

Unveiling the threads behind the global AI tech race

Background

The launch of OpenAI's ChatGPT has brought the semiconductor industry under the spotlight, resulting in a year of particularly high hype from investors, with constant stock spikes and new technological standards. With generative AI software gaining popularity over the past year, chipmakers' performance has begun to catch the attention again as markets grow confident in the companies' future performance. Open AI, one of the pioneers in large language models (LLMs), has nearly tripled its valuation in less than 10 months after completing a tender offer that valued the Microsoft-backed company at approximately \$80bn. Simultaneously, California-based NVIDIA [NASDAQ: NVDA], the primary GPU supplier to AI companies for the past two years, has seen its stock price surge almost 300% in the last year and closed 2023 as the best-performing stock in the S&P 500. Moreover, it has recorded a further price increase of over 30% since it announced its earnings on February 21st, 2024, which surpassed analysts' expectations. However, NVIDIA was not the only company affected by the semiconductor enthusiasm: another US chipmaker, AMD [NASDAQ: AMD], experienced a bright 2023, where its stock increased 128%.

These record valuations mirror investors' interest and eagerness for microprocessor designers as the implications of stronger chips continue to develop. Although GPUs were originally designed to play computer games, they have impacted various industries, from healthcare, where they are used to diagnose diseases with medical image analysis, to finance, where they power algorithmic trading systems and enable quick decision-making through real-time market data analysis. However, one of the industries where new improvements could potentially disrupt geopolitics is the national defense industry, where powerful semiconductors could lead to the integration of quantum technology and potentially revolutionize autonomous decision-making, threat assessment, and predictive maintenance of military equipment and logistics.

In this imminent AI revolution scenario, semiconductors play a crucial role. A semiconductor is a material product built with silicon - properties of which can be altered by introducing impurities to meet the needs of the specific electronic component in which it resides. Semiconductors are used to create central processing units (CPUs) and graphic processing units (GPUs), both of fundamental importance in processing tasks in order and parallel, respectively. The number of transistors on a chip impacts its speed. Thus, the industry aims to create smaller, faster, cheaper products. The pressure of constant improvement is reflected in the large research and development budgets, with an industry total of almost \$110bn in 2022.

Due to the many applications of high-end microchips, the political implications of being ahead in this race have brought this industry to the center of a technological rivalry between the US and China. Indeed, despite the US comprising 25% of total world semiconductor demand, its manufacturing capacity has decreased from 37% in the 1990s to just 12% in the 2020s. Given China's efforts to gain a foothold in this industry, this resulted in concerns of a national security threat. This article unveils these implications and the intricacies of an industry enabled by a global network involving major players.

Supply chain complexity

When analyzing the global artificial intelligence race, it is hard to avoid the topic of supply chain complexity. The journey – that starts with transistors and ends with LLMs – is one to unravel with great care and attention to inner connections between the upstream and downstream producers.



It all begins with chip-making equipment manufacturers that allow microchip production to commence, where ASML [NASDAQ: ASML], a Dutch lithography firm, is king. Europe's biggest public tech firm, valued at \$233.9bn, owns patents to advanced machines using "super fine extreme ultraviolet light" (EUV). Such technology allows it to work at very tiny scales and create complex patterns of transistors on the chips. The tech is one of a kind: it "can print lines on semiconductors 8 nanometers thick [...]. The thinner the lines, the more transistors you can fit on a chip. And the more transistors you can fit on a chip, the higher the processing speeds and memory". TSMC[NASDAQ: TSMC], Intel and Samsung are amongst ASML's largest customers, which makes it an exceptionally powerful company. ASML's share price has increased 51% year-over-year. Tokyo Electron [TYO: 8035] and Applied Materials [NASDAQ: AMAT], two competitors to ASML, have seen their shares jump 146.1% and 78.6% YoY, respectively.

Arm, another key player, focuses on chip designs. Recently, it gained plenty of media attention for unbelievable stock price growth: since its IPO in mid-September, shares soared 117% in value. The UK-based company is generally assumed to be in the "AI stocks"; however, the distinction remains unclear. While AI was mentioned 19 times in the firm's third-quarter investor presentation, there is room for misunderstanding. Arm gets its revenues from royalties, defined at 5% of the chip price, and refuses to discriminate by price. Despite Arm being a leader in the chip design space, the revenue growth behind the shares' surge does not necessarily come from boosted demand for AI. Arm's designs are already written into Apple, Microsoft, TSMC, Intel, NVIDIA, Amazon, Alphabet and Samsung products. In these upstream firms, the designs are used in smartphones, tablet computers, laptops, and other portable devices. Hence, one can quickly realize that perhaps strong growth in share prices comes from expected future demand for AI and Arm's designs. Interestingly, the aforementioned firms had all volunteered to anchor Arm's IPO, and NVIDIA tried to buy Arm for \$40bn before being blocked by FTC in 2020. The investment by strategic partners highlights the nature of the tight-knit AI companies' circle and the mutual benefits they reap from staying competitive this way.

Transitioning to physical goods directly connected to AI, the next step is the production of microchips itself. Of producers, TSMC comes to mind first – the Taiwan-based firm has been at the forefront of integrated circuit production for years and has recently been expanding quite rapidly. A Kumamoto, Japan plant is currently under construction, Dresden and Phoenix locations are confirmed and to be under construction soon. TSMC is the biggest contract chipmaker worldwide, which means that, unlike NVIDIA, AMD or Intel, it produces chips using clients' designs. The company, commanding the most state-of-the-art chip production and controlling 85% of the global market for most advanced made-to-order chips, is instrumental to the supply chain. Peter Hanbury, semiconductor specialist at Bain & Co., mentioned: "They [TSMC] basically control the most complicated part of the semiconductor system, and they're a near monopoly at the bleeding edge". While TSMC uses ASML's EUVs, they have been actively investing in R&D to invent a replacement technology to cut costs. TSMC's customers include the likes of AMD, Apple and NVIDIA. Moreover, C-suite executives of TSMC and NVIDIA are in close contact and have met before to discuss AI chip supply constraints, the former being a key provider of inputs for the latter. The relationship again highlights the symbioses within the AI supply chain and tech companies leveraging these to extract maximum value. TSMC's shares have surged 48.7% YoY.

The penultimate shelf in the supply chain is taken by GPU producers. Of these, the main AI boom benefiter is NVIDIA, which has recently taken a dominant position amongst competitors. Based in California, this firm held a whopping 90% market share of the global AI chip market as of January 26th, 2024. The features implemented in NVIDIA GPUs' chips, systems and software are ideal for efficient, high-performance machine learning, clearly diversifying the company from the rest of the industry. This Friday, 23rd of February, it became the third most valuable tech company after crossing a valuation of \$2tn — accumulating a YoY growth of 281.6%.



While start-ups such as Cerebras, Groq and Graphcore are trying to challenge the tech giant, NVIDIA has been quick to respond to adversities on its way. Last year, it intensified investments in upstream AI companies – those utilizing its GPUs – and participated in 35 deals. These included two of the biggest OpenAI's ChatGPT competitors, Inflection and Cohere. While executives denied that NVIDIA's portfolio companies get preferential access to its GPUs, it is not hard to see why such activity could raise eyebrows. Once again, this is an example of mutually beneficial support within the AI sphere.

Closing the chain are LLMs. Firms specialized in LLMs create programs that feed enough examples to develop neural networks that recognize and interpret human language and other data, a process otherwise referred to as machine learning. ChatGPT, the most famous LLM, has likely impressed every first-time user; combined with the multitude of applications of AI, the reasons for the artificial intelligence boom become apparent. LLMs can be trained to be extremely efficient at particular tasks when fed enough data. For example, AI is already transforming the future of farming: John Deere's R&D resulted in fully automated tractors that can cultivate, fertilize, harvest and plant with minimal intervention. But perhaps a more general application would be predictive analytics. Of course, AI can be used to predict crop growth and harvest, but if it was instead trained for the legal industry, it could streamline firms' operations and optimize resource allocation. Efficiency uses of AI are nearly limitless, and more companies will be seen leveraging this opportunity to cut costs. Hence the boom in demand for AI, GPU producers, chip designers and other downstream firms' products.

Political chessboard

Considered the brain of modern technology, microprocessor production relies on a complex and highly specialized supply chain spanning multiple countries worldwide. This global network is not just a logistical wonder but also a nexus of political implications. It is a battleground for international supremacy, where technological leadership equates to economic power, strategic advantage, and national security. More importantly, the international nature of this value chain means that disruptions in one part of the world can have cascading global effects, highlighting the fragility and interdependence of nations in high technology.

As previously mentioned, Taiwan stands out as one of the crucial players in the global semiconductor supply network, mainly due to the Taiwan Semiconductor Manufacturing Company. TSMC dominates the industry with around 85% of the market share in semiconductor production, followed by Samsung Electronics with about 10%. This grants Taiwan a unique geopolitical leverage called the "Silicon Shield." The "Silicon Shield" concept highlights the country's deterrent against aggressive actions by powers such as the United States and China. Maintaining a stable relationship with Taiwan is essential for both nations because any significant disruption to TSMC's operations would have a domino effect on the global technology sector. This would impact everything from consumer electronics to critical infrastructure and defense capabilities.

In this context, the United States and China have each developed different strategies to reduce vulnerabilities in response to the geopolitical landscape and the importance of steady supply. The U.S. aims to reduce dependence on foreign semiconductor production and has initiated efforts to bring chip manufacturing back to its soil. The country is incentivizing chipmakers to bolster domestic production facilities and national technological sovereignty. In its latest move in 2023, Intel bought a \$275m state-of-the-art lithography machine from ASML to return to the forefront of semiconductor development. Additionally, the U.S. constantly reinforces its military alliance with Taiwan. It has imposed restrictions on exports of advanced chipmaking equipment to rival countries to limit their advancement capabilities. One recent example is the US pressure on the Netherlands and Japan to end the sale of advanced silicon chipmaking equipment to China last year. After months of talks, The Hague agreed to implement export restrictions, hitting the sales of ASML.



On the other hand, there is a widespread belief that China could launch an invasion of Taiwan to strengthen its claim to the island as part of its territory and gain control over a vital hub in the AI supply chain. For this reason, China has been diversifying its financial approaches to minimize vulnerability and prepare for possible economic sanctions if the invasion happens. The country is conducting international trade negotiations in its currency, Yuan, to reduce reliance on the US dollar. Reducing dependence on the dollar and increasing the circulation of the yuan are two of the main lines of action of the Xi Jinping government's foreign policy. "This is an attempt to circumvent the structural power of the dollar in the international monetary system, which gives the United States, in practice, the power to interfere in the maneuvering radius of other States by exploiting dependence on the currency or possibly imposing sanctions," explains Isabela Nogueira, professor at the UFRJ Institute of Economics and coordinator of LabChina. This was demonstrated in the War in Ukraine, with the so-called "dollar bomb". Since the invasion of Ukrainian territory, Western states have imposed economic sanctions on Russia as a way of trying to contain the conflict. The blocking of US\$300 billion of the total US\$640 billion in Russian international reserves is a warning to China. The country is also selling off its U.S. Treasury bond reserves and significantly increasing its gold reserves. Overall, these measures are part of a broader strategy of both nations to guarantee technological dominance and limit rival nations' powers.

Currently, the United States dominates the race in the semiconductor and artificial intelligence sectors. The US is estimated to be around 24 months ahead of China in terms of development. Several key factors contribute to this advantage, including major industry players such as Microsoft, ARM and NVIDIA, who command a significant global market share. The annual investments by big tech of approximately \$162bn in processing capacity are also in stark contrast to China's \$9bn. On the other hand, China's development strategy in advanced computing components benefits from its more centralized authoritarian governance. Despite being counterintuitive, this structure allows for swift decision-making and targeted incentives in specific sectors, facilitating rapid development and implementation. The Chinese approach emphasizes balancing economic growth and political stability, focusing on cyber sovereignty, "national harmony," and maintaining the central party's power. Meanwhile, Western countries prioritize developing AI applications that champion decentralization and personal autonomy.

Outlook

A strong start for semiconductor companies in 2024 invites an investigation of the potential development of the industry, beginning with their high stock valuations. Take a look at Arm Holdings, whose impressive growth was mentioned previously. In its latest results, the company reported adjusted earnings per share (EPS) of \$0.29, compared to the expected \$0.25 predicted by analysts, and a revenue of \$824mn, against analysts' predictions of \$761mn. However, Arm's net income was \$87mn, far less than the \$182mn posted in the prior year. Although it is compared to NVIDIA as it might be the next trillion-dollar company, the latter has far stronger financials.

Arm's multiples are not the only ones that could be potentially worrying going forward. As of February 22nd, 2024, NVIDIA's P/E ratio was above 102x, above the industry average of 59.9x but far below Arm's P/E ratio of 1,456x. As their stock prices surge, so does the concern that this might be the next asset bubble. Chip providers have notoriously been victims of market hype and overly optimistic forecasts. Technological innovation like AI, with the potential to profoundly change the world, is a typical feature of asset bubbles. Obviously, not all semiconductor companies are equal in terms of value created and potentially deserving their valuations. For example, Super Micro Computer Inc [NASDAQ: SMCI], a company tangent to AI, saw its share price soar on February 16th, reaching almost a record \$1,078, before a rapid 20% decrease on the same day, closing at around \$803. Since then, prices have fluctuated, and SMCI closed above \$975 on February 22nd highlighting the high implicit volatilities. While



such price fluctuations indicate that this industry is susceptible to investors' sentiment, companies' financials suggest that the race for AI is set to fuel growth in a robust near-term momentum.

Lastly, 2024 is an important year for elections worldwide, some of which could affect the semiconductor industry's future. In Taiwan, the January 2024 election ended with the victory of the Democratic Progressive Party, which currently has no diplomatic relations with China. The outcome of this election could hurt the supply of semiconductors due to the probable increase in naval drills in the Taiwan Strait, airspace violations, and missile firings over Taiwan during senior US officials' visits or vice versa, as well as due to a possible invasion of Taiwan by China or damages to its submarine cables. Another election that could impact the current outcomes of this race is the US presidential elections in November and India's national elections. Prime Minister Modi, who has a high chance of serving his third term, is committed to developing India's semiconductor sector, which might positively change the global supply.

TAGS: AI, Semiconductor, NVIDIA, Arm, TSMC, Technology, Politics