Historical Geography of the Semiconductor Industry

David M. Byrne, Adrian Hamins-Puertolas, and Molly M. Harnish *

February 28, 2022

Abstract

The production and use of solid-state electronics–devices built using semiconductors–has played a central role in economic developments in the post-war era. Most recently, disruption of the global electronics supply chain has slowed emergence of the global economy from the COVID-19 Pandemic. Prior to that, concerns about a slowdown in electronics innovation–the end of "Moore's Law"–colored the outlook for productivity growth. Analysis of these and other topics requires data on the electronics sector that is fundamentally global in scope. We take a step in that direction, building a dataset for an industry central to the sector: semiconductor fabrication. We collect disparate sources of data on semiconductor fabrication plants, known as "fabs," where electronic devices and circuits are etched onto raw silicon wafers, and build a database covering the industry from its inception consistently measured over time and across countries. We then use the dataset to shed light on two questions: whether the industry has entered a period of slower innovation and whether the roles of the United States, Taiwan, China, and other countries have changed in recent years.

*Federal Reserve Board of Governors, Washington, D.C.: david.m.byrne@frb.gov. Additional contributors to this research, described using the CRediT taxonomy, are Andrea Garcia (data curation), Steve Young (funding acquisition).

1 Introduction

The production and use of solid-state electronics–devices built using semiconductors–has played a central role in economic developments in the post-war era. Most recently, disruption of the global electronics supply chain has slowed emergence of the global economy from the COVID-19 Pandemic. Prior to that, concerns about a slowdown in electronics innovation–the end of "Moore's Law"–colored the outlook for productivity growth. Going back further, the pace of semiconductor innovation and competition among locations for the activities in the electronics supply chain have driven the emergence of Silicon Valley as a global innovation hub, the battle between U.S., European, and Japanese national champions in the early days of the industry, memory chip dumping accusations against Japan in the 1980s, the emergence of the "Asian Tigers" (South Korea, Taiwan, Singapore, and Hong Kong), the struggle of China to climb up the value chain, and the tech boom of the 1990s.

Analysis of these topics requires data on the electronics sector that is fundamentally global in scope. We take a step in that direction, building a dataset for an industry central to the sector: semiconductor fabrication. The electronics sector is composed of distinct industries devoted to fabrication, circuit design, equipment manufacturing, specialized materials, microchip test and assembly, and final assembly of electronics. These plants are managed by firms encompassing a shifting collection of plants, often from multiple industries within the electronics sector. We focus on industry-level analysis, collecting disparate sources of data on semiconductor fabrication plants, known as "fabs," where electronic devices and circuits are etched onto raw silicon wafers, and build a database covering the industry from its inception consistently measured over time and across countries.

Leveraging these data, we provide indicators for the pace of technical advance in the industry, the rate of growth in real output of the semiconductor fabrication stage, the location of the leading edge of production, and the location of basic and applied R&D.

2 The Electronics Sector and the Semiconductor Fabrication Industry

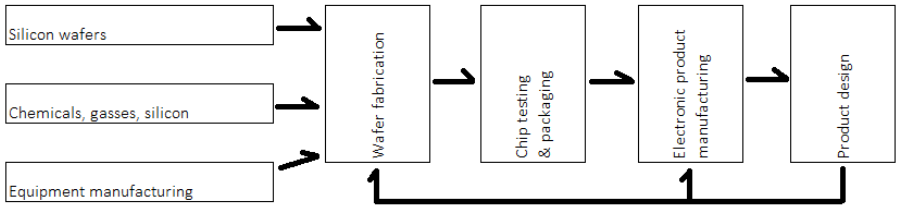
The manufacture of solid-state electronics, in contrast to the earlier technology based on signal transmission via gasses in vacuum tubes, consists of

creating electronic devices on a substrate of semiconducting material, most commonly silicon. These devices may be discrete components–individual resistors, capacitors, transistors, and so forth–or integrated circuits (ICs) which perform logical functions such as mathematical transformations (processors) or holding data to support processing (memory). Typically, multiple copies of the same device are created on a silicon wafer in a series of steps involving addition and removal of materials on the surface of the wafer. Technical advances focus in the main on achieving these devices on a smaller and smaller scale, leading to lower cost of production for a given device as well as greater performance for a given design, in the form of faster execution and lower power consumption..

There are, in effect, multiple industries at work in the construction of a semiconductor (fig. 1). The manufacturing of the specialized equipment used in the fabrication of electronic devices on silicon wafers plays a central role in the pace of technical advance. Most notably, advances in the equipment that performs the photolithography to deposit a film on the surface embodying the circuit design governs the level of miniaturization achievable. Wafer fabs house the ultra-clean environment where the equipment is used to conduct the dozens of steps needed to gradually build the device. Here, technical advance involves engineering the production line, careful calibration of the equipment, and refinement of the process to raise the yield–number of usable chips per wafer–for particular technologies and specific chip designs. The subsequent testing of each individual device on the completed wafer and the packaging of viable devices is a distinct industry in itself. The design process, where engineers use specialized software and engage with the end-user (a computer manufacturer, for example) and the chip manufacturer to create the functional identity of the final equipment in circuitry, is a separate process as well. Finally, there is the manufacture of specialized materials, notably including the manufacture of ultra-pure raw silicon wafers.

Early in the history of the industry, the entire supply chain was located under one roof (Bauer, 2020). Gradually, over time, the stages moved to different locations. Nowadays, materials, equipment, fabrication, test and assembly, and design most often take place at distinct establishments. The feasibility of physical separation and the different production functions in use at each stage argue in favor of treating each stage as a distinct industry as we do. That is, there are semiconductor industries, not a singular industry. Our focus in this paper is on the location and the output of the fabrication industry.

Figure 1: Electronics sector and its industries.



3 Data Construction

To address the drawbacks of data currently available from national statistical agencies, we construct a comprehensive, global, and consistent dataset of the fabrication industry. We primarily rely on data from two sources for the construction of our database. Gartner, a consultancy focused on high tech industries, published occasional comprehensive global lists of semiconductor fabs as part of their *Semiconductor Equipment and Materials Service* beginning in 1985. We have located, digitalized, and harmonized 22 of these lists for the period from 1996 to 2007. SEMI (Semiconductor Equipment and Materials Internationa), a trade association focused on semiconductor manufacturing equipment, publishes a comprehensive, quarterly-frequency database of fabs, the *Wafer Fab Forecast*. We use two vintages of this report that together provide history spanning the period from 2008 to 2022. Each data source provides plant-level information on location, company ownership, technology employed, wafer capacity, and product mix. Importantly, the reports do not provide information on realized production. Most of our analysis of industry trends focuses on capacity, though when we compare our indicators to real output indicators we endeavor to adjust for utilization.

From 1996 to 2007, Gartner published 22 global lists of semiconductor fabs as part of their *Semiconductor Equipment and Materials Service*. Those lists contain plant-level identifiers–the the name and location of the fab, the name of the company that owns the fab, and the year the plant began operations–as well as information on the production technology employed described in Section 3. Our second source, SEMI (Semiconductor Equipment and Materials International) provides comparable information. For this paper, we use the two most vintages of SEMIs *World Fab Forecast* report that together provide history spaning the period from 2008 to 2022.

Using the unique fab IDs provided in all SEMI reports and in some of the Gartner reports, we begin by linking the observations together to creating a time series for each plant from its year of first production the last year of production or through 2021. For three of the Gartner reports, fab IDs are not provided. In those cases, we match to fabs in reports containing IDs using company names, fab names, location (city, state, country) and year of production, and assign fab IDs. In cases where matching is unsuccessful, of which there were 24, we drop the plants and rely on interpolation from between adjacent years. Moreover, in the case of Gartner data which does not cover every quarter, we interpolate quarters with no data in a similar fashion. We exclude plants included in these databases not devoted to fabricating microchips. Among these exclusions are plants producing MEMS (microelectromechanical systems), epitaxial wafers (primarily used for solar cells), and "back end" plants (used for testing completed wafers). We also exclude plants recorded as planned or under development. After this process, we begin in 1996Q1 with 1,001 plant level observations and have 1,057 in 2007Q3. Data from SEMI contains 1,085 fabs in 2008Q1 and 1,217 in 2021Q4. SEMI has more exhaustive coverage of university and government R&D fabs, accounting for the difference in the number of fabs in 2007Q3 in Gartner and 2008Q1 in SEMI. We are in the process of linking individual plants between the Gartner and SEMI datasets. In the meantime, we construct aggregates for each period separately and link the databases at the aggregate level.

An advantage of both datasets is the provision of the data on when the plant was established. This information will provide the foundation for future efforts to push the dataset back to the dawn of the industry. For example, the earliest date for plant establishment in the data is 1953. The result is a global dataset containing plant-level information on the capacity and technology of the global fabrication industry at a quarterly frequency spanning 1996 to 2021.

Detailed information is available by plant for the scale of production and technology employed. In particular, the data provide plant capacity in the number of silicon wafers that can be processed in a month, diameter of wafer employed, and a measure of the maximum level of miniaturization achieved at the plant known as geometry. The geometry of a technology, sometimes referred to as linewidth, is a reference to the technical limitations of a technology generation. These generations, or nodes, were well characterized by the width of the transistor's gate–the on-off switch–through the 1990s. Since

that time, the meaning of the geometry measure has varied as the focus of the industry has shifted to other features, but it remains a barometer of the industry's progress and a focus of discussion. Geometry is the measure most commonly employed to characterize the state of technology in the industry, though the dimension of the largest wafer in use plays an important role as well.

We employ wafer start capacity and wafer diameter to construct 8-inch wafer equivalent capacity by plant. The capacity of the average plant in our dataset rises from 7,500 wafers per month in 1996 to 23,000 in 2021. Using geometry, we construct an indicator of the "transistor budget" capacity for each fab, dividing wafers by the square of the geometry to yield a measure which moves in proportion to the number of features that may theoretically be placed on a wafer. The average fab in 1996 was capable of producing about one trillion transistors per month by this measure.

We identify fabs as research by searching for the strings "R&D" and "research" in the fab description fields found in each database. R&D fabs are identified as government or academic using a list of keywords such as "university" in the fab description. As noted earlier, only a handful of the fabs in the Gartner database are identified as research, so our analysis of lab location is limited to the period from 2008 forward for the time being, though we anticipate augmenting the lab history in future iterations.

We then construct a set of aggregates for use in the analysis of the state of technology in the industry and the geography of the industry. Excluding R&D fabs, we calculate for each country the market share using 8-inch wafer capacity as well as two indicators of the technology employed in each national industry: the mean technology employed weighted by 8-inch wafer capacity and the most advanced technology employed at any fab located in the country. When calculating the most advanced technology we exclude fabs with reported wafer capacity of 2,000 wafers per month or less, a rough way to exclude small scale fabs and fabs during the ramp-up process. Small scale fabs focus on production of less common technologies and we wish to record that a country has reached production at a particular node only when they have completed the early stages of employing the technology. Our dataset shows increasing capacity and decreasing geometry over time, as one would expect: Average eight-inch equivalent wafer starts (capacity) increase from roughly 7,400 in 1996 to roughly 22,000 in 2021, while minimum geometry falls from 50nm in 1996 to 3 nm in 2021 (Table X).

4 The Pace of Technical Change

We consider first how the dataset can shed light on the question of the pace of technical change in the semiconductor industry. The indicators constructed reflect the scientific and engineering advances that enable miniaturization. Other forms of innovation, most notably advances in design, are not reflected, which we view as an opportunity rather than a drawback as it offers an opportunity to contrast our indicator with a measure of real output of the final product–a completed microchip ready for use in a computer, say–which does incorporate the value added of design.

4.1 Moore's Law

In 1965, Gordon Moore noted that as the complexity of integrated circuits increased–measured in the number of components, such as transistors–the cost of production fell (Moore, 2006). In particular, in place of components constructed as distinct products and connected into a system (on a circuit board, say), pulling components together onto a single piece of silicon lowered costs dramatically. This process, however, was limited by the rising probability that a component in the integrated circuit would fail as the number of components rose. At any time, given the state of technology, there was a minimum cost complexity, or number of components. Moore observed that "the complexity for minimum component costs has increased at a rate of roughly a factor of two per year" and speculated that technical change in semiconductor manufacturing would support a continuation of that trend–doubling of the number of transistors on an integrated circuit every year–for at least ten years. Revisiting the issue in 1975, he noted that engineers had exhausted the low hanging fruit represented by optimizing cirucit design to pack components together as tightly as possible and going forward, further increases in the pace would be more subdued, projecting a doubling of the number of components on integrated circuits every two years, a prediction commonly known as "Moore's Law").

The first integrated circuit to incorporate the entirety of a logic processor was Intel's 4004 microprocessor, released in 1971 and used in the electronic calculators that were the hot high-tech consumer product of the early 1970s. The size of the die–the area of the chip–was 12 square millimeters and the geometry employed was 10 microns, suggesting a value for our transistor budget indicator of 1.44 million. Importantly, this does not literally mean this

many electronic components (transistors, capacitors, and so on) could be fit on the chip and combined to create a functioning integrated circuit. (Creating integrated circuits is not an exercise in literally packing as many transistors onto a chip as physically possible.) In practice, the 4004 chip contained 2,250 processors. Starting from 1971, if Gordon Moore's speculation had come to pass 50 years into the future, the transistor budget indicator for an integrated circuit would be higher by a factor of 30 million in 2021.

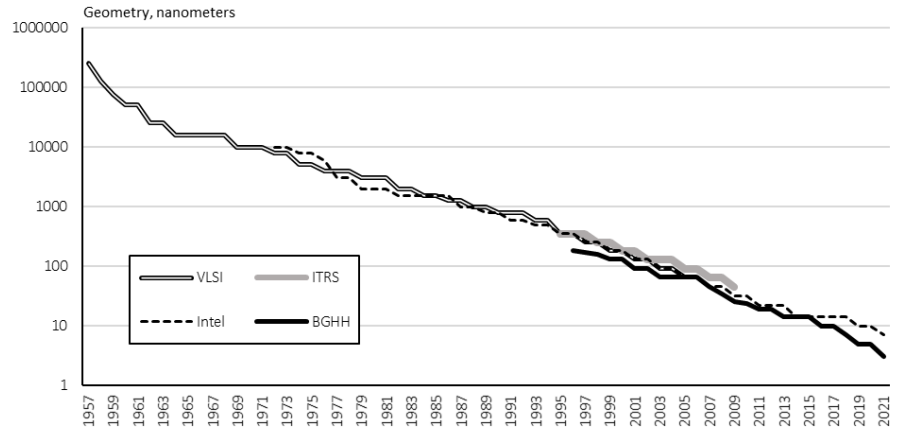
4.2 Geometry

The pace of advance in geometry is the most prominent focus of the fabrication and equipment portion of the industry, serving as a benchmark for the state of technology. The industry calibrates this technical pace in terms of nodes, which serve as focal points for basic and applied research as well as development of commercially viable products. Each node represents a roughly 70 percent shrink in the linear dimension of the features on a chip, implying that a given number of features occupies half the planar area on a wafer (squaring 70 percent yields roughly 50 percent). In practice, companies develop technologies in the neighborhood of these nodes, rather than the exact geometry, so there is a bit of judgement involved in concluding a particular node has been reached.

We consider first some available indicators of progress in the industry. Beginning in the 1980s, the industry published a roadmap identifying the technical hurdles required to reach the upcoming nodes, which served as a guide for basic and applied researchers contributing to the project, the ITRS (International Technology Roadmap for Semiconductors) (Semiconductor Industry Association, 2007). The node history from 1995 to 2009, cobbled together from ITRS documents–no single document provides the full history–is shown in fig. 2 on the following page. The pace of advance was irregular, but over time, the length of time between nodes in this period averaged a little over two years. The pace of technical advance for Intel is also shown, for which a longer history is available (**?**). Historically, nodes have sometimes been reached first by memory manufacturers, but among processor manufacturers, Intel is typically the first to achieve volume production at a particular node. Prior to the time period covered by the ITRS series, advances between nodes were less regular. Between 1971 and 1995, the average length of time between nodes was 4 years. Last, estimates from VLSI Research Inc., a con-

sultancy with deep history in the industry, is shown back to 1957.[1] These are broadly consistent, but not identical to the ITRS and Intel node history.

Figure 2: Semiconductor Technology Node History



Source: VLSI, Inc., International Technology Roadmap for Semiconductors, Intel, Inc., authors' calculations.
Note: Authors' calculations define a node as production using the geometry at a fab with capacity of at least 2,000 wafer starts per month.

Each measure is authoritative but has disadvantages from the point of view of economic research. The Intel measure is company-specific and since in recent history, Intel has been replaced by TSMC (Taiwan Semiconductor Manufacturing Corporation) as industry leader, not a good barometer of the leading edge. The VLSI measure represents the judgement of expert industry analysts, but is judgement nonetheless, and not algorithmic. The ITRS series does provide a clear definition, but one that is somewhat difficult to implement independently. In particular, they reference the status of DRAM production technology in the 2011 roadmap, stating: "Please note that, unless otherwise specified for a particular line item, the default year header still refers (as in previous Roadmaps) to the year when product shipment first exceeds thousands of units per month of ICs from a manufacturing site using 'production tooling.' Furthermore, a second company must begin production within three months. To satisfy this timing definition, ASIC production may represent the cumulative volume of many individual product line items processed through the facility."

[1]These estimates were kindly provided by G. Dan Hutcheson.

Finally, we show our results. From 1996 to 2021, we have reliable estimates at this juncture. The appeal of our estimates is we begin with industry data and apply a clear formula, such that the estimates are replicable without reference to the analysis of any company, trade group, or consultancy. We identify a node as reached when the geometry is employed in production at a facility with capacity of at least 2,000 wafer starts per month. Below this threshhold, one finds plants devoted to research, pilot lines, and boutique production. The average time between nodes for our estimates is [two and one-quarter years]. The time is somewhat variable, with a notable slowdown in the early 2000s.

What do we make of Moore's Law? Extending our estimates backward using the] timing provided by VLSI, Inc., we count 20 nodes, from 10 microns to 3 nanometers, yielding 2.5 years between nodes. The average rate of linear shrinkage of (ln(10000)-ln(3))/50 = -0.16 implies an annual areal shrinkage of -0.16*2 = -0.32, which is sufficient to double the density of components on a chip of a given size every -0.69/-0.32 = 2.15 years (fig. 3). Recall that Gordon Moore's observation referred to the number of features per integrated circuit, and integrated circuits have grown larger since 1971. While the Intel 4004 chip was 12 millimeters squared, a typical leading edge Intel processor in 2005, the Intel Montecito, occupies about 600 square millimeters, a pace of increase in the chip area of approximately 8 percent per year over that period, after which the die size has remained fairly stable. Allowing for the larger size of these chips, average time for the number of potential features per integrated circuit to double is 1.73 years.

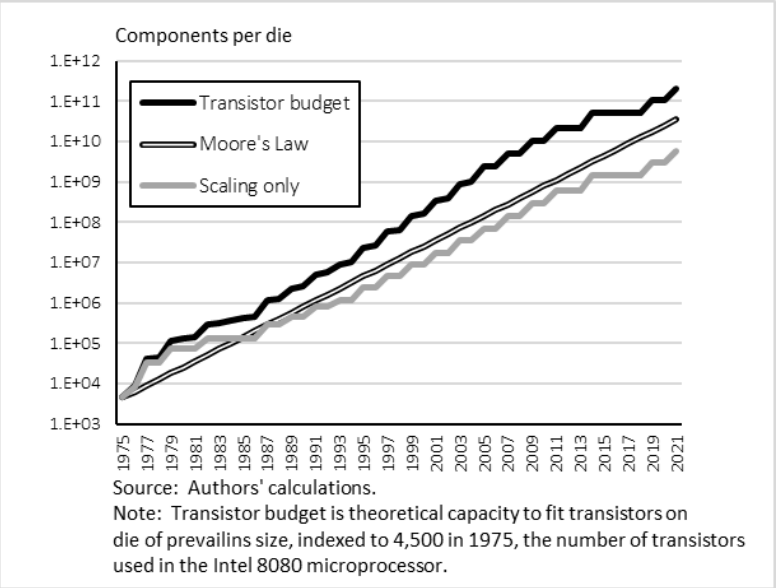
Figure 3: Moore's law calculations.

	1975-2021	1975-2003	2003-2021
Years to double components per die	1.73	1.53	2.03
Growth rate, components per die	0.40	0.45	0.34
Contribution from geometry	0.32	0.31	0.34
Contribution from die size	0.08	0.14	-
Memo:			
Years to double, geometry alone	2.15	2.20	2.03

Put differently, if one projects the number of components forward from the 1975, when the Intel 8080 was produced and Intel made his revised projection, using our measure of the transistor budget for ICs, and compares

that projection to Moore's law, the industry has exceeded his prediction, as shown in figure fig. 4.

Figure 4: Moore's law prediction vs. technical advance.



What of the uneven pace, however? It does appear that the pace of movement between nodes sped up in the mid-1990s, coinciding with the tech boom and intensified competition, most notably the battle between AMD and Intel for dominance of the processor market. Then, it slowed again in the early 2000s, remaining at 90nm for three years and 65nm for three years, leading many observers to remark on slowing innovation in the industry. At the time, processor producers intensified focus on design, rolling out multi-core processors, for example. However, the pace has sped up again recently, rolling out 8 nodes between 2006 and 2021, slightly faster than one every two years. On balance, it does appear that the pace of growth has slowed for the potential features per integrated circuit. While the number of years to double from geometry alone has actually decreased from 2.20 to 2.03, in the earlier period, the die size was increasing, and adding that effect, years to double has increased from 1.53 to 2.03 since 2003 (fig. 3).

4.3 Prices

Moore's Law and the pace of miniaturization are technical measures of the pace of technical advance. While driven ultimately by market forces, they are not a direct measure of what economists are interested in–the falling price of electronics, which ultimately measures welfare gains. While economists have made valiant efforts to measure these prices, the rapid changes in product quality confound these efforts to a great degree. For selected products, most notably memory chips (DRAM and Flash memory, especially) and microprocessors (MPUs), academic and government research has yielded plausible price indexes. But these products account for only 40 percent of global nominal shipments of semiconductors. For the remaining products, little progress has been made. Included in the group of products for which there has been no rigorous research on quality-adjusted prices are the application-specific processors employed in many of the most advanced products, such as smartphones, the digital signal at bridge the divide between analog real world signals (sound, for example) and the world of digital electronics, the sensors employed in motor vehicles and many other platforms, and many others.

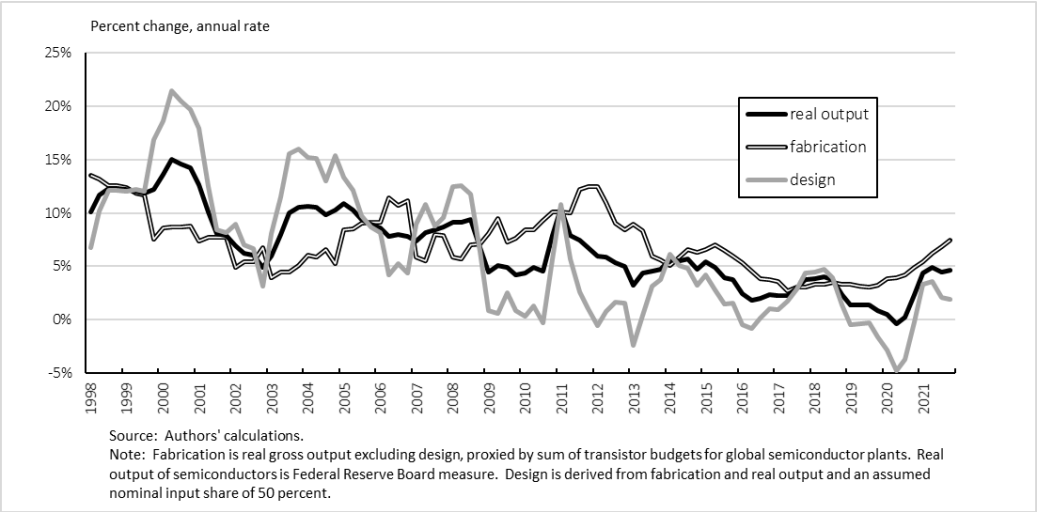
In any event, the aggregate price index for global total industry output employed by the Federal Reserve Board, indicates that prices for semiconductors, employing the quality adjustments available, have fallen 18 percent per year from 1992 to 2021. Measured prices fell at roughly twice this average pace from 1996 to 2001, during the tech boom when IT production and use accounted for the bulk of the surge in labor productivity in the United States. Since that time, prices have slowed, most notably declining only 4 percent from 2015 to 2021.

4.4 Prices and Geometry Together

In light of the sustained pace of advance in technology, the weak decline in quality-adjusted prices is puzzling. Employing the price index to deflate nominal output, **??** on page ?? depicts the rate of growth in the real gross output of the semiconductor industry.

Also shown is the rate of growth in an indicator for technical advance: The transistor budget for the aggregate global production of semiconductors. We arrive at production as follows. At the plant level, capacity in square inches and geometry in nanometers is available in our database. Converting and rescaling as appropriate, one arrives at an indicator for the number

Figure 5: Decomposition of real output of semiconductors.



of transistors the plant might theoretically produce in a given month. For example, a typical plant in 2021 might have a capacity of 10,000 eight inch wafers per month and employ a geometry of 25 nanometers. Assuming each transistor occupies a surface area of 25 nm squared, this implies one could create X million transistors on a typical wafer.[2] This is an indicator for the capacity for gross output of the fabrication stage of the industry, as discussed earlier. That is, the ability of the industry to provide a greater transistor budget for circuit design in a given surface area is a measure of technical change at the fabrication stage. To arrive at a production indicator, we scale the capacity indicator by the measure of capacity utilization for the U.S. industry provided by the Federal Reserve Board, assuming that the operating rate for the global industry is the same.

Showing the indicator for gross output of the fabrication stage in combination with the real output of the semiconductor industry as currently measured provides an interesting contrast. Taking these indicators at face

[2]We assume that the effect of factors that one would account for in a literal measure, such as the limitation on usable surface implied by creating square chips on a circular surface and the yield, are a fixed proportion over time. That is, we are constructing an indicator, not an exact measure. More importantly, the ability of the industry to accomodate more complicated designs over time, through a greater number of layers, is omitted. This makes the growth rate of the indicator for fabrication gross output biased downward, which, for the most part, does not undermine our broad conclusions.

value–that is, ignoring measurement issues for the moment–yields interesting implications for the story of the industry. The real gross output of the semiconductor industry is the sum of the contribution from the fabrication stage and the value added contributed by design, management, and test and assembly. Using the gross margin of major semiconductor firms as a barometer of the nominal weight of this residual category suggests that these remaining categories account for about 50 percent of the gross output of the industry. This implies an unseen pace of real value added from these remaining categories shown in the dashed line in the figure, which we refer to as "design, etc."

What story does this tell for the industry over the past 30 years? In the early 1990s, design advances were accounting for the lion's share of the gains in the industry. Then during the tech boom, fabrication and design were playing comparable roles. This comports with industry lore, at least for MPUs, in the sense that in the early 1990s, Intel was developing the x86 family of processors, then by the mid-1990s, Intel had released the Pentium and was churning out new generations with incremental improvements. Then in the early 2000s, the industry found that pushing the geometry envelop became increasingly hard and MPU design resorted to multi-core technology, which required far more design input than previous changes. By the late 2000s, despite the widespread predictions that we may have reached the limit of scaling, the pace of geometry began advancing rapidly again. Here, the story for design seems to go off the rails. Real output contribution to industry growth falls to negligible.

What should we conclude from this? Have advances in design in the semiconductor industry stalled? Returning to the issue of measurement, an alternative explanation is that design advances have shifted to chip categories that are relatively less well measured. Price indexes for the categories encompassing the design embedded in smartphones and other devices that rely relatively less on MPUs and memory and relatively more on custom processors, which account for roughly half of nominal output for the global industry. The rate of decline in the price index for these categories has slowed substantially in recent years, suggesting that product design, embedded in these chips to a great extent, has not improved as fast in recent years. This may indeed be true. Perhaps each incremental change in smartphones and televisions are not as valuable as the last. And, perhaps the advances embodied in the soaring electronic content of our vehicles is not as impressive as we might think. Or, indeed, these advances simply come with a price tag

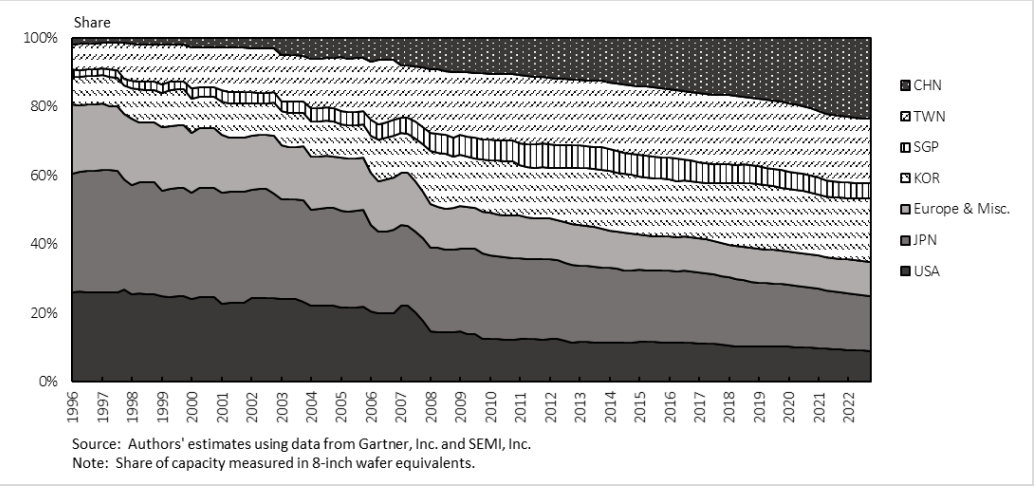
that implies a quality-adjusted price that only falls slowly. However, further inquiry into the performance gains of the chips embodied in these goods is in order. They have not been investigated with the rigorous approach used for memories and MPUs.

5 Industry Geography

5.1 Market Share

fig. 6 depicts the market share of key players in the industry in terms of the number of wafers processed.

Figure 6: Country market share, semiconductor wafer fabrication.

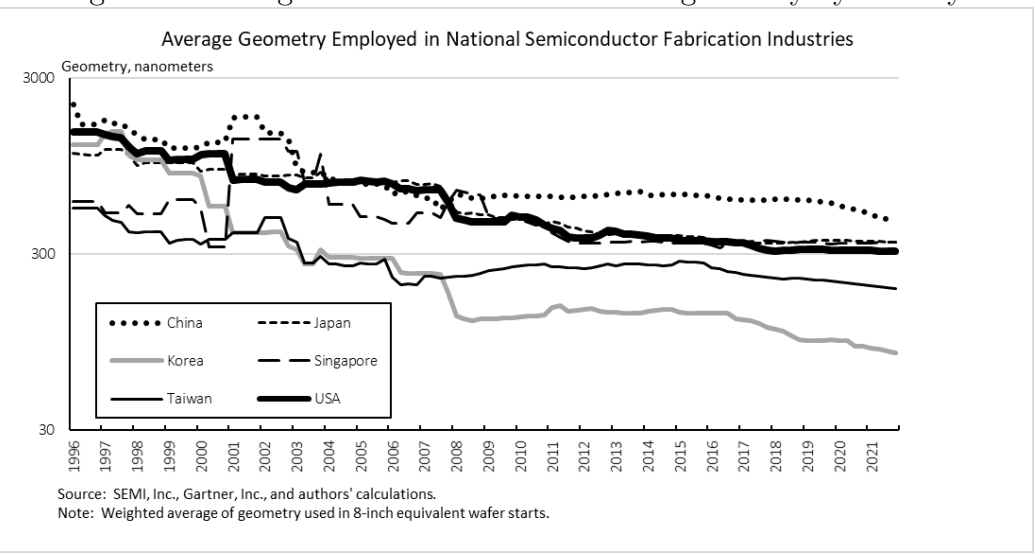


The role of different regions in the fabrication portion of the semiconductor industry has changed dramatically over time. In the 1960s and 1970s, Europe played a pivotal role in the early industry. In the 1980s, Japan emerged as a major player in the industry and Europe's role dwindled. The United States was also pushed out of the memory industry to a great extent, leading to accusations of dumping against Japan. However, the U.S. industry emerged as the key player in the processor sector and entering the 1990s, Japan and the United States accounted for the bulk of the industry with an important role remaining for Europe. Taiwan pushed into the industry in the 1990s, taking share from Japan and the United States. That trend

continued in the 2000s, while Korea and China emerged as well. The Taiwan share peaked at 20 percent, while the share of China has continued to rise, at the expense, primarily, of Japan and the United States.

Although China has the largest share of global capacity in terms of wafers shipped, the capacity is primarily at trailing nodes. fig. 7 depicts the average geometry by country and region and shows that the average China wafer has a geometry about 30 percent larger than the global average. The U.S. industry, by contrast, specializes more in cutting edge technology, with geometry averaging about 10 percent smaller than the global average. As countries introduce new fabs at the leading edge, the standing of other countries relative to the average can shift dramatically. Most recently, as new fabs have opened in Taiwan, the average geometry of that national industry has pushed more than 30 percent below the global average and Japan and China have fallen dramatically behind. The U.S. industry has slipped as well, though planned fabs will push them back to their conventional role. Trailing edge capacity plays an important role for national security concerns.

Figure 7: Average semiconductor fabrication geometry by country.

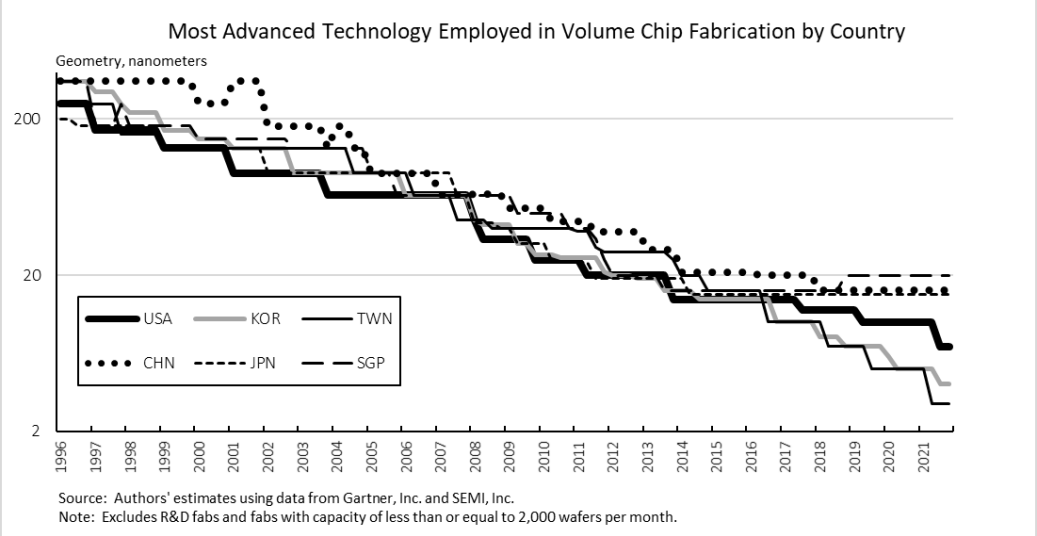


5.2 The Leading Edge

The presence of leading edge technology in a particular country is perceived to be a strategic advantage. Where have the technologies shown in

fig. 2 on page 8 first emerged? From at least 1996 until the late 2010s, the United States was this location (fig. 8). Since then, the Taiwan industry has emerged as the location for the leading edge.

Figure 8: Most advanced technology in volume chip fabrication by country.

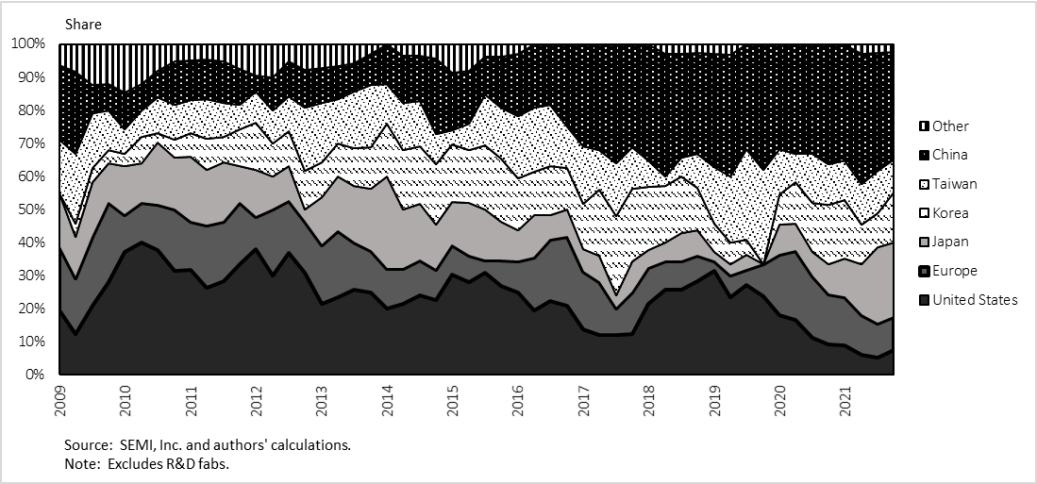


More broadly, the United States, Taiwan, and Korea have wrangled over the leading edge in recent years, led by their national champions–Intel, TSMC (Taiwan Semiconductor Manufacturing Corporation), and Samsung, respectively. Japan and Singapore had kept up with the leaders through the early 2000s but have since fallen behind. China, hampered by trade restrictions that preclude acquisition of leading edge manufacturing equipment, has been well behind the peloton throughout this period, but has fallen even further behind in recent years.

6 Location of Innovation

Volume production of semiconductors employing the most advanced technologies need not take place in the same location where those technologies are first developed. Wafer fabrication companies often designate facilities devoted to research into new fabrication technologies and development of the process that allows those technologies to be employed at scale. These

Figure 9: Location of new wafer fabrication plants.



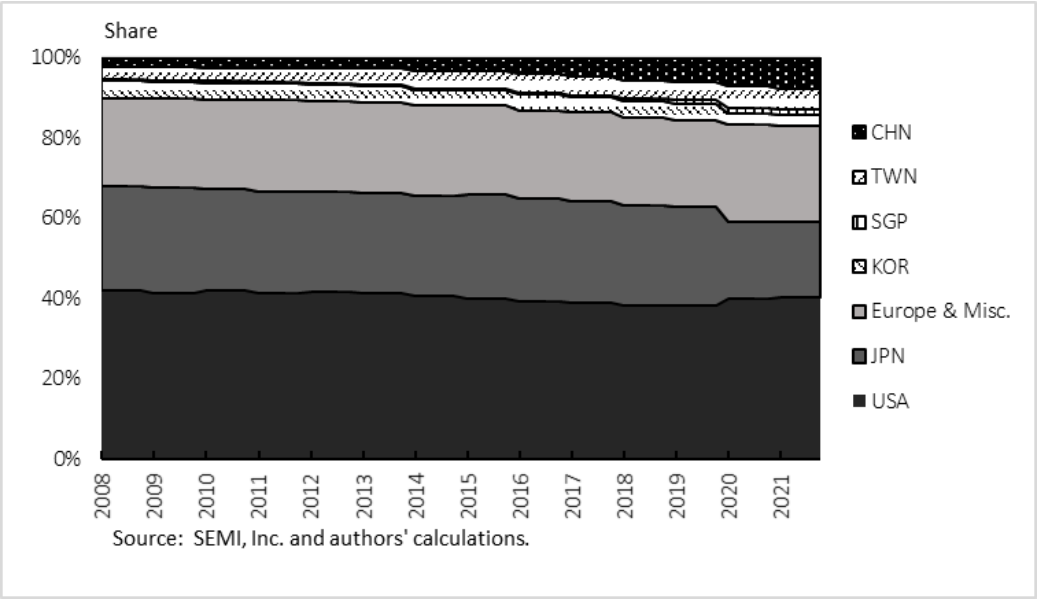
distribution of these corporate R&D fabs differs markedly from the distribution of production. In particular, the Asian locations accounting for roughly 60 percent of global wafer capacity in 2021–China, Taiwan, Singapore, and South Korea–accounted for less than 20 percent of the corporate R&D fabrication plants, though this share has increased in recent years. The United States, which accounted for only 12 percent of global wafer starts in 2021, accounted for more than 40 percent of corporate R&D labs, and Japan and Europe accounted for another 40 percent.

Academic and government fabs are an important part of R&D for the industry as well, with a particular focus on basic research. Here the role of Asia is disproportionately small as well: China, Taiwan, Singapore, South Korea, and Japan collectively are the location for roughly one in five of these fabs. Europe and the United States are by far the most important locations.

6.1 Where are the New Plants Emerging?

Addition of capacity to the industry and adoption of new technologies may take place through installation of equipment in clean room space at existing fabs. Eventually, though, companies face the decision of where to locate new fabs, a commitment of $10 billion dollars or more and consequently a revealing choice about the anticipated comparative advantage of countries for industry activity in years to come. Here the trend is decidedly

Figure 10: Corporate R&D fab locations.

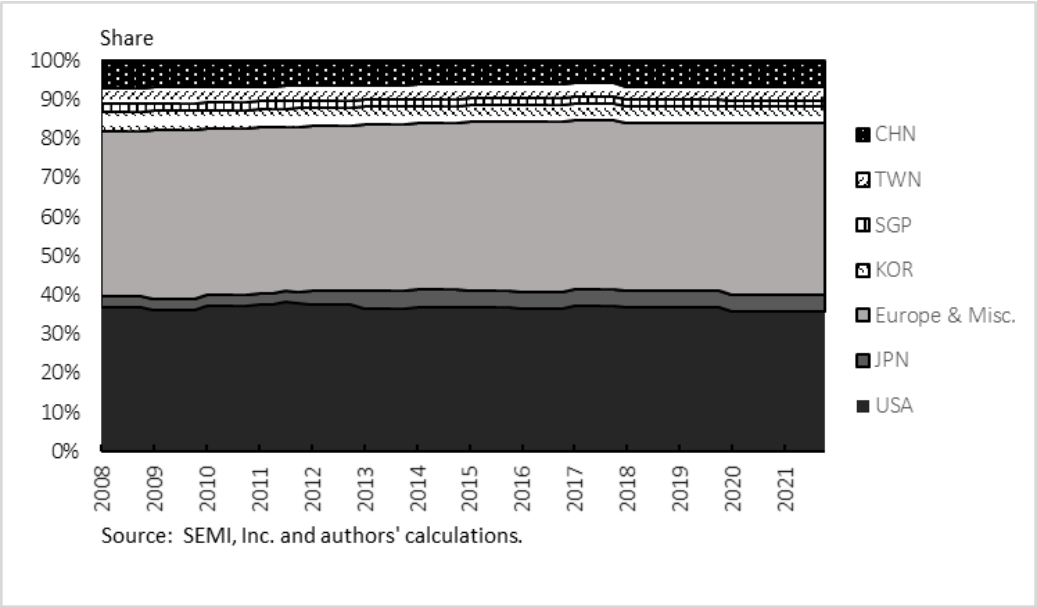


against location in Europe and the United States (fig. 9). These locations accounted for 50 percent of the new plants in 2011 and less than 20 percent in 2021. The role of China has more than doubled over that same time period, while the role of Japan, South Korea, and Taiwan has been little changed, on balance. Whether the recent flurry of new plant announcements will affect this distribution remains to be seen, as illustrated by the plans of the companies accounting for the lion's share of investment: Intel, TSMC, and Samsung. In a notable shift, all three companies have announced or broken ground on new fabs in the United States. In addition, Intel also has construction projects in Europe, TSMC has plans for new plants in Taiwan and Japan, and Samsung has expansion underway in South Korea.

7 Comparison to Official Statistics by Country

To be added

Figure 11: Academic & government fab locations.



8 Conclusion

The historical pace of technical advance in the semiconductor industry predicted by Gordon Moore has been sustained for fifty years by the metric he proposed: the number of components one can fit on an integrated circuit. Whether this measure remains a sufficient statistic for innovation in the industry is an open question–ICs come in a greater variety of forms and a produced using a wider array of technologies than was the case in 1971–but the ability to miniaturize devices continues to advance at a rapid pace. We offer a simple, well-defined measure to use to assess this question built from data on individual plant locations scrutinized by industry consultants and trade groups, which provides a useful counterpoint to measures that rely on judgemental appraisals by experts that are difficult to reproduce.

The industry's progress is to the credit of a wide array of contributors. Whereas corporate, government, and academic R&D continues to be located primarily in the United States, Europe, and Japan, as of 2021, the practical implementation of leading edge technology in volume production is dominated by Taiwan, with South Korea not far behind; the U.S. industry has fallen behind. Production in the United States as well as in Japan takes place

using a broad array of technologies, with the average technology in use incorporating appreciably less miniaturization than is the case for South Korea and Taiwan. And, in recent years, construction of new plants in the global industry has shifted toward China, though recent announcements point to a sustantial rise in the location of new advanced manufacturing locations in the United States.

The dataset used to establish these stylized facts about the industry, containing a long consistent history of plant-level data, will support future work on these production and R&D location decisions and the role played by such factors as the location of corporate headquarters and such policies as government subsidies and trade restrictions.

References

Moore, Gordon E, "Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp. 114 ff.," *IEEE solid-state circuits society newsletter*, 2006, *11* (3), 33–35.

Semiconductor Industry Association, *International Technology Roadmap for Semiconductors*, Austin, Texas: International SEMATECH, 2007.

WikiChip contributors, "Process Technology History - Intel," 2022. [Online; accessed 28-February-2022].

Wilder, E. Marshall Bauer Luc Olivier;, *The Microchip Revolution*, Independently published, 2020.