Spring

1987

Federal Reserve Bank of San Francisco

Economic Review

Number

2

The Federal Reserve Bank of San Francisco's Economic Review is published quarterly by the Bank's Research and Public Information Department under the supervision of Jack H. Beebe, Senior Vice President and Director of Research. The publication is edited by Gregory J. Tong, with the assistance of Karen Rusk (editorial) and William Rosenthal (graphics).

For free copies of this and other Federal Reserve publications, write or phone the Public Information Department, Federal Reserve Bank of San Francisco, P.O. Box 7702, San Francisco, California 94120. Phone (415) 974-3234.

Competition in the Semiconductor Industry

Randall Johnston Pozdena*

The semiconductor industry has played a key role in international trade disputes. Using data on Dynamic Random Access Memory devices and semiconductor chip fabrication facilities, this analysis examines the behavior of the industry for evidence of the influence of time-related technological change, economies of scale, learning curve behavior, and international differences in strategic pricing behavior. The analysis finds only weak evidence of anti-competitive behavior.

The pace of innovation in the field of electronics has been extremely rapid in the last thirty years, and high technology electronics has been a major source of strength for the American economy. The development of solid state devices — and integrated circuits in particular — has been the major contributor to the startling evolution of this field and the entry of high technology electronics into so many aspects of daily life. In addition, many place their hopes for continued growth of the national and regional economies on intensified innovation in and application of high technology electronics. Along with biotechnology, high technology electronics is seen as a kingpin of the future of the American economy.

The purpose of this paper is to evaluate popular claims that the semiconductor industry is susceptible to anticompetitive behavior, particularly on the part of foreign competitors. Specifically, we will examine the market for a particular integrated circuit (IC) device for evidence of imperfectly competitive performance. Learning and scale economies are found to be significant in this industry, and market structure — while not showing excessive concentration of market share — exhibited rigidity. Combined, these observations are consistent with what one would find where inefficient forms of strategic pricing behavior are practiced.

Production functions associated with integrated circuit fabrication facilities located in the United States and Japan are estimated to provide an insight into the origins of alleged international differences in pricing strategies. Only weak evidence is found to support the notion that Japanese integrated circuit (IC) fabrication costs are below those of their U.S. counterparts.

In Section I of this paper, a brief description of the semiconductor industry and its products is presented. Section II contains a description of the IC production process and an economic characterization of this process. Section III discusses potential implications of the production environment on industry structure and performance. In Section IV, several simple empirical investigations are performed to assess the importance of learning and scale economies in the IC industry and to investigate the origins of differences in U.S. and foreign firm pricing strategies. The paper concludes with a summary of findings and their implications for the future of the U.S. semiconductor industry.

* Assistant Vice President, Federal Reserve Bank of San Francisco. This paper benefitted from the valuable assistance of numerous people. I wish particularly to thank Dan Hutcheson of VLSI, Inc., Ione Ishii of the Semiconductor Industry Association and Lane Mason of Dataquest, Inc. for their generosity in making available much of the data used in this study. The paper also benefitted from the assistance provided by Bill Arnold of the Electronic Materials Report, Inc., Don Larsen of Tektronics, Inc., and Carolyn Rogers of Hambrecht and Quist, Inc. Finally, I wish to thank William M. Robertson for providing excellent and diligent research assistance.

I. The Industry and Its Products

The history of the semiconductor industry, its technology and products are discussed in a number of published sources[1]. It is useful, however, to review the basic features of the industry and its technology both to support the logic of subsequent discussion and to delimit the economic issues we will address.

The semiconductor industry is so named because it produces devices that exploit the special electrical characteristics of a class of natural elements and compounds known as "semiconductors" (such as silicon, germanium and gallium arsenide). The materials have the property that they can be made to behave alternately as conductors or barriers to the flow of electrical current. In the late 1940s, discovery of a means of managing the behavior of semiconductor crystals led to the development of the transistor — a device that uses small currents to control the conduction behavior of the material. Thus, the transistor can form the basis of an amplifier or electronic switch.

Through the 1950s and 1960s, the transistor rapidly replaced the vacuum tube because of its superior ruggedness, smaller size, lower power consumption, and ability to execute tasks more rapidly. Before 1958, functional electronic devices were built by connecting a number of transistors and other electronic components in a discrete manner. Then, two scientists developed the "integrated circuit" or IC, which is a single device combining the functions of a number of transistors. By so doing, ICs opened the possibility of constructing more efficient and compact electronic devices.

The first ICs were produced in commercial volume in the mid-1960s. They are produced by a complex process of etching, "doping" the crystalline material with other elements, and heat-treating the surface of a semiconductor crystal wafer. Today, a 5-inch diameter wafer of silicon can yield one hundred or more "chips", each of which may contain as many as 1 million transistors. Although "discrete" devices are still produced, the IC is now the dominant semiconductor product and has revolutionized industrial and consumer electronic products. In 1985, approximately $16.5 billion in IC shipments were made worldwide, against about $5 billion in shipments of discrete devices[2].

Despite wide variation in the types of functions that ICs can perform, the same basic production process is used in their manufacture. Microprocessors (the "brain" of computational devices), memory devices (for storing information), and a wide variety of standard circuits used in consumer electronics, telecommunications devices, and military hardware all involve similar production procedures[3]. By focusing our attention on ICs in general, and memory devices in particular later in the paper, we hope to make useful generalizations about the semiconductor industry.

II. The Economics of IC Production

The focus of this paper is on factors influencing the structure and future international competitiveness of the American semiconductor industry. We begin with a brief description of the IC production process. Certain aspects of this process are unusual and, when considered in light of U.S. patent law and the alleged industrial policies of foreign competitors, may be important determinants of the structure, performance, and international competitiveness of the American semiconductor industry.

Major Features of the Production Process

The production of integrated circuits involves very large pre-production investment. Such investment takes the form of circuit layout development, development of "maskworks" or templates to imbed the circuitry in the surface of the semiconductor material, and development and testing of prototypes. Because the prototypes often do not behave as modelled during the layout development process, many cycles of the prototype development process may be required before a useful design evolves. This basic circuit design process interacts with the design of the fabrication process and, in some cases, with the design of other chips or "firmware" (programming incorporated into ICs). In total, this preproduction investment may cost as much as $100 million in the case of a new micro-

processor chip[4].

Actual fabrication of the integrated circuits takes place in a fabrication line ("fab line") facility. Wafers of the semiconductor crystal (predominantly silicon) enter one end of the fab line and the fabricated IC exits the production process after various stages of chemical and heat treatment, "dicing" of the wafer into constituent chips, and electrical and physical attachment of the chip to its plug-like base.

It is conventional to describe the capacity and activity levels on a fab line in terms of "wafer starts" per week. The relationship between wafer starts and actual production flow of ICs, however, will depend upon the design of the device being fabricated, the size of the wafer stock, and the efficiency of the fabrication process, which generally is higher on lines with newer vintage fabrication equipment and higher quality labor.

Labor and capital are substitutable to some degree in most of the steps of the fabrication process. Once a fabrication process has been configured, however, significant changes in the process can be costly and time-consuming. Similarly, although a single fab line can, within limits, be used to produce a variety of devices, different types of devices involve different processing steps and sequences, new computer programs to guide those steps, and can involve changes in the degree of cleanliness of the fab line environment. Crossovers to radically different devices, therefore, also are costly[5].

Short-Run and Long-Run Costs of Production

The characterization of IC products and the production process made above can be re-stated in conventional economic terms as follows. First, the product in the IC industry is probably best thought of not as the IC itself, but rather, the units of memory storage, switching, or logical processing functions it provides. Although there are qualitative differences across IC devices providing these various functions (such as access speed in memory devices or the compactness of the IC device that contains them), it is helpful to think of the market as demanding memory storage or other functions rather than ICs *per se*. Then, within gross functional categories at least, the elemental unit of output relates to the fundamental electronic building block of the IC, namely the transistor.

In the short run, fab line capital and the capital representing the design of the IC (the maskworks) are fixed. Output is varied by the firm by manipulating labor and materials inputs. It seems clear that average total short-run costs decline sharply with increased output because of large, fixed maskwork and fab line capital costs. At production levels above the design capacity of a firm's fab line facilities, however, problems of congestion likely arise. Each of the 50 to 100 processing steps takes a finite amount of time and few opportunities exist in the short run to accelerate the processing or to improve the yield of useful output from wafer starts[6]. Thus, in the short run, rising average variable costs likely cause average total costs to rise at high output levels.

In the long run, both fab line capital and maskwork capital are variable, and there are several potential sources of increasing returns to scale. One is that larger fab line facilities offer lower unit fabrication costs than smaller ones. The industry's practice, however, has been to manipulate the number rather than the size of fab lines to alter fab line capacity, suggesting that individual fab line scale is not a major source of economies of scale generally. Of the 1,500 or so fab lines in existence in 1986, two-thirds had design capacities between 1,500 and 4,300 wafer starts per week[7]. If fabrication were an important source of scale economies, its effects, therefore, must be derived from firm-level synergies from operating multiple lines. (The issue of fabrication scale economies is explored further below.)

Increases in the firm's stock of "maskwork capital" also could result in lower long-run average costs. Conceptually, we might view improvements in maskworks and manipulation of processing steps (that is, alteration in the design of the IC) as either an increase in the employment of maskwork "capital" or a change in technology. Technological change is usually assumed to be exogenous to the firm's labor and capital allocation decisions (that is, technical change depends only upon the passage of time) whereas investments in what we are calling "maskwork capital" have been an important component of IC firms' cost-minimization strategy.

Indeed, the commitment of resources to chip (and fabrication process) design is probably responsible

for most of the widely touted, sharp declines in IC product costs that have been observed over time. IC engineers have succeeded in increasing the number of elemental components ("transistors") that can be accommodated by a single semiconductor chip of given physical dimension[8] and, hence, reducing the unit cost of fabricating IC products. (Empirical evidence will be presented in Section IV below on the relative contribution of scale economies and the passage of time to the decline in the cost of IC memory products.)

Technological Diffusion and Learning

Two other aspects of the IC production environment are relevant to understanding the current and likely future performance of the IC industry. The first is that property rights in "maskwork capital" historically have been poorly defined, making it difficult for one firm to prevent access to the fruits of its investment by other firms. It is relatively easy to reconstruct the design and manufacturing steps involved in an IC product through a process known as "reverse engineering"[9]. By a sequence of photographic analysis, disassembly by etching, and materials analysis, a rival firm can reconstruct the architecture of a functional chip and the maskworks and processing steps necessary to reproduce it. Such "reverse engineering" can cost as little as one-one thousandth of the original firm's investment[10] and permit "pirate" firms to enjoy lower total costs. The passage of the Semiconductor Chip Protection Act of 1984 foreclosed the possibility of precise "cloning" of maskwork capital by foreign or domestic competitors, although the more general practice of reverse engineering remains legal[11].

A second often-cited feature of the IC industry is the relevance of "learning curve" phenomena to IC production. The notion is simply that the cost of production may be related not only to the rate of output of a firm (economies of scale) and changes in technology over time but also to the independent effect of accumulated production experience. Such a phenomenon is considered to be relevant to complex manufacturing technologies: as output experience increases, the firm better understands the technology involved and technical efficiency increases[12].

Since integrated circuit manufacturing is an extremely complicated technical process, it seems likely *a priori* that learning-related cost adjustments may occur. The implications of learning phenomena and a test for their existence are presented below.

To summarize, the IC production process is a highly technical one involving large investments in maskwork capital that are difficult to recover if an enterprise fails and difficult to protect from exploitation by other firms. Potentially significant economies of scale are likely, and probably flow mainly from economies at the level of the firm rather than the plant (that is, the fab line). Costs also may decline over time because of technological progress and with accumulated output experience because of "learning curve" phenomena.

III. Implications for Industry Structure and Performance

The preceding discussion of the economic characteristics of the IC industry may help explain the likely structure and behavior of the industry, particularly whether the industry exhibits characteristics that may make it vulnerable to anticompetitive behavior. It has been widely alleged, for example, that Japanese producers have pursued predatory pricing strategies in certain IC products[13]. In this section, the implications of the postulated economic characteristics for structure and performance are discussed as a prelude to attempts at empirical verification.

Scale Economies and Contestability

The economic characteristics of the IC industry make it likely that the production of ICs is characterized by economies of scale. Significant scale economies, in turn, would mean that markets for IC products will tend to be concentrated in the long run, and thus have the potential for inefficiency.

Baumol and Bailey[14], however, have argued that high levels of concentration (or even monopoly) in production need not have serious effects on market efficiency if the market were "contestable". For an

industry to be considered contestable (in the sense that Bailey and Baumol use the term), it must be possible for new firms to enter a market displaying abnormal profits and earn normal profits or — if extant firms cut prices to thwart the new entry — leave the industry without losing the investment associated with entry.

At least one attribute of the IC industry suggests that it may not be ideally contestable: a major cost of entry — preproduction research and development — is difficult to recover if the firm is unsuccessful in competing against extant producers and must exit the market. In this respect, the IC industry contrasts with most manufacturing, transportation and service industries for which acquisition of re-sellable, fixed assets is the dominant cost of entering a market. Thus, "contesting" for markets may not be an effective means of imposing competitive discipline within the IC industry. This makes it especially important to investigate that industry's scale economies.

Learning and IC Market Efficiency

In Section I, we also postulated that IC production occurs in the presence of a learning curve. The existence of learning effects on unit production costs may have a bearing on both industry structure and pricing behavior and, thereby, on the efficiency of the IC industry. The logic of these effects in a model of dynamic entry and pricing behavior has been demonstrated rigorously by Spence[15]. The implications of his model will only be summarized briefly here.

First, if learning (production experience) reduces costs, Spence has demonstrated that, under certain theoretical conditions, learning can confer some protection from competitive entry to the first firm into a market, in effect, simulating an entry barrier. If "first movers" do enjoy such advantages in the IC industry, then market structure might be expected to be rigid over time — that is, show little change in the rank and share of firms in the market.

The second aspect of learning curve theory of interest here are the effects of the learning curve on pricing behavior. In essence, because production experience confers subsequent cost advantages on the firm, a firm maximizing long run profits in the presence of a learning curve will charge less (and produce more) in the learning stage of production than dictated by short run profit maximization considerations alone.

Not all environments are conducive to such learning curve pricing behavior, however. Whereas firms in an industry composed of just a few firms are able to exploit a learning-curve pricing strategy, Spence argues that strategy is less effective in unconcentrated production environments where there is assertion of competitive price discipline with successful entry.

In addition, whether learning is important at all to either pricing behavior or market structure depends upon how "rapidly" learning takes place. If learning were very rapid (that is, the effects of accumulated production experience are small relative to the effects of current production on costs), then there would be few strategic advantages to deviating from short run profit maximization. A related point concerns how rapidly learning reaches other firms. If such diffusion is very rapid, accumulated output might help explain cost and price trends for the industry as a whole, but current prices would be determined by current costs.

The implications of Spence's view for the IC industry might be summarized as follows. *If* learning (cumulative output) were important to firms in the IC industry, then early entrants able to survive initial periods of low prices might gain an (at least temporary) advantage over later entrants. Market structure would be less fluid than otherwise, and such "first movers" could enjoy higher profits than subsequent rivals. Firms, in turn, would have a strategic incentive to pursue an early-entry strategy.

International Competition in the IC Industry

For American IC producers, concerns over preproduction costs and scale economies, contestability, and learning phenomena seem to be at the root of current debates over the marketing strategies of their Japanese competitors. Japanese producers have gained a growing share of the world semiconductor market; their share of combined U.S.-Japanese production has risen from 33 percent in 1971 to over 50 percent in 1982. Since 1982, the U.S. share of world IC sales has fallen from about 60 percent to 50 percent, while the Japanese share has risen from 30 to about 40 percent[15a].

It is frequently alleged that the Japanese have obtained their growing share of IC sales by pursuing "predatory" pricing strategies. In particular, Japanese IC manufacturers have been accused of selling IC products in world markets at prices below their cost of production. In 1986, for example, there were three major International Trade Commission complaints alleging such behavior filed by the American IC industry and the U.S. government[16].

If true, one explanation for such pricing behavior would be the existence of exploitable learning curve advantages, although any IC firm — not only the Japanese ones — could exploit those advantages. Nevertheless, it is argued frequently that Japanese IC producers are better able to survive the early periods of low profitability necessary to secure market dominance in a "learning" dominated production environment. They are alleged to benefit from their affiliation with conglomerate manufacturing organizations, which underwrite early periods of low profitability, and the availability of subsidies from the Ministry of International Trade and Industry (MITI) and the banking industry[17]. Whether such subsidization occurs (or differs dramatically from support the U.S. IC industry has received from the military) has been debated extensively[18].

A second alleged reason for the growth in Japanese IC market share is that low technological and legal barriers to copying U.S. designs and processes have unfairly reduced the total costs faced by Japanese producers. Particularly egregious cases of apparent cloning indeed can be documented[19]. In addition, the property rights traditionally extended by the Japanese to foreign creators of intellectual property have been criticized as being weak by international standards. In the debate over software copyright reform from 1983 to 1985, for example, the Japanese proposed standards of protection were weaker than both international copyright standards and the standards applied to domestic (Japanese) copyrights[20].

As Dasgupta and Stiglitz[21] have pointed out, ill-defined intellectual property rights can reduce innovation below the socially optimal level. However, for low barriers to cloning to have a permanent effect favoring Japanese over U.S. production, the ability to "reverse engineer" a competitor's product must be asymmetric internationally which it is not, and some other factor must operate to "cement" market dominance once dominance is achieved by this means.

Finally, it is possible that growth in the Japanese IC market share flows from legitimate differences in fabrication cost. These differences could arise from lower costs for labor of a given quality or superior Japanese management of fabrication facilities. (It is unlikely that differences in materials or equipment costs would be as important since most of the wafer stock and fab line equipment has been manufactured by one country: the U.S.) The theory of international factor price equalization[22] argues against the lower labor cost argument, but foreign producers could still have the comparative advantage if they have a greater endowment of relevant production factors[23].

IV. Empirical Examination of the IC Industry

Several empirical investigations may inform our understanding of the structure and performance of the IC industry. First, a study of market share rigidity may shed light on the structure of the IC market. Market shares that appear to be rigid over time might indicate that the market is not easily contestable, or that learning phenomena operate to retard entry.

Second, we could test directly for the existence of learning phenomena by examining the response of costs to cumulative firm output — costs should decline with accumulated output experience. In addition, if a high degree of market concentration were associated with *lower* IC prices (an association not normally expected except in the presence of a learning curve or other strategic pricing considerations), learning in the IC industry would be more likely to be firm-specific (that is, it would not diffuse so rapidly that it did not influence firm behavior).

Third, the relationship between the scale of production and cost also would be of interest. If scale economies were not extremely great, the industry would be less likely to be concentrated, and the potential distortions caused by lack of contestability or other constraints on the fluidity of the industry

would be less important.

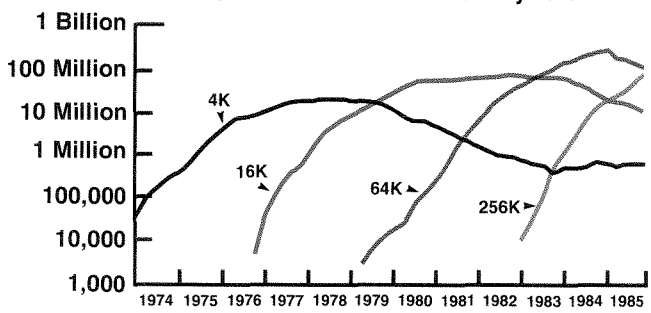
Finally, to address concerns about the behavior of foreign competitors, it would be helpful to compare fabrication costs in Japanese and U.S. facilities. If Japanese IC prices were below their American counterparts' but fabrication costs were the same, we might have evidence that Japanese producers were pricing below the full cost of production, including the cost of maskwork capital.

Unfortunately, the data available on the IC industry are not ideal for examining all of these relationships. No cost data are available and they would be suspect in any case, since the cost of producing a specific product is difficult to extract in a conglomerate enterprise. This is a particularly serious shortcoming in studying Japanese IC production costs. Fairly good price and output data are available by firm and device, however, as are data on the labor and capital employed on individual fabrication lines. In what follows, these data are exploited to provide rough information on the relationships of interest.

The Behavior of the DRAM Industry

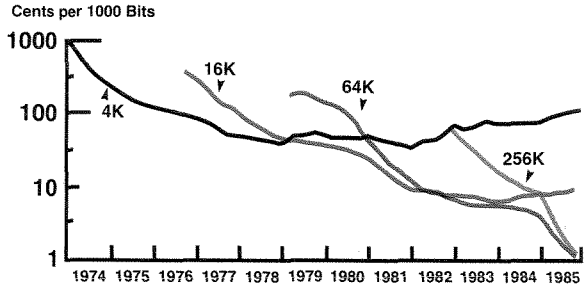
We explore here the issues of market structure rigidity, learning effects, and scale economies in the context of a particular type of IC device — the Dynamic Random Access Memory (DRAM) IC. This device stores binary bits of information in a randomly accessible manner. (The term "dynamic" simply refers to the requirement that DRAMs be powered continuously to retain implanted memory. The term distinguishes them from a related device — the Static RAM — that does not need continuous electrical power.) The memory capacity of DRAMs is measured in kilobits; each kilobit is 1,024 individual bits of memory capacity and is abbreviated by a "K". To date, DRAM devices have been

Chart 1A
Number of DRAM Devices Sold by Quarter



Source: Dataquest

Chart 1B
Time Trend in DRAM Prices



Source: Dataquest

manufactured in commercial volume in 4K, 16K, 64K and 256K capacities.

We focus on DRAMs for a number of reasons even though they represented only about 10 percent of total IC sales worldwide in 1985. First, the DRAM device is as close as the semiconductor industry gets to a "commodity"-type of device. Most other ICs have qualitative attributes that make them difficult to study over time or across firms. Second, unlike microprocessor ICs for example, DRAMs have been produced in significant volumes by non-U.S. firms, allowing some exploration of the influence of foreign entry on industry behavior. Indeed, DRAMs were involved in recent allegations of "dumping" by the Japanese[24]. Finally, as a practical matter, to expand sample sizes, it is necessary to combine data across devices. Such a combination is feasible with memory devices because they are unambiguously "generic" in their essential unit of service (the "bit"), and bits are substitutable across devices.

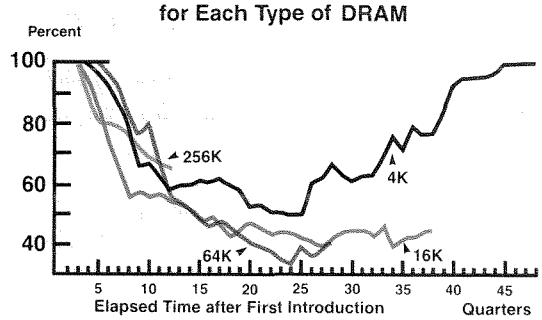
Substitutability across types of DRAMs is illustrated in Chart 1, which shows the actual quantity shipped in Panel A and actual prices per bit for four DRAM devices in Panel B. The sales of the 4K DRAM, for example, peak and decline sharply (note that all quantities are in log terms) when the price per bit of the successor device (16K DRAM) falls below the 4K price per bit. A similar pattern holds for subsequent generations of devices. The chart also illustrates vividly the observation made earlier that increasing the bit density on the chip has contributed importantly to the observed declines in price per bit of DRAM memory.

Market Rigidity

In examining the DRAM market for evidence of structural rigidity, it is instructive to trace the evolution of market structure in DRAM manufacture. As Chart 2 reveals, a new DRAM device typically is introduced by one or two firms with entry occurring gradually until concentration (as measured by the share of the market held by the largest 3 firms) declines to a relatively modest level.

In the cases of the early devices (such as the 4K DRAM), entry occurred more gradually than with subsequent generations of devices, and concentration levels did not decline below the 50 percent

Chart 2
3-Firm Concentration Ratios
for Each Type of DRAM



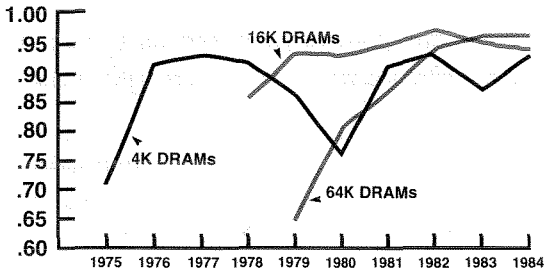
Source: Dataquest, FRBSF

level. These findings probably are consistent with the existence of important scale economies and, therefore, limited "room" in the market for additional firms. Indeed, as the market for 4K devices matured and declined, the concentration ratio for 4K DRAMS gradually increased as firms exited the market.

Although firm entry into the market for DRAMs appears capable of reducing levels of industry concentration for at least short periods of time, new entrants have difficulty dislodging "first movers". To illustrate this observation, the orderings of firms ranked by market share from one year to the next were compared and a statistic measuring the correlation of these ranks — the Spearman Rank Correlation Coefficient — was computed for each pair of adjacent years and for each device. (Unchanged year-to-year rank ordering produces a Spearman rank correlation coefficient of 1.0)[25]. The results are presented in Chart 3.

The high correlation of firm market share rankings from one year to the next suggests that DRAM market structure is not highly fluid after the initial period of entry, and has with each successive DRAM device reached this condition of structural stability more rapidly. Moreover, data not presented in the chart suggest that the first producers of a device not only retain pre-eminence in the market for that device but often are "first movers" into production of the next generation device. Turnover of producers is greatest among those firms that are not first entrants. These observations may suggest the existence of high fixed costs (either pre-produc-

Chart 3
Time Trend in Market Structure Rigidity as Measured by the Spearman Correlation of Market Share



Source: Dataquest, FRBSF

tion or production) or learning phenomena that benefit incumbent firms.

Scale Economies and Learning Effects

Without specific data on IC production costs, it is not possible to test directly for the existence and importance of learning in DRAM manufacture, or to explore directly the magnitude of scale economies. However, an examination of the behavior of DRAM *prices* in addition to the inferences drawn above from the behavior of market structure may shed some light on cost behavior. In particular, except in instances of coordinated or monopoly pricing, prices and average costs are likely to move together over time. This relationship suggests that some inferences about costs can be derived from price data if the circumstances that might lead to noncompetitive pricing can be controlled.

To explore these relationships more formally (and to control for the effects of the passage of time, accumulated production experience, and market structure), we studied a simple econometric relationship using quarterly data on DRAM devices over the 1976 to 1985 period. We studied the variable the price of a *bit* of DRAM, and we pooled time series data on four devices (the 4K, 16K, 64K and 256K DRAM) to expand the sample size. The general form of the relationship studied was:

$$\text{Price}(x,t) = h[\text{time}(x,t), \text{output}(x,t), \text{cumulated output}(x,t), \text{market structure}(x,t), \text{Japanese market share}(x,t) \text{ and } \text{size}(x,t)] \quad (1)$$

where

price(x,t) = price per bit, in U.S. dollars for device x at time t

time(x,t) = t = the date of the price observation on device x

output(x,t) = average industrywide output per firm (in number of devices), for device x at time t

cumulative output(x,t) = average device output per firm summed over time by device, for device x at time t

market structure(x,t) = 3-firm concentration ratio in market shares of device x at time t

Japanese market share(x,t) = fraction of total production by Japanese firms, for device x at time t

size(x,t) = the memory capacity in bits for device x at time t

and

x = 1, 2, 3, 4

t = 1 to 44 measured from 1974, quarter one

Table 1 presents a regression analysis of a specific configuration of Equation 1. In particular, ordinary least squares were used to estimate a relationship between the log of DRAM prices and the variables identified in Equation 1. Most of the coefficients are made a function of device size by interacting each variable with the device size measure. All of the coefficients in Table 1 are measured with considerable statistical precision, permitting a number of interesting observations.

First, the positive coefficient on the Size variable suggests that (if prices follow costs) increases in device size increase the cost of a bit of DRAM. That is, with a given technology, it is more costly (per bit) to produce large rather than small devices. The negative coefficients on Time and Time times Size, however, indicate that technological progress decreases bit price and that technological progress has been most important for large devices. The positive coefficients on Time Squared and Time Squared times Size indicate that the influence of technological change in reducing the bit price are diminishing with time and with increased device size over time. This is consistent with the idea of diminishing returns to size-related innovations and innovations generally.

Second, the negative coefficient on Output and the positive coefficient on Output Squared implies that the price per bit declines and then rises with increased output. This is consistent with the notion that there are economies of scale associated with IC production over *some* range. The rate of firm output at which prices begin to rise can be derived from the estimated coefficients and is 5.4 million devices per quarter for a 64 kilobit device. This rate is over four times the observed average output and close to the maximum of 8.53 million devices per quarter observed in the sample. It thus offers some comfort that the use of prices to study cost behavior may not be unreasonable.

Third, the negative coefficient on Cumulative Output indicates that, at least for the industry as a whole, cumulative output has an effect independent of that of current output. This finding supports the hypothesis that learning contributes to IC production behavior. Although it is not evident that

TABLE 1

Estimating the Effects of Market Structure, Learning, and Scale Economies on the Price of DRAMs

(Equation 1)

Dependent Variable: Natural log of DRAM price per bit (in $)

Variable	Estimated Coefficient*	T-ratio**
Constant	6.95	27.2
Size	.000156	9.66
Time	− .0741	6.47
Time Squared	.000931	4.45
Size x time	−7.12 D-06	9.16
Size x time squared	6.91 D-08	7.34
Output (devices per firm)	− .000154	2.42
Output Squared	2.87 D-08	3.44
Cumulative Output	−9.56 D-12	4.76
3-firm Concentration Ratio	1.28	5.29
Size x 3-firm Concentration Ratio	−2.79 D-05	4.90
Japanese Market Share	− .0282	10.57
Size x Japanese Market Share	4.94 D-07	7.64

Derived Estimates

(64K device)

Elasticity of price with respect to cumulative output:	−.08
with respect to 3-firm conc. ratio:	−.35
with respect to Japan market share:	+.13

Point at which increased output is associated with increase in DRAM prices: 5.4 million devices per quarter

R squared	.954
n	154

*Coefficients estimated using ordinary least squares. Where necessary the estimates are presented in scientific notion, with base-10 exponents.

**All of the estimated coefficients are statistically distinguishable from zero at the 95 percent confidence level or better.

Source: Dataquest, Inc., Federal Reserve Bank of San Francisco

individual firms are able to exploit learning curve phenomena in devising pricing strategies, it does indicate that the theoretical potential to do so exists.

Fourth, the positive coefficient on the Concentration Ratio and the negative coefficient on Size times the Concentration Ratio suggests that the influence of market share concentration on DRAM prices is positive for small devices, but negative for larger devices. (The effect becomes negative for a "size" greater than about 40 kilobits.) As was observed earlier in this paper, a negative effect of market share concentration on price is suggestive of firm-specific strategic pricing in the presence of a learning curve. The negative coefficient for 64K and 256K devices is interesting because these are the device sizes that came to be dominated by the Japanese, suggesting, perhaps, that the Japanese introduced a different pricing strategy into the DRAM IC market.

The independent effect of Japanese market presence on DRAM prices is given by the negative coefficient on Japanese Market Share and the positive coefficient on Size times Japan Market Share. It appears that, controlling for other market and production influences, the effect of the Japanese presence was to reduce prices for smaller devices (empirically, smaller than 57 kilobits), but to elevate prices for larger devices. Once again, the markets for larger devices are ones that the Japanese are said to dominate. Since we have controlled for industrywide learning by Cumulative Output, and early strategic "underpricing" may be accommodated by the concentration ratio variables, this finding could be seen as evidence that prices can be elevated by successfully dominating the market.

There are a number of important qualifications to these findings. An obvious difficulty lies with the study of average prices and average firm output to make inferences about what inherently is relevant only to individual firm costs and output. Many relationships that hold at an individual firm level are not appropriately aggregated or averaged. A similar criticism attaches to the use of average cumulative output instead of individual firm data to detect the presence of learning curve phenomena. More complex functional forms, separate specification of the estimated relationship for each device, and recognition of qualitative differences across devices also would be useful. The available data, however, does not allow us to resolve all of these potential sources of bias.

A Study of Fabrication Facilities

Our second empirical investigation focused on the fundamental unit of fabrication: the fab line. This investigation is of interest both to verify the casual observation made in Section II about the likely lack of scale economies in fabrication and to study differences in fabrication activity on fab lines operated by Japanese and American firms. If fabrication costs were the same in both countries, then such costs would have to be eliminated as a source of differences in pricing strategies.

Once again, quite severe data limitations restrict the type and quality of analysis that can be performed. Fab line cost data are not available; only data on installed capital equipment (in 1986 U.S. dollars), the number of employees engaged on the line, and wafer start activity are available. Data on the specific type of IC device produced on the line also are not available[26]. Nevertheless, the data do permit two simple empirical tests — a comparison of capital-labor ratios and a comparison of fab line production functions.

First, using data on 386 fab lines from June 1986, we computed the capital-labor ratios separately for American and Japanese fab lines[27], and found the ratio in Japanese fab lines to be approximately 2 percent less than the ratio on American lines[28]. This finding is likely an understatement of the actual difference because the Japanese work week is one day longer than the 5-day U.S. standard. Thus, for Japanese lines, the number of employees on the fab line is a downward-biased measure of labor input flows (in man-days).

The finding of a 2 percent difference implies that, assuming the same fabrication technology and quality of labor, the *unit* cost of labor relative to capital is lower for Japanese than American firms. Since the capital equipment costs likely are very similar (since much of the equipment is American in origin), this, in turn, could be consistent with the existence of absolutely lower labor factor costs in Japan [29].

The second use of the data was to estimate a fab line production function directly for a combined

sample of Japanese and American fab lines. The estimated functional form was the Cobb-Douglas representation of the production function:

$$Q = aL^bK^c \quad (2)$$

where

- Q = fab line output (measured in square inches of wafer starts per week)
- L = the number of fab line employees
- K = the dollar value of fab line capital (in millions of U.S. dollars).

The Cobb-Douglas representation has a number of well-known limitations, the most important of which is that the rate of substitution between factors is constrained to be equal to one. It has the advantage, however, that the exponents of labor and capital provide estimates of the marginal products of these respective factors. It also can be shown that the relative sizes of these two coefficients are related to the relative contribution of each factor to total product under certain assumptions, and that the sum of these coefficients is a measure of economies of scale[30]. (Specifically, if b plus c in equation 2 is greater than one, the production function exhibits economies of scale; if they sum to less than one, there are diseconomies to large scale production.) In addition, with simple assumptions about factor prices and the profit-maximizing behavior of the firm, cost functions can be derived[31].

The estimates of the coefficients of equation 2 were obtained by taking the logarithm of both sides and using ordinary least squares regression techniques. The results are presented in Table 2, with a dummy variable introduced to identify possible coefficient differences between American and Japanese fab lines. The coefficients on fab line employment and fab line capital are, respectively, .54 and .35 for the American fab lines. It thus appears that there are no scale economies from fabrication *per se*. Indeed, the point estimate of the coefficients imply very slight *diseconomies* of scale (.89 versus 1.00 for constant returns to scale) although the estimate cannot be distinguished from 1.0 statis-

TABLE 2

Estimating the Fab Line Production Function

(Equation 2)

Dependent Variable: Log of Achieved Wafer Output (in thousands of square inches per week)

Variable	Coefficient	t-ratio
Constant	3.59	(7.08)*
Country dummy	.84	(0.67)
Fab line employment	.54	(5.95)*
Fab line employment times Country dummy	.25	(1.16)
Fab line capital	.35	(3.62)*
Fab line capital times Country dummy	−.04	(0.17)
R-squared	.46	
n	381	

Country dummy equals 1 if a Japanese fab line, 0 otherwise.
Fab line employment is in thousands of employees, in log terms.
Fab line capital is in millions of 1986 US dollars, in log terms.
Asterisk (*) indicates that the estimated coefficient differs from zero at the 90 percent confidence level or better.

Data souce: VLSI, Inc., San Jose, CA

tically. The importance of the labor component of fabrication inputs also is illustrated by the estimates, which show that labor's contribution to total product is approximately fifty percent greater than that of capital[32].

American vs Japanese Production Functions

There also appear to be no differences in fab line production functions between American and Japanese facilities. All three of the coefficients on the variables designed to capture these differences (that is, the Japanese dummy variable and its interaction with fabrication employment and fabrication capital) are not statistically different from zero. From this evidence alone, there is little to suggest that fabrication economies can explain differences in final product prices between the two countries. In other words, if manufacturers in each country were profit-maximizing and faced the same labor and capital costs, there is no statistical evidence that total fabrication cost relationships would differ[33].

Important qualifications on this finding must be offered, however. First, the output measure used in estimating equation 2 is not the number of usable ICs completed but simply square inches of wafer processed. To the extent that Japanese and American firms differ in their ability to recover usable ICs from this process, effective cost per IC would differ[34].

Second, Japanese and American producers may emphasize different products not accommodated by the simple production function estimated here. Some data are available on coarse product categories associated with each fab line. However, the use of separate dummy variables for these categories did not significantly influence the estimated parameters, perhaps because the sample sizes for some of those product variations were small.

Finally, if the differences in capital-labor ratios on fab lines observed earlier in the two countries is indicative of absolute differences in labor and capital costs in the two countries, production costs would differ accordingly.

V. Conclusions

Although the available data do not permit definitive assessment of the factors that may affect firm behavior within the IC industry, some light has been shed on two major aspects of the performance of this industry. The first is whether the industry is prone to high levels of market share concentration or other features that may result in inefficient performance.

Pricing behavior was consistent with sizeable, but not extreme, overall scale economies, which include pre-production costs. Market structure in commodity-type DRAM devices appears to be concentrated only at low industry output levels.

Pricing behavior in the DRAM market is, however, consistent with the existence of a firm-specific learning effect. In addition, large, "sunk" pre-production investments are required to enter new device markets. Both phenomena would tend to give strategic advantages to incumbent firms and hence to firms (of any nationality) that might be supported through periods of negative earnings while they acquire the advantages of production experience and incumbency.

The second major issue confronting the IC industry is the conduct of foreign competitors compared to U.S. manufacturers. The fact that estimated fab line production functions did not uncover significant intercountry differences casts some doubt on differences in fabrication costs as a source of competitive advantage for Japanese producers.

Weighing against this view is our finding (in the context of fabrication lines) that firms in Japan behave as if their labor is less costly (or of lower quality) relative to capital than firms elsewhere. Since the conventional wisdom is that Japanese labor is not of lower quality, the behavior of Japanese firms argues in favor of a cost-advantage to Japanese production of ICs. Perpetuation of this disparity runs counter to the notion of international factor price equilibration predicted by trade theory, but without additional information, the argument that cost differences are the basis for the growing presence of the Japanese in the IC market must stand.

The Future of the U.S. Semiconductor Industry

Several of our findings suggest that semiconductor markets lost to foreign competition may be difficult to recover. For one, assuming that production experience confers cost-advantages on a firm and that a growing scale of "sunk" costs is associated with *de novo* entry, it follows that new nonsubsidized firms will find it increasingly difficult to dislodge incumbent firms, whether that incumbency was achieved through cost-advantages or subsidized operation. Second, although the DRAM market, at least, is not especially concentrated at this time, markets in DRAMs have tended to become rigidly structured over time.

On a more positive note, two recent policy changes have important implications for American firms. First, the passage of the Semiconductor Chip Protection Act, by giving property rights to designers of semiconductor chip maskworks, should reduce significantly the more egregious piracy practices. The Act will reduce the likelihood that different firms will face different effective costs of maskwork capital. If, as is popularly alleged, foreign firms previously acquired maskwork capital through reverse engineering at the expense of American firms, the improvement in property rights in IC maskworks should benefit American IC producers.

The second important policy initiative is the 1986 agreement reached between the U.S. and Japanese governments regarding, among other things, international semiconductor pricing policy. The agreement was reached by negotiation between the U.S. and Japanese governments to resolve complaints about Japanese IC pricing policy brought before the U.S. International Trade Commission and in petitions filed under Section 301 of the Trade Act of 1974[35].

The ITC complaints and petitions were dropped in return for agreements from the Japanese to cease the alleged practices of (1) retarding U.S. entry into Japanese markets and (2) "dumping" of Japanese products below cost in U.S. markets. In particular, the agreement provides firms in both countries protection against "subsidized" sale of semiconductors (priced below "company specific cost of production plus 8 percent"). Also, as part of the agreement, the Japanese government is charged with monitoring the relationship between firm costs and selling prices abroad.

The agreement potentially could provide a forum for resolving the debate about whether the Japanese producers are, in fact, subsidizing IC production. To the extent that the concept "company specific cost of production" is meant to refer to short-run average costs, the agreement also could retard learning curve pricing strategies and thereby improve prospects for American firms. At this writing, however, the agreement had broken down because of the alleged failure of the Japanese government to enforce its terms.

FOOTNOTES

1. See, for example, "The Solid State Era," Electronics, April 1980, M. S. Kiver, Transistor and Integrated Electronics, McGraw-Hill, 1972, and W. C. Hittinger, "Metal Oxide Semiconductor Technology," *Scientific American,* August 1973, pp. 48-57.

2. World Semiconductor Trade Statistics Committee, Semiconductor Industry Association, "Semiconductor Forecast Summary," September 1986.

3. Excellent summaries of the technical relationships among IC and other semiconductor devices are available in the annual reports of VLSI, Inc., a San Jose, California, research firm.

4. See, for example, the F. Thomas Dunlap, Prepared Statement, "The Semiconductor Chip Protection Act," USGPO, J-98-39, pp. 152-168, and "Intel's Development of 386 Chip Took 4 Years and $100 Million," *Wall Street Journal,* August 29, 1986, p. 4.

5. See footnote 3.

6. Wafer processing into ICs involves such steps as application and baking of special coating materials, deposition or doping selected portions of the surface, etching and numerous measurement and testing steps. The speed of each of these steps is not easily accelerated. In addition, the wafer is carried from one step to the next mechanically and production takes place, for many of the critical steps, in hoods or rooms with highly processed atmospheres. It is difficult to add equipment, work stations, or speed up processing at will in such a confined and sequence-driven environment.

7. These data refer to MOS technology fab lines. The source of the data is VLSI, Inc.

8. Specifically, the original contact photolithographic techniques for transferring circuit designs to the surface of the IC have been improved first through optical projection techniques and today, electron beam lithography techniques.

9. The prepared statement of the Semiconductor Industry Association, ibid, pp. 122-128, discusses the reverse engineering process.

10. From the testimony of F. Thomas Dunlap. See footnote 4.

11. Karen Ammer, "The Semiconductor Chip Protection Act of 1984," *Law and Policy in International Business,* Vol 17, 1985, pp. 395-420. See also, J. Chesser, "Semiconductor Chip Protection: Changing Roles for Copyright and Competition," *Virginia Law Review,* Vol 71, 1985, pp. 249-285.

12. This notion was studied in an early Federal Trade Commission report and studies performed by the Boston Consulting Group. See "Staff Report on the Semiconductor Industry," Bureau of Economics, Federal Trade Commission, January 1977 and "Perspectives on Experience," Boston Consulting Group, 1972.

13. See, for example, the complaint before the International Trade Commission regarding 256K DRAMs (ITC Docket number 731-TA-300).

14. See E. Bailey and W. Baumol, "Deregulation and the Theory of Contestable Markets," *Yale Journal on Regulation,* Vol 1, 1984, pp. 111-137.

15. A. Michael Spence, "The Learning Curve and Competition," *The Bell Journal of Economics,* Spring 1981, pp. 49-70. For an application of the learning curve evaluation process to another industry, see, for example, M. B. Lieberman, "The Learning Curve and Pricing in the Chemical Industries," *The Rand Journal of Economics,* Summer 1984, 213-239.

15a. Statistics on world semiconductor sales shares are not known with precision. The statistics on US and Japanese *semiconductor* market shares are from "International Competitiveness in Electronics," US Office of Technology Assessment, Washington, DC, 1983. The data on world IC shares are from "For Chipmakers, the Game has a New Set of Rules," *Business Week,* January 13, 1986, p. 90. The trade balance in ICs also reflects Japanese competitiveness. In 1979, the US/Japan trade balance in ICs was a surplus of $90 million. By 1984, this had deteriorated to a deficit of $884 million. (Source: *Industry Week,* November 25, 1985.)

16. These involved 64K DRAMS (ITC Docket #731-TA-270), 256K DRAMs (ITC Docket #731-TA-300) and EPROMs (ITC Docket #731-TA-288). DRAMs are Dynamic Random Access Memory ICs and EPROMs are Erasable Programmable Read Only Memory ICs.

17. See, for example, "The Effect of Government Targeting on World Semiconductor Competition: A Case History of Japanese Industrial Strategy and Its Costs for America," Semiconductor Industry Association, 1983.

18. This debate is presented in Okimoto, Sugano and Weinstein, eds, *The Competitive Edge: The Semiconductor Industry in the US and Japan,* Stanford University Press, 1984.

19. In one seemingly egregious instance, an American chip design was reproduced by a Japanese producer so precisely that even a microscopic error in the chip architecture was replicated. See the document cited in footnote 4 above.

20. See J. Chesser, ibid.

21. P. Dasgupta and J. Sitglitz, "Uncertainty, Industrial Structure, and the Speed of R&D," *The Bell Journal of Economics,* Spring 1984. See also J. E. Tilton, "International Diffusion of Technology: The Case of Semiconductors," Studies in the Regulation of Economic Activity, The Brookings Institution (Washington, DC), pp. 24-38. Gort and Konakayama study the diffusion of the production of an innovation using a simple model and several industries, including the transistor industry. See Gort and Konakayama, "A Model of Diffusion in the Production of an Innovation," *American Economic Review,* December 1982, pp. 1111-1119.

22. Miltiades Chacholiades, *International Trade Theory and Policy,* McGraw-Hill, 1978, Chapter 10.

23. M. Chacholiades, Chapter 11.

24. Specifically in the case of 64K and 256K DRAMs. See footnote 16 above.

25. The Spearman Rank Correlation Coefficient measures the degree of correspondence between two series of numbers by examining rank differences. This statistic is useful for studies of market rigidity because in a fluid

market, market shares held by individual firms would change over time as new entrants disturbed market shares and place-switching occurred among extant firms. The fact that the Spearman Correlations observed in the DRAM market are high is illustrated by a few simple examples. Assume that an industry is structured so that the largest firm has 30 percent market share, followed by firms with 25, 20, 15, 10, 5, and 2.5 percent respectively. If the first and fourth firms switch places, the correlation between the old and new structure would be only 67 percent. If, instead, the smallest firm exits the market (with the next smallest firm taking up its market share), the correlation would be 92 percent. Both of these correlations are less than the year-to-year correlation observed in a mature DRAM market.

26. The source of this data was VLSI, Inc., San Jose, California. The data are *not* available from this author, as per agreement with VLSI, Inc.

27. The sample was confined to metal oxide semiconductor technology lines for comparability.

28. The difference is of only marginal statistical significance. Specifically, the difference is different from zero at approximately the 80 percent confidence level.

29. The capital-labor ratio also could be affected by differences in the price of capital. Indeed, it is argued frequently that Japanese producers have access to subsidized financial capital. If true, the reduction in the capital-labor ratio stimulated by lower labor costs could be offset by lower user costs of fab line capital.

30. See, for example, H. R. Varian, *Microeconomic Analysis,* Norton Books, 1976, Chapter 4 for a discussion of the econometric problems encountered in direct estimation of production relationships. See also G. S. Maddala, *Econometrics,* McGraw-Hill, 1977, Chapter 13. Suffice it here to say that the endogeneity of the right hand side variables is ignored and that estimation of production frontiers generally raises problems associated with nonnormally distributed errors.

31. See H. Varian, Chapter 1.

32. That is, the marginal product of labor is greater than the marginal product of capital by about 50 percent. If constant returns to scale were exhibited, it can be shown that under competitive market conditions, the coefficients on labor and capital could be interpreted as their respective factor shares.

33. The derivation of cost functions from assumptions of profit maximization and Cobb-Douglas production technology are presented at length in H. Varian, *ibid,* Chapter 1.

34. Because of the extremely small physical size of the constituent elements of an IC, small impurities or imperfections in the processed surface of the wafer can cause an IC to perform inadequately, thereby reducing the effective yield of ICs per wafer start. There may be international differences in the ability of producers to improve effective yields. In addition to eliminating or reducing the factors that caused the imperfections in the first place, it is also possible to build in circuit redundancy and other means of "salvaging" processed, but imperfect, chips. Even under the best of circumstances, however, the ratio of the actual to potential number of ICs per wafer may be as low as 50 percent for very large scale integrated devices such as the 256K DRAM (*Business Week,* August 18, 1986, p. 66).

35. See, "Chip Fight is Settled," Los Angeles Times, August 1, 1986, p. 11, and Semiconductor Industry Association, "US, Japanese Governments Reach Agreement on Market Access, Prevention of Dumping," *SIA Circuit,* Autumn, 1986, p. 1.