



3rd Generation Intel® Xeon® Scalable Processor Family

(processors based on Ice Lake microarchitecture)

Instruction Throughput and Latency README

June 2021

Revision 1.0

Legal Notices and Disclaimers

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at Intel.com, or from the OEM or retailer.

No computer system can be absolutely secure. Intel does not assume any liability for lost or stolen data or systems or any damages resulting from such losses.

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.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

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.

This document contains information on products, services and/or processes in development. All information provided here is subject to change without notice. Contact your Intel representative to obtain the latest Intel product specifications and roadmaps.

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.

Warning: Altering PC clock or memory frequency and/or voltage may (i) reduce system stability and use life of the system, memory and processor; (ii) cause the processor and other system components to fail; (iii) cause reductions in system performance; (iv) cause additional heat or other damage; and (v) affect system data integrity. Intel assumes no responsibility that the memory, included if used with altered clock frequencies and/or voltages, will be fit for any particular purpose. Check with memory manufacturer for warranty and additional details.

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. For more complete information about performance and benchmark results, visit <http://www.intel.com/performance>.

Results have been estimated or simulated using internal Intel analysis or architecture simulation or modeling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your actual performance.

Copies of documents which have an order number and are referenced in this document may be obtained by calling 1-800-548-4725 or by visiting www.intel.com/design/literature.htm.

Intel, Xeon, and the Intel logo are trademarks of Intel Corporation in the U. S. and/or other countries.

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

Copyright © 2021, Intel Corporation. All Rights Reserved.

README

Review all information in this README file prior to using the "Throughput and Latency" csv file.

File format:

Iform - "Iform" is the XED term for variants of instructions. Please consider them experimental and subject to change.

Xed url: <https://intelxed.github.io/>

Example:

VADDPD_XMMdq_XMMdq_MEMdq

Instruction name: VADDPD

Instruction sources: src0 XMM, src1 XMM, src2 MEM

Data element size: dq – 128-bit

Regsize - source register size

Mask - this AVX512 instruction uses a mask (k register)

Throughput definition –

The number of clock cycles required to wait before the issue ports are free to accept the same instruction again.

This number can be lower than 1, e.g., 0.5 or 0.33, when instruction execution time is 1 cycle and the machine can execute multiple instructions within a single cycle. An instruction that executes without occupying a functional unit may have zero throughput value.

Latency definition -

The number of clock cycles that are required for the CPU to complete the execution of all of the μ ops that form an instruction.

The file contains partial throughput / latency data:

- a) Not all 3rd Generation Intel® Xeon® Scalable Processor Family instructions are appearing in the database.
- b) Not all instructions that do appear have both their throughput and their latency specified.

If you require data that is missing, we suggest using the values available in the Intel® Xeon® Scalable Processor throughput and latency document (available here: <https://software.intel.com/en-us/articles/intel-sdm#optimization>), that provides Skylake Server microarchitecture information as a reasonable approximation.

Note that actual throughput measured on Intel processors may vary by up to 0.1 cycles. Also, latency may sometimes vary due to dynamic micro-architectural conditions, and in this case we report the average rounded to the nearest integer value. In all cases we assume data read from DCU or written to store-buffers.