NVIDIA & Broadcom CPO, HBM4 & LPDDR6, TSMC Active LSI, Logic-Based SRAM, UCIe-S and More

ISSCC 2026 Roundup

APR 15, 2026
https://newsletter.semianalysis.com/p/isscc-2026-nvidia-and-broadcom-cpo

ISSCC 2026: NVIDIA & Broadcom CPO, HBM4 & LPDDR6, TSMC Active LSI, Logic-Based SRAM, UCIe-S and More

ISSCC 2026 Roundup

There are three major semiconductor conferences each year, IEDM, VLSI and finally ISSCC. We have covered the former two in great detail over the past few years. Today, we finally complete the trinity with our roundup on ISSCC 2026.

Compared to IEDM and VLSI, ISSCC has a much bigger focus on integration and circuits. Almost every paper comes with some form of circuit diagram, together with clear measurements and data.

In past years, ISSCC findings have been hit or miss when it comes to industry impact. This year was different, a significant number of papers and presentations were directly relevant to market trends. Topics covered range from the latest advancements in HBM4, LPDDR6, GDDR7, and NAND, to co-packaged optics, advanced die-to-die interfaces, and advanced processors from the likes of MediaTek, AMD, Nvidia, and Microsoft.

In this roundup, we will cover major categories such as Memory, Optical Networking, High-Speed Electrical Interconnects, Processors.

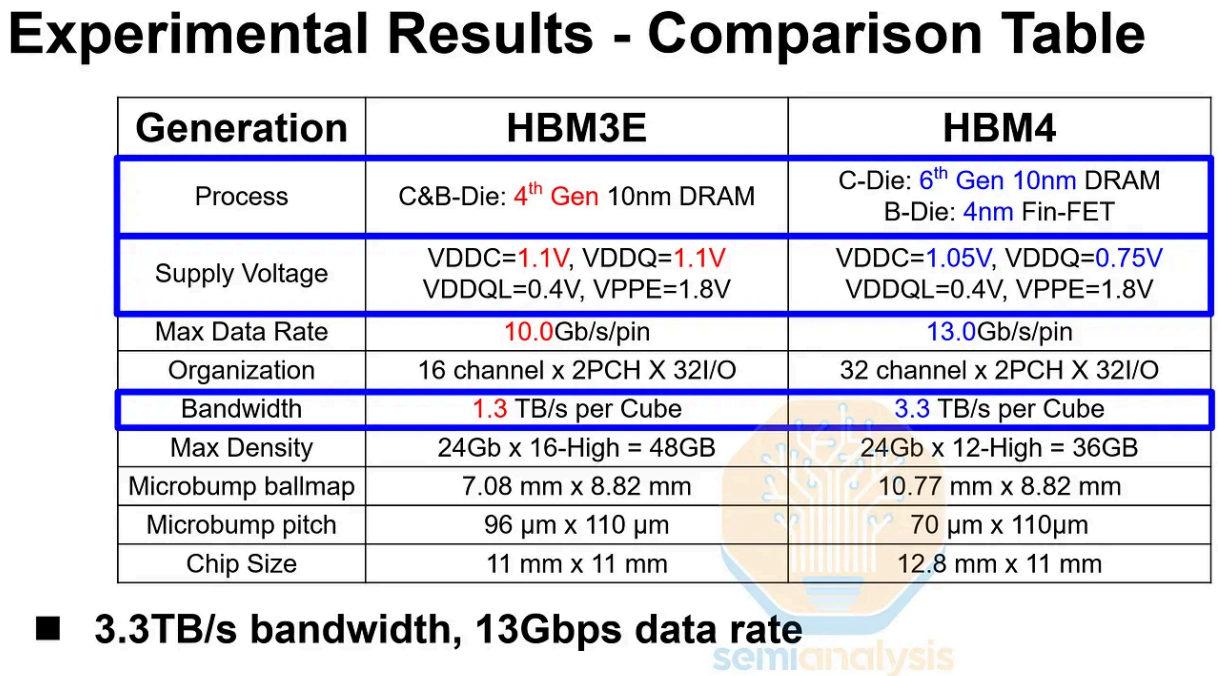
Memory

One key theme that caught our attention at this year's ISSCC was memory, including Samsung HBM4, Samsung and SK Hynix LPDDR6, and SK Hynix GDDR7. Other than DRAM, logic-based SRAM and MRAM also piqued our interest.

Samsung HBM4 - Paper 15.6

Samsung was the only one among the top three memory vendors to present a technical paper on HBM4. Before ISSCC, we noted in our Accelerator & HBM model that Samsung had made great improvements in their HBM4 generation over their HBM3E. The data presented at ISSCC confirmed our analysis, with Samsung posting best-in-class performance - we have also detailed this development months ago, in a model update note.

The technical details presented at ISSCC, combined with industry chatter we have gathered, clearly demonstrate that Samsung's HBM4 is competitive with its peers. Notably, it can meet the pin speed required for Rubin while staying below 1V. While Samsung still lags SK Hynix in terms of reliability and stability, the company has made meaningful progress in closing the gap on the technology front and could challenge SK Hynix's dominance in HBM. Their 1c-based HBM4 paired with an SF4 logic base die appears to deliver stronger performance in pin speed.



Experimental Results - Comparison Table

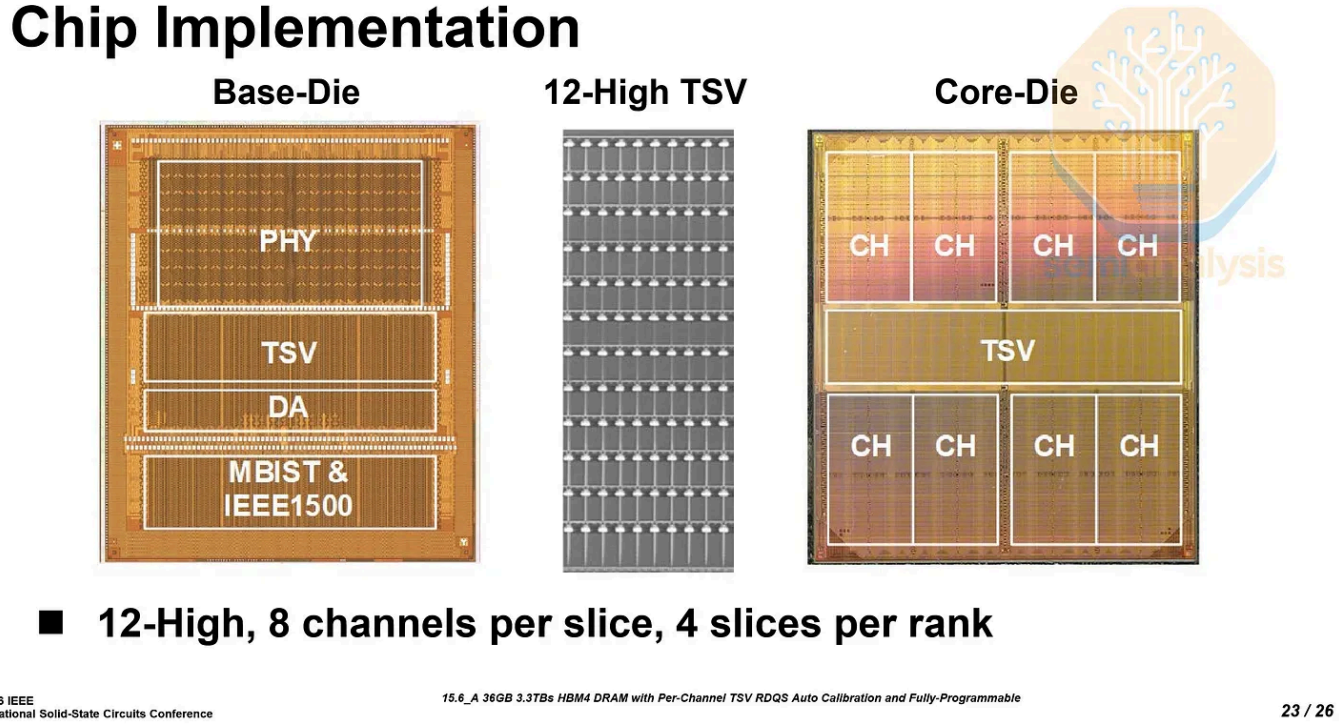
Generation	HBM3E	HBM4
Process	C&B-Die: 4th Gen 10nm DRAM	C-Die: 6th Gen 10nm DRAM B-Die: 4nm Fin-FET
Supply Voltage	VDDC=1.1V, VDDQ=1.1V VDDQL=0.4V, VPPE=1.8V	VDDC=1.05V, VDDQ=0.75V VDDQL=0.4V, VPPE=1.8V
Max Data Rate	10.0Gb/s/pin	13.0Gb/s/pin
Organization	16 channel x 2PCH X 32I/O	32 channel x 2PCH X 32I/O
Bandwidth	1.3 TB/s per Cube	3.3 TB/s per Cube
Max Density	24Gb x 16-High = 48GB	24Gb x 12-High = 36GB
Microbump ballmap	7.08 mm x 8.82 mm	10.77 mm x 8.82 mm
Microbump pitch	96 µm x 110 µm	70 µm x 110µm
Chip Size	11 mm x 11 mm	12.8 mm x 11 mm

■ 3.3TB/s bandwidth, 13Gbps data rate

semianalysis

© 2026 IEEE
International Solid-State Circuits Conference

Samsung HBM3E vs. HBM4 Specifications. Source: Samsung, ISSCC 2026

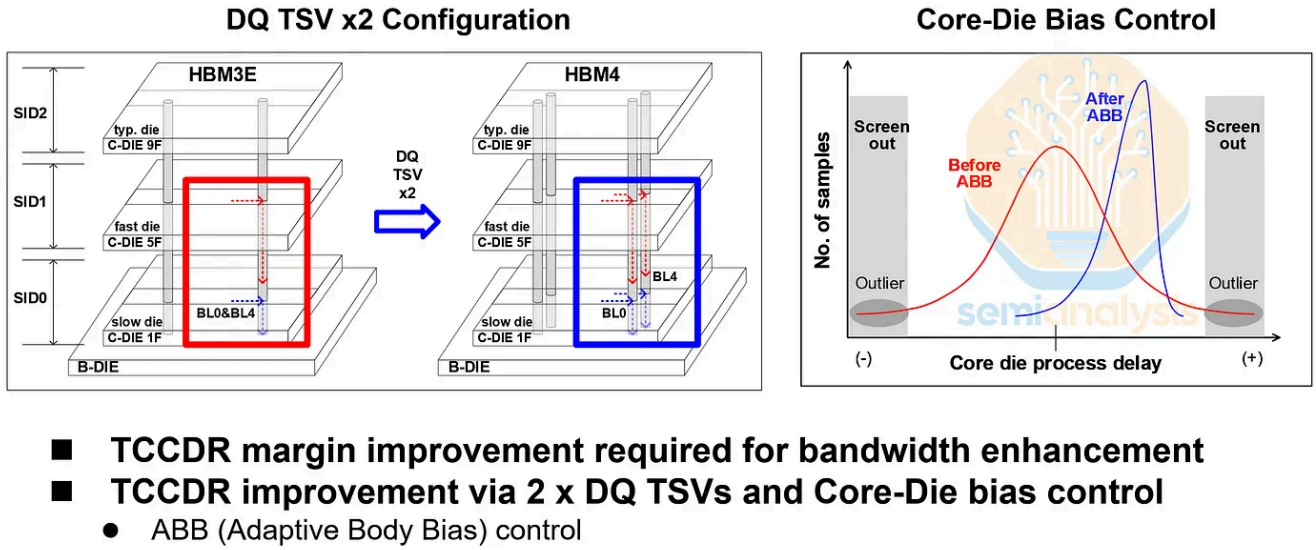


Samsung HBM4 Die Shots and Cross-Section. Source: Samsung, ISSCC 2026

Samsung demonstrated a 36 GB, 12-high HBM4 stack featuring 2048 IO pins and 3.3 TB/s of bandwidth, built using 6th-generation 10nm-class (1c) DRAM core dies paired with an SF4 logic base die.

The most obvious architectural change from HBM3E to HBM4 is the process technology split between the core DRAM dies and the base die. HBM4 uses the DRAM process node only for the core die while the base die is manufactured with an advanced logic node unlike previous generations of HBM that used the same process for both.

The key architectural challenge arises as AI workloads demand higher bandwidth and faster data rates from HBM. By moving the base die to the SF4 logic process, Samsung enables higher operating speeds and lower power consumption. The operating voltage (VDDQ) fell 32%, from 1.1V in HBM3E to 0.75V in HBM4. A logic-based base die provides higher transistor density, smaller device dimensions, and better area efficiency due to smaller transistors and larger metal-layer stack availability as compared to a base die fabricated on a DRAM process. This helps Samsung's HBM4 achieve — and significantly surpass — JEDEC's HBM4 standard that we explain more at the end of this section.



Samsung HBM4 Adaptive Body-Bias Control and Process Variation. Source: Samsung, ISSCC 2026

Combined with adaptive body-bias (ABB) control, which mitigates process variation across stacked core dies, the doubled TSV counts further improve timing margin. Together, Samsung's paper claimed that the ABB and the 4× higher TSV count allow their HBM4 to achieve operating speeds up to 13 Gb/s per pin.

The improvement brought by the SF4 base die and 1c DRAM core dies comes with a trade-off. Samsung's choice of SF4 for the logic base die comes at a higher cost compared with competing approaches even though Samsung Foundry can offer discounts for their internal base die usage. SK Hynix is adopting TSMC's N12 logic process for their HBM4 base die, while Micron relies on their internal CMOS base-die technology, both of which are lower-cost options than the near leading-edge SF4 node, even considering vertical integration cost advantages.

The 1c front-end manufacturing process has proved challenging for Samsung throughout 2025, especially given that the company skipped the 1b node and moved directly from 1a-based HBM3E to the 1c generation. Front-end yields for the 1c node were only around 50% last year, although they have been gradually improving over time. The lower yield poses a risk for their HBM4 margins.

Historically, Samsung's HBM has earned lower margins than those of their top competitor, SK Hynix, a dynamic that we model across all vendors comprehensively in our Memory Model. We

have detailed wafer volumes, yields, density, COGS, and more for each vendors HBM, DDR, and LPDDR across various nodes.

Samsung's strategy appears to be an aggressive adoption of a more advanced node for the base die to achieve superior performance and outpace their competitors, particularly as HBM requirements from leading customers such as NVIDIA continue to become more demanding.

Another key issue in HBM to address is tCCDR, the minimum interval required between consecutive READ commands issued across different stack IDs (SID). For AI workloads that rely heavily on parallel memory access across many channels, tCCDR directly impacts achievable memory throughput.

In a stacked DRAM architecture, multiple core dies are vertically integrated on top of a base die. This naturally introduces small delay differences across the stack, driven by factors such as process variation between the core dies and the base die, TSV propagation differences, and local channel variation.

The increased stack heights and channel counts, from 16 to 32, compound this challenge. As the channel counts and stack heights increase, the variation between the dies accumulates, causing larger timing mismatches across channels and dies that impact the achievable tCCDR and overall HBM performance.

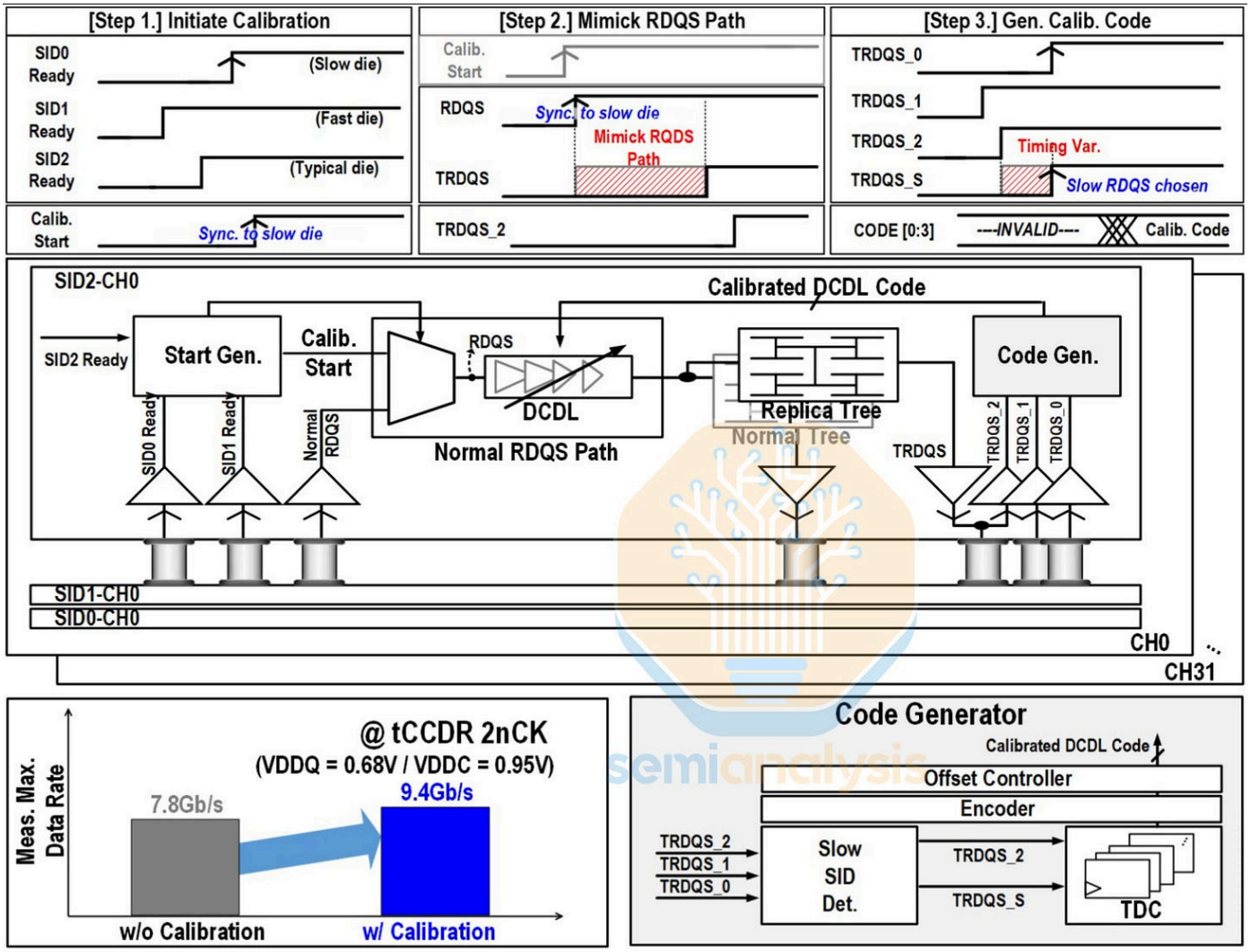


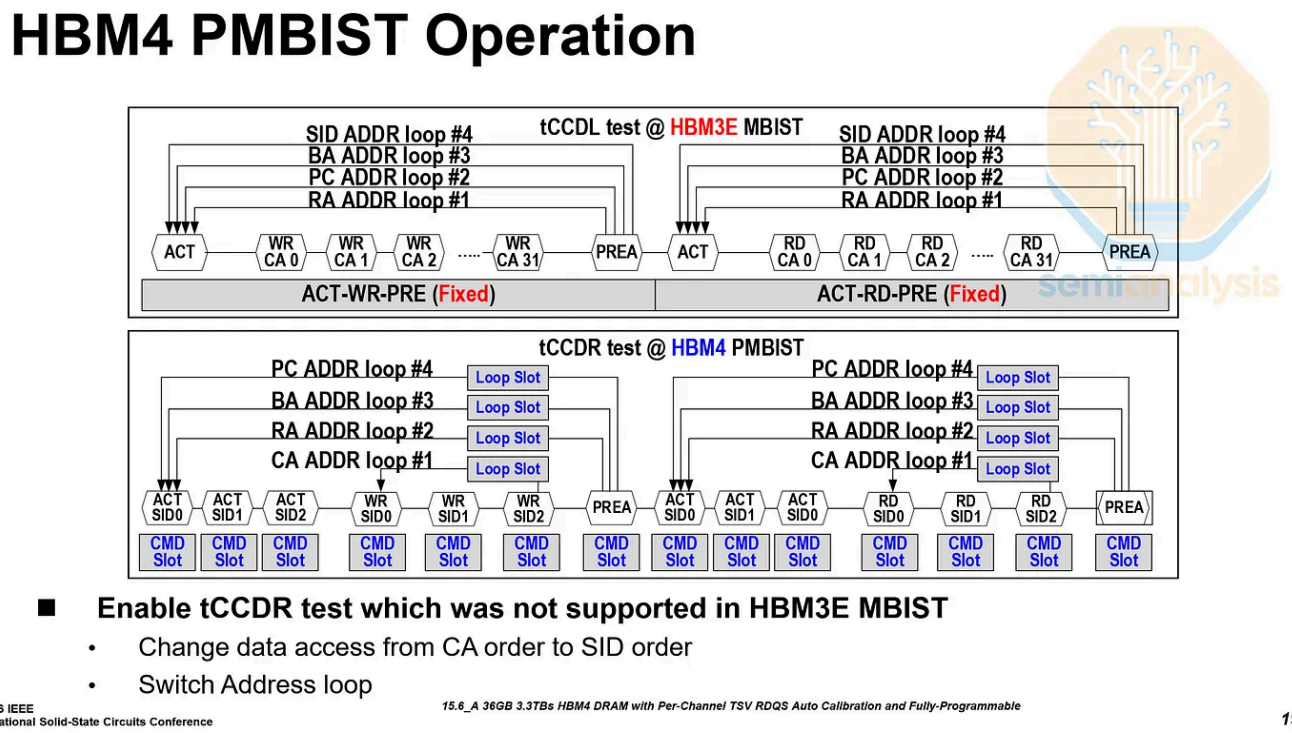
Figure 15.6.2: Conceptual illustration and block diagram of per-channel TSV t_{CCDR} RDQS auto-calibration scheme.

Samsung HBM4 Per-Channel TSV RDQS Auto-Calibration Scheme. Source: Samsung, ISSCC 2026

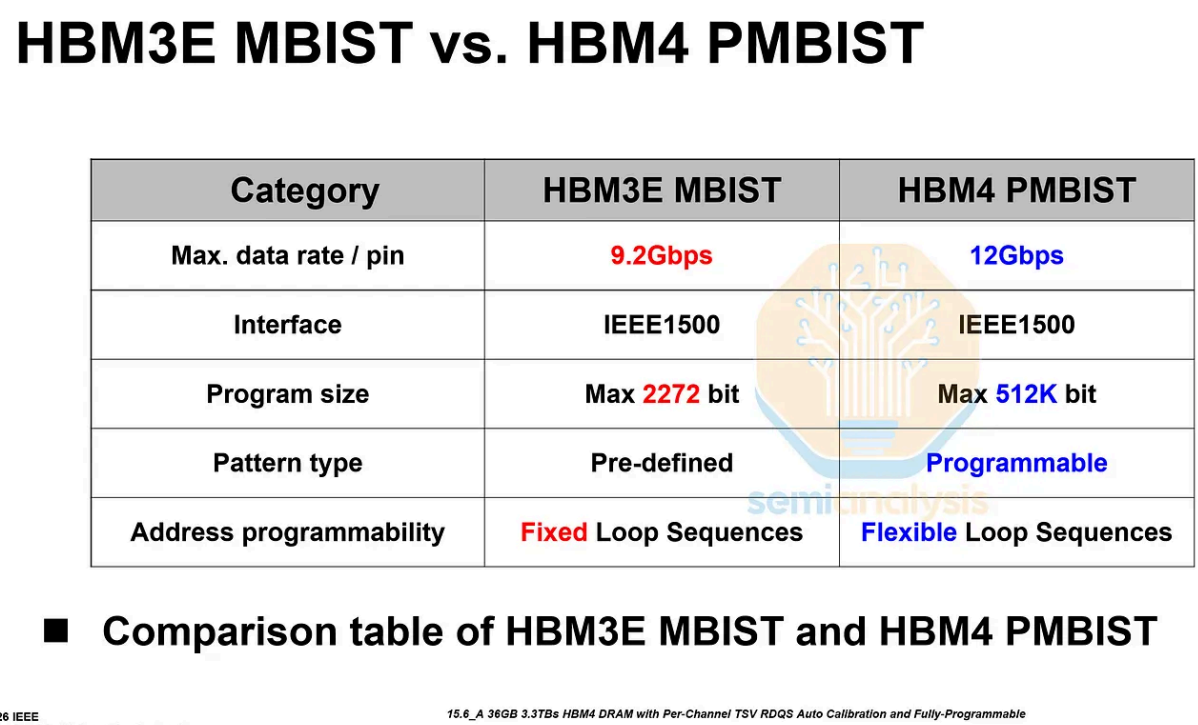
To address this issue, Samsung introduces a “per-channel TSV RDQS timing auto-calibration scheme.” After power-up, the system measures delay variation across channels using a replica RDQS path that mirrors the timing behavior of the real signal path. A time-to-digital converter (TDC) quantizes the timing differences, which are then compensated for using delay compensation circuits (DCDL) for each channel.

This calibration accounts for both global delay variation between stacked core dies and local per-channel variation, aligning timing across the stack. By compensating for these mismatches, Samsung significantly improves the effective timing margin and increases the maximum achievable data rate while maintaining the required tCCDR constraints. This scheme alone increased data rates from 7.8 Gb/s to 9.4 Gb/s.

Some of our readers who are well versed in memory technology may be asking: How is there enough die area to accommodate the significant increase in TSV counts? This is where the 1c node becomes important. Compared with the previous 1a node, 1c further shrinks the DRAM cell area, freeing up die space that can be used to integrate the larger number of TSVs required for HBM4.



Samsung HBM4 PMBIST Test Pattern Operation. Source: Samsung, ISSCC 2026

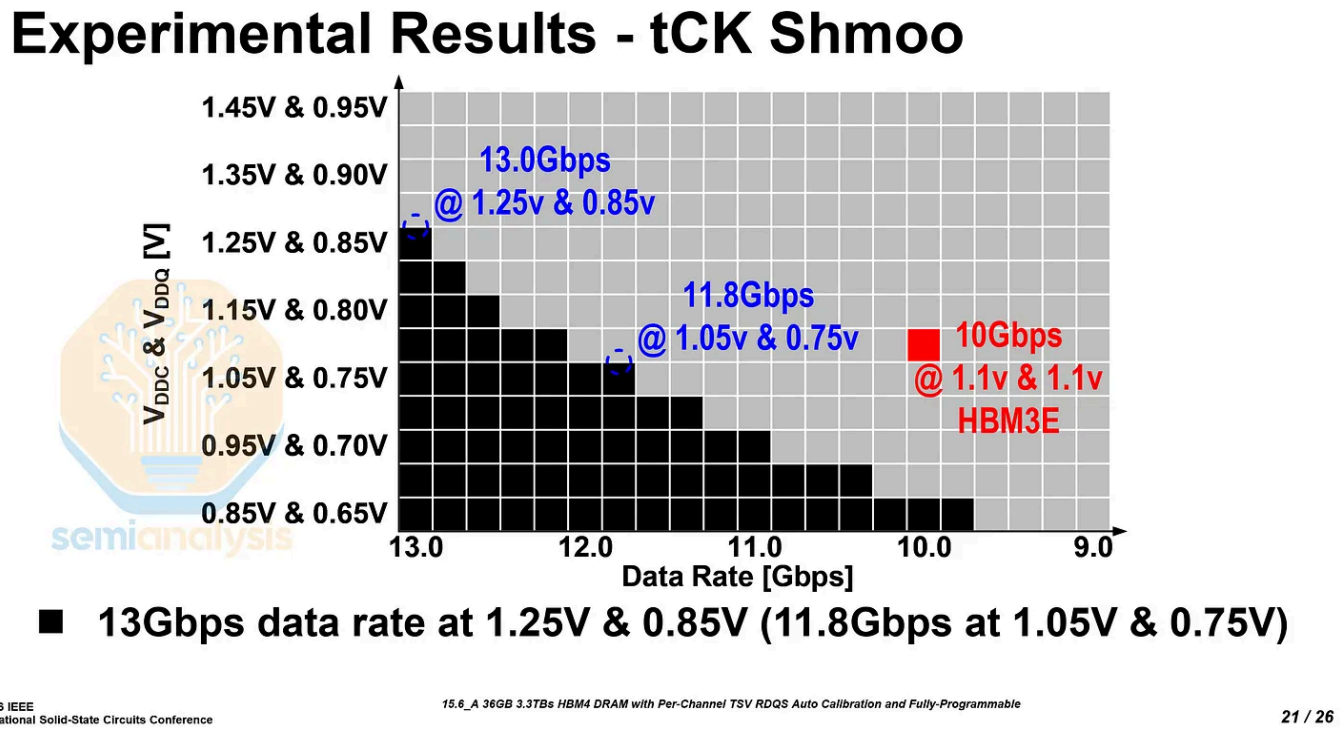


Samsung HBM4 PMBIST vs. HBM3E MBIST Comparison. Source: Samsung, ISSCC 2026

Another key innovation enabled by the logic base die is Samsung's Programmable Memory Built-In Self-Test (PMBIST) architecture. PMBIST allows the base die to generate fully programmable memory test patterns while supporting the complete JEDEC row and column command set, meaning the test engine can issue the same commands that a real system would generate and can do so at any clock edge and at full interface speed. In practical terms, this allows engineers to replicate complex real-world memory access patterns and stress the HBM interface under realistic operating conditions, which is difficult with traditional fixed-pattern test engines.

This approach represents a notable departure from HBM3E. As discussed earlier, the HBM3E base die is fabricated using a DRAM process, which imposed strict power and area constraints on the MBIST (Memory Built-In Self-Test) engine and limited testing to a small set of predefined patterns given the natural power and area disadvantage of DRAM against logic. By moving the base die to Samsung Foundry's SF4 logic process, Samsung enables a fully programmable testing framework capable of running complex test algorithms and flexible access sequences.

This enables much more robust debugging and better yield learning for HBM. Engineers can create targeted stress patterns to validate critical timing parameters such as tCCDR and tCCDS, identify corner-case failures earlier in manufacturing, and accelerate characterization during chip-on-wafer (CoW) and system-in-package (SiP) testing. Put simply, PMBIST improves test coverage, debug efficiency, and ultimately production yield as HBM stacks grow more complex and operate at higher speeds.



Samsung HBM4 Shmoo Plot. Source: Samsung, ISSCC 2026

Samsung also demonstrated strong pin speed results — their HBM4 is able to hit 11 Gb/s at sub-1V core voltage (VDDC), and up to 13 Gb/s at higher voltages. We have yet to see Samsung's peers demonstrate comparable performance albeit they do have better reliability and stability.

Samsung's implementation significantly exceeds the baseline specification of the official JEDEC HBM4 standard (JESD270-4), which specifies a maximum data rate of 6.4 Gb/s per pin and about 2 TB/s of bandwidth. Samsung demonstrates more than 2× the JEDEC-standard pin speed, reaching 13 Gb/s per pin and delivering 3.3 TB/s of bandwidth. Even at VDDC/VDDQ of 1.05V and 0.75V, the device can sustain a data rate of 11.8 Gb/s.

Samsung LPDDR6 - Paper 15.8

Both Samsung and SK Hynix showed off their LPDDR6 chips. We will discuss Samsung's chips first and turn to SK Hynix's later.

What's New in LPDDR5X vs. LPDDR6?

Category	Item	LPDDR5X	LPDDR6
Architecture	Channel / die	1 channel / die	2 sub-channels / die
	Pin Description	16 DQ + 2DMI / channel 7 CA + 1 CS / channel	12 DQ / sub-channel 4 CA + 1 CS / sub-channel
	Data Rate	Up to 10700Mb/s	Up to 14400Mb/s
	WCK to CK Ratio	4:1(2:1,<= 3200Mb/s)	2:1
	High Capacity Mode	Byte Mode (x8, row-address+1)	Static Efficiency Mode
Low Power Feature	2N Power Saving Mode	None	Dynamic Efficiency Mode
	Power for Peri-circuits	VDD1: 1.8V, VDDQ: 0.5V VDD2H: 1.05V, VDD2L: 0.9V	VDD1: 1.8V, VDDQ: 0.5V VDD2C: 1.0V, VDD2D: 0.85V-0.875V
	DVM	DVFSC, EDVFSC	DVFSL, DVFSQ, DVFSH/B
RAS	On-Die ECC	128+8 (SEC)	256+16 (SEC-DED)
	System Meta	None	Carved-out Meta
	RH Mitigation	RFM/A-RFM/D-RFM	PRAC

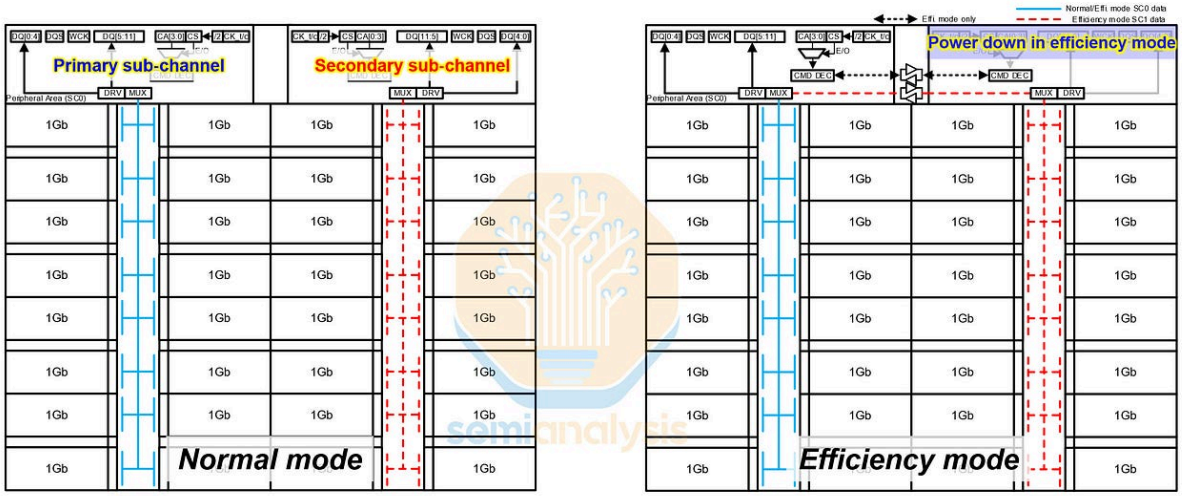
© 2026 IEEE
International Solid-State Circuits Conference

LPDDR5X vs. LPDDR6 Comparison. Source: Samsung, ISSCC 2026

Samsung presented their LPDDR6 architecture and detailed the power-saving techniques used.

Proposed Architecture

- **16 banks per sub-channel**
  - Efficiency mode for power saving / density scaling

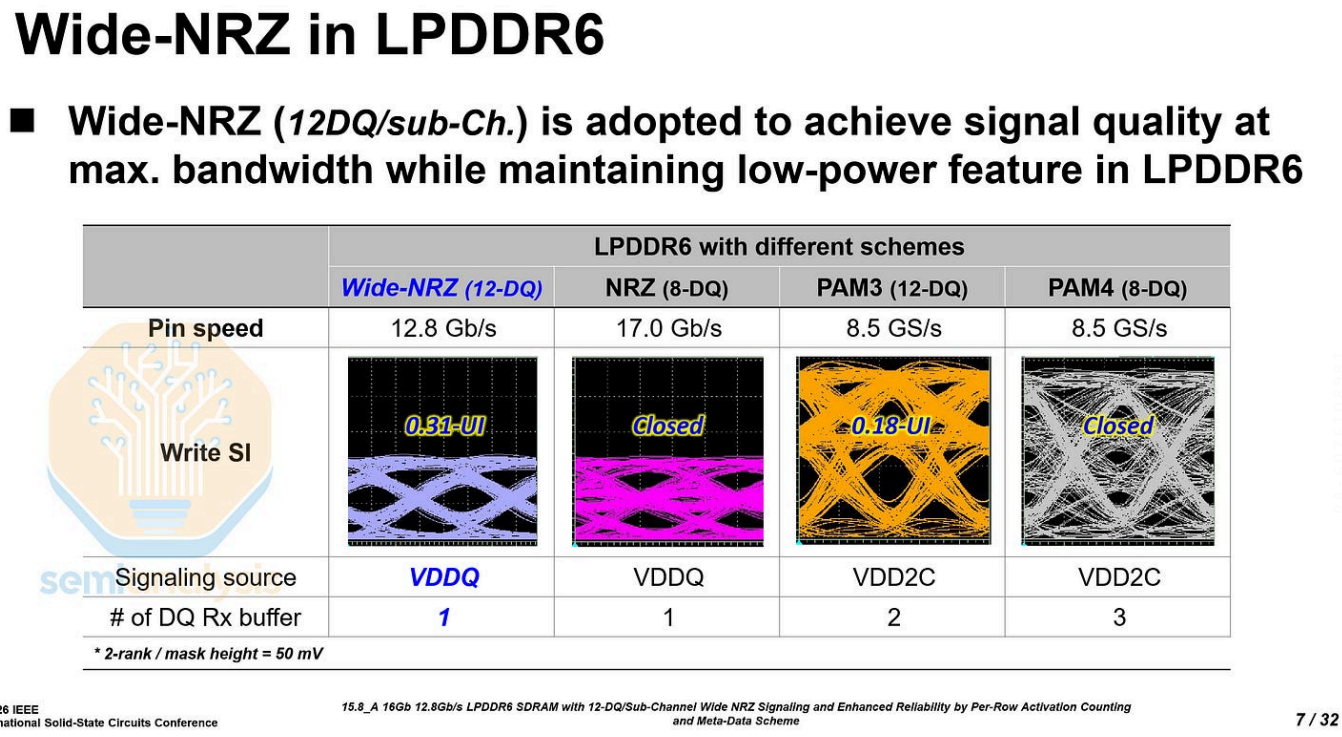


© 2026 IEEE
International Solid-State Circuits Conference

LPDDR6 Sub-Channel and Bank Structure. Source: Samsung, ISSCC 2026

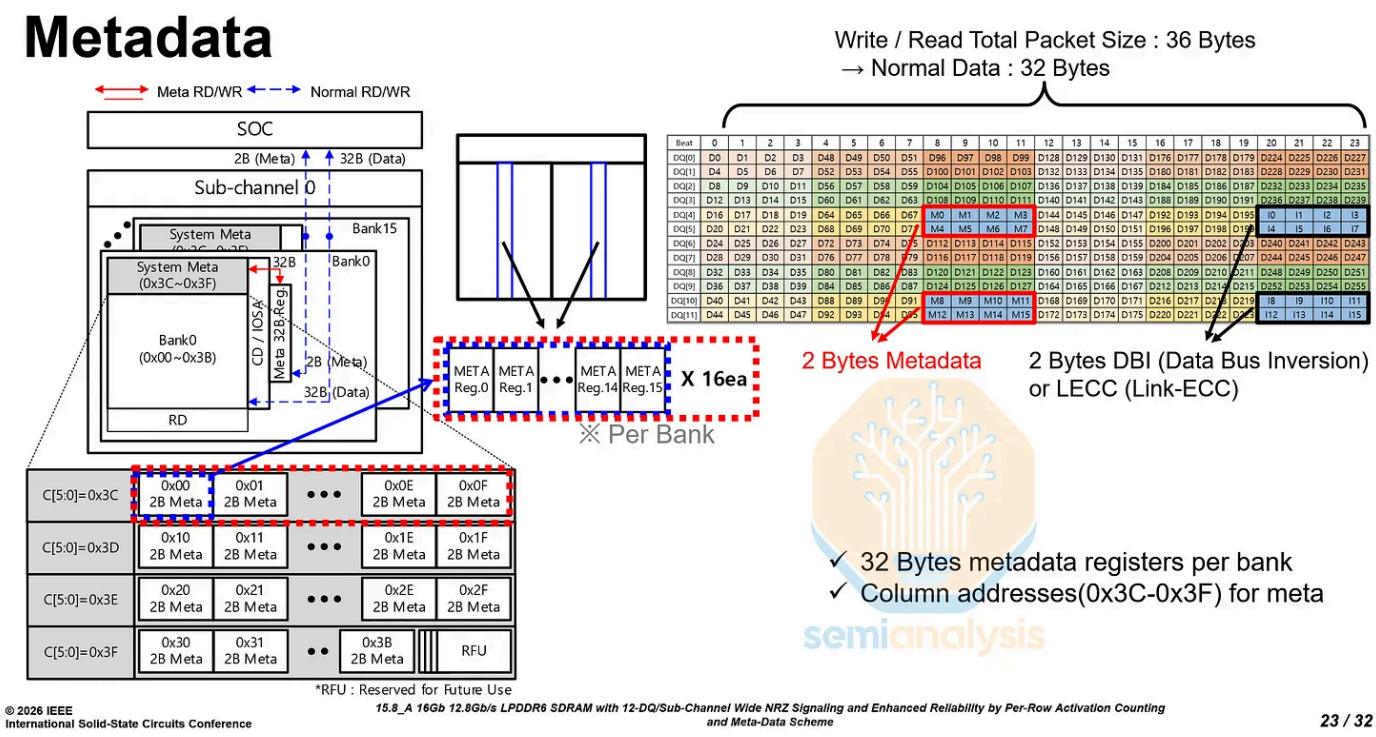
LPDDR6 adopts a 2 sub-channel per die architecture, with 16 banks in each sub-channel. It also features two modes: a normal mode and an efficiency mode. In the efficiency mode, the secondary sub-channel is powered down, and the primary sub-channel controls all 32 banks. However, there is a latency penalty for accessing data in the secondary sub-channel.

The dual sub-channel architecture also means that there is twice the amount of peripheral circuitry, such as command decoders, serialization and control. From the die shots provided by both Samsung and SK Hynix, the penalty is about 5% of the total die area, leading to a reduction in total bits per wafer.



LPDDR6 Signaling Options. Source: Samsung, ISSCC 2026

Unlike GDDR7, which uses PAM3 signaling, LPDDR6 will continue to use NRZ. However, it does not use standard NRZ as the eye would not have sufficient margin. It uses wide NRZ with 12 data (DQ) pins per sub-channel and a burst length of 24 per operation.

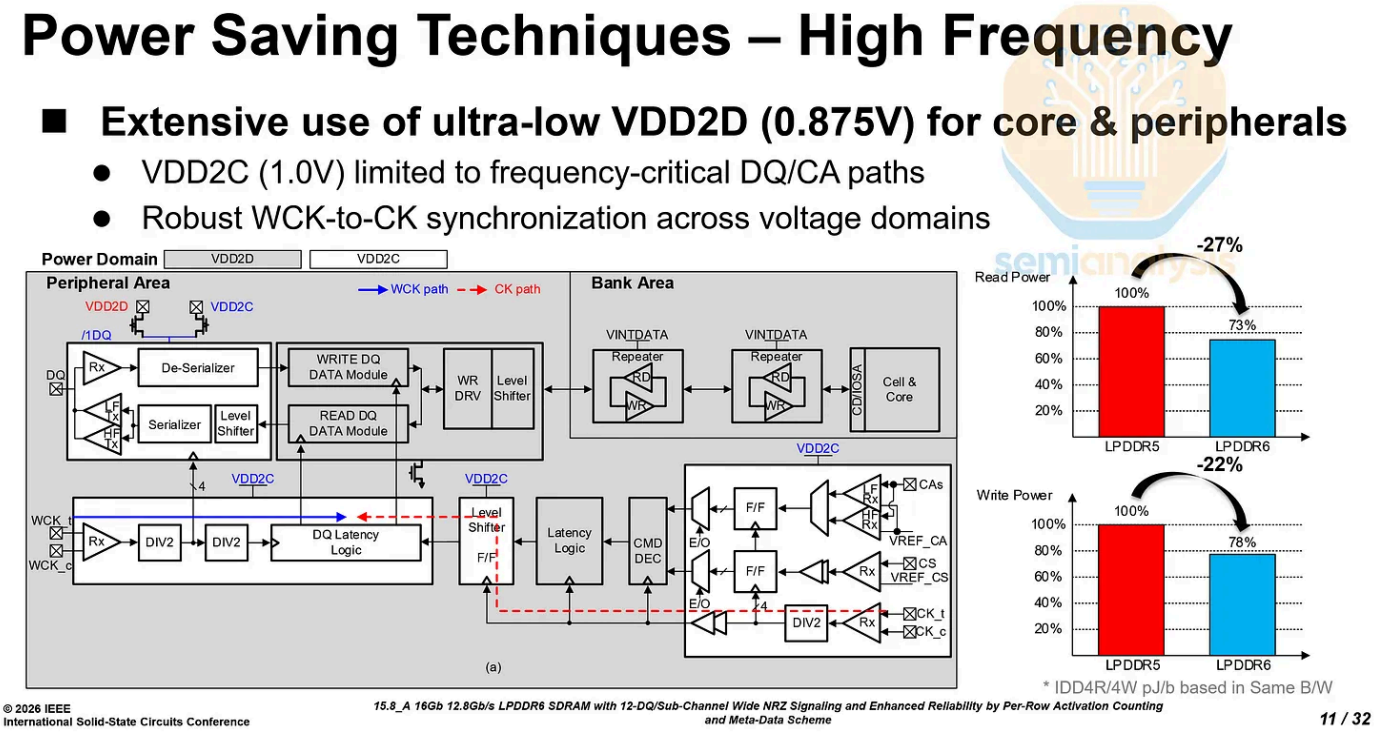


LPDDR6 Metadata and DBI Bit Allocation per Burst. Source: Samsung, ISSCC 2026

For those of you doing the math, 12×24 is 288, not a power of two. The remaining 32 bits are split into 2 use cases, 16 for metadata like ECC, and 16 for Data Bus Inversion (DBI).

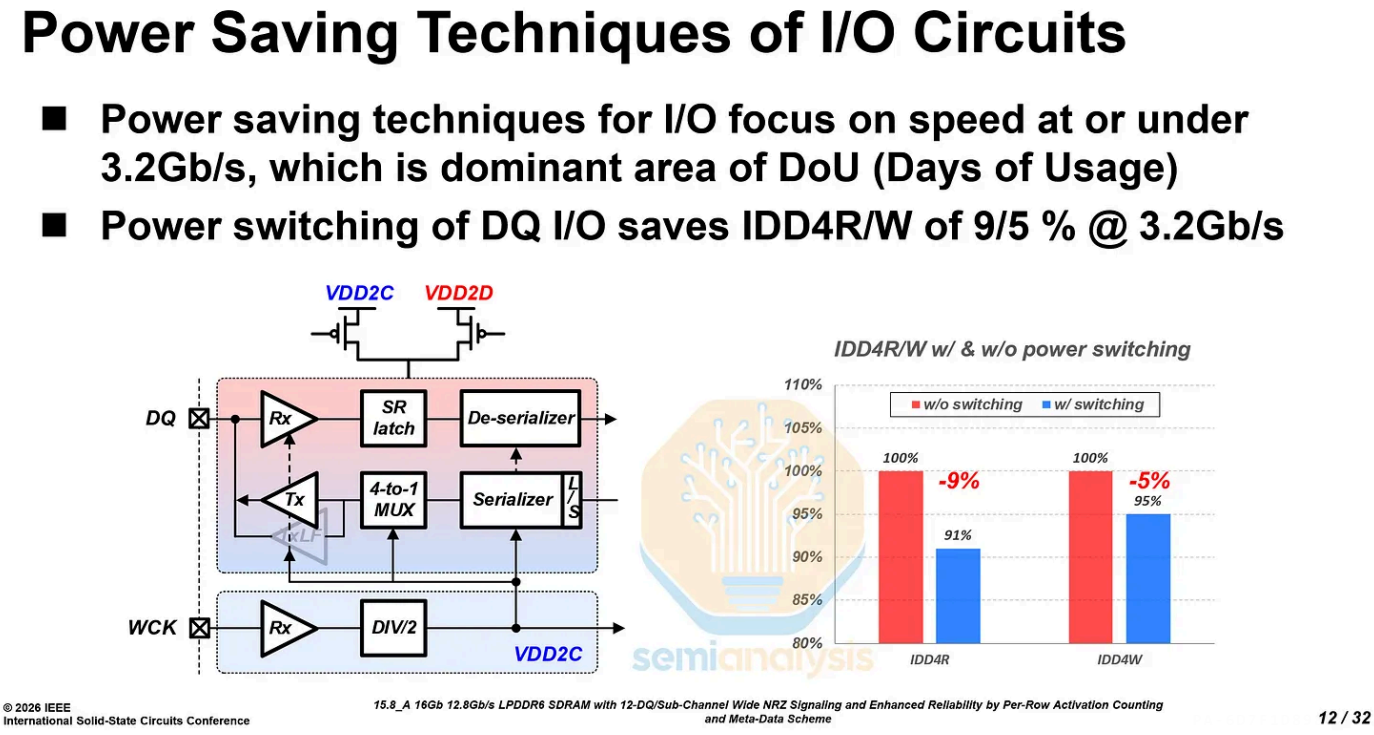
DBI is a power-saving and signal integrity mechanism. Before a burst is sent out, the controller checks if more than half the bits would switch state compared to the previous burst. If so, the controller inverts all the bits and sets a DBI flag, so that the receiver knows to invert them to get the actual data. This limits the maximum number of simultaneous switching outputs to half the bus width, reducing power consumption and supply noise.

To calculate the effective bandwidth, you must account for these metadata and DBI bits like so:
Bandwidth = Data Rate × Width (24 b) × Data (32 b) / Packet (36 b).
For 12.8 Gb/s, you get 34.1 GB/s, and for 14.4 Gb/s, you get 38.4 GB/s.

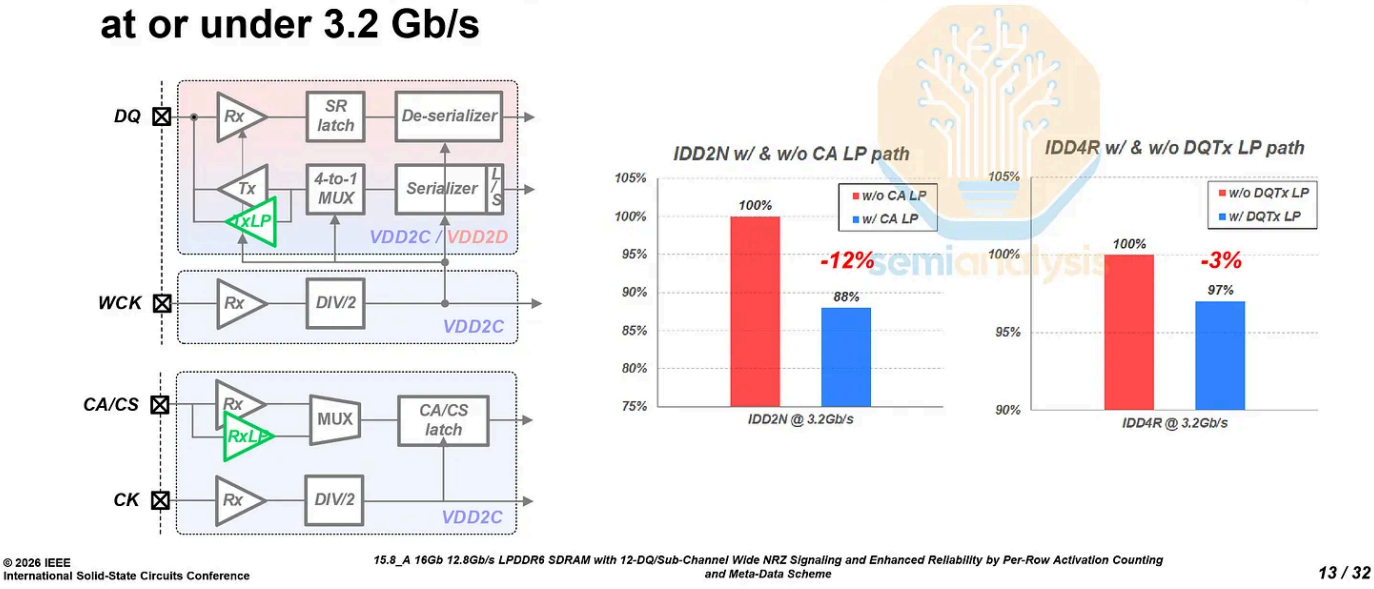


Samsung LPDDR6 High-Frequency Power Domain Optimization. Source: Samsung, ISSCC 2026

LPDDR6 has two constant power domains, VDD2C at 0.875V and VDD2D at 1.0V. By carefully choosing which peripheral logic is using which power domain, read power has been reduced by 27% and write power reduced by 22%.

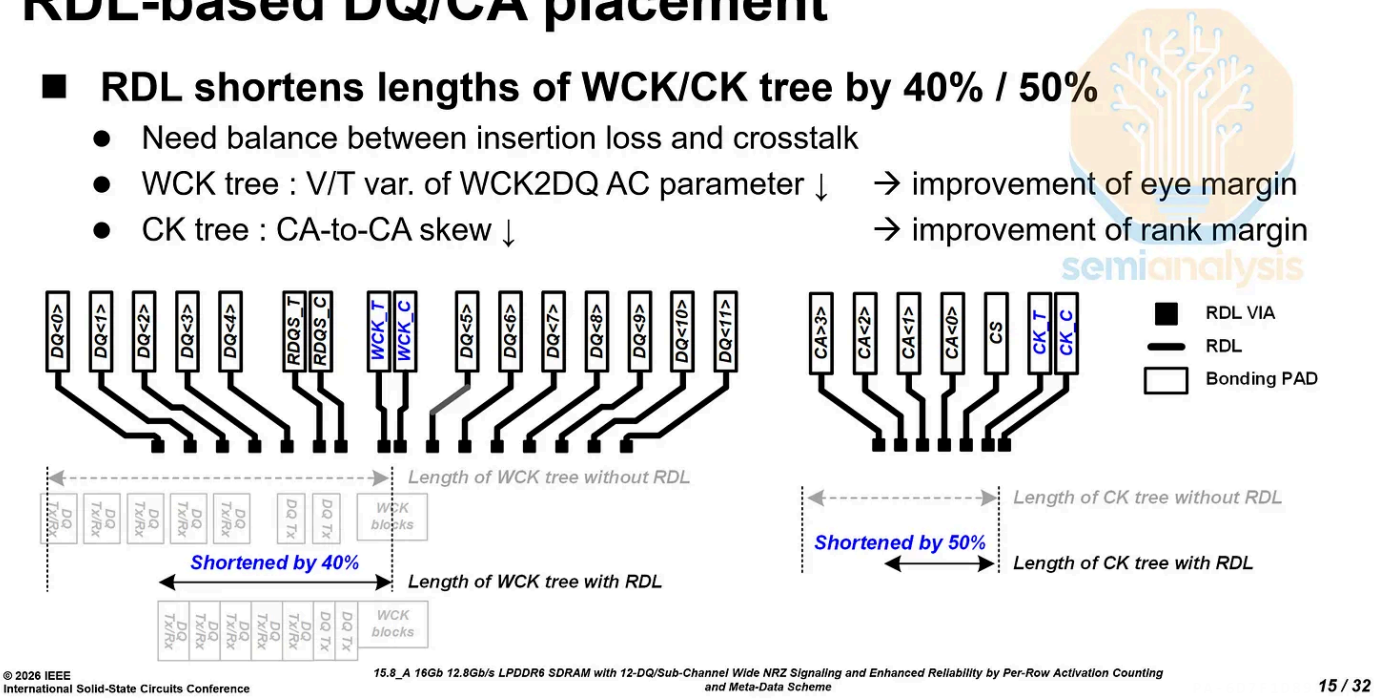


Samsung LPDDR6 I/O Power Switching at Low Data Rates. Source: Samsung, ISSCC 2026



Samsung LPDDR6 Additional Low-Power DQ/CA Paths. Source: Samsung, ISSCC 2026

LPDDR is primarily used at low data rates of 3.2 Gb/s and below when idling. Samsung focused heavily on saving power at these lower data rates through careful use of the voltage domains, reducing both standby and read/write power consumption.

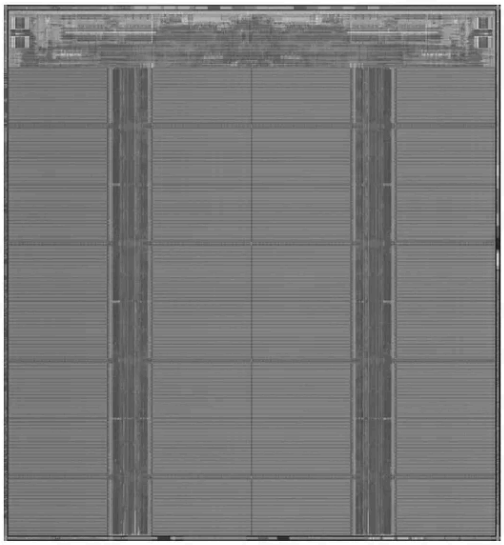


LPDDR6 RDL Timing and Layout Benefits. Source: Samsung, ISSCC 2026

By using a redistribution layer (RDL), Samsung can locate related circuits closer together physically. This shortens critical delay paths and reduces their sensitivity to voltage and temperature variation. At the high frequencies of LPDDR6, tighter timing and reduced variation are essential.

Implementation

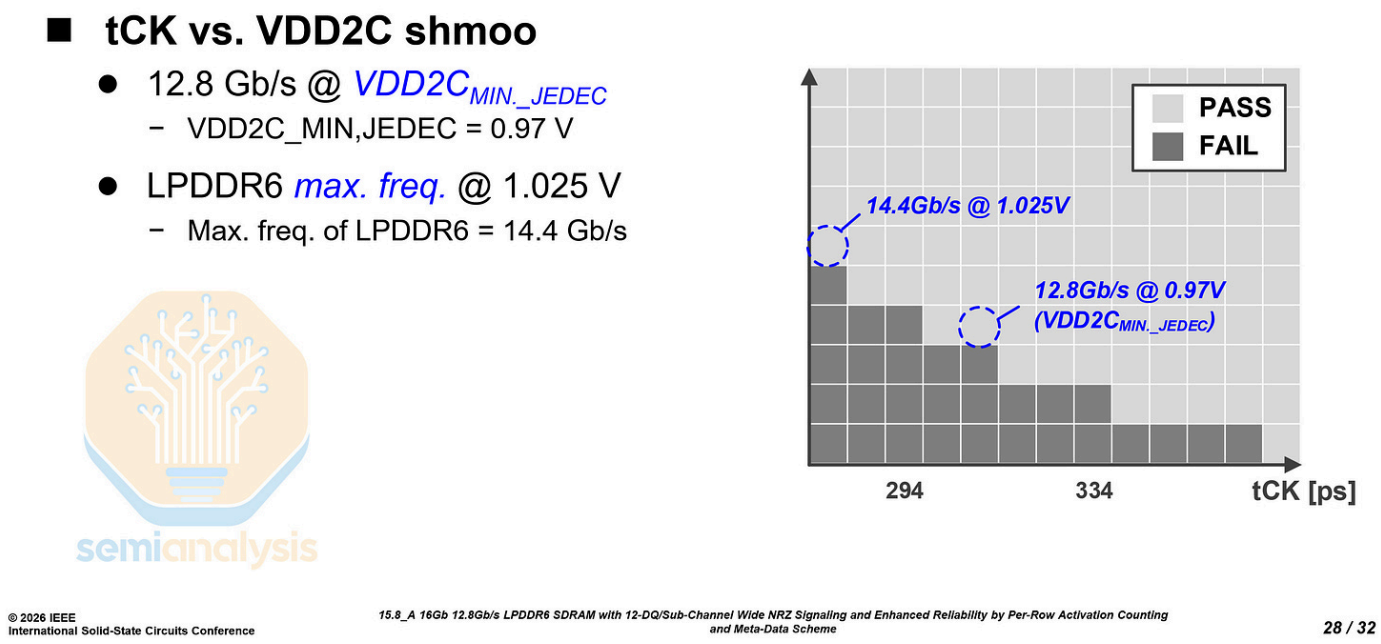
■ **16Gb DRAM micrograph in a 10nm-class DRAM process**



Process	10-nm Class DRAM Process
Supply Voltage	VDD2C=1.0V / VDD2D=0.875V
Speed	12.8Gb/s
Package Type	1295B FBGA
Density	16Gb
Chip Size	44.5mm^2 / die

© 2026 IEEE
International Solid-State Circuits Conference

Samsung LPDDR6 Specifications and Die Shot. Source: Samsung, ISSCC 2026

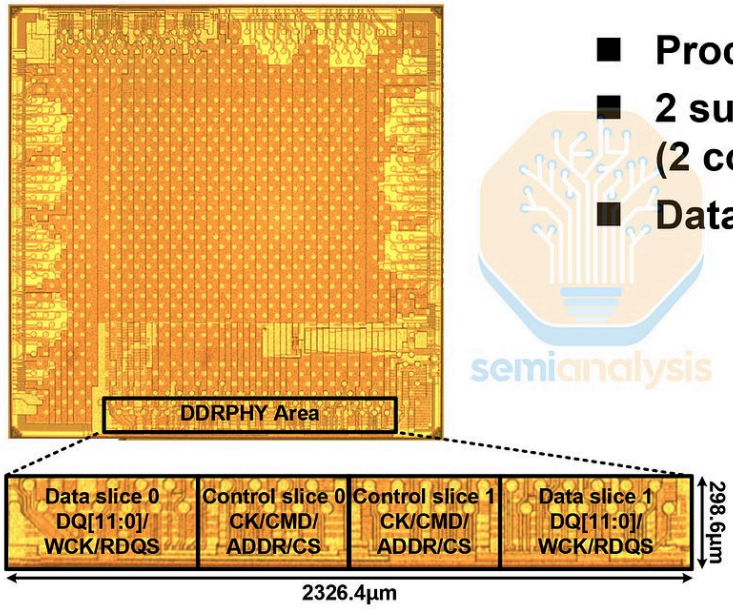


Samsung LPDDR6 Shmoo Plot. Source: Samsung, ISSCC 2026

Samsung's LPDDR6 can reach a data rate of 12.8 Gb/s at 0.97V, and up to 14.4 Gb/s at 1.025V. Each 16 Gb die is 44.5 mm², with a density of 0.360 Gb/mm² on an unknown 10nm-class process. This is considerably lower than the density of LPDDR5X on 1b at 0.447 Gb/mm² and only slightly higher than the density of LPDDR5X on 1a at 0.341 Gb/mm². While the area penalty from the dual sub-channel architecture does partially contribute, there seem to be other problems with the LPDDR6 as well. The memory density described leads us to believe that this prototype LPDDR6 chip was manufactured on their 1b process.

Samsung SF2 LPDDR6 PHY - Paper 37.3

Chip Implementation



- Process: 2nm MBCFET
- 2 sub-channels (2 control slices + 2 data slices)
- Data rate: ~14.4Gb/s/pin

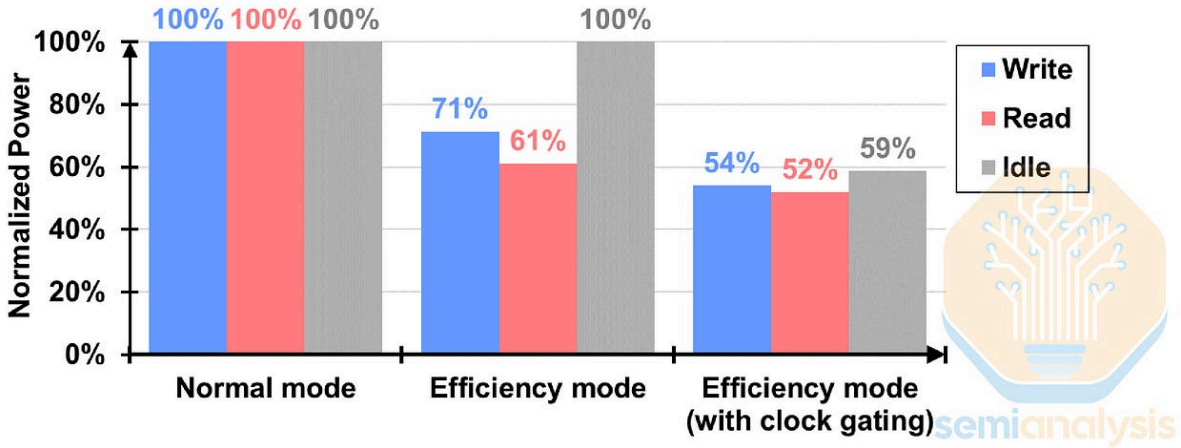
	This Work	[3] VLSI'19	[5] ISSCC'23
Interface	LPDDR6	LPDDR5	HBM3
Technology [nm]	2	8	4
Data Rate [Gb/s/pin]	14.4	7.3	9
VDD/VDDQ [V]	0.78/0.5	0.79/0.5	0.66/0.3

© 2026 IEEE
International Solid-State Circuits Conference

Samsung LPDDR6 PHY Test Chip Specifications and Die Shot. Source: Samsung, ISSCC 2026

Samsung also unveiled PHYs on the logic die interface with LPDDR6. The PHYs are fabricated on their new SF2 process and support up to 14.4 Gb/s. The PHYs take up 2.32 mm of shoreline and 0.695 mm² of area, with bandwidth densities of 16.6 Gb/s/mm and 55.3 Gb/s/mm² respectively.

Simulated PHY Power Reduction



- Without clock gating: Write 29% / Read 39% reduction
- With clock gating: Write 46% / Read 48% / Idle 41% reduction

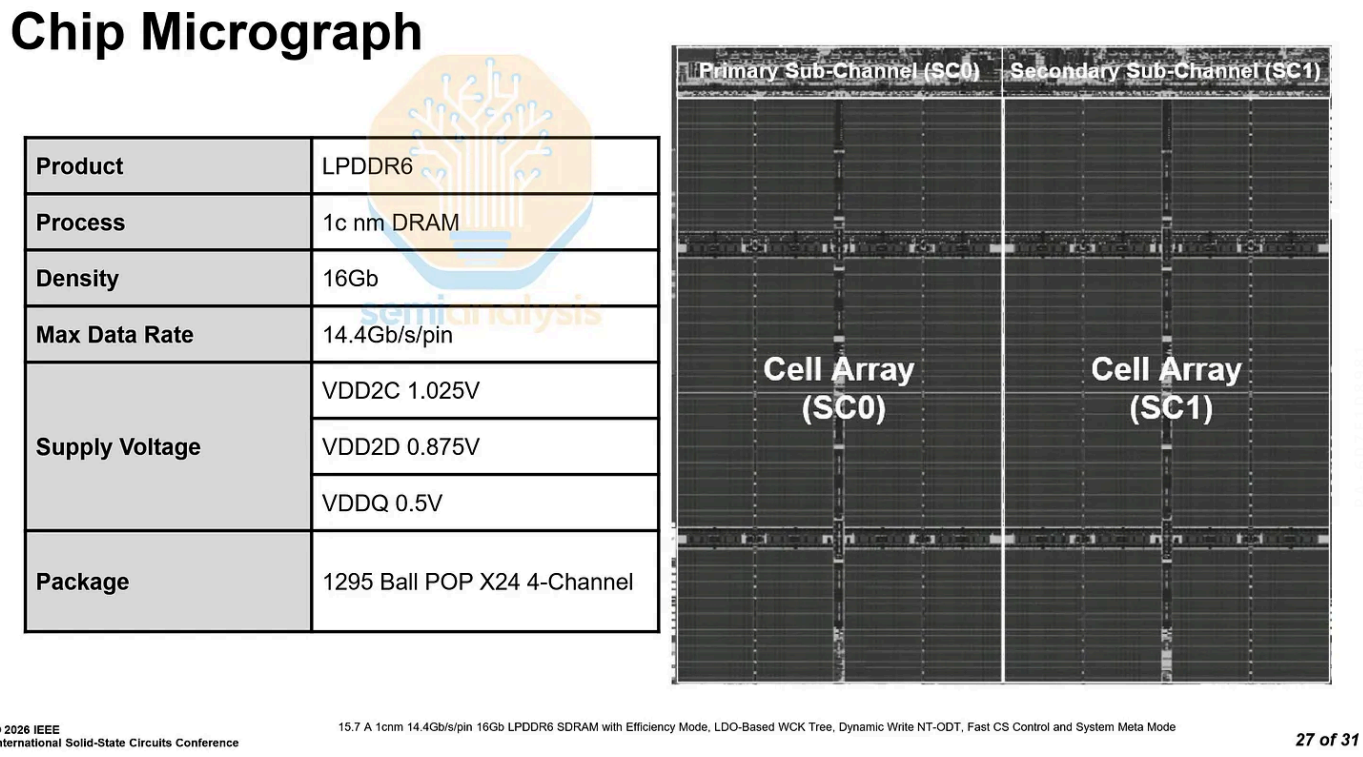
© 2026 IEEE
International Solid-State Circuits Conference

Samsung LPDDR6 PHY Efficiency Mode Power Reductions. Source: Samsung, ISSCC 2026

The PHYs also support the efficiency mode implemented by the LPDDR6 chips, which can reduce read power by 39% and write power by 29%.

The PHYs can augment the efficiency mode by gating the high-speed clock path for the inactive secondary sub-channel. With clock-gating, the power reduction reaches almost 50% for reading and writing, and idle power is reduced by 41%.

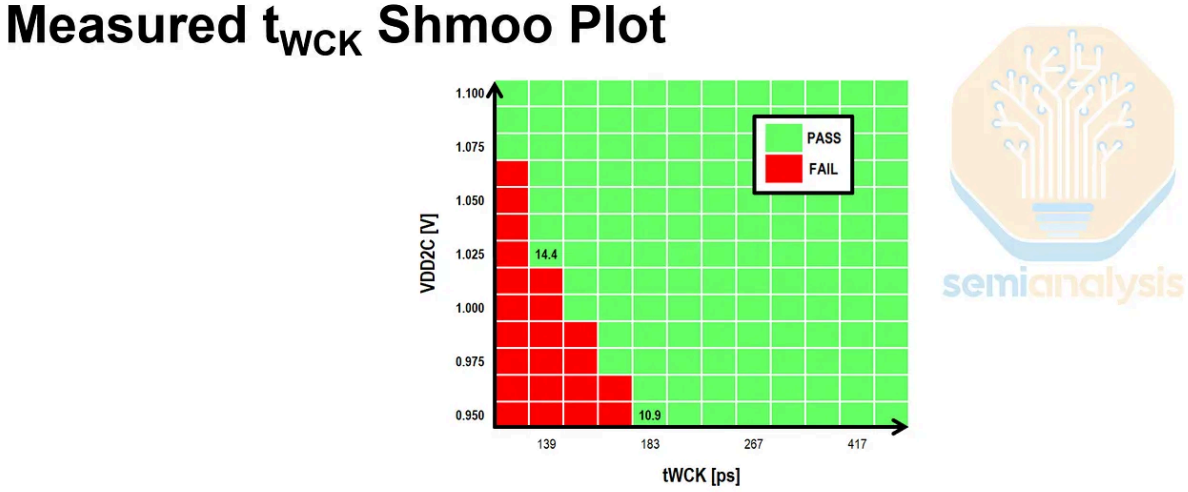
SK Hynix 1c LPDDR6 - Paper 15.7



SK Hynix LPDDR6 Specifications and Die Shot. Source: SK Hynix, ISSCC 2026

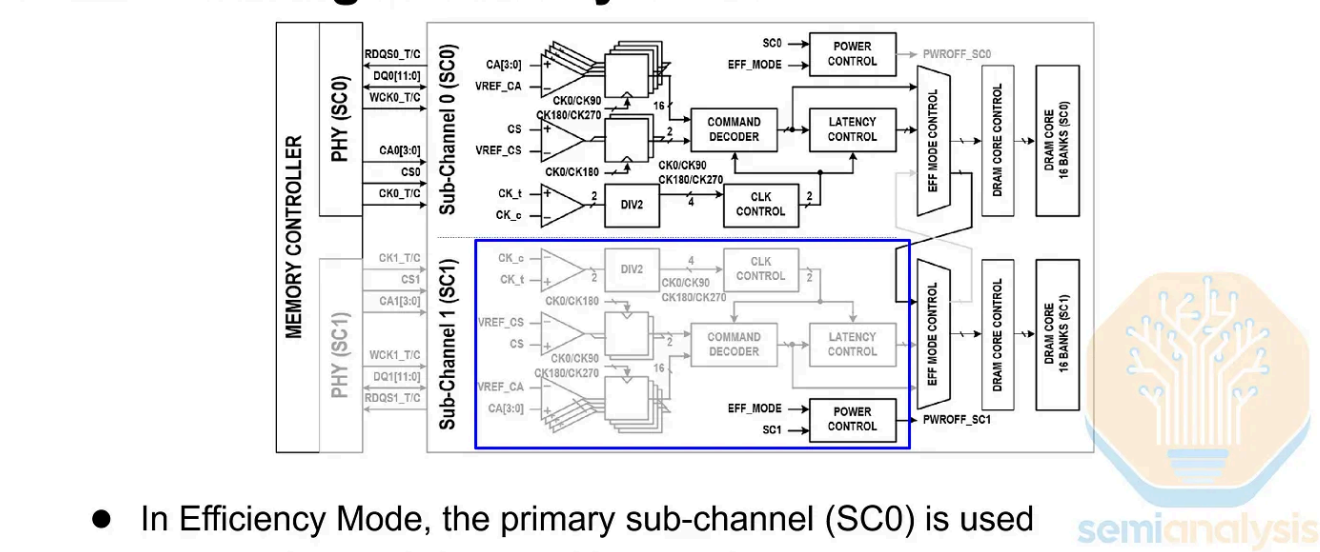
SK Hynix unveiled their first 1c DRAM products, both in LPDDR6 and in GDDR7 packages. Their LPDDR6 can operate at a data rate of up to 14.4 Gb/s, 35% faster than the fastest LPDDR5X, and at lower power.

Although SK Hynix did not state the area or density of the LPDDR6 chip, we estimate the bit density will reach 0.59 Gb/mm², based on the relative density increase of their GDDR7.

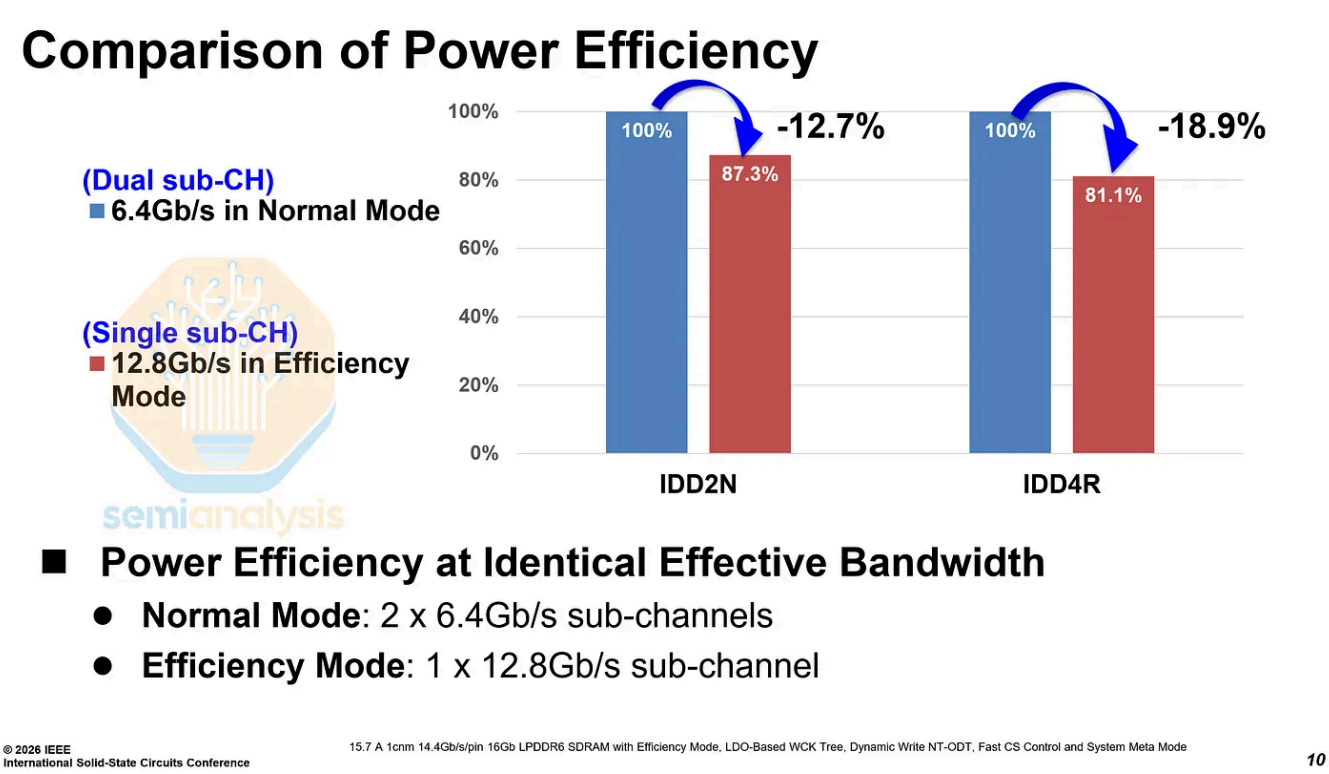


SK Hynix LPDDR6 Shmoo Plot. Source: SK Hynix, ISSCC 2026

In their shmoo plot, SK Hynix showed that they can reach a data rate of 14.4 Gb/s at 1.025V, the same as Samsung. However, they can only reach 10.9 Gb/s at 0.95V as compared to Samsung's 12.8 Gb/s at 0.97V. This indicates that SK Hynix may have worse power efficiency at lower pin speeds when compared to Samsung, having to run at higher voltages to maintain reliability.



SK Hynix LPDDR6 Efficiency Mode Architecture. Source: SK Hynix, ISSCC 2026

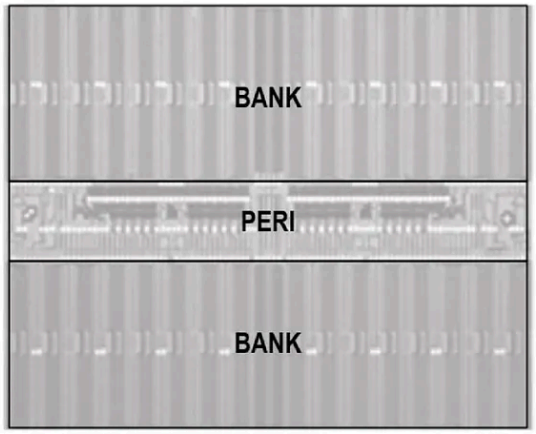


SK Hynix LPDDR6 Efficiency Mode Power Savings. Source: SK Hynix, ISSCC 2026

Like Samsung's LPDDR6, SK Hynix's LPDDR6 also features two modes, a normal mode and an efficiency mode. The efficiency mode runs at 12.8 Gb/s over a single sub-channel, with 12.7% and 18.9% lower standby and operational current draw respectively compared to normal mode.

SK Hynix 1c GDDR7 - Paper 15.9

Chip Implementation



Process	DRAM 1c nm
Supply [VDD/VDDQ]	1.2V/1.2V 1.05V/ 0.9V
Data rate	48Gbps @ 1.2V 30.3Gbps @ 1.05V/0.9V
Pkg. type	266B FBGA
Pkg. density	24Gb
Chip size	29.10mm^2

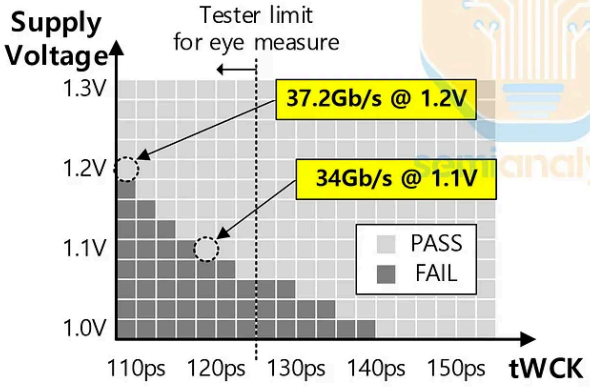
semianalysis

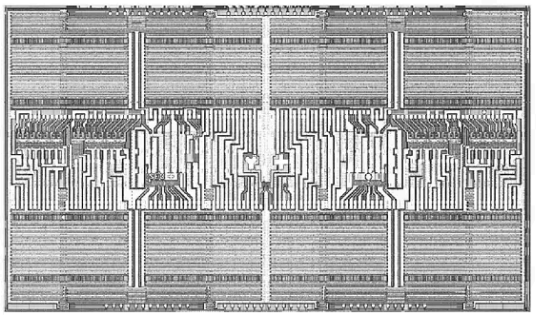
© 2026 IEEE
International Solid-State Circuits Conference

SK Hynix 1c GDDR7 Specifications and Die Shot. Source: SK Hynix, ISSCC 2026

While the LPDDR6 is a generational leap with new memory technology, SK Hynix's GDDR7 on their 1c process shows an even greater improvement, clocking up to 48 Gb/s at 1.2V/1.2V. Even at only 1.05V/0.9V, it can clock up to 30.3 Gb/s, higher than the 30 Gb/s memory in the RTX 5080.

Measurements



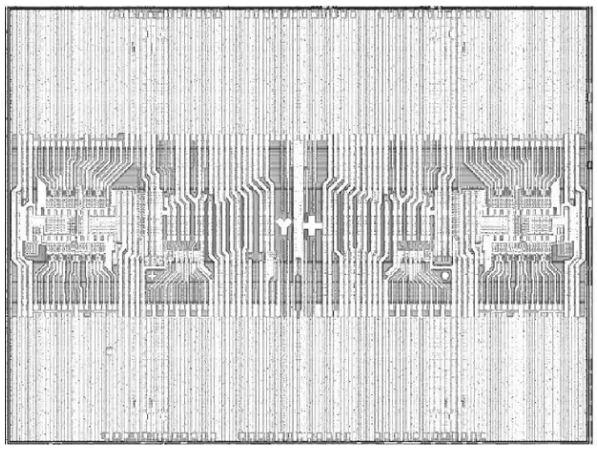


Process	1z nm DRAM
Density	16Gb/PKG
Data rate	37.2Gb/s at 1.2V 34Gb/s at 1.1V
Chip size	41.6mm²/die

© 2024 IEEE
International Solid-State Circuits Conference

Samsung 1z GDDR7 Shmoo Plot and Die Shot. Source: Samsung, ISSCC 2024

Chip Implementation



Process	DRAM 1b nm
Supply voltage	1.2V
Data rate	42.5Gb/s @1.2V 38.8Gb/s @1.1V
Package type	266B FBGA
Density	24Gb/PKG
Chip size	38.85mm²/die

© 2025 IEEE
International Solid-State Circuits Conference

Samsung 1b GDDR7 Specifications and Die Shot. Source: Samsung, ISSCC 2025

The bit density achieved is 0.412 Gb/mm², compared to 0.309 Gb/mm² on Samsung's 1b process, and 0.192 Gb/mm² on Samsung's older 1z process.

LPDDR5X and GDDR7 Densities (Gib/mm²)					
Manufacturer	Type	1z	1a/1α	1b/1β	1c/1γ
Samsung	LPDDR5X	0.283	0.341	0.436	-
SK Hynix	LPDDR5X		0.316	0.396	-
Micron	LPDDR5X		0.307	0.435	-
Samsung	GDDR7	0.192	-	0.309	-
SK Hynix	GDDR7		-	-	0.412
Micron	GDDR7		-	0.305	-

LPDDR5X vs. GDDR7 Density Comparison Across Vendors. Source: SemiAnalysis

GDDR7 has lower bit density than LPDDR5X, usually around 70% of the latter. Although it has much higher data rates, this comes at a cost, both in terms of power and area.

GDDR7's lower density is a result of the significantly higher periphery area for high access speeds. The actual memory arrays thus make up a smaller percentage of die area. This more complex logic control circuit is required for the PAM3 and QDR (4 symbols per clock cycle) signaling used in GDDR7.

GDDR7 is mainly used in gaming GPU applications that require high memory bandwidth at lower cost and capacity compared to HBM. NVIDIA had announced the Rubin CPX large-context AI processor in 2025 with 128GB of GDDR7, but this has all but vanished from the 2026 roadmaps as NVIDIA focuses on rolling out their Groq LPX solutions instead.

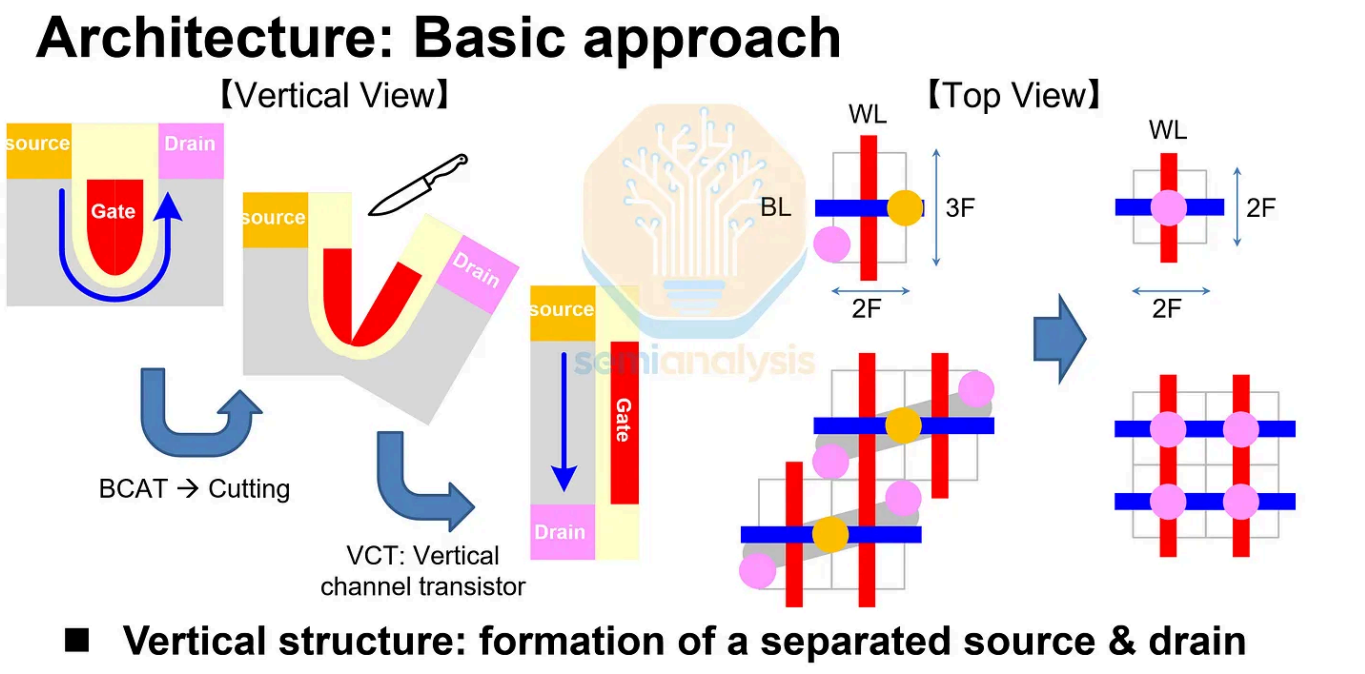
We have detailed wafer volumes, yields, density, COGS, and more in our memory model for HBM, DDR, and LPDDR across various nodes.

Samsung 4F² COP DRAM - Paper 15.10

We have extensively covered challenges in continuing to scale DRAM.

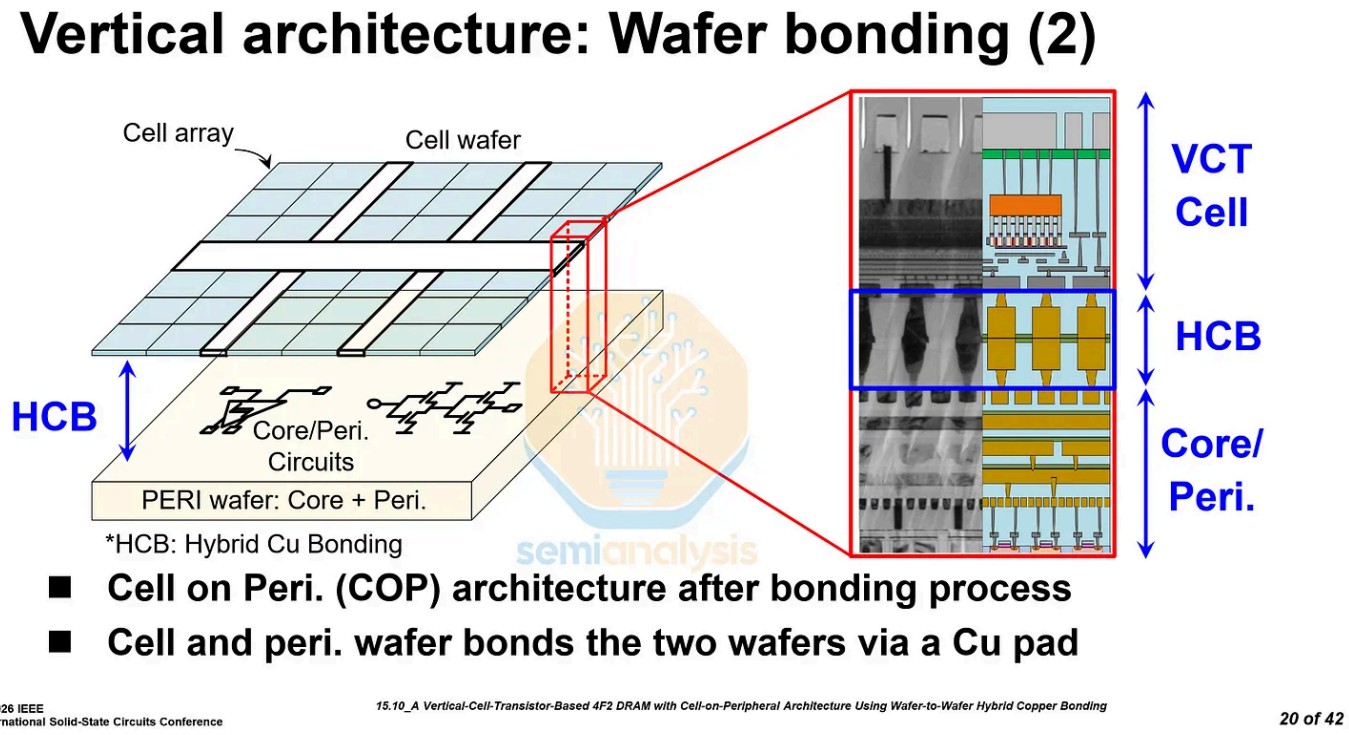
The Memory Wall: Past, Present, and Future of DRAM

At VLSI 2025, SK Hynix detailed their own 4F² Peri-Under-Cell (PUC) DRAM. At ISSCC, Samsung disclosed their own implementation of a 4F² Cell-on-Peripheral (COP) DRAM. PUC and COP are the same architecture with different names.



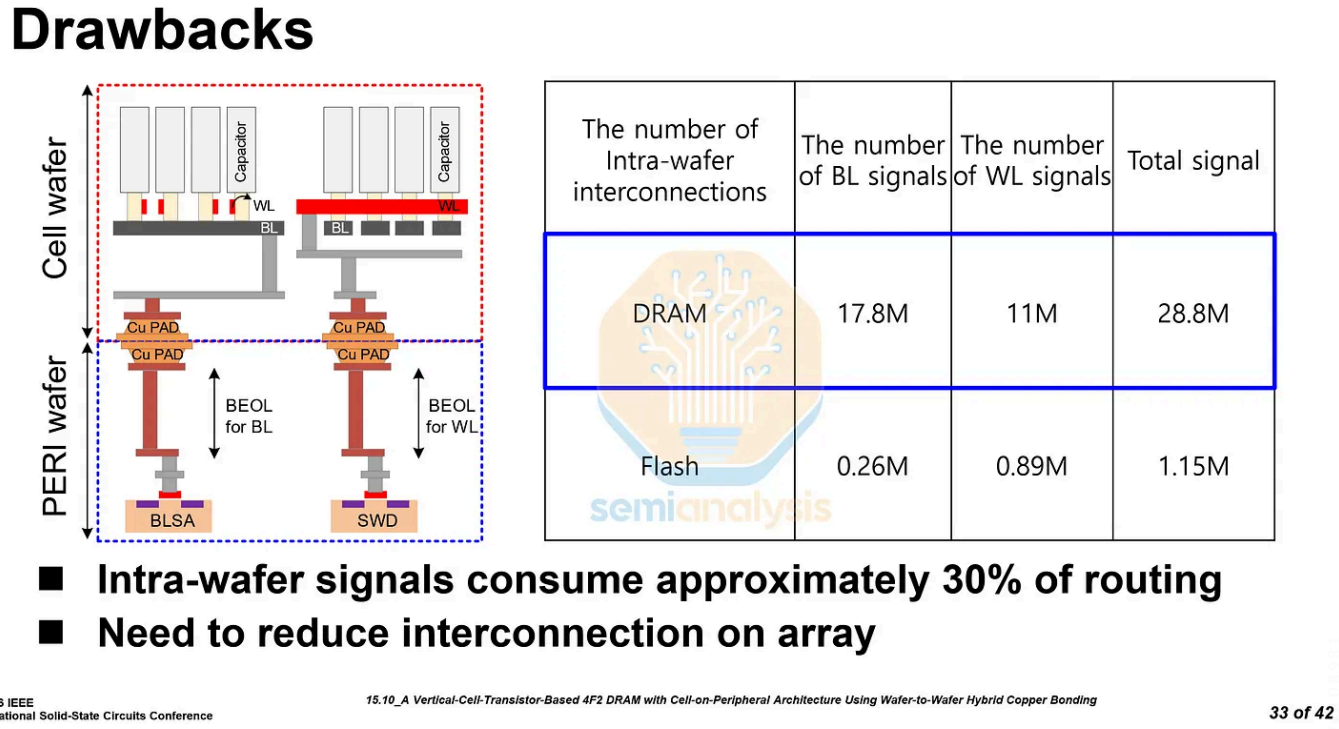
4F² VCT DRAM Cell Architecture. Source: Samsung, ISSCC 2026

The architecture for 4F² cells is the same as SK Hynix's, with vertical channel transistors (VCT), and capacitors above the drain.



Cell-on-Peripheral (COP) DRAM Stack Architecture. Source: Samsung, ISSCC 2026

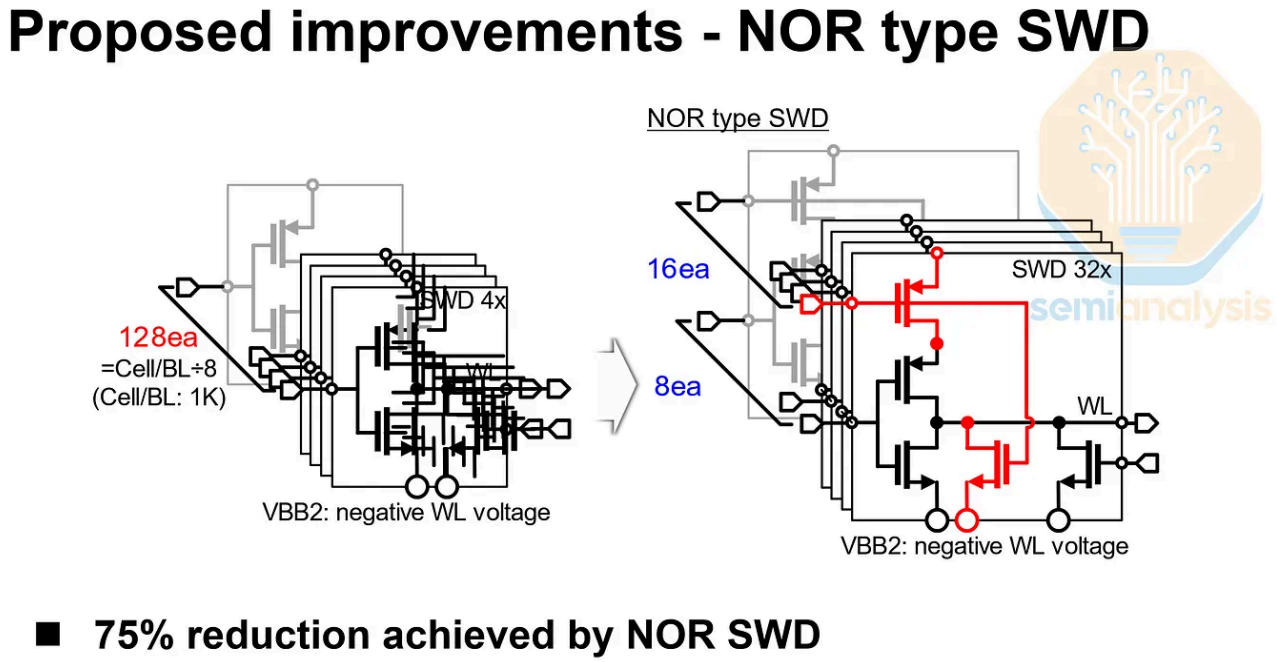
The vertical architecture presented by Samsung is essentially the same as that used by SK Hynix, with a cell wafer hybrid bonded on top of a peripheral wafer. With this architecture, it is possible to use a DRAM node for the cell wafer while using a more advanced logic node for the periphery.



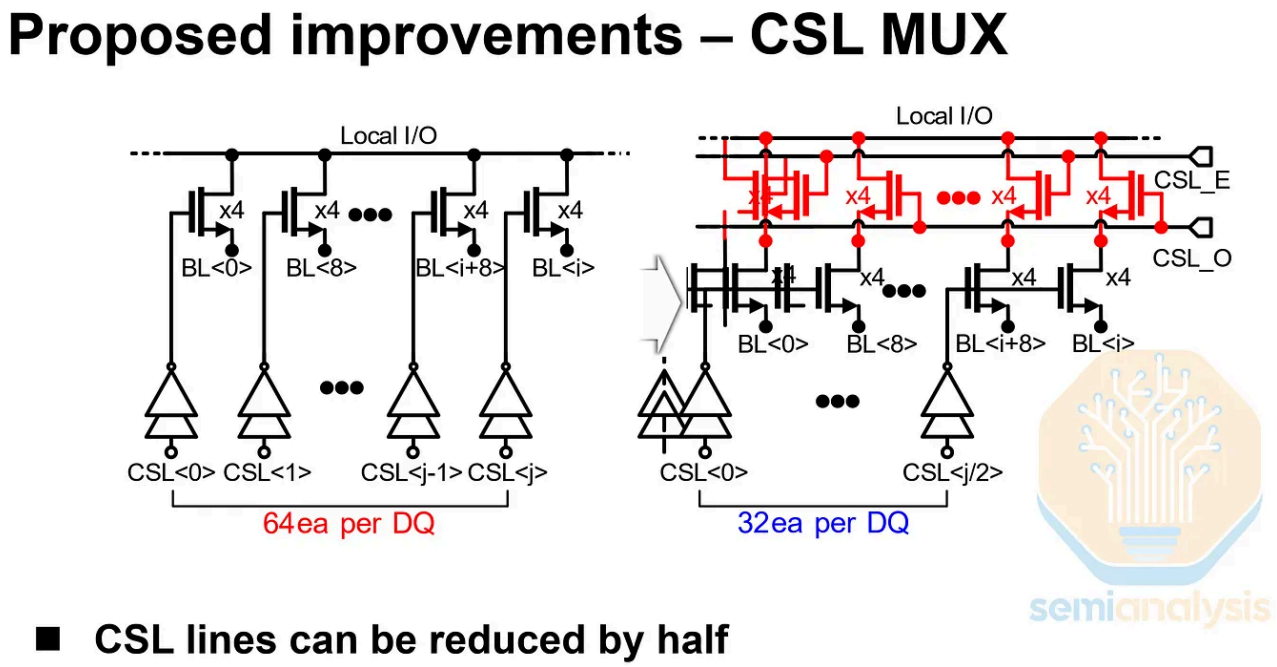
COP Architecture Comparison for DRAM vs. NAND. Source: Samsung, ISSCC 2026

Samsung notes that hybrid bonding for COP has already been used for NAND. This is true for other NAND manufacturers, but Samsung has not brought hybrid bonding for NAND into high volume production and is still years away from doing so.

Moreover, the number of inter-wafer connections for DRAM is an order of magnitude higher than for NAND and requires much tighter pitches. To reduce the number of inter-wafer interconnections, Samsung has employed two novel approaches.



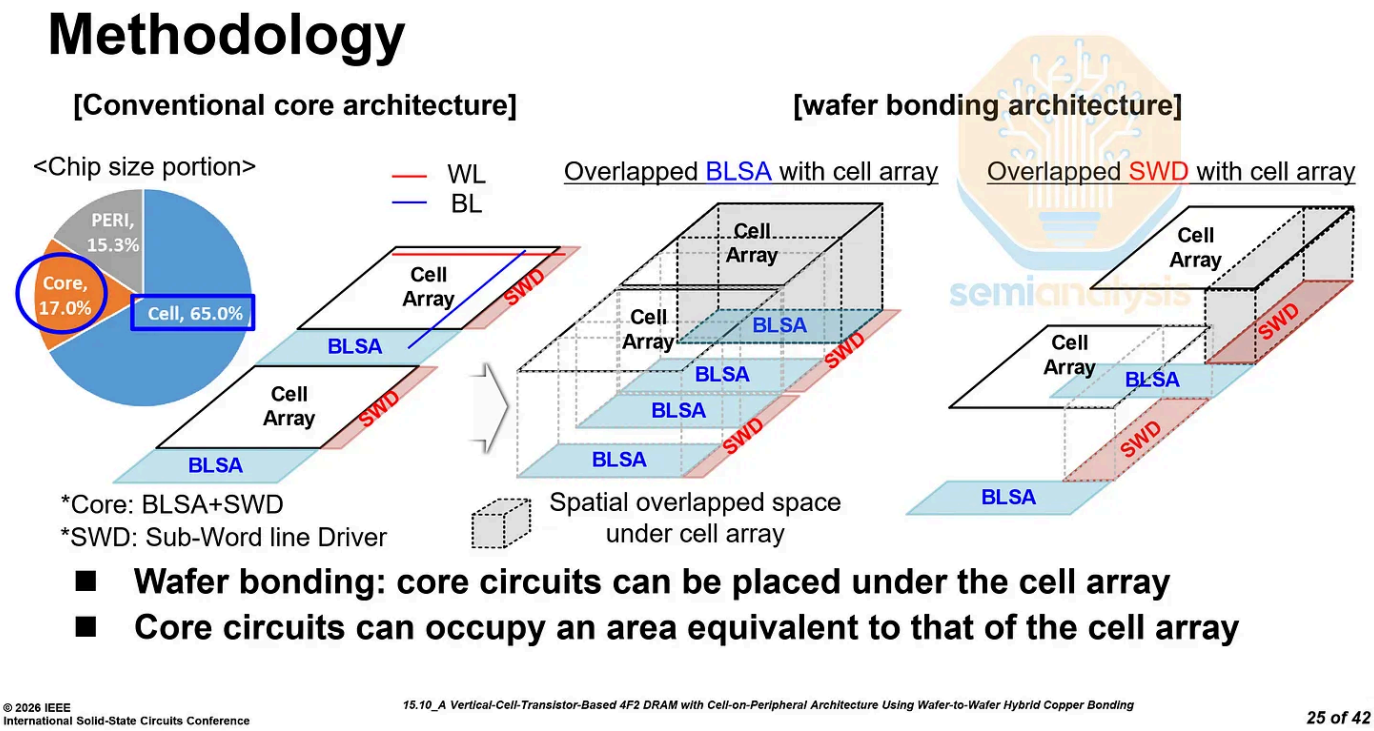
COP NOR-Type Sub-Wordline Driver Optimization. Source: Samsung, ISSCC 2026



COP Even/Odd Column Select MUX Optimization. Source: Samsung, ISSCC 2026

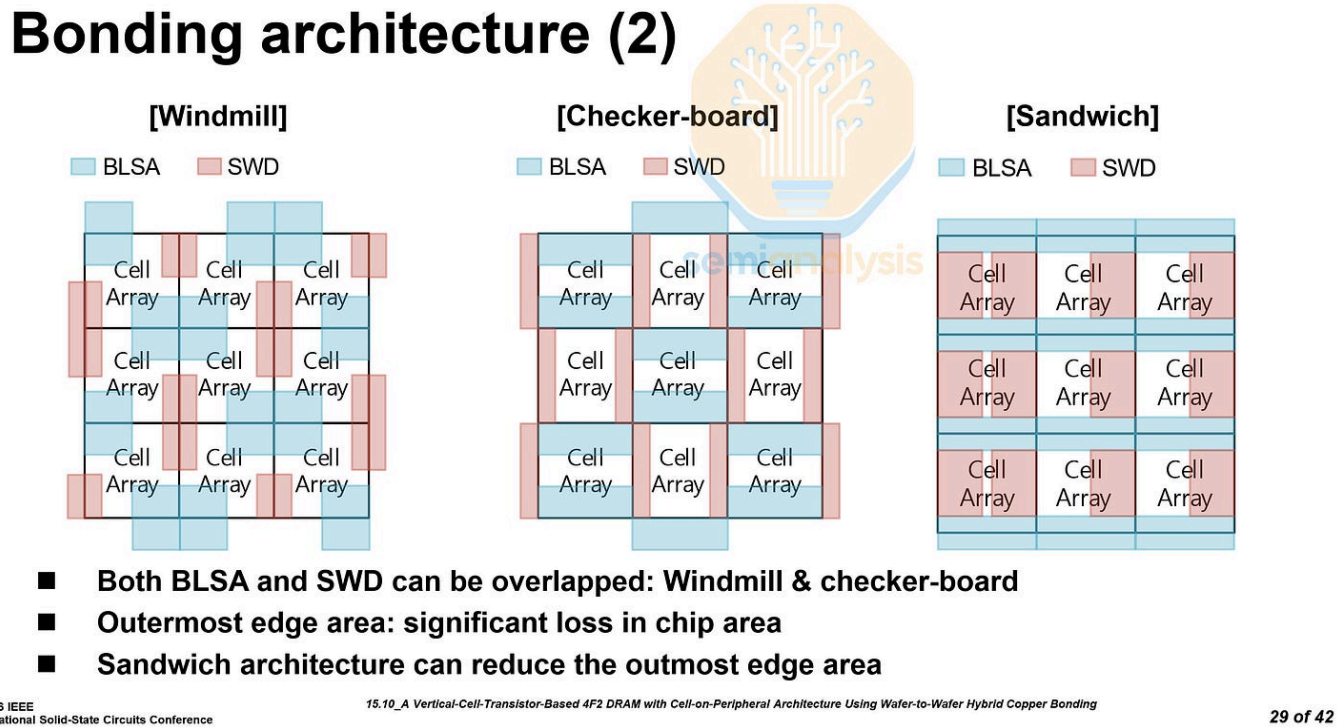
First, they have reorganized the sub-wordline drivers (SWD) from 128 per cell block to 16 groups of 8. This reduces the number of signals required for the SWD by 75%.

Next, they split the column select into an even and an odd path. This requires twice the multiplexers (MUX) but halves the column select line (CSL) count to 32 per data pin.



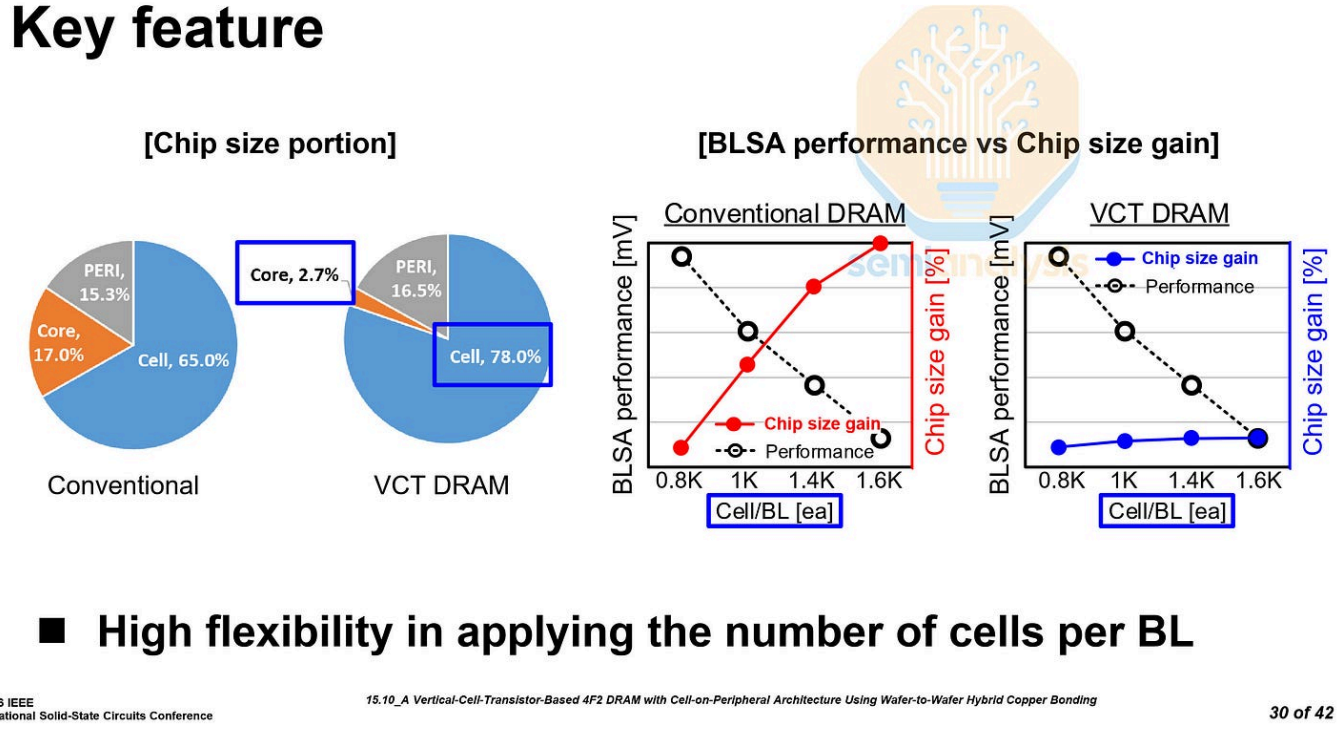
COP Core Circuitry Layout Under Cell Array. Source: Samsung, ISSCC 2026

With hybrid bonding, the core circuitry, that is, bitline sense amplifiers (BLSA) and SWD can be placed under the cell array. The goal is for the core circuits to occupy the same area as the cell array to increase overall density.



COP Core Circuitry Layout Options. Source: Samsung, ISSCC 2026

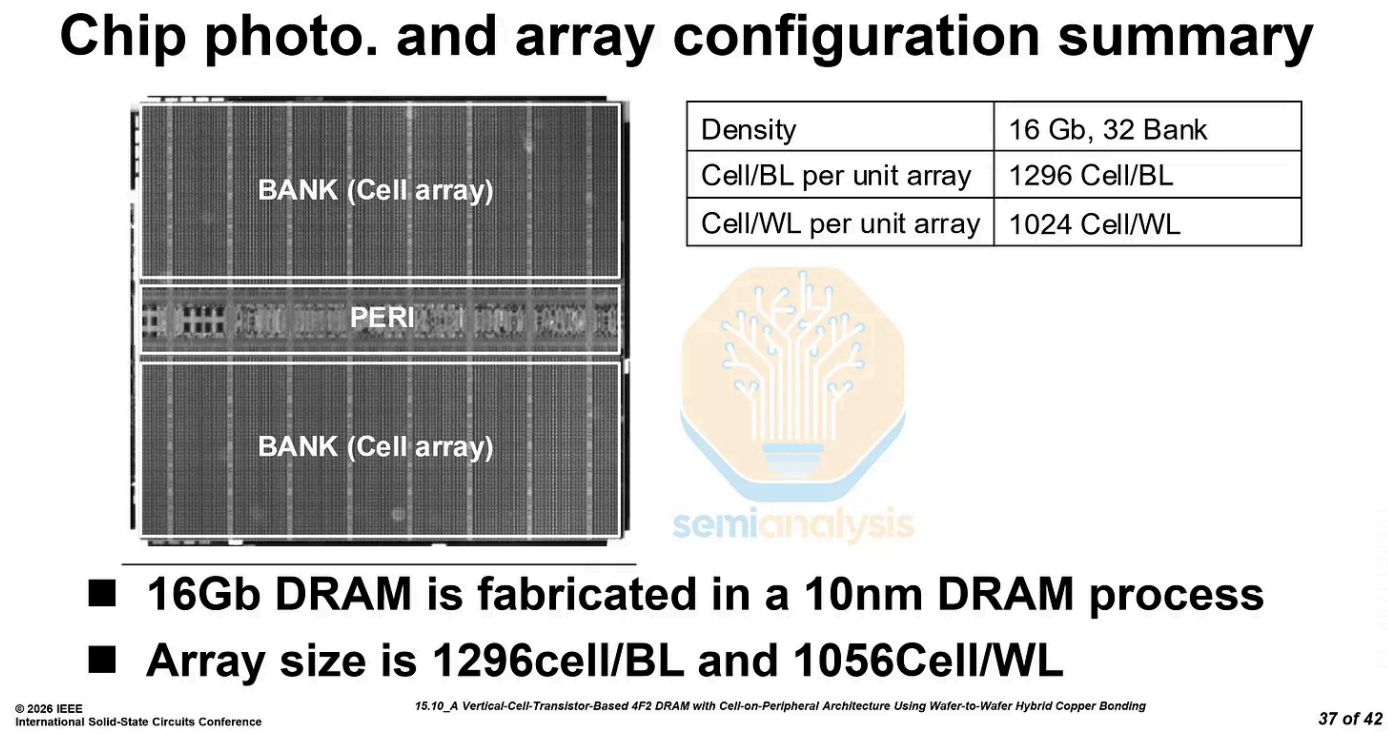
Samsung adopted the "sandwich" structure, which allows them to maximize the area efficiency of the core circuitry, and reduce the edge region area, which is not under any cells.



COP Sandwich Structure Area Efficiency. Source: Samsung, ISSCC 2026

The area used by the core circuitry was reduced from 17.0% down to only 2.7%, a significant improvement, directly translating to overall die size reductions.

In traditional DRAM, increasing the number of cells per bitline would result in a significant increase in chip area, while for VCT DRAM, the increase is almost negligible as the core circuitry is all below the cells.



Samsung 4F² COP DRAM Summary and Die Shot. Source: Samsung, ISSCC 2026

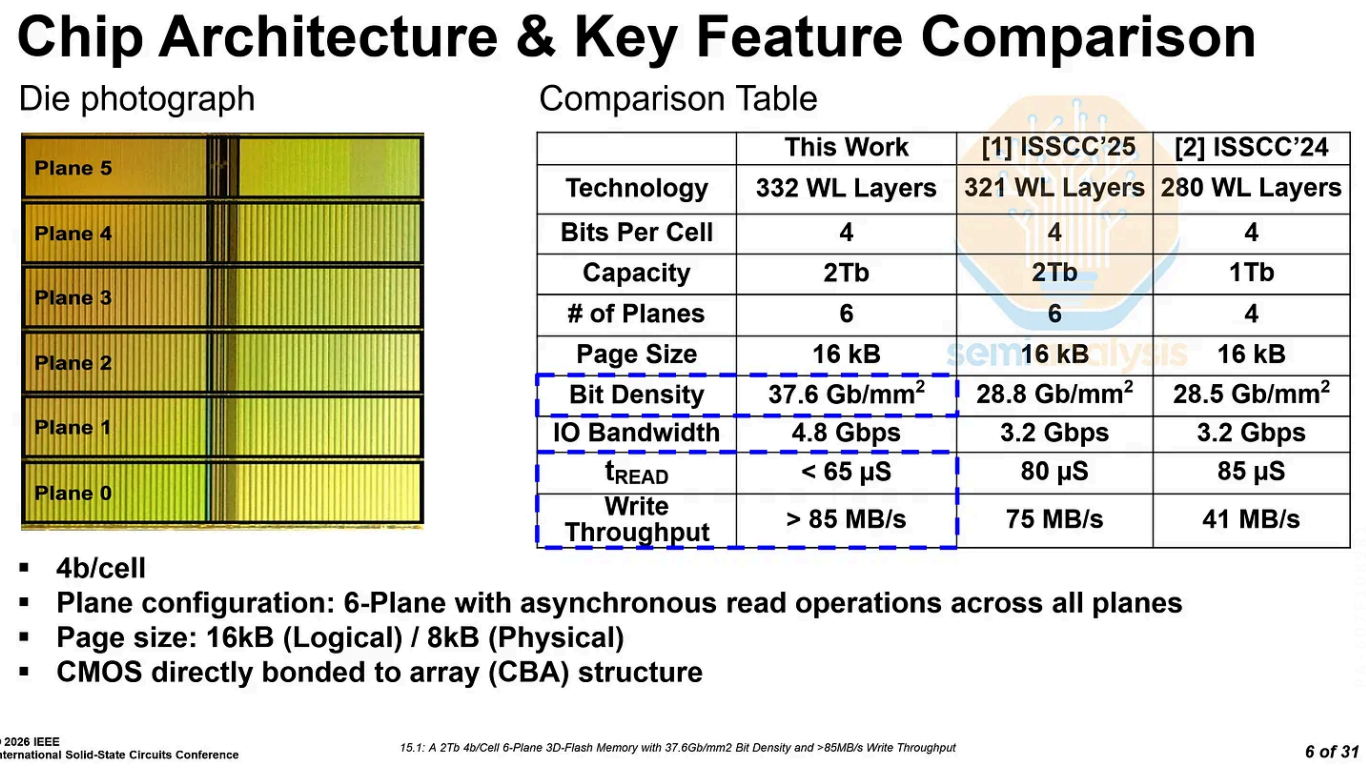
Samsung did not provide any density figures for this chip, only stating that it was a 16 Gb chip on a 10nm DRAM process.

Samsung noted that the VCT DRAM suffers from the floating-body effect, increasing leakage and reducing retention time. Mitigating this effect remains a key challenge for 4F² adoption.

Despite these challenges, we still expect 4F² hybrid bonded DRAM to arrive in the latter part of the decade as early as the generation after 1d. Our memory model tracks the timing and ramp of each node in detail. The current memory pricing landscape largely incentivizes the ramp and introduction of new nodes with higher bit densities to improve bit output per fab. On the otherhand, performance/$ of memory is highly sought after more than capacity for many use cases.

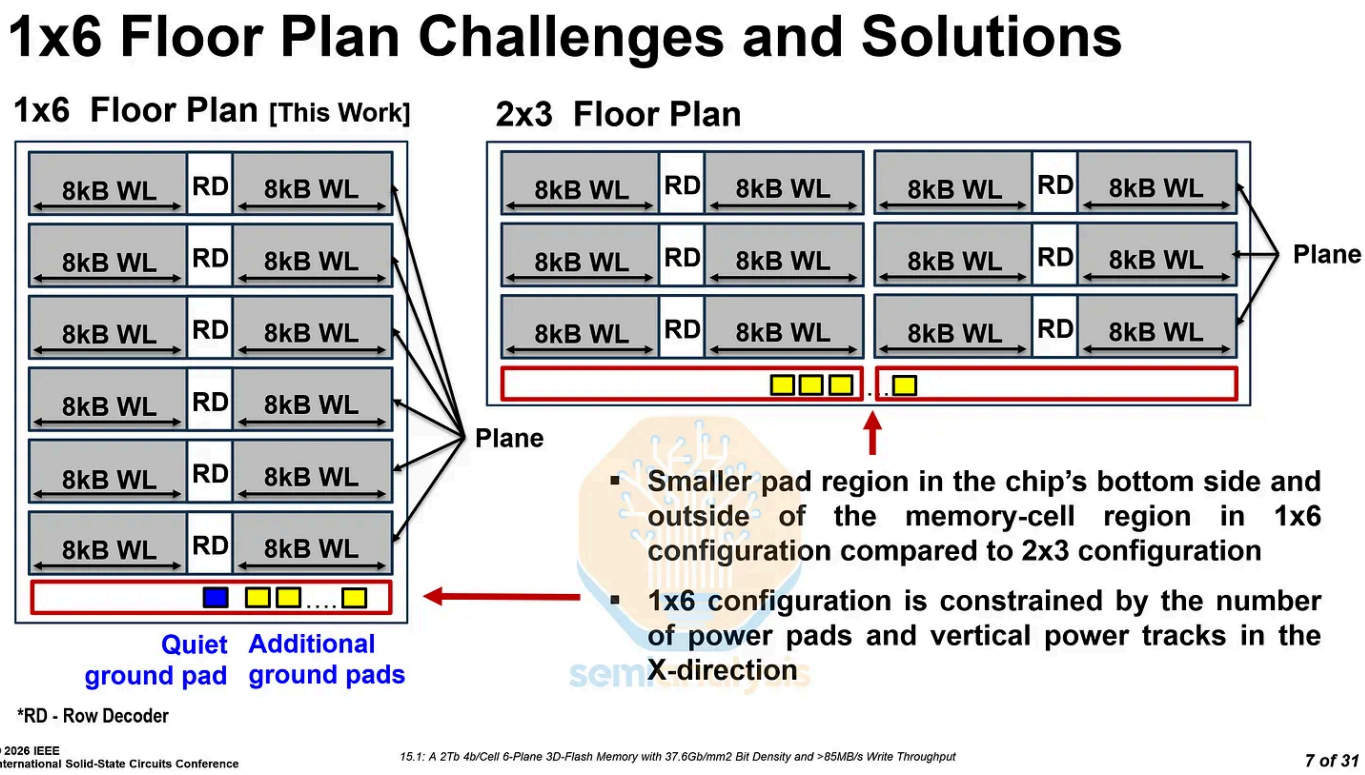
SanDisk/Kioxia BiCS10 NAND - Paper 15.1

SanDisk and Kioxia demonstrated their BiCS10 NAND, with 332 layers and 3 decks. This is the highest reported NAND bit density, at 37.6 Gb/mm², dethroning the previous champion, SK Hynix's 321L V9.



BiCS10 Die Shot and Density Comparison vs. SK Hynix and Samsung V9. Source: SanDisk/Kioxia, ISSCC 2026

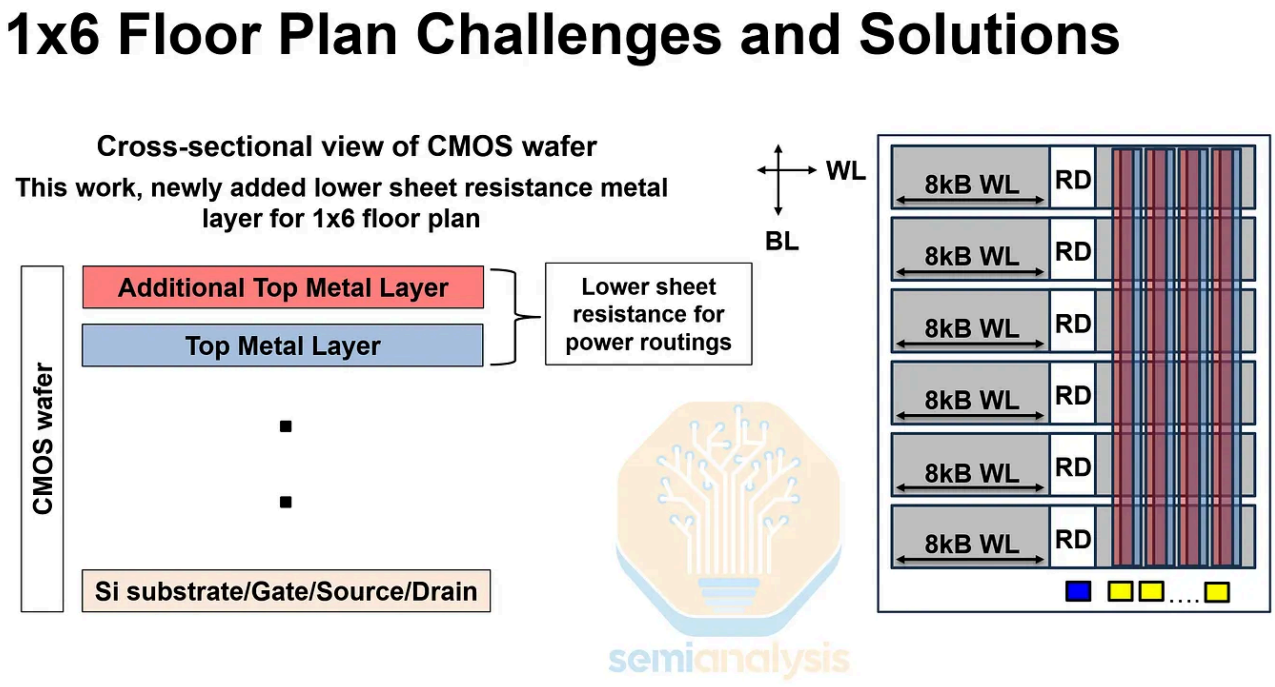
Despite using a similar architecture with 6 planes, 3 decks, and a similar number of layers, SK Hynix falls behind, with 30% lower bit density. In a QLC configuration, BiCS10 has a bit density of 37.6 Gb/mm², while SK Hynix's V9 has a bit density of only 28.8 Gb/mm². While in a TLC configuration, the densities are 29 and 21 Gb/mm² respectively, another example of SK Hynix's trailing position.



NAND 1×6 vs. 2×3 Plane Configuration Comparison. Source: SanDisk/Kioxia, ISSCC 2026

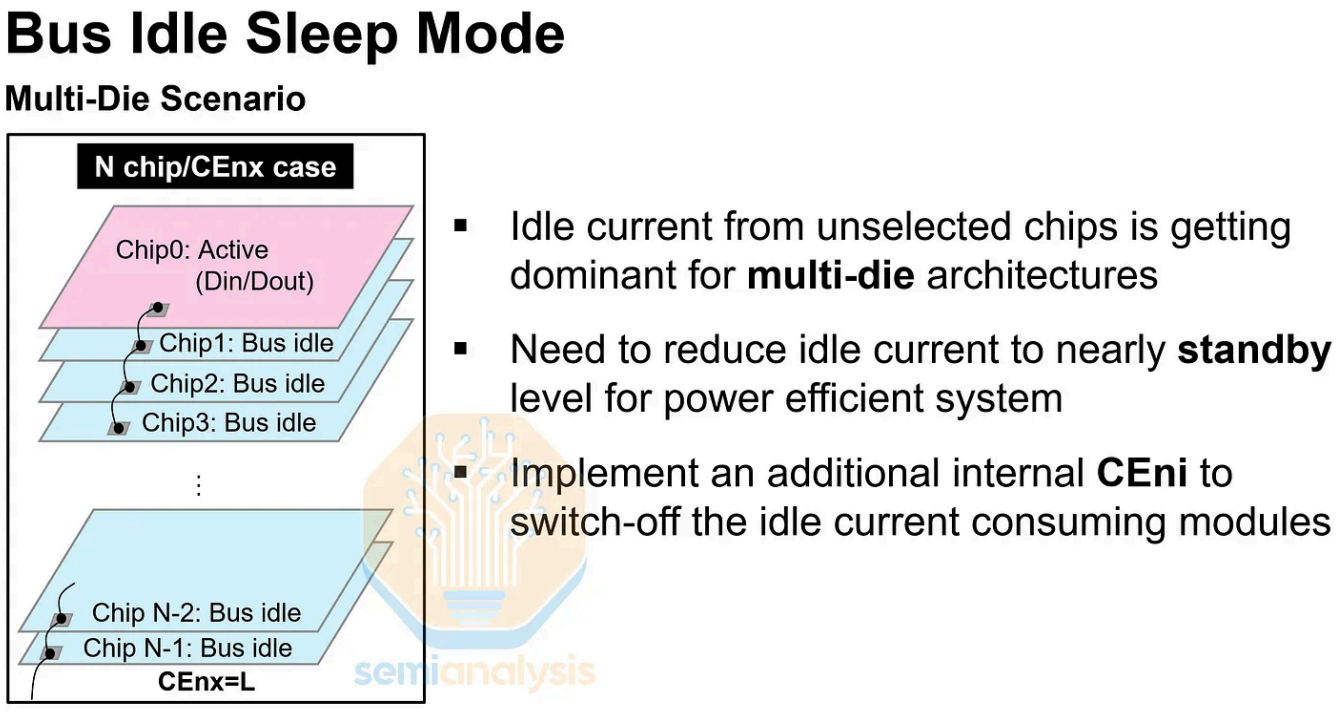
Moreover, BiCS10 features a 6-plane configuration, increasing IO bandwidth by 50%. There are two ways to implement a 6-plane configuration, 1×6 and 2×3. SK Hynix chose to use a 2×3 configuration, while SanDisk and Kioxia have decided to use a 1×6 configuration.

A 1x6 configuration has fewer ground pads and reduces area by 2.1%. However, the lower number of ground pads and vertical power tracks constrains power distribution.



BiCS10 CBA Additional Top-Metal Layer for Power Distribution. Source: SanDisk/Kioxia, ISSCC 2026

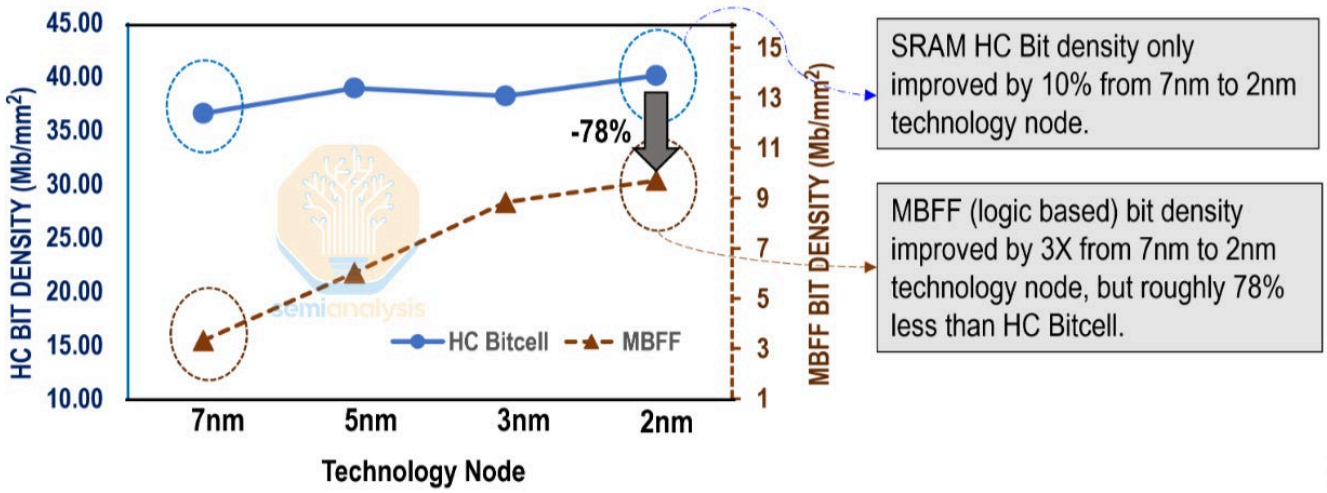
By using the CBA (Cell Bonded Array) architecture, SanDisk and Kioxia are able to customize the CMOS wafer process. By adding another top-metal layer in parallel to the existing one, they created a stronger power grid and overcame their power distribution constraints.



Multi-Die NAND Idle Power Penalty and Die-Gating Solution. Source: SanDisk/Kioxia, ISSCC 2026

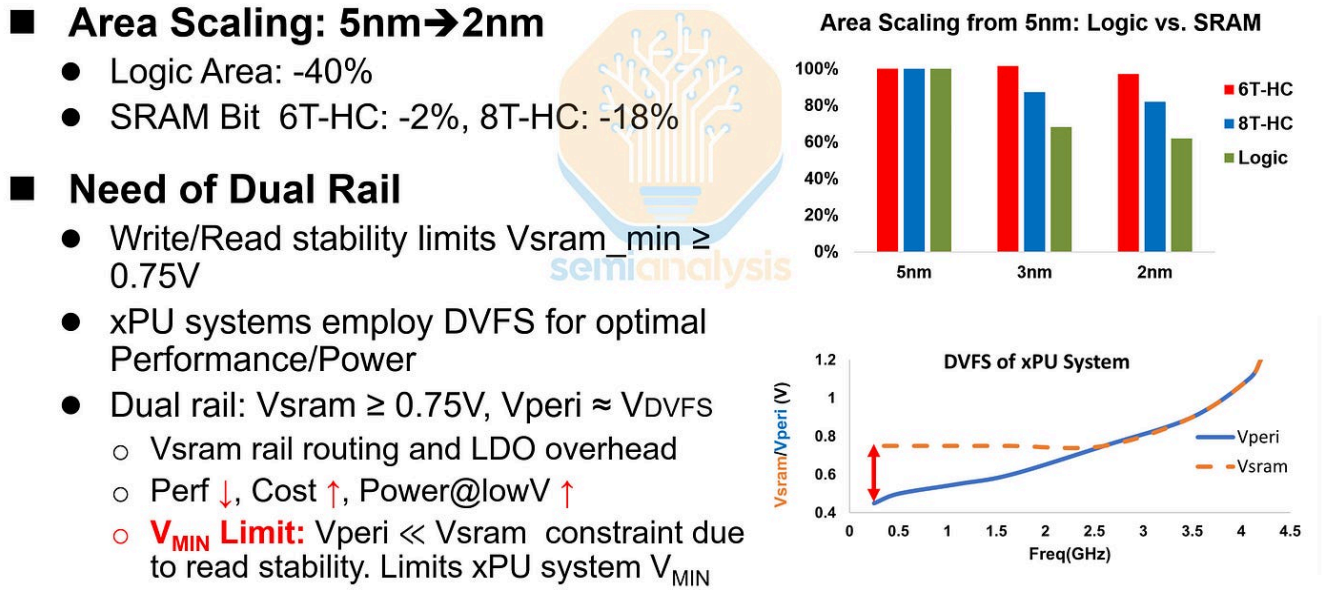
Stacking more dies is essential to increasing storage density. However, in multi-die architectures, the idle current from unselected dies is approaching the active current of the selected die. SanDisk implemented a gating system to fully shut down the data path of the unselected dies, reducing the idle current by two orders of magnitude.

MediaTek xBIT Logic-based Bitcell - Paper 15.2



SRAM HC Bitcell Density vs. Logic-Based MBFF Across Nodes. Source: MediaTek, ISSCC 2026

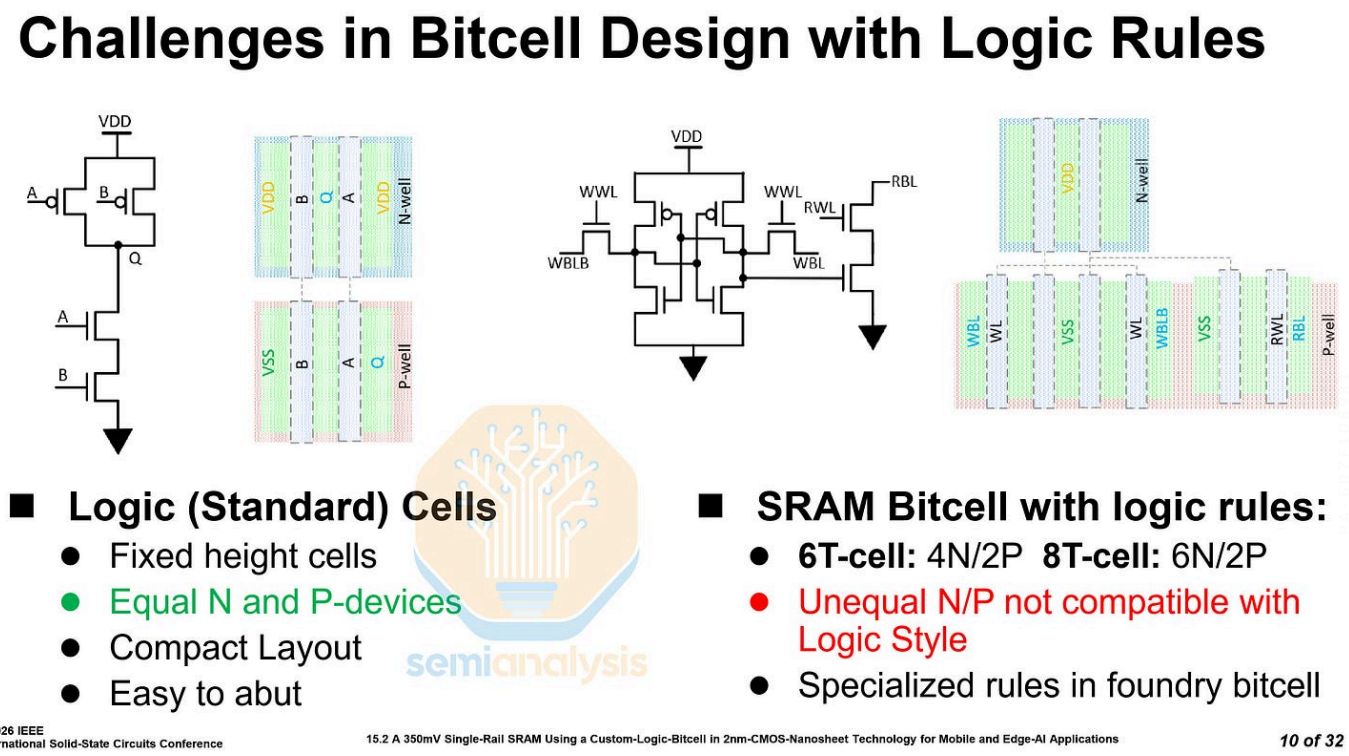
SRAM Bitcell Limitations: Scaling Issues



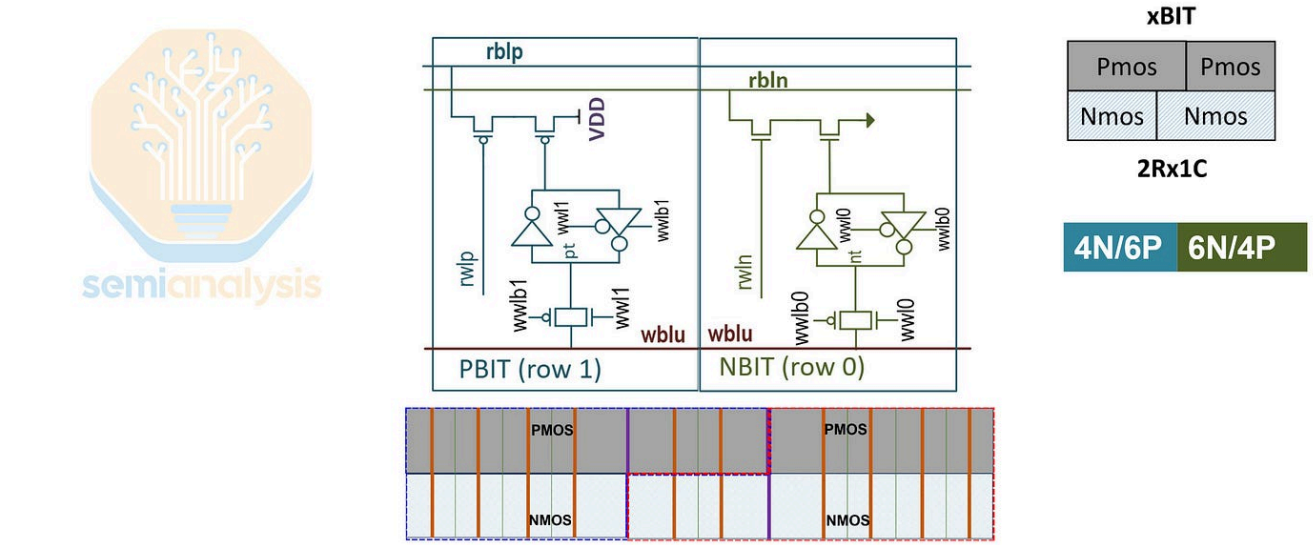
SRAM Bitcell Scaling Limitations: Area and Voltage Constraints. Source: MediaTek, ISSCC 2026

SRAM scaling is dead. Despite logic area decreasing by 40% from N5 to N2, 8-transistor high-current SRAM bitcells have only decreased in area by 18%. 6-transistor high-current (6T-HC) bitcells are even worse, only decreasing by 2%. Assist circuitry has scaled more, but it's not free lunch.

It is well known that N3E's high-density bitcell is a regression from N3B's, falling back to N5's density. In this paper, MediaTek shed some light on the high-current bitcell. N3E's high-current bitcell increased in area by 1-2% over N5. The density decreased from ~39.0 Mib/mm² to ~38.5 Mib/mm². Do note these figures do not account for assist circuitry overhead.



8T Bitcell NMOS/PMOS Layout Challenges with Logic Rules. Source: MediaTek, ISSCC 2026



MediaTek 10T xBIT Balanced Bitcell Circuit Design. Source: MediaTek, ISSCC 2026

In modern logic nodes, 6T bitcells have 4 NMOS and 2 PMOS transistors, while 8T bitcells have 6 and 2 respectively. The unequal number of NMOS and PMOS transistors requires specialized rules and makes layout more inefficient. MediaTek's novel bitcell is a 10-transistor cell, named the xBIT, with 4 NMOS and 6 PMOS transistors or vice versa. The two variants of the bitcell can be arranged together into a rectangular block, with 20 transistors, storing 2 bits.

xBIT SRAM Density and Energy Comparison

Density Comparison (Mux:1, Rows/BL:32)
xBIT vs. Foundry 8T Bitcell

Tech: 2nm	2-Port (1R1W)	
WordsxBits	8T (Mb/mm^2)	This Work (Mb/mm^2)
32x256	9.29	15.15
64x256	13.23	18.51
128x256	14.73	20.07
256x256	17.16	21.04

Energy Comparison (Mux:1, Rows/BL:32)
xBIT vs. Foundry 8T Bitcell

Tech: 2nm	2-Port (1R1W)	
Corner	Instance: 256x256	%(Gain Vs. 8T)
TT_0.5v_25C	Leakage (µW)	-29%
	Dynamic (pJ)	-31%
TT_0.9v_25C	Leakage (µW)	-9%
	Dynamic (pJ)	-36%

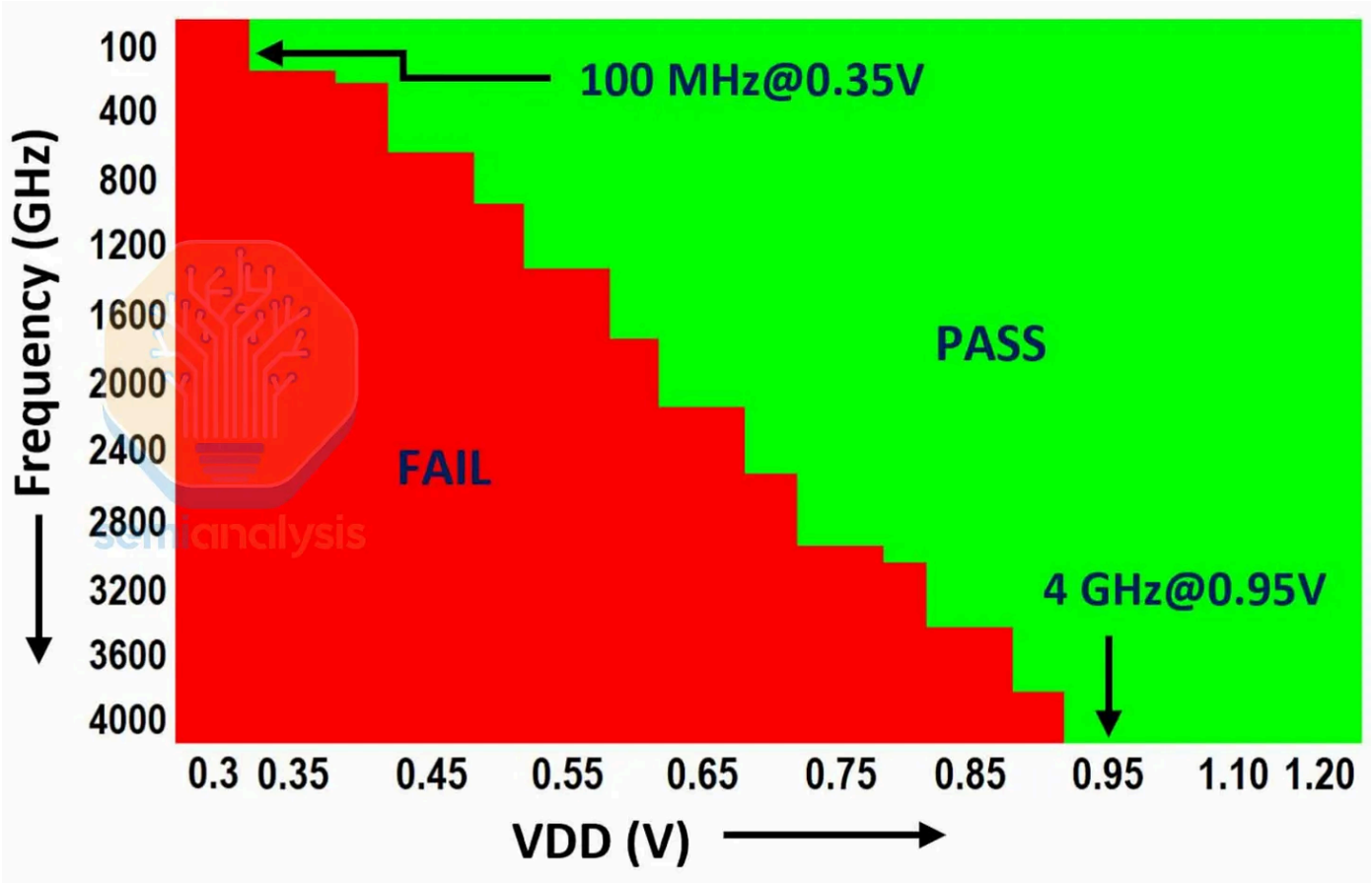
■ **SRAM Comparison: xBIT (This work) vs. 8T (In-house Compiler)**

- xBIT SRAM achieves higher density across: +22%(256x256), +63%(32x256)
- Dynamic power (Avg: R/W) reduced by ~30% and Leakage reduced by 29% at 0.5V
- High-V (0.9V): Similar GHz as 8T. Low-V (0.5V): 16% slower than 8T, but on-par with xPU logic speed and satisfy DVFS

© 2026 IEEE
International Solid-State Circuits Conference

xBIT vs. Foundry 8T Density and Power Comparison. Source: MediaTek, ISSCC 2026

When compared to the PDK's standard 8T bitcells, the xBIT achieved 22% to 63% higher density, with the largest gains at lower wordline widths. Power has also improved greatly, with the average read/write power reduced by over 30%, and leakage reduced by 29% at 0.5V. At 0.9V, performance was similar to an 8T bitcell and at 0.5V, although it was 16% slower than the 8T bitcell, it was fast enough not to be the bottleneck in a processor, and the voltage range was large enough for voltage-frequency scaling.



xBIT Shmoo Plot. Source: MediaTek, ISSCC 2026

MediaTek also showed a shmoo plot of the xBIT cell, going from 100 MHz at 0.35V, up to 4GHz at 0.95V.

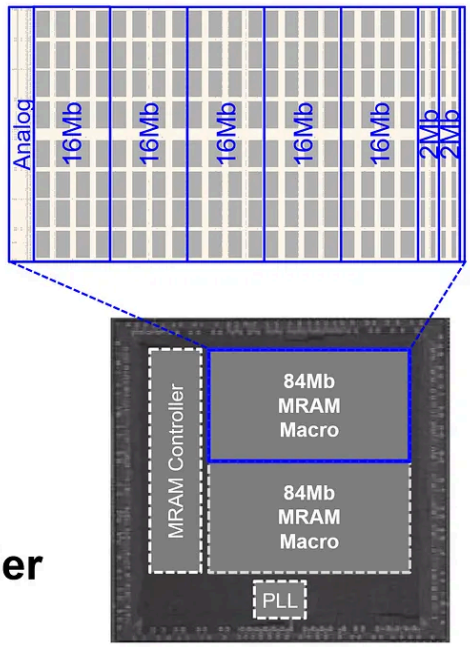
We will be doing a deep dive into SRAM and its scaling factors in an upcoming newsletter article.

TSMC N16 MRAM - Paper 15.4

TSMC presented an updated STT-MRAM on their N16 node, building on their previous work from ISSCC 2023. TSMC is positioning the MRAM as an embedded non-volatile memory (eNVM), for use in automotive, industrial, and edge applications, which do not need the most advanced technology but instead reliability.

Innovative Design Features

- **Monolithic Design**
  - "LEGO-like" module
- **Flexible Module Options**
  - 16Mb, 8Mb and 2Mb for macro integration
- **Dual Port Access**
  - Efficient Over-The-Air Updates
- **Interleaved Read**
  - Maximum read throughput
- **Coupling Cancellation Sense Amplifier**
  - Reliable and disturbance-free

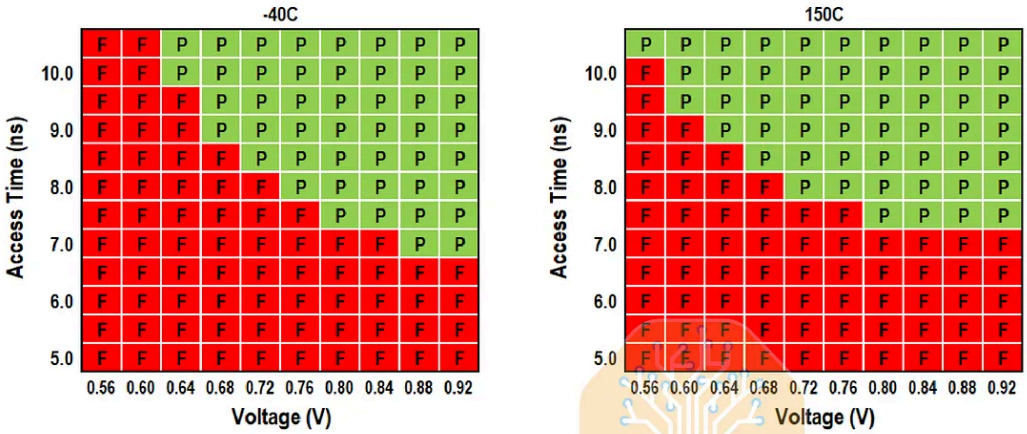


© 2026 IEEE
International Solid-State Circuits Conference

TSMC N16 MRAM Design Features and Die Floorplan. Source: TSMC, ISSCC 2026

The MRAM features dual-port access so reads and writes can occur simultaneously — critical for over-the-air (OTA) updates in automotive, where the system cannot halt reads while firmware is being written.

84Mb Macro Access Time

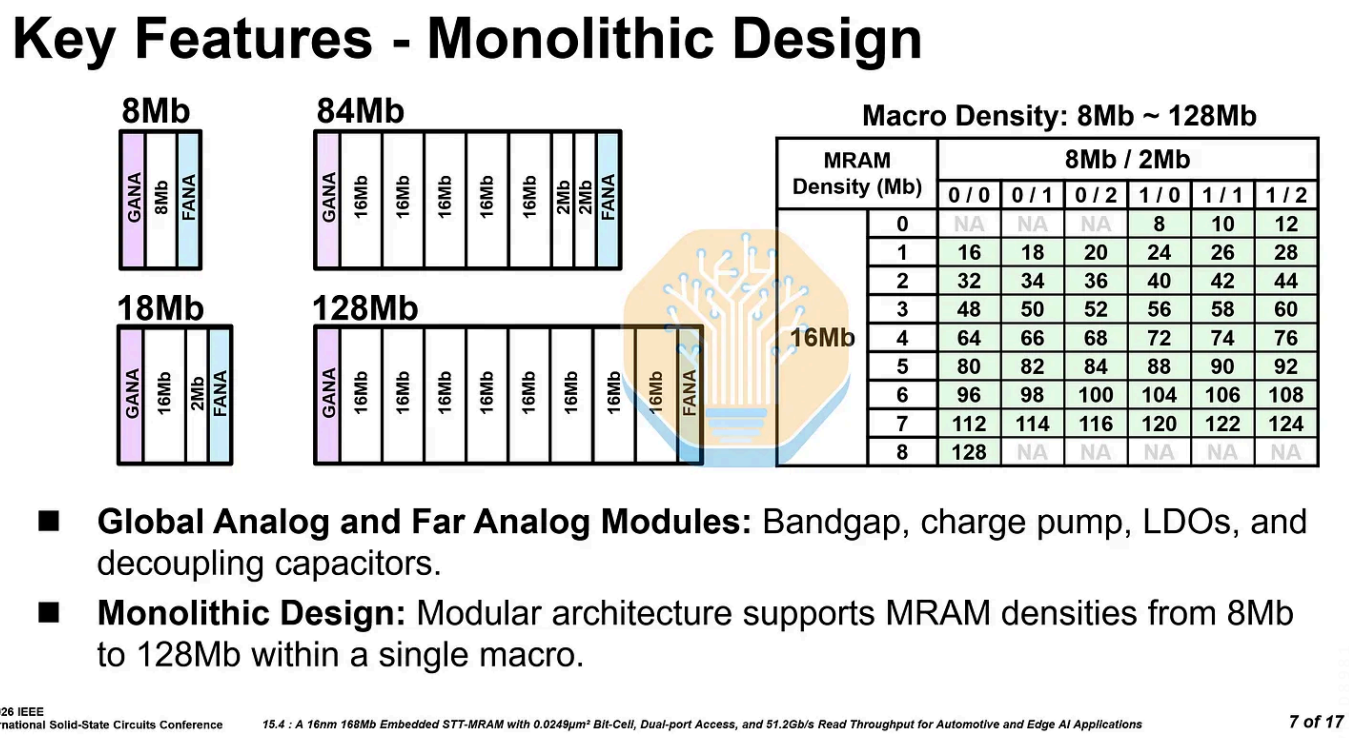


- **Read Speed:** Achieves 7.5 ns read speed at 0.8V across -40ºC to 150ºC for 84Mb macro density.
- **Read Throughput:** Toggling CLK[N:0] in pipeline boosts read frequency to 200 MHz, achieving 51.2 Gb/s throughput.

© 2026 IEEE
International Solid-State Circuits Conference

TSMC N16 MRAM Shmoo Plot at -40 °C and 150 °C. Source: TSMC, ISSCC 2026

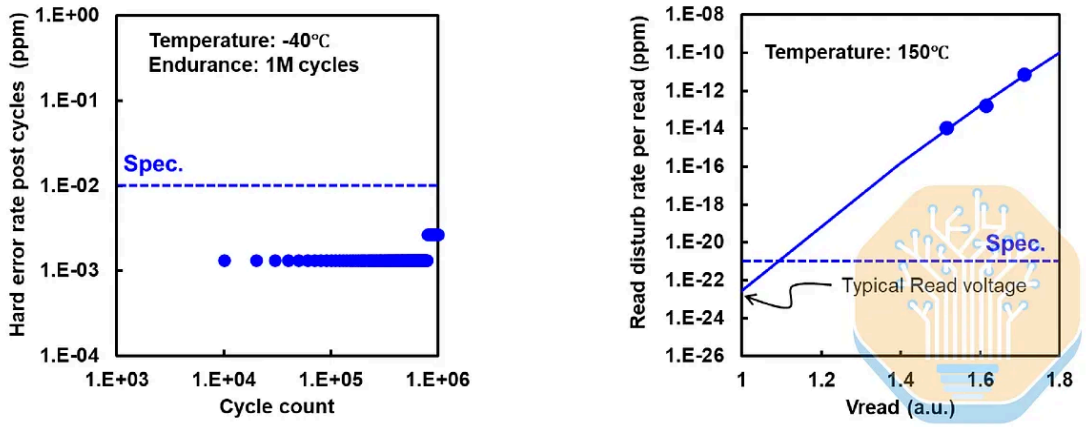
It features interleaved reads across modules with independent clocks, raising throughput to 51.2 Gb/s at 200 MHz. On silicon, the 84 Mb macro achieves 7.5ns read access time at 0.8V across -40 °C to 150 °C.



TSMC N16 MRAM Modular Macro Architecture. Source: TSMC, ISSCC 2026

The architecture is modular — configurable in 16 Mb, 8 Mb and 2 Mb modules that compose into macros from 8 Mb to 128 Mb. By combining large 16 Mb modules with a few smaller 2 Mb and 8 Mb modules, the capacity can be fine tuned to the needs of any design. For example, 5× 16 Mb modules and 2× 2 Mb modules form an 84 Mb macro.

Reliability Results



- **Hard error rate after 1M endurance cycles at -40˚C: <0.01 ppm.**
- **Read disturb rate per read at 150˚C: Extremely low, ensuring sufficient read margin.**

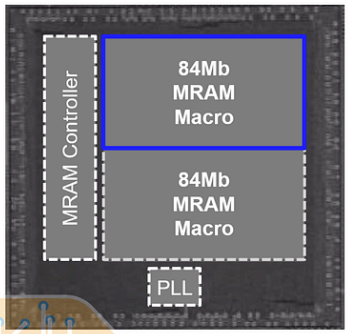
© 2026 IEEE
International Solid-State Circuits Conference

TSMC N16 MRAM Endurance and Reliability. Source: TSMC, ISSCC 2026

As stated earlier, reliability is where embedded MRAM lives or dies. After 1 million endurance cycles at -40 °C, the hard error rate stays well below 0.01 ppm — well within ECC correction range. Read disturb at 150 °C is below 10^{-22} ppm at typical read voltages, effectively negligible. The 168 Mb test chip passes reflow and supports 20-year retention at 150 °C, meeting stringent automotive requirements.

Key Features and Die Photo

	ISSCC 2023 [4]	This work
Technology	16nm MRAM	16nm MRAM
Physical Cell size (um^2)	0.033	0.0249
VDD (V)	0.72~0.88	0.72~0.88
VDIO (V)	1.62~1.98	1.62~1.98
Operation Temperature (Tj, ˚C)	-40˚C ~ 150˚C	-40˚C ~ 150˚C
Macro Density (Mb/mm^2)	12.8 (32Mb)	16.0 (32Mb) 14.8 (84Mb) 18.1 (128Mb)
Macro Read Speed (ns)	6ns (32Mb) -	5.5ns (32Mb) 7.5ns (84Mb)
4Mb Bank Read Access Time (ns)	-	3.97ns (4Mb)
Interleaved Read	N/A	200MHz @ 0.72V
Read Throughput (Gb/s)	32	51.2
Dual Port Access Feature	N/A	Yes
Endurance (cycle) (Code/Data)	1M	1M
Retention (yrs/˚C)	20/150	20/150
Reflow (120s @ 260˚C)	Yes	Yes

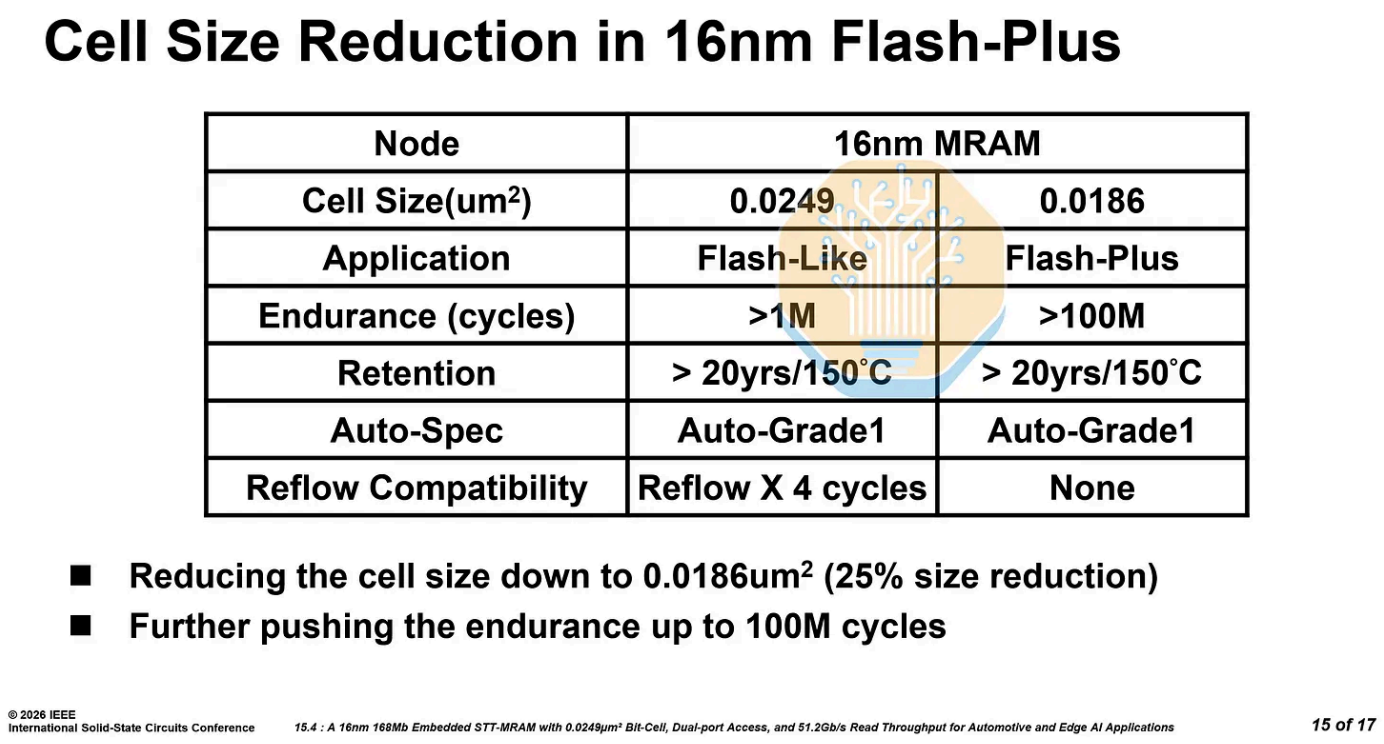


© 2026 IEEE
International Solid-State Circuits Conference

TSMC N16 MRAM Specification vs Prior Work. Source: TSMC, ISSCC 2026

Compared to the old MRAM on the same N16 node, the bitcell has shrunk 25% from 0.033 μm² to 0.0249 μm², and macro density increases to 16.0 Mb/mm² iso-capacity. The read speed drops from 6 ns to 5.5 ns iso-capacity, and the dual-port access and interleaved reads are entirely new.

While Samsung Foundry also published work on 8LPP eMRAM this year, TSMC's is far more promising. It targets the needed features, has great performance, and is on a cheaper N16 node.



TSMC N16 MRAM Flash-Plus Roadmap. Source: TSMC, ISSCC 2026

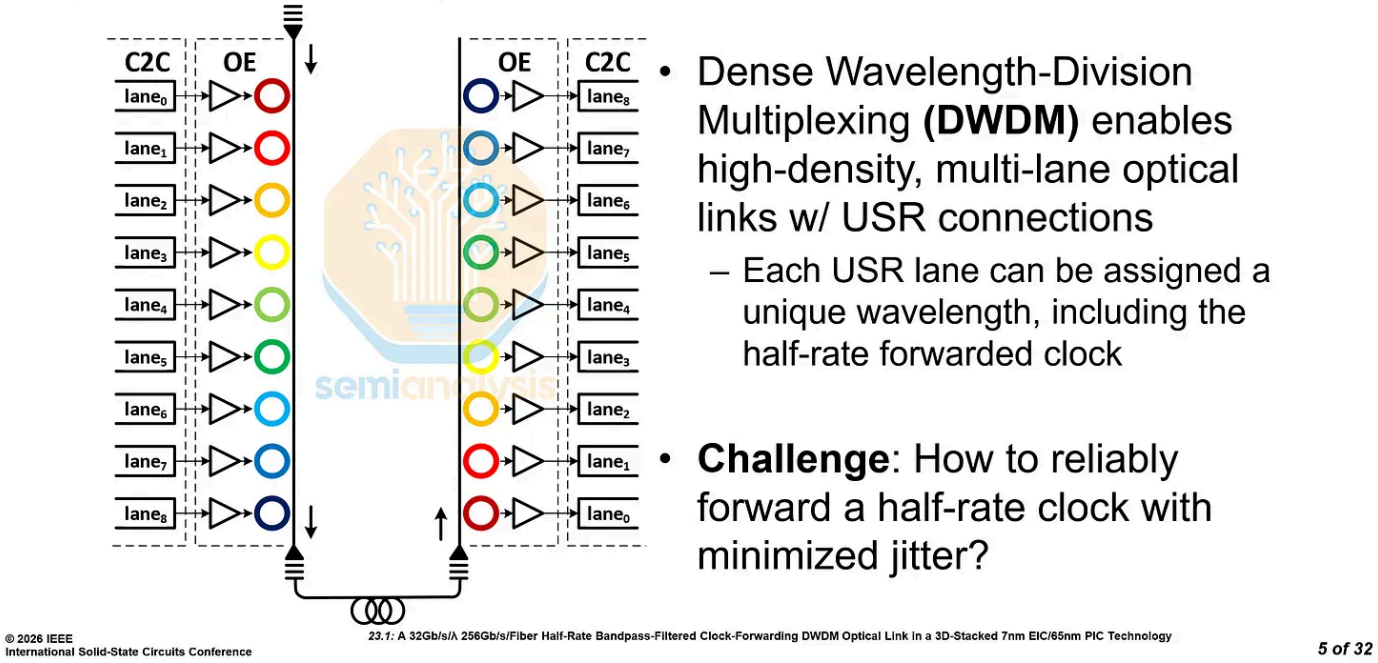
TSMC is already planning the next-generation “Flash-Plus” variant with a 25% smaller bitcell and 100× higher endurance.

Optical Networking

Several papers from major optics vendors tackled optical interconnects that will carry data between next-generation AI accelerators both within and between datacenters.

Nvidia DWDM - Paper 23.1

The choice of optical signaling format will affect the go-to-market timeline for scale-up co-packaged optics (CPO). Nvidia is ramping up production of COUPE optical engines that support 200G per lane PAM4 for scale-out switching in the near-term.



Nvidia DWDM Architecture Overview. Source: Nvidia, ISSCC 2026

However, at ISSCC, Nvidia proposed using 32 Gb/s per lambda, with 8 wavelengths multiplexed using DWDM. A 9th wavelength is used for clock forwarding at half rate — that is 16 Gb/s.

Clock forwarding means that SerDes can be made somewhat simpler by removing the Clock and Data Recovery (CDR) circuit as well as other circuitry, improving energy and chip shoreline efficiency.

Earlier in March, just before OFC 2026, the formation of the Optical Compute Interconnect MSA (OCI MSA) was announced, and it will focus on a 200 Gb/s Bi-directional link, with each of transmit and receive formed using 4 lambdas of 50G NRZ, which will be sent bi-directionally across the same fiber. Did I hear anyone mention OCS?

2. Optical Specifications

Each optical fiber port shall support 212.5 Gbps data rate bidirectionally with the transmit and receive signals on the same fiber at separate wavelength groups. The 212.5 Gbps electrical signals are formed by aggregating four DWDM 53.125 Gbps NRZ optical signals as described in the previous section. Figure 2-1 shows the DWDM wavelength channels and CWDM wavelength groups for transmitter and receiver supported by the same fiber. The Type A/B transceivers always transmit group A/B wavelengths which are specified in this section. A bidirectional link connects an A type transceiver on one side to a B type transceiver on the other end. While Figure 2-1 illustrates an example symmetric implementation with both Type A and Type B transceivers at both end points, this symmetry is not strictly required as part of the specification and an asymmetric physical implementation can also be compliant.

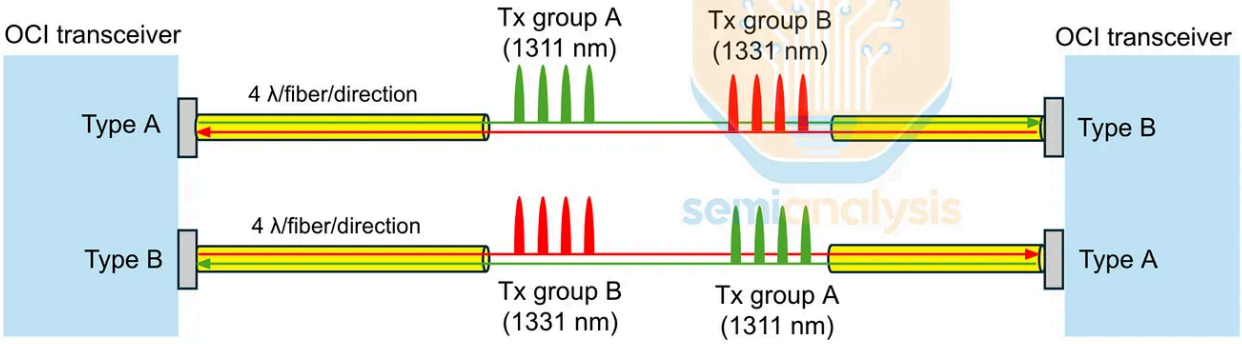


Figure 2-1. Illustration of DWDM wavelength channels and CWDM wavelength groups for a bidirectional link.

OCI MSA Optical Link Specifications. Source: OCI MSA

Interestingly, the OCI MSA did not make use of an additional wavelength for clock forwarding, and it appears that reserving all wavelengths for actual data transmission is the priority.

Much of Nvidia's published research on scale-up CPO has centered on DWDM, though today's CPO optical engines are oriented around 200G PAM4 DR Optics, which makes more sense for scale-out networking. The OCI MSA centering around DWDM for scale-up optics resolves this apparent contradiction, as it is now clear that Nvidia and others will center around the use of DWDM for scale-up and DR Optics for scale-out.

The OCI MSA also illustrates different implementations, an On-Board Optics (OBO), a version of CPO that is integrated via the substrate on the ASIC package, and a version where the optical engine is integrated directly on the interposer. The implementation illustrated in the middle figure (b) will be the most common one used for scale-up and scale-out CPO for the next few years, but it still requires some form of serialized links that can pass through the ASIC substrate and will still require some form of SerDes on both sides. For example, UCIe-S could be used as a protocol for such transmission.

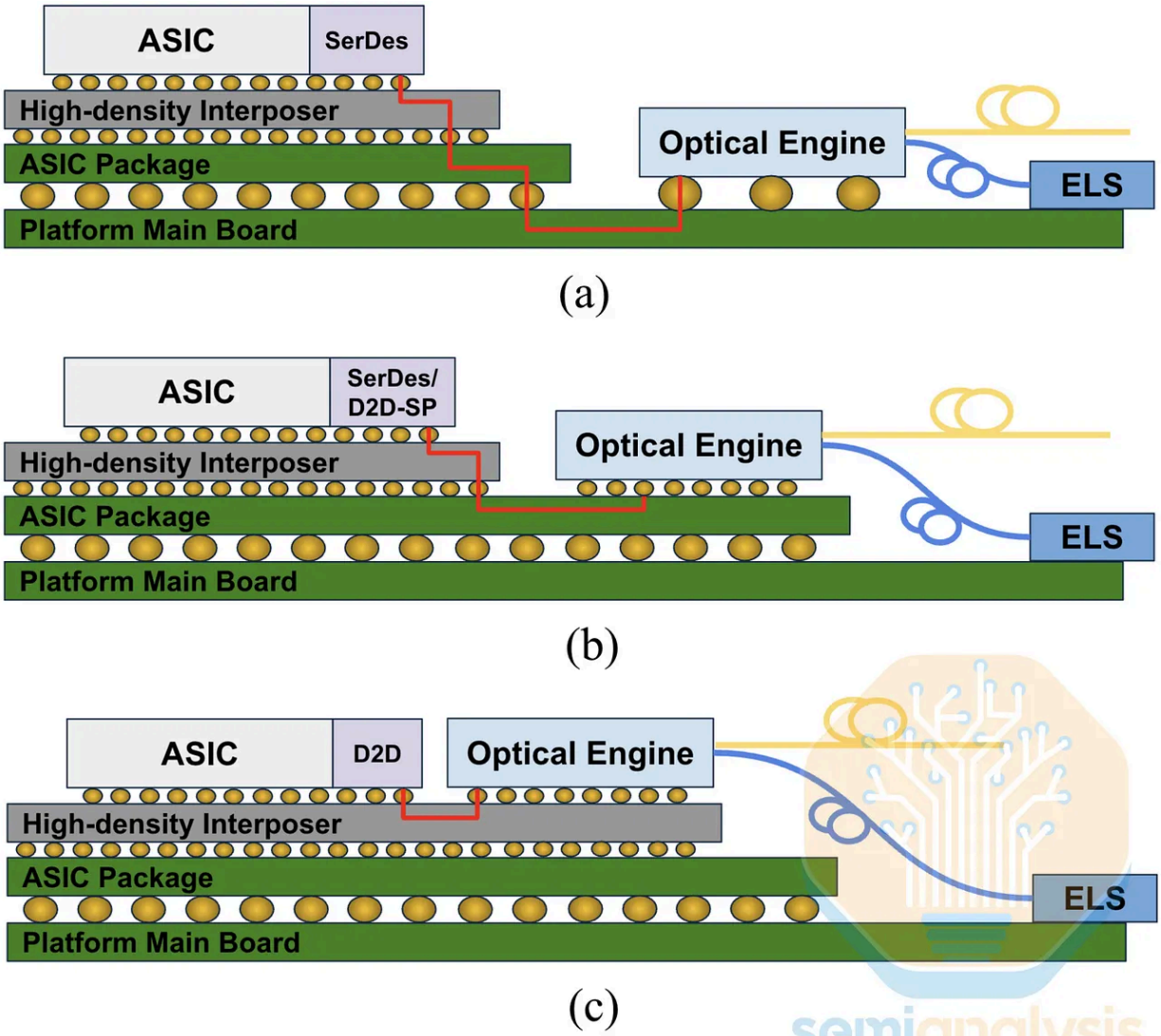
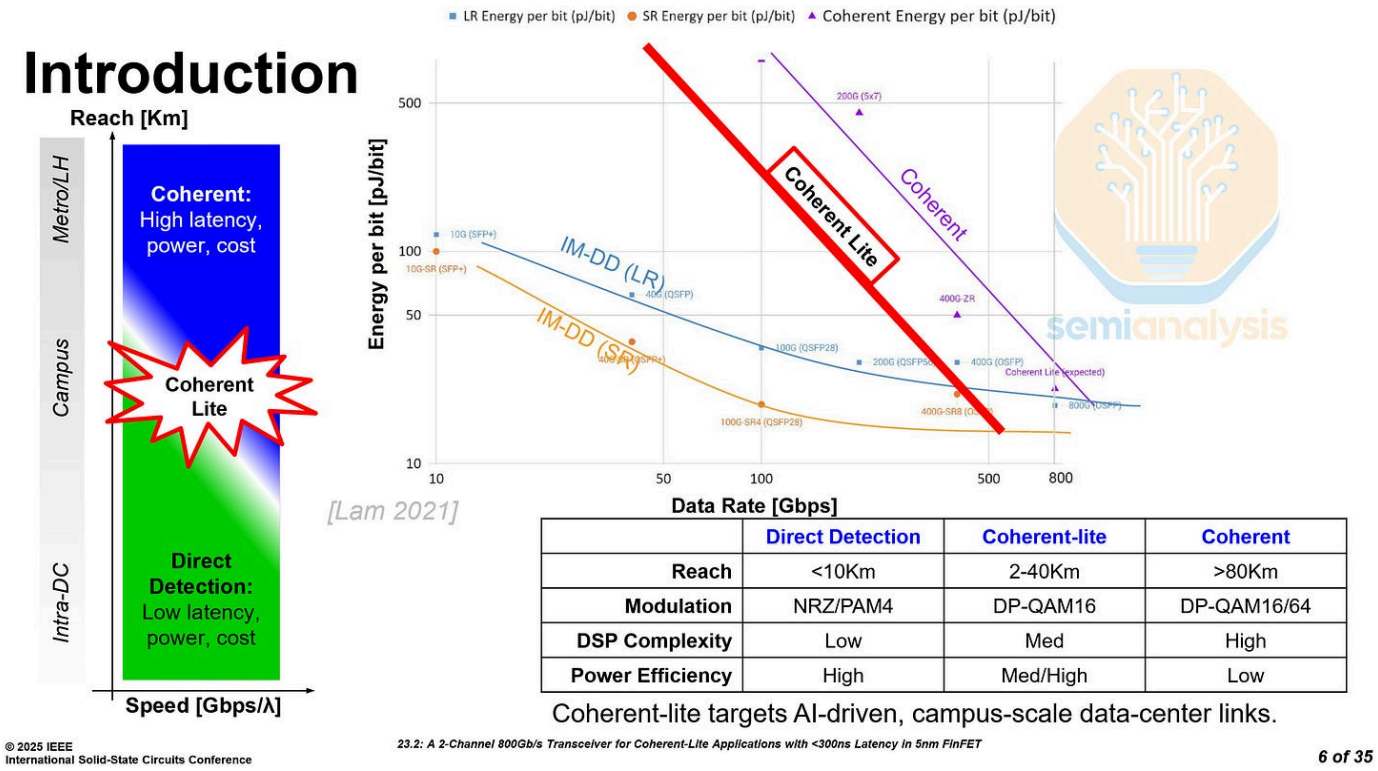


Figure 1. Illustration of three different physical OCI implementations highlighting the different levels of integration with the ASIC; where (a) shows an On-Board Optics (OBO) implementation, (b) shows a package integration implementation and (c) shows an interposer integrated implementation.

Optical Engine Integration Levels (OBO, Substrate CPO, Interposer CPO). Source: OCI MSA

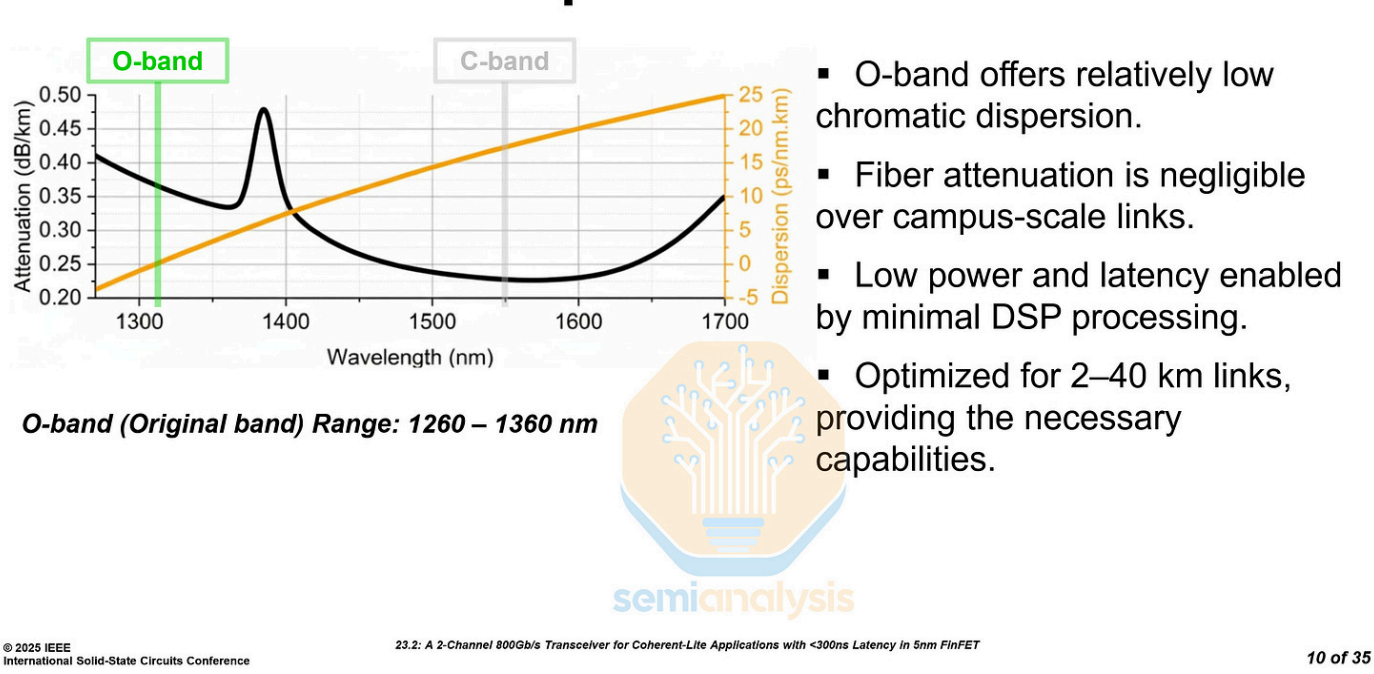
The “Final Boss” when it comes to implementing CPO will be when the optical engine can be integrated onto the interposer itself, connecting to the ASIC using a parallelized die to die (D2D) connection as depicted in (c) above. This could considerably improve shoreline bandwidth density, enable much higher radix and improve energy efficiency. This implementation thus unlocks benefits of CPO in ways that the other implementations cannot, but achieving it is still a few years away and requires further improvements in advanced packaging technology.

Marvell Coherent-Lite Transceiver - Paper 23.2



Direct Detection vs. Coherent-Lite vs. Coherent Transceiver Comparison. Source: Marvell, ISSCC 2026

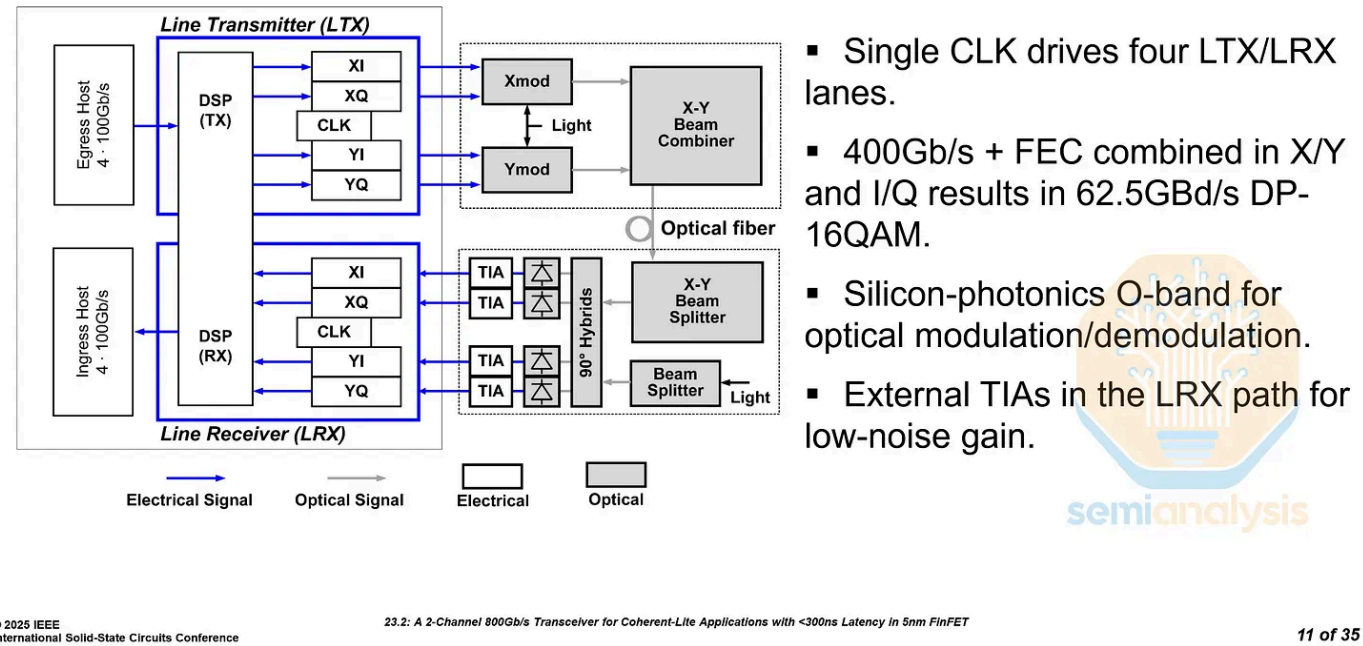
Marvell presented an 800G transceiver for coherent-lite applications. Traditional transceivers have a limit on how far they can reach, less than 10 kilometers. Coherent transceivers support much further reach, but they are complex, consume more power, and are more expensive. Marvell's Coherent-lite transceiver targets a middle ground with respect to power consumption, cost and range, which is perfect for large datacenter campuses with links spanning at most tens of kilometers.



Coherent and Coherent-Lite Optical Band Comparison. Source: Marvell, ISSCC 2026

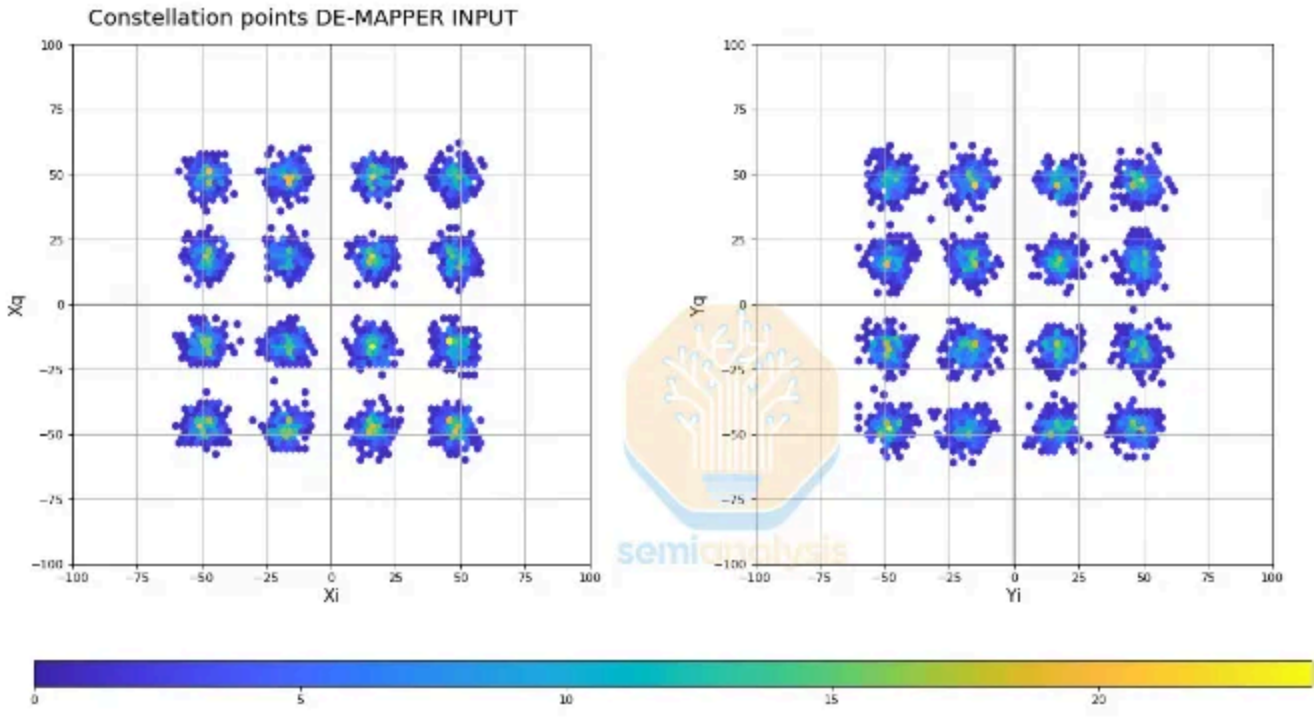
Coherent transceivers primarily use C-band wavelengths for their low attenuation. However, the long-haul links in which coherent transmission is used typically have a very high dispersion, needing heavy DSP processing. The long range of traditional Coherent optics can often be overkill for datacenter campuses with buildings that are only tens of kilometers apart.

Coherent-Lite transceivers instead use O-band wavelengths, which have near-zero dispersion over the relatively short distances on a datacenter campus. This enables minimal DSP processing, saving power and reducing latency.



Marvell Coherent-Lite Transceiver Architecture. Source: Marvell, ISSCC 2026

The Coherent-lite transceiver is a DSP-based pluggable module consisting of two 400G channels. Each 400G channel runs a dual polarization QAM and consists of two parallel modulation streams, X and Y.

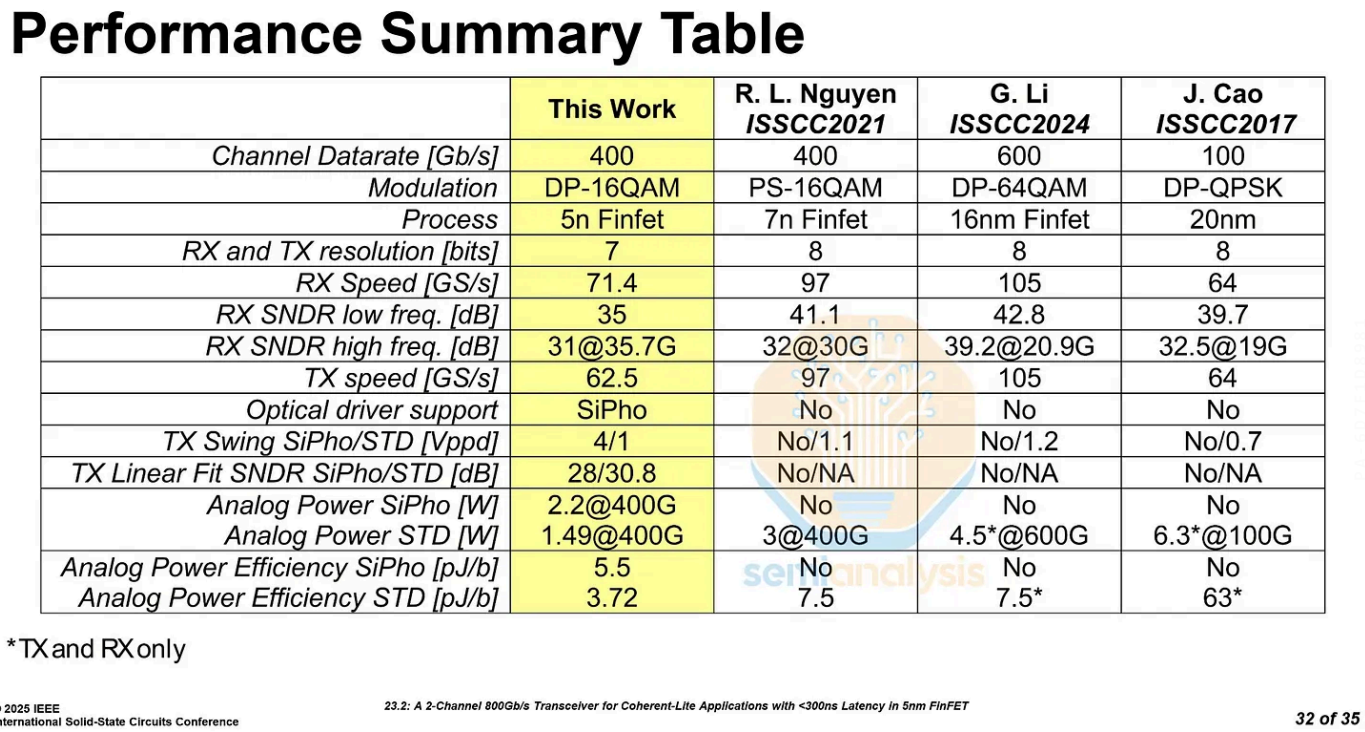


Marvell Coherent-Lite Measured Link Performance. Source: Marvell, ISSCC 2026

The key to this demonstration is highlighting other methods of scaling channel bandwidth that are optimized for the campus application.

Higher-order modulation coupled with the dual polarization using X and Y axes delivers 400G channel bandwidth. There are 8 bits per channel for a total of 32 constellation points as demonstrated above. These 8 bits times the 62.5 GBd signal rate equals ~400G of total bandwidth.

This modulation scheme is not entirely new to the industry, but it is now being brought into the datacenter campus setting for use in those shorter links.



Performance Summary Table

	This Work	R. L. Nguyen *ISSCC2021*	G. Li *ISSCC2024*	J. Cao *ISSCC2017*
Channel Datarate [Gb/s]	400	400	600	100
Modulation	DP-16QAM	PS-16QAM	DP-64QAM	DP-QPSK
Process	5n Finfet	7n Finfet	16nm Finfet	20nm
RX and TX resolution [bits]	7	8	8	8
RX Speed [GS/s]	71.4	97	105	64
RX SNDR low freq. [dB]	35	41.1	42.8	39.7
RX SNDR high freq. [dB]	31@35.7G	32@30G	39.2@20.9G	32.5@19G
TX speed [GS/s]	62.5	97	105	64
Optical driver support	SiPho	No	No	No
TX Swing SiPho/STD [Vppd]	4/1	No/1.1	No/1.2	No/0.7
TX Linear Fit SNDR SiPho/STD [dB]	28/30.8	No/NA	No/NA	No/NA
Analog Power SiPho [W] *Analog Power STD [W]*	2.2@400G 1.49@400G	No 3@400G	No 4.5*@600G	No 6.3*@100G
Analog Power Efficiency SiPho [pJ/b] *Analog Power Efficiency STD [pJ/b]*	5.5 3.72	No 7.5	No 7.5*	No 63*

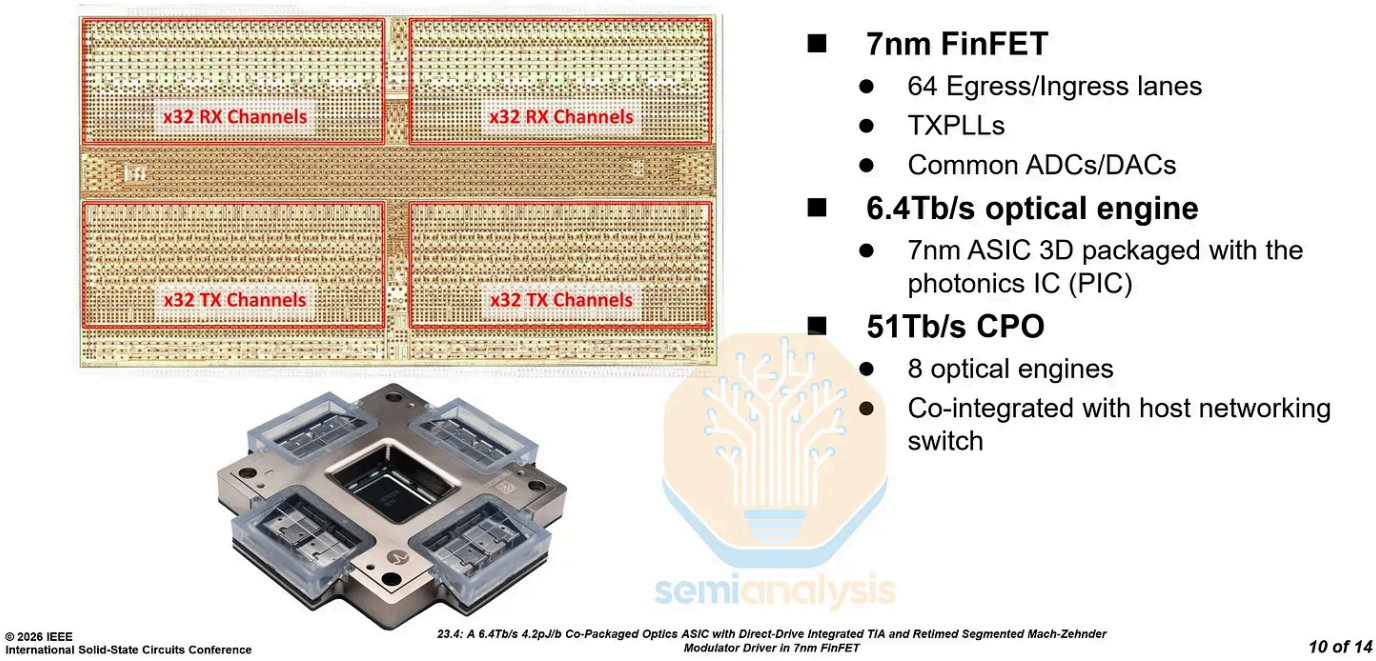
*TX and RX only

© 2025 IEEE
International Solid-State Circuits Conference

Marvell Coherent-Lite Performance vs. Prior Coherent Transceivers. Source: Marvell, ISSCC 2026

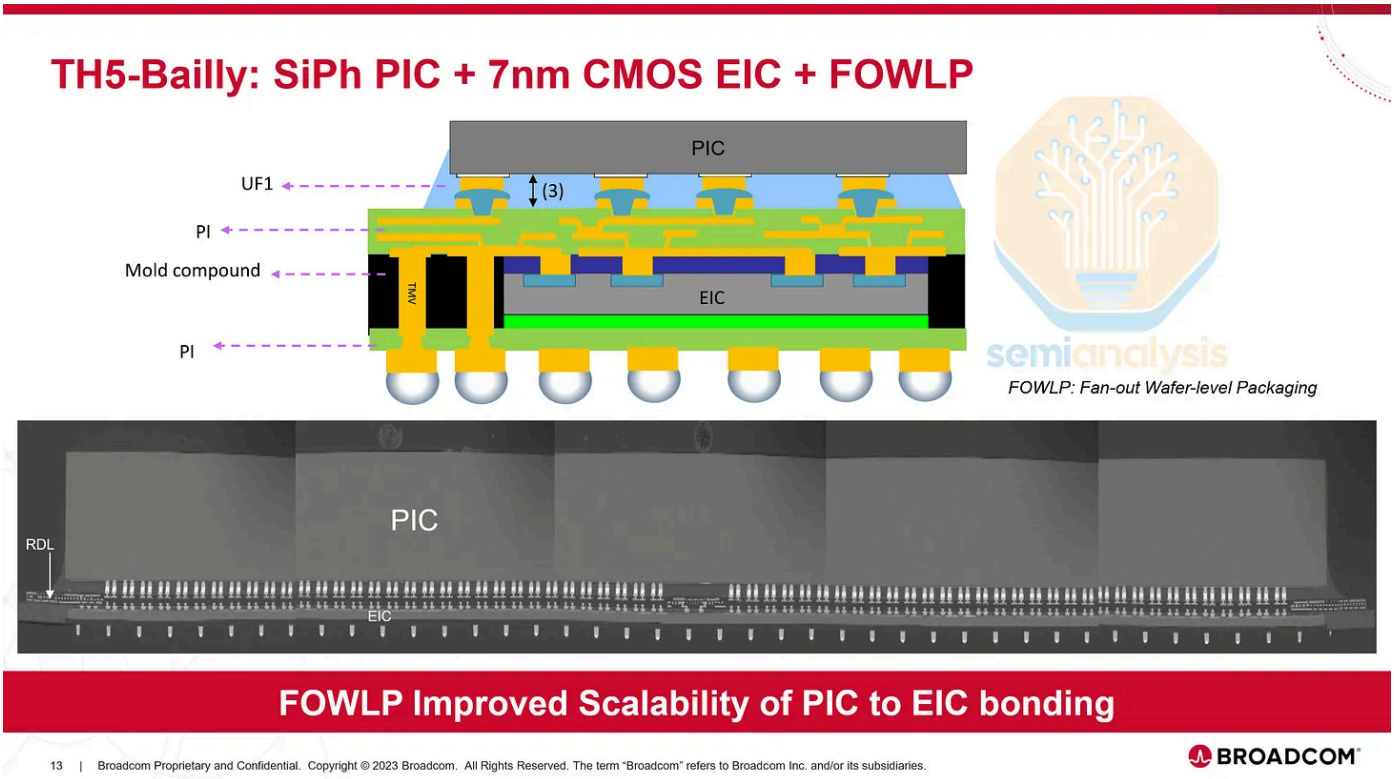
Marvell's approach significantly reduces power to only 3.72 pJ/b excluding silicon photonics, half of other full-fledged coherent transceivers. Their measurements were taken over a fiber length of 40km, with a latency of less than 300 ns.

Broadcom 6.4T Optical Engine - Paper 23.4



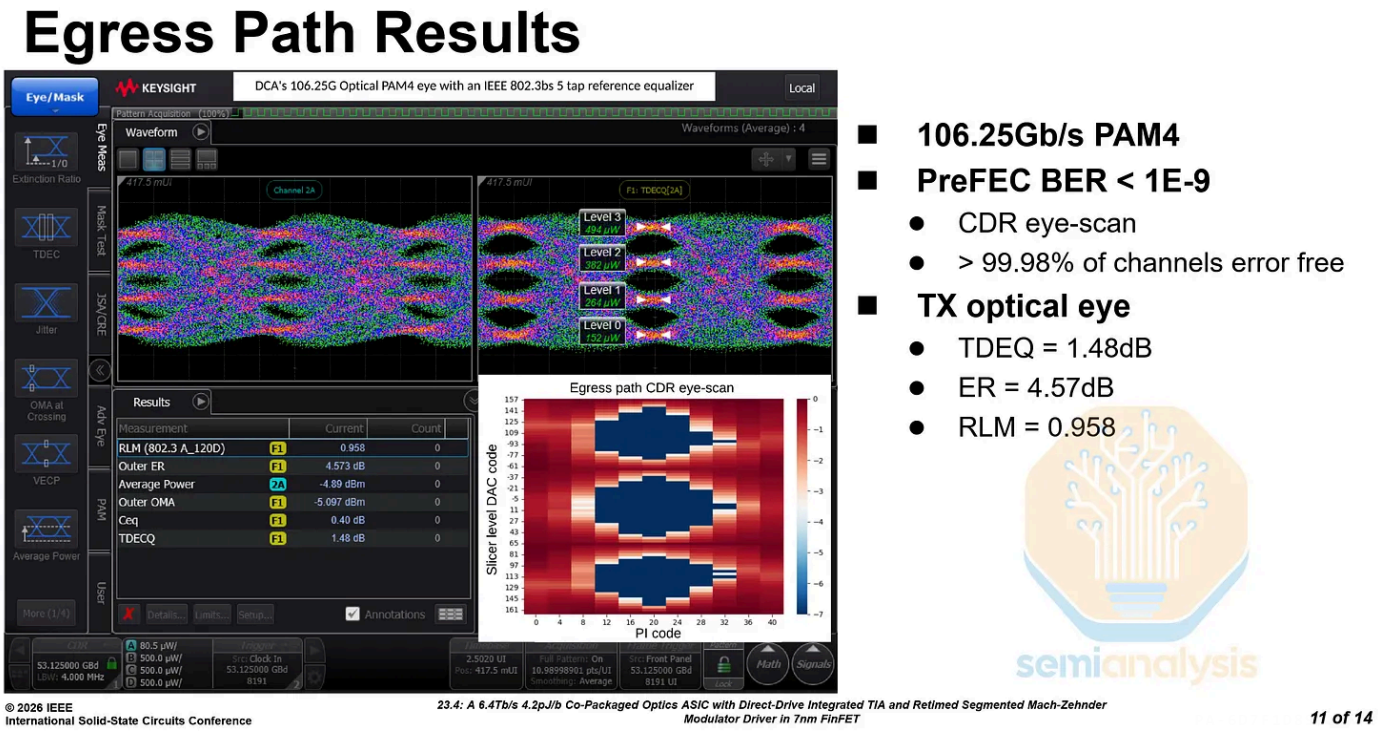
Broadcom Tomahawk 5 51.2T CPO Optical Engine Die Shot and Package. Source: Broadcom, ISSCC 2026

Broadcom showcased progress on their 6.4T MZM optical engine (OE) consisting of 64 lanes of ~100G using PAM4 modulation. The optical engines were tested in a Tomahawk 5 51.2T CPO system. One CPO package consists of eight 6.4T OEs, each with a PIC and an EIC, on TSMC's N7 process.



Broadcom Tomahawk 5 CPO Optical Engine Packaging. Source: Broadcom, Hot Chips 2024

While Nvidia uses COUPE, Broadcom uses a Fan-Out Wafer-Level Packaging approach for this OE. Broadcom will be switching to COUPE in the future, but older generation products like this OE still use other supply chain partners. Here are the promising results from their demonstration:

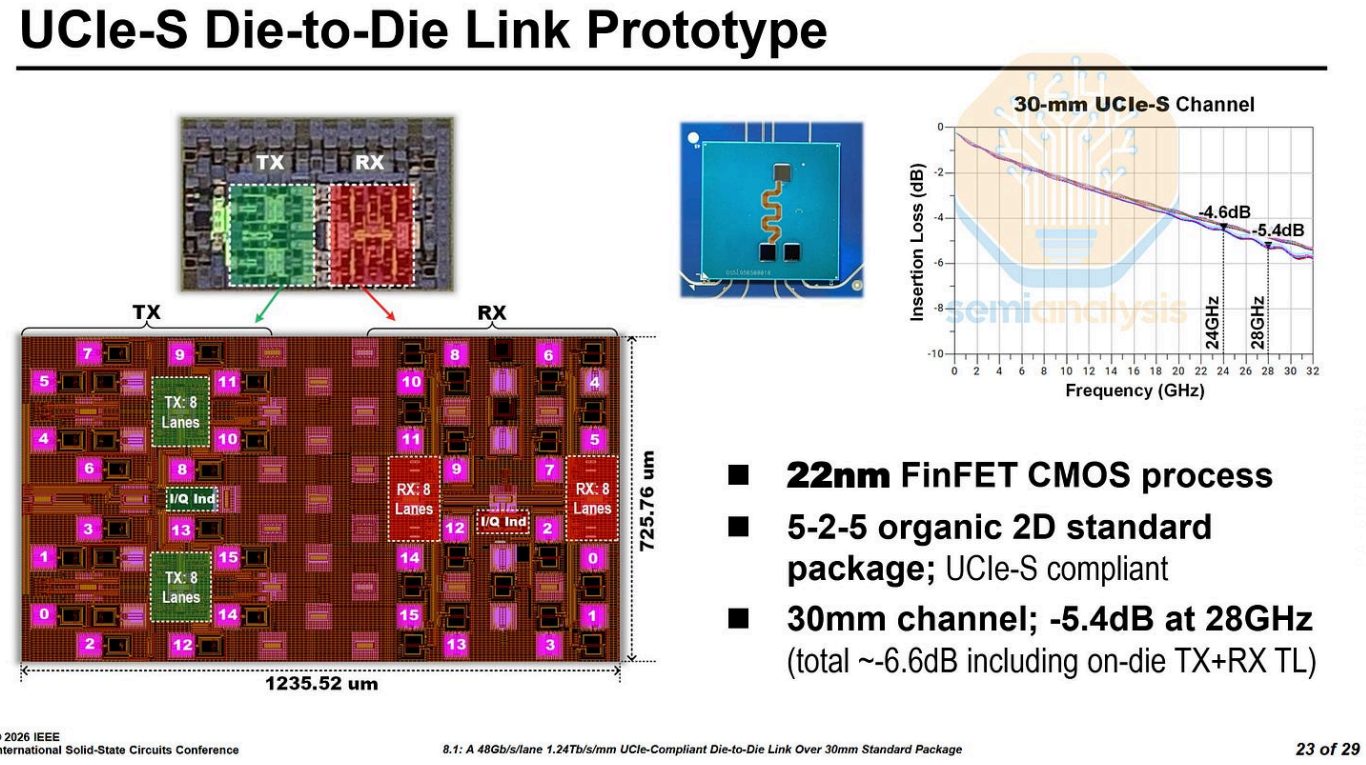


Broadcom 6.4T OE Egress Transmitter Performance. Source: Broadcom, ISSCC 2026

High-Speed Electrical Interconnects

As multi-die designs become the norm, die-to-die interconnects are a critical bottleneck. Major foundries and chip designers presented approaches to pushing bandwidth density and energy efficiency on both organic substrates and with advanced packaging.

Intel UCIe-S - Paper 8.1



Intel UCIe-S Die-to-Die Link Die Shot and Overview. Source: Intel, ISSCC 2026

Intel presented their UCIe-S compatible die-to-die (D2D) interface. It can reach up to 48 Gb/s/lane over 16 lanes with UCIe-S and up to 56 Gb/s/lane with a custom protocol. It works on a standard organic package for a distance of up to 30mm. Interestingly, it was manufactured on Intel's 22nm process.

Comparison: On-package Die-to-Die (D2D) Links

	Poulton, JSSC'19	McCollough, ISSCC'21	Poon, JSSC'22 [9]	Gangasani, ISSCC'22	Seong, ISSCC'24 [10]	Vandersand, VLSI'25 [11]	This Work
Process	16nm	7nm	7nm	5nm	4nm	3nm	22nm
Standard	Custom	Custom	Custom	Custom	Custom	UCIe-S	UCIe-S | Custom*
Package type	Organic	Organic	Organic	Organic	Organic	Organic	Organic
Modulation	NRZ	NRZ	NRZ | PAM4	PAM4	NRZ	NRZ	NRZ
Bump pitch [µm]	-	130	-	-	130	130	112.65
Data rate [Gb/s]	25	40	56 | 112	113	48	16	48Gb/s | 56Gb/s
Channel length [mm]	10 (80 off-pkg)	20	5-30	5-80	10	5-25	30mm
Shore-line BW [Tb/s/mm]	0.200	0.450	0.870	0.460	1.850	0.448	1.243 | 1.450
Energy efficiency [pJ/b]	1.17	1.7	1.24	1.55	0.67†	0.52	1.2pJ/b† | 1.13pJ/b†

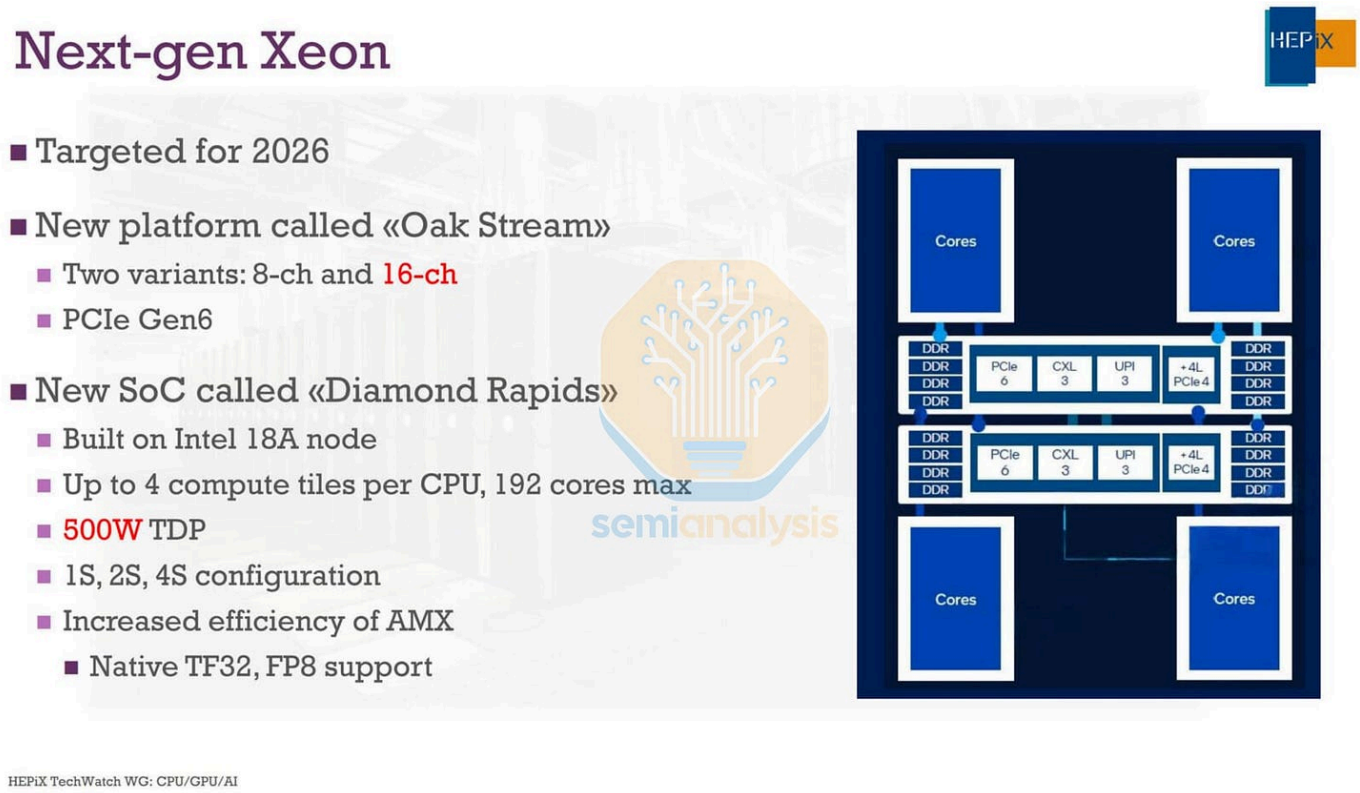
*UCIe-S compliant at 48 Gb/s; 56 Gb/s is an extended data-rate (not part of the current UCIe-S standard). †A PLL is not integrated in this work.

- **High D2D NRZ** rate; **better energy efficiency** in 22-nm vs. the 7-nm work [9]
- **High shoreline BW**; 2-layer UCIe-compliant pkg routing at 3× length vs. [10]
- **3× data rate, 2.8× shoreline density** vs. 16Gb/s UCIe-S in 3nm [11]

© 2026 IEEE
International Solid-State Circuits Conference

Intel UCIe-S vs. Other Die-to-Die Link Comparison. Source: Intel, ISSCC 2026

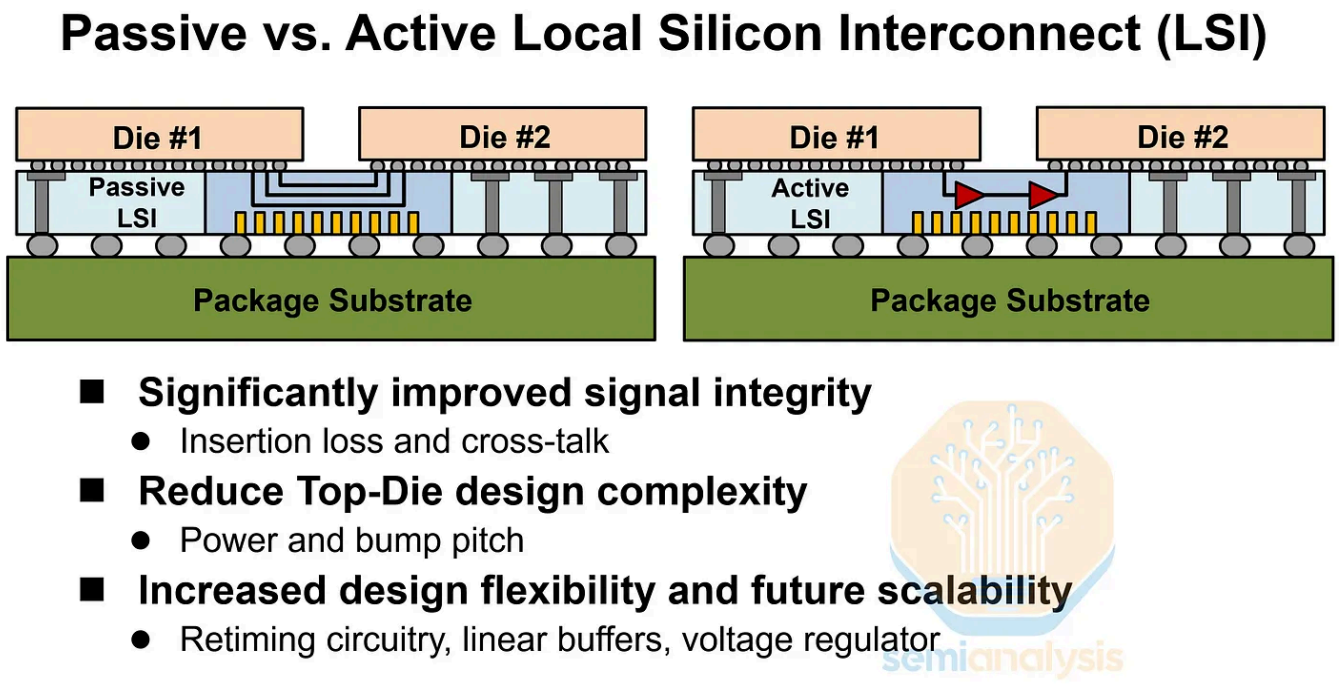
At VLSI 2025, Cadence demonstrated their own UCIe-S die-to-die interconnect on N3E. Despite the node disadvantage, Intel has managed to best Cadence's interconnect in data rate, channel length and shoreline bandwidth, only losing out in energy efficiency.



Intel Diamond Rapids Multi-Die Architecture Overview. Source: HEPiX via @InstLatX64

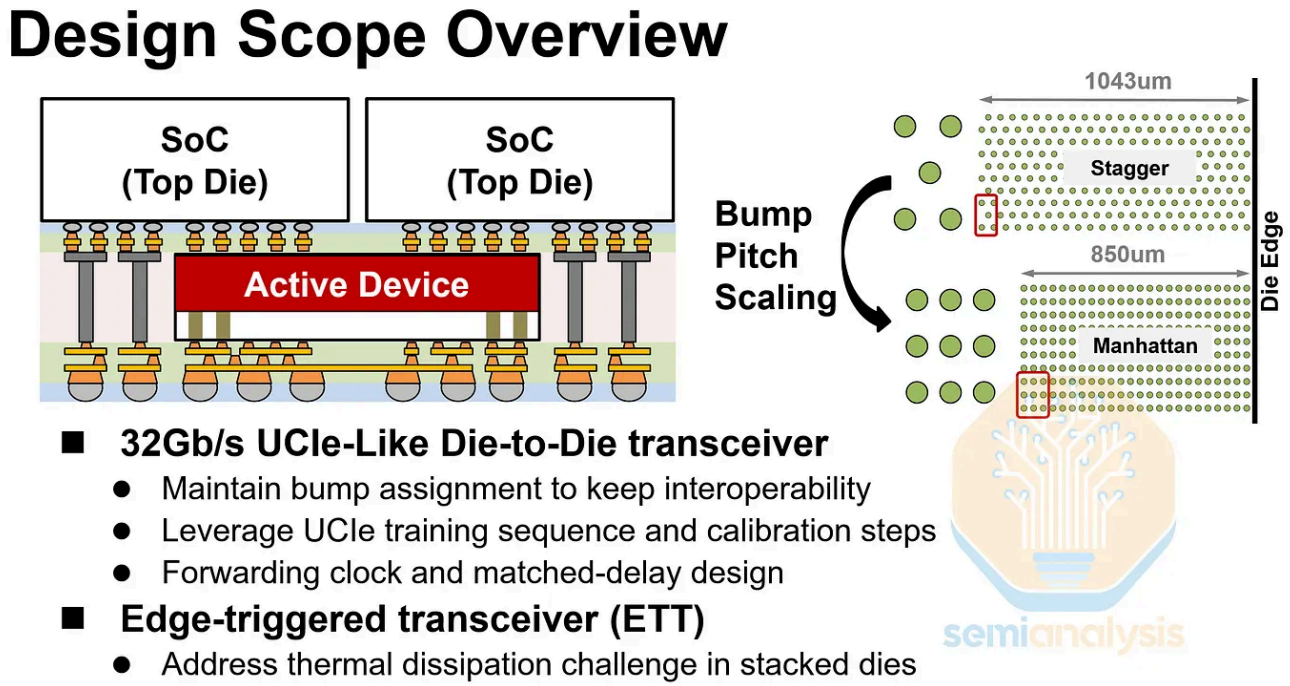
The interconnect presented by Intel is likely to be a prototype of what will be used on their Diamond Rapids Xeon CPUs. Efficiency should be much better when designed on the Intel 3 process compared to this 22nm test chip and could serve to replace the advanced packaging approaches like EMIB on Granite Rapids. As we have covered in our article on the Landscape of Datacenter CPUs, Diamond Rapids consists of two IMH dies, and 4 CBB dies. With the long traces between each CBB die to both IMH dies, we believe this link is a viable candidate to connect the dies over standard package substrate, negating the need for EMIB.

TSMC Active LSI - Paper 8.2



TSMC Passive vs. Active LSI Comparison. Source: TSMC, ISSCC 2026

TSMC's advanced packaging division presented their Active Local Silicon Interconnect (aLSI) solution. As opposed to standard CoWoS-L or EMIB, aLSI improves signal integrity and reduces the complexity of PHYs and SerDes on the top dies.



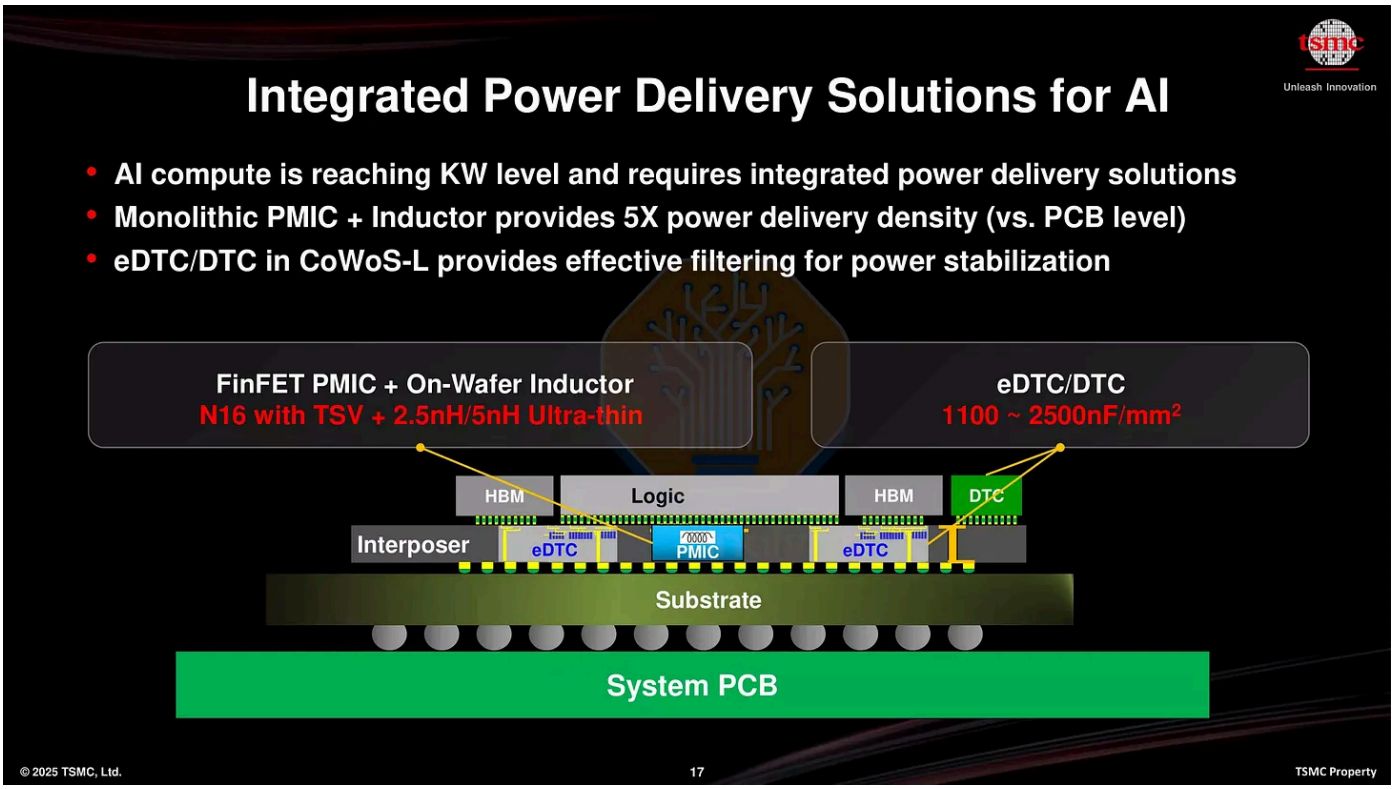
TSMC Active LSI Die-to-Die Link Overview. Source: TSMC, ISSCC 2026

The device that TSMC showed used a 32 Gb/s UCIe-like transceiver. Due to aLSI improving signal integrity, the area of the transceivers could be decreased, and the bump pitch could also be reduced from 45 μm to 38.8 μm. The combination of a tighter pitch and a switch to a Manhattan grid allowed them to reduce the PHY depth from 1043 μm to 850 μm, saving space that designers can reallocate to compute, memory, or IO, or use to shrink the die. The transceiver is only UCIe-like and not true UCIe, as UCIe mandates a hexagonal bump map rather than the Manhattan grid used here.

As designers eke out every bit of die space for next-generation AI accelerators, the switch to aLSI is inevitable.

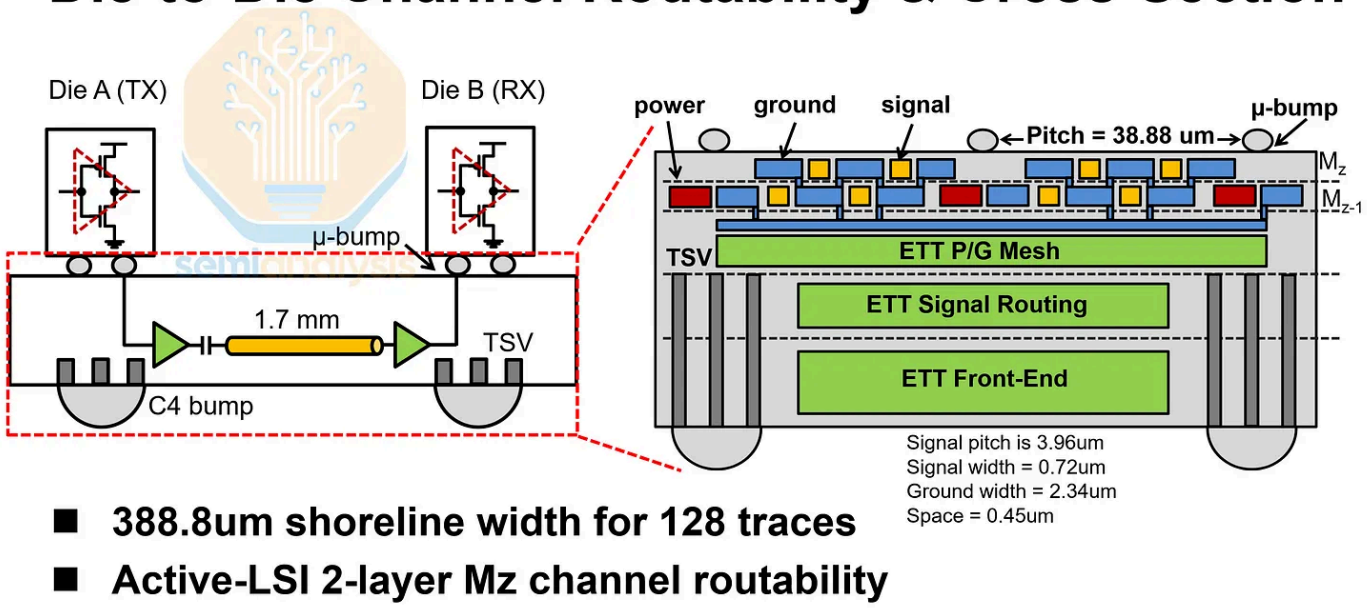
The 'active' part of aLSI comes from replacing the passive long-reach metal channel in the bridge die with active transistors forming an Edge-Triggered Transceiver (ETT) circuit to maintain signal integrity over longer reaches. This also reduces the signal driving requirements of the top die's Transmit/Receive ports. ETT circuits within the aLSI only add an additional 0.07pJ/b to the energy cost, minimizing thermal concerns from adding active circuits in stacked dies. By moving the signal conditioning circuits to the bridge die, PHY area on the top die TX/RX can be reduced by using smaller pre-drivers and clock buffers and remove the need for signal amplification on the receive end.

The ETT integrates a driver, an AC-coupling capacitor (Cac), an amplifier with both negative and positive feedback, and an output stage. Running the signal through the Cac introduces peaks in the signal transition edges, which is then picked up by the dual-loop amplifier, hence the edge-triggered nomenclature. The amplifier leverages both positive and negative feedback loops to stabilize the voltage level. In this design, Cac is set to be 180 fF for a 1.7 mm channel length, with 2kΩ resistance on die A and 3kΩ on die B respectively.



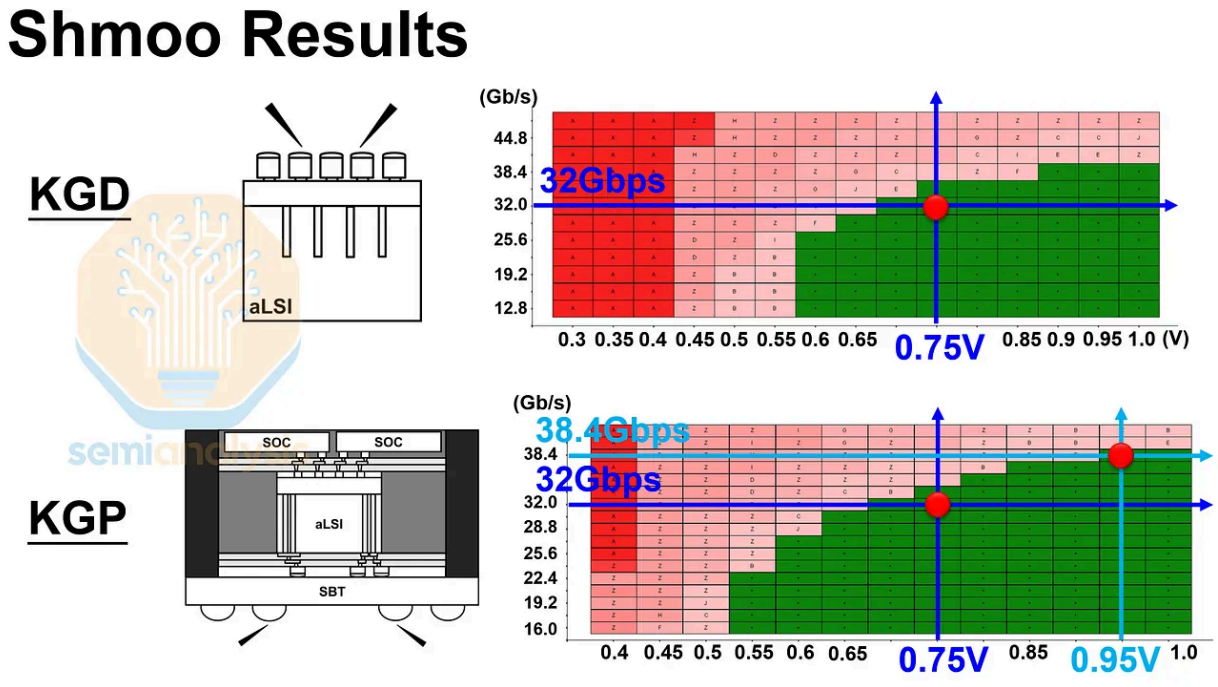
TSMC CoWoS-L Integrated Power Delivery with eDTC. Source: TSMC

These aLSI bridges can also integrate embedded deep trench capacitors (eDTC) along the front-end to improve power delivery to the PHY and D2D controllers. Instead of compromising the power grid by having a bridge die in the way, aLSI with eDTC improves both the power and signal routing along the D2D interfaces.



TSMC Active LSI Routability and Cross-Section. Source: TSMC, ISSCC 2026

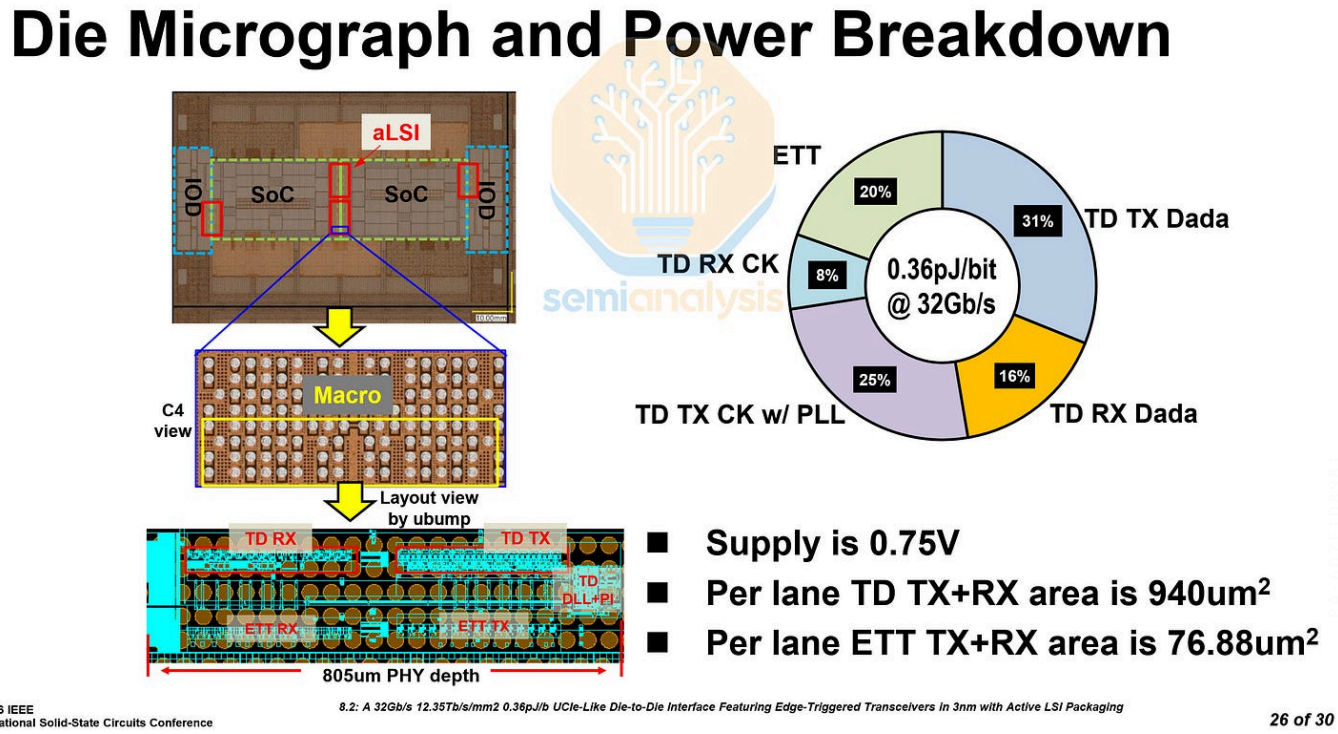
Only 388 μm of shoreline is required for 64 TX and 64 RX data lanes, coming out to a total area of 0.330 mm². Only the top 2 metal layers are required for routing the signals. The remaining metal layers can be used for the front-end circuitry.



TSMC Active LSI Shmoo Plots at KGD and KGP Stages. Source: TSMC, ISSCC 2026

TSMC explained how the Active LSI can be tested at multiple stages. The first is Known Good Die (KGD) with just the LSI for die validation. Next is Known Good Stack (KGS) with the SoCs connected by the LSI for stack functionality. Lastly is Known Good Package (KGP) with the full assembly to comprehensively verify functionality, performance and reliability.

They showed shmoo plots at the KGD and KGP stages, both showing the interconnect hitting 32 Gb/s at 0.75V and 38.4 Gb/s at 0.95V.



TSMC Active LSI Die Shot and Power Breakdown. Source: TSMC, ISSCC 2026

The package reveals two SoC dies and two IO dies. Interestingly, the test vehicle appears to match AMD's MI450 GPU's design, with 2 base dies connected to each other, 12 HBM4 stacks and 2 IO dies with Active LSI. Instead of each individual HBM4 stack having its own Active LSI, two HBM4 stacks share one.

As for the power, the total is only 0.36 pJ/b at 0.75V, with only 0.07 pJ/b being used by the ETT in the Active LSI. Below is a comparison with other D2D solutions.

Performance Comparison

	This work	[1] ISSCC'25	[2] ISSCC'25	[3] VLSI'24	[4] ISSCC'23	[5] VLSI'23
Technology	3nm	3nm	3nm	5nm	4nm	5nm
Module width (um)	388.8	388.8	388.8	300	250	330
Advanced Packages	Active interposer	Passive Interposer	Passive Interposer	Passive Interposer	Passive Interposer	On-chip Channel
Channel Length	1.7mm	1.7mm	1.4mm	2.0mm	3.0mm	1.2mm
Data Rate (Gbps/pin)	32	32	16	32	32	25.2
D2D bump pitch (um)	35	45	45	-	50	55
Energy Efficiency (pJ/bit)	0.36*	0.6*	0.29*	0.36	0.44	0.19**
Shoreline BW (Tbps/mm)	10.5	10.5	5.27	3.84	8.00	5.8**
Area BW density (Tbps/mm^2)	12.35	10.07	5.05	4.47	11.43	6.04**
FoM (Tbps/mm)/(pJ/bit)	29.41	17.50	18.17	10.67	18.18	30.53**

*PLL is included

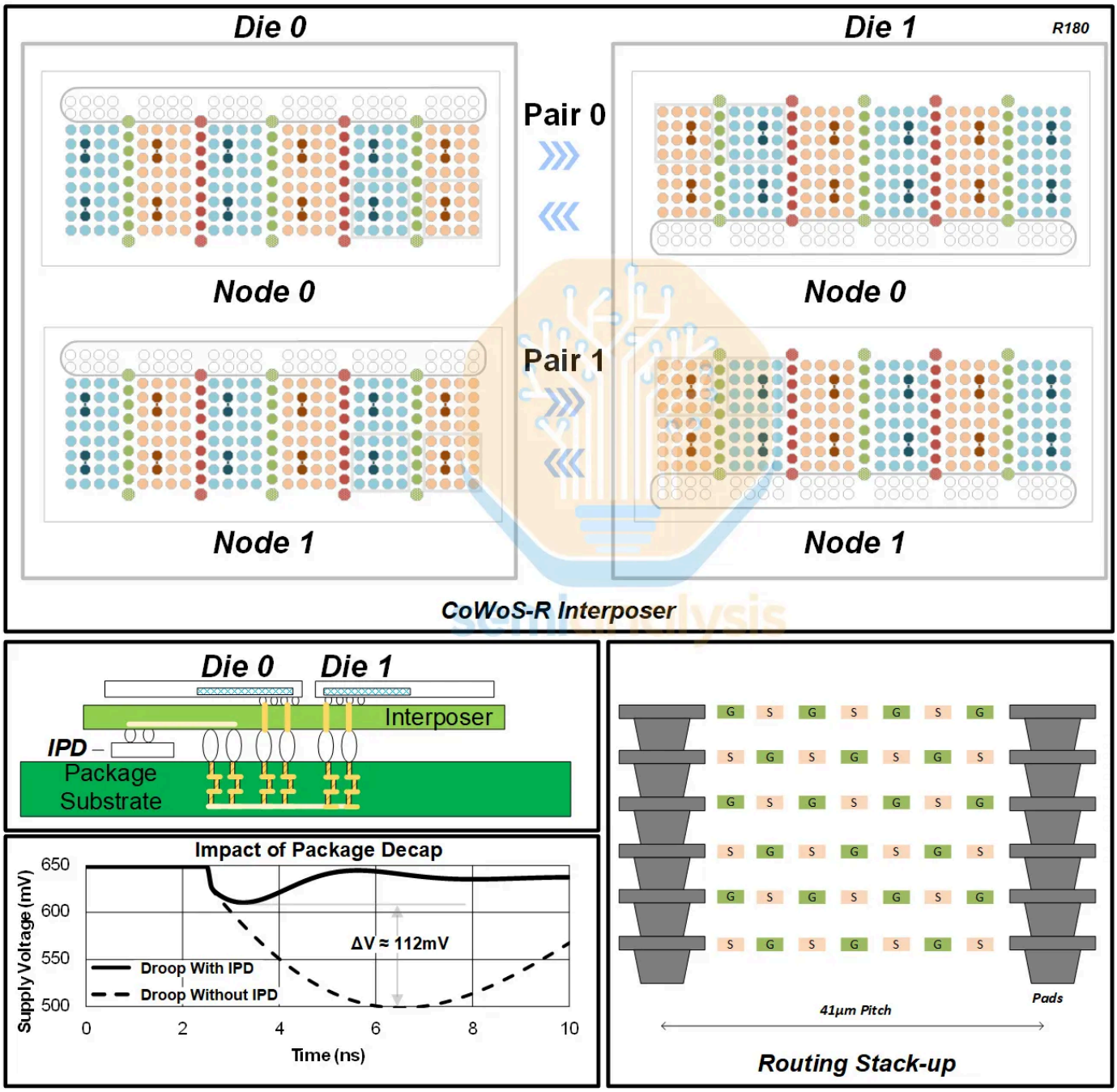
**Extrapolated data

semianalysis

© 2026 IEEE
International Solid-State Circuits Conference

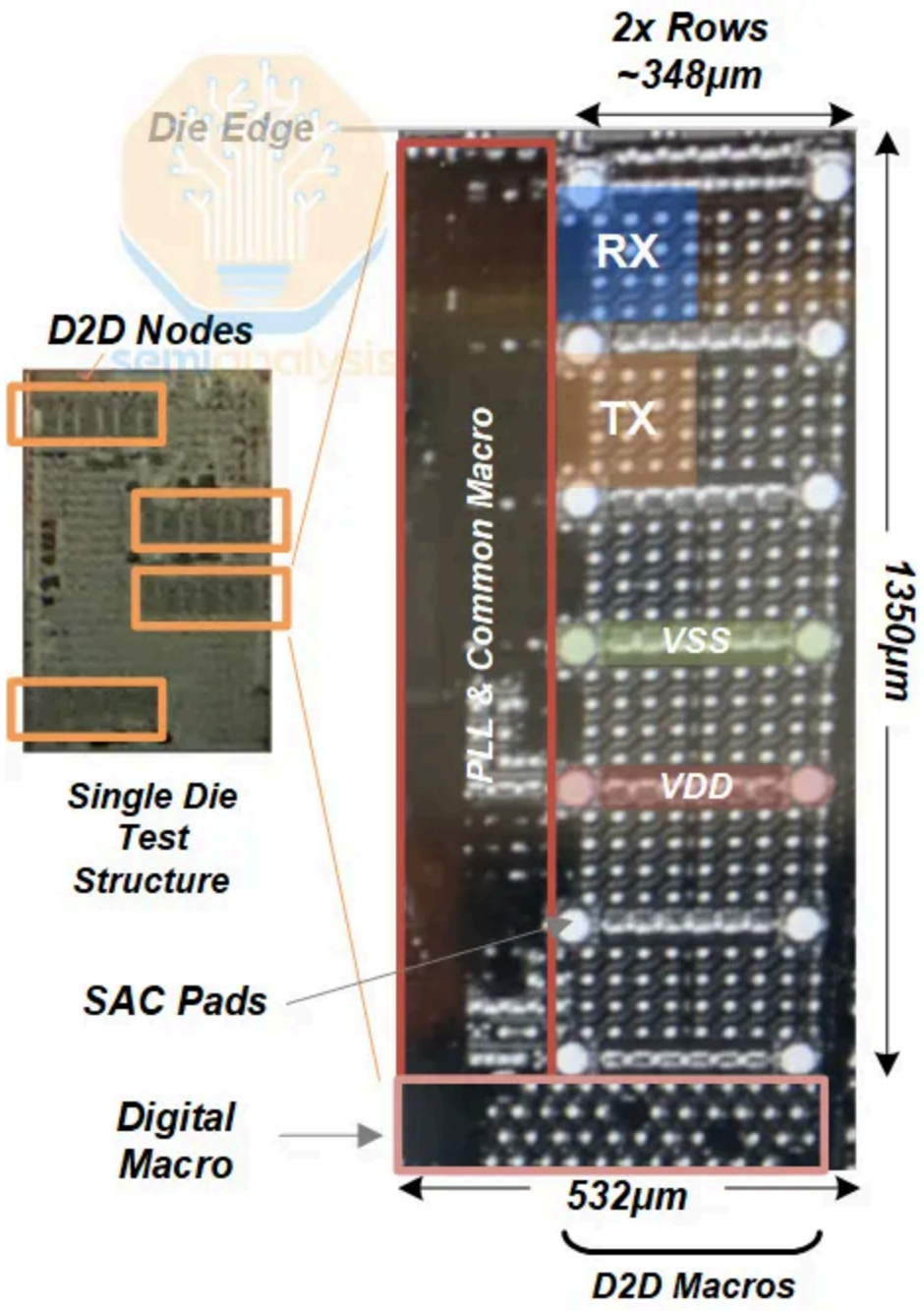
TSMC Active LSI vs. Other Die-to-Die Interconnects. Source: TSMC, ISSCC 2026

Microsoft D2D Interconnect - Paper 8.3



Microsoft D2D Test Vehicle Layout and Routing. Source: Microsoft, ISSCC 2026

Microsoft also detailed their die-to-die (D2D) interconnect. Their test vehicle includes two dies and two pairs of D2D nodes for interconnection. A full mock-up of the power delivery network and routing were included to mimic clock gating and crosstalk.



Microsoft D2D Interconnect Die Shot. Source: Microsoft, ISSCC 2026

The interconnect on their test die occupied 532 µm of shoreline and had a depth of 1350 µm. The test vehicle was fabricated on TSMC's N3P node, and the interconnect was tested at two data rates, 20 Gb/s at 0.65V, and 24 Gb/s at 0.75V.

Microsoft D2D Power Consumption Breakdown. Source: Microsoft, ISSCC 2026

Microsoft reported two power consumption figures, one with both analog and digital system power, and one with only analog power. The latter is what most die-to-die interconnects report. At

24 Gb/s, the system power is 0.33 pJ/b and the analog power is 0.226 pJ/b, while at 20 Gb/s, the system power is 0.25 pJ/b and the analog power is 0.17 pJ/b. The power at idle state is 0.05 pJ/b.

Microsoft D2D vs. Other Die-to-Die Interconnects. Source: Microsoft, ISSCC 2026

Microsoft also compared their interconnect to the same prior research as TSMC did for their Active LSI.

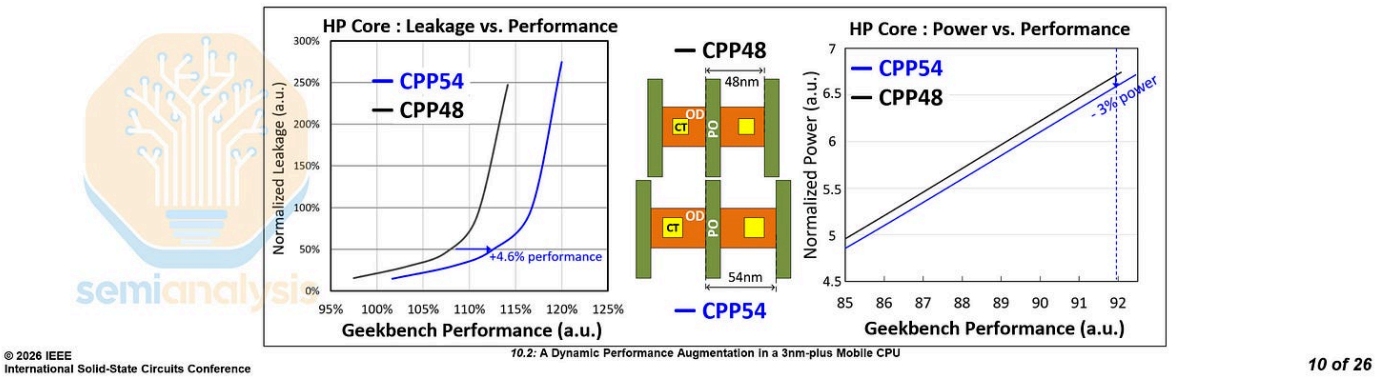
As we explained in a previous article, Microsoft's Cobalt 200 CPU features two compute chiplets connected by a custom high-bandwidth interconnect. We believe that this presentation details that exact interconnect.

Processors

From small mobile CPUs to large AI accelerators, ISSCC featured first-time architectural breakdowns from MediaTek, Intel, AMD, Rebellions, and Microsoft. Many even included die shots.

MediaTek Dimensity 9500 - Paper 10.2

Every year, MediaTek presents a different aspect of their flagship mobile CPUs. This year is no different, and the focus of this year's mobile CPU presentation was on boosting and thermal management.



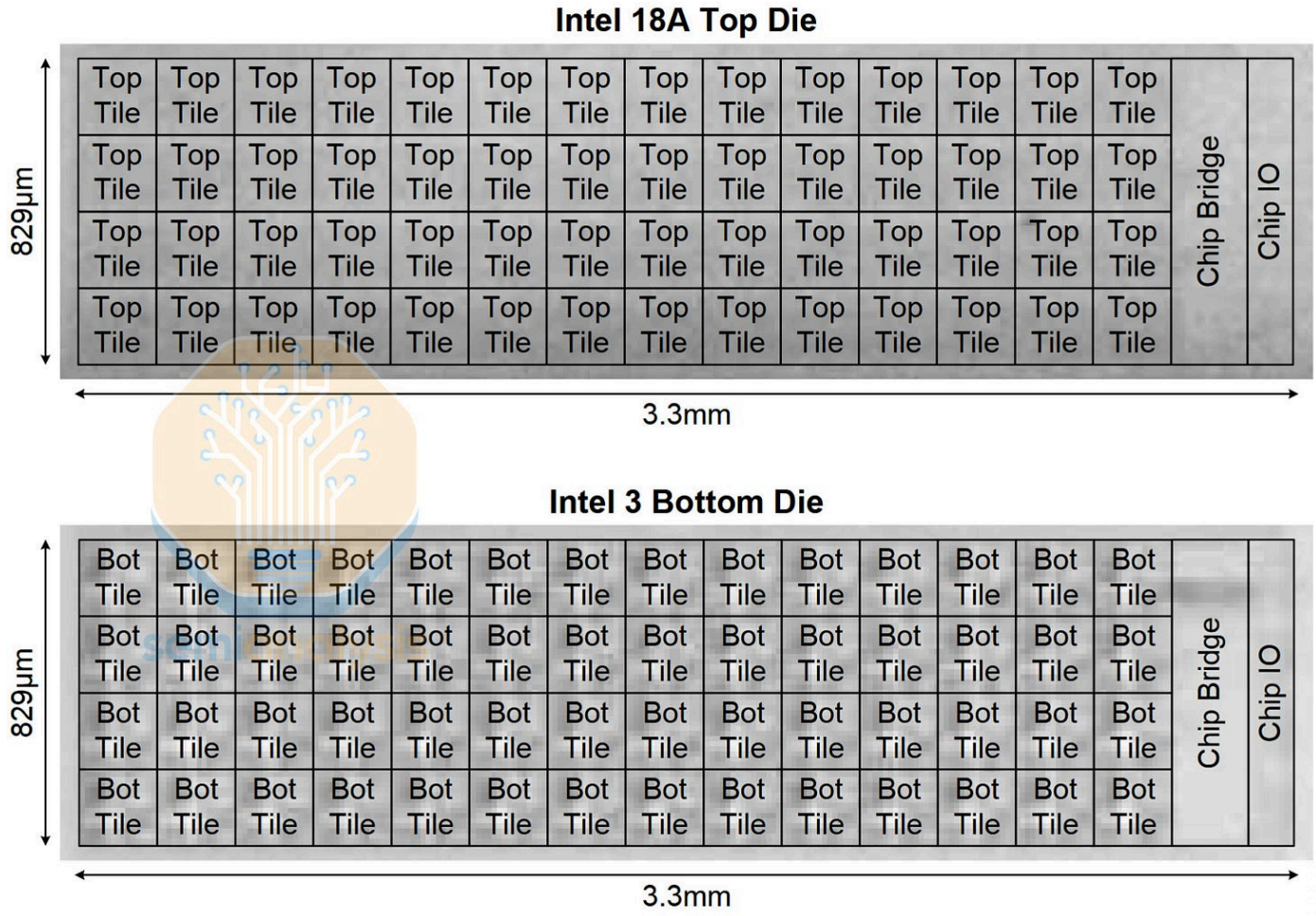
MediaTek Dimensity 9500 C1 Ultra Core Process Optimization. Source: MediaTek, ISSCC 2026

TSMC offers two different Contacted Gate Pitch (CGP) options for N3E and N3P, 48nm and 54nm. Across most chips, the narrower CGP of 48nm has been used, as it leads to smaller cell sizes and a greater die shrink. However, it also faces problems with leakage, routing and manufacturing due to the smaller critical dimensions.

MediaTek used the larger 54nm CGP on their Dimensity 9500's C1 Ultra high-performance cores, for the improved power efficiency. This allowed them to reach higher performance with less thermal penalty, with 4.6% more performance at iso-leakage or 3% less power at iso-performance.

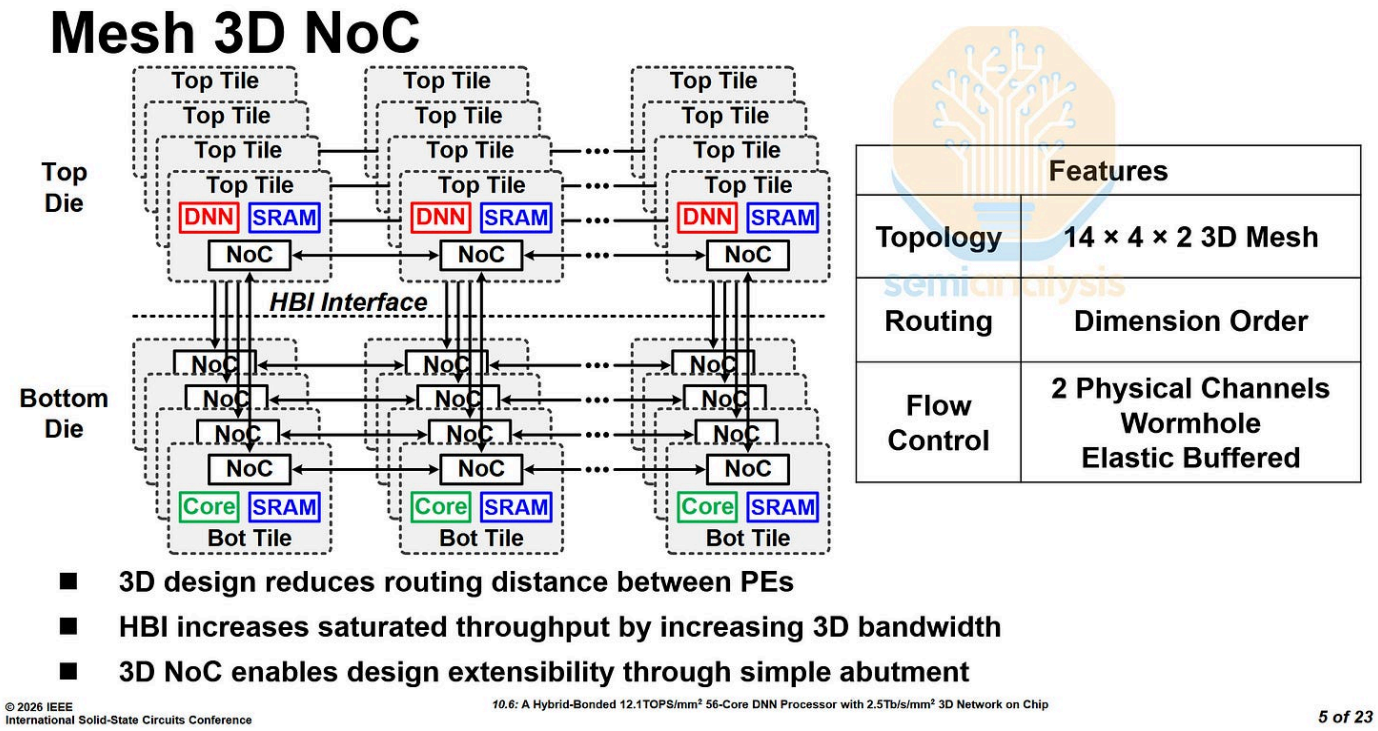
The rest of MediaTek's paper focuses on dynamic performance optimization by utilizing unused aging budget and reducing thermal overshoots. In total, they were able to increase boost clocks from 4.21 GHz to 4.4 GHz. If these optimizations sound interesting to you, we recommend you check out the paper, 10.2 A Dynamic Performance Augmentation in a 3nm-Plus Mobile CPU.

Intel 18A-on-Intel 3 Hybrid Bonding - Paper 10.6



Intel M3DProc 18A and Intel 3 Die Floorplan. Source: Intel, ISSCC 2026

Intel disclosed their first hybrid bonded chip, the M3DProc. It consists of an Intel 3 bottom die, and an 18A top die. Each die contains 56 mesh tiles, cores and DNN accelerator tiles, respectively. The two dies are bonded together with Foveros Direct, hybrid bonding at a 9µm pitch.



Intel M3DProc 3D Mesh Architecture. Source: Intel, ISSCC 2026

The mesh tiles are arranged in a 14×4×2 3D mesh, with SRAM being shared across both dies.

Intel M3DProc 2D vs. 3D Throughput and Energy Efficiency. Source: Intel, ISSCC 2026

Intel found that the 3D mesh reduces latency and increases throughput by almost 40%. They also tested the energy efficiency of transferring data, with 2D being within the 56 mesh tiles of the bottom die, and 3D being 28 adjacent mesh tiles across both dies. The results show that the Hybrid Bonding Interconnect (HBI) had a negligible impact on efficiency.

Intel M3DProc Tile Bonding Floorplan. Source: Intel, ISSCC 2026

Each tile has 552 pads, with just under half used for data, and just under a quarter used for power.

In terms of packaging, M3DProc is similar to Clearwater Forest (CWF). CWF has Intel 3 base dies, connected to 18A compute dies via 9μm Foveros Direct.

The M3DProc achieves 875 GB/s 3D bandwidth, while each CWF compute die only achieves 210 GB/s. This chip's 3D NoC has a significantly higher bandwidth density. CWF uses Foveros Direct to disaggregate the CPU core cluster's L2 cache from the base L3 with 6 clusters per top die at

35GB/s each for 210GB/s per top die. M3DProc's 875GB/s 3D bandwidth is aggregated over 56 vertical tile connections, or 15.6GB/s each over a far smaller area.

AMD MI355X - Paper 2.1

AMD MI300X vs. MI355X XCD Comparison. Source: AMD, ISSCC 2026

AMD presented on their MI355X GPUs. In conference presentations, AMD usually rehashes prior announcements while only introducing one or two new pieces of information. This paper was much better in that regard, explaining how the MI355X XCD and IOD were improved as compared to the MI300X.

AMD MI300X vs. MI355X XCD Area Efficiency. Source: AMD, ISSCC 2026

AMD detailed how they doubled the matrix throughput per CU while keeping the total area the same and the number of CUs largely similar. First, of course, was the move from N5 to N3P; this provided the bulk of the transistor density improvement. The additional two metal layers provided by N3P allowed for improved routing and thus, higher cell utilization. AMD designed their own standard cells, as they have with N5 before, to optimize the node for their HPC use case.

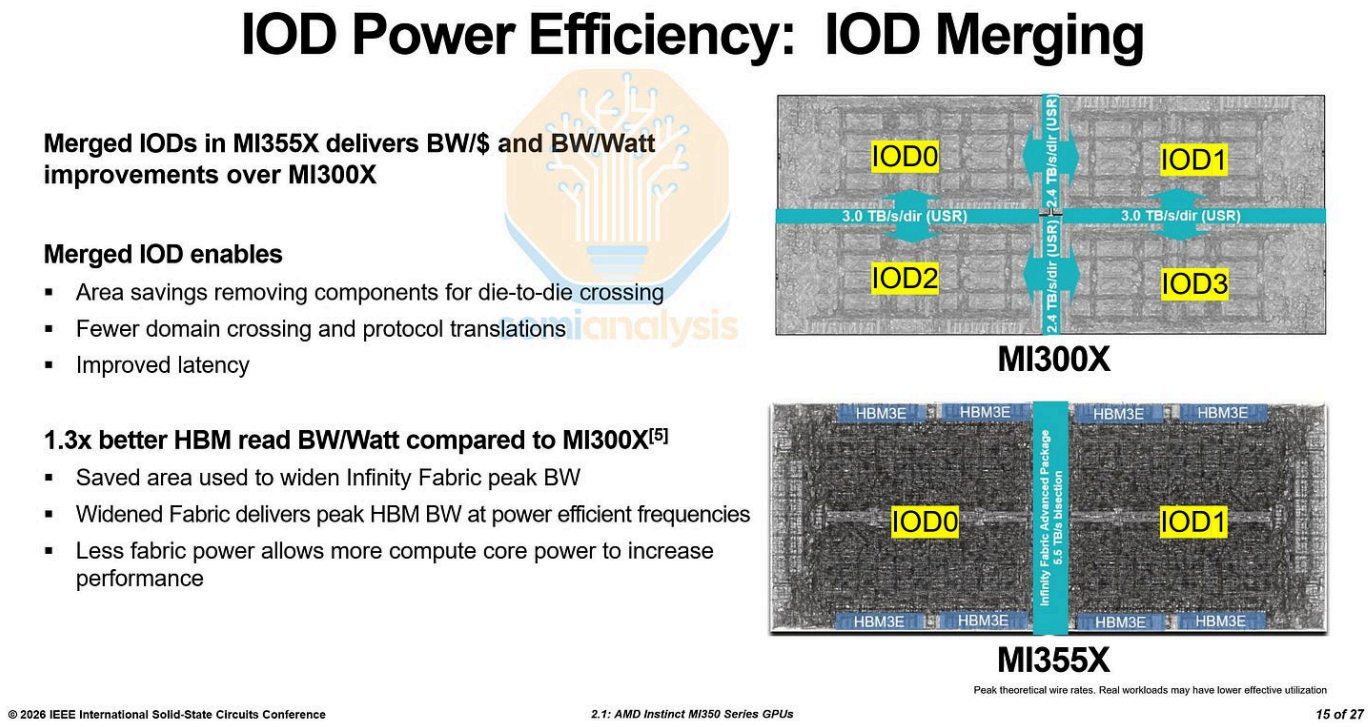
They also used denser placement algorithms, similar to how the Zen 4c cores used in EPYC Bergamo CPUs are much smaller than the Zen 4 cores used in EPYC Genoa CPUs.

There are two approaches when performing the same calculations with many different data formats like FP16, FP8, MXFP4, etc. The first is using shared hardware, where every format goes through the same circuits. However, this comes with a power cost as there is little optimization for each format. The second option is each data format using an entirely different set of circuits for calculations. However, this takes up a lot of additional space. Of course, the optimal approach is somewhere in the middle. This optimization was an important focus for AMD.

AMD MI355X XCD Frequency and Power Efficiency Gains. Source: AMD, ISSCC 2026

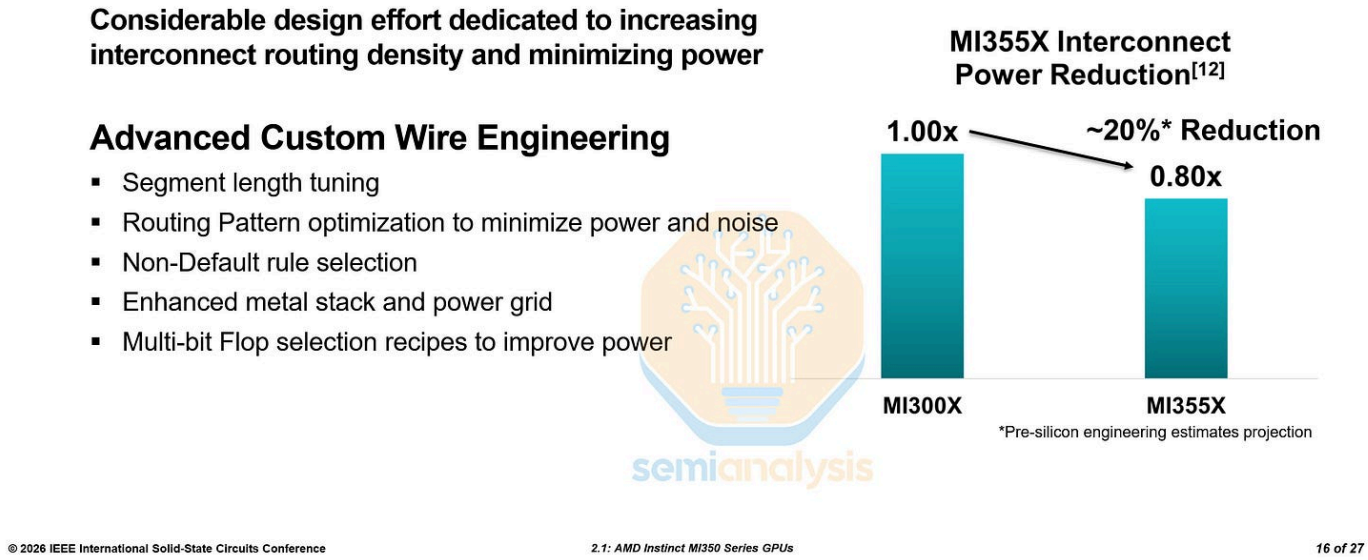
As the next node with improved transistors, N3P provides performance improvements by itself. Nevertheless, AMD was able to improve frequency iso-power by 5% before process node improvements. They also designed multiple flip-flop variants with varying power and

performance characteristics deployed across different areas of the chip depending on usage and architectural requirements.



AMD MI355X IOD Merging Power Efficiency. Source: AMD, ISSCC 2026

MI300X featured 4 IO Dies. MI355X cuts that down to two. By doing this, AMD saves area on die-to-die interconnects. A larger monolithic die improves latency and reduces SerDes and translations. Moreover, the efficiency of the HBM was also improved by increasing the interconnect width. The saved power could be reallocated to the compute dies to increase performance.



AMD MI355X IOD Interconnect Power Optimization. Source: AMD, ISSCC 2026

As a large die with many routing options between any two areas on the chip, AMD had to do a lot of work to optimize the wires and interconnects. Through custom engineering of the wires, AMD was able to reduce the interconnect power consumption by ~20%.

Rebellions Rebel100 - Paper 2.2

Rebellions is a South Korean startup working on AI accelerators. At ISSCC, they published the first architectural breakdown of their new accelerator, the Rebel100. Unlike other accelerators, which are generally manufactured at TSMC, Rebellions chose Samsung Foundry's SF4X node. With Nvidia, AMD, Broadcom and others hogging most of TSMC's capacity, this allows them more flexibility.

Rebellions Rebel-Quad (now Rebel100) Summary from Hot Chips 2025. Source: Rebellions via ServeTheHome

At Hot Chips 2025, Rebellions demonstrated the chip running Llama 3.3 70B. The specs have remained the same between Hot Chips and ISSCC. One key point to note is the use of Samsung's I-CubeS interposer technology. While the Hot Chips slide mentions the usage of TSMC's CoWoS-S, we have clarified that this was a mistake on the slide, and that it has always been I-CubeS.

We have recently mentioned that CoWoS-S capacity constraints have been easing. That said, Samsung may have offered steep discounts to bundle I-CubeS advanced packaging with their front-end process — sparing the startup from having to find and validate a separate advanced packaging supplier. Samsung may have also conditioned the availability of their HBM on using I-CubeS.

I-CubeS has not seen adoption among any of the leading AI accelerators, and this could be Samsung's attempt to break into the market. There are only 5 confirmed users of I-CubeS: eSilicon, Baidu, Nvidia, Rebellions and Preferred Networks.

The first is a networking ASIC by eSilicon on Samsung's 14LPP with HBM2. Baidu's Kunlun1 accelerator is similar, using Samsung's 14LPP process and 2 HBM2 stacks. When the CoWoS-S capacity was very tight back in 2023, Nvidia outsourced a small amount of H200 production to I-CubeS. Then, there is the Rebel100 and lastly, a planned accelerator from Preferred Networks on the SF2 process.

Rebellions Rebel100 Multi-Die Architecture. Source: Rebellions, ISSCC 2026

The Rebel100 uses 4 compute dies and 4 HBM3E stacks. Each die has 3 UCIe-A interfaces. However, only two are used on each die, clocked at 16 Gb/s.

Rebellions Rebel100 Package-Level Modularity. Source: Rebellions, ISSCC 2026

Rebellions claims that the design is reconfigurable at the package level, where additional IO or memory chiplets can be added to integrate with Ethernet for scale-up. This is where the remaining UCIe-A interface would be used.

Rebellions stated that the IO chiplets would be taped out by 1Q2026. There was no provided timeline for the memory chiplets.

Rebellions Rebel100 Summary and Roadmap. Source: Rebellions, ISSCC 2026

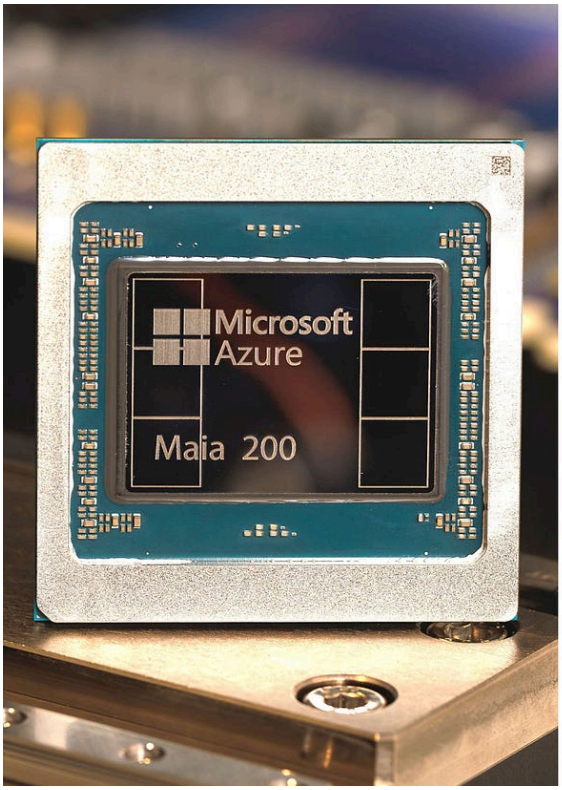
Rebellions Rebel100 Integrated Silicon Capacitors for HBM3E Power Quality. Source: Rebellions, ISSCC 2026

They have also integrated silicon capacitors beside each HBM3E stack to improve power quality for HBM3E and critical control blocks.

Microsoft Maia 200 - Paper 17.4

Microsoft detailed their Maia 200 AI accelerator. The paper was less of a research paper and more of a white paper, with only a single image, a spec sheet comparing it to the Maia 100. Makes sense given a lot of Maia 200 claims are dubious such as the flops/mm^2 and flops/w.

While the Maia 100 was designed in a pre-GPT era, Maia 200 was designed for the current age of models and specifically inference. Earlier this year, Maia 200 nodes were made generally available on Azure.



Maia 200 Introduction

Build an optimized perf/$ solution for AI Inference WL

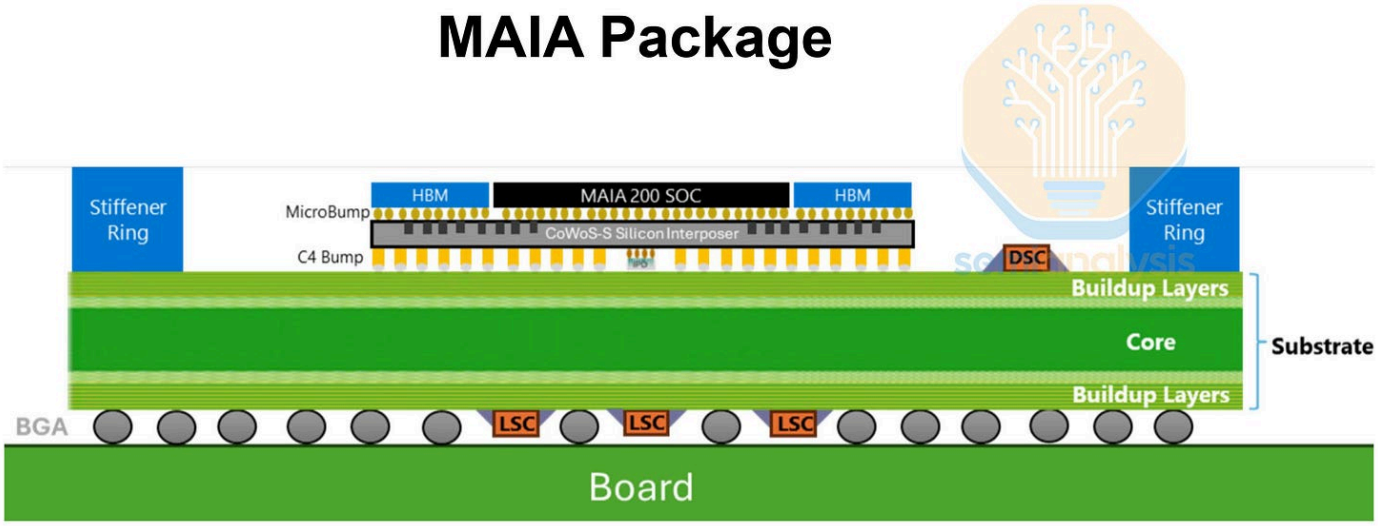
Peak Dense Tensor TOPS	FP4: 10,145 FP8: 5,072 BF16: 1,268
Peak HBM BW (TB/s)	7 TB/s
Number of HBM Stacks	6
HBM Capacity (GB)	216
Host PCIe BW (GB/s)	64
PCIe Configuration	Gen6 x8
Backend Network Bandwidth (GB/s)	1400
SOC Die Area (mm2)	~820
Package Size	75x75
Process Technology	TSMC 3nm
Package Technology	CoWoS-S
SoC TDP (provision)	750W

semianalysis

2

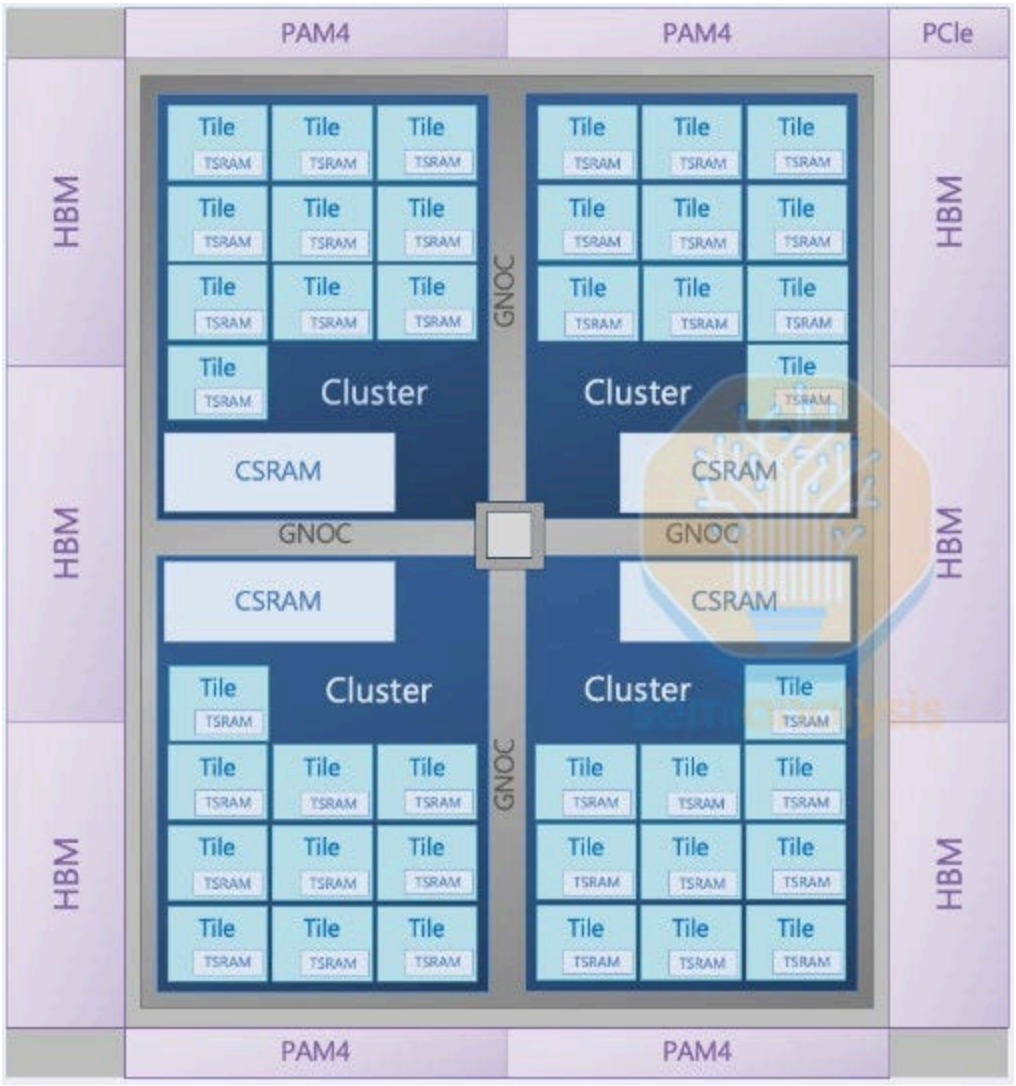
Microsoft Maia 200 Specifications Summary. Source: Microsoft, ISSCC 2026

Maia 200 is the last holdout of reticle-scale monolithic designs. Every major HBM-equipped training and inference accelerator has moved on to multi-chip designs with 2, 4, or even 8 compute dies per package. Every single mm² of the die has been hyper-optimized for one purpose. Unlike with an Nvidia or AMD GPU, there is no legacy hardware for media or vector operations. Microsoft has pushed the reticle-scale monolithic approach to its limit on TSMC's N3P process, packing in over 10 PFLOPs of FP4 compute, 6 HBM3E stacks, and 28× 400 Gb/s full-duplex D2D links.



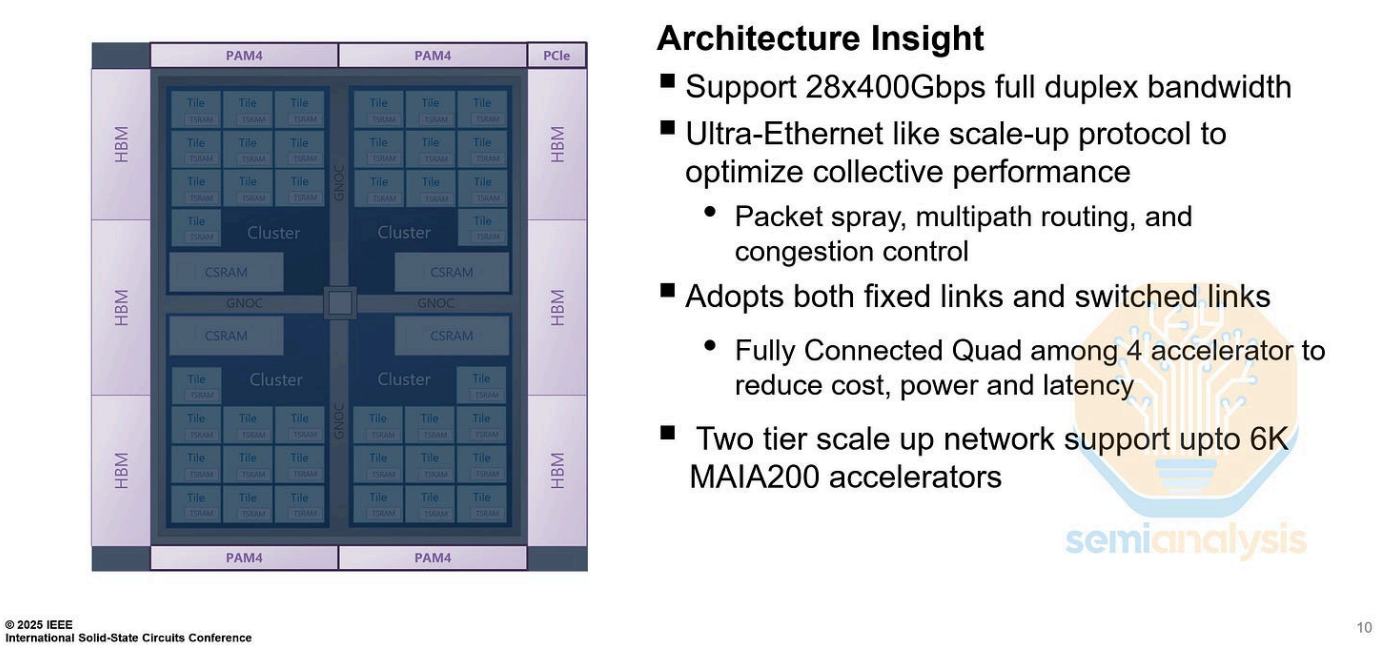
Microsoft Maia 200 Package Cross-Section. Source: Microsoft, ISSCC 2026

On the package-level, Maia 200 is very standard, mimicking the H100. A CoWoS-S interposer, with 1 main die, and 6 HBM3E stacks.



Microsoft Maia 200 Die Floorplan. Source: Microsoft, ISSCC 2026

The long sides of the chip are covered in 3 HBM3E PHYs each, while the short sides each have 14 of the 28 lanes of 400 Gb/s D2D links. In the center, there is 272 MB of SRAM, with 80 MB of TSRAM (L1) and 192 MB of CSRAM (L2).



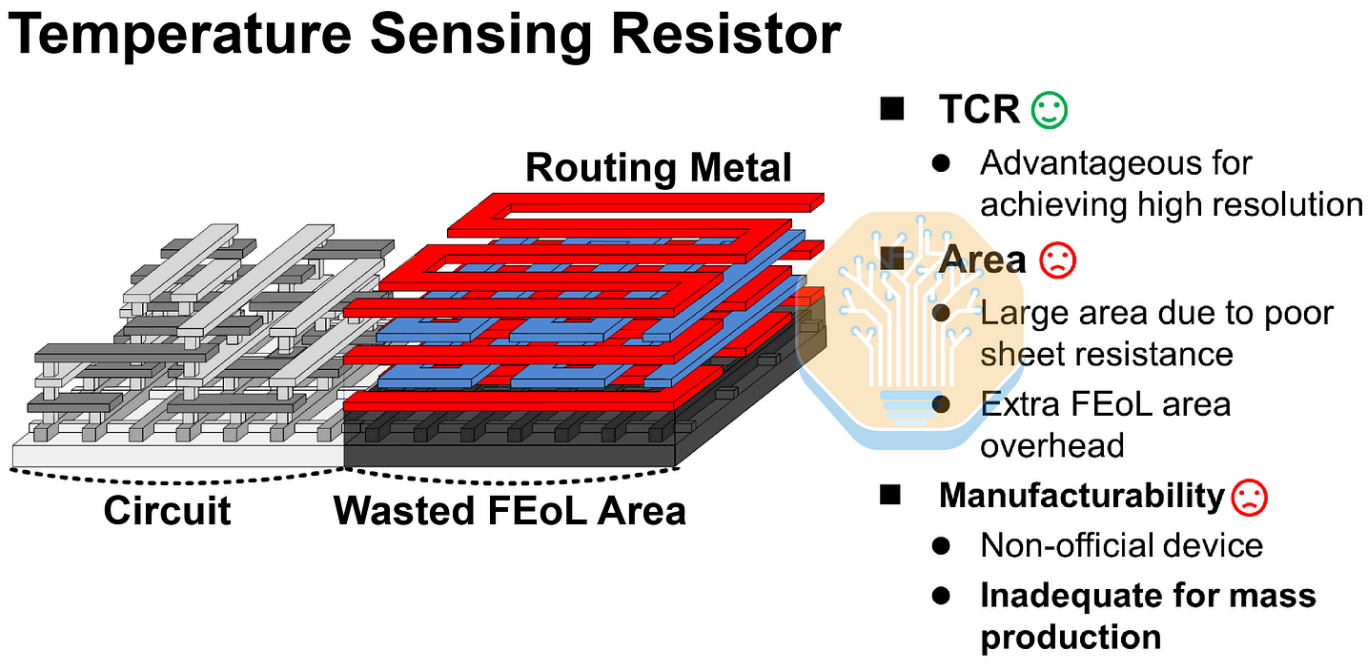
Microsoft Maia 200 Scale-Up Network and IO. Source: Microsoft, ISSCC 2026

The Maia 200 has two different kinds of links, fixed links between other chips in the same node, and switched links between a chip and a switch. 21 links are configured as fixed links, 7 to each other chip, while the remaining 7 links are configured as switched links to one of four in-rack switches.

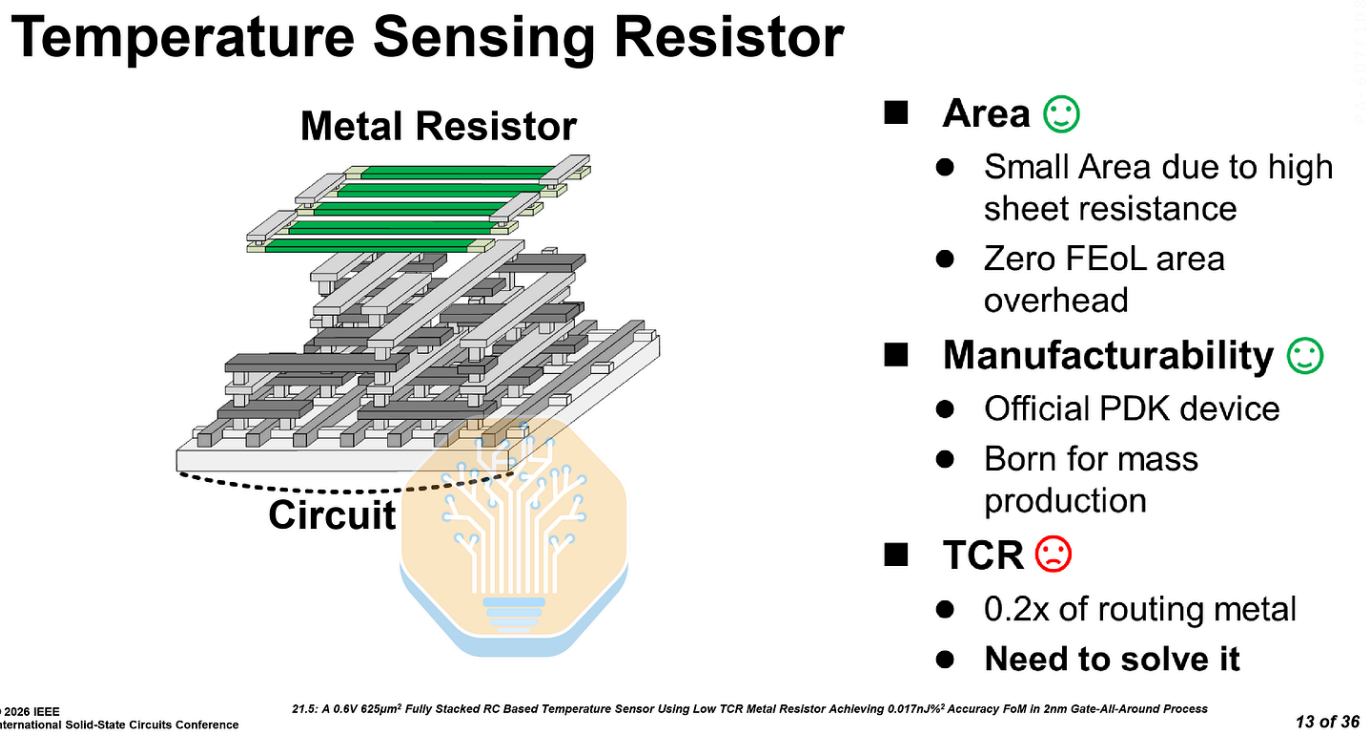
We will be publishing a deep dive into the Maia 200, its microarchitecture and network topology, for institutional subscribers.

Other Highlight

Samsung SF2 Temperature Sensor - Paper 21.5



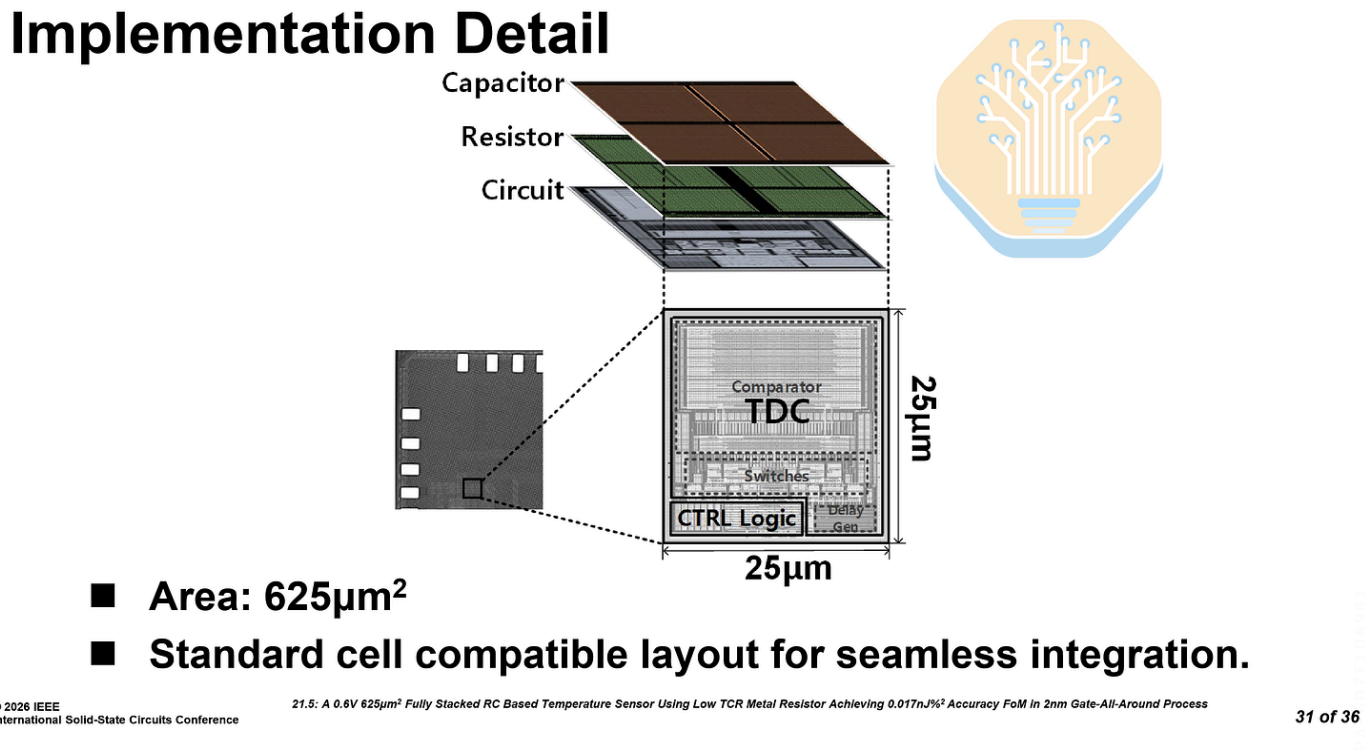
Traditional Temperature Sensor Tradeoffs. Source: Samsung, ISSCC 2026



Samsung SF2 Metal Resistor-Based Temperature Sensor Tradeoffs. Source: Samsung, ISSCC 2026

Samsung presented a compact temperature sensor on SF2, replacing the traditional bipolar junction transistor (BJT) approach with a BEOL metal resistor. This may not be as flashy as the next generation of memory or processors, but it is essential to making chips work.

The metal resistor offers 518× higher sheet resistance than an equivalent routing metal, requiring roughly 1% of the area for the same resistance. As it sits in the upper metal layers, it leaves plenty of room for any circuitry underneath and eliminates FEOL area overhead. Although it has a low resolution, the benefits more than make up for it.



Samsung SF2 Temperature Sensor Stacked Implementation. Source: Samsung, ISSCC 2026

The sensor uses a fully stacked capacitor-resistor-circuit structure with a total area of just 625 µm². As a characterized PDK element, its behavior is modeled by the foundry and validated. It is more suitable for mass production, where process variation must be tightly controlled. Even on a single chip, thousands of these sensors may be used near hotspots.

As mentioned earlier, the metal resistors have a lower temperature coefficient of resistance (TCR), just 0.2× that of routing metal — which limits sensing resolution. Samsung compensates for this by increasing the base resistance. However, this slows sensing time as the RC time constant grows. To address this, Samsung uses a time-offset compression technique: a low-resistance (0.1R) fast-charge path rapidly charges the RC filter, then the circuit switches to the full resistance for the temperature-sensitive portion of the waveform.

For the time-to-digital conversion (TDC), they replaced the large linear delay generator used in prior work with a compact ring oscillator-based (RO) TDC, cutting delay generator area by 99.1%. The RO also doubles as the system clock, with phase-interleaved counting preventing non-monotonicity.

Benchmarking

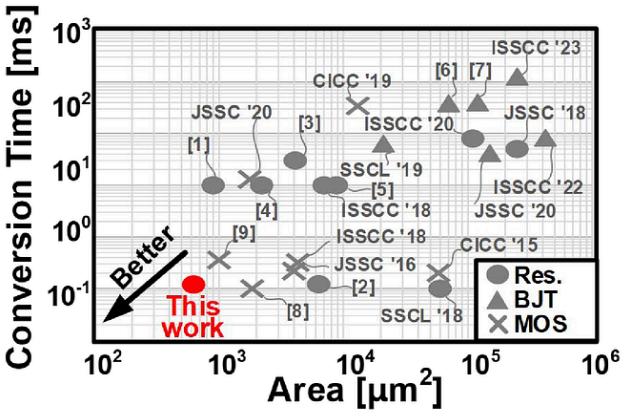
	This Work		ISSCC '24 [1]	ISSCC '22 [2]		ISSCC '25 [3]		ISSCC '21 [4]	ISSCC '20 [5]
Technology	2nm Gate-All-Around		3nm finfet	5nm finfet		28nm		65nm	65nm
Sensor Type	**Metal Res.**		Routing metal	Routing metal		Routing metal		Polysilicon Res.	Polysilicon Res.
Official PDK Device	**O**		X	X		X		O	O
Required Metal Layer for Resistor	**Upper only**		Lower to Upper	Lower		Lower to Upper		-	-
Wasted FEoL Area [μm²]	**0**		670[a]	510[a]		730[a]		1334[a]	3690[a]
Area [μm²]	**625**[b]		900[b]	1316[c]	6350	4100[b]		2210[b]	8800[b]
Read-Out Type	Ring-OSC TDC		Counter	SAR DLL		FLL		ΔΣADC	FLL
Conversion Time [ms]	**0.012**		1	0.012		3		1	1
Energy/Conversion [nJ]	0.18		20	0.18		31.5		28	45
Temperature Range [°C]	-40 to 125		-25 to 125	-40 to 150		-40 to 125		-55 to 125	-30 to 90
Calibration Points	1	2	1	1	2	1	2	1	1
Inaccuracy(3σ) [°C]	3.9	0.5	1.5	-	1.8	1.5	0.2	1.3	-
Inaccuracy(p2p) [°C]	1.9	0.8	1	2.7	-	-	-	-	0.72
Accuracy FoM[d] [nJ·%²]	**0.017**		20	0.16		0.46		14.6	16.2
Resolution FoM[e] [pJ·%²]	0.7		9.8	2.3		0.05		4.6	0.09

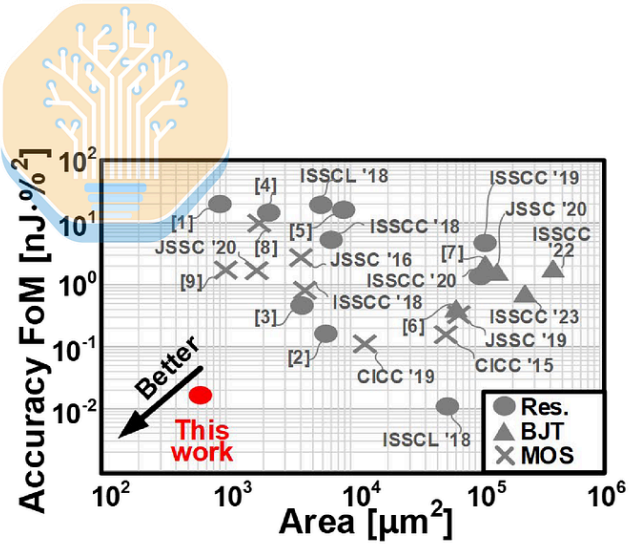
a) Estimated RC filter area
b) Excluding off-chip post-calculation block
c) Excluding off-chip post-calculation and read-out circuit
d) Accuracy FoM = energy/conversion x (relative inaccuracy)²
e) Resolution FoM = energy/conversion x (resolution)²

© 2026 IEEE
International Solid-State Circuits Conference

Samsung SF2 Temperature Sensor Conversion Time and Accuracy Comparisons. Source: Samsung, ISSCC 2026

Achievement





■ **Conversion time and accuracy FoM of the temperature sensor versus area.**

Accuracy FoM = energy/conversion × (relative inaccuracy)²

© 2026 IEEE
International Solid-State Circuits Conference

Samsung SF2 Temperature Sensor vs. Prior Work Comparison Table. Source: Samsung, ISSCC 2026

The new temperature sensor has an accuracy figure of merit (FoM) of 0.017 nJ·%², improving upon prior work on Samsung 5LPE, TSMC N3E and Intel 4 (JSSC 2025). Prior temperature

sensors could only optimize for one of these: area or speed. The sensor on N3E was small, at 900 μm^2 but took 1 ms, while the sensor on Samsung 5LPE was fast, at 12 µs but huge, at 6356 μm^2.

Recommend SemiAnalysis to your readers