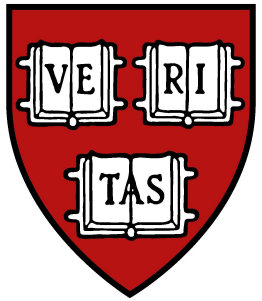
Life after Message Passing: Local Equivariant Interatomic Potentials

Preprint: arxiv.org/abs/2204.05249

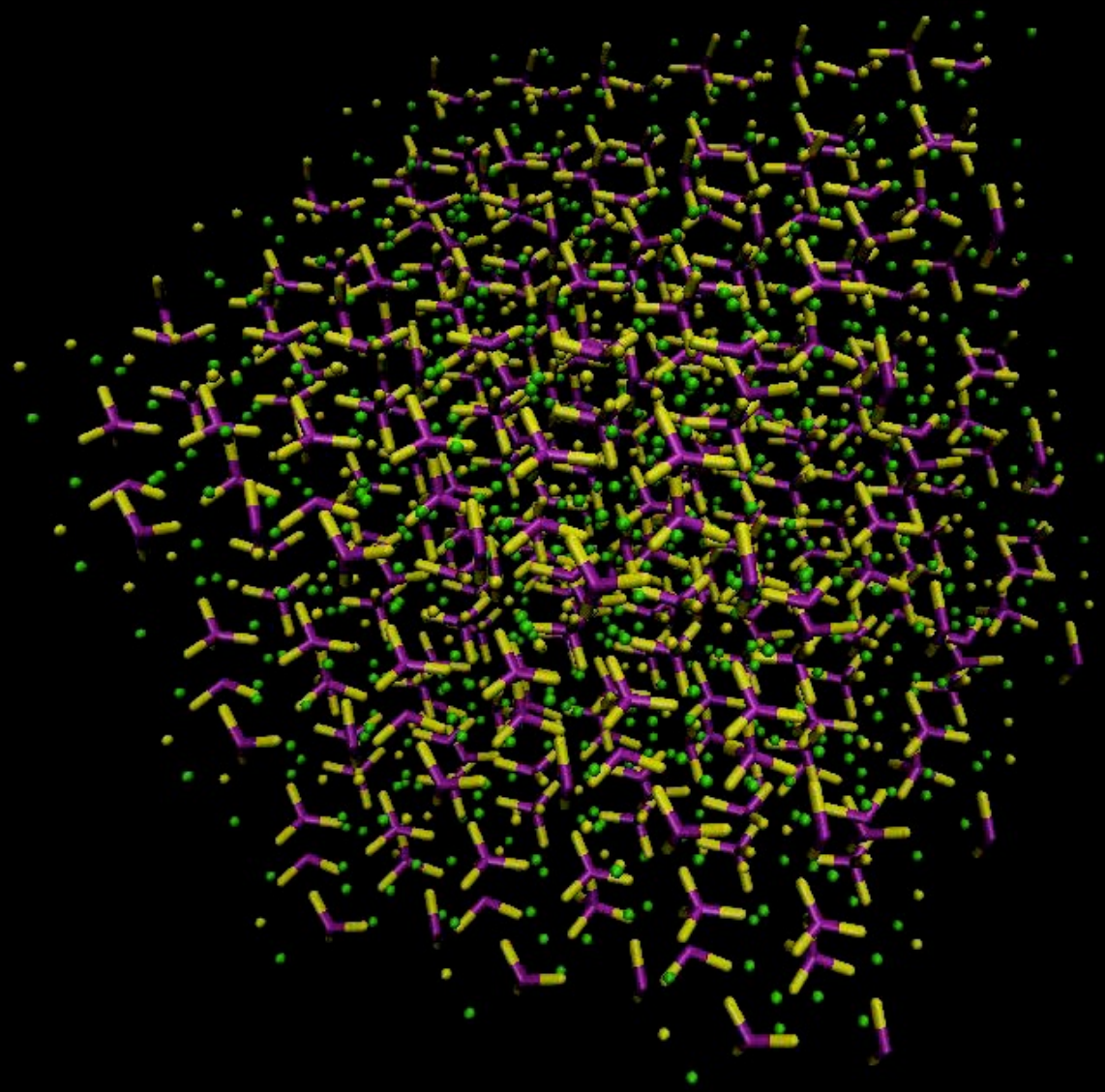
Albert Musaelian and Simon Batzner

Kozinsky Lab

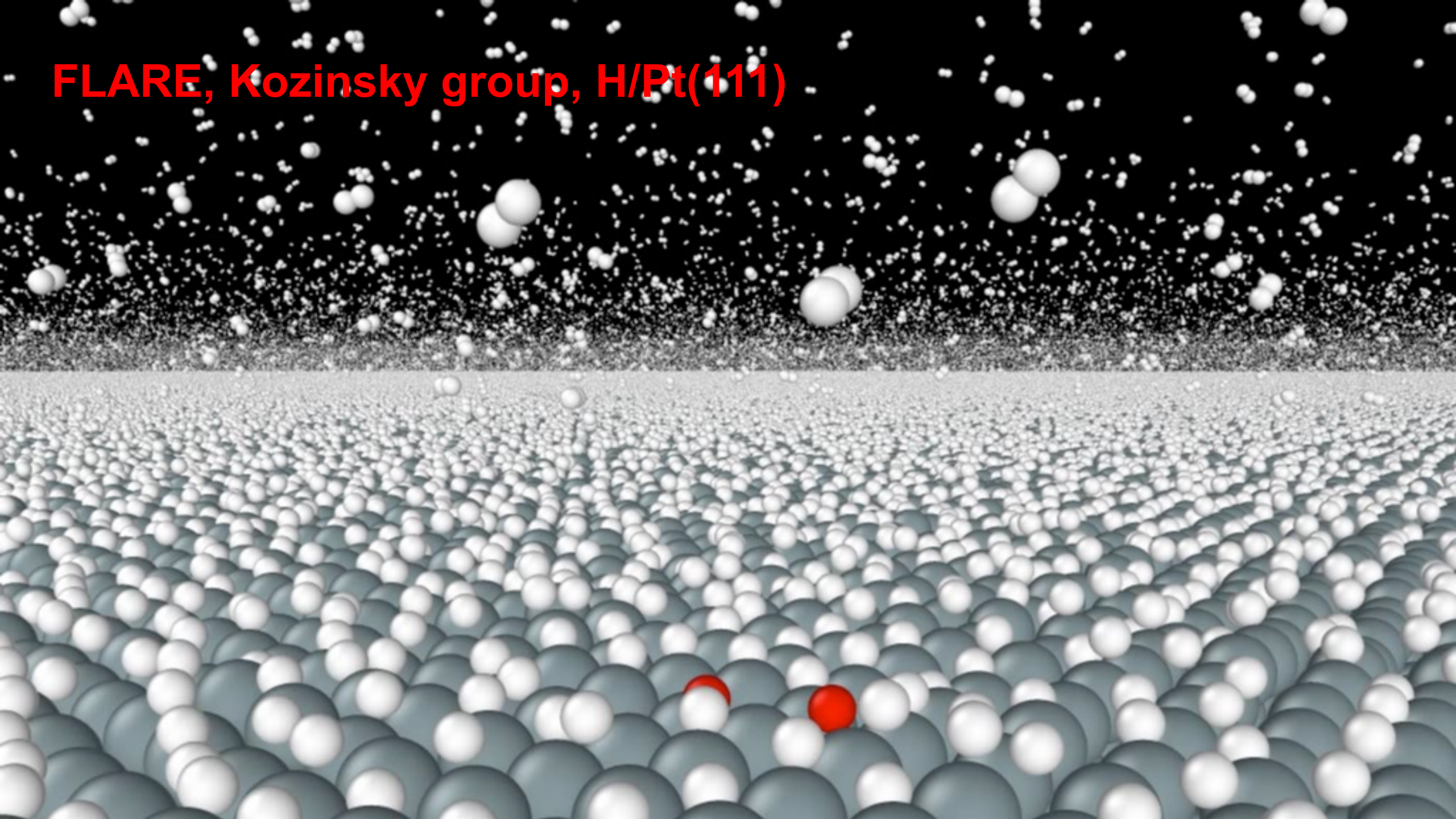


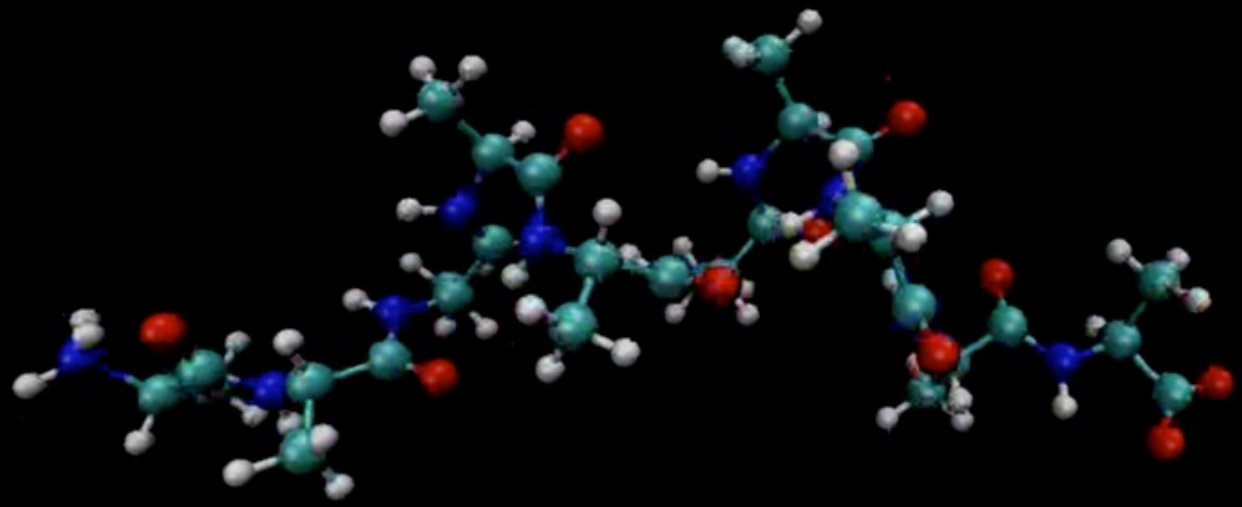
Molecular Dynamics

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i(\vec{r}_1, \ldots, \vec{r}_N) = -\frac{\partial E(\vec{r}_1, \ldots, \vec{r}_N)}{\partial \vec{r}_i}$$



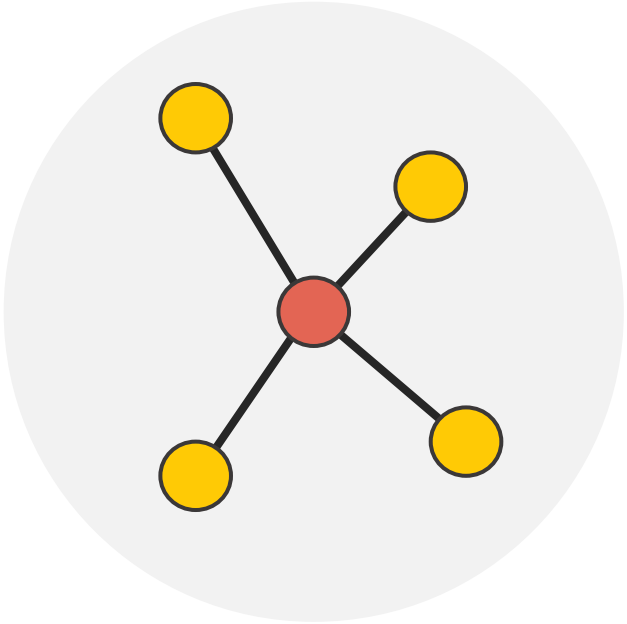
NequIP, Kozinsky group, LiPS


FLARE, Kozinsky group, H/Pt(111)



NequIP, Kozinsky group, Deca-Alanine

Molecular dynamics

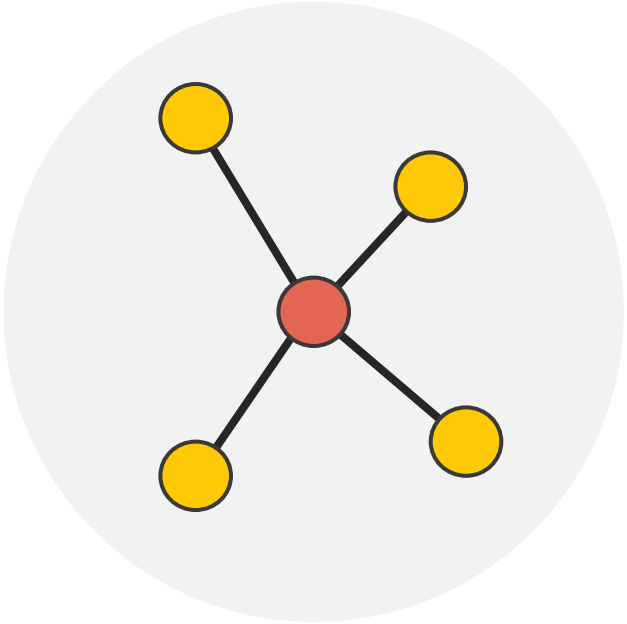


Potential energy function

$$E \longrightarrow \vec{F}_i = -\frac{\partial E}{\partial \vec{r}_i}$$

Integrate equations of motion: $m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i$

Molecular dynamics



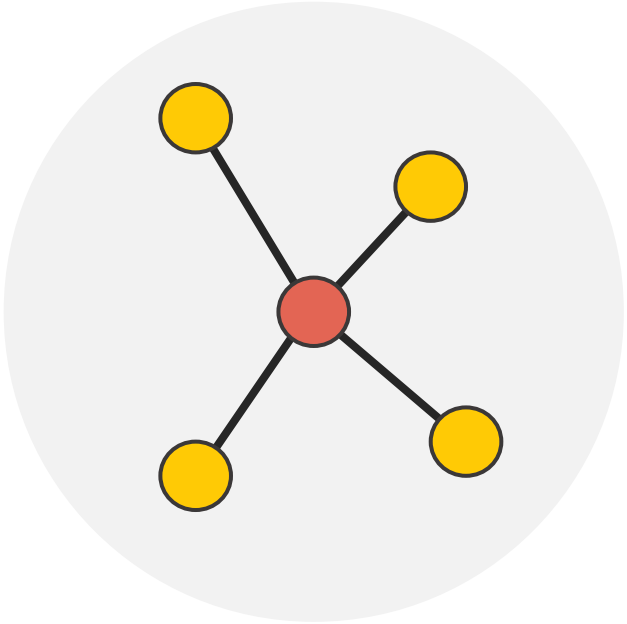
Potential energy function

$$E \longrightarrow \vec{F}_i = -\frac{\partial E}{\partial \vec{r}_i}$$

Integrate equations of motion: $m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i$

Challenge: Have to integrate billions to trillions of times

Molecular dynamics



Potential energy function

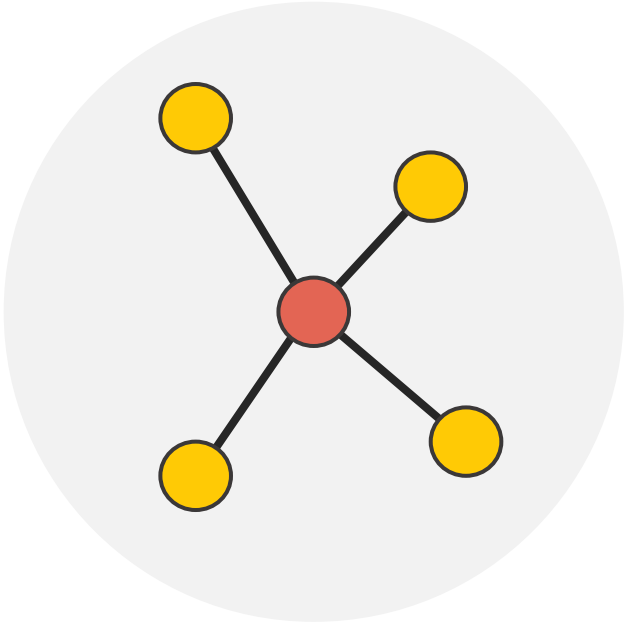
?

$$E \longrightarrow \vec{F}_i = -\frac{\partial E}{\partial \vec{r}_i}$$

Integrate equations of motion: $m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i$

Molecular dynamics

- Empirical
- Simple and fast
- Up to trillions of atoms
- Up to milliseconds of simulation



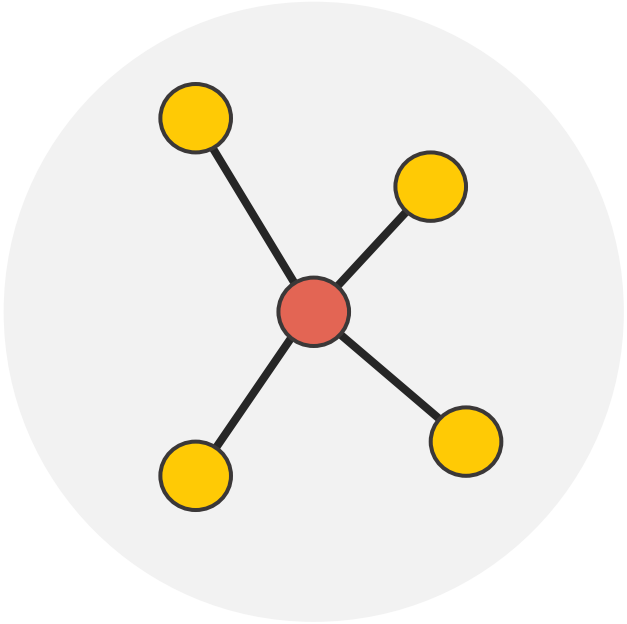
Empirical

$$E \longrightarrow \vec{F}_i = -\frac{\partial E}{\partial \vec{r}_i}$$

Integrate equations of motion: $m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i$

Molecular dynamics

- From first principles
- Often good agreement with experiment without fitting
- Very expensive and prohibitive scaling
- Limited to hundreds of atoms and 100s ps of simulations



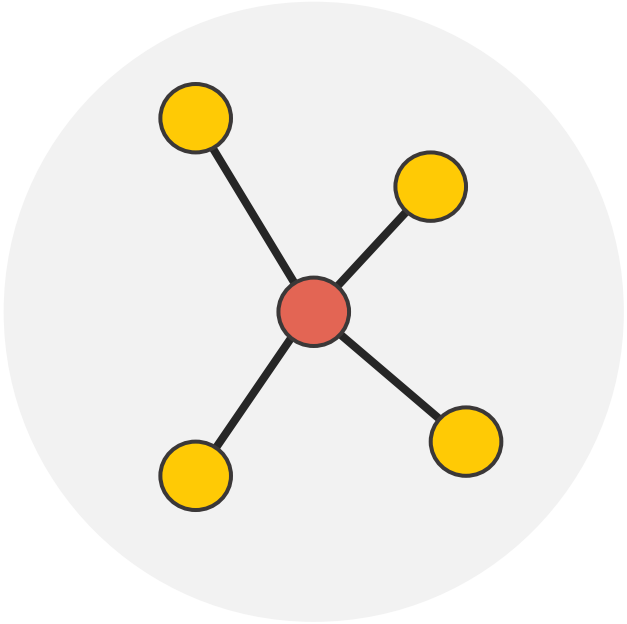
Quantum Mechanics

$$E \longrightarrow \vec{F}_i = -\frac{\partial E}{\partial \vec{r}_i}$$

Integrate equations of motion: $m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i$

Molecular dynamics

- Fit to a limited library of quantum calculations
- Promises near-quantum accuracy
- Significantly improved speed and scaling
- Classical potential speeds and scales at *ab-initio* accuracy?

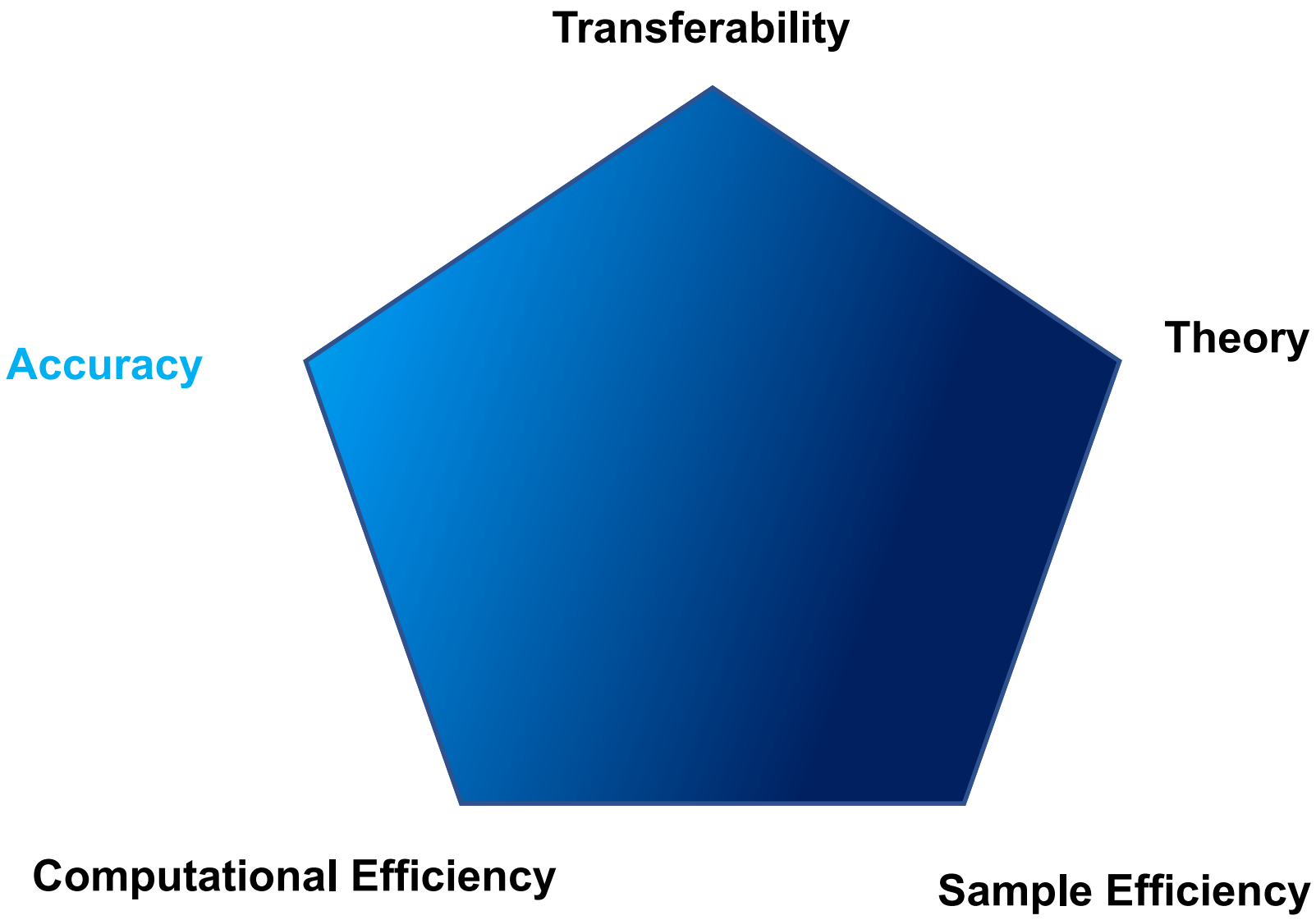


ML

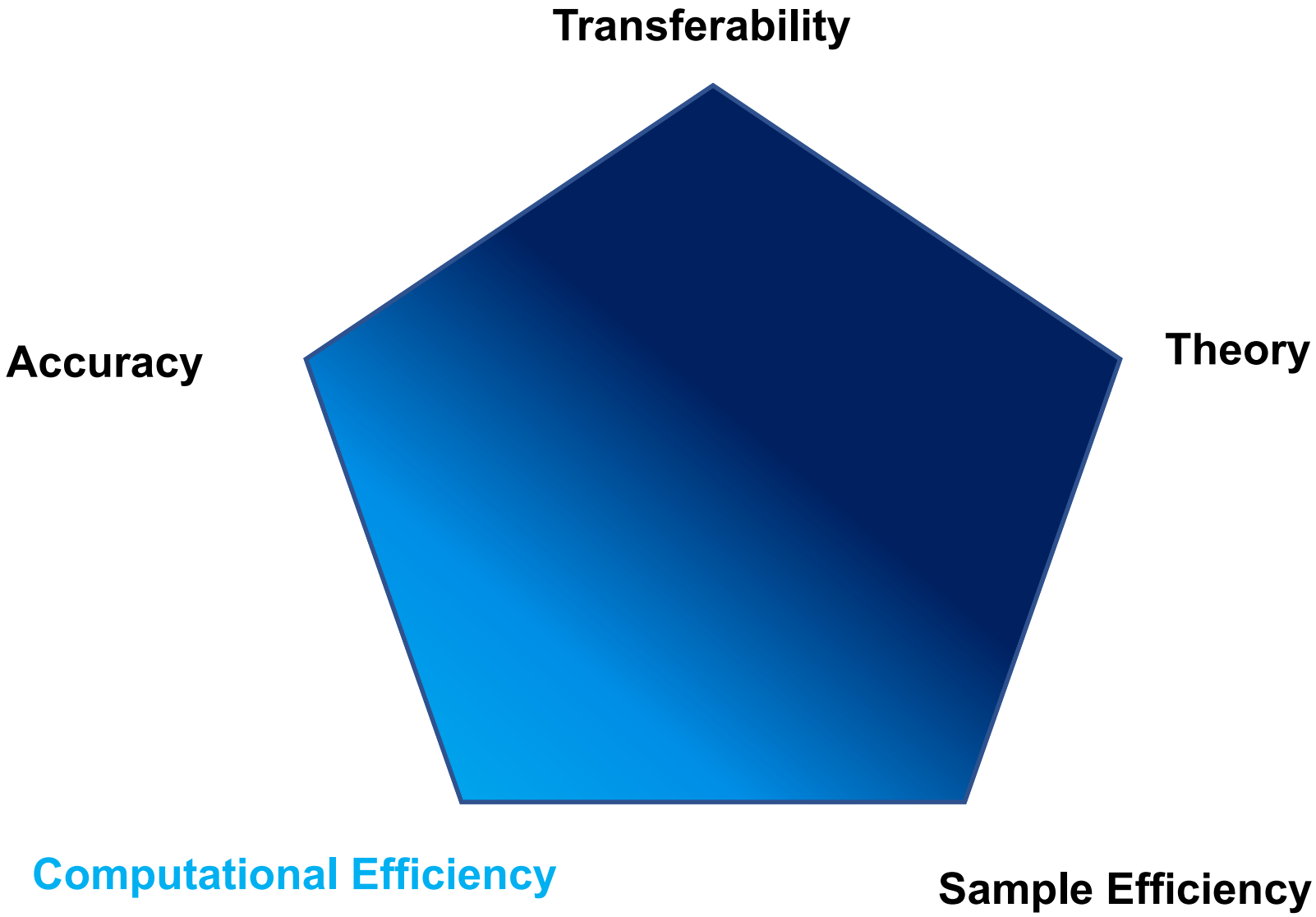
$$E \longrightarrow \vec{F}_i = -\frac{\partial E}{\partial \vec{r}_i}$$

Integrate equations of motion: $m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i$

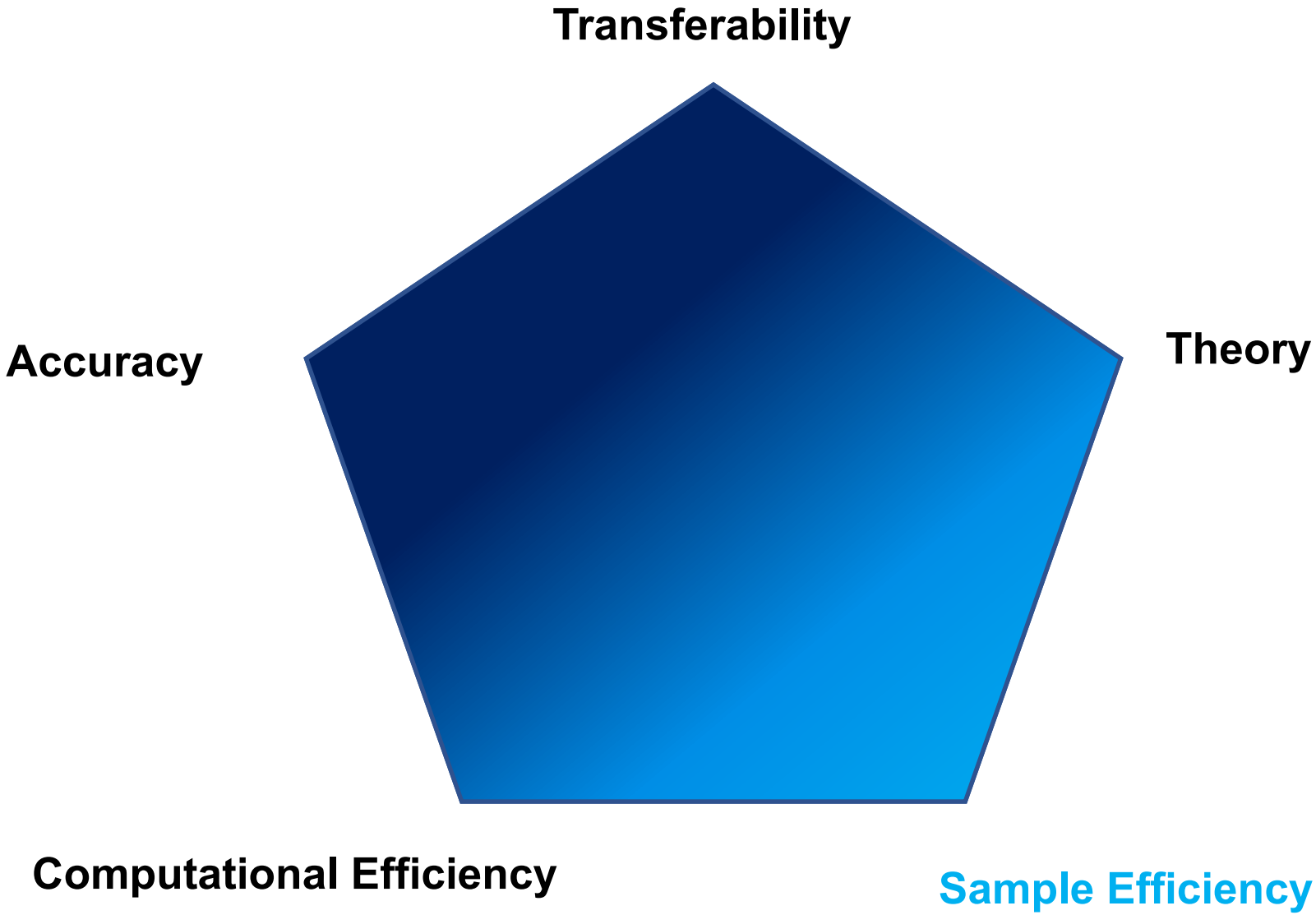
Five Properties of successful Machine Learning Interatomic Potentials



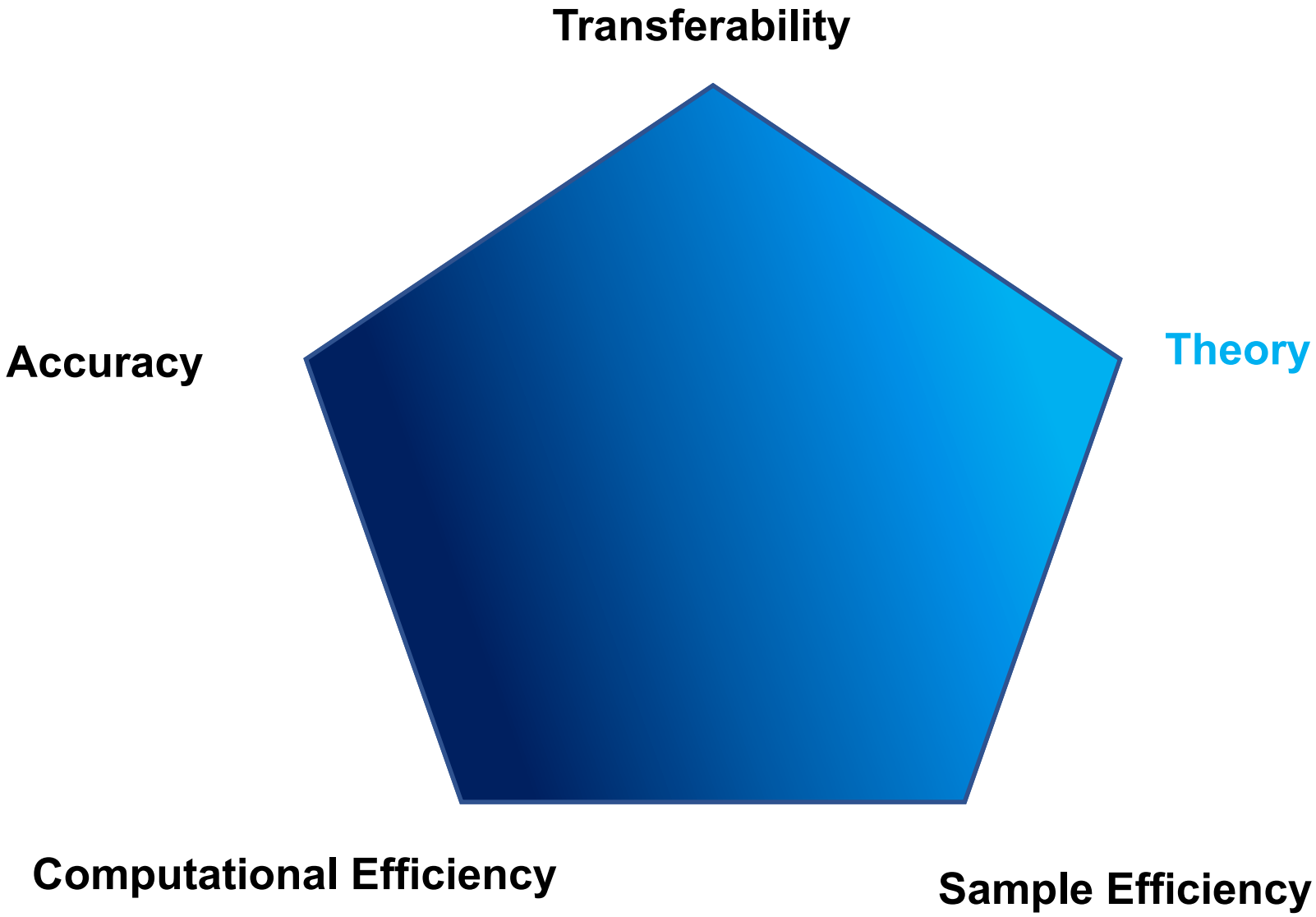
Five Properties of successful Machine Learning Interatomic Potentials



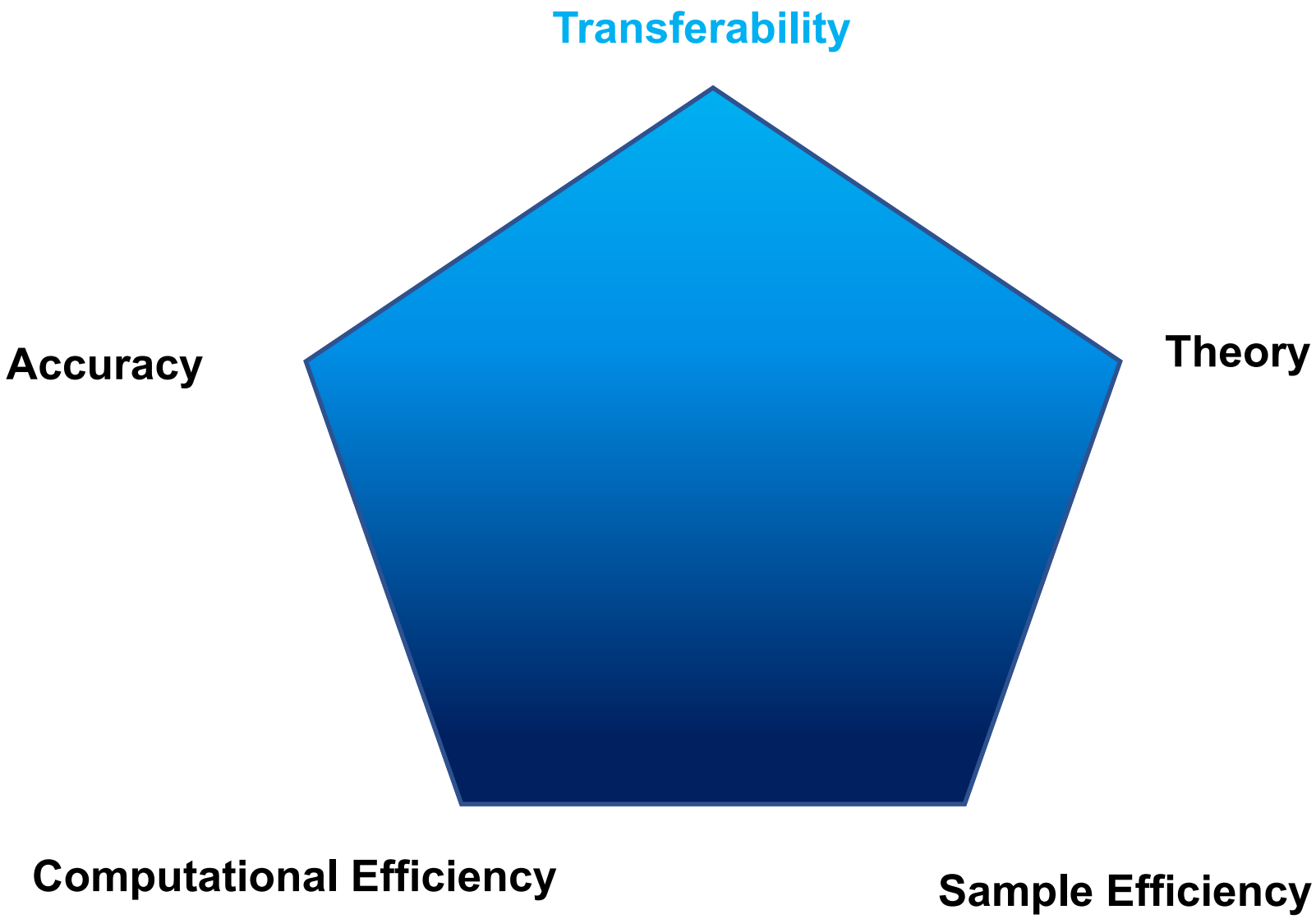
Five Properties of successful Machine Learning Interatomic Potentials



Five Properties of successful Machine Learning Interatomic Potentials

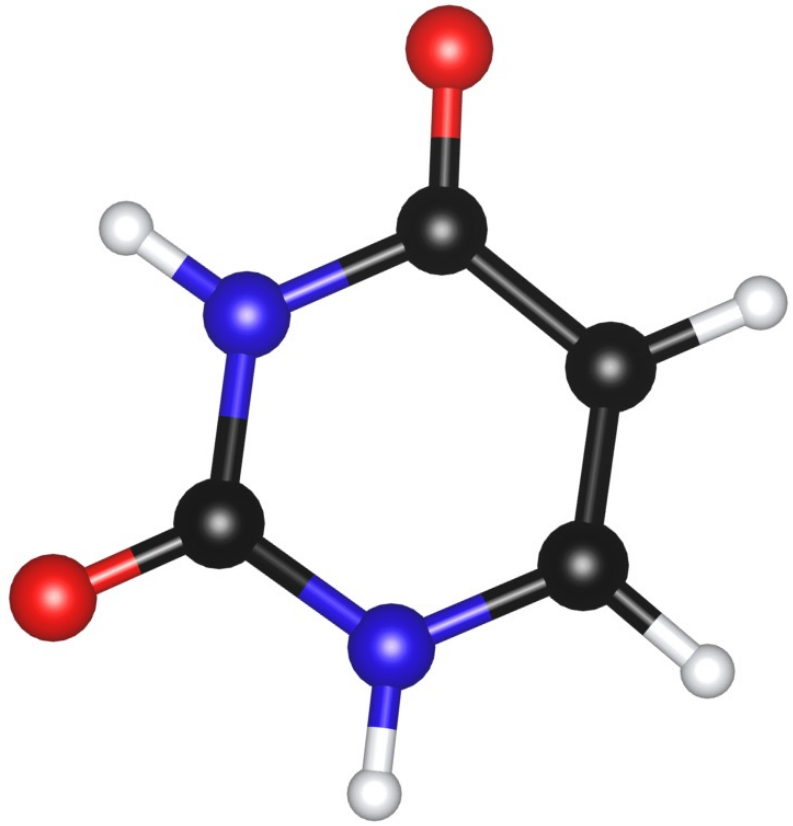


Five Properties of successful Machine Learning Interatomic Potentials

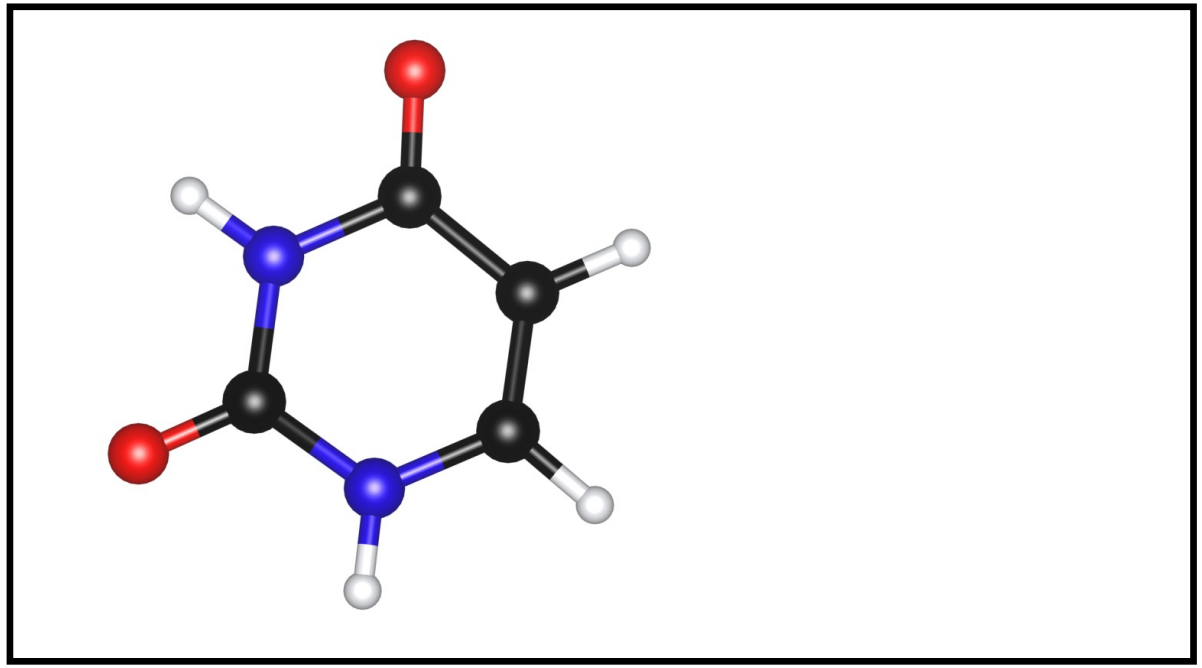


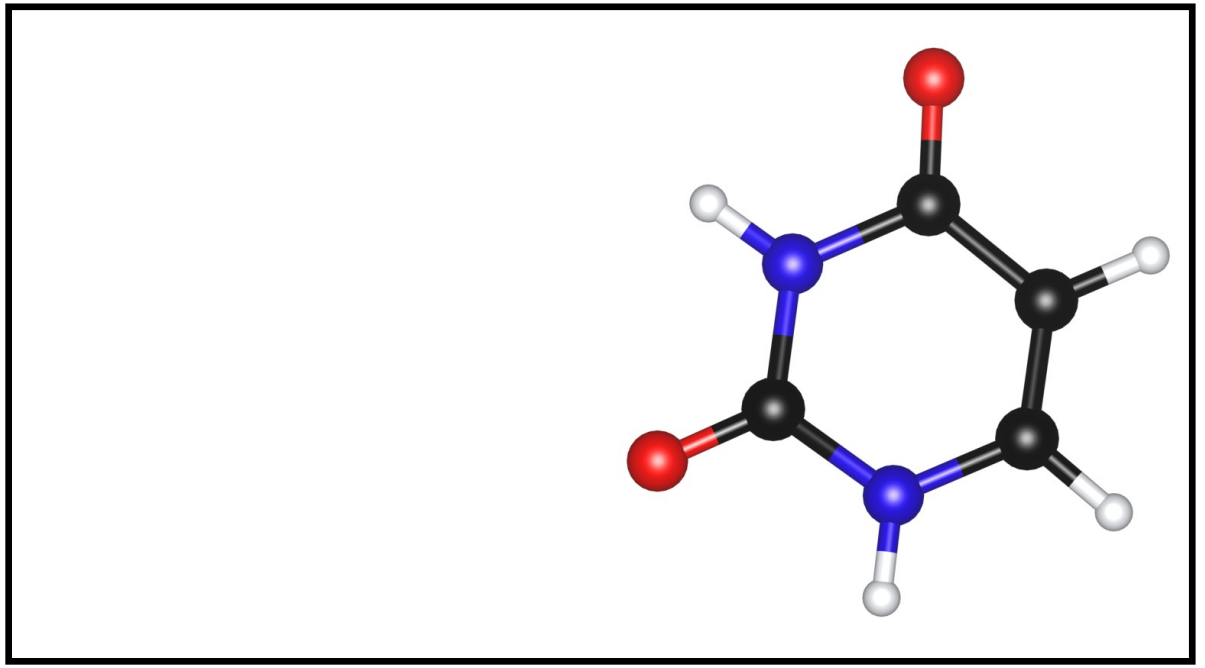
How to represent atomistic systems?

```
12
-260120.41022582943
C    1.63438356  0.29588831  -0.06029892
C    1.44408771  -1.03792413 0.04112652
N    0.15607651  -1.58651275 0.11493441
C    -1.00154586 -0.75181433 0.08175814
N    -0.81072938 0.645473    -0.02477061
C    0.49544837  1.21779532  -0.11049406
O    -2.11299038 -1.2549146  0.14927307
O    0.54437819  2.43752091  -0.19898519
H    2.28646928  -1.74881243 0.07996217
H    0.04292334  -2.59002107 0.16531753
H    -1.62925491 1.21958722  -0.04033683
H    2.64457191  0.71408144  -0.09981888
```

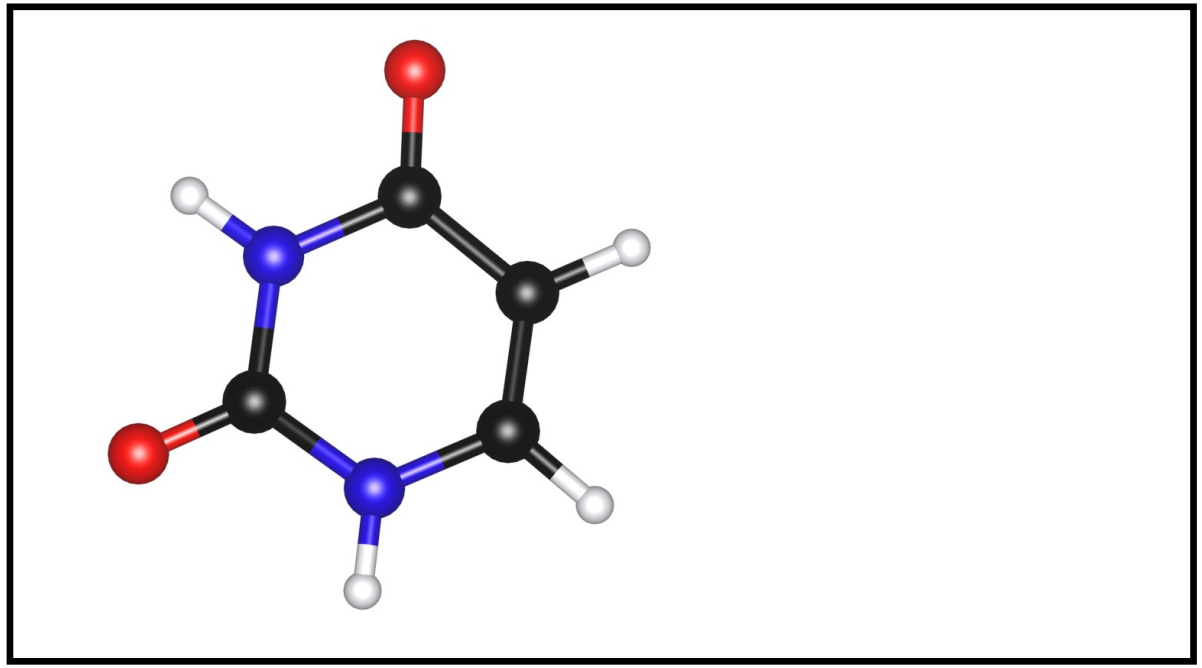


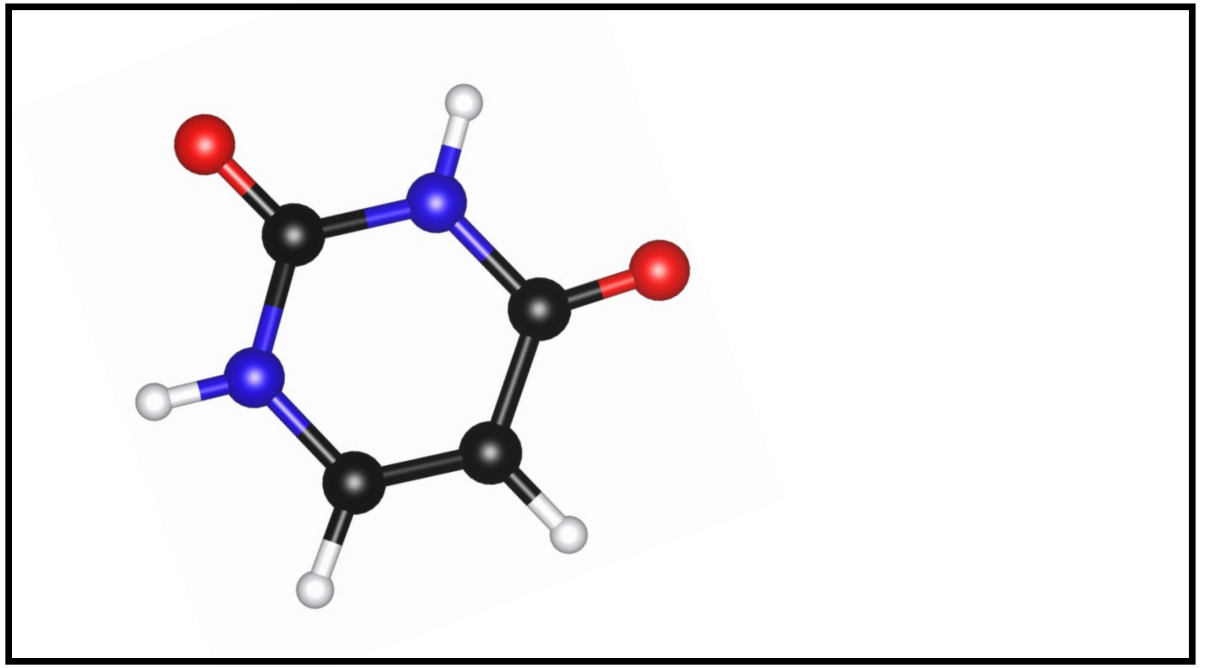
1. Translations



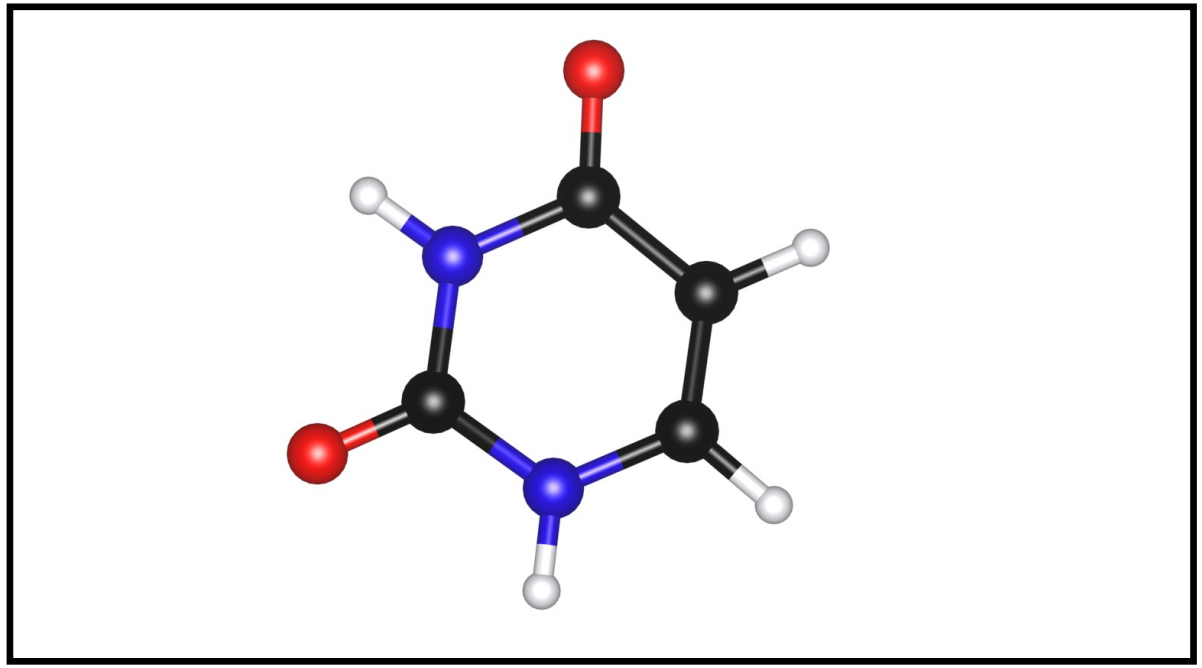


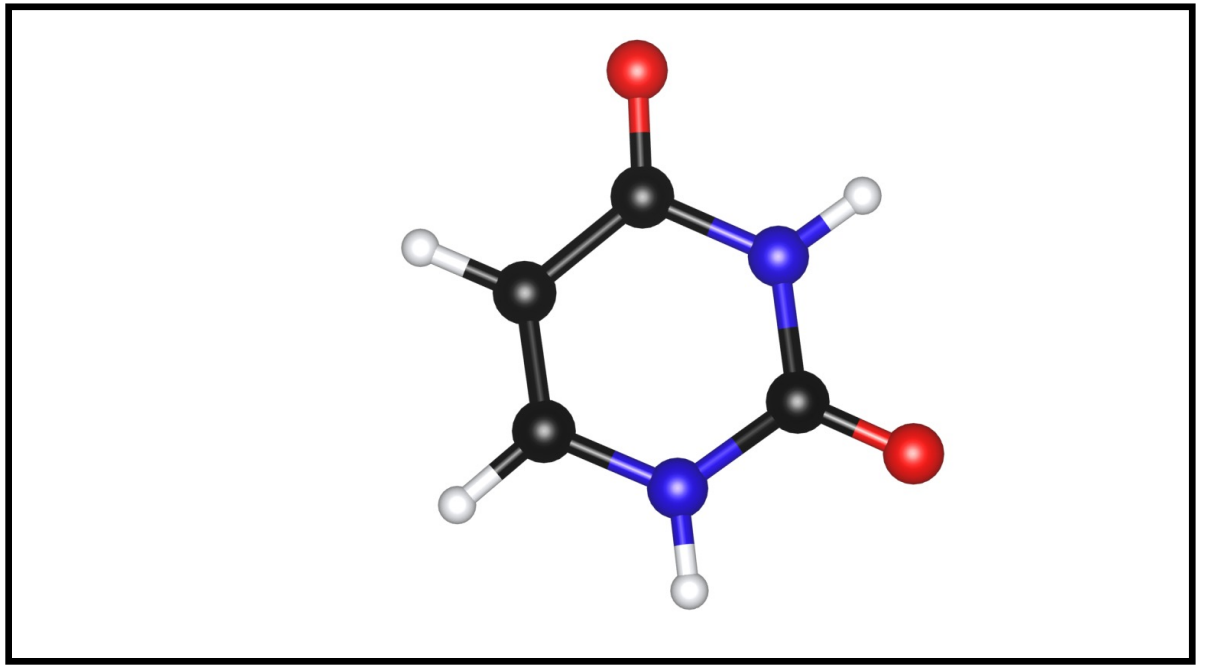
2. Rotations



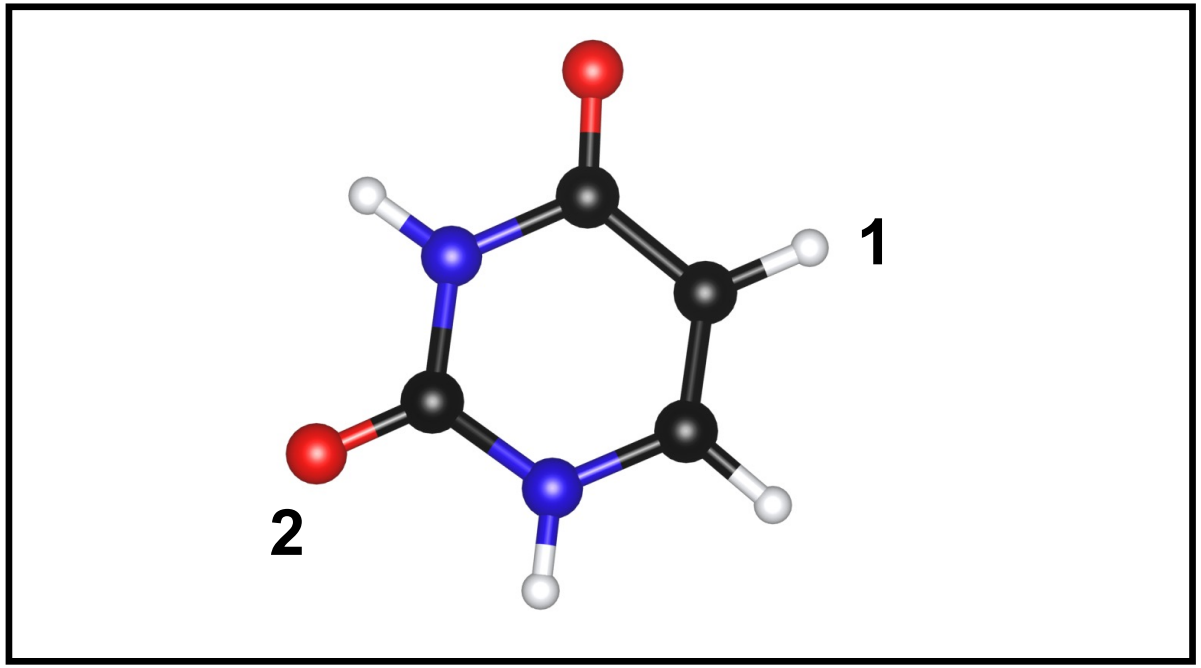


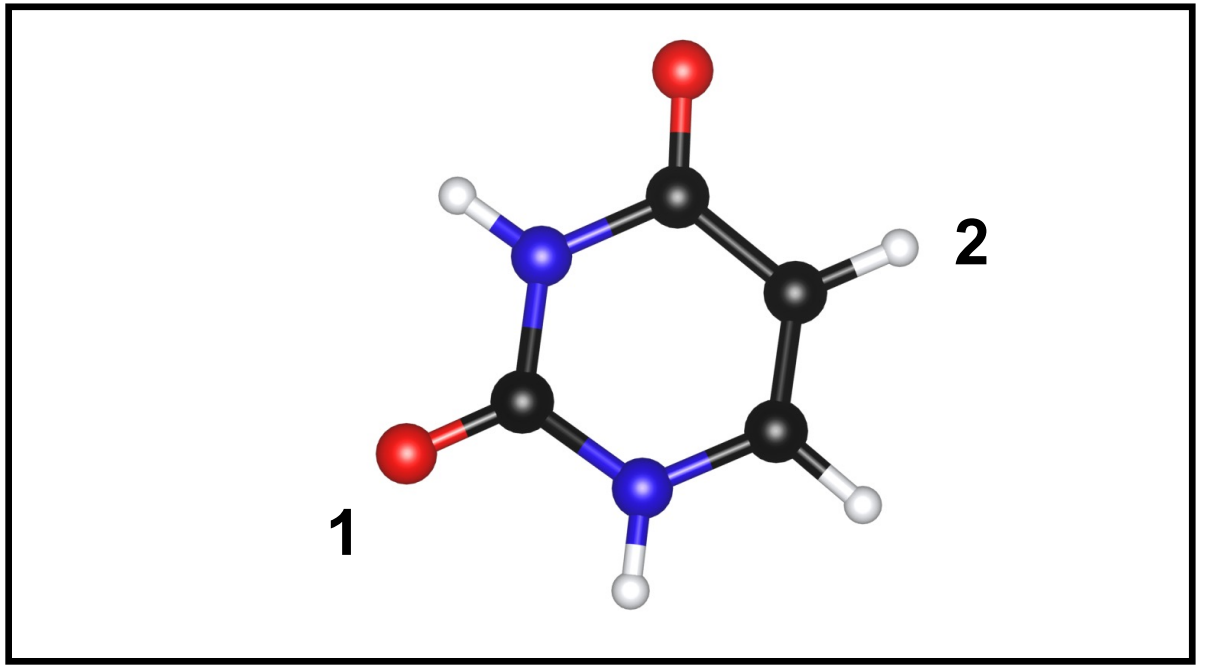
3. Reflections



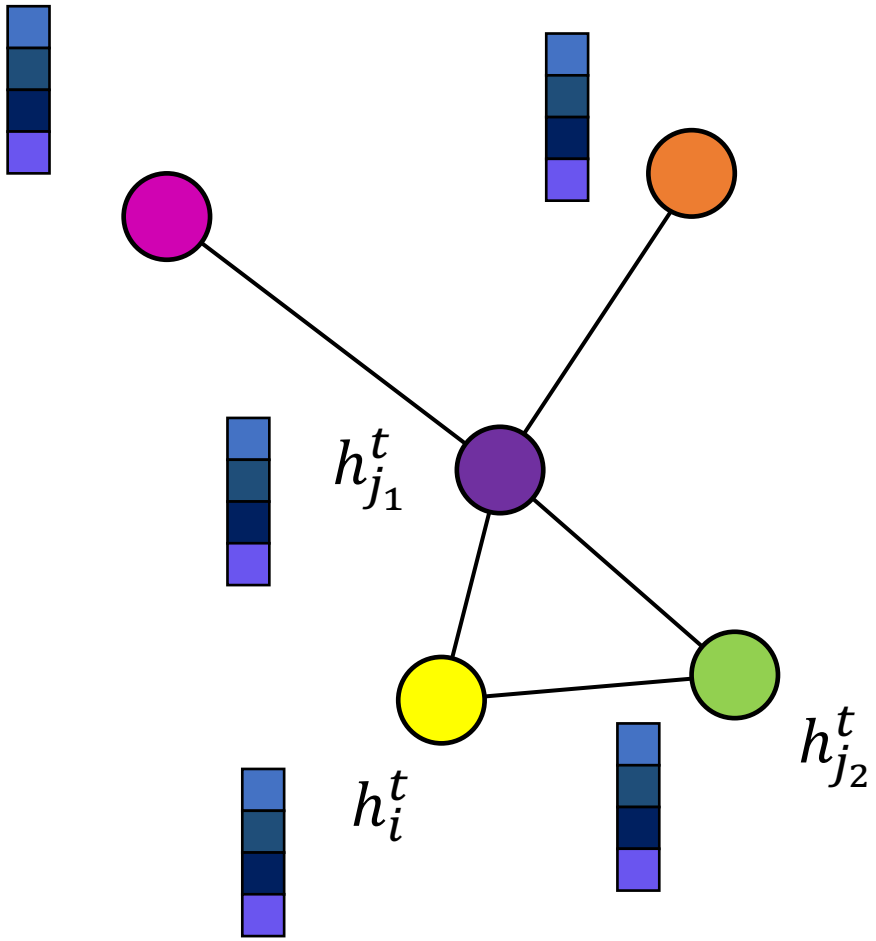


4. Permutation of atom indexing





Message Passing Neural Networks



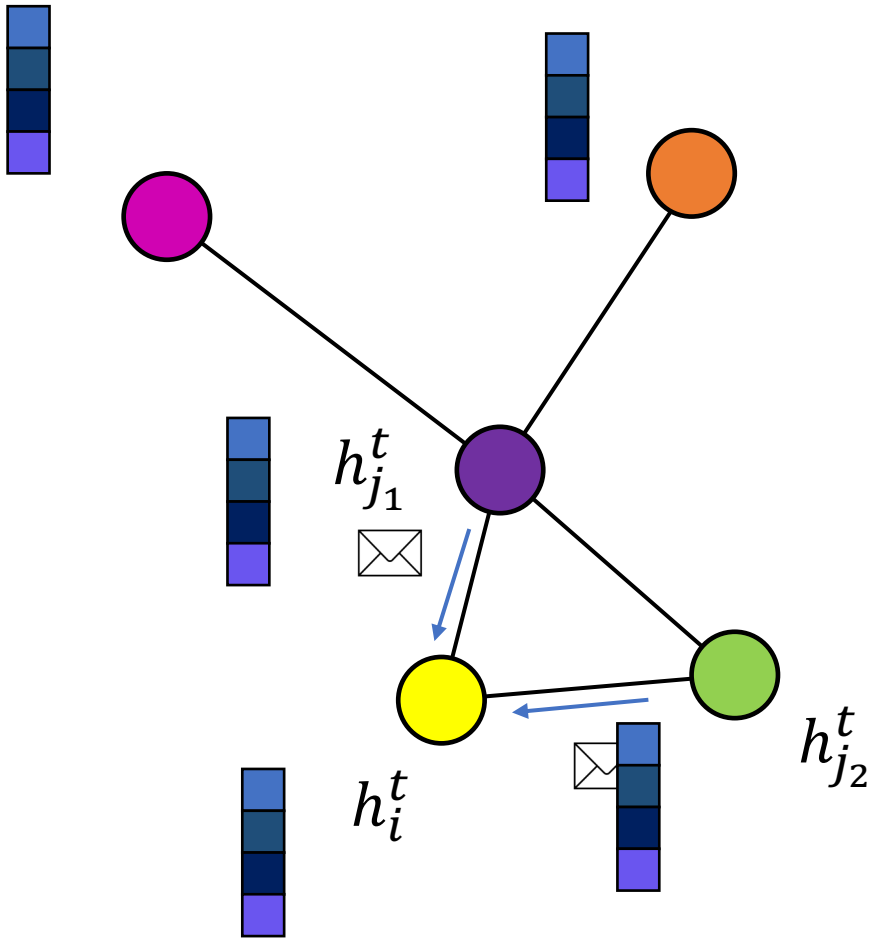
t: layer

i, j: atom indices

$\mathrm{h} \in \mathbb{R}^R$

[1] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017, July). Neural message passing for quantum chemistry. In *International Conference on Machine Learning* (pp. 1263-1272). PMLR.

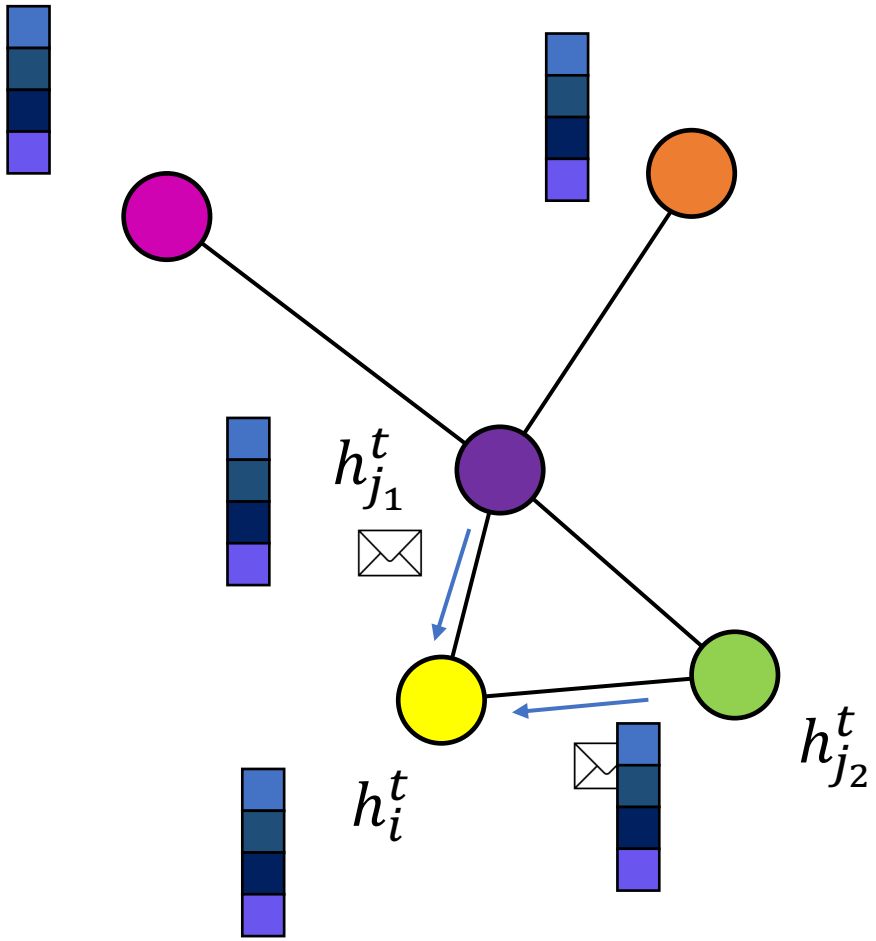
Message Passing Neural Networks



$$m_i^{t+1} = \sum_{j \in N(i)} M_t \left(h_i^t, h_j^t, e_{ij}\right)$$

[1] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017, July). Neural message passing for quantum chemistry. In *International Conference on Machine Learning* (pp. 1263-1272). PMLR.

Message Passing Neural Networks

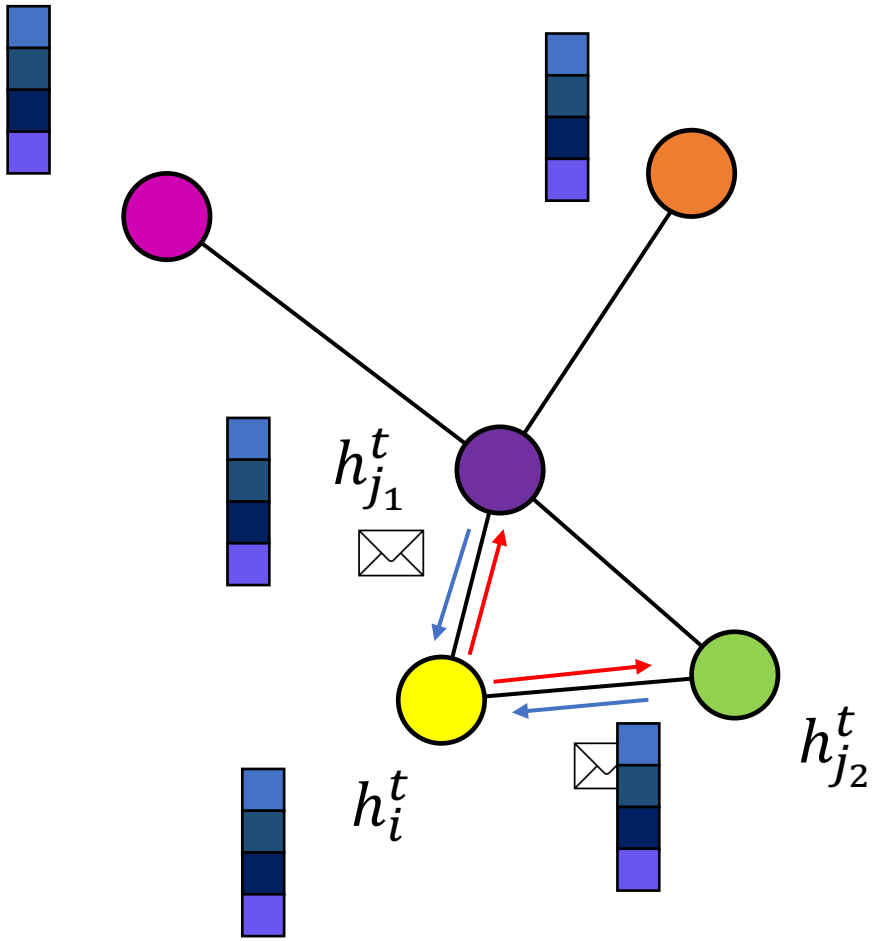


$$m_i^{t+1} = \sum_{j \in N(i)} M_t\left(h_i^t, h_j^t, e_{ij}\right)$$

$$h_i^{t+1} = U_t(h_i^t, m_i^{t+1})$$

[1] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017, July). Neural message passing for quantum chemistry. In *International Conference on Machine Learning* (pp. 1263-1272). PMLR.

Message Passing Neural Networks

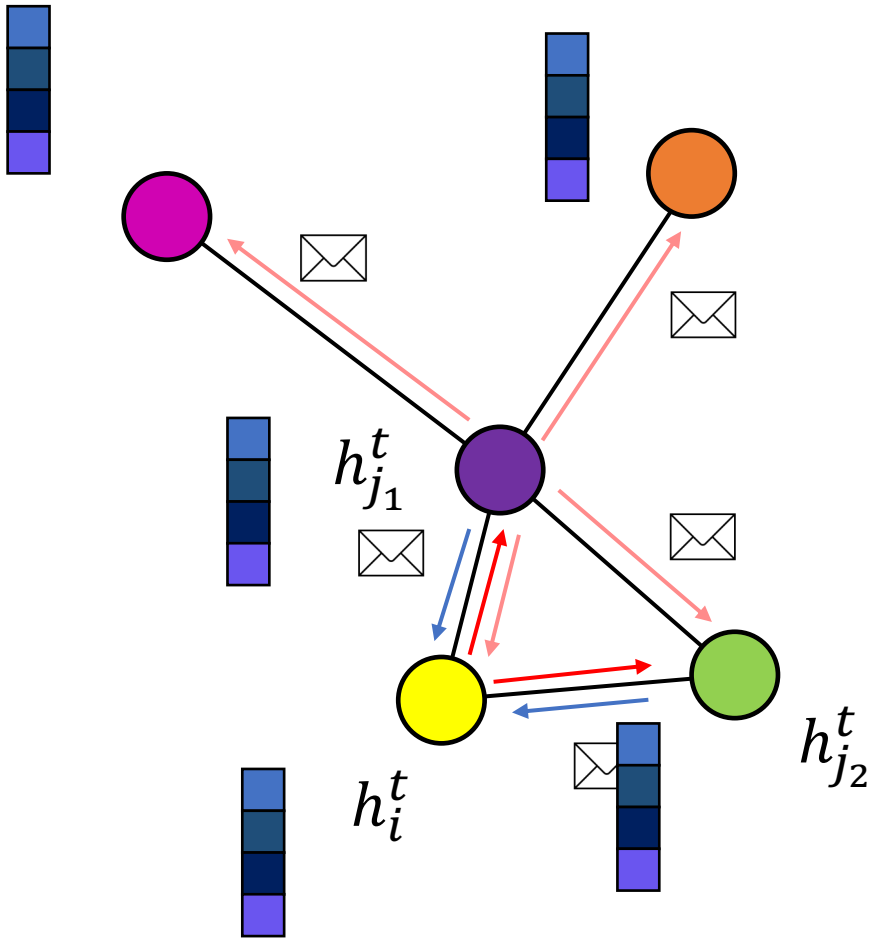


$$m_i^{t+1} = \sum_{j \in N(i)} M_t\left(h_i^t, h_j^t, e_{ij}\right)$$

$$h_i^{t+1} = U_t(h_i^t, m_i^{t+1})$$

[1] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017, July). Neural message passing for quantum chemistry. In *International Conference on Machine Learning* (pp. 1263-1272). PMLR.

Message Passing Neural Networks



$$m_i^{t+1} = \sum_{j \in N(i)} M_t\left(h_i^t, h_j^t, e_{ij}\right)$$

$$h_i^{t+1} = U_t(h_i^t, m_i^{t+1})$$

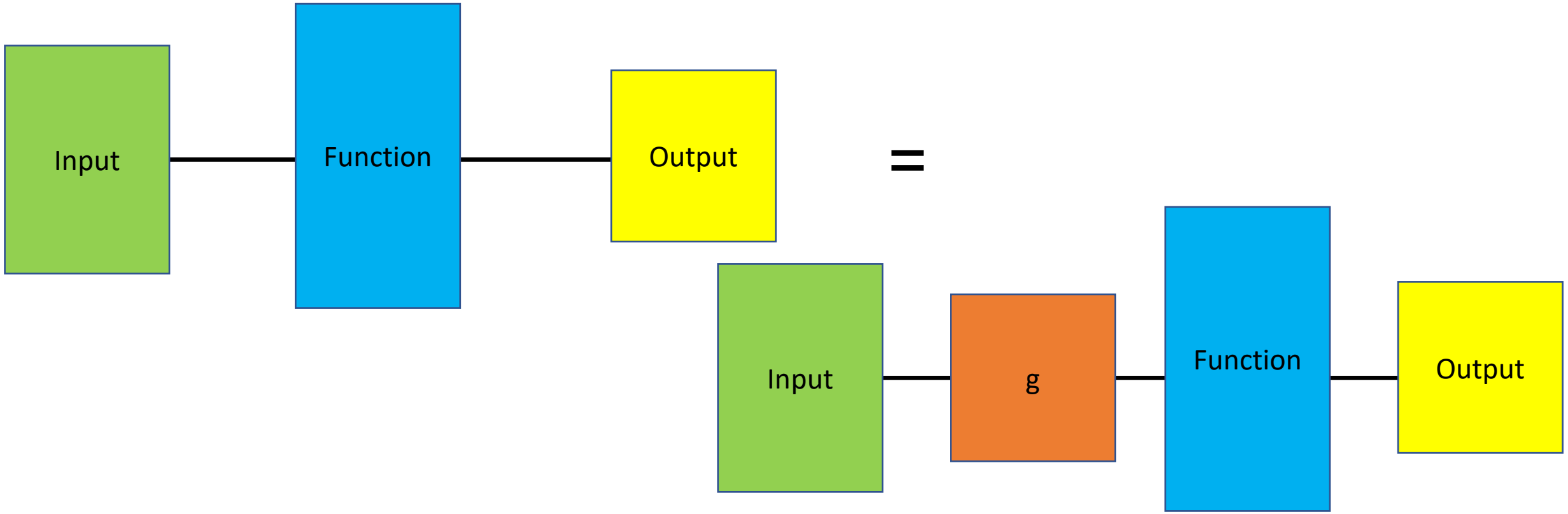
Existing Message Passing Neural Networks (SchNet, DimeNet, PhysNet, …) are invariant to E(3)

[1] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017, July). Neural message passing for quantum chemistry. In *International Conference on Machine Learning* (pp. 1263-1272). PMLR.

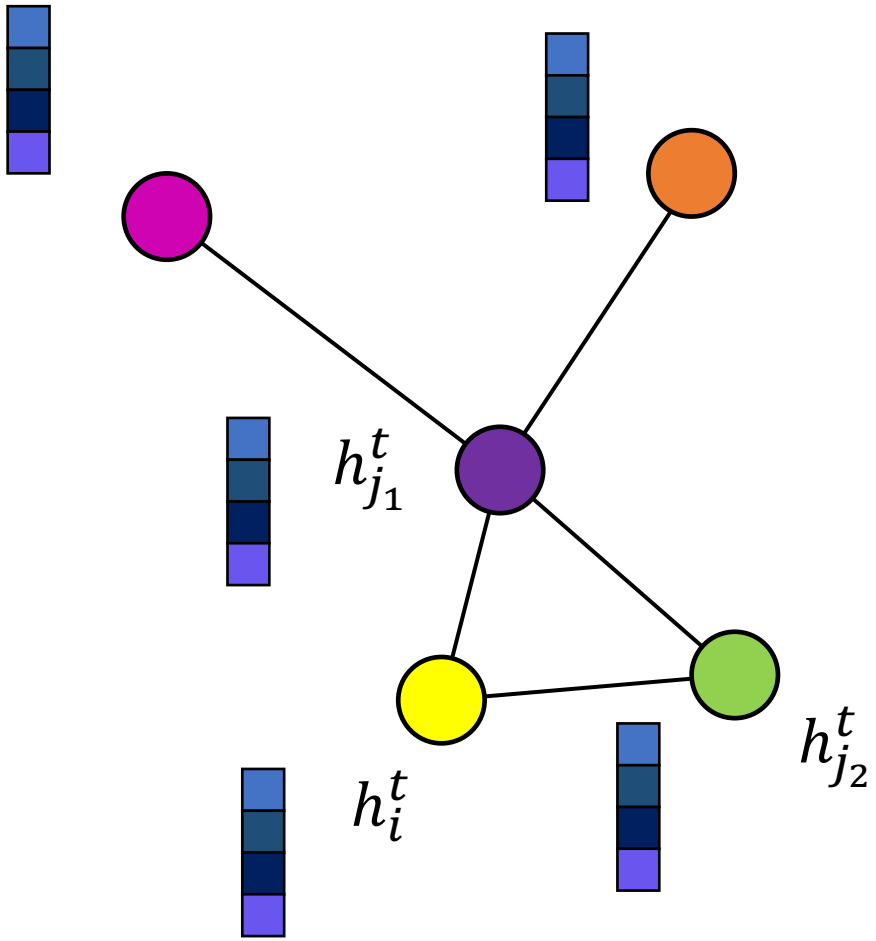
Invariance

$f: X \to Y$ is said to be **invariant** w.r.t. the action of the group G, if $\forall g \in G$ and $\forall x \in X$:

$$f(D_X(g)x) = If(x)$$



Message Passing Neural Networks



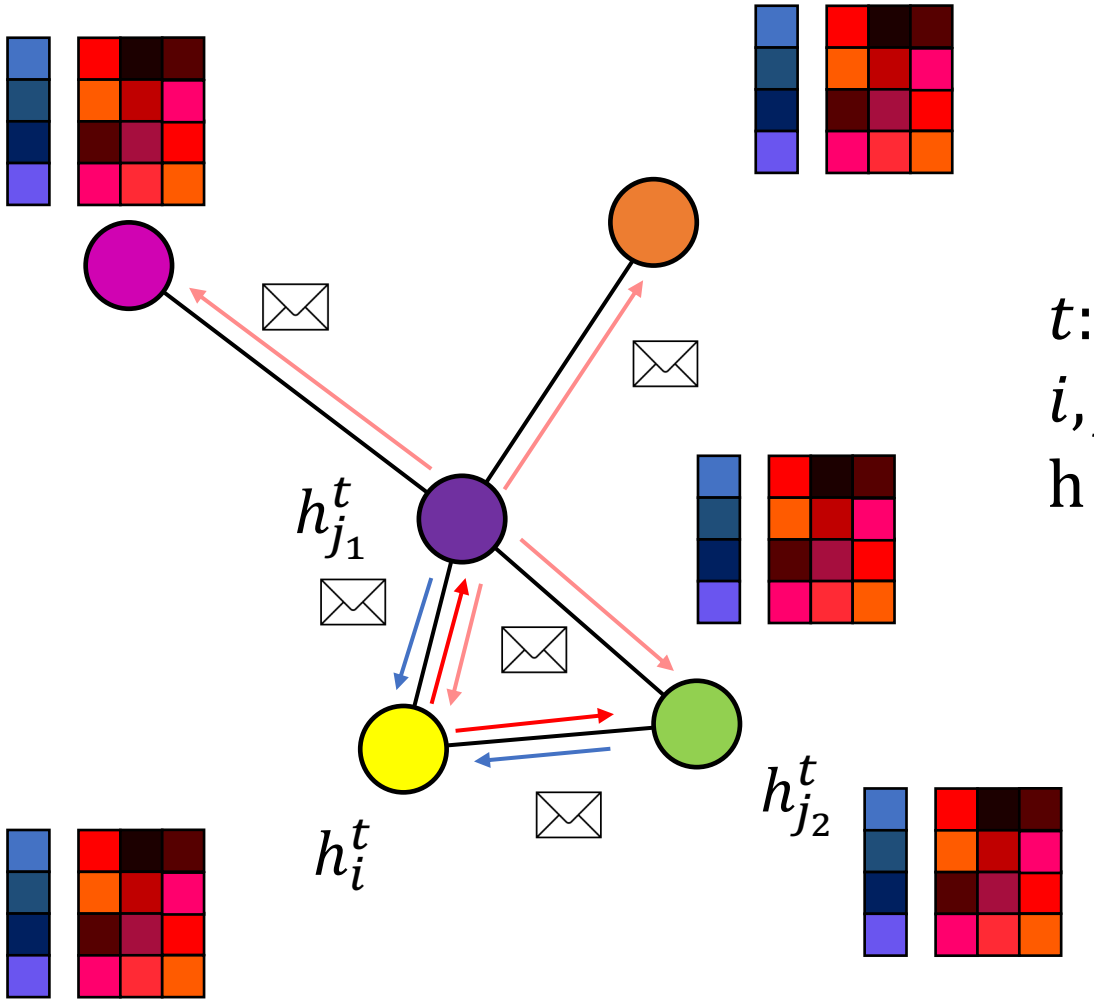
t: layer

i, j: atom indices

$\mathrm{h} \in \mathbb{R}^F$

[1] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017, July). Neural message passing for quantum chemistry. In *International Conference on Machine Learning* (pp. 1263-1272). PMLR.

Equivariant Message Passing Neural Networks [1, 2, 3]



t: layer

i, j: atom indices

$\mathrm{h} \in \mathbb{R}^c \oplus \mathbb{R}^{c\times 3} \oplus \cdots$

[1] Thomas, N., Smidt, T., et al. *arXiv preprint arXiv:1802.08219*.

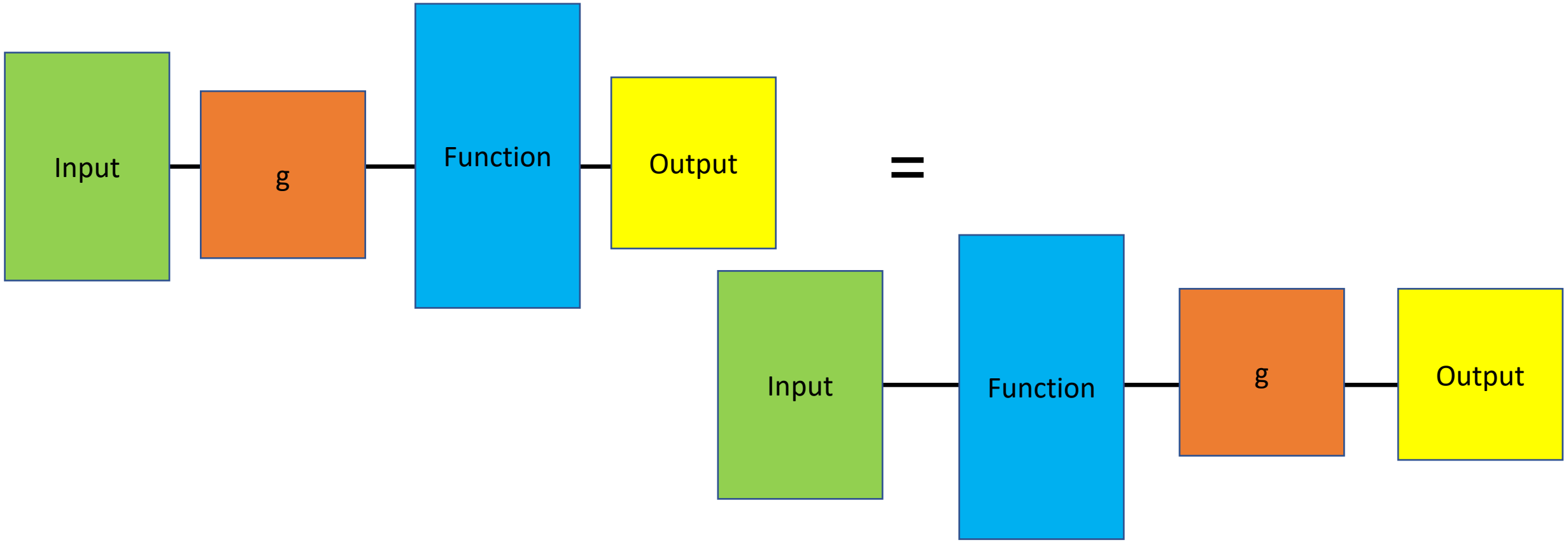
[2] Weiler, M., Geiger, M., et al. *Advances in Neural Information Processing Systems*, *31*

[3] Kondor, R., Lin, Z., & Trivedi, S. (2018*Advances in Neural Information Processing Systems*, *31*.

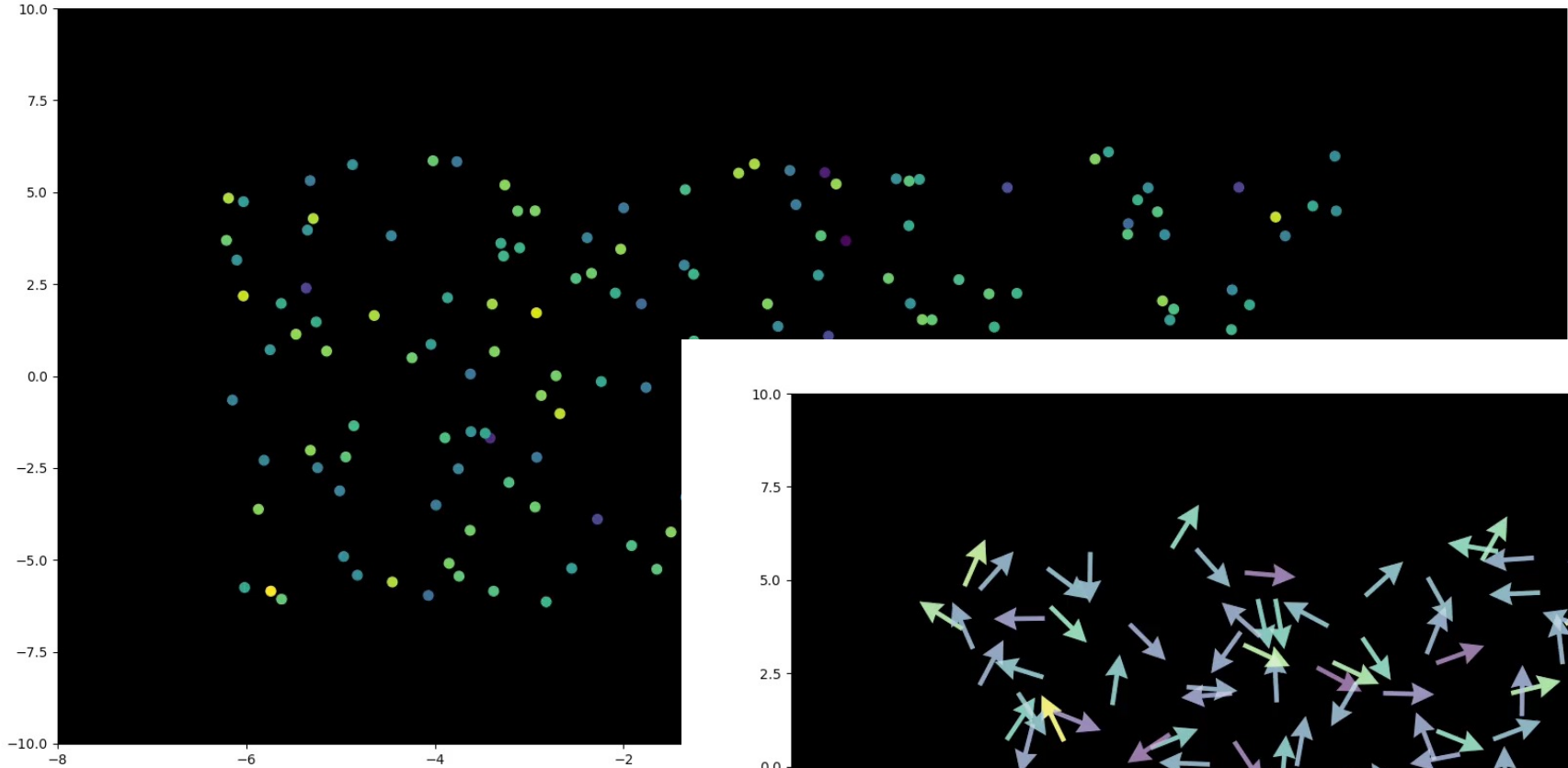
Equivariance

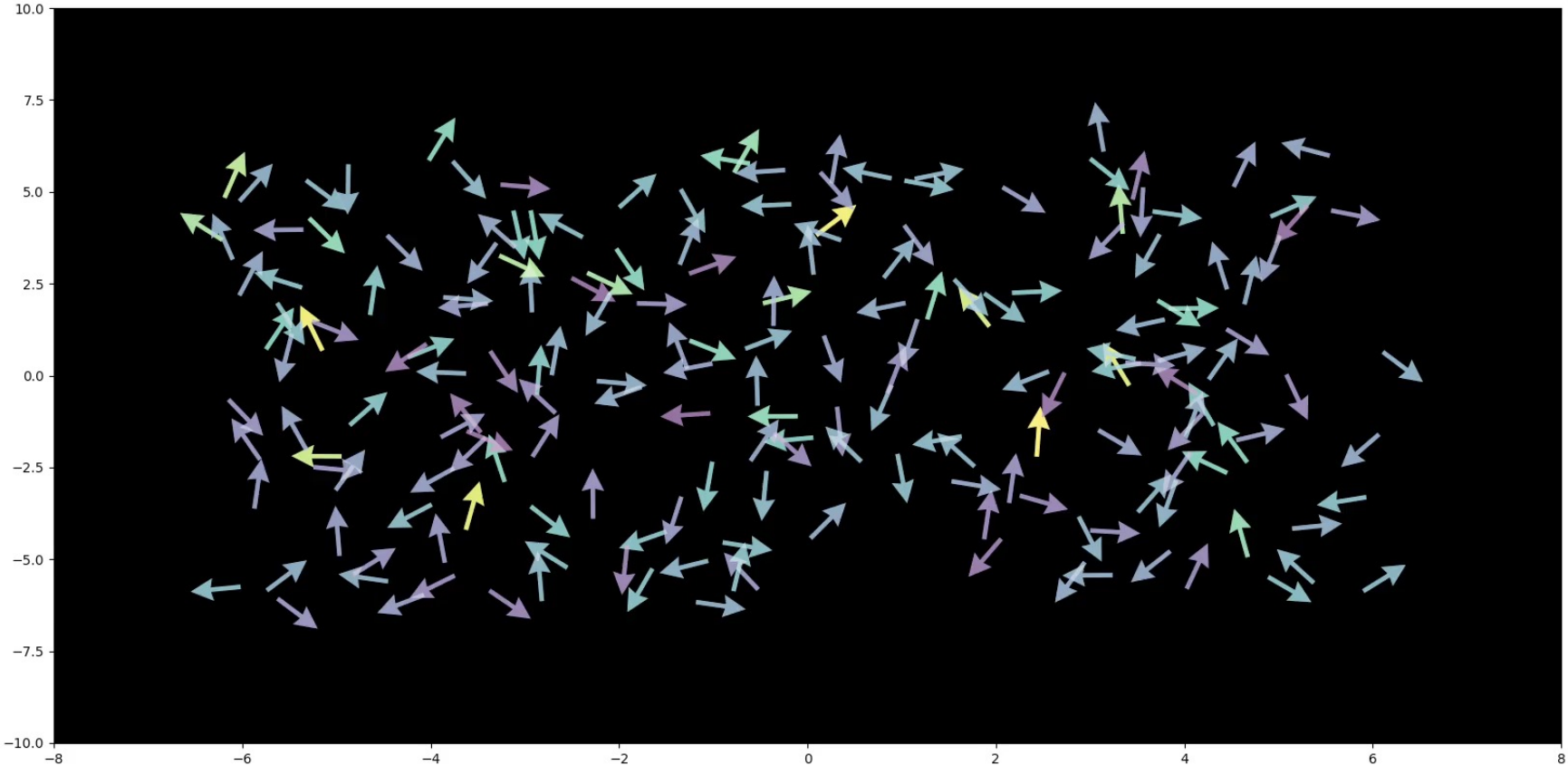
$f: X \to Y$ is said to be **equivariant** w.r.t. the action of the group G, if $\forall g \in G$ and $\forall x \in X$:

$$f(D_X(g)x) = D_Y(g)f(x)$$



Tensor Features transform with the geometry under E(3) group actions

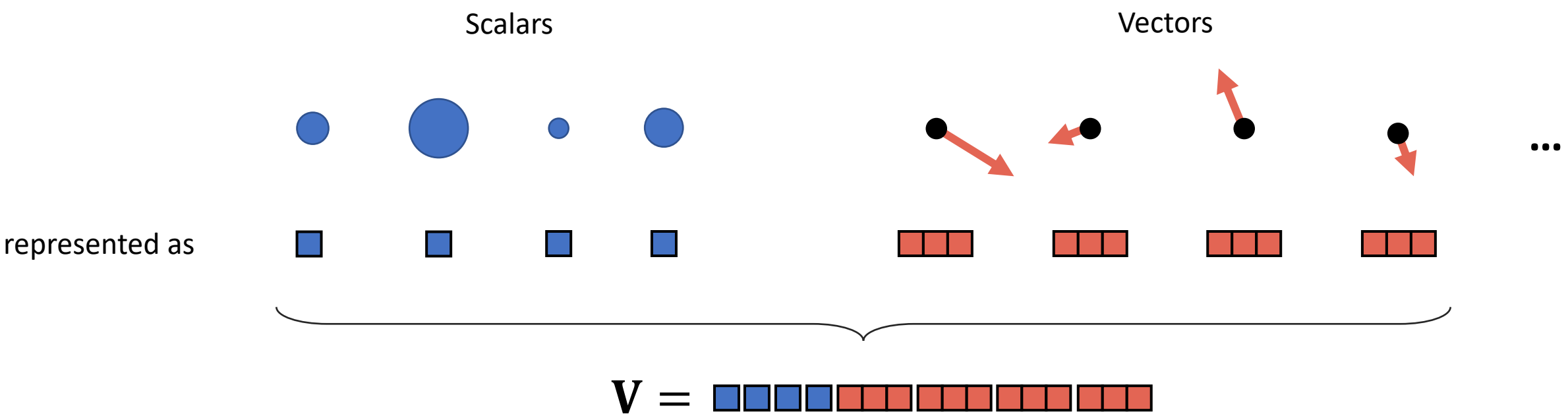




Equivariance: tensor features

The inputs, internal features, and outputs of the model are

collections of individual geometric tensors that transform variously under $O(3)$



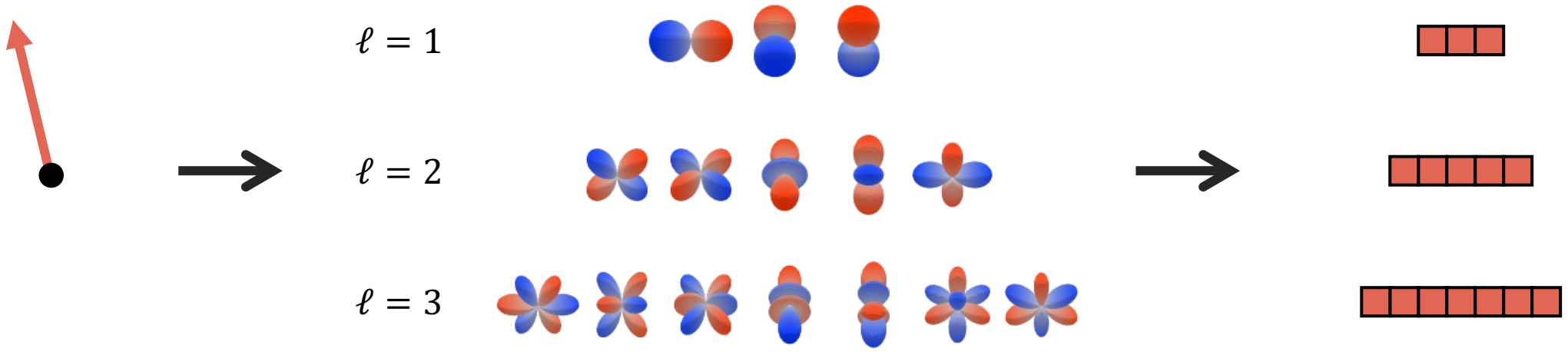
Equivariance: tensor features

- Formally, each tensor inhabits an irreducible representation (irrep) of $O(3)$
- The irreps are indexed by:
  - Rotation order $\ell \geq 0$:
    - $\ell = 0$: scalar
    - $\ell = 1$: vector
    - $\ell \geq 2$: tensor
  - Parity $p = \pm 1$:
    - $p = 1$: invariant under inversion
    - $p = -1$: changes sign under inversion
- A tensor of order ℓ has dimension $2\ell + 1$

[1] Thomas et al., “Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds”
[2] Geiger et al., e3nn documentation, e3nn.org

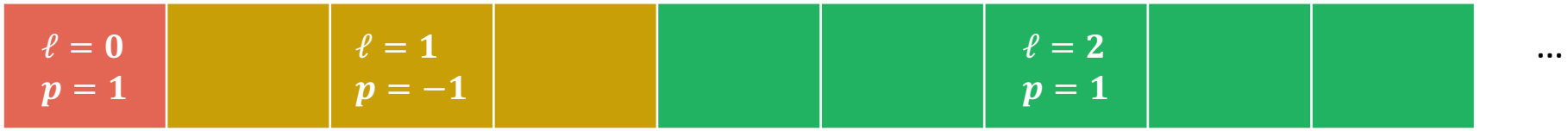
Equivariance: tensor features from spherical harmonics

- How can we encode data in these tensors?
- Spherical harmonics Y_ℓ^m are a basis for functions on the sphere
- They decompose functions into tensors of various ℓ and $p = (-1)^\ell$



Equivariance: tensor features

- An entire feature array

$\mathbf{V} =$ 

inhabits a direct sum of irreps

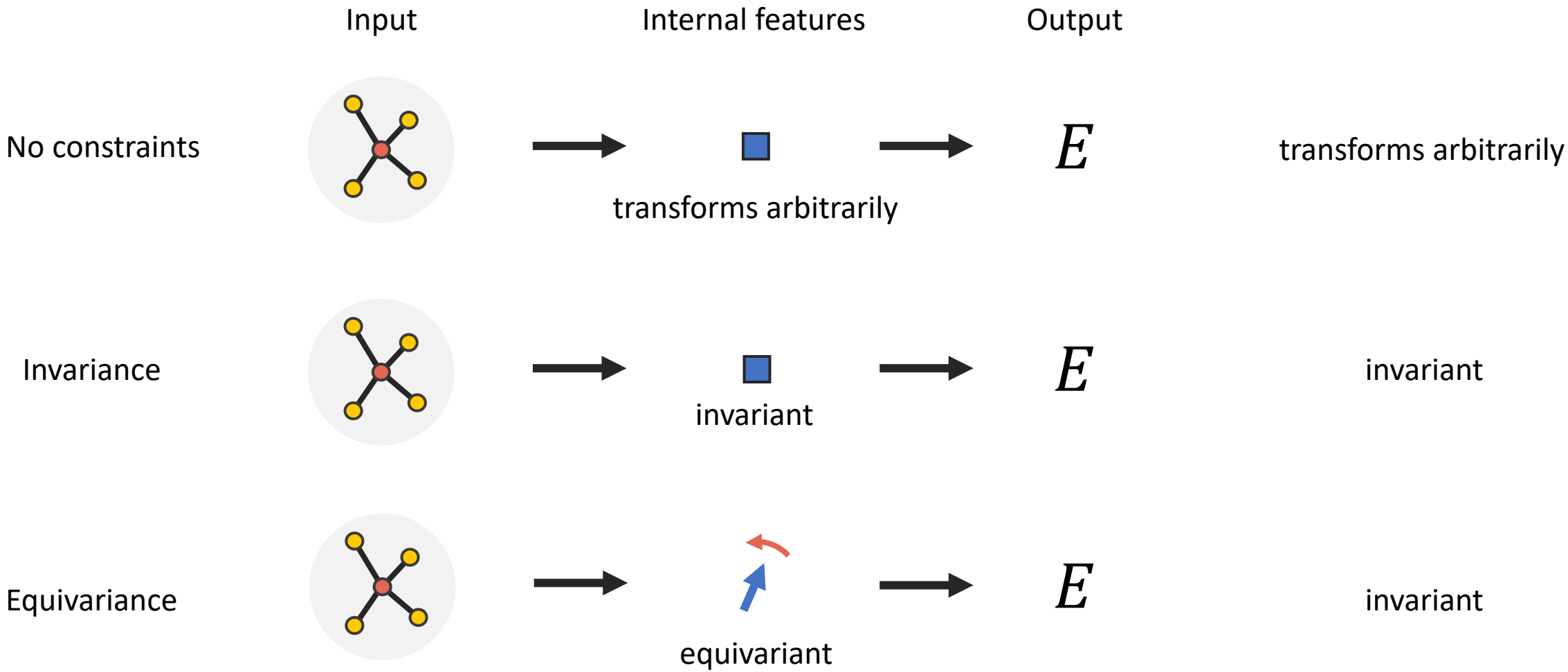
$$(\ell = 0, p = 1) \oplus (\ell = 1, p = -1) \oplus (\ell = 2, p = 1)$$

Very general:

- Any physical quantity transforms with a representation of $O(3)$
- Any representation of $O(3)$ decomposes into such a direct sum of irreps

[1] Thomas et al., “Tensor field networks: Rotation- and translation-equivariant neural networks for 3D point clouds”
[2] Geiger et al., e3nn documentation, e3nn.org

No constraints, invariance, equivariance



Tensor product

- A **bilinear**, **equivariant** operation combining two tensors

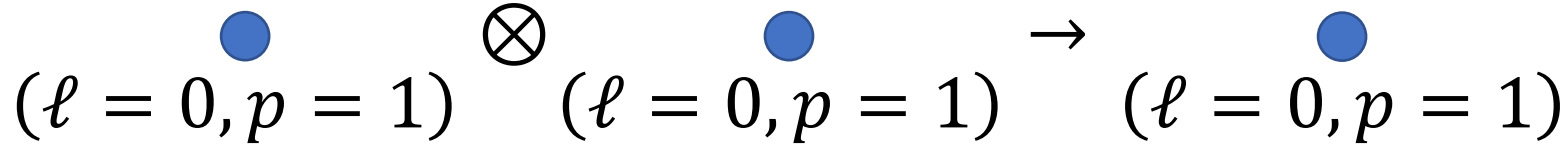
$$(\mathbf{x} \otimes \mathbf{y})_{\ell_{\text{out}}, m_{\text{out}}} = \sum_{m_1, m_2} \begin{pmatrix} \ell_1 & \ell_2 & \ell_{\text{out}} \\ m_1 & m_2 & m_{\text{out}} \end{pmatrix} \mathbf{x}_{\ell_1, m_1} \mathbf{y}_{\ell_2, m_2}$$

Wigner 3j coefficients

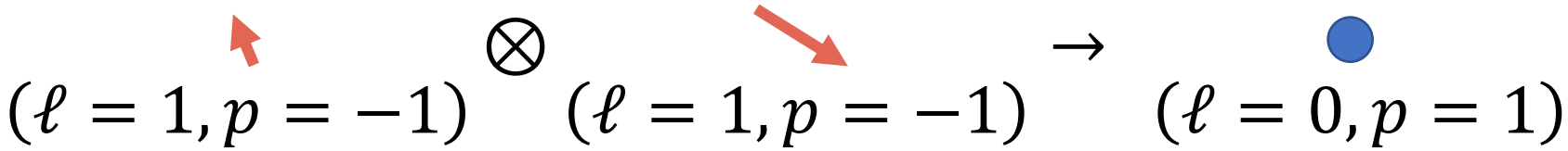
- The Wigner 3j are a change of basis from the product back into irreps
- Can produce any $|\ell_1 - \ell_2| \leq \ell_{out} \leq |\ell_1 + \ell_2|$ and $p_{out} = p_1 p_2$

Tensor product: examples

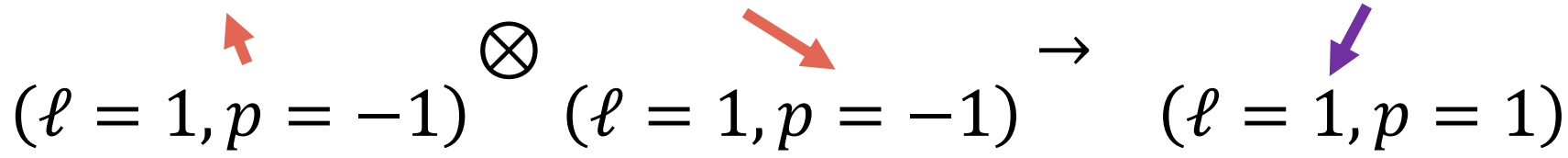
Scalar multiplication:



Vector-vector dot product:



Vector-vector cross product:



Equivariance dramatically improves MLIPs

- Examples: **NequIP** [0], PaiNN [1], UNiTE [2], EGNN [3], etc.
- All existing equivariant neural network MLIPs are message-passing

[0] Batzner et al. “SE(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials”

[1] K.T. Schutt, O.T. Unke, M. Gastegger. “Equivariant message passing for the prediction of tensorial properties and molecular spectra”

[2] Z. Qiao, A.S. Chirstensen, M. Welborn, F.R. Manby, A. Anandkumar, T.F. Miller III. “UNiTE: Unitary N-body Tensor Equivariant Network with Applications to Quantum Chemistry”

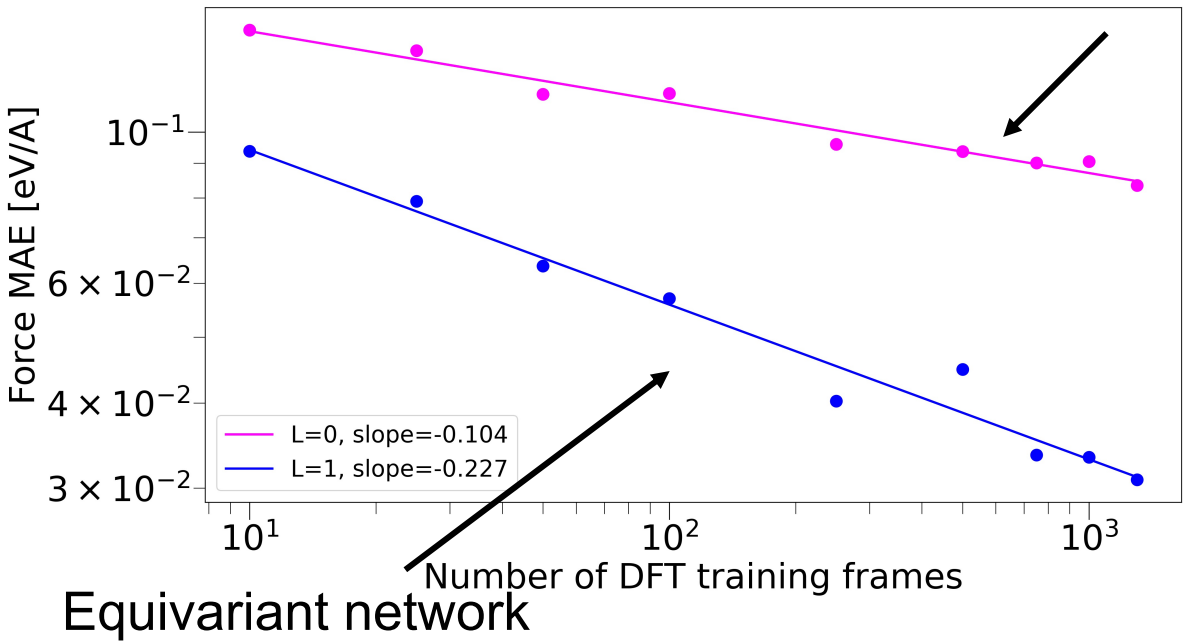
[3] V.G. Satorras, E. Hoogeboom, M. Welling. “E(n) Equivariant Graph Neural Networks”

NequIP demonstrated that equivariance leads fundamentally better molecular ML!

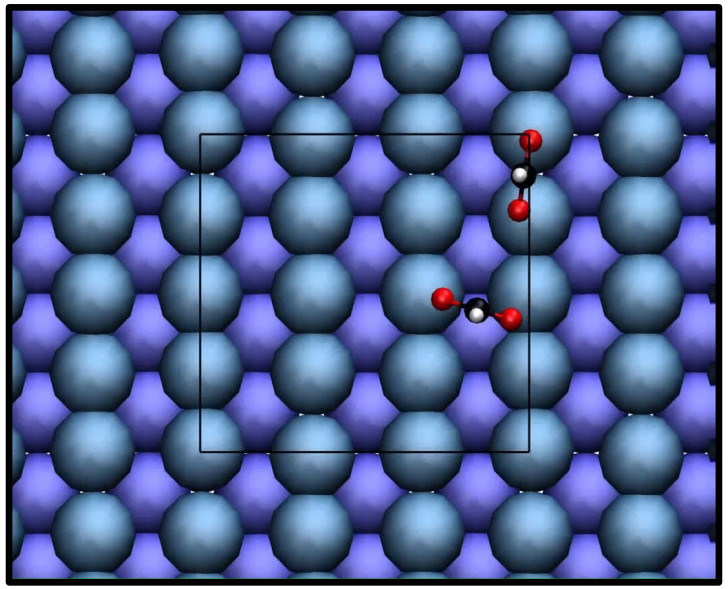
1000x fewer data

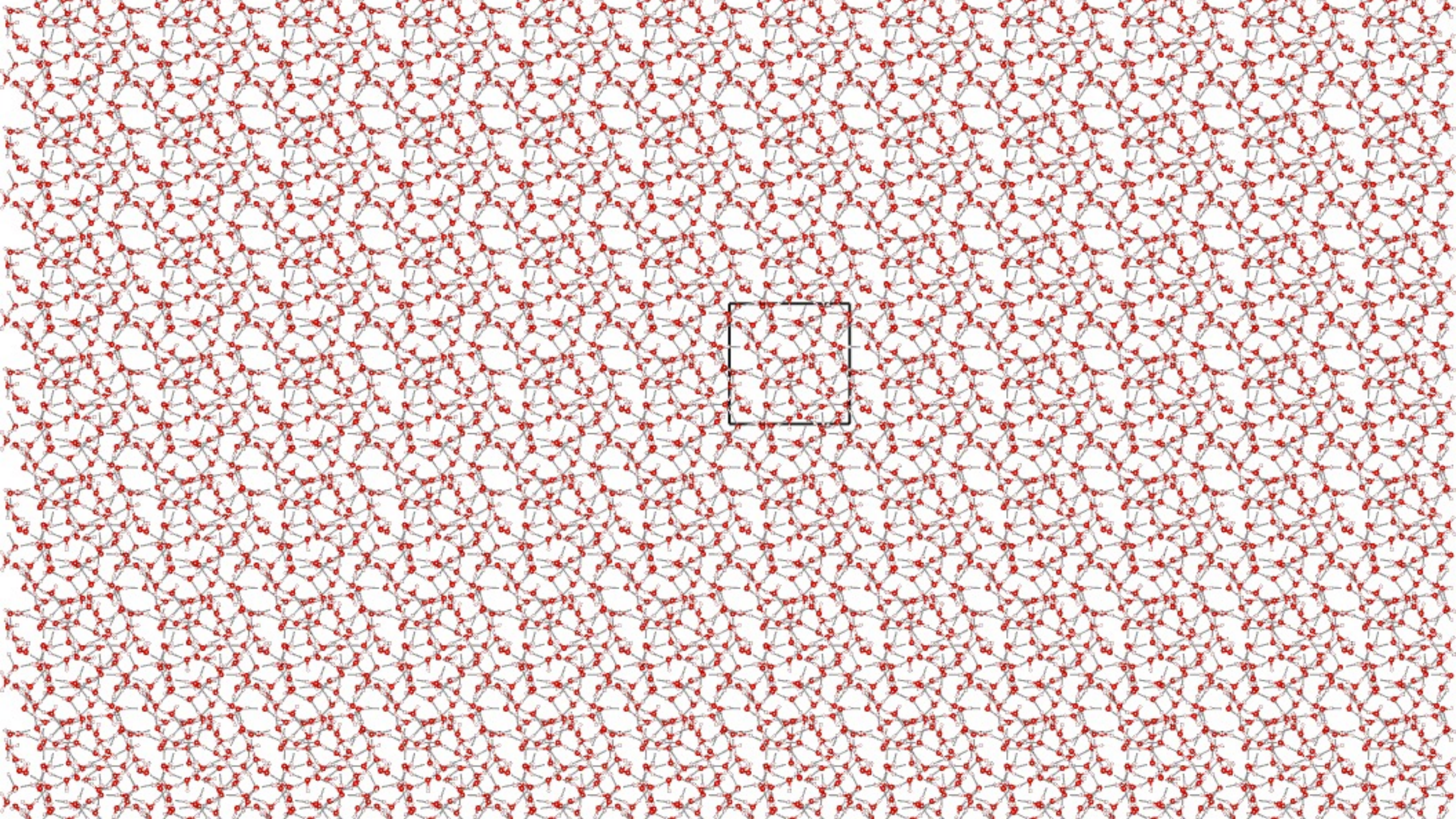
System	NequIP, trained on 133 structures	DeepMD, trained on 133,500 structures
Liquid Water	**11.9**	40.4
Ice Ih (b)	**10.2**	43.3
Ice Ih (c)	**12.0**	26.8
Ice Ih (d)	**9.8**	25.4

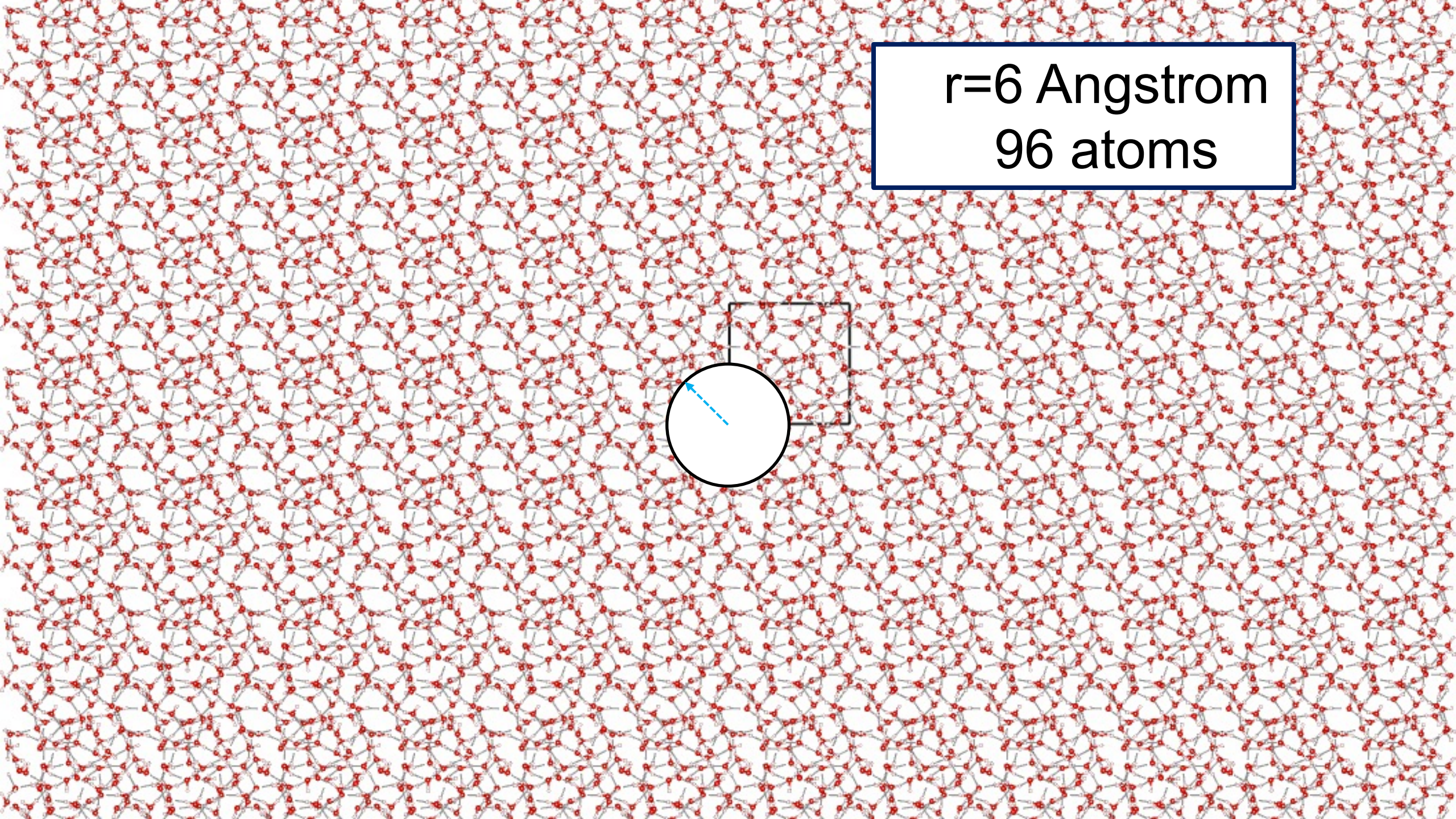
Different Scaling Laws

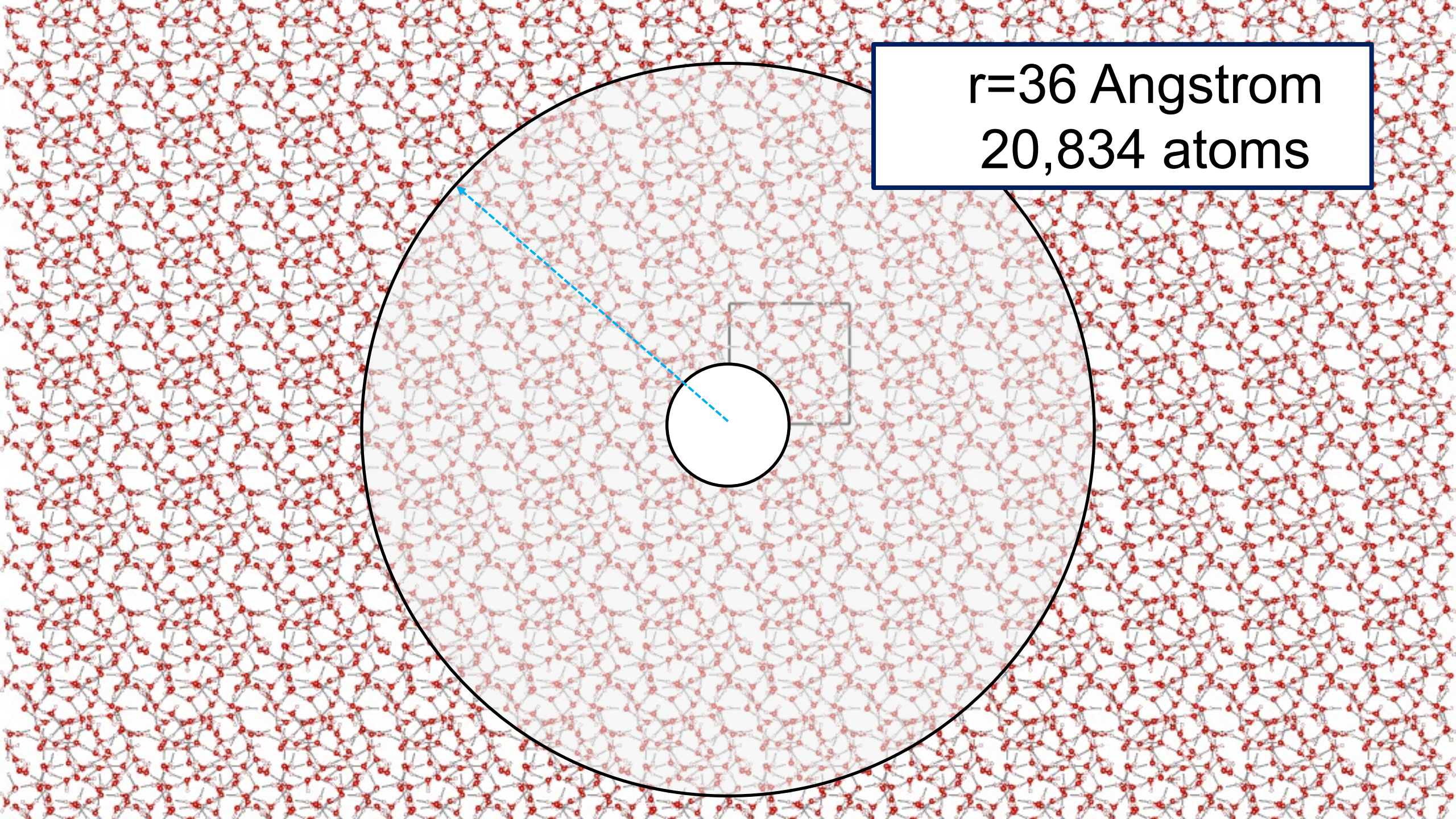


Complex, reactive systems






r=6 Angstrom
96 atoms

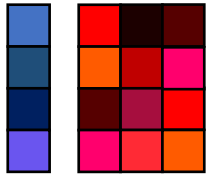

r=36 Angstrom
20,834 atoms

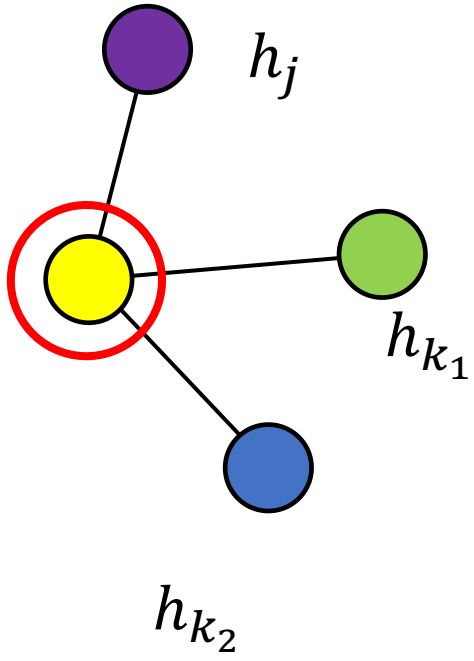
Equivariance 

Message Passing ?

Conventional GNN per-atom

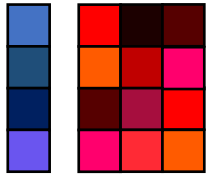
$$E = \sum_i E_i$$

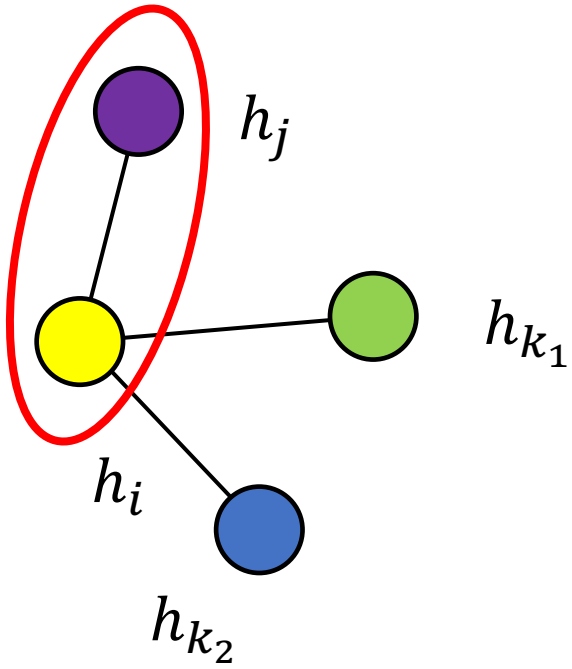
$h_i =$ 



Allegro per-pair

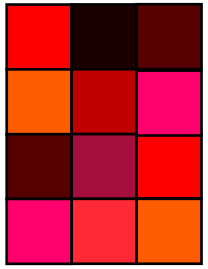
$$E = \sum_i \sum_{j \in N(i)} E_{ij}$$

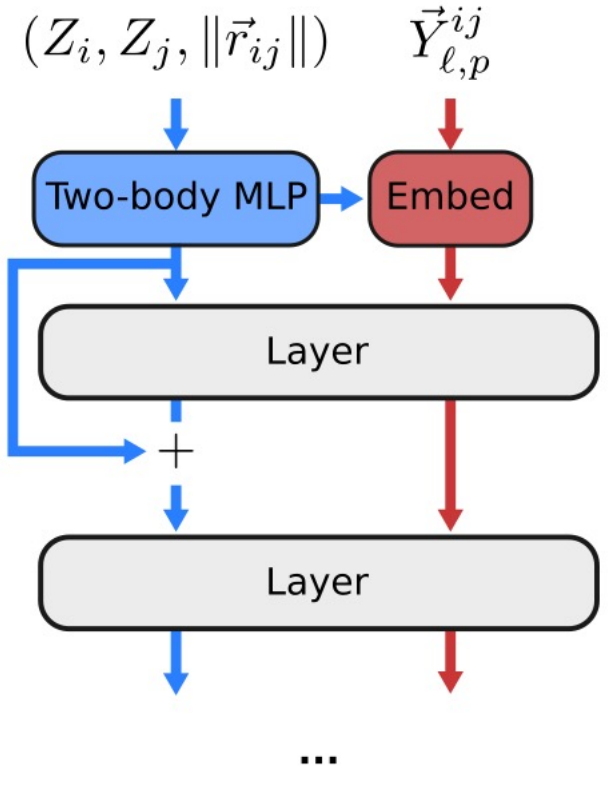
$h_{ij} =$ 



The Two-Track Architecture

Invariants **Equivariants**





Scalar / invariant track: $\mathbf{x}^{ij}$

All operations allowed

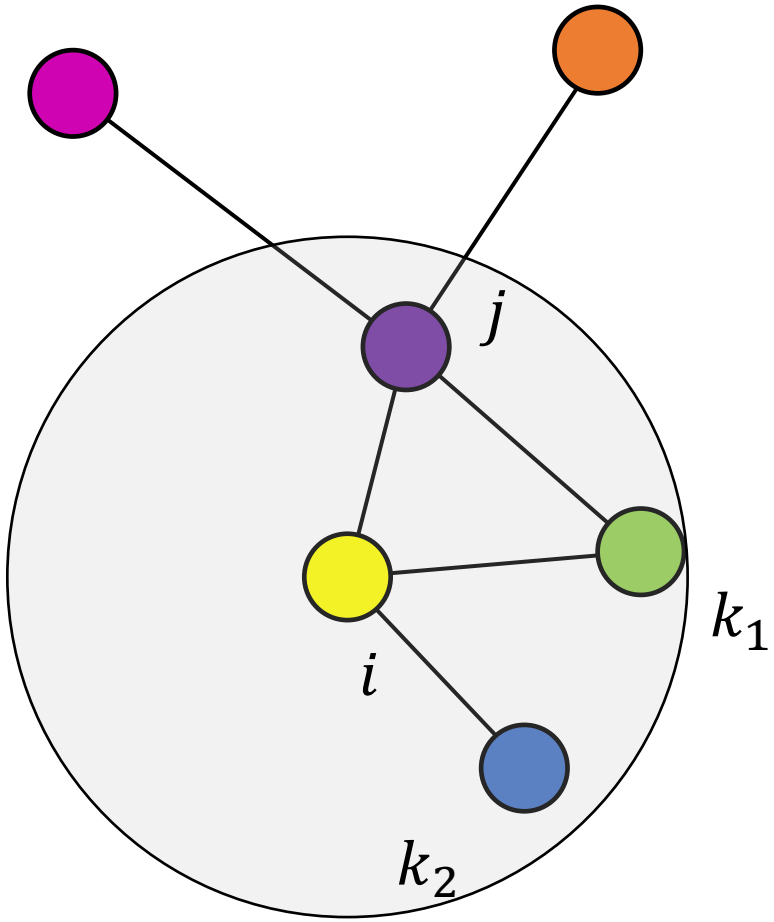
Tensor / equivariant track: $\mathbf{v}^{ij}$

Only E(3)-equivariant operations

Reasoning: scalars are cheap, tensors are expensive

Let large set of scalars control a small set of tensor operations!

Iterated tensor product increases correlation order

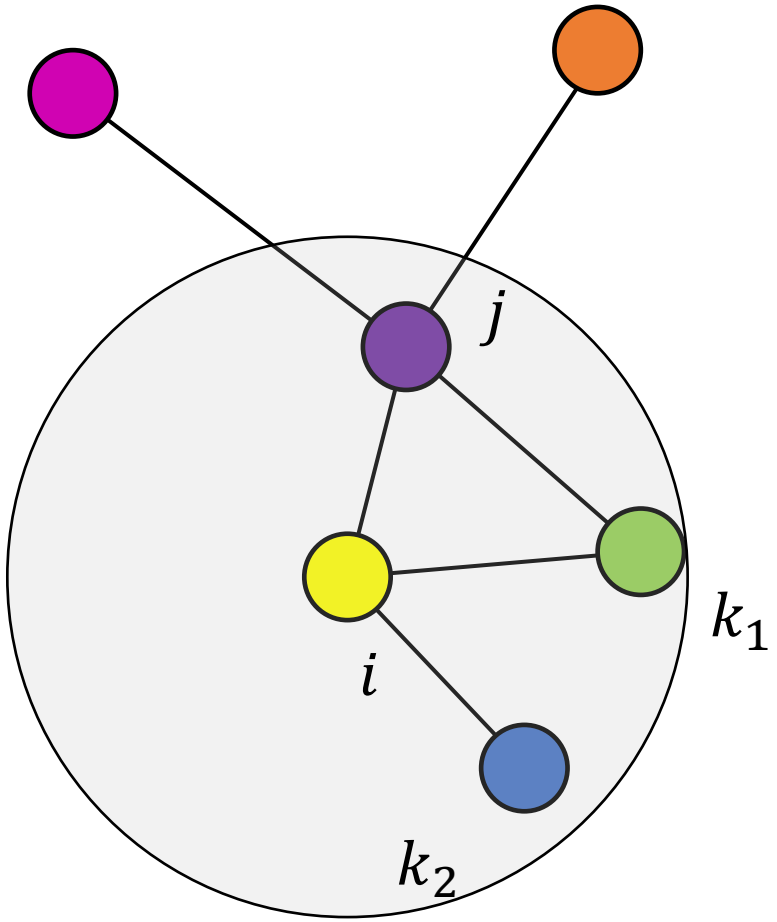


$$h_{i,j}^{L} = \sum_{k \in N(i)} w_{i,k}^{L} (h_{i,j}^{L-1} \otimes \vec{Y}_{i,k})$$

This induces a 3-body interaction $(i,j) \otimes (i,k) \rightarrow (i,j,k)$

Naively, this gives exponential scaling!

Density Trick[1] removes exponential scaling



Pair Feature

Pair Feature

$$h_{i,j}^{L} = \sum_{k \in N(i)} w_{i,k}^{L} \left(h_{i,j}^{L-1} \otimes \vec{Y}_{i,k} \right)$$

$$= \sum_{k \in N(i)} h_{i,j}^{L-1} \otimes w_{i,k}^{L} \vec{Y}_{i,k}$$

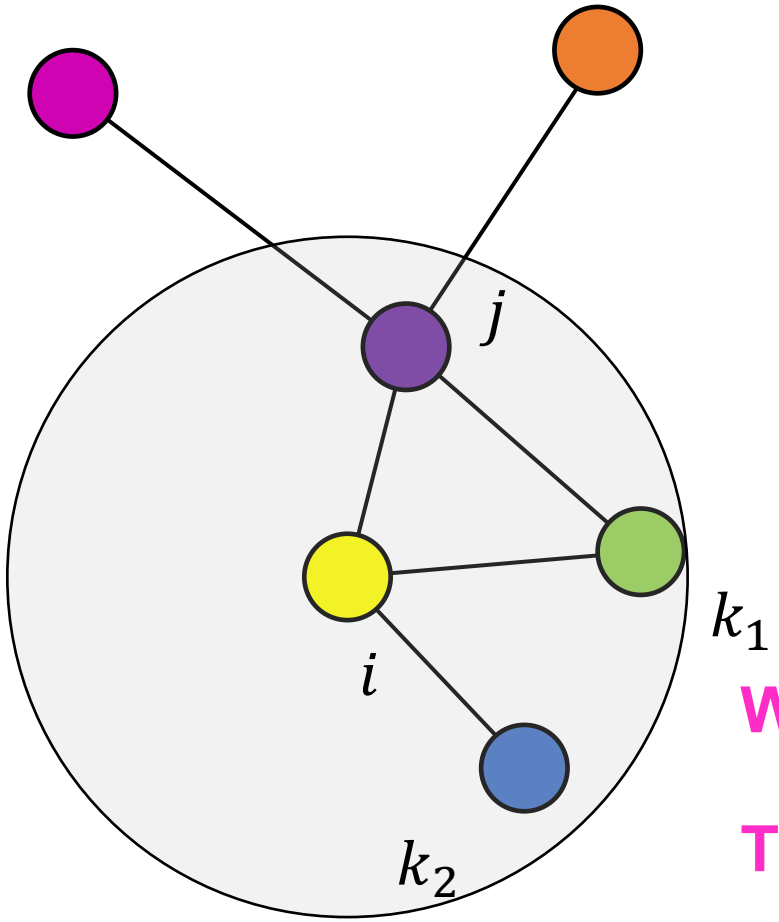
$$= h_{i,j}^{L-1} \otimes \sum_{k \in N(i)} w_{i,k}^{L} \vec{Y}_{i,k}$$

Pair Feature

Environment Feature

[1] Bartók, el al: On representing chemical environments. Phys. Rev. B: Condens. Matter Mater. Phys. 2013, 87, 1−16

Density Trick[1] removes exponential scaling



$$h_{i,j}^{L} = h_{i,j}^{L-1} \otimes \sum_{k \in N(i)} w_{i,k}^{L} \vec{Y}_{i,k}$$

Pair Feature

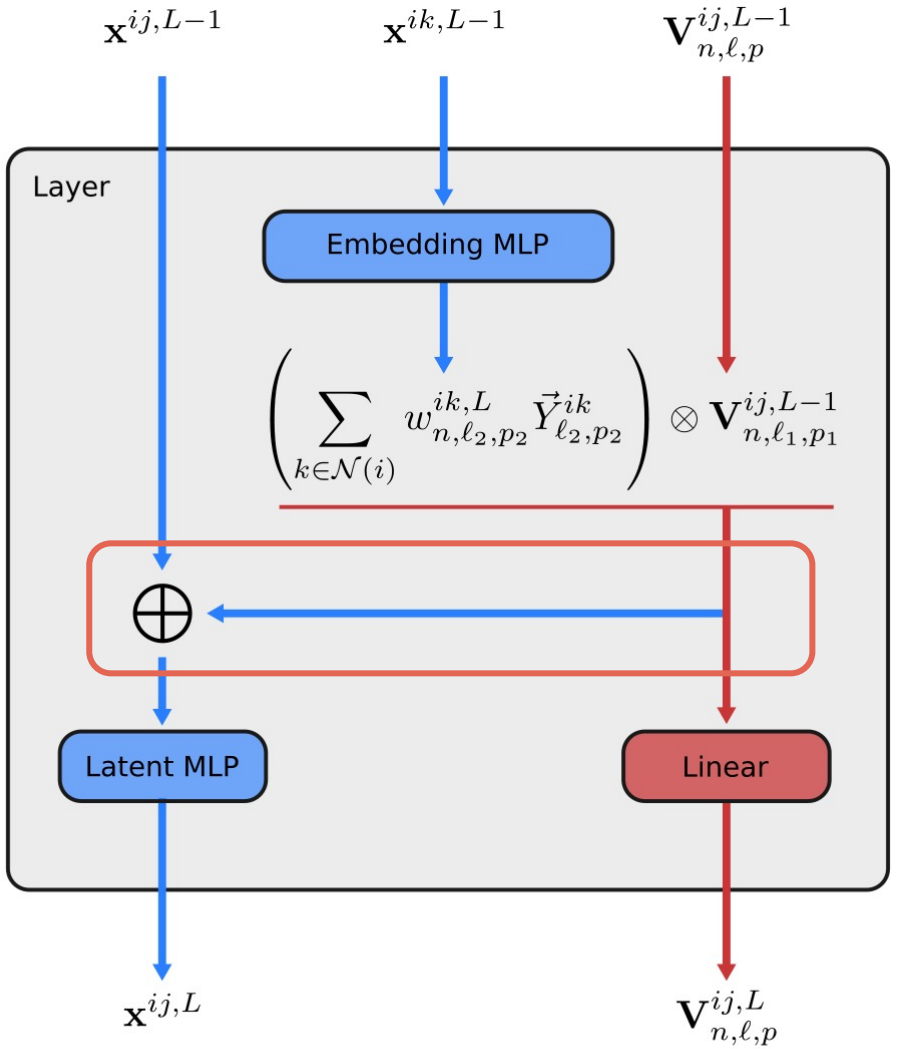
Environment Feature

We exploit the bilinearity of the tensor product

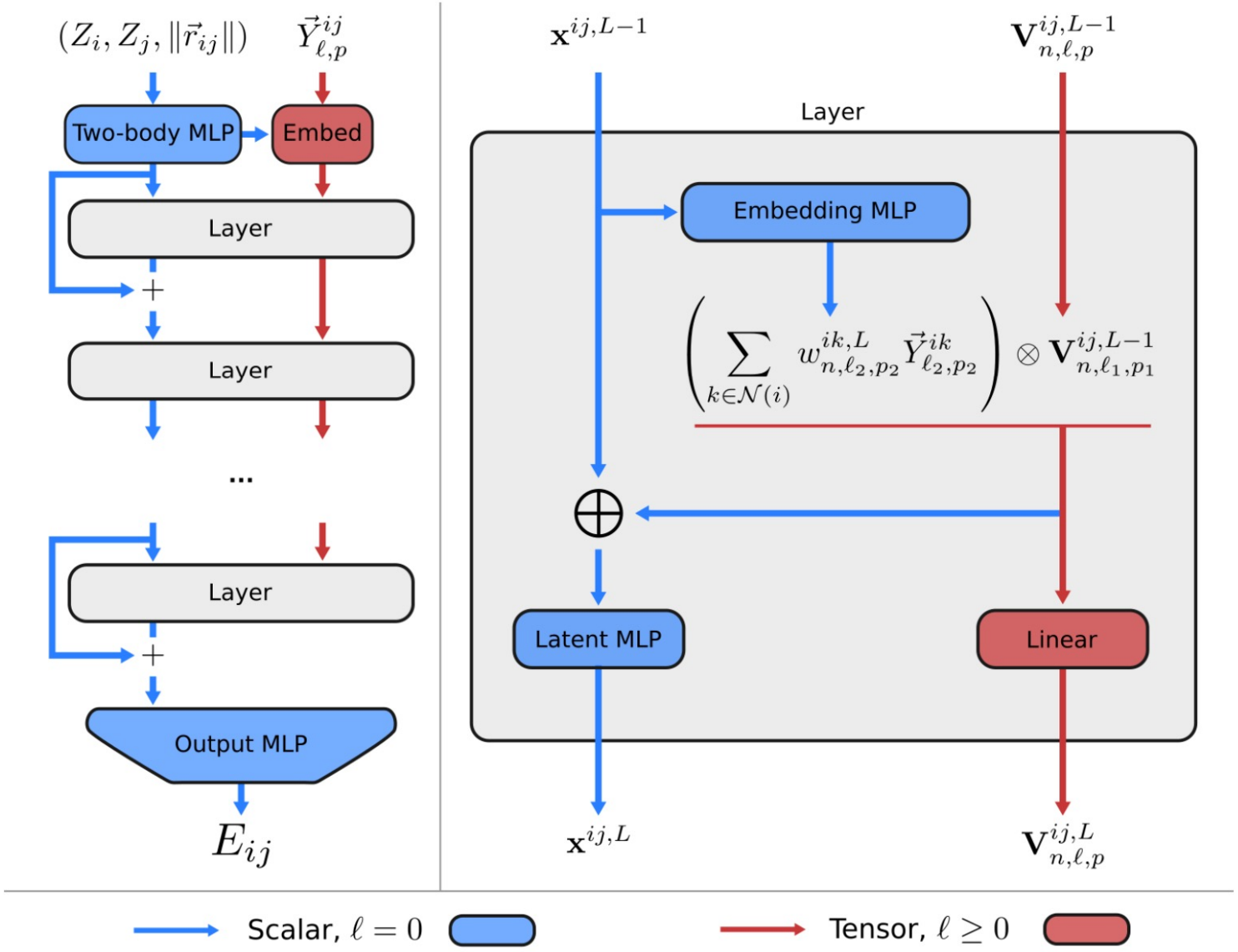
This gives linear scaling with the correlation order!

[1] Bartók, el al: On representing chemical environments. Phys. Rev. B: Condens. Matter Mater. Phys. 2013, 87, 1−16

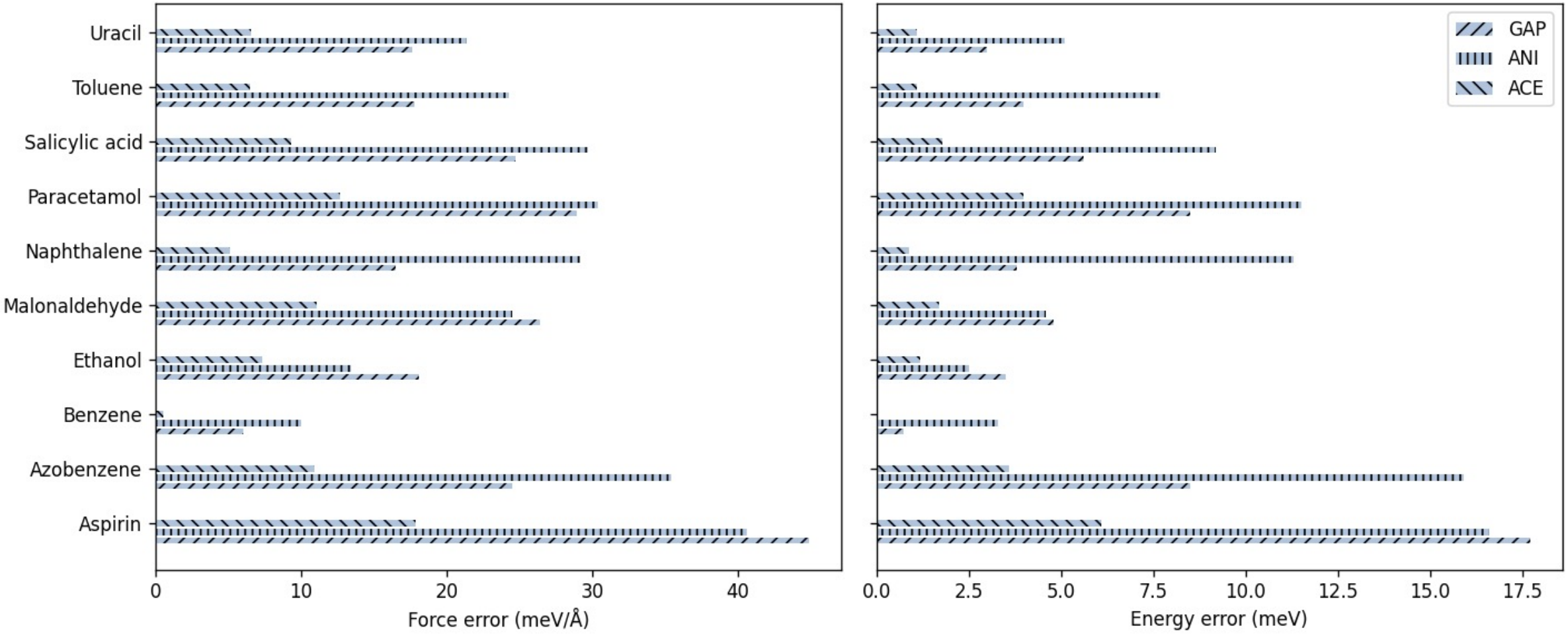
The full Tensor Product Layer



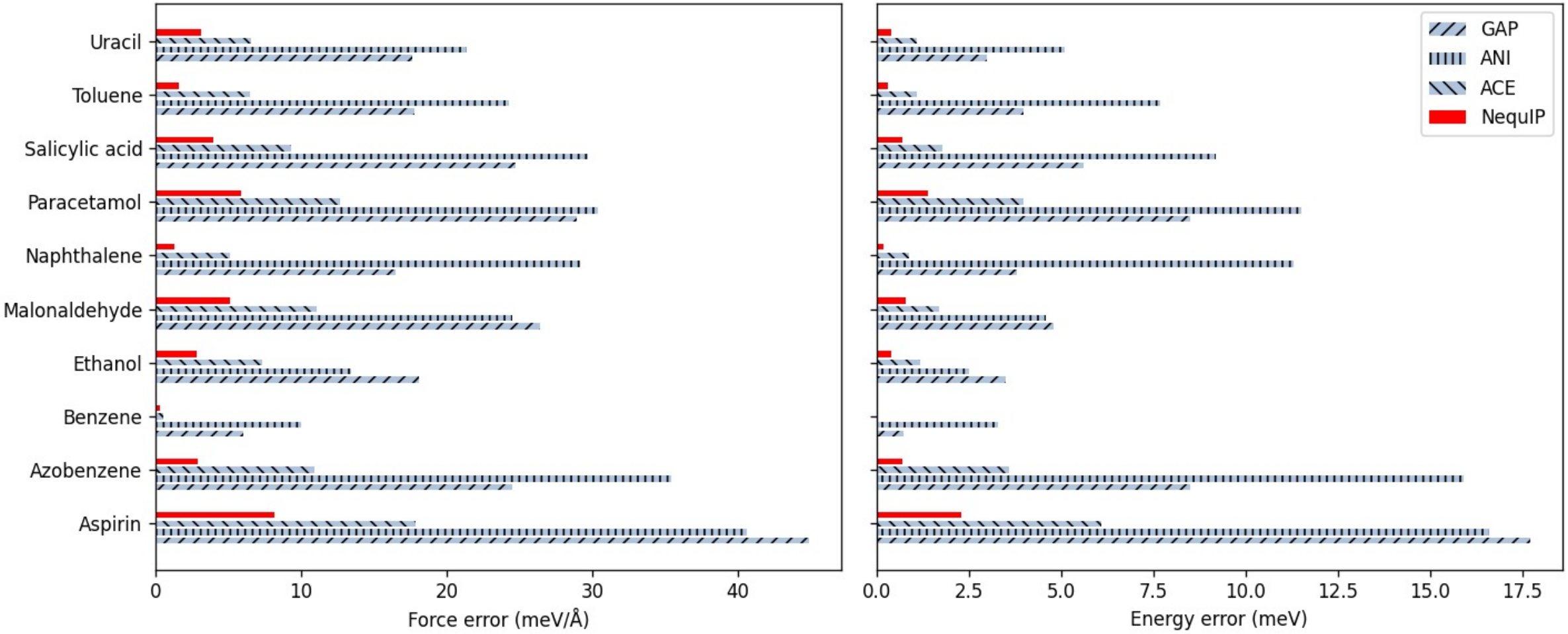
The full Allegro model



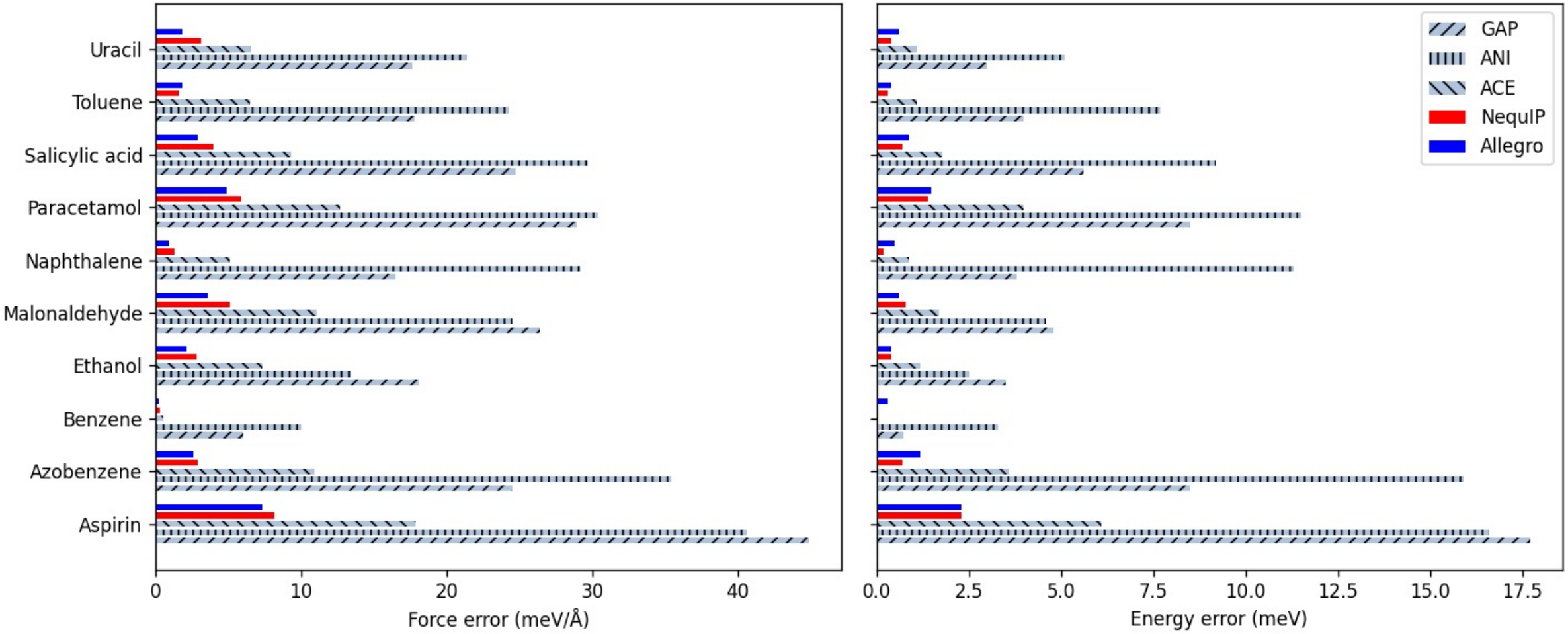
Allegro obtains state-of-the-art accuracy on revMD-17



Allegro obtains state-of-the-art accuracy on revMD-17



Allegro obtains state-of-the-art accuracy on revMD-17



Learning across compositional space, QM9

Model	U_0	U	H	G
Schnet [25]	14	19	14	14
DimeNet++ [54]	6.3	6.3	6.5	7.6
Cormorant [23]	22	21	21	20
LieConv [55]	19	19	24	22
L1Net [56]	13.5	13.8	14.4	14.0
SphereNet [57]	6.3	7.3	6.4	8.0
EGNN [32]	11	12	12	12
ET [40]	6.2	6.3	6.5	7.6
NoisyNodes [58]	7.3	7.6	7.4	8.3
PaiNN [27]	5.9	5.7	6.0	7.4
Allegro, 1 layer	5.7 (0.2)	5.3	5.3	6.6
Allegro, 3 layers	**4.7** (0.2)	**4.4**	**4.4**	**5.7**

1 layer →

Beyond accuracy: benchmarking the transferability of Allegro

Temperature Transferability

Test	———	T = 1200K
Test	———	T = 600K
Train	———	T = 300K

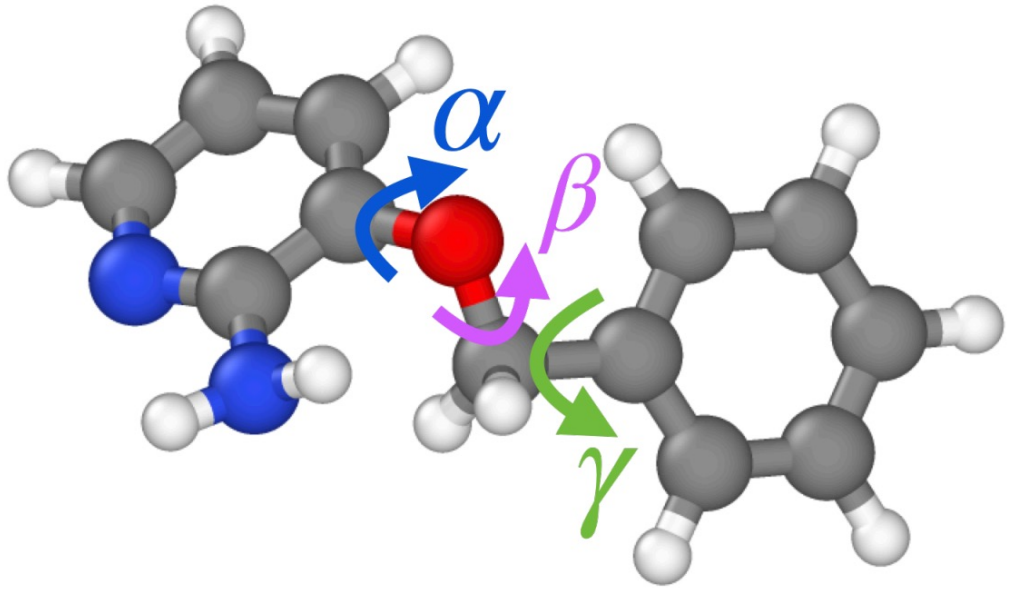
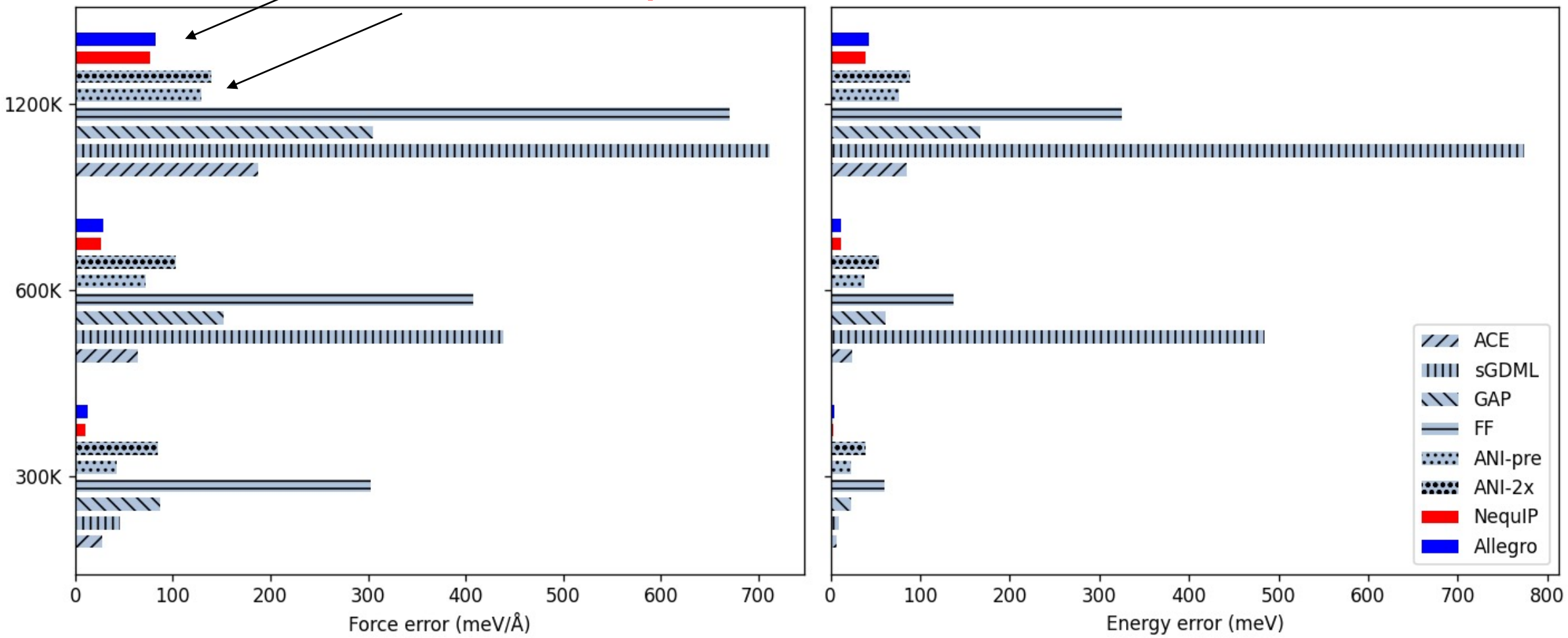


Figure from: Kovacs et al., 2021, *JCTC*

Allegro shows strong OOD-generalization

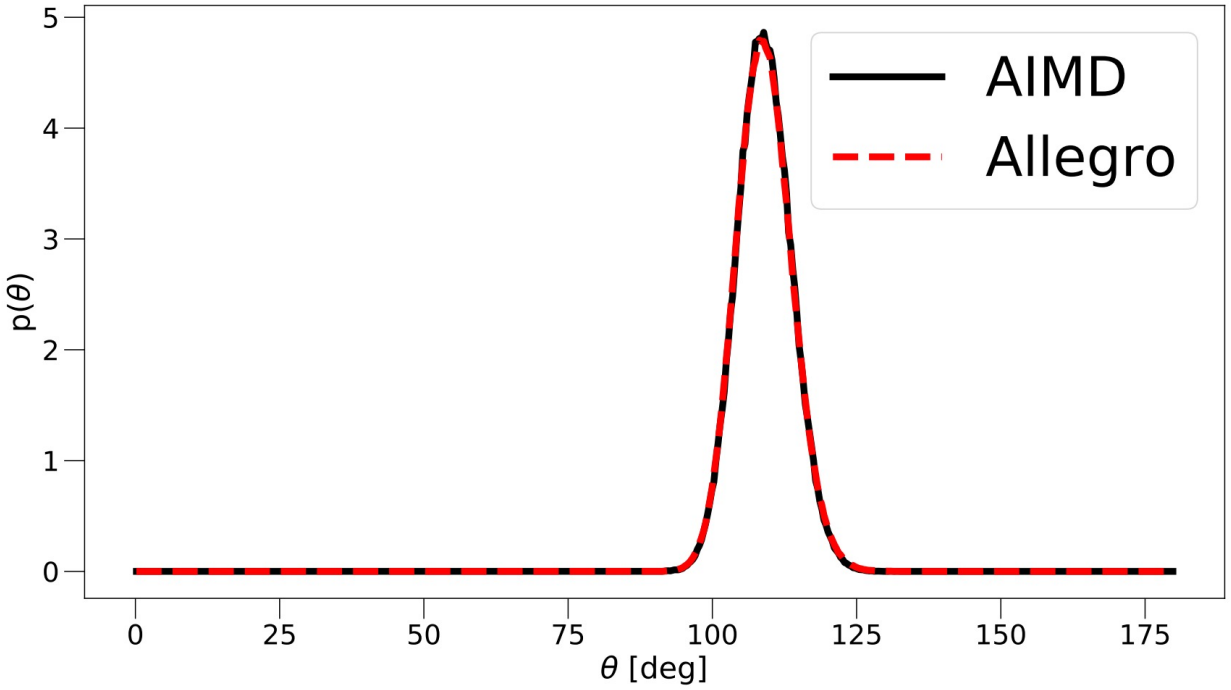
Allegro/NequIP models trained on 500 structures

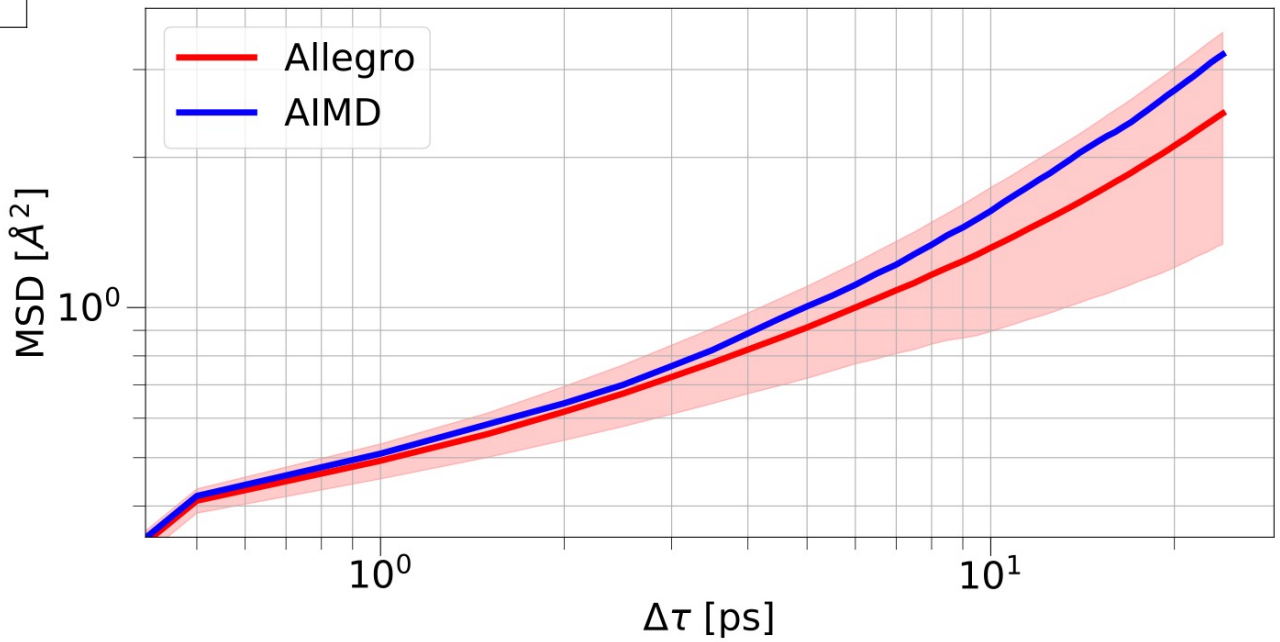
ANI models, pretrained on 8.9 million structures



Numbers for models besides NequIP and Allegro are from [1] Kovacs et al., 2021, *JCTC*

Allegro predicts the structures + kinetics of complex materials

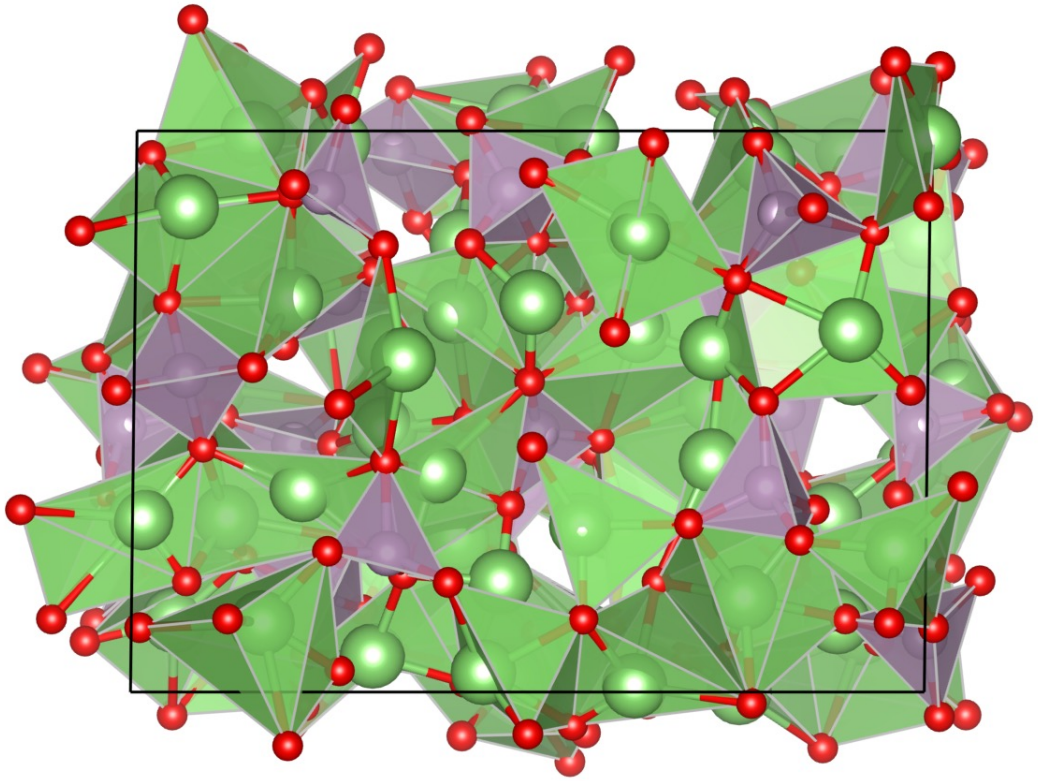




Li3PO4, quenched at T=600K
Top: tetrahedral ADF
Bottom: Li MSD

Allegro: speed

32.4 ns/day on a DFT sized system
(192 atoms)



Allegro's accuracy scales…

- $\boldsymbol{O(N)}$ in the **number of atoms**
  contrast: $O(N^2)$ global descriptors such as sGDML[1]

- $\boldsymbol{O(M)}$ in the **number of neighbors/atom**
  contrast: some $O(M^2)$ deep learning approaches such as DimeNet[2] or Equivariant Transformers[3]

- $\boldsymbol{O(1)}$ in the **number of chemical species**
  contrast: local descriptors like SOAP — $O(S^2)$ — and ACE[4]: $O(S^{\text{body order}-1})$

[1] Chmiela, S., Sauceda, H. E., Muller, K.-R. & Tkatchenko, A. Towards exact molecular dynamics simulations with machine-learned force fields. Nature Communications 9, 3887 (2018).

[2] Klicpera, J., Groß, J. & Gunnemann, S. Directional message passing for molecular graphs. arXiv preprint arXiv:2003.03123 (2020).

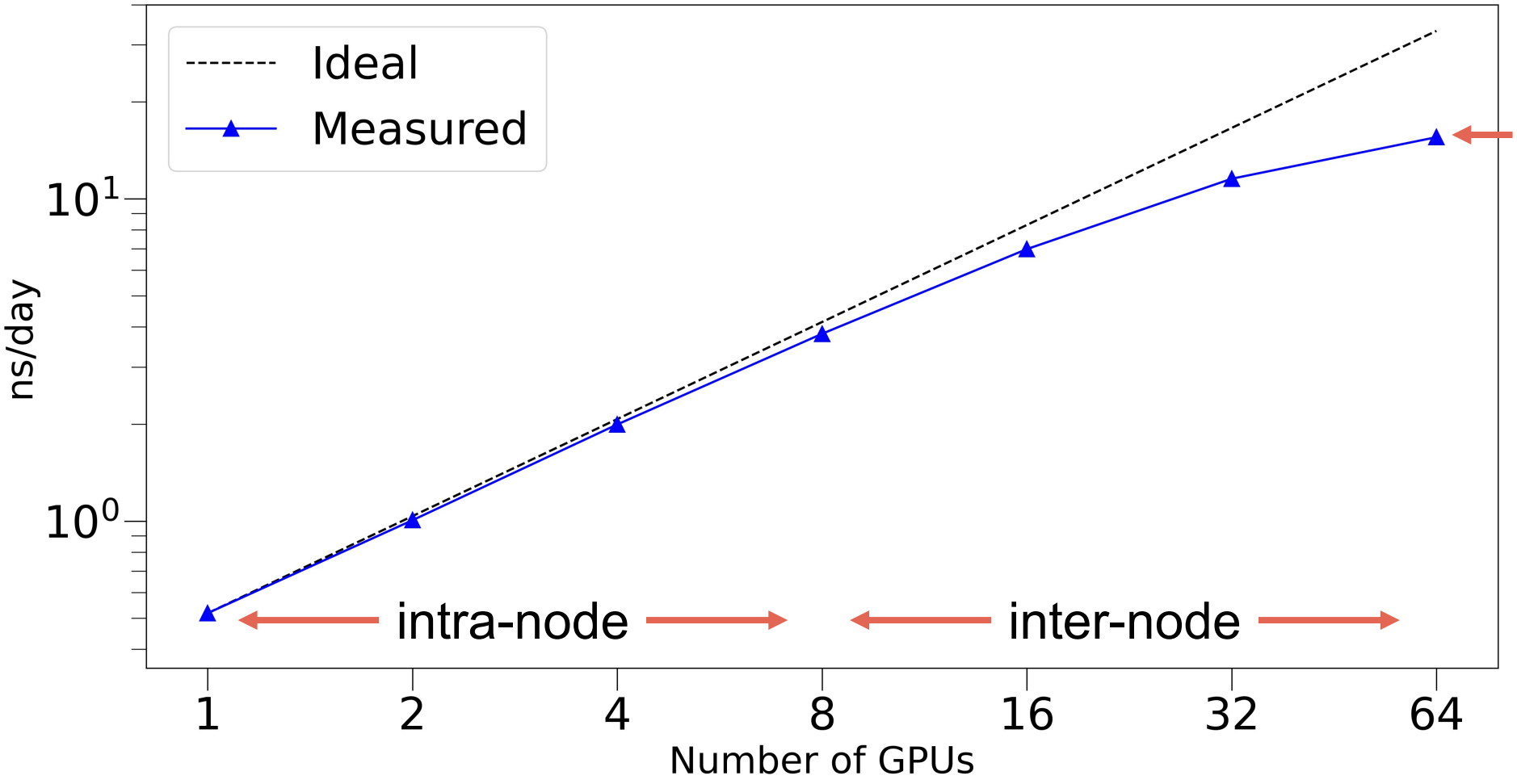
[3] Tholke, P. & De Fabritiis, G. Torchmd-net: Equivariant transformers for neural network based molecular potentials. arXiv preprint arXiv:2202.02541 (2022).

[4] Drautz, R. Atomic cluster expansion for accurate and transferable interatomic potentials. Physical Review B 99, 014104 (2019).

Allegro can **practically** scale…

…for a fixed system size

Allegro: strong scaling on 421,824 atoms



~16 ns/day
6,591 atoms/GPU
0.018 μs/atom/step

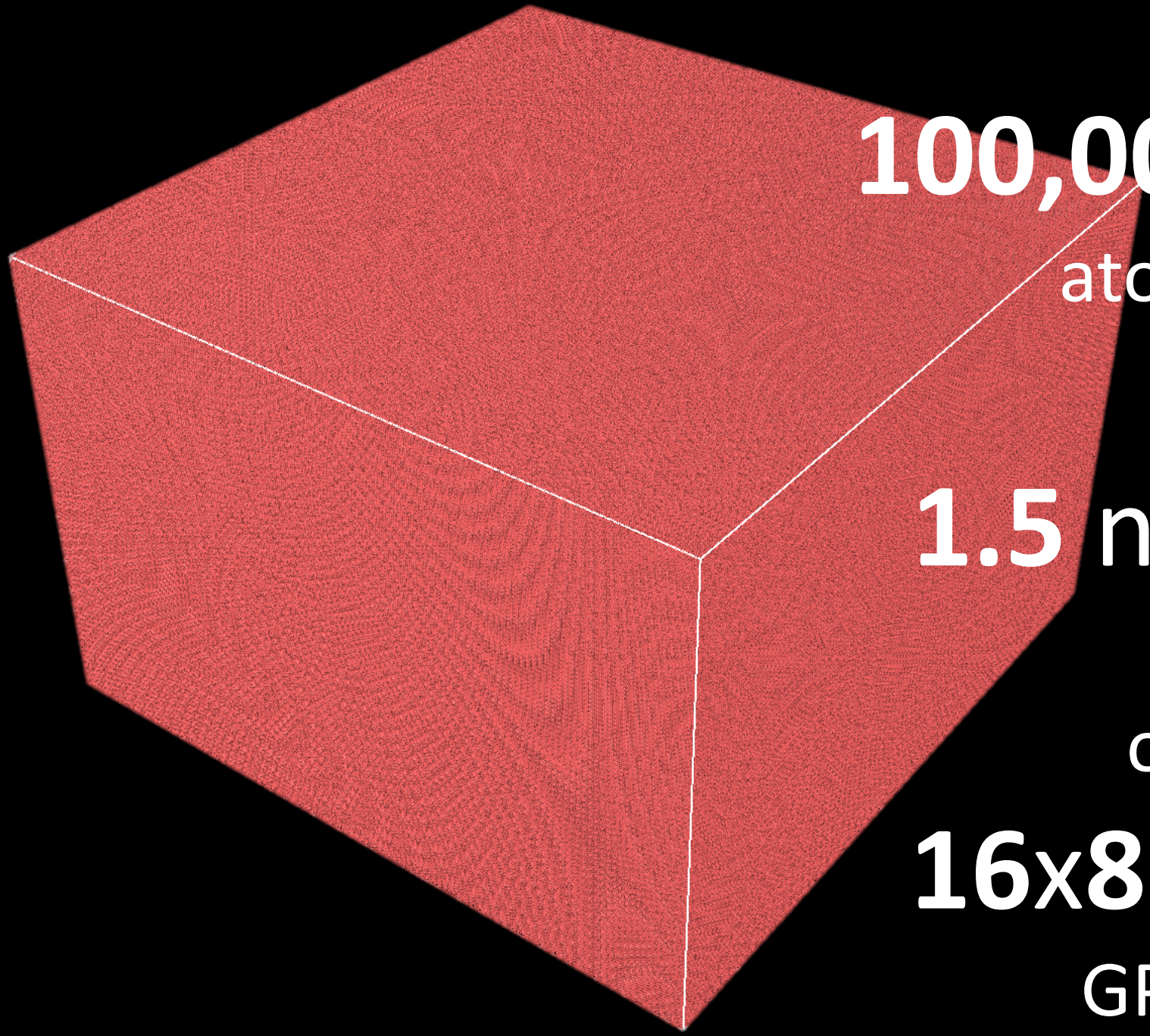
HARVARD UNIVERSITY

Anders Johansson

Simulations run in LAMMPS on NVIDIA A100 GPUs; 8 GPUs / node. Timestep: 2fs.

Allegro can **practically** scale…

…to extremely large systems

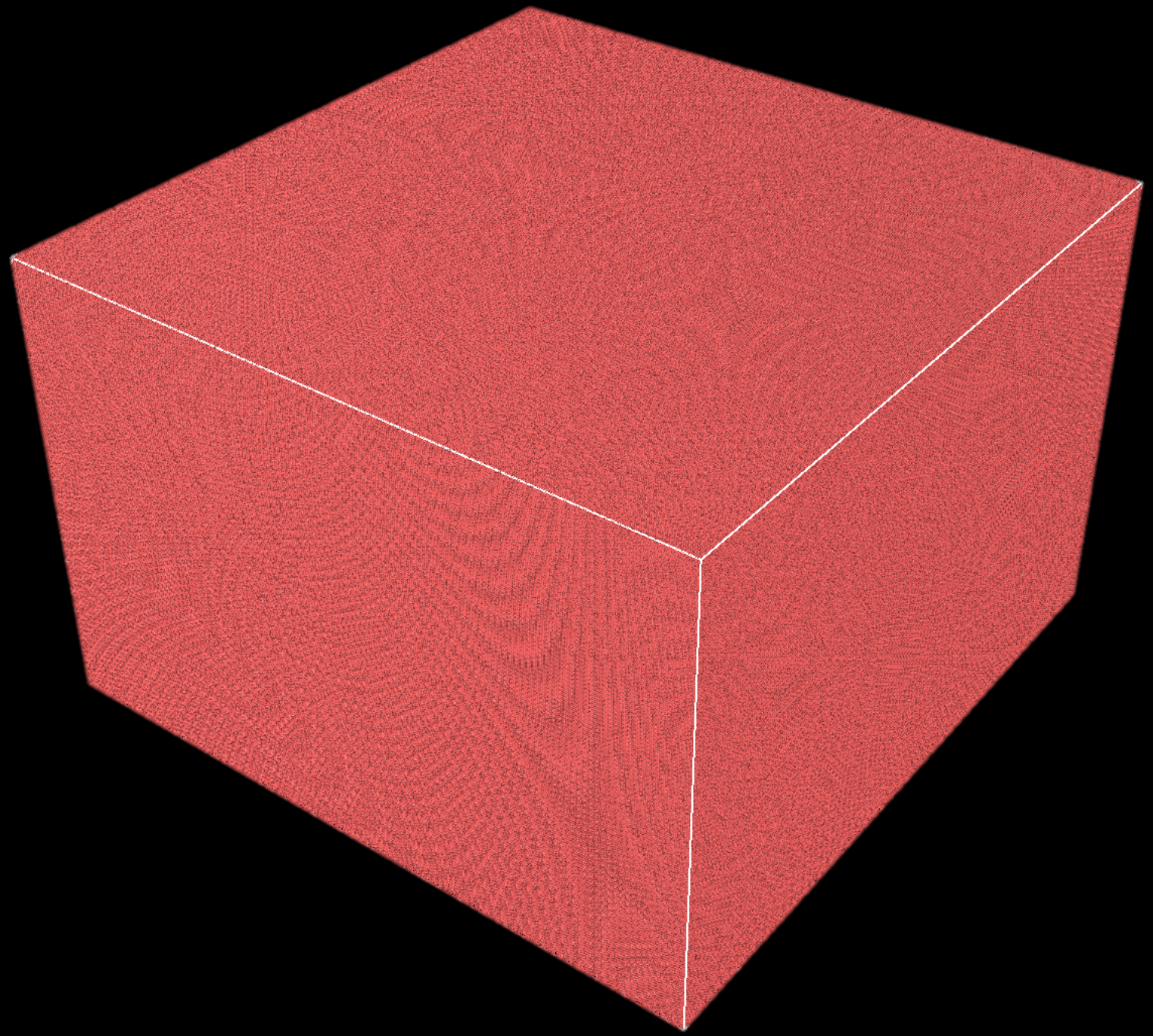


100,000,000
atoms

1.5 ns/day

on

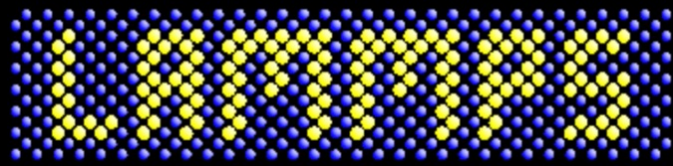
16x8 A100
GPUs



100,000,000
atoms

1.5 ns/day

integration with



Neural Equivariant Interatomic Potentials

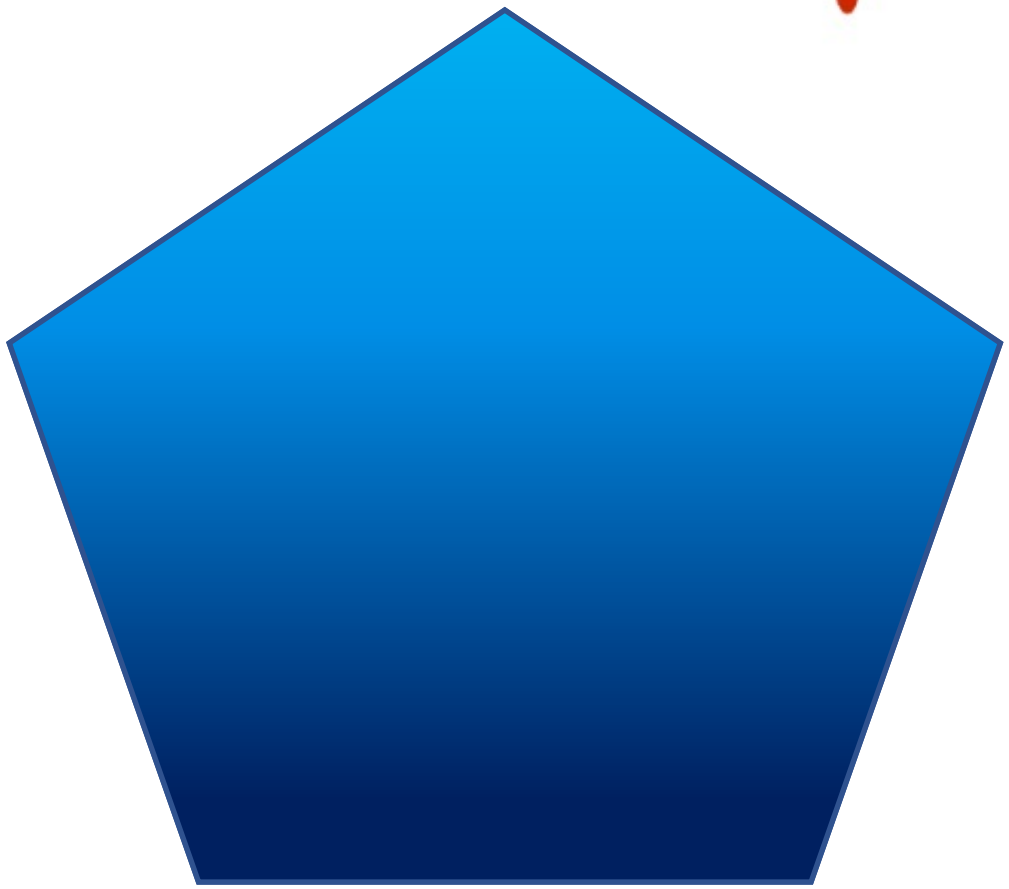
`github.com/mir-group/nequip`
`github.com/mir-group/allegro`

- Modular open-source framework for designing, training, testing, and deploying equivariant MLIPs
- Allegro is implemented as an extension package
- Optimized for GPUs with PyTorch
- Full TorchScript support for Python-free deployment, including to our LAMMPS plugin `pair_allegro`

NequIP

Allegro

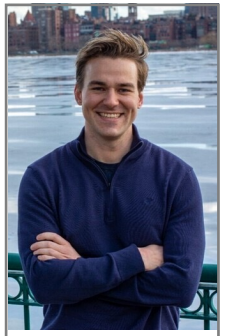
Is our community there yet?



Scale: Yes

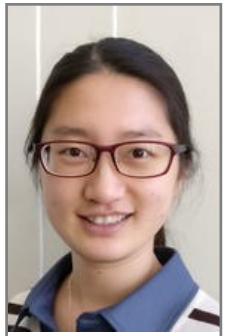
Speed: Not quite… (Allegro: 32.4 ns/day)

Albert Musaelian



Simon Batzner

Anders Johansson

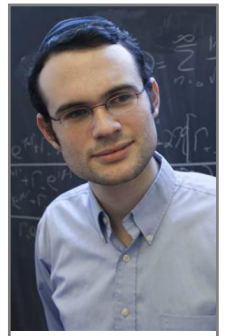


Lixin Sun

Cameron Owen

Nicola Molinari

Boris Kozinsky



Mordechai Kornbluth

BOSCH

Compute

FAS RC CANNON

Argonne NATIONAL LABORATORY

TACC

Funding

DOE CSGF

U.S. DEPARTMENT OF ENERGY | Office of Science

IMASC

BOSCH

Materials Intelligence Research

Thank you!

Harvard John A. Paulson School of Engineering and Applied Sciences