

## Computers and Semiconductors

Richard N. Langlois

### 10.1. Introduction

At the end of the twentieth century, it became common to talk of the “digital revolution,” a historical phenomenon worthy of its place among the various industrial revolutions of the previous two centuries. Underlying the digital revolution is the technology of the semiconductor, a device that emerged at the century’s half-way point. Although the digital revolution ramifies itself throughout the modern world, notably into telecommunications and consumer goods, its most signal embodiment is the digital computer, a technology born at almost exactly the same point in history.

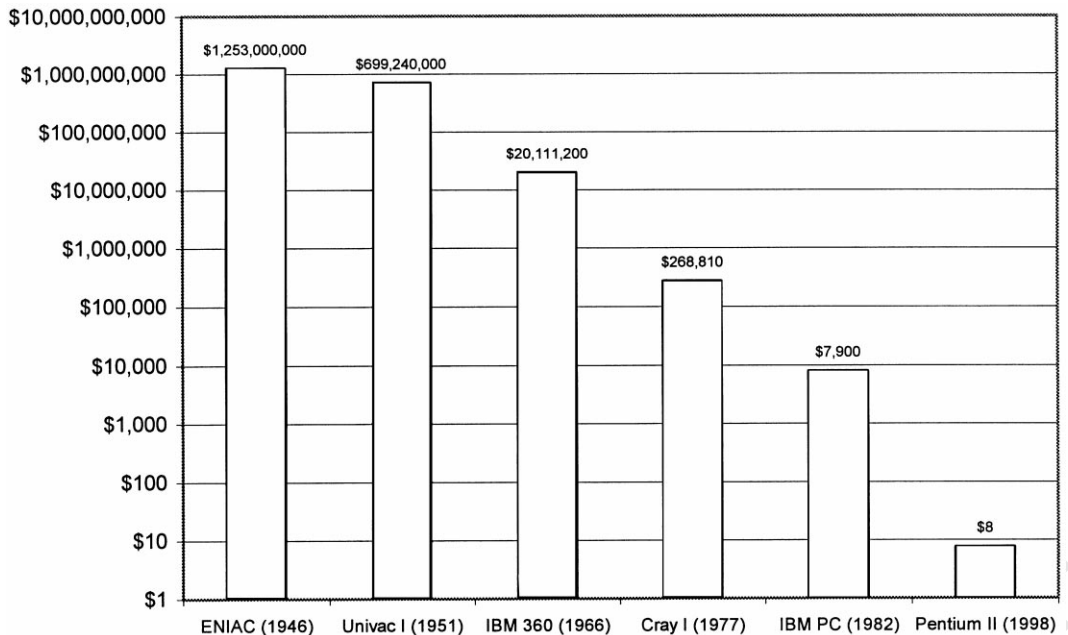
Figure 10.1 suggests why the progress of digital technology appears so revolutionary. One very broad measure of the power of a computer is the number of so-called floating-point operations (like adding together two numbers) a machine can perform in a second. The first truly digital computer, the ENIAC of 1946, cost some U.S.\$750,000 to produce—something like U.S.\$6,265,000 in 1998 dollars—and could perform 5,000 calculations per second. The circa 1998 Pentium II computer on which I am writing this chapter cost about U.S.\$1,500 and can perform 200 million calculations per second. That is about U.S.\$1.25 billion per million floating-point operations per second (MFLOPS) for the ENIAC—and about U.S.\$8 per MFLOPS for the Pentium II. This phenomenal decline is tied to the rapid improvement of the semiconductor technology on which computers now depend.<sup>1</sup>

<sup>1</sup> Data in this paragraph and in Figure 10.1 are from Kurzweil (1999: 320–1).

Using broad strokes, this chapter sketches the intertwined history of these two industries—semiconductors and computers. In so doing, it attempts to shed light on the sources of technological change in these industries and on the complex mechanisms through which that technological change has translated into economic growth.

A distinctive theme in this history will be the emergence and significance of *general-purpose technologies* (GPTs). Such technologies (and their attendant systems of skill and knowledge) typically develop in response to specific technological puzzles or bottlenecks, but they ultimately generate principles and techniques that are applicable to a wide variety of otherwise distinct output sectors of the economy (Bresnahan and Trajtenberg, 1995; Bresnahan and Gambardella, 1998). Nathan Rosenberg (1963) described this process as *technological convergence*. Because what is learned once can be reused many times, technological convergence generates the something-for-nothing effect economists call *increasing returns*, a phenomenon at the heart of economic growth.<sup>2</sup>

<sup>2</sup> By the concept of “returns” economists mean the following. If you double all your inputs and get exactly double the output, that is constant returns to scale. If you double your inputs and get less than double your output, that is diminishing returns to scale, the bane of economic growth feared by David Ricardo and the classical economists. If you double your inputs and get more than double your outputs, that is increasing returns to scale. Technological convergence generates increasing returns because one can double output without having to double one of the inputs (knowledge). Notice also that by economic growth economists usually have in mind intensive growth, that is, growth in real output per capita, rather than extensive growth, which is just plain growth in real output.



**Figure 10.1** The decreasing cost of computing power (1998 U.S.\$ per MFLOPS). *Source:* Kurzweil (1999: 320–1).

This chapter argues that the rapid performance improvements and price declines experienced by the semiconductor and computer industries can be traced to the status of these technologies as GPTs. Perhaps the most important GPT is the planar process, the basic technique for fabricating increasingly large numbers of transistors on a single chip. But there are other related GPTs as well, including the integrated circuit, standardized memory, the microprocessor, the von Neumann stored-program architecture of computers, and modular computer platforms.

The chapter also makes a number of more specific arguments about the sources of American success in digital technology.

- America benefited early on from federal demand for both semiconductors and computers, but the role of computer demand quickly supplanted government procurement as a driver of semiconductor technology. The two technologies codeveloped or coevolved in a virtuous cycle of technological progress, with declining costs in semiconductors driving increased demand for computers, and increased demand for computers driving cost declines in semiconductors.
- Universities played a crucial role in the birth of the digital computer and in developing the field of computer science, a body of knowledge complementary to the computer. But universities played little role in the birth and development of semiconductor technology, which was driven by the capabilities of semiconductor firms.
- Technological advance and diffusion of both semiconductor and computer technology depended on the *lack* of broad intellectual-property protection, especially of the principles underlying fundamental general-purpose technologies.
- Like its early success, the recent resurgence of the American semiconductor industry can be traced to the synergistic effect of demand from the American computer industry, in this case the microcomputer industry. Contrary to most pronouncements at the nadir of American fortunes in the mid-1980s, American institutions and industrial structure have not proven inherently inferior to those of Japan; indeed, in its ability to spur innovation, the vertically fragmented Amer-

ican industry may prove the more viable long-term model.

- Government policy toward the semiconductor and computer industries around the world has had a significant effect on the development of those industries, but the actual effects were often quite different from those intended. In both Japan and the United States, policy to foster cooperative research and development never proved effective, although in both cases it had some benefit in enhancing capabilities in the semiconductor equipment, rather than directly in the semiconductor manufacturing, industry. National policies to promote a mainframe computer industry on the model of IBM diverted resources from areas of national comparative advantage and were ultimately rendered irrelevant by the development of the microcomputer.

## 10.2. Semiconductors and computers: an interwoven history

### 10.2.1. *The origins of semiconductor technology*

Even though semiconductors and computers were born in the years immediately following World War II, their institutional origins were quite different. The invention of the computer involved both universities and direct government research funding. By contrast, the transistor—the basic building block of the semiconductor industry—emerged from private research at AT&T's Bell Labs. Because of its success and its secure status as the nation's telephone monopoly, AT&T was able to pursue a policy of research that, while arguably more focused toward commercial ends than basic research at universities, was nonetheless willing to indulge basic science and to envisage a research agenda quite far from commercial fruition.<sup>3</sup>

One of the long-run problems facing the Bell System was the expansion of a switching system

based on electromechanical relays. By the 1930s, Mervin Kelly, the research director at AT&T's Bell Labs, was voicing the opinion that electromechanical relays would eventually have to be replaced by an electronic alternative in order to handle the growing volume of traffic. William Shockley, one of the three Bell scientists to receive the Nobel Prize for the transistor, was impressed by this observation, and believed that the objective would be best realized with solid-state technology (Shockley, 1976).

Bell Labs announced the transistor in December of 1947.<sup>4</sup> Almost immediately transistor technology began spilling out to other firms. This was not, however, a process in which slippery knowledge leaked unintentionally to others but rather a deliberate and systematic attempt by AT&T to disseminate know-how through inexpensive licenses, technical symposia, and site visits (Tilton, 1971: 75–6; Braun and Macdonald, 1978: 54–5). The main driver of this policy was the consent decree AT&T had just signed with the Antitrust Division of the U. S. Justice Department, which specified how the company was to treat technology outside the scope of the company's primary mission.<sup>5</sup> But there is also reason to think that AT&T pursued the strategy of dissemination because the company saw value in taking advantage of the capabilities of others. AT&T was still primarily concerned with the usefulness of transistors to its own line of business, telephone switching. Although AT&T had developed the transistor and begun using it early in telephone devices and circuits, it was still an extremely immature technology. The company believed that if it allowed access to the transistor, telephony would reap the benefits of spillovers from the development of the capabilities of others in the electronics industry to an extent that would outweigh the foregone revenues of proprietary

<sup>4</sup> For detailed histories of the invention of the transistor, see Braun and Macdonald (1978), Morris (1990), and Nelson (1962).

<sup>5</sup> AT&T's strategy of dissemination may also have been motivated by a desire to preempt any thought the military might have had of classifying the technology (Levin, 1982: 58).

<sup>3</sup> For a classic account of how the research environment at Bell Labs led to the transistor, see Nelson (1962).

development<sup>6</sup> (McHugh, 1949; Bello, 1953; Braun and Macdonald, 1978: 54; Levin, 1982: 76–7).

One implication of this policy of easy access to the basic technology is that profits—or “rents,” as economists would put it—would accrue not to the inventors as much as to those who could make innovative commercial improvements in the basic technology. The unintended consequence was thus to create a large cohort of entrants intent on finding ways to commercialize this new technology (Mowery and Steinmueller, 1994).

The large vacuum-tube firms, as well as Bell Labs itself, continued to be major sources of transistor innovations through the 1950s, especially in the realm of process and materials. The work of this period led ultimately to a pivotal innovation that did allow for rapid experience-based improvements and cost reductions: the *planar process*, a development arguably responsible for the increasing-returns trajectory upon which the semiconductor industry now finds itself. But the planar process was not developed by Bell Labs or by any of the established vacuum-tube firms. Instead, in what would become a pattern characteristic of the American semiconductor industry, the new approach was developed by a small start-up organization.

Among the many Bell Labs researchers who had struck out on their own in the 1950s was Shockley, who returned home to the San Francisco peninsula to found Shockley Semiconductor Laboratories. Apparently prompted by dissatisfaction with the company’s orientation toward product breakthroughs at the neglect of the commercially richer area of process technology (Braun and Macdonald, 1978: 84; Holbrook, 1999), eight of Shockley’s team defected in 1957, and, with the backing of Long Island entrepreneur Sherman Fairchild,

founded the semiconductor division of Fairchild Camera and Instrument Corporation. The Fairchild group mounted an ambitious plan to produce silicon mesa transistors using technology developed at Bell Labs (Malone, 1985: 88; Lydon and Bambrick, 1987: 6). In attempting to overcome some of the limitations of this transistor design, one of the eight defectors, Jean Hoerni, found a way to create a “planar” device—that is, a device created by building up layers on a flat surface (Dummer, 1978: 143; Braun and Macdonald, 1978: 85; Morris, 1990: 38). The planar structure made it easy for Fairchild to devise a way to replace the mesa’s clumsy wires with metal contacts deposited on the surface.

The advantages of the planar process for transistor fabrication were overwhelming and recognized immediately throughout the industry (Sparkes, 1973: 8). It has become the basis of all semiconductor fabrication, including, of course, the integrated circuit (IC), for which the process is of critical importance. ICs are semiconductor devices containing an entire circuit of transistors and other devices on a single “wafer” or chip. The IC held out the promise of overcoming a developing bottleneck in the mass fabrication of transistor-based systems, what Braun and Macdonald (1978: 113) have aptly called the “tyranny of numbers.” As systems became more complex, requiring interconnections among hundreds of transistors, assembly costs mounted; more importantly, complex systems became vulnerable to the failure of any single connection or component. By fabricating an entire circuit using the techniques of semiconductor manufacture, the “monolithic” approach could yield greater reliability.

By 1961, two Americans, Robert Noyce of Fairchild and Jack Kilby of Texas Instruments (TI), had created prototype ICs.<sup>7</sup> Unlike Kilby, who had started with the monolithic idea and then sought to solve the problem of fabrication

<sup>6</sup> An AT&T vice president put it this way. “We realized that if this thing [the transistor] was as big as we thought, we couldn’t keep it to ourselves and we couldn’t make all the technical contributions. It was to our interest to spread it around. If you cast your bread on the water, sometimes it comes back angel food cake.” Quotation attributed to Jack Morton, in “The Improbable Years,” *Electronics* 41: 81 (February 19, 1968), quoted in Tilton (1971: 75–6).

<sup>7</sup> The idea of the integrated circuit was probably first propounded in 1952 by G.W.A. Dummer of the British Royal Radar Establishment (Braun and Macdonald, 1978: 108).

and interconnection, Noyce began with a process for fabrication and metallic interconnection—the planar process—and moved easily from that to the idea of the integrated circuit. Under pressure from the industry, TI and Fairchild forged a cross-licensing agreement in 1966 under which each company agreed to grant licenses to all comers in the range of 2–4 percent of IC profits (Reid, 1984: 94–5). This practice served to reproduce and extend the technology licensing policies of AT&T, again broadly diffusing the core technological innovation to all entrants and thereby reasserting the principle that innovative rents should flow to those who could commercialize and improve upon the key innovation.

As important as the innovation of the IC was, the planar process is arguably the more important technological breakthrough, not merely because it underlay the IC but because it provided the paradigm or technological trajectory the industry was to follow.<sup>8</sup> By either etching away minute areas or building up regions using other materials, semiconductor fabrication alters the chemical properties of a “wafer,” a crystal of silicon. Each wafer produces many ICs, and each IC contains many transistors. The most dramatic economic feature of IC production is the increase in the number of transistors that can be fabricated in a single IC. Transistor counts per IC increased from 10 to 4,000 in the first decade of the industry’s history; from 4,000 to over 500,000 in the second decade; and from 500,000 to 100 million in the third decade. The ten-million-fold increase in the number of transistors per IC has been accompanied by only modest increases in the cost of processing of a wafer, and almost no change in the average costs of processing the individual IC. This factor alone has been responsible for the enormous cost reduction in electronic circuitry since the birth of the IC. Electronic systems comparable in complexity to vacuum-tube or transistor systems costing millions of dollars can be constructed for a few hundred dollars, a magnitude of cost reduction that it is virtually unpre-

cedented in the history of manufacturing. The cheapness of electronic functions has reduced the costs of electronic systems relative to mechanical ones and lowered the relative price of electronic goods in general—developments that have had a major effect on the industrial structure of the electronics and IC industries.

Langlois and Steinmueller (1999) have pointed to the critical role of end-use demand in shaping industrial structure and competitive advantage in the worldwide semiconductor industry throughout its history. In the early years, demand in the United States came first from military sources and then importantly from the computer industry. Government procurement demand proved so valuable to the development of the industry not only because of its extent but also because of the military’s relative price-insensitivity and its insistence on reliability (Dosi, 1984). Commercial demand eventually grew more rapidly than military, however, and, by the mid-1970s, government consumption had declined to less than 10 percent of the market (Kraus, 1971: 91).

The American government also pushed the transistor and the IC through support of R&D and related projects. But scholarship on the subject is essentially unanimous that this activity was far less important for, and less salutary to, the industry than was the government’s procurement role. All the major breakthroughs in transistors were developed privately with the military market (among others) in mind. And, although the government tended to favor R&D contracts with established suppliers, notably the vacuum-tube firms, it *bought* far more from newer specialized semiconductor producers (Tilton, 1971: 91). The pragmatic policy of awarding work to those firms that could meet supply requirements was particularly important for encouraging new entry.

A significant feature of the transition to the IC was the virtual disappearance of the vertically integrated American electronics companies that had led in the production of vacuum tubes and that had been able to stay in the race during the era of discrete transistors. The

<sup>8</sup> Canonical sources here are Abernathy and Utterback (1978) and Utterback (1979).

market shares of those firms declined in the face of new entrants and the growth of relatively specialized manufacturers like TI, Fairchild, and Motorola. Why did the vertically integrated electronic system firms do so poorly in this era? Wilson et al. (1980) point out that the new leaders were either specialized start-ups or multidivisional firms (like TI, Fairchild, and Motorola) in which the semiconductor division dominated overall corporate strategy and in which semiconductor operations absorbed a significant portion of the attention of central management. By contrast, the semiconductor divisions of the integrated system firms were a small part of corporate sales and of corporate strategy, thereby attracting a smaller portion of managerial attention and receiving less autonomy.

### 10.2.2. *The birth of the digital computer*

The history of the digital computer has much in common with that of semiconductor technology, even if there are a number of important differences. Like the transistor, the digital computer was developed with a specific bottleneck in mind. But, unlike the transistor, the digital computer was developed not privately but at universities, with explicit government subsidy from the start.

During World War II, the U.S. Army contracted with J. Presper Eckert and John W. Mauchly of the Moore School at the University of Pennsylvania for a device “designed expressly for the solution of ballistics problems and for the printing of range tables”<sup>9</sup> (Stern, 1981: 15). By November 1945, they had produced the Electronic Numerical Integrator and Computer (ENIAC), the first fully operational all-electronic digital computer—a behemoth occupying 1,800 square feet, boasting 18,000 tubes, and consuming 174 kilowatts of electricity. Universities continued to play an important role throughout the early life of the technology, helping to create the wholly

new discipline of computer science. Indeed, Rosenberg and Nelson go so far as to call the computer “the most remarkable contribution of American universities to the last half of the twentieth century” (Rosenberg and Nelson, 1994: 331).

Like the transistor, the computer opened up wide possibilities for technological convergence.<sup>10</sup> In part, this convergence arose because of the falling cost of computation—attendant eventually on the falling cost of semiconductors—which allowed the device to be used in a wide range of applications requiring numerical computation and, later, information processing more generally. But a specific innovation in the design of digital computers was also central to the device’s wide potential. In the summer of 1944, the mathematician John von Neumann learned by accident of the Army’s ENIAC project. Von Neumann began advising the Eckert-Mauchly team, which was working on the development of a new machine, the EDVAC. Out of this collaboration came the concept of the stored-program computer: instead of being hard-wired, the EDVAC’s instructions were to be stored in memory to facilitate modification. A single hardware design could thus be quickly adapted to a variety of different uses through what came to be called *software*. Von Neumann’s abstract discussion of the stored-program concept (von Neumann, 1945) circulated widely and served as the logical basis for virtually all subsequent computers.<sup>11</sup>

Government, especially military, support for the computer remained significant throughout the 1950s, and government funding helped spur important technical developments like ferrite-core memory, which emerged from the military-funded Whirlwind project at MIT (Redmond and Smith, 1980; Pugh, 1984).

<sup>9</sup> In the event, the end of the war reduced the urgency of this goal, and the first major task given the ENIAC was actually to perform calculations for the development of the hydrogen bomb (Stern, 1981: 62).

<sup>10</sup> Rosenberg (1992: 382) explicitly likens the technological convergence of the digital computer to that of the nineteenth-century machine tool industry.

<sup>11</sup> The stored-program idea was also contained in the work of Turing in Britain, and the first functioning storable-program computer was run for the first time on June 21, 1948 at the University of Manchester.

But, as Bresnahan and Malerba (1999: 89–90) argue, government research support had little to do with the success of the *commercial* computer industry. Moreover, much of government policy, notably in the areas of R&D funding and antitrust, was actually aimed at forestalling the emergence of IBM as a dominant “national champion” in computers. As in semiconductors, however, the military’s pragmatic approach to procurement favored those firms who could deliver the goods, and in computers that meant IBM (Bresnahan and Malerba, 1999: 90; Usselman, 1993).

By the mid-1960s, however, IBM found itself riding herd on a multiplicity of physically incompatible systems—the various 700-series computers and the 1400 series, among others—each aimed at a different use. Relatedly, and more significantly, software was becoming a serious bottleneck. By one estimate, the contribution of software to the value of a computer system had grown from 8 percent in the early days to something like 40 percent by the 1960s (Ferguson and Morris, 1993: 7). And writing software for so many incompatible systems greatly compounded the problem. In what *Fortune* magazine called “the most crucial and portentous—as well as perhaps the riskiest—business judgment of recent times,” IBM decided to “bet the company” on a new line of computers called the 360 series. The name referred to all the points of the compass, for the strategy behind the 360 was to replace the diverse and incompatible systems with a single modular family of computers (Flamm, 1988: 96–9). Instead of having one computer aimed at scientific applications, a second aimed at accounting applications, etc., the company would have one machine for all uses. This was not to be a homogeneous or undifferentiated product; but it was to provide a framework in which product differentiation could take place while retaining compatibility.

As Timothy Bresnahan suggests, the 360 was the first major computer *platform*, by which he means “a shared, stable set of hardware, software, and networking technologies on which users build and run computer applications” (Bresnahan, 1999: 159). To put it another way, the 360 was a *modular system*, albeit one

that remained mostly closed and proprietary despite the efforts of the “plug compatible” industry to pick away at its parts. The essence of such a system is compatibility among the components, which, in the case of a computer platform, is maintained by (often de facto) interface standards (Langlois and Robertson, 1992). A large literature has arisen describing the positive-feedback character of technical standards: the more users adopt a platform, the more desirable that platform becomes to others, leading to a “virtuous circle” and pressure for the dominance of a single platform.<sup>12</sup> The IBM 360 did indeed become a dominant platform, a prototype form of general-purpose technology in the computer industry.<sup>13</sup>

As the market for computers picked up speed, the symbiosis between computers and semiconductors became stronger. In contrast to IBM, which did not begin using ICs until 1970, IBM’s competitors, such as RCA and Burroughs, adopted ICs more quickly in an effort to gain an advantage (Borras et al., 1983: 157). This led to a dynamic interaction in which competition among computer makers drove the demand for ICs, which lowered IC prices by moving suppliers faster down their learning curves, which in turn fed back on the price of computers, etc. The result was a self-reinforcing process of growth for both industries. Indeed, the falling prices of semiconductor logic fueled a second computer revolution, that of the minicomputer.

Minicomputers were smaller than mainframes and geared toward specialized scientific and engineering uses. Digital Equipment Corporation (DEC), founded in 1957, was the pioneer in the field. Among the other firms to enter the minicomputer market were Scientific Data Systems, Data General (founded in 1968 by defectors from DEC), Prime Computer, Hewlett-Packard, Wang, and Tandem (Flamm, 1988: 131).

<sup>12</sup> Useful entry points are David and Greenstein (1990) and Economides (1996).

<sup>13</sup> “The very idea of platform is associated with re-use across multiple sites, an inherent scale economy” (Bresnahan, 1999: 160).

### 10.2.3. *Memory races and the Japanese challenge*

The early history of innovation in semiconductors is largely an American story. But European and Japanese firms did enter the industry early, and the paths of development in those areas were guided in large part by rather different structures of end-use demand and government policy.

In terms of innovation, European firms trailed American firms in the early years of the transistor, but they nonetheless remained competitive in germanium transistors well into the 1960s by concentrating on the European market, where the dominant demand was for consumer and industrial, rather than military and computer, uses (Malerba, 1985: 75–80, 88–9). This structure of demand gave advantage in Europe to the large vertically integrated systems houses, who viewed transistors as a necessary input into electronic system products rather than as an end product. Significantly, the European firms tended to license technology almost exclusively from those American firms whom they most resembled—the large vacuum-tube firms—and almost not at all from the American merchant houses (Malerba, 1985: 65).

By the mid-1960s, Britain, France, and Germany had all begun efforts to foster national computer industries (Dosi, 1981: 27). As Bresnahan and Malerba (1999) point out, much of those European (and of Japanese) policies toward computers were aimed at forestalling IBM with preferential procurement policies as well as outright subventions. By subsidizing national computer makers, who were motivated if not constrained to buy from national semiconductor makers, the European computer initiatives thus attempted to create some indigenous demand for logic ICs. Moreover, all three countries initiated R&D programs in computers, some of which spilled over into semiconductors.<sup>14</sup> As Tilton (1971: 131) notes, these programs tended to favor a small number of large established firms—to a much greater extent than had American military R&D. Indeed, European government policy in this

period encouraged consolidation and rationalization. Especially in Britain and France, which did not initially have “national champions” the size of Philips or Siemens, a wave of mergers took place, both in computers and semiconductors, with government approval and sometimes government instigation. This policy of consolidation had the effect of reducing indigenous competition in the face of penetration by subsidiaries of American firms and generated “champions” that proved unfit to take on the Americans (Tilton, 1971: 131–2).

The early origins of the Japanese semiconductor industry are broadly similar to those of the European, albeit with some differences that may prove crucial in explaining the quite distinct path of Japanese development in later periods. As in Europe, the principal producers of transistors in the 1950s and 1960s were diversified systems houses, including firms that had previously produced vacuum tubes, rather than companies that were principally specialized into semiconductors. And, as in Europe, the main end-use for transistors in Japan in this period was consumer products rather than the military. At the same time, however, there were substantial differences from Europe at both the level of the firms themselves and at the level of government policy.

Japan responded to American competitive advantage with high tariffs, and in addition imposed quotas and registration requirements<sup>15</sup> (Tyson and Yoffie, 1993: 37). In contrast to European policies, moreover, the Japanese government essentially forbade foreign direct investment, which forced American firms to tap the Japanese market only through licensing and technology sales to Japanese firms rather than through direct investment. In addition, the rate of growth of the Japanese semiconduc-

<sup>14</sup> Several of these programs are described in Dosi (1981: 27).

<sup>15</sup> This is in contrast to European policy, which featured high tariffs but no prohibition on foreign direct investment. As a result, much of the European demand for semiconductors was satisfied by European subsidiaries of American companies. Japanese companies have typically supplied some 90 percent of the Japanese semiconductor market, whereas American firms—through imports or foreign direct investment—have supplied between 50 and 70 percent of the European market (Tyson and Yoffie, 1993: 34).



tor industry was much greater than that of the European simply because the Japanese started from a smaller base. And because the Japanese vacuum-tube firms were much smaller than their American or European counterparts at the beginning of the transistor era, they had less to lose in moving to the new technology. As Tilton (1971: 154) notes, the fact of rapid growth “also helped create a receptive attitude toward change on the part of the receiving tube producers by reducing the risks associated with new products and new technologies and by increasing costs, in terms of declining market shares, to firms content simply to maintain the status quo.” In many ways, then, Japanese systems firms faced many of the same constraints, and adopted many of the same approaches, as the aggressive American merchant firms rather than those of the American, or European, systems houses.<sup>16</sup>

Despite their early success in transistors, Japanese firms found themselves in a weak position by the 1970s. These firms were slow to make the transition to batch-produced silicon devices in the early 1960s, and, when they turned later in the decade to the production of bipolar ICs, they could not compete with the likes of Texas Instruments and National Semiconductor; some Japanese firms accused the Americans of “dumping” (Okimoto et al., 1984: 14–5). After 1967, indeed, the purchase of American ICs created a Japanese trade deficit in semiconductors (Malerba, 1985: 136).

How did Japanese industry move from this weak position in the 1970s to its dominant position by the mid-1980s? Until recently, the tacit assumption of most commentators had been that Japanese success was the result of some combination of (1) Japanese industrial structure, understood as superior to American industrial structure in a very general or even absolute sense, and (2) Japanese industrial policy, understood as a highly intentional—and even prescient—system of government industry planning and control. Langlois and Steinmueller (1999)

suggests a somewhat different picture. Although both industrial structure and government policy played important roles in the rise of the Japanese semiconductor industry, the benefits of that industrial structure were far less timeless than commentators supposed, and the effects of government policy were far less intentional, and perhaps somewhat less significant, than the dominant accounts suggested.

As in the earlier rise of the American semiconductor industry, the pattern of end-use demand was crucial in shaping the bundle of capabilities that Japanese industry possessed, as well as in narrowing and limiting the choices the Japanese firms had open to them. In this case, that end-use demand came largely from consumer electronics and, to a somewhat lesser extent, from telecommunications, especially purchases by NTT, Japan’s national telephone monopoly (see Table 10.1). Consumer demand helped place the Japanese on a product trajectory—namely MOS and especially CMOS ICs—that turned out eventually to have much wider applicability.<sup>17</sup> Moreover, Japanese firms adopted a strategy of specialization in high-volume production of one particular kind of chip—the DRAM. The DRAM, or dynamic random-access memory chip, is a technology that benefited from increasing returns to scale not only because of the volume effects of mass production but also because it is arguably a general-purpose technology of considerable importance—a device that can store digital information for a wide variety of purposes.<sup>18</sup>

Established American firms, accustomed to providing customized devices, were slow to recognize the cost-reduction advantages of a standardized memory chip (Wilson et al., 1980: 87; Dorfman, 1987: 193). Two new firms—National and Intel—quickly gained

<sup>17</sup> MOS stands for metal-oxide semiconductor, and CMOS for “complementary” MOS.

<sup>18</sup> DRAMs are “dynamic” in the sense that the electric charges containing the remembered information decay over time and need periodically to be “refreshed.” This stands in contrast to the static RAM (or SRAM), which does not require refreshing, but which therefore has disadvantages in size, cost, and power consumption because it requires more transistors per memory cell.

<sup>16</sup> Unlike European firms, the Japanese firms sought and received licenses from Texas Instruments, Fairchild, and other American merchant firms rather than limiting themselves to arrangements with American systems houses.

**TABLE 10.1**

Demand for Integrated Circuits by End-use Market, United States, Japan, and Western Europe, 1982 and 1985 (in percent)

<i>End-use</i>	<i>United States</i>		<i>Japan</i>		<i>Western Europe</i>	
	1982	1985	1982	1985	1982	1985
Computer	40	45	22	36	25	20
Telecommunications	21	10	10	13	20	29
Industrial	11	10	17	6	25	19
Military and Aerospace	17	18	0	0	5	7
Consumer	11	16	51	45	25	25

*Source.* OECD (1985).*Note.* Includes captive consumption.

advantage over their established competitors in the merchant market by moving more quickly into the production of high-volume standardized devices. Both firms were spin-offs from Fairchild—two of the first of what came to be called the “Fairchildren” (Lindgren, 1969). In pushing standardized DRAM chips, however, these firms precipitated a “memory race” in which Japanese firms were eventually to prove dominant. American firms led in the early—1K and 4K—DRAM markets. But an industry recession delayed the American “ramp-up” to the 16K DRAM, which appeared in 1976. Aided by unforeseen production problems among the three leaders, Japanese firms were able to gain a significant share of the 16K market. By mid-1979, sixteen companies were producing

DRAMs, and Japanese producers accounted for 42 percent of the market (Wilson et al., 1980: 93–4) (see Table 10.2). The opportunity opened for Japanese producers in the 16K DRAM market had proven sufficient for them to advance to a position of leadership in the 64K DRAM market. Japanese dominance accelerated in the 256K (1982) and one-megabit (1985) generations. Intense price competition, combined with the general recession in the U.S. industry in 1985, caused all but two American merchant IC companies to withdraw from DRAM production<sup>19</sup> (Howell et al., 1992: 29). In 1990, American market share had fallen to only 2 percent of the new generation 4-megabit DRAMs<sup>20</sup> (see Table 10.2).

Why did the Japanese succeed? In broad terms, circumstances had staked out for the Japanese industry a strategic path that fit well the existing competences of the firms—namely those in mass production and quality control—and supported the thrust of their final products, which, despite government efforts of to create a computer industry (Fransman, 1990), were still in consumer electronics and telecommunications.

**TABLE 10.2**

Maximum Market Share in DRAMs by American and Japanese Companies, by Device

<i>Device</i>	<i>Maximum market share (%)</i>	
	<i>United States</i>	<i>Japan</i>
1K	95	5
4K	83	17
16K	59	41
64K	29	71
256K	8	92
1M	4	96
4M	2	98

*Source.* Dataquest, cited in Methé (1991: 69)

<sup>19</sup> The exceptions were Texas Instruments, which produced in Japan, and Micron Technology, which produced in Idaho.

<sup>20</sup> These figures do not take into account the sizable captive production at IBM and AT&T.

Rather than feeling that they were on the verge of overtaking American companies, the Japanese saw their computer industry as relatively weak against IBM, and perceived that a key feature of IBM's advantage was technology, specifically its position in ICs. From the viewpoint of Japanese firms, the American IC industry was enormously innovative but did not share much of the manufacturing culture that had developed in the larger Japanese electronics companies, where quality, systematic capacity expansion, and long-term market position were regarded as key variables to control. The fact that Japanese IC producers were large companies in comparison with their American counterparts gave them one particular advantage: they were able to mobilize internal capital resources to make investments in the IC industry in a way that American companies could not.

James March (1991) has pointed out that there is a necessary tradeoff between exploration and exploitation—tradeoff between searching for new ideas and running with the old ones. As the technology leaders, the American firms found themselves with a full plate of alternatives to pursue, in both product and process technology. Sitting somewhat behind the frontier, Japanese firms could pick one item off the plate and run with it. Their morsel was the mass production of DRAMs.

#### 10.2.4. Personal computers and the American resurgence

American industry and politics certainly did not let these events go unnoticed, and alarms went up as early as the 64K generation. More worrisome than the loss of the memory market was the possibility that Japanese dominance in DRAMs would be translated into equal success in other kinds of chips. Although memories constituted at most 30 percent of the IC market, many believed them to be “technology drivers” essential for continued progress in increasing the number of transistors on an IC. If American firms could not use DRAM production to develop and gain experience in the next generation of technology, then Japanese producers would soon be able to climb up the design-complexity ladder and challenge American positions in logic markets (Ferguson, 1985; Forester, 1993).

In 1986, Japan's overall market share in semiconductors slipped ahead of that of the American merchants. Thus, in 1988 the American industry appeared to stand on the brink of oblivion, with no haven in product or process that could be counted to insure its survival into the 1990s. But the predicted extinction never occurred (see Figure 10.2). Instead, American firms surged back during the 1990s, and it now seems that it is the Japanese who are embattled.

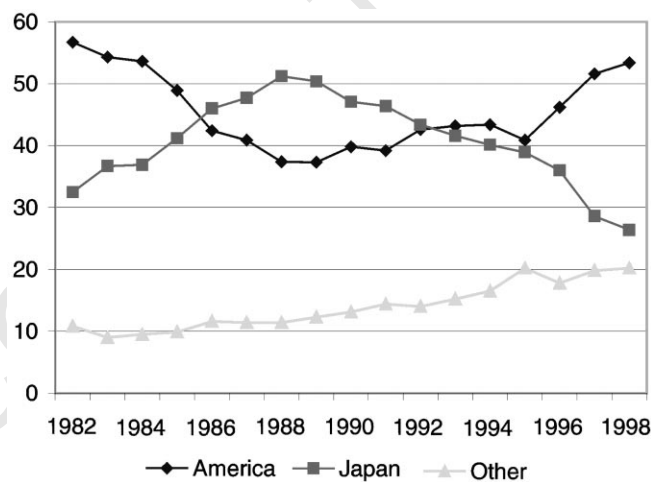


Figure 10.2 Worldwide semiconductor market shares (in percent), 1982–98. Source: Semiconductor Industry Association.

Langlois and Steinmueller (1999) argue that this resurgence is not the result of imitating Japanese market structure and policy but rather of taking good advantage of the distinctly American market structure and capabilities developed in the heyday of American dominance. Just as the innovation of, and the growing market for, the standardized DRAM had favored the Japanese, another semiconductor innovation, and the burgeoning market it created, came to favor the Americans. That innovation was the microprocessor, an IC designed not to store information (like the DRAM) but rather to provide on a single chip the information-processing capability of a digital computer.

In 1969, a Japanese manufacturer asked Intel to design the logic chips for a new electronic calculator. Marcian E. (“Ted”) Hoff, Jr., the engineer in charge of the project, thought the Japanese design too complicated to produce. The then-current approach to the design of calculators involved the use of many specialized hard-wired circuits to perform the various calculator functions. Influenced by the von Neumann architecture of minicomputers, Hoff reasoned that he could simplify the design enormously by creating a single programmable IC rather than the set of dedicated logic chips the Japanese had sought (Noyce and Hoff, 1981). By using relatively simple general-purpose logic circuitry that relied on programming information stored elsewhere, Hoff effectively substituted cheap memory (then Intel’s major product) for relatively expensive special-purpose logic circuitry (Gilder, 1989: 103). The result was the Intel 4004, the first microprocessor. A sixth of an inch long and an eighth of an inch wide, the 4004 was roughly equivalent in computational power to early vacuum-tube computers that filled an entire room. It also matched the power of a 1960s IBM computer whose central processing unit was about the size of a desk (Bylinsky, 1980: 7).

Intel gained an early lead in microprocessors that it never relinquished. Early on, Intel did not push patent protection, and, in Hoff’s view, “did not take the attitude that the microprocessor was something that you could file a patent claim on that covers everything” (quoted

in Malone, 1985: 144). Because the microprocessor is a general-purpose computer, there are many different ways to implement the microprocessor idea without infringing on a particular implementation. And the appropriation of rents in microprocessors has always depended on first-mover advantage rather than on patent protection for particular features of the system design or on the ability to produce a microprocessor that could not be emulated technically.

The microprocessor found uses in a wide variety of applications involving computation and computer control. But it did not make inroads into the established mainframe or minicomputer industries, largely because it did not initially offer the level of computing power these larger machines could generate using multiple logic chips. Instead, the microprocessor opened up the possibility of a wholly new kind of computer—the microcomputer.

The first microcomputer is generally acknowledged to have been something called the MITS/Altair, which graced the cover of *Popular Electronics* magazine in January, 1975.<sup>21</sup> Essentially a microprocessor in a box, the machine’s only input/output devices were lights and toggle switches on the front panel, and it came with a mere 256 bytes of memory. But the Altair was, at least potentially, a genuine computer. Its potential came largely from a crucial design decision: the machine incorporated a number of open “slots” that allowed for additional memory and other devices to be added later. These slots were hooked into the microprocessor by a network of wires called a “bus.” This extremely modular approach emerged partly in emulation of the design of minicomputers and partly because hobbyists and the small firm supplying them would have been incapable of producing a desirable (i.e., more-capable) nonmodular machine within any reasonable time. In effect, the hobbyist community captured the machine, and made it a truly open modular system. The first clone of the Altair—the IMSAI 8080—appeared within a matter of months, and soon the Altair’s archi-

<sup>21</sup> For a much longer and better-documented history of the microcomputer, see Langlois (1992), on which this section draws.

ecture became an industry standard, eventually known as the S-100 bus because of its 100-line structure.

The S-100 standard dominated the hobbyist world. But the machine that took the microcomputer into the business world adopted a distinctive architecture, built around a Motorola rather than an Intel microprocessor. Stephen Wozniak and Steven Jobs had started Apple Computer in 1976, quite literally in the garage of Jobs's parents' house. The hobbyist Wozniak, also influenced by the architecture of minicomputers, insisted that the Apple be an expandable system—with slots—and that technical details be freely available to users and third-party suppliers. With the development of word processors like WordStar, database managers like dBase II, and spreadsheets like VisiCalc, the machine became a tool of writers, professionals, and small businesses. Apple took in U.S.\$750,000 by the end of fiscal 1977, U.S.\$8 million in 1978, U.S.\$48 million in 1979, U.S.\$117 million in 1980 (when the firm went public), U.S.\$335 million in 1981, U.S.\$583 million in 1982, and U.S.\$983 million in 1983.<sup>22</sup>

Existing computer companies were slow to develop competing microcomputers, largely because they saw the machines as a small fringe market. But as business uses increased and microcomputer sales rose, some computer makers saw the opportunity to get a foothold in a market that was complementary to, albeit much smaller than, their existing product lines.<sup>23</sup> By far the most significant entry was that of IBM. On August 12, 1981, IBM introduced the computer that would become the paradigm for most of the 1980s.

In a radical departure, IBM decided to produce the machine outside the control of

company procurement policies and practices. Philip Donald Estridge, a director of the project, later put it this way. "We were allowed to develop like a start-up company. IBM acted as a venture capitalist. It gave us management guidance, money, and allowed us to operate on our own" (*Business Week*, October 3, 1983: 86). Estridge knew that, to meet the deadline he had been given, IBM would have to make heavy use of outside vendors for parts and software. The owner of an Apple II, Estridge was also impressed by the importance of expandability and an open architecture. He insisted that his designers use a modular bus system that would allow expandability, and he resisted all suggestions that the IBM team design any of its own add-ons. Because the machine used the Intel 8088 instead of the 8080, IBM needed a new operating system. A tiny Seattle company called Microsoft agreed to produce such an operating system, which they bought from another small Seattle company and rechristened as MS-DOS, for Microsoft Disk Operating System.

The IBM PC was an instant success, exceeding sales forecasts by some 500 percent. By 1983, the PC had captured 26 percent of the market, and an estimated 750,000 machines were installed by the end of that year. The IBM standard largely drove out competing alternatives during the decade of the 1980s. This happened in part because of the strength of the IBM name in generating network effects, principally because it created the expectation among users that the key vendor would continue to provide services long into the future and that a wide array of complementary devices and software would rapidly become available. But in large measure the "tipping" of the market to the IBM PC standard was a result of the openness of the IBM system, which could be easily copied by others, and the eagerness of Microsoft to license MS-DOS to all comers.

As it had with the 360/370 series, IBM had created a dominant computer platform. But, in the case of the PC, the dominance of the platform would not translate into a dominant market share for IBM. Because of the strategy of outsourcing and the standards it necessitated, others could easily imitate the IBM hard-

<sup>22</sup> Data from Apple Computer, cited in "John Sculley at Apple Computer (B)," Harvard Business School Case no. 9-486-002, revised May 1987: 26.

<sup>23</sup> Few people inside or outside IBM foresaw the sweeping changes the PC would make in computer markets. In April 1981, four months before the official announcement of the IBM PC, IBM gave presentations estimating it would sell 241,683 PCs over five years. In fact, IBM shipped 250,000 PCs in one month alone (Zimmerman and Dicarolo, 1999).

ware, in the sense that any would-be maker of computers could obtain industry-standard modular components and compete with IBM. A legion of clones appeared that offered IBM compatibility at, usually, a lower price than IBM. By 1986, more than half of the IBM-compatible computers sold did not have IBM logos on them. By 1988, IBM's worldwide market share of IBM-compatible computers was only 24.5 percent. IBM's choice of an open modular system was a two-edged sword that gave the company a majority stake in a standard that had grown well beyond its control. For reasons that are debated in the literature, but that likely have to do both with strategic mistakes by IBM and with the inherently strong positions of key suppliers in controlling their proprietary "bottleneck" technologies—the microprocessor and the operating system—Intel and Microsoft gained control of the standard that IBM had originally sponsored (Ferguson and Morris, 1993). The PC architecture is now often referred to as the "Wintel" (Windows/Intel) platform.

Langlois (1992) has argued that the rapid quality-adjusted price decline in microcomputers resulted not only from the declining price of computing power attendant on successive generations of Intel processors but also from the vibrant competition and innovation at the level of hardware components and applications software that resulted from the open modular design of the PC. A decentralized and fragmented system can have advantages in innovation to the extent that it involves the trying out of many alternate approaches simultaneously, leading to rapid trial-and-error learning. This kind of innovation is especially important when technology is changing rapidly and there is a high degree of both technological and market uncertainty (Nelson and Winter, 1977). Moreover, the microcomputer benefited from technological convergence, in that it turned out to be a technology capable of taking over tasks that had previously required numerous distinct—and more expensive—pieces of physical and human capital. By the early 1980s, a microcomputer costing U.S.\$3,500 could do the work of a U.S.\$10,000 stand-alone word-processor, while at the same time keeping track of the books

like a U.S.\$100,000 minicomputer and amusing the kids with space aliens like a 25-cents-a-game arcade machine.

The personal computer grew rapidly in a niche that existing mainframes and minicomputers had never filled. Quickly, however, the microcomputer's niche began to expand to encroach on the territory of its larger rivals, driven by the rapidly increasing densities and decreasing prices of memory chips and microprocessors. In the early 1980s, a class of desktop machines called workstations arose to challenge the dominance of the minicomputer in scientific and technical applications. As in the case of personal computers, the workstation market was driven by open technical standards and competition within the framework of what was largely a modular system (Garud and Kumaraswamy, 1993; Baldwin and Clark, 1997). Initially, these workstation used microprocessors and operating systems different from those of personal computers.<sup>24</sup> By the early 1990s, however, the same process of increasing power and decreasing cost began pushing the Windows-Intel platform into what is today a dominance of the workstation space. At the same time, workstations hooked together (or hooked to personal computers) began to take over many of the functions of larger minicomputers and mainframes. By the 1990s, networks of fast, cheap smaller machines were widespread, a development accelerated by the spectacular growth of the Internet.<sup>25</sup> This growth had a significant negative effect on the makers of larger computers, notably the Boston-area minicomputer makers. Many went bankrupt; and, in

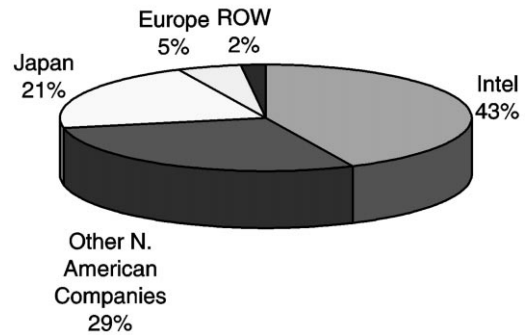
<sup>24</sup> So-called traditional workstations are built around Reduced-Instruction-Set-Computing (RISC) microprocessors and run variants of the UNIX operating system. Intel-platform workstations use high-end versions of the same microprocessors used in personal computers and typically run Microsoft's Windows NT or Windows 2000, which are compatible with Microsoft's operating systems for personal computers.

<sup>25</sup> In some respects, the demand for large websites created by the Internet has spurred demand for large central servers. Increasingly, however, even these servers are essentially high-powered workstations rather than traditional mainframes or minicomputers.

a telling development, the flagship maker of microcomputers—DEC—was acquired by Compaq, a maker of microcomputers. Bresnahan and Greenstein (1996, 1997) refer to this encroachment of smaller computers as the “competitive crash” of large-scale computing.

The losses incurred by the makers of large computers (including IBM) have been more than offset, however, by the growth of the personal computer industry and its suppliers. Principal among the beneficiaries has been the American semiconductor industry. The abandonment of the DRAM market by most American firms—including Intel—was a dark cloud with a bright silver lining. When Intel led the world industry in almost all categories, it and many of its American counterparts faced a full plate of product alternatives. With the elimination of mass memory as a viable market, these firms were impelled to specialize and narrow their focus to a smaller subset of choices. The areas in which American firms concentrated can generally be described as higher-margin, design-intensive chips. For such chips, production costs would not be the sole margin of competition; innovation and responsiveness would count for more. And innovation and responsiveness were arguably the strong suit of the “fragmented” American industry. As in the case of the personal computer industry, the decentralized structure of the American semiconductor industry permitted the trying out of a wider diversity of approaches, leading to rapid trial-and-error learning (Nelson and Winter, 1977). And the independence of many firms from larger organizations permits speedier realignment and recombination with suppliers and customers. Building on existing competences in design (especially of logic and specialty circuits) and close ties with the burgeoning American personal computer industry, American firms were able to prosper despite the Japanese edge in manufacturing technology (Ferguson and Morris, 1993).

The most important area of America specialization is microprocessors and related devices.<sup>26</sup> Between 1988 and 1994, a period in which merchant IC revenues grew by 121 percent, revenues from the microprocessor segment grew much faster than did memory



**Figure 10.3** Production of MOS microprocessors and related devices in 1996 (percent). *Source:* ICE (1998).

revenues (ICE, 1998). This evolution of the product mix in the industry has strongly favored American producers. In the microprocessor segment of the chip market, American companies accounted for 72 percent of world production in 1996, compared with a 21 percent share for Japanese companies (see Figure 10.3).

The importance of the microprocessor segment has meant that a single company, Intel, is responsible for much of the gain of American merchant IC producers. In 1996, Intel accounted for 43 percent of world output in the microprocessor segment (see Figure 10.4). Intel’s strategy for recovery, begun in the 1980s, has proven remarkably successful (Afuah, 1999). In the late 1980s, the firm consolidated its intellectual-property position in microprocessors by terminating cross-licensing agreements with other companies and, more importantly, began extending its first-mover advantage over rivals by accelerating the rate of new product introduction. These developments pushed Intel into the position of the largest IC producer in the world, with 1998 revenues of U.S.\$22.7 billion—more than the next three largest firms combined (see Table 10.3). Although Intel dominates the microprocessor market, it is not entirely without competitors; and it is significant that its principal

<sup>26</sup> This segment includes not only microprocessors but also microcontrollers (less sophisticated microprocessors that are used in embedded applications) and related “support” chips, such as memory controllers, that are necessary to assembling a microprocessor system.

**TABLE 10.3**  
Estimated 1998 Semiconductor Revenues (\$ million)

<i>Company</i>	<i>Revenues</i>
Intel	22675
NEC	8271
Motorola	6918
Toshiba	6055
Texas Instruments	6000
Samsung	4752
Hitachi	4649
Philips	4502
STMicroelectronics	4300
Siemens	3866
Fujitsu	3866

*Source:* Dataquest, cited in *Electronics Times* (1999: January 11, p. 3).

competitors in microprocessors are also American companies, notably AMD and Motorola.

The success of American firms in microprocessors and related chips has been reinforced by trends in end-use demand. In 1989, computer applications took 40 percent of merchant IC sales, followed by consumer and automotive applications at 28 percent.<sup>27</sup> By 1996, the respective shares were 50 percent for computer and 23 percent for consumer and automotive applications. The worldwide changes have led to increasing specialization. Between 1989 and 1994, North American use of ICs for computer applications soared from 15 to 24 percent of the total value of world merchant sales, while the Japanese IC market for consumer applications fell from 13 percent to 10 percent of world merchant sales. Thus, in contrast to rough parity (15 versus 13 percent) in 1989, an enormous gap has opened between IC demand for consumer and computer applications in the Japanese and American markets. Keep in mind that these figures are in terms of revenue not physical units, and much of the reversal of American fortunes has to do with the high value per component of microprocessors and other design-intensive chips, as against the low value

per unit of the mass-produced DRAMs on which Japanese firms long rested their strategies.

Another aspect of specialization that benefited the American industry was the increasing “decoupling” of design from production. Such decoupling is in many respects a natural manifestation of the division of labor in growing markets (Young, 1928); in this case, it was abetted by the development of computerized design tools (Hobday, 1991) and the standardization of manufacturing technology (Macher et al., 1998). On the one hand, this allowed American firms to specialize in design-intensive chips, taking advantage of an American comparative advantage that arguably arises out of the decentralized and “fragmented” structure of that country’s industry.<sup>28</sup> On the other hand, it also allowed many American firms to take advantage of growing production capabilities overseas. This “modularization” of the industry is spurring the kind of decentralized innovation from which the personal computer industry has benefited.

As globalization (broadly understood) has bolstered the fortunes of American firms, it has eroded those of the Japanese. Japanese firms were not the only ones who could understand the economics of capacity investment or productivity in manufacturing, and they were soon joined by Korean semiconductor producers and by larger American companies who matched Japanese productivity by the simple expedient of establishing Japanese plants. The result is a dilution of the control of capacity investment by Japanese producers. By the mid-1990s, a Korean firm had displaced Japanese firms as the leading producer of DRAMs in the world, and two other Korean firms had joined the top ten (see Table 10.4).

And what of the role of government policy in the American resurgence? The American response to the Japanese success of the early 1980s took two principal forms: (1) trade protection and (2) the funding of cooperative research,

<sup>27</sup> These and succeeding figures in this paragraph are from ICE (1990), ICE (1995), and ICE (1998).

<sup>28</sup> Perhaps surprisingly, the mid-1980s—that dark period for American fortunes—was actually the most fertile period in history for the start-up of new semiconductor firms, by a large margin. Most of these new firms were involved in design-intensive custom devices and ASICs (Angel, 1994: 38).



**TABLE 10.4**  
Worldwide Merchant-Market Sales of DRAMs (U.S.\$ million)

<i>Company</i>	<i>Country</i>	<i>1995</i>	<i>1996</i>
Samsung	Korea	6462	4805
NEC	Japan	4740	3175
Hitachi	Japan	4439	2805
Hyundai	Korea	3500	2300
Toshiba	Japan	3725	2235
LG Electronics	Korea	3005	2005
Texas Instruments	United States	3200	1600
Micron	United States	2485	1575
Mitsubishi	Japan	2215	1400
Fujitsu	Japan	2065	1350
Others		4999	1880

*Source:* ICE (1998).

notably the Sematech consortium. These policy responses both arguably had some effect on competition in semiconductors, but the effects were not necessarily the ones expected.

Trade protection came in the form of the Semiconductor Trade Agreement (STA), signed in September, 1986, which established what was effectively a price floor for DRAMs and EPROMs shipped to the United States. The agreement lasted through 1991, and was replaced by a somewhat weaker version that expired in 1996. The price floor catalyzed cartel behavior among Japanese producers by giving them a mechanism with which to coordinate their prices. Prices for DRAMs stabilized by 1986 and began to rise, reaching a peak in 1988–9. The price of EPROMs followed a similar pattern. Industry officials have claimed that the rents in EPROMs generated by the STA enabled Intel to develop the microprocessor line on which its current success rests<sup>29</sup>—and some have even claimed that many of the largest American companies would have gone bankrupt without those rents (Helm, 1995). Constructing counterfactuals is always a tricky

<sup>29</sup> Andrew Grove of Intel has also asserted that the pressure the STA exerted on Japan to increase the penetration of American chips led Japanese personal computer makers to adopt Intel microprocessors, which they might not otherwise have done (Siegmann, 1993).

business, however. What is clear is that the price rise in 1988–9 benefited Japanese DRAM producers at the expense of consumers. One estimate places these “bubble profits” (as they were called in Japan) at U.S.\$3–4 billion (Flamm, 1996: 277).

As with most complex policy interventions, the STA also had some unintended consequences. Early on, critics—and even some proponents of managed trade—pointed out that Japanese firms were plowing their bubble profits into research and development, which would strengthen those firms for further rounds of competition and the much-feared push into other semiconductor markets (Tyson, 1992: 117). Moreover, as Japanese firms are more vertically integrated than American ones, Japanese computer makers would have the advantage of internal transfer prices rather than market prices, giving them an edge over Americans in the computer arena.<sup>30</sup> It is largely this concern, indeed, that led Mowery and Rosenberg (1989: 114) to suggest that, if “the Semiconductor Trade Agreement thus far is an example of successful ‘managed trade,’ it is hard to know what might constitute a failure.”

In the event, however, the DRAM cartel generated a somewhat different set of unintended consequences—consequences much less happy for Japanese firms. By stabilizing DRAM prices and making that market so profitable, the cartel arrangement kept Japanese firms heavily invested in what was to become a low-margin commodity item. When the high prices attracted entry from Korea and Taiwan, prices and profits began to fall, and the cartel collapsed. By contrast, American firms like Intel were arguably well served in the medium term by their failure in DRAMs, a failure that left them free to pursue high-margin logic and specialty chips that would be in high demand by the burgeoning American personal computer market.

As we saw, much popular and professional opinion circa 1985 attributed the relative decline of American competitiveness to the

<sup>30</sup> In fact, this possibility did not materialize, partly because the cartel was short lived and partly because the structural disadvantages of the Japanese computer makers far outweighed any advantages from cheaper DRAMs.

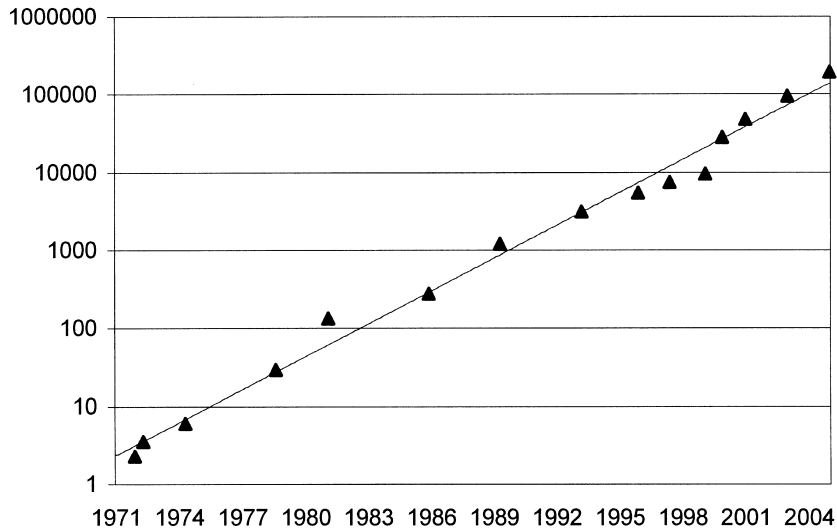
inherent inferiorities of American industrial structure relative to that of Japan. One widely touted aspect of the “Japanese model” was research coordination and collaboration in general and the much-touted VLSI Project specifically. As a result, another facet of the American policy response was an attempt to encourage cooperative research by indirect means as well as by direct subsidy.

Originally motivated by a desire to build a mainframe computer industry to rival IBM, the Very Large Scale Integrated Circuit (VLSI) Project was a pair of programs funded between 1975 and 1981 by the Japanese Ministry of International Trade and Industry (MITI) and by the Japanese telephone monopoly (NTT). These programs called for cooperative research among a number of leading Japanese semiconductor firms with the goal of improving manufacturing technology to challenge American dominance in semiconductors. Although contemporary accounts tended to heap praise on the VLSI project and to assign it most of the credit for Japanese success in DRAMs, recent scholarship has painted a rather different picture (Fransman, 1990; Callon, 1995). In planning the VLSI Project, MITI saw joint organization in a single laboratory as politically valuable, and pressed the companies to agree. This feature has attracted great attention and has been emulated in other consortia designs. It was also a feature that the companies vehemently opposed (Fransman, 1990: 63; Callon, 1995: 57). The companies reluctantly accepted MITI’s joint laboratory organization as the price of the private research subsidies they really wanted (Fransman, 1990: 64). One consequence of the resistance is that only 15–20 percent of the total budget went to the joint laboratories; 80–5 percent went to private research in company laboratories (Fransman, 1990: 80). To the extent that the VLSI project contributed to the improvement in Japanese manufacturing capabilities, it did so by bolstering the capabilities of supplier firms, notably Nikon and Canon in optical lithography (Flamm, 1996: 103).

The VLSI model was directly influential in the creation of research consortia in the United States, notably Sematech. In 1987, the Defense Science Board, a committee advisory

to the American Department of Defense, issued dire warnings that the decline of the American semiconductor industry would have serious repercussions for national defense. The committee proposed a manufacturing facility to be jointly owned by industry and government. In the same year, a committee of the SIA representing fourteen major semiconductor manufacturers issued a proposal for a research consortium to be funded by equal private and federal contributions. By the end of the year, the Defense Department agreed to fund such a consortium, with the fourteen firms uniting as the founding members of the Semiconductor Manufacturing Technology Consortium (Sematech). The organization was funded at a yearly level of U.S.\$100 million from federal sources and U.S.\$100 million from dues assessed to members.

Sematech set up shop in Austin, TX, staffed importantly by personnel on secondment from the member companies. The goal was to develop cutting-edge production technology of use to consortium firms. By 1989, a large-scale semiconductor fabrication facility had been completed at Sematech headquarters in record time. Largely because of problems of appropriability and proprietary information, however, the Sematech members were unable to agree on an appropriate research program for the facility (Grindley et al., 1994: 730). As a result, Sematech quickly reoriented its mission away from developing cutting-edge process technology for and with member companies toward improving the capabilities of the American semiconductor-equipment industry and strengthening cooperation between those firms and the semiconductor manufacturers they serve. This involved “contract R&D” with equipment suppliers, as well as programs to coordinate and set standards, in many cases through the offices of an organization called SEMI/SEMATECH that was set up at Sematech in 1987 to represent equipment makers. As in the case of the Japanese VLSI project, then, the ultimate virtue of Sematech may have lain not so much in the research it produced as in its role in reducing the transaction costs of research dissemination and in fostering closer “vertical” collaboration and coordination between manufacturers and equipment suppliers.



**Figure 10.4** Transistors per chip in microprocessors, 1971–2005 (thousands). *Source:* Intel Corporation, Semiconductor Industry Association (1999).

#### 10.2.5. Coda: Digital Technology and Economic Growth

Without much exaggeration, one could say that the engine of growth within digital technology derives from a single innovation, the planar process, and its logical extension, the integrated circuit. The planar approach to semiconductor fabrication created a technological trajectory of miniaturization that yielded genuinely astounding increases in the number of functions—bits of information stored or number of logical instructions processes—that could be fit on each chip, along with commensurate decreases in cost per function.<sup>31</sup> This phenomenon is encapsulated in the now-famous “law” promulgated by Intel co-founder Gordon Moore: that the number of functions that can be crammed on a chip doubles every 18–24 months.<sup>32</sup> This law of constant doubling time has held true since the beginning of IC technology, and will continue to do for the near future according to

<sup>31</sup> This is so because historically the cost of producing a chip has risen only about a third as fast as the number of functions per chip.

<sup>32</sup> Actually, Moore’s original formulation claimed a doubling time of one year (Moore, 1965).

the “technology roadmap” plotted out by the Semiconductor Industry Association (1999).

Consider the microprocessor. The Intel 4004 of 1971 contained some 2300 transistors. A Pentium III processor from late 1999 contains 28 million transistors. Figure 10.4 plots the number of transistors in a microprocessor over time using historical data for Intel microprocessors and projections from Semiconductor Industry Association (1999). The doubling time works out to a bit less than 26 months.

But there is a demand side as well as a supply side to the story. Moore’s Law is limited by the extent of the market, and, as Moore himself clearly recognized, it takes a “phenomenally elastic market” to soak up all the transistors produced (Moore, 1997). What generated the demand response to the phenomenal cost decline of semiconductors? The answer is in large measure that digital technology offered a variety of general-purpose technologies—technologies that could be adapted to a wide variety of both new and existing uses.

Some of these GPTs are indeed technologies in the narrow sense. The DRAM is a concrete device that can store an infinite variety of information. Others are “technologies” in the wider sense, like the von Neumann stored-program concept, as implemented first in large compu-

ters and then in the microprocessor. This created the possibility of a generic “brain” that could put its mind to an infinite variety of processing tasks. At another level are modular platforms like the IBM 360 or the Wintel platform, which extend demand by allowing consumers assemble exactly the components that best meet their needs. Such platforms also benefit from network effects, another source of increasing returns, as well as from the possibilities for rapid trial-and-error learning when the system is open to competition. The Internet (Mowery and Simcoe, chapter 9) is another general-purpose technology that extends the market for semiconductors and computers.

Economic historians debate whether technological change is really “revolutionary” and whether economic growth depends on such revolutions (Mokyr, 1990a). During the 1980s, this was a question of significance, as the technological manifestations of the digital revolution did not seem to translate into economic growth. In the well-known catch phrase attributed to Nobel laureate Robert Solow, “we see the computers everywhere but in the productivity statistics” (David, 1990: 355). By the end of the century, however, an almost unprecedented decade-long expansion in the United States had erased most remaining doubts about the ability of new technology to drive growth.

Gordon (chapter 3) shows that the acceleration in technical change in computers, peripherals, and semiconductors explains most of the acceleration in overall productivity growth in the American economy since 1995. In part, this acceleration reflects gains in the computer-producing sector: prices of computer hardware (including peripherals) declined at an average rate of 14.7 percent during 1987–95 and at an average rate of 31.2 percent during 1996–9. But the productivity gains also ramified themselves throughout the durable-goods manufacturing sector that uses computers.<sup>33</sup>

As Gordon and other authors in this volume suggest, the productivity gains experienced by the United States in recent years, especially those resulting from the adoption of computer technology, have not been as great elsewhere in the developed world. This suggests that the United States is enjoying considerable macroe-

conomic advantage from the success of its computer and semiconductor industries. As this chapter has argued, America’s early—and more recent—success in both of those industries is related to the codevelopment or coevolution of the two technologies, which led to virtuous cycles of increased productivity leading to increased demand leading to further increases in productivity.<sup>34</sup> Why in the United States? Gordon (chapter 3) discusses some of the general reasons for American success. In the specific context of semiconductors and computers, this chapter has pointed to: (1) the absolute size of the American market; (2) the early role of the federal government as a demander of semiconductors and computers, which gave way to a relatively more laissez-faire role as the technology matured; and (3) the relatively more diverse and open structure of American industry, which allowed for more rapid experimentation and learning than in other countries.

<sup>33</sup> However, it appears that computers had little productivity-enhancing effect outside the durable-goods sector. As Gordon himself hints, and as other economist would insist, this may be because gains in consumer welfare are harder to measure outside the durable-goods sector. One benefit of computers in the consumer sector has been the ability more finely to tailor product characteristics to the tastes of individual consumers—so-called mass customization. To the extent that computers have allowed greater diversity and variety of products, rather than lower prices for existing products, current data-gathering techniques may not register these gains (Cox and Alm, 1998).

<sup>34</sup> Gordon (chapter 3) argues that increases in demand for computers have actually been increases in quantity demanded, that is, movements along a relatively stable demand curve rather than shifts in a demand curve.