TAIWAN AND THE GLOBAL SEMICONDUCTOR SUPPLY CHAIN

Edited by:
Chen-Yuan Tung, Ph.D.
Representative
Taipei Representative Office in Singapore

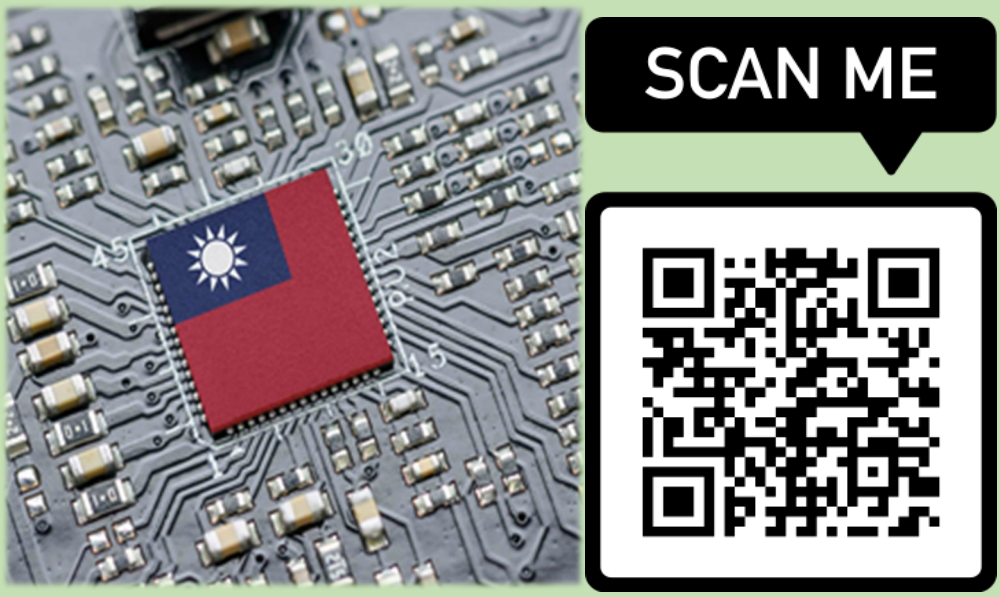
Please feel free to reach out to the Economic Division of the Taipei Representative Office in Singapore should you have any enquiries or are seeking partnership opportunities of investment or collaboration in the field of semiconductors in Taiwan.

Email: singapore@sa.moea.gov.tw
Telephone: +65 6500-0128

Published: Taipei Representative Office in Singapore
Address: 460 Alexandra Road, #23-00 mTower, Singapore 119963
Email: sgp@mofa.gov.tw
Telephone: +65 6500-0100

The Taipei Representative Office in Singapore provides monthly reports on Taiwan and global semiconductor supply chains. We welcome you and your friends to join our WhatsApp community, "**Taiwan Semiconductor Reports**".

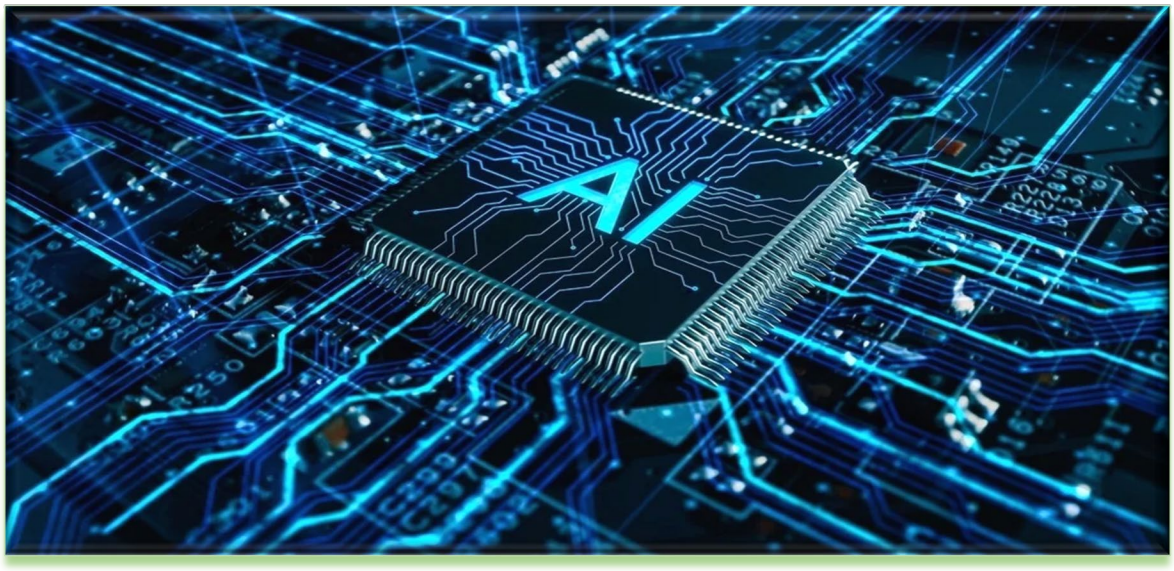
https://chat.whatsapp.com/BqwdAMgi1sUIGsujDx3YDk



IN THE SPOTLIGHT

Artificial Intelligence and the Semiconductor Industry

- **Advancements in Artificial Intelligence (AI) and its widespread adoption are opening up new frontiers in nearly all sectors.**
- **Semiconductor technology is at the core of all present and future AI-related innovations.**
- **TSMC, Samsung and Intel are major players in AI chip fabrication.**
- **The opportunities and challenges posed by AI and the semiconductor industry are of great interest to governments and policymakers worldwide.**



Source: Gorodenkoff, Shutterstock.com

OVERVIEW

Artificial Intelligence (AI) and semiconductors are two deeply interconnected fields that are driving technology and innovation. Semiconductors, the tiny chips that power electronic devices, play a crucial role in the functioning of AI systems. In recent years, progress in semiconductor technology, including the creation of smaller and more powerful chips and the advent of advanced semiconductor packaging techniques, has enabled tremendous technical progress in AI. In turn, AI is revolutionizing and creating new opportunities in the semiconductor industry, particularly in chip design, manufacturing technology and the supply chain.[1]

[1] Tim Culpan, "AI is helping create the chips that design AI chips," Bloomberg, May 25, 2023.

By stretching the limits of what is possible and opening up new frontiers in nearly all sectors, AI is shaping the lives of people across the globe, including the way we work, live and play. Indeed, AI and semiconductors are not only integral to our lives but are also shaping our future. Due to their transformative potential and the challenges that they pose, AI and by extension, semiconductors are an expanding area of research and a policy concern.[2]

WHAT IS ARTIFICIAL INTELLIGENCE?

The United Nations Educational, Scientific and Cultural Organization (UNESCO)'s Recommendation on the ethics of artificial intelligence notes that "Artificial Intelligence (AI) systems are information-processing technologies that integrate models and algorithms that produce a capacity to learn and to perform cognitive tasks leading to outcomes such as prediction and decision-making in material and virtual environments. AI systems are designed to operate with varying degrees of autonomy by means of knowledge modelling and representation and by exploiting data and calculating correlations."[3]

AI can be used both on the cloud and on the edge. Cloud AI involves running AI applications in a cloud computing environment, where AI algorithms and models are executed, and AI services are delivered using internet technologies. Examples of Cloud AI services include Google Cloud AI, Amazon AI Services, Microsoft Azure AI, and IBM Watson.[4] AI on the edge, on the other hand, operates on local devices such as wearables, Internet of Things (IoT) devices, or edge computing servers.[5] The use of AI both on the cloud and on the edge makes AI more prevalent and accessible.

Technologies such as machine learning, deep learning, and natural language processing are all part of the larger AI landscape.[6] These technologies are used in applications in our daily lives, including autonomous driving, virtual

[2] The White House, "Remarks by President Biden on Artificial Intelligence," July 21, 2023.
[3] UNESCO, Recommendation on the ethics of artificial intelligence, 2021, https://en.unesco.org/artificial-intelligence/ethics.
[4] Gartner Peer Insights, "Cloud AI Developer Services Reviews and Ratings, 2024, https://www.gartner.com/reviews/market/cloud-ai-developer-services.
[5] Nacho Palou, "Cloud AI vs. Edge AI: know their differences and choose the right approach for your AI project," Telefonica Tech, June 14, 2023.
[6] Artificial Intelligence Board of America, "AI Strategies: Using NLP, ML, and DL for Businesses," July 7, 2021.

assistants, financial trading, spam filtering and language translation. AI advances are also providing great benefits to our social wellbeing in areas such as precision medicine, environmental sustainability, education, and public welfare.[7]

A popular class of AI systems is deep neural networks (DNNs), which is capable of learning and recognizing patterns in large amounts of data. DNNs use algorithms, or models, to mimic neurons in the brain to identify complex patterns. The computational power required by DNNs for its complex mathematical modeling is provided by semiconductors.[8] DNNs involve two main phases: training and inference. During training, the AI model is fed data to identify patterns. In the inference stage, the trained model is used to enable predictions and guide decisions, such as autonomous driving systems detecting and avoiding obstacles on the road. In general, the accuracy of AI models increases with training on larger amounts of data, which in turn requires more computational power.

Traditional AI and Generative AI

Traditional AI and generative AI are two different approaches to artificial intelligence, each with its own strengths and applications. Traditional AI systems usually perform a specific task and include examples such as spam filters, voice assistants like Siri or Alexa, recommendation engines on Netflix or Amazon, or Google's search algorithm.[9] In contrast, generative AI is usually broader and can create new content, partly because generative AI tools are trained on larger and more diverse data sets than traditional AI. Furthermore, unlike traditional AI which is usually trained using supervised learning techniques, generative AI is trained without continually needing human intervention.[10] Examples of generative AI include chatbots like OpenAI's ChatGPT, Microsoft's Copilot and Google's Gemini.

While traditional AI is used in finance, healthcare, and manufacturing industries to automate processes and improve efficiency, generative AI opens

[7] U.S. Department of State, "Artificial Intelligence (AI)," 2020, https://www.state.gov/artificial-intelligence/.
[8] Congressional Research Service, "Semiconductors and Artificial Intelligence," September 18, 2023.
[9] Emily Heaslip, "What's the Difference Between Traditional and Generative AI?" CO by US Chamber of Commerce, October 16, 2023.
[10] Greg Pavlik, "What Is Generative AI? How Does It Work?" Oracle Singapore, September 15, 2023.

new doors of creativity and adaptability, making it a powerful tool in various industries. Industries relying most heavily on knowledge work are likely to see more disruption—and potentially reap more value from generative AI.[11] As AI research advances, both traditional AI and generative AI will complement each other, contributing to a more robust and versatile AI ecosystem.

Popular large language models, such as Generative Pre-trained Transformer 4 (GPT-4), are trained on trillions of text data to process and generate text.[12] AI models are trained on specialized hardware like graphics processor units, which draw more power and are generally more expensive than traditional central processing units. Given the large data sets associated with AI models, some of the largest AI models can take weeks or months to train, using thousands of chips and costing millions of dollars.[13] As powering and training of artificial intelligence models is energy intensive, the electricity required to operate the hardware contributes to the high costs, and there are also concerns on the environmental impact of the growing use of AI.[14]

AI models are becoming ever more expensive to develop, in terms of raw computational power and the training on massive datasets. According to OpenAI, the research organization responsible for developing GPT-3, the project's total cost is estimated to be around US$ 4.6 million.[15] Meanwhile, the newer GPT-4 model costs more than US$ 100 million to train during its initial development and requires about US$ 700,000 a day to run.[16] The typical cost of developing a large AI model may soon be in the billions of dollars.[17] As the financial cost of AI development rises, only a few select labs can afford to do it, and they will be the ones to set the agenda for what kinds of AI models get developed.

[11] McKinsey and Company, "The state of AI in 2023: Generative AI's breakout year," August 1, 2023.
[12] Ernestas Naprys, "Scientists to make their own trillion parameter GPTs with ethics and trust," Cybernews, November 28, 2023.
[13] Congressional Research Service, "Semiconductors and Artificial Intelligence," September 18, 2023.
[14] Lauren Leffer, "The AI Boom Could Use a Shocking Amount of Electricity," Scientific American, October 13, 2023.
[15] Kevin Pocock, "How much did GPT-3 cost?" PC Guide, August 11, 2023.
[16] Erik Brynjolfsson and Gabriel Unger, "The Macroeconomics of Artificial Intelligence," International Monetary Fund, December 2023.
[17] Craig Smith, "What Large Models Cost You – There Is No Free AI Lunch," Forbes, September 8, 2023.

SEMICONDUCTOR USE IN AI-BASED APPLICATIONS

AI models employ different types of chips. General chips such as logic chips and memory chips respectively process and store large amounts of data, making it possible to create and use AI.[18] In recent years, advanced semiconductor packaging techniques, by enabling the integration of various components from different wafers into a single unit, have played a crucial role in the development of AI-specific chips such as Application-Specific Integrated Circuits (ASICs) and Systems-on-a-Chip (SoCs).[19]

ASICs are single-purpose chips used for performing repetitive processing routines such as scanning a barcode. SoCs, on the other hand, are essentially integrator chips that combine many chips and circuits in a single chip. As both ASICs and SoCs are designed from ground up for AI, they can perform more computations per unit of energy, resulting in faster processing speeds and lower energy consumption compared to general-purpose chips.[20]

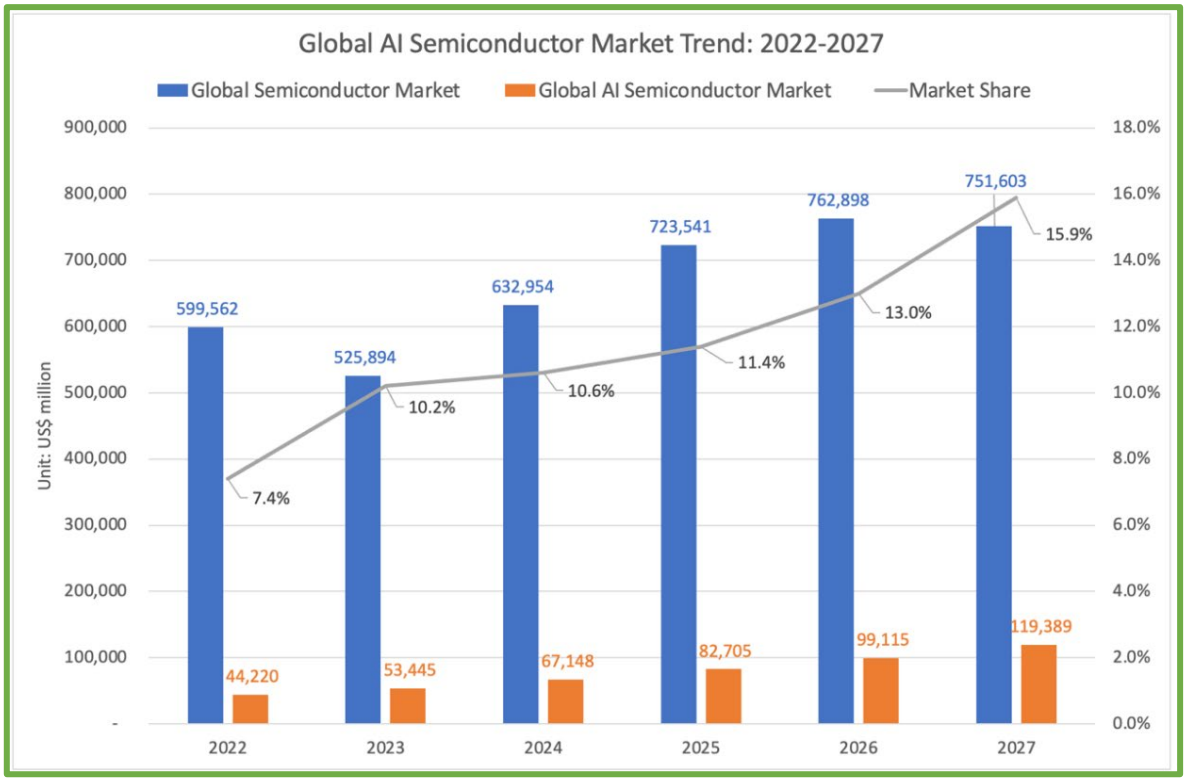
By delivering new functionalities, better performance and lower cost with each generation, advances in chips have spawned new products and transformed industries. The global semiconductor market is forecasted to grow from 2022 to 2027, albeit with some fluctuations in between years, as demand for consumer goods with AI applications amplifies across the globe. The global semiconductor market size which was valued at US$ 599.6 billion in 2022 is projected to fall to US$ 525.9 billion in 2023. From 2024 onwards, the market is expected to grow from US$ 633.0 billion to US$ 762.9 billion in 2026, before falling to US$ 751.6 billion in 2027 (See Figure 1).

2023 has been described as generative AI's breakout year.[21] Following its launch in November of 2022, OpenAI's ChatGPT set the record as the fastest-growing consumer application of all time and ushered generative AI to the mainstream. Other generative AI tools such as Google's Gemini and Microsoft's Copilot AI came quickly on the heels of ChatGPT's success. As the year progressed, numerous startups also emerged to innovate with generative

[18] Congressional Research Service, "Semiconductors and Artificial Intelligence," September 18, 2023.
[19] Ondrej Burkacky, Taeyoung Kim, and Inji Yeom, "Advanced chip packaging: How manufacturers can play to win," McKinsey, May 24, 2023.
[20] Saif M. Khan and Alexander Mann, Policy Brief: "AI Chips: What They Are and Why They Matter, An AI Chips Reference", Center for Security and Emerging Technology, April 2020.
[21] McKinsey, "The state of AI in 2023: Generative AI's breakout year," August 2023.

AI.[22] Consequently, AI semiconductor revenue is increased by the growing use of generative AI in data centers in the near term, and by the deployment of AI into PCs, smartphones, edge and endpoint devices in the longer term.[23]

Figure 1: Global AI Semiconductor Market Trend for 2022 to 2027



Source: Hsuan Chih Wang, "The AI Semiconductor Market Shows Strong Growth, with Emerging Consumer AI Applications," IEK, ITRI, October 2, 2023, p. 1.

Table 1: Global AI Semiconductor Market Growth Trend: 2023-2027

	2023	2024	2025	2026	2027
Global Semiconductor Market	-12.3%	20.4%	14.3%	5.4%	-1.5%
Global AI Semiconductor Market	20.9%	25.6%	23.2%	19.8%	20.5%

Source: Hsuan Chih Wang, "The AI Semiconductor Market Shows Strong Growth, with Emerging Consumer AI Applications," IEK, ITRI, October 2, 2023, p. 1.

Despite rapid technological progress in AI making predictions of the AI semiconductor market challenging, the AI semiconductor market is expected to show a rapid growth trend compared with the global semiconductor

[22] Lori Perri, "What's New in Artificial Intelligence from the 2023 Gartner Hype Cycle," Gartner, August 17, 2023.
[23] Alan Priestley, "Forecast: AI Semiconductors, Worldwide, 2021-2027, 3Q23 Update," Gartner Research, October 17, 2023.

market. When the global semiconductor market declined 12.3% to US$ 525.9 billion in 2023, the Al semiconductor market was forecasted to grow from US$ 44.2 billion in 2022 to US$ 53.4 billion in 2023 (see Figure 1), an increase of 20.9% from 2022, according to Gartner (see Table 1).[24]

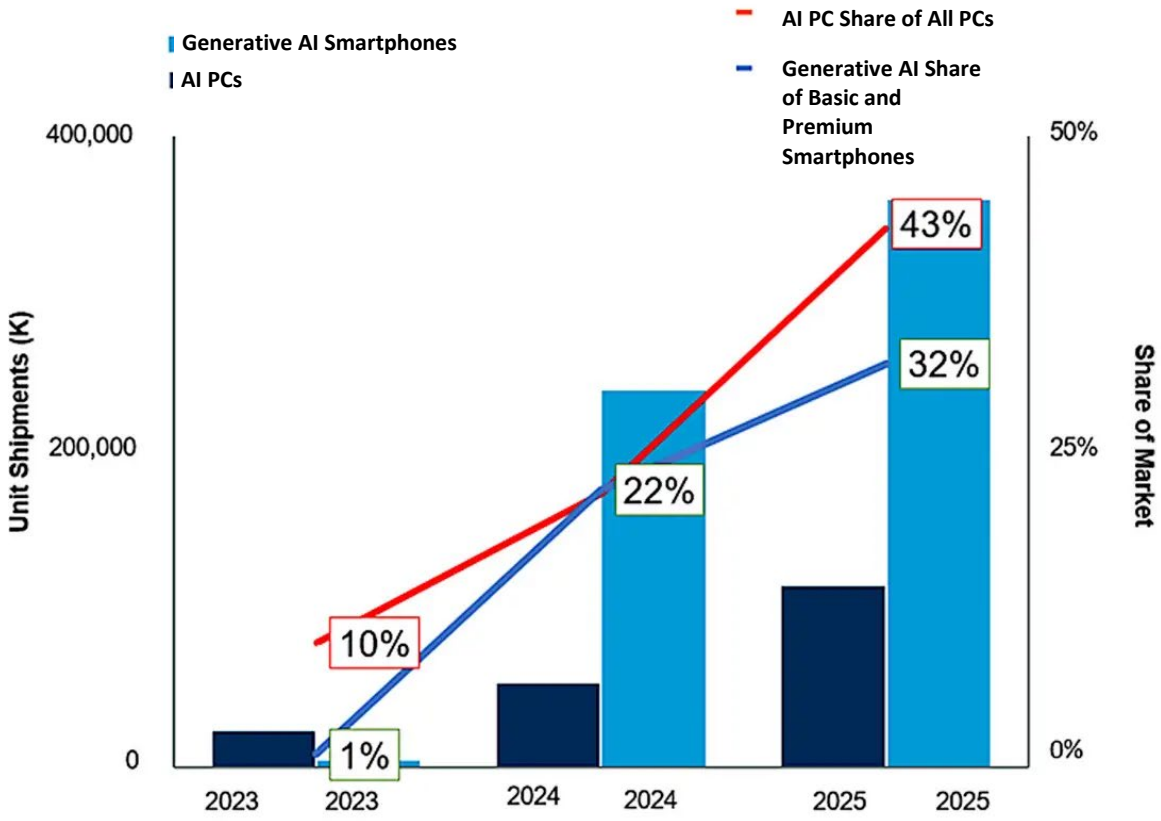
With more users experimenting directly with generative AI applications and the increasing use of a wide range AI-based applications in data centers, edge infrastructure and endpoint devices, the demand for high-performance graphics processing units (GPUs) and optimized semiconductor devices is soaring. In fact, worldwide shipments of AI PCs and generative AI smartphones are projected to total 295 million units by the end of 2024, up from 29 million units in 2023, according to a forecast from Gartner, Inc.[25] Gartner estimates 240 million generative AI smartphones and 54.5 million AI PCs will be shipped by the end of 2024 (see Figure 2). This will represent 22% of basic and premium smartphones and 22% of all PCs in 2024. Hence, AI semiconductor revenue is expected to continue to experience double-digit growth from 2023 to 2027 (See Table 1). By 2027, AI chips revenue is expected to be more than double the size of the market in 2023, reaching US$ 119.4 billion (See Figure 1).

In a nutshell, Al semiconductors' share in the global semiconductor market will see a steady increase – from 7.4% in 2022 to 15.9% in 2027. This increase of 8.5 percentage points from 2022 to 2027 reflects the growing adoption of Al technology. It also means that in the next few years, the Al semiconductor market will account for significantly more than 10% of the global semiconductor market. Clearly, the Al semiconductor market is becoming an integral part of the global semiconductor market (see Figure 1).

[24] Gartner, Press Release "Gartner Forecasts Worldwide AI Chips Revenue to Reach $53 Billion in 2023 by 2027, AI Chips Revenue Is Forecast to More Than Double," August 22, 2023.

[25] Gartner, Press Release: "Gartner Predicts Worldwide Shipments of AI PCs and GenAI Smartphones to Total 295 Million Units in 2024," February 7, 2024.

Figure 2: AI PCs and Generative AI Smartphones Market Share, Worldwide, 2023-2025



Source: Gartner, Press Release: "Gartner Predicts Worldwide Shipments of AI PCs and GenAI Smartphones to Total 295 Million Units in 2024," February 7, 2024.

KEY PLAYERS IN AI CHIP FABRICATION

The semiconductor industry's big three – Taiwan Semiconductor Manufacturing Company (TSMC), Samsung and Intel – are key players in fabricating chips, especially essential ones for AI development. According to TrendForce, TSMC had 57.9% of the market share of the third-party chip-manufacturing business, followed by Samsung with 12.4% and Intel Foundry Services with 1.0% in the third quarter of 2023.[26] All three are racing to produce 2 nanometer (nm) processor chips that will power the next generation of smartphones, data centers, and artificial intelligence by 2025.[27] Several media outlets have reported Samsung is targeting mass production at the 1.4

[26] Chang Chien-chung and Frances Huang, TSMC cements global lead as market share rises to 57.9% in Q3, Focus Taiwan, December 9, 2023.

[27] Yifan Yu, "Intel joins 1.4-nanometer chip race against TSMC and Samsung," Nikkei Asia, February 22, 2024.

nm level in 2027 while TSMC is aiming for 2027 to 2028. Of the Big Three, TSMC remains the analysts' favorite to maintain its global supremacy in the sector.[28]

Besides reducing nanometer size to yield more powerful and efficient chips, today's most powerful AI chips require advanced packaging methods to enable the greatest performance.[29] TSMC, with a cache of 2,946 advanced packaging patents, has developed the most expansive arsenal of patents surrounding advanced chip packaging, followed by Samsung Electronics (2,404) and then Intel (1,434), according to LexisNexis.[30]

TSMC's Chip-on-Wafer-on-Substrate (CoWoS®), Intel's Embedded Multi-Die Interconnect Bridge (EMIB), and Samsung's Interposer-Cube (I-Cube) are mature advanced semiconductor packaging technologies that are widely utilized in high-performance chips.[31] By 2024, suppliers are expected to focus on ramping up 2.5D packaging capacity to meet the rising demand for high computational power in applications like AI. Additionally, both TSMC and Samsung have also taken steps to establish 3D integrated circuit research centers in Japan, highlighting the critical role of packaging in the evolution of semiconductor technology.[32]

With their ability to innovate and adapt to the rapidly evolving AI landscape, these companies are poised to ride the AI boom and continue their lead in the years to come.

TSMC

TSMC is a major provider of logic chips for AI. As a leader in advanced chip packaging technology and the world's largest contract chipmaker, TSMC is a key partner to both established and startup AI chip designers. TSMC

[28] Christian Davies, Song Jung-a, Kathrin Hille and Qianer Liu, "Semiconductor giants race to make next generation of cutting-edge chips," Financial Times, December 11, 2023.
[29] Refer to the January issue of "Taiwan and the Global Semiconductor Supply Chain" for more information.
[30] Max A. Cherney, "TSMC leads in advanced chip packaging wars, LexisNexis patent data says," Reuters, August 2, 2023.
[31] Trendforce, Press Release: "TrendForce 2024: Riding the Wave of Revolutionary Tech Trends," October 17, 2023.
[32] TSMC, Press Release: "TSMC Japan 3DIC R&D Center Completes Clean Room Construction in AIST Tsukuba Center," June 24, 2022; Jin Eun-Soo, "Samsung to build $280 million chip research center in Japan," Korea JoongAng Daily, December 21, 2023.

produces chips designed by tech giants such as NVIDIA, the frontrunner in the AI chip race; Advanced Micro Devices (AMD), NVIDIA's rival; and Arm Holdings, which has a significant processor share across Google's Android and Apple's iOS devices.[33] Notable AI chip design startups such as US-headquartered SambaNova and UK-based Graphcore, which both specialize in AI chips for machine learning, are also among TSMC's customers.[34] Graphcore, for example, has used TSMC's wafer-on-wafer 3D integration technology to speed up AI processing by 40%.[35] TSMC also produces customized AI chips for data centers such as Amazon Web Services (AWS), Microsoft and Google, and other companies such as Apple.[36] TSMC's partnerships with these companies allow them to focus on designing innovative AI chips, while TSMC handles the manufacturing process.

At an investor conference on January 18, 2024, TSMC projected a year-on-year sales increase of 21-26% in 2024 due to the booming demand for high-end chips used in AI applications. In view of the company's lead over its rivals in advanced technology development, TSMC's 2024 sales projections are higher than an expected 20% in the global pure play wafer foundry industry.[37]

Additionally, TSMC has already disclosed the test results of its 2 nm prototype process to major clients such as Apple and NVIDIA, and Apple is set to become TSMC's inaugural customer for the 2 nm process when the mass production of its 2nm chips commences in 2025.[38]

Samsung

Samsung is the world's largest memory chipmaker. A surge in demand for high bandwidth memory (HBM), a critical component in all AI chips, has meant new growth for memory-chip makers. According to Morgan Stanley, the global HBM market, which grew from US$ 1.9 billion in 2022 to an estimated

[33] Aditya Soni and Roshan Abraham, "Arm's shares get helping hand as IPO banks weigh in with ratings," Reuters, October 9, 2023.
[34] SambaNova, Press Release: "SambaNova Unveils New AI Chip, the SN40L, Powering its Full Stack AI Platform," September 19, 2023.
[35] Samuel Moore, "Graphcore Uses TSMC 3D Chip Tech to Speed AI by 40%," IEEE Spectrum, March 3, 2022.
[36] Jai Vipra & Sarah Myers West, AI Now Institute, Computational Power and AI," September 27, 2023.
[37] Chang Chien-chung and Frances Huang, "TSMC forecasts sales to grow by over 20% in 2024," Focus Taiwan, January 18, 2024.
[38] Trendforce, Press Release: "TSMC's Primary Client for 2nm Chips Expected to be Apple, Set to Debut with iPhone 17 Lineup Next Year," January 26, 2024.

US$ 4 billion in 2023, is forecasted to more than double in 2024. The strong demand for HBM memory chips is a boon to Samsung, which is expected to command a 47-49% share of the global HBM market this year.[39]

Besides memory chips, Samsung also produces its own AI chips, such as Exynos and ISOCELL.[40] Samsung has also innovated new types of solid-state drives (SSDs) optimized for AI and machine learning applications.[41]

Several media have reported that Samsung is trying to seize the market for its 2 nm process by offering lower prices.[42] The first beneficiary of this discount is Preferred Networks (PFN), which has reportedly chosen Samsung over TSMC.[43] PFN is a Japanese startup known for its expertise in AI deep learning development and has attracted large-scale investments from major companies, including Toyota Motor, NTT and FANUC, a Japanese robot system maker. According to Korean media, PFN decided to forgo TSMC as Samsung can provide a comprehensive chip-making service on a turnkey basis, from chip design to production and advanced chip packaging.[44]

Intel

Intel has set itself a goal of ensuring that 50% of the world's semiconductors are built in the United States and Europe in a decade, compared with 20% today.[45] As part of its 'IDM 2.0' strategy, Intel is reorganizing into two units — Intel Foundry, which manufactures semiconductors, and Intel Products, involved in their design.[46] Launched on February 21, 2024 as the world's first systems foundry for the AI era, Intel Foundry positions itself as a 'systems foundry', which brings system-level

[39] Jeong-Soo Hwang, "Samsung eyes $8.8 billion 2024 chip profit: sources, The Korea Economic Daily, January 2, 2024.
[40] Jeong-Soo Hwang, "Samsung Electronics to up AI chip foundry sales to 50% by 2028," November 20, 2023.
[41] Samsung, "How Samsung Memory Is Powering the AI Revolution," May 16, 2023, https://semiconductor.samsung.com/us/news-events/tech-blog/how-samsung-memory-is-powering-the-ai-revolution/
[42] Christian Davies, Song Jung-a in Seoul, Kathrin Hille in Taipei and Qianer Liu, "Semiconductor giants race to make next generation of cutting-edge chips," Financial Times, December 11, 2023.
[43] Trendforce, Press Release: "Samsung Secures 2-Nanometer Order from Japanese AI Startup Preferred Networks," February 16, 2024.
[44] Jeong-Soo Hwang, "Samsung wins first 2 nm AI chip order from Japan's PFN; a snub to TSMC," The Korea Economic Daily, February 16, 2024.
[45] Michael Acton, "Intel to manufacture chips for Microsoft as AI drives demand," Financial Times, February 22, 2024.
[46] Ibid.

design, technology development, supply chain capabilities, packaging, connectivity fabric solutions, and cooling technology under one umbrella.[47] Intel has roped in support from ecosystem partners – including Synopsys, Cadence, Siemens and Ansys – to expedite chip designs for Intel Foundry customers using validated tools, design flows, and IP portfolios for Intel's advanced packaging and 18A process technologies. Under a US$ 15+ billion deal, Microsoft will be leveraging Intel's 18A process technology to manufacture an undisclosed AI chip that Microsoft itself designed.[48]

Serving both external and internal customers, Intel Foundry aims to be the second-largest foundry globally by 2030 and expects to beat an internal deadline of 2025 to overtake its biggest rival, TSMC, in advanced chip manufacturing. [49]

IMPACT OF AI ON THE SEMICONDUCTOR INDUSTRY

The rise of AI has led to a surge in demand for semiconductors, driving innovation and growth in the semiconductor industry. AI applications require chips that are faster, more power-efficient, and capable of handling vast amounts of data. This has led to the development of advanced semiconductor technologies, such as the 3 nanometer process nodes currently used by TSMC and Samsung,[50] the exploration of new materials beyond silicon,[51] and advanced packaging techniques.[52]

AI is increasingly prevalent in the semiconductor industry. One area where AI is playing a significant role is in the design of semiconductor chips. Advanced AI tools can test human designs by finding placement errors that increase power consumption, impede performance, or use space inefficiently; suggesting improvements; and then simulating and testing those. Additionally, AI tools are more efficient in defect detection. For example, AI can be used to

[47] Intel, Press Release: "Intel Launches World's First Systems Foundry Designed for the AI Era," February 21, 2024.
[48] Stephen Nellis and Max A. Cherney, "Intel signs Microsoft as foundry customer, says on track to overtake TSMC," Reuters, February 22, 2024.
[49] Ibid.
[50] Alan Patterson, "Samsung Seen Stumbling at Silicon's Leading Edge," EE Times Asia, December 29, 2023.
[51] Milton D'Silva, "Beyond Silicon – What Will Replace the Wonder Material?" EMEA Industry, February 28, 2023.
[52] Ondrej Burkacky, Taeyoung Kim, and Inji Yeom, "Advanced chip packaging: How manufacturers can play to win," McKinsey & Company, May 24, 2023.

improve fault detection by visual inspection of wafers by almost nine times.[53] Compared to traditional methods, advanced AI tools can enable faster and more efficient design of new chips.

Moreover, AI is increasingly being used in the semiconductor supply chain.[54] Chip companies can use AI to address supply chain challenges such as managing a network of outsourced semiconductor assembly and test (OSAT) providers. AI-based solutions also provide near-real-time insights into pricing and demand fluctuations, helping companies build resilience into their supply chains.[55] According to Gartner, more than 80% of new supply chain management technology applications will use AI and data science in some way by 2025.[56] Those who don't invest now risk falling behind.

TSMC, which projected upwards of 20% growth in 2024 driven by demand for artificial intelligence, leverages AI to enhance its operations. Besides using AI for chip design, demand forecasting and streamlining its operations, TSMC also uses AI for its trade secret management. As of January 2021, over 100,000 trade secrets have been registered and recorded by 30,000 employees in its trade secret management system.[57] In 2023, TSMC developed a customized AI talent scouting system called 'Trade Secret Intelligent Management Version 2.0: Innovation Talent Scouting Online Merge Offline Service ("ITS OMO")' to develop a greater pool of prospective talent and further encourage innovations.[58]

Like TSMC, Samsung also uses AI to improve its chip design and production processes. According to South Korean media reports, Samsung Electronics is planning to fully automate its semiconductor factories by 2030, with "smart sensors" set to control the manufacturing process. In doing so,

[53] Tobias Schlosser et al., "Improving automated visual fault inspection for semiconductor manufacturing using a hybrid multistage system of deep neural networks," Journal of Intelligent Manufacturing 33: pp. 1099–1123, 2022.

[54] Institute of Electrical and Electronics Engineers (IEEE), "International Roadmap for Devices and Systems, 2020 Edition: Semiconductors and Artificial Intelligence", 2020, https://irds.ieee.org/topics/semiconductors-and-artificial-intelligence.

[55] Francisco Caudillo, Michael Gordon, Wim Gysegom and Juan Sarmiento, "Chip hunting: The semiconductor procurement solution when other options fail," McKinsey & Company, April 4, 2023.

[56] Nick Carey, "Focus: Startups apply artificial intelligence to supply chain disruptions," Reuters, May 3, 2022.

[57] Yu-Chieh Wu, "TSMC Continues to Use Artificial Intelligence (AI) for Trade Secret Management to Sustain Innovation and Strengthen Competitiveness," TSMC ESG, January 21, 2021.

[58] Fortune Shieh, "TSMC Trade Secret Intelligent Management 2.0: Radical Innovation for Talent Development," TSMC ESG, September 19, 2023.

Samsung aims to create an "artificial intelligence fab" that operates without human labor.[59]

Intel has been harnessing AI for critical tasks such as yield optimization, defect detection, and process control for almost two decades.[60] Data on Intel's supply chain, which spans order taking, resource procurement, manufacturing, testing, and final delivery of products are analyzed using AI to optimize business processes, drive operational excellence, and provide higher levels of customer satisfaction.[61] Intel's warehouses, for example, are highly automated, with RFID and barcode scanners tracking shipments and inventory levels. Through predictive analytics, Intel can anticipate customer needs and optimize its supply chain operations.[62]

POLICIES ON AI CHIPS AND ARTIFICIAL INTELLIGENCE

US Exports Controls on AI Chips

The United States' goal of export controls on AI chips was to restrict China's ability to both purchase and manufacture certain high-end chips that could fuel breakthroughs in artificial intelligence and give China a military advantage.[63] This is especially important in the face of China's military-civil fusion strategy that targets technologies such as quantum computing, semiconductors, 5G, nuclear technology, aerospace technology, gene editing and artificial intelligence to achieve military dominance.[64] Furthermore, China has set 2030 as its target date to become a global leader in artificial intelligence, with the subsequent goal of putting the People's Liberation Army

[59] Scott Foster, "Samsung to build all-AI, no-human chip factories," Asia Times, January 16, 2024. Also refer to the November issue of "Taiwan and the Global Semiconductor Supply Chain for more information on the U.S. – China conflict over semiconductors.

[60] Rao Desineni and Eugene Tuv, White Paper: "High-Value AI in Intel's Semiconductor Manufacturing Environment," 2022, https://www.intel.com/content/dam/www/central-libraries/us/en/documents/ai-in-semiconductor-manufacturing-paper.pdf.

[61] Craig Chvatal and Anil Varhadkar, White Paper: "Transforming Intel's Supply Chain with Real-Time Analytics," Intel, September 2017.

[62] Ibid.

[63] U.S. Department of Commerce's Bureau of Industry and Security (BIS), Press Release: "Commerce Implements New Export Controls on Advanced Computing and Semiconductor Manufacturing Items to the People's Republic of China (PRC)," October 7, 2022.

[64] Nicholas R. Licata, "China's Military-Civil Fusion Strategy: A Blueprint for Technological Superiority," Foreign Policy Research Institute, December 19, 2023.

on par with the U.S. military by 2035 — a goal the export controls intend to complicate.[65]

The U.S. Department of Commerce's Bureau of Industry and Security (BIS) implemented a series of export controls on AI chips in October 2022 and October 2023.[66] The rules in October 2023 extended export controls to chips with fewer capabilities than previously restricted in October 2022, and included additional types of chip manufacturing equipment. Chipmakers including NVIDIA, AMD, and Intel, which make the most popular chips for the AI industry, were among those affected by the rules.[67]

NVIDIA's A800 and H800 chips, which can perform complex AI tasks and are in high demand in China after the U.S.'s October 2022 export controls, were hit by the updated rules. To preserve its market share in China, NVIDIA is planning to launch the H20, L20 and L2 chips, which include most of its newest features for AI work but with computing power cut back to comply with the new rules, in the second quarter of 2024.[68]

Additionally, the BIS's "Advanced Computing Foreign Direct Product Rule" means that chip manufacturing facilities globally were restricted from manufacturing certain advanced chips for Chinese-headquartered chip design firms without a license if the manufacturer uses U.S.-origin technology or software. As the United States is a global leader in the production of semiconductor manufacturing equipment, this rule would apply to most foreign chip manufacturing firms, including TSMC, which previously produced advanced chips for Chinese AI chip design companies such as Biren Technology and Cambricon Technologies.[69]

[65] Manya Koetse, "In the race for AI supremacy, China and the US are travelling on entirely different tracks," The Guardian, January 9, 2024.

[66] U.S. Department of Commerce's Bureau of Industry and Security (BIS), Press Release: "Commerce Implements New Export Controls on Advanced Computing and Semiconductor Manufacturing Items to the People's Republic of China (PRC)," October 7, 2022.; U.S. Department of Commerce's Bureau of Industry and Security (BIS), Press Release: "Commerce Strengthens Restrictions on Advanced Computing Semiconductors, Semiconductor Manufacturing Equipment, and Supercomputing Items to Countries of Concern," October 17, 2023.

[67] Kif Leswing, "U.S. curbs export of more AI chips, including Nvidia H800, to China," CNBC, October 17, 2023.

[68] Yelin Mo and Fanny Potkin, "Nvidia to launch China-focused AI chip in Q2 2024 – sources," Reuters, January 9, 2024.

[69] Debby Wu, "TSMC Suspends Work for Chinese Chip Startup Amid US Curbs," Bloomberg, October 23, 2022.

Both Biren Technology and Cambricon Technologies, which positioned themselves as challengers to industry giants like NVIDIA and AMD, faced significant challenges, including supply chain disruptions due to U.S. export controls. Being blacklisted by the U.S. government, they no longer have access to TSMC's advanced process technologies and must rely on Chinese foundries that are years behind those of TSMC. As a result, Biren Technology and Cambricon Technologies' products have become less competitive and relevant than before. Although Biren Technology recently received a US$ 280 million cash infusion from the Guangzhou government-backed investors, a series of founder departures has raised concerns about the company's stability and future direction.[70] Meanwhile, Cambricon Technologies is anticipating a net loss ranging from RMB 756 million (US$ 105.1 million) to RMB 924 million yuan (US$ 128.5 million) for 2023, marking the seventh consecutive year of losses for the company. Its problems got worse when Huawei, a major client, started designing its own AI processors, cutting it off. After undergoing multiple rounds of layoffs in 2023, Cambricon Technologies is encountering significant obstacles in promoting the adoption of Chinese-made AI chips, and profitability remains a distant prospect.[71]

Biren Technology and Cambricon Technologies are not the only AI GPU developers in China, but their struggles reflect that of the whole Chinese AI hardware in general. To survive the new odds, Biren Technology and Cambricon Technologies will now have to redouble their efforts to seek self-sufficiency in building chip-making tools and equipment.[72]

Meanwhile, the Chinese government is pushing for semiconductor self-sufficiency as a strategic move aimed at reducing dependence on imported semiconductors, fostering domestic innovation, and positioning China as a global leader in the semiconductor industry.[73] The Chinese government has been working to convince domestic technology firms to source their inputs

[70] Anton Shilov, "China's largest AI GPU maker loses its CEO despite generous government funding — development calls into question China's GPU advances," Tom's Hardware, January 25, 2024.

[71] China Money Network, "China's Nvidia Contender Cambricon Tech Struggles to Push Chinese-Made AI Chips as It Marks 7th Consecutive Year of Losses," February 9, 2024.

[72] Ann Chao, "Tech war: US sanctions on Biren and Moore Threads strike strong blow to China's GPU champions," October 18, 2023.

[73] Refer to the December issue of "Taiwan and the Global Semiconductor Supply Chain" for more information on China's semiconductor industry.

domestically rather than from U.S. suppliers.[74] According to a report from The Wall Street Journal in October 2023, the Chinese semiconductor industry's reaction to NVIDIA's downgraded chips was cool as major Chinese players like Alibaba and Tencent are redirecting some advanced semiconductor orders to domestic companies and relying more on internally developed chips. This trend is also observed with two other major chip buyers, Baidu and ByteDance.[75]

Despite impressive progress to date, indigenous Chinese chips will likely lag in performance compared to chips from the United States and its allies for years to come. There also exists the possibility of the U.S. further tightening AI chips export controls in the near future. So far, the U.S. controls targeted at AI chips are missing a key workaround component: cloud computing.[76] A gap in export control policy allows Chinese firms to access highly advanced computers physically located in different countries through the cloud. Additionally, one early study, from the Center for a New American Security (CNAS), a Washington-based non-profit organization, said in October 2023 that thousands of controlled AI chips could have been smuggled into China in 2023.[77] It was reported in July that BIS was considering a rule change, and its solicitation for public comment suggests that an additional update to current rules may be forthcoming.[78]

CONCLUSION

The relationship between semiconductors and AI is symbiotic. While semiconductors provide the necessary hardware for AI systems, the increasing use of AI is driving the growth and innovation in the semiconductor industry. As AI continues to evolve, we can expect to see further advancements in semiconductor technology, paving the way for more powerful, efficient, and intelligent systems.

[74] Brenda Goh and Katerina Ang, "China rushes to swap Western tech with domestic options as U.S. cracks down," Reuters, October 27, 2023.
[75] Trendforce, "Cooling Response to NVIDIA's Exclusive Chips for China, Lack of Interest in Downgraded Models by Customers," January 8, 2024.
[76] William Alan Reinsch, Matthew Schleich and Thibault Denamiel, "Insight into the U.S. Semiconductor Export Controls Update," Center for Strategic and International Studies, October 20, 2023.
[77] Tim Fist and Erich Grunewald, A Working Paper: "Preventing AI Chip Smuggling to China," Center for a New American Security, October 24, 2023.
[78] Yuka Hayashi and John D. McKinnon, "U.S. Looks to Restrict China's Access to Cloud Computing to Protect Advanced Technology," The Wall Street Journal, July 4, 2023.

The global semiconductor market is forecasted to grow from 2022 to 2027. 2023 marked a pivotal moment in the growth trajectory of AI, setting the stage for rapid advancements and widespread adoption in the years to come. Consequently, compared to the global semiconductor market, the AI semiconductor market is expected to grow even more rapidly, with chips designed to accelerate generative AI workloads constituting the fastest-growing market in semiconductor sales.

In addition, tech supremacy is closely intertwined with advancements in semiconductors and artificial intelligence. The on-going race for tech supremacy between the United States and China means the prospect of U.S. further tightening its export restrictions on high-performance computing chips to China and select locations is still hovering in the horizon, and that chip company revenues and growth could be negatively impacted. Furthermore, responses from restricted countries could also adversely affect companies, industries, and the global economy.

The export controls imposed by the United States on AI chips will have far-reaching implications for the Chinese advanced semiconductor sector. Among the most affected will be the Chinese AI data center companies, which rely heavily on cutting-edge chips for their operations. While these export controls will significantly harm Chinese AI data center companies in the short term, they will add new impetus for China to strengthen its domestic chip design ecosystem in the long run.

In a nutshell, the integration of AI and semiconductors has profound implications on both economic and military development, and the opportunities and challenges posed by AI and the semiconductor industry are of great interest to governments and policymakers worldwide. AI is not only transforming chip design and manufacturing but also reshaping the entire semiconductor industry landscape and unlocking new possibilities and economic value.

*****

TaiwanPlus

- **Curious about Taiwan?**
- **Planning a visit?**
- **Want to stay up to date on regional politics?**

TaiwanPlus is the premier international media platform providing English-language news and entertainment offering independent and impartial daily news from Taiwan. From its unique position at the nexus of geopolitics and international trade, TaiwanPlus also provides an inside look at Taiwan-China relations, with in-depth reporting and analysis, and showcases Taiwan's unique culture and lifestyle, offering some of the nation's best food, music, and travel programming. For more information, visit https://www.taiwanplus.com/

TaiwanPlus Videos on Semiconductors

2024/02/06
Taiwan's Mega Factories

Inside Micron Taiwan's Semiconductor Factory - Taiwan's Mega Factories EP1

In the first episode of Taiwan's Mega Factories, join us for a tour of Micron Technology's Taiwan chip manufacturing facilities to discover how chips are produced and how innovations in the semiconductor industry impact people and businesses all over the world. Micron leaders share how they are expanding their innovative approach into the area of resource management, enabling them to create more sustainable products.

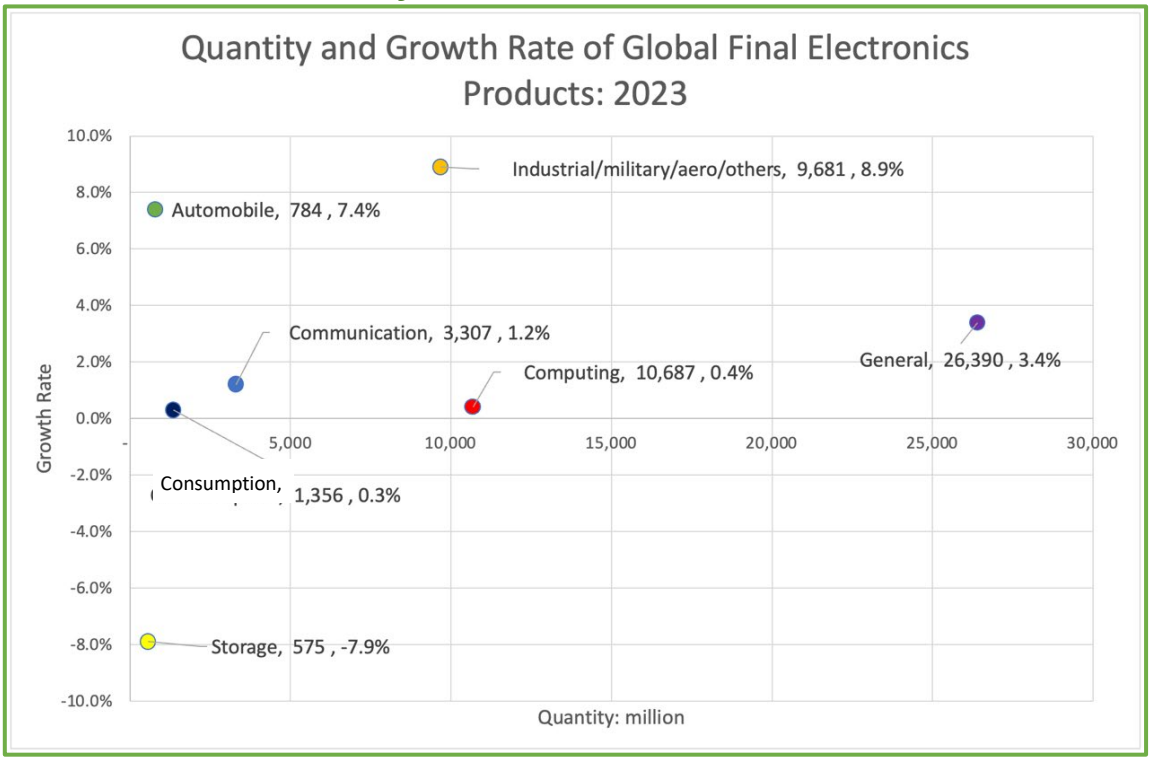
*****

SEMICONDUCTOR STATISTICS AT A GLANCE

GLOBAL TRENDS

End-User Products and Semiconductor Application

Figure 3: Worldwide Shipment Volume and Growth Rate of End-User Electronics Products: 2023



Source: Hui-Hsiu Huang, "Advanced Process Technologies Driving End- Application Innovation and Market Competition," IEK, ITRI, November 10, 2023, p. 1.

Semiconductor chips are deployed across a wide range of applications, transforming various sectors and industries. Figure 3 shows the worldwide shipment volume and growth rate of end-user electronics products in 2023.

According to Gartner's forecast in October 2023, the global shipment of *general electronics products* was the largest, reaching 26.4 billion units in 2023, representing an annual growth rate 3.4%. *Computing electronic products* at 10.7 billion units, accounted for the largest shipment, followed by the category of *industrial/military/aviation/other electronic products*, with shipments reaching 9.68 billion units in 2023.

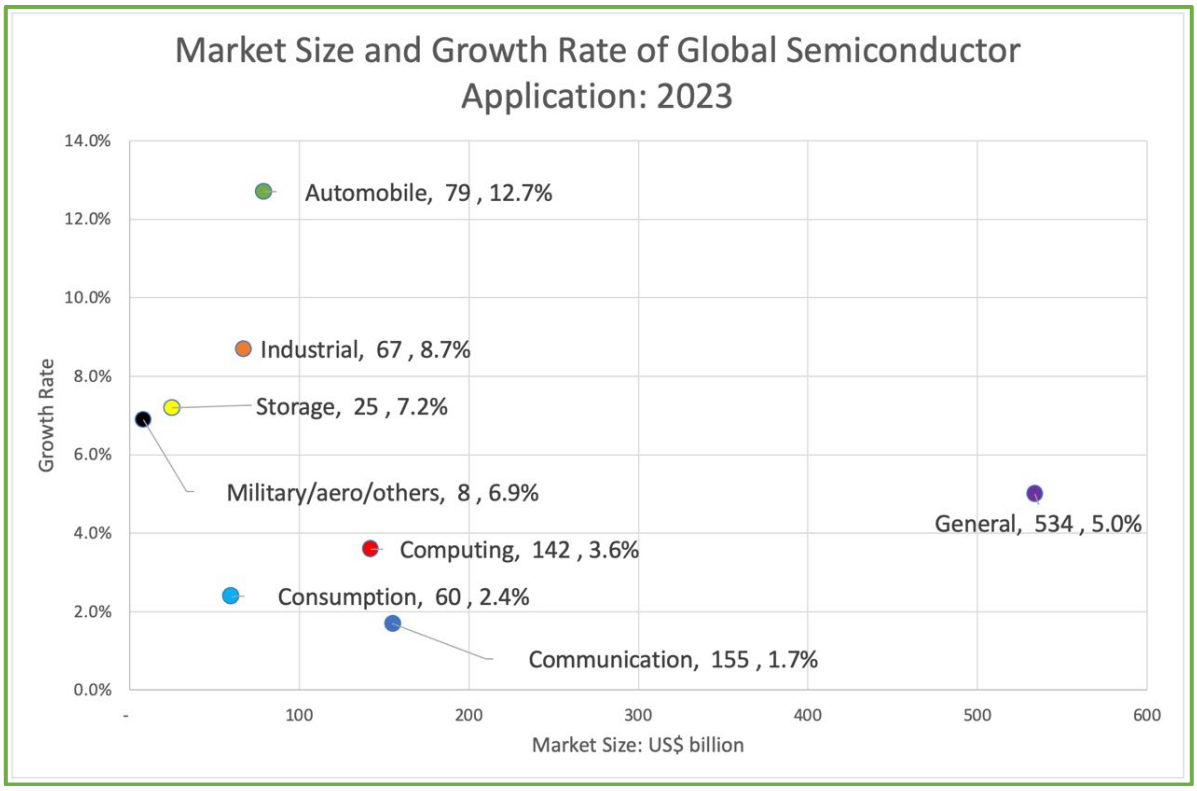
In terms of annual growth rate of end-user electronics products in 2023, the category of *industrial, military, aviation/other electronic products* had the highest annual growth rate, reaching 8.9%. Such electronic products require advanced semiconductors that provide both high performance and high reliability.

Coming in second to the category of *industrial/military/aviation/other electronic products* was *automotive electronic products* with an annual growth rate of 7.4%. Semiconductors are used in automotive electronic products in areas such as automobile safety, control and entertainment. With the improvement of automobile intelligence and electrification, the demand for automotive electronic products, and by extension, semiconductors, will continue to grow.

Other end-user electronic products, including *communications, computing*, and *consumer electronics products* witnessed positive annual growth rates in 2023. The category of *storage electronic products* was the only exception, with worldwide shipment of storage electronic products shrinking by 7.9% compared to the previous year. Likewise, the memory semiconductors industry, also faced a significant downturn in 2023.

Overall, the global end-user electronics market still maintained growth in 2023. In general, although the global semiconductor market was affected by factors such as the global economic downturn in 2023, it still showed positive growth. With the continued growth of high-end electronic products and the automotive semiconductor markets, the long-term growth trend of the global semiconductor market looks promising.

Figure 4: Global Market Size and Growth Rate of Semiconductor Application: 2023



Source: Hui-Hsiu Huang, "Advanced Process Technologies Driving End- Application Innovation and Market Competition," IEK, ITRI, November 10, 2023, p. 6.

Figure 4 shows the global market size and growth rate of semiconductor applications. According to data from Gartner in October 2023, the global semiconductor market size was estimated to be US$ 534.5 billion in 2023, a decrease of 10.9% compared to 2022. Of the products where semiconductors are deployed, the category of *communication products* is the largest market at US$ 154.6 billion, followed by *computing products* at US$ 142 billion, accounting respectively for 28.9% and 26.6% share of the global market.

In terms of market size, the automotive semiconductor market, valued at US$ 79 billion, is the largest. In fact, major players including NXP Semiconductors, which provide microcontroller and analog chips for various automotive applications, and Tesla, which is known for its electric vehicles and self-driving technology, have launched automotive processors using advanced processes to meet the increasingly complex automotive application needs. These advanced process automotive processors have higher performance, lower power consumption, and can provide safer and more advanced functions for cars.

It is also worth noting that the compound annual growth rate (CAGR) of the automotive semiconductor market from 2022 to 2027 is 12.7%, which is higher than the CAGR of the overall semiconductor market of 5.0%. It is the only market in the semiconductor application field to show double-digit growth in 2023, making it a significant driver of technological advancement in the automotive industry. The integration of semiconductors with AI and other cutting-edge technologies is fueling this growth, leading to innovations in safety, efficiency, and vehicle electrification. As the automotive sector continues to evolve, semiconductors play a pivotal role in shaping the future of mobility. With the improvement of automotive electronics and advances in artificial intelligence, it is foreseeable that the demand for automotive semiconductors will continue to grow in the future.

*****