WHITE PAPER
Intel® Advanced Vector Extensions
Processor Performance
September 2014

Optimizing Performance with Intel® Advanced Vector Extensions

Intel® Advanced Vector Extensions 2 instructions can provide significant performance increases.

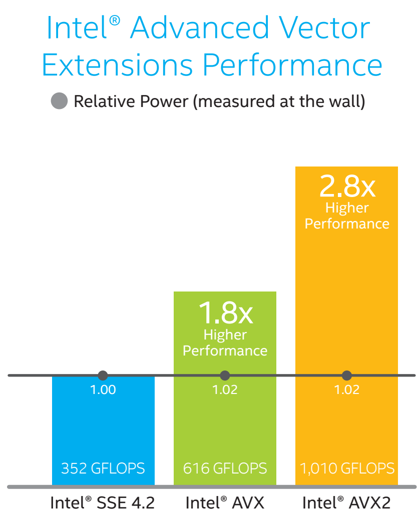


Figure 1. Measuring performance on the same processor using Linpack* benchmarks shows substantial increases from Intel® Streaming SIMD Extensions 4.2 (Intel® SSE 4.2) to Intel® Advanced Vector Extensions (Intel® AVX) and from Intel AVX to Intel® AVX2, with up to 2.8x the GFLOPS throughput when comparing Intel SSE 4.2 to Intel AVX2.[2]

Gregory Lento
Performance Marketing Engineer, Intel

Executive Summary

The release of the Intel® Xeon® processor E5 v3 family includes the Intel® Advanced Vector Extensions 2 (Intel® AVX2) instruction set, which can provide significant performance improvement over Intel® Advanced Vector Extensions (Intel® AVX)[1] and Intel® Streaming SIMD Extensions (Intel® SSE). The benefits include doubling the number of FLOPS (floating-point operations per second) per clock cycle, 256-bit integer instructions, floating-point fused multiply-add instructions, and gather operations.

Intel AVX instructions[¤] require more power to run. When executing these instructions, the processor may run at less than the marked frequency to maintain thermal design power (TDP) limits.

Beginning with the Intel Xeon processor E5 v3 family, Intel will provide two sets of frequency specifications:

- **Marked TDP and max all core turbo.** These are the frequencies that have traditionally been specified. Workloads that do not utilize Intel AVX instructions will operate in this frequency range.
- **AVX base and AVX max all core turbo.** These are the new frequencies for workloads that utilize Intel AVX instructions. These frequencies are defined in the "Intel® Xeon® Processor E5-1600 and E5-2600 v3 Product Families Specification Update."

Intel® Turbo Boost Technology continues to provide opportunistic frequency increases based on workload, number of active cores, temperature, power, and current. If there is power and thermal headroom, Intel AVX workloads could still opportunistically run at higher turbo frequencies. The actual frequency achieved will be opportunistic based on the factors listed above.

Because Intel AVX instructions generally consume more power, frequency reductions can occur to keep the processor operating within TDP limits. Intel is including additional AVX base and turbo frequency specifications to provide more clarity for these Intel AVX instructions. Performance of workloads optimized for Intel AVX instructions can be significantly greater than workloads that do not use Intel AVX instructions even when the processor is operating at a slightly lower frequency (see Figure 1).

[¤] Throughout this paper, "Intel AVX" refers to IntelAVX, Intel AVX2, and future versions of Intel AVX.

Table of Contents

History of Intel® SSE, Intel® AVX, and Intel® AVX2[3]

1999
Intel introduced Intel® Streaming SIMD Extensions (Intel® SSE), an SIMD (single instruction, multiple data) instruction set extension to the x86 architecture for Intel® processors. Intel SSE instructions work primarily on single-precision floating-point data, increasing performance when the same operations occur on multiple datasets. Typical applications are digital signal processing and graphics processing.

2011
Intel introduced Intel® Advanced Vector Extensions (Intel® AVX), a 256-bit instruction set extension to Intel SSE that is designed for applications that are floating-point intensive. Intel AVX improves performance due to wider vectors, new extensible syntax, and rich functionality. This performance improvement results in better management of data in general purpose applications such as image and audio/video processing, scientific simulations, financial analytics, and 3D modeling and analysis.

2013
Intel introduced Intel® Advanced Vector Extensions 2 (Intel® AVX2), instructions that expand integer data types to 256-bit SIMD. Intel AVX2 integer support is particularly useful for processing the visual data commonly encountered in consumer imaging and video-processing workloads.

Introducing Intel® Advanced Vector Extensions 2

Intel® processors can provide acceleration through the use of SIMD (single instruction, multiple data) instructions that include Intel® Streaming SIMD Extensions (Intel® SSE) instructions and Intel® Advanced Vector Extensions (Intel® AVX)[4] instructions. Typical uses for these instruction sets are image processing and complex algorithms (see the sidebar).

In 2014, Intel introduced the Intel® Xeon® processor E5-2600 v3 family, which includes the Intel® Advanced Vector Extensions 2 (Intel® AVX2) instruction set. Intel AVX2 extends Intel SSE and Intel AVX with 256-bit integer instructions and also adds support for floating-point fused multiply-add instructions, and gather operations. Intel AVX2 doubles the number of double-precision FLOPS (floating-point operations per second) per clock cycle, theoretically doubling the core's peak floating-point throughput (see Table 1).

Intel AVX[5] [¥] is designed to achieve higher throughput for certain integer and floating-point operations. Using these instructions may cause processors to operate at less than the marked TDP frequency. These reductions in frequency occur because high-power Intel AVX instructions require additional voltage and electrical current.

When the processor detects Intel AVX instructions, additional voltage is applied to the core. With the additional voltage applied, the processor could run hotter, requiring the operating frequency to be reduced to maintain operations within the TDP limits. The higher voltage is maintained for 1 millisecond after the last Intel AVX instruction completes, and then the voltage returns to the nominal TDP voltage level.

[¥] Subsequently in this paper, "Intel AVX" will refer to Intel AVX, Intel AVX2, and future versions of Intel AVX.

Table 1. Comparison of FLOPS Performance by Instruction Set

INSTRUCTION SET	SINGLE-PRECISION FLOPS PER CLOCK	DOUBLE-PRECISION FLOPS PER CLOCK
Intel® Streaming SIMD Extensions 4.2	8	4
Intel® Advanced Vector Extensions	16	8
Intel® Advanced Vector Extensions 2	**32**	**16**

New Intel® Advanced Vector Extensions Frequencies

Historically, Intel has specified a marked TDP frequency and a turbo frequency for all workloads. Now Intel is adding two new AVX frequencies:

- **AVX base** is the minimum frequency for workloads using Intel AVX instructions.
- **AVX max all core turbo** is the maximum frequency for workloads using Intel AVX instructions.

These new frequencies are defined in the "Intel® Xeon® Processor E5-1600 and E5-2600 v3 Product Families Specification Update."

As shown in Figure 2, Intel expects the following results:[6]

- Workloads using Intel AVX instructions may reduce processor frequency as far down as the AVX base frequency to stay within TDP limits.
- Some workloads that utilize Intel AVX instructions could achieve a turbo frequency above the AVX base frequency up to AVX max all core turbo.
- Workloads that contain no Intel AVX instructions could operate at the marked TDP frequency up to the max all core turbo frequency (non-AVX).

Workloads that utilize a very high percentage of Intel AVX2 instructions may operate closer to the AVX base frequency. Because Intel AVX2 doubles the number of FLOPS per clock, workloads that utilize a high percentage of Intel AVX2 instructions should still see a significant performance increase compared to workloads that use Intel AVX instructions running at the marked TDP frequency on previous generation processors (see Figure 3).

Frequency Range Comparison

FOR ILLUSTRATIVE PURPOSES ONLY

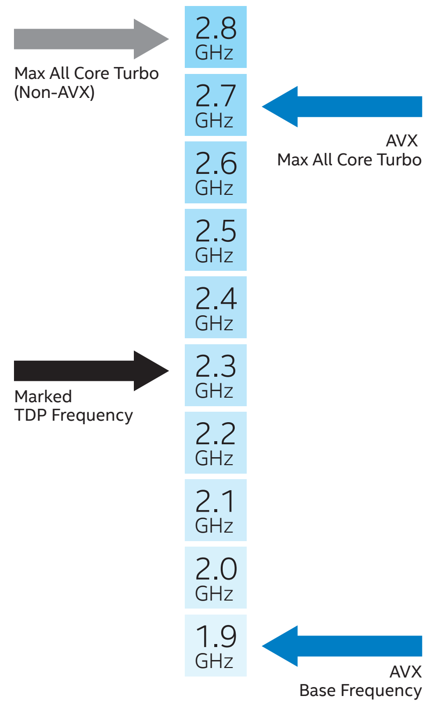
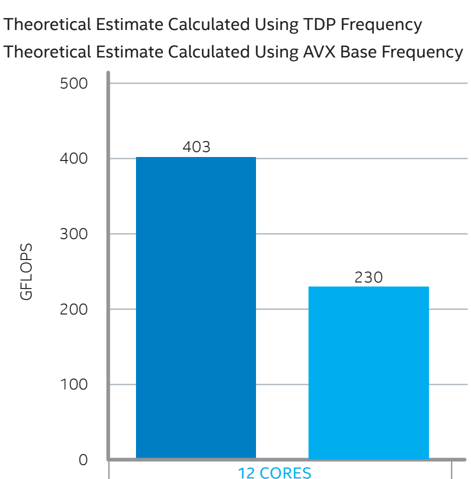


Figure 2. The new Intel® Advanced Vector Extensions (Intel® AVX) instructions could operate at or below the marked TDP frequency, while still providing up to two times the floating-point throughput.

Double-Precision Theoretical Peak FLOPS Comparison

Intel® Xeon® Processor E5 2600 v3 Calculated at the AVX Base Frequency Versus Intel® Xeon® Processor E5 2600 v2 Calculated at TDP Frequency

HIGHER IS BETTER



	12 CORES	
	Intel® Xeon® processor E5-2680 v3	Intel® Xeon® processor E5-2695 v2
Marked TDP Frequency (GHz)	2.5	2.4
AVX Base Frequency	**2.1 GHz**	**N/A**

UP TO A **1.7x** INCREASE IN PEAK GFLOPS

Figure 3. Calculating the theoretical peak FLOPS using the AVX base frequency results in up to a 1.7x increase in peak GFLOPS on Intel® Xeon® processors using Intel AVX2 over the value calculated using the TDP frequency on a previous generation processor with the same core count and similar frequency using Intel AVX.[7]

Intel® Turbo Boost Technology

Intel® Turbo Boost Technology 2.0[8] automatically and opportunistically enables processor cores to run faster than the marked TDP frequency if the processor is operating below power, current, and temperature specification limits. The availability and amount of turbo frequency depends on the workload and operating environment, specifically the following factors:

- Processor SKU
- Number of active cores in the C0 state (normal operation)
- Type of workload
- Estimated current consumption
- Estimated power consumption
- Processor temperature

Intel Turbo Boost Technology Opportunistic Frequency Upside

Intel Turbo Boost Technology frequencies are opportunistic and not guaranteed. Due to varying power characteristics, some parts with Intel Turbo Boost Technology 2.0 may not achieve maximum turbo frequencies when running heavy workloads and using multiple cores concurrently.

The behavior of Intel Turbo Boost Technology may have a positive impact on application performance as it provides opportunistic frequency upside above the marked TDP frequency when conditions allow.

DETERMINISTIC FREQUENCY

For customers who need deterministic processor frequency, the following options are available:

- Disable turbo boost in BIOS (contact the system supplier for instructions).
- Use the `TURBO_RATIO_LIMIT` Model Specific Register (MSR) to control the amount of turbo granted. Refer to the processor data sheet for additional information.
- Use the `UNCORE_RATIO_LIMIT` Model Specific Register (MSR) to run at a fixed uncore frequency. Refer to the processor data sheet for additional information.

Best Practices

Intel recommends using workloads optimized for Intel AVX because they can deliver substantial performance gains over non-AVX-enabled workloads. Developers should use dense, highly optimized Intel AVX code and avoid scattered instances.

To calculate theoretical peak FLOPS, Intel recommends using the marked TDP frequency as it best represents the frequency the processor could operate at over a range of workloads that utilize Intel AVX instructions. Customers may choose to use the AVX base frequency to calculate FLOPS if they think it better represents their system behavior.

For example, Linpack runs closer to the AVX base frequency on the Intel Xeon processor E5 v3 family, so using the Intel AVX base frequency to calculate theoretical peak FLOPS will more accurately represent Linpack efficiency.

Conclusion

Intel AVX2-enabled workloads could provide significant performance benefits over workloads using Intel SSE instructions, Intel AVX, and non-AVX instructions, theoretically doubling the peak floating-point throughput. Because Intel AVX instructions generally consume more power, frequency reductions can occur to keep the processor operating within TDP limits. Intel is including additional AVX base and turbo frequency specifications to provide more clarity for these Intel AVX instructions.

When compared to workloads that do not make use of Intel AVX instructions, workloads that are optimized for Intel AVX instructions could see significant increases in performance. Intel recommends using Intel AVX instructions where it makes sense for the specific workload, using performance measurements to assess the benefits.

Frequently Asked Questions about the New Intel® Advanced Vector Extensions (Intel® AVX) Frequencies

QUESTION	ANSWER
Why are new frequencies being defined on the Intel® Xeon® processors E5 v3 family?	Since Intel AVX instructions generally consume more power, which may cause the operating frequency of the processor to be reduced, additional AVX base and turbo frequency specifications are now included to provide more clarity of the expected processor frequency when running workloads that use Intel AVX instructions.
Do workloads using Intel AVX and Intel AVX2 provide performance benefits?	Yes, workloads using Intel AVX instructions may result in significant performance benefits compared to workloads not using Intel AVX instructions because they can execute more double-precision instructions per clock cycle (see Table 1. Comparison of FLOPS Performance by Instruction Set).
Why do workloads that utilize Intel AVX instructions run at a lower frequency than workloads that do not use Intel AVX instructions?	Intel AVX instructions require more power to run, which increases the temperature of the processor. This means that the processor frequency may need to be reduced to stay within thermal design power (TDP) limits.
Why did prior generation Intel® Xeon® processors that utilize Intel AVX instructions not run at a lower frequency?	Previous generation processors only support Intel AVX instructions. Intel AVX2 instructions consume additional power, which could result in a lower operating frequency.
Will workloads not using Intel AVX instructions also operate at a reduced frequency?	Workloads not using Intel AVX instructions will continue to operate at or above the TDP frequency.
Will running a small number of Intel AVX instructions reduce frequency below the regular marked frequency?	No, frequency will be reduced below the regular marked frequency only if a real power or thermal constraint is reached, not just due to the presence of Intel AVX instructions. Some workloads that utilize Intel AVX instructions could still achieve turbo above the marked TDP frequency.
What frequency will my processor operate at?	The frequency your processor will operate at will vary depending on number of active cores (in C0 state), the temperature of the part, the power consumption of the part, and the workload. Workloads that utilize Intel AVX instructions are more likely to operate at a reduced frequency to maintain TDP limits.
Why is Linpack* efficiency on the Intel® Xeon® processor E5 v3 family lower than expected compared to previous generation Intel Xeon processors?	On previous generation Intel Xeon processors, Linpack ran closer to the marked TDP frequency. Using the marked TDP frequency for theoretical peak FLOPS calculation gave an accurate representation of Linpack efficiency. With Intel® Xeon® processors E5 v3 family, Linpack optimized for Intel AVX2 runs closer to the AVX base frequency. Using the AVX base frequency to calculate theoretical peak FLOPS will likely give a more accurate representation of Linpack efficiency.

For more information about Intel® AVX2, visit **software.intel.com/en-us/intel-isa-extensions.**
For more information about Intel Turbo Boost Technology, visit **www.intel.com/go/turbo.**

[1,4] Intel® Advanced Vector Extensions (Intel® AVX) are designed to achieve higher throughput for certain integer and floating point 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 should consult your system manufacturer for more information.
* Intel® Advanced Vector Extensions refers to Intel® AVX, Intel® AVX2 or Intel® AVX-512. For more information on Intel® Turbo Boost Technology 2.0, visit www.intel.com/go/turbo.

[2] Benchmark: MP_LINPACK* 11.1.3 problem size N=120000.

Configurations: Intel® Wildcat Pass platform with two Intel® Xeon® processor E5-2699 v3 (18 cores, 2.3 GHz, 145 W), 1 x 1TB 7200RPM 64-MB cache. SATAIII HDD, 128-GB memory (8x 16 GB 2Rx4 DDR4-2133 RDIMM), Red Hat Enterprise Linux* 6.5 (kernel 2.6.32-431.el6.x86_64) MP LINPACK 11.1.3 on Linux AVX2 version. HT disabled, turbo enabled, HS snoop mode, Prefetch enabled, EIST enabled, C-state disabled, GRNDSDP1.86B.0032.R02.1405090848. Source: Intel internal testing as of July 2014. Score: 1009.87 GFLOPS. Power: 506W.

Intel Wildcat Pass platform with two Intel Xeon processor E5-2699 v3 (18 cores, 2.3 GHz, 145 W), 1 x 1TB 7200RPM 64-MB cache. SATAIII HDD, 128-GB memory (8x 16GB 2Rx4 DDR4-2133 RDIMM), Red Hat Enterprise Linux 6.5 (kernel 2.6.32-431.el6.x86_64) MP LINPACK 11.1.3 on Linux AVX version. HT disabled, turbo Enabled, HS snoop mode, Prefetch enabled, EIST enabled, C-state disabled, GRNDSDP1.86B.0032.R02.1405090848. Source: Intel internal testing as of July 2014. Score: 616.14 GFLOPS. Power: 504W.

Intel Wildcat Pass platform with two Intel Xeon processor E5-2699 v3 (18 cores, 2.3 GHz, 145 W), 1 x 1TB 7200RPM 64-MB cache. SATAIII HDD, 128-GB memory (8x 16GB 2Rx4 DDR4-2133 RDIMM), Red Hat Enterprise Linux 6.5 (kernel 2.6.32-431.el6.x86_64) MP LINPACK 11.1.3 on Linux SSE version. HT disabled, turbo Enabled, HS snoop mode, Prefetch enabled, EIST enabled, C-state disabled, GRNDSDP1.86B.0032.R02.1405090848. Source: Intel internal testing as of July 2014. Score: 352.15 GFLOPS. Power: 493W.

For more information go to www.intel.com/performance.

[3,6] Intel Instruction Set Architecture Extensions. software.intel.com/en-us/intel-isa-extensions.

[5] Availability and frequency upside of Intel Turbo Boost Technology 2.0 state depends upon a number of factors including, but not limited to the type of workload, number of active cores, estimated current consumption, estimated power consumption and processor temperature.

[7] Results were calculated for each processor based on the following formula: Clock frequency × number of cores × numbers of double-precision FLOPS per clock = theoretical peak FLOPS. These results are for informational purposes only. Any difference in formulas applied, or system hardware or software design or configuration may affect results.

[8] Requires a system with Intel® Turbo Boost Technology. Intel Turbo Boost Technology and Intel Turbo Boost Technology 2.0 are only available on select Intel® processors. Consult your system manufacturer. Performance varies depending on hardware, software, and system configuration. For more information, visit www.intel.com/go/turbo.

intel® Look Inside.™