Reference ID: 11227080610_Manz

Submission Date and Time: 12/16/2019 4:58:38 PM

Author Names & Affiliations

Submitting author: Thomas Manz - New Mexico State University

Additional authors: None

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

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

Chemical Engineering; computational chemistry; materials simulation and design; scientific computing; condensed matter and materials theory

Title of Response

Dedicated Computational Units to Compute Matrix Permanents

Abstract

Non-deterministic polynomial time (NP) problems are a class of problems that are hard to solve but easy to check. The sharp-P (#P) complexity class counts the number of accepting paths of a polynomial-time non-deterministic Turing machine. Computing the matrix permanent is a #P-complete problem, meaning

it can be mapped to every problem in the NP and #P complexity classes in polynomial time. Computing the matrix permanent efficiently would revolutionalize all computational science by providing an effective method to solve a wide range of heretofore insolvable NP and #P problems across all academic disciplines. A host of emerging technologies (i.e., artificial intelligence, autonomous vehicles, etc.) critically depend on solving this problem. Therefore, I argue the single most important unmet need in cyberinfrastructure is to develop and install dedicated matrix permanent computing units. These units would transform data mining, scientific computing, machine learning, and artificial intelligence. They could be developed and built during the next five years, thereby making computing in 2025 dramatically different today.

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.

Dedicated units to compute the matrix permanent would solve the single largest problem in computing today: solving computationally complex problems that have an almost unimaginably large number of combinations. These problems occur in all academic disciplines. Within the past year, my research group developed an algorithm to sample the matrix permanent in polynomial time. This algorithm statistically samples values whose expectation value exactly equals the matrix permanent. The sample average and standard deviation of the sample average provide a computationally experimentally measured value of the matrix permanent. This algorithm operates in polynomial time, making it suitable for computation. There is now a need to develop specialized and dedicated computing units to maximize the speed of executing this algorithm. This would be analogous the specialized units used to mine bitcoins or to the graphics processing units on computers. This would enable a large host of important data-science and computational-science problems to be solved that are presently unsolvable with today's computers.

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?

These dedicated computational units to compute matrix permanents should be initially deployed on the NSF-funded eXtreme Science and Engineering Discovery Environment (XSEDE). They could be mounted on compute nodes in a manner analogous to graphics processing units at existing XSEDE-funded supercomputing centers. XSEDE users could apply for access to these units in a manner analogous to how they currently request XSEDE allocations to GPU-enabled nodes. This will provide a sustainable mechanism for opening access to these units to the large number of academic researchers across the United States. This will immediately place the United States in the leadership position of computational solutions to the NP and #P problems, which will in turn make the United States the foremost leader in related computational technologies. US academics will pioneer the use of these units, which will then be widely adopted by US industries and US government organizations. Learning and workforce development could be accomplished through XSEDE and NSF coordinated training programs and grants. XSEDE could coordinate the allocation of computing time on these units and train users on the technical aspects of job submission to them. NSF could fund academic training grants to universities to train a workforce on the science related to them.

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.

-- End Submission –