Reference ID: 11215302351_Fraden

Submission Date and Time: 12/11/2019 4:53:33 PM

Author Names & Affiliations

Submitting author: Seth Fraden - Brandeis University

Additional authors: None

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

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

Soft matter

Title of Response

Data-driven needs of the Brandeis MRSEC and DMREF

Abstract

Data-driven methods for characterization, control and uncovering of underlying physics have become central components of soft materials research. These methods have proven to be effective for classifying complex, non-equilibrium spatiotemporal patterns and in comparing theory and experiment.

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.

The Brandeis MRSEC seeks to engineer active matter to create autonomous materials that perform work. Data-driven methods, including deep recurrent neural networks (DRNN) and spatiotemporal Koopman operators will be employed to characterize and control dynamics. Data-driven methods will also be used to uncover the underlying physics using methods such as Sparse Identification of Nonlinear Dynamics (SINDy).

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?

Needs at Brandeis for data-driven science and engineering. (1) Increased capacity of GPU clusters. Brandeis received a MRI for a GPU cluster in 2019. The upgraded cluster will have a total of 2608 CPU cores with a theoretical max performance of 101 Teraflops (double precision), 3444 Teraflops (single precision) with GPUs, and the storage node will have 340TB capacity. We anticipate needing to triple the size of the current cluster in the next 5 years to meet the increased demand as we begin students of active matter in 3D. (2) Students need instruction in data-driven science and engineering. We will add courses in our new applied math department and we will apply for a training grant to support this. (3) We will encourage the academic administration to hire data scientists in multiple disciplines, ranging from economics and sociology to math, computer science, and the physical and life sciences.

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.

Students need instruction in data-driven science and engineering. We will add courses in our new applied math department and we will apply for a training grant to support graduate study in data science.

-- End Submission –