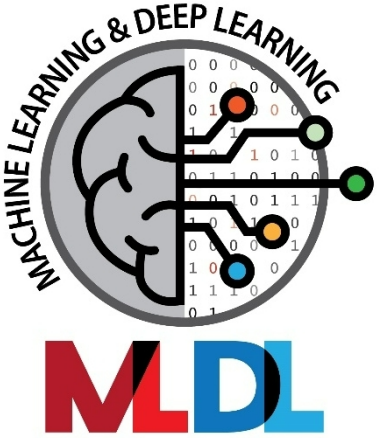
SAND2022-8952C



MLDL

Machine Learning and Deep Learning Conference 2022

Building a new generation of multiscale materials models with machine-learned interatomic potentials

- Megan McCarthy, Department 1444 – Computational Multiscale
- Mitchell Wood 1444, Aidan Thompson, 1444

- Funding: DOE, Office of Science, Fusion Energy Sciences

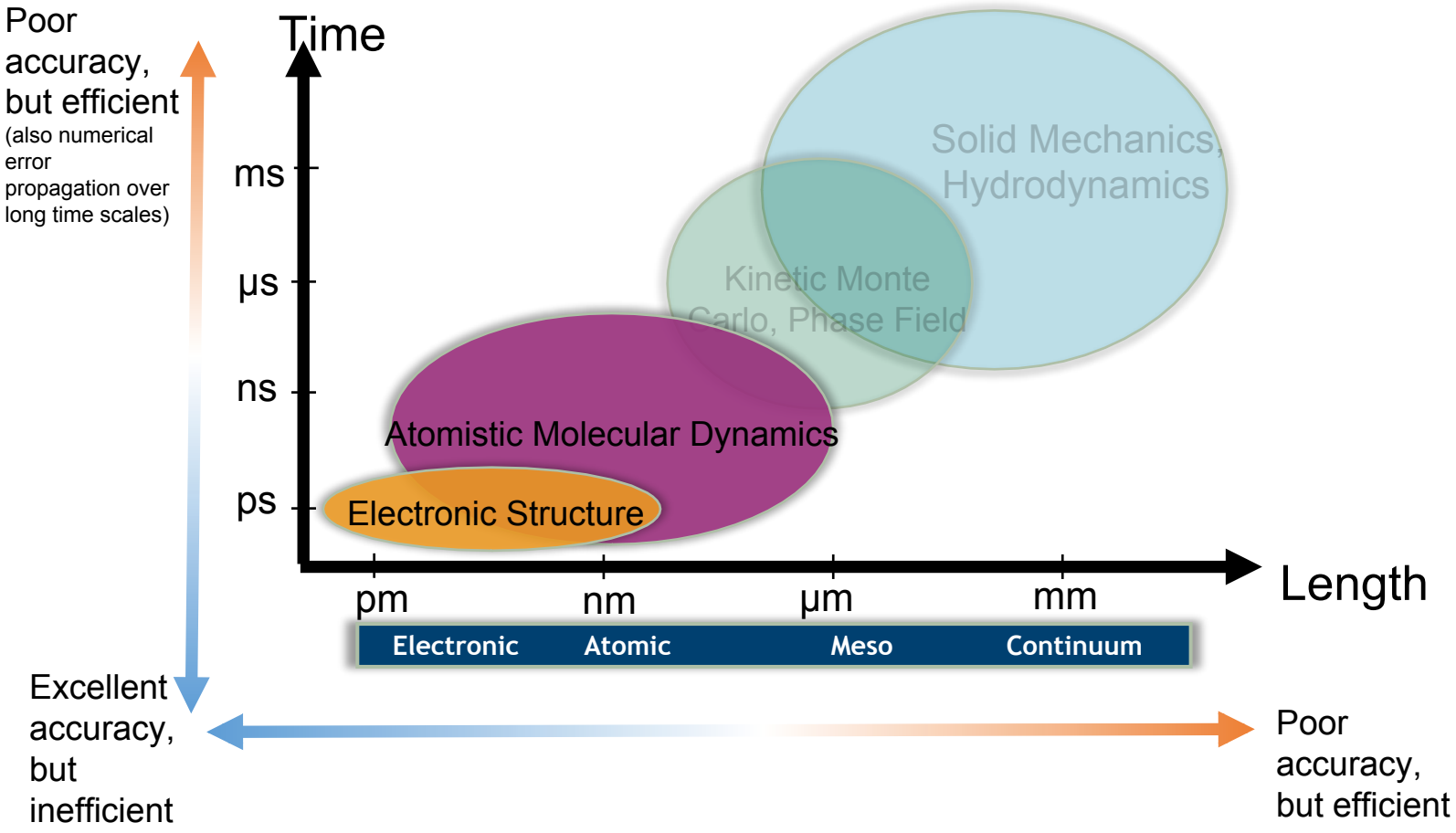
Abstract

Building a new generation of multiscale materials models with machine-learned interatomic potentials

Excellent physical models exist to simulate material behavior at specific length and time scales. However, many challenges still remain in linking materials behavior *across* scales. Applying the techniques of machine learning to the issue of multiscale simulation has led to a revolution in atomistic materials modeling. Machine-learned interatomic potentials (MLIAPs) connect extremely accurate but computationally expensive quantum modeling techniques with lower fidelity but fast classical molecular dynamics (MD) models. In other words, MLIAPs allow MD simulations to achieve quantum accuracy at unprecedented system sizes and time scales. The diversity of model forms and feature sets for MLIAPs, paired with popular MD codes such as LAMMPS, has also opened new pathways to accelerating MLIAP development. These developments have enabled significant progress to be made in challenging scientific and engineering areas, including modeling conditions inside future fusion reactors, understanding magnetic materials in extreme environments, and accelerated exploration of complex alloy systems.

Though MLIAPs represent a new paradigm in the atomistic modeling community, the traditional problems in ML/DL regarding retaining accuracy while extrapolating remain. This talk will overview our group's strategies to address these issues using the open-source and user-friendly FitSNAP MLIAP generation code. We will explore how current implementations of active learning, uncertainty quantification, and automated training set generation are systematically improving MLIAP simulation accuracy and extrapolative capabilities.

Physical models of materials are often constrained to limited length- and time-scale regimes.

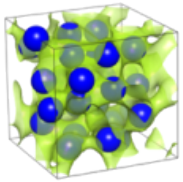


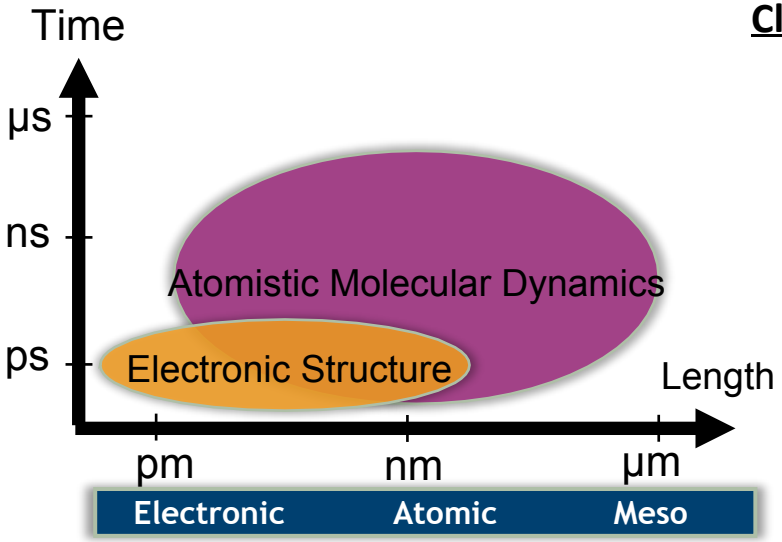
Accuracy/computational speed tradeoff in simulation

Machine learning is being used to successfully bridge quantum - and atomic-scale modeling and simulation.

Quantum Molecular Dynamics

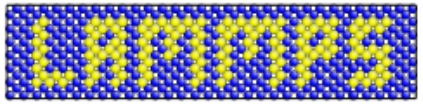
- Input: ion (atom) positions
- Output: **electronic structure**, energy, forces, stress
- **Immense predictive power**
- **Expensive**
- **O(N^3) scaling**
- **N ~ 100**



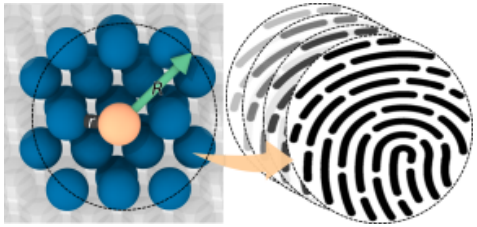


Classical Molecular Dynamics

- Input: atom positions, **interatomic potential**
- Output: energy, forces, stress
- **O(N) scaling**
- **N ~ millions, billions**
- **Accuracy is a problem**



Machine-Learned Interatomic Potential (MLIAP)

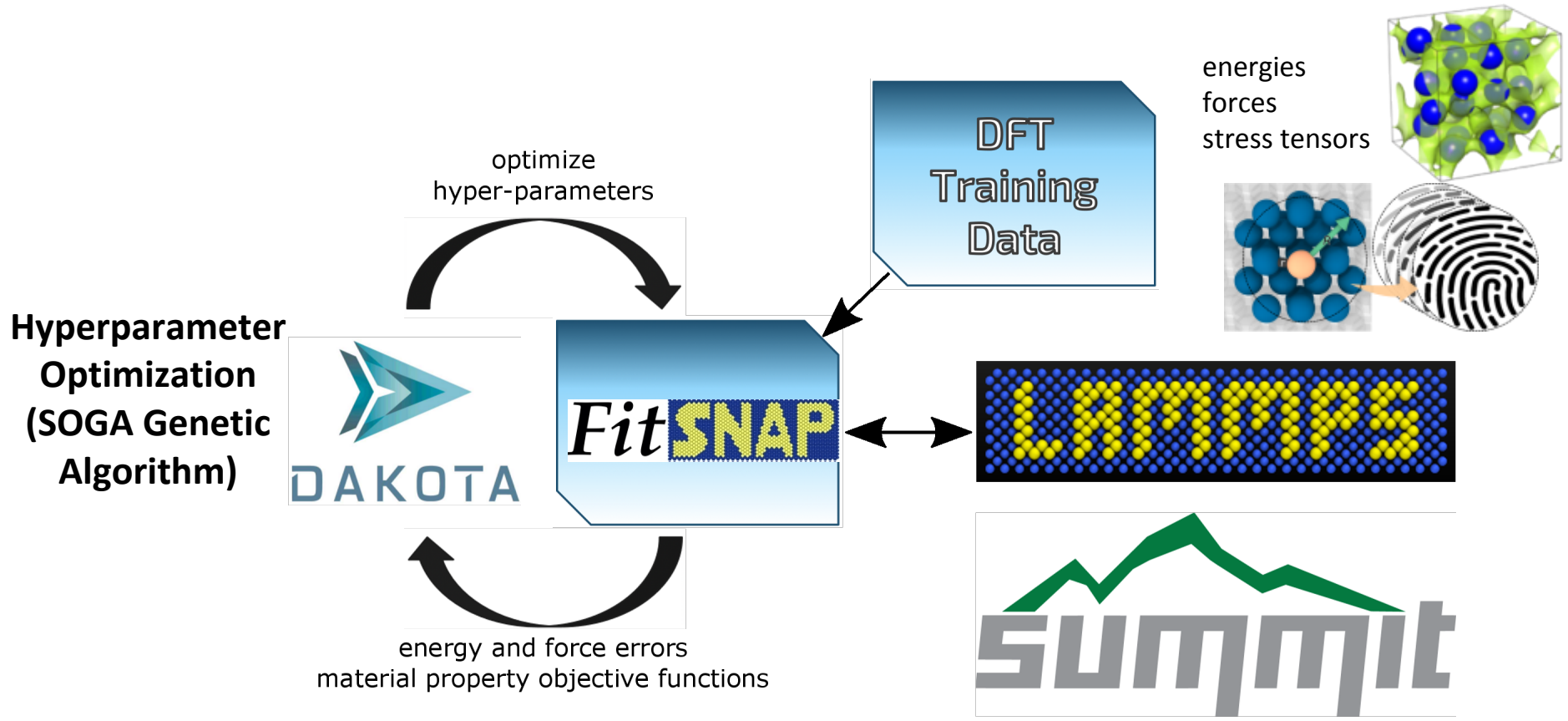


Ground truth

Reduced order model

$Energy_i = U(features)$

SNAP models are one means of developing MLIAPs for a variety of materials.



The underlying SNAP model form can be varied for different target applications and desired accuracy.

Linear SNAP
(weighted density, baseline)

Model Form

Energy of atom i expressed as a basis expansion over K components of the bispectrum (B_k^i)

$$E^i_{SNAP} = \beta \cdot \mathbf{B}^i + \frac{1}{2}(\mathbf{B}^i)^T \cdot \alpha \cdot \mathbf{B}^i$$

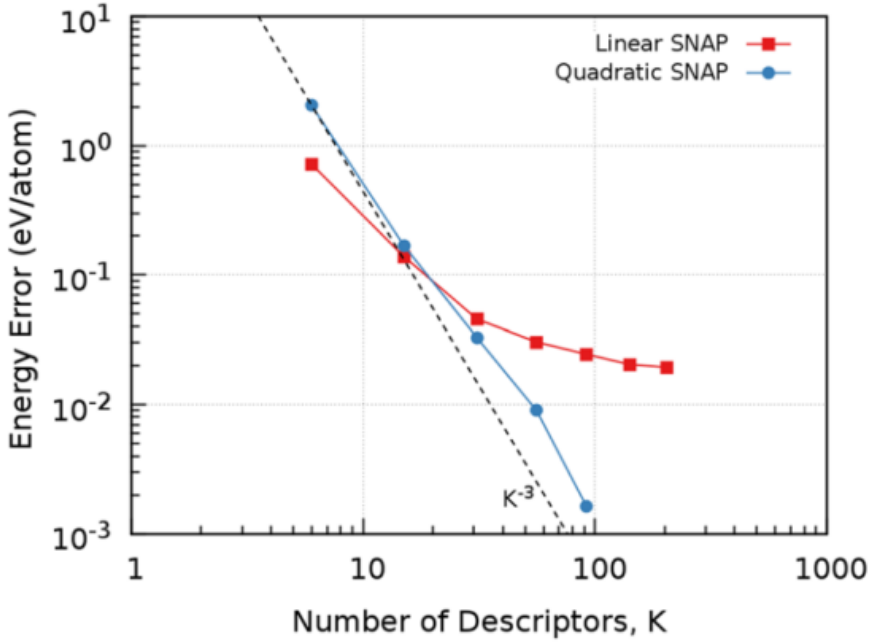
Regression Method

β vector fully describes a SNAP potential
Decouples MD speed from training set size

$$\min(||\mathbf{w} \cdot D\beta - T||^2 - \gamma_n \, ||\beta||^n)$$

Weights — Set of Descriptors — DFT Training — Regularization Penalty

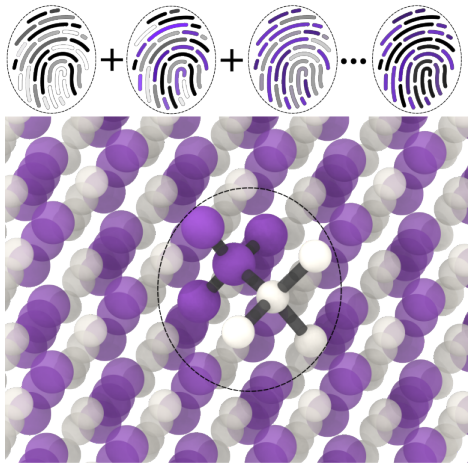
Quadratic SNAP



$$E^i_{SNAP}(\mathbf{r}^N) = \beta \cdot \mathbf{B}^i + \frac{1}{2}F''(\mathbf{a} \cdot \mathbf{B}^i)^2,$$
$$= \beta \cdot \mathbf{B}^i + \frac{1}{2}(\mathbf{B}^i)^T \cdot \alpha \cdot \mathbf{B}^i$$

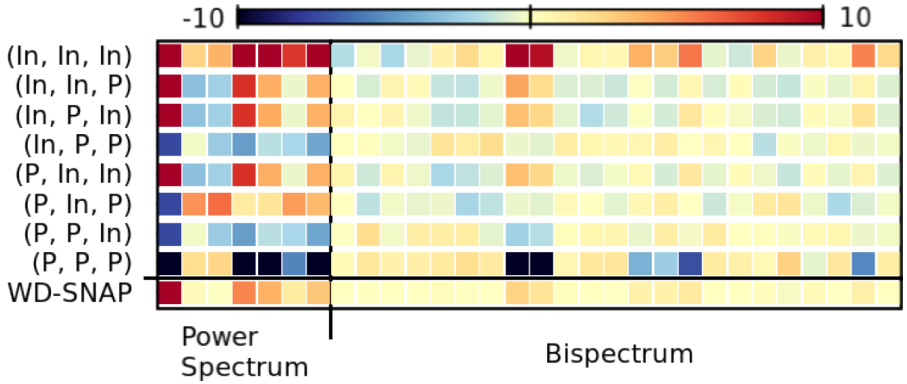
The underlying SNAP model form can be varied for different target applications and desired accuracy.

ChemSNAP



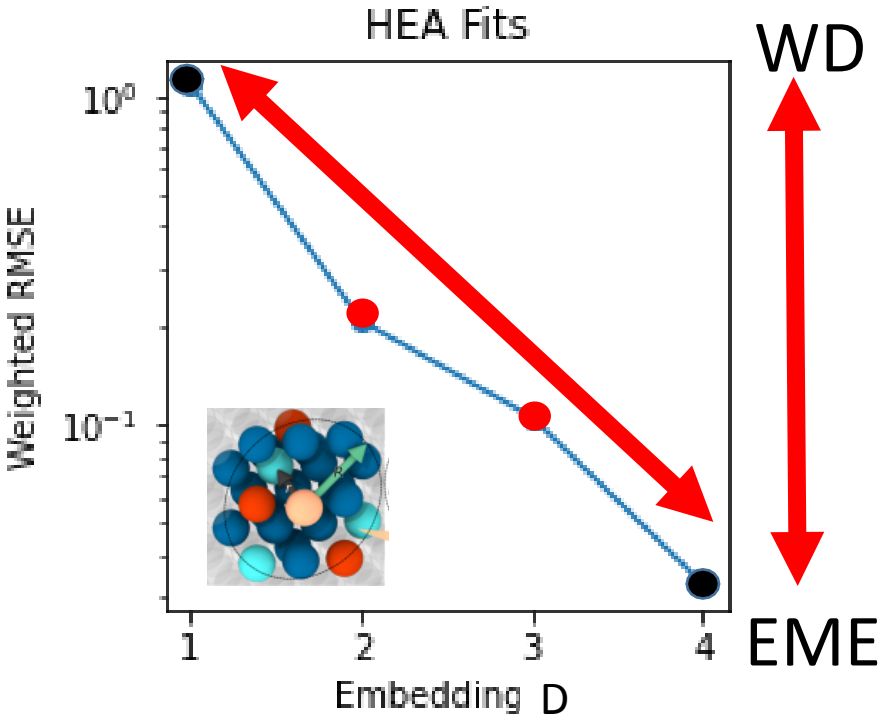
Sensitivity to Chemical Change

Antisite versus Bulk Zincblende



$$B_{j_1 j_2 j} = \sum_{\beta=1}^{N_{elem}} \sum_{\gamma=1}^{N_{elem}} \sum_{\delta=1}^{N_{elem}} B^{\beta\gamma\delta}_{j_1 j_2 j}$$

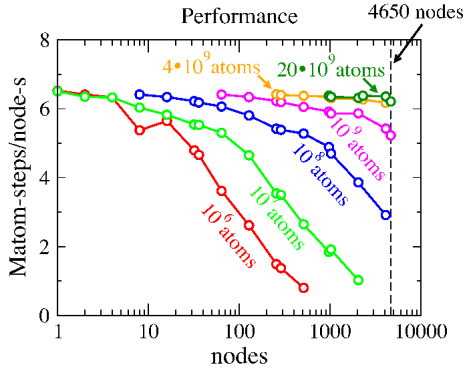
ChEmbed SNAP



Increases computational speed of ChemSNAP with some accuracy tradeoff

MLIAPs in general, and SNAP specifically, allow MD simulations to achieve quantum accuracy at unprecedented system sizes and time scales.

- Gordon Bell Prize Finalist, 2021
- Team from USF, Sandia, NERSC, NVIDIA, KTH : doi.org/10.1145/3458817.3487400
- **2.6 billion** atom diamond sample, $\mathbf{0.5 \times 1.5\ \mu m}$
- Shock wave in <110> direction initiated by piston, $v_p = 7\ km/s$.



Applications of SNAP models (SNAPplications!)

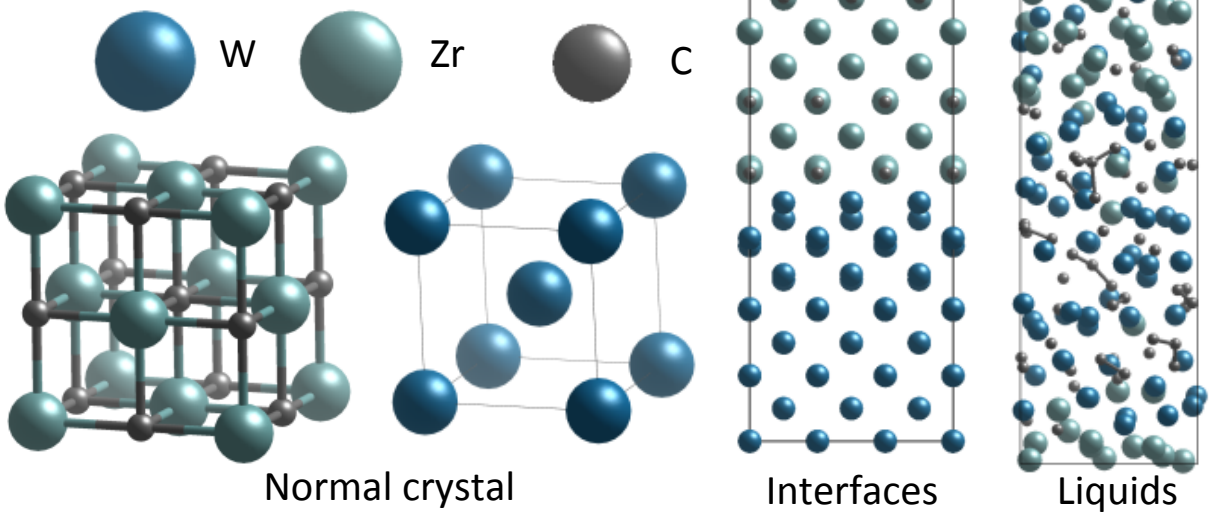
System	Year	Usage	N_{DoF}	$N_{Training}$	Descriptors
Ta	2014	Dislocation motion	31	363	Linear
InP	2015	Radiation damage, defe	31	665	Linear
WBeHe	2017	Plasma facing materials	56	25,052	Linear
Mo	2017	Phase diagram predictio	31	1000	Linear
Actinides	2018	Shock, phase transitions	56	20,000	Quadratic
NiMo	2018	Phase diagram predictio	31	2,000	Linear
LiN	2019	Super-Ionic Conductor	31	3,000	Lin+Charge
Various	2020	Accuracy/Cost comparis	10-130	1,000	Lin, Quad
InP	2020	Radiation damage, defe	241	1,000	EME
AlNbTi	2020	High entropy alloy desig	1596	7,250	Quadratic
Si	2020	Neural network SNAP	1596	>5,000	NN
NbMoTaW	2021	HEA alloy mechanics	124	~3,00	Linear

Applications of SNAP models (SNAPplications!)

System	Year	Usage	N_{DoF}	$N_{Training}$	Descriptors
A	2021	Predicting electron dens…	91	30	NN
Fe	2021	Magnetic phase transitic…	1596	683	Quad+Spin
WBeHN	-	Plasma facing materials	56*	>40,000	Linear
WZrC	-	Plasma facing materials	56*	>40,000	Linear
C	2022	Planetary impacts, shocl…	1596	30,000	Quadratic
C, V	2021	Metal plasmas	1596	10,000	Quadratic
MoNbTaTi	-	HEA alloy design	-	>5,000	EME
GeSe	-	Vitrification	-	>5,000	EME
LiMoS	-	Li-ion batteries	-	>5,000	-
SiGeSnPb	-	Thermoelectric materials	-	>5,000	-
W	2022	Model form selection	-	330,000	NN

Developing and curating MLIAP training sets is one of the field's biggest challenges.

Physics-informed training



Up to now:
Curated training sets created by hand, using physical intuition

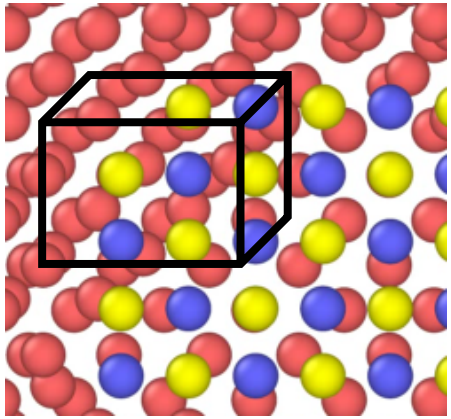
Works very well with targeted materials systems

Loses transferability, not easy to automate, need an expert

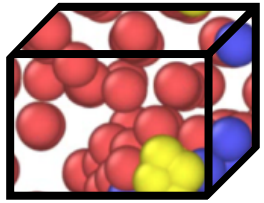
To improve extrapolative capability, 'iterative learning' with humans-in-the-loop has been used with success.

Hands-on Route:

Run MD



Carve out structures



Run DFT

Re-Train ML-IAP

optimize hyper-parameters

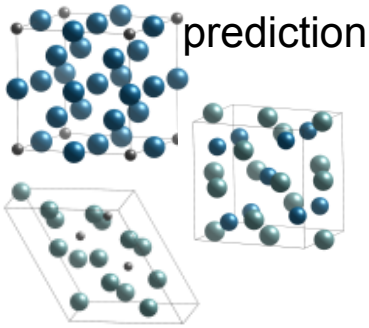
DAKOTA

FitSNAP

energy and force errors
material property objective functions

Automated (kind of):

Structure prediction from DFT



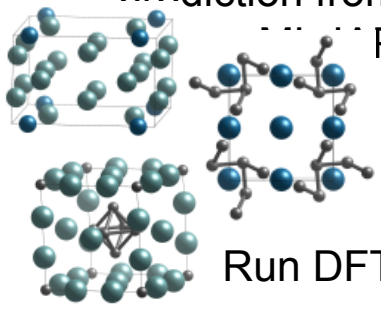
optimize hyper-parameters

DAKOTA

FitSNAP

energy and force errors
material property objective functions

Structure prediction from ML-IAP



Run DFT

USPEX : A. Oganov et al. / J. Chem. Phys. **124,** 244704 (

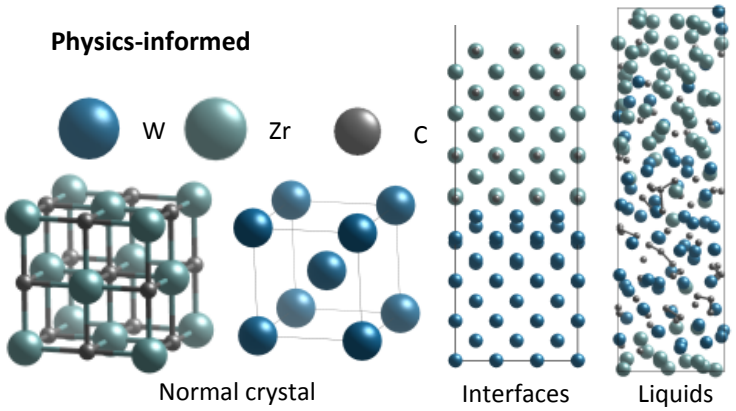
Automating training set creation and curation is crucial to developing more efficient and transferable MLIAPs.

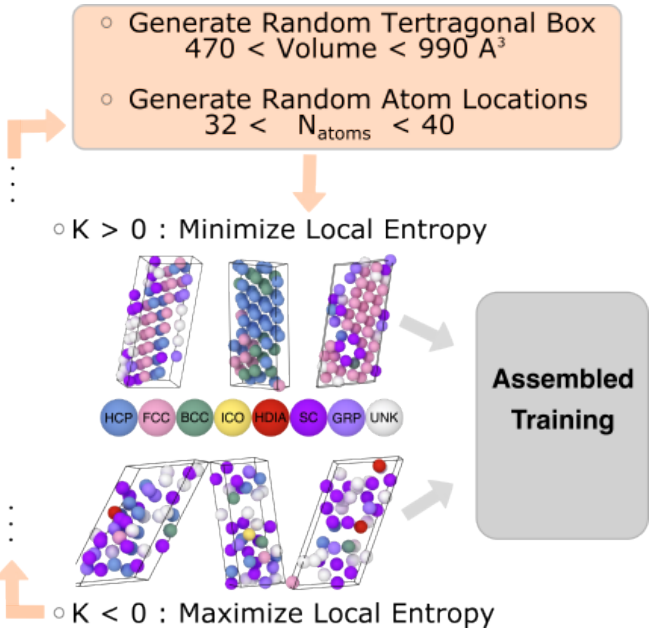
Up to now:
Curated training sets created by hand, using physical intuition

Works very well with targeted materials systems

Loses transferability, not easy to automate, need an expert

Physics-informed



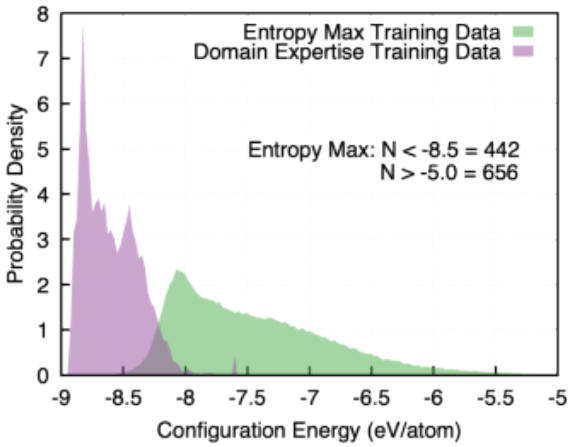


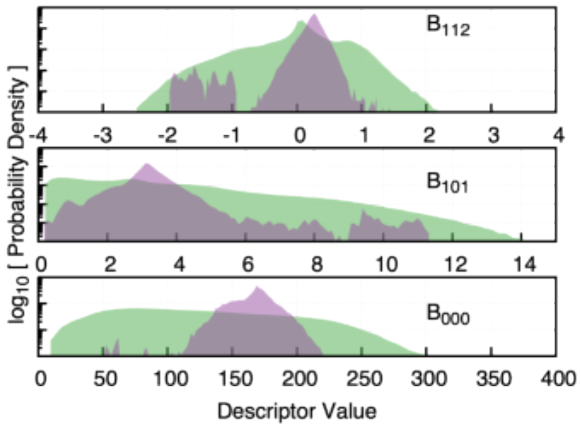
Karabin and Perez, JCP 153, 094110 (2020)

New techiques in development:
Information entropy maximization (right & bottom), active learning using uncertainty quantification (next slides)

No expert needed, shown to be transferable, what is needed created 'on the fly'

Not easily interpretable, groundwork still being laid, handshaking between software (LAMMPS, FitSNAP, UQ)

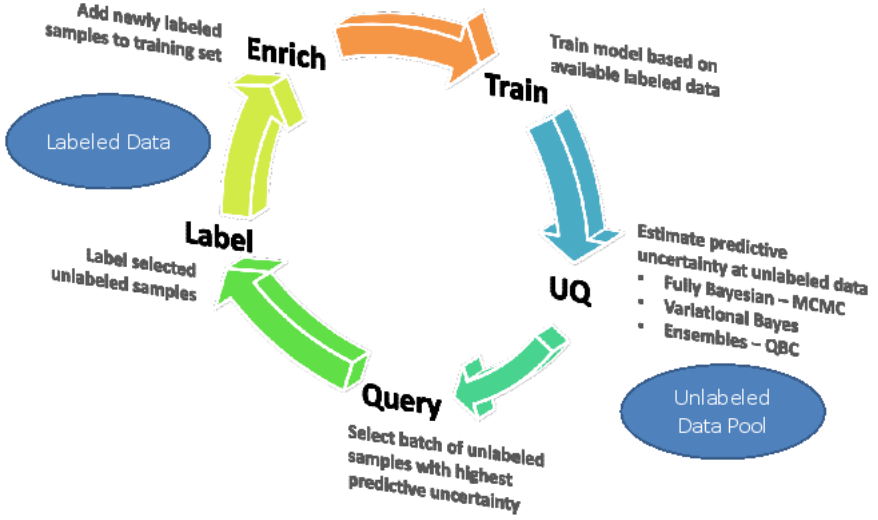


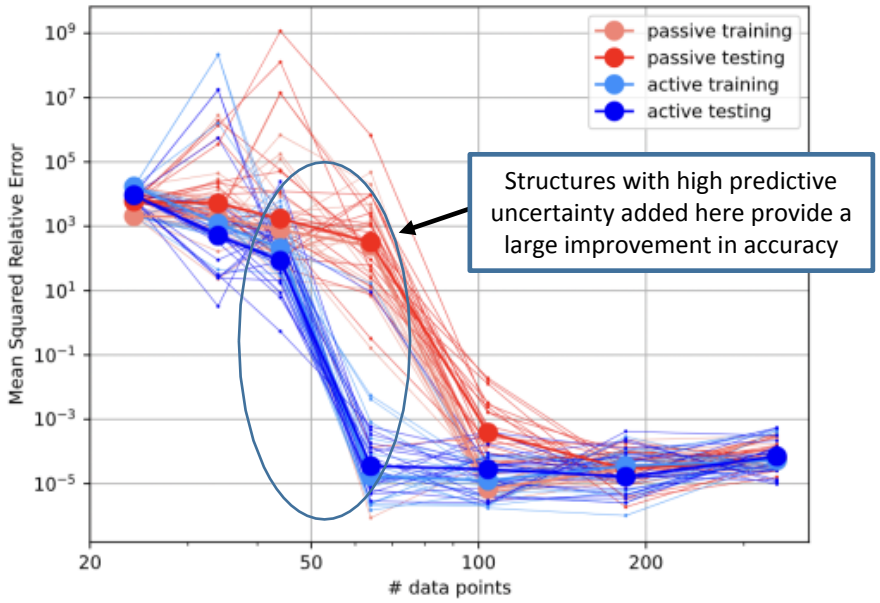


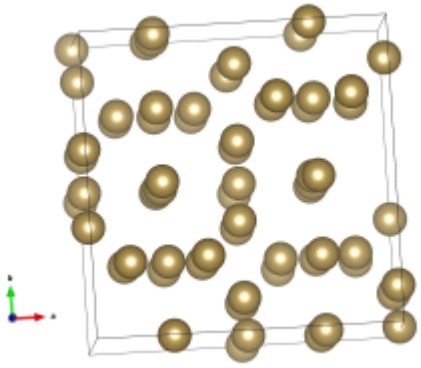
arXiv:2201.09829v2 (de Oca Zapiain, Pereyra, Lubbers, Perez, Thompson)

Active learning paired with uncertainty quantification can vastly improve the efficiency of training accurate MLIAPs.

Pool-based, uncertainty-informed active learning paradigm



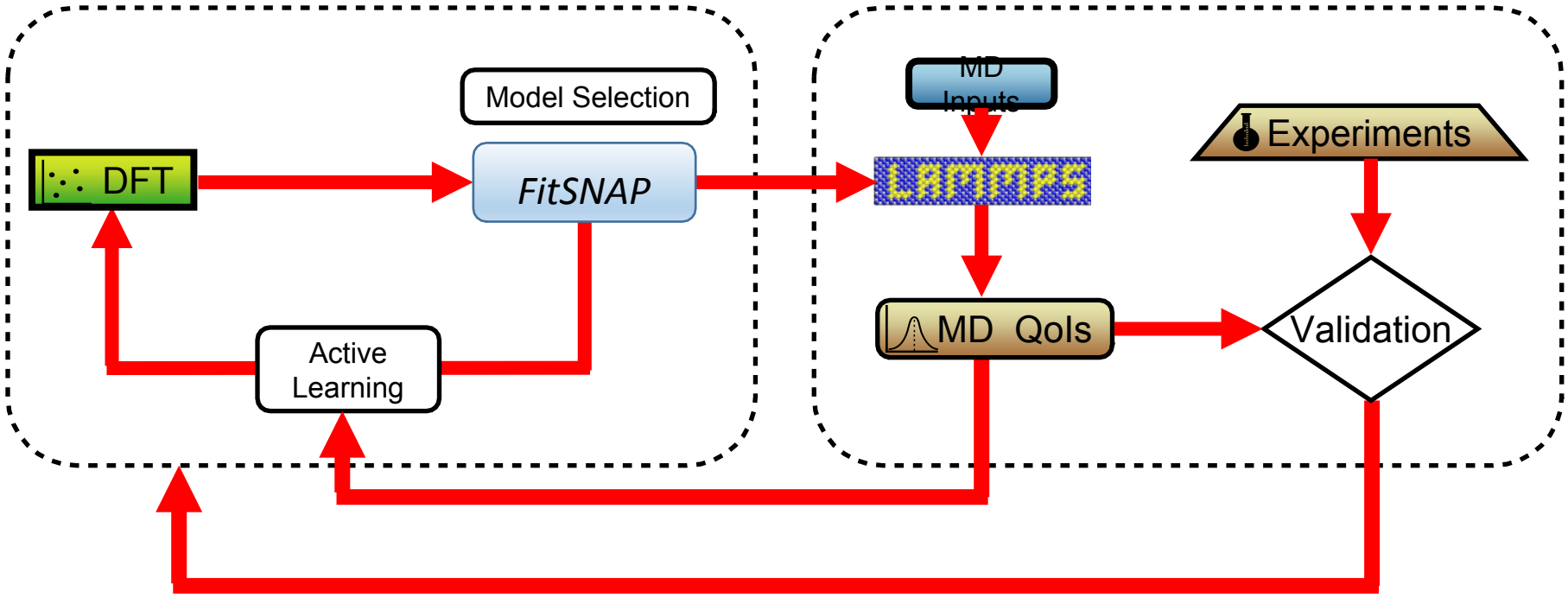




Representative structure added (at ~50 data points) with large uncertainty

Williams, Sargsyan, **Najm**, SNL

Active learning is being formally incorporated into the MLIAP development workflow.



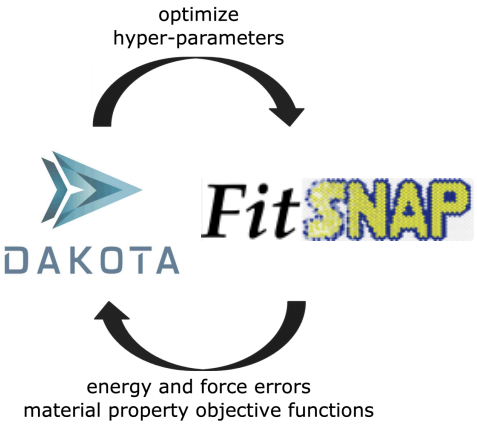
Williams, Sargsyan, **Najm**, SNL

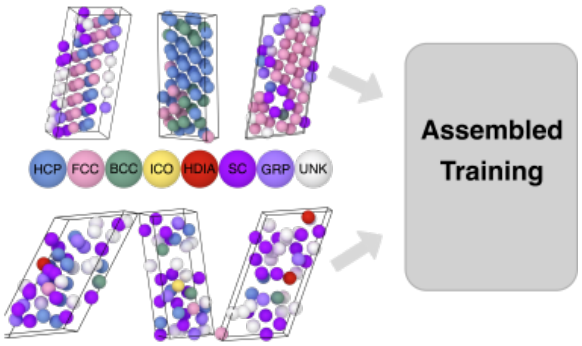
- Extend current AL loop to span MD predictions – goal oriented AL
- Major challenge is that the extensive computational costs of forward UQ in MD are now internal to the AL loop. A significant increase in AL costs.
- Use experimental data on MD observables to inform both data & model selection

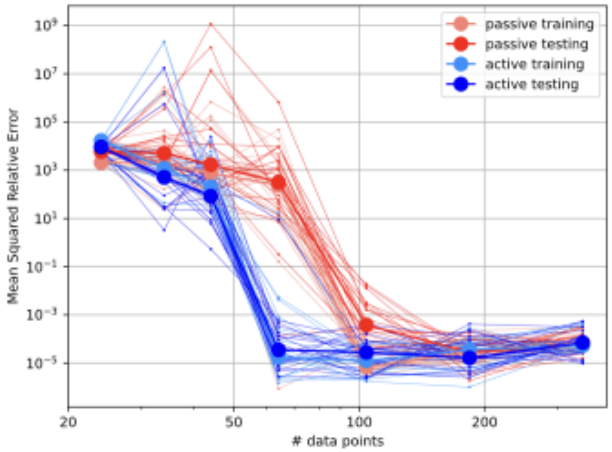
Conclusion

New developments in the field of machine-learned interatomic potentials will allow scientists to build materials models with more accuracy and efficiency.

This in turn will enable scientists to explore materials behavior (new and old!) at the nanoscale with more trust in the results and with higher throughput.







Sandia National Laboratories

Los Alamos NATIONAL LABORATORY EST. 1943

U.S. DEPARTMENT OF ENERGY | Office of Science