

David Tsoi, CFA, CAIA, FRM, CESGA, CAMS, *Lead Index Research Strategist*

Global Semiconductor Industry Year-over-year Sales Growth

The chart displays the year-over-year sales growth of the global semiconductor industry from January 2018 to January 2025. The y-axis represents the growth percentage, ranging from -30% to 40% in 10% increments. The x-axis shows time in quarters, with labels for Jan 18, Jul 18, Jan 19, Jul 19, Jan 20, Jul 20, Jan 21, Jul 21, Jan 22, Jul 22, Jan 23, Jul 23, Jan 24, Jul 24, and Jan 25. The growth was generally positive from 2018 to 2022, peaking at approximately 30% in late 2021. A sharp decline occurred in early 2023, reaching a low of about -22%, followed by a steady recovery to around 20% by early 2025.

Quarter	Year-over-year Sales Growth (%)
Jan 18	22
Apr 18	20
Jul 18	20
Oct 18	21
Jan 19	20
Apr 19	17
Jul 19	15
Oct 19	13
Jan 20	12
Apr 20	9
Jul 20	1
Oct 20	-5
Jan 21	-10
Apr 21	-12
Jul 21	-15
Oct 21	-14
Jan 22	-12
Apr 22	-10
Jul 22	-8
Oct 22	-18
Jan 23	-20
Apr 23	-22
Jul 23	-21
Oct 23	-18
Jan 24	-12
Apr 24	-5
Jul 24	5
Oct 24	12
Jan 25	15
Apr 25	18
Jul 25	15
Oct 25	18
Jan 26	20
Apr 26	22
Jul 26	20
Oct 26	18
Jan 27	17
Apr 27	18
Jul 27	22
Oct 27	20
Jan 28	20

1

fully design and target campaigns using AI tools by the end of next year. Based on the client's budget, these new tools would generate the entire advertisement, including images, videos and text, and deliver it to the targeted audience.⁶

In the past, AI demand has primarily focused on training workloads, particularly for frontier models. While leading tech companies continue pouring resources into building ever-larger AI models, they are also reallocating more investment toward inference. Inference is the stage where trained AI models process new data to generate insights, make predictions or support decision-making. While training a model is essentially a one-time expense, prompting a model (inference) produces tokens, each of which carries a cost. During the Google I/O 2025 keynote, Alphabet CEO Sundar Pichai shared that the firm processed 480 trillion tokens across its products and APIs in April 2025, 50 times more than the same month a year earlier.⁷ The rapid surge in token volume reflects growing usage and adoption of AI models, signaling a greater need for computing power and driving higher demand for chips.

The age of AI reasoning

The shift in investment toward inference has also gained momentum with the launch of new reasoning models. While traditional AI models respond swiftly and excel at pattern recognition, they often fail to understand broader contexts and struggle with complex reasoning. Reasoning models are built to deconstruct complex problems into smaller, manageable steps and solve them through explicit logical reasoning. They are specifically trained to show their work and follow a more structured thought process, which results in longer computation times for user queries. These models demand significantly more compute during inference to reason through intricate problems. This evolution from basic pattern recognition to structured reasoning is pivotal to AI, unlocking its potential to tackle complex real-world challenges effectively. As AI adoption rapidly expands, demand for inference will correspondingly intensify.

The ascending wave of AI agents

AI agents are positioned to revolutionize how organizations function, delivering breakthroughs in productivity and operational efficiency. They are intelligent systems designed to execute tasks independently by comprehending objectives, formulating decisions and taking actions to achieve predetermined goals. While humans define the desired outcomes, AI agents autonomously select optimal actions required to accomplish those goals. These agents boast a broad spectrum of uses, from supporting academic research and streamlining online purchases to planning leisurely vacations. Customer service, sales and marketing, and IT and cybersecurity are the three business functions where AI agents are most frequently deployed or planned for implementation in the next six months.⁸ As enterprises progressively integrate AI agents across diverse operational applications, demand for computational infrastructure is escalating dramatically.

The rise of custom AI chips

Hyperscalers are increasingly focused on ASIC (application-specific integrated circuits) infrastructure to meet surging AI demand. ASICs are custom-built for specific workloads and can execute those tasks far more efficiently and at a substantially lower cost than high-performance GPUs. Although the initial investment to develop ASIC infrastructure is considerable, the long-term cost of running GenAI workloads on these chips is expected to be lower once the upfront expense is absorbed. For example, in April 2025, Google unveiled Ironwood, its seventh-generation Tensor Processing Unit (TPU), specifically designed for inference workloads.⁹ While Google's in-house TPUs were once limited to internal use, the company is expanding external access to drive faster growth of its cloud business. Marvell Technology projects the custom computing device market will surge to US\$55.4 billion by 2028, more than eight times its size in 2023.¹⁰

⁶ <https://www.wsj.com/tech/ai/meta-aims-to-fully-automate-ad-creation-using-ai-7d82e249/>

⁷ <https://blog.google/technology/ai/io-2025-keynote/>

⁸ Source: PwC's AI Agent Survey (May 2025)

⁹ <https://blog.google/products/google-cloud/ironwood-tpu-age-of-inference/>

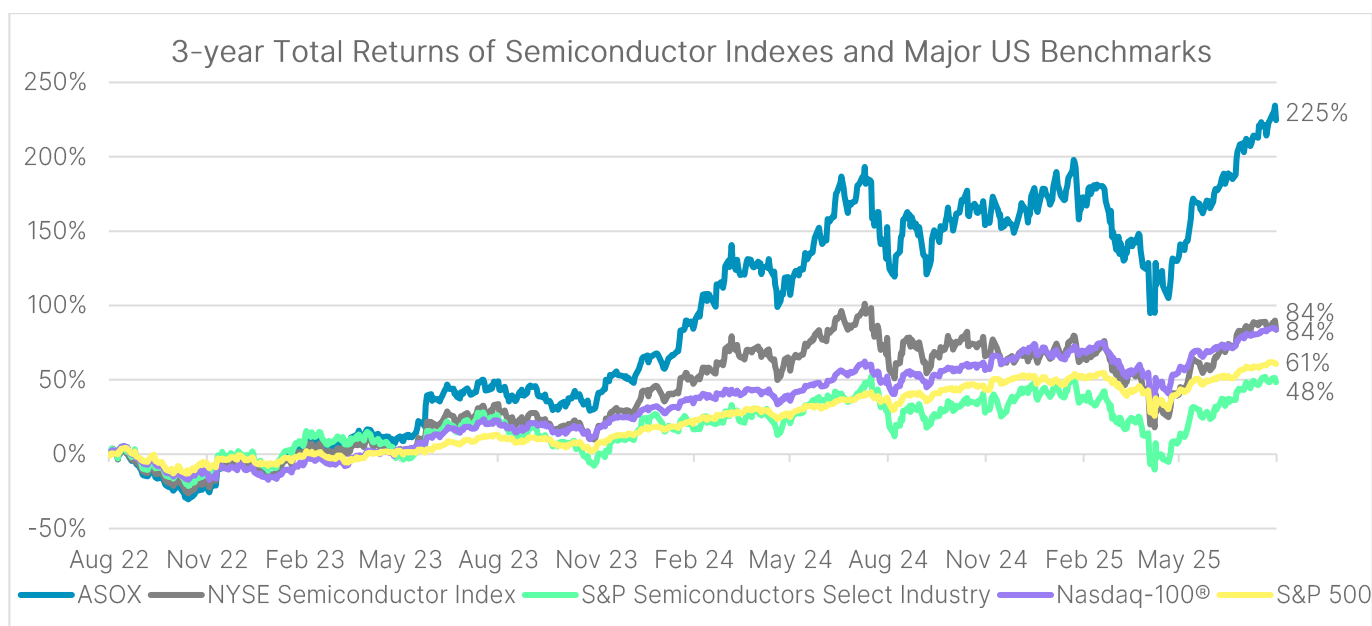
¹⁰ <https://www.marvell.com/content/dam/marvell/en/company/assets/marvell-custom-ai-investor-event-2025.pdf>

AI drives robust demand for high-bandwidth memory (HBM) technology

HBM represents a cutting-edge memory technology engineered to deliver faster data access while reducing energy consumption, which is critical for the performance of AI processing. HBM's market share in the dynamic random access memory (DRAM) segment is forecast to leap from 18% in 2024 to more than 50% by 2030.¹¹ Starting with the next-generation HBM4, the base die will be produced using logic processes, enabling lower power consumption and customizable features tailored to client requirements. Driven by escalating computational demands from AI training and inference workloads, HBM's market outlook remains strong. As the primary HBM supplier for Nvidia, holding a 62% share of global HBM shipments in Q2 2025¹², SK Hynix projects the global HBM market to expand by 30% annually through 2030.¹³

ASOX™ – the leading index tracking AI-focused semiconductor firms

Covering US-listed stocks and ADRs of companies engaged in the AI semiconductor value chain, Nasdaq's PHLX US AI Semiconductor™ Index (ASOX) delivered a total return of 225% over the past three years, nearly 2.7 times the return of the NYSE Semiconductor Index and more than 4.6 times that of the S&P Semiconductors Select Industry Index.



Source: Nasdaq Global Indexes, Bloomberg. As of July 31, 2025.

ASOX is a modified market capitalization-weighted index, with the top three constituents by market capitalization capped at 20%, 17% and 15%, respectively, and the rest capped at 12% during quarterly rebalancing. The number of constituents is limited to 20. Eligible firms must be classified under one of the five AI semiconductor value chain categories: (i) fabless, (ii) foundry, (iii) equipment, (iv) intellectual property (IP) and electronic design automation (EDA), and (v) packaging, assembly and testing. Three types of companies are excluded from the index:

- (i) integrated device manufacturers
- (ii) fabless firms with their top revenue source from analog ICs
- (iii) foundries without stated revenue from 7nm or smaller chips

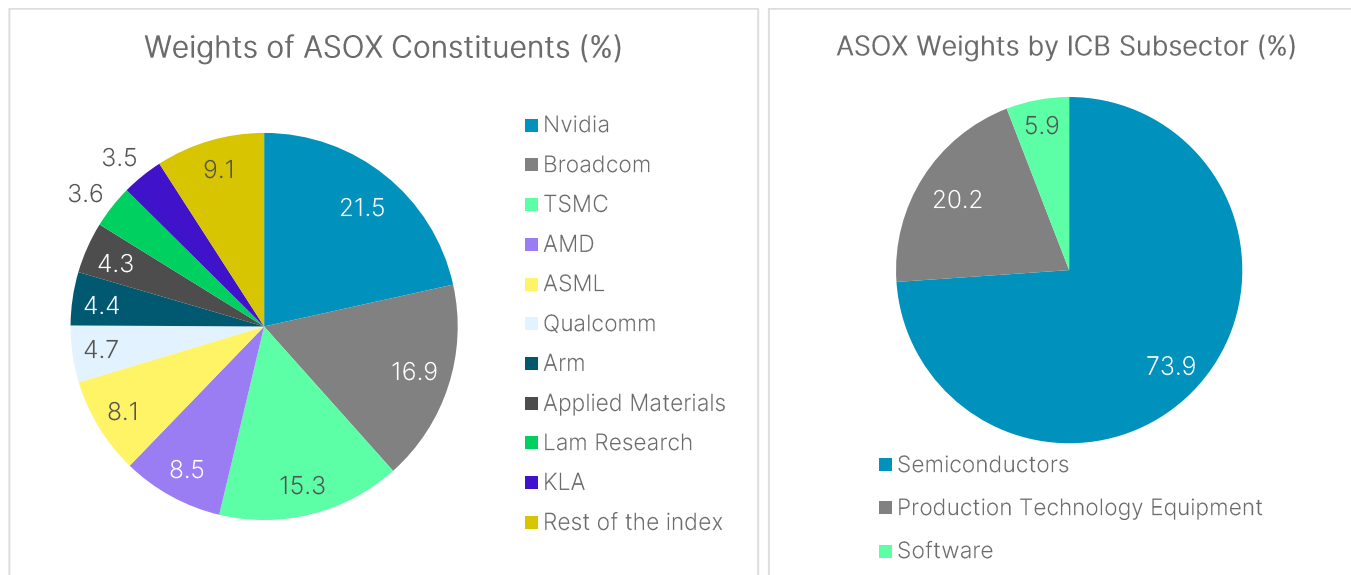
For the full index methodology, please visit our [website](#).

¹¹ <https://www.yolegroup.com/strategy-insights/memory-industry-at-a-crossroads-why-2025-marks-a-defining-year/>

¹² <https://www.counterpointresearch.com/en/insights/samsungs-q2-2025-memory-performance-disappoints-but-signals-h2-recovery/>

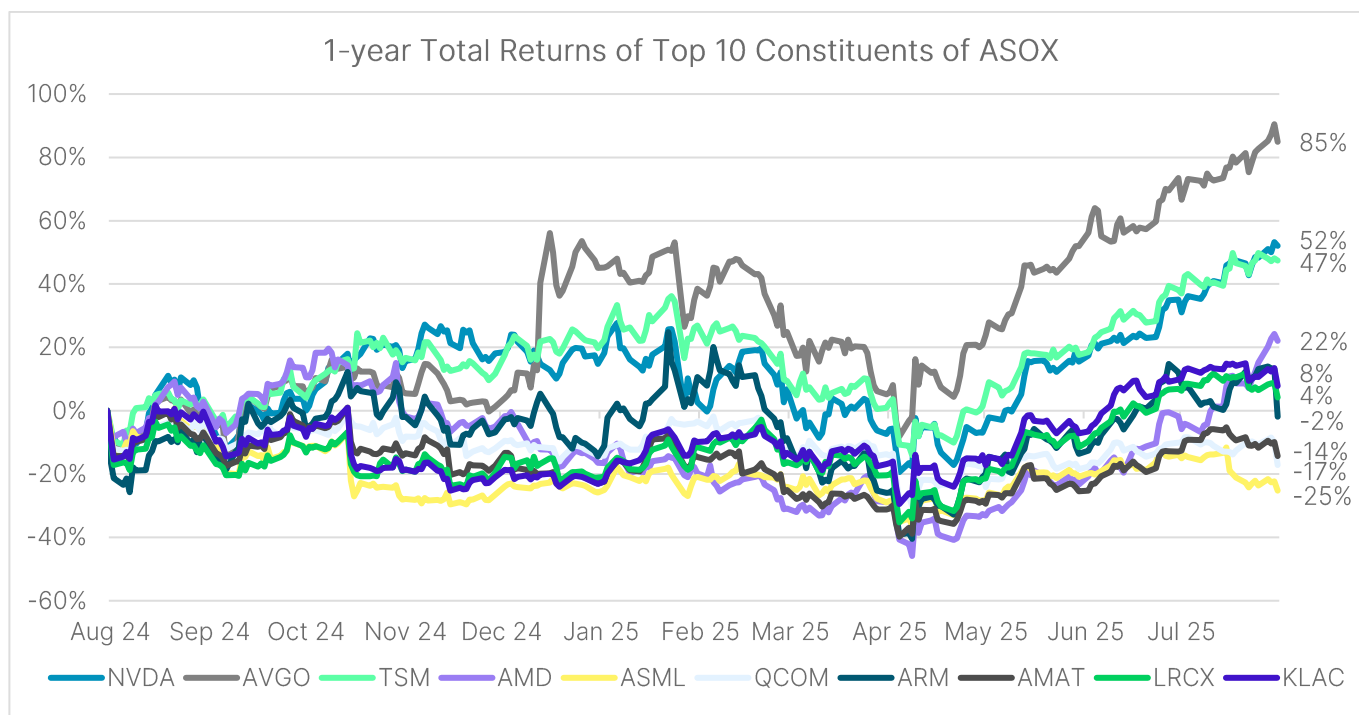
¹³ <https://www.reuters.com/world/asia-pacific/sk-hynix-expects-ai-memory-market-grow-30-year-2030-2025-08-11/>

As of the end of July 2025, ASOX included 18 stocks, with the five largest constituents accounting for 70.3% of the index weight. 73.9% of the index weight is in the Semiconductor Subsector and 20.2% is in the Production Technology Equipment Subsector, with the rest in the Software Subsector, according to the Industry Classification Benchmark (ICB) classification system.



Source: Nasdaq Global Indexes, FactSet. As of July 31, 2025.

Four of the five largest holdings posted positive total returns over the past 12 months. On average, the top five firms achieved a one-year total return of 36%. Within the same sector, stock performances varied significantly. The difference in one-year total returns between the best- and worst-performing stocks among the top 10 constituents (Broadcom and ASML) was a staggering 110 percentage points.¹⁴ This demonstrates the importance of diversification, even when investing in a single theme.



Source: Nasdaq Global Indexes, FactSet. As of July 31, 2025.

¹⁴ Source: FactSet. Difference in total return between Broadcom and ASML (US listing).

Nvidia (weight: 21.5%)

As the largest constituent in ASOX, Nvidia was the third-best performer among all constituents, gaining 52% over the past 12 months. It became the first publicly traded company in history to achieve a US\$4 trillion market valuation in July 2025, after joining the trillion-dollar club in May 2023.¹⁵ Nvidia's technological edge remains pronounced, as Blackwell shipments accelerate, driven by soaring AI reasoning demand and realized economies of scale. Beyond AI, the company identifies robotics as its most substantial addressable growth opportunity, with autonomous vehicles representing the first major commercial deployment. The firm recently unveiled new NVIDIA Omniverse libraries and NVIDIA Cosmos world foundation models that empower developers to build next-generation robots and autonomous vehicles by integrating AI reasoning with scalable, physically accurate simulations.¹⁶

Broadcom (weight: 16.9%)

As the second-largest constituent in ASOX, Broadcom emerged as the best performer among all index components, generating a one-year total return of 85% through July 2025. The company continues to dominate the AI ASIC and AI networking semiconductor markets. The chipmaker is engaged with the top seven hyperscalers for custom silicon products, including accelerators for Google, Meta and ByteDance. Moreover, Broadcom has significant exposure to enterprise software following its acquisition of VMware in 2023. Its infrastructure software division posted a 76% operating margin last quarter, up from 60% a year ago.¹⁷

TSMC (weight: 15.3%)

As the world's largest contract semiconductor manufacturer, TSMC ranks as the third-largest ASOX constituent, delivering a 47% total return over the trailing twelve months.¹⁸ With net revenue from its high-performance computing segment expanding to 60% from 52% a year ago, AI-driven demand remains the primary growth catalyst for TSMC, while its market dominance provides it with pricing power. Advanced chips with sizes 7nm or smaller accounted for 74% of the company's total wafer revenue in Q2 2025. Despite the US president's recent announcement of tariffs on semiconductor imports, Taiwan confirmed that TSMC secured an exemption from the latest levies given its substantial American manufacturing investments.¹⁹

Conclusion

Beyond the escalating demand for advanced training capabilities to support increasingly complex AI model architectures, AI inference has surfaced as a pivotal growth driver. The rapid increase in token volume signifies expanding usage and adoption of AI models. The proliferation of AI agents is poised to transform various industries and substantially increase compute demand. This dynamic landscape highlights the pivotal role of the semiconductor sector in driving the next wave of AI innovation.

Nasdaq's PHLX US AI Semiconductor Index (ASOX) delivered a total return of 225% over the three years through July 2025. The Mirae Asset TIGER PHLX US AI Semiconductor ETF (South Korea: 497570) tracks the index.

¹⁵ <https://www.reuters.com/technology/nvidia-sets-eye-1-trillion-market-value-2023-05-30/>

¹⁶ <https://nvidianews.nvidia.com/news/nvidia-opens-portals-to-world-of-robotics-with-new-omniverse-libraries-cosmos-physical-ai-models-and-ai-computing-infrastructure/>

¹⁷ <https://investors.broadcom.com/static-files/a5d6db22-6861-47e5-901b-13961fbc5321/>

¹⁸ Source: FactSet. Total return for TSMC (US listing).

¹⁹ <https://www.bloomberg.com/news/articles/2025-08-07/taiwan-chip-giant-surges-on-exemption-from-tough-new-trump-tariffs-on-chips>

Disclaimer:

Nasdaq®, PHLX US AI Semiconductor™, ASOX™ and Nasdaq-100® are registered trademarks of Nasdaq, Inc. The information contained above is provided for informational and educational purposes only, and nothing contained herein should be construed as investment advice, either on behalf of a particular security or an overall investment strategy. Neither Nasdaq, Inc. nor any of its affiliates makes any recommendation to buy or sell any security or any representation about the financial condition of any company. Statements regarding Nasdaq-listed companies or Nasdaq proprietary indexes are not guarantees of future performance. Actual results may differ materially from those expressed or implied. Past performance is not indicative of future results. Investors should undertake their own due diligence and carefully evaluate companies before investing. **ADVICE FROM A SECURITIES PROFESSIONAL IS STRONGLY ADVISED.**

© 2025. Nasdaq, Inc. All Rights Reserved.