WHITE PAPER
Internet Wide Area RDMA Protocol (iWARP)
NetEffect 10 Gbps Ethernet
Server Cluster Adapters
Technical and High-Performance Computing

Intel® iWARP Quantum ESPRESSO Performance Study

Julie Cummings
Intel Corporation

EXECUTIVE SUMMARY

The *Intel® iWARP Ethernet Performance Series* of test results intends to help close the gap between real user requirements and the micro-benchmarks promoted by other RDMA vendors. Each paper in the series demonstrates the real-world performance of Intel iWARP on an industry standard application.

This paper reports on Quantum ESPRESSO performance testing performed by the Research Computing and Cyberinfrastructure unit of Information Technology services at Penn State.

Introduction

RDMA enables direct, zero-copy data transfer between RDMA-capable server adapters and application memory, removing the need in Ethernet networks for data to be copied multiple times to operating system data buffers. The mechanism is highly efficient and eliminates the associated processor-intensive context switching between kernel space and user space. HPC applications can therefore reduce latency and perform message transfer very rapidly and consistently by directly delivering data from application memory to the network.

Both iWARP and InfiniBand use RDMA and a common API for HPC applications, however iWARP enables the use of RDMA over the familiar Ethernet fabric. Because iWARP runs over Ethernet TCP/IP, it enables both application and management traffic to operate over a single wire.

This paper reports on Quantum ESPRESSO performance testing performed by the Research Computing and Cyberinfrastructure unit of Information Technology services at Penn State to identify how well iWARP fabrics support workloads on widely used high performance computing applications compared to InfiniBand.

iWARP Features

Unlike InfiniBand, iWARP is an extension of conventional Internet Protocol (IP), so standard IT management tools and processes can also be used to manage the traffic and resources associated with iWARP, which implements the following key performance features:

- **Kernel-Bypass:** Enabling applications to interface directly to the Ethernet adapter removes the latency of the OS and the expensive CPU context switches between kernel-space and user-space.
- **Direct Data Placement**: Writing the data directly into user space eliminates the need for wasteful, intermediate buffer copies, thus reducing processing latency and improving memory bandwidth.
- **Transport Acceleration:** The TCP/IP and iWARP protocols are accelerated in silicon vs. host software stacks, thereby freeing up valuable CPU cycles for application compute processing.

iWARP Benefits

HPC applications can use iWARP technology with NetEffect™ Ethernet Server Cluster Adapters from Intel to provide a high-performance, low-latency Ethernet-based solution. By making Ethernet networks suitable for these high-performance clustering implementations, iWARP provides a number of benefits:

- **Fabric consolidation.** With iWARP technology, LAN and RDMA traffic can pass over a single wire. Moreover, application and management traffic can be converged, reducing requirements for cables, ports, and switches.
- **IP-based management.** Network administrators can use standard IP tools to manage traffic in an iWARP network, taking advantage of existing skill sets and processes to reduce overall cost and complexity.
- **Native routing capabilities.** Because iWARP uses Ethernet and the standard IP stack, it can use standard equipment and be routed across IP subnets using existing network infrastructure.
- **Existing switches, appliances, and cabling.** The flexibility of using standard TCP/IP Ethernet to carry iWARP traffic means that no changes are required to Ethernet-based network equipment.

iWARP vs. Infiniband Performance: Quantum ESPRESSO

Quantum ESPRESSO (opEn Source Package for Research in Electronic Structure, Simulation, and Optimization) is an integrated suite of computer codes for electronic-structure calculations and materials modeling at the nanoscale. It is based on density-functional theory, plane waves, and pseudopotentials (both norm-conserving and ultrasoft). It is freely available under the terms of the GNU General Public License*.

The package builds onto newly restructured electronic-structure codes (PWscf, PHONON, CP90, FPMD, Wannier) that have been developed and tested by some of the original authors of novel electronicstructure algorithms—from Car-Parrinello molecular dynamics to density-functional perturbation theory—and applied in the last twenty years by some of the leading materials modeling groups worldwide. The Quantum ESPRESSO distribution consists of a "historical" core set of packages and a set of plug-ins that performs more advanced tasks.

Quantum ESPRESSO is an initiative of the DEMOCRITOS National Simulation Center,8 (Trieste) and SISSA9 (Trieste), in collaboration with the CINECA National Supercomputing Center in Bologna,[1] the Ecole Polytechnique Fédérale de Lausanne,[2] the Université Pierre et Marie Curie,[3] Princeton University,[4] Massachusetts Institute of Technology,[5] and Oxford University.[6]

For more information, see the Quantum ESPRESSO website.[7]

Test Scenario

To broadly test the performance of iWARP and InfiniBand with Quantum ESPRESSO, the following workloads were developed, results for each of which are presented later in this section:

- Ausurf112 (DEISA benchmark): Self-consistent cycle of the surface of Ag made of 112 atoms, using a cut-off of 25 Ry and four kpoints (2x2x1).
- AI256 (1 kpoints): Two steps of molecular dynamics of liquid Al made of 256 atoms using a cutoff energy of 50 Ry and the gamma point.
- AI256 (8 kpoints): Identical to the previous description except using eight kpoints (2x2x2).

Conclusion

By providing realistic application performance instead of micro-benchmark test results, this report illustrates the danger of relying solely on synthetic benchmarking when evaluating networking options. HPC workloads behave much differently than, for example, a half round-trip latency test: using multiple connections in a switched environment with non-uniform I/O patterns.

The real-world application results shown in this report show the viability of iWARP Ethernet as an alternative to discrete, proprietary fabrics for HPC workloads. The fundamental advantages of a converged Ethernet network combined with easier IP-based management and native routing capability make iWARP a compelling solution for HPC use cases.

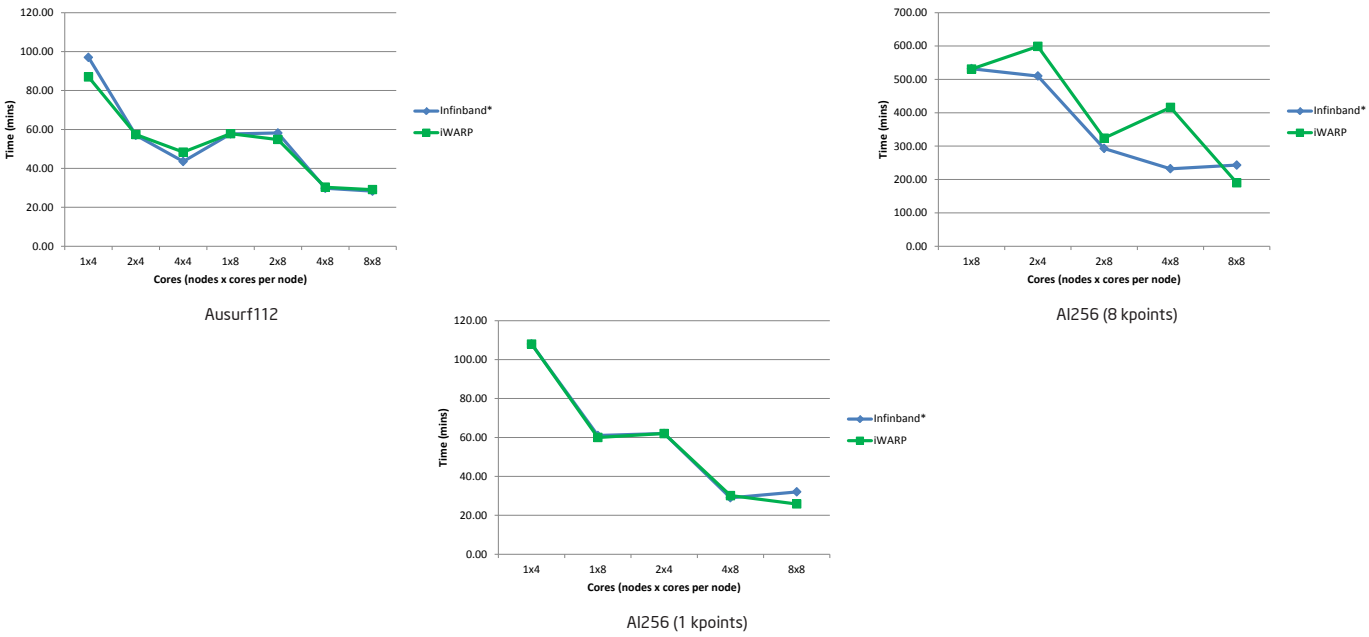

Figure 1. Quantum Espresso iWARP versus InfiniBand* Performance-Testing Results (lower y-axis figures are better)

2

References

[1] http://www.cineca.it/

[2] http://www.epfl.ch/

[3] http://www.impmc.upmc.fr/

[4] http://www.princeton.edu/

[5] http://web.mit.edu/

[6] http://www.materials.ox.ac.uk/

[7] http://www.quantum-espresso.org/

TEST ENVIRONMENT

All tests were performed by the Research Computing and Cyberinfrastructure unit of Information Technology services at Penn State.

The application software under test was Quantum ESPRESSO. (The results are shown in Figure 1.)

The test environment consisted of the following:

Servers

- Dell PowerEdge* R710 Server
- Two Intel® Xeon® processors X5560
- 48 GB RAM

Network Adapters

- 10 Gbps iWARP-enabled NetEffect Ethernet Server Cluster Adapter from Intel
- Mellanox Connect-X MT26428 QDR InfiniBand* Host Channel Adapter

System Software

- Red Hat Enterprise Linux* 5.6
- OpenFabrics Enterprise Distribution* 1.5.2
- OpenMPI 1.4.2

Switches

- iWARP: Arista 7148SX* with Jumbo Frames enabled
- InfiniBand: Mellanox MTS3600*

For more information on Intel® iWARP, please visit:

www.intel.com/go/ethernet

Printed in USA 0412/PDF/JC/SWU Please Recycle 327234-001US