

AMD INSTINCT™ MI210 ACCELERATOR

AMD
INSTINCT

Bringing Exascale-Class Technologies to Mainstream HPC and AI

Exascale-class Technologies

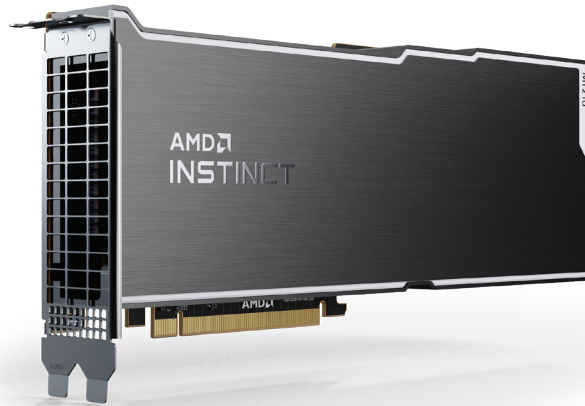
AMD Instinct™ MI210 leverages the same technology powering the **world's most powerful supercomputers**

Ultra-Fast HBM2e Memory

64GB High-bandwidth HBM2e memory with ECC Support at a bandwidth of **1.6 TB/s²**

416 Matrix Cores

enable **45.3 TFLOPs FP64** double precision matrix performance¹



181 TFLOPs

peak theoretical
FP16 and BF16 performance for AI training and machine learning

AMD Infinity Fabric™

technology enables up to **300 GB/s** (Dual) and **600 GB/s** (Quad) of total aggregate peak P2P theoretical I/O bandwidth in GPU hives³



AMD CDNA™ 2 Architecture

Powered by the 2nd Generation AMD CDNA™ architecture, the MI210 accelerator is purpose-built for HPC and AI, and delivers outstanding performance over competitive data center PCIe® GPUs today.¹



AMD Infinity Fabric™ Technology

3rd Gen AMD Infinity Fabric maximizes system efficiency throughout by enabling high speed direct GPU to GPU connectivity.



AMD ROCm™ Ecosystem

With the ROCm open software platform, HPC and AI developers can now access a range of open compute languages, compilers, libraries and tools for production-class deployments.

Visit [AMD.com/INSTINCT](https://www.amd.com/INSTINCT)

¹ Calculations conducted by AMD Performance Labs as of Jan 14, 2022, for the AMD Instinct™ MI210 (64GB HBM2e PCIe® card) accelerator at 1700 MHz peak boost engine clock resulted in 45.3 TFLOPs peak theoretical double precision (FP64 Matrix), 22.6 TFLOPs peak theoretical double precision (FP64), and 181.0 TFLOPs peak theoretical BF16 format precision (BF16), floating-point performance. Calculations conducted by AMD Performance Labs as of Sep 18, 2020 for the AMD Instinct™ MI100 (32GB HBM2 PCIe® card) accelerator at 1502 MHz peak boost engine clock resulted in 11.54 TFLOPs peak theoretical double precision (FP64), and 97 TFLOPs peak double precision (FP64) and 39 TFLOPs peak BF16 format precision (BF16), theoretical floating-point performance. <https://www.amd.com/content/dam/en-us/solutions/Data-Center/nvidia-ampere-architecture-whitepaper.pdf>, page 15, Table 1 MI200-41

² Calculations conducted by AMD Performance Labs as of Jan 27, 2022, for the AMD Instinct™ MI210 (64GB HBM2e) accelerator (PCIe®) designed with AMD CDNA™ 2 architecture 5nm FinFet process technology at 1,600 MHz peak memory clock resulted in 64 GB HBM2e memory capacity and 1.6384 TFLOPs peak theoretical memory bandwidth performance. MI210 memory bus interface is 4,096 bits and memory data rate is 3.20 Gbps for total memory bandwidth of 1.6384 TB/s (3.20 Gbps*(4,096 bits)/8). Calculations conducted by AMD Performance Labs as of Sep 18, 2020, for the AMD Instinct™ MI100 (32GB HBM2) accelerator (PCIe®) designed with AMD CDNA™ architecture 7nm FinFet process technology at 1,502 MHz peak clock resulted in 32 GB HBM2 memory capacity and 1.2288 TFLOPs peak theoretical memory bandwidth performance. MI210 memory bus interface is 4,096 bits and memory data rate is 2.40 Gbps for total memory bandwidth of 1.2288 TB/s (2.40 Gbps*(4,096 bits)/8). MI200-42

³ Calculations as of Jan 27th, 2022. AMD Instinct™ MI210 built on AMD CDNA™ 2 technology accelerators support PCIe® Gen4 providing up to 64 GB/s peak theoretical data bandwidth from CPU to GPU per card. AMD Instinct™ MI210 CDNA 2 technology-based accelerators include three Infinity Fabric™ links providing up to 300 GB/s peak theoretical GPU to GPU or Peer-to-Peer (P2P) bandwidth performance per GPU card. Combined with PCIe Gen4 support, this provides an aggregate GPU card I/O peak bandwidth of up to 364 GB/s. Dual-GPU hives: One dual-GPU hive provides up to 300 GB/s peak theoretical P2P performance. Four-GPU hives: One four-GPU hive provide up to 600 GB/s peak theoretical P2P performance. Dual four GPU hives in a server provide up to 1.2 TB/s total peak theoretical direct P2P performance per server. AMD Infinity Fabric link technology not enabled: One four-GPU hive provide up to 256 GB/s peak theoretical P2P performance with PCIe® 4.0. AMD Instinct™ MI100 built on AMD CDNA technology accelerators support PCIe® Gen4 providing up to 64 GB/s peak theoretical transport data bandwidth from CPU to GPU per card. AMD Instinct™ MI100 CDNA technology-based accelerators include three Infinity Fabric™ links providing up to 276 GB/s peak theoretical GPU to GPU or Peer-to-Peer (P2P) bandwidth performance per GPU card. Combined with PCIe Gen4 support, this provides an aggregate GPU card I/O peak bandwidth of up to 340 GB/s. One four-GPU hive provides up to 552 GB/s peak theoretical P2P performance. Dual four-GPU hives in a server provide up to 1.1 TB/s total peak theoretical direct P2P performance per server. AMD Infinity Fabric link technology not enabled: One four-GPU hive provides up to 256 GB/s peak theoretical P2P performance with PCIe® 4.0. Server manufacturers may vary configuration offerings yielding different results. MI200-43