TODAY - INTRODUCING

2ND GENERATION INTEL® XEON® SCALABLE PROCESSORS

intel
OPTANE DC
PERSISTENT MEMORY

intel

HOW INTEL SPEEDS HPC PERFORMANCE

$$\frac{\text{Work}}{\text{Time}} = \text{System Peak} \times \text{Efficiency} \times \text{Software Optimized}$$

2ND GENERATION INTEL® XEON® SCALABLE PROCESSORS SPEED HPC PERFORMANCE IN WAYS THAT MATTER

System Peak

- *AVX 512*
- *Memory Bandwidth*
- *Intel® DL Boost*
- *Intel® Optane™ DC persistent memory*

New technologies to exceed performance boundaries

Efficiency

- *Architecture Performance*
- *Multi-node Scaling*

Strong efficiency through strong hardware

Software Optimized

- *Software Tools*
- *Performance Libraries*
- *Development Tools*

Realize value with minimum work

2ND GEN INTEL® XEON® SCALABLE PROCESSORS ARCHITECTED FOR HPC & AI

Architecture Performance

Boost Applications

Amdahl's Law requires hardware improvement for real workload performance increases

Up to 1.7x better Floating Point perf/core[1]

Intel® AVX-512

Operation Efficiency

Special instructions optimized for HPC workloads and integrated into Intel® Xeon® Scalable Processors

2x better than AVX2[2]

Intel® Xeon® Platinum 9200 Processors

Compute Density

New class of Intel® Xeon® CPUs to boost performance for memory intensive HPC workloads

2x memory channels per CPU

Intel® DL Boost AI Convergence

Converge HPC & AI

Solutions to execute AI on HPC infrastructure combined with radical AI performance increases on Intel® Xeon®

14x over 1st Gen Intel® Xeon® SP[3]

Intel® Optane™ DC persistent memory for HPC

Lower Cost More Memory

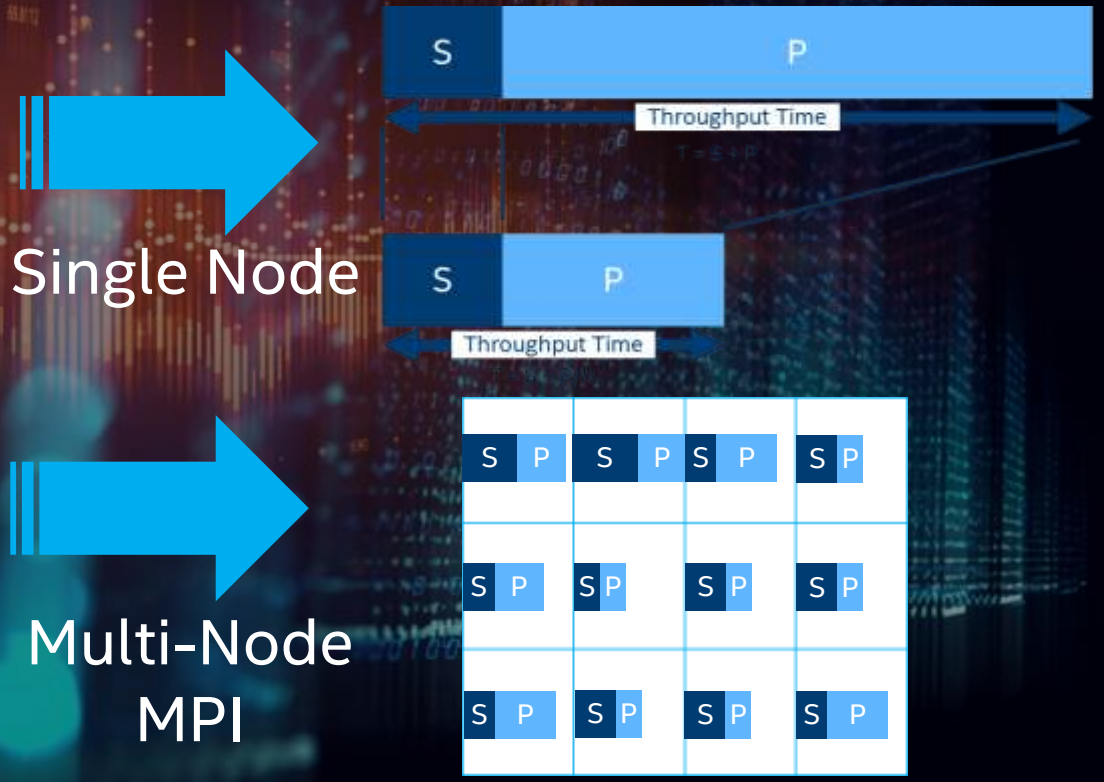
Use Intel® Optane™ DC persistent memory over DDR to solve larger problems at lower cost; Discover new use cases with Intel® Optane™ technology

[1]1-copy SPECrate2017_fp_base* 2 socket Intel 8280 vs 2 socket AMD EPYC 7601; config details slide 12
[2]Theoretical performance when comparing AVX512 to AVX2. Details on slide 11
[3]Platinum 8210 vs Plat 8180 at launch; config details on slide 12

intel

STRONG COMPUTE ARCHITECTURE MATTERS

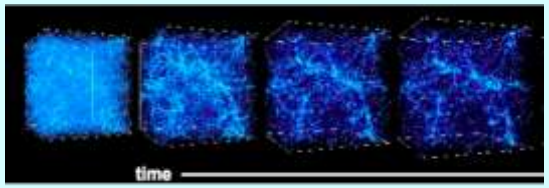
- **Amdahl's Law**
- Programs can only be accelerated as much as the serial component of that program



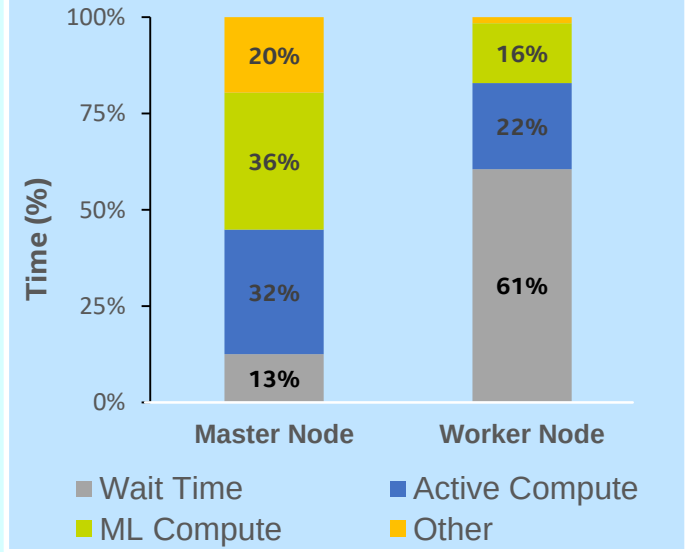
Real Performance Impact

CosmoFlow

Application predicting cosmological parameters by analyzing dark matter distributions in the universe



Single Node Breakdown[1]



Speeding up each node reduces the amount of "waiting time"

[1]Profile of time spent in various stages of the CosmoFlow application on a single Intel® Xeon Phi™ Processor 7250 (KNL) node. The stages are OpenMP spin time and overhead, non-convolutional computational time, 3D convolutions, CPE ML Plugin, other time, Linux kernel time, and TensorFlow framework time.. Configurations: See slide 11

INTEL® AVX-512: ACCELERATING HPC & AI APPLICATIONS

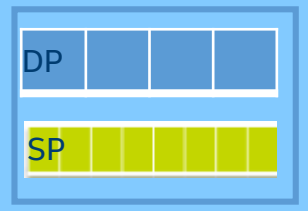
CPU Performance = # of CPU Cores × Core Frequency × Operation Efficiency

AVX-512

FP64: **8 FLOPs/cycle**
FP32: **16 FLOPs/cycle**

FP64: **32 FLOPs/cycle**
FP32: **64 FLOPs/cycle**
Int8: **256 OPs/cycle**

AVX-2 Vector,1 FMA



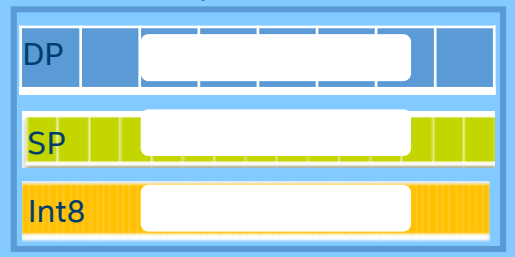
256 Wide Vector

Double wide vector
Double the FMAs

AVX-512 Vector, 2 FMA

512 (Double) Wide Vector

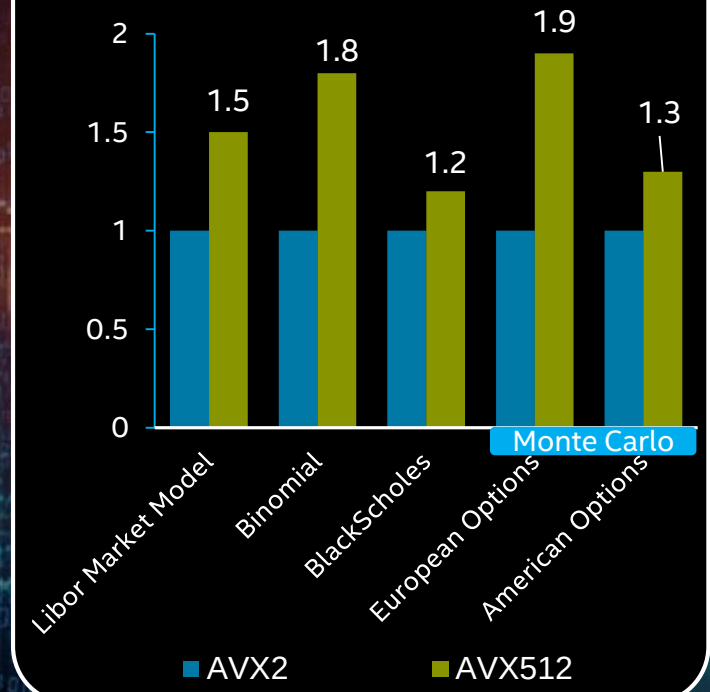
AVX-512 Optimized for HPC & AI



More operations per cycle

AVX2 vs AVX512 Real Performance Gain

8260L running FSI Apps
(higher is better)

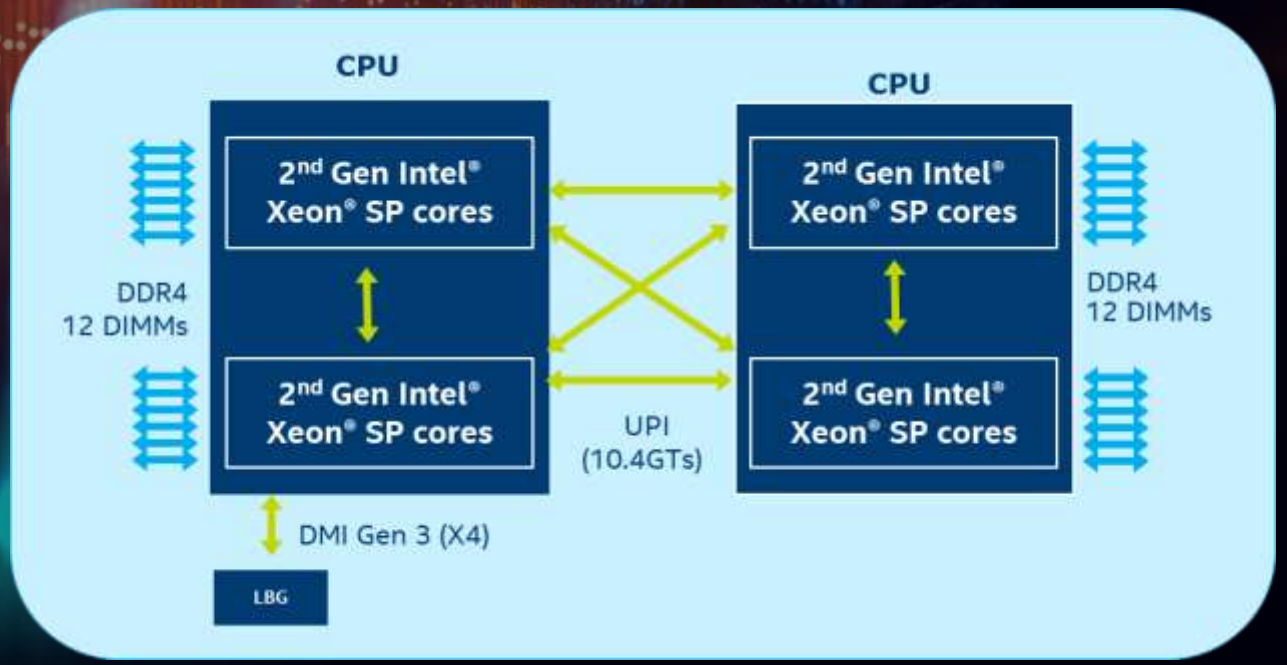


Performance results are based on testing as of 02/28/2019 and may not reflect all publicly available security updates. See configuration disclosure on slide 11. For more complete information about performance and benchmark results, visit www.intel.com/benchmarks.

A NEW CLASS OF ADVANCED PERFORMANCE

INTEL® XEON® PLATINUM 9200 PROCESSORS FOR YOUR MOST DEMANDING WORKLOADS

Extending Intel Xeon Processor Leadership for High Performant Workloads



PERFORMANCE LEADERSHIP

9200 SERIES WITH UP TO 56 CORES

UNPRECEDENTED MEMORY BANDWIDTH

12 DDR4 MEMORY CHANNELS PER CPU AT 2933 MT/S

OPTIMIZED MULTI CHIP PACKAGE

HIGH SPEED INTERCONNECT FOR INCREASED COMPUTE DENSITY
80 PCIE GEN3 LANES PER NODE

HPC↔AI CONVERGENCE

INCREASING AI PERFORMANCE ON INTEL® XEON® PROCESSORS

Intel® Optimizations for Caffe ResNet-50 Inference Throughput Performance

2S Intel® Xeon® Platinum 8180 processor (28 cores/S)

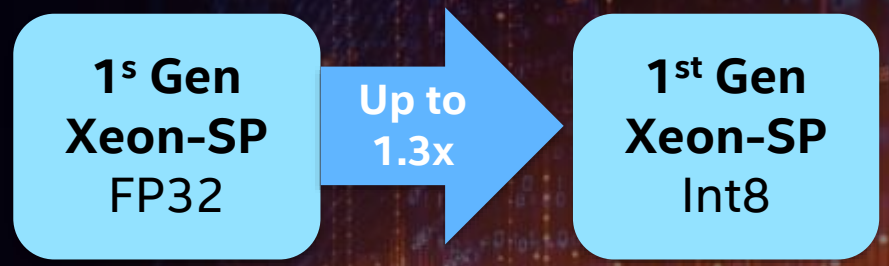
2S Intel® Xeon® Platinum 8280 processor (28 cores/S)

2S Intel® Xeon® Platinum 9282 processor (56 cores/S)

1st Generation Intel® Xeon® Scalable Processor

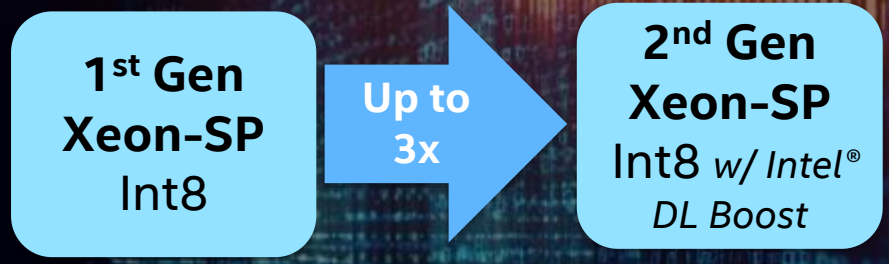
2nd Generation Intel® Xeon® Scalable Processor

Intel® DL Boost Theoretical Throughput per core over 1st Generation Intel® Xeon® Scalable Processors



3 Instructions
VPMADDUBSW
VPMADDWD
VPADDD

Faster throughput, but inefficient
Uses 3 instructions per operation



1 Instruction
VPDPBUSD

DL Boost fixes this, combines 3 instructions into 1

[1] Based on Intel internal testing: 1X,5.7x,14x and 30x performance improvement based on Intel® Optimization for Café ResNet-50 inference throughput performance on Intel® Xeon® Scalable Processor. See Configuration Details slide 13 Performance results are based on testing as of 7/11/2017(1x) ,11/8/2018 (5.7x), 2/20/2019 (14x) and 2/26/2019 (30x) and may not reflect all publically available security updates. No product can be absolutely secure. See configuration slide 13 Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. 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. For more complete information visit: http://www.intel.com/performance

RECENTLY ANNOUNCED:
INTEL TO BUILD FIRST EXASCALE SUPERCOMPUTER FOR U.S. DoE

Intel – partnered with Argonne National Laboratory – driving the convergence of HPC and AI

Aurora will accelerate the convergence of traditional HPC, data analytics, and AI

Intel's data centric portfolio at the heart of Aurora, integrated with Cray's "Shasta" system

Leading research & academia programs already engaged to harness Aurora and enable software ecosystem

"Achieving Exascale is imperative not only to better the scientific community, but also to better the lives of everyday Americans. Aurora and the next-generation of Exascale supercomputers will apply HPC and AI technologies to areas such as cancer research, climate modeling, and veterans' health treatments. The innovative advancements that will be made with Exascale will have an incredibly significant impact on our society."

- Rick Perry, US Secretary of Energy

intel


intel

NOTICES & DISCLAIMERS

- 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 product or component can be absolutely secure.
- Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. For more complete information about performance and benchmark results, visit http://www.intel.com/benchmarks .
- 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. For more complete information visit http://www.intel.com/benchmarks .
- 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. Performance varies depending on hardware, software, and system configuration and you can learn more at http://www.intel.com/go/turbo.
- Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.
- Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.
- Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.
- *Other names and brands may be claimed as property of others.
- Intel, the Intel logo, Xeon, and Optane are trademarks of Intel Corporation in the United States and other countries.
-