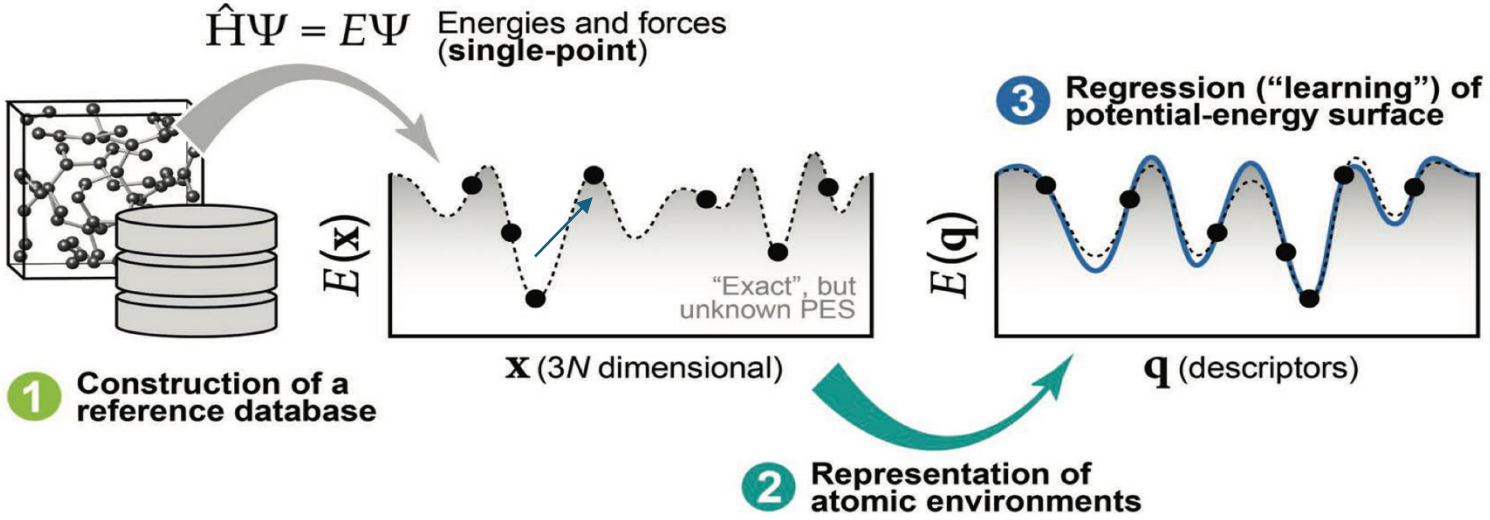
Introduction to Machine-learned Interatomic Potentials (MLIPs)

Nima Karimitari

Find the potential energy surface with a few Quantum chemistry calculations



$$E_{\text{pot}}(R, Z) = \sum_{i}^{N} E_i(\{\boldsymbol{r}_j, z_j\}_{j \in \mathcal{N}(i)})$$

Atomic positions — Chemical elements

- Fit a parametric function (Linear models, Neural Networks (NNs)):

$$E_i = V^{(1)}(\mathbf{r}_i) + \frac{1}{2}\sum_j V^{(2)}(\mathbf{r}_i, \mathbf{r}_j) + \frac{1}{3!}\sum_{jk} V^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \ \ldots$$

- Reference data : DFT, CCSD, MP2…
- Goal : Reference accuracy without the electrons

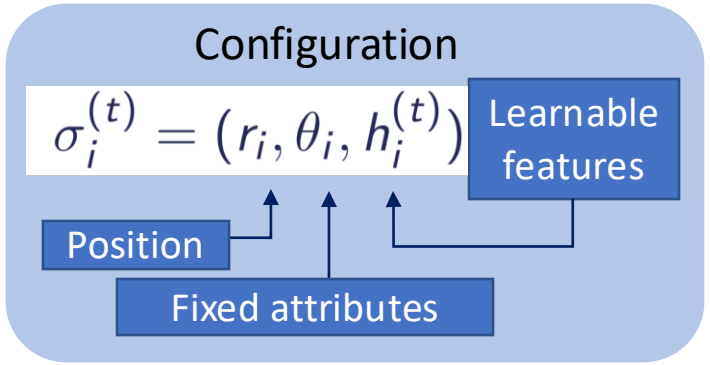
Adv. Mater. **2019**, *31*, 1902765

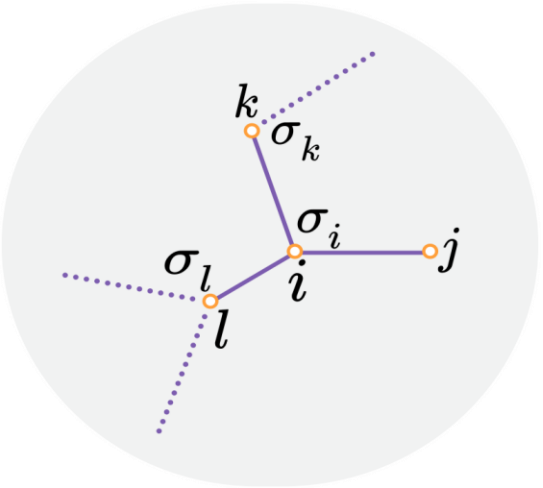
Active field in the past decade

- Neural Network
  - Behler–Parrinello neural network (BPNN) [1]
  - Deep potential for molecular dynamics (DeePMD) [2]
- Kernel methods
  - Gaussian approximation potentials (GAP) [3]
  - Adaptive, generalizable, and neighborhood informed (AGNI) force fields [4]
  - aenet [5]
- Linear models
  - Spectral neighbor analysis potential (SNAP) [6]
  - Atomic cluster Expansion (ACE) [7]
- Equivariant graph neural networks
  - Multi-ACE (MACE) [8]
  - NequiP [9]
  - Equiformer [10]

[1] Phys. Rev. Lett. 2007, **98**, 146401.
[2] Comput. Phys. Commun. 2018, **228**, 178.
[3] Phys. Rev. Lett. 2010, **104**, 136403.
[4] Phys. Rev. B 2015, **92**, 094306.
[5] J. Comput. Phys. 2015, **285**, 316.
[6] Comput. Mater. Sci.2016, **114**, 135.
[7] Phys. Rev. B **100**, 249901 (2019)
[8] J. Chem. Phys. 159, 044118 (2023)
[9] Nat Commun **13**, 2453 (2022)
[10] arXiv:2306.12059

Message passing neural networks (MPNN)





- MPNNs Potentials have 3 phases :
  1. The message construction phase,

$$\boldsymbol{m}_i^{(t)} = \frac{1}{\lambda} \bigoplus_{j \in \mathcal{N}(i)} M_t(\sigma_i^{(t)}, \sigma_j^{(t)})$$

  - $N(i)$: Neighbors of i within a cut-off radius
  - $\bigoplus$: Permutationally invariant pooling

  2. The update phase,

$$\sigma_i^{(t+1)} \equiv (\boldsymbol{r}_i, \boldsymbol{\theta}_i, \boldsymbol{h}_i^{(t+1)}) = (\boldsymbol{r}_i, \boldsymbol{\theta}_i, U_t(\sigma_i^{(t)}, \boldsymbol{m}_i^{(t)}))$$

  3. The readouts phase,

$$E_i = \sum_t \mathcal{R}_t(\sigma_i^{(t)})$$

Ilyes Batatia et al. The Design Space of E(3)-Equivariant Atom-Centered Interatomic Potentials arXiv:2205.06643

Multi-ACE (MACE)

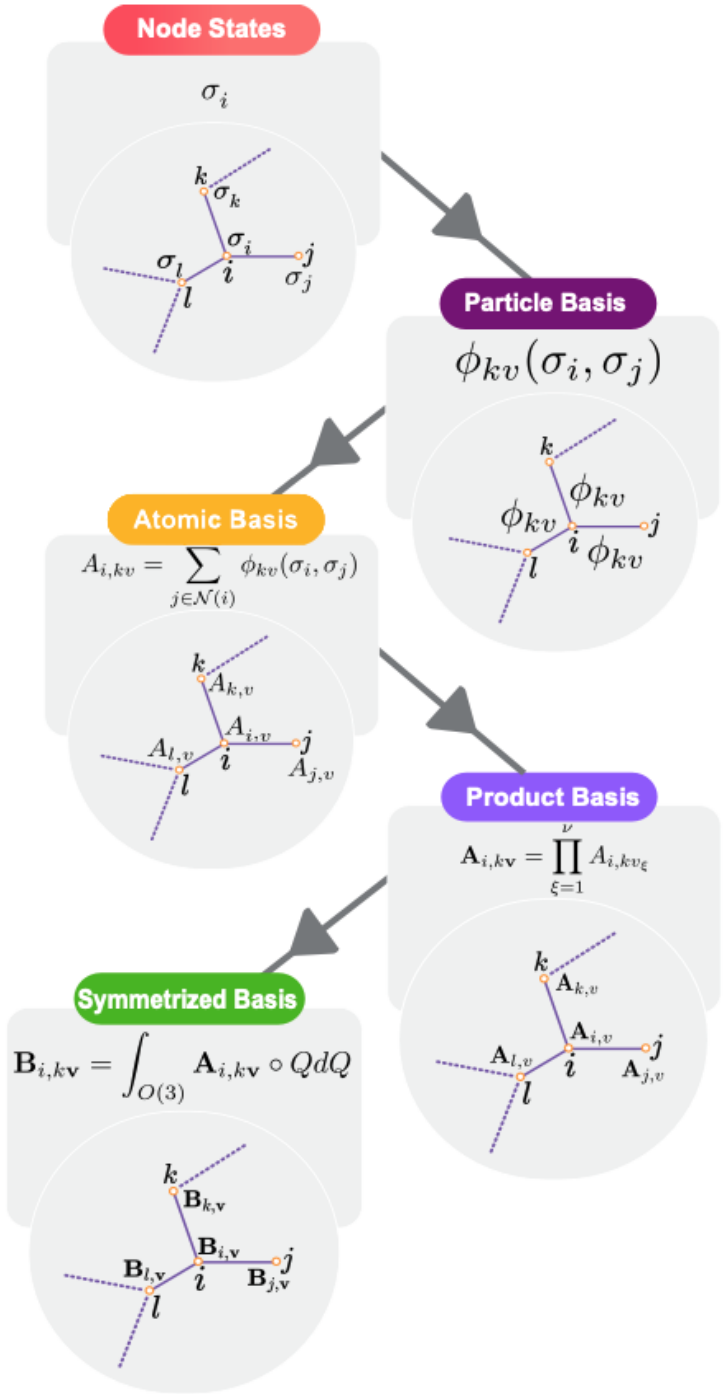
Higher order message passing in MACE

Construction of higher-body order of ACE basis

1. Neighborhood graph
2. The one-particle basis
   - Edge features
   - Directed graph
3. A basis
   - Pooling over the neighbors
   - Density trick
   - Higher order interaction
4. B basis
   - Averaging over all O(3) rotations
   - Complete basis

Using 4-body messages,

two layers => faster and parallelizable potential



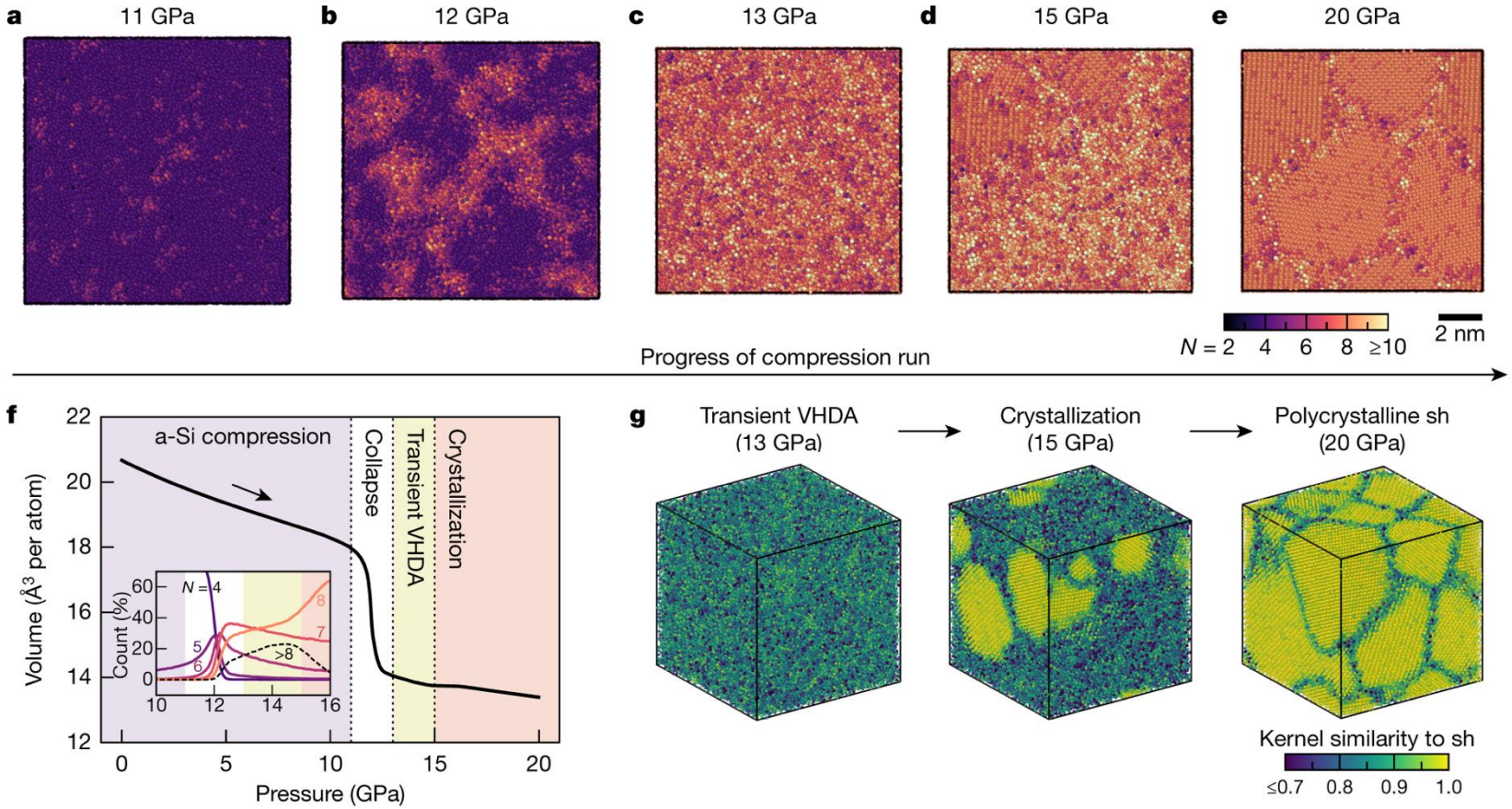
Ilyes Batatia et al. The Design Space of E(3)-Equivariant Atom-Centered Interatomic Potentials arXiv:2205.06643
Ilyes Batatia et al. Advances in Neural Information Processing Systems 35 (NeurIPS 2022)
Kovacs et al. J. Chem. Phys. 159, 044118 (2023)

What properties can you find with MLIPs and what are the limitations?

- Anything that only needs energy, forces, virials, etc.
  - Structural optimization
  - Molecular Dynamics
- Properties that do not need explicit electrons
  - Adsorption energy, Decomposition energy, formation energy
  - Phonons (harmonic and anharmonic)
  - Elasticity
  - Nudged Elastic Band calculations
  - Structural phase transitions
  - Charged systems (some recent progress on this subject)[1]
- When charge density is needed, you need beyond MLIP
  - Polarization tensor
  - Dielectric function

[1] arXiv:2406.10915

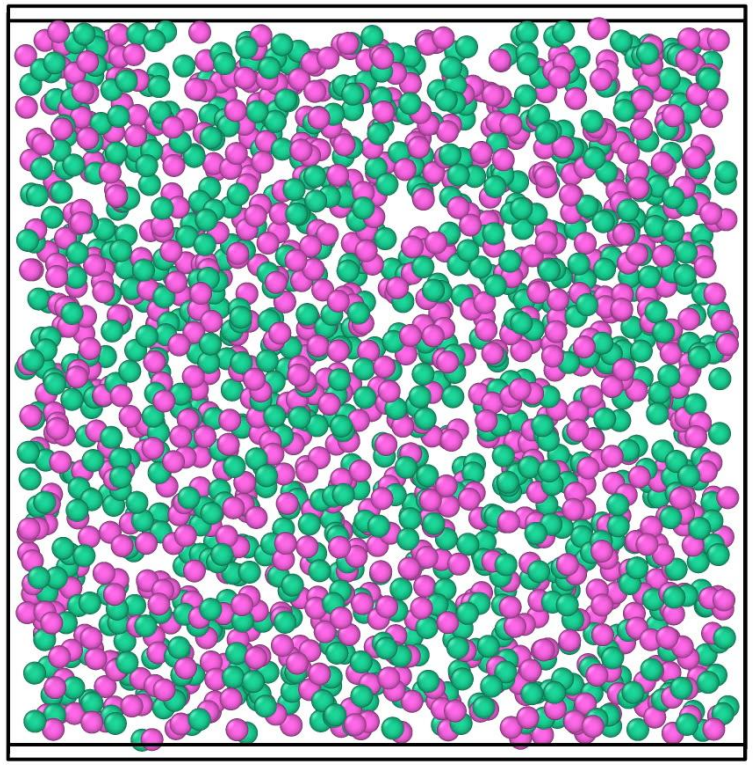
Silicon amorphous under very high pressure

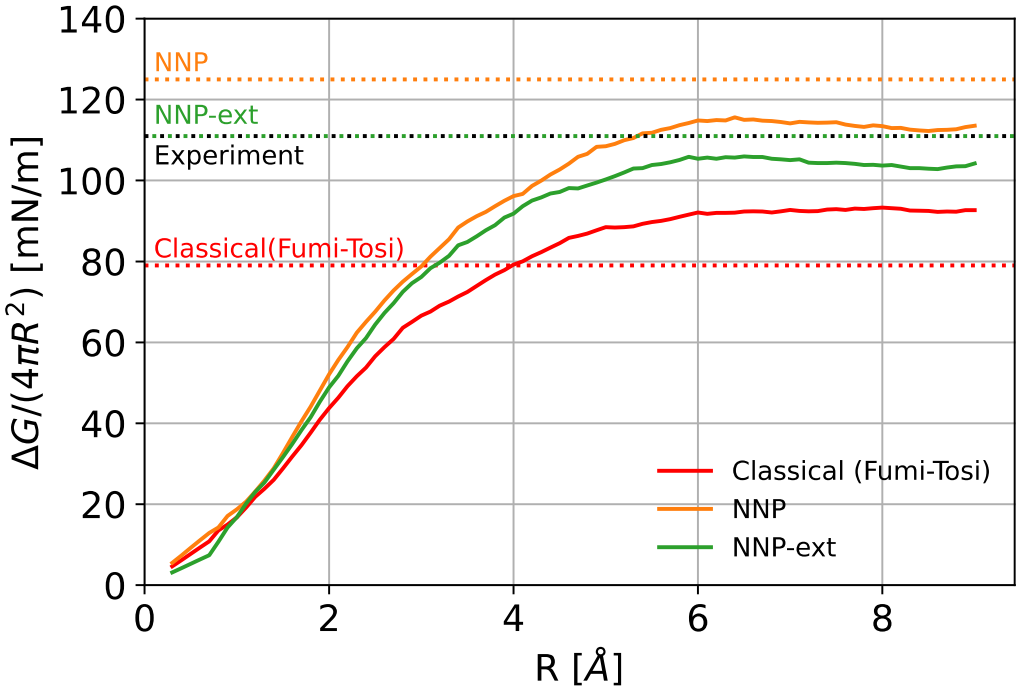


- 100,000 silicon atoms
- Transition from amorphous to very high density amorphous (VHDA
- Polycrystalline at high pressures
- Empirical methods do not observe either VHDA or polycrystalline

Nature **volume 589**, pages 59–64 (2021)

Modeling inhomogeneous response of liquids



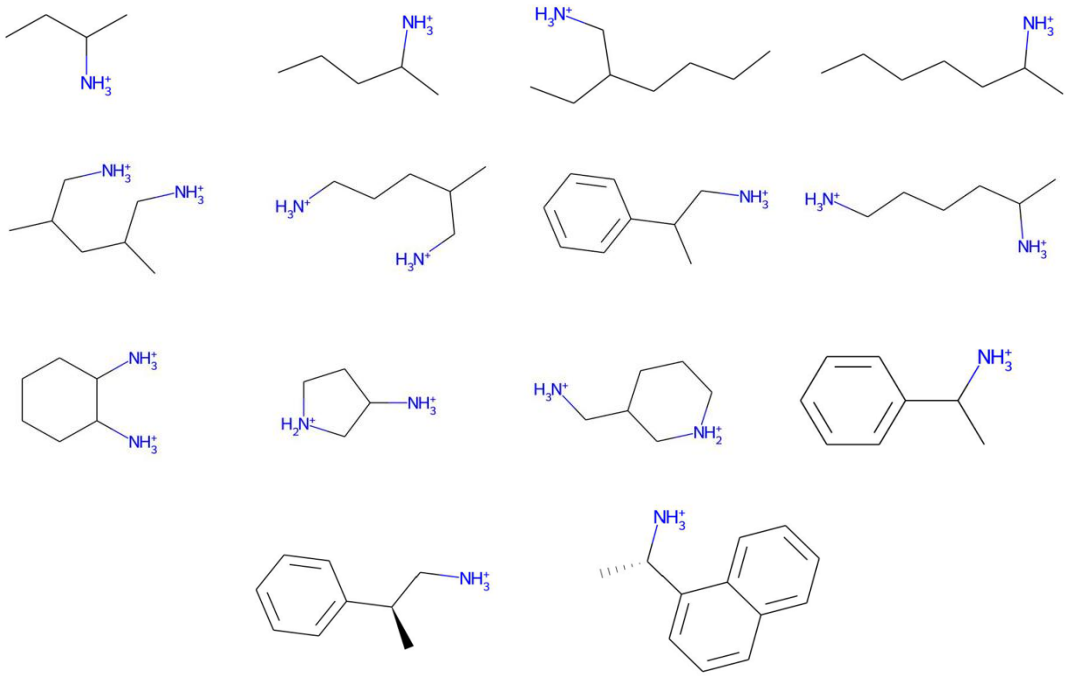


Cavitation free energy in agreement with experimental surface tension predictions

J Comput Chem. 2024;45:1821–1828.

Predict the structure for arbitrary combinations of organic cations and inorganic components

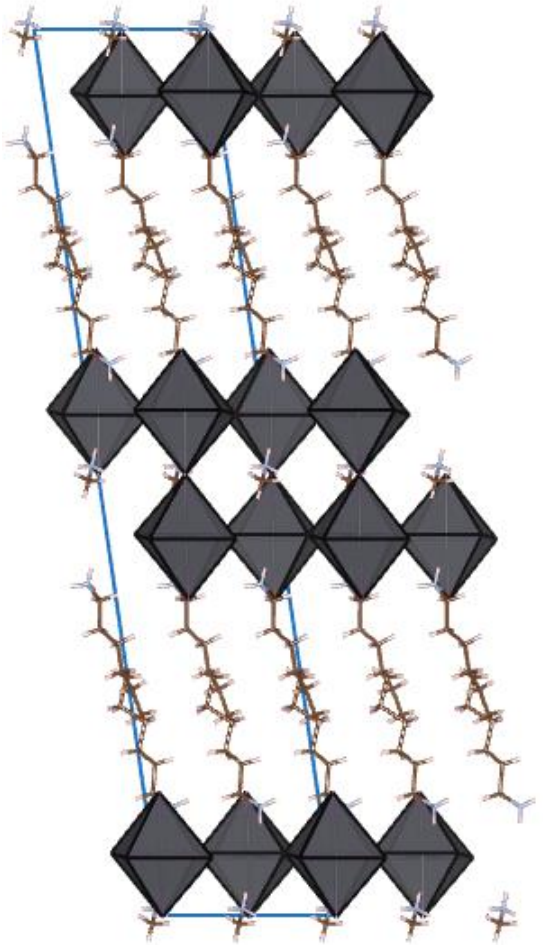
1. Select organic cation



2. Select metal cation and halide (I/Cl/Br)
   e.g., Mn, Fe, Co, Ni, Cu, Zn, Cd, Pb, Sn, Bi, Sb
   Or even B/B'

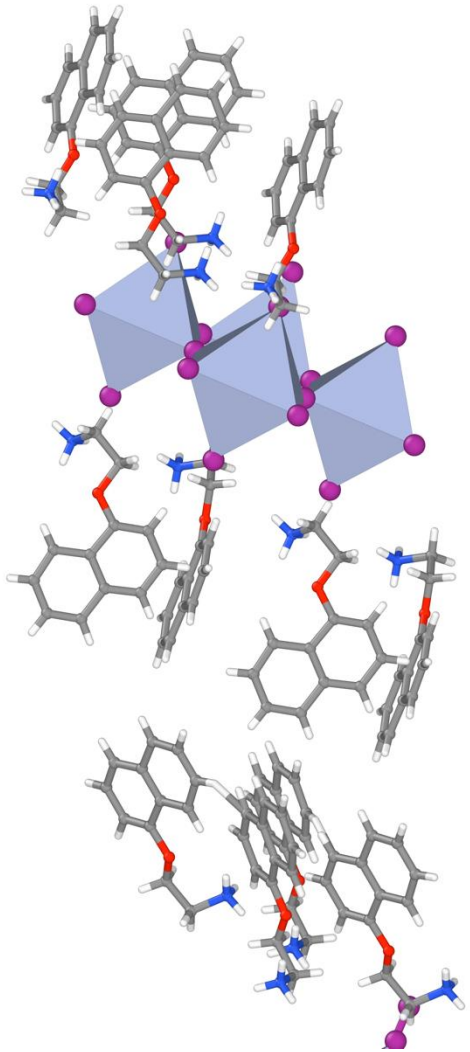
$f^{ML}(x)$

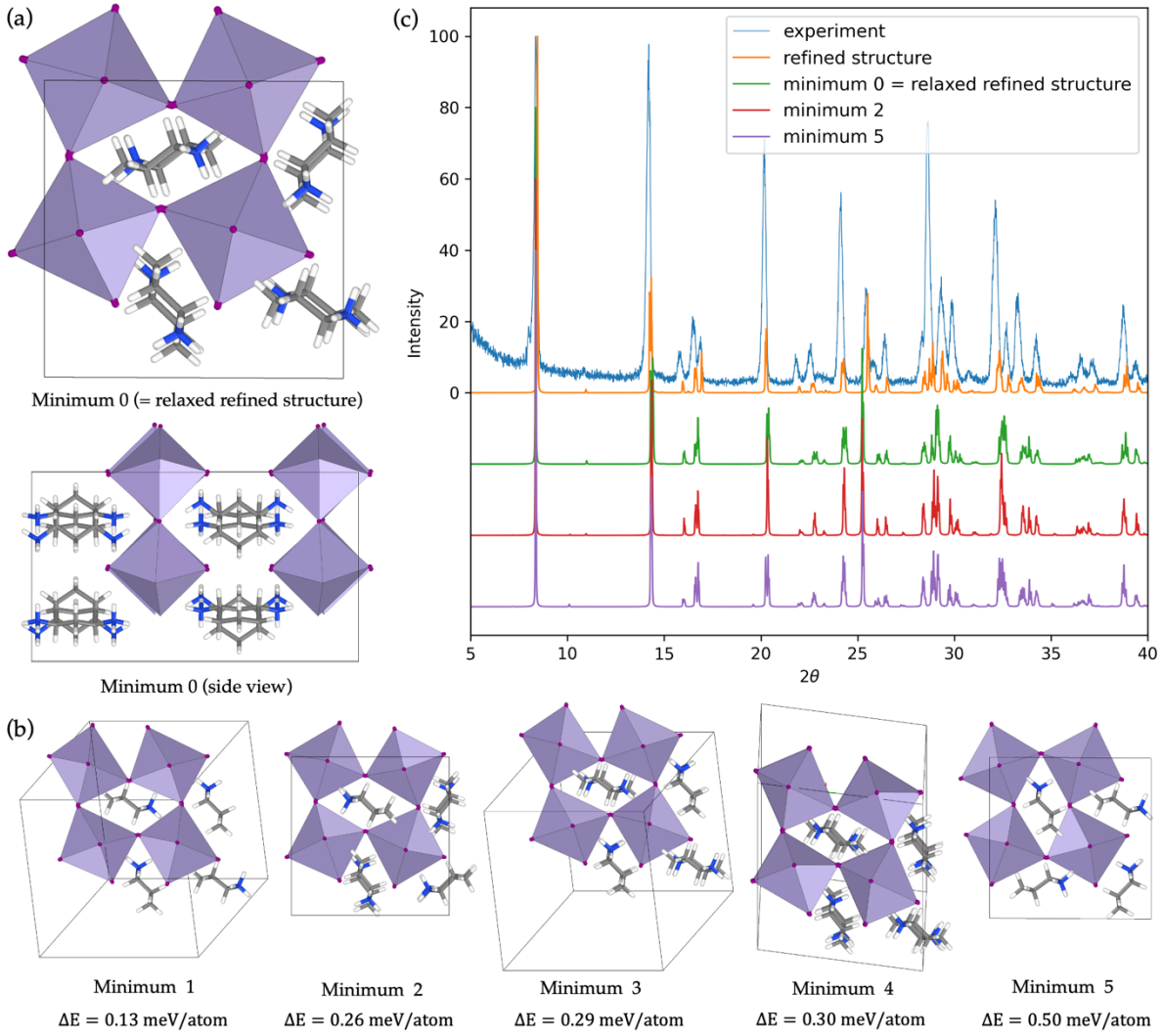
Predict structure



arXiv:2403.06955

Propose ne sstructures that have been synthesized before





arXiv:2403.06955

Foundation (Universal) models as way to go for training new models

- Models trained to materials project dataset
  - CHGNet [1]
  - M3GNET [2]
  - MACE-MP0 [3]
- Very stable but less accurate than local models
- Best suited for fine-tuning
  - A huge computational advantage in training new models
  - Small datasets with big computational cost (i.e. HSE06)
  - Generalizability to other environments
  - Explicit solvent in different env
- Multi-head approach
  - Train a to two datasets simultaneously
  - Predict two sets of results (e.g. PBE and HSE06 energies)

[1] Nature Machine Intelligence **volume 5**, 1031–1041 (2023)
[2] Nature Computational Science **volume 2**, 718–728 (2022)
[3] arXiv:2401.00096

The exercises that you will try today

- How to use a trained model with ASE objects
  - MACE-MP0
- Surface with adsorbate
  - Construct a surface slab from a relaxed bulk structure
  - Relax an adsorbate on the surface area
- Molecular Dynamics
  - NVT simulation for a box of water