HPC Solution Brief

5th Gen Intel® Xeon® Processors
High-Performance Computing

High Performance Computing with Built-In Accelerators: 5th Gen Intel® Xeon® Processors

Speed time to insights with accelerated high-performance computing (HPC) using 5th Gen Intel Xeon processors, featuring Intel® Accelerator Engines.

To help your business maximize the value of its infrastructure, 5th Gen Intel Xeon processors provide up to 1.22x higher average HPC performance versus the prior generation.[1] They also help lower costs and your carbon footprint with up to 1.08x higher average performance per watt vs. the prior generation.[1] Improve uptime and future-proof your IT with silicon-based security features and quality solutions that scale. The processors can also help accelerate the time to value by providing a common CPU platform across HPC-AI workloads.

TRADITIONAL HPC MARKET	
6.1% CAGR[2]	$51.2 Billion by 2027[2]

EXPANDED HPC + AI MARKET	
6.3% CAGR[2]	$60.4 Billion by 2027[2]

Whereas conventional HPC systems have been designed for monolithic applications, that approach is shifting toward more flexibility to reflect the new diversity of requirements. Likewise, this evolution intensifies the value that practitioners place on open-standards hardware and software, so that various types of solutions and workloads coexist and deliver optimal results on shared systems. Deployment of AI workloads on hardware not optimized for them may not meet performance-per-watt goals. Open, cross-architecture programming models have assumed new prominence, avoiding the need to recode for multiple architectures and heterogeneous accelerators. They also help avoid proprietary software lock-in and improve the longevity and future-readiness of scientific applications.

5th Gen Intel Xeon processors continue Intel's cadence of ecosystem enablement across hardware and software to deliver faster time to value on both traditional HPC and converged HPC-AI workloads. They are well suited to small or mid-sized specialized models with advanced transfer learning or fine-tuning requirements. Hardware-based acceleration for deep learning and general-purpose workloads handles real-time throughput with low latency for blended HPC applications.

The processor offers an increase up to 64 high-performance cores (128 threads) per socket, supported by up to 3x increased Last Level Cache with 5th Gen Intel Xeon processors vs. 4th Gen Intel Xeon processors.[3] These changes increase core utilization on demanding, highly parallel workloads such as those in electronic design automation (EDA) and computational fluid dynamics (CFD) domains. The processor provides DDR5 memory speeds up to 5600 MT/s,[4] and robust I/O includes up to 80 lanes of PCIe 5.0, to optimize latency and keep the cores supplied with data.

HIGHER PERFORMANCE

UP TO **1.22x** higher average HPC performance on 5th Gen Intel Xeon platform vs. prior gen[1]

UP TO **2.14x** higher average HPC performance across a broad set of HPC codes vs. prior generation Intel Xeon Platinum 8380 processor[1]

Balanced Platform Advances for HPC

- **High-throughput, high-efficiency execution resources.** Increase from 60 to 64 cores per socket and get up to 1.22x faster average HPC performance compared to the previous generation.[1]
- **Enhanced memory subsystem.** Up to 1.16x memory bandwidth improvement on 5th Gen Intel Xeon processors vs. 4th Gen Intel Xeon Scalable processors[4,5] and up to 3x increased Last Level Cache with 5th Gen Intel Xeon processors vs. 4th Gen Intel Xeon processors.[3]
- **Fast, high-capacity I/O.** Up to 80 lanes of PCIe per socket, with Intel® UltraPath Interconnect (Intel UPI) 2.0 speeds up to 20 GT/s and support for Compute Express Link (CXL) Types 1, 2 and 3.

Maximize the longevity and return on your IT investments by upgrading from the previous generation to new 5th Gen Intel Xeon processors, which are software- and pin-compatible with 4th Gen. The benefits are even more pronounced when upgrading from earlier generations of Intel Xeon processors. HPC-AI implementations also benefit from ongoing platform innovations that improve data movement and processing more broadly.

Intel Accelerator Engines

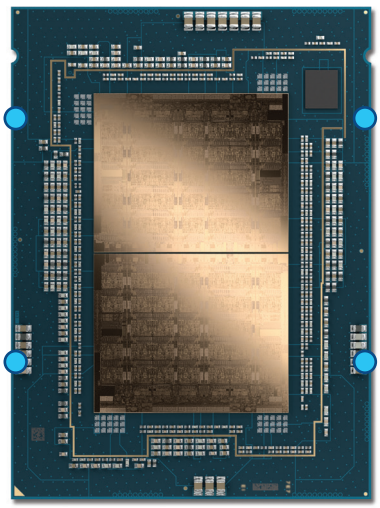
5th Gen Intel Xeon processors feature Intel Accelerator Engines, the industry's largest collection of built-in accelerators to increase throughput on critical tasks such as AI, HPC, analytics, networking and storage. Because they are built into the processor, they do not incur the latency of going off-chip to a discrete accelerator on the PCIe bus, with a corresponding savings in energy consumption as well, compared to discrete solutions or software-based ones running on the cores. Thus, Intel Accelerator Engines can help organizations achieve better performance as well as CapEx and OpEx savings:

- **Performance.** Specialized, purpose-built accelerators target delivering significant gains in throughput for their targeted workloads. In particular, Intel Advanced Matrix Extensions (Intel AMX) accelerates AI workloads on the CPU, improving throughput without additional dedicated hardware.
- **Operating and system costs.** Built in accelerators may reduce the need for additional system investment, and a reduced system footprint may provide significant energy savings.

HPC-AI solutions on 5th Gen Intel Xeon processors benefit from instruction set architecture (ISA) advances designed to accelerate common AI and machine learning tasks. For example, Intel Vector Neural Network Instructions (VNNI) accelerate inference by combining three instructions into one to complete multiply and accumulate operations in

Intel® Advanced Matrix Extensions (Intel AMX)
accelerate deep learning inference and training

Intel Advanced Vector Extensions 512 (Intel AVX-512)
optimize data processed per clock cycle



Intel QuickAssist Technology (Intel QAT)
accelerates encryption and compression operations

Intel Data Streaming Accelerator (Intel DSA)
optimizes streaming data movement and transformation

INT8. Support for new datatypes includes BF16, a 16-bit floating-point format that accelerates inference while maintaining model accuracy. It reduces the time to train a model (or the time to infer from it) by using lower precision, reducing compute requirements.

Ready-to-deploy tools for HPC-AI

The increased use of AI in HPC workflows across industries and scientific disciplines promises significant improvements in efficiency. Convergence between AI and HPC may occur at the application or the workflow level. In HPC applications, a traditional model can be replaced with an AI model for a task such as creating results on fine mesh from coarse mesh. The AI model can potentially complete this work in less time, possibly with higher accuracy. In HPC workflows, AI can be used for postprocessing to evaluate results and generate insights to users, or it can be used for preprocessing to improve input datasets used in HPC workloads.

Intel software developer tools, powered by oneAPI, include compilers, libraries, frameworks and performance tools to build, analyze and tune high-quality, cross-platform software optimized for Intel architecture. In particular, they streamline the adoption path for Intel Accelerator engines in their solutions, to pursue the highest performance and efficiency possible, across CPUs, GPUs and other hardware such as FPGAs. They include the Intel oneAPI Base Toolkit and Intel HPC Toolkit to build, analyze and scale applications across shared- and distributed-memory computing systems, as well as Intel AI tools to accelerate end-to-end data science and machine learning pipelines.

In addition, Intel is deeply committed to ecosystem enablement through open source involvement and upstreaming of new value-added optimizations, as well as co-engineering partnerships across the solutions ecosystem. Together, these developer enablement measures help reduce the time and effort required to align with Intel's technology roadmap and cost-effectively make solutions more performant, efficient and future-ready.

Enhanced performance and TCO advantages

5th Gen Intel Xeon processors increase throughput across a range of HPC benchmarks and workloads. These advantages are conferred primarily by the combination of higher core counts, larger last-level cache and increased memory bandwidth. In addition to increased value from system capital investment, the platform helps reduce energy spend for better TCO.

To handle the full spectrum of HPC-AI use cases, solution architects can adopt additional complementary Intel hardware technologies without porting or refactoring code, retaining software compatibility across the environment. To improve outcomes on the most demanding memory bandwidth-sensitive workloads, the Intel Xeon CPU Max Series integrates up to 64 GB of HBM2e high-bandwidth memory onto the processor package, reducing the need to traverse the memory bus for data. For solutions that benefit from GPU resources, the Intel Data Center GPU Max Series contains up to 128 Xe-HPC Cores, Intel's foundational GPU compute building block.

Diverse Intel hardware and software technologies give HPC-AI practitioners the power to innovate into the future, solving the world's largest compute problems more quickly and cost-effectively.

LIFE AND MATERIAL SCIENCE

UP TO **1.25x**

higher average Life & Material Science HPC performance on 5th Gen Intel Xeon platform vs. prior gen[1]

INDUSTRY BENCHMARK

UP TO **1.19x**

higher average industry standard performance on 5th Gen Intel Xeon platform vs. prior gen[1]

MANUFACTURING

UP TO **1.3x**

higher average manufacturing HPC performance on 5th Gen Intel Xeon platform vs. prior gen[1]

FINANCIAL SERVICES

UP TO **1.16x**

higher average FSI performance on 5th Gen Intel Xeon platform vs. prior gen[1]

LAMMPS

UP TO **1.31x**

higher LAMMPS performance on 5th Gen Intel Xeon platform vs. prior gen[6]

ANSYS LS-DYNA

UP TO **1.42x**

higher Ansys LS-DYNA performance on 5th Gen Intel Xeon platform vs. prior gen[7]

ANSYS FLUENT

UP TO **1.17x**

higher Ansys Fluent performance on 5th Gen Intel Xeon platform vs. prior gen[8]

Learn More

www.intel.com/xeon

www.intel.com/hpc

Resources

Intel Optimization Hub

Intel Developer Cloud

[1] See [H1] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.

[2] Intersect360, May 16, 2023. "Worldwide HPC and AI 2022 Total Market Size and 2023–2027 Forecast: Products and Services." https://www.intersect360.com/report/worldwide-hpc-and-ai-2022-total-market-size-and-2023-2027-forecast-products-and-services/.

[3] See [G11] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.

[4] Available on some SKUs.

[5] See [G12] at intel.com/processorclaims: 5th Gen Intel Xeon Scalable processors. Results may vary.

[6] See [H14] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.

[7] See [H7] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.

[8] See [H6] at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.

Performance varies by use, configuration and other factors. Learn more on the Performance Index site.

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.

Your costs and results may vary.

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

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

1223/JAW/MESH/353955-001US