## **What Is Moore's Law?**

Moore's Law refers to Moore's perception that the number of transistors on a microchip doubles every two years, though the cost of computers is halved. Moore's Law states that we can expect the speed and capability of our computers to increase every couple of years, and we will pay less for them. Another tenet of Moore's Law asserts that this growth is exponential.

## **Why Moore's Law is no longer valid**

Moore’s law is effectively dead because of thermal limitations *i.e.* the constraints of how much power can be dissipated from a chip die.

Since the dawn of the super-scalar out-of-order Intel architecture age, way back in 1995, more and more parallelism in form of additional transistors was thrown at the problem, with exponentially less utilization of those additional marginal transistors.

Nowadays most of those additional transistors sit idle with ever-decreasing fraction doing work for typical sequential workloads. For it would be physically impossible for all of them to be running at once as that would burn up the die.

It is now largely forgotten that clock scaling was the earliest and most well-known form of Moore’s law. It was truly stunning to keep hearing about chips doubling their FREQUENCY every 18 months. But that stopped about 15 years ago because CMOS transistors, which were rock-solid in terms of low power consumption (this is how they won over competing transistor technologies) became very leaky, dissipating much more power.

To compensate for this downer, new cores were invented as basically a marketing ploy to get masses to believe their new shiny processors were continuing to be supposedly exponentially better. Well, try to examine utilization of those additional cores in modern CPUs.

Keep in mind that multi-core Intel CPUs are hyper-threaded which actually means every core has the capability to run TWO hardware threads in the best case. And each of those threads is super-scalar out-of-order, meaning they can issue FIVE or more instructions in parallel simultaneously, in the best case.

It is kind of shocking to know that in a typical sequential instruction mix in a super-scalar CPU with five execution units the average utilization is about 1.7 meaning that out of five only one additional is used with 70% efficiency! All the others are needed to just to reach the 70%, on average, on the second one.

In summary, Moore’s law has been effectively dead for more than 10 years now, but that has been obscured and obfuscated by marginal parallelization. New computing paradigms such as AI and crypto offer some hope but there will be no free lunch until we find some new technology comparable to CMOS efficiency in the early days. Quantum computers are another such hope but the difficulty of their programming and just understanding what they do is truly unprecedented and off the charts.

Moore’s Law was an observation about process technology and economics. For half a century it drove the aspirations of the semiconductor industry. But the other limitation to packing more transistors onto to a chip is a physical limitation called [Dennard scaling](https://en.wikipedia.org/wiki/Dennard_scaling)– as transistors get smaller, their [power density](https://en.wikipedia.org/wiki/Power_density) stays constant, so that the power use stays in proportion with area. This basic law of physics has created a “Power Wall” — a barrier to clock speed — that has limited microprocessor frequency to around 4 GHz since 2005. It’s why clock speeds on your microprocessor stopped increasing with leaps and bounds 13 years ago. And why memory density is not going to increase at the rate we saw a decade ago.

This problem of continuing to shrink transistors is so hard that even Intel, the leader in microprocessors and for decades the gold standard in leading fab technology, [has had problems](https://www.theinquirer.net/inquirer/news/3031241/intel-delays-10nm-chips-yet-again-as-firm-suffers-yield-issues). Industry observers have suggested that [Intel has hit several speed bumps](https://semiaccurate.com/2018/08/02/intel-guts-10nm-to-get-it-out-the-door/) on the way to their next generation push to 10- and 7-nanometer designs and now is trailing TSMC and Samsung.

This combination of spiraling fab cost, technology barriers, power density limits and diminishing returns is the reason [GlobalFoundries](https://www.anandtech.com/show/13277/globalfoundries-stops-all-7nm-development) threw in the towel on further shrinking line widths . It also means the future direction of innovation on silicon is no longer predictable.

**It’s the End of the Beginning**

The end of putting more transistors on a single chip doesn’t mean the end of innovation in computers or mobile devices. (To be clear, 1) the bleeding edge will advance, but almost imperceptibly year-to-year and 2) GlobalFoundaries isn’t shutting down, they’re just no longer going to be the ones pushing the edge 3) existing fabs can make current generation 14nm chips and their expensive tools have been paid for. Even older fabs at 28-, 45-, and 65nm can make a ton of money).

But what it does mean is that we’re at the end of guaranteed year-to-year growth in computing power. The result is the end of the *type of innovation we’ve been used to for the last 60 years.* Instead of just faster versions of what we’ve been used to seeing, device designers now need to get more creative with the 10 billion transistors they have to work with.

It’s worth remembering that human brains have had 100 billion neurons for at least the last 35,000 years. Yet we’ve learned to do a lot more with the same compute power. The same will hold true with semiconductors — we’re going to figure out radically new ways to use those 10 billion transistors.

For example, there are [new chip architectures](https://semiengineering.com/big-changes-for-mainstream-chip-architectures/) coming (multi-core CPUs, [massively parallel CPUs](https://www.nextplatform.com/2018/08/30/intels-exascale-dataflow-engine-drops-x86-and-von-neuman/) and special purpose silicon for [AI/machine learning](https://www.cnbc.com/2017/09/19/ai-chip-start-ups-graphcore-and-mythic-raising-big-venture-rounds.html) and [GPU’s](https://en.wikipedia.org/wiki/Graphics_processing_unit) like [Nvidia](https://en.wikipedia.org/wiki/Nvidia)), new ways to package the chips and to interconnect memory, and even new types of memory. And other designs are pushing for extreme low power usage and others for very low cost.

**Lessons Learned**

* Moore’s Law — the doubling of every two years of how many transistors can fit on a chip — has ended
* Innovation will continue in new computer architectures, chip packaging, interconnects, and memory
* [5G networks](https://en.wikipedia.org/wiki/5G) will move more high-performance consumer computing needs seamlessly to the cloud
* New applications and hardware other than CPU speed ([5G networks](https://www.pcmag.com/article/345387/what-is-5g), displays, sensors) will now drive sales of consumer devices
* New winners and losers will emerge in consumer devices and chip suppliers
* Temperature increases as power increases.
* Power increases as transistor density increases.
* Voltage scaling reduces (dynamic) power consumption.
* Voltage scaling cannot prevent leakage power loss.
* Voltage scaling is limited due to noise or threshold voltage.