The strange reason the US lost control of chip manufacturing | Chris Miller

21 MIN · YOUTUBE · HTTPS://WWW.YOUTUBE.COM/WATCH?V=XTDOGRTLTCO
https://www.youtube.com/watch?v=XTDOgrTlTco

SUMMARY

The semiconductor industry has evolved into a highly specialized global supply chain, with distinct regions focusing on different aspects of chip design and manufacturing. The historical development of semiconductors, from early innovations at Bell Labs to the rise of companies like Intel and TSMC, highlights the interplay between technological advancement and market demands, as well as the geopolitical implications of chip production.

- *The semiconductor industry began with innovations at Bell Labs, leading to the invention of the transistor and the first integrated circuits.*
- *Companies like Texas Instruments and Fairchild Semiconductor were pivotal in the early growth of the chip industry, focusing on both government and commercial applications.*
- *Intel's shift to microprocessors revolutionized computing, allowing for programmable chips that could serve multiple functions.*
- *The industry has become globalized, with design primarily occurring in the U.S. and manufacturing concentrated in East Asia, particularly Taiwan.*
- *TSMC, founded by Morris Chang, specializes in chip manufacturing, becoming the largest and most advanced chipmaker globally by focusing solely on manufacturing rather than design.*
- *China's chip industry is growing but remains behind leaders like TSMC, particularly in advanced technology, despite significant investments in domestic manufacturing.*
- *India's semiconductor sector is rapidly developing, with potential to become a major player in the coming decades as it follows the paths of earlier successful nations in chip production.*
- *The complexity and high costs associated with semiconductor production necessitate a collaborative global supply chain, where no single region can independently produce cutting-edge chips.*

- The chip industry was a global industry
from really the earliest days. The processes that are used to both design and manufacturer semiconductors today were first pioneered by researchers working at Bell Labs. But because Bell Labs wasn't a computer company, they were able to take those technologies and either spin out their own startup or sell it to somebody else. And that's how many of the key technological advances undergirding semiconductors first emerged. But one of the things that's changed a lot over the past couple of decades is that, today, each region focuses on a different part of the semiconductor supply chain. I'm Chris Miller, a professor at The Fletcher School and author of, "Chip War: The Fight for the World's Most Critical Technology."

- [Announcer] The history of

the semiconductor revolution - In the middle of the 20th century, all telephones were managed by AT&T. They were a monopoly, and the government regulated them, and one of the rules was that their research lab had to share its inventions with the rest of the world. And they had some of the most brilliant physicists and chemists working in the world at that time, which they hired to improve the phone system. But in the process, they created some of the key inventions that drove technological progress in computing for decades to come. The transistor was one of the inventions that emerged out of Bell Labs, but actually many of the processes that are used to both design and manufacturer semiconductors today were first pioneered by researchers working at Bell Labs. But because Bell Labs wasn't a computer company, they were able to take those technologies and either spin out their own startup or sell it to somebody else. And that's how many of the key technological advances undergirding semiconductors first emerged. So William Shockley, John Bardeen, and Walter Brattain invented the first transistor while they were working at Bell Labs. They were initially planning to use these transistors as part of the telephone network. But in the late 1950s, the first engineers realized that you could take multiple transistors, and make them on a single piece of semiconductor material. And so that was the first chip, a piece of material with multiple transistors carved into it. And that was important, because if you had individual transistors, they were connected via wires in a way, that was okay if you had a handful of transistors. But if you had 1,000 connected together, you had a jungle of wires you had to manage. But the chip managed to have the electrical connection in a piece of material.

And so the jungle of connections was replaced by a single block of material, which was much more reliable, and also much more easy to shrink in its size. And so it was the invention of the chip that made it possible to deploy lots and lots of transistors together in a way that was economical, but also possible to engineer and avoided all of the wiring. The first chips were invented by engineers working at Texas Instruments and a company called Fairchild Semiconductor in Silicon Valley. They were invented simultaneously. Jack Kilby invented one in 1958 working in a Texas Instruments lab. And for a long time they were really at the cutting edge of chip manufacturing.

At first, they were building chips primarily for the U.S. government, for the space program, for example, and for weapon systems. But they realized early on you could take the exact same chips that the government wanted to guide spacecraft, and use them for commercial applications, like computers or pocket calculators. And that set the industry off into its first phase of growth in the 1960s and '70s and '80s. For the past 15 years, they've taken a different tack. They don't today produce chips that are used in computing, they're not, for example, in AI systems in a large way. Instead, they produce a lot of chips that are in industrial applications, or in automobile uses. And so Texas Instruments chips are all around you, but you don't see them because they're buried deep in your devices, making sure your windshield wipers work, for example, on your car, or that your windows move up and down when you press the button. Those are the types of use cases that Texas Instruments produces chips for. One of the first startups in Silicon Valley was created by one of the researchers who invented the first transistor, William Shockley, who was by all accounts, a brilliant physicist, but a horrible manager and a horrible person. And so he hired a very talented set of engineers in Silicon Valley. He moved to Palo Alto, California, where his mother lived, for the purpose. And although he hired lots of great people, they detested working for him. And so eight of them in the late 1950s went out on their own, and created Fairchild Semiconductor, which became one of the key startups that would give rise to Silicon Valley, and played a major role in Silicon Valley even being named Silicon Valley, because for a long time it was the absolute epicenter of chip design and manufacturing thanks to people at Fairchild Semiconductor. Robert Noyce, one of the two inventors of the integrated circuit, Gordon Moore, who later would go co-found Intel, and many others first started their career working at Fairchild. Intel was founded in 1969, and it initially planned to focus on making memory chips. But they realized early on that there was a potentially larger market for a type of chip that wouldn't just remember data, but would also process it, especially if that processing could be programmed in different ways for different use cases. And it quickly focused on making chips for personal computers, which at the time was a very small market, but they correctly bet that soon, everyone, would have a personal computer. And Intel, even today, is the world's largest producer of chips that go inside of PCs. Gordon Moore is one of the two co-founders of Intel.

He's most famous today probably for coining the
term Moore's Law, but he also played an absolutely critical role running Intel's R&D operations from the earliest days for many years. And when it came to the microprocessor, he was an early advocate of focusing on microprocessors at the expense of the more memory-focused chips that Intel had previously made. And so in some ways, he was the key figure in Intel in making the company focus on microprocessors. A tiny computer on a chip, as they originally called it. And it gave rise to the idea that you could deploy chips in lots of different use cases without having to redesign the chip itself, because the chips themselves could have a program running on top of them. Today, we take it for granted that you can have a chip in your phone, and a chip in your dishwasher, and a chip in your car. But at the time, that would've required many different chips for each of those purposes. Whereas, today, thanks to the microprocessor, we have programmable chips. And that was the main source of revenue for the chip industry, the main focus of technology, until about 20 years ago when the first smartphones began being produced. And today, smartphone chips are generally designed by one set of companies, but they're manufactured largely in Taiwan. So the largest designers of smartphone chips are Apple, which designs its own chips in California. Qualcomm, and other companies, almost all of them manufacture all of the chips that they design in Taiwan. And so today, the chip industry is split into two different parts. There's the chip designers, which, today, is essentially like a type of programming almost, programming where each of the transistors goes on the chip, and the actual manufacturing takes place generally in Taiwan or elsewhere in East Asia, where different companies specialize in manufacturing at precision scale. The chip industry was a global industry from really the earliest days. Fairchild Semiconductor was founded in Silicon Valley before it was even called Silicon Valley, but they opened their first facility in Hong Kong just a couple years later. So there was already a globalized nature to the chip industry from day one. But one of the things that's changed a lot over the

past couple of decades is that, today, each region focuses on a different part of the semiconductor supply chain. The first chips that were invented in the late '50s and early '60s were used for space programs and missile systems. So they were at the center of the Cold War competition. And the U.S. was ahead, but the Soviet Union realized that they also needed chips to guide their missiles more accurately or to help their spacecraft launch effectively. And so they were focused on building their own chip industry, but also on copying whatever they could from the West. And so since the earliest days of the Cold War, there were Soviet exchange students in physics, for example, studying at Stanford University, but also transmitting the knowledge that they gained back to the Soviet Defense industrial complex. And so there was a lot of copying, a lot of efforts to replicate what the U.S. was doing. But the Soviets made a couple of key errors. One was that they focused too much on copying, and not enough on innovating. And so they got very good at copying, but not so good at innovating, and that left them behind. And the second error they made was that they only focused on the military aspects. And the military was where the first chips were used, but today, most chips go to the private sector.

99% of chips that are made go into phones, or PCs, or data centers, not for defense equipment. And so if you only focus on the government and military uses, you've got a tiny market relative to the vast consumer market that was out there. U.S. firms were profit-seeking, they focused on the consumer market as early as they could. In the Soviet Union, they never made that shift, and so their chip industry was always tiny in comparison to the U.S., which meant they could invest less, they could hire fewer workers, and ultimately their technology fell behind even though they were pretty good at copying. So in the U.S. right now, most of the key chip firms only design chips.

Most of the manufacturing of chips happens in East Asia, in Taiwan, for example, or in Korea. Many of the chemicals that go into chipmaking come from Japan. And the machines that are used to make chips come from either Silicon Valley, where some of them are still made, or the Netherlands or Japan. So the industry has globalized, but it's also specialized in the process. And so there's not a single region today that can make cutting edge chips on its own. Everyone relies on this internationalized supply chain that brings together the U.S., Taiwan, Europe, Japan, and Korea. Japan was a major player in electronics assembly early in the 1950s and 1960s, so devices would be assembled in Japan because labor costs at the time were lower. But Japanese firms were fixated on moving up the value chain, producing more complex, more expensive types of goods. And Japanese firms realized very early on that consumer electronics could be a major growth area for them, where they could sell not just domestically, but all around the world. And so companies like Sony, which were among the leaders in the 1970s and 1980s, bet on the consumer market to produce the types of goods that would take advantage of the advanced chip technology that they were pursuing at the time. And so although we don't remember it much today, devices like the Sony Walkman in the 1980s was at the center of the tech industry, and it put Japan really on the map.

And at that point, Japan was, by a lot of metrics, just as capable as the United States when it came to building advanced chips and then deploying them in very profitable uses like the Sony Walkman. One of the places where the Japanese excelled was in video games, which most people might not think of as driving technological advances, but actually, the computing that's required to show graphics that look real life is extraordinarily complex. And so the Japanese companies like Sony, Nintendo is another one, were fixated on how to make better graphics, and it required more and more computing power to make better and better graphics. And today, they're no longer major players in that sphere, but NVIDIA, which is the central player in AI, actually started as a video game company, it made graphics cards for computers. And for most of their early history, they were selling chips primarily to gamers, because the graphics were better and rendered more rapidly. But it turns out that the same essential math that's used for showing graphics on a screen is pretty similar to the math that's used in training AI systems. And so NVIDIA was able to take chips that were made for video games, and made for computer games, and pivot them to be used in AI systems, which is why a video game company that was founded in the 1990s has now become not just any AI company, but the most important AI company in the world. In the 1980s, the South Koreans saw Japan becoming a major player in the chip industry and saw Japanese firms rise to the top, both in terms of technology and in terms of the amount of money they were making, selling both chips and devices that used them, and South Korea wanted to replicate Japan's strategy. So companies like Samsung and SK Hynix were founded to establish chip industries in Korea. And they replicated the Japanese model, they get very good at manufacturing, they competed very effectively on cost. They also represented an alternative to Japanese production. 'Cause U.S. firms in the 1980s were very worried that Japan was gonna take over the chip industry. So they were excited to have another option besides Japan, and shifted business towards Koreans, both because the Korean producers were cost competitive, but also because it provided a bit more diversification in the industry that would limit the ability of Japanese firms to dominate. One of the biggest European chip makers in the 1960s, '70s, and '80s was the Dutch company Phillips, which today still exists, but doesn't produce any semiconductors. They got out of the semiconductor business several decades ago. But one of the legacy units that they'd created was a unit that made the tools that make chips. And in particular, they focused on the lithography tools that are capable of patterning transistors on a chip. ASML was spun out of Phillips several decades ago, and at the time, most people thought it would likely fail, the Netherlands wasn't a big part of the chip industry, Silicon Valley was a long way away. But ASML took a series of pretty wild technological bets on technologies most people thought would fail. And the best example of this is the current cutting edge of lithography called extreme ultraviolet lithography, the tools that cost $350 million a piece to produce, everyone else thought that was a technology that would never work. It took three decades to commercialize, tens of billions of dollars of research and development money went into it, but ASML made that bet, and it was a bet that looked like a very bad bet for many years until about a decade ago when they first were able to build the initial EUV lithography machines.

So chip makers have always used lithography to manufacture semiconductors, but as transistors have gotten smaller and smaller, we've needed better and better lithography systems to print smaller transistors onto silicon chips. And several decades ago, it was clear that the cutting edge in lithography at the time was gonna be too broad in terms of the wavelength of light used to print tiny transistors. The cutting edge used light with a wavelength of 193 nanometers, which sounds really small, and it is really small. But if your transistors are measured in 10 nanometers, or 5 nanometers, 193 nanometers is still too broad of a brush with which to paint your transistors on the silicon chip. And so ASML bet on a new type of lithography system using light with a wavelength of 13.5 nanometers, much more narrow. Which sounds logical, but it was extraordinarily difficult to produce. Research started in the early 1990s, and it took 25 years before these machines were commercialized, because it required building a supply chain that involved these extraordinarily complex components, the flattest mirrors humans have ever made, the most powerful laser ever in a commercial device, all of these had to be invented in the process of making these machines work. So Taiwan was a major player in electronics assembly, and putting together transistor radios, for example, in the 1950s and '60s, or assembling televisions. And they did quite well on that, but there's not much money to be made in the assembly, the money is made in the manufacturing of the complex components involved. And so the Taiwanese government realized, as early as the 1970s, that they needed to move up the value chain and learn to do the more complex parts of electronics manufacturing. In 1987, there was a American engineer named Morris Chang who was passed over for the CEO job of Texas Instruments where he'd worked for several decades. And so he left TI, and was looking for something else to do, and he'd gotten to know the Taiwanese government for several years, because Texas Instruments, his former employer, operated a number of plants in Taiwan. And so the Taiwanese approached him and said, "Would you like to build a chip factory in Taiwan?" And he said yes. And he had an idea, which was to do manufacturing differently than anyone else. At the time, most chips were manufactured and designed by the same companies, but Morris Chang realized that manufacturing is getting more and more complex every single year, that if you specialized on manufacturing, you could manufacture better than your competitors.

And so he established TSMC in Taiwan in
1987 with the aim never of designing chips, only of manufacturing. His vision was sort of like to do for chips what Gutenberg had done for books. Gutenberg didn't write any books, he only printed them. Morris Chang didn't wanna design any chips, he only wanted to manufacture them. That's exactly what TSMC has done. And it's enabled TSMC to win among its customers, some of the largest companies in the world, Apple, NVIDIA, Qualcomm, AMD, they all rely on TSMC to produce its chips, which means that TSMC is the largest chipmaker in the world by far. And as a result, it's got more scale, it can drive down its costs, and it can hone its technology more than anyone else. And so TSMC, thanks to this unique business model, is both the largest and the most advanced chipmaker in the world. Today, China's the world's largest importer of chips. They spend as much money each year importing chips as they spend importing oil. There's nothing that China's more reliant on the outside world to purchase.

And China imports all these chips, both for its
own use, but also because most of the world's phones and computers and servers are assembled in China. So there's a flow of chips into China, they're assembled in the devices, and the many

of those devices are re-exported to the U.S., or to Europe, or to Japan, or to international markets. And so today, China's primary interface with the chip industry is by buying chips, assembling them, and then shipping them abroad.

But the Chinese government realizes this is not
the best place in the industry to be. They wanna do the higher value add parts of the industry, just like Taiwan did, just like Japan did to move up the value chain. And so for the past decade, China's been trying to build its own chip industry to manufacture more chips domestically. And right now it's having a lot of success when it comes to more low-end chips, the types of commodity chips that are in many different types of devices, where China is vastly expanding its manufacturing capacity and making real strides towards becoming a lot more self-sufficient. But at the cutting edge, the types of chips that are inside phones or in AI systems, China's still meaningfully behind industry leaders like TSMC. Right now, the most advanced Chinese firm, SMIC, is about five years behind TSMC, which might not sound like a lot, but that's two and a half Moore's Laws behind TSMC, which means that, for the most cutting edge applications, you really take a performance hit if you want to use a Chinese manufacturer versus a Taiwanese one. Until 2020, TSMC'S two largest customers were first, Apple, the biggest U.S. smartphone maker, and second Huawei, China's largest phone company. TSMC manufactured ships for both of their phones. But the United States is worried that Huawei is controlled by the Chinese government, it's worried about the surveillance capabilities that this might enable, and so the U.S. has been trying to limit Huawei's access to advanced technologies. And since 2020, it's prohibited Huawei from manufacturing advanced chips at TSMC. And so Huawei's had to turn to domestic suppliers to manufacture many of the chips that it needs. And this has been a challenge, it's possible to find Chinese domestic suppliers, but they're not as good as TSMC, the costs are higher, the performance is lower. And it's been a real headwind for Huawei over the past couple of years as they've tried to build their own supply chain to make up for the fact that they've lost access to the cutting edge in Taiwan. So until recently, India was a very small player in the chip industry. There's a couple of chip companies in India, but they're not at the cutting edge, and they're not that large. Much of the semiconductor manufacturing, as well as the rest of the supply chain, the assembly of phones, for example, of computers takes place in Southern India. Tamil Nadu, for example, is one of the key hubs for manufacturing.

And then Bangalore is a major center for
chip design inside of India. But right now, India is the country that's changing the most rapidly, I think, when it comes to investment in semiconductors. There's a series of new projects underway in India to put it more on the map of electronics manufacturing. And I think if you look at India today, you see what China looked like 30 years ago, or what Taiwan looked like 50 years ago, a country that's on the early stages of a major change in the types of manufacturing that happened there. And so I wouldn't be surprised at all if in 10 or 20 years we looked at India as a really central player in the production of all the computing and electronics that we rely on, because they're taking the exact same steps that China, and Taiwan, and Japan before them took when they were becoming major manufacturers. The irony of the chip industry is that it's simultaneously

globalized, and yet extraordinarily localized for certain types of production. And that's inevitable I think, because the engineering involved is so complicated, the dollar values required to spend are so vast that we need specialization. And specialization implies that we've got to rely on other people to help in the process. And so I think it's inevitable that U.S. firms will rely on manufacturing in Taiwan, and chemicals from Japan, and the rest of the supply chain for a very long time because no one has the capabilities they need to produce the chips that they require on their own.