



# Memory: **Almost** The Only Thing That Matters

A revolution in memory architecture for the data center

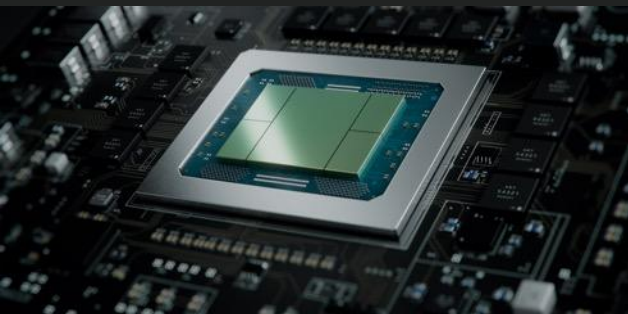
**Mark Kuemerle**  
VP of Technology  
August 25, 2025

# Forward-looking statements

Except for statements of historical fact, this presentation contains forward-looking statements (within the meaning of the federal securities laws) including statements related to future revenue, future earnings, and the success of our product releases that involve risks and uncertainties. Words such as “anticipates,” “expects,” “intends,” “plans,” “projects,” “believes,” “seeks,” “estimates,” “can,” “may,” “will,” “would” and similar expressions identify such forward-looking statements. These statements are not guarantees of results and should not be considered as an indication of future activity or future performance. Actual events or results may differ materially from those described in this presentation due to a number of risks and uncertainties.

Forward-looking statements are only predictions and are subject to risks, uncertainties and assumptions that are difficult to predict, including those described in the “Risk Factors” section of our Annual Reports on Form 10-K, Quarterly Reports on Form 10-Q and other documents filed by us from time to time with the SEC. Forward-looking statements speak only as of the date they are made. You are cautioned not to put undue reliance on forward-looking statements, and no person assumes any obligation to update or revise any such forward-looking statements, whether as a result of new information, future events or otherwise.

# Accelerated infrastructure portfolio for AI data centers



## Custom Compute

XPUs, CPUs, DPUs, NICs



## Interconnect

PAM, coherent, coherent-lite, and AEC DSPs  
PCIe retimers, DCI modules, TIAs, drivers



## Network Switching

Switches for scale-up  
and scale-out fabrics

**Full-lifecycle custom cloud solutions**

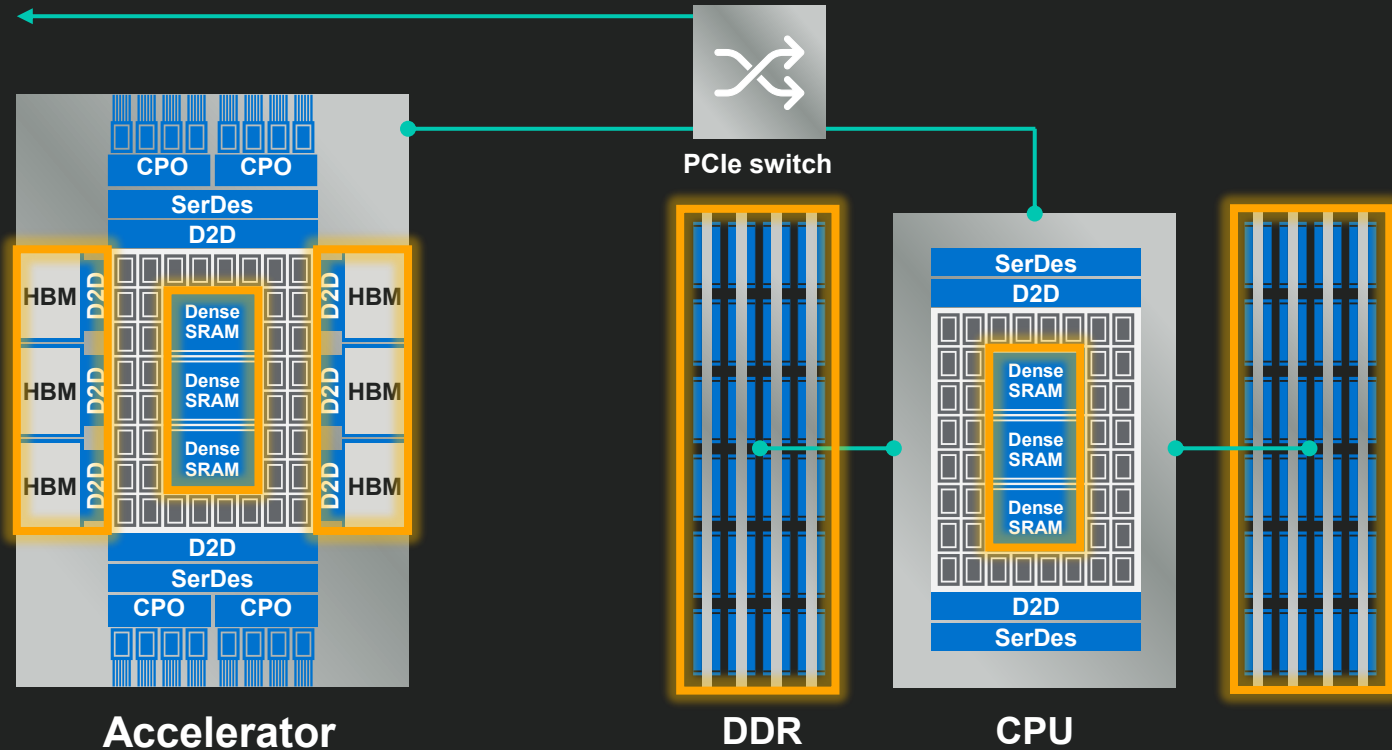
# The hot of about

?

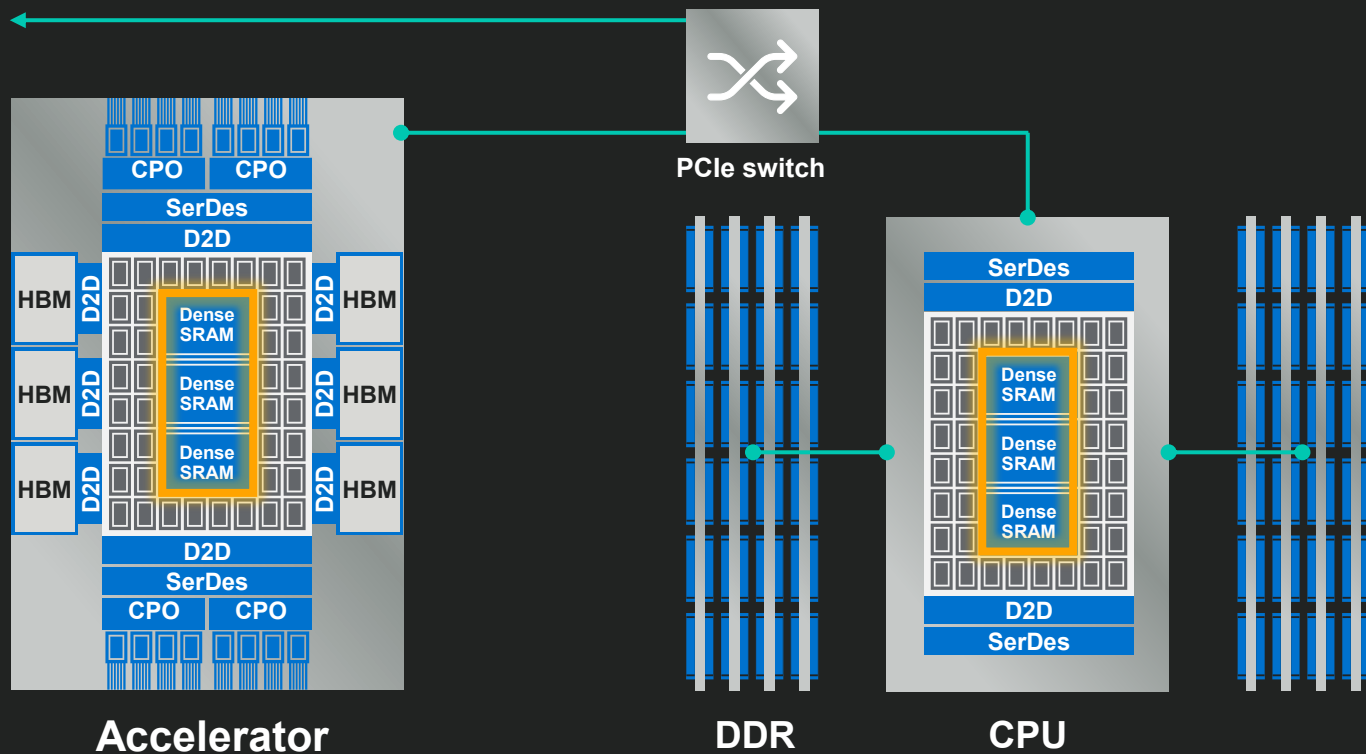
?

?

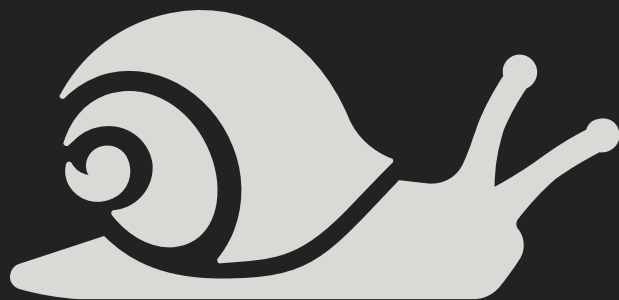
# Why is memory important?



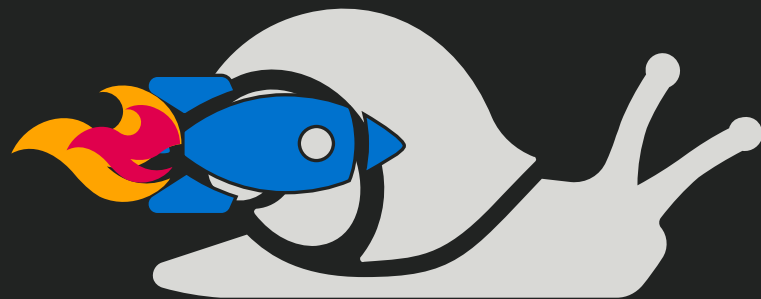
# Embedded SRAM



# It's all about bandwidth and capacity



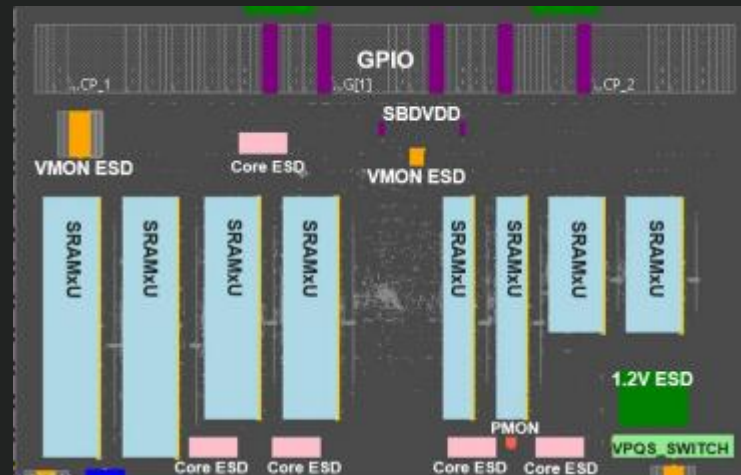
**SRAM bitcell scaling**



**Marvell innovation**

# Marvell® dense SRAM in 2nm

- Marvell key SRAM IP on first shuttle of every key technology node
- Optimized dense SRAM central to many key data center products, from XPU's to switches to NICs
- Marvell 2nm platform produced on TSMC 2nm technology

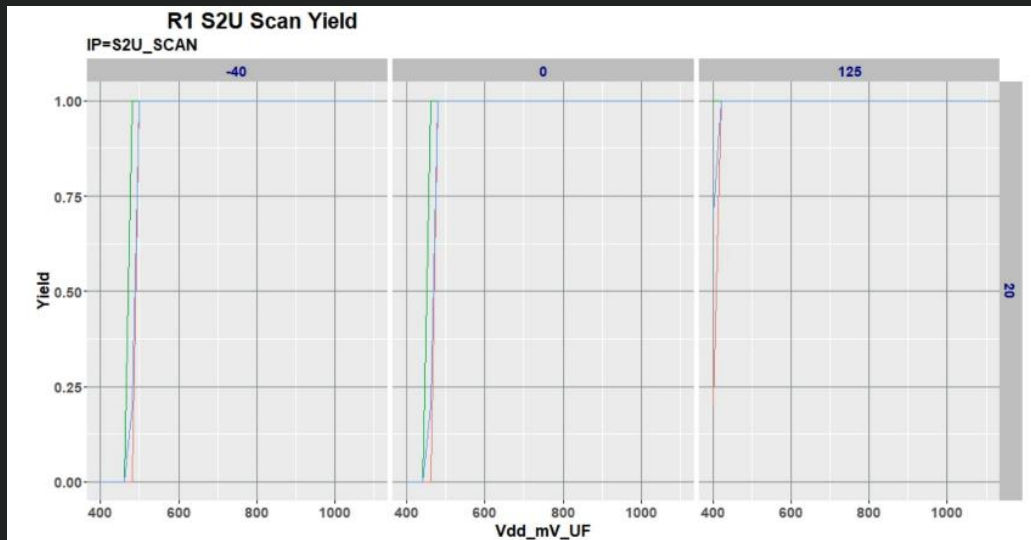




# Innovations to improve Vmins: Power = TCO

## Marvell implements key features to improve Vmins

- Write assist
- Stability assist
- Pioneered high-sigma design modeling to capture tail bits
- Row+column redundancy to enable low Vmins and high overall yields



## SRAM Vmins data from N2P hardware

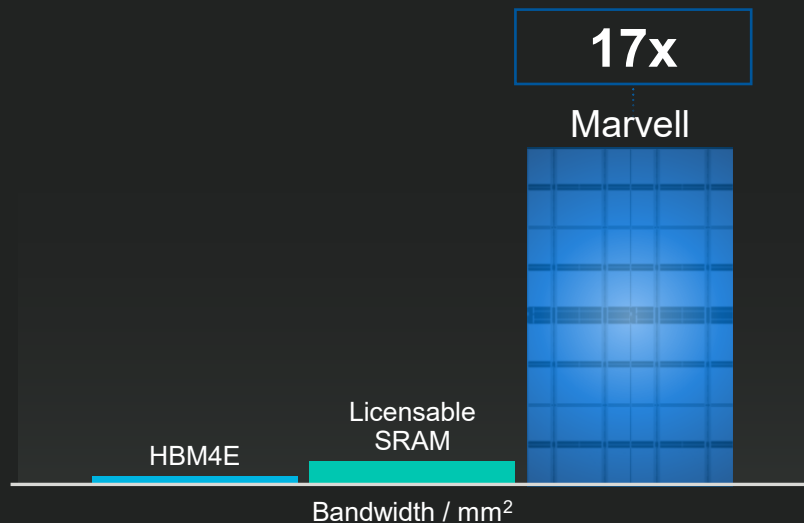
Lower than any previous gen ultra-dense bitcell memory

# Marvell custom SRAM

Key metrics for embedded memories

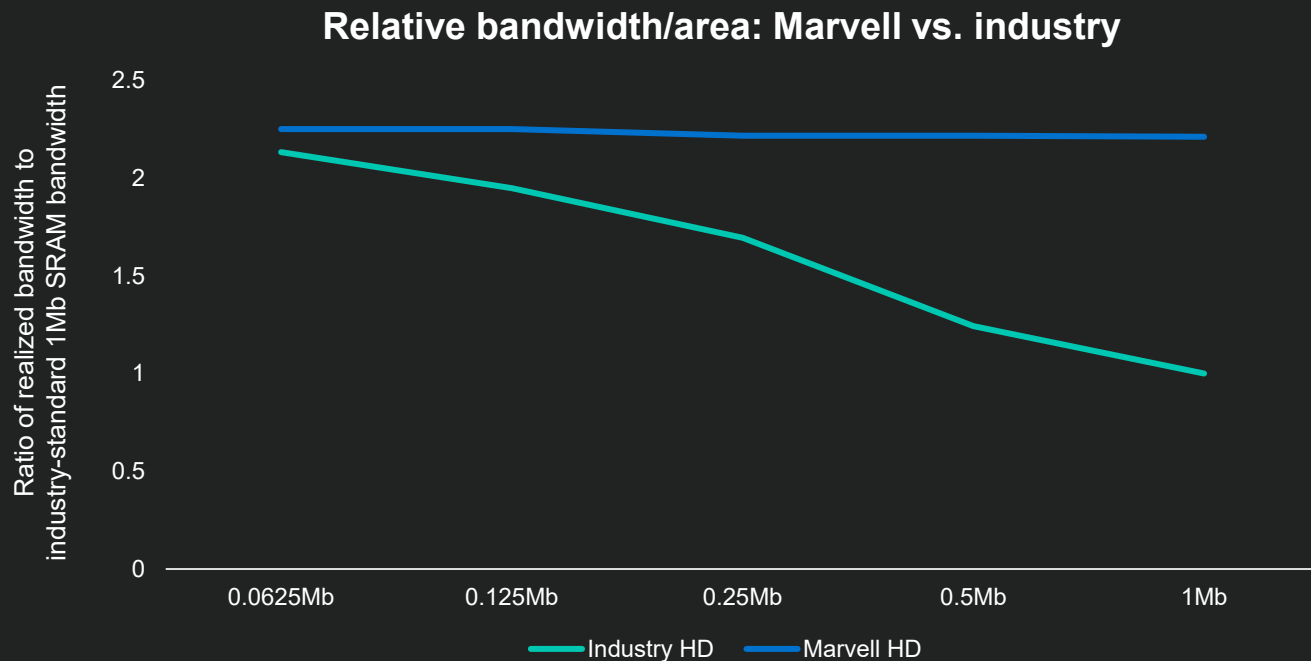
|                               | Marvell vs. licensable |
|-------------------------------|------------------------|
| <b>Area</b> at same bandwidth | <b>50% lower</b>       |
| Standby <b>power</b>          | <b>66% lower</b>       |
| <b>Bandwidth</b> at same area | <b>17x higher</b>      |

Bandwidth / mm<sup>2</sup> of memory options



Comparison of 1 Mb instance of third-party dense memories vs. Marvell

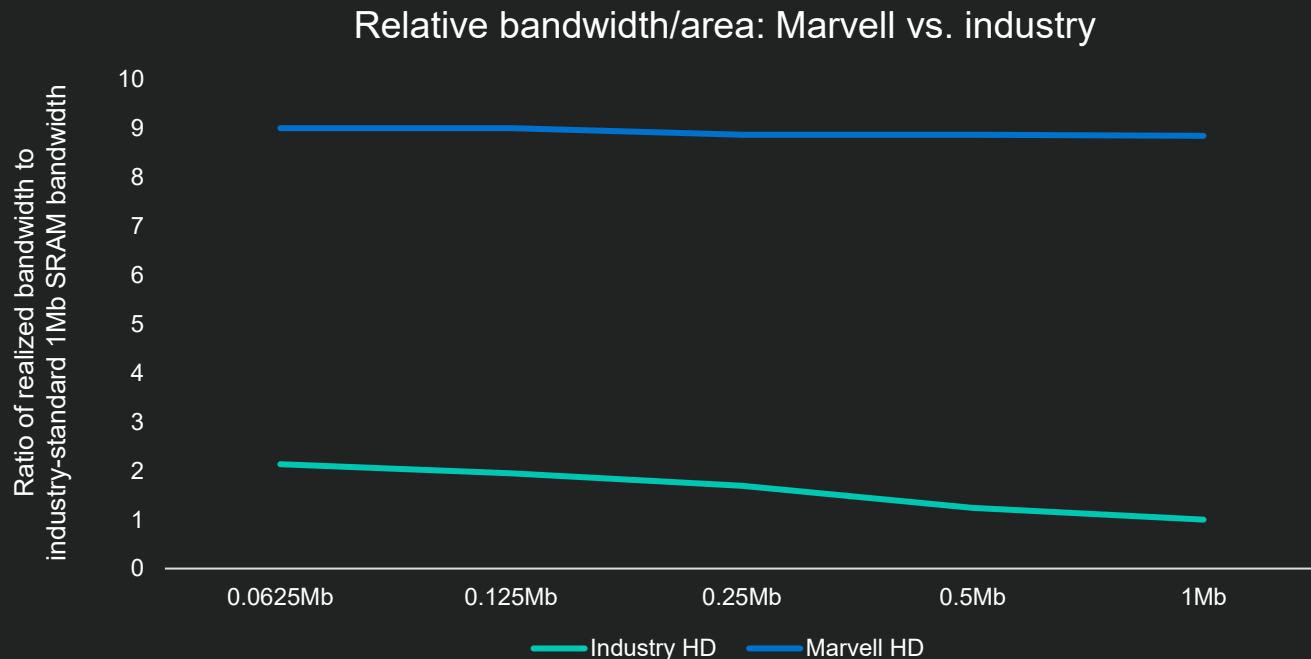
# How do we achieve this ridiculous bandwidth?



- **Run faster**
- Build wider
- More ports

**Use same array words/bits and bitcell**

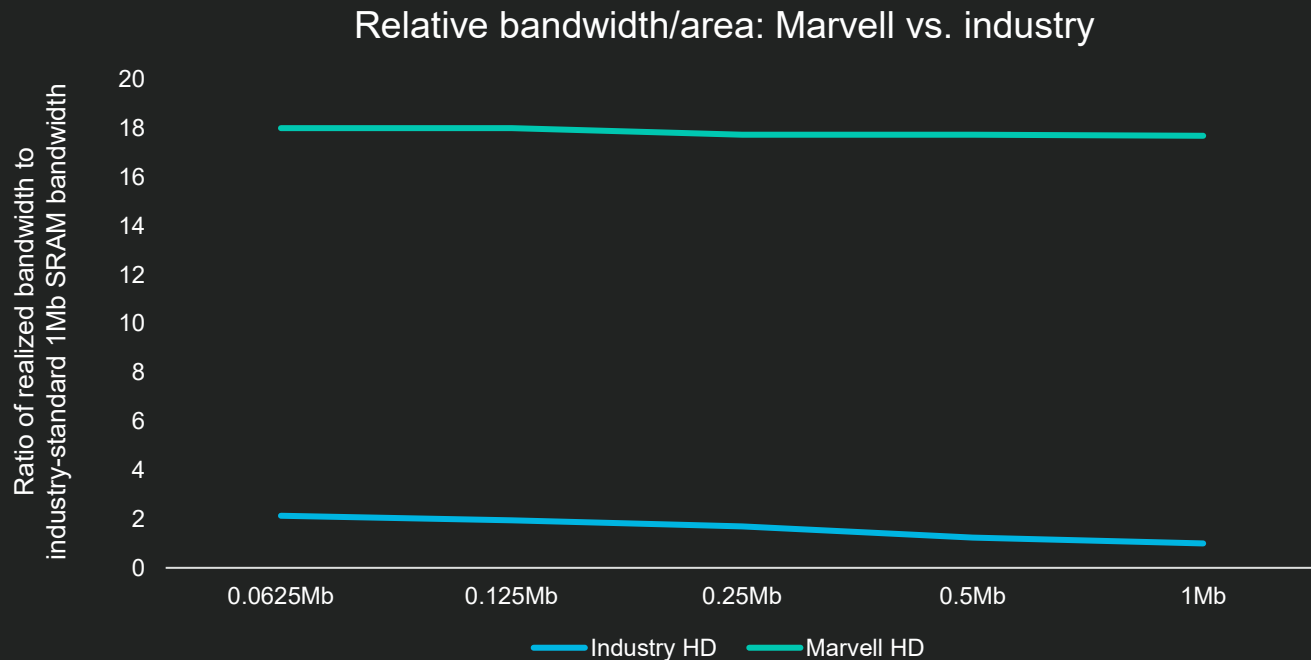
# How do we achieve this ridiculous bandwidth?



- Run faster
- **Build wider**
- More ports

**Use maximum Marvell memory width**

# How do we achieve this ridiculous bandwidth?

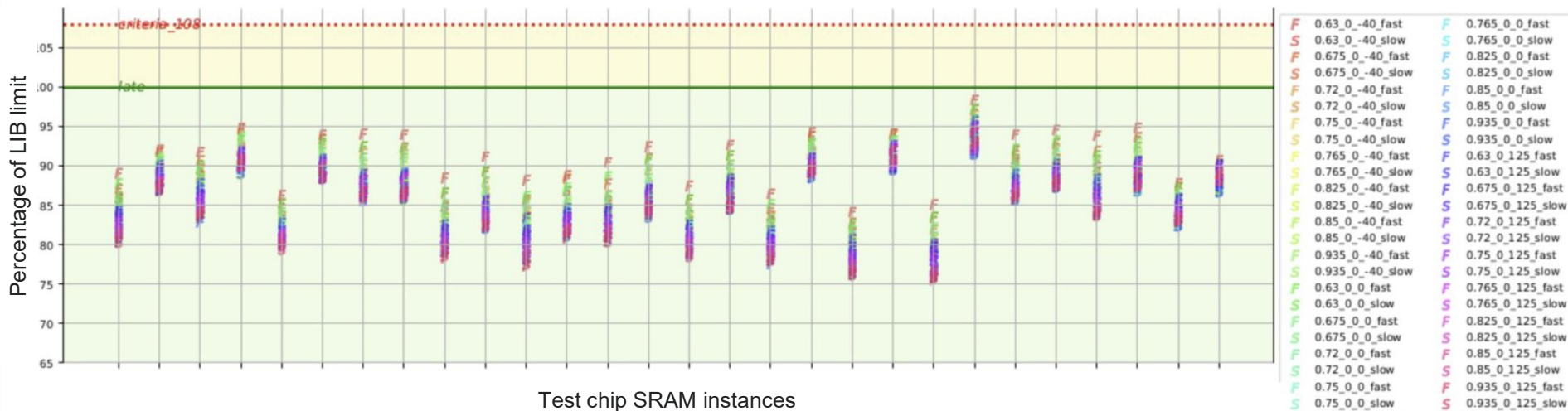


- Run faster
- Build wider
- **More ports**

**Use maximum width and ports**

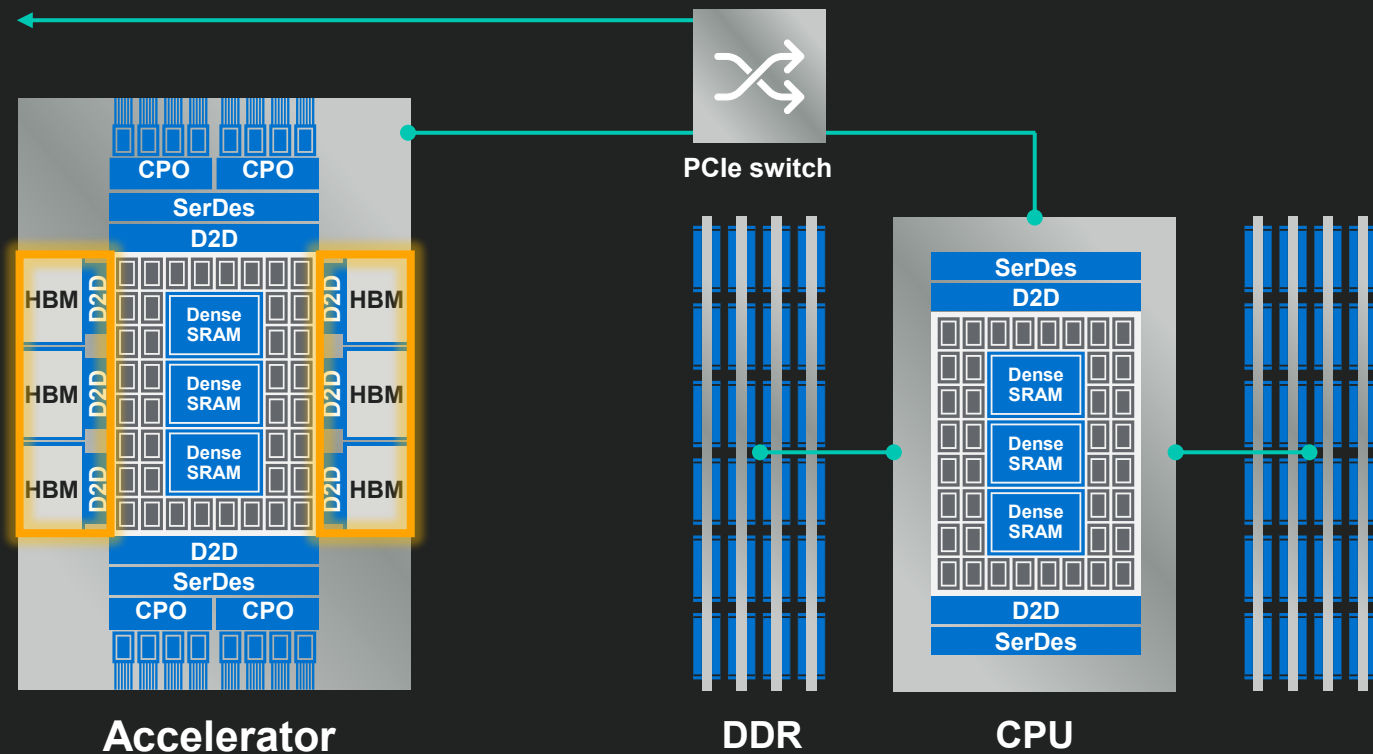
# Hot Chips is **all about hardware proof**

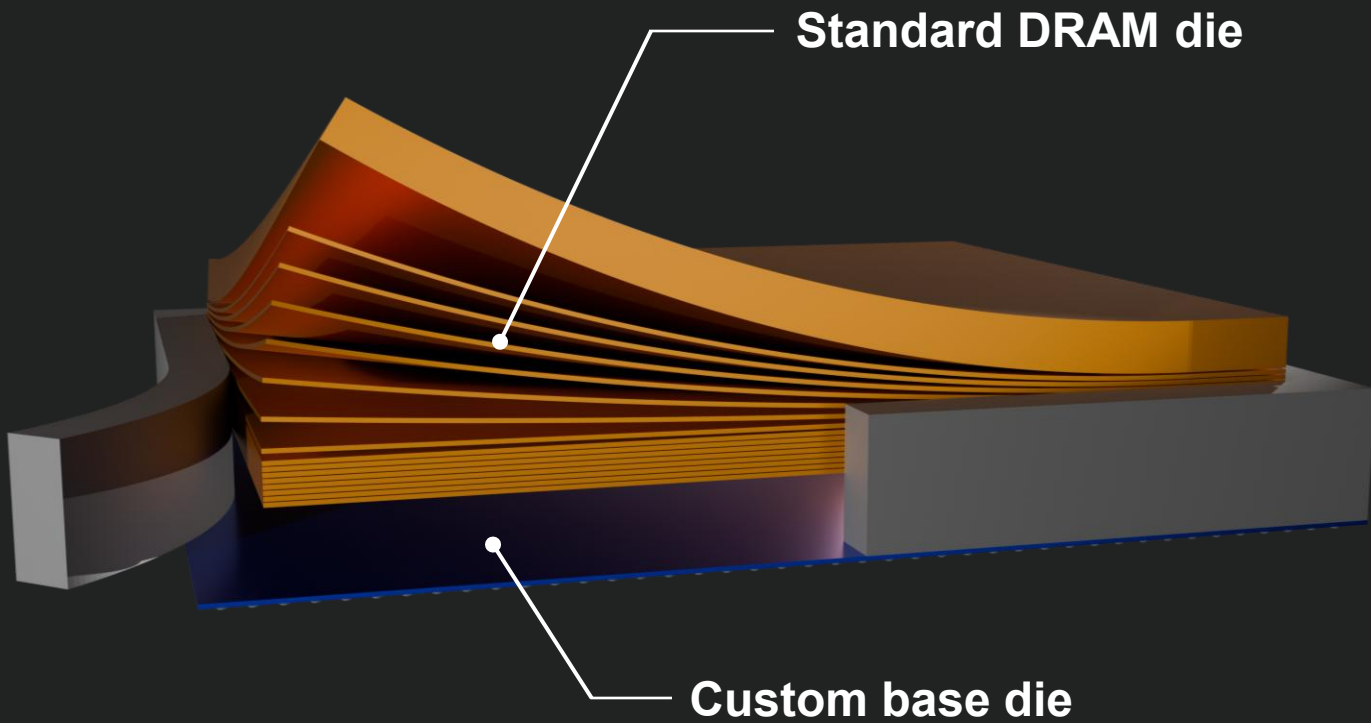
Marvell N3P dense SRAM performance vs. LIB limit across PVT (5-way split lot)



- Hardware characterization for 1Mb instance shown in comparison graphs
- Significant margin to LIBs without escapes
- Same LIBs used for comparison graphs

# Custom HBM base die

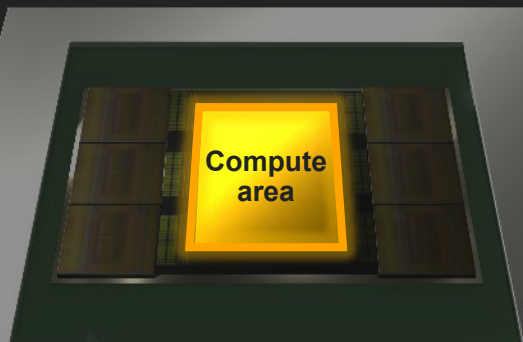






# Enhancing XPU's with custom HBM architecture

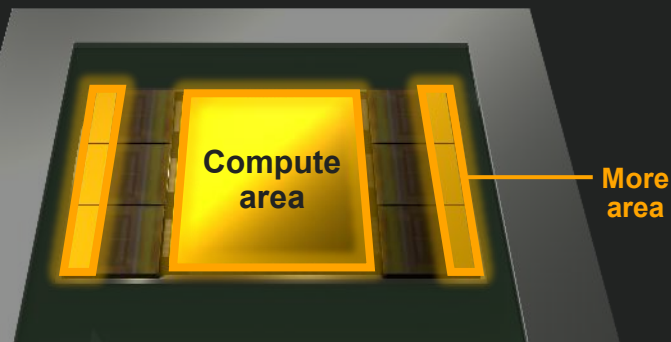
## Standard HBM No chiplets



Standard HBM XPU

- 1X useful compute area
- No compute area on HBM
- 1X power

## Marvell Custom HBM With I/O chiplets



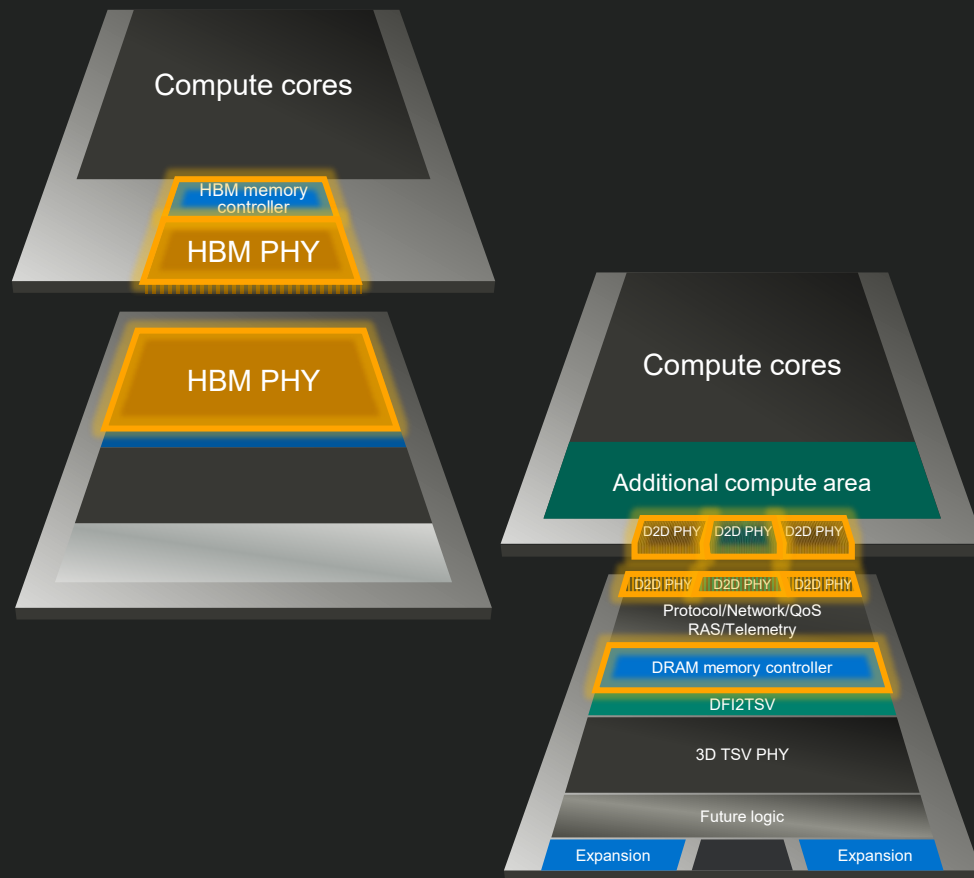
Custom HBM XPU

- **1.7X** useful compute area
- **More compute** area on HBM
- **75% lower** HBM & main die I/O power

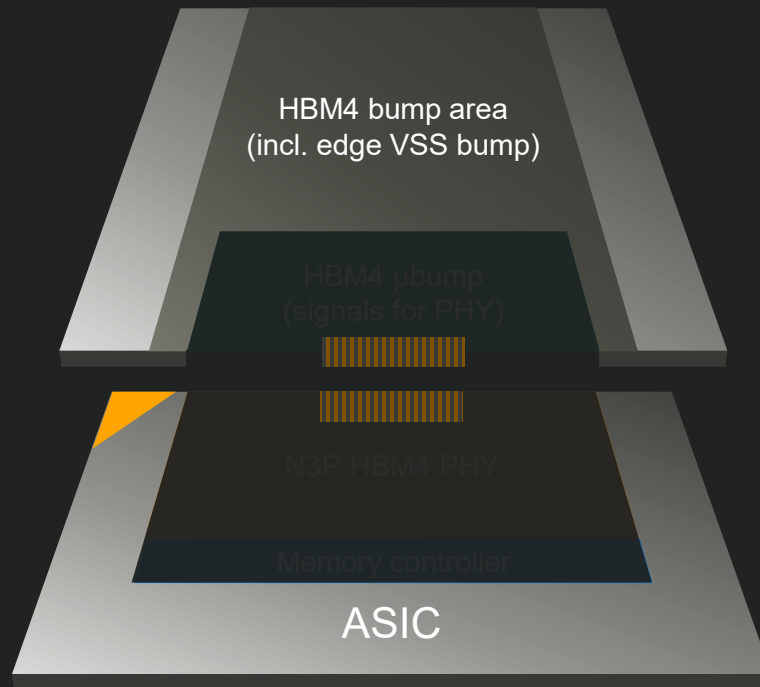
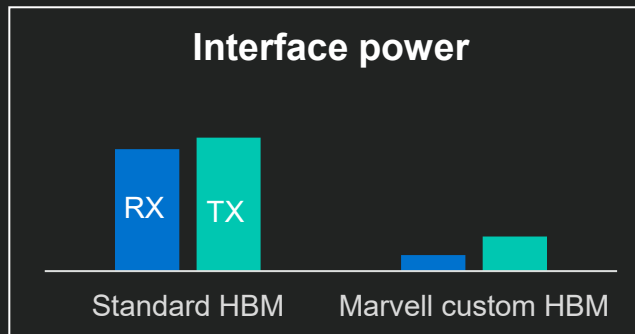
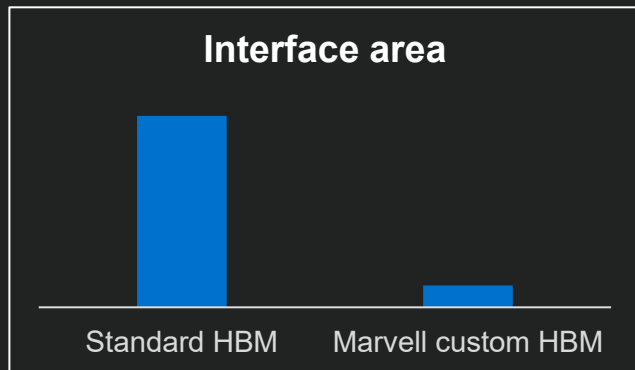
# Marvell custom HBM definition

## Key content

- Standard HBM PHY → D2D PHY
- HBM memory controller → logic base die
- AI-focused / optimized D2D
- RAS and telemetry
- QoS



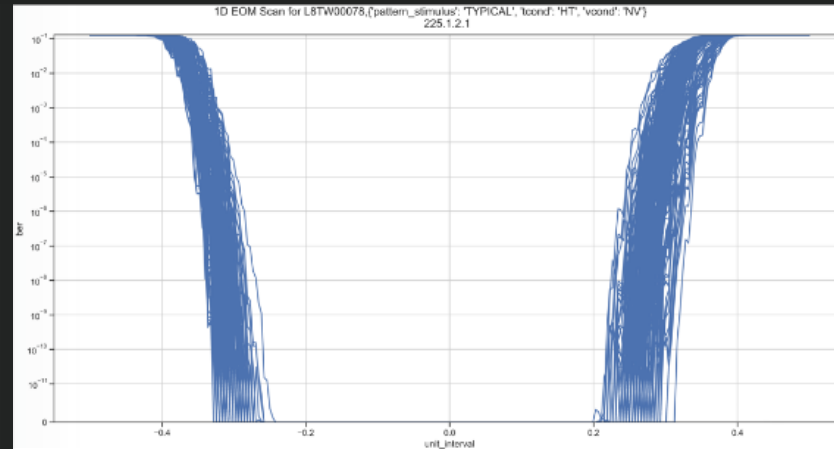
# Key comparisons



Dec 2024: <https://www.marvell.com/blogs/custom-hbm-what-is-it-and-why-its-the-future.html>

# Marvell D2D IP powers custom HBM

- Marvell D2D IP enables data rates a generation ahead of standards
- **32 Gbps** demonstrated in **Q2 2023** vs. standards-based tapeouts in 2025
- Reducing power with increased data rates, standards-based implementations increasing



Eye-opening plot for 216 lanes of 32 Gbps IP  
Extrapolated BER << 1E-30

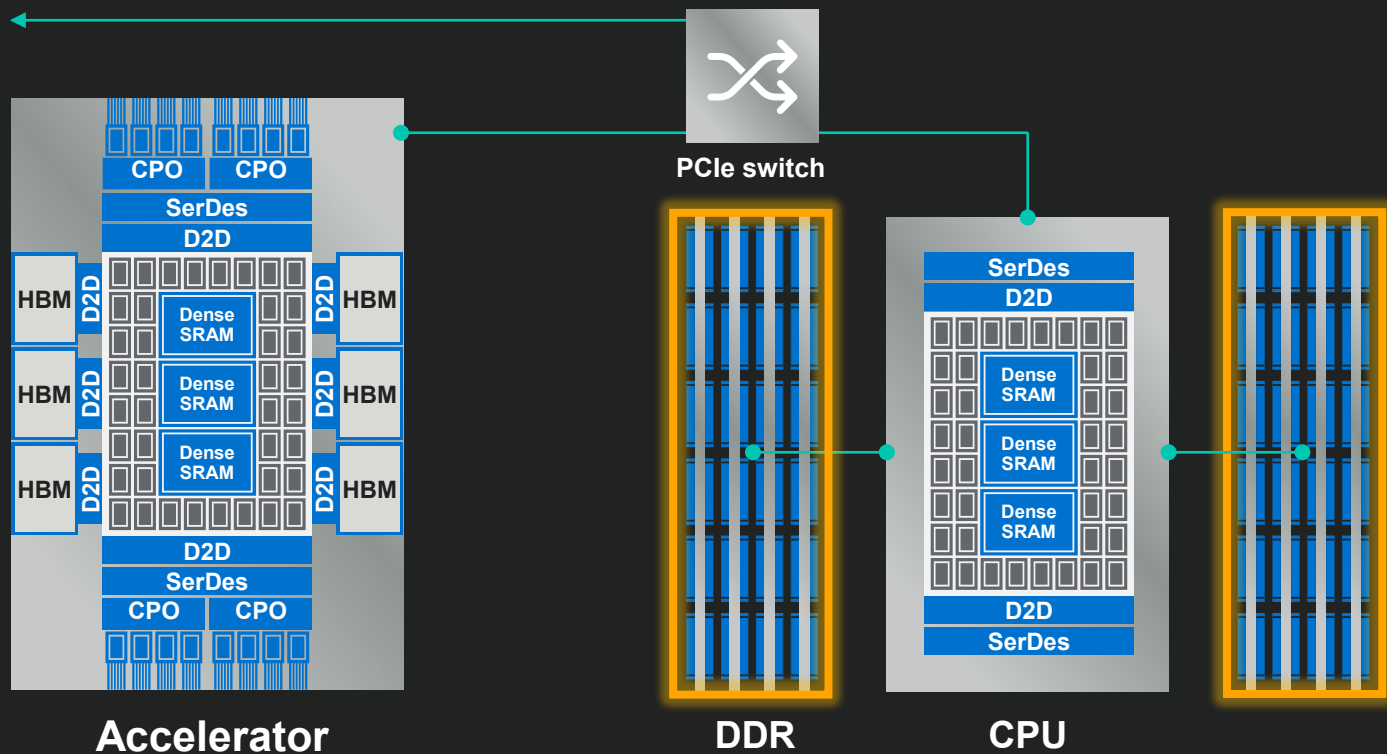
Custom HBM will use Marvell next-gen D2D at >30 Tbps/mm

# Marvell Unveils Industry's First 64 Gbps/wire Bi-Directional Die-to-Die Interface IP in 2nm to Power Next Generation XPU

- New D2D interface IP offers more than 3x bandwidth density of equivalent UCle interface while requiring far less silicon
- Advanced power management capability automatically adapts to bursty data center traffic, significantly lowering power consumption
- Delivers unmatched bandwidth, efficiency, and resiliency for AI infrastructure

August 26, 2025

# High-capacity memory

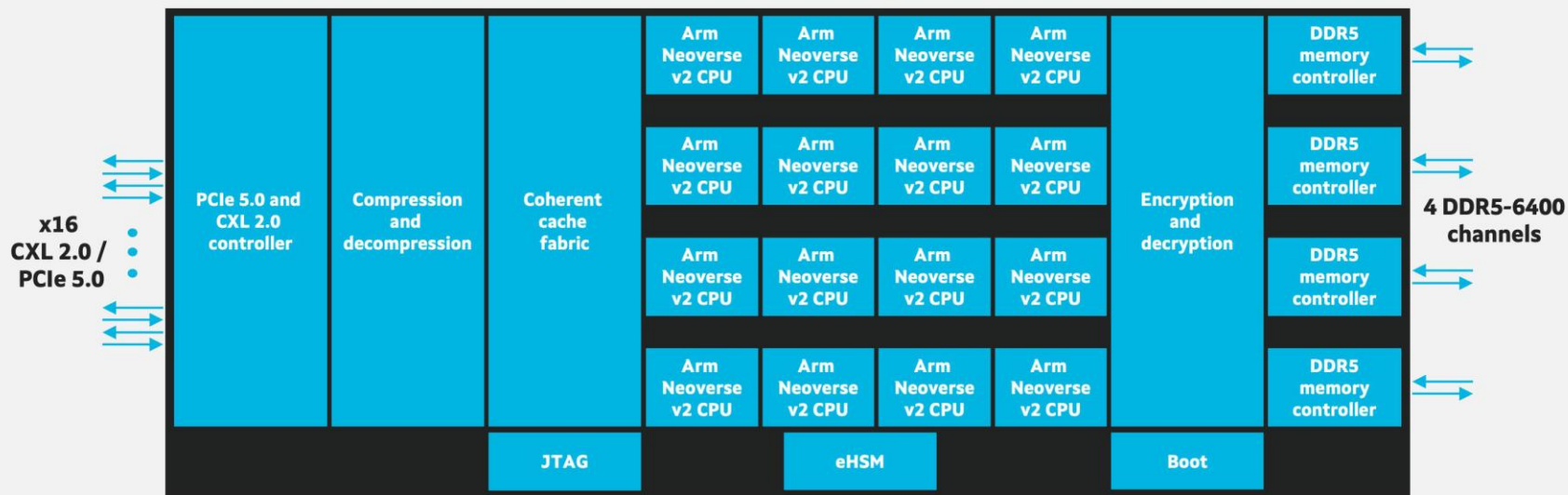


# Expand to high-capacity memory

- Lower latency to memory
- More bandwidth
- Pool memory between multiple devices
- Processor power (if needed)
  - More flexible use of memory
  - Remove CPU entirely



# Structera A near-memory accelerator in blocks



© 2025 Marvell. All rights reserved.

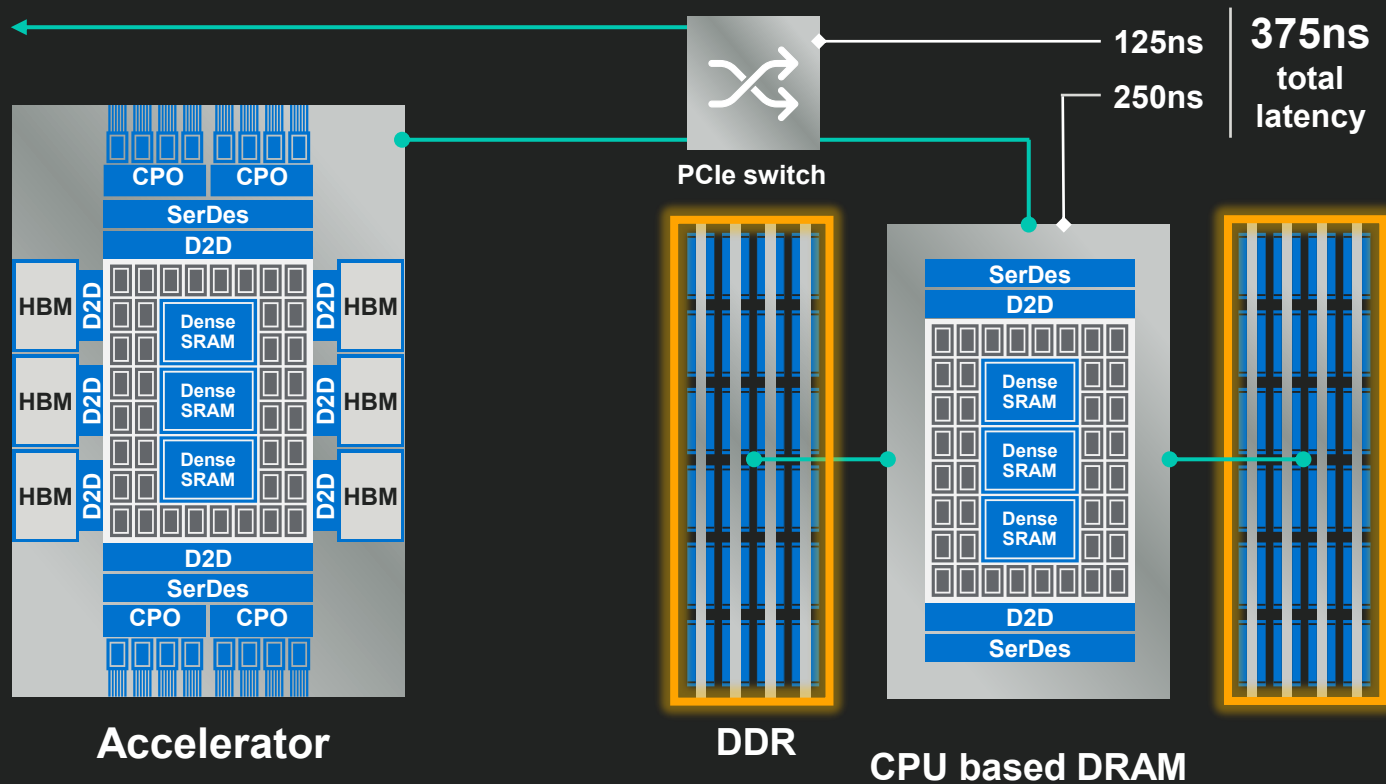


Essential technology, done right™

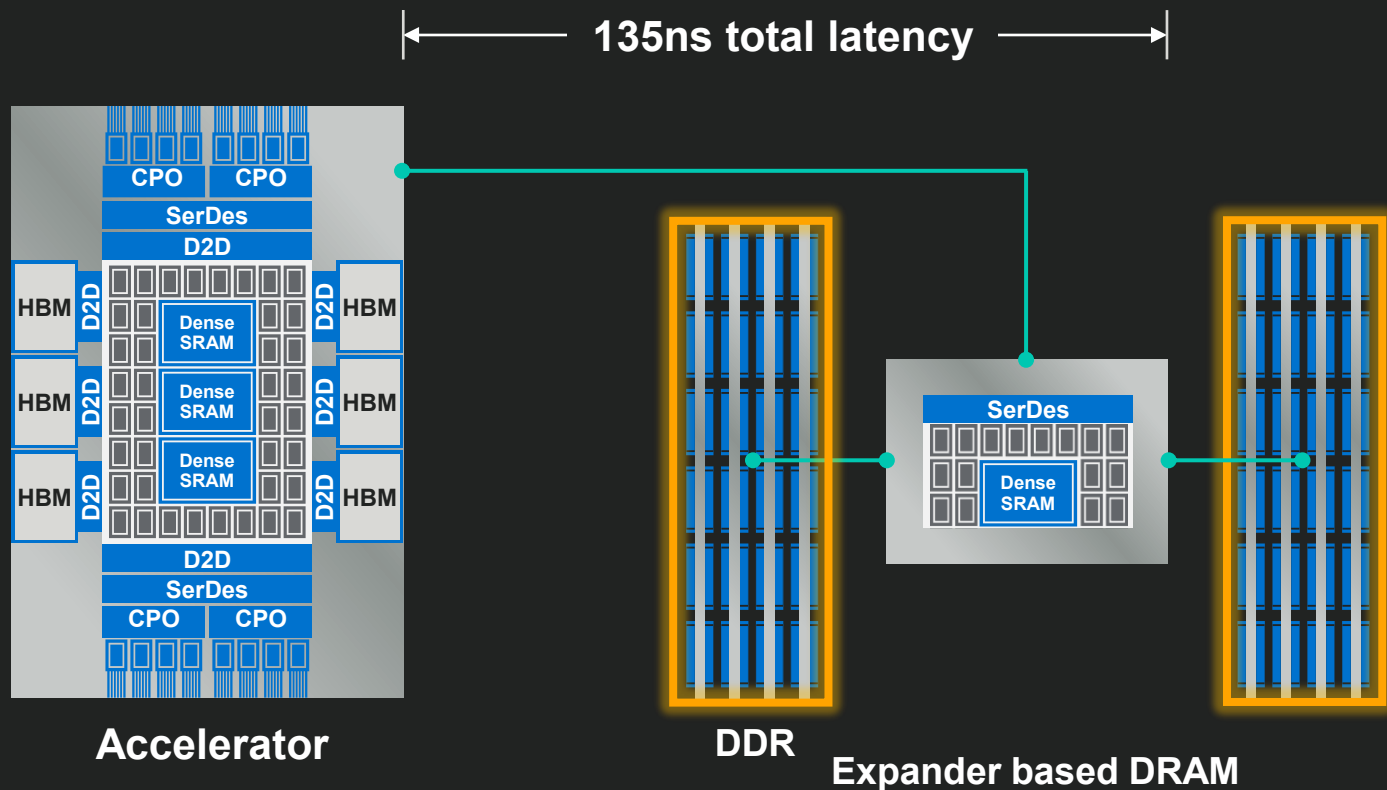
<https://www.marvell.com/blogs/five-ways-cxl-will-transform-computing.html>



# High-capacity memory (base case)



# High-capacity memory (expansion case)



# Summary: memory optimization at all levels

1

Dense SRAM architecture yields significant bandwidth improvement

2

D2D technology with low power, area, and error for custom HBM

3

Dense memory optimization with expansion and acceleration



Thank You





Essential technology, done right™



# Q&A