September 19, 2023

Intel Innovation 2023: Empowering Developers to Bring AI Everywhere

Artificial intelligence gives rise to the 'Siliconomy,' a new era of global expansion driven by the magic of silicon and software. NEWS HIGHLIGHTS

- Intel confirmed its five-nodes-in-four-years process technology plan remains on track, and it demonstrated the world's first multi-chiplet package using Universal Chiplet Interconnect Express (UCIe) interconnects.
- The company revealed new details on next-generation Intel® Xeon® processors, including major advances in power efficiency and performance, and an E-core processor with 288 cores. 5th Gen Intel® Xeon® processors will launch Dec. 14
- The AI PC arrives with the launch of Intel® Core™ Ultra processors on Dec. 14. With Intel's first integrated neural processing unit, Core Ultra will deliver power-efficient AI acceleration and local inference on the PC.
- A large AI supercomputer will be built on Intel Xeon processors and Intel® Gaudi®2 AI hardware accelerators, with Stability AI as the anchor customer.
- General availability announced for the Intel® Developer Cloud for building and testing high-performance applications like AI, including details that it is already in use by customers.
- New and forthcoming Intel software solutions, including the 2023.1 release of the Intel® Distribution of OpenVINO™ toolkit, will help developers unlock new AI capabilities.

SAN JOSE, Calif.--(BUSINESS WIRE)-- At its third annual Intel Innovation event, Intel unveiled an array of technologies to bring artificial intelligence everywhere and make it more accessible across all workloads, from client and edge to network and cloud.

This press release features multimedia. View the full release here: https://www.businesswire.com/news/home/20230919339712/en/

A photo shows a tray of Intel Core Ultra processors being assembled at Intel's advanced packaging facilities in Penang, Malaysia. The Intel Core Ultra processor, code-named Meteor Lake, is the first client processor manufactured on the new Intel 4 process node using its 3D high-performance hybrid architecture, and the first client tile-based design enabled by Foveros packaging technology and features CPU, GPU and NPU. (Credit: Intel Corporation)

"AI represents a generational shift, giving rise to a new era of global expansion where computing is even more foundational to a better future for all," said Intel CEO Pat Gelsinger. "For developers, this creates massive societal and business opportunities to push the boundaries of what's possible, to create solutions to the world's biggest challenges and to improve the life of every person on the planet."

More: Intel Innovation 2023 (Press Kit) | Innovation 2023: Bringing AI Everywhere (More

News)

In a keynote presentation to open the event targeting developers, Gelsinger showed how Intel is bringing AI capabilities across its hardware products and making it accessible through open, multi-architecture software solutions. He also highlighted how AI is helping to drive the “Siliconomy,” a “growing economy enabled by the magic of silicon and software.” Today, silicon feeds a $574 billion industry that in turn powers a global tech economy worth almost $8 trillion.

New Advances in Silicon, Packaging and Multi-Chiplet Solutions

The work begins with silicon innovation. Intel’s five-nodes-in-four-years process development program is progressing well, Gelsinger said, with Intel 7 already in high-volume manufacturing, Intel 4 manufacturing-ready and Intel 3 on track for the end of this year.

Gelsinger also showed an Intel 20A wafer with the first test chips for Intel’s Arrow Lake processor, which is destined for the client computing market in 2024. Intel 20A will be the first process node to include PowerVia, Intel’s backside power delivery technology, and the new gate-all-around transistor design called RibbonFET. Intel 18A, which also leverages PowerVia and RibbonFET, remains on track to be manufacturing-ready in the second half of 2024.

Another way Intel presses Moore’s Law forward is with new materials and new packaging technologies, like glass substrates – a breakthrough Intel announced this week. When introduced later this decade, glass substrates will allow for continued scaling of transistors on a package to help meet the need for data-intensive, high-performance workloads like AI and will keep Moore’s Law going well beyond 2030.

Intel also displayed a test chip package built with Universal Chiplet Interconnect Express (UCIe). The next wave of Moore’s Law will arrive with multi-chiplet packages, Gelsinger said, coming sooner if open standards can reduce the friction of integrating IP. Formed last year, the UCIe standard will allow chiplets from different vendors to work together, enabling new designs for the expansion of diverse AI workloads. The open specification is supported by more than 120 companies.

The test chip combined an Intel UCIe IP chiplet fabricated on Intel 3 and a Synopsys UCIe IP chiplet fabricated on TSMC N3E process node. The chiplets are connected using embedded multi-die interconnect bridge (EMIB) advanced packaging technology. The demonstration highlights the commitment of TSMC, Synopsys and Intel Foundry Services to support an open standard-based chiplet ecosystem with UCIe.

Increasing Performance and Expanding AI Everywhere

Gelsinger spotlighted the range of AI technology available to developers across Intel platforms today – and how that range will dramatically increase over the coming year.

Recent MLPerf AI inference performance results further reinforce Intel’s commitment to addressing every phase of the AI continuum, including the largest, most challenging generative AI and large language models. The results also spotlight the Intel Gaudi2 accelerator as the only viable alternative on the market for AI compute needs. Gelsinger

announced a large AI supercomputer will be built entirely on Intel Xeon processors and 4,000 Intel Gaudi2 AI hardware accelerators, with Stability AI as the anchor customer.

Zhou Jingren, chief technology officer of Alibaba Cloud, explained how Alibaba applies 4th Gen Intel® Xeon® processors with built-in AI acceleration to "our generative AI and large language model, Alibaba Cloud's Tongyi Foundation Models." Intel's technology, he said, results in "remarkable improvements in response times, averaging a 3x acceleration."[1]

Intel also previewed the next generation of Intel Xeon processors, revealing that 5th Gen Intel® Xeon® processors will bring a combination of performance improvements and faster memory, while using the same amount of power, to the world's data centers when they launch Dec. 14. Sierra Forest, with E-core efficiency and arriving in the first half of 2024, will deliver 2.5x better rack density and 2.4x higher performance per watt over 4th Gen Xeon and will include a version with 288 cores[2]. And Granite Rapids, with P-core performance, will closely follow the launch of Sierra Forest, offering 2x to 3x better AI performance compared to 4th Gen Xeon[2].

Looking ahead to 2025, the next-gen E-core Xeon, code-named Clearwater Forest, will arrive on the Intel 18A process node.

Introducing the AI PC with Intel Core Ultra processors

AI is about to get more personal, too. "AI will fundamentally transform, reshape and restructure the PC experience – unleashing personal productivity and creativity through the power of the cloud and PC working together," Gelsinger said. "We are ushering in a new age of the AI PC."

This new PC experience arrives with the upcoming Intel Core Ultra processors, code-named Meteor Lake, featuring Intel's first integrated neural processing unit, or NPU, for power-efficient AI acceleration and local inference on the PC. Gelsinger confirmed Core Ultra also will launch Dec. 14.

Core Ultra represents an inflection point in Intel's client processor roadmap: It's the first client chiplet design enabled by Foveros packaging technology. In addition to the NPU and major advances in power-efficient performance thanks to Intel 4 process technology, the new processor brings discrete-level graphics performance with onboard Intel® Arc™ graphics.

On stage, Gelsinger showed an array of new AI PC use cases, and Jerry Kao, chief operating officer of Acer, gave a sneak peek at an upcoming Acer laptop powered by Core Ultra. "We've been co-developing with Intel teams a suite of Acer AI applications to take advantage of the Intel Core Ultra platform," Kao said, "developing with the OpenVINO toolkit and co-developed AI libraries to bring the hardware to life."

Putting Developers in the Siliconomy Driver's Seat

"AI going forward must deliver more access, scalability, visibility, transparency and trust to the whole ecosystem," Gelsinger said.

To help developers unlock this future, Intel announced:

- **General availability of the Intel Developer Cloud:** The Intel Developer Cloud helps developers accelerate AI using the latest Intel hardware and software innovations – including Intel Gaudi2 processors for deep learning – and provides access to the latest Intel hardware platforms, such as the 5th Gen Intel® Xeon® Scalable processors and Intel® Data Center GPU Max Series 1100 and 1550. When using the Intel Developer Cloud, developers can build, test and optimize AI and HPC applications. They can also run small- to large-scale AI training, model optimization and inference workloads that deploy with performance and efficiency. Intel Developer Cloud is based on an open software foundation with oneAPI – an open multiarchitecture, multivendor programming model – to provide hardware choice and freedom from proprietary programming models to support accelerated computing and code reuse and portability.
- **The 2023.1 release of the Intel Distribution of OpenVINO toolkit:** OpenVINO is Intel's AI inferencing and deployment runtime of choice for developers on client and edge platforms. The release includes pre-trained models optimized for integration across operating systems and different cloud solutions, including many generative AI models, such as the Llama 2 model from Meta. On stage, companies including ai.io and Fit:match demonstrated how they use OpenVINO to accelerate their applications: ai.io to evaluate the performance of any potential athlete; Fit:match to revolutionize the retail and wellness industries to help consumers find the best-fitting garments.
- **Project Strata, and the development of an edge-native software platform:** The platform launches in 2024 with modular building blocks, premium service and support offerings. It is a horizontal approach to scaling the needed infrastructure for the intelligent edge and hybrid AI and will bring together an ecosystem of Intel and third-party vertical applications. The solution will enable developers to build, deploy, run, manage, connect and secure distributed edge infrastructure and applications.

This is just the beginning of the news from Intel Innovation. Tune in on the Intel Newsroom at 9:30 a.m. PDT Wednesday to hear from Intel Chief Technology Officer Greg Lavender on more ways Intel is opening opportunities in AI for developers and accelerating the convergence of AI and security.

Forward-Looking Statements

This release contains forward-looking statements, including with respect to Intel's business plans and strategy, process and product roadmaps, and current and future technologies, as well as the 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; changes in product mix; the complexity and fixed cost nature of our manufacturing operations; the high level of competition and rapid technological change in our industry; the significant upfront investments in R&D and our business, products, technologies, and manufacturing capabilities; 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; investment and transaction risk; intellectual property risks and risks associated with litigation and regulatory proceedings; evolving regulatory and legal requirements across many

jurisdictions; geopolitical and international trade conditions; our debt obligations; risks of large scale global operations; macroeconomic conditions; impacts of the COVID 19 or similar such pandemic; and other risks and uncertainties described in our earnings release dated July 27, 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.

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] Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.
[2] Based on architectural projections as of August 21, 2023 relative to 4th Gen Intel Xeon processors. Your results may vary.

© 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/20230919339712/en/

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

Source: Intel