# HBM is DRAM Rotated 90 Degrees

The chip that does the math is cheaper than the chips that feed it. Memory is 40-50% of the AI module cost.

Millimeters, not inches.

8 to 16 Layers of ordinary leaky-capacitor DRAM stacked vertically.

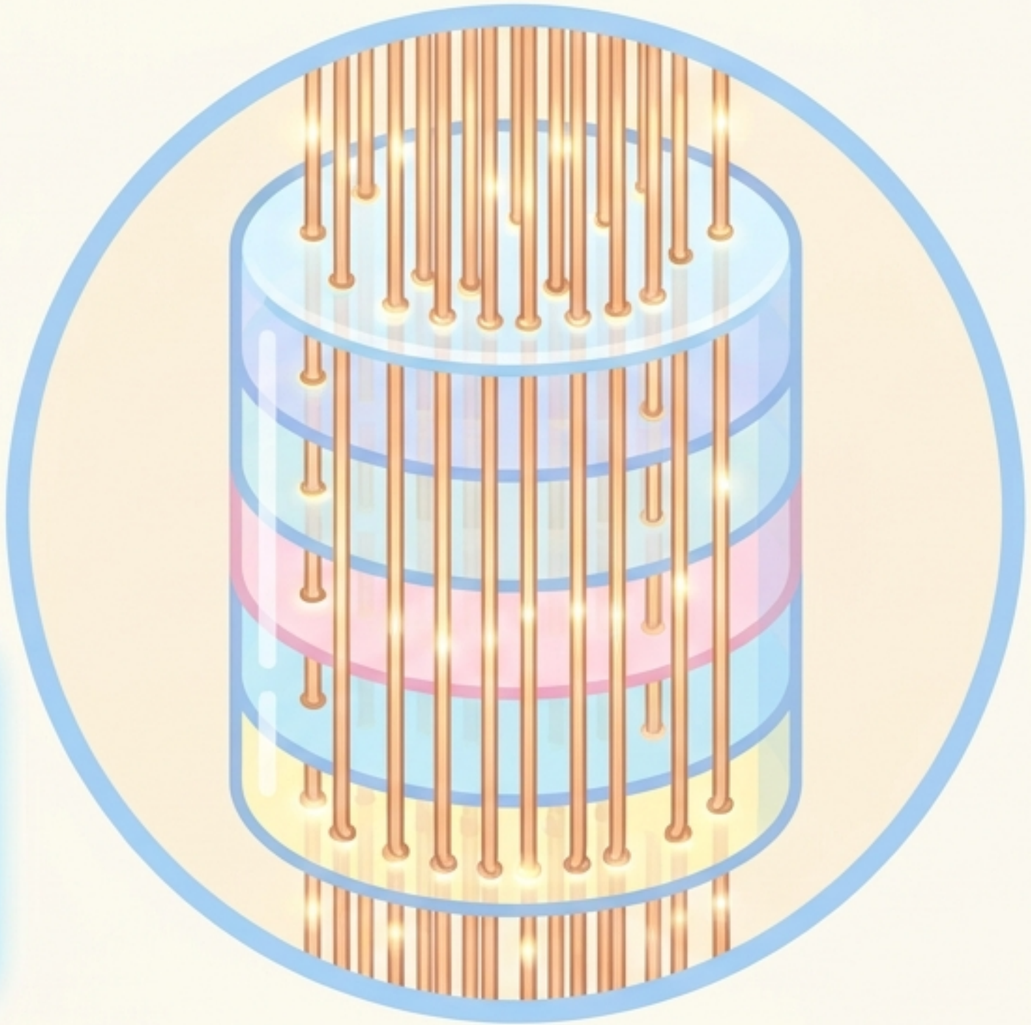
Logic Die (The Brains & Controller)

The HBM Tower

Standard DRAM lives flat. HBM stacks vertically to defeat the memory wall. Wide and close beats narrow and fast.

# Through-Silicon Vias (TSVs): The 1024-Bit Firehose

~1.2 TB/s peak bandwidth per stack (HBM3E)



### Traditional DDR5

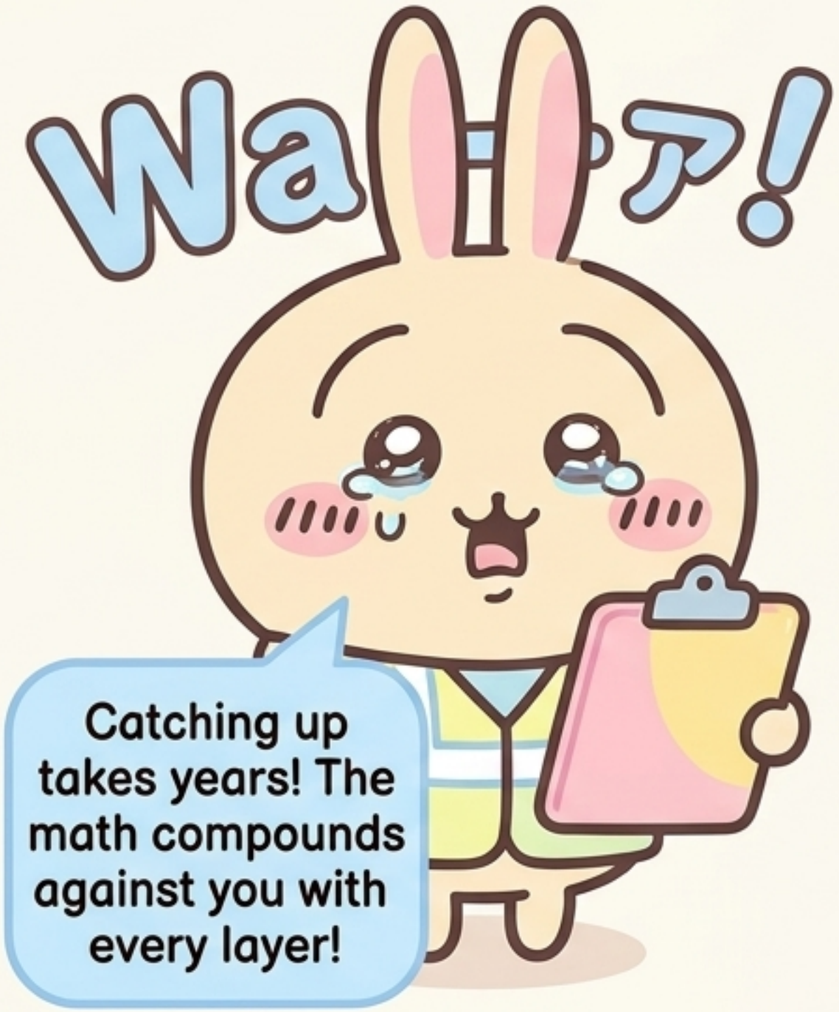
- 64-bit wide channel
- Long copper trace
- High frequency
- High picojoule-per-bit power burn

### HBM Stack

- 1,024-bit wide interface
- Microscopic vertical vias
- Low frequency
- Massive bandwidth-per-watt

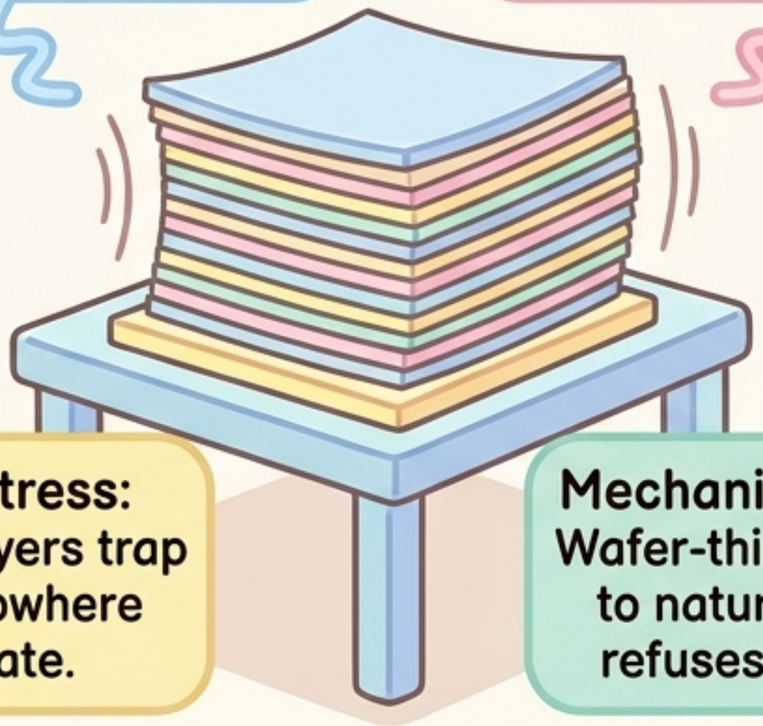
Moving data vertically saves the massive energy required to drive signals across circuit boards, keeping 100-kilowatt AI data center racks from melting.

# The Moat is Yield, Not Secrets




**Void-Free TSVs:**
Thousands of microscopic etched holes must align perfectly across 16 layers.

**Known-Good-Dies:**
One bad memory die in a 16-Hi stack scraps the entire multi-thousand-dollar tower.



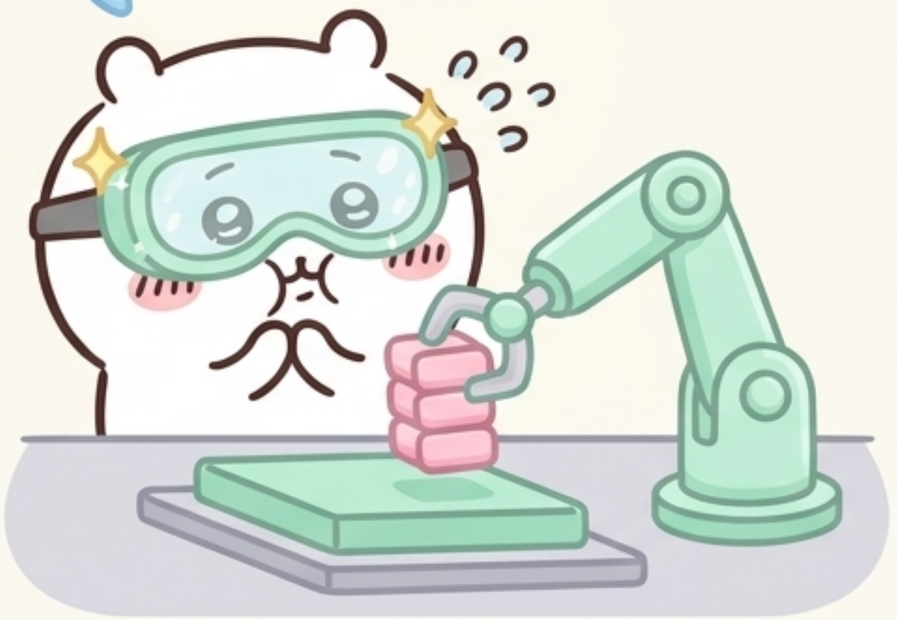
**Thermal Stress:**
Deep inner layers trap heat with nowhere to dissipate.

**Mechanical Warpage:**
Wafer-thin silicon wants to naturally bow and refuses to bond flat.

**Anyone can read the spec. Almost no one can build it at acceptable yield in volume. That is why only three companies control the entire global market.**

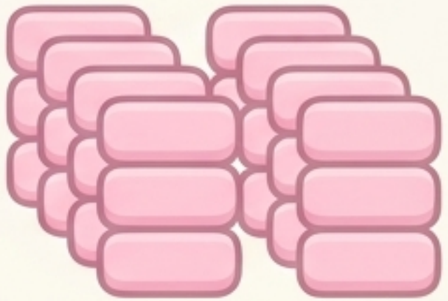
# Hostage to CoWoS Advanced Packaging

No CoWoS, no AI GPU.
The chip design is no longer the moat; the packaging is.



[GPU Logic Die]

+



[8x HBM Towers]

bonded onto

[Silicon Interposer (Ultra-fine wiring slab)]

=

**TSMC CoWoS Assembly**

HBM content roughly doubles every two years (80GB in 2022 → 288GB on the B300 in 2025). The interposer is the only way to route the 8,192+ wires between them.

The GPU shortage is rarely the GPU. It is the memory and the microscopic packaging that hold them together. Demand outpaces TSMC's rapidly doubling CoWoS capacity.

# The $1 Trillion Repricing

When the cheapest input becomes the scarcest, unclonable component.

**SK Hynix:** >50% market share. First to HBM3E. Primary Nvidia supplier.

**Samsung:** The integrated giant playing the in-house foundry card for HBM4.

**Micron:** The lone US maker ramping 12-Hi with reshoring tailwinds.

**The HBM4 Shift:** In HBM4, the base logic die becomes a custom TSMC foundry chip. HBM shifts from a memory commodity to a highly profitable semi-custom logic co-design.

Solving extreme physical constraints—like trapping heat in a 12-story silicon sandwich—turned a low-margin boom-bust commodity into a trillion-dollar bottleneck that decides who wins AI.