Testimony before the U.S. – China Economic and Security Review Commission

Date:
Thursday, May 10, 2012

Name of the panelist:
Horst D. Simon

Panelist's Title and Organization:
Deputy Laboratory Director, Lawrence Berkeley National Laboratory, Berkeley, CA 94720

Title of Hearing:
Assessing China's Efforts to Become an Innovation Society—A Progress Report.
Panel IV: Information and Communications Technology Innovation.

Outline:

1. My Background
2. General comments about the role of supercomputing and high performance computing
3. Assessment of supercomputing developments in China
4. Exascale computing and future developments
5. Summary and Conclusions

Many thanks to the Commission for the opportunity to testify at this important hearing. I am honored to have been asked and am happy to share my perspectives and insights, and hope that they are helpful to your deliberations.

1. My Background

I am the Deputy Laboratory Director of Lawrence Berkeley National Laboratory (Berkeley Lab or LBNL), a large (about 3400 FTE staff and $750M annual budget) Department of Energy multiprogram laboratory, managed by the University of California, located in Berkeley, California. LBNL's mission is to conduct research in the physical sciences, biosciences, and computing sciences in order to address some of the most challenging problems that the nation and the world face in the areas of energy and environment.

Previously, from 1996 to 2007, I was Director of the National Energy Research Scientific Computing Center (NERSC) and from 2003 to 2010 Associate Laboratory Director for Computing Sciences at LBNL. NERSC is one the largest supercomputer centers in world, supporting a broad base of scientific applications of relevance to the DOE mission. I have been an active researcher in supercomputing algorithms, performance evaluation, and applications throughout my career that has also included positions at Boeing, NASA Ames, and SGI. I am a member of the TOP500 team that publishes the bi-annual TOP500 list of the most powerful supercomputers in the world. This list has become a valuable tool to asses technology developments, and geographical and business trends, in high performance computing and supercomputing.

The current testimony is based on my 30 years of experience in the field having observed supercomputing almost since its beginning in the 1970s. I would like to thank my colleagues Jack Dongarra, Hans Meuer, and Erich Strohmaier from the TOP500 list, David Kahaner from the Asian Technology Information Program (ATIP), and colleagues from ICCS (International Center for Computational Science, Beijing, Berkeley, Heidelberg) for discussions and contributions. My testimony also includes some anecdotal information from a recent trip by Hemant Shukla (LBNL) and myself to Beijing and Tianjin, including visits to National Astronomy Observatory of China (Chinese Academy of Science), Electrical Engineering Department at Peking University and Kavli Institute for Astronomy & Astrophysics, National Supercomputing Center, Tianjin (computer named Tianhe 1A – ranked no. 2 in the world), and Institute of Process Engineering (computer named Mole 8.5).

2. General comments about the role of supercomputing and high performance computing

The term High Performance Computing (HPC) generally refers to all computing infrastructure and activities that contribute to the computational solution of difficult scientific and engineering problems.

HPC encompasses a very wide range of technologies and activities ranging from desktops to supercomputers.

Supercomputers are computing systems that provide close to the best possible computational performance at any given time. Supercomputers are often uniquely built systems that cost in the tens of millions to low hundreds of millions of dollars. By definition, at any given point in time, there are only about 50 to a 100 supercomputer worldwide.

Supercomputing refers to the various activities related to the design, manufacturing, and use of supercomputers, and is thus a subset of HPC. While supercomputing is in some sense a niche in the general computing world, it has tremendous impact because of its strategic importance for a country in three respects:

- Scientific competitiveness: as computation has become recognized as a mode of science of equal importance to theory and experiment, supercomputers have become an essential tool for basic science from nano science to cosmology
- National security: supercomputers essential tools for national security applications from the modeling of the nuclear stockpile to cryptanalysis
- Economic competitiveness: the use of supercomputer modeling in industrial applications from aerospace to geosciences, as well as the use of supercomputers for data intensive applications creates a competitive advantage.

In order to understand the relative state of supercomputing in a country it is very important to consider the ecosystem of supercomputing. The term "ecosystem" was used by the National Academies Study "Getting up to Speed, The Future of Supercomputin." Ecosystem refers to the fact that technologies, computer systems, software, applications, and human capital have to be developed simultaneously in order to make progress in supercomputing. They form an interlinked and mutual reinforcing system. I believe that the notion of ecosystem is essential to understand progress in the field, in particular as it relates to a nation such as China that is developing supercomputing capabilities.

3. Assessment of supercomputing developments in China

I will briefly address the components of the supercomputing ecosystem in China.

a) IT technology

The Chinese government has a consistent and long applied set of policies to encourage the development of local (domestic) IT companies and to build out the country's IT ecosystem, from device design to product development, system integration, standardization, and through to sales and service channels (modern IC fabrication facilities are a missing component). These policies include investments in basic and applied R&D, tax incentives, favorable procurements, transfer of experts from research organizations including universities and national institutes into start-up companies rolled out from research labs, etc. Examples of companies that have succeeded via this formula include Lenovo, Dawning, and Inspur. The "wall" between company and university is much less solid in China compared to the US.

China states that it is in their national interest to develop domestic IT IP for economic benefit, assurance of content, price, and national security. China has a very large and growing domestic market; products that can succeed locally might also be suitable for export. Chinese IT leaders believe that domestic hardware is only a few years behind world levels, but acknowledge that thus far, designs are largely following and adapting Western ideas. They also recognize that, regarding software applications, China is further behind and is only likely to catch up in limited sectors where local content can add significant value, such as GIS. Regardless, domestic developments not only benefit local organizations directly because of low price but also put price pressure on multinational products, and there is evidence that is a successful strategy.

b) HPC Systems

There are three, essentially independent, HPC developments; two utilizing modifications of Western processors and incorporating GPUs (accelerators). There is one indigenous processor effort. In all these systems the domestic engineering content is solid and can be directly traced to either a university or a research institute. Chinese government officials believe that development of both processors and software is necessary in order to build a domestic IT ecosystem. The national government is encouraging experimentation with different approaches to HPC architecture; at the same time there is some evidence that the technical community would prefer more standardization. The national government has also been encouraging and supporting the diffusion of HPC systems across the country. City (Beijing, Shanghai, etc) and provincial governments have been supplementing national support in order to acquire large systems. To some extent this is driven as a "build it and they will come" philosophy as well as competition between communities for HPC bragging rights. Nevertheless, systems sited locally are definitely enhancing local capabilities, especially if they include research and outreach components.

With today's processor technology, given adequate resources, it is relatively easy to build extremely large systems – hence China's recent entries into the well known Top 500 HPC list. In the Chinese equivalent (Top 100), domestic and multinational (primarily US) systems are equally represented, largely as a result of government support to domestic organizations. There is no indication this trend is abating.

c) Systems Software

China has done a good job leveraging open source software. This is particularly true regarding system software; Linux is the operating system (OS) of choice in nearly all large systems. The Chinese company, Red Flag is significant in this market at the low end and NFS at the server side (both spun out of the Chinese Academy of Science, Institute of Software). Another Chinese Linux derivative is Kylin produced by the National University of Defense Technology). Homegrown OS products are given priority in government procurements. In this context, the most likely growth is on the server side -- local server OS products have been gradually accepted by users and deployed in government and critical areas such as the state power grid due to their low price, controllable security, and lack of visibility to end users. There is also considerable government support for the development and deployment of Linux-based mobile OS because of that sector's potential market size.

d) Applications

Making effective use of any large system is challenging. Most supercomputer users are traditionally supported by government (including national defense, aerospace, weather, etc), basic science (physics, astronomy, geosciences, biology, etc), combined with automotive, mechanical, game/video, energy (oil/gas) and "knowledge" industries (e.g., search, social media).

Large-scale applications software is much weaker than hardware, although there is a considerable amount of local content. This is an important issue to the Chinese as multinational products are very expensive. In addition, key modules may be removed before importing. A typical domestic application software product can be viewed as a scale up of a small university or institute research effort. Examples are JASMIN, PHG, and GeoEast. The "scale" of applications, especially government research topics, is increasing but not yet at US levels. Further, while Chinese application software may be "good enough," compared to large US HPC applications the Chinese efforts lag considerably, not only in technical content, but in the relative lack of sophisticated use of modern software engineering, languages, and tools for their maintenance and expansion. Recently, government research investments have shifted toward software, encouraging the development of packages capable of running effectively on very large systems. This is potentially a significant development.

In spite of having access to large systems, so far there has been no winner or finalist in the annual ACM Gordon Bell Prize from China. I served on the selection committee in 2011, and there were several notable and very accomplished submissions. However, they did not yet quite reach the international competitive level in 2011, but in my judgment could have been a winner 10 years ago.

e) Human capital

I recently attended a workshop organized by ICCS in Beijing. It was evident that a lack of command of the English language remains a significant barrier for most students in becoming current and gaining expertise. Lack of learning materials further exacerbates the situation. However, the vigorous enthusiasm to learn is evident and therefore for many students who have crossed over the language barrier performance is commendable. The US remains the educational beacon for many aspiring students. While I don't want to generalize from a personal experience, it is clear that developing a community of experts in supercomputing requires times and is a task that will take 15 – 20 years.

A second issue that I can only speculate on is the fact that academic organizations China remain still hierarchical and stove piped. In a field such as supercomputing with rapid technology change and the need for large team work and multidisciplinary collaboration, an "academic" structure seems to a barrier to progress.

The state funding mechanism of Chinese supercomputers provides for the overall cost of building and hosting the system, while the facilities have to bear the operational costs. In order to raise annual operating costs, the facilities charge money for compute cycles – for example, Tianhe 1A charges one Yuan per core hour. This leads to an interesting bias towards industrial applications because the academic counterparts cannot afford the costs. Several researchers mentioned in conversations to me that they rather stay on their smaller local systems. This could be potentially a big barrier to the further development of supercomputing expertise.

4. Exascale computing and future developments

Since about 2007 US researches have set their sights on the next big goal in supercomputing: reaching the Exascale level. After several years of careful planning, workshop, and developing strategy documents, the U.S. efforts now face current federal budget realities that may lead to a more limited exascale program. In this context, Chinese developments are receiving increased attention, and each new technology announcement leads to exaggerated statements about "the Chinese are winning the Exaflops race".

Research toward the next level HPC target (Exascale) is underway, but cautiously, and Chinese researchers are carefully monitoring developments abroad, including participation, whenever possible, in open international fora. Thus far, there is no evidence that China will provide leadership in this direction.

Yes, there is national pride that Tianhe 1A was at least for one instance of the TOP500 list the #1 system in the world. But I would put this in the same category, as national pride over running a very successful Olympics. As a matter of fact, as one Chinese colleague remarked during my recent trip along these lines "... yes, China is investing in many areas of research and technology but nothing happens until the U.S. leads."

5. Summary and Conclusions

In summary, China's supercomputing activities should be viewed objectively. Recent Chinese successes neither indicate that US leadership is about to be supplanted, nor should they be minimized. The reality is that Chinese developments are strong and are improving rapidly, but there are many weak spots. China has people, resources, and commitment to succeed in supercomputing, and is deploying these strategically. An emerging economic superpower will also want to be a supercomputing power.