

Reference ID: 11226450791_Raj

Reference ID: 11226450791_Raj

Submission Date and Time: 12/16/2019 1:02:17 PM

This contribution was submitted to the National Science Foundation in response to a Request for Information, <https://www.nsf.gov/pubs/2020/nsf20015/nsf20015.jsp>. Consideration of this contribution in NSF's planning process and any NSF-provided public accessibility of this document does not constitute approval of the content by NSF or the US Government. The opinions and views expressed herein are those of the author(s) and do not necessarily reflect those of the NSF or the US Government. The content of this submission is protected by the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>).

Consent Statement: "I hereby agree to give the National Science Foundation (NSF) the right to use this information for the purposes stated above and to display it on a publicly available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>)."

Consent answer: I consent to NSF's use and display of the submitted information.

Author Names & Affiliations

Submitting author: Rishi Raj - University of Colorado Boulder

Additional authors: Rajiv Kalia, University of Southern California

Contact Email Address (for NSF use only): (hidden)

Research domain(s), discipline(s)/sub-discipline(s)

Computer Science, Materials Science, Defect Physics, Mechanical Engineering, Aerospace Science and Engineering

Title of Response

Materials Far from Equilibrium: from Structure and Properties to Scientific Algorithms to Broader Impact

Abstract

Conventional research in materials science is built upon phase diagrams. Recently, however, a new class of "ad-hoc" materials are emerging. They are made with multiple constituents and by processes that deploy electrical, magnetic, electromagnetic and supra-mechanical forces. Often, these new materials

lie outside the traditional realms of phase-diagrams. Energetic processing can induce high molar concentrations of defects, which now themselves become species thereby further increasing the complexity in the structure of these “energetic” materials. They manifest new spectra and patterns in X-ray diffraction and X-ray scattering, TEM, FTIR, UV-absorption, Raman spectroscopy, XPS, neutron scattering and EXAFS. These spectra do not fit known patterns; thus, we are unable to tell what may be the atomistic, the defect and the electronic structure of these novel materials. We will deploy machine learning approaches to go through the millions of iterations to identify the structure that best matches the details of the spectroscopic data. The discovery of high entropy, lithium containing oxides with gigantic ionic conductivity are just one example of the technological impact of the proposed approach. The synergy between machine-learning, the discovery, the science and the design of these complex materials will have a persistent “broader-impact” for the next several decades.

Question 1 (maximum 400 words) – Data-Intensive Research Question(s) and Challenge(s). Describe current or emerging data-intensive/data-driven S&E research challenge(s), providing context in terms of recent research activities and standing questions in the field. NSF is particularly interested in cross-disciplinary challenges that will drive requirements for cross-disciplinary and disciplinary-agnostic data-related CI.

Machine learning methods will provide ground-breaking results in understanding the fundamental atomistic, and electronic structure of new materials that are far from equilibrium. Electron microscopy and x-ray scattering experiments produce huge amounts of data. For example, the “4D Camera” (for Dynamic Diffraction Direct Detector) at Lawrence Berkeley National Laboratory can capture images with atomic scale resolution in microseconds. Processing of such large datasets is an enormous challenge that only AI-assisted tools can address. Recent advances in supervised and unsupervised deep learning show tremendous promise in analyzing these vast troves of electron microscopy and X-ray data. Deep learning and other AI tools can identify microstructural changes and how they affect material properties. These tools can also denoise data, impute missing data, identify relevant structural features, classify them, and provide statistical analysis. Furthermore, novel neural networks such as Equation Learners can extract physical laws from experimental data. Generative models and neural network solvers for partial differential equations can address inverse problems, e.g. prediction of structures including defect distributions from measurements of material properties. The emerging field of “flash sintering” serves as an example of the impact of the proposed coupling of experiments with machine learning for accelerated development of new materials that are far from equilibrium. Discovered less than a decade ago, when it was shown that modest electrical fields can sinter ceramics in mere seconds at low temperatures, this method has expanded into processing of complex materials, like high entropy oxides quickly, accelerating the processing of complex materials by orders of magnitude. What took weeks if not months to prepare and characterize new composition of complex materials, can now be completed many times over within one day. The bottle neck in the scientific and technological developments in this field is computational methods that can be empowered to quickly analyze the structure of these new materials, and predict their properties. Machine learning is indispensable in this pursuit.

Question 2 (maximum 600 words) – Data-Oriented CI Needed to Address the Research Question(s) and Challenge(s). Considering the end-to-end scientific data-to-discovery (workflow) challenges, describe any limitations or absence of existing data-related CI capabilities and services, and/or specific technical and capacity advancements needed in data-related and other CI (e.g., advanced computing, data services, software infrastructure, applications, networking, cybersecurity) that must be addressed to accomplish the research question(s) and challenge(s) identified in Question 1. If possible, please also consider the required end-to-end structural, functional and performance characteristics for such CI services and capabilities. For instance, how can they respond to high levels of data heterogeneity, data integration and interoperability? To what degree can/should they be cross-disciplinary and domain-agnostic? What is required to promote ease of data discovery, publishing and access and delivery?

Discovery of novel materials require integration of experimental methods, modeling and simulation, data science (DS) and machine learning (ML) tools. Faster and targeted discovery of materials hinges on automated methods to analyze raw experimental and simulation data that provide actionable information for rapid screening of material properties and processes. DS and ML will play a crucial role by capturing complex energy landscapes of functional materials and discovering transformation pathways from both simulation and experimental data. A major bottleneck in materials modeling, simulations and experiments is that datasets are so huge that they cannot be analyzed on-the-fly or transferred to supercomputers fast enough for AI-assisted modeling and analysis. As a result, most of the data is archived and not analyzed. Very likely, important discoveries will be missed because the current CI is hopelessly inadequate to tackle these problems. AI tools can couple petascale reactive and quantum dynamics simulations that generate unprecedented amounts of data for atomic trajectories, forces, wave functions, and electronic charge distributions. The datasets are prohibitively large for storage and analysis by manual, or even by semi-automated methodologies. Furthermore, upcoming exascale architectures are expected to be heavily optimized for atomic matrix-matrix operations, and therefore traditional MD force calculation requiring irregular memory access must be modified to scale on these new architectures. AI tools are needed for the development of experimentally-informed large-scale simulations. Going forward, the critical CI needs for material discovery are: (i) Highly efficient computing platforms that can be attached directly to instruments for rapid screening. (ii) Visualization platforms integrated with on-the-fly screening of data. (iii) Continuous growth in high-performance computing to address bottlenecks in first-principles calculations. (iv) Infrastructure for rapid data transfer from instruments to supercomputers. And, (v) Post exascale machine learning tools to analyze and mathematically model experimental data to discover physical laws hidden in the data. This proposal embarks on an unprecedented synergy between materials science, physics, chemistry, engineering and computer science, with the intensely focused objective of building the scientific foundation and the tools for next generation materials which are complex and made by energetic processes, and that are far from equilibrium, thereby being endowed with novel properties not present in conventional materials.

Question 3 (maximum 300 words) – Other considerations. Please discuss any other relevant aspects, such as organization, processes, learning and workforce development, access and sustainability, that need to be addressed; or any other issues more generally that NSF should consider.

This proposal is led by Professors Kalia of the Department of Computer Science at USC and Raj of Materials Science/Mechanical Engineering at the University of Colorado. This collaboration has been brewing over the last year starting from an ECI Conference in Tomar, Portugal in March 2019. Our understanding of the fundamental issues from materials science and computer science viewpoints is deep and unique given the “firewall” that often exists between these two communities. At USC, the AI Division of the Viterbi School of Engineering has close to 180 researchers which includes a large number of graduate students, underpinned by professional staff. Separately, the Department of Computer Science also has an AI group. The Kalia group at USC investigates the applications of machine learning to experiments and simulations in materials science. They have used deep learning tools, such as the restricted Boltzmann machine, to model synthesis of layered materials on a quantum computer. They have also used variational autoencoder to study crack propagation and phase transformations in layered materials. They have applied weighted constraint satisfaction model to impute missing data in electron microscopy experiments. Raj is the inventor of “flash sintering” which has captured the imagination of the materials science community. This method is now being deployed to process new materials of complex compositions in mere seconds (in comparison to several weeks needed by conventional processing). Thus, flash can radically accelerate the discovery of new materials; it can also serve as an experimental leg of the Genome initiative which remains largely a theoretical pursuit. Multicomponent materials made in this way are already showing unusual features in X-ray measurements, and properties, for example gargantuan conductivities in solid state electrolytes for lithium ion batteries.

-- End Submission --