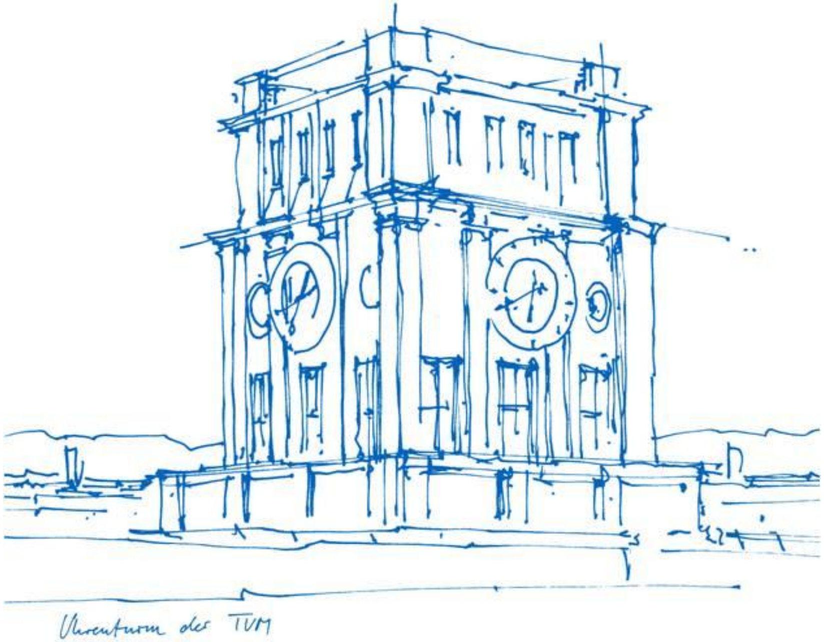
Equivariant Spatio-Temporal Attentive Graph Networks for Physical Dynamics

Celia Tundidor Centeno



Previously…

What are we discussing today?

How can we use machine learning to simulate physical systems with **high fidelity to their dynamics**?

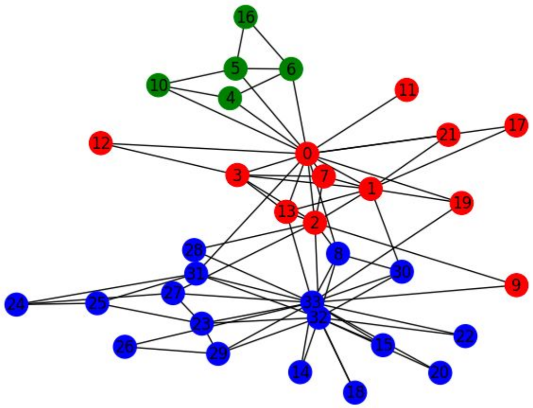
Contexts of application:

- topics: molecular dynamics, protein structure prediction, robotics…
- levels: macro, protein, smaller molecules…

Foundational concepts: Graph Neural Networks (GNNs)

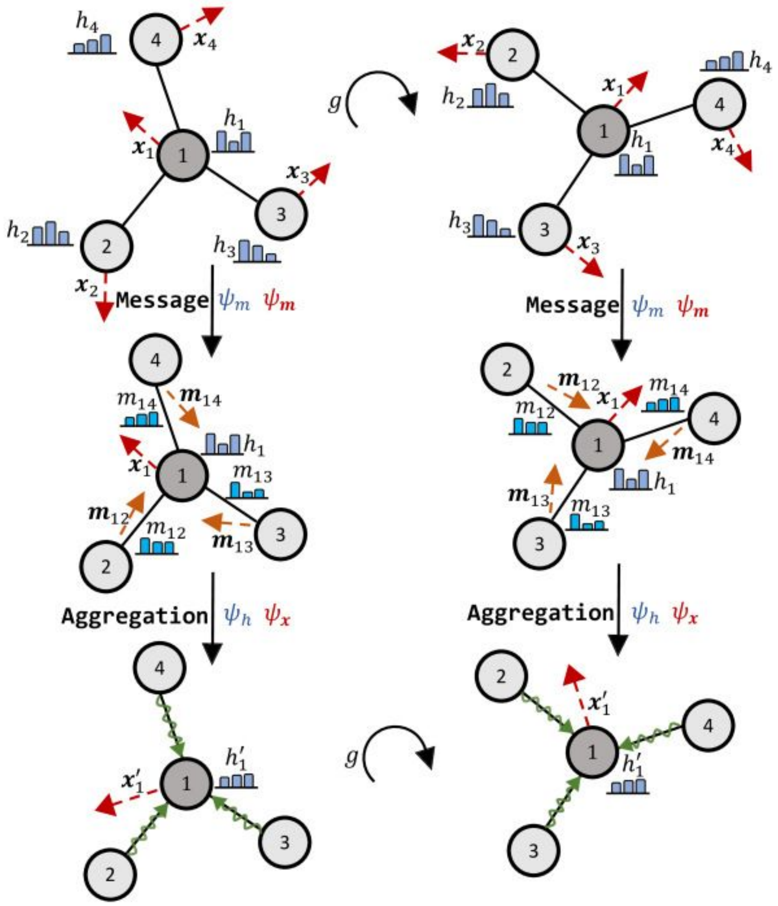
Naturally fit for physical system representation

- Unit elements as nodes (e.g., atoms)
- Relations as edges (e.g., chemical bonds)
- Latent interactions as message passing between these nodes with edges



Foundational concepts: Equivariance

Output reflects a predictable transformation equivalent to that of the input. Physical consistence irrespective of the coordinate system and view



State of the Art: equivariant GNNs

Spatially: generalising GNNs to fit the symmetry of our world

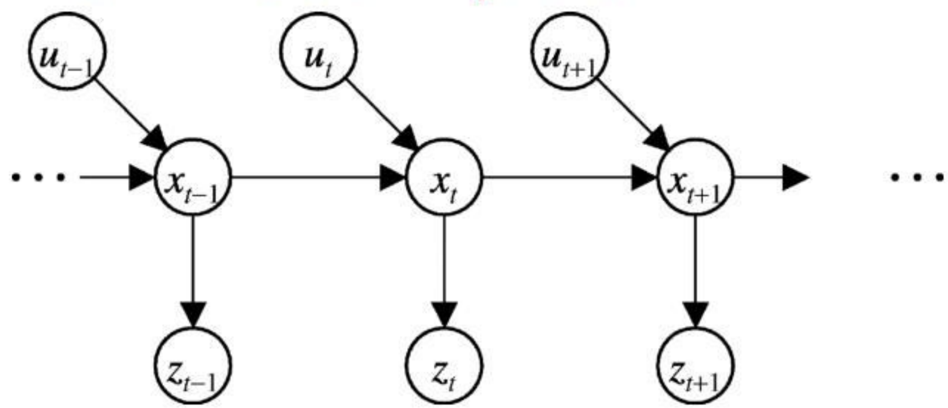
Temporally: **frame-to-frame forecasting**

E.g.:

- Tensor-Field Networks (TFN)
- SE(3)-Transformer
- LieTransformer and LieConv
- E(n)-equivariant GNNs (EGNN)
- Equivariant Graph Mechanics Networks (GMN)

The problem: the Markovian assumption

"The future state only depends on the current state, independent of all other past states"



$$p(z_t \mid x_{0:t}, z_{1:t}, u_{1:t}) = p(z_t \mid x_t)$$

$$p(x_t \mid x_{1:t-1}, z_{1:t}, u_{1:t}) = p(x_t \mid x_{t-1}, u_t)$$

The problem: the Markovian assumption

“The future state only depends on the current state, independent of all other past states”

Previous methods rely on this:

- A single input: system's conformation at a single frame.
- A fixed time step: they predict the future after a fixed time interval (frame-to-frame)

Why is the Markovian assumption problematic?

What if there are unobserved objects interacting with the system?

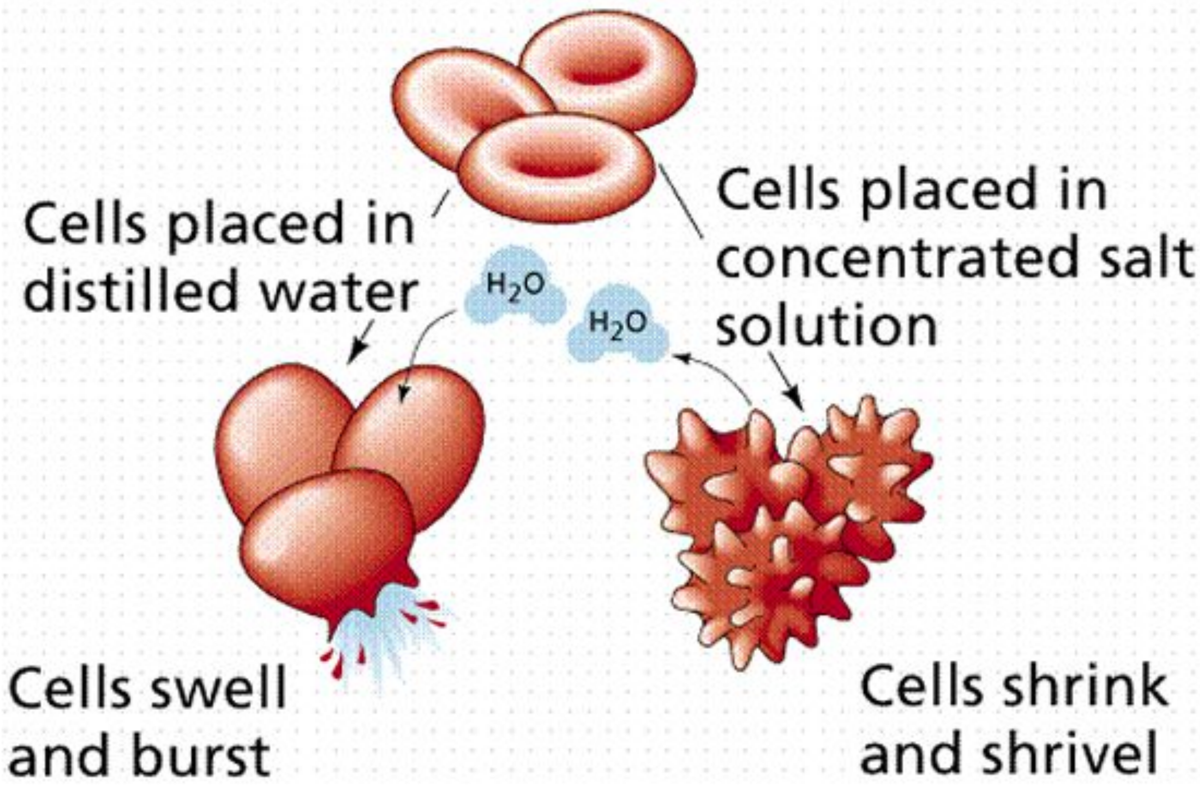
- Missed by the last frame
- Untracked

What if the effects induced by other objects are not constant or linear?

Why is the Markovian assumption problematic?

For molecular dynamics in particular:

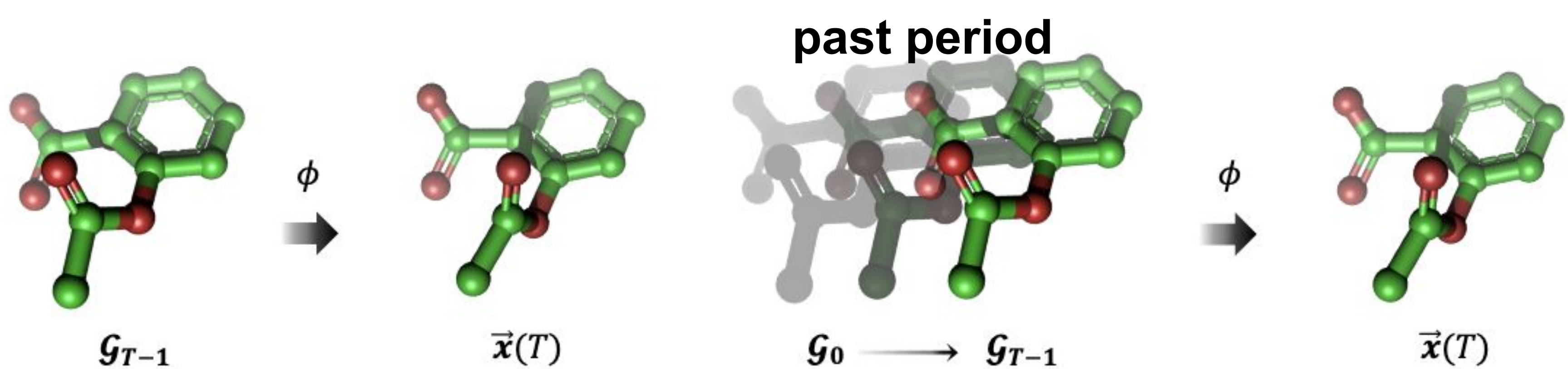
What about solvents (untracked object)?



Addressing Non-Markovian Dynamics

We define a **past period** (to be taking as input)

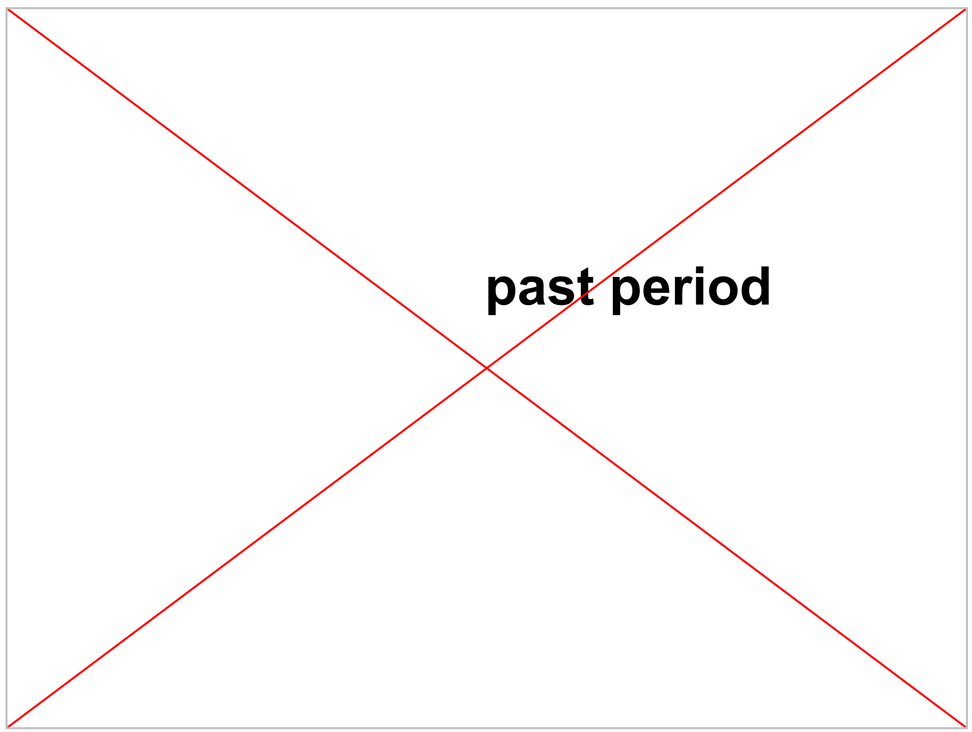
Idea: if the past period is sufficiently long, non-Markovian behaviour can be recovered



(***a***) Previous Setting (***b***) Our Setting

Addressing Non-Markovian Dynamics

We can also recover periodic motion (e.g. periodic thermal vibration)



Addressing Non-Markovian Dynamics

We can therefore use Spatio-Temporal Graph Neural Networks (STGNNs)...but they are unfit for Euclidean symmetry and physical laws

- traditional use case not on physical modelling (e.g. traffic forecasting)
- no 3D geometric equivariance

Enter ESTAG

Equivariant Spatio-Temporal Attentive Graph Networks (ESTAG):

- capturing non-Markovian behaviour (based on STGNNs)
- making STGNNs equivariant (for Euclidean symmetry)

ESTAG components

1. Equivariant Discrete Fourier Transform (EDFT): extracts periodic patterns
2. Equivariant Spatial Module (ESM): passes spatial messages.
3. Equivariant Temporal Module (ETM): aggregates temporal messages using forward attention and equivariant pooling

Equivariant Discrete Fourier Transform (EDFT)

Fourier Transform helps us understand the frequency domain

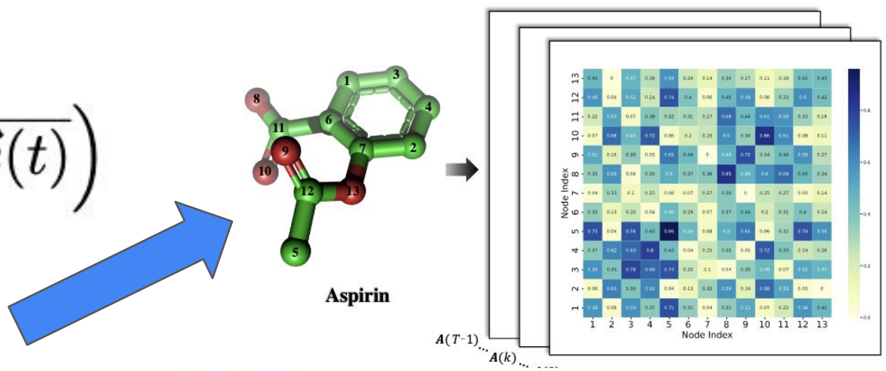
-> periodicity (node-wise temporal dynamics for the global context). c_i is the frequency amplitude of node i.

We can later use this information to check node cross-correlation (A)

A and c are E(3)-invariant!

$$\vec{\boldsymbol{f}}_i(k) = \sum_{t=0}^{T-1} e^{-i' \frac{2\pi}{T} kt} \left(\vec{\boldsymbol{x}}_i(t) - \overline{\vec{\boldsymbol{x}}(t)} \right)$$

$$\boldsymbol{c}_i(k) = w_k(\boldsymbol{h}_i) \| \vec{\boldsymbol{f}}_i(k) \|^2.$$



$$\boldsymbol{A}_{ij}(k) = w_k(\boldsymbol{h}_i) w_k(\boldsymbol{h}_j) |\langle \vec{\boldsymbol{f}}_i(k), \vec{\boldsymbol{f}}_j(k) \rangle|$$

Equivariant Spatial Module (ESM)

Encoding and passing the spatial geometry of each graph through each layer

EGNN + EDFT features:

+ correlation (Aij) to evaluate global temporal connections
+ amplitude (ci) to update hidden features at each node

Equivariant Spatial Module (ESM)

Process: compute messages, update hidden features, update positions

$$\boldsymbol{m}_{ij} = \phi_m\left(\boldsymbol{h}_i^{(l)}(t), \boldsymbol{h}_j^{(l)}(t), \|\vec{\boldsymbol{x}}_{ij}^{(l)}(t)\|^2, \boldsymbol{A}_{ij}\right),$$

$$\boldsymbol{h}_i^{(l+1)}(t) = \boldsymbol{h}_i^{(l)}(t) + \phi_h\left(\boldsymbol{h}_i^{(l)}(t), \boldsymbol{c}_i, \sum_{j \neq i} \boldsymbol{m}_{ij}\right),$$

$$\vec{\boldsymbol{a}}_i(t) = \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \vec{\boldsymbol{x}}_{ij}^{(l)}(t)\phi_x(\boldsymbol{m}_{ij}),$$

$$\vec{\boldsymbol{x}}_i^{(l+1)}(t) = \vec{\boldsymbol{x}}_i^{(l)}(t) + \vec{\boldsymbol{a}}_i(t),$$

note that these operations do not disturb equivariance

Equivariant Temporal Module (ETM)

Modelling self-correspondence with an attention mechanism

Forward temporal attention: we only rely on the past

Equivariant pooling

Equivariant Temporal Module (ETM)

Modelling self-correspondence with an attention mechanism

Forward temporal attention: we only rely on the past

Equivariant pooling: aggregates spatial and temporal information

$$\alpha_i^{(l)}(ts) = \frac{\exp(\boldsymbol{q}_i^{(l)}(t)^\top \boldsymbol{k}_i^{(l)}(s))}{\sum_{s=0}^{t} \exp(\boldsymbol{q}_i^{(l)}(t)^\top \boldsymbol{k}_i^{(l)}(s))},$$

attention weight

$$\boldsymbol{h}_i^{(l+1)}(t) = \boldsymbol{h}_i^{(l)}(t) + \sum_{s=0}^{t} \alpha_i^{(l)}(ts)\boldsymbol{v}_i^{(l)}(s),$$

hidden feature

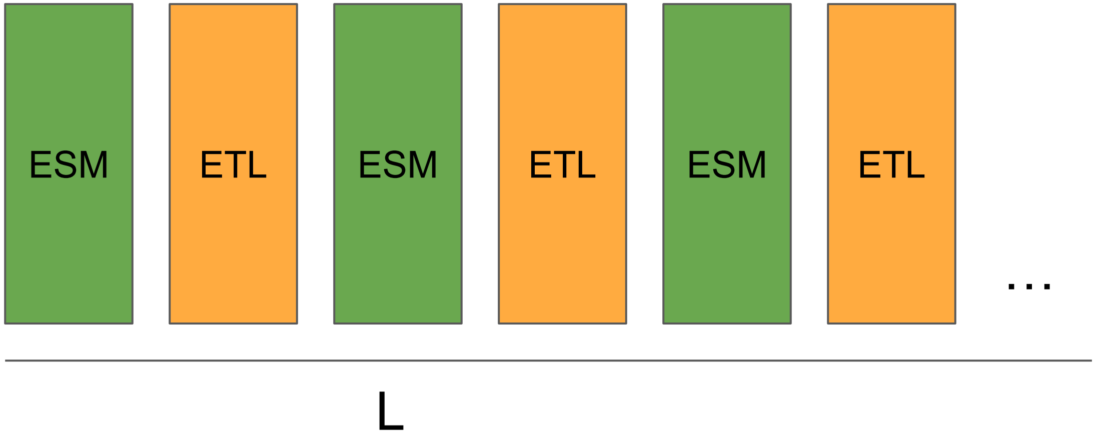
temporal displacement vector

$$\vec{\boldsymbol{x}}_i^{(l+1)}(t) = \vec{\boldsymbol{x}}_i^{(l)}(t) + \sum_{s=0}^{t} \alpha_i^{(l)}(ts)\vec{\boldsymbol{x}}_i^{(l)}(ts)\phi_x(\boldsymbol{v}_i^{(l)}(s)),$$

Equivariant Temporal Pooling

Equivariant pooling: apply a linear transformation to the updated coordinates

$$\vec{\boldsymbol{x}}_i^*(T) = \hat{\boldsymbol{X}}_i \boldsymbol{w} + \vec{\boldsymbol{x}}_i^{(L)}(T-1),$$



Training via MSE loss

$$\mathcal{L} = \sum_{i=1}^{N} \|\vec{\boldsymbol{x}}_i(T) - \vec{\boldsymbol{x}}_i^*(T)\|_2^2.$$

Architecture recap

Input: historical series of spatio-temporal graphs {Gt} from time t=0 to T-1

Equivariant Discrete Fourier Transform (EDFT): processes historical trajectory for each node. Extracts equivariant frequency features-> invariant node features (c) and adjacency matrix (A).

Stacked Modules: computes spatial and temporal relationships. L layers of alternating equivariant components (ESM, ETM)

Equivariant Temporal Pooling: pooling layer to combine time and space dependencies

Output: position of each node at time T

Architecture recap

EDFT:

$$\vec{\boldsymbol{f}}_i(k) = \sum_{t=0}^{T-1} e^{-i' \frac{2\pi}{T} kt} \left(\vec{\boldsymbol{x}}_i^{\alpha}(t) - \overline{\vec{\boldsymbol{x}}^{\alpha}(t)} \right),$$

$$\boldsymbol{A}_{ij}(k) = w_k(\boldsymbol{h}_i) w_k(\boldsymbol{h}_j) |\langle \vec{\boldsymbol{f}}_i(k), \vec{\boldsymbol{f}}_j(k) \rangle|,$$

$$\boldsymbol{c}_i(k) = w_k(\boldsymbol{h}_i) \|\vec{\boldsymbol{f}}_i(k)\|^2.$$

Architecture recap

ESM:

$$
\begin{aligned}
\boldsymbol{m}_{ij} &= \phi_m \left(\boldsymbol{h}_i^{(l)}(t), \boldsymbol{h}_j^{(l)}(t), \frac{(\vec{\boldsymbol{X}}_{ij}^{(l)}(t))^\top \vec{\boldsymbol{X}}_{ij}^{(l)}(t)}{\|(\vec{\boldsymbol{X}}_{ij}^{(l)}(t))^\top \vec{\boldsymbol{X}}_{ij}^{(l)}(t)\|_F}, \boldsymbol{A}_{ij} \right), \\
\boldsymbol{h}_i^{(l+1)}(t) &= \phi_h \left(\boldsymbol{h}_i^{(l)}(t), \boldsymbol{c}_i(k), \sum_{j \neq i} \boldsymbol{m}_{ij} \right), \\
\vec{\boldsymbol{A}}_i^{(l)}(t) &= \frac{1}{|\mathcal{N}(i)|} \sum_{j \in \mathcal{N}(i)} \vec{\boldsymbol{X}}_{ij}^{(l)}(t) \phi_{\boldsymbol{X}}(\boldsymbol{m}_{ij}), \\
\vec{\boldsymbol{X}}_i^{(l+1)}(t) &= \vec{\boldsymbol{X}}_i^{(l)}(t) + \vec{\boldsymbol{A}}_i^{(l)}(t).
\end{aligned}
$$

ETM:

$$
\begin{aligned}
\alpha_i^{(l)}(ts) &= \frac{\exp(\boldsymbol{q}_i^{(l)}(t)^\top \boldsymbol{k}_i^{(l)}(s))}{\sum_{s=0}^{t} \exp(\boldsymbol{q}_i^{(l)}(t)^\top \boldsymbol{k}_i^{(l)}(s))}, \\
\boldsymbol{h}_i^{(l+1)}(t) &= \boldsymbol{h}_i^{(l)}(t) + \sum_{s=0}^{t} \alpha_i^{(l)}(ts) \boldsymbol{v}_i^{(l)}(s), \\
\vec{\boldsymbol{X}}_i^{(l+1)}(t) &= \vec{\boldsymbol{X}}_i^{(l)}(t) + \sum_{s=0}^{t} \alpha_i^{(l)}(ts)\ \vec{\boldsymbol{X}}_i^{(l)}(ts) \phi_{\boldsymbol{X}}(\boldsymbol{v}_i^{(l)}(s)),
\end{aligned}
$$

where

$$
\begin{aligned}
\boldsymbol{q}_i^{(l)}(t) &= \phi_q \left(\boldsymbol{h}_i^{(l)}(t) \right), \\
\boldsymbol{k}_i^{(l)}(t) &= \phi_k \left(\boldsymbol{h}_i^{(l)}(t) \right), \\
\boldsymbol{v}_i^{(l)}(t) &= \phi_v \left(\boldsymbol{h}_i^{(l)}(t) \right).
\end{aligned}
$$

Equivariance details

Theorem A.1. *We denote ESTAG as* $\vec{\boldsymbol{X}}(T) = \phi\left(\{(\boldsymbol{H}(t), g \cdot \vec{\boldsymbol{X}}(t), \boldsymbol{A})\}_{t=0}^{T-1}\right)$*, then* ϕ *is E(3)-equivariant.*

Proof. **1.** We firstly prove that EDFT is E(3)-equivariant.

$$\boldsymbol{O}\vec{\boldsymbol{f}}_i(k) = \sum_{t=0}^{T-1} e^{-i'\frac{2\pi}{T}kt}\left(\boldsymbol{O}\vec{\boldsymbol{x}}_i(t) + \boldsymbol{b} - \overline{\boldsymbol{O}\vec{\boldsymbol{x}}(t) + \boldsymbol{b}}\right),$$

$$\boldsymbol{A}_{ij}(k) = w_k(\boldsymbol{h}_i)w_k(\boldsymbol{h}_j)|\langle \boldsymbol{O}\vec{\boldsymbol{f}}_i(k), \boldsymbol{O}\vec{\boldsymbol{f}}_j(k)\rangle|,$$

$$\boldsymbol{c}_i(k) = w_k(\boldsymbol{h}_i)\|\boldsymbol{O}\vec{\boldsymbol{f}}_i(k)\|^2.$$

2. We secondly prove the E(3)-equivariance of ESM.

$$\boldsymbol{m}_{ij} = \phi_m\left(\boldsymbol{h}_i^{(l)}(t), \boldsymbol{h}_j^{(l)}(t), \|\boldsymbol{O}\vec{\boldsymbol{x}}_{ij}^{(l)}(t)\|^2, \boldsymbol{A}_{ij}\right),$$

$$\boldsymbol{h}_i^{(l+1)}(t) = \phi_h\left(\boldsymbol{h}_i^{(l)}(t), \boldsymbol{c}_i(k), \sum_{j\neq i}\boldsymbol{m}_{ij}\right),$$

$$\boldsymbol{O}\vec{\boldsymbol{a}}_i(t) = \frac{1}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)}\boldsymbol{O}\vec{\boldsymbol{x}}_{ij}^{(l)}(t)\phi_x(\boldsymbol{m}_{ij}),$$

$$\boldsymbol{O}\vec{\boldsymbol{x}}_i^{(l+1)}(t) + \boldsymbol{b} = \boldsymbol{O}\vec{\boldsymbol{x}}_i^{(l)}(t) + \boldsymbol{b} + \boldsymbol{O}\vec{\boldsymbol{a}}_i^{(l+1)}(t).$$

Equivariance details

3. We then prove that ETM is E(3)-equivariant.

$$\boldsymbol{q}_i^{(l)}(t) = \phi_q\left(\boldsymbol{h}_i^{(l)}(t)\right),$$
$$\boldsymbol{k}_i^{(l)}(t) = \phi_k\left(\boldsymbol{h}_i^{(l)}(t)\right),$$
$$\boldsymbol{v}_i^{(l)}(t) = \phi_v\left(\boldsymbol{h}_i^{(l)}(t)\right),$$
$$\alpha_i^{(l)}(ts) = \frac{\exp(\boldsymbol{q}_i^{(l)}(t)^\top \boldsymbol{k}_i^{(l)}(s))}{\sum_{s=0}^{t} \exp(\boldsymbol{q}_i^{(l)}(t)^\top \boldsymbol{k}_i^{(l)}(s))},$$
$$\boldsymbol{h}_i^{(l+1)}(t) = \boldsymbol{h}_i^{(l)}(t) + \sum_{s=0}^{t} \alpha_i^{(l)}(ts)\boldsymbol{v}_i^{(l)}(s),,$$
$$\boldsymbol{O}\vec{\boldsymbol{x}}_i^{(l+1)}(t) + \boldsymbol{b} = \boldsymbol{O}\vec{\boldsymbol{x}}_i^{(l)}(t) + \boldsymbol{b} + \sum_{s=0}^{t} \alpha_i^{(l)}(ts)\ \boldsymbol{O}\vec{\boldsymbol{x}}_i^{(l)}(ts)\phi_x(\boldsymbol{v}_i^{(l)}(s)).$$

4. We finally prove that the linear pooling is equivariant:

$$\boldsymbol{O}\vec{\boldsymbol{x}}_i^{*}(T) + \boldsymbol{b} = \boldsymbol{O}\hat{\boldsymbol{X}}_i\boldsymbol{w} + \boldsymbol{O}\vec{\boldsymbol{x}}_i^{(L)}(T-1) + \boldsymbol{b}.$$

Experiments

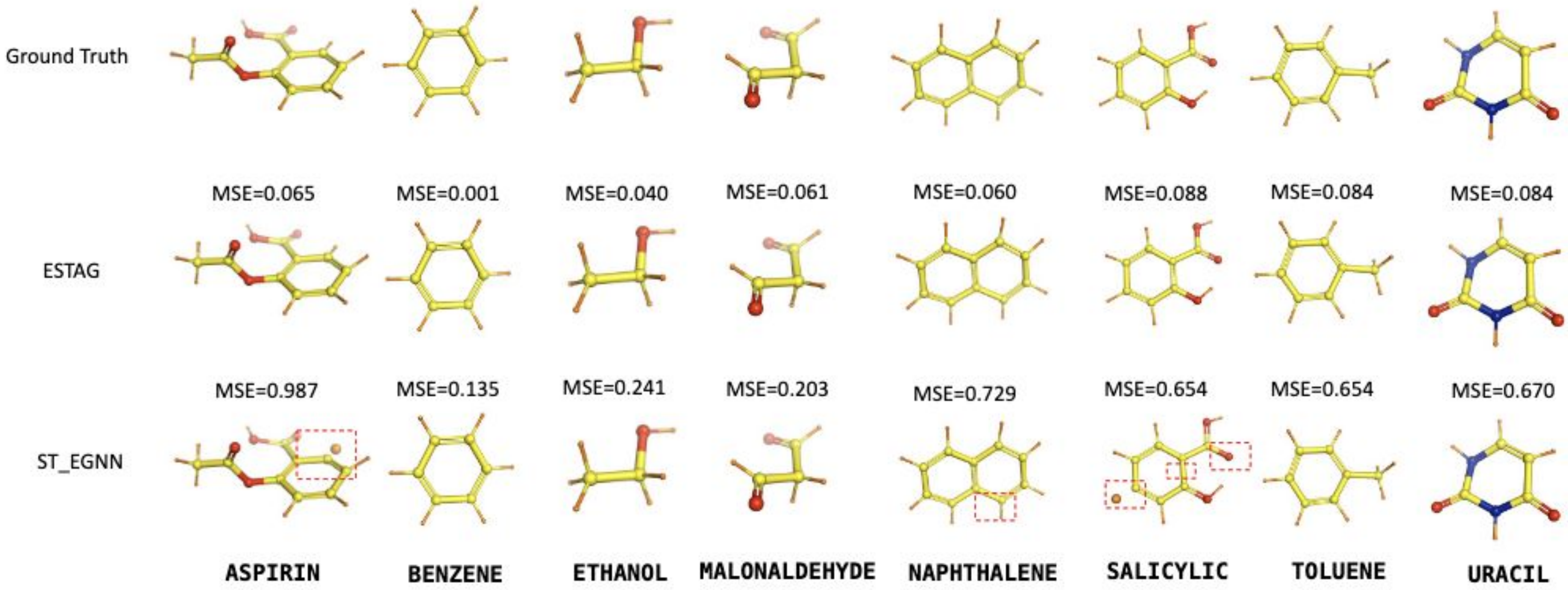
Testing on three datasets for the different levels:

Molecular: MD17, trajectories of small molecules (e.g., Aspirin, Benzene, Ethanol) generated by Molecular Dynamics simulation. External temperature and pressure are unobserved (non-Markovian behaviour)

Protein-level: AdK equilibrium trajectory dataset (protein dynamics). The dynamics of water and ions are unobserved (non-Markovian behaviour)

Macro-level: CMU Motion Capture Database (human motion trajectories) (e.g., walking, basketball). Environmental states are unobserved (non-Markovian behaviour)

Experimental results: molecular



Experimental results: molecular

Table 1: Prediction error ($\times 10^{-3}$) on MD17 dataset. Results averaged across 3 runs. We do not display the standard deviation due to its small value.

	Aspirin	Benzene	Ethanol	Malonaldehyde	Naphthalene	Salicylic	Toluene	Uracil
PT-s	15.579	4.457	4.332	13.206	8.958	12.256	6.818	10.269
PT-m	9.058	2.536	2.688	6.749	6.918	8.122	5.622	7.257
PT-t	0.715	0.114	0.456	0.596	0.737	0.688	0.688	0.674
EGNN-s	12.056	3.290	2.354	10.635	4.871	8.733	3.154	6.815
EGNN-m	6.237	1.882	1.532	4.842	3.791	4.623	2.516	3.606
EGNN-t	0.625	0.112	0.416	0.513	0.614	0.598	0.577	0.568
ST_TFN	0.719	0.122	0.432	0.569	0.688	0.684	0.628	0.669
ST_GNN	1.014	0.210	0.487	0.664	0.769	0.789	0.713	0.680
ST_SE(3)TR	0.669	0.119	0.428	0.550	0.625	0.630	0.591	0.597
ST_EGNN	0.735	0.163	0.245	0.427	0.745	0.687	0.553	0.445
EqMotion	0.721	0.156	0.476	0.600	0.747	0.697	0.691	0.681
STGCN	0.715	0.106	0.456	0.596	0.736	0.682	0.687	0.673
AGL-STAN	0.719	0.106	0.459	0.596	0.601	0.452	0.683	0.515
ESTAG	**0.063**	**0.003**	**0.099**	**0.101**	**0.068**	**0.047**	**0.079**	**0.066**

Experimental results: protein and macro

Method	MSE	Time(s)
Pt-s	3.260	-
Pt-m	3.302	-
Pt-t	2.022	-
EGNN-s	3.254	1.062
EGNN-m	3.278	1.088
EGNN-t	1.983	1.069
ST_GNN	1.871	2.769
ST_GMN	<u>1.526</u>	4.705
ST_EGNN	1.543	4.705
STGCN	1.578	1.840
AGL-STAN	1.671	1.478
ESTAG	**1.471**	6.876

Method	Walk	Basketball
Pt-s	329.474	886.023
Pt-m	127.152	413.306
Pt-t	3.831	15.878
EGNN-s	63.540	749.486
EGNN-m	32.016	335.002
EGNN-t	0.786	12.492
ST_GNN	0.441	15.336
ST_TFN	0.597	13.709
ST_SE(3)TR	0.236	13.851
ST_EGNN	0.538	13.199
EqMotion	1.011	<u>4.893</u>
STGCN	0.062	4.919
AGL-STAN	**0.037**	5.734
ESTAG	<u>0.040</u>	**0.746**

Ablation studies

Table 4: Ablation studies ($\times 10^{-3}$) on MD17 dataset. Results averaged across 3 runs.

	Aspirin	Benzene	Ethanol	Malonaldehyde	Naphthalene	Salicylic	Toluene	Uracil
ESTAG	**0.063**	**0.003**	**0.099**	**0.101**	**0.068**	**0.047**	**0.079**	0.066
w/o EDFT	0.079	**0.003**	0.108	0.148	0.104	0.145	0.102	**0.063**
w/o Attention	0.087	0.004	0.104	0.112	0.129	0.095	0.097	0.078
w/o Equivariance	0.762	0.114	0.458	0.604	0.738	0.698	0.690	0.680
w/o Temporal	0.084	**0.003**	0.111	0.139	0.141	0.098	0.153	0.071

Table 5: MSE on Ethanol $w.r.t.$ the number of layers L.

L	1	2	3	4	5	6
MSE ($\times 10^{-4}$)	1.25	0.990	1.096	1.022	1.042	1.028

Without EDFT: considerably worse performance. wk (learnable) shown to be beneficial as a spectral filter

Without attention: slightly worse performance

Without equivariance: considerably worse performance

Without temporal pooling: slightly worse performance

Paper analysis: contributions and advantages

- Time: modelling non-Markovian features, capturing periodicity, via EDFT and attention mechanism
- Space: Euclidean symmetry
- Good overall performance

Paper analysis: limitations and criticism

- Limited equivariance: missing embedded physical laws, e.g. no conservation of energy
- Limited benchmarks
- Inconsistent baseline comparisons (due to modifications)
- Ablation study interpretations (limited time effects?)
- Visualization as cherry-picking?

Thanks for listening!

Questions?