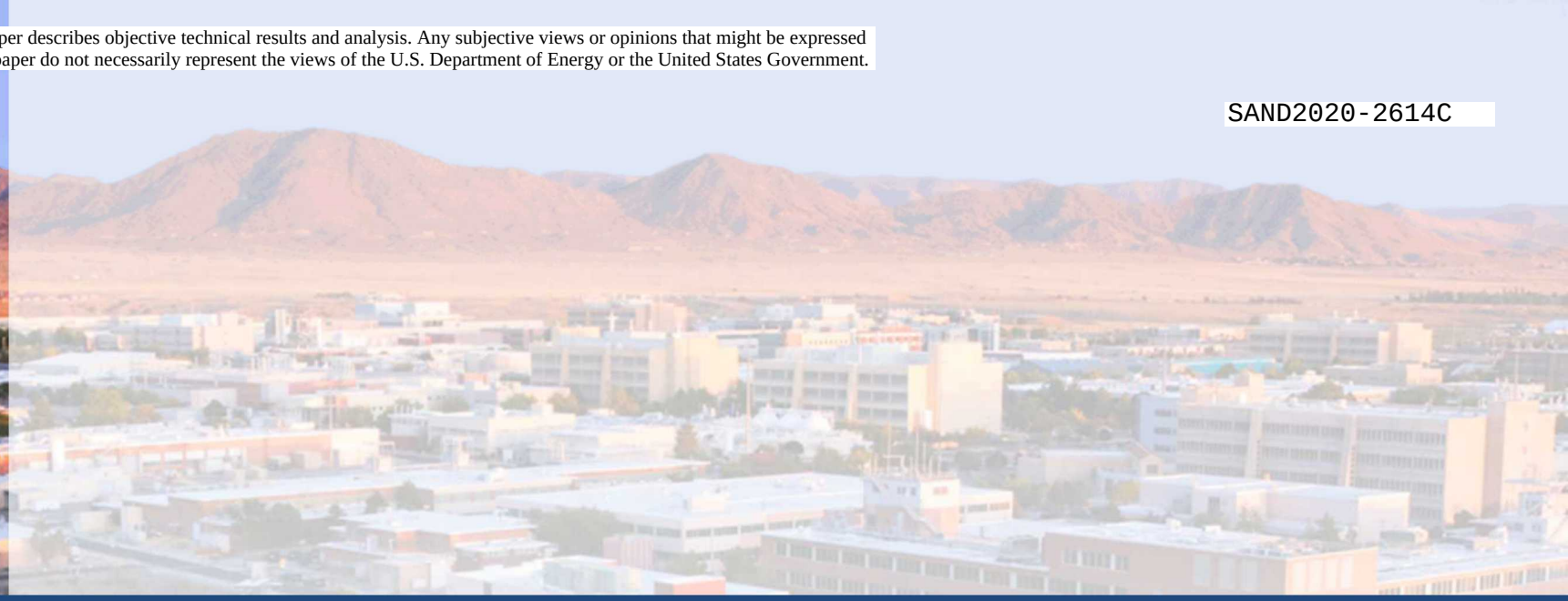
SAND2020-2614C



A Path to the Exascale for Atomistic Simulations with Improved Accuracy, Length and Time Scales

SAND2020-XXXX

Mitchell Wood, Aidan Thompson, Steve Plimpton

Center for Computing Research, Sandia National Labs

Anders Niklasson, Danny Perez

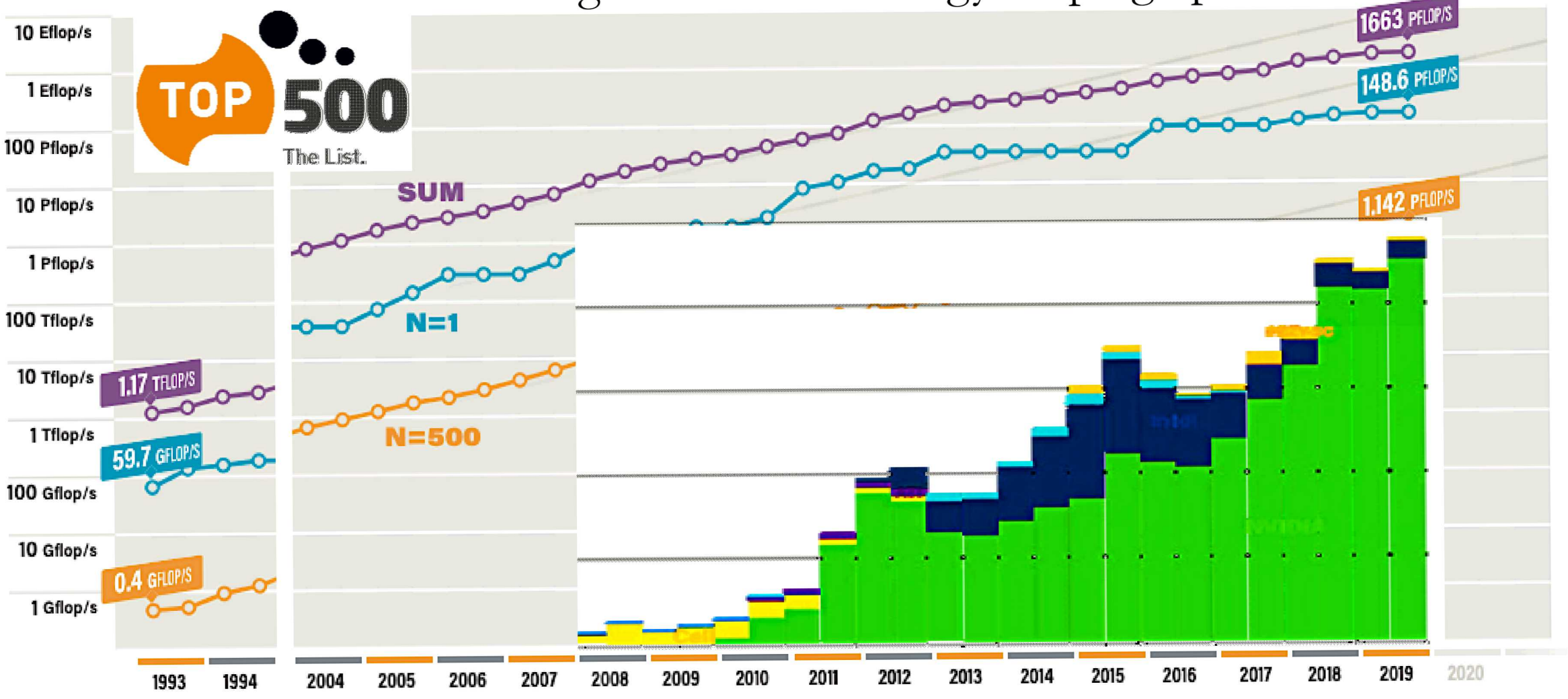
Theoretical Division, Los Alamos National Lab

APS March 2020 Meeting, Denver, CO

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.

Computing Status Report

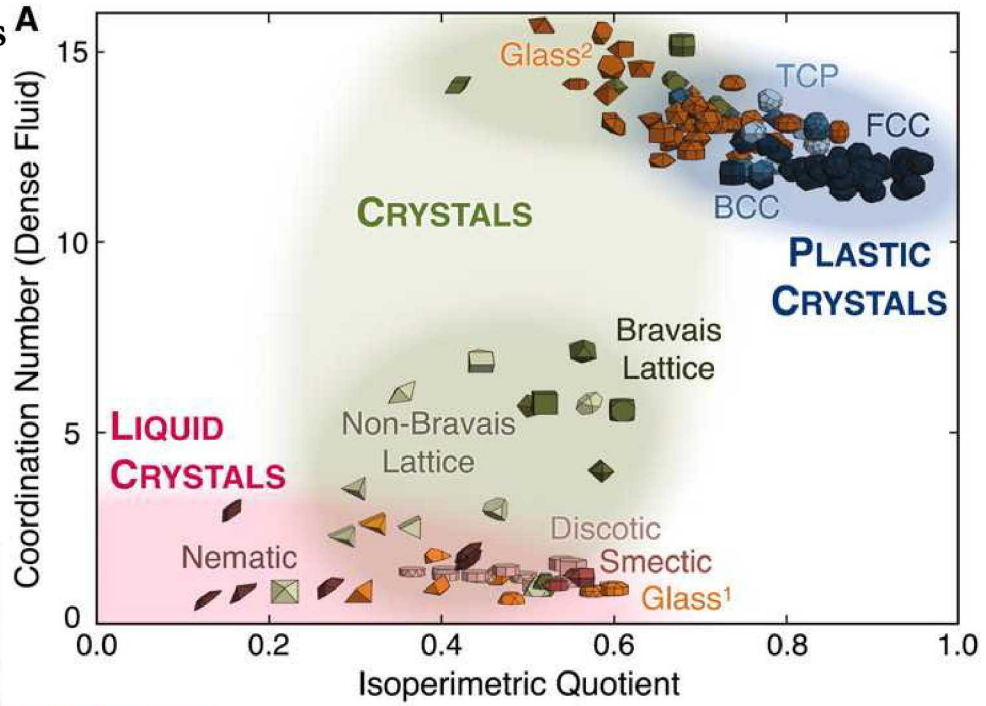
More powerful supercomputers are inevitable, but is our scientific usage of this technology keeping up?



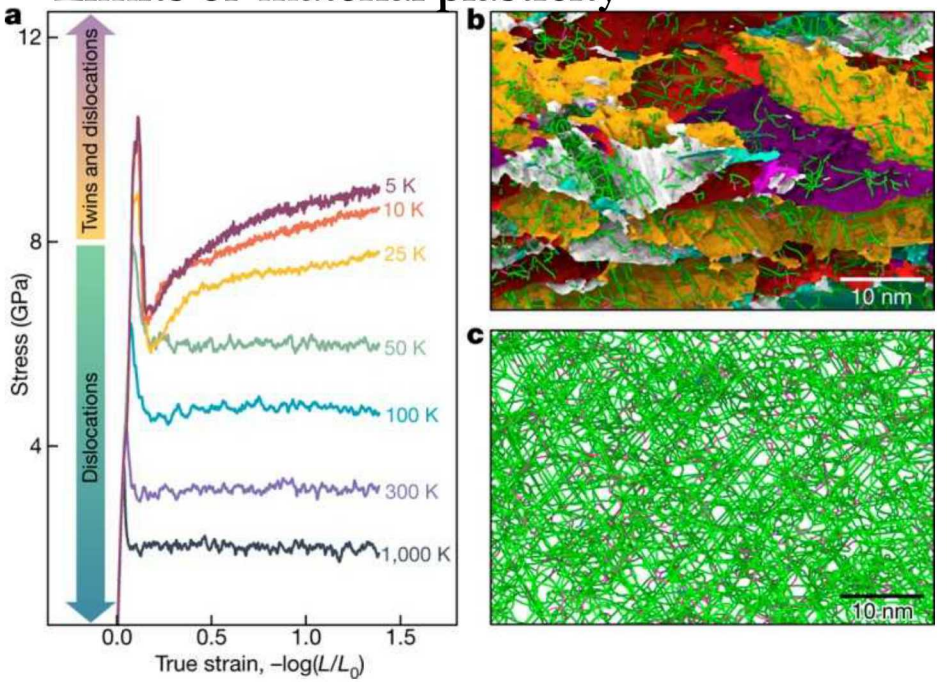
Examples of Petascale Achievement

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.

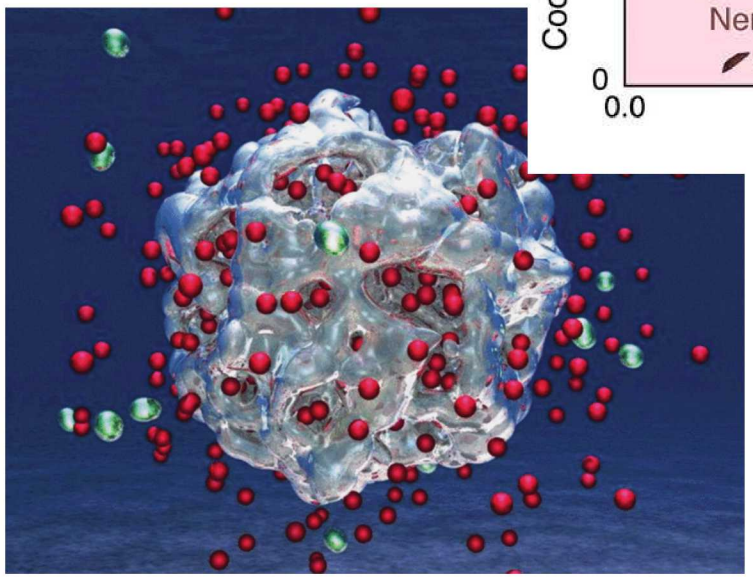


Limits of material plasticity



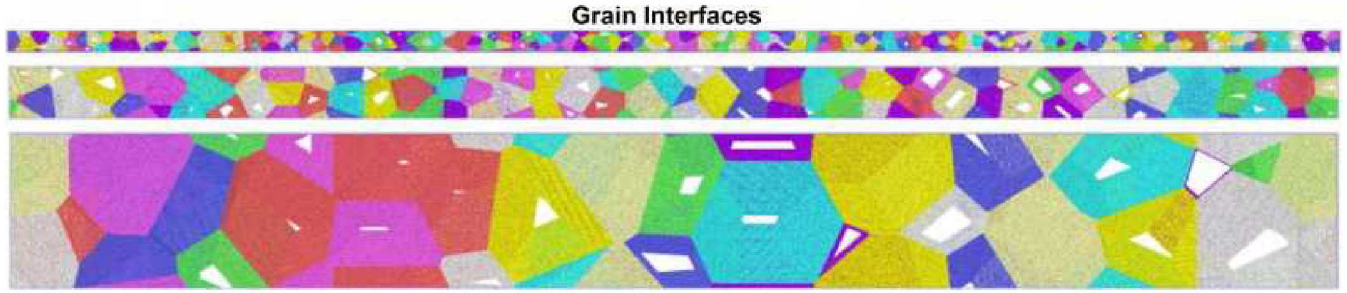
L A Zepeda-Ruiz *et al. Nature* **550**, 492–495 (2017) doi:10.1038/nature23472

H production in Water/Al QMD on 16k particles



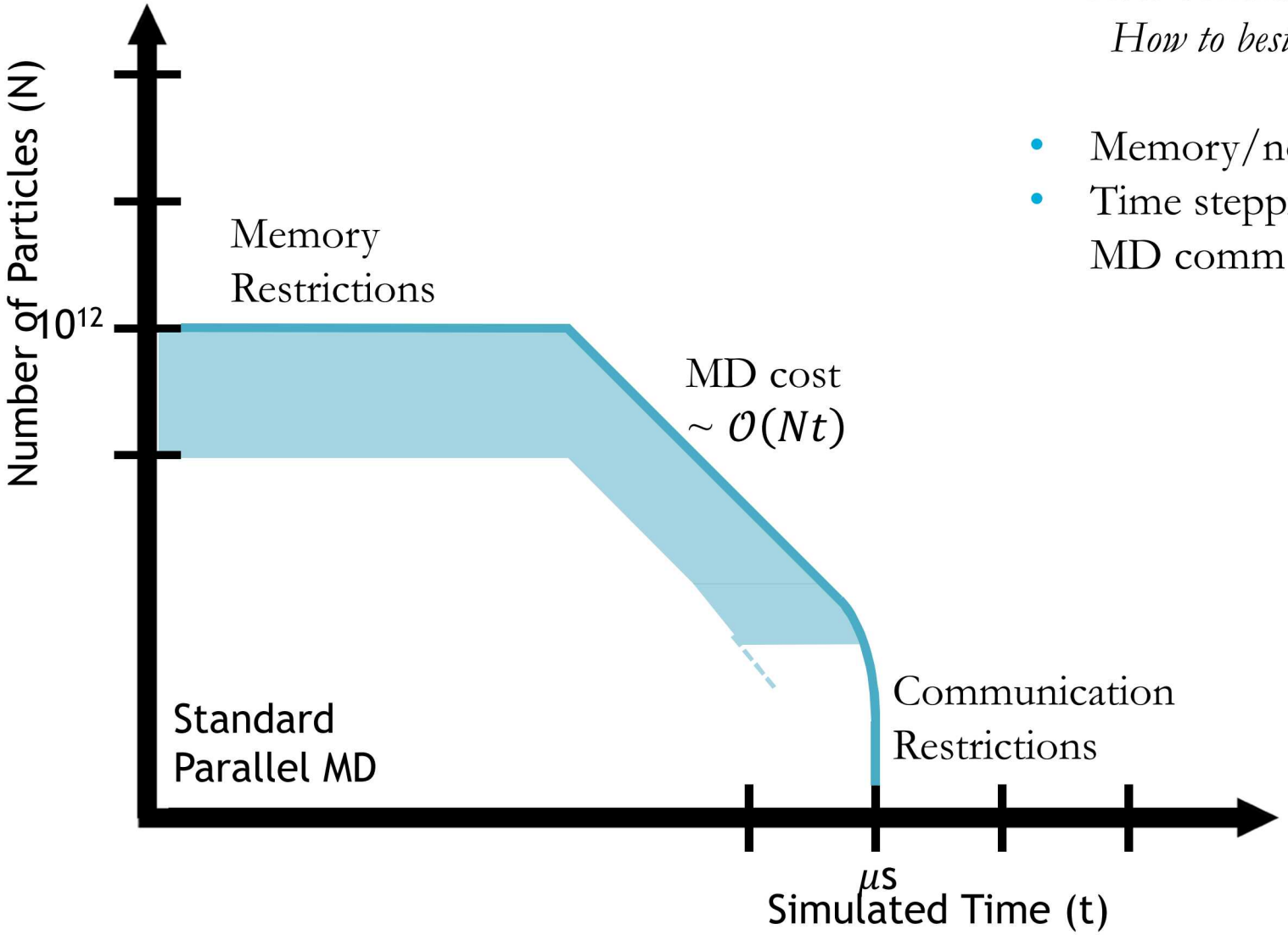
K. Shimamura et al., "Hydrogen-on-Demand Using Metallic Alloy Nanoparticles in Water," Nano Letters, vol. 14, no. 7,2014, pp. 4090–4096

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.

What is possible for MD at the Exascale?

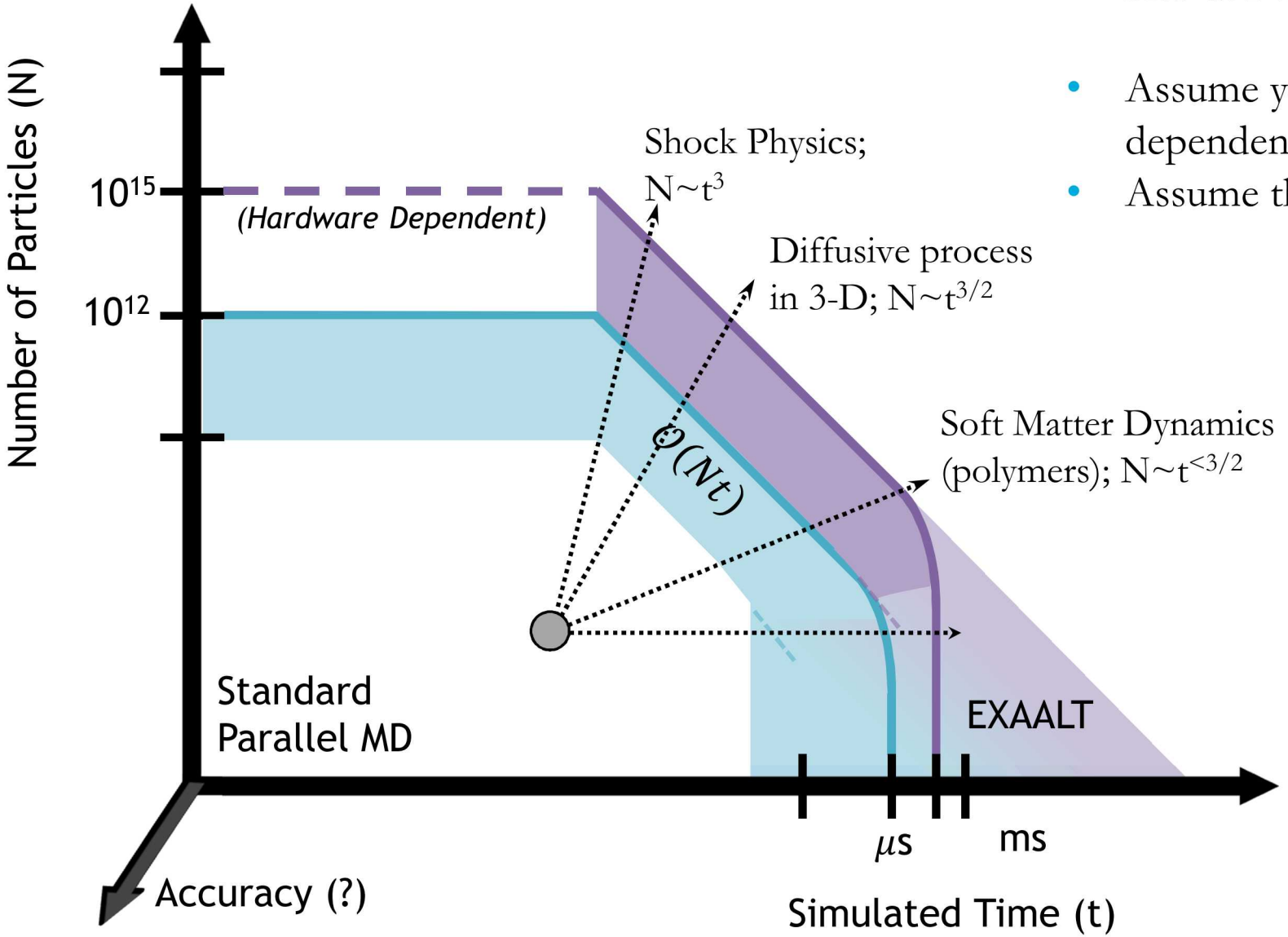


- Lets consider a 24hr allocation on a leadership platform
  How to best spend this computational budget on MD?

- 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?

The Master Plot



- *How does this affect the research done on these platforms?*
- Assume your problem has some specified length-scale dependence : $N \sim L^{\alpha}$
- Assume the associated time-scale goes as $t \sim L^{\gamma}$

- But what if you care about a system governed by rare event dynamics?

EXAALT Team and Scope

Team by Institution

- Los Alamos National Lab
  - Danny Perez(PI), Anders Niklasson(Co-PI), Marc Cawkwell, Toks Adedoyin, Christian Negre, Yu Zhang, Andrew Garmon, Enrique Martinez, Joshua Brown, Nicholas Lubbers, Tim Germann
- Sandia National Lab
  - Steve Plimpton(Deputy-PI), Aidan Thompson, Mitchell Wood, Mary Alice Cusentino, Stan Moore
- Oak Ridge National Lab
  - Brian Wirth, Li Yang
- External Collaborators
  - Rahul Gayatri, Yasaman Ghadar, Christopher Knight, Neil Mehta, Evan Weinberg, Art Voter(~retired)

ECP EXASCALE COMPUTING PROJECT

Los Alamos NATIONAL LABORATORY EST. 1943

Sandia National Laboratories

OAK RIDGE National Laboratory

NERSC

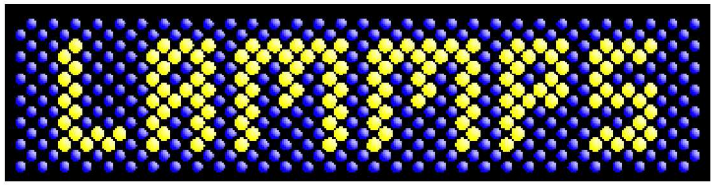
NVIDIA

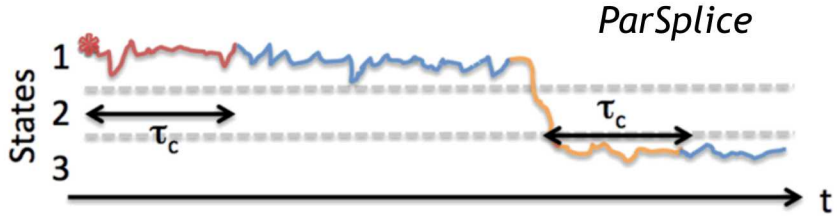
EXAALT Team and Scope

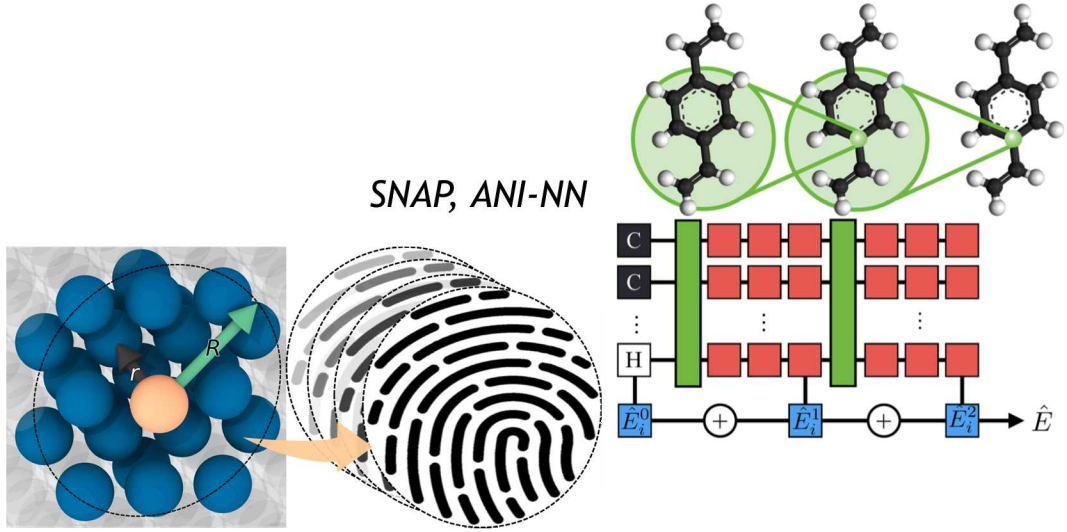
Team by Primary Role

- LAMMPS
  - Steve Plimpton, Aidan Thompson, Mitchell Wood
- LATTE
  - Toks Adedoyin, Marc Cawkwell, Christian Negre, Anders Niklasson, Yu Zhang
- ParSplice
  - Andrew Garmon, Enrique Martinez, Danny Perez
- Machine Learned Potentials
  - Joshua Brown, Mary Alice Cusentino, Nicholas Lubbers, Aidan Thompson, Mitchell Wood
- Domain Science
  - Brian Wirth, Li Yang
- Other Code Improvements
  - Tim Germann, Rahul Gayatri, Yasaman Ghadar, Christopher Knight, Neil Mehta, Stan Moore, Evan Weinberg

EXASCALE COMPUTING PROJECT



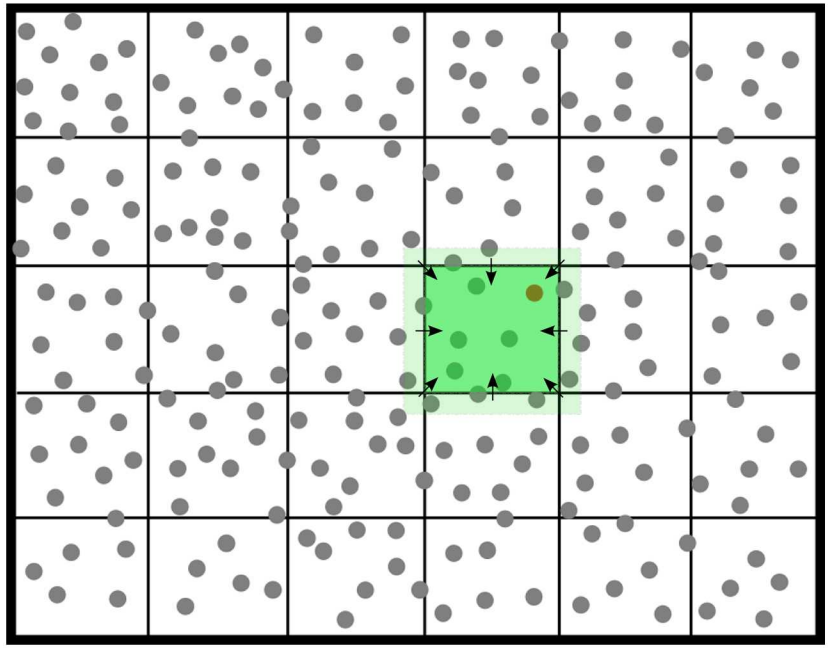


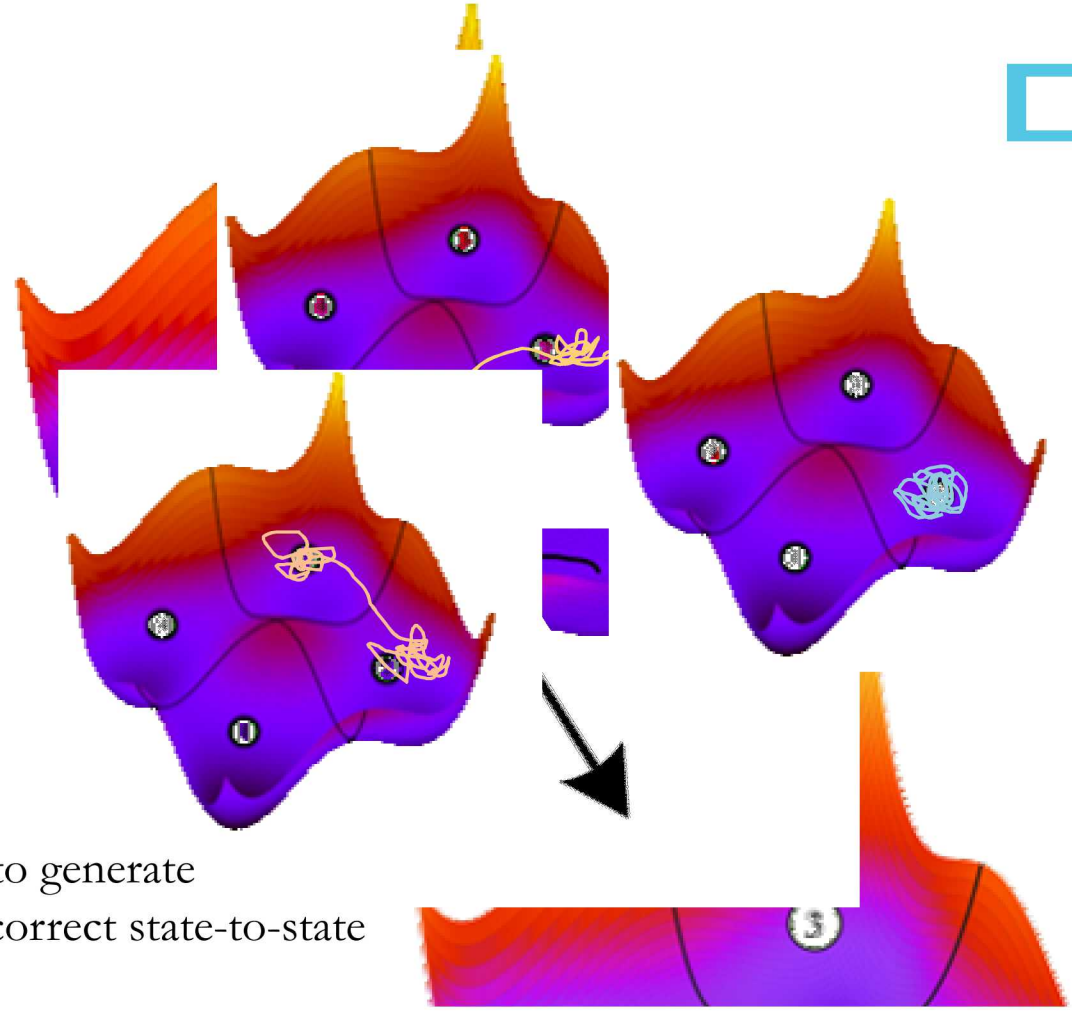


Parallel in Space, Time

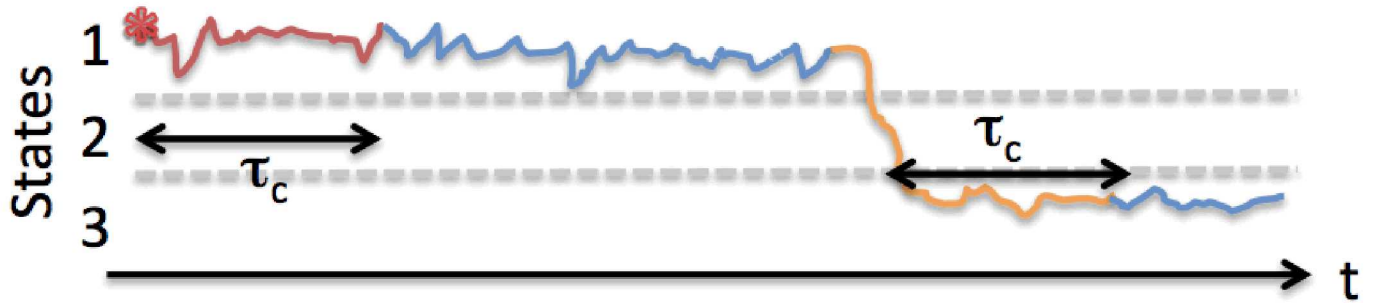
- Atoms/particles in space can be distributed across processors
- Need to track particles in nearby domains, reconstruct neighbor lists as particles move
- for all time;

  Compute forces, update atom positions





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



[Le Bris, Lelievre, Luskin, and Perez, MCMA 18, 119 (2012)]

Parallel in Time

Task Management

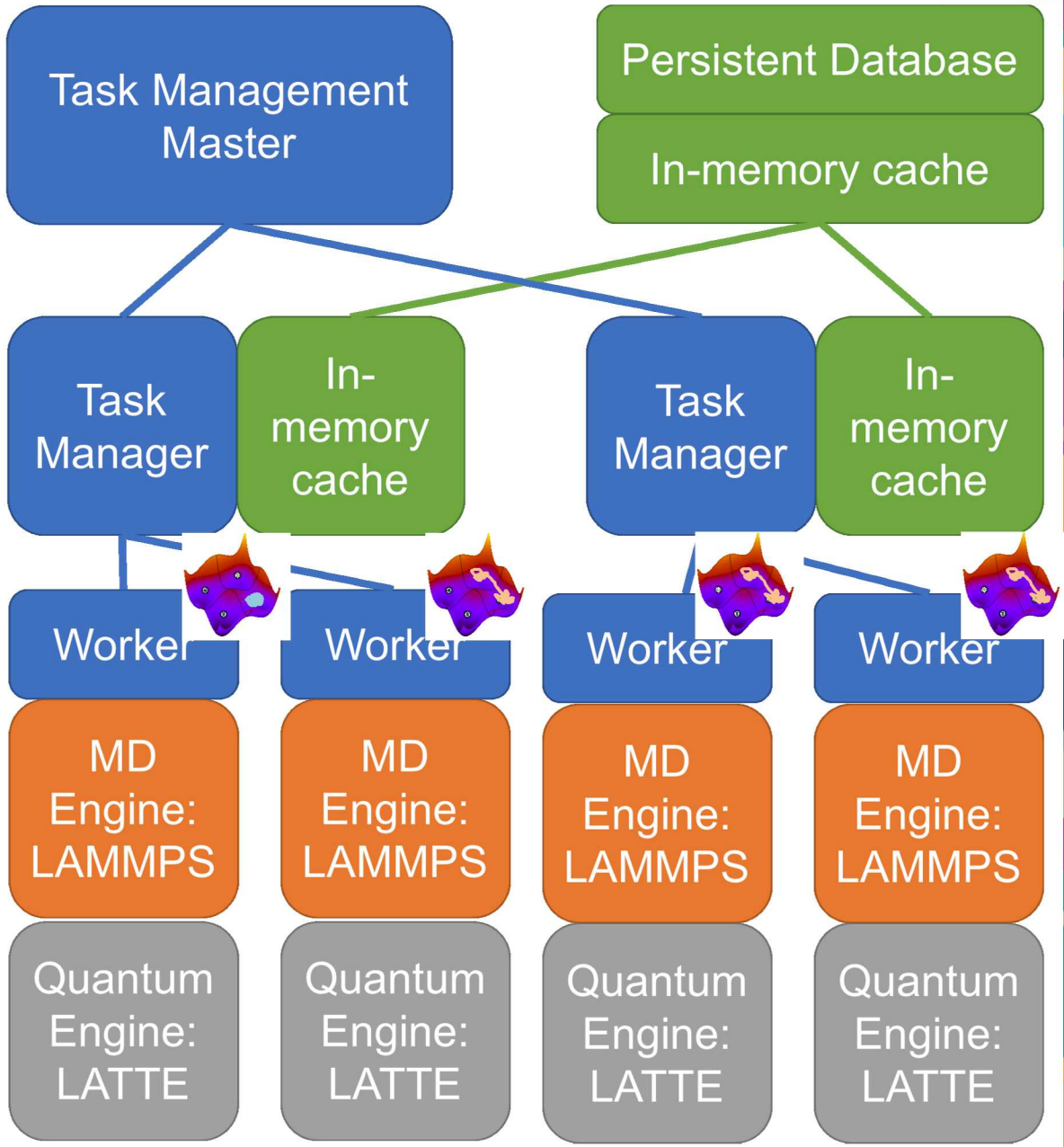
- Asynchronous master/slave architecture
- Master executes replica-based Accelerated MD algorithms
- Task managers prefetch tasks, fulfill data dependencies, consolidate results, and cache intermediates

Data Management

- Key-value store
- Persistent backend
- Hierarchical in-memory caches

Computational Engines

- Workers wrap instances of LAMMPS that run on ~1-1000 cores
- Quantum capabilities (at the DFTB level) provided by LATTE, upcoming binding to NWChem



Parallel in Time

Task Management

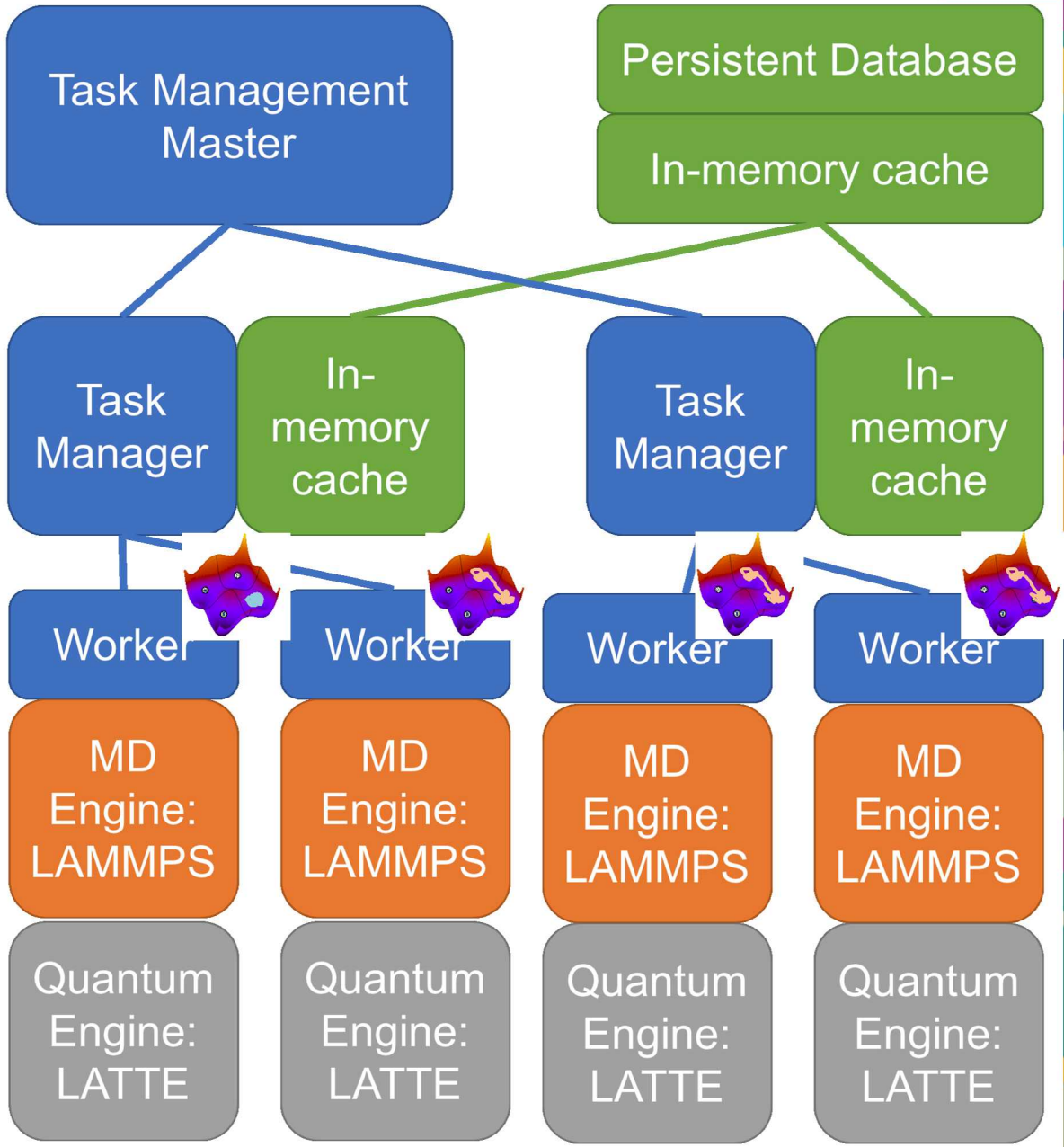
- Asynchronous master/slave architecture
- Master executes replica-based Accelerated MD algorithms
- Task managers prefetch tasks, fulfill data dependencies, consolidate results, and cache intermediates

Data Management

- Key-value store
- Persistent backend
- Hierarchical in-memory caches

Computational Engines

- Workers wrap instances of LAMMPS that run on ~1-1000 cores
- Quantum capabilities (at the DFTB level) provided by LATTE, upcoming binding to NWChem

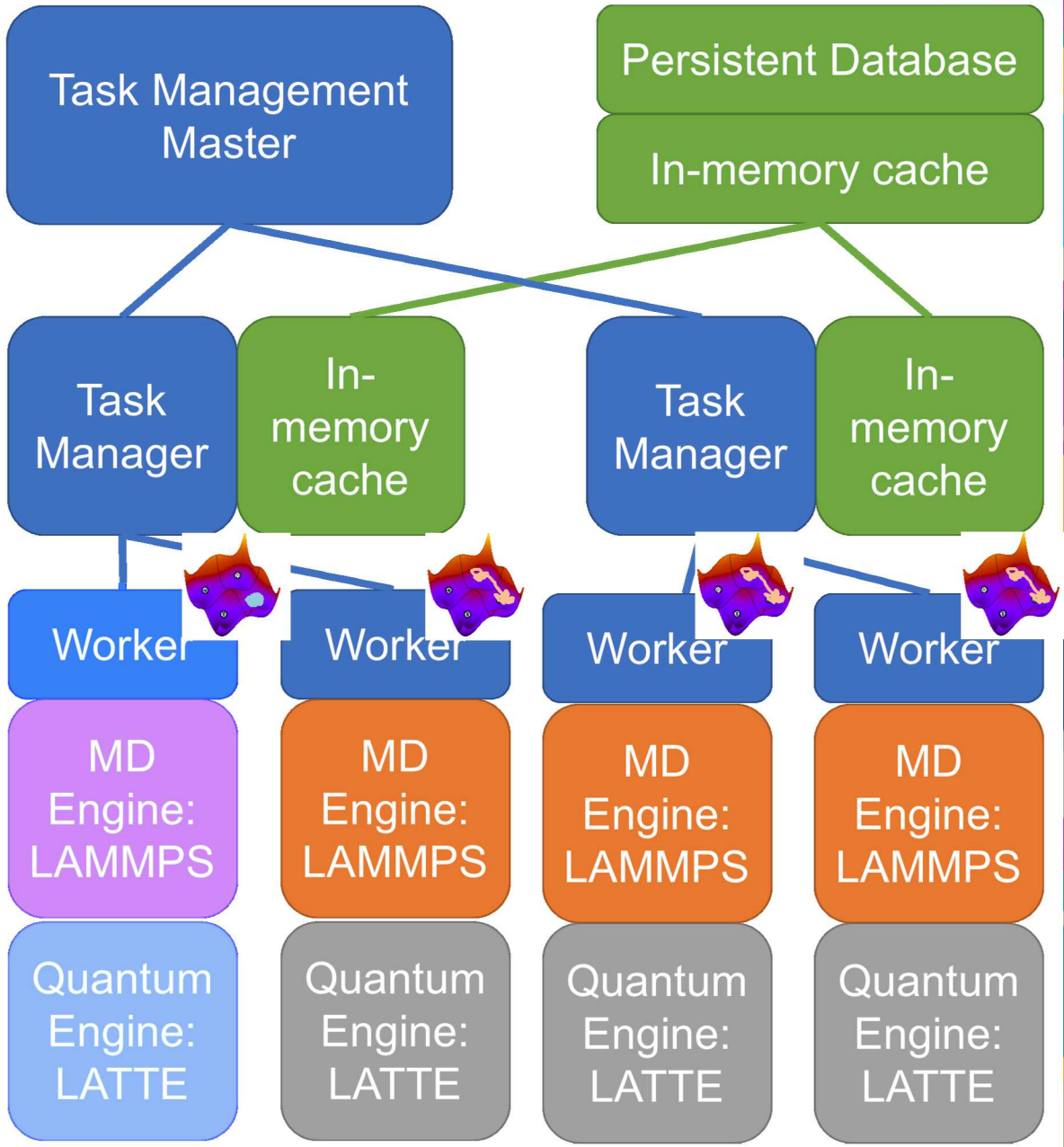


Parallel in Time

Computational Engines

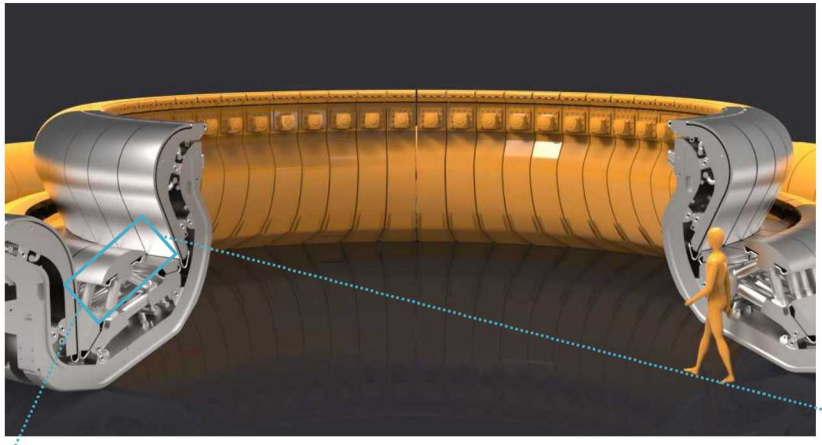
- Workers wrap instances of LAMMPS that run on ~1-1000 cores
- Quantum capabilities (at the DFTB level) provided by LATTE, upcoming binding to NWChem
- ParSplice can initialize workers in newly discovered states to improve diversity of spliced segments.

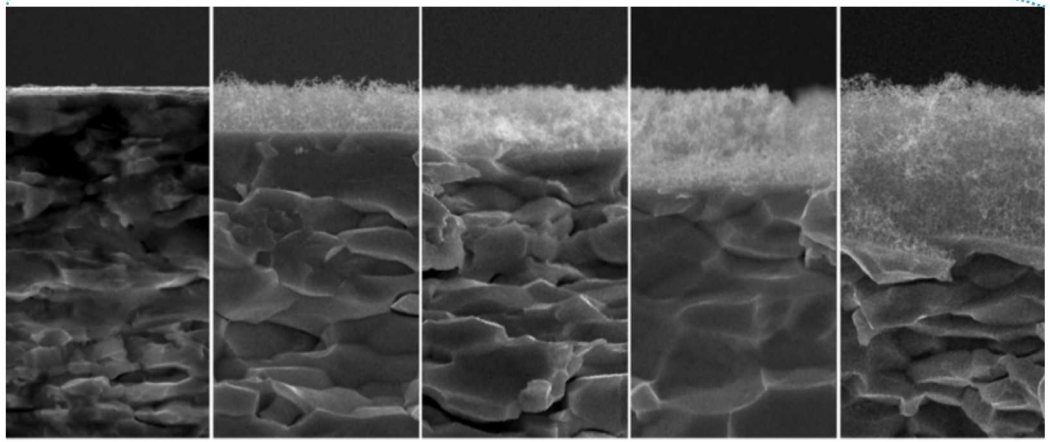
- Flowchart depicts all workers having similar tasks (MD), not a strict rule
- LAMMPS can act as a client that calls a server for information (forces)
- Error checking or multi-fidelity trajectories can be assembled (IAP, TB, DFT)



Applications of EXAALT

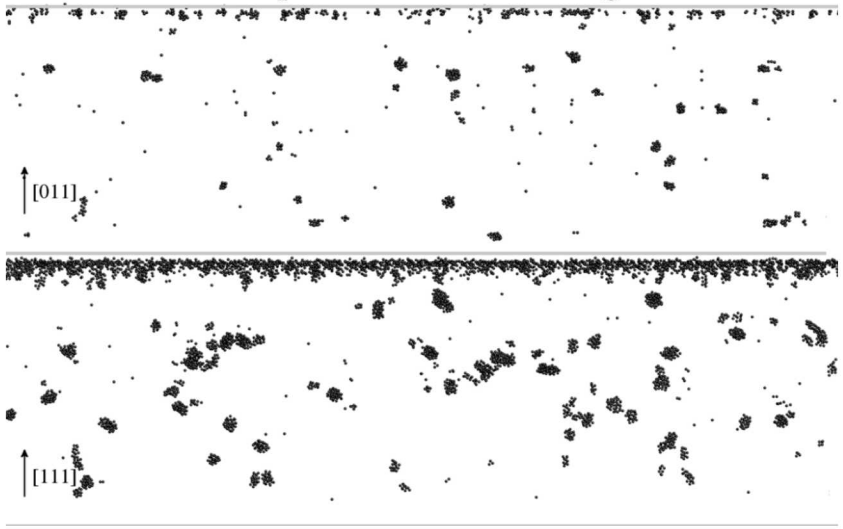
- ITER Fusion Reactor:



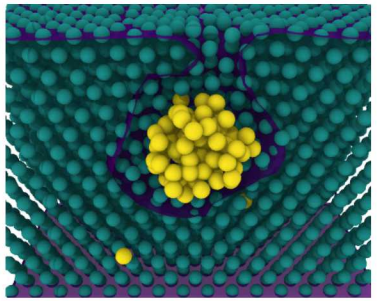


- Plasma facing material is W at divertor, Be at inner wall
- Exposure to He, H and other plasma species at high temperatures >1000K
- Fuzz buildup limits power output and useful lifetime of divertor.

Near surface implantation, clustering of He



nm-scale He bubble

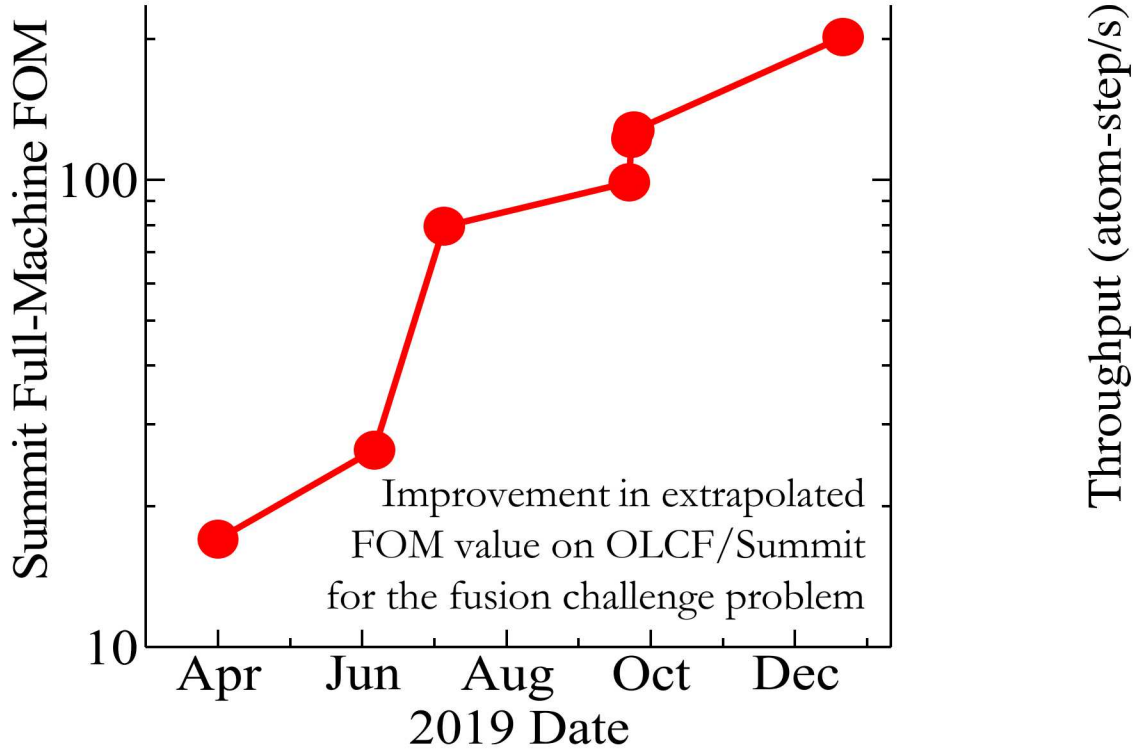


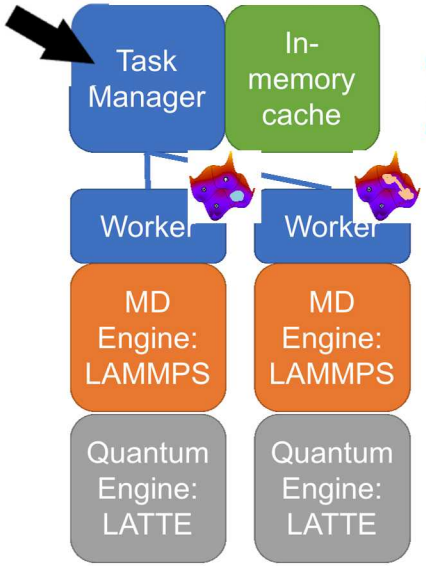
Luis Sandoval, Blas Uberuaga, Danny Perez, Art Voter, Phys. Rev. Lett. (2015)
Sefta, F., Hammond, K. D., Juslin, N., & Wirth, B. D. (2013). *Nuclear Fusion*, *53*(7), 073015.

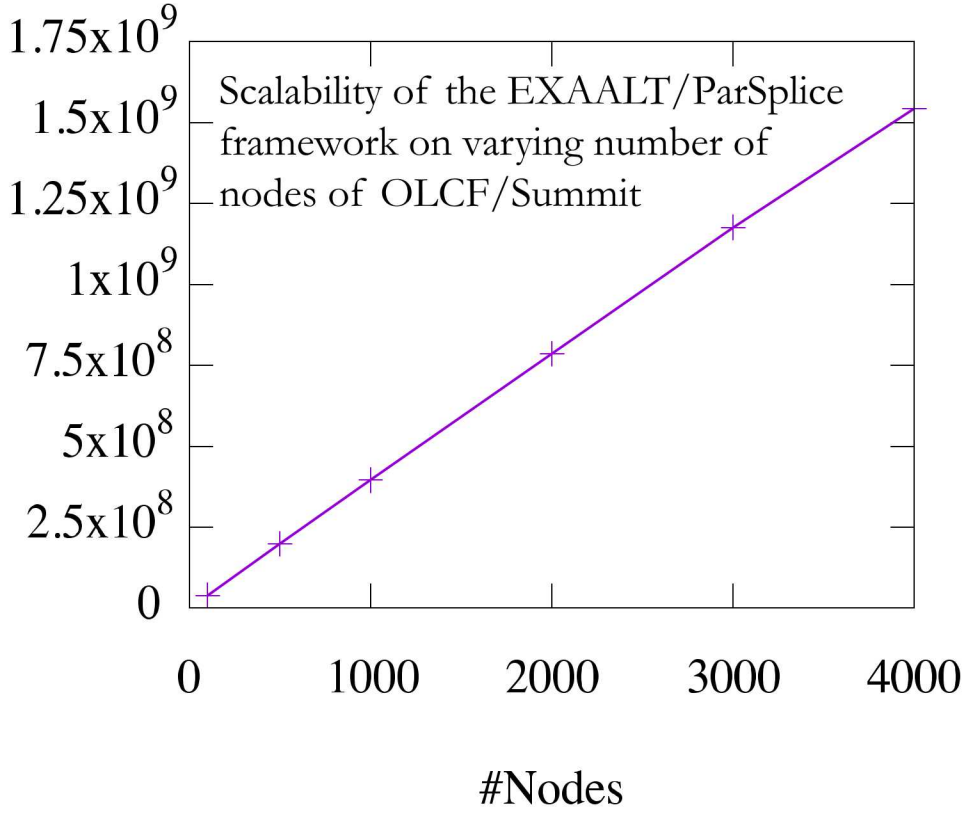
Performance Progress

Figure of Merit

- Atoms * Timesteps / second
  - Improved by atoms/replica, number of replicas, trajectory segments spliced together
- Fusion energy FOM uses SNAP with 205 bispectrum components

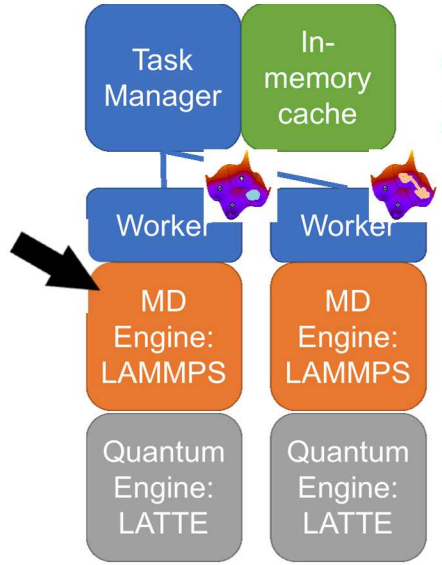






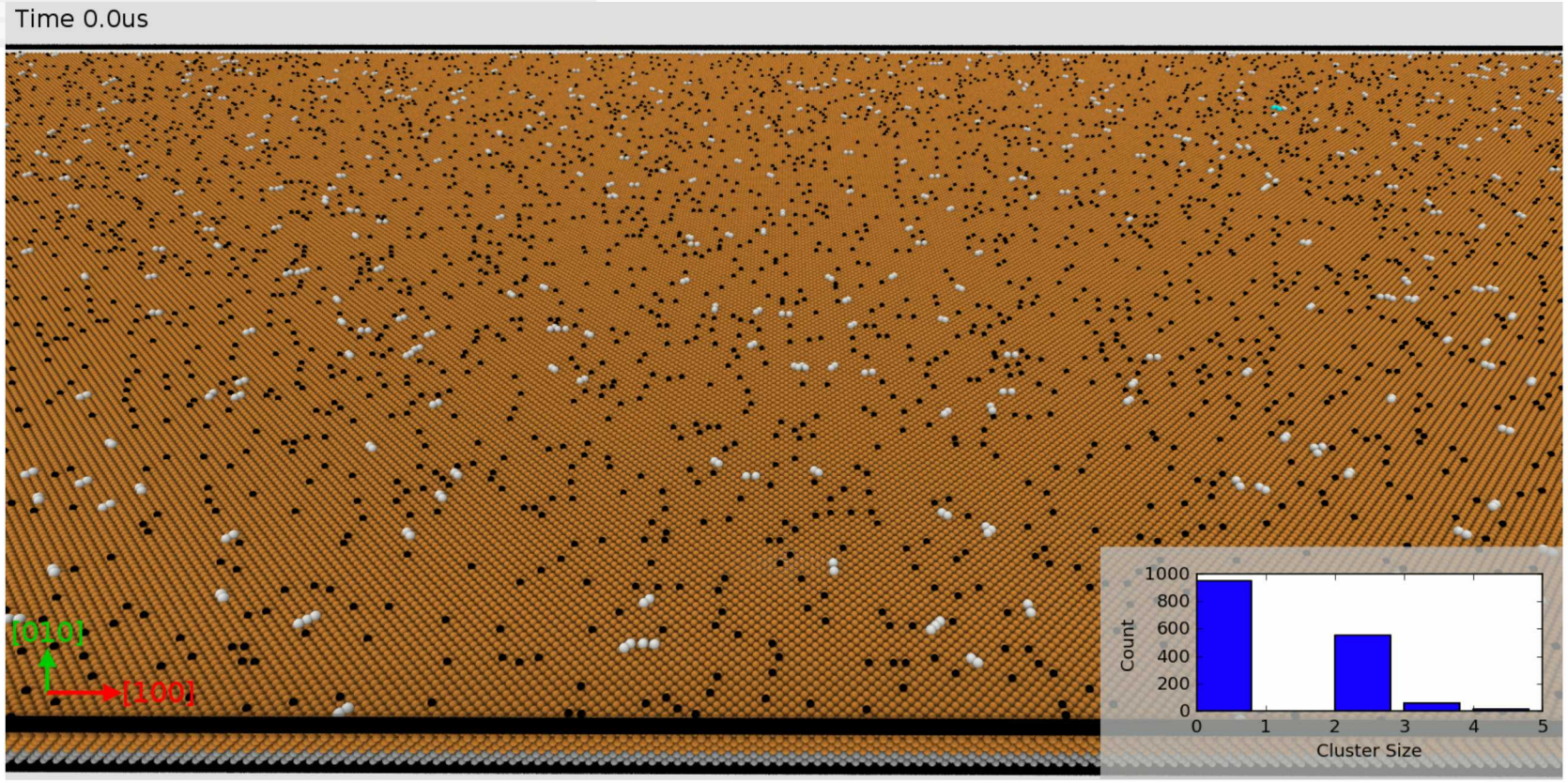
New Time Acceleration Methods

(Voter, Perez, Plimpton)



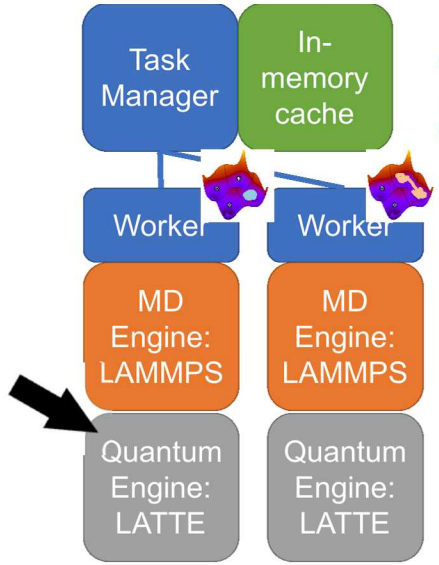
Local/Global Hyperdynamics

- Adds a bias potential to 'strained' atom pairs, enables more rapid events.
- Boost is proportional to energy barrier for rare events
  - 0.4eV barrier, run @ 400K → 500x boost

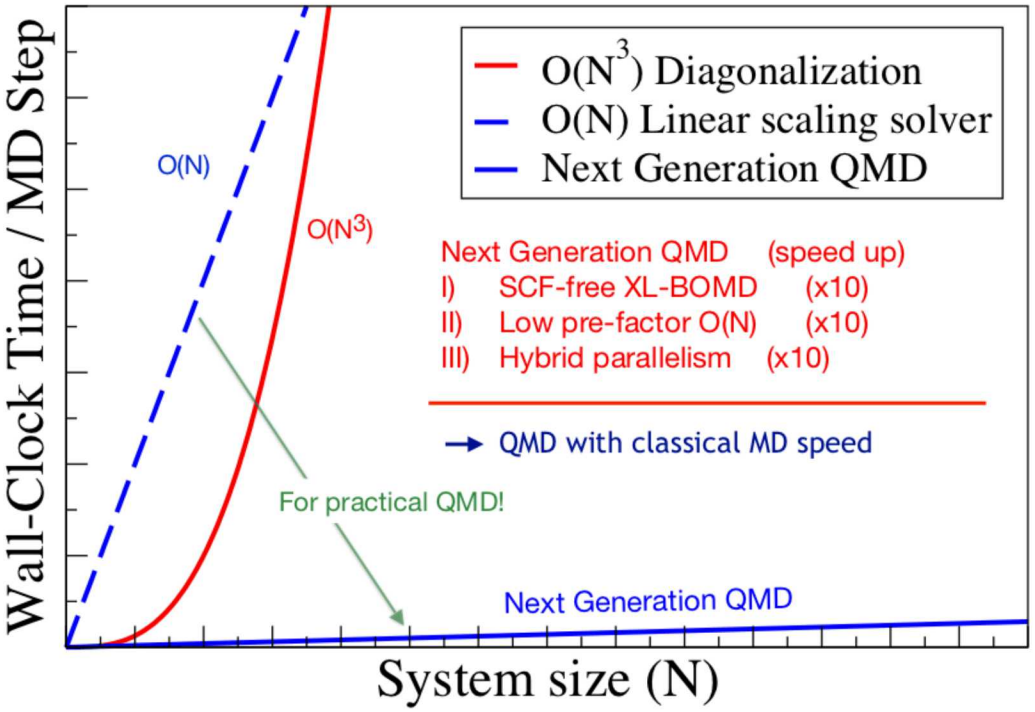


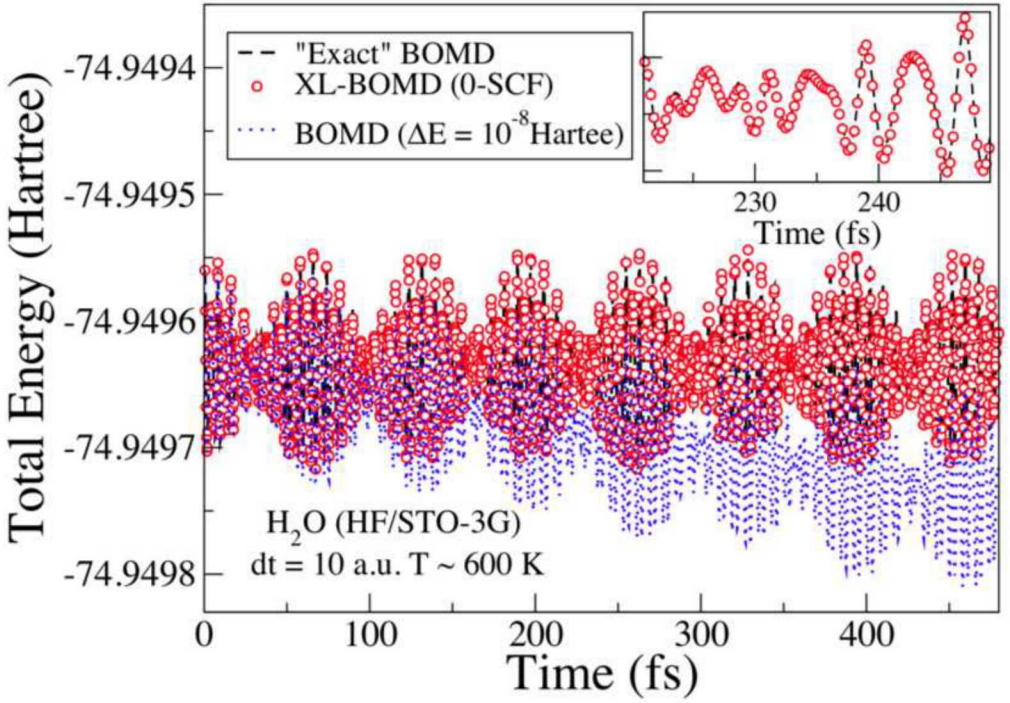
- Movie frames are 2us apart, 500us total
- 2.4M substrate atoms, 12k adatoms
- Run on 64 nodes of Theta (ALCF)

Efficient Quantum Engines (Niklasson, Cawkwell, Negre…)



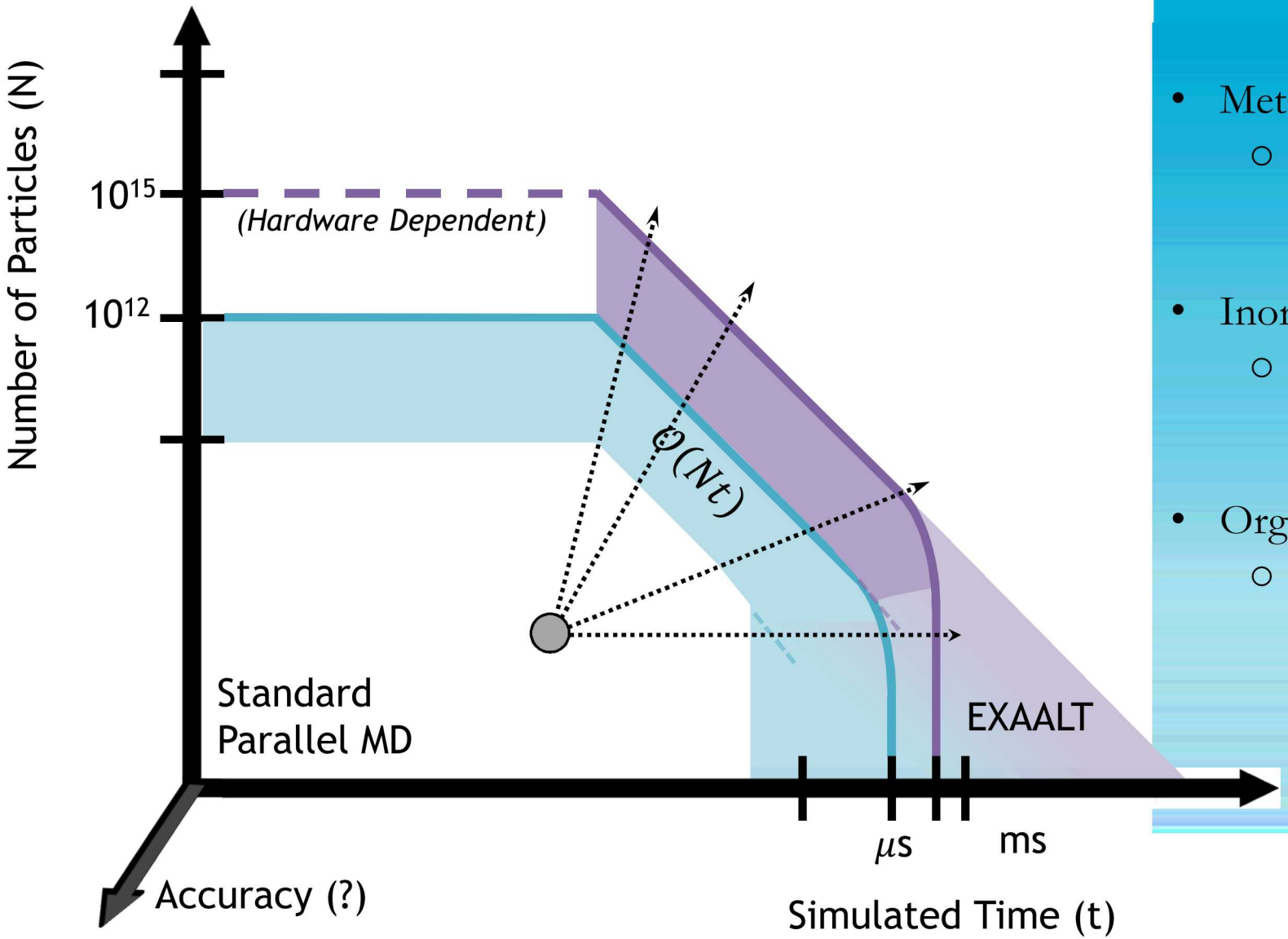
- Linear scaling quantum methods are useless without a low scaling prefactor
- Extended Lagrangian Born-Oppenheimer MD prodices stable trajectories, even without a self-consistent field calculation at each timestep.





Niklasson & Cawkwell JCP 141,164123 (2014); Niklasson JCP 054103 (2017)

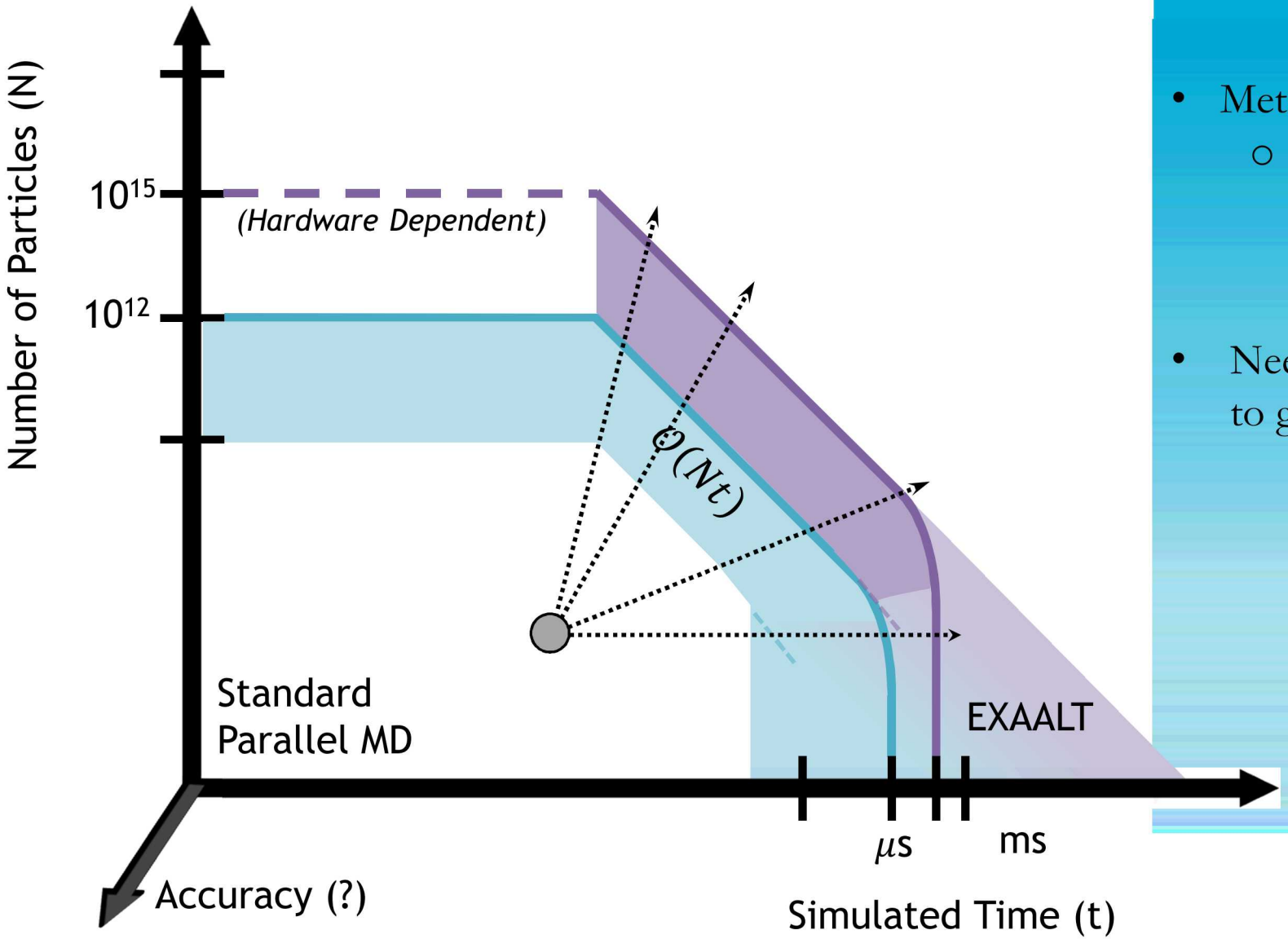
Return to the Master Plot



Classical, Empirical Potentials

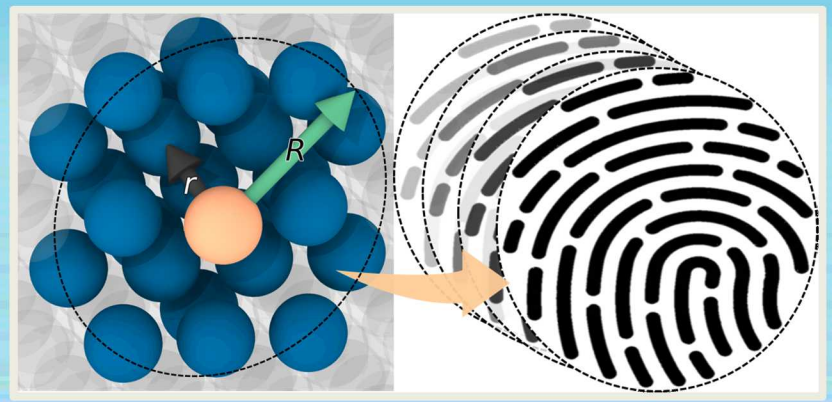
- Metals
  - EAM: Assume spherical electron density
    $E_i = F_\alpha\left(\sum_{j\neq i}\rho_\beta(r_{ij})\right) + \frac{1}{2}\sum_{j\neq i}\emptyset_{\alpha\beta}(r_{ij})$
- Inorganic
  - Stillinger-Weber: Assume 2,3-body harmonic springs
- Organic
  - ReaxFF: Assume covalent bonding, smooth bond-orders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



MD Approximations Change Over Time

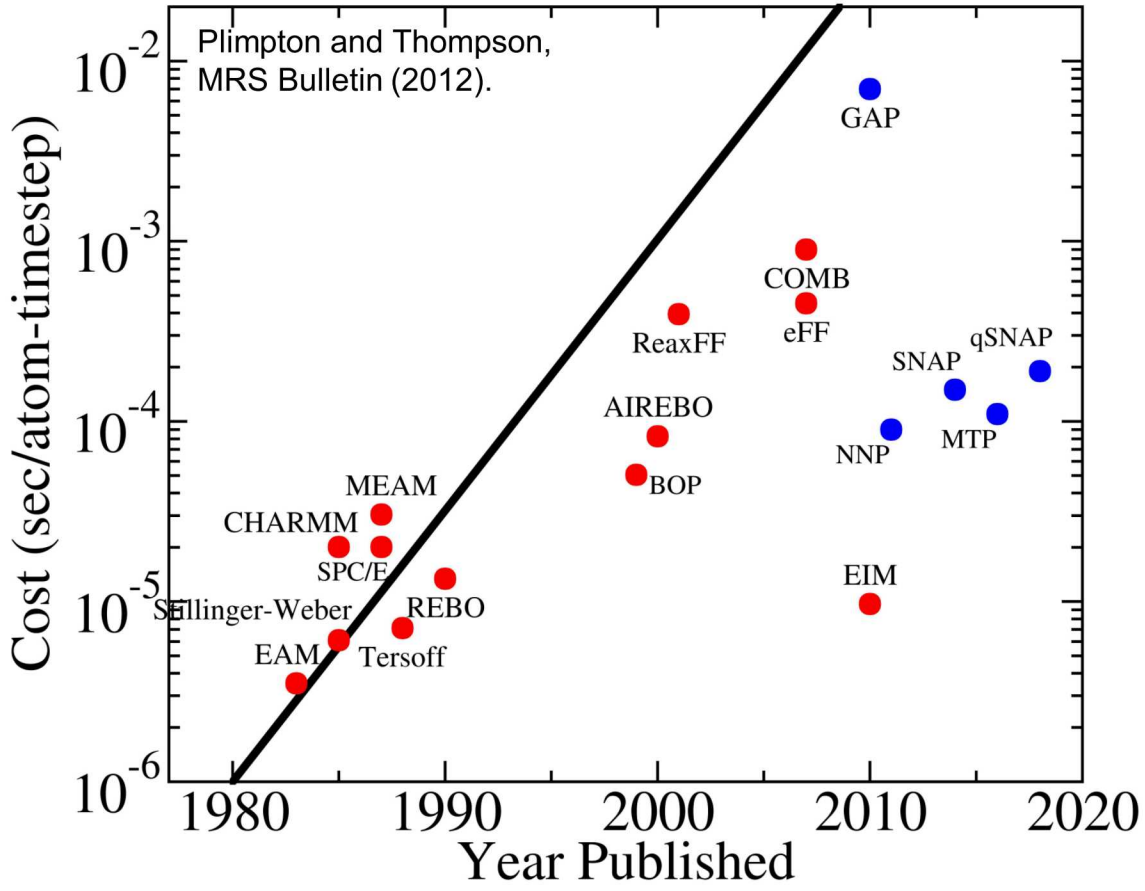
http://lammps.sandia.gov

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

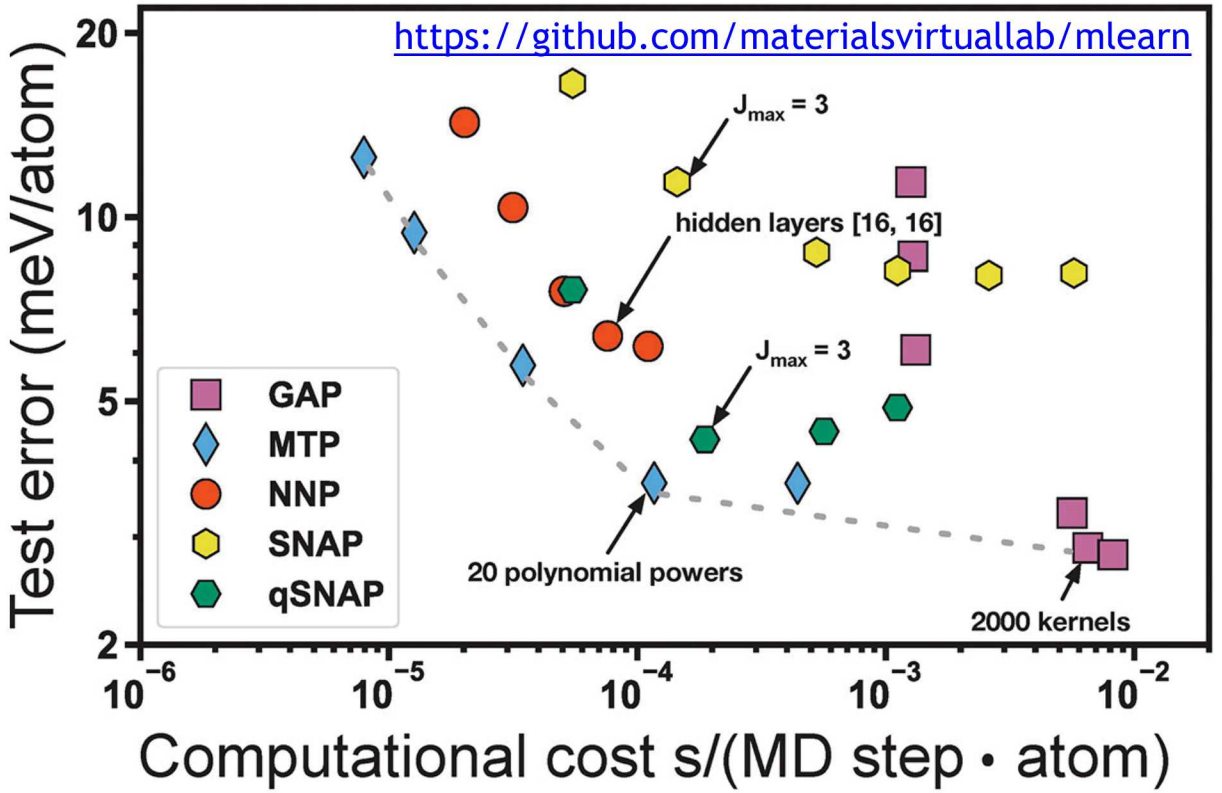
Manybody (1980s)
Stillinger-Weber, Tersoff, Embedded Atom Method

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

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



Resources are limited, which is your best choice?



But Empirical Potentials are Faster! For now…

- EXAALT benchmark uses 205 bispectrum coefficients, tungsten crystal
- Mira (IBM BG/Q) baseline: 0.182 Katoms-steps/s/node * 49152 Mira nodes

Mira baseline, 1 replica/node

- 2018 LAMMPS performance on Summit: 33.7 Katom-steps/s/node * 4608 Summit nodes
  Integrated more optimizations from TestSNAP into the Kokkos-LAMMPS

17.4x faster than Mira baseline

- Current LAMMPS on Summit: 262.0 Katom-steps/s/node * 4608 Summit nodes

134x faster than Mira baseline

- Evan Weinberg (NVIDIA) has added additional optimizations to Kokkos SNAP in LAMMPS
  New version being tested by CoPA and EXAALT project members
  Not yet released, but should be merged into master LAMMPS soon
- Unreleased LAMMPS on Summit: 407.7 Katom-steps/s/node * 4608 Summit nodes

210x faster than Mira baseline

And were not done yet!

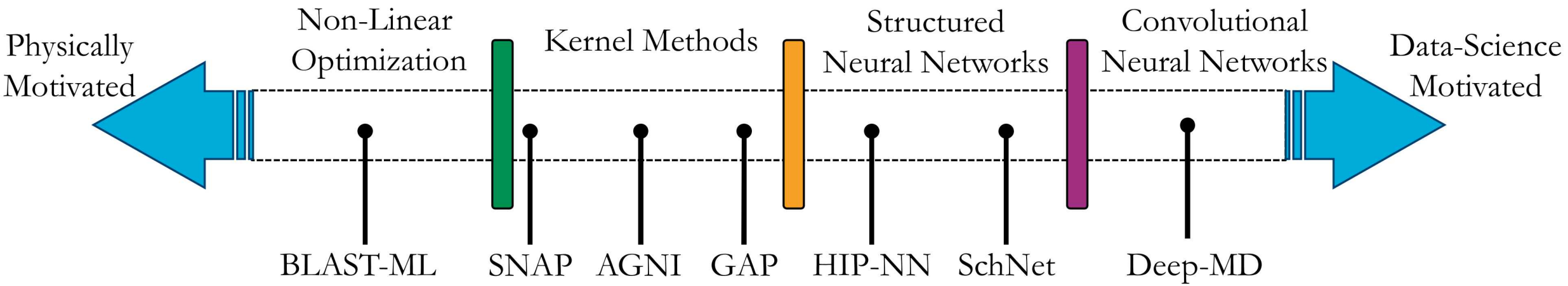
(Thompson, Moore)

(Lubbers, Germann)

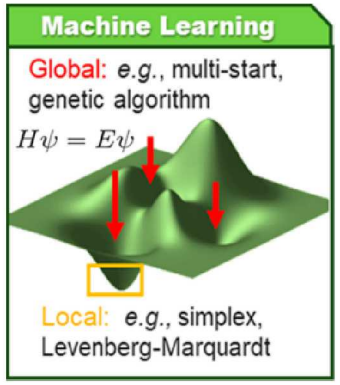
(Gayatri)

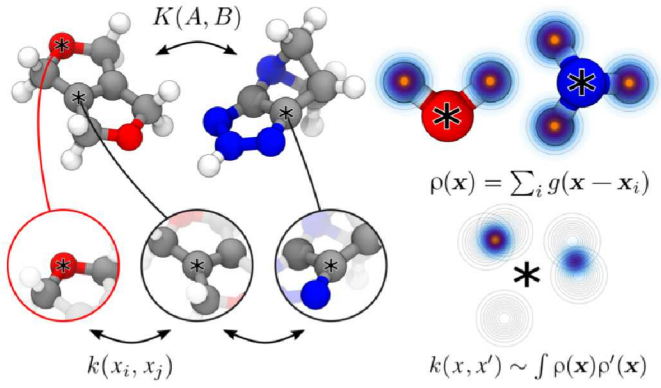
(Weinberg)

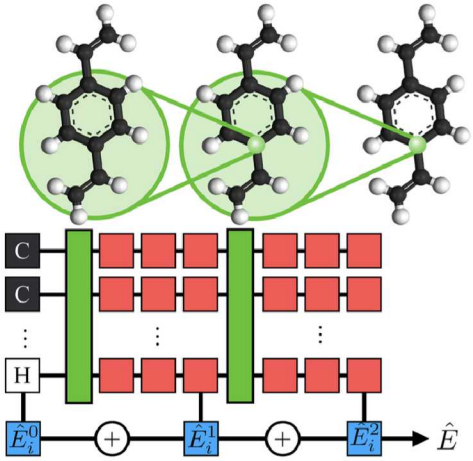
Environment of Machine Learning Techniques

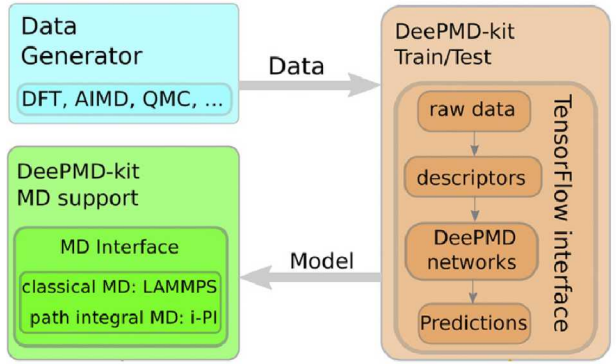


- Adoption of machine learning techniques within molecular dynamics has been varied





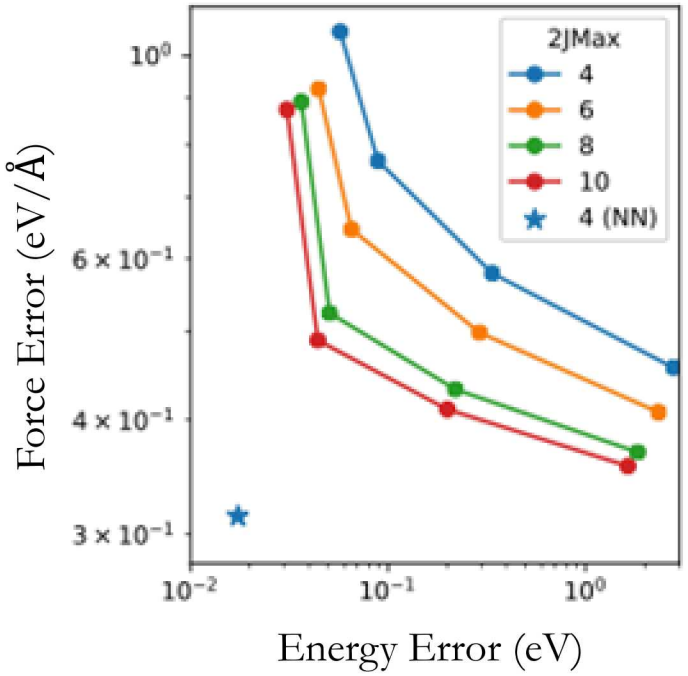


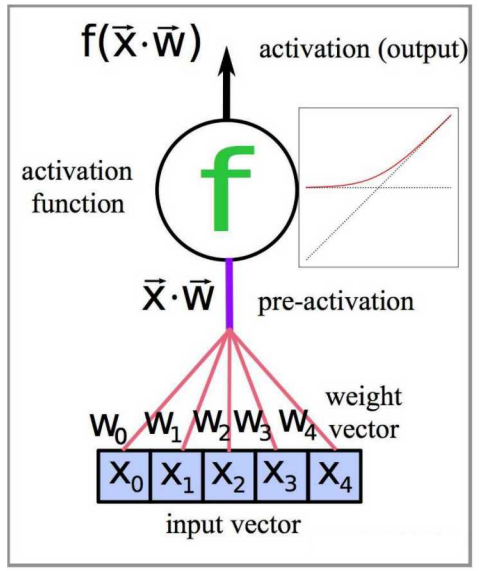


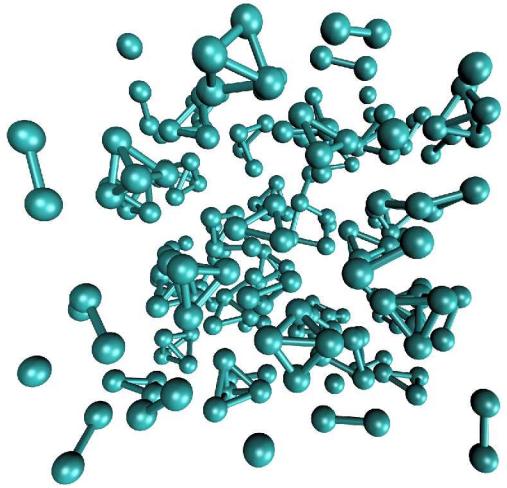
SNAP Development – Neural Networks

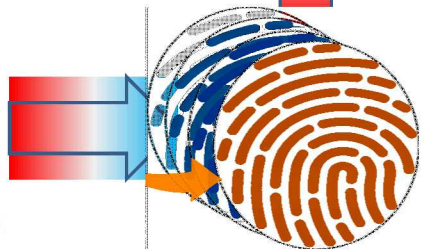
Motivation

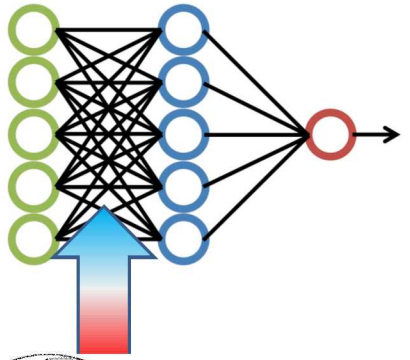
- Use Neural Networks to build better potentials using fewer descriptors
- Tackle big data: training to hundreds of thousands of atomic environments
- NN fitting techniques designed to naturally generalize to new configurations
- Preliminary NN/SNAP W/Be potentials show substantial improvements in force accuracy and dramatic improvements in energy accuracy











(Lubbers, Smith…)

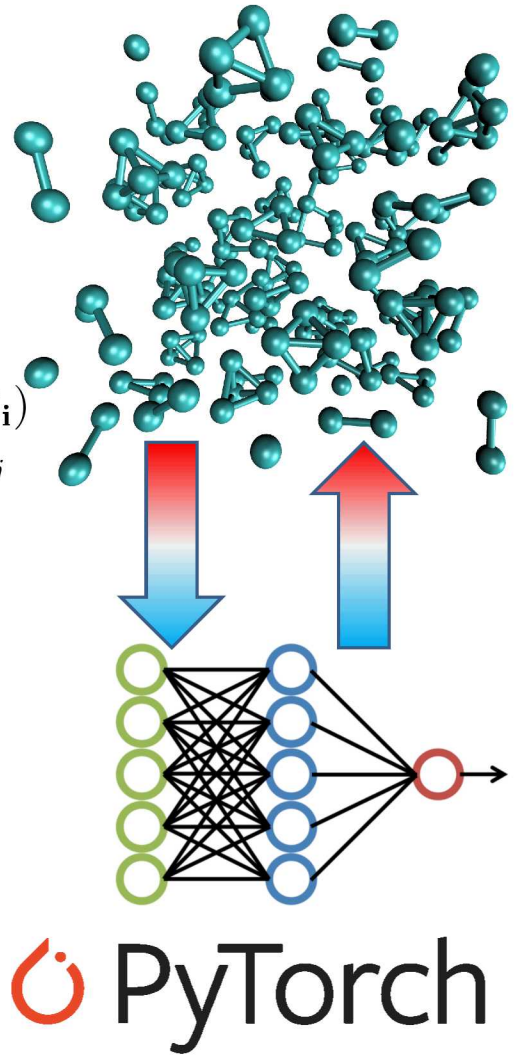
SNAP Development – Neural Networks

Infrastructure

- Efficiently couple LAMMPS and PyTorch using Cython
- Neural Networks in PyTorch for fast and flexible implementation of network structures
- Nonlinear SNAP models required reformulation of SNAP gradient calculations – lead to new insights into SNAP algorithm and computation speedups for Linear SNAP models as well
- Force fitting algorithm for nonlinear energy models

LAMMPS::ForceCompute():

for i in AtomList: $\mathbf{B}_i = \text{ComputeB}(i)$
\# Invoke callback to Python:
$E = \text{NN}(\vec{\mathbf{B}})$;
$\vec{\boldsymbol{\beta}} = \nabla_{\vec{\mathbf{B}}} E$
\# Back to LAMMPS
for j in FullAtomList: $\mathbf{F}_j = 0$
for i in AtomList:
  for j in NeighList(i):
    $\mathbf{F}_{ij} = \text{ComputeFij}(\boldsymbol{\beta}_i, \mathbf{B_i})$
    $\mathbf{F}_i \mathrel{+}= -\mathbf{F}_{ij}$; $\mathbf{F}_j \mathrel{+}= \mathbf{F}_{ij}$



Los Alamos NATIONAL LABORATORY EST. 1943

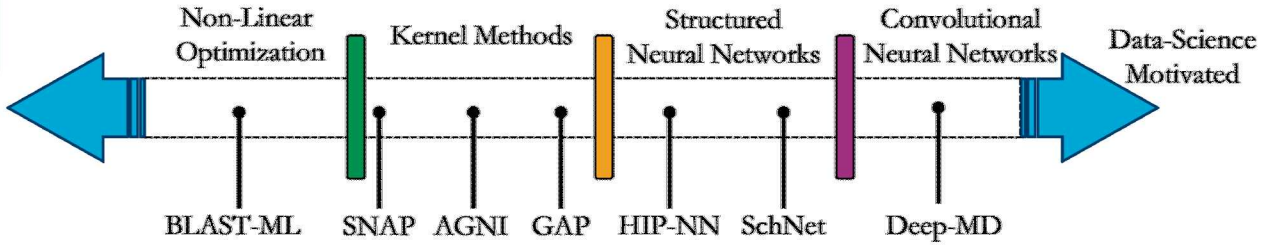
(Lubbers, Smith…)

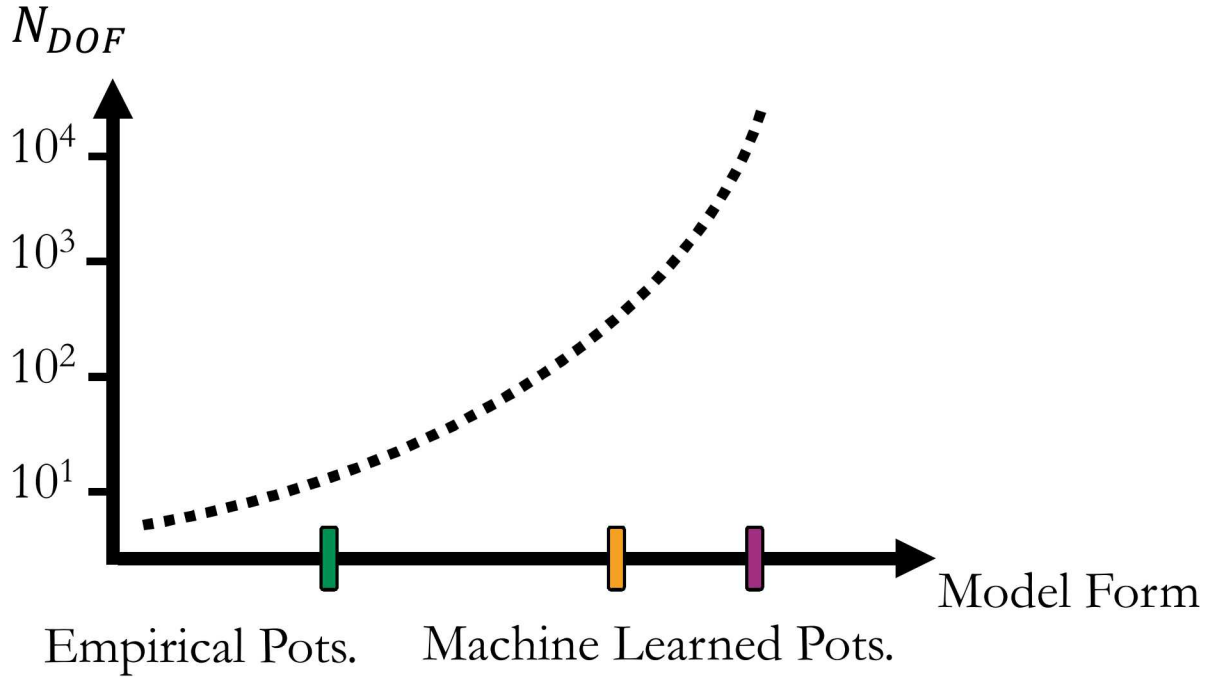
State-to-State Dynamics for Training

Designing Better Training

- Training data sets will be incomplete, ML-IAP need to be 'well-behaved' when extrapolating

- Structure of atomic configurations is highly correlated, still exploring what descriptors + training methods makes most sense.





Data Needs:

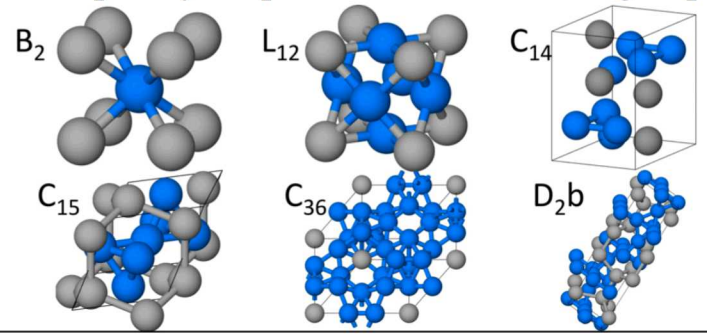
- When $N_{DOF} \sim N_{Train}$, high risk of overfitting → Poor Interpolation
- When training diversity is low → Poor Extrapolation
- RUNNING MD WILL EXPOSE THESE SHORTCOMINGS

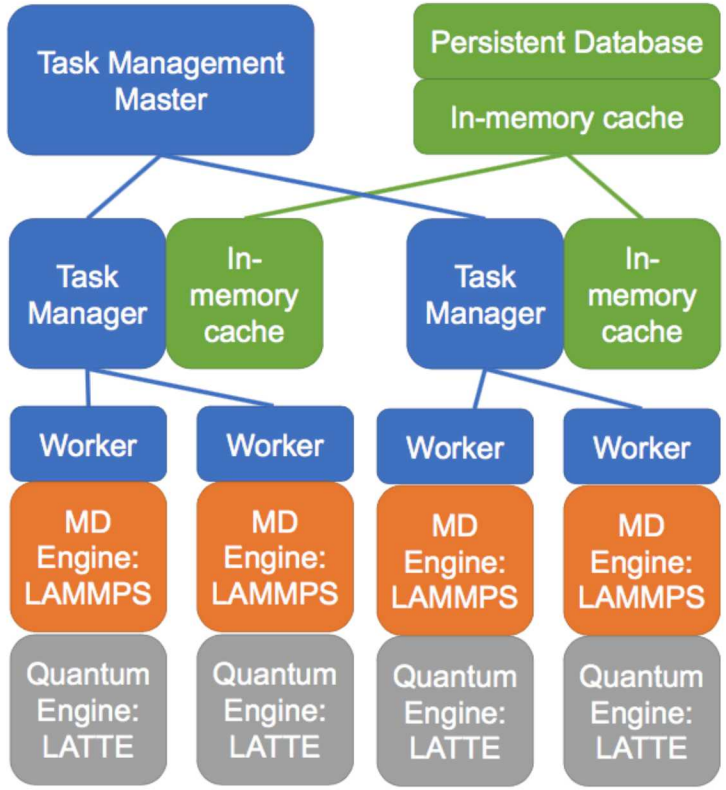
"Performance and Cost Assessment of Machine Learning Interatomic Potentials" J. Phys. Chem. A (2020)

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:





User Generated Training

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

Learn-on-the-Fly

- Framework of time acceleration tools can generate new training by running MD with lots of replicas
- Resource demand is VERY HIGH, but can produce the ideal general use potential.

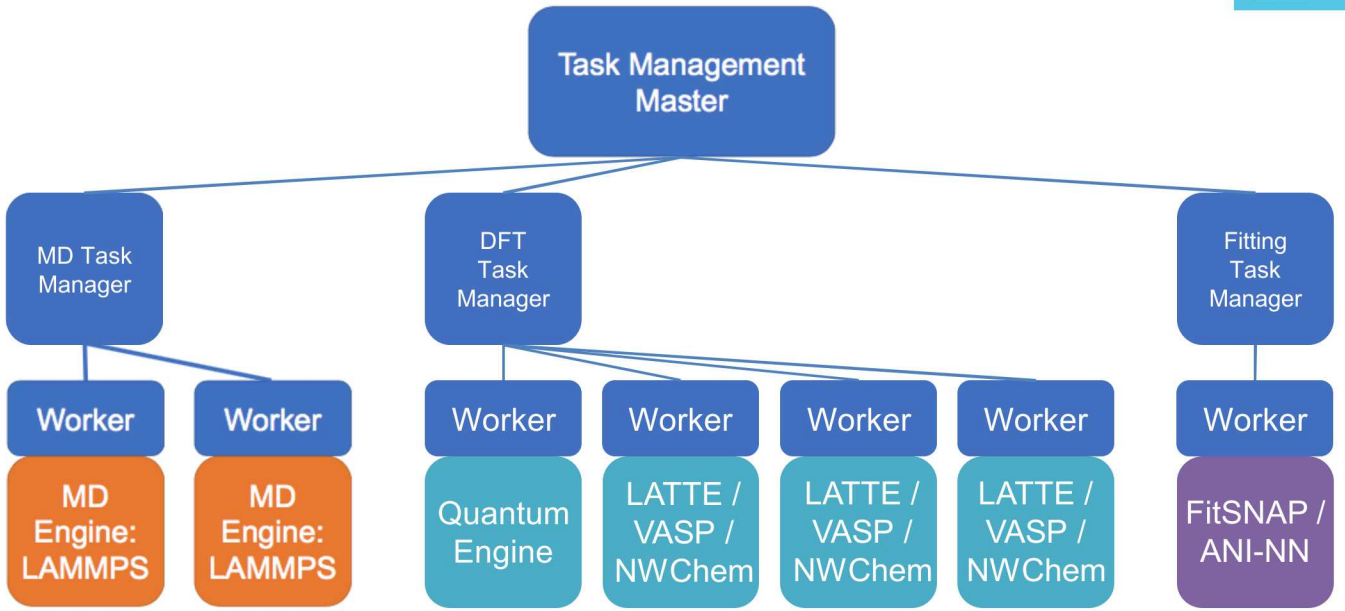
Learn On The Fly

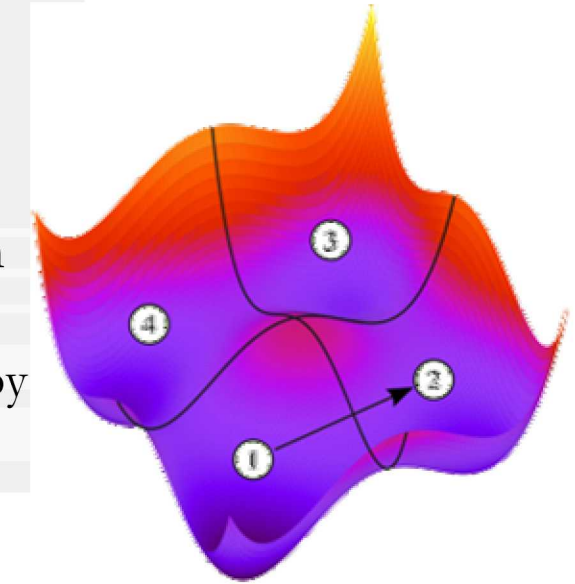
Workload Management

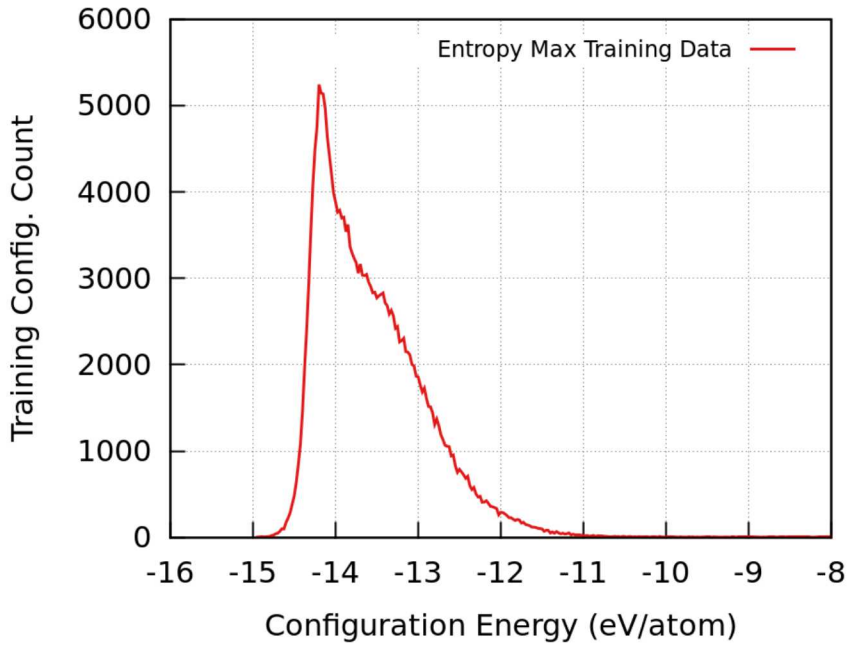
- Computing resources now need to be distributed proportional to the cost of the individual task (MD, DFT, Fitting)

- New training configurations from MD do not need to be reached in a time-accurate manner.
  - Currently still running small (~100atoms) problems for MD in order to simplify the handshake w/ DFT.

- An 'entropy maximized' training set has been generated using bispectrum components (SNAP descriptors) as an information entropy
  - ~230k unique configurations generated in 16h hours on Theta (ALCF)

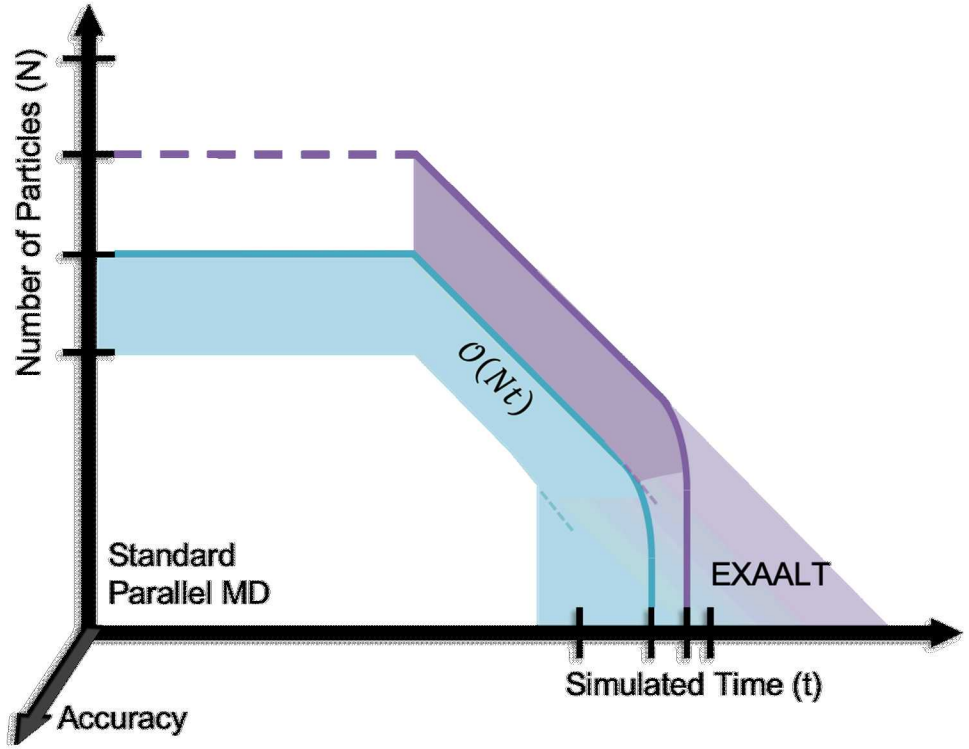






Conclusions and Path Forward

- 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



- New time accelerated methods (Hyperdynamics, sub-lattice ParSplice) made available
- Data-driven interatomic potentials (SNAP, SNAP-NN) allow for MD predictions of challenging material problems.

Contact Information:

mitwood@sandia.gov
danny_perez@lanl.gov

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