Why 10^{18} FLOPS will give us tools to know

By Luke Wadman
JUR Staff

This last February, the relationship between man and computer certainly changed. Watson, an IBM supercomputer designed to compete against humans in a game of *Jeopardy!*, beat the two greatest human champions of the famous quiz show. Ken Jennings and Brad Rutter fell to Watson to the tune of $77,147 versus $24,000 and $21,600 respectively.[1] While some, like Jennings, may see Watson's victory as a "welcome [to] our new computer overlords," I see a different and far less sinister conclusion.

Watson's achievement is more about computers being able to decipher the nuances of the human language offered on *Jeopardy!* - a challenge so great that few saw it as a remote possibility. A paper published in 2000 by IBM researchers described a QA (question answering) system that was a sort of predecessor to Watson; though this particular system could answer only 35% of complex, *Jeopardy!*-like inquiries correctly and would often take several minutes to come up with an answer.[2] The fact that computers like Watson can, for the most part, interpret questions given to it in natural human diction may herald intelligent QA systems, similar to "Computer" of Star Trek fame. This computer creates a sort of dialogue with the user, answering initial questions and then is able to answer follow-ups on the same topic.[1]

So, intelligent computer systems are well on their way, but they are a far cry from replacing (or enslaving) humanity. In the near future, computer systems like Watson will be integrated into systems like a physician's assistant, which will assist our doctors in diagnosing illnesses. The technology will also be used in businesses to direct customers and answer questions.[1]

So what else are these fantastic computers up to? To answer that question, I look back to 2000, and IBM's design of the ASCI White supercomputer. In July of that year, IBM sent 28 semi-trucks full of components to the Lawrence Livermore National Laboratory in California to begin construction of this supercomputer. This initial convoy provided only one quarter of the machine, which when fully assembled spanned an area the size of two basketball courts and needed 6 megawatts of electricity to both power and cool itself.[3] The ASCI White's main purpose was to simulate nuclear weapons blasts at a then-unprecedented speed of 12.3 trillion FLOPS (Floating point operations per second), or 12.3 teraFLOPS.[3] A FLOP is defined by any floating-point number calculation, whether it is addition, subtraction, division, and et cetera. The fastest supercomputer 5 years prior would have taken 60,000 years to perform a simulation that would take the ASCI White just 30 days.[3]

For a more applicable comparison to your average college student's laptop, the supercomputer had 8,192 processors, 6 terabytes of memory and 160 terabytes of disk storage. Today's average laptop may have one 4-core processor, 8 gigabytes of memory, and 500 gigabytes of disk storage.[3] In short, this 10-year-old supercomputer was 500-2000 times faster than today's average laptop, depending on which components you compare. It cost the LLNL over $110 million to produce but provided a priceless avenue for government researchers and university academics.[3]

Now, fast-forward 10 years. A new computing age based not on teraFLOPS (10^{12}), but on petaFLOPS (10^{15}) has arrived. In October 2010, the Tianhe-1A at the National Supercomputing Center in Tianjin, China, began operations at 2.566 petaFLOPS. In more conventional number terms, this comes out to 2.5 thousand million million (quadrillion) floating point operations per second.[4] The computer's uses include weather models, cancer research, and astrophysics. It is over 200 times faster than the 10-year old ASCI, yet uses less energy (4 MW) and costs less ($88 million) to produce.[5]

In a long, winding, number-permeated manner some may say is characteristic to an engineer (guilty), I've arrived at my main talking point of this editorial. Computers are continually getting smaller, faster, and more efficient. But where does this trend stop? Though matter is infinitely divisible down to elementary particles like quarks and bosons, we simply cannot continually shrink transistors to improve our computing capacity. This "Interconnect Bottleneck" point has supposedly been reached, and components supposedly cannot get any faster than they are today.

Researchers at Hewlett Packard (HP) may have found the answer. In order to keep up with increasing amounts of data needed digitalization and indexing, an article published in IEEE's (The Institute of Electrical and Electronics Engineers) Computer journal by an HP researcher proposes a solution.

This solution involves combining current technologies, rather than continuing the trend of shrinking and shrinking that has persisted in computing for the past 40 years. The nanostores, as they are called, feature memory storing centers called memristors, stacked along with conventional circuitry that will process the data stored within them. This differs from current computing setups which separate the processors and memory, in which large amounts of energy are needed to shuffle information from the memory to the microprocessor (where it is computed) then back to the memory for storage.[6] The article goes on to state that these nanostore setups could potentially store trillions of bytes of data along with heaps of processing power. In essence, the proposed memristor nanostores would combine the two most important components of a computer as opposed to the long standing trend of shrinking them individually. This reduces the need for energy to transport information large (in circuitry terms; the actual distances are on the scale of nano – 10^{-9} – meters) distances.

Cutting down on the energy costs required to perform complex mathematical complications is paramount for technological progress to continue growing at rates similar to those of the past 40 years. The energy required for an exaFLOPS (10^{18}) capable computer with today's current technology would be an astronomical 1.6 gigawatts, or about 150% what the average nuclear power plant provides.[7] Energy costs of the proposed HP technologies range around 1000 to 10000 picojoules per complex calculation, far lower than the most energy efficient of today's supercomputers.[7] With this proposed technology, our super (and regular) computers would set new standards for space and energy efficiency, while being thousands of times faster than our blazingly fast computers of today.

So now you've made it through all the numbers, statistics, words, and you're wondering to yourself, "Why does this matter? I don't need my computer to be 1,000 times faster than what it is; I already can't keep up with it." But the truth of the matter is this: these supercomputers built with nanostore technology can usher in a new era of human technological advancement, far faster and much different than the one we have seen since the 1960s. Being able to construct complex models to observe the physics, chemistry, and other processes behind nuclear explosions, galaxy formation, or cancerous tumor growth is of foremost importance to developing concrete scientific understanding of these topics.

Combined with the technology behind computers like Watson, new computers will be able to interact with humans much like we do with each other. They can also bring humanity's understanding of our questions and ourselves to a whole new level. There are a number of ethical questions and issues that come about at that point. These questions will seek to address where the line is drawn between humanity and technology. Bottom line, we should not fear that we could be replacing humans with these new machines. We should fear that these concerns might cloud our recognition of a monumental technological accomplishment; a platform capable of exploring and possibly answering our deepest questions far faster than we ever could. The more we know, the more we know.

References

[1]Markoff, J. (Feb. 2011) "Computer Wins on 'Jeopardy!': Trivial, It's Not" *The New York Times*. A1.

[2]Prager, J., Radev, D., Samn, V. (2000) "Ranking suspected answers to natural language questions using predictive annotation." *Proceedings of the 6th Conference on Applied Natural Language Processing*. IBM.

[3]Kahney, L (Jun. 2000) "IBM's Got A Big, Bad Computer" *Wired Magazine*.

[4](Oct. 2010) "China claims supercomputer crown" *BBC News*.

[5](Oct. 2010) "NVIDIA Tesla GPUs Power World's Fastest Supercomputer" *Nvidia.com* Nvidia. http://pressroom.nvidia.com. (Accessed 03/06/2011)

[6]Ranganathan, P. (Jan. 2011) "From Micro-processors to Nanostores: Rethinking Data-Centric Systems" *IEEE: Computer* 9162. 11. Pg 39-48.

[7]Markoff, J. (Feb. 2011) "Remapping Computer Circuitry to Avert Impending Bottlenecks" *The New York Times*. D3.