

Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-04 22:56:03

PAGE 1

REFERENCE NO: 219

This contribution was submitted to the National Science Foundation as part of the NSF CI 2030 planning activity through an NSF Request for Information, https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf17031. 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>).

Author Names & Affiliations

- Harley Johnson - University of Illinois at Urbana-Champaign
- Dallas Trinkle - University of Illinois at Urbana-Champaign

Contact Email Address (for NSF use only)

(Hidden)

Research Domain, discipline, and sub-discipline

Engineering, computational materials

Title of Submission

Cyberinfrastructure needs for computational materials

Abstract (maximum ~200 words).

Computational materials research is a key driver of engineering design, and it is poised for tremendous advances in coming years with the right investments in methods development, data science, computational power, and collaborative infrastructure. Development of methodology for bridging length and time scales will drive much of the emerging research in computational materials modeling. Data science and machine learning will enable new developments in computational materials methodology. Simultaneously, computational materials research will be advanced through new computational capabilities, which will include development of heterogeneous medium-scale parallel computing platforms and cloud computing infrastructure. Finally, development of collaborative mechanisms and opportunities for exchange of ideas, software, and expertise will be needed to fully engage the computational materials research community.

Question 1 Research Challenge(s) (maximum ~1200 words): Describe current or emerging science or engineering research challenge(s), providing context in terms of recent research activities and standing questions in the field.

So much of engineering design requires understanding of materials and mechanics. Current and emerging research in computational materials is driven largely by a single persistent problem: that of predictive modeling over multiple scales, both in length and time. Eight or more orders of magnitude separate atomistic structure and human length scales, and relevant time scales span some 15 or more orders of magnitude. These disparate scales are well known, but remain a problem in designing materials and understanding mechanics from first principles.

Scale-bridging efforts are necessary given today's computational limitations, and will continue to be necessary even with increasingly powerful computational resources in coming years; society simply needs to tackle ever-larger computational problems. Bridging length and

Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-04 22:56:03

PAGE 2

REFERENCE NO: 219

time scales generally requires robust surrogate or intermediate scale-bridging methods. These methods must be built upon trusted physics and chemistry models, and will rely on simultaneous advances in computer science. In addition, these methods must be validated to have any predictive capability, which requires that uncertainty quantification be built in from the beginning.

The science of scale-bridging in computational materials stands to evolve dramatically over the next few decades as data science and machine learning bring new tools to the table. Machine learning brings the possibility of entirely new types of empirical models, though some may obscure the underlying physical processes. To be truly predictive, future intellectual efforts in this arena must involve careful collaborations between both computational and data scientists, along with materials scientists, physicists, mechanicians, and chemists, to truly harness the enhanced modeling capabilities.

What makes scale-bridging most exciting right now is the prospect of computational power that will make new modeling approaches possible. Everything from Bayesian sampling approaches to Pareto optimization to Gaussian process modeling leverage the ability to sample over very large parameter sets to explore the "phase space" of models. The sampling of model phase space permits the selection of optimal models, but it is critical that we better understand the limitations of models, quantify uncertainties, and even develop guidance on how to provide additional data to improve the models.

Question 2 Cyberinfrastructure Needed to Address the Research Challenge(s) (maximum ~1200 words): Describe any limitations or absence of existing cyberinfrastructure, and/or specific technical advancements in cyberinfrastructure (e.g. advanced computing, data infrastructure, software infrastructure, applications, networking, cybersecurity), that must be addressed to accomplish the identified research challenge(s).

We expect the coming physical cyberinfrastructure needs in computational materials to be based on large, heterogeneous processor networks, but with what we may call "medium-scale" parallel processing. For example, one envisions many sets of 100-1000 cores that are very efficiently networked, operating in worker-manager systems, rather than homogeneous arrangements of 10^6 - 10^7 cores operating simultaneously. This expectation is based on (i) the types of research codes that are emerging, in which one or a few coupled physical descriptions of materials may be used on small or medium sized systems, and (ii) the many important analyses that require repeated realizations of physical configurations or processes over relatively small systems, running under different conditions or under the same conditions but where significant statistics are required.

A second trend that may emerge as critical in advancing computational materials research is the reliance on cloud computing platforms. As compute cycles become cheaper, and networking becomes faster, there is expected to be a move away from locally controlled parallel clusters for computational materials research. This may involve a combination of commercial operations (such as those now dominating software and storage aspects of computational science) along with academic operations, as research institutions themselves evolve toward less-centralized computational physical infrastructure. While petascale and exascale platforms will continue to serve many computational materials researchers, researchers will begin turning to cloud-based systems. This will necessitate attention to the software and tools used in research-level computational materials studies, in order to successfully compile and run research codes across extremely heterogeneous and delocalized cloud computing systems.

Question 3 Other considerations (maximum ~1200 words, optional): Any other relevant aspects, such as organization, process, learning and workforce development, access, and sustainability, that need to be addressed; or any other issues that NSF should consider.

We believe that "other considerations" may be the most important aspects to consider in developing cyberinfrastructure for computational materials. In our view, these aspects center on human issues, and not software or physical hardware development. Neither, however, is this an issue of workforce development or education. Instead, we see a need to promote collaboration, cross-pollination of ideas, training, and real sharing of expertise. While we now routinely build software repositories, online systems for collaborating on code development, and even training modules, etc., it is not clear that we are sufficiently investing in the side-by-side aspects of collaboration in computational science. This will be more important than ever as we attempt to push research codes built on increasingly heterogeneous platforms to computational infrastructures that themselves are increasingly heterogeneous. There is no direct technological fix for this. But future

Submission in Response to NSF CI 2030 Request for Information

DATE AND TIME: 2017-04-04 22:56:03

REFERENCE NO: 219

PAGE 3

investments in cyberinfrastructure should account for this missing element in computational materials research.

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 publically available website, consistent with the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>).”
-