Quantum-accurate Force Fields from Machine Learning of Large Materials Data

Xiang-Guo Li, Zhi Deng, Yunxing Zuo, Chi Chen,

Shyue Ping Ong

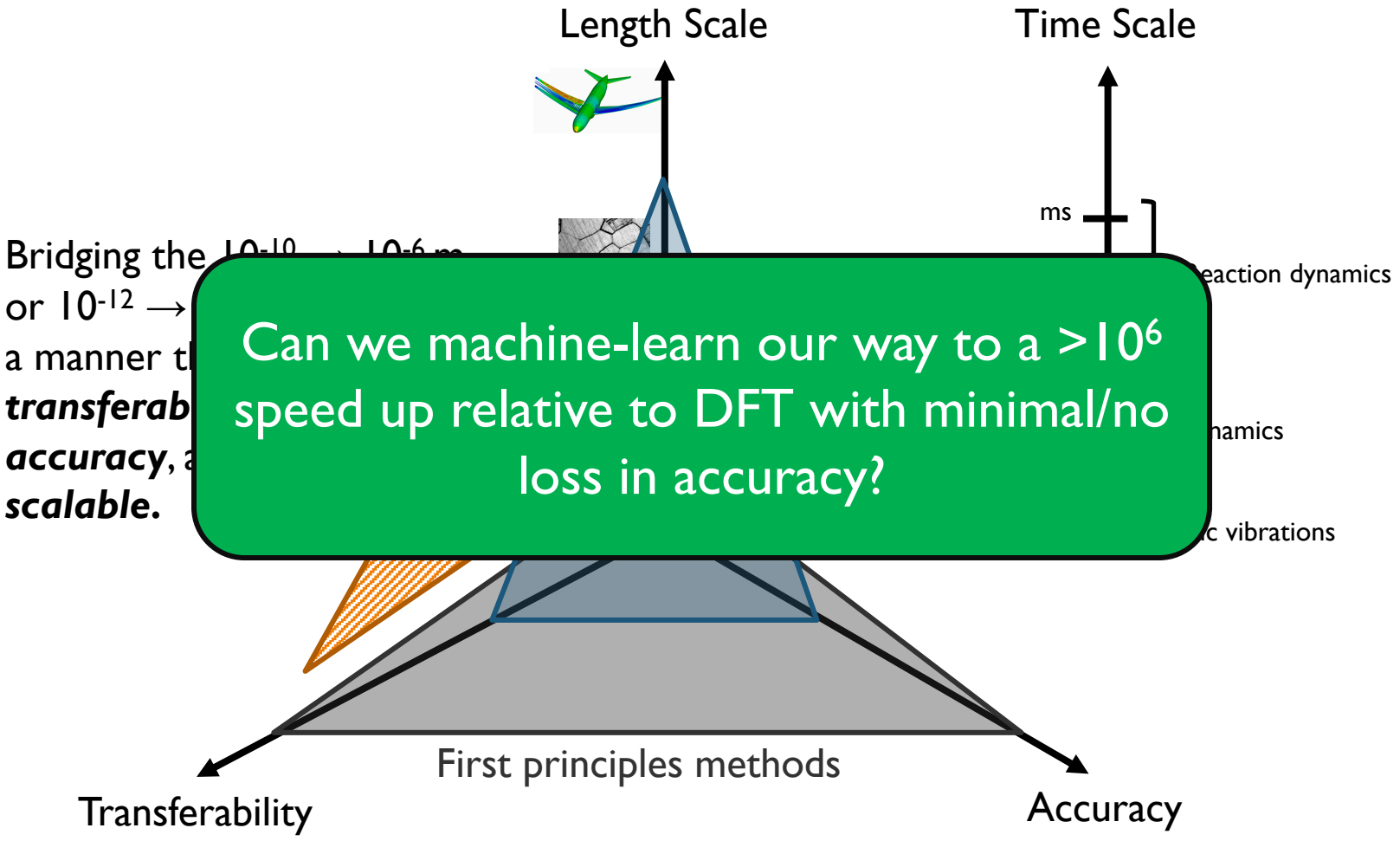
8/29/19

Albuquerque 2019

UC San Diego
Jacobs School of Engineering

materials virtual·lab

Scale Challenge in Computational Materials Science



General procedure

Sample a sufficiently large dataset

Describe local environment

Learn relationship between features and energy, force, etc.

Open databases

MATERIALS PROJECT

OQMD

pymatgen

FireWorks

ASE

❑ Requirements
- Invariance to rotation,

Spectral neighbor analysis potential (SNAP)

- Symmetry functions
- Bispectrum
- Smooth overlap of atomic positions
- Fragment descriptors
- …

gression

lge regression

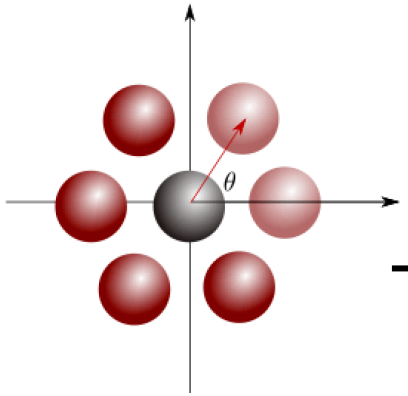
orest

Neural networks

❑ ….

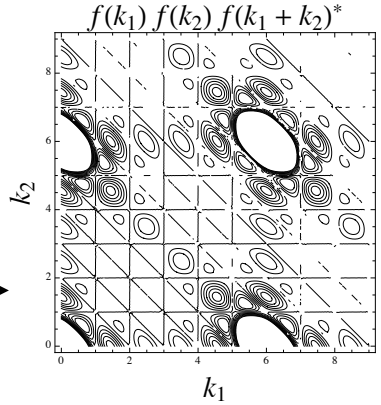
A many-body atomic environment descriptor: bispectrum coefficients

- Map neighbors into unit sphere in 4D
- Expand density in 4D spherical harmonics



$$\rho(\mathbf{r}) = \sum_{j=0,\frac{1}{2},\ldots}^{\infty} \sum_{m=-j}^{j} \sum_{m'=-j}^{j} u^j_{m,m'} U^j_{m,m'}(\theta, \phi, \theta_0)$$

Bispectrum



$$\rho_i(\mathbf{r}) = \delta(\mathbf{r}) + \sum_{r_{ii'} < R_c} f_c(r_{ii'}) w_{i'} \delta(\mathbf{r} - \mathbf{r}_{ii'})$$

$$B_{j_1,j_2,j} = \sum_{m_1,m_1'=-j_1}^{j_1} \sum_{m_2,m_2'=-j_2}^{j_2} \sum_{m,m'=-j}^{j} \left(u^j_{m,m'}\right)^* H^{jmm'}_{j_1 m_1 m_1' j_2 m_2 m_2'} u^{j_1}_{m_1,m_1'} u^{j_2}_{m_2,m_2'}$$

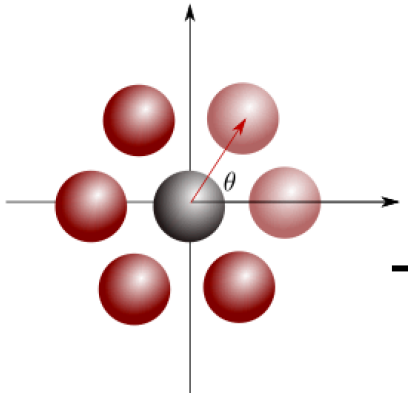
w^{atom} **distinguishes species**

j^{max} order of Bispectrum, set to 3.

R_c^{atom} **radius cutoff**

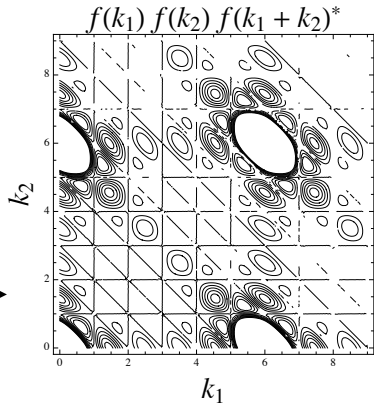
A many-body atomic environment descriptor: bispectrum coefficients

- Map neighbors into unit sphere in 4D
- Expand density in 4D spherical harmonics



$$\rho(\mathbf{r}) = \sum_{j=0,\frac{1}{2},\ldots}^{\infty} \sum_{m=-j}^{j} \sum_{m'=-j}^{j} u^j_{m,m'} U^j_{m,m'}(\theta, \phi, \theta_0)$$

Bispectrum



$$\rho_i(\mathbf{r}) = \delta(\mathbf{r}) + \sum_{r_{ii'}<R_c} f_c(r_{ii'}) w_{i'} \delta(\mathbf{r} - \mathbf{r}_{ii'})$$

$$B_{j_1,j_2,j} = \sum_{m_1,m_1'=-j_1}^{j_1} \sum_{m_2,m_2'=-j_2}^{j_2} \sum_{m,m'=-j}^{j} \left(u^j_{m,m'}\right)^* H^{jmm'}_{j_1 m_1 m_1' j_2 m_2 m_2'} u^{j_1}_{m_1,m_1'} u^{j_2}_{m_2,m_2'}$$

Spectral neighbor analysis potential (SNAP)[1]

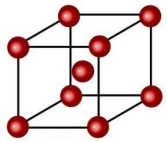
$$E_{\mathrm{SNAP}} = \beta_0 N + \boldsymbol{\beta} \cdot \sum_{i=1}^{N} \boldsymbol{B}^i$$

$$\mathbf{F}^j_{\mathrm{SNAP}} = -\boldsymbol{\beta} \cdot \sum_{i=1}^{N} \frac{\partial \boldsymbol{B}^i}{\partial \mathbf{r}_j}$$

$$\boldsymbol{\sigma}^j_{\mathrm{SNAP}} = -\boldsymbol{\beta} \cdot \sum_{j=1}^{N} \mathbf{r}_j \otimes \sum_{i=1}^{N} \frac{\partial \boldsymbol{B}^i}{\partial \mathbf{r}_j}$$

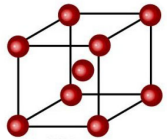
[1] Thompson et al. J. Comput. Phys. 2015, 285, 316–330

Extending the SNAP formalism



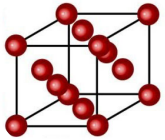
bcc
(Ta[1], W[2])

Sandia National Lab

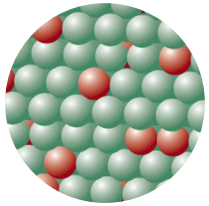


bcc
(Mo[3])

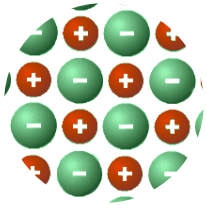
Materials Virtual Lab



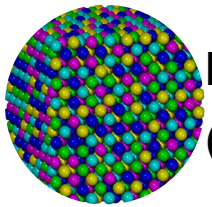
fcc
(Cu, Ni[4])



Binary Alloys
(Ni-Mo[4])



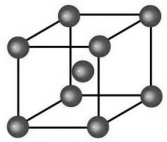
Ionic Crystals
(Li_3N[5])



MPE Alloys
(Ta-Nb-Mo-W)

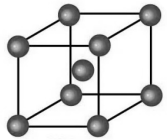
[1] Thompson et al. *J. Comput. Phys.* **2015**, *285*, 316–330.
[2] Wood et al. *arXiv preprint.* **2017**, *arXiv:1702.07042*.
[3] Chen et al. *Phys. Rev. Mater.* **2017**, *1*, 043603.
[4] Li et al. *Phys. Rev. B.* **2018**, 98, 094104.
[5] Deng et al. npj Comp Mat. **2019**, 5, 75

Extending the SNAP formalism



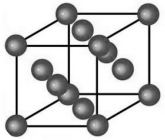
bcc
(Ta[1], W[2])

Sandia National Lab

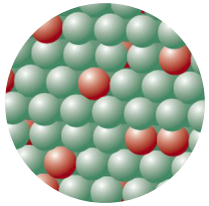


bcc
(Mo[3])

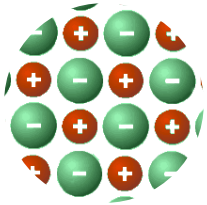
Materials Virtual Lab



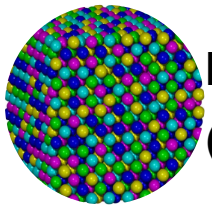
fcc
(Cu, Ni[4])



Binary Alloys
(Ni-Mo[4])



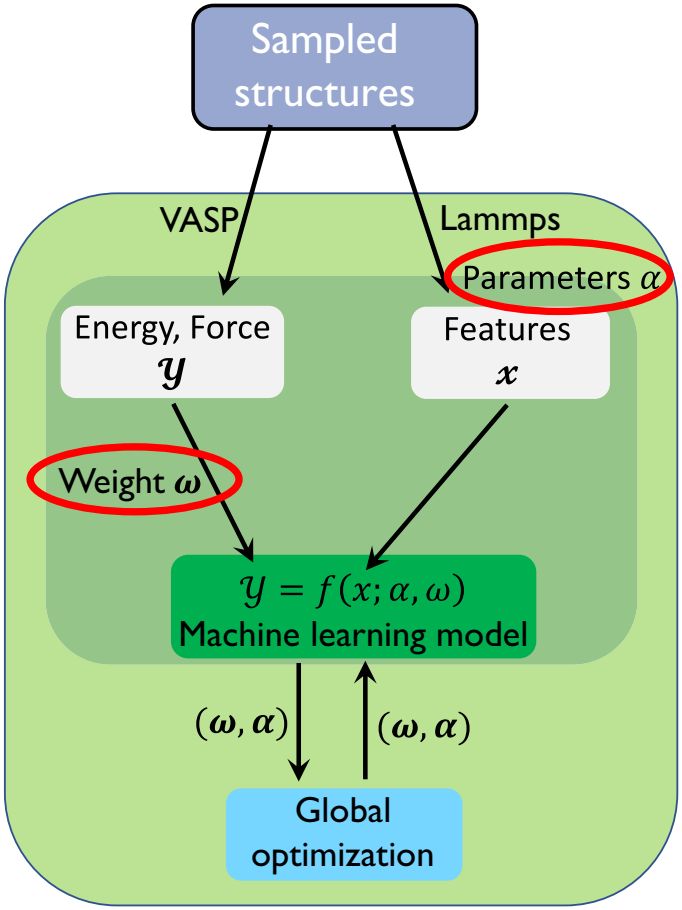
Ionic Crystals
(Li_3N[5])

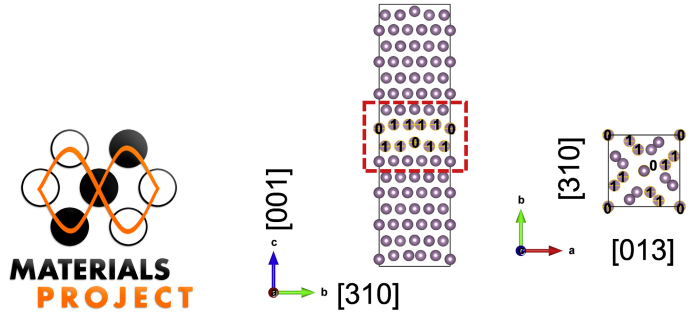


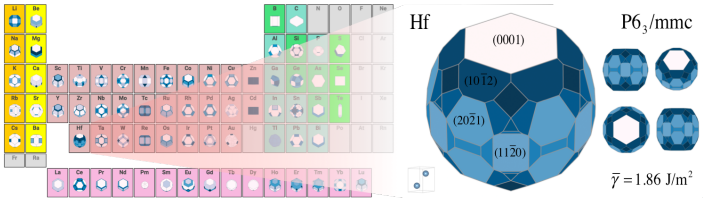
MPE Alloys
(Ta-Nb-Mo-W)

[1] Thompson et al. *J. Comput. Phys.* **2015**, *285*, 316–330.
[2] Wood et al. *arXiv preprint.* **2017**, *arXiv:1702.07042*.
[3] Chen et al. *Phys. Rev. Mater.* **2017**, *1*, 043603.
[4] Li et al. *Phys. Rev. B.* **2018**, 98, 094104.
[5] Deng et al. npj Comp Mat. **2019**, 5, 75

Optimization unit





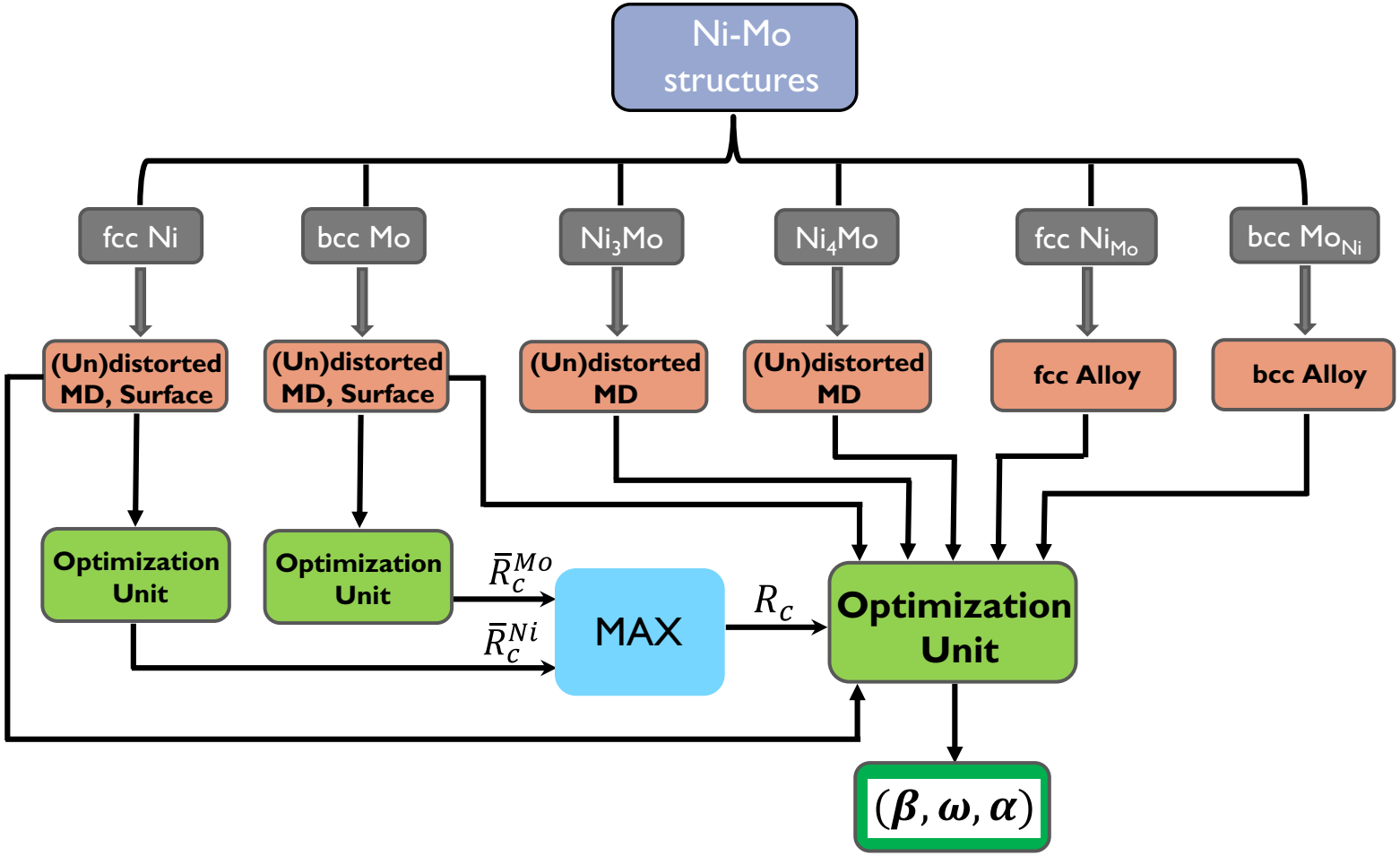


[1] Tran et al. Acta Mater. 2016, 117, 91–99 DOI: 10.1016/j.actamat.2016.07.005.
[2] Tran et al. Sci. Data 2016, 3, 160080 DOI: 10.1038/sdata.2016.80.
[3] Jain, A.; Ong, S. P.; et al. APL Mater. 2013, 1 (1), 11002 DOI: 10.1063/1.4812323.

$$\alpha = (R_{cut}^{atom}, w^{atom})$$

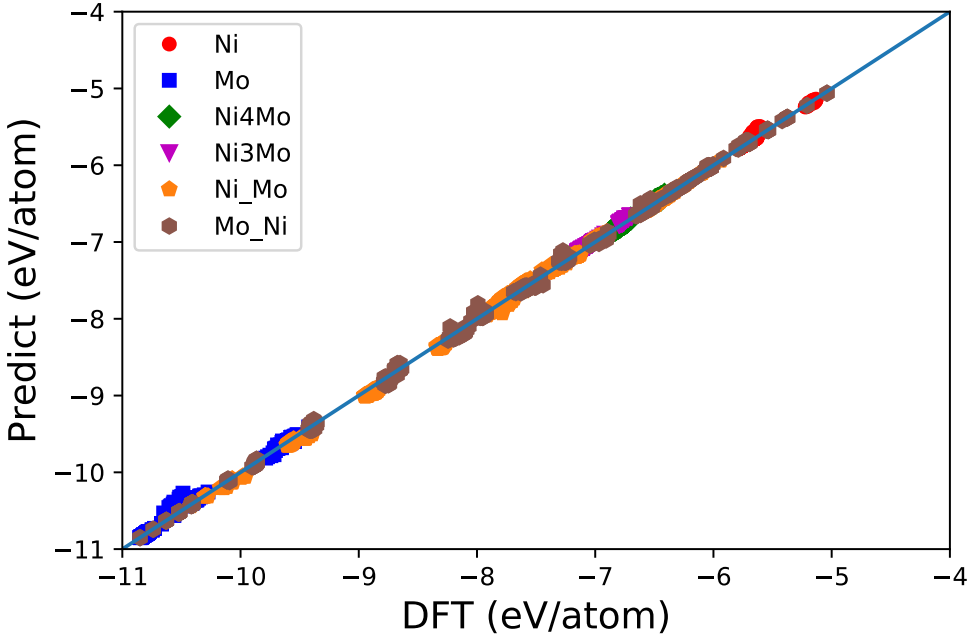
Li, X.; Hu, C.; Chen, C.; Deng, Z.; Luo, J.; Ong, S. P. *Phys. Rev. B.* **2018**, 98, 094104.

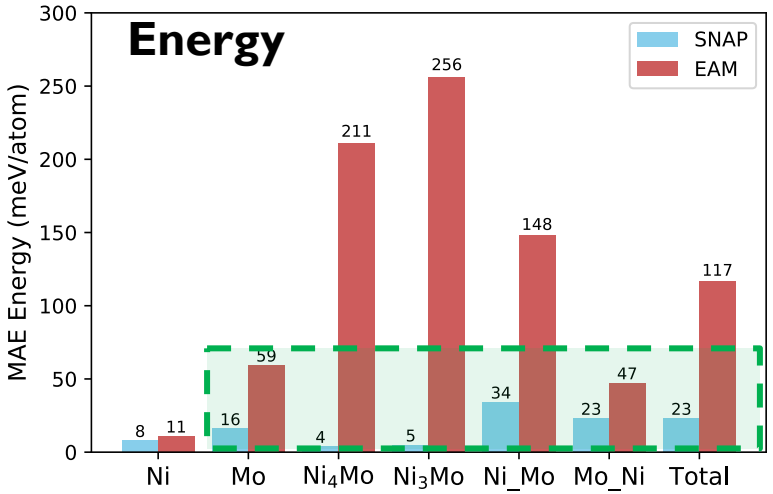
Workflow - alloy system

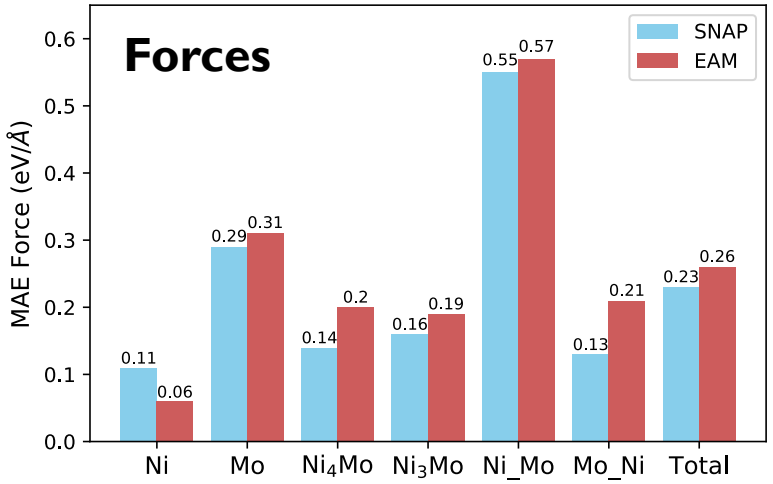


Li, X.; Hu, C.; Chen, C.; Deng, Z.; Luo, J.; Ong, S. P. *Phys. Rev. B.* **2018**, 98, 094104.

Ni-Mo alloy: SNAP model performance

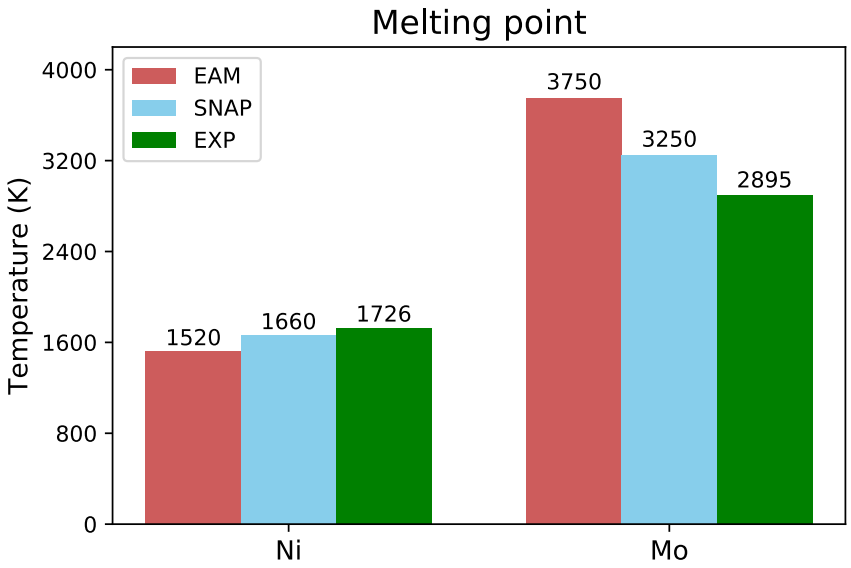




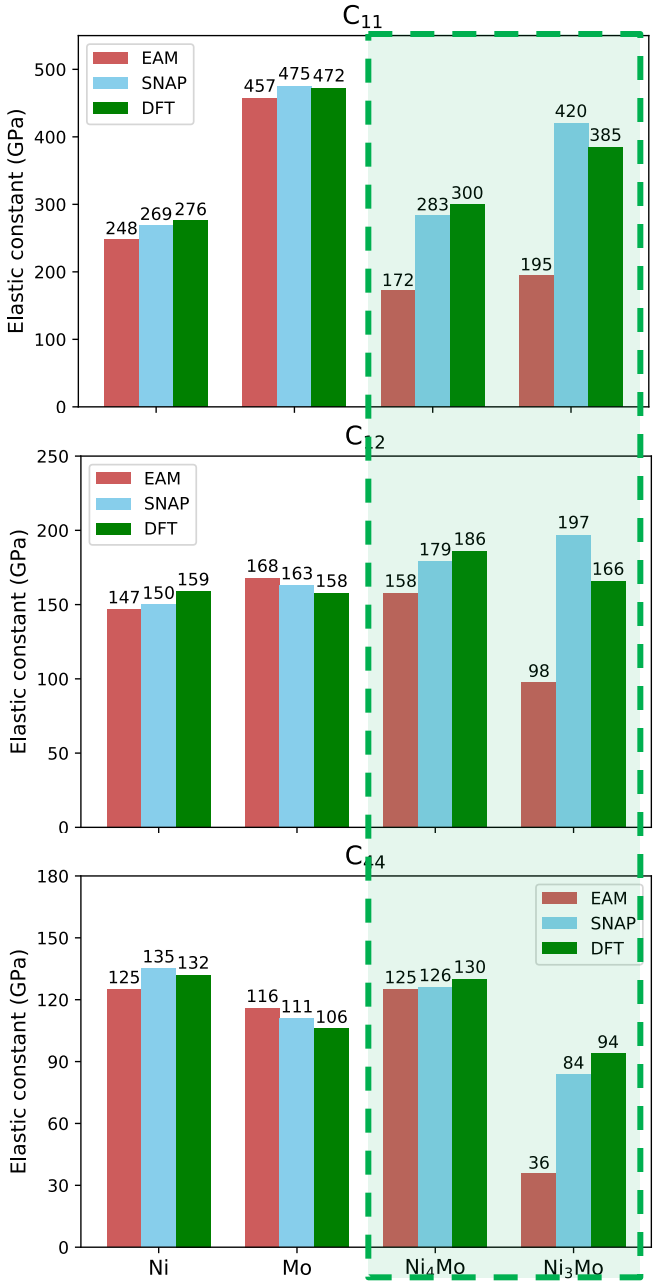


- SNAP significantly outperforms in binary and bcc Mo for energy.

Ni-Mo alloy: property matching

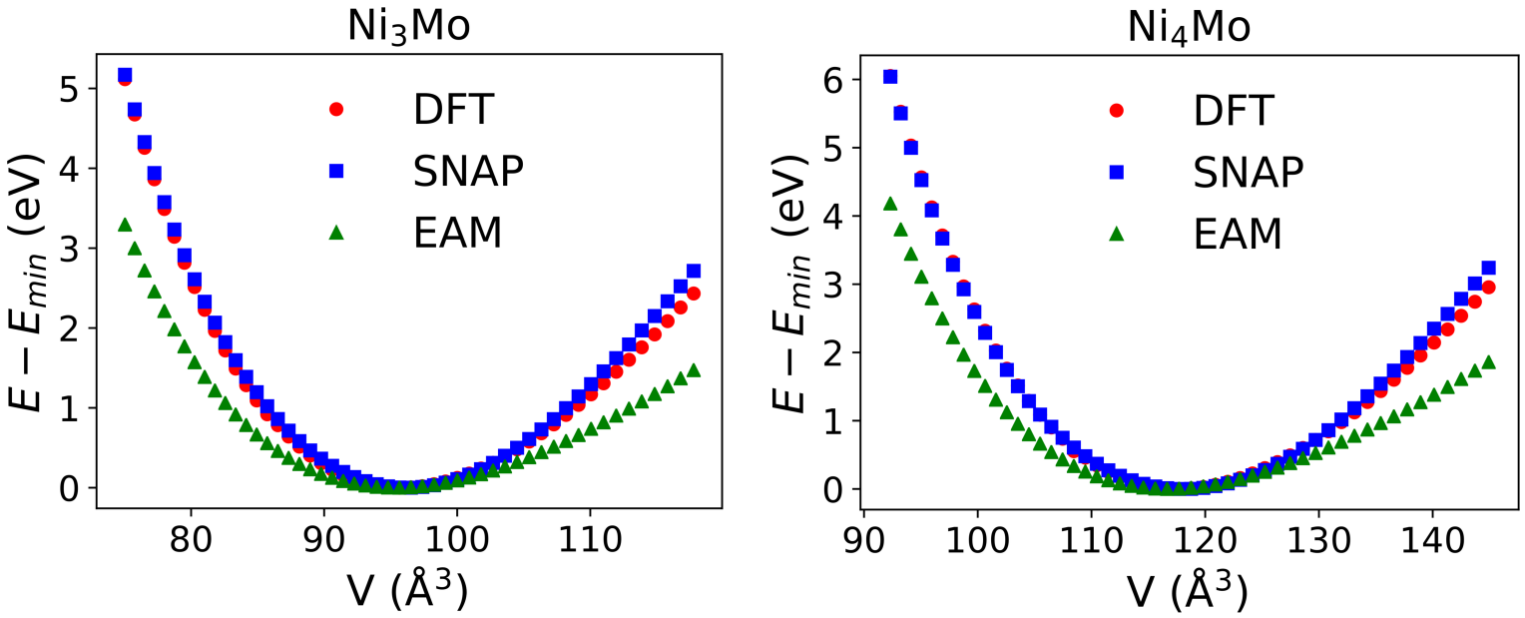


Elastic constants



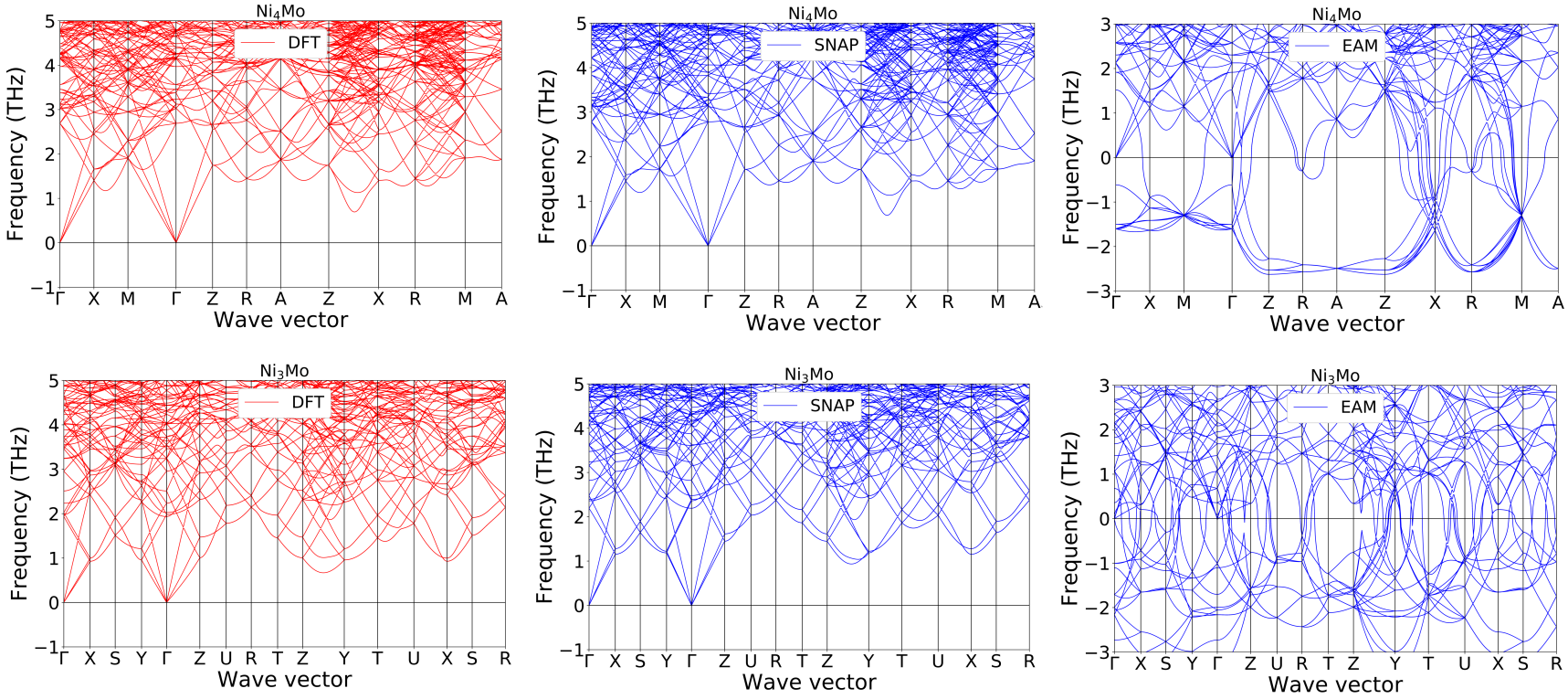
- EAM fails in binary bulk systems for elastic constants.

Ni-Mo alloy: Equation of State



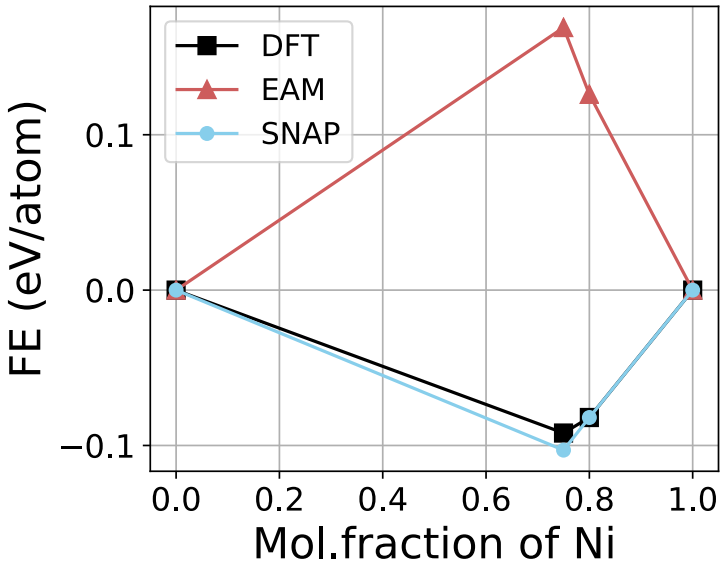
The EAM potential completely fails in the equation of state prediction for binary compounds.

Ni-Mo alloy: Phone Spectra



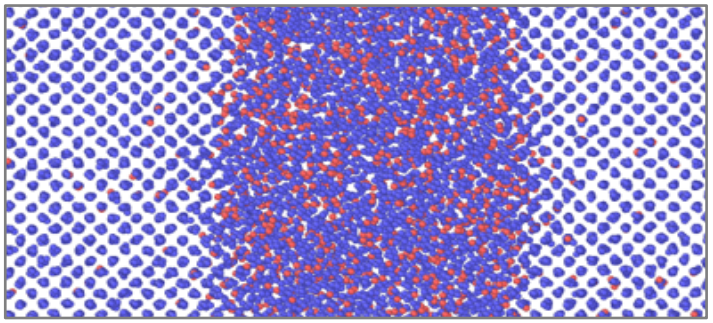
The EAM potential completely fails in the phonon spectra prediction for binary compounds.

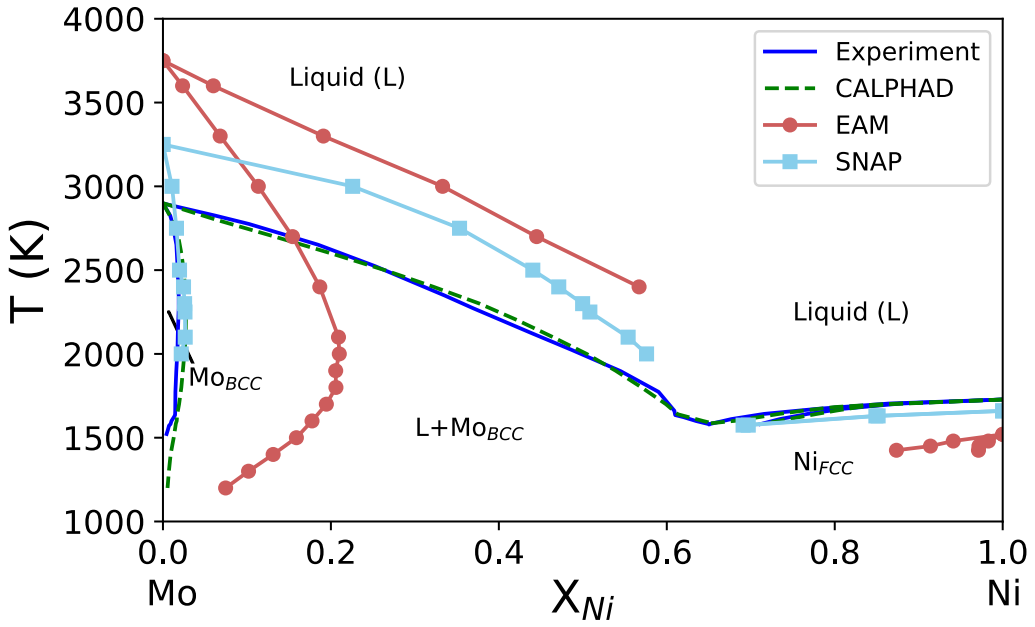
Ni-Mo phase diagram



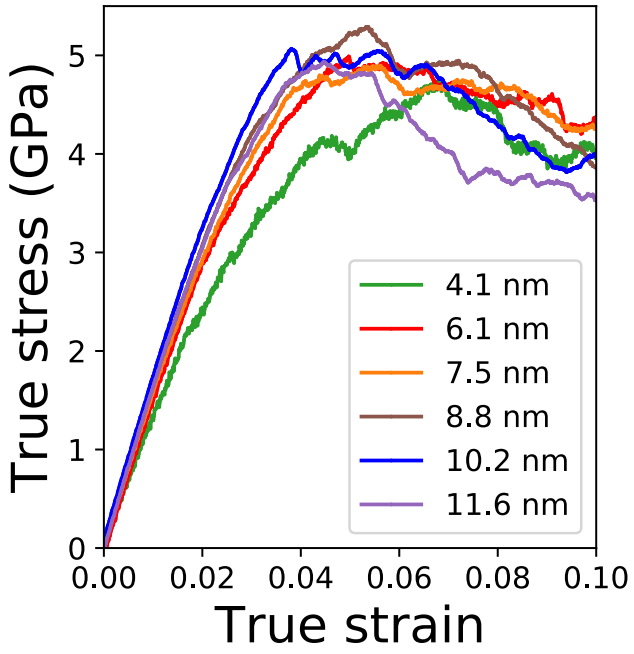
EAM completely fails to reproduce Ni-Mo phase diagram

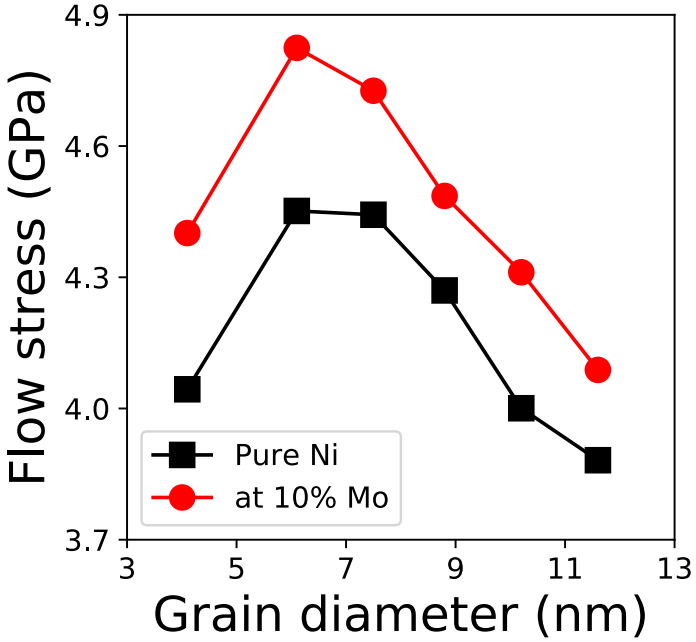
Solid-liquid equilibrium



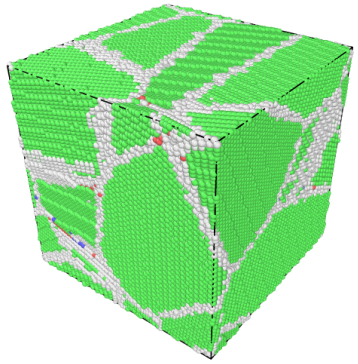


Application: Investigating Hall-Petch strengthening in Ni-Mo



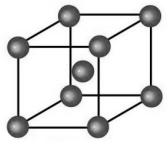


- ~20,000 to ~455,000 atoms
- Uniaxially strained with a strain rate of 5×10^8 s^{-1}
- SNAP reproduces the (inverse) Hall-Petch relationship, consistent with experiment[1].



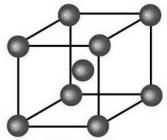
[1] Hu et al. *Nature*, **2017**, *355*, 1292

Extending the SNAP formalism



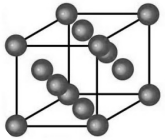
bcc
(Ta[1], W[2])

Sandia National Lab

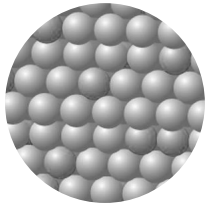


bcc
(Mo[3])

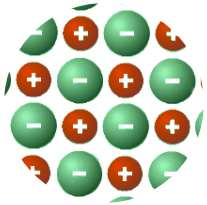
Materials Virtual Lab



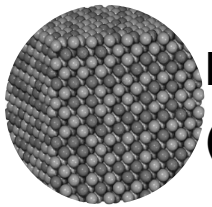
fcc
(Cu, Ni[4])



Binary Alloys
(Ni-Mo[4])



Ionic Crystals
(Li_3N[5])



MPE Alloys
(Ta-Nb-Mo-W)

[1] Thompson et al. *J. Comput. Phys.* **2015**, *285*, 316–330.
[2] Wood et al. *arXiv preprint.* **2017**, *arXiv:1702.07042*.
[3] Chen et al. *Phys. Rev. Mater.* **2017**, *1*, 043603.
[4] Li et al. *Phys. Rev. B.* **2018**, 98, 094104.
[5] Deng et al. npj Comp Mat. **2019**, 5, 75

Electrostatic SNAP (eSNAP) for ionic systems

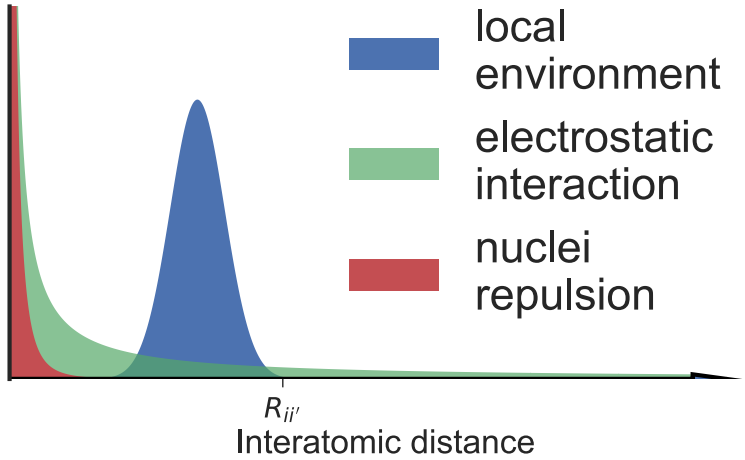
- Local environment

$$E_{\mathrm{SNAP}} = \beta_0 N + \boldsymbol{\beta} \sum_{i=1}^{N} B_i$$

- Electrostatic interaction

$$E_{el} = \frac{q_j q_k}{r}$$

  - Formal charge
  - Ewald summation



- eSNAP

$$E_p = \gamma E_{el} + E_{\mathrm{SNAP}}$$

$$\mathbf{F}_j = -\nabla_j E_p = -\gamma \nabla_j E_{el} - F_{j,\mathrm{SNAP}}$$

γ - screening parameter

- Nuclear repulsion - Ziegler-Biersack-Littmark (ZBL)

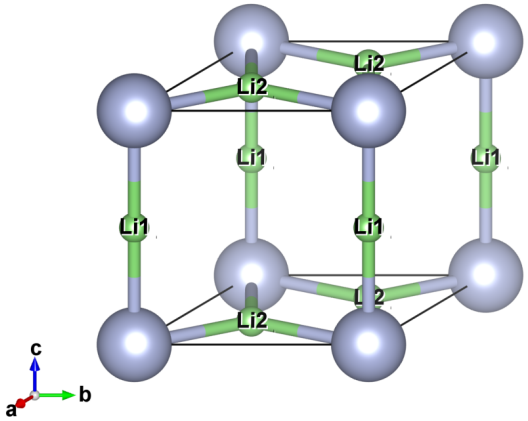
Deng et al. npj Comp Mat. **2019**, 5, 75

Generation of training data for Li_3N

Initial configuration pool

- ❑Unit cells with different lattice constant *a* and *c*
- ❑Unit cells with lattice distortions under different strains
- ❑Snapshots (3x3x3 supercells) taken from AIMD simulated below $1.2T_m$ (400 ~ 1200 K)

Static DFT to calculate reference energy and force



α-Li_3N

	Distorted unit cells	AIMD snapshots
N_{atoms}	4	108
$N_{configs}$	109	1000
w_E	10^3	1
w_F	0	10^{-3}

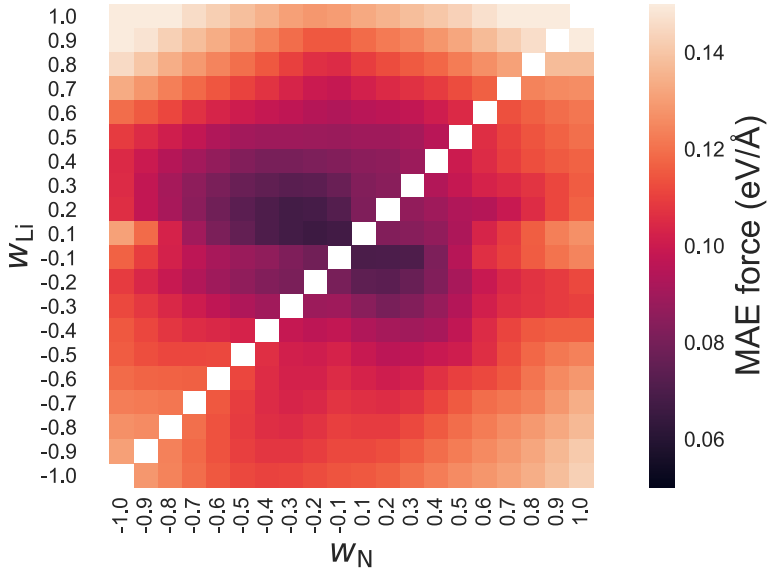
Deng et al. npj Comp Mat. **2019**, 5, 75

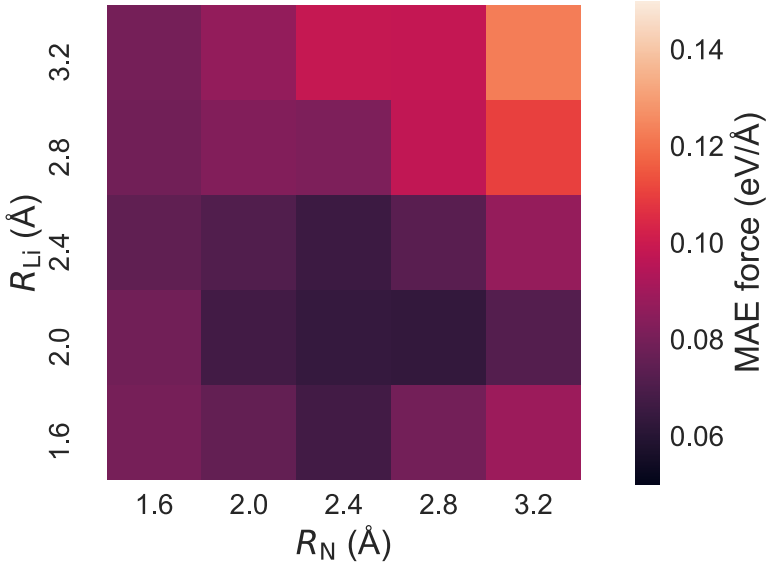
Grid search for hyperparameters of SNAP

$$\rho_i(\mathbf{r}) = \delta(\mathbf{r}) + \sum f_c(r_{ii'}) w_{i'} \delta(\mathbf{r} - \mathbf{r_{ii'}})$$

$$r_{ii'} < R_{ii'}$$

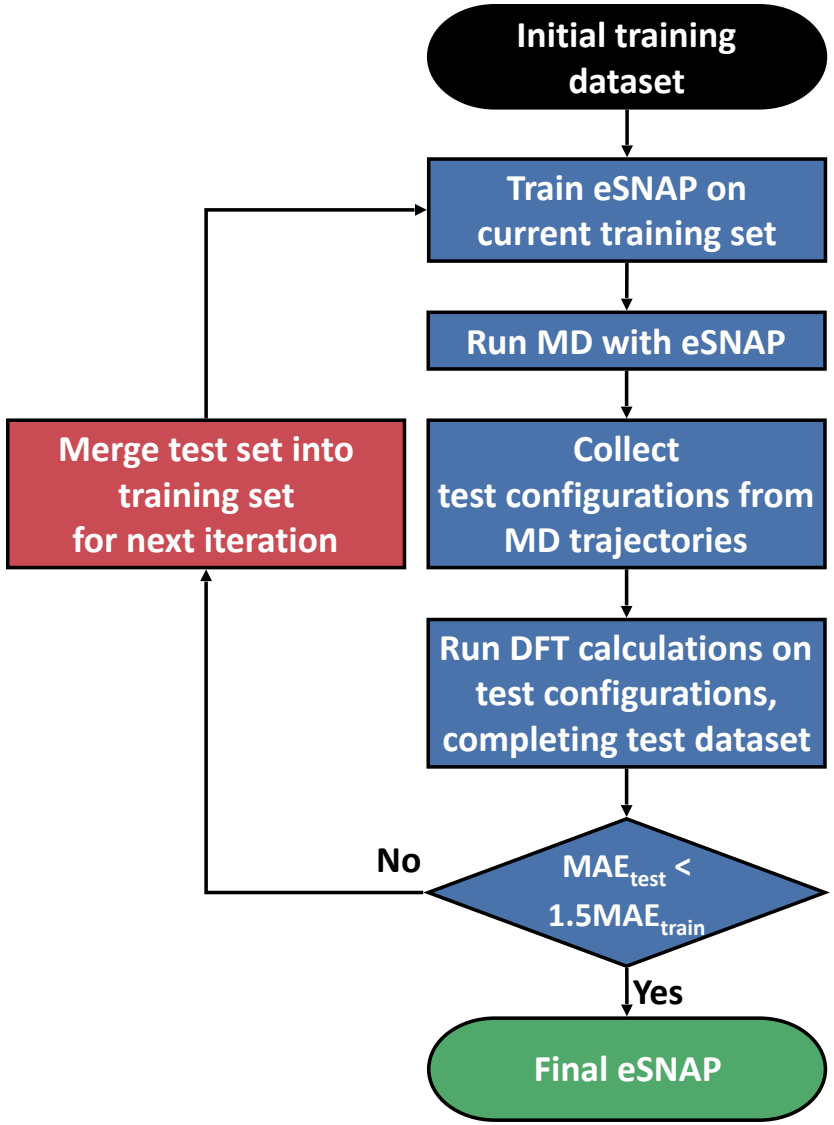
	w	*R* (Å)
Li	0.1	2.0
N	-0.1	2.8





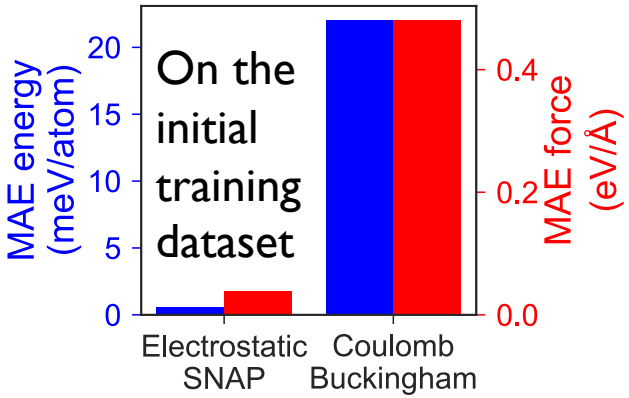
Training/test iteration

- Systematically improves the predictions on energy and force for MD simulations
- Leverages the benefit gained by adding more training instances and the associated costs for performing more DFT calculations

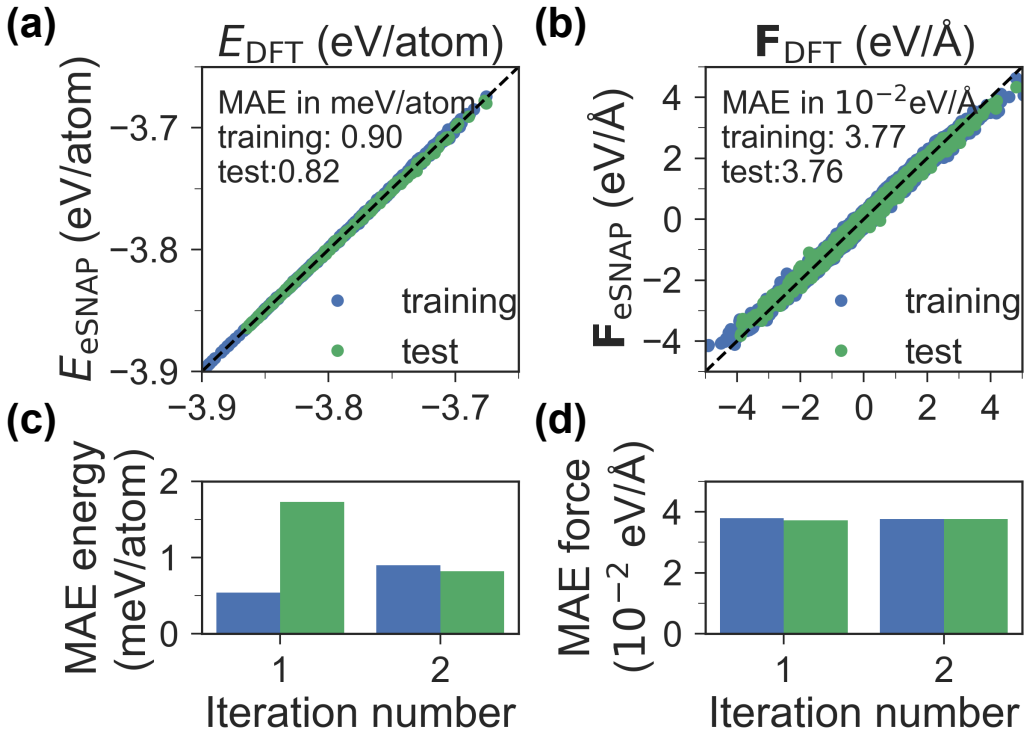


Energy and force prediction from eSNAP

- The iteration terminates with the training data expanded for once.
- eSNAP model has successfully captured the fundamental relationship between atomic environment and energy/force.



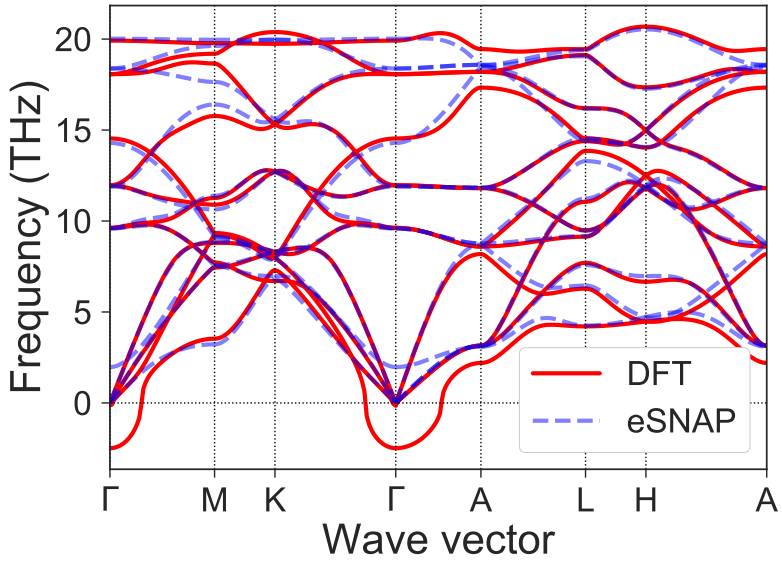
Coulomb-Buckingham:
Walker et al. *Philos. Mag. A* **1981**, *43* (2), 265–272.

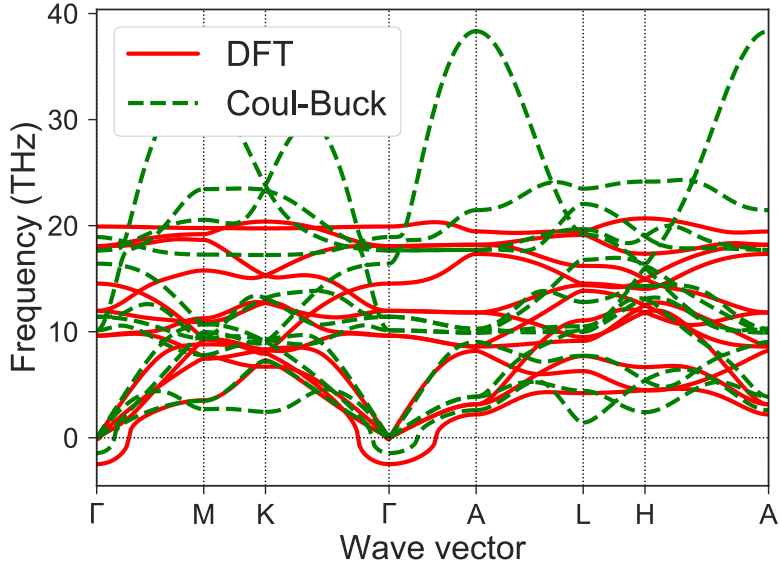


Structural property calculations using eSNAP

	DFT	eSNAP	Coul-Buck	Exp.
a (Å)	3.641	3.641	3.528	3.648
c (Å)	3.874	3.872	3.628	3.875
c_{11} (GPa)	123	116	165	114
c_{33} (GPa)	137	144	193	118
c_{44} (GPa)	17	17	19	17
c_{66} (GPa)	48	39	53	38
$E_{f,V2}$ (eV)	0.60	0.64	0.44	
$E_{f,V1}$ (eV)	0.51	0.63	0.46	

Coul-Buck: Walker et al. *Philos. Mag. A* **1981**, *43* (2), 265–272.
Exp:
Rabenau et al. *J. Less-Common Met.* **1976**, *50* (1), 155–159.
Kress et al. *Phys. Rev. B* **1980**, *22* (10), 4620–4625.

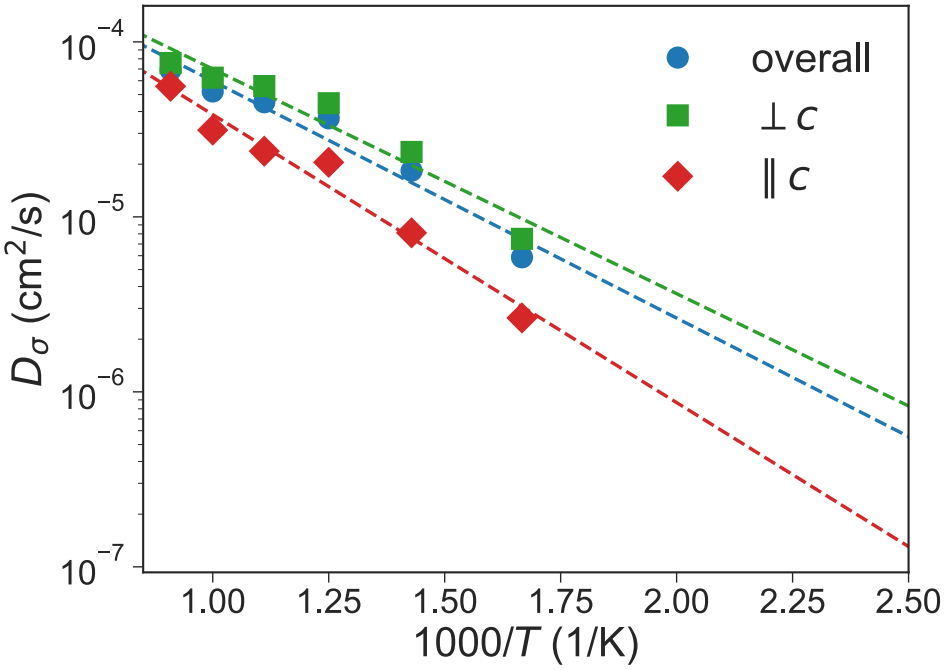




Deng et al. npj Comp Mat. **2019**, 5, 75

Bulk diffusion

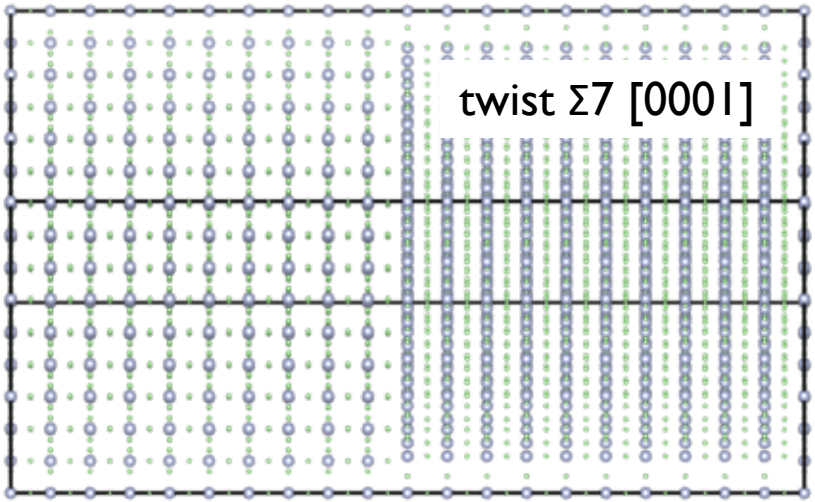
- Simulation box contains 4000 atoms (10x10x10 supercell)
- MD simulations performed from 600 to 1000 K in an *NVT* ensemble for 1ns

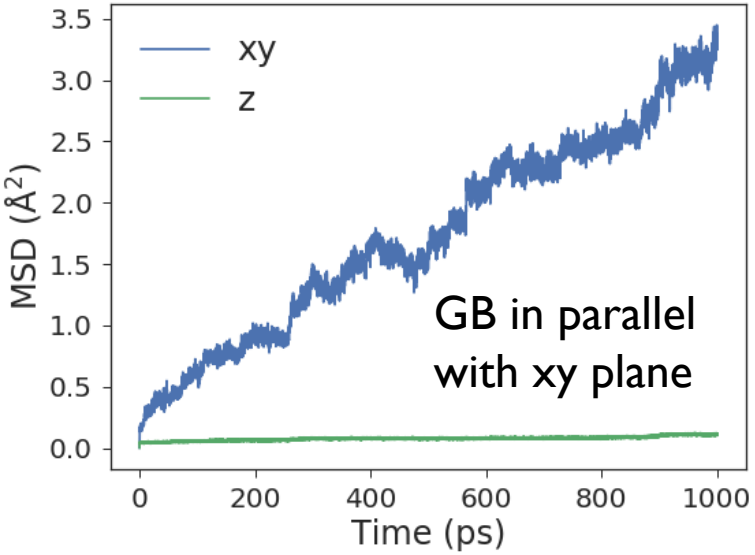


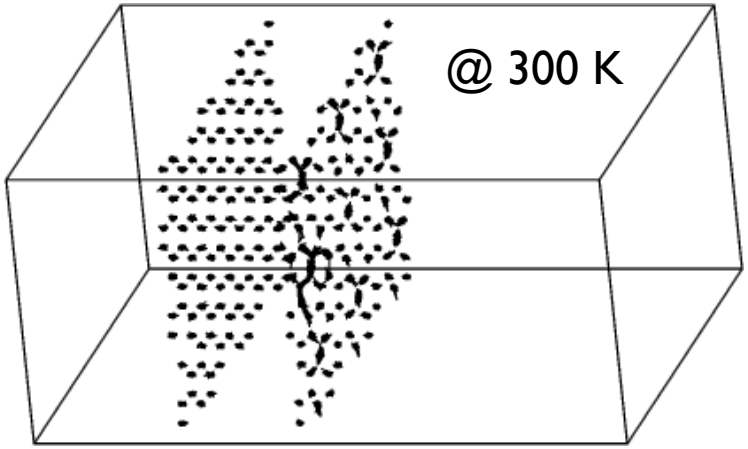
	E_a (eV)			σ_{RT} (mS/cm)		
	ab	c	total	ab	c	total
eSNAP	0.255	0.327	0.269	29.6	2.32	17.3
Exp.	0.290	0.490		1.20	0.01	

Exp.:Alpen et al. *Appl. Phys. Lett.* **1977**, *30* (12), 621–623.

Grain boundary diffusion







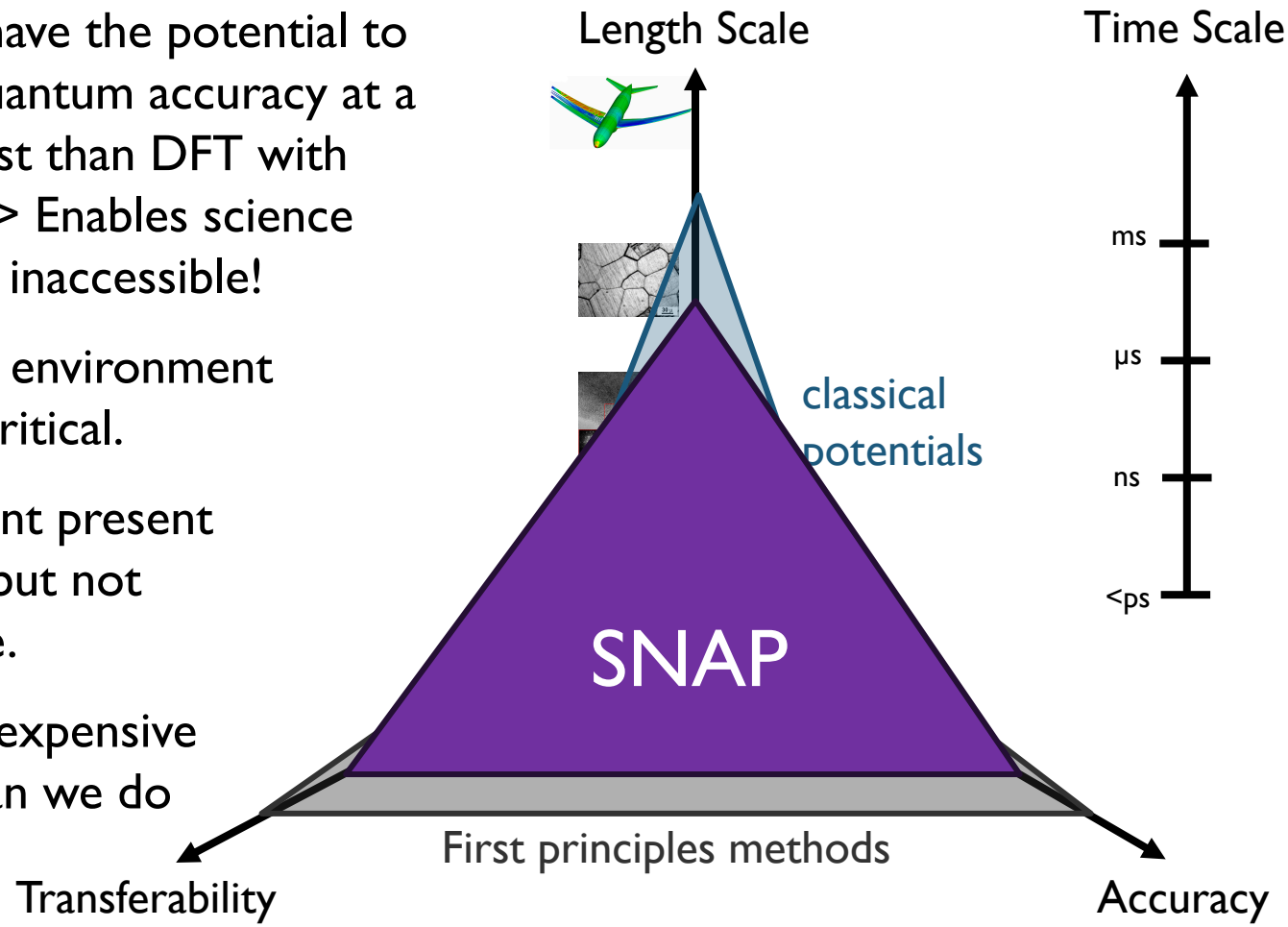
The presence of GB facilitates Li diffusion

*D** at 300 K

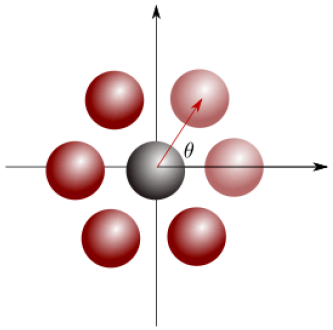
- GB: 7.09×10^{-8} cm^2/s
- Bulk (extrapolated): 2.24×10^{-8} cm^2/s

Conclusions

- ✓ ML potentials have the potential to achieve near-quantum accuracy at a much lower cost than DFT with linear scaling => Enables science that is hitherto inaccessible!
- ✓ Choice of local environment description is critical.
- ✓ Multi-component present complications, but not insurmountable.
- X 10-100x more expensive than MEAM (can we do better?)



Machine learning the potential energy surface



$$G_i^{\text{atom,rad}} = \sum_{j \neq i}^{N_{\text{atom}}} e^{-\eta(R_{ij}-R_s)^2} \cdot f_c(R_{ij}),$$

$$G_i^{\text{atom,ang}} = 2^{1-\zeta} \sum_{j,k \neq i}^{N_{\text{atom}}} (1 + \lambda \cos\theta_{ijk})^{\zeta} \cdot e^{-\eta'(R_{ij}^2+R_{ik}^2+R_{jk}^2)} \cdot f_c(R_{ij}) \cdot f_c(R_{ik}) \cdot f_c(R_{jk}),$$

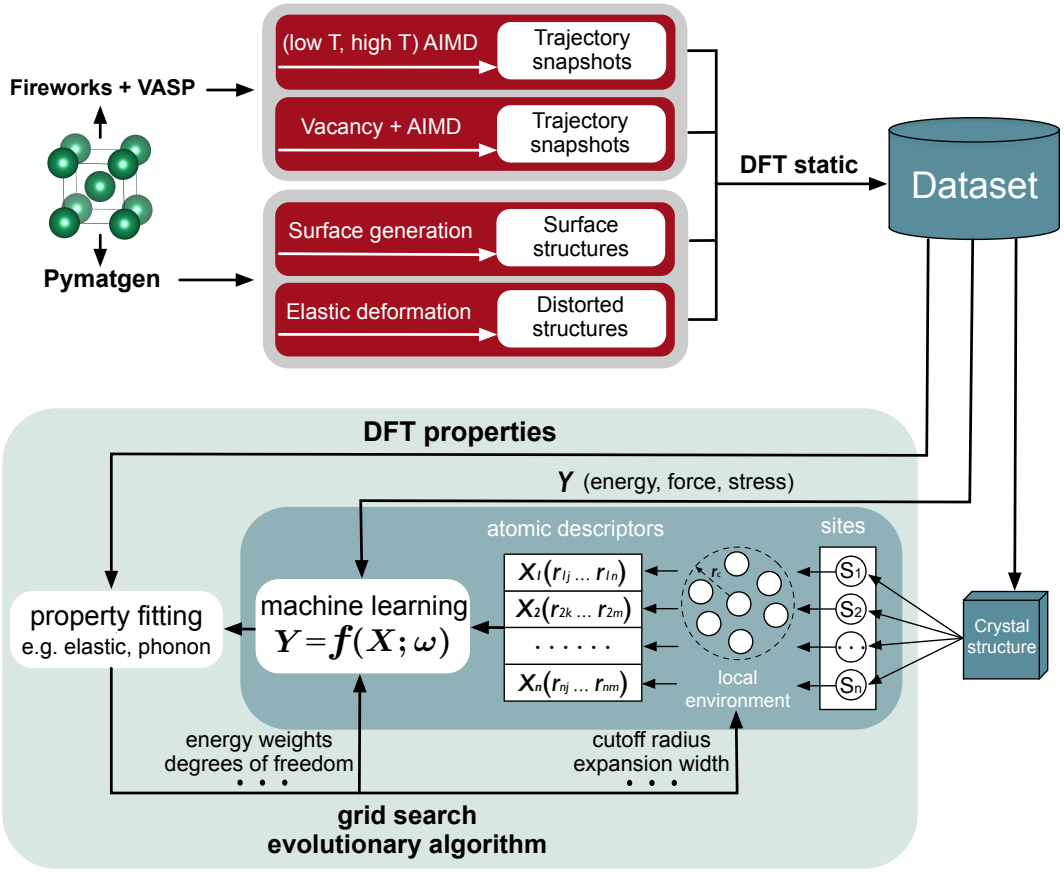
ACSF/MTP encodes atomic distances and angles.

$$\rho_i(\boldsymbol{R}) = \sum_j f_c(R_{ij}) \cdot \exp\left(-\frac{|\boldsymbol{R} - \boldsymbol{R}_{ij}|^2}{2\sigma_{\text{atom}}^2}\right)$$

SOAP/bispectrum encodes atomic neighbor density.

Local environment descriptors	**ML approach**	**Interatomic Potential**
Atom-centered symmetry functions (ACSF)	Neural networks	Behler-Parinello Neural Network Potential (NNP)
Moment tensors	Kernel regression	Moment tensor potential (MTP)
Smooth overlap of atomic positions (SOAP)	Polynomial / Linear regression	Gaussian Approximation Potential (GAP)
SO4 bispectrum		Spectral Neighbor Analysis Potential (SNAP)

Standardized workflow for ML-IAP construction and evaluation

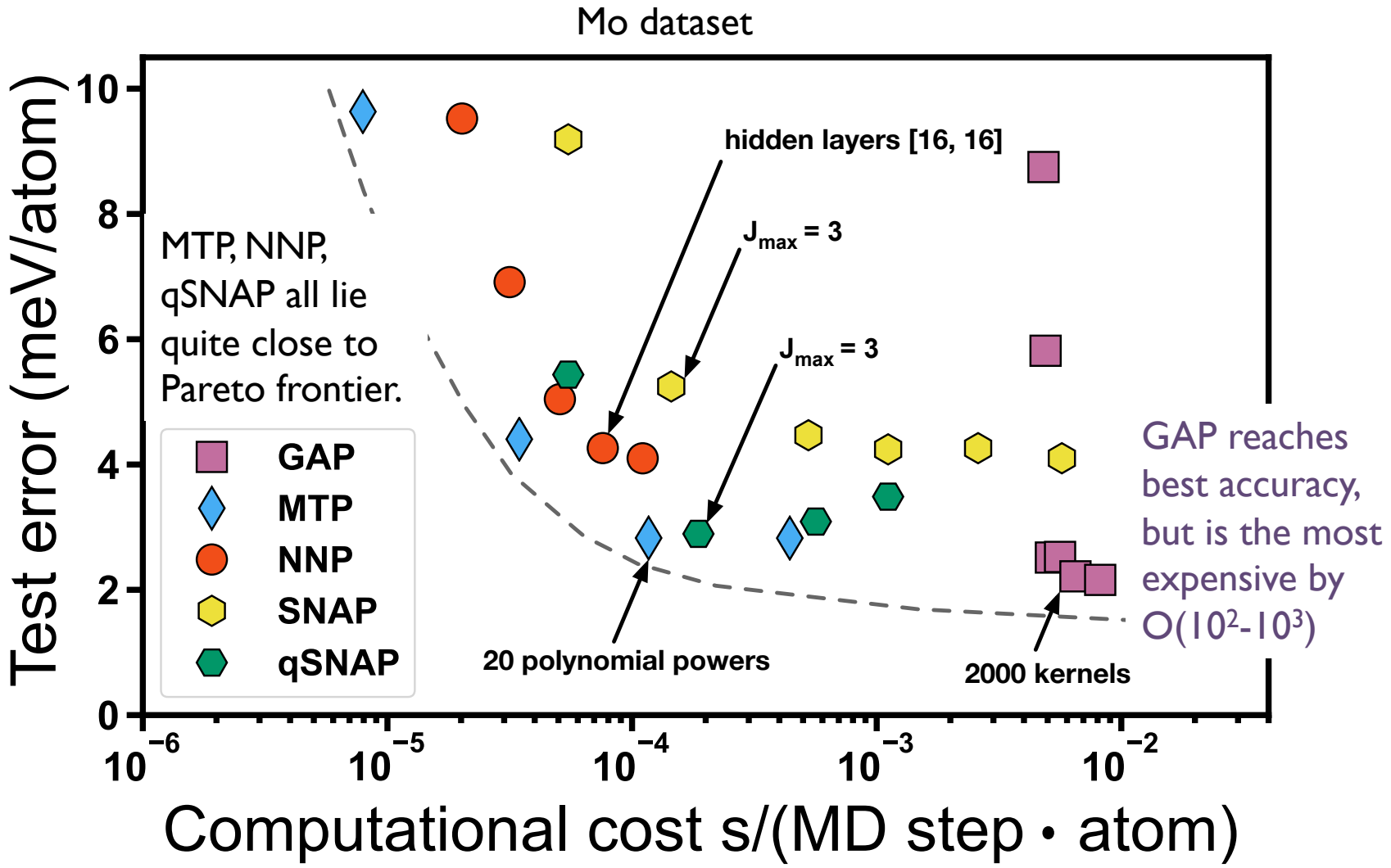


Test systems:

- Fcc Ni
- Fcc Cu
- Bcc Li
- Bcc Mo
- Diamond Ge
- Diamond Si

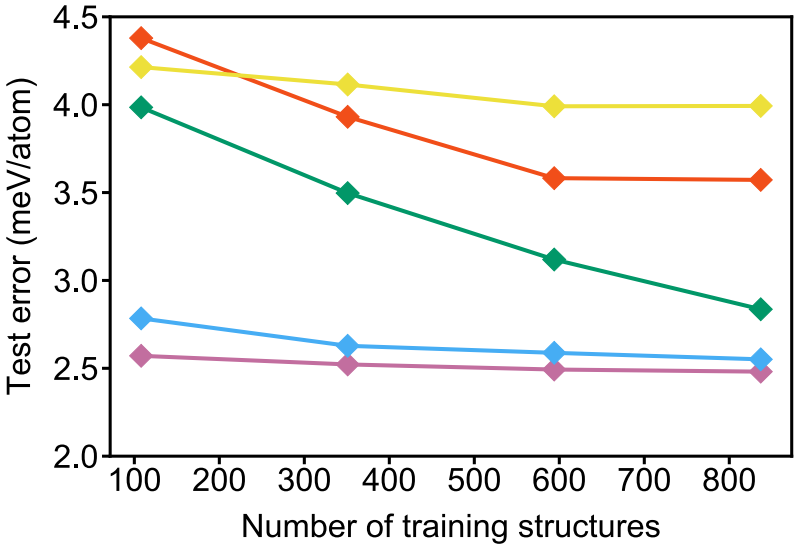
Available open source on Github: https://github.com/materialsvirtuallab/mlearn

ML-IAP: Accuracy vs Cost

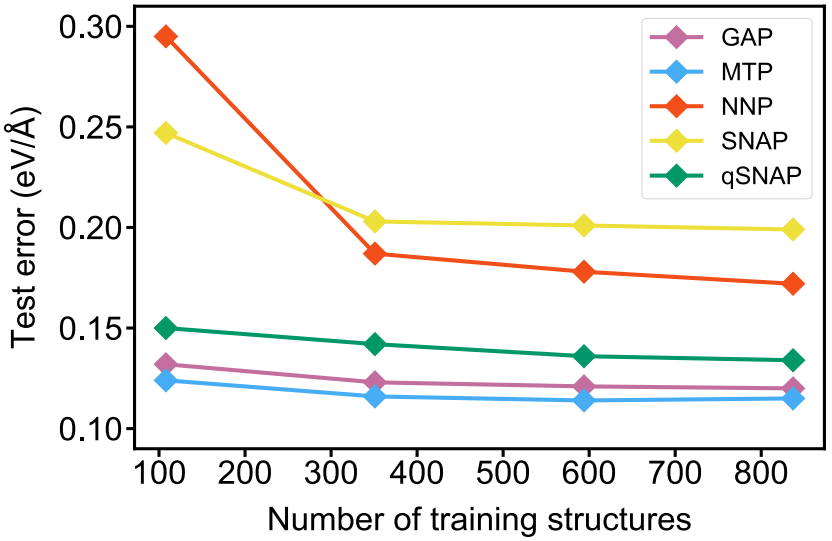


ML-IAP: Training Data Requirements

Energies

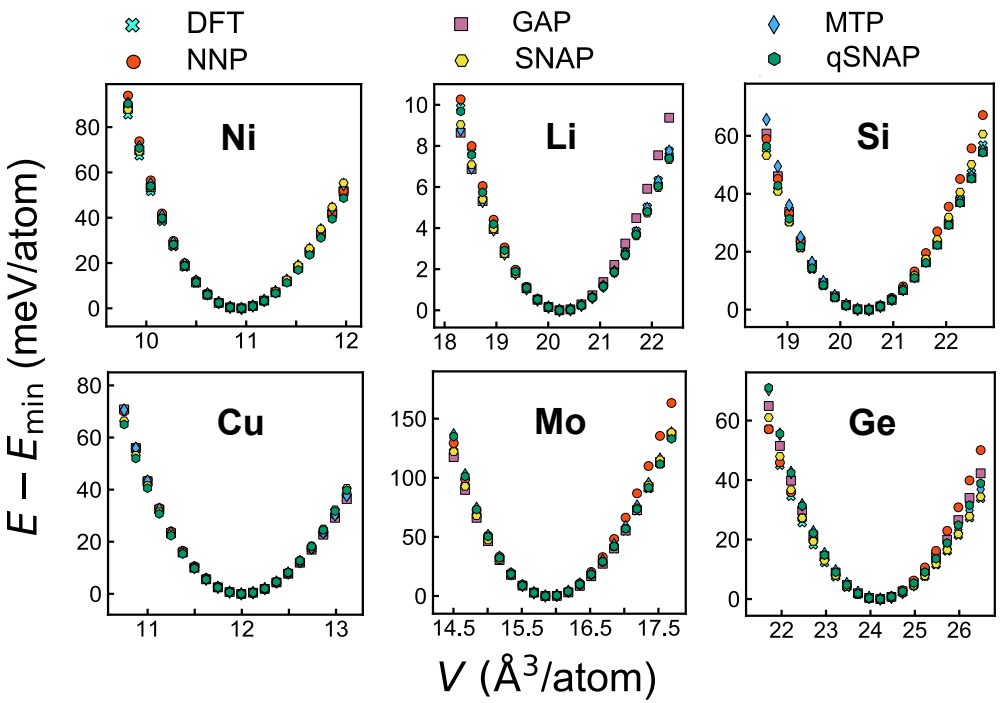


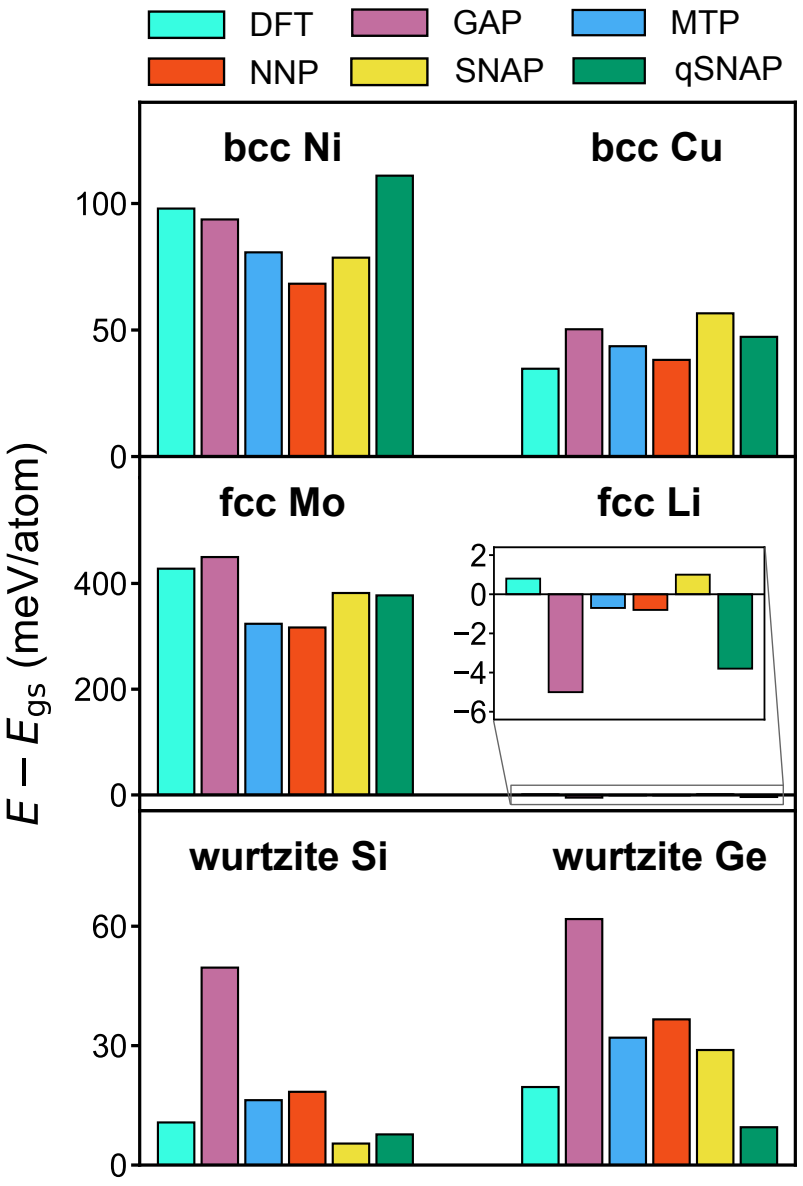
Forces



- Data quality is more important than data quantity - ~O(100) structures sufficient to converge
- NNP and qSNAP require much more training data than other models.

ML-IAP: Extrapolability



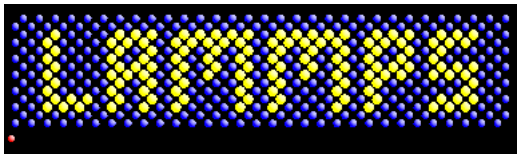


- The greater the ML complexity (e.g., NNP and GAP), the greater the issues with extrapolation.
- Linear SNAP performs surprisingly well on EOS and polymorph energy differences.

Acknowledgements

- **Shyue Ping Ong**
- Zhi Deng
- Yunxing Zuo
- Chi Chen

Where to get the potentials?



https://github.com/materialsvirtuallab/snap

- Xiang-Guo Li. *et al.* *Phys. Rev. B*, 98, 094104 (**2018**)
- Zhi Deng. *et al.* npj Comp Mat, 5, 75 (2019)
- Yunxing Zuo, *et al.* "A Performance and Cost Assessment of Machine Learning Interatomic Potentials", arXiv:1906.08888 (2019)

Office of Naval Research
USN ONR Science & Technology

XSEDE
Extreme Science and Engineering Discovery Environment

NERSC

materials virtual·lab