Machine learning *ab initio* calculations for Materials Science

Shyue Ping Ong, Chi Chen, Xiangguo Li, Zhi Deng

May 21, 2019

ABIDEV 2019

UC San Diego
Jacobs School of Engineering

materials virtual·lab

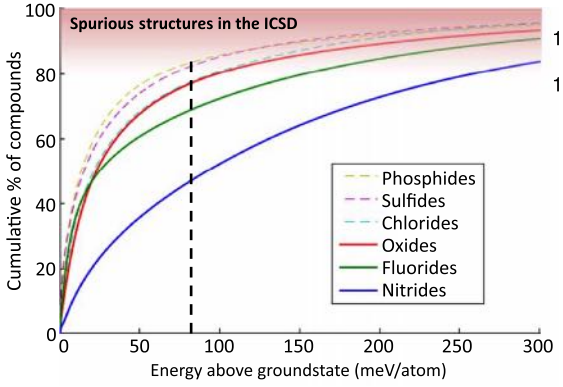
Electronic structure calculations are today *reliable* and *reasonably accurate*.

		average <Δ>	AE						
			Elk	exciting	FHI-aims/tier2	FLEUR	FPLO/T+F+s	RSPt	WIEN2k/acc
AE	Elk	0.6		0.3	0.3	0.6	1.0	0.9	0.3
	exciting	0.5	0.3		0.1	0.5	0.9	0.8	0.2
	FHI-aims/tier2	0.5	0.3	0.1		0.5	0.9	0.8	0.2
	FLEUR	0.6	0.6	0.5	0.5		0.8	0.6	0.4
	FPLO/T+F+s	0.9	1.0	0.9	0.9	0.8		0.9	0.9
	RSPt	0.8	0.9	0.8	0.8	0.6	0.9		0.8
	WIEN2k/acc	0.5	0.3	0.2	0.2	0.4	0.9	0.8	
PAW	GBRV12/ABINIT	0.9	0.9	0.8	0.8	0.9	1.3	1.1	0.8
	GPAW09/ABINIT	1.4	1.5	1.5	1.5	1.3	1.7	1.5	1.3
	GPAW09/GPAW	1.6	1.8	1.8	1.8	1.6	1.8	1.7	1.5
	JTH02/ABINIT	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.5
	PSlib100/QE	0.9	0.9	0.8	0.8	0.8	1.3	1.1	0.8
	VASPGW2015/VASP	0.6	0.4	0.4	0.4	0.6	1.0	0.8	0.3
USPP	GBRV14/CASTEP	1.1	1.2	1.2	1.2	1.1	1.4	1.3	1.0
	GBRV14/QE	1.1	1.2	1.2	1.2	1.1	1.4	1.3	1.0
	OTFG9/CASTEP	0.7	0.4	0.5	0.5	0.7	1.0	1.0	0.5
	SSSP/QE	0.5	0.4	0.3	0.3	0.5	0.9	0.8	0.3
	Vdb2/DACAPO	6.3	6.3	6.3	6.3	6.3	6.4	6.5	6.2
NCPP	FHI98pp/ABINIT	13.3	13.5	13.4	13.4	13.2	13.0	13.2	13.4
	HGH/ABINIT	2.2	2.2	2.2	2.2	2.0	2.3	2.2	2.1
	HGH-NLCC/BigDFT	1.1	1.1	1.1	1.1	1.0	1.2	1.1	1.0
	MBK2013/OpenMX	2.0	2.1	2.1	2.1	1.9	1.8	1.8	2.0
	ONCVPSP(PD0.1)/ABINIT	0.7	0.7	0.7	0.7	0.6	1.0	0.8	0.6
	ONCVPSP(SG15)1/QE	1.4	1.4	1.3	1.3	1.3	1.6	1.5	1.3
	ONCVPSP(SG15)2/CASTEP	1.4	1.4	1.4	1.4	1.3	1.6	1.5	1.4

Modern electronic structure codes give relatively consistent equations of state.

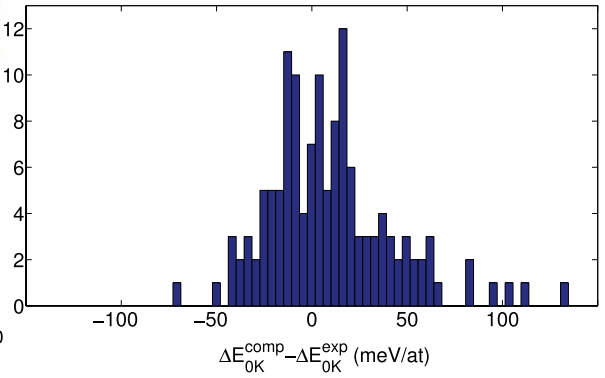
Lejaeghere et al. Science, 2016, 351 (6280), aad3000.

Phase stability



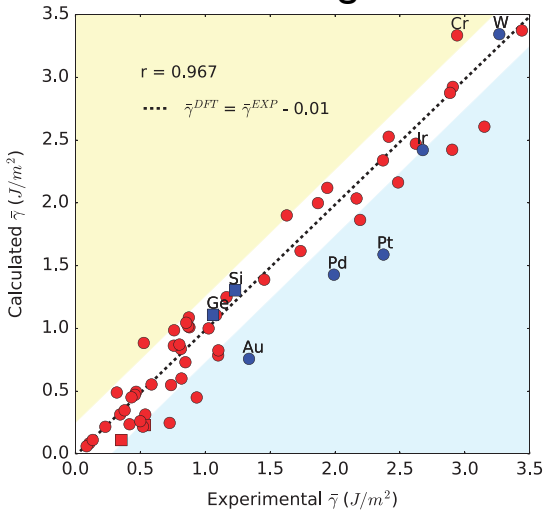
Sun, et al. Sci. Adv. 2016, 2 (11), e1600225.

Formation energies



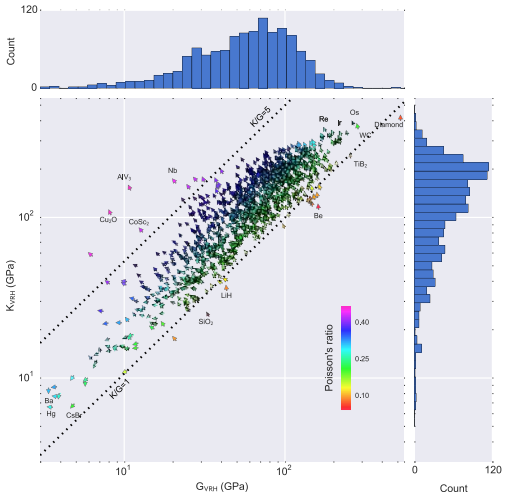
Hautier et al. Phys. Rev. B 2012, 85, 155208.

Surface energies



Tran, et al. Sci. Data 2016, 3, 160080.

Elastic constants



de Jong et al. Sci. Data 2015, 2, 150009.

With reliable electronic structure codes + great computing power, we have big databases….

Materials Project[1]

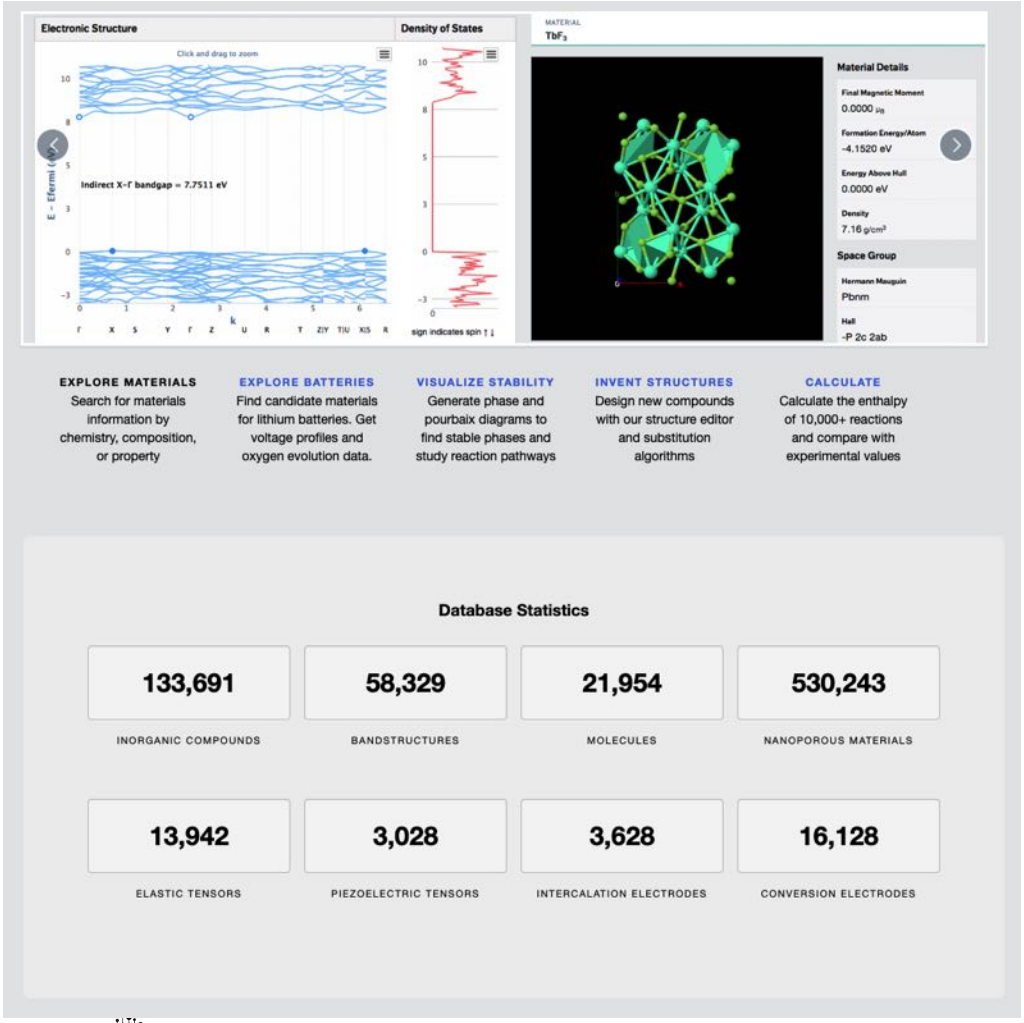
https://www.materialsproject.org

Powered by:

pymatgen[2]

Custodian[3]

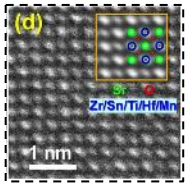
FireWorks



[1] Jain et al. APL Mater. 2013, 1 (1), 11002.
[2] Ong et al. Comput. Mater. Sci. 2013, 68, 314–319.
[3] Jain et al. Concurr. Comput. Pract. Exp. 2015, 27 (17), 5037–5059.

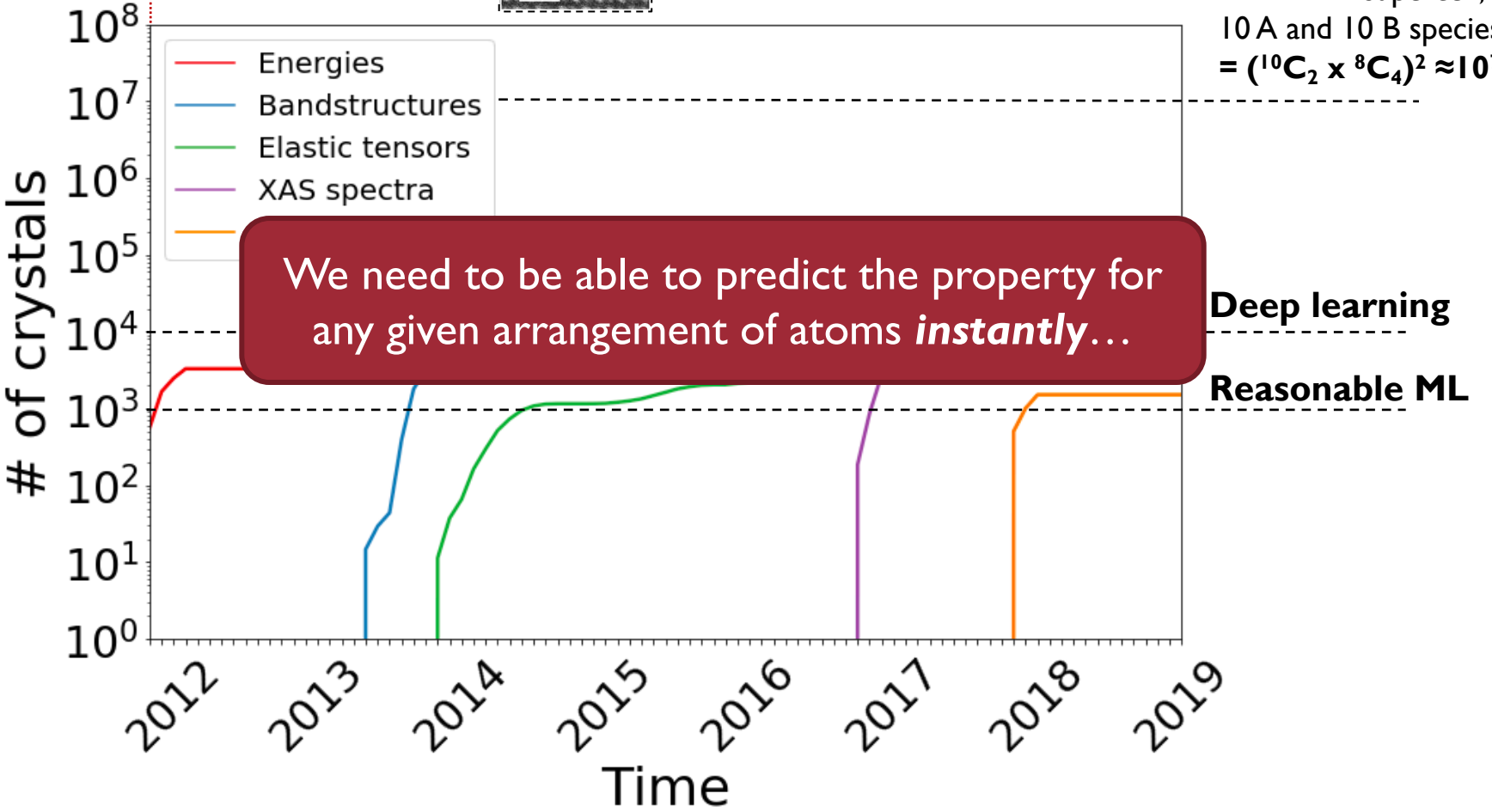
Why machine learning? (Part 1)

Materials design is *combinatorial*



Jiang et al. A New Class of High-Entropy Perovskite Oxides. *Scripta Materialia* **2018**, *142*, 116–120.

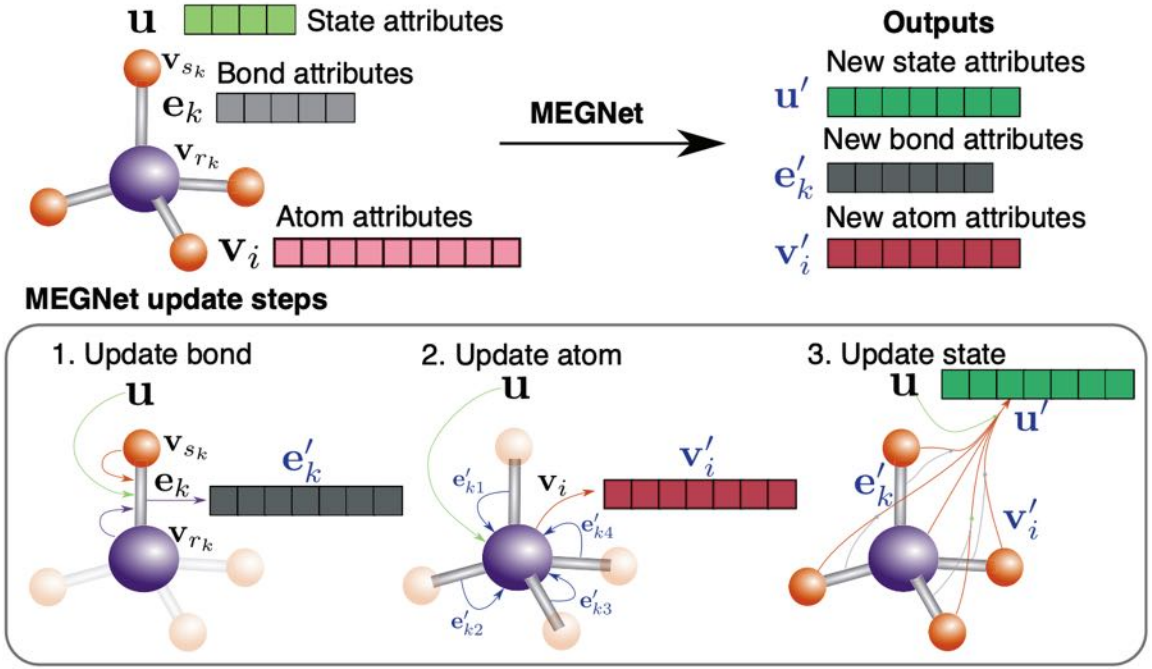
$(AA')_{0.5}(BB')_{0.5}O_3$ perovskite
2 x 2 x 2 supercell,
10 A and 10 B species
= $(^{10}C_2 \times {}^{8}C_4)^2 \approx 10^7$



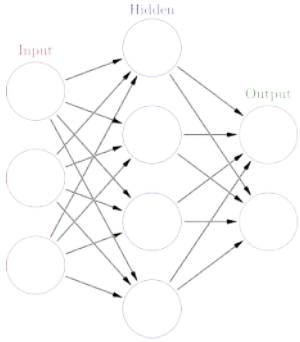
We need to be able to predict the property for any given arrangement of atoms ***instantly***…

Materials Project Timeline

MatErials Graph Network (MEGNet)



ϕ are update functions approximated using



Bond update

$$e'_{k1} = \phi_e(v_{rk} \oplus v_{sk} \oplus e_{k1} \oplus u)$$

Atom update

$$\overline{v_i^e} = \frac{1}{N_i^e} \sum_{k=1}^{N_i^e} \{e'_k\}_{r_k=1}$$

$$v'_i = \phi_v(\overline{v_i^e} \oplus v_i \oplus u)$$

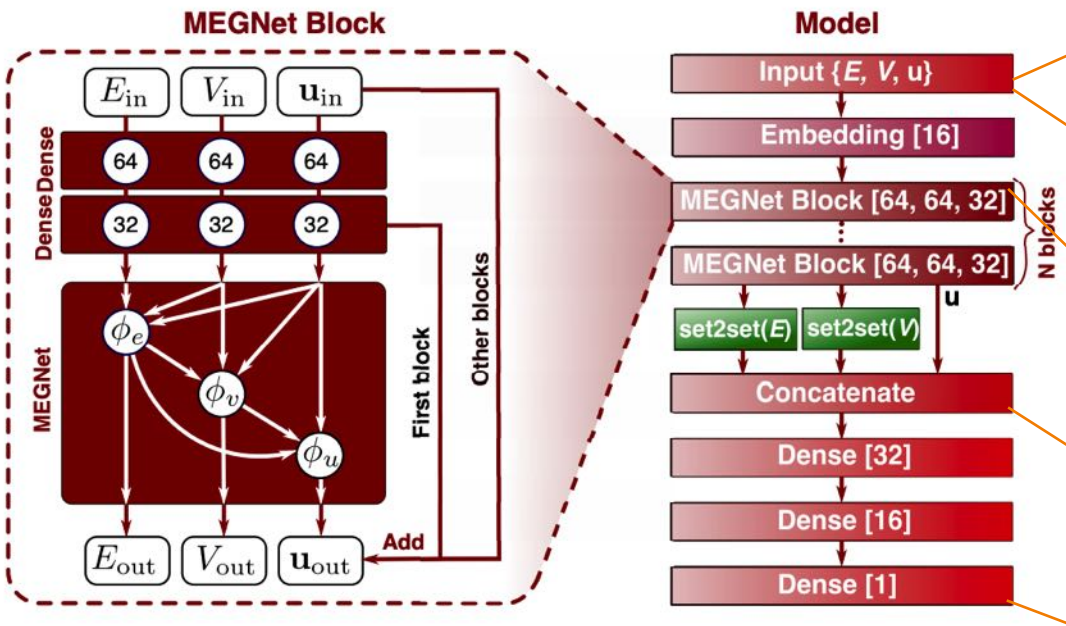
State update

$$\overline{u^e} = \frac{1}{N^e} \sum_{k=1}^{N^e} \{e'_k\}$$

$$\overline{u^v} = \frac{1}{N^v} \sum_{i=1}^{N^v} \{v'_i\}$$

$$u' = \phi_u(\overline{u^e} \oplus \overline{u^v} \oplus u)$$

Full model architecture



E: bonds attributes
V: atom attributes (basically *Z*)
u: state attributes

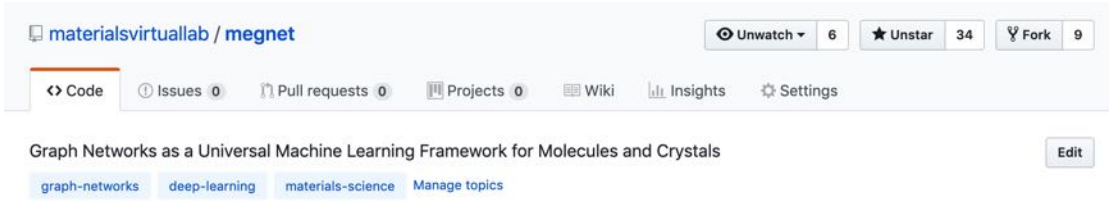
Atom number mapping to vector

Output a graph after each MEGNet block

Output a long vector for each structure

Output a single value for each structure

Implementation is open source at https://github.com/materialsvirtuallab/megnet.



Performance on 130,462 QM9 molecules

80%-10%-10% train-validation-test split

Property	Units	MEGNet-Full* (This Work)	MEGNet-Simple** (This Work)	Schnet[36]	enn-s2s[37]	Benchmark[32]	Target
ϵ_{HOMO}	eV	**0.038±0.001**	0.043	0.041	0.043	0.055 [a]	0.043
ϵ_{LUMO}	eV	**0.031±0.000**	0.044	0.034	0.037	0.064 [a]	0.043
$\Delta\epsilon$	eV	**0.061±0.001**	0.066	0.063	0.069	0.087 [a]	0.043
ZPVE	meV	**1.40±0.06**	1.43	1.7	1.5	1.9 [c]	1.2
μ	D	0.040±0.001	0.050	0.033	**0.030**	0.101 [a]	0.1
α	bohr^3	**0.083±0.001**	0.081	0.235	0.092	0.161 [b]	0.1
$\langle R^2 \rangle$	bohr^2	0.265±0.001	0.302	**0.073**	0.180	-	1.2
U_0	eV	**0.009±0.000**	0.012	0.014	0.019	0.025 [c]	0.043
U	eV	**0.010±0.000**	0.013	0.019	0.019	-	0.043
H	eV	**0.010±0.000**	0.012	0.014	0.017	-	0.043
G	eV	**0.010±0.000**	0.012	0.014	0.019	-	0.043
C_v	cal(molK)^{-1}	**0.030±0.001**	0.029	0.033	0.040	0.044 [c]	0.05
ω_1	cm^{-1}	**1.10±0.08**	1.18	-	1.9	2.71 [d]	10

11/13 properties reached chemical accuracy

11/13 properties state-of-the-art

Chemical Accuracy

Schnet: Schutt et al. J. Chem. Phys. 148, 241722 (2018)

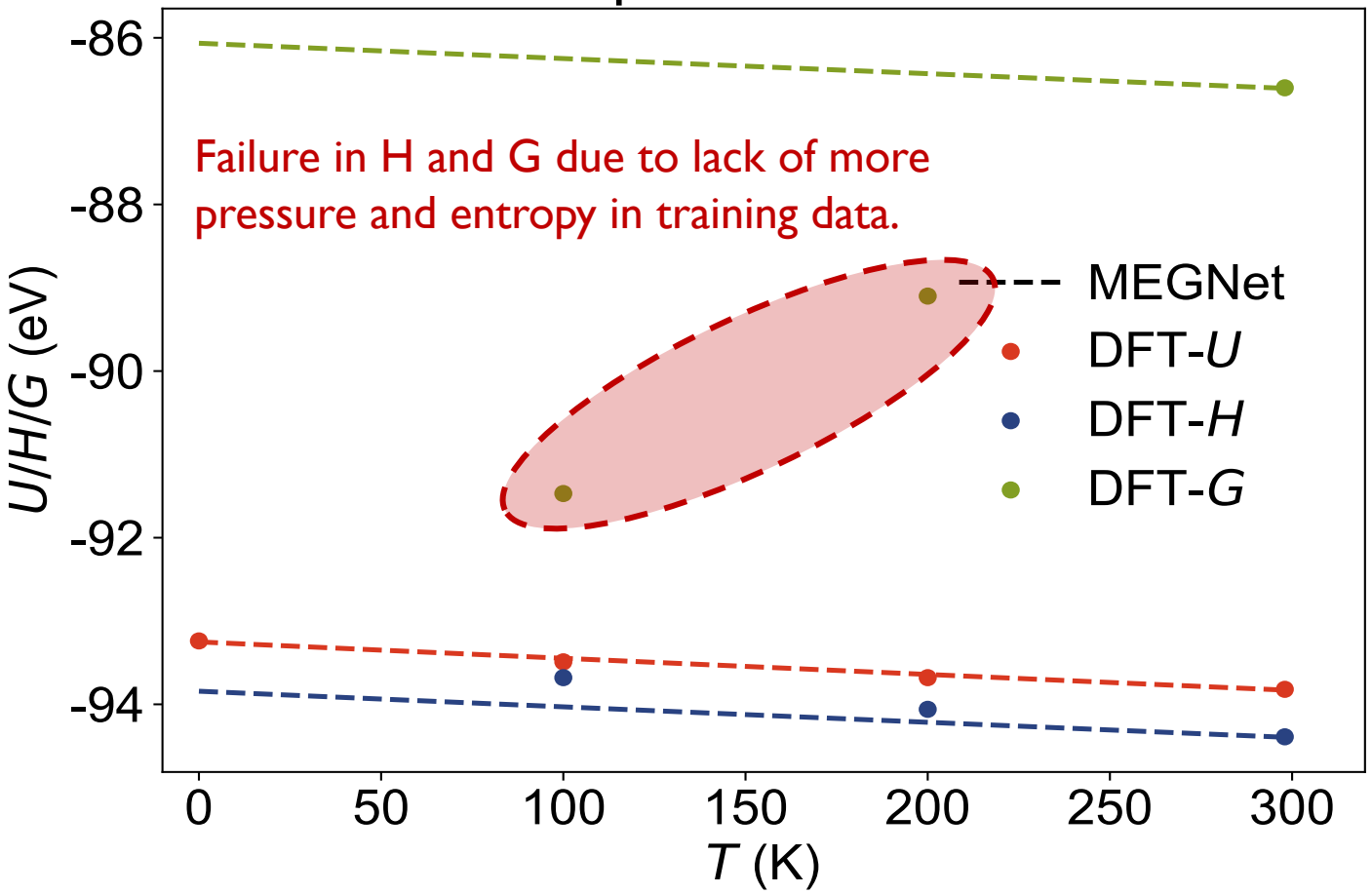
enn-s2s: Gilmer et al. *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 2017.

Unified Free Energy Model

Training data state:
U, H, G at 298 K
U at 0K (U0)

$$\emptyset = f(E, T, P, S))$$

qm9:099026



Performance on 69,000 Materials Project Crystals

"Noisy"
Dataset too small

	units	MEGNet	SchNet[36]	CGCNN[9]
elements		89	89	87
E_f	eV atom^{-1}	**0.028 ± 0.000** (60 000)	0.035 (60 000)	0.039 (28 046)
E_g	eV	**0.33 ± 0.01** (36 720)		0.388 (16 485)
K_{VRH}	$\log_{10}$ (GPa)	**0.050 ± 0.002** (4664)		0.054 (2041)
G_{VRH}	$\log_{10}$ (GPa)	**0.079 ± 0.003** (4664)		0.087 (2041)
metal classifier		78.9% ± 1.2% (55 391)		80% (28 046)
nonmetal classifier		90.6% ± 0.7% (55 391)		**95%** (28 046)

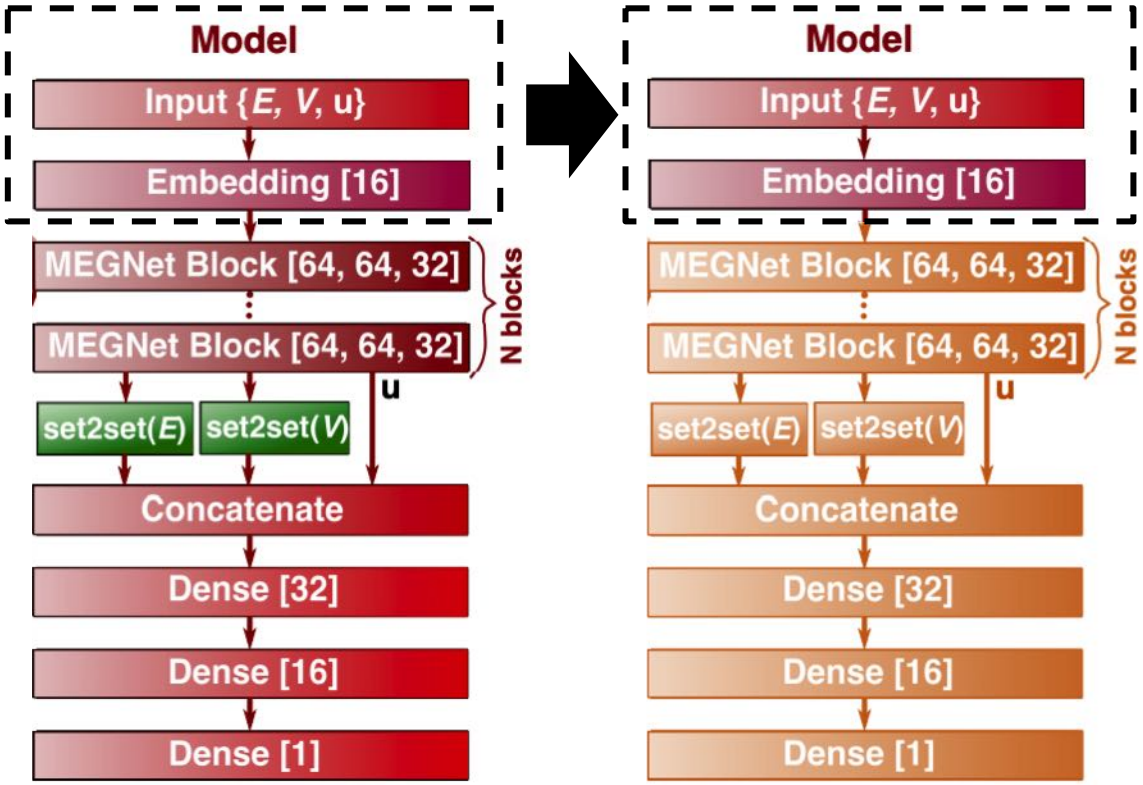
materials virtual·lab

Schutt et al. J. Chem. Phys. 148, 241722 (2018)
Xie et al. PRL. 120.14 (2018): 145301.

Transfer learning for improved convergence and speeed

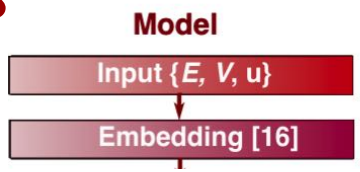
E_f model
(60,000 data points)

E_g model
(36,000 data points)

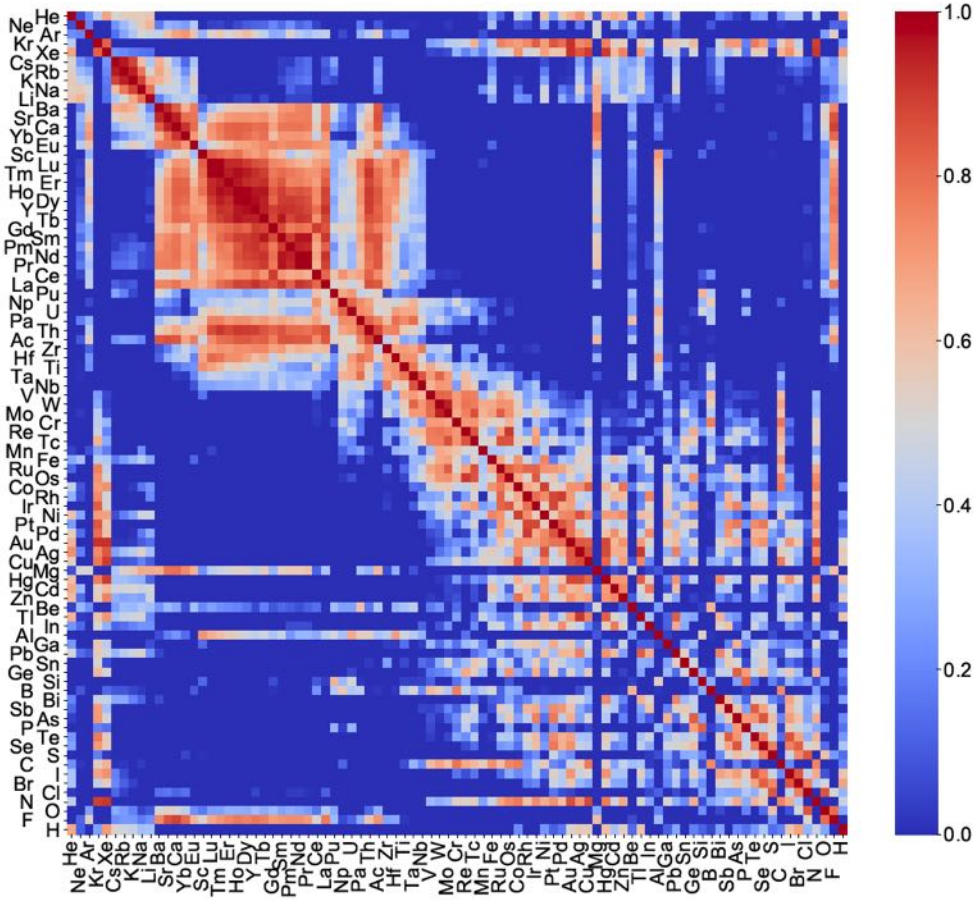


- ✓ MAE decreases from 0.38 eV to 0.32 eV
- ✓ Convergence speed x2

Extracting chemistry from machine-learned models

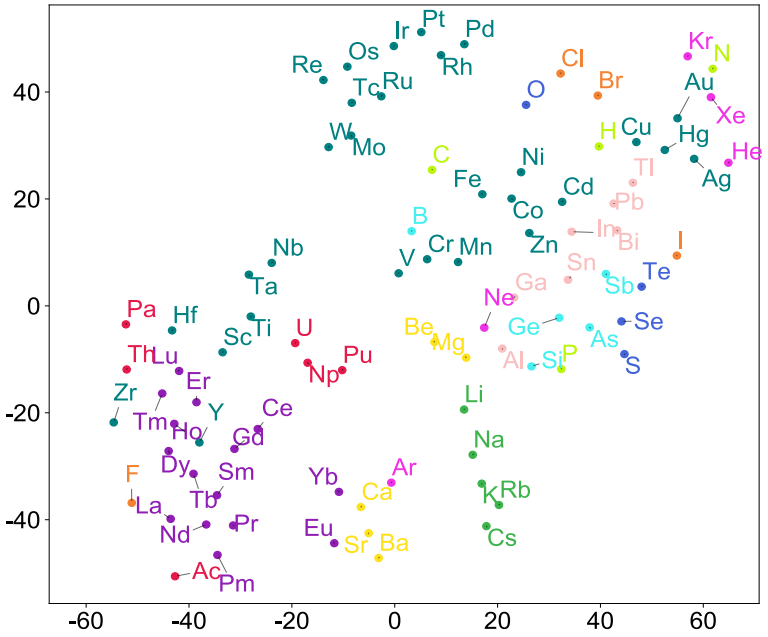


Pearson correlation



Sorted by Mendeleev number

t-SNE projection

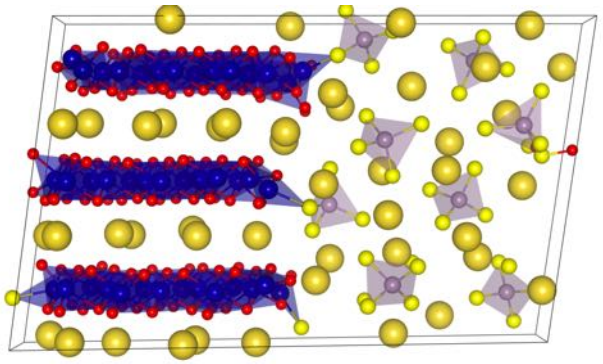


http://crystals.ai

Why machine learning? (Part 2)

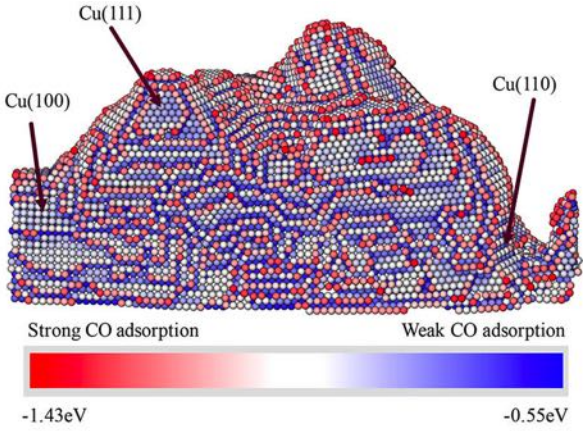
Many real-world materials problems are not related to bulk crystals.

Electrode-electrolyte interfaces



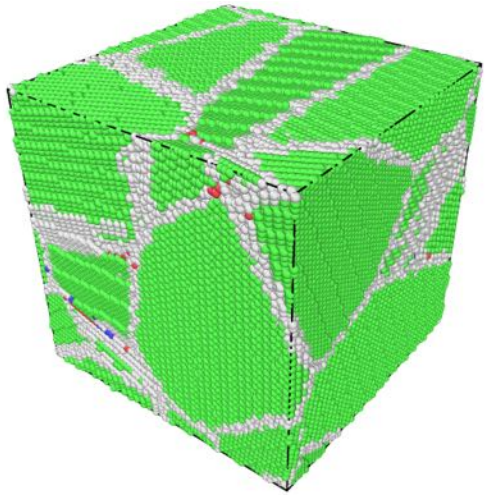
Tang et al. Chem. Mater. 2018, 30 (1), 163–173.

Catalysis



Huang et al. ACS Energy Lett. 2018, 3 (12), 2983–2988.

Microstructure and segregation



Need linear-scaling with ab initio accuracy.

General procedure

Sample a sufficiently large dataset → Describe local environment → Learn relationship between features and energy, force, etc.

Open databases

MATERIALS PROJECT, AFLOW, NOMAD, OQMD

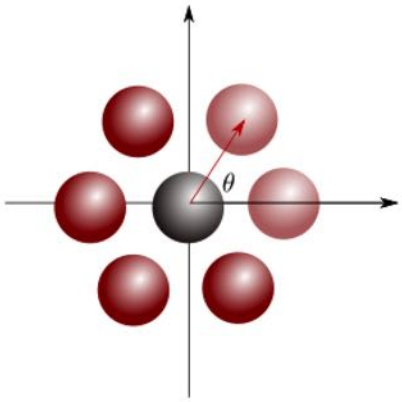
DIY

pymatgen, atomate, FireWorks, ASE

- Requirements
  - Invariance to rotation, reflection, translation, and permutation
  - Uniqueness
  - Differentiability
- Examples:
  - Coulomb matrix
  - Symmetry functions
  - Bispectrum
  - Smooth overlap of atomic positions
  - Fragment descriptors
  - …

- Linear regression
- LASSO
- Kernel ridge regression
- Random forest
- SVMs
- Neural networks
- ….

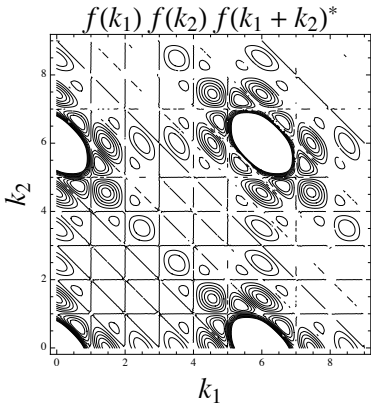
A many-body atomic environment descriptor: bispectrum coefficients



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

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

Bispectrum



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

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

Gaussian approximation potential (GAP)[1]

$$E_{\mathrm{GAP}} = \sum_n \alpha_n G(\boldsymbol{B}, \boldsymbol{B}_n)$$

Gaussian process regression (nonparametric model)

Spectral neighbor analysis potential (SNAP)[2]

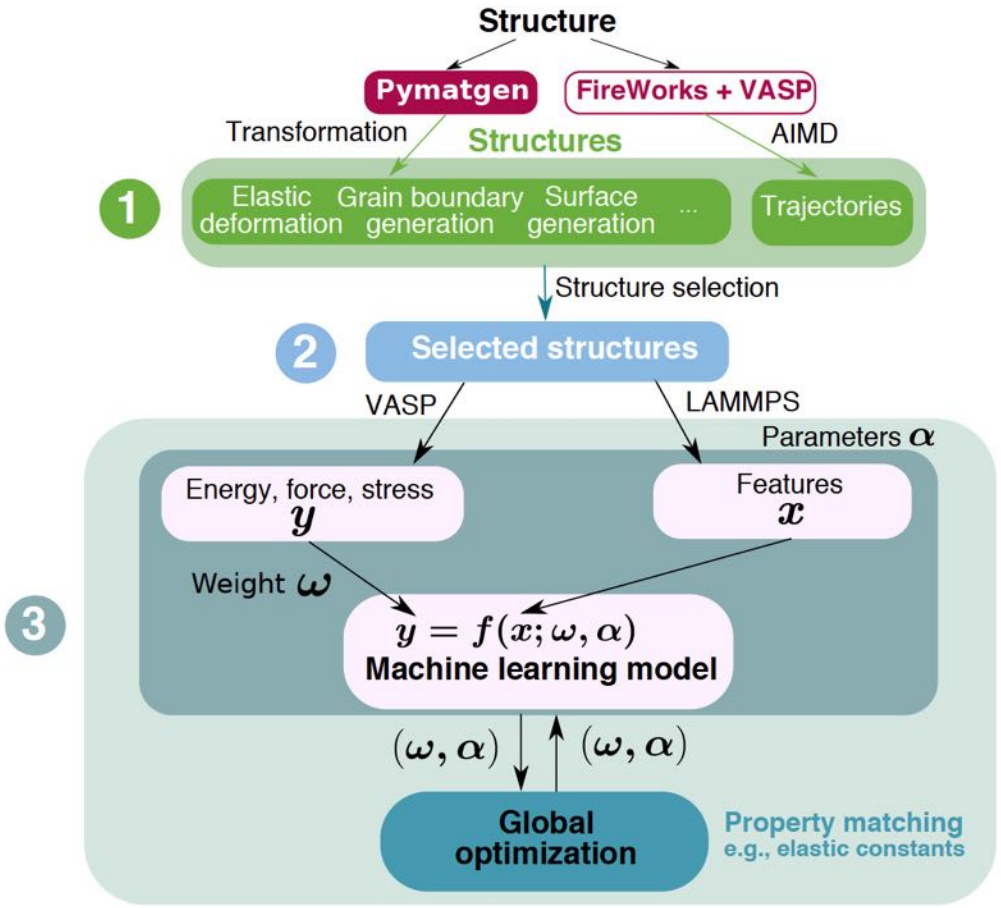
$$E_{\mathrm{SNAP}} = \beta_0 N + \boldsymbol{\beta} \cdot \sum_{i=1}^{N} \boldsymbol{B}^i$$

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

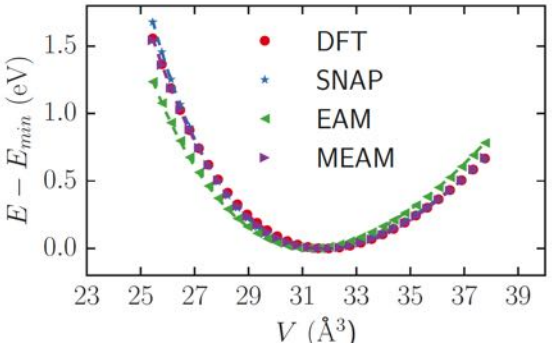
$$\boldsymbol{\sigma}^j_{\mathrm{SNAP}} = -\boldsymbol{\beta} \cdot \sum_{j=1}^{N} \mathbf{r}_j \otimes \sum_{i=1}^{N} \frac{\partial \boldsymbol{B}^i}{\partial \mathbf{r}_j}$$

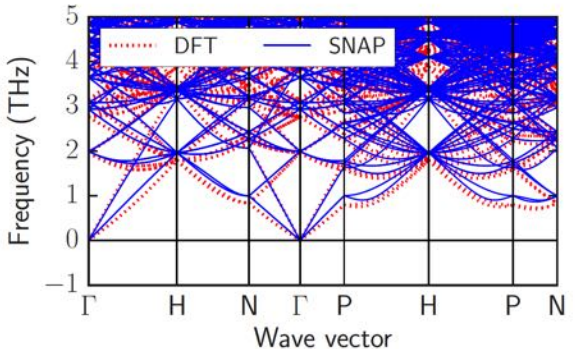
[1] Bartok et al. Phys. Rev. Lett. 2010, 104 (13), 136403.
[2] Thompson et al. J. Comput. Phys. 2015, 285, 316–330 DOI: 10.1016/j.jcp.2014.12.018.

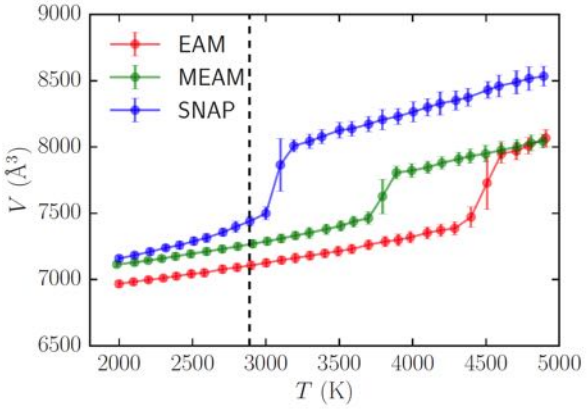
Models: Quantum-accurate force-field for Mo



Chen et al. Phys. Rev. Mater. 2017, 1 (4), 43603 DOI: 10.1103/PhysRevMaterials.1.043603.

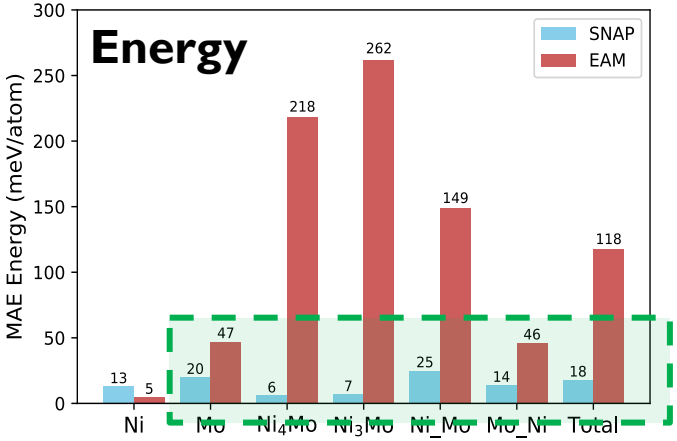


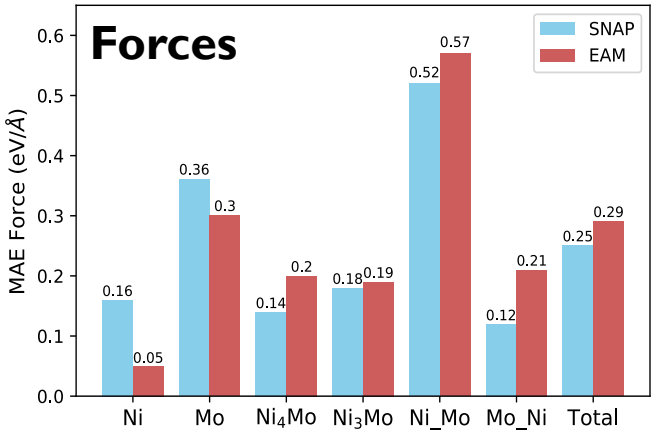




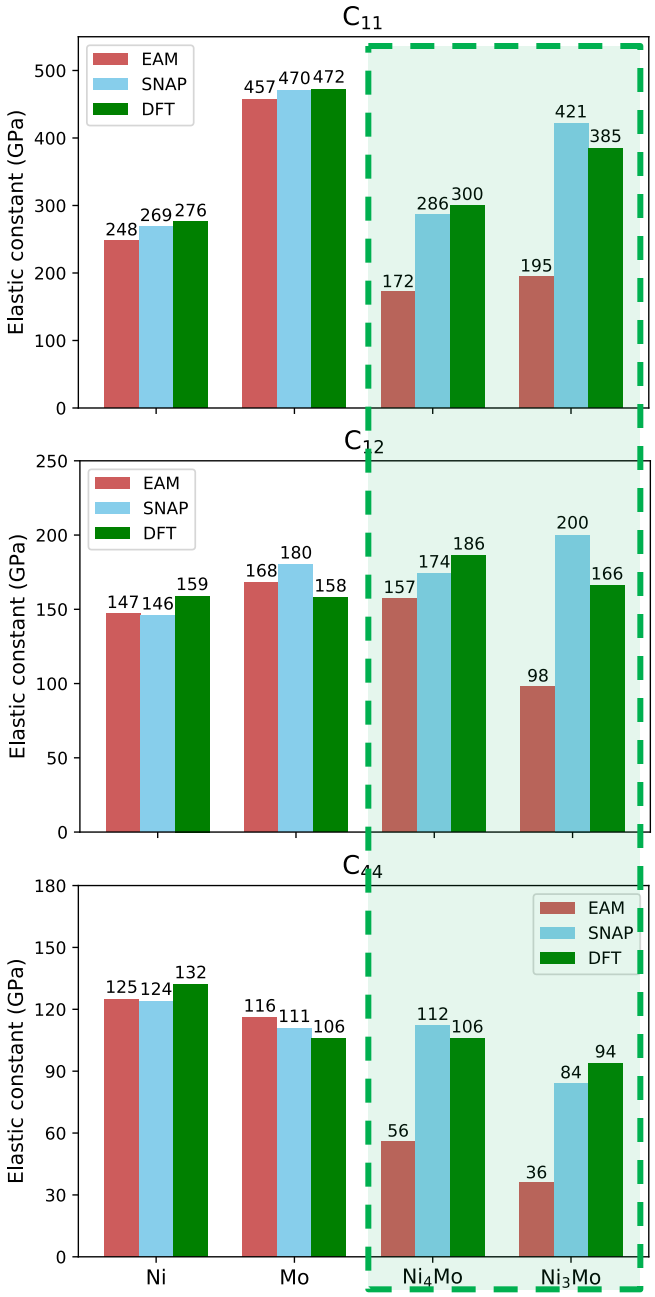
Ni-Mo SNAP performance

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

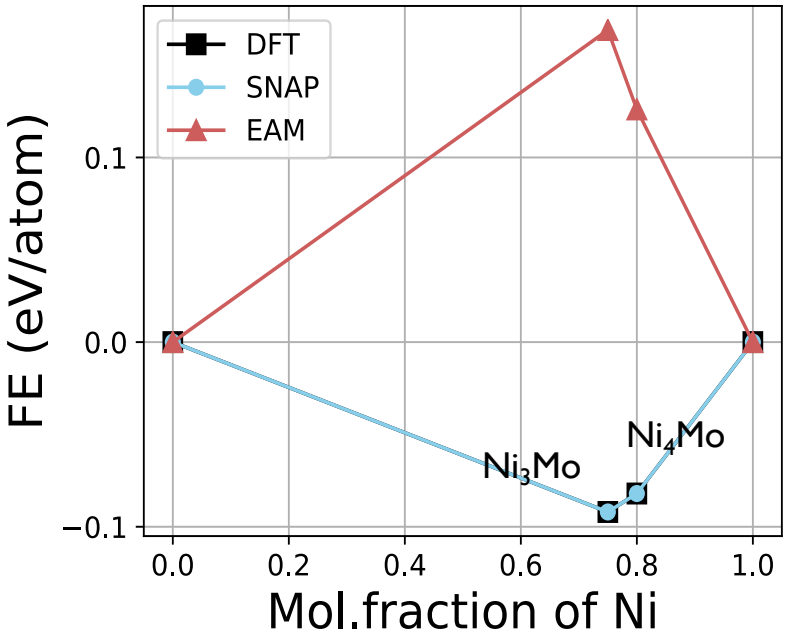




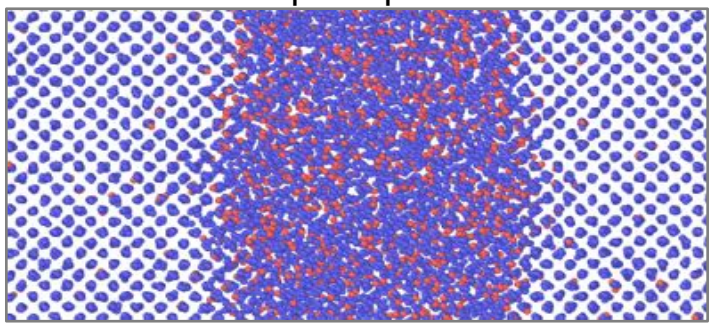
Elastic constants

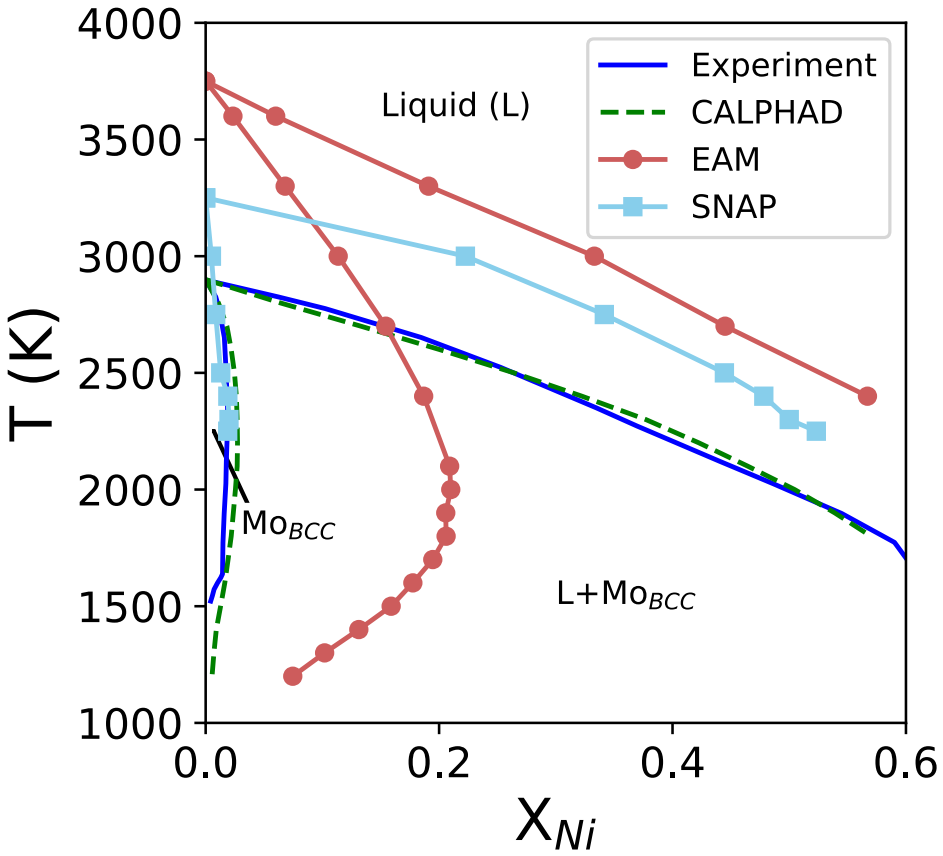


Ni-Mo phase diagram



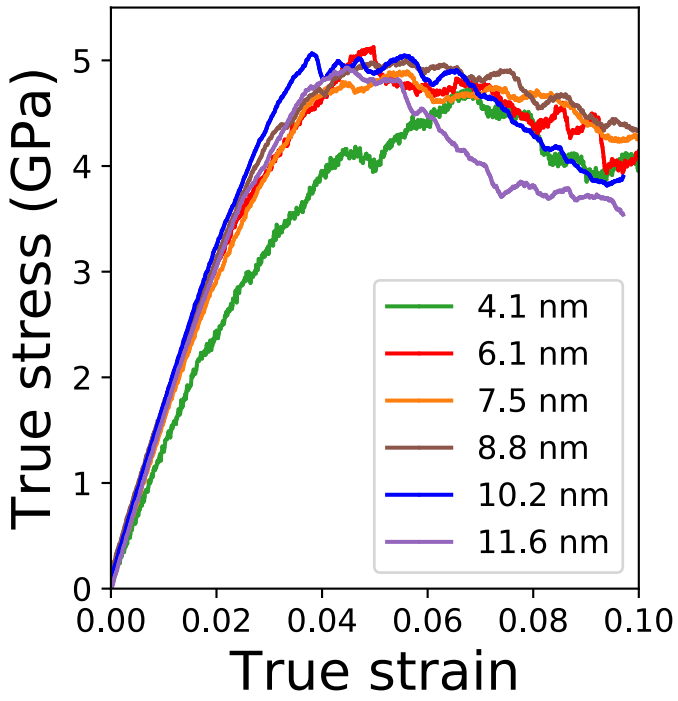
Solid-liquid equilibrium

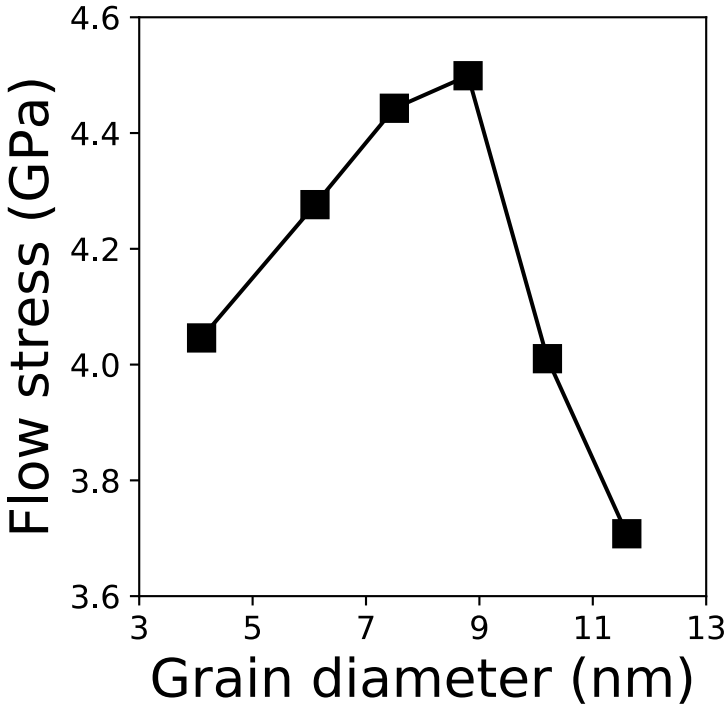




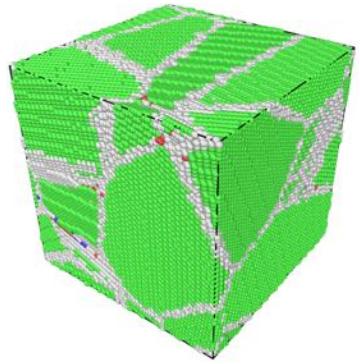
EAM completely fails to reproduce Ni-Mo phase diagram

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 Hall-Petch relationship, consistent with experiment[1].



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

Conclusions