Mellanox ScalableSHMEM

Support the OpenSHMEM Parallel Programming Language over InfiniBand

The SHMEM programming library is a one-side communications library that supports a unique set of parallel programming features including point-to-point and collective routines, synchronizations, atomic operations, and a shared memory paradigm used between the processes of a parallel programming application.

SHMEM Details

There are two types of models for parallel programming.   The first is the shared memory model, in which all processes interact through a globally addressable memory space. The other is a distributed memory model, in which each processor has its own memory, and interaction with another processors memory is done though message communication.   The PGAS model, or Partitioned Global Address Space, uses a combination of these two methods, in which each process has access to its own private memory, and also to shared variables that make up the global memory space.

SHMEM, which stands for SHared MEMory, uses the PGAS model to allow processes to globally share variables by allowing each process to see the same variable name, but each process keeps its own copy of the variable.   Modification to another process address space is then accom-plished using put/get (or write/read) semantics.

The ability of put/get operations, or one-sided communication, is one of the major differences between SHMEM and MPI (Message Passing Interface) which only uses two-sided, send/receive semantics.

Mellanox ScalableSHMEM

Mellanox ScalableSHMEM 2.0 is based on the

API defined by the OpenSHMEM.org consortium. The library works with the OpenFabrics RDMA for Linux stack (OFED), and also has the ability to utilize Mellanox Messaging libraries (MXM) as well as Mellanox Fabric Collective Accelera-tions (FCA), providing an unprecedented level of scalability for SHMEM programs running over InfiniBand.
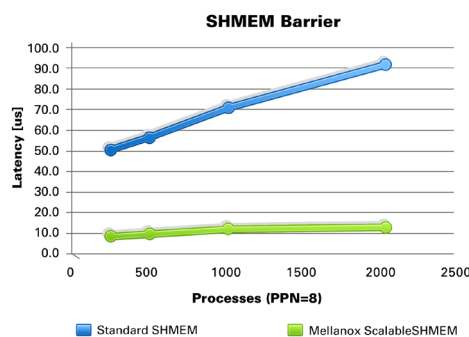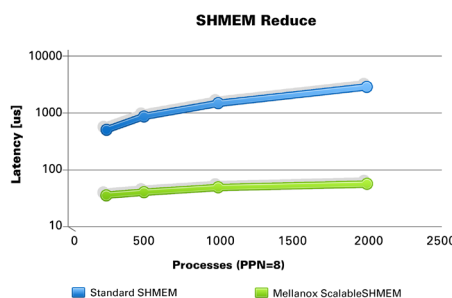

Figure 1.

Figure 2.

HIGHLIGHTS

FEATURES
– Use of symmetric variables and one-sided communication (put/get)
– RDMA for performance optimizations for one-sided communications
– Provides shared memory data transfer operations (put/get), collective operations, and atomic memory operations
– Hybrid model allows use of message passing and shared memory (MPI and SHMEM)
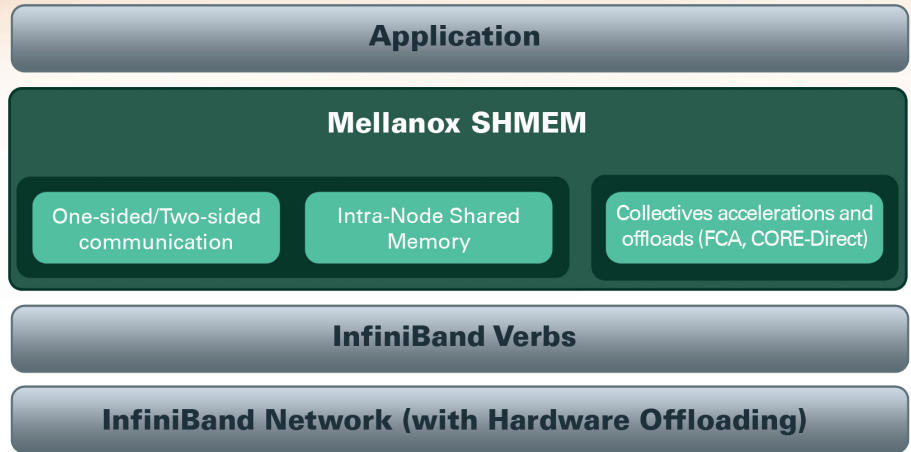
BENEFITS
– Provides a programming library for shared memory communication model extending use of InfiniBand to SHMEM applications
– Seamless integration with MPI libraries and job schedulers
– Improve collective scalability through integrations with Mellanox Fabric Collective Accelerator (FCA).
– Provides maximum performance and scalability SHMEM solution

In addition, the Mellanox MXM acceleration library allows for a high level of performance and scalability for the underlying put/get messages that SHMEM uses for its node-node communications.   MXM is integrated with the MLNX_OFED software stack, and ScalableSHMEM will automatically take advantage of the performance improvements offered by it when MXM presence is detected.   To further understand how these various pieces fit into the software stack refer to Figure 3.

Mellanox Advantage

Mellanox Technologies is a leading supplier of end-to-end servers and storage connectivity solutions to optimize high performance computing performance and efficiency. Mellanox InfiniBand adapters, switches, and software are powering the largest supercomputers in the world. For the best in server and storage performance and scalability with the lowest TCO, Mellanox interconnect products are the solution.

Figure 3.