

Intel Builds World's Largest Neuromorphic System to Enable More Sustainable AI

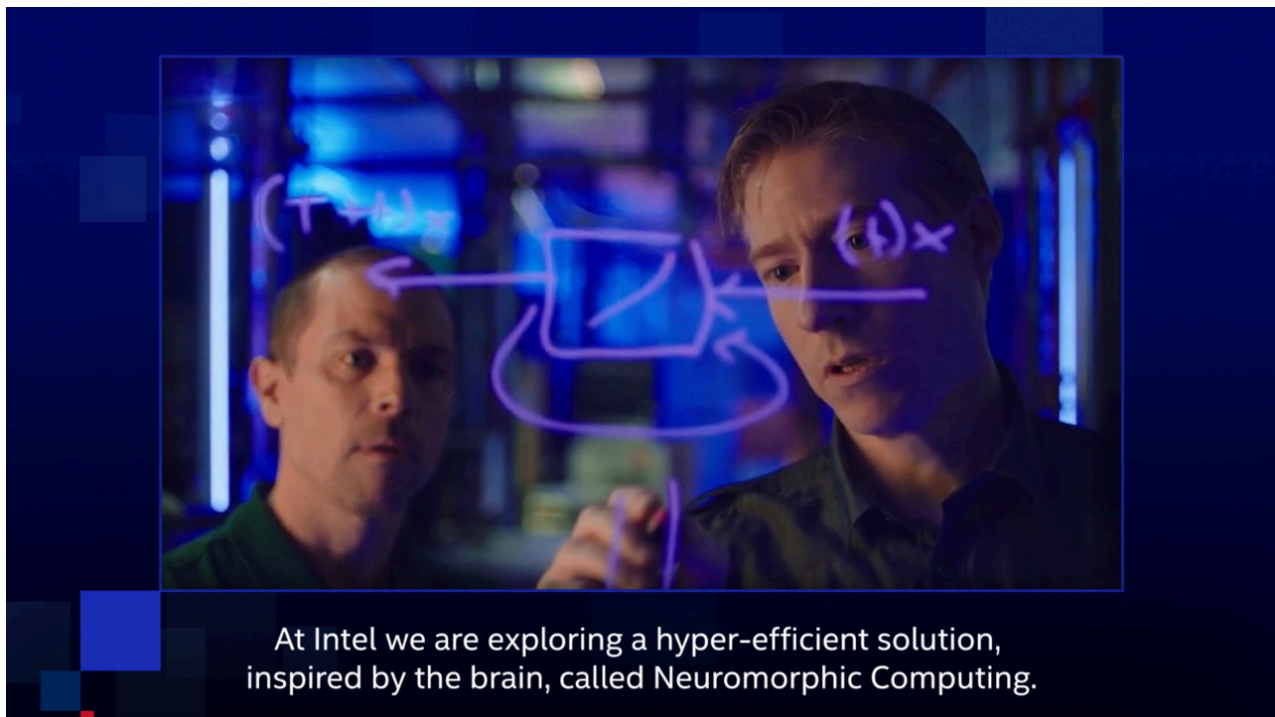
Hala Point, the industry's first 1.15 billion neuron neuromorphic system, builds a path toward more efficient and scalable AI.

News

- April 17, 2024
- [Contact Intel PR](#)
- Follow Intel Newsroom on social:



[More Artificial Intelligence News](#)



At Intel we are exploring a hyper-efficient solution, inspired by the brain, called Neuromorphic Computing.

What's New: Today, Intel announced that it has built the world's largest neuromorphic system. Code-named Hala Point, this large-scale neuromorphic system, initially deployed at Sandia National Laboratories, utilizes Intel's [Loihi 2](#) processor, aims at supporting research for future brain-inspired artificial intelligence (AI), and tackles challenges related to the efficiency and sustainability of today's AI. Hala Point advances Intel's first-generation large-scale research system, [Pohoiki Springs](#), with architectural improvements to achieve over 10 times more neuron capacity and up to 12 times higher performance.

"The computing cost of today's AI models is rising at unsustainable rates. The industry needs fundamentally new approaches capable of scaling. For that reason, we developed Hala Point, which combines deep learning efficiency with novel brain-inspired learning and optimization capabilities. We hope that research with Hala Point will advance the efficiency and adaptability of large-scale AI technology."

—Mike Davies, director of the Neuromorphic Computing Lab at Intel Labs

What It Does: Hala Point is the first large-scale neuromorphic system to demonstrate state-of-the-art computational efficiencies on mainstream AI workloads. Characterization shows it can support up to 20 quadrillion operations per second, or 20 petaops, with an efficiency exceeding 15 trillion 8-bit operations per second per watt (TOPS/W) when executing conventional deep neural networks. This rivals and exceeds levels achieved by architectures built on graphics processing units (GPU) and central processing units (CPU). Hala Point's unique capabilities could enable future real-time [continuous learning](#) for AI applications such as scientific and engineering problem-solving, logistics, smart city infrastructure management, large language models (LLMs) and AI agents.

How It will be Used: Researchers at Sandia National Laboratories plan to use Hala Point for advanced brain-scale computing research. The organization will focus on solving scientific computing problems in device physics, computer architecture, computer science and informatics.

"Working with Hala Point improves our Sandia team's capability to solve computational and scientific modeling problems. Conducting research with a system of this size will allow us to keep pace with AI's evolution in fields ranging from commercial to defense to basic science," said Craig Vineyard, Hala Point team lead at Sandia National Laboratories.

Currently, Hala Point is a research prototype that will advance the capabilities of future commercial systems. Intel anticipates that such lessons will lead to practical advancements, such as the ability for LLMs to learn continuously from new data. Such advancements promise to significantly reduce the unsustainable training burden of widespread AI deployments.

Why It Matters: Recent trends in scaling up deep learning models to trillions of parameters have exposed daunting sustainability challenges in AI and have highlighted the need for innovation at the lowest levels of hardware architecture. Neuromorphic computing is a fundamentally new approach that draws on neuroscience insights that integrate memory and computing with highly granular parallelism to minimize data movement. In [published results](#) from this month's International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Loihi 2 demonstrated orders of magnitude gains in the efficiency, speed and adaptability of emerging small-scale edge workloads¹.

Advancing on its predecessor, Pohoiki Springs, with numerous improvements, Hala Point now brings neuromorphic performance and efficiency gains to mainstream conventional deep learning models, notably those processing real-time workloads such as video, speech and wireless communications. For example, Ericsson Research is applying Loihi 2 to optimize [telecom infrastructure efficiency](#), as highlighted at this year's Mobile World Congress.

Hala Point, the World's Largest Neuromorphic System

Largest capacity: 1.15 billion neurons and 128 billion synapses

Scalable: 1,152 Loihi 2 processors with 140,544 neuromorphic cores and 2,304 x86 cores



Fast: 380 trillion synaptic operations per second, 240 trillion neuron operations per second, 16 petabytes per second of memory bandwidth

Efficient: 15 TOPS/W executing sparse deep neural networks with an equivalent 20 quadrillion 8-bit operations per second*

Introducing Intel's most advanced neuromorphic system to date, code-named Hala Point. This system demonstrates state-of-the-art computational efficiencies for more efficient and sustainable AI.

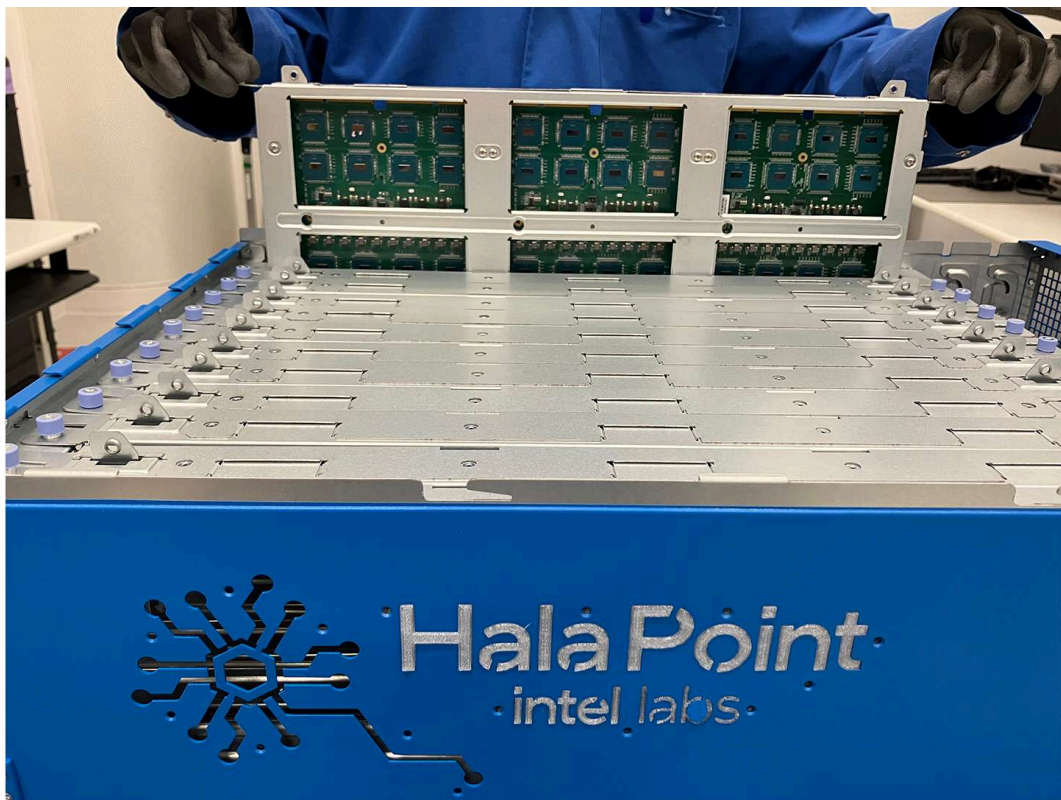
Hala Point could enable advancements in future real-time continual learning, inference, and optimization for applications like science and engineering problem-solving, logistics, smart city infrastructure management, large language models (LLMs) and AI agents.

*Characterization performed with a multi-layer perceptron (MLP) network with 14,784 layers, 2,048 neurons per layer 8-bit weights stimulated with random noise. The Hala Point implementation of the MLP network is pruned to 10:1 sparsity with sigma-delta neuron models providing 10% activation rates. Results as of testing in April 2024. Results may vary.

© Copyright Intel Corporation. All rights reserved. Intel and the Intel logo are trademarks of Intel Corporation in the U.S. and/or other countries. Other names and brands may be claimed as the property of others.

intel.







intel.

The world's largest neuromorphic research system, code-named

Hala Point



Source: Intel March 2024

Hala Point, the World's Largest Neuromorphic System

Largest capacity: 1.15 billion neurons and 128 billion synapses

Scalable: 1,152 Loihi 2 processors with 140,544 neuromorphic cores and 2,304 x86 cores



Fast: 380 trillion synaptic operations per second, 240 trillion neuron operations per second, 16 petabytes per second of memory bandwidth

Efficient: 15 TOPS/W executing sparse deep neural networks with an equivalent 20 quadrillion 8-bit operations per second*

Introducing Intel's most advanced neuromorphic system to date, code-named Hala Point. This system demonstrates state-of-the-art computational efficiencies for more efficient and sustainable AI.

Hala Point could enable advancements in future real-time continual learning, inference, and optimization for applications like science and engineering problem-solving, logistics, smart city infrastructure management, large language models (LLMs) and AI agents.

*Characterization performed with a multi-layer perceptron (MLP) network with 14,784 layers, 2,048 neurons per layer 8-bit weights stimulated with random noise. The Hala Point implementation of the MLP network is pruned to 10:1 sparsity with sigma-delta neuron models providing 10% activation rates. Results as of testing in April 2024. Results may vary.

© Copyright Intel Corporation. All rights reserved. Intel and the Intel logo are trademarks of Intel Corporation in the U.S. and/or other countries. Other names and brands may be claimed as the property of others.

intel.





[Download all images \(ZIP, 17 MB\)](#)

About Hala Point: Loihi 2 neuromorphic processors, which form the basis for Hala Point, apply brain-inspired computing principles, such as asynchronous, event-based spiking neural networks (SNNs), integrated memory and computing, and sparse and continuously changing connections to achieve orders-of-magnitude gains in energy consumption and performance. Neurons communicate directly with one another rather than communicating through memory, reducing overall power consumption.

Hala Point packages 1,152 Loihi 2 processors produced on Intel 4 process node in a six-rack-unit data center chassis the size of a microwave oven. The system supports up to 1.15 billion neurons and 128 billion synapses distributed over 140,544 neuromorphic processing cores, consuming a maximum of 2,600 watts of power. It also includes over 2,300 embedded x86 processors for ancillary computations.

Hala Point integrates processing, memory, and communication channels in a massively parallelized fabric, providing a total of 16 petabytes per second (PB/s) of memory bandwidth, 3.5 PB/s of inter-core communication bandwidth, and 5 terabytes per second (TB/s) of inter-chip communication bandwidth. The system can process over 380 trillion 8-bit synapses and over 240 trillion neuron operations per second.

Applied to bio-inspired spiking neural network models, the system can execute its full capacity of 1.15 billion neurons 20 times faster than a human brain and up to 200 times faster rates at lower capacity. While Hala Point is not intended for neuroscience modeling, its neuron capacity is roughly equivalent to that of an owl brain or the cortex of a capuchin monkey.

Loihi-based systems can perform AI inference and solve optimization problems using 100 times less energy at speeds as much as 50 times faster than conventional CPU and GPU architectures¹. By exploiting up to 10:1 sparse connectivity and event-driven activity, early results on Hala Point show the system can achieve deep neural network efficiencies as high as 15 TOPS/W² without requiring input data to be collected into batches, a common optimization for GPUs that significantly delays the

processing of data arriving in real-time, such as video from cameras. While still in research, future neuromorphic LLMs capable of continuous learning could result in gigawatt-hours of energy savings by eliminating the need for periodic re-training with ever-growing datasets.

What's Next: The delivery of Hala Point to Sandia National Labs marks the first deployment of a new family of large-scale neuromorphic research systems that Intel plans to share with its research collaborators. Further development will enable neuromorphic computing applications to overcome power and latency constraints that limit AI capabilities' real-world, real-time deployment.

Together with an ecosystem of more than 200 Intel Neuromorphic Research Community ([INRC](#)) members, including leading academic groups, government labs, research institutions and companies worldwide, Intel is working to push the boundaries of brain-inspired AI and progressing this technology from research prototypes to industry-leading commercial products over the coming years.

More context: [Intel Labs](#)

The Small Print:

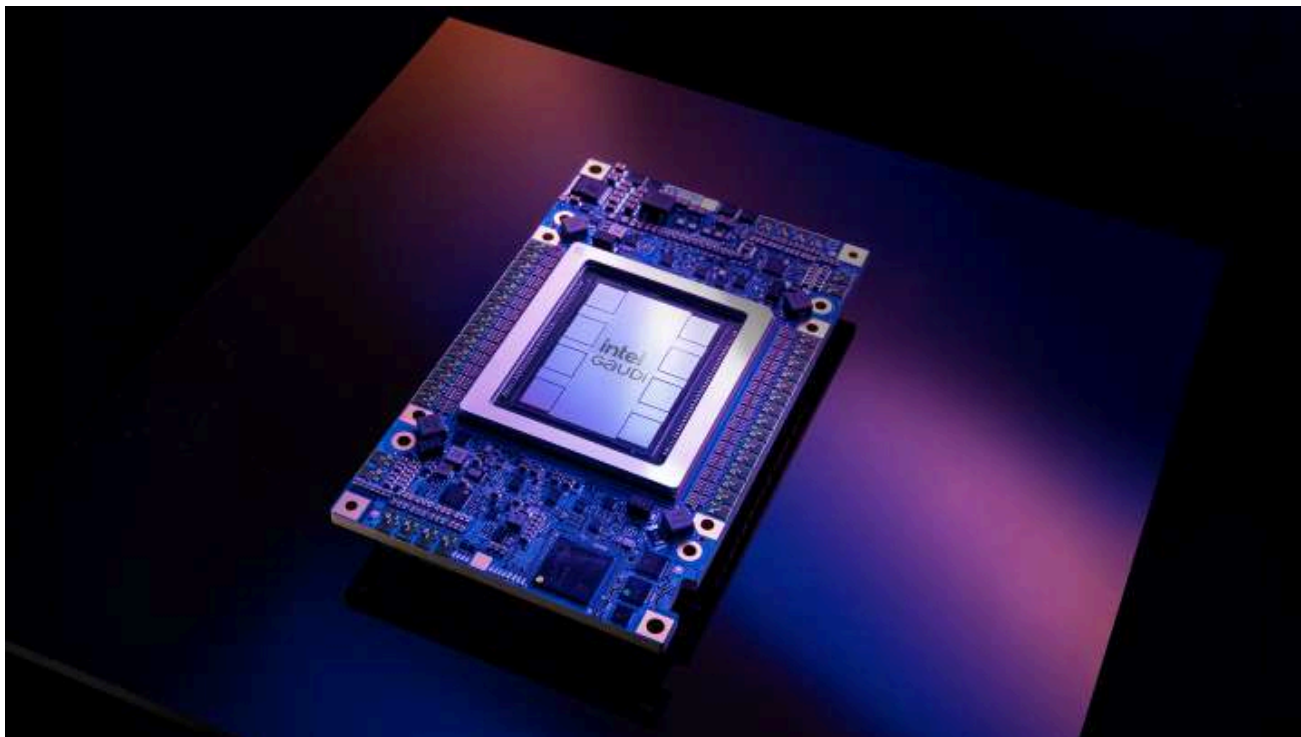
¹ See "[Efficient Video and Audio Processing with Loihi 2](#)," International Conference on Acoustics, Speech, and Signal Processing, April 2024, and "[Advancing Neuromorphic Computing with Loihi: Survey of Results and Outlook](#)," Proceedings of the IEEE, 2021.

² Characterization performed with a multi-layer perceptron (MLP) network with 14,784 layers, 2048 neurons per layer, 8-bit weights stimulated with random noise. The Hala Point implementation of the MLP network is pruned to 10:1 sparsity with sigma-delta neuron models providing 10 percent activation rates. Results as of testing in April 2024. Results may vary.

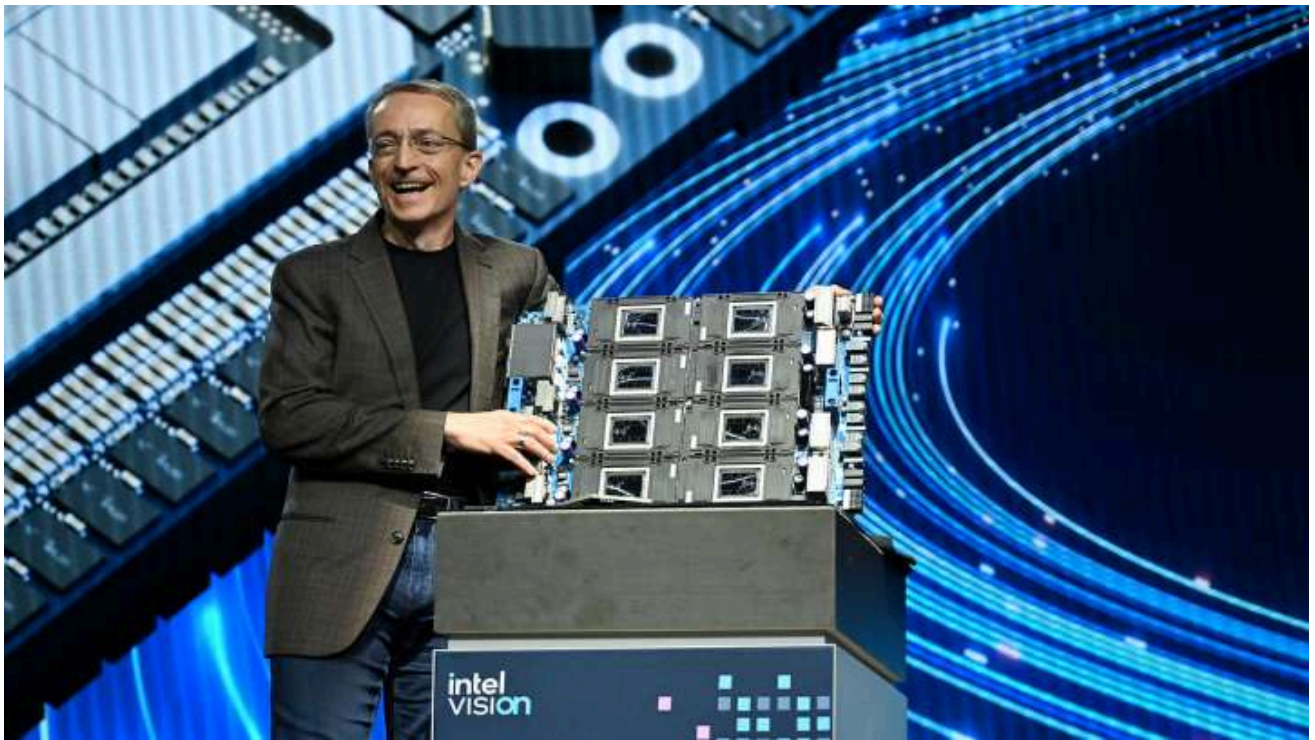
Tags

[Artificial Intelligence](#), [New Technologies](#)

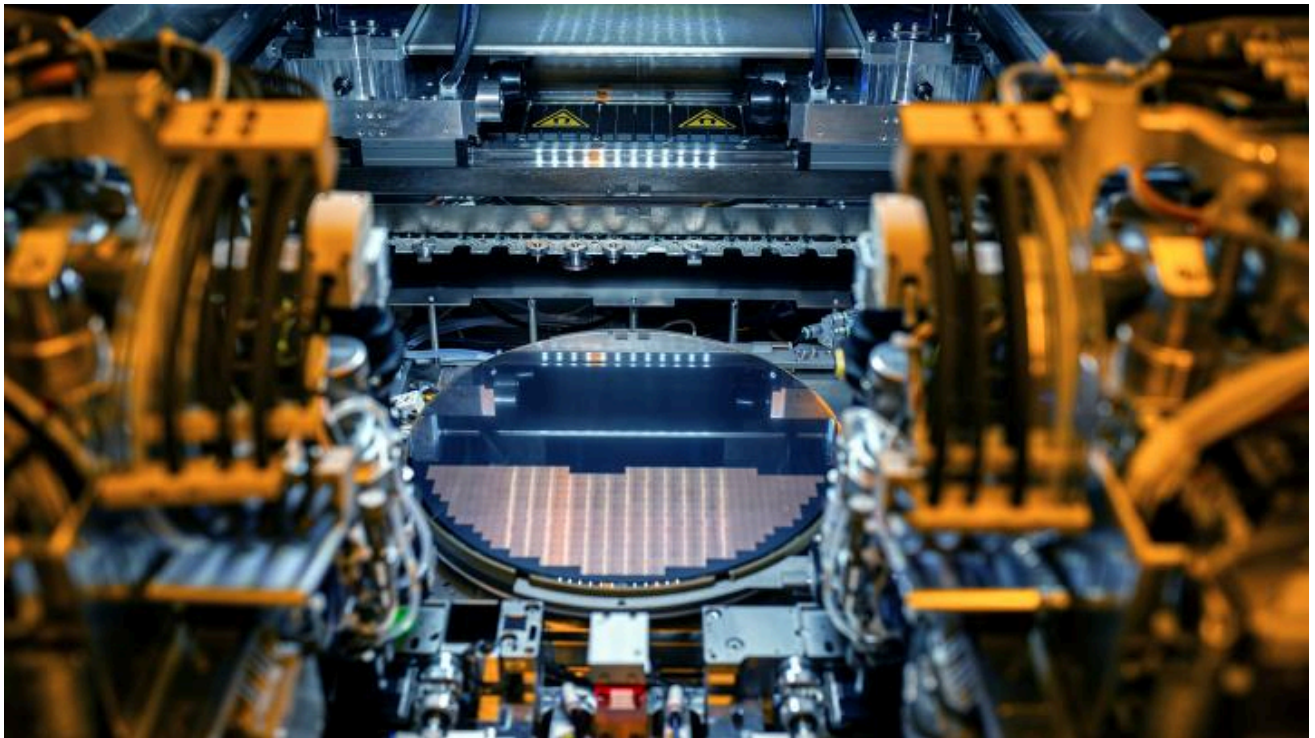
Related Stories



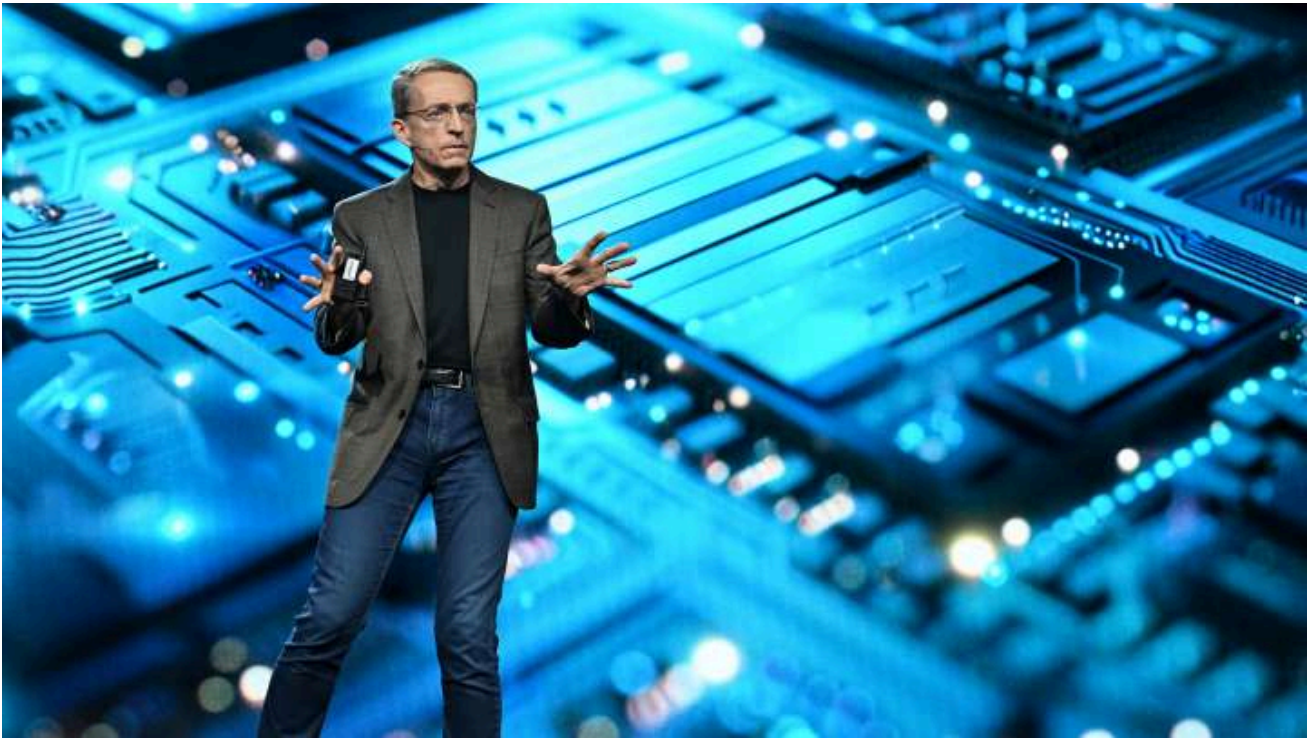
[Intel Breaks Down Proprietary Walls to Bring Choice to Enterprise...](#)



[Intel Unleashes Enterprise AI with Gaudi 3, AI Open Systems Strategy...](#)



[Intel Foundry Direct Connect Keynote \(Replay\).](#)



[Intel Vision 2024 Keynote \(Replay\)](#)

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.

© 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.