OLCF SPOCK TRAINING: AMD ROCM SOFTWARE

May 20th, 2021

AMD Public Use
Five Pillars of Exascale Software

OpenMP: Portable Compiler Directives

ROCm: Make Applications Open & Portable

HIP: ML Optimized

TensorFlow, Caffe2: 100% Open & Performance Portable

Unified CPU + GPU Tools
AMD ROCm Software

• ROCm = Radeon Open Compute Platform
  • Open-source collection of AMD software: drivers, compilers, and libraries

• ROCr is a hardware abstraction layer which provides HSA derived semantics
  • HSA = Heterogeneous System Architecture

• HIP = Heterogeneous Interface for Portability
  • HIP implements language runtime APIs in terms of ROCr APIs and semantics
  • HIP is a single source C++-like language

• LLVM-based compilers
SOFTWARE PUSH FOR EXASCALE IN 2021
A GROWING ECOSYSTEM

2019: ROCm 2.0
Building the Foundation

2020: ROCm 3.0
Growing Essential Capabilities

2021: ROCm 4.X
Production-Ready HPC & ML Stack
AMD ROCm 4.2

1. ROCm 4.2 is the latest release
2. Significant improvements in kernel launch latencies
3. Extends support to Stream Memory Operations
4. Cross-architectures:
   • HIP_PLATFORM_NVIDIA is defined if the HIP platform targets NVIDIA
   • _HIP_PLATFORM_NVCC__ was used previously if the HIP platform targeted NVIDIA
AMD ROCm References

- HIP Programming Guide v4.2
- HIP API Guide v4.2:
- HIP-Supported CUDA API Reference Guide v4.2
- HIP FAQ:
AMD OpenMP Compiler

- Compiles C/C++ code with OpenMP “target” pragmas
- LLVM-based clang driver (LLVM 11) – all the source is open!
- Flang - FORTRAN compiler (FORTRAN 2003 standard)
- Links with libomptarget to produce a binary for offloading to the GPU
- OMPD compliant implementation
- Support for ROC-GDB, Totalview, etc.
- File bugs for missing functionality!
AMD OpenMP Compiler

- Need the ‘rocm’ module loaded
- C: $ROCM_PATH/llvm/bin/clang
- Cpp: $ROCM_PATH/llvm/bin/clang++
- F90: $ROCM_PATH/llvm/bin/flang
- More information on the spock user-guide
Tools available today on Spock

- rocm-smi: ROCm System Management Interface
  - Observe power, frequencies, Infinity-Fabric link use
  - Which GPU is running?
- rocprof: command line profiling tool
  - Traces, performance counters, etc.
- Rocgdb: gdb-like debugging on the GPU
  - More than printf()
  - Breakpoints, stacktraces, tracepoints, etc.
Questions?
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.

©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 EPYC™ internal testing as of 02/01/2021, average performance improvement achieved on an EPYC™ 72F3 (8C/8T, 3.7GHz) compared to an EPYC™ 73F2 (8C/8T, 3.7GHz), per-core, single thread, using a select set of workloads including estimated SPECint®2017_base, SPECint®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 7.0.1 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™ OLTP workload comparison 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 SPECint®2017_int_base. The 2P AMD EPYC 7763 has 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/res2020q2/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 for more information.

MLN-008: Results as of 01/28/2021 using SPECint®2017_fp_base. The 2P AMD EPYC 7763 has an an 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 for more information.

MLN-016: Results as of 01/28/2021 using SPECint®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. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation.

MLN-017: Results as of 01/28/2021 using SPECint®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 for more information.

MLN-018: Results as of 02/20/2021 using SPECint®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 for more information.

MLN-040: Based on SPECint®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 (KMP-Q32); 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 using 2P Intel Gold 6258R, https://spec.org/cpu2017/results/res2020q3/cpu2017-20200525-23981.pdf. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See www.spec.org for more information.
ENDNOTES

MLN-0414: 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(KMPF-D3), 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 ACCE. Versus: the current highest score Intel Cascade Lake Refresh server with a score of 329 with a 2P Intel Gold 6258R based server, https://spec.org/cpu2017/results/res2004q3/cpu2017-2000915-25975.pdf. SPEC®, SPECrate® and SPEC CPU® are registered trademarks of the Standard Performance Evaluation Corporation. See www.spec.org for more information.

MLN-0403: 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-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 jobs/day 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 1T8 (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 version 2021 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-... version 2020.0 comparison based on AMD internal testing as of 02/05/2021 measuring the average 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 ~7% more (~1.7x the) per node or ~57% better per core jobs/day performance. Results may vary.

MLN-055: AMD EPYC 7003 CPUs with PCIe4 lanes have 2X the I/O throughput capability 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 + 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®2017fp_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 for more information.

13 Frontier Node Hardware Architecture | May 20th 2021 | Spock Training | AMD Public
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 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, up to 25% higher than the highest posted score 2P EPYC 7282 and 7303 powered servers with SPECrate®2017_int_base scores of 215 and 246 respectively (215×2=430, 246×2=492). 291/2=230.5, 26% more performance. 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 for more information.

MLN-060: EPYC 7643 and 7763 CPU powered 2P servers have measured estimated 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 for more information.


MLN-063K: 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 PC4-32000A-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 for more information.
MLNTO-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 ~$571468, 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 \[1 - (\text{AMD server count} / \text{Intel server count})\]; 25% less space \[1 - (\text{AMD rack count} / \text{Intel rack count})\]; 35% less power \[1 - (\text{AMD power cost} / \text{Intel power cost})\]; providing a 35% lower 4 year TCO \[1 - (\text{AMD TCO} / \text{Intel TCO})\].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. 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.