SAND2022-7063C

Sandia National Laboratories

FitSNAP : Scalable Solutions for Training Machine Learned Interatomic Potentials

Mitchell Wood

Sandia National Labs, Albuquerque, New Mexico, USA

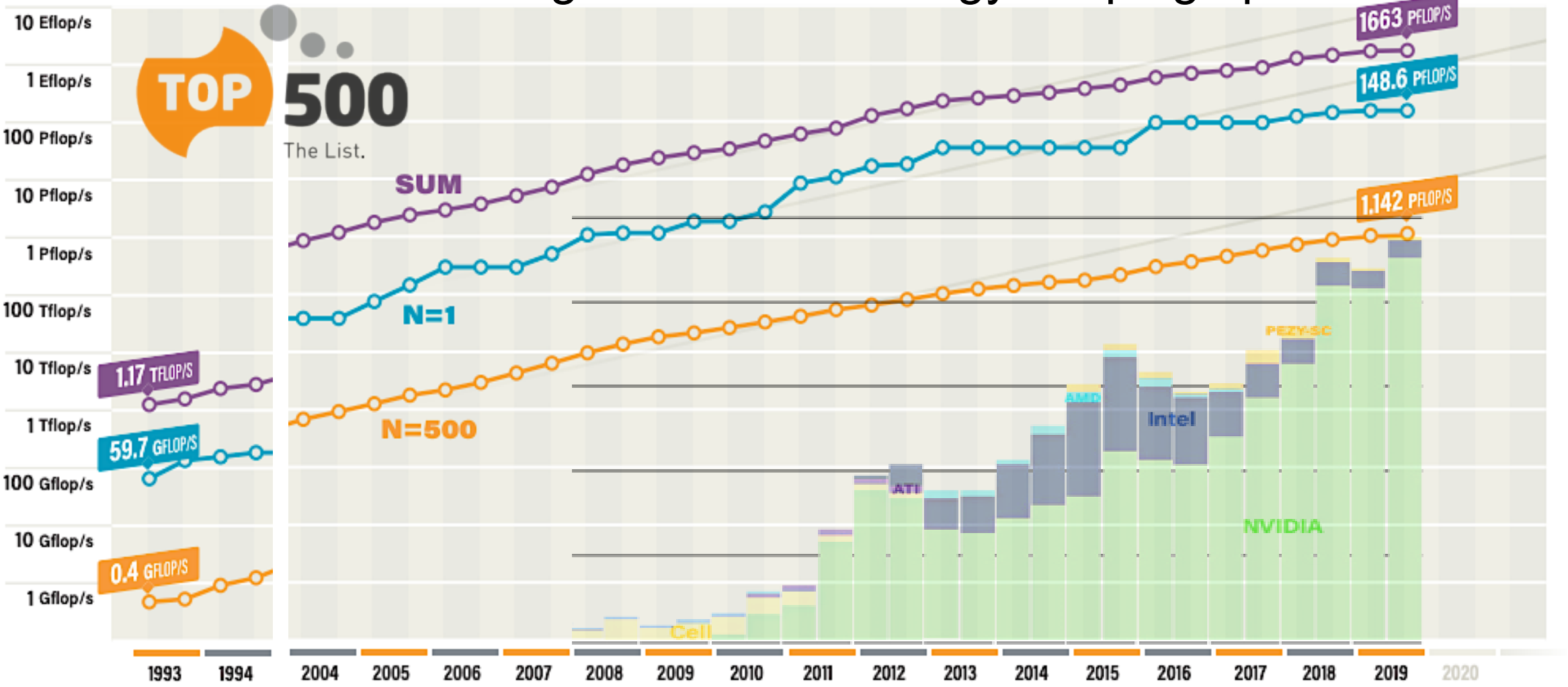
AMSET Workshop, May 23th, 2022

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 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.

Computing Status Report

More powerful supercomputers are inevitable, but is our scientific usage of this technology keeping up?



MD Approximations Change Over Time

http://lammps.sandia.gov

Twobody (B.C.)
Lennard-Jones, Hard Sphere, Coulomb, Bonded

Manybody (1980s)
Stillinger-Weber, Tersoff, Embedded Atom Method

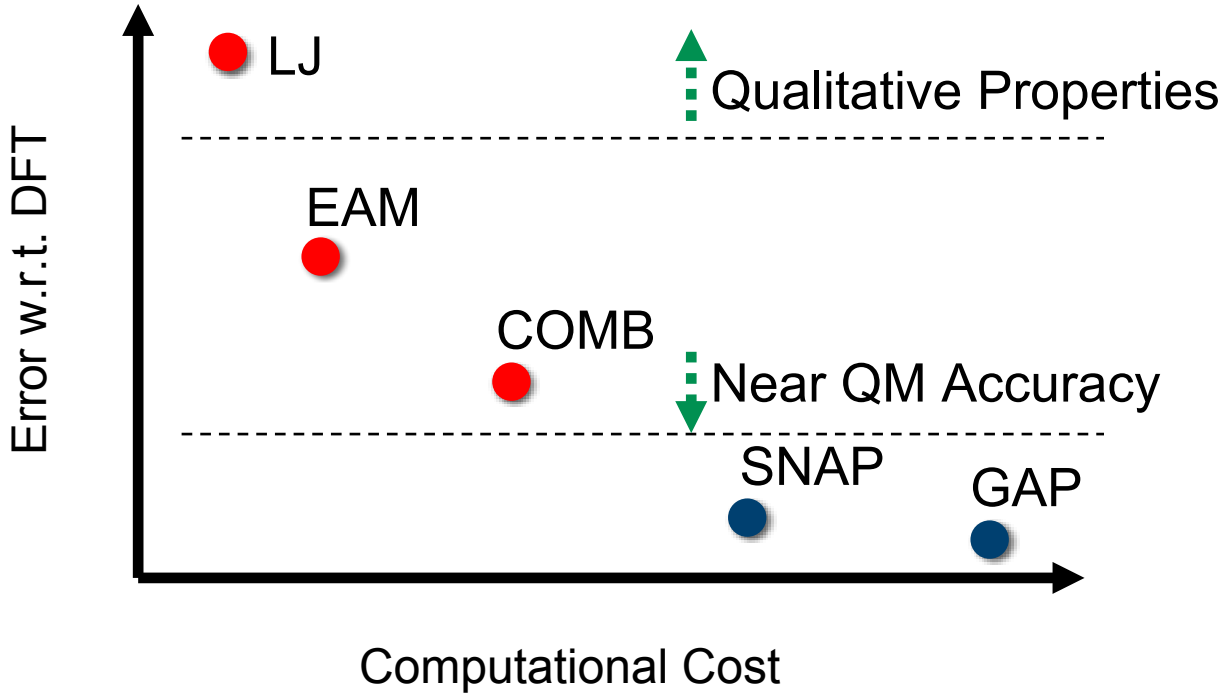
Advanced (90s-2000s)
REBO, BOP, COMB, ReaxFF

Big Data / Deep / Machine Learning (2010s)
GAP, SNAP, NN,…

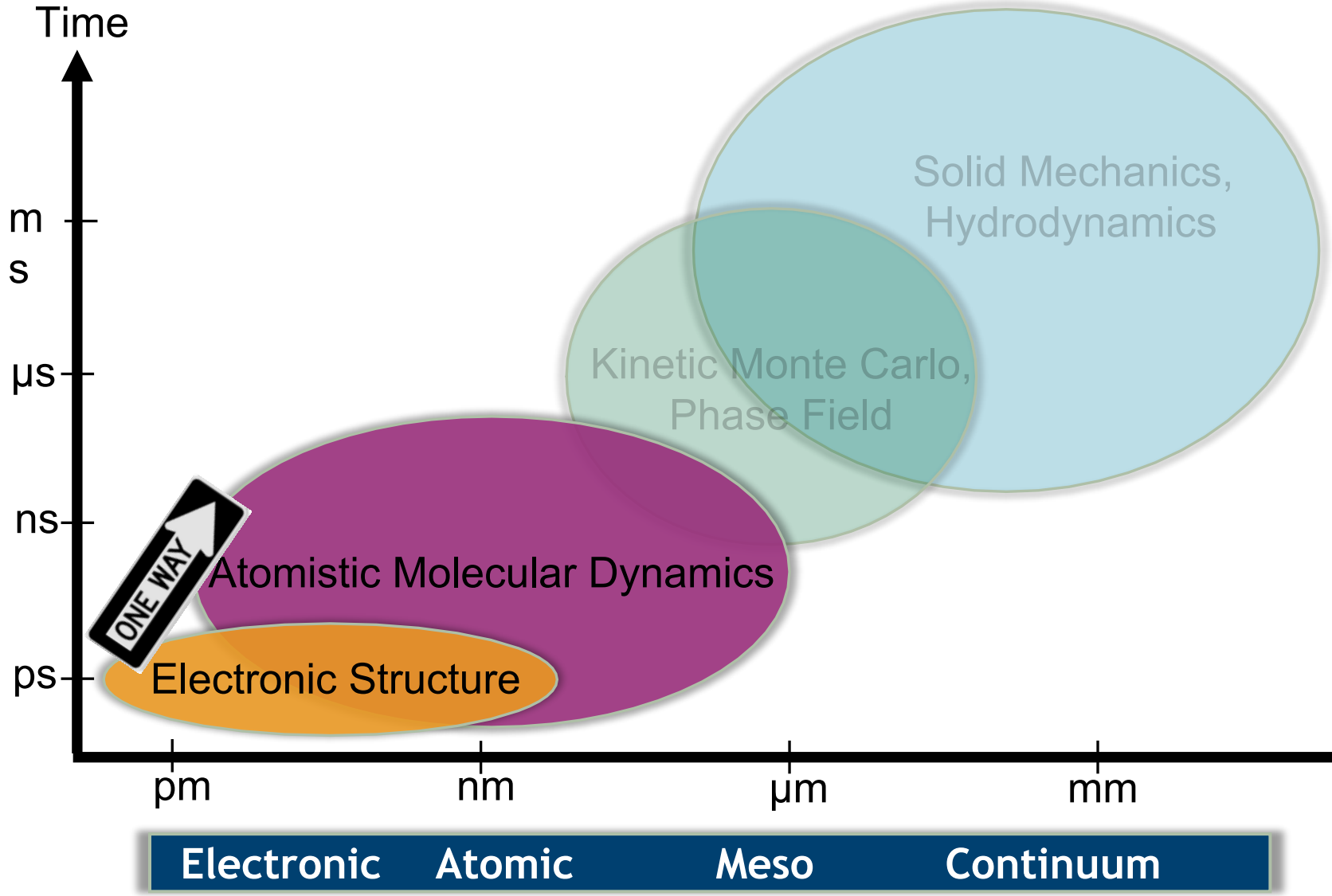
Plimpton and Thompson, MRS Bulletin (2012).

Resources are limited, which is your best choice?

https://github.com/materialsvirtuallab/mlearn



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

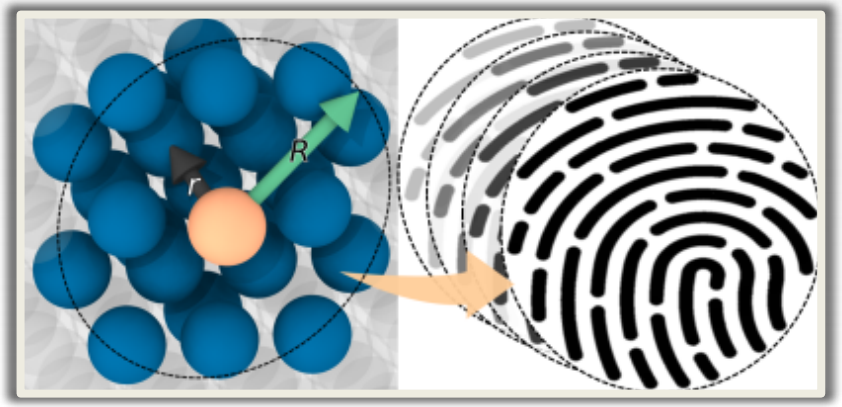
Machine Learning to Bridge Electronic-Atomic

Classical, Empirical Potentials

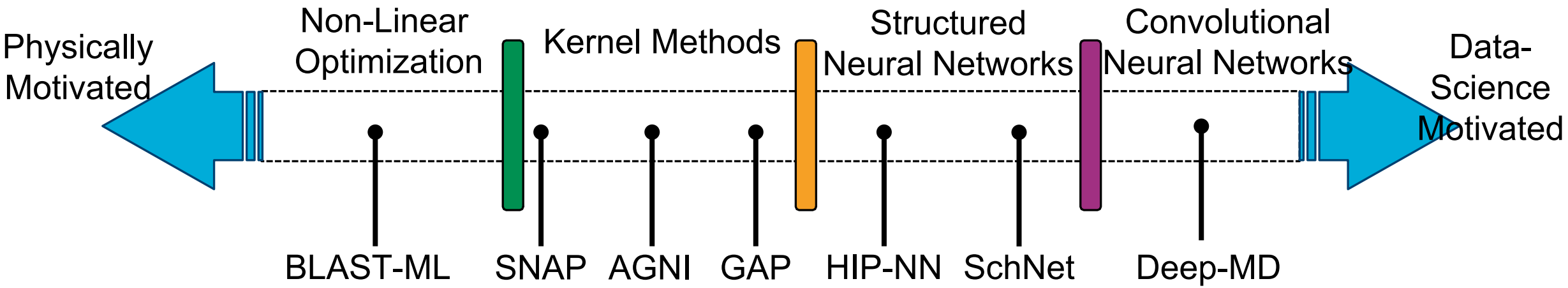
- Metals
  - EAM: Assume spherical electron density
    $$E_i = F_\alpha\left(\sum_{j\neq i}\rho_\beta(r_{ij})\right) + \frac{1}{2}\sum_{j\neq i}\emptyset_{\alpha\beta}(r_{ij})$$
- Inorganic
  - Stillinger-Weber: Assume 2,3-body harmonic springs
- Organic
  - ReaxFF: Assume covalent bonding, smooth bond-orders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

Machine Learned Potentials

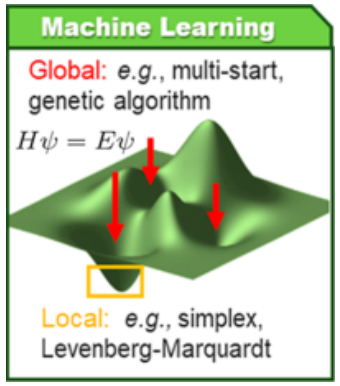
- Metals, Inorganic, Organic, etc.
  - Assume energy and forces are some function of local atomic neighborhood descriptors
- Needs reference data to be properly trained to get the ‘right’ energies and forces

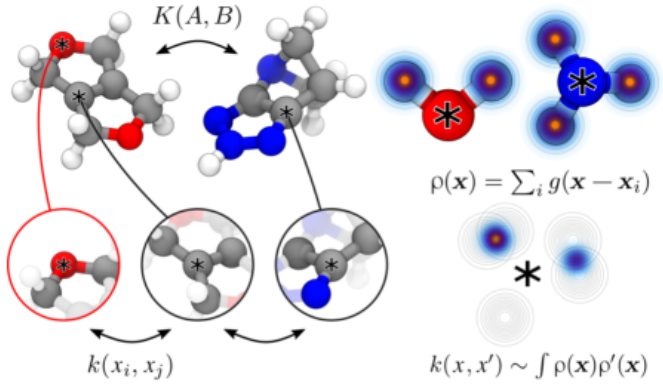


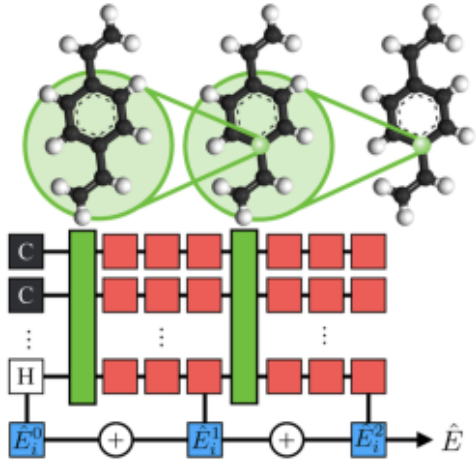
Environment of Machine Learning Techniques

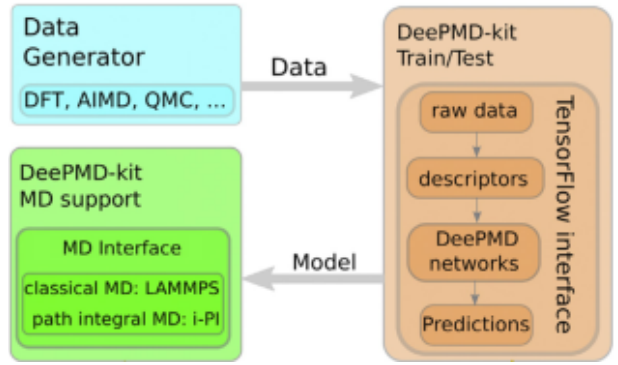


- Adoption of machine learning techniques within molecular dynamics has been varied



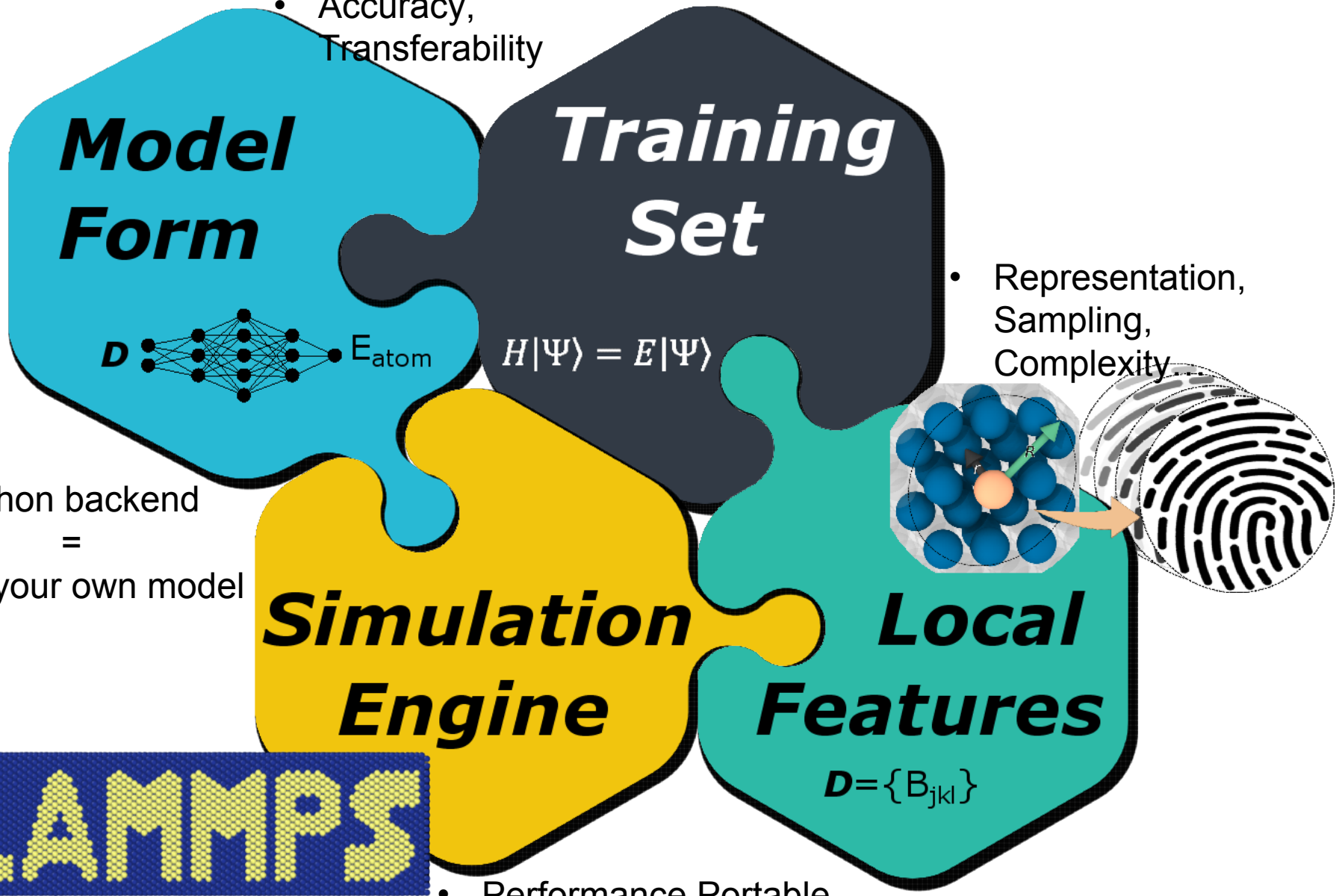






Components of ML-MD

https://github.com/FitSNAP/FitSNAP
http://lammps.sandia.gov



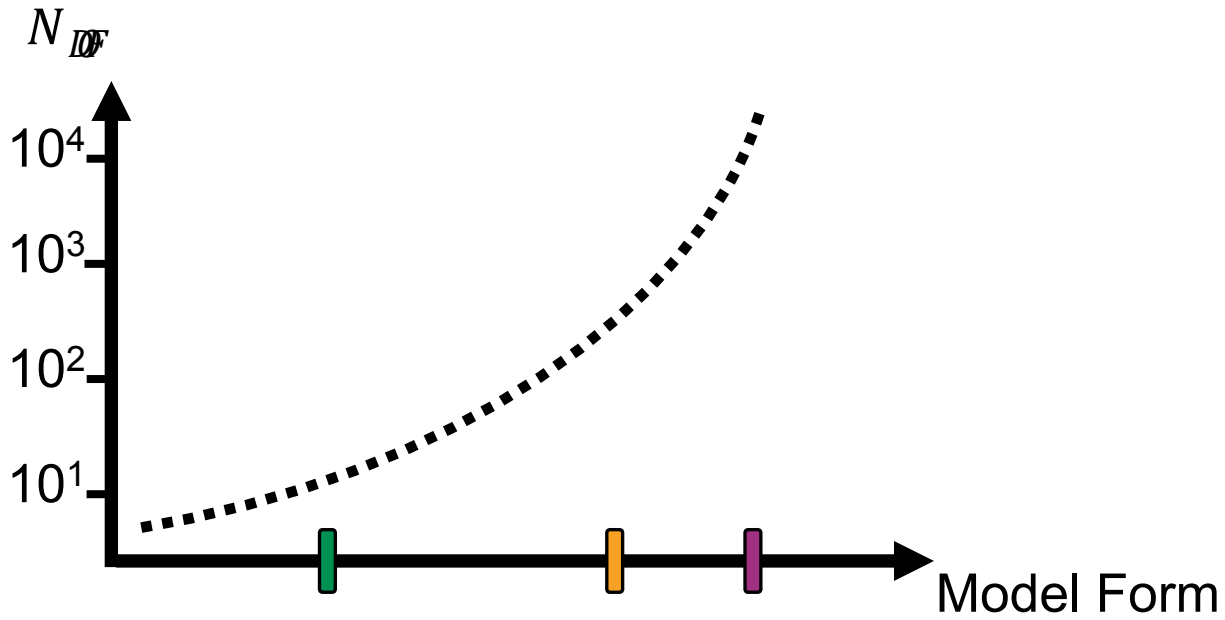
- Accuracy, Transferability
- Representation, Sampling, Complexity…
- Python backend = Bring your own model
- Performance Portable

Environment of Machine Learning Techniques

Descriptors	Training Technique
• 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:

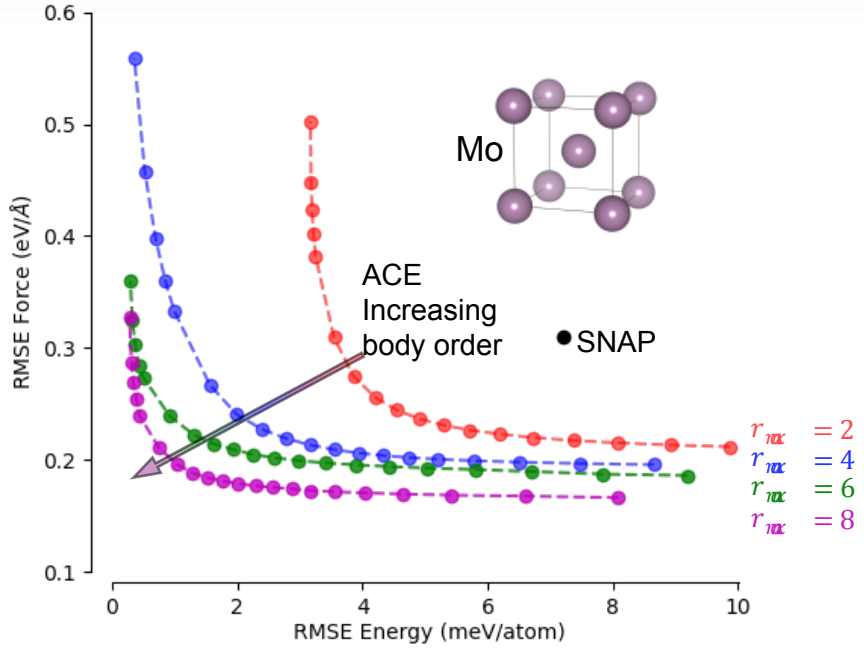
- When $N_{DOF} \sim N_{Train}$, high risk of overfitting → Poor Interpolation
- When training diversity is low → Poor Extrapolation
- RUNNING MD WILL EXPOSE THESE SHORTCOMINGS

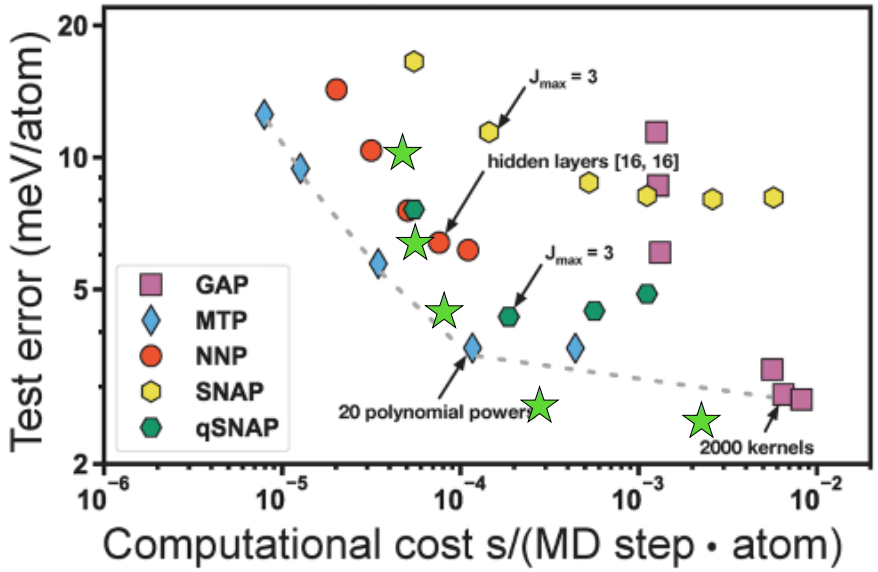


Accuracy Demands

Training Data Sets for Material Science

- S.P. Ong (UCSD) and group members generated ‘standardized’ training data for Cu, Ni, Mo, Li, Si, Ge
- Comparing ML-IAP implemented in LAMMPS, cost assessed on CPU only
- https://github.com/materialsvirtuallab/mlearn



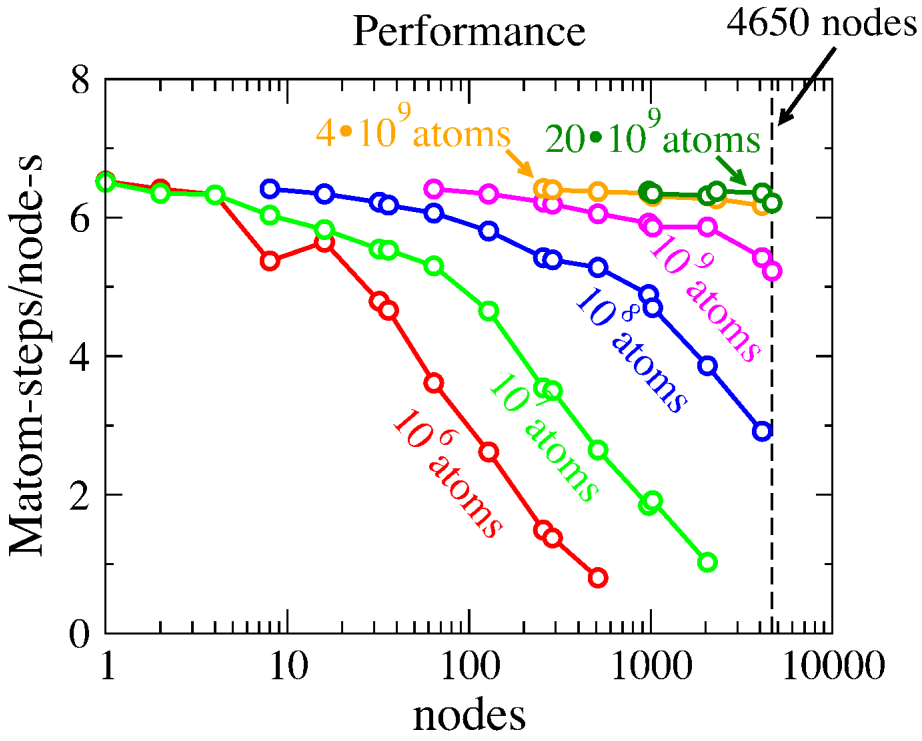


Zuo *et. al. J. Phys. Chem. A* (2020), 124, 4,

Efficiency Demands

2021 Gordon Bell Finalist

- Uses SNAP ML-IAP for high pressure Carbon
- Team from USF, Sandia, NERSC, NVIDIA, KTH : doi.org/10.1145/3458817.3487400

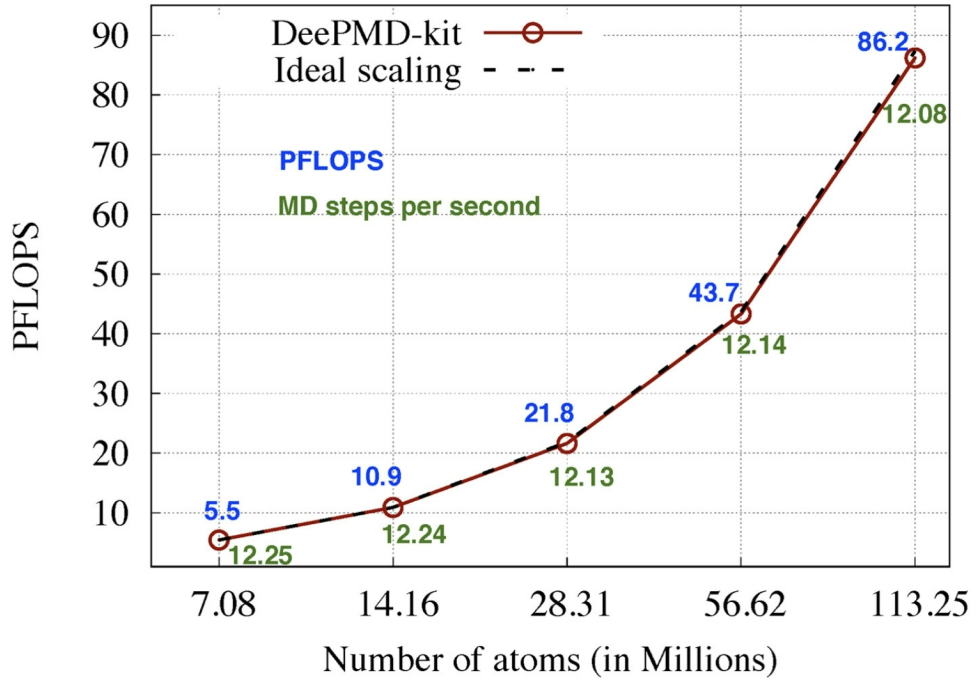


*Both heavily rely on GPU parallelization!

*Both interface with and run in LAMMPS!

2020 Gordon Bell Winner

- Neural Network based ML-IAP for Water, Cu
- Team from Princeton, Berkeley, LBNL, IAPCM(Beijing), Peking Univ.: doi.org/10.1016/j.cpc.2020.107624



ML-IAP in LAMMPS

- **Native LAMMPS**

ML-SNAP

- **LAMMPS Interfaces**

ML-HDNNP: Singraber, N2P2, Behler-Parrinello Descriptors, ANN Potentials

ML-QUIP: Bartok, Csanyi, GAP Potentials, SOAP Descriptors

ML-PACE: Lysogorskiy, Drautz, Atomic Cluster Expansion

ML-RANN: Dickel, NN potential with fast fingerprints

KIM: Tadmor, many ML potentials: DUNN, hNN, PANNA

USER-DEEPMD: Zhang, E, Car, Deep Network Potentials

USER-MLIP: Shapeev, Moment Tensor Potentials

USER-MLIP: Seko, Machine Learning Potential Repository

USER-PINN: Mishin, Physically informed neural network potential

USER-ANI: Barros, Smith, Lubbers, ANI ANN Potentials

USER-AENET: Artrith, Behler-Parrinello Descriptors, ANN Potentials

:

:

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 list)

System	Year	Usage	Origin	N_{DoF}	$N_{Training}$	Descriptors
WBeHN	-	Plasma facing materials	SNL, Cusentino	56*	>40,000	Linear
WZrC	-	Plasma facing materials	SNL, Sikorski	56*	>40,000	Linear
C	2022	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 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

Accelerating Model Development

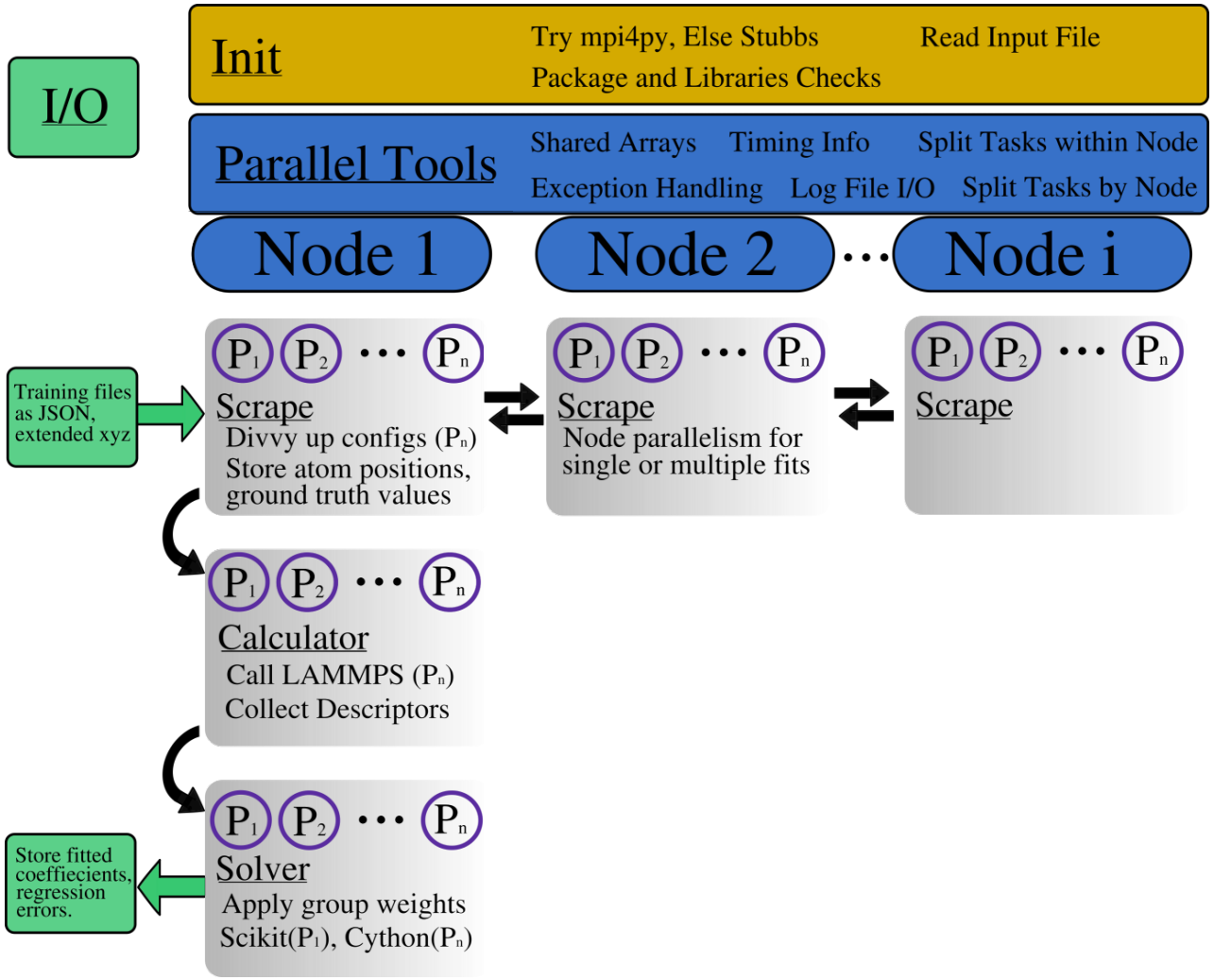
FitSNAP

FitSNAP.py Breakdown

- Software tools needed to realize the accuracy-cost tradeoff continuum
- Python interface for user ease of use Object oriented framework for developer ease of use

- ML-IAP can be ‘overlapped’ with other physical models (coulombic, magnetic spins, ion core repulsion)
- Training data storage as dictionaries:

```
Dataset =
{"Label":
  "Example containing 1 configurations, each with 2 atoms",
  "LatticeStyle": "angstrom","EnergyStyle": "electronvolt","StressStyle": "kB",
  "Data":
    [{"NumAtoms": 2,
      "Lattice": [[ 2.4223151206970215, 0.0, 0.0],[ 0.0, 2.139922618865967, 0.0
      "Energy": -14.30234864,
      "Stress": [[ -37.90633, 6.66761, 0.00398],[ 6.66761, 752.92696, 0.00498],
      "AtomTypes": [ "Fe", "Fe"],
      "Positions": [[ 0.0, 0.0, 0.0],[ 2.26141, 1.33745, 1.97781]],
      "Forces": [[ 0.809971, 0.872872, 0.000495],[ -0.809971, -0.872872, -0.000
      "Spins": [[ 0.10080178569846865, 0.03968183670838251, 0.9126822442927977,
      }]}}
```



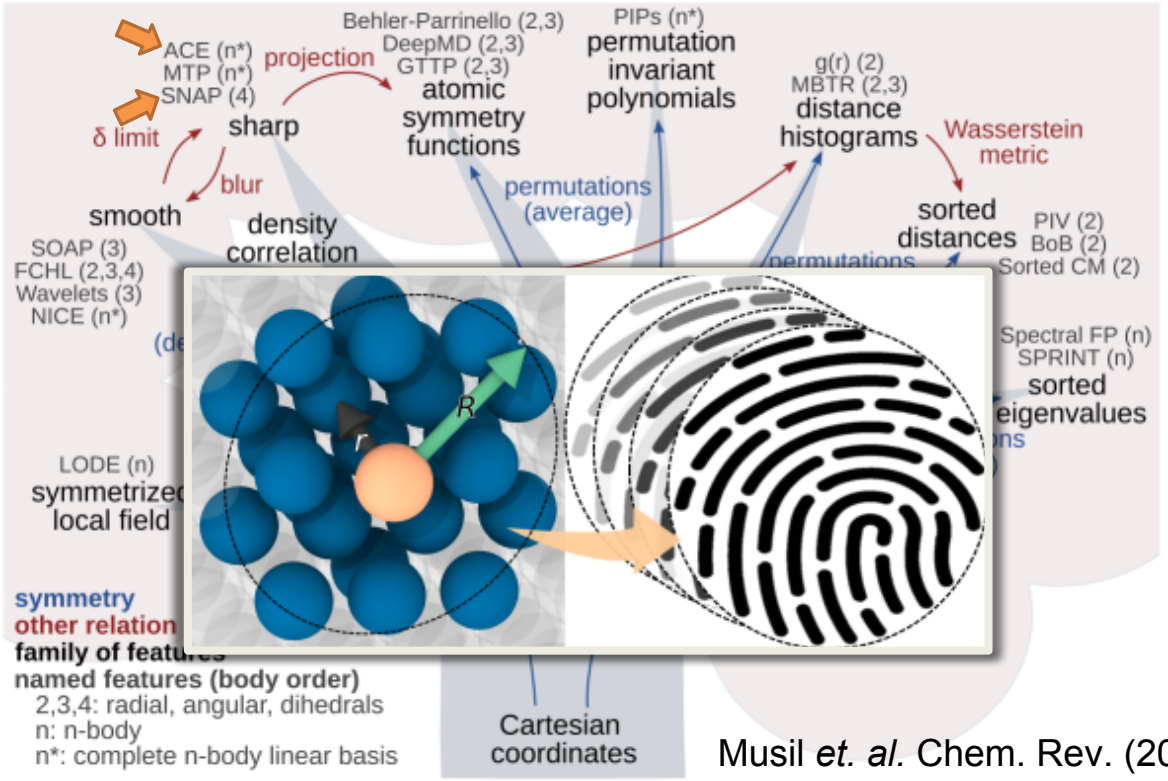
*Auto conversion from VASP OUTCAR coming soon!

Descriptors

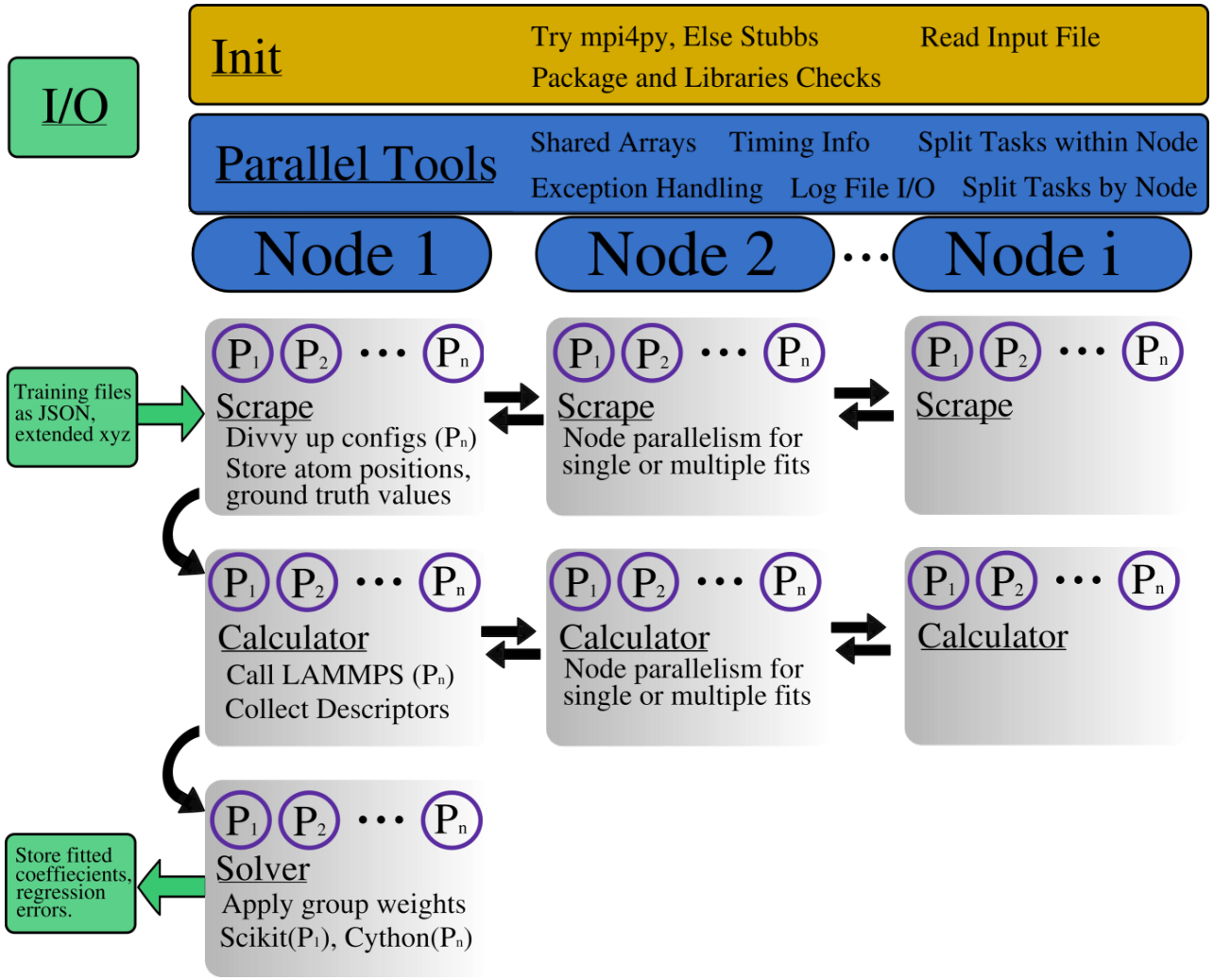
FitSNAP

LAMMPS Breakdown

- Calculator class calls LAMMPS to convert atomic coordinates into descriptors.
- Thread parallel implementation via Mpi4Py and LAMMPS python library interface.



Musil *et. al.* Chem. Rev. (2021) 121,



Regression

FitSNAP

Simple Model Form

$$E^i_{SNAP} = \alpha_0 + \sum_k \left[\alpha^{(1)}_k (B^i_k - B_{k_0}) \quad + \alpha^{(n)}_k (\dots)^n\right]$$

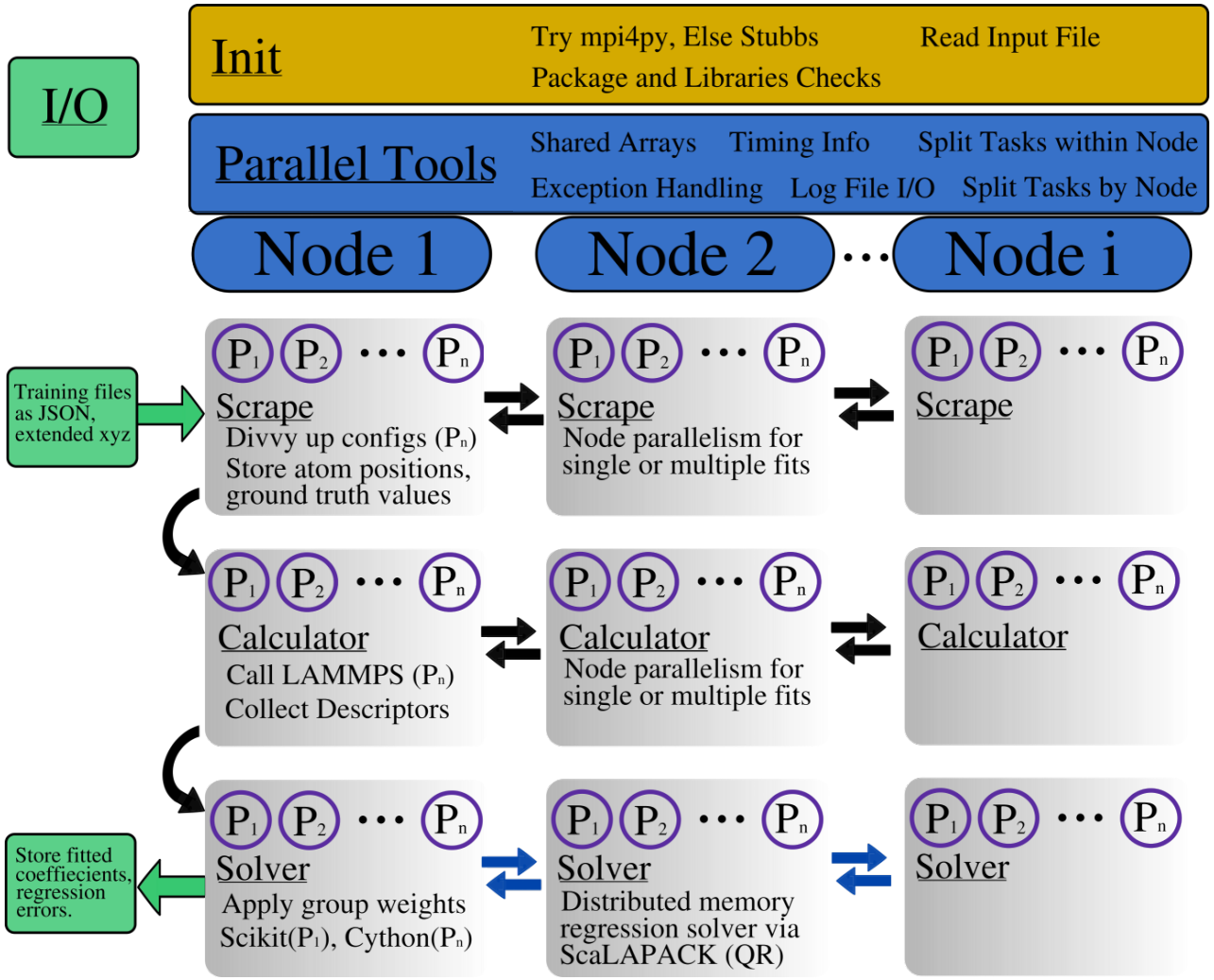
$$min\left(\|w \cdot D\alpha - T\|^2 - \gamma_l \|\alpha\|^l\right)$$

w : Weights

D : Set of descriptors

T : DFT training

- D is a N*M matrix
  - N Scales with number of training points
  - M Scales with the descriptor expansion
  - N>>M, can exceed local memory



*Coupling to AutoDiff and Pytorch available!

Parallelization

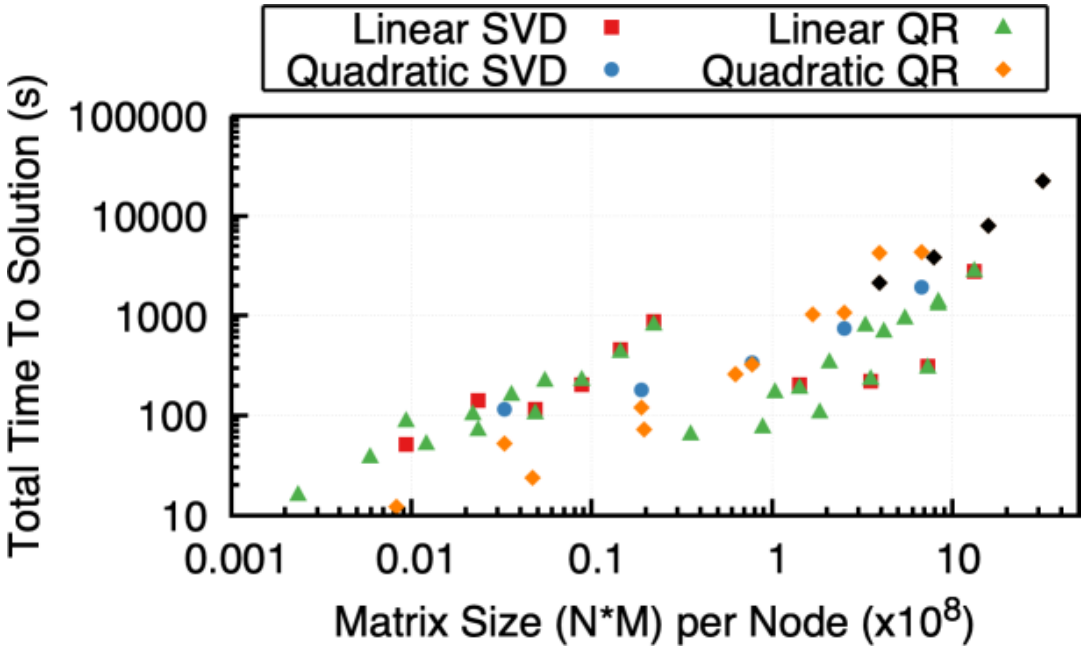
Timing Breakdown

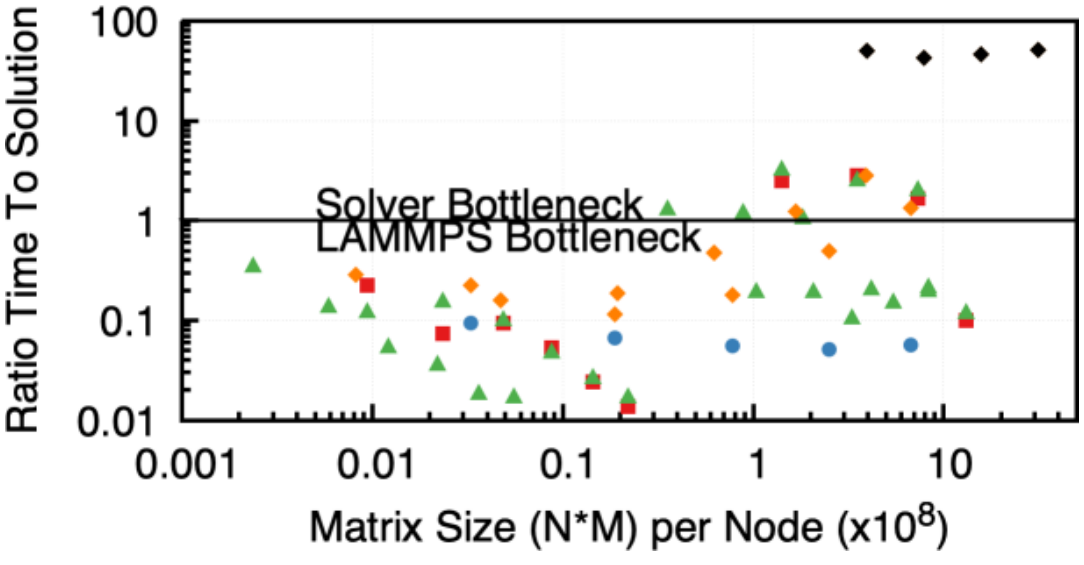
- Cython backend to Solver class allow for distributed memory regression → QR Decomposition via ScaLAPACK
- Each nodes' object handles its own set of training Data → ML-IAP fitting only limited by resource availability

- Points are increasing descriptor basis, Quadratic results in order of magnitude larger M sizes

◆ N*M matrix exceeds the 128Gb of local shared memory, only possible with a distributed solver.

- Gradient descent solvers are implemented, but are slow and add hyperparameters to fits



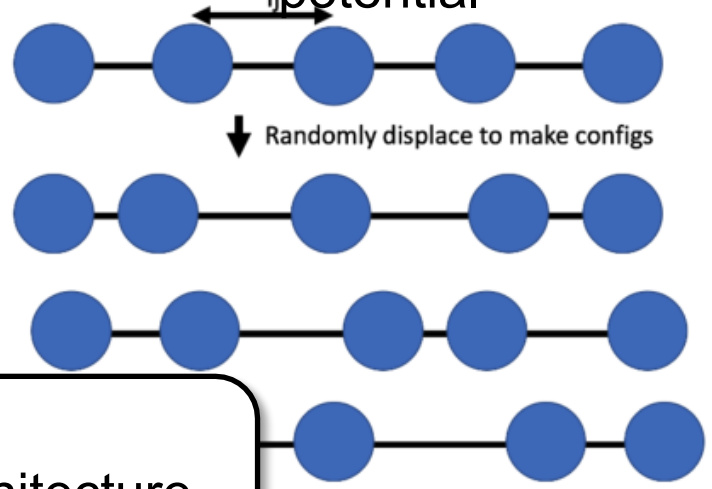


Neural Network Models

More Memory Needed

- Non-linear, descriptor-based models of atomic energy are simple to train, much more complicated for forces
  - i : central atom index
  - j : neighboring atom index
  - k : descriptor id

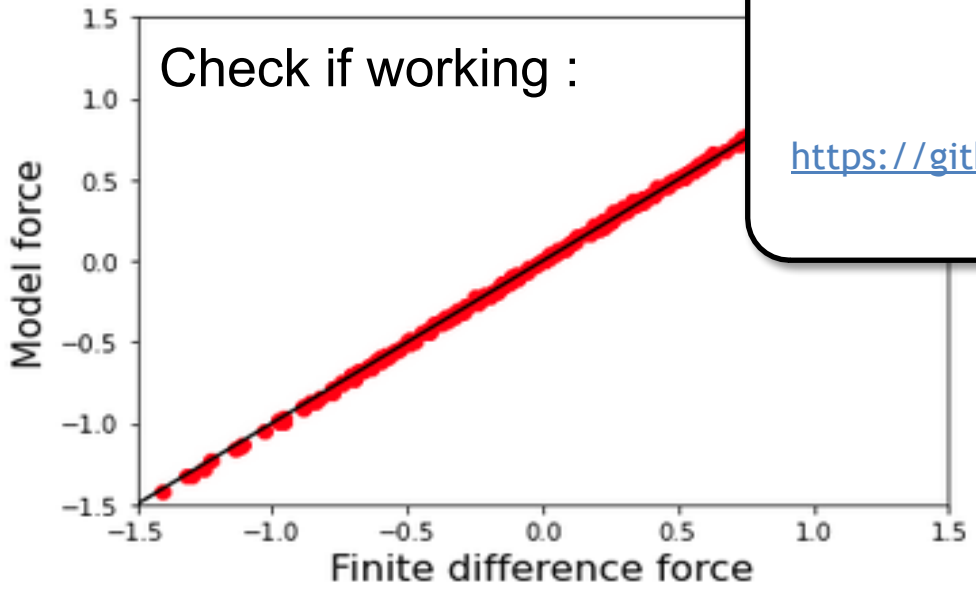
Generate simple configs modeled by harmonic potential



Same procedure/architecture as more realistic examples!

Example :
https://github.com/rohskopf/SimpleML-IAP

Check if working :



s using PyTorch or Jax

$$\sum_j \sum_k \frac{\partial D_{jk}}{\partial r_i} \frac{\partial E_j}{\partial D_{jk}}$$

Computed before fitting via neighbor lists.

Computed during fitting via autograd.

Software Development

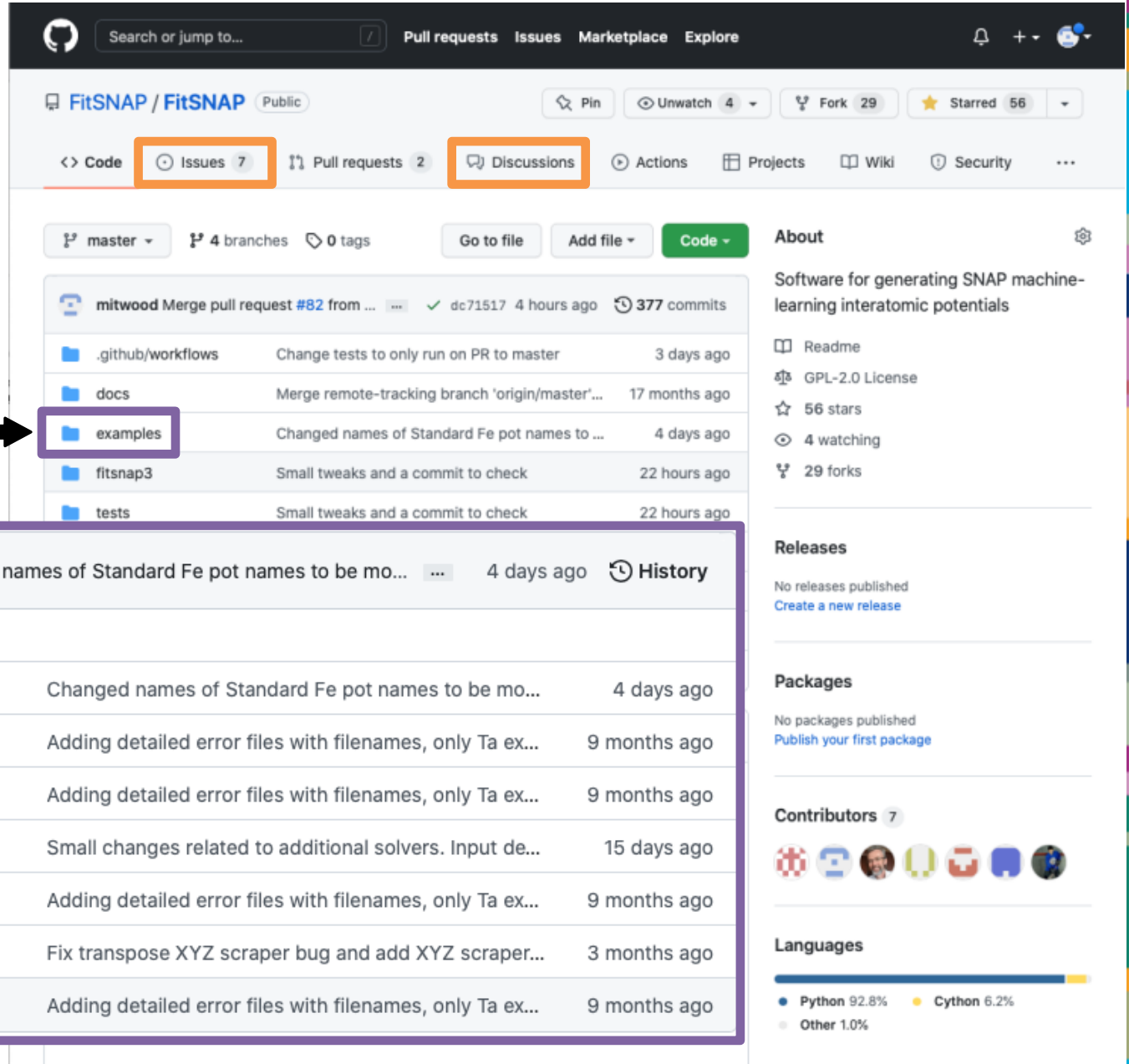
https://github.com/FitSNAP/FitSNAP

- pip install fitsnap
  or
  git clone https://…

- Open issues, request features, and discuss with developers/users via GitHub

- Example cases roughly follow our publications;

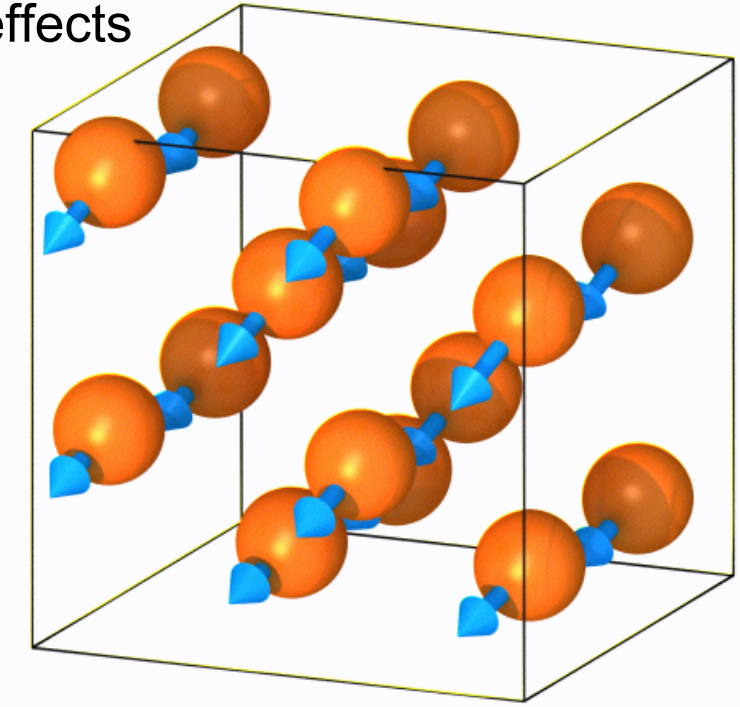
Fe: SNAP + Magnetic Spin
InP: Explicit Multi-Element Descriptors
Ta_PACE: Atomic Cluster Expansion Model
Etc.



Magnetic Materials Simulations

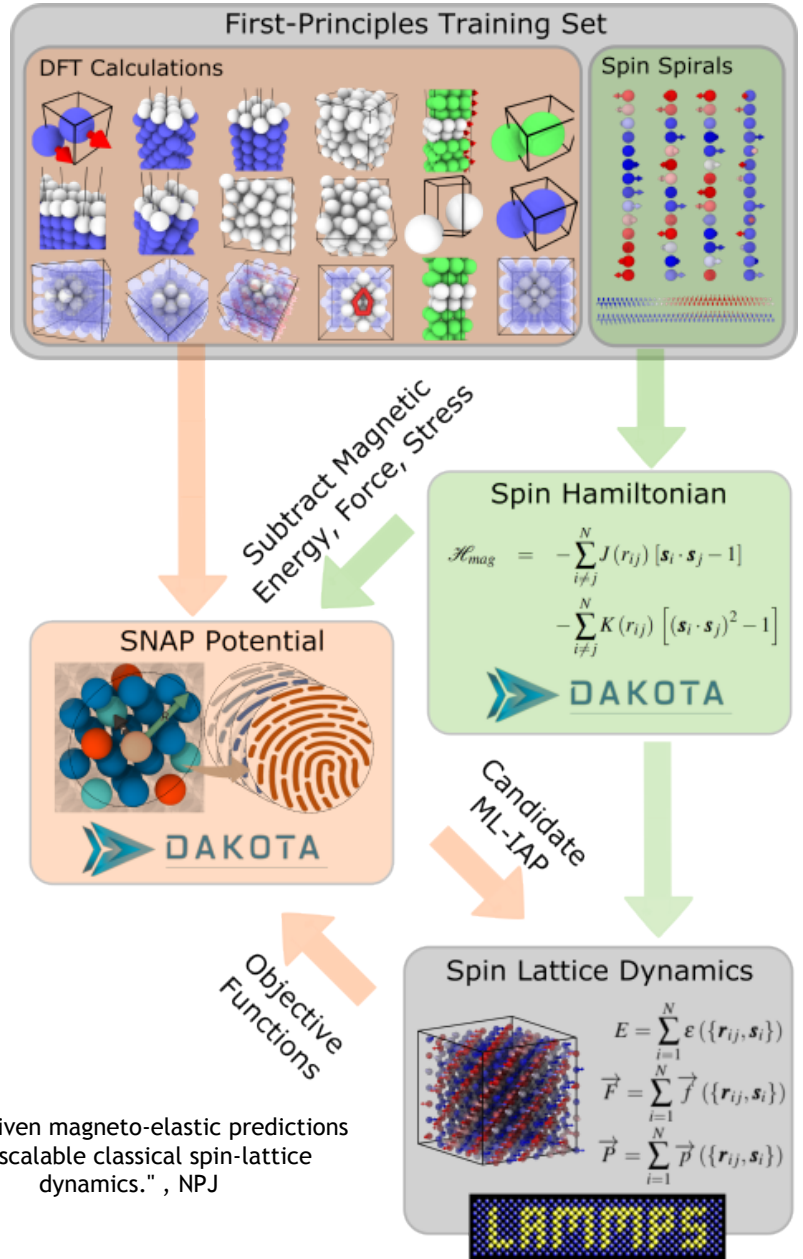
Chemical, Functional Complexity

- Electronic structure or Spin-Lattice simulations have been the only options for many decades.
- Former lacks scalability, latter lacks any real dynamic or finite temperature effects



- **Molecular Dynamics**
  - Atoms interact via nonmagnetic interatomic potential, $U(\mathbf{R})$
  - $\mathcal{H}_{MD} = \sum_i \frac{\mathbf{p}_i^2}{2m} + U(\mathbf{R})$
- **Spin Dynamics**
  - Atoms interact via exchange function which conserves total angular momentum
  - $\mathcal{H}_s = -\sum_{i,j}^{N} J_{ij}(\mathbf{R})\left[\vec{s_i}\cdot\vec{s_j} - 1\right] - \sum_{i,j}^{N} K_{ij}(\mathbf{R})\left[\left(\vec{s_i}\cdot\vec{s_j}\right)^2 - 1\right]$
- **Molecular-Spin Dynamics**
  - $\mathcal{H}_{MSD} = \mathcal{H}_l + \mathcal{H}_s = \sum_i \frac{\mathbf{p}_i^2}{2m} + U(\mathbf{R}) - \sum_{i,j}^{N} J_{ij}(\mathbf{R})\left[\vec{s_i}\cdot\vec{s_j} - 1\right] - \sum_{i,j}^{N} K_{ij}(\mathbf{R})\left[\left(\vec{s_i}\cdot\vec{s_j}\right)^2 - 1\right]$ (Implemented in LAMMPS!)

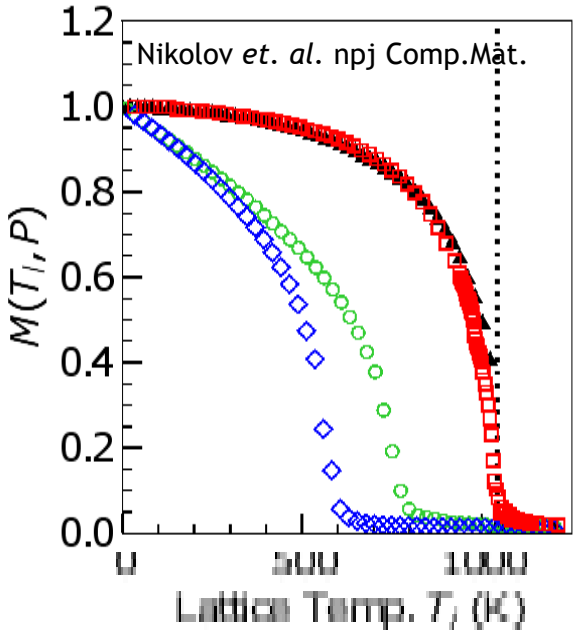
Magnetic Materials Simulations

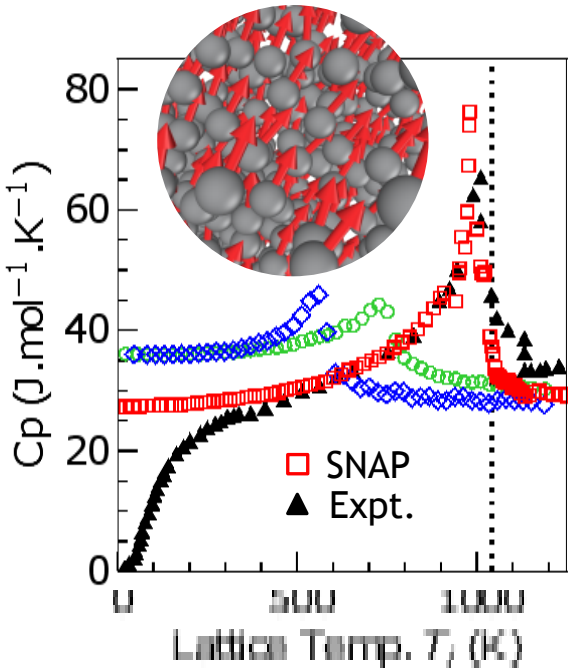


"Data-driven magneto-elastic predictions with scalable classical spin-lattice dynamics." , NPJ

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

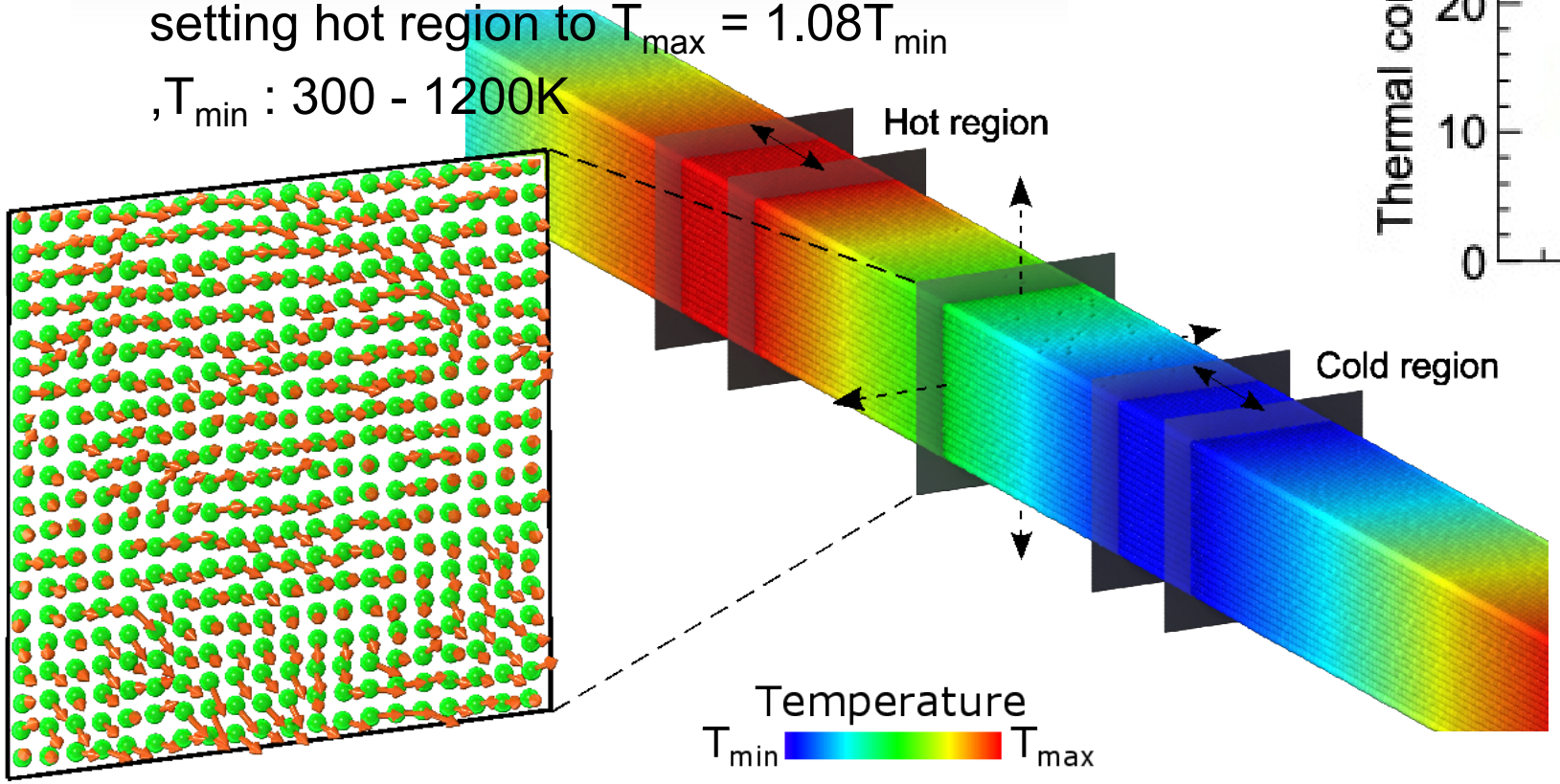


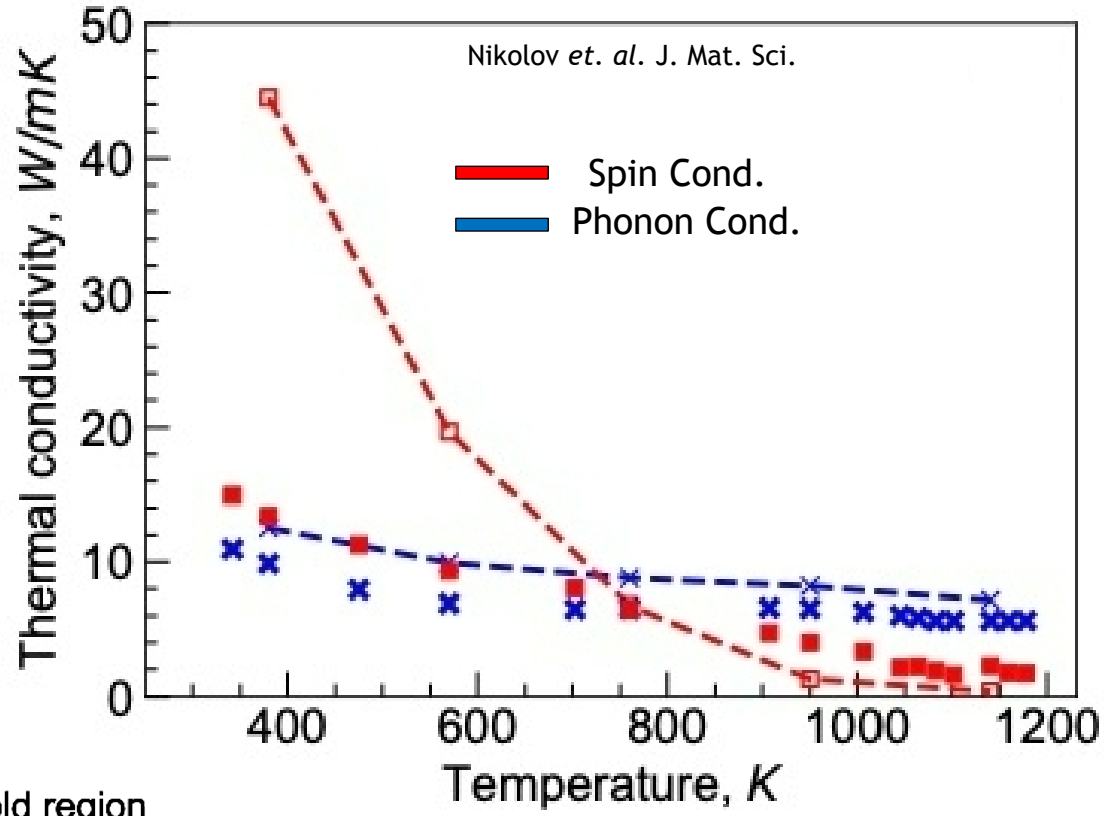


Magnetic Materials Simulations

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





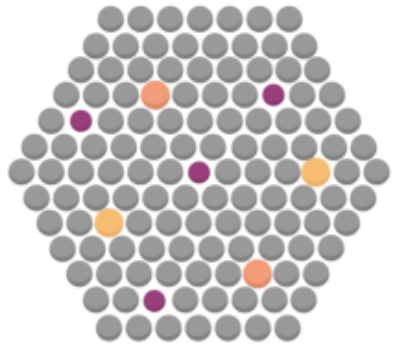
- Magnon-phonon scattering significantly reduces conductivity
- Magnons more conductive than phonons where T <

Alloy Design From MD

Alphabet Soup

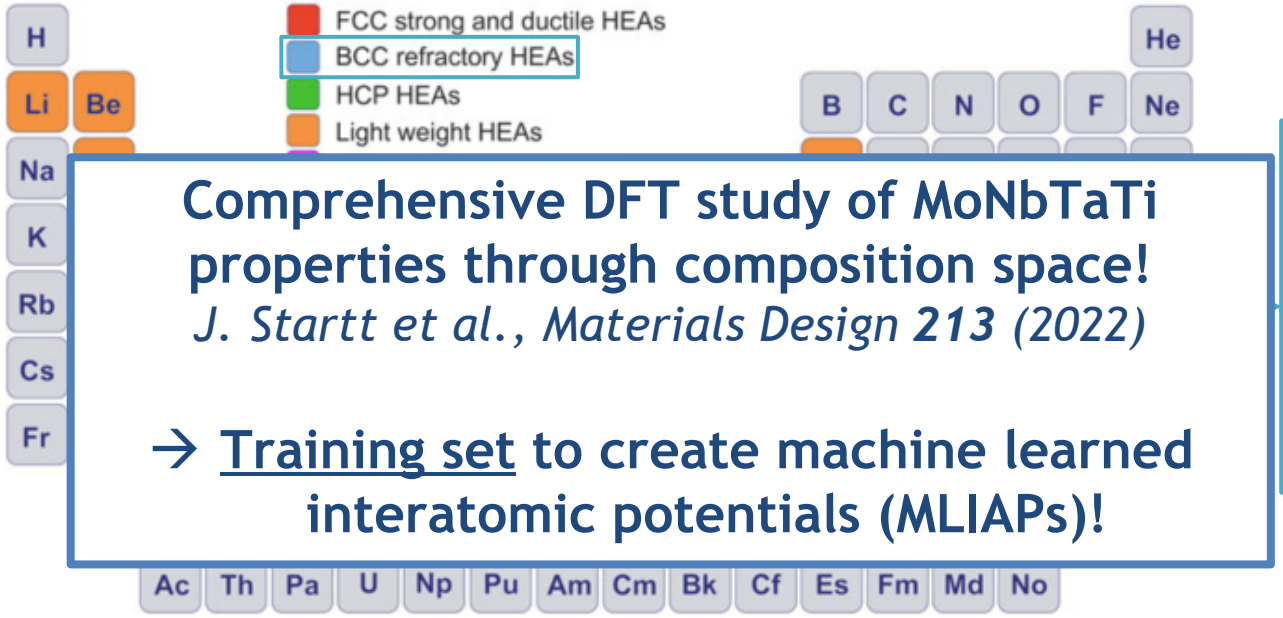
- Enormous design space, how can interatomic potentials help?
- Chemical transferability is paramount!

https://doi.org/10.21203/rs.2.15081/v1



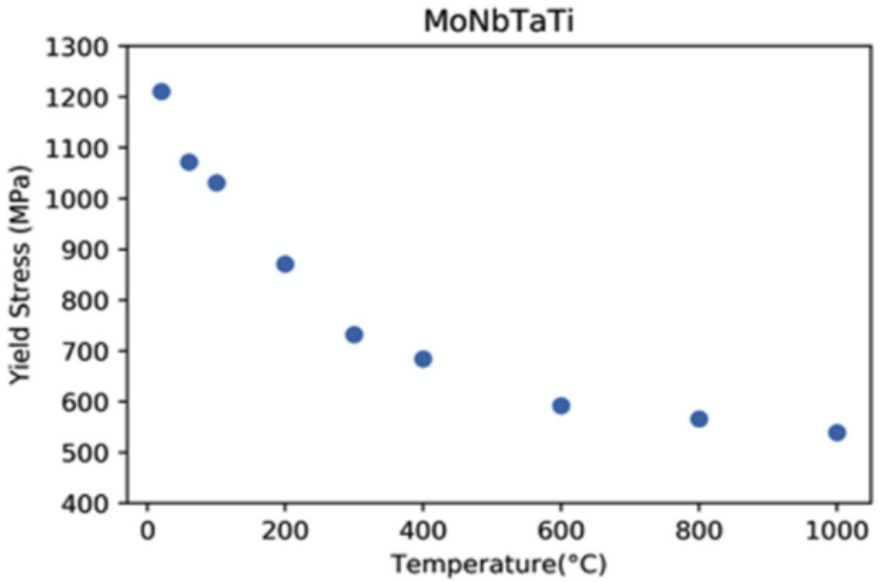
Conventional alloy

High-entropy alloy



Comprehensive DFT study of MoNbTaTi properties through composition space!
J. Startt et al., Materials Design ***213*** *(2022)*

→ **Training set to create machine learned interatomic potentials (MLIAPs)!**



Coury et al., Acta Mat. ***175*** *(2019)*

Li et al., J. Mater. Res., Vol. ***33****, (2018)*

Alloy Design From MD

Beyond DFT

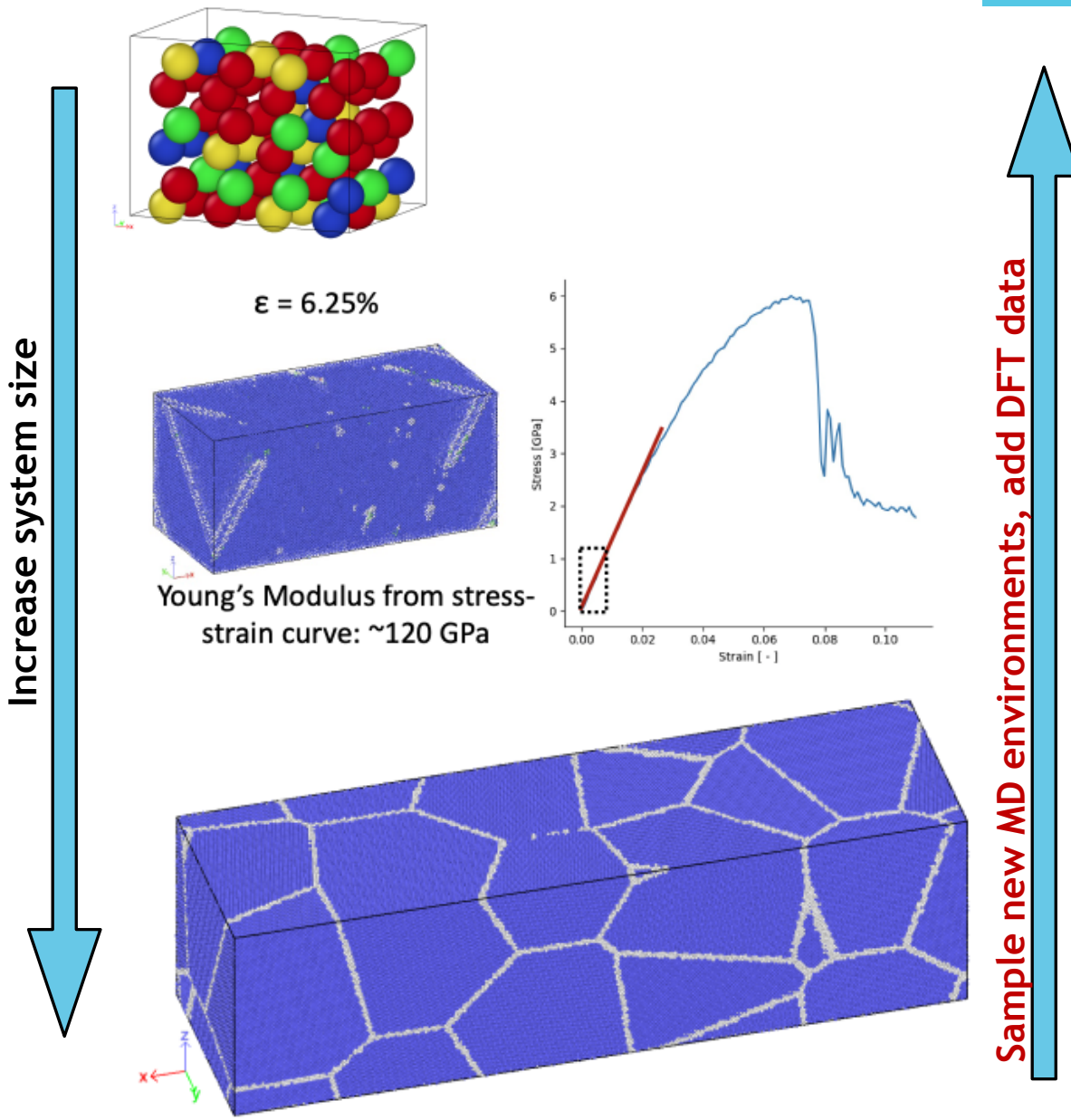
Untrained composition (vary 2 elements)

$Mo_{18.1}Nb_{23.6}Ta_{23.6}Ti_{34.7}$

Property	SNAP 2-element MLIAP	New DFT (not in training)	|SNAP - DFT|
C11 (GPa)	239.6	237.6	0.8 %
C12 (GPa)	143.6	129.7	10.7 %
C44 (GPa)	39.5	37.8	4.5 %
B (GPa)	175.6	165.7	6.0 %
G (GPa)	42.7	43.6	2.1 %
E (GPa)	118.5	120.1	1.3 %

*Startt et al., Materials & Design. **213** (2022)*

Active learning (AL) and uncertainty quantification (UQ) modules in development!

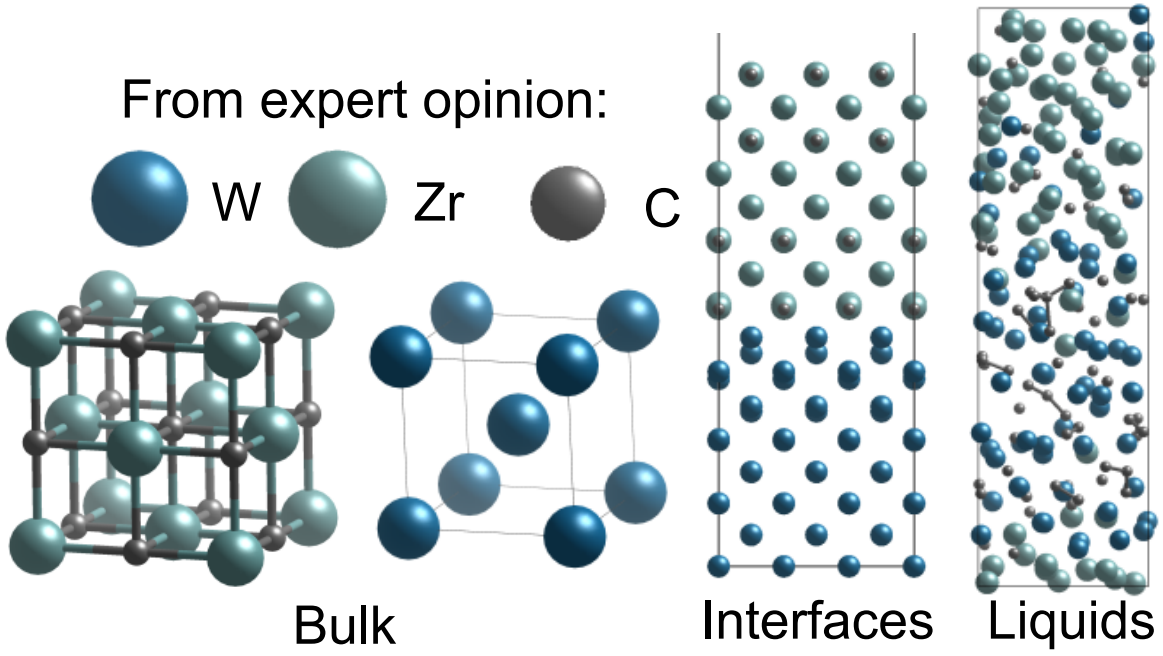


Active Learning

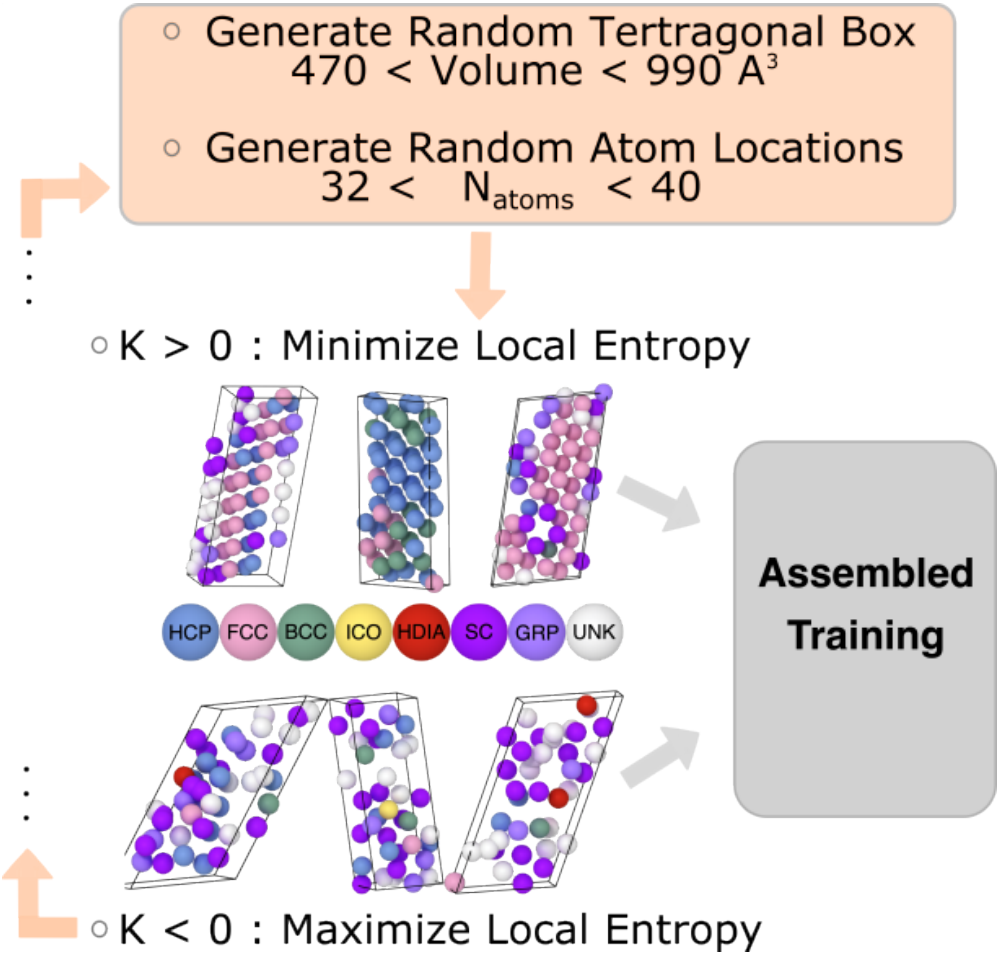
Unknown Unknowns

- Training structures can come from a variety of sources.
- ML-IAP are fit w.r.t. descriptors, not our physical intuition of what matters

From expert opinion:



From sampling the descriptor space:

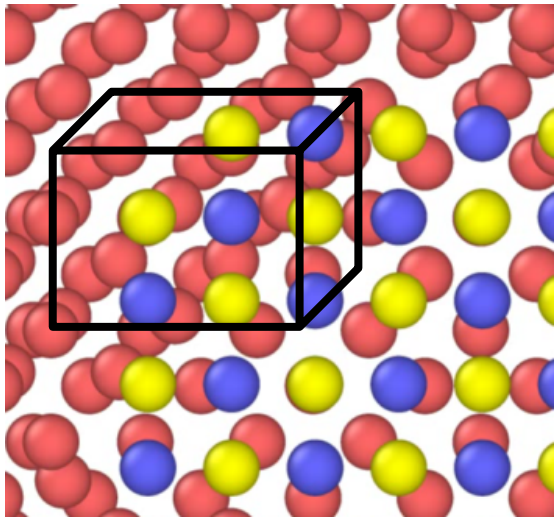


- ***But how to adapt to failures observed in MD***

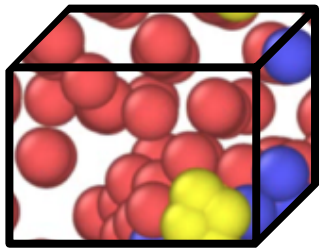
Active Learning

Hands-on Route:

Run MD

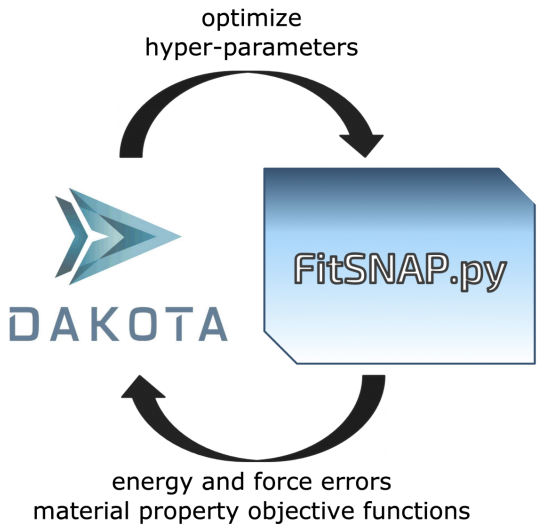


Carve out structures



Run DFT

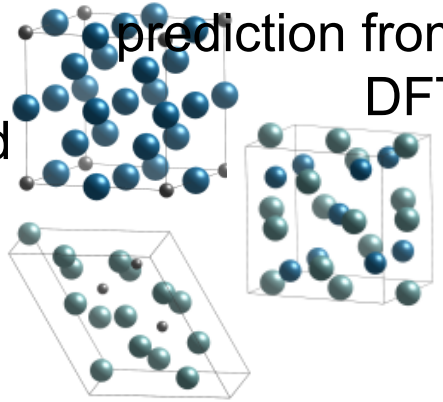
Re-Train ML-IAP

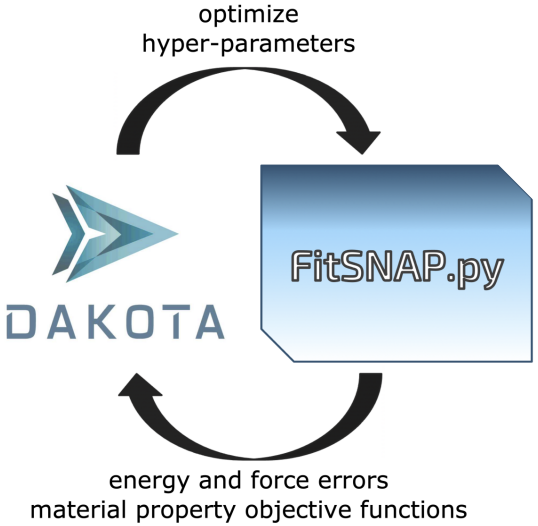


...

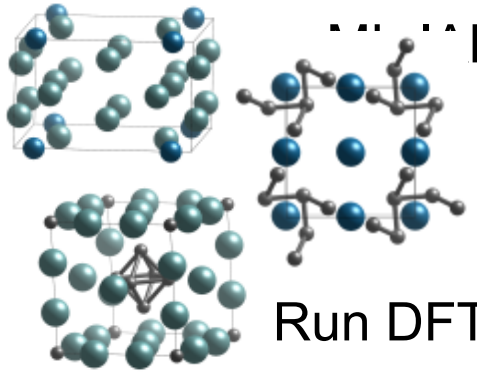
Automated (kind-of):

Structure prediction from DFT





Structure prediction from ML-IAP



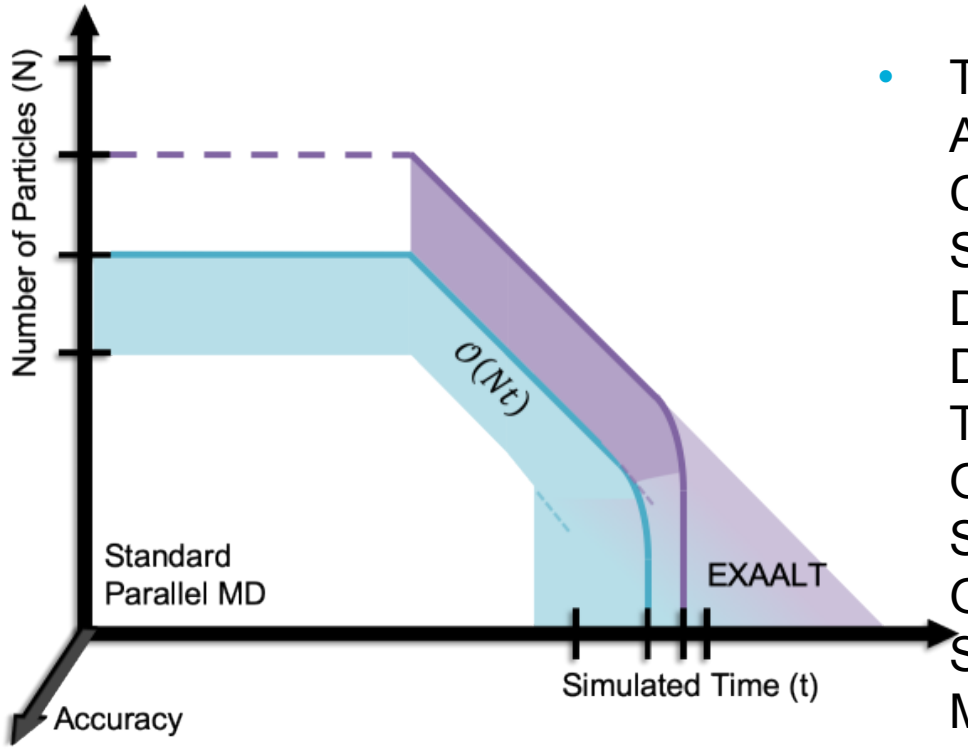
Run DFT

...

USPEX : A. Oganov et al. / J. Chem. Phys. **124,** 244704 (2006)

Conclusions and Path Forward

- Data-driven interatomic potentials (SNAP, SNAP-NN) allow for MD predictions of challenging material problems.
- While harder to quantify, the fidelity of our MD simulations needs to be a key consideration



- Thank you to all my collaborators: Aidan Thompson, Mary Alice Cusentino, Krupa Ramasesha, Svetoslav Nikolov, Charlie Sievers, David Montes do Oca Zapian, Danny Perez, Nick Lubbers, Julien Tranchida, Steve Plimpton, Ivan Oleynik, Jon Willman, Ember Sikorski, Megan McCarthy, James Goff, Drew Rohskopf, James Stewart, Carlos Pereyra, Nat Trask, Michael Sakano, and many others!

Contact Information:

mitwood@sandia.gov

EXASCALE COMPUTING PROJECT

Los Alamos NATIONAL LABORATORY EST. 1943

Sandia National Laboratories

NVIDIA

U.S. DEPARTMENT OF ENERGY | Office of Science

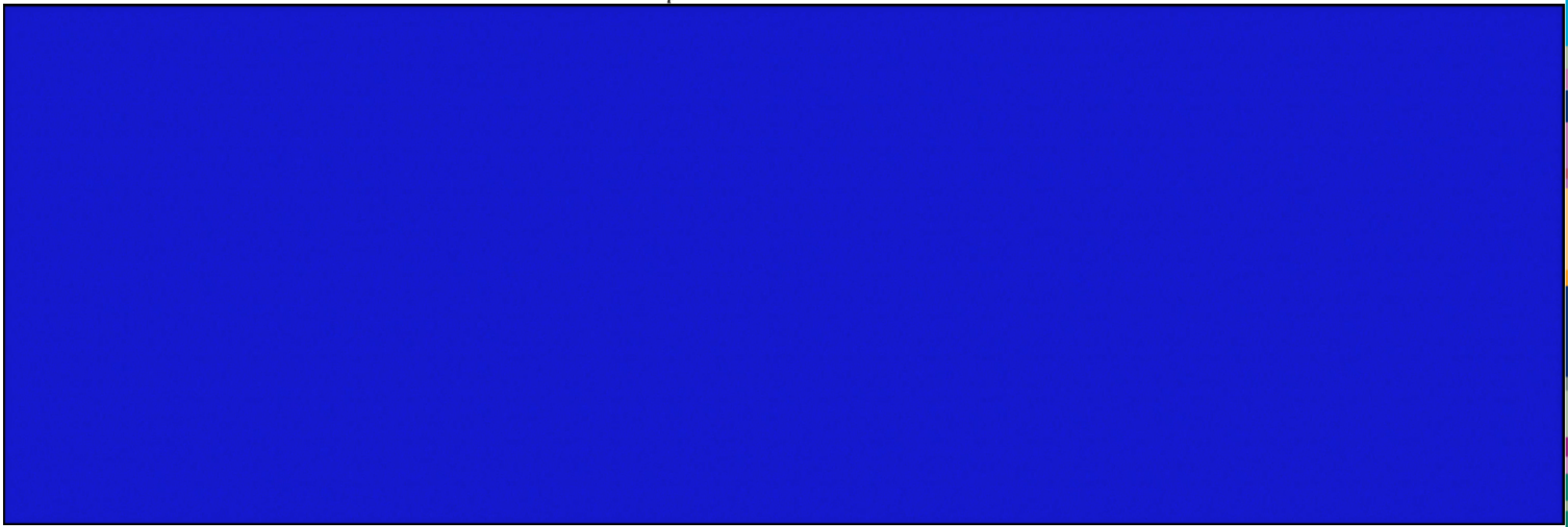
OAK RIDGE National Laboratory

NERSC

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