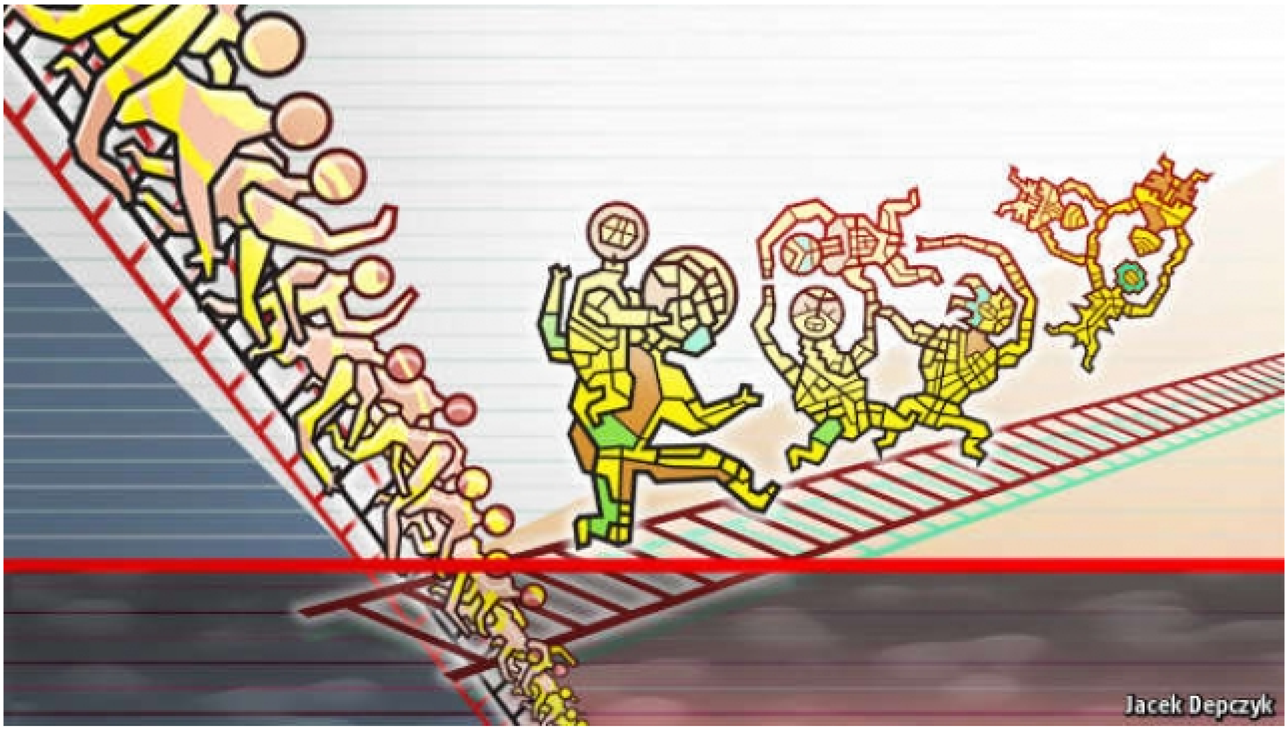
The Economist

Difference engine

More from Moore

Chipmaking: Moore's law may be running out of steam, but chip costs will continue to fall



Print edition | Technology Quarterly Sep 3rd 2015

THERE is a popular belief that Moore's law is coming to an end. The doubling of transistors on a chip every two years, for the same cost, has continued apace since Gordon Moore, one of Intel's founders, noted it back in 1965. At the time, a few hundred transistors could be crammed onto a sliver of silicon. Today's powerful chips contain billions.

Having become smaller and smaller over the decades, the crucial features within transistors are now approaching the size of atoms. Quantum and thermodynamic effects that occur at such microscopic dimensions have loomed large for several years, and these threaten to place physical limits on further shrinkage.

Until now, the integrated circuits in a chip have used a two-dimensional (planar) structure, with a metal gate mounted across a flat, conductive channel of silicon. The gate controls the current flowing from a source electrode at one end of the channel to a drain electrode at the other end. A small voltage applied to the gate lets current flow through the transistor. When there is no voltage, the transistor is switched off. These two binary states (on and off) are the ones and zeros of digital language.

However, when transistors are shrunk beyond a certain point, electrons flowing from the source can tunnel through the insulator protecting the gate, instead of flowing in a controlled manner direct to the drain. This wastes energy, raises the temperature and can cause the device to fail. Leakage becomes a serious problem when insulating barriers within transistors approach a few nanometres (billionths of a metre, or nm) in thickness. Below that, leakage increases exponentially, rendering the device almost useless.

Features this small are increasingly common in today's transistors. Intel's Broadwell chips, introduced in 2014, are made using 14nm process technology (which refers to the smallest "half-pitch" between identical features on a chip). Three-dimensional architecture helps to reduce leakage. Within the transistor itself, some features are considerably smaller than their half-pitch. The gate's insulating layer in Intel's chip is reckoned to be no more than 0.5nm thick—little more than a couple of silicon atoms across.

According to the tick-tock of Moore's law, Intel's next generation of processors, code-named Cannonlake, were to be fabricated using a 10nm technology and arrive in mid-2016. But the company recently admitted that the migration from 22nm to 14nm had been far more difficult than expected. As a result, 10nm chips are to be delayed until later in 2017. Suddenly, the doubling of processor performance is to take more than three years instead of two.

Many have declared this to be the end of Moore's law. But that is mistaken. If truth be told, Moore's law was never anything more than a rule of thumb for scheduling manufacturing targets. As such, it served as a metronome that helped Intel set the tempo of product announcements—and thereby encourage computer-makers to keep coming back every couple of years for increasingly powerful processors. Like it or not, the rest of the chip industry, usually a generation or two behind, was obliged to follow suit.

Moreover, Moore's law has always been as much about reducing the cost of transistors as about increasing performance. By doubling transistor density, individual chips got smaller, allowing more of them to be printed on a silicon wafer. After the wafer was sliced and diced, individual chips then cost less. Because leads and contacts on a chip cannot be shrunk as easily, doubling the density of transistors does not quite halve the price of individual devices. But it comes close.

Unfortunately, as transistors get smaller, more defects creep in. There is thus a trade-off between complexity and cost. And, while the cost per transistor is almost inversely proportional to the number of transistors crammed in a chip, there comes a point where the decrease in yield (percentage of good chips on a wafer) begins to outweigh the benefits of the chip's increasing complexity. In short, a minimum transistor cost exists for each particular generation of processing technology.

And here's the crunch: that minimum cost per transistor has been rising since 28nm chips hit the market five years or so ago. That is partly a result of decreasing yields, but also because of the escalating cost of the photolithography equipment needed to fabricate ever-smaller circuits. In short, the cost-effectiveness of chip manufacturing seems to have hit a sweet spot at about 28nm.

There is a lot to be said for sticking with legacy technology like 28nm. Transistor shrinkage over the years has left spare room on the chip to add specialised processing units for handling such services as graphics, video and cryptography.

The new system

As it is, the popularity of mobile communications and computing has encouraged semiconductor firms to embed as many features as possible in their processors. Such devices, known as "systems on a chip" (SOC), tend to devote around 65% of their real estate to memory, with the rest for everything else—including all the processor's logic gates, the necessary input/output circuitry, and numerous analogue functions needed to run a phone, tablet, laptop or whatever. While it is possible to shrink the size of the logic on an SOC, the memory components do not scale anything like as well, and the analogue circuitry barely at all.

That means Moore's law affects only a small portion of an SOC. As such a device is never going to gain significant cost or performance benefits from shrinking further, there is good reason to stick with mature processes like 28nm, with their minimal cost.

If transistor densities no longer double, will engineers continue to see chip costs halve every few years? With SOC devices based on mature technology, that is a distinct possibility. Eventually silicon will cede its semiconductor leadership to gallium arsenide or some other material, such as titanium trisulphide, being developed for the role. But that is probably years into the future.

In the meantime, the 50-year era of pushing down semiconductor costs through improvements in manufacturing know-how is about to be superseded by a new age of making chips cheaper, faster and better through smarter design, including systems on a chip. In so doing, Moore's law could get a new lease of life.