The Future of Computer Architecture

David Patterson
UC Berkeley

Version 3: December 12, 2008[1]

Computer Architecture is the field that designs computers, which sets the foundation for the entire IT industry.

Despite the tremendous resources at large companies such as IBM and Intel, there is a long track-record of breakthroughs from academic researchers in computer architecture that have led to new multi-billion-dollar industries. One reason is simply that many of the leading thinkers in computer architecture are in academia, as the number of awards, books, and papers document. Another reason is that academic researchers are not bound the business requirements of compatibility with legacy hardware and software. A third reason is that when information technology is changing rapidly, decades of experience may not be a huge asset, and bright young graduate students who don't know what can't be done are as likely to make an important contribution as those considerably more senior.

Thus, in areas where there is tremendous consensus on the guidelines on what to do, computer architects at Intel and IBM usually lead the way. When the directions are less clear, academics are often the path breakers even though they have fewer resources.

Today, we have hit the wall as to the practical limit to the amount of power that a microprocessor chip can dissipate; in the past each generation of chip used more power while getting more performance. We need to invent a new way to get more performance without more power. Moreover, the most interesting computers of the future are at the extremes in scale:

- The datacenter is the computer. Amazon, Google, Microsoft, and others are racing to construct buildings with 50,000 computers to run software as a service (SaaS).

- The cell phone is the computer. Millions of cell phones are shipped each day and they are increasing in functionality.

The "power wall" and the extremes in computer size mean that the old guidelines are out the window, so, if history is any guide, we're entering an era of increasing importance for academic computer architecture.

Note that in such an uncertain era there many chances for new multi-billion-dollar industries, and these new industries are likely to flourish close to where the researchers do

---

[1] For the most current version of this essay, as well as related essays, visit http://www.cra.org/ccc/initiatives

their work. We have seen the center of the IT universe shift geographically before during eras of change:

- New York in the mainframe era of the 1950s and 1960s;
- Boston in the minicomputer era of the 1970s and early 1980s;
- Silicon Valley in PC and Web era of the late 1980s to today.

Given the challenges mentioned above, other countries are investing tremendously in IT in general and in computer architecture specifically, since they see the opportunity in this era of change to shift the center of the IT universe and the jobs that come with it. Especially given the cut to academic IT research in general and to computer architecture specifically in the US, it's not hard to imagine the IT center moving further west in the next decade – to Beijing or Mumbai.

One area of tremendous opportunity in computer architecture is the manycore challenge. The goal is to invent computers that make it easy to write programs that are efficient, portable, correct, and scale as the number of cores per microprocessor increases – as easy as it has been to write programs for traditional computers. If this invention allows software to use many simple power-efficient cores instead of a single power-hungry core, this will reset the foundation for the IT industry for at least the next 30 years.

A second opportunity is inventing a new computer architecture that improves computer security and privacy, problems that plague the IT field. Architects could remove many of the vulnerabilities of today's computers if they were not bound by the legacy requirements of compatibility with today's computers. They could also provide new features to make it easier to build fast, secure, low-overhead virtual machines, making it easier and safer for software to migrate between the datacenter and the cell phone.

A third opportunity is invent computers that will remove the performance bottlenecks from new, highly productive programming environments such Ruby or Python. For example, the Ruby on Rails environment allows programmers to invent amazing new computer applications in just 1000 to 2000 lines of code—factors of 10 to 100 less than conventional approaches. Example application areas include personal health care, personal memory assistants, and personal digital educators. Alas, Ruby on Rails performance is factors of 3 to 10 worse than conventional systems. If we can invent computers that allow new programming systems like Ruby or Python to scale up to hundreds of cores while preserving their amazing programming productivity, we could unleash a new round of exciting applications that will lead to new multi-billion-dollar industries, just as we've done so many times in the past.

Intel and Microsoft recently funded two major academic centers to tackle these critical challenges – at UC Berkeley and the University of Illinois. Many other highly credible proposals were received, from strong teams pursuing diverse approaches. A national initiative would fund five more centers of excellence in computer systems and architecture, positioning the United States to maintain its preeminence in the IT field.