Overview of Tightly Coupled Accelerators (TCA) Architecture and HA-PACS/TCA

Tightly Coupled Accelerators (TCA) is a proof-of-concept GPU cluster based on TCA architecture equipped with not only TCA Interconnect but also InfiniBand QDR.

Performance of TCA Communication

CPU: CPU-to-CPU neighbor communication, GPU: GPU-to-GPU neighbor communication

MV2: MVAPICH2, MV2GDR: with GDR

HA-PACS/TCA Specification

Computation Node: CRAY 3623G4-SM
Motherboard: SuperMicro X10DRS-QF
CPU: Intel Xeon E5-2680 V2 (2.8 GHz, 10 core) x 2 socket
Memory: DDR3-1866MHz 4ch. 128 GB (19.4 GB/s)
Peak Performance: 448 GFLOPS/node
GPU: NVIDIA Tesla K20x x 4 GPU
Memory: GDDR5 2600MHz, 6 GB/GPU (250 GB/s/GPU)
Peak Performance: 5.24 TFLOPS/node
Interconnect: IB QDR x 2 rails ( Mellanox Connect-X-3)
TCA Interconnect: PEACH2 ( FPGA: Altera Stratix IV 530GX)

Himeno Benchmark Results

In 2 x 2 case, TCA improves the elapsed time 2x2.

TCA demonstrates good strong-scaling results compared with MPI thanks to the small communication overhead especially in the case of small dataset.

In 2 x 2 case, TCA improves the elapsed time with 38% and the performance with 62%.

Entire HA-PACS System Including

HA-PACS/TCA (5 racks x 2 rows) [Located at Univ. of Tsukuba]