FINANCIAL TIMES – 03 SETTEMBRE 2025

How AI has powered up memory chips

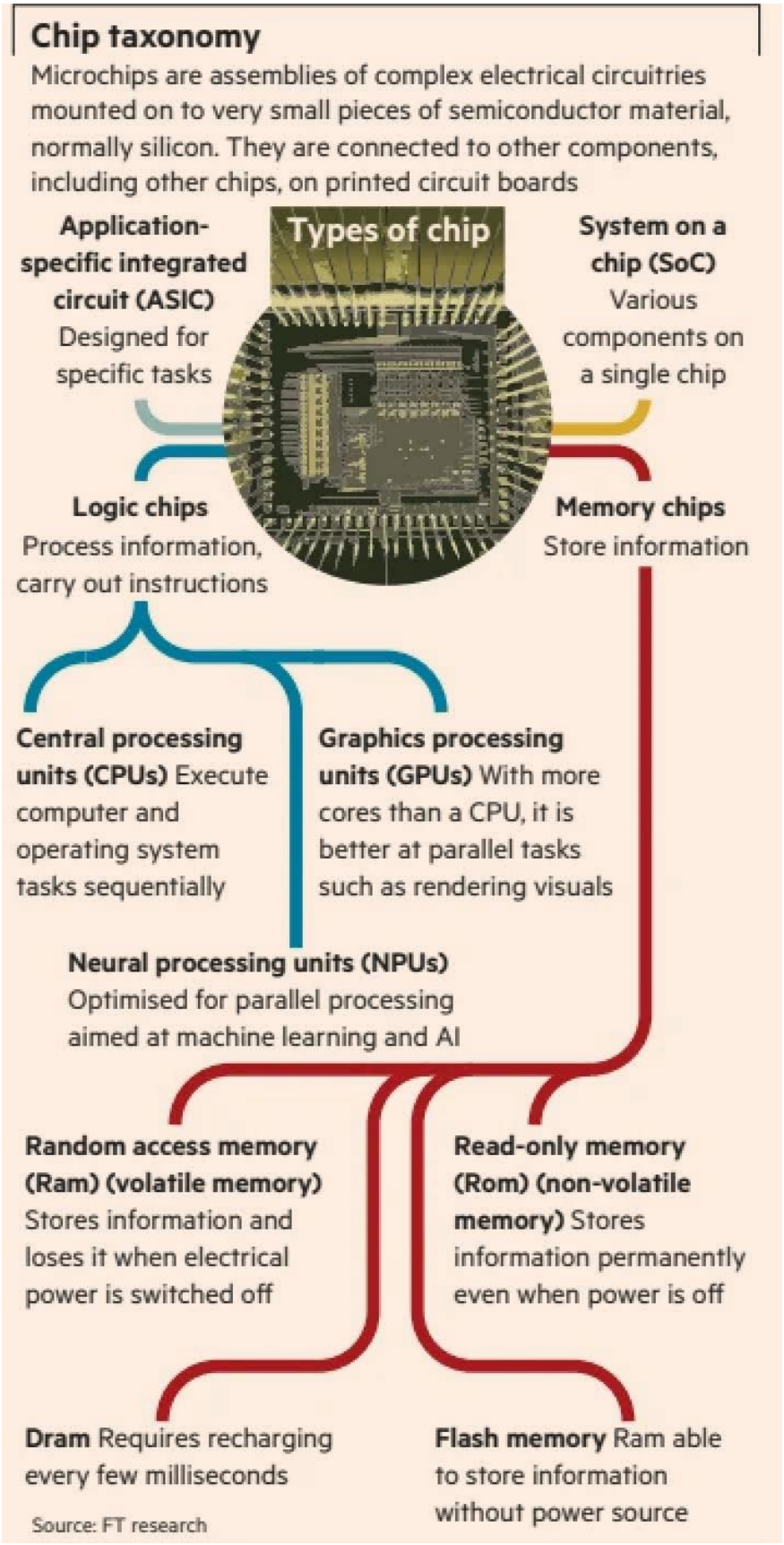
di Christian Davies

Innovation in the previously unglamorous and commoditised sector has resulted in Samsung losing its market leadership to SK Hynix — and put chipmakers in the middle of US-China tensions.

At SK Hynix's vast M14 chip fabrication plant, workers in clean-room suits inspect rows of machines as 700 robots zip along overhead rails, carrying silicon wafers between stages of the manufacturing process.

The factory, in the South Korean city of Icheon, produces high-bandwidth memory (HBM) chips, capable of transferring data equivalent to 200 featurelength movies every second.

For decades, memory chips were the unglamorous end of the semiconductor industry, overshadowed by the logic or processor chips that conduct calculations and control a device's operations.



But HBM designs, such as the HBM3E produced at the Icheon factory, are transforming the memory industry. Joon-yong Choi, vice-president and head of HBM business planning at SK Hynix, notes that whereas in conventional dynamic random-access memory (Dram), “cost was prioritised by customers over power and performance, with HBM power and performance are prioritised over cost”.

They are helping AI developers alleviate the effects of the “memory wall” — where limitations in storing and retrieving data impede improved performance — as well as boosting efficiency and lowering costs at thousands of data centres under construction around the world.

The growing importance of memory to AI has put the sector at the centre of intensifying competition between Washington, which is trying to restrict Chinese access to cutting-edge technology, and Beijing, which is nurturing a domestic semiconductor sector that can go toe-to-toe with global rivals.

It has also shifted the historic order of the industry’s top players. SK Hynix’s revenues from Dram, of which HBM is a subset, swelled from Won7.5tn ($5.4bn) in the second quarter of 2021 to Won17.1tn in the same quarter of 2025.

That pushed them past those of its rival Samsung for the first time since the two Korean groups started competing in memory in the 1980s, and “would have been unthinkable as recently as five years ago”, according to Chris Miller, associate professor at Tufts University and author of Chip War. “It would be like Dr Pepper suddenly becoming more popular than Coca-Cola.”

Miller adds that the past commoditisation of the memory chip market prompted many of the brightest minds in the sector, such as Nvidia’s Jensen Huang and Qualcomm’s Irwin Jacobs, to turn their attention to processor chips.

“But now, memory is back.”

Intel began life in the 1960s as a memory chip company, but exited the Dram sector in the 1980s under pressure from Japanese rivals Toshiba and NEC. They, in turn, were supplanted in the 1990s by Samsung and the chip division of Hyundai Electronics, or Hynix, which would later be acquired by the SK conglomerate. The two Korean groups and Micron, of the US, have dominated the sector ever since.

Samsung was until recently the undisputed leader in Dram chips, which are powered and store data temporarily while a processor is running.

Choi explains that while Dram and lower-value Nand chips — which store data for longer periods without power — were the dominant technologies, companies also experimented with more niche products.

HBM chips, which Hynix began developing in 2013, were among them. They involved stacking layers of Dram units connected by copper wires a tenth of the thickness of a human hair, like a multistorey library with lifts to quickly transport piles of books between floors.

That means HBM chips can offer many more pathways for sending data to and from a processor, Choi explains. “Think of it like the number of taps filling a water tank, or the number of lanes on a highway,” he says.

Ray Wang, lead semiconductor analyst at the Futurum Group consultancy, also cites Hynix’s early adoption of a bonding technology that prevents overheating, called mass reflow-moulded underfill, as key to its HBM success. Samsung and Micron had to settle for an inferior process that risks cracking the silicon layers, resulting in a higher failure rate for the chips.

Its superior product helped Hynix to become the principal supplier of HBM chips to Nvidia and allowed it to ride on the US company’s coat-tails as demand for AI chips exploded after OpenAI’s ChatGPT chatbot was released in 2022.

HBM’s share of Hynix’s overall Dram revenues went from about 5 per cent in the final quarter 2022, according to Bernstein research, to more than 40 per cent by the first quarter of 2025.

Myron Xie of the consultancy SemiAnalysis notes that while Micron’s current HBM3E chips have now passed the stringent qualification tests for use in Nvidia’s most advanced AI chips, Samsung’s equivalent is yet to do so.

People close to the company say its HBM3E chip is set to pass Nvidia’s tests “imminently”. But they also acknowledge that it was caught cold by the needs of AI players for memory solutions tailored to their specific requirements.

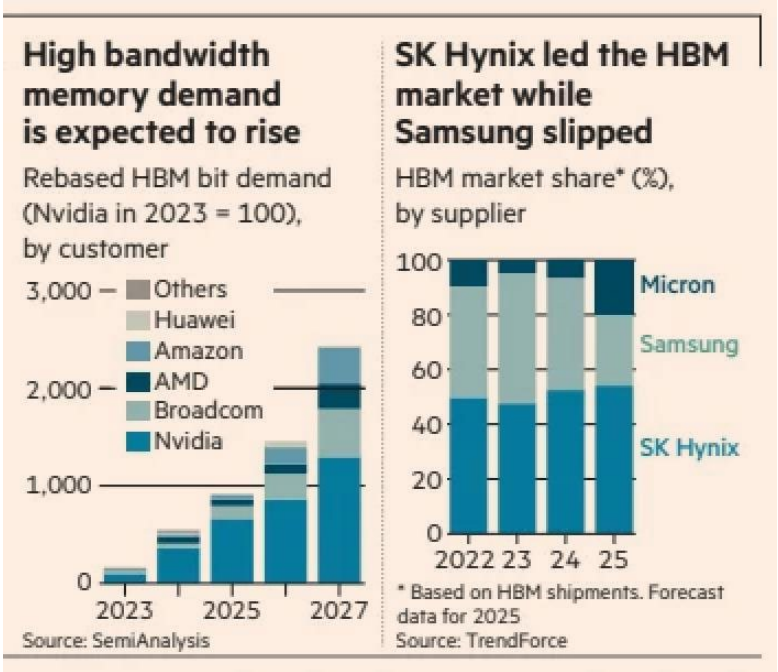
“Samsung is also struggling with the bread-and-butter task of producing the cutting-edge Dram chips that end up getting stacked into HBM,” says Xie. “Micron has done well, but it’s also pretty damning against Samsung for it to be in third place.”

Miller adds Samsung “was optimised for the smartphone era” rather than the AI one. “I think the entire organisation has struggled to conceive of what the world would look like when the smartphone was no longer the dominant product.” It has proved a costly stumble. Peter Lee, a Seoul-based semiconductor analyst at Citigroup, notes that HBM chips offer profit margins of about 50-60 per cent, compared with about 30 per cent for conventional Dram units.

Because each HBM chip needs to be designed to fit the AI graphics processing unit to which it is paired, orders must be placed a year before production, typically on one-year contracts.

“That gives memory companies much more pricing leverage over potential customers than when selling conventional Dram, which can be bought at a month’s or even a day’s notice and is easily swapped for the product of a rival chipmaker,” adds Lee.

Wang estimates that Samsung's missteps have cost it tens of billions of dollars a year in lost revenue since ChatGPT's release. "Underestimating the potential of HBM was a huge strategic mistake," he says.



If the rise of HBM has disrupted the old order at the top end of the memory market, another source of disruption is emerging from below, in the shape of ChangXin Memory Technologies, or CXMT. Based in China's province of Anhui, CXMT has increased its share of the global Dram market from close to zero in 2020 to 5 per cent last year, according to Shenzhen-based consultancy Qianzhan. However, it remains unclear whether CXMT's progress in conventional Dram will allow it to catch up with Hynix, Samsung and Micron in mass producing cutting-edge HBM chips — a development that would potentially reduce the dependence of Chinese AI developers and chipmakers on foreign companies for critical components.

Analysts and industry insiders remain sceptical that, without access to key equipment and materials that are subject to US export controls, CXMT can close the HBM gap in the near future.

"CXMT stockpiled a lot of the equipment it needed before the most recent round of controls," says Futurum Group's Wang. "But it cannot access extreme ultraviolet machines, and it is not clear that they have enough equipment to mass produce advanced HBM products at a similar scale to the leading memory players." He estimates CXMT is "three to four years behind" in HBM development.

Last week, the US government revoked waivers that had allowed Hynix and Samsung to send chip-making equipment to their manufacturing facilities in China without a licence, a decision that Wang says "underscores Washington's intent to further limit China's access to memory technology".

In a sign of the importance of HBM for China's AI ambitions, the country's tech groups and chipmakers rushed to stockpile Korean HBM chips before their availability was restricted by US export controls in December last year.

Despite those controls, many experts argue that US policymakers have been slow to recognise memory's centrality to AI performance, leaving Chinese companies with access to cutting-edge memory technologies long after restrictions were imposed on the most advanced processing chips.

While individual HBM chips above a certain standard can no longer be exported to China, more advanced chips can if they have been pre-packaged into AI chips that do not exceed certain performance criteria.

Xie, of Semi-Analysis, cites Nvidia's H20 chip, which US President Donald Trump has permitted to be sold into China on the basis that it is "obsolete". While its processing power is inferior to the flagship Nvidia H100, its memory performance is better than the H100 and Huawei's flagship Ascend 910c.

"People still focus on compute, but that's only one aspect of performance," he says, adding that memory is more relevant for deployment of AI models.

SK Hynix's current dominance of HBM could yet come under pressure from other quarters, analysts say. The Korean company plans to begin mass production this year of its next-generation HBM4 chip that is set to be used in Nvidia's forthcoming Rubin platform.

The HBM4 will use an advanced processor chip produced by TSMC as the "logic die" that regulates memory operations, rather than the relatively unsophisticated Dram chips used thus far.

Samsung's HBM4 offering will also use a processor chip produced by its own foundry division. A person familiar with its thinking told the FT that as the only company with cutting-edge capabilities in both processor and memory chips as well as advanced packaging — the process of integrating multiple chips closer together — it could offer customers a "one-stop shop".

The person adds that the company is also in "active discussions with key clients" about a technique called "hybrid bonding", an improved method of connecting chips that could offer greater bandwidth, power and signal integrity.

Wang, at Futurum Group, argues that whoever manages to incorporate hybrid bonding first "will decide who has leadership in the next generation of HBM".

Lee of Citigroup notes that as well as better performance, the use of processor chips as the logic die will increasingly allow HBM products to be tailored for specific tasks, making it even harder for customers to switch suppliers.

That raises the question of whether HBM4 will pave the way for Samsung to make a comeback. The larger company has had time to correct the mistakes it made on its HBM3E designs for HBM4, meaning it is likely to qualify as a supplier to Nvidia, says Lee.

But Wang notes that Hynix will continue to benefit from its recent close cooperation with Nvidia, as well as its long-standing relationship with industry leader TSMC. Today, Hynix unveiled high numerical aperture extreme ultraviolet lithography machines it has acquired from ASML, giving it a further edge over its HBM rivals.

In contrast, Xie at SemiAnalysis notes that Samsung's foundry and memory businesses have both been dogged by quality and production issues. "There's not much value to being a one-stop shop if all the individual parts are inferior."

Samsung said in a statement that it "continued to invest in the HBM business and is focusing on developing nextgeneration memory technologies".

Chinese AI chipmakers unable to circumvent US export controls are looking for ways to reduce their dependence on HBM as a means to boost performance. Huawei last month launched new AI software designed to allocate different memory tasks to different kinds of memory chip, thereby reducing HBM reliance, and has also unveiled three new "AI solid state drives" as alternative memory solutions.

Miller, of Tufts University, says that many tech companies are trying to develop alternatives to HBM, which remains energy-intensive and relatively expensive. They include Japanese tech investment group SoftBank, which is working with Intel to develop a Dram product that utilises a different wiring system to HBM.

Most analysts agree that HBM will probably dominate memory solutions for the next five years at least. But greater customisation is likely to mean deeper involvement of foundry companies, chip designers and the customers themselves in the design and manufacturing process.

That could threaten the memory companies' ambitions to capture more value from the supply chain, warns Xie. "The more parts of the HBM stack get outsourced to TSMC and the fabless design companies, the higher the risk for the memory companies that the sexy guys end up doing the hardest stuff," he says. "If that happens, then this moment could prove bittersweet."