Solution Brief

Live Video Production
Edge Cloud Video Division

Optimized Real-Time Video Transport using Intel® Data Streaming Accelerator

4th Gen Intel® Xeon® Scalable processors enhance video processing workloads with improved per-core performance, memory and I/O, as well as built-in accelerators including Intel Data Streaming Accelerator (Intel® DSA). Intel® Media Transport Library uses Intel DSA and the Data Plane Development Kit (DPDK) to optimize video movement, increasing the number of video streams at 60 fps by 2.25X per core.[1]

Executive Summary

Digital video continues to evolve with new technologies and market trends. Increased resolutions and higher bit rates drive the need for a range of new connectivity standards, especially for video over Internet Protocol (IP). Successive generations of media transport standards must meet growing demands for high-bandwidth content such as live video production, 4K and 8K video streaming, with new immersive forms such as AR/VR and 360-degree video on the near horizon. Video infrastructure must meet these escalating requirements while delivering excellent viewer experience within a steady cost envelope to remain competitive.

Intel DSA is a high-performance data copy and transformation accelerator that is native within 4th Gen Intel Xeon Scalable processors. It optimizes streaming data movement and transformation operations common with applications for high-performance storage, networking, persistent memory and various data processing applications. Intel DSA can provide better performance for data movement and transformation operations, while freeing up CPU cycles for higher level functions. The hardware platform also benefits from intelligent, high-performance network connectivity provided by Intel® Ethernet 800 Series Network Adapters.

Advanced software optimizations and enablement help deliver high throughput and low latency for video transport workloads:

- **Intel Media Transport Library** streamlines ST 2110 implementation while meeting rigorous objectives for flexibility, scalability and cost. It is highly optimized for 4th Gen Intel Xeon Scalable processors, taking advantage of the platform's foundational architecture improvements as well as Intel DSA for hardware acceleration of streaming media.
- **DPDK** further accelerates throughput in solutions based on the Intel Media Transport Library by simplifying the packet processing pipeline.

This solution brief presents performance results that characterize the benefits of Intel DSA in concert with the Intel Media Transport Library. It demonstrates the value of migrating media distribution workloads to servers based on 4th Gen Intel Xeon Scalable processors.

Intel solutions for real-time video transport

Intel solutions for live video production, processing and distribution include a range of hardware and software ingredients, as well as extensive ecosystem enablement. This brief focuses on 4th Gen Intel Xeon Scalable processors with Intel DSA, enabled by software based on Intel Media Transport Library and DPDK.

Next-generation server platform for high throughput and low latency

4th Gen Intel Xeon Scalable processors dramatically improve performance, reduce latency and reduce costs with improvements across the balanced platform and the industry's largest collection of built-in hardware accelerators. Enhanced per-core execution resources and up to 60 cores per socket (up to 240 threads in a two-socket platform) drive high throughput with outstanding energy efficiency.

An enhanced memory subsystem provides up to eight DDR5 channels operating at up to 4800 MT/s, a 1.5x improvement in bandwidth and speed over predecessors[2] that enables each node to hold more content in DRAM, for higher stream density. Faster, higher capacity I/O based on up to 80 lanes of PCIe 5.0 per socket, compared to 64 lanes of PCIe 4.0 in the previous platform, helps drive more sessions out to subscribers.

The large numbers of software-based memory-copy operations required in video pipelines conventionally occupy significant processor core resources and incur latency that adversely affects quality. By offloading almost all memory movement operations (including checksum, memory compare and checkpointing), Intel DSA relieves the CPU cores of overhead associated with moving data in and out of memory, storage and networking subsystems. Intel DSA optimizes the handling of streaming data across the CPU, memory and cache, as well as all attached memory, storage and network devices. The Intel DSA accelerator engine is much faster than software, which improves latency, and it requires less energy, which helps reduce total cost of ownership (TCO).

Network I/O for the platform is provided by Intel® Ethernet Network Adapter E810-2CQDA2 which delivers up to 200Gbps of total bandwidth and packet-sorting optimizations for high bandwidth video content.

Intel Media Transport Library, Intel DSA and DPDK

To reduce the compute overhead associated with transmitting and receiving high-bandwidth video between nodes, Intel Media Transport Library provides a streamlined path to implementation on Intel architecture for high throughput, low latency and high reliability. As video is received from the network adapter, Intel Media Transport Library functions process the IP packets and copy the payload of uncompressed media to the raw video frame buffer.

Intel DSA provides efficient offloading of these copy operations from the CPU cores to accelerate media data movement and transformation. Intel DSA enables this high performance by utilizing data mover capability to/from volatile memory, persistent memory, memory-mapped I/O and through a Non-Transparent Bridge (NTB) device to/from remote volatile and persistent memory on another node in a cluster. Working in concert, Intel Media Transport Library and Intel DSA provide higher media throughput while freeing up CPU core resources for other work, including higher-level functions.

The Intel Media Transport Library utilizes the open source DPDK and Intel Ethernet 800 Series Network Adapters to accelerate the transfer of packets by eliminating much of the kernel processing, interrupt handling and data copying typically required when sending and receiving network data. This set of user-space libraries and drivers bypasses the OS kernel network stack, as shown in Figure 2, which avoids the overhead of copying data between user space and kernel space, substantially improving throughput.

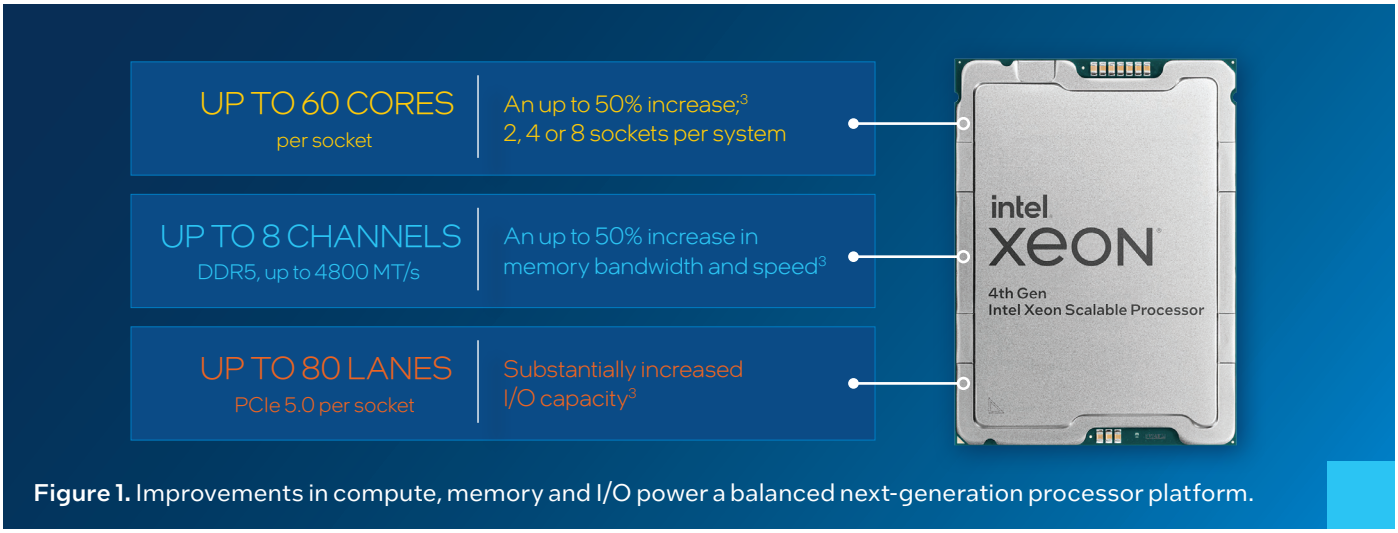

Figure 1. Improvements in compute, memory and I/O power a balanced next-generation processor platform.

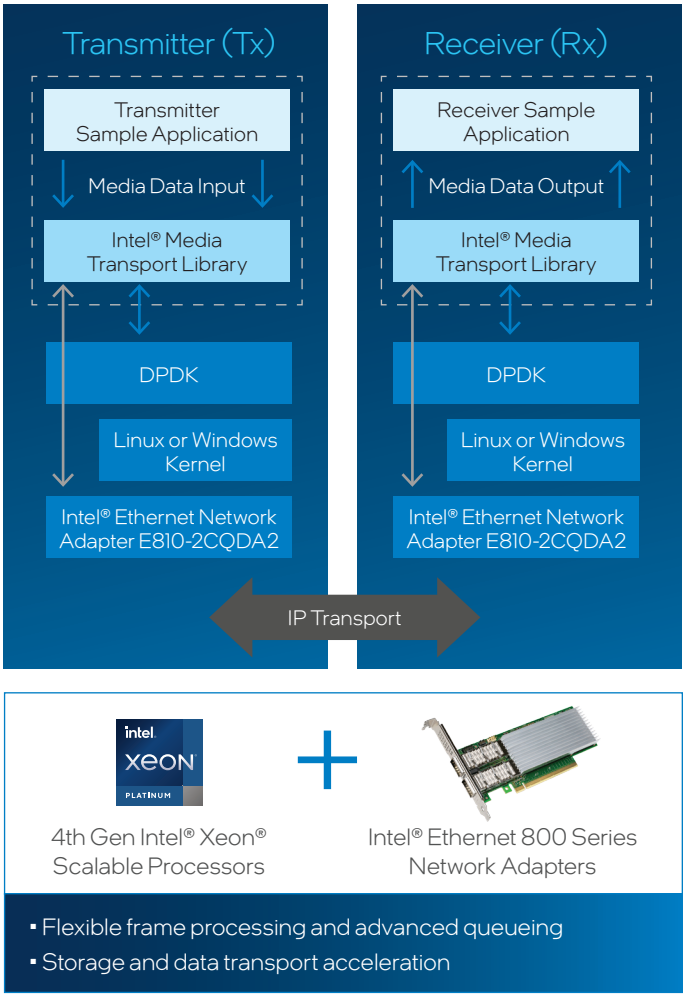

Figure 2. Accelerated IP transport on Intel architecture.

Real-time Video Transport Technology Requirements

The hardware and software reviewed here is designed to take full advantage of video acceleration capabilities based on Intel Media Transport Library using Intel DSA built into 4th Gen Intel Xeon Scalable processors. It addresses technology requirements for reduced latency and more efficient core resource usage to meet business requirements for live video production, processing and distribution with lower operating costs.

Optimizing cost with reduced hardware and bandwidth

Intel Media Transport Library interoperates with Intel Ethernet 800 Series Network Adapters to optimize multichannel IP video I/O workflows, meeting production needs for IP distribution. Ongoing refinement of the solution stack includes work to support more streams with less bandwidth and CPU core requirements, to reduce the cost of supporting multichannel inputs and outputs.

Meeting high-resolution and high-density requirements

The overhead of software-based bulk data-copy operations is a consistent bottleneck in the processing and delivery of high-bandwidth content such as 4K and 8K video in real-time, or during live production. Parsing video packets received from network adapters and copying the payload of uncompressed video to the raw video frame buffer commonly creates inefficiencies in existing solutions that must be addressed.

Video Throughput Performance

The team validated the benefits of Intel DSA for video throughput using a two-socket server based on high-end 4th Gen Intel Xeon Platinum processors, as described in Table 1.

Table 1. Hardware configuration for system under test.

Element	Description
Chassis	Intel rack-mount
CPU Model	Pre-production Intel® Xeon® Platinum 8490H processor
Sockets	2
Cores per Socket	60
Intel Hyper-Threading Technology	Enabled
Hardware Threads	240
Intel Turbo Boost Technology	Enabled
NUMA Nodes	2
Prefetchers	L2 HW, L2 Adj., DCU HW, DCU IP
Accelerators	QAT:8, DSA:8, IAA:8, DLB:8

Significant processor resources are consumed to copy the payload from RTP packets to the video frame buffer. Offloading copy operations using Intel DSA increases the number of 1080p video streams at 60 fps by using one core per NIC, as shown in Table 2 and Figure 3.[1]

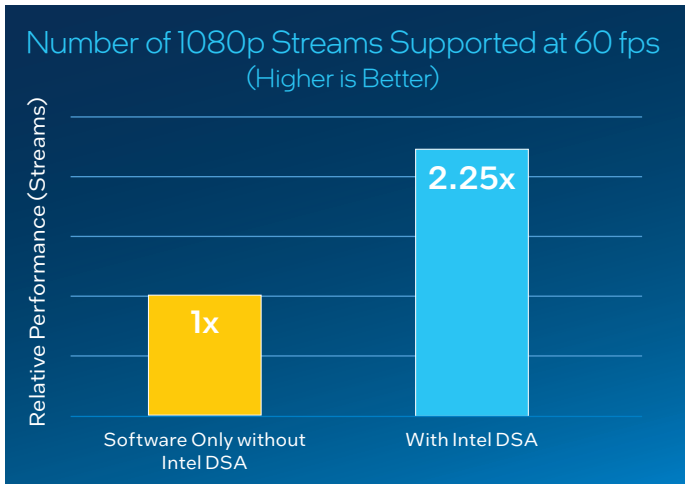

Figure 3. Rx sessions using one core per NIC.[1]

Table 2. Intel® Media Transport Library Rx sessions on single core.

CPU/Intel DSA Offload	Number of Network Adapters	Number of Cores per Network Adapter	1080p@60 # Sessions (max/core)	Total Max Sessions
CPU Cores Only	2	1	12	24 (2*12)
With Intel DSA	2	1	27	54 (2*27)

At the same time, the maximum network bandwidth can be achieved with fewer cores when offloaded to Intel DSA, as shown in Table 3 and Figure 4. Here, the team tested how many cores are required to maintain 200 Gbps network bandwidth using 54 streams of 1080p video content at 60 fps. In a CPU-only solution, six cores were needed. Also, the max sessions per core dropped down to 9 from 12 because the scalability of a CPU-only solution is deeply impacted by the availability of global system resources. For example, multiple cores use LLC and DDR bandwidth concurrently. With the addition of Intel DSA, the same workload throughput was maintained with just two cores, reflecting a 66% reduction in the number of cores.

Table 3. Intel® Media Transport Library Rx sessions on multiple cores.

CPU/Intel DSA Offload	Number of NICs	Number of Cores per NIC	1080p@60 # Sessions (max/core)	Total Max Sessions
CPU Cores Only	2	3	9	54 (2*3*9)
With Intel DSA	2	1	27	54 (2*1*27)

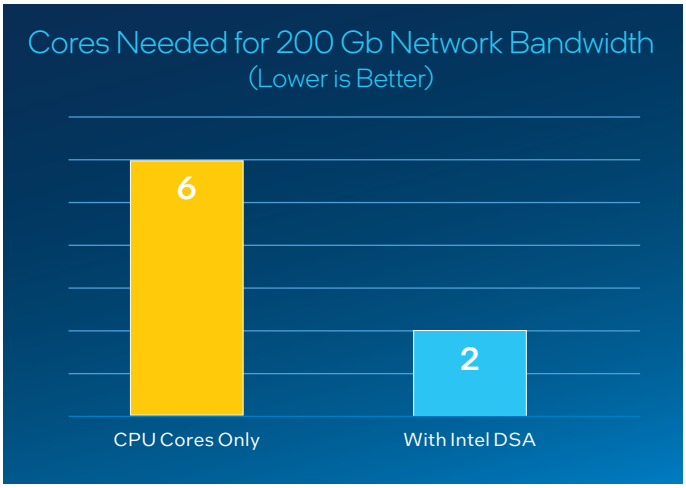

Figure 4. Rx sessions on multiple cores.

Conclusion

The throughput, latency and cost of video distribution pipelines can be improved using the combination of 4th Gen Intel Xeon Scalable processors with DPDK, Intel Ethernet 800 Series Network Adapters and Intel Media Transport Library. Implementing standards-based video transport over IP networks, Intel Media Transport Library takes advantage of Intel DSA, a built-in hardware accelerator for the memory-copy operations prevalent in streaming-media workloads. Together, these building blocks can enable each node to improve network bandwidth and provide leaner operations for video workloads.

Authors:

Eric Walker (eric.walker@intel.com)

Gordon Kelly (gordon.kelly@intel.com)

Ping Yu (ping.yu@intel.com)

Frank Du (frank.du@intel.com)

Jing Z Wang (jing.z.wang@intel.com)

Mrittika Ganguli (mrittika.ganguli@intel.com)

Raul Diaz (raul.diaz@intel.com)

Learn More

www.intel.com/xeon/scalable

www.intel.com/content/www/us/en/broadcast/overview.html

[1] BASELINE: Test by Intel as of 18th July 2022,1-node, pre-production 2x Intel® Xeon® Platinum 8490H, 60 cores, HT On, Turbo On, Total Memory, 224GB (14x16GB 4800 MT/s [4800 MT/s]),BIOS versionEGSDCRB1.86B.0083.D22.2206290535, 06/29/2022, ucode 0xaa0000a0, CentOS Stream 8, 5.15.0-spr.bkc.pc.8.8.5.x86_64,compiler gcc (GCC) 8.5.0 20210514 (Red Hat 8.5.0-13),Kahawai 22.06,score 24 1080p streams @60FPS.

NEW-1: Test by Intel as of 18th July 2022,1-node, pre-production 2x Intel® Xeon® Platinum 8490H , 60 cores, HT On, Turbo On, Total Memory, 224GB (14x16GB 4800 MT/s [4800 MT/s]),BIOS versionEGSDCRB1.86B.0083.D22.2206290535, 06/29/2022, ucode 0xaa0000a0, CentOS Stream 8, 5.15.0-spr.bkc.pc.8.8.5.x86_64,compiler gcc (GCC) 8.5.0 20210514 (Red Hat 8.5.0-13),Kahawai 22.06, DSA driver and SW BKC 60 --idxd driver 3.4.6.4.git873c591f , score 54 1080p streams @60FPS.

[2] 1.5x improvement based on prior gen DDR4 DRAM @ 3200 MT/s vs. current gen DDR5 DRAM @ 4800 MT/s.

[3] Compared to the previous generation.

Availability of accelerators varies depending on SKU. Visit the Intel Product Specifications page for additional product details.

Performance varies by use, configuration and other factors. Learn more at https://www.intel.com/PerformanceIndex.

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

Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

Your costs and results may vary.

Intel technologies may require enabled hardware, software or service activation.

You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Intel products described herein. You agree to grant Intel a nonexclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

0723/DL/MESH/350527-002US