CASE STUDY
Intel® Ethernet Converged Network Adapters
Low-Latency, High-Throughput Computing

Solutions for Ultra-low Latency, High-throughput Computing

MCORELAB

Specialized computing tasks often demand network and I/O performance beyond that of typical data center applications. Workloads such as high-frequency trading require ultra-low levels of latency and jitter. High-performance computing (HPC) requires high throughput and low latency. Today's network stacks based on Linux* and Microsoft Windows* are generally not well optimized to meet these requirements. Intel® Ethernet Controllers and network adapters have been well received in the financial services and HPC communities for their exceptional reliability, performance, and I/O virtualization capabilities. This report shows how financial services and HPC customers can benefit from the use of Intel Ethernet controllers and adapters in low-latency and high-performance applications.

Mcorelab has reported latency results using its MCoreRT* parallel processing software platform that are very close to the theoretical minimum latency of the Intel® Ethernet Converged Network Adapter X520. A team at Intel's Jones Farm Performance Lab undertook testing to verify those claims using the standard methods and processes that Intel applies to quantify the performance of Intel® Ethernet products. MCoreRT incorporates the following optimizations to support ultra-low latency computing on Intel® architecture-based hardware:

- **MCoreRT's kernel-bypass network stack** allows applications to directly access the network hardware. This capability avoids the overhead of the scheduler and other OS mechanisms, providing for low latency and high throughput.
- **MCoreRT's scalable I/O and event-processing system** works to optimize the Intel® platform by scalable utilization of multi-core resources, seamlessly feeding I/O and event streams to the processing cores and providing ultra-fast and scalable event processing and inter-processor communication to applications.
- **Processor quiescence analytics** proactively analyze all available processor cores within the system to help determine which core or cores are experiencing the least noise and interrupts. MCoreRT helps to identify and assign threads to specific processor cores on that basis, optimizing latency and throughput.
- **Optimized interrupt servicing** manages the order and timing of interrupt handling to reduce the impact of those events on network latency. The resulting predictability and consistency provide better overall support for real-time operations by reducing jitter.

The MCoreRT parallel processing software platform uses these capabilities to provide the ultra-low latency, minimal jitter, and high throughput needed to meet real-time processing demands.

This approach provides excellent performance on both Windows and Linux. Applications can take advantage of the hardware to deliver high throughput using MCoreRT's advanced parallel I/O and event-processing architecture. Alternatively, applications can use MCoreRT's kernel-bypass low-latency network stack in transparent mode, which provides unobtrusive latency improvements, requiring no modifications to application software except processor affinity binding.

Extremely Low Latency in the Standards-based Enterprise

Mcorelab, Inc. is the maker of the MCoreRT* parallel processing software platform, which is a kernel-bypass I/O and event-processing software stack for the multi-core environment of today's open-standards servers. This study reports on success using MCoreRT to optimize Intel® architecture solution stacks for extremely low latency, high packet throughput, and minimal jitter. These capabilities can dramatically improve results for usage models such as real-time transactional computing, security appliances, business intelligence, and algorithmic securities trading.

Testing Environment and Procedure

The Intel LAN Access Division performance team developed its test scenario around the MCoreRT solution and the best-selling Intel Ethernet Converged Network Adapter X520. The methodology specified that Spirent TestCenter* (v3.90) be connected directly to the adapter and server under test. In both Windows and Linux testing, non-required services were disabled; in Linux testing, for example, the User Space interrupt request (IRQ) balancer was disabled to keep IRQs static on the cores. The core affinity was determined by the methodology described below.

The Processor Quiescence Analytic tool measures the IRQ preemption over time to determine the CPU "noise" levels. The team used this tool to determine the most optimal physical core to which it would pin the Mcore stack. This approach was required to avoid disk preemption and reduce maximum latency. Pinning to the least noisy cores would reduce the maximum and average latency results. The team also made the following configurations and settings on the test platform:

- Disabled Intel SpeedStep® technology
- Disabled C3 and C6 states
- Enabled Intel® Hyper-Threading Technology (this setting should be inconsequential because of pinned cores)
- Disabled Intel® Virtualization Technology
- Disabled Intel® Virtualization Technology for Directed I/O
- Set CPU Power and Performance Policy setting to "Performance"

The team's goal was to run the tests they believed were most important in ultra-low latency applications to ascertain how well the combination of the Mcore driver and Intel hardware performs and to identify any limitations inherent in this solution. The system selected was based on the Intel® Server Board S2600GZ, powered by the Intel® Xeon® processor E5-2600 product family. This platform features Intel® Data Direct I/O, which also helps lower latency and improve performance. Further details about the test environment are given in Table 1.

Table 1. Details of the testing environment.

MCoreRT* Limitations – General LAN	▪ Spirent TestCenter* 2544 Latency Testing – TCP ▪ Force burst size of 1 to highlight request pipelining (UDP and TCP) ▪ Find max/min message rates where latency falters, using Spirent throughput tests ▪ netperf (Linux*) UDP_RR, TCP_RR ▪ MCoreRT sliding window (5 to 19 packets) – understand its impact on latency and throughput ▪ Scaling with Intel® Ethernet Server Adapter X520 multiple queues
Packet Capture Traces	Passive capture to analyze MCoreRT request bursting, etc. to hide latency
Configuration for Test	½RTT for UDP using Spirent TestCenter RFC 2544 latency test: 66B, 50 kpps–500 kpps

Test Results

Figure 1 illustrates a significant latency reduction that scales across the important payload sizes. The performance improvement is an optimal configuration across the cores, profiled and optimized largely by the MCoreRT tools.

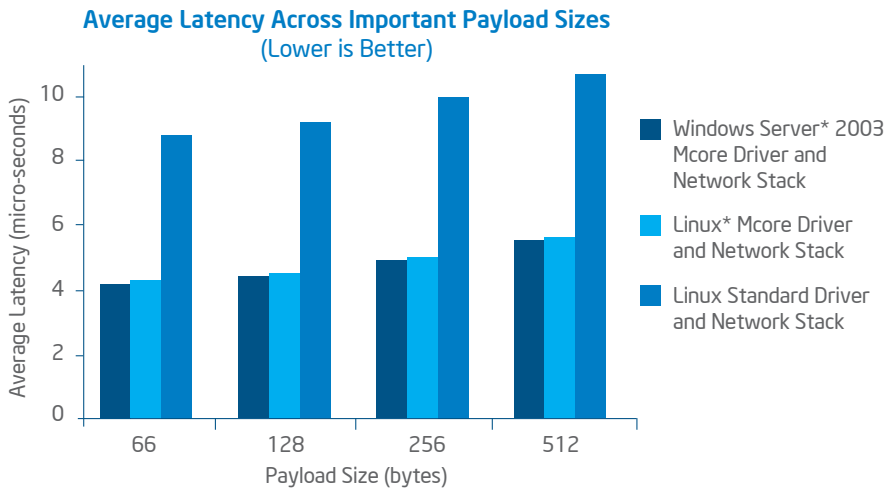

Figure 1. Latency at various common packet sizes on Windows* and Linux*, comparing Mcore drivers to the standard Linux driver.

To test the scalability of the MCoreRT and Intel architecture solution, performance engineers tested the effect on latency as throughput scaled upward, using both TCP and UDP workloads. The testing confirmed that it is indeed possible to get high throughput on both TCP and UDP without sacrificing latency, as illustrated in Figure 2.

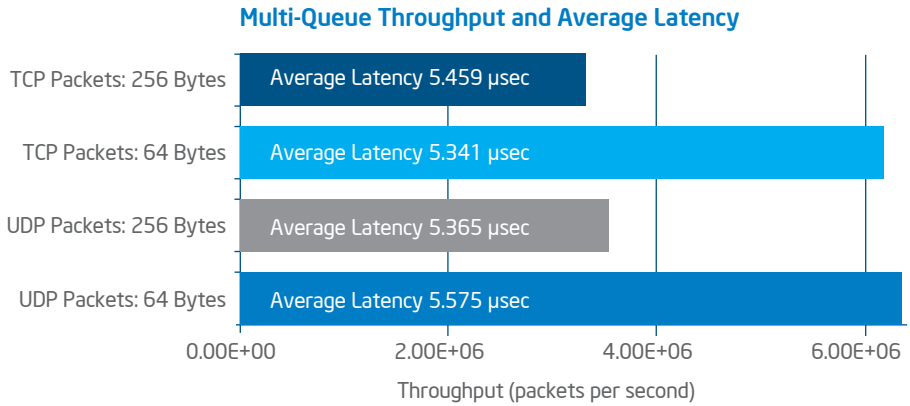

Figure 2. Scalable performance of TCP and UDP.

10 Gigabit Switching for Low Latency

The preferred networking technology for interconnecting highly optimized, latency-sensitive, high performance trading and HPC server nodes is 10 Gigabit Ethernet (10GbE). Operating in cut-through mode, 10GbE switching offers up to a 50X improvement over 100-Megabit and Gigabit switching, with port-to-port latency averaging less than 400 nanoseconds.[1] Equally compelling is the reduction in jitter, with near-flat-line-performance characteristics, irrespective of packet sizes and micro bursts between server nodes within the cluster.[1]

These performance characteristics, which are illustrated in Figure 3, make the Arista Networks 7150 low latency, feature-rich switch, based on the Intel® Ethernet Switch FM600 Series, an excellent complement to the Intel® Ethernet Converged Network Adapter X520 and MCoreRT's low latency network stack.

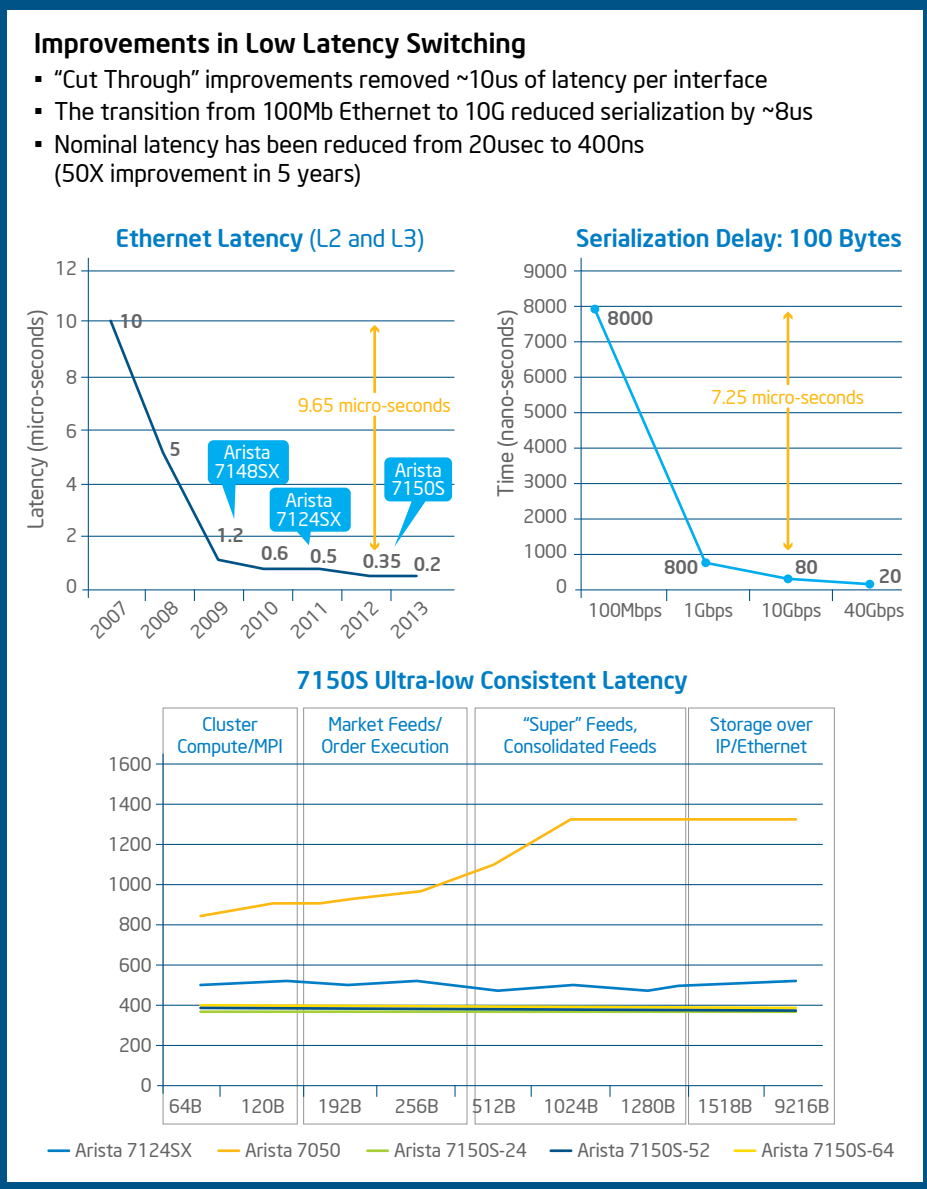

Figure 3. Dramatic latency reductions and low jitter from Arista switches.

Conclusion

These test results from the Intel Performance Lab verify that MCoreRT offers a highly optimized solution that delivers excellent latency and throughput results in combination with Intel architecture-based hardware. Customers that require the ultra-low latency capabilities that this solution provides should investigate evaluation options from Mcorelab and perform their own in-house evaluations of this unique product pairing.

To test MCoreRT in your own environment, register for an evaluation at
www.mcorelab.com/contact.html

SOLUTION PROVIDED BY:

intel® MCORELAB ARISTA

[1] Source: Arista Networks.

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. NO LICENSE, EXPRESS OR IMPLIED, BY ESTOPPEL OR OTHERWISE, TO ANY INTELLECTUAL PROPERTY RIGHTS IS GRANTED BY THIS DOCUMENT. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY, RELATING TO SALE AND/OR USE OF INTEL PRODUCTS INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT OR OTHER INTELLECTUAL PROPERTY RIGHT. UNLESS OTHERWISE AGREED IN WRITING BY INTEL, THE INTEL PRODUCTS ARE NOT DESIGNED NOR INTENDED FOR ANY APPLICATION IN WHICH THE FAILURE OF THE INTEL PRODUCT COULD CREATE A SITUATION WHERE PERSONAL INJURY OR DEATH MAY OCCUR.

Intel may make changes to specifications and product descriptions at any time, without notice. Designers must not rely on the absence or characteristics of any features or instructions marked "reserved" or "undefined." Intel reserves these for future definition and shall have no responsibility whatsoever for conflicts or incompatibilities arising from future changes to them. The information here is subject to change without notice. Do not finalize a design with this information. The products described in this document may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request. Contact your local Intel sales office or your distributor to obtain the latest specifications and before placing your product order. Copies of documents which have an order number and are referenced in this document, or other Intel literature, may be obtained by calling 1-800-548-4725, or by visiting Intel's Web Site www.intel.com.

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark* and MobileMark*, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more information go to www.intel.com/performance.

*Other names and brands may be claimed as the property of others.

Copyright © 2013 Intel Corporation. All rights reserved. Intel, the Intel logo, and Xeon are trademarks of Intel Corporation in the U.S. and other countries.

0213/WM/MESH/PDF 326880-002US