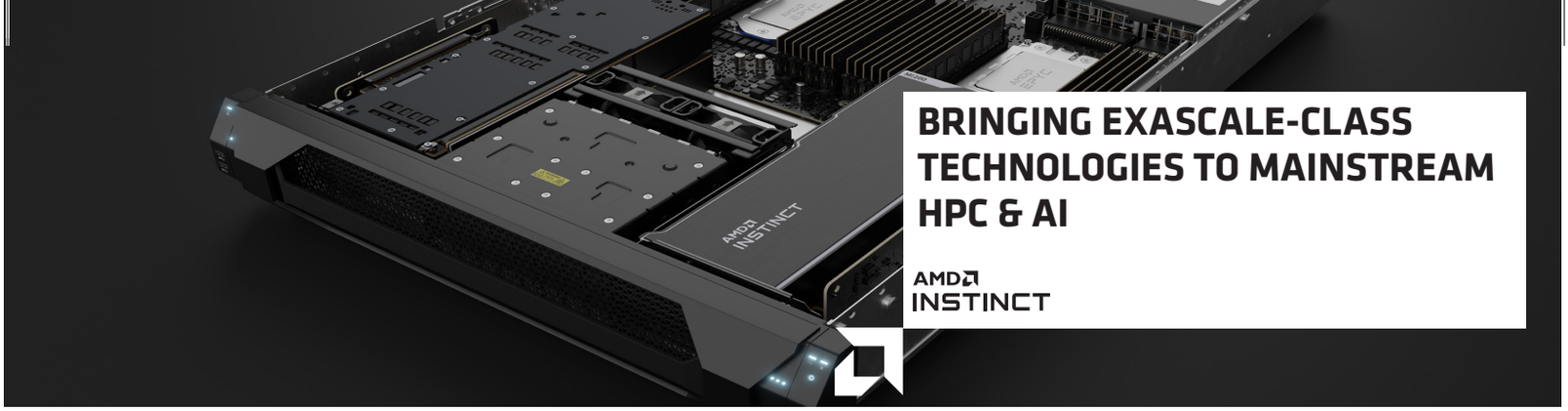


BRINGING EXASCALE-CLASS TECHNOLOGIES TO MAINSTREAM HPC & AI

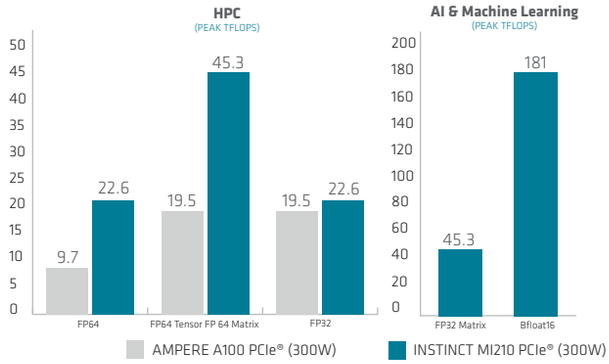
AMD INSTINCT

AMD INSTINCT™ MI210 ACCELERATOR

Exascale-Class Technologies for the Data Center

The AMD Instinct™ MI210 accelerator extends AMD industry performance leadership in accelerated compute for double precision (FP64) on PCIe® form factors for mainstream HPC and AI workloads in the data center.[1,2] Built on AMD Exascale-class technologies with the 2nd Gen AMD CDNA™ architecture, the MI210 enables scientists and researchers to tackle our most pressing challenges from climate change to vaccine research. MI210 accelerators, combined with the AMD ROCm™ 5 software ecosystem, allow innovators to tap the power of HPC and AI data center PCIe® GPUs to accelerate their time to science and discovery.

Superior Performance for HPC & AI



Graph 1: Peak TFLOPS across range of mixed-precision Compute[2]

Purpose-built Accelerators for HPC & AI Workloads

Powered by the 2nd Gen AMD CDNA™ architecture, AMD Instinct™ MI210 accelerator delivers HPC performance leadership over existing competitive PCIe® data center GPUs today with up to a 2.3x advantage over Nvidia Ampere A100 GPUs in FP64 performance delivering exceptional performance for a broad set of HPC & AI applications.[2] The MI210 accelerator is built to accelerate deep learning training, providing an expanded range of mixed-precision capabilities based on the AMD Matrix Core Technology, and delivers an outstanding 181 teraflops peak theoretical FP16 and BF16 performance to bring users a powerful platform to fuel the convergence of HPC and AI.

Innovations Delivering Performance Leadership

AMD innovations in architecture, packaging and integration are pushing the boundaries of computing by unifying the most important processors in the data center, the CPU, and the GPU accelerator. With our innovative double-precision Matrix Core capabilities along with the 3rd Gen AMD Infinity Architecture, AMD is delivering performance, efficiency and overall system throughput for HPC and AI using AMD EPYC™ CPUs and AMD Instinct™ MI210 accelerators.

Key Features

PERFORMANCE	MI210
Compute Units	104 CU
Stream Processors	6,656
Matrix Cores	416
Peak FP64/FP32 Vector	22.6 TF
Peak FP64/FP32 Matrix	45.3 TF
Peak FP16/BF16	181.0 TF
Peak INT4/INT8	181.0 TOPS

MEMORY	
Memory Size	64GB HBM2e
Memory Interface	4,096 bits
Memory Clock	1.6GHz
Memory Bandwidth	up to 1.6 TB/sec[3]

RELIABILITY	MI210
ECC (Full-chip)	Yes
RAS Support	Yes

SCALABILITY	
Infinity Fabric™ Links	up to 3
Coherency Enabled	Yes (Dual | Quad Hives)
OS Support	Linux™ 64 Bit
AMD ROCm™ Compatible	Yes

BOARD DESIGN	
Board Form Factor	Full-Height, Full-Length (Dual Slot)
Length	4.5" x 10.5" (11.43 CM x 26.67 CM)
Bus Interface	PCIe® Gen4 | Gen3 Support
SR-IOV Support	Yes (Passthrough Only)
Thermal	Passively Cooled
Max Power	300W TDP (EPS12V, 8-pin)
Warranty	Three Year Limited[5]

Ecosystem without Borders

AMD ROCm™ is an open software platform allowing researchers to tap the power of AMD Instinct™ accelerators to drive scientific discoveries. The ROCm platform is built on the foundation of open portability, supporting environments across multiple accelerator vendors and architectures. With ROCm 5, AMD extends its platform powering top HPC and AI applications with AMD Instinct™ MI200 series accelerators, increasing accessibility of ROCm for developers and delivering outstanding performance across key workloads.

HPC and MACHINE LEARNING APPLICATIONS

HPC | Life Sciences | Chemistry | Energy | Weather | Astrophysics | Automotive | Reinforcement Learning | Image | Object | Video Detection & Classification

OPEN PROGRAMING WITH CHOICE

OpenMP | HIP | OpenCL™ | Python

OPEN FRAMEWORKS

PyTorch | TensorFlow | ONNX | Kokkos | RAJA

OPTIMIZED LIBRARIES

BLAS | FFT | RNG | SPARSE | THRUST | MIOpen | RCCL

PROGRAMER AND SYSTEM TOOLS

Debuggers | Performance Analysis | System Management

2nd Generation AMD CDNA™ Architecture

The AMD Instinct™ MI210 accelerator brings commercial HPC & AI customers the compute engine selected for the first U.S. Exascale supercomputer. Powered by the 2nd Gen AMD CDNA™ architecture, the MI210 accelerator delivers outstanding performance for HPC and AI. The MI210 PCIe® GPU delivers superior double and single precision performance compared to the Nvidia Ampere A100 GPU for HPC workloads with up to 22.6 TFLOPS peak FP64|FP32 performance, enabling scientists and researchers around the globe to process HPC parallel codes more efficiently across several industries.[1]

AMD's Matrix Core technology delivers a broad range of mixed precision operations bringing you the ability to work with large models and enhance memory-bound operation performance for whatever combination of AI and machine learning workloads you need to deploy. The MI210 offers optimized BF16, INT4, INT8, FP16, FP32, and FP32 Matrix capabilities bringing you supercharged compute performance to meet all your AI system requirements. The AMD Instinct MI210 accelerator handles large data efficiently for training and delivers 181 teraflops of peak FP16 and bfloat16 floating-point performance for deep learning training.

AMD Infinity Fabric™ Link Technology

AMD Instinct MI210 GPUs provide advanced I/O capabilities in standard off-the-shelf servers with our AMD Infinity Fabric™ technology and PCIe® Gen4 support. The MI210 GPU delivers 64 GB/s CPU to GPU bandwidth without the need for PCIe® switches, and up to 300 GB/s of Peer-to-Peer (P2P) bandwidth performance through three Infinity Fabric links.[4] The AMD Infinity Architecture enables platform designs with dual and quad, direct-connect, GPU hives with high-speed P2P connectivity and delivers up to 1.2 TB/s of total theoretical GPU bandwidth within a server design.[4] Infinity Fabric helps unlock the promise of accelerated computing, enabling a quick and simple on-ramp for CPU codes to accelerated platforms.

Ultra-Fast HBM2e Memory

AMD Instinct MI210 accelerators provide up to 64GB High-bandwidth HBM2e memory with ECC support at a clock rate of 1.6 GHz. and deliver an ultra-high 1.6 TB/s of memory bandwidth to help support your largest data sets and eliminate bottlenecks moving data in and out of memory.[3] Combine this performance with the MI210's advanced Infinity Fabric I/O capabilities and you can push workloads closer to their full potential.