More Data, More Science and ... Moore’s Law?

Lecture: Friday, February 28, 2014
Time: 11:15 am-12:30 pm
Location: 2150 Torgersen Hall
Open to the general public.

Speaker: Dr. Katherine Yelick
Associate Laboratory Director for Computing Sciences at Lawrence Berkeley National Laboratory
Professor of Electrical Engineering and Computer Sciences at University of California Berkeley

Abstract:
In the same way that the Internet has combined with web content and search engines to revolutionize every aspect of our lives, the scientific process is poised to undergo a radical transformation based on the ability to access, analyze, and merge large, complex data sets. Scientists will be able to combine their own data with that of other scientists, validating models, interpreting experiments, re-using and re-analyzing data, and making use of sophisticated mathematical analyses and simulations to drive the discovery of relationships across data sets. This “scientific web” will yield higher quality science, more insights per experiment, an increased democratization of science, and a higher impact from major investments in scientific instruments.

What does this “big science data” view of the world have to do with HPC? The terms “high performance computing” and “computational science” have become nearly synonymous with modeling and simulation, and yet computing is as important to the analysis of experimental data as it is to the evaluation of theoretical models. Due to the exponential growth rates in detectors, sequencers and other observational technology, data sets across many science disciplines are outstripping the storage, computing, and algorithmic techniques available to individual scientists. Along with simulation, experimental analytics problems will drive the need for increased computing performance, although the types of computing systems and software configurations may be quite different.

In this talk I will describe some of the opportunities and challenges in extreme data science and its relationship to high performance modeling and simulation. One of those challenges (my own favorite) is the development of high performance, high productivity programming models. In both simulation and analytics, programming models are the “sandwich topic,” squeezed between application needs and hardware disruptions, yet often treated with some suspicion, if not outright disdain. But programming model research is, or at least should be, an exemplar of interdisciplinary science, requiring a deep understanding of applications, algorithms, and computer architecture in order to map the former to the latter. I will use this thread to talk about my own research interests, how I selected various research topics over the years, and the importance of teams and even complete communities of researchers when addressing one of these problems.

Bio:
Katherine Yelick is a Professor of Electrical Engineering and Computer Sciences at the University of California at Berkeley and the Associate Laboratory Director for Computing Sciences at Lawrence Berkeley National Laboratory. She is known for her research in parallel languages, compilers, algorithms, libraries, architecture, and runtime systems. She earned her Ph.D. in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology and has been on the faculty at UC Berkeley since 1991 with a joint research appointment at Berkeley Lab since 1996. She was the director of the National Energy Research Scientific Computing Center (NERSC) from 2008 to 2012 and in her current role as Associate Laboratory Director she manages a 300-person organization that includes NERSC, the Energy Science Network (ESNet), and the Computational Research Division. She recently received the ACM-W Athena award and is a member of the California Council on Science and Technology and the National Academies Computer Science and Telecommunications Board.