ME 5374-ST

Machine Learning for Materials Science and Discovery

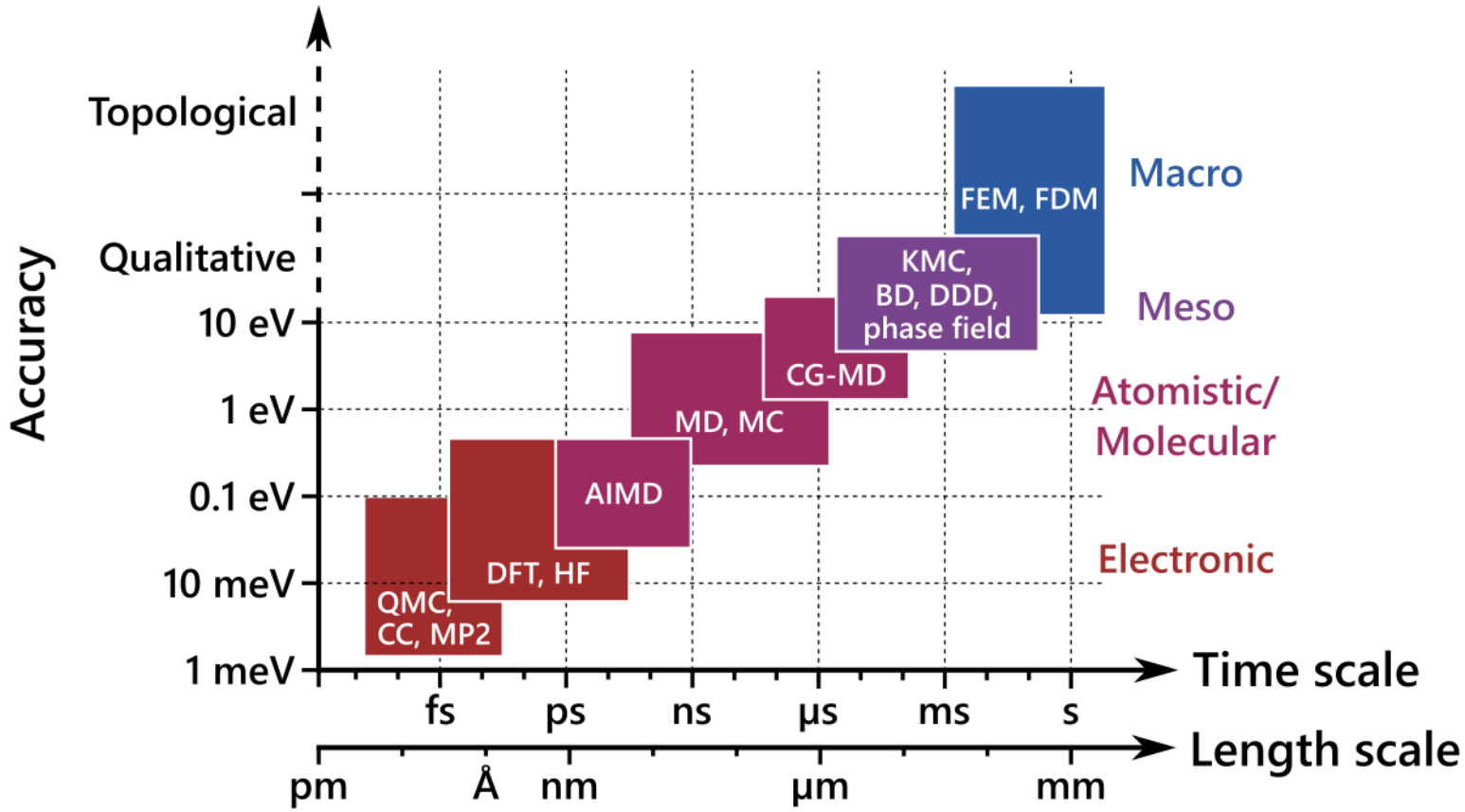
Fall 2025

Asst. Prof. Peter Schindler

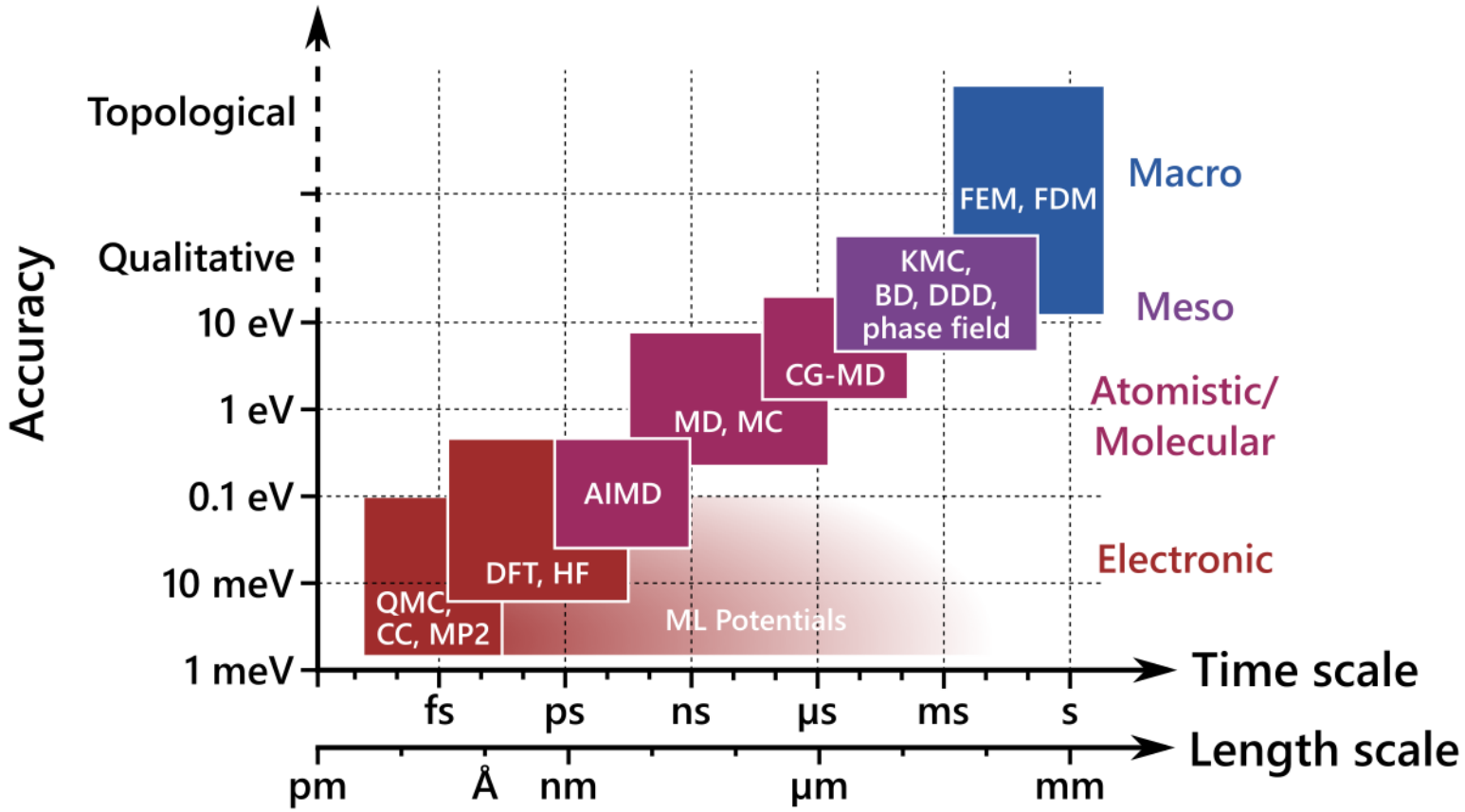
Lecture 17 – Machine Learning Interatomic Potentials (MLIPs)

- Enabling Large-scale MD Simulations with Near-Ab Initio Accuracy
- History and Development of MLIPs
- “Universal” MLIPs and their Training Datasets
- Properties Derivable from the Potential Energy Surface (PES)
- Benchmarking, Fine-tuning, and Future Directions

Computational Materials Science: Accuracy / Scale Tradeoff



MLIPs → "Affordable Accuracy"



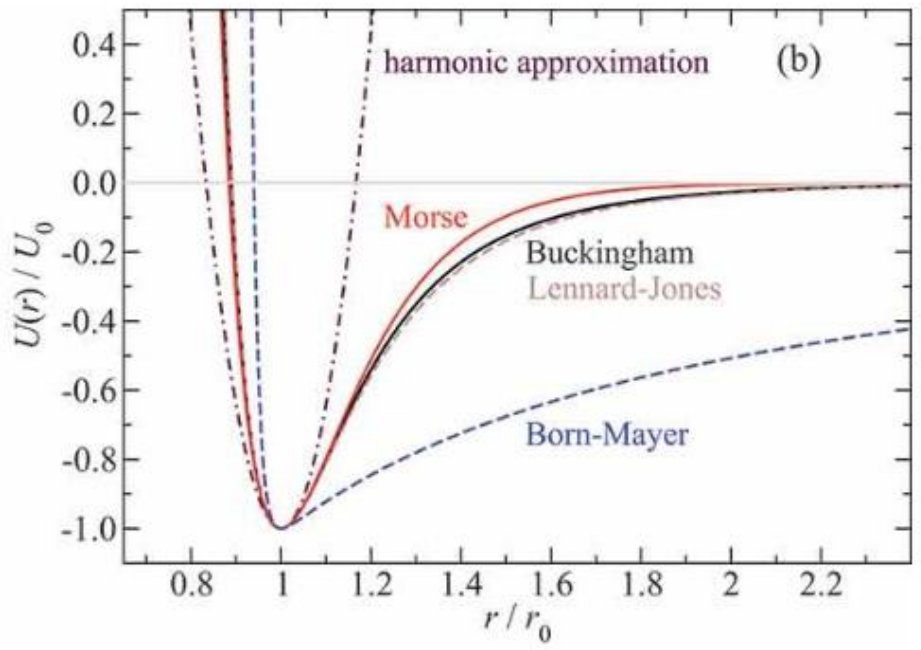
MD Recap: Interatomic Potentials

Attractive (e.g., Coulomb, dipole fluctuation)
+
Repulsive (Pauli exclusions) forces

$$V(r') = \int_{\infty}^{r'} F \cdot \mathrm{d}r \qquad F = \frac{\mathrm{d}V(r)}{\mathrm{d}r}$$

$$V(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

Lennard-Jones Potential



Two-body potentials. (Müser et al., 2022)

Recap: Ab initio Molecular Dynamics (AIMD)

Hellmann-Feynman theorem

$$\frac{\mathrm{d}E_\lambda}{\mathrm{d}\lambda} = \langle \psi_\lambda | \frac{\mathrm{d}H_\lambda}{\mathrm{d}\lambda} | \psi_\lambda \rangle$$

Gives us forces directly from the electron density:

$$F_{X,i} = -\frac{\partial E}{\partial X_i} = -\langle \psi | \frac{\mathrm{d}H}{\mathrm{d}X_i} | \psi \rangle = Z_i \left(\int \mathrm{d}\vec{r}\, \rho(\vec{r}) \frac{x - X_i}{\left| \vec{r} - \vec{R}_i \right|^3} - \sum_{\alpha \neq i}^{M} Z_\alpha \frac{X_\alpha - X_i}{\left| \vec{R}_\alpha - \vec{R}_i \right|^3} \right)$$

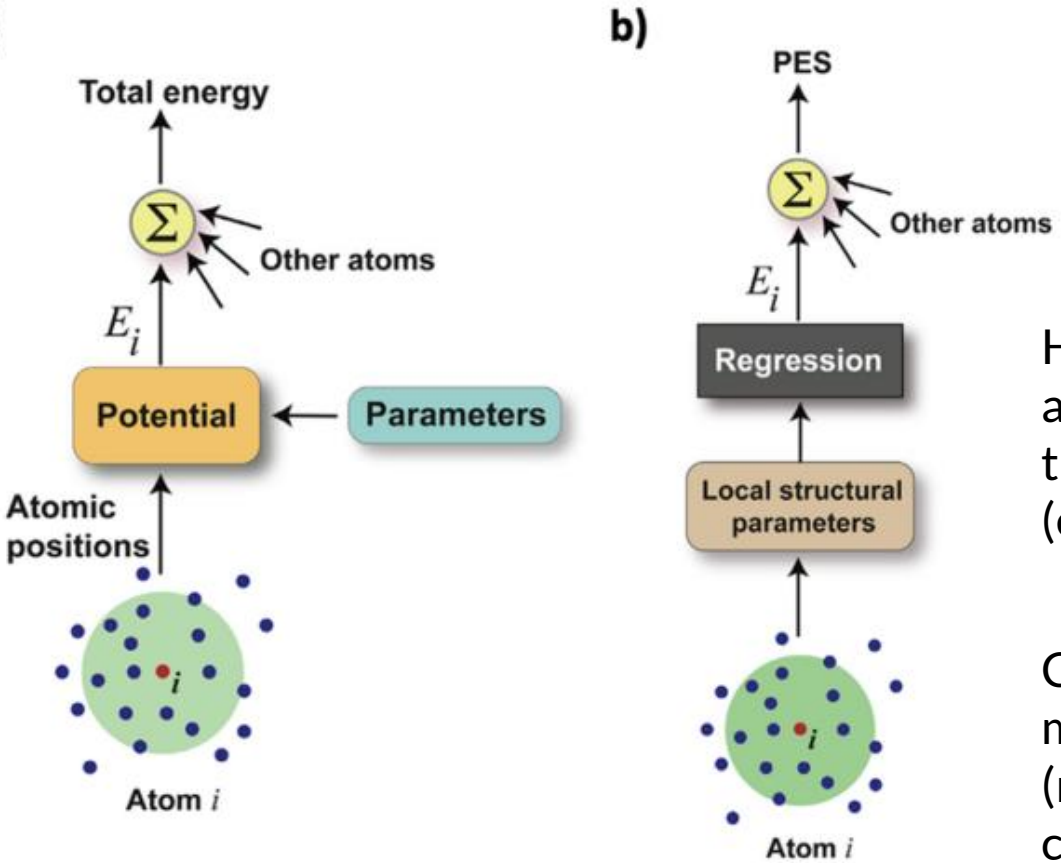
These forces **do not require parametrization** like classical molecular dynamics!

→ *"Ab initio Molecular Dynamics"*

But is computationally **expensive**!

Machine Learning Approach

Highly constrained functional form guided by physical intuition (**specific** to a system)



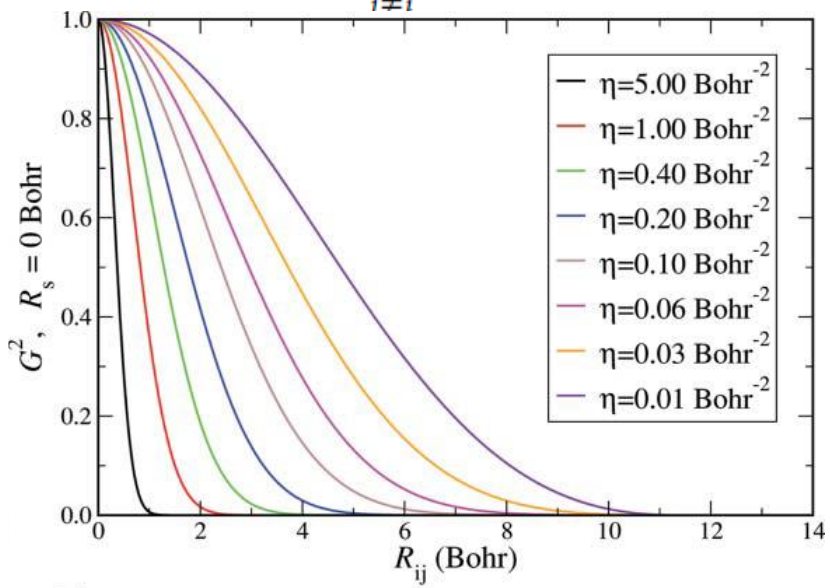
Highly flexible approach to represent the PES function (e.g., neural network)

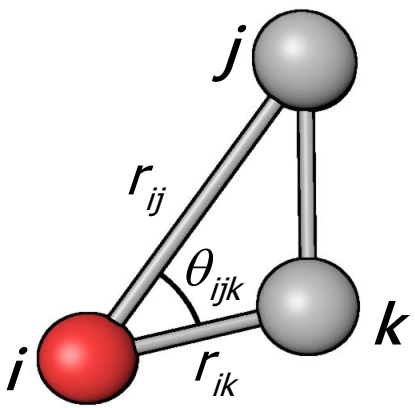
Can be applied to **any** materials system, in principle (metals, ionic, reactive chemistries, etc.)

2000s and 2010s: *Explicit* Atomic Environment Featurization

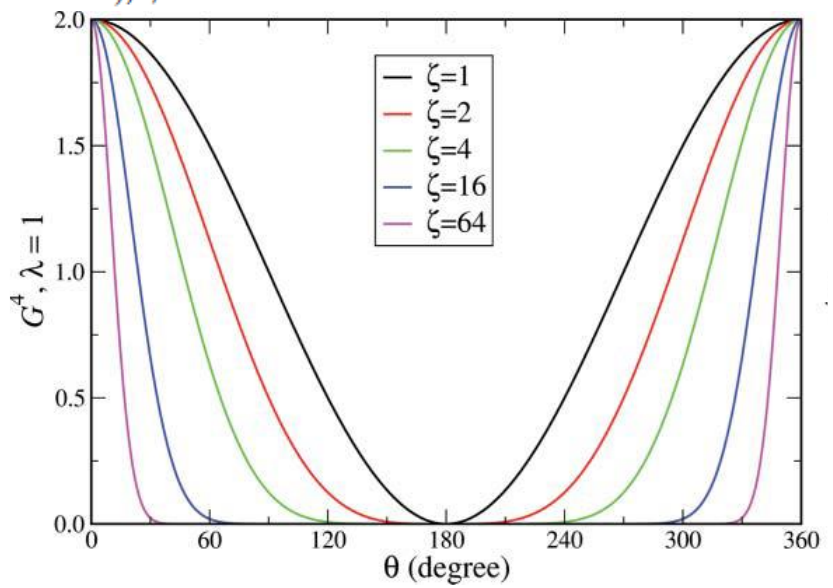
Atom-centered Symmetry functions, **ACSF** (Behler and Parinello) *"Hand-crafted"*

$$G_i^{\text{radial}} = \sum_{i \neq i} \mathrm{e}^{-\eta(r_{ij}-r_s)^2} f_c(r_{ij})$$





$$G_i^{\text{ang.n.}} = 2^{1-\zeta} \sum_{j,k \neq i} (1 + \lambda \cos\theta_{ijk})^{\zeta} \mathrm{e}^{-\eta(r_{ij}^2 + r_{ik}^2 + r_{jk}^2)} f_c(r_{ij}) f_c(r_{ik}) f_c(r_{jk})$$



$$E^{\text{pot}} = \sum_{j=1}^{N^a} E_j(\mathbf{G}_j)$$

$$\vec{F}_i = -\vec{\nabla}_i E^{\text{pot}} = -\sum_{j=1}^{N^a} \vec{\nabla}_i E_j = -\sum_{j=1}^{N^a} \sum_{k=1}^{N_j^s} \frac{\partial E_j}{\partial G_{j,k}} \vec{\nabla}_i G_{j,k}$$

First terms used: "*Neural Network Potentials*" (NNPs) / "*ML Force Fields*" (MLFFs)

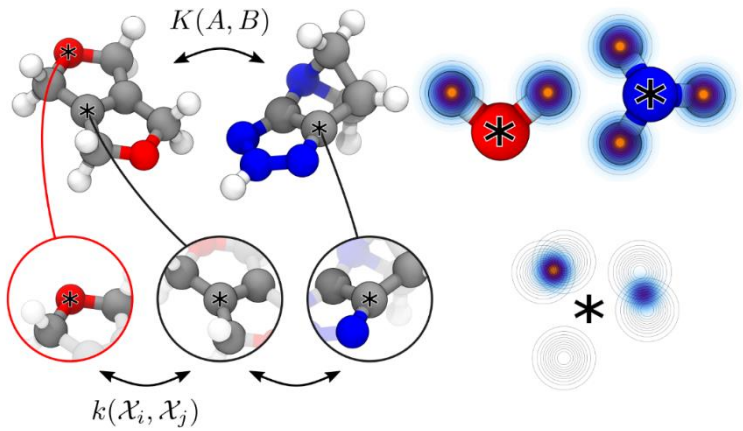
2000s and 2010s: *Explicit* Atomic Environment Featurization

Gaussian Approximation Potential (GAP)
The Smooth Overlap of Atomic Positions (SOAP) + sparse-Gaussian Process Regression (GPR)

The Spectral Neighbor Analysis Potential (SNAP)
Hyperspherical Bispectrum Functions (HBFs) + Linear Regression

The Moment Tensor Potentials (MTP)
Moment tensor functions (MTFs) + Linear Regression

The Atomic Cluster Expansion (ACE)
Radial functions and spherical harmonics + Linear Regression



$$\rho(\mathbf{r}) = \sum_i e^{-\alpha|\mathbf{r}-\mathbf{r}_i|^2}.$$

$$k(\rho, \rho\prime) = \int d\hat{R} \int d\mathbf{r}\, \rho(\mathbf{r})\rho\prime(\hat{R}\mathbf{r}).$$

Drawbacks:
Significant scaling issue!
Computational expense scales with $O((N_b S)^\nu)$

N_b Number of basis functions
S Number of elements
ν Body order (typically 3)

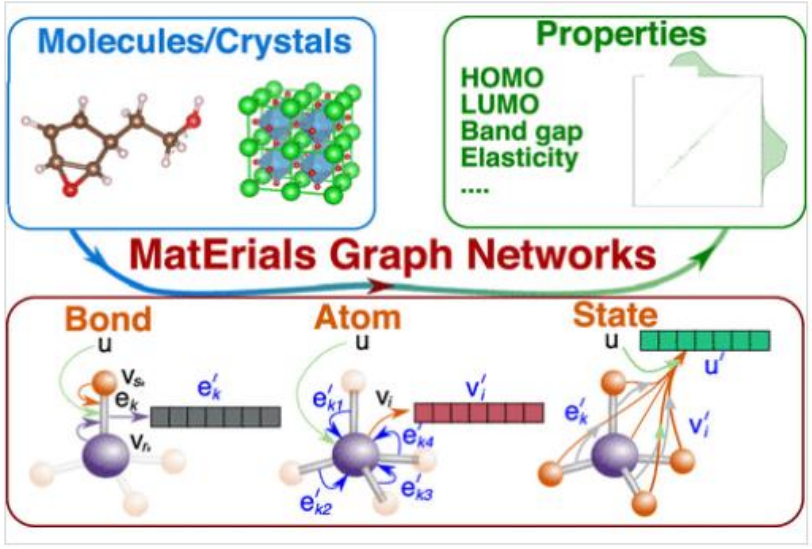
Towards 2020s: *Implicit* Atomic Environment Featurization

Graph Neural Networks (GNNs):

- **SchNet (2017)**: Continuous-filter convolutions (molecules)
- **MEGNet (2019):** Full node/edge/global convolutions (molecules & crystals)
- **DimeNet (2020)**: Directional message passing; 3-body angles (molecules)

Training Datasets:

QM9 (~133k molecules)
MD17 (10 molecules with ~100k steps each)
Materials Project (~60k materials)



2021-2023: E(3)-Equivariant GNNs

NequIP:
E(3)-equivariant message passing

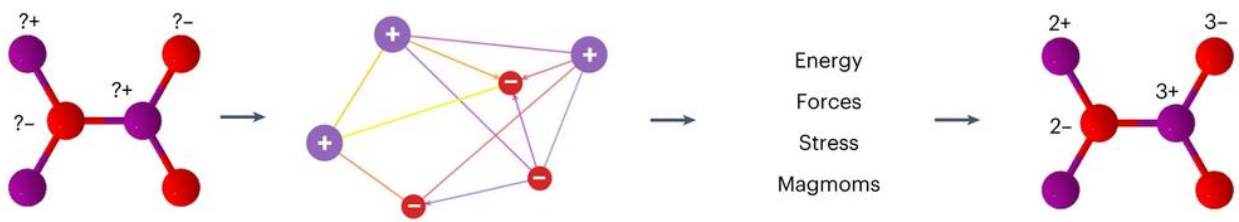
Allegro:
Local equivariant embeddings (avoids message passing) → Much faster

Multiplicative Atomic Cluster Expansion (MACE):
Combines equivariant GNNs with Atomic Cluster Expansion (ACE)
Introduces multiplicative message passing, encoding higher-order correlations

Partially equivariant architectures:

M3GNet:
Incorporates 3-body terms

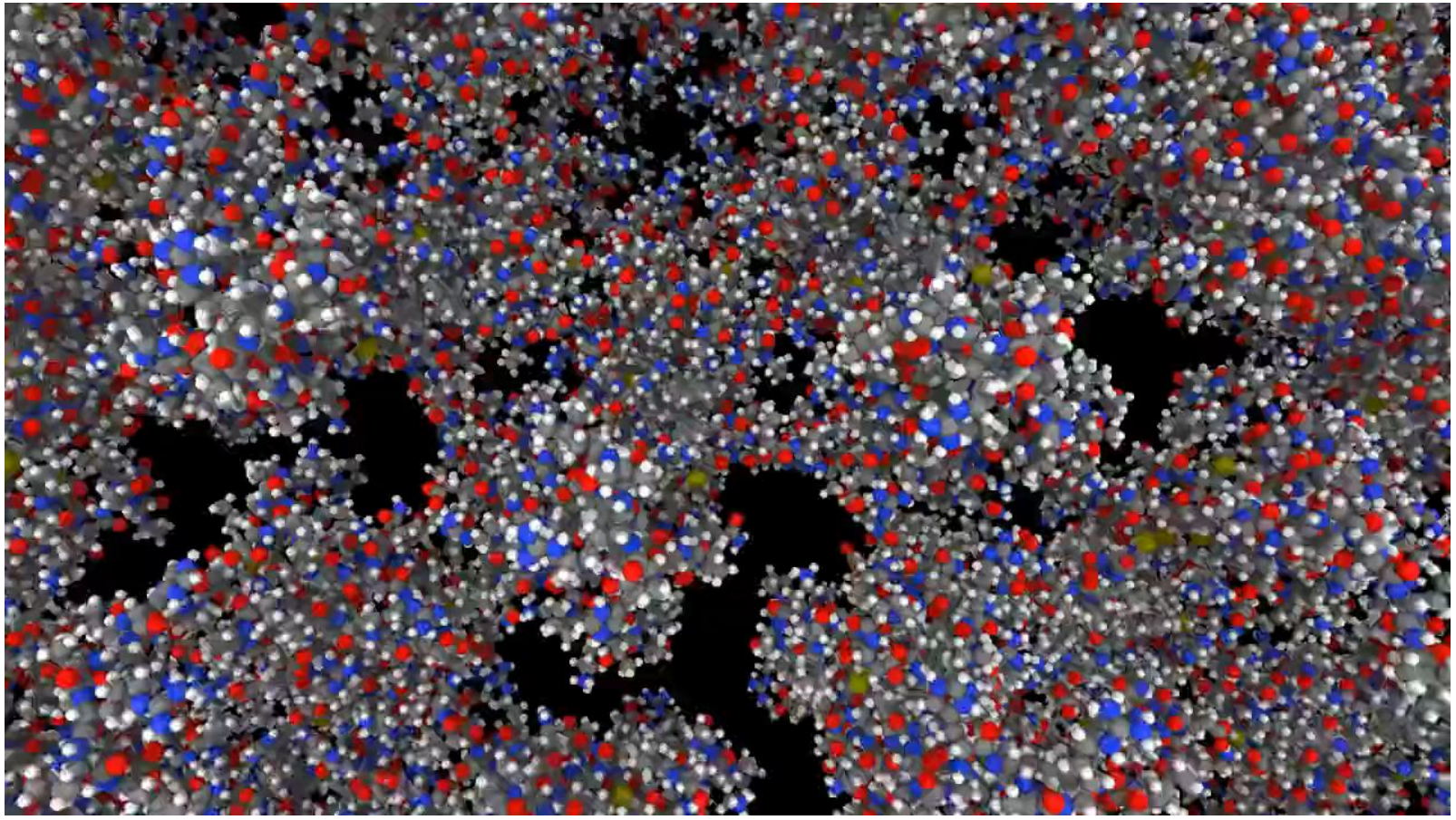
CHGNet:
Can also predict magnetic moments



Trained on **MPTraj** → First claimed "universal" potentials (89 elements)

Allegro: Large-scale AIMD Demonstration

HIV Capsid with 44 Million Atoms



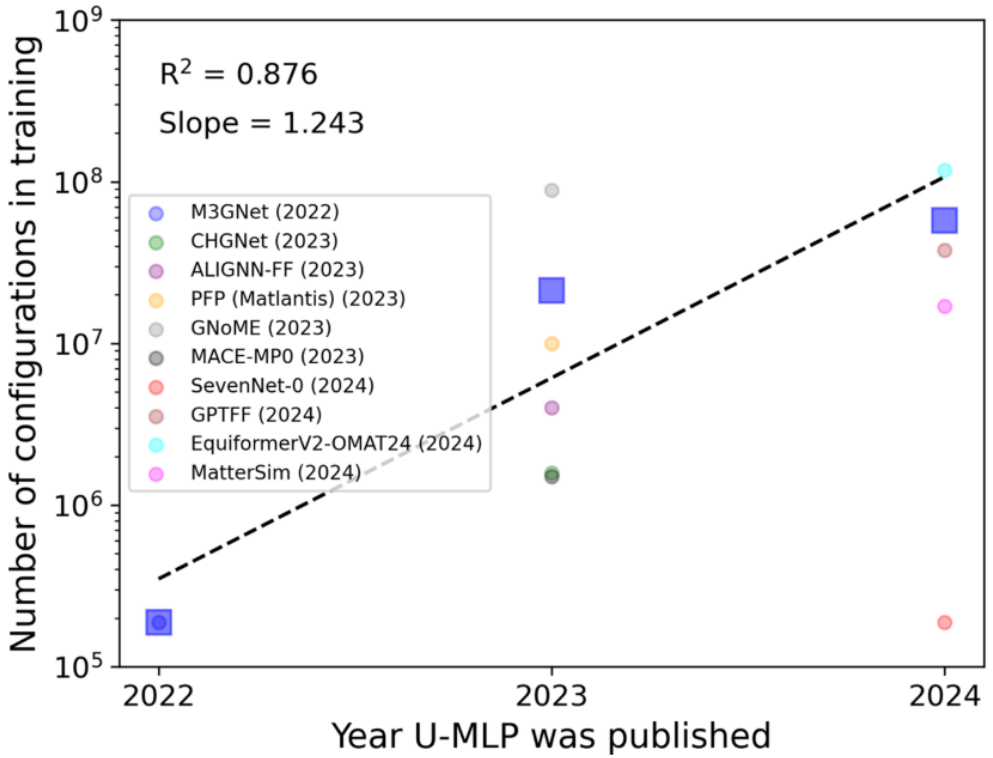
Dr. Simon Batzner

2023-Today: "*Universal*" MLIPs

"Foundational potentials" enabled by large-scale datasets that cover (most of) the periodic table

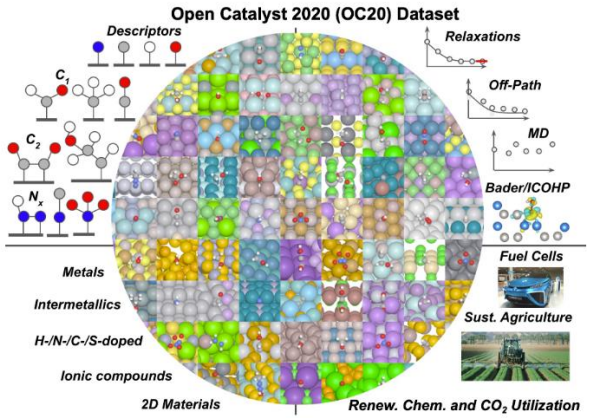
New *transformer*-based architectures:
EquiformerV2, eSEN, UMA

More importantly:
New *large-scale* materials datasets!



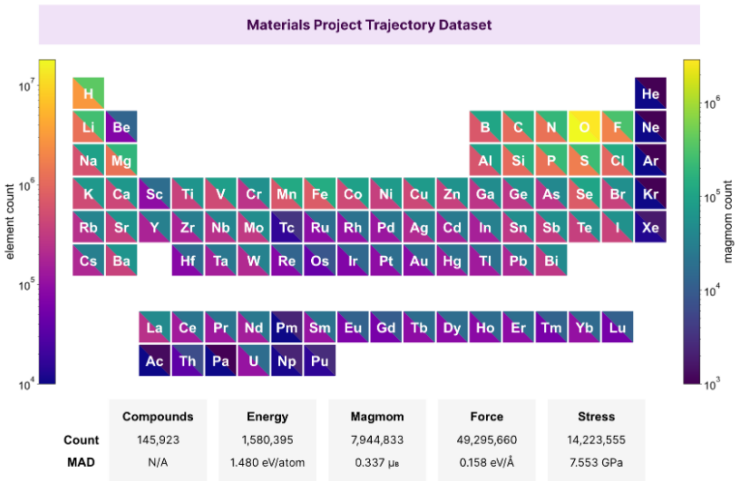
R. Jacobs et al. *Current Opinion in Solid State & Materials Science* 35 (2025) 101214

Large-Scale Materials Dataset



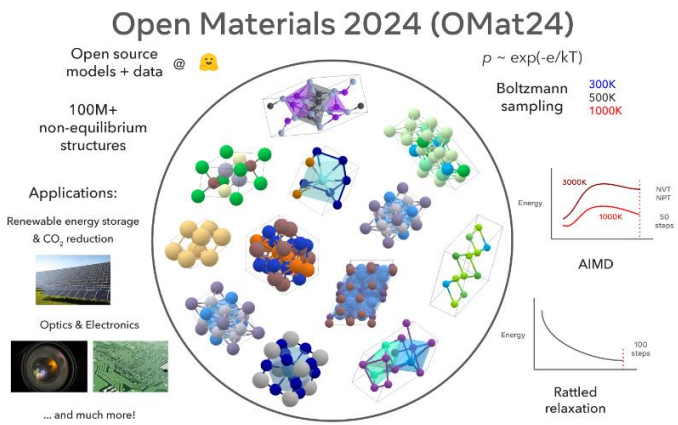
OC20
(2020)
1.28M
Surface+Adsorbate

MPtrj
(2023)
1.58M
89 Elements
Magnetic moments

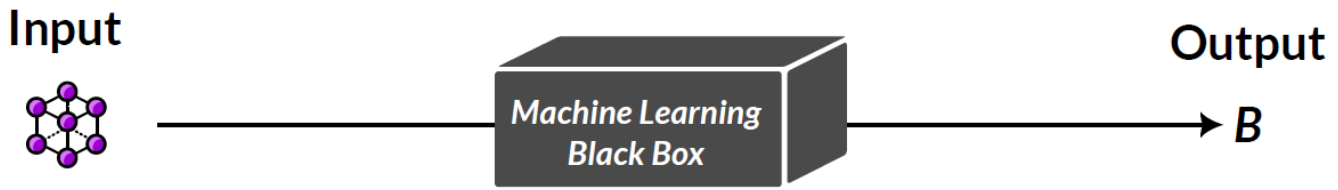


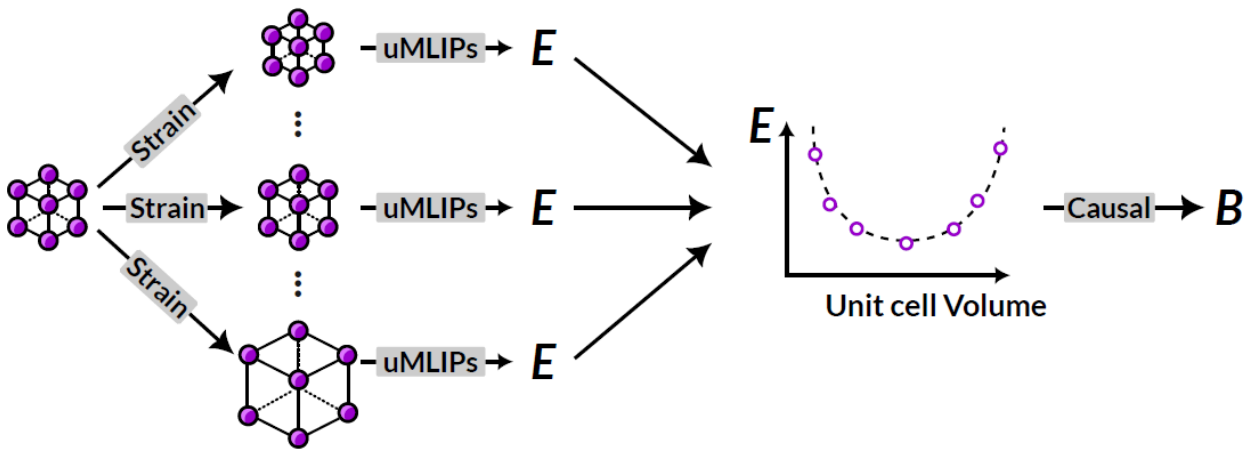
Alexandria
(2023)
5M
1D, 2D, 3D materials

OMat24
(2024)
118M
Extensive non-equilibrium sampling



Properties Derived from the Potential Energy Surface (PES)

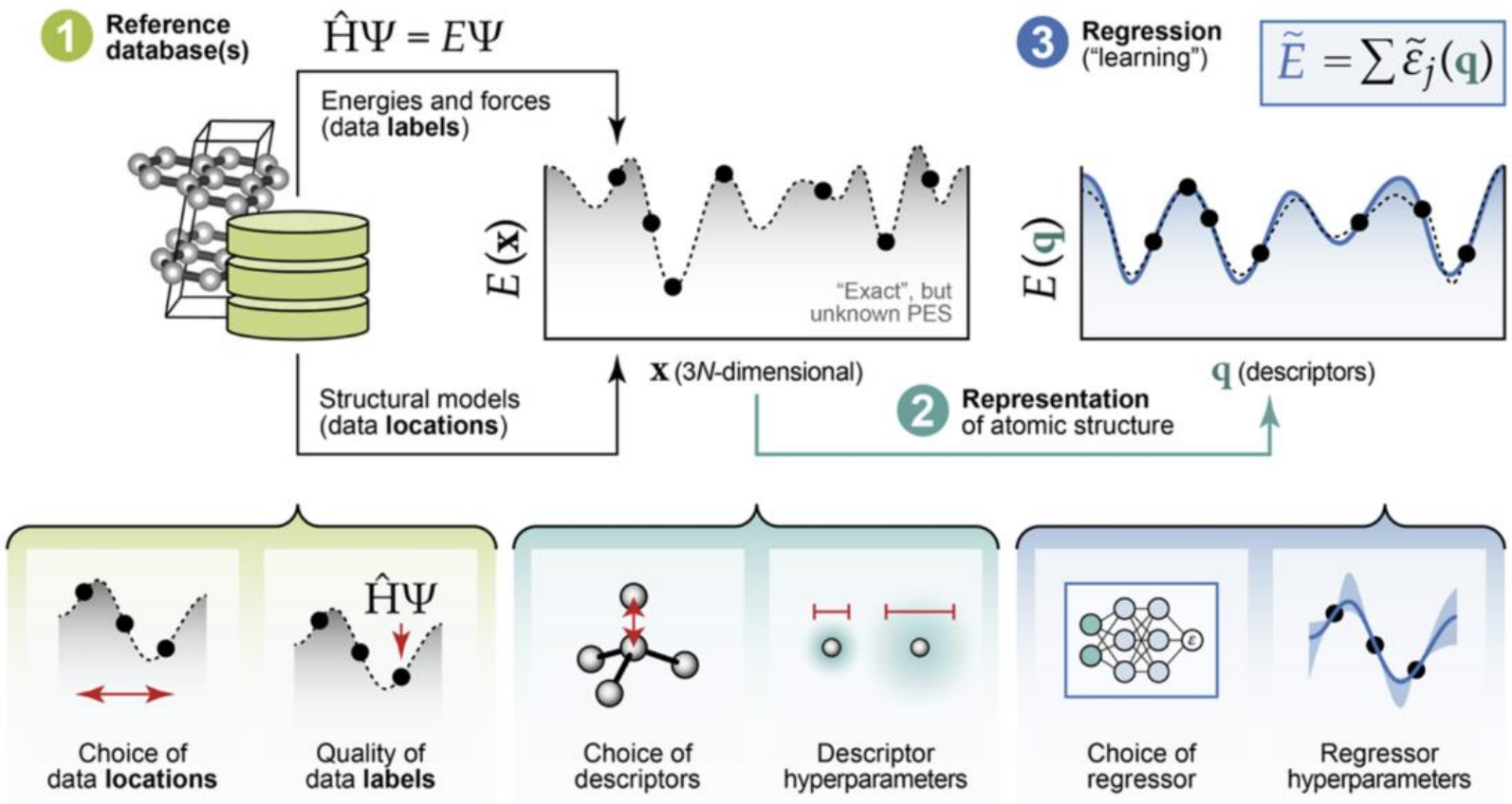




Bulk modulus can be computed using the *Equation of State* (EOS)

Other properties that can be derived from PES:
Formation energy, cohesive energy, elastic tensor, elastic modulus, Young's modulus, phonons (hence, thermal and thermodynamic properties), surface energy

MLIP Development

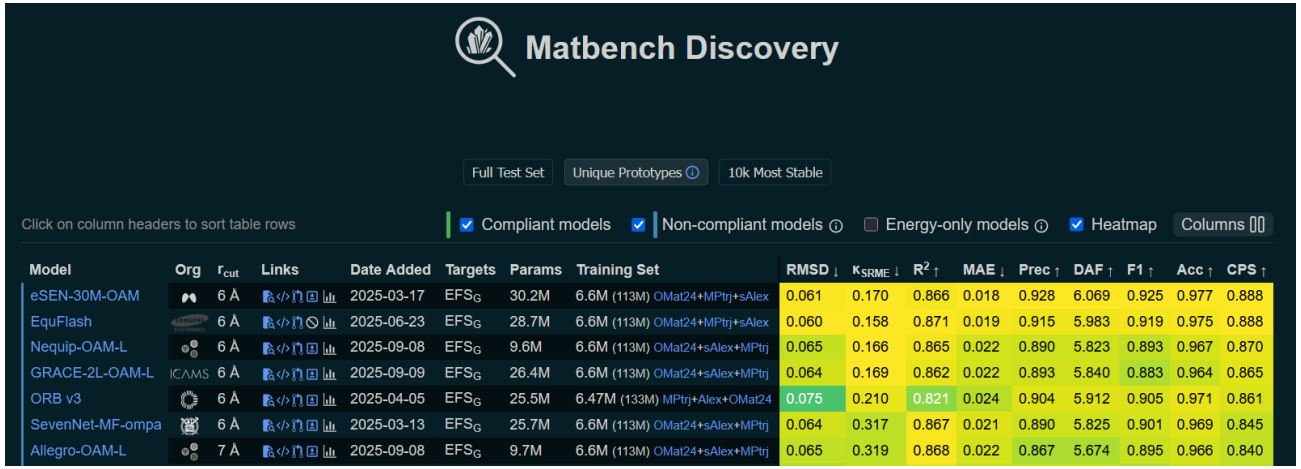


R. Jacobs et al. *Current Opinion in Solid State & Materials Science* 35 (2025) 101214

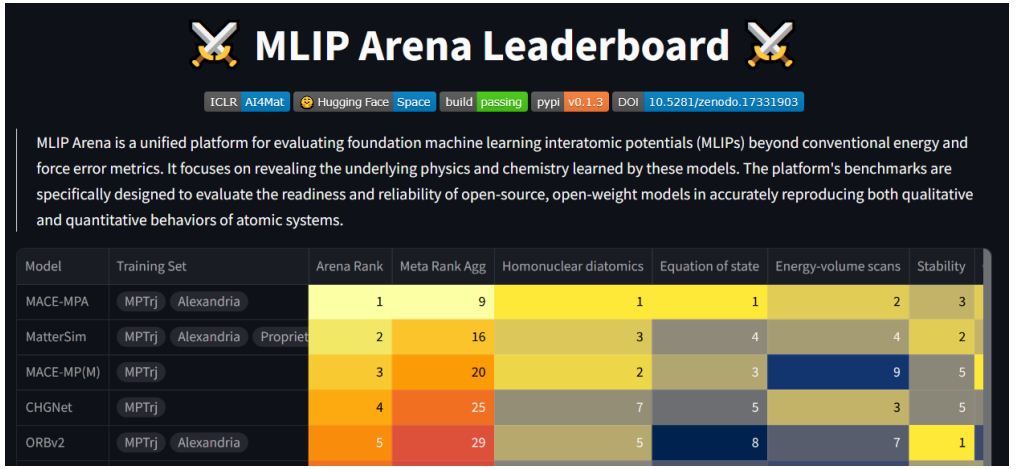
MatBench Discovery and MLIP Arena

https://matbench-discovery.materialsproject.org/

https://huggingface.co/spaces/atomind/mlip-arena



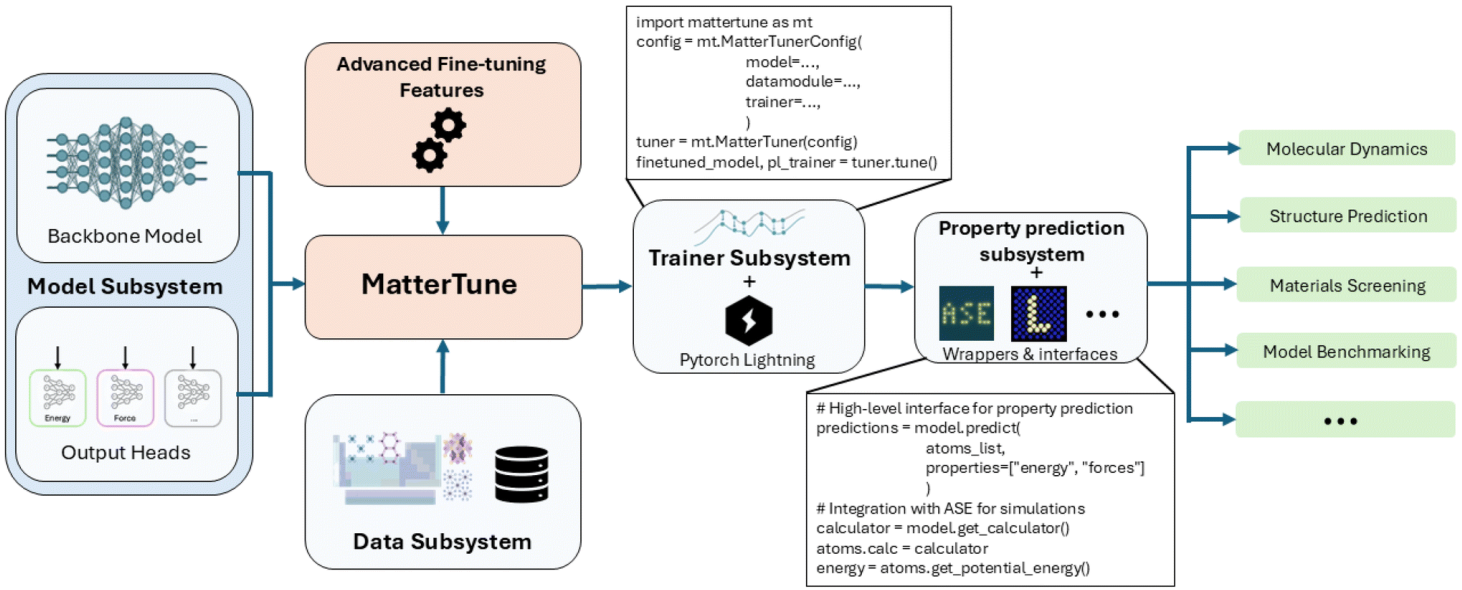
Model	Org	r_{cut}	Links	Date Added	Targets	Params	Training Set	RMSD ↓	κ_{SRME} ↓	R^2 ↑	MAE ↓	Prec ↑	DAF ↑	F1 ↑	Acc ↑	CPS ↑
eSEN-30M-OAM		6 Å		2025-03-17	EFS_G	30.2M	6.6M (113M) OMat24+MPtrj+sAlex	0.061	0.170	0.866	0.018	0.928	6.069	0.925	0.977	0.888
EquFlash		6 Å		2025-06-23	EFS_G	28.7M	6.6M (113M) OMat24+MPtrj+sAlex	0.060	0.158	0.871	0.019	0.915	5.983	0.919	0.975	0.888
Nequip-OAM-L		6 Å		2025-09-08	EFS_G	9.6M	6.6M (113M) OMat24+sAlex+MPtrj	0.065	0.166	0.865	0.022	0.890	5.823	0.893	0.967	0.870
GRACE-2L-OAM-L	ICAMS	6 Å		2025-09-09	EFS_G	26.4M	6.6M (113M) OMat24+sAlex+MPtrj	0.064	0.169	0.862	0.022	0.893	5.840	0.883	0.964	0.865
ORB v3		6 Å		2025-04-05	EFS_G	25.5M	6.47M (133M) MPtrj+Alex+OMat24	0.075	0.210	0.821	0.024	0.904	5.912	0.905	0.971	0.861
SevenNet-MF-ompa		6 Å		2025-03-13	EFS_G	25.7M	6.6M (113M) OMat24+sAlex+MPtrj	0.064	0.317	0.867	0.021	0.890	5.825	0.901	0.969	0.845
Allegro-OAM-L		7 Å		2025-09-08	EFS_G	9.7M	6.6M (113M) OMat24+sAlex+MPtrj	0.065	0.319	0.868	0.022	0.867	5.674	0.895	0.966	0.840



Model	Training Set	Arena Rank	Meta Rank Agg	Homonuclear diatomics	Equation of state	Energy-volume scans	Stability
MACE-MPA	MPTrj Alexandria	1	9	1	1	2	3
MatterSim	MPTrj Alexandria Propriet	2	16	3	4	4	2
MACE-MP(M)	MPTrj	3	20	2	3	9	5
CHGNet	MPTrj	4	25	7	5	3	5
ORBv2	MPTrj Alexandria	5	29	5	8	7	1

Fine-Tuning of MLIPs

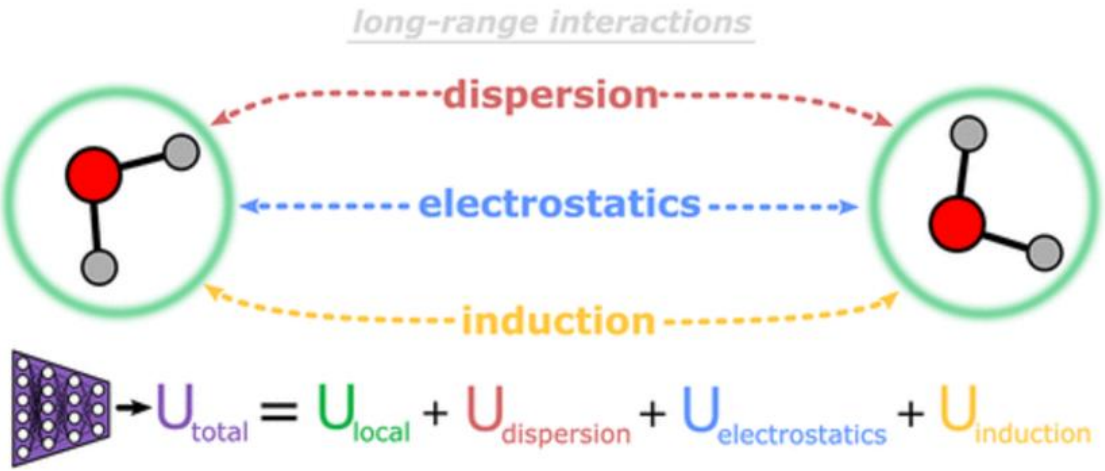
Common Approach:
Start with the Foundational Potentials and then **fine-tune on a specific subset** of data



L. Kong et al., Digital Discovery, 2025, 4, 2253-2262

Recent Developments and Future Directions for MLIPs

- Longe-range interactions (e.g., electrostatics)



- Higher-order functionals (e.g., R2SCAN)
- Model distillation
- Non-Born-Oppenheimer PES (e.g., magnetism, excited states)
- Foundational MLIPs for electronic structure (i.e., band diagrams)

D.M. Anstine, O. Isayev, Chem. A Eur. J. 127 (2023) 2417–2431

Lecture Feedback

Please, scan the QR code and take a minute to let me know how the lecture was and mention any **feedback/questions**

This form is **anonymous!**