

Understanding the HBM Industry

Industry Background

The memory industry has a habit of turning shortages into gluts. Demand for dynamic random-access memory (DRAM), which holds data while processors work, has historically followed spending on PCs, smartphones and servers. When prices rise, manufacturers invest in more capacity, which can arrive just as demand weakens. Factories still carry high fixed costs, leaving profits exposed when chip prices fall. Scale helps manufacturers survive the downturn but offers little protection from falling prices. As recently as fiscal 2023, Micron [NASDAQ: MU] reported a \$5.8bn net loss as revenue almost halved. Investors had good reason to be wary of the next expansion.

AI has changed what customers need from memory. Training a large language model, or running it to answer a prompt, requires processors to move enormous amounts of data. A GPU waiting for that data is expensive computing capacity sitting idle. High-bandwidth memory (HBM) addresses this problem by stacking DRAM chips vertically and placing them beside the processor in the same package. Thousands of connections create a much wider path for data to travel. Capacity determines how much information fits in memory; bandwidth determines how much can be transferred each second. Both matter and buying more of one does not automatically provide more of the other.

The increase in memory per accelerator has magnified demand. Nvidia's [NASDAQ: NVDA] Blackwell Ultra carries 288GB of HBM3E per GPU, compared with 141GB in the H200 and 80GB in the H100. Manufacturers therefore benefit from both more accelerators being deployed, and, across these generations, more memory being installed in each. Supply is unusually expensive to expand. In June 2024, Micron estimated that HBM3E requires roughly three times the wafer capacity of conventional DDR5 to produce the same number of bits at the same manufacturing node. Allocating more production to HBM consequently reduces the capacity available for ordinary DRAM, spreading the pressure beyond AI accelerators. While HBM3E required roughly 3 convention DRAM chips for the same amount of HBM, the higher you stack HBM chips, the more DRAM required, showcasing how this problem is extrapolated with scale.

Single Node HBM Capacity											
System	H100 HGX	H200 HGX	B200 HGX	GB200 NVL72	B300 HGX	GB300 NVL72	Vera Rubin NVL72	MI300 X	MI325 X	MI355 X	MI450 X
Package HBM Capacity (GB)	80	141	180	186	288	268	288	192	256	288	432
# of Chips per Node	8	8	8	72	8	72	72	8	8	8	72
System HBM Capacity (GB)	640	1,128	1,440	13,392	2,304	19,296	20,736	1,536	2,048	2,304	31,104
Memory Available for KV Cache (GB)	N/A	N/A	N/A	11,831	743	17,735	19,175	N/A	487	743	29,543
Fits K3 Weights*	No	No	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes

Source: SemiAnalysis, BSIC

The resulting shortage has changed how memory is bought. In its September 2026 results, Micron said it had already signed agreements covering most of its 2027 HBM bit supply, with significant price increases over 2026. SK hynix has also been signing multiyear agreements with major customers. For suppliers, these commitments

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgement at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

give better visibility when deciding whether to fund another factory. They also show how far procurement priorities have shifted: access to future output is itself worth negotiating well in advance.

Most of that business goes to SK hynix [KRX: 000660], Samsung Electronics [KRX: 005930] and Micron. Counterpoint estimates that SK hynix held 50% of HBM revenue in the second quarter of 2026, down from 58% in the first, with Samsung at 33% and Micron at 18%. This concentration reflects an expensive manufacturing process that rewards accumulated experience. HBM requires advanced DRAM production, precise stacking and packaging, and yields high enough to make the finished product economical. With HBM4, the base die at the bottom of the stack is moving to logic processes, and SK hynix has partnered with TSMC to make it, so each product is increasingly tailored to its customer. A new entrant must also convince customers that its memory will work reliably in an accelerator designed around particular performance and power requirements. A working sample is only the beginning. Samsung, long the leader in conventional DRAM, passed Nvidia's qualification for 12-layer HBM3E only in September 2025, about 18 months after completing development and after a string of failed attempts.

The cost of competing is visible in the incumbents' budgets. Micron reported \$27.4bn of net capital expenditure in fiscal 2026 across its memory and storage business. Even once the money is committed, new facilities take years to deliver meaningful output. These obstacles give established suppliers time to respond to challengers. China's ChangXin Memory Technologies (CXMT) nevertheless deserves attention. In August, The Information reported that it had begun producing HBM3E in small quantities for testing by Chinese chip designers, with expansion planned for 2027. Our view is that it remains an emerging competitor; trial production alone does not establish an ability to supply leading AI platforms at scale.

Investors have been quick to reward the change in demand. By late September, Micron's shares had almost quadrupled since the start of 2026, despite retreating from their summer peak. In June, SK hynix briefly overtook Samsung as South Korea's most valuable company, after its shares rose almost 900% in twelve months. The attraction is understandable: AI spending supports higher sales, while scarce supply gives manufacturers room to negotiate prices. Those conditions have encouraged enormous investment, and the new capacity will eventually compete for orders. Long-term contracts improve visibility, but the return on a factory still depends on demand over many years.

For HBM, that longer-term demand depends partly on choices made by a small number of accelerator designers. More AI computing does not imply a fixed quantity of memory per chip, and the balance between capacity, bandwidth and cost can change with the workload. If designers shift towards configurations that need fewer DRAM dies, the same wafers would yield far more HBM, and the shortage could ease sooner than current valuations assume. Understanding those choices, starting with why HBM matters for AI, is essential to judging how long the current shortage can last.

Importance of HBM for AI

HBM demand growth has come as a result of AI processors only being able to process data as quickly as memory can supply it. Modern GPUs are comprised of thousands of parallel compute units that are able to perform large quantities of matrix operations every second. These operations need model weights, activations and also other data to be moved all the time between memory and the processor. In a situation where the data can't be transmitted fast enough, a memory wall is reached, a phenomenon where expensive compute cores are left waiting and additional processing power provides limited incremental performance.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgement at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

HBM was created to help alleviate this bottleneck. While traditional DDR memory connects to a processor by using a narrow interface operating at high clock speeds, HBM, on the other hand, stacks multiple DRAM dies vertically and puts them right next to the accelerator in the same package. The layers are connected by thousands of Through-Silicon Vias (TSVs). The consequence of this is a much wider 1,024-bit data interface that uses less energy per bit and also allows significantly more data to be transferred in parallel.

This is important due to the fact that the value of an AI accelerator is based on not only its theoretical FLOPS, but also on how effectively those FLOPS can be used. HBM is basically the high-speed reservoir right next to the GPU that keeps its compute engines well supplied with data. For example, Micron's HBM3E delivers over 1.2TB/s of bandwidth per stack and the combination of Nvidia's H200 and HBM3E stacks can provide an even greater 4.8TB/s of aggregate memory bandwidth. This is largely due to Nvidia's switch from its H100 to its H200 that, despite using the same Hopper architecture, increased memory capacity to 141GB and bandwidth by 1.4 times to 4.8TB/s. The resulting outcome was substantial improvements in LLM inference due to the additional bandwidth relieving memory-bound parts of the workload and enabling the Tensor Cores to be used much more effectively.

It is important to note that HBM contributes to both capacity and bandwidth and understanding the difference between the two is crucial to understanding the future direction of the industry. Capacity relates to the quantity of data that can be kept close to the accelerator at any given time, while bandwidth is responsible for determining how fast that data can be accessed. Both of these resources are crucial for AI, buying more of one does not automatically provide more of the other, and their relative importance depends on the workload.

When training a model memory capacity is crucial because there is a large memory footprint required to store not only the model weights, but also activations, gradients, and other intermediate states. A lack of sufficient local memory increases communication overhead and reduces efficiency due to data being partitioned across more accelerators or having to be moved through slower levels of the memory hierarchy. This is ultimately what led to the industry's drive towards continuously larger HBM capacities with each new accelerator generation.

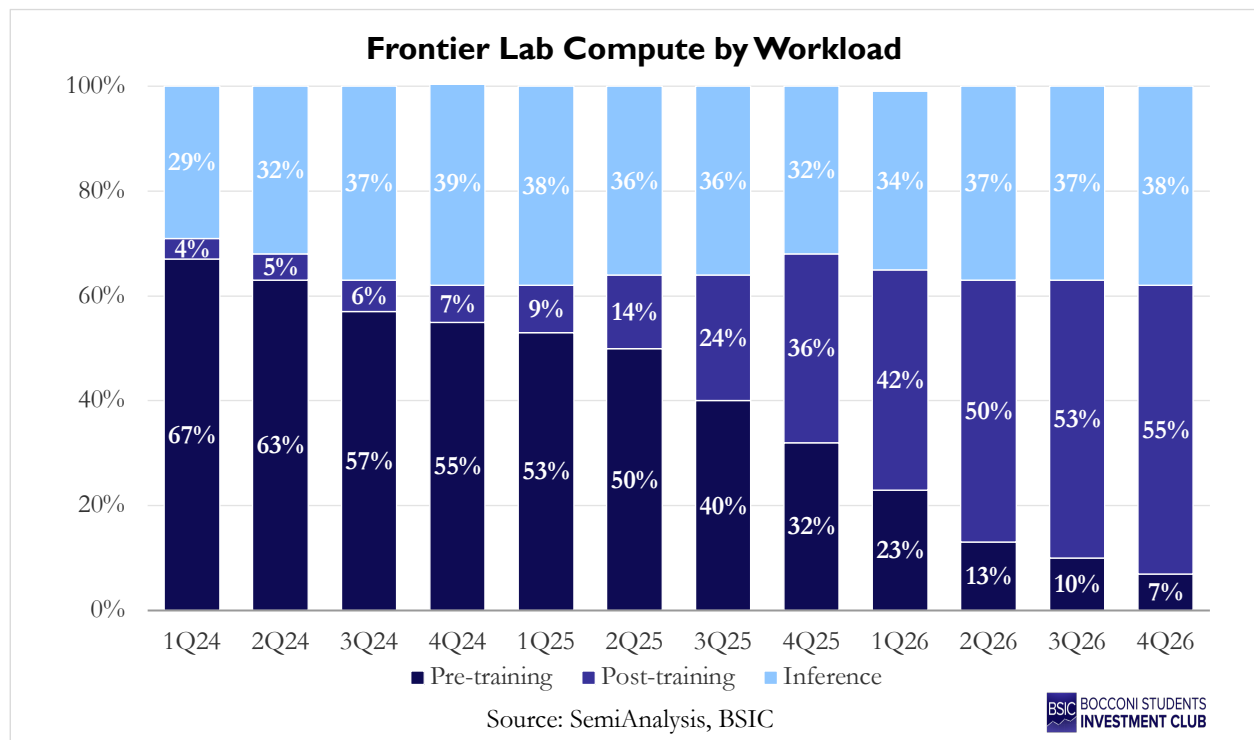
On the other hand, the economics of inference are different. After an LLM has already been trained, the generation of each additional token requires the accelerator to continuously read model weights and the relevant KV cache from memory. For this reason, the decode stage of LLM inference is memory-bandwidth bound as opposed to being compute bound. Nvidia describes the autoregressive decoding process explicitly as memory-bound, with transferring weights, keys, values, and activations dominating latency. Therefore, higher bandwidth is directly associated with higher tokens per second and reduced inference costs.

However, capacity is still important since the model weights have to fit somewhere as the KV cache expands with the number and length of simultaneous user requests. If the HBM capacity is insufficient, this can lead to performance penalties caused by workloads split among more accelerators or spilling into slower memory. It is important to note, though, that once sufficient capacity for the necessary working set is established, adding further memory doesn't improve performance. Bandwidth can still remain fully utilised even when some of the available capacity is unused.

This distinction has become increasingly relevant as AI infrastructure has moved from individual GPUs to more tightly connected rack-scale systems. Models are now no longer required to fit in the memory of a single accelerator and can now have their weights shared among hundreds of GPUs connected via high-speed scale-up networks like NVLink. As the memory available to a model grows, the minimum required amount of HBM for individual accelerators can fall despite total system capacity increasing. A paper written by SemiAnalysis makes the argument that this development, coupled with quantisation and increasingly inference-heavy workloads, is starting to weaken the historical relationship between newer accelerator generations and the need for increasing HBM capacity per

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgement at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

GPU. The marginal data centres being stood up around the world to meet the compute demand primarily focus on reducing inference time. This shifts the memory constraint away from capacity and towards bandwidth.



This begs the following question: If the primary value of HBM increasingly lies in bandwidth as opposed to maximum capacity, how much HBM does an AI accelerator actually need? Until recently, the industry's answer was simply as much HBM as possible, with each accelerator generation using more HBM cubes and denser memory dies, making memory an increasingly vital part of accelerator cost. This has been seen in the progression from 4-hi to 8-hi and, most recently, 12-hi, with 16-hi expected to follow soon. However, SemiAnalysis has argued that inference chips may be better served by shorter 4-hi or 8-hi stacks, which deliver similar bandwidth from fewer DRAM dies, suggesting that this assumption may need to be revisited.

The New Bear Case

In the last year, one of the most critical bottlenecks in the AI buildout has been the shortage of HBM supply. As a result, the three major suppliers are all increasing spending to buildout additional capacity in attempt to capture as much revenue and margin as possible while the demand is hot. SK Hynix has begun construction on its new Yongin 1 fab, which is expected to begin operating in 2027. Once that clean room begins operations, SK Hynix will break ground on an additional project on the same site, expected to begin producing in 2029. Samsung and Micron each have their own respective clean room fab projects underway that are expected to begin producing DRAM for HBM capacity over a similar time horizon. Since all of the HBM capacity is already purchased forward through 2026 and most of 2027, the industry is currently feeling the impacts of the extreme shortages that will start to decompress post the activation of these facilities from 2027 and 2028 onward. Despite the new capacity that will be introduced, the industry will remain in a shortage through 2030 based on most projections. This shortage and margin expansion in recent quarters (SK Hynix EBITDA margin: 81.4% 2Q26A) has allowed these stocks to explode over the last year (SK Hynix: +365.5%, Micron: +485.0%, and Samsung: +207.5%).

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgement at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

In August of 2026, SemiAnalysis published a report on what they think could be the future of HBM memory going forward: Nvidia decreasing the HBM capacity of the Vera Rubin Ultra that is expected to ship sometime in 2H27. SemiAnalysis gives two reasons for this. The first is the supply constrain that has delayed their base level Vera Rubin stacks to late 2026 after having announced them over a year ago now. With less HBM needed, Nvidia will be able to ship more product without facing the same supply constraints the market is currently navigating. Second, and more importantly, SemiAnalysis argues that higher stacks of HBM are not economically viable given the current state of AI and shorter stacks offer a cheaper alternative with only marginal technological loss. This situation, explained more in the previous section, create a brand-new bear case that the market has not priced in since the SemiAnalysis article was published as all the memory stocks were up between 10-15% through September.

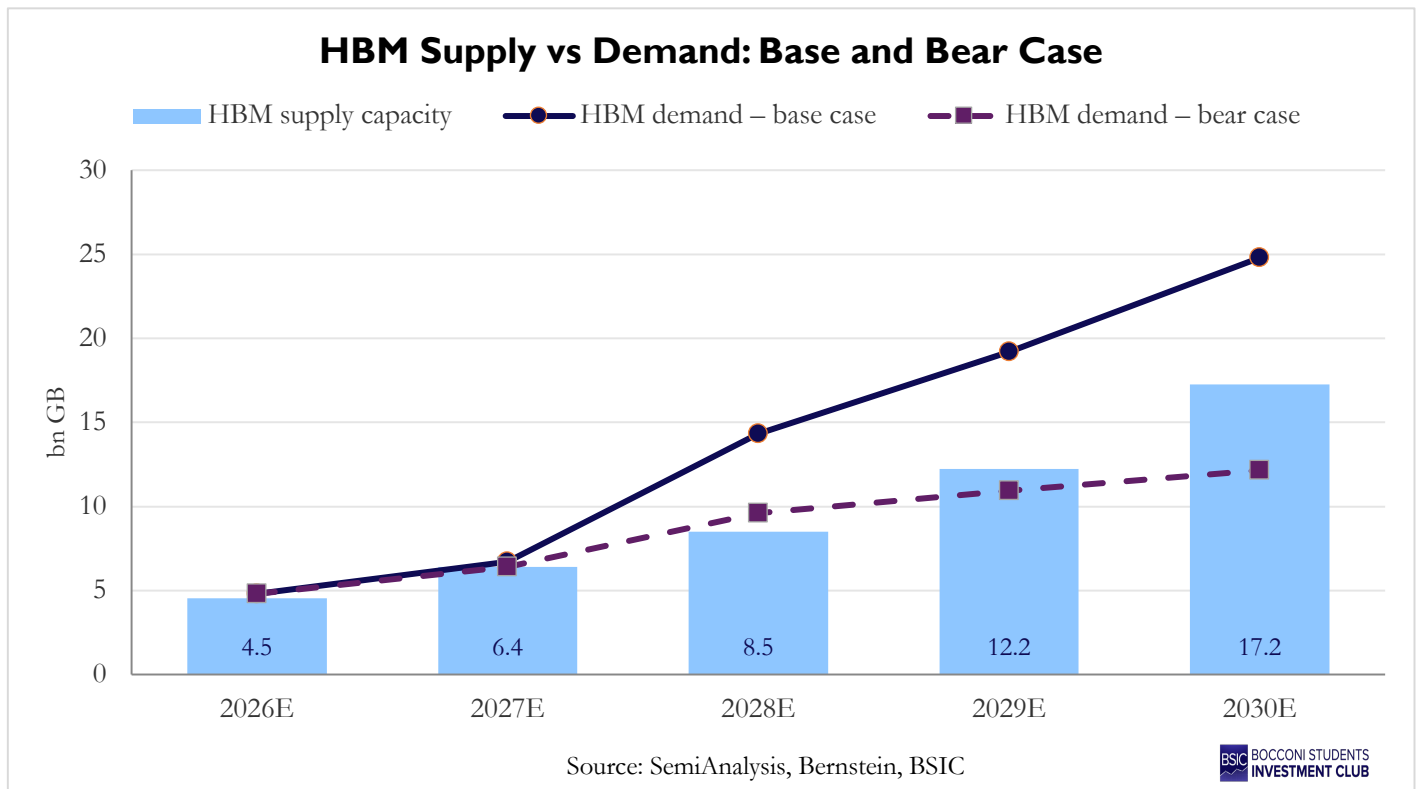
If this scenario plays out and the number of racks Nvidia ships does not significantly increase, we could see the demand of HBM per rack start to drop with Vera Rubin Ultra towards the end of 2027 and drop significantly in subsequent years as future racks shift toward to proposed norm of 4-hi or 8-hi stacks.

To test how far this changes the picture, we ran the SemiAnalysis scenario through our supply-demand model. We hold AI rack-equivalent shipments on our base-case path (160,000 in 2026 rising to 335,000 in 2030) and keep non-AI DRAM demand unchanged, so the only variable that moves is how much memory each rack carries. HBM content per rack sits 5% below our base case in 2027 as Vera Rubin Ultra ships with shorter stacks, then 33%, 43% and 51% below in 2028, 2029 and 2030 as 8-hi becomes the standard. In effect, HBM per rack stops growing at roughly 36TB from 2028, against 54TB rising to 74TB in our base case. Because the KV cache that no longer fits in HBM still has to sit somewhere, we raise conventional DRAM per rack by 5-15% over the same period.

The result is that HBM demand grows by roughly 24% a year between 2027 and 2030 rather than 55%, reaching 12.1bn GB in 2030 compared with 24.8bn GB in our base case. The capacity behind our supply estimates is already committed: SK hynix's Yongin site, Micron's Idaho and Tongluo fabs and Samsung's P5 all ramp over 2027-2030 regardless of what Nvidia decides for Vera Rubin Ultra. We estimate the industry's HBM supply capacity rises from 6.4bn GB in 2027 to 17.2bn GB in 2030, growth of roughly 39% a year that was sized for the base-case demand curve.

The gap between those two curves is what turns the shortage into a surplus. In our base case, HBM supply falls 41% short of demand in 2028 and is still 30% short in 2030, consistent with the market's view that the shortage lasts through the decade. In the bear case, 2027 is roughly balanced and the 2028 shortage narrows to 11%, as the new fabs are not yet at volume. By 2029 supply exceeds demand by 12%, and by 2030 the surplus reaches 42%: the industry could produce about 17.2bn GB of HBM against demand of 12.1bn GB, leaving close to 30% of the HBM capacity being built today without a buyer. The share of DRAM wafers used for HBM peaks at 37% in 2028 and falls back to 29% by 2030, well below the 41% the industry is preparing to allocate.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgement at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.



The excess does not stay within HBM. Wafers that are no longer needed for stacking return to conventional DRAM, and since each HBM bit uses about four times the wafer capacity of a conventional bit in our model (rising from 3.8x in 2027 to 4.4x in 2030), every wafer released from HBM produces several times as many conventional bits. In the bear case, conventional DRAM supply in 2030 is 21% higher than in our base case, and the conventional market moves from a 20% surplus to a 36% surplus, even after the additional DRAM per rack is included. Across all DRAM, wafer demand in 2030 falls to 29.9m from 39.8m in our base case, turning an 8% wafer shortfall into a 23% surplus.

If the bear case begins to play out upon future Nvidia announcements, it will be critical for the memory suppliers to react accordingly and slow down their capex and prevent themselves from being stuck with excess clean room capacity that is not going into making more DRAM and HBM chips. This would seriously impact the potential downside of the memory trade that is currently priced for continues growth.

Our Valuation

The supply-demand work above is ultimately useful only if it can explain what happens to the economics of an individual memory producer. For SK hynix, the transmission mechanism is relatively direct. AI infrastructure determines how much HBM and conventional DRAM the market needs; the amount of wafer capacity available determines how much of that demand can actually be served; and the gap between the two determines the price at which SK hynix can sell its output. Our valuation therefore starts with the industry rather than with a top-down revenue growth assumption.

Framework

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgement at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

We model SK hynix's DRAM revenue from the bottom up. On the demand side, the two most important variables are the number of AI systems deployed and the amount of memory required by each system. On the supply side, we forecast industry wafer capacity, the productivity of those wafers and the share allocated to HBM rather than conventional DRAM. SK hynix then receives a share of the resulting industry volumes based primarily on its available capacity.

The critical link between the two markets is the amount of wafer capacity consumed by HBM. An HBM bit requires several times the wafer capacity of a conventional DRAM bit, and this penalty rises as stacks become taller. Higher HBM demand therefore has two effects at once. It increases the amount of high-value HBM SK hynix can sell, but it also removes capacity from conventional DRAM production. The result is that an AI-driven HBM shortage can support pricing across the broader DRAM market rather than remaining isolated within HBM.

This is also why we do not model revenue by simply applying a growth rate to SK hynix's historical sales. When the industry is running close to its physical capacity ceiling, volumes cannot respond freely to demand. A stronger demand environment may result in only modestly higher shipments but significantly higher prices. Conversely, an improvement in effective supply can cause ASPs to decline sharply even while the underlying number of memory bits consumed continues to grow.

Pricing is therefore the main mechanism through which differences between our scenarios reach SK hynix's top line.

Scenario Logic

Our scenarios are built around three variables: AI infrastructure deployment, memory content per system and effective memory supply. We deliberately keep most other assumptions relatively stable. This allows us to isolate what we believe is the central debate for the industry: whether demand for DRAM wafers grows faster or slower than the capacity being built to serve it.

Our base case assumes that AI infrastructure continues to expand rapidly and that memory requirements per system continue rising, but that the current shortage gradually begins to ease as the new fabs already under construction come online. SK hynix, Samsung and Micron are all adding capacity, but construction timelines mean that this supply arrives with a significant lag. The result is several years in which demand remains ahead of supply, followed by a gradual move towards a more balanced market.

Importantly, this does not require today's exceptional pricing environment to continue forever. We assume that HBM and conventional DRAM prices eventually normalise as new supply becomes productive. What matters for equity value is how much cash SK hynix can generate before that normalisation takes place.

The bull case extends this period of scarcity. The most obvious source of upside is simply more AI infrastructure: more racks require more memory. However, the more powerful driver is memory intensity. If future accelerators continue following the historical pattern of carrying more HBM, each additional unit of compute consumes progressively more DRAM wafer capacity.

That creates a nonlinear effect on pricing. Once the industry's fabs are close to full utilisation, stronger demand cannot be met simply by producing proportionately more chips. Instead, customers increasingly compete for a fixed amount of available supply. Higher HBM demand also consumes wafers that would otherwise have produced conventional DRAM, tightening that market simultaneously. The bull case is therefore not primarily a story about SK hynix gaining market share; it is a story about scarcity lasting longer and being monetised at higher prices.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgement at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

There are plausible reasons why even this could prove conservative. AI deployment today is dominated by hyperscale data centres, but the range of potential memory-intensive workloads is expanding. More capable autonomous agents, increasingly long context windows, multimodal models and eventually large-scale robotics could all increase the amount of inference performed and the memory required to support it. In such a world, the industry may discover that the capacity currently being built for the late 2020s is absorbed almost as quickly as it becomes available.

The opposite outcome does not necessarily require an AI collapse.

The Bear Case

This is where the new bear case discussed above becomes important. Most traditional downside cases for memory assume weaker end-demand or excessive capital expenditure. The shorter-stack thesis introduces a different possibility: AI demand can remain strong while the amount of HBM required to satisfy it falls materially.

If bandwidth increasingly matters more than maximum local capacity for inference, the industry may no longer need to keep increasing the number of DRAM dies in every HBM stack. Moving from 12-hi towards 8-hi or 4-hi configurations would allow the same wafer input to produce substantially more HBM packages. This effectively increases supply without requiring a new fab.

That distinction is crucial because the physical capacity currently under construction cannot be cancelled easily. Yongin, Samsung's new capacity and Micron's expansion plans were designed around a world in which HBM demand continues consuming an increasing share of the industry's DRAM wafers. If memory content per rack instead levels off while those fabs continue to ramp, the supply-demand balance changes very quickly.

Our bear case therefore keeps AI rack deployment broadly intact but reduces HBM content per rack as shorter stacks become more prevalent. Some of the memory removed from HBM is shifted towards conventional DRAM, particularly where slower memory can support KV-cache requirements, but this does not offset the wafer impact. Because each HBM bit consumes several times more capacity, releasing a relatively small amount of HBM demand frees enough wafers to produce a much larger quantity of conventional DRAM.

The result is the dynamic shown in our supply-demand work above. A market that remains structurally short in the base case can move through balance and into excess capacity before the end of the decade. At that point the effect on SK hynix is no longer primarily lower shipment growth. It is lower pricing across both HBM and conventional DRAM.

This is what makes the bear case particularly damaging to the equity story. Memory companies have enormous operating leverage. A fab continues to incur depreciation, labour and other fixed costs regardless of where memory ASPs settle. Once incremental supply begins competing for a limited pool of demand, a relatively small change in industry utilisation can create a much larger change in prices, margins and free cash flow. The same operating leverage that is currently producing record profitability works in reverse.

From Supply-Demand to Revenue

The difference between the cases can therefore be reduced to a relatively simple chain: AI systems deployed → memory per system → wafer demand → industry utilisation → ASPs → SK hynix revenue.

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgement at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

The bull case pushes the first two variables higher and keeps utilisation elevated for longer. The bear case changes the third variable indirectly: shorter HBM stacks reduce the amount of wafer capacity required for a given level of AI compute, effectively increasing available supply. Our base case lies between the two, with strong structural demand but gradual normalisation as new capacity comes online.

This framework also explains why our revenue forecasts can move much more than our shipment forecasts. In a shortage, the incremental dollar of demand is increasingly captured through price rather than additional volume. In a surplus, the reverse happens. Producers can continue shipping more bits while generating less revenue per bit.

For NAND, where our thesis is less differentiated, we take a simpler approach. Rather than attempting to forecast a second semiconductor cycle in equal detail, we apply more conventional growth assumptions and leave the valuation primarily exposed to HBM and DRAM, where the supply-demand dynamics form the core of our view.

Valuation

We translate the resulting revenue paths through SK hynix's cost structure and into free cash flow before applying a DCF. The purpose of the DCF is therefore not to create the thesis through a terminal multiple; it is simply to value the cash flows produced by the supply-demand scenarios described above.

Our current selected case implies an equity value of approximately ₩1.85m per ordinary share, around 35% above the market price used in our model. The result is based on an approximately 11.8% cost of capital and a 2% terminal growth rate. We also cross-check the output against forward earnings and EBITDA multiples rather than relying exclusively on the terminal value.

More important than the precise target price, however, is what the valuation tells us about the distribution of outcomes. The market is attempting to value a company at a point when both its earnings and the industry's capital spending are far outside historical norms. Small differences in assumptions about HBM content or the timing of new supply compound into very different revenue and free-cash-flow trajectories.

At one extreme, continued growth in memory intensity allows demand to absorb each new wave of capacity, keeping utilisation high and extending SK hynix's pricing power. The industry begins to look structurally different from previous memory cycles, with HBM contracts, technological barriers and AI demand supporting returns for longer than investors are accustomed to.

At the other extreme, the industry could discover that it has spent tens of billions of dollars solving a bottleneck just as the architecture of AI systems begins to require less of the constrained resource. Demand for AI compute could continue increasing while HBM utilisation falls, releasing wafers into conventional DRAM and recreating the oversupply conditions that have historically defined the memory cycle.

That asymmetry is why we do not think the central question for SK hynix is simply whether AI continues growing. Almost every reasonable scenario assumes that it does. The more important question is how much DRAM wafer capacity is required for each incremental unit of AI compute.

Conclusion

The HBM boom has given memory manufacturers something they have historically struggled to achieve: visibility. Capacity is sold years in advance, customers are negotiating for access to supply, and the technical difficulty of

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgement at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.

producing leading HBM has concentrated the market among a handful of suppliers. For SK hynix, the result has been extraordinary revenue growth, margins and cash generation.

But visibility is not the same as permanence. The fabs being built today will continue arriving through the end of the decade. Whether they tighten or oversupply the market depends on a set of architectural decisions that have not yet been made. If future AI systems continue consuming more HBM per unit of compute, those fabs may struggle to keep up. If shorter stacks and more efficient memory architectures take hold, the same additions could turn today's shortage into tomorrow's excess capacity.

That is ultimately the framework behind our valuation. The long case is not simply that AI grows; it is that AI's demand for scarce DRAM wafers grows faster than supply. The bear case is that compute keeps growing while the wafer requirement does not. For a sector whose economics can swing from shortage to glut in a matter of years, that distinction is what matters most.

Appendix

- [1] Micron Technology, "[2025 Form 10-K](#)", 2025
- [2] Micron Technology, "[Fiscal 2023 Results](#)", 2023
- [3] Micron Technology, "[High Bandwidth Memory Overview](#)", 2026
- [4] Nvidia, "[Inside NVIDIA Blackwell Ultra: The Chip Powering the AI Factory Era](#)", 2025
- [5] Micron Technology, "[Fiscal Q3 2024 Prepared Remarks](#)", 2024
- [6] Micron Technology, "[Fiscal Q4 2026 Prepared Remarks](#)", 2026
- [7] SK hynix, "[Q2 2026 Business Results](#)", 2026
- [8] Counterpoint Research, "[Global DRAM and HBM Market Share](#)", 2026; Korea JoongAng Daily, "[Samsung Narrows Gap with No. 1 SK hynix in HBM Market Revenue in Q2](#)", 2026
- [9] Micron Technology, "[Fiscal Q3 2026 Results](#)", 2026
- [10] Micron Technology, "[Fiscal 2026 Results](#)", 2026
- [11] Reuters (via Boursorama), "[Report on CXMT HBM3E Production](#)", 2026
- [12] Investopedia, "[Micron Shares Ahead of Earnings](#)", 2026
- [13] Nvidia, "[H200 Tensor Core GPU](#)", 2024
- [14] SK hynix, "[SK hynix Partners with TSMC to Strengthen HBM Technological Leadership](#)", 2024
- [15] KED Global, "[Samsung HBM3E Qualification](#)", 2025
- [16] Fortune, "[SK hynix Market Value](#)", 2026
- [17] SemiAnalysis, "[Long Live the Short King](#)", 2026
- [18] Nvidia, "[Mastering LLM Techniques: Inference Optimization](#)", 2023

Tags: HBM, High-Bandwidth Memory, DRAM, Semi-conductors, AI, SK Hynix, Samsung, Micron, Nvidia, CXMT

All the views expressed are opinions of Bocconi Students Investment Club members and can in no way be associated with Bocconi University. All the financial recommendations offered are for educational purposes only. Bocconi Students Investment Club declines any responsibility for eventual losses you may incur implementing all or part of the ideas contained in this website. The Bocconi Students Investment Club is not authorised to give investment advice. Information, opinions, and estimates contained in this report reflect a judgement at its original date of publication by Bocconi Students Investment Club and are subject to change without notice. The price, value of and income from any of the securities or financial instruments mentioned in this report can fall as well as rise. Bocconi Students Investment Club does not receive compensation and has no business relationship with any mentioned company.