

AMD RADEON INSTINCT™ MI50 (16GB & 32GB)

UNLEASH DISCOVERY

The AMD Radeon Instinct™ MI50 server accelerator designed on the world's first 7nm FinFET technology process brings customers a full-feature set based on the industry's newest technologies. The MI50 is AMD's workhorse accelerator offering that is ideal for large scale deep learning. Delivering up to 26.5 TFLOPS of native half-precision (FP16) or up to 13.3 TFLOPS single-precision (FP32) peak floating point performance and INT8 support and combined with 16GB or 32GB of high-bandwidth HBM2 ECC memory, the AMD Radeon Instinct™ MI50 brings customers finely balanced performance needed for enterprise-class, mid-range compute capable of training complex neural networks for a variety of demanding deep learning applications in a cost-effective design.

Combine this outstanding performance with the MI50's advanced I/O capabilities, with support for PCIe® Gen 4 technology and the addition of AMD's dual Infinity Fabric™ Link technology providing up to 248 GB/s of aggregate GPU data transfer bandwidth¹; and the advanced efficiencies achieved with the use of AMD's 2nd Gen "Vega" 7nm technology; and you truly have one of the markets' most advanced, working-class server GPUs available for a variety of deep learning, VDI DaaS, cloud and general purpose HPC.

KEY FEAT URES

PERFORMANCE

Compute Units	60
Stream Processors	3,840
Peak INT8	Up to 53.6 TOPS
Peak FP16	Up to 26.5 TFLOPS
Peak FP32	Up to 13.3 TFLOPS
Peak FP64	Up to 6.6 TFLOPS
Bus Interface	PCIe® Gen 3 and Gen 4 Capable ²

MEMORY

Memory Size	16GB or 32GB HBM2
Memory Interface	4,096-Bits
Memory Clock	1 GHz
Memory Bandwidth	Up to 1 TB/s

RELIABILITY

ECC (Full-chip)	Yes ³
RAS Support	Yes ⁴

SCALABILITY

Infinity Fabric Links	2
MxGPU technology	Yes ⁵
OS Support	Linux® 64-bit
ROCm Compatible	Yes

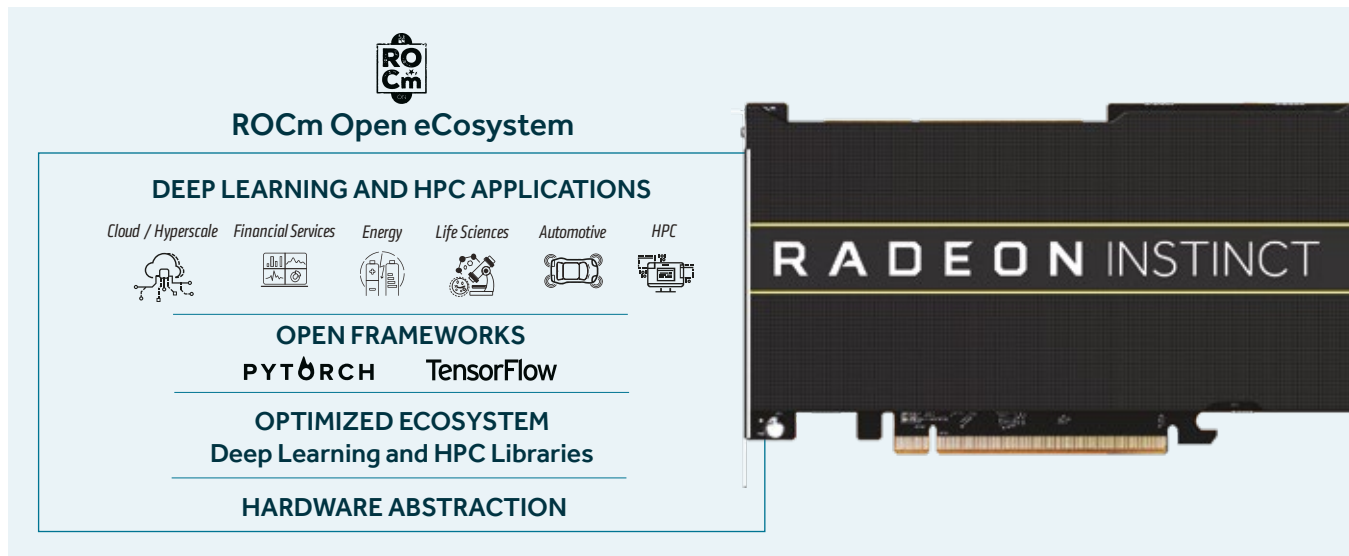
BOARD DESIGN

Board Form Factor	Full-Height, Dual Slot
Length	10.5" Long
Thermal	Passively Cooled
Max Power	300W TDP
Warranty	Three Year Limited ⁶

AMD RADEON INSTINCT™ MI50

ROCm Open Ecosystem

The ROCm Open eCosystem, is an open-source HPC/Hyperscale-class platform for GPU computing that's also programming-language independent. ROCm brings choice, minimalism and modular software development to GPU computing.



Optimized Deep Learning Operations

Optimized Deep Learning Operations with flexible mixed-precision **FP16, FP32, INT8 & INT4** capabilities bring customers supercharged compute performance to meet today's demanding system requirements of handling large data efficiently for training complex neural networks and running inference against those networks used in deep learning.

AMD Infinity Fabric Link

Two Infinity Fabric Links per GPU for high speed Direct-Connect GPU hives delivering up to **184 GB/s** GPU theoretical peer-to-peer bandwidth – 4.75x faster than using PCIe 3.0 alone.¹

Rapid FP64 Performance for HPC Workloads

The Radeon Instinct MI50 delivers great double precision performance with up to **6.6 TFLOPS** FP64 performance, enabling scientists and researchers across the globe to more efficiently process HPC parallel codes across several industries including life sciences, energy, finance, automotive and aerospace, academics, government, defense and more.

Ultra-Fast HBM2 Memory

With 16GB or 32GB HBM2 memory utilizing a four-stack memory configuration, the new **AMD Radeon Instinct MI50** delivers ultra-high memory bandwidth with up to **1 TB/s** memory bandwidth.

Footnotes:

1. As of Jun 19, 2019, Radeon Instinct™ MI50 and MI60 "Vega 7nm" technology-based accelerators support PCIe® Gen 4.0 providing up to 64 GB/s peak theoretical transport data bandwidth from CPU to GPU per card. Previous Gen Radeon Instinct compute GPU cards are based on PCIe Gen 3.0 providing up to 32 GB/s peak theoretical transport rate bandwidth performance. Peak theoretical transport rate performance is calculated by Baud Rate * width in bytes * # directions = GB/s per card. PCIe Gen3: 8 * 2 * 2 = 32 GB/s. PCIe Gen4: 16 * 2 * 2 = 64 GB/s.

Radeon Instinct™ MI50 and MI60 "Vega 7nm" technology-based accelerators include dual Infinity Fabric™ Links providing up to 184 GB/s peak theoretical GPU to GPU or Peer-to-Peer (P2P) transport rate bandwidth performance per GPU card. Combined with PCIe Gen 4 compatibility providing an aggregate GPU card I/O peak bandwidth of up to 248 GB/s. Performance guidelines are estimated only and may vary. Previous Gen Radeon Instinct compute GPU cards provide up to 32 GB/s peak PCIe Gen 3.0 bandwidth performance. Infinity Fabric Link technology peak theoretical transport rate performance is calculated by Baud Rate * width in bytes * # directions * # links = GB/s per card. Infinity Fabric Link: 23 * 2 * 2 = 92 GB/s. MI50/MI60 each have two links: 92 GB/s * 2 links per GPU = 184 GB/s. Refer to server manufacture PCIe Gen 4.0 compatibility and performance guidelines for potential peak performance of the specified server model numbers. Server manufacturers may vary configuration offerings yielding different results. <https://pcisig.com/>, <https://www.chipestimate.com/PCI-Express-Gen-4-a-Big-Pipe-for-Big-Data/Cadence/Technical-Article/2014/04/15>, <https://www.tomshardware.com/news/pcie-4.0-power-speed-express,32525.html> AMD has not independently tested or verified external/third party results/data and bears no responsibility for any errors or omissions therein. RIV-18

2. Works with PCIe® Gen 4.0 and Gen 3.0 compliant motherboards. Performance may vary from motherboard to motherboard. Refer to system or motherboard provider for individual product performance and features.

3. ECC support on 2nd Gen Radeon Instinct™ GPU cards, based on the "Vega 7nm" technology has been extended to full-chip ECC including HBM2 memory and internal GPU structures.

4. Expanded RAS (Reliability, availability and serviceability) attributes have been added to AMD's 2nd Gen Radeon Instinct™ Vega 7nm technology-based GPU cards and their supporting ecosystem including software, firmware and system level features. AMD's remote manageability capabilities using advanced out-of-band circuitry allow for easier GPU monitoring via I2C, regardless of the GPU state. For full system RAS capabilities, refer to the system manufacturer's guidelines for specific system models.

5. MxGPU SR-IOV technology for hardware-based virtualized compute deployments only available on specified Radeon Instinct™ server GPU cards (MI50 and MI60) when using KVM as a hypervisor. MxGPU SR-IOV virtualization usage beyond this requires further development with AMD for deployment. Current plans for "Vega 7nm" based product support will include 1 VM per GPU up to 8 GPUs in a VM.

6. The Radeon Instinct GPU accelerator products come with a three-year limited warranty. Please visit www.AMD.com/warranty for warranty details on the specific graphics products purchased. Toll-free phone service available in the U.S. and Canada only, email access is global.

© 2019 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, Radeon, Radeon Instinct, Infinity Fabric, and combinations thereof are trademarks of Advanced Micro Devices, Inc. OpenCL is a trademark of Apple Inc. used by permission by Khronos. PCIe and PCI Express are registered trademarks of PCI-SIG Corporation. Other product names used in this publication are for identification purposes only and may be trademarks of their respective companies.