Assessment 01

Moore's Law

Definition

Moore's law is the observation that over the history of computing hardware, the number of transistors on integrated circuits doubles approximately every two years.

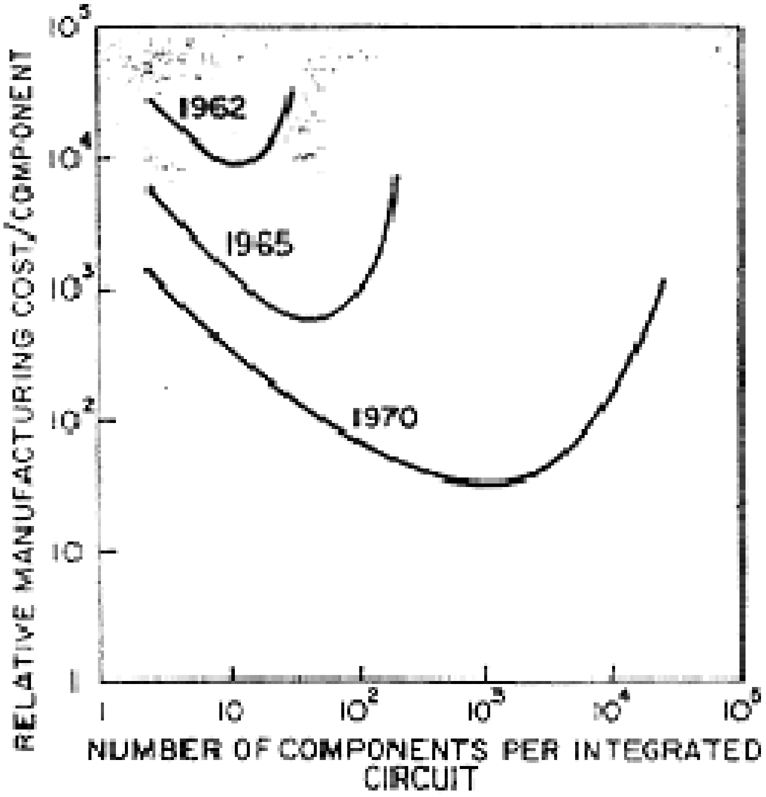
The period often quoted as "18 months" is due to Intel executive David House, who predicted that period for a doubling in chip performance (being a combination of the effect of more transistors and their being faster).

Multiple outcomes

Although commonly bound and related to the exponential transistor density/amount behavior, several approaches and outcomes were derived from Moore's original paper "Cramming More Components onto Integrated Circuits," published in Electronics magazine in 1965:

a. Transistor speed increases exponentially over time
b. Transistors per dollar grow exponentially over time
c. Transistor density grow exponentially over time
d. The number of transistors in a package grows exponentially over time

Again, although most commonly related and remembered as an observation on transistor density over time, the original 65 essay can also derive different approaches to the same rationale. Some scientists claim that the marginal transistor cost or transistor cost plateau cannot be left aside while exposing Moore's law. Such cost decreases: 1 - over time and 2 - given the optimal amount of transistors per chunk. The figure below taken from the original paper illustrates such behavior:



Consequences & Limitations

On 13 April 2005, Gordon Moore stated in an interview that the law cannot be sustained indefinitely: "It can't continue forever. The nature of exponentials is that you push them out and eventually disaster happens." He also noted that transistors would eventually reach the limits of miniaturization at atomic levels.

An overview of its limitations and consequences are:

1. The ensuing speed of technological change

    Technological change is a combination of more and of better technology. A recent study in the journal Science shows that the peak of the rate of change of the world's capacity to compute information was in the year 1998, when the world's technological capacity to compute information on general-purpose computers grew at 88% per year.

2. Transistor count versus computing performance

   The exponential processor transistor growth predicted by Moore does not always translate into exponentially greater practical CPU performance. Viewed even more broadly, the speed of a system is often limited by factors other than processor speed, such as internal bandwidth and storage speed, and one can judge a system's overall performance based on factors other than speed, like cost efficiency or electrical efficiency.

3. Importance of non-CPU bottlenecks

   As CPU speeds and memory capacities have increased, other aspects of performance like memory and disk access speeds have failed to keep up. As a result, those access latencies are more and more often a bottleneck in system performance, and high-performance hardware and software have to be designed to reduce their impact.

4. Parallelism and Moore's law

   Parallel computation has recently become necessary to take full advantage of the gains allowed by Moore's law. Now, to manage CPU power dissipation, processor makers favor multi-core chip designs, and software has to be written in a multi-threaded or multi-process manner to take full advantage of the hardware.

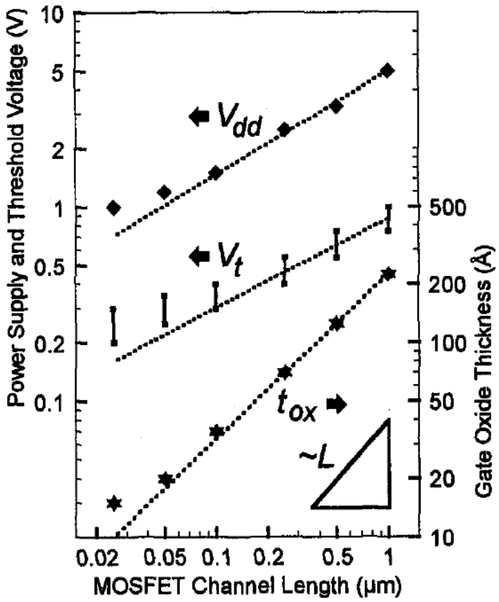
5. Temperature (Thermal limits)

   As a direct consequence of High Power density, any chip that exceeds a maximum operating temperature will be rapidly damaged beyond use. However, even operating close to a temperature limit can increase the degradation of transistors and wires on the chip - a problem that gets worse as they are increasingly miniaturised. The overall effect is to force the chip with many cores to shut most of these cores down to avoid overheating, and to move work from core to core to spread the heat across the chip.

6. Threshold voltage

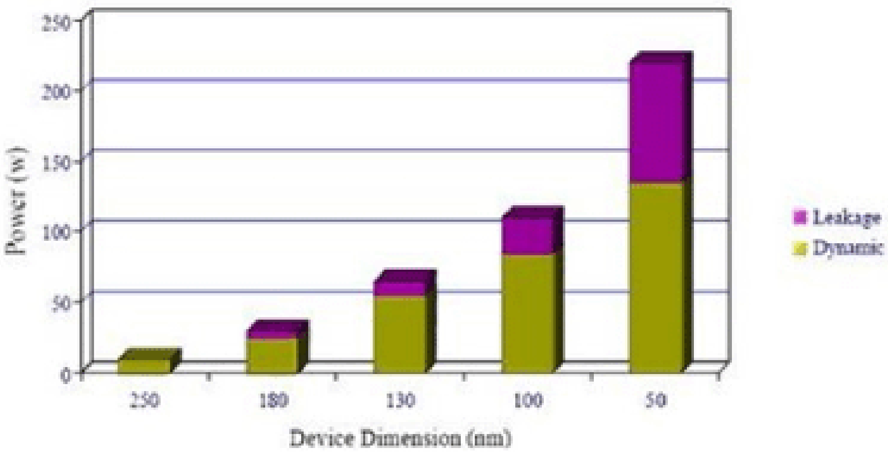
   There does not seem to be any fundamental physical limitation that would prevent Moore’s Law from characterizing the trends of integrated circuits. However, sustaining this rate of progress is not a straightforward achievement.

   Sub-threshold non-scaling and standby power limitations bound the threshold voltage to a minimum of 0.2 V at the operating temperature. Thus, a significant reduction in performance gains is predicted below 1.5 V due to the fact that the threshold voltage decreases more slowly than the historical trend, leading to more aggressive device designs at higher electric fields.



7. Leakage Power

Leakage power reduction is essential in sustaining the scaling of integrated circuits. While lowering of the threshold voltage leads to significant increase in sub-threshold leakage current, the increase in gate tunneling leakage current is caused by thinner gate oxides. While scaling improves transistor density, functionality, and higher performance on chip, it also results in power dissipation increase.



8. Lower Power Consumption

The speed of a circuit depends linearly on the supply voltage. The idea in dynamic voltage scaling is that during times when the circuit is not needing high performance, both its clock frequency and supply voltage can be scaled down. Because dynamic power dissipation depends on the square of the supply voltage and linearly on the frequency $P = f\,CV^2$, if both the supply voltage and frequency are scaled down, there is a cubic reduction in power consumption.