SAND2023-06577C

Sandia National Laboratories

Exascale Ready Molecular Dynamics Simulations With LAMMPS; Application to Fluid Instabilities at Liquid-Vapor Coexistence

Mitchell Wood

Center for Computing Research, Sandia National Labs

DSFD 2023 Meeting

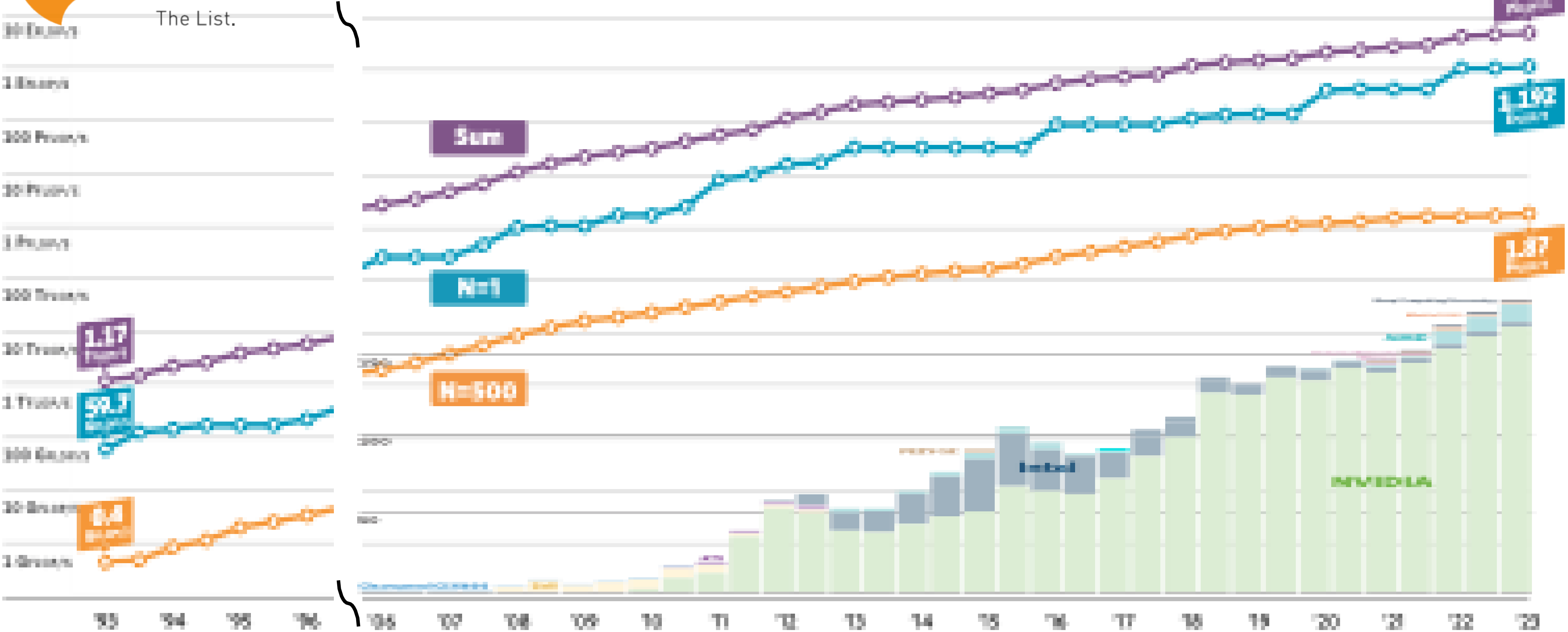
Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LL
U.S
under contract DE-NA0003525.

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.

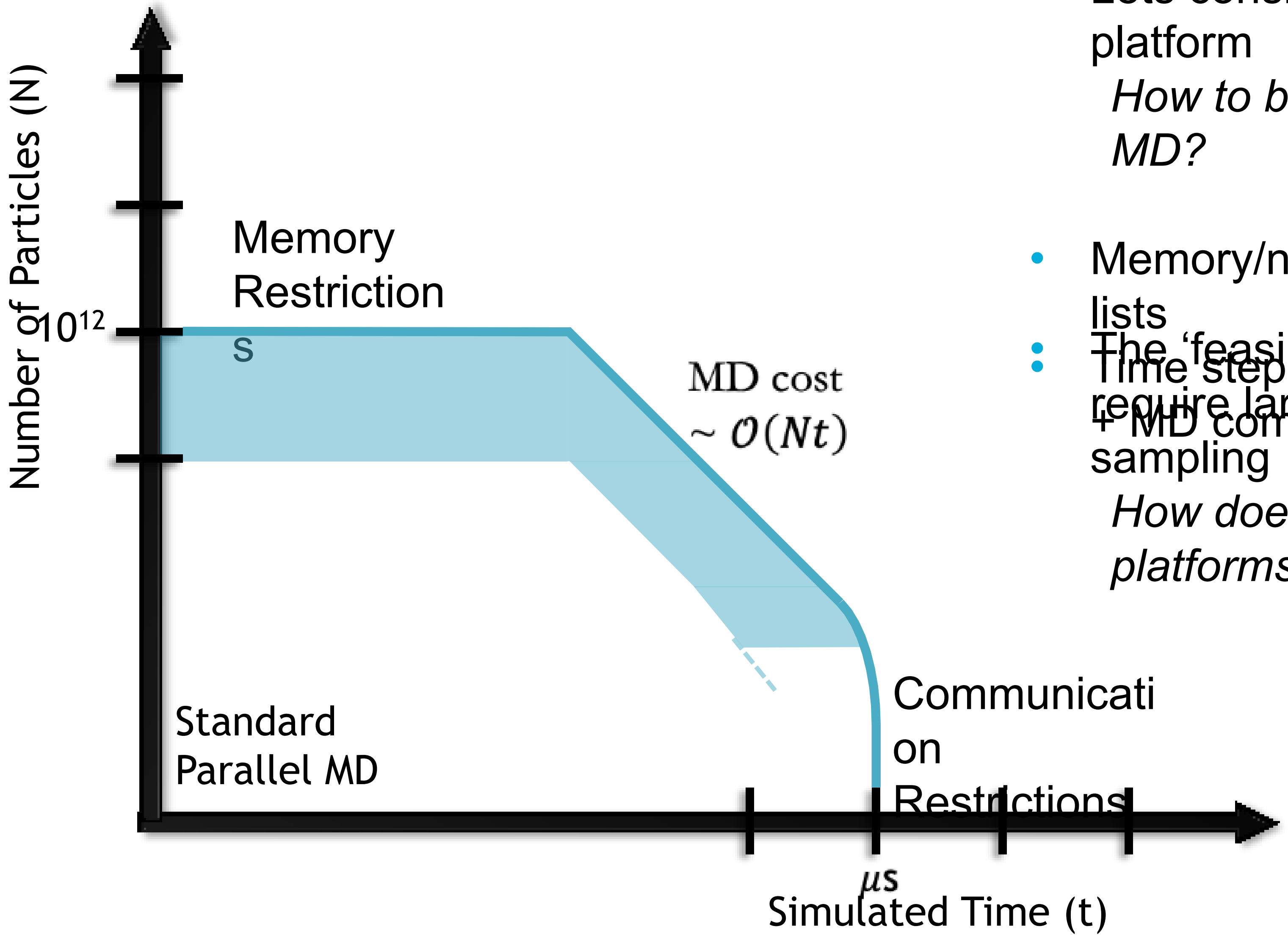
Computing Status Report

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

TOP500 The List.



What is possible for MD at the Exascale?



- Lets consider a 24hr allocation on a leadership platform
  How to best spend this computational budget on MD?
- Memory/node ~ particles/processor + neighbor lists
- Time stepping overhead from network bandwidth + MD comm pattern.
- The 'feasibility envelope' favors problems that require large atom counts over long time sampling
  How does this affect the research done on these platforms?

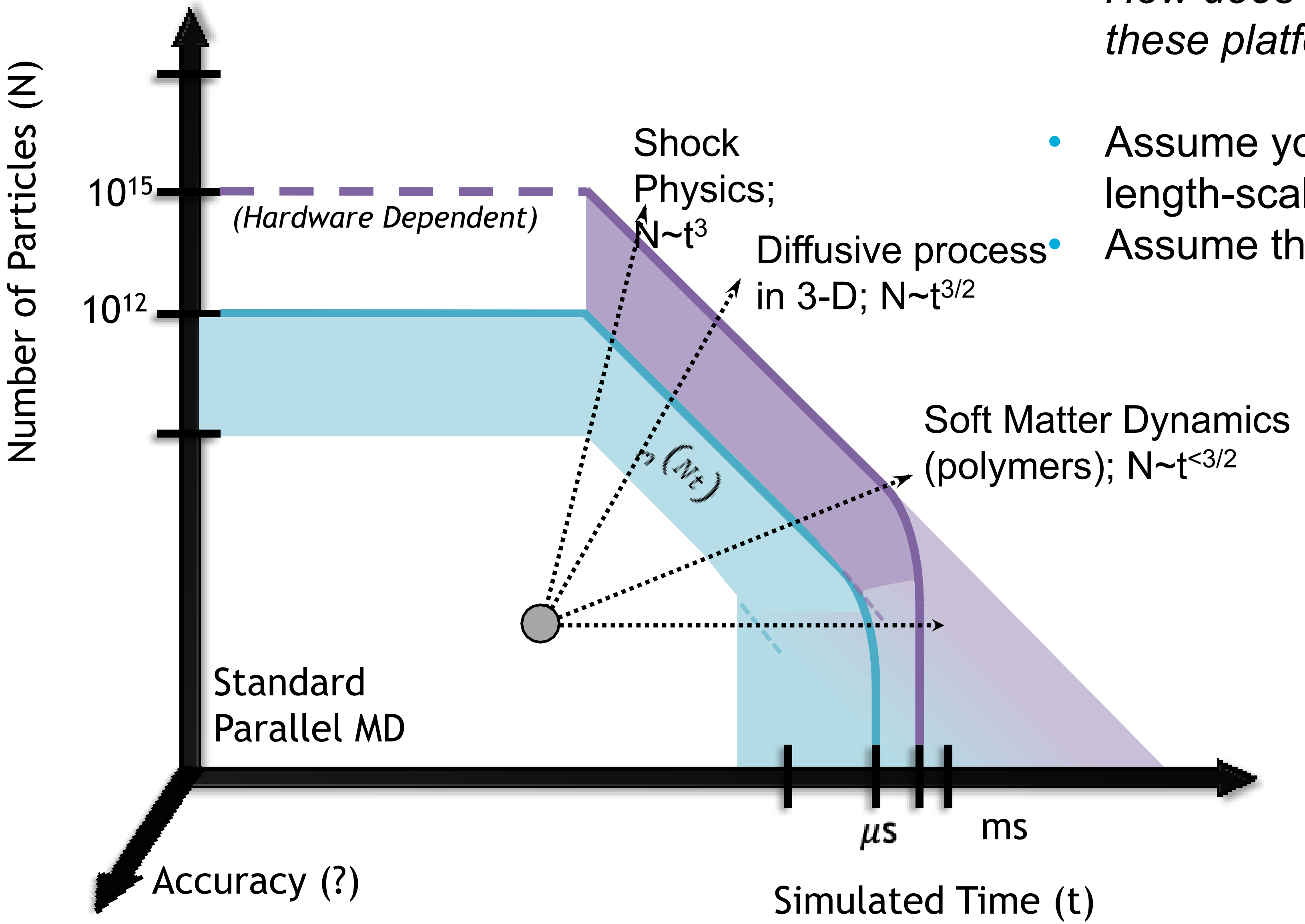
Newton's Equations:

$$\frac{d\mathbf{r}_i}{dt} = \mathbf{v}_i$$

$$\frac{d\mathbf{v}_i}{dt} = \frac{\mathbf{F}_i}{m_i}$$

$$\mathbf{F}_i = -\frac{d}{d\mathbf{r}_i} V\left(\mathbf{r}^N\right)$$

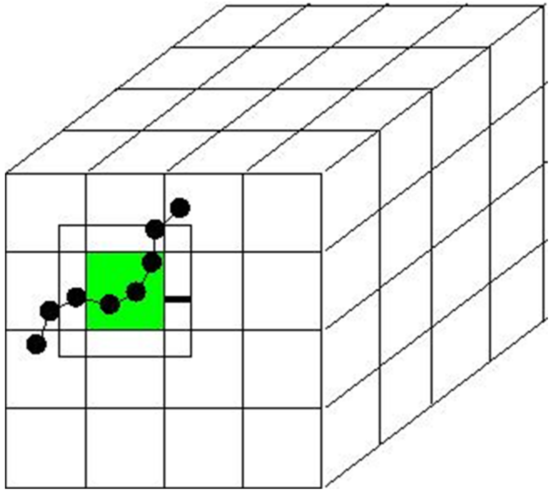
The Master Plot



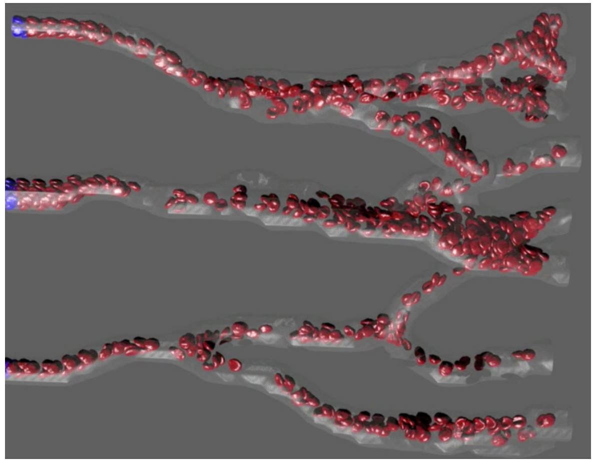
- *How does this affect the research done on these platforms?*
- Assume your problem has some specified length-scale dependence : N~L^{α}
- Assume the associated time-scale goes as t~L^{γ}
- But what if you care about a system governed by rare event dynamics?

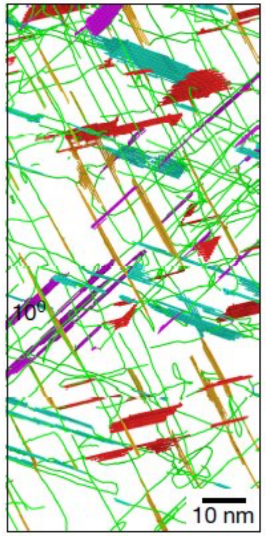
What is LAMMPS?

- Large-scale Atomic/Molecular Massively Parallel Simulator
  http://lammps.sandia.gov
- Open source, highly portable C++, free under GPL license
- Well documented with many examples, easily extendable for user specific
- Variety of boundary conditions, constraints, ensemble sampling methods
- Parallelism through spatial decomposition of simulation domain
- Short and long ranged interactions allowed/included
- CPU cost is (N/P) and communication is $(N/P)^{2/3}$



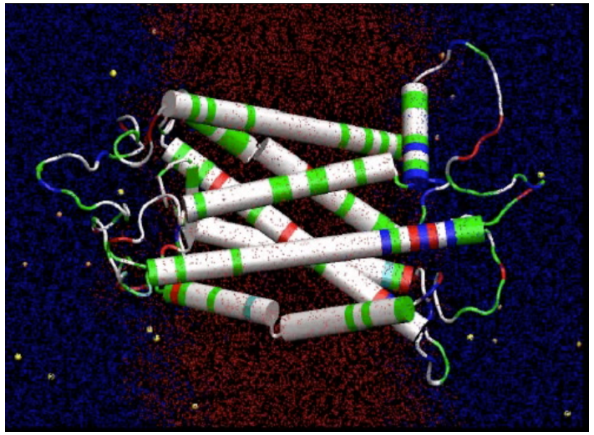
Blood Flow



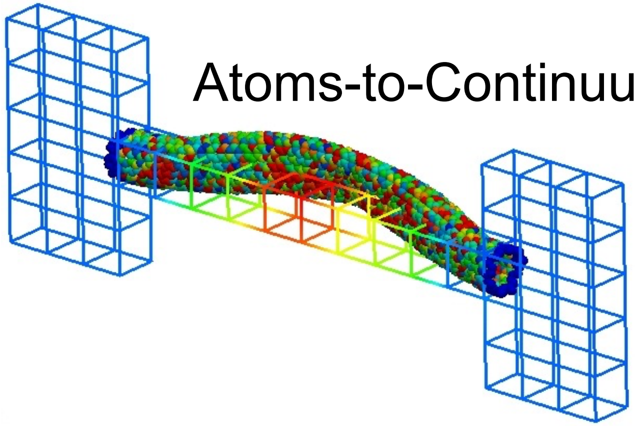


Dislocations in Materials

Proteins and Biophysics

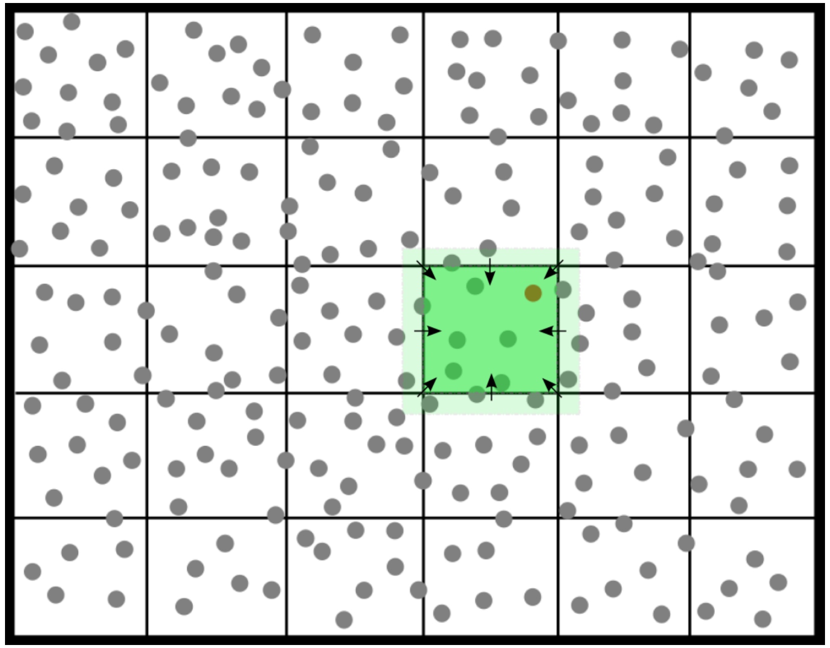


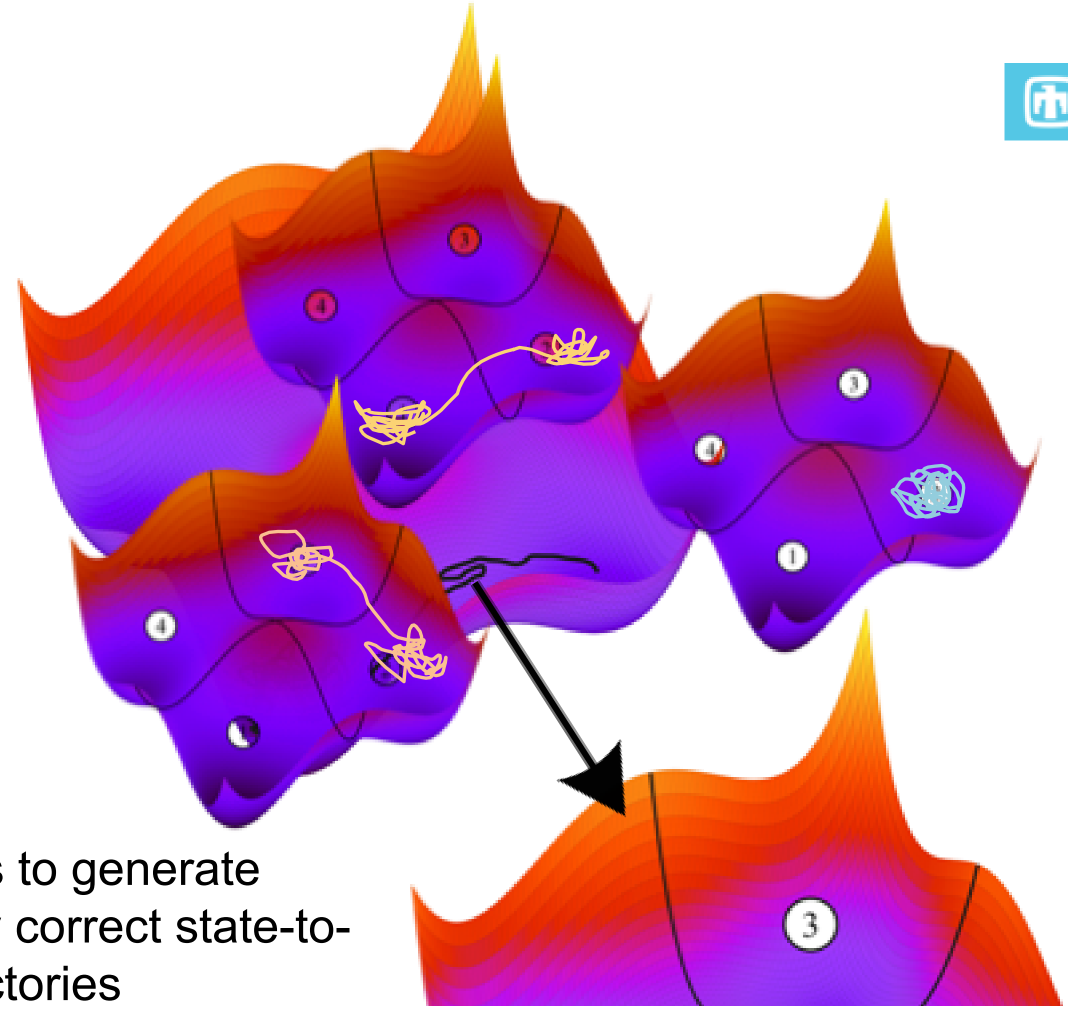
Atoms-to-Continuum



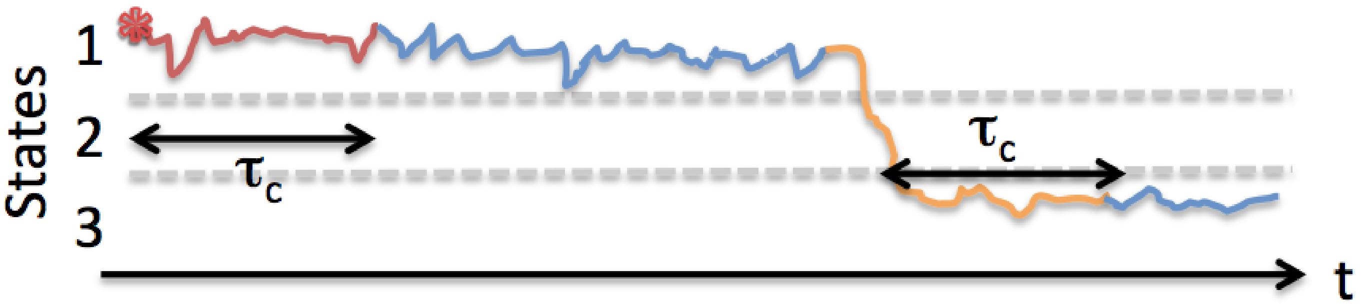
Parallel in Space, Time

- Atoms/particles in space can be distributed across processors
- Need to track particles in nearby domains, reconstruct neighbor lists as particles move
- for all time;
  <u>Compute forces</u>, update atom positions





- The goal is to generate statistically correct state-to-state trajectories



[Le Bris, Lelievre, Luskin, and Perez, MCMA 18, 119

MD Approximations Change Over Time

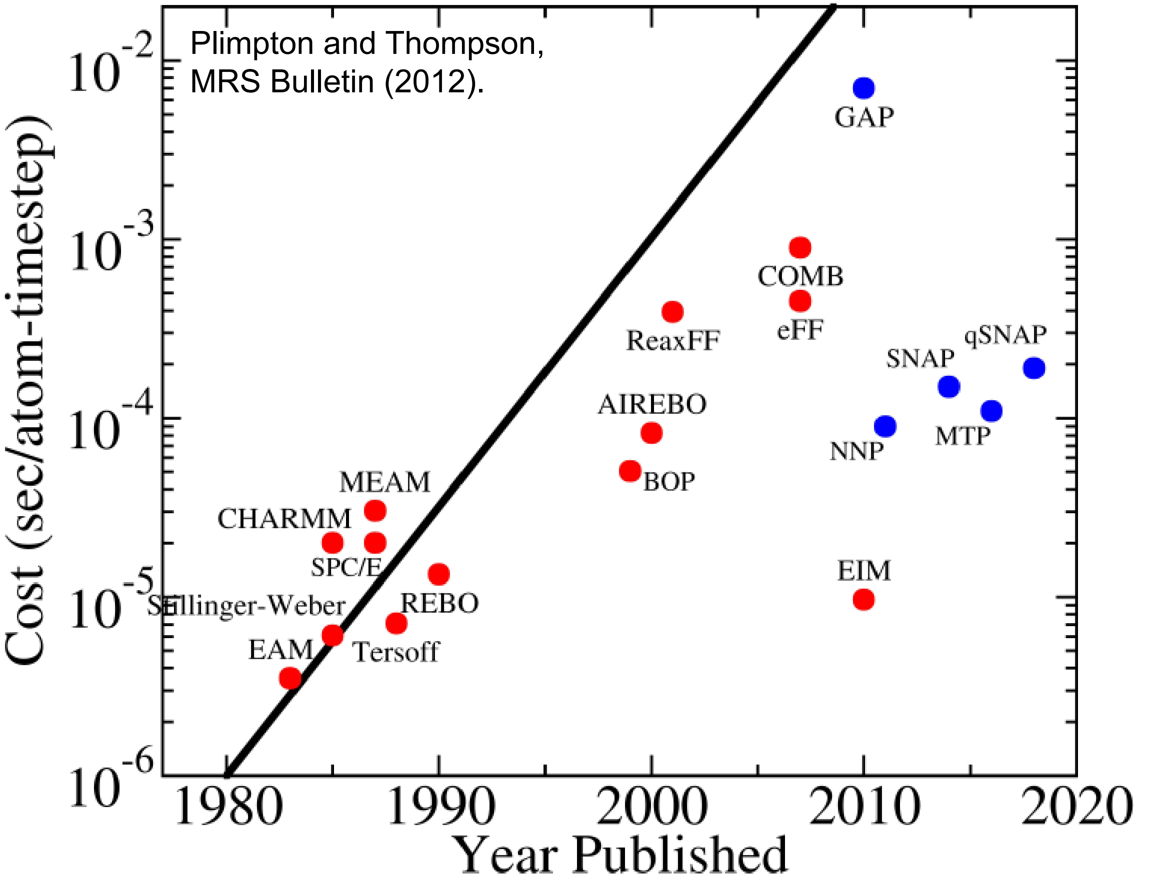
http://lammps.sandia.gov

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

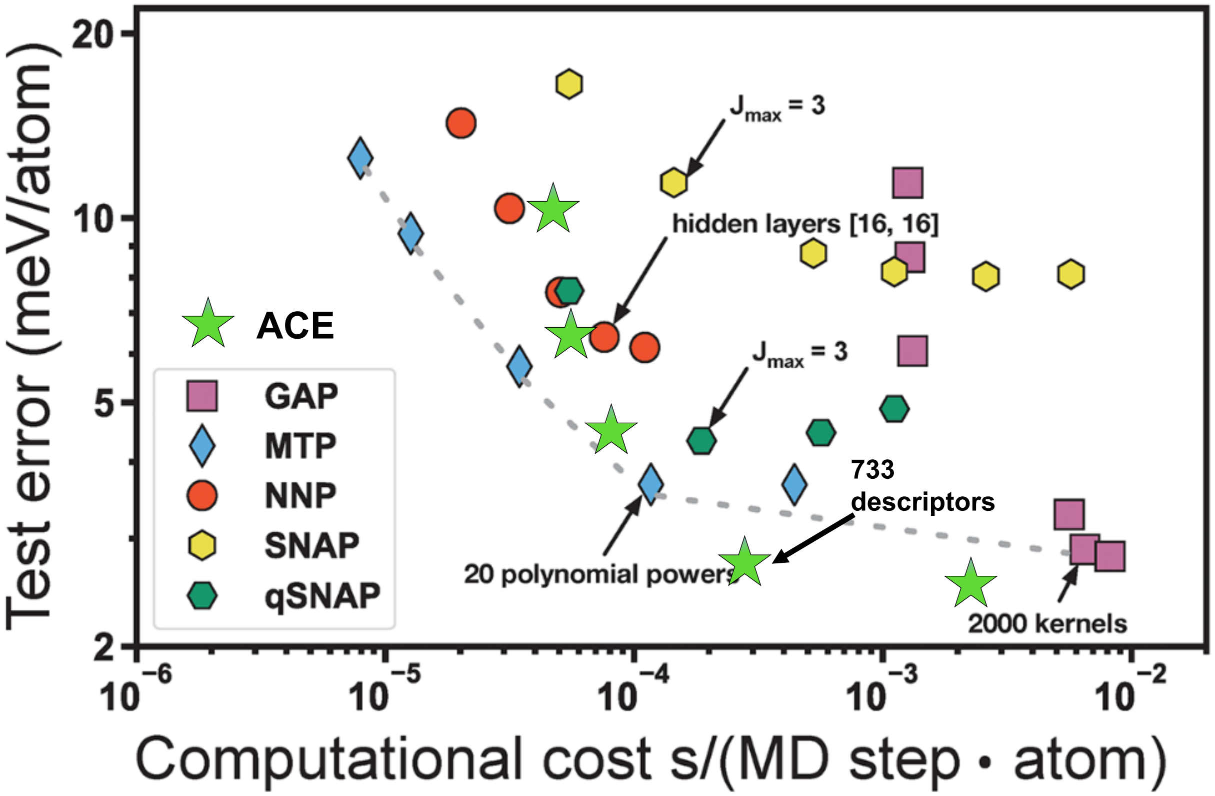
Manybody (1980s)
Stillinger-Weber, Tersoff, Embedded Atom Method

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

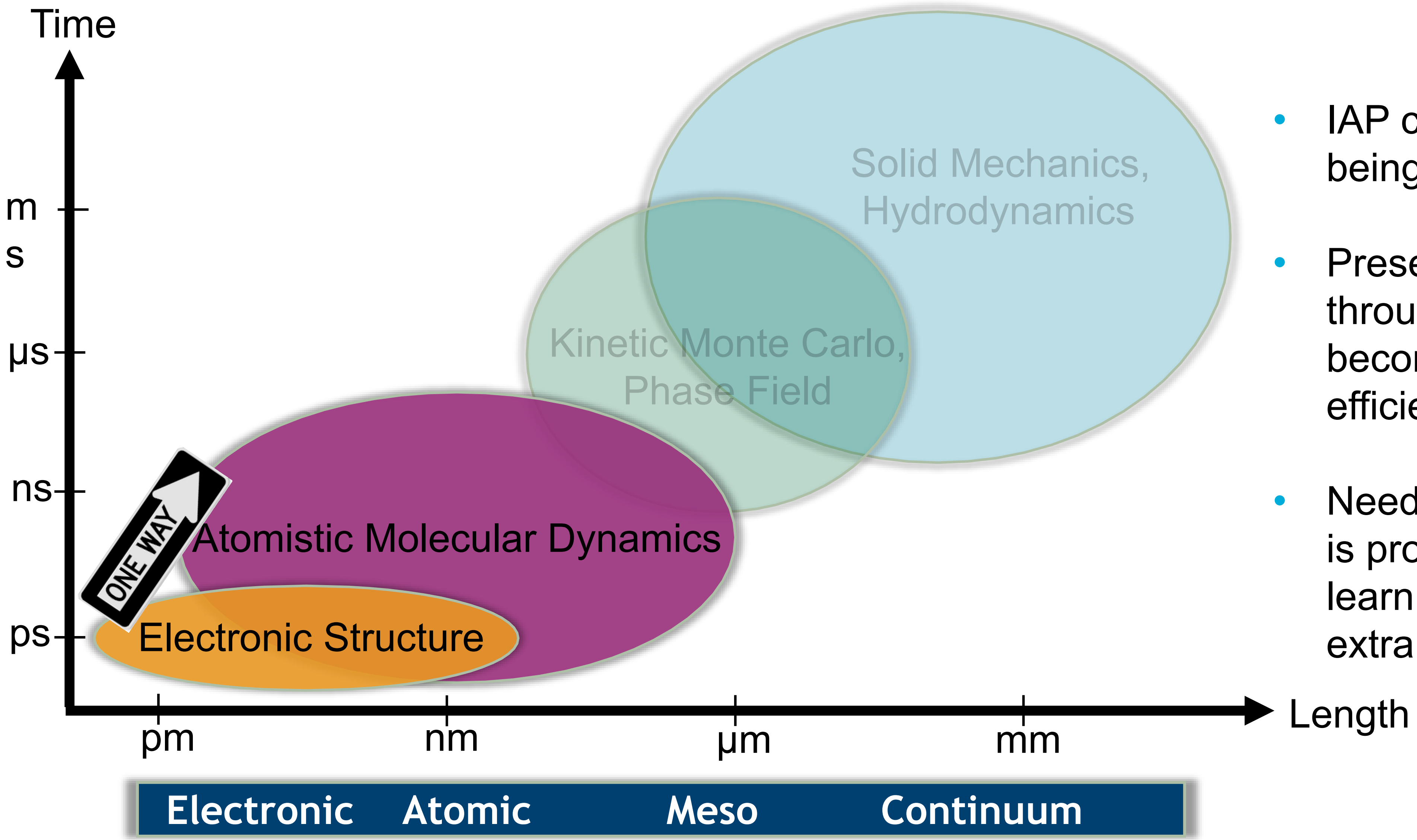
Big Data / Deep / Machine Learning (2010s)
GAP, SNAP, NN,…



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

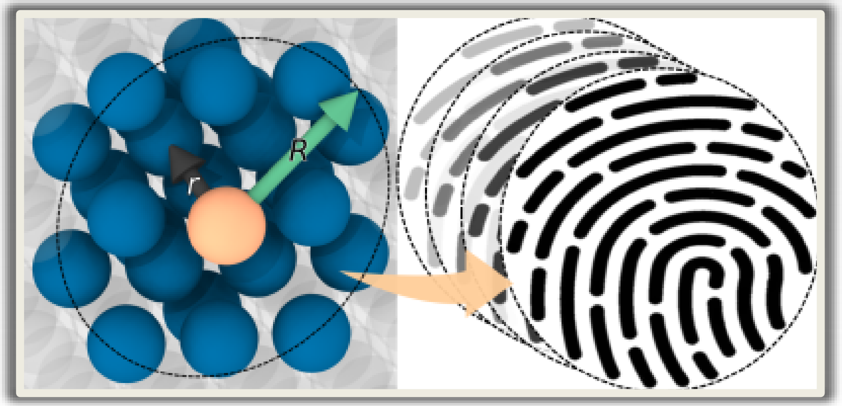
Machine Learning to Bridge Electronic-Atomic

Classical, Empirical Potentials

- Metals
  - EAM: Assume spherical electron density
    $E_i = F_\alpha\left(\sum_{j\neq i}\rho_\beta(r_{ij})\right) + \frac{1}{2}\sum_{j\neq i}\phi_{\alpha\beta}(r_{ij})$
- 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



Components of ML-MD

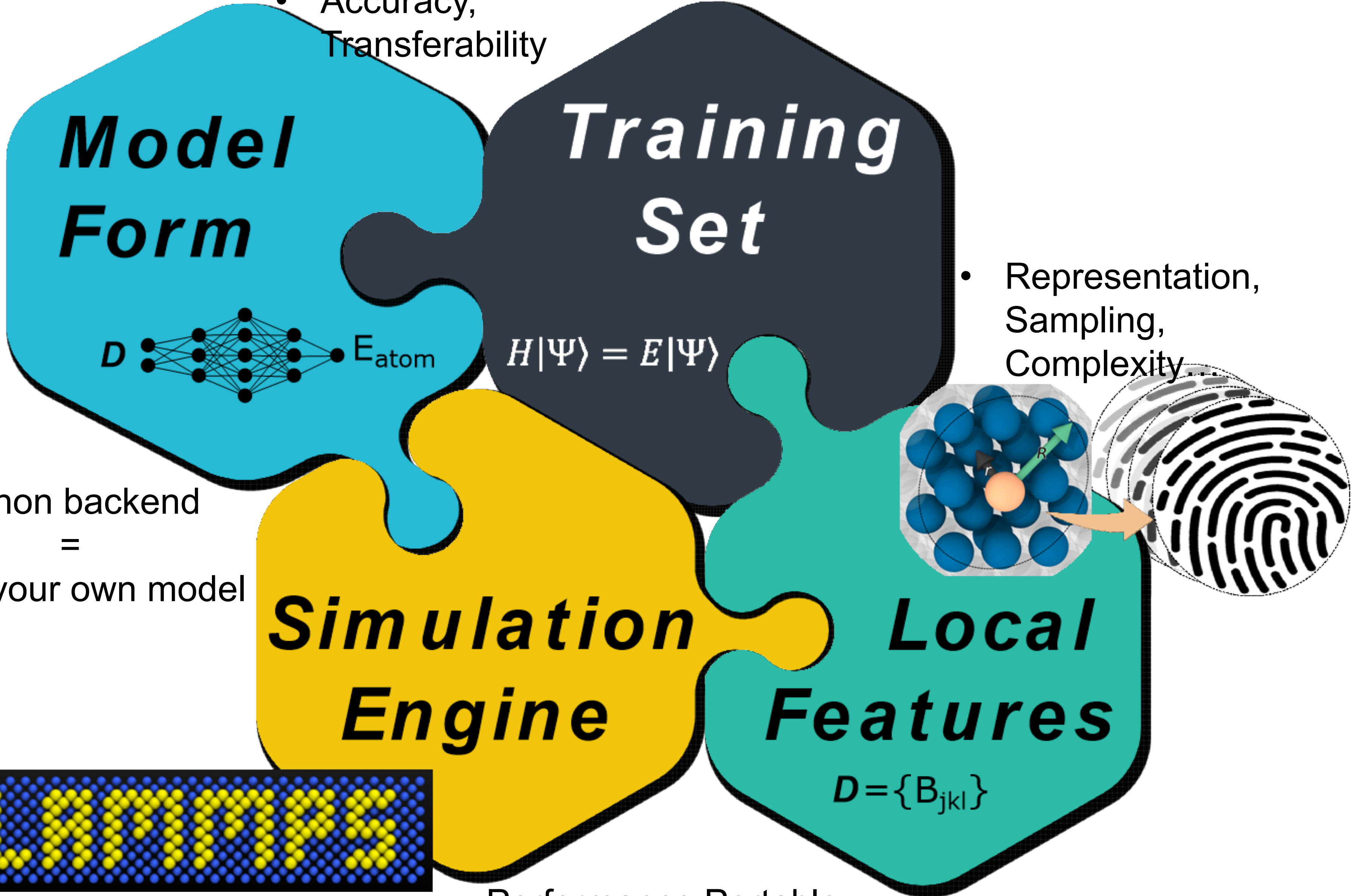
https://github.com/FitSNAP/FitSNAP

http://lammps.sandia.gov

- Accuracy, Transferability

- Representation, Sampling, Complexity…

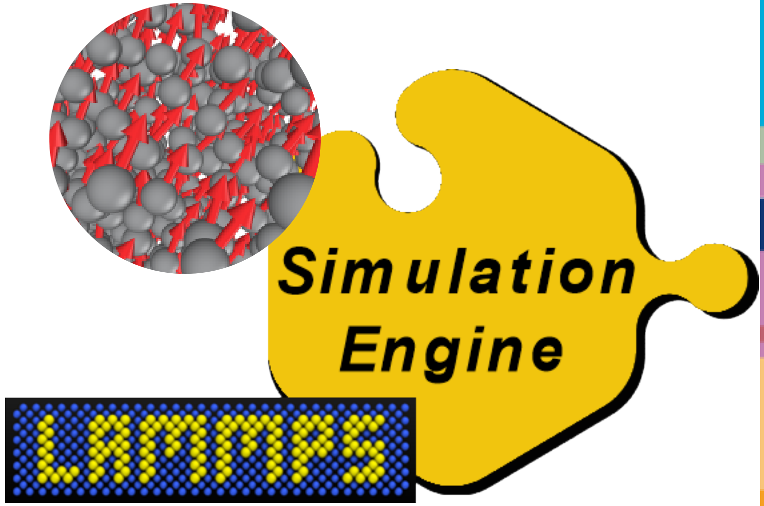
- Python backend = Bring your own model

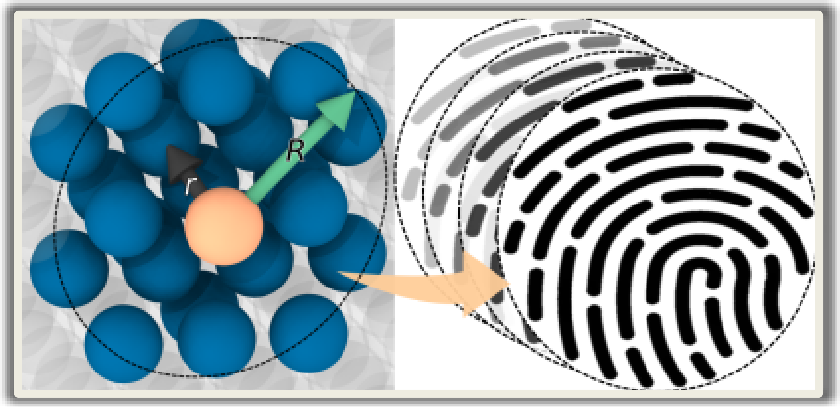


- Performance Portable

Unified Framework for MLIAP

- Provide a common API for many methods
  - Descriptor generates local fingerprint for each atom
  - Model computes energy as function of descriptors
  - Data handles LAMMPS interface and intermediate quantities e.g. gradients
- Descriptor and Model insulated from LAMMPS and each other
- **Allows mix-and-matching of Models and Descriptors**



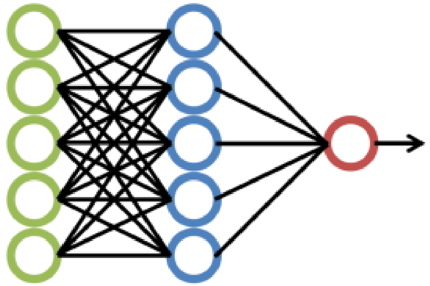


Descriptors

pair_MLIAP

Inputs	*Outputs*
Atoms	*Energy*
Elements	*Force*
Neighbors	*Stress*

Data



Model

Accelerating Model Development

A. Rohskopf (2023), Journal of Open Source Software

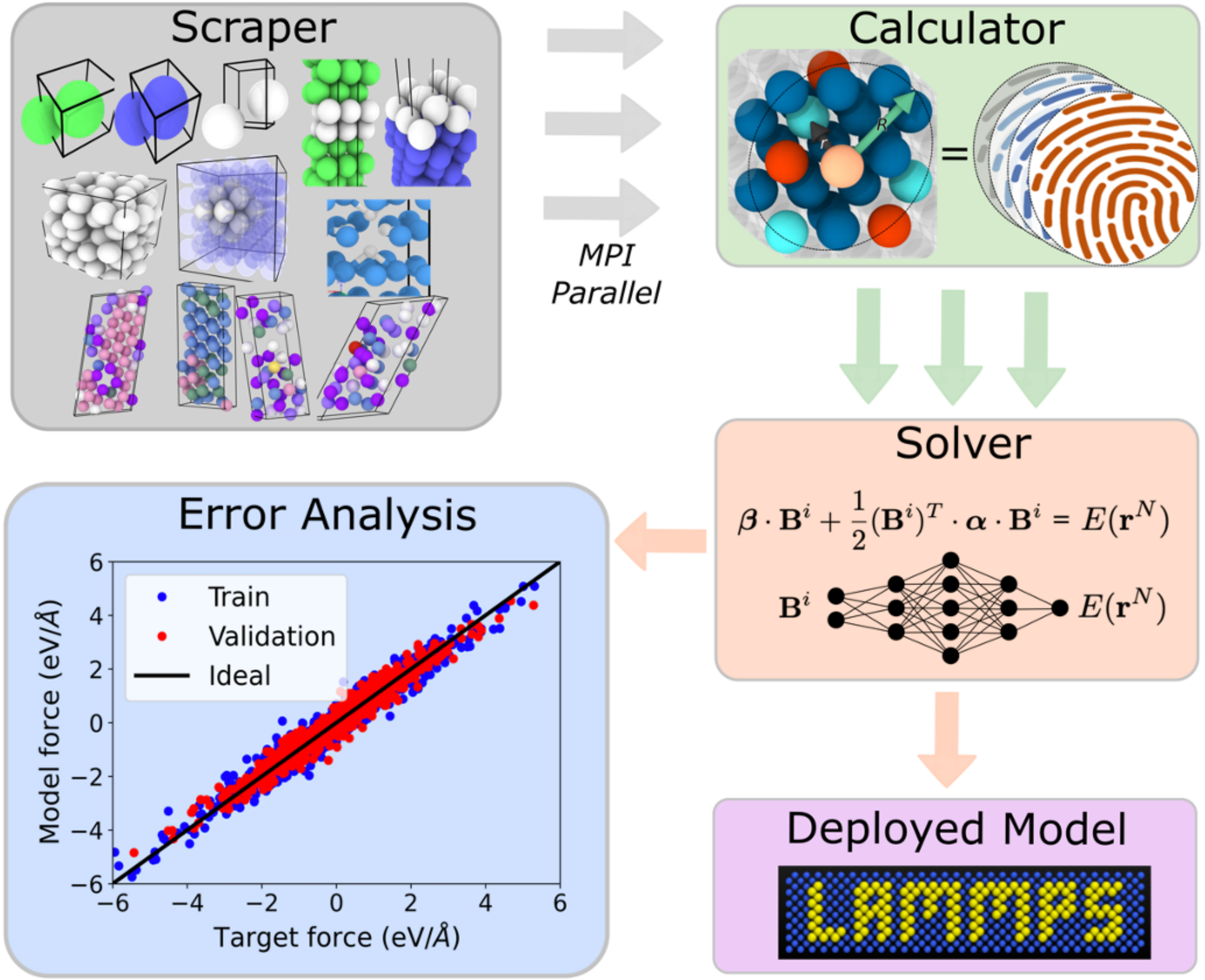
github.com/fitsnap/FitSNAP

Modular Code Structure

- Three main classes : Scrape, Calculate, Solve

- **Scraper** : Collects ground truth values from files on disk → (stores in dataframe)
- **Calculator** : Converts atomic structures into set of descriptors → (stores in dataframe)
- **Solver** : Performs regression commensurate with model form

- Adding functionality does not disrupt code flow because of object oriented structure
- Classes, and items thereof, can be called



Quick Install :

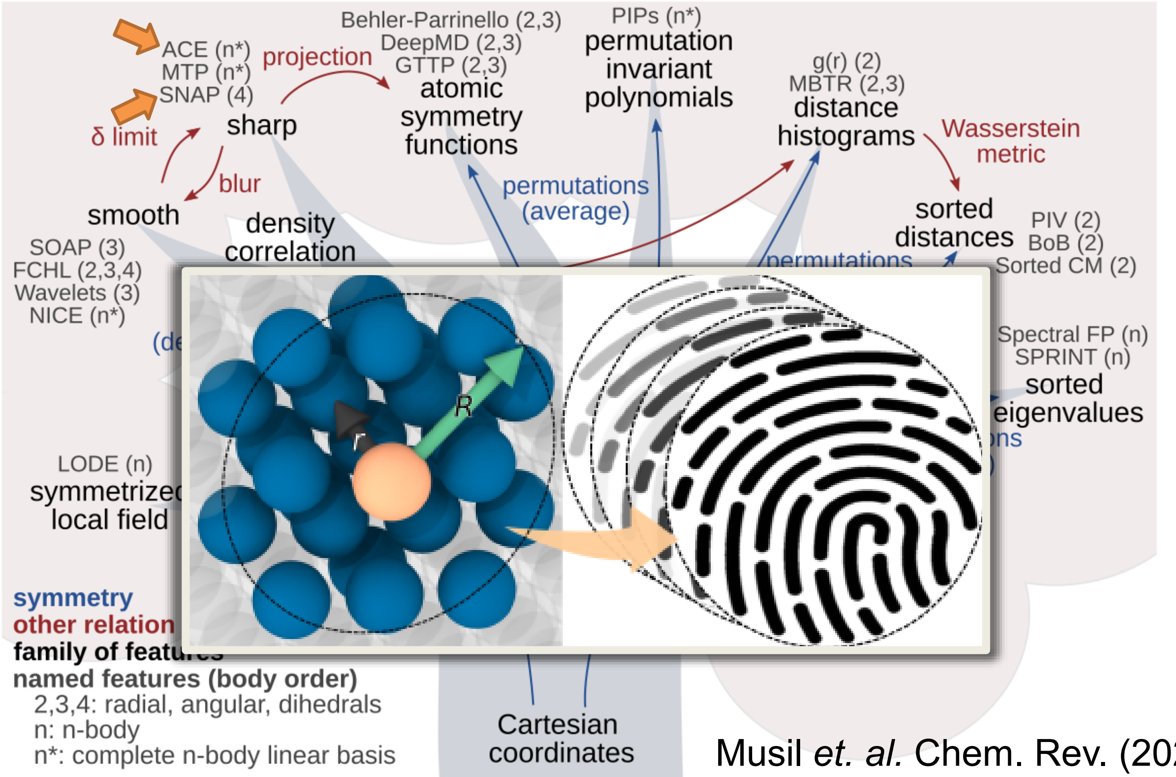
conda install -c conda-forge lammps

Calculator – Descriptor Sets

J. Goff (2022), arxiv.org/abs/2208.01756

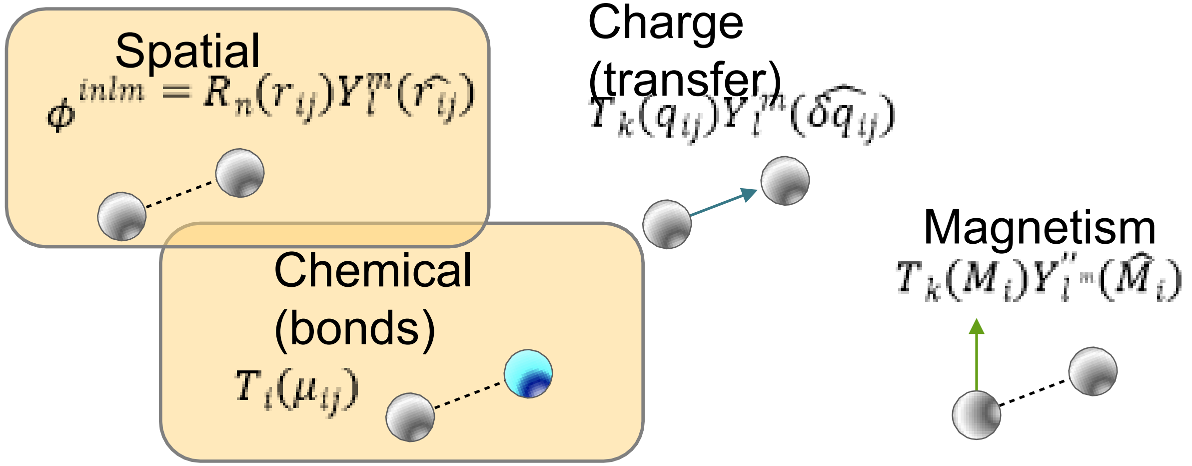
LAMMPS Breakdown

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

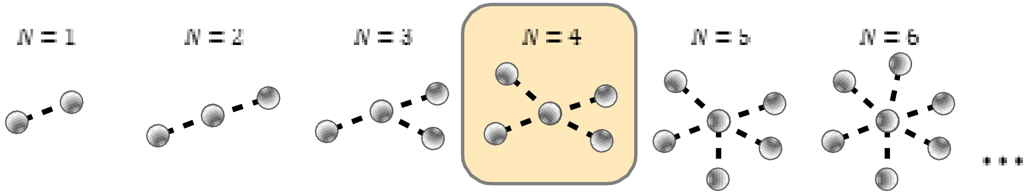


Musil *et. al.* Chem. Rev. (2021) 121,

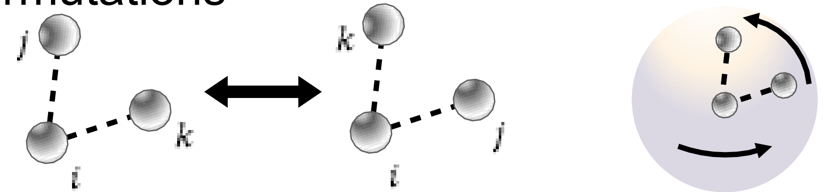
Complete, generalizable single-bond basis



Form a complete N-bond basis



Impose invariance w.r.t. rotations and permutations



Descriptor Improvements and Scalability

Flexible Model Form

- Energy (and forces) can be expressed as higher moments of the bispectrum (B_k^i)

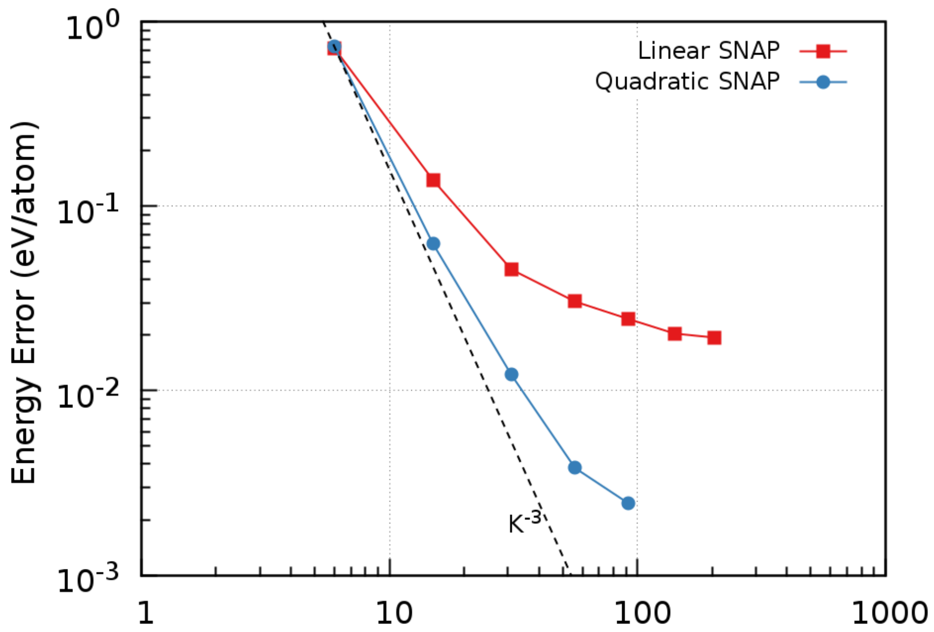
$$E_{SNAP}^i = \alpha_0 + \sum_k \left[\alpha_k^{(1)} \left(B_k^i - B_{k_0} \right) + \alpha_k^{(n)} (\ldots)^n \right]$$

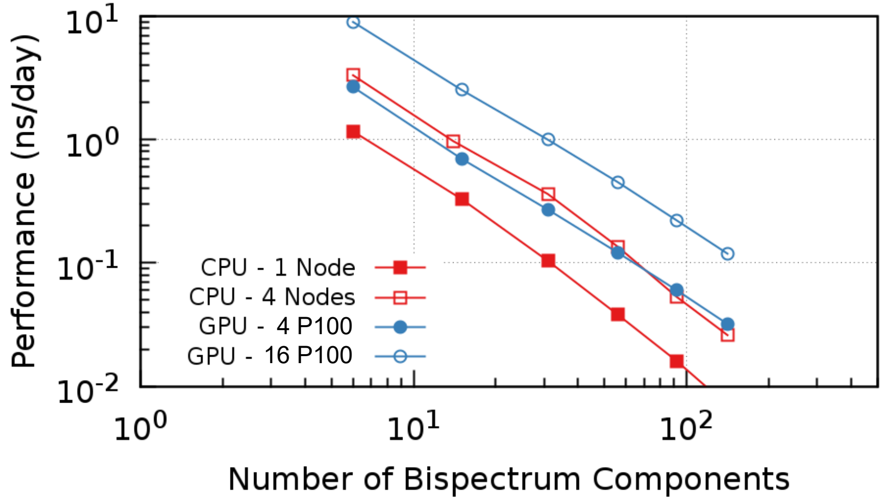
- Linear regression kernel can still be used for $\alpha_k^{(n)}$

Accuracy-Cost Tradeoff

- Cost of higher moment expansion is much cheaper than extending the sum on k. Accuracy gains either way.

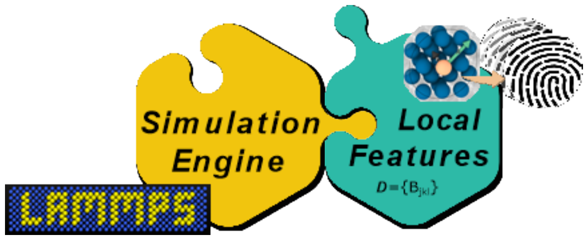
- GPU portable, 1 NVIDIA P100 ~1 Intel Dual-Broadwell





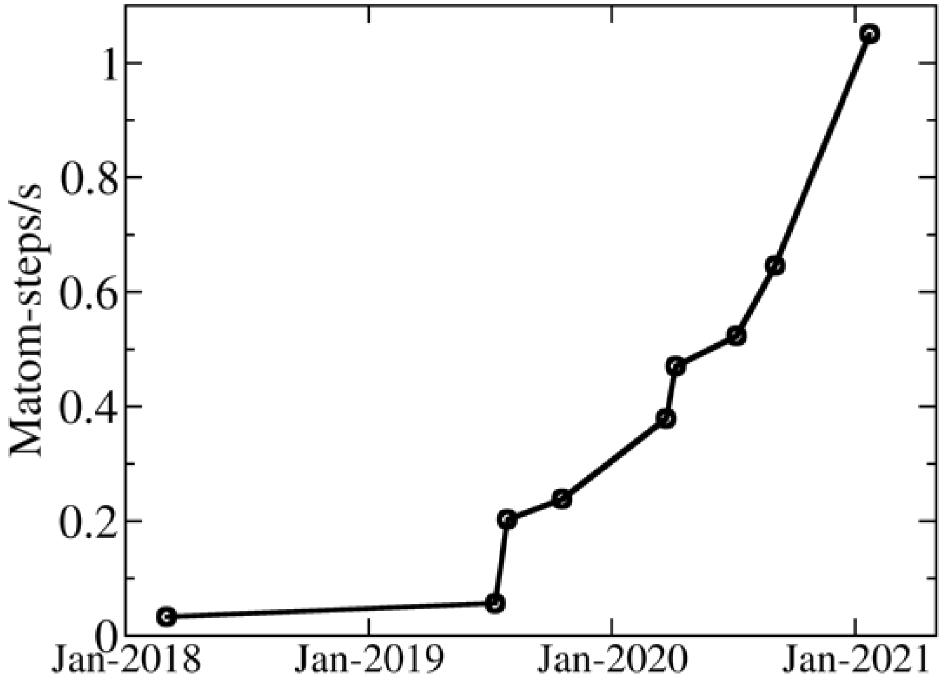
Exascale Ready Models

Breakdown of timing:



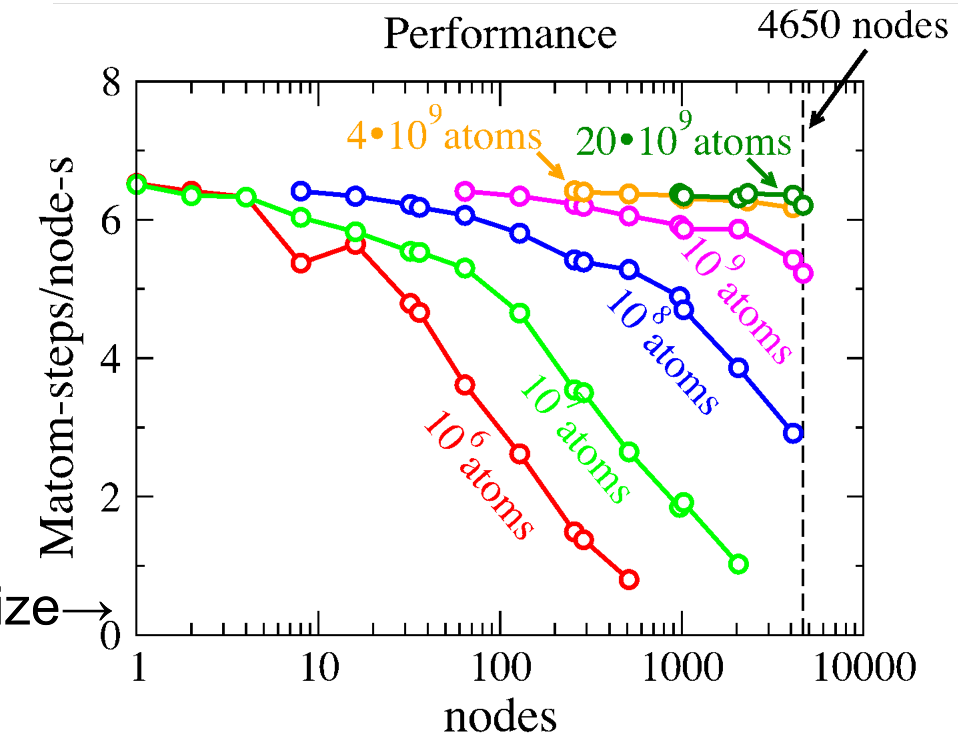
SNAP Performance

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



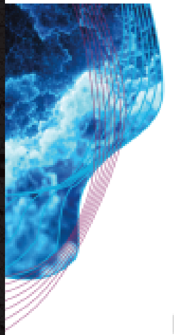
←Fixed Problem Size

Variable Problem Size→



Some Good Publicity

Where continuum theory breaks down, MD to the rescue!



LARGE-SCALE ATOMISTIC SIMULATIONS INVESTIGATE EXPANSION OF MOLTEN METAL

Team: Stan Moore, Mitchell Wood, Kyle Cochrane, John Carpenter, Aidan Thompson
Contributing Writer: Neal Singer

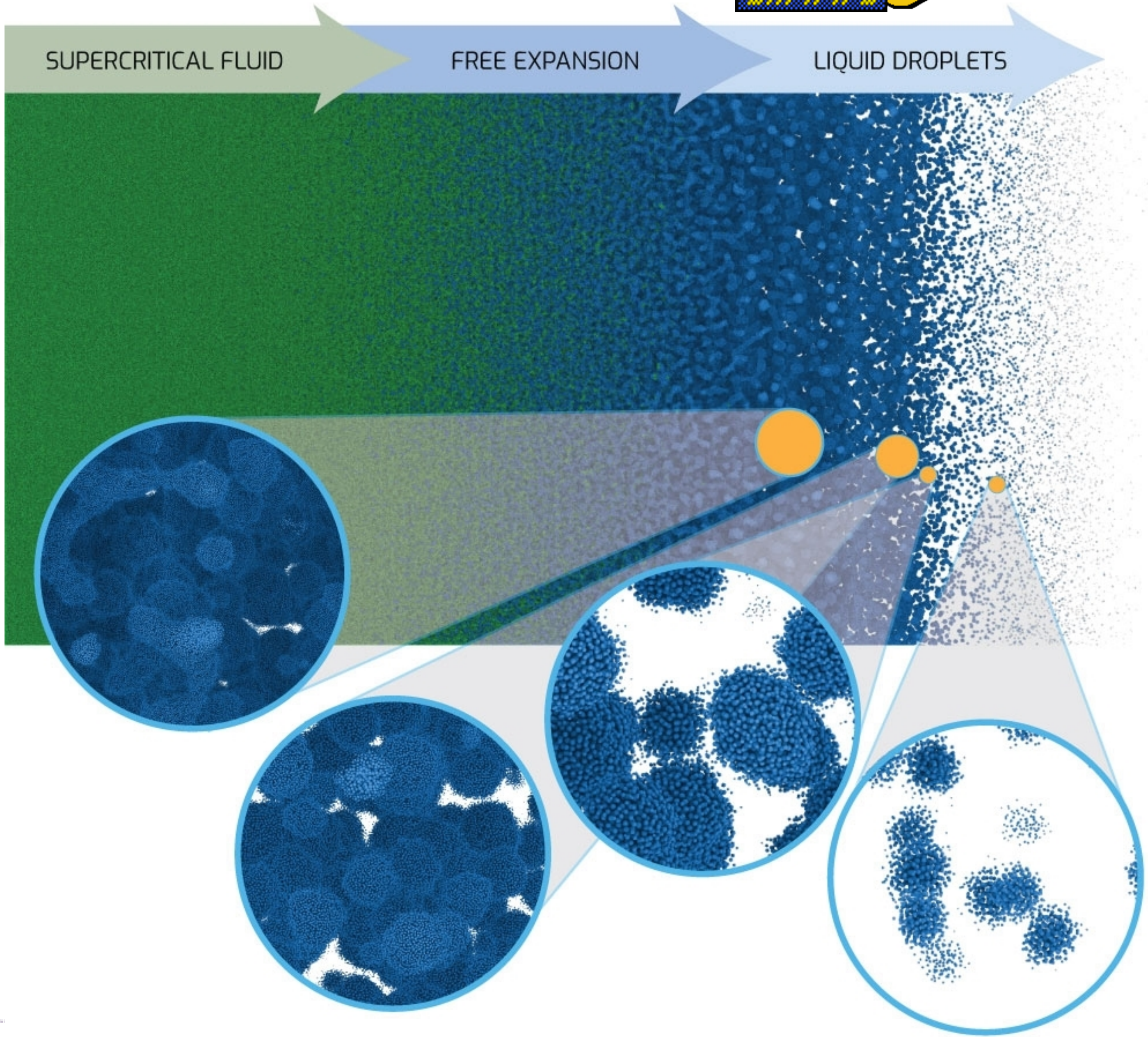
Can we visualize the moment between liquid and vapor at the atomic level?

As computing power grows, materials science researchers are developing higher fidelity models to capture smaller length and briefer time scales — where the action takes place — of materials under stress.

Ultimate accuracy in modeling at the atomic level comes from quantum chemistry calculations such as density functional theory (DFT), which solve Schrödinger's equation and simulate interactions between electrons. However, these methods don't scale well with large number of atoms; quantum system sizes are typically limited to less than a thousand atoms.

...ferred method called molecular dynamics (MD) ...'s equations of motion to describe atom behavior. MD tracks the motion of billions of atoms by repeatedly calculating the atomic forces over very many, exceedingly tiny, intervals of time. MD is much faster and scales better than DFT, but it does not explicitly include electrons, a weakness. Instead, MD typically uses simple empirical models to relate atom positions to forces. The model parameters can be adjusted to approximately mimic specific materials, but these models can not accurately match DFT results in detail.

10 SANDIA NATIONAL LABORATORIES



Simulation Engine LAMMPS

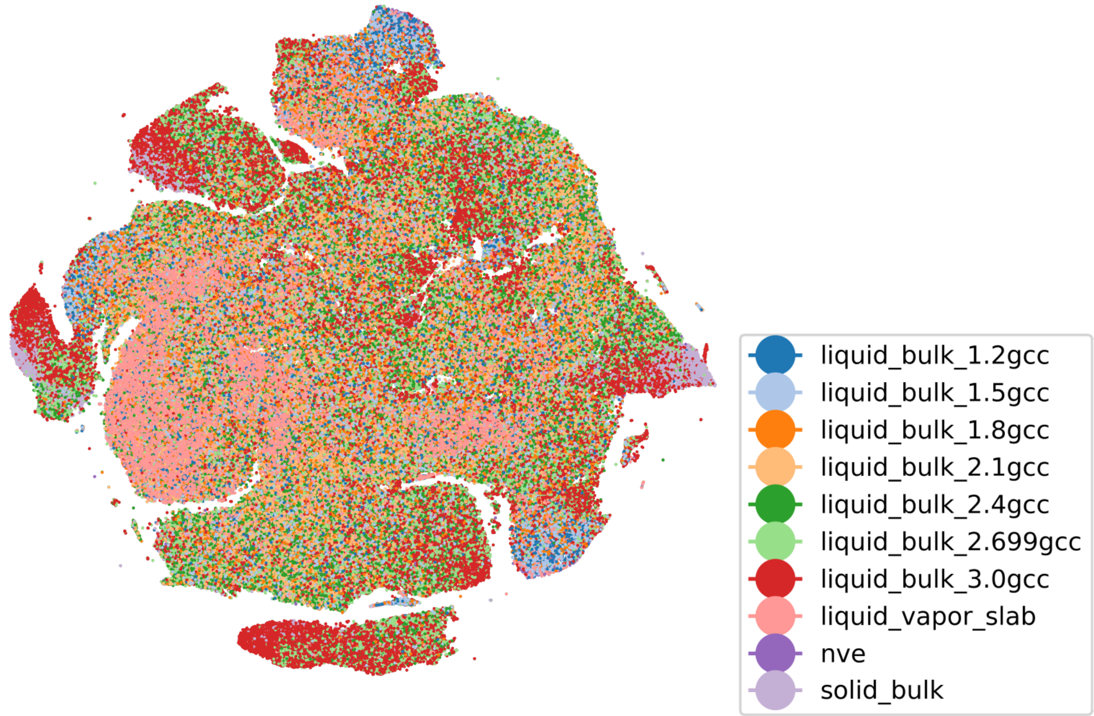
Google → Sandia HPC Annual Report
sandia.gov/news/publications/hpc-annual-reports/

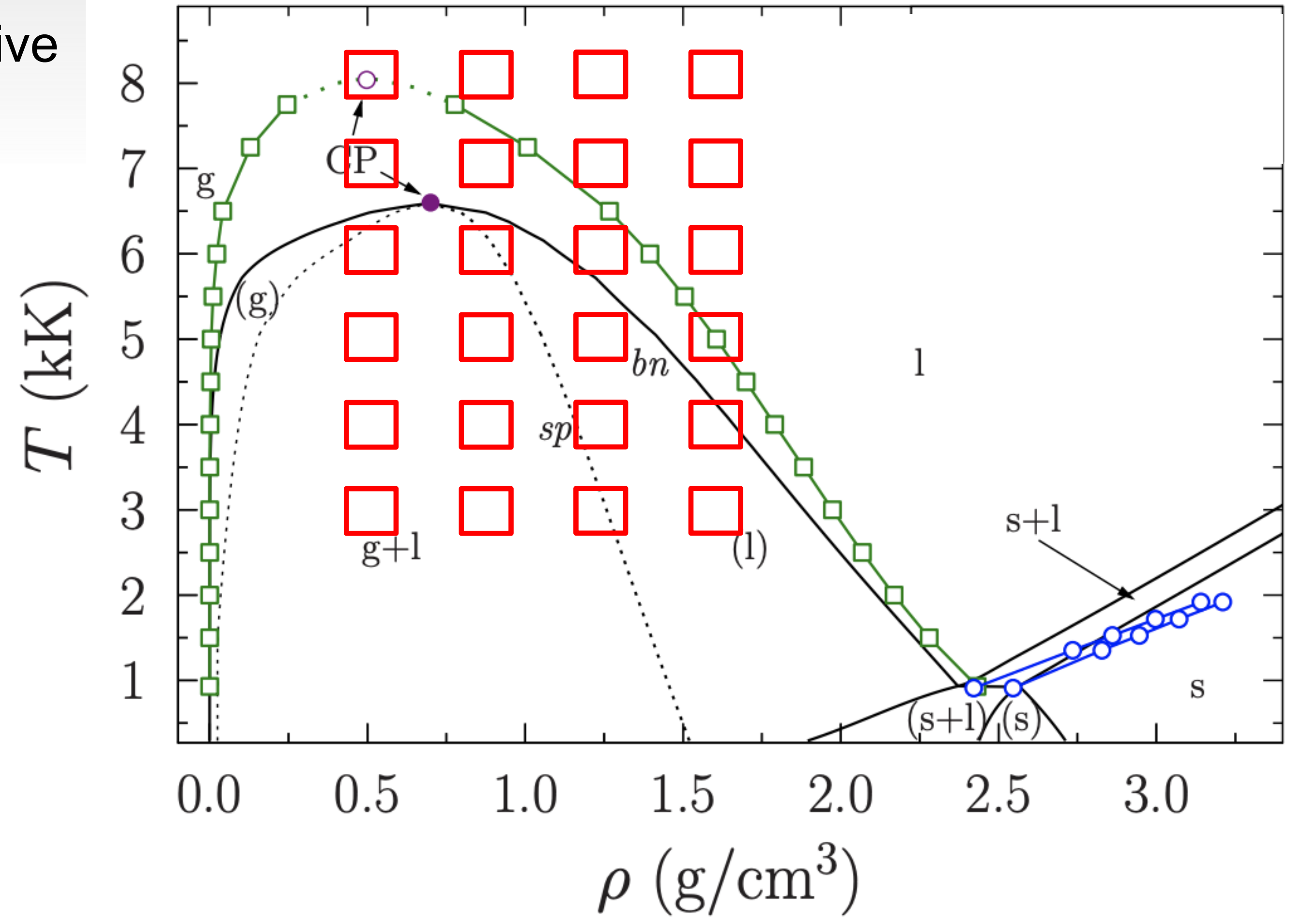
Trained to What?

Training Set Construction

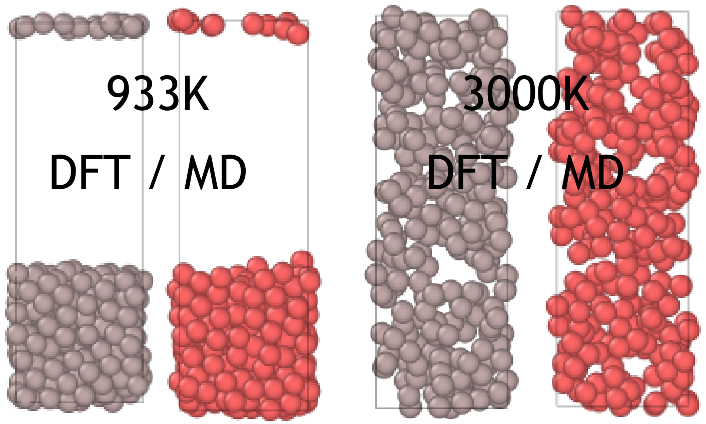
- Generated by running ab initio MD at various densities and temperatures
- How should we efficiently plan this expensive step?

t-SNE Projection:





Trained to What?

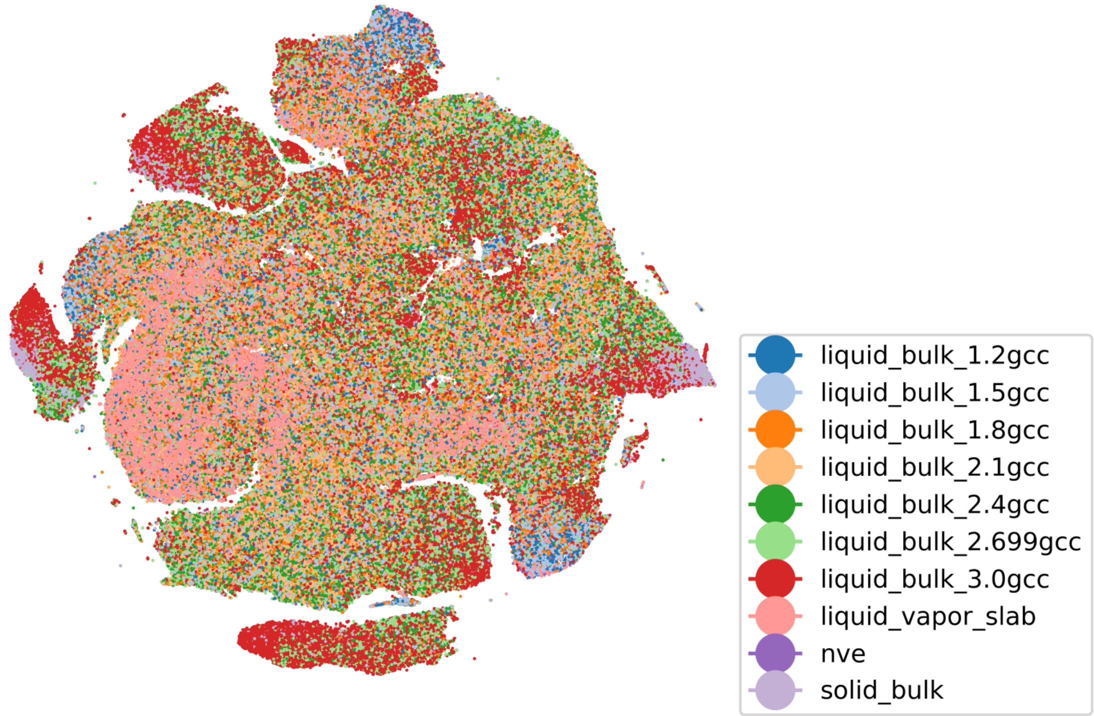


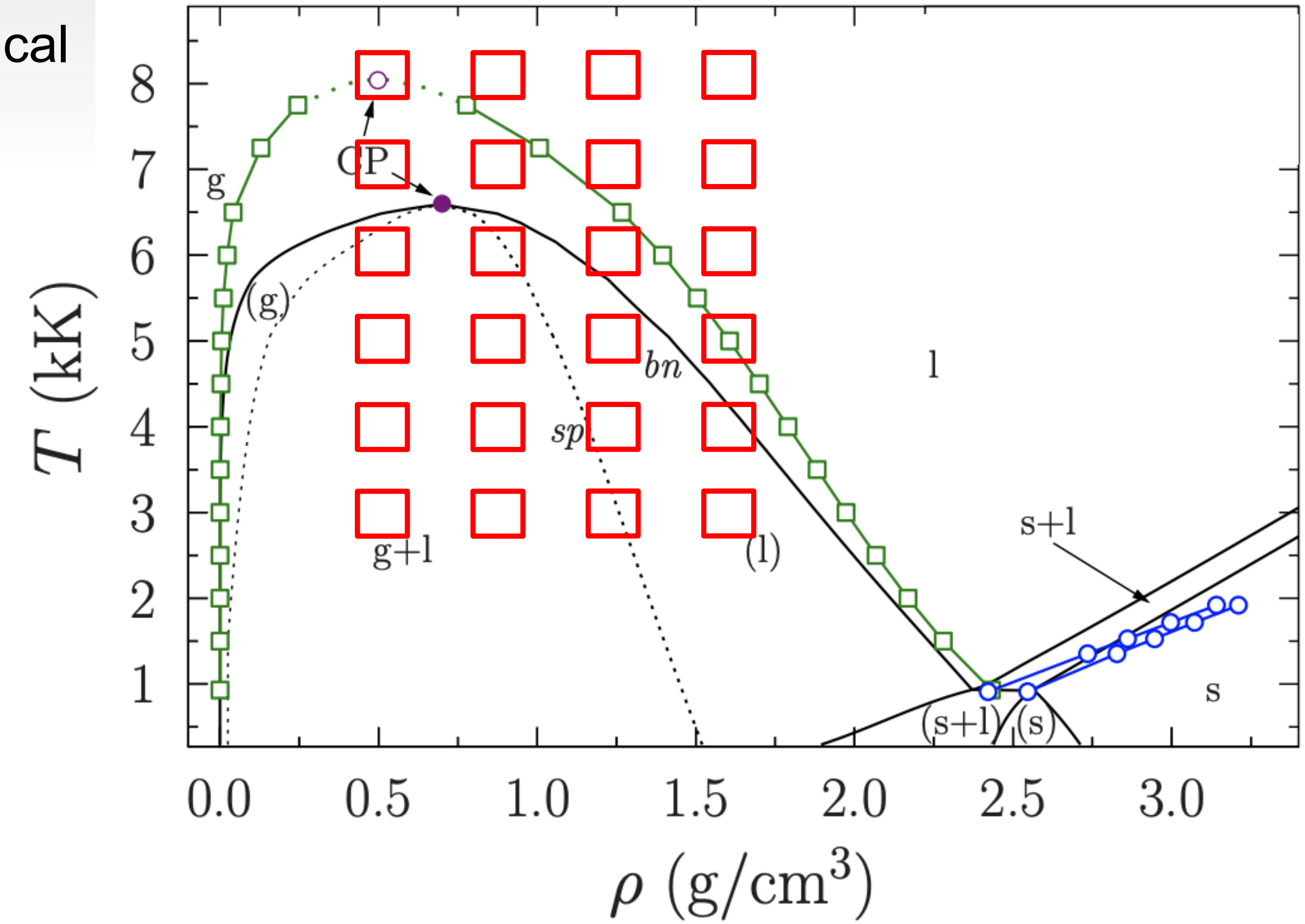
Training Set

Optimization

- Evaluated by ~30k atom simulations to map out the liquid-vapor coexistence region
- T_c and ρ_c were fit using universal Ising critical exponent β ≈ 0.326 and law of rectilinear diameter

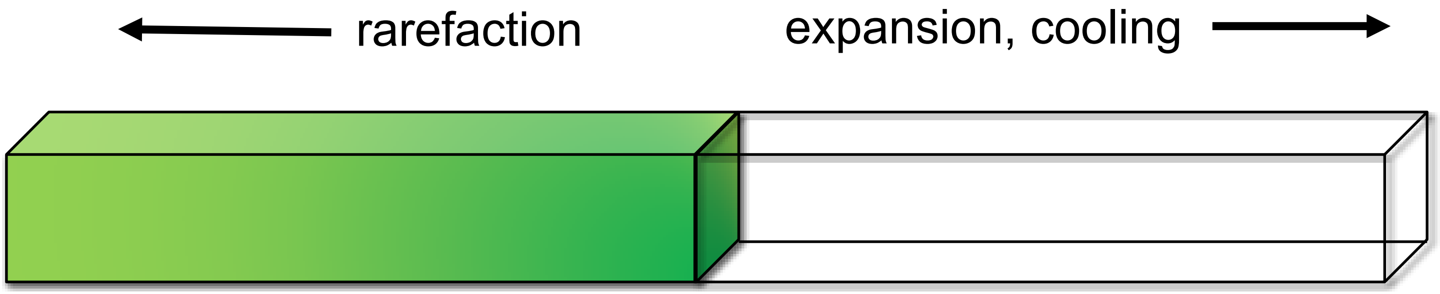
t-SNE Projection:



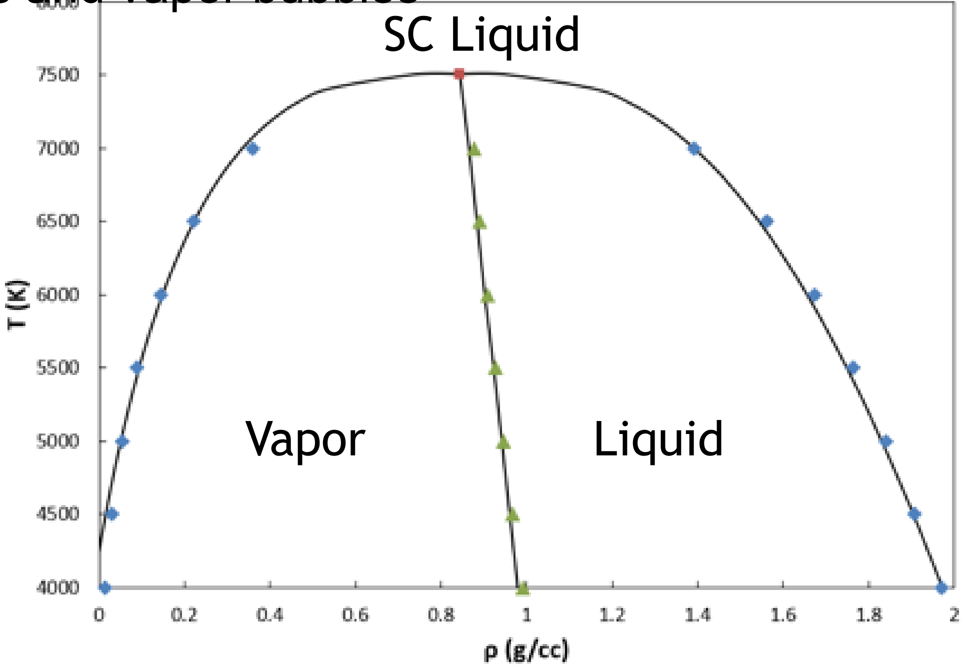


Computational / Physical Setup

- NNSA's ATS-2 Sierra Supercomputer
- 4320 nodes, 4 V100-16GB GPUs per node, IBM Power 9 CPUs



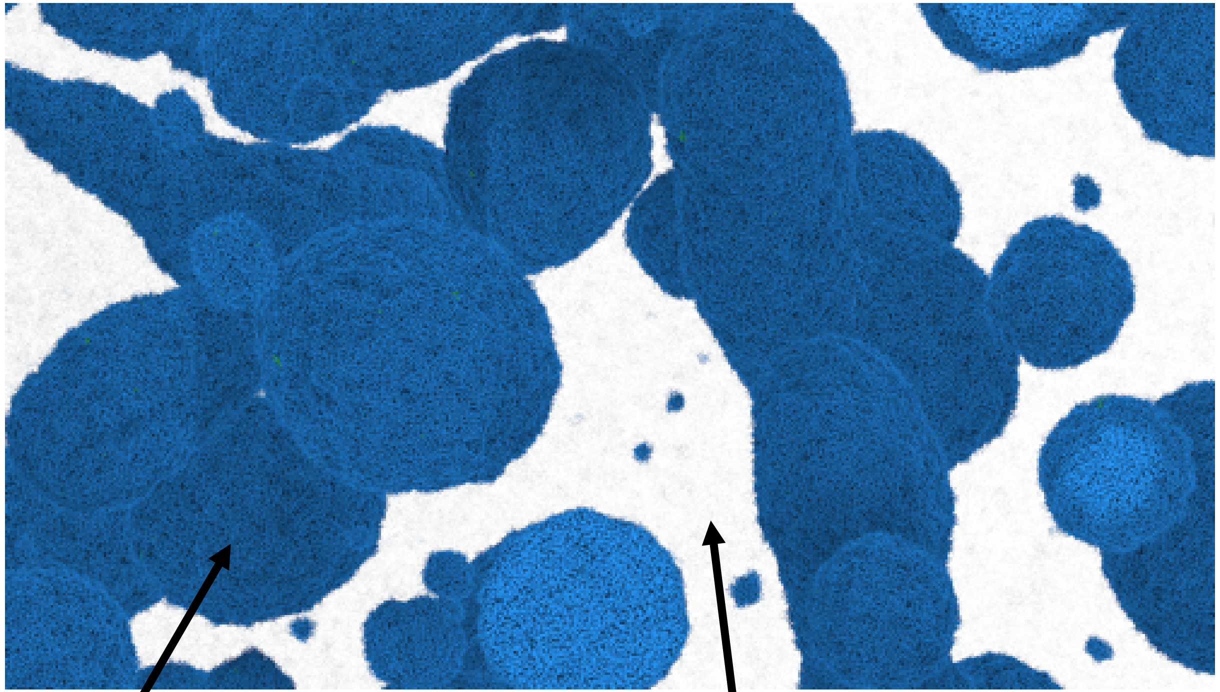
- When the supercritical fluid expands, the temperature drops below the critical temperature, and the fluid rapidly phase-separates into liquid droplets and vapor bubbles



Computational / Physical Setup

Visualization

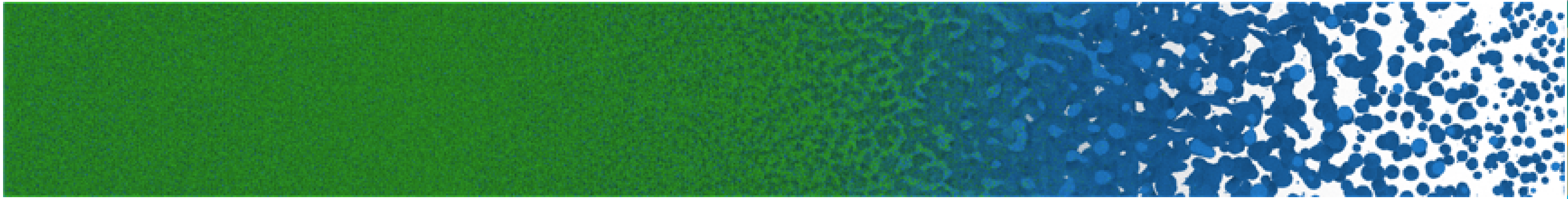
- Highly optimized for particle simulations, has direct support for LAMMPS output formats
- Produces high quality visualizations with ray tracing, ambient occlusion, etc.
- Highly scriptable with Python and useful for data post-processing and analysis in general (in addition to rendering images)



Visual size of an atom is scaled by local density

Green : Above T_c

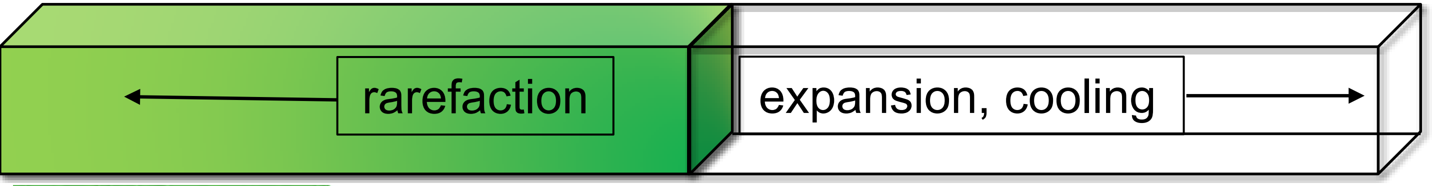
Blue : Below T_c

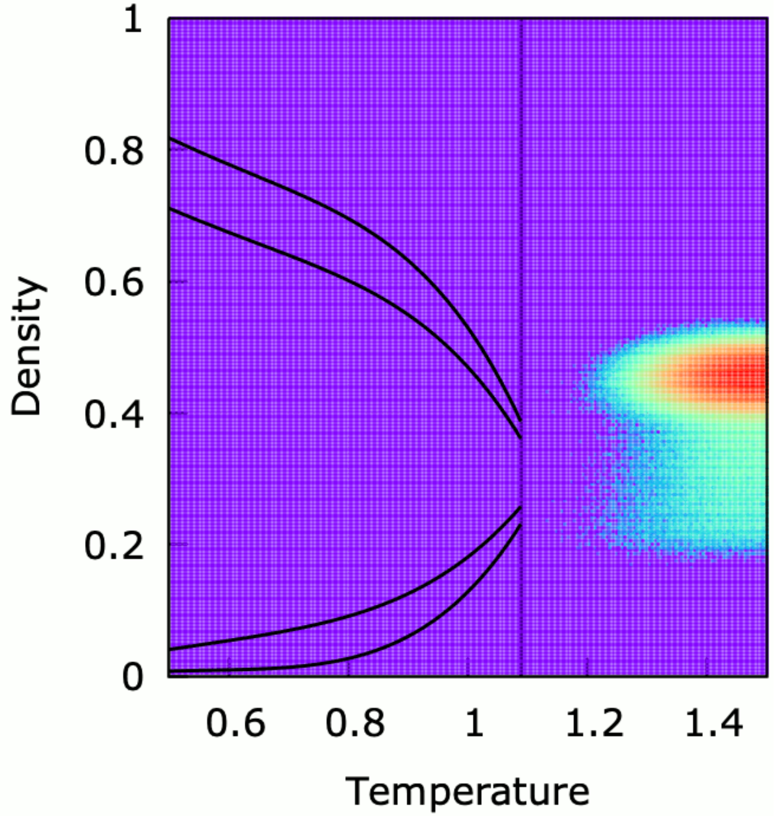


Aluminum Vapor Dome

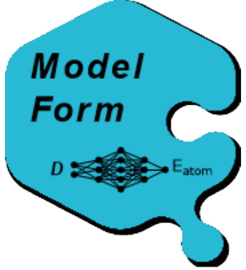
Exascale ML-MD

- 1.5B atoms, 8192GPUs (~47% of Sierra)
- $T_0 = 9000K$, $\rho_0 = 1.5\ {}^{g}/_{cm^3}$ 1.8um, 0.56ns captured





Model Form Selection

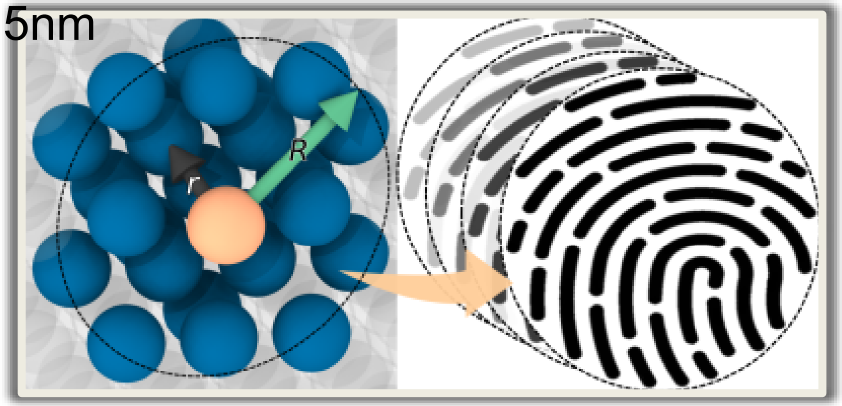


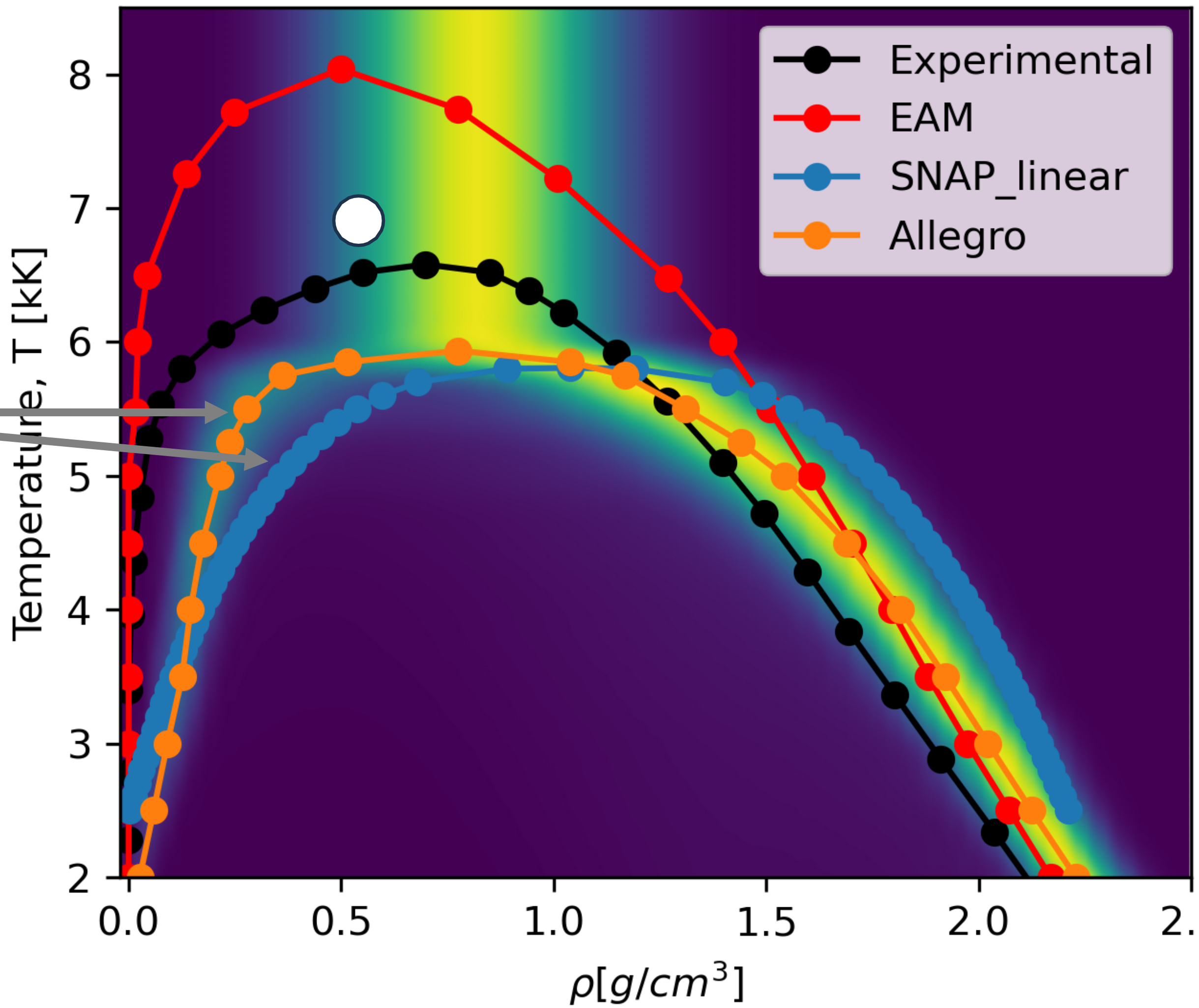
Right Answer, Right Reasons

- Looking at predicted vapor dome, some noticeable shortcomings
- EAM, Experiment is taken from Povarnitsyn *et. al.* PRB (2015)

- Underpredicting critical temperature
- Overpredicting vapor density

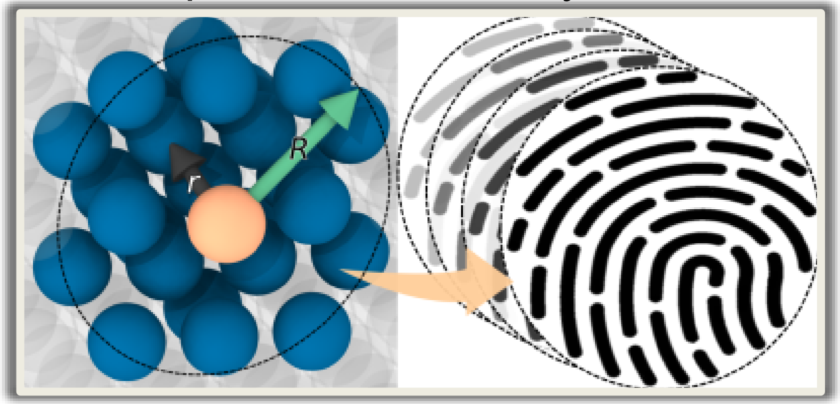
- Descriptor is a short ranged interaction, ~0.5nm





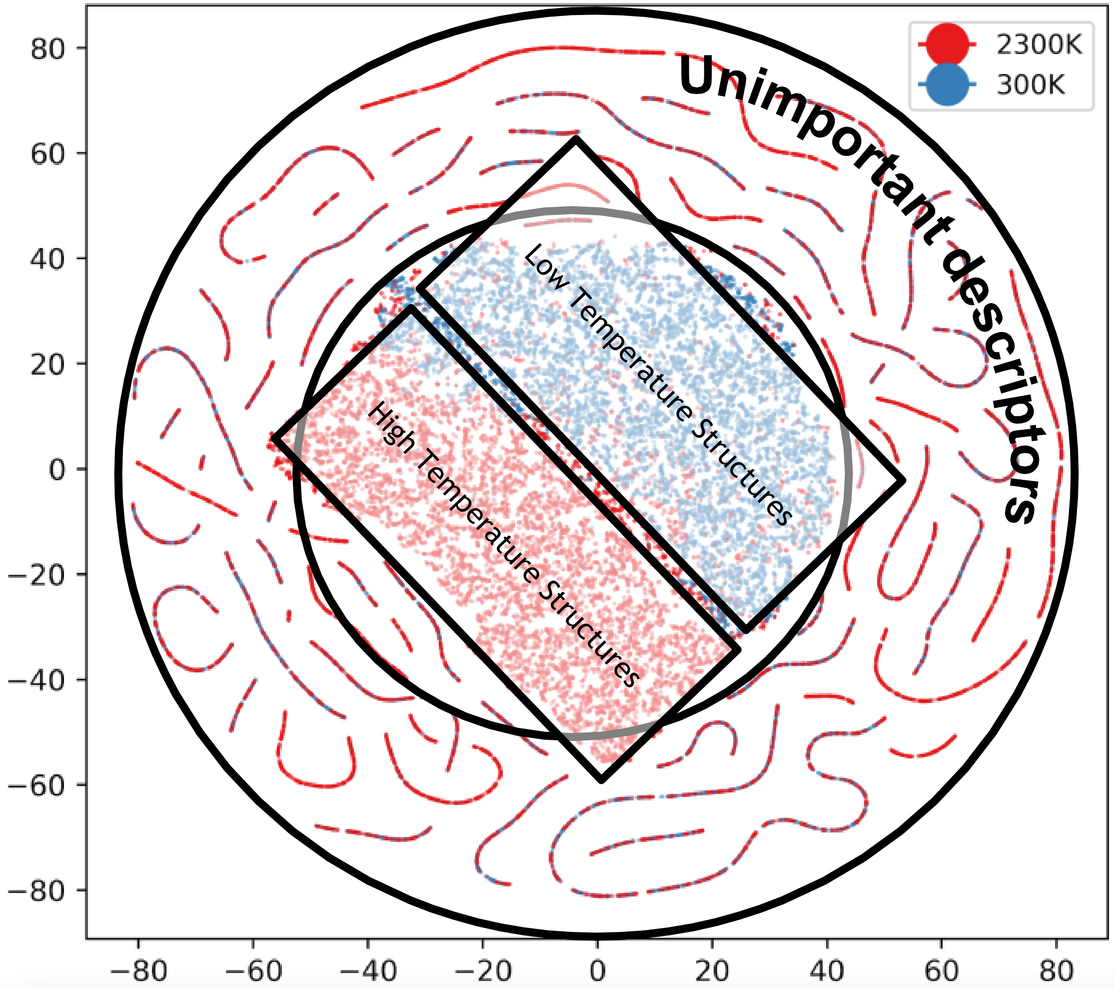
Descriptor Extrapolation

- Remember, this is what the model sees. Not temperature or density



- Extrapolations should be defined by the descriptors, allows for MD to be compared to DFT

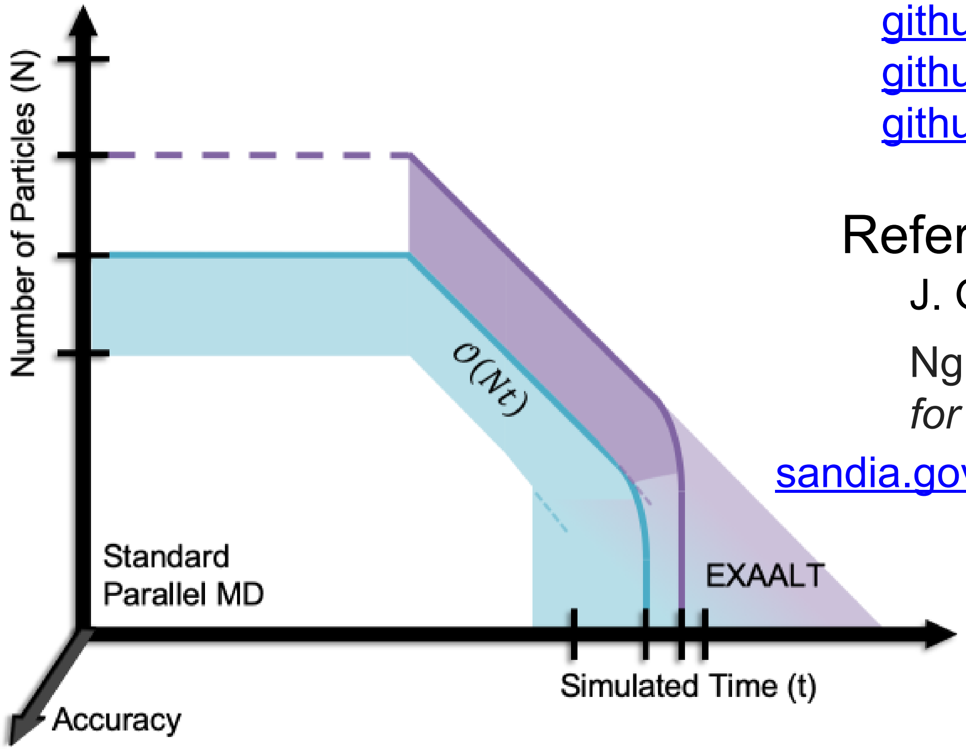
- **<u>Post-processing or real-time analysis of MD trajectory is possible to quantify extrapolations</u>**



Conclusions and Path Forward

Contact Information: mitwood@sandia.gov

- Data-driven interatomic potentials allow for MD predictions of challenging problems.
- While harder to quantify, the fidelity of our MD simulations needs to be a key consideration at the Exascale



Links:
github.com/FitSNAP/FitSNAP
github.com/lammps/lammps
github.com/materialsvirtuallab/mlearn

References:

J. Goff (2022) arXiv:2208.01756

Nguyen-Cong, K. *Proc. International Confere*
for High Performance Computing (2021).

sandia.gov/news/publications/hpc-annual-reports

- Thank you to all my collaborators:
Aidan Thompson, Stan Moore, Ember Sikorski, Steve Plimpton, Normand Modine, Dionysios Sema, Svetoslav Nikolov, Charlie Sievers, Danny Perez, James Goff, and many others!

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

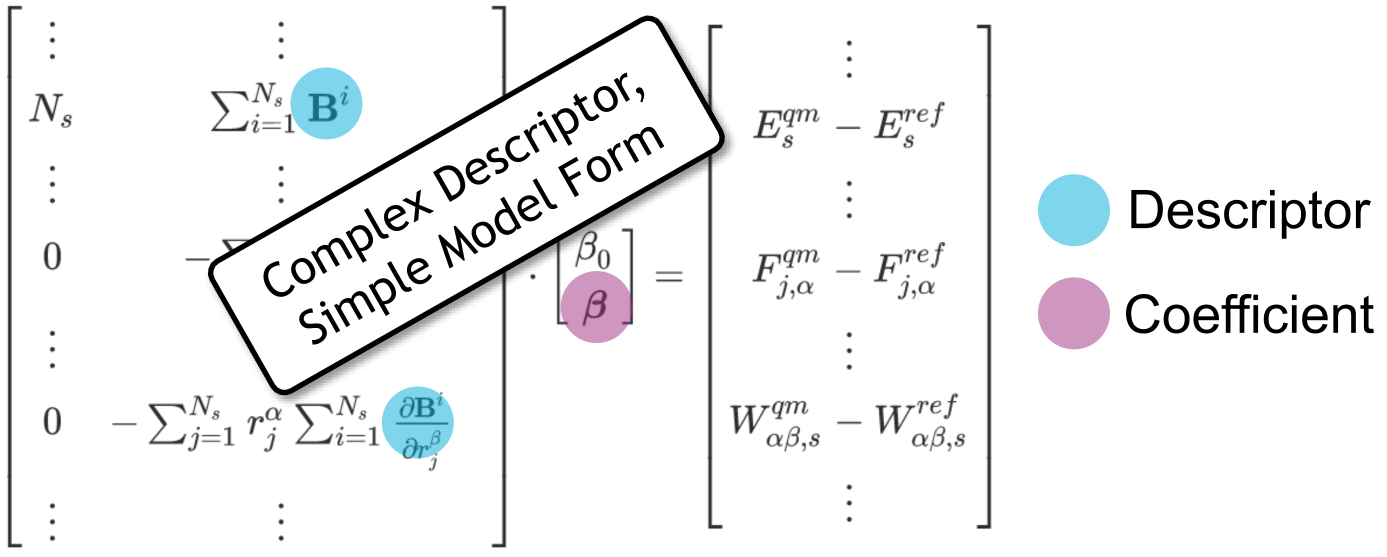
https://github.com/FitSNAP/FitSNAP

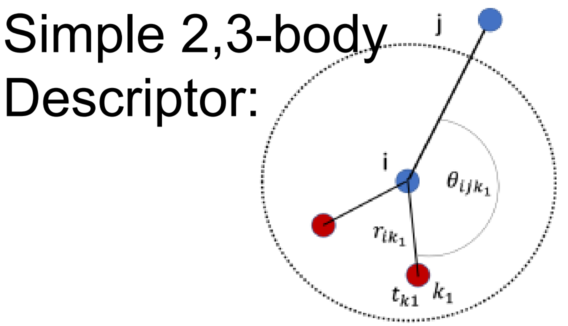
https://github.com/materialsvirtuallab/mlearn

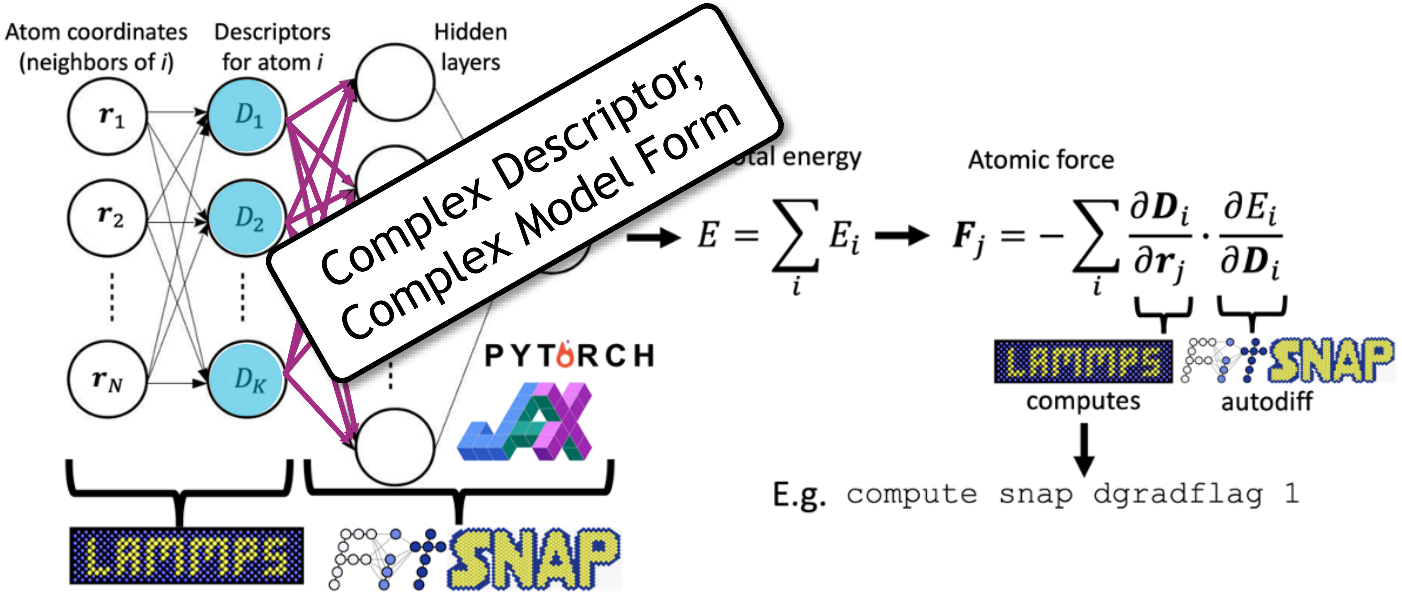
Contact Information:

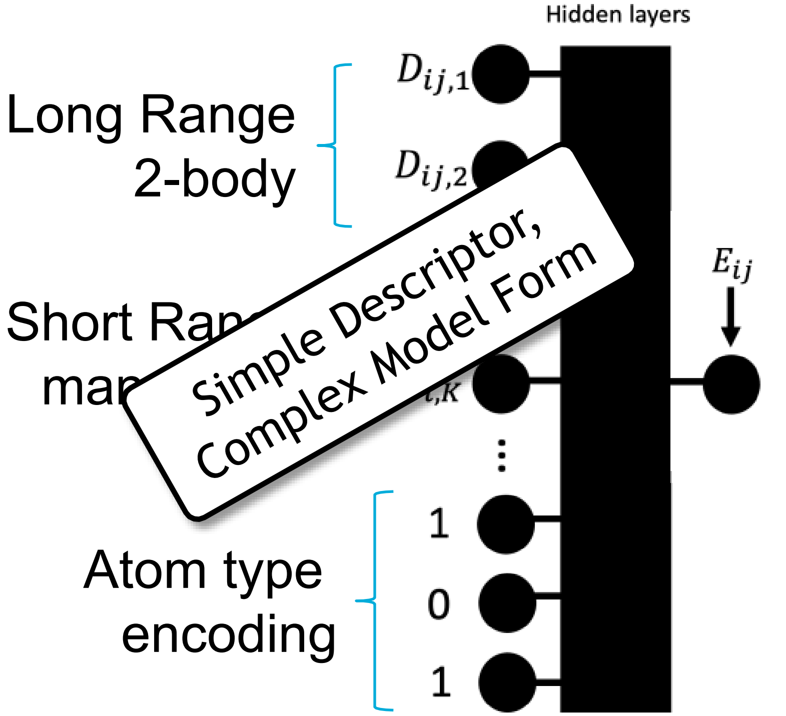
mitwood@sandia.gov

Model Form Selection



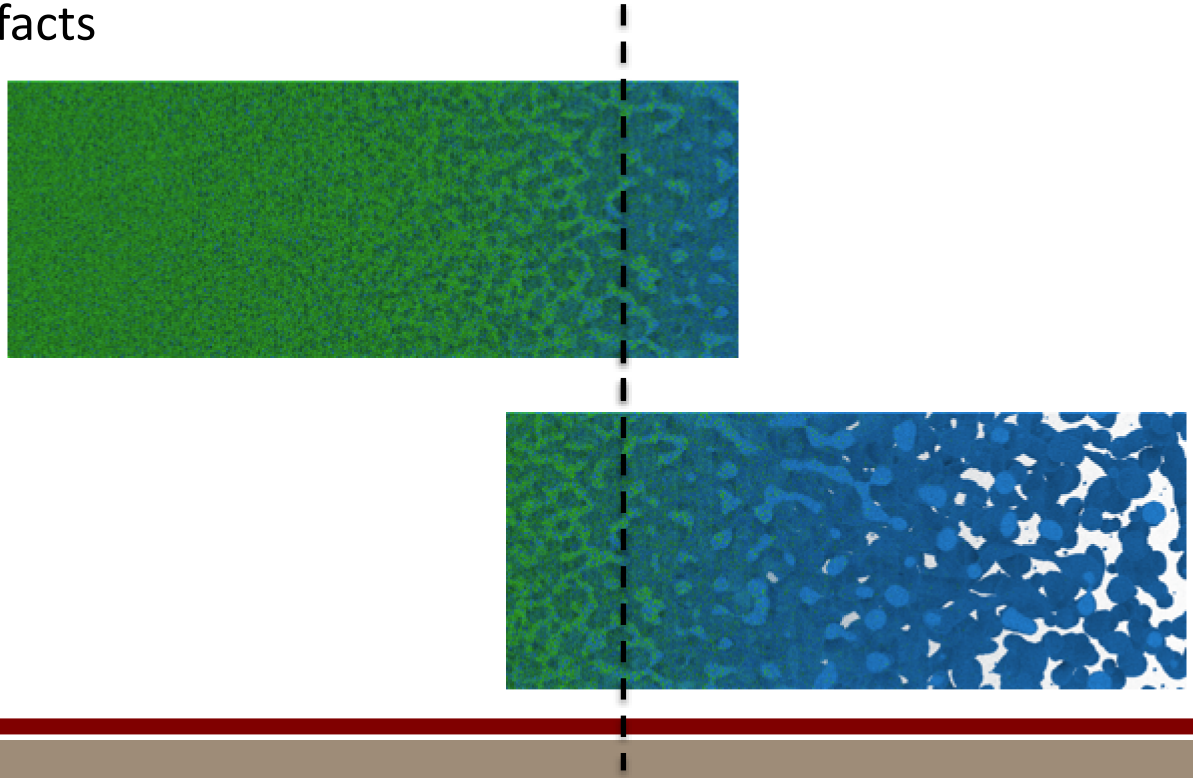






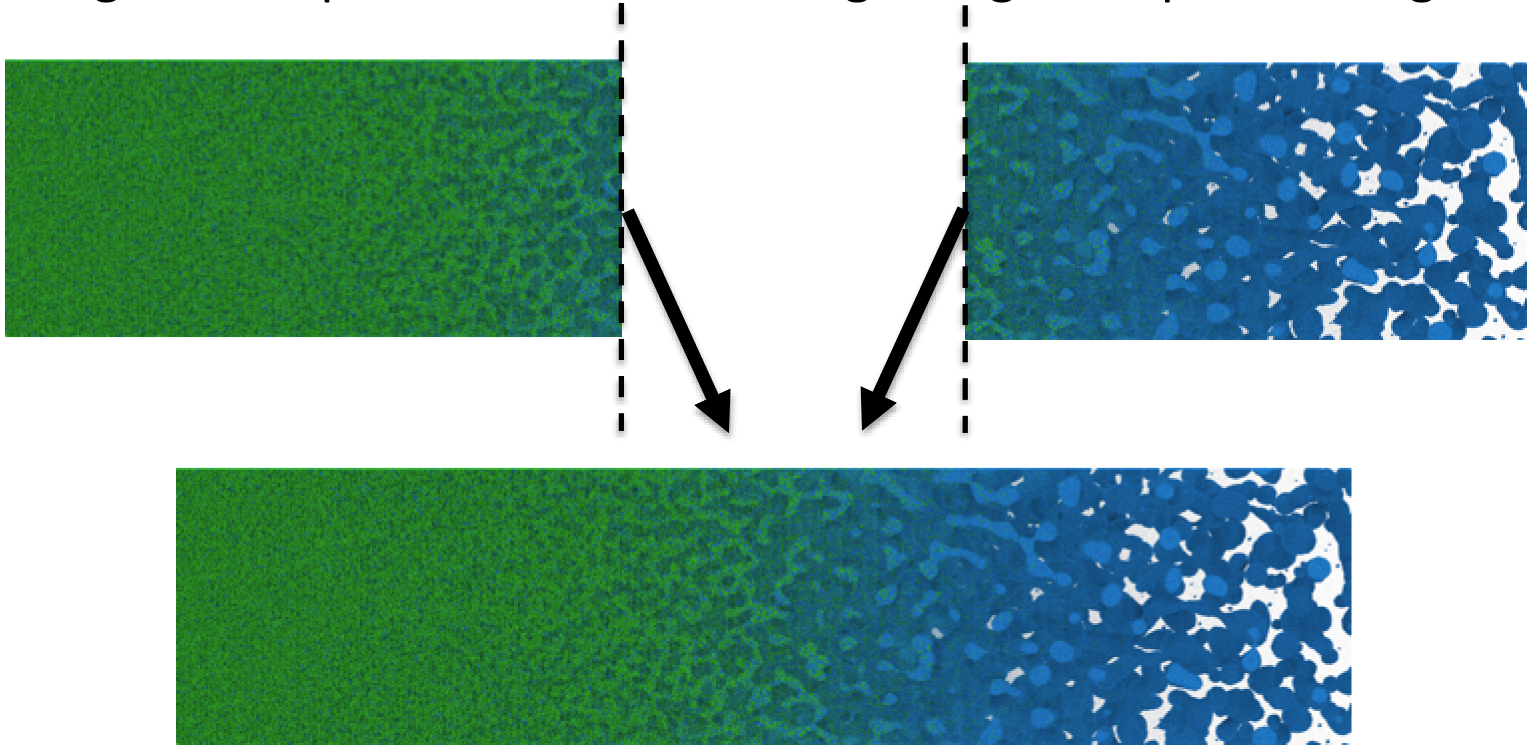
OVITO Parallelism Workaround

- LAMMPS reshuffles atom data between ranks so each rank has a “slice” of the simulation data in the x-direction
- Each rank outputs to a separate file (e.g. 8192 files total)
- MPI driver program launches separate instances of OVITO on many nodes
- Each OVITO instance loads atom data from “owned” slices, along with neighboring slice data to create a buffer zone to reduce visual edge artifacts



OVITO Parallelism Workaround (cont.)

- OVITO renders an image of the slice, including buffer zone, thenthe buffer region is cropped off
- Another MPI driver program stiches all the small slice images together in parallel to create a single large composite image



OVITO Parallelism Workaround (cont.)

Advantages:

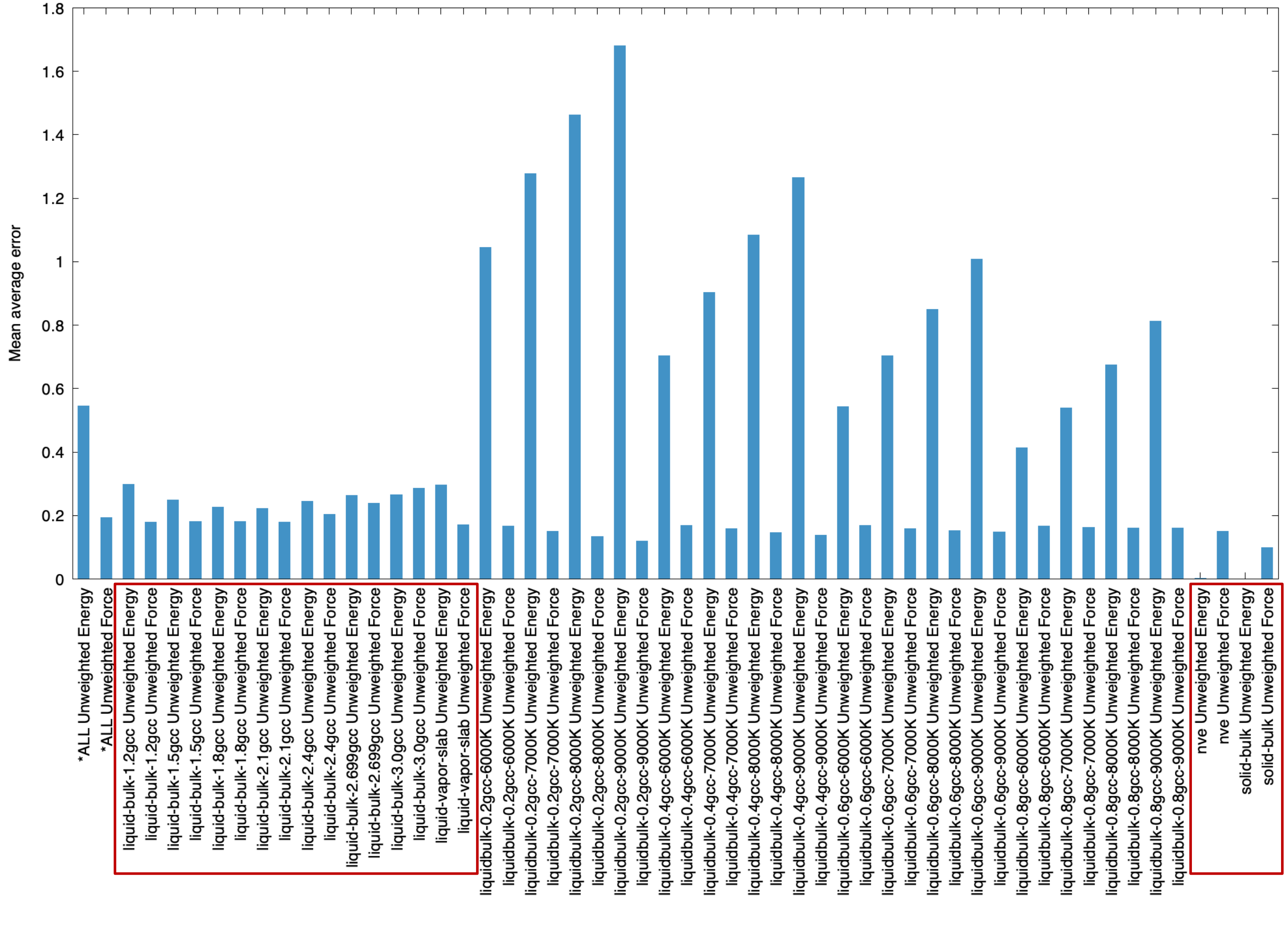
- Highly scalable: large images are rendered in an (almost) embarrassingly parallel manner
- Can render more than 2 billion atoms

Disadvantages:

- Minor artifacts in lighting/shadows, but overall produces nice, usable images in parallel
- Can only visualize a single face straight on (so everything lines up), no 3D perspective views

- Would like to also try Paraview in the future (less domain specific, but MPI-enabled so requires less workarounds)

AI Linear reduced May 16 – Force and Energy errors on new low density configurations



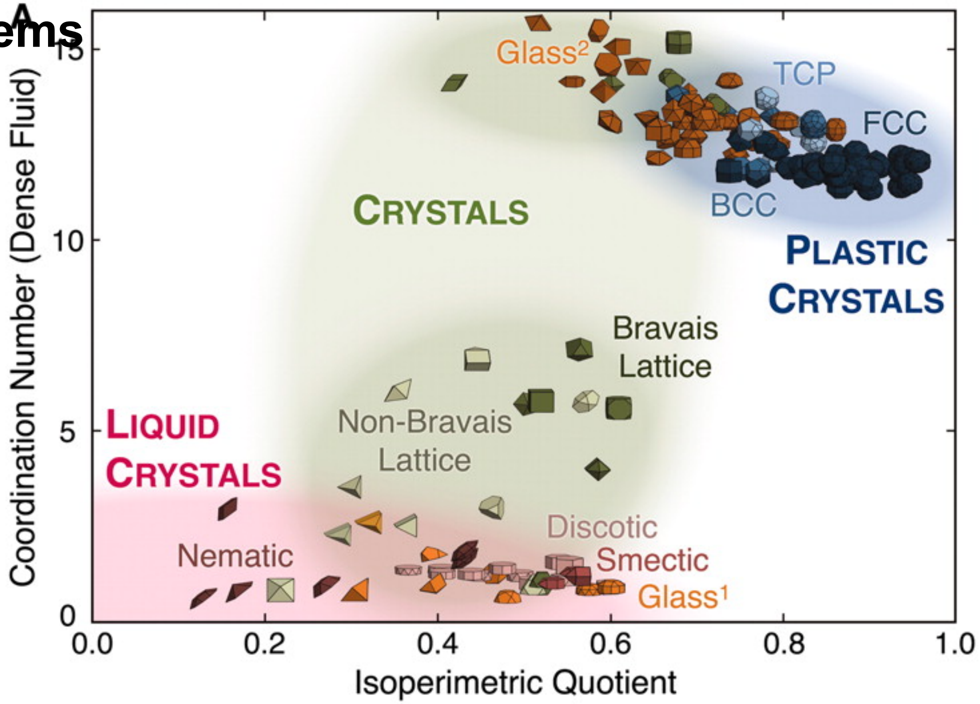
The May 16th potential performs poorly when extrapolating the <1 gcc configurations, especially on energies at higher temperatures.

Included in May 16th training set

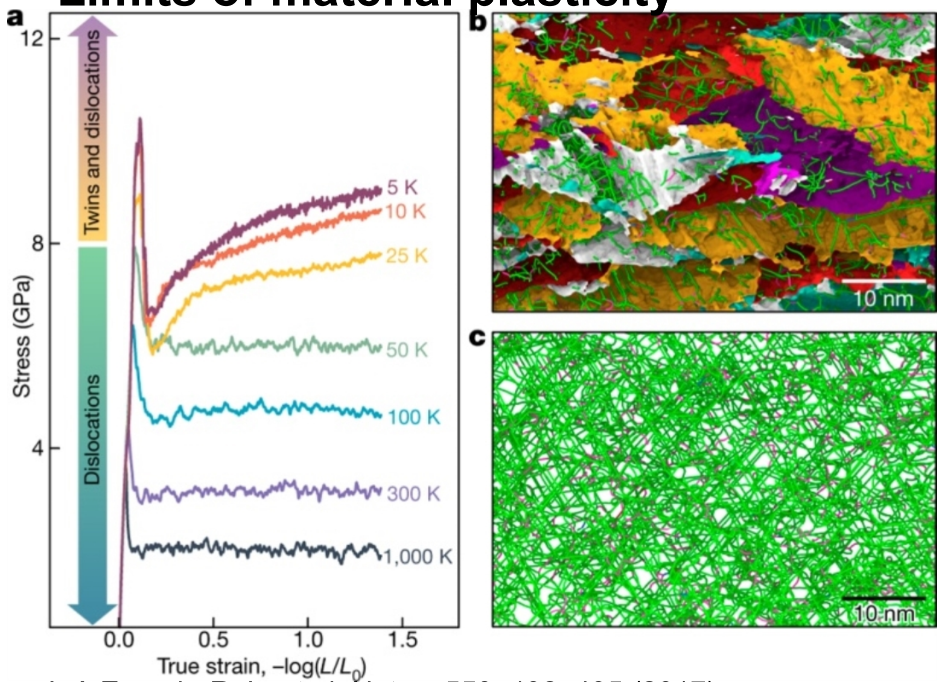
Examples of Petascale Achievement

Phases of granular systems

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

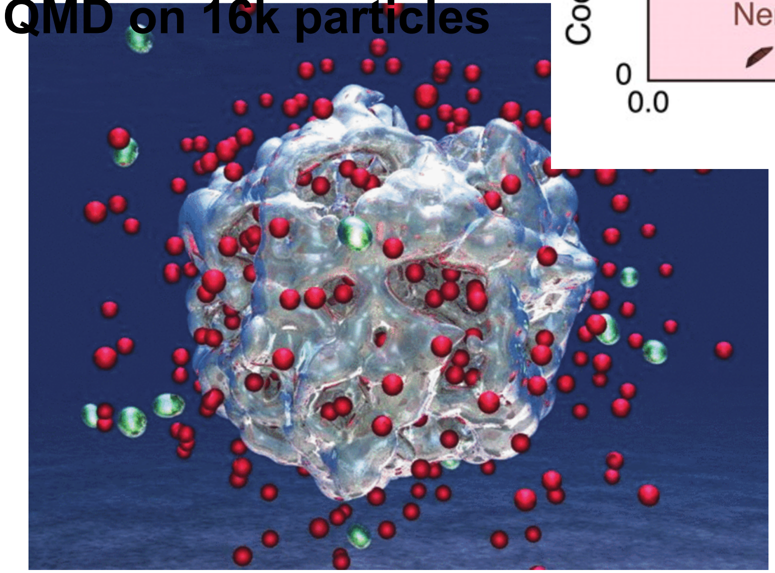


Limits of material plasticity



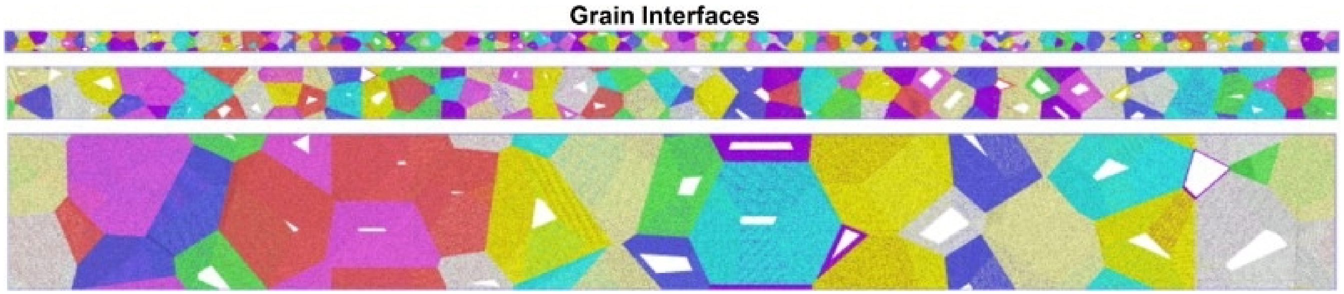
L A Zepeda-Ruiz *et al.* *Nature* **550**, 492–495 (2017) doi:10.1038/nature23472

H production in Water/Al QMD on 16k particles



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

Shock Response of coarse grained explosives



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.