HPC-AIの革新

ＨＰＣ事業開発部長

矢澤　克巳

飽くなきシミュレーションのニーズ

金融

製造

科学

地球モデル

人工知能

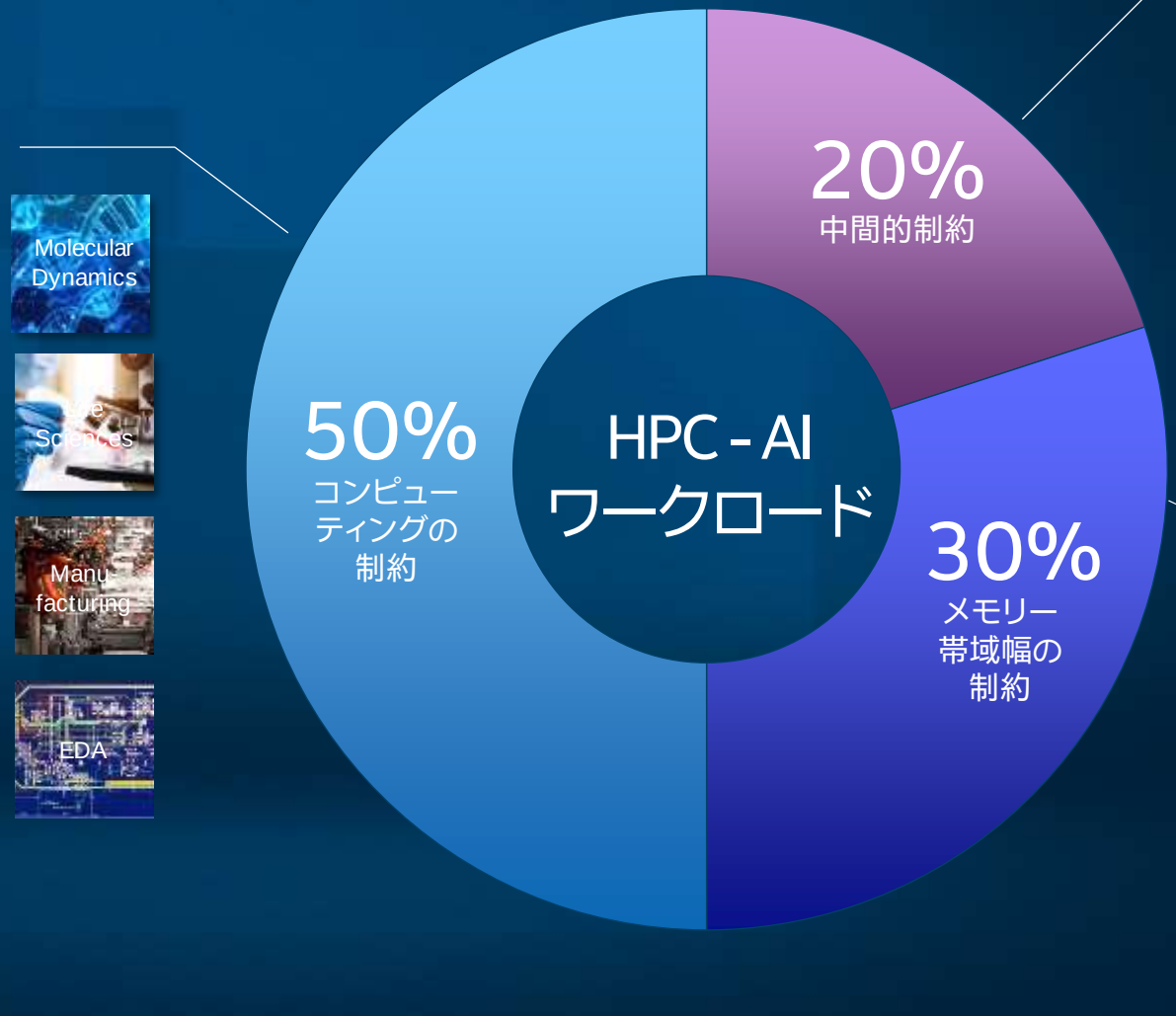
EDA

分子動力学

オイル&ガス

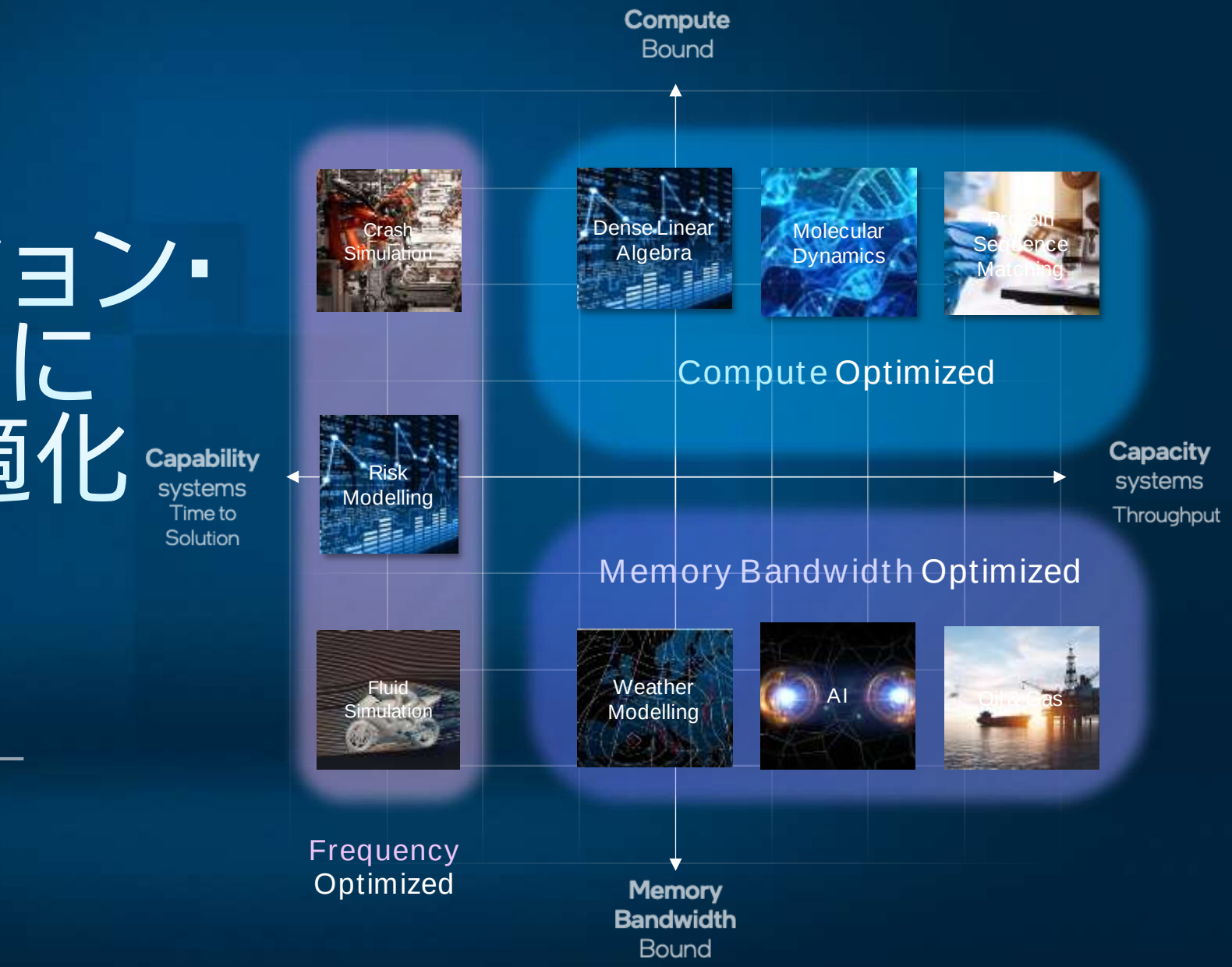
物理

ワークロードによって異なるボトルネック

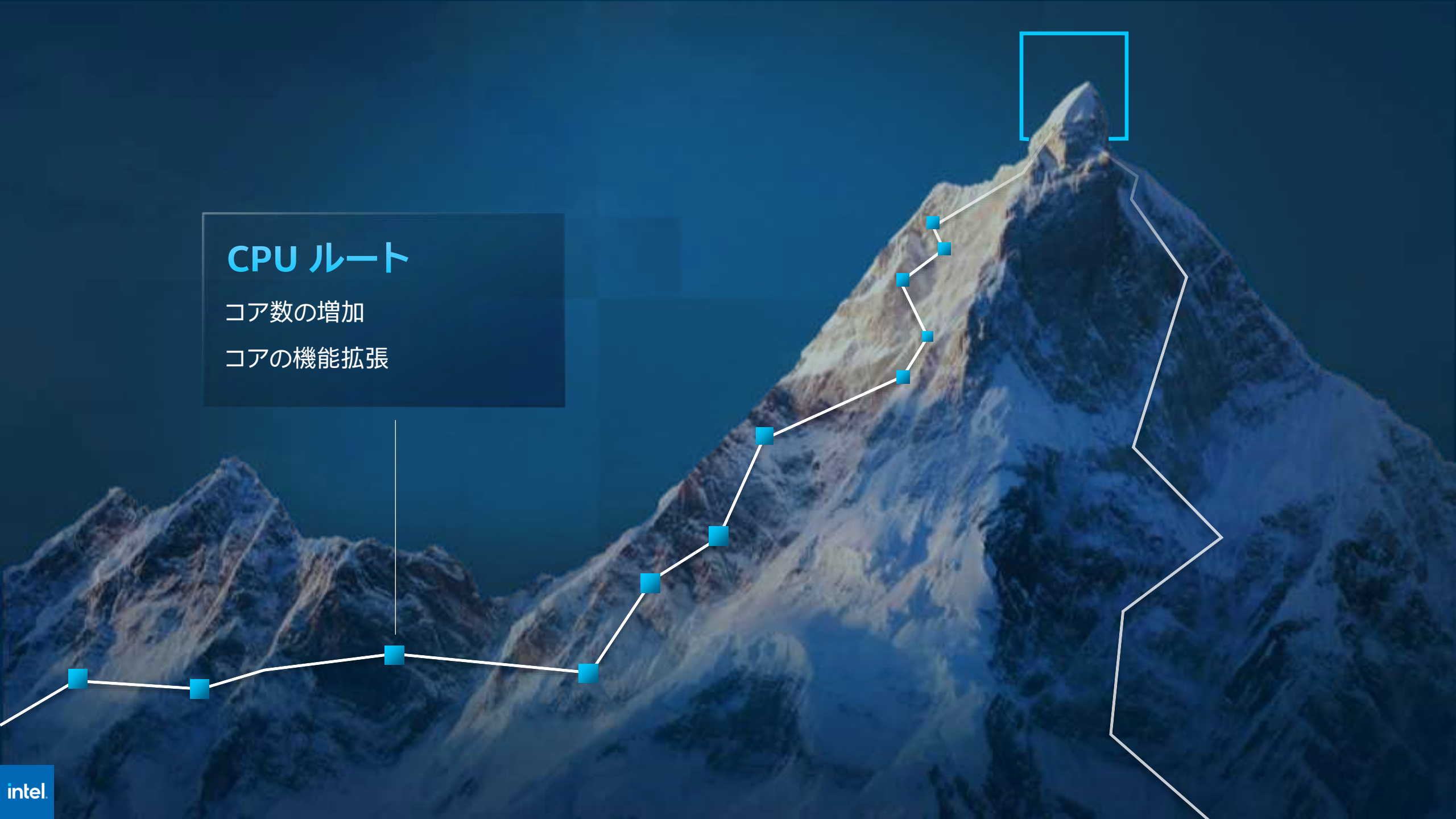


あらゆるシミュレーション・ワークロードに合わせた最適化

コンピューティングの急増
AI の出現
広帯域幅メモリーのニーズ
集積度、拡張性、サステナビリティー



アーキテクチャーごとに
頂上に達するルートは
複数

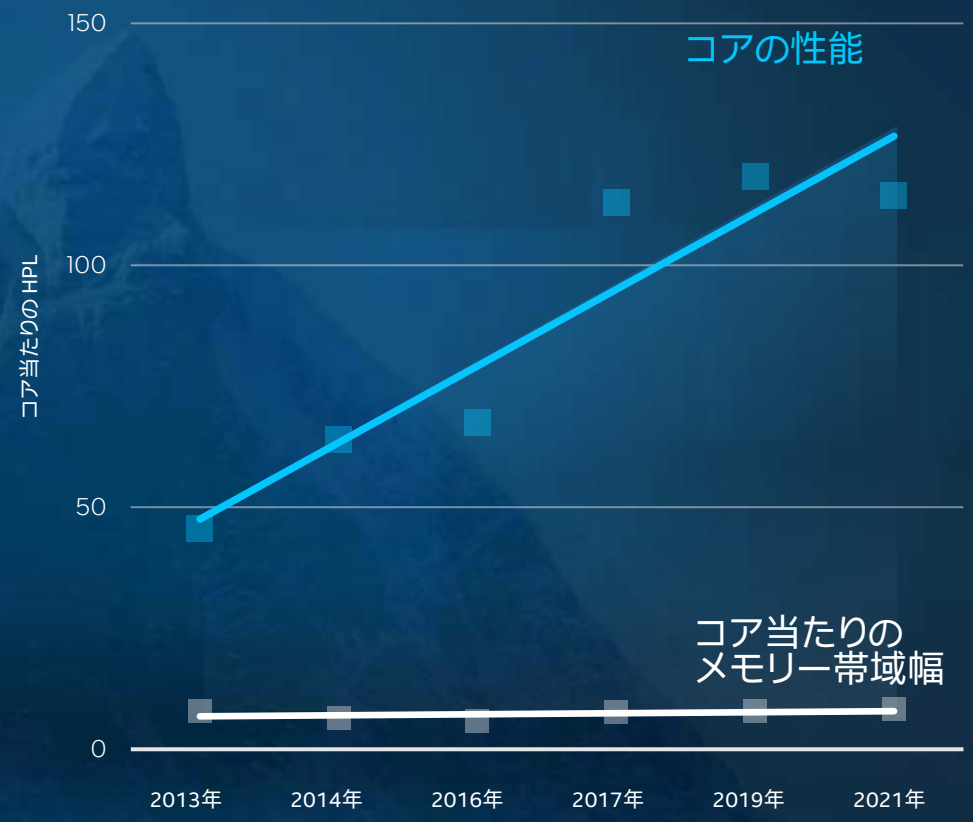

CPU ルート
コア数の増加
コアの機能拡張


GPU ルート
データ並列カーネルのオフロード
広帯域幅メモリー


CPU
スカラー
障壁
GPU
ベクトル
障壁
メモリー帯域幅

この 10 年間、HPC や AI のコンピューティング性能に比べてメモリー帯域幅の成長ペースは遅れており、さまざまなワークロードを高速化するうえでのボトルネックとなっています。

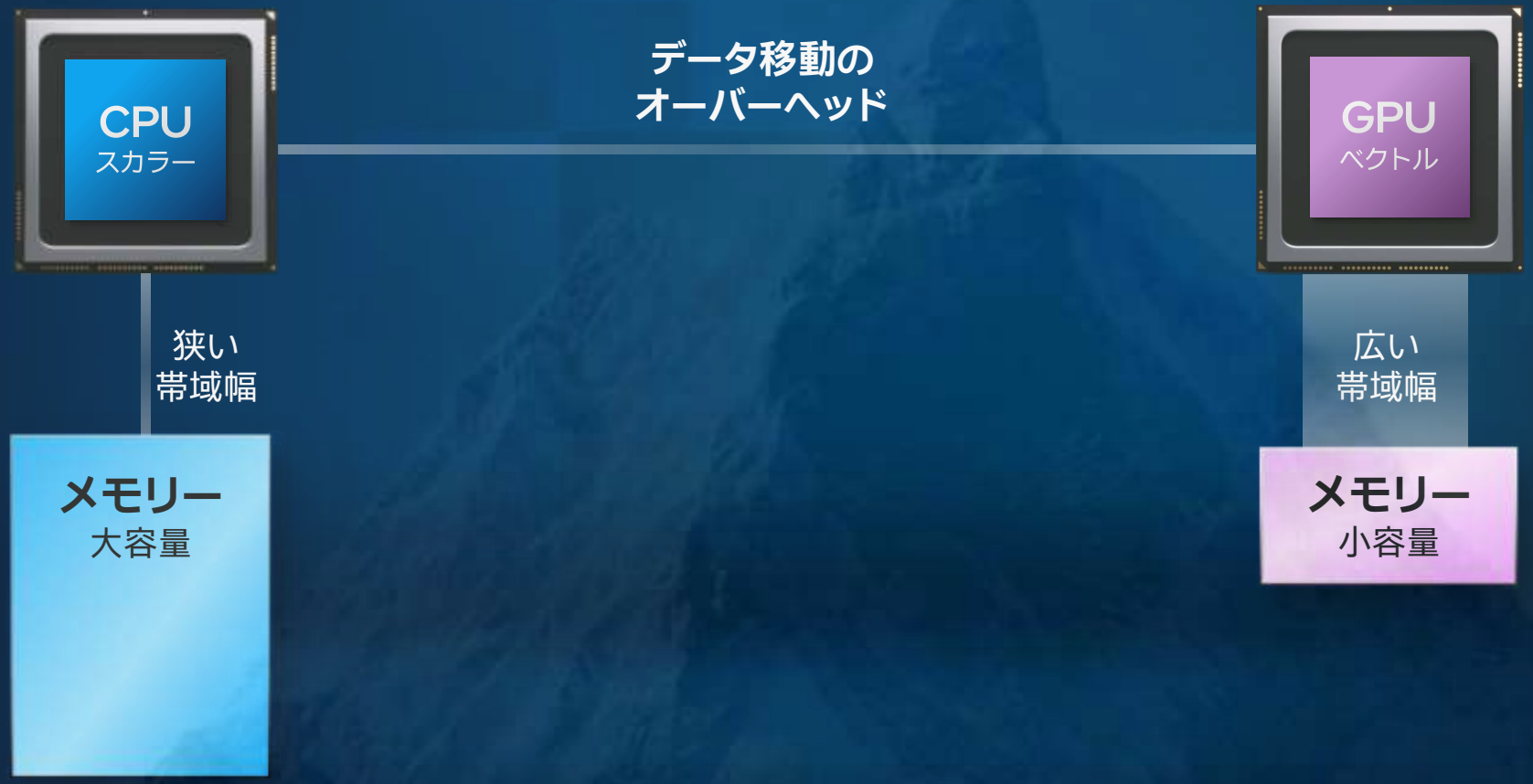
Hyperion Research CEO
Earl Joseph 氏



例示のみを目的としています。社内データに基づきます。


CPU
スカラー
障壁
GPU
ベクトル
障壁
メモリー容量
メモリー帯域幅
移植 & リファクタリング

HPC システムの状態


CPU
スカラー
データ移動の
オーバーヘッド
GPU
ベクトル
狭い
帯域幅
広い
帯域幅
メモリー
大容量
メモリー
小容量


どの HPC
ワークロードも
後れを取らず...

登場

可能性を最大限に引き出す

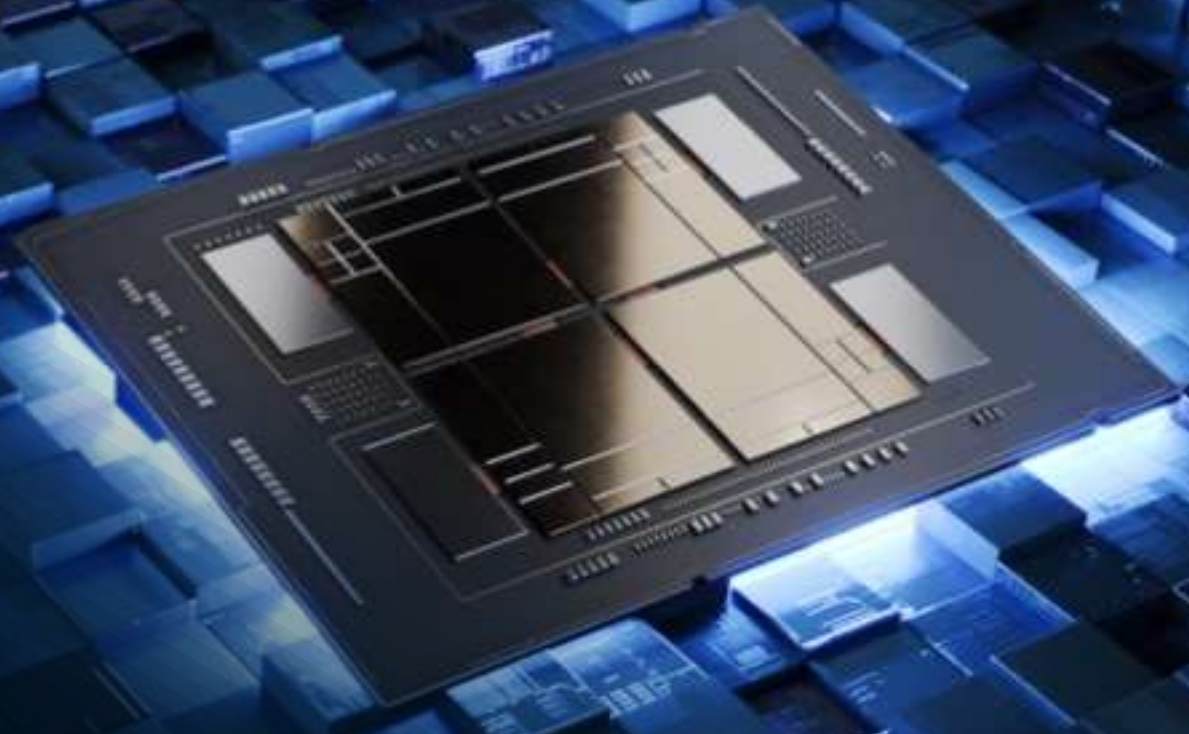
帯域幅を最大化
インパクトを最大化

intel

intel XEON MAX SERIES

インテル® Xeon® CPU マックス・シリーズ

開発コード名: Sapphire Rapids HBM



最大
56 P-cores

構造ベース:
エンベデッド・マルチダイ・インターコネクト・ブリッジ (EMIB)

350W
TDP

CXL Compute Express Link
1.1

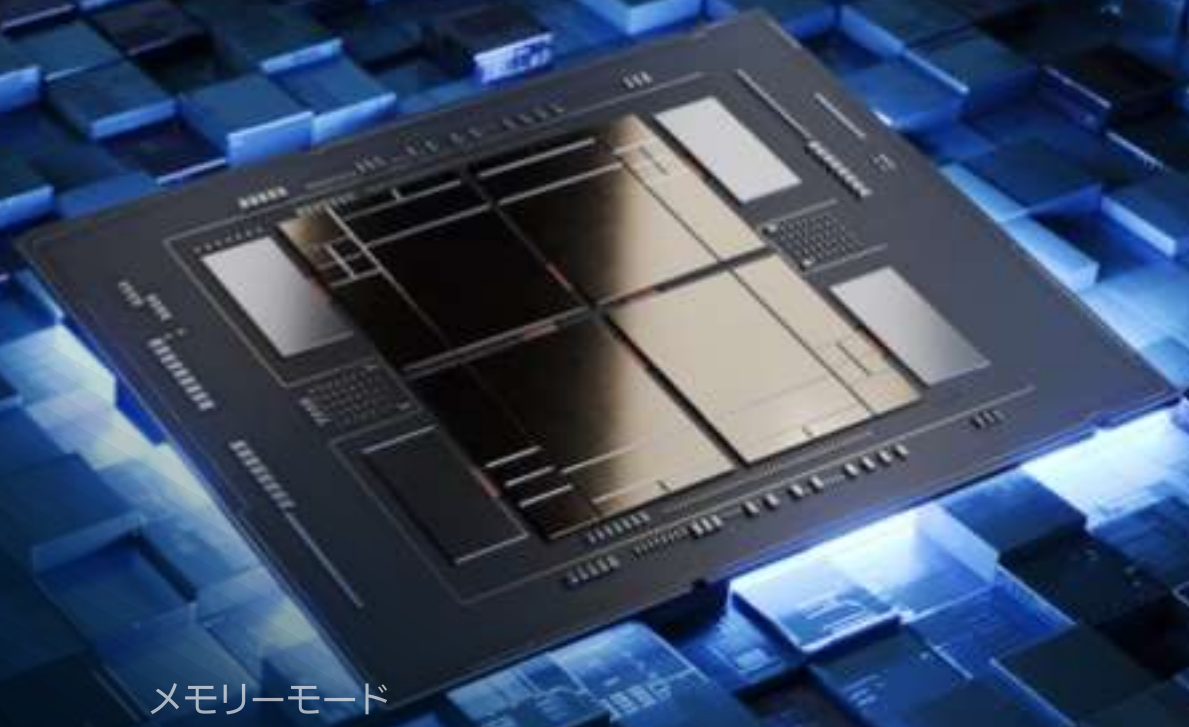
20
インテルの内蔵アクセラレーター・エンジン

intel

intel XEON MAX SERIES

初めて、かつ唯一の HBM 内蔵 x86 CPU

ニーズに適したメモリー構成を選択可能



64GB
HBM2e
16GB が 4 スタック

最大 1TB/s
メモリー
帯域幅

1GB 以上
コア当たりの
HBM

メモリーモード

HBM のみ
HBM からブート可能
コード変更不要

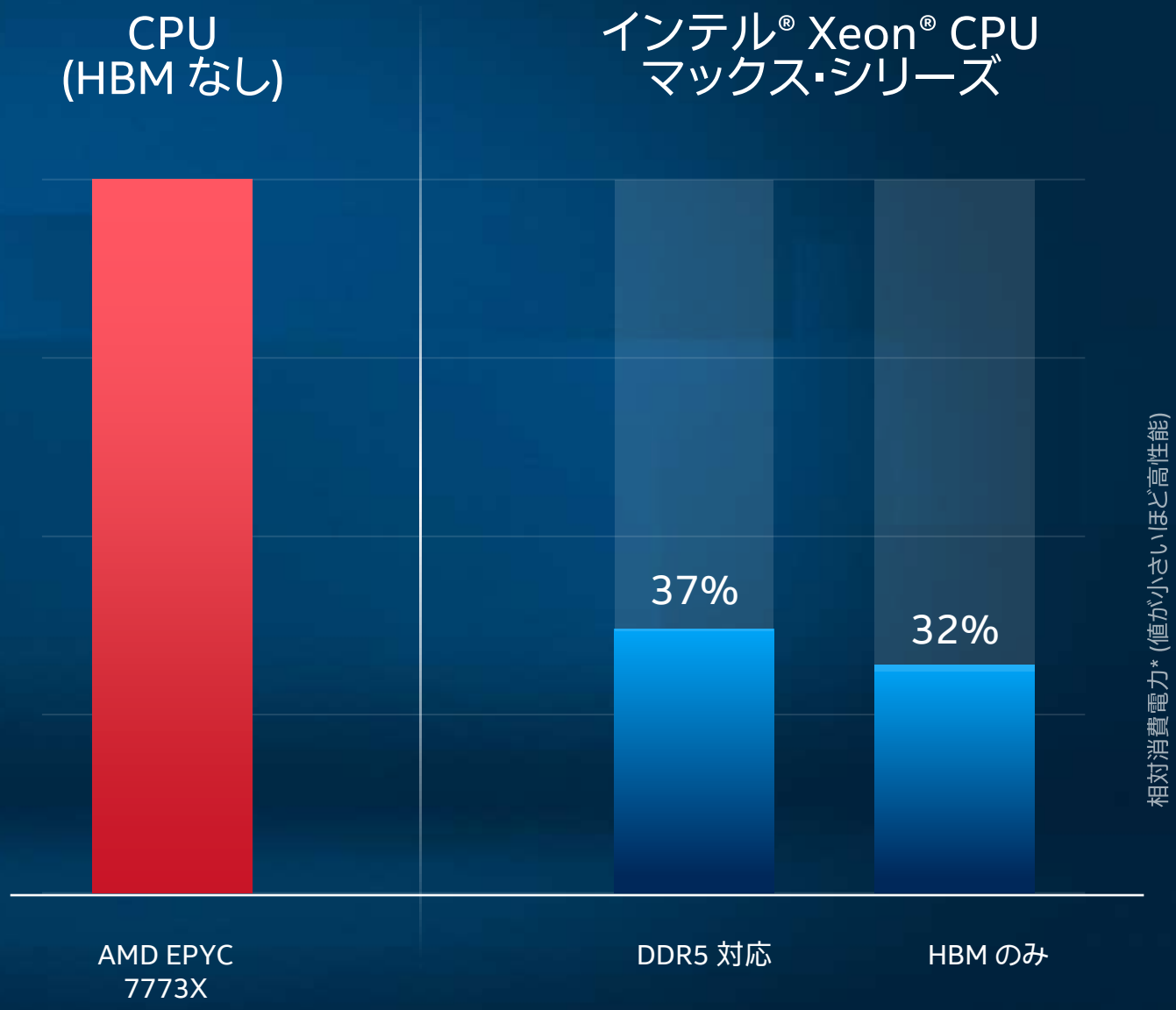
HBM フラット
2 つのメモリー領域
SW 最適化が必要

HBM キャッシング
DDR キャッシュとしての HBM
コード変更不要

intel

消費電力当たり性能の向上

Milan-X クラスターよりも
消費電力を 68% 抑えながら
同等の HPCG 性能



ワークロードと構成の詳細については、http://intel.com/PerformanceIndex/ (英語) の SuperComputing 22 ページを参照してください。結果は異なる場合があります。

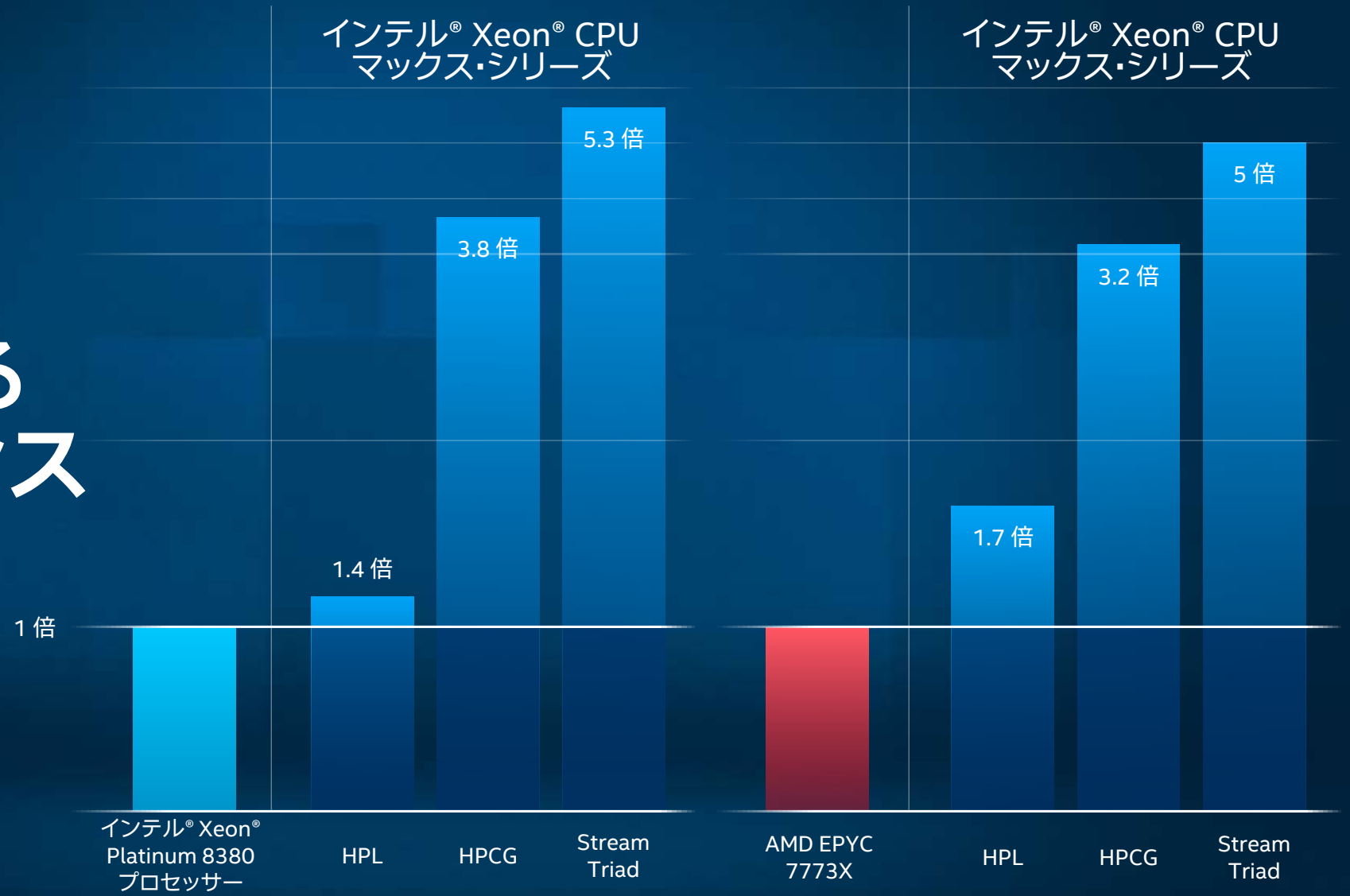
「京都大学の最新 HPC システム Camphor3 の設計では、インテルをはじめ、最先端技術を誇るさまざまな企業と連携しました。Camphor3 は計算科学者の研究の加速に役立っています。

京都大学では、パフォーマンスに対する広帯域幅メモリーの重要性を認識し、インテル® Xeon® CPU マックス・シリーズの導入を決定しました。広帯域幅メモリーを利用することで、研究者が科学や人間に関する発見の可能性を最大限に引き出せるようになると期待しています」

准教授、博士
深沢 圭一郎 氏

5 倍を超えるパフォーマンスの向上

2 ソケットのインテル® Xeon® CPU マックス・シリーズと
2 ソケットの AMD EPYC 7773X、
2 ソケットの第 3 世代インテル® Xeon® Platinum 8380 プロセッサーの比較



ワークロードと構成の詳細については、http://intel.com/PerformanceIndex/ (英語) の SuperComputing 22 ページを参照してください。結果は異なる場合があります。

intel XEON MAX SERIES

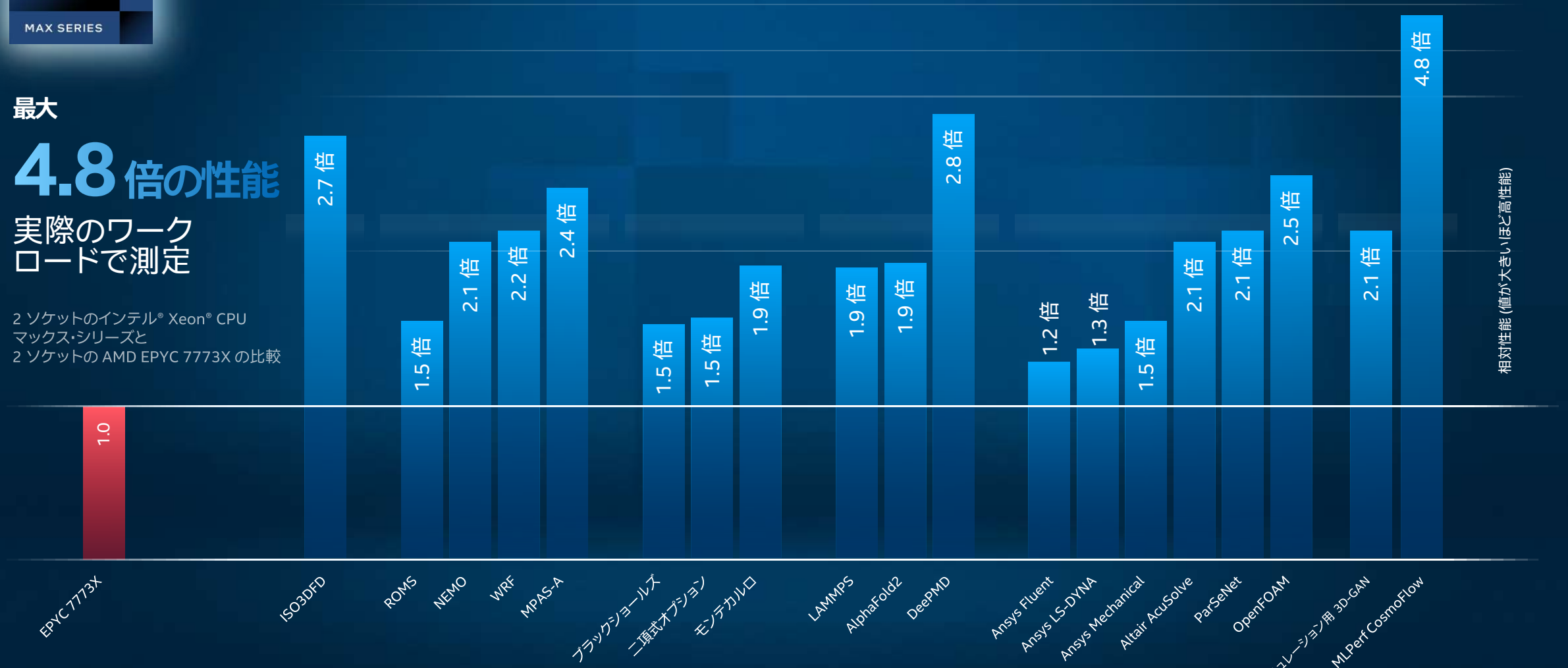
エネルギー　地球システムモデリング　金融サービス　ライフサイエンス / 物質科学　製造　物理

最大

4.8 倍の性能

実際のワークロードで測定

2 ソケットのインテル® Xeon® CPU マックス・シリーズと
2 ソケットの AMD EPYC 7773X の比較



ワークロードと構成については、補足資料を参照してください。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 の承認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。
MLPerf HPC-AI v 0.7 Cosmoflow 学習処理ベンチマークでの未検証のパフォーマンス向上。最適化済みの Tensorflow2.6 (FP32、FP16) および 2.9.1 (FP32、BP16) を使用。この結果の MLCommons Association による検証は未実施。未検証の結果は MLPerf レビューを経ておらず、検証結果に対する MLPerf の仕様とは一致しない測定方法やワークロード実装が使用されている可能性があります。MLPerf の名称およびロゴは、アメリカ合衆国およびその他の国における MLCommons Association の商標です。無断での引用、転載を禁じます。未承認の使用は固く禁じられています。詳細については、http://www.mlcommons.org を参照してください。

ワークロードと構成の詳細については、http://intel.com/PerformanceIndex/ (英語) の SuperComputing 22 ページを参照してください。結果は異なる場合があります。

intel

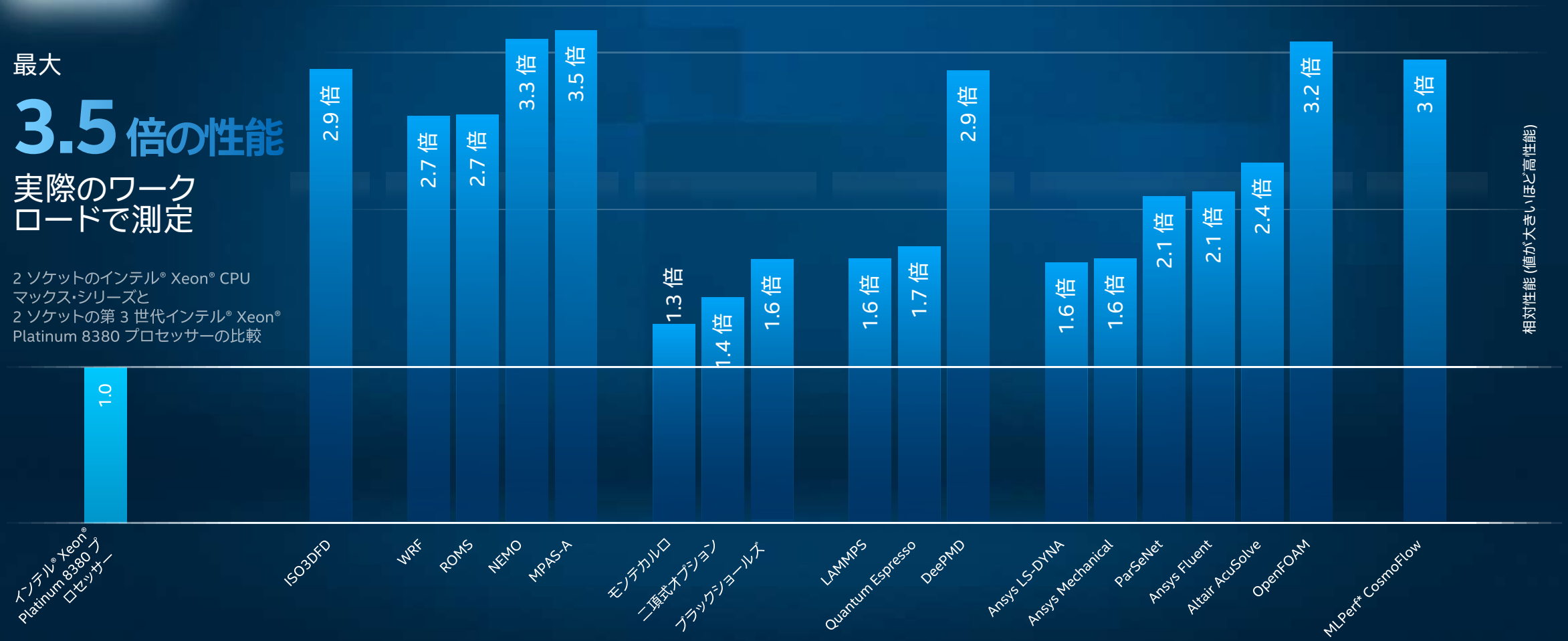
intel XEON MAX SERIES

最大

3.5 倍の性能

実際のワークロードで測定

2 ソケットのインテル® Xeon® CPU マックス・シリーズと
2 ソケットの第 3 世代インテル® Xeon® Platinum 8380 プロセッサーの比較



ワークロードと構成については、補足資料を参照してください。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。
MLPerf HPC-AI v 0.7 Cosmoflow 学習処理ベンチマークでの未検証のパフォーマンス向上。最適化済みの Tensorflow2.6 (FP32、FP16) および 2.9.1 (FP32、BP16) を使用。この結果の MLCommons Association による検証は未実施。未検証の結果は MLPerf レビューを経ておらず、検証結果に対する MLPerf の仕様とは一致しない測定方法やワークロード実装が使用されている可能性があります。MLPerf の名称およびロゴは、アメリカ合衆国およびその他の国における MLCommons Association の商標です。無断での引用、転載を禁じます。未承認の使用は固く禁じられています。詳細については、http://www.mlcommons.org を参照してください。

ワークロードと構成の詳細については、
http://intel.com/PerformanceIndex/ (英語) の
SuperComputing 22 ページを参照してください。
結果は異なる場合があります。

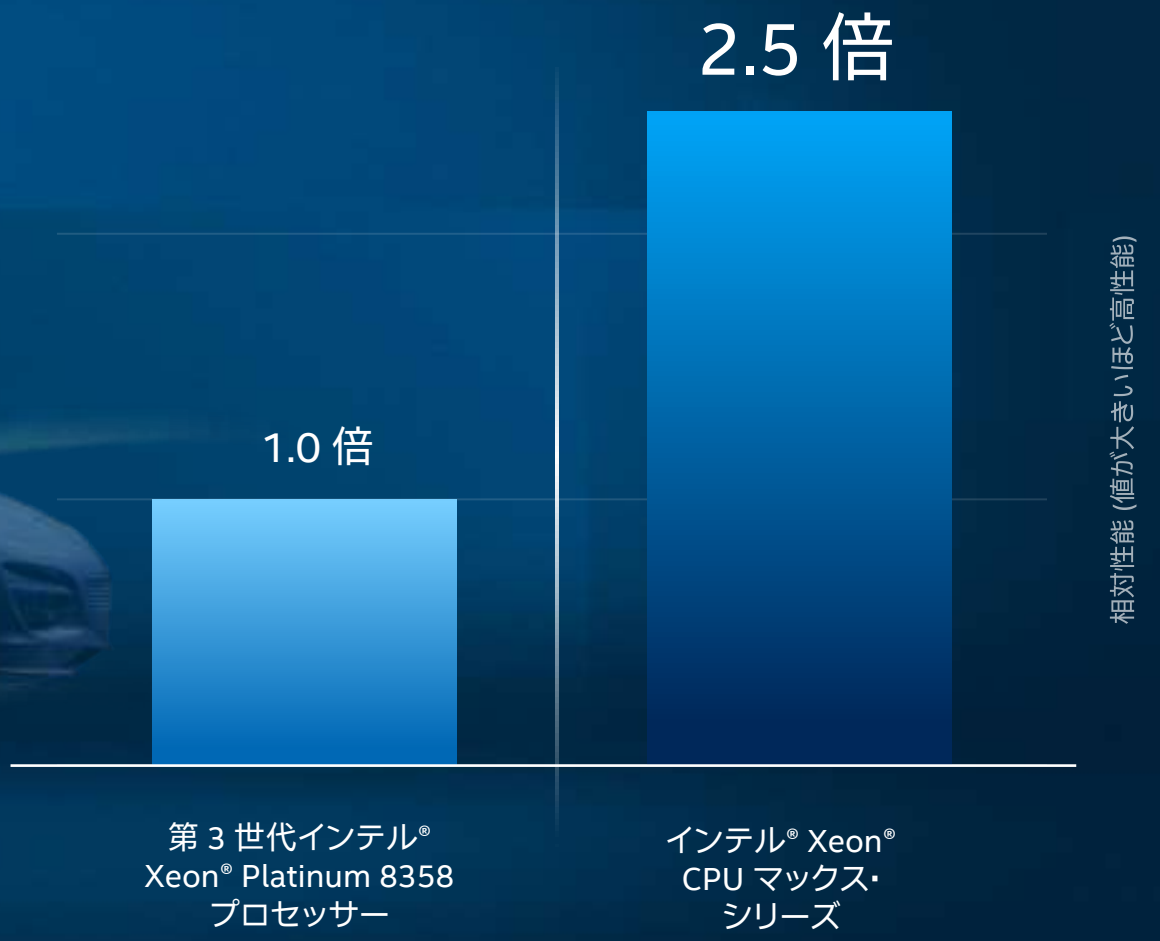
製造

数値流体力学

intel XEON MAX SERIES

単一ノードのパフォーマンス向上

ALTAIR AcuSolve* HQ

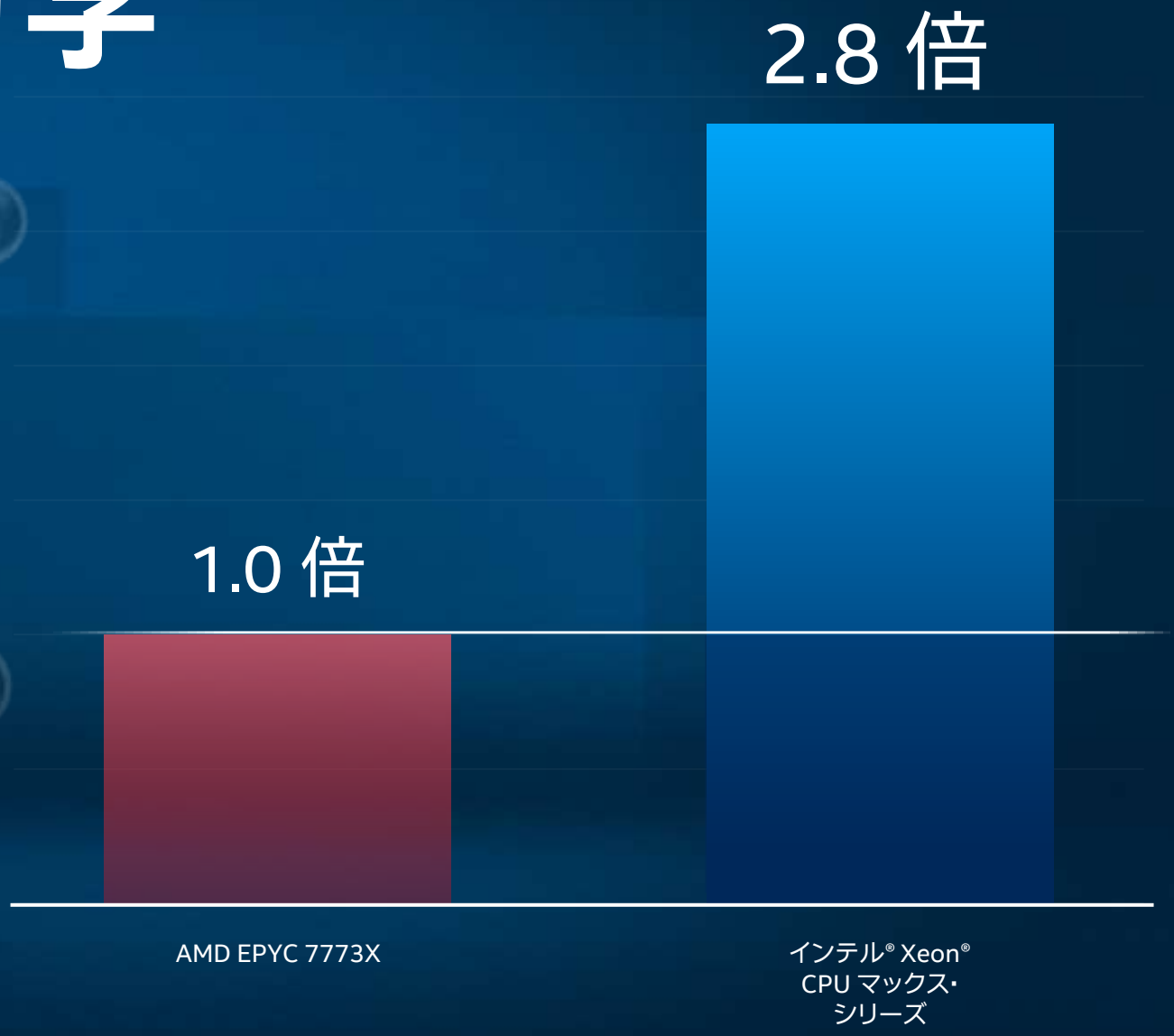


ワークロードと構成の詳細については、http://intel.com/PerformanceIndex/ (英語) の SuperComputing 22 ページを参照してください。結果は異なる場合があります。

分子動力学

分子動力学のモデル開発をスピードアップ

DeePMD - Kit



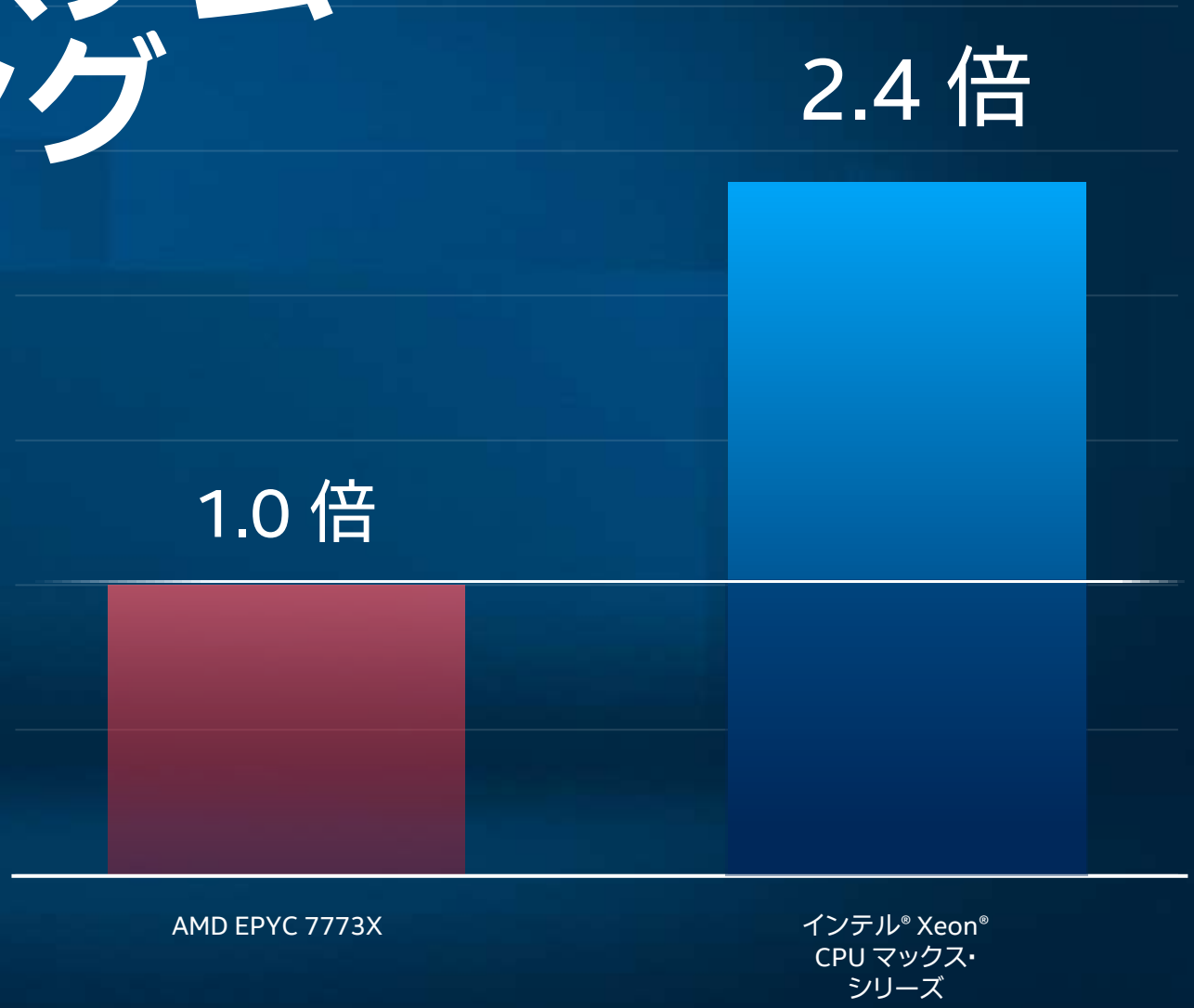
ワークロードと構成の詳細については、http://intel.com/PerformanceIndex/ (英語) の SuperComputing 22 ページを参照してください。結果は異なる場合があります。

地球システムモデリング

気象予報

地球システム・シミュレーションの高速化

MPAS-A (60-km)



ワークロードと構成の詳細については、http://intel.com/PerformanceIndex/ (英語) の SuperComputing 22 ページを参照してください。結果は異なる場合があります。

オープンで
フルスタックの
アプローチ

.....ベンダーを問わず、
マルチアーキテクチャーの
プログラミングを促進

ワークロード・
アプリケーション

ミドルウェア &
フレームワーク

oneAPI

仮想化

オペレーティング・
システム

レベルゼロ

intel

oneAPI

オープンな標準ベースの一体型プログラミング・モデル

intel

oneAPI

2023年の
新規
リリース

インテルのマックス・シリーズ製品に最適化したサポート

- SYCL と最新の Fortran 標準に対応しサポートを拡張
- OpenMP 5.1 対応のコンパイラーとツールのサポート
- TensorFlow と PyTorch への最適化
- CUDA から SYCL へのコード移行機能の強化

intel


CPU
スカラー
GPU
ベクトル
障壁
intel
XEON
MAX SERIES
メモリー容量
メモリー帯域幅
移植 & リファクタリング
intel

intel
DATA CENTER
GPU
MAX SERIES

インテル®
データセンター GPU
マックス・シリーズ

開発コード名: Ponte Vecchio

構造ベース:
EMIB と
Foveros

最大
128
Xe HPC
コア数

52TF
FP64
最大スループット

16

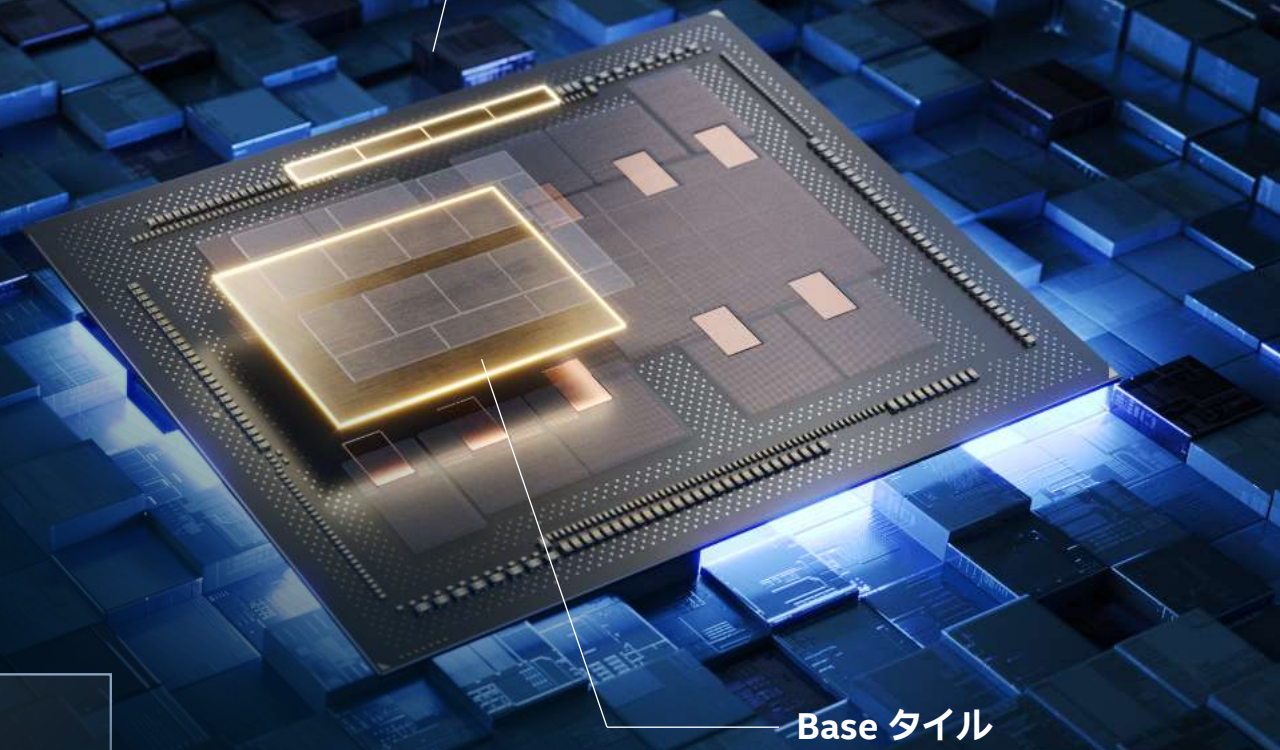
GPU 間通信用の
Xe Link 数

最高レベル
1 ソケット内の
コンピューティング密度

intel

intel
DATA CENTER
GPU
MAX SERIES

帯域幅
容量
メモリーを最大化



最大
128GB
HBM2e

最大
408MB
Rambo L2
キャッシュ

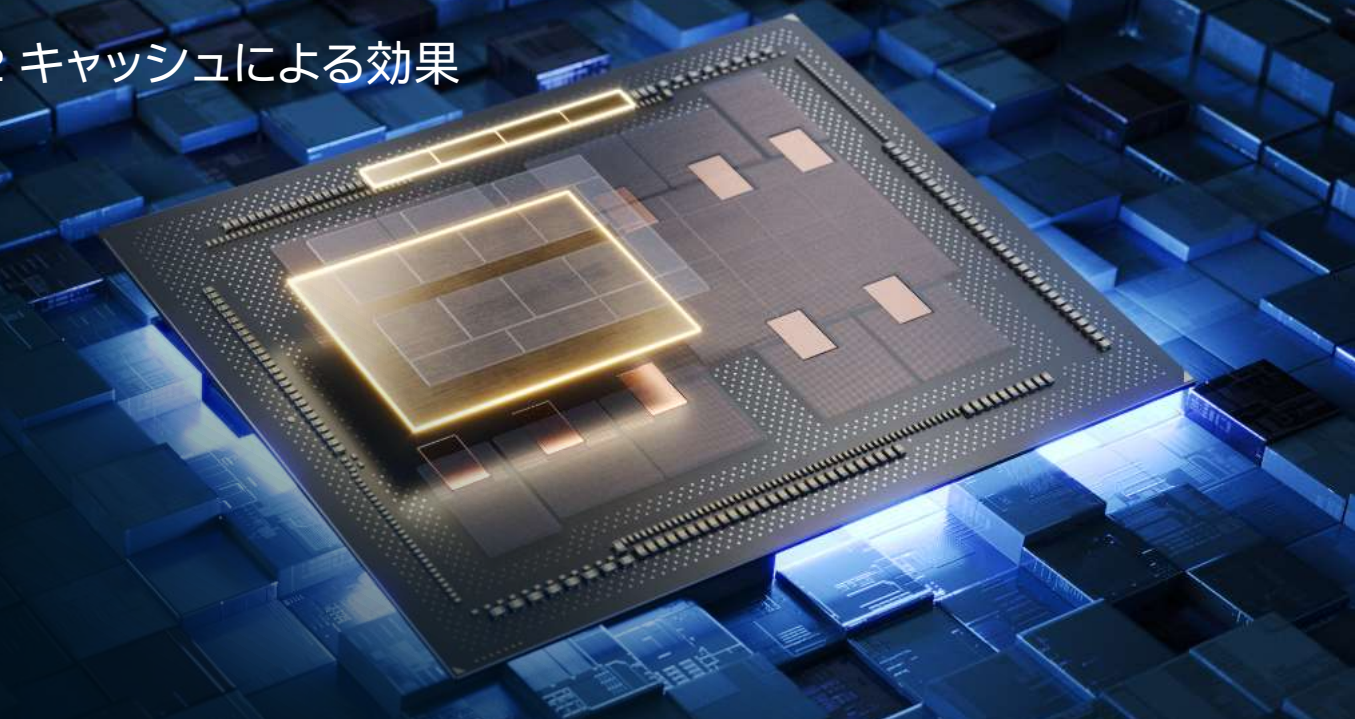
最大
64MB
L1 キャッシュ

intel

intel DATA CENTER GPU MAX SERIES

最大 2 倍のパフォーマンス

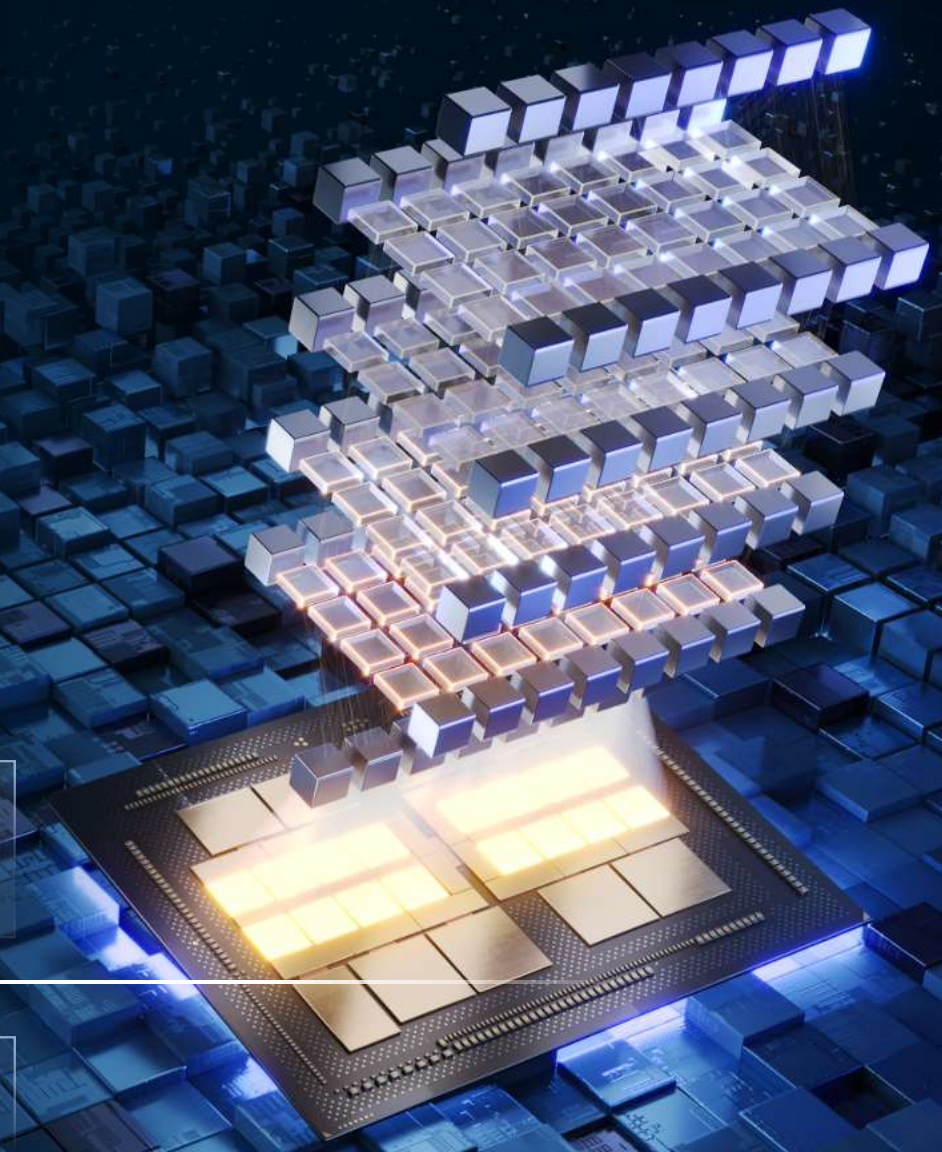
4Kx4K 2D-FFT DP での大規模 L2 キャッシュによる効果



この感度の測定では、インテル® GPU マックス・シリーズの 408MB L2 キャッシュを 80MB と 32MB に低く構成しています。
ワークロードと構成の詳細については、http://intel.com/PerformanceIndex/ (英語) の SuperComputing 22 ページを参照してください。結果は異なる場合があります。

インテル® Xe マトリクス・エクステンション

専用の内蔵 AI 機能
ほとんどの AI データタイプを高速化



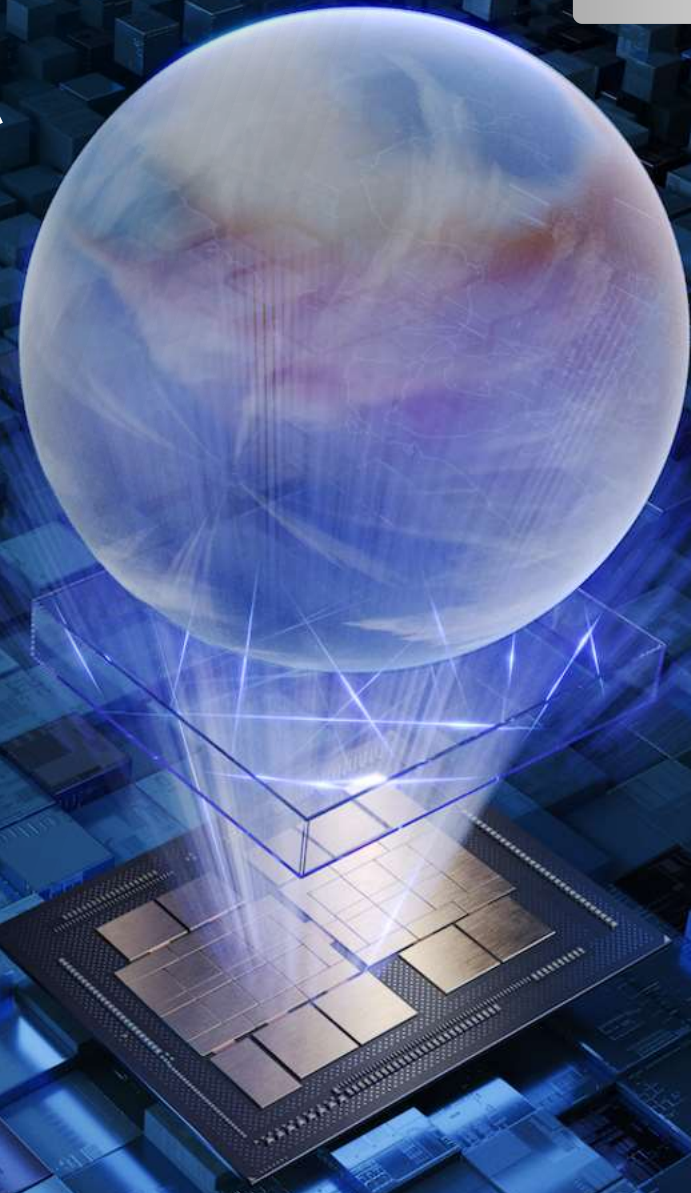
ベクトル OPS/CLK/Xe-core	**256** FP32	**256** FP64	**512** FP16
マトリクス OPS/CLK/Xe-core	**2048** TF32	**4096** FP16/BF16	**8192** INT8

intel

新発表

intel DATA CENTER GPU MAX SERIES

レイ・トレーシング



最大
128
レイ・トレーシング・ユニット (RTU) 数

6
ボックス・インターセクション数 / サイクル / RTU

1
トライアングル・インターセクション数 / サイクル / RTU

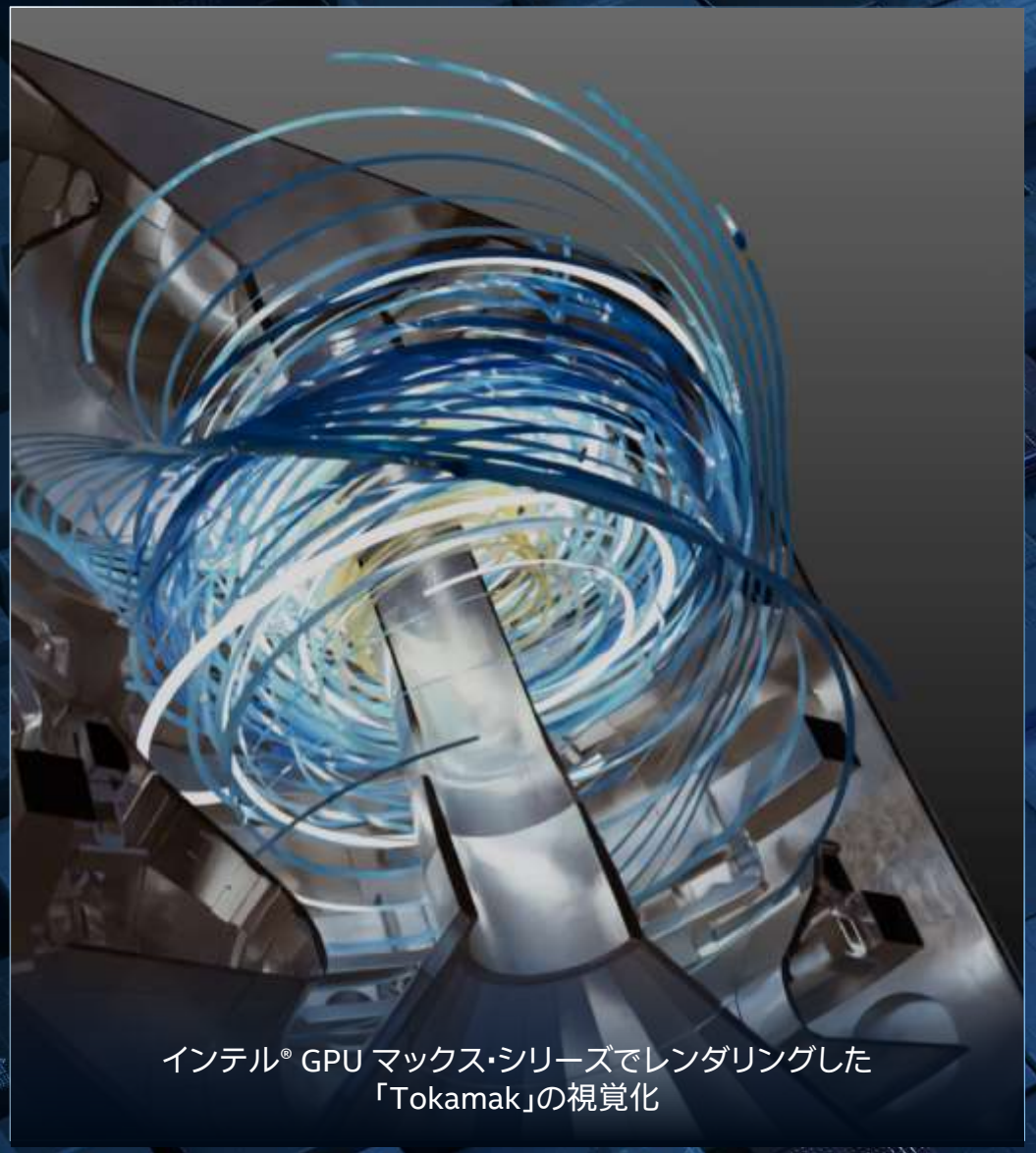
インテル® OSPRay
2023年にリリース

インテル® Embree 4
2023年にリリース

大容量 VRAM
大規模な HPC、M&E シーンに有効

intel

インテル® GPU マックス・シリーズでレンダリングした
「Moana Island」のシーン



インテル® GPU マックス・シリーズでレンダリングした
「Tokamak」の視覚化

HPC アプリをすぐに実行可能

インテル® Xeon® CPU マックス・シリーズに実装

拡大を続ける多様なアプリケーション

インテル® データセンター GPU マックス・シリーズに実装

ベンチマーク
- xGEMMs
- GUPS
- HPL
- SPEChpc
- STREAM
- MLBench
- SPECACCEL
- SpMV

ライブラリー & フレームワーク
- ELPA
- PETSc
- HYPRE
- HeFFTe
- Ginkgo

石油 & ガス
- ISO3DFD
- SPECFEM3D GLOBE

パフォーマンス移植性
- AMReX
- Kokkos
- RAJA
- OCCA
- YAKL

物理 & 製造
- MFIX-Exa
- HOTQCD
- Chroma
- MILC
- NekRS
- Shift
- PHASTA
- NAQMD
- HACC
- Gene
- Gem
- miniFE
- NEKBONE
- OpenMC
- CloverLeaf
- BookLeaf
- TeaLeaf
- GRID QCD
- QUDA
- XGC

金融サービス
- 二項式オプション
- ブラックショールズ
- モンテカルロ
- STAC-A2
- Riskfuel クレジットオプション価格決定

地球システムモデル
- SW4
- E3SM
- SeisSol

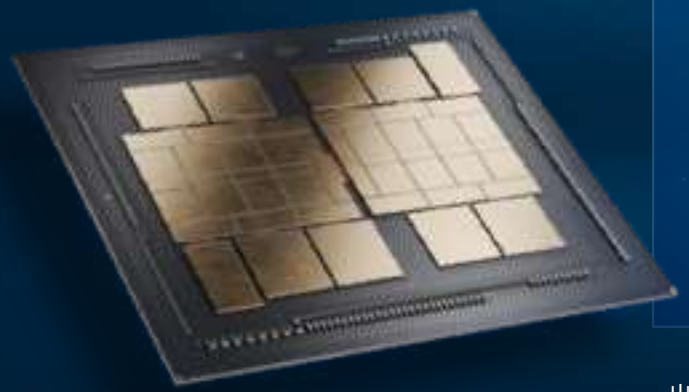
生物 & 化学科学
- Autodock
- CP2K
- LAMMPS
- GROMACS
- NAMD
- NWChem
- NWChemEx
- miniBUDE
- OpenMM
- ParSplice
- QMCPack
- Relion

AI & 新しいワークロード
- Candle ML
- 触媒計算科学 ML
- Atlas ML
- 材料科学 ML
- 核融合エネルギー ML
- コネクトミクス ML

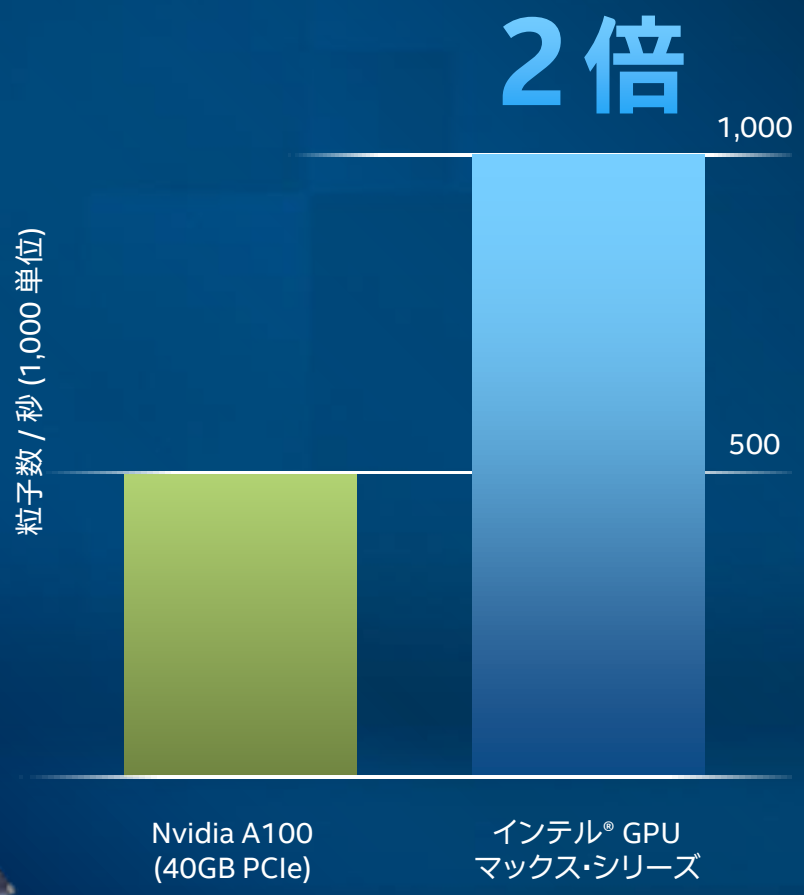
視覚化 & レンダリング
- Blender
- OSPRay
- Embree 4
- オープン・シェーディング・ライブラリー

ワークロードと構成の詳細については、http://intel.com/PerformanceIndex/ (英語) の SuperComputing 22 ページを参照してください。結果は異なる場合があります。

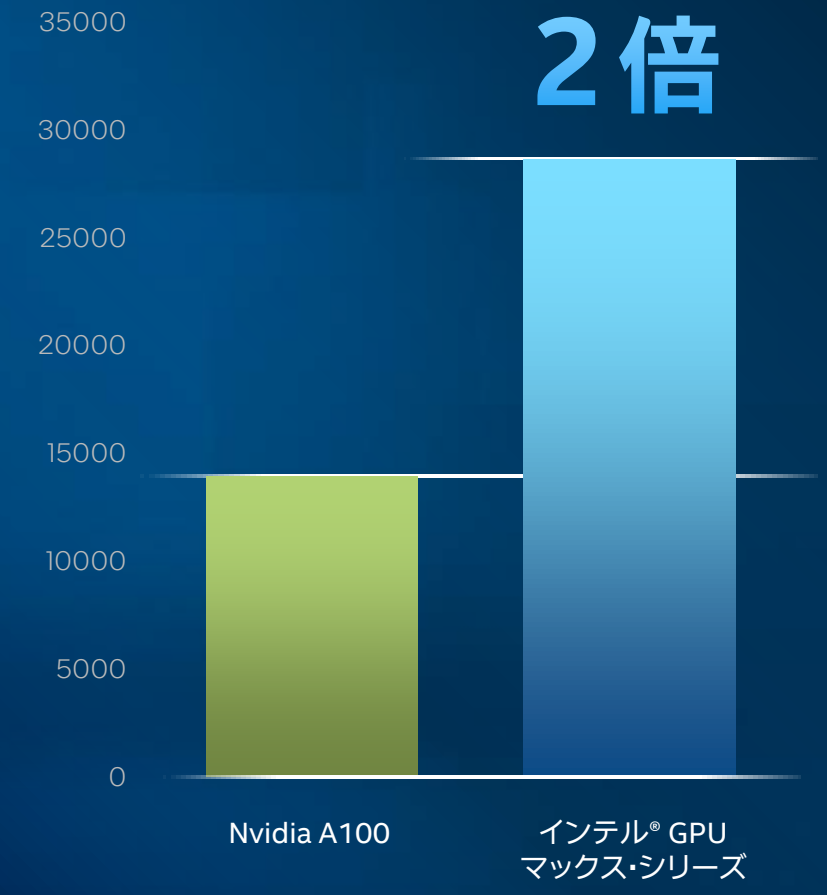
intel
DATA CENTER
GPU
MAX SERIES



OpenMC
モンテカルロ粒子輸送コードのエクサスケール演算



miniBUDE
Bristol University Docking Engine の基本的演算



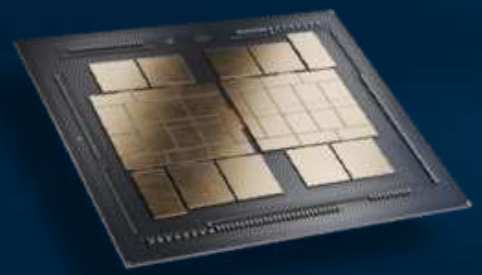
出荷前の測定値に基づきます。インテルは、サードパーティーのデータについて管理や監査を行っていません。ほかの情報も参考にしてデータの正確さを評価してください。

ワークロードと構成の詳細については、http://intel.com/PerformanceIndex/ (英語) の SuperComputing 22 ページを参照してください。結果は異なる場合があります。

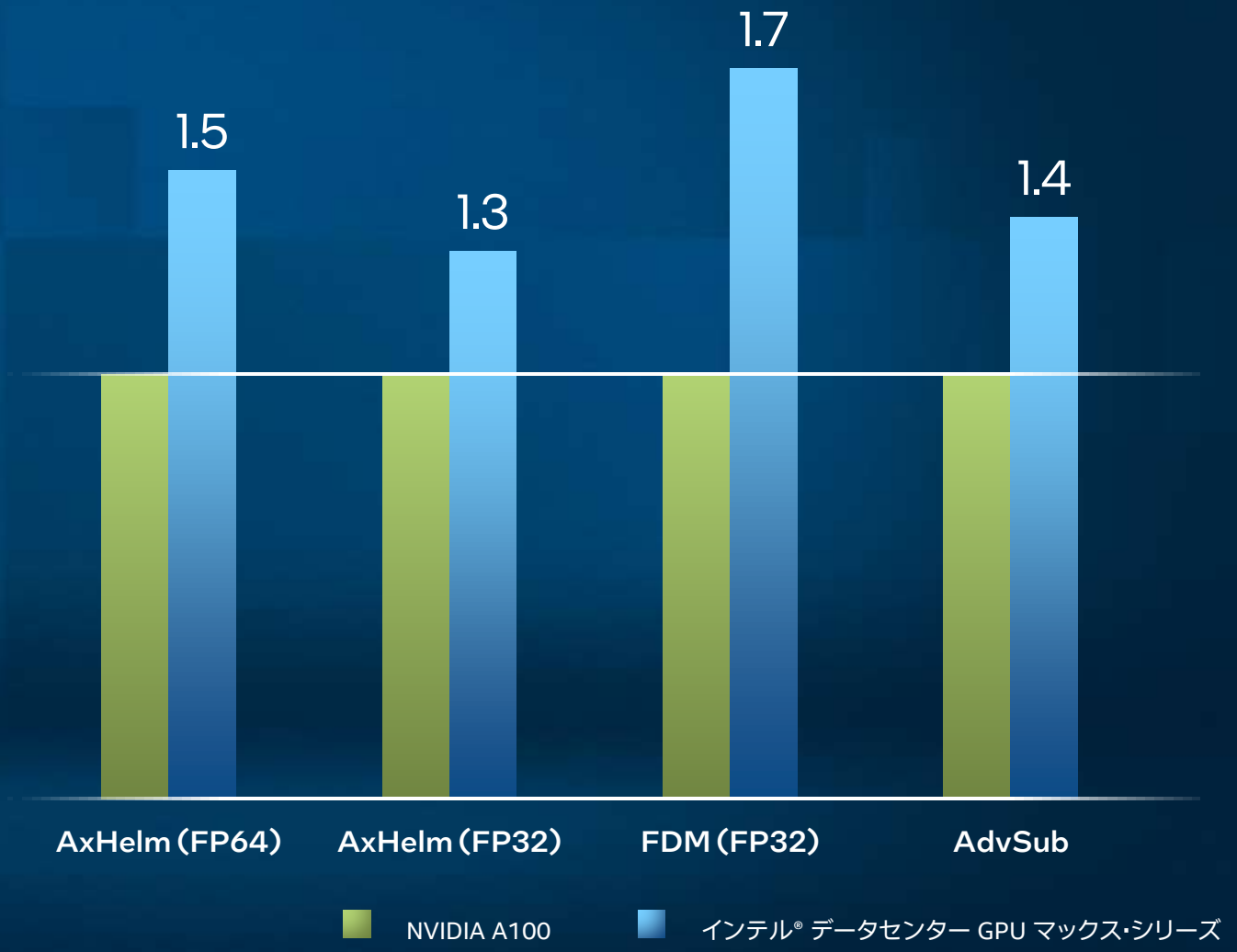
intel
DATA CENTER
GPU
MAX SERIES

仮想原子炉シミュレーションをスピードアップ

ExaSMR - NekRS パフォーマンス



1.5 倍の性能リード



NekRS ベンチマーク (テストサイズ: 8196) の相対性能。インテルの oneAPI SYCL 実装を使用。

ワークロードと構成の詳細については、http://intel.com/PerformanceIndex/ (英語) の SuperComputing 22 ページを参照してください。結果は異なる場合があります。

intel

新発表

intel DATA CENTER GPU MAX SERIES

インテル® データセンター GPU マックス 1100

ダブルスロット幅 PCIe カード

300W
TDP

56
Xe -core と RTU

48GB
HBM2e
メモリー

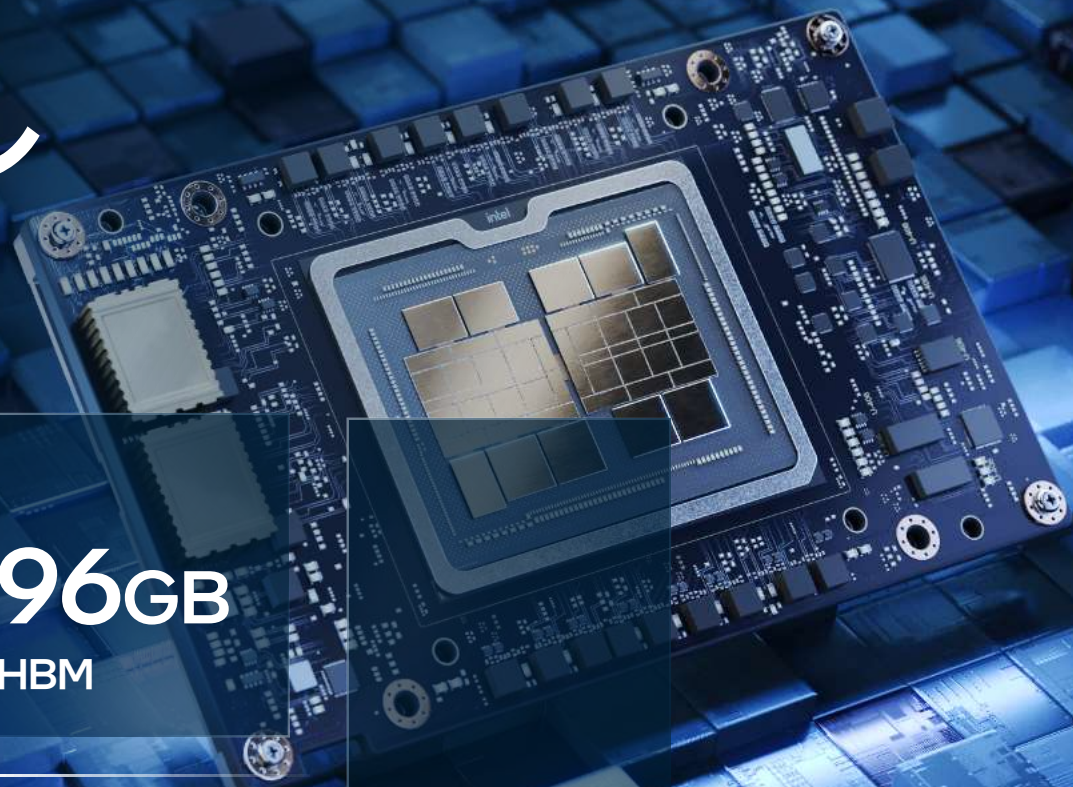
53G
SerDes
インテル® Xe Link
ブリッジ
最大 4 カード

intel

今後のリリース予定

intel
DATA CENTER
GPU
MAX SERIES

OAM モジュール



インテル® マックス・シリーズ 1350 GPU

- 450 W TDP
- 112 Xe-core
- 96GB HBM

インテル® マックス・シリーズ 1550 GPU

- 600 W TDP
- 128 Xe-core
- 128GB HBM

- 53G SerDes
- インテル® Xe Link ブリッジ
- 最大 8 OAM

intel

今後のリリース予定

intel
DATA CENTER
GPU
MAX SERIES

インテル® データセンター GPU マックス・シリーズのサブシステム

4 OAM モジュール

1,800W & 2,400W
TDP

最大
512GB

12.8TB/s
合計メモリー帯域幅

intel

最大限の選択肢を提供

intel
DATA CENTER
GPU
MAX SERIES

マックス・シリーズ GPU 1100

ダブルスロット幅 PCIe カード

300W

マックス・シリーズ GPU 1350

OAM フォームファクター

450W

マックス・シリーズ GPU 1550

OAM フォームファクター

600W

マックス・シリーズのサブシステム

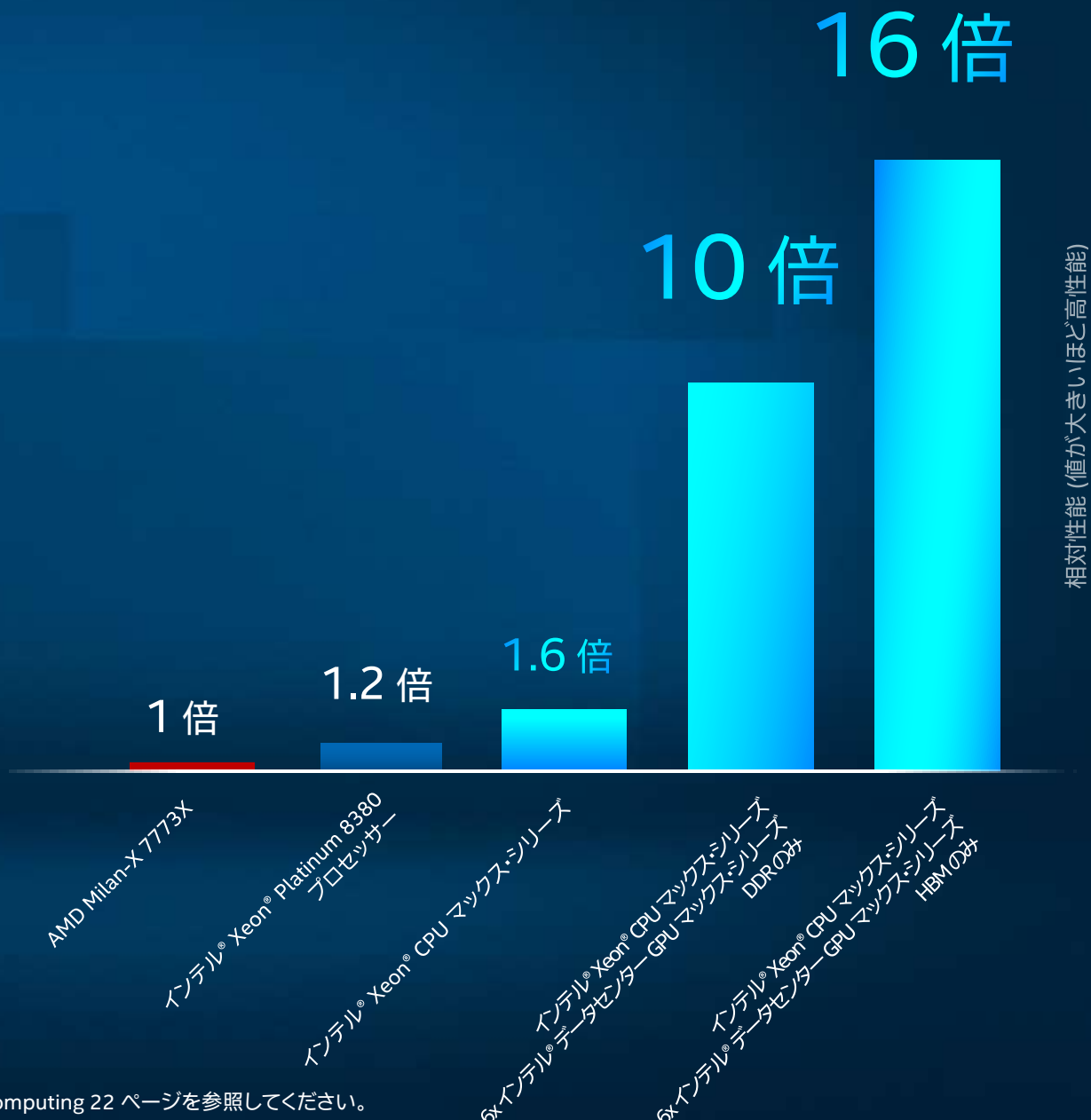
4 OAM

1,800W　2,400W

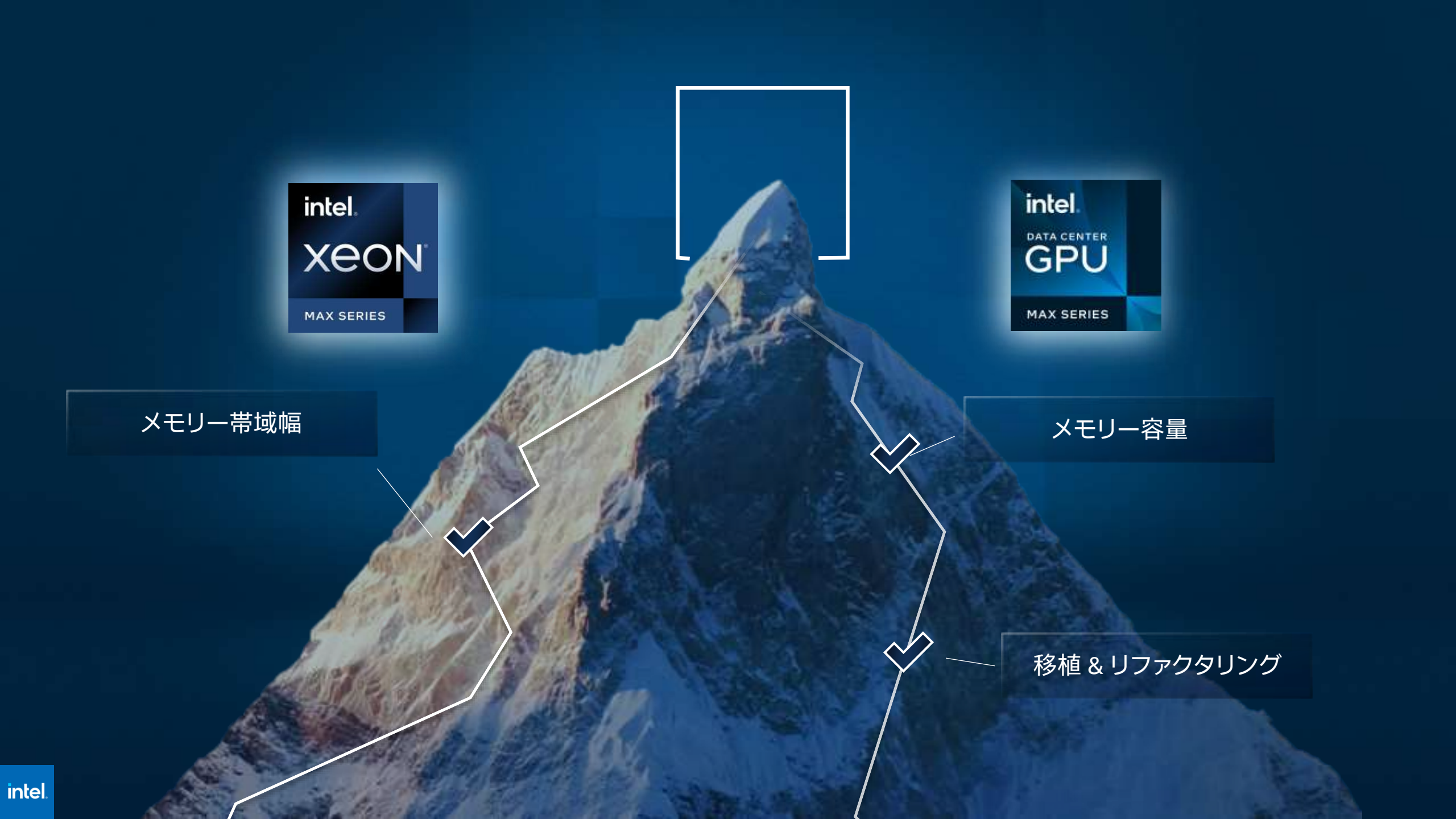
intel

ライフサイエンス / 物質科学

LAMMPS にコンピューティング密度と HBM を実装し、エクサスケールの新材料探索を高速化



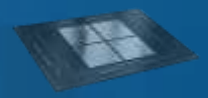
ワークロードと構成の詳細については、http://intel.com/PerformanceIndex/ (英語) の SuperComputing 22 ページを参照してください。
結果は異なる場合があります。


intel
XEON
MAX SERIES
intel
DATA CENTER
GPU
MAX SERIES
メモリー帯域幅
メモリー容量
移植 & リファクタリング
intel

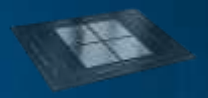
スーパーコンピューティング用シリコンのロードマップ

CPU
HPC

第 4 世代インテル® Xeon®
スケーラブル・プロセッサー



インテル® Xeon® プロセッサー
開発コード名: Emerald Rapids



インテル® Xeon® プロセッサー
開発コード名: Granite Rapids

インテル® Xeon® CPU
マックス・シリーズ

GPU
AI & HPC

インテル®
データセンター GPU
マックス・シリーズ

インテル®
データセンター GPU
マックス・シリーズ
開発コード名: Rialto Bridge

インテルの次の XPU、開発コード名:
Falcon Shores
柔軟性と拡張性に優れた
新しいタイルベースのアーキテクチャー

専用
DL

Habana®
Gaudi®
AI プロセッサー

Habana®
Gaudi®2
AI プロセッサー

Habana®
Gaudi®3
AI プロセッサー

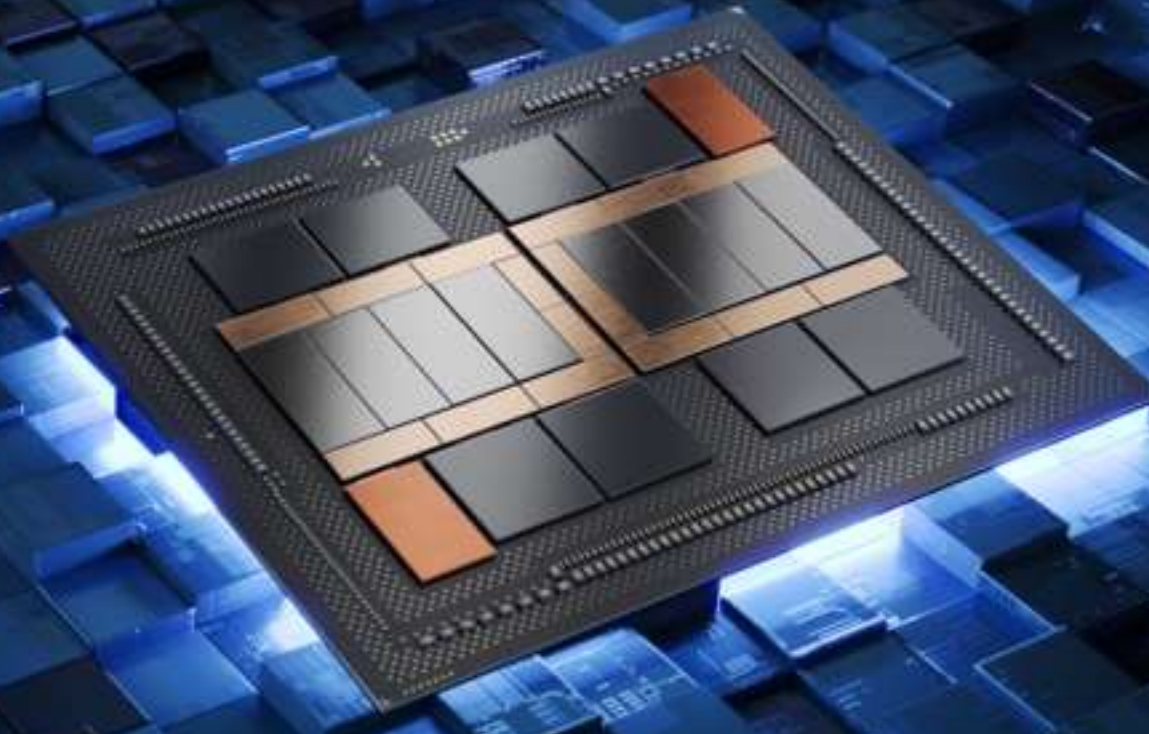
intel

intel DATA CENTER GPU MAX SERIES

まもなく登場するインテルのデータセンター向け GPU
開発コード名:

Rialto Bridge

DC サーバーの AI & HPC パフォーマンスを拡大

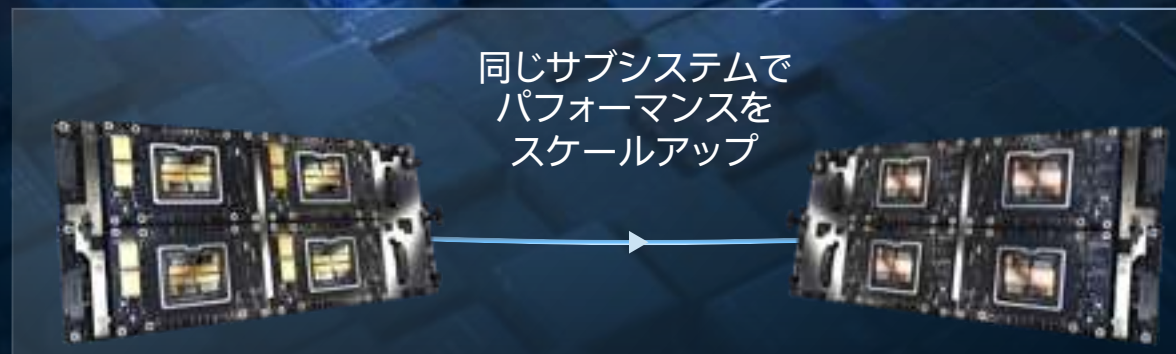


最大
800W
OAM 当たり、
液冷

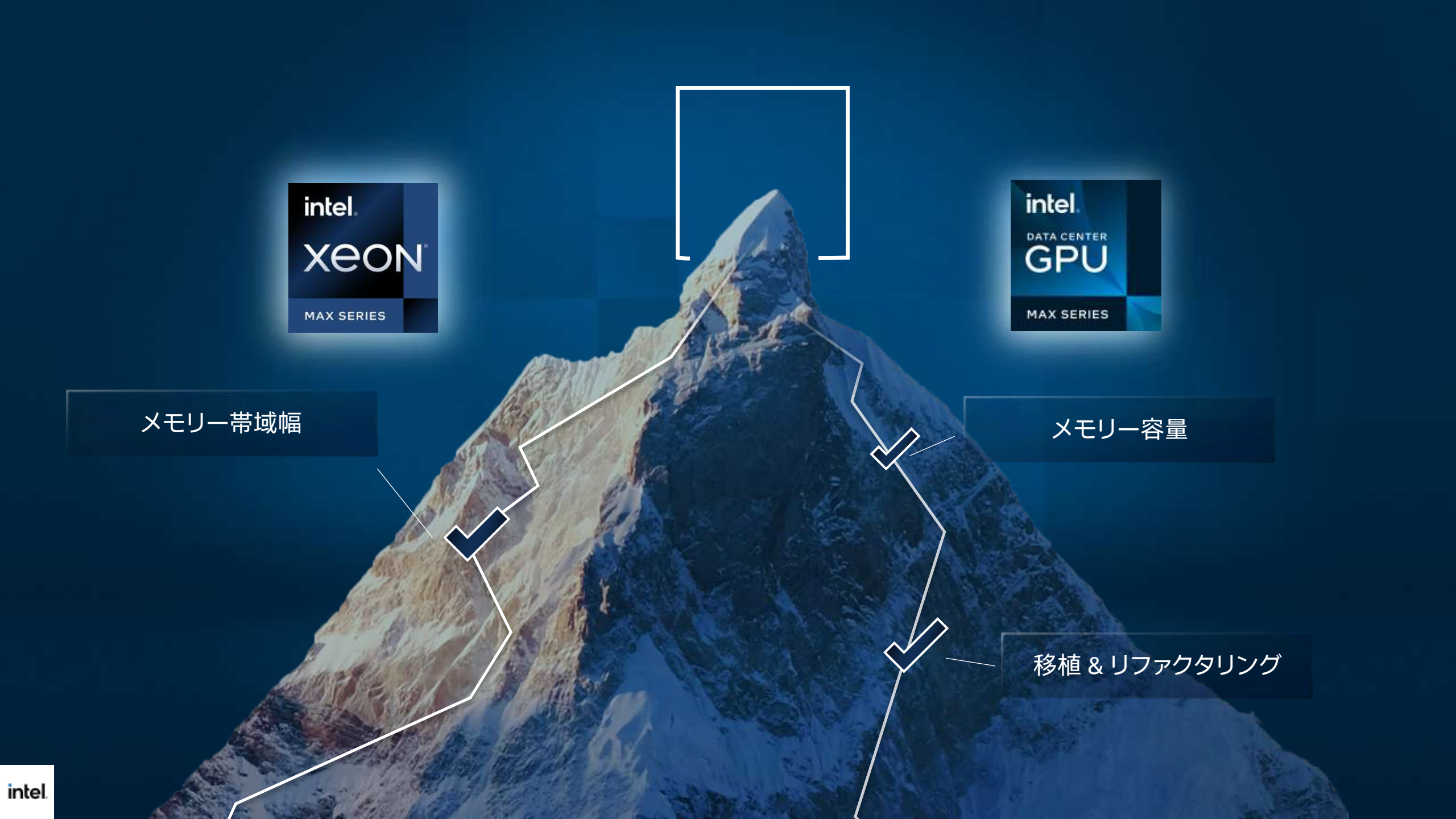
最大
160
HPC 対応
X^e -core の拡張

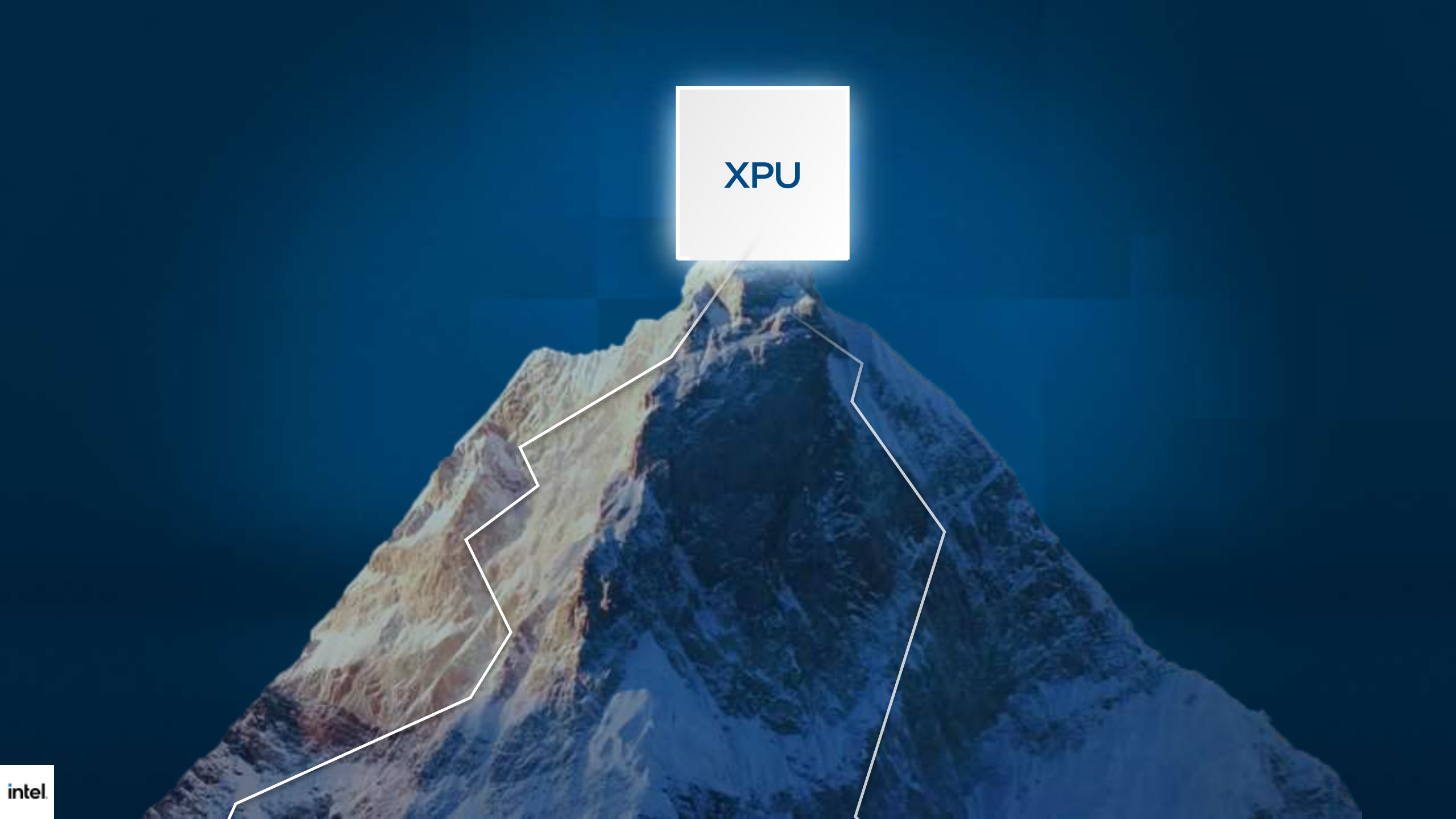
フォームファクター

- OAM
- PCIe
- サブシステム

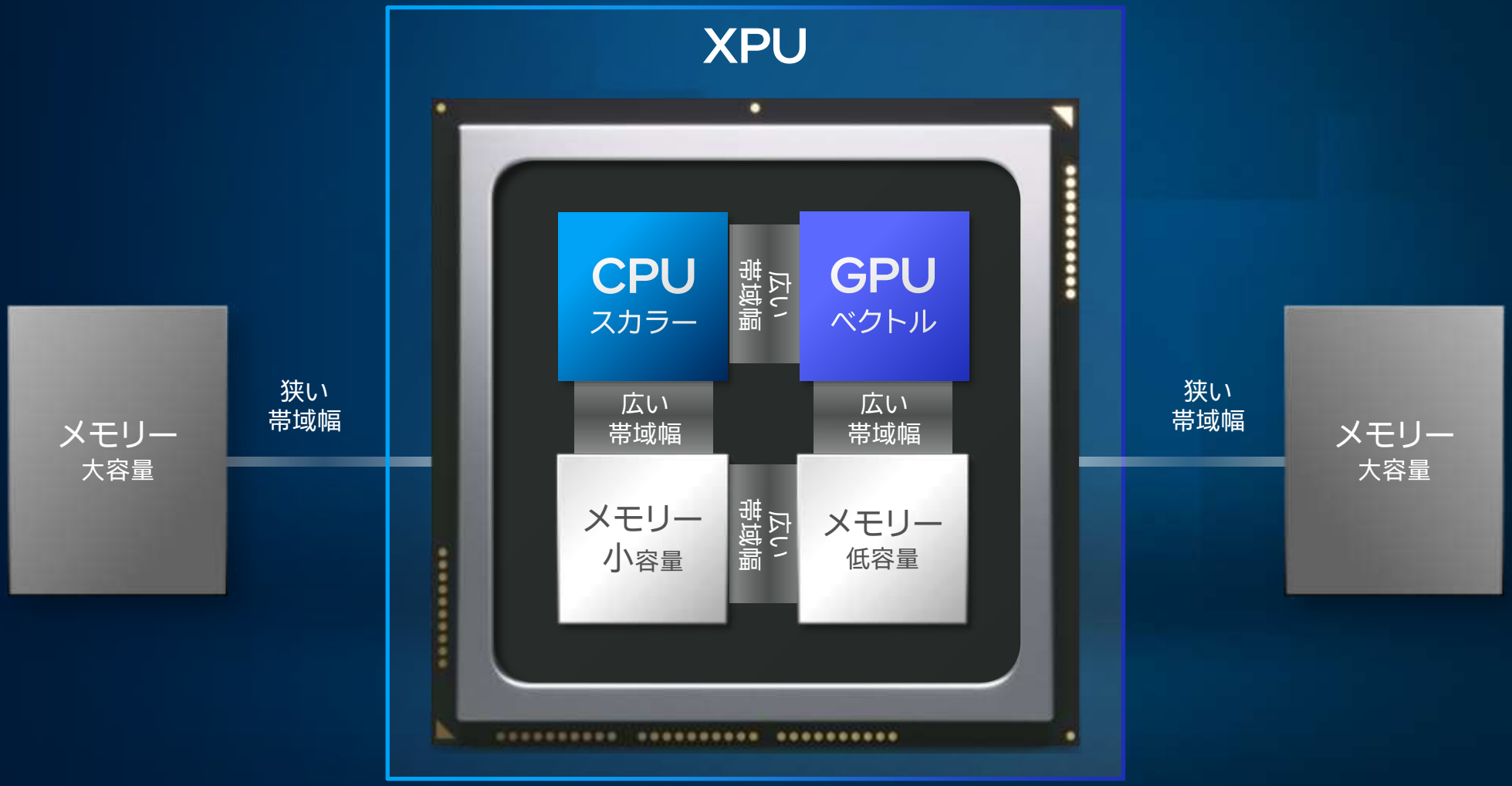


intel


intel
XEON
MAX SERIES
intel
DATA CENTER
GPU
MAX SERIES
メモリー帯域幅
メモリー容量
移植 & リファクタリング
intel


XPU

将来の HPC システム



柔軟性に優れた次世代アーキテクチャー

開発コード名:

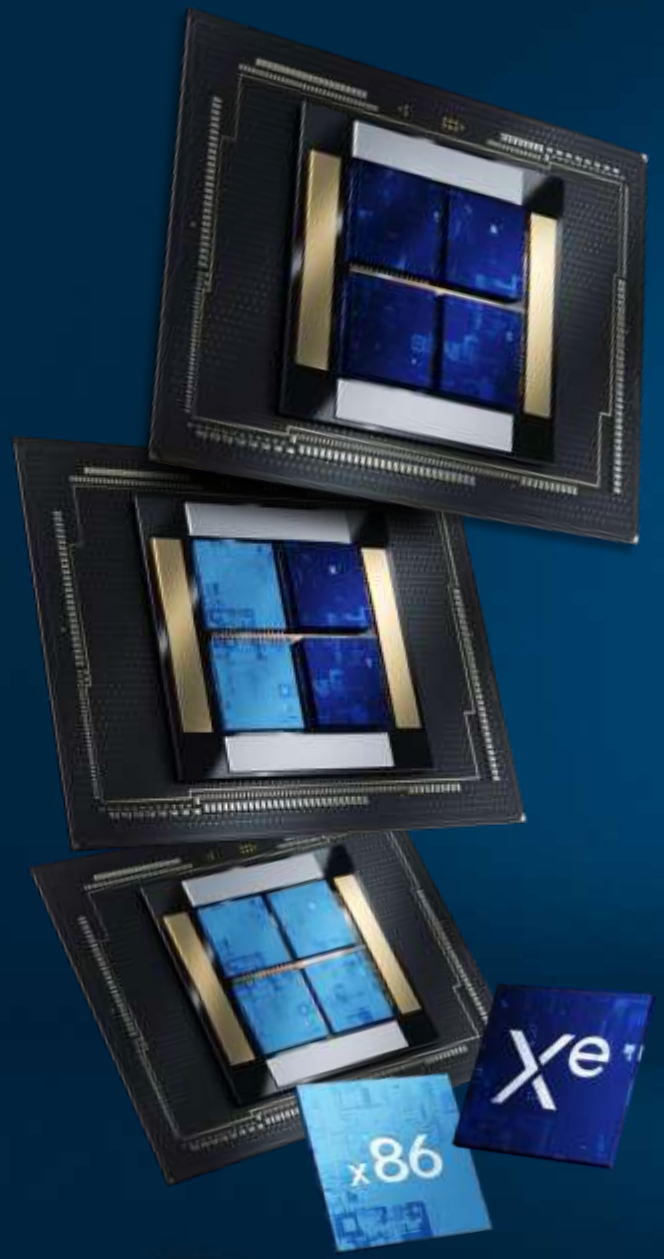
Falcon Shores

XPU

以下のすべての点で大幅に向上

- 消費電力当たり性能
- x86 ソケットでの演算密度
- メモリーの容量と帯域幅

IDM 2.0 で製造



この図は例示のみを目的とし、幅広いパフォーマンス要件にわたってフルインストール・ベースに対応するように設計された拡張可能なアーキテクチャーを表しています。

インテルのマックス・シリーズ CPU

64GB
HBM2e
16GB が
4 スタック

1GB 以上
P-core
当たりの
HBM2e

メモリーモード
- HBM のみ
- フラット
- キャッシュ

インテル® Xeon® CPU マックス・シリーズ ⟷ oneAPI ⟷ インテル® GPU マックス・シリーズ

シームレスにソフトウェアを拡張

2023年リリース | SYCL サポートの拡張 | CUDA から SYCL への移行を強化 | TF と PyT に最適化

インテルのマックス・シリーズ GPU

最大
64MB
L1 キャッシュ

最大
408MB
L2 キャッシュ

最大
128GB
HBM2e

30 以上
インテル® Xeon® CPU マックス・シリーズの システム設計

DELL Hewlett Packard Enterprise Lenovo SMC FUJITSU GIGABYTE inspur NEC hyve Atos ASUS

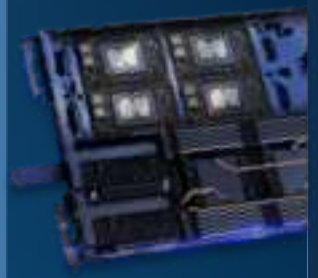
intel XEON MAX SERIES

intel DATA CENTER GPU MAX SERIES

可能性を最大限に引き出す

最大
128
レイ・トレーシング・ユニット

15 以上
インテル® GPU マックス・シリーズの システム設計



intel アクセラレーション・エンジン内蔵
HPC & AI 向け

- インテル® AVX 512
- インテル® DL ブースト
- インテルの DSA
- インテル® AMX

5 倍以上の性能
EPYC 7773-X、
インテル® Xeon® Platinum 8380 プロセッサーとの比較
Stream Triad

1.5 倍の性能
NVIDIA A100 との比較
ExaSMR - NekRS

2 倍以上の性能
幅広いワークロード全体 (幾何平均)
EPYC 7773-X との比較

CosmoFlow
Particle Shower SIM
OpenFOAM
Altair Acusolve
DeePMD
WRF
LAMMPS

素晴らしい組み合わせ
インテルの
マックス・シリーズ CPU/GPU
16 倍以上
の性能*

マックス・シリーズ 1100 GPU
300W

マックス・シリーズ 1350 GPU
マックス・シリーズ 1550 GPU
450W | 600W

マックス・シリーズのサブシステム
1,800W | 2,400W

*AMD EPYC 7773X との比較。分子動力学コード LAMMPS (液晶ワークロード) を実行。

ワークロードと構成の詳細については、http://intel.com/PerformanceIndex/ (英語) の SuperComputing 22 ページを参照してください。結果は異なる場合があります。


intel

注意事項および免責条項

性能は、使用状況、構成、その他の要因によって異なります。詳細については Performance Index サイトを参照してください。

性能の測定結果は、構成に示されている日付時点のテストに基づいています。また、現在公開中のすべてのアップデートが適用されているとは限りません。 構成の詳細については、補足資料を参照してください。 絶対的なセキュリティーを提供できる製品またはコンポーネントはありません。

実際のコストや結果は異なる場合があります。

インテルは、サードパーティーのデータについて管理や監査を行っていません。 ほかの情報も参考にしてデータの正確さを評価してください。

インテルのテクノロジーを使用するには、対応したハードウェア、ソフトウェア、またはサービスの有効化が必要となる場合があります。