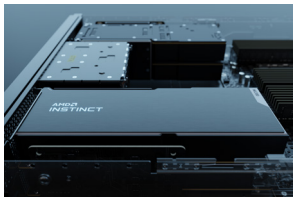
AMD INSTINCT™ GPU FAMILY

GPU ACCELERATORS IN EVERY FORM FACTOR TO PROPEL GENERATIVE AI, TRAINING, INFERENCE, AND HIGH-PERFORMANCE COMPUTING

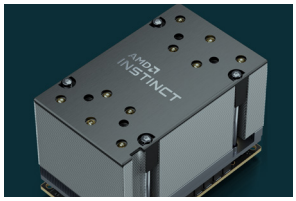
A GPU FAMILY FOR EVERY PURPOSE

Across three generations of AMD CDNA™ architecture, we have delivered amazing performance and energy efficiency for AI training, inference, and high-performance computing, with capabilities and form factors designed to support you wherever you need accelerated computing.



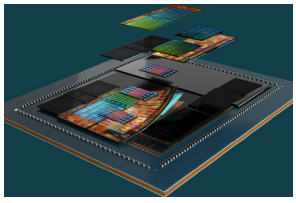
PCIE ACCELERATORS

AMD Instinct™ MI210 and MI250 accelerators are built in a standard full-height, full-length PCIe form factor that can propel workstations that data scientists use in workstations as well as inference servers in the core data center or in edge locations.



OAM ACCELERATORS

The OAM form-factor GPUs are the next step up in performance. Built as standard OAM modules, they are designed with coherent memory across a pod of up to four modules, and up to eight modules when deployed on a universal baseboard.



ACCELERATED PROCESSING UNITS

The Instinct MI300A accelerated processing unit (APU) is built from the ground up to overcome the challenges that discrete GPUs present: performance bottlenecks from the narrow interfaces between CPU and GPU. Designed to accelerate high-performance computing, the AMD Instinct MI300A integrates 'Zen 4' x86 CPU cores with high-throughput GPU compute units, with a single, shared address space between CPU and GPU. Slated for next-generation supercomputers, this technology is available through platforms offered by our solution partners.

INDUSTRY-STANDARD BASEBOARD FOR PURPOSE-BUILT SERVERS

For your most intensive datacenter AI and HPC needs, our OAM-form-factor accelerators are populated eight at a time onto an industry-standard universal baseboard (UBB) to fit into purpose-built servers. Both the MI350 and MI300 Series are enhanced with AI-specific functions including new data-type support, and photo and video decoding.

AMD ROCM ECOSYSTEM WITHOUT BORDERS

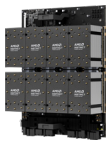
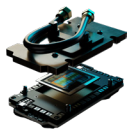
When you choose a variety of products from the AMD Instinct accelerator family, you need optimized performance from each device. That's not a worry with the AMD ROCm™ platform, which opens doors to new levels of freedom and accessibility. With proven scalability, ROCm software provides support for leading programing languages and frameworks for HPC and AI. With mature drivers, compilers and optimized libraries supporting AMD Instinct accelerators, ROCm provides an open environment that is ready to deploy when you are.

AMD INSTINCT








AI PEAK THEORETICAL PERFORMANCE	MI210	MI250	MI250x	MI300A		MI300X		MI300X Platform		MI325X		MI325X Platform		MI350X		MI350X Platform		MI355X		MI 355X Platform	
	tera-	tera-	tera-	tera-	sparsity peta-	tera-	sparsity peta-	tera-	peta-	tera-	sparsity tera-	peta-	sparsity peta-	tera-	sparsity	peta-	sparsity	tera-	sparsity	peta-	sparsity
FP16 (FLOPS)	181.0	362.1	383.0	980.6	1961.2	1307.4	20.9	10.5	20.9	1307.4	2614.9	10.5	20.9	2306.9	4613.8	18.4	36.8	2516.6	5033.2	20.1	40.3
BFLOAT16 (FLOPS)	181.0	362.1	383.0	980.6	1961.2	1307.4	20.9	10.5	20.9	1307.4	2614.9	10.5	20.9	2309.6	4619.2	18.5	36.9	2516.6	5033.2	20.1	40.3
INT8 (OPS)	181.0	362.1	383.0	1961.2	3922.3	2614.9	41.8	20.9	41.8	2614.9	5229.8	20.9	41.8	4613.7	9227.4	36.9	73.8	5033.2	10,066.4	40.3	80.5
INT4 (POPS)	181.0	362.1	383.0	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	4.6137	9.2274	36.9	73.8	5033.2	10.0664	40.3	80.5
FP8 (FLOPS)	181.0	N/A	NA	1961.2	3922.3	2614.9	41.8	20.9	41.8	2614.9	5229.8	20.9	41.8	4614	9227.4	36.9	73.8	5033.2	10,066.4	40.3	80.5
FP4 (FLOPS)	181.0	362.1	383.0	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	9227.5	18455	73.8	147.6	10,066.3	20,132.6	80.5	161.1

HPC PEAK THEORETICAL PERFORMANCE	MI210	MI250	MI250x	MI300A	MI300X	MI300X Platform	MI325X	MI325X Platform	MI350X	MI350X Platform	MI355X	MI 355X Platform
FP64 VECTOR TFLOPS	22.6	45.3	47.9	61.3	81.7	653.7	81.7	653.7	72.1	576.8	78.6	628.8
FP32 VECTOR TFLOPS	22.6	45.3	47.9	122.6	163.4	1307.4	163.4	1307.4	144.2	1.2	157.3	1.3
FP64 MATRIX TFLOPS	45.3	90.5	95.7	122.6	163.4	1307.4	163.4	1307.4	72.1	576.8	78.6	628.8
FP32 MATRIX TFLOPS	45.3	90.5	95.7	122.6	163.4	1307.4	163.4	1307.4	144.2	1.2	157.3	1.3

DECODERS AND VIRTUALIZATION†	MI210	MI250	MI250x	MI300A	MI300X	MI300X Platform	MI325X	MI325X Platform	MI350X	MI350X Platform	MI355X	MI 355X Platform
DECODER GROUPS				3	4	32	4	32	4	32	4	32
VIDEO TYPES				HEVC/H.265, AVC/H.264, V1, or AV1			HEVC/H.265, AVC/H.264, V1, or AV1		HEVC/H.265, AVC/H.264, VP9, or AV1	HEVC/H.265, AVC/H.264, V1, or AV1	HEVC/H.265, AVC/H.264, VP9, or AV1	HEVC/H.265, AVC/H.264, V1, or AV1
ZJPEG/MJPEG CODEC				24 cores, 8 cores per group	32 cores, 8 cores per group	256 cores, 8 cores per group	32 cores, 8 cores per group	256 cores, 8 cores per group	40 cores, 10 cores per group	320 cores, 10 cores per group	40 cores, 10 cores per group	320 cores, 10 cores per group
VIRTUALIZATION PARTITION SUPPORT WITH SR-IOV				Up to 3	Up to 8	Up to 64	Up to 8	Up to 64	Up to 8	Up to 8	Up to 8	Up to 64

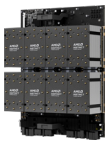
†Video codec acceleration (including at least the HEVC (H.265), H.264, VP9, and AV1 codecs) is subject to and not operable without inclusion/installation of compatible media players. GD-176











SPECIFICATIONS	MI210	MI250	MI250x	MI300A	MI300X	MI300X Platform	MI325X	MI325X Platform	MI350X	MI350X Platform	MI355X	MI 355X Platform
FORM FACTOR	PCIe Full-Height, Full-Length (Dual Slot)		OAM module	APU SH5 socket	OAM module	AMD universal baseboard (UBB)	OAM module	AMD universal baseboard (UBB)	OAM module	UBB 2.0 module	OAM module	UBB 2.0 module
FINFET LITHOGRAPHY I/O DIES (IODS)				5nm 6nm	5nm 6nm	5nm 6nm	5nm 6nm	5nm 6nm	TSMC 3nm/6nm FinFET 2 mirrored	TSMC 3nm/6nm FinFET 2 mirrored per module	TSMC 3nm/6nm FinFET 2 mirrored	TSMC 3nm/6nm FinFET 2 mirrored
GPU COMPUTE UNITS MATRIX CORES STREAM PROCESSORS PEAK ENGINE CLOCK (MHZ)	104 416 6,656	208 13,312		228 912 14,592 2100 MHz	304 1216 19,456 2100 MHz	2432 9728 155,648 2100 MHz	304 1216 19,456 2100 MHz	2432 9728 155,648 2100 MHz	256 1024 16,384 2200 MHz	2048 8192 131,072 2200 MHz	256 1024 16,384 2400 MHz	2048 8192 131,072 2400 MHz
X86 CPU CORES CORE TYPE	0	0	0	24 'Zen 4	0	0	0	0	0	0	0	0
MEMORY TYPE MEMORY CAPACITY MEMORY BANDWIDTH (MAX PEAK THEORETICAL) MEMORY COHERENCY MEMORY INTERFACE AMD INFINITY CACHE™ (LAST LEVEL) MEMORY CLOCK (GT/S)	HBM2e 64 GB Up to 1.6 TB/s No 4096 bits 1.6 GHz	HBM2e 128 GB Up to 3.2 TB/s No 8192 bits 1.6 GHz	HBM2e 128 GB Up to 3.2 TB/s Yes 8192 bits 1.6 GHz	HBM3 128 GB 5.3 TB/s I Yes 8192 bits 256 MB Up to 5.2	HBM3 192 GB 5.3 TB/s Yes 8192 bits 256 MB Up to 5.2	HBM3 1.5 TB 5.3 TB/s Yes 8192 bits 256 MB/GPU Up to 5.2	HBM3e 257 GB 5.3 TB/s Yes 8192 bits 256 MB Up to 5.2	HBM3e 2 TB 6 TB/s Yes 8192 bits 256 MB Up to 6	HBM3e 288 GB Up to 8 TB/s Yes (full chip) 8192 bits 256 MB Up to 8	HBM3e 2.3 TB 8 TB/s Yes 8192 bits 256 MB/GPU Up to 8	HBM3e 288 GB Up to 8 TB/s Yes (full chip) 8192 bits 256 MB Up to 8	HBM3e 2.3 TB 8 TB/s per GPU Yes 8192 bits 256 MB/GPU Up to 8
SCALE-UP INFINITY FABRIC™ LINKS LINK BANDWIDTH AGGREGATE BANDWIDTH PER GPU I/O TO HOST CPU I/O TYPE I/O BANDWIDTH	3 100 GB/s 300 GB/s 1 x16 PCIe Gen 4 Up to 64 GB/s	6 100 GB/s 600 GB/s 1 x16 PCIe Gen 4 Up to 64 GB/s	8 100 GB/s 800 GB/s 1 x16 PCIe Gen 4 Up to 64 GB/s	4 128 GB/s 256 GB/s 4 x161 PCIe Gen 5 128 GBs	7 128 GB/s 896 GB/s 1 x16 PCIe Gen 5 128 GB/s	7 128 GB/s 896 GB/s 8 x16 PCIe Gen 5 128 GB/s	7 128 GB/s 896 GB/s 1 x16 PCIe Gen 5 128 GB/s	7 128 GB/s 896 GB/s 8 x16 PCIe Gen 5 128 GB/s	8 144 GB/s 896 GB/s 5 x16 PCIe Gen 5 128 GB/s	7 (per GPU) 144 GB/s 1.2 TB/s 8 x16 PCIe Gen 5 128 GB/s	8 144 GB/s 896 GB/s 5 x16 PCIe Gen 5 128 GB/s	7 (per GPU) 144 GB/s 1.2 TB/s 5 x16 PCIe Gen 5 128 GB/s
RAS FEATURES	ECC memory and RAS support					Full-chip ECC memory, page retirement, page avoidance						
MAXIMUM TBP	300W	500W (air) 560W (DLC)	500W (air) 560W (DLC)	550W (air) 760W (DLC)	750W	750W per GPU	1000W	1000W per GPU	1000W	1000W per module	1400W	1400W per module

1 Assignable to host or GPU-to-GPU connectivity.

1. MI350 Family DS - PID#253461438

Footnote explanations are available at: https://www.amd.com/en/legal/claims/instinct.html

LE-92702-00 6/25