

AI Tailwinds and Memory Shortages: Navigating the New Semiconductor Cycle

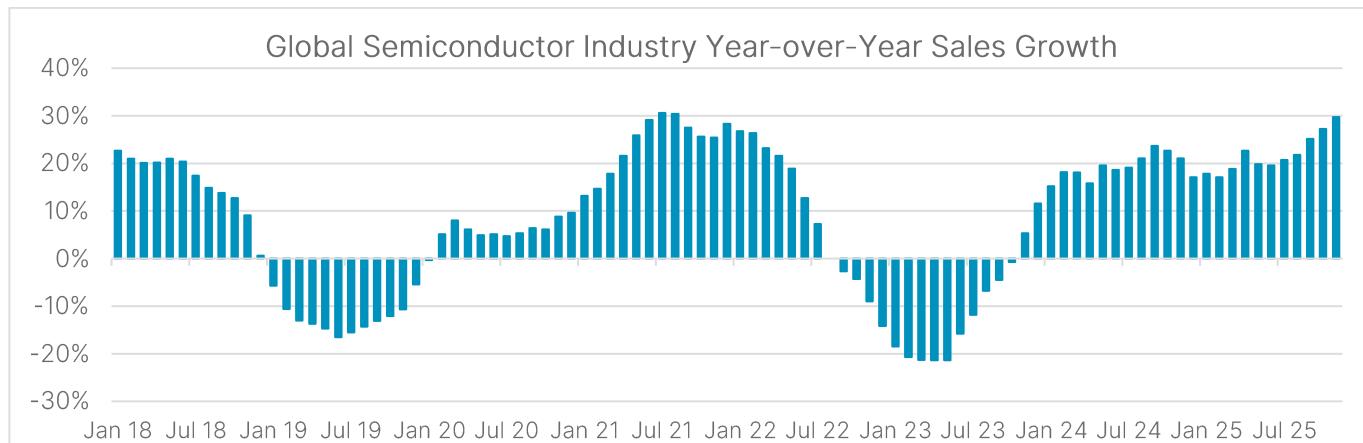
David Tsoi, CFA, CAIA, FRM, CESGA, CAMS, Head of Index Research, APAC

Artificial intelligence (AI) remains the defining technology of our era, with semiconductor companies at the forefront, powering transformative breakthroughs. After a strong rebound in 2024 fueled by demand for logic and memory chips, the global semiconductor market is projected to expand 22% in 2025, reaching US\$772 billion in total revenue, led by growth across the Americas and Asia Pacific. WSTS projects global semiconductor revenue to expand 26% in 2026, reaching US\$975 billion.¹



Source: World Semiconductor Trade Statistics. As of December 2, 2025.

Data center build-outs continue to be a major catalyst, particularly for firms driving AI and advanced semiconductor solutions. While 2024's recovery was uneven, 2025 marked a more balanced and broad-based upturn. Global semiconductor industry sales reached US\$75 billion in November 2025, up 30% year-over-year. Meanwhile, global sales of semiconductor manufacturing equipment by original equipment manufacturers (OEMs) are expected to hit a record US\$133 billion in 2025, up 14% year-on-year.²



Source: Semiconductor Industry Association. As of January 8, 2026.

¹ https://www.wsts.org/esraCMS/extension/media/f/WST/7310/WSTS_FC-Release-2025_11.pdf

² <https://www.semi.org/en/semi-press-release/global-semiconductor-equipment-sales-projected-to-reach-a-record-of-156-billion-dollars-in-2027-semi-reports>

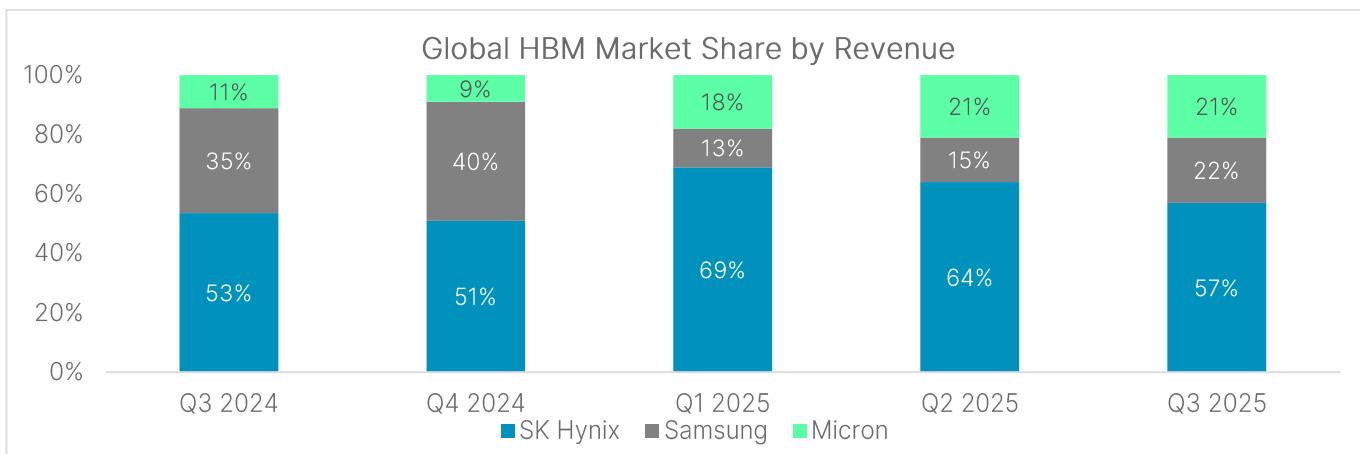
Sustained AI growth ahead

As AI-driven monetization opportunities materialize, hyperscaler capital spending continues to climb despite macroeconomic headwinds. Global data center capex surged 59% year-over-year in Q3 2025.³ Amazon, Google, Microsoft and Meta continue to raise data center capex expectations. Cloud providers are scaling accelerated compute, general-purpose servers and the supporting systems needed to deploy AI at production scale. Notably, this investment remains largely funded by operating cash flows, a critical distinction from the dot-com era, which should help ease near-term investor concerns. While hyperscaler capex has accelerated in recent years, it still trails levels seen in prior technology investment cycles. AI-related capex has recently equaled 0.8% of GDP, compared with peak levels of 1.5% or more during historic tech booms over the past 150 years.⁴

AI surge triggers global memory chip crunch

The global semiconductor industry faces an acute memory chip shortage, with cascading effects on device manufacturers and end users likely to persist beyond 2026. DRAM (Dynamic Random Access Memory) is the “working memory” inside personal computers (PCs) and servers. It is fast but loses data when power is off, and in AI systems it keeps models and their in-process calculations readily accessible for training and generating outputs. NAND flash is the “storage memory” used in solid-state drives and similar devices. It retains data even without power, making it well suited for storing AI datasets, saved model checkpoints and system logs. DRAM pricing has spiked as AI data center demand continues to outpace supply, creating a structural imbalance. Rapid expansion of AI infrastructure is fueling unprecedented requirements for high-performance, high-capacity memory and storage. Rather than scaling conventional DRAM and NAND for smartphones, PCs and consumer electronics, leading memory suppliers have reallocated capacity toward AI-centric products such as high-bandwidth memory (HBM) and DDR5. This pivot has constrained general-purpose memory availability and driven broad-based price inflation. Samsung and SK Hynix are reportedly poised to hike server memory prices by up to 70% in Q1, underscoring a decisive pivot to a seller’s market as relentless AI-driven demand strains global supply chains.⁵ Smartphone and PC segments now confront elevated costs and muted shipment growth. Global smartphone shipments are projected to fall 2.1% in 2026 as escalating semiconductor costs weigh on consumer demand.⁶

Micron projects HBM total addressable market to grow at approximately a 40% CAGR through 2028, rising from roughly US\$35 billion in 2025 to about US\$100 billion by 2028. This US\$100 billion milestone is now expected two years earlier than prior forecasts and exceeds the entire DRAM market size in 2024. The company has already secured pricing and volume commitments for its full 2026 HBM output.⁷ Micron’s global HBM revenue share has climbed sharply from 11% a year ago to 21% in the third quarter.



Source: Counterpoint Research. Note: Totals may not equal 100% due to rounding.

³ <https://www.delloro.com/news/data-center-capex-up-59-percent-on-continued-ai-investment-in-3q-2025>

⁴ <https://www.goldmansachs.com/insights/articles/why-ai-companies-may-invest-more-than-500-billion-in-2026>

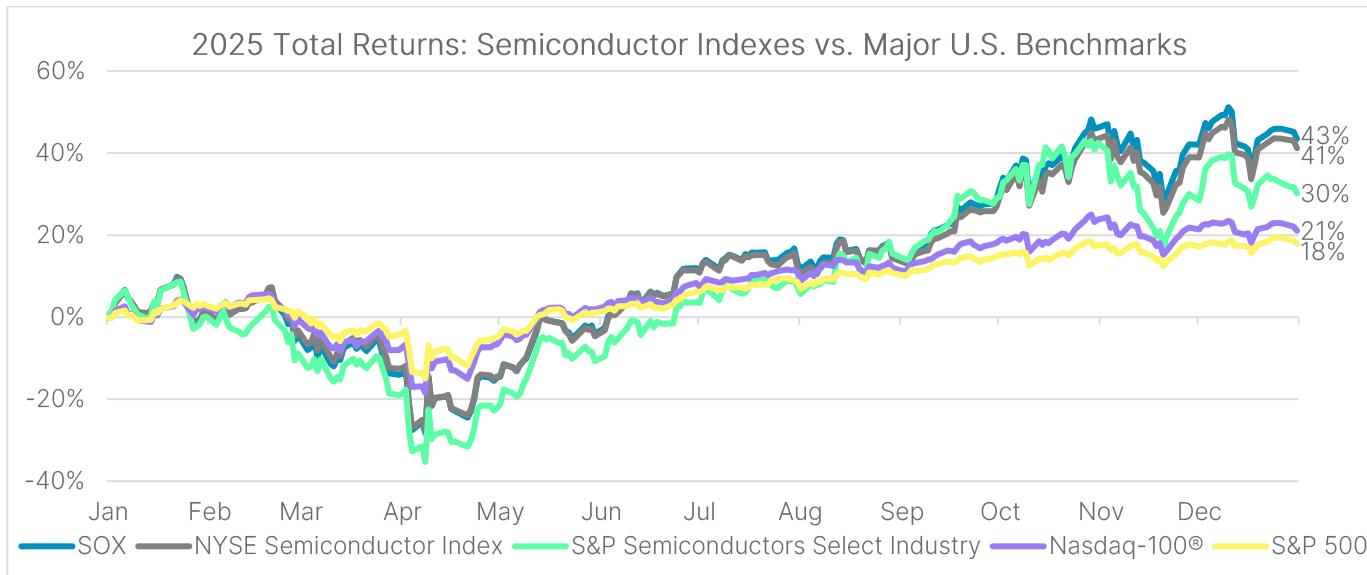
⁵ <https://www.kedglobal.com/korean-chipmakers/newsView/ked202601050006>

⁶ <https://counterpointresearch.com/en/insights/chip-crunch-to-curb-smartphone-output-in-2026>

⁷ <https://investors.micron.com/static-files/530bd7ed-a8c8-4687-af4a-8c129f740e09>

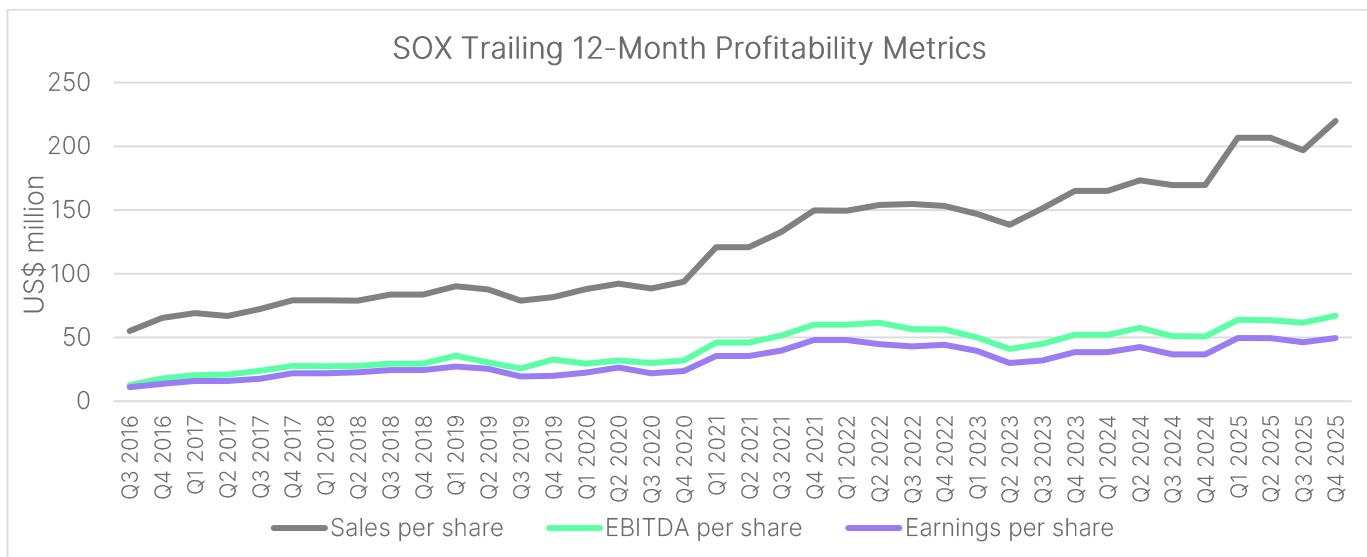
SOX™ – the leading index for the semiconductor industry

Semiconductors ranked among the strongest sectors in 2025 equity market performance. Representing the 30 largest U.S.-listed stocks and ADRs of companies engaged in semiconductor design, distribution, manufacturing and sales, Nasdaq's PHLX Semiconductor™ Index (SOX) delivered a 43.5% total return in 2025, 26 percentage points ahead of the S&P 500. SOX also outperformed the NYSE Semiconductor Index by 2 points and surpassed the equal-weighted S&P Semiconductors Select Industry Index by an impressive 13 points.



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

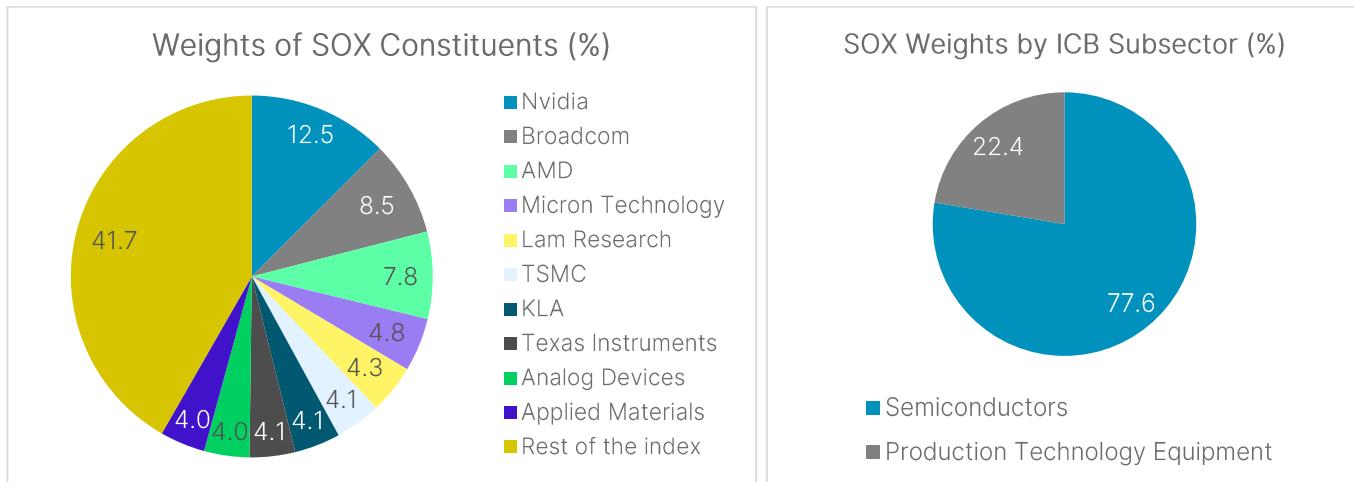
The journey, however, was not without challenges. Early in the year, SOX lagged the S&P 500 by 10 percentage points in Q1 amid new US tariff policies at both the country and sector level (including semiconductors), culminating in peak macro volatility at the beginning of April. These concerns gradually eased as companies implemented mitigation strategies and nations renegotiated lower tariffs in subsequent months, with occasional flare-ups specific to China and certain chips designed by Nvidia. A Section 232 investigation on semiconductors remains underway but is expected to grant meaningful exemptions to firms committing to expand U.S.-based production. This shift redirected investor attention to sector fundamentals, which remained solid, and triggered a sharp rebound after the April trough when SOX was down 28% YTD. By year-end 2025, SOX constituents' sales per share had climbed 44% over the past three years. Sentiment strengthened from Q2 onward as trade and macro risks receded, driving SOX to outperform the S&P 500 by 44 percentage points between Q2 and Q4.



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

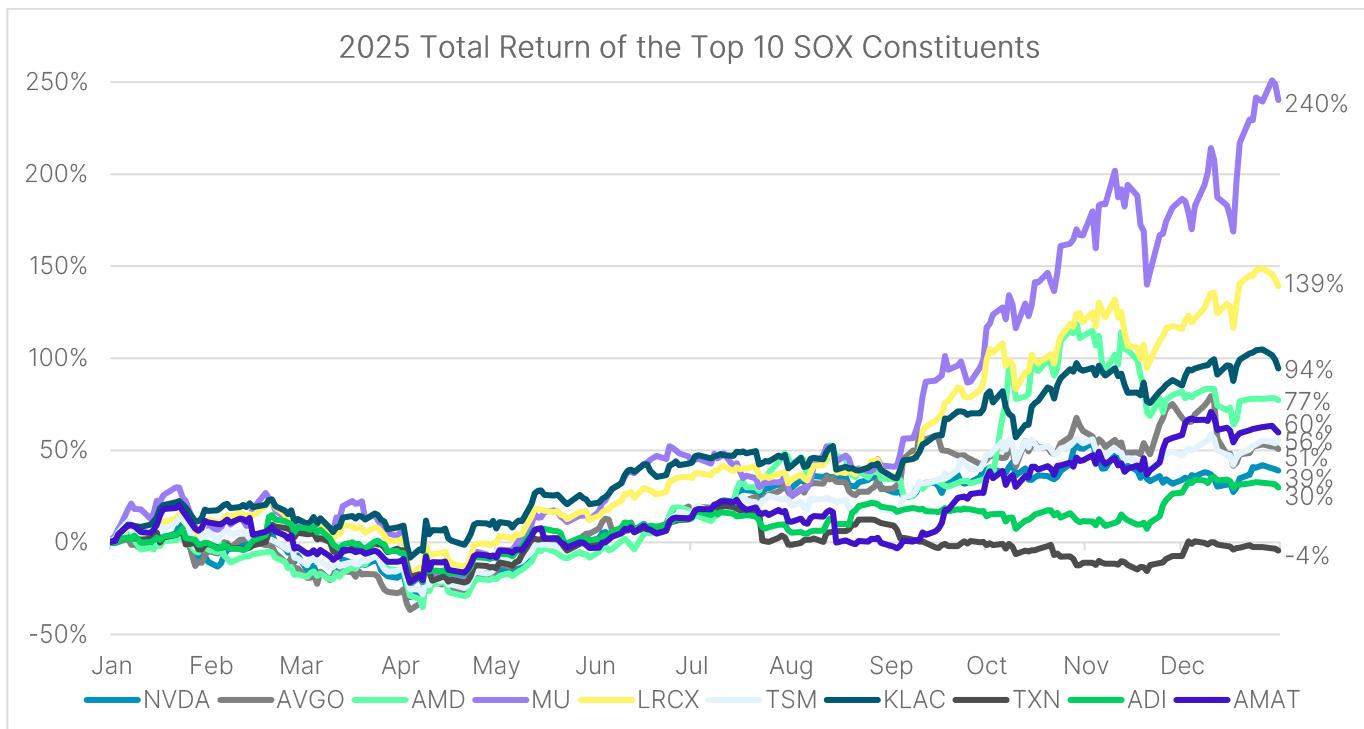
SOX is a modified market capitalization-weighted index, with the top three constituents by market capitalization capped at 12%, 10% and 8%, respectively, and the rest capped at 4% during quarterly rebalancing. For the full index methodology, please visit our [website](#).

The 10 largest constituents accounted for 58.3% of the index weight. 77.6% of the index weight is in the Semiconductor Subsector, with the rest in the Production Technology Equipment Subsector, according to the Industry Classification Benchmark (ICB) classification system.



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

SOX's top three performers in 2025 were Micron (+240%), Lam Research (+139%) and Credo (+114%). Nine of the ten largest holdings posted positive total returns in 2025. On average, the top 10 firms achieved a 78% gain. Within the same industry, stock performance varied significantly. The difference in total returns between the best- and worst-performing stocks among the top 10 constituents (Micron and Texas Instruments) was a staggering 245 percentage points. This demonstrates the importance of diversification, even when investing in a single sector or theme.

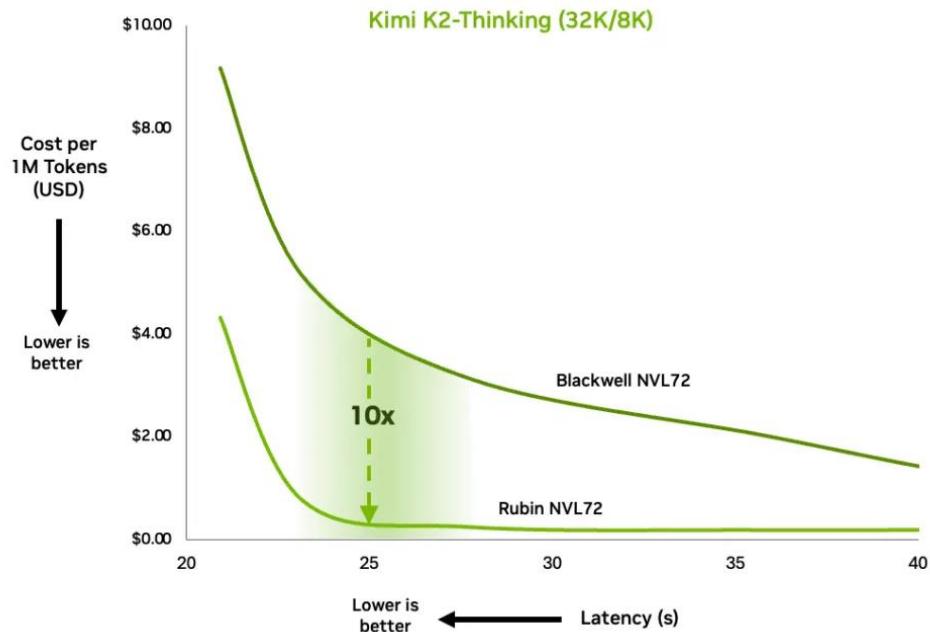


Nvidia (12.5% index weight)

As SOX's largest constituent, Nvidia advanced 39% in 2025, ranking eighth among the top ten holdings and 16th across all 30 index members. The company signed a non-exclusive license agreement with privately held Groq, acquiring its key assets, founder and other top talent.⁸ The deal reflects a recurring trend in recent years, with leading tech giants investing heavily in promising startups to secure technology and talent without pursuing full acquisitions. Groq specializes in low-latency AI inference chips, and Nvidia plans to integrate these processors into its AI factory architecture to extend support for a broader range of AI inference and real-time workloads. Groq asserts its language processing units (LPUs) deliver up to 10 times greater energy efficiency than graphics processing units (GPUs).⁹ According to PitchBook, the start-up was valued at US\$6.9 billion last September following a US\$750 million raise. As AI continues to scale, inference workloads will increasingly outpace demand for training, and Nvidia is aggressively positioning for this evolution. The partnership comes as the firm's major customers develop in-house AI silicon or adopt alternatives like Google's tensor processing units (TPUs).

Nvidia continues to push rack-scale AI systems, expand its open-source model library and advance physical AI applications such as autonomous vehicles. It maintains an annual cadence of platform upgrades, with the next generation, Vera Rubin, now in full production. The Rubin platform's "extreme codesign" integrates six chips into one system, cutting cost and power consumption while slashing inference token cost by up to tenfold. Compared with Blackwell, the Rubin GPU delivers five times the inference performance and 3.5 times the training throughput, alongside improved memory and networking bandwidth.¹⁰ Shipments are slated for the second half of this year.

Vera Rubin NVL72 Delivers 1/10 Inference Cost per Token vs. Blackwell NVL72



Source: Nvidia.

As AI becomes increasingly critical to autonomous vehicles, Nvidia is uniquely positioned to enable automakers in physical AI. The company introduced Alpamayo, a vehicle platform designed to let cars "reason" in real-world conditions. Carmakers can fine-tune the Alpamayo model for their needs, and the open-source solution aims to create vehicles that adapt to unexpected scenarios, such as traffic-light outages. Nvidia will debut its full-stack autonomous driving solution in the U.S. in Q1, starting with the Mercedes-Benz CLA.

⁸ <https://groq.com/newsroom/groq-and-nvidia-enter-non-exclusive-inference-technology-licensing-agreement-to-accelerate-ai-inference-at-global-scale>

⁹ <https://groq.com/blog/the-groq-lpu-explained>

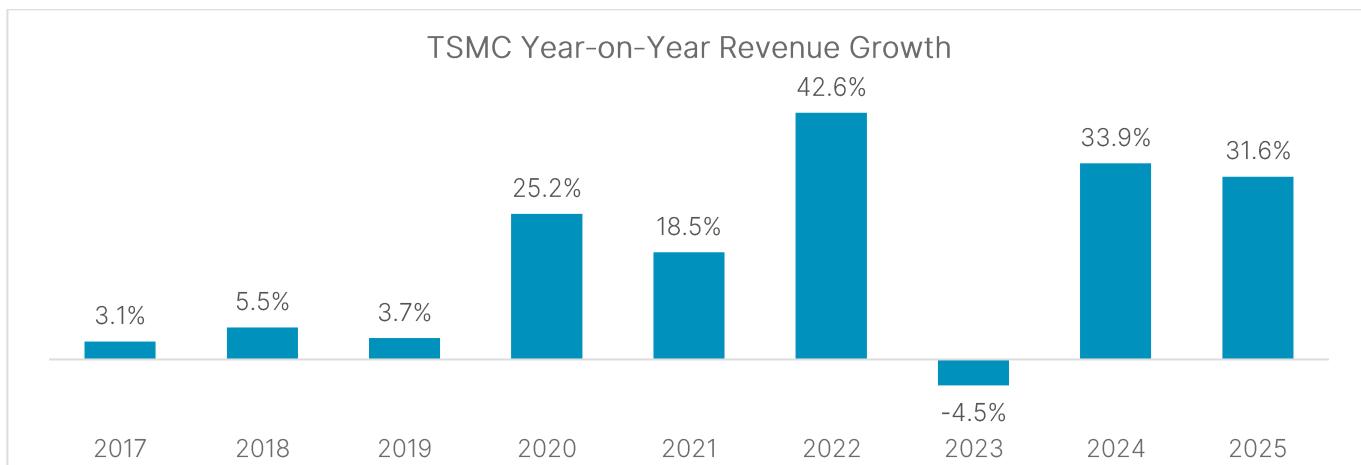
¹⁰ <https://developer.nvidia.com/blog/inside-the-nvidia-rubin-platform-six-new-chips-one-ai-supercomputer/>

Broadcom (8.5% index weight)

As SOX's second-largest constituent, Broadcom surged 51% in 2025, ranking seventh among the top ten holdings and 15th across all 30 index members. Its AI chip business is accelerating, fueled by demand for custom silicon amid a massive data center build-out. The Palo Alto-based company offers a broad portfolio spanning communications chips, networking components and software. Google remains its lead AI customer, with the latest TPU generation delivering exceptional performance and driving substantial orders. TPUs are no longer exclusive to Google, and significant demand from Anthropic¹¹ and Meta¹² is expected. Broadcom also secured an incremental US\$11 billion order from Anthropic and disclosed a fifth XPU (custom ASIC chips co-developed with major hyperscalers) customer with a US\$1 billion order for late-2026 delivery, though the client was not named.¹³

TSMC (4.1% index weight)

As SOX's sixth-largest constituent, TSMC ADR rose 56% in 2025, ranking sixth among the top ten holdings and 11th across all 30 index members. TSMC's structural growth drivers are expected to remain intact, anchored by its near-monopoly in AI accelerators and edge AI, and underpinned by a robust process roadmap and industry-leading advanced packaging. Volume production of its 2nm technology commenced in Q4 last year as scheduled.¹⁴ The Taiwan-based foundry reported full-year 2025 revenue of NT\$3.8 trillion (US\$120.6 billion), up 32% year-on-year.¹⁵ Strength likely stemmed from solid iPhone 17 volumes and data center chip demand, offsetting concerns that infrastructure buildout is outpacing AI adoption.



Source: TSMC, Bloomberg. As of January 9, 2026.

Conclusion

2025 delivered exceptional fundamental growth for the semiconductor sector, with Nasdaq's PHLX Semiconductor Index (SOX) posting a robust 43% total return, outperforming the S&P 500 by 26 percentage points. While AI remains the primary catalyst, investor focus is expected to shift toward diversification through broad-based exposure, narrowing the performance gap between AI-driven and non-AI names compared with prior years.

Funds tracking SOX include Invesco PHLX Semiconductor ETF (Nasdaq: SOXQ), Mirae Asset TIGER US PHLX Semiconductor Sector Nasdaq ETF (South Korea: 381180), Cathay PHLX Semiconductor ETF (Taiwan: 00830), Global X Semiconductor ETF (Japan: 2243) and Yurie PHLX Semiconductor Index Fund (South Korea: 7D01596). Mirae Asset TIGER Synth-US PHLX Semiconductor Sector Leverage ETF (South Korea: 423920) tracks SOX with 2x leverage.

¹¹ <https://www.anthropic.com/news/expanding-our-use-of-google-cloud-tpus-and-services>

¹² <https://www.wsj.com/tech/ai/meta-is-in-talks-to-use-googles-chips-in-challenge-to-nvidia-be390a51>

¹³ <https://www.morningstar.com/stocks/xnas/avgo/earnings-transcripts>

¹⁴ https://www.tsmc.com/english/dedicatedFoundry/technology/logic/l_2nm

¹⁵ <https://pr.tsmc.com/english/news/3278>

Disclaimer:

Nasdaq®, PHLX Semiconductor™, SOX™ 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.**

© 2026. Nasdaq, Inc. All Rights Reserved.