

BENEFITS OF INTEL® XEON® SCALABLE PROCESSORS FOR VIDEO COMPRESSION

E. Mora PhD (Research Engineer), J. Viéron PhD (Research Director), P. Larbier (CTO) – ATEME

KEY TAKEAWAYS

- New Intel® Xeon® Scalable processors enable implementation of demanding compression applications on standard dual-socket servers.
- Overall performance gains of up to 75 percent were observed on the Intel® Xeon® Platinum 8180 processor compared to the previous generation Intel® Xeon® processor E5-2697 v4 on the ATEME TITAN Live* broadcast video encoder.¹
- This performance gain was due to increases in the number of cores, memory bandwidth, and processor frequency. In addition, internal architecture changes to the processor bring significant gains.
- The advanced capabilities of Intel Xeon Scalable processors solve system complexity challenges faced by broadcasters and service providers. Real-time Ultra-High Definition video encoding is now possible using a single processor, providing additional capacity for adaptive-bitrate encoding applications required for OTT delivery on a dual-socket system.

Introduction

Digital video compression is such a compute-intensive task that for many years, using dedicated and highly specialized hardware was the only means to achieve real-time video encoding. But with advances in the performance and capabilities of general-purpose processors, software-based transcoding has become possible. Legacy video codecs such as MPEG-2 were initially implemented to compress Standard Definition videos (720 x 480 pixels at 30 frames per second). More advanced codecs such as H.264/MPEG-4 AVC are targeted for High Definition (1920 x 1080 pixels at 30 frames per second) encoding. Today, standard servers equipped with the dual-socket Intel® Xeon® processor E5 v4 product family can be used to compress Ultra-High Definition videos (3840 x 2160 pixels at 60 frames per second) in real time using the state-of-the-art HEVC codecs. In less than 15 years, tremendous improvements in both compression algorithms and processor technology have enabled the real-time software implementation of a codec that is far more complex, compressing videos that contain 48 times more pixels.

However, the vast benefit of software-based transcoding running on standards-based servers comes at a cost. Due to the complexity of the transcoding process, only one Ultra-High Definition channel could be compressed on a dual-socket server based on the Intel® Xeon® processor E5-2699 v4. There are simply not enough resources to perform other tasks, such as encoding lower-resolution versions of that channel. Therefore, adaptive bitrate formats that target over-the-top delivery require splitting operations over several synchronized servers, which leads to a significant increase in system complexity. In addition, specialized software is required to deliver the flexibility of pure software processing. Thus, as the demand for Ultra-High Definition compression rises, the capability to handle more than a single channel on a standard dual-socket server becomes increasingly desirable.

ATEME provides video compression systems to broadcasters and service providers. TITAN Live, the company's flagship software solution, is used to process thousands of live TV channels globally. The solution is commonly deployed in data centers, primarily on servers based on the Intel Xeon processor E5 family. For this reason, TITAN Live is highly optimized for Intel® architecture, and about 80 percent of processing time is spent on handcrafted assembly code. Hence, any important algorithmic change demands significant development efforts. Moreover, further program optimization may impact video quality, which is simply not an option for professional services.

Given that most data centers operated by broadcasters and MSOs run on Intel Xeon processors, the introduction of the new Intel® Xeon® Scalable processors has transformative potential for the media transcoding market. For ATEME, the numerous improvements offered by Intel Xeon Scalable processors over the previous generation, including more cores, memory, and I/O bandwidth, as well as the availability of the Intel® Advanced Vector Extensions 512 (Intel® AVX-512)² instruction set, have the potential to increase the number of video channels that can be processed on a single server, increasing density and reducing cost. This paper reports on testing to determine whether the new Intel Xeon Scalable processors can break the Ultra-High Definition barrier and solve compression bottlenecks.

Performance Improvement Measurements

Test set-up

To evaluate ATEME TITAN Live performance improvements using Intel Xeon Scalable processors, Intel provided a dual-socket server equipped with Intel® Xeon® Platinum 8180 processors. This high-end processor contains 28 cores operating at a base frequency of 2.5 GHz. Measurement results are compared to state-of-the-art dual-socket servers with the previous generation Intel® Xeon® processors E5-2697 v4 and E5-2699 v4. All of the tests are performed with Intel® Hyper Threading Technology enabled. The hardware specifications of the three platforms are summarized in Table 1.

Performance tests were realized using four standard progressive 10-bit Ultra-High Definition sequences³ (3840 x 2160 at 60 frames per second) representing a wide range of compression complexity: “Boxing Practice”, “Crosswalk”, “FoodMarket2”, and “RitualDance”. For each sequence, the minimal encoding speed was measured in terms of the

number of frames encoded per second. A configuration is proved to be suitable for real-time operation if the minimum encoding speed is above 60 frames per second.

The ATEME TITAN Live software was configured to compress the test sequences using HEVC Main 10 profile at a challenging constant bitrate of 40 Mbps. The quality level of the encoder was set to “Live Broadcast,” which provides the highest encoding speed available.

All tests were performed on a stock CentOS* 7.2 Linux* operating system. No modifications were made to the software to take advantage of a specific platform.⁴

Two configurations were tested: single-socket and dual-socket. Single-socket measurements were performed by setting affinity on one of the processors, disabling execution of the software on the cores associated with the other processor, and forcing memory I/O operations to be exclusively performed to and from the memory bank associated with the considered processor. Verifications were made to ensure that only the cores associated with the considered processor were active, that the Intel® QuickPath Interconnect (Intel® QPI) or Intel® Ultra Path Interconnect (Intel® UPI) links between the two processors were not utilized, and that there were no memory access misses (i.e., the processor only got data from its own memory bank). Since a single-socket task exclusively utilizes one processor’s resources, and since resources are symmetrical between processors (each processor has the same number of cores and the same memory bandwidth and capacity), the task’s performance is the same if it is executed on the other processor. Moreover, if both tasks are executed simultaneously, each on a different processor, the performance results also remain the same.

Table 1. Hardware specifications of the different platforms used for evaluation.

| Platform | Max Turbo Frequency | Processor Base Frequency | Physical Cores per Socket ^a | TDP | RAM per CPU | RAM Frequency |
|--------------------------------------|---------------------|--------------------------|--|-------|-------------|---------------|
| Intel® Xeon® Processor E5-2697 v4 | 3600 MHz | 2300 MHz | 18 | 145 W | 4 x 4 GB | 2400 MHz |
| Intel® Xeon® Processor E5-2699 v4 | 3600 MHz | 2200 MHz | 22 | 145 W | 4 x 8 GB | 2133 MHz |
| Intel® Xeon® Platinum 8180 Processor | 3800 MHz | 2500 MHz | 28 | 205 W | 6 x 16 GB | 2666 MHz |

^a Intel® Hyper-Threading Technology enabled

Table 2. Performance on Intel® Xeon® Processor E5 v4 product family-based platforms.

| Processor Configuration | | Ultra-High Definition Test Sequences (HEVC Main 10 @ 40 Mbps) | | | |
|-------------------------|--------------------------------------|---|-----------|-------------|-------------|
| | | BoxingPractice | Crosswalk | FoodMarket2 | RitualDance |
| Single-Socket | 1x Intel® Xeon® Processor E5-2697 v4 | 40 fps | 40 fps | 40 fps | 37 fps |
| | 1x Intel® Xeon® Processor E5-2699 v4 | 44 fps | 44 fps | 45 fps | 41 fps |
| Dual-Socket | 2x Intel Xeon Processor E5-2697 v4 | 65 fps | 64 fps | 65 fps | 62 fps |
| | 2x Intel Xeon Processor E5-2699v4 | 69 fps | 69 fps | 68 fps | 65 fps |

Measurements shown in Table 2 demonstrate that an HEVC Main 10 real-time encoding of Ultra-High Definition sequences on a single-socket server is not possible on either platform based on the Intel Xeon processor E5 v4 product family because the minimal encoding speed is lower than the frame rate for all sequences. Real-time encoding requires a dual-socket server, as shown in Table 2. With this dual-socket configuration, the CPU load is approximately 70 percent. However, trying to exploit the remaining CPU resources (e.g., encoding lower resolution channels) leads to real-time loss on Ultra-High Definition encoding.

This means that another Ultra-High Definition real-time encoding cannot be performed in parallel, because both processors are busy encoding the first channel. In addition, there is no room for adaptive bitrate applications, because no processing power remains available to encode additional lower-resolution channels. This leads to a poor tradeoff between density and server cost.

Performance on the Intel® Xeon® Platinum 8180 processor

The same tests performed on the two platforms based on the Intel Xeon processor E5 v4 product family were also conducted on a platform based on two Intel Xeon Platinum 8180 processors. The hardware specifications of the platform are summarized in Table 1. Single-socket measurements are shown in Table 3.

The significant coding performance improvement on the Intel Xeon Platinum 8180 processor-based platform in comparison to the platforms based on the previous generation Intel Xeon processor E5 v4 product family (about 60 percent higher than the Intel Xeon processor E5-2697 v4 and about 50 percent higher than the Intel Xeon processor E5-2699 v4) allows real-time transcoding of Ultra-High Definition sequences on a single-socket system because the minimal encoding speed is higher than the frame rate for all sequences.

Because real-time Ultra-High Definition video encoding is reached with a single-socket, the second socket could be used for either an additional Ultra-High Definition encoding or for adaptive bitrate encoding applications, effectively solving the main system complexity issue faced today.

Table 3. Platform performance using a single Intel® Xeon® Platinum 8180 processor.

| Platform | Ultra-High Definition Test Sequences (HEVC Main 10 @ 40 Mbps) | | | |
|--|---|-----------|-------------|-------------|
| | BoxingPractice | Crosswalk | FoodMarket2 | RitualDance |
| Intel® Xeon® Platinum 8180 Processor (Single-Socket) | 65 fps | 63 fps | 67 fps | 62 fps |

Table 4. Performance on the Intel® Xeon® Platinum 8180 processor-based platform with different memory configurations.

| Memory Configuration | Test Sequences (HEVC Main 10 @ 40 Mbps on a Single-Socket Platform Based on the Intel® Xeon® Platinum 8180 Processor) | | | |
|----------------------|---|-----------|-------------|-------------|
| | BoxingPractice | Crosswalk | FoodMarket2 | RitualDance |
| 4 x 16 GB @ 2.4 GHz | 49 fps | 47 fps | 50 fps | 49 fps |
| 6 x 16 GB @ 2.66 GHz | 65 fps | 63 fps | 67 fps | 62 fps |

Performance Improvements Analysis

Several factors impact the performance gains observed using Intel Xeon Platinum 8180 processor-based platforms compared to platforms based on the Intel Xeon processor E5 v4 product family. As Table 1 indicates, the number of cores, number of memory banks, memory frequency, and maximum processor frequency have all been increased. There is also a gain resulting from the internal architecture changes to the processor. Several tests were performed to measure the impact of each of these factors.

Impact of the increased memory bandwidth

Tests were conducted on the Intel Xeon Platinum 8180 processor-based platform, with two fewer memory banks on each CPU (four compared to six) and with a lower memory frequency (2400 MHz compared to 2666 MHz), to match the corresponding specifications of the platforms based on the Intel Xeon processor E5 v4 product family. Table 4 summarizes the results.

Results in Table 4 show that increasing memory performance provides an average improvement in encoding speed of 31 percent. This significant gain is explained by the fact that ATEME’s HEVC encoder is bound by external memory bandwidth. Therefore, increasing this bandwidth delivers a performance boost, as measured here.

The performance increase is lower than the 66 percent increase in memory bandwidth because the available bandwidth is higher than the current ATEME’s TITAN Live requirements. The remaining unused bandwidth could be used to improve video quality, for example, by adding an additional reference frame.

Note that the Intel Xeon Platinum 8180 processor’s memory capacity equals 96 GB (6 x 16 GB) per socket, while the memory capacities of the Intel Xeon processor E5-2697 v4 and E5-2699 v4 are 16 GB and 32 GB, respectively. However, the increase in capacity does not help to speed up the encoding. For a single-channel Ultra-High Definition encoding, 16 GB is sufficient capacity to host the source and reference images for ATEME’s TITAN Live encoder.

Impact of the increased core count

To evaluate the impact on the results of the number of available cores, tests were conducted using 18 and 28 physical cores on the Intel Xeon Platinum 8180 processor to match the corresponding specifications of the platform based on the Intel Xeon processor E5-2697 v4 platform for single-socket encodings. Furthermore, in both cases, CPU frequency was locked at 2.5 GHz to focus on the impact of core count on encoding speed. Results are provided in Table 5.

Table 5. Impact of the core count on the Intel® Xeon® Platinum 8180 performance results.

| Core Configuration | Test Sequences (HEVC Main 10 @ 40 Mbps on Single-Socket Intel® Xeon® Platinum 8180 Processor) | | | |
|--|---|-----------|-------------|-------------|
| | BoxingPractice | Crosswalk | FoodMarket2 | RitualDance |
| 18 Physical Cores (36 Threads) @ 2.5 GHz | 42 fps | 40 fps | 44 fps | 41 fps |
| 28 Physical Cores (56 Threads) @ 2.5 GHz | 57 fps | 55 fps | 58 fps | 54 fps |

Results show a 34 percent performance improvement, while the number of cores has been increased by 55 percent. This result suggests that the increased core count has a limited benefit to performance because perfect parallelism in video compression is very difficult to achieve. Indeed, video compression exploits neighboring dependencies, which imposes a wavefront coding block order. Such an order limits the number of threads capable of running in parallel.

Parallelizing HEVC encoding is easier than with previous standards, because HEVC provides dedicated tools (e.g., slices, tiles, wavefront parallel processing) that are more useful than tools provided in H.264/MPEG-4 AVC (i.e., only slices). However, although those tools may make parallelizing a video encoder easier and more efficient, there are still limitations on the number of threads that can work efficiently on a single frame.

For example, wavefront parallel processing allows a number of parallel threads lesser than or equal to the number of Coding Tree Blocks (CTBs) vertically, and lesser than or equal to half the CTBs horizontally. For an Ultra-High Definition video, it “only” amounts to 60 threads (68 32 x 32 CTBs vertically, but only 120/2=60 horizontally). One must also note that encoding speed is not linearly linked with the number of threads, since each thread beyond the first must wait for two CTBs to be encoded by the previous thread before starting.

Figure 1 illustrates both limitations on a 5 x 3 CTB frame, where CTBs with the same color are encoded in parallel. Only one thread is active while the first two CTBs (i.e. “0” and “1”) are encoded, then two threads with the next four CTBs (i.e., “2” and “3”), etc. In this example, the speed up from using three threads is only 1.66x (i.e., 15 CTBs encoded in nine time slots).

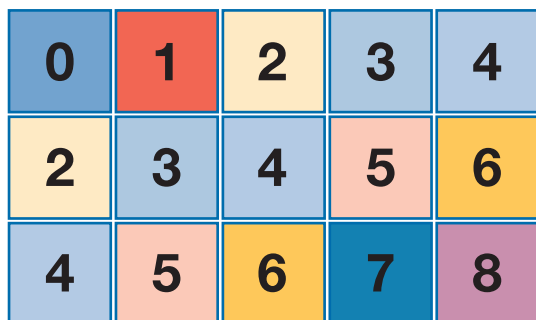


Figure 1. 5x3 CTB encoded with wavefront.

With the setups above, the actual speed-ups provided by threading are 27.3x for 36 threads and 31.7x for 56 threads, leading to a 15 percent theoretical speed increase. To achieve the measured 34 percent increase, ATEME’s TITAN Live encoder simultaneously uses tiles and wavefront coding block order.

Because of these threading limitations, cores are underutilized when encoding a single Ultra-High Definition channel in the current version of ATEME’s TITAN Live encoder. This underutilization provides capacity to perform other tasks, such as encoding a lower-resolution channel for adaptive bitrate applications. The platform could also be downscaled to have just enough cores to allow real-time encoding of one Ultra-High Definition sequence on a single socket. This potential reduction in the number of cores (26 cores would be enough) allows savings on server cost and power consumption.

Impact of the internal architecture changes

Some performance gains on the new platform come from the Intel Xeon Platinum 8180 processor alone. To isolate and evaluate the impact of the processor change, tests comparing platforms based on the Intel Xeon processor E5-2697 v4 and the Intel Xeon Platinum 8180 processor with the same external memory bandwidth and core count were conducted. To match the Intel Xeon processor E5-2697 v4 platform, only four memory banks at 2.4 GHz and only 18 physical cores per processor were used on the Intel Xeon Platinum 8180 processor platform. Furthermore, in both cases, CPU frequency was locked at 2.3 GHz (the base frequency of the Intel Xeon processor E5-2697 v4) to focus only on internal architecture changes. Table 6 summarizes the results.

Table 6 results show a compression speed-up of approximately five percent related to the architectural changes. ATEME’s TITAN Live encoder benefits mainly from the significant increase of the L2 cache size in the Intel Xeon Platinum 8180 processor, which is four times that of the previous generation Intel Xeon processor E5 v4 product family. More information pertaining to a coding block (neighboring information, motion vectors, etc.) can thus be stored in the cache for rapid access. The overall cache miss rate has been measured to be divided by a factor of 3.8 compared to the Intel Xeon processor E5 v4 product family.

Table 6. Impact of the internal architecture changes to the processor.

| Processor | Test Sequences (HEVC Main 10 @ 40 Mbps on a Single Socket) | | | |
|---|---|-----------|-------------|-------------|
| | BoxingPractice | Crosswalk | FoodMarket2 | RitualDance |
| 1x Intel® Xeon® Processor E5-2697 v4 @ 2.3 GHz | 34 fps | 33 fps | 34 fps | 32 fps |
| 1x Intel® Xeon® Platinum 8180 Processor @ 2.3 GHz | 36 fps | 35 fps | 35 fps | 34 fps |

Note that ATEME’s TITAN Live encoder can also benefit from other architecture improvements, the impacts of which are difficult to evaluate, such as improved branch predictor, faster instruction decoder, improved scheduler and execution engine, more load/store bandwidth, deeper load/store buffers, improved hardware prefetcher, and the mesh interconnect architecture.

Speedup could potentially be achieved in the current ATEME’s TITAN Live encoder by redesigning data memory layout to better benefit from the increased cache size.

Impact of the frequency increase

In addition to internal architecture changes, the Intel Xeon Platinum 8180 processor can reach higher frequency than the Intel Xeon processor E5 v4 product family.

Base frequencies of the Intel Xeon processor E5-2697 v4 and Intel Xeon processor E5-2699 v4 are 2.3 GHz and 2.1 GHz, respectively, while the base frequency of the Intel Xeon Platinum 8180 processor is 2.5 GHz, resulting in performance increases of nine percent and 19 percent, respectively.

Table 7. Operating frequency with Intel® Turbo Boost Technology on ATEME workload.

| | Operating Frequency |
|--|---------------------|
| Intel® Xeon® Processor E5-2697 v4 (18 Physical Cores) | 2700 MHz |
| Intel® Xeon® Processor E5-2699 v4 (22 Physical Cores) | 2630 MHz |
| Intel® Xeon® Platinum 8180 Processor (18 Physical Cores) | 3300 MHz |
| Intel Xeon Platinum 8180 Processor (28 Physical Cores) | 2850 MHz |

Accounting for Intel® Turbo Boost technology while running ATEME’s TITAN Live encoder workload⁵ (which makes use of the Intel® AVX2 instruction set), the measured operating frequencies are provided in Table 7.

Assuming 18 physical cores, Intel Turbo Boost technology increases the frequency gap between Intel® processor generations. Thus, the effective performance gain between the Intel Xeon processor E5-2697 v4 and the Intel Xeon Platinum 8180 processor jumps from 9 percent to 22 percent. Note that even when using 28 physical cores, the operating frequency of the Intel Xeon Platinum 8180 processor remains higher than the frequency of the Intel Xeon processor E5-2697 v4 (2850 MHz compared to 2700 MHz, nearly a six percent increase).

Impact of the Intel® AVX-512 instruction set

The Intel AVX-512 instruction set supported by the Intel Xeon Scalable processors has multiple benefits with potentially significant impact on coding performance. First, 512-bit registers provide double the processing capacity for some functions such as 32 x 32 transforms in comparison to Intel AVX2 instructions, which are limited to 256-bit registers. Intel AVX-512 technology also offers twice the number of registers as Intel AVX2 technology, enabling more efficient optimizations of some functions that involve large amounts of data, such as butterfly 32 x 32 transforms. Intel AVX-512 instructions will have limited impact on smaller block processing (8 x 8 or smaller); hence the benefit is biased toward higher-resolution encodings. It is not yet possible to accurately evaluate the impact of using Intel AVX-512 instructions on coding performance, because the optimized functions benefitting from these instructions have not yet been coded by ATEME and implemented in the encoder.

For ATEME’s TITAN Live encoder, functions that can be optimized using Intel AVX-512 instructions represent roughly 20 percent of the current encoding time. Speed-up expectations for these functions are between 50 percent and 100 percent. Thus, the overall gain of using Intel AVX-512 technology is estimated to be around eight percent. Table 8 summarizes the expected performance improvement.

Table 8. Estimate of the coding performance gain enabled by Intel® AVX-512 Technology.

| Processor | Test Sequences (HEVC Main 10 @ 40 Mbps on a Single Socket) | | | |
|--|--|-----------|-------------|-------------|
| | BoxingPractice | Crosswalk | FoodMarket2 | RitualDance |
| 1x Intel® Xeon® processor E5-2697 v4 Using Intel® AVX2 Technology | 40 fps | 40 fps | 40 fps | 37 fps |
| 1x Intel® Xeon® Platinum 8180 Using Intel AVX2 Technology | 65 fps | 63 fps | 67 fps | 62 fps |
| 1x Intel® Xeon® Platinum 8180 Using Intel® AVX-512 Technology (Estimate) | 70 fps | 68 fps | 72 fps | 67 fps |

Summary of performance gains

This section summarizes the performance gains provided by the Intel Xeon Platinum 8180 processor compared to the Intel Xeon processor E5-2697 v4 running ATEME’s TITAN Live video encoder. Performance gains include:

- Up to 31 percent improvement due to increased memory bandwidth
- Up to 34 percent improvement due to increased core count
- Up to five percent improvement due to architectural changes
- Up to six percent improvement due to frequency increase
- Up to eight percent improvement due to Intel AVX-512 technology

These performance gains are summarized in Figure 2. The overall performance gain is 75 percent¹. The overall gain cannot be computed as the summation of all the listed gains, because some of the improvements overlap (e.g., increased core count, architectural changes, and processor frequency increase are closely linked).

Another way to present the results is to proceed incrementally, starting with Intel Xeon processor E5-2697 v4 and Intel Xeon processor E5-2699 v4 platforms and working up to the Intel Xeon Platinum 8180 processor with optimization for Intel AVX-512 technology. That approach is illustrated in Figure 2, where:

- “Intel Xeon Processor 2697 v4” and “Intel Xeon Processor 2699 v4” represent baseline performance with those processors.
- “Intel Xeon Platinum 8180 Processor (Four Banks)” represents the Intel Xeon Platinum 8180 processor with four memory banks, RAM operating at 2.4 GHz (the same memory configuration as the predecessor platforms).
- “Intel Xeon Platinum 8180 Processor (Six Banks)” represents the Intel Xeon Platinum 8180 processor with its default memory configuration (six memory banks, RAM operating at 2.66 GHz).
- “Intel Xeon Platinum 8180 Processor (AVX-512)” represents the Intel Xeon Platinum 8180 processor with its default memory configuration and Intel AVX-512 optimizations performed in ATEME’s TITAN Live encoder.

This representation shows that increased memory bandwidth has the most significant impact on performance, which is explained by the fact that the ATEME’s TITAN Live encoder was already memory-bound on platforms based on the Intel Xeon processor E5 v4 product family. Switching to the new platform without increasing memory bandwidth only provides 24 percent performance gain (compared to the 75 percent overall gain).

Recommendations

Based on tests conducted on the Intel Xeon Platinum 8180 processor, ATEME recommends that broadcasters and service providers focus on three main paths to fully leverage available resources and improve performance of professional video encoders. Specifically, tuning and optimization should be conducted based on encoder internal threading strategy, Intel AVX-512 instructions, and cache usage.

To efficiently take advantage of the high core counts in the Intel Xeon Platinum 8180 processor, care must be taken with regard to threading in video-encoding applications. While it was challenging on previous Intel platforms to fully exploit all available cores while encoding a single channel, the challenge is greater with higher core counts and may require significant changes in threading strategy. ATEME’s approach was to reduce the sizes of the CTBs and to increase the number of tiles, but other strategies may also be valuable, such as multi-frame or chunk parallelism.

Another promising path for optimization is through use of the Intel AVX-512 Instruction set. This approach is not easily applicable to small block sizes (i.e., 16 x 16 or less), but functions working on 32 x 32 blocks or larger will benefit greatly. Such functions will need to be rewritten from scratch using Intel AVX-512 instructions, and code using those functions will need to be adjusted to meet the additional requirements of Intel AVX-512 technology, such as memory-data alignment.

Finally, regarding cache performance, ATEME recommends profiling cache misses of the video codec application on the Intel Xeon Platinum 8180 processor. Depending on the results, it may be beneficial to perform data-size and data-layout optimizations so that frequently accessed data remains in the cache during the entire encoding process. Note that such optimizations are relatively simple to perform on the Intel Xeon Platinum 8180 processor because of the large cache size.

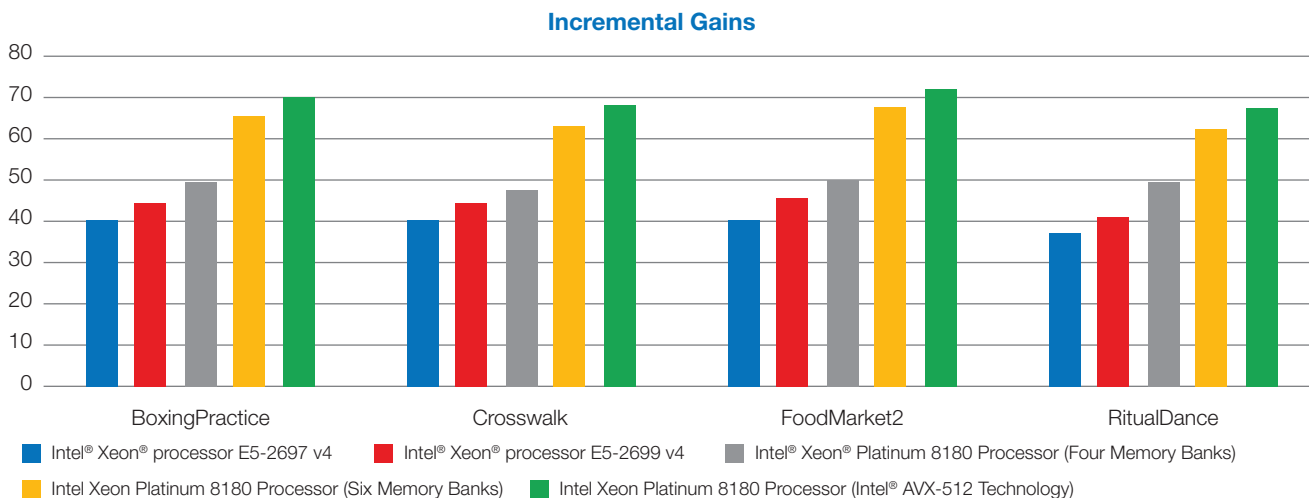


Figure 2. Incremental gains enabled by the Intel® Xeon® Platinum 8180 platform.

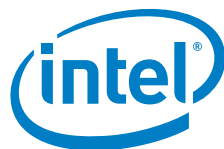
Conclusion

This paper demonstrates that the Intel Xeon Scalable processors provide significant benefits for video-compression applications. All reported tests are performed running ATEME's TITAN Live broadcast video encoder. Overall performance gains of up to 75 percent were observed on the Intel Xeon Platinum 8180 processor, compared to the predecessor Intel Xeon processor E5-2697 v4.

Several factors contribute to this performance gain, including increases to the number of cores, memory bandwidth, and maximum CPU operating frequency. In addition, changes to the processor's internal architecture deliver significant gains.

Importantly, these gains are possible without changing a single line of code. Despite a free initial development effort, the Intel Xeon Scalable processors offer room for improvements to professional codec developers. Specific recommendations regarding encoder internal threading strategy, use of Intel AVX-512 instructions, and cache usage optimizations are provided in this paper. Purely from a video-quality point of view, since it is very difficult to optimally utilize all physical cores, many visual quality improvements can be made by dedicating the additional computational power to video pre-processing.

These performance gains allow doubling of Ultra-High Definition channel density, assuming an HEVC codec. In addition, these new processors enable the implementation of the most demanding compression applications, such as Adaptive Bit Rate up to Ultra-High Definition on standard dual-socket servers. For example, 26 High-Definition channels (13 per socket) are now enabled, assuming an H.264/MPEG-4 AVC codec. The Intel Xeon Scalable processors pave the way to more demanding video applications.



¹ Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. Test and system configuration:

Baseline: One socket of a system based on 2S Intel® Xeon® processors E5 2697 v4, 2.3 GHz, 18 cores, Intel® Turbo Boost Technology and Intel® Hyper-Threading Technology enabled, BIOS "Dell Inc 2.3.3", 32 GB total memory, 8 slots / 4 GB / 2400 MT/s / DDR4 DIMM, 200 GB INTEL SSDSC1BG20, Debian® Linux® 8 kernel 3.16.7.

New: One socket of a system based on 2S Intel® Xeon® Platinum 8180 processors, 2.5GHz, 28 cores, Intel Turbo Boost Technology and Intel Hyper-Threading Technology enabled, BIOS "Intel Corporation SE5C620.86B.01.00.0412.020920172159", SMBIOS v2.8, 192 GB total memory, 12 slots / 16 GB / 2666 MT/s / DDR4 DIMM, 800 GB INTEL SSDSC2BA800G4, CentOS® Linux® 7.2.1511 kernel 3.10.0. All testing performed by ATEME in August 2017 using listed configurations. Intel does not control or audit third-party benchmark data or the web sites referenced in this document.

² 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.

³ Downloaded from <https://media.xiph.org/video/derf/ElFuente/> (publicly available).

⁴ Compilation of the code performed with g++ 4.7 and yasm 1.3. No specific flags.

⁵ Requires a system with Intel® Turbo Boost Technology capability. Performance varies depending on hardware, software, and system configuration. Learn more at <http://www.intel.com/go/turbo>.

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at intel.com/xeonscalable

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. All testing performed by ATEME in August 2017 using listed configurations.

Intel, the Intel logo, Intel Xeon, Intel® Advanced Vector Extensions, and Intel® Turbo Boost are trademarks of Intel Corporation in the U.S. and/or other countries.

Ateme and TITAN are registered trademarks of Ateme.

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