



Maximizing GPU Cluster Performance

TrueScale InfiniBand
Accelerates GPU Clusters

Optimal Performance Per Watt



Executive Overview

The High Performance Computing market's continuing need for improved time-to-solution and the ability to explore expanding models seems unquenchable, requiring ever-faster HPC clusters. This has led many HPC users to implement graphic processing units (GPUs) into their clusters. While GPUs have traditionally been used solely for visualization or animation, today they serve as fully programmable, massively parallel processors, allowing computing tasks to be divided and concurrently processed on the GPU's many processing cores. When multiple GPUs are integrated into an HPC cluster, the performance potential of the HPC cluster is greatly enhanced. This processing environment enables scientists and researchers to tackle some of the world's most challenging computational problems.

HPC applications modified to take advantage of the GPU processing capabilities can benefit from significant performance gains over clusters implemented with traditional processors. To obtain these results, HPC clusters with multiple GPUs require a high-performance interconnect to handle the GPU-to-GPU communications and optimize the overall performance potential of the GPUs. Because the GPUs place significant demands on the interconnect, it takes a high-performance interconnect, such

as InfiniBand, to provide the low latency, high message rate, and bandwidth that are needed to enable all resources in the cluster to run at peak performance.

QLogic® worked in concert with NVIDIA® to optimize QLogic TrueScale™ InfiniBand® with NVIDIA GPU technologies. This solution supports the full performance potential of NVIDIA GPUs through an interface that is easy to deploy and maintain.

Key Points

- Up to 44 percent GPU performance improvement versus implementation without GPUDirect™—a GPU computing product from NVIDIA that enables faster communication between the GPU and InfiniBand
- QLogic TrueScale InfiniBand offers as much as 10 percent better GPU performance than other InfiniBand interconnects
- Ease of installation and maintenance – QLogic's implementation offers a streamlined deployment approach that is significantly easier than alternatives

Ease of Deployment

One of the key challenges with deploying clusters consisting of multi-GPU nodes is to maximize application performance. Without GPUDirect, GPU-to-GPU communications would require the host CPU to make multiple memory copies to avoid a memory pinning conflict between the GPU and InfiniBand. Each additional CPU memory copy significantly reduces the performance potential of the GPUs.

QLogic's implementation of GPUDirect takes a streamlined approach to optimizing NVIDIA GPU performance with QLogic TrueScale InfiniBand. With QLogic's solution, a user only needs to update the NVIDIA driver with code provided and tested by QLogic. Other InfiniBand implementations require the user to implement a Linux® kernel patch as well as a special InfiniBand driver. The QLogic approach provides a much easier way to deploy, support, and maintain GPUs in a cluster without having to sacrifice performance. In addition, it is completely compatible with other GPUDirect implementations; the CUDA libraries and application code require no changes.

Optimized Performance

QLogic used AMBER molecular dynamics simulation software to test clustered GPU performance with and without GPUDirect. The following graph shows that there is a significant performance gain of up to 44 percent that results from streamlining the host memory access to support GPU-to-GPU communications.

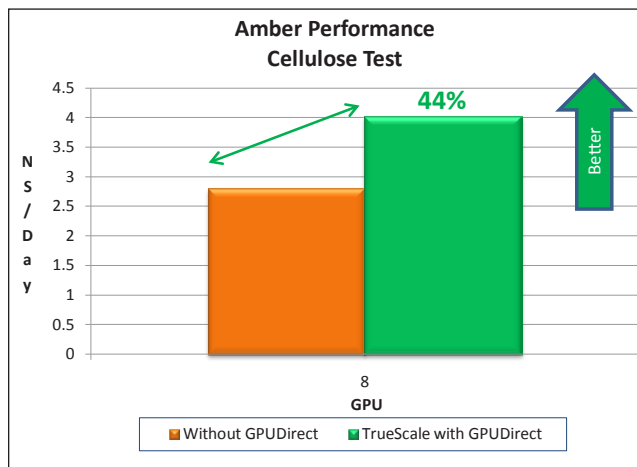


Figure 1. Performance with and without GPUDirect

Clustered GPU Performance

HPC applications that have been designed to take advantage of parallel GPU performance require a high-performance interconnect, such as InfiniBand, to maximize that performance. In addition, the implementation or architecture of the InfiniBand interconnect can impact performance. The two industry-leading InfiniBand implementations have very different architectures and only one was specifically designed for the HPC market—QLogic's TrueScale InfiniBand. TrueScale InfiniBand provides unmatched performance benefits, especially as the GPU cluster is scaled. It offers high performance in all of the key areas that influence the performance of HPC applications, including GPU-based applications. These factors include the following:

- Scalable non-coalesced message rate performance greater than 25M messages per second
- Extremely low latency for MPI collectives, even on clusters consisting of thousands of nodes
- Consistently low latency of one to two μ S, even at scale

These factors and the design of QLogic TrueScale InfiniBand enable it to optimize the performance of NVIDIA GPUs. The following tests were performed on NVIDIA Tesla™ 2050s interconnected with QLogic TrueScale QDR InfiniBand at QLogic's NETtrack Developer Center. The Tesla 2050 results for the industry's other leading InfiniBand are from the published results on the AMBER benchmark site (<http://ambermd.org/gpus/benchmarks.htm>).

Figure 2 shows performance results from the AMBER Myoglobin benchmark (2,492 atoms) when scaling from two to eight Tesla 2050 GPUs. The results indicate that **the QLogic TrueScale InfiniBand offers up to 10 percent more performance than the industry's other leading InfiniBand when both used their versions of GPUDirect.** As the figure shows, the performance difference increases as the application is scaled to more GPUs.

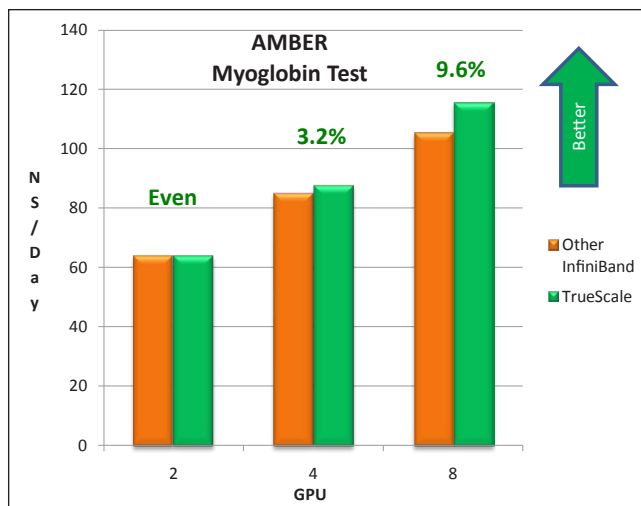


Figure 2. GPU Scalable Performance with the Industry's Leading InfiniBands

The next test shows the impact of the InfiniBand interconnect on the performance of AMBER across models of various sizes. The following Explicit Solvent models were tested:

- DHFR: 23,558 atoms
- FactorIX: 90,906 atoms
- Cellulose: 408,609 atoms

It is important to point out that the performance of the models is dependent on the model size, the size of the GPU cluster, and the performance of the InfiniBand interconnect. The smaller the model, the more it is dependent on the interconnect due to the fact that the model's components (atoms in the case of AMBER) are divided across the available GPUs in the cluster to be processed for each step of the simulation. For example, the DHFR test with its 23,557 atoms means that each Tesla 2050 in an eight-GPU cluster is processing only 2,945 atoms for each step of the simulation. The processing time is relatively small when compared to the communication time. In contrast, the Cellulose model with its 408K atoms requires each GPU to process 17 times more data per step than the DHFR test, so significantly more time is spent in GPU processing than in communications.

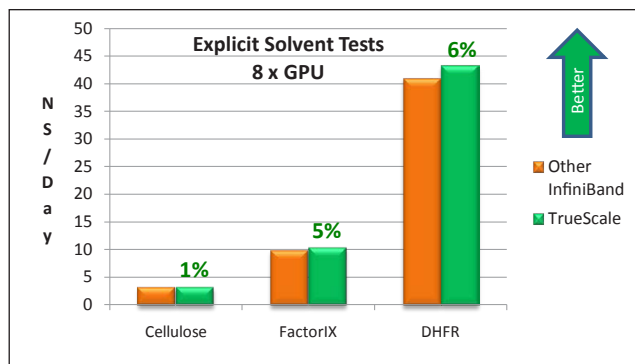


Figure 3. Explicit Solvent Benchmark Results for the Two Leading InfiniBands

The preceding tests demonstrate that the TrueScale InfiniBand performs better under load. The DHFR model is the most sensitive to the interconnect performance, and it indicates that TrueScale offers six percent more performance than the alternative InfiniBand product. Combining the results from Figure 1 and Figure 2 illustrate that TrueScale InfiniBand provides better results with smaller models on small clusters and better model scalability for larger models on larger GPU clusters.

Performance/Watt Advantage

Today the focus is not just on performance, but how efficiently that performance can be delivered.

This is an area in which QLogic TrueScale InfiniBand excels. The National Center for SuperCompute Applications (NCSA) has a cluster based on NVIDIA GPUs interconnected with TrueScale InfiniBand. This cluster is number three on the November 2010 Green500 list with performance of 933 MFlops/Watt (<http://www.green500.org/lists/2010/11/top/list.php>). This on its own is a significant accomplishment, but it is even more impressive when considering its original position on the SuperComputing Top500 list. In fact, the cluster is ranked at #404 on the Top500 list, but the combination of NVIDIA's GPU performance, QLogic's TrueScale performance, and low power consumption enabled the cluster to move up 401 spots from the Top500 list to reach number three on the Green500 list. This is the most dramatic shift of any cluster in the top 50 of the Green500. In part, the following are the reasons for such dramatic performance/watt results:

- Performance of the NVIDIA Telsa 2050 GPU
- Linpack performance efficiency of this cluster is 49 percent, which is almost 20 percent better than most other NVIDIA GPU-based clusters on the Top 500 list
- The QLogic TrueScale InfiniBand Adapter required 25 – 50 percent less power than the alternative InfiniBand product

Conclusion

The performance of the InfiniBand interconnect has a significant impact on the performance of GPU-based clusters. QLogic's TrueScale InfiniBand is designed and architected for the HPC marketplace, and it offers an unmatched performance profile with a GPU-based cluster. Finally, QLogic's solution provides an implementation that is easier to deploy and maintain, and allows for optimal performance in comparison to the industry's other leading InfiniBand.

Disclaimer

Reasonable efforts have been made to ensure the validity and accuracy of these performance tests. QLogic Corporation is not liable for any error in this published white paper or the results thereof. Variation in results may be a result of change in configuration or in the environment. QLogic specifically disclaims any warranty, expressed or implied, relating to the test results and their accuracy, analysis, completeness or quality.



Corporate Headquarters QLogic Corporation 26650 Aliso Viejo Parkway Aliso Viejo, CA 92656 949-389-6000

www.qlogic.com

International Offices UK | Ireland | Germany | India | Japan | China | Hong Kong | Singapore | Taiwan

© 2011 QLogic Corporation. Specifications are subject to change without notice. All rights reserved worldwide. QLogic, the QLogic logo, and TrueScale are trademarks or registered trademarks of QLogic Corporation. GPUDirect, Tesla, and NVIDIA are trademarks or registered trademarks of NVIDIA Corporation. InfiniBand is a registered trademark and service mark of the InfiniBand Trade Association. Linux is a registered trademark of Linus Torvalds. All other brand and product names are trademarks or registered trademarks of their respective owners. Information supplied by QLogic Corporation is believed to be accurate and reliable. QLogic Corporation assumes no responsibility for any errors in this brochure. QLogic Corporation reserves the right, without notice, to make changes in product design or specifications.