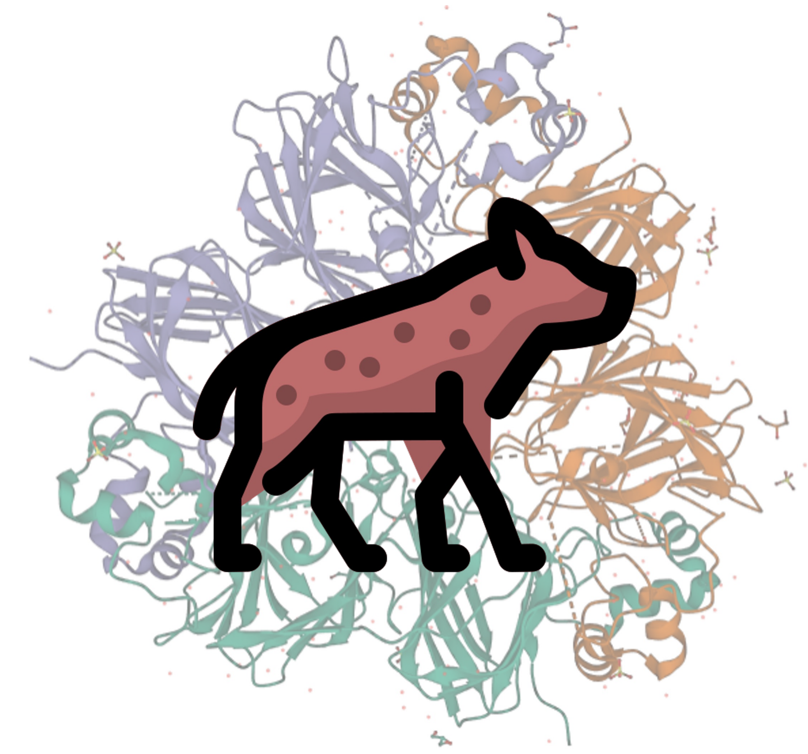


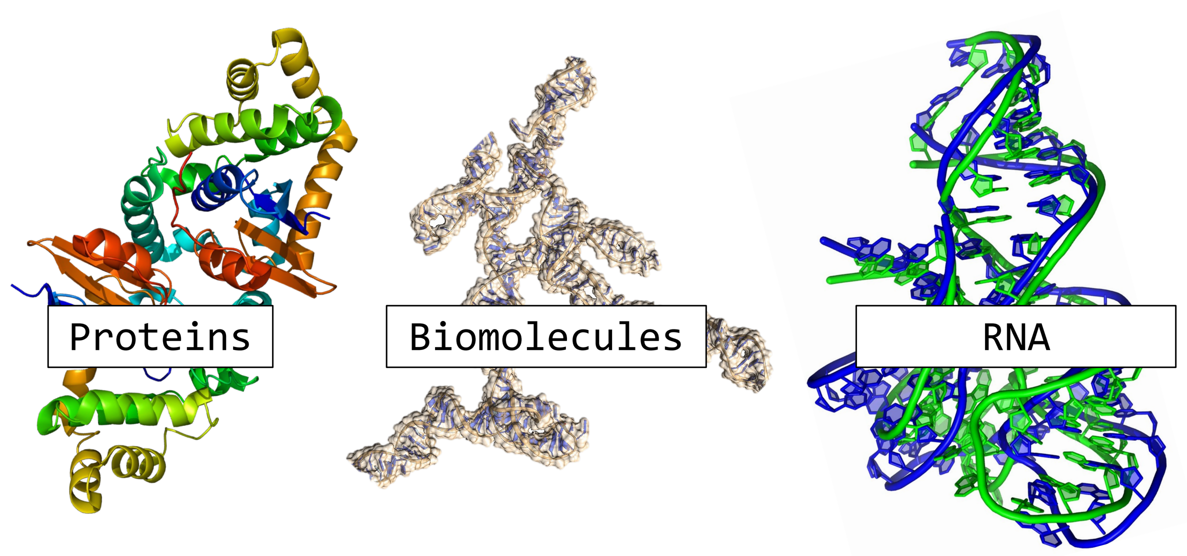
Geometric Hyena Networks for Large-scale Equivariant Learning

⭐ Spotlight ⭐

Artem Moskalev[1], Mangal Prakash[1], Junjie Xu[1,2], Tianyu Cui[1], Rui Liao[1], Tommaso Mansi[1]

[1]Johnson & Johnson Innovative Medicine, [2]The Pennsylvania State University

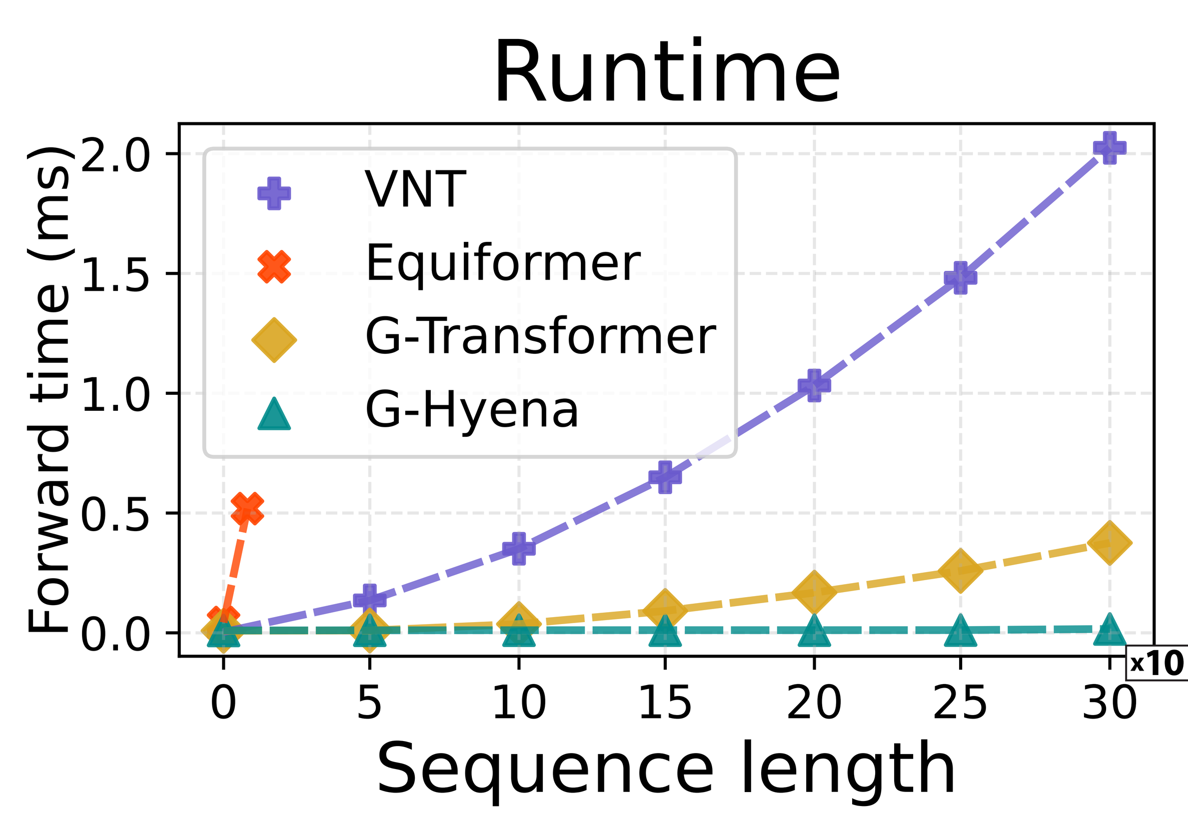
Large-scale systems

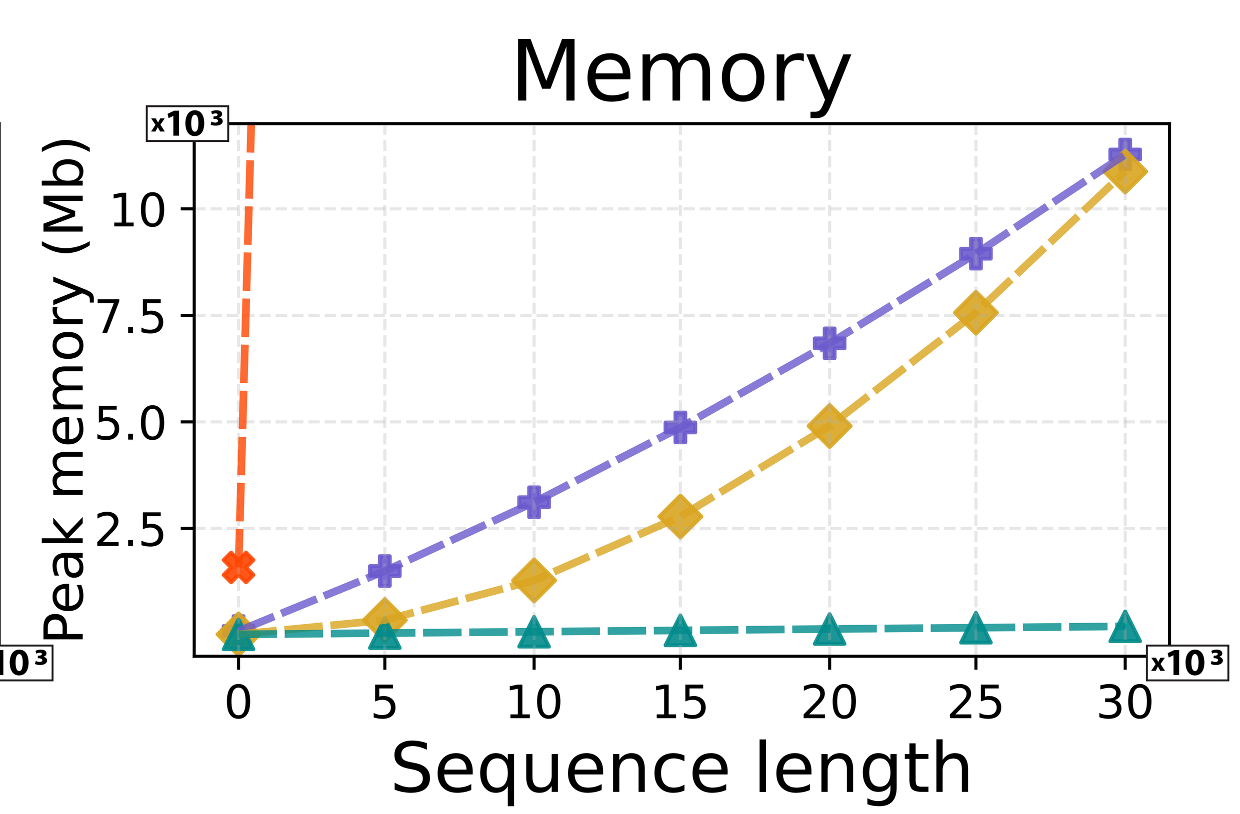


number of atoms: $\sim 10^3 - 10^8$

Quadratic complexity of global context with **equivariance** limits model **scalability** for large systems

2.7 million tokens on a single GPU (A10G)



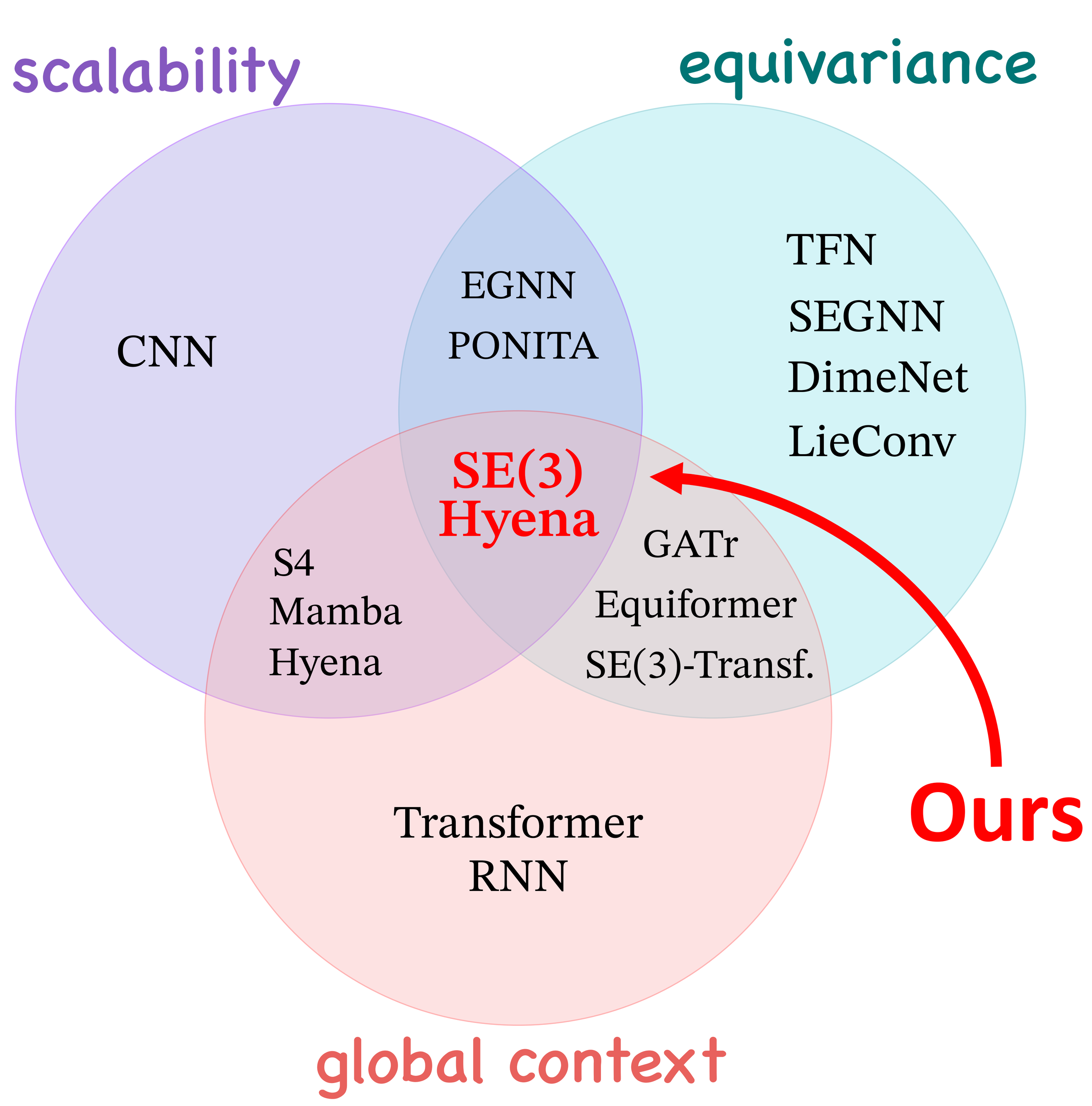


Global long context ✅

SE(3) Equivariance ✅

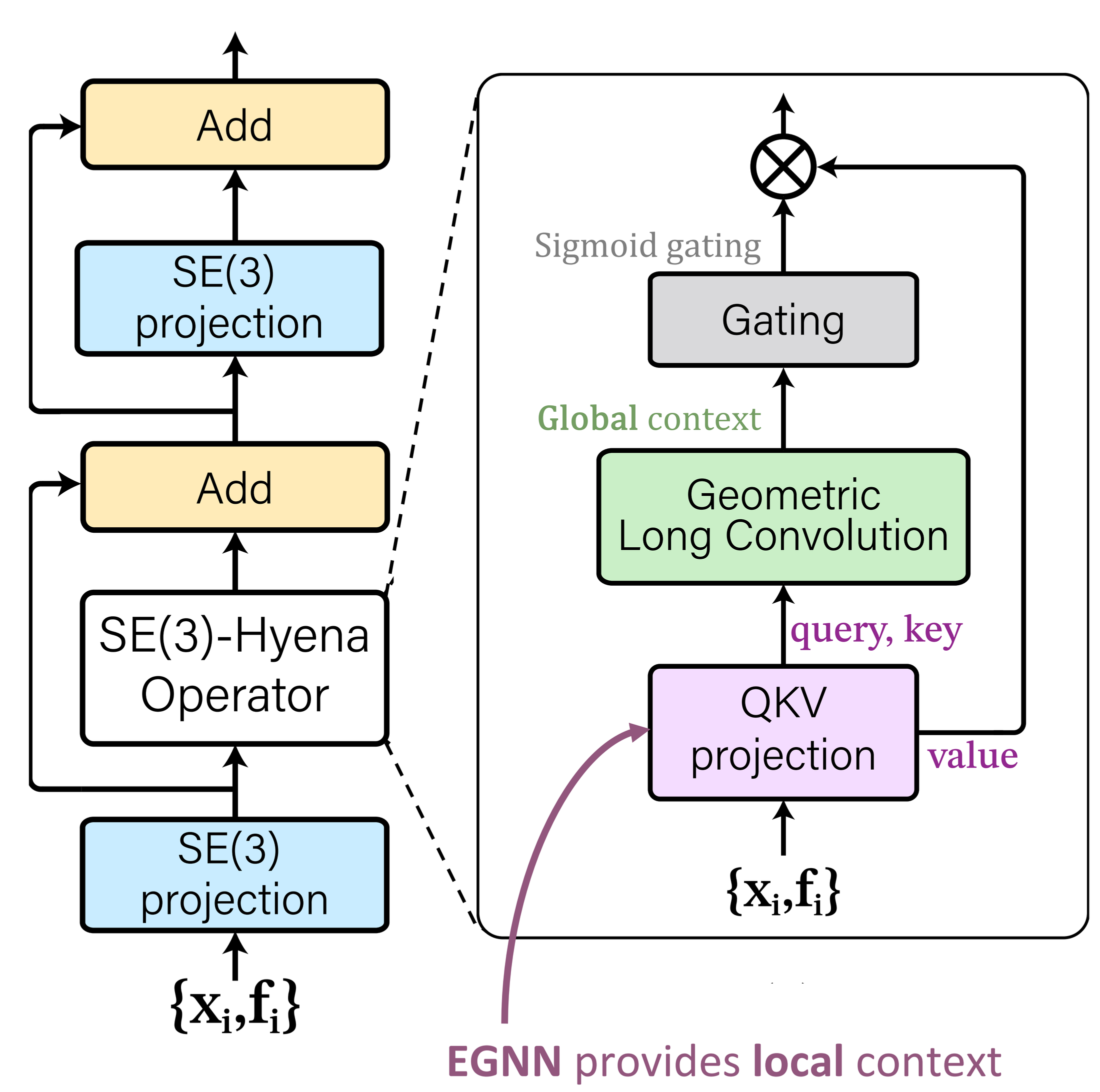
O(NlogN) complexity ✅

Other models



The first long-convolutional equivariant model

Architecture



Vector-convolution:

$$\mathbf{a}_i = (\mathbf{b} \circledast_\times \mathbf{c})_i = \sum_{j=1}^{N} \overbrace{\mathbf{b}_i \underbrace{\times}_{\text{cross prod.}} \mathbf{c}_{j-i}}^{\text{vectors}}$$

Scalar-decomposition:

$$\begin{bmatrix} b_x \\ b_y \\ b_z \end{bmatrix} \times \begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix} = \begin{bmatrix} b_y c_z - c_y b_z \\ c_x b_z - b_x c_z \\ b_x c_y - c_x b_y \end{bmatrix}$$

$$(\mathbf{b} \circledast_\times \mathbf{c})_i[x] = \sum_{y,z} \varepsilon_{xyz} \overbrace{\sum_{j=1}^{N} b_i[y]\, c_{j-i}[z]}^{\text{FFT-conv.: } O(N\log N)}$$

Higher-order steerable form:

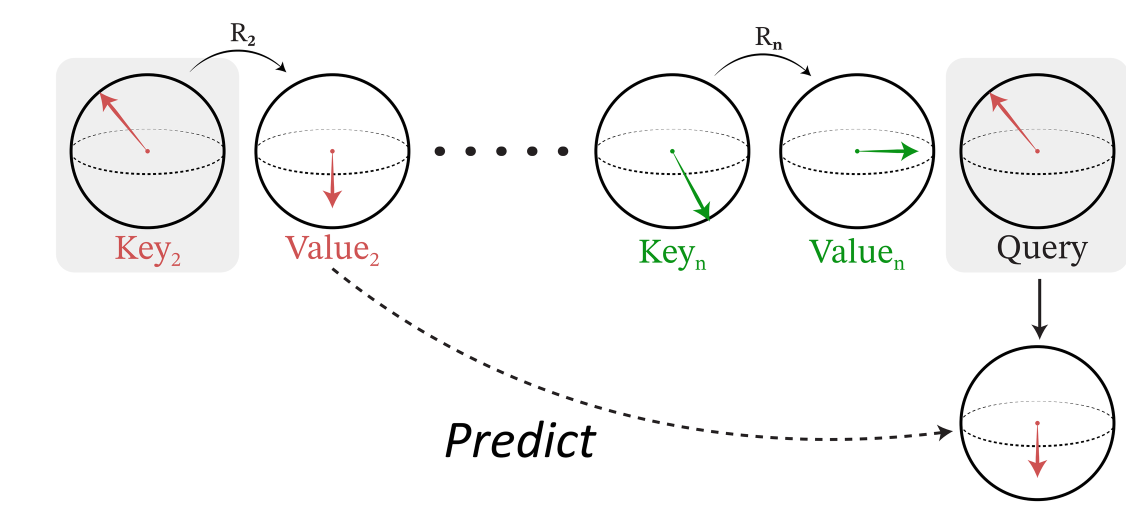
$$\underbrace{(\mathbf{b}^{l_1} \circledast_{CG} \mathbf{c}^{l_2})^{(l)}_{i,m}}_{\text{Tensor convolution}} = \sum_{j=1}^{N} \sum_{m_1=-l_1}^{l_1} \sum_{m_2=-l_2}^{l_2} \underbrace{C^{(lm)}_{(l_1 m_1)(l_2 m_2)}}_{\text{Clebsch–Gordan}} \mathbf{b}^{(l_1)}_{i,m_1} \mathbf{c}^{(l_2)}_{j-i,m_2}$$

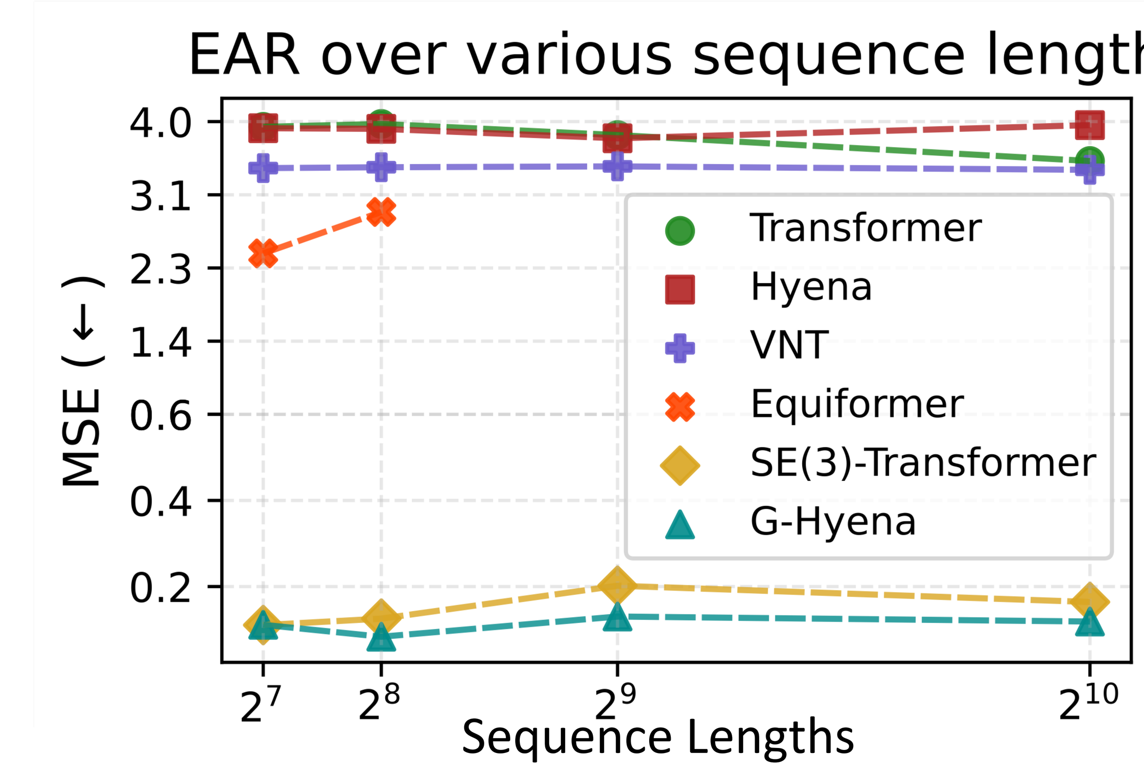
$$\mathbf{u}^{(l)}_{i,m} = \sum_{m_1=-l_1}^{l_1} \sum_{m_2=-l_2}^{l_2} C^{(lm)}_{(l_1 m_1)(l_2 m_2)} \left(\mathbf{b}^{(l_1)}_{m_1} \circledast \mathbf{c}^{(l_2)}_{m_2}\right)_i$$

‼️ **Assumption:** canonical ordering

Geometric long-convolution replaces equivariant self-attention

Equivariant Associative Recall (EAR)





Alternating global-local context is crucial

RNA degradation prediction

Data: *Large RNA biomolecules up to 11k atoms.* **Label:** *Degradation rate and Thermostability parameters per nucleotide* **Eval.:** *RMSE (↓)*

Model	Open Vaccine Covid-19 (Das et al., 2020)	Ribonanza-2k (He et al., 2024)
Backbone representation		
SchNet	0.515±0.005	0.911±0.008
TFN	0.522±0.006	0.927±0.006
EGNN	0.529±0.006	0.943±0.008
FastEGNN	0.519±0.012	0.912±0.032
LEFTNet	0.502±0.004	0.889±0.008
TMD-ET	0.500±0.006	0.781±0.006
Transformer	0.400±0.004	0.637±0.006
Hyena	0.447±0.037	0.810±0.124
VNT	0.401±0.005	0.659±0.005
G-Transformer	0.412±0.058	0.537±0.007
Equiformer	0.409±0.008	0.649±0.004
G-Heyna	**0.363**±0.045	**0.529**±0.005

Model	Open Vaccine Covid-19 (Das et al., 2020)	Ribonanza-2k (He et al., 2024)
All-atom representation		
SchNet	0.512±0.005	0.891±0.008
TFN	0.510±0.004	0.910±0.011
EGNN	0.511±0.005	0.928±0.022
FastEGNN	0.498±0.004	0.873±0.010
LEFTNet	0.501±0.005	0.880±0.008
TMD-ET	0.494±0.009	0.855±0.002
Transformer	0.399±0.004	0.633±0.007
Hyena	0.393±0.013	0.605±0.017
VNT	0.391±0.013	0.638±0.008
G-Transformer	0.391±0.071	0.592±0.043
Equiformer	OOM	OOM
G-Heyna	**0.339**±0.004	**0.546**±0.006

RNA switching factor prediction

Data: *Large RNA biomolecules up to few thousand atoms.*
Label: *RNA switching factor*
Eval.: *RMSE (↓)*

Model	Tc-Ribo (Groher et al., 2018) Backbone repr.	Tc-Ribo (Groher et al., 2018) All-atom repr.
SchNet	0.737±0.002	0.691±0.018
TFN	0.733±0.003	0.710±0.009
EGNN	0.728±0.001	0.729±0.002
FastEGNN	0.704±0.005	0.727±0.011
LeftNet	0.749±0.006	0.750±0.004
TMD-ET	0.750±0.004	0.751±0.003
Transformer	0.556±0.001	0.553±0.002
Hyena	0.560±0.002	0.569±0.001
G-Transformer	0.554±0.003	0.553±0.001
Equiformer	**0.550**±0.009	OOM
G-Heyna	**0.548**±0.008	**0.548**±0.001

Protein Molecular Dynamics

Data: *Backbone and All-atom proteins (up to 4k atoms)*
Label: *MD trajectory*
Eval.: *MSE (↓)*

Model	ProteinMD (Han et al., 2022) Backbone repr.	ProteinMD (Han et al., 2022) All-atom repr.
Linear	2.27±0.001	2.90±0.001
RF	2.26±0.001	2.85±0.001
TFN	2.26±0.002	2.85±0.002
EGNN	2.25±0.001	2.72±0.003
FastEGNN	1.84±0.002	NaN
Transformer	75.83±6.35	79.68±24.2
Hyena	48.94±9.03	55.34±5.51
G-Transformer	2.45±0.037	3.67±0.640
Equiformer	OOM	OOM
G-Heyna	**1.80**±0.009	**2.49**±0.037

G-Hyena outperforms other local and global* methods

*at a fraction of their computational cost