Deloitte.
Insights

ARTICLE • 18-MIN READ • 04 FEBRUARY 2025 • Deloitte Center for Technology, Media & Telecommunications

2025 global semiconductor industry outlook

Chip sales are set to soar in 2025, led by generative AI and data center build-outs, even as demand from PC and mobile markets may remain muted

The semiconductor industry had a robust 2024, with expected double-digit (19%) growth, and sales of US$627 billion for the year.[1] But that's even better than the earlier forecast of US$611 billion.[2] And 2025 could be even better, with predicted sales of US$697 billion,[3] reaching a new all-time high, and well on track to reach the widely accepted aspirational goal of US$1 trillion in chip sales by 2030. This suggests the industry only needs to grow at a compound annual growth rate of 7.5% between 2025 and 2030 (figure 1).[4] Assuming the industry continues to grow at that rate, it could reach US$2 trillion in 2040.

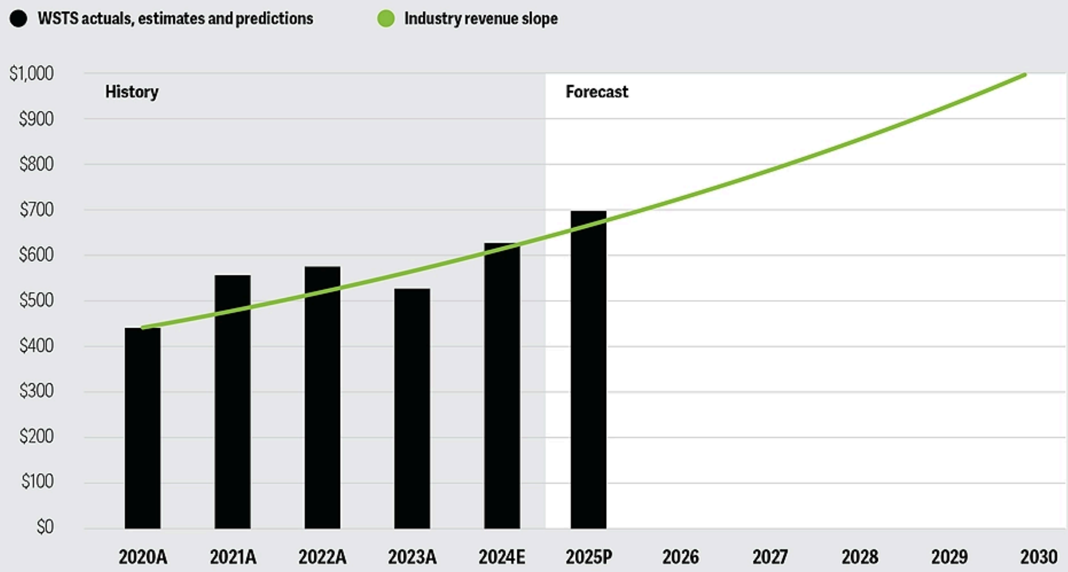
The stock market is often a leading indicator of industry performance: As of mid-December 2024, the combined market capitalization of the top 10 global chip companies was US$6.5 trillion—up 93% from US$3.4 trillion in mid-December 2023 and 235% higher than the US$1.9 trillion seen in mid-November 2022.[5] That said, it is worth noting that "average" chip stock performance in the last two years has been a "tale of two markets": Companies involved in the generative AI chip market outperformed that average, while companies without that exposure (automotive, computer, smartphone, and communications semiconductor companies, for example) underperformed.[6]

One driver of industry sales has been the demand for gen AI chips: a mix of CPUs, GPUs, data center communications chips, memory, power chips, and more. **Deloitte's 2024 TMT Predictions report** predicted that those gen AI chips collectively would be worth "more than" US$50 billion,[7] which was a much too conservative forecast, as the market was likely worth over US$125 billion in 2024—and represented over 20% of total chip sales for the year.[8] At the time of publication, we predict that gen AI chips will be over US$150 billion in 2025.[9] Further, Lisa Su, chief executive officer at AMD, moved her estimate for the total addressable market for AI accelerator chips up to US$500 billion in 2028—a number larger than sales for the entire chip industry in 2023.[10]

Figure 1

Revenues indicate the possibility of the chip industry hitting US$1 trillion in 2030

The path to $1 trillion in semiconductor revenues ($Billions)



Note: A = Actual, E = Estimate, P = Prediction.
Source: Deloitte analysis and extrapolation based on data from World Semiconductor Trade Statistics.

Deloitte. | deloitte.com/us/en/insights/research-centers/center-for-technology-media-telecommunications.html

In terms of end markets, after being flat at around 262 million units over 2023 and in 2024, PC sales are expected to grow in 2025 by over 4% to about 273 million units.[11] Meanwhile, smartphone sales are expected to grow at low single digits in 2025 (and beyond) to reach an estimated 1.24 billion units in 2024 (6.2% year-over-year growth).[12] These two end markets are important for the semi industry: In 2023, communication and computer chip sales (which include data center chips) made up 57% of overall semiconductor sales for the year compared to auto and industrial (which accounted for only 31% of sales combined, for example).[13]

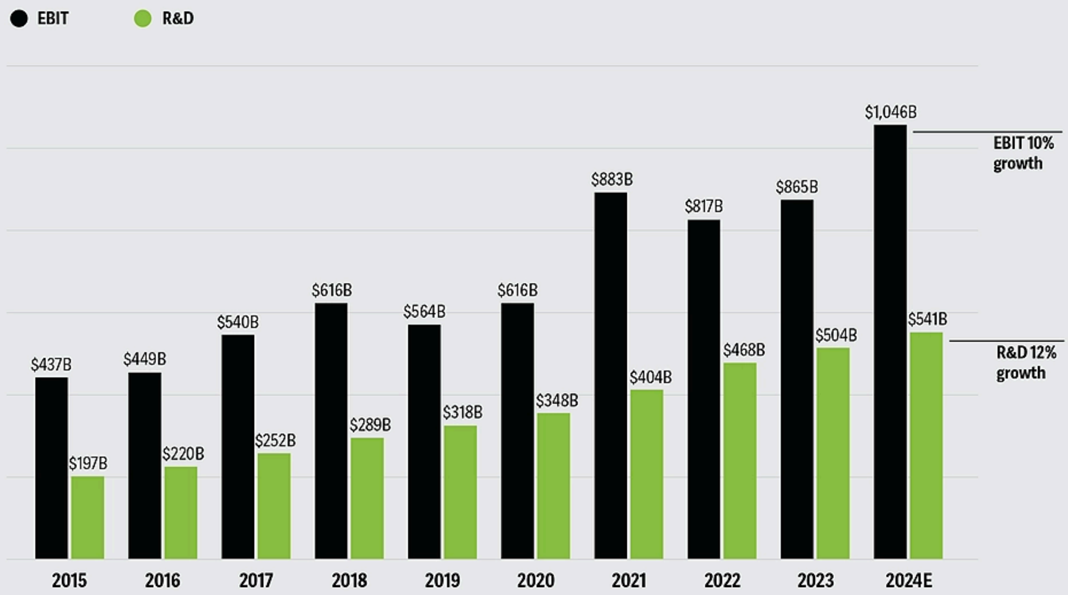
One challenge for the industry is that while gen AI chips and associated revenues (memory, advanced packaging, communications, and more) are responsible for outsized revenues and profits, they represent a small number of very high-value chips, meaning that wafer capacity—and therefore utilization—for the industry as a whole isn't as high as it might appear. In 2023, nearly a trillion chips were sold at an average selling price of US$0.61 per chip.[14] At a rough estimate, although gen AI chips might account for 20% of revenues in 2024, they were less than 0.2% of total wafers.[15] Even though global chip revenues for 2024 were forecast to rise 19%, silicon-wafer shipments for the year actually *declined* an estimated 2.4% for the year.[16] That number is expected to grow by almost 10% in 2025, fueled by demand for components and technologies used largely in gen AI chips, such as chiplets, as mentioned in our 2025 TMT Predictions report.[17] Of course, silicon wafers are not the only kind of capacity to track: Advanced packaging is growing even faster. As an example, some analysts estimate that TSMC's CoWoS (chip-on-wafer-on-substrate) 2.5D advanced packaging production capacity will reach 35,000 wafers per month (wpm) in 2024 and could increase to 70,000 wpm (100% year-over-year) and further by 30% year-over-year to 90,000 wpm by end of 2026.[18]

Further, driving innovation in the industry is not cheap. In 2015, the overall chip industry average for spending on research and development was 45% of its earnings before interest and taxes (EBIT), but by 2024, it was an estimated 52% of the same.[19] R&D seems to be growing at a 12% CAGR, white EBIT is only growing at 10% (figure 2).[20]

Figure 2

The semiconductor ecosystem now invests over half of its EBIT in driving innovation

R&D and EBIT growth, semiconductor ecosystem, 2015–2024E



Note: EBIT = Earnings before interest and taxes; E = Estimate.

Source: Graphic prepared by Deloitte based on ASML Investor Day presentation, Pg. 8 of "Small Talk 2024: Global market trends, Industry & ASML's technology roadmap, ESG," November 14, 2024. Analysis based on ASML Corporate Marketing (CMKT) analysis and company reports.

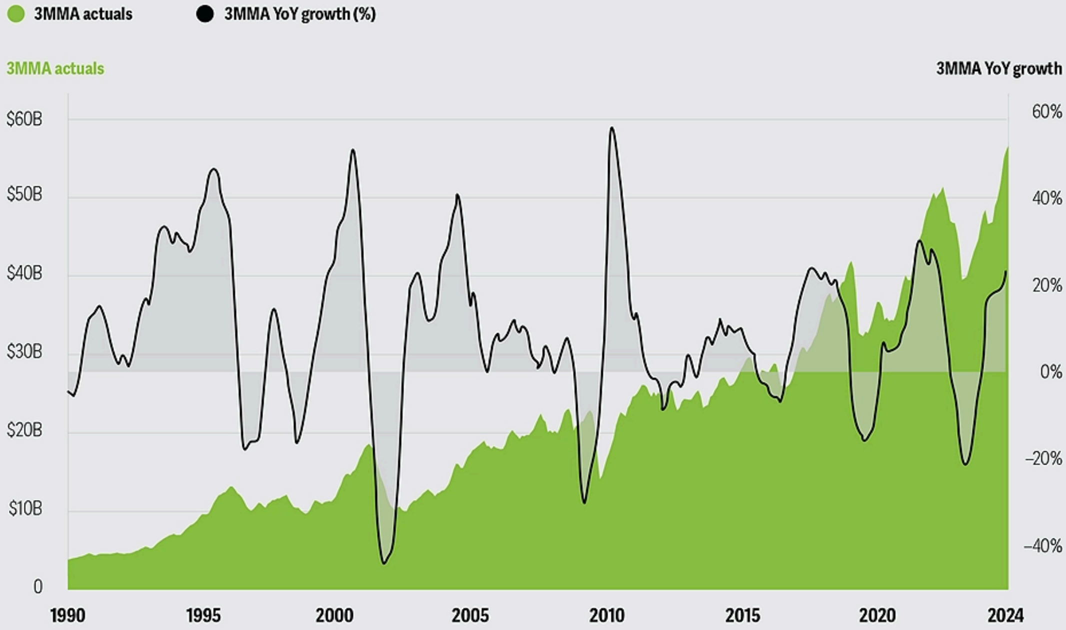
Deloitte. | deloitte.com/us/en/insights/research-centers/center-for-technology-media-telecommunications.html

Finally, it’s worth reminding readers that the chip industry can be notoriously cyclical. The industry has flipped from growth to shrinkage nine times in the last 34 years (figure 3).[21] So it may seem that the industry is seeing less extreme growth or shrinkage in the last 14 years, compared to 1990 to 2010, but the frequency of contractions seems to have increased. The year 2025 looks solid for now, it’s hard to tell what 2026 will bring.

Figure 3

The global chip industry has been extremely cyclical, flipping up and down over the past 34 years

Global semiconductor industry—historic billings monthly



Note: 3MMA = Three-month moving average; the average YoY growth during this period was 10.5%.
Source: World Semiconductor Trade Statistics Historical billings report as of October 2024.

Deloitte. | deloitte.com/us/en/insights/research-centers/center-for-technology-media-telecommunications.html

These trends and others play into our 2025 semiconductor industry outlook, where we drill down into four big topics for the year ahead: generative AI accelerator chips for PCs and smartphones and the enterprise edge; a new "shift-left" approach to chip design; the growing global talent shortage; and the need to build resilient supply chains amid escalating geopolitical tensions.

About Deloitte's TMT center outlooks

Deloitte's 2025 global semiconductor industry outlook seeks to identify the strategic issues and opportunities for semiconductor companies and other parts of the semiconductor supply chain to consider in the coming year, including their impacts, key actions to consider, and critical questions to ask. The goal is to help equip companies across the semiconductor ecosystem with information and foresight to better position themselves for a robust and resilient future.

Generative AI chips in PCs, smartphones, the enterprise edge, and the Internet of Things

Many of the chips that are being used for training and inference of gen AI cost tens of thousands of dollars and are destined for large cloud data centers. In 2024 and 2025, these chips or lightweight versions of these chips are also finding homes in the enterprise edge, in computers, in smartphones, and (over time) in other edge devices such as IoT applications. To be clear, in many cases, these chips are being used for either gen AI, traditional AI (machine learning) or, increasingly, a combination of both.

The enterprise edge market was already a factor in 2024, but the question in 2025 will be about smaller, cheaper, less powerful versions of these chips becoming a key part of computers and smartphones. What they lack in per-chip value, they can make up for in volume: PC sales are expected to be over 260 million units in 2025, while smartphones are expected to be over 1.24 billion units.[22] Sometimes, the "gen AI chip" can be a stand-alone single piece of silicon, but more commonly it's a few square millimeters of dedicated AI processing real estate that is tiny part of a much larger chip.

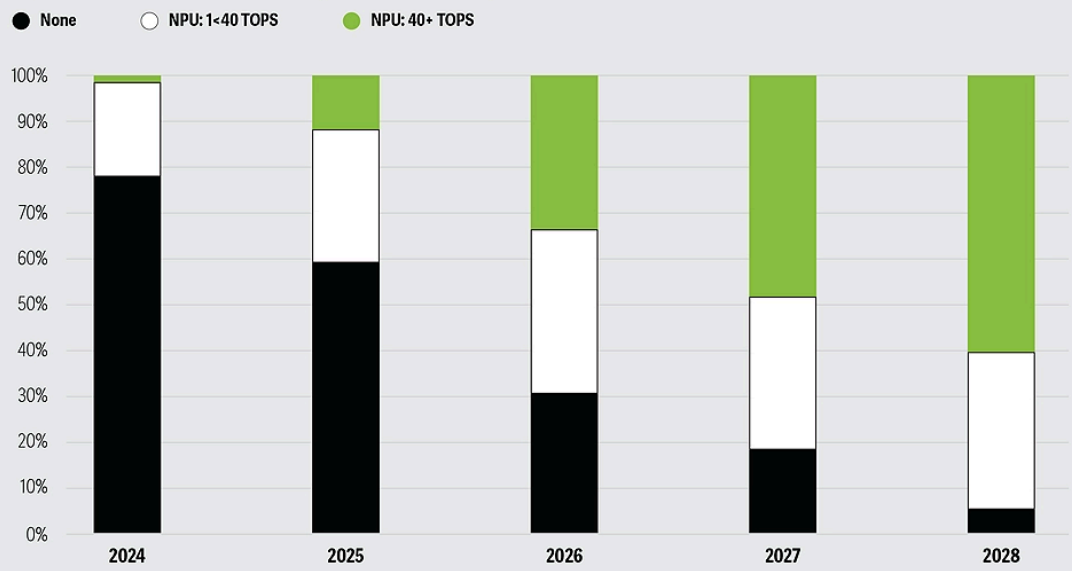
Enterprise edge: Although gen AI via the cloud will likely continue to be a dominant option for many enterprises, about half of the enterprises worldwide are predicted to add AI data-center infrastructure on-premises—an example of enterprise edge computing.[23] This could be, in part, to help protect their intellectual property and sensitive data and comply with data sovereignty or other regulations, but also to help them save money.[24] These chips are largely the same as those found in hyperscale data centers, with server racks costing millions of dollars and requiring hundreds of kilowatts of power. Although smaller than hyperscale chip demand, we estimate the chips for enterprise edge servers will likely be worth tens of billions of dollars globally in 2025.[25]

Personal computers: Sales of gen AI–powered PCs are predicted to comprise half of all PCs in 2025,[26] with some forecasts suggesting that almost all PCs will have at least some onboard gen AI processing—also known as neural processing units (NPUs)—by 2028 (figure 4).[27] These NPU-powered machines are expected to command a price premium of 10% to 15%,[28] but it's important to note that not all gen AI PCs are equal. There's a dividing line at the 40 TOPS (trillion operations per second) level, following a recommendation from major PC ecosystem companies that only computers with more than 40 TOPS be considered true AI-enabled PCs.[29] As at the time of writing, some buyers are cautious about these new PCs, either unwilling to pay the premium, or waiting until more powerful gen AI NPUs are introduced in the second half of 2025.[30]

Figure 4

Almost all PCs are expected to have some gen AI processing—also known as neural processing units (NPUs)—by 2028

Expected sales of NPU-enabled PCs of all PCs worldwide from 2024 to 2028



Note: TOPS = trillion operations per second.
Source: International Data Corporation Worldwide Quarterly Personal Computing Device Tracker, September 2024.

Deloitte. | deloitte.com/us/en/insights/research-centers/center-for-technology-media-telecommunications.html

As of December 2024, many of the installed base of PCs were running on x86 CPUs, with the balance being on CPUs based on the Arm architecture. MediaTek, Microsoft, and Qualcomm announced in 2024 that they would make Arm-powered PCs, specifically gen AI PCs.[31] It's unclear how successful these machines will be in the next 12 months, but it will likely be a key issue for the various chipmakers, with Qualcomm predicting it will sell US$4 billion worth of PC chips annually by 2029.[32]

Smartphones: Where PC NPUs might be worth tens of dollars in value, smartphone-equivalent gen AI chips may be worth much less, and we estimate the silicon on next-generation smartphone processors to be under US$1.[33] Even though the smartphone market is over a billion units sold annually, and even though we predict gen AI smartphones will be 30% of phones sold in 2025,[34] the semiconductor impact is likely smaller than PCs in dollar terms. Instead, an interesting angle for chipmakers could be to see if consumers are excited enough about new gen AI phones and features to shorten the replacement cycle. Consumers have been keeping phones longer before upgrading, and sales have been flat for years now.[35] If gen AI enthusiasm causes an uptick in smartphone sales, it could benefit all kinds of chip companies, not just those that make the gen AI chips themselves.

IoT: A gen AI chip in a data center might cost US$30,000. A gen AI chip on a PC might cost US$30. A gen AI chip on a smartphone might be US$3. For gen AI chips to work on the low-cost IoT market, they should cost about US$0.3. That's unlikely to happen anytime soon, but with the possibility of tens of billions of IoT endpoints needing AI processors, this is a market to watch for the longer term.

Strategic questions to consider

- Although gen AI chips for data centers are in demand now, given their importance to industry growth, are there any signs that demand is weakening, or that processing is moving away from data centers to edge devices?
- Given the success of gen AI chips in data centers, the market potential for various edge chips may drive mergers and acquisitions and attract further private equity, venture capital, and sovereign wealth fund interest: Chip companies are already aligning with financial players. Could we see more of this in 2025?
- Some analysts expect the market for gen AI inference to grow faster than training in 2025 and beyond: What implications could this have on various semiconductor sectors and players? As the cost of AI inference reduces at a faster pace, how can it affect semiconductor chips?
- With greater focus on sustainability and the heightening stress on power consumption due to an AI-driven surge in electricity demand, how can the industry strike a balance between power efficiency and performance in small form factors including laptops, mobile phones, and IoT devices?

Chip design 'shifts left' and calls for a greater collaboration across the industry

Deloitte predicted that, by 2023, AI would emerge as a powerful aid to human semiconductor engineers, assisting them on extremely complex chip-design processes, and enabling them to find ways to improve and optimize PPA (power, performance, and area).[36] As of 2024, gen AI has enabled rapid iterations to enhance existing designs and discover entirely new ones that can do it in less time.[37] In 2025, there will likely be more emphasis toward "shift left"—an approach to chip design and development where testing, verification, and validation are moved up earlier in the chip design and development process[38]—as optimization strategies could evolve from simple PPA metrics to system-level metrics like performance per watt, FLOPs (or "floating point operations per second") per watt, and thermal factors.[39] And the combination of advanced AI capabilities—graph neural networks and reinforcement learning—will likely continue to help design chips that are more power-efficient than typical chips produced by human engineers.[40]

Domain-specific and specialized chips are expected to continue to gain prominence over general-purpose ones, as several industries (such as automotive) and certain AI workloads would require customized approaches to designing chips.[41] However, a widespread adoption of application-specific integrated circuits[42] remains less clear, as the development and maintenance of such hardware can be costly and could divert focus from other AI advancements.[43] But here's where gen AI tools can allow companies to design more specialized and competitive products including custom silicon.[44]

3D ICs and heterogeneous architectures are introducing challenges related to arranging, assembling, validating, and testing the various chiplets, which can sometimes be preassembled.[45] This shift toward system design over individual product design can incorporate software and digital twins early on—stressing the importance of early and frequent testing.[46] By 2025, synchronizing hardware, system, and software development upstream in the process will likely help redefine future systems engineering and enhance overall efficiency, quality, and time to market.

To evolve and keep pace with the changing face of design, the industry may want to consider new ways to handle complex design processes. Already, the chip industry is exploring digital twins to emulate and visualize complex design processes step by step, including the ability to move around or swap chiplets to measure and assess performance of a multi-chiplet system.[47] And digital twins could increasingly be used to give a visual representation (via 3D modeling) of the physical end-device or the system to assist with all aspects of design, including mechanical as well as electrical (software and hardware). Designers should work with electronic design automation (EDA) and other hi-tech computer-aided design/computer-aided engineering companies to strengthen design, simulation, and verification and validation tools and capabilities for hybrid and complex heterogenous systems.[48] And they also should consider using and adapting model-based system engineering tools as part of the broader EDA “shift-left” approach.[49]

As design and software are expected to play crucial roles in the development of next-generation advanced chip products, bolstering cyber defense becomes more important, heading into 2025.[50] To help align with the shift-left approach, chip designers should integrate security and safety testing early in the chip-design process. They should implement redundancy and error-correction and -detection mechanisms to help ensure that systems can continue to operate even when some of the components fail, and hardware-based security features such as secure boot mechanisms and encryption engines.[51]

Strategic questions to consider

- As AI in chip design becomes more prevalent and common and EDA becomes more and more AI-enabled, how can the industry proactively ensure trust and transparency in the complex design process by always keeping human engineers in the loop and giving them a major role in the overall process?
- In the case of custom silicon design, what’s the nature of the relationship between original equipment manufacturers for devices, product designers, and chip designers? And what are some differentiating factors for chip companies and end customers? Does increased customization give scale advantages in terms of product pricing, or lower the cost to produce a prototype or accelerate prototype production?
- New tools and methodologies may require the broader chip industry, including EDA and design houses, to consider long-term direction and goals. In this context, what aspects should semi companies address from the standpoints of systems engineering and chip development/R&D?
- How might the growing demand to more quickly design more complex chips and at an increasingly faster pace affect manufacturing capabilities and capacity, especially for the back-end players (advanced packaging foundries and outsourced semiconductor assembly and test)?

The intensifying talent challenges in semiconductor industry

In **Deloitte’s 2023 semiconductor industry outlook**, we estimated that the industry needs to add a million skilled workers by 2030, or more than 100,000 every year.[52] Two years after, not only does that forecast hold good, but the talent challenge is expected to intensify further in 2025. Globally, countries are not producing enough skilled talent to meet their workforce needs.[53]

From core engineering to chip design and manufacturing, operations, and maintenance, AI may help alleviate some engineering talent shortages, but the skill gap looms (figure 5).[54] Attracting and retaining talent will likely continue to be a challenge for many organizations in 2025, and a big part of the problem is an aging workforce, which is more prominent in the United States and even Europe.[55] Add the complex geopolitical landscape and supply chain fragility to this equation, and it becomes clear that the availability of talent supply is under stress globally.[56]

With onshoring and reshoring of fabrication, assembly, and test in the United States and Europe, there will likely be pressure on chip companies and foundries as they source more of the talent locally in 2025. For example, talent challenges are contributing to delays in opening new plants.[57] On a related note, "friendshoring" (collaborating with companies from countries considered to be allies) can provide stability and resilience to supply chains, especially for the United States and European Union. But it also demands scouting for the right skills to help meet new capacity demands and talent roles in destinations such as Malaysia, India, Japan, and Poland.[58]

Chip companies can't continue to wrestle over the same finite talent pool and still expect to match up to the industry's pace of technological advancement and capacity expansion. So, what can semiconductor companies do in 2025 to address the talent conundrum?

To help attract AI and chip talent, chip companies should consider offering a sense of trust, stability, and projected market growth. With this, they can help make the industry more appealing to recent high school grads and fresh entrants to help reinvigorate talent pipelines.[59]

Countries aiming to benefit from their respective domestic chips acts should consider weaving in strategic goals and aspects related to workforce development and activation. Some examples could include training programs, expanded vocational and professional education, and employment opportunities that their local chip companies would commit to in order to receive funding. Semi companies should consider collaborating with educational institutions (high schools, technical colleges, and universities) and local government organizations to leverage chip funds to develop and curate targeted workforce training and development programs aligned with specific industry needs in the region.

Semi companies should design flexible upskilling and reskilling programs for career path flexibility to help address future workforce skills and gaps. Additionally, they should implement and leverage advanced tech and AI-based tools to assess diverse talent related factors such as supply, demand, and current and projected spend, to perform complex workforce scenario modeling to support strategic talent decision-making.[60]

Strategic questions to consider

- How should the workforce be characterized and segmented based on specialization areas, for example, design and intellectual property, and manufacturing, operator, engineering, and technical roles? And how can the industry customize talent sourcing and skill development strategies based on these roles, as well as based on specific geographic regions where hiring takes place?

- An emerging trend is agentic AI:[61] Could multimodal, multiagent AI be a partial solution to the looming talent shortage?

- When integrating new talent into a mainstream workforce, what nuances and factors should be considered to ensure a consistent corporate culture? And what associated risks and pitfalls related to talent retention issues and talent pipeline development gaps should be tackled?

- As part of future talent pipeline development, what adjacent skilled workforce types should be considered and what should the overall talent mix be, including full-time and gig workers, to help position the company strongly in the next one to two years?

Figure 5

Semiconductor talent shortage is a global issue and presents distinct challenges to the industry

80%

...of graduates in the **US** with a master's degree in semiconductor-related engineering fields do not stay in the country after completing their graduation.

90%

...of companies surveyed in **APAC** highlighted talent acquisition and development as a top priority to sustain industry growth and competitiveness.

63%

...of companies surveyed in **APAC** highlighted talent capability and retention as major industry risks.

Source: Sergey Shchemelev et al, "The global semiconductor talent shortage," Deloitte Global, October 2024.

Deloitte. | deloitte.com/us/en/insights/research-centers/center-for-technology-media-telecommunications.html

Building resilient supply chains amid geopolitical tensions

Deloitte's 2024 semiconductor outlook already talked about geopolitical tensions in depth, so what's new for 2025?

The same … but even more. As one example, in December of 2024, the outgoing administration issued a new list of US export restrictions mainly still focused on advanced nodes (despite some speculation that restrictions might be broadened to include some relatively less advanced nodes). These restrictions now include separate additional categories around advanced inspection and metrology.[62] Additionally, many (over 100) new entities (mainly Chinese) have been added to the restricted entity list.[63]

As part of these restrictions, the United States seems to be adopting the "small yard, high fence" approach toward semiconductor export restrictions.[64] This aims to impose a high level of restrictions on a relatively small subset of chip technologies, with a focus on those that defense, including advanced weapon systems, and advanced AI used in military applications.[65]

The new restrictions (if implemented by the new administration) go on to flag that AI advancements are increasingly being viewed as matters of national security. The day after those new restrictions, China announced further restrictions on the export of gallium and germanium (as well as other materials), both key for the manufacture of multiple semiconductors.[66] As we predicted in 2024, ongoing materials restrictions will likely pose a challenge for the chip industry, but also an imperative for the industry to do more recycling of e-waste.[67]

In mid-January 2025, the outgoing administration announced the Interim Final Rule on AI Technology Diffusion. The Interim Final Rule will impose new controls for chip exports.[68]

At the time of writing, it is unknown whether the incoming administration will roll back the December and January restrictions, modify them, or even propose additional restrictions.

Additionally, the new administration has proposed increasing its use of tariffs, including tariffs on goods from China, Mexico, and Canada.[69] Given the global nature of most semi supply chains, the proposed new AI-related chip-export controls (by the outgoing administration) and the planned higher tariffs would likely have an impact and could make supply chains more complex to administer, shifting profits, costs, and more. And the impact could be felt across the supply chain—including R&D and manufacturing—as well as affecting how industry policies are shaped across countries and regions.

Of course, there are additional geopolitical risks or changes: Conflicts in Ukraine/Russia and the Middle-East continue, potentially affecting semiconductor manufacturing, supply chains, and critical raw materials. But the chip industry has other vulnerable points: The December martial law order in South Korea highlighted the global supply chain's dependency on and concentration of certain types of semiconductors, especially in the most advanced technologies.[70] As an example of concentration, almost 75% of DRAM memory chips globally are made in South Korea.[71]

It's not just geopolitics that can interrupt key materials: 2024's Hurricane Helene briefly shut down two mines in North Carolina, which are sources for nearly all of the world's ultra-high-purity quartz, essential for making crucibles which are a key part of the chipmaking process.[72] With hurricanes, typhoons, and other extreme weather events projected to become more frequent and intense due to climate change, expanding sources for key materials is likely to remain a supply chain priority.[73]

It is worth noting that, as of late 2024, a key part of the export restrictions from the United States and allies is having an effect: The restrictions around extreme ultraviolet lithography machines seem to be posing a barrier, preventing Chinese companies from making advanced-node chips at scale and with acceptable yields. Although there are 7 nanometer and 6 nanometer chips being made in limited numbers using older deep ultraviolet technology, the volumes are low, yields are uneconomical, and that situation is expected to persist at least until 2026.[74]

To be clear, semiconductor supply chains worked well in 2024, even as the industry grew by almost 20%. At this time, there's no reason to believe 2025 supply chains will be less resilient, but as always, the risk is there. And given how important gen AI chips are expected to be in 2025 and beyond (up to 50% of sales, perhaps[75]) and the relatively higher concentration of processor, memory, and packaging required for cutting-edge chips, the industry may be more vulnerable to supply chain disruptions than ever before. Although the industry is likely to become less concentrated geographically thanks to various chips acts—and initiatives like onshoring, reshoring, nearshoring, and friendshoring are all still in their early days—the industry remains highly vulnerable for the next year or two, at least.

Strategic questions to consider

- Given the fluid geopolitical environment and escalating export restrictions, what should be the mix of reshoring versus offshoring? And how should the industry factor potential disruptions to any existing supply chain channel partner relationships in erstwhile friendly countries and allies—that is, friendshoring?
- As unpredictable climate-driven disruptions affect materials and components supplies, how could this aspect—coupled with an already complex geopolitical landscape—impact both the front-end wafer fab and back-end assembly and test, and packaging plants that are being ambitiously planned and rolled out in dozens of countries worldwide?
- If the trade wars continue to escalate, what could it mean for talent sourcing and availability? Can export restrictions further extend to, and eventually cascade into a much broader talent mobility challenge for countries locking horns in the chip race?
- How might countries with chip manufacturing capacity respond to potential additional US tariffs, given the incentive to shift activities to the United States? Can higher-value-add activities be ideal candidates to be shifted to the United States, given the higher cost, and might US-based companies rethink their offshore manufacturing investments and activities?

Signposts for the future

For 2025, semiconductor industry executives should be mindful of the following signposts:

1. There is currently a mismatch between very high spending on semiconductors for gen AI, and companies being able to monetize their gen AI offerings. For 2025, the argument of "the risk of underinvesting is greater than the risk of overinvesting" seems to be still dominant, but if that attitude shifts, demand for gen AI chips could become weaker than predicted.

2. Competition from agile chip startups could intensify, challenging incumbents in the broader semiconductor industry. Notably, AI chip startups secured a cumulative US$7.6 billion in venture capital funding globally during the second, third, and last quarter of 2024,[76] and several of these startups offer specialized solutions including customizable RISC-V-based applications, chiplets, LLM inference chips, photonic ICs, chip design, and chip equipment.

3. With interest rates in the United States and other major markets likely to drop further,[77] a favorable credit environment could act as a tailwind for the chip industry's M&A scene, which has already seen an uptick in 2024.[78] Moreover, with two different chip markets evolving—one for AI chips and one for all other types of chips—the industry may experience M&A and consolidation, especially if companies with valuable IP lag their peers and are seen as attractive targets. Nonetheless, potential tighter regulations and trade conflicts, globally, could potentially dampen the deal environment.

4. As geopolitical challenges ripple across the globe, chip companies should brace themselves for further disruptions. Traditional channel partner models and allyships could get upended, even as reshoring, friendshoring, and nearshoring have gained momentum. Prolonging regional conflicts and wars could further impact the flow of vital materials and inventories. All of these could disrupt semi companies' demand planning, requiring them to be more agile and adapt supply chain and sourcing contracts, and pricing terms.

5. A significant part of capex spending and revenues was driven by AI and the advanced wafers needed to produce those highly advanced AI chips. However, wafer demand from auto, industrial, and consumer segments continue to be lackluster, while there's some uptick in demand from mobile handset and other consumer products. Through 2025 and 2026, though overall revenue and capex seem to continue trending upward (at least for the next nine to 12 months), any downward movement in AI-related spending and components shortage could have an adverse impact ripping through the broader global semiconductor and electronics supply chain.

BY

Jeroen Kusters
United States

Deb Bhattacharjee
United States

Jordan Bish
Netherlands

Jan Thomas Nicholas
Malaysia

Duncan Stewart
Canada

Karthik Ramachandran
India

ENDNOTES

1. World Semiconductor Trade Statistics, "WSTS semiconductor market forecast fall 2024," news release, Dec. 3, 2024.

View in Article

2. Ibid.

View in Article

3. Ibid.

View in Article

4. Ibid.

View in Article

5. Deloitte analysis of public market capitalizations in 2024, 2023, and 2022.

View in Article

6. Deloitte analysis of semiconductor subsectors.

View in Article

7. Duncan Stewart, Christie Simons, Brandon Kulik, and Gillian Crossan, "Gen AI chip demand fans a semi tailwind … for now," *Deloitte Insights*, Nov. 29, 2023.

View in Article

8. Deloitte analysis of financial statements from the major generative AI chipmakers.

View in Article

9. Deloitte estimates based on company and analyst forecasts.

View in Article

10. Jacob Fox, "AMD's Dr. Lisa Su predicts AI GPU market will grow to $500 billion by 2028 or 'roughly equivalent to annual sales for the entire semiconductor industry in 2023'," PC Gamer, Oct. 30, 2024.

View in Article

11. IDC, "PC refresh cycle and tablets in emerging markets expected to spur demand in coming quarters, according to IDC," press release, Sept. 23, 2024.

View in Article

12. IDC, "Worldwide smartphone market forecast to grow 6.2% in 2024, fueled by robust growth for Android in emerging markets and China, according to IDC," press release, Nov. 26, 2024.

View in Article

13. Semiconductor Industry Association, "AI, auto, industrial markets spurred rebound in chip demand during second half of 2023," press release, March 18, 2024.

View in Article

14. Semiconductor Industry Association, "2024 state of the US semiconductor industry," Sept. 9, 2024. If global revenues were US$611 billion, then average price per chip was roughly $0.61.

View in Article

15. Deloitte calculation based on reported average selling price for gen AI chips.

View in Article

16. SEMI, "Global silicon wafer shipments to remain soft in 2024 before strong expected rebound in

16. SEMI, "Global silicon wafer shipments to remain soft in 2024 before strong expected rebound in 2025, SEMI reports," Oct. 21, 2024.

View in Article

17. Duncan Stewart, Karthik Ramachandran, and Prashant Raman, "Silicon building blocks: Chiplets could move Moore's Law forward," *Deloitte Insights*, Nov. 19, 2024.

View in Article

18. TrendForce, "[News] TSMC ramps up CoWoS capacity across Taiwan, projected to nearly triple by 2026," Dec. 13, 2024.

View in Article

19. Christophe Fouquet, "Global market trends: Industry & ASML's technology roadmap," ASML, Nov. 14, 2024.

View in Article

20. Ibid.

View in Article

21. World Semiconductor Trade Statistics, "Historical billings report," accessed Dec. 6, 2024.

View in Article

22. IDC, "PC refresh cycle and tablets in emerging markets expected to spur demand in coming quarters, according to IDC"; IDC, "Worldwide smartphone market forecast to grow 6.2% in 2024, fueled by robust growth for Android in emerging markets and China, according to IDC."

View in Article

23. Ariane Bucaille, Kevin Westcott, Gillian Crossan, and Lara Abrash, "TMT Predictions 2025: Bridging the gaps," *Deloitte Insights*, Nov. 19, 2024.

View in Article

24. Ibid.

View in Article

25. Deloitte analysis of brokerage and industry reports estimating the gen AI chip market demand.

View in Article

26. Chris Arkenberg, Duncan Stewart, Gillian Crossan, and Kevin Westcott, "On-device generative AI could make smartphones more exciting—if they can deliver on the promise," *Deloitte Insights*, Nov. 19, 2024.

View in Article

27. IDC, "PC refresh cycle and tablets in emerging markets expected to spur demand in coming quarters, according to IDC."

View in Article

28. Canalys, "AI-capable PCs forecast to make up 40% of global PC shipments in 2025," March 18, 2024.

View in Article

29. Paul Alcorn, "Intel confirms Microsoft's Copilot AI will soon run locally on PCs, next-gen AI PCs require 40 TOPS of NPU performance," Tom's Hardware, March 27, 2024.

View in Article

30. Deloitte author conversations with chief information officers.

View in Article

31. Ishan Dutt and Ben Yeh, “**Qualcomm’s vision of a reborn PC experience through Snapdragon X Elite**,” Canalys, June 13, 2024.

View in Article

32. Kif Leswing, “**Qualcomm says it expects $4 billion in PC chip sales by 2029, as company gets traction beyond smartphones**,” *CNBC*, Nov. 19, 2024.

View in Article

33. Deloitte analysis: smartphone CPUs cost about US$10 to US$30 each. The area of these chips is about 100 square millimeters, and based on die shots of the silicon from different chips, the NPU portion makes up about 3% to 5% of the area of the total chip.

View in Article

34. Arkenberg, Stewart, Crossan, and Westcott, “**On-device generative AI could make smartphones more exciting—if they can deliver on the promise**.”

View in Article

35. Sarah Barry James, “**Consumer checkup: Higher interest rates lead to longer tech replacement cycles**,” S&P Global, March 26, 2024.

View in Article

36. Jeff Loucks, Duncan Stewart, Christie Simons, and Brandon Kulik, “**AI in chip design: Semiconductor companies are using AI to design better chips faster, cheaper, and more efficiently**,” *Deloitte Insights*, Nov. 30, 2022.

View in Article

37. Deloitte finds gen AI can test human designs and identify placement errors that increase power consumption, impede performance, or use space inefficiently. And then, it can suggest, simulate, and test improvements. But what’s truly revolutionary is that gen AI can do all of these autonomously, guided by design principles, and can generate PPAs that surpass what humans create using typical EDA processes. For further reading, see: Deloitte, “**How generative AI is transforming the semiconductor industry**,” accessed Jan. 29, 2025.

View in Article

38. In a “shift-left” approach, any potential issues, bugs, and errors could be discovered earlier, thereby lowering the chances of escalating issues at later stages. Shifting left helps engineers to not only limit the risk of possible costly re-spins/re-designs, but also enhances design quality and reliability and speed up the design and verification processes. For further reading, see: Brian Bailey, “**Shift left is the tip of the iceberg**,” Semiconductor Engineering, Nov. 14, 2024.

View in Article

39. This change to greater emphasis on system-level metrics will likely be due to the shift toward new and advanced packaging methods, especially for gen AI and HPC environments that require scalability, modular designs, and multi-chip packages, reliability and security, as well as low-power design methods to manage thermal aspects better.

View in Article

40. Loucks, Stewart, Simons, and Kulik, “**AI in chip design**.”

View in Article

41. Bailey, “**Shift left is the tip of the iceberg**”; additionally, see: Bucaille, Westcott, Crossan, and Abrash, “**TMT Predictions 2025**.”

View in Article

42. Given their compact size, ASICs suit smartphones well, as they balance performance delivery with low-power needs of the small, constrained form-factor that is battery-powered; for further reading, see: Giuseppe Ciccomascolo, "**Future of AI is also a chip size matter: What's ASIC and why it's relevant now**," CCN, Jan. 5, 2024.

View in Article

43. ASICs demand custom and specialized design tools, and require specialized architects, designers, and verification engineers; for further reading, see: Danny Crichton, "**Will ASIC chips become the next big thing in AI?**," Lux Capital, Sept. 16, 2024.

View in Article

44. Ibid.

View in Article

45. Stewart, Ramachandran, and Raman, "**Silicon building blocks**."

View in Article

46. Once separate, software has become more integral to hardware design; for further reading, see: Bailey, "**Shift left is the tip of iceberg**."

View in Article

47. Ann Mutschler, "**Digital twins gaining traction in complex designs**," Semiconductor Engineering, June 27, 2024; Stewart, Ramachandran, and Raman, "**Silicon building blocks**."

View in Article

48. Ann Mutschler, "**Chip design digs deeper into AI**," Semiconductor Engineering, June 3, 2024; see further: Stewart, Ramachandran, and Raman, "**Silicon building blocks**."

View in Article

49. Don Dingee, "**Turning MBSE inside-out for an RF EDA shift left**," Semiconductor Engineering, April 19, 2022.

View in Article

50. EDA companies, chip designers, and security experts can devise ways to develop in-built functionality that could sense potential IP theft and cyber infringement at component level, and work with rest of the supply chain to help address the broader threat and attack parameters that could affect the distinct components that go into building larger systems for HPC and AI environments. For further reading, see: Stewart, Ramachandran, and Raman, "**Silicon building blocks**"; and Mutschler, "**Digital twins gaining traction in complex designs**."

View in Article

51. Deloitte analysis based on proprietary research and validated and confirmed with industry subject-matter specialists.

View in Article

52. Deloitte Global, "**2023 semiconductor industry outlook**," accessed Jan. 29, 2024.

View in Article

53. Deloitte, "**The global semiconductor talent shortage**," accessed Jan. 29 2024.

View in Article

54. Deloitte, "**How generative AI is transforming the semiconductor industry**"; Belle Lin, "**Designing chips is getting harder. These engineers say chatbots and AI can help**," *The Wall Street Journal*, Feb. 6, 2024.

Feb. 6, 2024.

View in Article

55. The US semiconductor workforce is older than other technology industries', with 55% of the workforce being above 45 and less than 25% under 35 years of age (as of July 2024), and in Europe, 20% of the workforce is above 55, with Germany expecting about 30% to retire over the next decade. For further reading, see: Anne-Françoise Pelé, "Semiconductor capacity is up, but mind the talent gap," EE Times Europe, April 15, 2024. Additionally, inconsistent knowledge management, and the lack of new talent to adopt institutional knowledge, present workforce barriers for many chip companies (see: Deloitte, "The global semiconductor talent shortage.").

View in Article

56. On one hand, the United States has not only restricted export of advanced node AI chips and chipmaking equipment, but also limits US persons from performing work for certain Chinese chipmakers without special licensing. To counteract, China has been aggressively recruiting expatriate talent, and is continuing to do so with high salaries, free homes, and more. To read further, see: Reuters, "China quietly recruits overseas chip talent as US tightens curbs," Aug. 24, 2023.

View in Article

57. Michelle Toh, "TSMC says its $40 billion chip project in Arizona faces a further delay," *CNN*, Jan. 19, 2024.

View in Article

58. Melanie Rojas, Adam Routh, Jesse Sherwood, John Buckley, and Akash Keyal, "Reshoring and 'friendshoring' supply chains," *Deloitte Insights*, March 24, 2022. See further in: Deloitte, "The global semiconductor talent shortage."

View in Article

59. Deloitte, "The global semiconductor talent shortage"; moreover, the semi industry historically has always struggled to demonstrate brand recognition and value proposition, as noted in the 2017 SEMI-Deloitte Workforce study (see: Chris Richard, Karthik Ramachandran, and Ivan Pandoy, "Looming talent gap challenges semiconductor industry," SEMI, accessed Jan. 29, 2025).

View in Article

60. Additionally, skills-based job architectures can be analyzed for opportunities to increase capabilities and efficiencies using AI, consolidating workforce gaps, and reducing total workforce spend.

View in Article

61. Jeff Loucks, Gillian Crossan, Baris Sarer, China Widener, "Autonomous generative AI agents: Under development," *Deloitte Insights*, Nov. 19, 2024.

View in Article

62. Bureau of Industry and Security, "Commerce strengthens export controls to restrict China's capability to produce advanced semiconductors for military applications," press release, Dec. 2, 2024.

View in Article

63. Ibid.

View in Article

64. Ibid.

View in Article

65. Ibid.

View in Article

66. Amy Lv and Tony Munro, “China bans export of critical minerals to US as trade tensions escalate,” Reuters, Dec. 3, 2024.

View in Article

67. Duncan Stewart, Gillian Crossan, Ariane Bucaille and Christie Simons, “A raw deal: Will materials shortages, supply chain challenges threaten tech's future?,” *Deloitte Insights*, Nov. 28, 2023.

View in Article

68. Bureau of Industry and Security, “Biden-Harris administration announces regulatory framework for the responsible diffusion of advanced artificial intelligence technology,” press release, Jan. 13, 2025.

View in Article

69. Costas Pitas, “Trump vows new Canada, Mexico, China tariffs that threaten global trade,” Reuters, Nov. 26, 2024.

View in Article

70. Robyn Mak, “South Korea curveball adds new 2025 risks,” Reuters, Dec. 4, 2024.

View in Article

71. Lino Jeng, “DRAM market dynamics—February 2024,” Omdia, Feb. 26, 2024.

View in Article

72. Megan Cassella, “How a tiny town hit by Helene could upend the global semiconductor chip industry,” *CNBC*, Oct. 3, 2024.

View in Article

73. Mark Poynting and Esme Stallard, “How climate change worsens heatwaves, droughts, wildfires and floods,” BBC, Nov. 14, 2024.

View in Article

74. Hassam Nasir, “US sanctions freeze Huawei's Ascend AI chips on older 7nm process node that’s eight years behind TSMC, stalling progress until at least 2026”, Tom’s Hardware, Nov. 20, 2024.

View in Article

75. If the US$500 billion in AI chip sales projection mentioned in endnote no. 9 occurs in 2028, that would be well over 50% of the roughly US$800 billion to US$900 billion in revenues, assuming the industry remains on track for US$1 trillion in 2030.

View in Article

76. Deloitte’s analysis based on data presented in Semiconductor Engineering’s articles on startup funding, Q2 2024, Q3 2024, and Q4 2024.

View in Article

77. Mike Wade, Michelle Gauchat, and Val Srinivas, “2025 banking and capital markets outlook,” *Deloitte Insights*, Oct. 13, 2024.

View in Article

78. Deloitte's analysis finds the top 10 M&A deals in the broader tech and chip industries as of October 2024 (year to date) involved three semiconductor company deals, totaling US$44.4 billion in value.

View in Article

ACKNOWLEDGMENTS

The authors wish to thank the following Deloitte colleagues for sharing their insights and perspectives: **Dan Hamling**, **Nina Zhang**, **Gillian Crossan**, **Sergey Shchemelev**, **Alan Huang**, **Dolly Zhang**, **Tomey Jiantao Lin**, **Steve Fineberg**, **Jeff Loucks**, **Brandon Kulik**, and **Christie Simons**.

Additionally, they would like to thank **Gautham Dutt** for his design support, **Shannon Rothacher** and **Alison Zink** for marketing and PR support, and **Michelle Dollinger** and **Shubham Oza** for helping with project management. Thanks also to the Deloitte Insights team (**Andy Bayiates**, **Prodyut Borah**, **Molly Piersol**) for their support to help us publish this report.

Cover image by: **Jaime Austin**; Adobe Stock