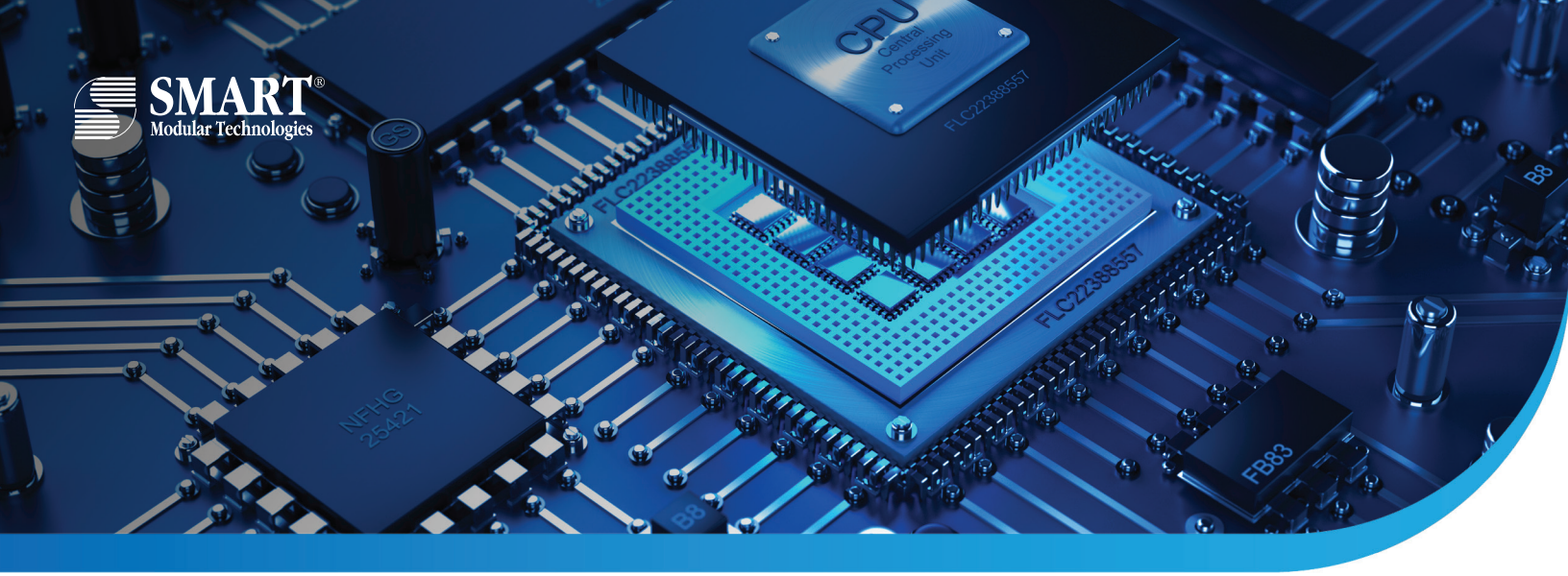


CCIX Memory Module (CMM)

Since the 1990s, DRAM Memory has been connected to CPUs using the parallel DDR bus defined by JEDEC now entering its fifth generation with DDR5. To increase DDR memory speeds, the number of available memory slots in servers has decreased but the number of CPU pins required has not changed. During this time the number of processor cores in CPUs has steadily increased. This has resulted in a performance bottleneck with per processor core memory bandwidth steadily decreasing.

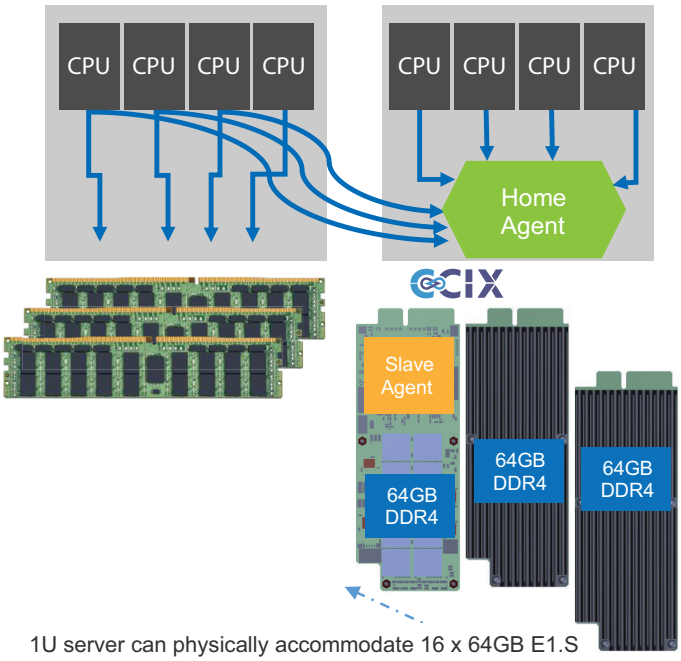
SMART Modular has designed a high-performance DRAM CMM (CCIX Memory Module) which solves the memory bandwidth problem. The CMM uses the CCIX high-speed, serial interface which provides high-speed memory access with a low pin count. The CMM includes the leading-edge Xilinx Versal™ ACAP FPGA to provide the host load/store, low-latency memory access. The on-board FPGA also supports CPU host offload or compute acceleration freeing the host CPU for use in other processing.

Key Features

- 64GB Cache-coherent DDR4 memory
- E1.S form-factor
- 8-lane CCIX front-end running over PCIe-Gen4 physical layer at 16GBps/lane
- On-board FPGA with dual-core ARM Cortex A72 processor for running mid to heavy workloads
- On-board quad-core ARM R5F processor for fixed function acceleration.
- Dedicated 128GB eMMC that can act as Storage backup or used for running application stack.
- Hardware engines to accelerate AI/ML workloads https://github.com/Xilinx/Vitis-AI

Benefits

- Low-latency memory expansion using serial interface frees up DDR slots
- Moving compute near the data. Offloads data-centric algorithms to FPGA and dual-core ARM A72
- Enables dynamic expansion of NUMA memory via E1.S hot-pluggable modules
- Utilizes application-specific acceleration libraries available from Xilinx https://xilinx.github.io/Vitis_Libraries/



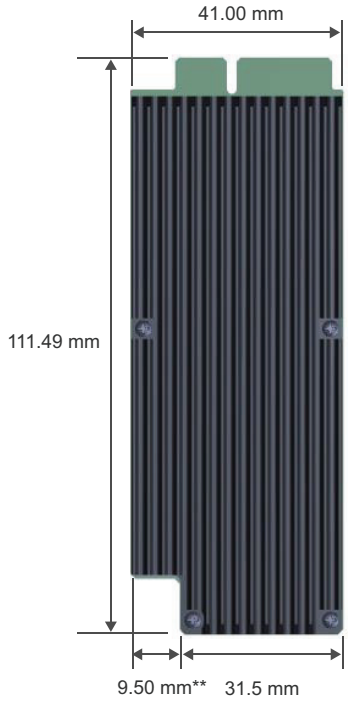
1U server can physically accommodate 16 x 64GB E1.S (CMM), provided there are sufficient CCIX lanes.

Use-cases

- AI/ML inference accelerators
- Offload and accelerate In-Memory database implementations to execute in parallel
- Storage cache for low cost HDD(s)
- Peer-to-Peer (P2P) data transfer enables implementing complex algorithms by pipelining each stage on individual memory modules

Physical Dimensions

Back View (with heatsink)



Front View

Side View

**For current version of E1.S prototypes, width is extended by 9.5mm to accommodate FPGA.
Prototype modules may require manual installation if heat-sink interferes with guide-rails in the E1.S bay.

Ordering Information

CCIX Memory Module (CMM)		
SMART Part Number	Density	Description
STCBS64GXB8D18MB	64GB	• CCIX Cache-coherent interconnect support for accelerators. • Dual-channel DDR4 • E1.S 2C (Gen4-x8) form-factor with 25mm heatsink

Contact: Arthur Sainio at Arthur.Sainio@smartm.com for any sales and sampling queries.

SMART® Modular Technologies

For more information, please visit: **www.smartm.com**

Headquarters/North America:
T: (+1) 800-956-7627 • T: (+1) 510-623-1231
F: (+1) 510-623-1434 • E: info@smartm.com

Latin America:
T: (+55) 11 4417-7200 • E: sales.br@smartm.com

Asia/Pacific:
T: (+65) 6678-7670 • E: sales.asia@smartm.com

EMEA:
T: (+44) 0 7826-064-745 • E: sales.euro@smartm.com

Customer Service:
T: (+1) 510-623-1231 • E: customers@smartm.com