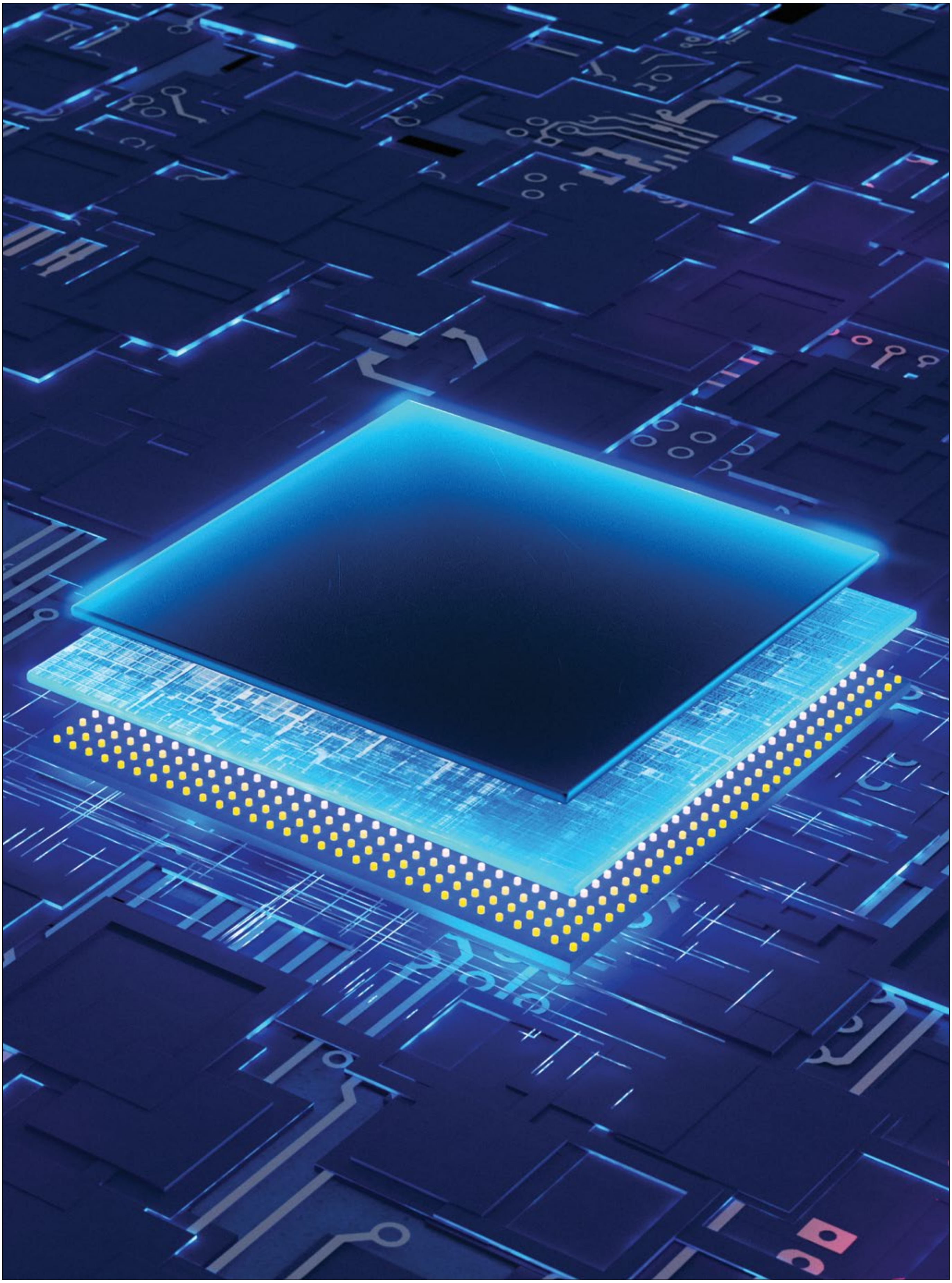
STANFORD UNIVERSITY

THE STANFORD EMERGING TECHNOLOGY REVIEW 2023

A Report on Ten Key Technologies and Their Policy Implications

CHAIRED BY Condoleezza Rice, John B. Taylor, Jennifer Widom, and Amy Zegart

DIRECTED BY Herbert S. Lin



08

SEMICONDUCTORS

KEY TAKEAWAYS

- Moore's law, which for fifty years has predicted rapid increases in semiconductor capabilities at decreasing costs, is now ending, raising profound implications for the future of hardware and software development.

- Recent research has identified methods that allow innovations in materials, devices, fabrication, and hardware to be added to existing processes or systems at low incremental cost. These methods need to be further developed since they will be essential to continue to improve the computing infrastructure we all depend on.

- Quantum computing may solve certain specialized problems, but experts debate whether it can ever achieve the rapid, consistent, predictable performance growth that semiconductors have enjoyed.

Overview

Semiconductors, or microchips, are crucial components used in everything from refrigerators and toys to smartphones, cars, computers, and fighter jets. These microchips are ubiquitous in modern medical equipment. Imaging devices such as magnetic resonance imaging (MRI), computed tomography (CT), and ultrasound use embedded computers to generate images from electromagnetic radiation and sound waves that penetrate or emanate from the human body. Precision robotic surgery would not be possible without digital-to-analog converter chips. Across these examples, innovation in chip design, new materials, and integration methods helps enable the performance, size, and efficiency of medical devices.[1]

A semiconductor is a material whose ability to conduct electricity lies between those of metals (with high electrical conductivity) and insulators (with low conductivity). Its conducting properties can be dramatically changed by adding impurities known as dopants into the material. The element silicon is the most commonly used semiconductor today, with the compound gallium nitride far behind in second place. This ability to change the state of the material allows one to build electrical devices such as transistors, which are the fundamental building block of all the information technology we use today.

Semiconductors are also used to create light-emitting diodes, or LEDs, for lighting and display applications, solar cells, electronics in power systems to control and manage the flow of electricity, and many other devices. This section focuses on semiconductors for computers and other information technology.

The first transistor was invented in 1949 and its size was measured in centimeters. Transistors were individual electronic devices that could be wired together along with other circuit components to do useful things, like function as a radio. About a decade later, the integrated circuit (IC), or chip, was invented; the IC integrates transistors and other circuit components into one physical package.

Different types of chips do different things in computers. Some chips are processors—they ingest digital data, perform computations on them, and output the results of those computations. Memory chips store information and are used with processors.

Many physical devices in daily life use chips, including computers, mobile phones and smartphones, cars, airplanes, washing machines, toasters, microwave ovens, televisions, refrigerators, lawn mowers, cameras, and so on. Chips control heating and cooling systems, elevators, fire suppression systems, and fire alarms in modern buildings. Traffic lights are controlled by chips. On farms, tractors and combines and irrigation systems are controlled by chips. Modern militaries could not function without chips in their weapons, navigation, and cockpit life support systems, or without computers to manage their complex logistics. The list goes on—in every aspect of modern life, chips are essential.

Two aspects of chips are important for our purposes here. Chips must be designed and then manufactured, both of which call for different skill sets. Chip design is primarily an intellectual task that requires tools and teams able to create and test systems containing billions of components. Fabrication has a physical component that calls for large factories, or "fab facilities," that can produce chips by the millions and billions. In 2022, a record 1.15 trillion chips were shipped worldwide.[2] But fabrication also entails a substantial degree of process engineering to continue to improve process technology and to achieve the stringent manufacturing standards. For example, the "clean rooms" in which chips are made require air that is one thousand times more particle-free than what is in a hospital operating room.[3]

Because chip design and chip fabrication are so different in character, only a very few companies do both. Intel is one such company. Many companies specialize in design—Qualcomm, Broadcom, Apple, and Nvidia. Such companies are also called "fabless" in recognition of the fact that they do the design work and outsource fabrication to others on the theory that the former has relatively high market value and the latter has relatively low market value.

Today, "others" usually means one company: Taiwan Semiconductor Manufacturing Company (TSMC), which is by far the world's largest contract chip manufacturing company. Samsung in South Korea is a distant second, and the United Microelectronics Corporation (UMC), also in Taiwan, ranks third. With a large fraction of the world's chip-manufacturing capacity in Taiwan, the global supply chain for chips is clearly fragile. US manufacturing capacity is losing ground. Fabrication plants in the United States accounted for 37 percent of global production in 1990 but dropped to just 12 percent by 2021.[4]

Key Developments

Moore's Law, Past and Future

For over half a century, information technology has been driven by improvements in the chip fabrication process. In 1965, Gordon Moore observed that the cost of fabricating a transistor was dropping exponentially with time—an observation that has come to be known as Moore's law. It's not a law of physics but rather a statement about the optimal rate at which economic value can be extracted from improvements in the chip fabrication process. As such, it depends not only on the sophistication and capabilities of the equipment and facilities used to fabricate chips but also on a set of economic conditions and decisions that make it financially sensible to invest in the expensive construction of new state-of-the-art fabrication facilities. Including the research and development needed to develop appropriate manufacturing processes, the cost of such a facility may be as much as $20 billion.

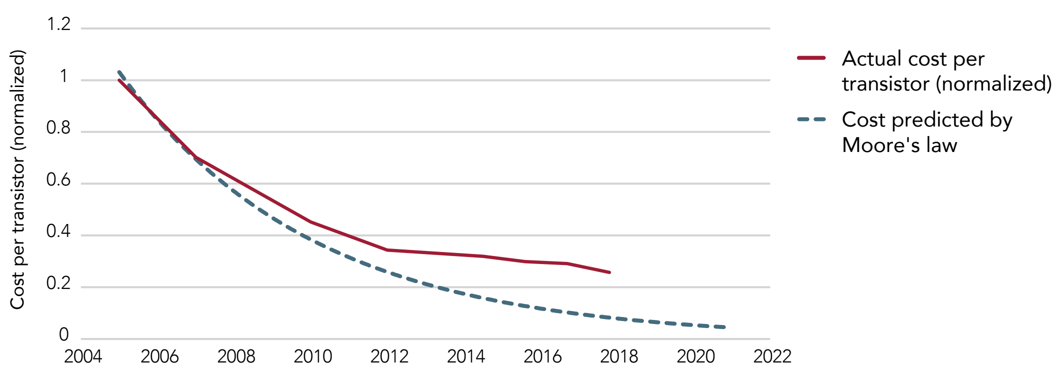
Although Moore's law is often stated as the number of transistors doubling every few years, historically the cost of making a chip was mostly independent of what it holds. This means that every few years, the same-size and approximately same-cost chip will have twice the number of transistors on it.

Moore's law scaling (i.e., the exponential increasing of the number of transistors on a chip) meant that each year one could build last year's devices for less money than before, or a more powerful system for the same cost. This scaling has been so consistent that everyone expects the cost of computing to decrease with time. The expectation is so pervasive that in almost all fields of work, people are working to develop more complex algorithms to achieve better results while relying on Moore's law to rescue them from the consequences of that additional complexity.

Unfortunately, the end of Moore's law cost scaling appears to be in sight, and the end of that technology trend in chip fabrication has profound implications for future systems and design.

As figure 8.1 shows, the actual cost of a chip per transistor (as shown by the solid red line) was tracking the cost predicted by Moore's law (dashed blue line) relatively well from 2004 to 2012.[5] However, the actual

FIGURE 8.1 Cost per transistor over time



Source: Data from Qualcomm Technologies Inc., https://eri-summit.darpa.mil/docs/Mollenkopf_Steve_Qualcomm_Final.pdf cost per transistor started to level off around 2012, and it has not kept up with Moore's law predictions since then.[6] Similar slowdowns in scaling the cost per bit of dynamic random-access memory (DRAM) and the cost per bit of mass disk memory have occurred as well. The cost per bit of flash memory has continued to drop, though not as rapidly as before, and it too may continue to slow.

Quantum Computing

While tremendous effort has been and is being invested to continue scaling semiconductor technology, the benefits gained from this scaling have dramatically decreased. This slowing has increased interest in alternative technologies that might have potential advantages over today's semiconductors. One such technology is quantum computing.

Quantum computing (QC) is a different way of performing computation—this is both its allure and its difficulty. Its different framing allows some computational tasks to take fewer steps than on today's computers, potentially making it much faster. Unfortunately, the different framing also means that the basic operations are different and more noise sensitive, which requires completely different hardware, and solving problems requires a different approach to algorithm design. Therefore, creating a useful quantum computer requires innovation across the entire quantum software/hardware stack: algorithms, compilers, control electronics, error correction, and quantum hardware.

A classical computer uses individual bits as the smallest unit of data, each being 0 or 1, and represents numbers, words, and even pictures as larger collections of bits. Analogously, QC uses quantum bits, or "qubits." Unlike normal bits, quantum mechanics allows a qubit to be in many different states beyond just 0 or 1. It can be both a 0 and a 1 at the same time, for example, allowing a quantum computer to process a large number of possibilities and problems all at once. Much has been written about this "quantum parallelism," but building algorithms that make use of it is difficult.

Qubits need to be isolated quantum systems, and many technologies are under consideration for the physical construction of qubits; these include trapped ion, superconducting, cold atom, photonic, crystal defect, quantum dot, and topological technologies. The most advanced quantum computing machines use either trapped ion or superconducting qubits, but neither technology has a clean scaling path to larger machines—so work continues on other possible technologies.

Current hardware is focused on increasing the number of qubits while decreasing the number of errors. The best quantum machines today have around fifty qubits and can do about three hundred 2-qubit operations between errors. Historically, increasing qubits increases error rates: machines with four hundred qubits exist but can do fewer than fifty operations. For comparison, conventional computers have billions of bits and can do more than a million billion (10^{15}) operations before errors occur.

Given the higher error rate, quantum error correction algorithms have been developed to work using noisy physical qubits. The essential idea is to encode a logical qubit among multiple physical qubits—maybe as few as ten, maybe as many as one thousand—and take advantage of the ensuing redundancy to detect and correct errors in the physical qubits so the logical qubits remain correct. Algorithms with smaller overheads require lower base error rates to operate correctly, and today's machines are approaching the point where error correction can be demonstrated.

For QC to be successful, it will be necessary to scale the number of qubits that can be used in a machine and decrease the error rates they experience. Like all technological development, this scaling will take greater financial investment. Moore's law was the result of a virtuous cycle where better technology increased the semiconductor market revenue, which in turn increased funding available to develop the next technology generation. For QC to flourish, it too will need a virtuous cycle created by a growing market that funds increasingly difficult technology development.

How to create a growing market for quantum computers is one of the biggest challenges in the field. While in theory a quantum computer can solve some problems—such as finding the prime factors of a number much faster than a conventional computer—solving this problem would require a computer that is much more advanced than what we can build today. To be able to refine the technology to that point, we first need to produce smaller useful machines that can generate the funding needed for continued technology development.

As of today, no one has found a commercial problem that a near-term quantum computer can solve that can't also be solved as effectively on a conventional computer. Given the large initial cost of a quantum computer relative to conventional computing, this situation means that we still don't have a commercial market for quantum computing. Factoring is still the holy grail of quantum computing, but the rough consensus in the field is that it will be at least a decade before it is possible to use quantum computing to factor large numbers—a problem that underlies the security offered by much of today's public-key cryptography.

Over the Horizon

Impact of New Technologies

The timeline for quantum computing is uncertain. But even if and when it does arrive and quantum computing is fully successful, QC machines will be useful for only a limited class of applications; they won't replace today's semiconductor technology.

If it is true that Moore's law is at an end, improvements in end-user applications will come primarily from better optimization of algorithm/hardware/technology to the application, rather than technology scaling. Such optimization will require innovation across the entire technology stack, in new materials, new technologies and devices, and design methods. Yet finding these new methods of building systems poses a dilemma for the industry that in some ways is burdened by its own success.

Moore's law has enabled us to produce computing systems of amazing power and complexity, but they also require huge and expensive design teams and must be manufactured in fabs costing billions of dollars to construct. As a result, the industry has consolidated, and the number of companies working in the hardware space has diminished, as has student interest.

The result is a paradox. Performing the necessary optimizations requires innovative researchers willing to try radical ideas that in the end might not succeed. But how can we find researchers and companies willing to take on these risks if every attempt costs $100 million or takes two years?

We must make it easier and cheaper and faster for people to explore those innovations. Many research teams are working on this problem, and it is an active research topic at Stanford. It calls for making the complexity of the design tasks proportional to the change being made, rather than to the complexity of the resulting system. Imagine how little home remodeling would be done, for example, if every idea for remodeling entailed revision of all the blueprints for the entire house, as though everything had to be redone from scratch. The latter more closely corresponds to the process of state-of-the-art chip design today.

The goal is to allow the prototyping of solutions at low cost. The specific tools and approach to accomplish this goal depend on which level in the semiconductor design stack is being addressed. For example, to foster hardware optimization to help with specific software applications, the designers of those software applications need to be able to explore different ways in which hardware improvements would help them to optimize their applications.

Software application developers should be able to test various custom accelerators without requiring a deep knowledge of hardware design. Because their primary concern is the potential benefits a new hardware accelerator could offer their applications, they should be able to direct the hardware design by exploring different structures and configurations in their code. This whole approach relies on the existence of a base hardware platform on which the application currently runs. Stanford hardware researchers interested in this approach are exploring tools and new software and hardware interfaces to the base platform that would allow an application developer to make a small addition to the platform to improve the application's performance. The goal is to build an infrastructure that allows application designers to add small hardware extensions to the base platform. This hardware extension is used only for the specific application of the software developer. In essence, just as a housing developer can build standard houses but also allow buyers to customize a room, this approach would enable software application developers to customize hardware for their applications by making a small addition to a standard base platform rather than starting from scratch.

Of course, this model works only if there are available platforms to use. Since building a platform is very expensive, for this idea to be a success it is critical to convince some firms that have complete working systems—for example, Nvidia, Apple, Intel, AMD, and Qualcomm—to participate in the effort. This approach bears substantial similarity to the model of the app store, which provides an open interface while keeping the base system proprietary. The app store model balances open innovation with the profit motives of companies. Advanced packaging technology makes this approach possible, yet many economic issues still need to be resolved.

Challenges of Innovation and Implementation

Historically, the US government has tended to refrain from funding R&D efforts that it believes are more properly supported by the private sector. These efforts are characterized by having relatively short time scales to benefit the specific competitive strategies that a private company might have. By contrast, it has been more generous in funding academic research with long time horizons on the grounds that such research is precompetitive in nature and enables "all boats to rise with the tide."

There are good theoretical, science-based arguments for why quantum computing is feasible in principle. It is further known that some types of useful problems are indeed solvable by quantum computing algorithms.

But there are two potential flies in the ointment. First, though QC may be feasible in principle, a large amount of engineering stands in the way of making QC feasible in practice. Second, the problem of QC's utility is not a matter of whether it can solve certain useful problems, but whether it can solve them more rapidly than conventional computers. This latter point means that QC's value depends not just on the status of the field of QC but also on the status of its competitors.

In any event, the private sector is unlikely to fund QC research at the level necessary to determine its actual utility because of the long time horizons to any possible payoff. Thus, if QC research is to flourish, sustained US government support will be necessary.

Policy, Legal, and Regulatory Issues

GEOPOLITICS

Taiwan is where over 60 percent of the world's semiconductor chips and over 90 percent of the most advanced chips are fabricated, most by Taiwan Semiconductor Manufacturing Corporation. Both China and the United States, as well as the rest of the world, depend on TSMC for the chips that power their advanced technologies.[7]

The fragility of the supply chain for advanced chips was demonstrated with the outbreak of the COVID-19 pandemic and its associated lockdowns. As of April 2021, more than 169 different industries were impacted by the lack of supply of semiconductors, according to an analysis by Goldman Sachs.[8] Given the stated desire of China to reintegrate Taiwan into the People's Republic of China—and its refusal to disavow the use of force to do so—the stakes for the economy and national security of the United States could not be higher. Few scenarios are more dire than physical PRC control over Taiwan and PRC global dominance in the manufacture of semiconductor chips.

In response to these concerns, the United States passed the CHIPS Act in 2022, which aims to increase domestic semiconductor production and research.[9] The act appropriated $52.7 billion for semiconductor manufacturing, research, and workforce development, along with a 25 percent tax credit for private investments made for capital expenses related to the manufacture of semiconductors—with an estimated value of $24.5 billion.[10]

The act was meant to provide incentives for companies to invest in American fabs. Companies have indeed announced $166 billion in investments in the one year since the law was enacted, though many of those projects are contingent on approval of federal aid.[11]

In addition, the US government in October 2022 imposed tighter export control restrictions on various semiconductor-related products that might otherwise be headed for China.[12] These products include advanced computing chips—such as the Nvidia chips used for AI computing—and equipment that could help to manufacture advanced semiconductor chips. It also reportedly persuaded nations such as Japan and the Netherlands to adopt similar export control measures, thus imposing a near-total blockage of China's ability to buy the equipment necessary to make leading-edge chips.[13] National Security Advisor Jake Sullivan noted that the rule shifted US export policy: "We previously maintained a 'sliding scale' approach that said we need to stay only a couple of generations ahead. That is not the strategic environment we are in today. Given the foundational nature of certain technologies, such as advanced logic and memory chips, we must maintain as large of a lead as possible."[14]

It is worth noting that the ending of Moore's law complicates this logic. Moore's law (i.e., the continued decrease in computational costs) is nearing its end because of limitations imposed by the laws of physics, which apply to all nations. This has several implications. First, pouring more money into chip fabrication won't increase transistor density forever, and Western nations will hit these limits sooner than nations like China because the West is currently ahead. Second, while the West may have previously benefited from restricting China's access to advanced fabrication technology, these controls are less impactful now. If newer technology offers diminishing returns, then older technology will not lag that far behind. Consequently, the Western lead will diminish over time. Export controls that limit China's access to Western technology also incentivize China to develop indigenous expertise and boost its domestic industry by providing a larger market for Chinese firms. Additionally, export controls can hamper US academic semiconductor research, potentially stalling innovation by limiting access for international talents.

Pouring more money into chip fabrication won’t increase transistor density forever.

TALENT

As noted above, regular hardware-based improvement for complex applications is unlikely to be sustained by Moore’s law for much longer, if at all. This reality puts a premium on generating new ideas for hardware improvements that are application specific. One new approach to integrate hardware improvements into software applications was described above, but equally important is the availability of human talent.

However, student interest in hardware design has dropped precipitously, as technically inclined students tend by overwhelming margins to favor software-oriented jobs. Some estimates suggest that by 2030, the semiconductor manufacturing employment sector will be able to fill only 30 percent of its needs.[15] Since appropriately trained people are the only real source of new ideas, these trends do not bode well for the industry.

As one possible data point, it is noteworthy that the CEO of TSMC pointed to “an insufficient amount of skilled workers with the specialized expertise required for equipment installation in a semiconductor grade facility” as an important reason that the construction of a planned fabrication facility in Arizona was significantly behind schedule.[16] To be fair, it is not clear whether he was referring to the high-end chip designers; others have suggested that TSMC’s difficulties stem from clashes of work culture between Taiwan and US unions that oppose bringing in non-union workers from Taiwan to help build the plant.[17]

CLASSIFICATION

In 2014, Edward Snowden leaked a US classified document pointing to the existence of an $80 million classified program in quantum computing, apparently for the purpose of codebreaking.[18] The scope and nature of classified US government programs currently pursuing quantum computing are unknown. Regardless, a robust unclassified program to support QC research will help to ensure that quantum scientists in the United States are able to draw on ideas from around the world—and maintain our leadership in the field.

NOTES

1. Semiconductor Industry Association, *From Microchips to Medical Devices: Semiconductors as an Essential Industry during the COVID-19 Pandemic*, 2020, https://www.semiconductors.org/wp-content/uploads/2020/10/From-Microchips-to-Medical-Devices-SIA-White-Paper.pdf.

2. Arjun Kharpal, "Global Semiconductor Sales Top Half a Trillion Dollars for First Time as Chip Production Gets Boost," CNBC, February 15, 2022, https://www.cnbc.com/2022/02/15/global-chip-sales-in-2021-top-half-a-trillion-dollars-for-first-time.html.

3. HPCwire, "Inside an Intel Chip Fab: One of the Cleanest Conference Rooms on Earth," March 29, 2018, https://www.hpcwire.com/off-the-wire/inside-an-intel-chip-fab-one-of-the-cleanest-conference-rooms-on-earth.

4. Semiconductor Industry Association, *2021 State of the U.S. Semiconductor Industry*, 2021, https://www.semiconductors.org/wp-content/uploads/2021/09/2021-SIA-State-of-the-Industry-Report.pdf.

5. Electronics Resurgence Initiative 2.0, "2019 Summit Agenda, Videos, and Slides," Defense Advanced Research Projects Agency, accessed August 30, 2023, https://eri-summit.darpa.mil/2019-archive-keynote-slides.

6. In fact, transistor costs are usually plotted in a semilog graph, where the log of the cost is plotted against time. In these plots the exponential decline in transistor costs becomes a straight line. The fact that the scale of the *y* axis is linear is a clear indication that Moore's law is over.

7. *The Economist*, "Taiwan's Dominance of the Chip Industry Makes It More Important," March 6, 2023, https://www-economist-com/special-report/2023/03/06taiwans-dominance-of-the-chip-industry-makes-it-more-important.

8. Allison Nathan, "The Global Chip Shortage: Impact, Outlook and Recovery," June 29, 2021, in *Exchanges at Goldman Sachs*, podcast produced by Goldman Sachs, MP3 Audio, 19:40, https://www.goldmansachs.com/intelligence/podcasts/episodes/06-29-2021-hari-yuzawa-hall.html.

9. Semiconductor Industry Association, "CHIPS for America Act and FABS Act," accessed August 30, 2023, https://www.semiconductors.org/chips.

10. Congressional Research Service, *Frequently Asked Questions: CHIPS Act of 2022 Provisions and Implementation*, April 25, 2023, https://crsreports.congress.gov/product/pdf/R/R47523.

11. White House, "Fact Sheet: One Year after the CHIPS and Science Act, Biden-Harris Administration Marks Historic Progress in Bringing Semiconductor Supply Chains Home, Supporting Innovation, and Protecting National Security," Statements and Releases, August 9, 2023, https://www.whitehouse.gov/briefing-room/statements-releases/2023/08/09/fact-sheet-one-year-after-the-chips-and-science-act-biden-harris-administration-marks-historic-progress-in-bringing-semiconductor-supply-chains-home-supporting-innovation-and-protecting-national-s.

12. Bureau of Industry and Security, "Commerce Implements New Export Controls on Advanced Computing and Semiconductor Manufacturing Items to the People's Republic of China (PRC)," US Department of Commerce, October 7, 2022, https://www.bis.doc.gov/index.php/documents/about-bis/newsroom/press-releases/3158-2022-10-07-bis-press-release-advanced-computing-and-semiconductor-manufacturing-controls-final/file.

13. "The United States Announces Export Controls to Restrict China's Ability to Purchase and Manufacture High-End Chips," *American Journal of International Law* 117, no. 1 (January 2023): 144–50, https://doi.org/10.1017/ajil.2022.89; Yuka Hayashi and Vivian Salama, "Japan, Netherlands Agree to Limit Exports of Chip-Making Equipment to China," *Wall Street Journal*, January 28, 2023, https://www.wsj.com/articles/japan-netherlands-agree-to-limit-exports-of-chip-making-equipment-to-china-11674952328.

14. White House, "Remarks by National Security Advisor Jake Sullivan at the Special Competitive Studies Project Global Emerging Technologies Summit," Speeches and Remarks, September 16, 2022, https://www.whitehouse.gov/briefing-room/speeches-remarks/2022/09/16/remarks-by-national-security-advisor-jake-sullivan-at-the-special-competitive-studies-project-global-emerging-technologies-summit.

15. Tom Dillinger, "A Crisis in Engineering Education—Where Are the Microelectronic Engineers?," SemiWiki, July 3, 2022, https://semiwiki.com/events/314964-a-crisis-in-engineering-education-where-are-the-microelectronics-engineers; Dylan Martin, "US Chip Industry Has Another Potential Shortage: Electronics Engineers," *The Register*, July 8, 2022, https://www.theregister.com/2022/07/08/semiconductor_engineer_shortage.

16. Tobias Mann, "TSMC Says Arizona Fab behind Schedule, Blames Chip Geek Shortage," *The Register*, July 20, 2023, https://www.theregister.com/2023/07/20/tsmc_arizona_fab.

17. Jacob Zinkula, "TSMC's Arizona Plant Is Delayed Over Poor Management, Not a Shortage of US Skilled Labor, the Workers Building It Say," Business Insider, August 28, 2023, https://www.businessinsider.com/tsmc-phoenix-arizona-chip-factory-taiwan-semiconductor-management-safety-workers-2023-8; Kevin O'Marah, "TSMC vs. Arizona Pipe Trades 469 Union: Everyone Loses," *Forbes*, August 17, 2023, https://www.forbes.com/sites/kevinomarah/2023/08/17/tsmc-vs-arizona-pipe-trades-469-union-everyone-loses.

18. Steven Rich and Barton Gellman, "NSA Seeks to Build Quantum Computer That Could Crack Most Types of Encryption," *Washington Post*, January 2, 2014, https://www.washingtonpost.com/world/national-security/nsa-seeks-to-build-quantum-computer-that-could-crack-most-types-of-encryption/2014/01/02/8fff297e-7195-11e3-8def-a33011492df2_story.html.