KASASV

KOREAN AMERICAN SEMICONDUCTOR ASSOCIATION IN SILICON VALLEY

January 2025

사람들이 당신을 좋아하게 만드는 방법은 무엇일까요?

다른 사람으로 하여금 자신이 중요한 사람이라는 느낌을 갖도록 노력하라. 존 듀이 교수는 중요한 사람이 되려는 욕망이야말로 인간 본성 중에서도 가장 깊은 충동이라고 말했다. 윌리엄 제임스 교수는 "인간 본성의 가장 깊은 곳에 자리 잡고 있는 원리는 인정받고 싶은 갈망이다"라고 하였다. 사람들은 자신들이 만나는 다른 사람들의 인정을 받고 싶어 한다. 적어도 자신의 세상 안에서는 자신이 중요한 사람이라는 느낌을 원한다.

마음에서 우러나오는 칭찬을 하고, 아낌없이 칭찬해 보시기를, 싸구려 칭찬이나 진지하지 않는 아첨이 아니라.... 데일 카네기

INTRO

E-mail 주소는 blind 로 나가기때문에 공유가 되지는 않습니다. 혹시 개인 메일주소를 원하시면 저에게 알려주시면 update 하도록 하겠습니다. 그리고 Newsletter distribuiton 을 원치않으시면, 알려주시기 바랍니다.

지난 12 월 27 일에는 2024 년도 4 번째 골프 토너먼트겸 연말모임이 있었습니다. 참석해주신 모든분들께 감사드립니다. 추가 내용관련해서는 저희 웹사이트 announcement 를 참조하시기 바랍니다.

2025 1 월초에는 경상국립대학교 세라믹 공학과 학부방문이 있어 저희와 미팅이 있을 예정입니다.

저희는 미국 Internal Revenue Code 501(c)(3)에 의거하여 Federal Income tax exemption 을 받은 비영리단체로 donation/협찬으로 운영이 되고 있습니다. 관심이 있으시면 저희 웹사이트 Partners tab 을 보시면 좀 더 자세히 보실수 있습니다. 올해부터는 off-line event 하고있으니, 스케쥴은 웹사이트(Announcement)를 참조하시기 바랍니다

– 사이먼 리 드림/협회 상근임원 www.kasainsv.com

감사합니다

How we will reach a 1 trillion transistor GPU > advances in semiconductors are feeding the AI boom

Dec 30, 2024 Samuel K. Moore

In 1997 the IBM Deep Blue supercomputer defeated world chess champion Garry Kasparov. It was a groundbreaking demonstration of supercomputer technology and a first glimpse into how high-performance computing might one day overtake human-level intelligence. In the 10 years that followed, we began to use artificial intelligence for many practical tasks, such as facial recognition, language translation, and recommending movies and merchandise.

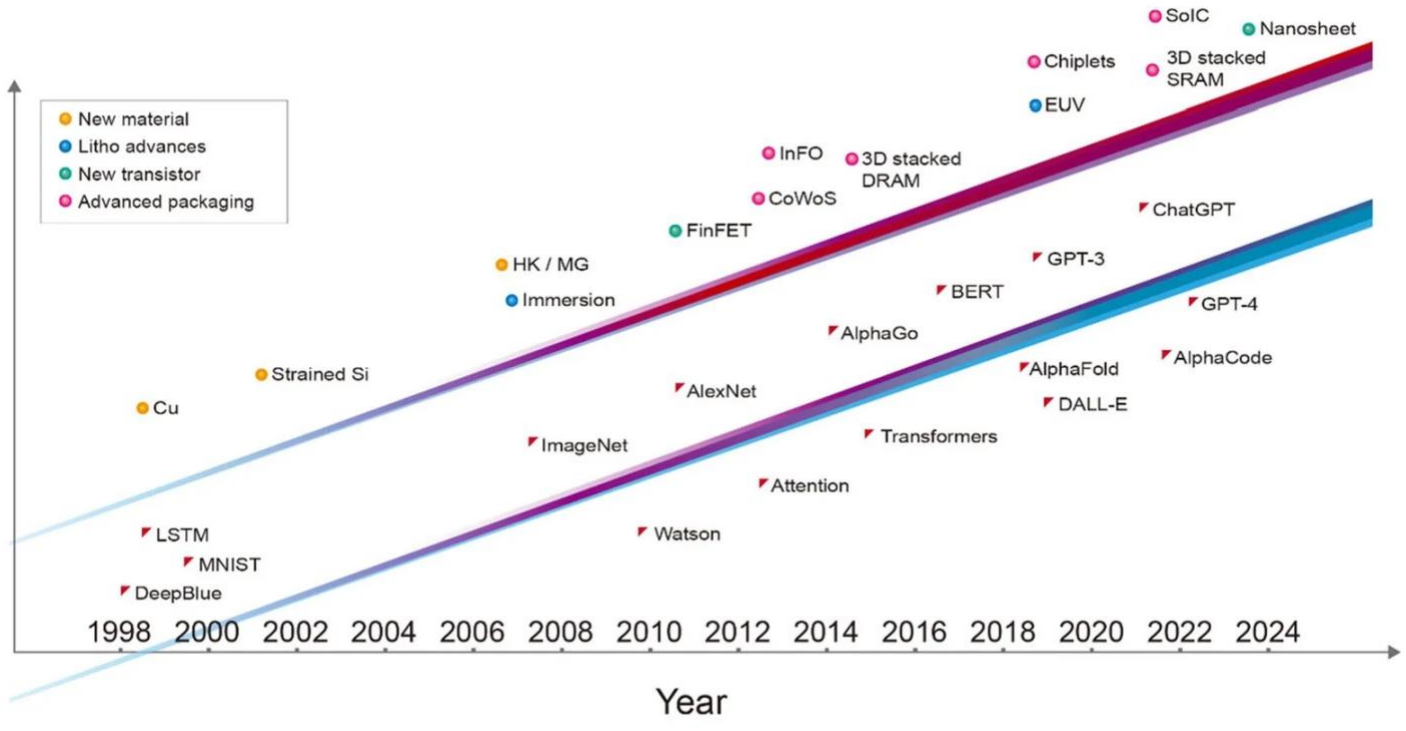
Fast-forward another decade and a half and artificial intelligence has advanced to the point where it can "synthesize knowledge." Generative AI, such as ChatGPT and Stable Diffusion, can compose poems, create artwork, diagnose disease, write summary reports and computer code, and even design integrated circuits that rival those made by humans.

Tremendous opportunities lie ahead for artificial intelligence to become a digital assistant to all human endeavors. ChatGPT is a good example of how AI has democratized the use of high-performance computing, providing benefits to every individual in society.

All those marvelous AI applications have been due to three factors: innovations in efficient machine-learning algorithms, the availability of massive amounts of data on which to train neural networks, and progress in energy-efficient computing through the advancement of semiconductor technology. This last contribution to the generative AI revolution has received less than its fair share of credit, despite its ubiquity.

Over the last three decades, the major milestones in AI were all enabled by the leading-edge semiconductor technology of the time and would have been impossible without it. Deep Blue was implemented with a mix of 0.6- and 0.35-micrometer-node chip-manufacturing technology. The deep neural network that won the ImageNet competition, kicking off the current era of machine learning, was implemented with 40-nanometer technology. AlphaGo conquered the game of Go using 28-nm technology, and the initial version of ChatGPT was trained on computers built with 5-nm technology. The most recent incarnation of ChatGPT is powered by servers using even more advanced 4-nm technology. Each layer of the computer systems involved, from software and algorithms down to the architecture, circuit design, and device technology, acts as a multiplier for the performance of AI. But it's fair to say that the foundational transistor-device technology is what has enabled the advancement of the layers above.

If the AI revolution is to continue at its current pace, it's going to need even more from the semiconductor industry. Within a decade, it will need a 1-trillion-transistor GPU—that is, a GPU with 10 times as many devices as is typical today.



Advanced in semiconductor technology [top line] – including new materials, advances in lithography, new types of transistors, and advanced packaging-have driven the development of more capable AI systems [bottom line]

Relentless Growth in AI Model Sizes

The computation and memory access required for AI training have increased by orders of magnitude in the past five years. Training GPT-3, for example, requires the equivalent of more than 5 billion billion operations per second of computation for an entire day (that's 5,000 petaflops-days), and 3 trillion bytes (3 terabytes) of memory capacity.

Both the computing power and the memory access needed for new generative AI applications continue to grow rapidly. We now need to answer a pressing question: How can semiconductor technology keep pace?

From Integrated Devices to Integrated Chiplets

Since the invention of the integrated circuit, semiconductor technology has been about scaling down in feature size so that we can cram more transistors into a thumbnail-size chip. Today, integration has risen one level higher; we are going beyond 2D scaling into 3D system integration. We are now putting together many chips into a tightly integrated, massively interconnected system. This is a paradigm shift in semiconductor-technology integration.

In the era of AI, the capability of a system is directly proportional to the number of transistors integrated into that system. One of the main limitations is that lithographic chipmaking tools have been designed to make ICs of no more than about 800 square millimeters, what's called the reticle limit. But we can now extend the size of the integrated system beyond lithography's reticle limit. By attaching several chips onto a larger interposer—a piece of silicon into which interconnects are built—we can integrate a system that contains a much larger number of devices than what is possible on a single chip. For example, TSMC's chip-on-wafer-on-substrate (CoWoS) technology can accommodate up to six reticle fields' worth of compute chips, along with a dozen high-bandwidth-memory (HBM) chips.

How Nvidia Uses CoWoS Advanced Packaging

CoWoS, TSMC's chip-on-wafer-on-silicon advanced packaging technology, has already been deployed in products. Examples include the Nvidia Ampere and Hopper GPUs. Each consists of one GPU die with six high-bandwidth memory cubes all on a silicon interposer. The compute GPU die is about as large as chipmaking tools will currently allow. Ampere has 54 billion transistors, and Hopper has 80 billion. The transition from 7-nm technology to the denser 4-nm technology made it possible to pack 50 percent more transistors on essentially the same area. Ampere and Hopper are the workhorses for today's large language model (LLM) training. It takes tens of thousands of these processors to train ChatGPT.

HBMs are an example of the other key semiconductor technology that is increasingly important for AI: the ability to integrate systems by stacking chips atop one another, what we at TSMC call system-on-integrated-chips (SoIC). An HBM consists of a stack of vertically interconnected chips of DRAM atop a control logic IC. It uses vertical interconnects called through-silicon-vias (TSVs) to get signals through each chip and solder bumps to form the connections between the memory chips. Today, high-performance GPUs use HBM extensively.

Going forward, 3D SoIC technology can provide a "bumpless alternative" to the conventional HBM technology of today, delivering far denser vertical interconnection between the stacked chips. Recent advances have shown HBM test structures with 12 layers of chips stacked using hybrid bonding, a copper-to-copper connection with a higher density than solder bumps can provide. Bonded at low temperature on top of a larger base logic chip, this memory system has a total thickness of just 600 μm.

With a high-performance computing system composed of a large number of dies running large AI models, high-speed wired communication may quickly limit the computation speed. Today, optical interconnects are already being used to connect server racks in data centers. We will soon need optical interfaces based on silicon photonics that are packaged together with GPUs and CPUs. This will allow the scaling up of energy- and area-efficient bandwidths for direct, optical GPU-to-GPU communication, such that hundreds of servers can behave as a single giant GPU with a unified memory. Because of the demand from AI applications, silicon photonics will become one of the semiconductor industry's most important enabling technologies.

Toward a Trillion Transistor GPU

As noted already, typical GPU chips used for AI training have already reached the reticle field limit. And their transistor count is about 100 billion devices. The continuation of the trend of increasing transistor count will require multiple chips, interconnected with 2.5D or 3D integration, to perform the computation. The integration of multiple chips, either by CoWoS or SoIC and related advanced packaging technologies, allows for a much larger total transistor count per system than can be squeezed into a single chip. We forecast that within a decade a multichiplet GPU will have more than 1 trillion transistors.

We'll need to link all these chiplets together in a 3D stack, but fortunately, industry has been able to rapidly scale down the pitch of vertical interconnects, increasing the density of connections. And there is plenty of room for more. We see no reason why the interconnect density can't grow by an order of magnitude, and even beyond.

Energy-Efficient Performance Trend for GPUs

So, how do all these innovative hardware technologies contribute to the performance of a system?

We can see the trend already in server GPUs if we look at the steady improvement in a metric called energy-efficient performance. EEP is a combined measure of the energy efficiency and speed of a system. Over the past 15 years, the semiconductor industry has increased energy-efficient performance about threefold every two years. We believe this trend will continue at historical rates. It will be driven by innovations from many sources, including new materials, device and integration technology, extreme ultraviolet (EUV) lithography, circuit design, system architecture design, and the co-optimization of all these technology elements, among other things.

Energy-efficient performance improves 3x every 2 years.
Drivers for future improvements include: 1. New transistors and materials 2. EUV lithography and design-technology co-optimization (DTCO) 3. Circuit and architecture innovations 4. More advanced packaging and system-technology co-optimization (STCO)

In particular, the EEP increase will be enabled by the advanced packaging technologies we've been discussing here. Additionally, concepts such as system-technology co-optimization (STCO), where the different functional parts of a GPU are separated onto their own chiplets and built using the best performing and most economical technologies for each, will become increasingly critical.

A Mead-Conway Moment for 3D Integrated Circuits

In 1978, Carver Mead, a professor at the California Institute of Technology, and Lynn Conway at Xerox PARC invented a computer-aided design method for integrated circuits. They used a set of design rules to describe chip scaling so that engineers could easily design very-large-scale integration (VLSI) circuits without much knowledge of process technology.

That same sort of capability is needed for 3D chip design. Today, designers need to know chip design, system-architecture design, and hardware and software optimization. Manufacturers need to know chip technology, 3D IC technology, and advanced packaging technology. As we did in 1978, we again need a common language to describe these technologies in a way that electronic design tools understand. Such a hardware description language gives designers a free hand to work on a 3D IC system design, regardless of the underlying technology. It's on the way: An open-source standard, called 3Dblox, has already been embraced by most of today's technology companies and electronic design automation (EDA) companies.

The Future Beyond the Tunnel

In the era of artificial intelligence, semiconductor technology is a key enabler for new AI capabilities and applications. A new GPU is no longer restricted by the standard sizes and form factors of the past. New semiconductor technology is no longer limited to scaling down the next-generation transistors on a two-dimensional plane. An integrated AI system can be

composed of as many energy-efficient transistors as is practical, an efficient system architecture for specialized compute workloads, and an optimized relationship between software and hardware.

For the past 50 years, semiconductor-technology development has felt like walking inside a tunnel. The road ahead was clear, as there was a well-defined path. And everyone knew what needed to be done: shrink the transistor.

Now, we have reached the end of the tunnel. From here, semiconductor technology will get harder to develop. Yet, beyond the tunnel, many more possibilities lie ahead. We are no longer bound by the confines of the past.

Micron's expansion $15 billion expansion

Dec 31, 2024 Don Day

Micron is well underway on its project to expand its Boise plant.

Boise-based Micron Technology has started its $15 billion project to expand its Boise site with an expanded fabrication plant and increased research and development space.

The project, which is the largest-ever construction project in Idaho, will combine 8,000 tons of steel to build 15 new buildings. The new fab will be supported by one of the state's largest-ever office buildings, a 2,800-space parking garage, water treatment plant and more.

Micron has seen a significant run-up in its stock over the past year, more than doubling in price—in part due to a boom in demand for artificial intelligence chips. Scott Gatzmeier, Micron's VP for expansion, told a crowd at the Boise Metro Chamber of Commerce gathering in Sun Valley earlier this spring that each of AI chipmaker Nvidia's leading-edge chips is supported by 64 Micron chips.

The new fab will be 12 times the size of the current facility on site, and will allow Micron to both test and develop new cutting-edge chips as well as put them into production. The Boise site will work in tandem with a new operation near Syracuse, New York that will mass-produce memory products.

There are already 11 cranes on site, and at full tilt, the project will have more than 30 cranes hoisting construction materials into the air.

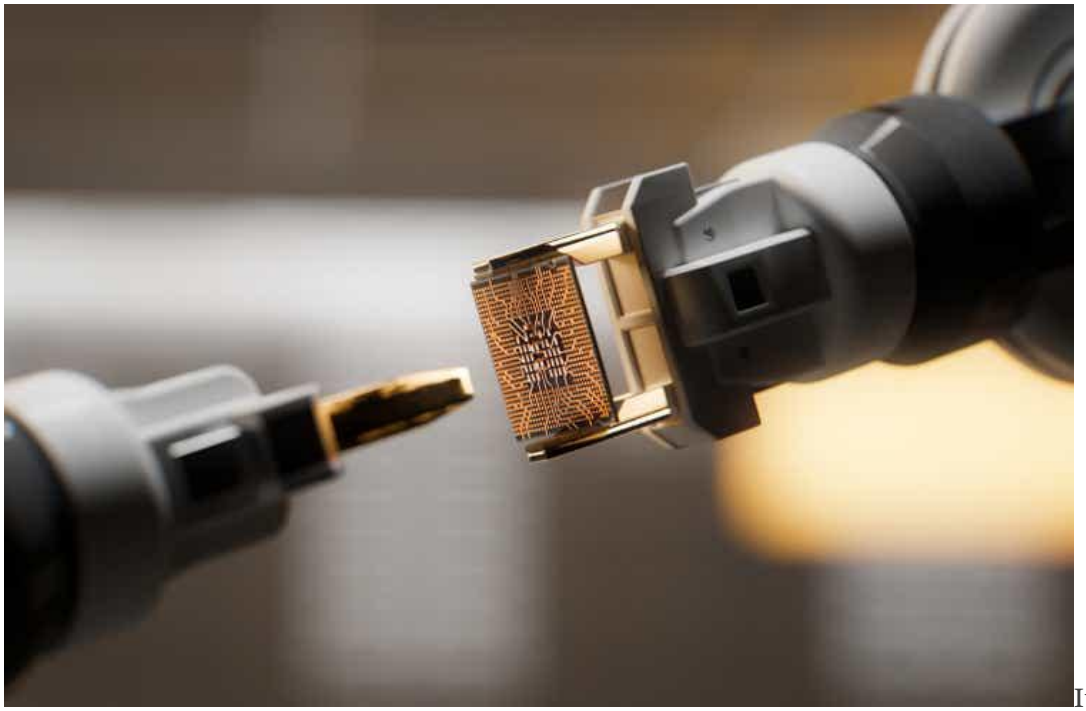
The project will be supported by as many as 4,000 construction workers, with three on-site concrete batch plants, and a rock-crushing operation.

Semiconductor Recovery Expands Beyond AI To Automotive, PCs

Dec 14., 2024 Markit

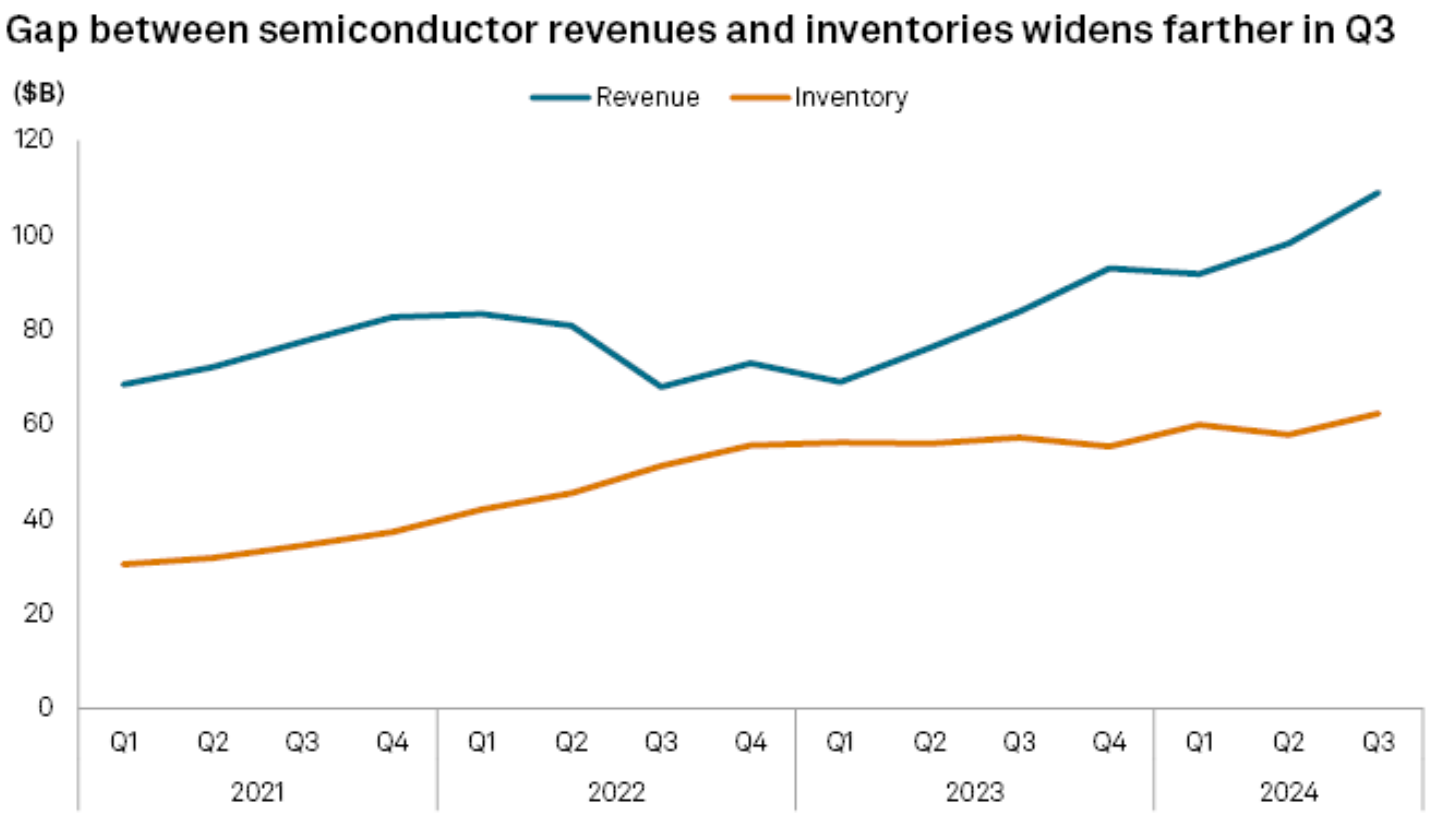
Summary

- For the first time in several quarters, the broader semiconductor sector is showing early signs of recovery in terms of clearing inventory.
- The gap between chip companies' revenue and inventories reached its highest point in the third quarter since 2021 when there was a widespread shortage of semiconductors of all types.
- Until the third quarter, semiconductor demand was mainly driven by chips used for AI, primarily those from advanced chipmaker NVIDIA.


Iuza studios

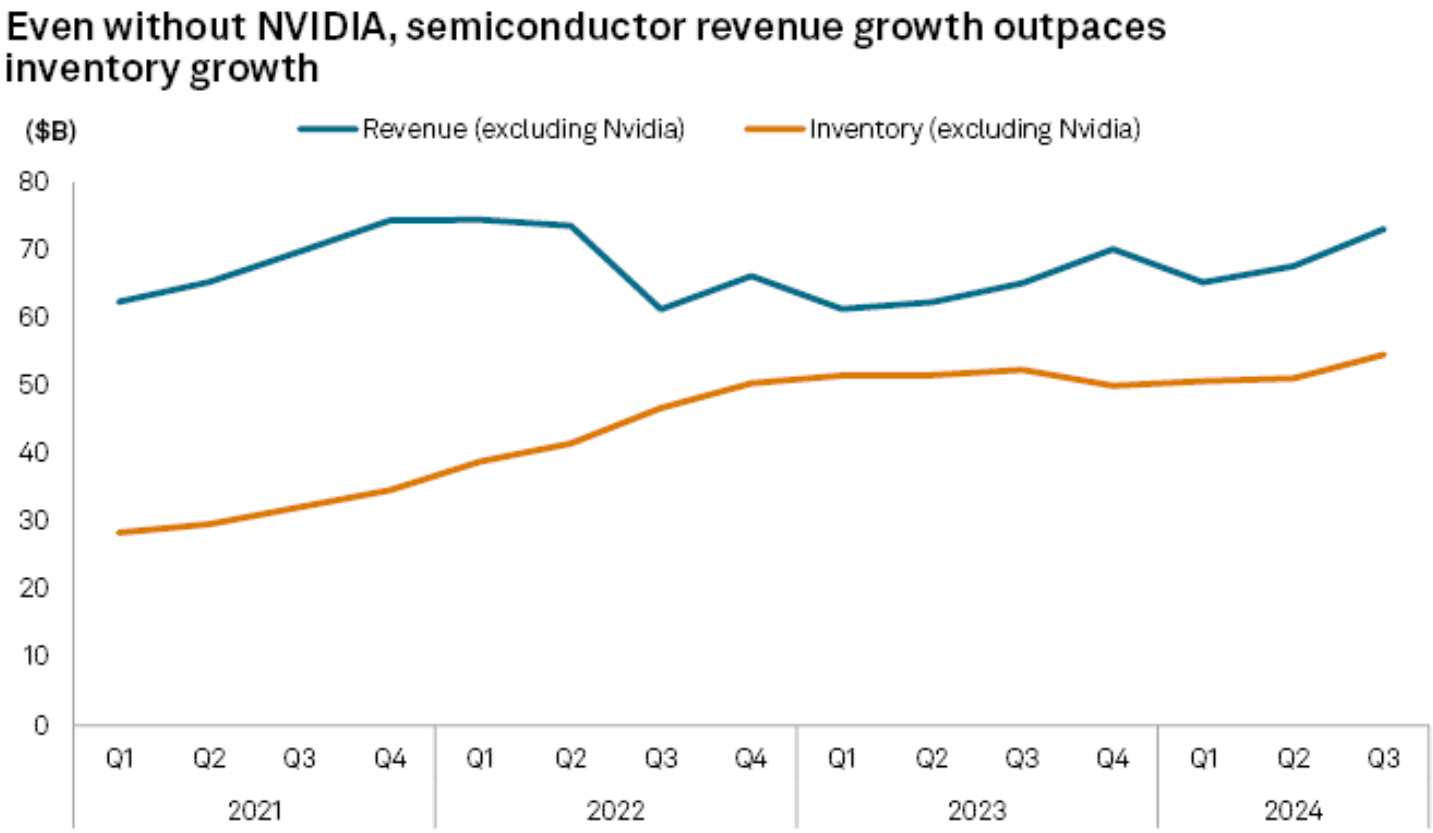
For the first time in several quarters, the broader semiconductor sector is showing early signs of recovery in terms of clearing inventory.

The gap between chip companies' revenue and inventories - a sign the companies are selling more chips than they are holding on to - reached its highest point in the third quarter since 2021 when there was a widespread shortage of semiconductors of all types. In reaction to outsize demand in 2021 and 2022, semiconductor companies built up inventories but found themselves with excess supply as demand collapsed.



Data compiled Dec. 12, 2024.
Includes constituents in S&P Semiconductors Select Industry Index.
Source: S&P Global Market Intelligence.

Until the third quarter, semiconductor demand was mainly driven by chips used for AI, primarily those from advanced chipmaker NVIDIA Corp. (NVDA). Excluding NVIDIA, which sees strong sales and weak inventories, the gap between revenue and inventories is smaller, although revenue is approaching the all-time highs of late 2021 and early 2022.



Data compiled Dec. 12, 2024.
Includes constituents in S&P Semiconductors Select Industry Index.
Source: S&P Global Market Intelligence.

Autos see strength

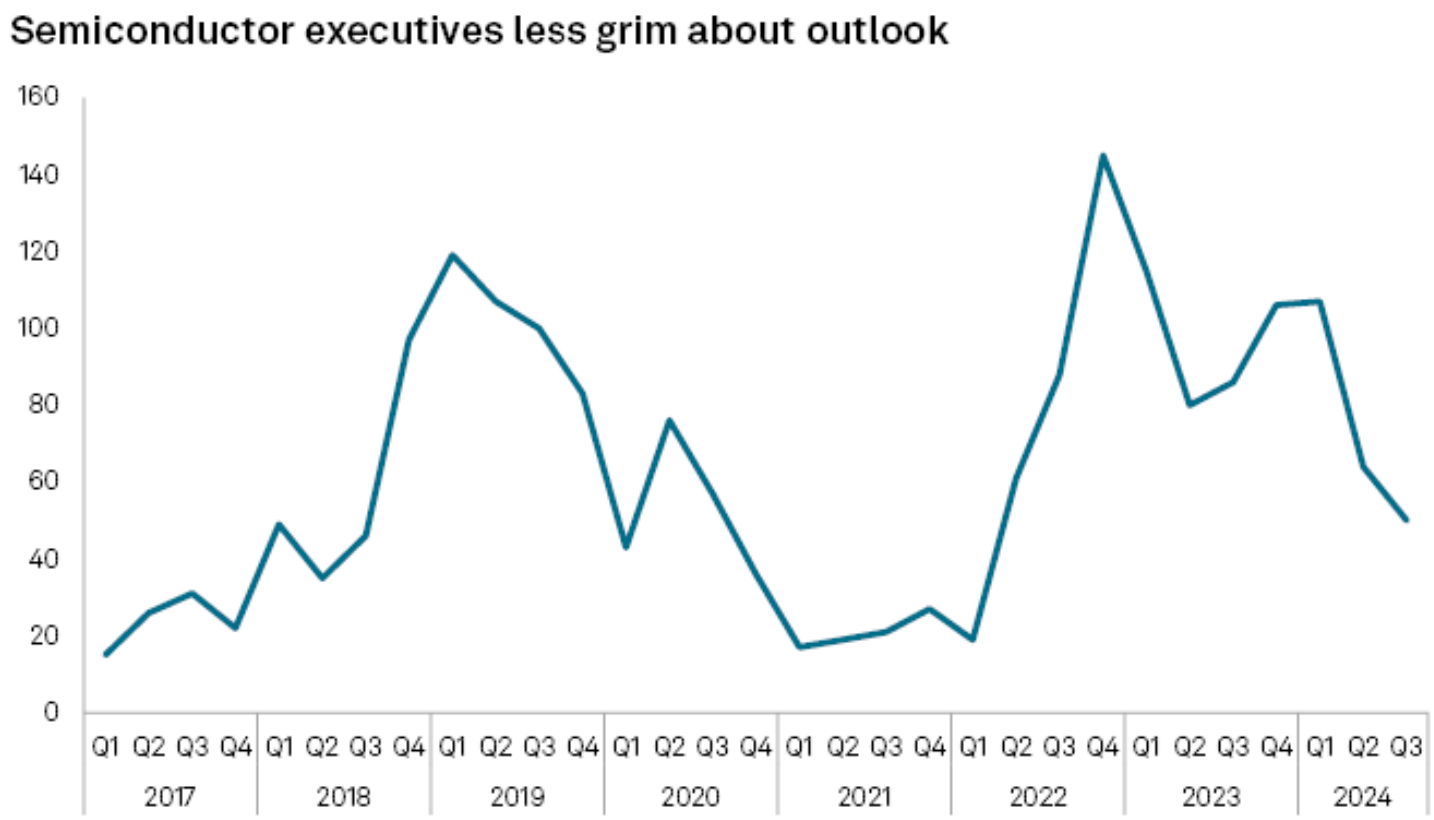
One strong end market in the third quarter was the automotive industry, which saw a notable recovery in the Chinese electric vehicle industry and a smaller ramp in the US. PCs and personal electronics also experienced an increase in demand, several semiconductor companies reported.

Analog Devices Inc. (ADI), Qualcomm Inc. (WCOM) and Texas Instruments Inc. (TXN) all said that the EV market drove revenue growth in the most recently ended quarter.

"Toward the end of [fiscal] Q3, bookings started to improve and that continued throughout our [fiscal] Q4 with stronger demand in China, reflecting EV volume growth, share gains and content growth," Analog Devices CFO Richard Puccio said on an earnings call. Analog Devices' fiscal 2024 ended Nov. 2.

Qualcomm President and CEO Cristiano Amon also said on the company's fiscal fourth-quarter earnings call that he saw "strength" in the auto market, partly due to market share gains. Qualcomm's fiscal fourth quarter ended Sept. 29.

A sentiment analysis by S&P Global Market Intelligence indicates that semiconductor CEOs are less gloomy about the outlook, with the number of negative keywords mentioned in earnings calls reaching a two-year low. However, executives from major US semiconductor companies caution that their customers remain in the inventory digestion phase.



Data compiled Dec. 9, 2024.
Includes earnings transcripts from Advanced Micro Devices Inc., Analog Devices Inc., Broadcom Inc., First Solar Inc, Intel Corp., Microchip Technology Inc., Micron Technology Inc., Monolithic Power Systems inc., NVIDIA Corp., NXP Semiconductors N.V., ON Semiconductor Corp., Qualcomm Inc., Qorvo Inc., Skyworks Solutions Inc. and Texas Instruments Inc.
Keywords: slowing, slow, softness, soft, weakness, weak, correction.
Source: S&P Global Market Intelligence.

AI broadening

Amid improvements in more traditional end markets, AI semiconductor revenue is expanding beyond major players such as NVIDIA and Broadcom Inc. (AVGO). Companies like Astera Labs Inc. (ALAB), Marvell Technology Inc. (MRVL) and Credo Technology Group Holding Ltd. (CRDO) benefited from strong demand for datacenter chips, with revenue growth surging in the third quarter.

For instance, Marvell expanded its strategic relationship with Amazon.com (AMZN) Inc.'s Amazon Web Services to assist in chip production, as the hyperscaler aims to reduce its reliance on NVIDIA.

Microsoft Corp. (MSFT) first introduced its own custom AI chip, the Azure Maia AI chip, in November 2023 to power its Azure datacenters. Additionally, it has continued to debut custom chips, including data processing units and hardware security modules.

These moves reflect a broader trend among hyperscalers to manufacture their own chips as they expand their datacenter footprint. Broadcom and Marvell serve as key enablers and beneficiaries of this trend.

Broadcom President and CEO Hock Tan said on an earnings call in September that hyperscalers are likely to move "towards creating as much as possible their own compute silicon."

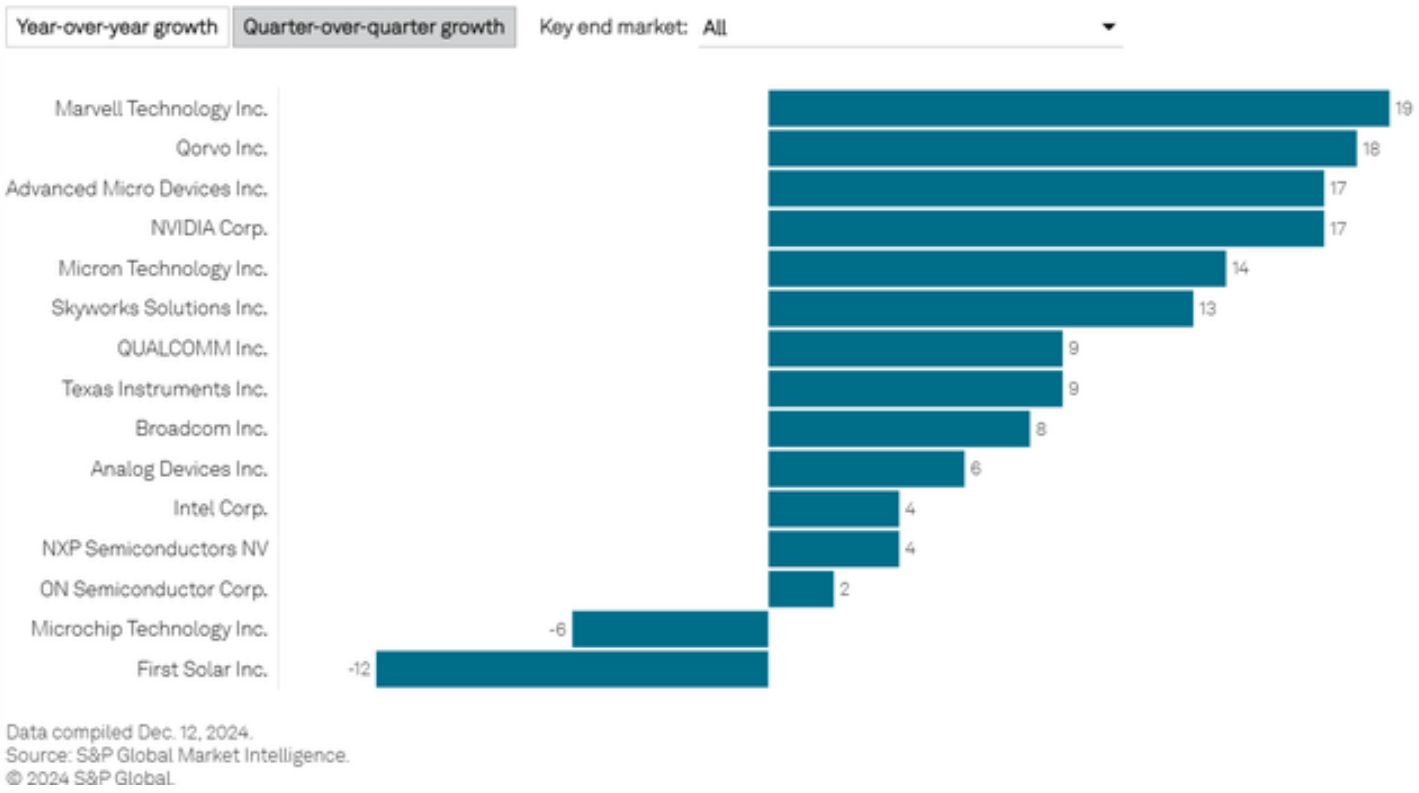
"We are in the midst of seeing that transition, which may take a few years for that to happen," Tan added.

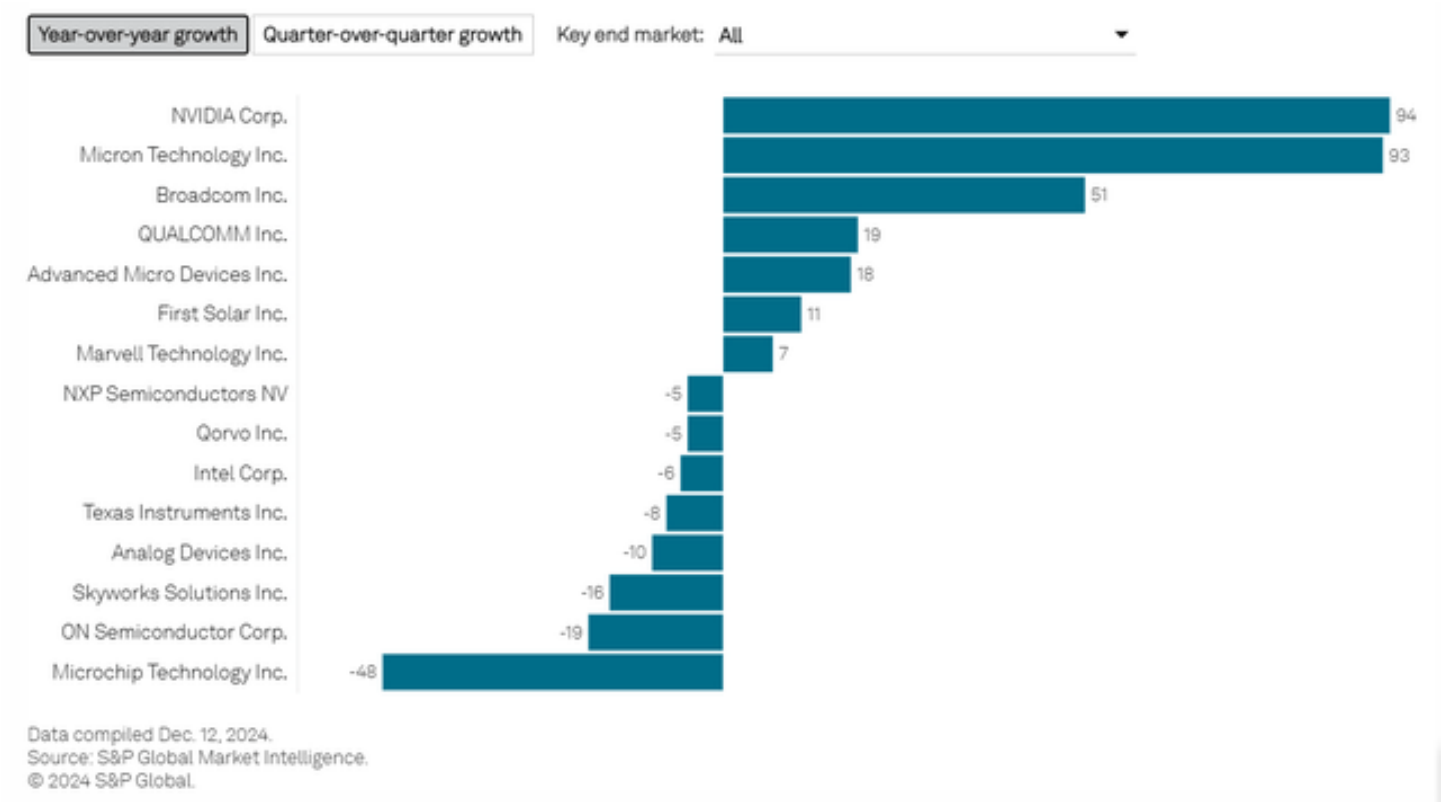
Most recently, The Information reported that Apple Inc. (AAPL) is working with Broadcom to develop its first server chip specifically designed for AI. The AI chip, internally known as Baltra, is expected to be ready for mass production by 2026.

On its latest earnings call, Broadcom said its "serviceable addressable market" in AI, which includes custom accelerator chips and networking chips, could grow to between $60 billion and $90 billion in fiscal 2027. Broadcom's AI revenue stood at $12.2 billion in fiscal 2024, up 220% versus fiscal 2023.

Looking at 15 key companies from the S&P 500 Semiconductors Index that represent various end markets, only two recorded quarter-over-quarter revenue declines, compared with four in the second quarter.

Revenue growth for key semiconductor companies, Q3 (%)



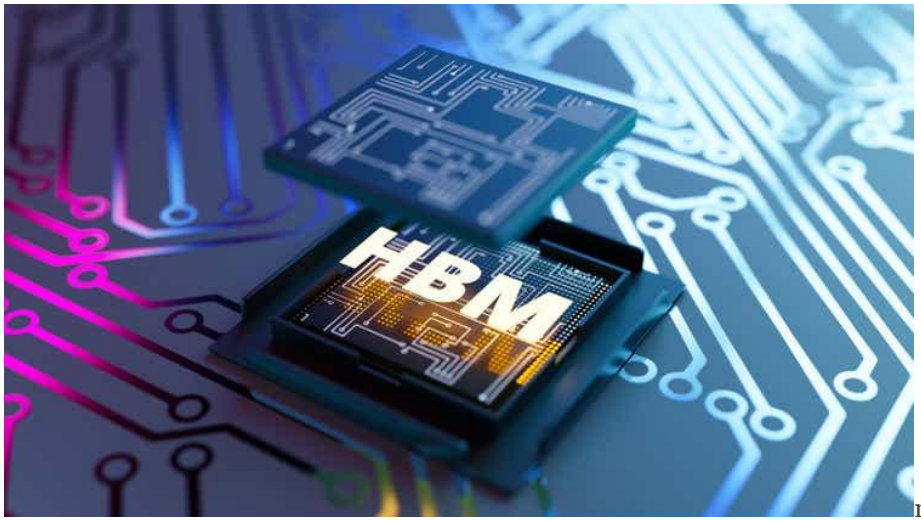


HBM Demand Is Surging Past Supply: Why Micron Is The Biggest Beneficiary

Dec 03, 2024 Hataf Capital

Summary

- High-bandwidth memory (HBM) demand is skyrocketing, projected to grow from $4B in 2023 to over $25B by 2025, positioning Micron as a critical AI enabler.
- FY2024 revenue rose 62% YoY, driven by AI-driven demand, with expanding margins and sold-out HBM production through 2025.
- Reduced industry capacity and AI-driven demand are stabilizing a traditionally cyclical market, creating sustainable growth opportunities.
- Micron's deep valuation discount and projected 51% EPS growth make it a compelling AI infrastructure investment.



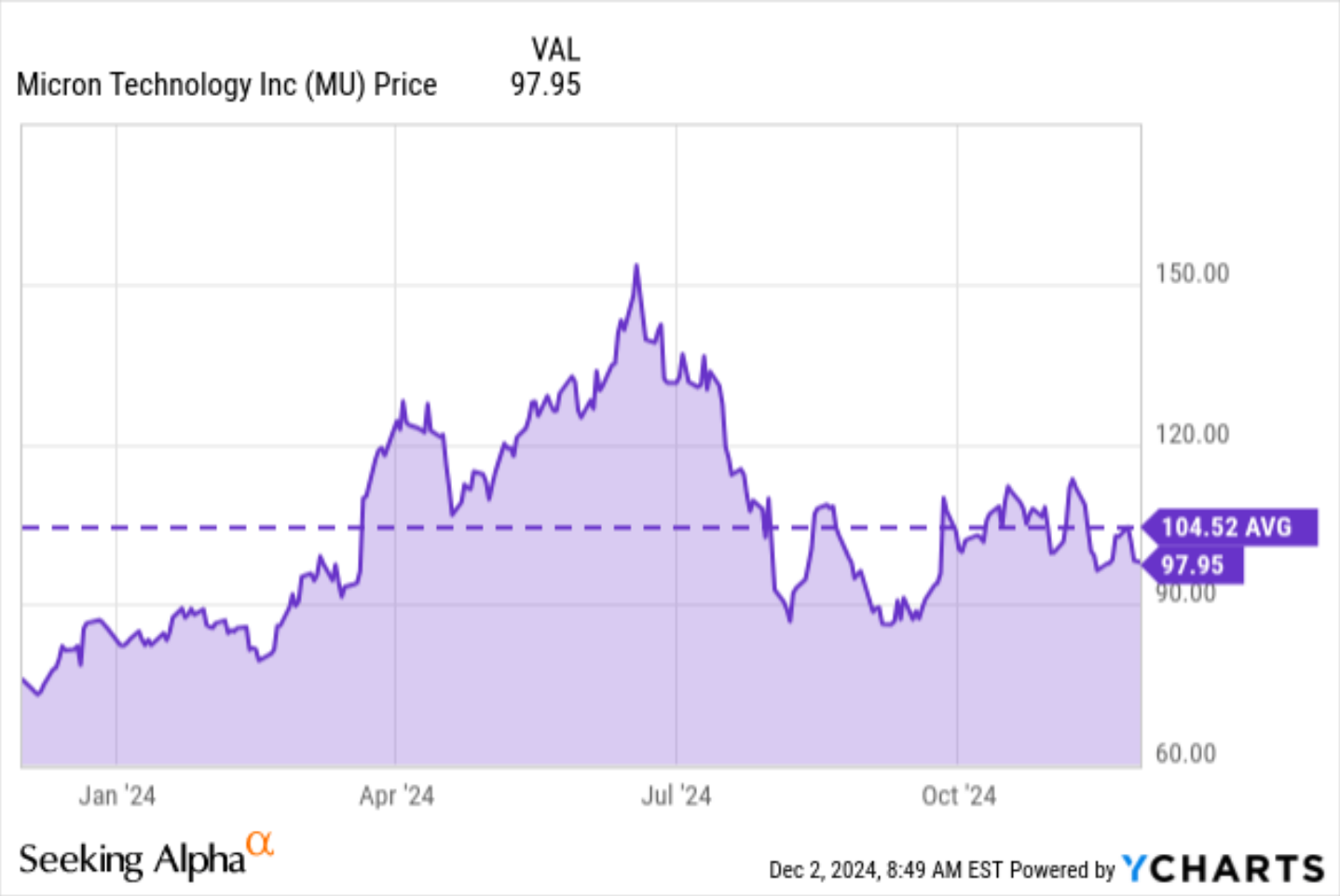
mesh cube

Last week during my research of the AI infrastructure supply chain. While everyone's been obsessing over GPUs and AI chips, I've found what, I believe, is the real unsung hero of the AI revolution: high-bandwidth memory (HBM). I was surprised too when I first realized just how critical this component is.

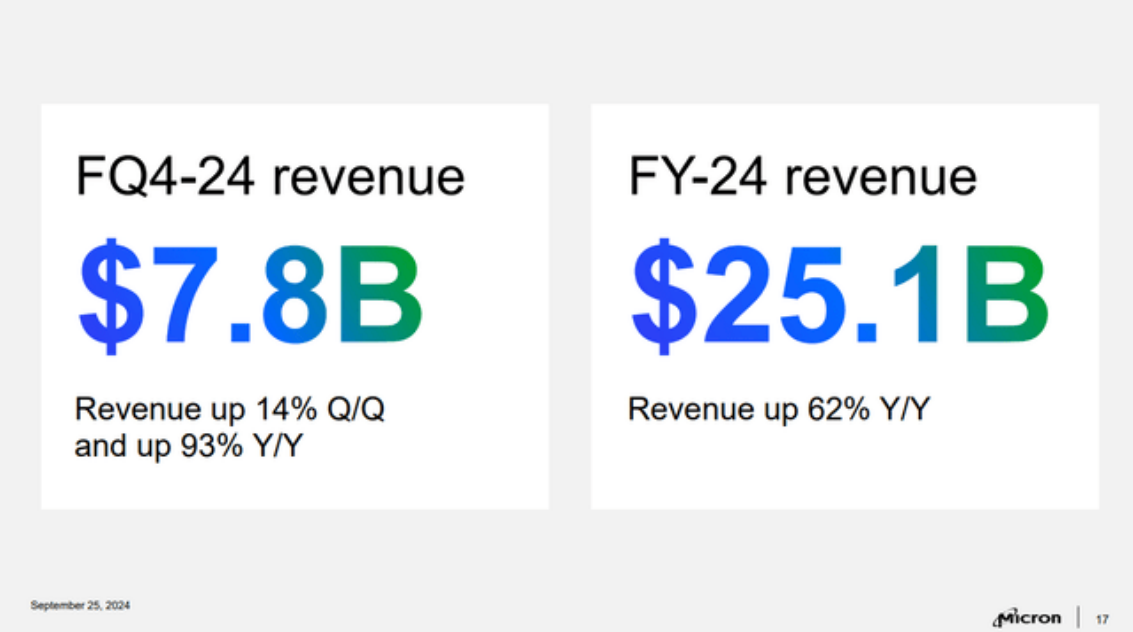
Think about it this way, it doesn't matter how powerful your GPU is if you can't feed it data fast enough. It's like having a Formula 1 car stuck in traffic. During my research, I kept coming back to this fundamental: the speed at which data moves between processors and memory.

This is where Micron (NASDAQ:MU) caught my eye. While everyone's paying premium prices for obvious AI plays like NVIDIA (NVDA) and AMD (AMD), Micron's trading at a fraction of those valuations despite its critical role in the AI supply chain.

Here's what I think most investors are missing: the memory market is undergoing a fundamental change due to AI. These workloads need specialized high-performance memory that only a few companies can make. Let me walk you through exactly why I believe Micron represents such a unique opportunity due to its impressive growth profile and demand for its HBM technology.

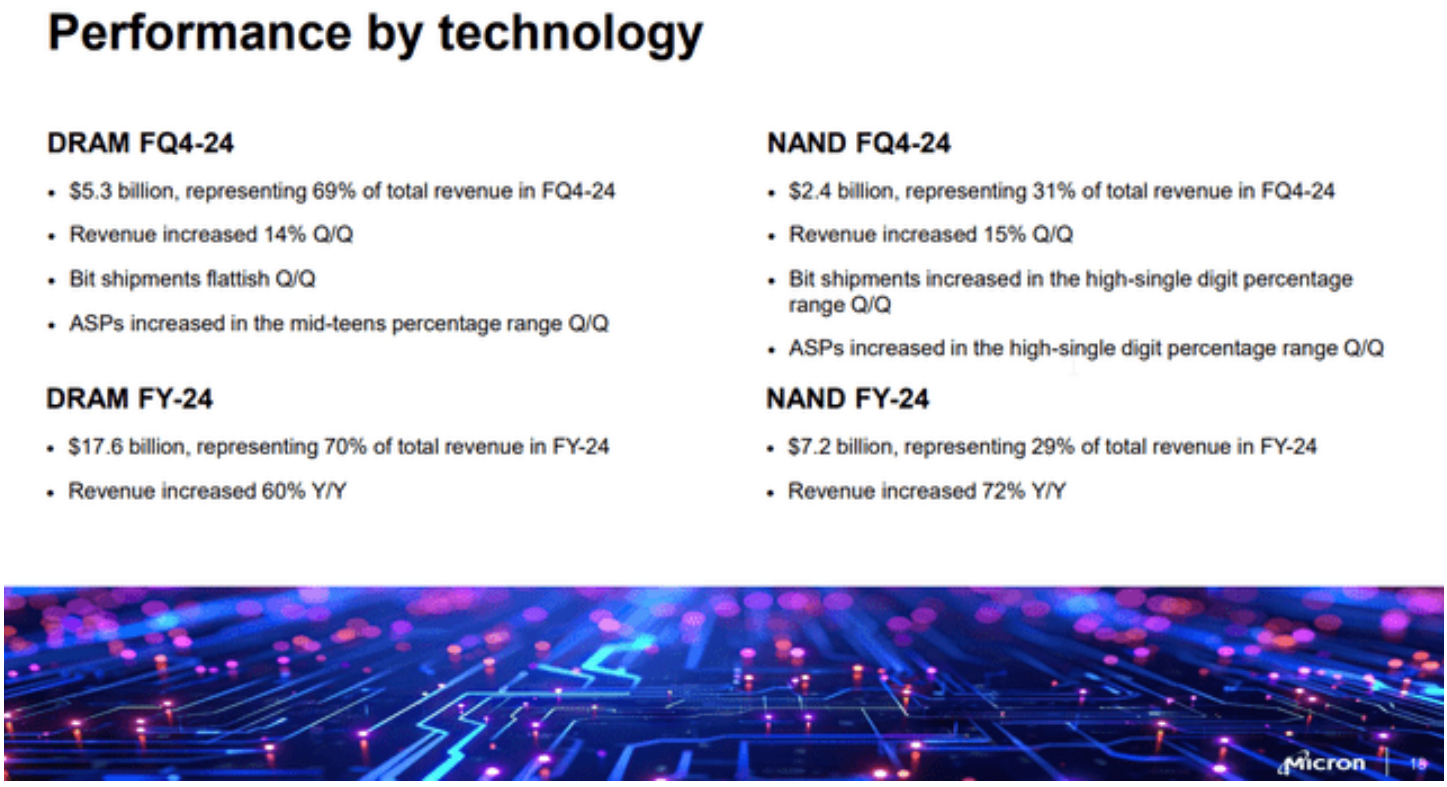


In FY2024 Micron delivered remarkable financial results with revenue growing 62% YoY to $25.1 billion, while expanding gross margins to 23.7% by 30 percentage points. This strong performance is just the beginning of what, I believe, will be a transformative growth cycle driven by AI infrastructure demands.



Micron Q4 Earning Deck

The company's technological leadership is evident in their successful ramp of 1-beta DRAM and G8/G9 NAND nodes, which will become an increasing portion of their mix through fiscal 2025. Their 1-gamma DRAM pilot production using EUV lithography is progressing well, with volume production on track for 2025. These technological advancement positions them at the forefront of the memory industry just as AI infrastructure demand is accelerating.



Micron Q4 Earning Deck

What's particularly interesting is the structural change occurring in the memory market. Historically, memory has been a highly cyclical industry but now I see a more sustainable demand environment emerging. The combination of reduced industry wafer capacity (below 2022 peak levels) and increasing HBM mix is creating a healthier supply and demand balance. Moreover, Micron's data center revenue reached record levels in fiscal 2024 and is expected to grow significantly in 2025, with each of their three main data center product categories - HBM, high capacity D5/LP5 solutions and data center SSDs projected to deliver multiple billions in revenue.

The market appears to be pricing Micron based on historical memory industry dynamics, failing to recognize the structural changes brought about by AI infrastructure demands and the company's strengthened competitive position. As these factors become more apparent in coming quarters, I expect a significant re-rating of the stock to better reflect its growth potential and strategic importance in the AI revolution.

The AI Memory Opportunity

Memory, particularly high-bandwidth memory (HBM) represents a critical bottleneck in AI infrastructure that's becoming increasingly important as AI workloads grow more complex and demanding.

According to Micron's management the HBM market is projected to expand from approximately $4 billion in calendar 2023 to over $25 billion in calendar 2025.

This growth trajectory is being driven by multiple vectors that are fundamentally reshaping AI memory demand: growing model sizes, increasing input token requirements, the emergence of multi-modality applications, multi-agent solutions, continuous training needs and the proliferation of inference workloads from cloud to the edge.

What makes this opportunity particularly compelling from an investment perspective is Micron's strong execution in this space. Their HBM3E 12-high 36GB solution demonstrates clear technological leadership, delivering 20% lower power consumption than competitors' HBM3E 8-high 24GB solutions while providing 50% higher DRAM capacity. This technical advantage is crucial for data center customers where power efficiency directly impacts operating costs and performance capabilities.

The financial impact of this shift is already becoming visible. In Q4 2024 Micron's HBM gross margins were accretive to both company and DRAM gross margins, even as overall DRAM margins improved. This margin expansion occurred during the early stages of HBM adoption, suggesting potential for further improvement as volumes scale up. The company's HBM production is sold out for both calendar 2024 and 2025.

The demand environment for AI memory extends beyond just HBM. Micron is seeing strong demand for their high capacity DDR5 and LPDDR5 solutions with increasing adoption of their high capacity mono-die-based 128GB DDR5 DIMM products. They're also pioneering the adoption of low-power DRAM for servers in the data center, offering unique features for enhanced reliability, availability and serviceability in server platforms. Their data center SSD revenue alone exceeded $1 billion in Q4 2024 with fiscal 2024 data center SSD revenues more than tripling YoY.

Looking ahead, I see this momentum accelerating. The company's strategic investments in manufacturing capacity, including new fabs in Idaho and New York, position them well to meet growing demand. The combination of technological leadership, strong execution and expanding market opportunity creates a compelling growth story that I believe is not fully reflected in current market valuations.

Market Opportunity Timeline

The artificial intelligence revolution is unfolding in distinct phases that create an expanding opportunity for Micron's memory solutions. In the current landscape, I see two critical phases that will drive sustained demand for high-performance memory products, particularly HBM.

The initial phase, spanning 2024 to 2030, centers on training large language models. This period is characterized by massive computational requirements for developing foundational AI models. During this phase, I expect to see extraordinary demand growth for High Bandwidth Memory, supported by Micron's recent projection that the HBM total addressable market will expand from $4 billion in 2023 to over $25 billion by 2025. This represents a compound annual growth rate of over 150% in just two years.

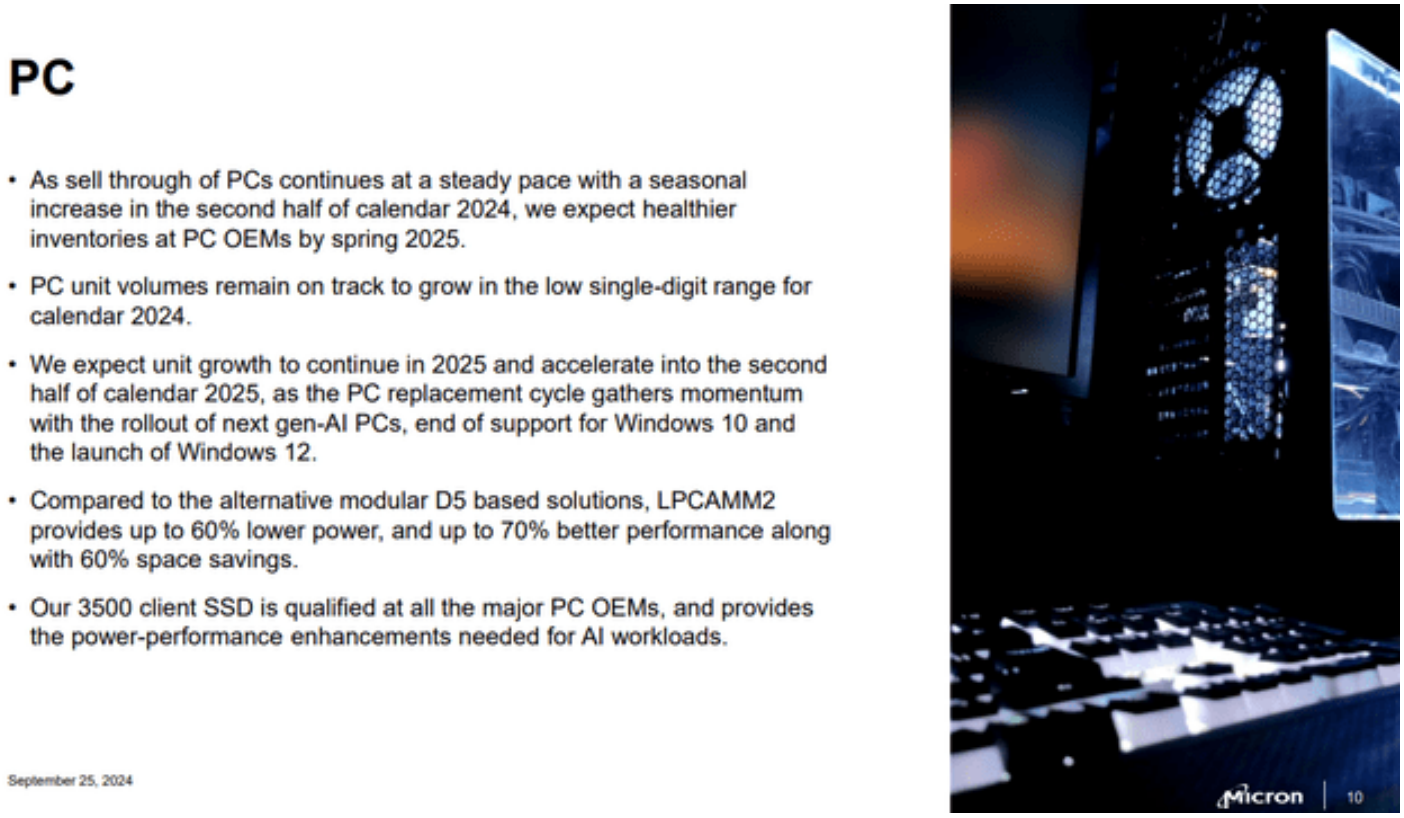
The next phase (2030-2035) will focus on reinforcement learning applications, where AI models are refined for specific tasks. This phase is particularly promising for Micron's advanced memory solutions, as reinforcement learning requires both high bandwidth for real-time processing and increased memory capacity for storing complex decision trees and model parameters. Micron is already positioning for this future with their HBM3E 12-high 36GB solution, which delivers 50% higher capacity than competing solutions while consuming 20% less power.

The financial implications of this evolution are substantial. Based on the company's latest guidance, I am expecting a forward revenue CAGR of 43.71% for the next three years supported by several key factors. First, Micron's data center revenue reached record levels in FY2024 with their data center SSD business alone generating over $1 billion in quarterly revenue. Second, the company's HBM products are sold out through 2025 providing excellent visibility into future growth.

Multiple Vectors Driving Micron's Growth

The convergence of AI infrastructure build out, data center expansion, margin improvements, and recovery in traditional markets creates a powerful growth narrative that is underappreciated by the market.

Beyond AI and data center markets I see substantial opportunity in the recovery of traditional PC markets. PC unit volumes expected to accelerate into the second half of FY2025. This acceleration will be driven by several factors: the rollout of next gen AI PCs, end of support for Windows 10 and the expected launch of Windows 12.



Micron Q4 Earning Deck

The AI PC opportunity is particularly interesting as these devices require more memory content than traditional PCs. Leading PC OEMs have announced AI enabled PCs with a minimum of 16GB of DRAM for the value segment and between 32GB to 64GB for the mid and premium segments, compared to an average of around 12GB last year.

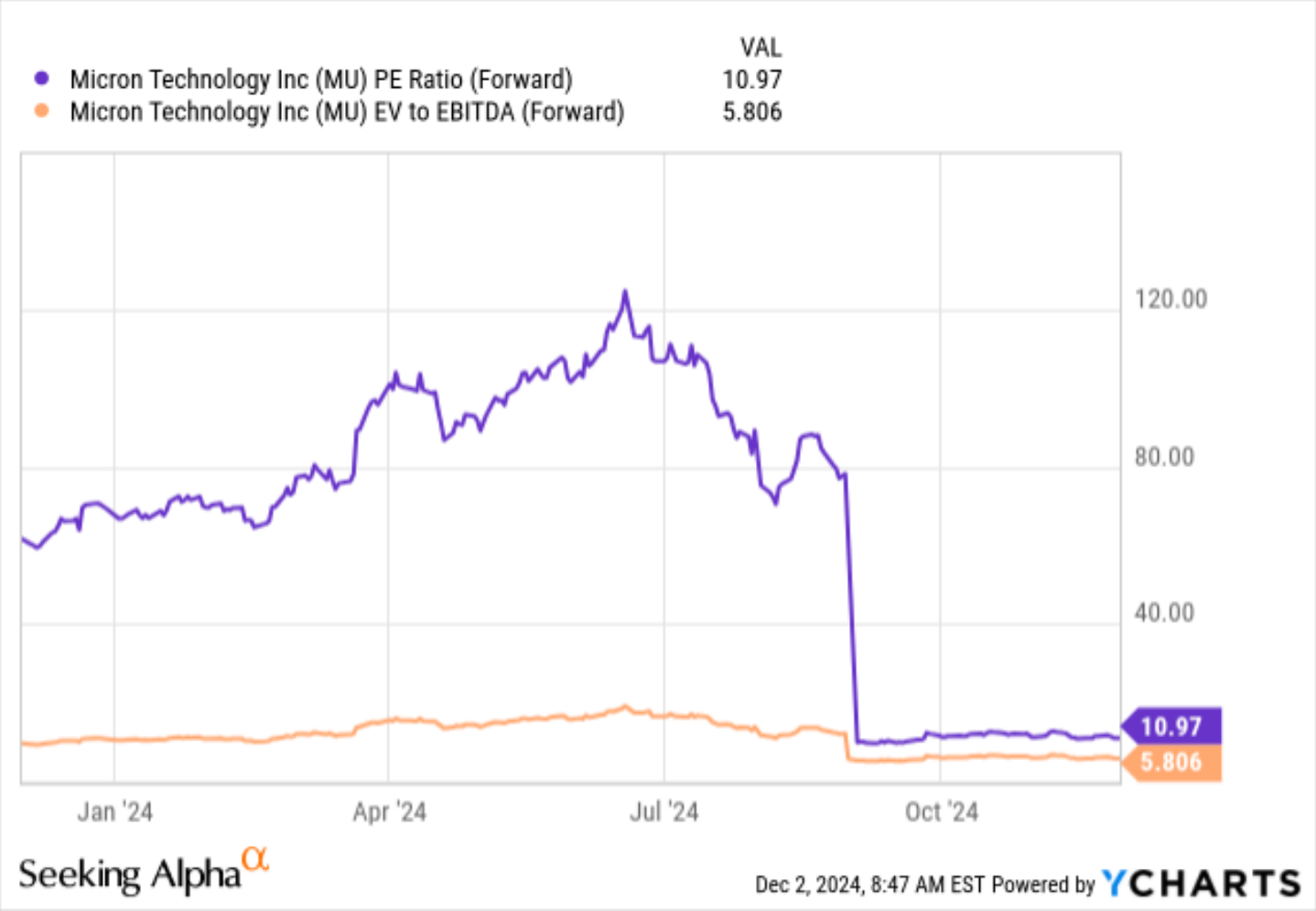
Industry supply dynamics further support these growth catalysts. Due to capital expenditure and supply reduction actions taken across the industry in 2023, industry wafer capacity in both DRAM and NAND in 2024 is expected to be below 2022 peak levels. The increasing mix of HBM wafers is reducing DRAM supply allocated to traditional products and contributing to a healthy industry supply and demand environment expected for calendar 2025. Given that HBM requires approximately three times more wafer capacity compared to traditional DRAM products this shift should support stronger pricing power and margins.

As these catalysts materialize over the coming quarters, I expect them to drive both revenue growth and margin expansion potentially leading to significant value appreciation for investors.

Financial Analysis and Valuation

Despite this strong performance and promising outlook Micron's valuation metrics suggest the market is undervaluing the company's prospects. The forward P/E ratio of 10.97x represents a 56.7% discount to the semiconductor sector median of 25.37x. The EV/EBITDA multiple tells a similar story. Micron's forward EV/EBITDA of 5.89x sits at a 62% discount to the sector median of 15.66x. This metric is particularly relevant as it accounts for different capital structures and provides a good basis for comparison across the semiconductor industry. I think this discount is mainly

because the market is still pricing Micron based on historical memory industry dynamics rather than its strengthening competitive position and growing exposure to secular growth trends in AI and data center markets.



However, the most compelling valuation metric to me is the forward PEG ratio of 0.21x compared to the sector median of 1.9x - an 88.7% discount. This is particularly notable because the PEG ratio factors in growth rates, essentially measuring how much investors are paying for growth. A PEG ratio below 1.0 typically suggests undervaluation and Micron's 0.21x is exceptionally low. For context, the company's long term EPS growth rate is projected at 51.37%, yet this strong growth outlook isn't reflected in the current valuation.

Looking ahead, management's guidance for Q1 2025, with revenue expected to reach $8.7 billion and Non-GAAP gross margins projected to expand to 39.5%. These forecasts combined with the company's deeply discounted valuation metrics create what, I believe, is an attractive entry point for long-term investors looking to participate in the AI infrastructure build-out at a reasonable valuation.

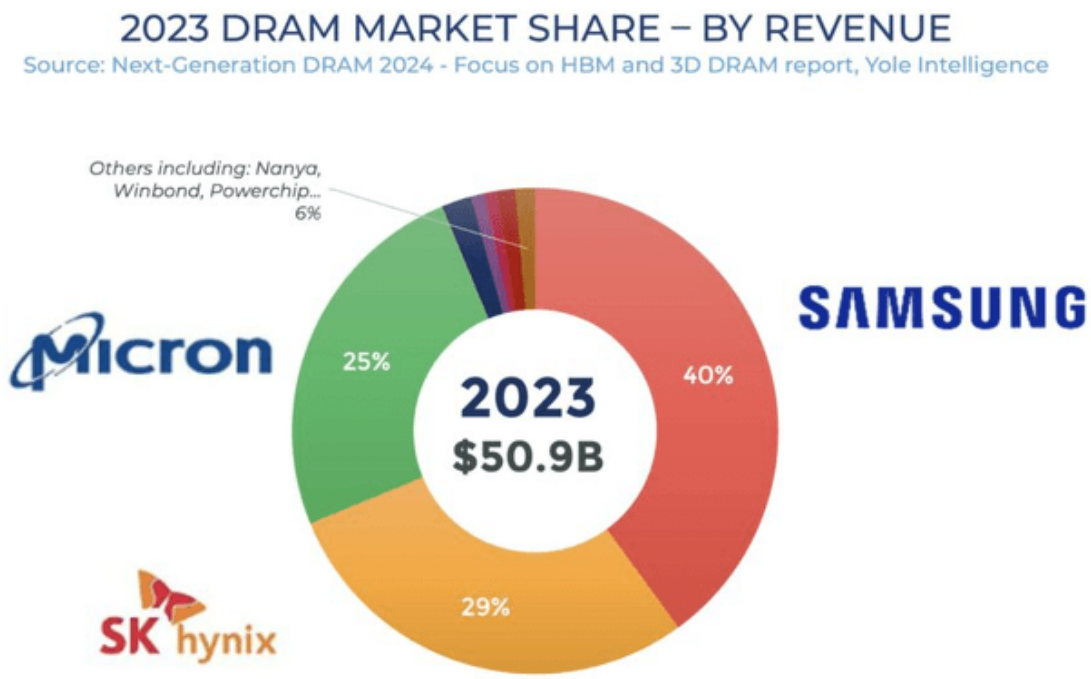
Risk Factors

The RAM chip market is dominated by three players Samsung Electronics, SK hynix and Micron, who together control almost the entire market. In the HBM segment, SK hynix currently dominates nearly 50% of the market, giving them scale advantages. Samsung's potential entry as the third supplier of AI memory chips for Nvidia could further intensify competition.

The capital intensive nature of memory manufacturing poses ongoing challenges. In fiscal 2024, Micron invested $8.1 billion in capital expenditures and management expects fiscal 2025 capex to be meaningfully higher at around the

mid-30s percentage range of revenue. While these investments are necessary to support future growth, they could pressure margins and free cash flow in the near term.

However, I believe these risks need to be weighed against several mitigating factors. The HBM market's projected growth from $4 billion in 2023 to over $25 billion in 2025 represents a significant expansion that should support multiple successful competitors. The company's financial strength provides some cushion against these risks. With $9.16 billion in cash and investments as of Q4 2024, Micron has substantial resources to weather industry cycles and fund necessary investments.



Yole Intellignece

To Sum Up:

Micron represents a rare opportunity to invest in a critical AI infrastructure component at a significant discount to intrinsic value. While the memory industry has historically been cyclical, I believe we're entering a new era where AI-driven demand will create a more sustainable growth trajectory. The combination of industry-leading technology, strategic manufacturing investments, and deeply discounted valuation makes Micron a compelling investment opportunity at current levels.

The company's strategic positioning in high-growth segments like HBM, combined with broad exposure to AI infrastructure build-out I see significant upside potential as the AI revolution continues to accelerate. For investors looking to participate in the AI boom at a reasonable valuation, I believe Micron offers one of the most attractive risk-reward propositions in the market today.

Broadcom Will Benefit From Taking Nvidia And AMD's Largest Customers (Rating Upgrade)

Dec. 13, 2024 The Value Portfolio

Summary

- Broadcom is nearing a $1 trillion valuation, driven by strong earnings growth and the acquisition of VMware, with a P/E ratio of ~30.
- The company's AI segment is booming, with 220% YoY growth and partnerships with three large customers, aiming for a $75 billion market opportunity by 2027.
- Broadcom's semiconductor segment saw 12% YoY growth, and infrastructure software revenue nearly tripled, indicating strong demand and future growth potential.
- Despite initial reservations about its valuation, Broadcom's innovative AI chips and customer shift from Nvidia position it for substantial future revenue and shareholder returns.

JHVEPhoto

Broadcom (NASDAQ:AVGO) is approaching the $1 trillion threshold after hours, making it potentially the first semiconductor company to do so. We last recommended against investing in the company here, discussing the threats faced by the company's acquisition focused strategy. However, as we'll see in this article, the company's ability to take the customers from existing dedicated GPU providers (NASDAQ:NVDA) (NASDAQ:AMD) could be the strength that justifies investing.

Broadcom Earnings

Broadcom announced strong YoY earnings growth primarily buoyed by the acquisition of VMware.

Fourth Quarter Fiscal Year 2024 Financial Highlights

	GAAP			Non-GAAP		
(Dollars in millions, except per share data)	Q4 24	Q4 23	Change	Q4 24	Q4 23	Change
Net revenue	$ 14,054	$ 9,295	+51 %	$ 14,054	$ 9,295	+51 %
Net income	$ 4,324	$ 3,524	+$ 800	$ 6,965	$ 4,810	+$ 2,155
Earnings per common share - diluted *	$ 0.90	$ 0.83	+$ 0.07	$ 1.42	$ 1.11	+$ 0.31

(Dollars in millions)	Q4 24	Q4 23	Change
Cash flow from operations	$ 5,604	$ 4,828	+$ 776
Adjusted EBITDA	$ 9,089	$ 6,048	+$ 3,041
Free cash flow	$ 5,482	$ 4,723	+$ 759

Boradcom press release

The company announced a 51% increase in revenue and a strong increase in non-GAAP net income. Earnings per share came in at roughly $6 annualized, with substantial YoY growth, putting the company at a P/E of ~30. It's a lofty-er valuation that indicates that the company will need continued growth to justify its valuation.

The company's annualized FCF came in at $22 billion, with adjusted EBITDA at more than $36 billion. The company's FCF yield, at roughly 2.5%, is also a yield that requires continued growth. For 2024 overall, the company saw more than $12 billion in AI revenue with 220% YoY growth, and we expect the company's portfolio here to give it strong growth.

Broadcom Segment Performance

Broadcom has continued to perform well segment by segment.

Net revenue by segment

(Dollars in millions)	Q4 24		Q4 23		Change
Semiconductor solutions	$ 8,230	59 %	$ 7,326	79 %	+12 %
Infrastructure software	5,824	41	1,969	21	+196 %
Total net revenue	$ 14,054	100 %	$ 9,295	100 %	

Broadcom press release

The company's semiconductor segment saw 12% YoY revenue growth, driven by the continued strength of artificial intelligence. Based on the company's earnings, artificial intelligence makes up 40% of the company's semiconductor revenue and has continued its substantial growth. Despite that, the acquisition of VMware means semiconductor solutions are now 59% of the company.

The company's infrastructure software revenue almost tripled YoY to almost $6 billion per quarter.

The company's 1Q 2025 guidance is for $14.6 billion in revenue with an adjusted EBITDA margin of 66% ($9.6 billion adjusted EBITDA). That would imply 7% QoQ adjusted EBITDA growth, showing continued demand for the company's products and strong growth. That will enable the company to grow into tis valuation.

Broadcom Artificial Intelligence

Here's the key segment from the earnings discussion on artificial intelligence.

"We see an opportunity over the next three years in AI," Tan told investors on the earnings call. "Massive specific hyperscalers have begun their respective journeys to develop their own custom AI accelerators."

Tan said Broadcom is currently developing AI chips with three very large customers, and he expects each of them to deploy 1 million AI chips in networked clusters by 2027. Tan said the total market opportunity for its AI chips, which it calls XPUs, as well as parts for AI networking could be between $60 billion and $90 billion by 2027.

- CNBC

We've discussed in our prior articles on Nvidia, the issue of Nvidia's massive customer concentration. For Nvidia, the company is reliant on massive sales to a handful of companies, and it needs those sales to continue increasing. Forecasts for 2025 Blackwell sales are 800K units for the first quarter ramping up to a total of ~5 million units for the year.

While the exact sales remain to be seen, almost half of the company's revenue comes from just 4 customers. So when Broadcom says they're currently developing AI chips with "three very large customers", initial data indicating Apple is one of them, there's very likely to be some overlap between these 3 large customers and Nvidia's largest customers.

The same is true for AMD given that the largest customers are the same sets of companies. Broadcom is saying, likely with strong internal data, that he expects each of these customers to deploy 1 million AI chips in networked clusters by 2027 with a total market opportunity of ~$75 billion by 2027. For perspective, the largest AI cluster in the world today is "Colossus" owned by xAI, with 100K Nvidia GPUs.

So in just a few years, these customers, working with Broadcom, will be developing clusters 10x the size of the largest today. That's not surprising to us. These "large customers" are massive, with deep pocket books, like in the realm of companies like Google, Amazon, Microsoft, Apple etc. They also already have a lot of their own tech prowess.

With Blackwell margins in the low-70s, and a cost per GPU of ~$35K, the difference between a 1 million equivalent GPU cluster at cost for Microsoft versus 1 million GPUs from Nvidia is $25 billion. That's massive savings, and it's no wonder that these companies are investing billions with Broadcom as a result.

It's an opportunity worth potentially $10s of billions to Broadcom. And it could cost Nvidia massively.

Why Nvidia? Why AMD?

Interested investors might be wondering Why Nvidia? Why AMD? Why won't the XPUs be independent. We highlight 3 reasons here:

1. Companies are becoming more thoughtful about GPU purchases.

Executives at Google and Meta have already stated they might be spending too much on GPUs, but they're continuing to do so because they're afraid of being left behind. One way to continue having the compute power you might require, without overspending, is to allocate your budget to XPUs instead of GPUs.

2. Large scale XPUs will decrease the moat of high-margin GPUs.

Nvidia GPUs are currently the best way to get hyper-scale GPU compute at volume with existing SW support. CUDA, Nvidia's SW stack, is considered a huge part of the company's moat. However, our view is that as we get 1+ million XPU systems, that moat will chip away, companies will need to support their own XPUs.

Given that many of these companies (Google Cloud) etc. offer their HW outwards, that could have a universal affect.

3. They have the present customers.

And then there's the last reason, and the most obvious. Nvidia and AMD have the largest source of customer's today. Broadcom can't take Intel's (INTC) lunch because Intel has no lunch today. Nvidia is making the lion's share of GPU revenue and that means that if there are changes in the market, they stand to lose the most.

Our View

We were initially against Broadcom given its lofty valuation. However, the company's packaging tech with XPUs is a new segment in the market, and Nvidia's massive margins are causing its largest customers to turn away and look at the alternatives. With millions of XPUs in demand, Broadcom is well positioned to help customers build their own silicon, as seen with its growth.

That could enable the company to earn billions, if not $10s of billions in reasonable revenue as customers move away from Nvidia. It can still have reasonably high margins and enable the company to grow into its valuation. That will enable Broadcom to generate strong shareholder returns.

Thesis Risk

The largest risk to our thesis is the company's largest present day customer, Apple (NASDAQ: AAPL). Rumors are that Apple is building its own chip to transition away from Broadcom, and Apple is Broadcom's largest customer, accounting for ~20% of FY'23 revenue. A shift away could hurt Broadcom's continued profits.

Conclusion

Broadcom is a massive company and might soon cross the $1 trillion valuation threshold. The company grew by double-digits after hours, with a strong earnings announcement. The company is seeing strength with an integrated VMware, and the company is seeing massive demand for its XPUs as Nvidia's largest customers build their own GPUs.

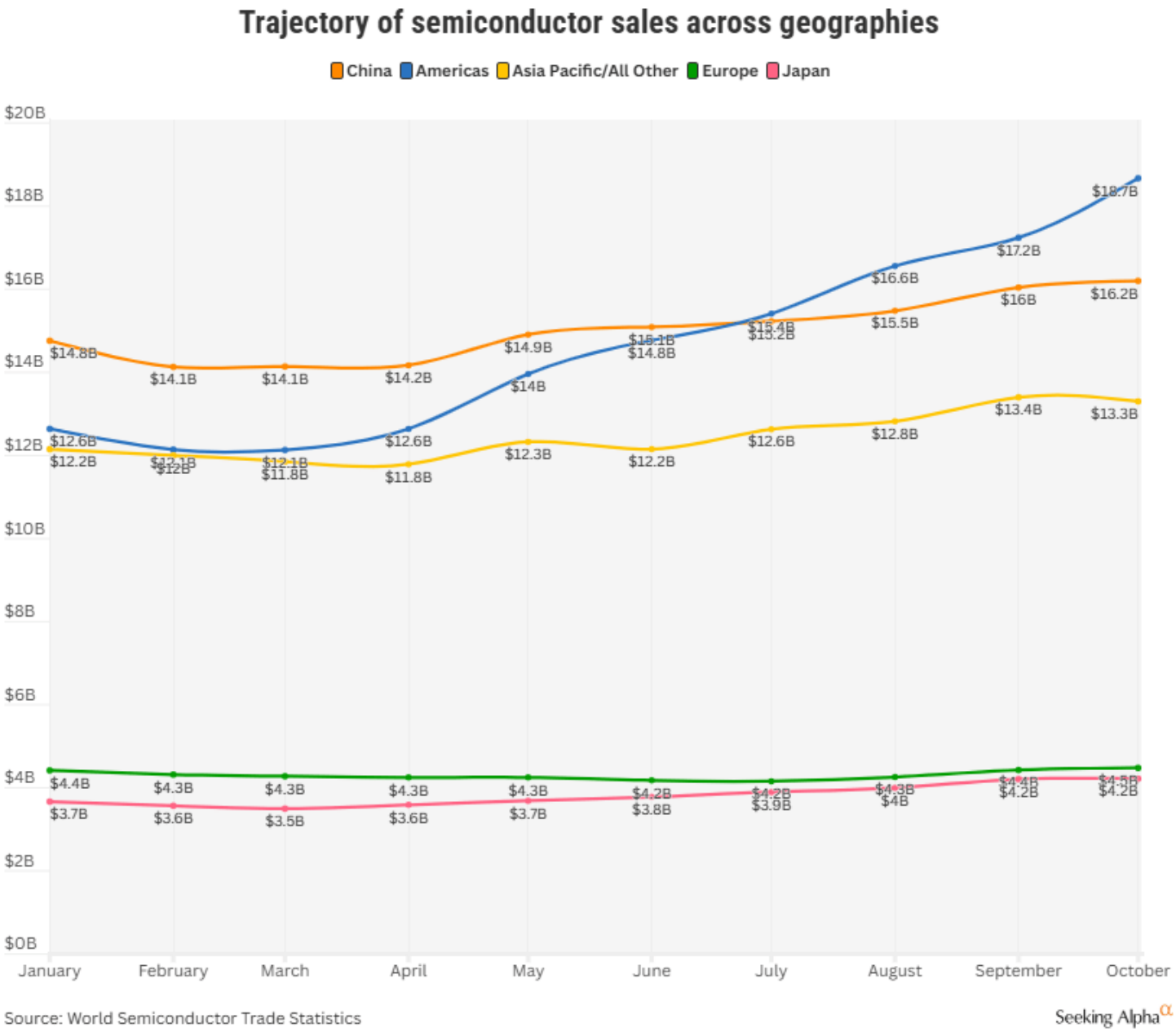
That could result in customers moving away from Nvidia, however, Nvidia's losses are Broadcom's gains. The company will be able to grow its FCF and earnings to justify its valuation, resulting in a ratings upgrade from our side. We now see Broadcom as a valuable investment proposition. Let us know your thoughts in the comments below.

SA Charts: Here's how global semiconductor sales have performed this year

Dec 16, 2024 Jaskiran Singh, SA

Semiconductor sales in October reached their highest-ever monthly total, posting a sales figure of $56.9B. This was an increase of 22.1% over the same period last year and 2.8% over September 2024.

"Total annual sales are now projected to increase by nearly 20% in 2024—higher than earlier forecasts—and then continue to grow by double-digits in 2025," said John Neuffer, Semiconductor Industry Association president and CEO.

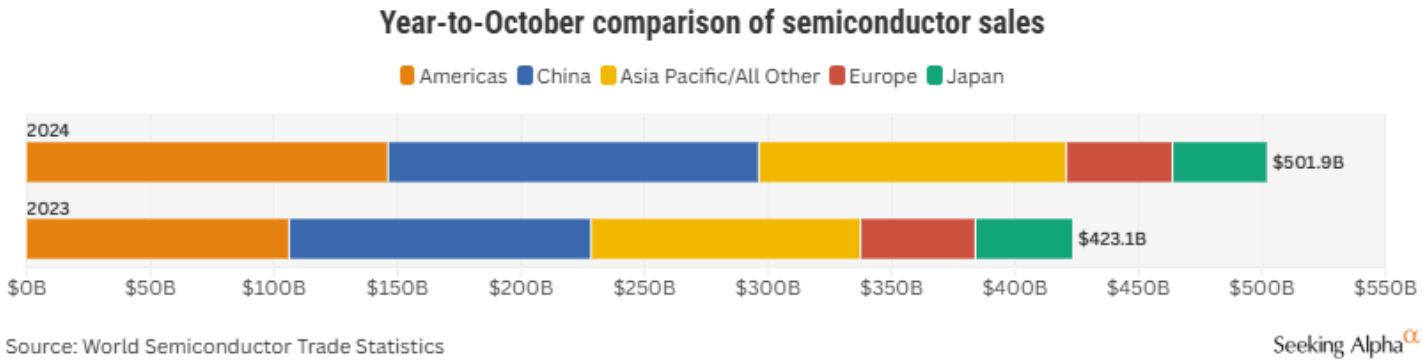


Taking the biggest chunk of semiconductor sales, the Americas further increased its monthly tally, recording sales of about $18.7B in October, a sequential increase of 8%, the highest among all regions.

Starting below China at $12.6B in semiconductor sales in January, the Americas has upped its monthly intake since April, overtaking China by June and comfortably sitting at a $2.5B lead over China in October.

Part of this has been due to the governmental support to chipmakers under the CHIPS and Science Act. Recently, the U.S. Department of Commerce is said to have finalized a $6.165B award to Micron Technology (MU) to support the construction of two fabs in Clay, New York, and one fab in Boise, Idaho. Before this, the CHIPS Act also granted Intel (NASDAQ:INTC) up to $7.86B in direct funding.

Regions EMEA, Japan, Japan, and China saw their tallies inch higher by about 1% over September. On the other hand, the Asia Pacific region saw a minor sequential dip in monthly sales at $13.30B.



Since the start of the year, Americas has secured about $146.2B in semiconductor sales, an increase of about 37.7% over last year. In October alone, the Americas recorded a 54% year-on-year jump in semiconductor sales.

On the other hand, China has recorded about $150.2B in the 10 months of 2024, however, compared to last year, sales have increased about 22.8%.

Sales in Asia-Pacific too have seen a 14.1% year-on-year increase to $124.4B. While semiconductor sales in Japan have remained somewhat stable, the figure came down for EMEA, recording sales of $43.03B, down 7.6%.