High Performance Computing with Intel

Nikolay Mester,

intel

HPC/Cloud AE, Intel



Intel[®] HPC Portfolio





Intel[®] Xeon[®] Scalable Processor The ONLY x86 Datacenter CPU with Built-in AI Acceleration

Intel® Advanced Vector Extensions 512 Intel® Deep Learning Boost Intel® Optane[™] persistent memory

Cascade Lake

14nm New AI acceleration built-in (Intel® DL Boost with VNNI) New memory storage hierarchy

Ice Lake

10nm New microarchitecture Increased memory bandwidth

Sapphire Rapids

10nm Enhanced SuperFin Next gen Intel® DL Boost (Intel® Advanced Matrix Extensions)

Accelerating Innovation from Edge to Cloud to Supercomputing

З

INTRODUCING

3rd Gen Intel[®] Xeon[®] Scalable processors Built specifically for our customers' needs

Optimized for Cloud, Enterprise, HPC, 5G and Edge

intel

PLATINUM

Xeon®

Built-in security with Intel Software Guard Extensions, Platform Firmware Resilience and Total Memory Encryption

Only data center processor with built-in AI (Intel DL Boost)

Built-in crypto acceleration reduces the performance impact of pervasive encryption

INTRODUCING

3rd Gen Intel[®] Xeon[®] Scalable processors Performance made flexible



Performance varies by use, configuration and other factors. Configurations see appendix [1,3,5,55]

INTRODUCING

3rd Gen Intel[®] Xeon[®] Scalable processors Performance made flexible



Flexible Performance for Most Demanding Workloads Outstanding gen-on-gen performance from intelligent edge to cloud



Performance varies by use, configuration and other factors. Configurations see appendix [5,7, 17, 19, 52]

Competitive Leadership

3rd Gen Intel[®] Xeon[®] Scalable processor vs. AMD EPYC Milan Superior performance at equal cores (32)

	Monte Carlo	RELI	ON	LAMMPS	
3rd gen Xeon vs EPYC Milan	Up to 105%	Up to 6	8%	Up to 5	7%
Up to 23% Better performance across 12 leading HPC applications and benchmarks	better performance	better performance better performance Binomia		formance better performance Binomial Options	
	IN/				
	Up to C	52%	Up to 3	37%	
	better pe	rformance	better per	formance	

Mainstream SKUs: Intel Xeon 8358 vs EPYC Milan 7543

Welcoming Ice Lake customers and partners 3rd Gen Intel ® Xeon® Scalable processor momentum



Intel® Optane[™] Persistent Memory 200 Series Persistent memory made flexible



eADR (Enhanced Asynchronous DRAM Refresh) improves performance of apps that use persistent memory by eliminating "cache flushes" – volatile data including the CPU caches save automatically, even if power fails

Intel Optane PMem 200 series is compatible with existing PMem SW ecosystem & it continues to grow

KATANA GRAPH

Computes up to **2X** faster graph analytics algorithms used in search, social networks, and fraud detection

vmware Lower infrastructure costs by up **25%** per VM while delivering the same performance

Performance varies by use, configuration and other factors. Configurations see appendix [53]

Intel® Optane[™] Persistent Memory 200 Series (Barlow Pass)



System Level Optimization



eADR automatically flushes caches eliminating CLFlush runtime use/overhead (Requires: Barlow Pass, Energy Store, App support)

Intel[®] Optane[™] Persistent Memory 200 Series

- Consistent developer target: capacity, latency...
- Increased bandwidth, better power efficiency
- Cross system innovation for increased application performance



Intel® Optane[™] SSD P5800X The world's fastest data center SSD



Performance varies by use, configuration and other factors. Configurations see appendix [46]

Intel® SSD D5-P5316 Massive storage capacity made flexible



Industry-leading density + no compromise quality and reliability enables customers to confidently deploy at scale in warm storage

Increased per server bandwidth accelerates access to large datasets in CDN, HCI, Big Data, AI, HPC, and Cloud Elastic Storage

Reduce storage footprint by up to 20x

Performance varies by use, configuration and other factors. Configurations see appendix [47]

Intel[®] Ethernet 800 Series Network Adapters Connectivity made flexible



Prioritizes application traffic to help deliver the performance required for high-priority, network intensive workloads

Fully programmable pipeline to enable frame classification for advanced and proprietary protocols

Supports **RDMA over iWARP or RoCEv2** protocols and NVMe over TCP with ADQ for high throughput, low latency storage, cloud and HPC clusters

New E810-2CQDA2 adapter targets high-performance workloads such as vRAN, NFV forwarding plane, storage, HPC, cloud and content delivery networks

Source: 2x more resources from 100Gbps to 200Gpbs

Storage



HPC

GRID

Software

THE PROBLEM WITH POSIX AND BLOCKS



Revolutionizing High Performance Storage



- Intel[®] Optane DC Persistent Memory is revolutionizing high performance storage by providing low-latency and fine-grained persistent storage.
 - ... but existing distributed storage software is a bottle neck
 - Optimized for millisecond rotating media
 - POSIX constraints limit performance Scale out storage needs to be built from the ground up for new NVM technology





18

pache

Datasets: DAOS container

- Manageable and coherent entities
 - Stored in a pool
 - Simplified data management
 - Cross-tier migration
 - Query capability to identify recently accessed containers
 - Container indexing
 - Snapshot and rollback support
 - Built-in producer/consumer workflow pipeline support
 - NFSv4-type ACLs







DAOS Container



Future products



HPC

• Al

- Interconnect
- Software

Next-Generation Intel Xeon Scalable Processors Unique Capabilities Optimized for HPC and Al Acceleration



Breakthrough Technology

DDR5 Increased Memory BW

High Throughput

CXL 1.1 Next-gen IO

Built-In AI Acceleration

Intel® Advanced Matrix Extensions (AMX)

Increased Deep Learning Inference and Training Performance

Agility and Scalability

Hardware Enhanced Security Intel[®] Speed Select Technology Broad Software Optimization

NEW

High Bandwidth Memory

Significant performance increase for bandwidth-bound workloads

Xe Architecture: Brought to Life



Freedom for Accelerated Compute Break Free from the Constraints of Proprietary Programming Models

Freedom of Choice in Hardware

Multi-Vendor Adoption Momentum

- Codeplay brings SYCL support for NVIDIA GPUs
- Fujitsu Fugaku uses oneAPI oneDNN on Arm
- Huawei Al Chipset supported by Data Parallel C++
- NERSC, ALCF, Codeplay partner on SYCL for next gen Supercomputer

Realize All the Hardware Value on XPUs

Optimized Libraries, Compilers, AnalysisTools & Intel® DevCloud



Confidently Develop Performant Code

Compatible with Existing Languages & Standards Intel® oneAPI Toolkits for HPC, AI, Rendering...





"Intel is providing the blended environment we need, with next-generation Intel Xeon Scalable processors (Sapphire Rapids), with built-in acceleration for new HPC and AI workloads plus Ponte Vecchio, Intel's upcoming GPU.

Since our users need to access data as quickly as possible, we'll be using Intel DAOS for fast, high bandwidth, low latency, and high IOPS storage, on a system with 3rd Gen Xeon processors and Optane persistent memory."

Prof. Dieter Kranzlmüller

Chairman of the Board of Directors of the Leibniz Supercomputing Centre (LRZ)



Accelerating HPC for the Future



Intel delivers heterogeneous architectures for today's & tomorrow's challenges



Intel's HPC portfolio has the flexibility to grow with changing customer demands



Notices and Disclaimers

Performance varies by use, configuration and other factors. Learn more at <u>www.Intel.com/PerformanceIndex</u>.

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

Intel contributes to the development of benchmarks by participating in, sponsoring, and/or contributing technical support to various benchmarking groups, including the BenchmarkXPRT Development Community administered by Principled Technologies.

Your costs and results may vary.

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

Some results may have been estimated or simulated.

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

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

Statements in this document that refer to future plans or expectations are forward-looking statements. These statements are based on current expectations and involve many risks and uncertainties that could cause actual results to differ materially from those expressed or implied in such statements. For more information on the factors that could cause actual results to differ materially, see our most recent earnings release and SEC filings at www.intc.com.

© 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.

Performance made flexible.