Geometric Algebra Transformer (GATr)

a versatile, scalable architecture for geometric data

Johann Brehmer*
Pim de Haan*
Sönke Behrends
Taco Cohen
*Equal contribution

Qualcomm AI research
An initiative of Qualcomm Technologies, Inc.

- GATr represents geometric data by combining the usual scalars with **geometric algebra** (GA) representations [1]
- We use a projective geometric algebra, which extends $\mathbb{R}^3$ to a graded **16-dimensional vector space**
- This GA can **express geometric objects** (points, lines...) and **transformations** (translations, rotations, ...)

Consists of grades: scalars, vectors, bivectors, trivectors, pseudoscalar

Object / operator	Scalar 1	Vector e_0	e_i	Bivector e_{0i}	e_{ij}	Trivector e_{0ij}	e_{123}	PS e_{0123}
Scalar $\lambda \in \mathbb{R}$	λ	0	0	0	0	0	0	0
Plane w/ normal $n \in \mathbb{R}^3$, origin shift $d \in \mathbb{R}$	0	d	n	0	0	0	0	0
Line w/ direction $n \in \mathbb{R}^3$, orthogonal shift $s \in \mathbb{R}^3$	0	0	0	s	n	0	0	0
Point $p \in \mathbb{R}^3$	0	0	0	0	0	p	1	0
Pseudoscalar $\mu \in \mathbb{R}$	0	0	0	0	0	0	0	μ
Reflection w/ normal $n \in \mathbb{R}^3$, origin shift $d \in \mathbb{R}$	0	d	n	0	0	0	0	0
Translation $t \in \mathbb{R}^3$	1	0	0	$\frac{1}{2}t$	0	0	0	0
Rotation expressed as quaternion $q \in \mathbb{R}^4$	q_0	0	0	0	q_i	0	0	0
Point reflection through $p \in \mathbb{R}^3$	0	0	0	0	0	p	1	0

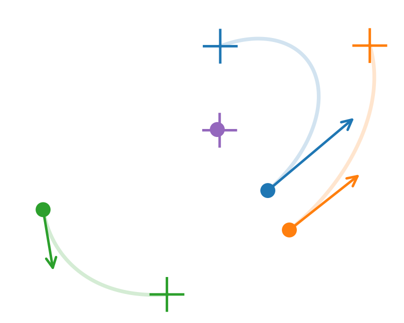
- GA also defines operations between GA elements:
  - Inner product / norm
  - **Geometric product**: multiplicative interaction that maps two GA elements to a new one (e.g. applying transformations to objects, mapping two points to distance, ...)
  - Join: Union-like operation needed for expressivity

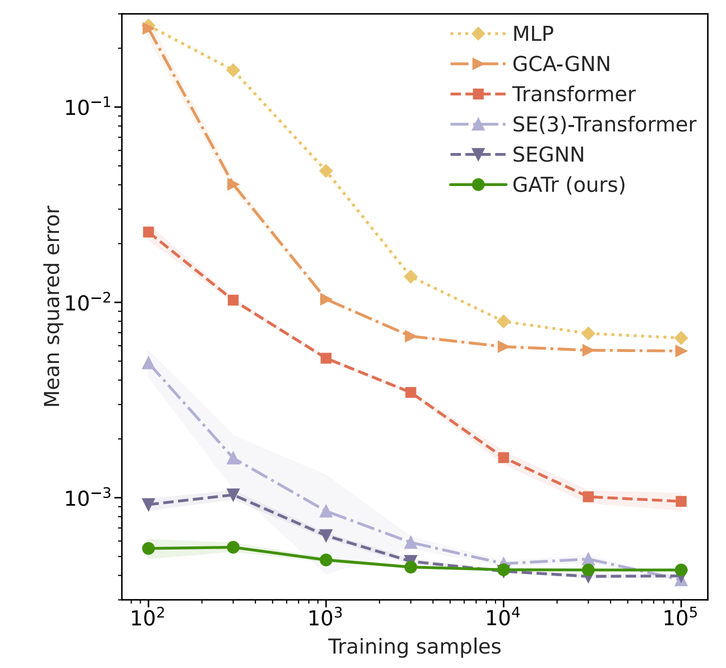
Further reading:

- Have a look at another GA-based, equivariant architecture [2]
- For a deep dive into different geometric algebras, see our **oral at the NeurReps workshop** on Saturday [3]

n-body modeling

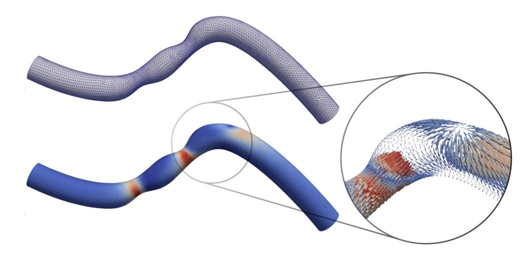
- GATr is more precise and sample-efficient than non-equivariant and equivariant baselines [4-6]

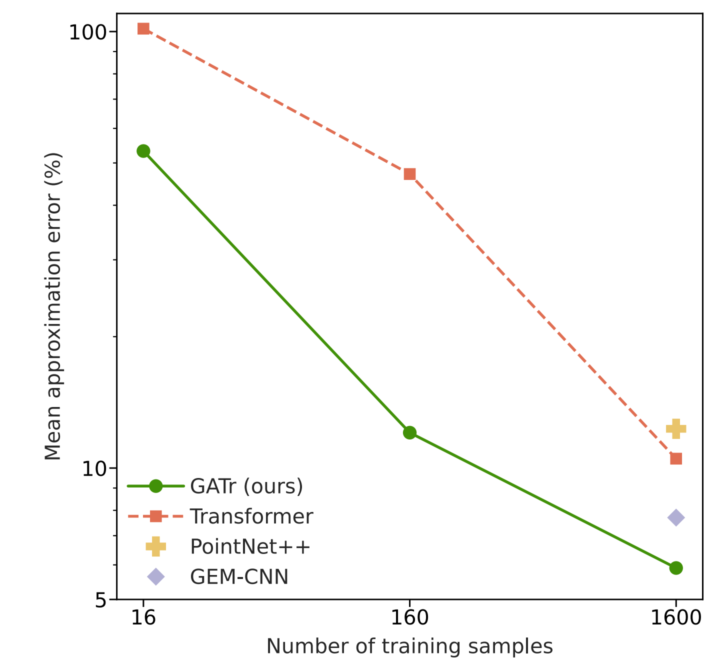




Predicting arterial wall shear stress

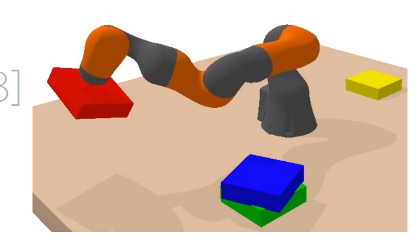
- Artery meshes of 7k nodes [7]
- GATr sets new SOTA

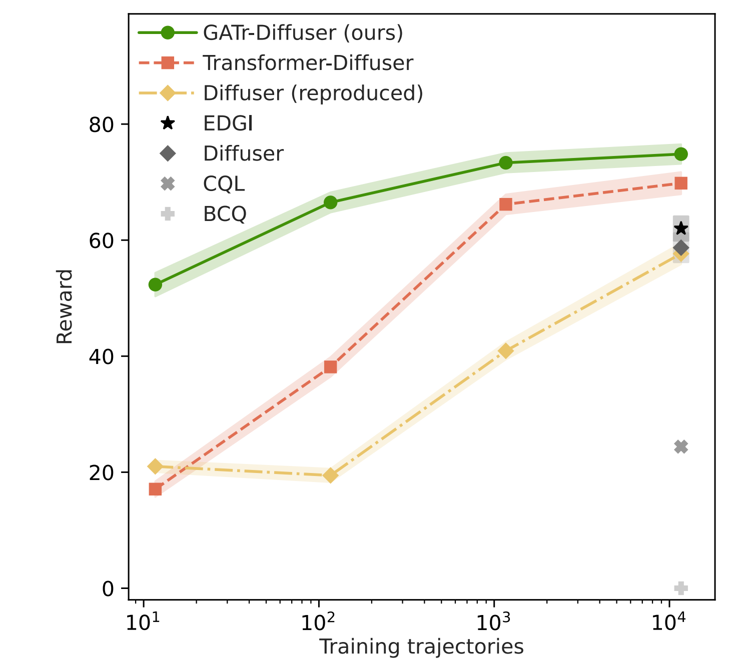




Robotic motion planning

- Imitation learning as generative modelling [8]
- GATr-based Diffuser outperforms baselines [8, 9]





Geometric algebra representations + E(3)-equivariant layers + Transformer architecture = Strong performance on diverse problems + Scalability to thousands of tokens

E(3)-equivariant layers

- GATr is **equivariant with respect to E(3)**, the symmetry group of 3D space
  - E(3) elements u act on GA as $\rho_u(x) = \begin{cases} uxu^{-1} & \text{if } u \text{ is even} \\ u\hat{x}u^{-1} & \text{if } u \text{ is odd} \end{cases}$

    Grade involution: flips sign for vector, trivector components
- Equivariance is achieved through **new layers**
  - Linear maps: $\mathrm{Linear}(x) = \sum_{k=0}^{d+1} w_k \langle x \rangle_k + \sum_{k=0}^{d} v_k e_0 \langle x \rangle_k$

    Projection to k-th grade

    (We prove that any equivariant linear map is of this form)
  - Gated nonlinearities: $\mathrm{GatedGELU}(x) = \mathrm{GELU}(x_1)x$

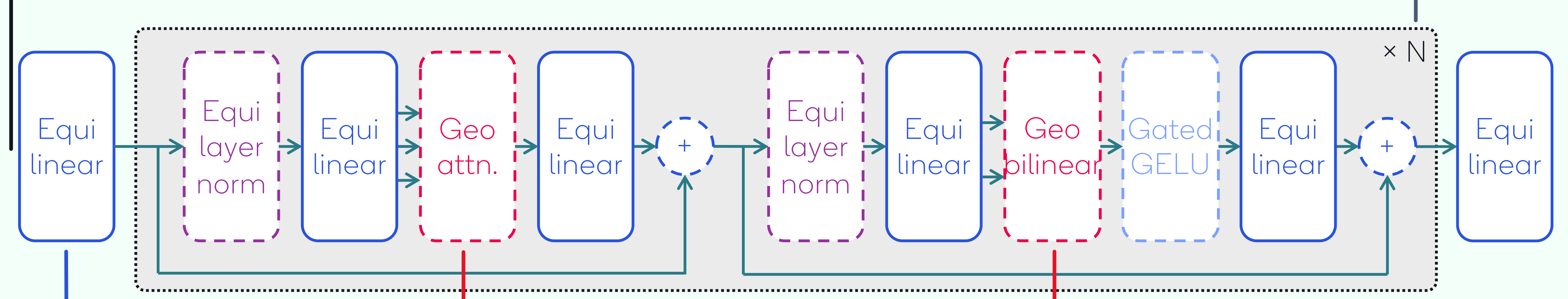
    Inner product
  - Normalization: $\mathrm{LayerNorm}(x) = x/\sqrt{\mathbb{E}_c\langle x, x\rangle}$
  - Attention: $\mathrm{Attention}(q,k,v)_{i'c'} = \sum_i \mathrm{Softmax}_i\left(\frac{\sum_c \langle q_{i'c}, k_{ic}\rangle}{\sqrt{8n_c}}\right) v_{ic'}$

    (We extend this mechanism with nonlinear features to make it sensitive to the Euclidean distance)
- GATr can **represent absolute positions faithfully and is equivariant to translations** – unlike most geometric architectures, which are only equivariant to rotations

Transformer architecture

Input and output data can have one or multiple token dimensions (for multiple token dimensions, blocks alternate between attending over different dimensions)

Attention blocks can be stacked to large depth



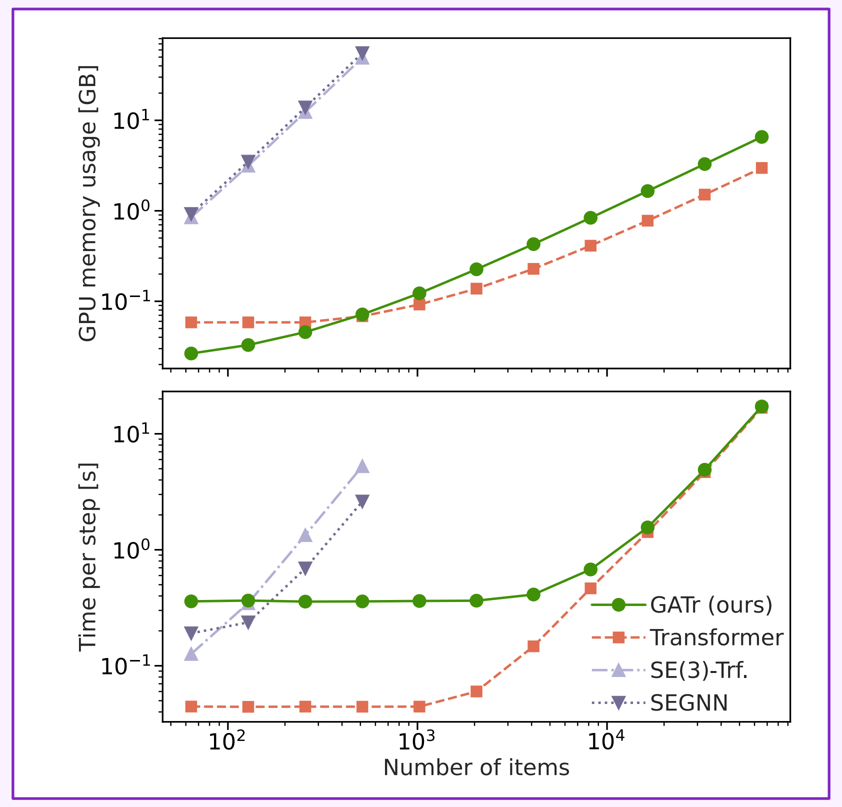
Linear layers between GA representations, with equivariance constraints

Geometric attention follows scaled dot-product attention recipe, but with GA representations and E(3) equivariance

Geometric bilinears combine GA's geometric product and join operation: construct new geometric types

Scalability to thousands of tokens

- Pairwise interactions through **scaled dot-product attention**
  - We can use **efficient backends** like Flash Attention [10]
- We can thus scale GATr to thousands of tokens, with fully connected interactions – impossible with most equivariant baselines!



[1] L. Dorst, "A guided tour to the plane-based geometric algebra PGA", 2020
[2] D. Ruhe et al, "Clifford group equivariant neural networks", NeurIPS 2023
[3] P. de Haan et al, "Euclidean, Projective, Conformal: Choosing a Geometric Algebra for Equivariant Transformers", NeurReps workshop at NeurIPS 2023
[4] F. Fuchs et al, "SE(3)-Transformers: 3D Roto-Translation equivariant attention networks", NeurIPS 2020
[5] J. Brandstetter et al, "Geometric and physical quantities improve E(3) equivariant message passing", ICLR 2022
[6] D. Ruhe et al, "Geometric clifford algebra networks", ICML 2023
[7] J. Suk et al, "Mesh neural networks for SE(3)-equivariant hemodynamics estimation on the artery wall", arXiv:2212.05023
[8] M. Janner et al, "Planning with diffusion for flexible behavior synthesis", ICML 2022
[9] J. Brehmer et al, "EDGI: Equivariant Diffusion for Planning with Embodied Agents", NeurIPS 2023
[10] T. Dao et al, "FlashAttention: Fast and memory-efficient exact attention with IO-awareness", NeurIPS 2022

Paper: Code: