This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

SAND2021-9633C

Sandia National Laboratories

SAND2021-XXXX X

Interatomic Potentials for Materials Science and Beyond; Advances in Machine Learned Spectral Neighborhood Analysis Potentials

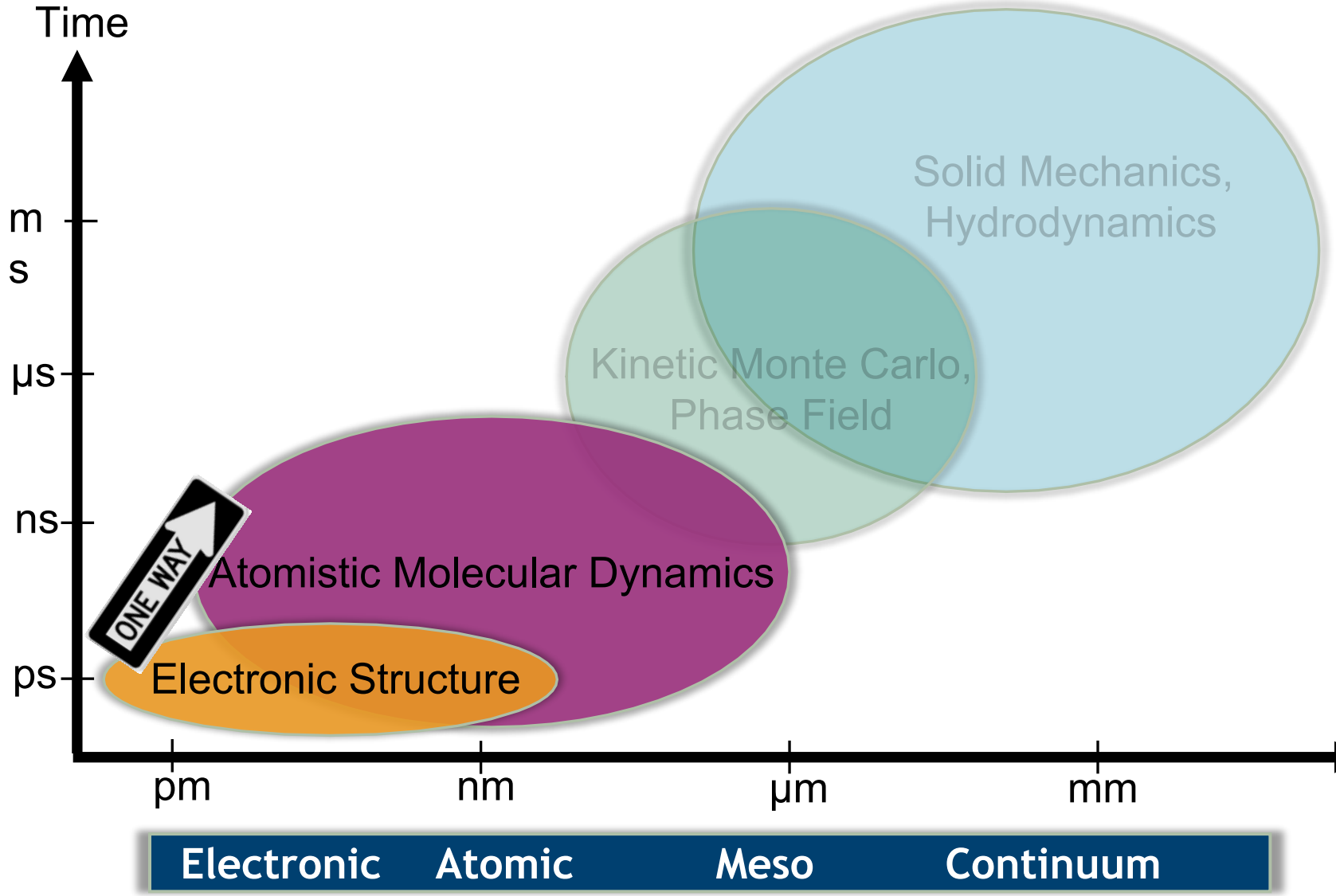
Mitchell Wood, Mary Alice Cusentino, David Montes de Oca Zapiain, Aidan Thompson
Sandia National Labs
Ivan Oleynik
University of South Florida

LAMMPS Users Conference, August 10-12th, 2021

Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

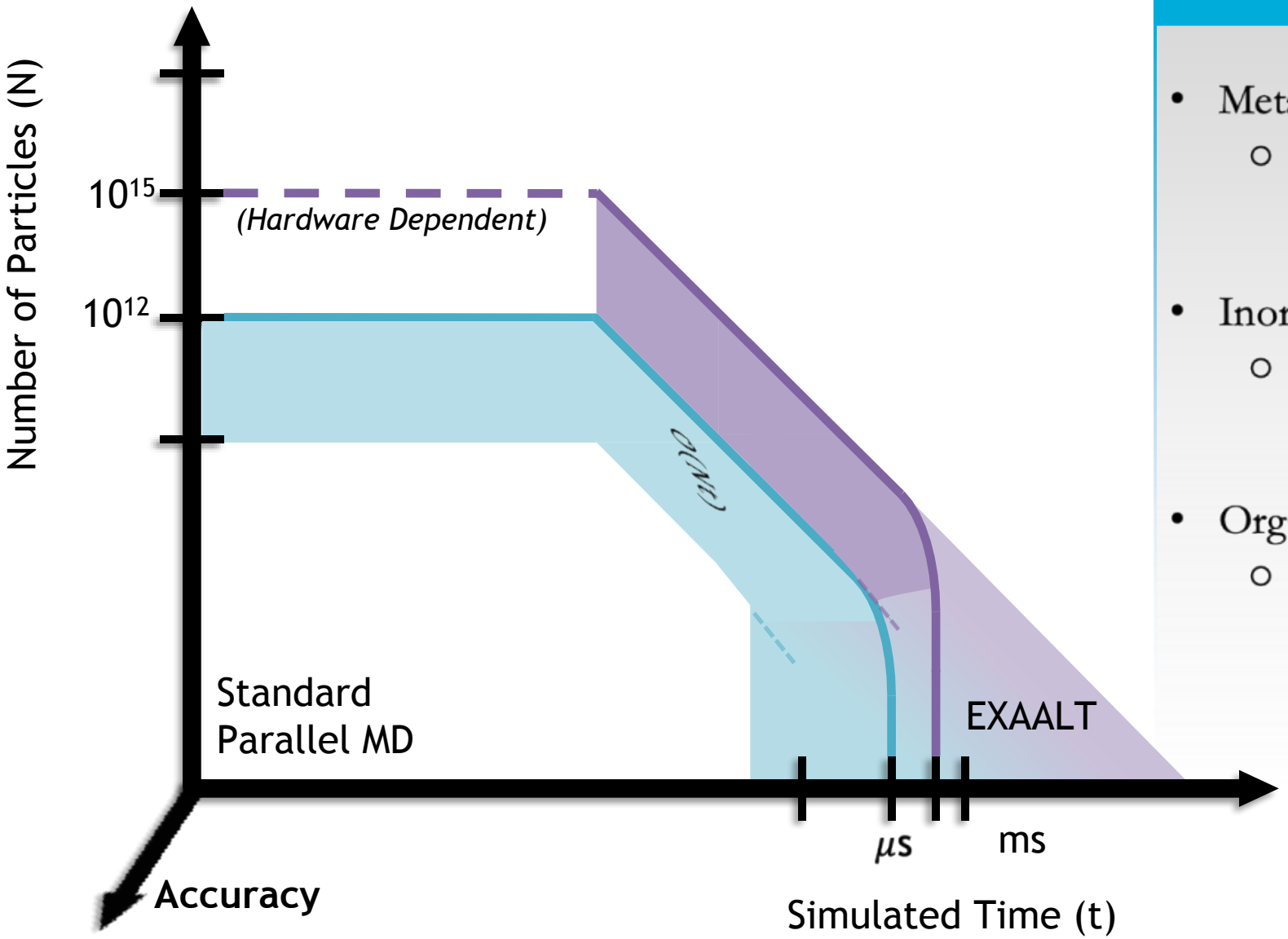
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Interatomic Potentials as Multi-Scaling



- IAP can be useful without being physically motivated
- Preserving accuracy through scales while becoming computationally efficient
- Need to be cautious of what is promised with machine learning, most of MD will be extrapolation

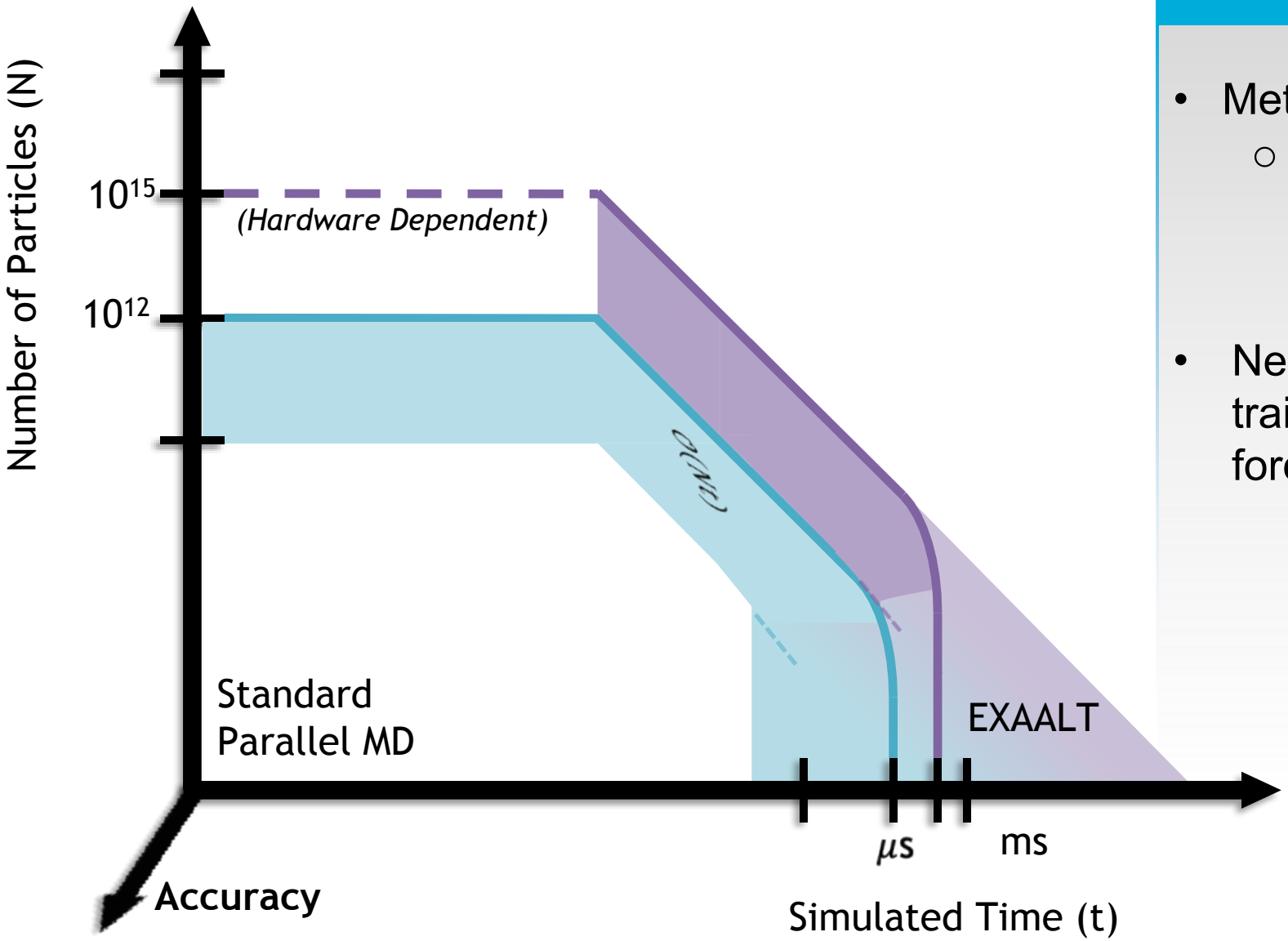
Return to the Master Plot



Classical, Empirical Potentials

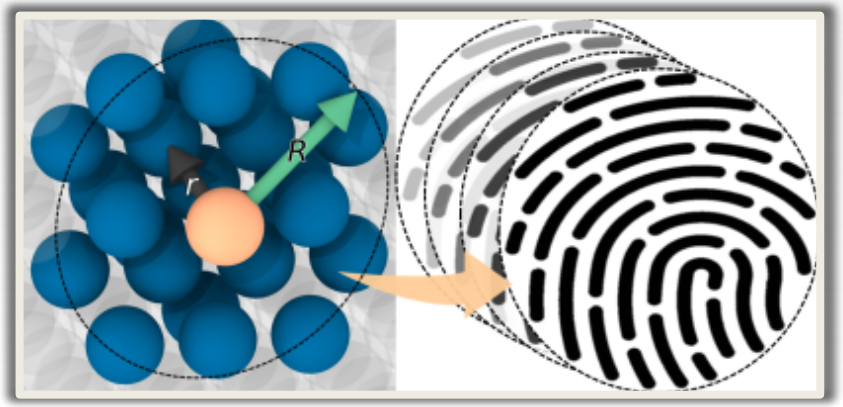
- Metals
  - EAM: Assume spherical electron density
    $$E_i = F_\alpha\left(\sum_{j\neq i}\rho_\beta(r_{ij})\right) + \frac{1}{2}\sum_{j\neq i}\emptyset_{\alpha\beta}(r_{ij})$$
- Inorganic
  - Stillinger-Weber: Assume 2,3-body harmonic springs
- Organic
  - ReaxFF: Assume covalent bonding, smooth bond-orders between all interacting atoms

Return to the Master Plot



Machine Learned Potentials

- Metals, Inorganic, Organic, etc.
  - Assume energy and forces are some function of local atomic neighborhood descriptors
- Needs reference data to be properly trained to get the 'right' energies and forces



SNAP Applications

SNL Involved, Independent

System	Year	Usage	Origin	N_{DoF}	$N_{Training}$	Descriptors
Ta	2014	Dislocation motion	SNL, Thompson	31	363	Linear
InP	2015	Radiation damage, defects	SNL, Thompson	31	665	Linear
WBeHe	2017	Plasma facing materials	SNL, Wood	56	25,052	Linear
Mo	2017	Phase diagram prediction	UCSD, Ong	31	1000	Linear
Actinides	2018	Shock, phase transitions	SNL/LLNL	56	20,000	Quadratic
NiMo	2018	Phase diagram prediction	UCSD, Ong	31	2,000	Linear
LiN	2019	Super-Ionic Conductor	UCSD, Ong	31	3,000	Lin+Charge
★ Various	2020	Accuracy/Cost comparison	UCSD/SNL	10-130	1,000	Lin, Quad
InP	2020	Radiation damage, defects	SNL, Cusentino	241	1,000	EME
AlNbTi	2020	High entropy alloy design	SNL, Tranchida	1596	7,250	Quadratic
Si	2020	Neural network SNAP	UNLV, Zhu	1596	>5,000	NN
Al	2021	Predicting electron density	SNL, Ellis	91	30	NN
Fe	2021	Magnetic phase transition	SNL, Nikolov	1596	683	Quad+Spin

SNAP Applications

SNL Involved, Independent

(more in the literature, not an exhaustive li

System	Year	Usage	Origin	N_{DoF}	$N_{Training}$	Descriptors
WBeHN	-	Plasma facing materials	SNL, Cusentino	56*	>40,000	Linear
★ C	-	Planetary impacts, shock	USF, Willman	1596	30,000	Quadratic
★ C, V	2021	Metal plasmas	SNL, Wood	1596	10,000	Quadratic
MoNbTaT	-	HEA alloy design	SNL, Startt	-	>5,000	EME
GeSe	-	Vitrification	UCD, Sievers	-	>5,000	EME
LiMoS	-	Li-ion batteries	UConn, Dongarre	-	>5,000	-
SiGeSnPb	-	Thermoelectric materials	GWU, Li	-	>5,000	-
★ W	-	Model form selection	LANL/SNL	-	330,000	NN

So what should you train a ML-IAP on?

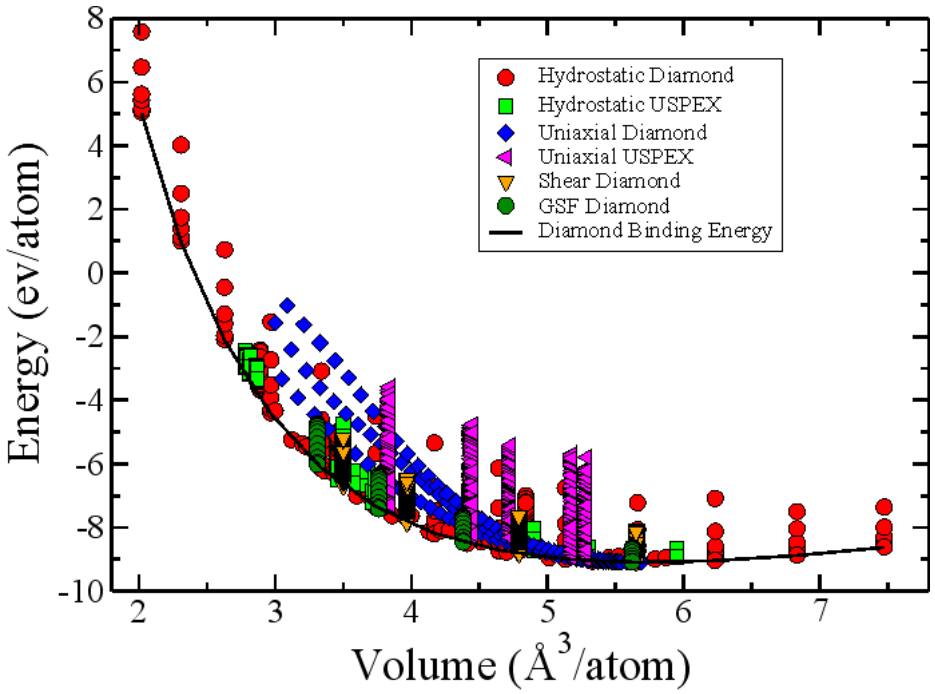
How do you recognize failures (poor extrapolatio

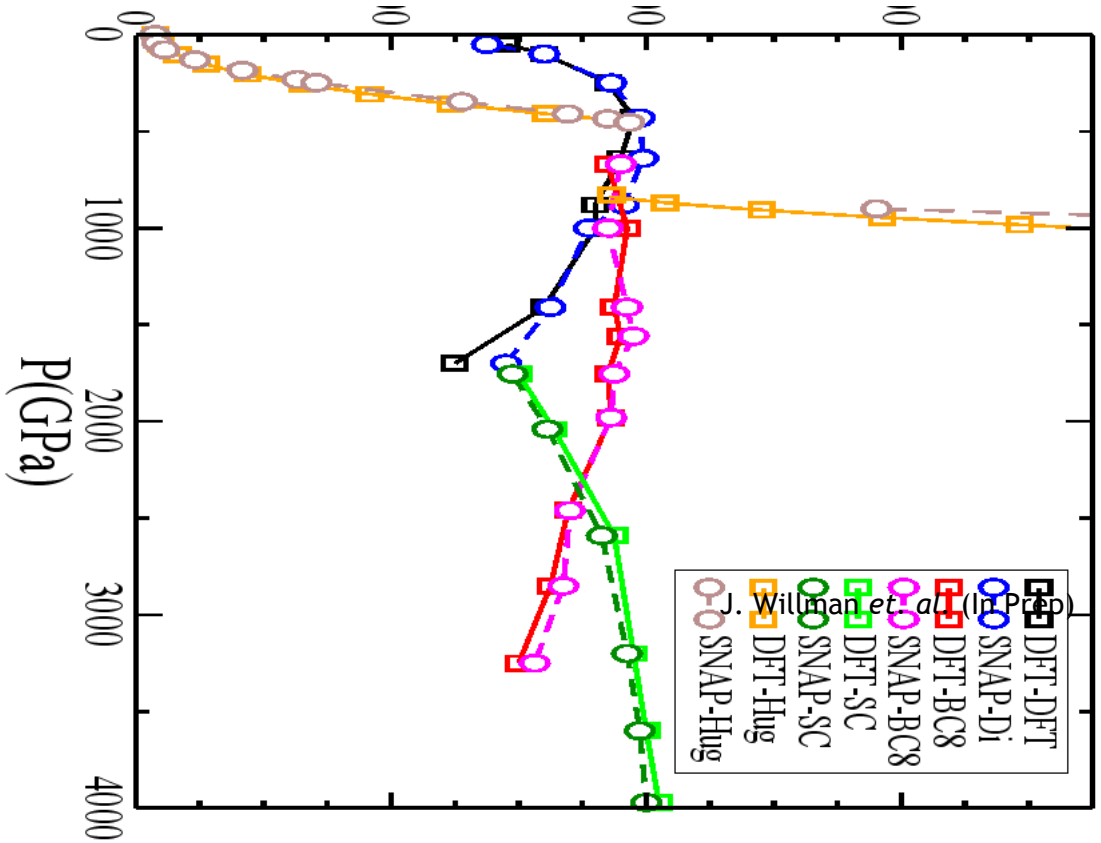
- Growing evidence that SNAP is a *general use* material model form, unlike any interatomic potential used in MD to date
- SNAP model training software now incorporated in Materials Design Inc. products

To be or Not to be Transferrable

Simple Model, Complex Descriptor

- A general use IAP is much more challenging to create.
- Phases of Carbon from 0-4TPa, 0-1.3eV reproduced because it was trained to do so.



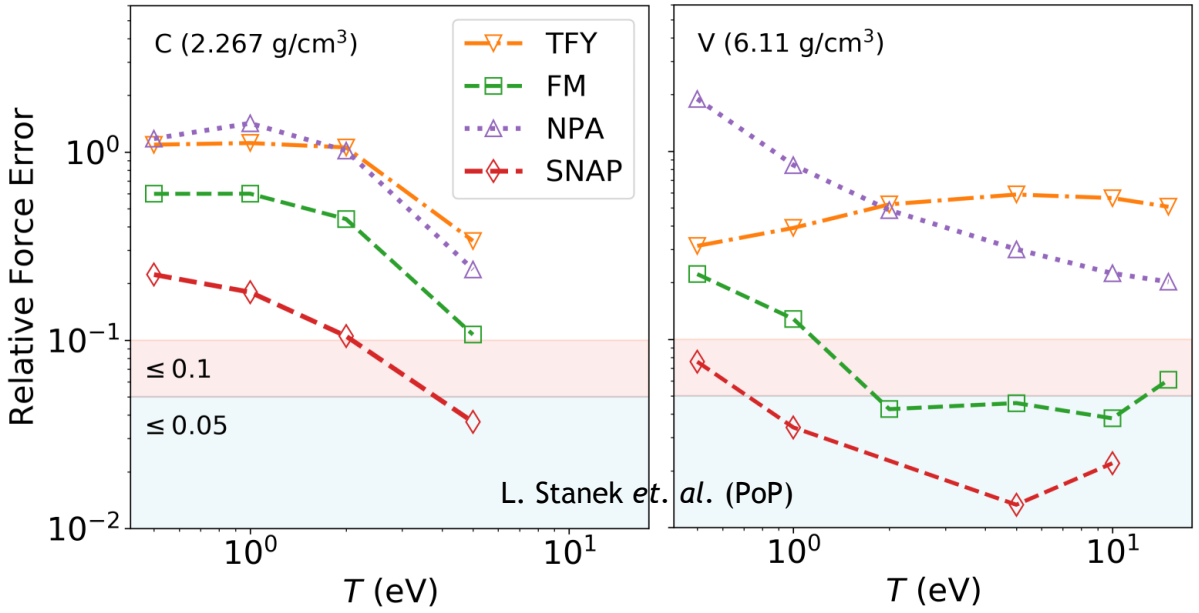


★ C	-	Planetary impacts, shock	USF, Willman	1596	30,000	Quadratic

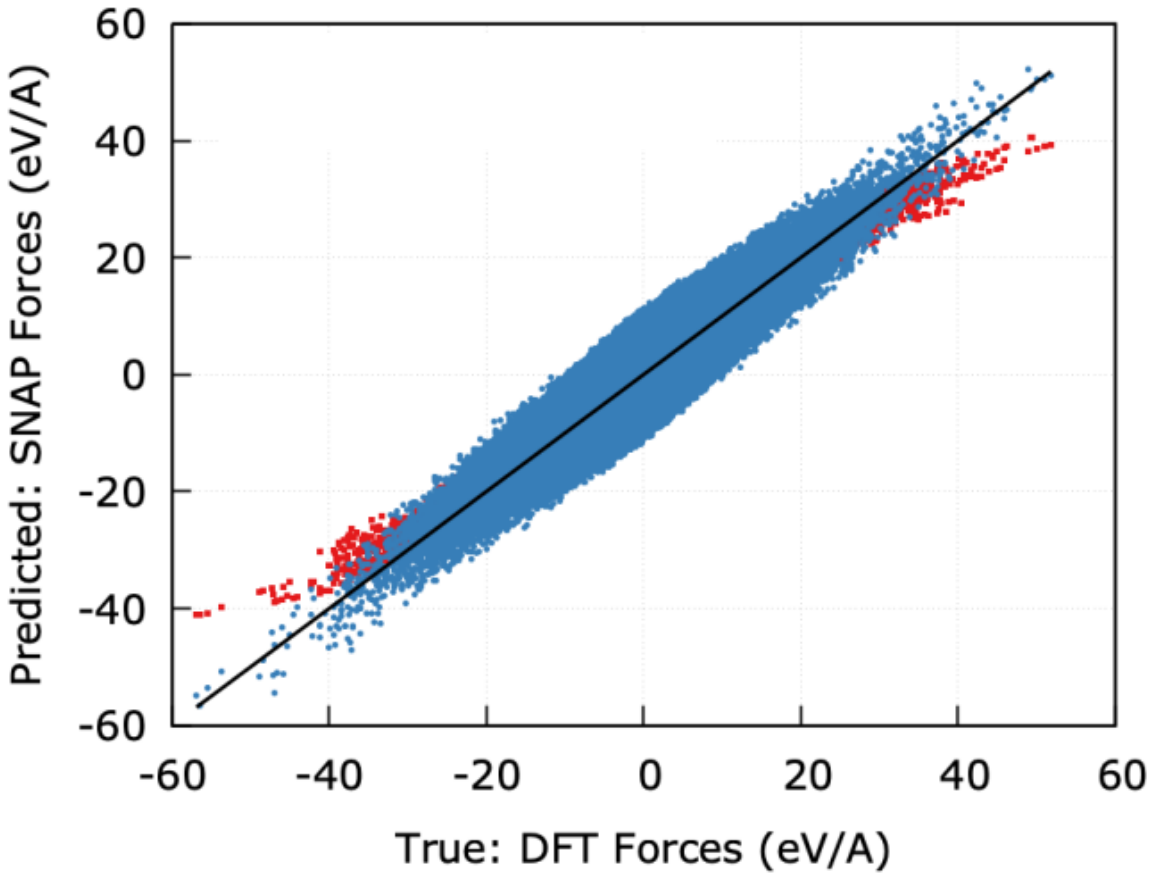
To be or Not to be Transferrable

Simple Model, Complex Descriptor

- A ‘Local’ potential is ideal for a narrow phase space, force matching IAP have done this for decades.
- Is pressure, temperature, composition a proper definition of training space of ML-IAP?



- **(red)** IAP Trained at T=1eV, evaluated at T=1eV
- **(blue)** IAP Trained at T=1eV, evaluated at T=5eV

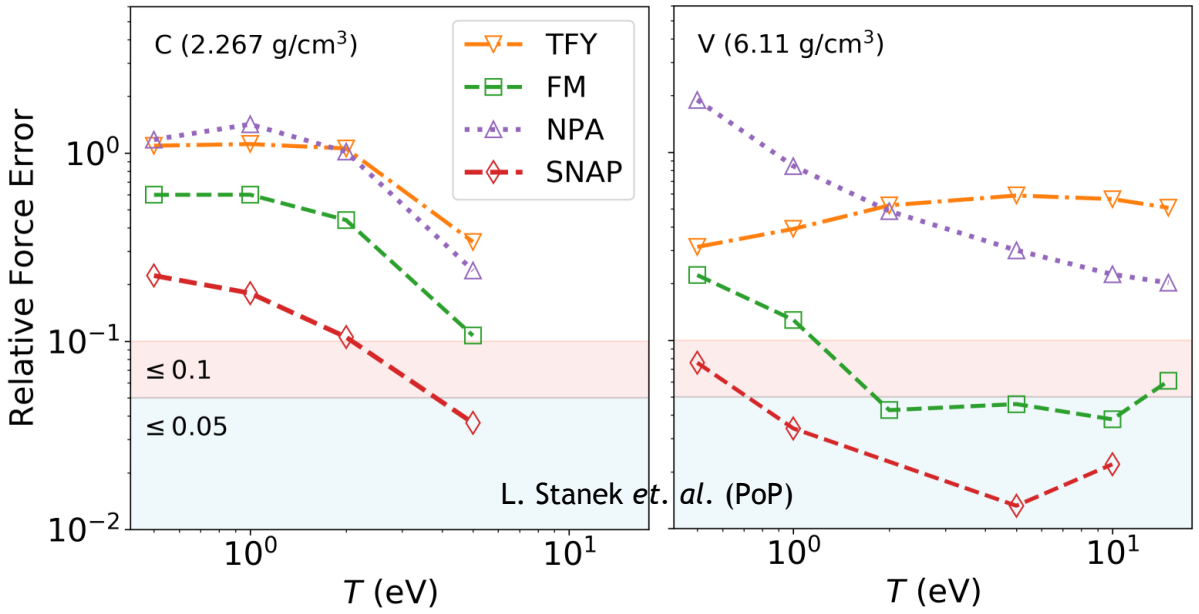


C, V	2021	Metal plasmas	SNL, Wood	1596	10,000	Quadratic

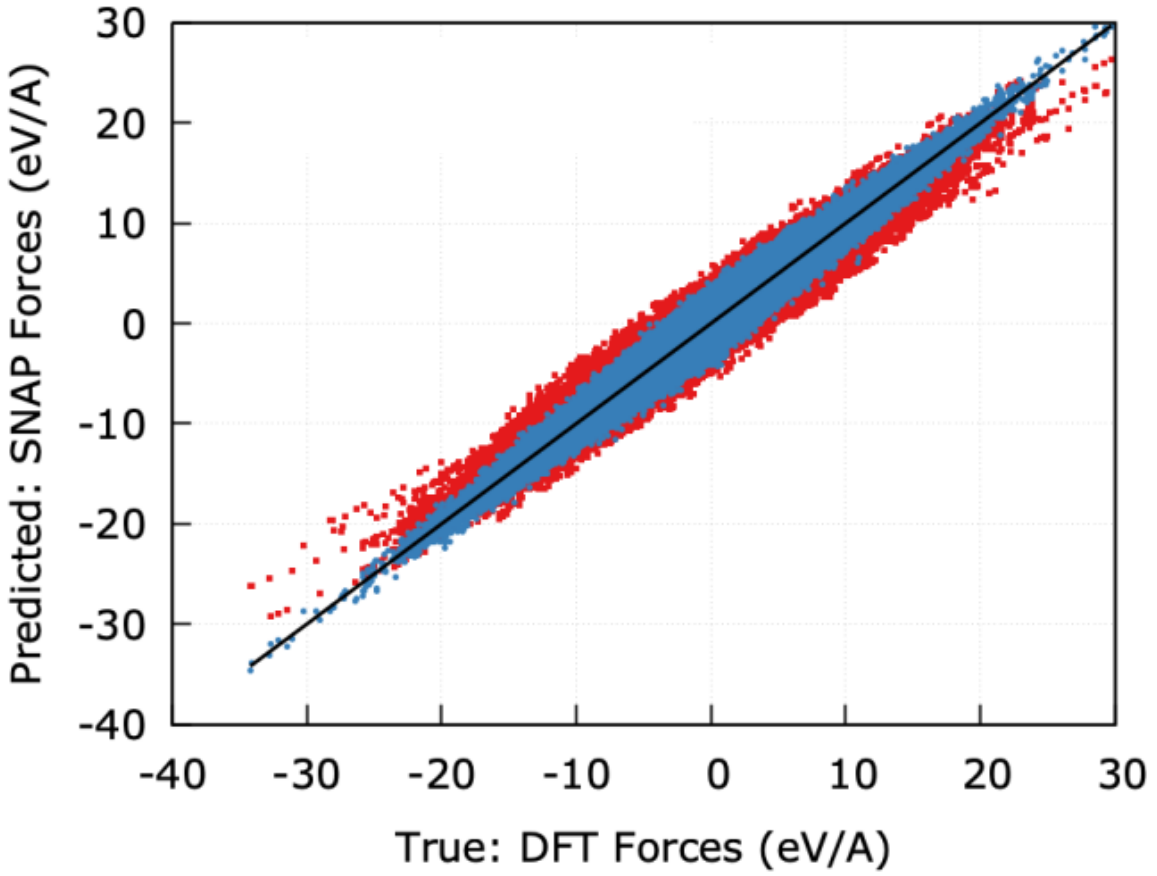
To be or Not to be Transferrable

Simple Model, Complex Descriptor

- A ‘Local’ potential is ideal for a narrow phase space, force matching IAP have done this for decades.
- Is pressure, temperature, composition a proper definition of training space of ML-IAP?

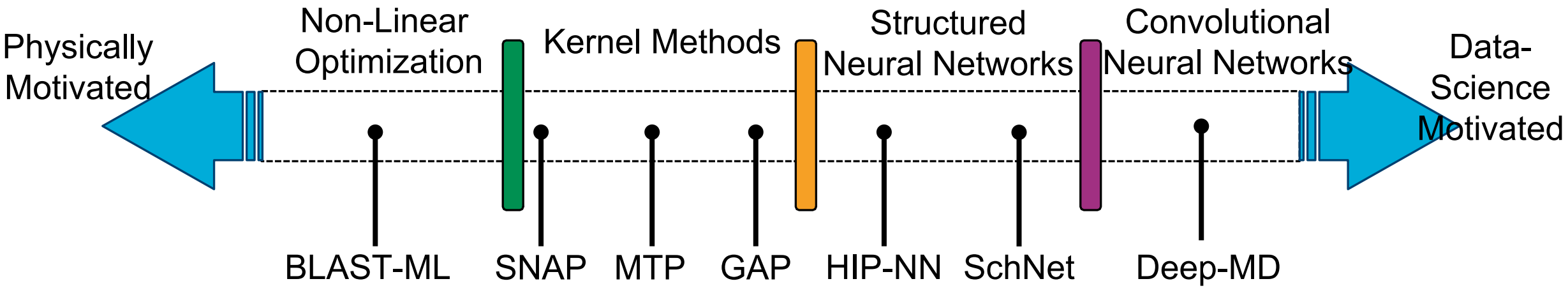


- **(red)** IAP Trained at T=5eV, evaluated at T=5eV
- **(blue)** IAP Trained at T=5eV, evaluated at T=1e

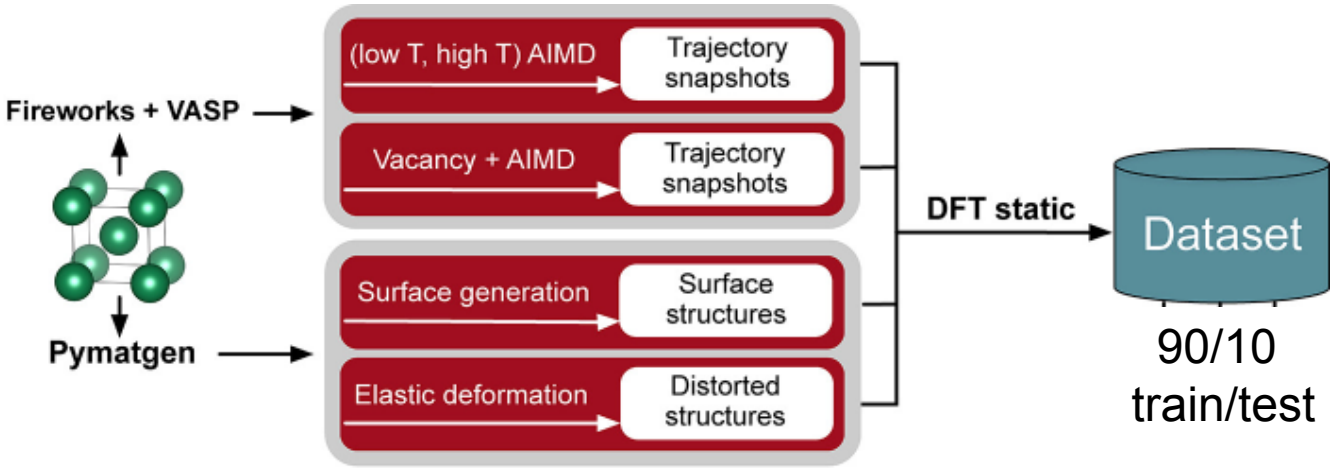


C, V	2021	Metal plasmas	SNL, Wood	1596	10,000	Quadratic

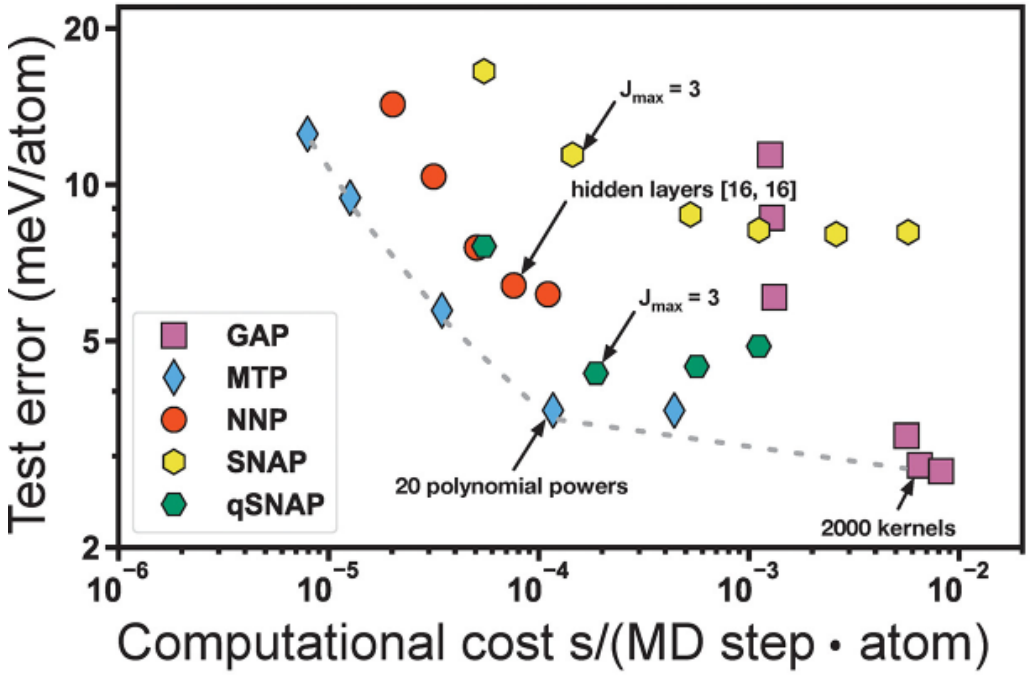
Environment of Machine Learning Techniques



https://github.com/materialsvirtuallab/mlearn



^ Credit to Y. Zou, X. Li, C. Chen and S.P. Ong



Entropy Maximizing Training Sets

Scalable Training Generation

- Self-entropy & cross entropy pseudo-potential. Maximize descriptor diversity over the whole training set. Prevent ‘bad’ configurations with strong short range repulsion.
- Generated 300,000 configurations of Tungsten in DFT, one of the largest training set ever assembled

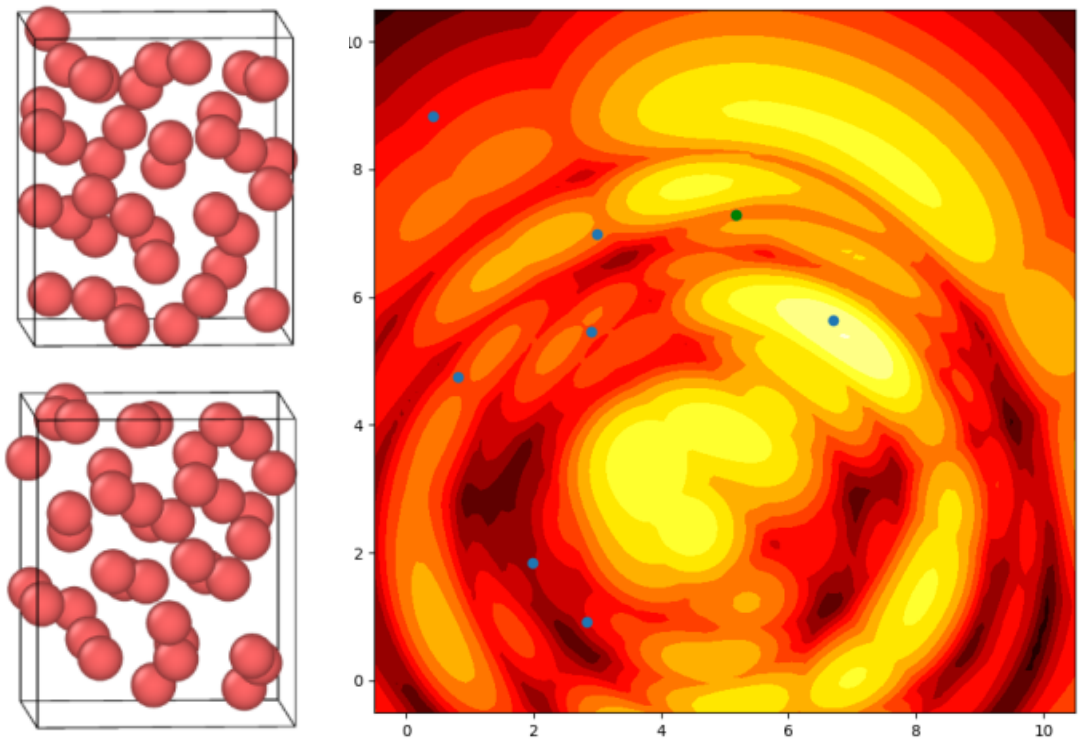
Advantages

- Applicable to any descriptor, not glued to SNAP in particular
- Prevents “holes” in descriptor space
- Finding configs is fast and scalable

J. Chem. Phys. 153, 094110 (2020)

$$V(x) = -H(x,y) \pm S(x)$$

Pseudo-potential | Cross-entropy | Self-entropy



Self-entropy landscape of one particle in a model system with one descriptor: the average interatomic distance

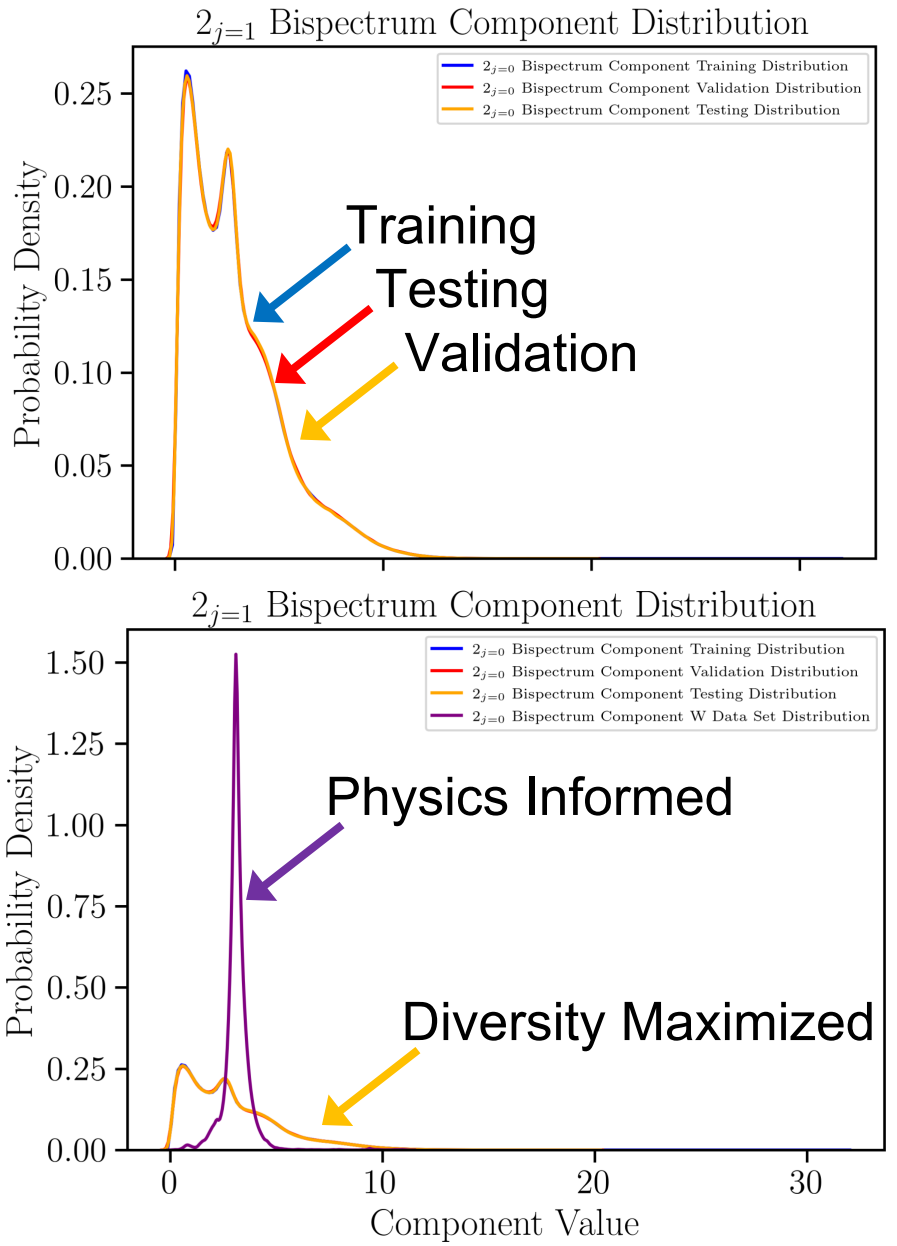
Model Form Selection

Which Model, Which Training

- ‘Simple’ training sets can be captured by nearly all model forms
- How should you choose your model form based on the generated training?

Physical training, or Entropy Maximized

- Comparing the entropy maximized training to the hand constructed training of Wood *et. al.* PRB 2019
- Automated, entropy maximized training contains descriptor space of hand tuned set.
- ‘Real’ correlations in descriptors will be lacking



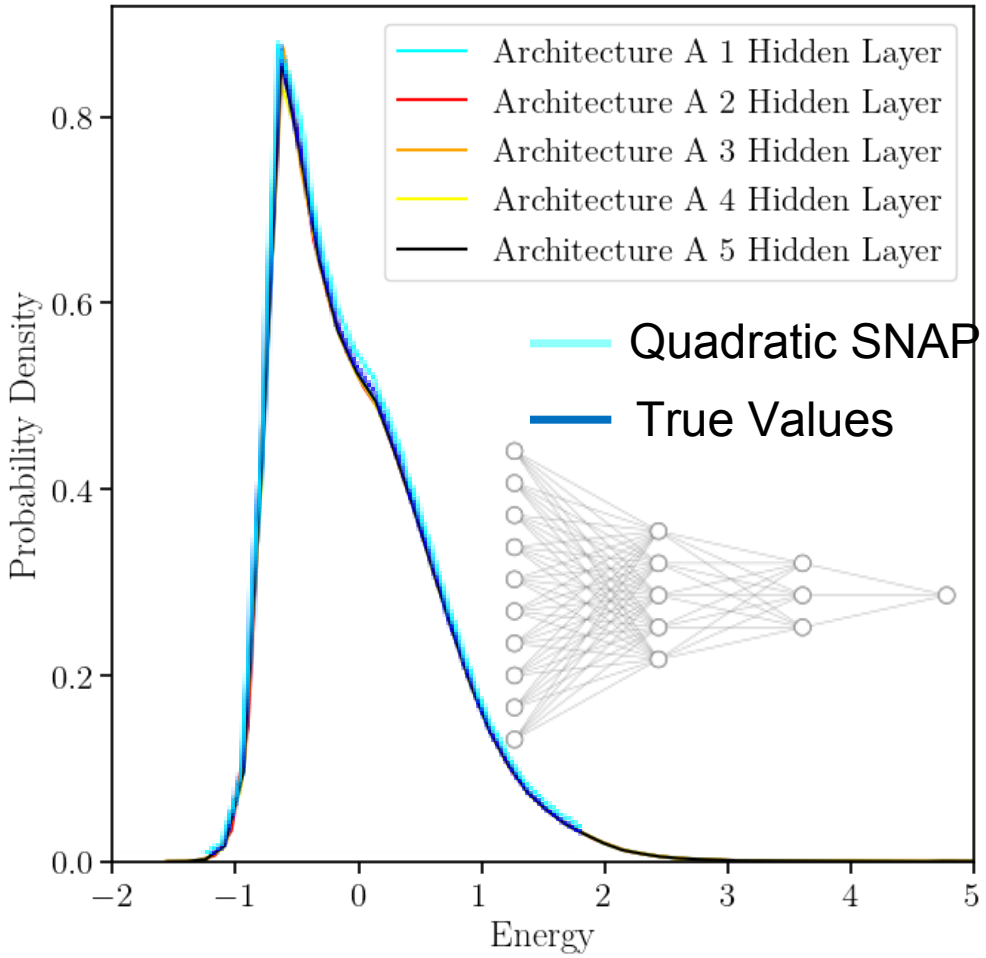
Comparing Model Complexity

Which Model, Which Training

- ‘Simple’ training sets can be captures by nearly all model forms
- How should you choose your model form based on the generated training?

Head-to-head Comparison

- Both NN and traditional SNAP models use bispectrum components as input descriptors
- All (Linear to deep-NNs) show similar performance on training set (?!)



NNs: Step-down, Soft+, 30 inputs, 3k-30k DoF

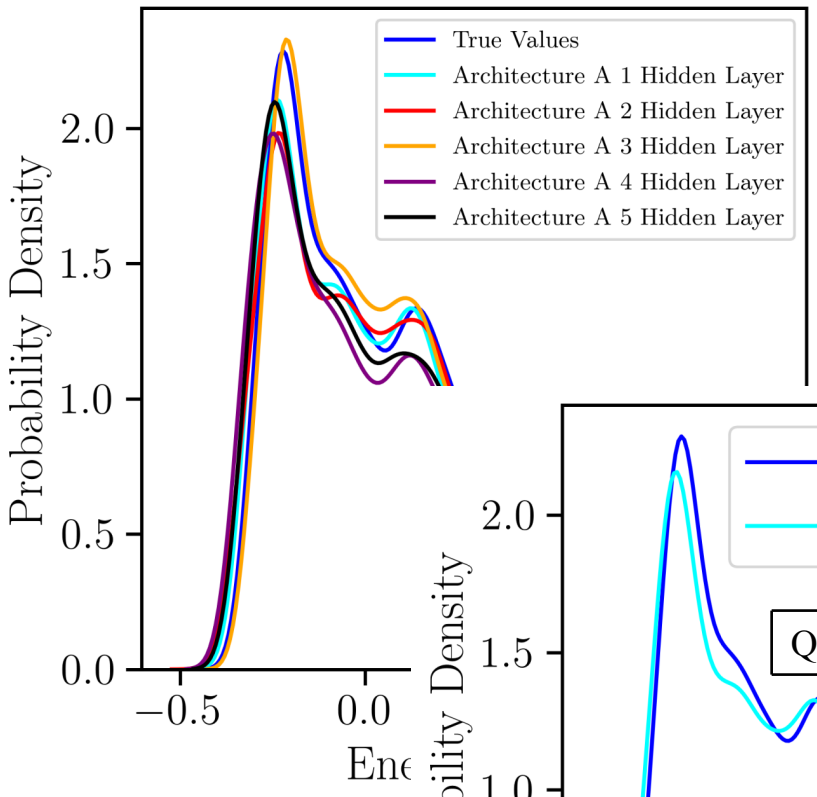
Quadratic SNAP: $\square^{8}_{(\#\,A\%}(r^{\#}) = \# \cdot B^{8} + \frac{1}{2} B^{8} \cdot '' \cdot B^{8}$

Back to Physics

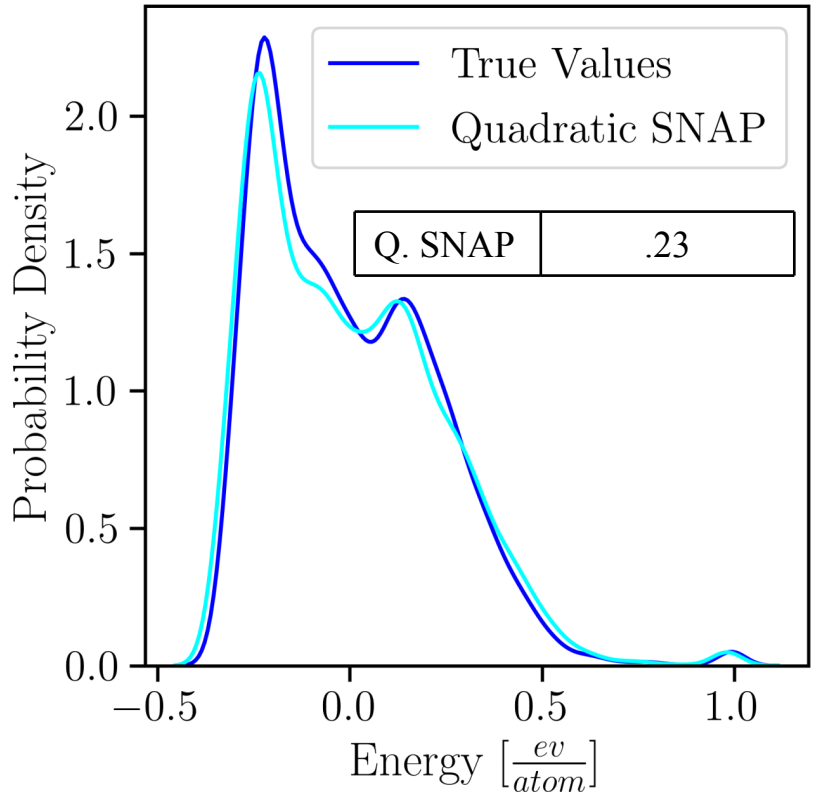
Physical training, or Entropy Maximized

- Comparing the entropy maximized training to the hand constructed W training of Wood *et. al.* PRB 2019

Description	ΔE(eV/atom)	ΔF(eV/Å)
W:		
Elastic deform	5.3×10^{-2}	0.0×10^{0}
Equation of state	1.4×10^{-1}	4.0×10^{-5}
DFT-MD	5.3×10^{-2}	6.0×10^{-2}
Surfaces	3.4×10^{-2}	2.8×10^{-1}
Self-interstitials	4.6×10^{-2}	9.5×10^{-2}
Liquids	$\mathbf{2.9 \times 10^{-1}}$	$\mathbf{4.8 \times 10^{-1}}$
Dislocations	5.0×10^{-2}	7.8×10^{-2}
Monovacancy	4.2×10^{-2}	9.8×10^{-2}
Divacancy	2.9×10^{-2}	8.7×10^{-2}
Γ surface	4.6×10^{-2}	2.5×10^{-1}
Γ surf.+vacancy	4.3×10^{-2}	1.7×10^{-1}

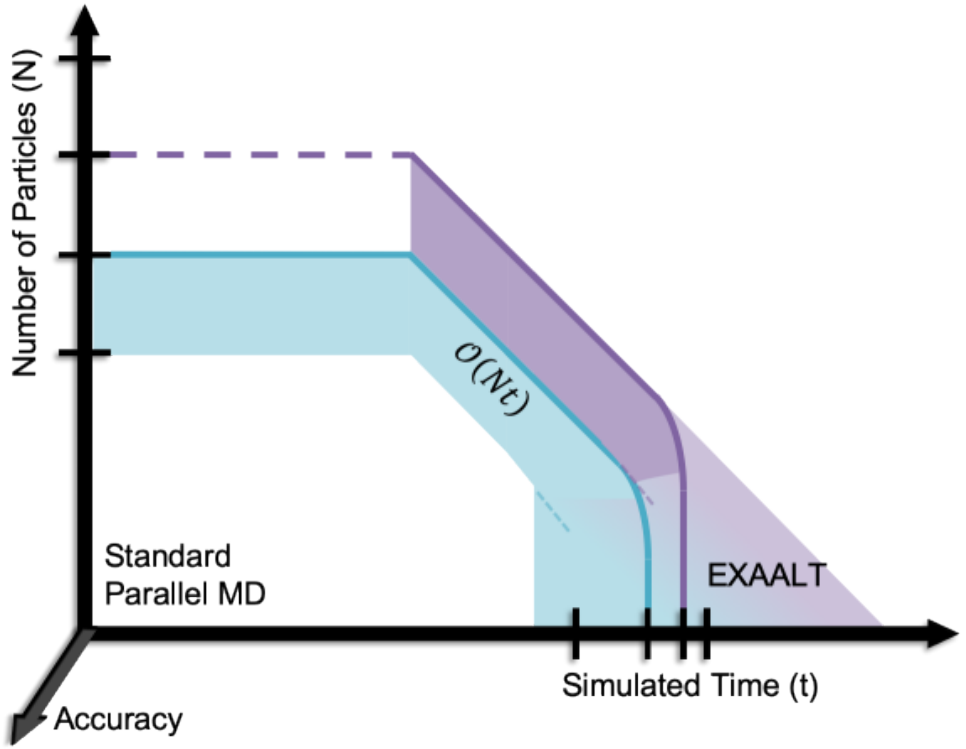


Arch.	RMSE $[\frac{ev}{atom}]$
1 HL	.031
2 HL	.043
3 HL	.017
4 HL	.042
5 HL	.033



Conclusions and Path Forward

- The EXAALT project is ensuring Exascale-ready MD software beyond the length, time-scales of standard MD
- While harder to quantify, the fidelity of our MD simulations needs to be a key consideration at the Exascale



- Data-driven interatomic potentials (SNAP, SNAP-NN) allow for MD predictions of challenging material problems.

Contact Information:

mitwood@sandia.gov

ECP EXASCALE COMPUTING PROJECT
Los Alamos NATIONAL LABORATORY EST. 1943
Sandia National Laboratories
NVIDIA
U.S. DEPARTMENT OF ENERGY | Office of Science
OAK RIDGE National Laboratory
NERSC

References and Repositories

Core Algorithms Papers

Plimpton, Steve. Fast parallel algorithms for short-range molecular dynamics. No. SAND-91-1144. Sandia National Labs., (1993).
Plimpton, Steven J., and Aidan P. Thompson. "Computational aspects of many-body potentials." *MRS bulletin* 37.5 (2012): 513-521.
Le Bris, Lelievre, Luskin, and Perez, *MCMA* 18, 119 (2012)
Perez, Cubuk, Waterland, Kaxiras, Voter, *JCTC* 12, 18 (2016)
Niklasson & Cawkwell JCP 141,164123 (2014)
Niklasson JCP 054103 (2017)

Impressive Particle Method Examples

L A Zepeda-Ruiz *et al. Nature* **550**, 492–495 (2017) doi:10.1038/nature23472

Glotzer, Sharon C., and Michael J. Solomon. "Anisotropy of building blocks and their assembly into complex structures." *Nature materials* 6.8 (2007): 557-562.

K. Shimamura et al., “Hydrogen-on-Demand Using Metallic Alloy Nanoparticles in Water,” *Nano Letters*, vol. 14, no. 7,2014, pp. 4090–4096

Mattox, Timothy I., et al. "Highly scalable discrete-particle simulations with novel coarse-graining: accessing the microscale." *Molecular Physics* 116.15-16 (2018): 2061-2069.

Uses of ParSplice/EXAALT

Defect evolution in fusion materials (w. Luis Sandoval, Blas Uberuaga, Art Voter). Up to 100,000 cores, ~10,000 atoms on ms [Sci. Rep. 7, 2522 (2017)]

Jogs in nickel (w. Lauren Smith, Tom Swinburne, Dallas Trinkle), ~1000 cores, ~10,000 atoms, tens of ms

Cation defect evolution in pyrochlores (w. Romain Perriot, Blas Uberuaga, Art Voter), ~200 cores, ~1000 atoms, tens of ms [Nature Comm., 8, 681 (2017)]

Shape evolution of metallic nanoparticles (w. Rao Huang, Art Voter). ~1000 cores, ~100 atoms, ms [JCP 147, 152717 (2017). JMR (in press)]

https://gitlab.com/exaalt

https://github.com/FitSNAP/FitSNAP

https://github.com/materialsvirtuallab/mlearn

Contact Information:

mitwood@sandia.gov
danny_perez@lanl.gov

EXASCALE COMPUTING PROJECT

Los Alamos NATIONAL LABORATORY EST. 1943

Sandia National Laboratories

NVIDIA

U.S. DEPARTMENT OF ENERGY | Office of Science

OAK RIDGE National Laboratory

NERSC