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

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, utilizing 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.

Yelick will discuss some of the opportunities and challenges in extreme data science and its relationship to high performance modeling and simulation. Including her research in the development of high performance, high productivity programming models. Her current research is largely focused on the problem of avoiding and minimizing the cost of communication, as data movement both to memory systems and between processors a major barrier to scalability and energy efficiency. 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. She will use this thread to talk about her research interests, how she selected various research topics, and what she sees as current open problems in parallel programming model design and implementation.