2020 Atomistic Simulations for Industrial Needs Workshop:

Program and Abstracts

Virtual event: August 5th to 7th, 2020

Contents

Note: Select titles were truncated to fit into the table of contents, see respective sections for the full titles

Day-1: Wednesday August-5th 2020:

Morning session: (9:00 am - 11:30 am EDT)

9:00 am: Introduction:

Carelyn Campbell

This workshop is designed to facilitate interactions between researchers in industry, government, and academia to identify and address obstacles to the wider use of atomic-scale simulation methods in quantitative industrial research and development. Topics for discussion include:

- The state of elemental and alloy interatomic potentials, including transferability.
- Current use cases and applications of interest analytics, machine learning, artificial intelligence, and uncertainty quantification.
- The use of atomistic and first-principles methods as inputs for phase-field, crystal plasticity and other higher-level models.
- Repositories of interatomic potentials, and workflows for comparing and validating potentials
- Increasing the availability and accessibility of community resources
- Interoperability of tools and information across projects Other topics of interest to the community

9:30 am: Machine Learned Interatomic Potentials

Gábor Csányi

I will make the somewhat bold claim that over the past 10 years, a new computational task has been defined and solved: this is the analytic fitting of the Born-Oppenheimer potential energy surface as a function of nuclear coordinates under the assumption of medium-range interactions, out to 5-10 Å. The resulting potentials are reactive, many-body, reach accuracies of a few meV/atom, with costs that are on the order of 1-10 ms/atom. This leaves the following challenges for ML potentials: treatment of long range interactions in a nontrivial way, e.g. consistency of treatment of open and periodic boundary conditions, environment dependent multipolar description, excited states (adiabatic surfaces), magnetism. We also still need a "shakedown" of the details among various approaches (neural networks, kernels, polynomials), and more standard protocols of putting together the training data. Tradeoffs between system- (or even project-) specific datasets and potentials vs. more general potentials will be ongoing. Further afield, another interesting question is what part of this technology can be reused to fit analytic surrogate models of *electronic* functions and functionals, such as reduced Hamiltonians, Green's Functions, density matrices etc, not to mention many body wave functions. There have been forays in this direction already.

10:00 am: Advancing Atomistic Capabilities for Heterogenous Systems

Difan Zhang, Yuxiang Wang, Alexandre F. Fonseca, and Susan B. Sinnott

Interactions between heterogeneous systems are challenging to model with high fidelity at the atomic scale across significant length scales. This presentation describes recent developments of third-generation charge optimized many-body (COMB3) potentials to enable metal-carbon interactions in classical molecular dynamics simulations. The potentials are applied to investigate the interaction of carbon materials, including metal-carbide-derived-carbons (CDCs) with titanium and aluminum, and the interactions of graphene and buckyballs with copper and aluminum surfaces. The results provide new insights into the bonding character at these interfaces and the role of defects, bond-angle, and temperature on interatomic interactions.

10:30 am: Achieving Machine-Learning-Potential Accuracy with Traditional Classical Potential Forms

Joshua A. Vita and Dallas R. Trinkle[a]

[a] University of Illinois, Urbana-Champaign

Classical interatomic potentials have long reflected the physics and chemistry of bonding in their functional forms, while the particular parameters for each element were the result of optimization to available data. This approach promises computational efficiency, transferability and interpretability, but is often criticized for insufficient accuracy compared to electronic structure calculations. The past decade has seen a revolution in the application of machine-learning to interatomic potentials, where large density-functional theory computed datasets are automatically constructed to optimize highly nonlinear, "black-box" functional forms. An increase in computational cost accompanies lost interpretability, but energy errors on the order of a few meV are now achievable. This accuracy is claimed to be a result of the highly nonlinear functional forms that can reproduce any function. Here, we show that the seemingly unprecedented accuracy compared with "traditional" classical potentials of machine-learning interatomic potentials is rather a result of large datasets combined with modern optimization algorithms. We use existing benchmark density-functional theory databases for Ni, Cu, Li, Mo, Si, and Ge, with the covariance matrix adaptation evolution strategy (CMA-ES) to optimize spline modified-embedded atom (MEAM) potentials. We compare both accuracy and computational efficiency of the spline MEAM potentials against five different machine-learning interatomic potentials,[1] and show that the traditional classical potential form offers comparable accuracy, and lie on the Pareto front for accuracy / efficiency optimization. These results suggest that interpretability and efficiency need not be sacrificed to achieve highly accurate interatomic potentials, and that similar results might be possible with other traditional classical potential forms.

[1] Yunxing Zuo, Chi Chen, Xiangguo Li, Zhi Deng, Yiming Chen, Jörg Behler, Gábor Csányi, Alexander V. Shapeev, Aidan P. Thompson, Mitchell A. Wood, and Shyue Ping Ong. "Performance and Cost Assessment of Machine Learning Interatomic Potentials." J. Phys. Chem. A 124, 731-745 (2020). https://doi.org/10.1021/acs.jpca.9b08723

11:00 am: Machine Learning for the Exploration of Energy Landscapes

Stephen Xie[a,b] , Shreyas Honrao[a,c] , and Richard G. Hennig[a,b]

[a] Department of Materials Science and Engineering, University of Florida, Gainesville, Florida
[b] Quantum Theory Project, University of Florida, Gainesville, Florida
[c] Department of Materials Science and Engineering, Cornell University, Ithaca, New York

Machine-learning can provide surrogate models that aid the search for new materials and physical relationships. We present our kernel approach to developing surrogate machine learning models for energy prediction. Using structurally and compositionally diverse materials generated with our genetic algorithm package GASP [1] and their formation energies from density functional theory, we train interatomic potentials using support vector regression. We show that radial and angular distribution functions can efficiently encode relevant physical information into machine-readable inputs and obey required constraints [2]. We demonstrate how augmenting the training data with local energies improves model performance [3]. These models can filter low-value candidates, reducing the genetic algorithm's computational cost by eliminating materials with a high probability of having higher energy.

[1] B. C. Revard, W. W. Tipton, and R. G. Hennig, GASP - Genetic algorithm for structure and phase prediction, https://github.com/henniggroup/gasp-python (2018), https://doi.org/10.5281/zenodo.2554076.

[2] S. Honrao, B. E. Anthonio, R. Ramanathan, J. J. Gabriel, and R. G. Hennig, Machine learning of ab-initio energy landscapes for crystal structure predictions, Comp. Mater. Sci. 158, 414 (2019), https://doi.org/10.1016/j.commatsci.2018.08.041.

[3] S. Honrao, S. Xie, and R. G. Hennig, Augmenting Machine Learning of Energy Landscapes with Local Structural Information, submitted (2020).

Break: (11:30 am - 1:00 pm EDT)

Afternoon session: (1:00 pm - 3:00 pm EDT)

1:00 pm: Machine-Learned Interatomic Potential Models for Practical Applications

Tim Mueller

Machine-learned interatomic potential models have the potential to revolutionize materials research by enabling accurate simulations at time and length scales that were previously inaccessible. I will present two examples of how machine learning can be used to develop interatomic potential models for practical use. In the first example, I will illustrate how moment tensor potentials, trained on the fly, enable identification of candidate coating materials for solid state lithium ion batteries by accelerating molecular dynamics calculations by about seven orders of magnitude. This approach results in calculated activation energies for lithium-ion diffusion that are in much better agreement with experimental values than those calculated using ab-initio molecular dynamics. In the second example, I will demonstrate how symbolic regression, in the form of genetic programming, can be used to identify accurate, transferable interatomic potential models with speeds comparable to the fastest available. The newly-discovered models are relatively simple, which makes them interpretable and limits the amount of data required to train them, enabling the use of highly accurate methods to generate training data at a reasonable computational cost. I will demonstrate this approach by presenting new interatomic potential models for copper and discuss the transferability of the newly-discovered functional forms to other elements.

1:30 pm: Development of General-Purpose Interatomic Potential with DFT Accuracy for Tantalum using Physically-Informed Artificial Neural Networks

Yi-Shen Lin, Ganga Purja Pun, Yuri Mishin

The physically-informed neural network (PINN) method allows development of interatomic potentials for general-purpose simulation with nearly-DFT accuracy by training a neural network on a large DFT database. This method equips interatomic potentials with superior transferability compared to the existing machine-learning potentials, thanks to the underlying analytical bond-order atomic interaction model which captures the physics of both metallic and covalent bonding. In this talk, we present a new PINN potential for tantalum, trained on a very diverse DFT database that encompasses thousands of atomic configurations. We demonstrate the power of the PINN method by showing its excellent transferability with nearly-DFT accuracy for many different environments. These include, but are not limited to, lattice vibrations, thermal expansion, a variety of crystal structures, deformation paths, point defects, surfaces, generalized stacking faults, dislocations, highly compressed states occurring in shock wave experiments, and the liquid phase. The PINN method provides a systematic workflow for the development of quality interatomic potentials that possess both high reliability and wide applicability, which are much needed for atomistic simulations in industry.

2:00 pm: Neural Network Reactive Force Field for C-H-N-O Systems

Pilsun Yoo, Michael Sakano, Saaketh Desai, Mahbubul Islam,

Peilin Liao$_a$, and Alejandro Strachan$_b$

[a] School of Materials Engineering and Birck Nanotechnology Center
[b] Purdue University, West Lafayette, Indiana, 47907 USA

Reactive force fields, parameterized for first principles simulations, have enabled an atomic level description of a wide range of phenomena, from chemistry at extreme conditions in reactive materials, to the operation of electrochemical devices and catalysis. While significant insight and semi-quantitative understanding have been drawn from such work, the accuracy of reactive force fields limits quantitative predictions. We developed a neural network reactive force field (NNRF) for CHNO systems to describe the decomposition and reaction of the high energy nitramine 1,3,5-Trinitroperhydro-1,3,5-triazine (RDX). NNRF was trained using energies and forces of a total of 3100 molecules (11941 geometries) and 15 condensed matter systems (32973 geometries) obtained from density functional theory calculations. The training set is generated via a semi-automated iterative procedure that enables refinement of the NNRF until a desired accuracy is attained. The RMS error of NNRF on a test ing set of configurations describing the reaction of RDX is one order of magnitude lower than current state of the art potentials.

2:30 pm: Atomistic Simulations as a Driver of Industrial Innovation

Ray Shan

From communication and information processing, to transport and medicine, to energy conversion and storage, materials provide the critical properties that govern applications and efficiency. Hence, the optimization and development of materials opens extraordinary opportunities. Many factors must be optimized: cost, long-term reliability, safety, sustainability and environmental impact. Atomistic simulation is increasingly employed as a component of the systematic development of optimal materials. Atomistic simulation provides the basis for systematic screening and detailed analysis of materials properties. This presentation describes the role of atomistic simulations in an industrial context and gives an insight into recent developments.

Virtual poster session: (3:00 pm - 3:30 pm EDT)

- **Synthesis, Thermal, and Rheological Evaluation of High C-Yield-Soluble Poly(phenylacetylene) Co-polymer Derivatives as New Ca**
  - Matt Agboola: Penn State University
- **Atomistic Simulations of Diffusion Induced Grain Boundary Migration in Aluminum**
  - Navjot Kaur: University of Manitoba
- **UV-EMF-Space Radiation Mitigation via Natural Compounds for All Living Cells**
  - Gaetano Lardieri: THCBD LL
- **Nano-structured Phoxonic Crystals for Midewave Infrared Sensing**
  - Anurag Sharma: Punjab Engineering College

Virtual poster session: (3:30 pm - 4:00 pm EDT)

- **Wetting Transition of a Nanodrop on Switchable Hydrophilic-Hydrophobic Surfaces**
  - Jyoti Roy Choudhuri: BMS Institute of Technology & Mgmt., Bangalore, Karnataka, India
- **Methane sorption in a family of qzd-MOFs: A multiscale computational study**
  - Mikhail Suetin: Karlsruhe Institute of Technology

- **Water Flow in Carbon Nanotubes: The Role of Tube Chirality**
    - Alan Sam: IIT Madras
- **SimPL: A Framework for Web-based Materials Design Platforms**
    - Kwang-Ryeol: Korea Institute of Science and Technology
- **Machine Learning Force Fields for Li-ion Cathodes: Validation and Outlook**
    - Joshua Gabriel: The University of Chicago, Argonne LLC

Day-2: Thursday August-6th 2020:

Morning session: (9:00 am - 11:30 am EDT)

9:00 am: The Interatomic Potentials Repository Project

Lucas Hale

The Interatomic Potentials Repository project has long served the atomistic simulation community by providing data and tools to assist researchers in discovering and using classical interatomic potentials. This presentation will highlight the resources developed by the project for discovering potentials, designing atomistic calculations, performing standard calculation tests, and running high-throughput workflows. Recent improvements have focused on making the data and tools more accessible. The underlying database for the repository is now open for exploration, and a Python API package has been created that interacts with the content hosted on the database. Updates have also been made with the calculation tools to simplify installation options and redesign methods to be more user friendly.

9:30 am: Molecular Simulations you can Trust and Reproduce: The OpenKIM Framework

Ellad B. Tadmor, Ryan S. Elliott, Daniel S. Karls, Yaser Afshar

Department of Aerospace Engineering and Mechanics University of Minnesota Minneapolis, MN 55455

The quality of classical molecular and multi-scale simulations hinges on the suitability of the employed interatomic model (IM) for a given application. Reproducibility of simulations depends on the ability of researchers to retrieve the original IM that was used. These two issues are addressed by the Open Knowledgebase of Interatomic Models project (link) . OpenKIM curates IMs with full provenance control, issues them DOIs so that they can be cited in publications, and tests them exhaustively using 'KIM Tests' that compute a host of material properties and 'Verification Check's on coding correctness. OpenKIM is integrated into major simulation packages (including ASE, DL_POLY, GULP and LAMMPS) allowing users to easily use OpenKIM IMs and query their predictions. Machine learning based tools for selecting an IM and assessing uncertainty are under development. OpenKIM functionality provides major benefits to researchers and promises to improve the reliability and reproducibility of molecular simulations of materials.

10:00 am: The Platform for the Interatomic Potentials Validation

Yury Lysogorskiy

Interatomic potentials (IP) are widely used in computational materials science, in particular for simulations that are too computationally expensive for density functional theory (DFT). A large number of IP is available for a wide range of chemical elements and their mixtures. Most IP have a limited application range and often there is very limited information available regarding their performance for specific simulations. We performed extensive tests for the majority of the available potentials from the OpenKIM and NIST repositories as well as from other sources. We present the online platform **atomistictools.org** for the validation of IP, where results of our tests could be

efficiently visualized and compared to the reference data (DFT and experiments, whenever available). Moreover, the platform would provide the possibility to access the tests and reference data, as well as the comparison protocols in order to run the tests locally on the user machine and perform the validation of the interatomic potentials under development.

10:30 am: Automated Ab-initio Determination of Materials Properties at Finite Temperatures with Pyiron

Jan Jannsen

A major challenge in predicting the properties of materials at realistic conditions is the accurate inclusion of finite temperature effects. Doing this on an ab-initio level often requires complex simulation protocols. These complex protocols, which often couple several specialized codes, make a quantitative description of error propagation and uncertainty quantification a critical issue. To handle this high level of complexity we have developed an integrated development environment (IDE) called pyiron[1] – http://pyiron.org. pyiron has been specifically designed to scale simulation protocols from the interactive prototyping level up to the high throughput level, all within the same software framework. To highlight the versatility and efficiency of this new framework we discuss two recent success stories: We demonstrate automated convergence for all key parameters in DFT codes and the calculation of melting points with a guaranteed precision of better 1K. These fully automated high-precision tools allow us to study trends over the periodic table in an efficient and systematic way. Examples of such high-throughput screenings will be given.

[1]: J. Janssen, et al., Comp. Mat. Sci. 161 (2019)

11:00 am: SEAMM: A productivity Environment for Computational Materials Science

Eliseo Marin-Rimoldi, Jessica Nash and Paul Saxe

Molecular Sciences Software Institute, Virginia Tech, Blacksburg, VA 24060, USA

The continuing progress in computational materials science (CMS) methods and software, coupled with the inexorable progress of computers, has been and will continue to drive the rapid progress in the field. However, it is increasingly difficult to bring the wide range of tools together in a single environment and have them work together to handle more sophisticated and complicated workflows -- and to run on the wide range of computers available.

There is an identical problem in the computational molecular science (CMS) community, and indeed there is considerable overlap between the two communities. The NSF established the Molecular Sciences Software Institute (MolSSI) to address these and other software issues in the community. One of the projects at the MolSSI is SEAMM, which is an open-source environment for computational molecular and materials science. The goals for SEAMM are as follows:

- Reduce the barrier for users to use the wide range of open source software available.
- Provide reproducible, shareable workflows.
- Encourage interoperability between codes.
- Encourage both software and computational best practices.

- Reduce the churn and rewriting of codes, scripts, etc.

This talk will cover how SEAMM is approaching these goals, the current state of the project, and future plans. The next part of the talk will focus on how using SEAMM can help users speed up their research while making it more reproducible and accessible. The last part of the talk will cover SEAMM from the point of view of software developers, covering the infrastructure and tools provided. These tools take care of many of the peripheral tasks that are necessary for using a code, but are not the principal focus of most development efforts -- the "necessary evils" of providing good user input, of translating data from different formats, presenting results in different file formats as well as graphically, connecting with other programs, and even running calculations on different computing resources and getting files back and forth.

Break: (11:30 am - 1:00 pm EDT)

Afternoon session: (1:00 pm - 2:30 pm EDT)

1:00 pm: Modeling Hydrogen-Oxygen Combustion and Hydrocarbon-in-Zeolite Systems via Neural Network Enhanced Programmable Potentials

Allan Avila

Although quantum scale simulations of chemical reactions offer an accurate description of the process, a multi-atom quantum simulation is unfeasible as it would not terminate in a scientist's lifetime. Multi-atom simulations are feasible at the molecular scale, however, the potential bond energies are inaccurate, and results often fail to match quantum simulations. We demonstrate the ability to utilize the programmable potentials methodology and a second-generation neural network enhanced programmable potentials methodology to develop quantum-accurate molecular-level potentials for several intermediate reactions involved in hydrogen-oxygen combustion and to describe the adsorption energies for various hydrocarbon-in-zeolite systems. We utilize sparse Electronic Structure Theory (EST) simulation data to train our programmable potentials and the developed potentials are then inputted into the molecular dynamics simulation package LAMMPS for verification. Our results demonstrate that the developed programmable potentials generalize beyond the EST training data set, can encode reactive reactions and can be combined with neural networks for improved accuracy.

1:30 pm: Atomistic and Coarse-Grained Polymer Modeling: Applications at the Army Research Laboratory and Beyond

Michael Salerno

Many fundamental and applied research projects at ARL rely on all-atom molecular dynamics (MD) simulations and also need to extend or couple to adjacent length and timescales. I will discuss two relevant modeling problems that link MD to smaller and larger length scales: First, ab-initio results on molecular photo-excitation are used to model photo-responsive molecules in glassy materials. Second, understanding the shear flow and rheology of polymers like high-molecular-weight polyethylene requires simulations orders of magnitude beyond the lengths and times of atomistic MD. One pitfall to mapping CG to all-atom models involves common CG model approximations that distort the intrinsic stiffness of a polymer, which I will argue is a fundamental polymer property.

2:00 pm: Compressibility of Nanoconfined Fluids:

Relating Atomistic Modeling to Ultrasonic Experiments

Gennady Gor

When fluids are confined to nanopores, many of the fluid's physical properties are altered. In this talk I will show that one such property is compressibility. The compressibility determines a material's response to mechanical loads and elastic wave propagation. I will demonstrate that the compressibility of a nanoconfined fluid can be readily calculated based on Monte Carlo or molecular dynamics simulations, and used for predictions of wave propagation in fluid-saturated nanoporous media. These results are important for exploratory geophysics for the analysis of wave-propagation in fluid-saturated nanoporous media, such as coal or shale.

Day-3: Friday August-7th 2020:

Hands-On session-1: (9:00 am - 12:30 pm EDT)

9:00 am: A "How-To" Tutorial on Using OpenKIM with LAMMPS

Ryan S. Elliott and Ellad B. Tadmor

The Open Knowledgebase of Interatomic Models (OpenKIM) contains, in part, a suite of curated interatomic potentials that are freely available (at https://openkim.org) for use in interatomic simulations by materials researchers and scientists across the world. OpenKIM "Models" are compliant with the KIM Application Programming Interface (KIM API) and may be used with any Simulator that also supports the KIM API, including LAMMPS, ASE and ASAP, DL_POLY, GULP and others. OpenKIM "Simulator Models" are curated interatomic potentials that do not conform to the KIM API and can only be used with specific simulators, such as LAMMPS.

OpenKIM is also a Knowledgebase containing thousands of predictions for important materials properties computed using each of the interatomic potentials (Models and Simulator Models) stored in OpenKIM. These predictions may be browsed and visualized at https://openkim.org. An exciting new feature of OpenKIM is "KIM Input File Queries" which allow a user to embed searches for OpenKIM data directly into their simulator input files. This makes it easy to accomplish tasks like obtaining the equilibrium lattice constant associated with the potential used for a particular simulation.

This tutorial will provide a practical demonstration of how a user may obtain OpenKIM Models and Simulator Models (in binary or source code form) on their computing system and use these with their preferred simulator. In particular, the use of OpenKIM Models, Simulator Models, and KIM Input File Queries within LAMMPS will be a focus of the discussion.

10:00 am: Atomman/iprPy

Lucas Hale

This demo will provide some quick examples of using atomman to generate defect structures and performing iprPy calculations.

11:00 am: Pyiron

Jan Jannsen

Pyiron is an integrated development environment (IDE) for computational materials science.

12:00 pm: Atomistictools.org

Yury Lysogorskiy

Demonstration of working with a website prototype atomistictools.org for interatomic potentials comparison and selection.

Break: (12:30 pm - 1:00 pm EDT)

Hands-On session-2: (1:00 pm - 4:00 pm EDT)

1:00 pm: NanoHUB: online simulation and data

Alejandro Strachan

Access to powerful simulation tools and hardware together with the ability to use machine learning tools to extract information from the large data sets is revolutionizing materials, nanotechnology, and related fields. Unfortunately, utilizing such tools requires significant expertise and has traditionally been restricted to computational experts, furthermore, the data generated has not been accessible or findable except to its creator. To address this gap, the NSF's nanoHUB turns powerful research-grade simulation codes into easy-to-use Apps, accessible via a web-browser, designed for end-users: researchers, instructors, and students. Today, over 600 community-contributed simulation & data tools and apps, each indexed by Web of Science and Google Scholar, serve 16,000+ users and 1,000,000 runs every year. Developers and end users benefit from HPC resources, a scientific software development environment, automatic uncertainty quantification, and support for Jupyter notebooks. This hands-on tutorial will introduce participants to the use of nanoHUB Apps and tools in research, education, and workforce development. In addition, we will describe resources for developers including the self-serve publication process.

2:00 pm: SEAMM

Paul Saxe

This demo/hands-on session will cover how to use SEAMM — to set up a flowchart and run it either by hand or via the Dashboard — and very briefly cover how to create plug-in. SEAMM is still in a preliminary state, and will not be ready for production use for a month or two, but this will give you an opportunity to understand how it works and also to make comments and suggestions for future directions. You may either view this as a demo, or you can install SEAMM on a Mac/Linux machine and follow along. Instructions for installing SEAMM will be sent prior to the session. While SEAMM works on Windows, some executables it uses are less readily available on Windows and may require a more manual installation.

3:00 pm: PYFIT-FF

James Hickman, A. Robinson, F. Tavazza, Y. Mishin

PYFIT-FF is a neural network interatomic potential training code written in Python. It relies heavily on the open source library PyTorch to optimize the neural network's fitting parameters and preform efficient linear algebra operations. This presentation will provide an introduction to PYFIT-FF and will be broken into three sections. The first section contains elementary background information relevant to understanding the code. This includes a brief overview of the following topics; regression and numerical optimization, neural networks, automatic differentiation, machine learning interatomic potentials, and the PyTorch library. The second section provides an overview of the PYFIT-FF code itself. Specifically, we cover how to download the code, the format of it's input and output files, the meaning of the most important input parameters, and a general outline of what the code is doing internally. In the third section there will be a demonstration showing how to use the code to train a neural network potential using a small silicon DFT training set and visualize the results.

Organizing committee:

- Lucas Hale, NIST, Material Measurement Laboratory
- James Hickman, NIST, Material Measurement Laboratory
- Ryan Elliott, University of Minnesota and OpenKIM project, Aerospace Engineering Mechanics
- Daniel Karls, University of Minnesota and OpenKIM project
- Ryan Thomas, Honeywell