December 14, 2023

Intel Accelerates AI Everywhere with Launch of Powerful Next-Gen Products

Intel Core Ultra and 5th Gen Intel Xeon processors expand Intel's unmatched AI portfolio, bringing AI to all

NEW YORK--(BUSINESS WIRE)-- At its "AI Everywhere" launch in New York City today, Intel introduced an unmatched portfolio of AI products to enable customers' AI solutions everywhere — across the data center, cloud, network, edge and PC.

This press release features multimedia. View the full release here: https://www.businesswire.com/news/home/20231214375413/en/

Michelle Johnston Holthaus, executive vice president and general manager of the Client Computing Group at Intel, holds the Intel Core Ultra mobile processor launched Dec. 14, 2023. (Credit: Intel Corporation)

Highlights include:

- The Intel® Core™ Ultra mobile processor family, the first built on the Intel 4 process technology and the first to benefit from the company's largest architectural shift in 40 years, delivers Intel's most power-efficient client processor and ushers in the age of the AI PC.
- The 5th Gen Intel® Xeon® processor family is built with AI acceleration in every core, bringing leaps in AI and overall performance and lowering total cost of ownership (TCO).
- Intel CEO Pat Gelsinger showed for the first time an Intel® Gaudi®3 AI accelerator, arriving on schedule next year.

"AI innovation is poised to raise the digital economy's impact up to as much as one-third of global gross domestic product[1]," Gelsinger said. "Intel is developing the technologies and solutions that empower customers to seamlessly integrate and effectively run AI in all their applications — in the cloud and, increasingly, locally at the PC and edge, where data is generated and used."

Gelsinger showcased Intel's expansive AI footprint, spanning cloud and enterprise servers to networks, volume clients and ubiquitous edge environments. He also reinforced that Intel is on track to deliver five new process technology nodes in four years.

More: 'AI Everywhere' 2023 (Press Kit) | Intel Core Ultra Ushers in the Age of the AI PC | Intel Core Ultra Processors (Press Kit) | New 5th Gen Intel Xeon Processors are Built with AI Acceleration in Every Core | 5th Gen Intel Xeon Processors (Press Kit) | Intel's 'AI Everywhere' Event (Event Video)

"Intel is on a mission to bring AI everywhere through exceptionally engineered platforms, secure solutions and support for open ecosystems. Our AI portfolio gets even stronger with

today's launch of Intel Core Ultra ushering in the age of the AI PC and AI-accelerated 5th Gen Xeon for the enterprise," Gelsinger said.

Intel Core Ultra Powers AI PC and New Applications

Intel Core Ultra represents the company's largest architectural shift in 40 years and launches the AI PC generation with innovation on all fronts: CPU compute, graphics, power, battery life and profound new AI features. The AI PC represents the largest transformation of the PC experience in 20 years, since Intel® Centrino® untethered laptops to connect to Wi-Fi from anywhere.

Intel Core Ultra features Intel's first client on-chip AI accelerator — the neural processing unit, or NPU — to enable a new level of power-efficient AI acceleration with 2.5x better power efficiency than the previous generation[2]. Its world-class GPU and leadership CPU are each also capable of speeding up AI solutions.

As important, Intel is partnering with more than 100 software vendors to bring several hundred AI-boosted applications to the PC market — a wide array of highly creative, productive and fun applications that will change the PC experience. For consumer and commercial customers, this means a larger and more extensive set of AI-enhanced applications will run great on Intel Core Ultra, particularly compared to competing platforms. For example, content creators working in Adobe Premiere Pro will enjoy 40% better performance versus the competition[3].

Intel Core Ultra-based AI PCs are available now from select U.S. retailers for the holiday season. Over the next year, Intel Core Ultra will bring AI to more than 230 designs from laptop and PC makers worldwide. AI PCs will comprise 80% of the PC market by 2028[4] and will bring new tools to the way we work, learn and create.

New Xeon Brings More Powerful AI to the Data Center, Cloud, Network and Edge

The 5th Gen Intel Xeon processor family, also introduced today, brings a significant leap in performance and efficiency[5]: Compared with the previous generation of Xeon, these processors deliver 21% average performance gain[6] for general compute performance and enable 36% higher average performance per watt across a range of customer workloads[7]. Customers following a typical five-year refresh cycle and upgrading from even older generations can reduce their TCO by up to 77%[8].

Xeon is the only mainstream data center processor with built-in AI acceleration, with the new 5th Gen Xeon delivering up to 42% higher inference and fine-tuning on models as large as 20 billion parameters[9]. It's also the only CPU with a consistent and ever-improving set of MLPerf training and inference benchmark results.

Xeon's built-in AI accelerators, together with optimized software and enhanced telemetry capabilities, enable more manageable and efficient deployments of demanding network and edge workloads for communication service providers, content delivery networks and broad vertical markets, including retail, healthcare and manufacturing.

During today's event, IBM announced that 5th Gen Intel Xeon processors achieved up to

2.7x better query throughput on its watsonx.data platform compared to previous-generation Xeon processors during testing [10]. Google Cloud, which will deploy 5th Gen Xeon next year, noted that Palo Alto Networks experienced a 2x performance boost in its threat detection deep learning models by using built-in acceleration in 4th Gen Xeon through Google Cloud. And indie game studio Gallium Studios turned to Numenta's AI platform running on Xeon processors to improve inference performance by 6.5x over a GPU-based cloud instance, saving cost and latency in its AI-based game, Proxi[11].

This kind of performance unlocks new possibilities for advanced AI – not only in the data center and cloud, but across the world's networks and edge applications.

AI Acceleration and Solutions Everywhere Developers Need It

Both Intel Core Ultra and 5th Gen Xeon will find their way into places you might not expect. Imagine a restaurant that guides your menu choices based on your budget and dietary needs; a manufacturing floor that catches quality and safety issues at the source; an ultrasound that sees what human eyes might miss; a power grid that manages electricity with careful precision.

These edge computing use cases represent the fastest-growing segment of computing — projected to surge to a $445 billion global market by the end of the decade — within which AI is the fastest-growing workload. In that market, edge and client devices are driving 1.4x more demand for inference than the data center[12].

In many cases, customers will employ a mix of AI solutions. Take Zoom, which runs AI workloads on Intel Core-based client systems and Intel Xeon based-cloud solutions within its all-in-one communications and collaboration platform to deliver best user experience and costs. Zoom uses AI to suppress the neighbor's barking dog and blur your cluttered home office, and to generate a meeting summary and email.

To make AI hardware technologies as accessible and easy-to-use as possible, Intel builds optimizations into the AI frameworks developers use (like PyTorch and TensorFlow) and offers foundational libraries (through oneAPI) to make software portable and highly performant across different types of hardware.

Advanced developer tools, including Intel's oneAPI and OpenVINO toolkit, help developers harness hardware acceleration for AI workloads and solutions and quickly build, optimize and deploy AI models across a wide variety of inference targets.

Sneak Peek: Intel Gaudi3 AI Accelerator

Wrapping up the event, Gelsinger provided an update on Intel Gaudi3, coming next year. He showed for the first time the next-generation AI accelerator for deep learning and large-scale generative AI models. Intel has seen a rapid expansion of its Gaudi pipeline due to growing and proven performance advantages combined with highly competitive TCO and pricing. With increasing demand for generative AI solutions, Intel expects to capture a larger portion of the accelerator market in 2024 with its suite of AI accelerators led by Gaudi.

With partners and a broad ecosystem, Intel is unlocking new growth opportunities fueled by AI, bringing AI everywhere. Today.

About Intel

Intel (Nasdaq: INTC) is an industry leader, creating world-changing technology that enables global progress and enriches lives. Inspired by Moore's Law, we continuously work to advance the design and manufacturing of semiconductors to help address our customers' greatest challenges. By embedding intelligence in the cloud, network, edge and every kind of computing device, we unleash the potential of data to transform business and society for the better. To learn more about Intel's innovations, go to newsroom.intel.com and intel.com.

[1] Projection based on Oxford Economics, BCG, and McKinsey data.
[2] As measured by Perf/Watt on UL Procyon AI benchmark while running an int8 model on Intel Core Ultra 7 165H NPU vs. Intel Core i7-1370P GPU. See www.intel.com/PerformanceIndex for workloads and configurations. Results may vary.
[3] As measured by PugetBenchPremier Pro Standard (v24.1) benchmark. Overall score on Intel Core Ultra 7 155H vs. AMD Ryzen 7 7840U. See www.intel.com/PerformanceIndex for workloads and configurations. Results may vary.
[4] Source: Boston Consulting Group
[5] Average performance gain as measured by the geomean of SPEC CPU rate, STREAM Triad, and LINPACK compared to 4th Gen Intel Xeon processor. See G1 at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[6] Average performance gain as measured by the geomean of SPEC CPU rate, STREAM Triad, and LINPACK compared to 4th Gen Intel Xeon processor. See G1 at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[7] See [T13] at intel.com/processorclaims: 5th Gen Intel Xeon Scalable processors. Results may vary.
[8] See [T7] at intel.com/processorclaims: 5th Gen Intel Xeon Scalable processors. Results may vary.
[9] Based on Intel's internal modeling as of December 2023. See A1, A2, A16 at intel.com/processorclaims: 5th Gen Intel Xeon processors. Results may vary.
[10] See [D7] at intel.com/processorclaims: 5th Gen Intel Xeon Scalable processors. Results may vary.
[11] NVIDIA Tesla M60 GPU: Tested by Gallium as of 11/17/2023. 1-node, 1x GPU on AWS g3s.xlarge, 4vCPU, 30.5 GB memory, 8 GB GPU memory, Ubuntu 22.04 Kernel 5.17, Gallium custom Word2vec, Sequence Length 64, Batch Size 64. 4th Gen Intel® Xeon® Scalable: Tested by Gallium as of 11/28/2023. 1-node, 2x Intel® Xeon® Platinum 8488C on AWS m7i.xlarge, 16GB memory, Ubuntu 22.04 Kernel 5.17, Numenta Platform for Intelligent Computing V1.0, Numenta-Optimized BERT-Large, Sequence Length 64, BF16, Batch Size 64.
[12] AI and Machine Learning Survey, 22.2, Evans Data Corp, 2023.

Forward-Looking Statements

This release contains forward-looking statements with respect to Intel's current and anticipated future products and processes, AI and other market trends, and anticipated benefits therefrom. Such statements involve many risks and uncertainties that could cause our actual results to differ materially from those expressed or implied, including: changes in demand for our products; the high level of competition and rapid technological change in our industry; the rapidly changing and uncertain evolution of the AI market; vulnerability to new

product development and manufacturing-related risks, including product defects or errata, particularly as we develop next generation products and implement next generation process technologies; risks associated with a highly complex global supply chain, including from disruptions, delays, trade tensions, or shortages; sales-related risks, including customer concentration and the use of distributors and other third parties; potential security vulnerabilities in our products; cybersecurity and privacy risks; intellectual property risks and risks associated with litigation and regulatory proceedings; evolving regulatory and legal requirements across many jurisdictions; geopolitical and international trade conditions; risks of large scale global operations; macroeconomic conditions; and other risks and uncertainties described in our earnings release dated October 26, 2023, our most recent Annual Report on Form 10-K and our other filings with the U.S. Securities and Exchange Commission. All information in this press release reflects Intel management views as of the date hereof unless an earlier date is specified. Intel does not undertake, and expressly disclaims any duty, to update such statements, whether as a result of new information, new developments, or otherwise, except to the extent that disclosure may be required by law.

© 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.

View source version on businesswire.com:
https://www.businesswire.com/news/home/20231214375413/en/

Kristie Taylor
1-512-924-0460
kristie.taylor@intel.com

Source: Intel