

PUBLIC EXHIBIT 19

US Patent No. 11,841,803

DI Representative Article B

(AMD Instinct MI250 OAM Module Accelerator Circuit Board Assembly)

Claim 1(pre): "A system, comprising:"

1(pre). A system, comprising:

The AMD Instinct MI250 OAM Module GPU Accelerator (the DI Representative Article B) (hereinafter, the "AMD Product") includes a circuit board assembly and is designed for use in, is used in, and is a material part of the claimed system.

Click image to enlarge.



**AMD INSTINCT MI250 OAM MODULE GPU ACCELERATOR 128GB
13312 STREAM CORES MEMORY INTERFACE 8192 BIT HBM2E
MEMORY BANDWIDTH 3.2TB/S PCI-E 4.0 X16 GENERAL PURPOSE
GRAPHICS PROCESSING UNIT GPGPU - ALDEBARAN CDNA 2.0
ARCHITECTURE (AMD INSTINCT MI250 OAM)**

Part Number: AMD INSTINCT MI250 OAM

Condition: New Other



Availability: **ETA 3-4 days**

Qty in Stock: Request Pricing

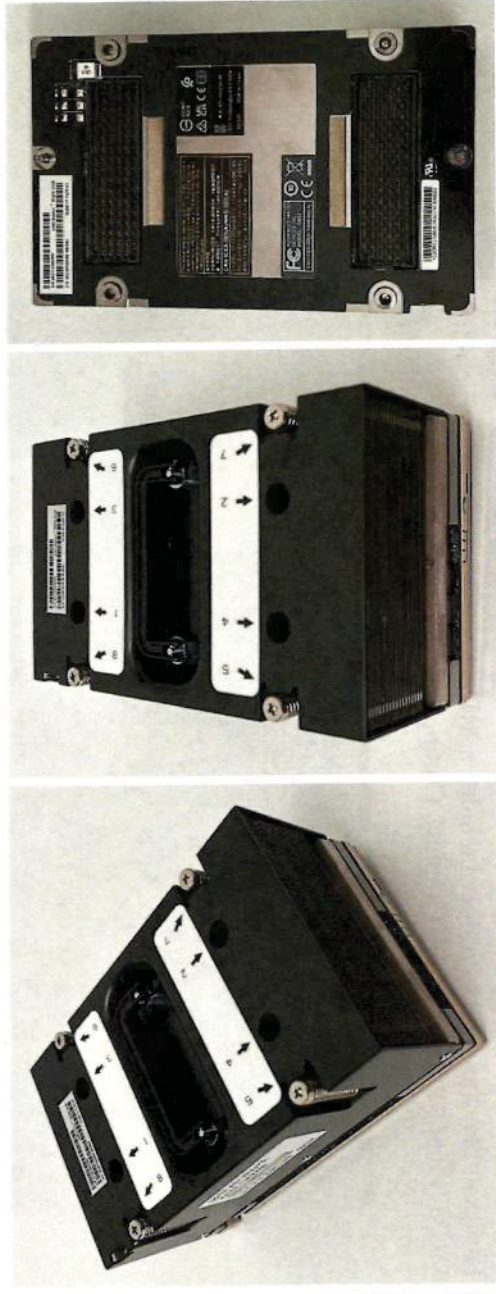
Product Details:

**AMD INSTINCT MI250 OAM MODULE GPU ACCELERATOR 128GB 13312 STREAM CORES MEMORY
INTERFACE 8192 BIT HBM2E MEMORY BANDWIDTH 3.2TB/S PCI-E 4.0 X16 GENERAL PURPOSE GRAPHICS
PROCESSING UNIT GPGPU - ALDEBARAN CDNA 2.0 ARCHITECTURE**

MANUFACTURER: AMD

PART NUMBER: **AMD INSTINCT MI250 OAM**

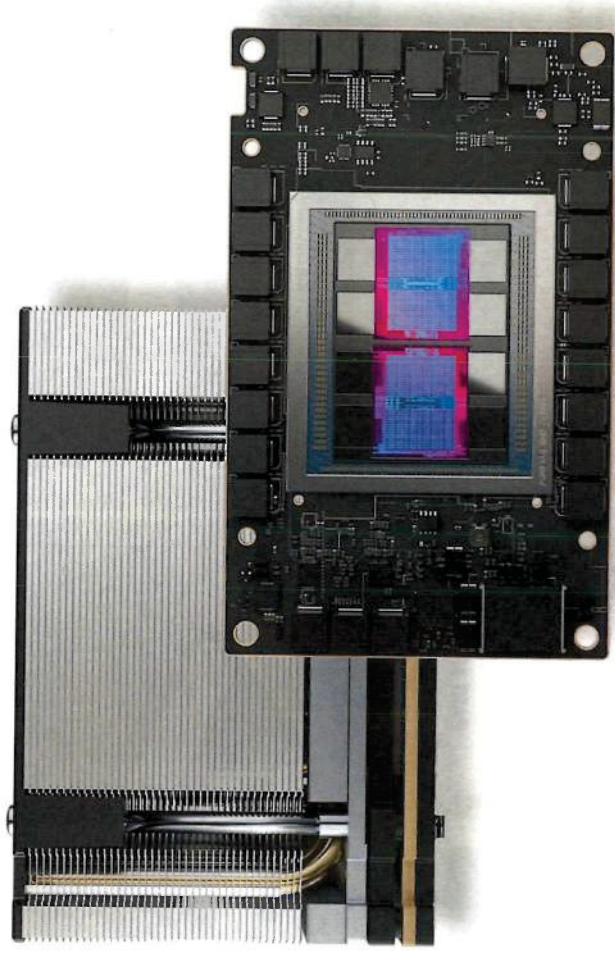
See <https://www.itcreations.com/product/150032>.



See Photographs for AMD MI250, IMG 5246.JPG, IMG 5245.JPG, IMG 5234.JPG.

Claim 1(pre): "A system, comprising:"

For example, the AMD Product includes a MI250 "multi-chip module" circuit board assembly ("MI250 MCM") that "comprises two Graphics Compute Dies (GCD)."



See AMD, AMD Instinct™ MI250
<https://www.amd.com/en/products/accelerators/instinct/mi200/mi250.html>.

Accelerators,

Claim 1(pre): "A system, comprising:"

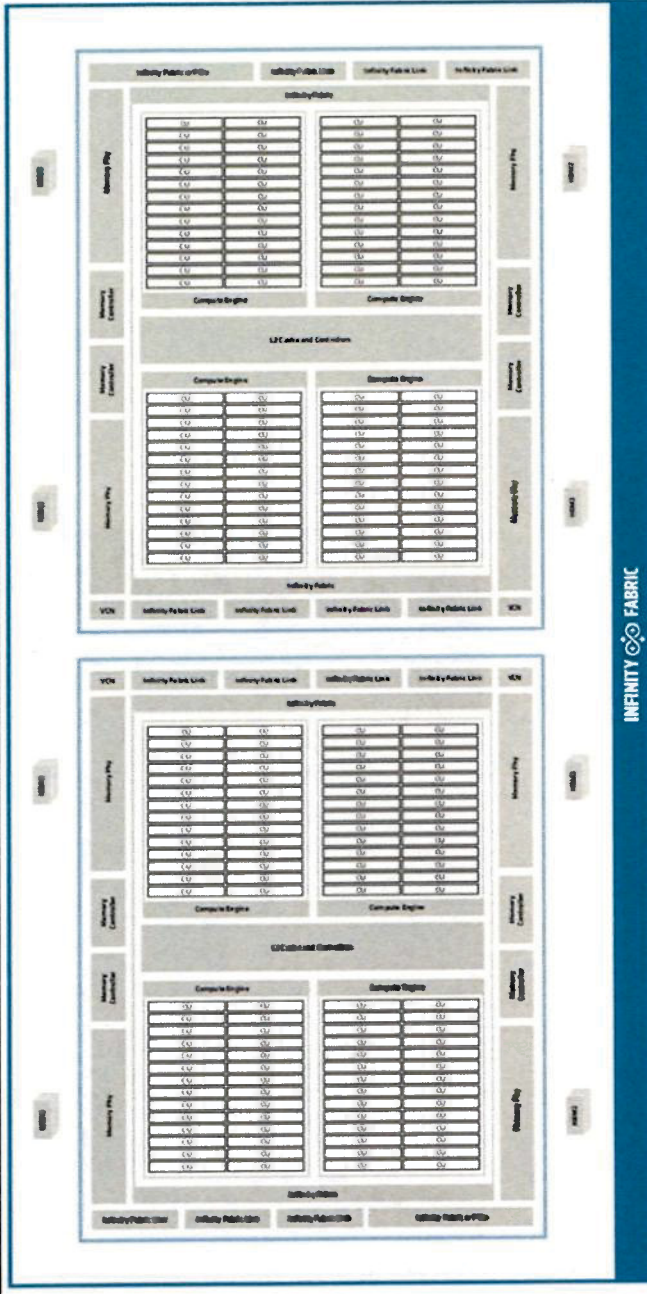
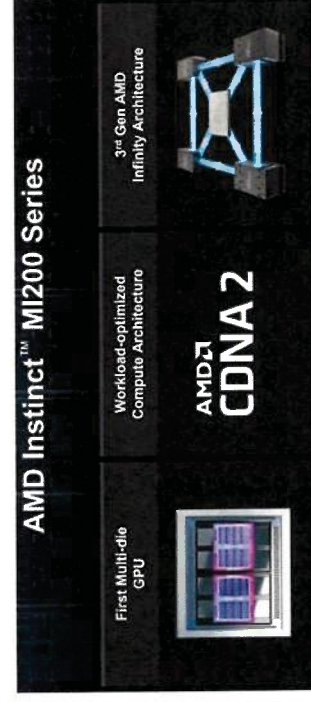


Figure 1b -Block diagram of the AMD Instinct™ MI200 multi-chip module (AMD Instinct™ MI250/MI250X) in the OAM form factor accelerator which comprise two Graphics Compute Dies (GCD) as illustrated.

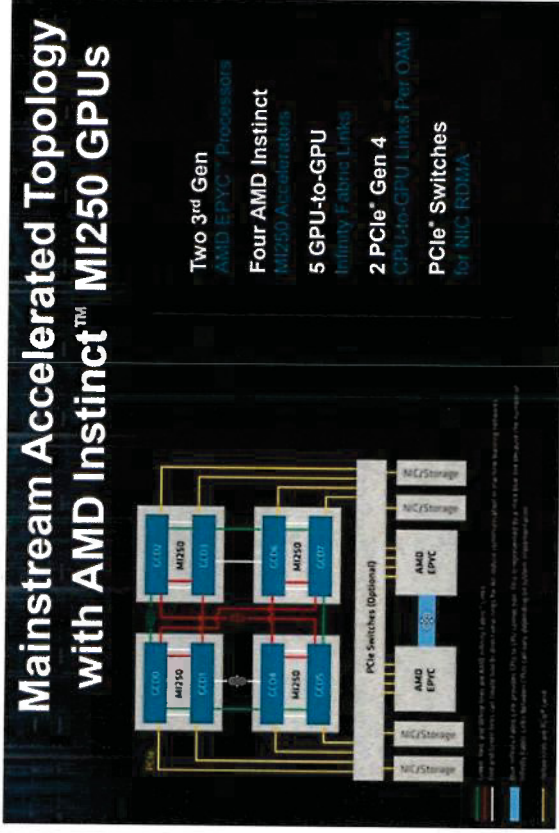
See AMD, Introducing AMD CDNA 2 Architecture, <https://www.amd.com/content/dam/amd/en/documents/instinct-business-docs/white-papers/amd-cdna2-white-paper.pdf>, p. 4.



See Hot Chips, AMD Instinct MI200 Series Accelerator and Node Architectures, <https://hc34.HotChips.org/assets/program/conference/day1/GPU%20HPC/HC2022.AMD.AlanSmith.v14.Final.20220820.pdf>, p. 3.

Claim 1(pre): “A system, comprising:”

Further, the AMD Product is designed for use in, is used in, and is a material part of a system with a CPU, and there is no substantial noninfringing use of the AMD Product other than used in a system with a CPU.



See also AMD, Introducing AMD CDNA 2 Architecture, <https://www.amd.com/content/dam/amd/en/documents/instinct-business-docs/white-papers/amd-cdna2-white-paper.pdf>, p. 7.



See HotChips, AMD Instinct MI200 Series Accelerator and Node Architectures, <https://hc34.hotchips.org/assets/program/conference/day1/GPU%20HPC/HC2022.AMD.AlanSmith.v13.Final.20220818.pdf>, p. 20.

Claim 1(a): “a central processing unit (CPU) communicably coupled to a first graphics processing unit (GPU) chiplet of a GPU chiplet array, wherein the GPU chiplet array includes: the first GPU chiplet communicably coupled to the CPU via a bus; and”

1(a). a central processing unit (CPU) communicably coupled to a first graphics processing unit (GPU) chiplet of a GPU chiplet array, wherein the GPU chiplet array includes: the first GPU chiplet communicably coupled to the CPU via a bus; and

The AMD Product is designed for use in, and is a material part of a system that includes a central processing unit (CPU) communicably coupled to a first graphics processing unit (GPU) chiplet of a GPU chiplet array, wherein the GPU chiplet array includes: the first GPU chiplet communicably coupled to the CPU via a bus.

For example, the MI250 MCM “comprise[s] two Graphics Compute Dies (GCDs)” that are “integrated in a single package.”

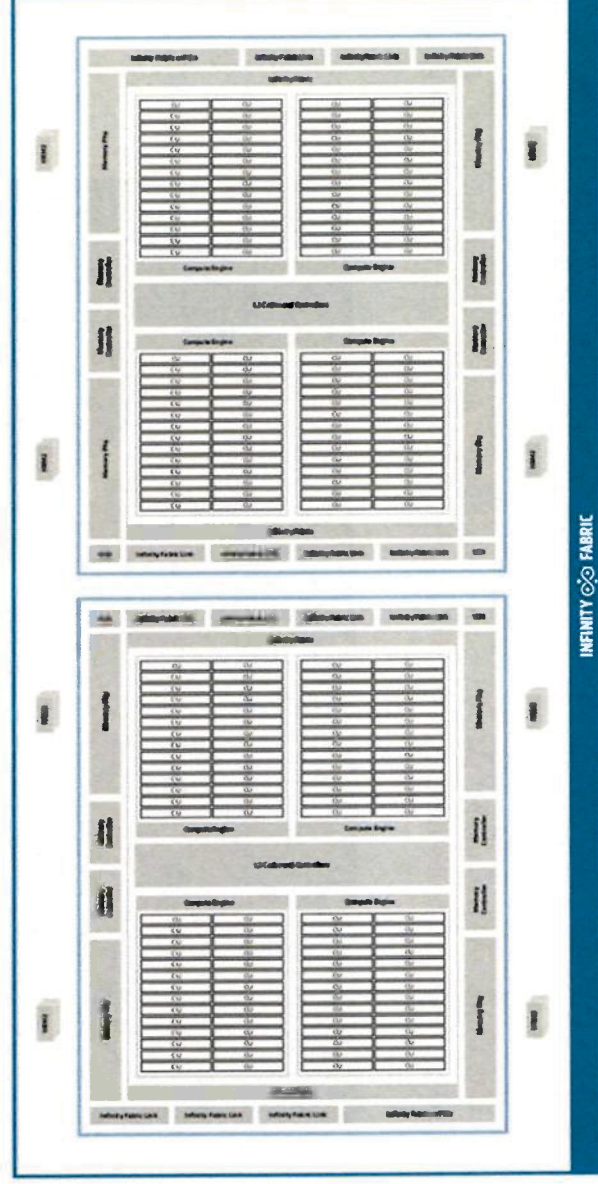


Figure 1b -Block diagram of the AMD Instinct™ MI200 multi-chip module (AMD Instinct™ MI250/MI250X) in the OAM form factor accelerator which comprise two Graphics Compute Dies (GCD) as illustrated.

The AMD CDNA 2 architecture is designed first and foremost for the most taxing scientific computing and machine learning applications. It powers the new AMD Instinct™ MI200 generation of products that target solutions ranging from compact single systems all the way to the world's largest exascale supercomputers with unique and highly differentiated programming models. Figure 1a illustrates the AMD Instinct MI200 Graphics Compute Die (GCD). The AMD Instinct™ MI200 is built on advanced packaging technologies, enabling two GCDs to be integrated into a single package in the OAM (OAM Accelerator Module) form factor in the MI250 and MI250X products, as shown in figure 1b. The AMD Instinct MI210 product enables one GCD in a traditional PCIe® form factor card for mainstream HPC and AI workloads in the data center. Each GCD is built on the AMD CDNA 2 architecture.

See AMD, Introducing AMD CDNA 2 Architecture, <https://www.amd.com/content/dam/amd/en/documents/instinct-business-docs/white-papers/amd-cdna2-white-paper.pdf>, p. 4.

Claim 1(a): “a central processing unit (CPU) communicably coupled to a first graphics processing unit (GPU) chiplet of a GPU chiplet array, wherein the GPU chiplet array includes: the first GPU chiplet communicably coupled to the CPU via a bus; and”

AMD Instinct™ MI200 Series

First Multi-die GPU

Workload-optimized Compute Architecture

3rd Gen AMD Infinity Architecture

AMD CDNA 2

See Hot Chips, AMD Instinct MI200 Series Accelerator and Node Architectures, <https://hc34.HotChips.org/assets/program/conference/day1/GPU%20HPC/HC2022.AMD.AlanSmith.v14.Final.20220820.pdf>, p. 3.

AMD Instinct™ MI200 Series

Key Innovations

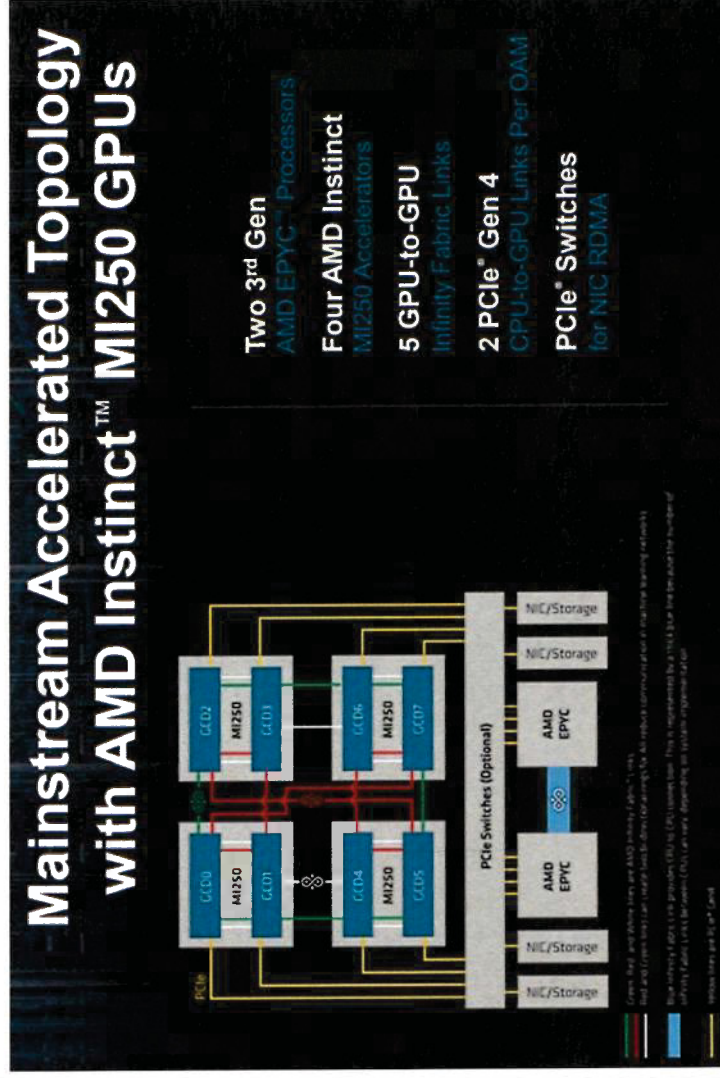
- Two AMD CDNA™2 Dies
- Ultra High Bandwidth Die Interconnect
- Coherent CPU-to-GPU Interconnect
- 2nd Gen Matrix Cores for HPC & AI
- Eight Stacks of HBM2E
- 2.5D Elevated Fanout Bridge (EFB)

AMD Instinct™ MI200 OAM Series

See Hot Chips, AMD Instinct MI200 Series Accelerator and Node Architectures, <https://hc34.HotChips.org/assets/program/conference/day1/GPU%20HPC/HC2022.AMD.AlanSmith.v14.Final.20220820.pdf>, p. 5.

Claim 1(a): “a central processing unit (CPU) communicably coupled to a first graphics processing unit (GPU) chiplet of a GPU chiplet array, wherein the GPU chiplet array includes: the first GPU chiplet communicably coupled to the CPU via a bus; and”

The AMD Product is designed for use in, is used in, and is a material part of a system that includes a CPU communicably coupled to a first processing unit (GPU) chiplet of a GPU chiplet array via a bus. For example, “[m]ainstream [a]ccelerated [t]opology with AMD Instruction MI250 GPUs” includes “2 PCIe Gen 4” “CPU-to-GPU Links Per OAM.”

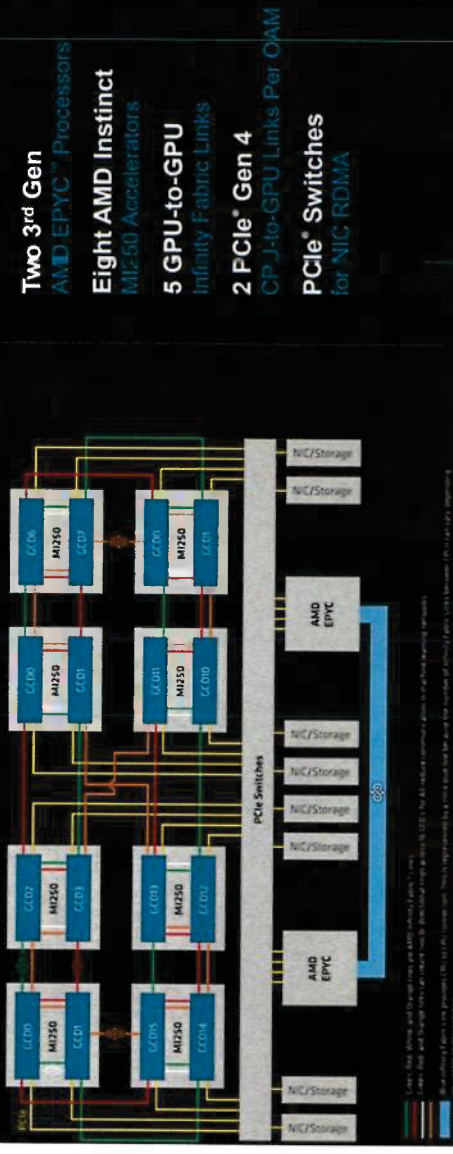


See Hot Chips, AMD Instinct MI200 Series Accelerator and Node Architectures, <https://hc34.HotChips.org/assets/program/conference/day1/GPU%20HPC/HC2022.AMD.AlanSmith.v14.Final.20220820.pdf>, p. 17; see also AMD, Introducing AMD CDNA 2 Architecture, <https://www.amd.com/content/dam/amd/en/documents/instinct-business-docs/white-papers/amd-cdna2-white-paper.pdf>, p. 7.

“Flagship [m]achine [l]earning [t]opology with AMD Instruction MI250 GPUs” includes “2 PCIe Gen 4” “CPU-to-GPU Links Per OAM.”

Claim 1(a): “a central processing unit (CPU) communicably coupled to a first graphics processing unit (GPU) chiplet of a GPU chiplet array, wherein the GPU chiplet array includes: the first GPU chiplet communicably coupled to the CPU via a bus; and”

Flagship Machine Learning Topology with AMD Instinct™ MI250 GPUs



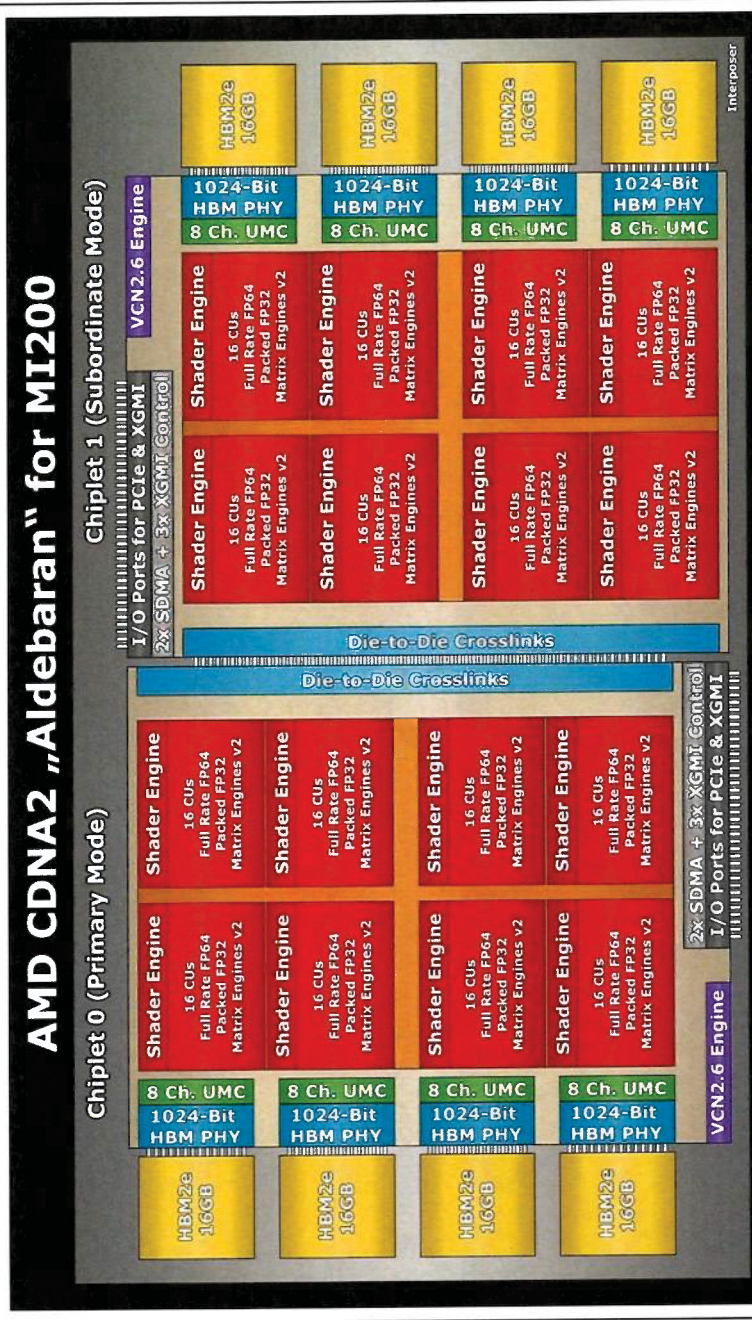
See Hot Chips, AMD Instinct MI200 Series Accelerator and Node Architectures, <https://hc34.HotChips.org/assets/program/conference/day1/GPU%20HPC/HC2022.AMD.AlanSmith.v14.Final.20220820.pdf>, p. 18; see also AMD, Introducing AMD CDNA 2 Architecture, <https://www.amd.com/content/dam/amd/en/documents/instinct-business-docs/white-papers/amd-cdna2-white-paper.pdf>, p. 8.

Claim 1(b): “a second GPU chiplet communicably coupled to the first GPU chiplet via a passive crosslink, wherein the passive crosslink is dedicated for inter-chiplet communications.”

1(b). a second GPU chiplet communicably coupled to the first GPU chiplet via a passive crosslink, wherein the passive crosslink is dedicated for inter-chiplet communications.

The AMD Product includes a second GPU chiplet communicably coupled to the first GPU chiplet via a passive crosslink, wherein the passive crosslink is dedicated for inter-chiplet communications.

For example, the AMD Product includes “Die-to-Die Crosslinks” coupling “Chiplet 0” and “Chiplet 1.”



See <https://www.notebookcheck.net/Leaked-diagram-reveals-the-AMD-Instinct-MI200-compute-GPU-with-dual-128-CU-dies-and-128-GB-HBM2e-RAM.549102.0.html>.

Further, for example, “Infinity Fabric Links” connect “GPU-to-GPU,” including “GCD0” and “GCD1,” “GCD2” and “GCD3,” “GCD4” and “GCD5,” and “GCD6” and “GCD7.”

Claim 1(b): “a second GPU chiplet communicably coupled to the first GPU chiplet via a passive crosslink, wherein the passive crosslink is dedicated for inter-chiplet communications.”

The IO die's Global Memory Interface (GMI) links are used by the Core Chiplet Dies (CCDs) of the Trento CPU to communicate and maintain cache coherency. The eXternal GMI² (or Inter-chip Global Memory Interconnect, xGMI) links are used for chip peer to peer communications (GPU/GPU, GPU/CPU CPU/GPU or CPU/CPU). The GMI and xGMI are the backbone of AMD's Infinity Fabric² (IF). A fabric can be seen as an abstraction of the underlying communication hardware. AMD's IF consists of two separate communication planes, the Infinity Scalable Data Fabric (SDF) and the Infinity Scalable Control Fabric (SCF). xGMI can be seen as a pumped up PCIe. It provides, at equivalent PCIe generation, slightly better transaction per second and (seemingly) better line code efficiency. xGM2 is ~ 18 GT/s, xGM13 is ~ 25 GT/s and xGMI 4 ~ 32 GT/s. In a similar vein, the PCIe Extended Speed Mode² (ESM) allows for higher throughput over similar PCIe 4 lane count and hardware due to higher PCIe clock. In fact, ESM is a side effect of the Cache Coherent Interconnect for Accelerators² (CCIX) protocol which is used by AMD and HPE-Cray on Adastr's nodes. The CCIX consortium standardized ESM to achieve up to 25 GT/s over traditional PCIe 4 (16 GT/s) lanes.

See INES, Adastr's Architecture, <https://dci.dci-gitlab.cines.fr/webextranet/architecture/index.html/>.