Episode Book

EPISODE #6679

Why 3D DRAM Beats Every RAM Replacement

PUBLISHED	19 November 2026 at 12:44 UTC
DURATION	21:15
LISTEN	myweirdprompts.com/episode/dram-replacements-3d-memory-02189/
HOSTS	Corn, Herman, Dorothy, Hilbert
SCRIPT MODEL	z-ai/glm-5.3
VOICES	omnivoice
PIPELINE	V5.36
BOOK GENERATED	19 November 2026 at 13:34 UTC

SYNOPSIS

With AI demand driving a prolonged DRAM shortage, engineers have spent twenty years trying to replace working memory with persistent alternatives. From Intel Optane's $559 million write-off to niche shipping chips like STT-MRAM and ReRAM, nothing matches DRAM on speed, density, and cost at once. Even promising lab contenders—like Micron's ferroelectric NVDRAM, imec's capacitorless IGZO cells, and UltraRAM—remain years away, leaving vertical 3D DRAM as the most realistic path forward.

TRANSCRIPT



Corn

The RAM shortage conversation isn't going away, so Daniel went back to it with the radical version of the question. We've done why fast SSDs still can't stand in for RAM, and we've done the shortage itself. His new angle is: stop competing for the stuff and replace it. He admits it sounds crazy, quote, "a bit crazy, but it's worth considering," and asks whether anyone has seriously tried to engineer a fundamentally different technology for fast working memory, and whether any of it is close enough to matter.



Herman

And "close enough to matter" is the part I like, because there's a real answer and it's mostly no.



Corn

How do you mean, mostly?



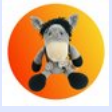
Herman

Well, there are things that are close on one axis. Something will be nearly as fast but way too low-density. Something will be dense enough but too slow. The "mostly no" is because nothing is close on all the axes at once, and the axes multiply together. You can't trade your way out of failing one of them.



Corn

Let's take the crazy seriously anyway. What have people actually tried?



Herman

So the framing is: DRAM is the working memory of basically every computer, AI demand has made it scarce and expensive, and the obvious response is to find something else that does the same job. Fast, dense, cheap, rewritable essentially forever. People have been chasing that for about twenty years, and the pattern is consistent. The candidates are mostly persistent memories, which sound great on paper. Your files survive a power cut. But every one of them trades away density, cost, or endurance to get that persistence, and none of them has displaced DRAM at volume.



Corn

So the non-volatility is the seduction and the density is the bill.



Herman

That's a good compression of it, actually. And there are really two live routes out of this. One is reinventing DRAM itself, going vertical, going three-dimensional, the way NAND did years ago. The other is architectural: use less DRAM, tier it, move more work onto on-chip memory. Both of those are more real than any replacement technology.



Corn

Before we get into the tombstones, give me the timing. Daniel's question is "is any of it close enough to matter." Close to what?



Herman

The shortage is forecast to last until about twenty thirty. SK hynix's CEO has reportedly said as much. New DRAM fabs are scheduled for between twenty twenty-seven and twenty thirty. And nothing in the labs reaches consumer products before roughly twenty twenty-eight at the earliest, more like twenty thirty-one for several of these. So the honest headline is that even the successful options don't help this shortage cycle.



Corn

The medicine arrives after the patient has recovered on its own.



Herman

In this case, yes. Doesn't mean the medicine isn't real. Let's start with the cautionary tale, because every serious discussion of this starts and ends with Intel Optane.



Corn

The one everyone bought zero of.



Herman

Intel wound the Optane memory business down in the second quarter of twenty twenty-two, and the filing that announced it included a five hundred and fifty-nine million dollar inventory impairment. Just to write off the chips sitting in warehouses. Analyst estimates of the ongoing losses run from around six hundred million to over two billion a year, and those are estimates, not Intel's numbers.



Corn

Half a billion just for the privilege of stopping.



Herman

Intel's filing doesn't give a reason, so this part is press and analyst reporting. Optane was 3D XPoint, and it reportedly landed in a gap between the two incumbents. Slower than DRAM, more expensive than NAND flash. Not fast enough to be your working memory, not cheap enough to be your storage. And because it never reached volume, the cost reportedly never came down. It also reportedly lost its manufacturing base when Micron, the co-developer, sold the fab, and Pat Gelsinger is reported to have pointed to a shift toward CXL.



Corn

So the lesson isn't that the physics failed. It's that "in between" is a place in the market that doesn't exist.



Herman

On the performance axes, no. There's no customer whose requirements are exactly halfway between RAM and storage. You buy RAM because it's fast, you buy storage because it's cheap, and anything in the middle has to justify itself to both crowds and convinces neither.



Corn

Give me a concrete picture of that, because I think people imagine there must be someone in the middle.



Herman

Think about what a middle product has to promise. To the RAM buyer, it says "I'm almost as fast", and almost is doing fatal work in that sentence. To the storage buyer, it says "I'm cheaper than RAM per bit", but not cheaper than flash. Both of them already have an answer, and the middle product is second best to each.



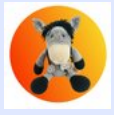
Herman

Some escape it by not trying. STT-MRAM, spin-transfer torque, is actually shipping, which already puts it ahead of most of this list. Everspin makes a one-gigabit part on a twenty-eight-nanometer process, DDR4-style interface, about two point seven gigabytes per second. It's fast. But the density is roughly nine point seven five megabits per square millimetre, which by my own arithmetic is tens of times below modern DRAM, so it can't be your working memory no matter how quick it is.



Corn

What's it for, then?



Herman

Persistent write buffers. Places where you need something that behaves like a small cache but remembers through a power cut. IBM uses it in one of their FlashCore storage modules. It's found a niche, and good for it, but a niche is what it is.



Corn

One gigabit in a world where the shortage is measured in gigabytes per GPU.



Herman

Right. And the other persistent technology making actual news is ReRAM, resistive RAM, but I want to be careful with it because it gets oversold. Weebit Nano licensed theirs to Texas Instruments in January of this year, and the company claims it adds about five percent to wafer cost where embedded flash, the thing it replaces, adds twenty percent or more. If that holds, it's a genuine win.



Corn

For what job, though?



Herman

Microcontrollers. Code storage on a chip. It replaces the little flash block next to the processor on embedded parts, it is not working memory, and nobody at Weebit claims it is. It's solving the cost of a completely different problem.



Corn

So in the shipping-today column we have a buffer and a microcontroller part. Daniel asked about replacing RAM. We're zero for two.



Herman

The interesting column is the lab one, though, and this is where I get excited, because the ferroelectric work is the closest anyone has publicly shown to non-volatile memory at DRAM-class density. Micron presented a paper at IEDM in December twenty twenty-three on what they called NVDRAM. Thirty-two gigabit, dual layer, three-dimensional stacked, ferroelectric capacitors at five point seven nanometres. Density above their own one-beta planar DRAM. That's non-volatile memory at DRAM-class density, on paper.



Corn

And it's a paper, not a product.



Herman

No product announced, full paper not even public, and analysts split on whether it ever becomes one. But it shows it can be built in a lab. And the ecosystem is moving. There's a Dresden outfit, the Ferroelectric Memory Company, working on hafnium-oxide ferroelectric memory, DRAM-class and SRAM-class. They raised a hundred million euros in November last year, and SK hynix is an investor.



Corn

The incumbent DRAM people funding the DRAM killer.



Herman

Which suggests they see it as a possible successor rather than a rival. imec showed a low-voltage ferroelectric capacitor with NAND-style vertical stacking at VLSI this year as a proof of concept, with talk of mainstream adoption over the next decade. That's the honest horizon. Years, not product cycles.



Corn

There's a second lab thread, the one I find sneakier, because it doesn't replace DRAM at all. Tell me about the capacitorless cell.



Herman

Ah, yes. This is my favourite part of the whole landscape. A DRAM cell today is one transistor and one capacitor. The capacitor is the awkward part. It's a tiny structure that leaks, which is why DRAM needs to refresh every sixty-four milliseconds. The whole refresh machinery, the constant rewriting, exists to keep charge in that capacitor. Capacitorless DRAM, the IGZO two-transistor cell imec has been working on, drops the capacitor entirely. The transistor itself holds the charge.



Corn

And the number I know you love is the retention.



Herman

I do love it, I'll admit that. In twenty twenty, the best published retention was around four hundred seconds. By twenty twenty-one, over a thousand seconds, with endurance above ten to the eleventh cycles and writes under ten nanoseconds. By twenty twenty-three, over four and a half hours. Against sixty-four milliseconds for conventional DRAM.



Corn

From milliseconds to hours in three years of papers.



Herman

That trajectory is the reason Yole put it on the long-term DRAM roadmap in twenty twenty-four. The catch, and there's always a catch, is hydrogen-related reliability, IGZO is touchy about hydrogen, and there's no commercial date. Nobody has committed to a product.



Corn

I notice you called it capacitorless DRAM, not a RAM replacement.



Herman

Because it isn't one. It's still a DRAM-style cell. It still needs the same fabs, the same manufacturing philosophy, probably the same companies. It's DRAM with a better-behaved cell, which is exactly the kind of revolution this industry will actually take.



Corn

Before we go further, I believe we have a call from your mother.



Herman

Mum! We haven't heard from her in weeks.

D

Dorothy

Hello dear, it's your mother. I've got my History of Toast fact for you. The first pop-up toaster for the home, the Toastmaster, went on sale in nineteen twenty-six. You'd think that would be a happy thing, but it mostly reminds me that things don't work out, dear, they just pop up and then they sit there. Anyway. Wear a cardigan, it's chilly. Ring me back when you can.



Herman

She's been on a real streak with these lately.



Corn

She's the only person covering the toaster beat with any rigour.



Corn

Two more from the lab file, then. NEO Semiconductor's 3D X-DRAM.



Herman

Proof of concept in April this year, built with 3D NAND tools, which is clever because it reuses an installed manufacturing base rather than inventing a new one. Company claims: latency under ten nanoseconds, endurance above ten to the fourteenth cycles, retention over a second at eighty-five degrees. All company figures, no production chip, and I'd want independent numbers before believing any of them.



Corn

And UltraRAM.



Herman

Quinas Technology, a Lancaster University spinout, with a compound-semiconductor approach. They finished a one point one million pound industrialisation project with IQE last year and are looking for foundry partners for a pilot line. April this year, work toward crossbar-array prototypes. No independent endurance or retention numbers I've seen. Promising physics, very early.



Corn

So that's the landscape. The graveyard has Optane in it. The shipping column has a niche buffer and a microcontroller part. The lab column has ferroelectrics, capacitorless DRAM, and two startups. Daniel's question was whether anyone has seriously explored alternatives. Answer: yes, for twenty years, seriously, with billions of dollars.



Herman

And none of it close. Now, the part I think actually matters for the shortage is neither column, and I want to hear your read on it. What's the realistic revolution?



Corn

I think it keeps the capacitor. 3D DRAM.



Herman

That's where I've landed too, and I don't land there happily, because it's the least exciting sentence in this whole field. Samsung showed a sixteen-gigabit 4F2 DRAM with vertical channel transistors and wafer-to-wafer hybrid bonding at ISSCC in February. Press reports say qualification in twenty twenty-seven, mass production twenty twenty-eight. That's press reporting, Samsung hasn't committed.



Corn

And the other two?



Herman

Micron reportedly told investors last month they'll skip the 4F2 layout and extend their existing design before going 3D. SK hynix timing gets reported differently depending on the outlet. So the three DRAM makers are not even marching in step on the thing that's supposed to save them.



Corn

But here's the thing I want to say about 3D DRAM, and stop me if you disagree. It doesn't break the dependence at all. You still need DRAM-class fab capacity. You still need new tooling. You've made the same chips in the same buildings by the same three companies. It packs more bits into the same footprint, which is real, but it's evolution, not escape.



Herman

No, I agree completely, and that's actually the point. The industry has looked at the alternatives for twenty years and decided the safest bet is DRAM itself, restructured. Everything else is either dead, niche, or a decade out. The capacitor is sixty-year-old physics and it keeps winning on the combination of speed, density, and cost. Nothing beats it on all three at once. There are things that beat it on one.



Corn

You said earlier there were two routes. What's the second one, the one that doesn't invent a new memory?



Herman

Use less of it. And this is the one that's actually live right now. CXL, Compute Express Link. It's a protocol for tiering memory. You put a slower, cheaper pool of memory behind the fast DRAM, and the system moves data between them on demand. The clever version is what Meta did, by their own account. They took retired DDR4, memory that had come out of old servers and would otherwise be scrapped, and run it behind CXL on their own silicon. They reported up to twenty-five percent fewer inference servers for the same workload.



Corn

They turned their own e-waste into capacity.



Herman

Twenty-five percent fewer servers is not a rounding error. And Marvell has a CXL part claiming two to two and a half times effective capacity per server, which is a vendor figure and should be treated as such.



Corn

And how much of the world is actually doing this today?



Herman

Essentially none. Yole's estimate is near zero percent of servers today, thirteen percent by twenty thirty. That's an analyst estimate.



Corn

Thirteen percent in four years, from zero. So the realistic architectural solution is also slow, and also, presumably, CXL parts themselves need DRAM to sit behind.



Herman

They do. The pool behind CXL is usually regular DRAM. So you haven't replaced it, you've made the same DRAM go further. That's the honest summary of the whole architecture column.



Corn

Which is actually the answer to Daniel's question, isn't it. Not can we replace DRAM, but can we make less DRAM feel like more.



Herman

And the other branch of that is the Groq-style approach. Nvidia licensed Groq's technology in December last year for a reported twenty billion dollars, and Groq's chips do inference with on-chip SRAM. Huge amounts of it for its size, so the working data sits in fast memory on the chip rather than in DRAM or HBM.



Corn

Let me push on that one, because it gets reported as "the end of the DRAM bottleneck."



Herman

And it isn't. SRAM is far less dense than DRAM. Six transistors per bit versus one transistor and a capacitor. You can fit enormously less of it in the same silicon. So a Groq-style chip doesn't remove the need for memory, it swaps which memory you're short of. You've moved the scarcity from the DRAM supply chain to the fab's ability to pour



Herman (continued)

silicon into SRAM arrays. It's a real trade with real wins, but it's not an escape from scarcity, it's a different scarcity.



Corn

So the answer to whether any of this relieves the shortage is, in the near term, plainly no. The alternatives are niche, dead, or lab-stage with no product dates. The realistic relief is using less DRAM and stacking DRAM higher, and both of those still run through the same fabs the labs are competing for.



Herman

That's where I am. The shortage persists. But the engineering response is real, it's incremental, and it's coming from inside the existing supply chain rather than around it.



Corn

Twenty years of trying to replace it, and the best anyone's managed is to use it more cleverly.



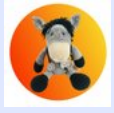
Corn

Hilbert's been listening from the back and looks like he has a view.



Hilbert

I do. You two kept talking about physics and cost. My view is that there's a third thing upstream of both, which is qualification. A new memory type means every customer has to validate it before they'll put it in a product, and that takes years of reliability data. As far as I can tell, that's a big part of why nothing in the middle ever gets a chance.



Herman

That's a fair point I didn't give enough weight to. It's a view rather than something in our research, but it fits: the qualification burden, not just the physics, plausibly keeps new memory types out of servers.



Corn

And it favours the DRAM-in-disguise options, doesn't it.



Hilbert

It does. The ferroelectric and capacitorless work get partway round it by being close cousins of DRAM. Whether the hydrogen problem in the oxide cells validates quickly is the bit I'd want answered.



Herman

On the plus side, no capacitor, no problem.



Hilbert

There's always a capacitor somewhere.



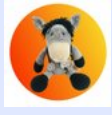
Herman

So if the shortage runs to twenty thirty, what's the consumer-facing consequence of all this? I'm looking at this and thinking, ordinary buyers shouldn't expect relief from a new memory type this cycle, because the architectural fixes, CXL tiering and SRAM-heavy inference chips, are datacenter plays.



Corn

And by the time the ferroelectric or capacitorless stuff actually ships, the shortage will be a story people tell at parties.



Herman

Which doesn't make it pointless work, it just means it's for the next shortage, not this one.



Corn

The real revolution, it turns out, isn't a new kind of memory. It's the boring discipline of needing less of the old kind.



Herman

If you want the groundwork, go back to "Why Fast SSDs Still Need Slow RAM" and "When AI Eats the World's Memory". This has been My Weird Prompts.



Corn

If you like the show, leave us a review. If you've got a weird prompt, send it in.



Herman

We'll be back soon. See you then.