

インテル® Agilex™ FPGA : データ中心の世界を牽引する柔軟性と機敏性の融合

インテルのイノベーションを最大限に活用して FPGA を再定義

著者

インテル プログラマブル・ソリューションズ事業本部、テクニカル・スタッフ・シニアメンバー
Martin S. Won

「インテルは、コンピューティング、メモリー、通信の領域で幅広いヘテロジニアス・インテグレーションを可能にしています。決められた消費電力枠で、小さなフットプリントに多彩なテクノロジーを接続し、積み重ねることによって、非常に優れた柔軟性ととともに、独自のコスト性とパフォーマンス特性を実現できます。異なる種類のコンピューティング要素や異なるプロセスに基づく多様性に富んだ機能、さらにはデータセンター・クラスのテクノロジーを最も効率的なパッケージに統合できたときの可能性を想像してみてください」

ー インテル コーポレーション、テクノロジー / システム・アーキテクチャー & クライアント・グループ、グループ・プレジデント兼最高エンジニアリング責任者
Venkata (Murthy) M. Renduchintala 博士

目次

はじめに : データの急増という課題.....	2
理想的なソリューションの設計.....	3
インテル® Agilex™ FPGA の構成要素.....	4
多様なテクノロジーの統合.....	8
インテル® Agilex™ FPGA のアプリケーション開発の促進.....	9
ユースケースの例.....	9
まとめ.....	9

概要

エッジからコア、さらにはデータセンターまで、多様な市場における急速な変化により、ソリューション・プロバイダー、アプリケーション開発者、業界、企業は、かつてないスピードで大規模なイノベーションを実現することを求められています。その結果、エッジにおける組込みデバイスによるほぼリアルタイムに処理された情報の提供、ネットワーク・コアではネットワーク機能仮想化 (NFV) による大量のデータの集約と処理、またデータセンターにおいては分析、メモリー、ストレージの増大への対応が急務となっています。ネットワークのエッジからコアを経由して、データセンターに到達するまでのあらゆるポイントで今、インフラストラクチャーにはデータが氾濫しつつあるのです。すべての市場セクターが一斉にデータの構造化と処理を推進し、その変化のスピードが増した結果、柔軟性に対する要求も高まっています。新しいインテル® Agilex™ FPGA は、単なる最新のプログラマブル・ロジック製品ではありません。インテル® テクノロジーが牽引するさまざまな分野における技術革新が結集されたインテル® Agilex™ FPGA は、エッジからデータセンターに及ぶさまざまな変革を通じて新たな価値や意味を引き出すという、新たな機会を創出します。

エッジ / 組込み機器	ネットワーキング / NFV	データセンター
		
目標 : リアルタイムに処理された情報の提供 ニーズ : カスタマイズされた接続性と低レイテンシーなコンピューティング	目標 : 広帯域なアグリゲーション / 処理によるネットワーク機能の仮想化 ニーズ : ネットワーク・ワークロードの高速化と、データ・スループットの最大化	目標 : 急増するデータの管理、整理、処理 ニーズ : 多様なワークロードにおける、カスタマイズされた低レイテンシーなアクセラレーション

図 1. インテルのテクノロジーによって、多様な市場において、エッジからネットワーク、クラウドまでのデータ処理の最適化および高速化が可能に

はじめに: データの急増という課題

大量のデータの生成、さらには分析の強化、実用的なインサイトの獲得がビジネスの競争において不可欠となった今、データこそがコンピューティング・テクノロジーの将来を決定付ける存在となっています。データを処理するほぼすべての分野のインフラストラクチャーが、こうした市場の大きな変化に直面しています。データの種類の多様化とデータから価値を引き出すための新しい手法やアルゴリズムとともに、機敏性と柔軟性の向上に対するニーズが高まっています。データの新たな処理方法が登場し、新しいデータサービスが次々に生み出されるため、意思決定者、ソリューション・プロバイダー、アプリケーション開発者にとって複雑さが増しています。

例えば、AIの実用的なアプリケーションは日々進化を続けていますが、多くの分野で、AIをどのように導入すればよいのか、今後どのように導入されるのかはまだ明らかになっていません。こうした不確実性は、これらの進化するAIワークロードを処理するために柔軟性が必要になることを意味します。また、多くのAIアプリケーションにどの程度の精度が必要なのかも未だ不明です。さまざまな精度に対応するためには、柔軟なアプローチを採用しなければなりません。

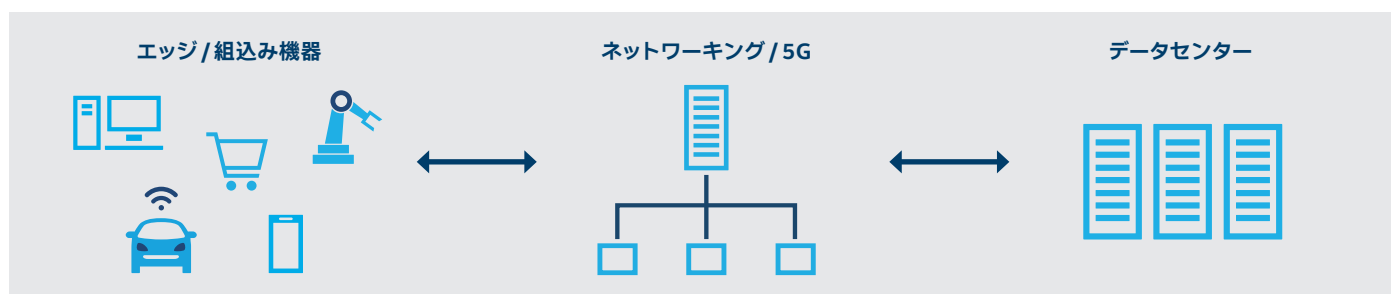


図2. 大量のデータの生成により、インフラストラクチャー全体で需要が急増

エッジと組み込み機器

さまざまな企業や業界がIoTとそのコネクテッド・デバイスに関与することで、その市場規模は2020年までに約310億米ドル規模に達することが見込まれています。¹ ネットワーク・エッジでは、より多くのデータが生み出され、処理され、送信されます。構造化/非構造化データを活用し、ほぼリアルタイムでインサイトを生み出すことが求められています。フィルター処理されていない生データをエッジで処理し、それを構造化して、ディープラーニング推論を実行するには、新たなレベルの性能と容量が必要となります。

ネットワーク・コア

エッジで取り込まれたデータがネットワーク・コアへと移動することにより増加し続けるトラフィックは、適切に処理されなければなりません。ネットワークの能力が、集中管理されたハードウェアに依存している場合、付加価値を提供するための機能は、特定のハードウェアが用意された場所にしか実装できません。しかし今日、大量のデータの転送と処理をより迅速に行うために、ネットワーク・レベルの重要なポイントにおいて、より高度なインテリジェンスと柔軟性が求められています。つまり、これまで特定の機器に集中していた機能の分散化と、ニーズに応じた柔軟な実装を実現する必要があります。

その結果、高い使用率と多様なネットワーキング・ワークロードに対応する高度なプログラマビリティを確保して、必要なリソースの柔軟性を提供するための手段として、ネットワーク機能仮想化(NFV)が登場しました。NFVを標準的な既製製品(COTS)によるインフラストラクチャー上に構築することで、コスト効率の高い導入を実現できます。しかし、大量のデータ、低いレイテンシー、広帯域幅のアプリケーションの増加により、仮想化されたネットワーク機能では、サービス品質(QoS)の維持に限界が生じる可能性があります。FPGAアクセラレーションは、インテル®アーキテクチャー・ベースのNFVの使用率とコスト効率の高いスケーリングを最適化し、ユーザー、

データ、アプリケーションの経時的な増加に伴い、SmartNICの機能を拡張してネットワーキングを高速化します。

5Gの到来もネットワークに新たな要求を突き付けます。膨大な数のコネクテッド・デバイスと、多様なユースケースによって生じるデータ・トラフィックの急増やさまざまなサービス品質レベルに対応するために、ネットワークには高可用性を維持し、効率的にスケーリングできることが求められます。しかし、特殊なアプライアンスに依存した既存のソリューションでは、将来の5Gのニーズに十分に対応することはできません。独自の規格を採用したハードウェアは、幅広く導入するにはコストがかかりすぎる場合があります。また、進化する顧客のニーズに応え、開発者によるエコシステムを活用できるだけの柔軟性を備えていない可能性もあります。ここでも、変化する顧客環境と実装/導入のさまざまな選択肢に対応するには柔軟性が必要となります。

データセンター

財務分析、データベース・アクセラレーション、ゲノミクスからビデオ・トランスコーディング、ネットワーク/ストレージ・アクセラレーション、セキュリティまでさまざまな業種のユースケースにおいて、データセンターで大規模なデータセットを処理し、重要な分析が行われるようになりました。

さらに、さまざまな種類のデータのリアルタイム分析に対する顧客のニーズが増え続けていることから、データセンター内で発生するデータ処理の種類は常に変化しています。また、データセンター事業者が、顧客がどのようなデータサービスを期待しているかを把握できない場合もあるかもしれません。特定の演算やアルゴリズムのみを高速化するソリューションでは、将来のアクセラレーション・ニーズに柔軟に対応できなくなります。まだ定義されていない新たな処理に対応するためには、周到な計画が欠かせません。これが、ハードウェア・レベルの柔軟性(つまり、FPGAが提供する柔軟性)が不可欠であるもう1つの理由です。

データセンター・リソースの効率的な利用は非常に重要であり、アクセラレーションはデータ中心のコンピューティングの実現に効果を発揮します。インテル® Xeon® スケーラブル・プロセッサは世界中の多くのデータセンターで活用されているため、可能な限り低遅延、低レイテンシーでインテル® Xeon® スケーラブル・プロセッサが実行する機能を高速化して、できる限りリソース効率の高い方法で、極めて価値の高い演算に対して最大限の性能を実現することが求められます。

ニーズ: 市場固有の要件に対応できる柔軟性

エンドカスタマーがサービスやソリューションを拡大するにつれて、データが生成、処理、転送、または格納される場所に関係なく、データ処理のより詳細なカスタマイズが必要になります。こうした高度な処理には、最適な電力効率と性能を備え、目的に特化した機能が求められます。データ処理の要件は、さまざまな市場のアプリケーション・クラスごとに異なります。これらの要件には、エッジと組み込みデバイスまたは自動車、ネットワークでの大量のデータ・トラフィックとイー

サネット速度への対応、データセンターでの広帯域幅、低レイテンシーのコンピューティング・アクセラレーションの提供など、消費電力、スペース、コストが重視されるスモール・フォーム・ファクターや複雑なデザインに対応する、電力効率とコスト効率に優れた機能が含まれる場合があります。

製品開発において特定のニーズを満たす際には、標準的な既製品を使用することも、カスタム ASIC を設計して導入することも可能です。しかし、標準的な製品は、業界の特定の要件に合わなかったり、市場で十分な差別化ができなかったりする可能性があります。カスタム ASIC はこれらの特定の機能に対応できますが、開発には多くの時間とコストを伴うため、採算性の確保については ROI の高いハードルがあります。

理想的なソリューションとは、特定の機能を実現するために、FPGA の柔軟性と ASIC レベルの性能効率および電力効率を兼ね備えたソリューションです。

理想的なソリューションの設計

これらの課題に対処するには、新たなアーキテクチャー・アプローチが必要となります。インテル® Agilex™ FPGA は、新しい FPGA アーキテクチャー・アプローチを採用し、各アプリケーション・クラスに固有の課題への対処を目的としたカスタマイズが可能な FPGA 製品として構築されました。

理想とされるのは、柔軟性と最大限の電力 / 性能効率を兼ね備えたソリューションです。その実現と次世代のプログラマブル・ロジックの構築に向けて、インテルの革新的な FPGA アーキテクチャー・アプローチでは、さまざまな半導体素子を単一のシステムインパッケージ (SiP) に集積することを可能にしました。

このアプローチでは、インテルの 10nm 製造プロセスを採用した高性能な FPGA コアダイを、特定機能のチップレットと組み合わせ、先進の 3D パッケージングによってすべての要素を 1 つの製品にヘテロジニアスに統合します。その結果、カスタマイズされた柔軟なソリューションによって、多様なアクセラレーション・アプリケーションや他のアプリケーションへの対応が可能になります。チップレットは、PCIe* Gen 5、112G トランシーバー、インテル® Xeon® スケーラブル・プロセッサのキャッシュ・コヒーレント・インターフェイスなどの機能を提供します。その他のトランシーバー・タイプ、カスタム I/O、カスタム・コンピューティング機能など、他のチップレットも利用できます。エンベデッド・マルチダイ・インターコネクト・ブリッジ (EMIB) や、インテル独自の他の最先端集積 / パッケージング・テクノロジーを活用する新しいアーキテクチャー・アプローチにより、従来の FPGA ダイと専用の半導体ダイを組み合わせ、ターゲット・アプリケーション用としてデバイスを最適化することができます。

インテル® Agilex™ FPGA & SoC は、前世代の FPGA と比較して、ファブリック性能の向上、消費電力の削減、デジタル信号処理 (DSP) 機能の向上、設計者の生産性の向上を実現することによって、次世代の高性能アプリケーションを実現します。インテル® Agilex™ FPGA は、データ中心のコンピューティングの多種多様な課題に対処し、ビジネスと産業分野の新たな可能性を切り開きます。汎用ファブリックによって柔軟性を実現すると同時に、シリコンレベルでの非常に効率的な処理により、市場ごとに求められるカスタマイズされた機能を提供します。

また、独自のインテル® eASIC™ デバイス・テクノロジーにより、インテル® Agilex™ FPGA デバイスにカスタム機能を迅速に統合することもできます。インテル® eASIC™ テクノロジーを使用すると、顧客の FPGA デザインを、ASIC レベルの性能効率と電力効率を実現する特定機能のダイに変換し、高度なカスタマイズのための他の機能とともに 1 つのコンポーネント・パッケージに統合できます。

柔軟性と機敏性の融合によって、ターゲット・アプリケーションの要件に対応

FPGA は、進化する市場の要件に対応する柔軟性が高く評価されています。ただし、機能が安定するにつれ、それほど柔軟性が必要とされなくなるケースは少なくありません。そのような機能はできるだけ「ハード化」し、最大限の電力効率と性能を得ることが望まれます。柔軟性の向上は、電力効率の低下、消費電力の増加というトレードオフを常に伴うため、特定の機能をハード化することには一定のメリットがあります。

従来、FPGA は、単一のモノリシック・ダイまたは同じ種類のモノリシック・ダイの複数のインスタンスを使用してデザインされていました。現在では、インテルの新しい先進のパッケージング・テクノロジーにより、複数の異なるシリコンダイを 1 つのパッケージに統合することが可能になりました。インテルは、プロセスの種類や機能が異なるダイを統合することで、かつてない柔軟性とカスタマイズを実現しています。専用ダイの例を次に示します。

- 低レイテンシーのキャッシュ・コヒーレントなプロセッサ・アクセラレーションのためのインターフェイス
- 112G トランシーバー (XCVR) やデータ・コンバーターなどの高度なアナログ機能
- アプリケーション固有の機能に対応するカスタム・コンピューティング・エンジン
- ロジック・ファブリックに近接して繋がる、さまざまな種類と構成のメモリー

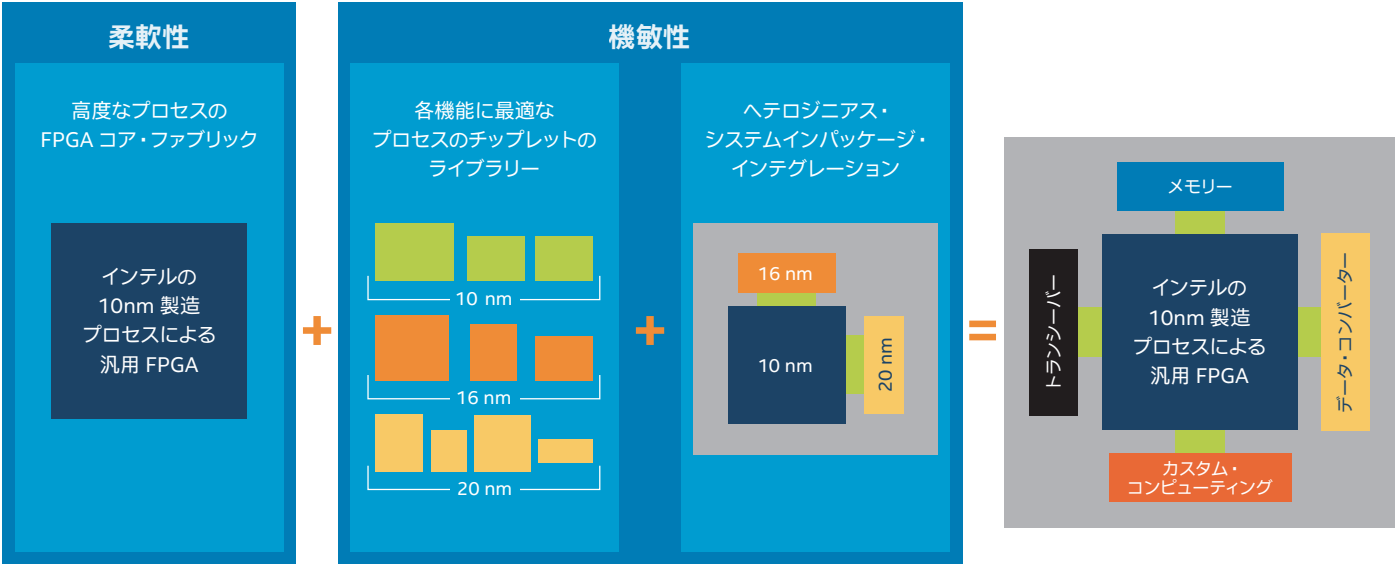


図 3. インテル® Agilex™ FPGAの主な領域にインテル独自のアーキテクチャー・イノベーションを結集

コンピューティング領域全体で市場のニーズに対応するターゲットを絞った最適化		
エッジと組み込み機器	ネットワーク・コア	データセンター
電力 / 面積効率に優れた AI 推論 カスタムデータの前処理と取り込み	最大 112 Gbps の高速トランシーバー 最大 400 G のイーサネット・ブロック	低レイテンシーのインテル® Xeon® スケーラブル・プロセッサ・アクセラレーション 電力効率に優れた DSP ブロックによる AI などのアルゴリズムのアクセラレーション

インテル® Agilex™ FPGAの構成要素

高度な10nm FPGA ファブリック

すべてのインテル® Agilex™ FPGA デバイスの中心となる FPGA ファブリック・ダイは、インテルが誇る非常に高度な FinFET プロセスである 10nm チップ製造プロセス・テクノロジーを採用しています。このファブリック・ダイでは、10nm のトップクラスの性能に最適化されたハイパーレジスターと呼ばれるレジスターを FPGA 全体で使用する、第 2 世代のインテル® Hyperflex™ FPGA アーキテクチャーを採用しています。第 2 世代のインテル® Hyperflex™ アーキテクチャーでは、インテル® Quartus® Prime 開発ソフトウェアと組み合わせることで、次世代システムに不可欠なパフォーマンスと生産性が実現されます。

また、FPGA ファブリックは、前世代の FPGA よりも高い DSP 集積度に加えて、半精度浮動小数点 (FP16) と BFLOAT16 を専用回路でサポートし、AI 機能と DSP 演算を高速化するアーキテクチャーの最適化にも対応しています。

インテル® Agilex™ FPGA では、固定小数点および浮動小数点 DSP 演算を非常に効率的に実装できます。DSP ブロックでは、9x9 乗算器の数が前世代の 2 倍になっています。これにより、インテル® Agilex™ FPGA が DSP ブロックごとに提供できる INT8 演算の量も 2 倍になります。FP16 の新しいモードの追加により、FP32 での実装に比べ、デバイス使用率と消費電力を抑えつつ、特定の AI ワークロード (画像 / 物体検出のための畳み込みニューラル・ネットワーク (CNN) など) に対応する非常に効率的な実装がサポートされます。

第 2 世代インテル® Hyperflex™ アーキテクチャー

革新的な第 2 世代のインテル® Hyperflex™ アーキテクチャーは、従来型のアーキテクチャーでは不可能だった高いレベルの性能をサポートします。

第 1 世代と同様に、第 2 世代インテル® Hyperflex™ アーキテクチャーも、コア・ファブリック全体でハイパーレジスターと呼ばれる追加のレジスター・リソースを備えています。これらのレジスターは、配線構造とすべての機能ブロックの入力で使用できます。ハイパーレジスターは、いかに帯域幅を拡大し、面積 / 電力効率を向上させるかという問題に対するきめ細かいソリューションをもたらします。ハイパーレジスターを使用してこれらの手法を実装することにより、他のすべての FPGA ロジックリソースをロジック機能のために使用でき、従来のように LUT レジスターに到達するためフィードスルー・セルとして犠牲になるようなことはなくなります。

このアーキテクチャーの第 2 世代では、消費電力を最小限に抑えながら全体的なファブリック性能を向上させるためにいくつかの進歩を遂げています。最も大きな改善点の 1 つは、図 4 に示すように、ハイパーレジスターへの高速バイパスの追加です。

図 4 の左側は、インテル® Stratix® 10 FPGA のハイパーレジスターを表しています。レジスターを通過する信号パスとレジスターをバイパスする別の信号パスがあることが分かります。どちらの信号パスも、コンフィグレーション RAM によって制御されたマルチプレクサーを通過します。第 2 世代のインテル® Hyperflex™ アーキテクチャーの改善点の 1 つは、ハイパーレジスターのバイパス用パスを高速化したことです。

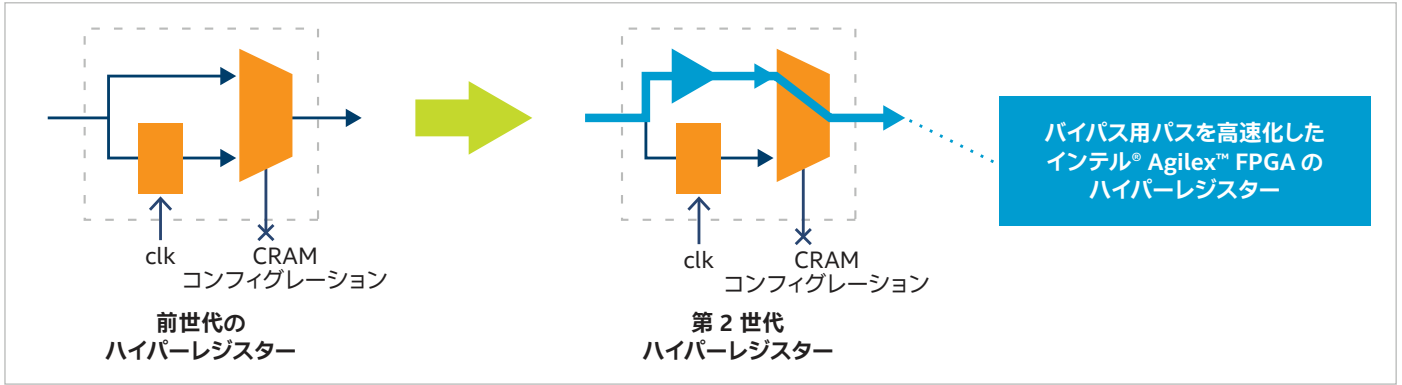


図 4. 高速バイパスによるハイパーレジスターの高速化によってファブリック性能が向上

この改善により、インテル® Hyperflex™ アーキテクチャーに最適化されたデザインのみならず、最適化されていないデザインの性能も向上します。

図5に、2つのデザイン例を示します。上の例はインテル® Hyperflex™ アーキテクチャーに最適化されており、下の例は最適化されていません。インテル® Agilex™ FPGA デバイスのアダプティブ・ロジック・モジュール (ALM) を大きなブルーのボックスで示し、ハイパーレジスターを2色で示します。オレンジのボックスがハイパーレジスターを使用していないもの、グレーの輪郭のブルーのボックスが使用しているものです。

上のデザインではハイパーレジスターを使用しており、インテル® Hyperflex™ アーキテクチャーに最適化されていることを示しています。一方、下のデザインではハイパーレジスターを使用しておらず、インテル® Hyperflex™ アーキテクチャーに最適化されていないことを示しています。

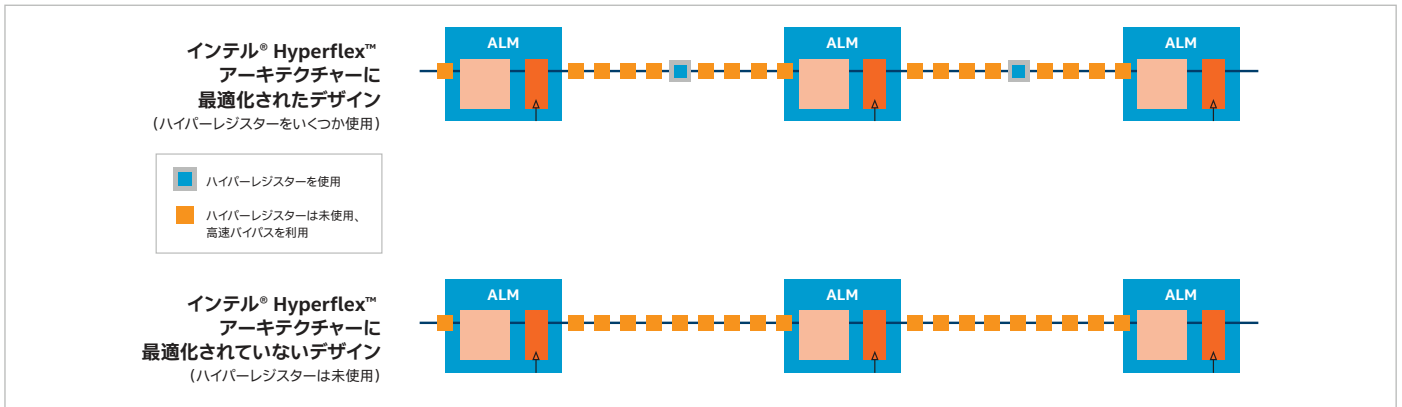


図 5. インテル® Hyperflex™ アーキテクチャーに最適化されたデザイン (上) と最適化されていないデザイン (下)

図6に、各デザインにおけるレジスター間の信号遅延を示します。信号遅延によってクリティカル・パスが決まり、最終的に各デザインのFmax（最高動作周波数）が決まります。

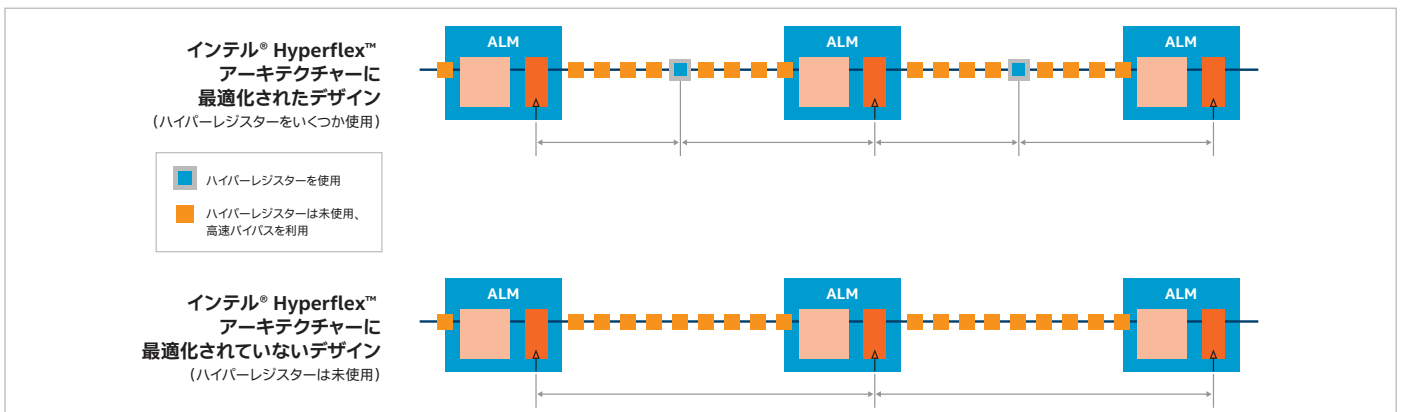


図 6. 各デザインのFmaxを決定するレジスター間の信号遅延

図7では、ハイパーレジスターの高速化されたバイパスとALMの性能向上により、2つのデザインタイプのどちらにおいても潜在的なクリティカル・パスが高速化されることが分かります。信号は、未使用のハイパーレジスター（オレンジのボックス）を高速で通過し、同期クリアとクロックイネーブルがないALMも高速で通過します。このように、第2世代インテル® Hyperflex™ アーキテクチャーにより、デザインがインテル® Hyperflex™ アーキテクチャーに最適化されているかどうかにかかわらず、すべてのデザインの性能が向上します。ただし、最適化されているデザインなら、最大限の性能向上が得られます。

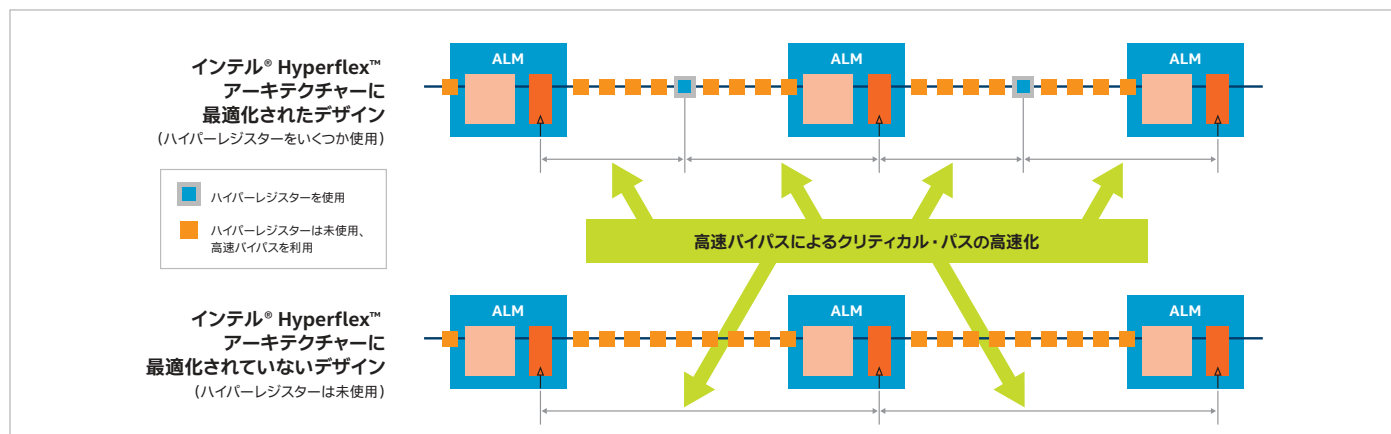


図7. インテル® Hyperflex™ アーキテクチャーの追加により、両方のデザインタイプのクリティカル・パスが高速化

第2世代インテル® Hyperflex™ アーキテクチャーの改善は、インテル® Agilex™ FPGA デバイスがもたらす性能上の利点に大きく貢献しています。

DSP アクセラレーション機能

より高精度の信号処理のニーズに対応するために、インテルは業界初の変換精度 DSP ブロックを開発しました。前世代のインテル® FPGA では、変換精度 DSP ブロックを強化し、単精度浮動小数点 (FP32) のサポートを追加しました。インテル® Agilex™ FPGA では、変換精度 DSP ブロックを引き続き強化し、新たに半精度浮動小数点 (FP16) と BFLOAT16 (BF16) をサポートしました。これらのハード化された DSP ブロックにより数千個の浮動小数点演算子が利用でき、インテル® Agilex™ FPGA は最大 40 TFLOPS (FP16 または BF16)、または最大 20 TFLOPS (FP32) の DSP 性能を実現します。³

チップレット・ベースのアーキテクチャーによる開発の利点

従来、FPGA では、トランシーバー、メモリー、I/O などの個々の機能をプログラマブル・ロジックと同じダイに追加して、機能をモノリシックに集積していました。ただし、この方法には3つの制限があります。第1に、FPGA 上の個々の機能をすべて同じプロセスで開発する必要があります。第2に、個々の機能のいずれかが更新または改善されるたびに新しいテープアウトが必要になります。そして第3に、個々の機能のリソースのさまざまな組み合わせや比率を得るために（同じ量のロジックでのトランシーバー数の増減など）、新しいテープアウトが必要になります。テープアウトには時間とコストがかかるため、製品イノベーションのスピードを加速するには、テープアウトを最小限に抑える必要があります。

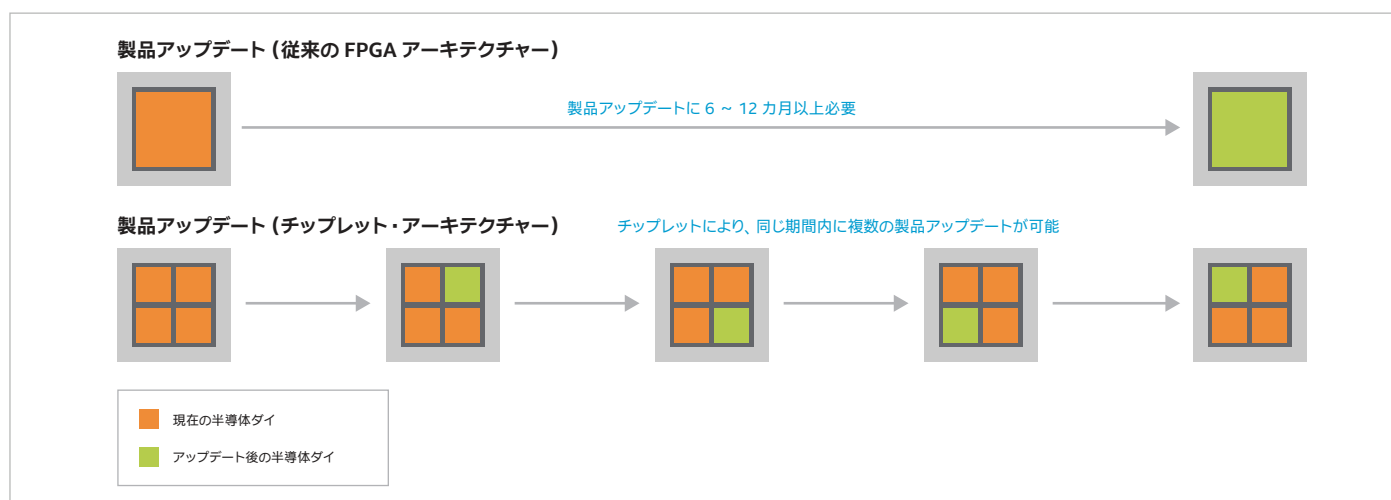


図8. 従来のFPGAアップデートのスケジュールとチップレット・ベースのアップデート

業界の共通認識では「小さいほど良い」とされるため、14nm、10nmと微細化が進んでいます。ただし、一部の機能において必ずしも最新の半導体プロセスを必要としないケースも存在します。例えば、一部のアナログ機能とメモリー機能では、汎用ロジックと同様にスケールアップされません。また、一部の従来のシリコン製品のIP機能は古いプロセスノードに基づいて開発された可能性があり、新しいノードに移行しても、移行に伴うコストと時間に見合うメリットはありません。インテルのチップレット手法を採用すると、各機能をその実装に適した半導体プロセスで構築できます。チップレットを組み合わせて、市場の特定のニーズに適したソリューションを実現することが可能になります。

インテルのチップレットを使用することで、汎用性が高く、より優れたアプローチを実現できます。チップレット（あるいはタイル）は、パッケージレベルの集積とインターフェイスの標準化によって統合可能な物理IPブロックです。まるでLEGO*ブロックのように柔軟な組み合わせが可能なインフラストラクチャーを実現します。インテル® FPGAで112GトランシーバーとPCIe* Gen 5テクノロジーを早くからサポートしていることから分かるように、チップレットを使用すると、機能性と機敏性を高めて、市場投入までの時間を短縮して新しい製品を構築できます。図8に、この利点を示します。

チップレットは、各機能に最適な半導体プロセスで構築されます。チップレットはインテル® Agilix™ FPGA ファブリックと通信します。チップレットを組み合わせ、特定のアプリケーション・クラスで必要とされる組み合わせを作成できます。

製造プロセスでは、基板として機能する高度なパッケージ技術を利用して、複数のチップレットを1つのパッケージに統合できます。例えば、チップレットを使用して、複数のコア、I/O、FPGA、アナログ、ネットワークと通信、メモリー・アクセラレーターを統合することが可能になります。サードパーティー・エコシステムのテクノロジーをインテルのシリコンと組み合わせることも可能です。

すべての要素を1つのパッケージにまとめるには、それぞれ機能の異なるダイをコスト効率よく集積するテクノロジーが必要です。従来、1つのパッケージに1つのシリコンダイが使用されていました。今は、先進のインテル® テクノロジーにより、プロセスノードの異なる複数のダイを同じパッケージにまとめることができます。例えば、インテル® Agilex™ FPGA が採用している先進の3D パッケージング技術の1つが、エンベデッド・マルチダイ・インターコネクト・ブリッジ (EMIB) です。図9に示すように、EMIBはチップレットをFPGAダイと一緒に同じパッケージに集積する高性能でコスト効率の高い手段を提供します。

インテルのパッケージング・テクノロジーにおけるこれらの最新イノベーションにより、ヘテロジニアスなエレメントを1つのパッケージに集積することができます。専用シリコンで市場の特定のニーズに対応し、柔軟なコア・ファブリックでダイを結合することで、大規模に集積されたシステムが実現されます。

インテルは、FPGA ファブリックを専用シリコンと組み合わせ、特定のアプリケーション領域にターゲットを絞った製品を構築しています。結果として構築されたアーキテクチャーは、さまざまなユースケースや市場セクターのニーズを満たし、特定の領域専用の特定の機能に対応できます。

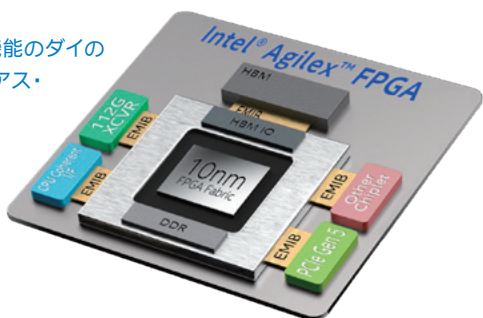


図9. 複数の特定機能のダイの柔軟なヘテロジニアス・パッケージング

Intel® Xeon® スケーラブル・プロセッサーでのプロセッサ当たりのコア数の増加やメモリとI/Oの帯域幅の大幅な拡大により、チップ内インターコネクトに対する需要の増大が性能を制限する存在となる可能性があります。

プロセッサ機能の高性能のインライン/オフロード・アクセラレーションを容易にするために、インテル® Agilex™ FPGAでは、PCIe® Gen 5やCompute Express Link (CXL) など、最新世代の高性能プロセッサ・インターフェイスをサポートしています。

Compute Express Link

インテル® Agilex™ FPGAは、Compute Express Link (CXL) をサポートします。これは、CPUとワークロード・アクセラレーター間の高性能かつ低レイテンシーなキャッシュ/メモリー・コヒーレント・インターフェイスです。CXLテクノロジーは、CPUメモリー空間と接続されたデバイスのメモリー間でメモリー・コヒーレンシーを維持することで、リソースの共有を可能にし、性能の向上、ソフトウェア・スタックの複雑さの軽減、全体的なシステムコストの削減を実現します。これにより、ユーザーはアクセラレーターの冗長なメモリー管理ハードウェアではなく、ターゲット・ワークロードだけに集中することができます。CXLの詳細については、<http://www.computeexpresslink.org/> (英語) を参照してください。

インテル® Agilex™ FPGA は、PCI Express* インターフェイスを Gen 5 までサポートします。PCIe* Gen 5 を使用すると転送速度が飛躍的に向上します。汎用性に優れたこのテクノロジーを使用するアプリケーションの中でも、PC インターコネクト、グラフィックス・アダプター、チップレベルの通信の開発においてパフォーマンスが向上します。

高度なメモリ階層

メモリの容量と帯域幅は、次世代プラットフォームの重大なボトルネックになる可能性があります。FPGAやCPUなどのコンピューティング要素は、さまざまなサイズやアクセス要件のデータタイプで動作します。そのため、処理の効率と性能を最大限に高めるためには、さまざまな種類とサイズのメモリが必要になります。

インテル® Agilex™ FPGAでは、エンベデッド・メモリ・リソース、インパッケージ・メモリ、専用インターフェイスによるオフチップメモリなど、幅広い階層のメモリリソースをサポートしています。図10に、この階層を示します。

この階層は、MLAB、ブロックRAM、eSRAMなどの内蔵オンチップメモリから始まります。これらのメモリは、さまざまな処理要件に対応するために、容量がそれぞれ異なります。

High Bandwidth Memory (HBM) は、SiPテクノロジーを使用してインテル® Agilex™ FPGA デバイスに組み込まれた次世代の高速メモリです。HBMは、他のソリューションでは実現不可能な広いメモリ帯域幅を実現します。複数のDRAM層がベースI/O層に接続され、インテル® Agilex™ FPGA デバイスに組み込まれたハード・メモリ・コントローラーに接続されて直接制御される3D高速メモリを形成します。HBMダイをFPGAパッケージに直接統合することで、ボードサイズとコストが削減され、電力要件が簡素化および削減されます。

インテル® Agilex™ FPGAでは、コア・ファブリックの隣にHBM2Eメモリを配置しています。DDRのようなディスクリート・メモリと比較して、コア・ファブリックとメモリ間のインターコネクトが大幅に

短縮され、これまで長いプリント基板(PCB)トレースに費やされていた電力が削減されます。トレースを終端していないので容量性負荷が減少し、その結果としてI/O消費電流が低減されます。最終的にシステム消費電力が削減され、消費電力当たり性能が最適化されます。さらにインテル® Agilex™ FPGAは、デバイスの外部のメモリ・コンポーネント(DDR4/5、QDR、RLDRAMや、FPGAユーザーにとって画期的な選択肢となる最新のインテル® Optane™ DC パーシステント・メモリなど)へのインターフェイスを搭載します。

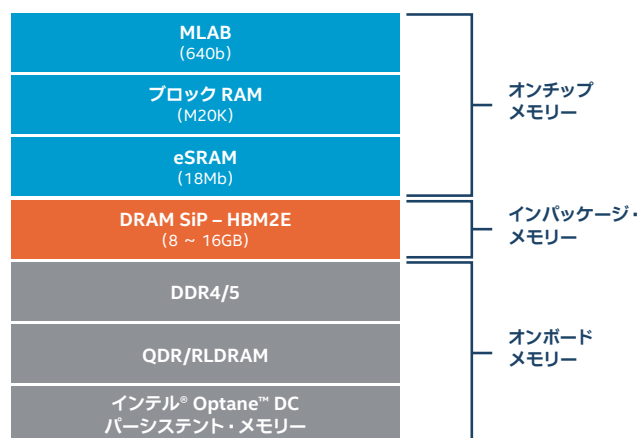


図10. インテル® Agilex™ FPGAは、オンチップ、インパッケージ、オンボードの各メモリをサポートする高度なメモリ階層をサポート

多様なテクノロジーの統合

インテル® Agilex™ FPGAは、インテル® FPGA & SoCのイノベーションを組み合わせ、かつてないカスタマイズ性と柔軟性を実現します。インテル® Agilex™ FPGAファブリックと革新的なチップレット・アーキテクチャー(単一システム内のさまざまなチップレットを1つのパッケージに統合する、インテルのEMIBパッケージング・テクノロジーによって実現)により、幅広いデバイス集積度にわたってスケーリングできる、拡張性に優れたFPGAプラットフォームが実現します。その主な機能と利点を以下に示します。

- 次世代データセンター向けのレーン当たり32GT/秒のデータレートを提供したPCIe® Gen 5 x16のサポートなど、最新世代の広帯域プロセッサ・インターフェイス・インターコネクト
- Compute Express Linkの専用サポートによりキャッシュ・コヒーレントな動作を高速化する、高性能なプロセッサ・インターフェイス
- 半精度FP16とBFLOAT16のための専用の構造とDSPブロック集積度の向上により、最大40 TFLOPS (FP16/BFLOAT16) または最大20 TFLOPS (FP32) のDSP性能を実現し、AI機能とDSP演算を高速化するアーキテクチャーを最適化
- 次世代のデータセンター/エンタープライズ/ネットワーキング環境において厳しい帯域幅要件を支える112Gbpsシリアル・トランシーバー・リンク
- コスト効率と電力効率を向上させるためのインテル® eASIC™ デバイスオプションを使用した顧客のIPの迅速な統合

- 最大40%の性能向上あるいは最大40%の消費電力削減を実現する、第2世代インテル® Hyperflex™ アーキテクチャー(インテル® Stratix® 10 FPGAとの比較)^{1,2}
- トランシーバー・タイルからプロセッサ・インターフェイスまで、専用の半導体チップレットを活用して、特定のアプリケーションや市場セグメント向けのターゲットを絞った機能を提供するチップレット統合機能
- メモリ容量の増加と、DDR5やインテル® Optane™ DC パーシステント・メモリなど、高度なメモリのサポート
- 幅広いアプリケーションに対応するシステム・インテリジェンスと柔軟な組込み機能を実現するクアッドコアA53 HPS

インテル® Agilex™ FPGAのアプリケーション開発の促進

インテル® Agilex™ FPGAの設計開発サポートは、複数のオプションによって提供されます。1つ目は、インテル® Quartus® Prime 開発ソフトウェア・ツールによるサポートです。インテル® Quartus® Prime 開発ソフトウェアは、デザインエントリ、論理合成から、最適化、検証、シミュレーションまで、インテル® FPGAとSoCのデザインに必要なすべてを備えています。数百万個のロジックエレメントを搭載し、大幅に性能向上した最新デバイスは、次世代デザイン設計の要求を満たす最適なプラットフォームを設計者に提供します。

さらにインテルは、OpenVINO™ ツールキットのインテル® ディストリビューションから始め、One APIに移行することで、ソフトウェア開発者にFPGA開発の機会を提供しています。インテルのOne API開発環境は、CPU、GPU、FPGA、AI、その他のアクセラレーターの多様なコンピューティング・エンジンのプログラミングを簡素化します。One APIには、最適化されたアプリケーション、ミドルウェア、フレームワークの最適なアクセラレーションを可能にする、ハードウェアにソフトウェアをマップするための開発ツールの包括的な統合ポートフォリオが含まれています。

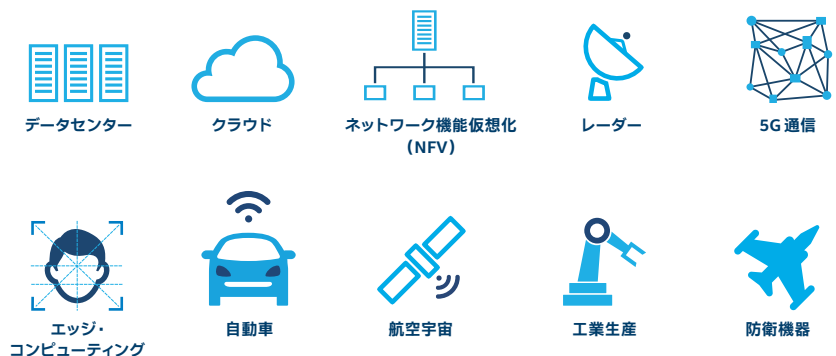
One APIにより、より多くの設計者がインテル® Agilex™ FPGAを利用できるようになります。One APIを使用すると、開発者はハードウェアの多様な選択肢を対象としてデザインのさまざまな実装を検討できるので、各実装の特定の電力機能と実行機能を理解したうえで、最終的には具体的なニーズに最適なハードウェアを選択できるようになります。

ソリューション・コンポーネント

インテル® Agilex™ FPGA
 インテル® Xeon® スケーラブル・プロセッサ
 OpenVINO™ ツールキットのインテル® ディストリビューション
 インテル® Hyperflex™ アーキテクチャー
 インテル® Quartus® Prime 開発ソフトウェア

ユースケースの例

インテル® Agilex™ FPGAは、データが生成、転送、格納、処理される場所に関係なく、エッジからコア、クラウドまで、さまざまなアプリケーションのアクセラレーションとコンピューティングのニーズをターゲットとし、特定の市場セグメントやアプリケーションに合わせたカスタマイズが可能です。



まとめ

重要な技術領域におけるインテルの深い専門知識を結集し、インテル® Agilex™ FPGA & SoCに実装されたデザインとアーキテクチャーのイノベーションは、卓越した性能、精度、柔軟性、カスタマイズ性を実現し、エッジからネットワーク、クラウドまで、さらに大きな可能性を生み出そうとしています。

詳細情報

インテル® Agilex™ FPGAの詳細については、<http://www.intel.co.jp/agilex/> を参照してください。

インテル® FPGAの詳細については、<http://www.intel.co.jp/fpga/> を参照してください。



¹ <https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/> (英語)

² インテル® Stratix® 10 FPGAと比較してパフォーマンスが最大40%向上: インテル® Quartus® Prime 開発ソフトウェアを使用して、インテル® Stratix® 10 デバイスで達成された最大動作周波数(Fmax)とインテル® Agilex™ デバイスで達成されたFmaxを比較した、サンプル・デザイン・スイートのベンチマークに基づいています。2019年2月に実施されたテストでは、インテル® Agilex™ FPGAの最速スピードグレードで動作するデザインは、インテル® Stratix® 10 デバイスの最も一般的なスピードグレード(-2スピードグレード)で動作する同じデザインと比較して、Fmaxが平均で40%向上します。

³ 最大40 TFLOPSのDSP性能(FP16コンフィグレーション): 各インテル® Agilex™ DSPブロックは、クロックサイクル当たり2つのFP16浮動小数点演算(FLOPS)を実行できます。FP16コンフィグレーションの合計FLOPSは、単一のインテル® Agilex™ FPGAで提供されるDSPブロックの最大数x2と、そのブロックに指定できる最大動作周波数を乗算して算出されました。

結果はインテル社内での分析、アーキテクチャーのシミュレーション、およびモデリングに基づくものであり、情報提供のみを目的としています。システム・ハードウェア、ソフトウェア、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。

テストは、特定システムでの特定テストにおけるコンポーネントのパフォーマンスを測定しています。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。購入を検討される場合は、ほかの情報も参考にして、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。性能や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。

性能の測定結果は2019年2月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セキュリティ・アップデートが適用されているとは限りません。詳細については、公開されている構成情報を参照してください。絶対的なセキュリティを提供できる製品はありません。

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。ベンチマークと性能テストの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。

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コンピューター・システムはありません。詳細については、各システムメーカーまたは販売店にお問い合わせいただくか、<http://www.intel.co.jp/fpga> 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。

Intel, インテル, Intel ロゴ, Agilex, eASIC, Hyperflex, Intel Optane, OpenVINO, Quartus, Stratix, Xeon は、アメリカ合衆国および/またはその他の国における 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。