WHITEPAPER

JANUARY 2025

IMPACT OF AI ON IP AND CHIP DESIGN

Global Semiconductor Alliance

Intellectual Property Interest Group

Contents

1. Executive Summary

This whitepaper aims to explore the impact of AI on intellectual property (IP) and Chip Design, providing comprehensive details and multifaceted data to cover all aspects of the semiconductor industry. It highlights AI growth trends, market predictions, and current silicon chip design innovations. The paper also examines key industries and their supply chains while emphasizing the critical importance of security in supporting massive AI workloads. This evolution demands a closer look at legal considerations and the need for specific skill sets that have significantly impacted labor and immigration on a global scale, which are highlighted here. The paper recommends areas where cross-industry collaboration is needed to address major AI challenges and recommends standards that make AI in semiconductors a sustainable, scalable, and secure solution.

2. AI Growth Trend and Market CAGR prediction

The global artificial intelligence market size was estimated at USD 196.63 billion in 2023 and is projected to grow at a CAGR of 36.6% from 2024 to 2030. The continuous research and innovation directed by tech giants drive the adoption of advanced technologies in industry verticals, such as automotive, healthcare, retail, finance, and manufacturing. For instance, in December 2023, Google LLC launched 'Gemini' - a large language AI model made available in three sizes: Gemini Nano, Gemini Pro, and Gemini Ultra. Gemini stands out from its competitors due to its native multimodal characteristics.

AI has proven to be a significant revolutionary element of the upcoming digital era. Tech giants like Amazon.com, Inc.; Google LLC; Apple Inc.; Facebook; International Business Machines Corporation; and Microsoft are investing significantly in AI research and development (R&D), thus increasing the artificial intelligence market cap. These companies are working to make AI more accessible for enterprise use cases. Moreover, various companies adopt AI technology to provide a better customer experience and improve their presence in the artificial intelligence industry 4.0.

The essential fact accelerating the rate of innovation in AI is accessibility to historical datasets. Since data storage and recovery have become more economical, healthcare institutions and government agencies build unstructured data accessible to the research domain. Researchers are getting access to rich datasets, from historic rain trends to clinical imaging. The next-generation computing architectures, with access to rich datasets, are encouraging information scientists and researchers to innovate faster.

Furthermore, progress in profound learning and Artificial Neural Networks (ANN) has also fueled the adoption of AI in several industries, such as aerospace, healthcare, manufacturing, and automotive. ANN works in recognizing similar patterns and helps in providing modified solutions. Tech companies like Google Maps have been adopting ANN to improve their route and work on feedback received using ANN. ANN is substituting conventional machine learning systems to evolve precise and accurate versions.

For instance, recent advancements in computer vision technology, such as Generative Adversarial Networks (GAN) and Single Shot MultiBox Detector (SSD), have led to digital image processing techniques. For instance, images and videos taken in low light or low resolution can be transformed into HD quality by employing these techniques. Continuous research in computer vision has built the foundation for digital image processing in security & surveillance, healthcare, and transportation, among other sectors. Such emerging methods in machine learning are anticipated to alter the manner AI versions are trained and deployed.

The COVID-19 outbreak stimulated market growth of next-generation tech domains, including artificial intelligence, owing to the mandated work-from-home (WFH) policy due to the pandemic. For instance, LogMeIn, Inc., a U.S.-based company that provides Software-as-a-Service (SaaS) and cloud-based customer engagement and remote connectivity & collaboration services, has experienced a significant increase in new sign-ups across its product portfolios amid the pandemic.

A recent outlook by Bloomberg estimates that AI could add $6 billion to the electronic design-automation market for chip design.

3. How AI demands innovation in Semiconductor from concept to production

This topic will include Any Cost and Time Efficiency data/projection with AI enabling. Here are the two design-flow charts to use as framework:

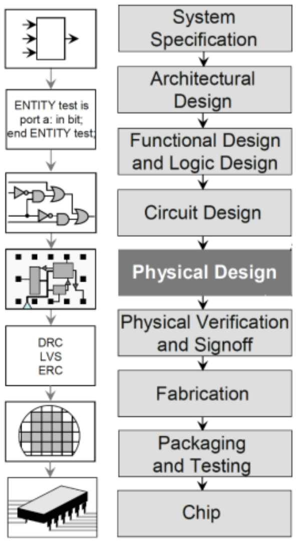

Figure 1: A typical System Design Flow, Source: Andrew B. Kahng, et al., "VLSI Physical Design: From Graph Partitioning to Timing Closure," Springer (2011)

Achieving the next level of productivity demands innovation from system-level design all the way to design for manufacturing (see Figure 1), and literally, no part of the flow remains untouched by AI/ML.
Even before any line of code is written, Generative AI techniques can be applied to analyze written specifications and deductions of requirements for later checking. EDA vendors have reported on early experiments to automatically generate RTL code from textual specifications.

Starting at the system-design level, simulation time is a key issue that can be addressed using AI/ML techniques to set up system-design scenarios and, during architectural analysis sweeps, apply the correct heuristics as early as possible. This can be combined with high-level synthesis, generating different hardware configurations based on parameters like performance, power consumption, and area, allowing designers to identify optimal solutions quickly. Extending to 2.5D and 3D multi-die designs, AI-driven optimization maximizes system performance and quality of results, expediting design partitioning, floor planning, and implementation processes to optimize for thermal, signal, and power integrity.

For Printed-Circuit Board (PCB) system design, AI/ML based approaches allow faster design closure and improve routability of PCBs. The DARPA IDEA program in 2018 spawned early results that, at this point, have been productized and included in commercial products.

Once the architecture of a chip is more defined, functional verification – using both dynamic simulation and static formal techniques – is an NP-Complex problem that is literally never done. AI/ML techniques can significantly reduce the number of required simulation cycles while achieving the same coverage by smartly arranging regression runs. In the world of formal verification, using AI/ML to optimize the application of SAT solvers, for instance, can lead to significantly faster out-of-the-box proofs.

Later in the flow, during digital implementation, the application of AI/ML can lead to dual-digit better PPA (Performance Power Area), and up to 10x productivity has been reported as an abundance of computing cycles allows to simply run many more implementation options, searching for optimization targets in vast solution spaces of chip design, utilizing reinforcement learning to enhance power, performance, and area, and applying re-enforcement learning allows to identify and optimize the trends in digital implementation settings. In addition, AI/ML-optimized approaches allow better layout timing predictions.

For testing, AI/ML techniques help users achieve fewer patterns, higher coverage, and reduced turnaround time, minimizing test cost and time-to-market for complex designs.
For Custom IC implementation, AI/ML provides accurate response surface models of analog devices or blocks and applications in layout group prediction optimized custom IC design even further.
In design for manufacturing and library characterization of semiconductor technology to feed higher level flows, AI/ML-based approaches significantly accelerate library development using interpolation of libraries, allow hotspot prediction, and enable in-design detection and fixing of those hotspots, improving yield.

Application of AI/ML to EDA metrics and tool-flow data is an emerging field to enable comprehensive data visibility, machine-intelligence-guided design optimization, and optimization of signoff-closure. Extending to fab manufacturing, AI/ML-based data analytics allows quick and accurate decision-making, improving overall fab efficiency using process control that maximizes yield and throughput.

3.1 Hardware Obsolescence for AI

Artificial Intelligence (AI) is generating increasing interest both in industry and research fields. This enthusiasm has led to a growing number of projects and research efforts, driving rapid advancements in AI algorithms. To fully leverage these increasingly powerful algorithms, it is crucial to have hardware that is sufficiently powerful and well-suited. Studies have shown that executing an algorithm efficiently requires designing specialized and optimized chips for specific tasks. However, the time required to manufacture these electronic chips is relatively long compared to the speed at which AI algorithms evolve. As a result, by the time specialized hardware is produced, it is often already outdated, as the algorithms it was designed for have evolved, rendering the chip obsolete. A significant aspect of this challenge is the deployment of edge AI, which involves executing AI algorithms locally on edge devices rather than relying on centralized cloud servers. This hardware obsolescence, which has already been observed in fields such as telephony, is even more pronounced in AI due to its breakneck pace of advancement. In this chapter, we will explore various hardware solutions for executing AI algorithms and analyze how they are impacted by obsolescence. We will also examine emerging technologies, such as embedded FPGAs, that aim to mitigate this challenge.

3.1.1 Hardware Deployment of AI at the Edge

Deploying edge AI algorithms, especially during the inference phase, is increasingly common in applications such as autonomous vehicles, smart cities, and IoT devices. Given the vast amount of data generated at the edge, moving data from sensors to the cloud for processing and expecting real-time responses becomes challenging. One solution is to bring AI to edge devices for local data processing. This approach reduces latency by performing processing locally, avoids transferring all data to a centralized server, improves privacy by processing sensitive data locally, and enhances reliability by reducing dependency on potentially unreliable internet connectivity. Additionally, it cuts costs because edge devices are generally inexpensive, cloud server resources are reduced, and there's no need for extensive infrastructure for data transfer. Different hardware solutions for edge AI can be categorized into three types: general-purpose, specialized AI, and programmable systems.

3.1.1.1 General-Purpose Processors

General-purpose processors, including CPUs and GPUs, are widely used for edge AI due to their versatility and processing power. CPUs are often inadequate for time-sensitive AI applications due to extended run times unless used with large computational infrastructures. GPUs provide significant speedups for neural network inference and can handle numerous parallel computations, making them ideal for fast, efficient processing of large data volumes. However, GPUs are challenging to use in edge AI scenarios because of their important power consumption and prohibitive costs. Consequently, general-purpose processors may quickly become obsolete in edge AI applications where hardware must adapt rapidly to evolving algorithms and constraints.

3.1.1.2 Application-Specific Integrated Circuits (ASICs)

Specialized hardware chips, such as Application-Specific Integrated Circuits (ASICs), are designed with specific architectures to accelerate AI applications. Some ASICs, like Google Tensor Processing Units (TPUs), are used in data centers for training deep learning algorithms. In contrast, others, such as Intel Movidius Myriad and Google Coral, are tailored for edge AI. ASICs excel in providing high performance and efficiency for their intended applications and can be more cost-effective in high-volume production compared to general-purpose processors. However, ASICs face limitations in flexibility and scalability. They are designed for specific functions and cannot be reconfigured for different tasks. ASICs can quickly become outdated as AI algorithms evolve, requiring frequent updates or replacements, which can be resource intensive.

3.1.1.3 Reconfigurable FPGA Hardware

FPGAs (Field-Programmable Gate Arrays) offer high flexibility and can be reconfigured to implement custom logic tailored to specific AI algorithms. This adaptability makes them well-suited for edge AI applications, where power and resource constraints demand efficient and optimized hardware. By programming an FPGA to implement a specific algorithm, designers can achieve significant performance improvements and power efficiency gains compared to running the same algorithm on a traditional CPU. However, while FPGAs offer flexibility, they are not immune to obsolescence. The need for frequent reconfiguration to keep up with rapidly advancing AI algorithms can lead to challenges in maintaining optimal performance and efficiency over time.

3.1.2 Emerging Hardware Solutions from FPGA Lens

Addressing the challenge of hardware obsolescence in AI requires a multifaceted approach, incorporating strategies to ensure that hardware can keep pace with evolving AI demands. One promising solution is the adoption of more flexible and reconfigurable hardware platforms that can be updated or modified to accommodate new AI algorithms. This approach can extend the useful life of hardware, reduce the frequency of replacements, and lower overall costs.

Among emerging technologies, embedded Field-Programmable Gate Arrays (eFPGAs) stand out as a particularly promising option. Unlike traditional FPGAs, eFPGAs are integrated directly into System-on-Chip (SoC) designs, providing a compact, energy-efficient solution that can be reprogrammed as needed. This integration allows eFPGAs to offer high performance while addressing space and power constraints typical of edge AI applications. Their reconfigurability means they can adapt to evolving AI algorithms without needing external hardware updates. Although eFPGAs may not be the sole solution to hardware obsolescence, they represent a significant advancement in developing adaptive, future-proof hardware for AI applications, offering a balance between flexibility, efficiency, and adaptability.

4. Impact of AI on Key Industry segments

4.1 Aspects of Developing AI Silicon- Hardware Feature Focused

With the disruptive shift in the semiconductor industry towards AI, hardware strategy needs focused shifting. The escalating performance requirements of AI, specifically deep learning, have caused a renaissance in hardware architecture for neural networks, with many varied business models. Google has invested considerable resources in building out a singularly focused computing stack for deep learning - Tensorflow-XLA-TPU (although predictably, they are now adding PyTorch), while Nvidia is enhancing its product lines to dominate AI silicon sales at high margins. There are dozens of startups, as well as established companies (Intel, Qualcomm, Huawei), building AI hardware with many different architectural approaches. Amazon, Baidu, Apple, Microsoft, Facebook, and others are all known to be working on various non-commodity accelerated AI. There is a paradigm shift from expensive GPUs to higher-performance custom AI processors. This selection of Graphics core is critical due to its market position, bandwidth, programmability, scalability, and performance to meet heavy AI demand across every market segment from Military, Healthcare, Banking & Finance, Retail & E-Commerce, Legal, Gaming, IT and Tech Sector and so forth. A robust Hardware and AI centric Datacenter is needed to support all this demand. Heavy Investments and focused effort happening to be ready to meet and excel the market need.

Key Hardware Components of AI Silicon:

- **Processing Units**: These include CPUs, GPUs, NPUs and specialized AI accelerators like TPUs (Tensor Processing Units) designed to handle AI workloads efficiently.
- **Memory**: High-speed memory (e.g., HBM, DDR) is crucial for storing and quickly accessing large datasets in AI computations.
- **Interconnects**: These are high-bandwidth connections that facilitate rapid data transfer between different components of the AI silicon.
- **Power Management**: Efficient power management systems are essential to ensure the silicon operates within thermal limits while maintaining high performance.
- **Cooling Solutions**: Advanced cooling mechanisms are necessary to dissipate the heat generated by intensive AI computations.

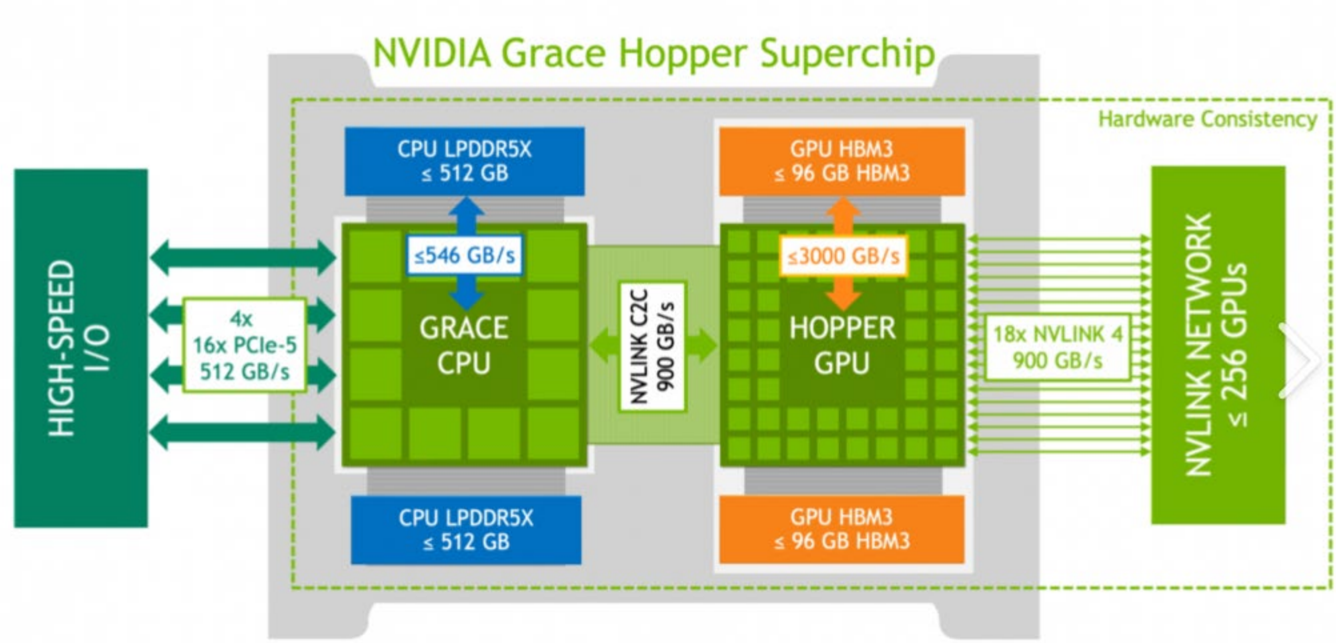

Figure 2: Using AI for developing silicon

The algorithms and frameworks continue to evolve rapidly, increasing the cost of staying competitive if the agility and programmability of the engine aren't sufficient (e.g., Nvidia is rumored to have over 1,000 people dedicated to maintaining their AI software moat). Google is leveraging its perceived lead in AI, particularly using heavy marketing of its TPU chips, to increase its cloud attach rate. Distributed training of neural networks using data and model parallel techniques on hundreds of AI accelerators in large scale-out systems is becoming an increasingly popular approach to cut down training time drastically. Computational throughput requirements for emerging AI scenarios like AGI will increase by orders of magnitude (Up to an Exaflop and more) in the next few years and will require data center scale systems to fulfill those demands. Thus, leading capabilities in AI have implications broader than just AI platform sales. Silicon must provide leading performance in inference and training, with low latency and low $/transaction.

Developing AI-based silicon involves several critical aspects that ensure the hardware can efficiently support the demanding computational needs of AI applications. Here are some key considerations:

- **Custom Design:** Companies like Google, Meta, and Amazon are increasingly developing custom AI chips tailored to their specific needs. This allows for optimized performance and efficiency, as these chips are designed to handle the unique workloads of the AI model.
- **Performance and Efficiency:** AI chips need to manage large data volumes and quickly perform complex calculations. This requires high-performance computing capabilities and efficient power consumption to minimize heat generation and energy use.
- **Integration with Software:** Effective AI silicon development involves close collaboration between hardware and software teams. This integration ensures that the chips can fully leverage the capabilities of AI algorithms and models, leading to better performance and more relevant user experiences

- **Scalability:** As AI applications grow, the demand for computational power increases. AI chips must be scalable to handle larger datasets and more complex models without significant performance degradation.
- **Sustainability**: The environmental impact of AI computing is a growing concern. Developing energy-efficient AI chips is crucial to reduce the carbon footprint associated with training and deploying AI models.
- **Innovation in Materials:** Leveraging the material properties of silicon and exploring new materials can lead to more sophisticated and capable AI systems. This involves continuous research and development to push the boundaries of what AI chips can achieve.

These aspects highlight the multi-faceted approach required to develop AI-based silicon that meets the evolving needs of AI applications.

5. Consequences of the Rollout of AI-based Solutions on the Chipset Supply Chain

There's little doubt that AI and technologies such as generative AI (Gen AI) and machine learning (ML) are enabling supply chain teams to streamline operations, improve decision-making, and enhance productivity.

- **Enhanced supplier relationship management (SRM):** AI can analyze supplier performance data to identify patterns and trends – thus enabling companies to select suppliers based on factors including quality, cost, reliability, and sustainability credentials. The result is often stronger, more trustworthy supplier relationships and, ultimately, a more resilient supply chain.
- **Better demand forecasting:** Sticking with the theme of data analysis, AI can be leveraged to examine sales records, market trends, and external factors – such as weather patterns or economic indicators – to forecast demand more accurately. AI enables companies involved in the supply chain of semiconductors to better anticipate customer needs and adjust inventory levels, reduce overstock and stockouts.
- **Improved inventory management:** Tied in with forecasting and planning improvements, organizations can optimize inventory levels by predicting exactly when and where their products are needed with AI-driven analytics. One savings will come from reducing the carrying costs of excess inventory.
- **End-to-end visibility:** The integration of AI provides companies with end-to-end visibility within their supply chains by integrating data from multiple sources. As a result, firms can track goods in real time, manage inventory across multiple locations, and, crucially, respond quickly to any disruptions.
- **Automation of routine tasks:** AI enables the automation of routine tasks along the whole supply chain. Activities such as order processing, invoicing, and inventory tracking are prime targets for automation. Automation eliminates human error and speeds up processes. It frees employees to focus on more strategic tasks that add value to operations.
- **Sustainability and environmental impact:** AI helps the logistics element of operations by optimizing routes and thus reducing fuel consumption and emissions output. It helps companies track and manage their carbon footprint and adhere to environmental regulations, contributing to an all-round more sustainable supply chain.
- **Predictive maintenance:** AI can predict when equipment or other machinery will likely fail, allowing companies to perform maintenance before a breakdown occurs.

6. Criticality of Security with AI in Semiconductors

The semiconductor industry faces the challenge of ensuring the security of hardware products. With the increasing complexity of hardware systems and the rising threats of cyber-attacks, traditional methods of testing and securing semiconductor products are becoming less effective. Artificial Intelligence has become a key tool among cybersecurity solutions.

- **AI-Powered Testing: A New Frontier:** AI introduces a paradigm shift to implement hardware security in the semiconductor industry. With their ability to learn and adapt, machine learning algorithms provide a dynamic approach to testing and securing hardware. Unlike static testing methods, AI-powered systems continuously evolve, learning from new data and patterns, making them more effective in identifying and responding to new threats.
- **Deep Learning for Deep Inspection:** Deep learning, a subset of AI, enables a more thorough inspection of hardware at a granular level. Deep learning algorithms can identify potential security breaches and vulnerabilities by analyzing patterns and anomalies with unprecedented precision.

- **Predictive Analytics for Proactive Security:** AI enables predictive analysis to detect future vulnerabilities. Predictive analytics checks for trends and patterns to forecast potential security issues before they manifest. This proactive approach to hardware security can save the semiconductor industry from costly and damaging breaches.
- **Automating the Tedious: Efficiency and Accuracy:** The automation capabilities of AI can handle repetitive and complex tasks with greater efficiency and accuracy than human testers. This not only speeds up the testing process but also reduces the likelihood of human error, ensuring more reliable security measures.

7. Impact on Skillset, Immigration, and Labor Requirements

AI has the potential to revolutionize semiconductor manufacturing. However, to fully realize this, the industry must also address the intense competition for skilled labor driven by a global talent shortage in this field. Key areas of focus can be summarized as follows:

- **Understanding Cross-Border Labor Mobility:** A recent review of immigration processes in key jurisdictions focused on the semiconductor sector revealed the insights below. Questions focused on speed and efficiency of processes, predictability of outcome, and also on the types of moves- be it internships, long-term relocation, or skill exchanges/ training.

Overall ranking for:	Country's overall immigration processes		How easy is it for a semiconductor company to bring in workers they need
Estonia	4.5	Chile	4.4
Mexico	4.5	Costa Rica	4
Israel	4.25	Mexico	3.8
Ireland	4.25	Japan	3.8
Malaysia	4	Portugal	3.8
China	3.75	China	3.6
Czech Republic	3.75	Malta	3.6
Germany	3.5	Estonia	3.4
UK	3.5	Malaysia	3.4
Colombia	3.5	Ireland	3.4
Taiwan	3	Colombia	3.4
Costa Rica	3	Taiwan	3.2
Malta	3	Israel	3.2
Japan	3	Germany	3.2
Portugal	2.75	Czech Republic	3
Italy	2.75	Greece	3
Chile	2.5	UK	2.8
Greece	2.5	Italy	2.2
Vietnam	1.75	Vietnam	2.2

Figure 3:Cross Broader Labor Mobility

The United States is absent from the above list. The newly elected U.S. administration's AI policies vis-à-vis immigration are still coming into focus. President-elect Trump has stated he will revoke President Biden's Executive Order on AI, which he viewed as too restrictive on businesses' development of AI. While President-elect Trump has expressed an intent to protect U.S. workers, he has shown a clear desire to keep the United States in the AI forefront. Undoubtedly, effective engagement with government policymakers is essential to shaping immigration programs that support agility in talent mobility, a key driver of successful outcomes. The substantial financial commitments from major players such as the United States, European Union, China, and South Korea underscore the strategic importance governments place on securing a strong foothold in the semiconductor industry. To maximize the impact of these investments, it is imperative that labor market access and skills shortages remain central to these discussions and negotiations.

With significant proportions of new hires across the sector now requiring the hiring of foreign talent, it is also increasingly important to see potential locations as opportunities. For example, while a consistent talent pipeline from India has been notable for many decades, a notable shift towards the opportunities for businesses to establish a presence there is growing as India continues to increase its foothold in the industry.

- **Generating a consistent and impactful graduate talent pipeline:** Ongoing collaboration with universities and technical institutions is critical to developing and sustaining a robust talent pipeline. Whether through engineers serving as part-time professors to facilitate effective knowledge transfer, industry-sponsored PhD programs aimed at retaining key talent, or strategic regional partnerships directly with academic institutions, consistent and deliberate investment in these initiatives is indispensable. Such efforts not only strengthen the sector but also ensure its long-term resilience.
- **Upskilling existing technician-level workers and broader workforce, with a significant focus on women and underrepresented STEM groups:** At the heart of successfully navigating the challenges posed by labor shortages are the critical teams in Learning & Development (L&D) and Talent Acquisition (TA). L&D plays a vital role in cross-training and upskilling existing employees, helping to retain valuable institutional knowledge, reducing turnover, and ensuring that the workforce remains relevant and effective. Likewise, fostering internal mobility, supported by TA, is essential to developing a flexible and resilient workforce. By creating strong networks for underrepresented groups, providing clear career pathways, and offering continuous learning opportunities, organizations can address current attrition rates and build a more sustainable and diverse talent pool for the future.

8. Areas where Industry Collaboration is needed to resolve challenges

AI's rapid growth presents significant challenges across various aspects of the chip industry. To address these challenges, the industry could benefit from collaborative efforts in several key areas. Below are some of the examples:

- **Joint Research and Development**: Chip manufacturers, AI companies, and research institutions could collaborate on R&D to develop new architectures and materials that improve chip performance, energy efficiency, and integration for AI workloads.
- **Open Standards and Interoperability**: Developing open standards for AI-specific chips and components would promote interoperability across devices and reduce fragmentation.
- **Supply Chain Transparency and Resilience**: Chipmakers, foundries, and suppliers can work together to improve transparency and share data to anticipate and mitigate disruptions. This collaboration could involve blockchain for traceability. In addition, multi-regional production, with manufacturing and packaging spread across different countries, can reduce reliance on single points of failure.
- **Green AI and Sustainable Practices**: Companies could collaborate on designing AI chips optimized for energy efficiency, balancing performance with sustainability. Collaboration on enabling Edge AI capability on renewable energy facilities (e.g., solar photovoltaic stations) is also a hot topic that distributes AI computation tasks to energy resources so that the energy transmission loss could be deducted as much as possible.

9. Legal Considerations

The AI boom has created enormous opportunities for the semiconductor industry, both in terms of supplying the industry with advanced, AI-capable chips and in terms of improving chip design processes. These opportunities raise several legal issues that companies must bear in mind to ensure they capture value and avoid costly legal disputes.

Three key issues arise when thinking about legal challenges. Most obvious is the issue of using AI in chip design. In particular, the provenance of training data and the question of ownership of the outputs of AI systems are hot topics in the field of large-language model (LLM) development.

Secondly, given the rapidly developing field of AI-capable hardware, with many participants in different markets, the question of who owns what IPRs will evolve quickly. Patent disputes are anticipated to increase significantly in the coming years.

Finally, a key point in terms of IPR development will be the issue of standards, especially open standards and open hardware.

- **Use of AI in chip design**

The semiconductor industry is well ahead of most industries when it comes to using AI to improve efficiency and reduce development costs. EDA tools have been a central driving force behind Moore's Law for several decades. However, the move towards more advanced AI technologies, such as reinforcement learning, genetic algorithms, adversarial networks, and, most recently, transformers, raises new questions about the legal implications of using such technologies. Many of these techniques may avoid using training data or at least use proprietary training data. However, if training data is being sourced externally, companies need to be careful about the provenance and ownership of such. As widely reported, OpenAI is being sued by multiple parties over its use of third-party copyrighted material. Semiconductor EDA and design companies must be careful that the data they use to train AI systems does not use third-party semiconductor designs that may be subject to copyright or other intellectual property rights (IPRs).

In addition to the above, companies must be careful about what third-party AI tools their employees use as part of their design work. In 2022, Samsung had to ban the use of some LLMs after they found that employees were entering confidential information into ChatGPT.

Companies also need to consider the impact of using AI on issues such as the ownership of IPRs. In most jurisdictions, IPR ownership begins with the employee and passes to the employer by agreement or by virtue of national laws. If the "inventor" is an AI process, who owns the IPRs? Several courts have looked at this issue, with the general consensus being that only humans can be inventors. Inventions made by AI processes are essentially owned by the employees running the AI systems. However, this is a rapidly evolving area that companies must keep a close eye on.

- **Race for the leading technologies and IPRs**

There is currently a race to develop AI-capable hardware for the next generation of AI applications. One consideration is how specific the chip should be. For example, companies have in the past designed processors with convolutional neural network architectures in mind for computer vision tasks, whereas vision tasks are increasingly being performed using vision transformers.

Many state-of-the-art AI chips now focus on transformers, the architecture behind large language models (LLMs). Groq's language processing unit (LPU) is one of the most specialized and performant. Nvidia's GPUs and Google's TPUs are more general and rely on additional low-level software (e.g., using CUDA in the case of Nvidia) to perform appropriate scheduling for a given architecture. Nevertheless, Nvidia's most performing chips (e.g., H100) are billed as transformer engines. Big bets are being made on transformers, but the field is rapidly developing, and a new architecture could replace transformers, at least in some domains. Furthermore, other architectures like diffusion models (used by most image generators) may be better run on more general GPUs.

The speed of development and the number of different companies involved in chip design have arguably never been higher. This raises many questions about which technology or technologies will dominate. Central to this will be who owns key IPRs. Given the strategic importance of these issues across borders, we can expect patent litigation in this area to be prevalent in the coming years. The industry has seen this in the past, for example, with the memory patent disputes of the early 2000s.

---

[1] https://www.bbc.co.uk/news/technology-67826601
[2] https://techcrunch.com/2023/05/02/samsung-bans-use-of-generative-ai-tools-like-chatgpt-after-april-internal-data-leak/
[3] https://en.wikipedia.org/wiki/DABUS
[4] https://www.eetimes.com/analysis-patent-suit-loss-batters-rambus-business-model/

In view of these issues, it will be important for companies to ensure they are properly protecting themselves by filing patents and maintaining appropriate trade secrets around their core technologies.

- **Development of open standards and open-source hardware**

The chip industry has seen an increased move towards open standards and open hardware design in order to improve interoperability and provide access to smaller players in the market. RISC-V is a key example. AI-specific hardware is likely to see a similar drive. While companies can build proprietary systems on top of open standards and open source, the IPR risks and the approach to IPR management require careful consideration. For example, open-source software is largely based on copyright, and the rights that underpin the licenses pass effectively from source to executable code. In contrast, patents tend to be a more appropriate form of IPR protection for hardware. Furthermore, the software industry generally sees less patent litigation, with the open-source community being seen as typically a bad target for corporate IPR disputes. In the hardware game, many more patents are in play, making the chance of patent litigation higher. Companies should not expect a free ride from the big hardware technology companies just because they are building system based on open standards and designs.

10. Conclusions

AI's impact on IP & Chip Design is multifaceted, with significant Growth Trends and scaling Market Predictions; the need for innovations in Silicon development is a must across all aspects, from architecture, design, process technology, and material to fleet and supply chain. AI/ML techniques are being used to address key issues in system-level design, functional verification, digital implementation, and custom IC implementation. These innovations lead to significant improvements in performance, power, and area (PPA) and productivity. The rapid evolution of AI algorithms often renders specialized hardware obsolete by the time it is produced. Emerging technologies like embedded FPGAs aim to mitigate this challenge by offering reconfigurable and adaptable hardware solutions. The shift towards AI has caused a renaissance in hardware architecture for neural networks, with many varied business models. Companies are investing heavily in building AI hardware to meet the escalating performance requirements of AI across various market segments. Every Cloud Service Provider and major industrial sector is investing and exploring innovative ways to drive toward the adoption of advanced technologies in multiple verticals. The chip industry has seen an increased move towards open standards and open hardware design to improve interoperability and provide access to smaller players in the market. RISC-V is a key example. AI-specific hardware is likely to trend towards similar drives.

AI enables better demand forecasting, improved inventory management, and enhanced supplier relationship management, leading to a more resilient supply chain. AI helps optimize routes, reduce fuel consumption, and manage carbon footprints, contributing to a more sustainable supply chain.

We clearly see the need for a closer look at legal considerations due to the evolution of AI and its impact on the semiconductor industry. This includes addressing intellectual property (IP) rights, data privacy, and compliance with international regulations. The rapid advancements in AI technology necessitate updated legal frameworks to protect innovations and ensure the ethical use of AI.

The need for specific skillsets to support the growing demands of AI in the semiconductor industry. This includes expertise in AI/ML techniques, system-level design, functional verification, digital implementation, and custom IC design. Professionals with these skills are essential for driving innovation and maintaining competitive advantage in the industry. AI is playing a vital role in ensuring cybersecurity.

The significant impact of AI on labor and immigration on a global scale. The demand for specialized skillsets has led to increased labor mobility and the need for immigration policies that support the movement of skilled professionals. This includes facilitating work visas and ensuring that immigration policies align with the industry's needs to attract and retain top talent.

11. Contributors

Tulika Jha / Editor

Andrew Thompson

Marc Canel

Nassim Abderrahmane

Deirdre Murray

Frank Schirrmeister

Gary Xu