SAND2020-2640C

Sandia National Laboratories

Predictive Atomistic Simulations of Materials using SNAP Data-Driven Potentials

Aidan Thompson
Center for Computing Research,
Sandia National Laboratories,
Albuquerque, New Mexico

APS March Meeting, Denver, CO

(3/3/2020)

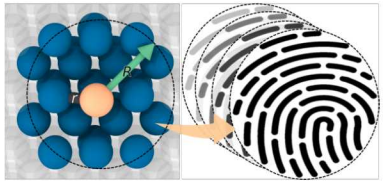
U.S. DEPARTMENT OF ENERGY NNSA

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.

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.

SNAP Training Workflow (FitSNAP)

https://github.com/FitSNAP/FitSNAP



Model Form

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

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

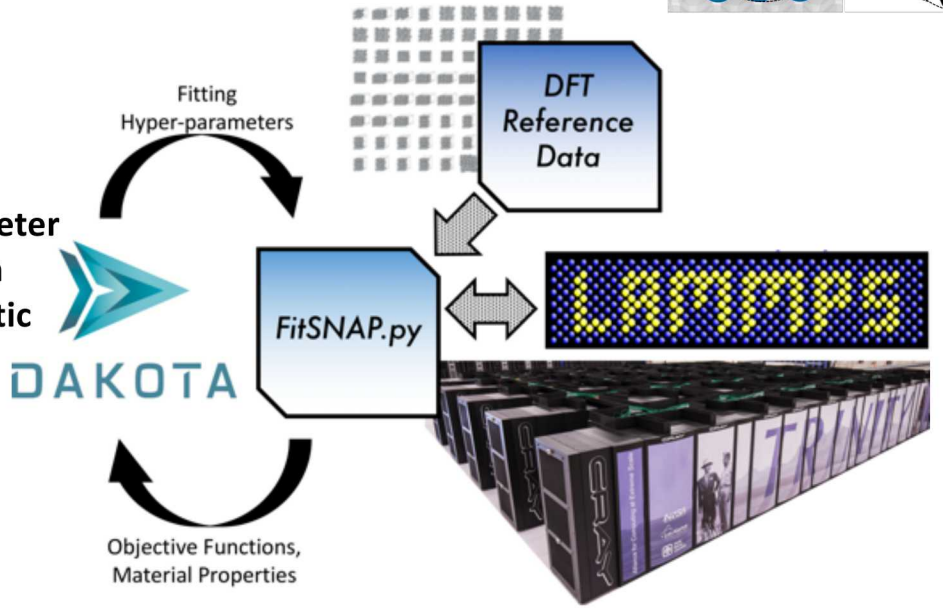
Regression Method

- $\boldsymbol{\beta}$ vector fully describes a SNAP potential
- Decouples MD speed from training set size

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

Weights — Set of Descriptors — DFT Training — Regularization Penalty

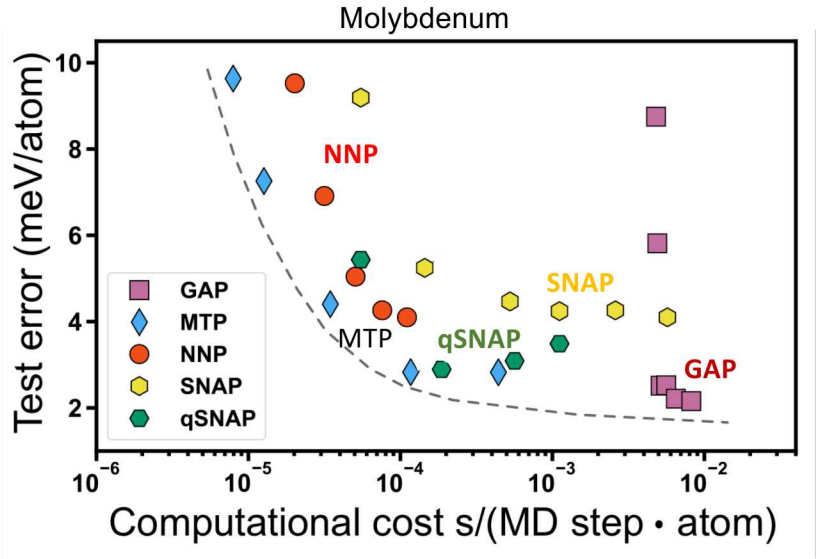
Hyperparameter Optimization (SOGA Genetic Algorithm)



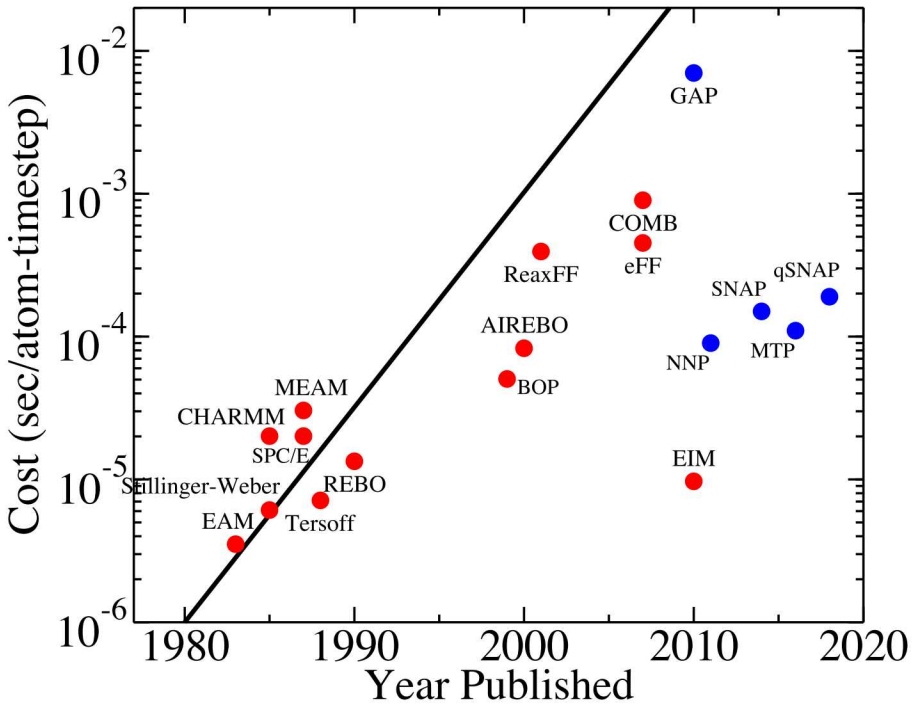
Comparing ML Potentials

https://arxiv.org/abs/1906.08888
https://github.com/materialsvirtuallab/mlearn

"Performance and Cost Assessment of Machine Learning Interatomic Potentials," *J.Phys.Chem A (2020)* , Shyue Ping Ong (UCSD), with: Csanyi (2010), Shapeev (2015), Behler(2007), Thompson (2015), Wood (2018)



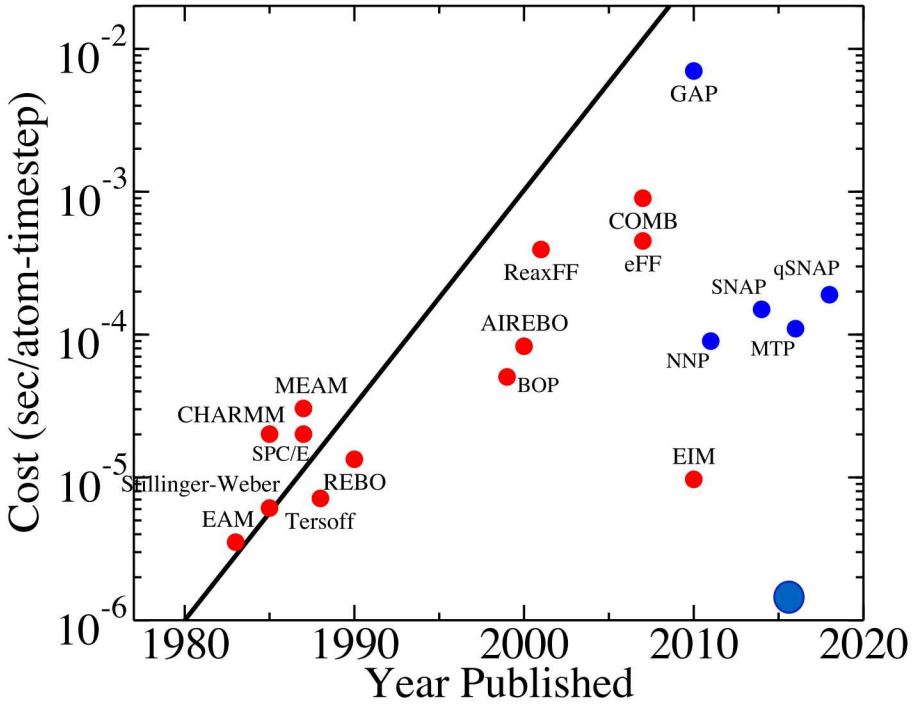
Plimpton and Thompson, MRS Bulletin (2012).



SNAP GPU Performance Improvements

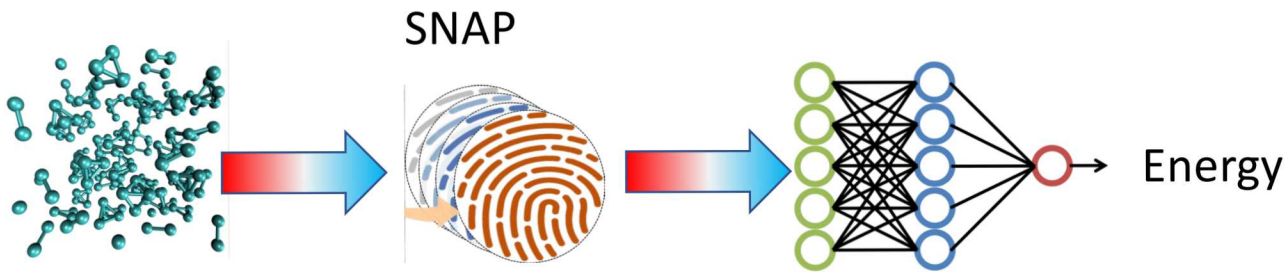
- Joint effort by Aidan Thompson (EXAALT), Stan Moore (CoPA), Rahul Gayatri (NESAP), Sarah Anderson (Cray), Evan Weinberg (NVIDIA)
- Created stripped-down proxy code (TestSNAP)
- Completely rewrote TestSNAP to reduce flops and memory
- Explored many different GPU strategies, using OpenACC and CUDA
- Greatly improved memory-access patterns
- Ported best implementation back to production code with Kokkos
- More than 10x improvement relative on V100 GPUs
- ORNL Summit node (2000 atoms/node spread over 6 V100 GPUs)

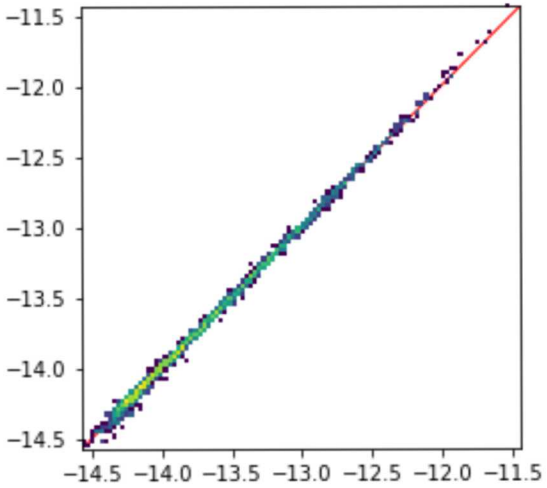
- 2M atom-steps/s ~ 1000 ns/day

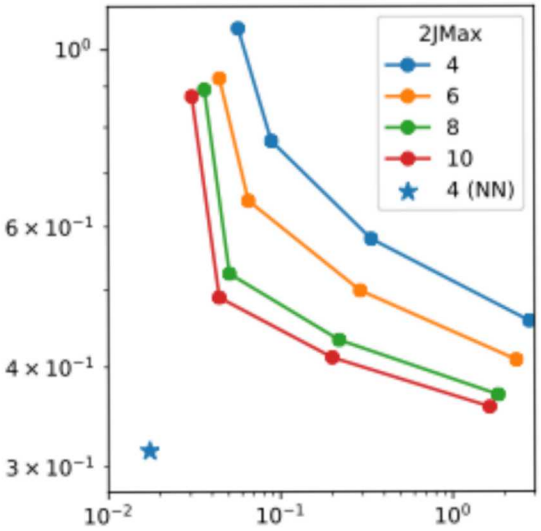


SNAP with Neural Networks

- Initial Results
  - **Integration of LAMMPS with PyTorch using Cython**
  - Energy regression performance of the NN/SNAP model.
  - Dataset of 20k disordered tungsten systems.
- Ongoing Work
  - Optimize ANN architecture
  - Train on large dataset (~20M points)
  - Integrate DFT into workflow
  - Active learning



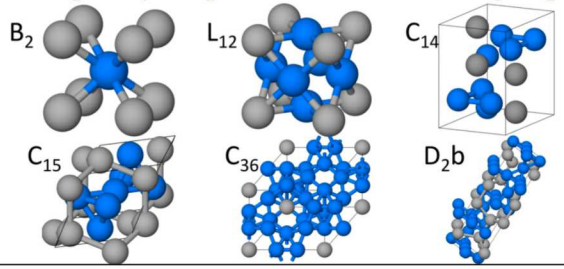


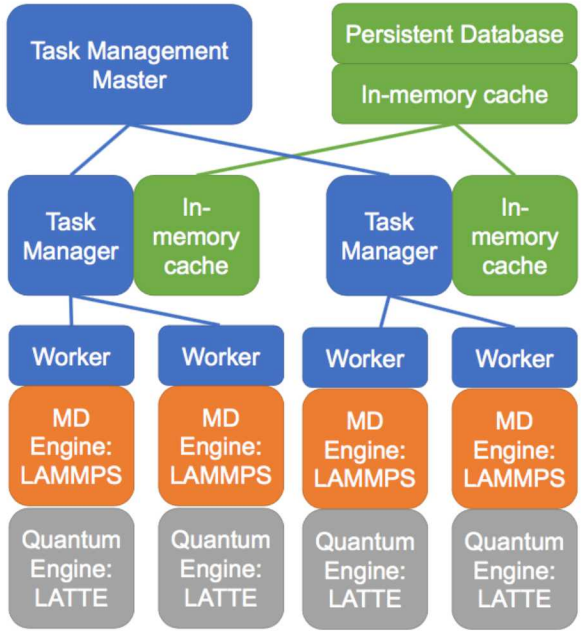


SNAP Active Learning

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.

ChemSNAP: Explicit Multi-Element SNAP

Partial Neighbor Density

$$\rho^{\beta}(\mathbf{r}) = w_{\alpha\beta}^{self}\delta(\mathbf{0}) + \sum_{\substack{r_j < R_{cut}^{\alpha\beta} \\ j \in \beta}} f_c(r_j; R_{cut}^{\alpha\beta}) w_{\beta}\delta(\mathbf{r}_j)$$

Partial Basis Function

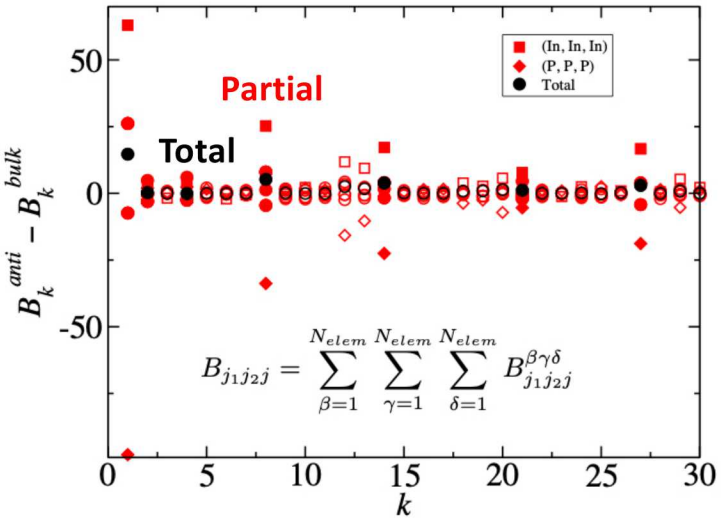
$$u_{jmm'}^{\beta} = w_{\alpha\beta}^{self} U_{jmm'}(0,0,0) + \sum_{\substack{r_j < R_{cut}^{\alpha\beta} \\ j \in \beta}} f_c(r_j; R_{cut}^{\alpha\beta}) w_{\beta} U_{jmm'}(\theta_0, \theta, \phi)$$

3-Element Partial Bispectrum Descriptor

$$B_{j_1 j_2 j}^{\beta\gamma\delta} = \frac{1}{2j+1} \sum_{m,m'=-j}^{j} \sum_{m_1,m_1'=-j_1}^{j_1} \sum_{m_2,m_2'=-j_2}^{j_2} (u_{jmm'}^{\beta})^* H_{j_1 m_1 m_1' \; j_2 m_2 m_2'}^{jmm'} \; u_{j_1 m_1 m_1'}^{\gamma} u_{j_2 m_2 m_2'}^{\delta}$$

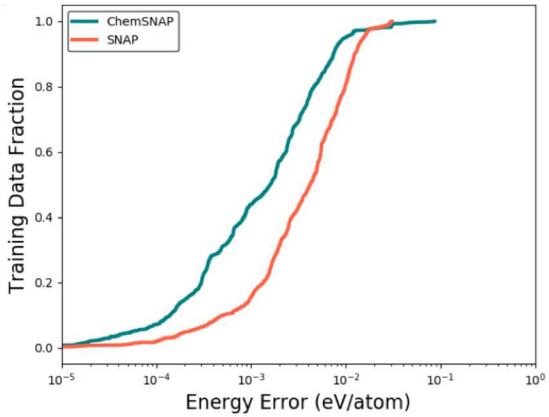
- Number of Partial Descriptors $\sim N_{elem}^3$
- Force cost $\sim N_{elem}^2$

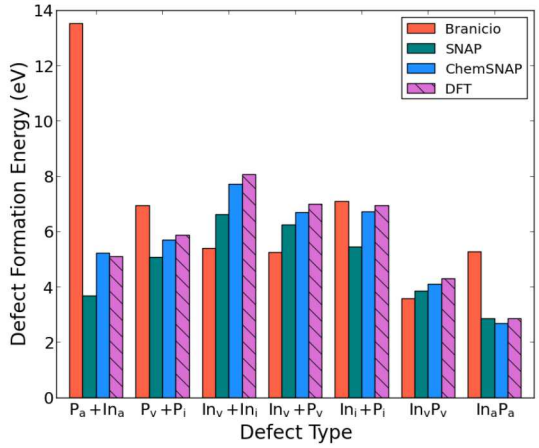
Sensitivity to Chemical Change
Antisite versus Bulk Zincblende



ChemSNAP: Indium Phosphide Relaxed Defects

- Compare SNAP relaxed defect formation energies to DFT
- Branicio potential shows very large defect formation energies for some structures
- Original SNAP formulation failed to reproduce defect formation energies within 1 eV difference from DFT
- ChemSNAP performs much better
  - Defect formation energies show much reduced error from DFT
  - Largest difference is ~0.2 eV
- ChemSNAP also represents standard properties accurately: Lattice constants, elastic constants, polymorphs





Conclusions

- ML interatomic potentials are driving a broad transition in the role of large-scale atomistic materials modeling from qualitative accuracy to quantitative accuracy
- Many challenges remain:
  - Robustness
  - On-the-fly accuracy estimate (hard, because no QM query on large-scale)
  - ML surrogate for QM (allows QM-like query on medium scale)
  - Active learning
  - Combining SNAP and ANNs
  - Descriptors (feature selection)
  - Many-element, chemically-active materials
- Long-term Goal: Integrated HPC workflow that iteratively generates a trusted ML potential for each materials modeling application

Ackowledgements

LAMMPS

- Steve Plimpton, Stan Moore
- Axel Kohlmeyer (Temple U.)

SNAP

- Mitch Wood, Julien Tranchida, Mary Alice Cusentino
- Danny Perez, Nick Lubbers (LANL)
- Rahul Gayatri (NERSC)