Lorentz-Equivariant Geometric Algebra Transformers for Top-Quark Event Reconstruction

Abasov E., Dudko L., Iudin E., Markina A.,

Perfilov M., Volkov P., Vorotnikov G., Zaborenko A.

Skobeltsyn Institute of Nuclear Physics, Lomonosov Moscow State University

This study was conducted within the scientific program of the National Center for Physics and Mathematics, section #5 «Particle Physics and Cosmology». Stage 2026-2027

Introduction

- Modern HEP events contain many final-state particles, and their physical meaning is encoded in relations between four-momenta.
- Standard neural networks can use these momenta directly, but they usually have to learn Lorentz symmetry from data.
- Geometric algebra gives a single mathematical language for scalars, vectors, planes, volumes, and oriented four-dimensional structures.
- This makes GA a natural candidate for building neural networks with physics symmetries built into the architecture.

Geometric algebra

Standard HEP kinematics uses three separate operations: scalar product pq, antisymmetric tensor $p^\mu q^\nu - p^\nu q^\mu$, and Levi-Civita contraction $\epsilon_{\mu\nu\rho\theta}$.Geometric algebra (GA) unifies all three into one associative, invertible product.

- GA operates on multivectors
- Multivectors can be decomposed into grades (analogous to tensor rank)
- Main operation between multivectors – geometric product, which can be decomposed into grade-decreasing (dot) and grade-increasing (wedge) products

$$pq = p \cdot q + p \wedge q$$

$$p \cdot q = \tfrac{1}{2}(pq + qp), \qquad p \wedge q = \tfrac{1}{2}(pq - qp)$$

Dot product – lower grade Wedge product – higher grade

Example in HEP: $W \to l\nu$ decay:

$$p_\ell \cdot p_\nu = \frac{m_W^2}{2} \quad \text{(invariant mass)}, \qquad p_\ell \wedge p_\nu = \text{6-component bivector (decay plane)}$$

GA can be defined over any real vector space with a bilinear metric. In HEP Cl(1,3) is used (Minkowski space)

- Basis vectors γ_i

$$\gamma_\mu\gamma_\nu + \gamma_\nu\gamma_\mu = 2\,\eta_{\mu\nu}, \qquad \eta = \mathrm{diag}(+1,-1,-1,-1) \qquad \gamma_0^2 = +1, \qquad \gamma_i^2 = -1 \quad (i = 1,2,3) \qquad I = \gamma_0\gamma_1\gamma_2\gamma_3, \qquad I^2 = -1$$

Geometric algebra

- Every grade can be universally transformed with a single multivector – rotor – using a sandwich product

$$X \longmapsto R\,X\,\widetilde{R}, \qquad R\widetilde{R} = 1$$

$$R = \exp\left(\frac{B}{2}\right), \qquad B \in \langle \mathrm{Cl}(1,3)\rangle_2$$

$$B^2 = -1 \;\Rightarrow\; \text{spatial rotation}, \qquad B^2 = +1 \;\Rightarrow\; \text{Lorentz boost}$$

Examples:

2D rotation

$$R = e^{\theta \mathbf{e}_1\mathbf{e}_2/2} = \cos\frac{\theta}{2} + \sin\frac{\theta}{2}\,\mathbf{e}_1\mathbf{e}_2$$

$$\mathbf{v}' = R\,\mathbf{v}\,\widetilde{R}$$

Lorentz boost

$$R_{\text{boost}} = e^{\alpha\,\gamma_3\gamma_0/2} = \cosh\frac{\alpha}{2} + \sinh\frac{\alpha}{2}\,\gamma_3\gamma_0$$

Since every grade transforms with the same rotor – neural network using GA can be equivariant

$$\Phi\big(\Lambda \cdot x\big) = \Lambda \cdot \Phi(x) \longleftrightarrow \Phi\big(R\,X\,\widetilde{R}\big) = R\,\Phi(X)\,\widetilde{R}$$

Geometric algebra in HEP

- Event grade decomposition

Grade k	$\binom{4}{k}$	**Name**	**HEP meaning**
0	1	scalar	masses m_i^2, Mandelstam $p_i \cdot p_j$
1	4	vector	4-momentum p^μ
2	6	bivector	decay planes $p_i \wedge p_j$, Lorentz generators
3	4	trivector	3-body oriented volumes $p_i \wedge p_j \wedge p_k$
4	1	pseudoscalar	CP-odd sign, chirality

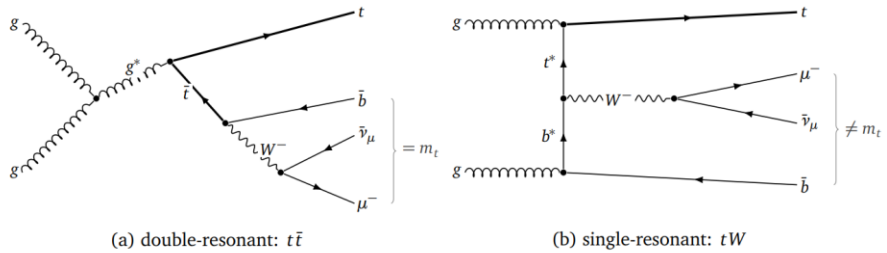
- Cayley–Menger collapse - Any Lorentz-invariant scalar built from a higher-grade blade collapses to a polynomial in $\{p_i \cdot p_j, m_i^2\}$ – GA does not generate new Lorentz-invariant scalars beyond what you already have
- Main benefit of GA is not in new inputs (the only exception is CP-odd pseudoscalar), but in forcing Lorentz-equivariance by construction
- Additional example – top-quark spin correlations
  - Standard approach: (1) boost 4-momenta to $\bar{t}t$ rest frame, (2) boost to respective top/antitop rest frame, (3) project lepton directions onto spin-quantization axis – complex operation needs to be learned
  - GA approach – one operation

$$B_t = p_b \wedge p_\ell \wedge p_\nu, \qquad B_{\bar{t}} = p_{\bar{b}} \wedge p_{\bar{\ell}} \wedge p_{\bar{\nu}}$$

$$\cos \Delta\phi_{\text{spin}} \sim \frac{\langle B_t \widetilde{B}_{\bar{t}} \rangle_0}{|B_t|\,|B_{\bar{t}}|}$$

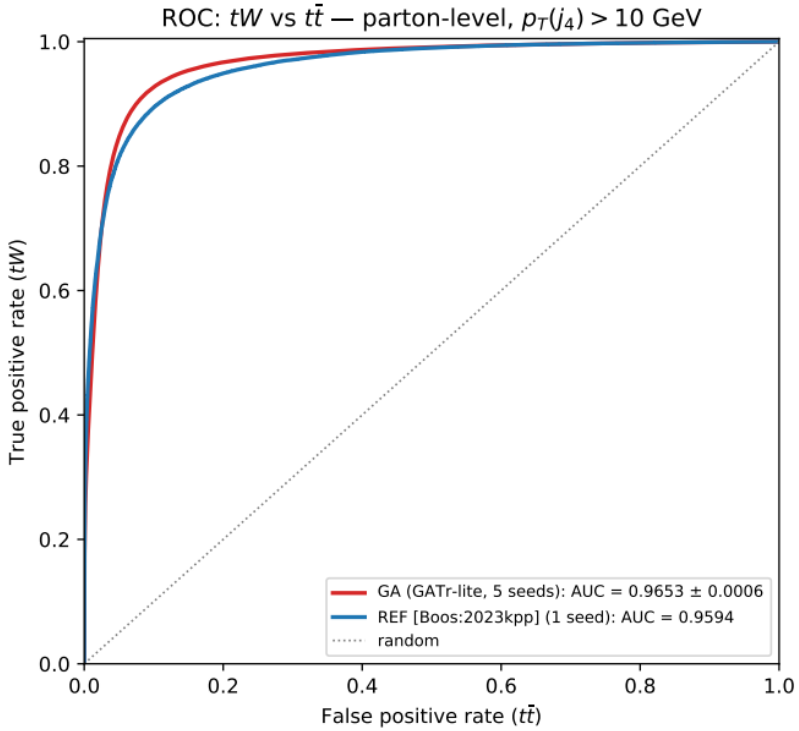
LGaTr proof-of-concept: $\bar{t}t$ / tWb

- LGaTr was tested on a simpler task - $\bar{t}t$ / tWb separation



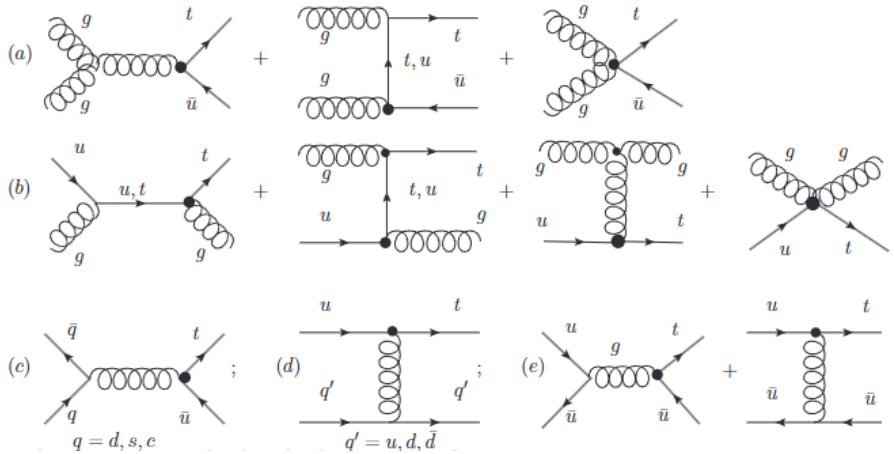
(a) double-resonant: $t\bar{t}$ (b) single-resonant: tW

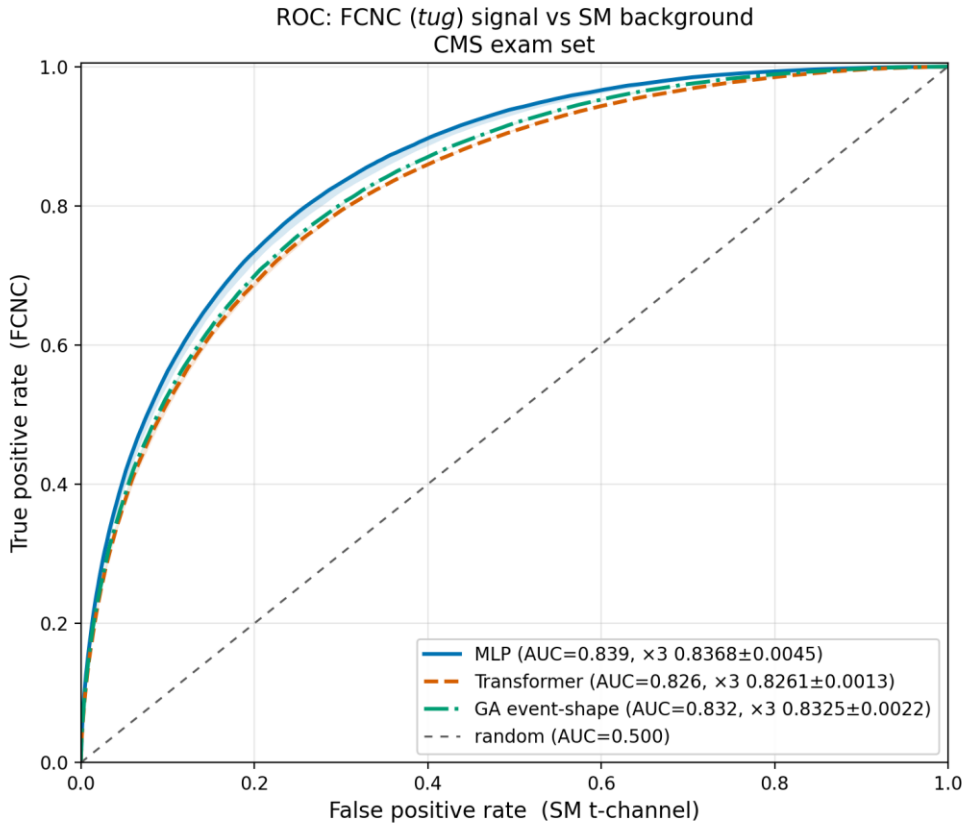
- Event multivector
  - Grade-0 – one-hot encoding of the object type $(\mu^-, \overline{\nu_\mu}, u, \bar{d}, b\ \bar{b})$
  - Grade-1 – 4-momentum
  - Grade-2/3 (new) – wedge products of 4-momenta,
- Baseline – MLP on hand-crafted scalar features
- Results show LGaTr beating the baseline with a significant margin



LGaTr proof-of-concept: FCNC

- Second test - real CMS analysis - separation of tug FCNC from SM events in single top-quark production





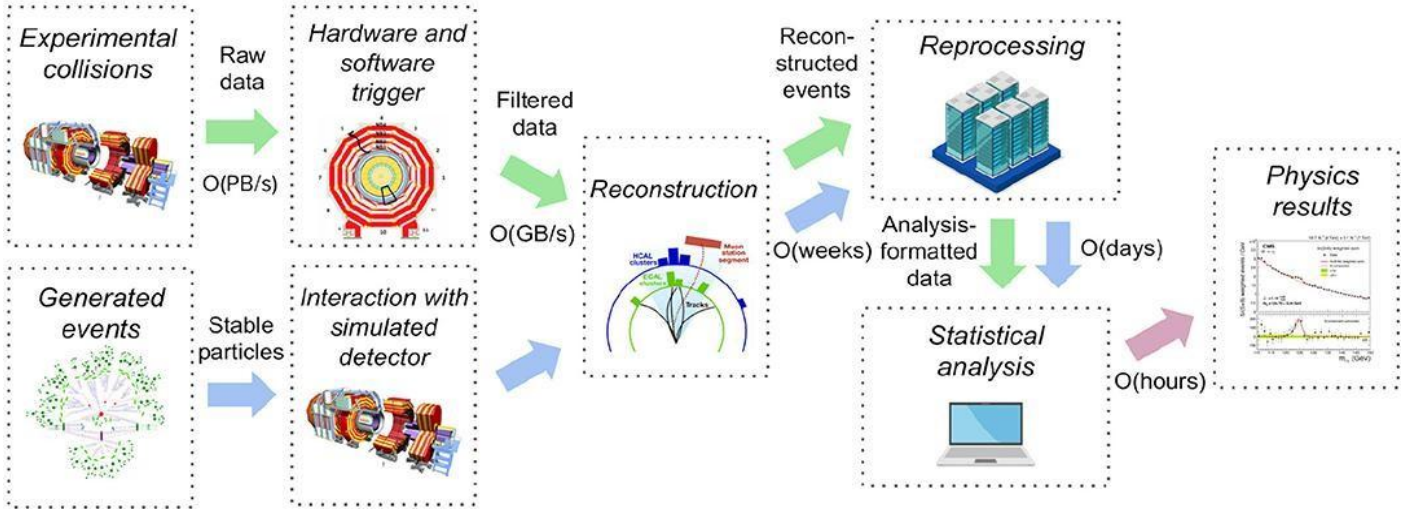
- In this task hand-crafted features for MLP input prove to be superior to raw 4-momenta being given to LGaTr
- For comparison, standard transformer-encoder was also trained, using the same inputs as LGaTr, demonstrating the advantage of GA approach on the same input space

Universal foundational model

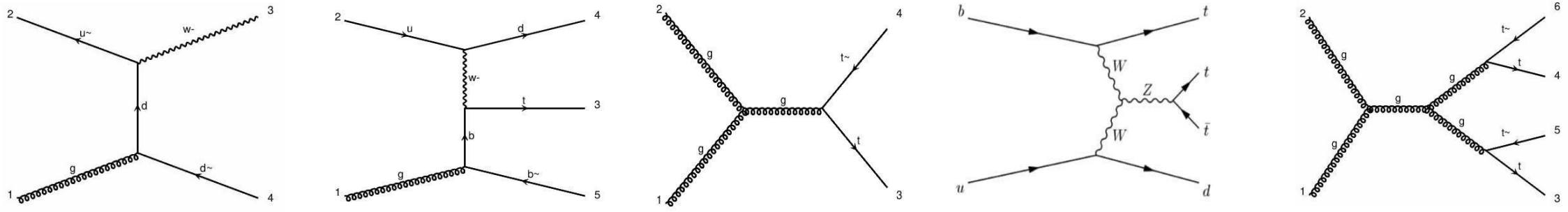
- Modern experiments in High Energy Physics generate enormous amounts of data
- Classical ML models require complex manual customisation for each task

- CV and NLP domains have seen the development of “foundational models”, which are pretrained to infer basic rules of the domain and then finetuned for specific downstream tasks

- Our goal is to apply the “foundational model” approach to HEP

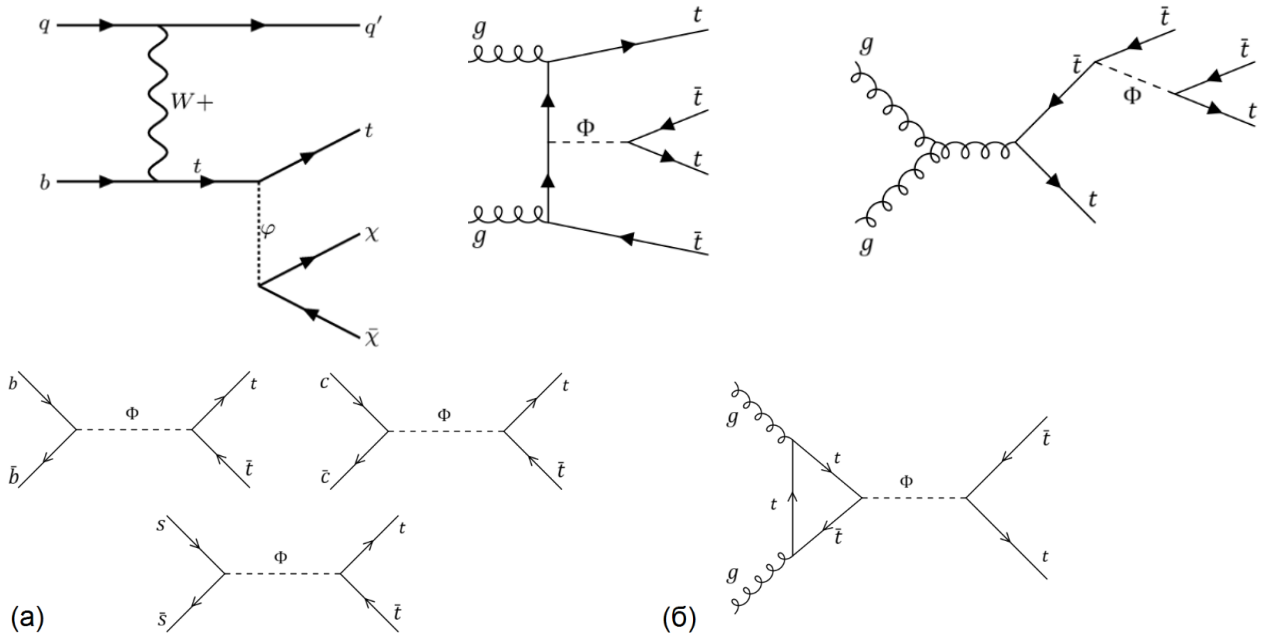


Data



SM processes

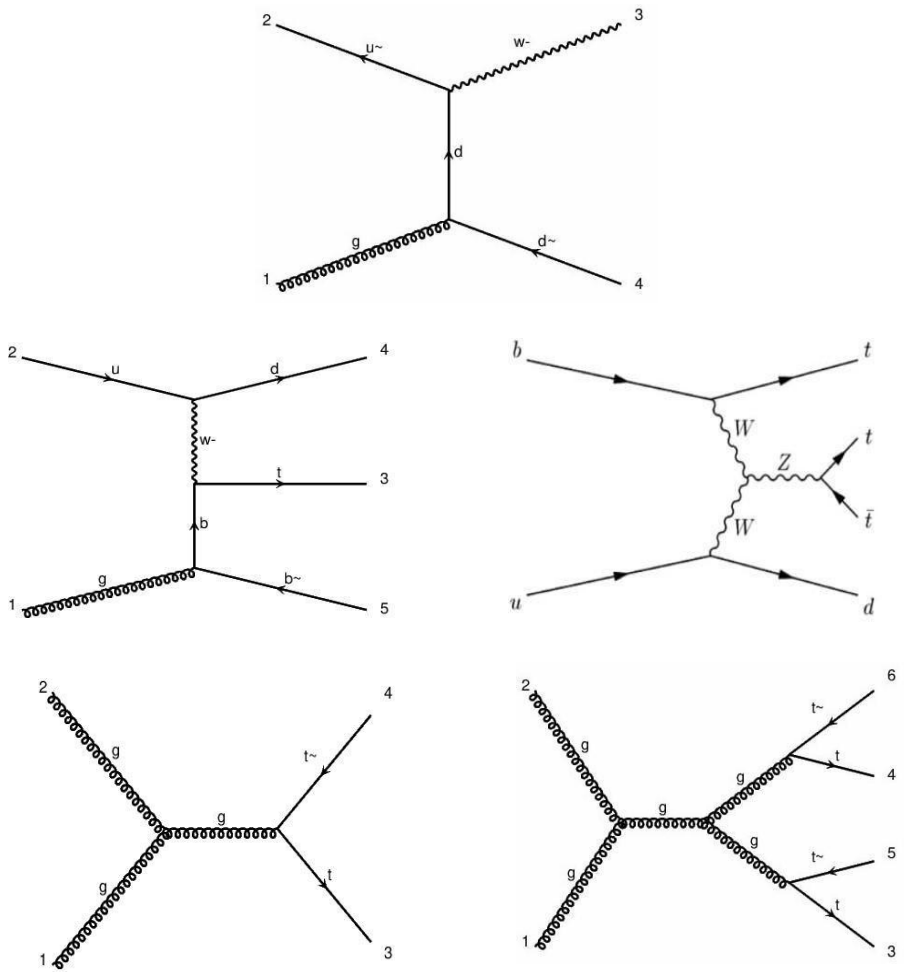
- 5 process groups (0, 1, 2, 3, 4 top-quarks) in order to cover wide range of parameters and final states
- Generators: MadGraph5 and CompHEP
- ~8M events in total



BSM processes

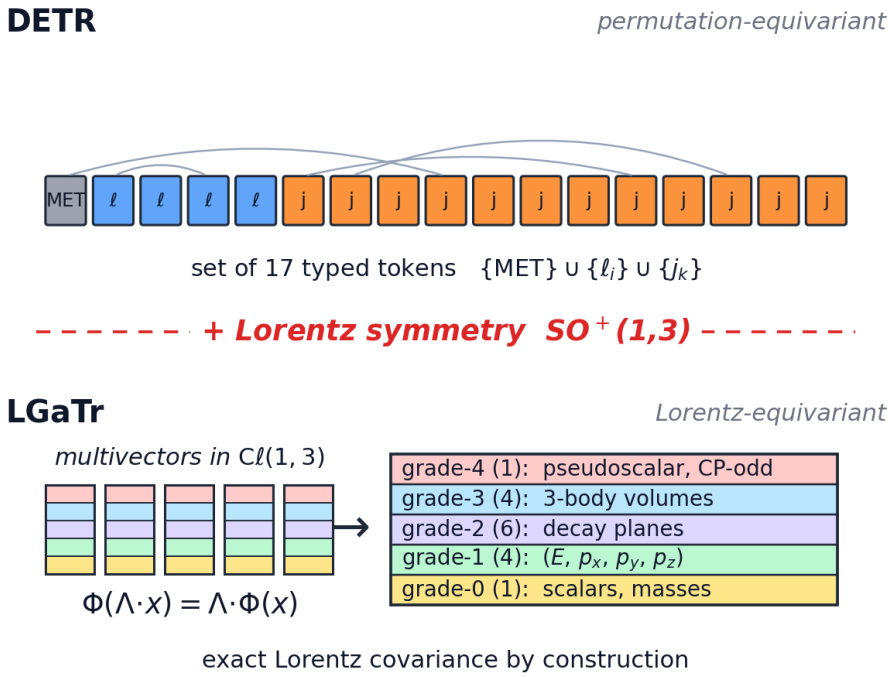
Universal foundational model

- 5 process groups (0, 1, 2, 3, 4 top-quarks) in order to cover wide range of parameters and final states
- Inputs – detector-level variables
- Outputs – properties of parton-level particles
  - p_T, η, φ, E – target variables for the regression task
  - Flavour/charge categorical tags – targets for classification task
  - "High-level" variables (M_tW, H_t, pairwise variables)
  - Decay chain reconstruction (assignment task) – which jets/leptons originated from a given t/W



Models

1. DEtection Transformer (DETR) (from the A.Zaborenko talk) – *set + decay-tree prior*
   - Event → 17 tokens {MET, ℓ, j}; attention is permutation-equivariant within each type
   - Lorentz symmetry still only learned
2. Lorentz Geometric Algebra Transformer (LGaTr) - *Lorentz-equivariant prior*
   - Each token = multivector in Cℓ(1,3)
   - Encoder is Lorentz-equivariant by construction
   - Invariant masses and angles are exact, not learned



LGaTr: concepts

Data: each detector token → multivector in Cl(1,3)

- Grade-1 channel: raw (E, p_x, p_y, p_z) / E_scale, E_scale = 100 GeV
- Grade-0 channels: b-tag, lepton flavour, particle type (5 scalar channels)
- Potentially grade-2 pairwise features

Architecture: GATr encoder blocks:

- Equivariant linear maps: mix multivector channels, preserve grade structure

$$Y^{(c')} = \sum_{c=1}^{C} W_{c'c}\, X^{(c)}, \qquad W_{c'c} \in \mathbb{R}$$

- Geometric product layer: channel-wise pairwise products

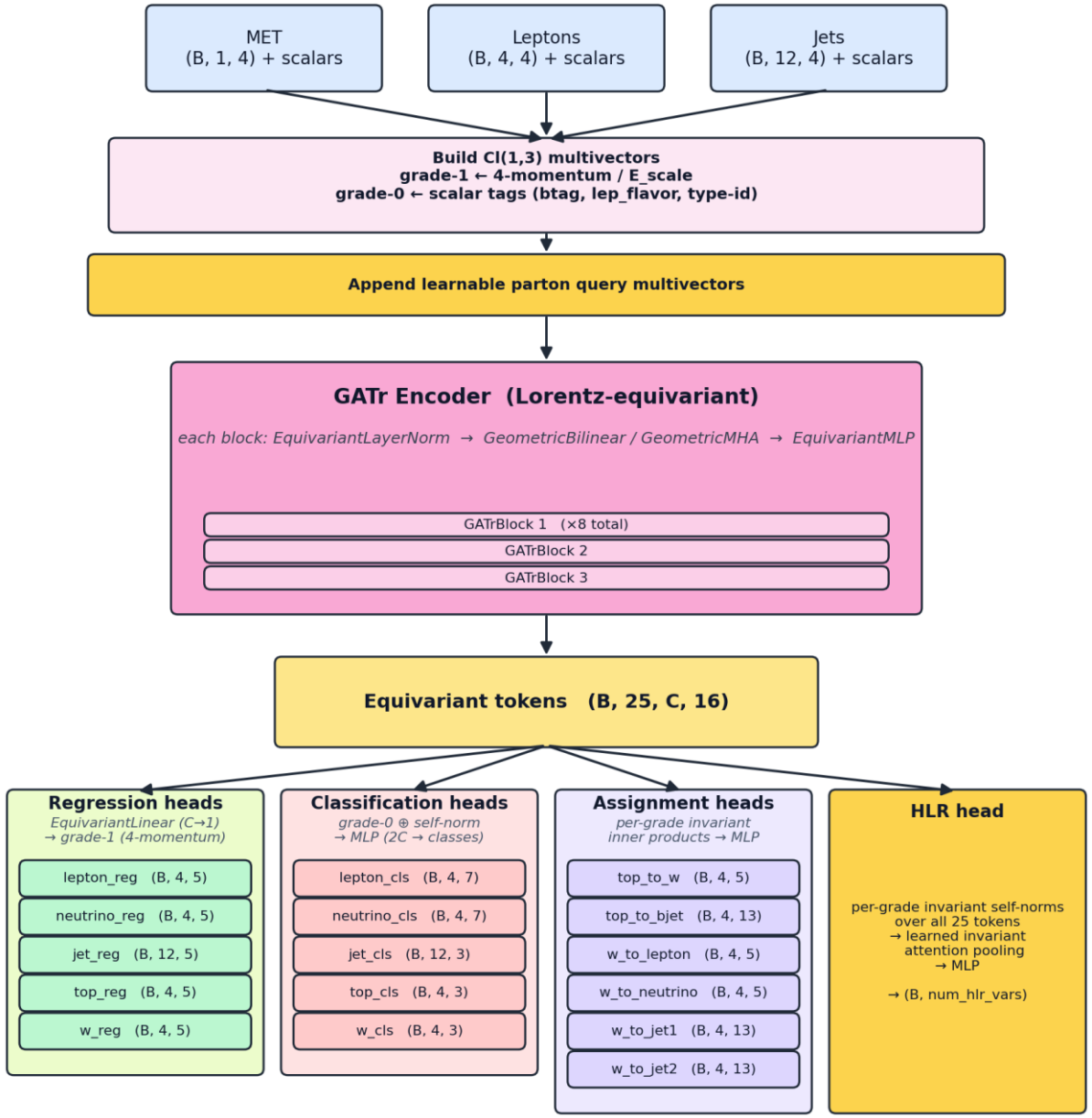
$$Z^{(c_1,c_2)} = X^{(c_1)}\, X^{(c_2)}, \qquad \langle Z \rangle_k = \sum_{\substack{r,s \geq 0 \\ |r-s| \leq k \leq r+s}} \langle X^{(c_1)} \rangle_r \langle X^{(c_2)} \rangle_s$$

- Gated nonlinearity on scalar norm

$$X \longmapsto \sigma\big(\langle X \tilde{X} \rangle_0\big) \cdot X$$

- Attention: Q, K, V are multivectors

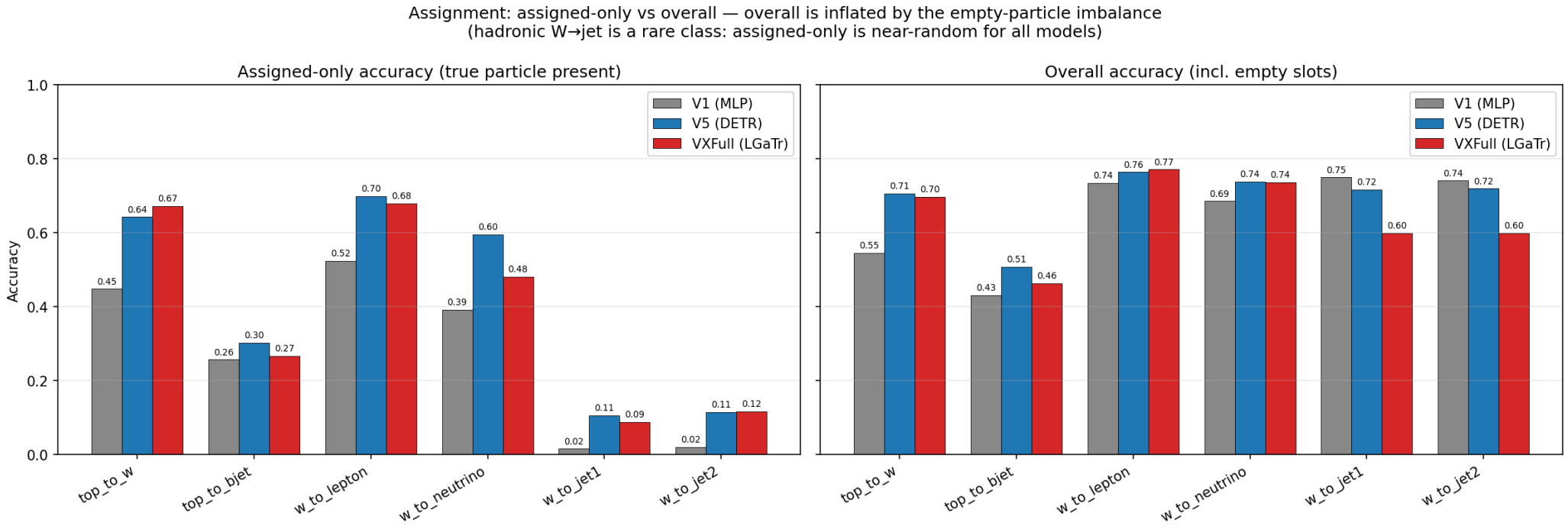
$$\mathrm{score}_{ij} = \langle Q_i\, \tilde{K}_j \rangle_0, \qquad \mathrm{Attn}(Q,K,V)_i = \sum_j \mathrm{softmax}\left(\frac{\mathrm{score}_{ij}}{\sqrt{d}}\right) V_j$$



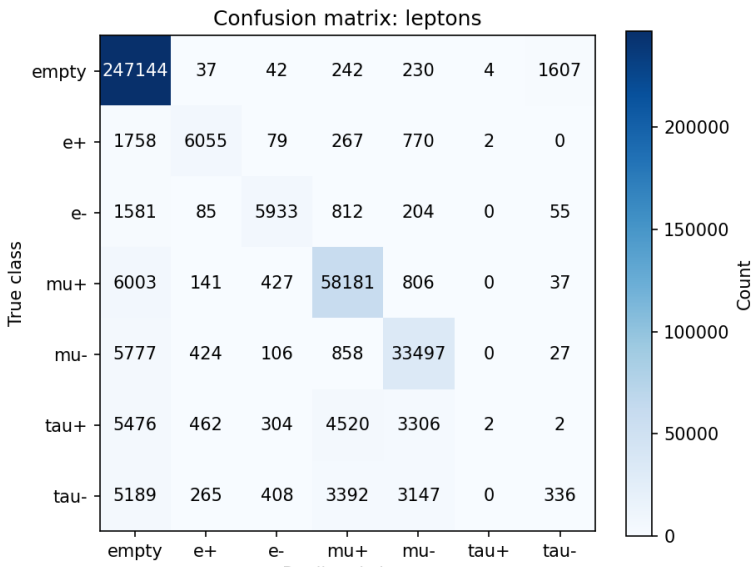
Single Lorentz-equivariant stack over detector + parton-query tokens
all heads read Lorentz invariants of the multivector tokens, so outputs transform covariantly with boosts.

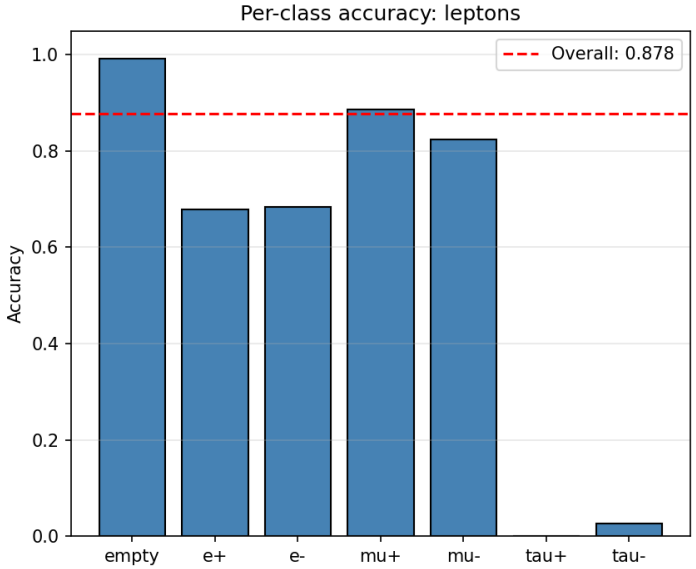
Results: Assignment

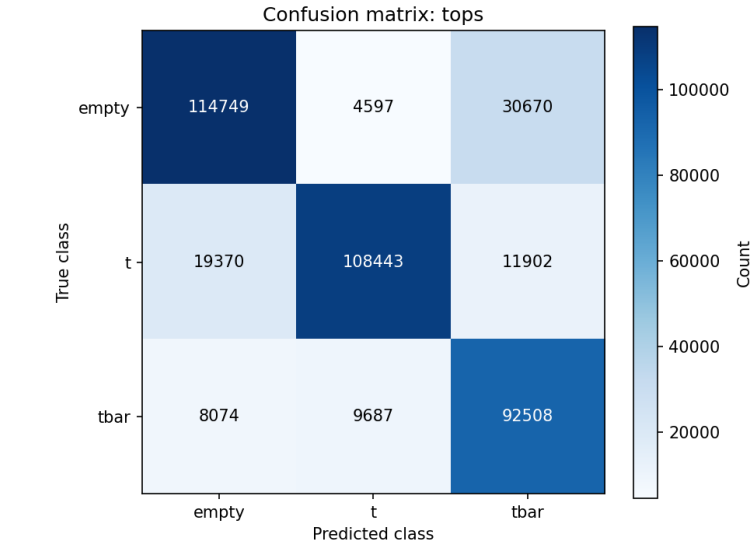
- LGaTr performs on par with DETR-like architecture in the assignment task
- Poor W -> jet performance is caused by low amount of such events in the training sample

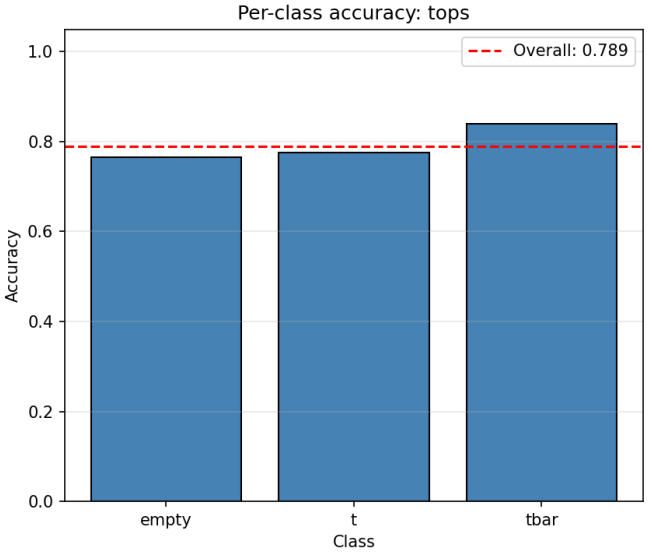


Results: Classification



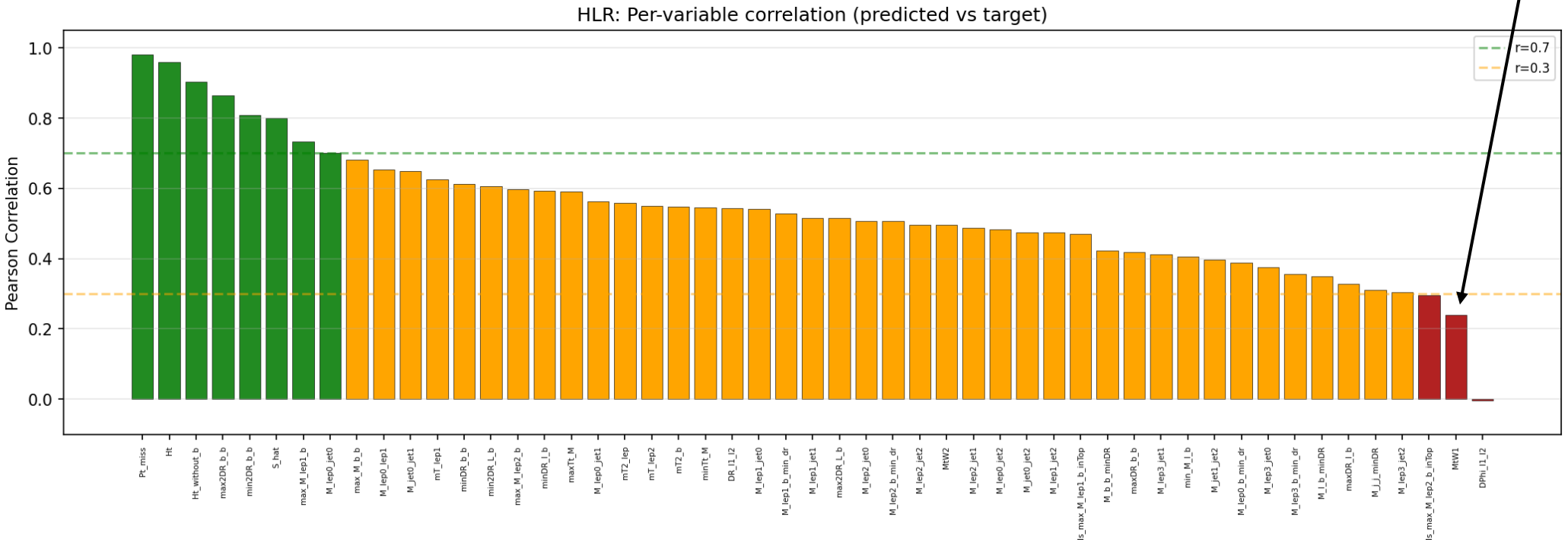






Results: HLR

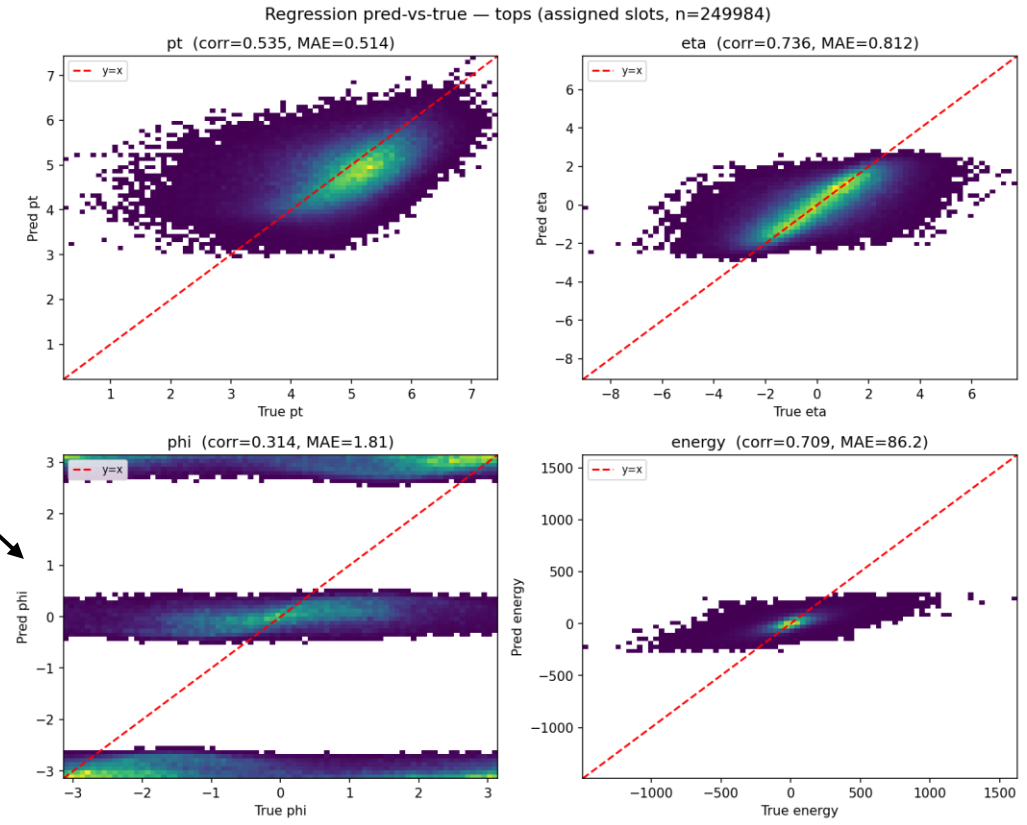
Poor performance; linked to low t->b assignment quality



Results: Regression

All-in-all – DETR here is on par or better for all tasks, probably due to the many available architectural tricks applied to it. LGaTr for now lacks these improvements and is expected to exhibit at least similar performance once these are implemented.

Even with circular phi loss LGaTr tends to favor discrete phi values



Conclusion

- Geometric algebra provides further physics-based improvements by introducing Lorentz-covariance and CP-odd observables
- LGaTr is promising in proof-of-concept discrimination tasks, but in the full reconstruction setup DETR-like models are still stronger
- The next step is architectural improvement and controlled ablation, to determine where GA equivariance gives a real advantage for HEP foundation models
- A more detailed study is published in the preprint 2605.15910

GA variables

#	Variable	Source	Grade	Frame	Inv. subgroup
1	m_i^2	[4]	0	inv.	Y
2	$\hat{s}$	[4,5]	0	inv.	Y
3	$\hat{t}$	[4]	0	inv.	Y
4	m_{ij}	[4,5]	0	inv.	Y
5	p_T^i	[5]	1	beam	H_{LHC}
6	η_i, y_i	[5]	1	beam	$\mathrm{SO}(1,1)_z$
7	$\phi_i{}^{\mathrm{a}}$	[5]	1	beam	$\mathrm{SO}(2)_\phi$
8	H_T, S_T	[4,5]	1	beam	H_{LHC}
9	$\cos\theta_\ell^*$	[4,20]	0	hel.	Y
10	$\Delta\phi_{ij}$, ΔR_{ij}	[5,11]	1→0	beam	H_{LHC}
11	$C_{ij}^{\mathrm{Bernreuther}}$	[21,22]	0/2	inv.	raw: N; inv. proj.: Y
12	E_T^{miss}	[5]	$1_{(\text{masked } p_z)}$	beam	H_{LHC}-only
13	b-tag, τ-tag	[5]	V_{flav}	n.a.	trivial
14	charge Q_ℓ	[5]	V_{flav}	n.a.	trivial
15	$T^\star_{ijk}$ trivector dual	this work; cf. [18]	3→1	inv.	raw: H_{LHC} for T^{123}; covariant: Y
16	$\mathrm{sgn}\langle p_1p_2p_3p_4\rangle_4$	this work; cf. [18,22]	4	inv.	**N** (CP-odd 1-bit)
17	sphericity S, aplanarity A	event-shape (standard)	n.a.	lab	n.a. (orth. complement)
18	thrust T, Fox–Wolfram H_ℓ	event-shape (standard) [23,24]	n.a.	lab	n.a. (orth. complement)