Forwards or Backwards

https://forwardorbackwards.com



The Memory Shortage: Why RAM Prices Doubled

Watch on YouTube: https://www.youtube.com/watch?v=Ne3-SVPrcxw

A common memory chip cost about twelve dollars in November 2025. By late September 2026 it cost about forty-six dollars, and the machines built around it - phones, PCs, cars - are getting dearer as a direct result.

This explains why. First, what memory actually is: how a DRAM cell holds a bit as a trace of electrical charge, how a NAND cell holds one without power, and how both combine into the module inside every computer. Then who makes it - Samsung, SK Hynix and Micron, nearly the whole world market between three companies - and the decades of price crashes that thinned the field down to those three.

Then the technology at the centre of the shortage: high-bandwidth memory, HBM, the stacked memory built to sit beside an AI accelerator chip. HBM is a genuine engineering achievement, and it is also, by Micron's own account, about three times as expensive in wafer capacity as ordinary DDR5 memory for the same number of bits - so every wafer a factory gives to HBM is a wafer of ordinary RAM that does not get made. Nvidia and the major cloud computing companies have locked up years of that supply in long-term agreements, which is why device makers are left competing for what is left.

Then who pays: IDC's own 2026 forecasts, smartphone shipments down 16.7% and PC shipments down 11.3%, and the older DDR4 chip - still built into cars and appliances - caught in the same squeeze. Then the new fabs under construction at SK Hynix, Samsung and Micron, and why a fab takes two to three years from its first working cleanroom to meaningful output. And finally, the industry's own history of boom followed by crash - 1988, 2008, 2019, 2023 - and the real risk that today's shortage could flip into tomorrow's glut.

The finished-memory inventory figure used here is an analyst's estimate, from KB Securities, named as such rather than presented as a disclosure from either memory maker.

Educational documentary. Not financial or investment advice.

Chapters

1. The price shock
2. What DRAM and flash are
3. Three companies, one market
4. The boom-and-bust cycle
5. What HBM is
6. Why HBM starves ordinary RAM
7. AI demand and long-term deals
8. Who pays: phones, PCs, cars
9. New fabs and how long they take
10. When does supply catch up?
11. Forwards or backwards?

1. The price shock



Something that used to be cheap has become expensive, in less than a year, and finding out why is the point of what follows.

Ordinary memory chips – the kind inside every phone, laptop, games console and car – have risen sharply in price since late 2025. RAM and SSD prices have not doubled once. They have doubled and then doubled again, and the machines they go into are getting dearer as a direct result.

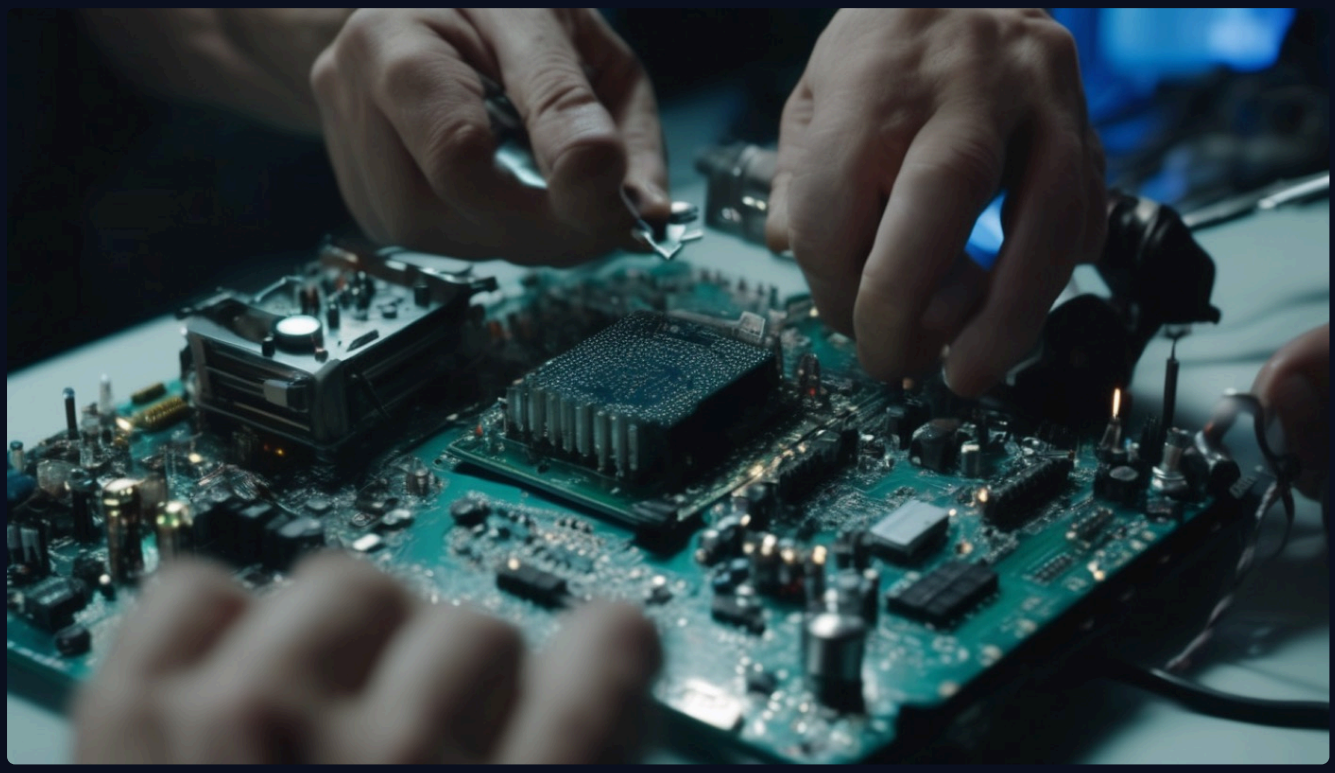
One chip's price makes the pattern plain. A standard DDR4 memory chip, a common type used in ordinary computers, traded at about twelve dollars in November 2025. By late September 2026, the same chip cost about forty-six dollars. That is one specific chip's spot price – the price for buying right now, rather than under a long contract – and it is the clearest single line in this story.

The people who track the memory market for a living do not think this is a temporary spike. An analyst at the South Korean bank KB Securities estimated in September 2026 that Samsung and SK Hynix, the two biggest memory makers, were each holding fewer than ten days of finished memory in stock.

"Beyond a simple recovery in demand, a situation could arise in which the volume available for sale itself is depleted." Those are the analyst's own words, not a company statement. Neither maker publishes its inventory in days of stock, so this is treated here as an estimate rather than a disclosed fact.

So a chip that used to cost about the same as a sandwich now costs about the same as a meal for four, and an analyst who watches this market closely thinks that is a taste of things to come. Why memory, and why now, is what comes next.

2. What DRAM and flash are



To understand why this is happening, start with what memory actually is, because "memory" quietly covers two different things.

The first is DRAM, dynamic random-access memory. It is the memory a device thinks in while it runs: open applications, half-typed messages, everything the processor needs close at hand. A single DRAM cell is almost nothing: one tiny capacitor that stores a bit as a trace of electrical charge, and one transistor that switches it onto a wire so the charge can be read. That charge leaks away within a fraction of a second. So the whole chip refreshes every cell, thousands of times a second, for as long as the power is on. Turn the power off, and DRAM forgets everything instantly.

The second kind is flash memory, usually built as NAND. A NAND cell stores its bit differently, as charge trapped inside an insulated gate, with nothing needed to keep it there once it is set. That is why flash remembers with the power off, which is what makes a phone, an SSD or a memory card work at all.

Millions of these cells are built onto a single memory chip. A handful of chips are then soldered onto a small circuit board with a row of gold contacts, which is the memory module you would recognise from inside a laptop. Every computer, every phone and every car needs both kinds: DRAM to think with, and flash to remember with.

That is why a shortage in either one reaches almost everything with a processor inside it.

3. Three companies, one market





Only three companies make nearly all of the world's DRAM, and that concentration is itself part of the story.

Samsung Electronics, of South Korea, is the largest, followed by SK Hynix, also South Korean, and then Micron Technology, based in the United States. In the first quarter of 2026 those three between them held close to ninety per cent of all DRAM revenue.

Samsung's main memory operation is its Pyeongtaek campus in South Korea, one of the largest chip-manufacturing sites in the world. A plant this size takes years to build and represents a commitment of tens of billions of dollars, which is part of why so few companies have ever managed to enter this business and stay in it.

Getting down to three survivors took decades. Dozens of companies made DRAM through the 1970s and 1980s; price crashes, mergers and exits thinned that number, which is exactly the history covered next. With three makers left, one company's decision moves the whole world's price of memory. Building a new fab, or holding one back, is enough to do it.

4. The boom-and-bust cycle



Steep price rises like this one have happened before, and every time, they have been followed by a crash.

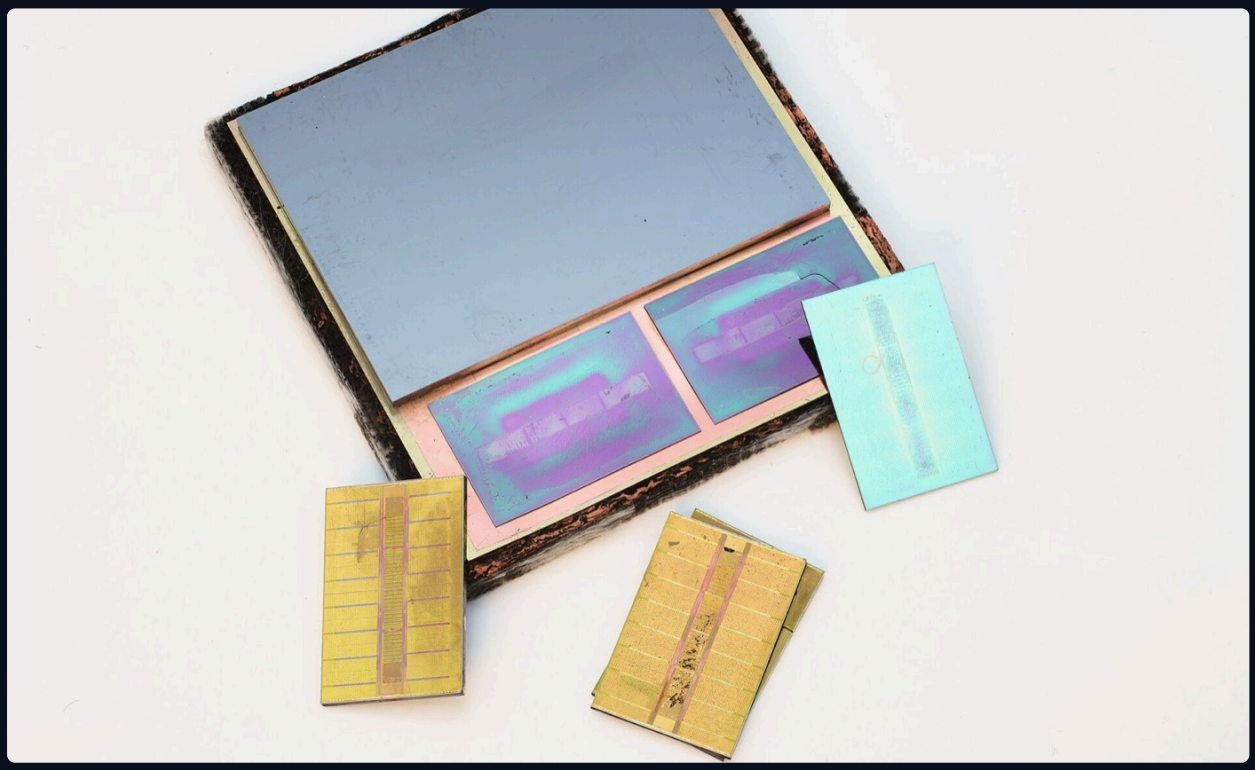
Japanese manufacturers built the DRAM industry through the 1970s and 1980s, and by nineteen eighty-eight they held about ninety per cent of the world market. Trade disputes, a strong currency and new competition then cut that share down over the following decade.

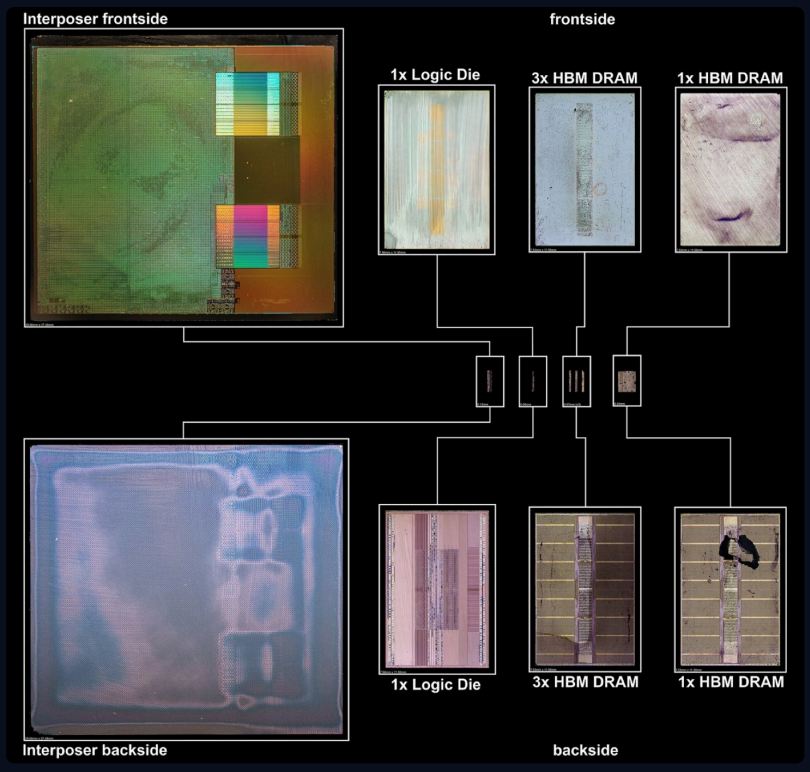
Three more crashes are marked on this timeline. Two thousand eight came first, when the global financial crisis emptied out demand almost overnight. Two thousand nineteen came next, after an earlier boom had brought a wave of new fabs that all switched on at once. And two thousand twenty-three was the most recent, after the surge in pandemic buying of laptops and servers finally ran out.

Every one of those busts followed the same shape: a shortage drives prices up, every maker builds new capacity to catch the high prices, and by the time that capacity switches on, there is too much memory chasing too few buyers. That history is exactly why the three survivors are cautious now. Building a new fab today means betting that today's shortage is still there in three years, when the fab is finally ready — and the last three cycles all ended the same way.

5. What HBM is







Everything so far describes ordinary memory. High-bandwidth memory, or HBM, is a different and much newer way of building it, made for one job: feeding an AI accelerator chip as fast as physically possible.

An ordinary memory module sits on a circuit board, some distance from the processor. HBM instead stacks eight, twelve or sixteen individual memory dies directly on top of each other, joined by thousands of microscopic connections drilled straight down through the silicon, called through-silicon vias. The whole stack then sits immediately beside the processor, on the same small package, rather than across the board.

This is what that stack looks like once it has been taken apart and photographed.

Layer after layer of memory silicon, each one thinned down and drilled through, bonded into a single block only a few millimetres tall. The picture shows an HBM stack sitting beside its logic chip on a graphics processor, with the closeness between them the whole point of the design.

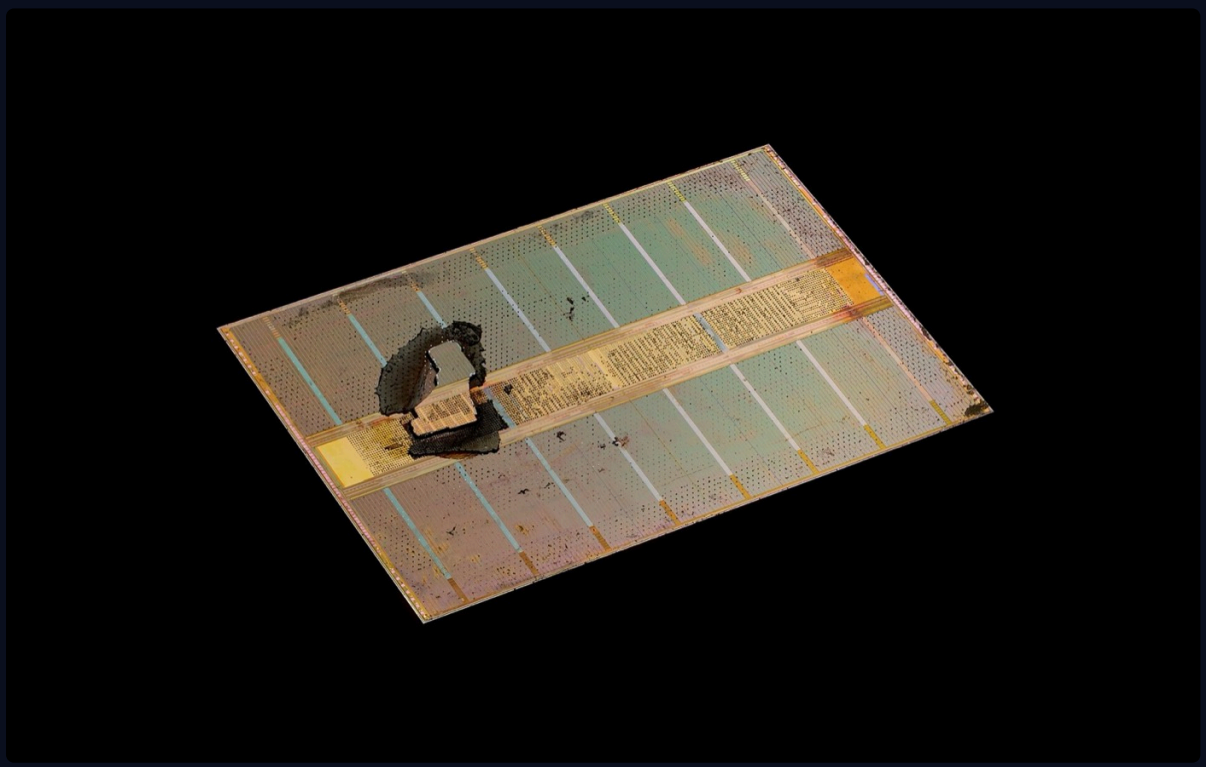
The stack and the processor both sit on a further piece of silicon called an interposer, a wiring plate finer than any circuit board could be. It is what makes that huge number of connections possible at all, in the small space beside the processor.

Here is that interposer next to an ordinary HBM stack, photographed side by side. It is the fine wiring plate that lets thousands of connections cross between the stack and the processor in the space of a few millimetres.

None of that comes cheap, in money or in manufacturing effort, and what it costs the rest of the memory supply is next.

6. Why HBM starves ordinary RAM





Every wafer of silicon a factory processes can become ordinary DDR memory, or it can become HBM, and once it is committed to one, it cannot become the other.

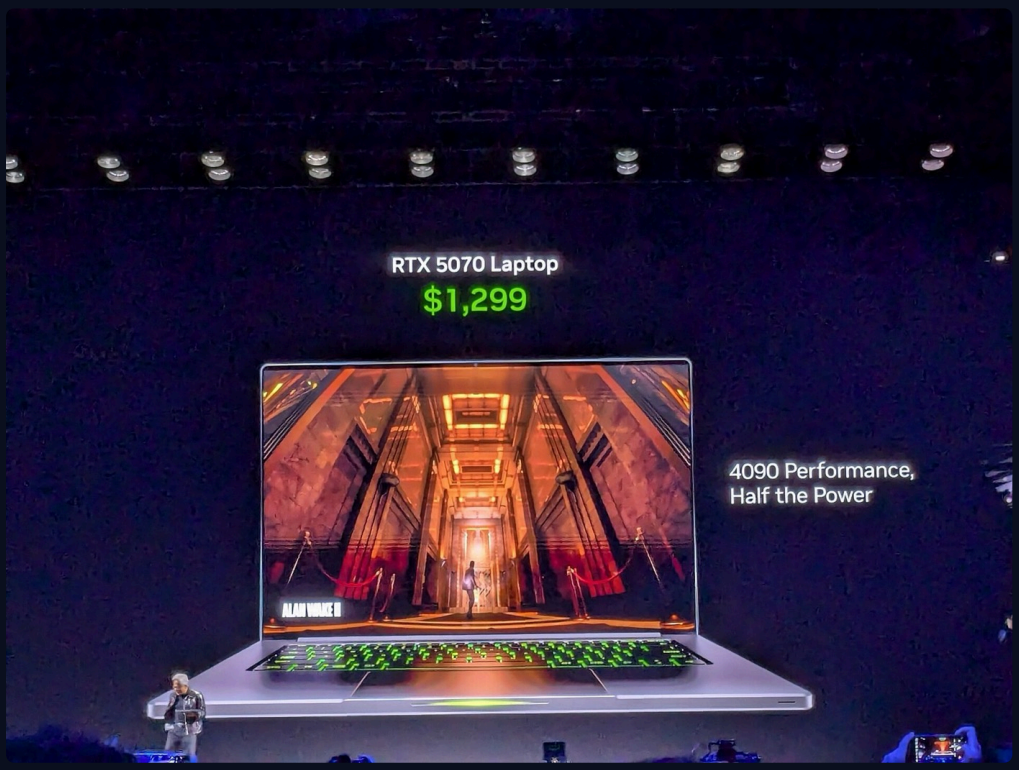
That trade-off is worse than it sounds, because HBM is expensive in a way that goes beyond money. Micron told an industry conference in August 2026 that producing HBM uses roughly three times the wafer capacity of ordinary DDR5 memory, for the same number of bits delivered at the end. Every stack has to be thinned to a fraction of a normal wafer's thickness, drilled through with thousands of vertical connections, and bonded layer by layer. Each of those steps loses some of the silicon before it ever reaches a customer.

So when a factory chooses to build more HBM, it is not adding new capacity — it is converting existing capacity away from ordinary memory. Industry tracker TrendForce estimates that HBM took about eighteen per cent of the top three makers' wafer supply in 2025. By 2027 that share is expected to reach about thirty per cent, while returning a much smaller share of the actual bits produced. Every percentage point that shifts toward HBM is a percentage point of ordinary RAM that does not get made.

Here is a single HBM die, the kind that goes into one of those stacks, photographed close up after being taken apart from a finished chip. Multiply the waste built into making one of these by the number of layers in a stack, and the three-times figure stops sounding abstract.

7. AI demand and long-term deals







The other side of this shortage is demand, and almost all of the new demand is coming from one place: building computers for artificial intelligence.

An AI accelerator, the kind of chip that trains and runs large AI models, needs HBM sitting right beside it to work at all, and it needs a great deal of it. Nvidia is the dominant maker of these accelerators.

Nvidia does not simply buy HBM on the open market. It and the large cloud computing companies that buy its chips have signed long-term supply agreements directly with Samsung, SK Hynix and Micron, locking in output years ahead of when the chips will actually ship.

SK Hynix has said its DRAM, NAND and HBM output is essentially sold out for 2026. More recent reports say 2027 capacity is following the same pattern.

Jensen Huang, seen here at a trade event in January 2025, runs the company that buys more HBM than anyone else in the world. Memory makers now plan years of factory output around orders like his, before a single ordinary computer maker gets a turn.

That is what a long-term agreement does: it takes memory off the table before the shortage even shows up in a shop. The servers those agreements fill look like the one in this next picture.

This is a technician working on exactly this kind of server rack, at a US national laboratory's computing centre. It stands for a growing share of the world's memory supply, ahead of the phone or laptop still waiting on a shelf.

8. Who pays: phones, PCs, cars



All of this eventually reaches a shop shelf, in the price of the finished device rather than the price of the chip inside it.

Market researcher IDC now forecasts that global smartphone shipments will fall by sixteen point seven per cent in 2026. That is a record decline, one IDC has revised upward twice this year as the shortage worsened. Personal computers are hit almost as hard: the same researcher expects PC shipments down eleven point three per cent for the year, with the worst quarter still to come.

Both of those figures come from the same market researcher, and both point the same way: fewer phones, fewer PCs, and higher prices for the ones that do get made. The chart shows both declines side by side, because the point is not simply that shipments fall — it is that both fall together, driven by the same shortage in the same materials.

Phones and PCs use the newest kind of memory, so they absorb the price rise directly and immediately. Cars and household appliances are different: many of them are still built around the older DDR4 chip from the opening chapter, because a car's electronics are certified years in advance and cannot simply switch to whatever memory is cheapest this quarter.

That is why an older, supposedly obsolete memory chip has become one of the hardest things for a carmaker to secure: it cannot be substituted, and the newest fabs are not being built to make more of it. The chip nobody expected to matter has become one everybody needs.

9. New fabs and how long they take





If this is mostly a supply problem, then new factories should fix it, and they will, eventually. The question is how long "eventually" takes.

A new fab is not simply a big building. It has to be built to an exactness measured in nanometres. It is filled with tools that themselves take more than a year to manufacture and install. And every one of those tools has to be proven, machine by machine, before the fab can be trusted with a single working chip.

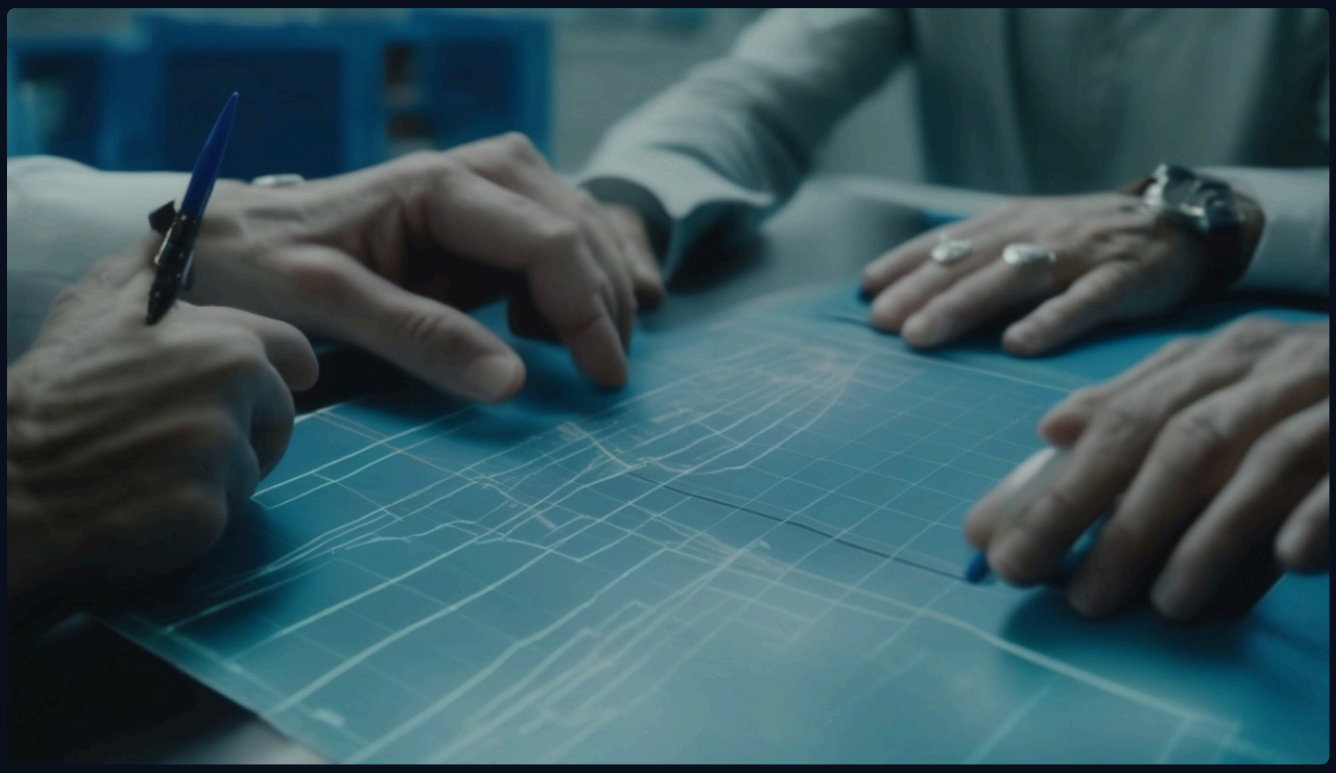
Four sites carry most of this new capacity. SK Hynix is expanding at Cheongju, where a fab called M15X is being built, and at Yongin, a much larger site still under construction. Samsung's newest line, called P5, is at Pyeongtaek, the same site in the next photograph. And Micron's main US HBM production sits in Boise, Idaho.

This is inside Samsung's Pyeongtaek campus, the same site that P5 is being added to: one of the largest single chip-manufacturing complexes anywhere, already running, with more being built beside it.

The timing tells the real story. SK Hynix's M15X fab reached its first working cleanroom in 2026 and is aimed at full production by 2027. Its much bigger Yongin site is following a similar path, with its first phase also targeted for 2027, and later phases stretching out toward the end of the decade. Samsung's P5 finished its cleanroom construction in 2026 too. Reports disagree on whether its meaningful output starts in 2027 or in 2028.

So from a fab's first working cleanroom to the point where it is actually adding meaningful supply typically takes two to three years, which is why nothing announced this year will touch the shortage happening right now.

10. When does supply catch up?



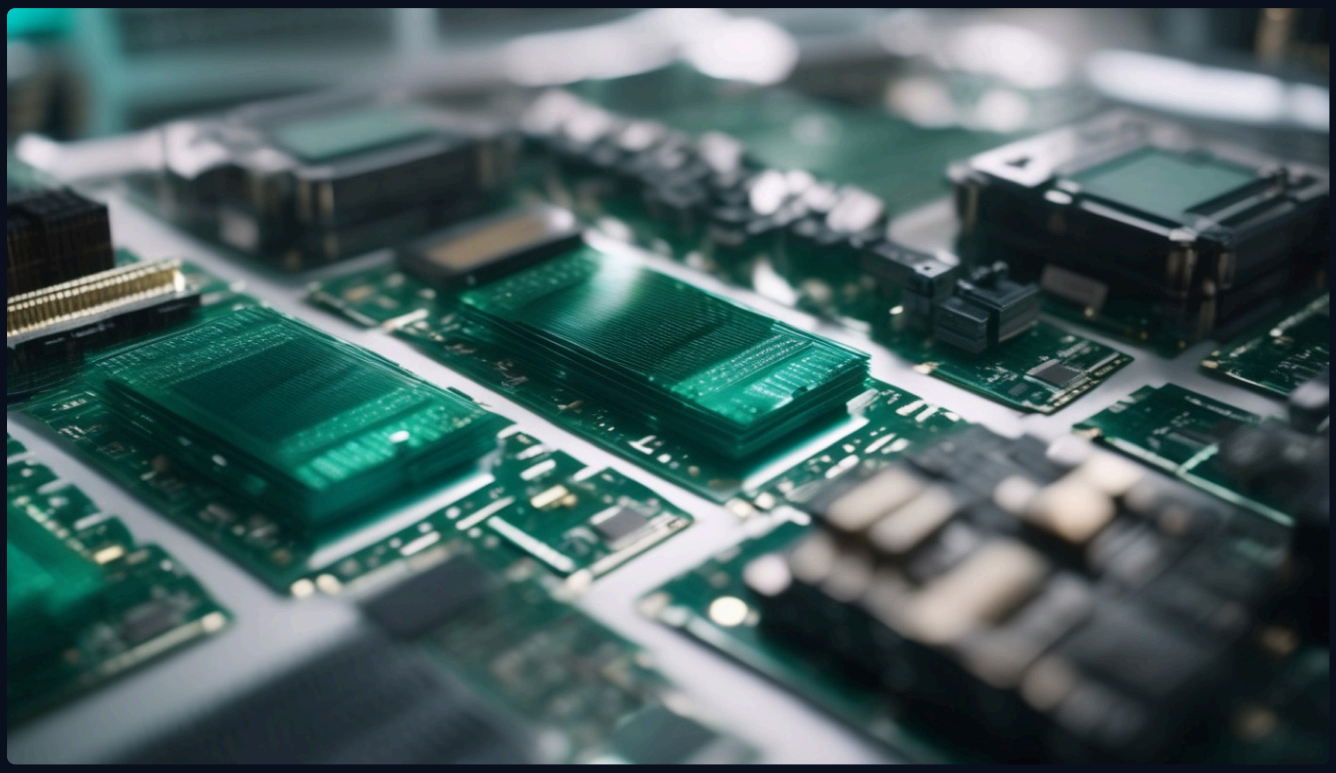
So when does this actually end?

TrendForce, the market research firm behind most of the figures here, forecast in July 2026 that conventional DRAM contract prices would keep rising. The forecast: a further thirteen to eighteen per cent in the third quarter of 2026. That is still a rise, but a much slower one than the near-doubling seen earlier in the year, because long-term supply agreements already lock in a large share of what the fabs produce.

Longer term, the new fabs are the only real answer, and none of them is expected to add serious volume before 2027 at the earliest. Until then, the shortage does not ease — it simply gets managed, chip by chip, contract by contract.

There is a genuine risk sitting on the other side of all this. If AI spending slows even a little, while all these new fabs are still coming on line, the industry could tip from a shortage into the kind of oversupply that has crashed prices before. That kind of crash has happened more than once, and each time the industry did not see it coming until prices were already falling.

11. Forwards or backwards?



So: forwards, or backwards?

Forwards, without question, for the technology feeding artificial intelligence. HBM is a genuine engineering achievement, and the money now flowing into new fabs will eventually make far more memory of every kind than exists today.

But backwards, right now, for almost everyone else. A phone, a laptop or a car does not care that the shortage has a good reason behind it — it only sees a higher price, or a wait. The chip that used to be the invisible, boring half of every computer has become the thing deciding how much that computer costs.

Four figures carry the whole story. A common memory chip's price rising from about twelve dollars to about forty-six dollars, in under a year. HBM costing roughly three times the ordinary wafer supply for the same number of bits. And two device markets, phones and PCs, both being cut by double digits because of it.

None of that is permanent. The new fabs will eventually catch up, the way they always have. When they do, today's shortage will look like one more turn of a forty-year cycle. Until then, the honest answer is that memory is moving forwards and backwards at the same time, depending on which machine you are asking about.

Sources and credits

Primary sources

- TrendForce, 'Memory Spot Price Update' releases, 12 Nov 2025 through 24 Sep 2026 - the DDR4 1Gx8 3200MT/s spot-price series ($12.18 to $46.11).
- TrendForce, 'Rapid Contract Price Surge Drives 1Q26 DRAM Industry Up 81% QoQ', 1 Jun 2026 - the 93-98% QoQ contract-price rise, $97bn industry revenue and the Samsung/SK Hynix/Micron revenue split.
- TrendForce, 'Long-Term Agreements Cap Price Increases; Server DRAM Contract Prices Expected to Rise 13-18% QoQ in 3Q26', 9 Jul 2026.
- TrendForce, 'AI Reportedly to Consume 20% of Global DRAM Wafer Capacity in 2026', 26 Dec 2025, and 2026 follow-ups - HBM's 18/22/30% share of top-three wafer input and 8/9/13% share of bit output, 2025-2027.
- Digitimes, 'HBM4 squeezes DRAM supply as Samsung, SK Hynix inventories fall below 10 days', 7 Sep 2026; Tech Times, 'Memory Runs Dry', 7 Sep 2026; Seoul Economic Daily, 7 Sep 2026 - the KB Securities analyst estimate and Kim Dong-won's quoted remarks.
- Micron Technology, remarks at Hot Chips 2026 (Aug 2026), as reported by partgenie.ai and MLQ News - HBM's roughly 3x wafer capacity per bit against DDR5.
- IDC, 'Smartphone Shipments Set for Record 16.7% Drop in 2026', 26 Aug 2026.
- IDC, 'PC Market Enters Volatile Territory as Memory Shortage Persists Through 2027', 2 Jun 2026 - PC shipments down 11.3% in 2026.
- TrendForce and The Elec/Korea Times reporting, 2026 - SK Hynix M15X and Yongin fab timelines, Samsung P5 timeline.
- JEDEC Solid State Technology Association, HBM4 standard JESD270-4, Apr 2025.
- U.S. Congressional Office of Technology Assessment report on the DRAM industry (1990) and general industry-history accounts - Japan's ~90% DRAM share in 1988.

Photo credits (Wikimedia Commons)

- Samsung Electronics Pyeongtaek Campus semiconductor plant: Office of the President of the United States, Public domain - https://commons.wikimedia.org/wiki/File%3APresident_Biden_and_Yoon_Suk-yeol_at_the_Samsung_Electronics_Pyeongtaek_Campus.jpg
- Samsung Electronics Pyeongtaek Campus, inside the fab: Office of the President of the United States, Public domain - https://commons.wikimedia.org/wiki/File%3APresident_Biden_visited_the_Samsung_Electronics_Pyeongtaek_Campus_%281%29.jpg
- HBM stack with logic die on AMD Vega 10 (photo): Fritzchens Fritz, CC0 - https://commons.wikimedia.org/wiki/File%3AAMD%4014nm%40GCN_5th_gen%40Vega10%40Radeon_RX_Vega_64%40ES-Sample%40_Stack-DSC01160-DSC01176_-_ZS-retouched.jpg
- HBM stack and GPU interposer comparison: Fritzchens Fritz, CC0 - https://commons.wikimedia.org/wiki/File%3AAMD%4014nm%40GCN_5th_gen%40Vega10%40Radeon_RX_Vega_64%40ES-Sample%40_DSCx5_Interposer_HBM_comparison.jpg
- HBM DRAM die, close-up photo: FritzchensFritz, CC0 - https://commons.wikimedia.org/wiki/File%3AAMD%4014nm%40GCN_5th_gen%40Vega10%40Radeon_RX_Vega_64%40HBM_DRAM_Die%40_Stack-DSC08838-DSC08973_-_ZS-retouched.jpg
- Jensen Huang, Nvidia CEO, CES 2025: Pronoia, CC0 - https://commons.wikimedia.org/wiki/File%3AJensen_Huang_-_RTX_5070_Laptop_-_Nvidia_Keynote_-

_CES_2025_Las_Vegas.jpg

- Data centre server rack (technician at NERSC): Derrick Coetzee from Berkeley, CA, USA, CC0 - https://commons.wikimedia.org/wiki/File%3ATechnician_with_laptop_working_on_server_rack_at_NERSC.jpg

Not regulated financial advice.