

# The future of computing

The era of predictable improvement in computer hardware is ending. What comes next?



IN 1971 the fastest car in the world was the Ferrari Daytona, capable of 280kph (174mph). The world's tallest buildings were New York's twin towers, at 415 metres (1,362 feet). In November that year Intel launched the first commercial microprocessor chip, the 4004, containing 2,300 tiny transistors, each the size of a red blood cell.

Since then chips have improved in line with the prediction of Gordon Moore, Intel's co-founder. According to his rule of thumb, known as Moore's law, processing power doubles roughly every two years as smaller transistors are packed ever more tightly onto silicon wafers, boosting performance and reducing costs. A modern Intel Skylake processor contains around 1.75 billion transistors—half a million of them would fit on a single transistor from the 4004—and collectively they deliver about 400,000 times as much computing muscle. This exponential progress is difficult to relate to the physical world. If cars and skyscrapers had improved at such rates since 1971, the fastest car would now be capable of a tenth of the speed of light; the tallest building would reach half way to the Moon.

The impact of Moore's law is visible all around us. Today 3 billion people carry smartphones in their pockets: each one is more powerful than a room-sized supercomputer from the 1980s. Countless industries have been upended by digital disruption. Abundant computing power has even slowed nuclear tests, because atomic weapons are more easily tested using simulated explosions rather than real ones. Moore's law has become a cultural trope: people inside and outside Silicon Valley expect technology to get better every year.

But now, after five decades, the end of Moore's law is in sight (see *Technology Quarterly*). Making transistors smaller no longer guarantees that they will be cheaper or faster. This does not mean progress in computing will suddenly stall, but the nature of that progress is changing. Chips will still get better, but at a slower pace (number-crunching power is now doubling only every 2.5 years, says Intel). And the future of computing will be defined by improvements in three other areas, beyond raw hardware performance.

## Faith no Moore

The first is software. This week AlphaGo, a program which plays the ancient game of Go, beat Lee Sedol, one of the best human players, in the first two of five games scheduled in Seoul. Go is of particular interest to computer scientists because of its complexity: there are more possible board positions than there are particles in the universe (see page 73). As a result, a Go-playing system cannot simply rely on computational brute force, provided by Moore's law, to prevail. AlphaGo relies instead on "deep learning" technology, modelled partly on the way the human brain works. Its success this week shows that huge performance gains can be achieved through new algorithms. Indeed, slowing progress in hardware will provide stronger incentives to develop cleverer software.

The second area of progress is in the "cloud", the networks of data centres that deliver services over the internet. When computers were stand-alone devices, whether mainframes or desktop PCs, their performance depended above all on the speed of their processor chips. Today computers become more powerful without changes to their hardware. They can draw upon the vast (and flexible) number-crunching resources of the cloud when doing things like searching through e-mails or calculating the best route for a road trip. And interconnectedness adds to their capabilities: smartphone features such as satellite positioning, motion sensors and wireless-payment support now matter as much as processor speed.

The third area of improvement lies in new computing architectures—specialised chips optimised for particular jobs, say, and even exotic techniques that exploit quantum-mechanical weirdness to crunch multiple data sets simultaneously. There was less need to pursue these sorts of approaches when generic microprocessors were improving so rapidly, but chips are now being designed specifically for cloud computing, neural-network processing, computer vision and other tasks. Such specialised hardware will be embedded in the cloud, to be called upon when needed. Once again, that suggests the raw performance of end-user devices matters less than it did, because the heavy lifting is done elsewhere.

## Speed isn't everything

What will this mean in practice? Moore's law was never a physical law, but a self-fulfilling prophecy—a triumph of central planning by which the technology industry co-ordinated and synchronised its actions. Its demise will make the rate of technological progress less predictable; there are likely to be bumps in the road as new performance-enhancing technologies arrive in fits and starts. But given that most people judge their computing devices on the availability of capabilities and features, rather than processing speed, it may not feel like much of a slowdown to consumers.

For companies, the end of Moore's law will be disguised by the shift to cloud computing. Already, firms are upgrading PCs less often, and have stopped operating their own e-mail servers. This model depends, however, on fast and reliable connectivity. That will strengthen demand for improvements to broadband infrastructure: those with poor connectivity will be less able to benefit as improvements in computing increasingly happen inside cloud providers' data centres.

For the technology industry itself, the decline of Moore's law strengthens the logic for centralised cloud computing, already dominated by a few big firms: Amazon, Google, Microsoft, Alibaba, Baidu and Tencent. They are working hard to improve the performance of their cloud infrastructure. And they are hunting for startups touting new tricks: Google bought Deepmind, the British firm that built AlphaGo, in 2014.

For more than 50 years, the seemingly inexorable shrinking of transistors made computers steadily cheaper and more capable. As Moore's law fades, progress will be less metronomic. But computers and other devices will continue to become more powerful—just in different and more varied ways. ■