Global AI Convergence Reassessing Industry Trends and Investment Logic

Editor's Note:

Since April 2025, global technology stocks have experienced a broad-based rally, with AI-related companies across the A-share, Hong Kong, U.S., Japanese, and Korean markets rising in tandem.

What makes this rally distinctive is that it is no longer simply a concept-driven trade, nor is it an isolated rally confined to any single country. For the first time, major global equity markets have converged around the same technological theme, with industry capital expenditure, earnings growth, and equity-market valuation expansion reinforcing one another in a broad-based cycle of convergence.

However, in July this year, the sector underwent a temporary correction amid factors including market rumors surrounding Meta's business plans and deleveraging by leveraged investors. Concerns over the sustainability of capital expenditure and potential oversupply of computing capacity also increased.

Against this backdrop of heightened volatility, has the underlying industry fundamentally changed? Does the structural logic behind this global convergence remain intact?

At a recent client seminar, Xu Binyi, a technology-sector researcher at Rosefinch Fund, systematically reviewed the industrial logic, market dynamics, and investment opportunities across different segments of the current AI cycle. The following is an edited version of the discussion.

1. Four Historical Technology Bull Markets

Among previous technology-driven bull markets, one of the earliest representative episodes was the Internet bubble. In the U.S., leading beneficiaries included Cisco, AOL, and various dot-com companies, with network equipment and Internet portals at the center of the rally.

In Japan, SoftBank and NTT were among the major beneficiaries, with investment companies and telecommunications operators leading the market. In Korea, the rally was concentrated in telecommunications stocks and the KOSDAQ market. In mainland China, computer and broadband penetration remained relatively low, and dial-up connections were still the dominant form of Internet access.

Although global equity markets rose simultaneously during this period, there was no genuine global industrial convergence. The core of the global supply chain was concentrated in the U.S., while Chinese companies remained highly dependent on overseas hardware and equipment. China experienced a concept-driven market rally, while overseas markets were able to capture more of the underlying industrial benefits.

The next major wave was the smartphone revolution from 2009 to 2013. In the U.S., Apple and Qualcomm were among the core beneficiaries, and the broader mobile ecosystem experienced a significant rally.

In Japan, companies such as Murata Manufacturing benefited from the miniaturization of components. MLCCs, passive components, and upstream materials experienced sustained growth, with Murata's share price rising substantially during the period.

In Korea, Samsung was the key beneficiary, riding the Android ecosystem to generate a major market rally. In mainland China, the first generation of Apple supply-chain companies emerged, particularly in acoustic and optical components.

This cycle represented a stronger link between equity markets and the underlying industrial supply chain, although China's share of the global industry value chain remained relatively limited at the time.

From 2014 to 2015, mainland China experienced an independent "Internet Plus" rally. The market initially took off through brokerage stocks, followed by a broad-based rally in the first half of 2015, with Internet companies among the strongest performers.

In the U.S., the main theme remained the continuation of the platform Internet cycle. Japan's equity market rose under Abenomics, but its leading sectors were largely disconnected from those in China and the U.S. Korea's market declined overall and showed little correlation with the other three markets.

In other words, although several countries experienced simultaneous equity-market gains, there was no meaningful convergence in the underlying real economy. The major investment themes in each market remained largely independent.

The 2016-2018 semiconductor cycle was centered on data-center semiconductors. In 2017, the memory sector experienced a particularly strong rally.

Japan's leading beneficiaries included semiconductor equipment and upstream materials companies, while Korea's rally was concentrated in memory chips. China's A-share semiconductor sector also benefited from the global cycle, but the STAR Market had not yet been established, leaving a limited number of investable semiconductor companies.

This was therefore a period when the U.S., Japan, and Korea experienced a relatively synchronized semiconductor upcycle, while China's industry benefited from the same trend but lacked sufficient listed companies through which investors could fully participate.

From 2019 to 2021, semiconductors rebounded strongly after the 2018 correction. However, the underlying driver was arguably more related to global supply-chain disruptions caused by the COVID-19 pandemic than to a powerful structural industry cycle.

Email service@rosefinch.cn

The Philadelphia Semiconductor Index (SOX) rose approximately threefold during the period, with cloud computing and electric vehicles providing additional momentum later in the cycle.

Japanese semiconductor equipment and materials companies continued to rise, with earnings per share increasing alongside share prices. Carry-trade activity also contributed to the rally. In Korea, memory chips remained in a strong cycle, while battery companies benefited from the new-energy vehicle theme and saw substantial share-price gains.

In mainland China, two major developments became important.

First, following the implementation of overseas sanctions in May 2019, the consumer electronics sector bottomed out, driving a broad re-rating of the technology sector. Valuations of consumer electronics companies subsequently moved into the 20-30x range.

Second, the launch of the STAR Market, together with the intensification of trade tensions, made supply-chain security a major policy and investment theme. This created an important opportunity for the localization of semiconductor production.

Following the initial consumer electronics rally, domestic semiconductor substitution and new-energy sectors became the next major themes. China's technology sector experienced strong performance across multiple areas, but the drivers remained relatively fragmented. In semiconductors, the core investment thesis was primarily domestic substitution rather than a full-fledged global industry convergence.

From 2021 to 2022, global technology markets experienced a deep correction, bottoming in October 2022. The first major rebound, however, had little to do with AI and was driven primarily by overseas macroeconomic factors.

It was not until the launch of ChatGPT in November that the AI investment cycle truly began. GPUs and semiconductor stocks surged, with the SOX Index eventually rising from around 2,000 to 6,000.

In this cycle, U.S. model developers and cloud service providers benefited simultaneously. Japan's semiconductor equipment and materials companies continued to perform strongly, while Korean memory manufacturers and semiconductor equipment companies also rallied.

China initially lagged. From the launch of ChatGPT through September 2024, China's semiconductor index delivered relatively modest gains. The industry cycle had started, but there was still insufficient synchronization between the industry and equity markets.

The rally that began in 2025, particularly from April onward, represents a much more genuine case of global industrial convergence.

In the U.S., GPU, semiconductor, cloud, connectivity-chip, and optical communications companies all rallied together. In Korea, memory chips, semiconductor equipment, and AI supply-chain companies rose sharply. Japan saw continued strength in memory, semiconductor equipment, and materials companies.

Meanwhile, Chinese companies leveraged their strong manufacturing and service capabilities to enter the global AI supply chain. Optical modules, PCBs, domestic large language model companies, computing servers, and energy infrastructure-related companies all participated in the rally.

This is arguably the first time in history that China, the U.S., Korea, and Japan have participated simultaneously in a market cycle centered on the same technology, with industrial capital expenditure, earnings expansion, and equity-market valuation expansion reinforcing one another.

2. Two Core Investment Themes in AI

The current market can broadly be divided into two major investment themes.

The first is the pricing-driven theme, including memory, MLCCs, upstream materials, and various components. Returns in these areas are partly driven by sustained increases in product prices.

The second is the technology-innovation theme, whose underlying framework can be traced to the architecture presented at NVIDIA's GTC conference. This framework can broadly be divided into three stages: Scale Up, Scale Out, and Scale Across. Current market attention is primarily focused on Scale Up and Scale Out, although the two themes overlap in several areas.

Scale Up primarily includes GPUs and various XPUs, particularly application-specific integrated circuits (ASICs). Two key supporting areas are advanced process technologies and advanced packaging.

In memory, the initial driver has been HBM, with demand subsequently spreading to conventional DRAM and NAND flash, as well as hard disk drives.

At the board-level interconnect layer, PCBs are a core medium for chip-to-chip connectivity, alongside optical modules and technologies such as CPO and NPO.

Scale Out, which refers to interconnection between racks, is driving higher demand for high-frequency, high-speed PCBs and optical modules. Longer-distance interconnection corresponds to the Scale Across stage, where the supply chain includes optical-interconnect companies, PCB manufacturers, and complete switch-system manufacturers.

Overall, we believe the risk-reward profile of the technology-innovation theme is relatively more attractive than that of the pricing-driven theme.

Pricing-driven segments are subject to long-term contracts and cyclical factors, as reflected in the recent correction in memory stocks. By contrast, technology-innovation segments have clearer structural growth drivers

[021]2030 5888 [021]2030 5999
58 Floor, New Bund Center, NO.555 West Haiyang Road/588 Dongyu Road, Pudong New Area, Shanghai 上海市浦东新区海阳西路 555 号/东育路 588 号前滩中心第 58 层，邮编：200126

and potentially much larger long-term addressable markets.

As long as valuations remain within a reasonable range, technology-innovation segments may therefore represent a more stable long-term allocation.

3. Key Market Concern: Can Cloud Capex Continue to Rise?

Against the backdrop of the current market correction, investors have become increasingly focused on whether cloud service providers can continue raising capital expenditure.

Capital expenditure by cloud providers is a major source of revenue for the AI supply chain. Any downward revision or failure to meet growth expectations could therefore have a significant impact on the broader industry.

The key catalyst for the rally since April 2025 was the upward revision of capital expenditure guidance by major cloud providers. Sector performance has consequently been highly correlated with expectations for capex growth.

However, after repeated upward revisions, expectations may become increasingly stretched. Market sensitivity to optimistic guidance may decline, while any stabilization or downward revision in capex could trigger a rapid deterioration in sentiment.

Another concern relates to the operating cash flow of cloud providers. Some companies have already reached levels of capital expenditure above their operating cash flow, although the major players still appear capable of supporting their current investment levels.

Companies can also use equity financing, debt financing, and off-balance-sheet structures such as SPVs to maintain investment. However, if the returns on these investments fall short of expectations, capex growth could slow and valuations across the AI supply chain could come under pressure.

According to Bank of America estimates, combined capex by the five major cloud providers could approach $800 billion in 2026, with relatively strong growth potentially continuing into 2027. This suggests that near-term pressure on industry investment remains manageable.

There were two additional catalysts for the July market correction.

First, Meta announced plans to enter the cloud-computing business. The market initially interpreted this as a potential sign that Meta might reduce investment in its own models, implying an oversupply of computing capacity and a possible reduction in capex.

However, Meta has not yet begun renting out its computing capacity and continues to invest heavily in proprietary large language models. There is also little evidence of a meaningful oversupply of computing capacity, while rental prices for computing cards have continued to rise.

朱雀基金 Rosefinch Fund
[021]2030 5888 [021]2030 5999
58 Floor, New Bund Center, NO.555 West Haiyang Road/588 Dongyu Road, Pudong New Area, Shanghai 上海市浦东新区海阳西路 555 号/东育路 588 号前滩中心第 58 层，邮编：200126

Second, the correction was amplified by deleveraging in Korea's leveraged ETF market. By the end of June, leveraged positions in the Korean equity market had reached historically high levels and were concentrated in large-cap stocks. The subsequent market decline triggered substantial forced deleveraging.

4. Investment Logic Across the AI Hardware Supply Chain

4.1 GPU and ASIC: Evolution of the Computing-Chip Landscape

Over the next three years, GPUs are likely to remain the core of AI computing, with their dominant market position largely intact in the near term.

At the same time, customized ASICs are likely to gradually increase their share of the computing-chip market.

At the supply-chain level, GPUs and ASICs share many common components and production processes, including core substrate materials and ABF films.

ABF films are used in ABF substrates, and capacity remains tight. High-end computing chips can require substrates measuring approximately 10 cm × 10 cm, integrating two computing chips and six to eight HBM stacks. Substrates may contain more than 20 layers.

Because the thermal expansion coefficients of organic substrate materials and silicon differ, large substrates are more susceptible to warping and other manufacturing issues, resulting in a relatively tight supply-demand balance.

The overlap between GPU and ASIC supply chains creates some degree of competition on the supply side. However, global demand for computing chips continues to expand, and chip shipments are rising accordingly.

By 2028, the total market size of customized ASICs could potentially approach that of general-purpose GPUs.

Major cloud providers are continuing to develop proprietary chips. The fundamental motivation is supply-chain security. If computing-chip supply becomes overly concentrated in a single supplier, the bargaining power of downstream customers becomes significantly constrained, similar to the current situation in memory, where suppliers capture a relatively large share of industry profits.

The development of customized ASICs should therefore provide positive support for the broader AI hardware ecosystem.

As of June, rental prices for H-, A-, and B-series computing cards remained stable or continued to rise, with no meaningful decline. This further suggests that computing capacity has not yet entered a state of oversupply.

4.2 Agents Could Trigger a Surge in CPU Demand

Email service@rosefinch.cn

The investment logic behind the recent CPU rally is relatively straightforward: it is being driven primarily by the rise of AI agents.

Previously, AI applications were dominated by chatbots, which relied heavily on GPUs. With the emergence of agents, the importance of CPUs within the overall computing architecture has increased significantly.

For example, a human can only search a limited number of webpages simultaneously, whereas an AI agent can search hundreds or thousands of webpages in parallel. This dramatically increases thread requirements.

GPU resources are primarily used for code generation, while code execution and large-scale information retrieval rely more heavily on CPUs.

Within an AI computing cluster, CPUs can broadly be divided into three categories based on their functions:

Server CPUs

Scheduling CPUs

Workload CPUs

Server CPUs are paired with B200 computing boards, with one CPU typically serving two GPUs, creating a relatively fixed ratio.

Scheduling CPUs create isolated sandboxes and distribute task instructions. A single CPU can schedule a very large number of threads, but overall demand is relatively limited.

Workload CPUs handle webpage retrieval, code execution, and multithreaded parallel tasks. This is where incremental demand could be particularly strong.

A CPU may be responsible for opening and rendering 100 webpages. But actually understanding the textual content of those webpages requires a large language model, which relies on GPUs.

Previously, the system might have processed one article. Now it may need to process 100 simultaneously. GPU demand therefore increases roughly in line with the number of tasks, while demand for HBM, KV cache, and eventually conventional RAM behind those GPUs could be amplified further.

4.3 The Cyclical Nature of Memory May Gradually Weaken

Memory has been one of the strongest-performing technology sectors since last year.

The key question is whether the current memory upcycle can remain sustainable.

Spot prices are still rising, while server manufacturers generally procure memory under long-term contract pricing.

Even if contract prices place some constraints on price increases, tightness in the spot market could persist.

The fundamental reason is the continued displacement of conventional DRAM capacity by HBM.

Why does HBM place such pressure on overall memory capacity?

First, each HBM stack requires multiple DRAM dies, meaning that HBM production itself consumes a large amount of DRAM capacity.

Second, the planar density of an individual HBM die is lower than that of conventional DRAM because HBM requires through-silicon vias, or TSVs, to be drilled through the silicon. These structures consume significant amounts of usable wafer area and further reduce effective production capacity.

In other words, each wafer produces less effective memory capacity when allocated to HBM.

HBM has now progressed toward 12-layer, 16-layer, and even 20-layer configurations, suggesting that capacity constraints are unlikely to ease significantly in the short term.

On the demand side, even when manufacturers secure capacity through long-term agreements, actual available supply remains insufficient relative to demand.

HBM has become a bottleneck for the broader computing ecosystem because its manufacturing process is highly dependent on advanced packaging. High-density connections within a very limited physical space require interposers, placing significant demands on advanced-packaging capacity.

The question, therefore, is whether memory can remain in a strong cycle.

To answer this, we need to return to the cyclical nature of the memory industry itself.

Memory chips are among the largest commodity-like products in the electronics industry and exhibit some characteristics similar to futures markets.

A CPU purchased ten years ago may now be technologically obsolete. A hard drive purchased ten years ago, however, can still function today. This durability and standardization are among the reasons why memory has historically experienced such pronounced inventory and price cycles.

The commodity characteristics of memory create both inventory cycles and price cycles.

Memory is one of the largest individual semiconductor product categories. In 2025, the global semiconductor market was approximately $792.4 billion, with memory accounting for roughly $230 billion.

A more optimistic view is that long-term contracts can lock in prices and volumes, thereby smoothing the cycle. However, the market is still waiting to see how effectively this plays out in practice.

One factor that could structurally change the cyclical nature of memory is the increasing customization of HBM. As HBM becomes less standardized and more application-specific, its traditional cyclical characteristics may gradually weaken.

NAND also presents an interesting contrast with the 2017 cycle.

When NAND first experienced a major upcycle in 2017, the market was transitioning from hard disk drives to solid-state drives, creating significant cyclical fluctuations.

Today, the situation is different. HBM is consuming DRAM capacity, which in turn puts pressure on NAND capacity. Combined with the inherently consumptive nature of NAND storage, this creates structural support for a relatively tight supply-demand balance.

Data stored in NAND does not disappear automatically. Cloud providers generally do not actively delete stored data, while user-generated content such as videos continues to be written and retained.

Once data is written to NAND, the capacity is effectively "consumed." This creates substantial room for continued demand expansion.

This structural logic helps explain why NAND has also experienced a strong rally during the current cycle, broadly comparable to the strength seen in DRAM.

4.4 Advanced Packaging, PCBs, and Optical Communications: High Visibility of Demand

First, advanced-packaging capacity is likely to expand significantly in the coming years.

Both computing chips and memory chips increasingly depend on advanced packaging. At the same time, Moore's Law is slowing, prompting the Chinese semiconductor industry to pursue alternative paths for technological advancement, with advanced packaging being one of the key routes.

Leading overseas OSAT companies have significantly increased capital expenditure, in some cases to levels above their on-balance-sheet cash flow. Strong demand, however, has allowed downstream customers to support packaging companies' capacity expansion through advance payments.

In mainland China, packaging and testing companies have also announced significant increases in planned capital expenditure, suggesting that industry conditions are improving in parallel with global trends.

Second, PCB demand has relatively high visibility. Goldman Sachs estimates that the market could reach several hundred billion yuan by 2028.

The recent correction in PCB stocks was mainly driven by third-party reports suggesting delays and design changes involving new architectures at overseas semiconductor companies.

We believe the impact of any individual architecture delay or redesign on the overall PCB market should be limited. Demand from AI applications for high-density and high-layer-count PCBs remains intact.

These products require more sophisticated manufacturing processes and command higher unit prices, suggesting that the long-term value of PCB content per system could continue to increase.

In addition, substrate supply remains tight. The industry is exploring CoWoP as an alternative to traditional substrate solutions, potentially simplifying the architecture from a four-level structure to a three-level structure by moving directly from the chip and interposer to the PCB.

This could effectively turn the PCB into part of the packaging architecture and further increase the long-term value per square meter of PCB.

Third, the expansion of computing capacity directly drives demand for optical modules, with a nonlinear multiplier effect.

Each additional computing chip needs to establish connections with other chips in the cluster, meaning that interconnection requirements increase nonlinearly as the architecture scales.

On the supply side, a significant share of global optical-module capacity is concentrated among Chinese companies, giving the domestic supply chain a strong competitive position.

There is ongoing debate over the relative merits of optical-module, CPO, and NPO architectures. However, these technological debates are unlikely to change the fundamental long-term growth logic of optical interconnection.

The medium- to long-term outlook for optical communications therefore remains supported by structural demand growth.

5. Industry Valuations and the Risk of an AI Bubble

We currently do not believe that the AI supply chain is experiencing an Internet-bubble-level valuation excess.

At the peak of the Internet bubble, leading companies traded at forward P/E ratios of approximately 130-140x. Today, leading AI-related companies trade at forward P/E ratios of around 22x, while other major overseas technology companies generally trade in the 18-25x range. Memory stocks are trading at less than 10x earnings.

From this perspective, current valuation levels remain far removed from those seen during the peak of the Internet bubble.

A defining characteristic of a bubble is that share prices become detached from earnings growth over an extended

Rosefinch Fund
[021]2030 5888 [021]2030 5999
58 Floor, New Bund Center, NO.555 West Haiyang Road/588 Dongyu Road, Pudong New Area, Shanghai 上海市浦东新区海阳西路 555 号/东育路 588 号前滩中心第 58 层，邮编：200126

period.

At present, however, share prices and corporate earnings are rising broadly in tandem, forward P/E multiples have remained relatively stable, and supply-demand conditions remain tight. The underlying structural growth thesis therefore retains meaningful support.

That said, we must remain alert to the risk that industry fundamentals and market narratives can temporarily diverge, or even move in opposite directions.

In many cases, the stronger the market narrative becomes, the greater the potential for a temporary disconnect between market pricing and fundamentals.

In other words, even if current valuations do not constitute a bubble, the market's strong consensus around the future of AI could itself push share prices temporarily beyond the pace justified by underlying fundamentals.

Overall, relative to the current stage of industry development, we believe valuations across the AI supply chain remain broadly reasonable. The long-term structural growth thesis has not been fundamentally impaired, and the sector continues to offer attractive investment opportunities.

At the same time, investors should remain disciplined and maintain a clear perspective on the potential gap between short-term narrative-driven premiums and the eventual return of valuations to underlying fundamentals.

Email service@rosefinch.cn