November 8, 2021

# New AMD Instinct™ MI200 Series Accelerators Bring Leadership HPC and AI Performance to Power Exascale Systems and More

*— With new AMD CDNA™ 2 architecture, AMD Instinct MI200 series accelerators deliver ground-breaking 4.9x advantage in HPC performance[1] compared to competing data center accelerators, expediting science and discovery —*

*— MI200 series accelerators are first multi-die GPU, first to support 128GB of HBM2e memory, and deliver a substantial boost for applications critical to the foundation of science —*

SANTA CLARA, Calif., Nov. 08, 2021 (GLOBE NEWSWIRE) -- AMD (NASDAQ: AMD) today announced the new AMD Instinct™ MI200 series accelerators, the first exascale-class GPU accelerators. AMD Instinct MI200 series accelerators includes the world's fastest high performance computing (HPC) and artificial intelligence (AI) accelerator,[1] the AMD Instinct™ MI250X.

Built on AMD CDNA™ 2 architecture, AMD Instinct MI200 series accelerators deliver leading application performance for a broad set of HPC workloads.[2] The AMD Instinct MI250X accelerator provides up to 4.9X the performance of competitive accelerators for double precision (FP64) HPC applications and surpasses 380 teraflops of peak theoretical half-precision (FP16) for AI workloads to enable disruptive approaches in further accelerating data-driven research.[1]

"AMD Instinct MI200 accelerators deliver leadership HPC and AI performance, helping scientists make generational leaps in research that can dramatically shorten the time between initial hypothesis and discovery," said Forrest Norrod, senior vice president and general manager, Data Center and Embedded Solutions Business Group, AMD. "With key innovations in architecture, packaging and system design, the AMD Instinct MI200 series accelerators are the most advanced data center GPUs ever, providing exceptional performance for supercomputers and data centers to solve the world's most complex problems."

**Exascale With AMD**

AMD, in collaboration with the U.S. Department of Energy, Oak Ridge National Laboratory, and HPE, designed the Frontier supercomputer expected to deliver more than 1.5 exaflops of peak computing power. Powered by optimized 3rd Gen AMD EPYC™ CPUs and AMD Instinct MI250X accelerators, Frontier will push the boundaries of scientific discovery by dramatically enhancing performance of AI, analytics, and simulation at scale, helping

scientists to pack in more calculations, identify new patterns in data, and develop innovative data analysis methods to accelerate the pace of scientific discovery.

“The Frontier supercomputer is the culmination of a strong collaboration between AMD, HPE and the U.S. Department of Energy, to provide an exascale-capable system that pushes the boundaries of scientific discovery by dramatically enhancing performance of artificial intelligence, analytics, and simulation at scale,” said Thomas Zacharia, director, Oak Ridge National Laboratory.

**Powering The Future of HPC**

The AMD Instinct MI200 series accelerators, combined with 3rd Gen AMD EPYC CPUs and the ROCm™ 5.0 open software platform, are designed to propel new discoveries for the exascale era and tackle our most pressing challenges from climate change to vaccine research.

Key capabilities and features of the AMD Instinct MI200 series accelerators include:

- **AMD CDNA™ 2 architecture** – 2nd Gen Matrix Cores accelerating FP64 and FP32 matrix operations, delivering up to 4X the peak theoretical FP64 performance vs. AMD previous gen GPUs. [1,3,4]
- **Leadership Packaging Technology** – Industry-first multi-die GPU design with 2.5D Elevated Fanout Bridge (EFB) technology delivers 1.8X more cores and 2.7X higher memory bandwidth vs. AMD previous gen GPUs, offering the industry’s best aggregate peak theoretical memory bandwidth at 3.2 terabytes per second. [4,5,6]
- **3rd Gen AMD Infinity Fabric™ technology –** Up to 8 Infinity Fabric links connect the AMD Instinct MI200 with 3rd Gen EPYC CPUs and other GPUs in the node to enable unified CPU/GPU memory coherency and maximize system throughput, allowing for an easier on-ramp for CPU codes to tap the power of accelerators.

**Software for Enabling Exascale Science**

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 5.0, AMD extends its open platform powering top HPC and AI applications with AMD Instinct MI200 series accelerators, increasing accessibility of ROCm for developers and delivering leadership performance across key workloads.

Through the AMD Infinity Hub, researchers, data scientists and end-users can easily find, download and install containerized HPC apps and ML frameworks that are optimized and supported on AMD Instinct accelerators and ROCm. The hub currently offers a range of containers supporting either Radeon Instinct™ MI50, AMD Instinct™ MI100 or AMD Instinct MI200 accelerators including several applications like Chroma, CP2k, LAMMPS, NAMD, OpenMM and more, along with popular ML frameworks TensorFlow and PyTorch. New containers are continually being added to the hub.

**Available Server Solutions**

The AMD Instinct MI250X and AMD Instinct MI250 are available in the open-hardware compute accelerator module or OCP Accelerator Module (OAM) form factor. The AMD Instinct MI210 will be available in a PCIe® card form factor in OEM servers.

The AMD MI250X accelerator is currently available from HPE in the HPE Cray EX Supercomputer, and additional AMD Instinct MI200 series accelerators are expected in systems from major OEM and ODM partners in enterprise markets in Q1 2022, including ASUS, ATOS, Dell Technologies, Gigabyte, Hewlett Packard Enterprise (HPE), Lenovo, Penguin Computingand Supermicro.

### MI200 Series Specifications

| Models | Compute Units | Stream Processors | FP64 \| FP32 Vector (Peak) | FP64 \| FP32 Matrix (Peak) | FP16 \| bf16 (Peak) | INT4 \| INT8 (Peak) | HBM2e ECC Memory | Memory Bandwidth | Form Factor |
|---|---|---|---|---|---|---|---|---|---|
| AMD Instinct MI250x | 220 | 14,080 | Up to 47.9 TF | Up to 95.7 TF | Up to 383.0 TF | Up to 383.0 TOPS | 128GB | 3.2 TB/sec | OCP Accelerator Module |
| AMD Instinct MI250 | 208 | 13,312 | Up to 45.3 TF | Up to 90.5 TF | Up to 362.1 TF | Up to 362.1 TOPS | 128GB | 3.2 TB/sec | OCP Accelerator Module |

### Supporting Resources

- Learn more about AMD Instinct™ Accelerators
- Learn more about AMD HPC Solutions
- Learn more about AMD CDNA™ 2 architecture
- Learn more about the AMD EPYC™ processors
- Become a fan of AMD on Facebook
- Follow AMD on Twitter
- Connect with AMD On LinkedIn

### About AMD

For more than 50 years AMD has driven innovation in high-performance computing, graphics and visualization technologies — the building blocks for gaming, immersive platforms and the data center. Hundreds of millions of consumers, leading Fortune 500 businesses and cutting-edge scientific research facilities around the world rely on AMD technology daily to improve how they live, work and play. AMD employees around the world are focused on building great products that push the boundaries of what is possible. For more information about how AMD is enabling today and inspiring tomorrow, visit the AMD (NASDAQ: AMD) website, Facebook, LinkedIn and Twitter pages.

### CAUTIONARY STATEMENT

This press release contains forward-looking statements concerning Advanced Micro Devices, Inc. (AMD) such as the features, functionality, performance, availability, timing and expected benefits of AMD products including the AMD Instinct™ MI200 series accelerators, which are made pursuant to the Safe Harbor provisions of the Private Securities Litigation Reform Act of 1995. Forward-looking statements are commonly identified by words such as "would," "may," "expects," "believes," "plans," "intends," "projects" and other terms with similar meaning. Investors are cautioned that the forward-looking statements in this press release are based on current beliefs, assumptions and expectations, speak only as of the date of this press release and involve risks and uncertainties that could cause actual results to differ materially from current expectations. Such statements are subject to certain known and unknown risks and uncertainties, many of which are difficult to predict and generally beyond AMD's control, that could cause actual results and other future events to differ materially from those expressed in, or implied or projected by, the forward-looking information and statements. Material factors that could cause actual results to differ materially from current expectations include, without limitation, the following: Intel Corporation’s dominance of the microprocessor market and its aggressive business

practices; global economic uncertainty; loss of a significant customer; impact of the COVID-19 pandemic on AMD's business, financial condition and results of operations; competitive markets in which AMD's products are sold; market conditions of the industries in which AMD products are sold; cyclical nature of the semiconductor industry; quarterly and seasonal sales patterns; AMD's ability to adequately protect its technology or other intellectual property; unfavorable currency exchange rate fluctuations; ability of third party manufacturers to manufacture AMD's products on a timely basis in sufficient quantities and using competitive technologies; availability of essential equipment, materials, substrates or manufacturing processes; ability to achieve expected manufacturing yields for AMD's products; AMD's ability to introduce products on a timely basis with expected features and performance levels; AMD's ability to generate revenue from its semi-custom SoC products; potential security vulnerabilities; potential security incidents including IT outages, data loss, data breaches and cyber-attacks; uncertainties involving the ordering and shipment of AMD's products; AMD's reliance on third-party intellectual property to design and introduce new products in a timely manner; AMD's reliance on third-party companies for design, manufacture and supply of motherboards, software and other computer platform components; AMD's reliance on Microsoft and other software vendors' support to design and develop software to run on AMD's products; AMD's reliance on third-party distributors and add-in-board partners; impact of modification or interruption of AMD's internal business processes and information systems; compatibility of AMD's products with some or all industry-standard software and hardware; costs related to defective products; efficiency of AMD's supply chain; AMD's ability to rely on third party supply-chain logistics functions; AMD's ability to effectively control sales of its products on the gray market; impact of government actions and regulations such as export administration regulations, tariffs and trade protection measures; AMD's ability to realize its deferred tax assets; potential tax liabilities; current and future claims and litigation; impact of environmental laws, conflict minerals-related provisions and other laws or regulations; impact of acquisitions, joint ventures and/or investments on AMD's business, including the announced acquisition of Xilinx, and ability to integrate acquired businesses; AMD's ability to complete the Xilinx merger; impact of the announcement and pendency of the Xilinx merger on AMD's business; impact of any impairment of the combined company's assets on the combined company's financial position and results of operation; restrictions imposed by agreements governing AMD's notes and the revolving credit facility; AMD's indebtedness; AMD's ability to generate sufficient cash to meet its working capital requirements or generate sufficient revenue and operating cash flow to make all of its planned R&D or strategic investments; political, legal, economic risks and natural disasters; future impairments of goodwill and technology license purchases; AMD's ability to attract and retain qualified personnel; AMD's stock price volatility; and worldwide political conditions. Investors are urged to review in detail the risks and uncertainties in AMD's Securities and Exchange Commission filings, including but not limited to AMD's most recent reports on Forms 10-K and 10-Q.



Additional benchmark data is available on AMD.com

1. World's fastest data center GPU is the AMD Instinct™ MI250X. Calculations conducted by AMD Performance Labs as of Sep 15, 2021, for the AMD Instinct™ MI250X (128GB

HBM2e OAM module) accelerator at 1,700 MHz peak boost engine clock resulted in 95.7 TFLOPS peak theoretical double precision (FP64 Matrix), 47.9 TFLOPS peak theoretical double precision (FP64), 95.7 TFLOPS peak theoretical single precision matrix (FP32 Matrix), 47.9 TFLOPS peak theoretical single precision (FP32), 383.0 TFLOPS peak theoretical half precision (FP16), and 383.0 TFLOPS peak theoretical Bfloat16 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 1,502 MHz peak boost engine clock resulted in 11.54 TFLOPS peak theoretical double precision (FP64), 46.1 TFLOPS peak theoretical single precision matrix (FP32), 23.1 TFLOPS peak theoretical single precision (FP32), 184.6 TFLOPS peak theoretical half precision (FP16) floating-point performance. Published results on the NVidia Ampere A100 (80GB) GPU accelerator, boost engine clock of 1410 MHz, resulted in 19.5 TFLOPS peak double precision tensor cores (FP64 Tensor Core), 9.7 TFLOPS peak double precision (FP64). 19.5 TFLOPS peak single precision (FP32), 78 TFLOPS peak half precision (FP16), 312 TFLOPS peak half precision (FP16 Tensor Flow), 39 TFLOPS peak Bfloat 16 (BF16), 312 TFLOPS peak Bfloat16 format precision (BF16 Tensor Flow), theoretical floating-point performance. The TF32 data format is not IEEE compliant and not included in this comparison. https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/nvidia-ampere-architecture-whitepaper.pdf, page 15, Table 1. MI200-01
2. AMD Instinct MI250X accelerator application and benchmark performance can be found at https://www.amd.com/en/graphics/server-accelerators-benchmarks.
3. Calculations conducted by AMD Performance Labs as of Sep 15, 2021, for the AMD Instinct™ MI250X accelerator (128GB HBM2e OAM module) at 1,700 MHz peak boost engine clock resulted in 95.7 TFLOPS peak double precision matrix (FP64 Matrix) theoretical, floating-point performance. Published results on the NVidia Ampere A100 (80GB) GPU accelerator resulted in 19.5 TFLOPS peak double precision (FP64 Tensor Core) theoretical, floating-point performance. Results found at:https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/nvidia-ampere-architecture-whitepaper.pdf, page 15, Table 1.MI200-02
4. Calculations conducted by AMD Performance Labs as of Sep 21, 2021, for the AMD Instinct™ MI250X and MI250 (128GB HBM2e) OAM accelerators designed with AMD CDNA™ 2 6nm FinFet process technology at 1,600 MHz peak memory clock resulted in 128GB HBM2e memory capacity and 3.2768 TFLOPS peak theoretical memory bandwidth performance. MI250/MI250X memory bus interface is 4,096 bits times 2 die and memory data rate is 3.20 Gbps for total memory bandwidth of 3.2768 TB/s ((3.20 Gbps*(4,096 bits*2))/8).The highest published results on the NVidia Ampere A100 (80GB) SXM GPU accelerator resulted in 80GB HBM2e memory capacity and 2.039 TB/s GPU memory bandwidth performance.https://www.nvidia.com/content/dam/en-zz/Solutions/Data-Center/a100/pdf/nvidia-a100-datasheet-us-nvidia-1758950-r4-web.pdf MI200-07
5. The AMD Instinct™ MI250X accelerator has 220 compute units (CUs) and 14,080 stream cores. The AMD Instinct™ MI100 accelerator has 120 compute units (CUs) and 7,680 stream cores. MI200-027
6. Calculations conducted by AMD Performance Labs as of Sep 21, 2021, for the AMD Instinct™ MI250X and MI250 (128GB HBM2e) OAM accelerators designed with AMD CDNA™ 2 6nm FinFet process technology at 1,600 MHz peak memory clock resulted in 3.2768 TFLOPS peak theoretical memory bandwidth performance. MI250/MI250X memory bus interface is 4,096 bits times 2 die and memory data rate is 3.20 Gbps for total memory bandwidth of 3.2768 TB/s ((3.20 Gbps*(4,096 bits*2))/8). Calculations by AMD Performance Labs as of OCT 5th, 2020 for the AMD Instinct™ MI100 accelerator designed with AMD CDNA 7nm FinFET process technology at 1,200 MHz peak

memory clock resulted in 1.2288 TFLOPS peak theoretical memory bandwidth performance. MI100 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) MI200-33

Contacts:
Aaron Grabein
AMD Communications
+1 512-602-8950
Aaron.Grabein@amd.com

Laura Graves
AMD Investor Relations
+1 408-749-5467
Laura.Graves@amd.com

Source: Advanced Micro Devices, Inc.