

# STRETCHING SUPERCOMPUTERS TO THE LIMIT

BY KATHERINE BOURZAC

The race is on to build the world's first 'exascale' machines. But the technological challenge is immense.

**A**t the end of July, workers at the Oak Ridge National Laboratory in Tennessee began filling up a cavernous room with the makings of a computational behemoth: row upon row of neatly stacked computing units, some 290 kilometres of fibre-optic cable and a cooling system capable of carrying a swimming pool's worth of water. The US Department of Energy (DOE) expects that when this US\$280-million machine, called Summit, becomes ready next year, it will enable the United States to regain a title it hasn't held since 2012 — home of the fastest supercomputer in the world.

Summit is designed to run at a peak speed of 200 petaflops, able to crunch through as many as 200 million billion 'floating-point operations' — a type of computational arithmetic — every second. That could make Summit 60% faster than the current world-record holder, in China.

But for many computer scientists, Summit's completion is merely one lap of a much longer race. Around the world, teams of engineers and scientists are aiming for the next leap in processing ability: 'exascale' computers, capable of running at a staggering 1,000 or more petaflops. Already, four national or international teams, working with the computing industries in their regions, are pushing towards this ambitious target. China plans to have its first exascale machine running by 2020. The United States, through the DOE's Exascale Computing Project, aims to build at least one by 2021. And the European Union and Japan are expected to be close behind.

Scientists anticipate that exascale computers will enable them to solve currently intractable problems in fields as varied as climate science, renewable energy, genomics, geophysics and artificial intelligence. That could include pairing detailed models of fuel chemistry and combustion engines in order to more quickly identify improvements that could lower greenhouse-gas emissions. Or it might allow for simulations of the global climate at a spatial resolution as high as a single kilometre. With the right software in hand, "there will be a lot of science we can then do that we can't do now", says Ann Almgren, a computational scientist at the Lawrence Berkeley National Laboratory in California.

But reaching the exascale regime is a tremendous technological challenge. The exponential increases in computing performance and energy efficiency that once accompanied Moore's law are no longer guaranteed, and aggressive changes to supercomputer components are needed to keep making gains. Moreover, a supercomputer that performs well on

a speed test is not necessarily one that will excel at scientific applications.

The effort to push high-performance computing to the next level is forcing a transformation in how supercomputers are designed and their performance measured. "This is one of the hardest problems I've seen in my career," says Thomas Bretin, a computer scientist at the Argonne National Laboratory in Illinois, who is working on medical software for exascale machines.

## ACCELERATED HARDWARE

Broader trends in the computing industry are shaping the path to exascale computers. For more than a decade, transistors have been so tightly packed that computing chips can't be made to run at faster rates. To circumvent this, today's supercomputers lean heavily on parallelism, using banks of chips to create machines with millions of processing units called 'cores'. A supercomputer can be made more powerful by stringing together more of these chips.

But as these machines get bigger, data management becomes more of a challenge. Moving data in and out of storage, and even within cores, takes much more energy than the calculations themselves. By some estimates, as much as 90% of the power supplied to a high-performance computer is used for data transport.

That has led to some alarming predictions. In 2008, in a report for the US Defense Advanced Research Projects Agency, a team headed by computer scientist Peter Kogge concluded that an exascale computer built from foreseeable technologies would need gigawatts of power — perhaps from a dedicated nuclear plant (see [go.nature.com/2hs3x6d](http://go.nature.com/2hs3x6d)). "Power is the number one, two, three and four problem with exascale computing," says Kogge, a professor at the University of Notre Dame in Indiana.

In 2015, in light of technological improvements, Kogge reduced this estimate down to between 180 and 425 megawatts. But that is still substantially more power than today's top supercomputers use; the system that leads the world rankings today — China's Sunway TaihuLight — consumes about 15 megawatts.

"Peter's report was important because it raised the alarm bell," says Rick Stevens, associate laboratory director for computing, environment and life sciences at Argonne. Thanks in part to Kogge's projections, he says, "there's been a lot of intellectual ferment around reducing power".

But in recent years, Stevens says, a host of new technologies has helped

Japan's K supercomputer is currently tenth-fastest on the Top500 list, but first by an alternative measure.

to bring down power consumption. A key advance has been bringing memory closer to computing cores to reduce the distance that data must traverse. For similar reasons, engineers have also built upward, stacking arrays of high-performance memory instead of spreading them across two dimensions. Supercomputers are also increasingly incorporating flash memory, which does not require power to maintain data, as some other, widely used types of memory do. And circuit designers have made it possible to shut down circuits in chips when they are not in use, or to change their voltage or frequency, to save on power.

More-fundamental changes to processors have also made a difference. One major development has been the adoption of general-purpose versions of graphics-processing units, or GPUs, which excel at the kind of data-intensive number-crunching needed for applications such as video-game rendering. Computers that incorporate GPUs, together with central processing units (CPUs) to direct traffic, are particularly proficient at physical simulations. From a programming point of view, says Katherine Yelick of Lawrence Berkeley National Laboratory, the calculations needed to realistically animate ocean waves in a film such as *Finding Nemo* are not dramatically different from simulating atmospheric dynamics in a climate model.

Other supercomputers have been built with 'lightweight' processors, which jettison some capabilities in favour of speed and energy efficiency. China used the lightweight scheme to build Sunway TaihuLight. The machine took the top spot with home-grown processors not long after the United States imposed a trade embargo (in 2015) on selling chips to supercomputing centres in China. The lightweight Sunway processors are not radically different from garden-variety CPUs, says Depei Qian, a computer scientist at Beihang University in Beijing, who is helping to manage China's exascale efforts. The individual cores are simplified, with limited local memory and lower speeds. But with many working together, the overall machine is faster.

The DOE's electricity-use target for its first exascale system, called Aurora, is 40 megawatts — with leeway for an absolute maximum of 60 megawatts. Computing giant Intel has been tasked with making the chips for the machine, and supercomputing company Cray, based in Seattle, Washington, has been subcontracted to assemble the full system. Details regarding how that target will be achieved are not yet public. But Al Gara, chief architect of high-performance and exascale computing at Intel in Santa Clara, California, says that the company is working on a new platform — including a new chip microarchitecture — that is designed to minimize power use.

Others have more-aggressive goals. Qian says that China will target as little as 30 megawatts for its first exascale system. With a later deadline of 2022 or 2023 and so more time to work on its system, the European

**“POWER IS THE NUMBER ONE, TWO, THREE AND FOUR PROBLEM WITH EXASCALE COMPUTING.”**

project might get down to 10 megawatts, says Jean-Philippe Nominé, a high-performance-computing specialist at CEA, the French Alternative Energies and Atomic Energy Commission in Saclay near Paris. But energy efficiency is only one factor: there is also the matter of performance.

The meaning of 'exascale' has become a matter of soul-searching for computer scientists. The simplest definition is a computer that can process a specific set of linear-algebra equations at a rate of 1 exaflops — equivalent to 1,000 petaflops. A group of researchers has used this benchmark, called LINPACK, to rank supercomputers on the Top500 list since 1993.

LINPACK has become shorthand for supercomputer performance, and since June 2013, supercomputers built in China have topped the list (see 'Steady leaps'). But speed isn't everything, says Jack Dongarra, a computer



scientist at the University of Tennessee in Knoxville and a founder of the Top500 list. “Everybody wants bragging rights,” Dongarra says. But he compares peak supercomputer ratings to the highest speed on a car’s speedometer. Although the ability to hit 300 kilometres per hour might seem impressive, what really gives most cars value is how they perform during daily drives at the speed limit.

In a similar manner, a computer’s speed at zipping through specific linear-algebra operations doesn’t necessarily reflect its ability to predict drug activity, train neural networks or perform complex simulations. All place different demands on processing power, on which sorts of calculations can be tackled in parallel and on how much data must be moved around. The Top500 “doesn’t measure how well the hardware is going to work on real applications”, says Barbara Helland, associate director for advanced scientific-computing research in the DOE’s Office of Science.

Despite that, today’s top supercomputers have been “built to deliver the highest LINPACK performance”, says Shekhar Borkar, a computer scientist who retired from Intel last year. A real-world scientific application might make use of 10% of that speed — but just 1.5–3% is more typical, Borkar says. He expects that this limitation will persist at the exascale.

In the United States, growing concern about this disconnect between peak speeds and utility has led to a different, applications-driven definition of exascale computing. The DOE aims for its first exascale computers to perform about 50 times better than the United States’ current fastest system: the 17.6-petaflops (as measured by LINPACK) Titan. That might mean, for example, screening 50 times as many potential solar materials in a given time, or modelling the global climate with a factor of 50 greater spatial resolution.

To pursue these gains, the DOE is working with hundreds of researchers from academia, government and industry. It has set up 25 teams, each tasked with devising software that could exploit an exascale machine to solve a specific scientific or engineering question, such as engine design. Stevens says the primary metric of success for US exascale supercomputers will be a “geometric mean” of their performance on the 25 applications.

In developing such computers, the agency is also trying to improve collaboration between people who use the supercomputers, those who write the software and the semiconductor companies responsible for building hardware. With the DOE’s exascale project, “we’re bringing these communities together. We can force that convergence,” says Doug Kothe, an Oak Ridge National Laboratory computer scientist who is leading the project. This strategy of uniting users and builders, called co-design, is not new. But, Kothe says, “it hasn’t been done in as broad and deep a way as it’s being done now”.

“I’ve been in this 20 years. This is the first time I’ve seen this kind of coordination and support,” says Thuc Hoang, programme manager for supercomputing research and operations at the National Nuclear Security Administration (NNSA) in Washington DC.

The United States is not alone in fostering collaborations between scientists and engineers in these disparate fields. China’s supercomputing programme, which has been criticized for prioritizing raw speed over science, is also using co-design in its exascale efforts, with a focus on 15 software applications. “We have to connect the hardware and software development with the domain scientists,” Qian says.

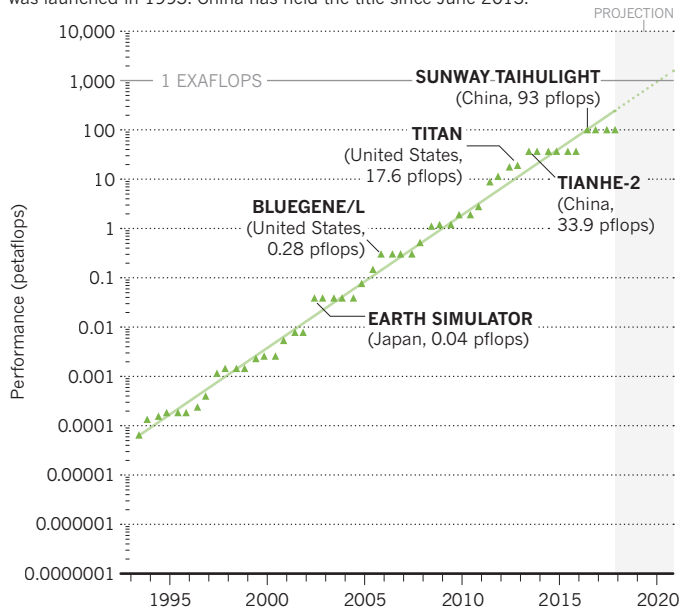
## FUTURE PROOF

But Borkar and some other observers are concerned that the first exascale systems in China and the United States might be stunt machines that won’t work well for real applications. “Delivering higher application performance would mean designing the machines differently, more realistically,” Borkar says. That, he adds, “would definitely compromise LINPACK performance, making them look bad from [a] marketing standpoint”. (Borkar notes that, although he still consults for the US government and for private companies, these views are his own.)

Borkar says he wishes that the United States, in particular, had stuck with plans first forged in 2008, which would have used the exascale shift as a chance to rethink computing more radically. “Evolutionary approaches will fail,” he says. “You need a revolutionary approach.” Stevens says that

## STEADY LEAPS

The performance of the most powerful supercomputer — rated by the Top500 list’s speed test and measured in petaflops — has been improving since the list was launched in 1993. China has held the title since June 2013.



big changes are happening behind closed doors. The DOE will complete its official contract with Intel around or after Christmas, he expects. Until then, he says, “I can’t tell you what we’re doing, but it’s very innovative”.

But there are limits to how aggressively supercomputing can be pushed forward. With each new generation of supercomputers, programmers must build on the software they have. “We have legacy code,” says Hoang. The programme she operates at the NNSA relies on supercomputers to maintain the United States’ arsenal in compliance with the ban on testing nuclear weapons. “Because of what my office is responsible for, we can’t just drop old codes that took us a decade to develop and validate.”

Budgetary constraints have also dictated US exascale plans. Aurora was intended to be a 180-petaflops machine, and to begin operation at Argonne in 2018. But the agency did not have enough money to begin commissioning exascale hardware. Instead of issuing a public request for proposals, the DOE changed Intel and Cray’s contract for Aurora to an exascale machine, to be supplied by 2021. Stevens is confident that they have the technology in the works to deliver.

Meanwhile, other exascale programmes are making progress. Still on target to reach exascale first, in 2020, is China. The country is weighing up three prototypes. Two, being built at supercomputing facilities that house that country’s fastest machines, are likely to be variations on the lightweight architecture the country has pioneered, says Dongarra. The third is being constructed by Sugon, a computing company in Beijing that has a relationship with high-performance chip developer AMD, and so has access to AMD’s workhorse microarchitectures. This machine, Dongarra thinks, will probably have new features and differ from the lightweights.

At the same time, researchers are considering what it will take to surpass the exascale and achieve even-faster and better-performing supercomputers in the coming decades. Producing that next generation of supercomputers might mean adopting technologies that are still in their earliest stages today: neuromorphic circuits, perhaps, which are modelled on the operation of neurons in the brain, or quantum computing.

But many researchers’ main concern is making sure they can deliver the promised exascale systems — and that scientific applications developed for them will work the moment they’re powered on. “Making [exascale] work,” says Helland. “That’s what keeps me up at night.” ■

**Katherine Bourzac** is a freelance journalist based in San Francisco, California.