

From Information Geometry to Jet Substructure: A Triality of Cumulant Tensors, Energy Correlators, and Hypergraphs

Aritra Bal,^{1,3} Markus Klute,¹ Benedikt Maier,² and Michael Spannowsky^{3,4}

¹*Institute of Experimental Particle Physics, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany*

²*Blackett Laboratory, Imperial College of Science, Technology and Medicine, London, SW7 2AZ, United Kingdom*

³*Institute for Theoretical Physics, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany*

⁴*Institute for Quantum Materials and Technologies, Karlsruhe Institute of Technology, Karlsruhe 76131, Germany*

E-mail: aritra.bal@kit.edu, markus.klute@kit.edu,
benedikt.maier@cern.ch, michael.spannowsky@kit.edu

ABSTRACT: Pairwise Fisher graphs capture local covariance information, but they cannot distinguish an irreducible multi-observable radiation pattern from a collection of ordinary pairwise correlations. We show that this missing structure is naturally supplied by the higher Fisher tensors. In a finite basis of binned EECs, ECFs, or EFPs, and in the natural exponential-family coordinates generated by that basis, the same local tensor has three equivalent interpretations: it is a coefficient in the local Kullback–Leibler expansion, a connected cumulant of the chosen correlator observables, and a signed weight on a hyperedge linking those observables. This gives an exact Fisher–correlator–hypergraph triality in the local exponential-family embedding.

The triality provides a direct construction of physics-informed hypergraphs from measured or simulated correlator data. Extending the quadratic Fisher matrix to the first non-trivial higher tensor identifies genuinely connected multi-observable radiation patterns, supplies hyperedge weights for higher-order Laplacians and message passing, and gives a principled criterion for compressing observable bases beyond pairwise information. We develop these constructions and spell out why the exact cumulant interpretation is special to natural exponential-family coordinates.

We illustrate the framework in four applications. In a minimal local-KL study, including the cubic Fisher tensor reduces the KL truncation error by about 30x near the reference point and isolates the dominant triplet structure. In a $W \rightarrow q\bar{q}$ versus $t \rightarrow bq\bar{q}$ substructure benchmark, the hypergraph selector improves compressed-basis classification. In a 33-observable basis-design problem, the Fisher hypergraph retains more third-order local response at twelve observables, with $\bar{R}_{(3)} = 0.937$ compared with 0.871 for the pairwise graph. A low-capacity learning benchmark then shows how the same Fisher hyperedges can be used as an interpretable inductive bias for message passing on correlator observables.

Contents

1	Introduction	2
2	Fisher Information Tensors	4
2.1	Natural coordinates and cumulant tensors	4
3	Energy Correlators	6
3.1	EECs and ECFs	6
4	Hypergraphs	7
4.1	Weighted hypergraphs and Laplacians	7
4.2	Learning rules	9
5	The Fisher–Correlator–Hypergraph Triality	9
5.1	Fisher tensors and energy correlators	9
5.2	Fisher tensors and hypergraphs	10
5.3	Energy correlators and hypergraphs	10
6	Applications of the Triality Relation	11
6.1	Simulation Setup	12
6.2	Minimal Study of Local KL and Triplet Structure	12
6.2.1	Setup	12
6.2.2	Third Order Improves the Local Divergence Model	13
6.2.3	The Hypergraph Sees Structure the Graph Smears Out	14
6.3	Substructure Probes and Fisher-Hypergraph Compression	14
6.3.1	Benchmark and Compression Rule	16
6.3.2	Classification Performance Under Compression	18
6.3.3	Interpretable Prong-Sensitive Hyperedges	19
6.4	Observable Discovery and Higher-Order Basis Design	21
6.5	Physics-Informed Hypergraph Learning	24
7	Summary and Conclusions	26
A	Generic-Coordinate Form of the Cubic KL Coefficient	27

1 Introduction

Energy correlators provide one of the cleanest ways to describe radiation patterns at colliders. The energy–energy correlation (EEC) measures pairwise energy flow as a function of angular separation, while higher-point energy correlation functions (ECFs) and the energy-flow polynomial (EFP) basis extend the same logic to multi-prong structure [1–5]. These observables are therefore a natural language for jet substructure, but the way their information is organised matters. A pairwise graph can encode covariances among observables, but it cannot tell whether a triangle of strong pairwise links came from one irreducible three-observable fluctuation or from three unrelated pairwise effects.

Hypergraphs are designed precisely for this kind of multiway structure. Unlike ordinary graphs, whose edges join only pairs of vertices, a hypergraph allows one hyperedge to connect an entire subset of observables, particles, or features at once [6, 7]. This is why hypergraph methods have become useful in machine learning and collider reconstruction [8–10]. What is missing, however, is a principled physics prescription for the hyperedge weights. In collider applications one would like a hyperedge not to be merely an architectural choice, but to correspond to a definite radiation pattern and to a definite contribution to local distinguishability.

Information geometry supplies the relevant hierarchy. The Fisher information matrix is the quadratic coefficient in the local expansion of the Kullback–Leibler divergence, but it is only the rank-two member of a tower of higher Fisher tensors [11–15]. The higher-rank tensors encode the first departures from a purely Gaussian, pairwise description of local statistical response. In machine learning, the same Fisher geometry also underlies natural-gradient methods [16]; here we use it instead as a bridge between collider observables and higher-order combinatorial structure.

The central observation of this paper is that the bridge becomes exact in a local exponential-family embedding generated by the chosen observable basis. We choose a finite set of binned EECs, ECFs, or EFPs as sufficient statistics and write a local model of the form $p_\theta(x) = h(x) \exp[\theta^a \phi_a(x) - \psi(\theta)]$. This is not a claim that the full space of collider event distributions is globally an exponential family. Rather, it is a local chart around a reference sample, in which the chosen observables generate the deformations being probed. In this natural coordinate system the log-partition function $\psi(\theta)$ is a cumulant-generating function: its derivatives are simultaneously coefficients in the local KL expansion, connected cumulants of the correlator basis, and weights on hyperedges linking the corresponding observables. We call this equivalence the *Fisher–correlator–hypergraph triality*.

The practical consequence is direct. A connected triplet of correlators acquires a statistical meaning because it contributes to local distinguishability, a physical meaning because it represents an irreducible radiation-pattern fluctuation, and an algorithmic meaning because it specifies a hyperedge that should be retained or propagated jointly. The first genuinely new layer begins at rank three, where the cubic Fisher tensor can expose structure that a pairwise Fisher graph necessarily smears into ordinary edges.

Concretely, this paper makes three contributions:

1. We derive the higher-order KL expansion in natural exponential-family coordinates

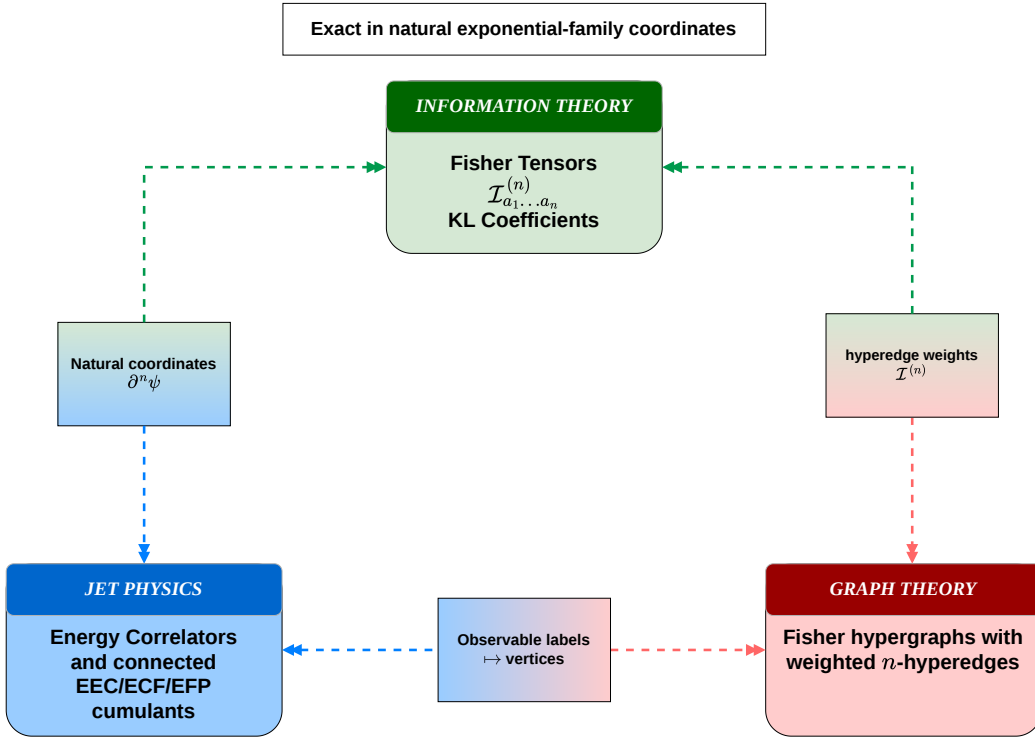


Figure 1. The central triality of the paper. In an exponential-family embedding built from energy correlators, the same tensor is simultaneously a KL coefficient, a connected correlator cumulant, and a hyperedge weight.

and identify the resulting Fisher tensors with connected cumulants of a finite correlator basis.

2. We use the same tensors to construct signed Fisher hypergraphs, together with the corresponding hypergraph Laplacians and message-passing operators.
3. We demonstrate the construction in four applications: a minimal local-KL study, a prong-substructure compression benchmark, a higher-order basis-design problem, and a low-capacity physics-informed learning benchmark.

Figure 1 summarises this logic and serves as a roadmap for the rest of the paper.

We organise the paper by introducing the three ingredients separately before proving their relation. Section 2 develops the Fisher tensor hierarchy from the KL expansion and states precisely when the cumulant interpretation is exact. Section 3 reviews the EEC, higher-point ECFs, and the EFP basis as the collider observables of interest. Section 4 summarises the weighted hypergraph language, Laplacians, and learning operators that will later inherit physics-motivated weights. Section 5 then proves the triality by establishing the three pairwise relations shown in Figure 1. Section 6 gives four concrete applications, beginning with a minimal local-KL study. Appendix A records the generic-coordinate caveat that motivates our use of exponential coordinates.

2 Fisher Information Tensors

In information geometry, one may expand a smooth statistical divergence around two nearby probability distributions and read off a hierarchy of local tensor fields from its coefficients. For the Kullback–Leibler divergence, the quadratic term yields the Fisher information metric, the cubic term gives the Amari–Chentsov tensor, namely the rank-three member of the Fisher-tensor hierarchy, and the same construction extends to higher symmetric tensors [11–15, 17]. In exponential-family coordinates this hierarchy becomes especially concrete, because the KL coefficients are generated directly by derivatives of the log-partition function. Equivalently, the log-partition function $\psi(\theta)$ is the generating function for the full local hierarchy: its successive derivatives generate the Fisher tensors, just as a cumulant-generating function generates connected moments. We therefore begin by recalling this construction in a form suited to the collider observables used later in the paper.

2.1 Natural coordinates and cumulant tensors

Let x denote an event and let $\phi_a(x)$, with $a = 1, \dots, p$, be a chosen observable basis. We consider a model from the exponential family,

$$p_\theta(x) = h(x) \exp[\theta^a \phi_a(x) - \psi(\theta)], \quad (2.1)$$

which is one of the standard ways of writing a probability distribution when a chosen set of observables is meant to carry the relevant information. In this form, $h(x)$ is a reference distribution, the $\phi_a(x)$ are the observables or features whose expectation values characterise the event sample, the parameters θ^a control how strongly the distribution responds to those observables, and $\psi(\theta)$ is the log-normalisation factor that ensures the total probability integrates to one. For readers with a collider background, the main point is simple: once a basis of measured observables has been chosen, the exponential-family form provides a controlled way to describe how the event distribution changes as one tilts the sample towards or away from particular radiation patterns. One may also view the parameters θ^a as source-like couplings and $\psi(\theta)$ as the associated generating function: differentiating with respect to θ generates the connected response of the observable basis, and in the present setting those generated objects are precisely the Fisher tensors.

This representation is especially useful here because it makes the link between observables and information geometry explicit. The score components, namely the derivatives of the log-likelihood with respect to the parameters, are

$$s_a(x; \theta) = \partial_a \log p_\theta(x) = \phi_a(x) - \partial_a \psi(\theta). \quad (2.2)$$

where it can be shown that $\partial_a \psi(\theta) = \langle \phi_a(x) \rangle_{p_\theta}$, by a single differentiation of the normalisation requirement for Eq. (2.1). They are therefore just the chosen observables with their mean values subtracted. In other words, the score tells us how unusual a given event looks relative to the average event in the direction of observable a . They are centred by construction, $\langle s_a \rangle_\theta = 0$.

The central structural statement is that the full tower of derivatives of ψ is the cumulant tower of the observables, and therefore also the Fisher-tensor hierarchy itself [18].

Theorem 1 (Fisher tensors as cumulants). *For the exponential family in Eq. (2.1),*

$$\partial_{a_1} \cdots \partial_{a_n} \psi(\theta) = \kappa_\theta(\phi_{a_1} \cdots \phi_{a_n}) \equiv \mathcal{I}_{a_1 \cdots a_n}^{(n)}(\theta), \quad n \geq 1. \quad (2.3)$$

In particular,

$$\mathcal{I}_{ab}^{(2)}(\theta) = \text{Cov}_\theta(\phi_a, \phi_b) = \langle s_a s_b \rangle_\theta, \quad (2.4)$$

$$\mathcal{I}_{abc}^{(3)}(\theta) = \kappa_\theta(\phi_a \phi_b \phi_c) = \langle s_a s_b s_c \rangle_\theta. \quad (2.5)$$

Proof. Normalization implies

$$\psi(\theta + t) - \psi(\theta) = \log \langle \exp(t^a \phi_a) \rangle_\theta. \quad (2.6)$$

Expanding the right-hand side as a multivariate cumulant generating function gives [19]

$$\psi(\theta + t) - \psi(\theta) = \sum_{n=1}^{\infty} \frac{1}{n!} \kappa_\theta(\phi_{a_1} \cdots \phi_{a_n}) t^{a_1} \cdots t^{a_n}. \quad (2.7)$$

Matching coefficients of $t^{a_1} \cdots t^{a_n}$ proves Eq. (2.3). The identities in Eqs. (2.4) and (2.5) follow from $s_a = \phi_a - \langle \phi_a \rangle_\theta$. \square

The rank-two tensor $\mathcal{I}^{(2)}$ is the ordinary Fisher information matrix. For $n \geq 3$, $\mathcal{I}^{(n)}$ is not a metric tensor, but it still carries invariant local information about how the model departs from the Gaussian approximation around θ . In natural exponential coordinates, the cubic tensor $\mathcal{I}^{(3)}$ coincides with the Amari–Chentsov tensor, that is, with the rank-three Fisher tensor [14, 15, 20].

Corollary 2 (Exact KL expansion). *For the same exponential family, let δ denote a small parameter displacement so that $p_{\theta+\delta}$ is a nearby distribution, let $D_{\text{KL}}(p_\theta \| p_{\theta+\delta})$ be the Kullback–Leibler divergence from p_θ to $p_{\theta+\delta}$, and let $\mathcal{I}_{a_1 \cdots a_n}^{(n)}(\theta) = \partial_{a_1} \cdots \partial_{a_n} \psi(\theta)$ be the rank- n Fisher tensor at θ . Then*

$$D_{\text{KL}}(p_\theta \| p_{\theta+\delta}) = \sum_{n=2}^{\infty} \frac{1}{n!} \mathcal{I}_{a_1 \cdots a_n}^{(n)}(\theta) \delta^{a_1} \cdots \delta^{a_n}. \quad (2.8)$$

Proof. Using Eq. (2.1),

$$D_{\text{KL}}(p_\theta \| p_{\theta+\delta}) = \psi(\theta + \delta) - \psi(\theta) - \delta^a \partial_a \psi(\theta). \quad (2.9)$$

Expanding $\psi(\theta + \delta)$ about θ and canceling the linear term gives Eq. (2.8). \square

Equation (2.8) is the precise sense in which the higher Fisher tensors control local distinguishability. The quadratic truncation is the familiar Fisher metric approximation. The cubic term is the first correction that can distinguish a skew local response from a purely Gaussian one.

Remark 1 (Why we insist on exponential coordinates). For a generic parametric family, the cubic coefficient of the KL expansion is not simply $\langle s_a s_b s_c \rangle$. The exact identity mixes third score moments with expectations of second derivatives of the log-likelihood. Appendix A records that formula explicitly. The exponential-family embedding makes the higher-order KL coefficients equal to connected observable cumulants without residue terms.

3 Energy Correlators

Energy correlators are infrared- and collinear-safe observables that characterise how the energy carried by an event is distributed over angles, and for that reason they provide a direct probe of QCD radiation patterns. The energy-energy correlator (EEC) is the basic two-point example: it measures pairwise angular energy flow, has a long history as a precision observable in e^+e^- annihilation, and has also become central in conformal collider theory and in modern perturbative studies of event structure [1, 4, 5, 21]. Higher-point energy correlation functions (ECFs) extend the same idea to genuinely multiparticle patterns and have become standard tools for resolving jet substructure, especially when multi-prong structure is the discriminating feature [2, 22–24]. Energy flow polynomials (EFPs) organise this energy-flow information into a systematic linear basis for infrared- and collinear-safe observables, making it possible to treat a broad class of collider measurements within one common language [3]. Related geometric perspectives on the full space of collider events have also been developed [25]. In the present work these observables provide the physically motivated basis from which the Fisher tensors and the associated hypergraph weights will be constructed.

3.1 EECs and ECFs

For an event with constituents $i = 1, \dots, N$, let

$$z_i = \frac{E_i}{\sum_j E_j} \quad \text{or} \quad z_i = \frac{p_{T,i}}{\sum_j p_{T,j}}, \quad (3.1)$$

depending on whether the observable is defined in e^+e^- or hadron-collider kinematics. The n -point energy correlation function with angular exponent $\beta > 0$ is

$$e_n^{(\beta)}(x) = \sum_{1 \leq i_1 < \dots < i_n \leq N} \left(\prod_{r=1}^n z_{i_r} \right) \left(\prod_{r < s} \theta_{i_r i_s}^\beta \right), \quad (3.2)$$

where θ_{ij} denotes an opening angle or ΔR_{ij} .

The eventwise EEC is the differential pairwise correlator

$$\text{EEC}_x(\chi) = \sum_{i < j} 2z_i z_j \delta(\cos \chi - \cos \chi_{ij}). \quad (3.3)$$

The factor of two together with the ordered sum $\sum_{i < j}$ is equivalent to the unordered convention $\sum_{i \neq j} z_i z_j \delta(\cos \chi - \cos \chi_{ij})$, and is chosen here so that the kernel relation in Proposition 3 carries the conventional $\frac{1}{2}$ prefactor. This relation is relevant for what follows because it puts the EEC and the ECFs on the same footing. The EEC resolves pairwise energy flow differentially in the angular variable χ , whereas $e_2^{(\beta)}$ is an integrated observable obtained by weighting that same distribution with a kernel. Consequently, a basis built from EEC bins and a basis built from two-point ECFs probes the same underlying pairwise radiation pattern at different levels of resolution. This is the reason one can later reinterpret cumulants of ECF observables as kernel-smearred cumulants of the EEC itself. The simplest instance of this statement is the following identity.

Proposition 3 (The two-point ECF as an EEC moment). *For e^+e^- kinematics,*

$$e_2^{(\beta)}(x) = \frac{1}{2} \int_{-1}^1 d(\cos \chi) \left(2 \sin \frac{\chi}{2}\right)^\beta \text{EEC}_x(\chi). \quad (3.4)$$

Proof. Substituting Eq. (3.3) into the right-hand side of Eq. (3.4) gives

$$\frac{1}{2} \sum_{i < j} 2z_i z_j \int_{-1}^1 d(\cos \chi) \left(2 \sin \frac{\chi}{2}\right)^\beta \delta(\cos \chi - \cos \chi_{ij}). \quad (3.5)$$

The delta function localises the integral at the pairwise angle χ_{ij} , so the expression reduces to

$$\sum_{i < j} z_i z_j \left(2 \sin \frac{\chi_{ij}}{2}\right)^\beta, \quad (3.6)$$

which is precisely the spherical definition of $e_2^{(\beta)}$ in Eq. (3.2) for $n = 2$. \square

The proposition makes precise the sense in which the two-point ECF is not an un-related observable but a coarse-grained version of the EEC. This point is used explicitly in Section 5, where observables that are linear functionals of the EEC inherit an interpretation in terms of kernel-smearred connected EEC cumulants. Higher ECFs and EFPs similarly arise from kernels acting on higher-point energy-flow observables [3]. Throughout the rest of the paper we work with a finite correlator basis $\{\phi_a\}_{a=1}^p$ chosen from binned EECs, ECFs, and EFPs.

4 Hypergraphs

Hypergraphs are combinatorial structures in which the basic relation is allowed to involve more than two objects at once. Formally, a hypergraph $\mathcal{H} = (V, E)$ consists of a vertex set V together with a family E of subsets of V , called hyperedges; unlike an ordinary graph edge, which always joins exactly two vertices, a hyperedge may connect any number of vertices [6, 7]. This extension is important whenever the relevant structure is genuinely multiway. A graph can represent only pairwise links, so a collective three-way dependence must be projected onto several ordinary edges and is thereby stripped of its irreducible character. Hypergraphs have therefore become a natural language for higher-order structures in combinatorics and network science, for spectral clustering and embedding methods, and for modern hypergraph neural networks [7–9]. More recently they have also been proposed as collider-physics architectures precisely because they can retain higher-order correlations that pairwise graph constructions smear out [26].

4.1 Weighted hypergraphs and Laplacians

Let V be a finite set of vertices. In the present paper these vertices will eventually label observables, but for the moment it is enough to think of them as the nodes of the combinatorial structure. For a fixed order $r \geq 2$, an r -uniform weighted hypergraph consists of a collection $E^{(r)}$ of subsets $e \subseteq V$ together with a nonnegative weight $w_e^{(r)}$ assigned to each

such subset. The condition $|e| = r$ means that every hyperedge in $E^{(r)}$ connects exactly r vertices. Thus $r = 2$ reproduces an ordinary weighted graph, while $r = 3$ allows one edge to connect three vertices simultaneously.

It is useful to encode this information in matrices. The incidence matrix $B^{(r)} \in \{0, 1\}^{|V| \times |E^{(r)}|}$ records which vertices belong to which hyperedges: its rows are labelled by vertices $a \in V$, its columns are labelled by hyperedges $e \in E^{(r)}$, and its entries are

$$B_{ae}^{(r)} = \begin{cases} 1, & a \in e, \\ 0, & a \notin e. \end{cases} \quad (4.1)$$

The diagonal matrix $W^{(r)} = \text{diag}(w_e^{(r)})$ stores the hyperedge weights. From it one defines the vertex degree

$$d_a^{(r)} = \sum_{e \in E^{(r)}} B_{ae}^{(r)} w_e^{(r)}, \quad (4.2)$$

which measures the total order- r weight incident on vertex a . Collecting these degrees into the diagonal matrix $D_v^{(r)} = \text{diag}(d_a^{(r)})$, and writing $\delta_e^{(r)} = |e| = r$ for the degree of a hyperedge itself, one also defines the diagonal hyperedge-degree matrix $D_e^{(r)} = \text{diag}(\delta_e^{(r)})$. With this notation, a standard normalised hypergraph Laplacian is [8]

$$\mathcal{L}^{(r)} = I - \left(D_v^{(r)}\right)^{-1/2} B^{(r)} W^{(r)} \left(D_e^{(r)}\right)^{-1} \left(B^{(r)}\right)^\top \left(D_v^{(r)}\right)^{-1/2}. \quad (4.3)$$

For readers more familiar with ordinary graphs, $\mathcal{L}^{(r)}$ plays the same role as the normalised graph Laplacian: it compares a feature at one vertex to a weighted average of features on vertices that participate in the same hyperedges. The ordinary graph is recovered as the special case $r = 2$.

In many applications several hyperedge orders are relevant at once. A multi-order hypergraph therefore retains several values of r and combines their Laplacians as

$$\mathcal{L}_{\text{multi}} = \sum_{r=2}^{r_{\text{max}}} \alpha_r \mathcal{L}^{(r)}, \quad \alpha_r \geq 0. \quad (4.4)$$

The coefficients α_r control how much each order contributes. For example, α_2 weights pairwise information, while α_3 weights irreducible three-way information.

This distinction is important for the work at hand because a graph projection can discard exactly the higher-order information that the Fisher tensors are meant to capture. A standard reduction, called a clique expansion, replaces one r -hyperedge by the full set of $\binom{r}{2}$ pairwise edges among the same vertices. After that replacement, however, the original meaning of the structure is lost: one can no longer tell whether the resulting triangle of pairwise edges came from a single irreducible three-way correlation or from three separate pairwise correlations. Thus, in more concrete terms, the difference is the following. If three observables become informative only when they fluctuate together as a triplet, then one 3-hyperedge records that collective effect directly. An ordinary graph can only draw the three pairwise links between them, and that picture looks identical to a situation in which the three pairs were unrelatedly strong. Whenever the first informative structure appears at order three or higher, the hypergraph is therefore the natural object and the graph is only an approximation.

4.2 Learning rules

The same ingredients define higher-order message-passing operators. If $H^{(\ell)} \in \mathbb{R}^{|V| \times d_\ell}$ denotes node embeddings at layer ℓ , a simple multi-order hypergraph layer is

$$H^{(\ell+1)} = \sigma \left[\sum_{r=2}^{r_{\max}} \alpha_r \left(D_v^{(r)} \right)^{-1/2} B^{(r)} W^{(r)} \left(D_e^{(r)} \right)^{-1} \left(B^{(r)} \right)^\top \left(D_v^{(r)} \right)^{-1/2} H^{(\ell)} U_r \right], \quad (4.5)$$

with trainable matrices U_r and nonlinearity σ [9]. Equation (4.5) is standard once the hyperedge weights have been specified. In applications with signed weights, one may propagate with $|w_e^{(r)}|$ or $(w_e^{(r)})^2$ while carrying the sign as an auxiliary edge feature, and sparsification can be imposed by thresholding small entries. Related collider work has already argued that hypergraph architectures are natural for IRC-safe multipoint correlator structure [10, 26]. For collider applications the vertices can label EEC bins, ECFs at different β , or selected EFP monomials; what the present paper adds is a principled way to choose the hyperedge weights from physics itself.

5 The Fisher–Correlator–Hypergraph Triality

Having introduced the three ingredients separately, we can now connect them pairwise. The next three subsections correspond to the three edges of Figure 1. Taken together, they establish the full triangle.

5.1 Fisher tensors and energy correlators

We now return to the exponential-family model of Eq. (2.1) and choose its sufficient statistics from a finite correlator basis,

$$\phi_a \in \left\{ \text{binned EECs, } e_n^{(\beta)}, \text{ EFPs} \right\}. \quad (5.1)$$

The statement below is immediate from Theorem 1, but it is the core bridge between information geometry and collider observables, including the higher-point energy correlators studied in [22].

Theorem 4 (Energy correlator embedding). *Let the sufficient statistics in Eq. (2.1) be chosen from a set of EECs, ECFs, or EFPs. Then*

$$\mathcal{I}_{a_1 \dots a_n}^{(n)}(\theta) = \kappa_\theta(\phi_{a_1} \dots \phi_{a_n}), \quad n \geq 2. \quad (5.2)$$

If ϕ_a is a linear functional of the EEC, as in Eq. (3.4), then the same tensor is a kernel-smearred connected EEC cumulant.

Proof. The first statement is Theorem 1 specialised to a correlator basis. For the second statement, write $\phi_a = \int K_a(\chi) \text{EEC}_x(\chi) d(\cos \chi)$. Multilinearity of cumulants then gives

$$\kappa_\theta(\phi_{a_1} \dots \phi_{a_n}) = \int \prod_{r=1}^n d(\cos \chi_r) K_{a_r}(\chi_r) \kappa_\theta(\text{EEC}(\chi_1) \dots \text{EEC}(\chi_n)). \quad (5.3)$$

□

This is the first edge of the triangle. In natural coordinates, the higher Fisher tensors are the connected fluctuations of familiar collider observables.

5.2 Fisher tensors and hypergraphs

The second edge of Figure 1 maps the same tensor hierarchy into combinatorial language. Introduce a vertex for every observable label a . At fixed order, the full Fisher tensor contains components with repeated indices. For example, at cubic order,

$$\frac{1}{6}\mathcal{I}_{abc}^{(3)}\delta^a\delta^b\delta^c = \frac{1}{6}\sum_a\mathcal{I}_{aaa}^{(3)}(\delta^a)^3 + \frac{1}{2}\sum_{a\neq b}\mathcal{I}_{aab}^{(3)}(\delta^a)^2\delta^b + \sum_{a<b<c}\mathcal{I}_{abc}^{(3)}\delta^a\delta^b\delta^c. \quad (5.4)$$

The first two terms are retained in the exact KL expansion, but they are not ordinary set-valued 3-hyperedges: they correspond respectively to single-observable skewness and mixed repeated-observable couplings. Equivalently, the exact tensor defines a weighted multihypergraph whose hyperedges may contain repeated vertices. In the hypergraph constructions below we use the square-free projection, retaining only all-distinct index sets as ordinary hyperedges.

For every order $n \geq 2$ and every unordered all-distinct index set $e = \{a_1, \dots, a_n\}$, define a signed weight

$$\tilde{w}_e^{(n)} = \mathcal{I}_{a_1 \dots a_n}^{(n)}. \quad (5.5)$$

When a diffusion operator or Laplacian needs nonnegative weights, we use

$$w_e^{(n)} = \left| \tilde{w}_e^{(n)} \right| \quad \text{or} \quad w_e^{(n)} = \left(\tilde{w}_e^{(n)} \right)^2, \quad (5.6)$$

and retain the sign as an auxiliary attribute. This is necessary because odd-order cumulants can be negative. We will refer to the square-free order- n hypergraph obtained from the weights $\tilde{w}_e^{(n)} = \mathcal{I}_{a_1 \dots a_n}^{(n)}$ as the *Fisher n -hypergraph*, or simply as the *Fisher hypergraph* when the order is clear from context.

Proposition 5 (Fisher tensors define Fisher hypergraphs). *For each order $n \geq 2$, the all-distinct components of $\mathcal{I}_{a_1 \dots a_n}^{(n)}$ define an n -uniform signed hypergraph on the observable index set. The full tensor, including repeated-index components, defines the corresponding weighted multihypergraph and supplies the order- n local divergence coefficients in Eq. (2.8).*

Proof. The square-free hypergraph is defined by Eq. (5.5). Corollary 2 identifies the full tensor components as the coefficients of the exact KL expansion. Restricting to all-distinct components gives the set-valued hypergraph used below, while keeping repeated indices gives the exact multihypergraph representation. \square

The pairwise Fisher graph is the special case $n = 2$. The cubic Fisher tensor is the first genuinely new layer, because it defines 3-hyperedges that cannot be represented faithfully by pairwise edges without losing the distinction between irreducible three-way structure and a dense collection of ordinary correlations.

5.3 Energy correlators and hypergraphs

The remaining edge of the triangle is then a direct physical interpretation of the hypergraph. Because each vertex labels a correlator observable, a hyperedge records an irreducible joint fluctuation among a set of correlators.

Proposition 6 (Correlator hyperedges). *With the square-free construction above, the ordinary n -hyperedge $e = \{a_1, \dots, a_n\}$, with all indices distinct, is present precisely when the connected cumulant $\kappa_\theta(\phi_{a_1} \cdots \phi_{a_n})$ is nonzero. If each ϕ_a is a linear functional of the EEC, then the hyperedge weight is the corresponding kernel-smeared connected EEC cumulant.*

Proof. The first statement follows by combining Eq. (5.5) with Eq. (5.2). For the second statement, use the kernel representation $\phi_a = \int K_a(\chi) \text{EEC}_x(\chi) d(\cos \chi)$ together with multilinearity of cumulants, exactly as in the proof of Theorem 4. \square

This subsection completes Figure 1 at the level of its edges. A hyperedge is not an arbitrary machine-learning primitive: in the present setting it is a connected multipoint fluctuation of collider observables.

Theorem 7 (Triality). *For an exponential-family model built from energy correlators, the tensor $\mathcal{I}_{a_1 \dots a_n}^{(n)}$ has three equivalent interpretations:*

1. *it is the coefficient of $\delta^{a_1} \cdots \delta^{a_n} / n!$ in the exact KL expansion of Eq. (2.8);*
2. *it is the connected cumulant of the correlator observables, Eq. (5.2);*
3. *it is the signed weight of the n -multihyperedge $\{\{a_1, \dots, a_n\}\}$ in the Fisher multihypergraph, with the ordinary Fisher hypergraph obtained by restricting to all-distinct indices.*

Proof. Item 1 is Corollary 2. Item 2 is Theorem 4. Item 3 is Proposition 5. All three refer to the same tensor components. \square

The theorem illustrates that a measured or simulated correlator cumulant can be interpreted immediately as an information-geometric coefficient and as a hyperedge weight. Conversely, a hypergraph learning rule initialised from $\mathcal{I}^{(3)}$ is initialised from the cubic term of the exact local divergence. At this stage, the three edges displayed already in Figure 1 have all been established explicitly.

6 Applications of the Triality Relation

Once the triality has been established, the same tensor hierarchy can be used in several different ways. In this section we focus on four applications. The first is a minimal toy study, included because it shows in the cleanest possible setting how the cubic Fisher tensor improves the local KL model and simultaneously defines the first non-trivial hypergraph. The second is a prong-substructure compression benchmark, where the triality is used to retain the observables most sensitive to two- versus three-prong jet structure. The third is observable discovery: the hypergraph identifies which correlator collections are genuinely irreducible and therefore worth promoting to a compact analysis basis. The fourth is physics-informed learning: instead of choosing a hypergraph architecture heuristically, one can initialise it directly from the measured or simulated Fisher tensors. All four applications are developed numerically, and all follow from the same map between local

distinguishability, correlator cumulants, and hyperedge weights. Before progressing to the actual studies, a quick description of the simulation setup that is used for the generation of Monte-Carlo (MC) samples is provided in the following subsection.

6.1 Simulation Setup

Simulated event samples for all processes considered in this study are generated using a common chain of Monte Carlo tools. Hard-scattering matrix elements are computed at leading order (LO) with MADGRAPH5_AMC@NLO [27] using the Standard Model with CKM quark mixing (SM-CKM model). The NNPDF3.1_nnlo_as_0118_luxqed parton distribution function (PDF) set [28, 29] is adopted throughout. Parton showering and hadronization are performed with PYTHIA 8 [30], employing the CMS CP5 underlying-event tune [31]. The resulting particle-level events are subsequently passed through DELPHES 3 [32], a fast parametric detector simulation, configured with the official CMS detector card distributed with the package. The specific processes and their corresponding sample sizes are described in the sections that follow.

6.2 Minimal Study of Local KL and Triplet Structure

We begin with a minimal class-mixture example because it gives the cleanest low-dimensional view of the full triality. In this four-observable setting one can see directly that the cubic Fisher tensor both improves the local divergence model and isolates the first genuinely irreducible triplet in correlator space.

6.2.1 Setup

This study is intentionally simple. We use 100,000 events, retaining up to the 25 highest- p_{T} constituents, from a 22% admixture of QCD jets and hadronic top decay jets. For both samples, we apply the following kinematic selections to select sufficiently boosted events in a mass window around the top quark mass:

- $p_{\text{T,jet}} > 300 \text{ GeV}$
- $m_{\text{SD}} \in [140, 220] \text{ GeV}$

The resulting feature distribution is visibly non-Gaussian and therefore has a nonzero cubic Fisher tensor.

The observable basis is

$$\phi_1 = \text{EEC}_{\text{narrow}} = \sum_{i < j} z_i z_j \mathbf{1}_{0.06 < \Delta R_{ij} \leq 0.20}, \quad (6.1)$$

$$\phi_2 = \text{EEC}_{\text{wide}} = \sum_{i < j} z_i z_j \mathbf{1}_{0.20 < \Delta R_{ij} \leq 0.80}, \quad (6.2)$$

$$\phi_3 = e_2^{(1)}, \quad (6.3)$$

$$\phi_4 = e_3^{(1)}. \quad (6.4)$$

The first two observables are binned EECs; the last two are ECFs. This basis already mixes the pairwise EEC language with higher-point correlators, making it a minimal setting in which to test the triality.

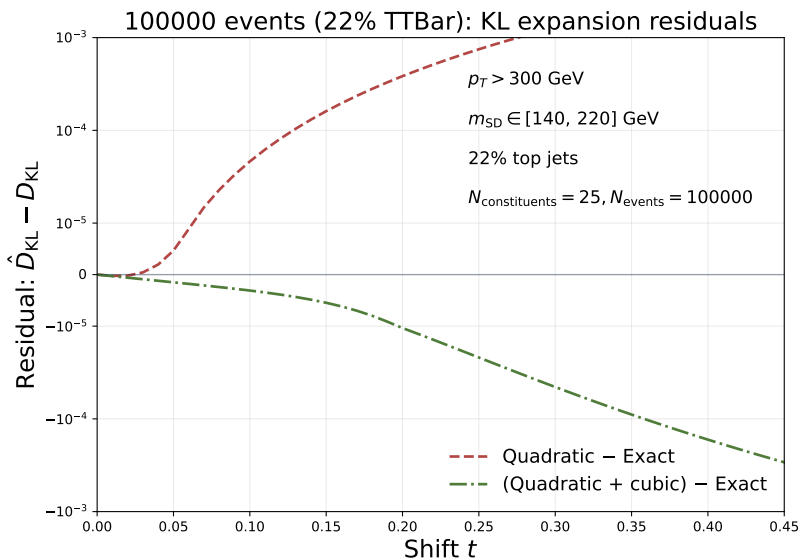


Figure 2. Residuals of the KL divergence along the feature direction in Eq. (6.5), for the quadratic Fisher approximation and the approximation including the cubic Fisher tensor, with respect to the exact KL expansion. The cubic term extends the range over which the local information model tracks the true divergence, leading to an approximately 30x reduction in error at the evaluation point $t = 0.15$.

t	exact KL	quadratic	quadratic+cubic	error reduction
0.15	0.023219	0.023380	0.023213	29.5x
0.30	0.092232	0.093518	0.092187	28.4x
0.45	0.206216	0.210416	0.205922	14.3x
0.60	0.364615	0.374073	0.363421	7.9x

Table 1. Selected points from the exact KL curve. “Error reduction” is the ratio of the quadratic truncation error to the quadratic-plus-cubic truncation error. The cubic term is most useful in the genuinely local regime and remains competitive up to moderate shifts.

6.2.2 Third Order Improves the Local Divergence Model

We standardise the feature vector and define the physically relevant local direction by the class-mean difference,

$$v \propto \mu_{\text{top}} - \mu_{\text{QCD}} \quad (6.5)$$

Equation (2.8) then gives the exact KL series along v . Figure 2 compares the exact divergence to its quadratic and quadratic-plus-cubic truncations.

Table 1 makes the same point numerically. The quadratic approximation systematically underestimates the divergence because the local feature distribution is positively skewed along v . Adding the cubic Fisher tensor repairs most of that bias for small and moderate shifts. This is the operational meaning of $\mathcal{I}^{(3)}$ in the present setting: it is the first controlled correction to the Fisher-matrix approximation of local distinguishability.

The three legs of the triality are already active in this minimal example. On the Fisher side, $\mathcal{I}^{(2)}$ and $\mathcal{I}^{(3)}$ are the coefficients of the local KL expansion, and Fig. 2 shows that the cubic term has direct operational meaning. On the correlator side, those same coefficients are estimated from the EEC/ECF basis as connected moments, so the skewed response along v is tied to a concrete radiation-pattern fluctuation. On the hypergraph side, the same $\mathcal{I}_{abc}^{(3)}$ will become the weight of a 3-hyperedge, allowing the dominant triplet to be read off combinatorially rather than only statistically.

6.2.3 The Hypergraph Sees Structure the Graph Smears Out

The pairwise Fisher graph is built from the standardised covariance matrix, while the order-three Fisher hypergraph is built from the distinct cubic weights $\mathcal{I}_{abc}^{(3)}$. The left panel of Figure 3 shows that with standardised features, the pairwise Fisher graph is saturated at second order: all large off-diagonal correlations sit close to ± 1 , so pairwise rankings cannot, in this simple study, tell us which observables carry genuinely independent higher-order information. To isolate the third-order content, we move to a whitened basis,

$$z_a = (I^{-1/2})_{ab} (\phi_b - \langle \phi_b \rangle_\theta), \quad (6.6)$$

where I_{ab} is the second-order Fisher (covariance) matrix. By construction this diagonalises the second-order Fisher tensor, $\tilde{I}_{ab} = \delta_{ab}$, so that all pairwise correlations among the new features vanish identically; any remaining structure in the third-order tensor \tilde{I}_{abc} must therefore be irreducible three-point information, decoupled from second-order content.

The whitened features are linear combinations of the original observables, but as shown in Table 2 the rotated features remain dominated by a single original feature in most cases: F_0 by the narrow-angle EEC, F_1 by the wide-angle EEC, and F_3 by $e_3^{(1)}$. The exception is F_2 , whose dominant contribution comes from the higher-order $e_3^{(1)}$ rather than from $e_2^{(1)}$, though these contributions are almost equal. Whitening forces all pairwise centralities to be trivially equal and isolates the genuinely third-order centralities, shown in the right panel of Figure 3. The feature whose dominant contribution comes from $e_3^{(1)}$ is promoted to the most central node of the hypergraph, in contrast to the pairwise graph where $e_2^{(1)}$ ranks highest. Physically, the hypergraph correctly recognises that $e_3^{(1)}$ participates in the strongest irreducible three-prong-sensitive triplet. Figure 4 confirms this: the strongest irreducible triplet hyperedge ties together the wide-angle EEC (dominant contributor to F_1) with the second- and third-order ECFs (the latter dominantly contributing to F_2 and F_3).

This compact study already closes the triangle. The cubic Fisher tensor is simultaneously a cubic correction to the local divergence, a connected three-observable cumulant of the correlator basis, and a 3-hyperedge weight used by the learning operator in Eq. (4.5). The same measured tensor therefore governs local inference, observable interpretation, and hypergraph propagation.

6.3 Substructure Probes and Fisher-Hypergraph Compression

The second application asks whether the triality is sensitive to the prong substructure of hadronically decaying massive particles. Resolving the number of hard prongs inside a

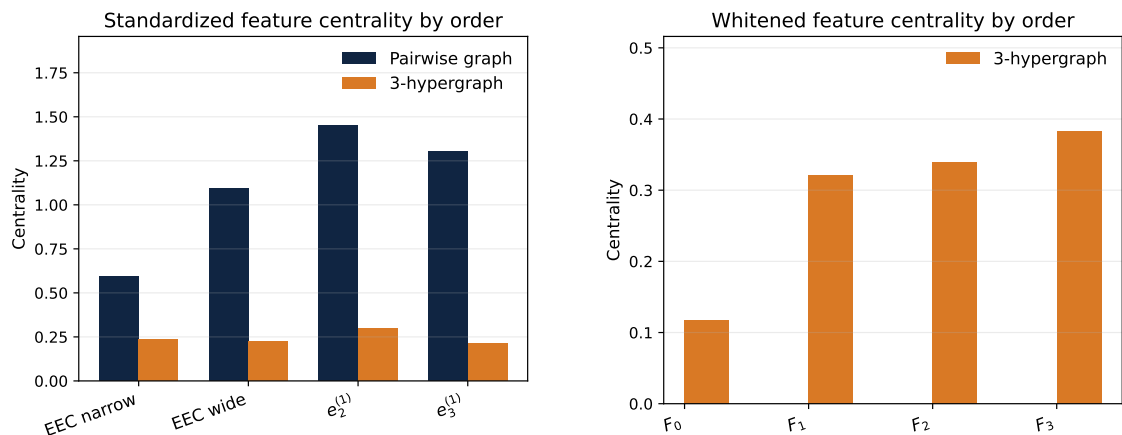


Figure 3. Node centralities before and after removing pairwise structure. Left: in the standardised observable basis, the pairwise Fisher graph is saturated, while the orange bars show the corresponding third-order Fisher-hypergraph centralities. Right: in the whitened basis, the second-order Fisher matrix is the identity, so the pairwise graph has no non-trivial off-diagonal edges and the blue centrality bars vanish; the orange bars are therefore the remaining irreducible third-order hypergraph centralities. The scales in both plots are chosen to highlight the relative values.

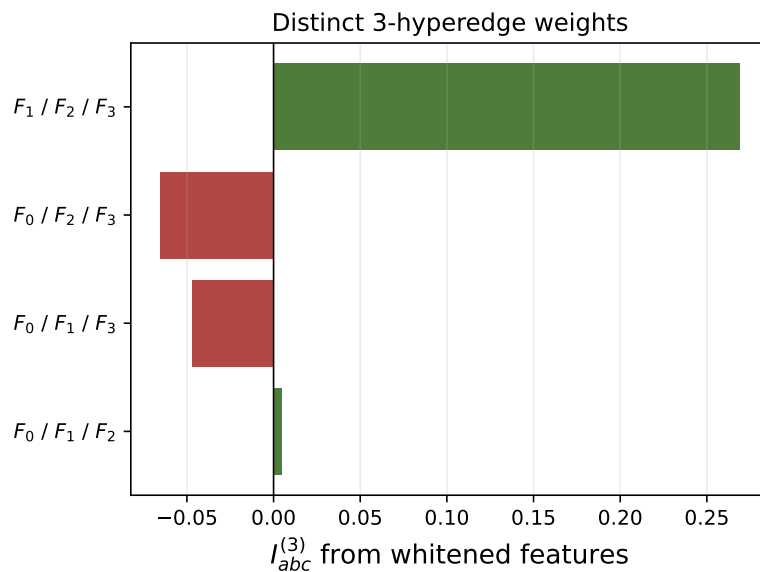


Figure 4. Distinct 3-hyperedge weights in the whitened basis F_0, \dots, F_3 defined in Table 2. Since whitening removes the non-trivial pairwise Fisher graph, the plotted coefficients are irreducible cubic Fisher weights. The dominant positive hyperedge $F_1/F_2/F_3$ is aligned, through the whitening loadings, with the wide-angle EEC and the second- and third-order ECF directions.

	F_0	F_1	F_2	F_3
$\text{EEC}_{\text{narrow}}$	0.8547	0.1147	0.0093	-0.0008
EEC_{wide}	0.1049	0.6708	-0.0637	0.0013
$e_2^{(1)}$	0.0252	-0.1878	0.4100	-0.0719
$e_3^{(1)}$	-0.0153	0.0266	-0.5169	0.9260

Table 2. Contributions to the whitened features F_i in terms of the original observables from equations (6.1) to (6.4), with the contributions normalised such that their absolute values sum to 1.

fat jet underlies essentially every modern boosted-object tagger, from dedicated substructure observables [23, 33] and shower deconstruction [34] to large-parameter deep-learning architectures [35], and it is therefore a natural setting in which to ask whether the higher-order geometric structure captured by the triality carries discriminating information beyond what the Fisher graph alone encodes. The strategy is to construct two benchmark samples with cleanly separated hard substructure, estimate the Fisher information matrix and the leading non-trivial higher-order tensor, and convert them respectively into a graph and a hypergraph of observable correlations. We then select a basis of observables designed to probe two- and three-body structure together with the intra-jet radiation pattern [3, 23, 24], and retain the smallest sub-basis that preserves the structural divergence. The simplest physically motivated instance of this construction, a two-prong $W \rightarrow q\bar{q}$ decay versus a three-prong $t \rightarrow bq\bar{q}$ decay [36, 37], is used to illustrate the procedure in the following sections.

6.3.1 Benchmark and Compression Rule

The benchmark dataset comprises 100,000 jets each for the W boson and hadronic top quark signal classes, clustered using the anti- k_t algorithm [38] with $R = 0.8$ (AK8). Jets are generated using the simulation setup described in Section 6.1, with a transverse momentum requirement of $p_T > 300$ GeV and mass-window selections centred on the respective W boson and top quark masses. This sample serves as a concrete, physically motivated reference point at which the local Fisher tensors are estimated; the subsequent analysis should therefore be understood as a local compression benchmark evaluated on a realistic jet topology rather than a generalised information-geometric study. The observable basis

is:

$$\phi_1 = \text{EEC}_1 = \sum_{i < j} z_i z_j \mathbf{1}_{0.02 < \Delta R_{ij} \leq 0.1}, \quad (6.7)$$

$$\phi_2 = \text{EEC}_2 = \sum_{i < j} z_i z_j \mathbf{1}_{0.1 < \Delta R_{ij} \leq 0.25}, \quad (6.8)$$

$$\phi_3 = \text{EEC}_3 = \sum_{i < j} z_i z_j \mathbf{1}_{0.25 < \Delta R_{ij} \leq 0.4}, \quad (6.9)$$

$$\phi_4 = \text{EEC}_4 = \sum_{i < j} z_i z_j \mathbf{1}_{0.40 < \Delta R_{ij} \leq 0.85}, \quad (6.10)$$

$$\phi_{5\dots 13} = e_2^{(\beta)}, \quad \beta \in [0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 2.0, 2.5, 3.0] \quad (6.11)$$

$$\phi_{14\dots 19} = e_3^{(\beta)}, \quad \beta \in [0.5, 1.0, 1.25, 1.5, 2.0, 2.5] \quad (6.12)$$

Rather than claim a realistic shower-level substructure tagger, we isolate the local compression problem itself by encoding the $W \rightarrow q\bar{q}$ versus $t \rightarrow b\bar{q}$ distinction through a normalised benchmark score direction defined using the class-mean differences between the two-prong and three-prong samples

$$\mathbf{u} = \mu_{\text{top}} - \mu_W, \quad (6.13)$$

where μ represents the mean of the 19-observable basis vector in the order written above. For a retained feature subset S , we define u_S by setting the discarded components of u to zero.

The pairwise graph baseline ranks observables by the quadratic node score

$$s_a^{(2)} = \left| u_a \mathcal{I}_{ab}^{(2)} u_b \right|, \quad (6.14)$$

while the multi-order Fisher hypergraph uses the additional cubic score

$$s_a^{(3)} = \left| u_a \mathcal{I}_{abc}^{(3)} u_b u_c \right|. \quad (6.15)$$

We normalise both score vectors to unit sum,

$$\hat{s}_a^{(r)} = \frac{s_a^{(r)}}{\sum_b s_b^{(r)}}, \quad r = 2, 3, \quad (6.16)$$

and define the direction-aligned multi-order node score

$$s_a^{\text{multi}} = \hat{s}_a^{(2)} + \lambda_3 \hat{s}_a^{(3)}, \quad \lambda_3 = 3. \quad (6.17)$$

In the present nineteen-observable benchmark, the two largest quadratic scores identify the pairwise-graph core

$$S_0^{(2)} = \{e_2^{(1.25)}, e_2^{(1)}\}, \quad (6.18)$$

while the multi-node score identifies the hypergraph core

$$S_0^{(\text{multi})} = \{e_3^{(1)}, e_3^{(1.25)}\}. \quad (6.19)$$

We use the node scores above for interpretation and for the left panel of Fig. 6, but the compression itself is performed by greedy forward selection. This makes the substructure sensitivity study structurally consistent with the basis-design benchmark in Section 6.4.

For each selector we denote its starting core by S_0 , with $S_0 \equiv S_0^{(2)}$ for the pairwise-graph baseline and $S_0 \equiv S_0^{(\text{multi})}$ for the Fisher hypergraph. For a retained subset $S \supseteq S_0$ and a candidate observable $a \notin S$, define the mean pairwise redundancy

$$\rho(a, S) = \frac{1}{|S|} \sum_{b \in S} |\mathcal{I}_{ab}^{(2)}|, \quad (6.20)$$

the redundancy-suppressed quadratic marginal gain

$$\Delta_2(a | S) = \frac{|u_a^2 \mathcal{I}_{aa}^{(2)}| + |2u_a \sum_{b \in S} \mathcal{I}_{ab}^{(2)} u_b|}{1 + \rho(a, S)}, \quad (6.21)$$

and the aligned cubic completion gain

$$\Delta_3(a | S) = \sum_{\{b,c\} \subseteq S} |u_a u_b u_c \mathcal{I}_{abc}^{(3)}|. \quad (6.22)$$

To compare the quadratic and cubic gains on the same scale, we introduce the fixed rescaling factor

$$r_3 = \frac{\max_{a \notin S_0} \Delta_2(a | S_0)}{\max_{a \notin S_0} \Delta_3(a | S_0)}. \quad (6.23)$$

The pairwise graph baseline then grows the basis by

$$a_{\text{graph}}^*(S) = \arg \max_{a \notin S} \Delta_2(a | S), \quad (6.24)$$

while the Fisher hypergraph uses

$$a_{\text{hyper}}^*(S) = \arg \max_{a \notin S} \left[\Delta_2(a | S) + \lambda_3 r_3 \Delta_3(a | S) \right], \quad \lambda_3 = 3. \quad (6.25)$$

With each selector starting from a core of two features, defined by the highest respective node scores, each subsequent iteration is a non-trivial selection, and cubic triplet completion becomes relevant.

6.3.2 Classification Performance Under Compression

To compare the compressed observable bases selected by the two strategies, we train a Boosted Decision Tree (BDT) classifier of maximum depth 3 and 200 estimators on each retained subset, using the classification of jet events as arising from either a hadronic top quark decay ($t \rightarrow bq\bar{q}$, three-pronged) or a W boson decay ($W \rightarrow q\bar{q}$, two-pronged) as the target task. Classification performance is quantified by the Area Under the ROC Curve (AUC) score, evaluated using a stratified k -fold cross-validation strategy with $k = 6$ folds, ensuring robust performance estimates against statistical fluctuations in the training sample. This protocol enables a direct, interpretable comparison of the substructure sensitivity retained by each compressed basis at each compression level.

The results are shown in Figure 5. The hypergraph selection yields higher AUC scores throughout the compressed and intermediate regimes, demonstrating that the multi-node scoring strategy retains superior sensitivity to jet substructure when the observable basis is restricted to a small number of features. The performance gap is most pronounced at intermediate compression levels, where the two selection strategies diverge most strongly in their chosen observables, and closes once almost the full basis is retained. This improvement is not an artifact of the classifier or the training procedure, but reflects the intrinsic informativeness of the selected observables with respect to the prong multiplicity of the jet.

The origin of this gain is identified in the left plot of Figure 6, which contrasts the per-observable scores assigned by the pairwise (quadratic) and hypergraph selectors. The hypergraph scoring correctly elevates higher-order Energy Correlation Functions (ECFs) to the top of the ranked list, observables that the pairwise selector systematically undervalues. These higher-order ECFs encode irreducible three-point correlation information among jet constituents, which is precisely the structure that distinguishes a three-pronged hadronic top decay from the two-pronged W topology. The pairwise selector, by construction sensitive only to marginal two-body correlations, cannot capture the collective discriminative value of such higher-order observables, and consequently creates a compressed basis that is structurally misaligned with the substructure signal. The hypergraph selector, by rewarding observables for their multi-point correlation content, recovers this information and translates it into measurable gains in classification performance.

The three legs of the triality are all used explicitly in this benchmark. On the Fisher side, $\mathcal{I}^{(2)}$ and $\mathcal{I}^{(3)}$ define the quadratic and cubic gains, and therefore the expected sensitivity to jet substructure. On the correlator side, those same coefficients are estimated from the EEC/ECF basis as connected fluctuations, so the selected structures can be read as concrete radiation patterns rather than as abstract scores. On the hypergraph side, $\mathcal{I}^{(2)}$ supplies the pairwise baseline while $\mathcal{I}^{(3)}$ supplies the 3-hyperedges whose completion changes the greedy selection. The compression gain, therefore, comes from a prong-sensitive pattern that is irreducibly triplet-valued.

6.3.3 Interpretable Prong-Sensitive Hyperedges

The compression gain is not a black-box effect. The right panel of Figure 6 shows the direction-aligned node scores together with the leading 3-hyperedges ranked by

$$\left| u_a u_b u_c \mathcal{I}_{abc}^{(3)} \right|. \quad (6.26)$$

The dominant triplet is $(e_3^{(1)}, e_3^{(1.5)}, e_3^{(2)})$, comprising three-point Energy Correlation Functions evaluated at angular exponents $\beta = 1, 1.5, 2$, followed by closely related triplets that couple two members of this family to lower-order two-point moments. This pattern carries a clear physical interpretation: the classification task is not driven by any single pairwise energy-flow observable, but by a collective signature of genuine three-pronginess. The $e_3^{(\beta)}$ family is, by construction, sensitive to three-body angular correlations among jet constituents, and its joint variation across multiple β values provides a multi-scale characterisation of the three-prong geometry that is irreducible to any combination of two-point

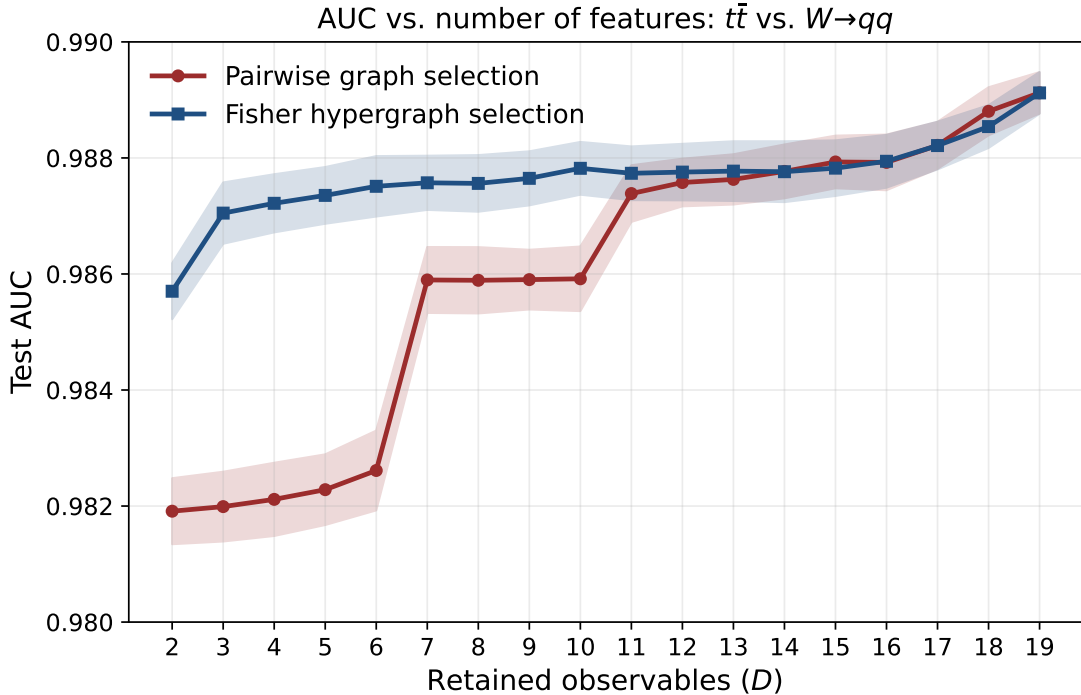


Figure 5. Classification performance for the $t\bar{t}$ versus $W \rightarrow q\bar{q}$ benchmark under greedy forward selection. The curves show mean test AUC as a function of the number of retained observables, with shaded bands indicating the spread across the cross-validation folds. The Fisher hypergraph selector performs better in the compressed regime, while the two selections coincide once almost the full basis is retained.

observables. Critically, this triplet structure is invisible to selection strategies based solely on the quadratic ($N = 2$) Fisher information matrix, whose score functions reward pairwise discriminative power and cannot resolve the collective gain arising from the simultaneous inclusion of correlated three-point observables. It is only when the selection score is defined using higher-order information tensors at order $N > 2$, as in the hypergraph construction, that the greedy selector correctly identifies this triplet as the dominant informative structure, and consequently assembles a compressed basis that is geometrically aligned with the three-prong structure driving the $t \rightarrow bq\bar{q}$ versus $W \rightarrow q\bar{q}$ classification.

This shows the practical use of the triality. The same tensor simultaneously quantifies local distinguishability, identifies the physically important correlator triplets, and prescribes the higher-order combinatorial structure on which a learning or compression algorithm should operate. While a more realistic analysis with more complex algorithms could be designed, this simple benchmark demonstrates how the full chain would work in practice. The minimal class-mixture study in Section 6.2 provides a complementary low-dimensional illustration.

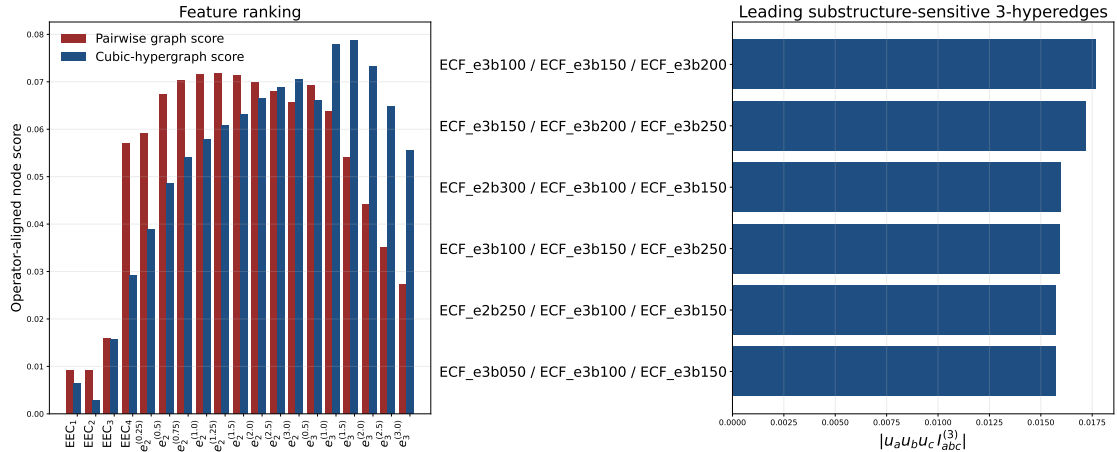


Figure 6. Left: direction-aligned feature scores for the pairwise graph and the multi-order Fisher hypergraph in the benchmark prong-sensitivity study. Right: the leading substructure-sensitive 3-hyperedges ranked by $|u_a u_b u_c \mathcal{I}_{abc}^{(3)}|$. The dominant hyperedge ties together the three three-point energy correlators.

6.4 Observable Discovery and Higher-Order Basis Design

A third application is to the design of the observable basis itself. In a realistic jet analysis one often begins with tens or hundreds of binned EECs together with a selected ECF/ECF basis, and the main practical question is not how to fit one fixed set of observables but which observables should survive into the final analysis. This is closely related to information-theoretic approaches to observable design [39]. The triality provides a principled answer because it attaches a local statistical meaning to every connected correlator pattern. Given a task direction Δ^a in feature space, one may rank hyperedges by the aligned score

$$S_e^{(n)} = \left| \Delta^{a_1} \dots \Delta^{a_n} \mathcal{I}_{a_1 \dots a_n}^{(n)} \right|, \quad e = \{a_1, \dots, a_n\}. \quad (6.27)$$

Because the same coefficient is also a connected correlator cumulant, a large value of $S_e^{(n)}$ identifies not merely a large covariance but an irreducible multipoint radiation pattern that matters for the task at hand. This is precisely the information that is lost if one ranks observables by pairwise correlations alone.

To make this concrete we now consider a higher-dimensional benchmark. The reference sample is the boosted top hadronic decay jet (three-pronged) used in the substructure study of Section 6.3, but the basis-design question is now posed against a *family* of nearby deformations of the reference sample rather than a clear and easily resolvable difference in the jet substructure itself. We use three benchmark deformations with clear collider interpretations: an opening-angle shift of the hard prongs, an inter-prong bridge component and a soft fourth-prong contamination. The candidate observable library contains eighteen binned EECs spanning $0.02 < \Delta R \leq 0.85$, nine two-point ECFs, and six three-point ECFs,

$$\{\text{EEC}_1, \dots, \text{EEC}_{18}, e_2^{(0.25)}, e_2^{(0.5)}, \dots, e_2^{(3)}, e_3^{(0.5)}, \dots, e_3^{(3)}\}. \quad (6.28)$$

For each deformation $m = 1, \dots, N_\Delta$ with $N_\Delta = 3$, let $\Delta_{(m)}^a$ denote the standardised mean-shift direction relative to the reference sample. We construct three benchmark deformations of the reference $t\bar{t}$ distribution through phase-space selections on per-jet substructure features. For each jet, we compute three hard prong axes $\{a_1, a_2, a_3\}$ via exclusive k_T reclustering with $N_{\text{exclusive}} = 3$, and define $z_i = p_T^{a_i} / \sum_k p_T^{a_k}$, the pairwise opening angles ΔR_{ij} , and the N-subjettiness ratio τ_4/τ_3 . Each deformation k is realised by selecting the subset S_k of events lying in the upper (or lower) tail fraction $q \in (0, 1)$ of a deformation score s_k , as summarised in Table 3. The corresponding mean-shift direction in the 33-dimensional observable space is then

$$\Delta^{(k)} = \langle \theta \rangle_{S_k} - \langle \theta \rangle_{\text{ref}}, \quad (6.29)$$

where $\langle O \rangle_{\text{ref}}$ is computed over the remaining background sample. We use $q = 0.1$ as a reasonable choice.

Deformation	Score s_k	Tail
Opening-angle shift	ΔR_{12}	top q
Inter-prong bridge	$\sum_j p_T^j \mathbf{1}(R_{\text{col}} < \Delta R_{j,\text{axis}} < R_{\text{wide}})$	top q
Soft 4th-prong contamination	τ_4/τ_3	bottom q

Table 3. Deformation scores and tail selections used to construct the benchmark mean-shift directions. Default radii: $R_{\text{wide}} = 0.4$, $R_{\text{col}} = 0.15$.

A basis-design algorithm should suppress redundancy: once a retained subset already captures a given pairwise fluctuation pattern, adding another observable from the same pairwise ladder should produce only a small marginal gain. We therefore replace static top- k ranking by a greedy forward selection rule. For a retained subset S , let $\Delta_{(m),S}$ denote the direction obtained from $\Delta_{(m)}$ by setting the discarded components to zero, and define the retained quadratic Fisher response

$$Q(S) = \frac{1}{N_\Delta} \sum_{m=1}^{N_\Delta} \frac{\Delta_{(m),S}^a \mathcal{I}_{ab}^{(2)} \Delta_{(m),S}^b}{\Delta_{(m)}^a \mathcal{I}_{ab}^{(2)} \Delta_{(m)}^b}. \quad (6.30)$$

Because Eq. (6.30) is evaluated on the current subset rather than on isolated node scores, it penalises redundant pairwise picks automatically. The pairwise graph baseline therefore grows the basis by the marginal-gain rule

$$a_{\text{graph}}^*(S) = \arg \max_{a \notin S} \left[Q(S \cup \{a\}) - Q(S) \right]. \quad (6.31)$$

To track genuinely irreducible three-way structure, we complement $Q(S)$ by the aligned triplet coverage

$$C_3^{\text{align}}(S) = \frac{\sum_{m=1}^{N_\Delta} \sum_{\{i,j,k\} \subseteq S} \left| \Delta_{(m)}^i \Delta_{(m)}^j \Delta_{(m)}^k \mathcal{I}_{ijk}^{(3)} \right|}{\sum_{m=1}^{N_\Delta} \sum_{i < j < k} \left| \Delta_{(m)}^i \Delta_{(m)}^j \Delta_{(m)}^k \mathcal{I}_{ijk}^{(3)} \right|}. \quad (6.32)$$

The Fisher-hypergraph selector uses the same redundancy-suppressing greedy search, but augments the quadratic gain by the aligned cubic coverage,

$$a_{\text{hyper}}^*(S) = \arg \max_{a \notin S} \left[Q(S \cup \{a\}) - Q(S) + \lambda_3 (C_3^{\text{align}}(S \cup \{a\}) - C_3^{\text{align}}(S)) \right], \quad \lambda_3 = \frac{3}{2}. \quad (6.33)$$

The two selectors therefore differ only in whether the cubic triality information is allowed to influence the marginal gain.

The same three ingredients are therefore active throughout this basis-design problem. The Fisher tensors provide the objectives: $Q(S)$ measures retained quadratic response, while $C_3^{\text{align}}(S)$ measures retained rank-three response across the deformation family. Because these same tensors are also connected correlator cumulants, a large cubic marginal gain identifies a genuine multi-observable radiation pattern rather than a redundant refinement of an already covered pairwise ladder. The hypergraph language then turns those same cubic coefficients into triplets that the selector tries to close, which is why the final basis favors wide-angle and three-point structure over extra short-distance pairwise bins.

Because the aim of this subsection is basis design across several nearby tasks, we evaluate each compressed basis using the deformation directions, assessing the final information retention with the corresponding Fisher-aligned natural-parameter direction. For each deformation we measure the retained local third-order divergence fraction

$$R_{(3)}^{(m)}(S) = \frac{\frac{1}{2} \Theta_{(m),S}^\top \mathcal{I}_S^{(2)} \Theta_{(m),S} + \frac{1}{6} \mathcal{I}_{S,ijk}^{(3)} \Theta_{(m),S}^i \Theta_{(m),S}^j \Theta_{(m),S}^k}{\frac{1}{2} \Theta_{(m)}^\top \mathcal{I}^{(2)} \Theta_{(m)} + \frac{1}{6} \mathcal{I}_{ijk}^{(3)} \Theta_{(m)}^i \Theta_{(m)}^j \Theta_{(m)}^k}, \quad (6.34)$$

where:

$$\Theta_{(m),S} = \mathcal{I}_S^{(2)-1} \Delta_S, \quad \Theta_{(m)} = \mathcal{I}^{(2)-1} \Delta_{(m)}.$$

We summarise this family of tasks by its mean and median,

$$\bar{R}_{(3)}(S) = \frac{1}{N_\Delta} \sum_{m=1}^{N_\Delta} R_{(3)}^{(m)}(S), \quad R_{(3)}^{\text{med}}(S) = \text{median}_m R_{(3)}^{(m)}(S). \quad (6.35)$$

Figure 7 shows the result. Once redundancy is handled at the selection stage, the pairwise graph and Fisher hypergraph no longer collapse onto nearly identical top- k lists. The two selectors coincide only for the first few obvious picks and then separate over the full range $k = 6$ –18. The hypergraph basis improves the exact local diagnostics at second and third order, and the aligned triplet coverage simultaneously throughout that regime. At twelve retained observables, the pairwise graph keeps $\bar{R}_{(3)} = 0.871$, $R_{(3)}^{\text{med}} = 0.899$, and $C_3^{\text{align}} = 0.449$, whereas the Fisher hypergraph keeps 0.937, 0.938, and 0.490, respectively. The correct interpretation is therefore that the observed gain is not only a change in ordering: but also that the hypergraph selector retains more of the local Fisher/KL information while also preserving more deformation-aligned cubic structure. This supports using the hypergraph selector as the preferred means of constructing compressed bases for an analysis with a large number of possible observables to choose from, correctly maximising the sensitivity to deformations of the reference sample space.

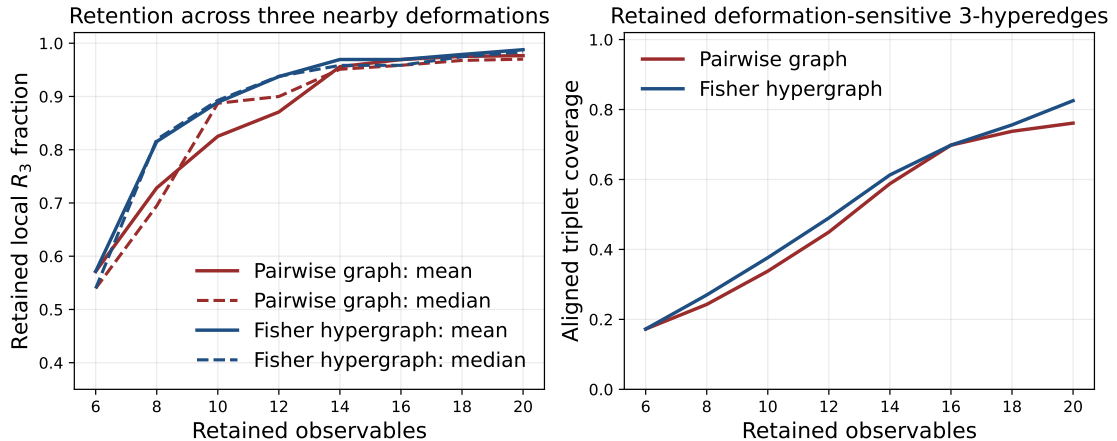


Figure 7. Greedy basis design across three nearby radiation-pattern deformations in a 33-observable EEC/ECF library. Left: mean and median retained local $R_{(3)}$ fractions, evaluated with Eq. (6.34). Right: aligned triplet coverage from Eq. (6.32). Once redundancy is suppressed by greedy marginal-gain selection, the Fisher hypergraph keeps more exact local discrimination power and visibly more deformation-sensitive rank-three structure over a broad range of feature budgets.

6.5 Physics-Informed Hypergraph Learning

The fourth application turns Eq. (4.5) into a concrete low-capacity learning benchmark. Graph-based collider architectures such as ParticleNet and energy-weighted message passing have already shown the effectiveness of physics-aware message passing on jet data [40, 41]. For the present purpose a nearby-deformation task is more informative than the QCD-versus-top separation of Section 6.2, because in the four-feature toy benchmark even simple linear classifiers saturate quickly. We therefore return to the 33-observable EEC/ECF library and the three-prong reference family of Section 6.4, and classify the reference sample against the soft fourth-prong deformation. This is precisely the regime where the graph and hypergraph priors differ most. In a representative reference training sample the leading pairwise Fisher graph is dominated by the adjacent e_2 ladder, whereas the leading 3-hyperedges are $(e_3^{(1.5)}, e_3^{(2)}, e_3^{(3)})$, $(e_3^{(1)}, e_3^{(2)}, e_3^{(3)})$, and $(e_3^{(1)}, e_3^{(1.5)}, e_3^{(3)})$.

For an event represented by the standardised node-feature vector $x \in \mathbb{R}^{33}$, we keep the order-specific propagators fixed and train only a one-layer readout,

$$h(x) = \sigma(xU_0 + U_2(\mathcal{P}_2x) + U_3(\mathcal{P}_3x)), \quad s(x) = \beta^\top \bar{h}(x), \quad \bar{h}(x) = \frac{1}{|V|} \sum_{a \in V} h_a(x), \quad (6.36)$$

with the fixed order- r propagation matrices

$$\mathcal{P}_r = \left(D_v^{(r)}\right)^{-1/2} B^{(r)} |W^{(r)}| \left(D_e^{(r)}\right)^{-1} \left(B^{(r)}\right)^\top \left(D_v^{(r)}\right)^{-1/2}. \quad (6.37)$$

U_2 and U_3 are simple multi-layer perceptrons (MLPs) with a single hidden layer and about 500 trainable parameters each. The pairwise graph baseline sets $U_3 = 0$. The triality-informed hypergraph keeps both orders, with the order-two layer built from the top 150

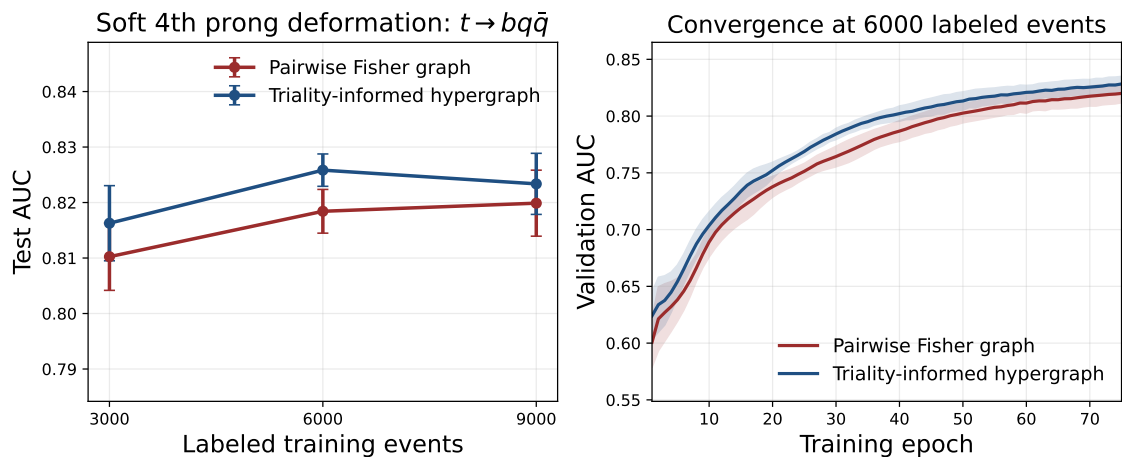


Figure 8. Physics-informed hypergraph learning for the soft fourth-prong deformation in the 33-observable EEC/ECF library. Left: mean test AUC versus the number of labelled training events for a pairwise Fisher graph and a triality-informed Fisher hypergraph; error bars show one standard deviation over five random splits. Right: mean validation AUC versus epoch at 6000 labelled training events, with shaded bands showing one standard deviation over the same splits. The faster convergence and higher validation score of the triality-informed hypergraph are consistent with retaining irreducible triplet structure that the graph baseline smears out.

entries of $|\mathcal{I}_{ij}^{(2)}|$ and the order-three layer from the top 200 entries of $|\mathcal{I}_{ijk}^{(3)}|$ estimated on the reference training sample.

The benchmark uses balanced training samples of 3000, 6000, and 9000 labelled events, together with fixed validation and test sets, and Figure 8 reports means and standard deviations over five random splits.

The result is modest in absolute size, because the classifier is intentionally tiny, but it is systematic. At 3000 labelled training events the Fisher hypergraph reaches a mean test AUC of 0.8163 ± 0.0068 , compared with 0.8102 ± 0.0061 for the pairwise graph. At 6000 labelled events the same ordering persists, with 0.8258 ± 0.0029 for the hypergraph versus 0.8184 ± 0.0039 . The right panel of Figure 8 shows the same effect dynamically: at fixed model size the triality-informed hypergraph converges to a higher validation AUC than the graph baseline. Therefore, once the signal depends on a subtle deformation of the radiation pattern rather than on an easy top-vs-QCD split, the Fisher 3-hyperedges become an informative inductive bias rather than a decorative add-on.

This benchmark uses all three legs of the triality in a direct way. The Fisher tensors supply the pairwise and triplet weights that define the two propagation channels. The correlator hierarchy explains why the informative 3-hyperedges sit in the high- β e_3 sector rather than in another redundant refinement of the pairwise e_2 ladder. The hypergraph layer then turns that information-geometric and correlator structure into a concrete learning prior. The gain in Figure 8 is therefore not a generic statement that “hypergraphs are better”; it is the practical benefit of using a hypergraph whose topology and weights have been fixed by the local Fisher–correlator data.

7 Summary and Conclusions

We have shown that, once a collider event family is represented in the natural exponential-family coordinates associated with a chosen finite basis of observables, higher Fisher tensors, connected correlator cumulants, and weighted hyperedges are the same local tensor written in three different languages. In this representation the log-partition function generates the full hierarchy, so that the coefficients governing the local Kullback–Leibler expansion also determine the connected multi-observable radiation structure and the weights of the corresponding hyperedges. This provides an exact Fisher–correlator–hypergraph triality for a finite basis of EECs, ECFs, and EFPs.

The scope of this statement is important. The identification is coordinate-specific: it is exact in the canonical chart generated by the chosen sufficient statistics. In a generic parametrisation the local information geometry remains well defined, but from cubic order onwards the KL coefficients contain additional Hessian contributions, as shown in Appendix A, and no longer coincide directly with connected observable cumulants. For applications this does not require the full collider problem to be globally exponential-family. It requires a local embedding around a reference sample in which the chosen observable basis captures the deformations relevant to the task.

The four applications show that the cubic level of this hierarchy is already useful in practice. In the QCD-versus-top local-KL study of Section 6.2, the cubic tensor improves the local KL approximation and isolates a dominant irreducible triplet coupling wide-angle EEC support to the two- and three-point ECFs. In the prong-substructure compression benchmark of Section 6.3, the hypergraph selector improves BDT classification of $W \rightarrow q\bar{q}$ versus $t \rightarrow bq\bar{q}$ jets in the compressed regime by elevating higher-order ECFs associated with the leading three-point correlator triplets.

In the three-deformation basis-design study of Section 6.4, the same cubic information improves compressed observable selection. At twelve observables the Fisher hypergraph reaches $\bar{R}_{(3)} = 0.937$, $R_{(3)}^{\text{med}} = 0.938$, and $C_3^{\text{align}} = 0.490$, compared with 0.871, 0.899, and 0.449 for the pairwise graph. In the physics-informed learning benchmark of Section 6.5, a one-layer classifier initialised from the pairwise Fisher graph and Fisher 3-hypergraph reaches a mean test AUC of 0.8258 ± 0.0029 at 6000 labelled events for the soft fourth-prong deformation, slightly above the pairwise graph baseline of 0.8184 ± 0.0039 at the $\sim 1.5\sigma$ level over five splits.

Taken together, these results show that the benefit of the triality is not merely that one can attach a hypergraph interpretation to a higher Fisher tensor. The main point is that the same measured or simulated correlator data can carry a common quantitative interpretation across local inference, physics interpretation, basis compression, and learning architecture design. Once a large cubic coefficient is identified, one learns that a specific three-observable radiation pattern is genuinely connected, that it contributes to local distinguishability beyond the Fisher-matrix approximation, and that the corresponding triplet should be retained or propagated jointly in a hypergraph representation. In this way the triality turns higher Fisher tensors from abstract information-geometric objects into quantities that guide which observables to keep, which multi-observable structures to interpret as

physically irreducible, and how to initialise higher-order architectures in a way tied directly to the underlying radiation pattern.

Several extensions are natural. For collider analyses, the triality suggests new strategies for analysis-basis construction, feature compression, and the interpretation of wide-angle and multi-prong radiation patterns in both precision and new-physics searches [42, 43]. For machine learning, it provides a physics-motivated recipe for building higher-order architectures whose topology and weights are initialised from measured or simulated local structure rather than from generic architectural heuristics [16, 40, 41]. For information geometry, collider observables offer a concrete setting in which higher Fisher tensors have an immediate operational interpretation as local divergence coefficients, connected cumulants, and combinatorial weights. On the algorithmic side, the main open questions are how best to truncate, sparsify, regularise, and propagate the higher-order hierarchy in larger models and larger observable libraries.

A Generic-Coordinate Form of the Cubic KL Coefficient

In the main text the triality is formulated in natural exponential-family coordinates, where derivatives of the log-partition function generate both the local KL coefficients and the connected cumulants of the chosen observable basis. Here, we show what changes when one works in a completely generic parametrisation of the same family of probability distributions.

The key point is that the Fisher matrix at quadratic order is special: it can still be written in a clean way in generic coordinates. Starting at cubic order, however, the coefficients of the local KL expansion are no longer determined only by moments of the score. They also contain terms involving second derivatives of the log-likelihood, which measure how the chosen coordinate chart itself bends through the statistical model. The exact equality used in the main text is therefore not a generic coordinate identity, but a property of the natural exponential-family representation.

Let $\ell_\theta(x) = \log p_\theta(x)$ and $s_a = \partial_a \ell_\theta$. Expanding the KL divergence around $\delta = 0$ gives

$$D_{\text{KL}}(p_\theta \| p_{\theta+\delta}) = - \sum_{n=1}^{\infty} \frac{1}{n!} \langle \partial_{a_1} \cdots \partial_{a_n} \ell_\theta \rangle_\theta \delta^{a_1} \cdots \delta^{a_n}. \quad (\text{A.1})$$

Equation (A.1) is completely general. It says that the local KL coefficients are expectation values of higher derivatives of the log-likelihood. The first term vanishes because the KL divergence is minimised at $\delta = 0$, and the quadratic term reduces to the usual Fisher information matrix after the standard normalisation identities are used. Those identities are

$$0 = \langle s_a \rangle_\theta, \quad (\text{A.2})$$

$$0 = \langle \partial_{ab} \ell_\theta + s_a s_b \rangle_\theta. \quad (\text{A.3})$$

The first relation is simply the statement that the score has zero mean. The second says that the expected Hessian of the log-likelihood is the negative Fisher matrix. Up to this

order, the expansion behaves in the familiar way. The first place where the coordinate dependence becomes non-trivial is the cubic coefficient. Differentiating the second identity once more yields

$$0 = \langle \partial_{abc}\ell_\theta + s_a\partial_{bc}\ell_\theta + s_b\partial_{ac}\ell_\theta + s_c\partial_{ab}\ell_\theta + s_as_bs_c \rangle_\theta. \quad (\text{A.4})$$

This identity shows that the third derivative of the log-likelihood is not determined by the score-cubic moment alone, but mixes with products of scores and Hessians. Therefore the cubic KL coefficient in generic coordinates is

$$\frac{1}{6} \langle s_as_bs_c + s_a\partial_{bc}\ell_\theta + s_b\partial_{ac}\ell_\theta + s_c\partial_{ab}\ell_\theta \rangle_\theta \delta^a\delta^b\delta^c. \quad (\text{A.5})$$

Equation (A.5) reveals that the cubic local KL coefficient generally contains two kinds of information at once. One part is the genuine third-order fluctuation of the score, which is the quantity that becomes the connected observable cumulant in the exponential-family setting. The other part comes from the curvature of the chosen parametrisation, encoded in the Hessian terms. Only in natural exponential coordinates does that second contribution reduce in such a way that the cubic KL coefficient matches the observable cumulant directly. This is why the triality theorem in the main text is stated in the exponential-family embedding rather than for a completely arbitrary parametrisation.

Equivalently, one may say that natural exponential coordinates are the chart in which the local statistical geometry is generated directly by the sufficient statistics. In that chart the cubic coefficient has a clean physical interpretation as a connected three-observable fluctuation and, at the same time, as a hyperedge weight. In a generic chart the same local distinguishability is still present, but the decomposition into physically interpretable cumulants and purely coordinate-dependent terms is mixed. This is the precise sense in which the triality is exact only in the exponential-family representation.

Acknowledgments

The authors acknowledge the support of Schmidt Sciences, the Alexander von Humboldt Foundation, and IPPP Durham, as well as the usage of computing resources at the Institute of Experimental Particle Physics (ETP), KIT.

References

- [1] C. L. Basham, L. S. Brown, S. D. Ellis, and S. T. Love. Energy Correlations in electron-Positron Annihilation in Quantum Chromodynamics: Asymptotically Free Perturbation Theory. *Phys. Rev. D*, 19:2018, 1979. doi: 10.1103/PhysRevD.19.2018.
- [2] Andrew J. Larkoski, Gavin P. Salam, and Jesse Thaler. Energy Correlation Functions for Jet Substructure. *JHEP*, 06:108, 2013. doi: 10.1007/JHEP06(2013)108.
- [3] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler. Energy flow polynomials: A complete linear basis for jet substructure. *JHEP*, 04:013, 2018. doi: 10.1007/JHEP04(2018)013.

- [4] Diego M. Hofman and Juan Maldacena. Conformal collider physics: Energy and charge correlations. *JHEP*, 05:012, 2008. doi: 10.1088/1126-6708/2008/05/012.
- [5] Ian Moulton and Hua Xing Zhu. Simplicity from Recoil: The Three-Loop Soft Function and Factorization for the Energy-Energy Correlation. *JHEP*, 08:160, 2018. doi: 10.1007/JHEP08(2018)160.
- [6] Claude Berge. *Hypergraphs: Combinatorics of Finite Sets*, volume 45 of *North-Holland Mathematical Library*. North-Holland, 1989.
- [7] Alain Bretto. *Hypergraph Theory: An Introduction*. Mathematical Engineering. Springer, 2013. doi: 10.1007/978-3-319-00080-0.
- [8] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in Neural Information Processing Systems 19*, 2007.
- [9] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. *CoRR*, abs/1809.09401, 2018. URL <http://arxiv.org/abs/1809.09401>.
- [10] Callum Birch-Sykes, Brian Le, Yvonne Peters, Ethan Simpson, and Zihan Zhang. Reconstructing short-lived particles using hypergraph representation learning. *Phys. Rev. D*, 111(3):032004, 2025. doi: 10.1103/PhysRevD.111.032004.
- [11] C. Radhakrishna Rao. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.*, 37:81–91, 1945.
- [12] Shinto Eguchi. A differential geometric approach to statistical inference on the basis of contrast functionals. *Hiroshima Mathematical Journal*, 15:341–391, 1985. doi: 10.32917/hmj/1206130775.
- [13] Shinto Eguchi. Geometry of minimum contrast. *Hiroshima Mathematical Journal*, 22(3): 631–647, 1992. doi: 10.32917/hmj/1206128508.
- [14] Shun-ichi Amari. *Differential-Geometrical Methods in Statistics*, volume 28 of *Lecture Notes in Statistics*. Springer, 1985. doi: 10.1007/978-1-4612-5056-2.
- [15] Shun-ichi Amari and Hiroshi Nagaoka. *Methods of Information Geometry*, volume 191 of *Translations of Mathematical Monographs*. American Mathematical Society and Oxford University Press, 2000.
- [16] Shun-ichi Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10(2): 251–276, 1998. doi: 10.1162/089976698300017746.
- [17] Nihat Ay, Jürgen Jost, Hông Vân Lê, and Lorenz Schwachhöfer. *Information Geometry*, volume 64 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge*. Springer, 2017. doi: 10.1007/978-3-319-56478-4.
- [18] Peter McCullagh. *Tensor Methods in Statistics*. Chapman and Hall, 1987.
- [19] Ole E. Barndorff-Nielsen. *Information and Exponential Families in Statistical Theory*. Wiley Series in Probability and Mathematical Statistics. Wiley, 1978.
- [20] N. N. Cencov. *Statistical Decision Rules and Optimal Inference*, volume 53 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 1982. English translation of the 1972 Russian original.
- [21] Lance J. Dixon, Ian Moulton, and Hua Xing Zhu. Collinear limit of the energy-energy correlator. *Phys. Rev. D*, 100(1):014009, 2019. doi: 10.1103/PhysRevD.100.014009.

- [22] Hao Chen, Ming-Xing Luo, Ian Moulton, Tong-Zhi Yang, Xiaoyuan Zhang, and Hua Xing Zhu. Three point energy correlators in the collinear limit: symmetries, dualities and analytic results. *JHEP*, 08(08):028, 2020. doi: 10.1007/JHEP08(2020)028.
- [23] Jesse Thaler and Ken Van Tilburg. Identifying Boosted Objects with N-subjettiness. *JHEP*, 03:015, 2011. doi: 10.1007/JHEP03(2011)015.
- [24] Andrew J. Larkoski, Ian Moulton, and Benjamin Nachman. Jet Substructure at the Large Hadron Collider: A Review of Recent Advances in Theory and Machine Learning. *Phys. Rept.*, 841:1–63, 2020. doi: 10.1016/j.physrep.2019.11.001.
- [25] Patrick T. Komiske, Eric M. Metodiev, and Jesse Thaler. Metric Space of Collider Events. *Phys. Rev. Lett.*, 123(4):041801, 2019. doi: 10.1103/PhysRevLett.123.041801.
- [26] Partha Konar, Vishal S. Ngairangbam, and Michael Spannowsky. Hypergraphs in LHC phenomenology — the next frontier of IRC-safe feature extraction. *JHEP*, 01:113, 2024. doi: 10.1007/JHEP01(2024)113.
- [27] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, O. Mattelaer, H.-S. Shao, T. Stelzer, P. Torrielli, and M. Zaro. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 07:079, 2014. doi: 10.1007/JHEP07(2014)079.
- [28] Richard D. Ball et al. Parton distributions from high-precision collider data. *Eur. Phys. J. C*, 77(10):663, 2017. doi: 10.1140/epjc/s10052-017-5199-5.
- [29] Valerio Bertone, Stefano Carrazza, Nathan P. Hartland, and Juan Rojo. Illuminating the photon content of the proton within a global PDF analysis. *SciPost Phys.*, 5(1):008, 2018. doi: 10.21468/SciPostPhys.5.1.008.
- [30] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to PYTHIA 8.2. *Comput. Phys. Commun.*, 191:159–177, 2015. doi: 10.1016/j.cpc.2015.01.024.
- [31] Albert M Sirunyan et al. Extraction and validation of a new set of CMS PYTHIA8 tunes from underlying-event measurements. *Eur. Phys. J. C*, 80(1):4, 2020. doi: 10.1140/epjc/s10052-019-7499-4.
- [32] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. DELPHES 3: a modular framework for fast simulation of a generic collider experiment. *JHEP*, 02:057, 2014. doi: 10.1007/JHEP02(2014)057.
- [33] Jonathan M. Butterworth, Adam R. Davison, Mathieu Rubin, and Gavin P. Salam. Jet substructure as a new Higgs search channel at the LHC. *Phys. Rev. Lett.*, 100:242001, 2008. doi: 10.1103/PhysRevLett.100.242001.
- [34] Davison E. Soper and Michael Spannowsky. Finding physics signals with shower deconstruction. *Phys. Rev. D*, 84:074002, 2011. doi: 10.1103/PhysRevD.84.074002.
- [35] Gregor Kasieczka et al. The Machine Learning landscape of top taggers. *SciPost Phys.*, 7:014, 2019. doi: 10.21468/SciPostPhys.7.1.014.
- [36] David E. Kaplan, Keith Rehermann, Matthew D. Schwartz, and Brock Tweedie. Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks. *Phys. Rev. Lett.*, 101:142001, 2008. doi: 10.1103/PhysRevLett.101.142001.

- [37] Roman Kogler et al. Jet Substructure at the Large Hadron Collider: Experimental Review. *Rev. Mod. Phys.*, 91(4):045003, 2019. doi: 10.1103/RevModPhys.91.045003.
- [38] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti- k_t jet clustering algorithm. *JHEP*, 04:063, 2008. doi: 10.1088/1126-6708/2008/04/063.
- [39] Andrew J. Larkoski, Jesse Thaler, and Wouter J. Waalewijn. Gaining (Mutual) Information about Quark/Gluon Discrimination. *JHEP*, 11:129, 2014. doi: 10.1007/JHEP11(2014)129.
- [40] Huilin Qu and Loukas Gouskos. ParticleNet: Jet Tagging via Particle Clouds. *Phys. Rev. D*, 101(5):056019, 2020. doi: 10.1103/PhysRevD.101.056019.
- [41] Partha Konar, Vishal S. Ngairangbam, and Michael Spannowsky. Energy-weighted message passing: an infra-red and collinear safe graph neural network algorithm. *JHEP*, 02:060, 2022. doi: 10.1007/JHEP02(2022)060.
- [42] Benjamin Nachman. Anomaly Detection for Physics Analysis and Less than Supervised Learning. 10 2020.
- [43] Andrew Blance, Michael Spannowsky, and Philip Waite. Adversarially-trained autoencoders for robust unsupervised new physics searches. *JHEP*, 10:047, 2019. doi: 10.1007/JHEP10(2019)047.