A Study of LS-DYNA® Implicit Running the Rolls-Royce® Large Representative Engine Model with Intel® Optane™ DC Persistent Memory Technology

Nick Meng[1], Roger Grimes[2], Robert Lucas[2], Francois-Henry Rouet[2], James Ong[3]

1 Intel Corporation
2 Livermore Software Technology (LST), an ANSYS company
3. Rolls-Royce Holdings Corp.

Abstract

In this paper we discuss Intel's continued efforts to optimize the performance of LS-DYNA Implicit. We focus on the Rolls-Royce® Large Representative Engine Model (LREM), the largest implicit model known to Livermore Software Technology. Performance analysis indicated three opportunities for improvement: shared memory parallelization of the LS-GPart reordering code, optimization of the multifrontal linear solver, and usage of Intel® Optane™ DC Persistent Memory. We present results taken while running the LREM model with a tuned hybrid version of LS-DYNA R12.r144413 HYBRID_DP on 2nd Generation Intel® Xeon™ Platinum 8260L scalable processor (formerly Cascade Lake) cluster with Intel's Optane persistent Memory. We depict the benefits of Intel Optane persistent Memory technology and discuss the techniques needed to optimize LS-DYNA.

Introduction

In 2015, Rolls-Royce created a large representative engine model (LREM) with **66 million** finite elements (approximately 200M degrees of freedom) that could be shared with its collaborators [1],[2]. Figure 1 depicts a cross-section of the representative engine model. An initial implicit load calculation with the large (LREM/200M DOFs) model was performed on an internal Rolls-Royce Linux cluster. The initial run was solved using out-of-core solver with MMD [3] reordering method successfully and took **160 hours** to complete with 16 MPI, 14 OMP thread each, on a 16-node, 448-core, Linux cluster.

To address these long run times, Rolls-Royce formed a consortium with Cray, The University of Illinois' National Center for Supercomputing Applications (NCSA), and Livermore Software Technology Company, now Livermore Software Technology, an ANSYS company (LST), to examine the performance of LS-DYNA R11 running the large representative engine model on increasing numbers of processors of the NCSA Blue Waters and the ORNL Titan Cray XE/XK supercomputers. Figure 2 shows the improvement of LS-DYNA's scalability achieved by the Spring of 2019. The minimum elapsed time of the LREM reached was 2625 seconds (0.73 hours) running on 2048 nodes (16,384 cores) of a Cray XE supercomputer**.** The LREM now runs in 5.63 hours using 36 MPI ranks, 14 OMP thread**s** each, on a new Rolls-Royce 36-node, 1008-core, Linux Cluster.

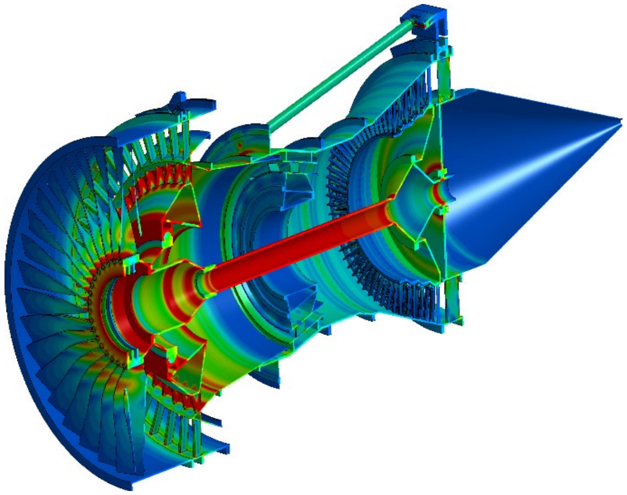
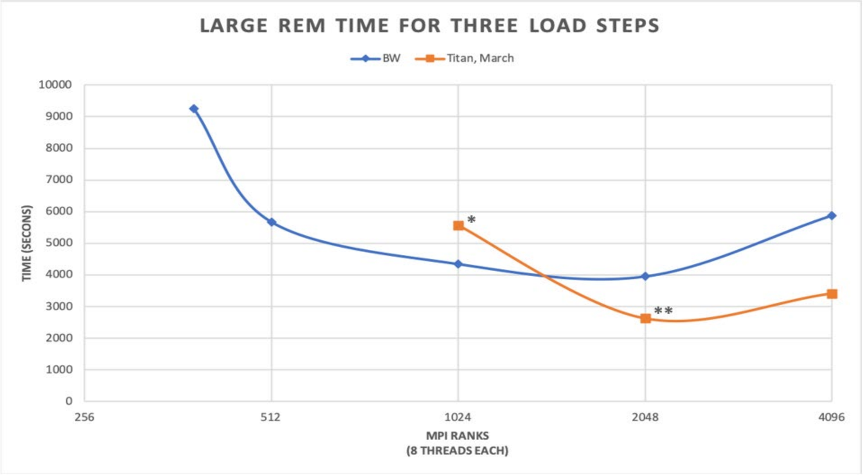


Figure 1: Cross-section of the Rolls-Royce Large Representative Engine Model (LREM)



*Out-of-core, Titan has 32 GB memories

**We believe Titan's workload was less I/O and communication bound than Blue Waters'

Figure 2: Strong scaling experiment, 3 load steps of the LREM/200M DOFs model (Source: [1])

In 2019, Rolls-Royce, LST, and Intel partnered to further optimize the performance of LS-DYNA Implicit on large models such as the LREM, exploring new system solutions based on Intel's latest hardware and software technologies. In this paper, optimization of LS-DYNA R12.144413 Implicit with Intel® Math Kernel Library (Intel® MKL) leveraging Intel® Advanced Vector Extensions AVX 512 (Intel® AVX-512) and improvements to LS-GPart (a major time-consuming area of code) are discussed. Additionally, the performance of LREM/200M DOFs model is evaluated on an Intel® Xeon™ 8260L processor cluster with Intel Optane persistent memory. We demonstrate superior performance when compared to a similar cluster, without the Optane persistent memory, which has to go out-of-core.

Intel® Math Kernel Library

A key optimization strategy for HPC applications on Intel platforms is to leverage the Intel® MKL/BLAS library where possible. Intel® MKL is optimized for Intel® SSE, AVX2 and AVX-512, and can dynamically use the appropriate vector instruction set at run time depending on which Intel Architecture processor it runs on. LS-DYNA R9.1 HYBRID_DP and later versions are linked with an Intel® MKL library supporting for Intel® AVX-512 instructions. To quantify the performance impact of Intel® MKL for LS-DYNA HYBRID_DP Implicit a performance comparison of Intel® MKL using Intel® SSE/AVX2/AVX-512 on 2nd generation Intel® Xeon® Scalable Processor Platinum 8260. LS-DYNA R9.3 HYBRID_DP binary built without Intel® MKL serves as the baseline. LS-DYNA HYBRID_DP Implicit was run with in hybrid mode with 2 MPI ranks and 16 OpenMP threads per rank using 2.5M DOF CYL1E6 model. Figure 3 shows a significant performance improvement for LS-DYNA R9.3 HYBRID_DP Implicit (double precision), 9.9x, using Intel® MKL AVX-512 vs. without Intel® MKL.

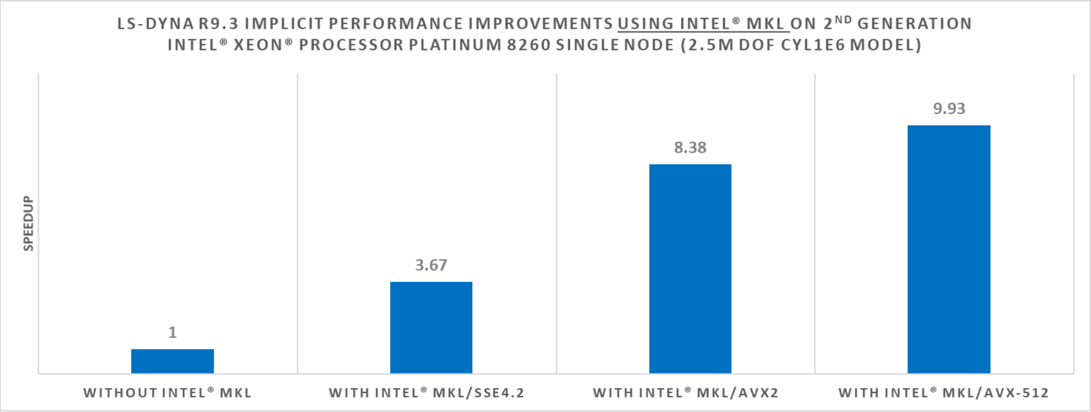


Figure 3: LS-DYNA R9.3 HYBRID_DP Implicit performance impact using Intel® MKL/AVX-512 on 2nd generation Intel® Xeon® Scalable Processor Platinum 8260L
(System configuration: Single node 2-socket Intel® Xeon® Platinum 8260L processor, 384GB DDR4@2933Mhz, single 1.2TB Intel® 3710 SSD, Intel® Turbo Boost enabled, Intel® Hyper-Threading Technology disabled. 2 MPI ranks x 16 OMP threads, Intel® MPI 2018.)

As Figure 3 demonstrates, a similar benefit of AVX-512 is gained running LREM/200M DOFs model with LS-DYNA R11.r139948 HYBRID_DP and R12.r144413 HYBRID_DP on Intel® Xeon™ Platinum 8260 processor cluster with Intel Optane persistent memory.

Intel® Optane™ DC Persistent Memory

Intel® Optane™ DC Persistent Memory is an innovative memory technology that delivers a unique combination of affordable large capacity and persistence (i.e., non-volatility), filling a gap between DRAM and NAND (Figure 4). The Intel Optane persistent memory can be provisioned in two modes. The first is called the Memory Mode. With Memory Mode, applications get a high capacity main memory solution at substantially lower cost and power, while providing performance close to that of DRAM, depending on the workload. No modifications are required to the application. The operating system sees the persistent memory module capacity as system main memory with DRAM memory acting as a cache.

In Memory Mode, even though the media is persistent it will look volatile to application software. The second provisioning mode is called App Direct Mode where applications can access persistent memory capacity as non-volatile memory. In particular, the STorage Over App Direct mode (SToAD) allows access to persistent memory capacity using standard file APIs. No modifications to application code, or the file systems that expect block storage devices is required. SToAD provides high-performance block storage, without the latency of moving data to and from the I/O bus.

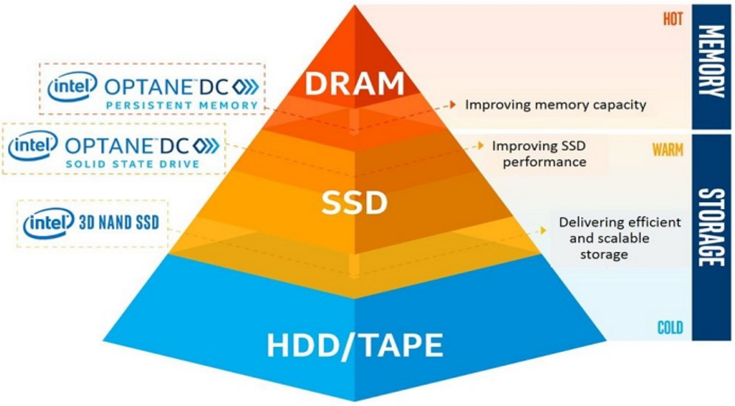

intel OPTANE DC PERSISTENT MEMORY
intel OPTANE DC SOLID STATE DRIVE
intel 3D NAND SSD
DRAM
SSD
HDD/TAPE
Improving memory capacity
Improving SSD performance
Delivering efficient and scalable storage
HOT
WARM
COLD
MEMORY
STORAGE

Figure 4: Storage/memory performance vs. capacity pyramid. Source: Reference [5]

As discussed in the Ashcraft, et.al. [2] paper, LST researchers added dynamic memory management improvements to LS-DYNA R11 Implicit, to reduce its overall memory footprint. Even with these improvements, memory requirements remain very high in large implicit models. I/O performance usually is a serious bottleneck in existing systems when the implicit solver goes out-of-core, even if fast local file system is available. Hence memory capacity is a critical factor for implicit solvers such as LS-DYNA implicit. If an in-core solution is possible, there is no file I/O, resulting in a big performance boost.

Rolls-Royce LREM model performance analysis and optimizations with LS-DYNA R11.r139948 HYBRID_DP

In the first phase of Intel, LST and Rolls-Royce collaboration, we simply evaluated LS-DYNA R11.r139948 HYBRID_DP Implicit on two Intel Xeon 8260L processor (2-socket per node) clusters using the Rolls-Royce LREM model: a 6-node cluster with 6TB Intel Optane persistent memory per node and a 12-node cluster with 1.5TB Intel Optane persistent memory per node. The results are shown in Figure 5. The first bar (20,340 seconds) in the graph is the baseline from Rolls-Royce on their 36-node (2-socket per node) Intel Xeon E5-2695v3 processor (formerly known as Haswell) cluster using in-core solver. The second bar (19,226 seconds) is the run time from the 6-node Intel Xeon 8260L processor (2-socket/6TB Intel Optane persistent memory per node) cluster, also using in-core solver, comparable in performance to the Rolls-Royce baseline. The large 6TB Intel Optane persistent memory per node allowed us to run significantly more MPI ranks per node with much better CPU utilization than the baseline. This configuration also allowed us to determine from the run log file that we roughly needed a total of **12TB** for the LREM problem. The third bar in the graph (9,430 seconds) shows we can scale linearly from 6 to 12-nodes. Overall, the use of 1.5TB Intel Optane persistent memory per node delivered over **2x** performance improvement on 12-nodes compared to Rolls-Royce baseline of 36-nodes.

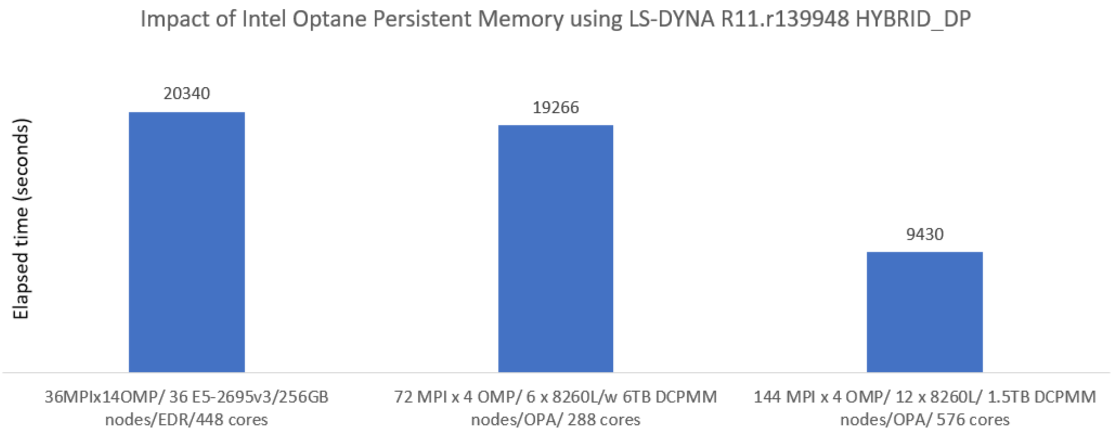


Figure 5: Initial R11.r139948 HYBRID_DP benchmark results of LREM/200M DOFs model on Intel Xeon processor 8260L cluster with Intel Optane persistent memory
(System configuration: 6 node 2-socket Intel® Xeon® Platinum 8260L 24C@2.4Ghz processor, 6TB DCPMM memory, single 1.2TB Intel® 3710 SSD, Intel® Turbo Boost enabled, Intel® Hyper-Threading Technology disabled, dual port OPA-100 network, 72 MPI ranks x 4 OMP threads, Intel® MPI 2018u5. 12 node 2-socket Intel® Xeon® Platinum 8260L 24C@2.4Ghz processor, 1.5TB DCPMM memory, single 1.2TB Intel® 3710 SSD, Intel® Turbo Boost enabled, Intel® Hyper-Threading Technology disabled, dual port OPA-100 network. 144 MPI ranks x 4 OMP threads, Intel® MPI 2018u5)

As noted earlier, two areas of LS-DYNA Implicit take significant % of run time (LS-GPart and factorization). For example, on the 6-node 6TB Intel Optane persistent memory configuration, running 24 MPI ranks x 12 OpenMP threads, LS-GPart and factorization take 35% and 31% of the overall run time respectively, Trading off MPI ranks and OpenMP threads also provided some insights to the behaviors of LS-GPart and factorization (see Figure 6). Factorization did not change much in performance with increased ranks as it is dominated by time in Intel MKL which utilized all the cores fully (ranks x threads). However, LS-GPart improved with increasing ranks since it does not scale well with threads due to for example, serial sections.

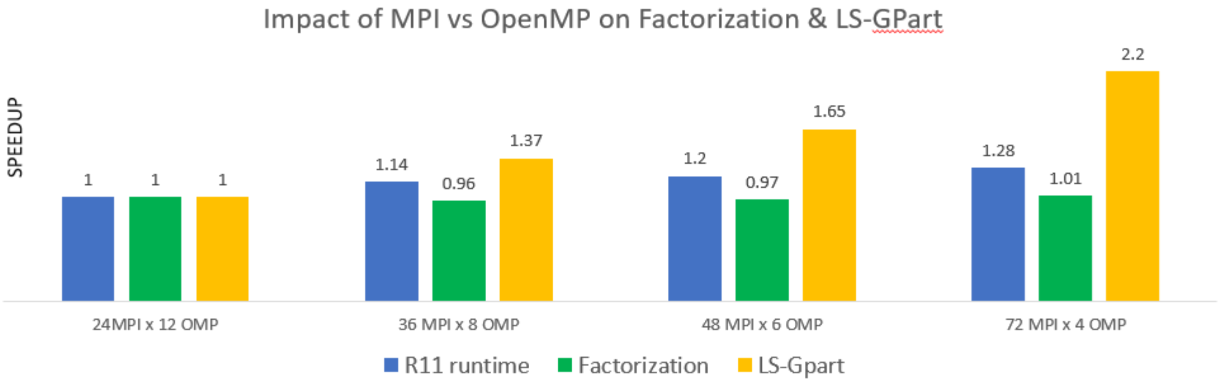


Figure 6: Impact of MPI vs. OpenMP choices on R11.r139948 HYBRID_DP runtime and compute regions (Factorization & LS-GPart)
(System configuration: 6 node 2-socket Intel® Xeon® Platinum 8260L 24C@2.4Ghz processor, 6TB DCPMM memory, single 1.2TB Intel® 3710 SSD, Intel® Turbo Boost enabled, Intel® Hyper-Threading Technology disabled, dual port OPA-100 network)

Rolls-Royce LREM model optimizations with LS-DYNA R12.r144413 HYBRID_DP

In the second phase of Intel, LST and Rolls-Royce collaboration, LST added shared memory parallel version of LS-GPart from R11 to R12. In addition, Intel parallelized and vectorized hot OpenMP loops in Multi Frontal 2 (MF2) solver. We also discovered the impact of CACHE_BLK parameter defined in the Multi Frontal solver in improving Intel MKL performance using Intel AVX-512 (see Figure 7). For example, doubling this parameter from default 128 to 256, improved LREM model by 18%.

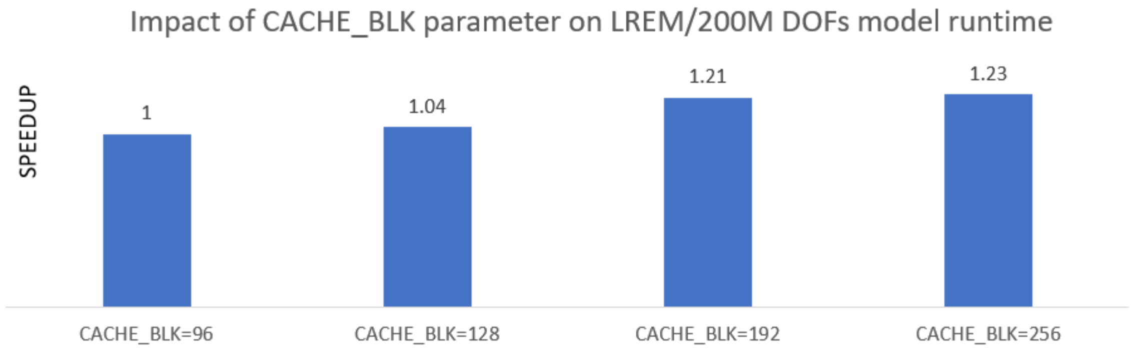


Figure 7: Improving LREM/200M runtime by adjusting CACHE_BLK parameter with LS-DYNA R12.r144413 HYBRID_DP

(System configuration: 6 node 2-socket Intel® Xeon® Platinum 8260L 24C@2.4Ghz processor, 6TB DCPMM memory, single 1.2TB Intel® 3710 SSD, Intel® Turbo Boost enabled, Intel® Hyper-Threading Technology disabled, dual port OPA-100 network)

The results of putting all our optimizations and Intel's new platforms including Intel memory are shown in Figure 8. The first 3 bars are the same as in Figure 5 based on LS-DYNA Implicit R11.r139948 HYBRID_DP. The 4th bar (5,627 seconds) represents an improvement of 68% strictly due to software optimizations in R12.r144413 HYBRID_DP vs. R11.r139948 HYBRID_DP. The final bar (4,640 seconds) shows additional improvements possible by increasing MPI ranks per node resulting in 2x improvement (9,430 seconds vs. 4,640 seconds). In Figure 9, we plot our current best result on the graph published earlier by Ashcraft, et al [2].

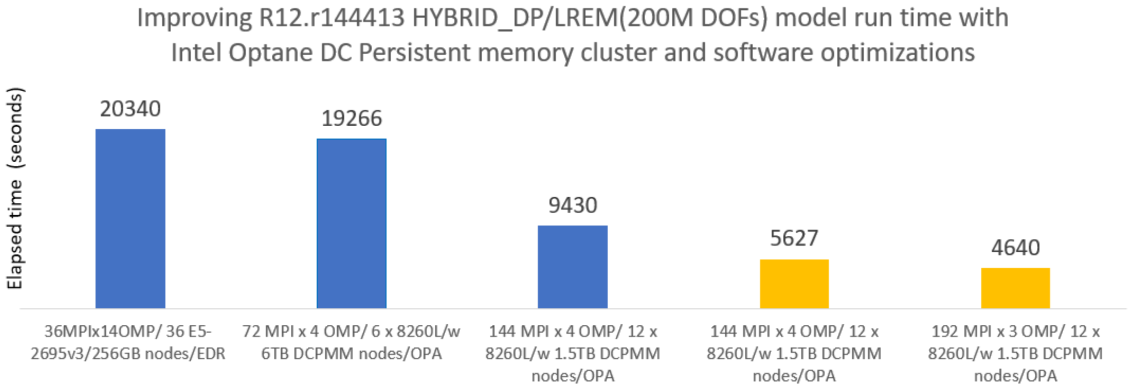


Figure 8: Improving LREM/200M DOFs model runtime using Intel Optane persistent memory cluster and LS-DYNA R12.r144413 HYBRID_DP software optimizations

(System configuration: 6 node 2-socket Intel® Xeon® Platinum 8260L 24C@2.4Ghz processor, 6TB DCPMM memory, single 1.2TB Intel® 3710 SSD, Intel® Turbo Boost enabled, Intel® Hyper-Threading Technology disabled, dual port OPA-100 network, 72 MPI ranks x 4 OMP threads, Intel® MPI 2018u5. 12 node 2-socket Intel® Xeon® Platinum 8260L 24C@2.4Ghz processor, 1.5TB DCPMM memory, single 1.2TB Intel® 3710 SSD, Intel® Turbo Boost enabled, Intel® Hyper-Threading Technology disabled, dual port OPA-100 network. 144 MPI ranks x 4 OMP threads, Intel® MPI 2018u5 and 192 MPI ranks x 3 OMP threads, Intel MPI 2018u5)

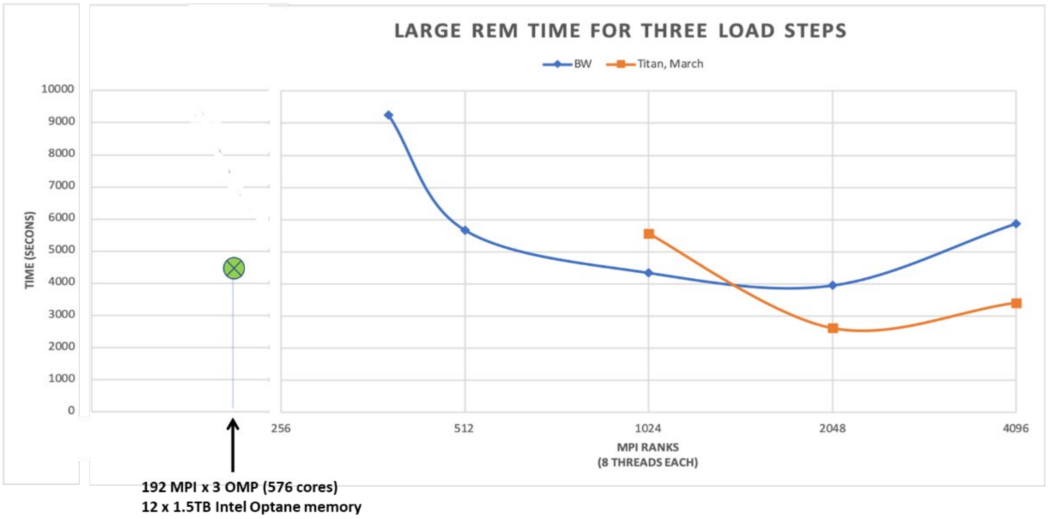


Figure 9: Comparison of LREM/200M R12.r144413 HYBRID_DP performance on Intel Xeon processor 12-node cluster with Intel Optane memory to earlier published performance results
(System configuration: 12 node 2-socket Intel® Xeon® Platinum 8260L 24C@2.4Ghz processor, 1.5TB DCPMM memory, single 1.2TB Intel® 3710 SSD, Intel® Turbo Boost enabled, Intel® Hyper-Threading Technology disabled, dual port OPA-100 network. and 192 MPI ranks x 3 OMP threads, Intel MPI 2018u5)

Conclusion

In this paper we demonstrated the value of the collaboration between Intel, LST, and Rolls-Royce by optimizing LS-DYNA Implicit in order to approach its performance milestone with far fewer compute nodes. This collaboration highlights our software optimization efforts and use of Intel technologies such as Intel MKL using Intel AVX-512, 2nd generation Intel Xeon processors and Intel® Optane™ DC persistent memory. In particular, we discussed how Intel® Optane™ persistent memory could help large memory applications such as LS-DYNA Implicit increase compute density in a node and thereby requiring fewer compute nodes, potentially improving total cost of ownership.

References

[1] Ashcraft C., Dawson J., Grimes, R., Guleryuz E., Koric S., Lucas R., Ong J., Rouet, F-H., Simons T., Zhu T-T., "Increasing the Scale of LS-DYNA Implicit Analysis", 15th International LS-DYNA Users Conference, Dearborn, MI, USA
[2] Ashcraft C., Dawson J., Grimes, R., Guleryuz E., Koric S., Lucas R., Ong J., Rouet, F-H., Simons T., Zhu T-T., "Running Jet Engine Models on Thousands of Processors with LS-DYNA Implicit", 12th European LS-DYNA Conference 2019, Koblenz, Germany
[3] J.W.H.Liu, Modification of the minimum degree algorithm by multiple elimination, ACM Transactions on Mathematical Software, 11(1985), pp 141-153
[4] Ashcraft C., Rouet F.-H., A global, distributed ordering library, SIAM Workshop on Scientific Computing, 2016.
[5] Intel® Optane™ DC Persistent Memory: A Major Advance in Memory and Storage Architecture (https://software.intel.com/en-us/blogs/2018/10/30/intel-optane-dc-persistent-memory-a-major-advance-in-memory-and-storage-architecture

Notices & Disclaimers

All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.

Intel processors of the same SKU may vary in frequency or power as a result of natural variability in the production process.

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 complete information visit www.intel.com/benchmarks.Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

Performance results are based on testing as of dates shown and may not reflect all publicly available security updates. See configuration disclosure for details. No product or component can be absolutely secure.

Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Notice Revision #20110804.
Intel® Advanced Vector Extensions (Intel® AVX)* provides higher throughput to certain processor operations. Due to varying processor power characteristics, utilizing AVX instructions may cause a) some parts to operate at less than the rated frequency and b) some parts with Intel® Turbo Boost Technology 2.0 to not achieve any or maximum turbo frequencies. Performance varies depending on hardware, software, and system configuration and you can learn more at http://www.intel.com/go/turbo.