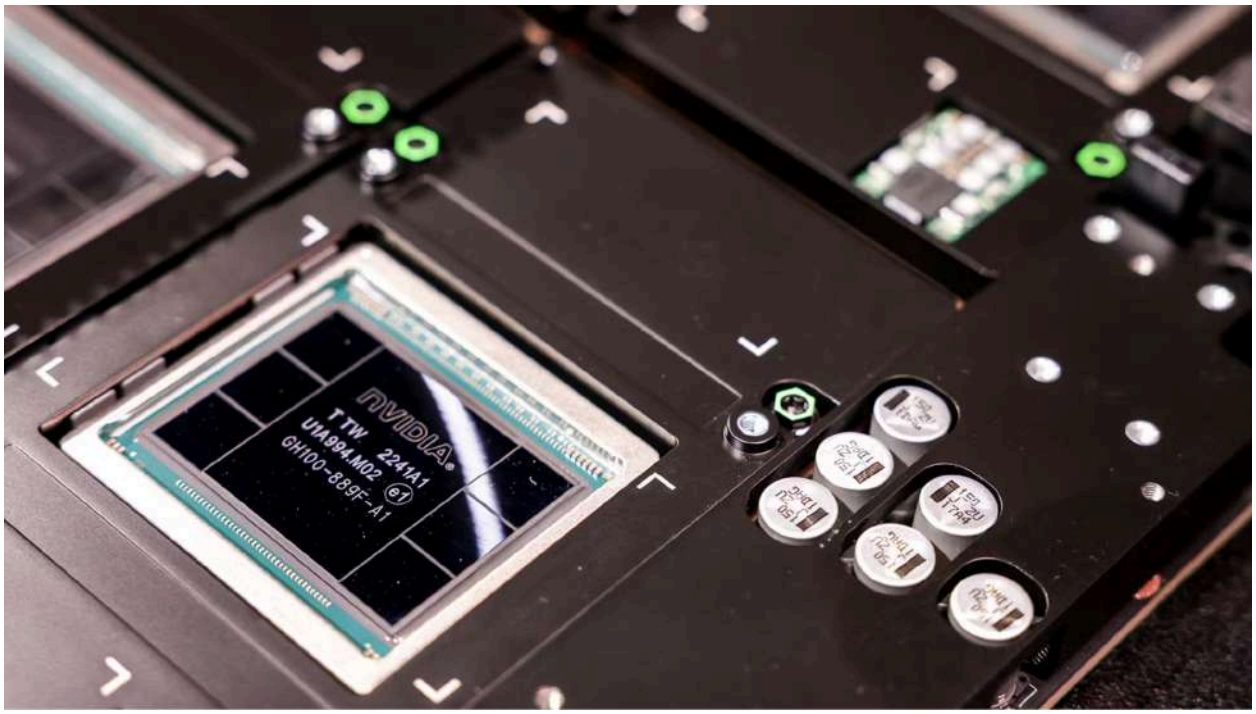
The Information

Nvidia AI Chip Smuggling to China Becomes an Industry



A Nvidia chip. Photo by Getty

By Qianer Liu

Aug 12, 2024

Several months ago, an electric appliance company in eastern China put in a $120 million order for 300 servers powered by eight of Nvidia's cutting-edge H100 chips. The order was for chips that U.S. export rules bar from sale in China.

To get around those rules, the company didn't go to one of Nvidia's authorized distributors but to a chip broker in Malaysia. The broker arranged for the Chinese buyer to establish a shell company in Malaysia, concealing any link to the parent company in China.

The Takeaway

A large-scale industry has evolved to get Nvidia's most cutting-edge chips to Chinese buyers, via a network of smugglers using shell companies and phony data centers.

He also helped set up a corresponding corporate website and corporate email addresses to enhance the fake company's legitimacy. The broker even rented space in a Malaysian data center to temporarily house the servers when they arrived, as a way of fooling Nvidia staffers who wanted to check if such a larger order of servers was installed properly.

In a matter of weeks, the servers were in China, having first passed through Malaysia, according to the broker, who gave his first name as William and who didn't want to be identified by his full name.

The episode demonstrates how smuggling of Nvidia chips into China has become an organized, multinational enterprise this year—partly made possible by the complicated distribution network Nvidia uses to sell its chips—since the U.S. Commerce Department expanded the number of artificial intelligence chips that can't be sold to companies in China.

While smuggling of Nvidia chips into China has been documented in media reports over the past year by Reuters, The Wall Street Journal and other outlets, those reports have portrayed smuggling as a small-scale operation—a couple of units packed in travelers' suitcases moving across Chinese borders.

But in recent months, both the scale and the sophistication of smuggling have taken off. Scores of mom-and-pop firms in Asia have concocted elaborate schemes that involve setting up shell companies and disguising chips in truckloads filled with manufacturing machinery or other goods, and they in turn oftentimes rake in north of $100 million from a single order. Their clientele has also expanded from small, obscure companies to businesses with ties to

well-known, publicly listed tech companies, according to procurement documents viewed by The Information.

This account of how Nvidia chips have been smuggled into China is based on interviews with eight chip smugglers, several employees at Nvidia, their suppliers and authorized distributors, people in the U.S. Commerce Department, export policy researchers, and procurement documents viewed by The Information.

In addition to buying through this gray market, Chinese companies have also managed to access some of Nvidia's most advanced chips via renting servers at data centers outside China from Google, Microsoft and other global firms, as cloud computing services are not subject to U.S. export rules, The Information reported previously. Together, these workarounds underscore the gaps in Washington's attempts to contain Beijing's AI development and the challenges for U.S. regulators to maintain airtight export controls in a globalized economy.

It is unclear whether Nvidia and its CEO, Jensen Huang, are aware of how advanced chip smuggling into China has become. In several private meetings with employees and U.S. policymakers last year, Huang said he had no idea what was going on in China when asked about the company's sales in the country through both authorized and smuggled channels, according to several Nvidia employees and people close to the Commerce Department who were present.

Nvidia doesn't like export control regulations because it hurts business, said Gregory Allen, director of the Wadhwani Center for AI and Advanced Technologies. "Every smuggling sale is a better alternative than no sale from their financial perspective as long as it's not Nvidia that engages in the smuggling," he said.

Nvidia disputed the claim and said in a statement that "any unauthorized deviation of previously owned products, including any gray market resales, would be a burden on our business, not a benefit."

"We have products designed and manufactured for different regions, and we run our supply, sales, and support operations accordingly. We insist that our customers and partners strictly adhere to all export control restrictions," the statement said.

Complex Web

Nvidia generally doesn't sell its products directly to customers. Instead it authorizes server companies such as Dell Technologies and Super Micro Computer (aka Supermicro) as well as specialized dealers to sell them. Big distributors then farm out products to smaller ones, which in turn pass chips to even smaller ones, eventually giving rise to an extensive web of wholesalers and resellers of which even Nvidia can't fully keep track.

While selling through distributors has been common in the semiconductor industry for decades, in the case of Nvidia it has created a loophole vulnerable to chip smugglers.

Normally, authorized distributors are responsible for verifying a potential buyer's creditworthiness and performing other procurement-related compliance checks on behalf of Nvidia, including ensuring that advanced AI chips do not end up in the countries banned by U.S. export control.

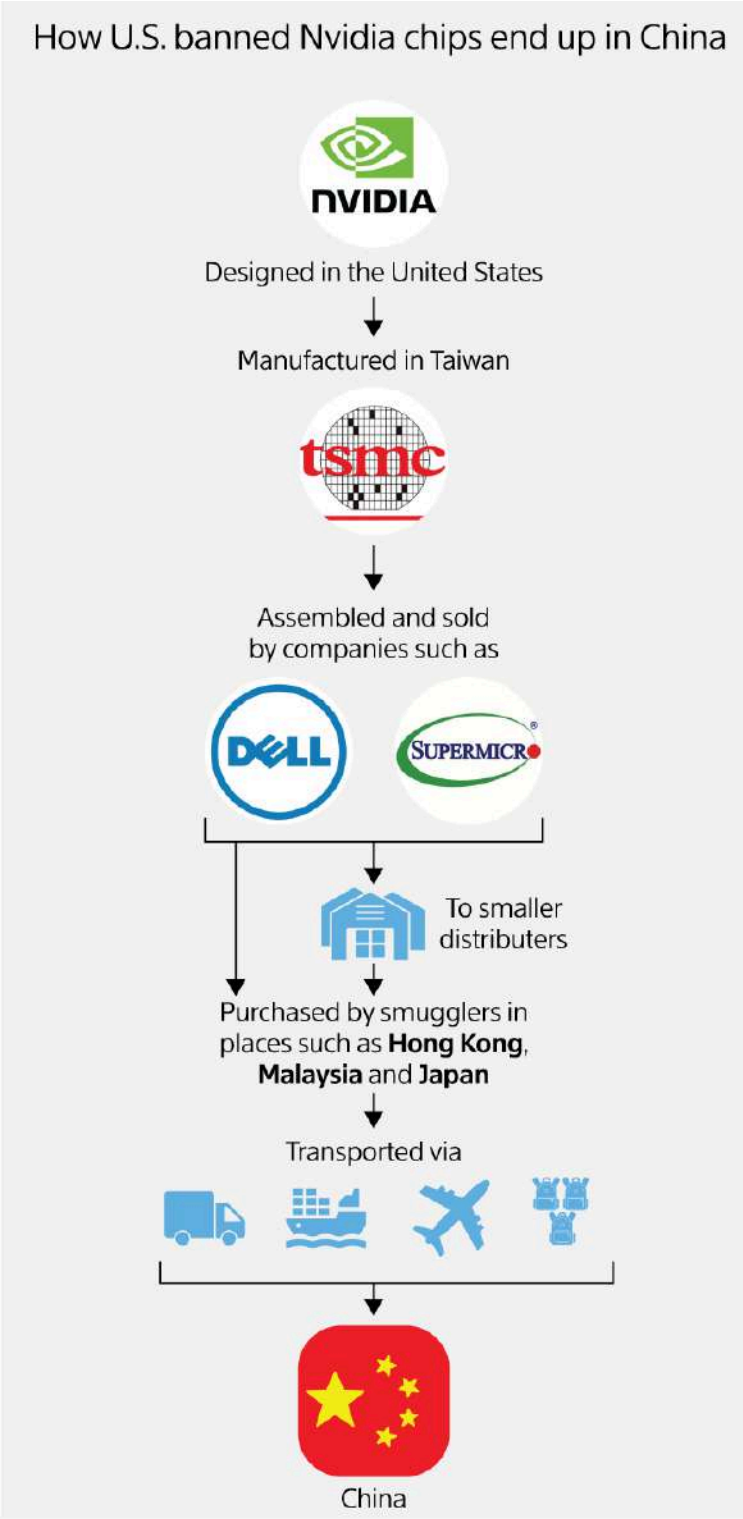
However, some distributors have been lax in reviewing buyers' information, allowing Chinese chip brokers to purchase restricted chips through shell companies that mask the true buyer's Chinese identity, according to eight chip brokers.

The brokers all said that by using these tactics to disguise their clients, they had been able to successfully buy Nvidia chips during the past two years.

When the broker who calls himself William was arranging the $120 million order for the Chinese electrical appliance company, the order was so big that it triggered a mandatory on-site inspection by personnel from Nvidia, who wanted to ensure that the costly equipment was properly installed.

In response, William says, he coordinated the rental, installation and activation of the servers in a spare data center facility in Johor Bahru, a Malaysian town adjacent to the Singapore border and home to large clusters of data centers. Nvidia inspectors checked the servers there and left. Shortly afterward, the servers were whisked away to China via Hong Kong.

In some cases, brokers buy directly from Dell and Supermicro, but to get the amount they want, brokers say they often have to turn to smaller distributors.



One such chip broker is a Hong Kong–based businessman only willing to be identified by his first name, Mike. He said he was able to get his hands on thousands of restricted Nvidia chips from server makers and distributors including Dell and Supermicro, thanks to what he called "strong personal relationships" with sales representatives at these firms.

When The Information visited in May, Mike's warehouse in Hong Kong held 600 servers from Dell and Supermicro, each containing eight Nvidia H100 chips, which are popular with Microsoft and Meta Platforms and highly coveted by their Chinese peers as well.

Mike utilized different shell companies registered in countries not restricted by U.S. regulations to purchase Nvidia's advanced AI chips. His personal relationships with Nvidia's distributors have so far guaranteed that his purchase orders have been filled.

Those orders were delivered to the countries where the shell companies were set up, and from there, Mike and his team shipped them into a warehouse in northern Hong Kong, a stone's throw from the Chinese border.

The 4,800 Nvidia H100 chips were valued at approximately $180 million, and Mike sold them to Chinese buyers for about $230 million, pocketing a handsome return.

Mike is careful not to leave any traces behind. At his warehouse, the servers are taken out of the Dell- or Supermicro-branded cartons and put in ones that bear other logos. On customs forms, he labels the contents as something else, depending on what cargo he can mix with the chips on any given day, Mike said.

A spokesperson for Dell said the company "complies with global regulations, including U.S. export controls on advanced computing products. We maintain a strict trade compliance program, and as part of our robust qualification process, we require our distributors and resellers to follow all applicable global regulations and export controls."

The company added, "When we become aware of partners who are not adhering to these obligations, we take appropriate action, including up to termination of our relationship."

A spokesperson for Supermicro said in a statement: “Supermicro provides rack-scale total solutions and is not a distributor of individual components. We follow all US export control requirements on the sale, service, support, and export of GPU systems to regions and parties that require licenses under the Export Administration Regulations. If Supermicro becomes aware that a third party has exported or reexported without the required licenses, it investigates the matter and takes appropriate action.”

Long Delays

Initially, many Chinese internet and AI companies refrained from purchasing smuggled chips for fear of jeopardizing their relationships with Nvidia.

But late last year, when the U.S. government widened the list of Nvidia chips subject to the ban, Nvidia canceled numerous orders from Chinese companies. Weeks later, it released three new, less powerful chips, the H20, L20 and L2, that the U.S. government allowed it to sell to China. However, Nvidia’s customers had to place new orders and wait for several months to get the new chips.

Instead of waiting, some tech companies resorted to buying the banned chips, according to three chip brokers involved in the deals and the procurement documents. Some of the purchases were signed in the name of employees working at the largest publicly listed tech companies in China, as a way to avoid drawing any direct connection to the actual buyers, the documents showed.

State-owned enterprises and local governments are among the major buyers, too.

Some of William’s customers are Chinese state-owned companies, which have their own corporate governance and procurement compliance guidelines that forbid purchasing from illicit sellers. But those buyers had little choice when they were struggling to purchase Nvidia AI chips from authorized channels.

To get around those internal rules, those transactions happen through intermediaries that are government-approved suppliers so the Chinese companies can maintain deniability. “Sometimes I didn’t even get to talk to the actual customers even after millions of dollars worth of orders were completed,” William said.

Another chip broker who caters to government buyers, going by the first name of Andy, is based in Shanghai. He conducts business meetings in opulent teahouses. Andy has become a familiar face to government officials, as he's contributed to establishing numerous advanced AI computing facilities for state-owned enterprises and government departments, said the chip broker who shared some of his business records with The Information.

Unlike tech companies, government-backed clients purchase smuggled chips for various purposes, such as renting the chips to AI companies and building data centers for regional research labs. Andy said that in addition to AI development, the most important use is building sufficient data centers to reduce the computing power gap with the U.S.

Closing the Loophole

The swelling underground market of smuggled Nvidia chips calls into question the effectiveness of U.S. export control policy and has prompted discussion on what Nvidia can do to prevent it.

The Commerce Department has been trying to stop the illegal export of Nvidia chips into China. The department has placed certain Chinese companies involved in this activity on the Entity List, prohibiting U.S. firms from engaging in business with them.

Commerce officials told The Information they were following up on possible violations of the export controls, including letting their staff stationed overseas do end-customer checks to identify potential vulnerabilities in distribution chains. They added that the department also has the regulatory authority to issue nonpublic directives to companies regarding specific products, destinations, and customers and then place restrictions on those categories.

The department plans to impose additional regulations aimed at curbing chip smuggling in its annual update of export control measures in October. These will include efforts to scrutinize the trade routes used for smuggling chips into China, according to two people involved in the discussion.

Some policy researchers suggested that the Commerce Department should require chip companies including Nvidia to take more responsibility for preventing smuggling.

“I would not expect Nvidia to do anything that is not explicitly specified in the guidance,” said Allen of the Wadhwani Center for AI and Advanced Technologies.

Another proposal put to the Commerce Department by two policy advisers at outside think tanks is that Nvidia could track the movement of the banned chips and remotely disable them. Nvidia can theoretically shut down servers remotely when customers download software updates from the company, but it hasn’t been required to do so.

Nvidia didn’t have a comment on these proposed measures..

Despite threats from the U.S. to further tighten export controls, William, the Malaysia-based chip broker, remains upbeat. “We hope to do more deals before the U.S. updates the export controls. Every dollar of profit counts.”

William is currently assisting clients in establishing shell companies in Australia, another country that’s not subject to U.S. chip export regulations.

Digital Version:
https://www.theinformation.com/articles/nvidia-ai-chip-smuggling-to-china-becomes-an-industry

The Information

U.S. Probes TSMC's Dealings With Huawei

Photo by Getty

By Qianer Liu

Oct 17, 2024

The U.S. Commerce Department is investigating whether Taiwan Semiconductor Manufacturing Co. has been making artificial intelligence or smartphone chips for Chinese tech giant Huawei Technologies in what would be a breach of U.S. export rules, said two people with direct knowledge of the matter.

Since 2020, the U.S. has banned Huawei from buying chips manufactured using American equipment, citing national security concerns. The U.S. restrictions also block Huawei from using

U.S. technology to make its own chips without approval from the Commerce Department. In recent weeks the department has contacted TSMC to ask whether it was involved in making either smartphone chips or AI chips for Huawei, the people said.

The Takeaway

- The U.S. is investigating whether TSMC produced advanced chips for Huawei
- U.S. sanctions block Huawei from buying U.S. chips or chip technology
- Last year the U.S. fined Seagate $300 million for selling tech to Huawei

One possibility the department is looking into is whether Huawei bought chips from TSMC indirectly using intermediary companies with a different name to place orders on its behalf, the two people said. The department is scrutinizing whether TSMC conducted the required know-your-customer due diligence when accepting orders.

The inquiry could be politically sensitive. TSMC is the most important chipmaker in the world, manufacturing chips designed by companies such as Apple and Nvidia, among many others. The Biden administration has given TSMC $6.6 billion to help finance the company's construction of chipmaking facilities in Arizona as part of the administration's push to revive a domestic chip industry.

If the Commerce Department found that TSMC breached export rules in its dealings with Huawei, it could impose a fine or more severe penalties, such as temporarily restricting TMSC's ability to access U.S. tech. Last year, for example, the department fined Seagate Technology $300 million for selling hard disk drive technology to Huawei.

At the time of the Seagate fine, a Commerce Department official, John Sonderman, said violations of the export rules "will be investigated and charged as appropriate."

The TSMC investigation is still in its early stage, the people said. It remains unclear how long it will take the Commerce Department to gather information and reach a conclusion. A spokesperson for the department declined to comment. Huawei did not have a comment.

TSMC said the company is “committed to complying with all applicable rules and regulations, including applicable export controls” and maintains an “export system for monitoring and ensuring compliance.” The company said it would take action to ensure compliance if there are potential issues, including conducting investigations and proactively communicating with relevant parties as necessary.

In 2020, after the U.S. imposed sanctions on Huawei, TSMC said it stopped taking new orders from Huawei.

AI and Smartphones

The current inquiry focuses partly on whether TSMC has been involved in manufacturing AI chips designed by Huawei. Its AI server chips have emerged as a popular alternative to Nvidia-made chips for Chinese customers, who are prevented by U.S. export rules from buying Nvidia chips.

The department is also looking into whether TSMC made smartphone chips for Huawei phones. Huawei is a major manufacturer of smartphones, and before the U.S. sanctions were implemented it was a big customer of TSMC.

Huawei’s global market share took a hit after the U.S. imposed the sanctions. But the Chinese company showed some recovery last year when it released its flagship smartphone series, Mate 60, powered by a processor chip designed and manufactured in China by Semiconductor Manufacturing International Corp., a Chinese company.

The Commerce Department is looking into whether TSMC also made some of the smartphone chips used in the phone, according to the two people with direct knowledge of the matter.

Chip manufacturers like TSMC rely heavily on U.S. equipment for chip fabrication. As a result, U.S. export rules require them to monitor whether their customers are circumventing regulations: for instance, by producing chips that exceed specified power limits or by reselling chips to entities that are on the trade blacklist, like Huawei.

Huawei has hit production hurdles in making chips for the new flagship smartphone Mate 70 series due to the continued U.S. government crackdown, which is blocking the company and its chipmaker from procuring equipment and components, The Information reported.

Digital Version:
https://www.theinformation.com/articles/u-s-probes-tsmcs-dealings-with-huawei

The Information

TSMC's Push to Be Tech's Switzerland in Doubt as U.S.-China Tensions Grow

Taiwan's chipmaking powerhouse is facing increasing U.S. pressure to cut off Chinese customers, while also walking a careful line to avoid offending Beijing.

Photo by Getty

By Qianer Liu

Nov 26, 2024

If Taiwan Semiconductor Manufacturing Co. had its way, the chip manufacturing powerhouse would be a neutral party in the battle between the U.S. and China for technological supremacy—a semiconductor version of Switzerland, one TSMC executive told The Information. Unfortunately, the company may find it increasingly tough to avoid picking a side.

Last month, the U.S. Commerce Department launched an investigation into whether TSMC was still producing chips for Chinese tech giant Huawei Technologies in a potential violation of U.S. sanctions that have been in force since 2020, as The Information first reported.

As a result, TSMC in recent weeks undertook a thorough review of its Chinese customer base, combing through employee emails and chip design documents, said two people with direct knowledge of the probe. During that probe, TSMC cut off supply to multiple clients with suspicious chip orders and destroyed the wafers they ordered, said the two people.

The Takeaway

- TSMC has conducted a sweeping review of Chinese customers since a U.S. investigation into possible Huawei links.
- Chinese officials have asked TSMC about the U.S. investigation and a potential tightening of export controls.
- TSMC executives told Commerce Department officials that U.S. export regulations were difficult to enforce.

The Commerce Department is exploring the possibility that Huawei bought chips from TSMC indirectly through intermediaries.

Earlier this month, the Commerce Department made the unusual move of ordering TSMC to halt shipments of advanced chips to Chinese clients, rather than waiting months to update formal policies to close loopholes and make other changes, according to three people with direct knowledge of the matter. (Reuters was the first to report the Commerce Department order.)

Meanwhile, TSMC is doing what it can to avoid getting under Beijing's skin, especially amid constant chatter of a possible China invasion of Taiwan. China considers the self-ruled island a breakaway province, and President Xi Jinping has not ruled out the use of military force to achieve reunification, putting the island under persistent threat of a potential invasion.

The company operates two factories in China and produces chips for many clients from the region. Disruption in the relationship with the Chinese government could have immediate consequences for its bottom line and ripple throughout the rest of the semiconductor industry.

In the U.S., TSMC's future may look complicated under the incoming Donald Trump administration. During his reelection campaign, Trump lamented that Taiwan had "stolen" the American chip industry, arguing that the U.S. was giving too much to Taiwan without getting enough back. He also suggested that the U.S. might impose high tariffs on imported chips, which could disrupt the semiconductor supply chain.

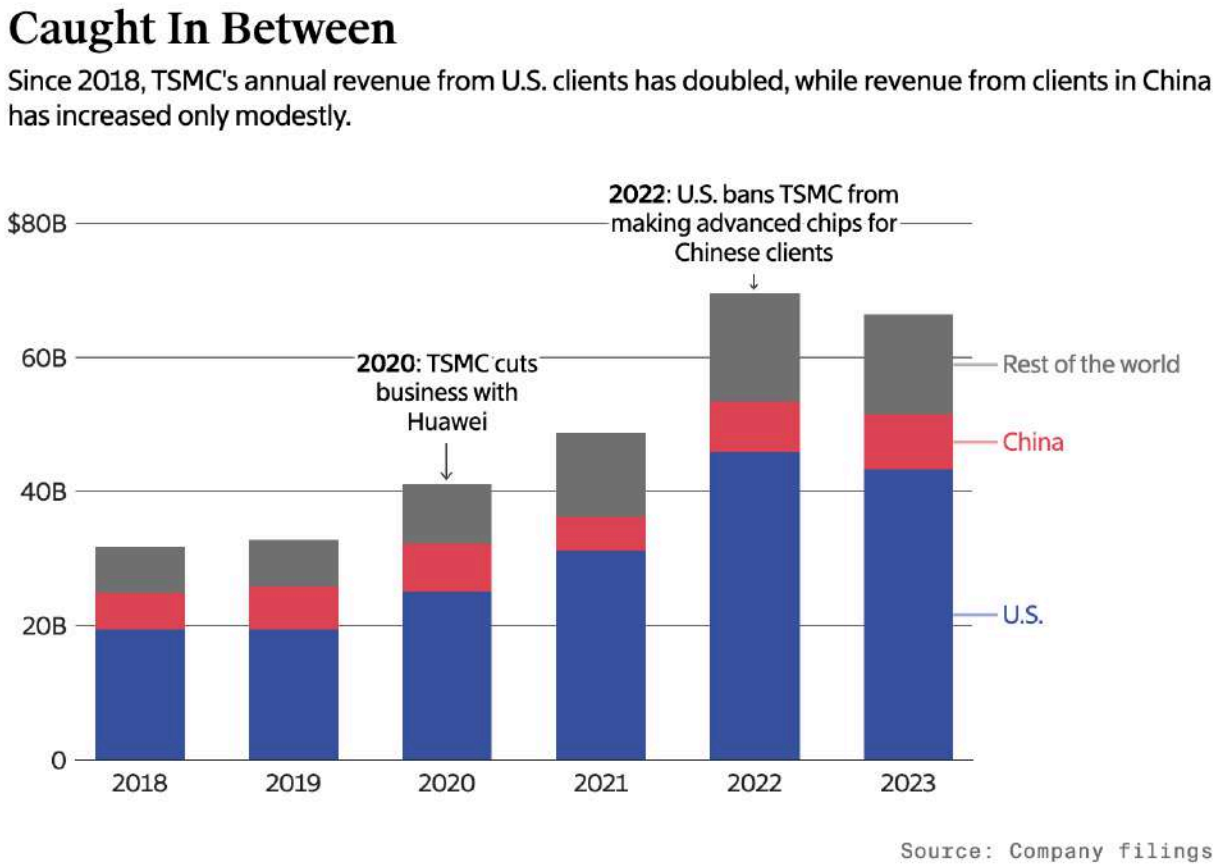
Recently, Chinese officials quizzed TSMC executives about the U.S. investigation and the U.S.'s possible tightening of future export controls, said two people with direct knowledge of the communications. TSMC told them the U.S. calls the shots and the company is just playing by its rules, the people said.

But the TSMC executives reassured the officials that the company would do only what the regulations required and no more, the people said.

TSMC said it maintains an export system for monitoring and ensuring compliance and will take prompt action if there are potential issues. "TSMC is a law-abiding company and we are committed to complying with all applicable rules and regulations, including applicable export controls," the company said in an emailed statement.

If the Commerce Department finds that TSMC flouted U.S. law, it will be a moment of reckoning for a company that has become a linchpin of the global tech industry. TSMC has become the leading manufacturer of advanced chips for artificial intelligence, cloud computing and smartphones, through skill and a single-minded focus on chip manufacturing, which helped differentiate it from rivals like Intel and Samsung that also design their own chips.

As far as the geopolitical pressures facing TSMC goes, the U.S. has the upper hand. Over 65% of the company's revenue now comes from American clients, including its two biggest customers, Apple and Nvidia. That's up from 59% in 2019.



That shift has occurred in part because the U.S. forced TSMC to stop working with some Chinese clients. Between 2019 and last year, the amount of TSMC's revenue derived from China fell to 12% from 19%. The most notable of those former Chinese clients was Huawei, which was the company's second-largest customer until the U.S. Commerce Department banned it from making unauthorized purchases of chips made with U.S. technologies due to national security concerns.

In recent quarters, Chinese companies contributed just over 10% of TSMC's revenue. The share will decline further if, as is likely, the U.S. imposes more restrictions. But TSMC still doesn't want to get on the wrong side of China because of the government's power and the size of its markets.

In a recent conversation with U.S. officials, though, TSMC executives have warned that the country's export regulations are difficult for the company to enforce, according to two people with direct knowledge of the matter. For instance, U.S. rules require chipmakers to gather chip performance data before accepting production orders from clients.

Manufacturers like TSMC, however, usually don't have access to such data until those products are made and tested. As a result, TSMC has to rely on its customers to report the details themselves, creating opportunities for them to manipulate or withhold crucial data, the people added.

A spokesperson for the Commerce Department said it has imposed more restrictions on the exports of advanced AI chips to China than any other administration while continuing to evaluate the effectiveness of and update its export controls. "No Commerce Department in history has been tougher on China," the spokesperson said.

Child of Globalization

The connections between TSMC and the U.S. run deep.

It was founded in 1987 by Morris Chang, a China-born, U.S.-educated engineer who previously led Texas Instruments' chip division. In the mid-1980s, Chang, who was then in his mid-50s and had spent all of his adult life in the U.S, was invited by the Taiwanese government to boost the island's semiconductor industry.

Chang moved to Taiwan and set about creating the world's first dedicated semiconductor foundry, which would make chips for other companies so they didn't need to build their own factories. The foundry enabled other firms to focus on chip design development without having to make the significant capital investment required for manufacturing facilities, filling a critical gap in the industry at that time.

In the early days, TSMC's management and technology reflected its mix of U.S. and Taiwanese roots. Chang licensed its initial chipmaking technology, inherited from IBM, while many of TSMC's executives—such as current chair C.C. Wei—shared his educational and work journey. They had earned their bachelor's and master's degrees in Taiwan, pursued their doctorates in the U.S. and gone on to work at major American tech companies.

TSMC rapidly scaled up its chipmaking know-how, attracting some powerful customers along the way. Apple became a hugely important one after it cut a deal with TSMC to manufacture the iPhone maker's in-house chip designs, with the first TSMC-made Apple processor showing up in the iPhone 6 in 2014. Soon TSMC-made chips were shipping in hundreds of millions of iPhones each year.

The iPhone's success—and the key role Apple's processors played in it—prompted other companies to seek out TSMC's services.

TSMC also relies on equipment from the U.S. and other Western countries. One of them is Lam Research, maker of some of the most advanced etching machines, which remove layers of material from wafers to create circuit patterns.

At the same time, TSMC has made significant investments in China. In the early 2000s, TSMC began establishing factories in China as part of its strategy to tap into one of the world's biggest consumer markets and expand its manufacturing capabilities. The company now has factories in two eastern cities in China, Shanghai and Nanjing, to cater to the growing demand for chips within the country.

‘Globalization Is Almost Dead’

While TSMC prospered through the growth of free global trade, it has had to adjust to a new reality in the U.S. in recent years.

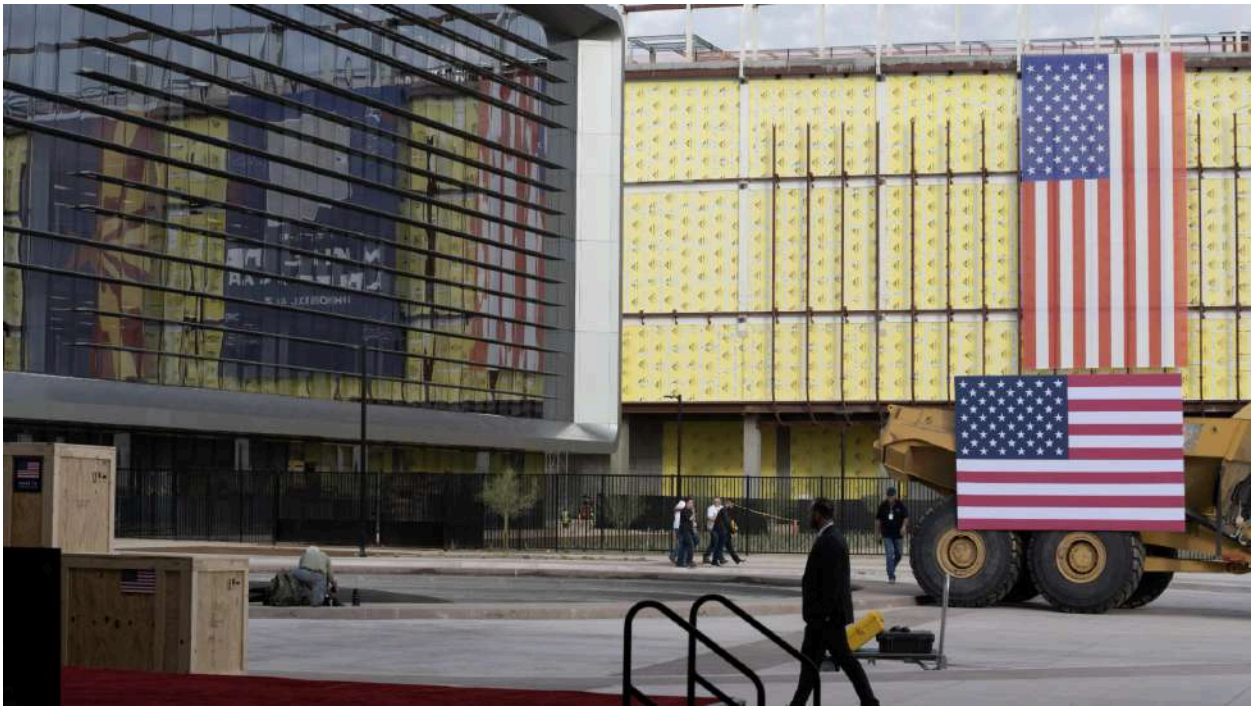
In 2020, TSMC announced it would build factories in Arizona with financial support from the U.S. government. The three factories—one of which is finished while the other two are still in the planning phase—are critical parts of the U.S. government’s strategy to bring chip manufacturing back home and reduce its reliance on Taiwan, based on the risk of a potential Chinese invasion.

But many critics viewed the $40 billion chip projects in the U.S. as unwise. Among them was Chang, TSMC’s 93-year-old founder, who is now retired.

Speaking on a Brookings Institution podcast in 2022, he described the U.S. push for increased chip manufacturing as a “wasteful, expensive exercise,” and said manufacturing costs were about 50% higher in the U.S. than in Taiwan. And at an event to celebrate the initial installation of equipment in TSMC’s first Arizona factory in 2022, Chang bemoaned how dramatically geopolitics had reshaped trade over the past couple of decades.

“Globalization is almost dead and free trade is almost dead,” he said. “A lot of people still wish they would come back, but I don’t think they will be back.”

There have also been issues with the construction of the Arizona factories, including equipment delivery delays and labor issues. The lack of a chip and electronics manufacturing base in Phoenix, Ariz., has posed staffing hurdles.



A TSMC factory in Arizona under construction in Dec. 2022. Photo by Bloomberg via Getty

A shortage of experienced workers in the U.S. prompted TSMC to bring in staff from Taiwan, the two people said. But the increasing number of Taiwanese workers also brought cultural conflicts.

A talent acquisition executive at TSMC, Deborah Howington, sued the company in federal court in Northern California last August, alleging that TSMC discriminated against people "who are not Asian and not Taiwanese citizens." Twelve former employees have joined the lawsuit as plaintiffs.

The new Trump administration adds further uncertainty to the company's future as a manufacturer in the U.S. During his campaign, Trump criticized the program, known as the CHIPS and Science Act, under which TSMC is set to receive $6.6 billion in grants and $5 billion in loans. This month, House Speaker Mike Johnson hinted at a possible repeal of the law, though he later retracted that statement.

Avoiding Fines

TSMC isn't the only company receiving notifications from the Commerce Department about export restrictions.

In 2022, the Commerce Department sent letters to Nvidia and AMD, limiting their ability to export advanced AI chips to China. These rules later became part of the official export regulations, and the department applied them to other companies as well.

TSMC could face a wide range of fines if the Commerce Department determines it violated U.S. rules. Recently, the department fined GlobalFoundries $500,000 for shipping chips to an affiliate of a sanctioned Chinese company, Semiconductor Manufacturing International Corp. In 2023, Seagate Technology received a hefty $300 million fine for selling hard drives to Huawei.

TSMC hopes it can avoid heavy fines and only get a warning by proving how challenging it is to identify suspicious clients while also promising to comply with the department's rules as best it can, said two people involved in the investigation.

Jacob Feldgoise, a data research analyst at Georgetown University's Center for Security and Emerging Technology, said the Commerce Department's Bureau of Industry and Security, which is responsible for enforcing the export control rules, could decide to show TSMC leniency.

"The level of cooperation between the company and [the] BIS, if the company has been suspected of a violation, can make a big difference in how they are penalized or whether they are penalized down the road," said Feldgoise.

Digital Version:
https://www.theinformation.com/articles/tsmcs-push-to-be-techs-switzerland-in-doubt-as-u-s-china-tensions-grow

The Information

Chinese Regulators Tell Local Tech Firms to Buy Fewer Nvidia Chips

Nvidia's booth at a trade show in Hangzhou, China in October 2023. Photo by FeatureChina via AP.

By Qianer Liu

May 13, 2024

Chinese authorities have a message for the local tech industry: Buy Chinese.

Officials with government agencies in recent months have told tech firms, including TikTok parent ByteDance, Tencent, Alibaba and Baidu, to cut back on their purchase of foreign-made artificial intelligence chips in favor of buying more domestically made chips, said two people who work with the tech giants.

While foreign-made hardware, especially Nvidia's graphic processing units, has been the go-to choice for most Chinese tech companies for AI, the Chinese government now wants tech companies to buy equal numbers of domestic and foreign-made AI chips for new internet data centers, the people said. It's the first time the government has set specific guidelines for where companies should buy their AI chips.

The Takeaway

- Two Chinese government agencies told local tech firms to buy Chinese chips
- Guidelines will force tech companies to cut back on U.S. chip purchases
- Directive reflects an effort to develop technological self-sufficiency

The directive highlights how China is trying to balance a desire for technological self-sufficiency with a goal of global tech leadership. Blocking Chinese companies from buying any U.S. chips could impede China's progress in the realm of AI and cloud computing, which runs counter to Beijing's vision of competing with the U.S. for tech leadership. Foreign-made servers, running on the graphics processing units from Nvidia, are crucial for Chinese tech firms developing large language models.

But the escalating series of export controls put in place by the Biden administration has prompted the Chinese government to focus on ways to ensure its companies are more self-sufficient.

Branches of the National Development and Reform Commissions—China's top economic planner—and the Ministry of Industry and Information Technology, which regulates telecom,

internet, and other areas, shared the guidance verbally in recent months when companies registered their plans to build new or expand existing data centers. National Development and Reform Commissions and the Ministry of Industry and Information Technology did not respond to requests for comments.

The guidelines could hurt the sales of new versions of Nvidia's specialized AI chips designed to comply with the tightened export controls the Biden administration put in place last year, blocking the sale of certain chips to Chinese companies. Nvidia then engineered the new chips for the Chinese market. Since their release, companies in China have been testing and evaluating the new GPUs—known as H20, L20 and L2—to figure out how many to buy.

Nvidia declined to comment. ByteDance, Alibaba, Tencent and Baidu did not reply to requests for comments.

Some firms are quite enthusiastic about the H20, according to two Chinese engineers, having realized that it can be deployed in large quantities to power AI as efficiently as Nvidia's more advanced chips, which aren't available to Chinese firms anymore.

If companies opt for ordering more foreign chips over their homegrown equivalents, they will have to detail in writing how many U.S. chips they have ordered for deployment and justify the decision, said three people directly involved in the conversations between the agencies and the companies.

The new guidance hasn't yet been strictly enforced. Many AI companies and cloud service providers that depend heavily on foreign chips are waiting to see whether the government will penalize anyone who doesn't strictly follow those rules. If the Chinese government levies penalties or imposes stronger policies forcing Chinese companies to cut back on procurement of foreign chips, that could delay the development of cloud computing networks and AI in China.

Even if the government doesn't take such steps, the new guidelines could hurt the sales of Nvidia's new chips tailored for the China market as well as server chips made by AMD and Intel in the Chinese market.

Many Chinese companies will struggle to comply with the government's guidelines, given the substantial performance gap that now exists between domestic and overseas AI chips. Adopting some Chinese-made chips will also force some companies to move off Nvidia's CUDA software, which works with Nvidia chips to help developers improve the performance of their applications. But CUDA doesn't work well with some Chinese chips.

Chinese companies have ordered at least 350,000 Nvidia H20 chips, worth about $4 billion, for delivery in 2024, according to four people involved in the chip sale. That's down from the $5 billion Chinese firms ordered from Nvidia last year, although some of those were not delivered after the U.S. tightened export controls.

Tencent, for instance, has begun building several clusters of Nvidia's new H20 chips. The largest computing cluster may consist of more than 60,000 H20 GPUs, said two people close to Tencent.

Nvidia has witnessed a significant downturn in its data center–related revenue from China in the last quarter, and attributes it to U.S. export restrictions. This resulted in China's share of Nvidia's data center business dwindling to a mid–single-digit percentage, after it had previously accounted for 19% in fiscal year 2023. The company's latest earnings report is due for release on May 22.

Digital Version:
https://www.theinformation.com/articles/chinese-regulators-tell-local-tech-firms-to-buy-fewer-nvidia-chips