Technology Brief 2: Moore's Law and Scaling

In 1965, Gordon Moore, co-founder of Intel, predicted that the number of transistors in the minimum-cost processor would double every two years (initially, he had guessed they would double every year). Amazingly, this prediction has proven true of semiconductor processors for 40 years, as demonstrated by Fig. TF2-1.

In order to understand ***Moore's Law***, we have to understand the basics about how transistors work. As we will see later in Section 3-7, the basic switching element in semiconductor microprocessors is the ***transistor***. All of the complex components in the microprocessor (including logic gates, arithmetic logic units, and counters) are constructed from combinations of transistors. Within a processor, transistors have different dimensions depending on the component's function; larger transistors can handle more current, so the sub-circuit in the processor that distributes power may be built from larger transistors than, say, the sub-circuit that adds two bits together. In general, the smaller the transistor, the less power it consumes and the faster it can switch between binary states (0 and 1). Hence, an important goal of a circuit designer is to use the smallest transistors possible in a given circuit. We can quantify transistor size according to the smallest drawn dimension of the transistor, sometimes called the ***feature size***. In the Intel 4004, for example, the feature size was approximately 10 μm, which means that it was not possible to make transistors reliably with less than 10-μm features drawn in the CAD program. In modern processors, the feature size is 0.065 μm or 65 nm. (Remember that 1 nm $= 10^{-9}$ m.)

The questions then arise: How small can we go? What is the fundamental limit to shrinking down the size of a transistor? As we ponder this, we immediately observe that we likely cannot make a transistor smaller than the diameter of one silicon or metal atom (i.e., $\sim$ 0.2 to 0.8 nm). But is there a limit prior to this? Well, as we shrink transistors down to the point that they are made of just one or a few atomic layers ($\sim$ 1 to 5 nm), we run into issues

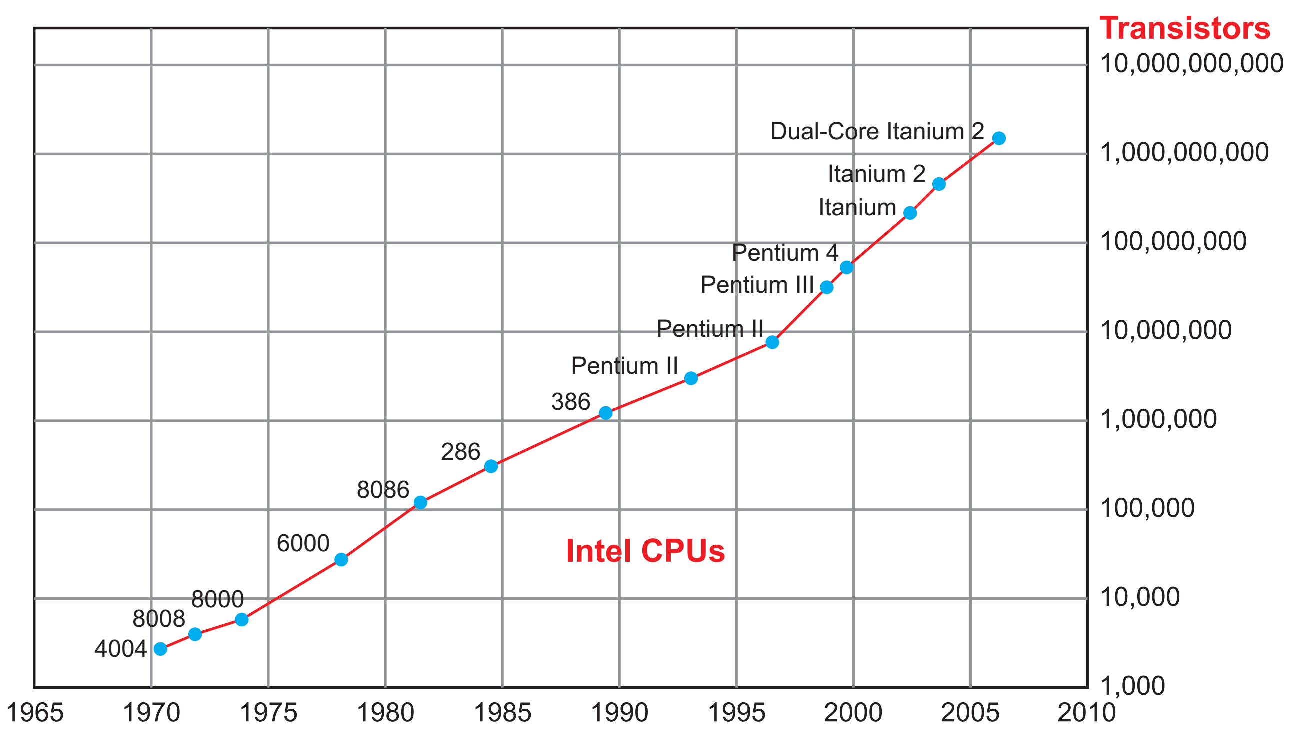
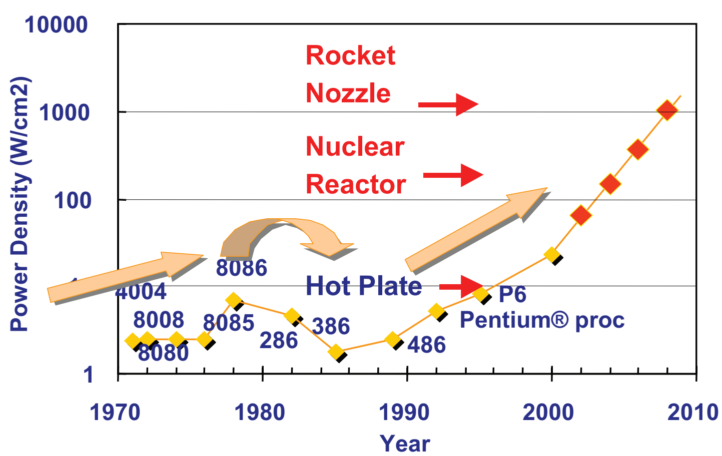


Figure TF2-1: Moore's Law predicts that the number of transistors per processor doubles every two years.

related to the stochastic nature of quantum physics. At these scales, the random motion of electrons between both physical space and energy levels becomes significant with respect to the size of the transistor, and we start to get spurious or random signals in the circuit. There are even more subtle problems related to the statistics of yield. If a certain piece of a transistor contained only 10 atoms, a deviation of just one atom in the device (to a 9-atom or an 11-atom transistor) represents a huge change in the device properties! (Can you imagine your local car dealer telling you your sedan will vary in length by ± 10 percent when it comes from the factory!?) This would make it increasingly difficult to economically fabricate chips with hundreds of millions of transistors. Additionally, there is an interesting issue of heat generation: Like any dissipative device, each transistor gives off a small amount of heat. But when you add up the heat produced by 100 million transistors, you get a very large number! Figure TF2-1 compares the power density (due to heat) produced by different processors with the heat produced by rocket engines and nuclear reactors.

None of these issues are insurmountable. Challenges simply spur driven people to come up with innovative solutions. Many of these problems will be solved, and in the process, provide engineers (like you) with jobs and opportunities. But, more importantly, the minimum feature size of a processor is not the end goal of innovation: It is the means to it. Innovation seeks simply to make *increasingly powerful* processors, not smaller feature sizes. In recent years, processor companies have lessened their attempts at smaller, faster processors and started lumping more of them together to distribute the work among them. This is the idea behind the dual and quad processor cores that power the computers of the last few years. By sharing the workload among various processors (called ***distributed computing***) we increase processor performance while using less energy, generating less heat, and without needing to run at warp speed. So it seems, as we approach ever-smaller features, we simply will transition into new physical technologies and also new computational techniques. As Gordon Moore himself said, "It will not be like we hit a brick wall and stop."



	Light Bulb	Integrated Circuit
Power dissipation	100 W	50 W
Surface area	106 cm^2 (bulb surface area)	1.5 cm^2 (die area)
Heat flux	0.9 W/cm^2	33.3 W/cm^2

Figure TF2-2: The power density generated by an IC in the form of heat is approaching the densities produced by a nuclear reactor. (Courtesy of Jan Rabey, University of California, Berkeley.)