

ZeroPoint

Remove the waste.
Release the power.

About ZeroPoint Technologies

Prof. Per Stenström – CSO
Co-founder/inventor



- Professor at Chalmers University of Technology
- Internationally renowned memory expert
- Senior industry experience at Sun Microsystems
- Wide industry network

Dr. Angelos Arelakis – CTO
Co-founder/inventor



- Memory architecture and ultra-fast data compression expert
- Recipient of King Carl XVI Gustaf's award for science, technology and environment

Company founded 2016, over 50 man-years invested in R&D

20 people in R&D team

- Compression Algorithm Research & Development
- Memory Architecture
- System Software
- High Volume ASIC Design

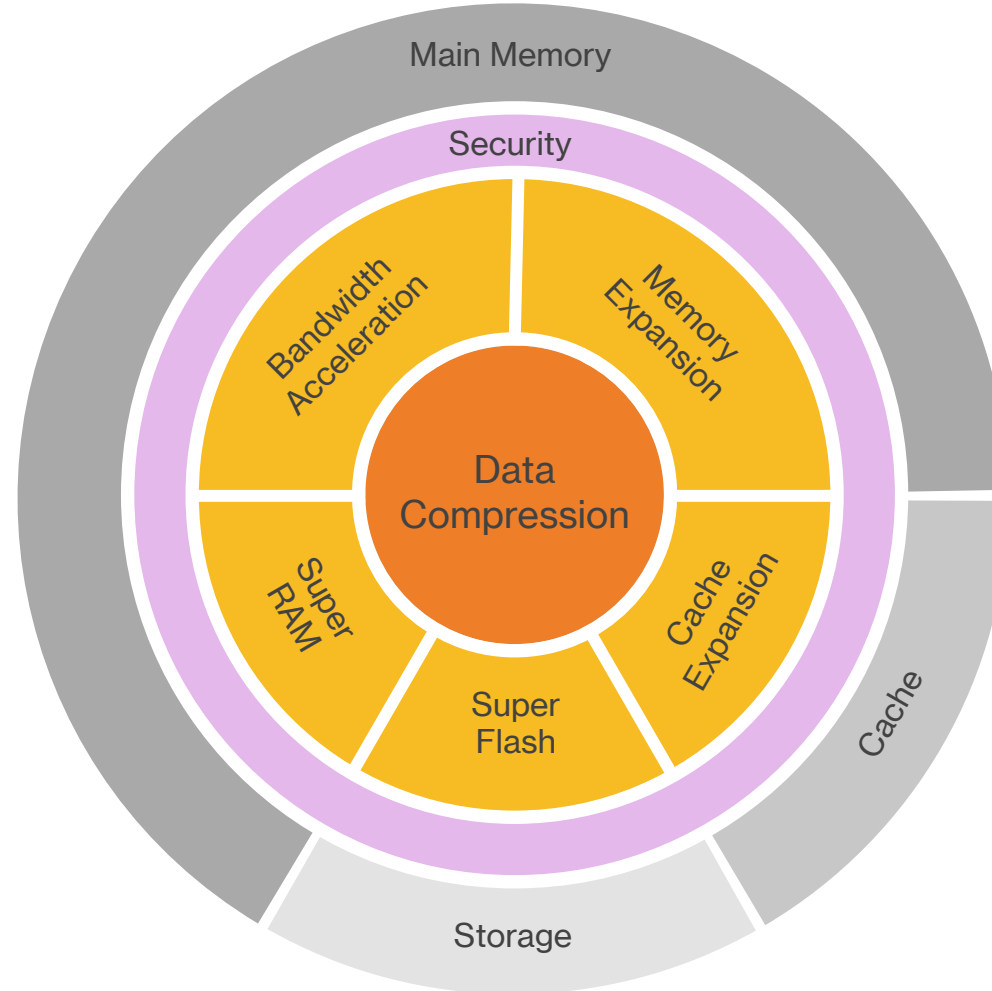
arm



The background of the slide is an aerial photograph of a city, likely Los Angeles, showing a dense grid of streets and buildings. The buildings are color-coded in various shades including blue, green, yellow, orange, red, and purple, creating a vibrant, abstract pattern. The perspective is from a high angle, looking down at the city grid.

Product overview.

Product overview.



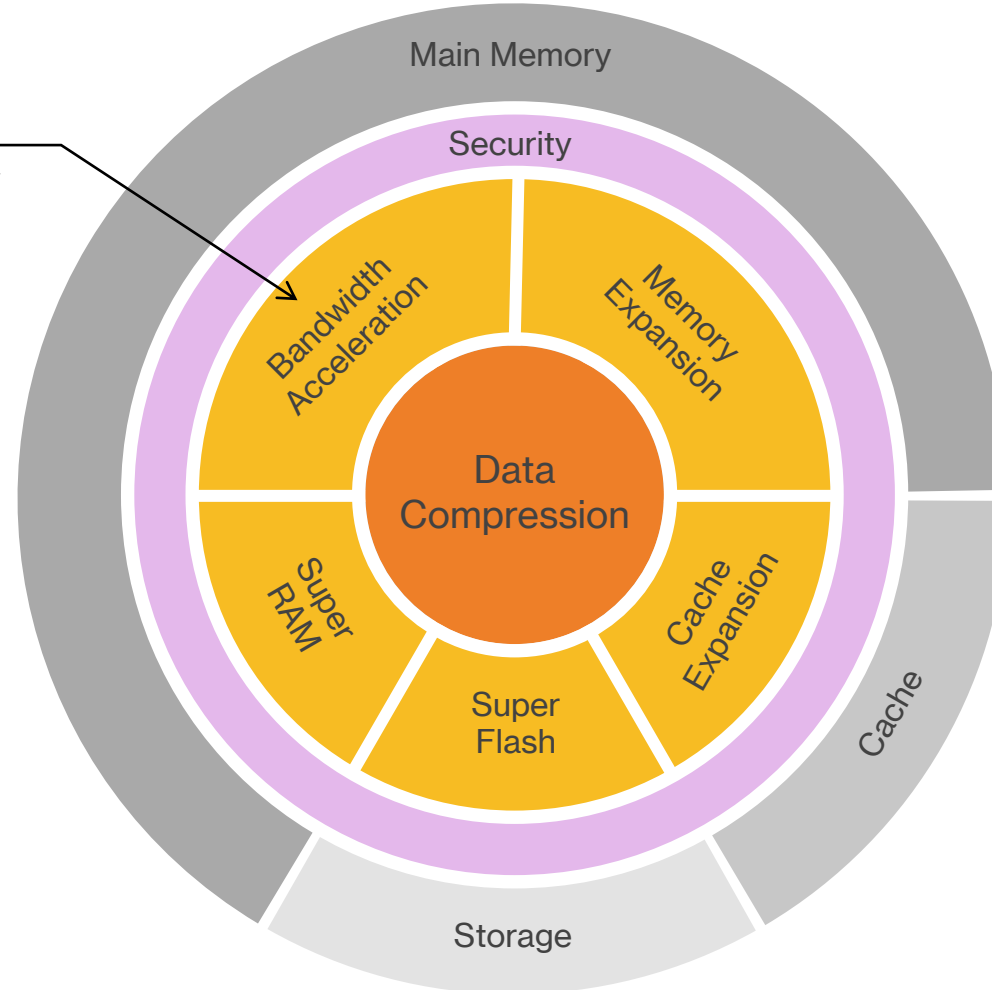
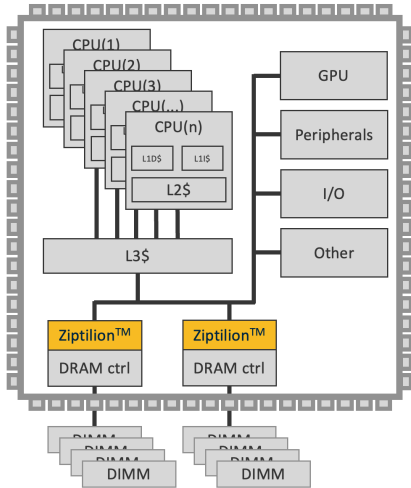
Product overview.

Ziptilion™ Bandwidth

Up to 50% main memory
bandwidth acceleration

Scheduled to be released: Q2-2023

Part of the EPI EPAC  European
Processor
Initiative



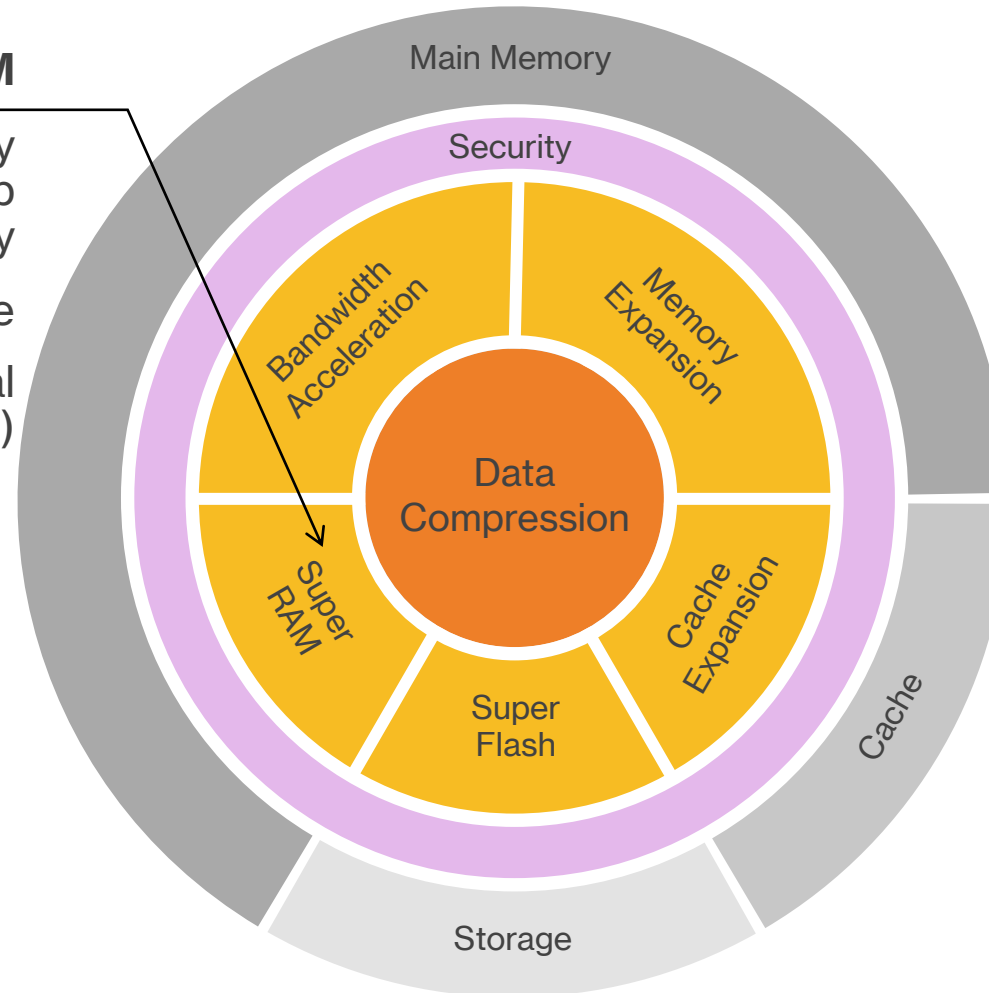
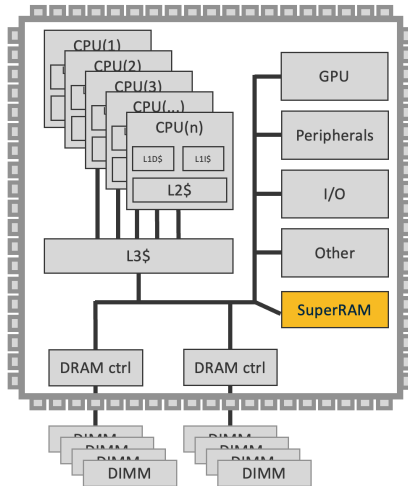
Product overview.

SuperRAM

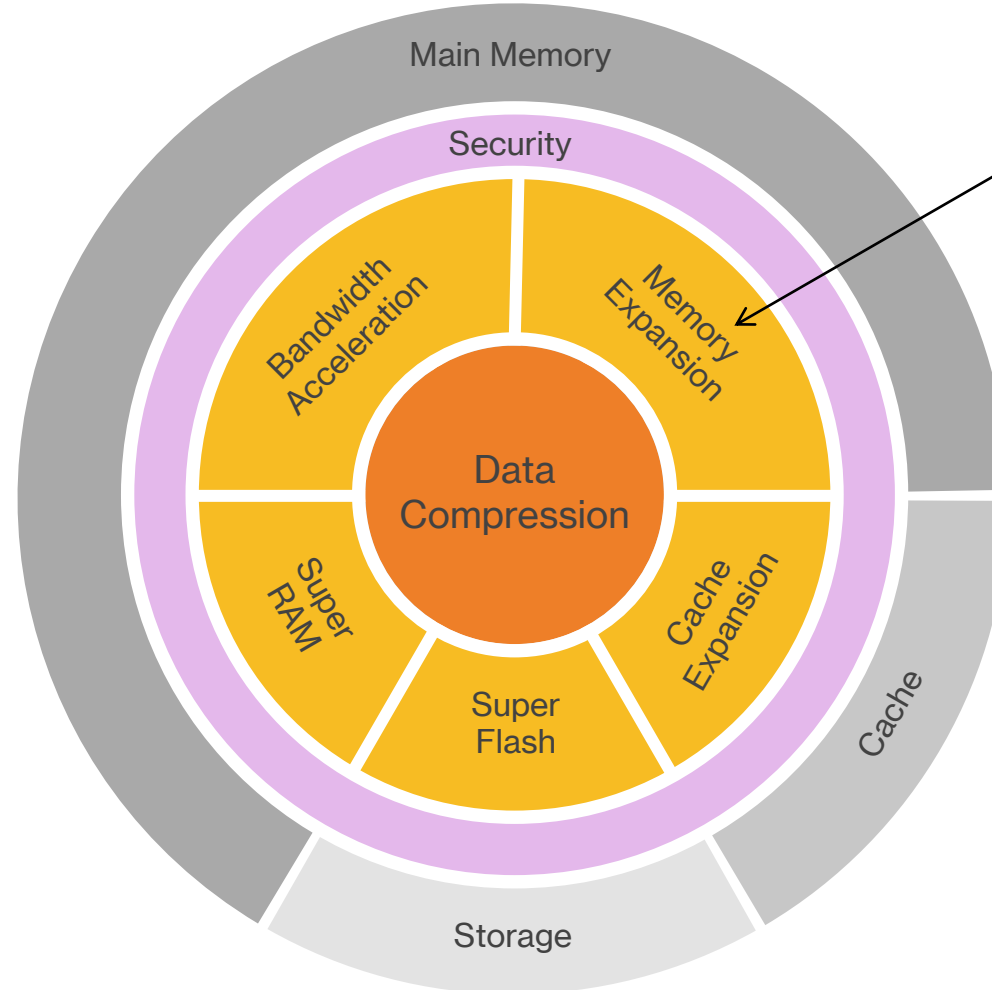
High performance and low latency
hardware accelerated zram/zswap
at unmatched power efficiency

Released – Product available

Due to be tape-out in a commercial
chip (TSMC 5nm)



Product overview.

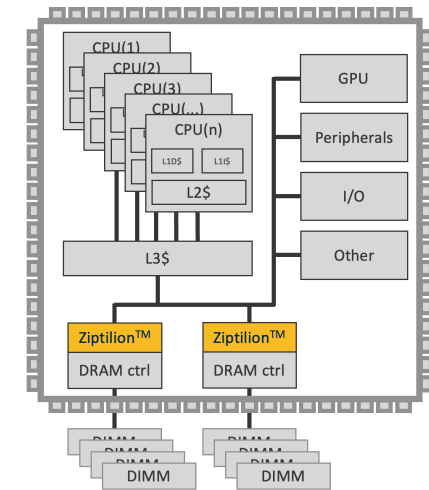


Ziptilion™ Memory Expansion

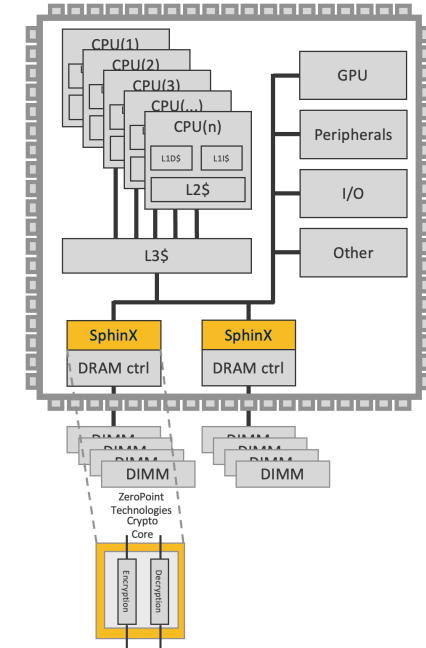
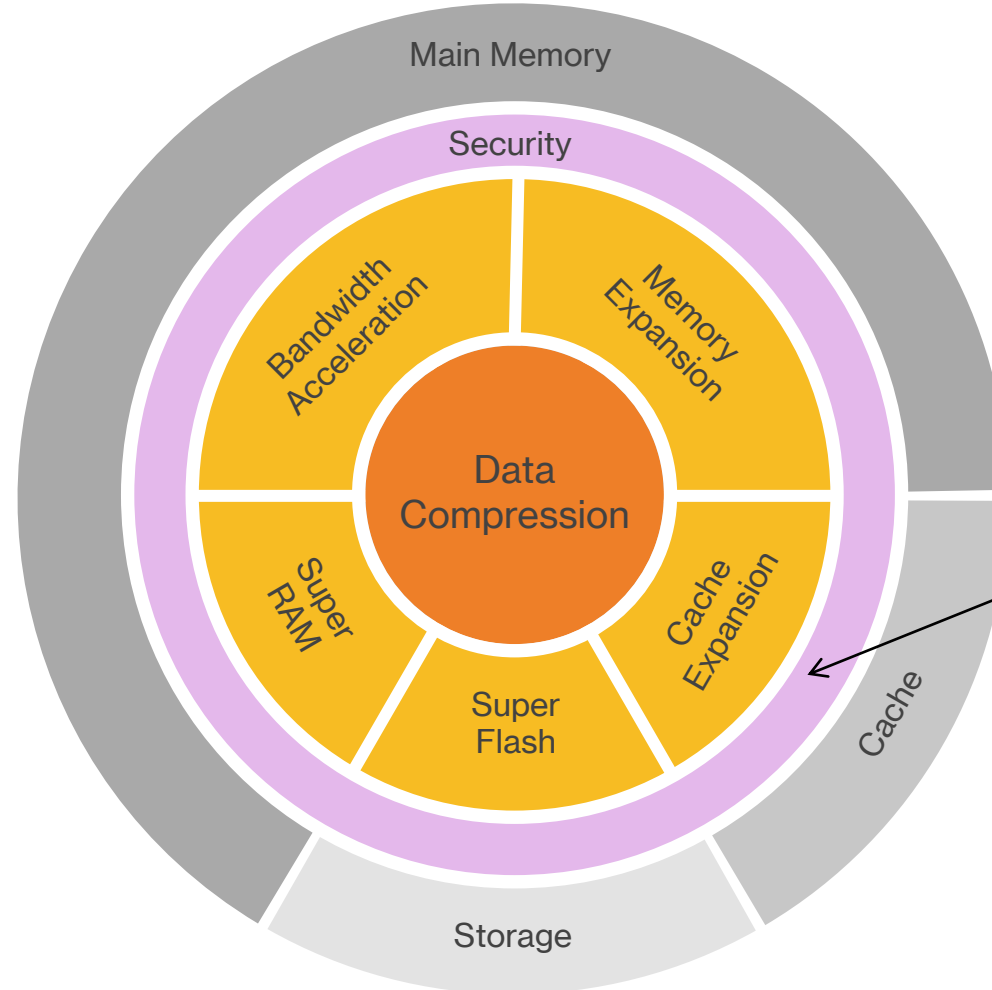
2-3x main memory expansion

For SoCs and CXL devices

Scheduled to be released:
(contact for details)



Product overview.



SphinX AES-XTS Security IP

An easy to integrate off the shelf
AES-128/256-XTS high throughput,
low latency server main memory
security solution

Released – Product available

Verified on silicon in a commercial
server chip (TSMC 7nm)

DenseMem™

Hardware-accelerated memory compression and management
for CXL-based memories

Why CXL?

Why CXL™?

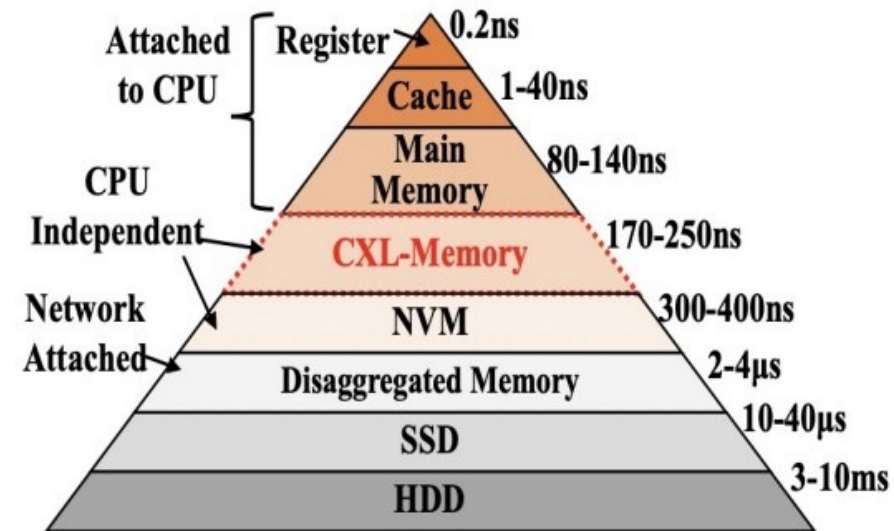
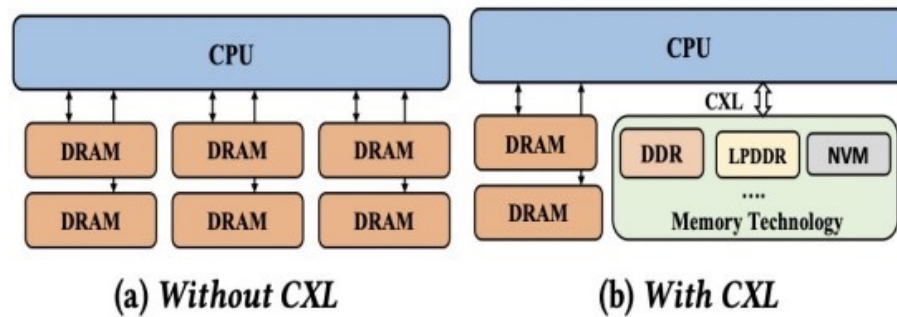
Future data centers need heterogeneous compute, new memory and storage hierarchy, and an agnostic interconnect to tie it all together

CXL can overcome CPU-memory and memory-storage bottlenecks faced by computer architects

Maintains memory coherency between the processor memory space and memory on attached devices

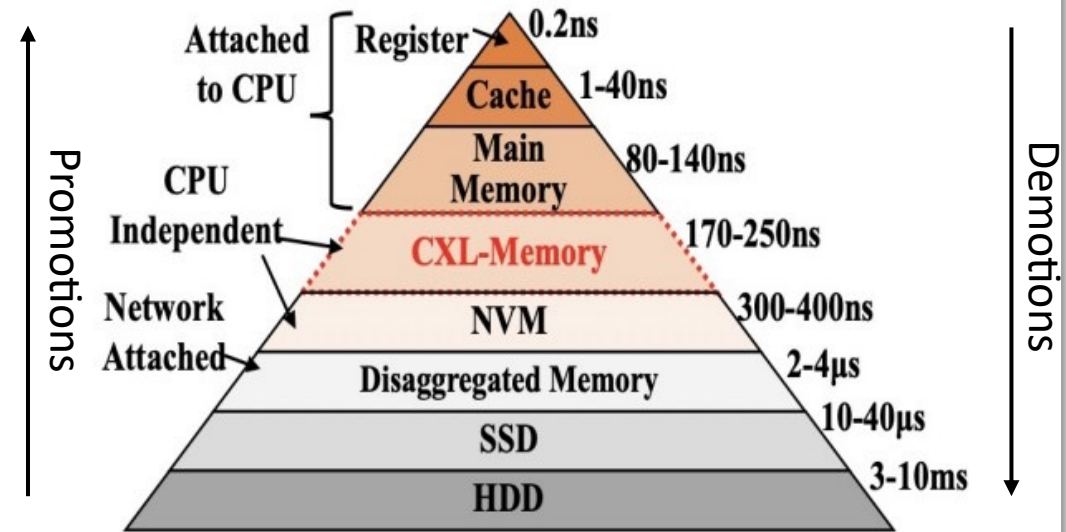
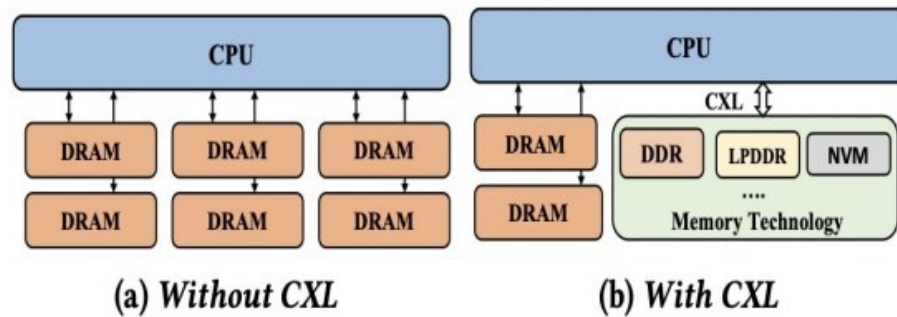
Enables pooling and sharing of resources to provide higher performance, reduce software stack complexity, and lower overall system cost.

Tiers of happiness



Reference: <https://www.computeexpresslink.org/>

Tiers of happiness



Reference: <https://www.computeexpresslink.org/>

DenseMem™

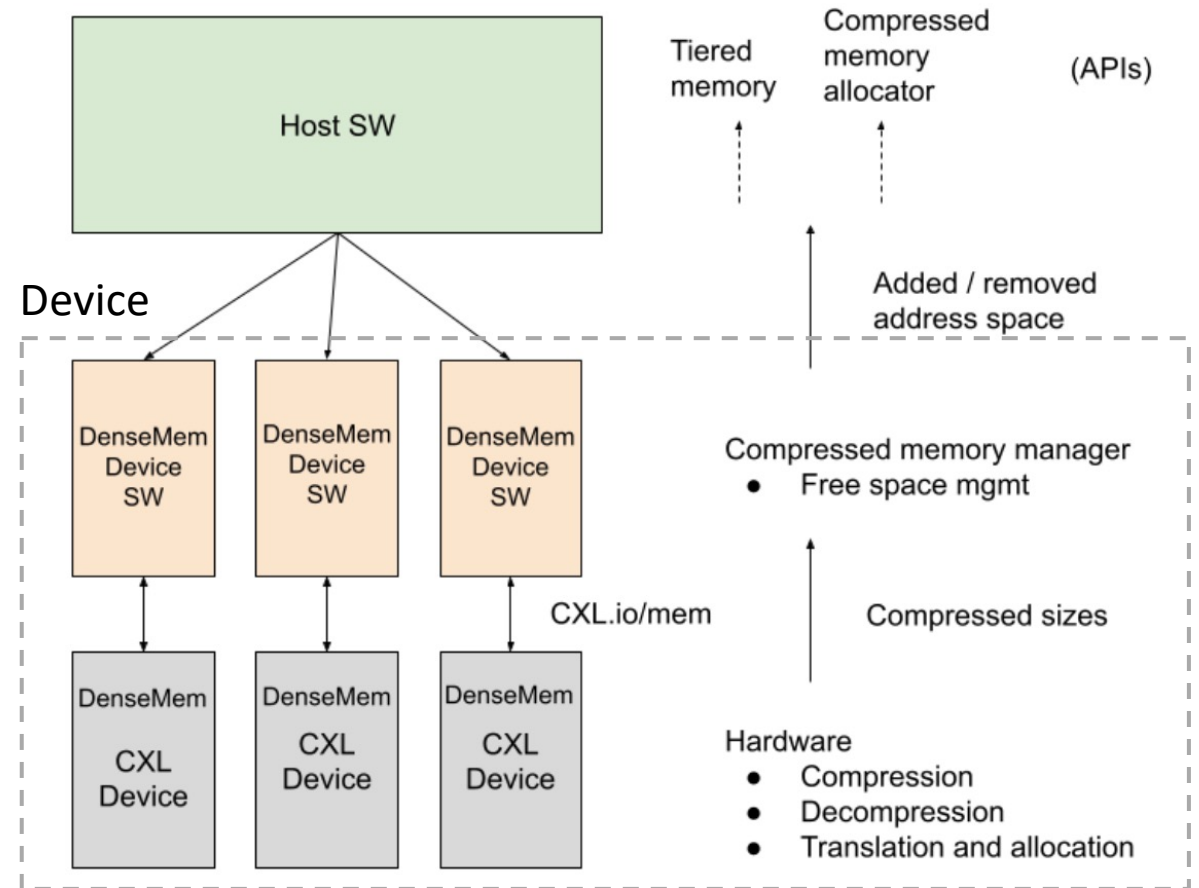
Overview

DenseMem™ solution overview

CXL - type 3: Baseline memory expansion or memory pooling

DenseMem: Memory compression and compressed memory manager

- DenseMem hardware
 - executes compression and decompression
 - handles transparently the compressed memory space
- DenseMem device driver
 - releases the free space to host system
- The compressed memory appears in the system as a memory tier



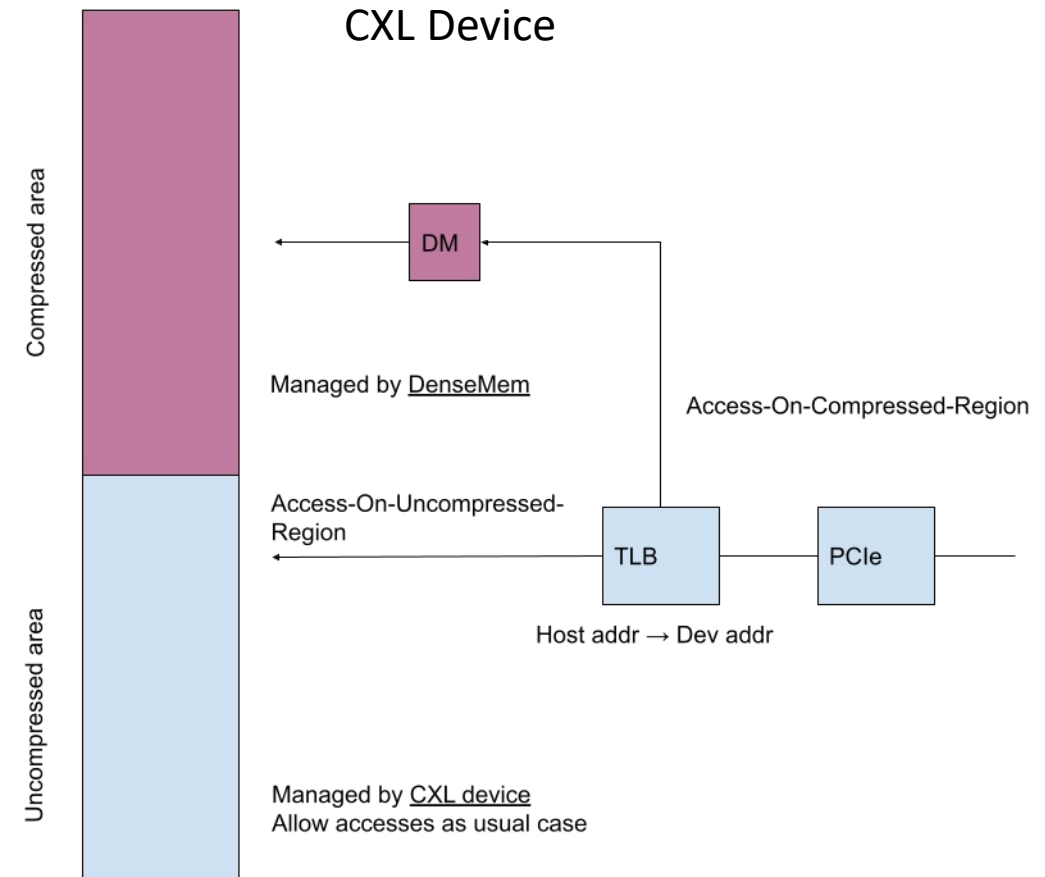
CXL device: Uncompressed and compressed area

CXL - type 3

Physical address space is reserved in a CXL device to be compressed, compacted and managed by the DenseMem IP

Two high-level memory areas

- **Uncompressed area:** Directly accessed and managed by the CXL device
- **Compressed area:** Accessed through the DenseMem (DM) IP
 - Reserved area, not directly accessible from device



DenseMem™: Exposing the compressed memory as Tier 3

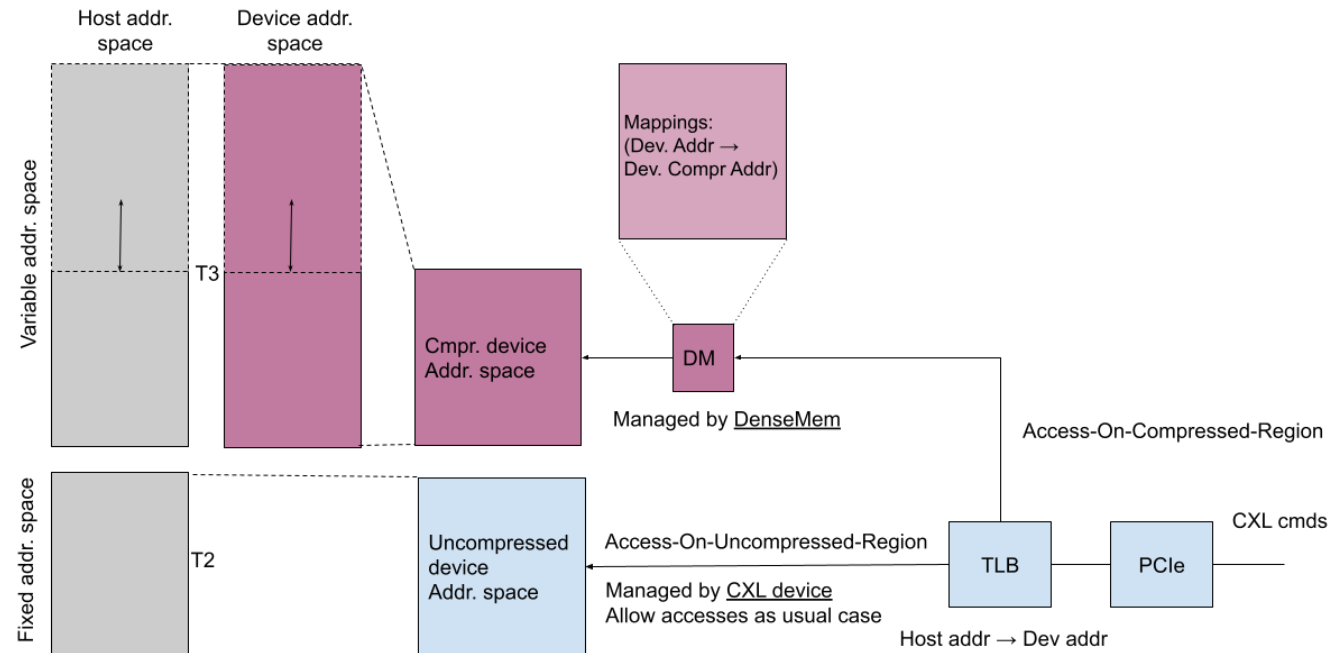
Compressed memory appears to host as a new Tier (T3), lower than the CXL uncompressed memory

T3

- Variable address space managed through ballooning and hot add/remove (CXL 2.0)
- The hot-add triggers a registration of an allocated host address space and respective mapping to 1:1 device address space

DenseMem

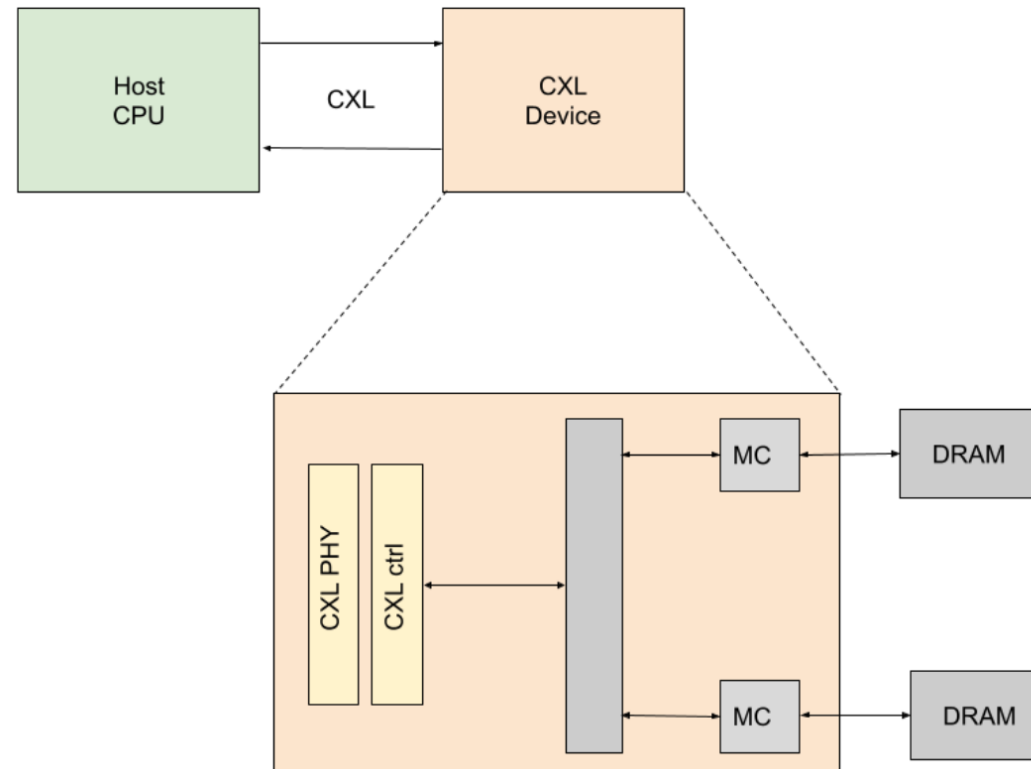
- Responsible to translate into the denser compressed address space
- Demotions/promotions to/from T3 are compressions/decompressions



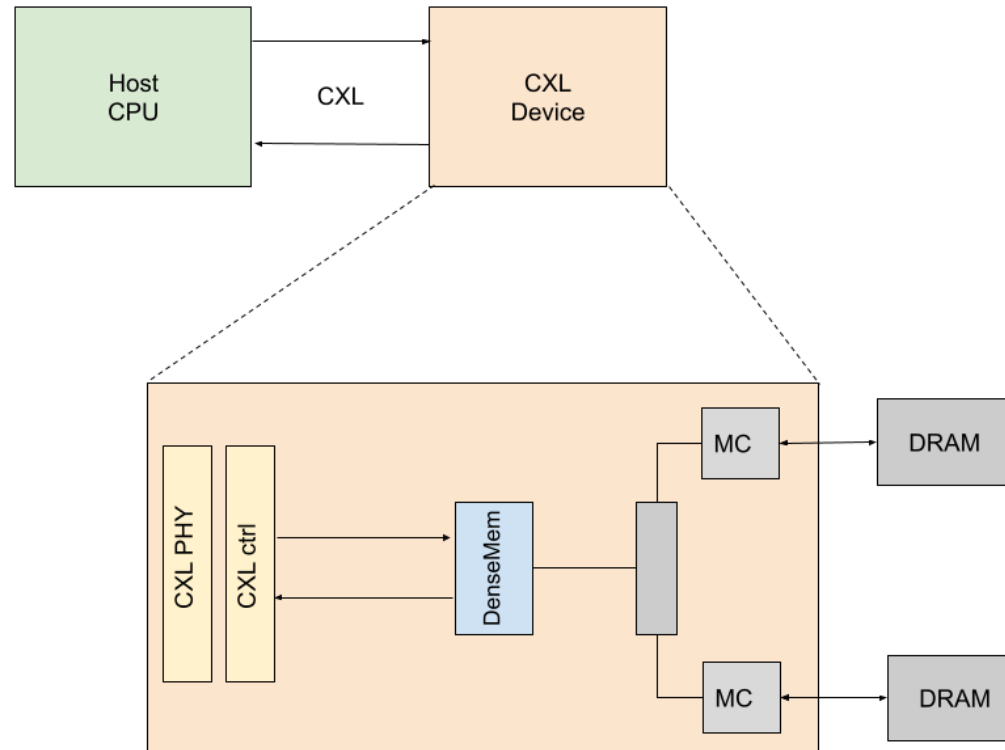
DenseMem™

Integration

Baseline CXL device



DenseMem™ – Integration in a CXL device

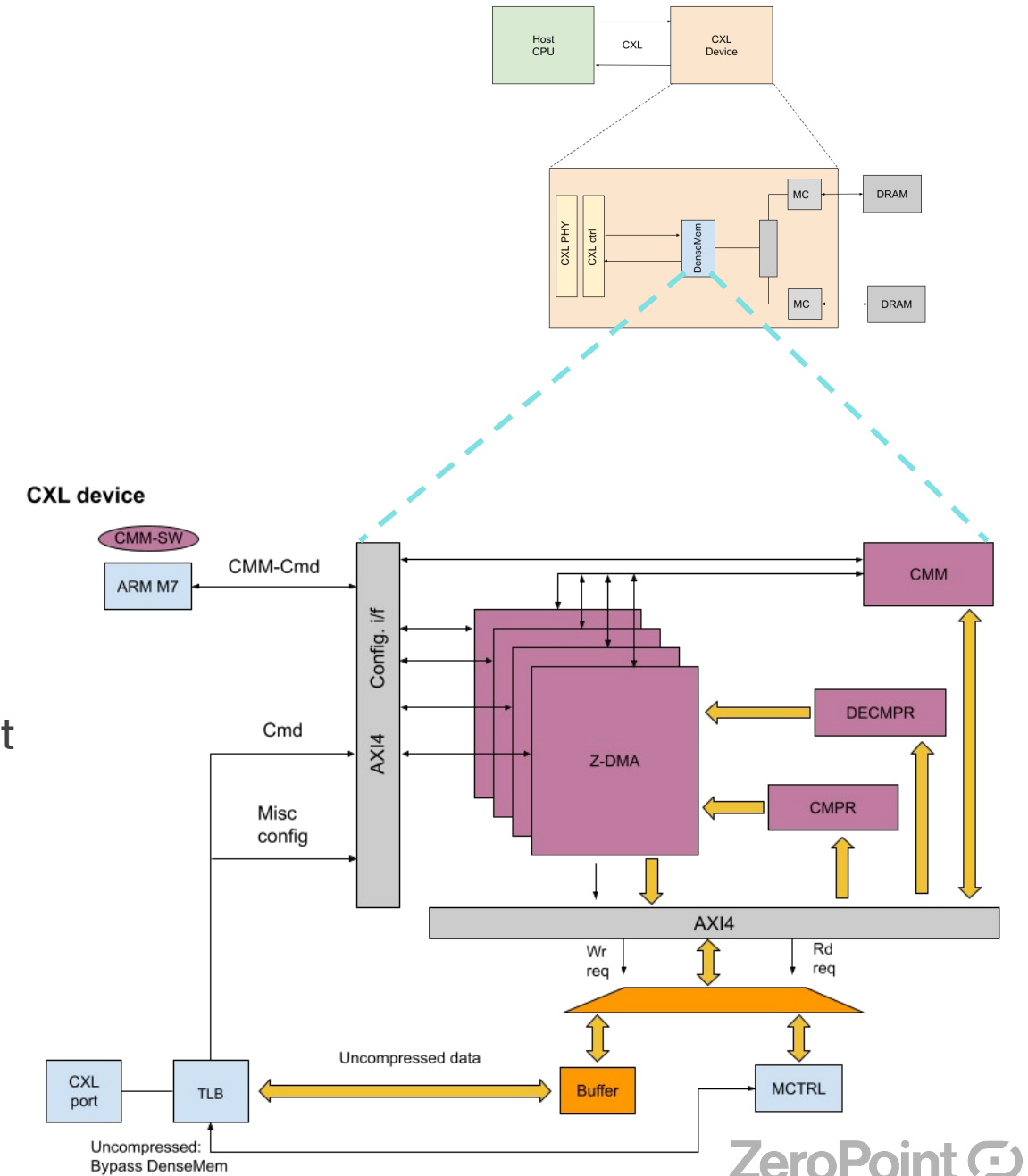


Inside DenseMem™

Blue blocks: Exist in the baseline CXL device

Orange blocks: Outside the DenseMem IP block but inside the CXL device

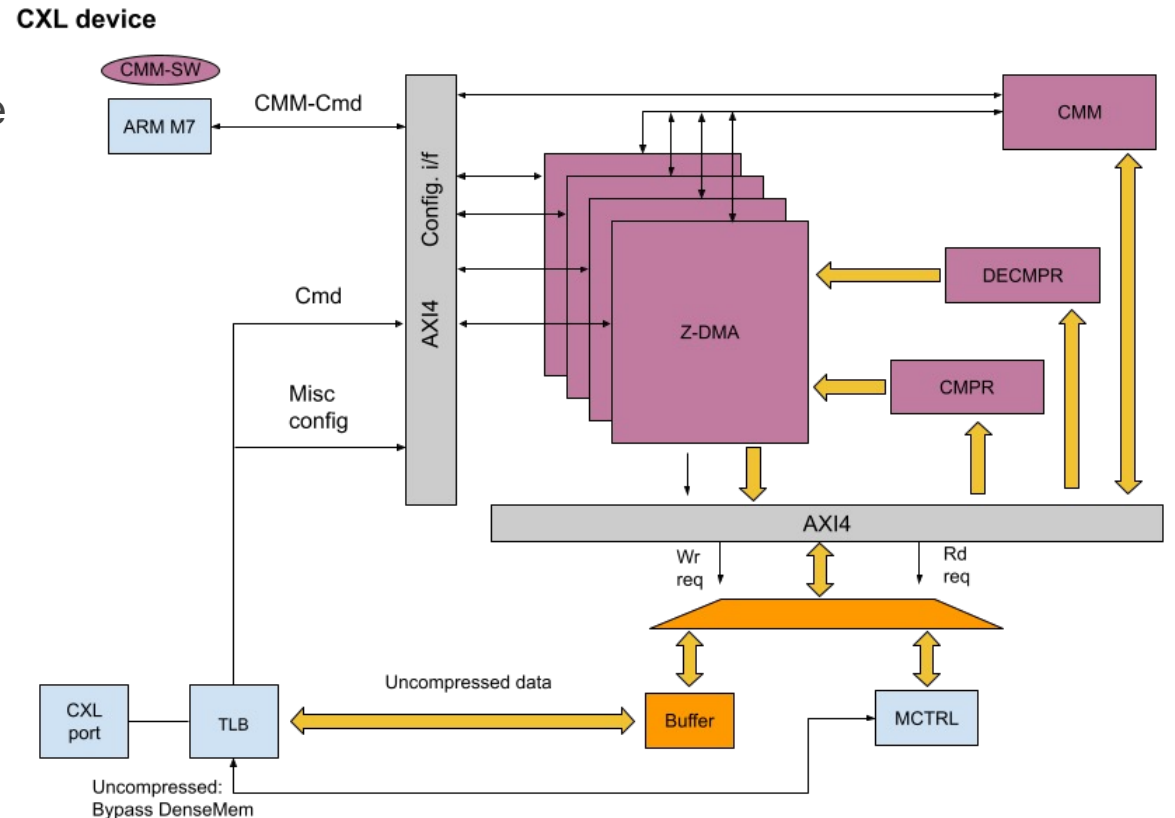
Purple blocks: Part of the DenseMem IP block



Inside DenseMem™ Operations

DenseMem handles promotion and demotion operations from/to the host as well as within the device

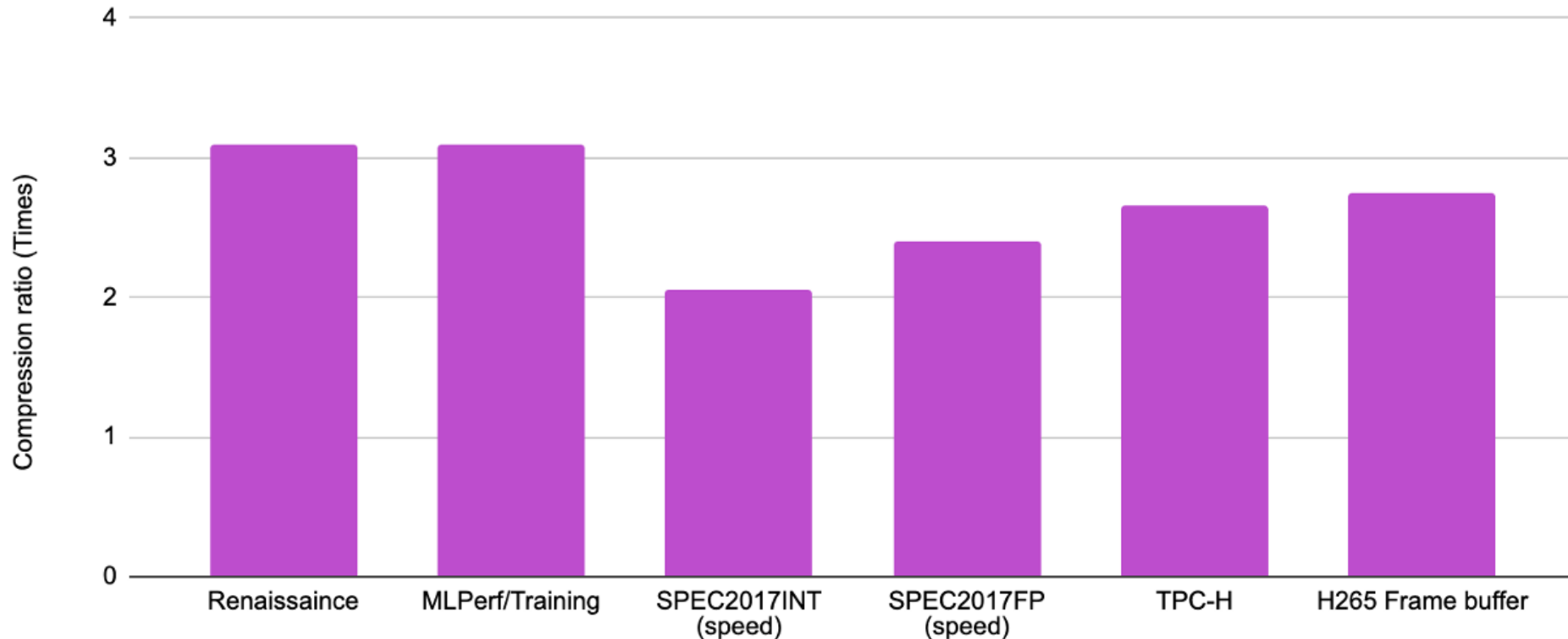
- Host triggers promotion/demotion operations by programming the DenseMem host-software
 - DenseMem APIs
- DenseMem host-software converts operations to CXL commands (VDM / CXL.io memory req)
- CXL device maps commands to DenseMem commands over interface (e.g., AXI4)
 - CMD (compress, decompress)
 - SRC addr
 - DST addr
- CXL.mem commands to transfer data to/from host



DenseMem™

KPIs

Compression performance on various server data-sets. Higher is better



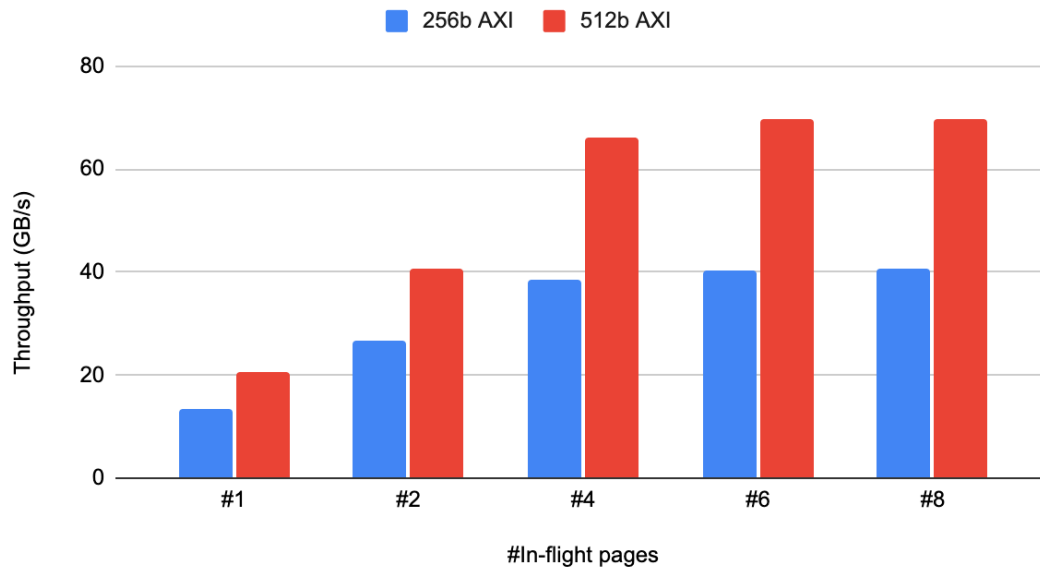
Robust compression across diverse data sets
Compression ratio: >2.5x on average

Page compression and decompression throughput

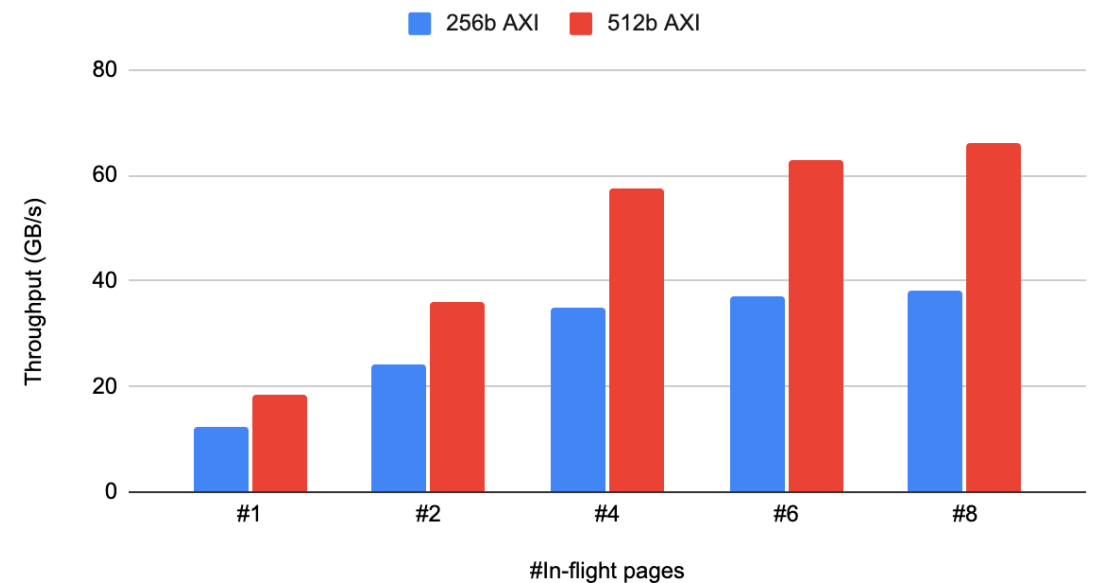
256b and 512b bus width

- #In-flight pages correspond to number of parallel operations
- DenseMem accelerator throughput (GB/s) – RTL implementation
 - IP clock frequency: 1.6GHz

Compression throughput (average)



Decompression throughput (average)



Summary

DenseMem™: The first real-time memory compression accelerator for CXL memory devices

Characteristics	Value
Clock frequency	1.6GHz (N5 technology node)
DDR memories supported	DDR4 and DDR5
Compression / Decompression granularity	4 KB block*
IP Area	0.2 – 0.3mm ² (TSMC N5)
Maximum number of parallel operations	4 (Can be upscaled/downscaled at design time)
Compression memory manager	Hardware-accelerated + software to control and perform maintenance
Support for adaptive compression based on target data (training-based compression)	Hardware-accelerated + software to control and tune
Compression algorithms	ZeroPoint proprietary

* With possibilities to support random CXL.mem 64-B reads w/o decompressing the full page

Visit our booth

