Doubling Down through miniaturization

Source: https://www.linkedin.com/pulse/doubling-down-through-miniaturization-tamoghna-sengupta-i2dzc/
Generated on: 2026-01-15

The paper – “THE IMPORTANCE OF (EXPONENTIALLY MORE) COMPUTING POWER” published by MIT Computer Science and AI Lab says increased computational power enables more accurate and timely weather forecasts. The global leader in weather forecasting models - the European Centre for Medium-Range Weather Forecasts (ECMWF) has increased the resolution of their models from 210 kilometers in 1979 to 9 kilometers in 2016. As per the article – “Improvements in these three dimensions of resolution are computationally expensive, with each doubling requiring an 8-fold rise in the amount of computation”.

Now, when I take a step back and think about – what it means to humanity? Imagine, the families of fishermen staying on the coast of an ocean, and people staying in weak houses vulnerable to tornados – It’s a 20x rise in their chances of surviving the next calamity (even after discounting the negligence of any of the local authorities).

Very recently, we witnessed the miraculous speed of the design of mRNA vaccines to combat the SARS-CoV-2 genome. Throughout the entire process, in my opinion, the unsung hero was high-performance computing which allowed rapid sequencing of the SARS-CoV-2 genome and finally quickly designing the mRNA vaccines by identifying the optimal mRNA sequences to encode the viral spike protein.

This raw enormous computational power has not been generated or developed overnight. The capacity to produce the huge computational power that we take for granted today has built up over time – through miniaturization of chips, exponentially increasing the number of transistors in IC.

Ok, let us hang in there. Let us take a further step back – the animalistic computational power we see today can be traced back to the power of underlying semiconductor chips.

What is the underlying principle that has driven the rise of the power of semiconductors for decades – It is Moore’s Law, which states that that the number of transistors on a microchip doubles approximately every two years, leading to exponential growth in computing power and performance while reducing cost per transistor.

Again, I am getting ahead of myself. Let me stop and start from the beginning.

Way back in the 1930s, electronic devices functioned using vacuum tubes. Vacuum tubes were used for amplification and switching.

As per Wikipedia – “By the end of its operation in 1956, ENIAC contained 18,000 vacuum tubes, 7,200 crystal diodes, 1,500 relays, 70,000 resistors, 10,000 capacitors, and approximately 5,000,000 hand-soldered joints. It weighed more than 30 short tons (27 t), was roughly 8 ft (2 m) tall, 3 ft (1 m) deep, and 100 ft (30 m) long, occupied 300 sq ft (28 m2) and consumed 150 kW of electricity.”

That was quite overwhelming, right? William Shockley, a physicist working at Bell Labs felt the same. He became interested in solid-state physics and semiconductors and had the belief that they could replace the monstrous vacuum tubes. To achieve what he believed was possible, he assembled a team that included Bardeen and Brattain.

They achieved their breakthrough on December 23, 1947, at Bell Labs (At that time it was jointly owned by Western Electric and AT&T). They demonstrated the first type of transistor, which was called the point-contact transistor. The semiconductor used in the case of the point-contact transistor was Germanium – Note, that it's not silicon yet.

People had just started celebrating the advent of transistors through Germanium semiconductors, only to realize that it is not sustainable. Germanium started to melt at high temperatures.

This is when Morris Tanenbaum from Bell Labs came to the rescue – He was the first one to develop silicon transistors on January 26, 1954.

Texas Instruments engineers were the first to commercially build and market silicon transistors to power electronic devices.

We need to pause here and understand what just happened – Bell Labs was the first organization to develop and demonstrate the capabilities of silicon transistors, but they did not pursue it further – thinking it would not be commercially viable. Texas Instruments latched onto an opportunity created by someone else – Gordon Teal, who led the development of mass-producing silicon transistors and made the first commercially available silicon transistors available on April 14, 1954. Look at the foresight – fast forward today - and we have Silicon Valley!

So far so good. However, the technology world quickly realized powering electronic devices using transistors is not scalable. How could they climb up the complexity curve of electronic devices if they need to keep packing more transistors, resistors, and capacitors into a device? – Size, portability, and the amount of heat generated became an issue.

In comes Jack Kilby. Jack Kilby was an engineer at Texas Instruments who invented the first IC (Integrated Circuit) on September 12, 1958. However, he built the prototype using germanium, which we discussed earlier has its challenges. Around the same time, Robert Noyce at Fairchild Semiconductor developed a similar IC demonstrated by Kilby using Silicon.

That is, it – Now you have a chip, which can pack multiple transistors, resistors, and capacitors and can comfortably power computing.

Ok, Hang on. Now we can pack multiple transistors, resistors, and capacitors into a single IC. But how to connect them together and make them work in tandem? Well, the planar process patented by Jean Hoerni while working at Fairchild Semiconductor in 1959 was the answer. Robert Noyce built on Hoerni's work to build scalable and reliable IC for the first time – This in a way was the inception of Lithography and the miniaturization race which sent technological innovation into a frenzy.

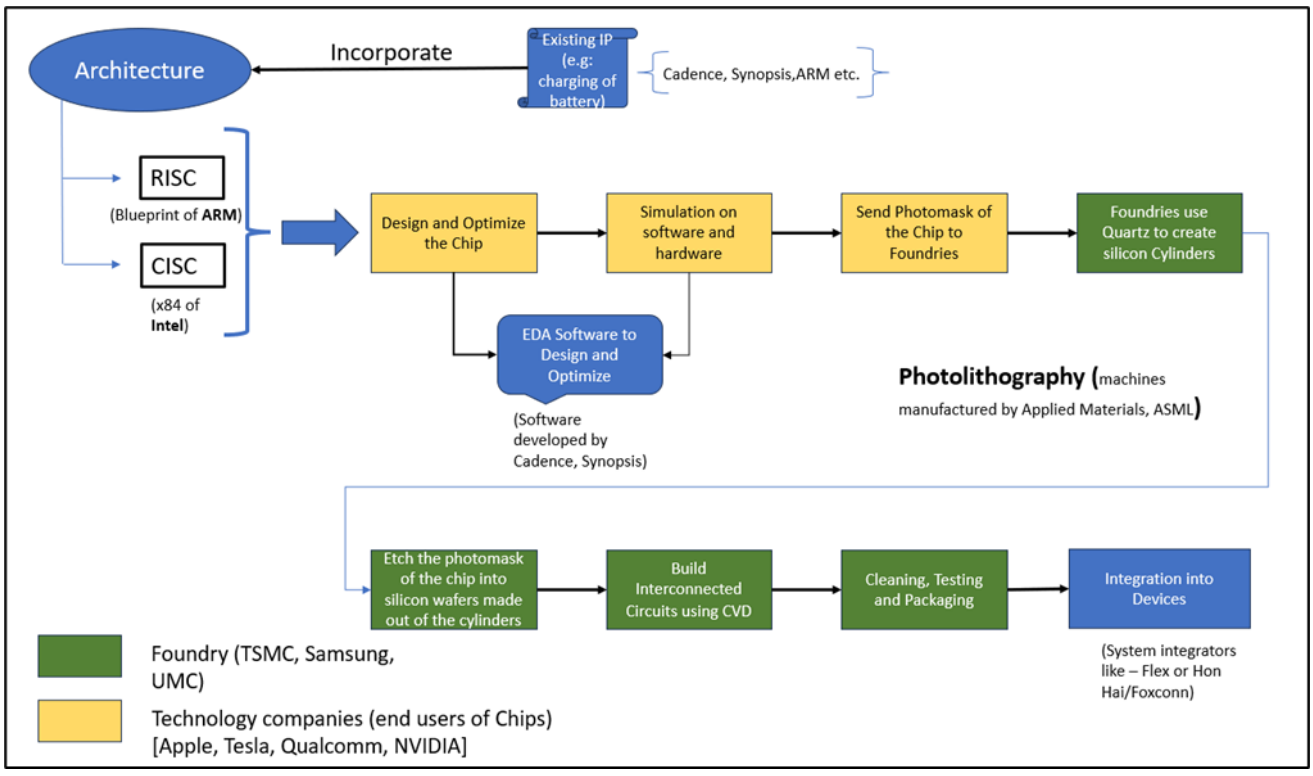
Now, while Robert Noyce was doing all these, Gordon Moore was working as the director of research at Fairfield Semiconductor. Gordon Moore was an interesting person, he was an

engineer and Businessman, which I believe is an amazing blend. He was asked by Electronics Magazine to predict what he thought might happen in the semiconductor components industry over the next ten years, in response, he published the article – “Cramming More Components onto Integrated Circuits” where he had put up an intriguing future roadmap of miniaturization for the semiconductor industry. This is what he says in the section “Costs and Curves” of the article, which we now call Moore’s Law.

> *“The complexity for minimum component costs has increased at a rate of roughly a factor of two per year (see graph). Certainly, over the short term this rate can be expected to continue, if not to increase. Over the longer term, the rate of increase is a bit more uncertain, although there is no reason to believe it will not remain nearly constant for at least ten years. That means by 1975, the number of components per integrated circuit for minimum cost will be 65 000 .”*

Moore’s Law paved the way for stacking more components into chips and increasingly making them more powerful. The beast thus formed - computing power kept getting bigger and chips kept getting smaller.

The instruction set driving these chips moved from circuit (CISC) to compiler (RISC), but the chip miniaturization process never stopped. Photolithography, which germinated in the form of a planner process kept getting even more complicated, so much so, that today there are only one or two companies that can perform EUV (Extreme Ultraviolet) Lithography – which I believe is even more complicated a technology than rocket science.



One final thought: Each of the companies mentioned here – Fairchild Semiconductor, Bell Labs, and Texas Instruments was an NVIDIA of their time.