

最先端のパフォーマンスを実現する インテル® XPU 戦略

力 翠湖 (Midoriko Chikara)

データセンター・セールス・グループ / HPC テクニカル・ソリューション・スペシャリスト

注意事項および免責条項

インテル® テクノロジーの機能と利点はシステム構成によって異なり、対応するハードウェアやソフトウェア、またはサービスの有効化が必要となる場合があります。実際の性能はシステム構成によって異なります。

絶対的なセキュリティを提供できる製品やコンポーネントはありません。

テストでは、特定のシステムでの個々のテストにおけるコンポーネントの性能を文書化しています。ハードウェア、ソフトウェア、システム構成などの違いにより、実際の性能は掲載された性能テストや評価とは異なる場合があります。パフォーマンスおよびベンチマーク結果の詳細については、<http://www.intel.com/benchmarks/> (英語) を参照してください。

性能に関するテストに使用されるソフトウェアとワークロードは、性能がインテル® マイクロプロセッサ用に最適化されていることがあります。SYSmark* や MobileMark* などの性能テストは、特定のコンピューター・システム、コンポーネント、ソフトウェア、操作、機能に基づいて行ったものです。結果はこれらの要因によって異なります。製品の購入を検討される場合は、他の製品と組み合わせた場合の本製品の性能など、ほかの情報や性能テストも参考にして、パフォーマンスを総合的に評価することをお勧めします。詳細については、<http://www.intel.com/benchmarks/> (英語) を参照してください。

インテル® アドバンスド・ベクトル・エクステンション (インテル® AVX)* は、特定のプロセッサ演算で高いスループットを示します。プロセッサの電力特性の変動により、AVX 命令を利用すると、a) 一部の部品が定格周波数未満で動作する、b) インテル® ターボ・ブースト・テクノロジー 2.0 を使用する一部の部品が任意または最大のターボ周波数に達しない可能性があります。パフォーマンスは、ハードウェア、ソフトウェア、システム構成により異なります。詳細については、<http://www.intel.com/go/turbo/> を参照してください。

インテル® コンパイラーでは、インテル® マイクロプロセッサに限定されない最適化に関して、他社製マイクロプロセッサ用に同等の最適化を行えないことがあります。これには、インテル® ストリーミング SIMD 拡張命令 2、インテル® ストリーミング SIMD 拡張命令 3、インテル® ストリーミング SIMD 拡張命令 3 補足命令などの最適化が該当します。インテルは、他社製マイクロプロセッサに関して、いかなる最適化の利用、機能、または効果も保証いたしません。本製品のマイクロプロセッサ依存の最適化は、インテル® マイクロプロセッサでの使用を前提としています。インテル® マイクロアーキテクチャーに限定されない最適化のなかにも、インテル® マイクロプロセッサ用のものがあります。この注意事項で言及した命令セットの詳細については、該当する製品のユーザー・リファレンス・ガイドを参照してください。

記載されているコスト削減シナリオは、指定の状況と構成で、特定のインテル® プロセッサ搭載製品が今後のコストに及ぼす影響と、その製品によって実現される可能性のあるコスト削減の例を示すことを目的としています。状況はさまざまであると考えられます。インテルは、いかなるコストもコスト削減も保証いたしません。

インテルは、本資料で参照しているサードパーティーのベンチマーク・データまたはウェブサイトについて管理や監査を行っていません。本資料で参照しているウェブサイトアクセスし、本資料で参照しているデータが正確かどうかを確認してください。

Intel、インテル、Intel ロゴ、その他のインテルの名称やロゴは、Intel Corporationまたはその子会社の商標です。

その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

© 2021 Intel Corporation. 無断での引用、転載を禁じます。

HPC の"今"

これまで以上にシステムへの要求が高まっている



ライフ
サイエンス



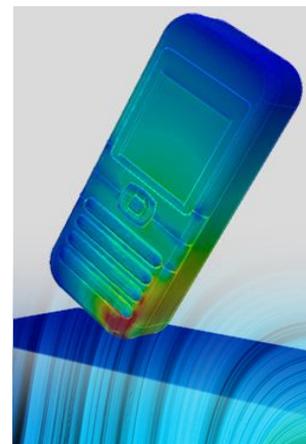
金融
サービス



エネルギー



ビジネス
解析



工業



クラウド



AI

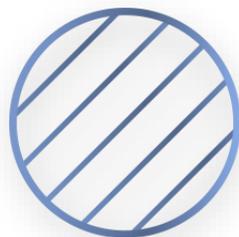
Heterogeneity 異種混合へ

ワークロードの最適化と加速

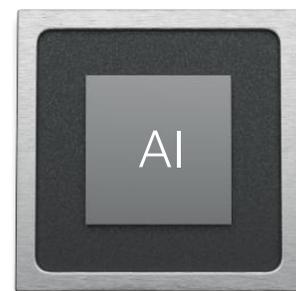
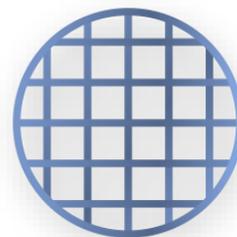
Scalar



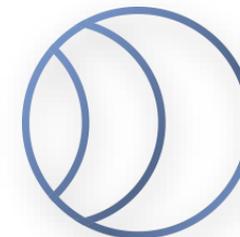
Vector



Matrix



Spatial



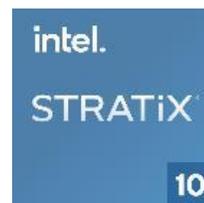
インテル® XPU アーキテクチャー



GPU for
HPC & AI



FPGA

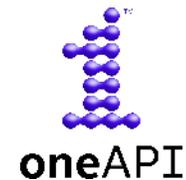


VPU



クロス・アーキテクチャーの開発のための oneAPI プラットフォーム

インテル® HPC ポートフォリオ



XPU アーキテクチャー



ソフトウェアツール



メモリー



ストレージ



インターコネク



セキュリティ

インテル® Xeon® スケーラブル・プロセッサー

融合を実現する最適化された唯一のメインストリームCPU

Intel Advanced Vector Extensions 512
Intel® Deep Learning Boost
Intel Optane™ persistent memory

Cascade Lake

14nm

新しいAIアクセラレーション
(Intel® DL Boost with VNNI)
新しいメモリーストレージ階層

Ice Lake

10nm

新しいマイクロアーキテクチャ
メモリバンド幅の増加

Sapphire Rapids

新しい世代の Intel® DL Boost
(Intel® Advanced Matrix Extensions)

イノベーションを加速
エッジからクラウドまで含めたスーパーコンピューティングへ

第3世代 インテル® Xeon® スケーラブル・プロセッサー

Performance made flexible

AI & セキュリティー機能を備えた
唯一の x86 データセンター・プロセッサー



Targeted for IS-2S systems

先進的なセキュリティー・ソリューション

スケーラブル・フレキシブル・カスタマイズ可能

Intel Software Guard Extensions

Intel Crypto Acceleration

Intel Total Memory Encryption

Intel Platform Firmware Resilience

Intel Deep Learning Boost

Intel Speed Select Technology

Intel AVX-512

oneAPI POWERED
Optimized Software

次世代のインテル® Xeon® スケーラブル・プラットフォーム

最大 6TB システムメモリー容量 (ソケットあたり) DRAM + PMem

最大 8CH DDR4-3200 2 DPC (ソケットあたり)

最大 2.6X 第2世代と比較したメモリー容量

最大 64 Lanes PCI Express 4 (ソケット当たり)

Breakthrough Data Performance

OPTANE
インテル® Optane™ パーシステント・メモリー 200 シリーズ

OPTANE
インテル® Optane™ SSD P5800X シリーズ

インテル® SSD D シリーズ

Faster, Flexible, Data Scale

インテル® イーサネット 800 シリーズ ネットワーク・アダプター

インテル® Agilix™ FPGA ソリューション

intel SELECT SOLUTIONS

intel MARKET READY

実アプリケーションにおいて、
前世代から飛躍した性能向上を実現

LAMMPS

60%

Improvement
Xeon® 8380 vs prior gen

GROMACS

64%

Improvement
Xeon® 8380 vs prior gen

RELION

61%

Improvement
Xeon® 8380 vs prior gen

HPCG

41%

Improvement
Xeon® 8380 vs prior gen

NAMD

57%

Improvement
Xeon® 8380 vs prior gen

HPL

38%

Improvement
Xeon® 8380 vs prior gen

STREAM TRIAD

47%

Improvement
Xeon® 8380 vs prior gen

WRF

58%

Improvement
Xeon® 8380 vs prior gen

BINOMIALOPTIONS

28%

Improvement
Xeon® 8380 vs prior gen

OPENFOAM

51%

Improvement
Xeon® 8380 vs prior gen

BLACK SCHOLES

67%

Improvement
Xeon® 8380 vs prior gen

MONTE CARLO

70%

Improvement
Xeon® 8380 vs prior gen

AI パフォーマンス

20 のワークロードの平均で

50%

高い AI パフォーマンス
(DL ブーストを搭載した第 3 世代インテル® Xeon® スケーラブル・プロセッサ (40C) と AMD EPYC 7763 (64C) との比較)

20 のワークロードの平均で

30%

高い AI パフォーマンス
(DL ブーストを搭載した第 3 世代インテル® Xeon® スケーラブル・プロセッサと NVIDIA A100 GPU との比較)

Performance Made Flexible with AI

AI 機能を実装した第 3 世代インテル® Xeon® スケーラブル・プロセッサ



Data Practitioner

Productivity

1
oneAPI

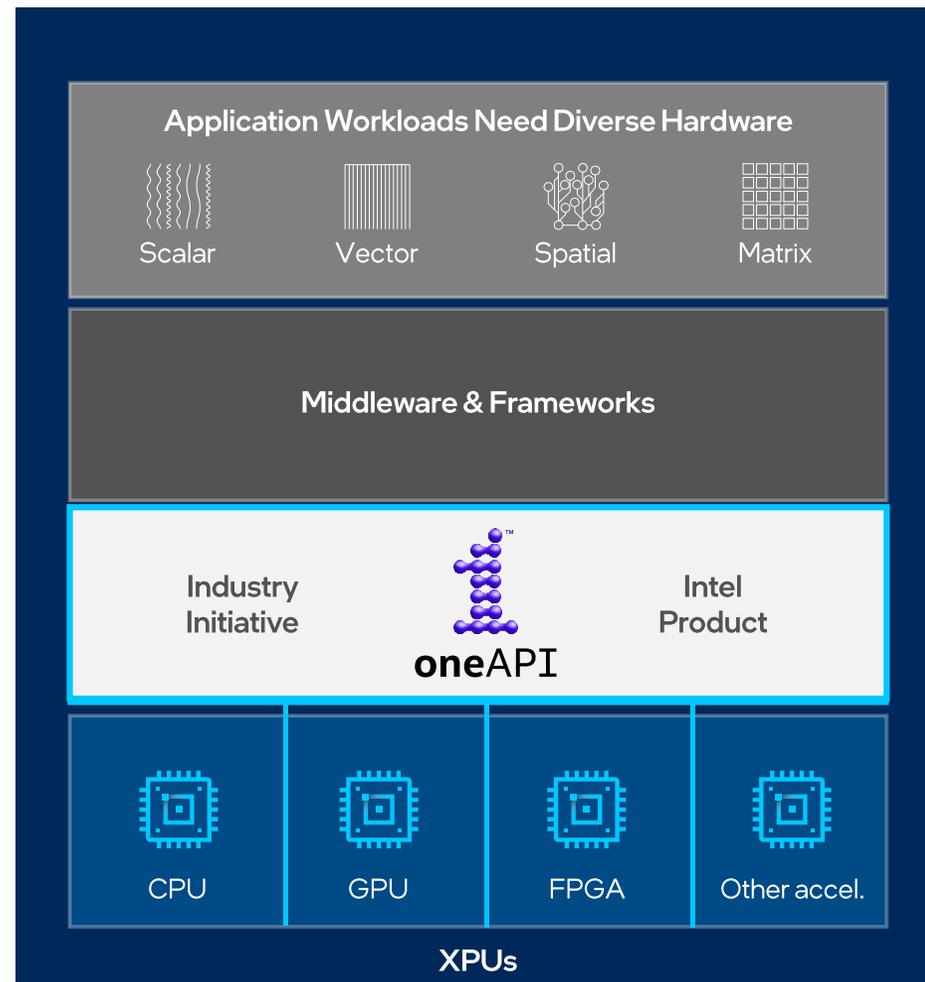
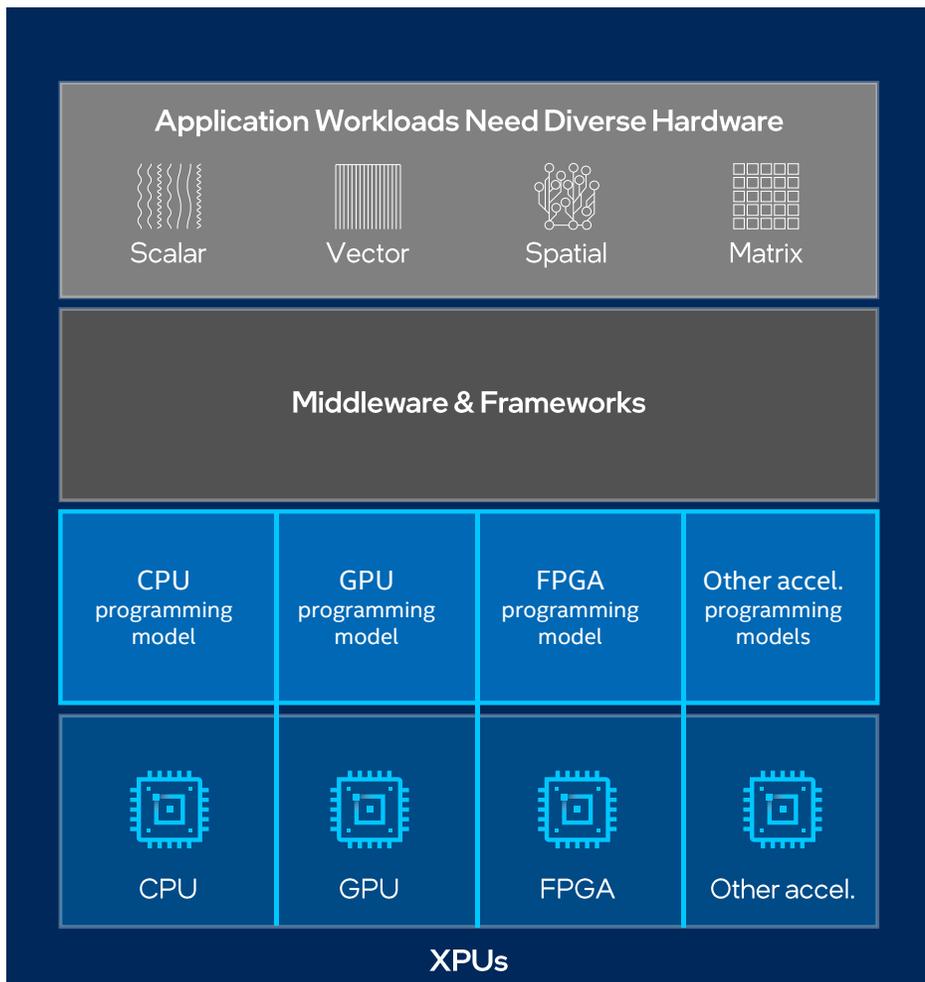
Simplicity

200 enterprise
ISV solutions

For workloads and configurations see [43, 44] at <https://www.intel.com/3gen-xeon-config>. Results may vary.

インテル® oneAPI

アーキテクチャ選択の自由を提供する業界標準とオープン仕様



インテル® oneAPI ツールキット

A complete set of proven developer tools expanded from CPU to XPU



Intel® oneAPI Base Toolkit

Native Code Developers

A core set of high-performance tools for building C++, Data Parallel C++ applications & oneAPI library-based applications



Add-on Domain-specific Toolkits

Specialized Workloads



Intel® oneAPI Tools for HPC

Deliver fast Fortran, OpenMP & MPI applications that scale



Intel® oneAPI Tools for IoT

Build efficient, reliable solutions that run at network's edge



Intel® oneAPI AI Analytics Toolkit

Accelerate machine learning & data science pipelines with optimized DL frameworks & high-performing Python libraries



Intel® oneAPI Rendering Toolkit

Create performant, high-fidelity visualization applications

Toolkit powered by oneAPI

Data Scientists & AI Developers



Intel® Distribution of OpenVINO™ Toolkit

Deploy high performance inference & applications from edge to cloud

データセンター・プロセッサ・ロードマップ

Cascade Lake†



intel
2nd Gen
Intel® Xeon®
Scalable
Processor

Cooper Lake†



intel
3rd Gen
Intel® Xeon®
Scalable
Processor

- Intel® DL Boost
- BFLOAT16

Ice Lake†



ICX

- TME
- PCIe* Gen 4
- 8 memory channels
- Crypto accel.
- ICX-D

Sapphire Rapids†



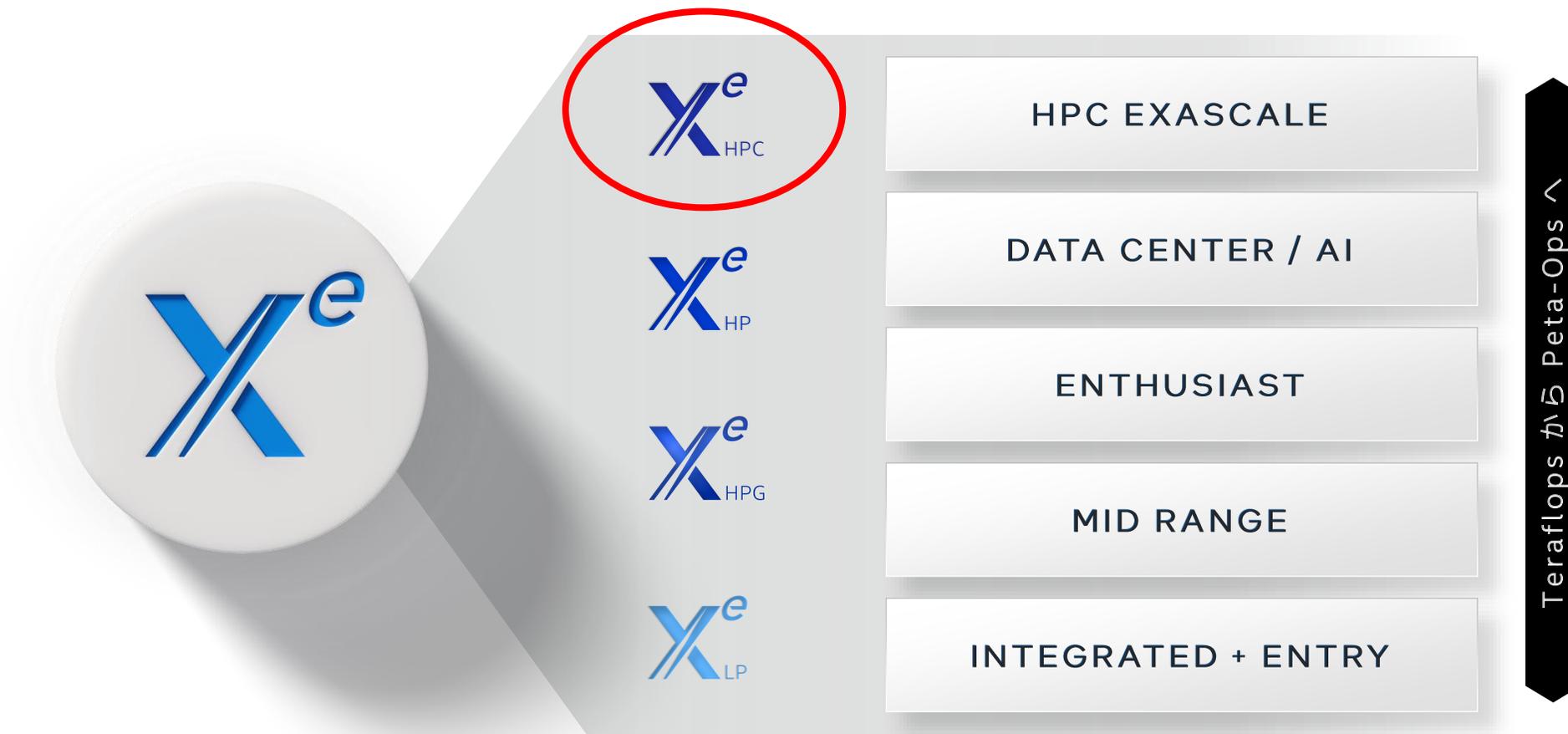
SPR

- DDR5
- PCIe* Gen 5
- CXL 1.1
- AMX
- DSA

†開発コード名

GPU Architecture Strategy

1つのアーキテクチャーと4つのマイクロアーキテクチャー



X^e-HPC (コードネーム:Ponte Vecchio)

HPC と AI に最適化された最初の X^e ベースの GPU

スケーラブル・アーキテクチャー

Compute

最高の計算密度：ソケットとノード



Data Parallel Matrix Engine

ベクトルおよび行列計算のための高効率 ISA

SIMD & SIMT

(CPU STYLE)

(GPU STYLE)

データパス制御による最大のパフォーマンス

Memory

キャッシュと帯域幅に敏感なワークロード用に最適化されたメモリー階層



RAMBO Cache

大規模な L3 キャッシュにより大型モデルのパフォーマンスを押し上げ



Compute-Memory Fabric

統一されたメモリーを提供する Xe Link とメモリー 構造

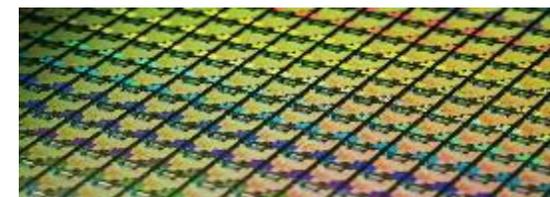
Technology

新しい境界線を定義する革命的な製品オプション



FOVEROS Advanced Packaging

最終的な組み立てとテストはインテルが保証

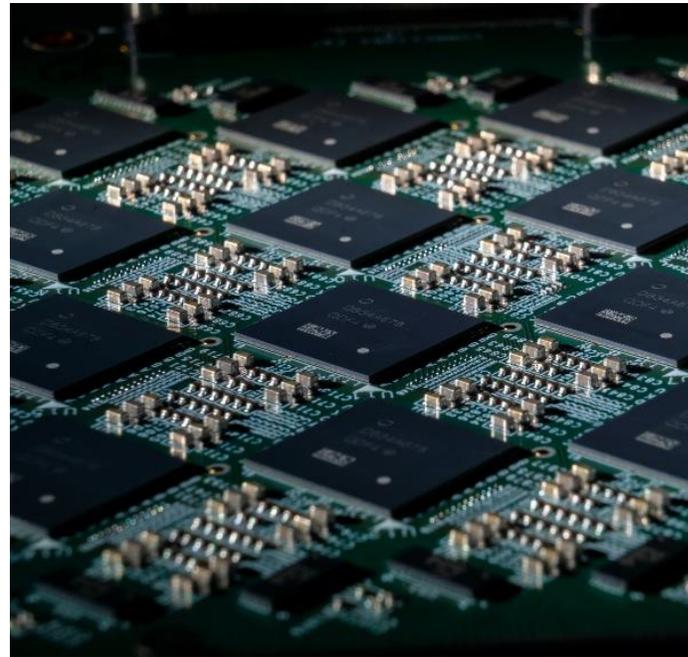


Leading Process Technology

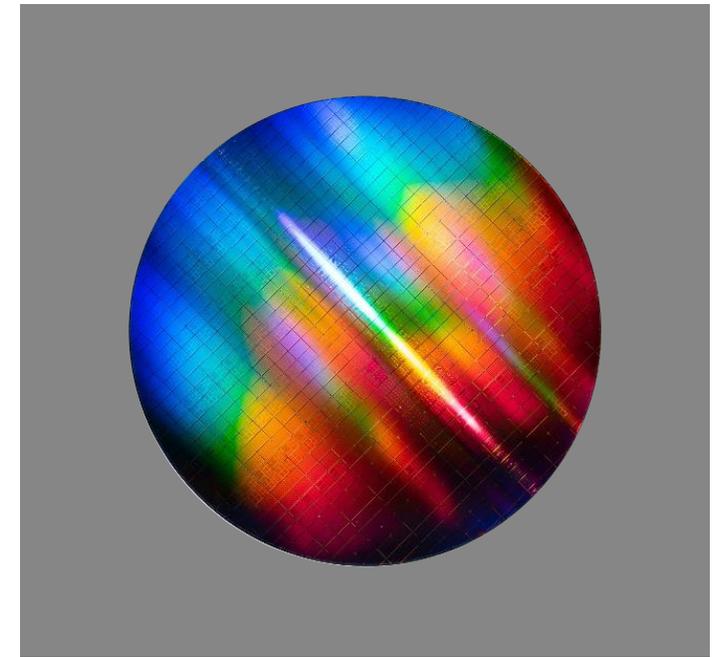
To the 10^{18} and Beyond エクサとその先へ



Exascale
Computing

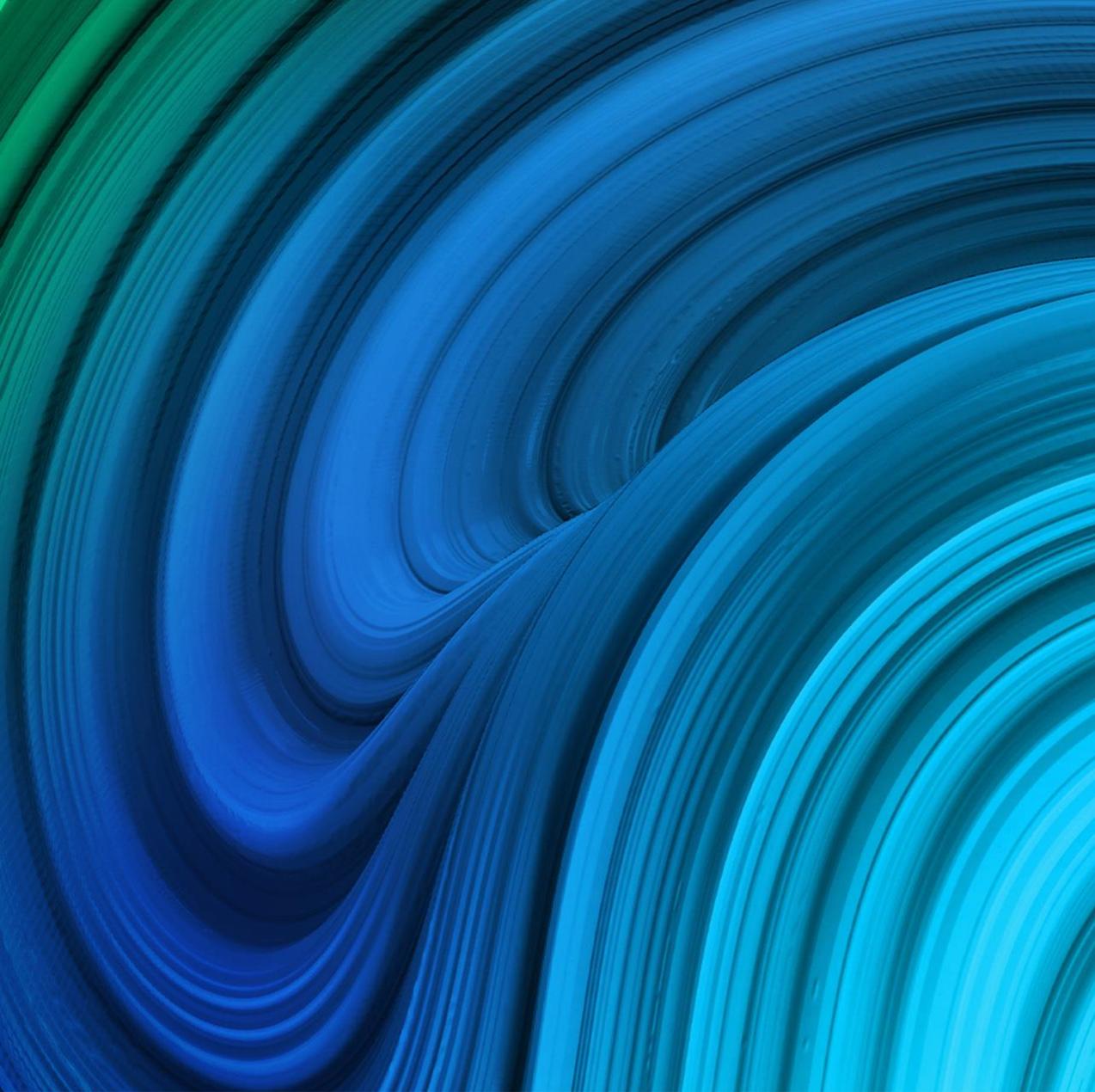


Neuromorphic
Computing



Quantum
Computing

See our virtual talks, and demos
at: www.hpcwire.com/solution_channel/intel/



Thank you!
For more information:
intel.com/hpc