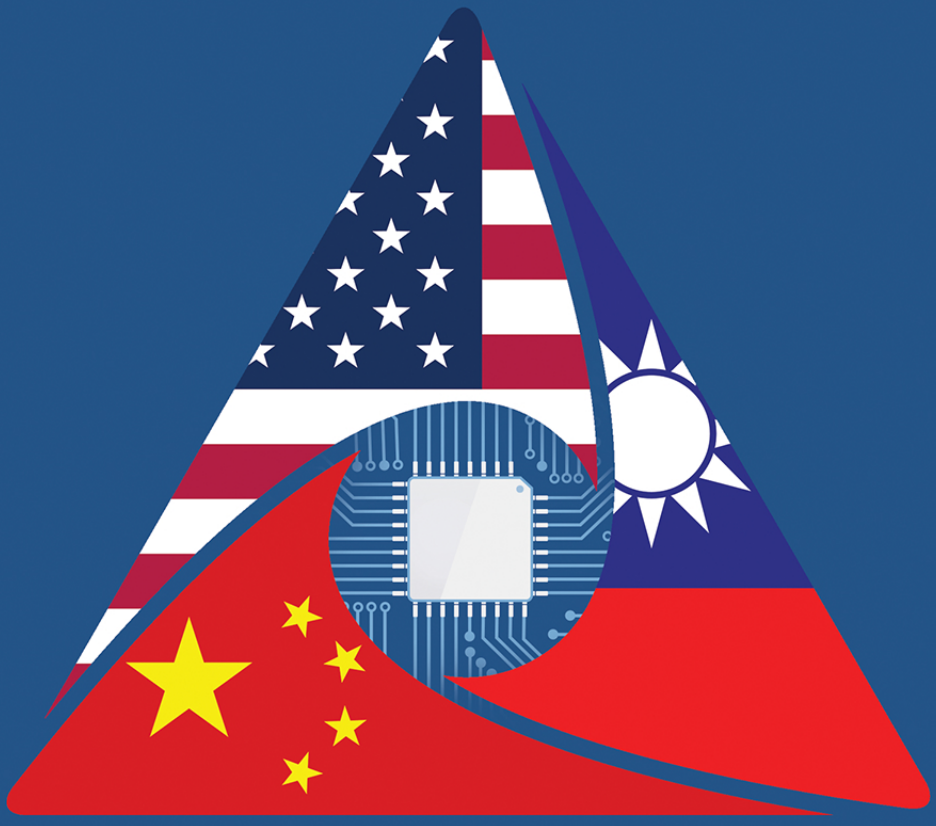

Silicon Triangle

The United States, Taiwan, China, and Global Semiconductor Security

Edited by Larry Diamond, James O. Ellis Jr., and Orville Schell

CHAPTER TWO

Implications of Technology Trends in the Semiconductor Industry

H.-S. PHILIP WONG AND JIM PLUMMER

Today's semiconductor industry is not static—it undergoes constant reinvention, and it is built on mutual interdependencies. This chapter offers background on and discusses the implications of semiconductor technology and industry trends.

• • •

Chip Types and Uses

Semiconductor technology covers a very broad range of technologies, such as logic, memory, power electronics, sensors, actuators, analog, and high-frequency/radio frequency (RF), as shown in table 2.1. Crucially, in discussing semiconductor technology, one must be cognizant that the entire semiconductor space is much broader than advanced-node logic chips, which have been the focus of attention recently. Table 2.1 provides an overview of the structure of the global semiconductor market.

A word on logic chip nomenclature: Chips are often referred to on a "nanometer scale," which has become a proxy for complexity and computing power. While that nanometer measurement can be thought of as referring to the length of the smallest component on the chip, these nanometer-branded "process nodes" are now umbrella terms that manufacturers use to represent successive generations of upgraded

Table 2.1. Semiconductor Market Segmentation

TYPE	% OF 2022 INDUSTRY REVENUE	FUNCTION	EXAMPLES
Logic	44%	Digital processors that act as the "brain" of modern computing	CPU (central processing units) GPU (graphics processing units)
Memory	23%	Short- and long-term storage of digital information	DRAM (dynamic random access memory) acts as the computer's "working memory." NAND Flash memory acts as long-term storage for computers and devices.
Discrete, analog, and other (DAO)	33%	Interact with the physical world by generating or transforming signals from electricity to radio waves or light, for example	Chips that enable such functions as charging a battery, electric vehicle motors, and phone calls (by accessing radio waves)

Source: 2022 data from the World Semiconductor Trade Statistics (WSTS) global industry forecast for calendar year 2022, released November 2022; "logic" category includes WSTS "logic" and "micro" categories. See WSTS, "The Worldwide Semiconductor Market Is Expected to Slow to 4.4 Percent Growth in 2022, Followed by a Decline of 4.1 Percent in 2023," press release, November 29, 2022.

production processes. This disconnect between branding and measurements is particularly true as chips have increasingly become complex three-dimensional structures.[1] While the metric is still in common use (mainly for marketing purposes), the nanometer node designations from different companies cannot be directly compared—a snag that can complicate industry assessments. For example, US-based Intel's 10nm and 7nm nodes are said to be roughly equivalent to Taiwan's Taiwan Semiconductor Manufacturing Company's (TSMC) or Korea's Samsung's 7nm and 5nm technology, respectively, due to a similar basket of transistor specification metrics adopted by each.[2] Apart from logic chips, memory technologies also commonly adopt a nanometer-based nomenclature, while storage technologies can be referred to by the number of three-dimensionally stacked layers they have.

Generally speaking, the performance of logic and memory chips improves at smaller nanometer measurements: chips with a greater density of transistors have greater computational power and (to a lesser extent) memory capacity. The performance of analog and discrete chips, however, is not directly correlated with the nanometer scale. Their "performance" instead refers to the overall beneficial attributes of a chip technology, including speed, power and energy efficiency, and density (not just speed).

For advanced logic, the current state of the art in production is 3nm technology—TSMC introduced 3nm commercial-scaled production in early 2023. The most-advanced-logic chips are used in CPUs and GPUs as well as field-programmable gate arrays (FPGAs) and application-specific processors, such as those in cell phones.[3] Currently, the world's largest and most profitable chip manufacturer, TSMC, has focused its investments on fabrication of the most-advanced nodes, with its 7nm and 5nm production lines accounting for more than half its sales in 2022.[4]

Currently, only TSMC, Samsung, and Intel have logic fabs (chip fabrication facilities) capable of manufacturing chips below 10nm at commercial scale, while major foundries United Microelectronics Corporation (UMC; Taiwan) and GlobalFoundries (US/UAE) have chosen not to invest in competing at the leading edge. China's Semiconductor Manufacturing International Corporation (SMIC), as of 2022, has generally not commercially produced chips below 10nm, with the exception of a cryptocurrency-mining chip that is claimed by third parties to exhibit some features consistent with manufacturing below 10nm.[5]

While leading-edge logic chips are profitable and central to advancing the technology frontier, in 2019 less than 5 percent of global manufacturing capacity was actually for nodes below 10nm.[6] To build a complete electronic system, one needs more than the logic chips—at a minimum, memory and storage are required, and depending on applications a system might also require analog devices, sensors, high-frequency/radio frequency (RF) devices, and power devices. Leading-edge logic is important for peak-speed performance as well as

overall industry revenue—but securing these chips is not sufficient for the totality of electronics end uses.

While emerging technologies and high-end consumer applications—e.g., supercomputers, gaming computers, cloud computing infrastructure, neural network accelerators for AI applications, and smartphones—require leading-edge chips, many parts of the economy operate on trailing-edge mature chips. Mature nodes are often defined as manufacturing processes at 28 or 40nm and above, used in the production of many automotive semiconductors, image signal processors for digital cameras, and other chips such as LCD (liquid crystal display) and LED (light-emitting diode) drivers and power-management controllers. A single car will use hundreds or even thousands of chips. A 40 or 65nm logic chip, for example, may be embedded within a larger assembly of sensors (i.e., discrete/analog/optoelectronic [DAO] devices) to allow the vehicle to function.[7]

It's a common misconception that such chips are simply older versions of the advanced nodes and the only difference is their lower cost. This misconception arises because the label "trailing-edge mature nodes" actually consists of two categories of chips: (a) digital logic chips of legacy nodes, and (b) specialty technologies.

Because that second category of specialty technologies is often derived from a digital logic platform, it is easy to conflate them with digital logic of mature nodes. These specialty technologies include, for example, sensors and actuators; power electronics; embedded memory; analog/mixed-signal and RF devices; power management integrated circuits (PMIC); and high-temperature/high-reliability and radiation-hardened technologies used in aerospace applications. While such specialty technologies use fabrication processes that derive from trailing-edge "mature" logic nodes, significant efforts must still be expended to develop and qualify them for these tailored applications. These technologies are a special category per se and should not simply be interpreted as cheaper products. For example, the claim that the US military's use of microelectronics often requires trailing-edge "mature" nodes typically refers to use of these specialty chip technologies and not necessarily (old) digital logic chips of mature or legacy nodes. The

world's third-largest contract chip manufacturer, UMC, now focuses its investment on fabrication of these chip technologies at 28 or 40nm process nodes for a variety of these specialized applications.

Cost also matters. Today, as in the past, consumers of semiconductors must weigh the functional gains of more-advanced chips against the much greater costs. A product such as an iPhone 13, which is designed around a 5nm chip, simply could not exist without using that advanced technology. Others point to the popularity of chips produced using 28nm technologies as the "sweet spot" between cost and function. It is perhaps most accurate to say, however, that for each product segment, there is a suitable technology node given the cost and performance trade-off. One has to meet both the performance and the cost targets.

After logic chips, memory and storage chips are the second-largest category of semiconductors, representing 32 percent of global manufacturing capacity and 26 percent of revenue in 2019.[8] While memory chips often do not receive the same attention as logic chips, they are similarly ubiquitous in enabling the function of electronic devices. As such, they merit similar supply chain resilience attention. The dominant memory technology today is dynamic random access memory (DRAM). Samsung (Korea) is the dominant supplier of DRAM (44 percent market share) followed by Micron (US, though with most production overseas) and SK hynix (Korea), with about 22 to 27 percent each. Meanwhile, NAND Flash is the dominant storage technology. Samsung also dominates NAND storage (35 percent market share), followed by Kioxia (Japan, formerly Toshiba) and its joint venture partner Western Digital (US, though NAND is produced in Japan with Kioxia), SK hynix, and Micron sharing market shares in the teens. While significant innovation occurs in both fields—one may hear, for example, of NAND storage being progressively stacked in three-dimensional configurations of 176 or 232 "layers"—memory and storage chips are generally considered to be more commoditized than logic chips. They are often combined interchangeably from different vendors within a finished electronic system, given that these chips are more likely to be produced to industry-wide common specifications representing different device architectures and manufacturing methods.

This standards-based interchangeability is in part why China's emerging DRAM and NAND manufacturers have been able to make better progress than China's logic chip manufacturers.[9]

Defense Needs

Military chip needs are of particular interest. In addition to having corporate information technology (IT)–type chip and consumer electronic demands much like any other complex global organization, the US Department of Defense (DoD) is concerned with the procurement and maintenance of specialty transport, communications, and weapons platforms that have unique semiconductor capabilities and security requirements. Javelin missiles—sent by the thousands to Ukraine, for example—rely on over 250 chips to manufacture before reaching the shoulder of the warfighter.[10] A deployed soldier himself may carry upward of six Global Positioning System (GPS) chips for his radios, range finders, and other equipment, with each GPS chip relying on other semiconductors for specific capabilities.[11] Although many of the components in such systems are similar to semiconductors used in consumer electronics—a weapons system, much like an automobile, may rely on hundreds of distinct logic chips, memory chips, communication chips, and sensors—the US defense apparatus also requires chips with higher levels of reliability and performance for unpredictable environments of conflict.

DoD chip-security concerns are therefore broad. First, like other chip consumers, DoD is concerned with supply chain resilience. In other words, because it relies on foreign suppliers, its supply may be cut off through global disruptions, or the threat of intentional disruption could be used as strategic leverage against US interests.[12] Second is a more unusual information-security concern: the risk that its chip designs or specifications may be leaked to adversaries through the production process, or the risk that hidden vulnerabilities could be inserted into a chip through a foreign supply process. Finally, DoD has been concerned that semiconductor-related capabilities and know-how—which underpin much of the so-called third offset of the US military's

comparative strength and which were largely invented in the United States—could fall into foreign hands. This last concern has largely already come to pass with the spread of leading-edge chip research and development (R&D) around the world and the migration of large portions of the semiconductor supply chain overseas; hence DoD is no longer always the first to capitalize on chip advances.

Because of these special performance requirements and security concerns, many military-grade microchips are subject to a higher level of production oversight, testing, and quality control than those used in consumer electronics. Even so, while the defense industry is reliant on chips, the chip industry as a whole is no longer reliant on the defense customers. In its infancy, the semiconductor industry got its start and was nurtured by US defense needs.[13] Today, US DoD and contractor chip needs are about two billion chips per year, estimated to be less than 2 percent of the market.[14] Reconciling these special needs with a relatively small purchasing power has led to a unique portfolio of supply streams. Those defense uses, in rough order of increasing specialization, include these:

- *Purchases of commercial off-the-shelf semiconductors*, including analog, memory chips, or GPUs, produced in the United States, Korea, or Taiwan. Such chips are subject to the same global supply chain resilience concerns as consumer products.
- *Field-programmable gate arrays (FPGAs)*, which are application-agnostic upon manufacture but can then be programmed or reprogrammed by the chip integrator to perform the functions needed for that application. FPGAs have large commercial market applications in data centers and communication switching networks. But the use of modular FPGAs is also attractive in the defense industry because of the small volume of chips often needed for specific use cases: using modular FPGAs, the customer is able to purchase relatively advanced–logic chips from a commercial fabricator without needing a high-volume, custom-designed production run; further, there are fewer security concerns with the supply chain, as the chip designer and fabricator do not need

to have full visibility on final circuit configurations (which could otherwise reveal characteristics of the weapons platform in which the chip is being used). The flexibility of FPGA chips does come at some cost: while easier to produce at small volumes, these chips are less dense (in terms of logic gates per square centimeter) and generally slower than the more optimized, application-specific chips described below. The US firm Xilinx—a pioneer in FPGAs, acquired by US chip design firm AMD in 2022—designs such chips and sells them to DoD users; Xilinx chips are fabricated at least in part by leading Taiwanese fabs UMC and TSMC.[15] Intel is another provider of FPGAs, through its 2015 acquisition of Altera (US); Intel/Altera has historically used both TSMC and Intel itself for fabrication of its FPGAs.

- *Application-specific integrated circuits (ASICs)* whose functions are optimized from the beginning for particular platform needs. Because ASICs are designed and produced to match specific end uses, there are more security concerns surrounding their production—those involved in their creation could gain information about the strengths and weaknesses of the weapons they enable. In part for this reason, twenty years ago DoD established its "Trusted Foundry" program to provide for domestic design, manufacturing, and assembly of very small volumes of classified or radiation-hardened chips meeting high security standards—and for a premium price.[16] The Trusted Foundry program certifies (and provides availability payments to) a constellation of suppliers—all of them currently US based. These suppliers range across the chip supply chain, from designers to IP block vendors to mask producers, fabricators, and testing; Trusted Foundry certifies each participant to be able to handle what DoD terms "Critical Program Information." The important downside of this approach to security is that the extra cost and overhead needed to maintain and certify such protections—which can affect things like staffing of facilities[17]—combined with the small volumes of chips that are needed from it (thought to be as little as 2 to 10 percent of DoD's own needs[18]) means that the most-advanced commercial

chip firms choose not to participate in it. In turn, DoD is left with slower innovation cycles and older technologies underlying its most secure chips, including for chips that are meant to underpin next-generation weapons systems. Recent initiatives—including DoD's multibillion-dollar Rapid Assured Microelectronics Prototypes using Advanced Commercial Capabilities (RAMP) and State-of-the-Art Heterogeneous Integrated Packaging (SHIP) programs—are intended to more flexibly access commercial semiconductor capabilities, given DoD security needs.[19] While real technical hurdles still stand in the way of a complete transition from a "trusted" to a "zero trust" (quantifiable assurance) model for DoD chip buyers, that desired end goal is the correct one, and accelerated efforts toward quantifiable assurance would contribute to US national security.[20]

Two other specific classes of defense industry chips that are small in market volume but have important area applications are these:

- *Compound and wide-bandgap semiconductors*, which are ideal for high-power and high-frequency applications such as radios and microwaves used in defense and aerospace. Compound semiconductor chips are produced using gallium arsenide (GaAs), silicon carbide (SiC), and gallium nitride (GaN), in addition to the conventional silicon substrate of typical commercial semiconductors. Skyworks is one US-based producer and fabricator of such chips; Skyworks' chips are also manufactured on a contract foundry basis by WIN Semiconductors (Taiwan).[21]
- *Radiation-hardened (rad-hard) semiconductors*, meanwhile, are needed to perform reliably in high-radiation environments—including in outer space and in nuclear accident environments—and in strategic nuclear weapons systems.[22] Chips operating in such environments are subject to Single Event Effects (SEEs) stemming from the interaction of atmospheric neutrons produced from cosmic rays or alpha particles from radioactive decay of thorium and uranium.[23] While less likely, these sorts

of interactions also pose potential concern in high-reliability ground systems (e.g., autonomous vehicles, electric vehicles, unmanned aerial systems, or smart grid). Without hardening or other resilience to SEEs, affected chips can malfunction or return unexpected outputs.

Major rad-hard chip producers include Microchip Technology (US), BAE Systems (UK), Honeywell (US), Renesas (Japan), Crane Aerospace & Electronics (US), and Infineon (Germany). Such chips can be produced by device and technology design and with the use of physical hardening materials. Their production is small volume and expensive—and thus the use of semiconductors in these environments is often many generations behind the commercial state of the art. Alternately, they can be certified for rad-hard resilience through "serendipity"—that is, when a commercial off-the-shelf component, when tested in relevant environments, happens to have good radiation performance without the traditional physical hardening processes. For example, the 7nm Xilinx Versal FPGA-type chip, fabricated by TSMC, was not designed to be rad-hard but performs well in space and other high-radiation environments.[24] Rad-hard by serendipity—or the use of self-checking and redundant processing architectures—is of growing importance to the market because of the growing number of space systems being constructed and sent to orbit.[25] Even so, the overall radiation-hardened chip market remains small, expected to be worth only $1.8 billion by 2027.[26]

In short, the US defense industry wants to feel comfortable with the security of its chip supplies, but it also wants access to the latest chip technologies. Getting that balance right has been difficult. Many observers argue that the defense industry has gone too far in the direction of security. Having access to advanced domestically produced chips (subsidized by measures such as the CHIPS Act) would in a way be an "easy answer" to DoD's quandary, as opposed to more fully pursuing a more flexible "quantifiable assurance" model of chip procurement.

The Commercial Semiconductor Value Chain

The semiconductor industry demands high levels of both R&D and capital expenditure (capex). These demands have created commercial incentives for a globally distributed and highly specialized global supply chain. The value chain can be summarized as having four production steps, each with various inputs:

Production Steps

1. *Chip Design*: Semiconductor design firms use technology-proven units of intellectual property called "IP cores"—which are previously designed circuit blocks known to function correctly—and electronic design automation (EDA) software to design chips for specific end uses (e.g., AI accelerators or chips for smartphone memory). This stage involves close collaboration between the design firm and the end customer (such as a systems integrator or original equipment manufacturer [OEM]), and chip design firms compete to develop the highest-performance or most efficient chips, or desirable application-specialized chips. Large systems companies such as Apple, Alphabet, and Amazon have also started designing their own chips.

    - *Design stage software inputs*: EDA software is a collection of powerful computer-aided design tools to map out the components on an individual chip, simulate and verify designs, optimize chip layouts for performance, assess manufacturing margins, and create physical masks for the manufacturing process. Today's EDA tools allow chip designers to start from high-level descriptions of desired system behavior instead of designing every transistor circuit explicitly, thus allowing the design of chips with hundreds of billions of transistors.
    - *Design stage IP inputs*: The fundamental IP building blocks are used as starting points in the semiconductor design process. Key examples are the Advanced RISC Machine (ARM) architecture for mobile devices and the x86 processor architecture for CPUs. These specialized firms continuously invest in and upgrade their IP blocks to remain competitive.

2. *Production Technology Development*: Just as chips themselves need to be designed, so do the manufacturing processes themselves. If a particular chip design is like a recipe for a dish, and the fabrication step (below) is the cooking of that dish, then this middle step can be thought of as the conceptualization of the restaurant, the scope of its menu, and the design of its kitchen. Often ignored in policy discussions, the manufacturing technology development step is difficult and expensive, often learned over time and sustained though tacit knowledge—and therefore poses high barriers to entry.

    - *Customer service and business coordination*—that is, working with end-use system integrators (whether internal to the firm or external, as in a contract foundry model) to identify the technology specification and cost trade-offs for a commercially viable chip technology, given the application needs. This close, trust-based process also involves working with semiconductor equipment manufacturers to road-map new tool capabilities for fabricating the desired chip technology within an overall production process.
    - *In-house design of fabrication processes*: These production protocols can run into hundreds or thousands of steps.
    - *Simulation and experimental prototyping*: Combined device and process technologies are tested at small scale to achieve technology targets, followed by ramping up of those prototype technologies deemed feasible into high-volume production with high yield. Such prototyping capability is expensive to set up. It is often done in the same physical location as the fabrication facility, and uses some of the same skilled workers, to ensure a smooth handoff between R&D and manufacturing.

3. *Fabrication*: Chip designs are then manufactured at specialized facilities called "fabs" or "foundries." The fabs use specialized equipment to print the geometric circuit patterns onto silicon wafers, which are then treated with chemicals to etch or deposit the pattern onto the wafer.[27] The customer of a stand-alone fab

company will often be a chip design firm, which will then sell the finished chip on to the system integrator/OEM.

- *Fabrication stage equipment inputs*: Semiconductor manufacturing equipment (SME) is a category of tools required for manufacturing (such as lithography and etching tools, including stencil-like masks that are specific to a chip design) and for metrology (tools that allow high-precision monitoring and measurement of the manufacturing process).
- *Fabrication stage chemicals and material inputs*—that is, specialty chemicals, gases, and materials that are used in the manufacturing process
- *Fabrication stage wafer inputs*—that is, the silicon wafers onto which individual chips are etched and deposited

4. *Test and Assembly*: After fabrication, the printed wafer is tested to ensure function, cut into individual integrated circuits (die), and packaged alongside complementary chips into specific product applications, itself an increasingly complex process.

As the industry has developed, six regional hubs have emerged in the semiconductor value chain: the United States, South Korea, Japan, China, Taiwan, and Europe. Broadly speaking, the United States currently specializes in many of the less capital-intensive (and more profitable) parts of the value chain, such as EDA software, intellectual property (IP), chip design, and manufacturing equipment. The US advantage in these areas derives from a leading global talent pool, a hub of world-leading universities, and high levels of government investment in basic research. The countries of East Asia, meanwhile, lead in capex-intensive activities, such as production technology development and fabrication as well as packaging, assembly, and testing.[28] These countries tend to have strong government incentives to establish facilities, as well as a larger, cheaper pool of both low-skilled labor and high-skilled talent. In earlier decades, the United States also led in these activities, but over time it has outsourced them, largely to Asian economies.

The often used, somewhat simplified motif of design versus manufacturing belies the fact, however, that developing new generations of semiconductor technology (the technology development step outlined above) or developing increasingly related advanced packaging technology also requires very large, colocated R&D efforts. While the United States excels in basic science research, East Asian countries often do very well in translating such research into practical technologies, and their governments often have incentives and infrastructures that facilitate such technology translations. So, while upper echelons of these firms' engineering development teams are in fact often staffed at least in part by experts trained in the United States, they oversee the work of hundreds or thousands of local, highly skilled R&D staff who enable that continuous process of translation from basic science to applied commercial technology.

Industry Structure

The global chip industry structure now exhibits both specialization for efficiency among the different segments of the value chain described above and consolidation of players within each segment.

Today, we are down to three major players in leading-edge logic (Intel, Samsung, and TSMC), a second tier of perhaps three major players in mature logic (Taiwan's UMC, US/UAE's GlobalFoundries, and China's SMIC), and three to four major players in memory (Korea's Samsung and SK hynix, US-based Micron, and the Japanese-US Kioxia/Western Digital). The semiconductor equipment companies have also consolidated into five major players (US-based Applied Materials, Lam Research, and KLA; Japan's Tokyo Electron; and ASML in the Netherlands). The EDA software companies have also consolidated into three players (US-based Cadence and Synopsys, and German-US Siemens/Mentor Graphics). FPGAs—which as described above are used in data centers and many military applications—once designed by Altera and Xilinx and manufactured at foundries domestically or abroad, are now part of Intel and AMD (which uses TSMC as its foundry), respectively.

Because of the high cost of capital investment and the long time horizon for maturing the technology, startups in both semiconductor manufacturing equipment and chip production technologies or chip

manufacturing itself (e.g., foundries) are almost nonexistent in the United States. And while there have been a number of US memory device startups, none has been successful. Rather, whatever companies are left are quickly consolidating. The few successful startups in the semiconductor manufacturing equipment space (e.g., US-based Cymer and Inpria) are all part of larger companies now (ASML and Japan's JSR, respectively). In software EDA, US startups with a key innovation (often in an algorithm) or a niche application often get acquired by one of the big three incumbent chip software companies (Cadence, Synopsys, or Siemens/Mentor Graphics). These startups in EDA software typically no longer organically grow into larger companies in the United States—it is often difficult to remain independent for long because their products need to be plugged into the larger, more comprehensive design infrastructure dominated by the large firms.

One exception to this trend is in so-called fabless chip design companies: instead of manufacturing its own chips, fabless firms produce their designs for sale to customers using a third-party foundry's production lines (e.g., TSMC or UMC) on a contract basis. The capital needs of fabless firms are lower, and there are many startups. In a way, foundries play the role of venture capitalists: foundries "invest" in the startups by offering wafer capacity, with the goal that those wafers that are used to prove out a product will eventually turn into larger wafer orders down the line. Fabless chip design firm US-based Nvidia's use of TSMC's manufacturing capacity for its groundbreaking GPU chips was a prime example of this symbiotic relationship.

Even here, however, there are emerging warning signs regarding the health of this startup ecosystem—namely, a primary bottleneck for fabless startups has become lack of access to foundry capacity to prove out their ideas in fabricated chips. In a tight supply-demand environment, leading logic fabs prefer to instead allocate access to prime wafer capacities to established customers (such as Qualcomm or Apple) with large wafer volumes that lead to surefire profits. Increasingly, whatever access smaller chip design startups then have is often a few technology generations (nodes) behind. This dynamic limits the pace of innovation in a segment of the chip supply chain where the United States has traditionally dominated.

Regional Value Chain Concentration

The small handful of countries and regions holding major concentrations in the chip supply chain (see table 2.2) has driven concerns about the resilience of supply to external shocks and geopolitical tensions. Chapter 6 in this report delves deeper into some of the regional specialties, and their future ambitions to extend or to diversify from current strengths.

Table 2.2. Countries with Leading Positions in Different Segments of the Semiconductor Supply Chain

CHOKE POINT	COUNTRY	COMPANIES	DESCRIPTION
Semiconductor design	US	Qualcomm, Nvidia, Broadcom (and systems companies such as Apple)[a]	The US is home to 10 out of 20 top global semiconductor design companies, which account for 50% of global revenue.[b] US firms account for >90% of market share for the design of advanced-logic products.
EDA software	US	Cadence, Siemens/Mentor, Synopsys	The US is home to the three largest EDA firms, which account for 85% of global market. Near-term alternatives to these three firms are likely infeasible.[c] Mentor is now owned by Siemens (Germany), but its HQ remains in the US.
Manufacturing equipment (SME)	US, Japan, Netherlands	Applied Materials, Lam Research, KLA-Tencor, & others (US) Tokyo Electron (Japan) ASML (Netherlands)	US firms collectively account for >50% of global market share in 5 of the major manufacturing process equipment categories.[d] ASML has 100% global share in EUV lithography equipment, which conveys a major advantage in leading-edge manufacturing (at 5nm and below).
Technology development and fabrication of leading-edge logic chips	Taiwan	TSMC	TSMC has a lead of 2 to 3 years in leading-edge logic manufacturing technology over all other industry competitors.

Table 2.2. (*continued*)

CHOKE POINT	COUNTRY	COMPANIES	DESCRIPTION
Technology development and fabrication of memory (DRAM) and flash storage (NAND) chips	South Korea	Samsung, SK hynix	South Korean integrated device manufacturers are dominant in the design, fabrication, and assembly of memory chips. They have 75% of the global DRAM market and 45% of the global NAND market.[e] But China-based memory manufacturers have rapidly been gaining capabilities and market share.
Wide-bandgap and compound semiconductors	US, Europe, Japan	Wolfspeed/Cree, ON Semiconductor (US) Infineon, STMicroelectronics (EU) ROHM, Mitsubishi Electric (Japan)	A variety of products and applications across power electronics, RF, and LED lighting. There is no clear market leader among the major players in the US, Germany, Netherlands, and Japan. China has identified power electronics as a focus area to reduce reliance on Western producers.
Photoresist processing equipment	Japan	JSR, TOK, Sumitomo Chemical, Shin-Etsu	Japanese companies have ~90% share in the global photoresist processing market.[f]
IP cores	UK	ARM Holdings	ARM architecture and processor cores are dominant in the mobile and tablet market. A $40 billion acquisition of ARM by Nvidia was abandoned under regulatory pressure in early 2022.[g]

[a]Systems companies such as Google, Facebook, Amazon, and Microsoft have started designing their own chips.

[b]Antonio Varas, Raj Varadarajan, Jimmy Goodrich, and Falan Yinug, *Strengthening the Global Semiconductor Supply Chain in an Uncertain Era* (Boston, MA: Boston Consulting Group and Semiconductor Industry Association, April 2021).

[c]Nurzat Baisakova and Jan-Peter Kleinhans, "The Global Semiconductor Value Chain: A Technology Primer for Policy Makers," Stiftung Neue Verantwortung, October 2020.

[d]The five categories are deposition tools, dry/wet etch and cleaning, doping equipment, process control, and testers. Varas et al., *Strengthening the Global Semiconductor Supply Chain*.

[e]The White House, *Building Resilient Supply Chains, Revitalizing American Manufacturing, and Fostering Broad-Based Growth: 100-Day Reviews under Executive Order 14017*, June 2021.

[f]Baisakova and Kleinhans, "Global Semiconductor Value Chain."

[g]Nvidia, "NVIDIA and SoftBank Group Announce Termination of NVIDIA's Acquisition of Arm Limited," press release, February 7, 2022.

Chip manufacturing or fabrication capabilities in particular are at the center of geopolitical tensions over semiconductors, along with controlling access to the technology that enables design and manufacturing. The chip fabrication stage of the supply chain has the following key features:

- *Production is highly concentrated*. Enormous R&D and capex costs of leading-edge production have seen regional and industrial concentration. Leading-edge fabs cost up to $20 billion to build.[29]
- Leading-edge logic volumes are very low yet generate substantial portions of revenue and device integrator/OEM economic activity. One source shows that less than 5 percent of global volumes in 2019 were below 10nm (although precise measurements are not possible given the difficulty of directly comparing the process technology of different companies).[30]
- *At the leading edge for logic, Samsung and TSMC dominate*. Only TSMC and Samsung are producing commercial volumes of the leading-edge 3nm and 5nm chips. TSMC is one to two years ahead of Samsung and two to three years ahead of Intel.
- *Intel has fallen behind in leading-edge logic*. Intel encountered delays with its 14nm and 10nm technologies, and its 7nm (roughly equivalent to TSMC's 5nm) production has been further delayed. These cumulative delays partly explain TSMC's recent ascension to leadership alongside its business strategy to focus on these more profitable leading-edge logic chips since securing Apple as a key customer a decade ago.[31]
- *SMIC (China) is pursuing both leading-edge and mature-logic fabrication*. China's national chip manufacturing champion has achieved commercial production at 14nm, and it may have shipped small volumes of products with aspects of 7nm technology by early 2022.[32] It has also invested heavily in less-profitable mature-logic manufacturing capacity.
- *EUV equipment conveys a major advantage at the leading edge*. Commercial production at 5nm and below is greatly facilitated (and made profitable) by extreme ultraviolet (EUV) lithography

equipment, for which the Netherlands' ASML is the monopoly supplier.[33] China does not have access to this technology due to an export control agreement between the US and Netherlands governments.

The United States has never had a credible pure-play (contract) foundry company. The foundry concept was pioneered by TSMC in Taiwan in 1987, and today, essentially all pure-play foundries are in Asia. It is worth noting, however, that while TSMC is headquartered in Taiwan and manufactures most of its chips there, the firm is publicly traded: about 75 percent of its shares are foreign owned (with US entities as top shareholders), and half its board members are US citizens.[34] Meanwhile, US-headquartered (and majority foreign-owned[35]) GlobalFoundries is a much smaller player (6 percent of the global contract foundry market) than TSMC (56 percent) and Samsung (16 percent), and it does not compete in the latest technology nodes.[36] Despite some attempts, even during periods when Intel was successful in manufacturing chips for its own use, Intel was never successful in the contract foundry model—a failure that has been attributed to company cultural as opposed to technical barriers.

While Intel's capabilities as a traditional vertically integrated device manufacturer (IDM)—functioning across the value chain of both chip design and manufacturing—gave the United States a strong position in logic chip manufacturing for many years, it has repeatedly stumbled in the past five to ten years. Intel held a three-year lead (at least one node generation) until recently. But its 14nm node was one year late, its 10nm node was three years late, and now its 7nm node is expected be at least two years late. These delays are cumulative—so what was a three-year lead six years ago is now a three-year lag. TSMC (and, to a lesser degree, Samsung) has more advanced digital technology today than Intel does; even Intel now outsources the manufacture of its most advanced chips to TSMC since it cannot build them in-house.[37] It is the opinion of many US industry observers that this situation is unlikely to change in the near term despite Intel's stated plans to regain leadership. While Intel now claims to have fixed the internal problems that led to

high defect rates and delays, the real test will be whether it ships its newest technologies.

Meanwhile, the vast majority of memory (and, more recently, storage) chip production has been in Asia for decades. In the chip manufacturing domain, Intel stopped making memory chips (DRAM) more than thirty years ago, and sold its NAND Flash business to SK hynix in 2012. While the jury is still out, memory and storage has emerged as an area of notable early success for China's chipmakers.

For example, the China-based storage (NAND Flash) company Yangtze Memory Technologies Co. (YMTC) started from a blank slate in Wuhan, China, in 2016. It pursued a technology that was rather new to the NAND Flash storage companies at the time—the use of copper-to-copper hybrid bonding to stack a conventional logic wafer on top of a flash storage array wafer. The mainstream storage chip companies generally ignored this approach. Today, however, all the flash storage companies are considering using the same approach. YMTC has traditionally offered low-end products (e.g., USB storage sticks), but more recently it gained attention when Apple considered using YMTC NAND Flash chips for China's domestic market iPhones (but dropped its plan due to US export controls imposed on YMTC).[38] Established global competitors are all watching how YMTC might grow and eat into the higher-end markets as well.

Similarly, the DRAM memory company ChangXin Memory Technologies (CXMT) was started in Hefei, China, in 2016—a surprising market to enter, as DRAM is considered a difficult segment in which to make money (hence the general absence of US firms). Today, CXMT offers DDR4 DRAM products. While its market share is small (a few percent), CXMT now has a strong plausible growth story: first, leading-edge DRAM technology development has significantly slowed down due to an inability to continue to miniaturize the memory cell (now progressing at just one or two nanometers of improvement per generation, while a paradigm shift from two- to three-dimensional memory architecture has not yet happened); and second, commodity DRAM memory products have standard interfaces, so there is little distinction among manufacturers as long as the products meet industry

specification standards. As of mid-2022, company leadership was confident that CXMT, despite being a very late entry in the space, could catch up to the global leading edge within three to four product generations (from today's 14–15nm leading node for DRAM memory to a future 10nm leading node).

US firms (e.g., Texas Instruments, Analog Devices, and ON Semiconductor) competitively manufacture specialty products like analog chips or wide-bandgap power management devices. These are worldwide segments with production distributed among many more players than in other parts of the semiconductor supply chain. While these specialty products use lower-resolution lithography, the process technologies themselves are sophisticated and require substantial R&D efforts to sustain commercial development. Complex systems (e.g., vehicles or weapons systems) require these technologies, and the United States remains a world-class manufacturer of these products. Even so, despite the more distributed global production of these analog and power management chips, there are growing concerns that China's efforts at chip self-sufficiency—and associated subsidies of lower-margin or unprofitable domestic manufacturers—may one day flood the global market with these older-node and specialty products. Research and development of these specialty technologies does not require sophisticated (Western) equipment—just talented people. As a result, China's focused efforts in this area, intentionally or not, could kill off market-based Western or other Asian competition via consistent underpricing to gain a controlling advantage over what was once a distributed global supply chain.

It is also helpful for industry outsiders to appreciate the deep sense held by semiconductor manufacturers—especially when chip shortages are in the news and on the minds of customers—that the semiconductor industry has been a boom-and-bust industry. This mentality is borne out by the financial histories of even today's leading and very profitable manufacturing firms such as TSMC or Samsung. One driver of this boom-bust cycle is the fact that chip manufacturers cannot gradually add incremental capacity: a new fab costs around $20 billion and provides a huge increase in capacity; customer markets, meanwhile,

tend to grow gradually, so bringing a new fab online almost guarantees that there will be overcapacity for some period of time.[39] Even with the broad growth in future demand for semiconductors from seemingly every sector of the economy and across countries, this underlying boom-and-bust dynamic isn't likely to change going forward—at least for leading-edge technology. This boom-bust phenomenon also is blamed for the resistance of manufacturing firms to invest in new capacity of more mature trailing-edge chips, which (given their ubiquity on final products) can be the source of many supply bottlenecks, but which (given their commodity nature) are generally even less profitable than leading-edge chips. In short, once there is a shortage, market conditions will change by the time a new fab can come online.

Beyond fabrication, as described in table 2.2 above, the United States has a strong position in semiconductor manufacturing equipment (with companies such as Applied Materials, Lam Research, KLA, and others). ASML is the only supplier of EUV lithography tools but also supplies manufacturing equipment for a variety of mature nodes, including widely sold deep ultraviolet (DUV) lithography machines.[40] And as for chip manufacturing inputs, the silicon wafers on which chip designs are fabricated are mostly manufactured in Asia, while Japan is a strong player in the variety of pure chemicals needed for semiconductor manufacturing steps.

The United States further has a strong position in electronic design automation software tools, with Synopsys and Cadence being the leading worldwide suppliers. But given the perceived possibility in recent years that the US government could declare these software tools to be "foundational technologies" and prevent their sale to companies in China, indigenous software tools and indigenous equipment manufacturers have become an attractive area of private sector commercial investment in China, driven by the assumption that such firms could effectively capture a hugely subsidized and rapidly growing domestic chip industry. Developing design software is easier—or at least cheaper—than building a fab, and all it takes is time and talented people who are not subject to export controls. Moreover, there are already competent electronic design automation software startups in China,

and China-based companies will become competitive in tools sooner rather than later.

Trends in Commercial Technology

"Moore's Law"—named after Intel's Gordon Moore—observes a general trend in the semiconductor industry: the number of transistors on commercial microchips doubles every two years while costs fall. This is not a natural law, of course, but rather a self-fulfilling prophecy that has been borne out by R&D, continued investment by leading companies, and intense competition.

For the past fifty years, a primary enabler of such advancement has been the continued "two-dimensional" reduction in chip element sizes, from the larger to the smaller nanometer node sizes described earlier. Two-dimensional downscaling—that is, making devices (such as transistors, memory, and the wires) smaller and smaller—allows manufacturers to pack more components on the same chip area and thereby achieve lower cost per function. This pathway has given structure and predictability to the semiconductor industry, offering a clear road map of the state of the art year by year.

Recently, however, that pathway of two-dimensional scaling is reaching saturation, primarily driven by the escalating cost of pushing to ever-smaller process nodes. But there are many other avenues for chip technologies to progress. This diversification of chip technology pathways has led to discussion not of the death of Moore's Law but of the "post-Moore era," when innovation will be driven instead by the way chips are fabricated, stacked, and packaged, and by how networks of chips (sometimes called "chiplets") can be made to interact with each other or deployed for application-specific purposes.[41] For example, to provide more components on the same chip, one can go to three dimensions, like (analogously) building high-rises in Manhattan. As another example: instead of using silicon transistors to perform all the functions desired of a chip, one can use other materials and other devices that are fabricated specifically for a certain function, and then integrate these functions together on the same chip. It may take more

or different kinds of innovation to build high-rises and to come up with new materials and devices, but there are no physical limits to doing so. In that sense, the future of chip technologies is full of possibilities. Certainly, three-dimensional chips, advanced packaging (chiplets), and application-tailored "heterogenous compute" device technology and chip architectures will play a role. But the path forward is more diversified than it used to be.[42] And in this new paradigm, it's not clear who has the technology advantage. Whoever figures out how to make progress will likely become the leader, which has implications for the use of any policy tool that seeks to encourage—or impede—technological progress.

That changing semiconductor technology landscape is likely to affect industry structure. For example, it used to be the case that the development of chip-level manufacturing technologies could be carried out somewhat independently from the design of the system where the chip would be used (e.g., a smartphone). The abstraction boundary has worked quite well in the past, in that these two activities could occur in parallel because the trajectories of the two activities were clear and predictable (due to the predictable trend of two-dimensional device miniaturization of approximately 0.7 times per generation). Now, because performance (broadly defined as not just the speed of computation but all aspects of performance, including energy efficiency and power consumption) gains are harder to achieve, firms increasingly need to codesign the system with the chip technology. Both sides need to optimize the engineering trade-offs together. The result is that chip design companies (such as Apple, Nvidia, AMD, and Qualcomm) are working even more closely with the foundries throughout the entire design cycle, from early conception to final product. Among other things, this collaboration requires a substantial degree of trust among leading designers, fabs, and system integrators: they need to share not only product road maps but also innovation ideas that are not yet proven.

What might this trend mean for the relative future competitiveness of today's national semiconductor manufacturing champions?

One possibility is that, as two-dimensional scaling slows down across the industry, progress is going to come from system-specific

(or domain-specific) technologies. This possibility means that system companies—not chip manufacturers—are increasingly likely to steer technology directions. For example, companies like Apple may start to work with new companies that specialize in advanced packaging as a service, so that they can use chiplets from multiple suppliers and build their own "2.5"-dimensional and three-dimensional technologies. More focus on heterogenous—i.e., specialized—computing applications may even drive large customers to develop more specialized IP blocks such as AI accelerators, moving beyond, for example, general-purpose ARM or x86 architectures as the main compute core.

Another possibility is essentially the opposite—that due to the consolidation of fabricators (including advanced packaging), chip production technology developers and manufacturers may occupy a more commanding position in the value chain due to their integrations with both designers and systems integrators/OEMs. This outcome would further increase the capital-intensity and barriers to entry at the leading edge of the chip industry.

Still one more possibility is that an incumbent leader like TSMC may slow down through such a complex transition, which could make it easier for Intel or other challengers to catch up.

There may even be new business models due to the changing industry dynamics. For example, chip customers are coinvesting up front in building manufacturing capacity with the foundries to ensure adequate supply. Indeed, automobile manufacturers—which often obtain chips through chip design companies (e.g., Infineon, NXP, and Renesas) and thus are second-tier customers of foundries—are increasing interaction with foundries directly to take more control of the supply of this increasingly central component of their products.[43]

What about the implications for China's semiconductor technology competitiveness as it attempts to rapidly advance? Unlike in the past, there are now many possible paths for advancing chip technology—so it is certainly possible that China may pick the right path and pull ahead of the rest of the world. Still, China does not appear to be focusing on the development of particular technology pathways that are more promising than those other countries are pursuing. The semiconductor

research community, much like the broader semiconductor supply chain itself, is global and is highly competitive.

Beyond this likely transition from single-minded two-dimensional miniaturization to a plethora of ways to advance chip technology, are there more fundamental technology "leapfrog" applications that could enable China's chip firms to seize a more commanding market position?

Recently, for example, the People's Republic of China's (PRC) government has trumpeted the arrival of "third-generation" semiconductors.[44] These refer specifically to wide-bandgap semiconductors, which use more exotic materials—for example, silicon carbide (SiC), gallium nitride (GaN), and diamonds. The term *wide-bandgap* derives from the closer spacing of the atoms in these materials, which results in stronger atomic bonds and wider electrical bandgaps. But while the term *third-generation* suggests an evolution or even an advance from the first- and second-generation semiconductors, these wide-bandgap third-generation semiconductors are not a replacement for or successor to the foundational silicon-complementary, metal-oxide semiconductor (CMOS) material system, nor do they necessarily constitute a straight advancement of semiconductor technology from first generation (silicon) to second generation ("III-V" materials used in optoelectronics and microwaves). Each of these so-called generations serves very different applications and markets, and they are under the big tent of semiconductor technology. The applications of these wide-bandgap semiconductors, for example, are principally in high-voltage and high-power electronics. These are use/case-specific technologies that are also very important for electric vehicles, the electric power grid, and battery management—intrinsically very important, and likely increasingly so, but not replacements for other semiconductor applications.

Quantum computing is another emergent technology that is often mentioned in discussions of "leapfrog" potential. It is a very specialized technique to solve a very limited—though very important—class of problems. An analogy for quantum computing might be along the lines of the use of lasers in light—extremely useful for certain applications and enabling of new technologies, but not a replacement for general

lighting. Thus, in the foreseeable future, quantum computing should not be considered an alternative to semiconductor technology, and the use of quantum computing will not be as ubiquitous as conventional semiconductor chip technology; revenues from the quantum computing industry round to zero percent of the semiconductor market, and that will be true for some time. On the contrary, for quantum computing to become a practical technology, it would need very sophisticated semiconductor technology to serve the control and signal-processing functions required for a practical system, and it would likely use the same fabrication infrastructure as today's microelectronics.

In summary, the technology trends today and going forward suggest a closer coupling among various parts of the semiconductor value chain. The emergence of the foundry and fabless model heralded a decoupling of the various parts of the value chain. Now, we see the opposite. Fabless chip design companies have to work closely with chip manufacturing companies, and the chip manufacturing companies in turn have to work closely with materials and equipment suppliers. While we may not see the revival of firm-level integrated device manufacturers like Intel or Samsung due to the need for economy of scale and the high degree of specialization, technology trends will reward tighter integration overall. Tighter integration will change the dynamics and ecosystem of the entire industry. In short, we are still at the beginning of this evolution, and we do not know where it will take us.

Trends in Research and Development

The United States has enjoyed a leading position in university-based R&D in semiconductors for decades. Today, however, China has at least as broad a set of university programs as the United States does, and China likely has many more PhD students working on traditional semiconductor (silicon) technology and devices. The leading semiconductor R&D journals (published by the Institute of Electrical and Electronics Engineers, IEEE) are now dominated either by university-based submissions from China or by the Belgium-based R&D consortium imec, which represents the industrial R&D community.

This shift is arguably the result of a severe underfunding of academic research on semiconductors in the United States for over two decades. For example, after nurturing the early semiconductor industry through government procurement, and later funding major R&D efforts through the 1980s and again in the 1990s (through the industry consortium Semiconductor Manufacturing Technology [SEMATECH]), the DoD's semiconductor R&D funding focus diverged from that of the much larger commercial R&D market to more narrowly targeted, longer-term niche technologies with specific defense applications.[45] The Department of Energy, meanwhile, primarily funds fundamental, basic science and high-performance computing (i.e., building supercomputers), but it typically does not sponsor research between these two extremes. The National Science Foundation, meanwhile, institutionally focuses on "science" and "discovery" while generally undervaluing "engineering" and "technology" and the translation to industry, which is most important to semiconductors. Overall, as compared to the 1990s, semiconductor research in US academia has been stagnant: an analysis of the combined top paper presentations at the International Electron Devices Meeting (IEDM), the International Solid-State Circuits Conference (ISSCC), and the Symposia on VLSI Technology and Circuits (VLSI) shows that, between 1995 and 2020, US-authored papers sustained a roughly 40 percent share of the total, while combined papers from Taiwan and South Korea grew from just 6 percent to 26 percent, and those from China grew from nothing to 10 percent.[46]

Meanwhile, the same changes in technology development described in the commercial industry—slowing progress in two-dimensional lateral scaling (miniaturization) of device dimensions—will also affect R&D patterns now. This result was predictable, and the challenges were universally understood. In fact, for over a quarter century, an actual international road map (ITRS, the International Technology Roadmap for Semiconductors) guided chip R&D in industry and in universities. This road map coordinated R&D programs, including national programs such as the Semiconductor Research Corporation (SRC), because everyone had a common picture of where the industry

was heading. The last edition of the ITRS was updated in 2013, and there is no equivalent industry road map today. As discussed above, with the slowing down of lateral scaling because of physical limits, the path forward is a lot less clear.

Many of the possible paths forward involve technologies that are difficult for universities to contribute to. For example, advanced packaging tools are not commonly found in university labs, and currently there is no national shared facility that researchers can access to work on these kinds of R&D problems. Furthermore, the solutions will likely be specific to particular systems. For example, getting better system performance in data centers is likely a different problem than getting better system performance in power management. Single-point solutions such as improving traditional silicon CMOS transistor density, which used to be the solution, are likely not the solutions of the future.

This all may require a rethinking of university R&D structures and approaches. In particular, academia should work even more closely with the chip industry; while industry by and large has known what needed to be done in the past five decades, today the future paths for advancing chip technology are more ambiguous. And industry may increasingly rely on academic and lab-based research to explore possible paths forward, because university research is more nimble and less costly.

R&D and manufacturing form a close symbiotic relationship. Manufacturing without R&D is not sustainable because a company must have a pipeline of future products. And R&D without manufacturing is like building a bridge to nowhere—research in isolation may lead to technologies that are not manufacturable.

In sum, despite the narratives of the past two decades, semiconductor knowledge and advancements today do in fact provide foreign countries a fair degree of asymmetric advantage over the United States—especially when combined with the ability to manufacture.

Trends in Workforce

While the challenges of STEM education in the United States in general are well documented, the semiconductor industry exhibits specific

structural problems that more fellowships, internships, and stipends for STEM graduate students will not solve. Given current and expected needs, it is not accurate to say that the United States has a workforce shortage problem in semiconductors. Rather, there are structural and incentive problems in the industry.

The word *manufacturing* may conjure the image of traditional factory work that can be performed with skill levels at the technician level—but a majority of the work in modern microelectronics manufacturing requires a relatively high skill level. While TSMC is sometimes regarded in the industry as being particularly reliant on advanced-skill workers given its emphasis on the leading edge of logic chip innovation, it is nonetheless an important benchmark: as of 2022, 79 percent of TSMC employees had at least a bachelor's degree, (strikingly) 51 percent had at least a master's degree, and 4 percent had a PhD.[47] Going forward, advanced-semiconductor manufacturing will increasingly rely on automation, data analytics, and artificial intelligence (AI)—and the R&D for those next-generation technologies, for which the path forward is less clear (as discussed earlier), will increasingly require PhD-level engineers.

Given the foundational nature of semiconductors in the economy, without continued advancements in semiconductor technology, it will be difficult to fulfill the high expectations now placed on adjacent future technology applications, such as AI, 5G, quantum computing, or self-driving cars. The phrase *continued advancement* is therefore often used in industry because semiconductors must constantly improve to showcase their value to society. In other words, semiconductors are not a commodity like oil—the value of semiconductors is in their ability to do more year after year. Manufacturing and R&D are both needed to achieve these next generations. Thus, a highly skilled talent pool is required.

It is important to recognize that engineering graduates in the United States (including foreign graduates of US universities) are not decreasing—they are in fact increasing. So the semiconductor workforce issue is not so much the total number of technical graduates in the United States, but rather the choices those graduates are making about career paths they wish to follow—and in which countries they follow them.

Consider that in the US technology sector today (and even within the US semiconductor value chain itself), talent tends to be attracted to end-product consumer-facing companies, and not the companies that make the components (such as the chips themselves, or the tools and equipment that produce the chips). The skills of graduates are transferable across the value chain—for example, someone who is skilled in algorithms and consumer software is equally valuable for a chip electronic design automation software company as for a social media company. Algorithm and data science skills that are useful for developing next-generation AI-enabled chip manufacturing systems are similarly valuable to financial technology (fintech) companies. And someone who is skilled in manufacturing semiconductors is also valuable to companies that design consumer-facing electronics products using those same chips. To construct fabs, one needs a variety of both technicians with trade skills and highly educated engineers. To run a fab at the leading edge, technicians are still required, but firms mostly need engineers with the ability to understand and analyze data from the fabrication process, to report problems, and to make decisions on the fly. In Taiwan's experience, that means most fab engineers have at least a bachelor's or master's degree. And culturally, working for a fab in Taiwan today means good pay and high social cachet—TSMC is seen as one of Taiwan's most desirable employers, and new hiring windows make national news.[48]

Meanwhile, in the United States, highly skilled engineering and technology graduates tend instead to be consumed by companies that make the end products—such as Apple, Google, or even Nvidia—and many of them are software or systems applications companies. Students can imagine how, working at these companies, their engineering talents can lead to exciting products. Even though chips are at the heart of these products, they are nonetheless largely invisible. Perhaps more importantly, mirroring trends in service versus manufacturing sectors elsewhere, profits tend to accrue to these firms that produce differentiated customer-facing end products, not the companies that make the chips. As a crude illustration, consider that the price difference between an iPhone that has 256 GB versus one with 128 GB of storage is $100, but the price difference between the chips themselves is $8. Thus, it's easy

to see why companies like Apple or Google or Facebook might have nicer cafeterias and higher compensation than semiconductor companies. The value capture and compensation differences are substantial. Indeed, a 2022 McKinsey analysis noted that US employees rank semiconductor firms lower than consumer technology or even automotive employers not just in compensation, but also in spillovers to work-life balance, perceived quality of senior management, firm culture, and overall perceived career opportunities.[49]

Assuming the demand side of semiconductor workforce development is eventually solved, and assuming there is a new interest from US students wanting to be trained in semiconductors, the best way to meet that demand will be to then increase R&D funding in semiconductors and to build up semiconductor research and teaching facilities on university campuses. This approach will grow the number and quality of professors and instructors in the field of semiconductors, making available both quality course offerings and hands-on training opportunities that are vital to excellence in technology development and manufacture. Quality professors and hands-on training experiences can also help today's technical graduates see how careers in the chip industry might let them apply their skills toward meaningful, global-scale societal problems such as the environment, human health, or AI that they consider more important than consumer electronics.

In sum, the semiconductor industry arguably has an unaddressed structural issue: even though developing semiconductor technology requires top talent, US chip companies are generally not able to offer the money or the excitement to acquire it. Improving this situation, as discussed further in chapter 4 of this report, should not be left to market dynamics—it should be an objective of policy.

NOTES

1. H.-S. Philip Wong, Kerem Akarvardar, Dimitri Antoniadis, Jeffrey Bokor, Chenming Hu, Tsu-Jae King-Liu, Subhasish Mitra, James D. Plummer, and Sayeef Salahuddin, "A Density Metric for Semiconductor Technology [Point of View]," *Proceedings of the IEEE* 108, no. 4 (April 2020): 478–82.

2. Given this, there have been proposals to replace the nanometer naming convention with a density metric instead, but this approach has yet to gain industry traction.
3. FPGAs are modular, multipurpose chips that can be flexibly programmed and reprogrammed by the end user for various tasks. They are commonly used today in data center applications; see below on FPGA defense sector applications as well.
4. TSMC, *2022 Second Quarter Earnings Conference*, July 14, 2022.
5. TechInsights (blog), "SMIC 7nm Technology Found in MinerVa Bitcoin Miner," accessed May 13, 2023.
6. Antonio Varas, Raj Varadarajan, Jimmy Goodrich, and Falan Yinug, *Strengthening the Global Semiconductor Supply Chain in an Uncertain Era* (Boston, MA: Boston Consulting Group and Semiconductor Industry Association, April 2021).
7. Cars with autonomous or driver-assistance capabilities also now increasingly require leading-edge logic chips, given the computational complexity of these real-time processes. For more on contemporary automotive chip needs, see Jack Ewing and Neal E. Boudette, "A Tiny Part's Big Ripple: Global Chip Shortage Hobbles the Auto Industry," *New York Times*, April 23, 2021.
8. Varas et al., *Strengthening the Global Semiconductor Supply Chain*.
9. Both DRAM chips and NAND Flash chips can be referred to as either memory or storage chips in that they both store digital information. The differences are in how fast they can be written and read, how long they can store information, the density of storage they provide, and the cost per bit of storage. Generally, DRAM (which is quite fast) is placed in close architectural proximity to a computer's CPU logic chip. NAND Flash chips are slower and are further away (architecturally speaking, not necessarily physically speaking). Generally, the slower a memory technology is, the higher the density of storage, and hence the lower cost per bit it provides.
10. Yuka Hayashi, "Chip Shortage Limits US's Ability to Supply Weapons to Ukraine, Commerce Secretary Says," *Wall Street Journal*, April 27, 2022.
11. Vikram Mittal, "US Soldiers' Burden of Power: More Electronics Means Lugging More Batteries," *Forbes*, October 26, 2020.
12. See, for example, comments in US Government Accountability Office, *Semiconductor Supply Chain: Policy Considerations from Selected Experts for Reducing Risks and Mitigating Shortages*, GAO-22-105923 (Washington, DC: US GAO, July 2022).
13. These beginnings are well documented by Chris Miller in his history of the industry, *Chip Wars* (New York: Scribner, 2022).
14. Jon Y, "The Government Semiconductor Chip Buying Problem," *The Asianometry Newsletter*, November 3, 2021.

15. Shujai Shivakumar and Charles Wessner, "Semi-Conductors and National Security: What Are the Stakes?," Center for Strategic and International Studies, June 8, 2022.
16. Representing eighty-one suppliers as of August 2022. For more details, see "DMEA Trusted IC Program," Defense Microelectronics Activity, accessed May 13, 2023.
17. Mark Lapedus, "A Crisis in DoD's Trusted Foundry Program?," Semiconductor Engineering, October 22, 2018.
18. Michaela D. Platzer, John F. Sargent Jr., and Karen M. Sutter, *Semiconductors: US Industry, Global Competition, and Federal Policy* (Washington, DC: Congressional Research Service, October 26, 2020).
19. See Brad R. Williams, "DoD Seeks $2.3B to Bolster US Chip Making," *Breaking Defense*, June 4, 2021.
20. A recent public review by the DoD Office of Inspector General reports some delays in this process; a provision of the 2023 National Defense Authorization Act mandates additional reviews to this end. See Department of Defense Office of Inspector General, "Evaluation of the Department of Defense's Transition from a Trusted Foundry Model to a Quantifiable Assurance Method for Procuring Custom Microelectronics (DODIG-2022-084)," May 4, 2022.
21. Eric Lee, "How Taiwan Underwrites the US Defense Industrial Complex," *The Diplomat*, November 9, 2021.
22. Keith Holbert and Lawrence Clark, "Radiation Hardened Electronics Destined for Severe Nuclear Reactor Environments," US Department of Energy, February 19, 2016.
23. Jonathan Pellish, "A New Market for Terrestrial Single-Event Effects: Autonomous Vehicles," NASA, May 2019.
24. Xilinx calls this chip an adaptive compute acceleration platform (ACAP).
25. Michael Johnson, ed., "A New Approach to Radiation Hardening Computers," NASA, last updated August 12, 2022.
26. "Radiation Hardened Electronics Market by Component," MarketsandMarkets, May 2022.
27. Examples include ultraviolet lithography tools, which use certain wavelengths of light to print the circuit pattern onto the silicon wafer.
28. Including South Korea, Japan, China, and Taiwan. Southeast Asian nations such as Singapore, Vietnam, Malaysia, and the Philippines are also starting to play larger roles, especially in outsourced assembly, packaging, and testing (OSAT). See chapter 6 for more detail.
29. Semiconductor Industry Association, "US Needs Greater Manufacturing Incentives," July 2020.
30. Varas et al., *Strengthening the Global Semiconductor Supply Chain*.

31. Pushkar Ranade, "The Apple-TSMC Partnership," Bits and Bytes, March 6, 2022.
32. Given that this SMIC 7nm chip lacks static random access memory (SRAM), which typically takes up half the area of a processor chip, it is probably inaccurate to describe this as a bona fide 7nm technology. Claims of 7nm originally reported in "SMIC 7nm Technology Found in MinerVa Bitcoin Miner," *TechInsights*, accessed May 13, 2023.
33. It is possible to produce 5nm technology without EUV, but not profitably at scale. The development of EUV by ASML was supported by investments from Intel, Samsung, and TSMC. Intel was the most visible early supporter.
34. TSMC, *Annual Report 2021(I)*, March 12, 2022, 66.
35. Wallace Witkowski, "GlobalFoundries IPO: 5 Things to Know about the Chip Company Going Public in a Semiconductor Shortage," *MarketWatch*, last updated October 23, 2021.
36. Foundry revenue figures as of the third quarter of 2022. See TrendForce, "Global Top 10 Foundries' Total Revenue Up 6% in 3Q 2022," *EE Times Asia*, December 16, 2022.
37. Industry trade press in 2021 reported that Intel had contracted with TSMC for a significant fraction of TSMC's 3nm production capacity. See, for example, Monica Chen and Jessie Shen, "TSMC to Make 3nm Chips for Intel, Sources Claim," *DIGITIMES Asia*, January 27, 2021.
38. Reuters, "Apple Freezes Plans to Use China's YMTC Chips—Nikkei," October 17, 2022.
39. Indeed, following the widely reported chip shortages of 2021 (and broad growth in semiconductor industry profits), financial results of Q2 2022 showed market softening. In fact, on the same day that the CHIPS and Science Act of 2022 passed the Senate, Intel announced its first net loss in thirty years and a reduction in expected capex on account of weaker demand. Global industry financial results were soft across Q4 2022 and into Q1 2023. See Dylan Patel, "Intel Cuts Fab Buildout by $4B to Pay Billions in Dividends," Semianalysis, July 28, 2022; and Nicholas Gordon, "Chip Glut Batters Semiconductor Industry as Intel Shares Lose Almost All Their 2023 Gains after Dismal Earnings," *Fortune*, January 27, 2023.
40. ASML has suggested that 90 percent of equipment it has ever sold is still in use on chip manufacturing lines.
41. Samuel K. Moore, "The Transistor of 2047: Expert Predictions," *IEEE Spectrum*, November 21, 2022.
42. Mark Liu, "TSMC Chairman Mark Liu Describes How the World's Largest Chipmaker Is Reimagining the Semiconductor Industry," *Fortune*, June 8, 2022.

43. See, for example, the Sony-TSMC joint ventures on research and fabrication in Japan, and the GM-GlobalFoundries joint investment. Masaharu Ban, "Sony Begins Funding TSMC's Japanese Chip Plant," *Nikkei Asia*, January 26, 2022; and Jane Lee, Joseph White, and Steven Nellis, "GM Inks Deal with GlobalFoundries to Secure US-Made Chips," Reuters, February 9, 2023.
44. Shiyin Chen, Yuan Gao, and Abhishek Vishnoi, "Xi Jinping Picks Top Lieutenant to Lead China's Chip Battle against US," *Bloomberg News*, June 16, 2021.
45. Marko M. G. Slusarczuk and Richard Van Atta, "The Tunnel at the End of the Light: The Future of the US Semiconductor Industry," *Issues in Science and Technology* 28, no. 3 (Spring 2012).
46. Japan saw a major decline during this period alongside its loss of global semiconductor manufacturing leadership, from 39 percent in 1995 to just 10 percent in 2020, equaling China. See Jan-Peter Kleinhans, Julia Hess, Pegah Maham, and Anna Semenova, "Who Is Developing the Chips of the Future?," Stiftung Neue Verantwortung, June 16, 2021.
47. TSMC, *Annual Report 2021(I)*.
48. Cheng Hung-ta and Frances Huang, "TSMC, Evergreen, Fubon Financial Dream Employers for Job Changers: Poll," *Focus Taiwan*, January 14, 2023.
49. Ondrej Burkacky, Marc de Jong, and Julia Dragon, "Strategies to Lead in the Semiconductor World," McKinsey & Company, April 15, 2022.

A report of the Working Group on
Semiconductors and the Security of the
United States and Taiwan, a joint project of
the Hoover Institution and the Asia Society
Center on U.S.-China Relations