SAND2022-2445C

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

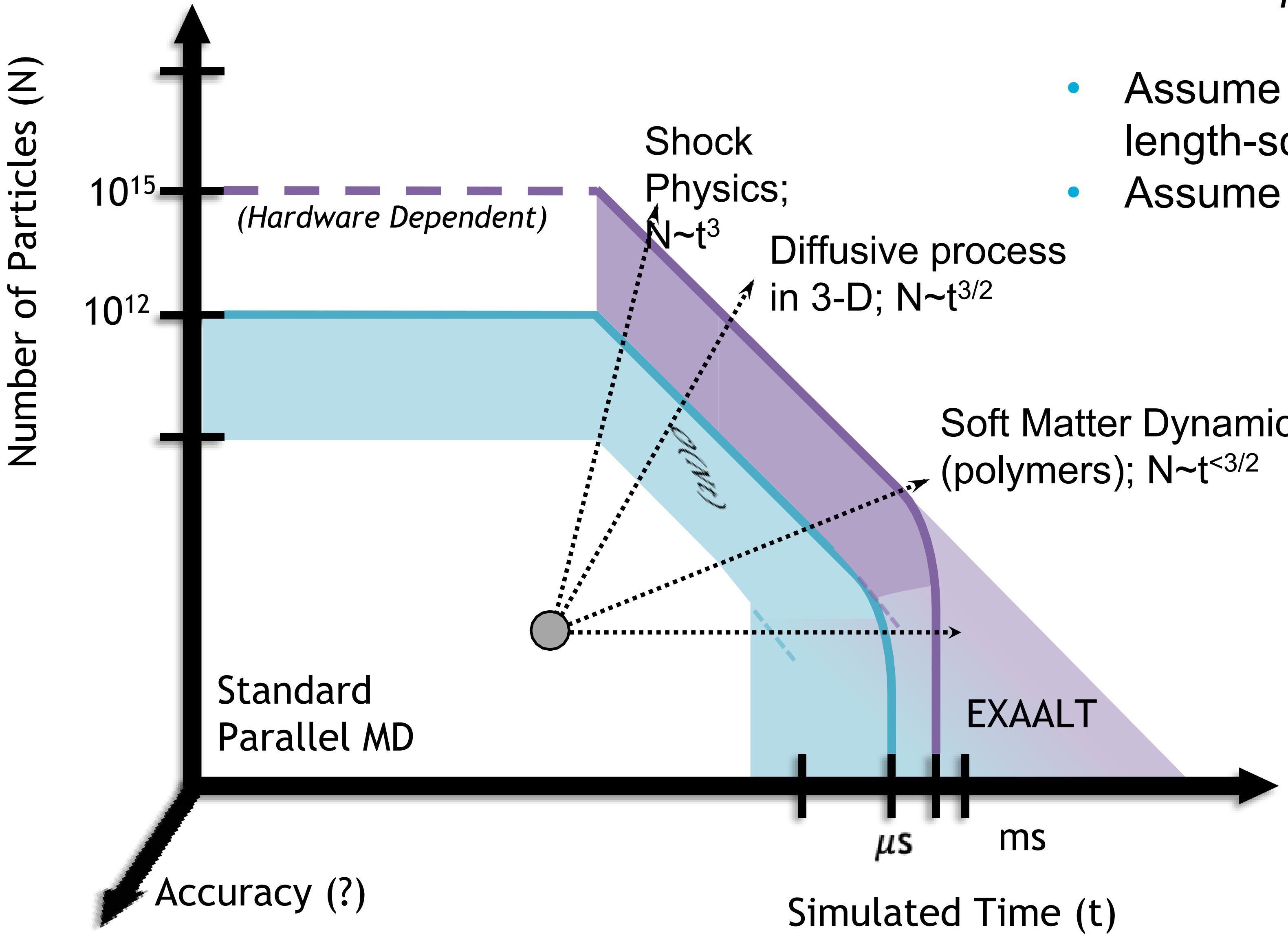
Mitchell Wood, Mary Alice Cusentino, Aidan Thompson
Sandia National Labs, Albuquerque, New Mexico, USA
Ivan Oleynik
University of South Florida, Tampa, Florida, USA

TMS Meeting, March 14th, 2022

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 Securit
under contract DE-NA0003525.

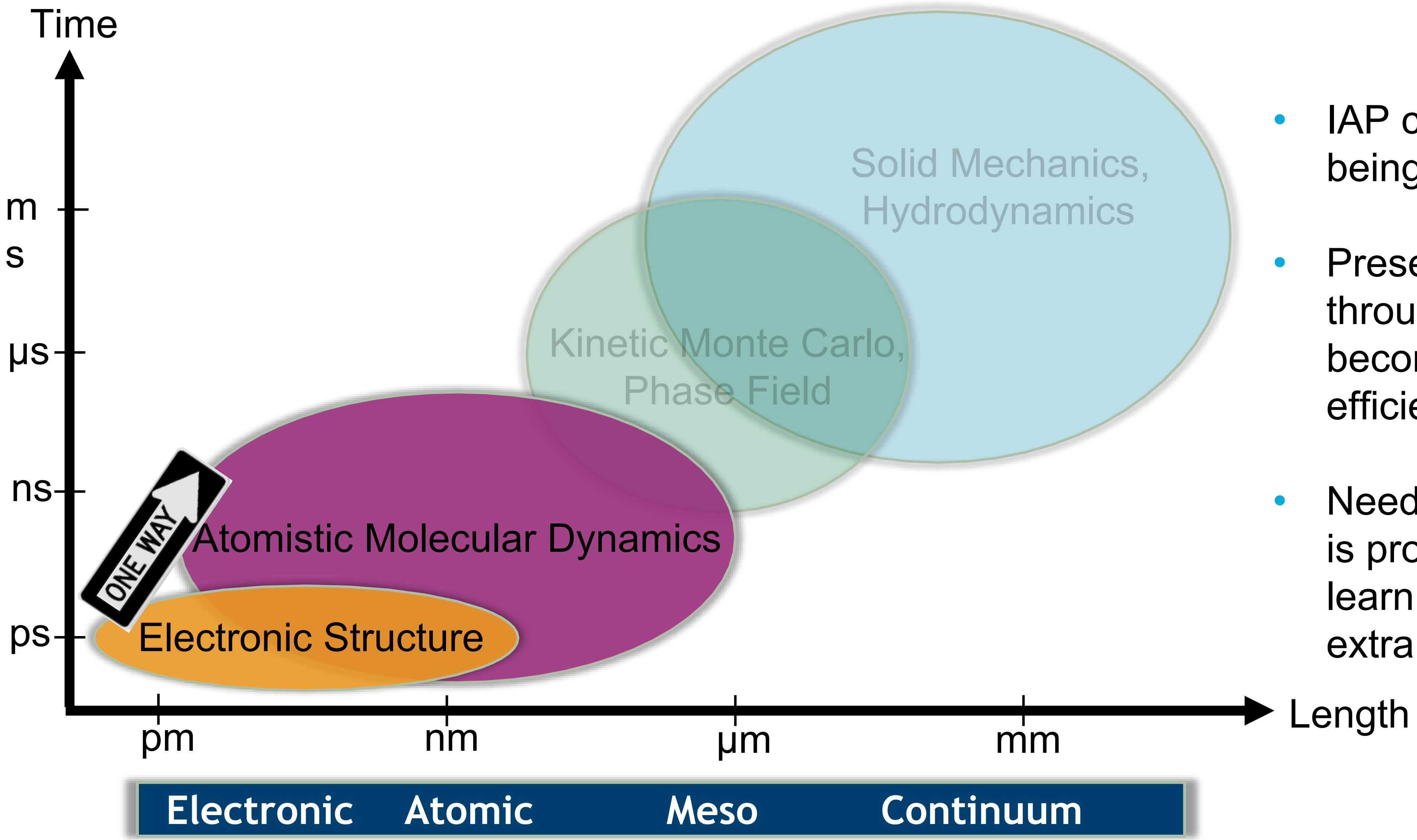
Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

The Master Plot



- *What is possible for MD at the Exascale?*
- Assume your problem has some specified length-scale dependence : N~L^{α}
- Assume the associated time-scale goes as t~L^{γ}
- But what if you care about a system governed by rare event dynamics?

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

Environment of Machine Learning Techniques

Descriptors	Training Technique / Model Form
• Bond distances, angles [Behler] • Moment tensors [Shapeev] • Bispectrum components [Cysani, Thompson] • Smooth overlap of atomic positions [Cysani] • Atomic cluster expansion [Drautz] …	• Linear Regression • Stochastic Gradient Descent • Gaussian Process Regression …

Data Needs:

- When $N_{DOF} \sim N_{Train}$, high risk of overfitting → Poor Interpolation
- When training diversity is low → Poor Extrapolation
- Running MD will expose these short comings (energy drift, instabilities, unphysical behavior)

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, McCarthy	-	>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 extrapolatic

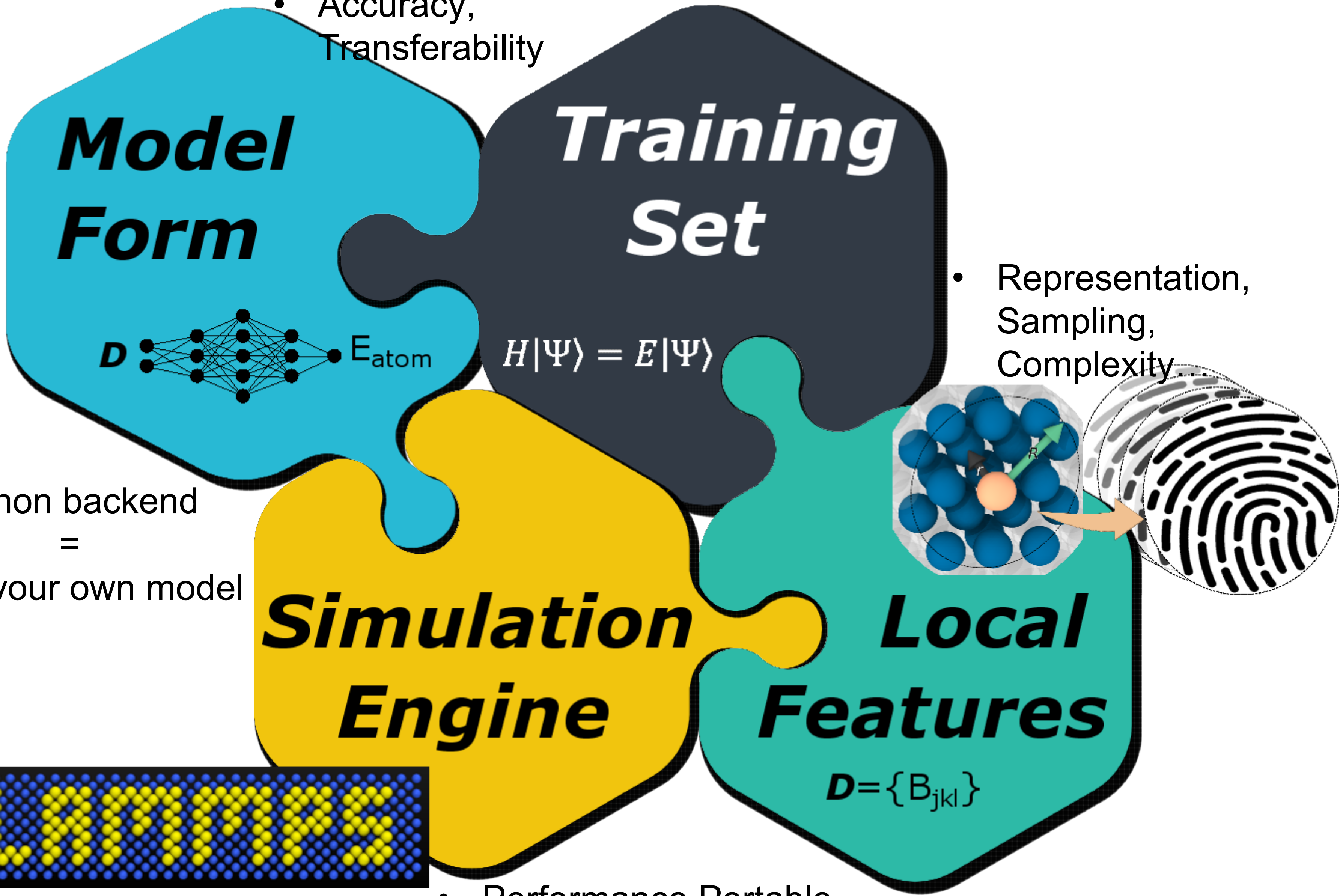
- Growing evidence that SNAP is a *general use* material model form, unlike any interatomic potential used in MD to date
- SNAP model training software now incorporated in [Materials Design Inc.](#) products

Components of ML-MD

https://github.com/FitSNAP/FitSNAP

http://lammps.sandia.gov

- Accuracy, Transferability



- Representation, Sampling, Complexity…

- Python backend = Bring your own model

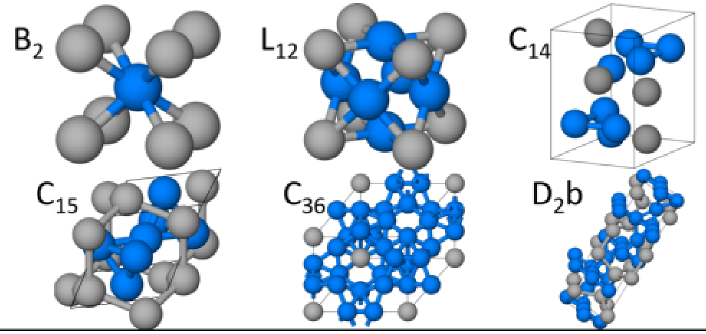
LAMMPS

- Performance Portable

Assembling a Better Training Set

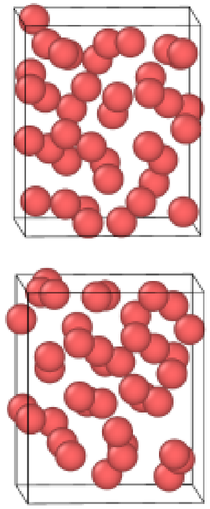
Description	N_E	N_F	σ_E	σ_F
W-Be:				
Elastic Deform†	3946	68040	$\mathbf{3 \cdot 10^5}$	$2 \cdot 10^3$
Equation of State†	1113	39627	$2 \cdot 10^5$	$4 \cdot 10^4$
DFT-MD†	3360	497124	$7 \cdot 10^4$	$6 \cdot 10^2$
Surface Adhesion	381	112527	$2 \cdot 10^4$	$\mathbf{9 \cdot 10^4}$

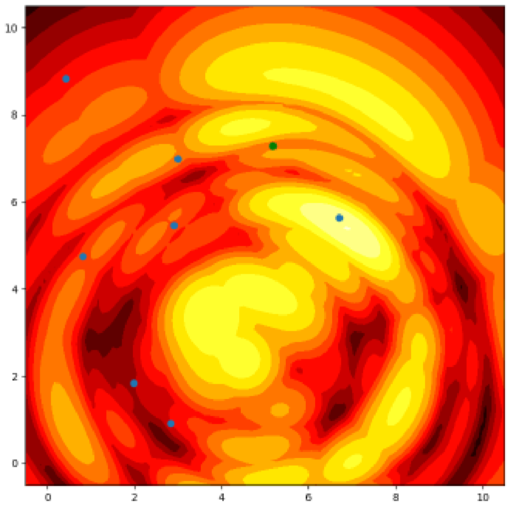
† Multiple crystal phases included in this group:



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

Pseudo-potential · Cross-entropy · Self-entropy





Domain Expertise Training

- Use cases for the potential are known, run DFT on representative configurations
- Intrinsically biased to a small region of configuration space

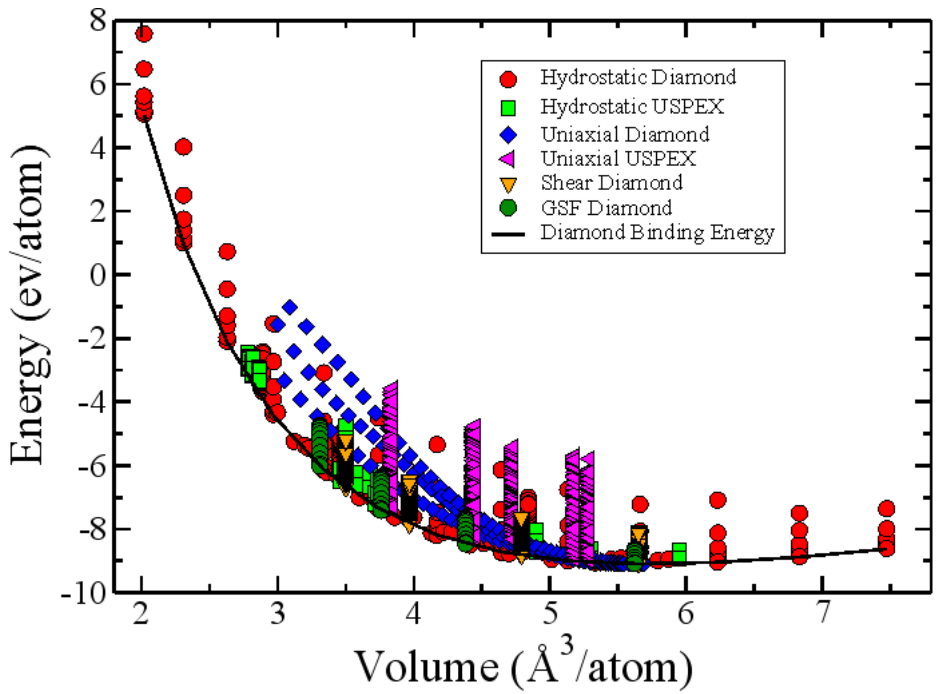
Maximizing Diversity

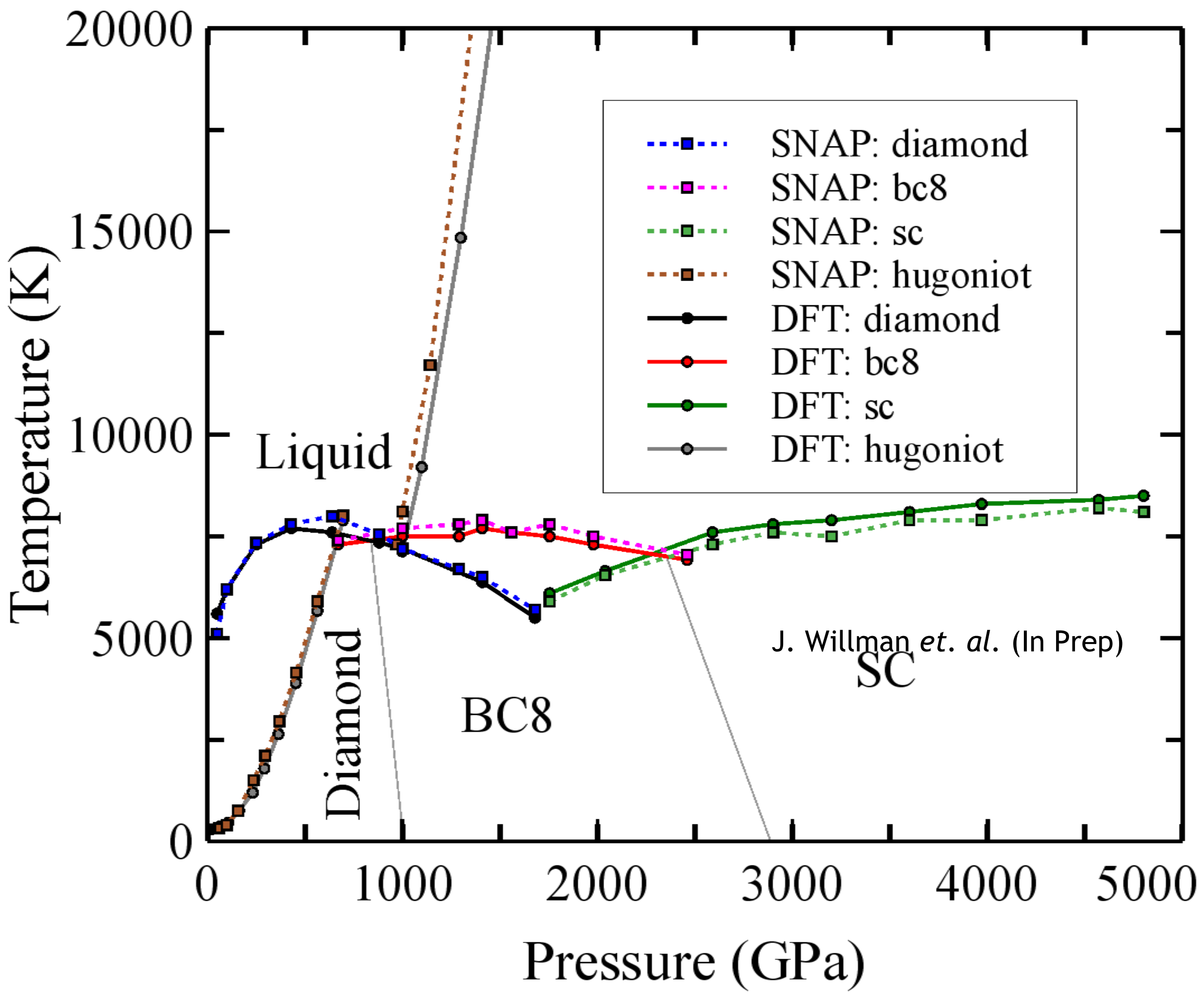
- Framework of time acceleration tools can generate new training by running MD with lots of replicas
- (above) Self-entropy landscape of the average interatomic distance

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, 0-15,000K reproduced because it was trained to do so.

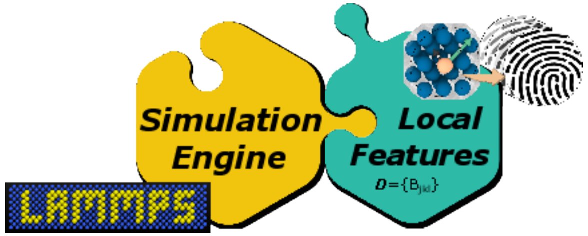




★ C	-	Planetary impacts, shock	USF, Willman	N_{DoF}=1596	N_{train}	Quadratic

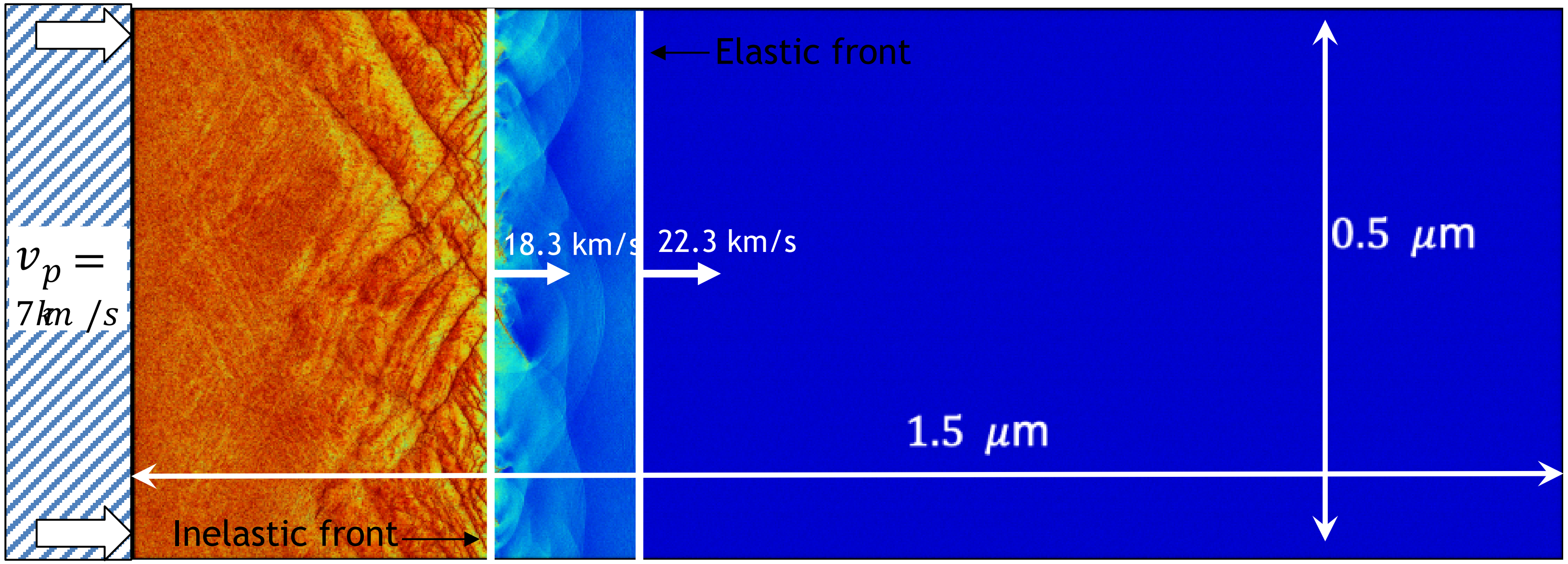
Exascale ML-MD

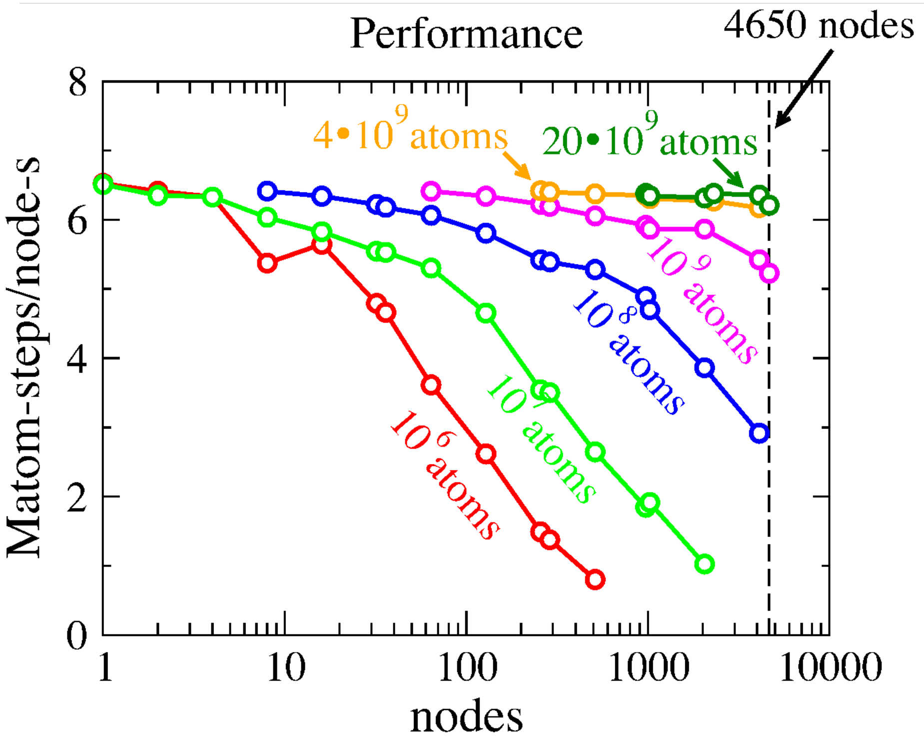
Breakdown of timing:



Gordon Bell Finalist

- ML-IAP cost will be dictated by the descriptors of the local atom environment
- Team from USF, Sandia, NERSC, NVIDIA, KTH : doi.org/10.1145/3458817.3487400

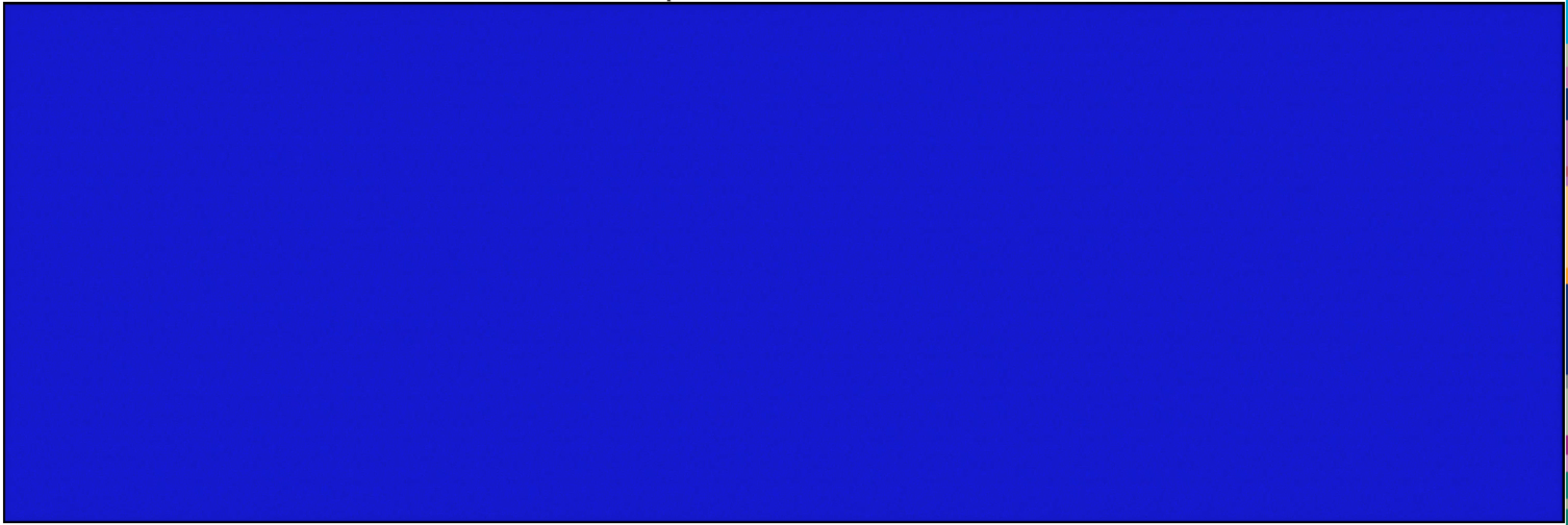




Exascale ML-MD

doi.org/10.1145/3458817.3487400

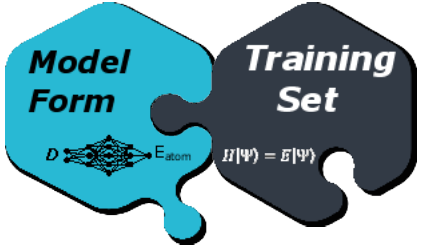
- 2.6 billion atom diamond sample, $0.5{\times}1.5\ \mu$m
- Shock wave in <110> direction initiated by piston, $v_p = 7\ km/s$.



- Novel mechanism of inelastic deformations observed for the 1st time – multiple cracks create multiple sound waves which interfere while propagating towards the elastic front

Transformative opportunity - direct atomic-scale insight by running simulations at experimental time and length scales

Maximized Diversity in Training Sets

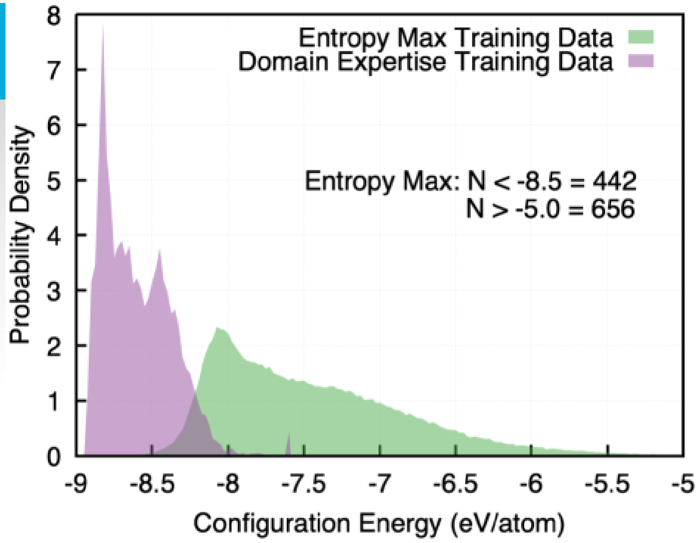


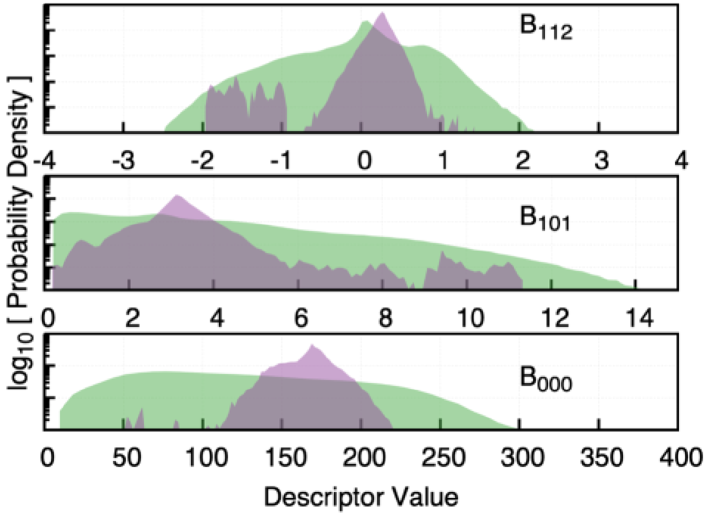
Montes, Perez, Lubbers, Pereyra

arXiv:2201.09829v1

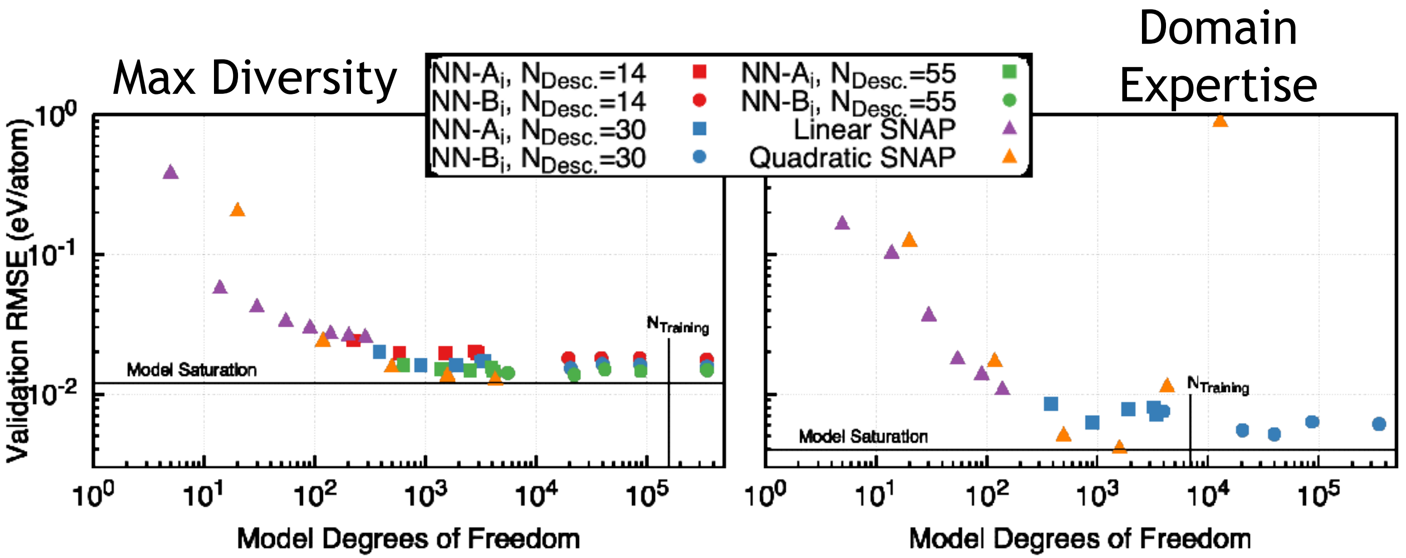
Model form – Training Pairing

- SNAP models are really only tied to bispectrum components as descriptors, model form is flexible
- How complex of a model is needed to capture the training set? Linear? Deep Neural Network?

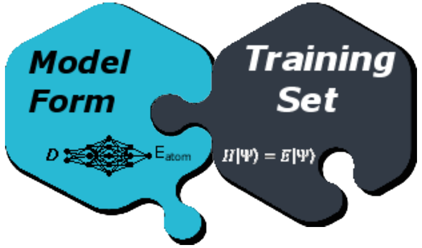




- Accuracy of ***all*** model forms saturates, true of simple linear and NN models!
- Observed for user constructed and automated training set generation!



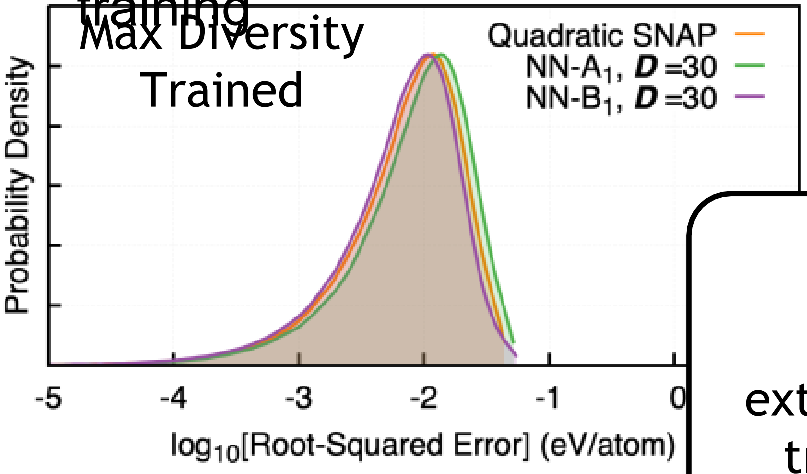
Interpolation vs. Extrapolation

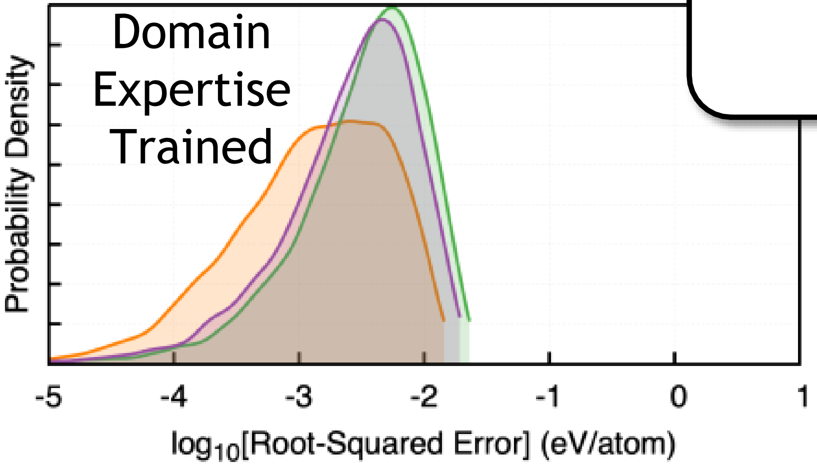


Montes, Perez, Lubbers, Pereyra

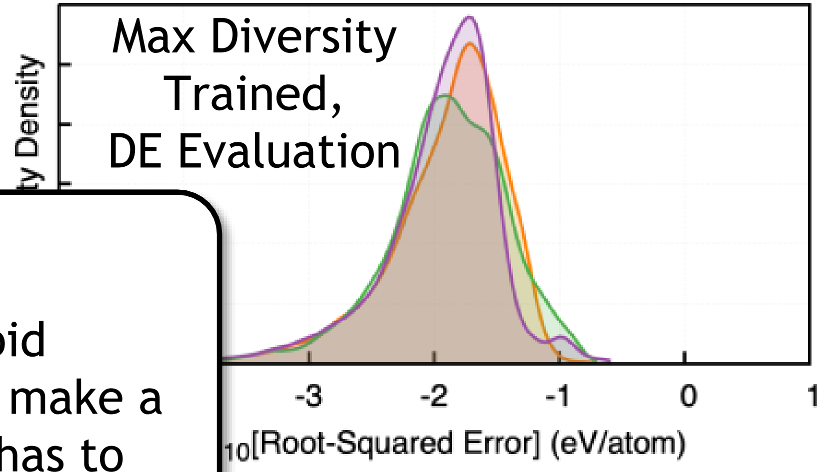
arXiv:2201.09829v1

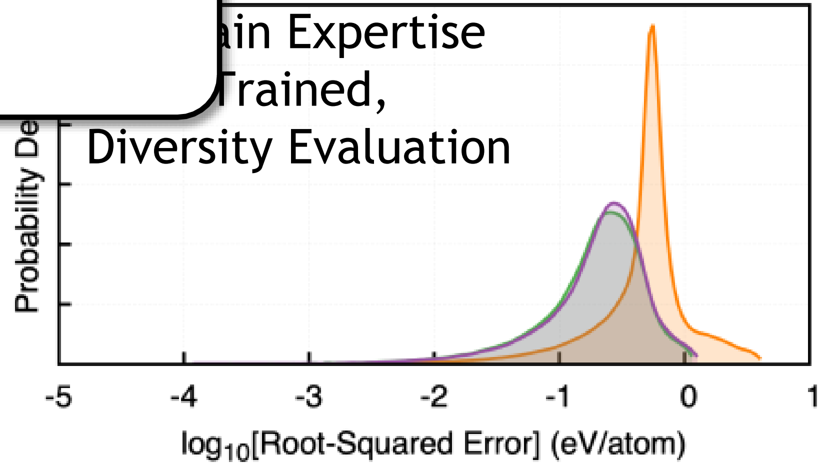
- Interpolation : Fed the **same** 'kind' of validation data as training





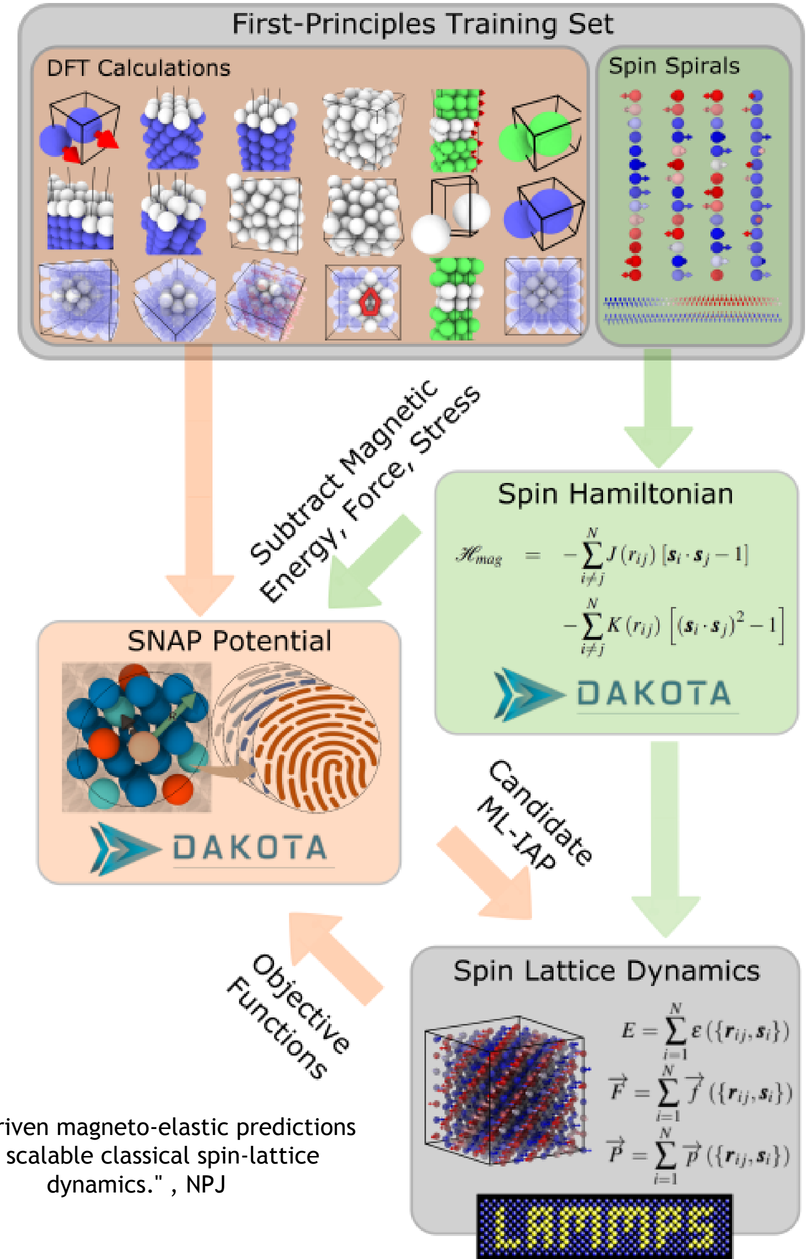
- Extrapolation : Fed the **opposite** 'kind' of validation data as training





The best way to avoid extrapolation errors is to make a training set that never has to extrapolate!

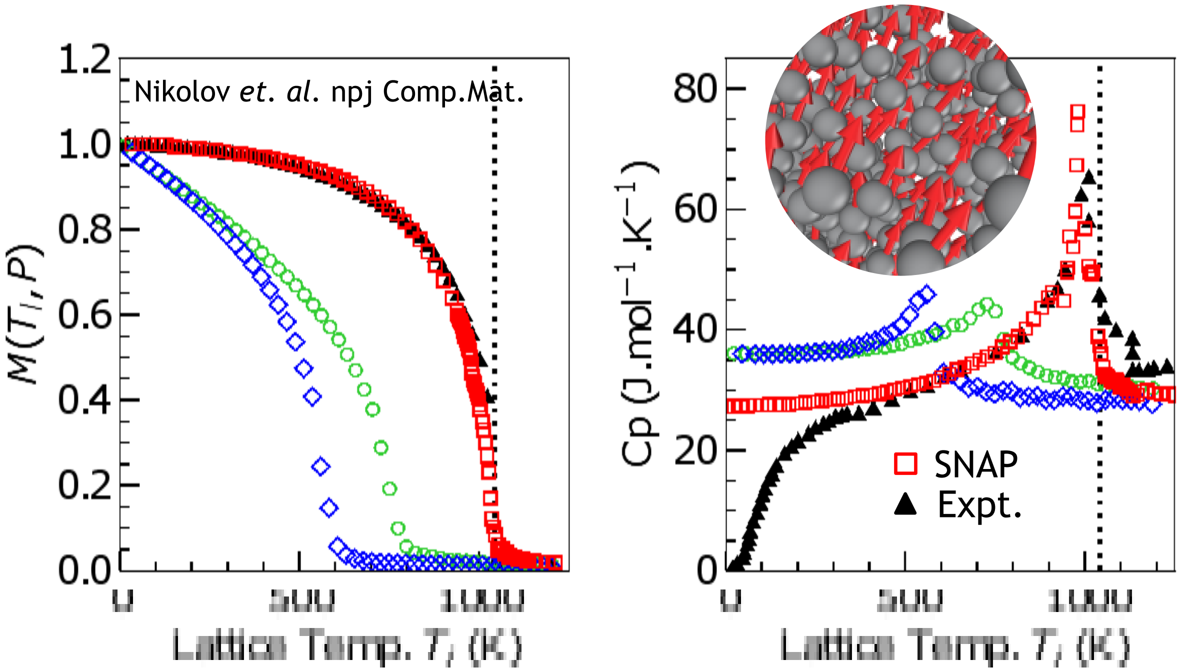
Magnetic Materials Simulations



"Data-driven magneto-elastic predictions with scalable classical spin-lattice dynamics." , NPJ

Fe; Everyone's Favorite

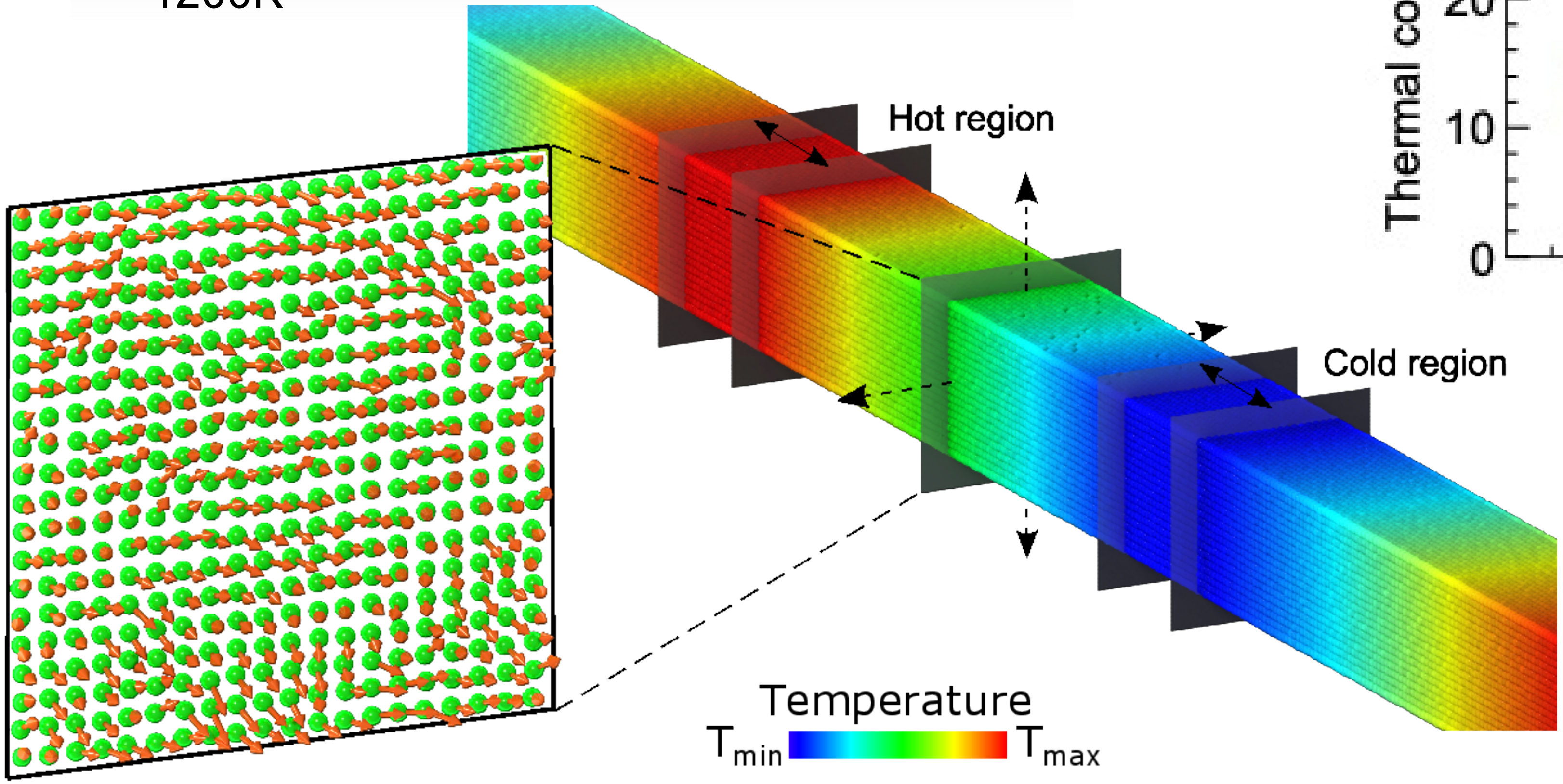
- Transformational capability to study magnetic materials at the grain scale
- Explicit treatment of spin dynamics captures the second order phase transition at Curie temperature

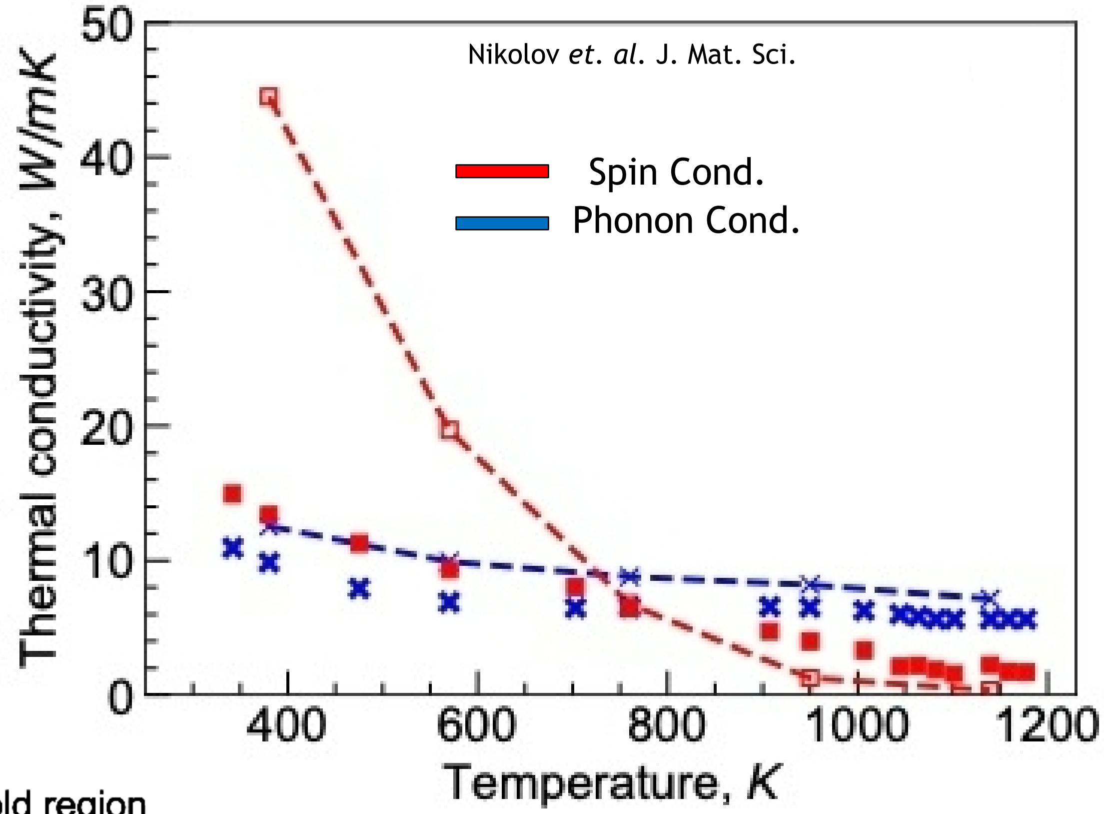


Magnetic Materials Simulations

Finite Temperature Magnetism

- Hot/cold regions are spaced 28.8 nm apart
- Thermal gradient established by setting hot region to $T_{max} = 1.08T_{min}$,T_{min} : 300 - 1200K

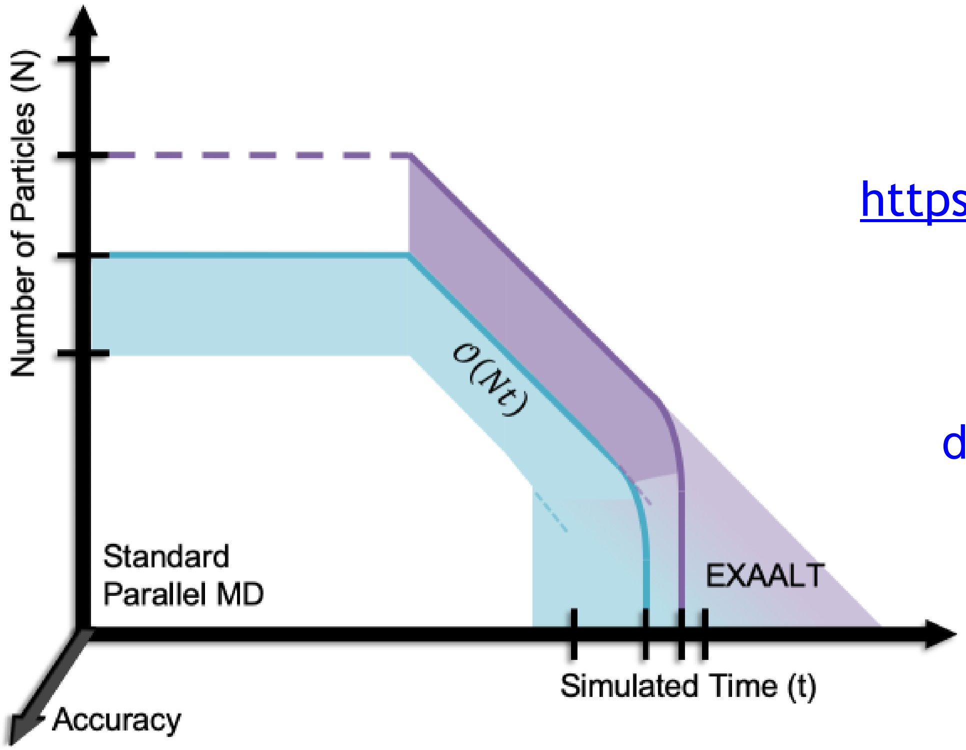




- Magnon-phonon scattering significantly reduces conductivity
- Magnons more conductive than phonons where $T < 0.5T_{Curie}$

Conclusions and Path Forward

- 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.



https://github.com/FitSNAP/FitSNAP

http://lammps.sandia.gov

Gordon Bell Finalist Paper:
doi.org/10.1145/3458817.3487400

Entropy Maximization Paper:
arXiv:2201.09829v1

Contact Information:

mitwood@sandia.gov