits on the mirror Twin Higgs, we consider two scenarios: the minimal Standard Model neutrino mass spectrum of $m_\nu = [0.0, 0.009\text{ eV}, 0.06\text{ eV}]$ and a degenerate spectrum of $m_\nu = [0.1\text{ eV}, 0.1\text{ eV}, 0.1\text{ eV}]/3$ from (2). In Figure 4.4 we plot the predictions of the mirror Twin Higgs for $\Delta N_{\rm eff}$ and $m_\nu^{\rm eff}$ for both types of spectra, as well as for both Dirac and Majorana masses (which scale differently with $f/v$). As is plainly evident, the mirror Twin Higgs is ruled out cosmologically, no matter the choices of neutrino masses one makes, if only for the presence of the twin photon. In the standard cosmology, the twin sector will have roughly

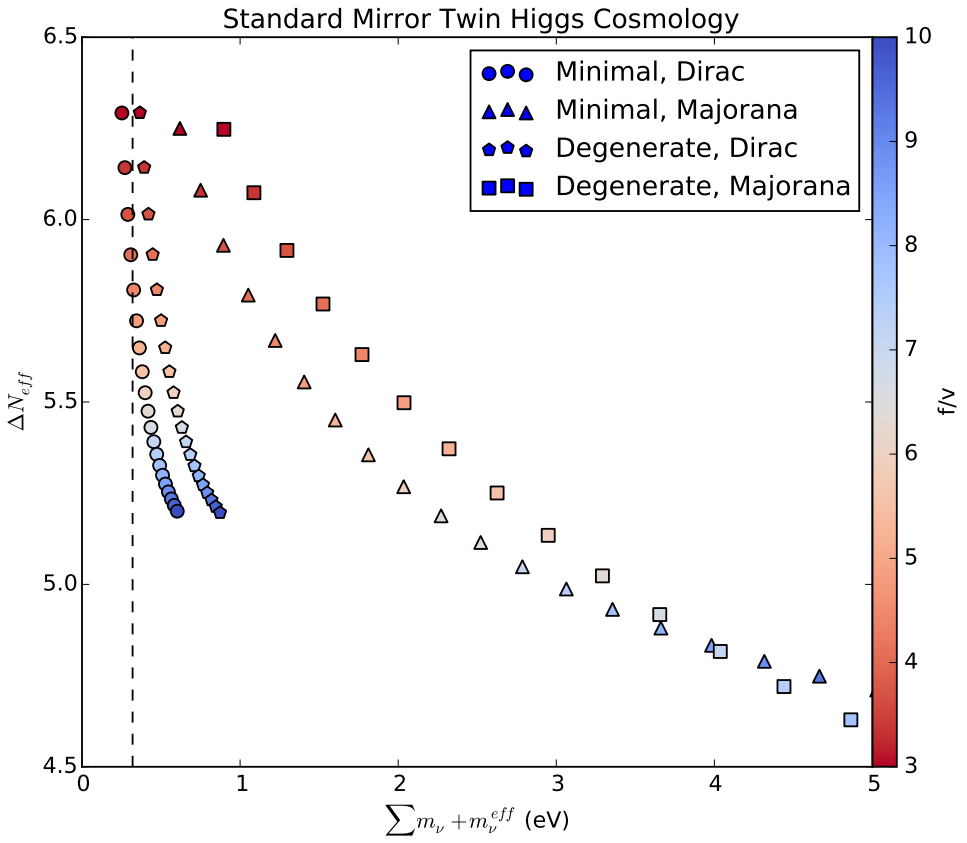



Figure 4.4: Predicted values of $\Delta N_{\rm eff}$ and $\sum m_\nu + m_\nu^{\rm eff}$ for minimal and degenerate neutrino mass spectra with both Dirac and Majorana masses for $f/v$ from 3 to 10. The Planck 2015 constraint(2) is the dashed line; the corresponding $N_{\rm eff}$ upper bound is well below the bottom of the plot. All points are excluded by the combination of bounds on $\Delta N_{\rm eff}$ and $\sum m_\nu + m_\nu^{\rm eff}$.

the same temperature as the SM, giving $4.6 \lesssim \Delta N_{\rm eff} \lesssim 6.3$ for $f/v < 10$, according to the definition of (4.21). This range depends upon $f/v$ through the twin neutrino decoupling temperature (4.19), which determines the extent to which the twin photons are reheated relative to the twin neutrinos after twin electron/positron annihilations. This is sufficiently large that even the cold dark matter fraction cannot be adjusted to keep matter-radiation equality fixed, resulting inevitably in changes to the height and shape of the first acoustic peak. The energy density in neutrinos is predicted to be above the present observational upper bounds for most neutrino mass configurations, with the

exception of the minimal values permitted by neutrino oscillation measurements with $f/v \lesssim 6$. We therefore discuss cosmological mechanisms in which the twin radiation is diluted to levels compatible with these observational bounds in the subsequent sections of this paper.

## 4.4 Reheating by the decay of a scalar field

We now turn to simple scenarios that reconcile the mirror Twin Higgs with cosmological bounds, while taking care to respect the softly-broken $\mathbb{Z}_2$ symmetry. We begin with the out-of-equilibrium decay of a particle with symmetric couplings to the Standard Model and twin sectors, in which the desired asymmetry is generated kinematically. That is to say, the dimensionless couplings between the decaying particle and the two sectors are equal, and asymmetric energy deposition into the two sectors is a direct consequence of the asymmetric mass scales. In this respect, the scenario is philosophically similar to $N$naturalness (39), albeit with a parsimonious $N = 2$ sectors. See also (40), (41) and (24) for other recent related ideas of using long-lived particles for the dilution of dark sectors.

For simplicity, here we will focus on the case of a real scalar $X$ coupled symmetrically to the $A$ and $B$ sector Higgs doublets. Due to the difference in masses between the sectors after electroweak symmetry breaking, simple kinematic effects give $X$ a larger branching ratio into the Standard Model. This occurs over a range of $X$ masses within a few decades of the weak scale. If $X$ decays out-of-equilibrium below the decoupling temperature of the two sectors, this injects different amounts of energy into the two sectors, effectively suppressing the temperature of the twin sector relative to the Standard Model. This relative cooling suppresses the contribution of the light degrees of freedom of the mirror Twin Higgs to below cosmological bounds. Insofar as the asymmetry is driven entirely

by kinematic effects arising from $v \ll f$, the resulting temperature inequality between the two sectors is proportional to powers of $v/f$.

The requisite suppression of the twin sector temperature relative to the Standard Model temperature necessitates that the $X$ dominate the cosmology before it decays. Our main discussion will follow the simplest case of an $X$ which dominates absolutely before it decays, comprising all of the energy density of the universe and effectively acting as a 'reheaton'. Afterwards, we will discuss the possibility of a 'thermal history' for $X$ – a scenario where $X$ is in thermal equilibrium with the two sectors, then chemically decouples at some high temperature and grows to dominate the cosmology before it decays. This scheme will result in additional stringent constraints on the viable parameter space.

### 4.4.1 Asymmetric Reheating

A $\mathbb{Z}_2$-even scalar $X$ which is a total singlet under the SM and twin gauge groups admits the renormalisable interactions

$$V \supset \lambda_x X(X+x)\left(|H_A|^2 + |H_B|^2\right) + \frac{1}{2}m_X^2 X^2, \tag{4.26}$$

where $m_X$ is the mass of $X$ (neglecting corrections from mixing that will be shown below to be tiny), $\lambda_x$ is a dimensionless coupling and $x$ is a dimensionful parameter, which one may imagine identifying as a vacuum expectation value (vev) of $X$ in an UV theory. Note that these interactions preserve the accidental $SU(4)$ symmetry of the Twin Higgs. The $X$ field may additionally possess self-interactions, which we omit here as they do not play a significant role in what follows.

The interactions in (4.26) allow $X$ to decay into light states in the Standard Model and twin sectors. If $X$ reheats the universe through out-of-equilibrium decays, the reheating

temperatures of the two sectors will be determined by its partial decay widths, assuming that the decay products do not equilibrate. In the instantaneous decay approximation, $X$ decays when the Hubble parameter falls to its decay rate $\Gamma_X \sim H$. As we will show in Section 4.4.2, in order to evade cosmological constraints we need the $X$ to decay mostly into the SM, so we may estimate $\Gamma_X \sim \Gamma(X \to \mathrm{SM})$. Then the energy that was contained in the $X$ is transferred into radiation energy density, with the resulting temperature of the radiation given by (see (42))

$$T \sim 1.2\sqrt{\frac{\Gamma_X M_{\mathrm{pl}}}{\sqrt{g_\star}}} \tag{4.27}$$

where $g_\star$ is the effective number of relativistic degrees of freedom, as defined in Section 4.3, of the particles that are being reheated. Our numerical calculation of the reheating temperature, which will be presented in Section 4.4.2, indicates that the approximation $T \sim 0.1\sqrt{\Gamma_X M_{\mathrm{pl}}}$ reliably reproduces the reheating temperature over the range of interest.

As shown in Section 4.3.2, the two sectors thermally decouple when the temperature falls below $T_{\mathrm{decoup}} \sim 1$ GeV, so reheating must take place to below this temperature. At even lower temperatures, big bang nucleosynthesis (BBN) places strong constraints on energy injected into the SM at temperatures below $\mathcal{O}(1-10)$ MeV (43). Requiring that the SM reheating temperature is above $\sim 10$ MeV, these constraints on the SM reheating temperature become constraints on the decay rate of the $X$ into the SM, which in the above approximation becomes

$$5 \times 10^{-21}\ \mathrm{GeV} \lesssim \Gamma_X \lesssim 3 \times 10^{-16}\ \mathrm{GeV}. \tag{4.28}$$

This then constrains the couplings $\lambda_x$ and $x$ of the $X$ to the Higgs sector. Importantly, it means that $X$ must couple very weakly, in order to be long-lived enough to reheat to

a low temperature, as will be shown below.

The asymmetry in partial widths arises from different effects depending upon the mass of $X$. For masses below the SM Higgs threshold, it is predominantly differences in mass mixing with the two Higgs doublets that produces the asymmetry, where the size of the mixing angles determines the effective coupling of $X$ to the SM and twin particles and therefore its branching fractions. For masses below the twin scale, the relative size of the mixing scales inversely with the vevs in each sector. Thus the hierarchy $v \ll f$ already present in the Higgs sector can automatically gives rise to a hierarchy in partial widths. Note that additional threshold effects can enhance the asymmetry further, in particular when $X$ has mass above threshold for a significant decay channel in the SM, but below the corresponding mass threshold in the twin sector. Decays into on-shell Higgses complicate this picture further. In what follows, we first give an analytic calculation of the mass mixing effect, then present a more precise calculation of the decay widths into each sector.

To lowest order, $X$ decays via its interactions with the SM and twin Higgs, and only to other fermions and gauge bosons through its mass mixing with the Higgs scalars. Expanding the $X$ potential after the $SU(4)$ is spontaneously broken, the mixing term between $X$ and $h_A$ in the scalar mass matrix is $\sqrt{2}\lambda_x x v_A$, while that between $X$ and $h_B$ is $\sqrt{2}\lambda_x x v_B$. The $h_A$ and $h_B$ components of the $X$ mass eigenstate, which we denote respectively as $\delta_{XA}$ and $\delta_{XB}$, can then be determined. The expressions for the mixing angles are in general complicated, but they simplify in limits $m_X < f$ and $m_X \gg f$:

$$(\delta_{XA}, \delta_{XB}) \approx \begin{cases} \frac{4\lambda_x x v_A}{m_X^2 - m_h^2} \left(\frac{1}{\sqrt{2}}, \frac{v_A}{f}\right) & m_X < f \\ \frac{\lambda_x x f}{m_X^2} \left(\frac{\sqrt{2} v_A}{f}, 1\right) & m_X \gg f \end{cases} \tag{4.29}$$

to lowest order in $(v/f)^2$ and $\kappa/\lambda$. The partial width for the decay of $X$ into SM states

(excluding the Higgs) is

$$\Gamma(X \to \text{SM}) \approx |\delta_{XA}|^2 \, \Gamma_h(m_h = m_X), \tag{4.30}$$

where $\Gamma_h(m_h = m_X)$ denotes the decay width of a SM Higgs if it were to have mass $m_X$. Note that the Higgs partial width must be computed using the vev $v_A \approx v/\sqrt{2}$ to determine the masses and couplings of the SM particles. The partial width of the $X$ into twin states is computed the same way using $\delta_{XB}$ and the vev $v_B \approx f/\sqrt{2}$.

From the mixing angles (4.29), it is already apparent over what mass range asymmetric reheating from $X$ decays will work. These give

$$\frac{\Gamma(X \to \text{SM})}{\Gamma(X \to \text{Twin})} \sim \begin{cases} f^2/v_A^2 \gg 1 & m_X < f \\ v_A^2/f^2 \ll 1 & m_X \gg f. \end{cases} \tag{4.31}$$

Thus when the mass of $X$ is less than the twin scale, the Standard Model will be reheated to a higher temperature than the twin sector, but in the large mass limit this mechanism works in the opposite direction and would appear to lead to preferential reheating of the twin sector.

More precise statements about the relative branching ratios and resulting temperatures require additional care. In addition to decaying through mass mixing, $X$ can decay into the Higgs mass eigenstates themselves if above threshold. As the energy is ultimately transferred to the SM and twin sectors, we then need to consider how these states decay and account for the further mixing of the Higgs mass eigenstates into Higgs gauge eigenstates.

For $m_X > 2m_h$, decay can occur into the lighter (SM-like) Higgs mass eigenstate $h$

with partial width

$$\Gamma(X \to hh) \approx \frac{\lambda_x^2 x^2}{16\pi m_X}\sqrt{1-\left(\frac{2m_h}{m_X}\right)^2}. \tag{4.32}$$

Similarly, for $m_X > 2m_H$, decays can proceed into $HH$ with a similar partial width, but with the $h$ mass replaced with that of the $H$. Above the intermediate threshold $m_X > m_h + m_H$, there is also the mixed decay

$$\Gamma(X \to hH) \approx \frac{\lambda_x^2}{2\pi m_X}\sqrt{1-\left(\frac{m_H+m_h}{m_X}\right)^2}(f\delta_{AX} + 2v_A\delta_{BX})^2. \tag{4.33}$$

Here, $\delta_{AX} \approx -\delta_{hA}\delta_{XA} - \delta_{hB}\delta_{XB}$ is the component of the $h_A$ gauge eigenstate in the $X$ mass eigenstate and $\delta_{BX} \approx \delta_{hB}\delta_{XA} - \delta_{hA}\delta_{XB}$ is the corresponding component of the $h_B$ gauge eigenstate, where $\delta_{hA}$ and $\delta_{hB}$ are, respectively, the components of the SM Higgs in the $h_A$ and $h_B$ gauge eigenstates to zeroth order in $\lambda_x$. Combining all ingredients, this decay width is of order $\lambda_x^4 x^2$. Since it is only the total decay width that is constrained to be small by the demand that the SM reheating temperature lie in the required window, this fixes only a product of $\lambda_x$ and $x$. If $x \sim v$, then the mixed decay to $hH$ is effectively second order in the small coupling $\lambda_x^2$ and can be neglected relative to the other partial widths. Conversely if $x \ll v$, then $\lambda_x$ is much larger and this decay cannot be neglected. In what follows we will work in the region of parameter space where mixed decays to $hH$ are negligible.

The rate of heat flow into each sector may be well approximated by adding the decay rates of $X$ into each channel and weighting these by the fraction of energy transferred into the particular sector. Of course, when $X$ decays into Higgs particles, these in turn decay out of equilibrium into both the Standard Model and twin sectors. As the Higgs decays are almost instantaneous, the fraction of energy transferred into each sector is simply that carried by the Higgs decay products multiplied by their branching fractions

for each sector. The total rate at which $X$ particles are transferred into the SM plasma is

$$W(X \to SM) \approx \Gamma(X \to SM) + \Gamma(X \to hh)Br(h \to SM) + \Gamma(X \to HH)(Br(H \to SM) + Br(H \to hh)Br(h \to SM)). \quad (4.34)$$

The corresponding rate for energy deposition into the twin sectors is simply given by the replacement of $SM \mapsto$ Twin. The first term is the rate at which $X$ decays directly into the SM through mass mixing with the Higgs. The second is the fraction of $X$ energy that is transferred into lighter Higgs states that subsequently decay into the SM. The third is the analogous term for decays into the heavy Higgs, where cascade decays of the $H$ into the $h$ and subsequently other SM particles must be included. Note that decays of the heavy Higgs into the light Higgs make up a majority of decay width, because of the large quartic coupling required for the twin Higgs potential.

Below the $hh$ threshold, it is possible for $X$ to decay via one on-shell and one off-shell Higgs boson. The partial width for off-shell Higgs production was calculated for $X \to hh^* \to hb\bar{b}$ and found to be negligible compared to two-body decays through mass mixing and so we omit three-body decay widths in what follows.

Ultimately, the complete partial widths for the decay of $X$ into the Standard Model and twin sectors includes the sum of decays into Higgs bosons $h$ and $H$ and direct decays into the fermions and gauge bosons of the two sectors. We compute the latter to an intended level of accuracy of $\sim 10\%$ (including, e.g., NLO QCD corrections to decays into light-flavor quarks), mostly following (44). The resulting partial widths into the Standard Model and twin sectors are shown as a function of $m_X$ in Figure 4.5 with the ratio of branching fractions displayed in Figure 4.6.

Over much of the space below the Higgs mass, the branching ratio exhibits the ex-

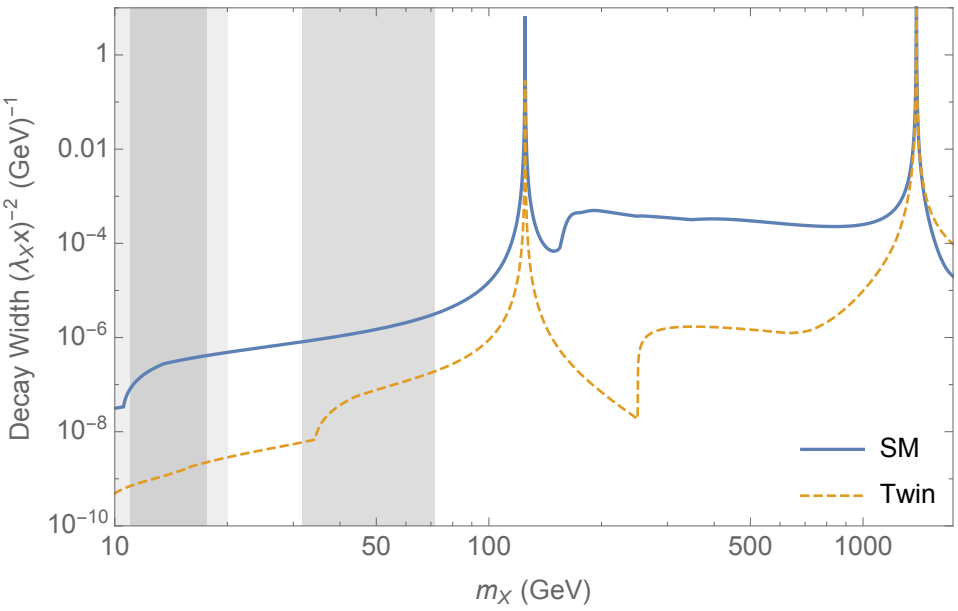



Figure 4.5: The partial widths of the $X$ into the SM (solid blue line) and twin sector (dashed orange) for $f/v = 3$ in units of $(\lambda_x x)^2$. The light gray bands indicate regions of QCD-related uncertainty in the SM calculation, while the darker gray bands indicate the corresponding regions of uncertainty for the twin calculation.

pected $(f/v)^2$ scaling from the mass mixing. Below $\sim 40$ GeV, suppression of the twin partial width arises because the twin bottom quark pair production threshold is crossed. As $m_X$ nears $m_h$, the SM branching fraction grows by $\sim 4$ orders of magnitude as the $WW^*$, $ZZ^*$, and then $WW$ and $ZZ$ decays go above threshold. Since the analogous thresholds are at much higher energies in the twin sector, the enhancement is not paralleled by decays into the twin sector until $m_X$ is close to the twin scale. There is therefore a large range of masses $m_h \lesssim m_X \lesssim m_H$ over which the SM branching fraction dominates by several orders of magnitude.

Above the $X \to hh$ threshold, the ratio of decay widths is roughly constant in mass up to the $HH$ threshold. The twin sector decay rate is dominated by decays of on-shell light Higgs into twin states, $\Gamma(X \to \text{Twin}) \approx \Gamma(X \to hh)\text{Br}(h \to \text{Twin}) \propto 1/m_X$ as in (4.32). If the SM were also predominantly reheated through this channel, then the

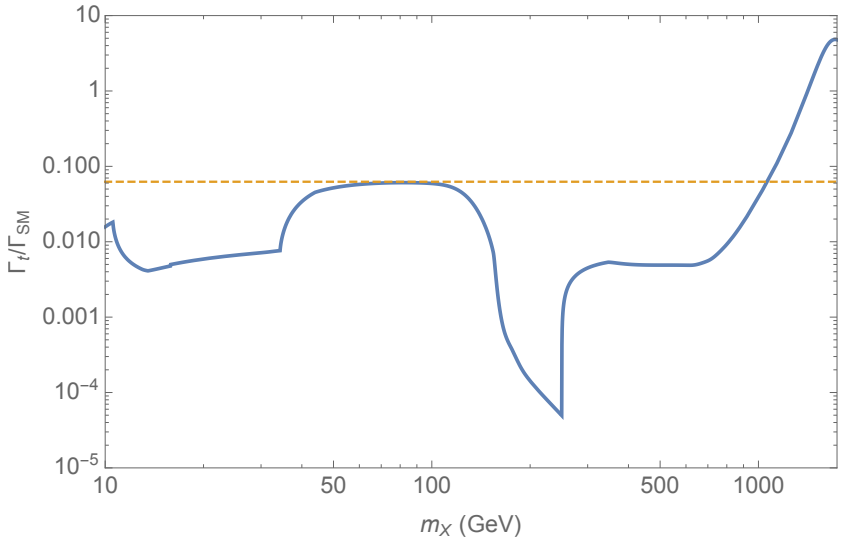



Figure 4.6: The ratio of branching fractions of the $X$ into the SM and twin sectors at $f/v = 3$. The dashed line gives the expected $(v/f)^2$ scaling from the mass mixing; deviations are due to various mass threshold effects.

ratio of branching fractions would again be approximately $\delta_{hA}^2/\delta_{hB}^2 \approx (f/v)^2$. However, the SM decay width also receives a larger contribution from decays through mass mixing between the $X$ and the Higgs gauge eigenstates.

For masses $m_X > 2m_h$, decays through mass mixing are dominated by the SM $WW$ and $ZZ$ channels. In this mass region, the decay rate of a Higgs into longitudinally polarized vector bosons scales as $\Gamma(h \to WW, ZZ) \sim m_X^3$, but the mixing angle scales as $\delta_{AX}^2 \sim 1/m_X^4$ (as in (4.29)), resulting in the same $\sim 1/m_X$ scaling and thus a roughly constant ratio in this range of masses. Near $m_X \sim 1$ TeV, decays into twin vector bosons through mass mixing begin to dominate, and there is no favourable asymmetry in the branching fractions, as discussed in this section. Even at higher masses, the effects of heavy Higgs decays into light Higgs do not compensate sufficiently, as this partial width scales with $m_X$ in the same way as the partial width for longitudinally polarised weak bosons.

The constraint on the decay width from the required reheating temperature (4.28)

translates into a constraint on the size of the coupling $\lambda_x x$. For $m_X \gtrsim m_h$, this gives $10^{-8.5}\,\text{GeV} \lesssim \lambda_x x \lesssim 10^{-6}\,\text{GeV}$, while for lower masses, this range increases to $10^{-7}\,\text{GeV} \lesssim \lambda_x x \lesssim 10^{-5.5}\,\text{GeV}$ at $m_X \sim 20\,\text{GeV}$.

The gray bands in Figure 4.5 highlight regions where our analytic estimates of the partial widths encounter enhanced uncertainties arising from the bottom and charm thresholds in both sectors. Over most of these ranges, we estimate the size of these uncertainties to be either $\sim 10\%$ or confined to very small subregions. The thicknesses of these bands have been chosen conservatively, and ultimately the branching ratios should be accurate to within a factor of $\pm\Lambda_{QCD}$ of the bottom and charm mass thresholds. In particular, the prescription of (45) has been followed for approximating the bottom partial width close to the open flavour threshold. Resonant decay into gluons from bottomonia mixing has been neglected, although these resonant mass ranges are expected to be only $\sim$ MeV wide at the CP-even, spin-0 bottomonia masses $m_X = m_{\chi_{b_i}}$ (see (45) and (46)). It should be noted, however, that at temperatures above that of the QCD phase transition, the quark decay products behave differently compared to that expected in a low temperature environment. In particular, for hot enough temperatures, the $b$ or $c$ quarks may not hadronise and the partonic partial widths may more reliable. The applicability of the treatment of the flavour thresholds used here may therefore not be valid if the decay occurs in the hot early universe. However, it is only very close to the threshold itself (within several GeV) that this uncertainty becomes significant. Finally, quark masses have been neglected in the gluon partial width. For $m_X$ close to the flavour thresholds, this approximation breaks down, but the gluon branching fraction is only $\sim 10\%$ and so the error does not contribute to the uncertainty of the total width by more than this order (it is this uncertainty that is responsible for most of the extension of the length of the gray bands about the flavour threshold).

Close to the charm threshold, the analogous uncertainties are even more poorly under-

stood. Below the charm threshold, hadronic decays of a light scalar are highly uncertain (see (47) for discussion). We avoid these regions altogether by restricting our considerations to $m_X$ roughly above the twin charm threshold. Note that below the SM charm threshold, the smaller decay rate of a Higgs-like scalar necessitates larger couplings $\lambda_X x$ for $X$ to have a lifetime within the required reheating window. The larger couplings then imply potentially stronger constraints from invisible mesonic decays. See (46–48) for further discussion and recent analysis of the pertinent experimental constraints.

Taken together, the results in Figures 4.5 and 4.6 bear out the expectation that a scalar $X$ with symmetric couplings to the Standard Model and twin sectors may nonetheless inherit a large asymmetry in partial widths from the hierarchy between the scales $v$ and $f$. Across a wide range of masses $m_X$, the asymmetry is proportional to (or greater than) $v^2/f^2$, tying the reheating of the two sectors to the hierarchy of scales.

Before proceeding to our computation of cosmological observables, we comment on an alternative variation on the reheating mechanism presented here that involves having $X$ odd under the twin parity. This permits two renormalisable interactions with the Higgses to give a Higgs potential of the form:

$$\begin{aligned}\mathcal{V} \supset\, & m_0^2 \left(|H_A|^2 + |H_B|^2\right) + \lambda_0 \left(|H_A|^4 + |H_B|^4\right) \\ & + \epsilon X^2 \left(|H_A|^2 + |H_B|^2\right) + \tilde{\epsilon} X \left(|H_A|^2 - |H_B|^2\right). \end{aligned} \tag{4.35}$$

If $X$ then acquires a vev at some scale, it may be possible to arrange for the resulting spontaneous breaking of the $\mathbb{Z}_2$ to give that required in the Higgs potential. However, we find that, in order for $X$ to be long-lived and reheat the universe, its couplings to the Higgs must be highly suppressed and therefore that the resulting vev of $X$ required to explain the soft $\mathbb{Z}_2$-breaking in the Higgs potential must be many orders of magnitude above the twin scale. If this is to be identified with the characteristic mass scale of $X$,

then a UV-completion of the twin Higgs is required for anything further to be said of the prospects of this possibility. However, if such a UV completion has similar structure to the couplings in (4.35), then asymmetric reheating may require a cancellation between the odd and even couplings of $X$ to the Higgs potential in order to suppress its twin-sector branching fraction (because the odd coupling appears with opposite signs in the coupling between $X$ and the $h_A$ and $h_B$ states). We do not consider this possibility further.

### 4.4.2 Imprints on the CMB

For appropriate values of $m_X$, the out-of-equilibrium decay of $X$ reheats the two sectors to different temperatures and effectively dilutes the energy density in the twin sector. We obtain an analytic estimate of the effects of the $X$ decay on the number of light degrees of freedom observed from the CMB by approximating both the decay of $X$ and the decoupling of species as instantaneous in Section 4.4.2. We then demonstrate that this estimate is reliable over most of the parameter space of interest with a numerical calculation in Section 4.4.2. In Section 4.4.2 we consider neutrino masses and their joint constraints with $N_{\rm eff}$.

#### Analytic estimate of $N_{\rm eff}$

If $X$ dominates the energy density of the universe and then decays, depositing energy $\rho_{\rm SM}$ and $\rho_{\rm t}$ into the SM and twin sectors respectively, then the temperature ratio is determined by

$$\frac{\rho_{\rm t}}{\rho_{\rm SM}} = \frac{g_\star^{\rm t}(T_{\rm reheat}^{\rm t})}{g_\star^{\rm SM}(T_{\rm reheat}^{\rm SM})} \left(\frac{T_{\rm reheat}^{\rm t}}{T_{\rm reheat}^{\rm SM}}\right)^4 \approx \frac{\Gamma(X \to \text{Twin})}{\Gamma(X \to \text{SM})}, \tag{4.36}$$

where $T_{\rm reheat}^{\rm SM}$ and $T_{\rm reheat}^{\rm t}$ are the reheating temperatures for each sector, while $g_\star^{\rm SM}$ and $g_\star^{\rm t}$ are the SM and twin effective number of relativistic degrees of freedom, respectively. We have assumed that the two sectors are cool enough that they have already decoupled.

We point out that not only does the number of effective degrees of freedom in each sector need to be evaluated at the temperature of that sector, but that $g_\star^{\rm t}$ and $g_\star^{\rm SM}$ differ as functions of temperature due to the differences in the spectra of the sectors, as seen in Figure 4.1. As is well-known (42), reheating is a protracted process that occurs over a time-scale given by the lifetime of the reheaton. During this time, the temperature of the plasma cools slowly because, while the energy is being replenished by the decay of the reheaton, it is simultaneously diluted and redshifted with the expansion of the universe. It is assumed in (4.36) that any primordial energy density in either sector is subdominant.

The temperatures of both sectors then redshift in the same way, so the only additional differences between their temperatures arise from changes to the effective number of degrees of freedom in each sector. By conservation of comoving entropy within each sector, each evolves as $T_{eq}^i/T_{\rm reheat}^i = \left(g_\star^i(T_{\rm reheat}^i)/g_\star^i(T_{eq}^i)\right)^{1/3} a(T_{\rm reheat})/a(T_{eq})$ where $T_{eq}^i$ is the temperature of the sector at matter-radiation equality, which the CMB probes as explained in Section 4.3.3, and $a(T)$ is the scale factor as a function of temperature. In the mirror Twin Higgs model, the two sectors have the same number of light degrees of freedom at recombination (three neutrinos and a photon, assuming that the neutrinos are still relativistic), so

$$\left(\frac{T_{eq}^{\rm t}}{T_{eq}^{\rm SM}}\right)^4 = \left(\frac{T_{\rm reheat}^{\rm t}}{T_{\rm reheat}^{\rm SM}}\right)^4 \left(\frac{g_\star^{\rm t}(T_{\rm reheat}^{\rm t})}{g_\star^{\rm SM}(T_{\rm reheat})}\right)^{4/3} = \frac{\Gamma(X\to{\rm Twin})}{\Gamma(X\to{\rm SM})}\left(\frac{g_\star^{\rm t}(T_{\rm reheat}^{\rm t})}{g_\star^{\rm SM}(T_{\rm reheat})}\right)^{1/3}. \quad (4.37)$$

As our range of reheat temperatures encompasses the QCD phase transitions of both sectors, the factors of $g_\star$ can be important.

Given the temperatures of the two sectors after $X$ decays, we can obtain a simple estimate of the contribution to $N_{\rm eff}$ that neglects the impact of masses of the twin neutrinos

discussed in Section (4.3.3),

$$(\Delta N_{\rm eff})_{m_\nu=0} = \frac{4}{7}\left(\frac{11}{4}\right)^{4/3} g_\star^{\rm SM}(T_{eq}^{\rm SM})\frac{\rho_{\rm t}(T_{eq}^{\rm t})}{\rho_{\rm SM}(T_{eq}^{\rm SM})} \tag{4.38}$$

$$\approx 7.4 \times \frac{{\rm Br}(X \to {\rm Twin})}{{\rm Br}(X \to {\rm SM})}\left(\frac{g_\star^{\rm t}(T_{\rm reheat}^{\rm t})}{g_\star^{\rm SM}(T_{\rm reheat}^{\rm SM})}\right)^{1/3}. \tag{4.39}$$

In this limit the most recent Planck data give a $2\sigma$ bound of $\Delta N_{\rm eff} \lesssim 0.40$ assuming pure $\Lambda$CDM+$N_{\rm eff}$ (2). This translates into the requirement $\frac{\rho_{\rm t}(T_{eq}^{\rm t})}{\rho_{\rm SM}(T_{eq}^{\rm SM})} \approx \frac{\Gamma(X\to{\rm Twin})}{\Gamma(X\to{\rm SM})} \lesssim 0.05$, ignoring possible differences in $g_\star$.

Of course, as discussed in Section 4.3, the twin neutrino masses are relevant at the temperature of matter-radiation equality, so we can obtain a more meaningful estimate of $\Delta N_{\rm eff}$ using the results of Section 4.3.3 evaluated at the twin temperature determined above:

$$\Delta N_{\rm eff} = \left(\frac{11}{4}\right)^{4/3}\frac{120}{7\pi^2\left(T_{eq}^{\rm SM}\right)^4}\left(\rho_\gamma^{\rm t}\left(T_{eq}^{\rm t}\right) + \sum_\alpha 3w_{\nu_\alpha}^t\left(T_{eq}^{\rm t}\right)\rho_{\nu_\alpha}^{\rm t}\left(T_{eq}^{\rm t}\right)\right) \tag{4.40}$$

$$T_{eq}^{\rm t} = T_{eq}^{\rm SM}\left(\frac{\Gamma(X\to{\rm Twin})}{\Gamma(X\to{\rm SM})}\right)^{1/4}\left(\frac{g_\star^{\rm t}(T_{\rm reheat}^{\rm t})}{g_\star^{\rm SM}(T_{\rm reheat}^{\rm SM})}\right)^{1/12} \tag{4.41}$$

with $T_{eq}^{\rm SM} \approx 0.77$ eV (2) the photon temperature. While the right-hand side of this equality has implicit dependence on $T_{eq}^{\rm t}$ through $g_\star^{\rm t}$, this is only important if the reheating occurs between the SM and twin QCDPTs and the neglecting of the factors of $g_\star$ is otherwise reliable. With the further inclusion of Standard Model neutrino masses or an extra sterile neutrino, the bound described above weakens to $\Delta N_{\rm eff} \lesssim 0.7$. As discussed in Section 4.3.3, we are not aware of any analyses specific to our model involving both pure dark radiation and three sterile neutrinos with masses of order the photon decoupling temperature of the CMB and possibly cooler temperatures. In the absence of such an analysis, we use the inequality $\Delta N_{\rm eff} \lesssim 0.7$ to indicate where the present CMB

measurements are likely to constrain the light degrees of freedom of this model, leaving a more detailed analysis of the CMB constraints as future work. In this case, the bound on the decay width ratio is $\frac{\Gamma(X\to\text{Twin})}{\Gamma(X\to\text{SM})} \lesssim 0.09$. The next generation of CMB experiments are projected to strengthen this constraint to $\Delta N_{\text{eff}} \lesssim 0.02$ at the $1\sigma$ level (49).

**Numerical Calculation of $N_{\text{eff}}$**

A more precise study of the effect of $X$ decay on the number of effective neutrino species at recombination may be performed by numerically solving a system of differential equations for the entropy in $X$ and the two sectors as a function of time. Following the analysis of Chapter 5.3 of (42) we have

$$H = \frac{1}{a}\frac{da}{dt} = \sqrt{\frac{1}{3M_{Pl}^2}(\rho_X + \rho_{SM} + \rho_t)} \tag{4.42}$$

$$\frac{d\rho_X}{dt} + 3H\rho_X = -\Gamma_X \rho_X \tag{4.43}$$

$$\rho_i = \frac{3}{4}\left(\frac{45}{2\pi^2 g_\star^i}\right)^{1/3} S_i^{4/3} a^{-4} \tag{4.44}$$

$$S_i^{1/3}\frac{dS_i}{dt} = \left(\frac{2\pi^2 g_\star^i}{45}\right)^{1/3} a^4\Big(\rho_X \Gamma_{X\to i} + \frac{dq_{j\to i}}{dt}\Big), \tag{4.45}$$

where $S_i$ are comoving entropy densities and it has been assumed that $X$ is cold by the time it decays so that $\rho_X = m_X n_X$ with number density $n_X$ (this is reliable as we only consider $m_X > 10$ GeV, which is above the decoupling temperature of $\sim 1$ GeV). The rate of heat flow from sector $j$ to $i$ per proper volume, $\frac{dq_{j\to j}}{dt}$, is defined in (4.7). To account for the temperature-dependence of the effective number of relativistic degrees of freedom in each sector, these equations are solved iteratively in the profiles of $g_\star^i(T^i)$.

The equations are solved in three stages: before, during and after the decoupling of the SM and twin sectors. The ratio $f/v$ is fixed to 4 for this analysis. Initial conditions were chosen with $\rho = 10^{-12}\rho_X$, for combined SM and twin energy densities $\rho$. However,

it is only the requirement that the initial energy density of $X$ dominates over that of the SM and twin sectors that is important for simulating the cosmology over the times of interest here, as the entirety of the latter is then generated by the subsequent decay. The results close to the decoupling and reheating epochs are otherwise insensitive to the initial conditions and ultimately match onto the standard outcome (42) expected by equating the Hubble rate with the decay rate of X. The sectors are assumed to be in thermal equilibrium and sharing entropy until a temperature of 10 GeV, below which they are evolved separately with the heat flows $\frac{dq_{i\to j}}{dt}$ switched on. Elastic scatterings were neglected from the heat flow rate to accelerate the computation. It was verified for the results found below that their contribution to the heat flow was always $\lesssim 10\%$ while the heat flow was itself not dominated by the Hubble rate. Heat flow was switched off again once the twin temperature reaches 0.1 GeV, by which time thermal decoupling is long-since complete, and the sectors are subsequently evolved separately. Again, although the strengthening of the colour force and the QCDPT make the perturbative tree-level computation of the scattering rates unreliable at temperatures below $\sim 1$ GeV, as found in Section 4.3.2 and also in the results below, the sectors decouple above these temperatures. Notably, the impact of $X$ on the expansion rate causes decoupling to occur at slightly hotter temperatures than expected from the analysis of Section 4.3.2 for the decoupling in the standard cosmology.

The ratio of energy densities in each sector determines $N_{\text{eff}}$, from (4.40). A plot of this ratio over time is shown in Figure 4.7, with the expectation under the approximations of the previous section shown as well. This approximation is reliable as long as the lifetime of $X$ is much longer than the temperature at which decoupling concludes, here $\sim 1$ GeV. The larger asymptotic value of the ratio of the blue line arises because the lifetime lies close to the decoupling period, so that a significant fraction of the energy is transferred while the sectors are thermalised or partially thermalised and does not contribute toward

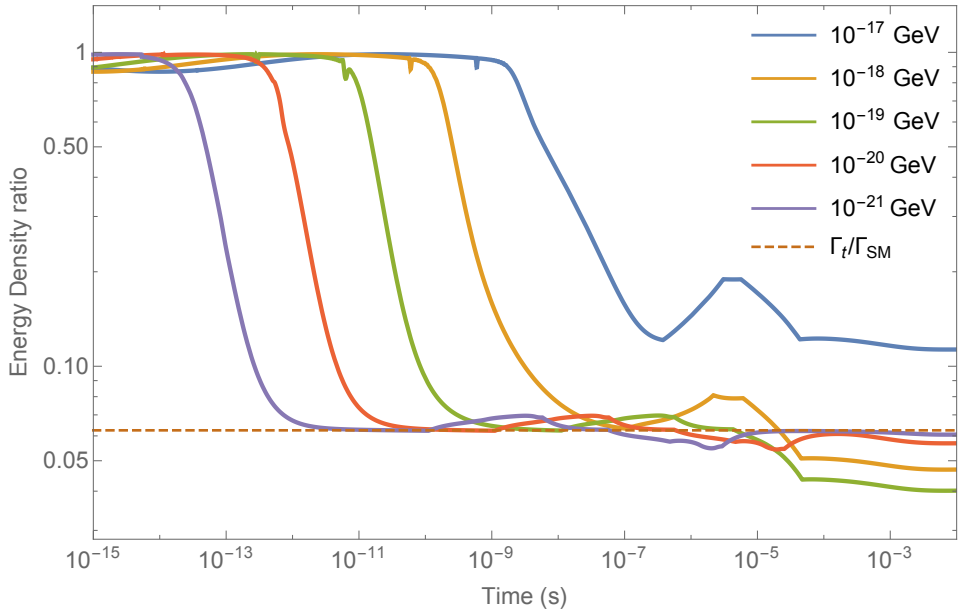



Figure 4.7: Ratio of twin to SM energy densities throughout decoupling and reheating, for different decay rates $\Gamma_X$. The dashed line corresponds to the prediction of from the ratio of decay widths, here selected to be 1/16.

asymmetric reheating. Equivalently, as will be discussed below, insufficient time elapses between decoupling and reheating for the twin energy density to dilute and be repopulated by the decays to the level predicted by (4.36). The subsequent bump represents the period between the reheating of the twin sector by its QCD phase transition followed by that of the SM. The green and orange lines correspond to reheating temperatures that lie between SM and twin QCD phase transitions. In these cases, the reheating of the SM from the subsequent SM QCD phase transition raises its energy density relative to the twin sector above that expected from the ratio of branching fractions. As this occurs after the lifetime of the reheaton, the estimate of the reheating temperatures presented in (4.37) is still good as subsequent changes in the ratio due to the evolution of $g_\star$ are accounted for in our analysis of the reheating scenarios.

The steep drop in the energy density ratio corresponds to the brief period during which the energy density of the twin sector present at decoupling dilutes and redshifts,

which continues until it reaches a comparable size to the energy density that is being replenished by reheating. If the twin-sector branching fraction is highly suppressed, as can occur in the "valley" region in Figure 4.6 with $m_h \lesssim m_X \lesssim 2m_h$, then a longer time is required for this to happen, especially close to the decay epoch where the diminishing of the $X$ population also contributes to a reduced reheating rate. These effects can prolong the time required for the energy density ratio to converge to the asymptotic prediction of (4.36).

Contour plots of $\Delta N_{\text{eff}}$ as a function of $m_X$ and $f/v$ appear in Figure 4.8, along with current and predicted bounds using the analytic results of Section 4.4.2. The minimum neutrino mass configuration with Dirac masses has also been assumed, although the results are relatively insensitive to this provided that the twin neutrino masses are not well above the eV scale. A SM reheating temperature of 0.7 GeV has been assumed. At this temperature, we have verified using the numerical calculation of Section 4.4.2 that the twin sector reheating temperature is always roughly above the twin neutrino decoupling temperature over the parameter space of the figure, ensuring that the neutrinos thermalise once produced in the decays and hence that the predictions of Section 4.4.2 are valid. A treatment of the case in which the twin neutrinos are produced below their decoupling temperature is beyond the scope of this analysis, but would involve the computation of the phase space spectrum of the neutrino decay products of the $X$.

Also, as discussed in Section 4.3.2, a large temperature difference may partially relax back if reheating occurs close to sector decoupling. However, a reliable calculation of the heat flow at the temperatures of interest here must incorporate non-perturbative effects. We do not perform such a computation, but note that, at a slightly higher SM reheating temperature of 2 GeV where this computation is more reliable, $\Delta N_{\text{eff}}$ in Figure 4.8 can be raised by up to an order of magnitude in the region with $f/v \lesssim 4$ and $150\,\text{GeV} \lesssim m_X \lesssim 200\,\text{GeV}$, notably where the twin sector partial width is suppressed relative to the

SM by several orders of magnitude. The resulting $\Delta N_{\rm eff}$ prediction is, nevertheless, still out of observable reach. At the lower SM reheating temperature assumed in Figure 4.8, it is expected that decoupling will be further advanced and the enhancement in $\Delta N_{\rm eff}$ would be weaker.

We emphasize that, if the lifetime of $X$ is sufficiently close to the time of decoupling, or equivalently, that the reheating temperature is sufficiently close to the decoupling temperature, then the residual twin energy density left-over may be comparable to or greater than that regenerated by reheating. Consequently, the suppression in $\Delta N_{\rm eff}$ would be less than that predicted in (4.37). In this respect, the projection of Figure 4.8 should be regarded as a lower bound on $\Delta N_{\rm eff}$. In the regions of high suppression, such as the "valley" region, the full asymmetry may not be generated before the complete decay of $X$ when the reheating temperature is of similar order as the decoupling temperature. In particular, for the reheating temperature chosen here of 0.7 GeV and branching fraction $\mathrm{Br}(X \to \mathrm{Twin}) \sim 10^{-5}$, the numerical calculation of the energy density ratio saturates at $\sim 4 \times 10^{-5}$. We do not include this effect in Figure 4.8 as its only impact is to mildly shift the unobservably small $\Delta N_{\rm eff} = 10^{-4}$ contour. Lower reheating temperatures would agree with the prediction of (4.36) were it not for the caveat that the twin neutrinos may be produced out of equilibrium. However, this minimum value at which $\Delta N_{\rm eff}$ is saturated can grow significantly with hotter reheating temperatures upon which it is highly dependent.

CMB-S4 observations will be able to probe a large portion of the most natural parameter space, save the region $m_h \lesssim m_X \lesssim 2m_h$ where decays into the Standard Model dominate well beyond the ratio $f^2/v^2$, as previously discussed. Significantly, precision Higgs coupling measurements at the LHC are unlikely to probe the mirror Twin Higgs model beyond $f \sim 4v$, so that the observation of additional dark radiation may be the *first* signature of a mirror Twin Higgs.

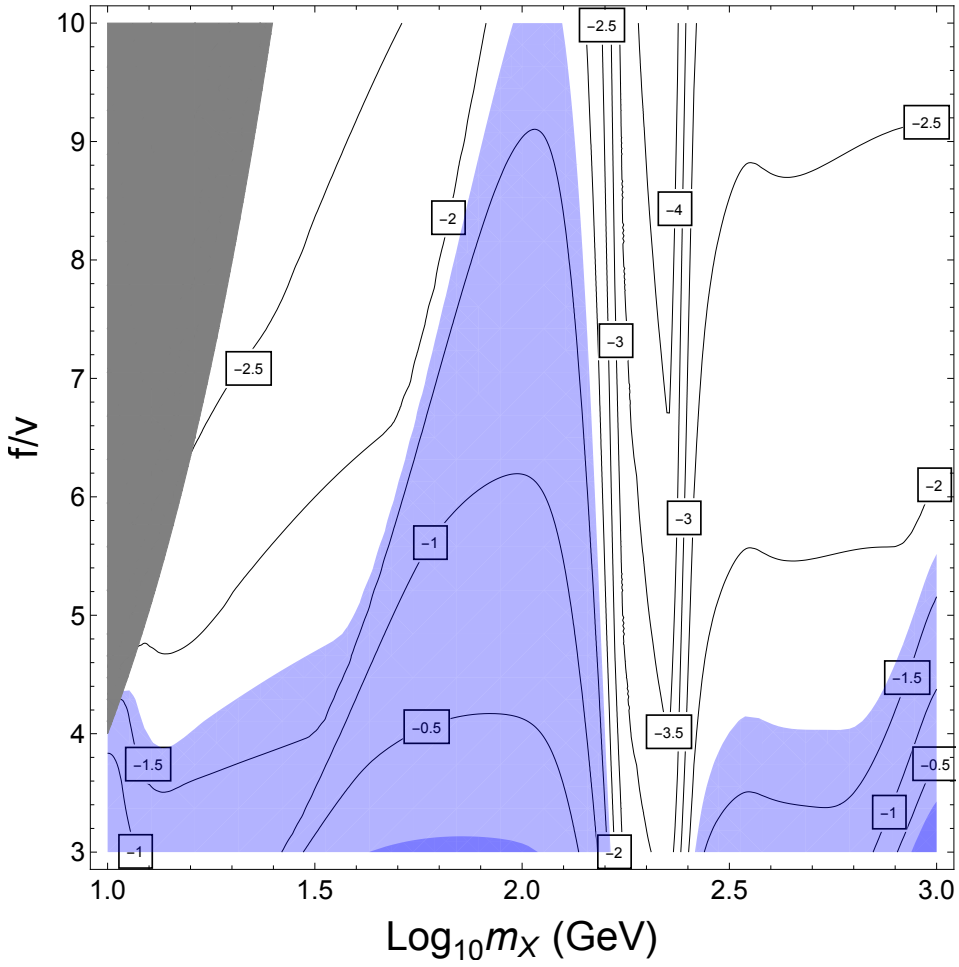



Figure 4.8: Contours of $\log_{10}\Delta N_{\rm eff}$ as a function of $m_X$ and $f/v$, for $T^{\rm SM}_{\rm reheat} = 0.7$ GeV. The dark blue region is in tension with Planck, while the light blue region will be tested by CMB-S4. Gray regions are where the $X$ mass is below the twin charm threshold and our calculation of the twin sector partial width is unreliable.

### Neutrino Masses

In addition to the bounds on $N_{\rm eff}$, we must also respect the bounds on neutrino masses. The analysis remains nearly the same as in Section 4.3.3, but now with the twin neutrinos at a lower temperature, as determined above. As mentioned above, for large enough $f/v$ and SM reheating temperature sufficiently close to the lower bound, the reheating temperature of the twin sector may be below the twin neutrino decoupling temperature and the resulting energy density would be more difficult to compute. For

simplicity, we choose $\lambda_x x$ large enough such that the twin reheating temperature is always above the twin neutrino decoupling temperature.

As before, we compute $m_\nu^{\rm eff}$ as

$$m_\nu^{\rm eff} = \frac{n_\nu^{\rm t}}{n_\nu^{\rm SM}} \sum_\alpha m_{\nu_\alpha}^{\rm t}. \tag{4.46}$$

In relating the scale factors at neutrino decoupling in each sector, we now have to use the above temperature ratio to find, analogously to Section 4.3.3, that

$$m_\nu^{\rm eff} = \left(\frac{\Gamma_{\rm t}}{\Gamma_{\rm SM}}\right)^{3/4} \left(\frac{g_\star^{\rm t}\left(T_{\rm reheat}^{\rm t}\right)}{g_\star^{\rm SM}\left(T_{\rm reheat}^{\rm SM}\right)}\right)^{1/4} \left(\frac{f}{v}\right)^n \sum_\alpha m_{\nu_\alpha}^{\rm SM}, \tag{4.47}$$

where, again, $n = 1$ for Dirac masses and $n = 2$ for Majorana masses. Interestingly, if the branching ratios scale as $\Gamma_{\rm t}/\Gamma_{\rm SM} = (v/f)^2$, then we have $m_\nu^{\rm eff} \propto (f/v)^{-3/2+n}$, so the contribution grows with $f/v$ for Majorana masses, but is suppressed for Dirac masses.

As before, we consider the minimal mass spectrum of $m_\nu = [0.0, 0.009, 0.06\ \text{eV}]$ and a degenerate spectrum of $m_\nu = [0.1\ \text{eV}, 0.1\ \text{eV}, 0.1\ \text{eV}]/3$. In Figure 4.9 we plot the predictions of the $X$ reheating for $\Delta N_{\rm eff}$ and $m_\nu^{\rm eff}$ for both spectra and both Dirac and Majorana masses using the approximations of Section 4.3.3, for $f/v$ from 3 to 10 and assuming the $\frac{\Gamma_{\rm t}}{\Gamma_{\rm SM}} \sim (v/f)^2$ scaling; there are regions in the space of $m_X$ where the suppression of $m_\nu^{\rm eff}$ would be much higher.

Dashed lines indicate the rough locations of present experimental limits from Planck 2015, and projected bounds from CMB-S4. As mentioned in Section 4.3.3, we are unaware of any study of bounds on both $m_\nu^{\rm eff}$ and $\Delta N_{\rm eff}$ treated jointly. In the absence of this, we show present and projected constraints on $N_{\rm eff}$ and $\sum m_\nu$ from (50) and (27), ignoring correlations, as described in Section 4.3.3.

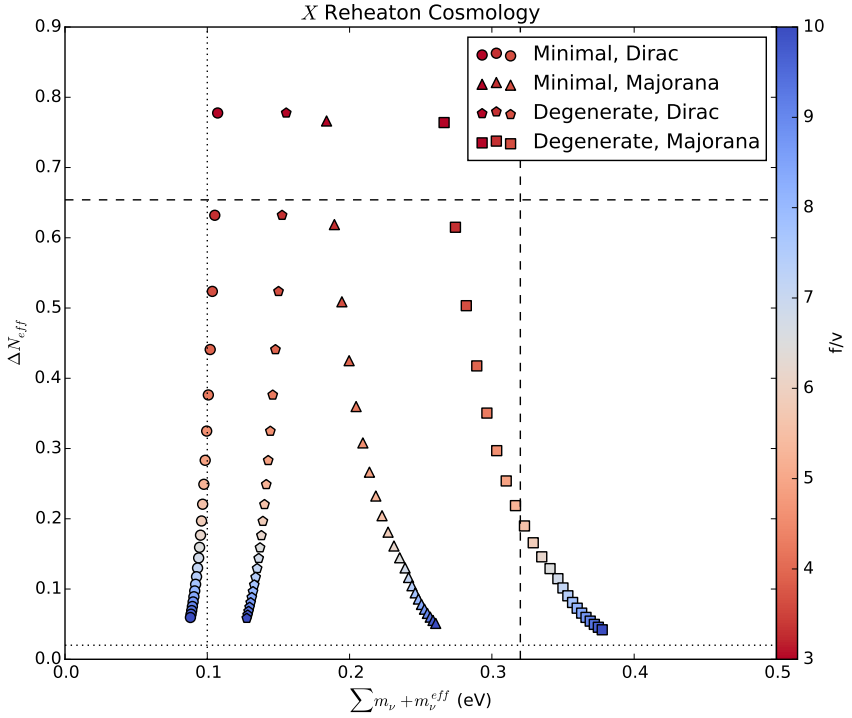



Figure 4.9: Predicted values of $\Delta N_{\text{eff}}$ and $\sum m_\nu + m_\nu^{\text{eff}}$ for minimal and degenerate neutrino mass spectra with both Dirac and Majorana masses for $f/v$ from 3 to 10. The Planck 2015 (2) bounds on $\sum m_\nu$ and $N_{\text{eff}}$, as discussed in Section 4.3.3, are represented by the dashed lines, and the projected CMB-S4 constraints are given by the dotted lines. It has been assumed that $\frac{\Gamma_t}{\Gamma_{\text{SM}}} \sim (v/f)^2$. Note however, that, from Figure 4.8, this scaling of the partial widths holds only for the mass range $50\,\text{GeV} \lesssim m_X \lesssim 120\,\text{GeV}$, outside of which the twin partial width is more suppressed and the model is only testable through $\Delta N_{\text{eff}}$ over a smaller range in $f/v$.

### 4.4.3 Thermal Production

In our discussion up to this point, we have been agnostic about the origin of the cosmic abundance of $X$ and have operated under the assumption that it absolutely dominates the cosmology before it decays. Here, we consider the possibility that $X$ was thermally produced through freeze-out and subsequently dominates the universe as a relic before decaying. This thermal history is viable, but places strong constraints on the mass and couplings of the $X$.

The energy density of relativistic species redshifts as $\rho_r \propto a^{-4} \propto T^4$, while the energy

density of non-relativistic, chemically decoupled matter scales as $\rho_m \propto a^{-3}$. The energy density contained in the $X$ can therefore only grow relative to the energy density in the thermal bath once it becomes non-relativistic. We found in Section 4.4.2 that by recombination, $\rho_{\rm t}/\rho_{\rm SM} \lesssim 0.09$ is needed to evade current bounds on $\Delta N_{\rm eff}$. Thus we need to have the energy density in the $X$ dominate over the SM and twin plasmas by more than this factor when it decays. If $X$ becomes non-relativistic instantaneously at the moment that its temperature reaches some fraction $c \sim \mathcal{O}(0.1)$ of its mass, then, as $T \propto 1/a$ and $\rho_X$ is $\sim 1/g_\star$ of the total energy density, the mass is required to satisfy $m_X \gtrsim 10/c \times g_\star\left(T = m_X\right) T^{\rm SM}_{X\rm reheat}$. Since the SM reheating temperature is strongly constrained to be above BBN, this effectively puts a lower limit on the mass of the $X$. Importantly, $X$ must freeze-out when relativistic or its energy density will be further Boltzmann suppressed. The lower limit on the mass of the $X$ becomes an upper limit on the $X$'s couplings - if it couples too strongly to the thermal bath, then it won't freeze out early enough to be hot.

In fact the situation is somewhat less favorable than the above analysis suggests, because it is relevant operators that must keep $X$ in thermal equilibrium. For an $X$ with the interactions introduced in Section 4.4.1, the annihilations have rates that scale with temperature as $\Gamma \sim n_X \langle\sigma v\rangle \sim T$ for $T \gtrsim m_X, m_h$ (where $n_X$ is the number density of $X$ and $\langle\sigma v\rangle$ is its thermally averaged annihilation cross section). However, in a radiation-dominated universe, $H \sim T^2$. Thus, at high enough temperatures, $X$ is not in thermal equilibrium with the plasma and it is only once the universe cools enough that it may thermalise. Then, as the temperature drops, $XX \to q\bar{q}$ annihilations become suppressed by the Higgs mass and subsequent Boltzmann suppression causes $X$ to freeze-out. Note that the rates of these annihilation processes are controlled by the coupling $\lambda_x$, independently of $x$, which is unconstrained by itself (other processes mediated by $\lambda_x x$ are found to be subdominant in the ensuing analysis, for the range of $\lambda_x$ over which

thermal production is successful). If the coupling is too weak to begin with, then the $X$ never thermalises and thermal production cannot happen. Thermal production therefore requires a careful balancing of parameters - small coupling $\lambda_x$ is preferred for $X$ to freeze-out hot and as early as possible, but the coupling is bounded from below by the requirement that $X$ reach thermal equilibrium. This combination of constraints severely restricts the size of the parameter space over which thermal production is viable to cases in which the coupling is selected so that $X$ enters and departs from thermal equilibrium at close to the same temperature.

To obtain numerical predictions for this scenario, the calculation of Section 4.4.2 was modified to account for the time after the freeze-out of $X$ before it becomes non-relativistic. During this period we use (4.16) and (4.17) for the energy density of the $X$, approximating decays as being negligible, before switching over to (4.43) when the temperature drops below the mass of the $X$. The approximation that the $X$ does not decay appreciably while it is relativistic must be good if there is to be sufficient time for it to grow to dominate between becoming non-relativistic and decaying. The decay width of $X$ was fixed to $5 \times 10^{-21}$ GeV, corresponding to a reheating temperature close to the $\sim 10$ MeV lower limit, in order to maximise the amount of time over which the energy density of $X$ may grow relative to the SM plasma, thereby providing the greatest possible reheating.

The predictions for $\Delta N_{\text{eff}}$ from a thermally produced $X$ are shown in Figure 4.10 for the small regions of parameter space where this is viable, with $f/v = 4$. We find that the dominant annihilation channels over this region are $XX \to t\bar{t}$ and $XX \to b\bar{b}$, mediated by the light Higgs, as well as their twin analogues, mediated by the heavy Higgs. As expected, the primordial energy density in the twin sector is too large compared to that generated by the $X$ for the asymmetric reheating to be effective when $m_X$ is too light ($\lesssim 100$ GeV in this case). Similarly, when the coupling is too strong, the X is held

in equilibrium for longer and freezes-out underabundant compared to the twin energy density. However, when the coupling is too weak (the gray region), $X$ never thermalises to begin with (close to the boundary with this region, $X$ freezes-out almost immediately after thermalising). The peak in the contours occurs because of the "$H$-funnel" in which the twin Higgs resonantly enhances annihilations into twin quarks. All of this region will be testable by CMB-S4.

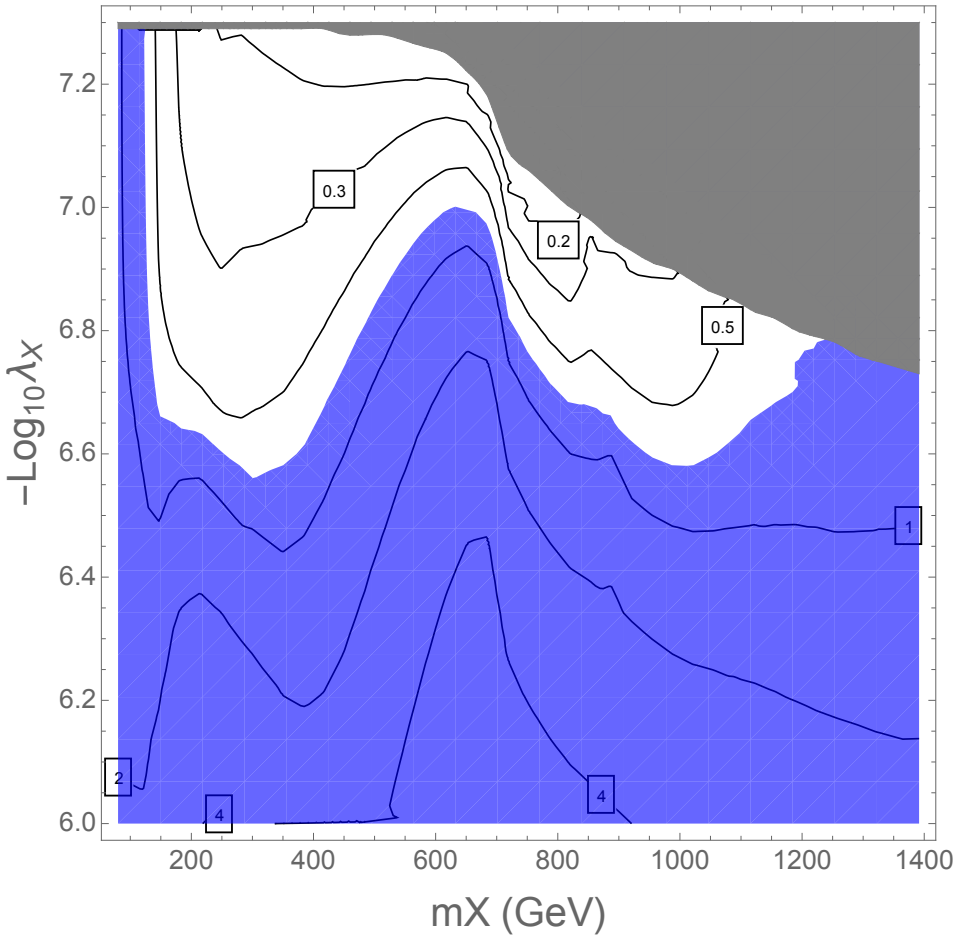



Figure 4.10: Parameter space where thermal production of $X$ gives a large enough relic abundance to dilute the twin sector, for $f/v = 4$. In the gray region, the coupling is too weak for $X$ to ever reach thermal equilibrium. The blue region is in tension with recent Planck measurements of $\Delta N_{\rm eff}$, whereas all of the white region will be tested by CMB-S4. Predictions presented here for $\Delta N_{\rm eff}$ close to the gray boundary are more uncertain because of the high sensitivity of the freeze-out temperatures to the coupling.

## 4.5 Twinflation

As an alternative to the model presented above of late, out-of-equilibrium decays of a $\mathbb{Z}_2$-symmetric scalar, one may imagine that the field driving primordial inflation reheats only the Standard Model to below the decoupling temperature of the two sectors. Production of the twin particles then ceases at some time after the temperature drops below the decoupling temperature during reheating.

To make this consistent with a softly-broken $\mathbb{Z}_2$ symmetry, we extend the inflationary sector and introduce a 'twinflaton' that couples solely to the twin sector. The combined inflationary and twinflationary sectors respect the $\mathbb{Z}_2$ symmetry. However, if the two sectors are entirely symmetric then one generally expects both inflationary dynamics to happen coincidentally, which would result in identical reheating. We therefore rely on soft $\mathbb{Z}_2$-breaking to give an asymmetry between the two sectors that causes the twinflationary sector to dominate the universe first. With the right arrangement we can end up with two distinct periods of inflation - a first caused by the "twinflaton" and a second that then reheats the Standard Model to below the decoupling temperature, having diluted the sources of twin-sector reheating from the first period.

One simple mechanism for $\mathbb{Z}_2$-breaking which is well-suited for introducing asymmetry to inflationary sectors is to introduce an additional $\mathbb{Z}_2$-odd scalar field $\eta$ (as was done in (51)). This admits linear and quadratic interactions to antisymmetric and symmetric combinations of the inflationary sector fields, respectively. When $\eta$ acquires a vev, this introduces an asymmetry in the fields to which it was coupled, dependent on the combination of its vev and its couplings. If $\eta$ is coupled to both the inflationary sectors and the Higgs sectors, it could be the sole source of $\mathbb{Z}_2$-breaking in a twinflationary theory. One may generally imagine that, in some UV completion, the mechanism that softly breaks the symmetry in the Higgs potential could also be the origin of the soft breaking of the

inflationary sector.

Cosmologically, this possibility may have similar observational signatures as the model discussed in Section 4.4, where the amount of twin-sector dark radiation is determined by the partial widths of the inflaton of the second inflationary epoch. If this dominantly couples to the SM, then $\Delta N_{\text{eff}}$ will be suppressed which, while successfully resolving the cosmological problems of the Mirror Twin Higgs, may also be observationally inaccessible. However, additional, distinctly inflationary signatures may make this potentially testable by other cosmological observations.

The mechanism of twinflation completes a catalog of models of asymmetric reheating by late decays, which may be indexed by representations of the twin parity: the case of a $\mathbb{Z}_2$-even particle, in which a kinematic asymmetry in the partial widths provides the reheating asymmetry, the case of a $\mathbb{Z}_2$-odd particle, which can also provide the spontaneous $\mathbb{Z}_2$-breaking required in the Higgs potential, and the case where two distinct, long-lived particles couple to each sector, which may also be related to inflation.

### 4.5.1 Toy Model

As a toy model we here consider 'twinning' the simple $\varphi^2$ chaotic inflation scenario. The inflationary dynamics in this case are easy to understand and we have the additional benefit that this inflationary model has been considered in the literature before as 'Double Inflation' (see (52), (53) and (54)). We furthermore specialize to 'double inflation with a break', where there are two distinct periods of inflation which produces a step in the power spectrum, and we consider the constraints that this places on our model. In this case, it is assumed that each inflaton field couples and therefore decays dominantly into the sector to which it belongs. We will comment briefly on the case without a break and the additional signals one could look for in that case.

The potential of the inflationary sector for inflaton $\varphi_A$ and twinflaton $\varphi_B$ is

$$V = \frac{1}{2}m_A^2\varphi_A^2 + \frac{1}{2}m_B^2\varphi_B^2, \tag{4.48}$$

where $m_A \neq m_B$ may arise from soft $\mathbb{Z}_2$- breaking, perhaps related to the soft $\mathbb{Z}_2$-breaking in the Higgs potential. In order for the 'twinflation' to occur first, we require that the energy of the $B$ field initially dominates the energy density of the universe. We take the initial positions of the fields to be the same and $m_B^2 \gg m_A^2$.[5] Call $\varphi_A(0) = \varphi_B(0) = n\sqrt{2}M_{\rm pl} = n\varphi_c$, where $\varphi_c$ is the critical value at which inflation stops and $m_B = rm_A = rm$ with $n, r > 1$. The inflationary dynamics are then those of slowly-rolling scalar fields. At some point in the early universe we imagine that the slow-roll approximation holds for both fields and the inflationary sector dominates the universe. The dominating field then slow-rolls down its potential for $\frac{n^2-1}{2}$ $e$-folds, while the lighter field's velocity is suppressed by approximately $\frac{\varphi_A}{r^2\varphi_B}$. Solving the system numerically reveals that the motion of $\varphi_A$ during this period can be neglected entirely.

After $\varphi_B$ reaches the critical value $\sqrt{2}M_{\rm pl}$, it stops slow-rolling and begins oscillating around the minimum of its potential. For there to be two distinct periods of inflation, there must be a period where these oscillations dominate the universe, which requires that the energy densities of each inflaton $\rho_A$ and $\rho_B$ satisfy $\rho_B(\varphi_c) = r^2m^2M_{\rm pl}^2 > \rho_A(\varphi(0)) = n^2m^2M_{\rm pl}^2$ and therefore $r > n$. For a $\varphi_B^2$ potential, the energy in these oscillations redshifts as $\rho_B \sim a^{-3}$. Eventually, the energy density in $\varphi_B$ drops below that of $\varphi_A$ and a new epoch of inflation, driven by $\varphi_A$, begins. This provides a further $\frac{n^2-1}{2}$ $e$-folds of inflation to give $n^2 - 1$ in total, while the $B$-sector energy density is diluted away.

Note that in order for our toy model to reheat below the decoupling temperature of the

[5]Note that merely giving the twin field a much larger initial condition does not instigate twinflation. The dynamics of the subdominant field in this case are such that it will track the dominant field and both will reach the critical value at the same time. This is easily confirmed numerically.

two sectors, reheating must occur well after the end of inflation. If, during the coherent oscillation of an inflaton, it becomes the case that the inflaton decay width $\Gamma \sim H$, then reheating will occur and result in temperature $T_{\rm reheat} \sim 0.1\sqrt{\Gamma M_{\rm pl}}$. However, if $\Gamma \gg H$ when inflation ends, then all of the energy in the inflaton is immediately transferred and we instead have reheating temperature $T_{\rm reheat} \sim 0.1\sqrt{m_\alpha M_{\rm pl}}$ for an inflaton of mass $m_\alpha$. But in order for $T_{\rm reheat} \lesssim 1$ GeV, it is required that $m_\alpha \lesssim 10^{-7}$ eV, so this possibility that the inflaton is short lived is not viable. The procedure of twinning inflationary potentials may be generalised to other, more realistic models, provided that this constraint upon the reheating temperature can be satisfied.

### 4.5.2 Observability

One could always make a twinflationary scenario consistent with observational constraints by letting the second inflationary period of inflation last long enough. In our toy model, this would correspond to setting $n$ high enough that the momentum modes which left the horizon during the first inflation have not yet re-entered the horizon - such a scenario would look exactly like single-field chaotic inflation.

Alternatively, we may also allow for $n$ small enough that all the momentum modes that left the horizon during the second inflation are currently sub-horizon. In this case, fluctuations at large enough wavenumbers (equivalently, small enough length scales) are 'processed' (cross the horizon) at a different inflationary energy scale than those that were processed earlier, giving a step in the power spectrum. While Planck has measured the primordial power spectrum for modes with $10^{-4}$ Mpc$^{-1} \lesssim k \lesssim 0.3$ Mpc$^{-1}$ (where the lower bound is set by the fact that smaller modes have not yet re-entered the horizon), proposed CMB-S4 experiments will increase this range (27) somewhat, as will be discussed further below. We wish to show that the power spectrum of our toy model is

not ruled out and, furthermore, may be observed in the coming decades.

The height of the step in the primordial power spectrum is determined by the energy scale of each period of inflation, so modes crossing the horizon in the second inflationary period should be suppressed by a factor of $r^2 > n^2 \gtrsim 25$ compared to those exiting in the first period. This degree of suppression is ruled out by Planck for the range of modes over which it has reconstructed the power spectrum (50). A computation of the primordial power spectrum for double inflation was given in (53). It was found that significant damping does not occur for modes which cross outside the horizon during the first inflationary period, re-enter during the inter-inflationary period and again cross the horizon during the second inflationary period. It is only those scales which first cross the horizon during the second inflationary period that are significantly damped (although other features in the shape, such as oscillations, may be present for modes that are subhorizon during the intermediate period).

The relation of this characteristic scale to present-day observables is easily done using the framework given in (55). Let the subscripts $a, b, c, d, e$ respectively correspond to the beginning of the first inflationary period, the end of that period, the beginning of the second inflationary period, the end of that period, and the beginning of radiation domination. During the coherent oscillation periods, the inflaton acts as matter and the energy density falls as $\rho \propto a^{-3}$. Let $k_i$ be the momentum whose mode is horizon-size at the $i$ epoch; $k_i = a_i H_i$. The scales $k_i$ can be related using the number of $e$-folds in each period, which are themselves determined from the first Friedmann equation. Denoting $N_{ij} = \ln \frac{a_j}{a_i}$, we have $k_a = e^{-N_{ab}} n k_b$, $k_b = e^{\frac{1}{2}N_{bc}} k_c$ and similarly for the other characteristic modes, where, in particular, slow-roll inflation predicts that $N_{ab} = N_{cd} = \frac{n^2-1}{2}$. The evolution of the characteristic momentum scales is shown schematically in Figure 4.11.

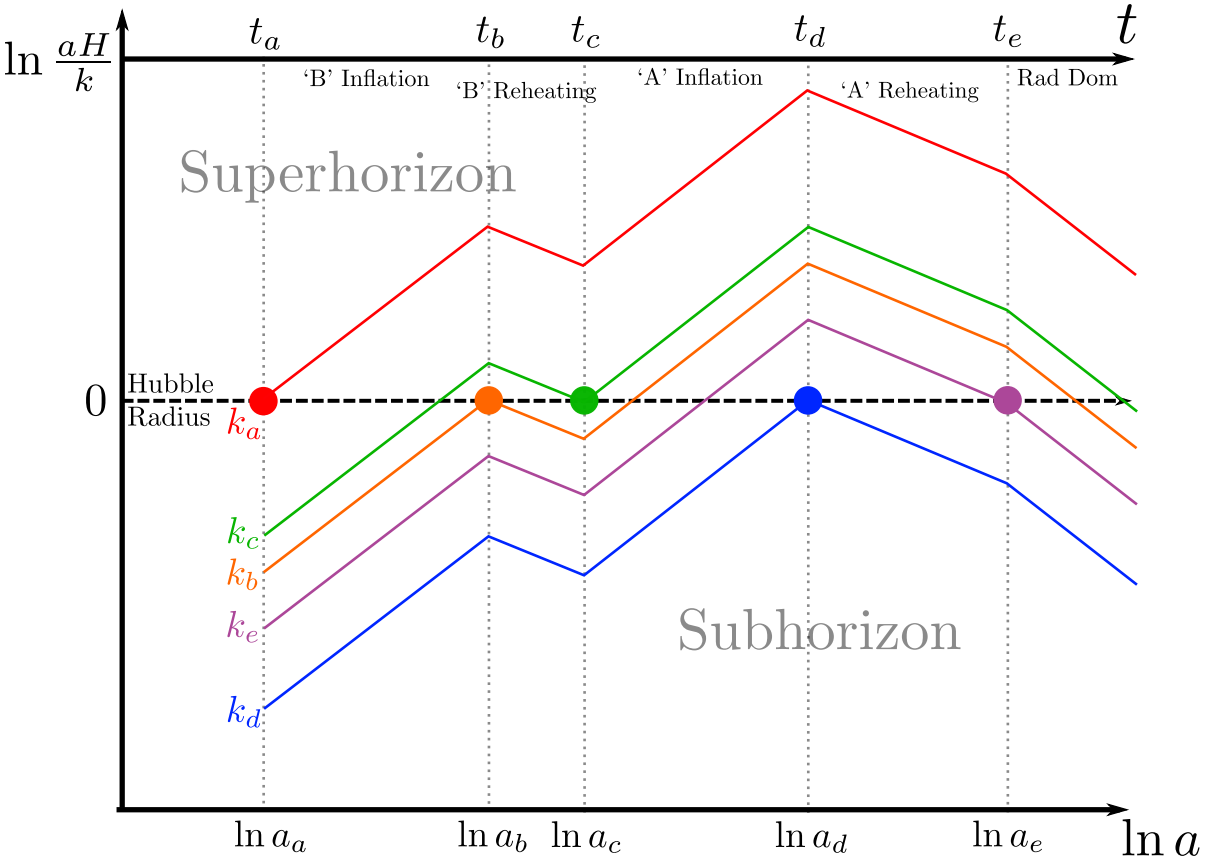



Figure 4.11: Schematic evolution of the characteristic scales in Twinflation, as seen by comparing wavenumbers to the Hubble radius over time. Note that the time axis is not a linear scale.

Finally, $k_e$ can be determined using the conservation of comoving entropy:

$$k_e = \frac{\pi g_\star^{1/3}(T_0) g_\star^{1/6}(T_{\rm reheat}) T_0 T_{\rm reheat}}{3\sqrt{10} M_{\rm pl}}, \tag{4.49}$$

where $T_0$ and $a_0$ are the temperature and scale factor today and $T_{\rm reheat}$ is the reheating temperature (which is sufficiently low that only SM particles are produced). We work explicitly with the convention $a_0 = 1$. The characteristic modes associated with the break can then be determined.

As mentioned above, (53) shows that damping occurs for modes that exit the horizon only during the second inflationary period, so we should take the characteristic damping

scale to be the smallest such scale, which here corresponds roughly to $k_b$ This can be determined as

$$\begin{aligned} k_b &= n e^{\frac{1}{2}N_{bc} - N_{cd} + \frac{1}{2}N_{de}} k_e \\ &= n \left(\frac{r}{n}\right)^{1/3} \exp\left(-\frac{n^2-1}{2}\right) \left[\frac{\frac{1}{2}m^2 M_{\rm pl}^2}{\frac{\pi^2}{30} g_\star(T_{\rm reheat}) T_{\rm reheat}^4}\right]^{1/6} \frac{\pi g_\star^{1/3}(T_0) g_\star^{1/6}(T_{\rm reheat}) T_0 T_{\rm reheat}}{3\sqrt{10} M_{\rm pl}} \end{aligned} \tag{4.50}$$

where $k_c$ only differs by the factor of $(r/n)^{1/3}$ (which is roughly close to unity). Once again, between $k_b$ and $k_c$ are oscillatory features, so $k_b$ should merely be taken as the rough characteristic scale of the damping.

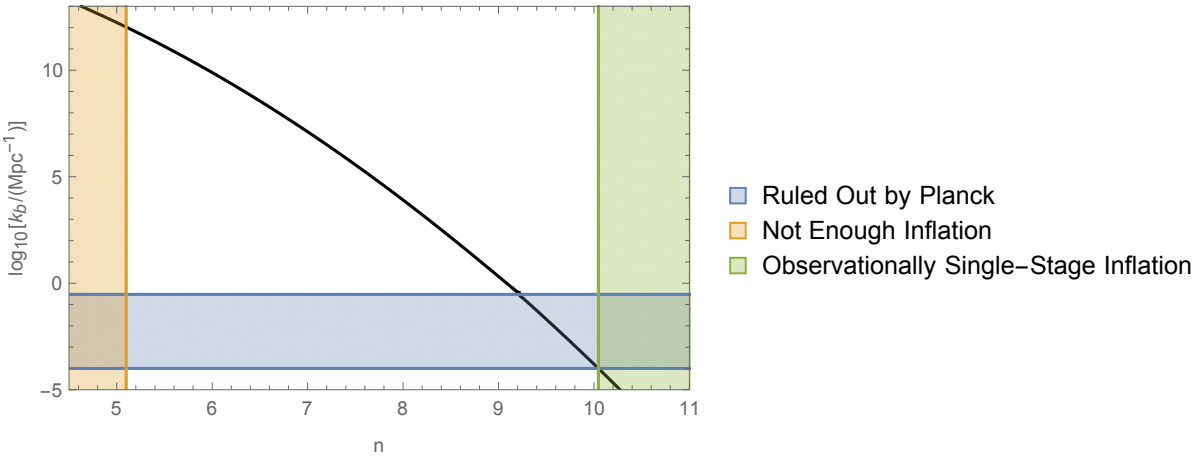



Figure 4.12: The prediction for the characteristic suppression scale as a function of the initial values of the fields. The mapped regions should be interpreted not as having hard boundaries, but rather fuzzy endpoints where they break down. Here we have used $T_{\rm reheat} = 10$ MeV and $r = 2n$.

Now the characteristic damping scale is determined by $m$, $n$, $r$, and $T_{\rm reheat}$. Our observational bound on $k_b$ is that Planck has not seen this suppression on momentum scales at which it has been able to reconstruct the primordial power spectrum from the angular temperature anisotropy power spectrum, which is roughly $k \lesssim 0.3$ Mpc$^{-1}$. We have constraints on the reheating temperature from rethermalization of the twin sector or interrupted big bang nucleosynthesis $10\,\text{MeV} \lesssim T_{\rm reheat} \lesssim 1$ GeV, on having a period

of intermediate matter domination between the two inflations $r > n$ and on the total number of $e$-folds $n^2 - 1 \gtrsim 25$ to solve cosmological problems. Note that we require fewer $e$-folds of inflation than is typically assumed in the standard cosmology. Since the low reheating temperature gives fewer $e$-folds from reheating up to today, less inflation is needed to explain the large causal horizon and flatness.

The normalization of the spectrum provides a further constraint, the most recent measurement of which come from Planck (50). The scalar power spectrum at $k_\star = 0.05\ \mathrm{Mpc}^{-1}$ is measured to be $\mathcal{P}_\mathcal{R}(k_\star) = e^{3.094\pm0.034} \times 10^{-10}$. Then for $k_\star < k_c$ (i.e. $k_\star$ having left the horizon during the first period of inflation and not re-entered before the second, so no deviation from single-field inflation would be seen at this scale), the spectrum of (53) yields the constraint

$$2.03 \times 10^{-6} = \frac{r^2 m^2}{M_{\mathrm{pl}}^2} \ln\left(\frac{k_b}{k_\star}\right)\left(\ln\frac{k_b}{k_\star} + \frac{n^2}{2}\right). \tag{4.51}$$

The characteristic scale (4.50) depends much more strongly on $n$ than it does on any of the other parameters. In Figure 4.12, we give a rough idea of the scale as a function of $n$, having set $T_{\mathrm{reheat}} = 10$ MeV and $r = 2n$, while $m$ is chosen to satisfy the normalization condition. We also show the constraint on $k_b$ set by Planck. Note again that the region described as "observationally single-stage inflation" *does* still provide a solution to the problem of reconciling cosmology with the mirror Twin Higgs.

CMB-S4 will improve the constraint on $k_b$ through its improved measurement of polarization anisotropies (27). With only precision measurements of temperature anisotropies, the un-lensed power spectrum cannot be so easily reconstructed from the lensed spectrum. The effects of gravitational lensing of CMB place an upper limit on the size of primordial temperature anisotropies that can be measured (56), which Planck has saturated. However, the polarization anisotropy power spectrum allows the removal of lensing

noise from the temperature spectrum so that higher primordial modes can be detected. The polarization power spectrum itself also gives us another window into the high-$\ell$ modes of the primordial power spectrum, as the signal does not become dominated by polarized foreground sources until higher scales near $\ell \sim 5000$. CMB-S4 is projected to make cosmic variance limited measurements of both the temperature and polarization anisotropy power spectra up to the modes where they become foreground-contaminated and so provide additional information on the shape of the primordial power spectrum (27). The map from measurements of angular modes $\ell$ to contraints on spatial modes $k$ depends on the evolution of the power spectrum between inflation and the CMB, so forecasting constraints requires careful study. However, these improvements will not test most of the parameter space presented in Figure 4.12, where the step is predicted on extremely small distance scales.

We have discussed a twinflationary model of double inflation with a break for simplicity, but there is a parametric regime where double inflation without a break gives the required amount of asymmetric reheating into the Standard Model. With two periods of inflation, the second period dilutes the energy density of the heavier field sufficiently that there is no observable signal of it produced in reheating. However, even with only one period, inflation can continue for long enough after the inflaton turns the corner in field space such that, at late times, the fraction of the inflaton in the $B$ state relative to the $A$ state is small enough that the expected energy densities that are transferred into each sector satisfy $\rho_B/\rho_A < 0.1$. This occurs as long as $r \gtrsim 1.2$, assuming that the mixing angle of the slow-rolling field with the $\varphi_A$ and $\varphi_B$ fields entirely determines the fraction of its energy that reheats each sector. There is thus a much larger range of $r$ where this toy model of inflation passes $N_{\text{eff}}$ bounds than our above analysis shows. The resulting imprint on the CMB could resemble that of the long-lived decay model of Section 4.4, with $\Delta N_{\text{eff}}$ again being related to the ratio of branching fractions, although

this is dependent upon the UV completion of the Twin Higgs.

When there is only one period of inflation, the step is smoothed out and less pronounced and it is necessary to locate the feature numerically. Furthermore, having multiple degrees of freedom available allows for non-trivial evolution of momentum modes after they become super-horizon, which does not occur in single-field inflation but may be calculated from the full solution to the field equations (54). While a twinned potential leading to two periods of inflation generally predicts a step in the power spectrum, when there is no break the predictions, and thus constraints, this prediction become more model-dependent. Therefore we leave detailed predictions in that case for future study using realistic models and merely state that the range of $r = 1$ to $n$ interpolates between the single field spectrum and that with a step, as one would expect.

There are also at least two other detectable effects one might expect in double inflation without a break and in general realistic twinflationary models. Interactions between inflaton fields may produce primordial non-Gaussianities, while the presence of additional oscillating degrees of freedom may produce isocurvature perturbations. These do not appear in our toy model because the heavy field is exponentially damped during the second inflation. CMB-S4 is projected to improve Planck's bounds on non-Gaussianities by a factor of $\sim 2$ and on isocurvature perturbations by perhaps an order of magnitude (though model-independent projections have not been made), so may be able to detect or place useful constraints on realistic twinflationary models (27).

We have introduced twinflation as a mirror Twin Higgs model which suppresses the cosmological effects of twin light degrees of freedom. It extends the mirror symmetry to the inflationary sector. The soft $Z_2$ symmetry-breaking of the Higgs sector may be used in the inflationary sector to cause distinct periods of inflation. There exists a parametric region where this is cosmologically indistinct from single-stage inflation, but also another in which it may be observable. As the direct product of inflation and the Mirror Twin

Higgs, this is in some sense a minimal solution.

## 4.6 Conclusion

In this work we have considered scenarios in which cosmology provides meaningful insight on solutions to the electroweak hierarchy problem. In particular, we have demonstrated several simple mechanisms in which the cosmological history of a mirror Twin Higgs model is reconciled with current CMB constraints and provides signatures accessible in future CMB experiments. In the case of out-of-equilibrium decays, we have found that decays of $\mathbb{Z}_2$-even scalars sufficiently dilute the energy density in the twin sector without the addition of any new sources of $\mathbb{Z}_2$-breaking. In much of the parameter space, the residual contribution to $\Delta N_{\text{eff}}$ is directly proportional to the ratio of vacuum expectation values $v^2/f^2$ parameterizing the mixing between Standard Model and twin sectors (as well as the tuning of the electroweak scale), and may be within reach of CMB-S4 experiments. In the case of twinflation, we have found that a (broken) $\mathbb{Z}_2$-symmetric inflationary sector may successfully dilute the energy density in the twin sector, as well as potentially leave signatures in the form of a step in the primordial power spectrum or in departures of primordial perturbations from adiabaticity and Gaussianity. In both cases, these models raise the tantalizing possibility that signatures of electroweak naturalness may first emerge in the CMB, rather than the LHC.

There are a variety of possible directions for future work. Here we have focused on the cosmological consequences of late-decaying scalars and twinned inflationary sectors without specifying their origin in a microscopic model. It would be interesting to construct complete models (where, e.g., supersymmetry or compositeness protect the scale $f$ from UV contributions) in which the existence and couplings of late-decaying scalars arise as intrinsic ingredients of the UV completion. Likewise, we have considered only

a toy model of twin chaotic inflation; it would be interesting to see if twinflation may be realized in complete inflationary models that match the observed spectral index and constraints on the tensor-to-scalar ratio.

While we have taken care to ensure that our scenarios respect the well-measured cosmological history beneath $T \sim 1$ MeV, we have not addressed the origin of the observed baryon asymmetry. In the case of out-of equilibrium decays, there are a number of possibilities. It is plausible that a somewhat larger baryon asymmetry is generated through various conventional mechanisms and diluted by late decays. Alternatively, the decay mechanism itself may possibly be expanded to generate a baryon asymmetry or some other late decay may generate the baryon asymmetry below $\sim 1$ GeV. In the case of twinflation, inflationary dilution of pre-existing baryon asymmetry requires that baryogenesis occur in association with reheating or via another mechanism at temperatures below $\sim 1$ GeV. It would be worthwhile to study models for the baryon asymmetry consistent with these scenarios. Steps in this direction have been taken in (17), which attempted to relate this to asymmetric dark matter in the twin sector.

Likewise, any investigation of dark matter, be it related directly to the twin mechanism or otherwise, must also address implications of the dilution. Previous work attempting to construct dark matter candidates in the twin sector (11–18)) has relied upon explicit $\mathbb{Z}_2$-breaking that is not present in the mirror model. Dark matter may alternatively be unrelated to the Twin Higgs mechanism, such as a a WIMP in some minimal extension of the electroweak sector that freezes-out as an overabundant thermal relic and is then diluted to the observed density during reheating. Alternatively, it may be that the dark matter abundance is produced directly during reheating. It would be interesting to study extensions of our scenarios that incorporate dark matter candidates directly related to the mechanism of dilution.

Finally, we have only approximately parameterized Planck constraints and the reach

of CMB-S4 on twin neutrinos and twin photons. Ultimately, more precise constraints and forecasts may be obtained via numerical CMB codes. This strongly motivates the future study of CMB constraints on scenarios with three sterile neutrinos and additional dark radiation whose temperatures differ from the Standard Model thermal bath.

# Acknowledgments

We thank Zackaria Chacko, Yanou Cui, Paddy Fox, Daniel Green, and Roni Harnik for useful discussions, and Nima Arkani-Hamed for enlightening remarks on $N$naturalness. The work of NC, SK, and TT is supported in part by the US Department of Energy under the grant DE-SC0014129. NC acknowledges the hospitality of the Kavli Institute for Theoretical Physics and the corresponding support of the National Science Foundation under Grant No. NSF PHY11-25915.

# Bibliography

[1] Z. Chacko, H.-S. Goh and R. Harnik, *The Twin Higgs: Natural electroweak breaking from mirror symmetry*, *Phys. Rev. Lett.* **96** (2006) 231802, [hep-ph/0506256].

[2] Planck collaboration, P. A. R. Ade et al., *Planck 2015 results. XIII. Cosmological parameters*, 1502.01589.

[3] R. H. Cyburt, B. D. Fields, K. A. Olive and T.-H. Yeh, *Big Bang Nucleosynthesis: 2015*, *Rev. Mod. Phys.* **88** (2016) 015004, [1505.01076].

[4] N. Craig, S. Knapen and P. Longhi, *Neutral Naturalness from Orbifold Higgs Models*, *Phys. Rev. Lett.* **114** (2015) 061803, [1410.6808].

[5] N. Craig, S. Knapen and P. Longhi, *The Orbifold Higgs*, *JHEP* **03** (2015) 106, [1411.7393].

[6] M. Geller and O. Telem, *Holographic Twin Higgs Model*, *Phys. Rev. Lett.* **114** (2015) 191801, [1411.2974].

[7] R. Barbieri, D. Greco, R. Rattazzi and A. Wulzer, *The Composite Twin Higgs scenario*, *JHEP* **08** (2015) 161, [1501.07803].

[8] M. Low, A. Tesi and L.-T. Wang, *Twin Higgs mechanism and a composite Higgs boson*, *Phys. Rev.* **D91** (2015) 095012, [1501.07890].

[9] N. Craig, A. Katz, M. Strassler and R. Sundrum, *Naturalness in the Dark at the LHC*, *JHEP* **07** (2015) 105, [1501.05310].

[10] N. Craig, S. Knapen, P. Longhi and M. Strassler, *The Vector-like Twin Higgs*, *JHEP* **07** (2016) 002, [1601.07181].

[11] R. Barbieri, L. J. Hall and K. Harigaya, *Minimal Mirror Twin Higgs*, 1609.05589.

[12] I. Garcia Garcia, R. Lasenby and J. March-Russell, *Twin Higgs WIMP Dark Matter*, *Phys. Rev.* **D92** (2015) 055034, [1505.07109].

[13] N. Craig and A. Katz, *The Fraternal WIMP Miracle*, *JCAP* **1510** (2015) 054, [1505.07113].

[14] I. Garcia Garcia, R. Lasenby and J. March-Russell, *Twin Higgs Asymmetric Dark Matter*, *Phys. Rev. Lett.* **115** (2015) 121801, [1505.07410].

[15] M. Farina, *Asymmetric Twin Dark Matter*, *JCAP* **1511** (2015) 017, [1506.03520].

[16] M. Freytsis, S. Knapen, D. J. Robinson and Y. Tsai, *Gamma-rays from Dark Showers with Twin Higgs Models*, *JHEP* **05** (2016) 018, [1601.07556].

[17] M. Farina, A. Monteux and C. S. Shin, *Twin mechanism for baryon and dark matter asymmetries*, *Phys. Rev.* **D94** (2016) 035017, [1604.08211].

[18] V. Prilepina and Y. Tsai, *Reconciling Large And Small-Scale Structure In Twin Higgs Models*, 1611.05879.

[19] S. Chang, L. J. Hall and N. Weiner, *A Supersymmetric twin Higgs*, *Phys. Rev.* **D75** (2007) 035009, [hep-ph/0604076].

[20] N. Craig and K. Howe, *Doubling down on naturalness with a supersymmetric twin Higgs*, *JHEP* **03** (2014) 140, [1312.1341].

[21] M. Drees, F. Hajkarim and E. R. Schmitz, *The Effects of QCD Equation of State on the Relic Density of WIMP Dark Matter*, *JCAP* **1506** (2015) 025, [1503.03513].

[22] D. C. Kelly, *Collision Operator for a Relativistic Lorentz Gas*, *Phys. Fluids* **12** (1969) 799.

[23] J. Edsjo and P. Gondolo, *Neutralino relic density including coannihilations*, *Phys. Rev.* **D56** (1997) 1879–1894, [hep-ph/9704361].

[24] P. Adshead, Y. Cui and J. Shelton, *Chilly Dark Sectors and Asymmetric Reheating*, *JHEP* **06** (2016) 016, [1604.02458].

[25] G. Mangano, G. Miele, S. Pastor, T. Pinto, O. Pisanti and P. D. Serpico, *Relic neutrino decoupling including flavor oscillations*, *Nucl. Phys.* **B729** (2005) 221–234, [hep-ph/0506164].

[26] Z. Hou, R. Keisler, L. Knox, M. Millea and C. Reichardt, *How Massless Neutrinos Affect the Cosmic Microwave Background Damping Tail*, *Phys. Rev.* **D87** (2013) 083008, [1104.2333].

[27] CMB-S4 Collaboration, *CMB-S4 Science Book - First Edition*, 1610.02743.

[28] C. Brust, D. E. Kaplan and M. T. Walters, *New Light Species and the CMB*, *JHEP* **12** (2013) 058, [1303.5379].

[29] S. Dodelson, A. Melchiorri and A. Slosar, *Is cosmology compatible with sterile neutrinos?*, *Phys. Rev. Lett.* **97** (2006) 041301, [astro-ph/0511500].

[30] D. Baumann, D. Green, J. Meyers and B. Wallisch, *Phases of New Physics in the CMB*, *JCAP* **1601** (2016) 007, [1508.06342].

[31] M. Archidiacono, S. Hannestad, A. Mirizzi, G. Raffelt and Y. Y. Y. Wong, *Axion hot dark matter bounds after Planck*, *JCAP* **1310** (2013) 020, [1307.0615].

[32] T. D. Jacques, L. M. Krauss and C. Lunardini, *Additional Light Sterile Neutrinos and Cosmology*, *Phys. Rev.* **D87** (2013) 083515, [1301.3119].

[33] J. Lesgourgues and S. Pastor, *Neutrino mass from Cosmology*, *Adv. High Energy Phys.* **2012** (2012) 608515, [1212.6154].

[34] S. Dodelson, E. Gates and A. Stebbins, *Cold + hot dark matter and the cosmic microwave background*, *Astrophys. J.* **467** (1996) 10–18, [astro-ph/9509147].

[35] C. Howlett, A. Lewis, A. Hall and A. Challinor, *CMB power spectrum parameter degeneracies in the era of precision cosmology*, *jcap* **4** (Apr., 2012) 027, [1201.3654].

[36] E. Di Valentino, S. Gariazzo, M. Gerbino, E. Giusarma and O. Mena, *Dark Radiation and Inflationary Freedom after Planck 2015*, *Phys. Rev.* **D93** (2016) 083523, [1601.07557].

[37] M. Costanzi, B. Sartoris, M. Viel and S. Borgani, *Neutrino constraints: what large-scale structure and CMB data are telling us?*, *JCAP* **1410** (2014) 081, [1407.8338].

[38] S. Gariazzo, C. Giunti, M. Laveder, Y. F. Li and E. M. Zavanin, *Light sterile neutrinos*, *J. Phys.* **G43** (2016) 033001, [1507.08204].

[39] N. Arkani-Hamed, T. Cohen, R. T. D'Agnolo, A. Hook, H. D. Kim and D. Pinner, *Nnaturalness*, 1607.06821.

[40] M. Reece and T. Roxlo, *Nonthermal Production of Dark Radiation and Dark Matter*, 1511.06768.

[41] L. Randall, J. Scholtz and J. Unwin, *Flooded Dark Matter and S Level Rise*, *JHEP* **03** (2016) 011, [1509.08477].

[42] E. W. Kolb and M. S. Turner, *The Early Universe*, *Front. Phys.* **69** (1990) 1–547.

[43] P. F. de Salas, M. Lattanzi, G. Mangano, G. Miele, S. Pastor and O. Pisanti, *Bounds on very low reheating scenarios after Planck*, *Phys. Rev.* **D92** (2015) 123534, [1511.00672].

[44] A. Djouadi, *The Anatomy of electro-weak symmetry breaking. I: The Higgs boson in the standard model*, *Phys. Rept.* **457** (2008) 1–216, [hep-ph/0503172].

[45] M. Drees and K.-i. Hikasa, *Heavy Quark Thresholds in Higgs Physics*, *Phys. Rev.* **D41** (1990) 1547.

[46] M. Baumgart and A. Katz, *Implications of a New Light Scalar Near the Bottomonium Regime*, *JHEP* **08** (2012) 133, [1204.6032].

[47] J. D. Clarke, R. Foot and R. R. Volkas, *Phenomenology of a very light scalar (100 MeV ¡ $m_h$ ¡ 10 GeV) mixing with the SM Higgs*, *JHEP* **02** (2014) 123, [1310.8042].

[48] U. Haisch and J. F. Kamenik, *Searching for new spin-0 resonances at LHCb*, *Phys. Rev.* **D93** (2016) 055047, [1601.05110].

[49] Topical Conveners: K.N. Abazajian, J.E. Carlstrom, A.T. Lee collaboration, K. N. Abazajian et al., *Neutrino Physics from the Cosmic Microwave Background and Large Scale Structure*, *Astropart. Phys.* **63** (2015) 66–80, [1309.5383].

[50] Planck collaboration, P. A. R. Ade et al., *Planck 2015 results. XX. Constraints on inflation*, 1502.02114.

[51] Z. G. Berezhiani, A. D. Dolgov and R. N. Mohapatra, *Asymmetric inflationary reheating and the nature of mirror universe*, *Phys. Lett.* **B375** (1996) 26–36, [hep-ph/9511221].

[52] J. Silk and M. S. Turner, *Double Inflation*, *Phys. Rev.* **D35** (1987) 419.

[53] D. Polarski and A. A. Starobinsky, *Spectra of perturbations produced by double inflation with an intermediate matter dominated stage*, *Nucl. Phys.* **B385** (1992) 623–650.

[54] D. Wands, N. Bartolo, S. Matarrese and A. Riotto, *An Observational test of two-field inflation*, *Phys. Rev.* **D66** (2002) 043520, [astro-ph/0205253].

[55] S. E. Hong, H.-J. Lee, Y. J. Lee, E. D. Stewart and H. Zoe, *Effects of thermal inflation on small scale density perturbations*, *JCAP* **1506** (2015) 002, [1503.08938].

[56] W. Hu and T. Okamoto, *Principal power of the CMB*, *Phys. Rev.* **D69** (2004) 043004, [astro-ph/0308049].