Integral Formulas for Vector Signal Tensor Products

Valentin Heyraud [1] **Zachary Weller-Davies** [1] **Jules Tilly** [1]

Abstract

We derive integral formulas that simplify the Vector Signal Tensor Product recently introduced by Xie et al. (2026), which generalizes the Gaunt Tensor Product to antisymmetric couplings. In particular, we obtain explicit closed-form expressions for the antisymmetric analogues of the Gaunt coefficients. This enables us to simulate the Clebsch-Gordan tensor product using a single Vector Signal Tensor Product, yielding up to a $9\times$ reduction in the required tensor product evaluations. Our results enable efficient and practical implementations of the Vector Signal Tensor Product, paving the way for applications of this generalization of Gaunt tensor products in $\mathrm{SO}(3)$-equivariant neural networks. Moreover, we discuss how the Gaunt and the Vector Signal Tensor Products allow to control the expressivity-runtime tradeoff associated with the usual Clebsch-Gordan Tensor Products. Finally, we investigate low rank decompositions of the normalizations of the considered tensor products in view of their use in equivariant neural networks.

1. Introduction

Clebsch-Gordan Tensor Products (CGTP) are the most prominent workhorses of $\mathrm{SO}(3)$-equivariant neural networks, which have proven effective for a wide range of applications involving geometric data. In such models, feature vectors transform according to irreducible representations of $\mathrm{SO}(3)$, and nonlinear interactions between these must preserve this transformation law. The CGTP provides a standard combination mechanism that respects rotational symmetry. In practice, this operation decomposes the interaction between two features into several admissible coupling paths, each corresponding to a different way in which their representations can combine.

[1]InstaDeep, 1 Triton Square, London, NW1 3BF, United Kingdom. Correspondence to: Valentin Heyraud <v.heyraud@instadeep.com>.

Proceedings of the 43rd International Conference on Machine Learning, Seoul, South Korea. PMLR 306, 2026. Copyright 2026 by the author(s).

Unfortunately, CGTP operations are costly; they naively admit a $\mathcal{O}(L^6)$ scaling with the representation order L of the features on which they act. To address this limitation, Luo et al. (2024) proposed using Gaunt Tensor Products (GTP) to accelerate tensor products involving feature vectors with multiple irrep types. Their approach relies on an integral formula for the Clebsch-Gordan coefficients which allows to rewrite the tensor product between feature vectors in terms of triple integrals of the corresponding signals on the sphere.

Subsequently, Xie et al. (2025) investigated the computational and representational properties of GTPs. They compared GTP to alternative tensor product constructions and proposed two efficient implementations, the first one evaluating the required triple integrals using a spherical design (Hardin & Sloane, 1996), and the second, called S2FFT, relying on a spherical design and a $\mathbb{S}^2$ fast Fourier transforms algorithm. As noted by Xie et al. (2025), however, a key limitation of the GTP is that it cannot reproduce the antisymmetric components of the CGTP, preventing key underlying tensor product operations such as the cross product. Preserving all coupling paths is important for the expressive power of equivariant neural networks, as these paths capture the different symmetry-preserving ways in which features can interact.

To circumvent this issue, Xie et al. (2026) recently introduced Vector Signal Tensor Products (VSTP). The VSTP generalizes GTP by constructing vector-valued signals on the sphere. The VSTP exhibits an advantageous complexity scaling in the features irreps order over the CGTP and can encompass both the symmetric and anti-symmetric cases. Specifically, a main result of (Xie et al., 2026) is that a collection of VSTPs can be used to simulate CGTPs, meaning that it is capable of producing non-zero outputs for all admissible coupling paths.

However, the construction proposed in (Xie et al., 2026) yield a cumbersome implementation. While the authors show that the VSTP can be used to recover CGTP up to a factor, each irrep label admits up to three admissible internal angular-momentum couplings. Consequently, simulating the CGTP requires computing a VSTP for every pair of couplings, leading to as many as $3\times 3 = 9$ VSTP operations in practice.

In this paper, we build on the results of (Xie et al., 2026) and

derive simple closed-form integral expressions that serve as antisymmetric analogues of the integral GTP formula (Luo et al., 2024). Our main result, given in Equation (13), shows that the antisymmetric components of the CGTP can be encoded as an integral over the sphere, and we obtain explicit closed-form expressions for the corresponding antisymmetric Gaunt coefficients in Appendix A. Moreover, via Equation (16), we show that the symmetric and antisymmetric contributions of the GTP can be combined into a single integral representation, yielding a simple and practical recipe to simulate a CGTP with only one VSTP, while retaining the favorable scaling properties of the VSTP framework. We also discuss the advantages and limitations of such integral-based tensor products, identifying the regimes in which they can effectively replace standard CGTP.

2. Background

In this section, we briefly summarize the findings from the references (Luo et al., 2024; Xie et al., 2025; 2026) on which we base our results. In the following we assume a basic knowledge of $\mathrm{SO}(3)$-equivariant neural networks and of the representation theory of $\mathrm{SO}(3)$, for which we refer the to Refs. (Unke & Maennel, 2024; Geiger & Smidt, 2022) and references therein.

We work with the basis of real spherical harmonics, as it is the basis commonly used in python libraries for $\mathrm{SO}(3)$-equivariant neural networks (Geiger & Smidt, 2022; Unke & Maennel, 2024). We denote Y_{lm} the real spherical harmonic polynomials and $C^{l_3m_3}_{l_1m_1,l_2m_2}$ the Clebsch-Gordan coefficients in this basis.

2.1. Clebsch-Gordan Tensor Products

The CGTP allows to couple equivariant features of irreps type (l_1, l_2) to produce new equivariant of irrep type l_3 provided the triplet (l_1, l_2, l_3) satisfy the triangular condition

$$|l_1 - l_2| \leq l_3 \leq l_1 + l_2 \,. \tag{1}$$

For feature vectors h^{l_1}, h^{l_2} it is defined using the Clebsch-Gordan coefficients as

$$(h^{l_1} \otimes h^{l_2})^{l_3m_3} := \sum_{m_1,m_2} C^{l_3m_3}_{l_1m_1,l_2m_2} h^{l_1m_1} h^{l_2m_2} \,. \tag{2}$$

The Clebsch-Gordan coefficients satisfy the symmetry relation (Varshalovich et al., 1988)

$$C^{l_3m_3}_{l_1m_1,l_2m_2} = (-1)^{l_1+l_2-l_3} C^{l_3m_3}_{l_2m_2,l_1m_1} \,, \tag{3}$$

from which it appears that the CGTP is anti-symmetric whenever the triplet $l_1 + l_2 - l_3$ is odd and symmetric whenever it is even.

2.2. Gaunt Tensor Products

Luo et al. (2024) proposed GTPs as fast alternative to CGTP. We discuss the regime in which GTPs can be beneficial in Section 4. GTPs rely on the following formula relating Gaunt coefficients, defined as a triple integral of spherical harmonic polynomials, to the Clebsch-Gordan coefficients

$$\begin{aligned} G^{l_3m_3}_{l_1m_1,l_2m_2} :&= \int_{\mathbb{S}^2} Y_{l_1m_1}(\hat{r}) Y_{l_2m_2}(\hat{r}) Y_{l_3m_3}(\hat{r}) \mathrm{d}\mu_{\mathbb{S}^2}(\hat{r}) \\ &= \tilde{G}^{l_3}_{l_1,l_2} C^{l_3m_3}_{l_1m_1,l_2m_2} \,. \end{aligned} \tag{4}$$

Here $\hat{r}$ is the unit vector of spherical coordinates $(\theta, \phi) \in [0, \pi] \times [0, 2\pi)$ and $\mathrm{d}\mu_{\mathbb{S}^2}(\hat{r}) = \sin(\theta)\mathrm{d}\theta\mathrm{d}\phi$ is the uniform measure on the sphere $\mathbb{S}^2$. To simplify notations, in the following we sometimes omit the $\hat{r}$ argument in integrals.

GTPs are defined by replacing the Clebsch-Gordan coefficients by the Gaunt coefficients

$$(h^{l_1} \otimes_{\mathrm{Gaunt}} h^{l_2})^{l_3m_3} = \sum_{m_1,m_2} G^{l_3m_3}_{l_1m_1,l_2m_2} h^{l_1m_1} h^{l_2m_2} \,, \tag{5}$$

such that

$$(h^{l_1} \otimes_{\mathrm{Gaunt}} h^{l_2})^{l_3m_3} = \tilde{G}^{l_3}_{l_1,l_2} (h^{l_1} \otimes h^{l_2})^{l_3m_3} \,. \tag{6}$$

One can rely on the integral formula in Equation 4 to rewrite the GTP as

$$\begin{aligned} &\left(h^{l_1} \otimes_{\mathrm{Gaunt}} h^{l_2}\right)^{l_3m_3} \\ &= \sum_{m_1,m_2} h^{l_1m_1} h^{l_2m_2} \int_{\mathbb{S}^2} Y_{l_1m_1} Y_{l_2m_2} Y_{l_3m_3} \mathrm{d}\mu_{\mathbb{S}^2} \,. \end{aligned} \tag{7}$$

This expression can be further factorised by introducing the signal on the sphere associated with the features h^{l_1}, h^{l_2}. In fact, an equivariant feature h^l yields a signal of the sphere of the form

$$\begin{aligned} F_{h^l} : \mathbb{S}^2 &\to \mathbb{R} \\ \hat{r} &\mapsto \langle h^l, Y_l(\hat{r}) \rangle \end{aligned} \tag{8}$$

where $\langle h^l, Y_l(\hat{r}) \rangle := \sum_{m=-l}^{m} h^{lm} Y_{lm}(\hat{r})$. In this notation, the GTP can be rewritten more compactly as

$$(h^{l_1} \otimes_{\mathrm{Gaunt}} h^{l_2})^{l_3m_3} = \int_{\mathbb{S}^2} F_{h^{l_1}} F_{h^{l_2}} Y_{l_3m_3} \mathrm{d}\mu_{\mathbb{S}^2} \,. \tag{9}$$

This integral form of the GTP allows to factorize sums of tensor products corresponding to different irreps paths $(l_1, l_2) \to l_3$, which can lead to a significant speed-up. We discuss the benefits and limitations of this approach in Section 4.

Luo et al. (2024) originally proposed to use a method based on Fast Fourier Transform to estimate the triple integral

defining the GTP. Instead, Xie et al. (2025) proposed to use a spherical t-cubature (McLaren, 1963; Hardin & Sloane, 1996; Womersley, 2018; Price & McEwen, 2024), namely a finite set of points $\{\hat{r}_i\}_{1\leq i\leq N} \in \mathbb{S}^2$ on the sphere with associated weights $\{w_i\}_{1\leq i\leq N} \in \mathbb{R}$ such that for any polynomial $P : \mathbb{S}^2 \mapsto \mathbb{C}$ of degree t,

$$\sum_{i=1}^{N} w_i P(\hat{r}_i) = \int_{\mathbb{S}^2} P(\hat{r}) \mathrm{d}\mu_{\mathbb{S}^2}(\hat{r}) \,. \tag{10}$$

Remarking that the spherical harmonics Y_{lm} are uniform polynomials of degree l, it suffices to use a spherical $(l_1 + l_2 + l_3)$-cubature to evaluate the triple integral. Given that a spherical L-cubature contains roughly $\mathcal{O}(L^2)$ points (McLaren, 1963), Xie et al. (2025) estimate the runtime of the GTP with this evaluation method to be $\mathcal{O}(L^2 \log(L))$ for a single triplet of irreps $l_1 = l_2 = l_3 = L$. The authors also propose an evaluation method based on the fast spherical transform algorithm of Healy Jr et al. (2003) with a similar runtime for a single triplet but yielding an improved asymptotic runtime in the case of multiple input and output irreps which we discuss in Section 4.

From the definition of the Gaunt coefficients and the symmetry property of the Clebsch-Gordan coefficient of Eq. (3), one clearly sees that $\tilde{G}^{l_3}_{l_1,l_2} = 0$ whenever $l_1 + l_2 - l_3$ is odd. Consequently, the GTP fails to reproduce the CGTP for anti-symmetric cases.

2.3. Vector Signal Tensor Products

Xie et al. (2026) recently proposed a generalization of the GTP approach encompassing both the symmetric and the anti-symmetric cases of the CGTP. They suggest to replace the signals on the sphere used in GTP by irreducible tensor-valued signals. To this end they propose to use the so-called Tensor Spherical Harmonics (TSH)

$$Y^{ls}_{jm_j} : \mathbb{S}^2 \to \mathbb{R}^{2s+1} \tag{11}$$

which entries are defined as

$$\left(Y^{ls}_{jm_j}\right)_{m_s} := \sum_{m_l} C^{j,m_j}_{l,m_l,s,m_s} Y_{lm_l} \,. \tag{12}$$

TSH are indexed by the integers (j, l, s, m) where the Clebsch-Gordan selection rules impose $|j - s| \leq l \leq j + s$ and $-j \leq m_j \leq j$. These tensors are used in physics, where they correspond to a coupling of a spin angular momentum s with an orbital momentum l to obtain a total angular momentum j (Varshalovich et al., 1988). When $s = 1$, TSH are refered to as Vector Spherical Harmonics (VSH). Note that for a given value of j there are only three VSH corresponding to the only values of l allowed by the selection rule, namely $j - 1, j$ and $j + 1$. VSH can be decomposed in terms involving the gradients of the associated spherical harmonics, which we leverage to demonstrate our main results presented in Section 3, see Appendix A.

The authors suggest to replace the usual equivariant feature h^l by tensor features $h^{(j,l)}$, which contraction with THS yields irreps-valued signal on the sphere. They then define a generalization of GTP for these irreps-valued tensor, which they call Irrep Signal Tensor Product (ISTP). The authors then derive a formula to expand the tensor product of TSH generalizing existing formula for $s = 1$. This allows them to obtain a generalization of the integral formula of GTP to ISTP, where point-wise product of the signals are replaced by tensor products of the corresponding irrep-valued signals. They further show that to obtain all possible tensor products interactions, it suffices to consider ISTP restricted to VSH, which they call Vector Signal Tensor Product (VSTP). Crucially, they prove that VSTP are enough to simulate all the CGTP corresponding to different values of j for the inputs and outputs. In particular, they show that for a CGTP corresponding to an irrep path $(j_1, j_2) \to j_3$, at least one of the $3 \times 3 = 9$ interactions between the input's VSH yields non-zero output of type j_3.

The VSTP introduced in Ref.(Xie et al., 2026) inherits the same efficient asymptotic runtime scaling in L as the GTP and it encompasses both the symmetric and anti-symmetric cases of the CGTP. However, practical applications of VSTPs are hindered by their cumbersome implementation. In fact, VSTPs replace standard spherical-harmonics-based features with VSH-derived equivariant features, and they necessitate computing nine cross products between the irreps-valued signals on the sphere associated with each input. This large number of operations might also harm the potential speed-up of VSTP in the low L regime, which is the one mostly used in current architecture.

2.4. Other related works

Other methods have been proposed to accelerate tensor product in a variety of specific setups. In particular, Passaro & Zitnick (2023) proposed a method to accelerate tensor products between generic equivariant features and a vector of spherical harmonic features, relying on a well-chosen rotation of the feature vectors. Li et al. (2025) considered a similar setup in the context of equivariant graph neural networks. Their proposed method replaces edge-based tensor products with node-based operations by relying on a binomial expansion of the spherical harmonic feature vector. Unlike for the GTP and VSTP, these methods cannot by directly be applied to tensor products between generic feature vectors, and we do not discuss them in this work.

Shortly after the initial submission of this work, Bochkarev et al. (2026) independently proposed a similar method to generalize GTP to anti-symmetric cases.

3. Integral Formula for Anti-Symmetric Tensor Products

In this section we present our main findings, namely integral formulas involving gradients of the spherical harmonics that allow us to simplify VSTP and to propose an easy-to-implement extension of GTPs to anti-symmetric cases of CGTP. Our first result is the following integral equation.

Theorem 3.1. *Let (l_1, l_2, l_3) be a triplet satisfying the anti-symmetric condition, i.e. $l_1 + l_2 - l_3$ is odd. Then there exists $\tilde{V}^{l_3}_{l_1,l_2} \neq 0$ such that*

$$\int_{\mathbb{S}^2} ((\nabla Y_{l_1 m_1} \times \nabla Y_{l_2 m_2}) \cdot \hat{r}) Y_{l_3 m_3} \mathrm{d}\mu_{\mathbb{S}^2} = \tilde{V}^{l_3}_{l_1,l_2} C^{l_3 m_3}_{l_1 m_1, l_2 m_2}, \tag{13}$$

Proof. See Section A.2 of the appendix.

This formula yields an integral expression of the CGTP in the anti-symmetric cases. Since it involves the cross product of spherical harmonics gradients, which can be expressed in terms of VSH, this expression corresponds to a specific interaction involved in the VSTP proposed by Xie et al. (2026). Given this connection, we propose redefining the VSTP as

$$\int_{\mathbb{S}^2} \left((\nabla \langle h^{l_1}, Y_{l_1} \rangle \times \nabla \langle h^{l_2}, Y_{l_2} \rangle) \cdot \hat{r} \right) Y_{l_3 m_3} \mathrm{d}\mu_{\mathbb{S}^2} := \left(h^{l_1} \otimes_{\mathrm{VSTP}} h^{l_2} \right)^{l_3 m_3} . \tag{14}$$

The previous theorem yields

$$\left(h^{l_1} \otimes_{\mathrm{VSTP}} h^{l_2} \right)^{l_3 m_3} = \tilde{V}^{l_3}_{l_1,l_2} \left(h^{l_1} \otimes h^{l_2} \right)^{l_3 m_3} . \tag{15}$$

Note that the above formula hold true for all triplets (l_1, l_2, l_3) provided one set $\tilde{V}^{l_3}_{l_1,l_2} = 0$ whenever the triplet corresponds to a symmetric CGTP.

Using the previous result, we derive the following integral formula encompassing both the symmetric and the anti-symmetric cases.

Theorem 3.2. *For all triplets (l_1, l_2, l_3), there exists $\Gamma^{l_3}_{l_1,l_2} \neq 0$ such that*

$$\left(h^{l_1} \otimes h^{l_2} \right)^{l_3 m_3} = \Gamma^{l_3}_{l_1,l_2} \int_{\mathbb{S}^2} \left(V^{l_1} \cdot W^{l_2} \right) Y_{l_3 m_3} \mathrm{d}\mu_{\mathbb{S}^2} , \tag{16}$$

with

$$\begin{cases} V^{l_1}(\hat{r}) := \langle h^{l_1}, Y_{l_1} \rangle \hat{r} + \hat{r} \times \nabla \langle h^{l_1}, Y_{l_1} \rangle \\ W^{l_2}(\hat{r}) := \langle h^{l_2}, Y_{l_2} \rangle \hat{r} + \nabla \langle h^{l_2}, Y_{l_2} \rangle . \end{cases} \tag{17}$$

Proof. Using the identity of the mixed product under cyclic permutation

$$(\vec{a} \times \vec{b}) \cdot \vec{c} = (\vec{c} \times \vec{a}) \cdot \vec{b} \tag{18}$$

one can rewrite the VSTP of h^{l_1}, h^{l_2} as

$$\int_{\mathbb{S}^2} \left((\hat{r} \times \nabla \langle h^{l_1}, Y_{l_1} \rangle) \cdot (\nabla \langle h^{l_2}, Y_{l_2} \rangle) \right) Y_{l_3 m_3} \mathrm{d}\mu_{\mathbb{S}^2} . \tag{19}$$

To encompass both the symmetric and the anti-symmetric case in a single formula, one can simply sum the GTP and VSTP, which gives

$$\left(h^{l_1} \otimes_{\mathrm{Gaunt}} h^{l_2} \right)^{l_3 m_3} + \left(h^{l_1} \otimes_{\mathrm{VSTP}} h^{l_2} \right)^{l_3 m_3} = \left(\tilde{G}^{l_3}_{l_1,l_2} + \tilde{V}^{l_3}_{l_1,l_2} \right) \left(h^{l_1} \otimes h^{l_2} \right)^{l_3 m_3} . \tag{20}$$

As the coefficients $\tilde{G}^{l_3}_{l_1,l_2}$ and $\tilde{V}^{l_3}_{l_1,l_2}$ are non-zero on distinct triplets, their sum is always non-zero so we can define

$$\Gamma^{l_3}_{l_1,l_2} := \left(\tilde{G}^{l_3}_{l_1,l_2} + \tilde{V}^{l_3}_{l_1,l_2} \right)^{-1} , \tag{21}$$

and we have $\Gamma^{l_3}_{l_1,l_2} \neq 0$ for all (l_1, l_2, l_3). Using the facts that

$$\begin{aligned} (\hat{r} Y_{l_1 m_1}) \cdot (\nabla Y_{l_2 m_2}) &= 0 , \\ (\hat{r} Y_{l_1 m_1}) \cdot (\hat{r} \times \nabla Y_{l_2 m_2}) &= 0 , \\ (\hat{r} Y_{l_1 m_1}) \cdot (\hat{r} Y_{l_2 m_2}) &= Y_{l_1 m_1} Y_{l_2 m_2} \end{aligned} \tag{22}$$

and expanding the terms in the theorem's integral, we recognize the integral expressions of the GTP (Eq. 9) and the VSTP (Eq. 19), which proves the result. □

These results simplify the VSTP proposed by Xie et al. (2026) by reducing the number of VSTP interactions to consider when simulating a given CGTP from up to 9 possibilities to a single one. Moreover it allows for a clear and simple implementation involving only standard irrep features $h^l \in \mathbb{R}^{2l+1}$ instead of the tensor-valued features used in the VSTP proposed by Xie et al. (2026). In fact, one can straightforwardly adapt the spherical cubature implementation introduced in Ref. (Xie et al., 2025) to the previous formulas.

Our result provide an intuitive interpretation of the VSTP in terms of signals on the sphere, which is represented in Figure 1. The GTP let the features interact by multiplying their corresponding signals pointwise on the sphere. As this operation is symmetric, so are the GTP. Instead, the VSTP let signals interact through their Poisson bracket (Lee, 2003) given by the cross product of the gradients projected along the radial vector. Similarly to the GTP, the VSTP inherits the anti-symmetry of the cross product and it allows to recover the anti-symmetric part of the CGTP.

4. Expressivity-Runtime Tradeoff of Integral-Based Tensor Products

In this section we discuss the expressivity-runtime tradeoff of GTP and VSTP, building on the insights of Refs. (Xie

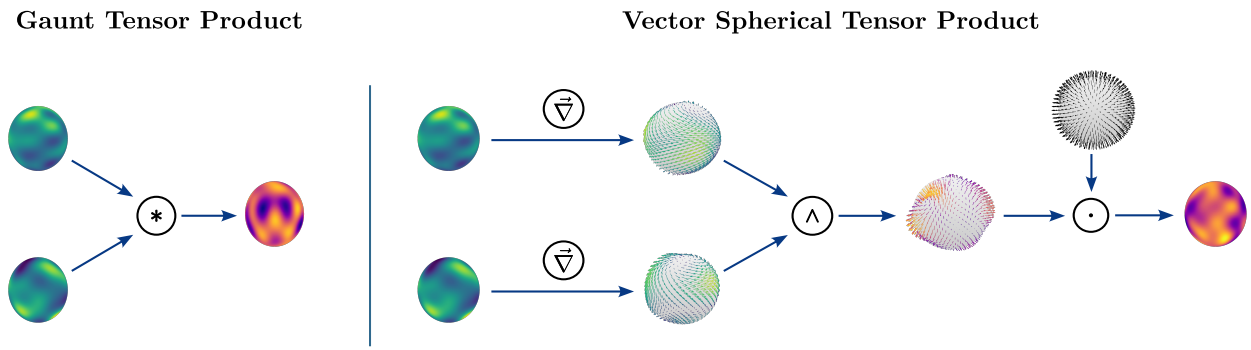

Figure 1. Sketch of the Gaunt Tensor Product and our simplified Vector Signal Tensor Products. The inputs irrep features are first transformed into scalar signals on the sphere using spherical harmonics. Gaunt Tensor Products operates the pointwise product of the signals and the resulting signal is turned back into an irrep features by projecting on spherical harmonics. To obtain the anti-symmetric counter part, VSTP takes the cross product of the signals gradients and project the result pointwise on the radial vectors. The resulting scalar signal is then projected back to irrep features as for Gaunt Tensor Products.

et al., 2025; 2026). We focus on GTP to simplify the presentation, although the discussion holds true for all similar integral-based tensor products, including VSTP and our approach.

In the context of $\mathrm{SO}(3)$-equivariant neural networks, tensor products are used to build equivariant bilinear layers with learnable parameters. Xie et al. (2025) rely on this observation to define a measure of expressivity for the layers built on different types of tensor products. They then compare this measure of expressivity to the corresponding runtime. The same authors extend their analysis to VSTP and detail their findings depending on the number of irreps in the input and output in Ref. (Xie et al., 2026).

Here we only discuss the case of tensor products between input features with multiple irrep types, as it is the one for which the integral formulations of tensor products find their relevance. This setup is refereed to as a multiple-input-multiple-output (MIMO) setup in Refs. (Xie et al., 2025; 2026).

Suppose one wants to build an equivariant bilinear layer acting on inputs

$$h^1 = \bigoplus_{l_1=0}^{L_1} h^{l_1}\,, \quad h^2 = \bigoplus_{l_2=0}^{L_2} h^{l_2} \tag{23}$$

and producing an output feature

$$h^3 = \bigoplus_{l_3=0}^{L_3} h^{l_3}\,. \tag{24}$$

Due to the universal property of the CGTP, the most general form of such a layer is given by

$$h^3 = \bigoplus_{l_3=0}^{L_3} \left(\sum_{l_1=0}^{L_1} \sum_{l_2=0}^{L_2} w^{l_3}_{l_1,l_2} (h^{l_1} \otimes h^{l_2})^{l_3} \right)\,, \tag{25}$$

where $w^{l_3}_{l_1,l_2}$ are learnable weights parameterizing the layer. To simplify the discussion, we suppose that $L_1 = L_2 = L_3 = L$. As discussed in Refs. (Xie et al., 2025; 2026), the runtime associated with this operation scales as $\mathcal{O}(L^5)$ when leveraging the sparsity of the Clebsch-Gordan coefficients.

An important aspect of the previous layer is that the CGTP allows to have different weights $w^{l_3}_{l_1,l_2}$ for each possible irreps path $(l_1, l_2) \to l_3$. Integral representations of the tensor product allows to reduce the complexity of the previous layer upon assuming the learnable weights can be factorized, i.e.

$$w^{l_3}_{l_1,l_2} = w^{l_3} w_{l_1} w_{l_2}\,. \tag{26}$$

In this case, replacing the CGTP by a GTP and using the corresponding integral formula yields

$$\begin{aligned} h^{l_3} &= \sum_{l_1}^{L} \sum_{l_2}^{L} w^{l_3}_{l_1,l_2} (h^{l_1} \otimes_{\text{Gaunt}} h^{l_2})^{l_3} \\ &= \sum_{l_1=0}^{L} \sum_{l_2=0}^{L} w^{l_3} w_{l_1} w_{l_2} \int_{\mathbb{S}^2} F_{h^{l_1}} F_{h^{l_2}} Y_{l_3} \mathrm{d}\mu_{\mathbb{S}^2} \\ &= w^{l_3} \int_{\mathbb{S}^2} \left(\sum_{l_1=0}^{L} w_{l_1} F_{h^{l_1}} \right) \left(\sum_{l_2=0}^{L} w_{l_2} F_{h^{l_2}} \right) Y_{l_3} \mathrm{d}\mu_{\mathbb{S}^2}\,. \end{aligned} \tag{27}$$

This formula has the advantage that the signals corresponding to the different irreps order of the different input features are factorized and can be summed before the product and the integration are performed. As shown by Xie et al. (2025), this significantly reduce the cost of the corresponding layer to $\mathcal{O}(L^3)$ for a simple spherical design evaluation and to $\mathcal{O}(L^2 \log^2(L))$ for the S2FFT evaluation proposed in Ref. (Xie et al., 2025). Note that the authors mention that their S2FFT evaluation method is more efficient only asymptotically in L and that it doesn't result in a practical

speed-up below $L \sim 1000$, which is far beyond the values of L used in current architectures.

The quadratic asymptotic speed-up in L of GTP relies on the weights factorization, which has a strong impact on the expressivity of the corresponding layer. Xie et al. (2025) define the expressivity of the layers above as the dimension of the manifold of the associated bilinear maps. In the case of CGTP, the weights do not need to be factorized so this dimension is of the order of the number of path allowed by the Clebsch-Gordan selection rules, which scales as $\mathcal{O}(L^3)$. On the other hand, the GTP advantageous runtime assumes the factorization of the weights, which yields a corresponding dimension scaling in $\mathcal{O}(L)$. Consequently, Xie et al. (2025) conclude that the CGTP and the GTP layers share the same $\mathcal{O}(L^2)$ asymptotic runtime-expressivity ratio.

Although the analysis of Xie et al. (2025) offers a clear insight on the asymptotic expressivity-runtime tradeoff, we argue that the GTP can still prove useful when the learnable weight $w^{l_3}_{l_1,l_2}$ admits a rank R decomposition (Kolda & Bader, 2009)

$$w^{l_3}_{l_1,l_2} = \sum_{k=1}^{R} w^{(k)l_3} w^{(k)}_{l_1} w^{(k)}_{l_2} \,. \tag{28}$$

For such weights the GTP layer yields a runtime scaling as $\mathcal{O}(RL^3)$, whereas the corresponding CGTP layer does not leverage the weights decomposition and yields the same $\mathcal{O}(L^5)$ runtime. Thus it appears that the expressivity measure presented by Xie et al. (2025) is a worst-case scenario measure which assumes the learnable weights have full rank $R = \mathcal{O}(L^2)$.

This analysis shows that integral-based tensor products as the ones considered in this paper allow to control the runtime-expressivity tradeoff by enabling shorter runtime for bilinear layers which parameters $w^{l_3}_{l_1,l_2}$ admit a low rank decomposition. It would be interesting to investigate this tradeoff numerically in future work.

5. Normalization

When the antisymmetric GTP in Equation (13), or the full tensor product in Equation (16), are used as learnable layers in an equivariant neural network, it is important that the coupling coefficients are properly normalized. One typically wishes the magnitude of the resulting features to be comparable to those produced by the standard CGTP, so that layer initialization remains well-conditioned and does not introduce artificial scale differences across angular momentum channels.

While the Clebsch-Gordan tensor $C^{l_3 m_3}_{l_1 m_1\, l_2 m_2}$ is orthonormal by construction, the symmetric and antisymmetric coefficients $T_{l_1,l_2,l_3} \in \{\tilde{G}_{l_1,l_2,l_3}, \tilde{V}_{l_1,l_2,l_3}\}$ introduce an angular-momentum-dependent scale. Consequently, if the tensor product is used directly in a neural network layer, different (l_1, l_2, l_3) channels may produce outputs with substantially different magnitudes.

A natural strategy is therefore to explicitly normalize the tensor product by multiplying each channel by the inverse factor $T^{-1}_{l_1 l_2 l_3}$. However, in an efficient implementation of GTP (Luo et al., 2024; Xie et al., 2025), as described in Section 4, the spherical integral is evaluated using a factorized representation that separates the dependence on l_1, l_2, and l_3. Introducing a fully dense normalization tensor would destroy this factorization and therefore remove the computational advantage.

To preserve the factorized structure, in Appendix A.4, we instead seek approximate low-rank decompositions (Kolda & Bader, 2009) of the inverse coupling tensors of the form

$$(T_{l_1 l_2 l_3})^{-1} \approx \sum_{r=1}^{R} a_{l_1,r}\, b_{l_2,r}\, c_{l_3,r}, \tag{29}$$

where the vectors $a_{l_1,r}$, $b_{l_2,r}$ and $c_{l_3,r}$ depend only on a single angular momentum index. Substituting this form into the tensor product preserves the separability of the computation and allows the normalization to be implemented with negligible additional cost (a factor of R, which can be evaluated in parallel). As shown in Appendix A.4, we find empirically that $(T^{l_3}_{l_1,l_2})^{-1}$ is intrinsically low rank, at least if we only care about order of magnitude accuracy . In particular, a rank-2 decomposition of $\tilde{V}^{-1}_{l_1 l_2 l_3}$ reproduces normalization of the full tensor over a wide range of angular momenta that were tested (up to $L_{max} < 20$). By contrast, a rank-1 approximation of $\tilde{V}$ fails qualitatively and cannot reproduce the structure of the tensor. Conversely, we find that $\tilde{G}^{-1}_{l_1,l_2,l_3}$ is well approximated by a rank-1 tensor. The difference between the approximate ranks of the antisymmetric and symmetric Gaunt coefficients likely arises from the two distinct coupling pathways $l_3 \pm 1$ appearing in Equation (42), which introduces two dominant modes in the coupling tensor.

In practice, this low-rank representation provides a way to initialize Gaunt tensor-product layers with a scale comparable to CGTP while preserving the factorization structure that enables efficient implementations. For example, using Equation (28), a simple way to initialize a neural network with the same scale as the Clebsch-Gordan coefficients would be to choose the initialization

$$w^{l_3}_{l_1,l_2} = \sum_{k=1}^{R} a_{k,l_1} b_{k,l_2} c_{k,l_3} w^{(k)l_3} w^{(k)}_{l_1} w^{(k)}_{l_2} \,. \tag{30}$$

where $a_{k,l_1} b_{k,l_2} c_{k,l_3}$ are the low rank matrices found in Equation (29) and $w^{(k)l_3} w^{(k)}_{l_1} w^{(k)}_{l_2}$ are initialized around the identity.

6. Conclusion

In this work, we have arrived at an integral representation (Equation (13)) that generalizes the Gaunt coefficient formula for Clebsch-Gordan coefficients to the antisymmetric case. From this result we obtained Equation (16), a single universal integral expression that encompasses both the antisymmetric and symmetric components of the Clebsch-Gordan coefficients. Such integral formulas have well known and efficient implementation schemes (Luo et al., 2024; Xie et al., 2025) that can be directly applied to this work. Our formulation further shows that a single VSTP is enough to simulate the CGTP (Xie et al., 2026). Crucially, the integral formula uses only standard irrep features h^l, instead of the more general tensor-valued features, which ultimately reduces the number of VSTPs required by a factor of up to 9.

Beyond computational efficiency, having analytic expressions for the even and odd-parity Gaunt coefficients is particularly important for applications to equivariant neural networks. The expressions derived in Appendix A.2 allow one to directly study the magnitude and scaling behavior of the coupling coefficients across the angular momentum channels. Using the explicit form of the coefficients derived in this work, we showed in Appendix A.4 that the resulting normalization tensors admit accurate low-rank approximations. This enables efficient normalization schemes that preserve the factorized structure of the tensor product and therefore maintain the computation advantages of the VSTP construction.

Together, these results provide a practical recipe for incorporating integral-based tensor products into $\mathrm{SO}(3)$-equivariant neural networks and for controlling the speed-expressivity tradeoff in equivariant architectures. A natural direction for future work is to evaluate the proposed implementation on concrete learning tasks, for example within machine-learning interatomic potential (MLIP) models (Batatia et al., 2022; Wood et al., 2026; Brunken et al., 2025), where efficient and well-conditioned equivariant tensor products are critical for scalability and performance.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

Acknowledgements

We thank the anonymous reviewers for their careful reading and constructive comments, which helped improve the clarity and presentation of this manuscript.

References

Batatia, I., Kovacs, D. P., Simm, G., Ortner, C., and Csanyi, G. Mace: Higher order equivariant message passing neural networks for fast and accurate force fields. In Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., and Oh, A. (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 11423–11436. Curran Associates, Inc., 2022.

Bochkarev, A., Lysogorskiy, Y., and Drautz, R. Fast contracted clebsch–gordan tensor products for equivariant graph neural networks, 2026.

Brunken, C., Peltre, O., Chomet, H., Walewski, L., McAuliffe, M., Heyraud, V., Attias, S., Maarand, M., Khanfir, Y., Toledo, E., Falcioni, F., Bluntzer, M., Acosta-Gutiérrez, S., and Tilly, J. Machine learning interatomic potentials: library for efficient training, model development and simulation of molecular systems, 2025.

Geiger, M. and Smidt, T. e3nn: Euclidean neural networks, 2022.

Hardin, R. H. and Sloane, N. J. Mclaren's improved snub cube and other new spherical designs in three dimensions. *Discrete & Computational Geometry*, 15(4):429–441, 1996.

Healy Jr, D. M., Rockmore, D. N., Kostelec, P. J., and Moore, S. Ffts for the 2-sphere-improvements and variations. *Journal of Fourier analysis and applications*, 9(4): 341–385, 2003.

Jackson, J. D. *Electrodynamics*. John Wiley and Sons, Ltd, 2007.

Kolda, T. G. and Bader, B. W. Tensor decompositions and applications. *SIAM review*, 51(3):455–500, 2009.

Lee, J. M. Smooth manifolds. In *Introduction to smooth manifolds*, pp. 578–579. Springer, 2003.

Li, Y., Huang, L., Ding, Z., Wei, X., Wang, C., Yang, H., Wang, Z., Liu, C., Shi, Y., Jin, P., Qin, T., Gerstein, M., and Zhang, J. E2former: An efficient and equivariant transformer with linear-scaling tensor products. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Luo, S., Chen, T., and Krishnapriyan, A. S. Enabling efficient equivariant operations in the fourier basis via gaunt tensor products. In *The Twelfth International Conference on Learning Representations*, 2024.

McLaren, A. D. Optimal numerical integration on a sphere. *Mathematics of Computation*, 17(84):361–383, 1963.

Passaro, S. and Zitnick, C. L. Reducing SO(3) convolutions to SO(2) for efficient equivariant GNNs. In Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., and Scarlett, J. (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 27420–27438. PMLR, 23–29 Jul 2023.

Price, M. A. and McEwen, J. D. Differentiable and accelerated spherical harmonic and wigner transforms. *Journal of Computational Physics*, 510:113109, 2024.

Unke, O. T. and Maennel, H. E3x: $\mathrm{E}(3)$-equivariant deep learning made easy, 2024.

Varshalovich, D. A., Moskalev, A. N., and Khersonskii, V. K. *Quantum Theory of Angular Momentum*. WORLD SCIENTIFIC, October 1988.

Womersley, R. S. *Efficient Spherical Designs with Good Geometric Properties*, pp. 1243–1285. Springer International Publishing, Cham, 2018.

Wood, B. M., Dzamba, M., Fu, X., Gao, M., Shuaibi, M., Barroso-Luque, L., Abdelmaqsoud, K., Gharakhanyan, V., Kitchin, J. R., Levine, D. S., Michel, K., Sriram, A., Cohen, T., Das, A., Rizvi, A., Sahoo, S. J., Ulissi, Z. W., and Zitnick, C. L. Uma: A family of universal models for atoms, 2026.

Xie, Y., Daigavane, A., Kotak, M., and Smidt, T. The price of freedom: Exploring expressivity and runtime tradeoffs in equivariant tensor products. In *Forty-second International Conference on Machine Learning*, 2025.

Xie, Y., Daigavane, A., Kotak, M., and Smidt, T. Asymptotically fast clebsch-gordan tensor products with vector spherical harmonics, 2026.

A. Appendix

Unless otherwise stated, in this appendix we use complex spherical harmonics which we denote Y^l_m, and we write $C^{l_3 m_3}_{l_1 m 1; l_2 m_2}$ the Clebsch-Gordan coefficients in this basis. The spin components of the vector spherical harmonics are defined in the spherical coordinates basis. We follow the notations of Ref. (Varshalovich et al., 1988) and denote respectively $\cdot$ the Euclidian dot product and $\times$ the cross product in the Cartesian basis. We show explicitly how to transform our results to the real spherical harmonics basis used in the main text in Section A.3.

A.1. Vector Harmonics

In this Section, we introduce the relevant background on vector spherical harmonics used in the appendix. For a more detailed overview coverage of vector harmonics we recommend (Varshalovich et al., 1988).

As in (Varshalovich et al., 1988; Xie et al., 2026), we define the tensor (spin-s) spherical harmonics by coupling orbital angular momentum l with spin s

$$\left(Y^{l,s}_{jm}(\hat{r})\right)_{m_s} = \sum_{m_l} C^{jm}_{lm_l,\, sm_s}\, Y^l_{m_l}(\hat{r})\,. \tag{31}$$

The vector harmonics arise from the $s = 1$ components, and for notational convenience, we will write the $s = 1$ vector harmonic with components m_s as $(\vec{Y}_{j,l,m})_{m_s} := (Y^{j1}_{lm})_{m_s}$. Note that for the vector harmonics, the selection rules dictate that $j \in \{l-1, l, l+1\}, j \geq 0$. The tensor harmonics obey the orthonormality condition

$$\int_{\mathbb{S}^2} Y^{l,s}_{jm}(\hat{r}) \cdot Y^{l',s'}_{j'm'}(\hat{r})^*\, d\mu_{\mathbb{S}^2}(\hat{r}) = \int_{\mathbb{S}^2} \sum_{m_s} \left(Y^{l,s}_{jm}(\hat{r})\right)_{m_s} \left(Y^{l',s'}_{j'm'}(\hat{r})\right)^*_{m_s} d\mu_{\mathbb{S}^2}(\hat{r}) = \delta_{ll'}\, \delta_{ss'}\, \delta_{jj'}\, \delta_{mm'}. \tag{32}$$

The conjugation of the harmonics with the dot product arises because we use complex spherical coordinates, and so the correct inner product is the Hermitian inner product (Varshalovich et al., 1988).

For our purposes, it is useful to relate the vector harmonics to the spherical harmonics through application of standard differential operators. In particular, the $j = l$ component of the vector harmonics is known to arise from application of the angular momentum operator $\hat{L} = -i(\hat{r} \times \nabla)$ to the spherical harmonics (Varshalovich et al., 1988)

$$\hat{L} Y^l_m = \sqrt{l(l+1)}\, \vec{Y}_{l,l,m}. \tag{33}$$

Similarly, the $j = l \pm 1$ components of the vector harmonics have known relations in terms of the gradient of the spherical harmonics (Jackson, 2007; Varshalovich et al., 1988)

$$\begin{aligned} \vec{Y}_{l-1,l,m} &= -\sqrt{\frac{2l+1}{l}}\, [l\, \hat{r} - r\, \nabla]\, Y^l_m, \qquad l \neq 0\,, \\ \vec{Y}_{l+1,l,m} &= -\sqrt{\frac{2l+1}{l+1}}\, [(l+1)\, \hat{r} - r\, \nabla]\, Y^l_m\,, \end{aligned} \tag{34}$$

where $r = ||\hat{r}||$.

A.2. Proof of Theorem 3.1

In this Section, we prove our main result in Equation (13). We first show that for complex spherical harmonics

$$\int_{\mathbb{S}^2} (\nabla Y^{l_1}_{m_1} \times \nabla Y^{l_2}_{m_2}) \cdot (\hat{r} Y^{l_3}_{m_3})^* \mathrm{d}\mu_{\mathbb{S}^2}(\hat{r}) = \Lambda_{l_1 l_2 l_3} C^{l_3 m_3}_{l_1 m_1, l_2 m_2}, \tag{35}$$

where $\Lambda_{l_1,l_2,l_3} \neq 0$ is non-zero whenever a l_1, l_2, l_3 path is allowed by the Clebsch-Gordan coefficients and $l_1 + l_2 - l_3 = 2k + 1$. The proof follows from a number of useful tensor identities that can be found in (Varshalovich et al., 1988; Xie et al., 2026). We derive Equation (35) using the complex harmonics in a spherical coordinate system, and hence the Clebsch-Gordan coefficients should be understood in this basis according to the Condon–Shortley convention. In Section A.3, we transform this equation to real harmonics in Cartesian coordinates and relate it to the rotated Clebsch-Gordan coefficients.

We first relate Equation (35) to the vector spherical harmonics using the identities of Section A.1. For $\vec{a}, \vec{b} \perp \hat{r}$, we have that

$$(\hat{r} \times \vec{a}) \times (\hat{r} \times \vec{b}) = ((\vec{a} \times \vec{b}).\hat{r})\hat{r}. \tag{36}$$

Hence, for $||\hat{r}|| = 1$ we get

$$\begin{aligned}(\nabla Y^{l_1}_{m_1} \times \nabla Y^{l_2}_{m_2}) \cdot \hat{r} &= (\hat{r} \times \nabla Y^{l_1}_{m_1}) \times (\hat{r} \times \nabla Y^{l_2}_{m_2}) \cdot \hat{r} \\ &= -(\hat{L} Y^{l_1}_{m_1}) \times (\hat{L} Y^{l_2}_{m_2}) \cdot \hat{r} \\ &= -\sqrt{l_1 l_2 (l_1+1)(l_2+1)} \left(\vec{Y}_{l_1,l_1,m} \times \vec{Y}_{l_2,l_2,m_2} \right) \cdot \hat{r},\end{aligned} \tag{37}$$

where we use Equation (33) between the second and the third line. It therefore remains to evaluate the integral

$$\int_{\mathbb{S}^2} \left(\vec{Y}_{l_1,l_2,m} \times \vec{Y}_{l_2,l_2,m_2} \right) \cdot \left(\hat{r} Y^{l_3}_{m_3} \right)^* \mathrm{d}\mu_{\mathbb{S}^2}(\hat{r}). \tag{38}$$

To proceed, we can use the identity introduced in Equation 101, page 222 of (Varshalovich et al., 1988)

$$\left(\vec{Y}_{l_1,l_1,m_1} \times \vec{Y}_{l_2,l_2,m_2} \right) = i\sqrt{\frac{3}{2\pi}}(2l_1+1)(2l_2+1) \left[\sum_{j,l} \begin{Bmatrix} l_1 & l_1 & 1 \\ l_2 & l_2 & 1 \\ j & l & 1 \end{Bmatrix} C^{l0}_{l_1 0, l_2 0} C^{jm}_{l_1 m_1, l_2 m_2} \vec{Y}_{j,l,m} \right] \tag{39}$$

where the curly bracket terms $\{\}$ are the Wigner-9j symbols. To find the integral, we will use orthogonality of the vector harmonics, and so we first find it useful to re-write $\hat{r} Y^{l_3}_{m_3}(\hat{r})$ in terms of the irreducible vector harmonics. From Equation 70, page 219 of (Varshalovich et al., 1988), we have that

$$\hat{r}\, Y^{l_3}_{m_3}(\hat{r}) = \sqrt{\frac{l_3}{2l_3+1}}\, \vec{Y}_{l_3,l_3-1,m_3}(\hat{r}) - \sqrt{\frac{l_3+1}{2l_3+1}}\, \vec{Y}_{l_3,l_3+1,m_3}(\hat{r})\,. \tag{40}$$

Using the orthonormality of the tensor harmonics in Equation (32) we now subsitute Equation (40) into Equation (39) and integrate over the sphere to find

$$\begin{aligned}&\int_{\mathbb{S}^2} \left(\vec{Y}_{l_1,l_1,m_1} \times \vec{Y}_{l_2,l_2,m_2} \right) \cdot \left(\hat{r} Y^{\ell_3}_{m_3} \right)^* d\mu_{\mathbb{S}^2}(\hat{r}) \\ &= i\sqrt{\frac{3}{2\pi}} \frac{(2l_1+1)(2l_2+1)}{\sqrt{2l_3+1}} \left[\sqrt{l_3} C^{(l_3-1)0}_{l_1 0, l_2 0} \begin{Bmatrix} l_1 & l_1 & 1 \\ l_2 & l_2 & 1 \\ l_3 & l_3-1 & 1 \end{Bmatrix} - \sqrt{l_3+1} C^{(l_3+1)0}_{l_1 0, l_2 0} \begin{Bmatrix} l_1 & l_1 & 1 \\ l_2 & l_2 & 1 \\ l_3 & l_3+1 & 1 \end{Bmatrix} \right] C^{l_3 m_3}_{l_1 m_1, l_2 m_2},\end{aligned} \tag{41}$$

Recalling Equation (35), Equation (41) implicitly defines $\Lambda_{l_1 l_2 l_3}$ in Equation (35)

$$\begin{aligned}\Lambda_{l_1 l_2 l_3} = &- i\sqrt{\frac{3}{2\pi} \frac{l_1 l_2 (l_1+1)(l_2+1)}{(2l_3+1)}}(2l_1+1)(2l_2+1) \\ &\times \left[\sqrt{l_3} C^{(l_3-1)0}_{l_1 0, l_2 0} \begin{Bmatrix} l_1 & l_1 & 1 \\ l_2 & l_2 & 1 \\ l_3 & l_3-1 & 1 \end{Bmatrix} - \sqrt{l_3+1} C^{(l_3+1)0}_{l_1 0, l_2 0} \begin{Bmatrix} l_1 & l_1 & 1 \\ l_2 & l_2 & 1 \\ l_3 & l_3+1 & 1 \end{Bmatrix} \right].\end{aligned} \tag{42}$$

It remains to study the selection rules of $\Lambda_{l_1 l_2 l_3}$ and to show that its non-vanishing for odd parity $l_1 + l_2 - l_3 = 2k+1$.

To show this, it is sufficient to study the components with the two Wigner-9j terms. Our method will be to first show that $C^{(l_3-1)0}_{l_1 0, l_2 0}$ and $C^{(l_3+1)0}_{l_1 0, l_2 0}$ are always strictly non-vanishing for odd parity and have opposite signs. Our proof then follows by showing the Wigner-9j symbols appearing in Equation (35) always have the same sign and never vanishes simultaneously.

The explicit form of the Clebsch-Gordan coefficients for $m_1 = m_2 = m_3 = 0$ can be read off from Equation 32 page 251 of (Varshalovich et al., 1988)

$$C^{l_3 0}_{l_1 0, l_2 0} = \begin{cases} 0, & \text{if } l_1 + l_2 + l_3 = 2g+1, \\ \dfrac{(-1)^{g-l_3}\sqrt{2l_3+1}\, g!}{(g-l_1)!(g-l_2)!(g-l_3)!} \left[\dfrac{(2g-2l_1)!(2g-2l_2)!(2g-2l_3)!}{(2g+1)!} \right]^{\frac{1}{2}}, & \text{if } l_1 + l_2 + l_3 = 2g. \end{cases} \tag{43}$$

We immediately see that $\Lambda_{l_1 l_2 l_3}$ vanishes unless

$$l_1 + l_2 + l_3 \pm 1 = 2g, \tag{44}$$

which is equivalent to the parity condition, given l_1, l_2, l_3 are integers. The sign of $C^{l_3 0}_{l_1 0, l_2 0}$ is thus determined by $(-1)^{g-l_3}$. We immediately see that

$$\begin{aligned} \text{Sign}\left(C^{(l_3-1)0}_{l_1 0, l_2 0}\right) &= (-1)^{\frac{l_1+l_2-l_3+1}{2}} \\ \text{Sign}\left(C^{(l_3+1)0}_{l_1 0, l_2 0}\right) &= (-1)^{\frac{l_1+l_2-l_3-1}{2}} \\ \text{Sign}\left(\frac{C^{(l_3-1)0}_{l_1 0, l_2 0}}{C^{(l_3+1)0}_{l_1 0, l_2 0}}\right) &= (-1)^{\frac{1}{2}+\frac{1}{2}} = -1. \end{aligned} \tag{45}$$

Since the $m = 0$ Clebsch-Gordan coefficients have opposite signs, it remains to show that the Wigner-9j coefficients have the same sign and cannot be simultaneously vanishing. This fact follows from Equations for the form of Wigner $9j$ symbols of the form in Table 10.9 page 383 of (Varshalovich et al., 1988). By taking the transpose of the formula, and then swapping the first two columns, we find:

$$\begin{Bmatrix} a & a & 1 \\ b & b & 1 \\ c & c+1 & 1 \end{Bmatrix} = -2(c+1)\left[\frac{(a+b+c+2)(a+b-c)(a-b+c+1)(-a+b+c+1)(2a-1)!(2b-1)!(2c)!}{3(2a+2)!(2b+2)!(2c+3)!}\right]^{1/2}. \tag{46}$$

$$\begin{Bmatrix} a & a & 1 \\ b & b & 1 \\ c & c-1 & 1 \end{Bmatrix} = -2c\left[\frac{(a+b+c+1)(a+b-c+1)(a-b+c)(-a+b+c)(2a-1)!(2b-1)!(2c-2)!}{(2a+2)!(2b+2)!(2c+1)!}\right]^{1/2}. \tag{47}$$

It is clear that both terms have the same sign. Then, if any of the terms $(a+b-c), (a-b+c), (-a+b+c)$ is zero, $a+b+c$ must be even. By contraposition, none of these terms can vanish whenever $a+b+c$ is odd. As a consequence, assuming $a+b+c$ is odd, the Wigner-9j symbol of Eq. (47) is equal to zero if and only if $a+b-c+1 = 0$, which yields $a+b = c-1$. In this case, we get $a-b+c+1 = a-b+a+b+2 = 2(a+1) > 0$ and $-a+b+c+1 = -a+b+a+b+2 = 2(b+1) > 0$, such that the Wigner-9j symbol of Eq. (46) is non-zero. Reciprocally, assuming $a+b+c$ is odd, the Wigner-9j symbol of Eq. (46) is equal to zero if and only if $a-b+c+1 = 0$ or $-a+b+c+1 = 0$. If $a-b+c+1 = 0$ we have that $b = a+c+1$, and it comes $a+b-c+1 = a-c+1+a+c+1 = 2(a+1) > 0$. Symmetrically, if $-a+b+c+1 = 0$ we get $a = b+c+1$ such that $a+b-c+1 = 2(b+1) > 0$. As a result, the Wigner-9j symbol of Eq. (47) is non-zero whenever the one of Eq. (46) is. This concludes our proof that Equation (35) holds, where $\Lambda_{l_1 l_2 l_3}$ (as defined by Equation (42)) is non-vanishing for odd parity $l_1 + l_2 + l_3 = 2k+1$, and zero otherwise.

Equation (35) is a tensor equation that holds in the complex spherical harmonic basis. In Section A.3 we explicitly transform it to the real spherical harmonics basis. With this choice of basis it is possible to show that the rotated Clebsch-Gordan coefficients for odd parity are purely imaginary, which explains the factor of i appearing in Equation (42).

A.3. Transforming from complex to real harmonics

In this section we show explicitly how Equation (35) transforms as we change from the complex spherical basis to the real Cartesian basis, and from complex spherical harmonics to the real spherical harmonics.

Let $\{e_x, e_y, e_z\}$ denote the Cartesian basis and $\{e_{+1}, e_0, e_{-1}\}$ the complex spherical basis

$$e_{\pm 1} = \mp\frac{1}{\sqrt{2}}(e_x \pm i e_y), \qquad e_0 = e_z. \tag{48}$$

The change of basis is described by the unitary matrix

$$W^{(1)} = \begin{pmatrix} -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{i}{\sqrt{2}} & 0 & \frac{i}{\sqrt{2}} \\ 0 & 1 & 0 \end{pmatrix}. \tag{49}$$

There are two separate rotations we can do on Equation (35). Firstly, we can rotate the tangent space vectors with $W^{(1)}$. These vectors arise from the m_s spin indices of Equation (31). Separately, we can also choose a different basis for the spherical harmonics themselves via a rotation on the m indices. The fact that we have both these choices arises because the vector harmonics form a basis for irreducible representations of vector fields on the sphere.

Since Equation (35) is a scalar equation in the tangent space, we know that it transforms as a scalar under changes of basis of the tangent space, and so the change of basis from spherical to Cartesian coordinates on the tangent space is trivial. However, Equation (35) has free m labels, and so the equation does transform when we change the basis from complex to real harmonics.

Specifically, let $W^{(l)}$ be a unitary change-of base matrix from complex spherical basis to the real spherical basis Y_{lm}

$$Y_{lm} = \sum_m W^{(l)}_{mm'} Y^l_{m'}. \tag{50}$$

$W^{(l)}$ can be defined by the following relations

$$\begin{aligned} &Y_{lm} = \frac{(-1)^m}{\sqrt{2}} \left(Y^l_m + (-1)^m Y^l_{-m}\right), \;\; Y_{l,-m} = \frac{1}{\sqrt{2}\, i} \left(Y^l_m - (-1)^m Y^l_{-m}\right), \;\; m > 0 \\ &Y_{l0} = Y^l_0. \end{aligned} \tag{51}$$

Under such a rotation, both the left and right hand side of Equation (35) transform. In particular we find

$$\int_{\mathbb{S}^2} (\nabla Y_{l_1 m_1} \times \nabla Y_{l_2 m_2}) \cdot (\hat{r} Y_{l_3 m_3}) \mathrm{d}\mu_{\mathbb{S}^2}(\hat{r}) = \Lambda_{l_1 l_2 l_3} \left(C^{\mathbb{R}}\right)^{l_3 m_3}_{l_1 m_1\, l_2 m_2}, \tag{52}$$

where we have defined the real basis Clebsch-Gordan tensor

$$\left(C^{\mathbb{R}}\right)^{l_3 m_3}_{l_1 m_1\, l_2 m_2} = \sum_{m'_1, m'_2, m'_3} W^{(l_1)}_{m_1 m'_1} W^{(l_2)}_{m_2 m'_2} C^{l_3 m'_3}_{l_1 m'_1\, l_2 m'_2} \left(W^{(l_3)\dagger}\right)_{m'_3 m_3}. \tag{53}$$

Note that the left hand side of Equation (52) is now real, and so the right hand side must also be. However, there is a puzzling factor in i in Equation (42) for Λ_{l_1,l_2,l_3}. This is accounted for when we note that the real-basis Clebsch-Gordan tensor is purely imaginary for odd parity. More generally we have

$$\left(\left(C^{\mathbb{R}}\right)^{l_3 m_3}_{l_1 m_1\, l_2 m_2}\right)^* = (-1)^{l_1 + l_2 - l_3} \left(C^{\mathbb{R}}\right)^{l_3 m_3}_{l_1 m_1\, l_2 m_2}. \tag{54}$$

so

$$\left(C^{\mathbb{R}}\right) \in \begin{cases} \mathbb{R} & l_1 + l_2 - l_3 \text{ even}, \\ i\,\mathbb{R} & l_1 + l_2 - l_3 \text{ odd}. \end{cases} \tag{55}$$

We now give a short proof of Equation (54).

Consider, the complex conjugate of Equation (53)

$$\left(\left(C^{\mathbb{R}}\right)^{l_3 m_3}_{l_1 m_1\, l_2 m_2}\right)^* = \sum_{m'_1, m'_2, m'_3} W^{(l_1)*}_{m_1 m'_1} W^{(l_2)*}_{m_2 m'_2} C^{l_3 m'_3}_{l_1 m'_1\, l_2 m'_2} W^{(l_3)T}_{m'_3 m_3}. \tag{56}$$

For the complex to real spherical harmonic basis change, the matrix $W^{(l)}$ satisfies

$$W^{(l)*}_{mm'} = (-1)^{m'} W^{(l)}_{m,-m'}, \tag{57}$$

which follows from the identity $Y^{l*}_m = (-1)^m Y^l_{-m}$. We now use the known identity (see Equation 24, page 247 (Varshalovich et al., 1988)) that

$$C^{l_3 m_3}_{l_1 m_1\, l_2 m_2} = (-1)^{l_1 + l_2 - l_3} C^{l_3, -m_3}_{l_1, -m_1\, l_2, -m_2} \tag{58}$$

Substituting Equation (58) into Equation (56) and relabeling the summation indices we find

$$\left((C^{\mathbb{R}})^{l_3 m_3}_{l_1 m_1\, l_2 m_2}\right)^* = (-1)^{l_1+l_2-l_3} \sum_{m_1',m_2',m_3'} W^{(l_1)}_{m_1 m_1'}\, W^{(l_2)}_{m_2 m_2'}\, C^{l_3 m_3'}_{l_1 m_1'\, l_2 m_2'} \left(W^{(l_3)\dagger}\right)_{m_3' m_3}, \tag{59}$$

where we have also used that $(-1)^{m_1'+m_2'+m_3'} = (-1)^{2m_3'} = 1$ via selection rules. Recognizing the original definition of $(C^{\mathbb{R}})$ in Equation (53) we find

$$\left((C^{\mathbb{R}})^{l_3 m_3}_{l_1 m_1\, l_2 m_2}\right)^* = (-1)^{l_1+l_2-l_3} (C^{\mathbb{R}})^{l_3 m_3}_{l_1 m_1\, l_2 m_2}. \tag{60}$$

which completes the proof.

Note that in computational libraries such as e3nn (Geiger & Smidt, 2022), the rotation of the Clebsch-Gordan coefficients is defined with an extra factor of i in according to $\tilde{W}^{(l)} = (-i)^l W^{(l)}$. With this in mind, the real e3nn Clebsch-Gordan coefficients are defined by

$$(\tilde{C}^{\mathbb{R}})^{l_3 m_3}_{l_1 m_1 l_2 m_2} = i^{l_1+l_2-l_3} (C^{\mathbb{R}})^{l_3 m_3}_{l_1 m_1 l_2 m_2} = i^{l_1+l_2+l_3} (-1)^{l_3} (C^{\mathbb{R}})^{l_3 m_3}_{l_1 m_1 l_2 m_2} \tag{61}$$

which is always real. With these conventions, our result Eq. (52) reads

$$\int_{\mathbb{S}^2} (\nabla Y_{l_1 m_1} \times \nabla Y_{l_2 m_2}) \cdot (\hat{r} Y_{l_3 m_3}) \mathrm{d}\mu_{\mathbb{S}^2}(\hat{r}) = \mathrm{Im}(\Lambda_{l_1 l_2 l_3})\, (-i)^{l_1+l_2+l_3-1} (-1)^{l_3} (\tilde{C}^{\mathbb{R}})^{l_3 m_3}_{l_1 m_1\, l_2 m_2}, \tag{62}$$

which is always real for odd parity. Finally, identifying the above formula with Eq. (13) of the main text, we get

$$\tilde{V}^{l_3}_{l_1,l_2} = (-1)^{\frac{l_1+l_2+l_3-1}{2}+l_3} \mathrm{Im}(\Lambda_{l_1 l_2 l_3}). \tag{63}$$

A.4. Low-Rank Decomposition of Coupling Coefficients

In this Section we discuss the normalization of the antisymmetric Gaunt coupling coefficients introduced in the main text. When these quantities are used within $O(3)$-equivariant neural networks, it is desirable that the associated tensor products have magnitudes comparable to standard Clebsch-Gordan tensor products so that network layers are well initialized and numerically stable.

As discussed in Section 4, the efficient implementation of an integral tensor product formula relies on a factorized representation of the spherical integral that separates the dependence on l_1, l_2, and l_3. If the tensor product is explicitly normalized by multiplying each channel by a general tensor, this factorization would be destroyed unless the normalization itself can be expressed in a separable form.

In particular, practical use of the factorized integral formula discussed in the main text benefits from an explicit low-rank decomposition of the coupling tensor if one wants to explicitly normalize the tensor product. We therefore wish to approximate normalization factors of the form

$$(T^{l_3}_{l_1 l_2})^{-1} \approx \sum_{r=1}^{R} a_{l_1,r}\, b_{l_2,r}\, c_{l_3,r} \tag{64}$$

for a small R, so that we can initialize layers to be approximately the same magnitude as the Clebsh-Gordan tensor product and benefit from the factorization trick. In Equation (64), $T^{l_3}_{l_1,l_2}$ can be either the anti-symmetric or symmetric Gaunt coefficients, $T^{l_3}_{l_1,l_2} \in \{\tilde{V}^{l_3}_{l_1,l_2}, \tilde{G}^{l_3}_{l_1,l_2}\}$.

Empirically we find that $(T^{l_3}_{l_1 l_2})^{-1}$ is intrinsically low rank. Figure 2 compares the quality of rank-1 and rank-2 approximations across $L_{\max}$ for $T = \tilde{V}$, the antisymmetric Gaunt coefficients. A rank-2 decomposition accurately reproduces the tensor over a wide range of angular momenta. In particular, a rank-2 approximation achieves approximately 10% per-entry accuracy for all tested values up to $L_{\max} = 19$. By contrast, a rank-1 approximation fails qualitatively and cannot reproduce the structure of the tensor. Rank 1 is qualitatively incorrect ($\sigma_{\log} \sim 1$–2, and we also noticed sign errors over large subspaces), whereas rank 2 achieves $\sigma_{\log} < 0.04$ at all tested $L_{\max}$, with 100% of entries accurate to within a factor of two. This behavior likely arises from the two distinct coupling pathways $l_3 \pm 1$ appearing in Eq. (42), which introduce two dominant modes in the coupling tensor.

Figure 3 compares fits of $(\tilde{V}^{l_3}_{l_1 l_2})^{-1}$ with fits of the even-parity Gaunt coefficients $(\widetilde{G}^{l_3}_{l_1,l_2})^{-1}$. The Gaunt tensor depends on a single Clebsch–Gordan coefficient coupling, and appears approximately rank 1 ($\sigma_{\log} < 0.13$). This perhaps explains why the authors of (Luo et al., 2024) did not encounter issues with normalization.

The decomposition in Equation by minimizing the mean relative squared error

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \frac{(y_i - \hat{y}_i)^2}{y_i^2}, \tag{65}$$

where the sum runs over the N non-zero entries permitted by the selection rules. This loss treats all tensor elements equally on a relative scale, which is important because the $\tilde{V}^{l_3}_{l_1 l_2}$ span several orders of magnitude. To quantify the quality of the approximation we report the logarithmic ratio spread

$$\sigma_{\log} = \mathrm{std}[\log_{10} |y_i/\hat{y}_i|] \tag{66}$$

which measures the dispersion of multiplicative errors. In this metric, $\sigma_{\log} = 0$ corresponds to exact agreement, $\sigma_{\log} = 0.3$ corresponds to a factor-of-two scatter, and $\sigma_{\log} = 0.04$ corresponds to approximately 10% relative error.

Optimization was performed using L-BFGS-B with numerically estimated gradients, convergence tolerances $f_{\mathrm{tol}} = 10^{-15}$ and $g_{\mathrm{tol}} = 10^{-10}$, and a maximum of 300 iterations per run. In practice the optimization converged reliably, and alternative optimization methods yielded similar results.

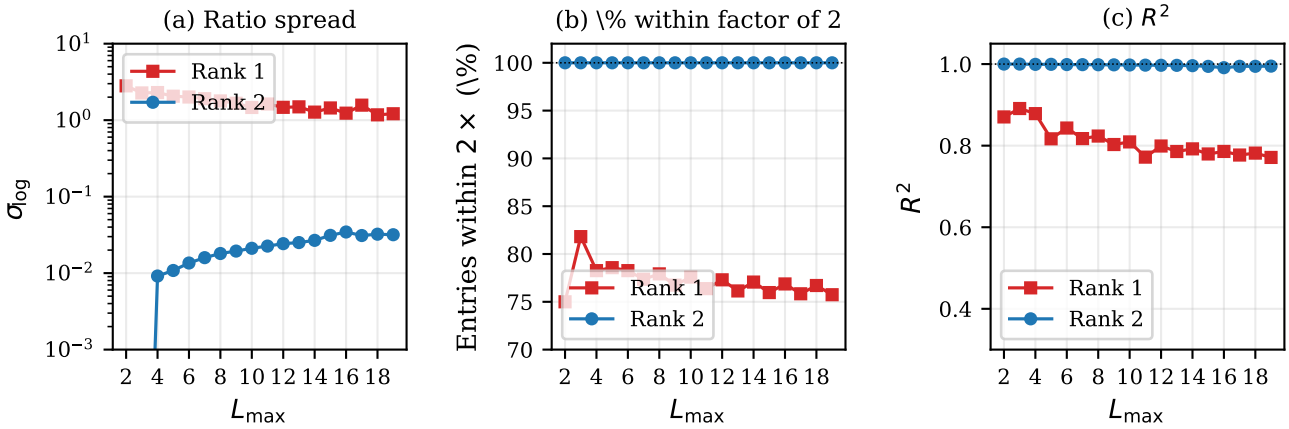

Figure 2. Decomposition quality for $(\tilde{V}^{l_3}_{l_1 l_2})^{-1}$ vs $L_{\max}$. **(a)** $\sigma_{\log}$ (log scale): rank 2 remains below 0.04, while rank 1 exhibits order-of-magnitude errors. **(b)** 100% of rank-2 entries lie within a factor of 2. **(c)** R^2 for the rank-2 fit remains above 0.9.

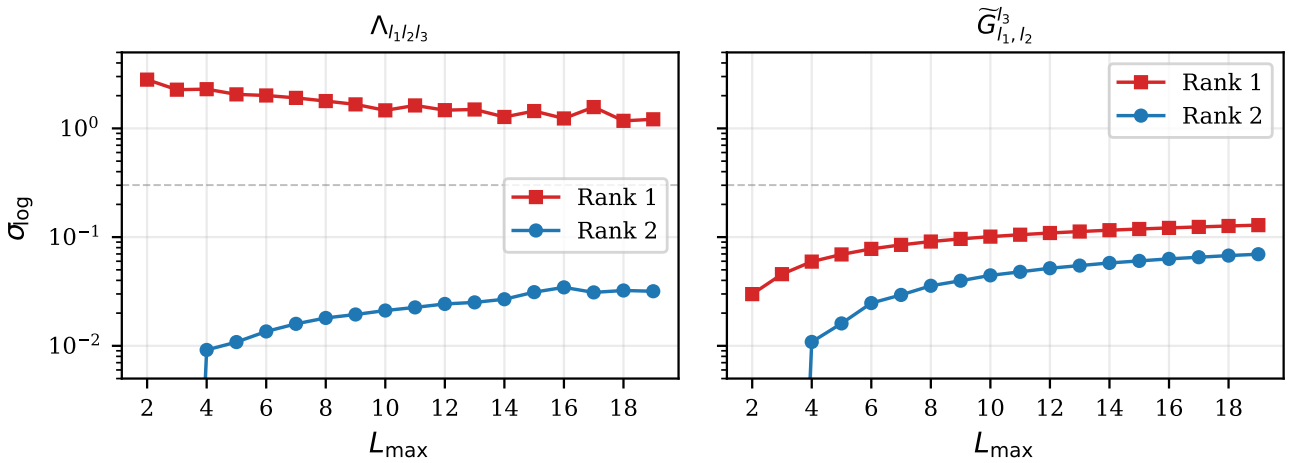

Figure 3. $\sigma_{\log}$ vs $L_{\max}$ for $(\tilde{V}^{l_3}_{l_1 l_2})^{-1}$ (left) and $(\widetilde{G}^{l_3}_{l_1,l_2})^{-1}$ (right). The dashed line indicates $2\times$ scatter. We see that $\tilde{V}^{-1}$ requires rank 2, whereas $\widetilde{G}^{-1}$ a single rank approximation is sufficient for normalization in a neural network.