

# Future-of-chipmaking

TWO YEARS shy of its 60th birthday, Moore's law has become a bit like Schrödinger's hypothetical cat—at once dead and alive. In 1965 Gordon Moore, one of the co-founders of Intel, observed that the number of transistors—a type of electronic component—that could be crammed onto a microchip was doubling every 12 months, a figure he later revised to every two years.

That observation became an aspiration that set the pace for the entire computing industry. Chips produced in 1971 could fit 200 transistors into one square millimetre. Today's most advanced chips cram 130m into the same space, and each operates tens of thousands of times more quickly to boot. If cars had improved at the same rate, modern ones would have top speeds in the tens of millions of miles per hour.

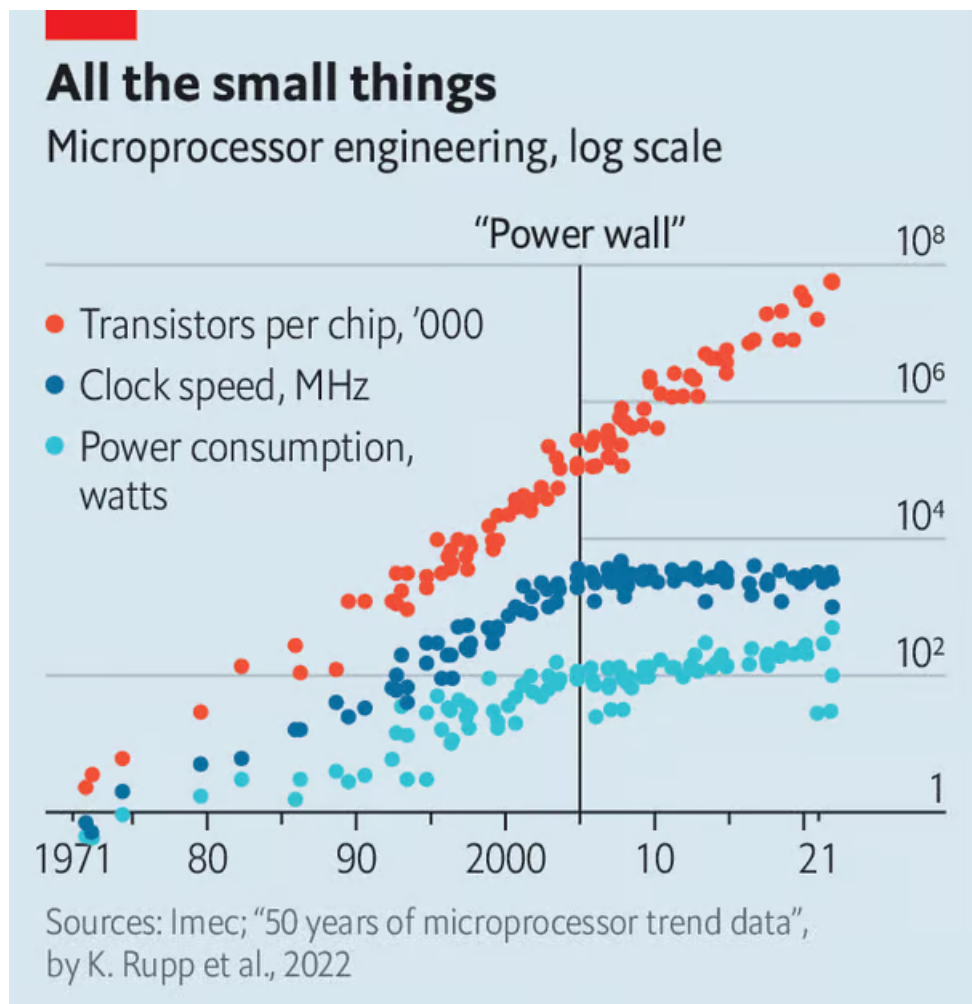
Moore knew full well that the process could not go on for ever. Each doubling is more difficult, and more expensive, than the last. In September 2022 Jensen Huang, the boss of Nvidia, a chipmaker, became the latest observer to call time, declaring that Moore's law was "dead". But not everyone agrees. Days later, Intel's chief [Pat Gelsinger](#) reported that Moore's maxim was, in fact, "alive and well".

Delegates to the International Electron Devices Meeting (IEDM), a chip-industry shindig held every year in San Francisco, were mostly on Mr Gelsinger's side. Researchers showed off several ideas dedicated to keeping Moore's law going, from exploiting the third dimension to sandwiching chips together and even moving beyond silicon, the material from which microchips have been made for the past half-century.

A transistor is to electricity what a tap is to water. Current flows from a transistor's source to its drain via a gate. When a voltage is applied to the gate, the current is on: a binary 1. With no voltage on the gate, the current stops: a binary 0. It is from these 1s and 0s that every computer program, from climate models and ChatGPT to Tinder and Grand Theft Auto, is built.

## Small is beautiful

For decades transistors were built as mostly flat structures, with the gate sitting atop a channel of silicon linking the source and drain. Making them smaller brought welcome side benefits. Smaller transistors could switch on and off more quickly, and required less power to do so, a phenomenon known as Dennard scaling.



By the mid-2000s, though, Dennard scaling was dead. As the distance between a transistor's source and drain shrinks, quantum effects cause the gate to begin to lose control of the channel, and electrons move through even when the transistor is meant to be off. That leakage wastes power and causes excess heat that cannot be easily disposed of. Faced with this "power wall", chip speeds stalled even as transistor counts kept rising (see chart).

In 2012 Intel began to build chips in three dimensions. It turned the flat conducting channel into a fin standing proud of the surface. That allowed the gate to wrap around the channel on three sides, helping it reassert control (see diagram). These transistors, called "finFETs", leak less current, switch a third faster and consume about half as much power as the previous generation. But there is a limit to making these fins thinner and taller, and chipmakers are now approaching it.

## There's still some room at the bottom

### Projected transistor development, 2022-36



**2022**  
**N3**



Type FinFET

Metal pitch\* 22nm

**2024**  
**N2**



GAA nanosheet

21

**2028**  
**A10**



GAA forksheet

16

**2032**  
**A5**



CFET

16-12

**2036**  
**A2**



CFET atomic

16-12

\*Smallest distance between components, in nanometres (billionths of a metre)

Source: Imec

The next step is to turn the fins side on such that the gate surrounds them completely, giving it maximum control. Samsung, a South Korean electronics giant, is already using such transistors, called “nanosheets”, in its newest products. Intel and TSMC, a Taiwanese chip foundry, are expected to follow soon. By stacking multiple sheets and reducing their length, transistor sizes can drop by a further 30%.

Szuya Liao, a researcher at TSMC, compares going 3D to urban densification—replacing sprawling suburbs with packed skyscrapers. And it is not just transistors that are getting taller. Chips group transistors into logic gates, which carry out elementary logical operations. The simplest is the inverter, or “NOT” gate, which spits out a 0 when fed a 1 and vice versa. Logic gates are made by combining two different types of transistor, called n-type and p-type, which are produced by “doping” silicon with other chemicals to modify its electrical properties. An inverter requires one of each, usually placed side by side.

At IEDM Ms Liao and her colleagues showed off an inverter called a CFET built from transistors that are stacked on top of each other instead. That reduces the inverter’s footprint drastically, to roughly that of an individual transistor. TSMC says that going 3D frees up room to add insulating layers, which means the transistors inside the inverter leak less current, which wastes less energy and produces less heat.

The ultimate development in 3D chip-making is to stack entire chips atop one another. One big limitation to a modern processor’s performance is how fast it can receive data to crunch from memory chips elsewhere in the computer. Shuttling data around a machine uses a lot of energy, and can take tens of nanoseconds, or billionths of a second—a long time for a computer.

Julien Ryckaert, a researcher at [imec](#), a chip-research organisation in Belgium, explained how 3D stacking can help. Sandwiching memory chips between data-crunching ones drastically reduces both the time and energy necessary to get data to where it needs to be. In 2022 AMD, an American firm whose products are built by TSMC, introduced its “X3D” products, which use 3D technology to stick a big blob of memory directly on top of a processor.

As with cities, though, density also means congestion. A microchip is a complicated electrical circuit that is built on a circular silicon wafer, starting from the bottom up. (Intel likens it to making a pizza.) First the transistors are made. These are topped with layers of metal wires that transport both electrical power and signals. Modern chips may have more than 15 layers of such wires.

As chips get denser, routing those power and data lines gets harder. Roundabout routes waste energy, and power lines can interfere with data ones. 3D logic gates, which pack yet more transistors into a given area, make things worse.

To untangle this mess, chipmakers are moving power lines below the transistors, an approach called “backside power delivery”. Transistors and data lines are built as before. Then the wafer is flipped and thick power lines are added to the bottom. Putting the power wires along the underside of the chip means fundamental changes to the way expensive chip factories operate. But shortening the length of the power lines means less wasted energy and cooler-running chips. It also frees up nearly a fifth of the area above the transistors, giving designers more room to squeeze in extra data lines. The end result is faster, more power efficient devices without tinkering with transistor sizes. Intel plans to use backside power in its chips from next year, though combining it with 3D transistors in full production is still a while away.

Even making use of an extra dimension has its limits. Once a transistor’s gate length approaches ten nanometres the channel it governs needs to be thinner than about four nanometres. At these tiny sizes—mere tens of atoms across—current leakage becomes much worse. Electrons slow down because silicon’s surface roughness hinders their movement, reducing the transistor’s ability to switch on and off properly.

Some researchers are therefore investigating the idea of abandoning silicon, the material upon which the computer age has been built, for a new class of materials called transition metal dichalcogenides (TMDs). These can be made in sheets just three atoms thick. Many have electrical properties that mean they leak less current from even the tiniest of transistors.

Three TMDs in particular look promising: molybdenum disulphide, tungsten disulphide and tungsten diselenide. But while the industry has six decades of experience with silicon, TMDs are much less well understood. Engineers have already found that their ultra-thin profile makes it difficult to connect transistors made from them with a chip’s metal layers. Consistent production is also tricky, particularly at the scales needed for reliable mass production. And the materials’ chemical properties mean it is harder to dope them to produce n-type and p-type transistors.

## The atomic age

Those problems are probably not insurmountable. (Silicon suffered from doping problems of its own in the industry’s early days.) At the IEDM, Intel was showing off an inverter built out of TMDs. But Eric Pop, an electrical engineer at Stanford University, thinks it will be a long while before they replace silicon in commercial products. For most applications, he says, silicon remains “good enough.”

At some point, the day will arrive when no amount of clever technology can shrink transistors still further (it is hard to see, for instance, how one could be built with less than an atom's worth of stuff). As Moore himself warned in 2003, "no exponential is for ever." But, he told the assembled engineers, "your job is delaying for ever". Chipmakers have done an admirable job of that in the two decades since he spoke. And they have at least sketched out a path for the next two decades, too.