



HPC/AI開発の常識を覆す インテルの最新テクノロジー

インテル株式会社

インダストリー事業本部 シニア・ソリューション・アーキテクト

高藤 良史

Notices and Disclaimers

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. Results have been estimated or simulated.

Your costs and results may vary.

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

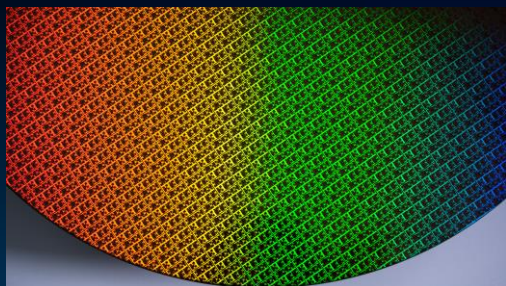
Intel does not control or audit third-party data. You should consult other sources to evaluate accuracy.

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

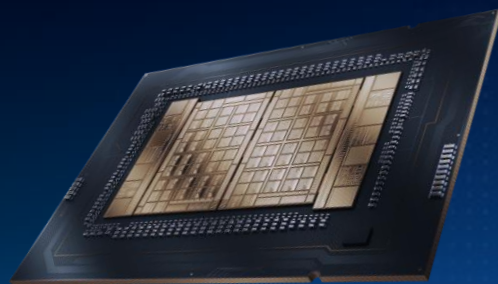
Code names are used by Intel to identify products, technologies, or services that are in development and not publicly available. These are not "commercial" names and not intended to function as trademarks.

© 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.

インテルの最新テクノロジー



ファウンドリー
技術

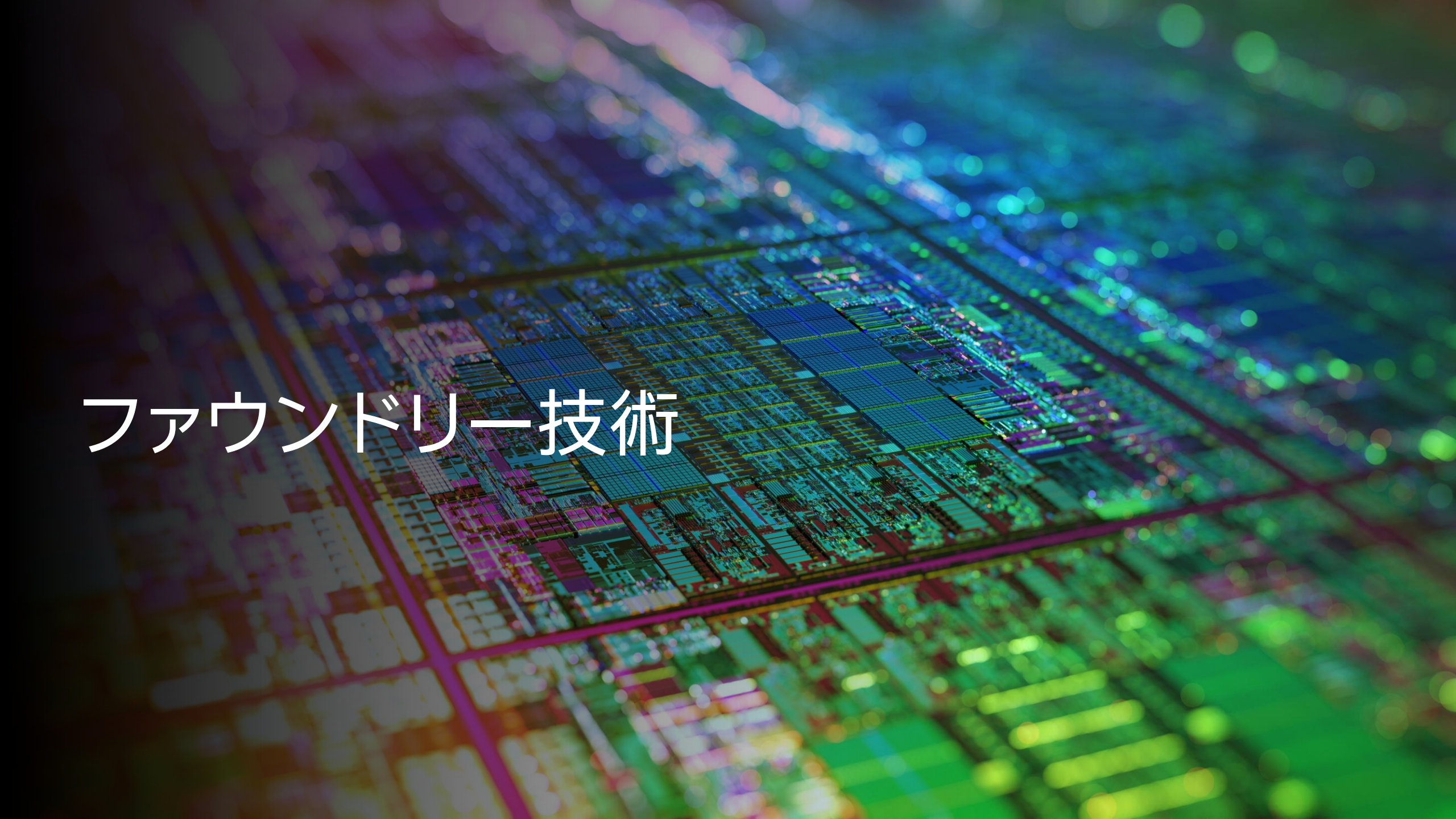


インテル® Xeon® 6
プロセッサ



インテル® Gaudi® 3
AI アクセラレーター

ファウンドリー技術



Intel Foundry Process Roadmap

2025 Availability

2026

2027

2028

RibbonFET +
Backside Power

Intel 14A

Industry 1st
High-NA EUV
PowerDirect

intel
14A

intel
14A-E

Future*
Derivative

Intel 18A

Industry 1st
PowerVia

intel
18A

intel
18A-P

intel
18A-PT

FinFET

Intel 3

intel
4

intel
3

intel
3-T

intel
3-E

Future*
Derivative

Mature
Nodes

intel
7

intel
16

intel
16-E

UMC
12
intel

Future*
Derivative

P = Performance improvement, T = Through silicon vias for 3D stacking, E = Feature Extension

*In planning

All product and service plans, roadmaps, and performance estimates are subject to change without notice.

Foundry
Enabled

intel foundry

インテル® Xeon® 6 プロセッサ

Imperial College London が インテル® Xeon® 6 を最新のHPCに採用

水冷システムにより、環境への影響を抑えつつ、科学研究の計算性能を向上

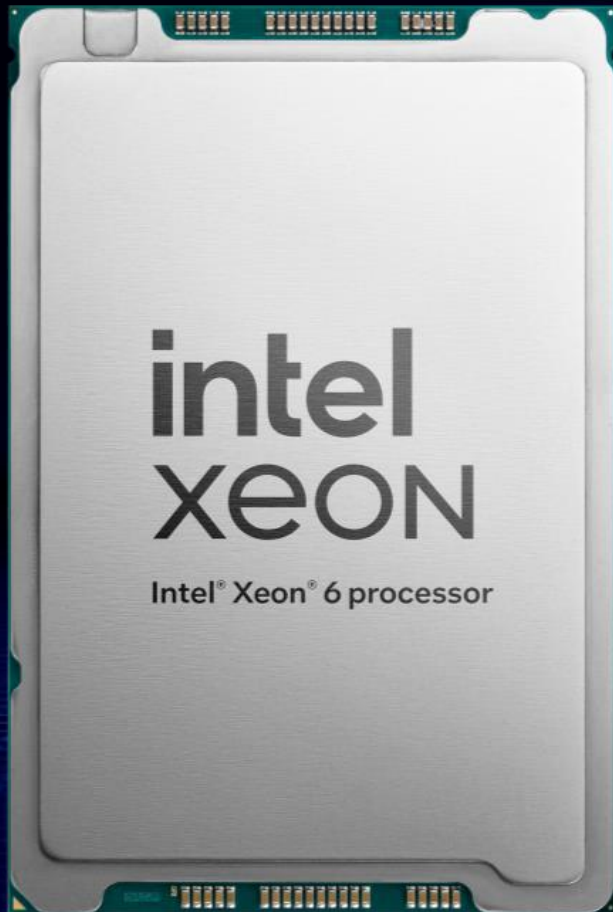


インテルはインペリアル・カレッジ・ロンドンに新しいHX2スーパーコンピュータの計画を発表。
Lenovoサーバー上に構築され、インテル® Xeon® 6プロセッサ (P-cores採用)によって運用されます。
水冷システムは、共同投資ICICLEコラボレーションの一部として設計されており、今年導入され、次世代の高性能コンピューティング(HPC)と人工知能(AI)の性能向上を提供します。

HX2スーパーコンピュータは、研究を加速し、HPC/AIキャリアに向けた準備を整えるための装備を提供します。この新しいシステムは、商業コロケーションデータセンターにおけるLenovoのThinkSystem SC750 V4 Neptuneサーバーの直接液冷の英国初の学術導入となります。

インテル® Xeon® 6900P プロセッサー

パフォーマンスと効率性の多様な要件を満たす最高水準のプロセッサー



最大 6400 MT/s のDDR5

8800 MT/s のMRDIMMメモリ

最大 128 個のパフォーマンスコア

最大24 GT/sの6つのUPI 2.0リンク

最大 96レーン のPCIe 5.0/CXL® 2.0

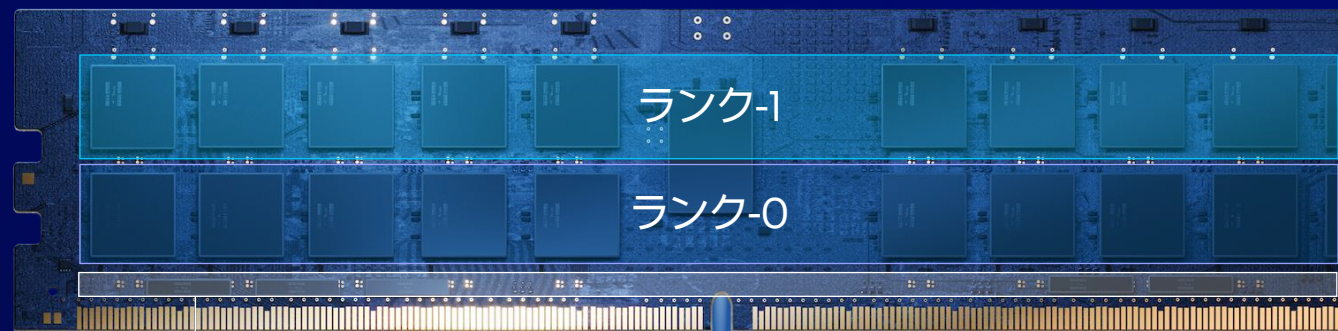
最大 504 MB のL3キャッシュ

FP16をサポートする
インテル® アドバンスド・マトリクス・エクステンション (インテル® AMX)

多重化ランクDIMM (MRDIMM) による 大幅な帯域幅の拡大

インテル® Xeon® 6 (P-cores採用) シリーズ

MRDIMMモジュール



多重化
チップ

2倍の
DDR5転送

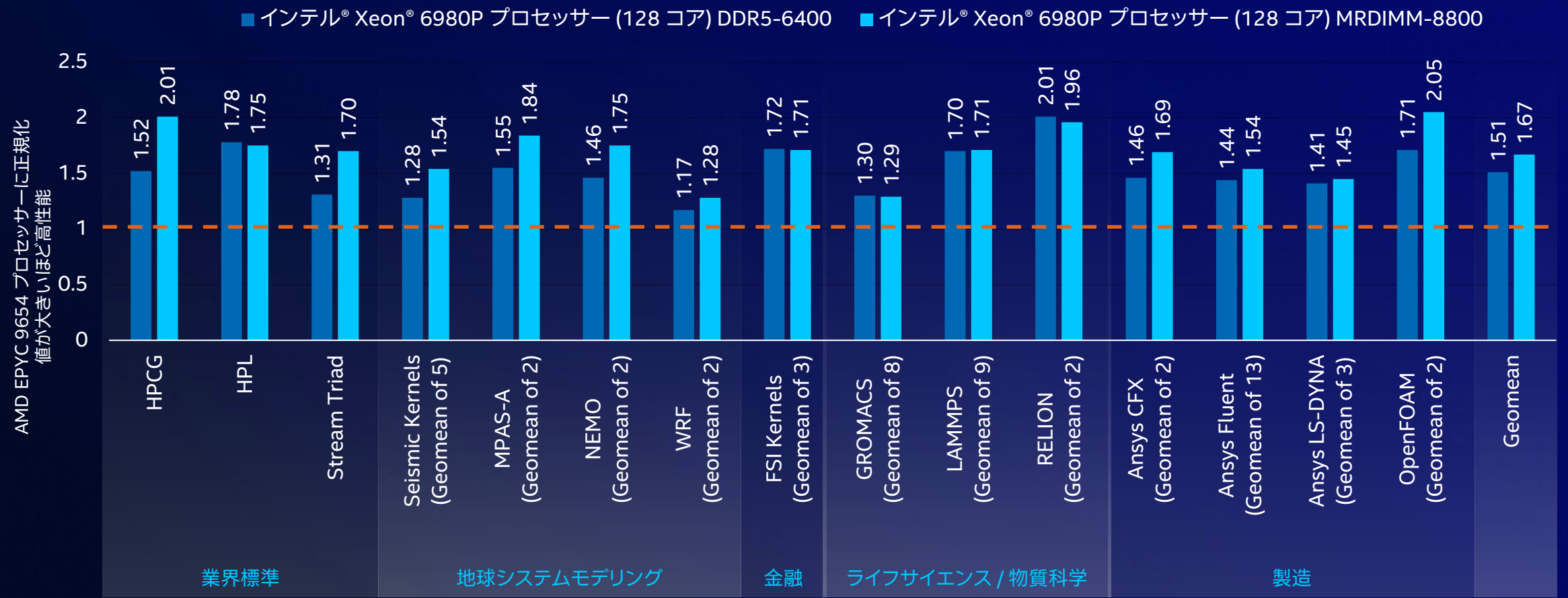
最大
8800 MT/s
データレート

メモリー帯域幅
スケーリングの
改善

メモリー・バウンド
ワークロードに
最適

HPC ワークロードにおけるトップレベルのパフォーマンス

コア数の増加とメモリー帯域幅の拡大



<https://www.intel.com/processorclaims/> (英語): Intel® Xeon® 6 の [9H5, 9H6] を参照。結果は異なる場合があります。
ここで記載する内容は OpenCFD Limited の承諾または承認を受けたものではありません。OpenCFD Limited は、OpenFOAM ソフトウェア (<http://www.openfoam.com>) の開発元およびディストリビューターであり、OpenFOAM、OpenCFD の商標を所有します。

CXL[®]アタッチメモリー (フラット・メモリー・モード)

インテル[®] Xeon[®] 6 プロセッサではコスト効率が良く、柔軟なソリューションに

CXL NUMAノード



2 NUMAノード

- SW (ハイパーバイザー/OS/アプリ) 支援ティアリング
- 4KBページ転送、移動オーバーヘッドあり

CXLヘテロ・インターリーブ



1 NUMAノード

- HWベースのメモリーインターリーブ
- DRAMとCXLメモリーアドレス空間のインターリーブ

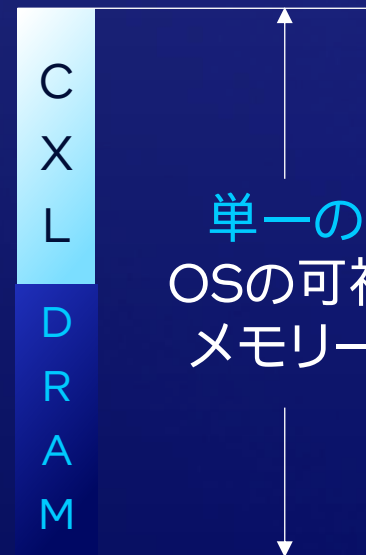
P-cores 搭載 インテル[®] Xeon[®] 6900/6700シリーズプロセッサのみ

フラット・メモリー・モード

50% CXL
メモリー



50%
DRAM



- HW支援ティアリングにより、完全にSWを透過
- 単一のネイティブDDR + CXLメモリー領域
- 1:1 近/遠メモリー比

インテル® TDX Connect コンフィデンシャルAIの次の一歩

CPUとPCIeデバイス間に高性能の暗号化接続を提供し、
直接メモリアクセスと低オーバーヘッドを実現



インテル® Xeon® プロセッサ

AIアクセラレーション・システム向けで最も採用されているホストCPU

AIアクセラレーション・
システムのパフォーマンスを
最大化するホストCPUの役割:

- モデルトレーニングデータの準備
を行うデータの前処理
- GPUへのデータ伝送による
並列コンピューティング
- システムメモリへの
チェックポイント管理
- 同一のインフラストラクチャーで
実行している複数のワークロード
を処理するための柔軟性

ホストCPUの技術要件

インテル Xeon 6 を選ぶべき特徴

優れたI/O性能

最大192レーンのPCIe 5.0
前世代より20%多いレーン数と高いI/O帯域幅を実現¹

高い単一スレッド性能

高いコア性能、最大ターボ周波数

高いメモリ帯域幅と容量

MRDIMMメモリにより30%高い帯域幅²
CXLによるメモリ容量拡張

大規模システム向けのRAS

高度なRASサポート
(例: PCIe強化ダウンストリームポートコンテ
インメント-eDPC)

多様なフォームファクターをサポート

DC-MHS、NVIDIA® MGX™ をサポート

インテル® ソフトウェア開発ツール

oneAPIによって実現されるフレキシブルで包括的なオープン・ソフトウェア・スタック



intel.
**AI
TOOLS**

1
oneAPI

Data Analytics at Scale:

MODIN

pandas

NumPy

SciPy

DL Inference and Training:

TensorFlow

PyTorch

OpenVINO

Intel® Neural
Compressor

Classical ML:

sklearn

dmlc
XGBoost

python™

Intel® Distribution
for Python

Package & Environment
Managers

Data Processing & Modeling
Packages

Machine Learning Packages
Advanced Programming Packages

Python Interpreters & Compilers
Development Packages & Runtimes

intel.

1
oneAPI

HPC
TOOLKIT

Base Toolkit Components +

Intel® Fortran Compiler

Intel® MPI Library

Intel® SHMEM Library

intel.

1
oneAPI

BASE
TOOLKIT

Tools

Intel® DPC++/C++
Compatibility Tool

Intel® VTune™ Profiler

Intel® Advisor

Intel® Distribution
for GDB

Performance Libraries:

oneMKL

oneDNN

oneDAL

oneCCL

oneTBB

oneDPL

Intel® IPP
Intel® Cryptography Primitives

Compilers:

Intel® oneAPI DPC++/C++ Compiler

Direct Programming:

C++ with SYCL

C++

C

Python

OpenMP

OpenCL

Fortran

CPU

GPU

FPGA¹

NPU¹

Download at intel.com/oneAPI or run tools on the Intel® Tiber™ AI Cloud at cloud.intel.com

*Other names and brands may be claimed as the property of others. SYCL is a trademark of the Khronos Group Inc.

1. Limited software support is available. For more information on FPGA support, see Intel oneAPI DPC++/C++ Compiler [Release Notes](#).

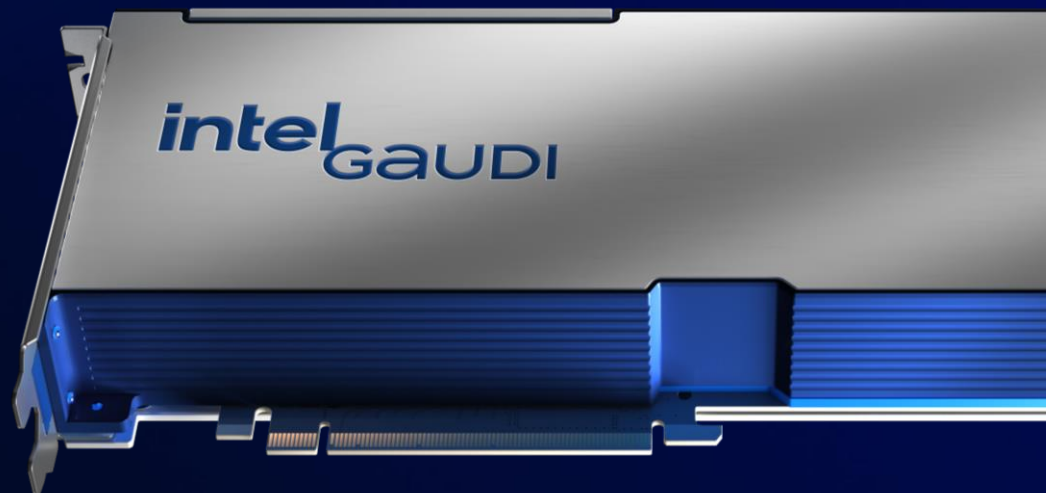
A photograph of a server rack in a data center. The rack is filled with server units, and the scene is illuminated with blue and green light. The text "Bringing AI Everywhere" is overlaid in white.

Bringing AI Everywhere

intel Gaudi

PCIe CEM

HL-338 PCIe アドインカード



1,678

TFlops
FP8 (最大)

128GB

HBM2e

600W

TDP

8

行列乗算
エンジン

18ポート

200GbE
RDMA NIC

デュアルスロット
フルハイト
長さ10.5インチ

インテル® Gaudi® 3 AI アクセラレーター: 価格・パフォーマンス比で優れた選択肢

最大

43%

高いスループット
(トークン毎秒)

IBM Granite-3.1-8B-Instruct
でのNVIDIA H200の比較
コンテキストサイズ: 小

最大

335%

高いコスト効率
(トークン毎単価)

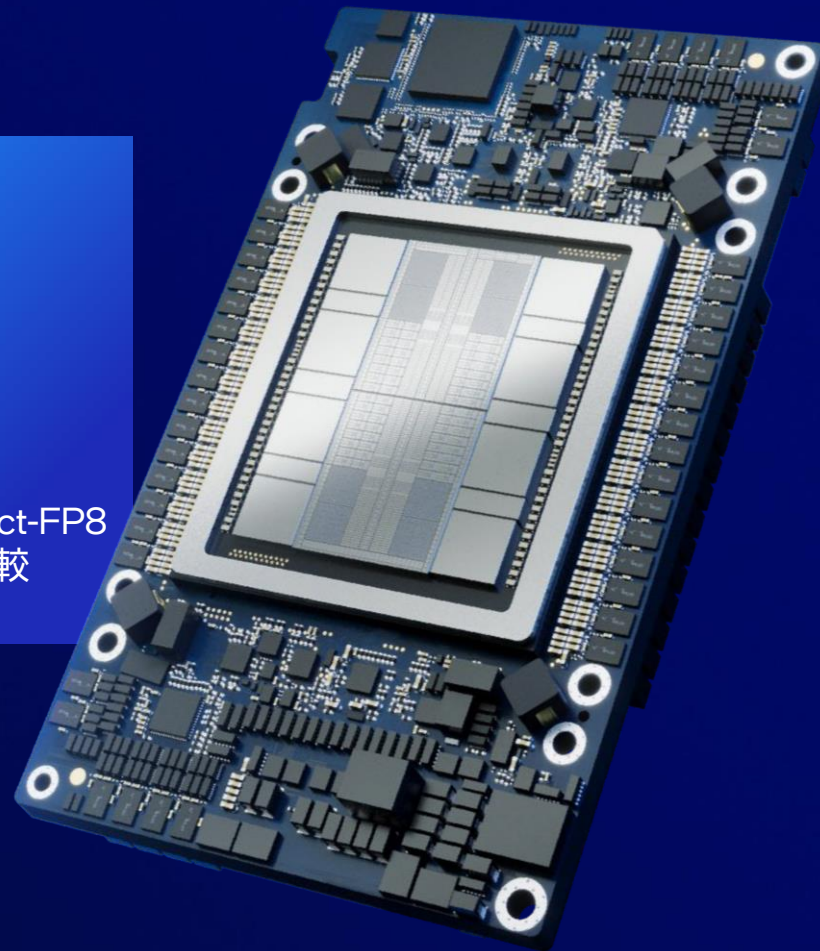
Llama-3.1-405B-Instruct-FP8
でのNVIDIA H100の比較
コンテキストサイズ: 大

最大

92%

高いコスト効率
(トークン毎単価)

Llama-3.1-405B-Instruct-FP8
でのNVIDIA H100の比較
コンテキストサイズ: 大

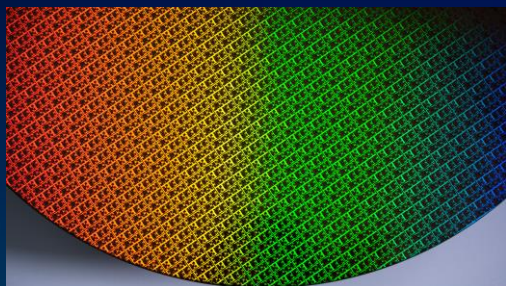


出典: [Signal65 ラボ インサイト レポート - Intel® Gaudi® 3 AI Accelerator at Scale on IBM Cloud](#)

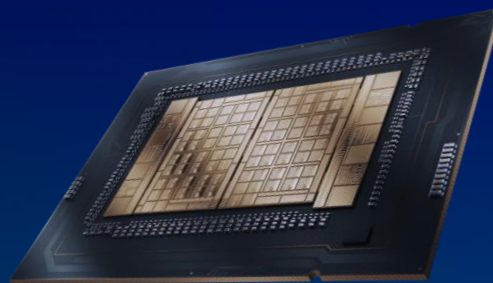
インテルの委託による研究 Signal65による 2025年4月公開。

[ワークロードと構成についてはSignal65のレポートを参照](#)。結果は異なる可能性があります。

インテルの最新テクノロジー



ファウンドリー
技術



インテル® Xeon® 6
プロセッサ



インテル® Gaudi® 3
AI アクセラレーター

The Intel logo is centered on a dark blue background. It features the word "intel" in a white, lowercase, sans-serif font. A small, solid blue square is positioned above the first vertical stroke of the letter "i".

intel