Maximizing LS-DYNA® Performance and Scalability with In-Network Computing Acceleration Engines

Ophir Maor, Gerardo Cisneros, David Cho, Yong Qin, Gilad Shainer
HPC Advisory Council

Abstract

The Co-Design Collaboration is a collaborative effort among industry leaders, academia and manufacturers, whose mission is to reach the next level of application performance by exploiting system efficiency and optimizing performance. The above is achieved through creating a synergy between the hardware and the software. One of the major outcomes of this collaboration is In-Network Computing technology. This technology enables data algorithms, traditionally managed by the software on general processors, to be managed and executed by the data center interconnect, utilizing dedicated hardware components. This new approach dramatically improves application performance and overall data center return on investment (ROI). In this paper we describe and test the performance of LS-DYNA, benchmarked over the new architecture, and demonstrate its scaling and efficiency capabilities.

Introduction

High-performance computing (HPC) is a critical tool for automotive design and manufacturing, specifically for computer-aided engineering (CAE): from component-level design to full vehicle analysis of crash simulations, structure integrity, thermal management, climate control, engine modeling, exhaust, acoustics, and much more. HPC helps large enterprises drive faster time-to-market, realize significant cost reductions over laboratory testing and achieve tremendous flexibility. HPC's strength and efficiency hinges on its ability to achieve sustained top performance by driving the CPU performance toward its limits. The motivation for high-performance computing in the automotive industry has long been its tremendous cost savings and product improvements; the cost of a high-performance compute cluster can be just a fraction of the price of a single crash test, and the same cluster can serve as the platform for every test simulation going forward.

The recent trends in high-performance computing cluster environments, ranging from multi-core CPUs, GPUs, and advanced high speed, to low latency interconnect with offloading capabilities, are changing the dynamics of cluster-based simulations. Software applications are being reshaped for higher degrees of parallelism and multi-threading, and hardware is being reconfigured to solve new emerging bottlenecks to maintain high scalability and efficiency. LS-DYNA software from Livermore Software Technology Corporation is a general-purpose structural and fluid analysis simulation software package, capable of simulating complex real-world problems. It is widely used in the automotive industry for the analysis of crashworthiness, occupant safety, metal stress and much more. In most cases, LS-DYNA is used in cluster environments, as they provide better flexibility, scalability, and efficiency for such simulations, allowing for larger problem sizes and speeding up time to results.

LS-DYNA relies on Message Passing Interface (MPI), the de-facto messaging library for high performance clusters that is used for node-to-node inter-process communication (IPC). MPI relies on a fast, unified server and storage interconnect to provide low latency and high messaging rate. Performance demands from the cluster interconnect increase exponentially with scale due, in part, to all-to-all communication patterns. This demand is even more dramatic as simulations involve greater complexity to properly simulate physical model behaviors.

In-Network Computing

The latest revolution in HPC is the effort around the Co-Design collaboration, a collaborative effort among industry thought leaders, academia, and manufacturers to reach Exascale performance by taking a holistic system-level approach to fundamental performance improvements. Co-Design exploits system efficiency and optimizes performance by creating synergies between the hardware and the software components, and between the different hardware elements within the data center.

Co-Design recognizes that the CPU has reached the limits of its scalability, and offers an intelligent network as the new “co-processor” to share the responsibility for handling and accelerating application workloads. By placing data-related algorithms on an intelligent network, one can dramatically improve data center and application performance.

Smart interconnect solutions are based on an offloading architecture, which can offload all network functions from the CPU to the network, freeing up CPU cycles and increasing the system’s efficiency. These kind of interconnect solutions have proven in the past to enable performance leadership and better scalability. With more recent efforts in the Co-Design approach, the interconnect will also include data algorithms that will be managed and executed within the network, allowing users to run data algorithms on the data as it is being transferred within the system interconnect, rather than waiting for the data to reach the CPU. This technology is referred to as In-Network Computing, which is the leading approach to achieve performance and scalability for Exascale systems. In-Network Computing transforms the data center interconnect into a “distributed CPU” and “distributed memory,” overcoming performance walls and enabling faster and more scalable data analysis.

SHARP - Scalable Hierarchical Aggregation and Reduction Protocol

SHARP is a technology that enables data reduction and aggregation operations on the interconnect components. SHARP has been implemented in the latest generation of EDR 100Gb/s InfiniBand solutions. With increases in the amount of data that need to be analyzed and with higher simulation complexity, the traditional concept of analyzing data only on the compute elements has reached a latency wall. Adding more cores to handle the various data reduction or aggregation operations does not result in any performance improvement. SHARP helps to overcome the performance wall by migrating these operations to the network and by performing them while the data is being transferred (Figure 1).

The goal of the In-Network Computing architecture is to optimize completion time of frequently used global communication patterns and to minimize their CPU utilization. The first set of patterns being targeted are global reductions of small amounts of data, including barrier synchronization and small data reductions. SHARP provides an abstraction describing data reduction. The protocol defines aggregation nodes (ANs) in an aggregation tree, which are basic components of in-network reduction operation offloading. In this abstraction, data enters the aggregation tree from its leaves, and makes its way up the tree, with data reductions occurring at each AN, with the global aggregate ending up at the root of the tree. This result is distributed in a method that may be independent of the aggregation pattern. Much of the communication processing of these operations is moved to the network, providing host-independent progress, and minimizing application exposure to the negative effects of system noise. The implementation manipulates data as it traverses the network, minimizing data motion. The design benefits from the high degree of network-level parallelism, with the high-radix InfiniBand switches enabling the use of shallow reduction trees.

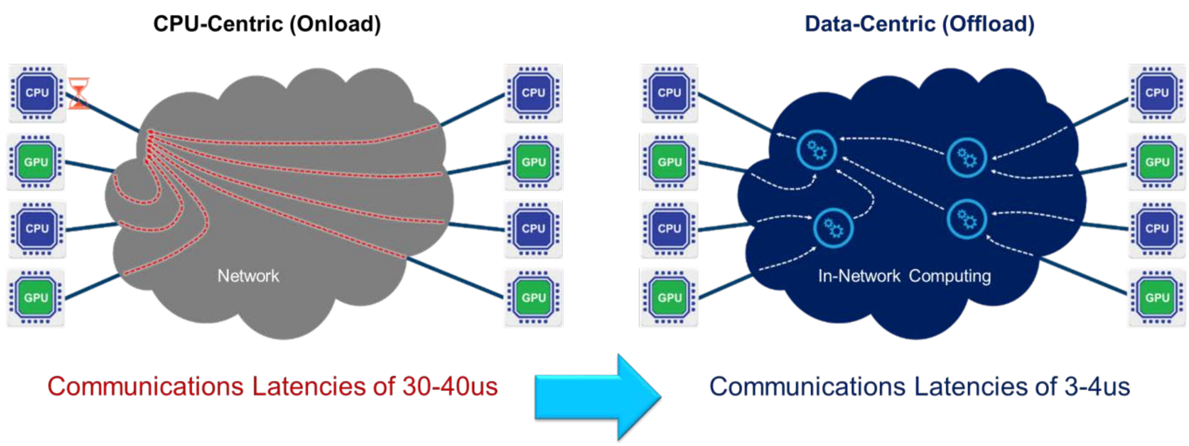

Figure 1 – Illustration of SHARP

Other In-Network Computing elements include interconnect-based, hardware-based MPI tag matching, MPI rendezvous offloads, and more.

LS-DYNA Performance Evaluation with In-Network Computing

The following performance tests were conducted using the resources of the HPC Advisory Council HPC cluster center. We used a cluster based on a combination of Supermicro and Foxconn servers. Each server consists of dual socket Intel(R) Xeon(R) Gold 6138 CPUs at 2.00GHz, Mellanox ConnectX-5 EDR 100Gb/s InfiniBand adapters, Mellanox Switch-IB 2 SB7800 36-Port 100Gb/s EDR InfiniBand switches, memory of 192GB DDR4 2677MHz RDIMMs per node and 1TB 7.2K RPM SSD 2.5" hard drive per node. The LS-DYNA version used was ls-dyna_mpp_s_r10_0_118302_x64, and the operating system was Red Hat Enterprise Linux 7.4.

The first set of tests compared two MPI (message passing interface) libraries – HPC-X and OpenMPI. The HPC-X MPI suite is based on OpenMPI with the addition of the available In-Network Computing technology that is part of the latest EDR InfiniBand solutions. The comparison to OpenMPI was done in order to try and isolate the advantages of the In-Networking Computing elements.

We tested both the Neon_Refined_Revised and 3Cars benchmarks. The performance for these tests is shown in Figure 2 and Figure 3, respectively, both for HPC-X and OpenMPI, as well as a percentage difference line. The unit of performance used for the graphs is jobs per day – how many similar jobs can be executed in a 24-hour period (higher is better).

The results showcase the performance and scalability advantages of In-Network Computing (HPC-X). For the Neon_Refined_Revised benchmark, HPC-X provides up to 130% higher performance versus OpenMPI and better scalability, as OpenMPI does not scale beyond 4 server nodes.

For the 3Cars benchmark case, HPC-X exhibits up to 126% higher performance and better scalability versus OpenMPI, where the latter did not scale beyond 12 server nodes, and exhibits poor scaling when increasing node count up to 12 server nodes.

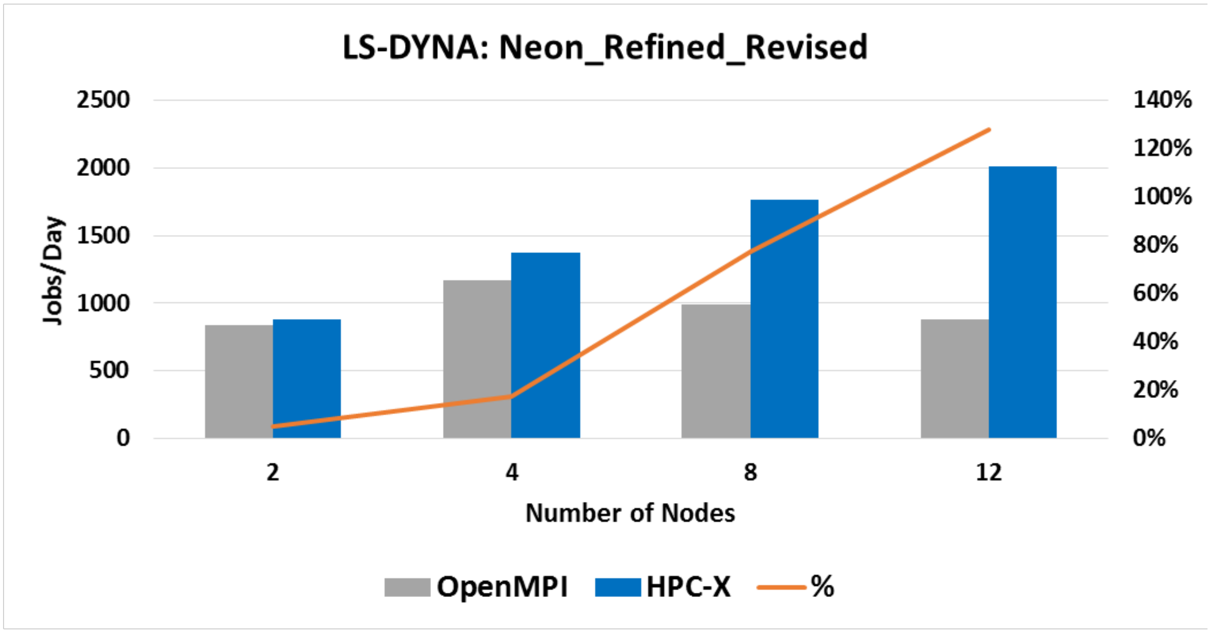

Figure 2 – In-Network Computing Performance Advantages for Neon_Refined_Revised Benchmark

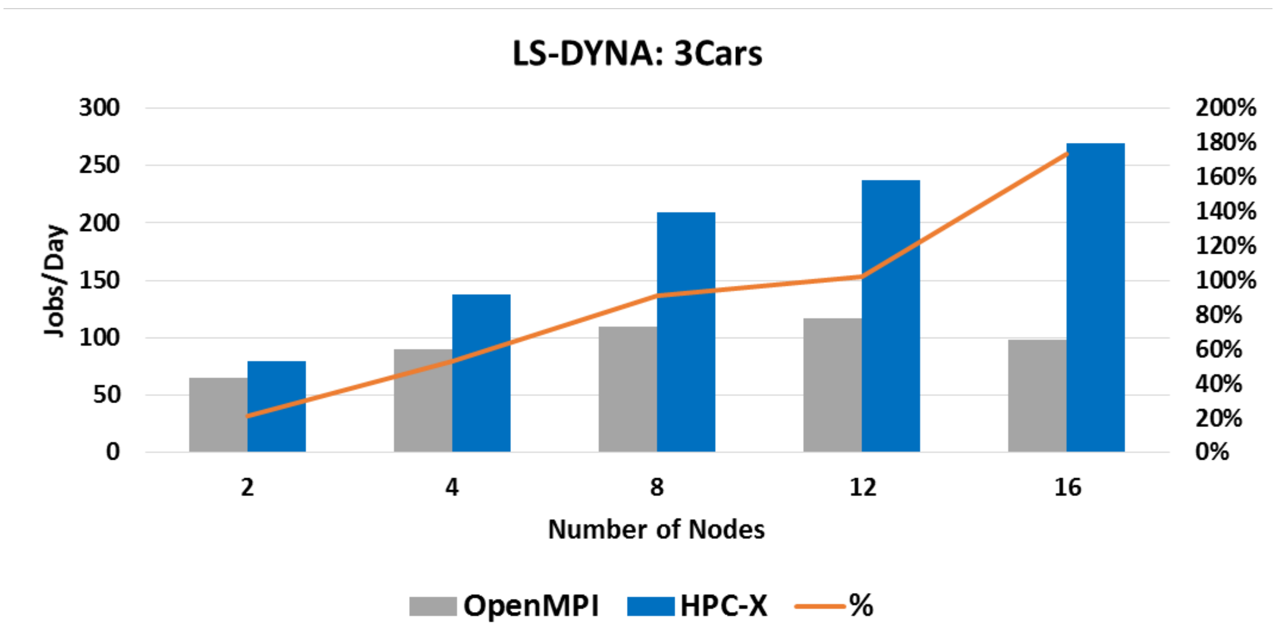

Figure 3 – In-Network Computing Performance Advantages for 3Cars Benchmark

The second part of the testing drew a comparison between InfiniBand and Omni-Path. Omni-Path is an "onload" network architecture, an architecture that relies on the CPU to manage and execute the network operations, while InfiniBand is an "offload" network architecture that manages and executes the network operations at the network level, therefore frees expensive CPU cycles to be used for other applications.

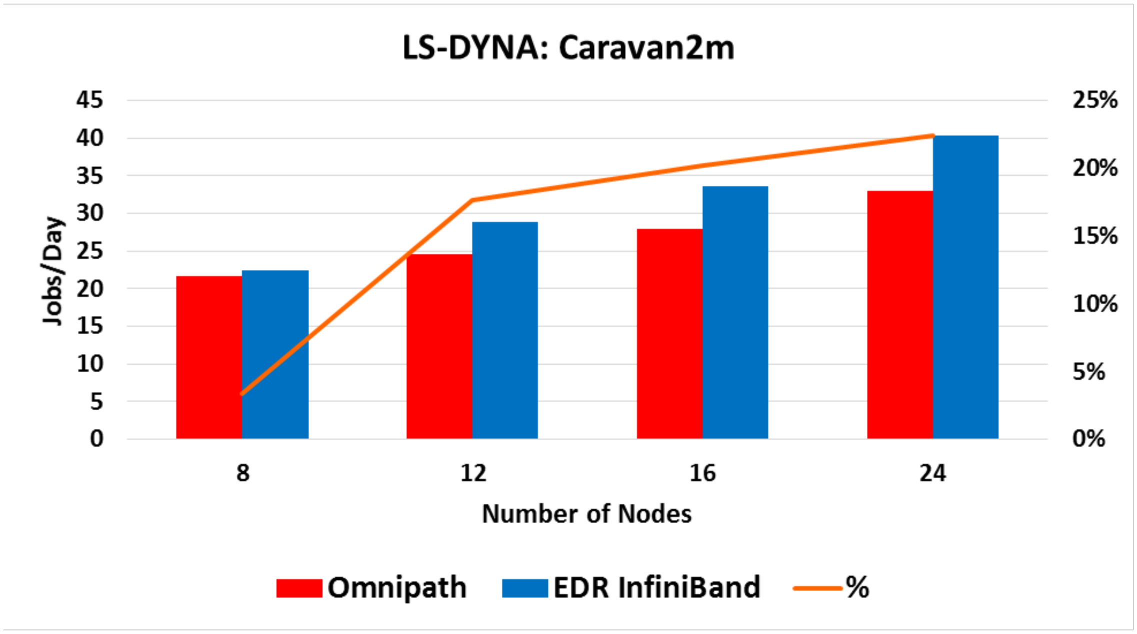

Figure 4 – InfiniBand advantages over Omni-Path for the Caravan2m benchmark

Figure 4 shows the performance results for EDR InfiniBand and Omni-Path, as well as a performance comparison percentage line. The performance metric is again the number of jobs per day (higher is better). The results showcase the advantage of the In-Network Computing technology in providing higher performance and data center efficiency. InfiniBand extracts 23% higher productivity from the cluster, resulting in a higher number of jobs that can be run during a given time period, for an overall better return on investment. As Caravan2m is a benchmark case that demonstrates good scaling capability, we believe that the gap between InfiniBand and Omni-Path will continue to increase with cluster size.

Conclusions

HPC cluster environments impose high demands for connectivity throughput and low latency with low CPU overhead, network flexibility, and high-efficiency. Fulfilling these demands allows to maintain a balanced system that can achieve high application performance and high scaling. With the increase in number of CPU cores and application threads, insimulation complexity and in data volume requiring analysis, there is a need to develop a new HPC cluster architecture—one that will be data-focused instead of the traditional CPU- focused architecture. The Co-Design collaborations enable the development of In-Network Computing technology that breaks the performance and scalability barriers, and move us to the next generation of HPC systems.

Livermore Software Technology Corporation (LSTC) LS-DYNA software was benchmarked for this study to understand the advantages of In-Network Computing technology that was implemented in the latest EDR InfiniBand interconnect. In all cases, LS-DYNA demonstrated higher performance and scalability with EDR InfiniBand.