Open UCX on Arm

Pavel Shamis / Pasha
Principal Research Engineer

© 2018 Arm Limited
Scaling

HPE Apollo 70: This dense, scalable platform and HPE’s first ARM-based HPC system brings more choice and flexibility to HPC customers. It provides easy access to HPC technology with support for standard HPC processing, cluster management and performance software. The Apollo 70, using Cavium’s 64-core ARM® A72™ ThunderX²® Server Processor, is purpose-built for memory intensive HPC workloads and delivers up to 23 percent more memory bandwidth than today’s industry standard servers. The Apollo 70 also provides access to HPE’s partnership ecosystem delivering key HPC components including Intel® Enterprise Lustre®, SGI® Linux® Enterprise Server for ARM, and HIPAC-1000x power-optimized ARM® SoC processors.

Announcing the GW4 Tier 2 HPC service, “Isambard” named after Isambard Kingdom Brunel.
Open UCX on Arm

- Works seamlessly on Arm v8
- uArch level optimizations for ThunderX2 and other Arm implementations
- Optimized for NEON SIMD
- Optimized memory pulling routines for Arm
- Tested with multiple Arm v8 implementations
- Tested with InfiniBand, ROCE, TCP, shared memory
Hardware Software Stack Example

Applications

Open MPI, MPICH, SHMEM (OSHMEM, OSSS)

Open UCX API

XPMEM

Verbs API

Arm

InfiniBand
**Testbed**

- HPE Comanche platform with Cavium ThunderX2 (pre-production)
- Mellanox ConnectX-4 EDR (100Gb/s) HCA
- Mellanox 100Gb/s Switch
- OpenMPI 3.0 with OpenUCX 1.3 (MLNX_OFED_LINUX-4.3-1.0.1.0)
- OSU Micro-Benchmarks
Cavium Thunder X2 SINGLE core InfiniBand Bandwidth

Higher is better
Cavium Thunder X2 MPI Ping-Pong Latency with InfiniBand

Lower is better
Cavium Thunder X2 MPI Message Rate with InfiniBand (28 cores)

Higher is better
SHMEM_WAIT()

Latency of 'put' ping-pong with shmem_wait()

- ARM Optimized shmem_wait()
- Regular shmem_wait()

Latency (microseconds)

Total Benchmark Instruction count

- ARM Optimized shmem_wait()
- Regular shmem_wait()

Instruction count

Total Benchmark Cycle count

- ARM Optimized shmem_wait()
- Regular shmem_wait()

Cycle count

- 73% increase
- 35% increase