



The Converged Vector ISA: Intel[®] Advanced Vector Extensions 10

Technical Paper

December 2023

Revision 2.0

Order Number: 356368-002US

Notices & Disclaimers

This document contains information on products in the design phase of development. The information here is subject to change without notice. Do not finalize a design with this information.

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

No product or component can be absolutely secure.

Results have been estimated or simulated.

Your costs and results may vary.

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 non-exclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

All product plans and roadmaps are subject to change without notice.

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.

Intel disclaims all express and implied warranties, including without limitation, the implied warranties of merchantability, fitness for a particular purpose, and non-infringement, as well as any warranty arising from course of performance, course of dealing, or usage in trade.

Code names are used by Intel to identify products, technologies, or services that are in development and not publicly available. These are not "commercial" names and not intended to function as trademarks.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document, with the sole exception that a) you may publish an unmodified copy and b) code included in this document is licensed subject to the Zero-Clause BSD open source license (0BSD), <https://opensource.org/licenses/0BSD>. You may create software implementations based on this document and in compliance with the foregoing that are intended to execute on the Intel product(s) referenced in this document. No rights are granted to create modifications or derivatives of this document.

© Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.

CONTENTS

PAGE

CHAPTER 1

CONVERGED VECTOR ISA: INTEL® ADVANCED VECTOR EXTENSIONS 10

1.1	BACKGROUND	1-1
1.2	INTRODUCTION TO INTEL® AVX10	1-1
1.3	ENUMERATION	1-2
1.4	PERFORMANCE BENEFITS	1-2
1.5	AVAILABILITY	1-3
1.6	CONCLUSION	1-3

FIGURES

Figure 1-1. Intel® AVX-512 Feature Flags Across Intel® Xeon® Processor Generations vs. Intel® AVX10 1-2
Figure 1-2. Intel® ISA Families and Features 1-3

Revision History

Revision Number	Description	Date
1.0	Initial release of the document.	June 2023
2.0	Updated Section 1.2, "Introduction to Intel [®] AVX10," to remove the 32-bit mask register limitation.	December 2023

CHAPTER 1

CONVERGED VECTOR ISA: INTEL® ADVANCED VECTOR EXTENSIONS 10

Intel® Advanced Vector Extensions 10 (Intel® AVX10) introduces a modern vector Instruction Set Architecture (ISA) that will be supported across future Intel® processors. This new ISA includes all the richness of the Intel® Advanced Vector Extensions 512 (Intel® AVX-512) with additional features and capabilities enabling it to seamlessly run across Performance-cores and Efficient-cores, delivering performance and consistency across all platforms. It also introduces a new enumeration approach based on version and supported vector lengths, reducing the burden on the developer to check multiple feature bits for the platform. Intel AVX10 extends and enhances the capabilities of Intel AVX-512 to benefit all Intel® products and will be the vector ISA of choice moving into the future.

1.1 BACKGROUND

In 2016, Intel launched a major update to its vector instruction set with the launch of a high-performance vector ISA named Intel Advanced Vector Extensions 512 (Intel AVX-512). The Intel AVX-512 ISA included several new features and capabilities over the Intel® Advanced Vector Extensions 2 (Intel® AVX2) ISA including 512-bit vector registers, a discrete feature enumeration methodology, 16 additional vector registers, 8 mask registers, 512-bit vector length embedded rounding, and a large suite of new instructions. Over time, Intel AVX-512 evolved to include support for shorter vector length versions of instructions (128 and 256 bits) along with many additional instructions, each with its own CPUID feature flag, driving performance and capabilities for Performance-core (P-core) targeted vector workloads.

The Intel® AVX family of instruction sets (Intel AVX, Intel AVX2, and Intel AVX-512) have successfully gained wide industry adoption for a variety of applications including video processing, cryptography, HPC, AI, gaming, and others. Building on this momentum, Intel is announcing the next generation Intel AVX10 as the standard for ISA, supported by our future Efficient-cores (E-cores) and Performance-cores (P-cores). Intel AVX10 will enable the ecosystem to seamlessly integrate solutions across products and platforms and innovate for future generations of our products for years to come.

1.2 INTRODUCTION TO INTEL® AVX10

Today we are announcing the most impactful vector ISA evolution since the introduction of Intel AVX-512: Intel Advanced Vector Extensions 10 (Intel AVX10). Intel AVX10 includes all the capabilities and features of the Intel AVX-512 ISA, both for processors that feature 256-bit maximum vector register sizes, as well as for processors that feature 512-bit vector registers. In addition, this ISA includes several new capabilities and supports a new enumeration scheme that reduces the number of CPUID feature flags needing to be checked for feature support. Intel AVX10 is designed to run on future Intel P-core and E-core-based processors, allowing applications to seamlessly move across platforms.

There are three motivating factors for Intel AVX10:

1. To continue to support a high performance, vector ISA with all the richness of features of the existing Intel AVX-512 ISA.
2. To create a converged vector ISA based on Intel AVX-512 that will be supported on all future Intel processors.
3. To ease the developer task of verifying CPUID feature support.

The converged version of the Intel AVX10 vector ISA will include Intel AVX-512 vector instructions with an AVX512VL feature flag, a maximum vector register length of 256 bits, **eight 64-bit mask registers**, and new versions of 256-bit instructions supporting embedded rounding. This converged version will be supported on both P-cores and E-cores. While the converged version is limited to a maximum 256-bit vector length, Intel AVX10 itself is not limited to 256 bits, and optional 512-bit vector use is possible on supporting P-cores. Thus, Intel AVX10 carries forward all the benefits of Intel AVX-512 from the Intel® Xeon® with P-core product lines, supporting the key instructions, vector and mask register lengths, and capabilities that have comprised the ISA to date. Future P-

core based Xeon processors will continue to support all Intel AVX-512 instructions ensuring that legacy applications continue to run without impact.

1.3 ENUMERATION

The developer community has provided feedback that the current Intel AVX-512 enumeration method has become increasingly unwieldy over time. As new instructions were introduced, they were assigned a new CPUID feature flag that would need to be checked to determine processor support. As of future Intel Xeon processors with P-cores, codenamed Granite Rapids, there are expected to be more than 20 discrete Intel AVX-512 feature flags. To address this, Intel AVX10 introduces a new versioning approach to enumeration: a Vector ISA feature bit specifying Intel AVX10 support, an Intel AVX10 ISA Version Number, and three bits enumerating 128-, 256-, and 512-bit vector length support in the product.

The Intel AVX10 ISA Version Number will be inclusive and monotonically increasing. A developer can expect that Intel AVX10 Version N+1 will include all the features and capabilities included in Version N. With the stated goal of minimizing developer impact, a new version of the Intel AVX10 ISA can be expected to include a significant suite of new instructions and capabilities, delivering sufficient additional value to justify the associated software enablement effort. In rare cases, a discrete CPUID feature flag may be allocated for a segment-specific feature or in the case of an interim launch in between new Intel AVX10 versions.

The Intel AVX-512 ISA will be frozen as of the introduction of Intel AVX10 and all CPUID feature flags will continue to be enabled on future P-core processors for legacy support. All new subsequent vector instructions will be enumerated only as part of Intel AVX10. Apart from a few special cases, those instructions will be supported at all vector lengths, with 128-bit and 256-bit vector lengths being supported across all processors, and 512-bit vector lengths additionally supported on P-core processors.

							AVX10
					AVX512_FP16	AVX512_FP16	
				AVX512_VPOPCNTDQ, AVX512_VBMI2, VAES, GFNI, VPCLMULQDQ, AVX512_BITALG	AVX512_VPOPCNTDQ, AVX512_VBMI2, VAES, GFNI, VPCLMULQDQ, AVX512_BITALG	AVX512_VPOPCNTDQ, AVX512_VBMI2, VAES, GFNI, VPCLMULQDQ, AVX512_BITALG	
			AVX512_BF16	AVX512_BF16	AVX512_BF16	AVX512_BF16	
		AVX512_VNNI	AVX512_VNNI	AVX512_VNNI	AVX512_VNNI	AVX512_VNNI	
	AVX512_VBMI, AVX512_IFMA	AVX512_VBMI, AVX512_IFMA	AVX512_VBMI, AVX512_IFMA	AVX512_VBMI, AVX512_IFMA	AVX512_VBMI, AVX512_IFMA	AVX512_VBMI, AVX512_IFMA	
AVX512F, AVX512CD, AVX512BW, AVX512DQ	AVX512F, AVX512CD, AVX512BW, AVX512DQ	AVX512F, AVX512CD, AVX512BW, AVX512DQ	AVX512F, AVX512CD, AVX512BW, AVX512DQ	AVX512F, AVX512CD, AVX512BW, AVX512DQ	AVX512F, AVX512CD, AVX512BW, AVX512DQ	AVX512F, AVX512CD, AVX512BW, AVX512DQ	
AVX/AVX2		AVX/AVX2		AVX/AVX2		AVX/AVX2	
Intel® Xeon® Scalable Processors	Intel® Core™ Processors	2nd Gen Intel® Xeon® Scalable Processors	3rd Gen Intel® Xeon® Scalable Processors for 4S/8S Servers	3rd Gen Intel® Xeon® Scalable Processors	4th Gen Intel® Xeon® Scalable Processors	5th Gen Intel® Xeon® Scalable Processors	Future Intel P-core and E-core Processors

Figure 1-1. Intel® AVX-512 Feature Flags Across Intel® Xeon® Processor Generations vs. Intel® AVX10

1.4 PERFORMANCE BENEFITS

In addition to the previously stated usability benefits, several additional performance-based benefits of Intel AVX10 include:

- Intel AVX2-compiled applications, re-compiled to Intel AVX10, should realize performance gains without the need for additional software tuning.
- Intel AVX2 applications sensitive to vector register pressure will gain the most performance due to the 16 additional vector registers and new instructions.
- Highly-threaded vectorizable applications are likely to achieve higher aggregate throughput when running on E-core-based Intel Xeon processors or on Intel® products with performance hybrid architecture.

Existing Intel AVX-512 applications, many of them already using maximum 256-bit vectors, should see the same performance when compiled to Intel AVX10/256 at iso-vector length. For applications that can leverage greater

vector lengths, Intel AVX10/512 will be supported on Intel P-cores, continuing to deliver the best-in-class performance for AI, scientific, and other high-performance codes. New Intel® AVX10 libraries, compilers, and tool support will also be provided to help application developers realize the best achievable performance for all vector lengths and processor targets.

1.5 AVAILABILITY

Intel AVX10 Version 1 will be introduced for early software enablement and supports the subset of all the Intel AVX-512 instruction set available as of future Intel Xeon processors with P-cores, codenamed Granite Rapids, that is forward compatible to Intel AVX10. This version will not include the new 256-bit vector instructions supporting embedded rounding or any of the new instructions and will serve as the transition base version from Intel AVX-512 to Intel AVX10.

Intel AVX10 Version 2 will include the 256-bit instruction forms supporting embedded rounding as well as a suite of new Intel AVX10 instructions covering new AI data types and conversions, data movement optimizations, and standards support. All new instructions will be supported at 128-, 256-, and 512-bit vector lengths with limited variances. All Intel AVX10 versions will implement the new versioning enumeration scheme.

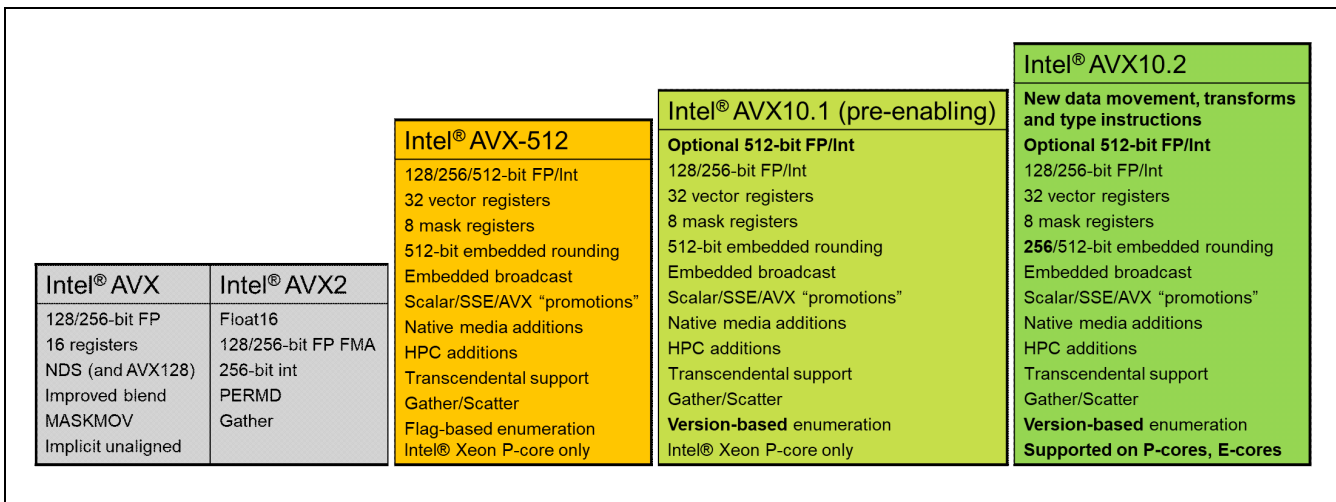


Figure 1-2. Intel® ISA Families and Features

1.6 CONCLUSION

Intel AVX10 represents a major shift to supporting a high-performance vector ISA across future Intel processors. It allows the developer to maintain a single code-path that achieves high performance across all Intel platforms with the minimum of overhead checking for feature support. Future development of the Intel AVX10 ISA will continue to provide a rich, flexible, and consistent environment that optimally supports both Server and Client products.