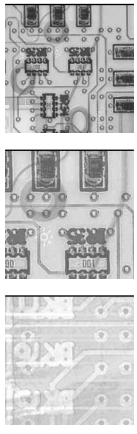


CHAPTER 7

Moore's Law at 40

Gordon E. Moore

Following a paper that I wrote in 1965 and a speech that I gave in 1975, the term "Moore's law" was coined as a name for a type of prediction that I had made. Over time, the term was used much more broadly, referring to almost any phenomenon related to the semiconductor industry that when plotted on semilog graph paper approximates a straight line. In more recent years, Moore's law has been connected to nearly any exponential change in technology. I hesitate to focus on the history of my predictions, for by so doing I might restrict the definition of Moore's law. Nevertheless, in my discussion, I will review the background to my predictions, the reasoning behind them, how these predictions aligned with actual industry performance, and why they did. I will close with a look forward at the future prospects for the prediction.

Overview

Moore's law is really about economics. My prediction was about the future direction of the semiconductor industry, and I have found that the industry is best understood through some of its underlying economics. To form an overall view of the industry, it is useful to consider a plot of revenue versus time. As Figure 1 indicates, the semiconductor industry has been a strong growth industry: it has grown a hundredfold during Intel's existence. However, from my point of view, this plot of revenue growth really underestimates the true rate of growth for the industry.

I prefer a manufacturing viewpoint, analyzing the industry from the perspective of the products we have made. I started with this approach several years ago, looking at the worldwide production of all semiconductor devices, estimating the number of transistors in these devices, and looking at the growth in the total number of transistors shipped in working electronic devices (Figure 2).

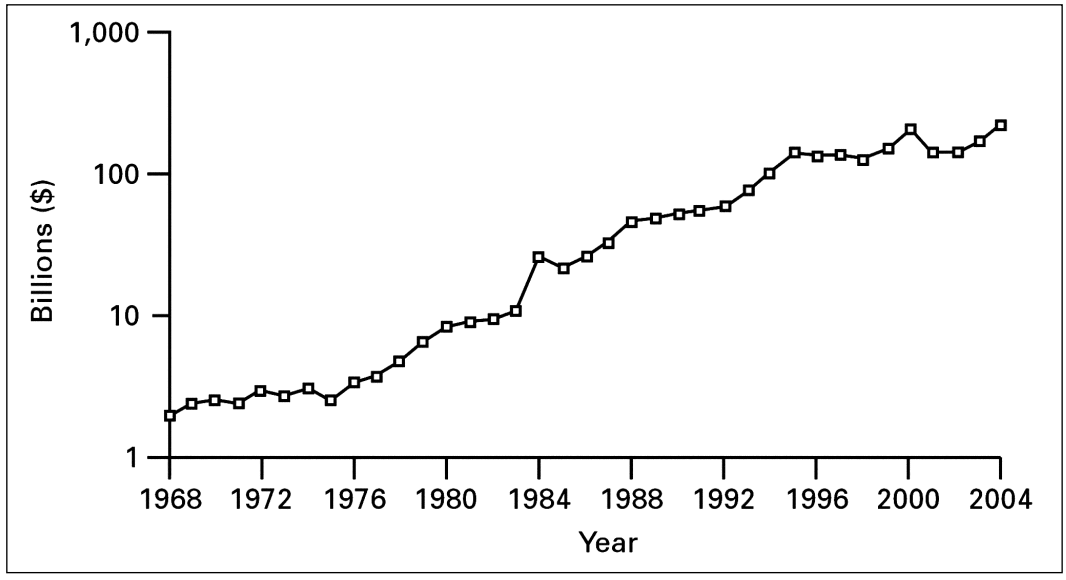


FIGURE 1. Global semiconductor industry revenues (1968–2004). Source: Intel/WSTS, May 2005.

This is *rapid* growth! In fact, there was even a period during the 1970s when the industry was *more than doubling* the total number of transistors ever made every year, so that more electronics were built each year than existed at the beginning of the year. The pace has slowed recently but is still on a good growth curve. Interestingly, there are no bumps and wiggles in this transistor output curve (Figure 2) as there are in the plot of revenue over time (Figure 1).

Transistor output has steadily expanded. Today we have reached over 10^{18} transistors a year. That is a hard number to contemplate. Patrick Gelsinger of Intel estimates that present transistor output equals the number of grains of rice produced globally each year. Over the years, I have used a variety of similar comparisons. At one stage, Edward O. Wilson, the well-known naturalist at Harvard, had estimated that there were perhaps 10^{16} to 10^{17} ants on earth. In the early 1990s, then, the semiconductor industry was producing a transistor for every ant. Now, the poor little ant has to carry a hundred of them around if he is going to get his share.

I have also estimated that the total number of printed characters produced globally every year—including all newspapers, magazines, photocopies, and computer printouts—is between 10^{17} and 10^{18}. Today, the semiconductor industry makes more transistors than the world's output of printed characters, and we sell them for less. This cost dimension is the key factor. To make this dimension clear, dividing annual revenue by transistor output provides a plot of the average price of a transistor (Figure 3).

Today, the cost of an average transistor has dropped to about a hundred nano-dollars. For transistors in dynamic random access memories (DRAMs), the cost is less. For transistors in microprocessors, the cost is a bit more. This cost reduction curve represents an amazing rate of change, and it is the basis for the large impact the semiconductor industry has had in making electronics so much more available.

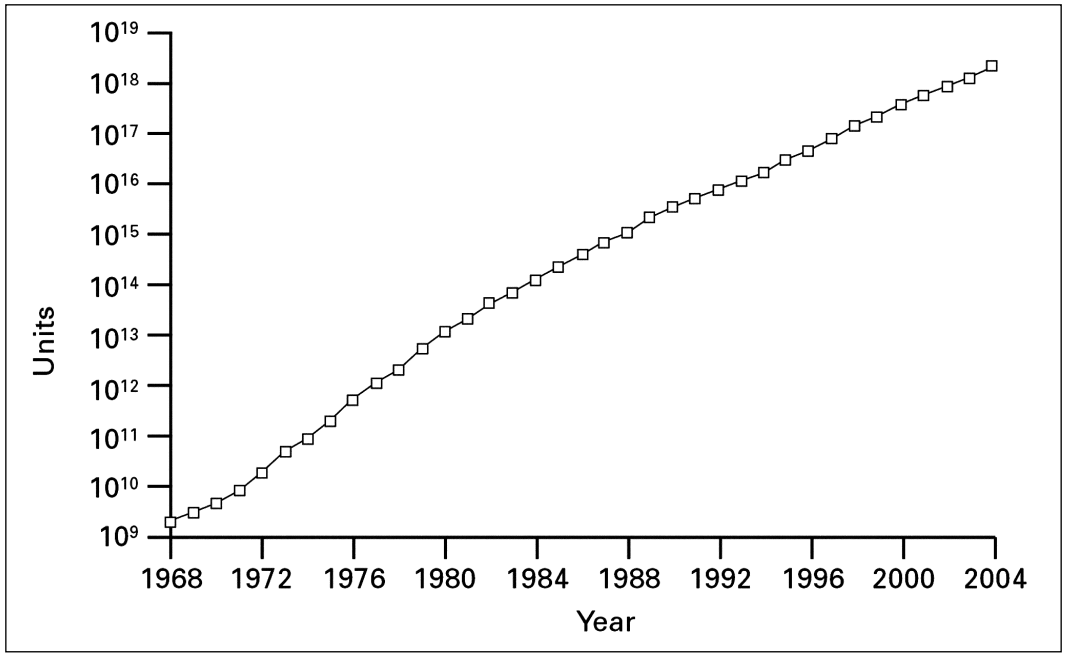


FIGURE 2. Total number of transistors shipped by the semiconductor industry (1968–2004). Source: Intel/WSTS, May 2005.

BACKGROUND

My interest in thinking about these kinds of plots dates back at least to 1964, when I was writing the paper that contains the first version of what became known as Moore's law. I was not alone in making projections. At a conference in New York City that same year, the IEEE convened a panel of executives from leading semiconductor companies: Texas Instruments, Motorola, Fairchild, General Electric, Zenith, and Westinghouse. Several of the panelists made predictions about the semiconductor industry. Patrick Haggerty of Texas Instruments, looking approximately ten years out, forecast that the industry would produce 750 million logic gates a year. I thought that was a huge number, and puzzled, "That is really perceptive. Could we actually get to something like that?" Harry Knowles from Westinghouse, who was considered the wild man of the group, said, "We're going to get 250,000 logic gates on a single wafer." At the time, my colleagues and I at Fairchild were struggling to produce just a handful. We thought Knowles's prediction was ridiculous. C. Lester Hogan of Motorola looked at expenses and said, "The cost of a fully processed wafer will be $10."

When you combine these predictions, they make a forecast for the entire semiconductor industry. If Haggerty were on target, the industry would produce 750 million logic gates a year. Using Knowles's "wild" figure of 250,000 logic gates per wafer meant that the industry would only use 3,000 wafers for this total output. If Hogan was correct, and the cost per processed wafer was $10, that would mean that the total manufacturing cost to produce the yearly output of the semiconductor industry would be $30,000! Somebody was wrong.

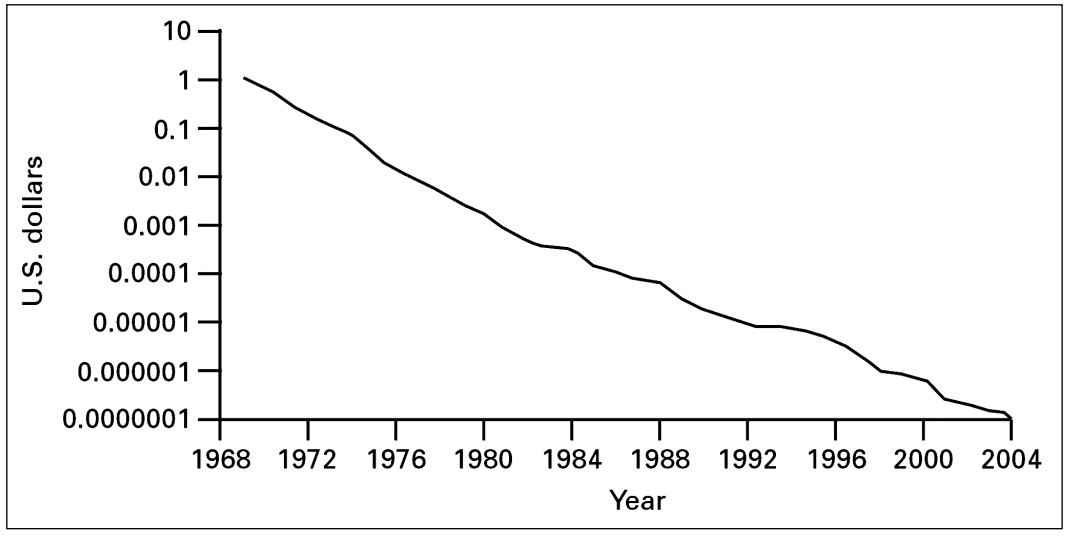


Figure 3. Average price of a transistor (1968–2004). Source: Intel/WSTS, May 2005.

As it turned out, the person who was the "most wrong" was Haggerty, the panelist I considered the most perceptive. His prediction of the number of logic gates that would be used turned out to be a ridiculously large underestimation. On the other hand, the industry actually achieved what Knowles foresaw, while I had labeled his suggestion as the ridiculous one. Even Hogan's forecast of $10 for a processed wafer was close to the mark, if you allow for inflation and make a cost-per-square-centimeter calculation. Today, the industry does not "do" $10 wafers, but we use wafers that are very much larger than the one-inch wafers that Hogan was talking about in 1964. Using a cost-per-area calculation, Hogan's prediction really was in the ballpark.

1965 Paper

The suggestions of Haggerty, Knowles, and Hogan reflected the general views of the semiconductor industry around the time I was working on my 1965 projection. *Electronics* magazine had asked me to forecast what would happen in the next ten years to the semiconductor components industry. This was very early into the semiconductor integrated circuits era. The primary user of integrated circuits was the military. Integrated circuits were too expensive for use in commercial systems, costing significantly more than the equivalent circuit built out of individual components. Potential customers also had a variety of other objections. They were concerned with ensuring the reliability of integrated circuits when you could no longer measure the parameters of each element—the transistors, the resistors, and so on. With the integrated circuit, only the reliability of the whole device could be measured.

Critics also argued that, with integrated circuits, our yields would vanish. Yield is a critical concept in the semiconductor industry, meaning the percentage of acceptable devices actually produced on a wafer out of the total number of potential working devices. These critics knew that, at the time, we made transistors at yields in the

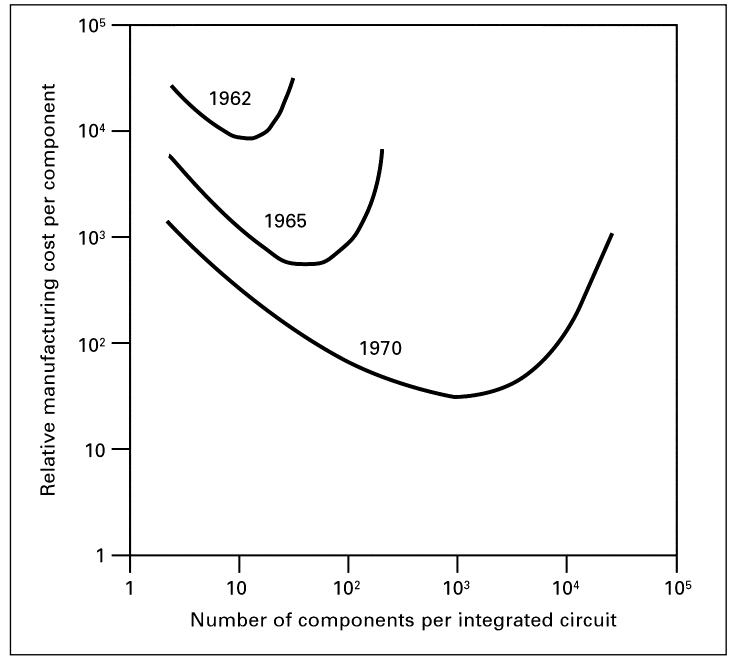


FIGURE 4. Manufacturing cost per component versus number of components per integrated circuit, from 1965 *Electronics* article.

10 to 20 percent range. They argued that for a circuit with eight transistors, taking 0.2 to the eighth power, you got an awfully small number for yield. Moreover, the integrated circuit's performance was below that obtained by using individual components, because of parasitics and other factors in the integrated circuits. The low yields argument reflected the fact that potential purchasers did not think that they would be able to get the actual supplies that they would need.

From my different perspective, as the director of the research laboratory at Fairchild Semiconductor, I could see some of the major developments that were coming. In my article in *Electronics*, I wanted to send the message that, looking forward, integrated circuits were going to be the route to significantly cheaper products. That was the principle message I was after.

To sharpen this economics message I analyzed the cost per component versus circuit complexity for integrated circuits. I plotted and projected this relationship in a series of curves (Figure 4), which suggested that, at a given time, there was a minimum manufacturing cost per component that was achieved by using a particular degree of complexity. At a lower complexity, one was not taking full advantage of the processing technology, and therefore costs increased. Beyond the optimal complexity point, the yields also dropped considerably, and hence costs increased. Importantly, I saw that the minimum cost per component point had been coming down quickly over several years, as the manufacturing technology improved. From this observation, I took my few data points and plotted a curve, extrapolating out for the ten years I had been asked to predict (Figure 5).

The last data point in this graph, the 1965 point, represents a device that we had at the time in the Fairchild laboratory with approximately sixty components, which

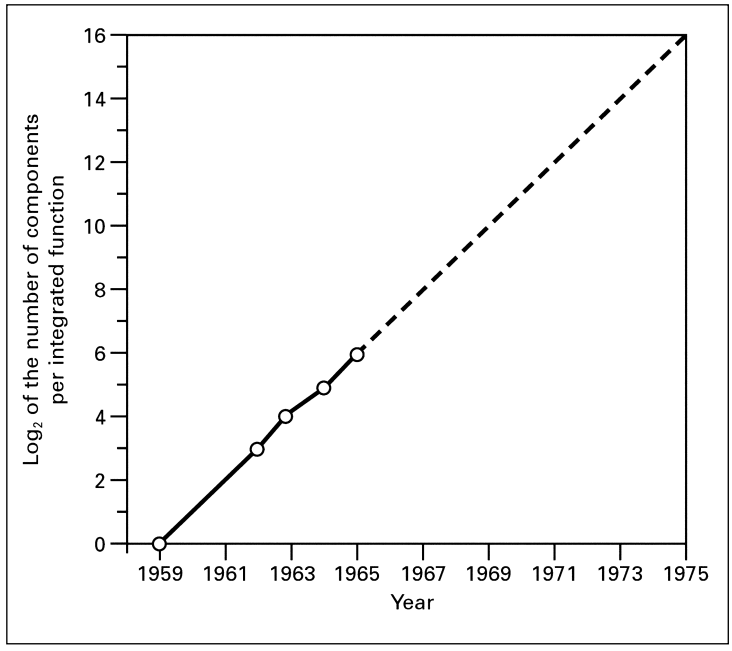


Figure 5. 1965 projection of number of components on an integrated circuit.

the company was going to introduce soon. The first data point on the graph represents the original planar transistor that we introduced in 1959. The planar transistor was the starting point for the integrated circuit's basic technology, so it deserved to be on this curve also.

Between the point for our 1959 planar transistor and our 1965 new device with sixty components were several points representing the Micrologic family of integrated circuits that Fairchild Semiconductor had introduced. Plotting these points using a log-base-two scale, I saw that the points fell closely along a line representing a doubling of complexity every year through 1965. To make my requested prediction, I simply extrapolated this same line for another decade, thereby predicting a thousandfold increase in complexity. The rather obscure log-base-two scale that I used on the vertical element of the graph made it a bit difficult to see that I was extrapolating from sixty to sixty thousand components. Nevertheless, the curve did seem to make sense with the existing data and some people who looked at this line and said, "That's a reasonable extrapolation."

1975 Speech

I never expected my extrapolation to be very precise. However, over the next ten years, as I plotted new data points, they actually scattered closely along my extrapolated curve (Figure 6). At the end of these ten years, I gave a talk at the IEEE International Electron Devices Meeting to show what had actually happened since my 1965 prediction, to analyze how the semiconductor industry had accomplished that degree of progress, and to make a prediction for upcoming years. To do this, I broke down the complexity curve into several contributing factors (Figure 7).

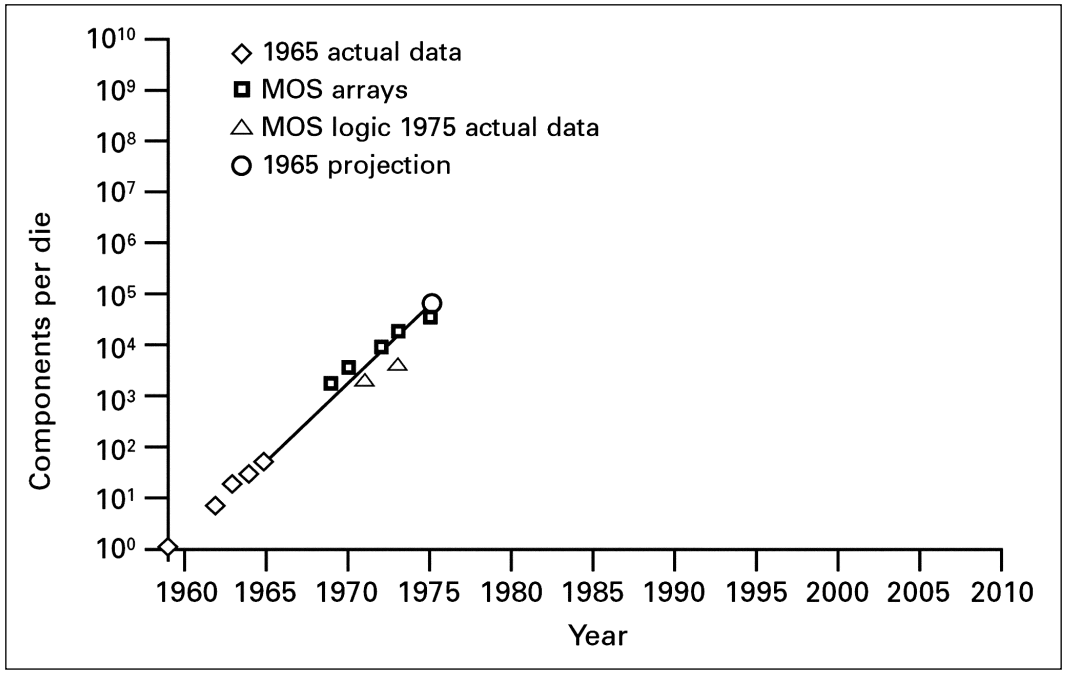


Figure 6. Integrated circuit complexity (1959–1975). Source: Intel.

One of the factors I named the "die size contribution." In the semiconductor industry, the term *die* is used for the area on a processed wafer that contains a single device. After a wafer is completely processed, the wafer is cut in to many "dice," each containing a single integrated circuit. The "die size contribution" factor in Figure 7 reflects how the semiconductor industry was making larger devices (with increased die sizes) and therefore had more area onto which to put components. A second, slightly larger contribution to the complexity increase was "dimension reduction." This was the shrinking of component dimensions, which led to an increase in the density. Multiplying these two contributions results in a curve that represents the combined effect on complexity growth of "die size and dimensions." This combined contribution was responsible for more than half of the progress that the industry had made on the complexity curve, but there remained a very considerable element that came from some other factor. On the graph, I labeled this factor the "contribution of device and circuit cleverness." This factor I identified with squeezing waste space out of the chip, getting rid of isolation structures and a variety of other things.

The last data point that I had for my 1975 talk was the component count for a charge-coupled device (CCD) memory that we were working on at Intel. With CCDs, the active areas are as close to one another as possible. There was no room left to squeeze. As a result, my argument was that, sometime soon after 1975, we were going to lose this "cleverness" factor, a factor that had contributed nearly half of the progress on the complexity curve. For simplicity's sake, I rounded this contribution to half of the total. With this loss, then, the complexity curve was going to change from doubling

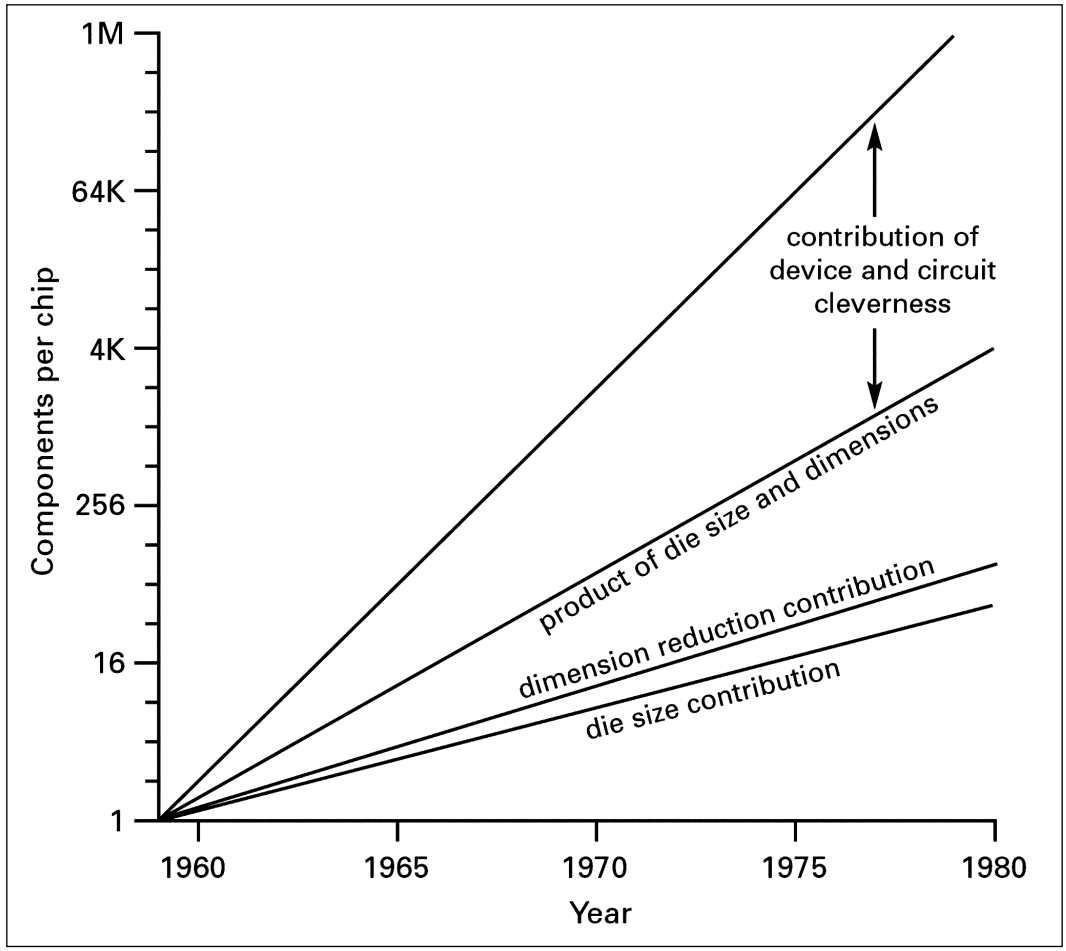


FIGURE 7. Resolution of complexity increase into contributing factors. Source: Intel.

every year to doubling every two years, and we would have to rely only on the two factors of increased die size and finer dimensions.

I knew too much. I was looking at our CCD memories, and the device we had nearly ready to go into production was 32 kilobits. We also had a 64 kilobit CCD memory coming along, and a 256 kilobit not too far behind the 64. I believed that those CCD memories were going to keep us doubling every year for another few years. I thought, "Well, I'm not going to change the slope right away. I'll give the rate a five-year rollover time" (Figure 8).

What I did not realize was that CCD memories were going to be a disaster. The same property that makes CCDs good imaging devices in such products as digital cameras makes them terrible memory devices: they are very sensitive to radiation. An alpha particle generated out of the packaging material for a CCD memory can completely wipe out several bits. This was a major problem. These were non-repeatable errors, with occasional random losses of bits of information, and we started to find

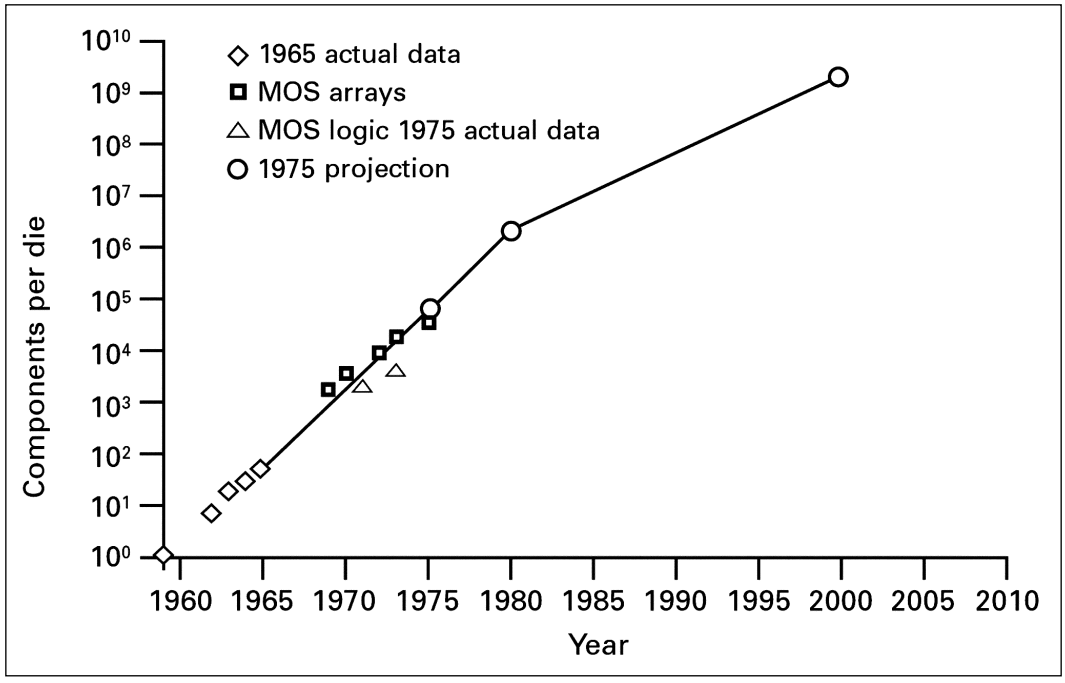


FIGURE 8. Integrated circuit complexity, 1975 projection. Source: Intel.

them in DRAMs as well. Our CCDs turned out to be very valuable for studying this alpha particle phenomenon, finding out what the problem was and getting to some solutions. However, we did not introduce any CCD memories after our first CCD memory product. The net result of the CCD experience was that, while I had predicted a five-year hiatus before the complexity slope would change, in fact the slope changed right away. Had I started the new slope, representing a doubling every two years, in 1975 instead of after the five-year rollover, my prediction would have been much more accurate. But I didn't (Figure 9).

Other "Exponentials": Wafer Size

I made a number of other extrapolations; some were just to demonstrate how ridiculous it is to extrapolate exponentials. In my 1975 talk, I described the contribution of die size increase to complexity growth and wrote: "In fact, the size of the wafers themselves have grown about as fast as has die size during the period under consideration and can be expected to continue to grow." One of my colleagues at Intel caught wind of that extrapolation and let me know that the 57-inch wafer predicted for the year 2000 did not quite come to pass.

Nonetheless, wafer size has grown dramatically, and I have to say that I am agreeably surprised by the size of the 300-mm wafers that the semiconductor industry uses today. To make the first planar transistor at Fairchild in 1959, we used three-quarter-inch wafers. In fact, one of my contributions to the semiconductor industry at that time was to show that if wafer size went above three-quarters of an

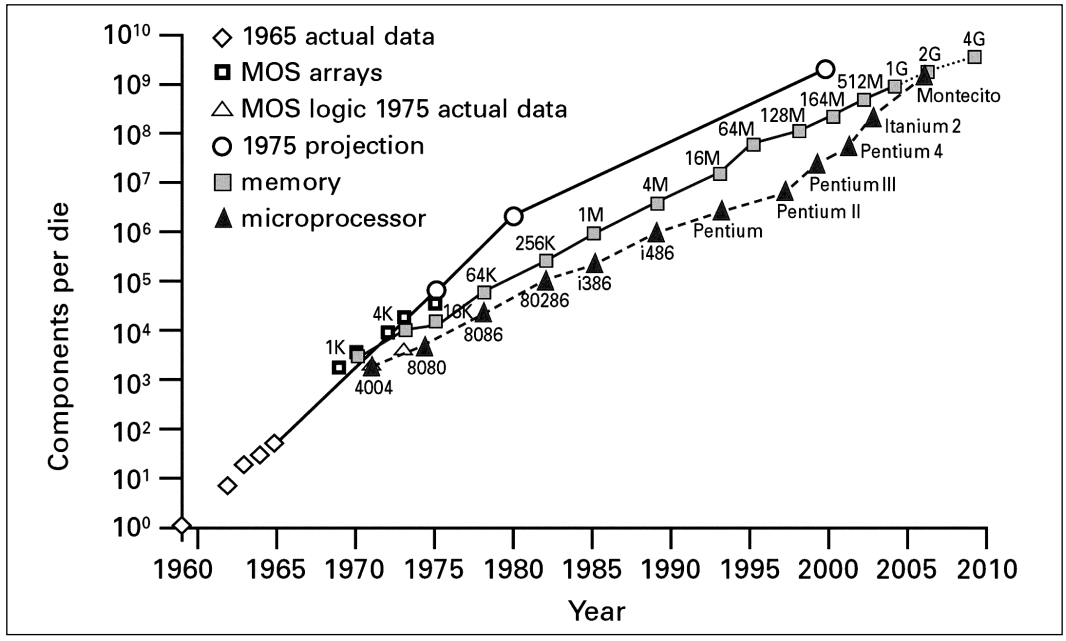
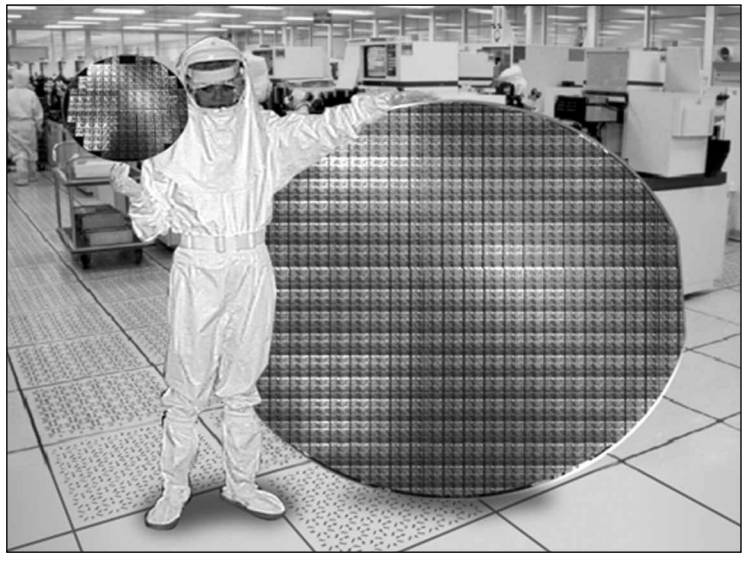
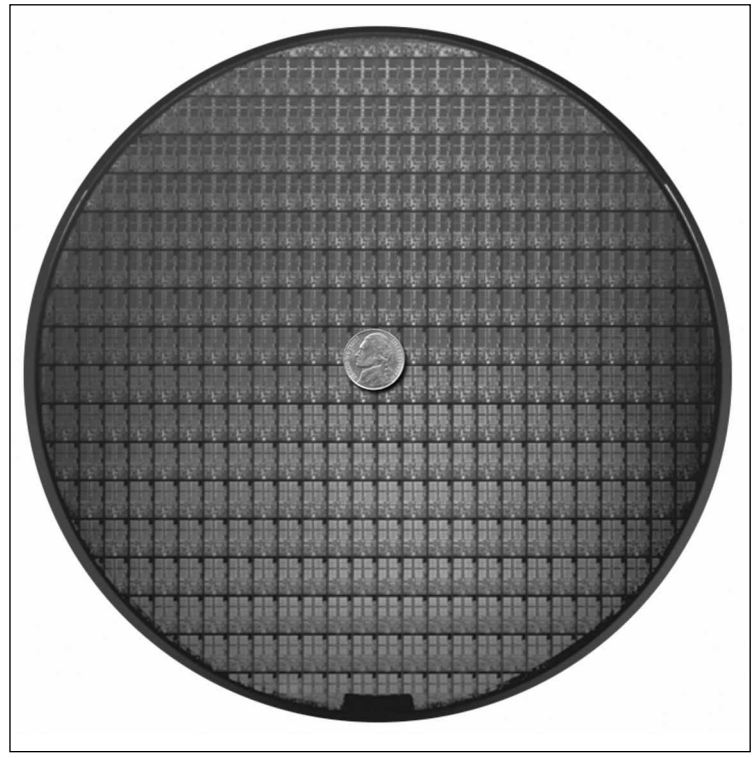


FIGURE 9. Integrated circuit complexity, actual data compared with 1975 projection. Source: Intel.

inch then yields would drop to zero because the quality of the material deteriorated so rapidly. The amount of technology that has gone into growing and slicing single crystals, with rapidly expanding diameters, is fantastic. Our next wafer size, 450 mm, will be the size of the kind of pizzas that can be bought at Price Club—about 18 inches. Those are monster pizzas.



A digitally manipulated photograph showing the ficticious 57-inch wafer. Courtesy of Intel.



A 300mm wafer with U.S. nickel (approximate size of 1959 standard ¾-inch wafer) for scale. Courtesy of Intel.

Other "Exponentials": Interconnections

Just as wafer size has grown dramatically over the past forty years, so too has the "complexity" of interconnection technology, the system of metal pathways that connect the many components of an integrated circuit. As with wafer development, this growth in interconnection technology has required an impressive amount of materials innovation. The intricacy of contemporary interconnects can be visualized by examining an electron micrograph of the copper interconnections for a device where all the insulating regions have been dissolved away in order to highlight the complexity of the interconnection system. The crystal grains of the copper in the top level interconnects are visible. Moving down the levels, the interconnects become smaller still.

A more modern process technology, the 90 nanometer generation, is now being introduced into production. It has seven layers of metal interconnections, separated by low dielectric-constant insulators, with a very thin layer of active silicon buried at the very bottom level (Figure 10). This is an amazingly complex structure that we have evolved.

There are many materials involved in a contemporary transistor (Figure 11, left): nickel silicide in some areas, silicon nitride in others. Moreover, we use strained silicon in these devices, producing the straining in one direction by adding germanium to the silicon, using a silicon carbon-nitride mixture for straining in the other direction. In some devices, the silicon is compressed, in others it is expanded, depending

A region of copper interconnects for an Intel logic device from 2001. Courtesy of Intel.

on whether it is the mobility of holes or electrons that one is trying to increase. Exotic compounds like tantalum nitride are used as barriers to stabilize the performance of the copper interconnections (Figure 11, right). The simple, old, silicon-oxide-aluminum system for semiconductor devices has been replaced by a much more complex system of materials.

DECREASING DIMENSIONS

These complex material systems have become necessary for maintaining one of the principal factors that produces the complexity curve: the continual decrease in the dimensions of components.

The top curve in Figure 12 represents what Intel has actually achieved in reducing feature size. This progress has been along a fairly constant slope, with a new generation of technology introduced about every three years, and each generation doubling the component density. This is the general developmental algorithm the industry has followed. In the beginning the industry did not analyze the pace explicitly. Increasingly we recognized that there was a definite pattern in what was happening, and we deliberately tried to continue it. With the advent of the International Technology Roadmap

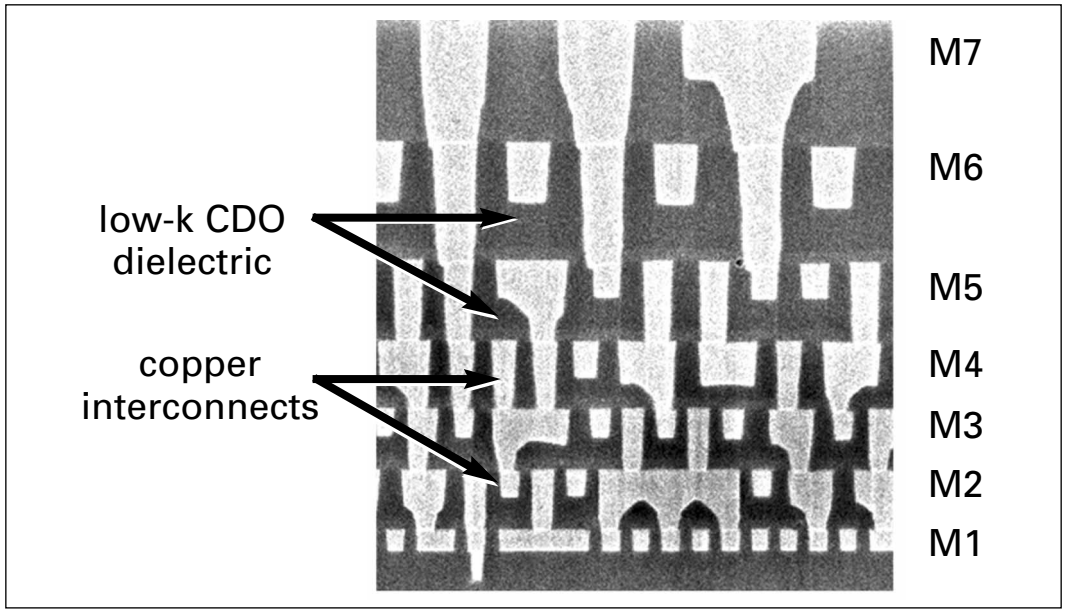


FIGURE 10. Side view of seven layers of metal interconnects in the 90-nm process technology. Source: Intel.

for Semiconductors produced by the Semiconductor Industry Association, the goal of continuing this slope has been formalized, with a new generation of process technology coming on line every three years. That was a reasonable goal to set.

The nature of the semiconductor business is such that companies have to be at the leading edge of the technology to be competitive. The reason is that the semiconductor industry is really selling real estate. The price of that real estate has been nearly constant for as long as I have been in the business: on the order of a billion dollars an acre. It used to be a few billion dollars an acre, and microprocessors now are about

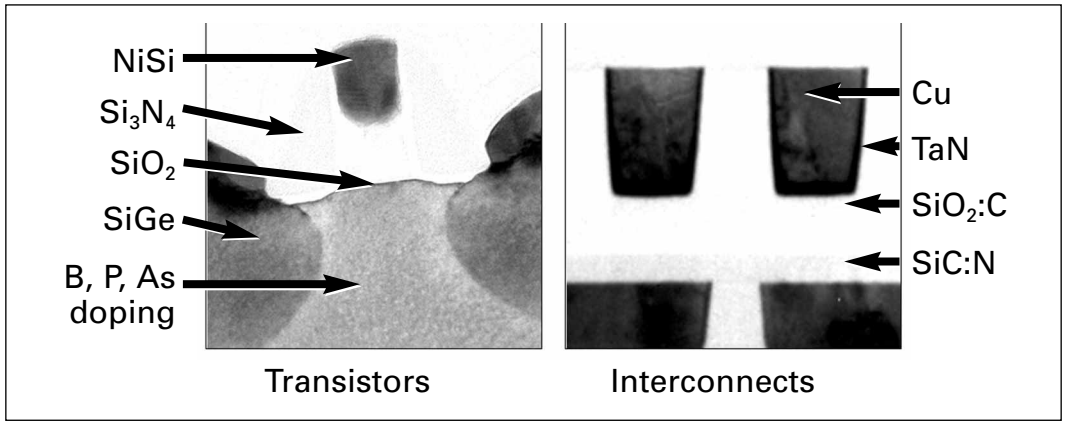


FIGURE 11. Micrographs of a transistor and interconnects created using the 90-nm process. Source: Intel.

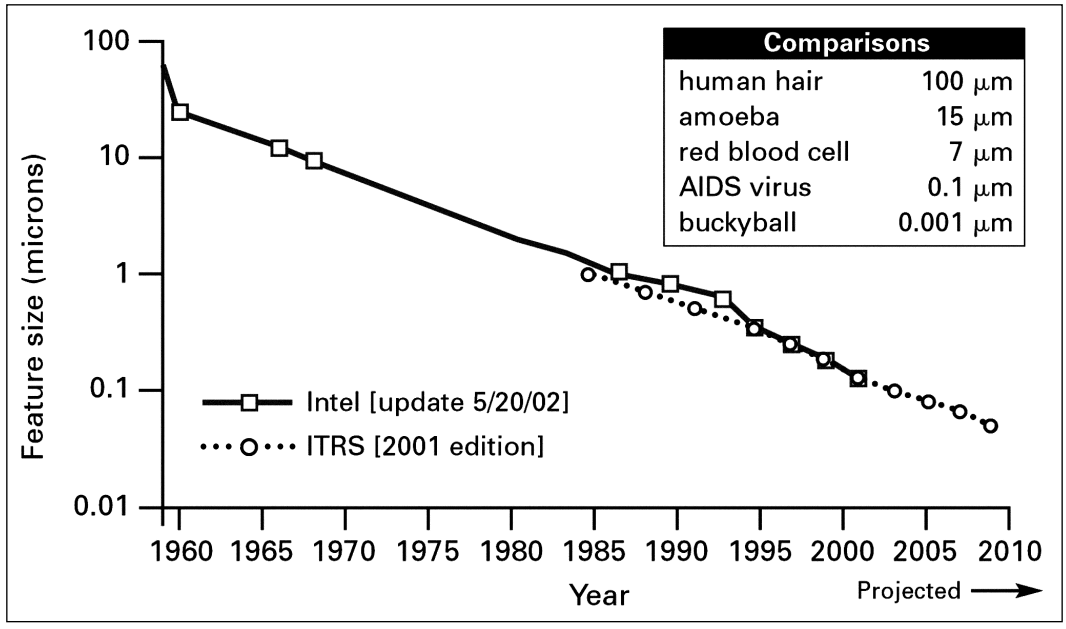


FIGURE 12. Decreasing dimensions: minimum feature size of integrated circuits (1963–2010). Source: Intel, post 1996 trend data provided by *SIA International Technology Roadmap for Semiconductors* (ITRS). (ITRS DRAM Half-Pitch vs. Intel "Lithography").

$2 billion an acre. Memory today is about $0.8 billion. On balance, silicon real estate has been steady on the order of a billion dollars an acre. I used to joke that this was why Japanese firms were such formidable competitors: silicon real estate was about the price of land in Tokyo in the 1980s.

As the real estate perspective shows, companies that do not stay near the leading edge of process technology suffer from a cost disadvantage. They are not exploiting the available enhanced densities, thereby are not making the most out of some very expensive real estate. In addition, in the semiconductor industry, the most profitable products have been leading edge devices. If companies do not keep to the leading edge, their products suffer a performance disadvantage. Straying from the most advanced technology, the combination of cost and performance disadvantages is competitively catastrophic.

What happens with this push to the leading edge? What happened, of course, is that we changed the slope (Figure 13). When the industry fully recognized that we were truly on a pace of a new process technology generation every three years, we started to shift to a new generation every two years to get a bit ahead of the competition. As a result, instead of slowing, this trend to smaller and smaller dimensions has actually accelerated as a result of people recognizing the slope. I think that this is a strong example of how awareness of Moore's law-type trends has driven the industry. Everybody recognizes that they have to keep up with this curve, or they will fall behind.

While the complexity curve can be understood intellectually, it is interesting to approach it from a more tangible point of view in order to get a feel for what this

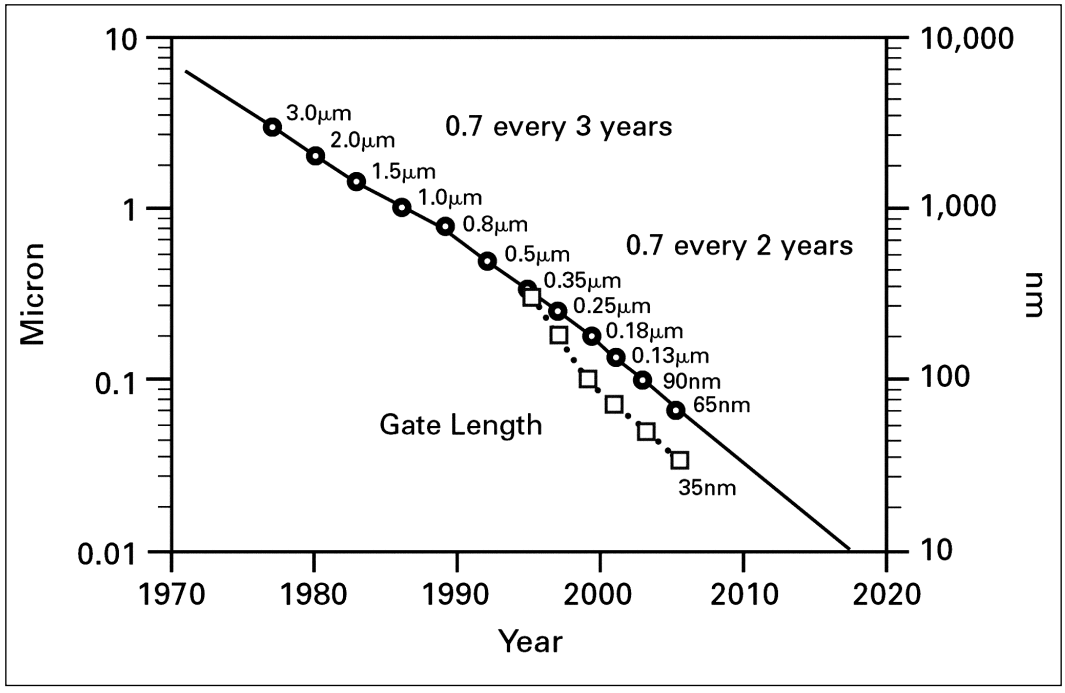


FIGURE 13. Plot of new technology generation introductions (1975–2005). Source: Intel.

pace of change has produced. Figure 14 shows one contact produced on a device using the generation of process technology in 1978. Twenty-four years later, an entire six-transistor memory cell occupies just a fraction of the area of the 1978 vintage single contact. A series of incremental changes, over a sufficient period, results in dramatically different products.

Equipment Costs

Keeping to the complexity curve for semiconductor devices has entailed a corresponding increase in the complexity of semiconductor manufacturing equipment. Modern 193 nanometer exposure systems use argon fluoride excimer lasers, among other things; these are really very sophisticated and complicated pieces of equipment. Nevertheless, to keep on the complexity curve, we are going to have to make a step beyond this equipment before too long. The industry is working on a 13 nanometer, extreme ultraviolet exposure system: essentially more than an order of magnitude decrease in the wavelength. These 13 nanometer machines require completely reflective optics. No materials are fully transparent in this wavelength range, and mirrors are not very good reflectors of these wavelengths either. One of these 13 nanometer systems uses thirty-plus reflections, each with only something like 0.6 or 0.7 reflectivity. This presents a definite challenge. The optical surfaces have to be better than those of the Hubble Space Telescope, in order for us to get the kind of performance that we want.

These equipment challenges have kept the semiconductor industry on another exponential, in the cost of lithography equipment (Figure 15). In fact, you can plot a

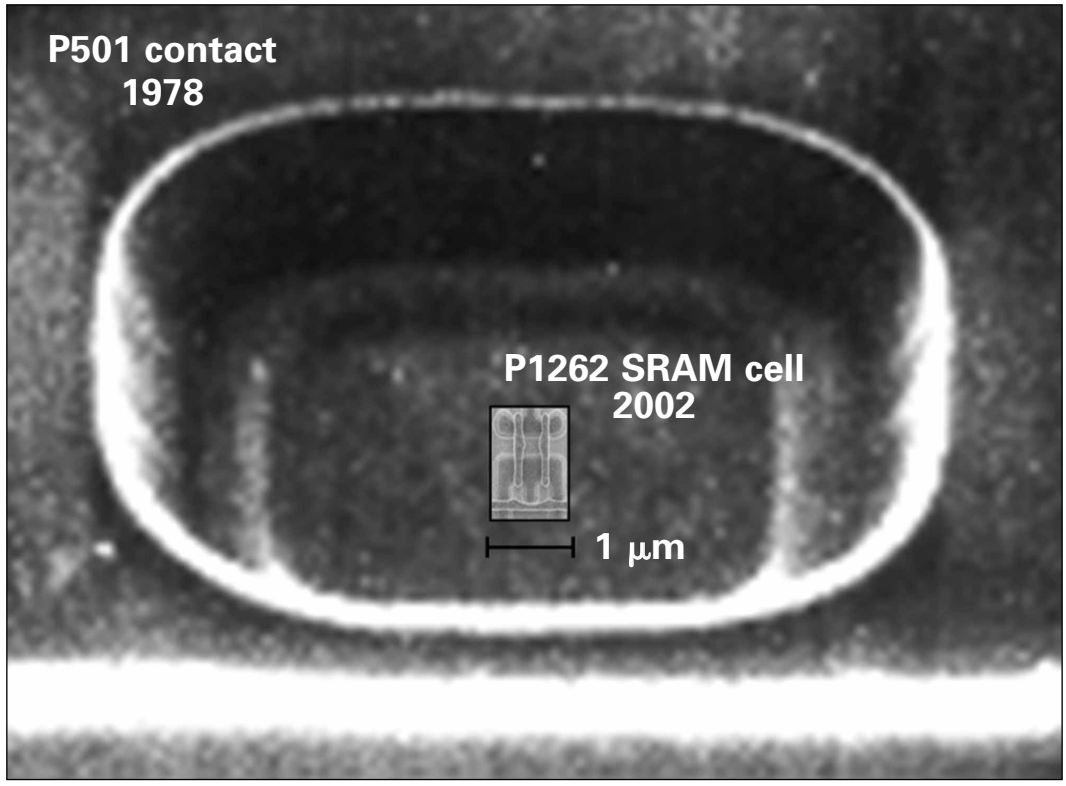


FIGURE 14. Size comparison of a single contact from 1978 with a full SCRAM cell from 2002. Source: Intel.

price exponential for most types of semiconductor manufacturing equipment. This presents an interesting economic challenge. The equipment keeps going up in cost exponentially, but the semiconductor industry is not growing as rapidly anymore. Capital costs are rising faster than revenue. Nevertheless semiconductor companies have to stay on the leading edge of process technology or they suffer cost and performance disadvantages.

Materials Challenges

In addition to the equipment challenges for producing smaller and smaller features, we also encounter materials challenges. As component size decreases, we use thinner and thinner layers of insulators. With this thinness, electrical leakage and other factors are of greater concern. To see how dramatic this issue is, consider a transmission electron micrograph of a gate area on a component produced by the 90 nanometer production process (Figure 16, left).

At the bottom of the pictured area, the individual atoms in the silicon substrate are visible. The middle section, the insulating silicon dioxide layer, is no thicker than a couple of molecular layers. At the top, the gate is again formed by silicon. This leading edge technology has a problem: leakage current because of quantum mechanical

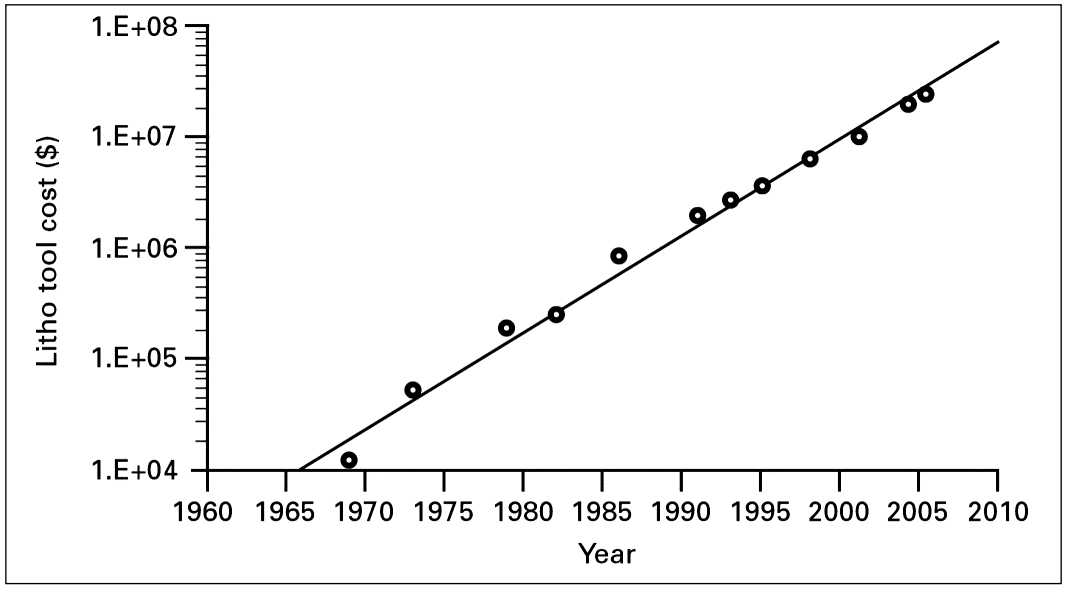


FIGURE 15. Cost of process equipment for photo lithography (1969–2005). Source: Intel.

tunneling through the thin layer. This problem can be minimized if we change to a new insulating material with a high dielectric constant. We have a material that we are working on now that allows us to make this kind of dramatic change (Figure 16, right). With this new insulating material, the capacitance is preserved. In fact, it increases, which means a higher electrical field is transferred to the silicon substrate, resulting in better performance. More remarkably, the leakage current goes down a

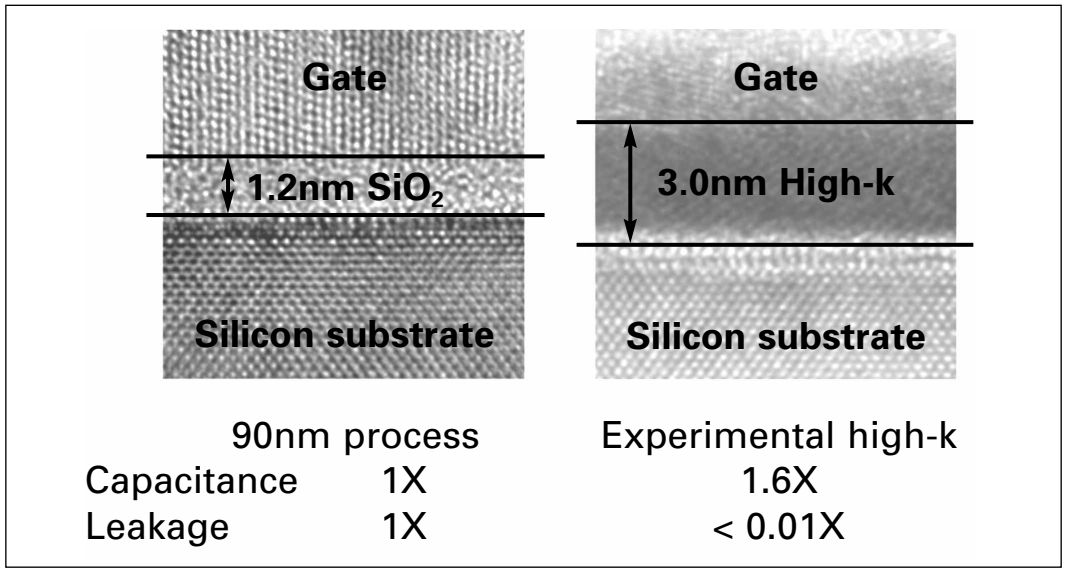


FIGURE 16. Materials challenges: leakage and dielectrics. Source: Intel.

hundredfold. Those are the kinds of changes that new materials allow us to make. They are not easy. A tremendous amount of work went into finding a material with the correct dielectric capabilities that was also stable enough to withstand processing and could tolerate these high electric fields.

LOOKING FORWARD

Having outlined my general approach to understanding the semiconductor industry and having identified some key factors for keeping on the complexity curve, one might ask, "When is it all going to end?" I have been asked that question at least a hundred times this year. The answer is: "Not very soon." I can see the complexity curve lasting for at least as long now as I ever could in the past. I always could see what we were going to do to make the next two or three technology generations happen on the curve. Today, as I speak with the Intel research and development staff members, they are looking out even further. Now I can see what we are going to do for the next four generations, which is further than we have ever been able to look out before. Many advances have to happen to make those future generations occur, but we are confident the problems are going to be solved in time to make it all happen. It is amazing what a group of dedicated scientists and engineers can do, as the past forty years have shown. I do not see an end in sight, with the caveat that I can only see a decade or so ahead.