



# OLCF SPOCK TRAINING: AMD HARDWARE AND SOFTWARE

May 20th, 2021

AMD Public Use

# NEW FRONTIER FOR HPC AMD FUELING THE ERA OF EXASCALE



Oak Ridge Frontier Supercomputer  
1.5 Exaflops Powering the World's Leading Open Science System

- Custom EPYC™ CPU  
Optimized for HPC & AI
- HPC-Customized  
Compute Engines in  
Instinct™ GPU
- High Bandwidth, Low  
Latency CPU & GPU for  
Coherency
- Open Software  
Platform

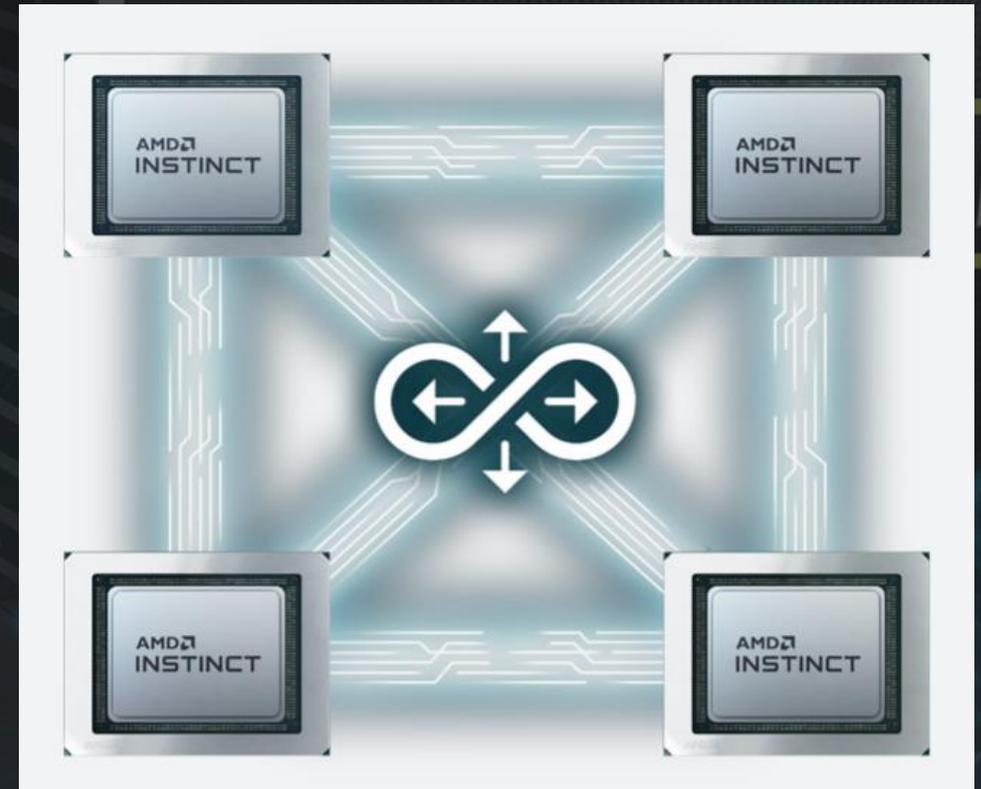


[https://www.olcf.ornl.gov/wp-content/uploads/2019/05/frontier\\_specsheet.pdf](https://www.olcf.ornl.gov/wp-content/uploads/2019/05/frontier_specsheet.pdf)

# The Frontier Node at a Glance

- Custom EPYC™ HPC-optimized processor
- Four Instinct™ accelerators
- Coherent connectivity:
  - Via Infinity Fabric™ interconnect
  - Tightly integrated
  - Unified memory space

<https://www.amd.com/en/products/exascale-era>



# 3RD GEN AMD EPYC™ PROCESSORS AT A GLANCE

## COMPUTE

- AMD “Zen3” x86 cores (64 core / 128 threads)
- Up to 32MB L3 cache / core, shared by each chiplet
- Flatter NUMA domain, reduced latency w/ smaller system diameter
- TDP range: 120W-280W

## MEMORY

- 8 channel DDR4 with ECC up to 3200 MHz  
Option for 6 channel Memory Interleaving<sup>1</sup>
- RDIMM, LRDIMM, 3DS, NVDIMM-N
- 2 DIMMs/channel capacity of 4TB/socket (256GB DIMMs)

## PERFORMANCE

- +Increased socket performance, single threaded performance, performance per core\*
- Infinity Fabric™ Gen 2 (xGMI-2)



## INTEGRATED I/O – NO CHIPSET

- 128 lanes PCIe™ Gen3/4
  - Used for PCIe, SATA, and Coherent Interconnect
  - Up to 32 SATA or NVMe™ direct connect devices
  - 162 lane option (2P config)
- Server Controller Hub (USB, UART, SPI, LPC, I2C, etc.)

## SECURITY

- Dedicated Security Subsystem
- Secure Boot, Hardware Root-of-Trust
- SME (Secure Memory Encryption)
- SEV-ES (Secure Encrypted Virtualization & Register Encryption)
- SNP (Secure Nested Paging)

# AMD INSTINCT™ MI100

GPU Architecture	MI100
Compute Units	120
Peak FLOPS	185 TF FP16   46 TF FP32   11.5 TF FP64
Memory Capacity	32GB HBM2 w/ ECC
Peak Memory Bandwidth	1.2 TB/s
Guest OS	Ubuntu , RHEL, SUSE SLES
API Support	ROCm , OpenCL™, HIP, OpenMP
Framework Support	TensorFlow, PyTorch/Caffe2, Caffe, Kokkos, RAJA
Form Factors	PCIe
Typical Board Power / Input Power	300W / 12V
Infinity Fabric Links / Topology	3 Links / Up to 4 GPUs – Fully Connected Quads

# MI100 FP INSTRUCTION RATES AND MFMA INSTRUCTIONS

For reference: MI50/MI60 FP instruction rates:

- ▲ 64 FP32 FMA instructions (128 FP32 operations) / CU / Cycle.
- ▲ 32 FP64 FMA instructions (64 FP64 operations) / CU / Cycle (half rate double precision).

MI100 has the same instruction rate, but it has more CUs per GPU (120 for MI100 vs 64 for MI60 vs 60 for MI50).

In addition to that, MI100 has new MFMA (Matrix Fused Multiply-Add) operations.

# MI100 FP INSTRUCTION RATES AND MFMA INSTRUCTIONS

- ▲ MI100 has new hardware units on the CUs for doing matrix multiplications instructions. These instructions operate on a per-wavefront basis, rather than on a per-thread basis.
- ▲ The new instructions support various basic matrix block sizes, number of blocks per instruction, and data types.
  - ▲ Input FP32 to output FP32
  - ▲ Input FP16 to output FP32
  - ▲ Input BF16 to output FP32
  - ▲ Input Int8 to output Int32

# MI100 FP INSTRUCTION RATES AND MFMA INSTRUCTIONS

Generic instruction format:  $V\_MFMA\_FO\_M \times N \times KFI$

- ▲  $FO$  is the output data type (F32, I32)
- ▲  $FI$  is the input data type (F32, F16, BF16, I8)

$M$ ,  $N$ , and  $K$  specify matrix dimensions: the instruction performs  $b$  independent computations  $D = A \times B + C$ , where

- ▲  $D$  and  $C$  are  $M \times N$  matrices
- ▲  $A$  is a  $M \times K$  matrix
- ▲  $B$  is a  $K \times N$  matrix
- ▲ Number of blocks  $b$  is not encoded in the instruction name; see tables on next slides for a specification.

# MI100 FP INSTRUCTION RATES AND MFMA INSTRUCTIONS

Matrix instructions with instruction rates (number of blocks in brackets)

128 FMA (256 flops) / CU / Cycle

V\_MFMA\_F32\_32X32X1F32 (2)  
 V\_MFMA\_F32\_16X16X1F32 (4)  
 V\_MFMA\_F32\_4X4X1F32 (16)  
 V\_MFMA\_F32\_32X32X2F32 (1)  
 V\_MFMA\_F32\_16X16X4F32 (1)

512 FMA (1024 flops) / CU / Cycle

V\_MFMA\_F32\_32X32X4F16 (2)  
 V\_MFMA\_F32\_16X16X4F16 (4)  
 V\_MFMA\_F32\_4X4X4F16 (16)  
 V\_MFMA\_F32\_32X32X8F16 (1)  
 V\_MFMA\_F32\_16X16X16F16 (1)

256 FMA (512 flops) / CU / Cycle

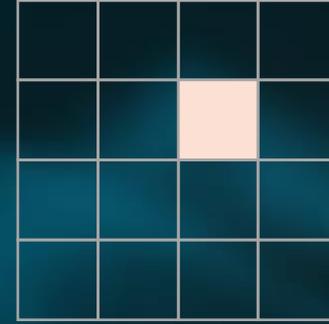
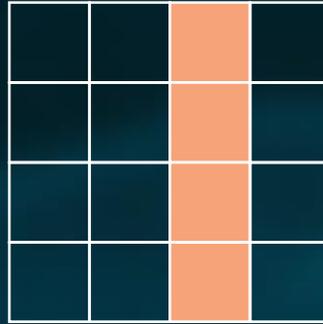
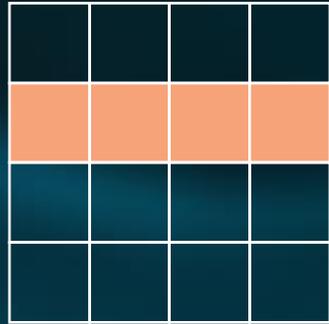
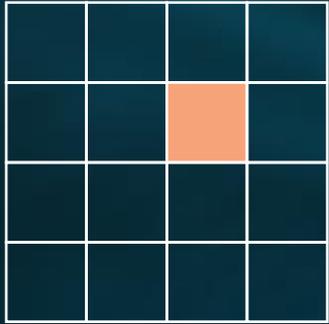
V\_MFMA\_F32\_32X32X2BF16 (2)  
 V\_MFMA\_F32\_16X16X2BF16 (4)  
 V\_MFMA\_F32\_4X4X2BF16 (16)  
 V\_MFMA\_F32\_32X32X4BF16 (1)  
 V\_MFMA\_F32\_16X16X8BF16 (1)

512 FMA (1024 iops) / CU / Cycle

V\_MFMA\_I32\_32X32X4I8 (2)  
 V\_MFMA\_I32\_16X16X4I8 (4)  
 V\_MFMA\_I32\_4X4X4I8 (16)  
 V\_MFMA\_I32\_32X32X8I8 (1)  
 V\_MFMA\_I32\_16X16X16I8 (1)

# INTERMEZZO: LINEAR ALGEBRA 101

Computing a matrix multiplication using dot products:



$$D = A \times B + C$$

Every entry in D is computed as the dot product of a row from A and a column of B (plus an entry of C)

This is **not** what the MFMA instructions do.

# INTERMEZZO: LINEAR ALGEBRA 101

Computing a matrix multiplication by accumulating outer products:

V\_MFMA\_F32\_4x4X1F32, applied four times



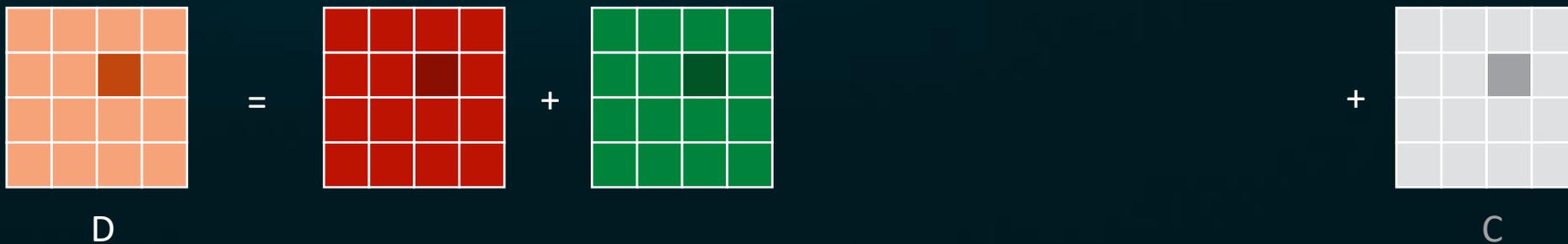
D

C

# INTERMEZZO: LINEAR ALGEBRA 101

Computing a matrix multiplication by accumulating outer products:

V\_MFMA\_F32\_4x4X2BF16, applied twice



# LEVERAGING THE MFMA INSTRUCTIONS

Current support for using MFMA instructions:

- ▲ AMD libraries: rocBLAS
  - ▲ Library launches GPU kernel(s) and decides on which pipeline to use, what matrix sizes to use for decomposition
- ▲ Write kernel completely in assembly
  - ▲ Oh boy...
- ▲ Sprinkle kernels with some inline assembly or intrinsics
  - ▲ Intrinsics: straightforward, but sub-optimal
  - ▲ Inline assembly: hard to get correct, likely to be even less optimal than intrinsics.

Not currently supported:

- ▲ Libraries of device functions, utilizing the matrix operations, that can be called from kernels
- ▲ Abstraction frameworks (Kokkos, Raja, OCCA)
  - ▲ These would have to use one of the other mechanisms internally

# POTENTIAL USE CASES FOR THE MFMA INSTRUCTIONS

## ▲ Kripke

- ▲ Representative of deterministic particle transport codes
- ▲ Key performance component: LPlus (RHS)
- ▲ Performed as batched DGEMM operations

## ▲ Laghos (CEED Center)

- ▲ Representative of high-order FEM
- ▲ Mass matrix, assuming partial assembly

## ▲ Other ideas

- ▲ Libraries for spline interpolation
- ▲ Batched linear transformations such as rotation, reflection, or (using homogeneous coordinates) translation

## ▲ Bring us your use cases!

# AMD MI100 Key Takeaways

- MI100 is available now
- Shares Instruction Set Architecture (ISA) with future GPUs
  - Graphics Compute Next (GCN)
  - GCN ISA:  
[https://developer.amd.com/wp-content/resources/CDNA1\\_Shader\\_ISA\\_14December2020.pdf](https://developer.amd.com/wp-content/resources/CDNA1_Shader_ISA_14December2020.pdf)
  - “GCN ISA Reference Guide”
- Shares core software and tools environment with future GPUs
  - Use libraries for BLAS

**Application porting can begin now**

# FRONTIER



Questions?

# DISCLAIMERS AND ATTRIBUTIONS

The information contained herein is for informational purposes only and is subject to change without notice. Timelines, roadmaps, and/or product release dates shown in these slides are plans only and subject to change. “Polaris”, “Vega”, “Radeon Vega”, “Navi”, “Zen” and “Naples” are codenames for AMD architectures, and are not product names.

While every precaution has been taken in the preparation of this document, it may contain technical inaccuracies, omissions and typographical errors, and AMD is under no obligation to update or otherwise correct this information. Advanced Micro Devices, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this document, and assumes no liability of any kind, including the implied warranties of noninfringement, merchantability or fitness for particular purposes, with respect to the operation or use of AMD hardware, software or other products described herein. No license, including implied or arising by estoppel, to any intellectual property rights is granted by this document. Terms and limitations applicable to the purchase or use of AMD’s products are as set forth in a signed agreement between the parties or in AMD’s Standard Terms and Conditions of Sale.

The information contained herein is for informational purposes only and is subject to change without notice. While every precaution has been taken in the preparation of this document, it may contain technical inaccuracies, omissions and typographical errors, and AMD is under no obligation to update or otherwise correct this information. Advanced Micro Devices, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this document, and assumes no liability of any kind, including the implied warranties of noninfringement, merchantability or fitness for particular purposes, with respect to the operation or use of AMD hardware, software or other products described herein. No license, including implied or arising by estoppel, to any intellectual property rights is granted by this document. Terms and limitations applicable to the purchase or use of AMD’s products are as set forth in a signed agreement between the parties or in AMD’s Standard Terms and Conditions of Sale. GD-18

©2021 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, Ryzen, EPYC, Radeon, Radeon Instinct, and combinations thereof are trademarks of Advanced Micro Devices, Inc. Other product names used in this publication are for identification purposes only and may be trademarks of their respective companies.

# ENDNOTES

EPYC-18: Max boost for AMD EPYC processors is the maximum frequency achievable by any single core on the processor under normal operating conditions for server systems.

GD-83: Use of third party marks / logos/ products is for informational purposes only and no endorsement of or by AMD is intended or implied.

GD-183: AMD Infinity Guard features vary by EPYC™ Processor generations. Infinity Guard security features must be enabled by server OEMs and/or Cloud Service Providers to operate. Check with your OEM or provider to confirm support of these features. Learn more about Infinity Guard at <https://www.amd.com/en/technologies/infinity-guard>.

MLN-001: AMD EPYC™ 7003 Series processors require a BIOS update from your server or motherboard manufacturer if used with a motherboard designed for the AMD EPYC™ 7002 Series processors. A motherboard designed at minimum for EPYC 7002 processors is required for EPYC 7003 Series processors.

MLN-003: Based on AMD internal testing as of 02/1/2021, average performance improvement at ISO-frequency on an AMD EPYC™ 72F3 (8C/8T, 3.7GHz) compared to an AMD EPYC™ 7F32 (8C/8T, 3.7GHz), per-core, single thread, using a select set of workloads including estimated SPECrate®2017\_int\_base, SPECrate®2017\_fp\_base, and representative server workloads.

MLN-004: Login VSI™ Pro v4.1.40.1 comparison based on AMD internal testing as of 02/01/2021 measuring the maximum “knowledge worker” desktop sessions within VSI Baseline +1000ms response time using VMware ESXi 7.0u1 and VMware Horizon 8 on a server using 2x AMD EPYC 7763 versus a server with 2x Intel Xeon Gold 6258R for ~112% more max [~2.1x the] performance. Results may vary.

MLN-006: HammerDB 4.0 OLTP comparison based on AMD internal testing on Oracle® 19c RDBMS as of 02/01/2021 on a server using 2x AMD EPYC 75F3 versus a server using 2x AMD EPYC 7542 for ~19% more [~1.2x the] performance. TPROC-C: OLTP workload profile in HammerDB derived, from the TPC-C specification using 2000 Warehouses. Results may vary.

MLN-007: Results as of 01/28/2021 using SPECrate®2017\_int\_base. The 2P AMD EPYC 7763 a measured estimated score of 798, versus the current highest score Intel Cascade Lake Refresh server with a score of 397 using 2P Intel Gold 6258R, <https://spec.org/cpu2017/results/res2020q3/cpu2017-20200915-23981.pdf>. OEM published score(s) for EPYC may vary. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See [www.spec.org](http://www.spec.org) for more information.

MLN-008: Results as of 01/28/2021 using SPECrate®2017\_fp\_base. The 2P AMD EPYC 7763 has an a measured estimated score of 614.7 versus the current highest score Intel Cascade Lake Refresh server with a score of 309 and 2P Intel Gold 6258R, <https://spec.org/cpu2017/results/res2020q3/cpu2017-20200915-23979.pdf>. OEM published score(s) for EPYC may vary. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See [www.spec.org](http://www.spec.org) for more information.

MLN-016: Results as of 01/28/2021 using SPECrate®2017\_int\_base. The AMD EPYC 7763 estimated score of 798 is higher than the current highest 2P server with an AMD EPYC 7H12 and a score of 717, <https://spec.org/cpu2017/results/res2020q2/cpu2017-20200525-22554.pdf>. OEM published score(s) for EPYC may vary.

MLN-017: Results as of 01/28/2021 using SPECrate®2017\_int\_base. The AMD EPYC 75F3 a measured estimated score of 546 has up to 23% higher than a comparable 2P EPYC 7002 CPU powered server, the 7532 with a score of 444, <https://spec.org/cpu2017/results/res2020q3/cpu2017-20200622-23002.pdf>. OEM published score(s) for 3rd Gen EPYC may vary. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See [www.spec.org](http://www.spec.org) for more information. "

MLN-018: Results as of 02/20/2021 using SPECrate®2017\_int\_base. The AMD EPYC 7763 a measured estimated score of 804 which is higher than the current highest 2P server with an AMD EPYC 7H12 and a score of 717, <https://spec.org/cpu2017/results/res2020q2/cpu2017-20200525-22554.pdf>. OEM published score(s) for EPYC may vary. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See [www.spec.org](http://www.spec.org) for more information.

MLN-040k: Based on SPECrate®2017\_int\_base on 02/20/2021, a server powered by two 64c AMD EPYC 7763 CPUs has a score of 839 in a compliant result run on an ASUS RS720A-E11(KMPP-D32); with Memory: 1 TB (16 x 64 GB 2Rx4 PC4-3200AA-R); OS: SUSE Linux Enterprise Server 15 SP2 (x86\_64) Kernel 5.3.18-22-default; Compiler: C/C++/Fortran: Version 3.0.0 of AOCC. Versus the current highest score Intel Cascade Lake Refresh server with a score of 397 using 2P Intel Gold 6258R, <https://spec.org/cpu2017/results/res2020q3/cpu2017-20200915-23981.pdf>. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See [www.spec.org](http://www.spec.org) for more information.

# ENDNOTES

MLN-041k: Based on SPECrate®2017\_fp\_base on 02/20/2021, a server powered by two 64c AMD EPYC 7763 CPUs has a score of 651 a compliant result run on an ASUS RS720A-E11(KMPP-D32); with Memory: 1 TB (16 x 64 GB 2Rx4 PC4-3200AA-R); OS: SUSE Linux Enterprise Server 15 SP2 (x86\_64) Kernel 5.3.18-22-default; Compiler: C/C++/Fortran: Version 3.0.0 of AOCC. Versus the current highest score Intel Cascade Lake Refresh server with a score of 309 with a 2P Intel Gold 6258R based server, <https://spec.org/cpu2017/results/res2020q3/cpu2017-20200915-23979.pdf>. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See [www.spec.org](http://www.spec.org) for more information.

MLN-043: WRF version 4.1.5 comparison based on AMD internal testing completed on 2/17/2021 on a reference platform with 2x EPYC™ 75F3 (32C) compared to an Intel server on a production system with 2x Intel® Xeon® Gold 6258R (28C) processors. Results may vary.

MLN-044A: SPECjbb®2015-MultiJVM critical-jOPS comparison based on highest system results published as of 03/11/2021. Configurations: 2x AMD EPYC 7763 (301297 SPECjbb2015-MultiJVM critical-jOPS, 359067 SPECjbb2015-MultiJVM max-jOPS, <https://spec.org/jbb2015/results/res2021q1/jbb2015-20210224-00610.html>) versus 2x Intel Xeon Platinum 8280 (138942 SPECjbb2015-MultiJVM critical-jOPS, 169,598 SPECjbb2015-MultiJVM max-jOPS, <https://spec.org/jbb2015/results/res2019q2/jbb2015-20190314-00428.html>) for 117% higher [~2.2x the] performance. SPEC®, and the benchmark SPECjbb® are registered trademarks of the Standard Performance Evaluation Corporation. Learn more at [spec.org](http://spec.org).

MLN-046: STREAM Triad GB/s comparison based on AMD internal testing as of 02/01/2021 on a server with 2x AMD EPYC 7763 versus the 2x AMD EPYC 7742 processors score. Results may vary.

MLN-047: STREAM Triad GB/s comparison based on AMD internal testing and a published competitive Intel result as of 02/01/2021 Configurations: on a server with 2x AMD EPYC 75F3 (371.5 GB/s) versus the 2x Intel Xeon Gold 6258R processors score at (224 GB/s, <https://newsroom.intel.com/news/product-fact-sheet-accelerating-5g-network-infrastructure-core-edge>) for ~66% more [~1.7x the] performance. Results may vary.

MLN-048: ANSYS® CFX® 2021.1 comparison based on AMD internal testing as of 02/05/2021 measuring the time to run the Release 14.0 test case simulations (converted to jobs/day - higher is better) using a server with 2x AMD EPYC 75F3 versus 2x Intel Xeon Gold 6258R. The External Flow Over a LeMans Car test case individually was 112% [2.1x the] per node or 85% per core performance. Results may vary.

MLN-048A: ANSYS® CFX® 2021.1 comparison based on AMD internal testing as of 02/05/2021 measuring the time to run the Release 14.0 test case simulations (converted to jobs/day - higher is better) using a server with 2x AMD EPYC 75F3 utilizing 1TB (16x 64 GB DDR4-3200) versus 2x Intel Xeon Gold 6258R utilizing 384 GB (12x 32 GB DDR4-3200). The External Flow Over a LeMans Car test case individually was 112% [2.1x the] per node or 85% per core performance. Results may vary.

MLN-049A: ANSYS® LS-DYNA® version 2021.1 comparison based on AMD internal testing as of 02/05/2021 measuring the time to run 3cars, test case simulation (converted to jobs/day - higher is better) Configurations using a server with 2x AMD EPYC 75F3 versus a server with 2x Intel Xeon Gold 6258R utilizing 384 GB (12x 32 GB DDR4-3200). The 3cars test case gain individually was 126% [~2.26x the] per node or ~98% per core jobs/day performance. Results may vary.

MLN-050: ESI Virtual Performance Solution (VPS better known as PAM-CRASH®) version 2020.0 comparison based on AMD internal testing as of 02/05/2021 measuring the neon test case simulation (converted to jobs/day - higher is better) using a server with 2x AMD EPYC 75F3 versus a server with 2x Intel Xeon Gold 6258R for ~43% more [~1.4x the] per node or ~25% per core jobs/day performance. Results may vary.

MLN-053: Star-CCM+ 2020.3 comparison based on AMD internal testing as of 02/05/2021 measuring the average seconds to complete 11 test cases and converted to jobs/day (higher is better) using a server with 2x AMD EPYC 75F3 versus a server with 2x Intel Xeon Gold 6258R. The KCS Marine Hull with No Rudder in Fine Waves test case individually was ~79% more [~1.7x the] per node or ~57% better per core performance. Results may vary.

MLN-055: AMD EPYC 7003 CPUs with PCIe4 lanes have 2X the I/O throughput capacity per lane than any Intel Xeon Scalable CPU which use PCIe3. PCIe4 provides 16GB/s of link bandwidth versus PCIe3 with 8Gb/s, <https://pcisig.com/pci-express-delivering-needed-bandwidth-open-compute-project>.

MLN-056: Each AMD EPYC 7003 processor has 8 memory channels. Each Intel Xeon Scalable processor has 6 memory channels.  $8 - 6 = 2 \div 6 = 0.33$  AMD EPYC has 33% more memory bandwidth. Class based on industry-standard pin-based (LGA) X86 processors.

MLN-057K: Based on SPECrate®2017\_int\_base on 02/20/2021, a server powered by two 8c AMD EPYC 72F3 CPU has a measured estimated score of 176 with a per core score of 11.00 which is a higher per core performance score than any currently posted in any SPEC.org publication. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See [www.spec.org](http://www.spec.org) for more information.

# ENDNOTES

MLN-058: A 2P AMD EPYC 72F3 8 core CPU powered server has a measured estimated SPECrate®2017\_int\_base score of 220 yielding a per core score of 13.75. The posted score on SPEC.org as of 02/20/2021 yielding the highest per core performance is a server with one AMD EPYC 7F32 8 core CPU with a per core score of 12.875, from a published score of 103, <https://spec.org/cpu2017/results/res2020q2/cpu2017-20200316-21228.pdf>. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See [www.spec.org](http://www.spec.org) for more information.

MLN-059: EPYC 7313 and 7343 CPU powered 2P servers have measured estimated SPECrate®2017\_int\_base scores of 287 and 295 respectively (287+295=582, 582/2=291), is up to 25% higher than than highest posted score 2P EPYC 7282 and 7302 powered servers with SPECrate®2017\_int\_base scores of 215 and 246 respectively (215+246=461, 461/2=230.5). 291/230.5=1.26. 16 core EPYC 7003 CPUs have 126% the perf or 26% more performance of 16c 7002 CPUs. OEM published score(s) for 3rd Gen EPYC may vary. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See [www.spec.org](http://www.spec.org) for more information.

MLN-060: EPYC 7643 and 7763 CPU powered 2P servers have measured estimated SPECrate®2017\_fp\_base scores of 510 and 614.7 respectively (average score 562.35), is up to 15% higher than than 2P EPYC 7552 and 7662 powered servers with SPECrate®2017\_fp\_base scores of 435 and 546 respectively (average score 490.5). OEM published score(s) for 3rd Gen EPYC may vary. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See [www.spec.org](http://www.spec.org) for more information.

MLN-061K: As of Feb. 20, 2021, the Intel log trendline from top SPECrate®2017\_int\_base published scores to date for 2P Intel based Xeon SP (LGA socketed) servers for each of 2017, 2018, 2019, 2020, and 2021. The AMD log trendline from top SPECrate®2017\_int\_base published score to date, for 2P Intel based AMD EPYC servers for each of 2017, 2018, 2019, and 2020, and for 2021 the measured estimate score for the EPYC 7763 for SPECrate®2017\_int\_base. The lines below are organized as: Year, CPU model, SPEC score, URL. 2017, Intel 8180, 302, <https://spec.org/cpu2017/results/res2017q4/cpu2017-20170928-00070.pdf> 2018, Intel 8180, 304, <https://spec.org/cpu2017/results/res2018q3/cpu2017-20180709-07701.pdf> 2019, Intel 8280L, 364, should be 8280L <https://spec.org/cpu2017/results/res2019q2/cpu2017-20190429-12779.pdf> 2020, Intel 6258R, 397, <https://spec.org/cpu2017/results/res2020q3/cpu2017-20200915-23981.pdf> 2021, Intel 6258R, 397, <https://spec.org/cpu2017/results/res2020q3/cpu2017-20200915-23981.pdf>. 2017, AMD EPYC 7601, 275, <https://spec.org/cpu2017/results/res2017q4/cpu2017-20171211-01594.pdf> 2018, EPYC 7601, 282, <https://spec.org/cpu2017/results/res2018q3/cpu2017-20180827-08666.pdf> 2019, EPYC 7742, 701, <https://spec.org/cpu2017/results/res2019q4/cpu2017-20191125-20001.pdf> 2020, EPYC 7H12, 717, <https://spec.org/cpu2017/results/res2020q2/cpu2017-20200525-22554.pdf> 2021, EPYC 7763, 839 a compliant result run on an ASUS RS720A-E11(KMPP-D32); with Memory: 1 TB (16 x 64 GB 2Rx4 PC4-3200AA-R); OS: SUSE Linux Enterprise Server 15 SP2 (x86\_64) Kernel 5.3.18-22-default; Compiler: C/C++/Fortran: Version 3.0.0 of AOCC. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See [www.spec.org](http://www.spec.org) for more information.

MLN-062: SPECrate®2017\_int\_base comparison based on internally measured estimates of 2P 3rd Gen AMD EPYC™ CPUs powered server as of 2/18/21; and the highest performing 2P Intel Xeon Gold, 2P Intel Xeon Silver and select 2P 2nd Gen AMD EPYC powered servers published at [spec.org](http://spec.org) as of 1/28/21. 3rd Gen AMD EPYC measured estimates 72F3=176, 7313=287, 7343=295, 73F3=320, 7413=377, 7443=397, 74F3=432, 7453=406, 7513=453, 7543=503, 75F3=546, 7643=617, 7663=657, 7713=695, 7763=802; Published URLs: Intel Xeon 6258R – 397, <http://spec.org/cpu2017/results/res2020q3/cpu2017-20200915-23981.html>; Intel Xeon 4216 – 193 <http://spec.org/cpu2017/results/res2020q3/cpu2017-20200804-23653.html>; EPYC 7532 – 444, <http://spec.org/cpu2017/results/res2020q3/cpu2017-20200622-23002.html>; EPYC 7352 – 335, <http://spec.org/cpu2017/results/res2020q2/cpu2017-20200608-22792.html>; EPYC 7282 – 215, <http://spec.org/cpu2017/results/res2020q2/cpu2017-20200608-22784.html>; EPYC 7272 – 171, <http://spec.org/cpu2017/results/res2020q2/cpu2017-20200608-22780.html>; EPYC 7262 – 135, <http://spec.org/cpu2017/results/res2020q2/cpu2017-20200608-22777.html>; EPYC 7252 – 119, <http://spec.org/cpu2017/results/res2021q1/cpu2017-20201214-24568.html> OEM published score(s) for 3rd Gen EPYC may vary. SPEC® and the benchmark name SPECrate® are registered trademarks of the Standard Performance Evaluation Corporation.

MLN-083K: Based on SPECrate®2017\_int\_base on 02/20/2021, a server powered by two 64c AMD EPYC 7763 CPUs has a score of 839 which is higher than any currently posted SPEC 2P server score. Per socket score would be 839/2=419.5 which is higher than any per socket score (score / sockets). This is a compliant result run on a Lenovo ThinkSystem SR645; with Memory: 2 TB (32 x 64 GB 2Rx4 PC4-3200AA-R), OS: SUSE Linux Enterprise Server 12 SP5 (x86\_64) Kernel 4.12.14-120-default; Compiler: C/C++/Fortran: Version 3.0.0 of AOCC. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See [www.spec.org](http://www.spec.org) for more information.

# ENDNOTES

MLNCO-001 The Bare Metal TCO (total cost of ownership) Estimator solution compares the selected AMD EPYC™ and Intel® Xeon® CPU based server solutions required to deliver a TOTAL PERFORMANCE of 25000 unit of integer performance based on published the SPECrate®2017\_int\_base scores for Intel and AMD measured estimated scores for AMD EPYC 7003. This analysis is based on tool VERSION: 02/20/2021 v0.9982. This estimation reflects a 4 year time frame. This analysis compares a 2 CPU AMD EPYC EPYC\_7763 powered server with a measured estimated SPECrate®2017\_int\_base score of 802; compared to a 2 CPU Intel Xeon Gold\_6258R based server with a SPECrate®2017\_int\_base score of 397, <https://spec.org/cpu2017/results/res2020q3/cpu2017-20200915-23981.pdf>. Both AMD EPYC and Intel based servers use the same estimated cost for the following elements of the analysis: server chassis size of 2RU at a cost of \$2500 per chassis; internal storage \$380; physical servers managed per admin: 30; fully burdened cost per admin \$110500; server rack size of 42; space allowance per rack of 27 sq feet; monthly cost of data center space \$20 per sq foot; cost per kW for power \$0.12; power drop per rack of 12kW; and a PUE (power usage effectiveness of 2). The EPYC powered solution estimates are: 32 2P EPYC 7763 powered total servers at a hardware only acquisition cost of \$19232 per server, which includes total system memory of 768GB, which is 6GB of memory / core and a total system memory cost of \$3072; internal storage cost of \$380. The total AMD EPYC hardware acquisition cost for this solution is \$615424. Each server draws ~611kWhr per month. For the 4 years of this EPYC powered solution analysis the: total solution power cost is ~\$225240 which includes the PUE factor; the total admin cost is ~\$471468, and the total real estate cost is ~\$77760. The total 4 year TCO estimate for the AMD solution is \$1389892. The Intel based solution estimates are: 63 2P Xeon Gold 6258R based total servers at a hardware only acquisition cost of \$12316 per server, which includes total system memory of 384GB, which is 6.9GB of memory / core and a total system memory cost of \$1536; internal storage cost of \$380. The total Intel hardware acquisition cost for this solution is \$775908. Each server draws ~476kWhr per month. For the 4 years of this Intel based solution analysis the: total solution power cost is \$345460 which includes the PUE factor; the total admin cost is ~\$928200, and the total real estate cost is ~\$103680. The total 4 year TCO estimate for the Intel solution is \$2153248. Delivering 25000 of estimated SPECrate®2017\_int\_base performance, produces the following estimated results: the AMD EPYC solution requires 49% fewer servers [1-(AMD server count / Intel server count)]; 25% less space [1-(AMD rack count / Intel rack count)]; 35% less power [1-(AMD power cost / Intel power cost)]; providing a 35% lower 4 year TCO [1-(AMD TCO / Intel TCO)]. AMD processor pricing based on 1KU price as of February 2021. Intel® Xeon® Scalable processor data and pricing from <https://ark.intel.com> as of September 2020. All pricing is in USD. Results shown here are estimates and actual results may vary. Product and company names are for informational purposes only and may be trademarks of their respective owners. SPECrate® scores as of 02/20/2021. AMD EPYC performance numbers based on AMD internal estimates and are subject to change based on actual results. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See [www.spec.org](http://www.spec.org) for more information. AMD EPYC performance numbers based on AMD measured internal estimates and are subject to change based on actual results. Results generated by the AMD EPYC™ BARE METAL SERVER TCO ESTIMATION TOOL, VERSION: 02/20/2021 v0.9982.

# ENDNOTES

The information contained herein is for informational purposes only, and is subject to change without notice. While every precaution has been taken in the preparation of this document, it may contain technical inaccuracies, omissions and typographical errors, and AMD is under no obligation to update or otherwise correct this information. Advanced Micro Devices, Inc. makes no representations or warranties with respect to the accuracy or completeness of the contents of this document, and assumes no liability of any kind, including the implied warranties of noninfringement, merchantability or fitness for particular purposes, with respect to the operation or use of AMD hardware, software or other products described herein. No license, including implied or arising by estoppel, to any intellectual property rights is granted by this document. Terms and limitations applicable to the purchase or use of AMD's products are as set forth in a signed agreement between the parties or in AMD's Standard Terms and Conditions of Sale.

Timelines, roadmaps, and/or product release dates shown in these slides are plans only and subject to change. "Rome", "Milan", "Zen", "Zen2", and "Zen3" are codenames for AMD architectures, and are not product names.

©2021 Advanced Micro Devices, Inc. All rights reserved. AMD, the AMD Arrow logo, EPYC, Infinity Fabric, and combinations thereof are trademarks of Advanced Micro Devices, Inc. Other product names used in this publication are for identification purposes only and may be trademarks of their respective companies. SPEC®, SPECjbb® and SPECrate® are registered trademarks of Standard Performance Evaluation Corporation. Learn more at [spec.org](https://www.spec.org).

**AMD** 