SOLUTION BRIEF

Simulation
Software Optimization

intel®

Accelerate LS-DYNA Software with the Latest Intel® Technologies

Benchmark studies show that Intel technologies can improve performance in LS-DYNA computations and simulations by up to 9.9x.[1]

Abstract

This paper discusses how Intel and Livermore Software Technology Corporation (LSTC) partnered to optimize LS-DYNA software, and it demonstrates the impact of 2nd Generation Intel® Xeon® Scalable processor technologies for enhancing the performance of LS-DYNA software. The paper focuses on two different approaches for making use of Intel Advanced Vector Extensions 512 (Intel AVX-512) with LS-DYNA:

- LS-DYNA Explicit using Intel compiler vectorization techniques

- LS-DYNA Implicit using Intel Math Kernel Library (Intel MKL) for accelerating dense matrix computational kernels

The paper also explores the numerical accuracy of simulation results for LS-DYNA Explicit, comparing Intel Streaming SIMD Extensions 2 (Intel SSE2) and Intel AVX-512. Finally, the paper examines the benefits of Intel® Optane™ persistent memory (PMem) for LS-DYNA Implicit simulations. The findings in this paper are drawn from TopCrunch benchmarks (ODB-10M and car2car models) for LS-DYNA Explicit and from AWE benchmarks (CYL1E6 and CL2E6 models) for LS-DYNA Implicit.

Introduction

LS-DYNA from LSTC is an advanced finite element analysis (FEA) tool used for simulating complex real-world problems in the automotive, aerospace, construction, manufacturing, and bioengineering industries. LS-DYNA is used in a wide range of engineering applications such as structural analysis, fluid flow, wave propagation, and heat transfer, among many others. The NASA JPL Mars Pathfinder mission, for example, used an LS-DYNA simulation of how the space probe used airbags to aid in its Mars landing.

As problem sizes grow larger, simulations become more complex. The typical problem size of an LS-DYNA Explicit model has grown to more than 10 million elements, and implicit models with hundreds of millions of degrees of freedom (DOF) have been developed for turbine-engine and bioengineering analyses.

LS-DYNA is optimized for scale across multiple cores and shared distributed memory. As models grow larger, the amount of memory and number of cores used for simulation workloads needs to increase to reduce the turnaround time of these simulations. The scaling of LS-DYNA at the lowest cost and highest performance is becoming more important. The process of improving application performance on a single core and then scaling those improvements across many nodes is a major focus area for both Intel and LSTC.

Authors

Nick Meng
Intel

Ram Ramanujam
Intel

Jason Wang
Livermore Software Technology Corporation (LSTC)

Robert Lucas
Livermore Software Technology Corporation (LSTC)

Solution Brief | Accelerate LS-DYNA Software with the Latest Intel® Technologies

The latest Intel technologies can enhance the performance of LS-DYNA, enabling users to get the performance needed for their large analyses. These technologies include:

- 2nd Generation Intel Xeon Scalable processors with Intel AVX-512
- Intel Omni-Path Architecture (Intel OPA) interconnect fabric
- Intel Optane PMem (a new class of persistent memory and storage technology)
- Enhanced Intel Fortran Compiler 2018
- Intel MPI Library 2018
- Intel MKL 2018 with Intel AVX-512 support

This range of technologies enables the most effective integrated scalable solution for LS-DYNA customers, both in cost and performance. These Intel technologies provide performance advantages both for explicit analysis, with Intel AVX-512, and for implicit analysis, with Intel MKL, Intel AVX-512, and Intel Optane PMem.

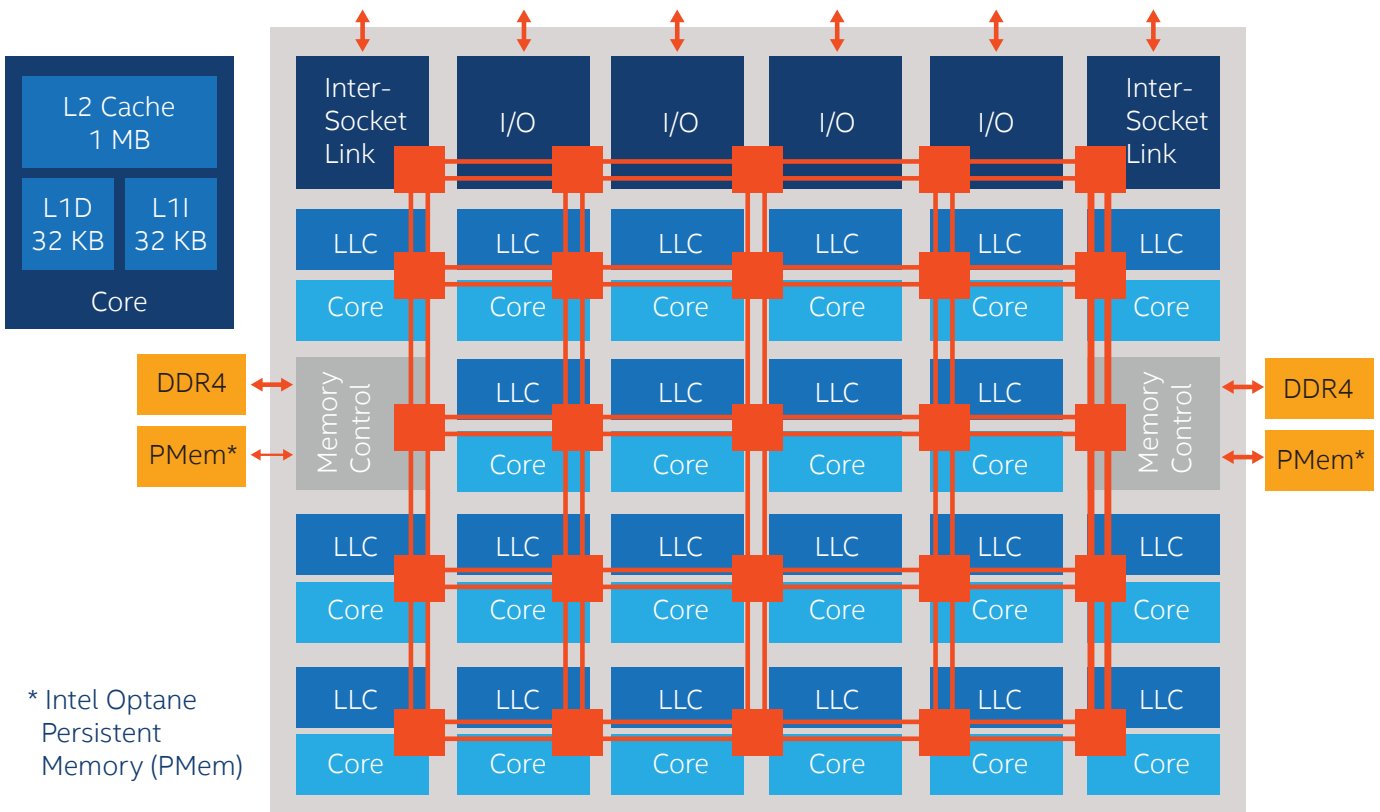## 2nd Generation Intel Xeon Scalable Processors

2nd Generation Intel Xeon Scalable processors introduce many innovative features compared to previous-generation Intel Xeon processors.[2,3] These features include, for example, a higher number of processor cores—up to 56 cores per processor in Intel Xeon Platinum 9200 processors. Other features include higher memory bandwidth (up to 2,933

megatransfers per second [MT/s] in Intel Xeon Platinum 8200 processors, compared to the previous 2,663 MT/s) and up to 12 memory channels in Intel Xeon Platinum 8200 processors compared to the previous 6 memory channels. Also, 2nd Generation Intel Xeon Scalable processors support greater memory capacity (up to 36 TB system memory using Intel Optane PMem), Intel Speed Select Technology (Intel SST) for configurable core/frequency processor attributes, and continued improvements to the Intel AVX-512 instruction set architecture (for example, Vector Neural Network Instructions [VNNI]).

2nd Generation Intel Xeon Scalable processors continue to make use of a mesh interconnect (see Figure 1), introduced in the previous generation, to reduce memory latency.
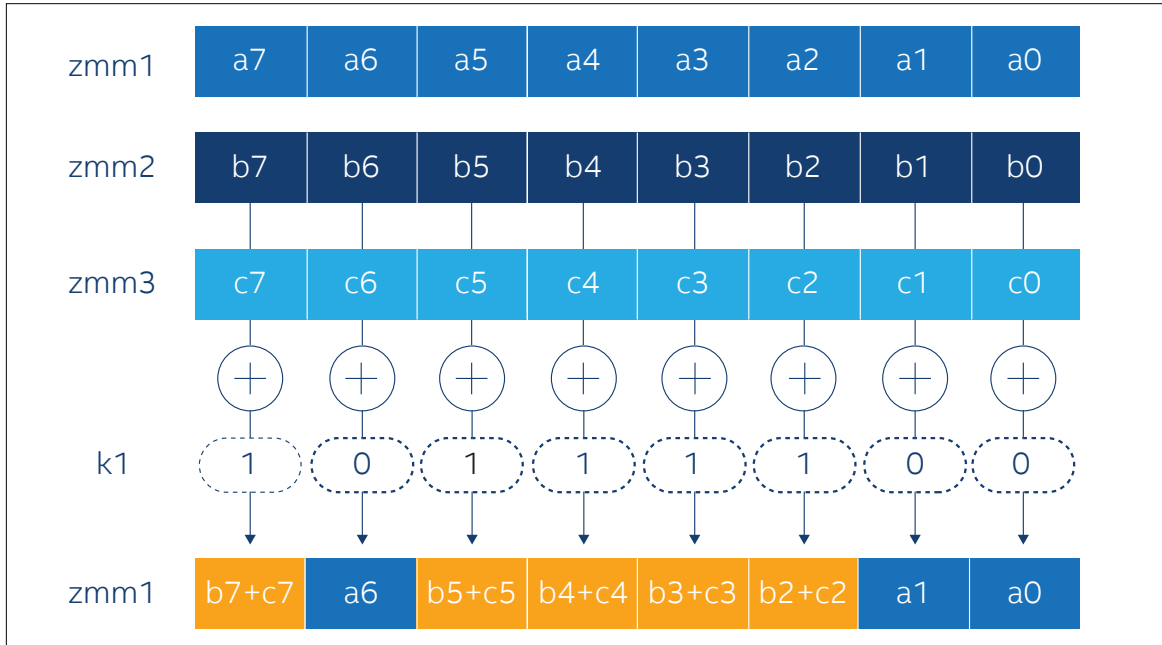
Previous-generation Intel Xeon Scalable processors introduced several Intel AVX-512 instruction groups: Intel AVX-512 conflict detection (Intel AVX-512CD), Intel AVX-512 Foundation (Intel AVX-512F), Intel AVX-512 Byte and Word Instructions (Intel AVX-512BW), Intel AVX-512 Doubleword and Quadword Instructions (Intel AVX-512DQ), and Intel AVX-512 Vector Length Extensions (Intel AVX-512VL).

Intel AVX-512 is an extension to the CPU vector extensions that includes 32 registers, each 512 bits wide, and eight dedicated mask registers that help in efficient Intel AVX-512 instruction generation for branching codes. Intel AVX-512 also doubles the width of the register compared to its predecessor, Intel AVX2, thus computationally doubling the single- and double-precision floating point operations per second (FLOPS) per clock cycle.

Figure 1. A generalized CPU layout and mesh interconnect for Intel Xeon Scalable processors, intended to help illustrate the concepts (not a definitive representation of the microarchitecture)[4]

VADDPD zmm1 {k1}, zmm2, zmm3

zmm1	a7	a6	a5	a4	a3	a2	a1	a0
zmm2	b7	b6	b5	b4	b3	b2	b1	b0
zmm3	c7	c6	c5	c4	c3	c2	c1	c0
	+	+	+	+	+	+	+	+
k1	1	0	1	1	1	1	0	0
zmm1	b7+c7	a6	b5+c5	b4+c4	b3+c3	b2+c2	a1	a0

Figure 2. Vector addition of packed double-precision data elements using masking with the Intel AVX-512 instruction set

Figure 2 shows the addition of eight double-precision data elements in two 512-bit-wide source registers in a single Intel AVX-512 VADDPD instruction with masking. Note that, in contrast, Intel AVX2 and Intel SSE use 256-bit-wide and 128-bit-wide registers, respectively. 2nd Generation Intel Xeon Scalable processors provide improvements to the Intel AVX-512 instruction-set architecture with VNNI. Intel also continues to enhance Intel compiler vectorization technology to make use of Intel AVX-512 instructions, which provides benefits for applications like LS-DYNA.[2]

## Performance Improvements for LS-DYNA with Intel AVX-512

Intel and LSTC continue to collaborate to optimize LS-DYNA, making use of Intel SSE2, Intel AVX2, and Intel AVX-512 instructions for improved performance. A key focus area has been to make source changes to allow Intel compilers to vectorize performance-critical loops, generating vector instructions including Intel AVX-512 (for example, using the -xCORE-AVX512 compiler option) for high performance.

The result of such optimizations is shown in Figure 3, with LS-DYNA R9.3_SP Explicit using the ODB-10M model on a 192-core cluster of 2nd Generation Intel Xeon Scalable processors. The baseline Intel SSE2 binary is built with Intel Fortran Compiler 13. Switching to Intel Fortran Compiler 16 with Intel SSE2 yields a performance increase of about 4 percent, indicating improvements in compiler vector-code generation in newer versions of the Intel compiler.[5] To get even more performance, switching to Intel Fortran Compiler 18 with Intel AVX-512 yields a net 22 percent improvement for LS-DYNA over the baseline. Intel has seen similar improvements for newer versions of LS-DYNA.

(Performance chart showing values: Intel SSE2/Intel Fortran Compiler 13 = 1, Intel SSE2/Intel Fortran Compiler 16 = 1.04, Intel AVX-512/Intel Fortran Compiler 18 = 1.22)
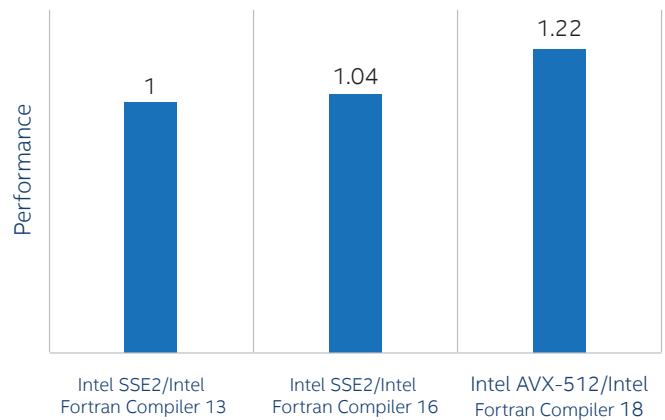
Figure 3. LS-DYNA R9.3_SP Explicit performance comparison of Intel AVX-512 compared to Intel SSE2 on a 192-core 2nd Generation Intel Xeon Platinum 8260L processor–based cluster using the ODB-10M model[5]

Figure 4 demonstrates the performance benefits of the 2nd Generation Intel Xeon Platinum 9242 processor (with 48 cores, running at 2.3 GHz) compared to the Intel Xeon Platinum 8260 processor (with 24 cores, running at 2.4 GHz) for LS-DYNA R9.3_SP Explicit on a single two-socket node— almost a linear scaling, with 1.95x speedup with the ODB-10M benchmark.[6]

3
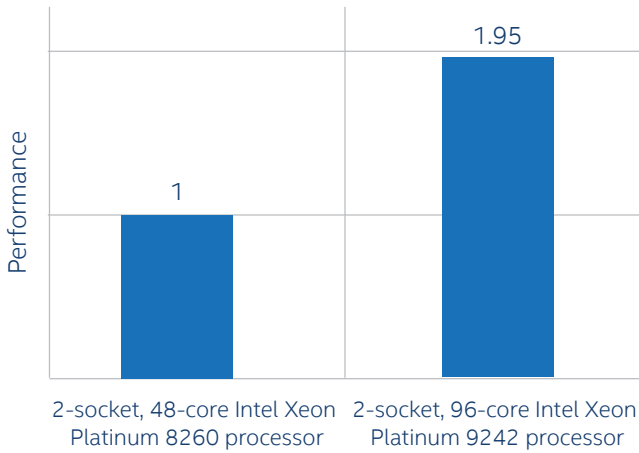
Figure 4. LS-DYNA R9.3_SP Explicit performance comparison between 2nd Generation Intel Xeon Platinum 9242 processor and Intel Xeon Platinum 8260 processor with the ODB-10M model[6]

A key optimization strategy for high-performance computing (HPC) applications on Intel platforms is to make use of the Intel MKL BLAS library where possible. Intel MKL is optimized for Intel SSE/SSE2, Intel AVX/AVX2, and Intel AVX-512, and it can dynamically use the appropriate vector instruction set at runtime, depending on which Intel processor it runs on. LS-DYNA R9.1_DP and later versions are linked with Intel MKL with support for Intel AVX-512 instructions.

To quantify the performance impact of Intel MKL for LS-DYNA Implicit, a performance comparison of Intel MKL using Intel SSE, Intel AVX, and Intel AVX-512 on 2nd Generation Intel Xeon Platinum 8260 processors and running an LS-DYNA R9.3_DP binary built without Intel MKL serves as the baseline. LS-DYNA Implicit was run in hybrid mode with two MPI ranks and 16 OpenMP threads per rank using 2.5M DOF in the CYL1E6 model. Figure 5 shows a significant performance improvement of 9.9x for LS-DYNA R9.3_DP Implicit (double precision) using Intel MKL and Intel AVX-512 compared to running without Intel MKL.[1]

The performance improvements described in this section—as high as 9.9x above the baseline—are demonstrated on standard LS-DYNA benchmarks. In addition, Intel and LSTC collaborate in optimizing proprietary customer workloads, broadening benefits for the wider LS-DYNA user community.
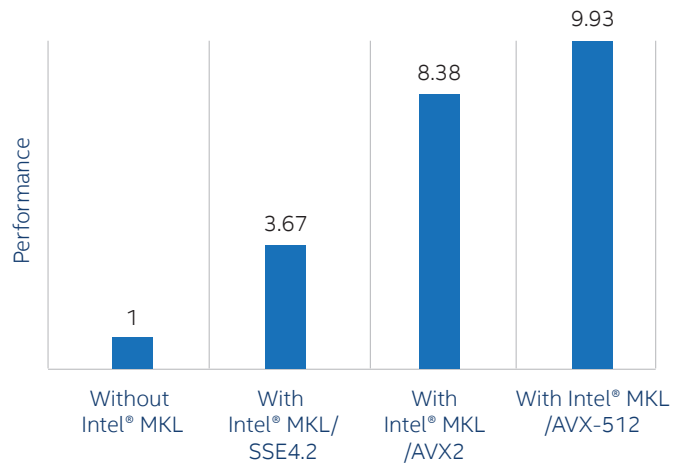

Figure 5. LS-DYNA R9.3_DP Implicit performance impact using Intel MKL and Intel AVX-512 on a 2nd Generation Intel Xeon Platinum 8260 processor using the CYL1E6 model[1]

LS-DYNA Explicit Numerical-Accuracy Comparisons with Intel AVX-512 and Intel SSE2

Even though the performance of LS-DYNA Explicit improves with the Intel technology optimizations shown earlier in this paper, numerical accuracy of LS-DYNA Explicit model is still the top concern for all LS-DYNA users. LS-DYNA users expect to see repeatable, consistent results when they move from Intel SSE2 to Intel AVX-512. In theory, it's hard to get repeatable, consistent results comparing an Intel SSE2 binary and an Intel AVX-512 binary, because the order of operations changes, and thus the accumulation of rounding errors varies. However, in practice, consistent results can be generated with an Intel AVX-512 binary. This is demonstrated with the car2car/120 ms model. Intel ran the simulation with Intel SSE2 and Intel AVX-512 binaries on the 2nd Generation Intel Xeon Platinum 9242 processor. During this simulation, internal energy (Figure 6), resultant displacement in node 5,341,465 (Figure 7), and the X-acceleration in node 341,513 (Figure 8) were checked. As seen from Figures 6–8, the curves are closely matched, showing that the numerical accuracy remains effectively unchanged.
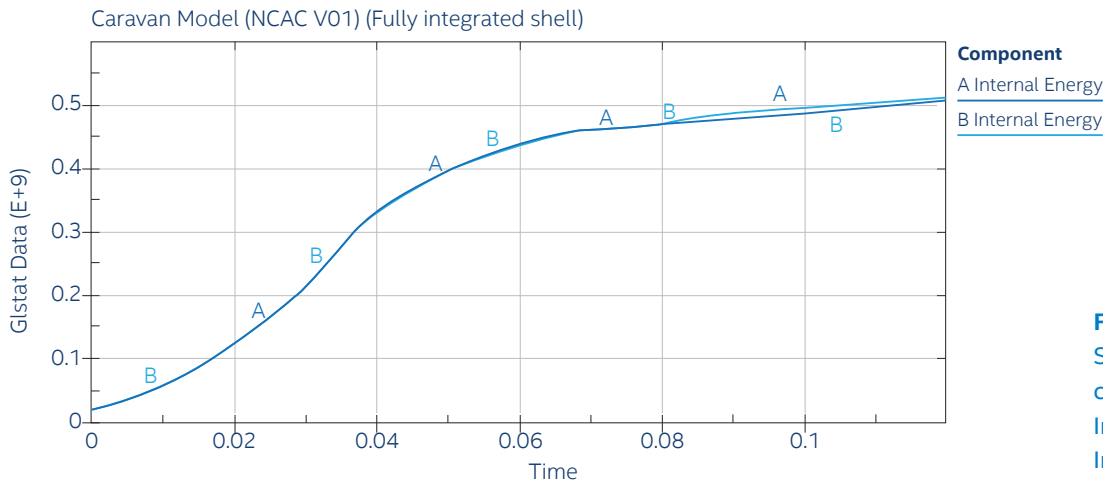
Figure 6. LS-DYNA R9.3_SP Explicit internal-energy comparison between using Intel AVX-512 (green) and Intel SSE2 (red)[7]
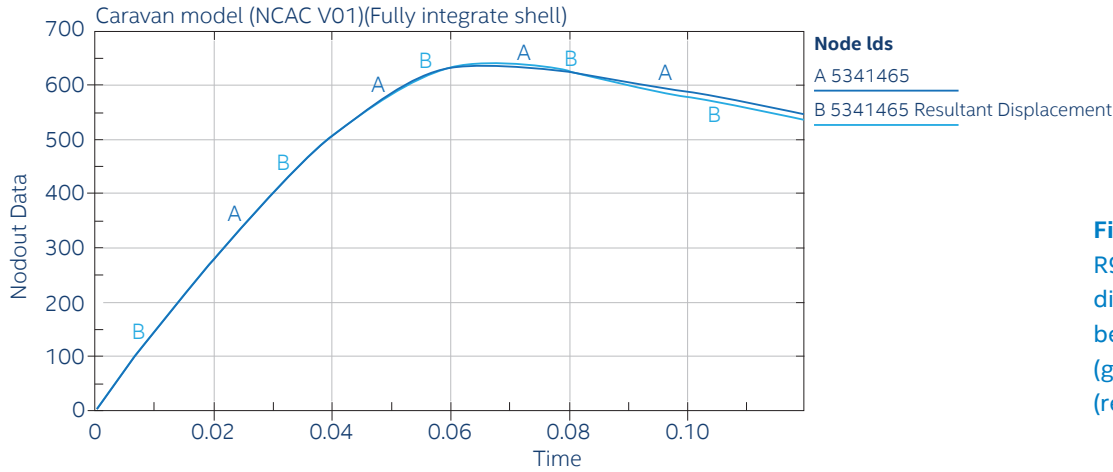
Figure 7. LS-DYNA R9.3_SP Explicit resultant-displacement comparison between Intel AVX-512 (green) and Intel SSE2 (red) at node 5,341,465[7]
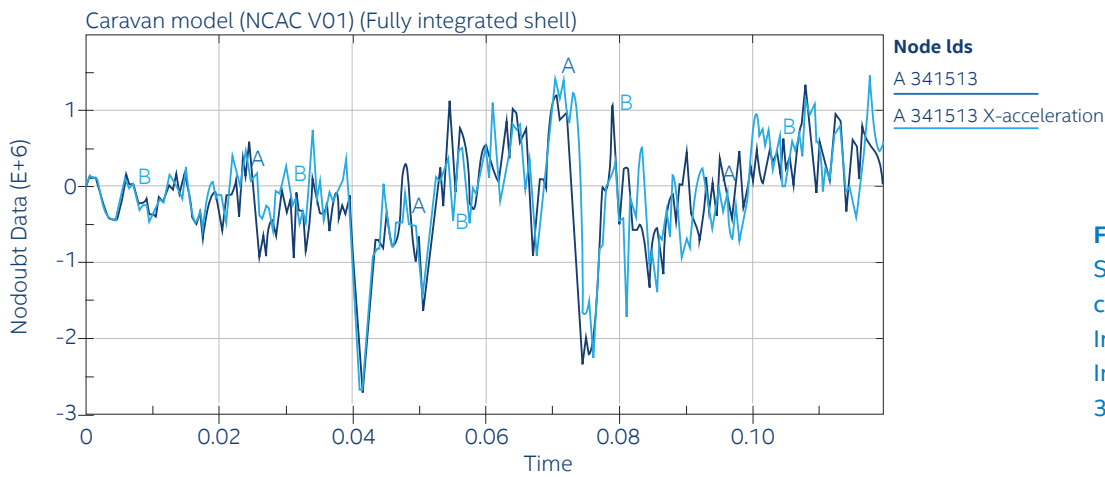
Figure 8. LS-DYNA R9.3_SP Explicit X-acceleration comparison between Intel AVX-512 (green) and Intel SSE2 (red) at node 341,513[7]

LS-DYNA Implicit with Intel Optane Persistent Memory

As the problem size of LS-DYNA Implicit models increases rapidly, input/output (I/O) performance becomes a serious bottleneck in existing systems when the solver uses an out-of-core algorithm. A fast local file system or larger memory is required to address the I/O performance bottleneck. To avoid the high cost of increasing DDR4 memory capacity, customers today tend to choose a fast local file system such as RAID 0. Intel Optane PMem now offers a cost-effective way to scale up memory instead.

Intel Optane PMem is an innovative memory technology that delivers a unique combination of affordable large capacity and persistence (that is, non-volatility), filling a gap between DRAM and NAND (see Figure 9). Intel Optane PMem can be provisioned in two modes. The first is called Memory Mode. With Memory Mode, applications get a high-capacity main-memory solution at substantially lower cost and power than DRAM, while providing performance close to that of DRAM, depending on the workload. No modifications are required for applications to use Memory Mode; the operating system sees the Intel Optane PMem module's 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, in which applications can access persistent memory capacity as non-volatile memory. In particular, the Storage over App Direct (SToAD) mode allows access to persistent memory capacity using standard file APIs. No modifications are needed to application code or the file systems that are normally required for block storage devices. SToAD provides high-performance block storage, without the latency of moving data to and from the I/O bus.

To demonstrate the impact of Intel Optane PMem on the performance of LS-DYNA Implicit, Intel ran the AWE CYL2E6 model (6M DOF) on a 2nd Generation Intel Xeon Gold 6248M processor in three modes (RAID 0 mode, SToAD mode, and Memory Mode). The baseline configuration used 192 GB DDR4 memory with solid state drive (SSD)-based RAID 0. The second configuration used 192 GB DDR4 memory and 1.5 TB of Intel Optane PMem with SToAD mode. The third configuration used 192 GB DDR4 memory and 1.5 TB of Intel Optane PMem in Memory Mode. The benchmark testing found that the performance of Intel Optane PMem in SToAD mode is 14 percent better than the RAID 0 SSD setup.[9] The performance of Intel Optane PMem in Memory Mode is 75 percent faster than the baseline RAID 0 SSD setup. This significant performance jump is primarily due to LS-DYNA Implicit being able to use an in-core solver with the large memory capacity of Intel Optane PMem (Figure 11).
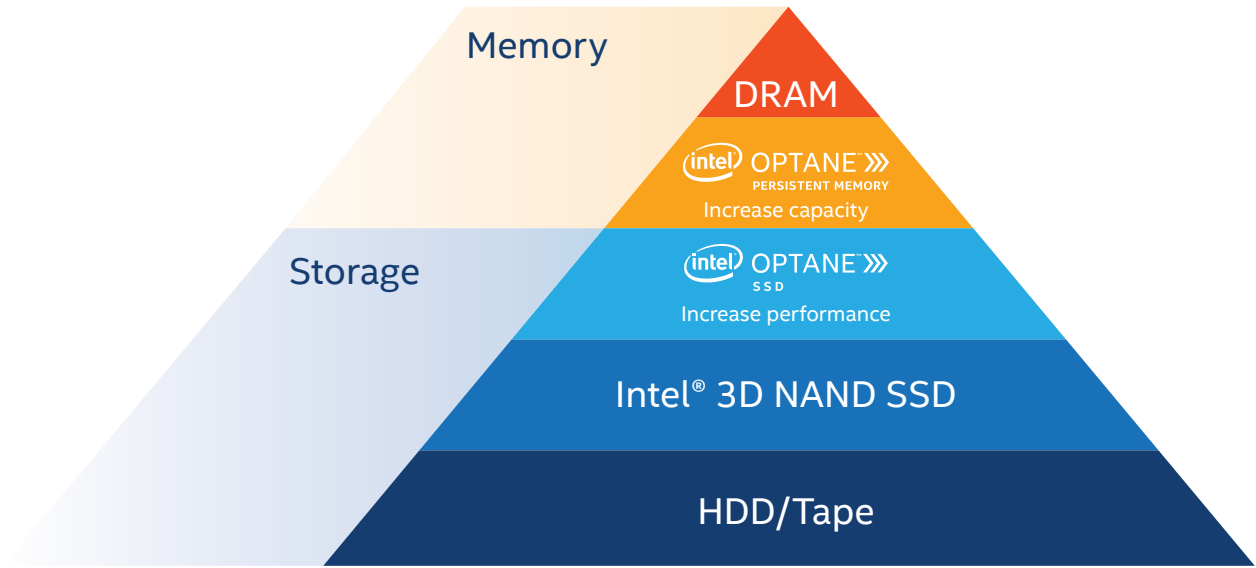
Figure 9. Storage/memory performance compared in the capacity pyramid[8]
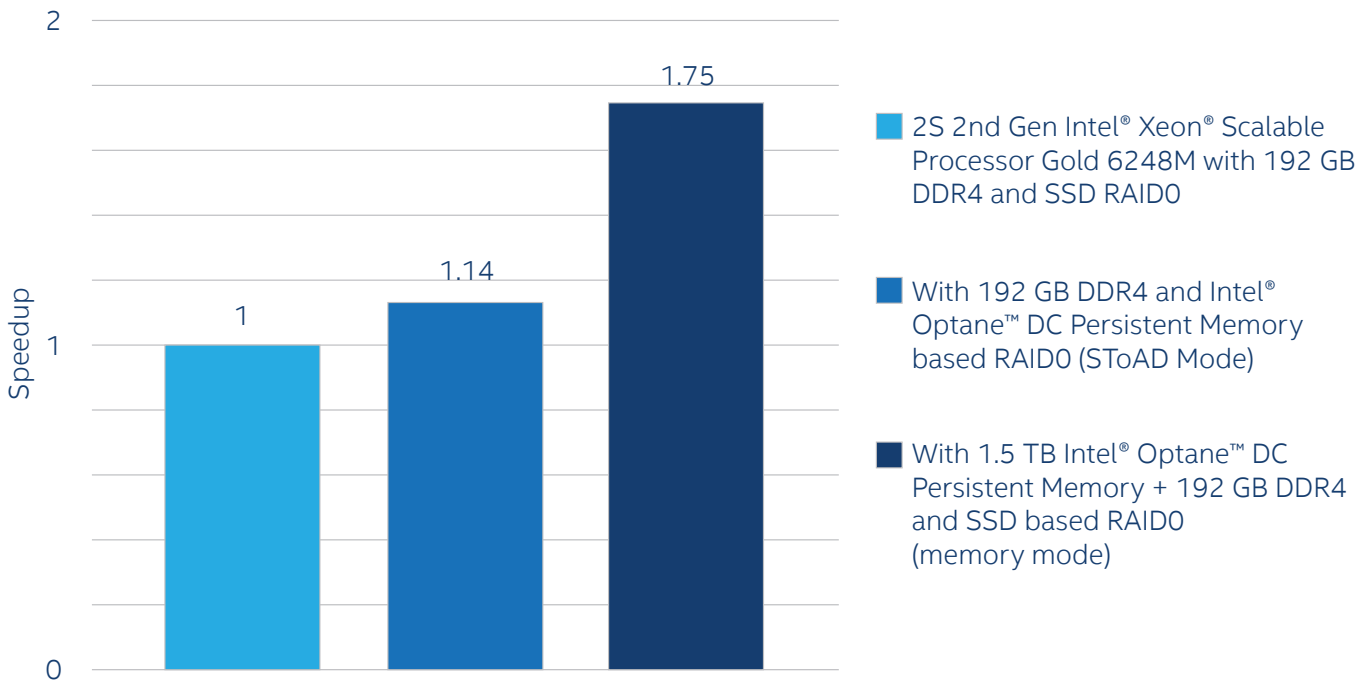

Figure 10. LS-DYNA R9.3_DP Implicit performance improves 1.75x using Intel Optane persistent memory in Memory Mode compared to a baseline RAID 0 SSD setup[9]

Conclusion

The performance of both LS-DYNA Explicit and LS-DYNA Implicit continue to improve using the latest Intel technologies, including 2nd Generation Intel Xeon Scalable processors, Intel Fortran Compiler 18, Intel MKL, Intel AVX-512, and Intel Optane PMem. In addition, LS-DYNA Explicit numerical accuracy with Intel AVX-512 is shown to match closely with the Intel SSE2 version. Intel and LSTC remain committed to deeper collaborations, optimizing LS-DYNA R9.3, R11, and other versions to bring the best value to LSTC customers on Intel architecture–based platforms.

¹ Based on Intel testing as of March 22, 2019. System configuration: one node, two-socket Intel Xeon Platinum 8260 processor, 384 GB DDR4 memory at 2,933 MHz, 1 x 1.2 TB Intel SSD DC S3710, Intel Turbo Boost Technology enabled, Intel Hyper-Threading Technology (Intel HT Technology) disabled, 2 MPI ranks x 16 OpenMP threads, Intel MPI Library 2018.

² Intel. „Intel Xeon Processor Scalable Family Technical Overview." July 2017. https://software.intel.com/en-us/articles/intel-xeon-processor-scalable-family-technical-overview.

³ Intel. „Second Generation Intel Xeon Scalable Processors." intel.com/content/www/us/en/products/docs/processors/xeon/2nd-gen-xeon-scalable-processors-brief.html.

⁴ Intel. „Intel VTune Amplifier Tuning Guide for the Intel Xeon Processor Scalable Family, 2nd Gen." 2019. https://software.intel.com/sites/default/files/managed/bf/4a/TuningGuide_IntelXeonProcessor_ScalableFamily_2ndGen.pdf.

⁵ Based on Intel testing as of February 19, 2019. System configuration: four nodes, two-socket Intel Xeon Platinum 8260L processor, 384 GB DDR4 memory at 2,667 MHz, 1 x 1.2 TB Intel SSD DC S3710, Intel OPA, Intel Turbo Boost Technology enabled, Intel HT Technology disabled, 48 MPI ranks per node, Intel MPI Library 2018 Update 1, Linux release 3.10.0-957.5.1.el7.crt1.x86_64.

⁶ Based on Intel testing as of February 19, 2019. System configuration: four nodes, two-socket Intel Xeon Platinum 8260L processor, 384 GB DDR4 memory at 2,667 MHz, 1 x 1.2 TB Intel SSD DC S3710, Intel OPA, Intel Turbo Boost Technology enabled, Intel HT Technology disabled, 48 MPI ranks per node, Intel MPI Library 2018 Update 1, Linux release 3.10.0-957.5.1.el7.crt1.x86_64.

⁷ Based on Intel testing as of March 22, 2019. System configuration: one node, two-socket Intel Xeon Platinum 8260L processor, 384 GB DDR4 memory at 2,667 MHz, 1 x 1.2 TB Intel SSD DC S3710, Intel Turbo Boost Technology enabled, Intel HT Technology disabled, 48 MPI ranks, Intel MPI Library 2018 Update 4, Linux release 3.10.0-957.5.1.el7.crt1.x86_64 compared to single node, two-socket Intel Xeon Platinum 9242 processor, 384 GB DDR4 memory at 2,667 MHz, 1 x 1.2 TB Intel SSD DC S3710, Intel Turbo Boost enabled, Intel HT Technology disabled, 96 MPI ranks, Intel MPI LIbrary 2018 Update 4, Linux release 3.10.0-957.5.1.el7.x86.

⁸ Intel. „Intel Optane DC Persistent Memory: A Major Advance in Memory and Storage Architecture." January 2019. https://software.intel.com/en-us/blogs/2018/10/30/intel-optane-dc-persistent-memory-a-major-advance-in-memory-and-storage-architecture.

⁹ Based on Intel testing as of [Date]. System configuration: two-socket Intel Xeon Gold 6248M node, 192 GB DDR4 memory at 2,933 MHz, 1.5 TB Intel Optane PMem and RAID 0 with 2 x 1.2 TB Intel SSD DC S3710, Intel Turbo Boost Technology enabled, Intel HT Technology disabled, 2 MPI ranks x 12 OpenMP threads, Intel MPI Library 2018, Linux kernel: 4.18.19-100.fc27.x86_64).