

Why the computer is doomed

By Omar El Akkad

Some time around the end of this decade, it will become impossible to improve computing power at the rate it has advanced for the past three decades

IBM engineers are currently putting the finishing touches on a beast of a computer.

The machine, code-named Blue Waters and set for delivery to the University of Illinois later this year, is the product of work completed in myriad IBM offices around the world. At 10 petaflops, it will be about five times faster than the fastest supercomputer in the world today.

To get a sense of how fast a peta-scale computer is, think of every human being on Earth doing a million calculations each. A peta-scale computer can do that every second. This is the kind of computer you use if you want to measure what every atom in a person's digestive system is doing, or if you are trying to predict what the Earth's climate will look like in 100 years.

Two years from now, IBM hopes to double that power, delivering a 20-petaflop machine to the Lawrence Livermore National Laboratory in California. By 2018, the company plans to build the world's first exa-scale supercomputer - think of every human being on Earth doing a billion calculations a second.

A mind-boggling display of digital brawn, to be sure, but right around this point everything will change for computer engineers. The fundamental laws that brought modern computing this far will begin to break down.

"If you take the fastest computer in the world circa 2005 ... that was consuming two megawatts of electricity. In 2018, that system will be 3,000 times more powerful," says Dave Turek, vice-president of Deep Computing at IBM. "If you simply do the math, that's six gigawatts of power. Two gigawatts is a moderate-sized nuclear power plant.

Powering a computer of that size using traditional methods "would be a disaster in terms of power and cooling. Forget what the system looks like, the first order of argument is that you can't afford to power it."

Some time around the end of this decade, one of the most profound transformations in the history of computer science will begin to take shape. It will simply become impossible to improve computing power at the rate it has advanced for the past three decades. The ceiling won't be a result of cost - in their current configurations, computer chips can only be made so small before running into the basic laws of physics.

The implications for the computer industry are enormous. It may be years away, but software programmers, circuit makers and computer manufacturers are nonetheless staring at a brick wall in the distance.

For decades, hardware and software engineers depended on something called Moore's Law, which essentially states that the number of transistors that can be placed on a circuit - the brains of a computer - will double roughly every two years. Although Moore's Law deals with circuit density, it is widely viewed in pop culture as a proxy for the speed of computers. In general terms, as long as

Moore's Law holds, the power of everything digital has the potential to continue to grow at an exponential rate.

We went from the chest-thumping notion of 'I have more gigahertz so I'm better,' to 'I have more cores so I'm better' *Dave Turek, vice-president of Deep Computing at IBM*

But Moore's Law is in some ways analogous to folding a piece of paper in half. The first few folds are easy, but eventually the effort required to make the next fold becomes too great. In the world of computing, that's what will happen some time between 2015 and 2020 -circuitry can only be made so thin before it becomes too thin for the very electrons travelling through it. Simply put, Moore's Law takes a back seat to the laws of physics.

As Mr. Turek points out, computer manufacturers actually saw this coming several years ago. Until early 2004, personal computer advertising was largely dominated by claims about processor speeds, measured in kilohertz a few decades ago and gigahertz today. But while customers and manufacturers could reasonably assume a quick turnaround between, say, a 300-megahertz system and a 600-megahertz system hitting the market, that doubling time doesn't necessarily hold from 3 gigahertz to 6 gigahertz - if only because the increase in price inevitably makes such exponential growth prohibitively expensive for the vast majority of consumers, likely well before manufacturers hit any technical limitations.

Indeed, manufacturers have spent more time exploiting the other end of Moore's law, taking chips that used to be state-of-the-art and putting them in new machines for a much lower price tag.

"That's been the sweet spot," says Deloitte Canada analyst Duncan Stewart "Using technology that was top-of-the-line five or six years ago and now goes for 40 bucks."

With the surge in mobile computing devices such as smart phones and tablets, the strategy became much more profitable. Because mobile computer users don't tend to do processor-intensive work such as video editing or graphics rendering on their devices, a less-powerful processor at a lower cost proves particularly attractive.

At the high-end, computer-makers stopped expanding vertically with ever-faster chips, and instead began expanding horizontally by selling computers with multiple processors, or cores.

"We went from the chest-thumping notion of 'I have more gigahertz so I'm better,' to 'I have more cores so I'm better,'" Mr. Turek says.

But as with vertical expansion, horizontal expansion has its limits: Simply doubling the number of processors in a computer doesn't result in double the performance, much in the same way that adding a second identical engine doesn't double a car's speed.

And as more cores are added, power usage and the potential for processor failure create eventually insurmountable barriers. The IBM exa-scale computer, for example, is projected to contain some 10 million cores. If engineers don't figure out new ways of managing a system of that scale, Mr. Turek says, some of the individual components of the supercomputer - such as the fabric connecting the processors - could end up consuming much more power than the entirety of the fastest supercomputer available today.

"One of things we've all been taught to ignore is reliability of integrated circuits," he says. " But you have to begin to look at systems from a total engineering perspective. Last decade was about simplistic assembly; you can't do that any more."

But regardless of how many innovations engineers can come up with, the laws of physics will eventually win. If the exponential growth in power and speed defined by Moore's Law is to continue in the coming decades, researchers will have to do something entirely different - design a competently new kind of computer.

Omar El Akkad is The Globe and Mail's technology reporter.