The semiconductor industry in the AI era

Innovating for tomorrow's demands

Capgemini RESEARCH INSTITUTE

#GetTheFutureYouWant

Table of contents

Executive summary

AI adoption is powering a surge in demand for semiconductors:

While semiconductor industry organizations forecast a 15% rise in two years, the downstream organizations (those reliant on semiconductor supply for their products or services and operations) anticipate their demand for chips to increase at a higher growth rate of 29%. Increased adoption of artificial intelligence (AI) and generative AI (Gen AI) is driving the need for specialized neural processing units (NPUs) and high-performance graphics processing units (GPUs) that can handle massive computations and large datasets efficiently. Additionally, downstream organizations expect their demand for AI chips, custom silicon chips, and memory-intensive chips to increase over the next 12 months.

Amid buoyant demand lies concern:

Over half of downstream organizations doubt the semiconductor industry's ability to meet their needs. Technological advancements, including GPU computing for AI and machine learning and inference acceleration are important to these industries, which are continually seeking to enhance customization, introduce more comprehensive application programming interfaces (APIs) and software development kits (SDKs), as well as reinforcing security features. Consequently, one in three downstream organizations is exploring or is actively engaged in in-house chip design, enabling greater customization while also gaining more control over their supply chains. Further, sustainability, supply chain resilience, and security are critical concerns for downstream organizations going forward.

Amid "softwarization" challenges, innovation shines through:

The semiconductor industry continues to excel in innovation across a number of areas. Although various players stand out in specific areas, our analysis reveals three types of innovation that are consistently prioritized across the industry:

- **Design innovation:** Advances in chip architectures, such as 3D integrated circuit (IC) design and multi-die integration are pushing the boundaries of performance and energy efficiency, while half of design organizations are investing in Gen AI to shorten design cycles.

Executive summary

- **Manufacturing innovation:** Advances in extreme ultraviolet (EUV) lithography and the shift towards smaller process nodes (i.e., 3 nanometer and 2 nm) enable the production of more powerful and efficient chips. Nearly half of manufacturers also rely on AI and ML to optimize processes.
- **Packaging innovation:** 3D packaging and the use of chiplets (tiny integrated circuits that can be combined to create complex components) are enhancing functionality and performance without increasing physical footprint.

Hardware security remains paramount, with significant investment in secure chip design, hardware-based encryption, and root of trust (RoT, a trusted source within a cryptographic system) technologies. But while there is steady progress in integrating software and hardware to create more adaptable and programmable semiconductor solutions, monetization remains a hurdle.

However, the "softwarization of semiconductors" is falling short of the industry's expectations. While this innovation is crucial for semiconductor companies to expand use cases, extend the chip lifecycle, and enhance customization in an evolving market, the industry finds monetizing its software a challenge.

Focus on supply chain resilience and sustainability:

Only two in five semiconductor organizations are confident in the resilience of their supply chains. Organizations focus on onshoring and "friendshoring" (basing supply chains in countries that are geopolitical allies) to enhance stability and reduce single-region dependency. Consequently, the industry anticipates domestic sourcing will improve by 17% over the next two years. Three-quarters (74%) of organizations expect to increase their US investments, and 59% will increase investment in the EU.

Besides its continued focus on improving power efficiency, the industry is becoming more eco-friendly by cutting energy consumption, implementing water recycling and reuse systems, using less toxic alternative chemicals, and minimizing waste.

Executive summary

Recommendations for the semiconductor industry:
To capitalize on emerging opportunities, semiconductor organizations should consider the following:

- Utilize **AI and Gen AI** to automate design processes, improve production efficiency, and optimize performance to meet the specialized needs of emerging applications.
- Invest in **cutting-edge fabrication methods** such as 3D chip stacking and accelerate research in emerging fields such as advanced silicon photonics integration.
- **Diversify supplier networks** across multiple regions while investing in R&D for alternative materials and technologies. Implement **sustainable manufacturing practices** such as green chemistry and utilize renewable energy sources to minimize carbon footprint.
- **Coordinate strategies with governmental initiatives** such as grants for R&D and collaborate within the industry ecosystem to drive shared innovation and standardization.
- Enhance cybersecurity measures to protect data integrity, use **advanced security** to safeguard proprietary technologies, and advocate for stronger intellectual property (IP) laws to deter infringement and protect innovation-led competitive advantage.
- Adopt **open standards** and open-source collaboration to drive semiconductor innovation.

Who should read this report and why?

This report will be relevant for decision-makers across the semiconductor ecosystem and its down-stream industries. Specifically, this will be useful for:

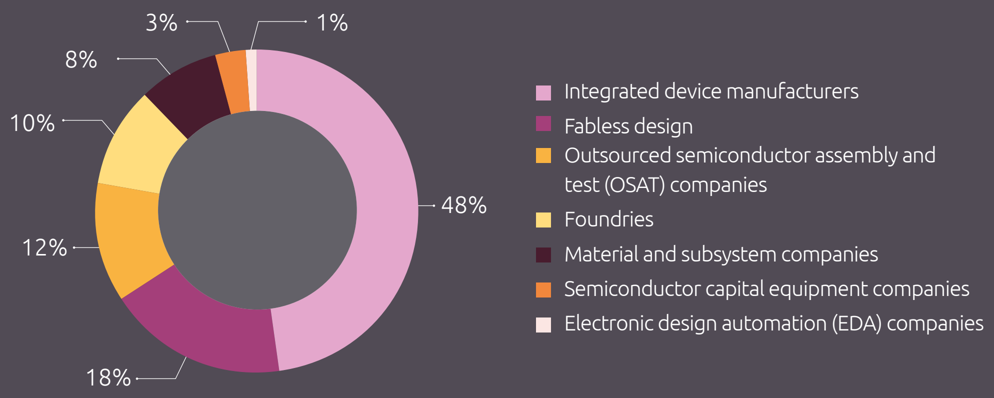
1. Executives in the semiconductor industry: Integrated device manufacturers, fabless design firms, foundries, OSAT companies, material and subsystem companies, and semiconductor capital equipment manufacturers will gain a strategic outlook on industry trends, including advancements in design, manufacturing, and packaging. The report also addresses critical topics like cybersecurity, softwarization, supply chain resilience, and sustainability—enabling executives to align their strategies for the future.

2. Leaders in downstream industries: Professionals in automotive, consumer electronics, retail, telecom, aerospace and defense, high tech, medical devices/medical electronics, industrial equipment, financial services, and energy industries will gain a deeper understanding of how semiconductor trends impact their industries. Insights into custom silicon, in-house chip design, sustainability, and supply chain dynamics provide guidance for aligning technology adoption with organizational goals.

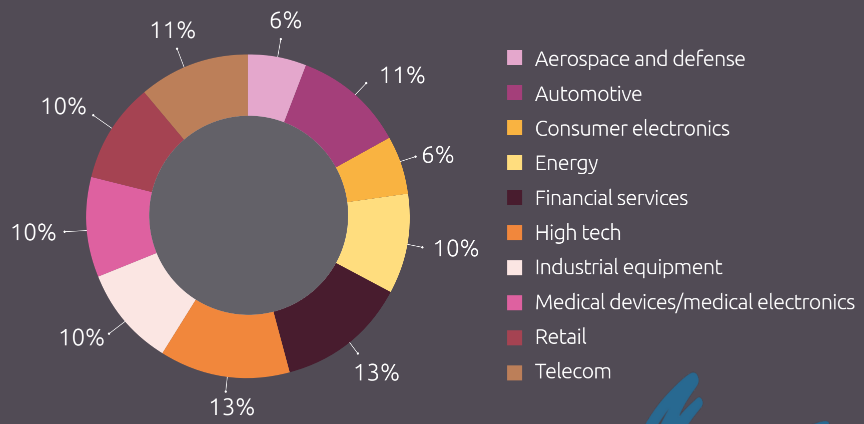
By connecting perspectives from both semiconductor executives and downstream organizations, this report equips stakeholders with the knowledge and strategies needed to thrive in a rapidly evolving landscape.

Snapshot of research methodology

1. Survey of 250 semiconductor industry organizations



2. Survey of 800 downstream organizations



3. Twelve in-depth interviews with executives from the semiconductor industry and downstream industries.

Definitions

- **Neural processing units (NPUs):**[1] NPU architecture simulates the neural network of the human brain. It processes large amounts of data simultaneously, performing trillions of operations per second. It uses less power and is far more efficient than a CPU or GPU, while freeing these up for other tasks. Combining an NPU with machine learning (ML) offers lightning-fast, high-bandwidth AI in real time.
- **Data processing units (DPUs):**[2] A DPU is a "system-on-a-chip" (SoC) that combines:
  – An industry-standard, high-performance, software-programmable, multi-core CPU, tightly coupled with the other SoC components,
  – A high-performance network interface capable of parsing, processing, and efficiently transferring data at line rate, or at least network speed, to GPUs and CPUs,
  – A rich set of flexible and programmable acceleration engines that offload and improve applications performance for AI and ML, zero-trust security, telecommunications and storage, among others.
- **Graphics processing units (GPUs):**[3] A GPU is an electronic circuit that can perform mathematical calculations at high speed. Computing tasks such as graphics rendering, ML, and video editing require similar operations on a large dataset. GPUs can perform the same operation on multiple data values simultaneously. This increases processing efficiency for many compute-intensive tasks.
- **AI chips:**[4] AI chips are specialized computing hardware used in the development and deployment of AI systems. AI chips are essential for meeting the demand for greater processing power, speed and efficiency.
- **Custom chips:**[5] Custom or application-specific integrated circuit (ASIC) chips are designed for a specific purpose, such as image processing or inference, at a lower monetary and resource cost than a general-purpose processor. ASICs enable ML and other typically high-cost functionality in situations where it would otherwise be impractical. ASICs will not always be the appropriate solution, but are worth consideration.
- **Softwarization:**[6] Softwarization is the concept of developing a more standardized, limited set of base chips that can be customized for various industries and solutions by reducing the number of unique hardware designs and instead using software to provide industry-specific functionalities – essentially moving the "logic" from silicon to software.
- **Downstream industries:** While nearly all industries rely on semiconductors for their products or services and operations, the scope of this research includes automotive, consumer electronics, retail, telecom, aerospace and defense, high tech, medical devices/medical electronics, industrial equipment, financial services, and energy.

01 Organizations anticipate surging semiconductor demand

29%

expected increase in demand for semiconductors by downstream organizations, to the end of 2026

Semiconductors are the backbone of the digital world, powering smartphones, Gen AI, computers, cars, satellites, and virtually every electronic device in use today. In 2023, nearly 1 trillion semiconductors – more than 100 times the number of people on the planet – were sold worldwide.[7] Despite a cyclical market decline in the first half of 2023, worldwide sales recovered in the second half of the year to $527 billion.[8] By 2030, it is anticipated that the semiconductor market size will surpass $1 trillion.[9]

Semiconductor manufacturers and downstream industries expect demand to rise

The semiconductor industry and downstream industries anticipate significant growth in demand over the next two years. The main drivers of this increase are the growing usage of sophisticated electronics, the rise in electric and driverless cars, and the development of smart technologies and high-speed internet. Growth in data-driven applications and demand for energy-efficient solutions are also helping to perpetuate this trend.

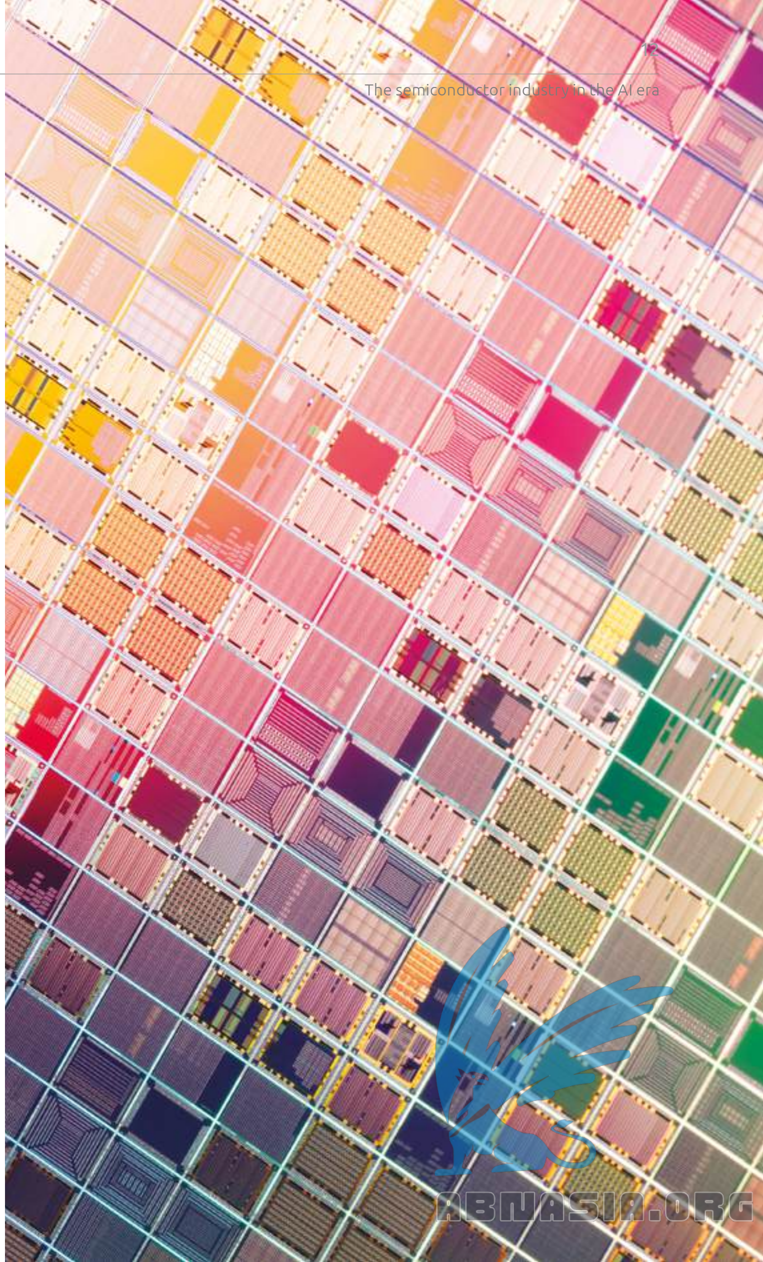
While the semiconductor industry is cyclical by nature, our research indicates that the semiconductor industry anticipates demand to increase by 15% by 2026, while downstream organizations expect an increase of almost 30% (see Figure 1). While we did not ask the respondents for their projections beyond 2026, extrapolating from the current increase of 15% suggests that the market size could reach approximately $930 billion by 2030, nearing the $1 trillion estimate mentioned earlier. The semiconductor industry's cautious outlook reflects a slow recovery in markets like automotive despite strong AI-driven demand.

Figure 1.

Downstream industries estimate demand for semiconductors to increase at double the rate of the semiconductor industry's expectation

Expected semiconductor demand increase in two years to the end of 2026

Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 190 semiconductor industry organizations (includes Integrated Device Manufacturers (IDMs),fabless design firms and foundries), N = 800 downstream organizations.

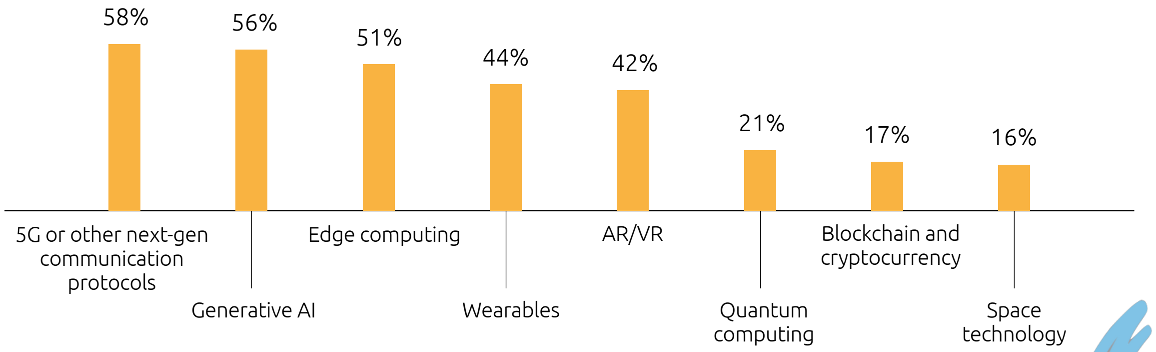


Rapid technological progress is driving demand for more powerful, efficient, and customized chips, prompting semiconductor manufacturers to invest heavily in R&D. Our research indicates that 58% of semiconductor organizations believe that 5G or other next-gen communication protocols are impacting their strategies, while 56% believe that Gen AI is a strong influence. 5G and other next-generation communication protocols are foundational to enabling a wide range of emerging technologies and markets, including IoT, autonomous vehicles, AR/VR, and edge computing. They drive demand for advanced semiconductor solutions that require higher performance, energy efficiency, and integration.

Figure 2.

Nearly three in five semiconductor organizations say that 5G (or other next-gen communication protocols) and Gen AI are impacting their strategy

Technology domains impacting semiconductor manufacturing strategy



Source: : Capgemini Research Institute, Semiconductor survey, November 2024, N = 250 semiconductor organizations.

The adoption of AI/Gen AI is driving GPU/NPU demand

The adoption of AI and Gen AI is driving demand for model training and inference capabilities, and data centers, while the growth of on-device AI applications further underscores the need for specialized semiconductor solutions. Subi Kengeri, VP of AI Systems Solutions at Applied Materials, says, *"The AI era marks a new wave of growth for the semiconductor industry, propelled by the high returns on investment generated by AI's economic value. For AI systems, the key metric remains Total Cost of Ownership, while for Silicon, it is Perf/Watt/$."*

To supply the demand rising from AI/ Gen AI, organizations are ramping up their NPU, GPU, and memory capabilities. A Senior Director at a US-based IDM says, *"We focus on optimizing the interplay of compute, memory, and network components to achieve system-level efficiency as customers adopt AI to unlock its benefits. Computational efficiency minimizes the time required for matrix multiplication, the memory subsystem ensures data is readily available for the compute engine, and the network subsystem enhances throughput, whether in small or large clusters."*

Qualcomm's NPU, designed specifically to take on AI workloads, is an essential enabler of on-device Gen AI capabilities. The NPU offers optimal performance, power, and space efficiency to handle complex ML operations.[10] ChatGPT exemplifies the transformative role of GPUs in AI. Leveraging thousands of NVIDIA GPUs, the training and inference processes for its large language model (LLM) demonstrate the unparalleled efficiency and scalability GPUs bring to AI workloads. This infrastructure supports Gen AI services for over 100 million users, underscoring the critical contribution of GPU technology to cutting-edge semiconductor applications in AI-driven innovation.[11] A Senior Director at a US-based IDM explains, *"One of the key advancements in the semiconductor industry is performance enhancement through parallel computing, exemplified by GPUs. As AI accelerators evolve, various parallelization techniques, including expert parallelism, pipeline parallelism, tensor parallelism, and context parallelism, are being employed. These optimization and architectural innovations are designed to boost overall throughput, whether for training or inference workloads."*

Furthermore, today's Gen AI models require more data to enhance outcomes and exploit new opportunities. Inferencing LLMs such as ChatGPT benefits from advanced memory solutions such as Micron's HBM3E (one of many products on the market), which can optimize performance and reduce CPU offload during AI processing, allowing faster training and greater responsiveness to queries.[12]

In our research, 58% of semiconductor organizations mentioned they expect higher demand for NPUs to accompany growth in Gen AI adoption, with 57% anticipating increased need for high-performance chips and 56% for memory-intensive chips, signaling a shift toward advanced processing solutions.

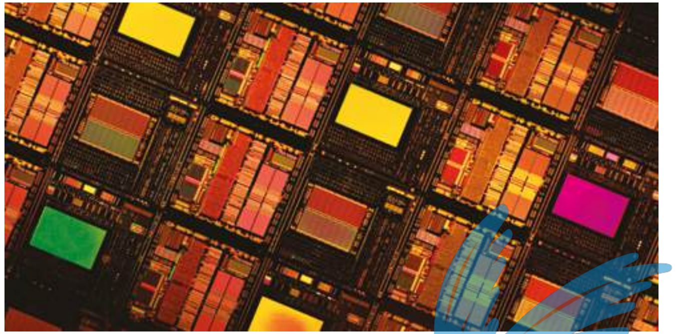
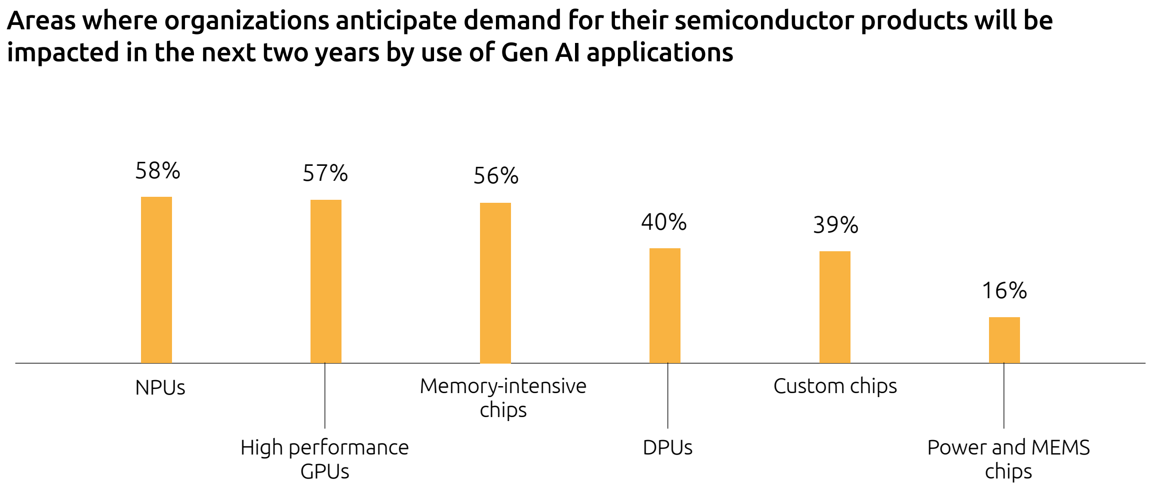
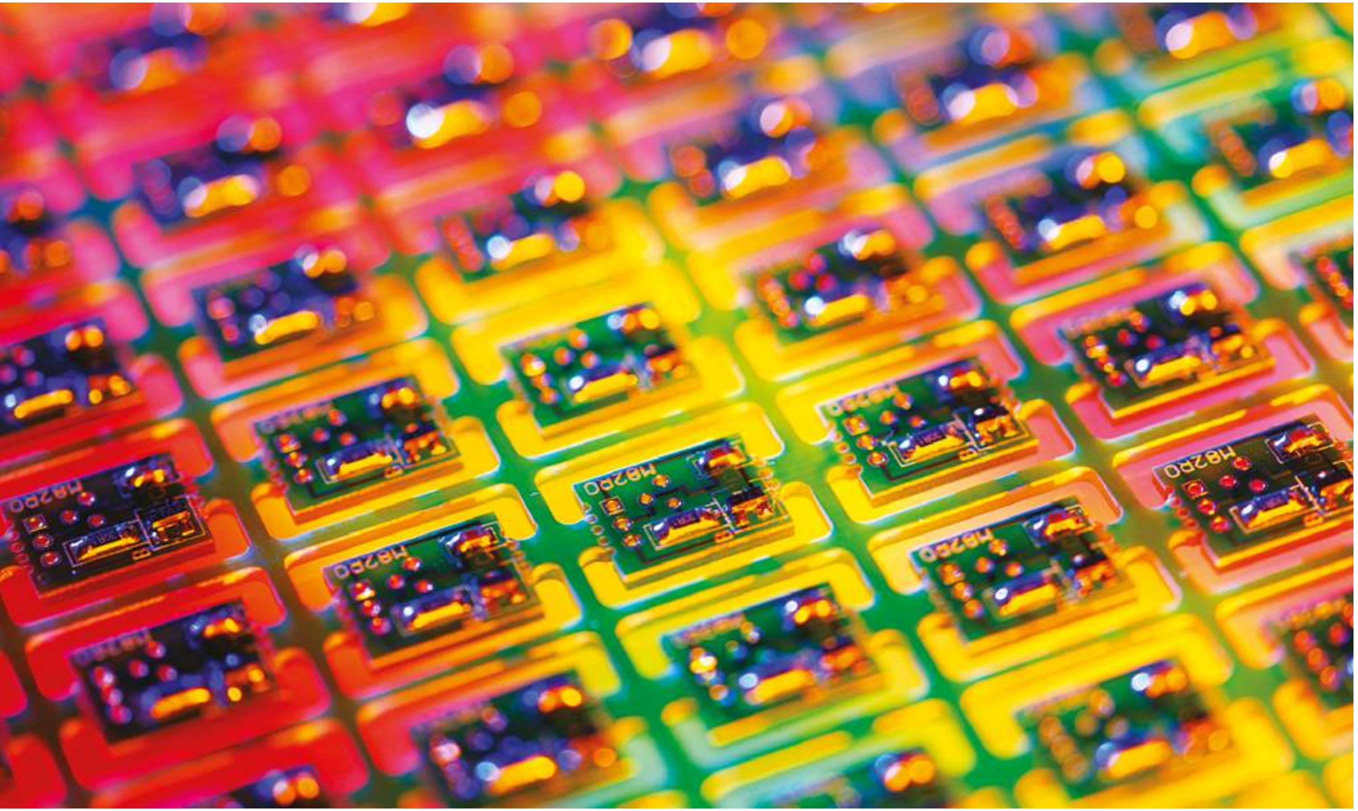


Figure 3.

Due to Gen AI adoption, nearly three in five semiconductor organizations are seeing increased demand for NPUs, high-performance GPUs, and memory-intensive chips



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 250 semiconductor organizations.



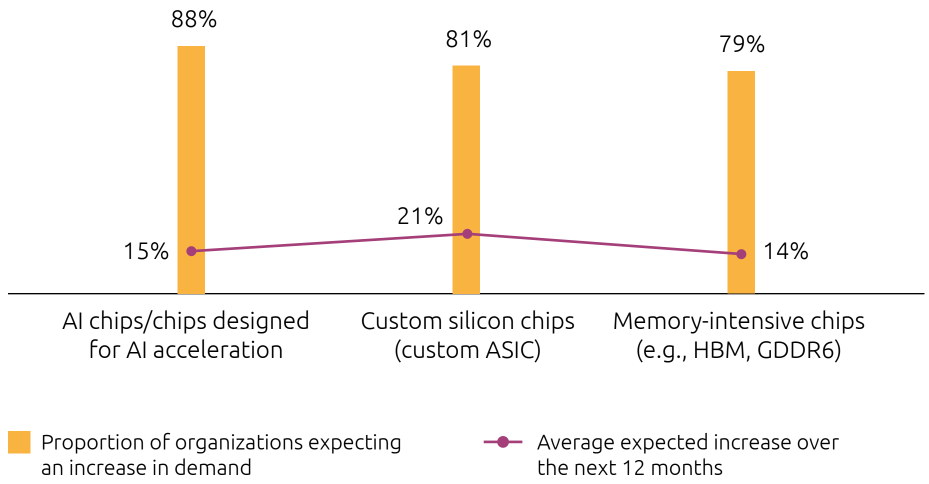
Organizations project high demand for AI chips and custom silicon chips

The market is also witnessing a surge in diverse applications for AI chips and custom-designed chips, underlining their transformative potential. Intel and AWS announced a multi-year co-investment in custom chip design that would encompass Intel products and wafers.[13] Sam Geha, EVP of IoT, Compute and Wireless business at Infineon Technologies, explains, *"Just a few years ago, our role was simply to produce chips, leaving it to customers to determine how to use them. Today, however, we are expected to deliver customized solutions. Software has emerged as a critical differentiator, particularly as our chips have become increasingly complex. Beyond general-purpose software, we now provide specialized solutions for AI and edge AI, enabling customers to effectively train and deploy models. Alternatively, we can offer services to manage the training and deployment for them."*

Our research shows that most downstream organizations foresee heightened demand in the next two years: 88% anticipate a rise in AI chip needs, 81% foresee heightened calls for custom chips, and 79% predict an upturn in demand for memory-intensive chips. As figure 3 shows, according to 39% of semiconductor organizations, Gen AI is expected to drive demand for custom chips in the next two years.

Figure 4.
Nearly four out of five downstream organizations anticipate increased demand for AI chips, custom silicon chips, and memory-intensive chips over the next 12 months

Expected demand for chips



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 800 downstream organizations.

"The AI era marks a new wave of growth for the semiconductor industry, propelled by the high returns on investment generated by AI's economic value. For AI systems, the key metric remains Total Cost of Ownership, while for Silicon, it is Perf/Watt/$."

Subi Kengeri
VP of AI Systems Solutions,
Applied Materials

Rise of custom silicon

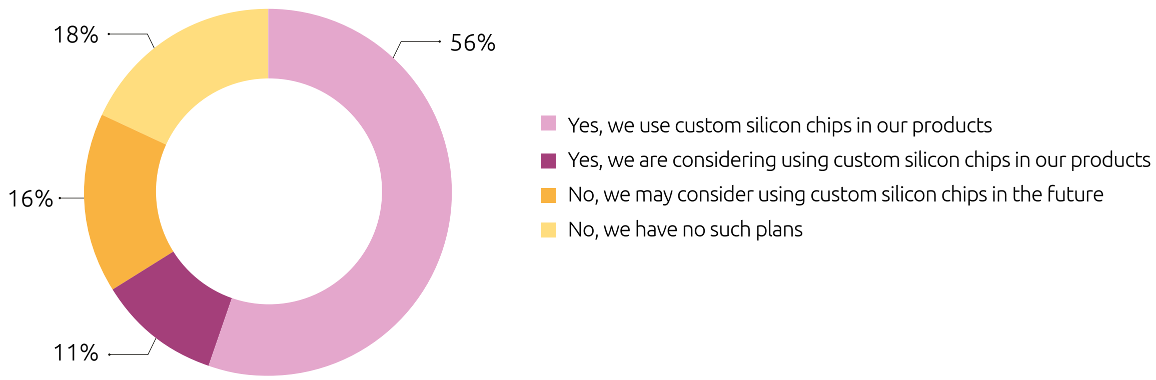
As AI workloads ramp up energy demand, electricity and infrastructure suppliers must adapt quickly to support the growing needs of data centers. Meanwhile, companies such as NVIDIA with their commanding market share in AI chips, continue to exerts significant influence over pricing. The need for optimized performance, energy efficiency, and differentiation in competitive markets such as automotive, high-tech, and medical devices is driving the use of custom silicon. Custom chips allow organizations to better meet specific application demands, reduce costs at scale, and leverage advances in AI and IoT, fueling widespread adoption across industries.

According to our research, 56% of downstream organizations are already using custom chips, while 11% are considering the possibility. A large majority of industrial equipment organizations (85%), and telecom organizations (82%) currently rely on custom silicon chips. Tech giants Microsoft, Amazon, and Meta are also developing in-house chips tailored for AI inferencing.

Figure 5.

Two in three downstream organizations are either already using custom silicon chips or are considering using them in their products

Usage of custom silicon chips in products



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 800 downstream organizations.

"Advanced platforms and software are no longer just enablers but critical differentiators in the semiconductor industry, driving efficiency and scalability in design, manufacturing, and deployment. With the growing complexity of AI, IoT, and edge computing applications, the ability to integrate domain-specific software with hardware accelerators will define leadership. To stay competitive, semiconductor players must embrace co-optimization across the stack, from chip architecture to application interfaces, ensuring they can meet the escalating demands of data-intensive, low-latency markets."

Jiani Zhang
EVP, Chief Software Officer,
Capgemini Engineering

02 Downstream industries express concerns over semiconductor supply

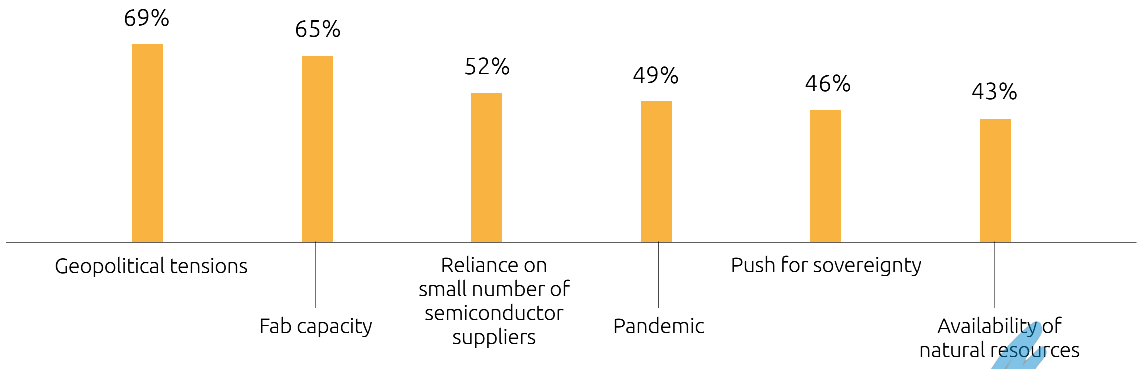
Over half of all downstream organizations are uncertain that the semiconductor industry can cope in 2025

As nations compete for control over vital technologies and resources, geopolitical tensions continue to impact the global semiconductor supply chain. The flow of components, materials, and completed semiconductor products has been hindered by international trade disputes, export restrictions, and tariffs. For example, Taiwan's TSMC, the world's largest semiconductor manufacturer, with a market share of about 55%, is the producer of the world's most advanced chips.[14] Consumer supply networks that rely on TSMC could be seriously disrupted by any military escalation involving China and Taiwan. Deteriorating US-China ties have also given rise to setbacks in the form of prohibitions on certain products and more stringent controls. In 2022, the US introduced export controls that restrict the People's Republic of China's (PRC's) ability to obtain advanced computing chips, develop and maintain supercomputers, and manufacture The The

Figure 6.

More than three out of five organizations believe that geopolitical tensions and inadequate fab capacity impact the reliability of the semiconductor supply chain

Factors impacting the reliability of semiconductor supply chain



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 800 downstream organizations.

advanced semiconductors.[15] These rules were revised in 2023, and in December 2024, the US Commerce Department further expanded the list of Chinese technology companies subject to export controls and included many that make equipment used to make computer chips, chipmaking tools, and software. China, as a response, announced that it is banning exports to the United States of gallium, germanium, antimony and other key high-tech materials with potential military applications.[16] Our research indicates that 69% of downstream organizations believe that geopolitical tensions significantly impact the reliability of the semiconductor supply chain.

Additionally, 65% of downstream organizations consider fab capacity to have a strong impact, while 52% feel that reliance on a small number of semiconductor suppliers impacts their reliance on the semiconductor supply chain.

The COVID-19 pandemic exposed vulnerabilities in the global semiconductor supply chain, disrupting logistics, demand, and production, while heightening concerns over product availability and rising costs. Our research shows that 49% of downstream organizations consider this impact to be ongoing, while 47% had to curtail some product/feature launches due to chip shortages during the pandemic.

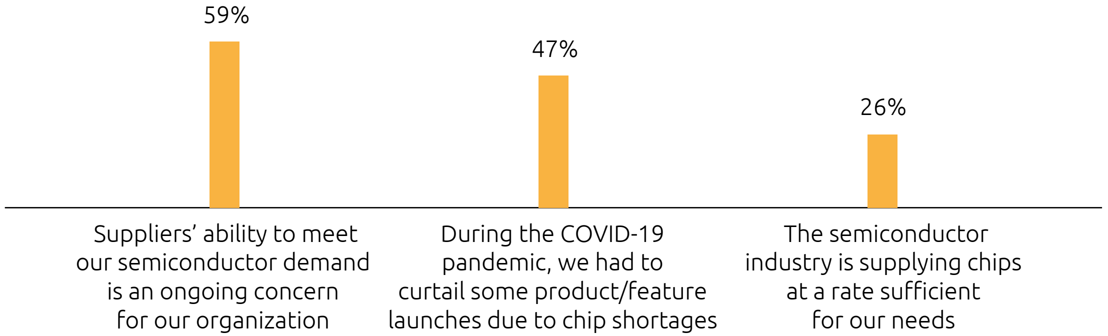
Downstream industries share this uncertainty. Around 59% of downstream organizations believe that suppliers' ability to meet their semiconductor demand is an ongoing concern. Similarly, only around one-quarter (26%) feel that supply is sufficient. This is particularly prominent among sectors such as A&D (14%) and organizations headquartered in Sweden (10%) and the United States (11%).

69%

Percentage of downstream organizations that believe geopolitical tensions impact the reliability of the semiconductor supply chain

Figure 7.
Fewer than three in ten downstream organizations believe chip supply is sufficient

Downstream industries' perception of chip supply/demand



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 800 downstream organizations.

GPU computing and AI/ML acceleration are the most relevant advancements for downstream organizations

Semiconductor technology breakthroughs have spurred innovation in consumer industries, breeding smarter, more efficient products. AI/ML acceleration and GPU processing have the potential to revolutionize downstream operations. GPUs outperform because they provide high throughput and parallel processing, streamlining real-time inference and model training, in particular for AI and ML applications.

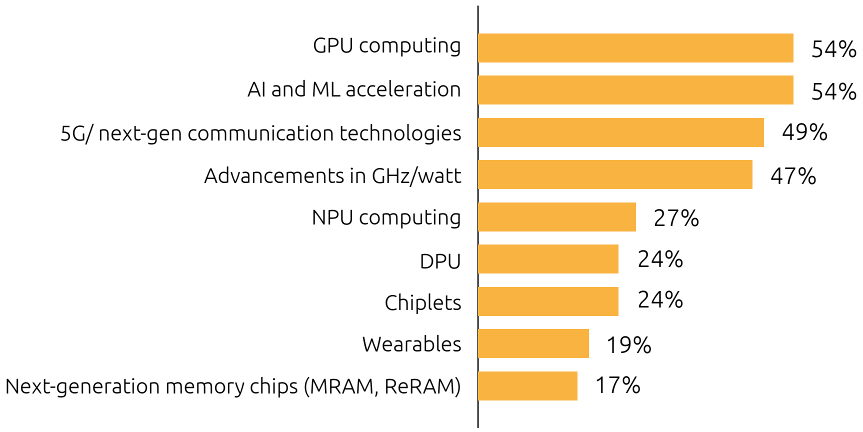
Our research suggests that 54% of downstream organizations (those that rely on fast data processing and AI-powered automation) believe that GPU computing and AI/ ML accelerations are the most relevant advancements for them. Alessandro Miranda, Senior Director of Radio Access Network (RAN) Design and Optimization, at ZTE, explains, *"We need specialized hardware and architectures designed to accelerate processing and optimize algorithms. Graphics* *processing units (GPUs), for instance, which were originally developed solely for rendering images and videos in video games, have now become the cornerstone of AI processing due to their ability to handle parallel data processing efficiently."*

Dell Technologies showcases the practical application of GPU computing and AI/ML acceleration by integrating NVIDIA's AI-ready GPUs, networking solutions, and tools like AI Enterprise and Omniverse with its own hardware and expertise. This collaboration offers communications service providers (CSPs) the tools needed to efficiently run AI workloads across networks.[17]

Figure 8.

More than half of downstream organizations believe that advancements in GPU computing and AI/ML acceleration can bring most value

Most relevant semiconductor advancements for downstream industries



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 800 downstream organizations.

“We need specialized hardware and architectures designed to accelerate processing and optimize algorithms. Graphics processing units (GPUs), for instance, which were originally developed solely for rendering images and videos in video games, have now become the cornerstone of AI processing due to their ability to handle parallel data processing efficiently.”

Alessandro Miranda
Senior Director of Radio Access Network (RAN) Design and Optimization, ZTE

Downstream organizations expect enhanced customization, more comprehensive APIs and SDKs, and stronger security

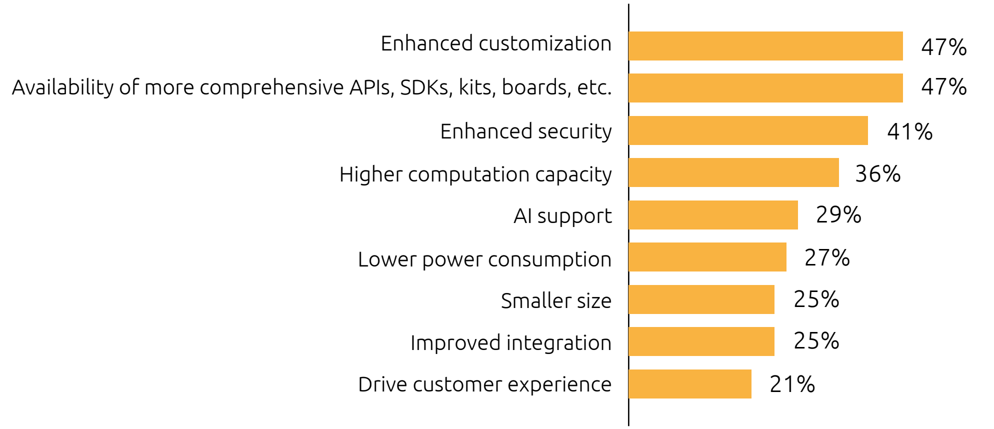
The semiconductor industry has seen significant enhancements in recent years, particularly in areas such as customization, security, APIs, software development kits (SDKs), kits, and boards. These advancements are helping businesses develop more tailored, secure, and efficient products. Our research indicates that 47% of downstream organizations have ranked enhanced customization and availability of more comprehensive APIs, SDKs, kits, boards, etc., in the top three innovations that they look forward to in the semiconductor industry, while 41% have ranked enhanced security among the top three.

Companies such as Intel[18] design custom ASICs for their CPUs and GPUs, integrating both digital logic and analog components to meet the specific performance and power requirements of their products.

Figure 9.

Nearly half of downstream organizations are looking for enhanced customization and more comprehensive APIs and SDKs

Top ranked technology innovations/improvements desired by downstream organizations



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 800 downstream organizations.

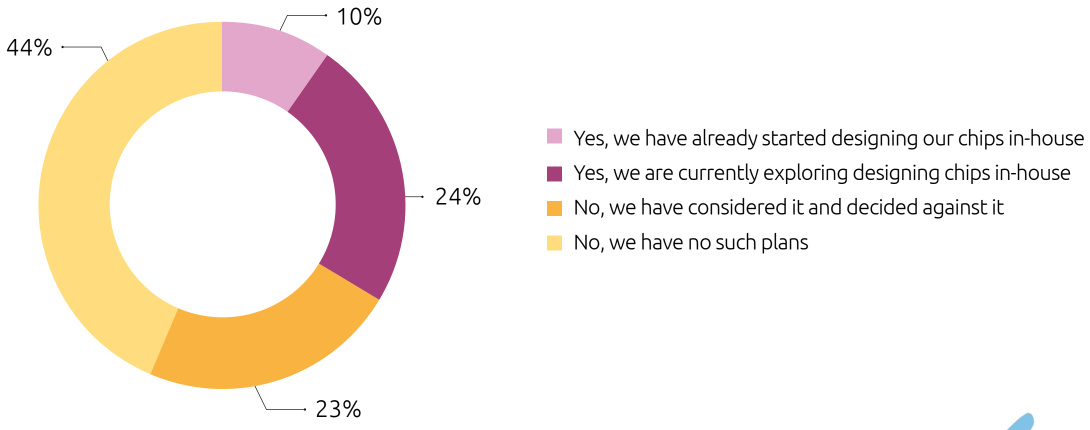
One in three downstream organizations is designing chips in-house (or currently exploring the option)

In-house chip design allows organizations to tailor semiconductors to their unique specifications, minimizing reliance on external vendors and preserving control over their intellectual property (IP), while boosting speed, power efficiency, and compatibility with other hardware and software. Because tech giants use SoCs and custom processors to differentiate their goods, this strategy has become more popular. As part of a broad initiative to integrate AI capabilities into its devices, Apple will implement several of its future AI features using data centers furnished with its in-house processors.[19] Our research indicates that 34% of downstream organizations are either already designing chips in-house or are exploring options in this area.

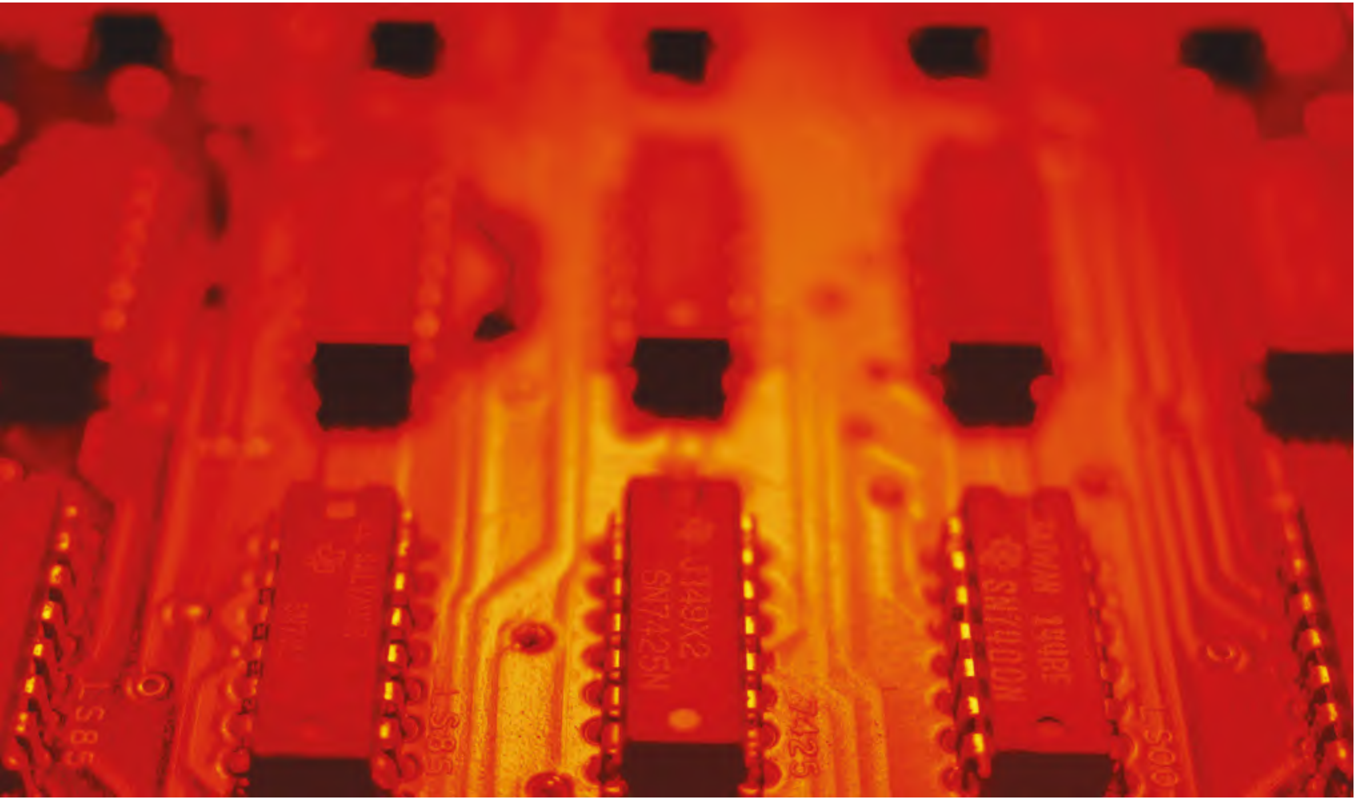
Figure 10.

More than three out of ten downstream organizations are either already designing their chips in-house or are considering it

Has your organization considered/been considering designing your own semiconductor chips?

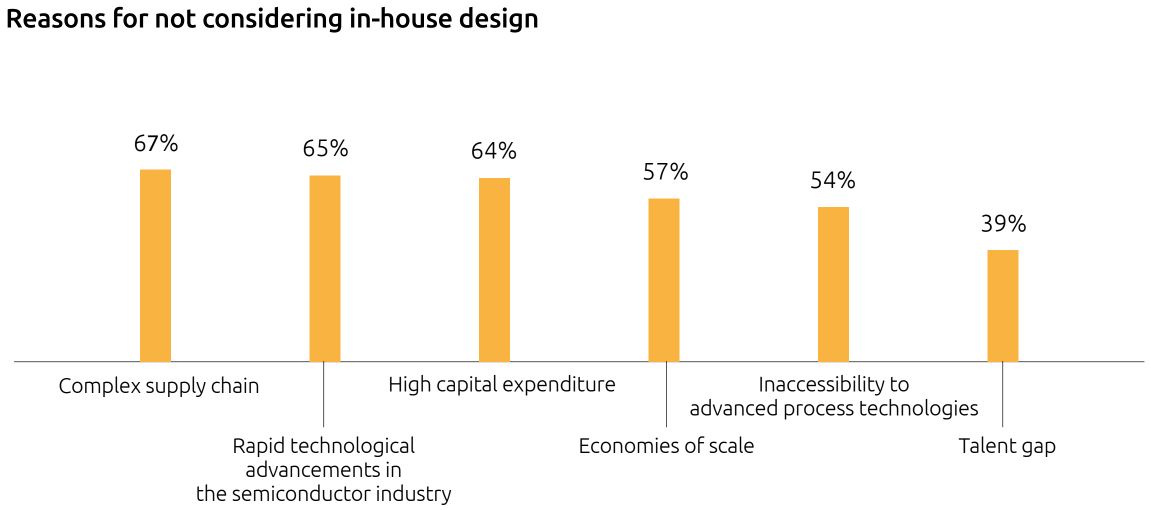


Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 800 downstream organizations.



However, developing a complex design cycle for in-house chips requires a significant investment in talent, cutting-edge design tools, and testing equipment. Supply chain issues, including securing raw materials and managing production, further complicate these choices. Our research suggests that 23% of downstream organizations have decided against in-house chip design, while 44% have not considered it.

Figure 11.
More than three out of five downstream organizations consider supply chain complexity, rapid technological advancements and high capital expenditure as primary reasons to decide against in-house chip design



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 529 downstream organizations.

In today's feverishly competitive tech market, the opportunity to control the hardware ecosystem is a strong selling point. Although semiconductor design organizations can fall back on decades of experience and IP, they should avoid complacency in a changing industry. The semiconductor and information technology industries are likely to continue to develop in tandem as a result of Meta, Microsoft, Apple, Google, Amazon, and other significant tech companies switching to in-house semiconductor chip design in order to reduce supply chain risk, capitalize on in-house innovation, and keep costs low.

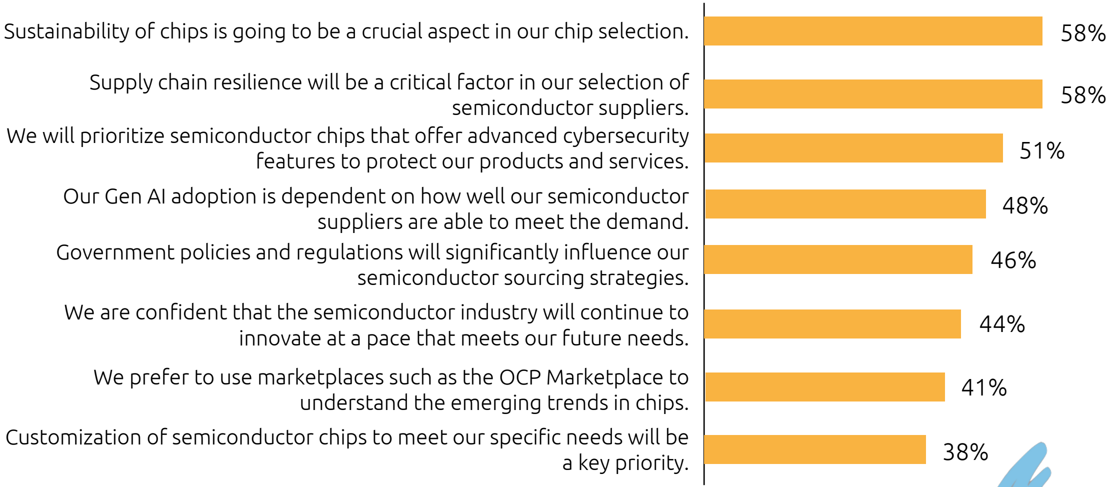
Over the next two years, sustainability, supply chain resilience, and security will be customers' top priorities

Ensuring a sustainable, flexible, and resilient supply chain involves adopting eco-friendly practices. Volkswagen,[20] for instance, reorganized its procurement with Cross Operational Management Parts & Supply Security (COMPASS) to address the daily shortage of semiconductors and questionable supply chain resilience. Security is equally important, as the semiconductor industry is a prime target for cyber threats and IP theft. Organizations must implement stringent security measures to safeguard the integrity of the supply chain and boost customer trust. Together, these elements form the backbone of a reliable, future-proof semiconductor industry. Our research shows that 58% of downstream organizations cite sustainability and supply chain resilience as top priorities, while 51% will prioritize cybersecurity for the coming two years.

Figure 12.

More than half of downstream organizations plan to prioritize chip sustainability, supply chain resilience, and cybersecurity features in the next two years

Do you agree with the below statements for your organization over the next two years?



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 800 downstream organizations.

03 The semiconductor industry is innovating, but softwarization remains a challenge

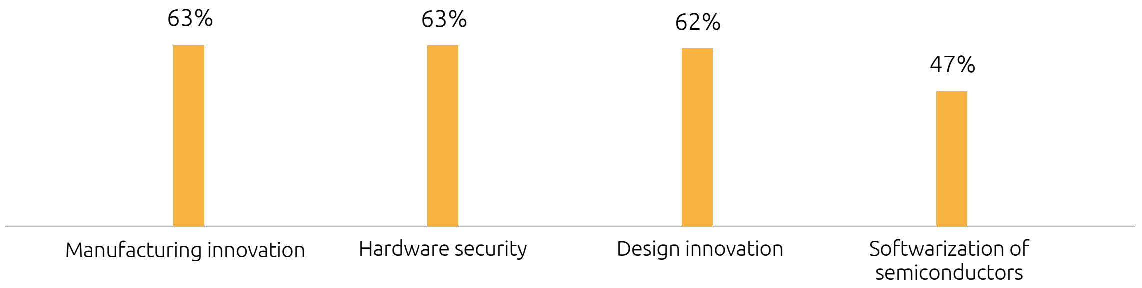
Figure 13 highlights the semiconductor industry's design and manufacturing innovation, with nearly two in three organizations rating their abilities in these areas as high. Similarly, hardware security remains a strong area, with nearly two in three excelling in it.

> *"Chiplets offer a way to address the complexities stemming from Moore's law. They enable the integration of diverse technologies, which is particularly beneficial for AI applications that demand increased computational power."*

Gregg Bromley
Director, AMD

Figure 13.
While the semiconductor industry excels in innovation and security, softwarization poses a challenge

Organizations that rated their abilities high in the below areas



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 250 semiconductor organizations.

The semiconductor industry continues to excel in design and manufacturing innovation

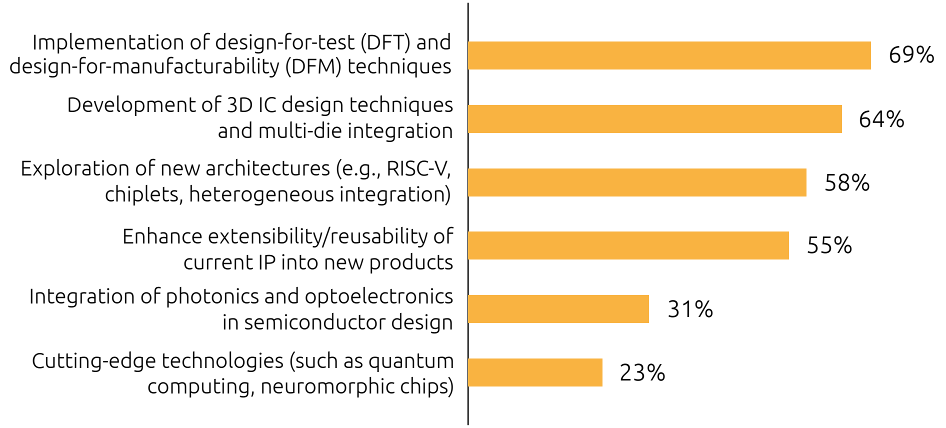
Design innovation

The semiconductor industry is turning to advanced architectures and methodologies to meet demand from downstream organizations for increasingly powerful components. Several organizations already rely on the flexibility of chiplet technology. Having already implemented multi-chiplet architecture for its CPUs and data center GPUs, AMD has filed a patent for a GPU design that is split across multiple GPU chiplet sets. This modular GPU design enables scalable adjustment of GPU resources and performance to match different operating modes.[21] Our research shows that the development of 3D IC techniques and multi-die integration is one of the top three areas of focus for semiconductor organizations. The industry is also exploring open-source architectures such as RISC-V, enabling cost-effective customization. (Designing for testing and manufacturability nevertheless remains the top focus area.)

Figure 14.

Designing for testing and manufacturability remains the top focus area for semiconductor design

Please select the top three areas where your organization intends to invest to enhance design innovation/engineering in the next two years



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 250 semiconductor organizations.

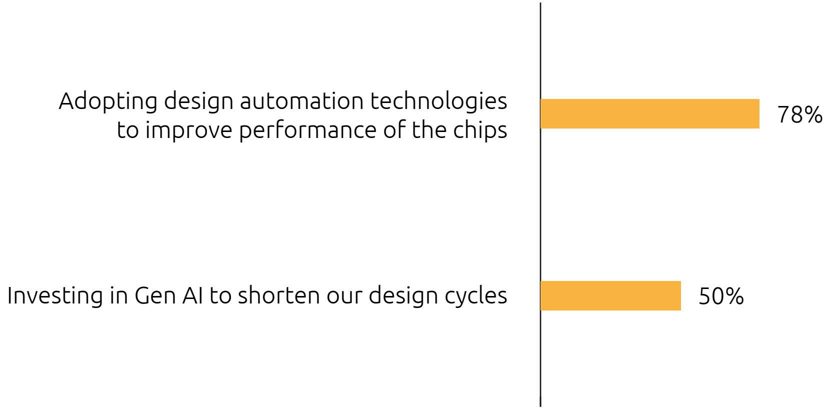
In addition to AI and ML, around half of semiconductor design organizations use Gen AI to shorten their design cycles. Thy Phan, a Senior Director at Synopsys says, *"EDA tools are leveraging AI to enhance PPA (performance, power, area) and development time by automating iterative design processes. These tools can concurrently optimize performance, power, and area, which are key factors of chip design, to find the best balance among them so that the specifications can be met or exceeded."* NVIDIA's engineers have also created a custom LLM for internal use, which is trained on internal data to generate and optimize software and assist human designers. The LLM serves multiple purposes, including generating code snippets in two design languages and automating bug monitoring.[22] Bill Dally, NVIDIA 's Chief Scientist, says: *"This marks an important first step in applying LLMs to the complex work of designing semiconductors. It shows how even highly specialized fields can use their internal data to train useful generative AI models."*

Our research further revealed that the industry is expecting its R&D budget to increase by around 10% over the next two years. In 2023 alone, Intel spent $16.5 billion on R&D, followed by Qualcomm ($8.9 billion), NVIDIA ($7.3 billion), and AMD ($5.9 billion).[23]

Figure 15.

One in two semiconductor organizations is investing in Gen AI for design

Is your organization investing in the design innovations below?



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 167 IDMs, fabless design firms and EDA firms.

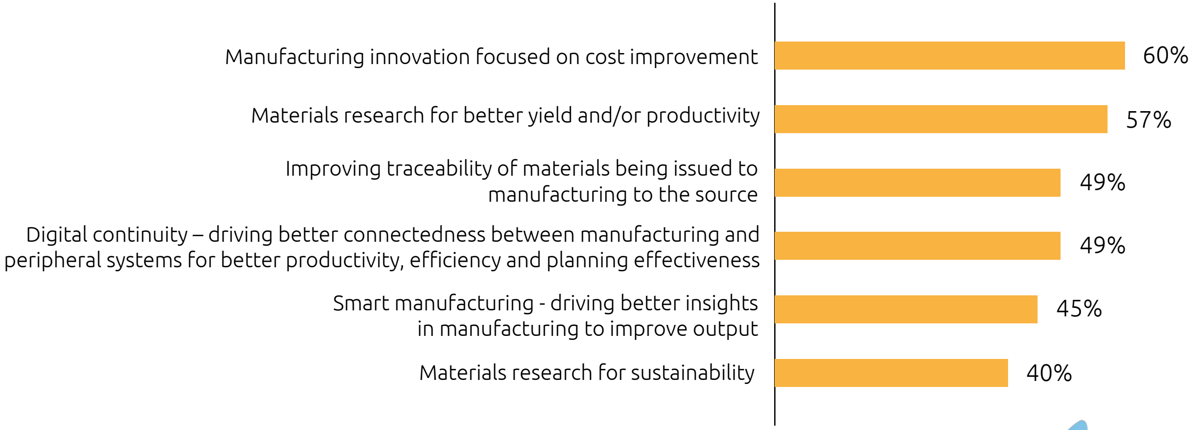
Manufacturing innovation

To keep up with Moore's law, semiconductor manufacturers continue to push the boundaries of physics. Extreme ultraviolet (EUV) lithography, for instance, employs light with shorter wavelengths (13.5 nm) to create intricate patterns on silicon wafers. This allows the production of chips with transistors as small as 3 nm or 5 nm, meaning manufacturers can pack more transistors onto a single chip. With ASML being the sole supplier of these EUV machines,[24] and each machine costing $150 million,[25] semiconductor manufacturers face a bottleneck in their supply chains. Our research shows that cost improvement is the top area of focus for 60% of the industry, while 57% are turning to materials research for better yield. Antimonide- and bismuthide-based semiconductors are employed in advanced infrared sensors for medical and military applications. Similarly, pyrite, which is abundant and non-toxic, can replace the rare earth element cadmium telluride, currently widely used in solar cells.[26] Subi Kengeri from Applied Materials, says, *"Organic substrates tend to have mechanical reliability and yield challenges beyond a certain size. To surpass a specific size, such as an 80 x 80 mm monolithic substrate, alternative materials are needed for packaging. This is why the industry is turning to advanced substrates such as glass, which offers more rigidity."*

Figure 16.

Lowering costs and materials research are the top focus areas for semiconductor manufacturing

Please select the top three areas where your organization intends to invest in the next two years to enhance manufacturing

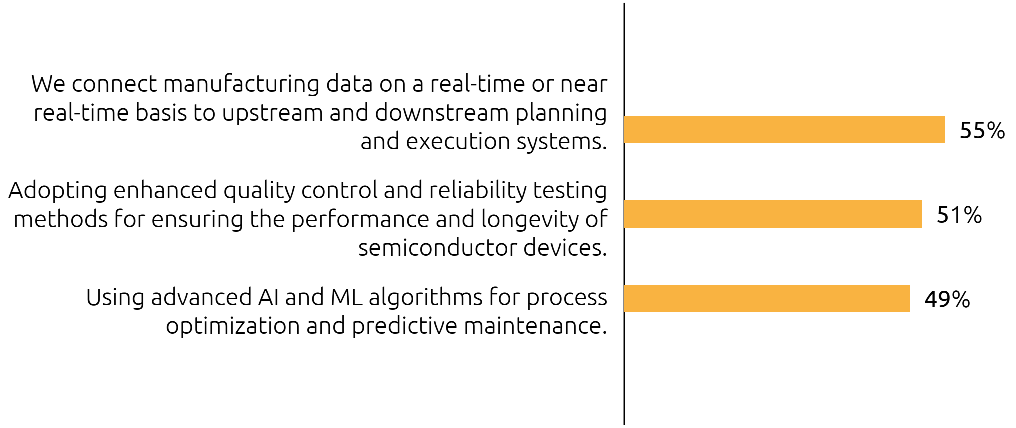


Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 250 semiconductor organizations.

Our research has found that more than half of semiconductor manufacturers use techniques such as real-time data sharing across upstream and downstream systems to improve agility and enhance quality control and reliability testing. TSMC has integrated intelligent mobile devices, Internet of Things (IoT) technologies, augmented (AR) and mixed reality (MR) applications, and mobile robots into its intelligent automated material handling systems (AMHS). This integration consolidates wafer data collection and analysis and efficiently utilizes resources, maximizing manufacturing effectiveness.[27]

Figure 17.
Nearly half of all semiconductor manufacturers rely on AI and ML

Please indicate whether your organization uses the manufacturing innovations below:



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 182 integrated device manufacturers (IDMs), foundries, outsourced semiconductor assembly and test (OSAT) firms and semiconductor capital equipment firms.

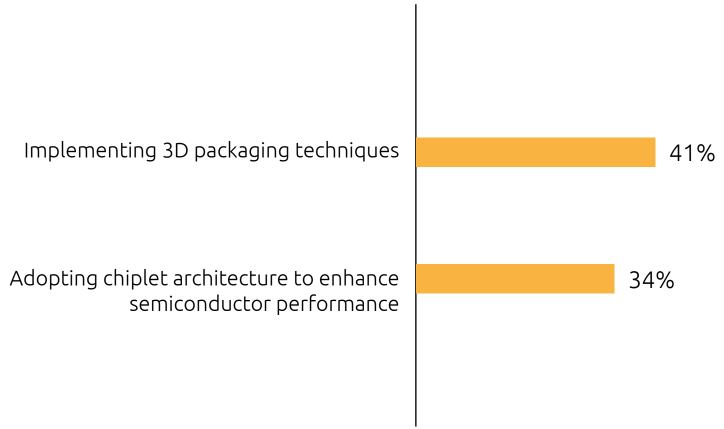
Packaging innovation

TSMC's Chip-on-Wafer-on-Substrate (CoWoS) packaging technology allows 2.5D and 3D stacking of components, streamlining efficiency.[28] TSMC is doubling its CoWoS capacity for 2025, with NVIDIA set to receive 50% of this expanded supply, although the Taiwanese manufacturer still anticipates that AI demand will continue to outpace supply capacity.[29] Around 40% of organizations in our research are implementing 3D-packaging techniques, while one in three focuses on chiplets. Omer Dossani, VP of Global Test Services, at Amkor Technologies, says, *"Silicon continues to grow increasingly complex. Packaging has become more critical than ever as we strive to keep pace with Moore's law. The challenge lies in managing high power demands while addressing space constraints simultaneously."* Gregg Bromley, a Director at AMD, adds, *"Chiplets offer a way to address the complexities stemming from Moore's law. They enable the integration of diverse technologies, which is particularly beneficial for AI applications that demand increased computational power."*

Figure 18.

Around two in five organizations deploy 3D packaging for semiconductors

Please indicate whether your organization uses the packaging innovations below:



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 149 IDMs and OSAT firms for the top statement and N = 197 IDMs, fabless design firms, OSAT firms and EDA firms for the bottom statement.

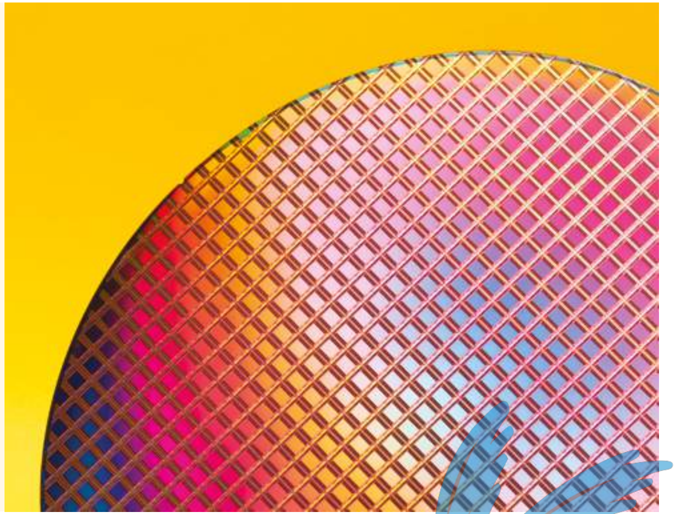
Hardware security remains a critical area of focus

The semiconductor industry's highly complex and interdependent supply chain spans suppliers of IP, equipment, and materials. This means that any vulnerability can jeopardize the entire ecosystem that depends upon semiconductor supply. A US-based semiconductor firm attributed a potential loss of $250 million in sales in 2023 to a cybersecurity incident at one of its suppliers.[30] The industry also faces potential data breaches, IP theft, physical security issues, and attacks on the manufacturing process. In another notable incident, researchers identified a vulnerability in a series of chips from a prominent company that allowed attackers to extract secret encryption keys during cryptographic operations (known as a "side-channel attack"). The vulnerability stems from the microarchitectural design of the silicon; to mitigate it would require a significant trade-off in performance.[31]

Our research has shown that three in five semiconductor design organizations prioritize secure firmware and software integration to prevent vulnerabilities. Additionally, 58% also use cryptographic protection, such as hardware-based encryption, to safeguard the data within the chip. Sam Geha from Infineon comments, *"We are observing significant industry trends around security, with customers increasingly requesting features like secure boot and secure firmware over-the-air updates. There is a growing need for enhanced security across various applications, and we believe this trend will only continue to strengthen."*

60%

Percentage of semiconductor organizations focusing on secure firmware and software integration when designing hardware-based chip security

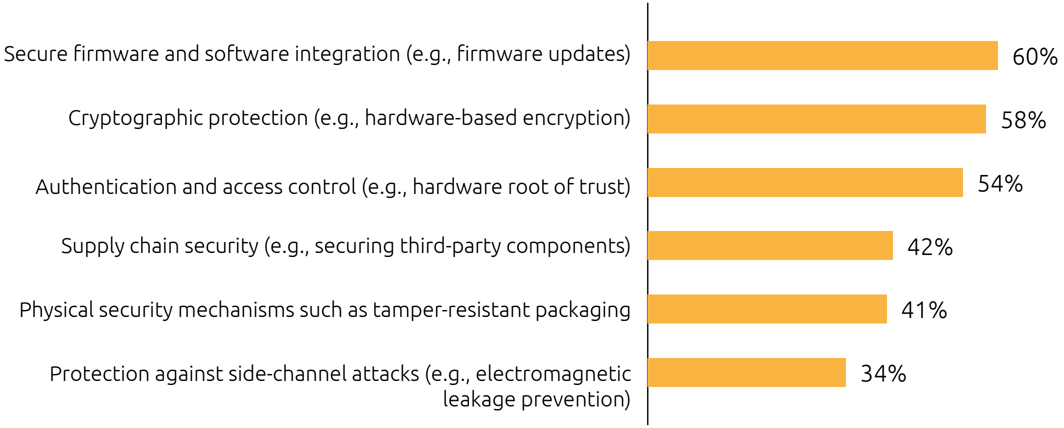


“We are observing significant industry trends around security, with customers increasingly requesting features like secure boot and secure firmware over-the-air updates. There is a growing need for enhanced security across various applications, and we believe this trend will only continue to strengthen."

Sam Geha
EVP of IoT, Compute and Wireless business, Infineon Technologies

Figure 19.
Nearly three in five semiconductor design organizations focus on cryptographic protection to enhance chip security

Which of the below security measures are considerations when designing hardware-based chip security?



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 167 IDMs, fabless design firms and EDA firms.

Steady progress in softwarization, but monetization challenges persist

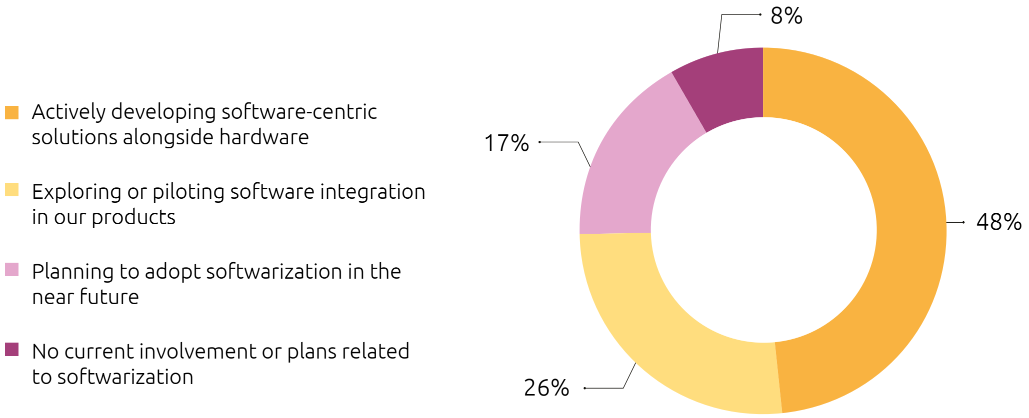
Softwarization allows for the creation and updating of custom software for different industries without needing to redevelop hardware. This leads to agile adaptation, extended product lifecycles, and tailored solutions across various applications, with significant cost and efficiency benefits. Our research found that nearly half of all semiconductor design firms are actively developing software-centric solutions, while another quarter are exploring or piloting the integration of software. For instance, NVIDIA's CUDA [Compute Unified Device Architecture] platform supports parallel processing, using GPU acceleration to develop and optimize high-performance applications.[32] Similarly, Qualcomm is working with AWS to enable downstream automotive organizations to integrate cloud technology for simulating, testing, validating, and verifying new features and functions prior to traditional hardware testing, expediting the development process.[33]

Brett Bonthron, Industry Leader for Tech and Digital from Capgemini, adds, *"Softwarization is transforming the semiconductor industry, enabling unprecedented flexibility and scalability. By decoupling hardware from software, companies can rapidly adapt to evolving demands, optimize performance across diverse applications, and unlock new capabilities through updates. This shift not only accelerates innovation but also fosters collaboration across the ecosystem, as software becomes the bridge connecting hardware excellence with cutting-edge functionality."*

Figure 20.

Nearly half of semiconductor organizations are actively developing software-centric solutions

What is your company's current level of involvement in the softwarization of chips?



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 165 IDMs and fabless design firms.

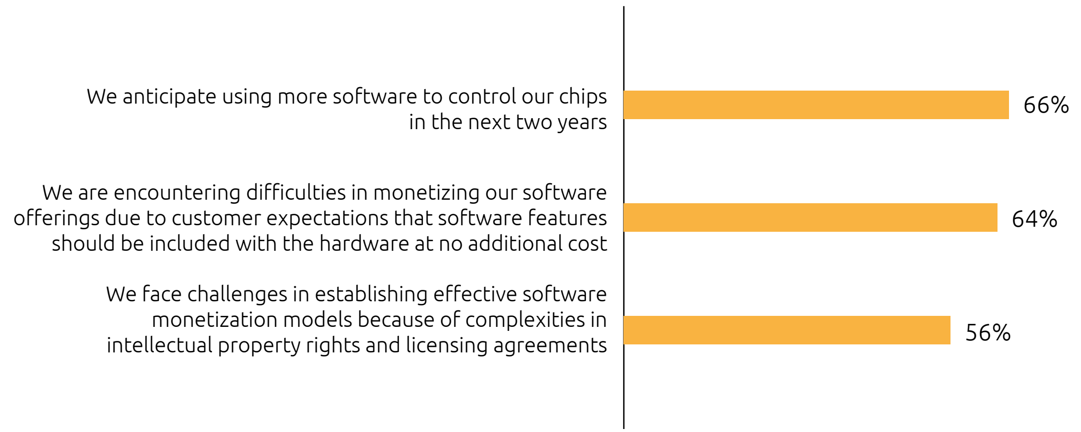
Our research found that two in three semiconductor organizations anticipate increasing their use of software in the next two years. However, around 60% of organizations find monetizing these software-based offerings a challenge, due either to customer expectations or IP complexities.

66%

Percentage of semiconductor organizations anticipating the use of more software to control chips in the next two years

Figure 21.

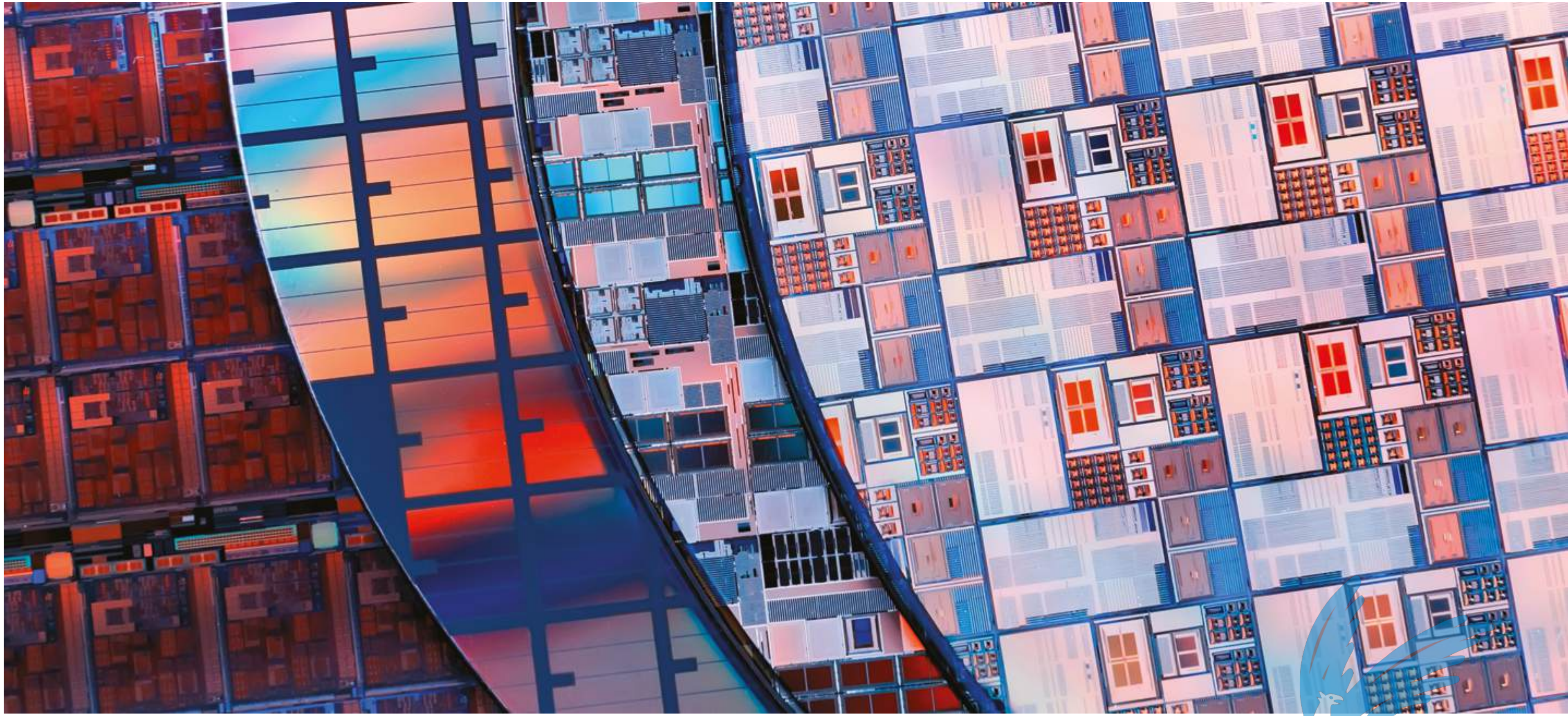
Monetization of software remains a challenge for three in five semiconductor organizations



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 250 semiconductor organizations.

“Silicon continues to grow increasingly complex. Packaging has become more critical than ever as we strive to keep pace with Moore's law. The challenge lies in managing high power demands while addressing space constraints simultaneously.”

Omer Dossani
VP of Global Test Services,
Amkor Technologies



04 Resilience and sustainability gather momentum

The global semiconductor chip shortage in 2020–23 wrought devastation in many other economic areas, underscoring the far-reaching significance of the industry. Factory shutdowns, US trade restrictions on Chinese chip imports, extreme weather and fires disrupting global production, and a neon shortage caused by the Russian invasion of Ukraine all exacerbated supply chain issues. Conversely, the surge in cryptocurrency markets, stay-at-home orders, and the shift to remote work drove significant spikes in demand. Only two out of five organizations rate their resilience as high, while less than half (46%) rate their sustainability achievements as high.

Organizations turn to onshoring and friendshoring to enhance resilience

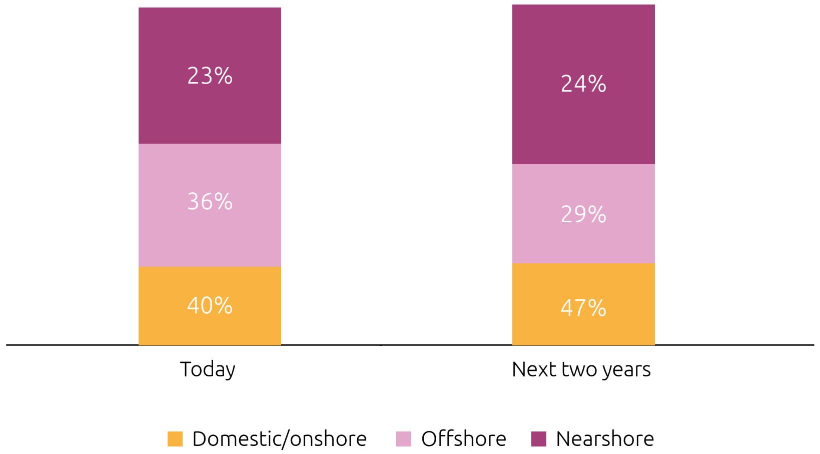
Our research found that the semiconductor industry expects to improve its domestic sourcing from 40% currently to 47% over the next two years to mitigate the risks associated with international logistics. The challenge is reliance on a limited set of suppliers or regions for certain components or pieces of equipment. ASML, for instance, is the only supplier of EUV lithography machines (see page 36).[34] As of 2022, Ukraine was the supplier of 90% of the highly purified, semiconductor-grade neon for chip production in the US industry.[35] On the manufacturing front, TSMC has a market share of above 60% in contract manufacturing and over 90% in advanced chip manufacturing.[36] Similarly, Spruce Pine, a small town in North Carolina, US, is home to two mines that produce more than half of the world's high-purity quartz, and production at these mines was shut down for two weeks due to Hurricane Helene in 2024.[37]

17%

Semiconducor industry's anticipated improvement in domestic sourcing over the next two years

Figure 22.
Semiconductor industry anticipates domestic sourcing to improve by 17% over the next two years

Sourcing capacity by location



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 202 IDMs, foundries, OSAT firms, capital equipment firms, material and subsystem organizations.

Our research also found that the industry anticipates a significant shift to nearshoring over the next two years (around one-quarter of sourcing and one-fifth of manufacturing capacity). Three-quarters of organizations expect to increase US investments, while 59% plan to increase EU investments. In 2022, the US allocated nearly $53 billion in loans, grants, and incentives for US semiconductor research, development, manufacturing, and workforce development via the CHIPS and Science Act, a US federal statute signed into law by President Joe Biden that is intended to strengthen domestic semiconductor production.[38]

The Act awarded $8.5 billion in grants and $11 billion in federal loans to Intel, and grants and federal loans to firms including GlobalFoundries, Micron, TSMC, and Amkor Technology, among others.[39] With up to $6.6 billion in direct funding allocated through the CHIPS Act, TSMC is investing more than $65 billion in three US greenfield fabs. This investment will support the development of chips for AI, high-performance computing, and 5G/6G communications. One of these fabs is scheduled to manufacture chips using

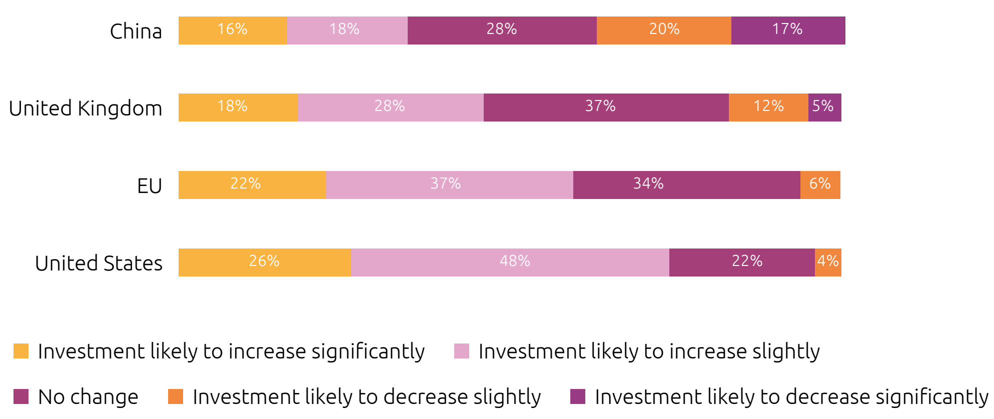
2 nm or even more advanced process technology.[40] Since the introduction of the CHIPS Act, semiconductor organizations have announced more than 90 new US manufacturing projects, representing nearly $450 billion in investment capital.[41]

In a similar vein, the EU in April 2023 announced a 43 billion subsidy plan through the EU CHIPS Act, with the objective of building large-scale capacity and innovation, ensuring the self-sufficiency of the EU, and galvanizing its resilience against future supply crises.[42]

Organizations are also deploying strategies to ensure production continuity. Four in five semiconductor organizations plan to maintain strategic inventory levels, while three-quarters are prioritizing supplier diversification.

Figure 23.

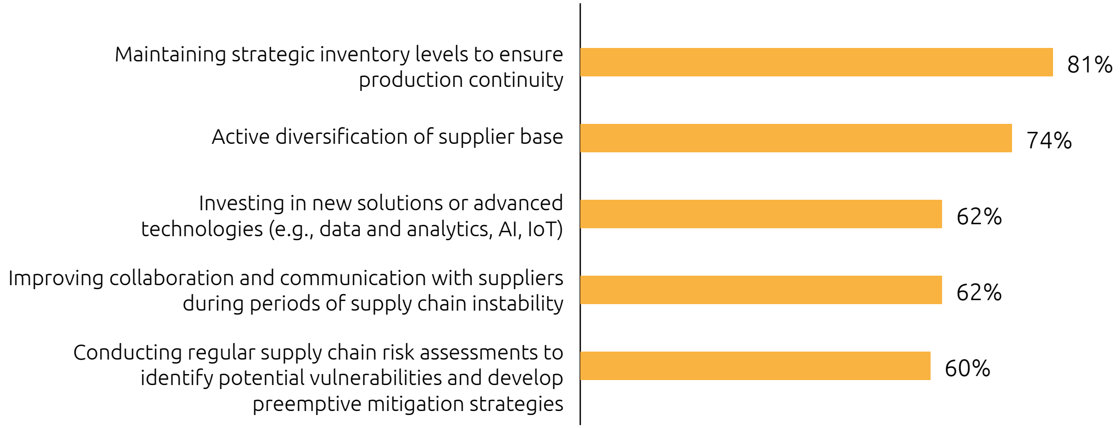
Investment will focus on the US and the EU over the next two years



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 250 semiconductor organizations.

Figure 24.
Three-quarters of semiconductor organizations are focusing on active diversification of their supplier bases

Which resilience strategies is your organization implementing or planning to implement to combat supply chain disruptions?



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 250 semiconductor organizations.

10%

Expected increase of around 10% in semiconductor industry's R&D budget over the next two years

The industry is making strides with various sustainability initiatives

For a number of years, the semiconductor industry has continued its focus on improving power efficiency. Our research found that the semiconductor industry is also focusing on energy conservation, water use reduction, hazardous chemicals management, and waste minimization. Alessandro Miranda from ZTE highlights, *"We are heavily investing in more energy-efficient manufacturing processes, striving to reduce waste and water usage each year. This effort is a global initiative, implemented across all our manufacturing sites worldwide."* Figure 25 shows the top three deployments across each of these categories. Intel, for instance, pledged to achieve net-zero greenhouse gas (GHG) emissions in its global operations by 2040.[43] To achieve this, the organization plans to invest around $300 million by 2030, targeting 4 billion cumulative kilowatt hours (KwH) of energy savings. Additionally, it will launch a cross-industry R&D initiative to identify greener chemicals and develop new abatement equipment. ASML is focusing on circularity and aspires to send zero waste from its operations to landfills or incineration. [44]

Chiplet-based modular design allows for interchangeable component configurations, boosting flexibility while cutting manufacturing costs and waste. By producing its fourth-generation EPYC CPUs with eight separate compute chiplets, rather than a single monolithic die, AMD was able to save ~50,000 metric tons of CO_2e in 2023 by cutting wafer manufacturing – roughly the same as its operational CO_2e footprint in 2022.[45] As chiplets can be stacked in a package, they allow the creation of energy-efficient processors as an alternative to data centers. By upgrading its server processors to AMD EPYC, STMicroelectronics slashed its electricity consumption by 33% while increasing compute performance.

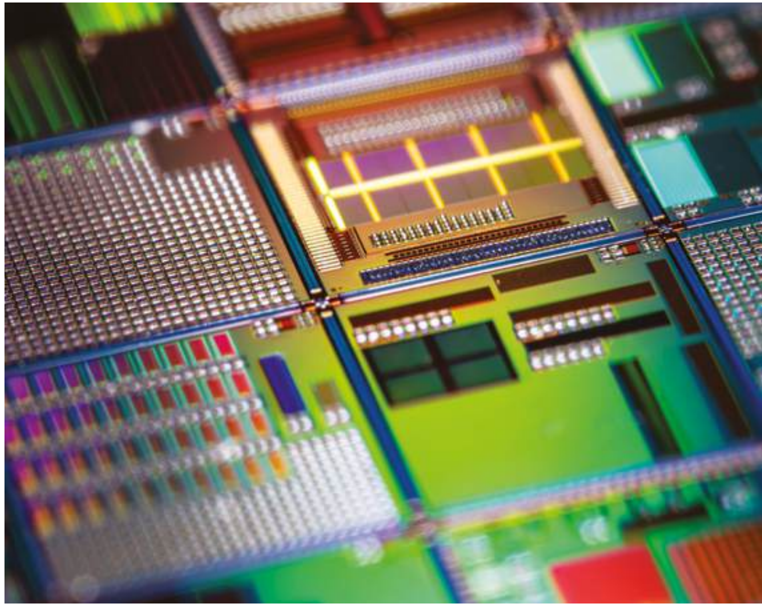


Figure 25.

Top three initiatives currently being deployed in each category

Energy conservation

- Upgraded to energy-efficient machinery and equipment
- Implemented energy-management systems
- Reduced water usage

Reducing water usage and promoting water circularity

- Implemented water recycling and reuse systems
- Modified processes to reduce water consumption
- Achieved zero waste by reusing all wastewater

Managing hazardous chemicals

- Replaced hazardous chemicals with less toxic alternatives
- Implemented closed-loop systems for chemical reuse
- Recycled or reclaimed chemicals for reuse in processes

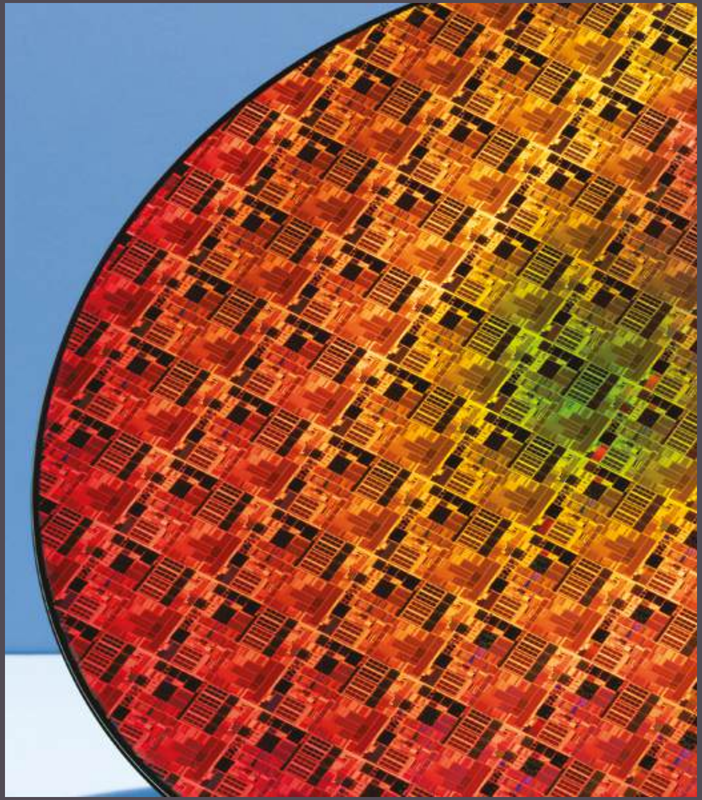
Waste reduction

- Implemented waste-minimization programs
- Partnered with organizations on waste-to-resource initiatives
- Designed products and processes to reduce material use
- Utilized end-of-life products as raw materials (closed-loop recycling)

Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 182 IDMs, foundries, OSAT firms and semiconductor capital equipment firms.

"EDA tools are leveraging AI to enhance PPA (performance, power, area) and development time by automating iterative design processes. These tools can concurrently optimize performance, power, and area, which are key factors of chip design, to find the best balance among them so that the specifications can be met or exceeded."

Thy Phan
Senior Director, Synopsys



Exploring big tech's shift to in-house chip design

Alphabet

- As part of a broader strategy to reduce reliance on external suppliers, Alphabet used AlphaChip technology to design its own TPU AI accelerators, which has been used in several initiatives:[46]
  – TPU AI accelerator has helped Google to drive large-scale AI models and cloud services, including Gemini and Imagen.
  – This AI accelerator has been used to improve the design of each successive processor chips, the latest one being Trillium 6th Generation chips being used extensively in data centers worldwide, which offers 4.7X enhanced computational performance while being 67% more energy efficient as compared to the previous version.[47]
  – TPU AI accelerator has also played a significant role in development of Dimensity 5G system-on-chips for MediaTek, which are used extensively in smartphones.
- AlphaChip AI has also been used in chip design recently, where reinforcement learning (RL) has been used to optimize chip layouts for computers, cutting down design times and boosting performance.
- Google's custom-built Tensor G4 chips are used to power Google AI in Pixel smartphones, and enhance user experience.[48]
- Google has integrated Tensor A1 chips into Pixel Buds Pro 2 earbuds to enhance its noise cancellation capabilities.[49]

In 2024, Google's parent company, Alphabet, allocated a considerable portion of its $13 billion capital expenditure to expanding its technical infrastructure, including data centers and servers, to support production of its custom Tensor Processing Unit (TPU) AI chips.[50]

Amazon

- Amazon has launched its 4th generation general-purpose processor Graviton4, which offers 30% higher computational performance and 75% increase in memory bandwidth compared to its previous version. It is planned to be used in Amazon Elastic Compute Cloud (Amazon EC2) R8g to manage databases, in-memory caches, and large-scale data analytics.[51]
- The company has also launched Trainium 2 for AI training. It offers 4X faster model training speed, 3X memory capacity and 2X performance per watt as compared to the previous version.[52]
- Amazon has developed Inferentia chips, which power Amazon EC2 Inf1 instances for deep learning and generative AI applications. It offers 2.3X higher throughput at up to 70% lower cost per inference.[53]
- Intel and AWS are seeking to expand their collaboration, including a co-investment in custom chip design, under a multi-year, multi-billion-dollar framework.[54] These chips serve more than 50,000 customers, including Formula 1, Pinterest, and SAP. Graviton4 also enhances security and energy efficiency, playing a critical role in AWS's strategy to lower costs for its cloud clients.

In 2024, Amazon's planned capital expenditure rose sharply to $75 billion, driven largely by Gen AI investments, and up from $48.4 billion in 2023.[55] The organization expended $30.5 billion in CapEx in H1, principally on infrastructure upgrades to accommodate AI and ML demand.[56]

Apple

- In November 2020, Apple announced Apple Silicon, in-house-designed chips for iPhone, iPads and Apple watches.[57]
- Apple launched the M1 chip in 2020, following with the M2 and, recently, the M3 chip, offering substantial improvements in performance and energy efficiency with each iteration. The M3 chip, used in devices such as the MacBook Air, marks a leap in chip design, integrating 3-nanometer technology for better computational power and efficiency.[58]
- TSMC is producing Apple's custom-designed chips, including the A16 and A17 series, at the former's new Arizona plant, as part of Apple's broader strategy to localize its US semiconductor manufacturing.[59]
- In 2023, Apple confirmed that Amkor would package Apple silicon chips produced at a nearby TSMC facility, as part of a US manufacturing push.[60]

Meta

- In order to avoid excessive reliance on external vendors, Meta is working on building Artemis chips in-house for its data centers customized to support the AI development programs, especially the inferencing part.[61]

"Semiconductors are enabling the deployment of Artificial Intelligence (AI) across enterprises. However, unlocking AI's full potential requires collaboration among companies throughout the semiconductor value chain, as well as partnerships with downstream industries. This collective effort will accelerate innovation, reduce risks, and eliminate inefficiencies, ultimately lowering costs and enhancing overall performance for businesses."

Sanjiv Agarwal
Semiconductor Industry Leader, Capgemini

- Meta's first-generation custom AI chip, MTIA [Meta Training and Inference Accelerator] is tailored for in-house AI workloads, and offers greater efficiency and computational power than other CPUs in the market.[62] It is focused on reducing power consumption while handling large-scale AI workloads, using TSMC's 7-nm process. It is current working on second-generation MTIA to be deployed across its data centers, and will use high-bandwidth memory to boost performance.[63]
- Meta has deployed 3D chip integration technology in AR devices, which increases computational power without increase size of the devices.[64]

In 2024, Meta announced plans to spend $38 billion–$40 billion on AI infrastructure, marking a significant increase from previous guidance. This investment is earmarked for enhancing its LLM capabilities, including training Llama 3, and readying its infrastructure for more advanced AI workloads. Meta's chip-development efforts are central to its independence strategy.[65]

Microsoft

- Microsoft has developed Azure Maia 100, Microsoft's first in-house AI accelerator, to support its cloud-based AI workloads and optimize Azure AI infrastructure.[66] Microsoft added a rack-level, closed-loop liquid cooling system, which smooths integration of these Maia 100 systems into Microsoft's existing data center infrastructure, without increasing their footprint.[67]
- It has also built Arm-powered Azure Cobalt CPU, a 128-core chip, to support general cloud services on Azure, such as Teams, SQL servers, and plan to offer virtual machines for different kinds of workloads.[68]
- Microsoft has signed a $15 billion deal with Intel to produce custom chips.[69]

Partnerships are taking the industry forward

Partnerships allow semiconductor organizations to share costs, mitigate reducing financial risks, and more effectively meet growing demand for specialized components. In our research, we identify several industry collaborations and highlight five themes emerging from these partnerships (see figure 26).

Sanjiv Agarwal, Semiconductor Industry Leader from Capgemini, adds, *"Semiconductors are enabling the deployment of Artificial Intelligence (AI) across enterprises. However, unlocking AI's full potential requires collaboration among companies throughout the semiconductor value chain, as well as partnerships with downstream industries. This collective effort will accelerate innovation, reduce risks, and eliminate inefficiencies, ultimately lowering costs and enhancing overall performance for businesses."*

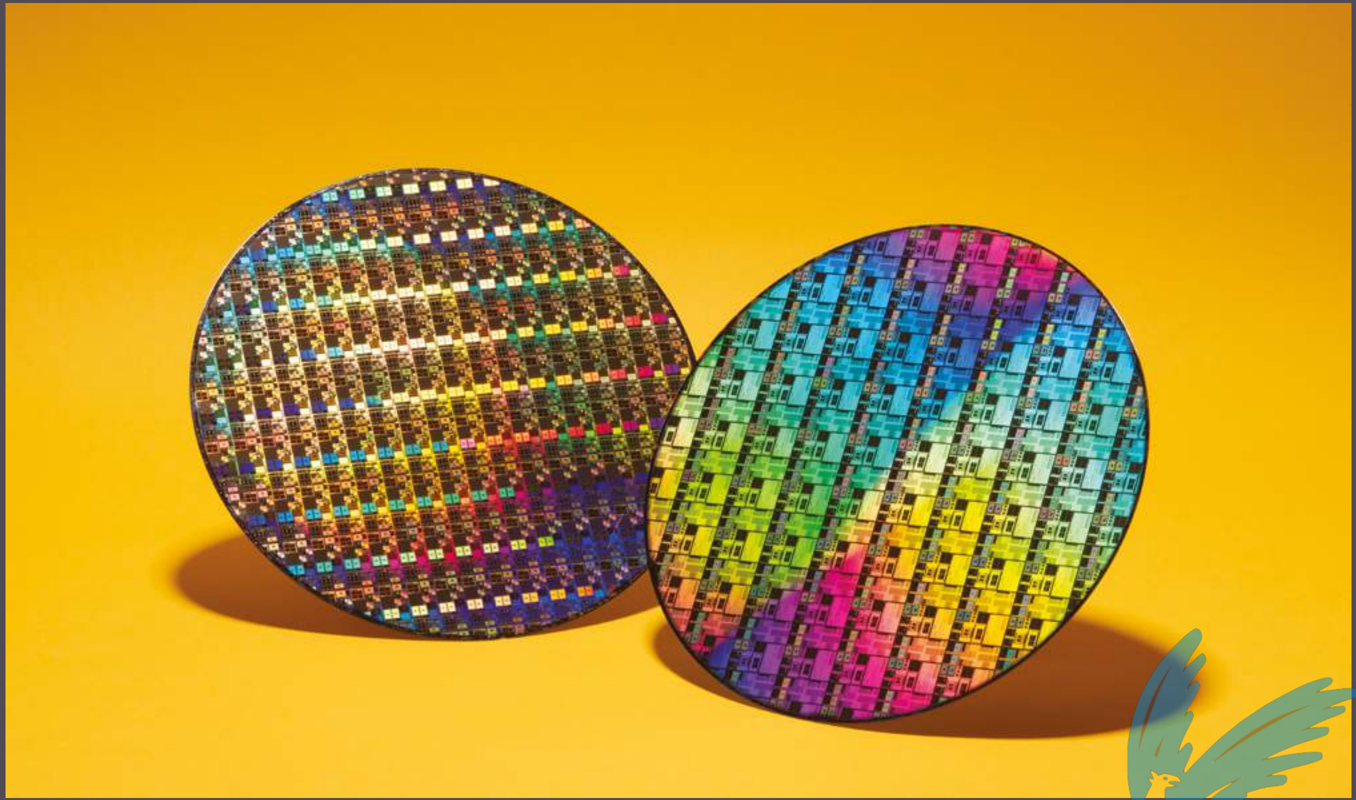


Figure 26.
Five major themes of partnerships in the semiconductor industry

Partnerships

Supply chain

- **TSMC, Bosch, Infineon, and NXP** established the European Semiconductor Manufacturing Company (ESMC) joint-venture in Dresden, Germany, to bring an advanced semiconductor manufacturing presence to Europe.
- **GlobalFoundries and STMicroelectronics** announced the creation of a new, jointly-operated, high-volume semiconductor manufacturing facility in Crolles (France).
- **TSMC** will build its second factory in Japan with backing from **Sony and Toyota.**

Innovation

- **Intermolecular** is working with **Merck** to prototype and test new materials, accelerating innovation and commercial production.
- **Synopsys and Microsoft** formed a multi-year collaboration to apply hyperscale cloud and AI to electronic design automation (EDA), including enhanced chip development.
- **Samsung Semiconductor Europe** showcased its latest foundry innovations for high-performance, low-power semiconductors targeting the AI and automotive industries. These include advanced Gate-All-Around (GAA) technology, Bipolar-CMOS-DMOS (BCD) solutions, and eMRAM, enabling industries to fully leverage AI potential.

Source: Capgemini Research Institute analysis.

To be cont.

Partnerships

Sustainability

- **Intel and Siemens** have collaborated to advance factory efficiency and sustainability across Scopes 1 to 3 of the value chain, while supporting global industry ecosystems.
- **Infineon and Amkor** have partnered to engage with common suppliers to help them drive decarbonization strategies and boost greener supplier engagement.
- **Merck in partnership with Intel** funded a new three-year academic research program in Europe to leverage AI and machine learning technologies to drive breakthroughs in innovative and more sustainable semiconductor manufacturing processes and technologies.

Softwarization

- **Intel and Advanced Micro Devices,** among others, will partner to ensure compatibility across their chips. Partners will contribute technical insights to identify "essential functions and features" for AMD and Intel chips.

Cybersecurity

- **TSMC** has named **TXOne**, a cybersecurity provider to the semiconductor industry, as a collaborator on OT cybersecurity.
- **AMD** has teamed with **Microsoft** and OEMs to implement layered PC defences with AMD Ryzenä processors, helping to improve security from the endpoint and the edge to the cloud.
- **NXP** leverages **Scalys** as a cyber security specialist, providing the organization with the right support for cybersecurity innovations.

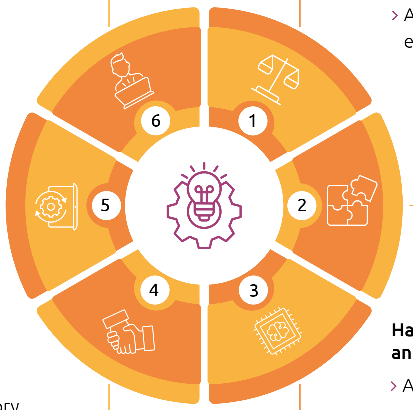
05 How the semiconductor industry can capitalize on emerging opportunities

26%

Percentage of downstream organizations which agree that the semiconductor industry is supplying chips at a rate sufficient for their needs

To navigate the bountiful but complex landscape of semiconductor opportunities, we propose the following recommendations for the semiconductor industry, based on our research, in-depth interviews, and experience.

Figure 27.
Focus areas for the semiconductor industry



1. Invest in advanced manufacturing technologies and innovation
- Adopt next-generation fabrication methods
- Accelerate research and development in emerging technologies

2. Galvanize semiconductor supply chains through diversification and adoption of eco-friendly processes
- Diversify and secure supplier networks
- Implement sustainable manufacturing practices
- Minimize ecological footprint

3. Harness AI and Gen AI for advanced chip design and manufacturing
- Automate design processes using AI
- Use intelligent systems to improve production efficiency and yields
- Develop AI-optimized semiconductor products

4. Align with government policies and foster open innovation
- Coordinate strategies with regulatory frameworks
- Collaborate and share innovation

5. Protect intellectual property and enhance security measures
- Strengthen cyber defenses and protect data integrity
- Safeguard proprietary technologies

6. Adopt open standards to foster innovation
- Promote industry-wide collaboration for cross-platform interoperability
- Encourage open-source design and development in semiconductor innovation

Source: Capgemini Research Institute analysis.

Invest in advanced manufacturing technologies and innovation

Adopt next-generation fabrication methods

To remain competitive in a changing market, semiconductor manufacturers must innovate continuously to create more powerful, efficient chips at lower costs. The following technology-driven manufacturing techniques can support this drive:

3D chip stacking: Vertical integration allows semiconductor organizations to increase the density and performance of chips without increasing the physical footprint. 3D chip stacking supports heterogeneous integration, allowing the combination of many chip types in a single versatile and powerful package. The method enhances performance and bandwidth, creating new opportunities for advanced computing applications. Our research suggests that 64% of semiconductor organizations will prioritize the development of 3D IC design techniques and multi-die integration in the next two years.

Automated fabrication processes: AI can help businesses reduce the incidence of human error by automating complex and repetitive tasks, resulting in more dependable, higher-quality semiconductor products. Automation also allows faster production cycles to help manufacturers meet rising demand with enhanced accuracy. Almost half (49%) of semiconductor manufacturers say they are already using advanced AI and ML algorithms for process optimization and predictive maintenance.

Advanced EUV lithography: This technique supports the creation of incredibly fine patterns on silicon wafers and the production of smaller, more complex semiconductor chips and is consequently crucial to the development of technologies such as 5G networks, IoT, and AI. Higher transistor densities greatly improve electronic device performance and energy efficiency. Additionally, by reducing the required number of patterning steps, EUV lithography reduces complexity and cost. Major semiconductor manufacturers such as Intel[70] and TSMC[71] have invested heavily in EUV technology.

Accelerate R&D in emerging technologies

Showcasing a strong commitment to R&D ensures long-term growth and resilience. Semiconductor organizations should invest in the following R&D initiatives:

Neuromorphic computing hardware: By imitating the structure of the human brain, neuromorphic computing hardware can process complex, unstructured data more effectively than conventional hardware. Parallel computing with neuromorphic circuits may open up opportunities in fields such as robotics, driverless cars, and healthcare. Additionally, because neuromorphic computing can handle data more effectively and requires less processing power, it drastically lowers AI's carbon footprint. Semiconductor organizations can obtain a competitive advantage through early adoption of this technology. Intel has recently announced its neuromorphic system, code-named Hala Point. This utilizes Intel's Loihi 2 processor to support research for future brain-inspired AI and tackles challenges related to the efficiency and sustainability of today's AI.[72]

Advanced silicon photonics integration: Advanced silicon photonics integration can help modern computing and communication systems meet the increasing demand for quicker and more effective data transfer. Photonic devices can offer greater bandwidths and reduced latencies, boosting operations at 5G networks, data centers, and AI applications.

Process node miniaturization: Miniaturization allows more transistors to be packed onto a single chip, resulting in more functional and powerful integrated circuits. Businesses can lower the cost per transistor and increase total cost efficiency by streamlining the process. Miniaturization of high-density circuits that can execute complex tasks helps address the growing demand for mobile devices, AI, and high-performance computing. Our research reveals that 31% of semiconductor manufacturers are prioritizing semiconductor performance and miniaturization by embracing advanced lithography techniques.

Galvanize semiconductor supply chains through diversification and adoption of eco-friendly processes

Diversify supplier networks

Below are a few strategies for semiconductor organizations to diversify their relationships and invest in technologies to enhance and secure the supply chain:

Foster international partnerships: Global collaboration is an effective supply chain risk-mitigation strategy. Joint ventures with local firms can allow semiconductor organizations to access new markets and navigate unfamiliar regulatory systems more easily, as well as increase the customer base and revenue. Through partnerships, organizations can access a wider range of technologies and expertise and mitigate potential supply chain bottlenecks. For instance, TSMC, Robert Bosch, Infineon, and NXP Semiconductors announced their European Semiconductor Manufacturing Company (ESMC) joint venture in Germany to provide advanced semiconductor manufacturing services within the EU. According to Jochen Hanebeck, CEO of Infineon Technologies, *"Our joint investment is an important milestone to bolster the European semiconductor ecosystem. Dresden is strengthening its position as one of the world's most important semiconductor hubs and is already home to Infineon's largest front-end site."*[73] Collaboration also facilitates cost-sharing for ambitious projects and fosters resilience through diversification.

Develop alternative materials: The manufacturing of various high-performance components requires rare earth elements, the supply of which is often concentrated in a limited number of countries, leading to geopolitical and economic vulnerabilities. Our research suggests that the limited availability of critical materials such as silicon wafers, inert gases, and rare earth elements affects 67% of semiconductor supply chains. Investment in research to find viable alternative materials can mitigate such risks. Furthermore, diversifying material sources can enhance resilience against market fluctuations and trade restrictions.

31%

Percentage of semiconductor manufacturers prioritizing semiconductor performance and miniaturization by embracing advanced lithography techniques

Implement digital twin technology: Digital twins enable proactive simulation of various scenarios, allowing the anticipatory identification of vulnerabilities. By analyzing these simulations, organizations can develop mitigation strategies and optimize their operations, ensuring continuity of supply chain operations. Intel collaborated with supply chain software organization AIMMS to optimize its supply chain using digital twin. This has improved product flow, reduced delivery times, increased chip availability, and managed cost.[74] Digital twins offer a competitive edge by boosting agility and responsiveness in an industry where precision, quality, and timely delivery are critical.

Implement sustainable manufacturing practices

Lowering the environmental impact of the chip manufacturing process is of the utmost importance to the semiconductor industry. The following actions address urgent market demand and regulatory requirements:

Using sustainable materials and packaging: Traditional packaging materials are often non-biodegradable and environmentally harmful. TSMC has collaborated with suppliers to optimize the materials used in drum packaging, proactively replacing raw materials that contain substances of very high concern (SVHC), such as phosphite ester (TNPP), and refining drum-bottom design to reduce chemical residue by 60% while maintaining quality and safety standards. Following full implementation by 2025, TSMC expects to reduce chemical residue by 3.2 million liters annually and cut carbon emissions by around 570 tons, increasing environmental sustainability and green manufacturing capabilities.[75]

Implementing green chemistry practices: The semiconductor industry uses a variety of potentially harmful chemicals. By shifting to environmentally friendly alternatives, organizations can reduce their ecological footprints. Furthermore, adopting environmentally responsible practices helps to ensure compliance with increasingly stringent regulations, avoiding costly fines and legal issues. Our research reveals that 63% of organizations are either already deploying such practices across some regions/products or are actively scaling up the use of less toxic alternatives.

Minimize ecological footprint

Commitment to sustainability is not just a trend but an essential step in ensuring the industry's long-term viability. This motivates organizations to balance technological advancement with environmental responsibility by switching to renewable energy sources and designing energy-efficient chips.

Utilize renewable energy sources: Reducing reliance on fossil fuels and mitigating the impact of fluctuating energy prices can lead to long-term cost savings. Switching to renewable energy sources can also ensure a more stable and reliable energy supply, supporting the continuous operation of manufacturing processes. To enable this, semiconductor organizations are already collaborating with partners to source renewable energy. Micron, for instance, signed a virtual power purchase agreement for its US operations with Terra-Gen. Under the agreement, Micron will receive around 178 megawatts (MW) of wind electricity capacity and associated renewable energy certificates annually. This contract will support Micron's US target to reach 100% renewable energy by the end of 2025 and marks the latest milestone in Micron's journey to achieve net-zero GHG emissions by 2050.[76]

Design energy-efficient chips: According to AMD's CEO, Lisa Su, *"Driving performance gains over the next decade requires relentless focus on energy efficiency."*[77] Semiconductor organizations can reduce energy consumption by focusing on energy-efficient chips. This also extends the battery life of portable devices, enhancing user experience and convenience. Additionally, energy efficiency brings a competitive edge by meeting regulatory standards and consumer expectations for eco-friendly products.

Harness AI and Gen AI for advanced chip design and manufacturing

Automate design processes using AI and Gen AI

AI-led automation accelerates development cycles. ML algorithms and Gen AI can optimize various stages of the chip design workflow, from initial architecture to physical layout. For instance:

- Google's AlphaChip uses RL to optimize chip layouts in hours rather than the weeks or months it would otherwise require and has been used to design the last three generations of TPUs.[78]
- Design Space Optimization AI (DSO.ai) from Synopsys is a generative optimization paradigm that uses RL technology to autonomously search design spaces, optimizing for 'power, performance, and area' (PPA). Samsung's SoC designers used DSO.ai to fine-tune the design and maximize yields for its first 3-nm smartphone SoC.[79]
- Electronic design automation (EDA) firms employ ML algorithms to optimize chip design. Cadence introduced Cadence Cerebrus to improve productivity through a unique RL engine. In one instance, the ML model reduced overall optimization time from 18 to less than nine days.[80]

Use intelligent systems to improve production efficiency and yields

Through intelligent operations such as AI and ML, semiconductor manufacturers can enhance production efficiency, optimize complex fabrication processes, reduce the incidence of defects, and accelerate time-to-market for advanced chips. Examples include:

- TSMC deployed AI and ML to develop systems for precise fault detection and classification, and intelligent advanced equipment and process control, to ensure the consistency of tool matching and process stability. This allows the organization to meet 5G's stricter quality requirements for mobile, high-performance computing (HPC), automotive, and IoT.[81]
- ASML relies on AI to improve the performance of its lithography machines. The organization uses ML software to predict deviations in the lithography process.[82]
- Intel is relying on AI to analyze bug reports and categorize bugs to allow engineers to respond immediately to issues with new designs. It also uses AI to investigate minute variations in chips coming off production lines, including clock speed, cache memory size, and power consumption. The AI system scrutinizes this complex collection of details

"Softwarization is transforming the semiconductor industry, enabling unprecedented flexibility and scalability. By decoupling hardware from software, companies can rapidly adapt to evolving demands, optimize performance across diverse applications, and unlock new capabilities through updates. This shift not only accelerates innovation but also fosters collaboration across the ecosystem, as software becomes the bridge connecting hardware excellence with cutting-edge functionality."

Brett Bonthron
Industry Leader for Tech and Digital,
Capgemini

and finds the best fit for these processors in one of Intel's products. Intel has further begun testing Gen AI tools that process the detailed specs for a chip and create tests to ensure the chip design meets them.[83]

Develop AI-optimized semiconductor products

As downstream industries integrate AI into their products and services, the semiconductor industry must focus on products optimized for AI workloads. These specialized chips are designed to deliver enhanced performance, energy efficiency, and scalability. For instance:

- NVIDIA's H100 Tensor Core GPU offers significant advances in computational capabilities for training and inference of large-scale AI models, including Gen AI applications such as GPT-3 and GPT-4.[84] Major cloud service providers, including Google Cloud, Microsoft Azure, and AWS, have integrated H100 GPUs into their data centers.
- NVIDIA is also collaborating with AWS to offer a low-cost inference for Gen AI with Amazon SageMaker integrated into NVIDIA's NIM inference microservices. This allows customers to deploy pre-compiled and optimized FMs at speed, reducing time to market for Gen AI applications.[85]
- AMD launched its MI300X accelerator in 2023, a GPU designed for Gen AI workloads and HPC applications, featuring a large memory capacity to handle extensive language models and complex neural networks.[86]

Intel's AI accelerator Gaudi 2 offers a benchmarked alternative to NVIDIA's H100 by driving enhanced performance for deep learning training workloads.[87]

Align with government policies and foster open innovation

Coordinate strategies with regulatory frameworks

Government policy can have a significant impact on the operations of a semiconductor organization within its territory. Organizations should take the following steps to ensure they are aligned with local frameworks and ready to take advantage of the benefits they offer.

Stay informed and adaptive: Government policy has a strong influence on availability of raw materials for semiconductor manufacturing, including critical resources and minerals. These policies also impact trade agreements and tariffs, influencing cost and supply chain dynamics. Organizations should monitor government policies carefully, adjusting business strategies to align with regulatory changes, compliance, market access, and opportunities.

Collaborate with government initiatives: National programs such as tax credits or subsidies for R&D can offer significant benefits, helping semiconductor organizations to offset the extreme costs associated with chip manufacturing. Tax benefits against R&D allow organizations to deploy more resources on innovation, as well as ensuring a reliable and resilient supply chain. TSMC Arizona and the US Department of Commerce announced up to $6.6 billion in direct funding under the CHIPS and Science Act, fulfilling the goal of bringing advanced chip manufacturing to the US.[88]

Collaborate and share innovation

Strategic collaborations with organizations in similar fields, research institutions, academic bodies, etc., help semiconductor organizations to pool resources and expertise to address complex challenges and accelerate innovation. Such collaborations facilitate the creation of cutting-edge solutions, from advanced chip design to new manufacturing techniques, and lay the foundation for a more interconnected, resilient, and innovative semiconductor ecosystem. Areas where semiconductor organizations can participate are:

Industry consortia: Partnerships can address the growing complexity of technology development. Pooling knowledge and expertise and harnessing complementary strengths and advantages (e.g., production capacity) can accelerate innovation, promote industry standardization, help streamline manufacturing operations, and improve interoperability. Such relationships also strengthen collective resilience to supply chain disruption.

Ecosystem partnerships: Strategic alliances with suppliers, customers, and technology partners enhance innovation and extend market reach. Suppliers can help semiconductor organizations access cutting-edge raw materials or components. Customers can help organizations understand market demand and direct curation of products specific to requirements. Technology partners can boost the integration of cutting-edge technologies, such as AI, into their operations or product offerings.

Collectively addressing growth challenges: New technologies and other innovations bring with them new challenges in terms of talent and sustainability requirements. Working together, industry, government, academia, and civil society can meet these requirements in a way that benefits all parties. Industry can drive innovation and enrich the talent pool through rigorous training programs, while governments can boost investment and innovation with sympathetic policy formulation. Academia helps develop talent at the tertiary educational level, and civil society imposes checks and balances by pressuring for healthy sustainability practices.

Protect IP and enhance security measures

Strengthen cyber defenses and protect data integrity

The complexity of semiconductor supply chains attracts a critical threat landscape. Implementing robust security measures to every link of the semiconductor supply chain is important for enhancing cybersecurity and data integrity. The threat of more sophisticated cyberattacks means organizations should continue to innovate to safeguard the integrity of both their products and data. To bolster cybersecurity and maintain data integrity organizations should:

Ensure trusted supply chain components: Maintaining the integrity and reliability of products using chips requires a trusted semiconductor supply chain. TSMC upgraded the security of its factories by requiring new equipment to comply with standards set out in procurement contracts, strengthening the cybersecurity of its supply chain.[89] Sourcing parts from verified, secure suppliers helps prevent the infiltration of counterfeit parts, which can jeopardize product functionality and safety. Organizations should implement robust supplier-verification processes, including regular audits and certifications, such that they adhere to quality standards and provide traceability to meet industry specifications. Technologies such as blockchain can further secure the supply chain by providing transparent, immutable records of component origins.

Implement robust cybersecurity measures: Organizations can prevent unauthorized access and data breaches by using advanced security protocols, which ensure that sensitive information remains confidential while avoiding financial losses associated with IP theft.

Invest in hardware-based security: Semiconductor organizations should prioritize investment in hardware-based security to boost the overall resilience and trustworthiness of products. Incorporation of encryption and hardware security into chip design can provide robust protection against a wide range of cyber threats. Our research reveals that 58% of semiconductor organizations use cryptographic protection when designing hardware-based security for chips. Additionally, integration of security features at the hardware level can enhance performance and reduce latency compared with software-only solutions. As cyber threats evolve, establishing a strong hardware security foundation becomes essential to maintaining trust and safeguarding critical information.

Harness AI's threat-detection potential: Semiconductor organizations can proactively monitor and respond to security threats using AI and ML. AI-driven systems can analyze vast amounts of data in real time, identifying patterns and anomalies that could indicate potential threats in a dynamic defense mechanism that avoids costly security incidents.

Safeguard proprietary technologies

Implement advanced security measures: A comprehensive approach will be most effective in safeguarding IP. Organizations should deploy end-to-end encryption for both data at rest and in transit to secure designs, fabrication processes, and trade secrets. Multi-factor authentication and strict access controls can help limit internal threats. Similarly, AI-powered intrusion detection systems can help monitor networks in real time.

Regular security audits and vulnerability assessments can identify weaknesses in security operations. A few organizations also conduct "red-team" exercises to simulate cyberattacks, assisting the preparation for real offensives. Implementing secure boot processes and code signing can further protect firmware and software layers.

Employ technological solutions for IP management: Advanced technologies can further safeguard industry IP. Blockchain technology can create immutable and transparent records to store patent registrations, licensing agreements, and transaction details, helping them to resolve disputes. IPwe created a blockchain-based platform that helps organizations manage their patent assets.[90]

Organizations can deploy AI and ML to monitor and detect potential IP infringements, and quantum-resistant cryptography (QRC) can withstand the advanced hacking capabilities of quantum computers. KSE, a security IP platform, has integrated QRC, allowing semiconductor organizations to safeguard SoC products (and the devices that use them) from emerging quantum computing threats.[91]

Enhance legal protection for IP security: IP fuels innovation and secures competitive advantage. Enhancing legal measures for IP security is essential to combat the increasing sophistication of IP theft, cyber espionage, and the global nature of semiconductor supply chains. Strengthened legal frameworks protect proprietary technologies from infringement, unauthorized use, and counterfeiting, as well as protecting the financial and resource-based R&D investment and the return on that investment.

The semiconductor industry should advocate for stronger IP laws and enforcement mechanisms, both domestically and internationally. It should collaborate with government agencies to update existing legislation, ensuring it keeps pace with technological advancements in nanotechnology and quantum computing. Organizations could also push for stricter penalties for IP violations and faster patent approval.

Given the global nature of semiconductor supply chains and operations, international treaties and agreements would benefit all organizations. The ongoing legal battle between Arm and Qualcomm over Qualcomm's interpretation of the licensing agreement Arm had with Nuvia (which Qualcomm acquired)[92] highlights the complex nature of intra-industry litigation, a complication that standardized industry regulation could help avoid. Organizations should also clearly outline IP ownership, licensing rights, and confidentiality obligations in their contractual agreements with partners, suppliers, and customers.

Adopt open standards to foster innovation

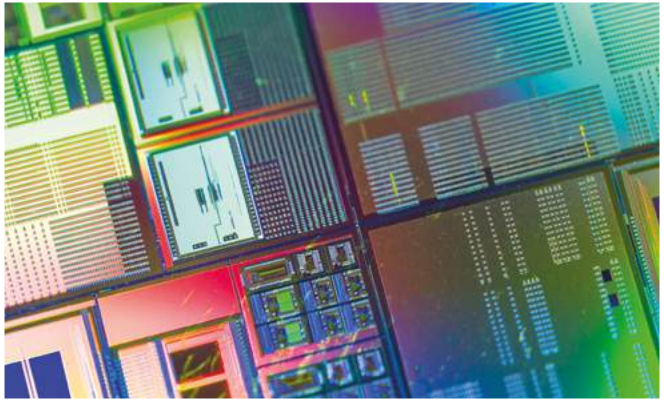
Open standards are essential for interoperability, innovation, and reducing proprietary lock-in as technology evolves rapidly. Despite challenges like market demands and integration complexities, adopting widely recognized standards can streamline semiconductor development, foster collaboration, and enhance performance across the supply chain.

Promote industry-wide collaboration for cross-platform interoperability

orming consortia with manufacturers, developers, and end-users is vital to ensure compliance with open standards. These groups can drive cross-platform interoperability by standardizing interfaces, protocols, and communication methods. Unified frameworks, like common instruction sets and physical layer standards, reduce fragmentation, cut costs, and enhance scalability across the industry. Notable examples include a consortium that contributes to cooperation for open architecture and design, such as the RISC-V[93] and the Open Compute Project (OCP).[94]

Encourage open-source design and development in semiconductor innovation

The semiconductor industry should embrace open-source models for chip design and development, enabling collaboration through shared design files, tools, and processes. This approach fosters innovation, accelerates product development, and ensures broader testing and validation across diverse use cases. Open-source initiatives such as OpenCores[95] and RISC-V showcase how shared chip designs can drive adaptability, affordability, and scalability while reducing monopolistic advantages. Transparent development processes enhance compatibility with open standards, simplify complex architectures, and invite contributions from companies, universities, and other stakeholders, broadening talent pools and accelerating innovation.



Conclusion

The semiconductor industry is at an inflection point. The anticipated surge in demand, fueled by the widespread adoption of AI and Gen AI technologies, means that semiconductor manufacturers will need to reconsider their supply chain strategies. Addressing the supply concerns of downstream industries is paramount. Semiconductor organizations must ramp up their capacity and agility to meet growing demand, confirming their role as the backbone of modern innovation.

Our research highlights that, while the industry excels in design, manufacturing innovation, and security, there are significant challenges related to softwarization of semiconductors and monetization of software-driven value additions. By embracing software-hardware integration and developing sustainable business models around it, organizations can unlock new revenue streams.

Supply chain resilience and sustainability have emerged as critical focus areas. Geopolitical tensions and supply shortages have highlighted the vulnerabilities of concentrated supply chains. By diversifying sourcing strategies through onshoring and friendshoring, and by bolstering cybersecurity measures, semiconductor organizations can mitigate risks and ensure uninterrupted operations. Eco-friendly manufacturing practices will assuage regulatory and societal demands for sustainability.

Strategic collaboration with governments and alignment with policy initiatives can provide additional support and R&D funding. Engaging in open innovation and industry partnerships can accelerate technological advancements and promote interoperability and standardization, benefiting the entire ecosystem.

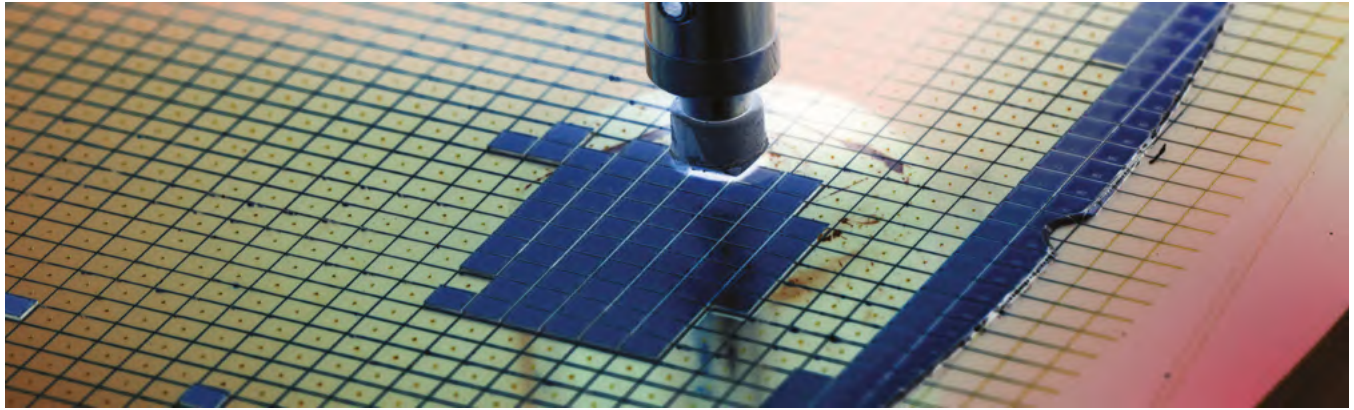
Protecting IP through fortified legal frameworks and advanced security measures is essential to safeguarding innovation and preserving competitive advantage. Furthermore, harnessing AI and Gen AI in design and manufacturing processes can significantly improve efficiency, reduce time-to-market, and optimize products for the next generation of technological applications.

In conclusion, the semiconductor industry's ability to adapt and innovate in response to these challenges will determine its success in the coming years. By investing in advanced technologies, strengthening supply chains, embracing sustainability, fostering collaboration, protecting intellectual assets, embracing open standards, and harnessing AI, the semiconductor industry can not only meet surging demand but also drive the next wave of global technological advancement. The actions we recommend in this report provide a strategic roadmap for industry leaders to navigate the complexities of the current landscape and to seize the opportunities that lie ahead, ensuring sustained growth and profitability in a rapidly evolving market.

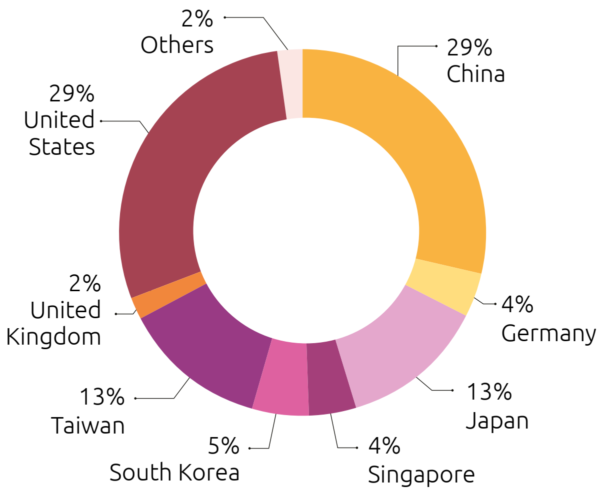
Research methodology

The research methodology is structured into three parts.

1. We surveyed 250 executives, at the director level or above, from the semiconductor industry (including integrated device manufacturers, fabless design, foundries, OSAT companies, EDA companies, capital equipment companies, and material and subsystem companies) across 11 countries in Asia–Pacific, Europe, and North America. These organizations each have annual revenues of $500 million and over. We carried out the global survey in November 2024. We provide the distribution of these respondents and their organizations below.

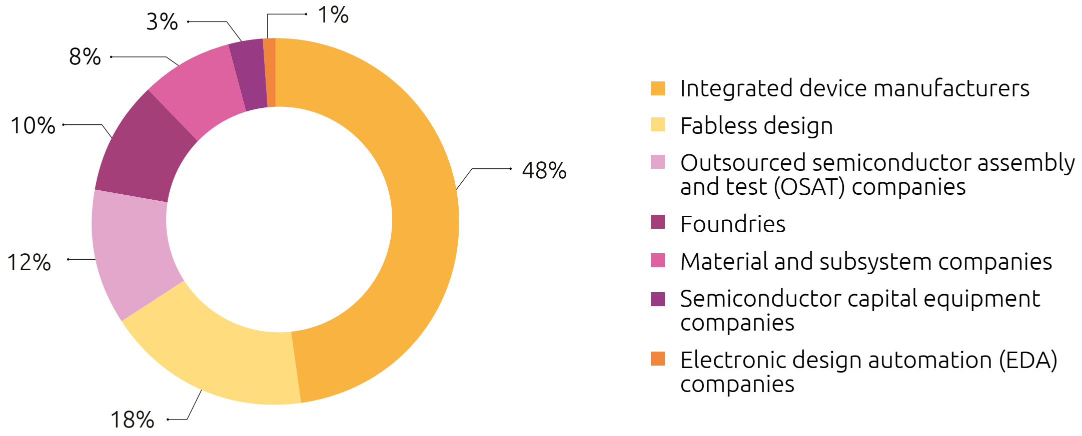


Organizations by headquarters



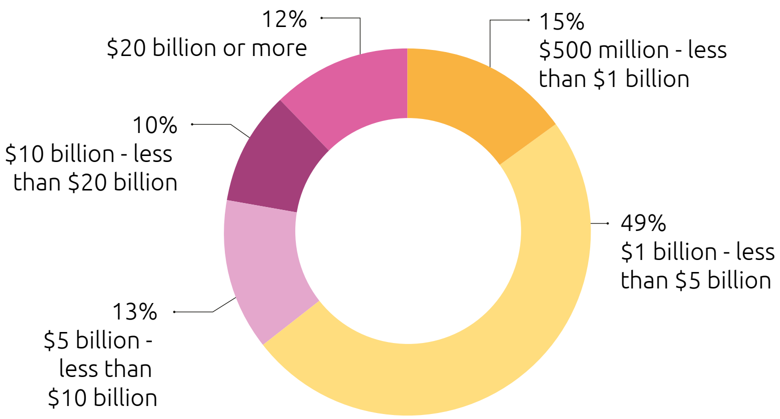
Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 250 semiconductor organizations. 'Others' includes France, Italy, and the Netherlands.

Organizations by type



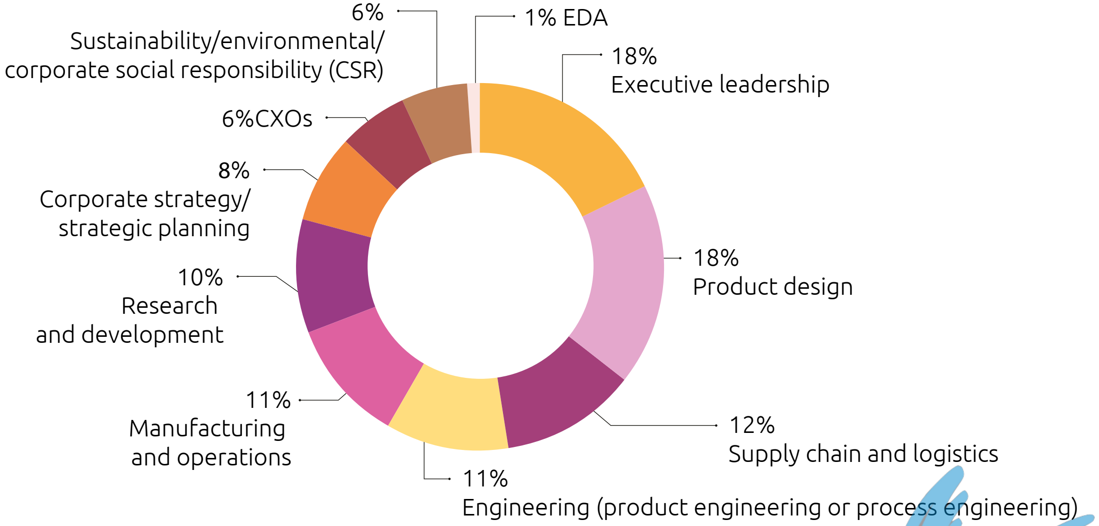
Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 250 semiconductor organizations.

Organizations by revenue



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 250 semiconductor organizations.

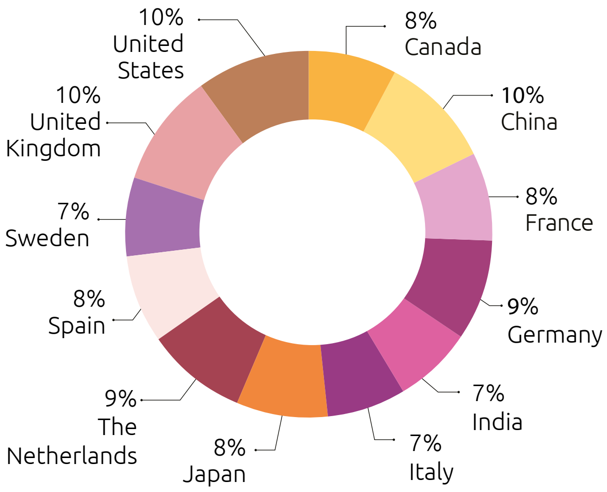
Executives by function



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 250 semiconductor organizations.

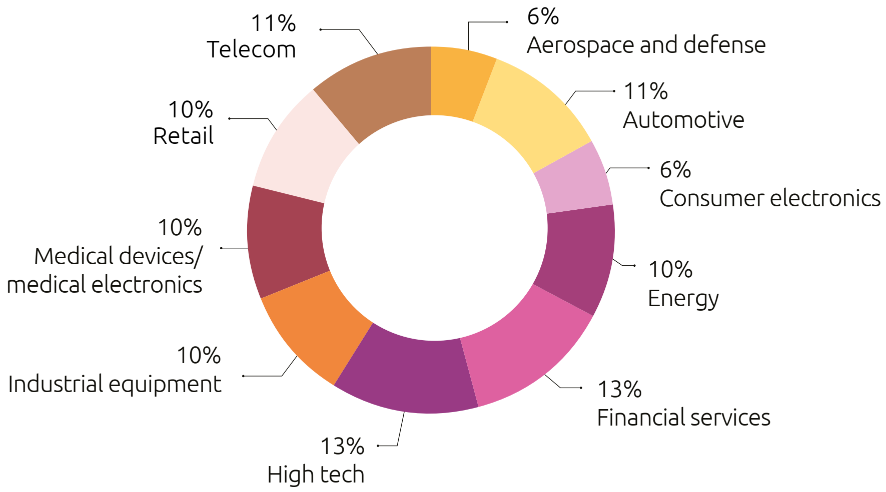
2. We further surveyed 800 executives, at the director level or above, from ten downstream industries across 12 countries in Asia–Pacific, Europe, and North America. These organizations each have annual revenues of $1 billion and over. We carried out the global survey in November 2024. We provide the distribution of these respondents and their organizations below.

Organizations by headquarters



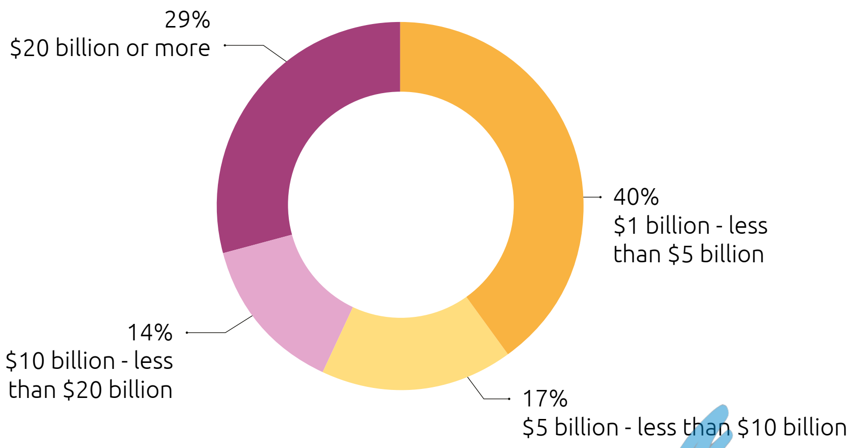
Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 800 downstream organizations.

Organizations by industry



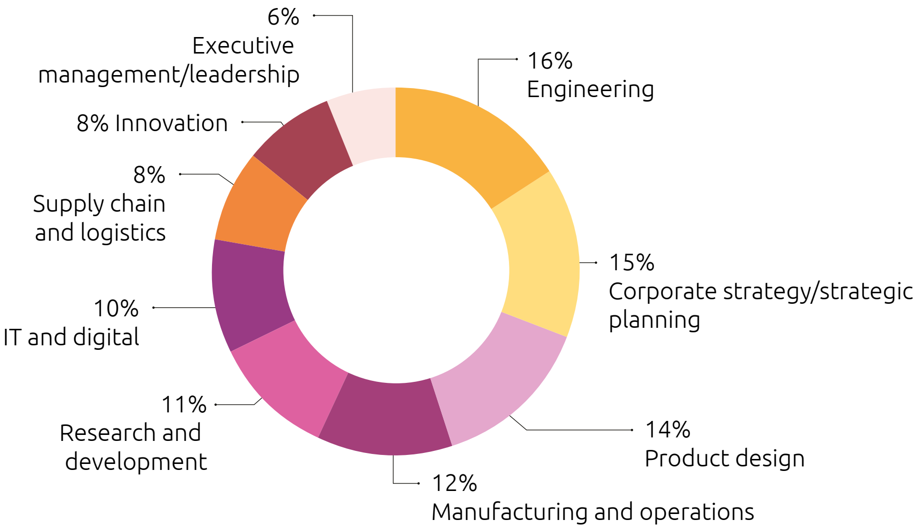
Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 800 downstream organizations. High tech sector includes software, internet, enterprise datacenter, and networking organizations.

Organizations by revenue



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 800 downstream organizations.

Executives by function



Source: Capgemini Research Institute, Semiconductor survey, November 2024, N = 800 downstream organizations.

3. To complement the survey findings, we conducted in-depth discussions with 12 executives from the semiconductor industry and downstream industries.

The study findings reflect the views of the respondents to our online questionnaire for this research and are intended to provide directional guidance. Please contact one of the Capgemini experts listed at the end of the report to discuss specific implications.

References

1. Support.microsoft.com, *All about neural processing units (NPUs) - Microsoft Support*. (n.d.). Retrieved November 13, 2024, from https://support.microsoft.com/en-us/windows/all-about-neural-processing-units-npus-e77a5637-7705-4915-96c8-0c6a975f9db4
2. Deierling, K., "What Is a DPU?" NVIDIA Blog, May 21, 2020. https://blogs.nvidia.com/blog/whats-a-dpu-data-processing-unit/
3. Amazon Web Services, What is a GPU? - *Graphics Processing Unit Explained* (n.d.). Retrieved November 14, 2024, from https://aws.amazon.com/what-is/gpu/
4. Bulletin.com, *"AI chips: What are they?"* | Built In. (n.d.). Builtin.com. Retrieved November 14, 2024, from https://builtin.com/articles/ai-chip
5. Arrow Electronics. (February 2024). *Custom silicon is here: benefits today and for the future*. Arrow.com; Arrow.com. https://www.arrow.com/en/research-and-events/articles/custom-silicon-chips-benefits
6. Capgemini, "Softwarization for Semiconductors," March 2024.
7. Semiconductor Industry Association (SIA), "State of the U.S. semiconductor industry," 2024.
8. Ibid.
9. Ibid.
10. Qualcomm, "Unlocking on-device generative AI with an NPU and heterogeneous computing," February 2024.
11. Nvidia blog, "Why GPUs are great for AI," December 4, 2024.
12. HBM3E. Micron Technology Inc. (n.d.). Retrieved November 4, 2024, from https://www.micron.com/products/memory/hbm/hbm3e
13. Intel blogs, "Intel and AWS expand strategic collaboration, helping advance U.S.-based chip manufacturing," September 16, 2024.
14. Taipei Representative Office in Singapore, "Taiwan and the global semiconductor supply chain," January 2024.
15. U.S. Embassy & Consulates in China, "Commerce Implements New Export Controls on Advanced Computing and Semiconductor Manufacturing Items to the PRC," October 11, 2022.
16. AP News, "China bans exports to US of gallium, germanium, antimony in response to chip sanctions," December 2024.
17. Dell blogs, "Introducing Dell AI for Telecom," 25 September 2024.
18. *Intel® eASIC™ Devices*. (2024a). Intel. Retrieved November 5, 2024, from https://www.intel.com/content/www/us/en/products/details/easic.html
19. Bloomberg, "Apple plans to make its own server chips to power AI tools this year," May 31, 2024.
20. Volkswagen blogs, "Volkswagen Group reorganizes semiconductor procurement," August 24, 2023.
21. TechSpot, "AMD patent unveils multi-chiplet GPU designs for unrivaled flexibility," June 15, 2024.
22. NVIDIA blogs, "Silicon Volley: Designers tap Generative AI for a chip assist," October 30, 2023.
23. Yahoo Finance, "Intel spends more on R&D than Nvidia and AMD combined, et continues to lag in market cap — Nvidia spends almost 2X more than AMD," October 2024.

24. *MIT Technology Review,* “How ASML took over the chipmaking chessboard,” April 1, 2024.

25. Reuters, “Intel hails 'landmark' as high-volume EUV production begins at Irish plant,” September 29, 2023.

26. *Semiconductor Materials - IEEE IRDS™*. (n.d.). Irds.ieee.org. Retrieved November 6, 2024, from https://irds.ieee.org/topics/semiconductor-materials

27. TSMC, Agile and Intelligent Operations -. (n.d.). Www.tsmc.com. Retrieved November 6, 2024, from https://www.tsmc.com/english/dedicatedFoundry/manufacturing/intelligent_operations

28. TSMC, *CoWoS®* -. (n.d.). 3dfabric.tsmc.com. Retrieved November 6, 2024, from https://3dfabric.tsmc.com/english/dedicatedFoundry/technology/cowos.htm

29. Forbes, “TSMC to double production based on Nvidia numbers and overall demand,” November 2024.

30. *CPO Magazine,* “Major semiconductor firm applied materials hit by supply chain attack; Ransomware impact will cost $250 million,” February 2023.

31. Ars Technica, “Unpatchable vulnerability in Apple chip leaks secret encryption keys,” March 2024.

32. Investor’s Business Daily, “CUDA, Woulda, shoulda: How this platform helps Nvidia dominate AI,” September 5, 2024.

33. Qualcomm, “Qualcomm and AWS commit to long-term co-innovation to advance the future of software-defined mobility,” September 5, 2023.

34. *MIT Technology Review*, “How ASML took over the chipmaking chessboard,” April 1, 2024.

35. CSIS, “Russia's Invasion of Ukraine impacts gas markets critical to chip production,” March 14, 2022.

36. *The Economist*, “Taiwan’s dominance of the chip industry makes it more important,” March 2023.

37. The News & Observer, " Largest Spruce Pine quartz mine restarts after Helene battered mountain town,” October 10, 2024.

38. The White House, “FACT SHEET: CHIPS and Science Act will lower costs, create jobs, strengthen supply chains, and counter China,” August 9, 2022.

39. CSIS, “A world of Chips Acts: The future of U.S.-EU semiconductor collaboration,” August 20, 2024.

40. U.S. Department of Commerce, “Biden-Harris administration announces preliminary terms with TSMC, expanded investment from company to bring world’s most advanced leading-edge Technology to the U.S.,” April 8, 2024.

41. SIA, “2024 State of the Industry report underscores opportunities and challenges for U.S. chip industry,” September 12, 2024.

42. CNBC, “Europe approves its $47 billion answer to Biden’s CHIPS Act — here’s everything that’s in it,” April 2023.

43. Intel, “Intel commits to net-zero greenhouse gas emissions in its global operations by 2040,” April 2022.

44. *Circular economy - Minimizing waste, maximizing resources*. (n.d.). Www.asml.com. Retrieved November 7, 2024, from https://www.asml.com/en/company/sustainability/circular-economy

45. AMD, “Thinking small to think big: AMD modular architecture is advancing sustainability," April 22, 2024.

46. Toms Hardware, “Google unveils AlphaChip AI-assisted chip design technology – chip layout as a game for a computer,” September 28, 2024.

47. Hot Hardware, "Google Cloud unveils Trillium, its 6th-Gen TPU with a huge 4.7X AI performance leap," May 15, 2024.
48. Google Store, "Google Tensor: the brains behind Pixel phones," accessed on December 5, 2024.
49. Google, "How we built Pixel Buds Pro 2 for even better sound and comfort," November 22, 2024.
50. Data Center Dynamics, "Alphabet,' quarterly capex hits $13bn, Google Cloud revenues jump to $10.32bn," July 24, 2024.
51. Toms Hardware, "Amazon solidifies push into custom chips that replace Intel and AMD processors, unveils Graviton4 and Trainium2 cloud servers," November 29, 2023.
52. Toms Hardware, "Amazon solidifies push into custom chips that replace Intel and AMD processors, unveils Graviton4 and Trainium2 cloud servers," November 29, 2023.
53. AWS, "AWS Inferentia," accessed on December 5, 2024.
54. Market Screener, "Intel and AWS to expand strategic collaboration on chip manufacturing," September 16, 2024.
55. CNBC, "Amazon CEO pledges AI investments will pay off as capital expenditures surge 81%," October 31, 2024.
56. Data Center Dynamics, "AWS revenue hits $26.3bn for Q2 2024, Amazon capex reaches $30.5bn," August 2, 2024.
57. TechRepublic, "Apple Silicon: Why developers don't need to worry about the transition from Intel-based Macs," November 11, 2020.
58. Apple, "Apple unveils M3, M3 Pro, and M3 Max, the most advanced chips for a personal computer," October 30, 2023.
59. Mac Daily News, "TSMC is making Apple chips in Phoenix, Arizona," October 16, 2024.
60. Mac Rumors, "Apple chip manufacturing in U.S. to expand following new partnership," October 4, 2024.
61. Data Centre Magazine, "How Meta will power data centers with its in-house AI chips," February 6, 2024.
62. Data Center Magazine, "Top 10 chip providers for the data centre industry," November 8, 2023.
63. The Register, "Untangling Meta's plan for its homegrown AI chips, set to actually roll out this year," February 2, 2024.
64. IEEE Spectrum, "3D chip tech is key to Meta's AR goals," February 21, 2024.
65. TechRadar, "Meta's new AI card is one step to reduce its reliance on Nvidia's GPUs," April 15, 2024.
66. Microsoft Azure blogs, "Azure Maia for the era of AI: From silicon to software to systems," April 3, 2024.
67. Ibid.
68. The Verge, "Microsoft is finally making custom chips — and they're all about AI," November 15, 2023.
69. Intel, "Intel launches world's first systems foundry designed for the AI era," February 2024.
70. Intel blogs, "High NA EUV at Intel," April 18, 2024.
71. Nikkei Asia blogs, "TSMC to receive ASML's next-gen chip machines this year," November 1, 2024.
72. Intel blogs, "Intel builds world's largest neuromorphic system to enable more sustainable AI," April 17, 2024.
73. TSMC blog, "TSMC, Bosch, Infineon, and NXP establish joint venture to bring advanced semiconductor manufacturing to Europe," August 8, 2024.
74. *LinkedIn*. (2024). Linkedin.com. Retrieved November 11, 2024, from https://www.linkedin.com/posts/aimms_aimms-helps-intel-to-optimize-its-supply-activity-7064246188920725504-t2vr/
75. TSMC blogs, "TSMC collaborates with suppliers to innovate chemical material packaging to enhance green benefits," July 5, 2024.
76. Micron blogs, "Micron signs 178-megawatt wind power agreement with Terra-Gen," June 8, 2023.
77. Applied Materials blogs, "Materials engineering: The true hero of energy-efficient chip performance," August 1, 2024.
78. Google DeepMind, "How AlphaChip transformed computer chip design," September 26, 2024.
79. AnandTech, "Samsung tapes out its first 3nm smartphone SoC, gets a boost from Synopsys AI-enabled tools," May 3, 2024.

80. Cadence blogs, "What's behind the 5% die-area shrink and 12% power saving by MediaTek?" August 22, 2022.

81. TSMC, Annual report 2022, March 2023.

82. *How to measure the quality of a microchip.* (n.d.). Www.asml.com. Retrieved November 12, 2024, from https://www.asml.com/en/technology/lithography-principles/measuring-accuracy

83. CNET, "AI helps chipmakers design the very processors that speed up AI," December 13, 2023.

84. *NVIDIA H100 Tensor Core GPU.* (n.d.). NVIDIA. Retrieved November 14, 2024, from https://www.nvidia.com/en-us/data-center/h100/

85. CRN, "Nvidia's 10 New Cloud AI Products For AWS, Microsoft And Google," March 2024.

86. *AMD Instinct™ MI300X Accelerators.* (November 2024). AMD. https://www.amd.com/en/products/accelerators/instinct/mi300/mi300x.html

87. *Intel® Gaudi® 2 AI Accelerator.* (2023). Intel. https://www.intel.com/content/www/us/en/products/details/processors/ai-accelerators/gaudi2.html

88. TSMC blogs, "TSMC Arizona and U.S. Department of Commerce announce up to US$6.6 Billion in proposed CHIPS Act direct funding, the company plans third leading-edge fab in Phoenix," April 8, 2024.

89. TSMC blog, "TSMC Strengthens "Specification for Cybersecurity of Semiconductor Equipment", Realizing Mutual Industry Benefits," September 14, 2023.

90. Spangenberg, E. (February 2021). *Disrupting the patent ecosystem with blockchain and AI.* Ibm.com. https://www.ibm.com/think/insights/disrupting-the-patent-ecosystem-with-blockchain-and-ai

91. *Kudelski IoT Launches Quantum-Resistant Security IP, Future-Proofing Semiconductors Against Emerging Quantum Threats.* (2024). Kudelski-Iot.com. https://www.kudelski-iot.com/news/kudelski-iot-launches-quantum-resistant-security-ip-future-proofing-semiconductors-against-emerging-quantum-threats

92. Forbes, "Why protecting IP in the semiconductor market is vitally important," January 17, 2023.

93. About RISC-V international. RISC. (n.d.). Retrieved December 9, 2024, from https://riscv.org/about/

94. Open Compute Project. (n.d.). Open Compute Project. Retrieved December 9, 2024, from https://www.opencompute.org/

95. OpenCores. (n.d.). Retrieved December 9, 2024, from https://opencores.org/

Authors

Meet the experts

Brett Bonthron
EVP, Global industry leader for High-tech, Capgemini
brett.bonthron@capgemini.com

Brett has over 35 years of experience in high-tech, across technical systems design, management consulting, start-ups, and leadership roles in software. He has managed many waves of technology disruption from client-server computing to re-engineering, and web 1.0 and 2.0 through to SaaS and the cloud. He is currently focusing on defining sectors such as software, computer hardware, hyper-scalers/platforms, and semiconductors. He has been an Adjunct Faculty member at the University of San Francisco for 18 years teaching Entrepreneurship at Master's level and is an avid basketball coach.

Sanjiv Agarwal
Vice President, Group industry leader for Semiconductors, Capgemini
sanjiv.agarwal@capgemini.com

Sanjiv is the leader of Capgemini's Global Semiconductor Industry practice, bringing over 30 years of experience in delivering business transformation programs across diverse industries, including technology, manufacturing, and banking. He collaborates with semiconductor value chain companies to drive end-to-end transformation strategies. Sanjiv has extensive background in helping organizations optimize operations, embrace technological innovation, and drive sustainable business growth.

Authors

Meet the experts

Nicolas Rousseau
EVP, MD for Digital Engineering and Manufacturing,
Capgemini Engineering
nicolas.a.rousseau@capgemini.com

Nicolas drives business for "intelligent industries" by integrating product, software, data, and services. He leads a team that helps clients innovate business models, optimize operations, and prepare for digital disruptions, enhancing customer interaction, R&D, engineering, manufacturing, and supply chains. Nicolas spent 15 years with Altran (now Capgemini) in various leadership roles. Nicolas also worked as a Visiting Researcher at Stanford University after graduating from Ecole Nationale Supérieure des Mines de Nancy, where he studied earth sciences and engineering.

Khalid Sebti
EVP, MD for Americas,
Capgemini Engineering
khalid.sebti@capgemini.com

Khalid is responsible for driving business across various verticals including Technology, Semiconductor, Media, Telecommunications, Manufacturing, Aerospace, Automotive, Life Sciences, Energy, Consumer Product Goods, Services, and Public Sector. He is a proven business leader with extensive experience in products and engineering domains. Previously, Khalid was Vice President of Software Engineering & Cybersecurity at Altran, overseeing product engineering, system verification & validation, and cybersecurity. He has led major engineering transformation projects for various clients. Khalid holds an MSc. in Computer Science from the French Aeronautics & Space Engineering School and an MBA from INSEAD, France.

Laurence Noël
Executive Vice President, Head of Global Automotive Industry, Capgemini
laurence.noel@capgemini.com

Laurence Noël is a distinguished professional with a robust background in the automotive industry. She has more than 25 years of experience and leadership and was previously Senior Vice President of the Circular Economy Business Unit at Stellantis. Leveraging a diverse background across management of business units, strategic planning and engineering, Laurence empowers clients to shape profitable and successful strategies, achieve transformation, and meet automotive sustainability goals. Laurence also holds several influential Board positions in the Automotive industry that are relevant to Capgemini's portfolio of clients.

Authors

Meet the Capgemini Research Institute

Jerome Buvat
Head,
Capgemini Research Institute
jerome.buvat@capgemini.com

Jerome is the head of the Capgemini Research Institute. He works closely with industry leaders and academics to help organizations understand the business impact of emerging technologies.

Subrahmanyam KVJ
Senior Director,
Capgemini Research Institute
subrahmanyam.kvj@capgemini.com

Subrahmanyam is a director at the Capgemini Research Institute. He loves exploring the impact of technology on business and consumer behavior across industries in a world being eaten by software.

Ramya Krishna Puttur
Associate Director,
Capgemini Research Institute
ramya.puttur@capgemini.com

Ramya has over twelve years of experience in consulting and digital transformation and is an Indian Institute of Technology alumnus with gold medals at both post-graduation and under-graduation levels. She has co-authored numerous publications for the Capgemini Research Institute on the impact of digital technologies across industries and developed proprietary tools for assessing clients' data and digital maturity.

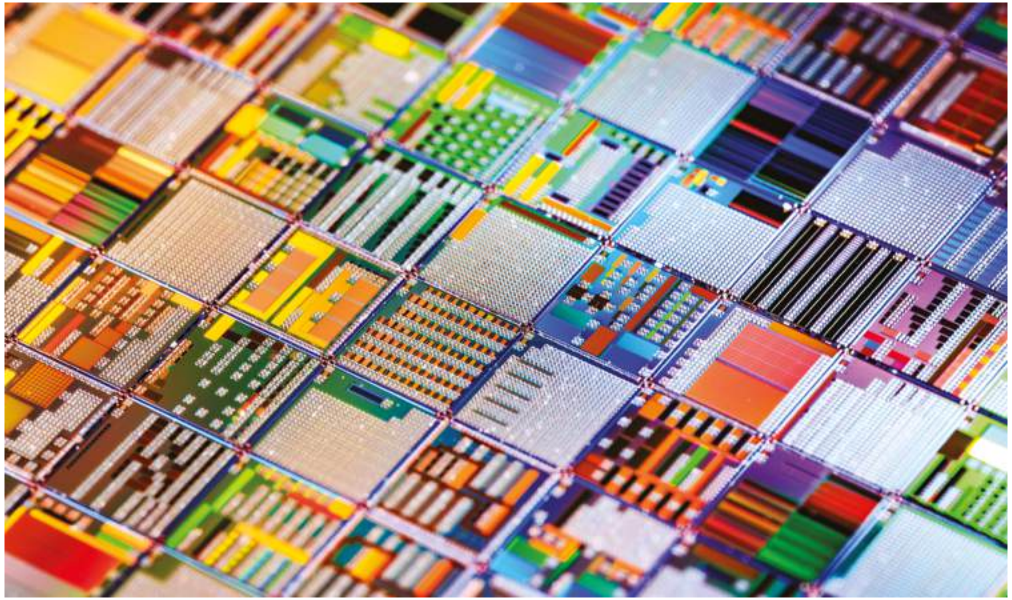
Sroddha Patra
Senior Consultant,
Capgemini Research Institute
sroddha.patra@capgemini.com

Sroddha is a research consultant at the Capgemini Research Institute, where she works alongside industry leaders to understand the impact of technologies on businesses across sectors. Her expertise lies in strategic research, advisory, and developing insights that help organizations stay ahead in the evolving technology landscapes.

The authors would like to thank the following people for their contributions to the research:

Pascal Brier	Group Chief Innovation Officer, Capgemini
William ROZE	CEO, Capgemini Engineering
Mourad Aberbour	VP, CTO, Silicon Engineering, Capgemini Engineering
Loic Hamon	CMO, Silicon Engineering, Capgemini Engineering
Jonathan Nussbaumer	VP, Silicon Engineering, Capgemini Engineering
Shiv Tasker	VP, ER&D, TMT, Capgemini Engineering
Henrik Ljungström	EVP, Head of Business Unit, Germany, Capgemini
Jiani Zhang	EVP-Head of Portfolio and Product, Capgemini Engineering
Sandeep Kamath	VP, TMT Business Unit, Capgemini
Prem Kumar	Senior Director, Engineering and RD Services, Capgemini Engineering
Elaine Outler	Senior Director - Portfolio and Product Director, Engineering and RD Services, Capgemini Engineering
Ravi Gupta	Senior Director - Industry and Innovation, Capgemini
Karl Bjurstroem	EVP, Global head of Tech and Telecom, Invent, Capgemini
Thorsten Rall	EVP, Global lead for life sciences, Capgemini
Hema Gandhi	Senior Director - Industry and Innovation, Capgemini
Sebastian Scholz	Senior Manager, Capgemini Engineering

Richa Bubna	Account Manager, Capgemini
Claudio Sinisi	Senior Manager, Capgemini Engineering
Ananya Das	Director - Industry and Innovation, Capgemini
Imttiaz Ali	Senior Director, Chief of Staff, Engineering and RD Services
Adrien Girard	Chief of Staff, Capgemini Engineering
Vincent Godiner	Chief of Staff, Capgemini
Nico Wolf	Director, Capgemini Engineering
Subhash K Nair	Director, Engineering and RD Services, Capgemini Engineering
Martin Przewloka	Head of Strategy and Technology Application Services for SAP, CTO, Capgemini
Mark Murrietta	Senior Manager, Engineering and RD Services, Capgemini Engineering
Stephanie Berthe	VP, Global Head of M&C, Capgemini Engineering
Ralf Bus	Global Director of Technology Marketing, Capgemini
Frederic ARQUIER	Chief Marketing Officer, Capgemini Engineering
Claudia Hytrek	Marketing Manager, Capgemini
Amrita	Senior Manager, Capgemini Research Institute
Nikhil Sai Kongonda	Manager, Capgemini



The authors would like to thank all industry experts and executives who participated in this research. They would also like to thank Laura Immonen Beatty, Punam Chavan, Manish Saha, Vibha Palekar, Amitabha Dutta and Mithun Dutta for their contributions to the research.

About the Capgemini Research Institute

The Capgemini Research Institute is Capgemini's in-house think tank on all things digital. The Institute publishes research on the impact of digital technologies on large traditional businesses. The team draws on the worldwide network of Capgemini experts and works closely with academic and technology partners. We are proud to have been ranked #1 in the world for the quality of our research by independent analysts for six consecutive times - an industry first.

Visit us at www.capgemini.com/researchinstitute/

Transforming the semiconductor industry with innovation and expertise

At Capgemini, we are at the *forefront of the semiconductor revolution*, driving **performance** and **excellence** across the industry. Our comprehensive suite of services is designed to meet the unique challenges and opportunities of the semiconductor sector, ensuring your business stays ahead in a rapidly evolving market. We are the global leader in semiconductor and silicon design services, offering end-to-end solutions from chip to industry.

We help develop products and services for tomorrow, from concept through industrialization.

Why industry leaders trust us:

- Strategic transformation partner for innovation leaders across industries
- Unmatched domain expertise with decades of vertical industry experience – **85%** of the **top 200** global companies are Capgemini clients
- Trusted by 9 of the top 10 semiconductor giants
- 21+ years of experience in semiconductor engineering and R&D
- Scalable and global delivery, with a worldwide network of 6000+ engineers & world-class engineering centers in 13 countries
- Expertise in cloud, data & AI, and cybersecurity
- A new type of semiconductor service company, offering clients the full expertise of a fabless semiconductor company, from specification to silicon and volume production
- Part of the *Arm Total Design* program

Partner with Capgemini to transform your semiconductor business with our unparalleled expertise and innovative solutions.

For more information: *https://www.capgemini.com/industries/high-tech/semiconductor/*

For more information:

Silicon Engineering

Softwarization for Semiconductors

Sustainability for Semiconductors

Follow us on LinkedIn: *Capgemini High Tech*

For more information, please contact:

Global

Brett Bonthron
EVP, Global industry leader for High-tech, Capgemini
brett.bonthron@capgemini.com

Sanjiv Agarwal
VP, Group industry leader for Semiconductors, Capgemini
sanjiv.agarwal@capgemini.com

South Central Europe

Guido Kamann
EVP, MD for Switzerland, Capgemini
guido.kamann@capgemini.com

Henrik Ljungström
EVP, Head of Business Unit, Germany, Capgemini
henrik.ljungstroem@capgemini.com

North Central Europe

Jeannine Peek
MD, Netherlands Business Unit, Capgemini
jeannine.peek@capgemini.com

United States

Darshan Naik
EVP, Chief Growth Officer for TMT industry, Capgemini
darshan.naik@capgemini.com

APAC

Jun Cheng
Vice President, China
jun.cheng@capgemini.com

Masamitsu Murata
Senior Director, Japan
masamitsu.murata@capgemini.com

Vivekanandan Vadivelu
Senior Director, India
vivekanandan.vadivelu@capgemini.com

Ravi Gupta
Senior Director - Industry and Innovation, India
ravi.b.gupta@capgemini.com

More Capgemini Research Institute publications

New defenses, new threats: What AI and Gen AI bring to cybersecurity

A world in balance 2024: Accelerating sustainability amidst geopolitical challenges

Data-powered enterprises 2024

Reindustrialization of Europe and the US

Conversations for Tomorrow #9: Generative AI

Autonomous networks

The Eco-Digital Era™: The dual transition to a sustainable and digital economy

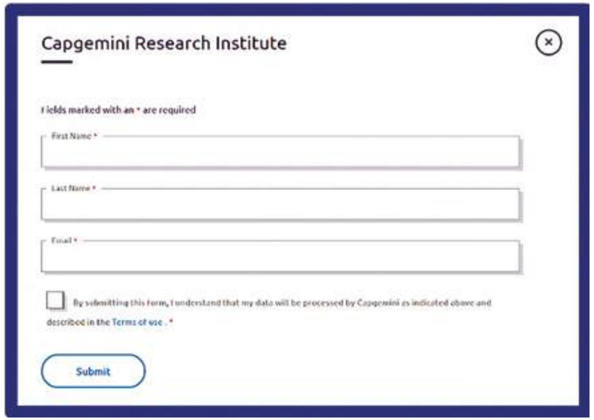
Digital twins: Adding intelligence to the real world

Generative AI in organizations 2024

Quantum technologies

Subscribe to latest research from the Capgemini Research Institute

Receive copies of our reports by scanning the QR code or visiting https://www.capgemini.com/capgemini-research-institute-subscription/


Capgemini Research Institute

Fields marked with an * are required

First Name *

Last Name *

Email *

By submitting this form, I understand that my data will be processed by Capgemini as indicated above and described in the Terms of use . *

Submit

Notes

About Capgemini

Capgemini is a global business and technology transformation partner, helping organizations to accelerate their dual transition to a digital and sustainable world, while creating tangible impact for enterprises and society. It is a responsible and diverse group of 340,000 team members in more than 50 countries. With its strong over 55-year heritage, Capgemini is trusted by its clients to unlock the value of technology to address the entire breadth of their business needs. It delivers end-to-end services and solutions leveraging strengths from strategy and design to engineering, all fueled by its market leading capabilities in AI, cloud and data, combined with its deep industry expertise and partner ecosystem. The Group reported 2023 global revenues of €22.5 billion.

Get the Future You Want | **www.capgemini.com**