arm

ホワイトペーパー

Arm Total Compute

未来のワークロードに対応する設計

目次

1 はじめに

モバイルデバイスは、多くの人に限りない楽しみ、生産性、成功をもたらすコンテンツ制作プラットフォームとなりました。片手に収まるデバイスで、迫力あるデジタル体験を作ることも視聴することもでき、そのために必要な処理能力はますます増加しています。高い性能、インテリジェンス、視覚的/感覚的操作に対する需要もとどまるところを知りません。

スマートフォン、タブレット、ノート PC、ウェアラブルデバイス、スマートホームデバイスの処理能力に対する需要がこれほど高く、複雑だったことはありません。まもなく XR（拡張現実 [AR] と仮想現実 [VR]）などの技術が普及すれば、新しい体験を提供するために、全体的なパフォーマンス、ワット性能、効率の大幅な引き上げが必要となるでしょう。しかも従来どおりの限られた発熱量でです。

2 Total Compute について

Arm の IP は、リリースごとに低コストデバイスのパフォーマンスを引き上げてきました。しかしハード面の限界はあります。デバイスの発熱と消費電力による壁は越えることができません。消費者がますます高度でイマーシブなスマートデバイス体験を求める現在、コンピューティングシステムの設計も変わる必要があります。Arm Total Compute は、ソリューションに重点を置いた総合的な SoC（システムオンチップ）設計アプローチを採用し、個別の IP ではなく、システム全体を設計および最適化します。

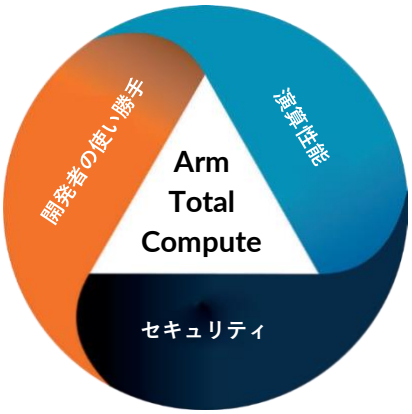
Total Compute は、IP に対する Arm の思考と設計方針の転換です。未来のデバイスのユーザーが、インテリジェントな AI を活用し、最高の画像や映像を撮影、作成、視聴できるようにするものです。パフォーマンスの向上のほか、システム全体にわたる多層型ソリューションを通じて実際のデバイスからクラウドサービスまでのセキュリティ強化を可能にし、エンドツーエンドで個人データを保護します。

そして根本的には、開発者の使い勝手を高め、デバイスのパフォーマンスとセキュリティ機能の活用を容易にすることで、複数のコンピューティングプラットフォームにわたる効率と有効性に優れた高度なソリューションの設計と構築に貢献し、未来の市場やデバイスのニーズを満たします。

この**演算性能、セキュリティ、開発者の使い勝手**という 3 本の柱が Total Compute を支えています。

3 デジタルイマージョン

現在の消費者向けデバイスは、コミュニケーション、買い物、銀行取引などの日常作業から動画のストリーミング、ゲーム、XR などの複雑なワークロードまで、人々のすべての行動の中心です。Arm はこのような体験を「デジタルイマージョン」と呼んでいます。



「コンピューティングの第 5 波」、すなわち 5G、AI、IoT（モノのインターネット）内の重要な技術が成熟するにつれ、これらはさらに高度かつ安全で充実したものになるでしょう。

デジタルイマージョンは演算性能と通信帯域幅の不断の向上、そしてデータのセキュリティを守る消費者への責任に依存しています。同時に、形状やバッテリー寿命などの物理的制約にも対処が必要です。

現在、デジタルイマージョン体験を左右しているのはスマートフォンですが、他の消費者向けデバイスの機能やニーズによる影響も常に受けています。業界は多様な新しいユースケースや消費者向けの体験を生み出しています。

3.1 多様な新しいユースケース

複雑なユースケースに向かう動きは、ノート PC、タブレットからスマートホームや車載プラットフォームまで、さまざまな市場のさまざまなデバイスに影響を与えています。たとえば Netflix、Amazon プライム、Disney+などのストリーミングサービスが主なコンテンツ配信経路として家庭に普及したことで、スマート TV の性能は 10 年前と比べて何十倍にも向上しました。

そして現在のスマート TV の画面の裏に強力な Arm プロセッサーが詰め込まれているにもかかわらず、複雑なユースケースは際限なく生まれてきます。消費者が高い解像度（8K 以上）、高いフレームレート、それにクラウドゲームなどのまったく新しい強度のユースケースを求めるにつれ、次世代のスマート TV はさらに高い演算性能を必要とすると予想されます。そのうえ多くのテレビメーカーは AI を使用して体験を最適化している（アップスケーリングなどの画質向上や音声アシスタント）ため、今後のスマート TV には AI ワークロードを処理できるコンピューティングシステムが必要です。

AI に対応することはすでに多くのスマートデバイスの必須要件です。たとえば計算写真学は機械学習（ML）を使用して人、物体、風景を特定しますし、画質の最適化やカメラのズームにも使用されています。背景をぼかす場合も背景をすべて取り除く場合も、デュアルカメラからデプスデータを生成し、リアルタイムの修正や認識を可能にしています。今やスマートフォンでは普通のことですが、産業用 IoT の動画センサーや動画による赤ちゃんモニターにも採用されつつあります。

「Arm テクノロジーを前提としたスマートフォンエコシステムのアプリ数は 2008 年には約 500、2021 年には 350 万に成長しました。」

AR は消費者にはまだそれほど馴染みがありませんが、小売、デザイン、製造、メンテナンス、医療トレーニング、建設などのビジネスでは人気が高まっています。AR はリアルな 3D 物体をリアルタイムで描画し、カメラとジャイロスコープデータを ML アルゴリズムと併用してその仮想物体を物理環境に置きます。

AR/VR 体験を提供する新しい消費者向けウェアラブル機器が市場化され、デジタルイマージョンの新しい波が始まれば、消費者の関心も高まると予想されます。VR は今後も基本的にゲーミングプラットフォームになると思われますが、AR は見るもの、することすべてにコンテキストを追加し、常時オンの体験を提供する可能性を秘めています。しかし、AR と VR の根本的な課題は、長時間快適に装用できるような小型で軽量のデバイス（未来のスマートグラスなど）に、高い演算性能と適切な電源を詰め込むことです。

もっと小型のウェアラブル（スマートウォッチなど）について言えば、現在のフィットネス用小型機器を充実した医療機器に進化させるなどの可能性もあります。

これをさらに後押しするのが、まったく新しいコネクティビティ、つまり 5G です。ネットワーク速度の向上は、ウェブ閲覧や動画ストリーミングなど既存のユースケースや体験の速度と利便性をはるかに高め、レイテンシを低減します。しかしデバイスで収集されるデータと情報が増えるにつれ、未来のワークロードの複雑性と処理量はさらに増えます。

デジタルイマージョンの成長の一部として、スマートフォン開発者のエコシステムにも注目することが重要です。昨年、ストリーミングやモバイルゲームの主流デバイスとしてスマートフォンの市場規模は 920 億ドルに達し、モバイルアプリの売上も 4,300 億ドルを超えました。デバイスで使用される技術（生成 AI など）も高度化しています。

このような多くのアプリケーションにより多様なデジタルデータが 2 年ごとに倍増しています。

スマートフォンユーザーがアクティブになるにつれ、高いパフォーマンスを必要とするアプリとデータが増えます。また、ユーザーは敏感であり、シームレスで包括的な体験を求めています。ユーザーはサービスの選択肢を求め、そのサービスをスマートフォン以外のデバイスでも利用したいと考えます。個人データ証跡の知識と親密性が高まるにつれ、消費者の好みや関心に応じたターゲティングが容易となります。つまり現在の「データの洪水」は、個人の意思決定プロセスを幅広く左右し、社会的な影響が大きいという意味です。データとアプリケーションの両方を保護するシステム全体のセキュリティ基盤が重要となるのはこのためです。デバイスに対する信用不足は将来のデジタルイマージョン体験を阻害することになるでしょう。

3.2 デジタルイマージョンのニーズを満たす

ムーアの法則のスピードが落ちている現在、CPU、GPU、NPU など、SoC を構成する演算コンポーネントを最適化し続けるだけでは追いつきません。新しいワークロードはますます複雑になり、最善のユーザー体験を提供するには最適化されたコンピューティングシステムを必要とします。演算コンポーネントを隔離して設計するだけでは、このような新しい複雑なワークロードに対応できません。一歩進んで SoC をシステムレベルのソリューションの観点から考え、デジタルイマージョンのユースケースと体験を設計の中心に据える必要があります。未来のコンピューティングプラットフォームは 1 つの効率的なコンピューティングシステムとして最善のユーザー体験を提供する必要があります。たとえば Arm は 1 つのシステム内で Immortalis や Mali GPU と Cortex-A CPU の協調性を高めることで、バッテリー寿命を最大限に維持しつつ、AAA ゲームを快適にプレイできるような技術を設計しています。

Arm の IP は、リリースごとに低コストデバイスのパフォーマンスを引き上げてきました。しかし AI が広く普及し、ユースケースが複雑化するとともに、アプリに AI や ML のワークロードを統合する開発者も増加しています。

これはコンピューティング IP だけの問題ではありません。エンジニアや開発者は高性能のセキュリティ機能、そして簡単に実装できてエコシステム全体で利用できるソフトウェアやツールを必要とします。セキュリティとは、ハードウェア、ソフトウェア、オペレーティングシステム、アプリケーション、サービスを守るものを意味します。一方、開発者は、すぐに高性能が得られ、問題があれば一貫した方法で解決できることを望みます。つまり、演算機能やセキュリティ機能にアクセスしやすく、開発プロセスをスピードアップする高性能のソフトウェアやツールをすぐに利用できるということです。

この結果、ハードウェアからソフトウェアまで SoC 開発フレームワーク全体に、一貫性、拡張性、信頼性、セキュリティ、安定性を備えたテクノロジースタックが求められます。これには、エンジニアや開発者がソリューションベースの異種のコンピューティング IP ブロックを活用し、縛られることなくシステムやエンドアプリケーション全体を検討できるよう、エコシステム全体における SoC 設計アプローチの転換が必要です。

新しい、複雑なユースケースを実現し、デジタルイマージョンの成長に対応するには、この SoC 設計上の課題に正面から取り組まねばなりません。これがデスクトップコンピューターなら、強力なプロセッサー、グラフィックスカード、冷却システム、強力な電源を追加するかもしれません。しかしモバイルデバイスではそう簡単にいきません。小さな軽量デバイスの消費電力の制約に対応するには、新しいアプローチが必要です。

Arm ではこれを Total Compute と呼びます。

4 システム設計へのアプローチ転換

これまで Arm は、シリコンパートナーにそれぞれの複雑な SoC（システムオンチップ）に組み込んでいただけるよう、各 IP「コンポーネント」の開発と最適化に力を尽くしてきました。

しかし今後、各コンピューティングコンポーネントを隔離して設計し、最適化するだけでは十分ではありません。ソリューションに重点を置いた総合的な SoC 設計アプローチを採用し、個別の IP ではなく、システム全体を設計および最適化する必要があります。

Total Compute は、特定のユースケースに最適化した総合的な SoC ソリューションを提供する Arm の方向性を示しています。

この方法なら、次世代のデバイスでクラス最高のパフォーマンスと効率を実現することができます。基本的なアーキテクチャは共通のため、パートナー各社もソフトウェアエコシステム全体も製品を市場化しやすくなり、最先端のデジタルイマージョン体験を想定したパフォーマンスを確保できます。

SoC 設計における多大な課題については強く認識していますが、AR およびそれが実現するユースケースや体験に重点を置くことは、Total Compute が実際にどのように作用するのかを示す良い例だと思います。CPU は電力効率の高い形でパフォーマンスを引き上げます。GPU はグラフィックスを処理します。AI はユーザーの位置、具体的な物体、建物などの検出に使用されます。次にこの IP をまとめてシステム内でシームレスに協調させる必要があります。低消費電力の制約を満たす優れたシステムを構築するために、インターコネクトやコントローラなどのシステム IP が必要なのはこのためです。こうして初めて SoC 設計者がシステムソリューションを最適化し、多くの時間を差別化、新しいアイデアの開発、技術的改善、顧客のためのソリューションに割くことが可能となります。

パフォーマンスの向上と開発者の使い勝手に加え、セキュアに体験を提供することも必要です。デバイス本体だけでなく、デバイスのエコシステム全体でです。Arm は、デバイス上のソリューションからデバイスエコシステム全体のサポートまで、Total Compute のあらゆる側面にセキュリティを組み込んでいます。これはエンドポイントデバイス（ネットワークのエッジにある物理的デバイス）のセキュリティに対する Arm の取り組みを前進させたものであり、ハードウェア、ファームウェア、ソフトウェア、オペレーティングシステム、アプリケーション、サービスへの「多層防御」セキュリティを意味します。

強力な基盤

あらゆる意味でこれはアプローチの転換ですが、Arm にとってゼロから始めることではありません。Arm には、電力制御フレームワークおよびマイクロコントローラ制御の電力管理スキームなど、ユースケースや消費電力に応じて最適なパフォーマンスを提供するために長年培ったテクノロジーがあります。特に厳しいユースケースについては給電にも注目しています。

シリコンに実装する前の SoC 設計の不可欠な部分として、フィジカル設計があります。Arm はさまざまなフィジカル IP を提案しています。中でも Arm POP IP は Cortex-A CPU とシリコンプロセステクノロジーの橋渡しです。POP IP は、最新プロセスノードでの業界をリードする PPA（性能、消費電力、実装面積）を実現します。メモリとロジックは消費電力の制限内で Arm IP に最適化され、Total Compute の演算性能の柱に貢献します。

「システム内のボトルネックを解消するとともに、全 IP 境界（CPU、GPU、NPU、メモリなど）にわたってシステムレベルで最適化することが必須です。」

POP IP とは、製品化期間を短縮し、技術的リスクやプロジェクトリスクを抑えるためのコアハードニング用アクセラレーション・テクノロジーです。この IP の境界を越えた最適化が、デバイスでのエッジコンピューティングに対するニーズの変化の中で、未来のワークロードにどう対応すべきかを示しています。

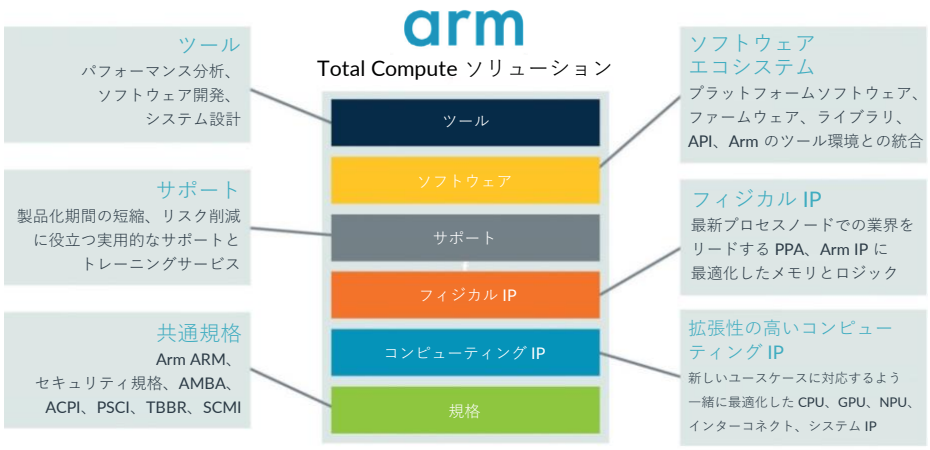
ソリューション重点型のアプローチを採用する理由

- 複雑なユースケースには高いパフォーマンスが必要ですが、それにはすべての IP がシームレスに協調することが前提となります。IP レベルでパフォーマンスを上げるだけでは足りません。システム内のボトルネックを解消するとともに、全 IP 境界（CPU、GPU、NPU、メモリなど）にわたってシステムレベルで最適化することが必須です。これによりソリューション全体が好循環を生み、同じ消費電力で複雑なユースケースに必要なパフォーマンスを提供し、持続します。まさに全体（ソリューション）は部分（IP）の総和に勝る、です。多様な高性能 IP コンポーネントを結ぶインターコネクトや帯域幅の要件がレイテンシの問題を引き起こし、SoC 自体がボトルネックとなる場合もあります。
- セキュリティやデータプライバシーのニーズが高まる中、エンドツーエンドの総合セキュリティ機能による一貫したセキュリティアプローチを採用する上で、多様な IP コンポーネントで構成する SoC は設計上の課題を生じる場合があります。
- ソリューション重点型のアプローチは、使いやすくて高性能の IP とソフトウェア/ツールを開発者に提供し、スムーズでスピーディーな開発を可能にするとともに、消費者にはイマーシブなアプリケーションを提供します。
- ソフトウェア開発者にとって、供給元の異なる IP コンポーネントを使用した複数のハードウェアアーキテクチャに対処するのは非常に時間とコストのかかることであり、ソフトウェアエコシステムの断片化の助長にもなります。

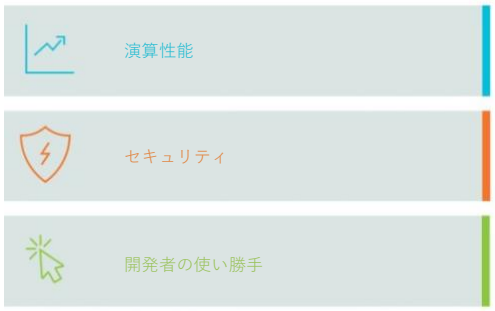
しかし、デバイスが複雑化するにつれ、セキュリティ要件も厳しくなります。セキュアなデバイスを実現するには、セキュリティが二の次にならないようなハードウェアとソフトウェアの共生関係が必要です。この土台となるのが共同エコシステム内で充実しつつあるセキュア IP とソフトウェアのポートフォリオです。

Total Compute の主な要素

- 新しいユースケースのニーズに対応し、協調動作に最適化したコンピューティング IP。
- 共通規格。エコシステム全体で一貫した導入しやすいテクノロジー。
- ハードウェアと共同設計され、面倒な設定なくデバイス上でテクノロジーを即座に活用できるソフトウェアコンポーネント。
- システム設計、ソフトウェア開発、パフォーマンス最適化をサポートするツール。
- プロセスノードと Arm IP の複雑化にもかかわらず、競争力の高い PPA を提供し、シリコンの製品化期間を短縮する Arm の専門知識を活用したフィジカル IP（POP コンポーネントを含む）。
- エンジニアリングチームを最新のソリューションやテクノロジーに早く慣れさせ、リスクや製品化期間を抑えるサポートとトレーニング。



この IP の境界を越えた最適化が、デバイスでのエッジコンピューティングに対するニーズの変化の中で、未来のワークロードにどう対応すべきかを示しています。



5 Total Compute の 3 本柱

5.1 演算性能

ソリューション重点型のアプローチで性能向上を促進

ムーアの法則がフィジカル設計におけるパフォーマンスの壁に突き当たる一方、パフォーマンスの急速な向上へのニーズは高まっています。新しい複雑なワークロードは同じ限られた発熱量で高性能のハードウェアを必要とします。

多様な IP コンポーネントを SoC に統合する際の課題は、アクティブなダイ面積の拡大が発熱量や消費電力の増大につながることです。システム全体に重点を置き、パフォーマンス、効率、データ交換を考慮した共通の基本アーキテクチャで各 IP ブロックを開発する必要があるのはこのためです。

Total Compute の目的は、SoC 設計に包括的なシステムレベルのアプローチを採用することにより、クラス最高の電力効率と最小の実装面積で性能を向上させることです。綿密な電力管理、システムキャッシュの使用、そしてもちろんセキュリティやプライバシー機能の実装により、性能および帯域幅向上への需要に対応します。

各 IP の最適化ではなく SoC 全体をシステムレベルのソリューションとして捉えることは、次世代デバイスでのユースケースや体験を重視し、システム全体のシームレスな協調によって最大限のパフォーマンスと効率を提供することにつながります。

では、ユースケースに重点を置いた SoC システムの開発と最適化に、どう着手すれば良いのでしょうか？Total Compute 開発サイクルの第一歩はユースケースの選択と分析です。次に、それらを主なワークロードに分解します。ユースケースには、セキュリティ、ML、あるいは所定の消費電力内における特定アプリケーション（8K、AAA ゲーム、高リフレッシュレートなど）のコンピューティング要件などが挙げられます。この分析結果から IP のコンピューティングドメインやソフトウェアフレームワークにわたるアーキテクチャを選択します。ワークロードは、ユーザー体験、プラットフォーム（GPU、CPU、メモリ帯域幅など、各サブシステム項目のパフォーマンス対効率）、テクノロジー（キャッシュなど）の要件分析を可能にします。

Total Compute システム設計で欠かせないのは、異なる IP コンポーネントやコンピューティングドメイン間でどのように連携データとコンピューティングを配分するかの分析です。複数の IP 間でのデータ圧縮の統合が成功への鍵となります。

次に、システムの評価と分析を通じて、システムレベルでハードウェアとソフトウェアでこれらを最適化します。明日のワークロードをサポートし、しかも昨日のベンチマークではない強力かつ効率的なマイクロアーキテクチャも開発します。

データ圧縮の適切な使用とともに、メモリ帯域幅使用の削減が、より処理量の多いユースケースを可能にし、消費電力を引き下げます。たとえば圧縮は外部のメモリボトルネックを克服し、動画の解像度やフレームレートを向上させる可能性があります。これにより、同じワークロードの消費電力削減、あるいは外部メモリ帯域幅の狭い低価格のデバイスでの機能実装も実現します。

たとえばモバイルゲームを考えてみましょう。IP コンポーネントをいったんシステムに組み込んだ後、各コンポーネントの合成ベンチマークの改善を有意義なパフォーマンスや効率の改善に結び付けることは非常に困難です。Total Compute を中心とした高性能ゲームワークロードのシステム分析では、コンピューティングプラットフォームの新しい基準を特定することが可能です。たとえば特定の持続的な消費電力で所定の 1 秒あたりのフレーム数（FPS）を達成することです。

これで全体的なゲーム体験の改善に向けて、各コンピューティングコンポーネントに必要な新しい機能や特長に注目することができます。簡単に言えば、トップダウン式にシステムを見つめ、コンピューティング IP コンポーネントを改善することが、Total Compute アプローチのポイントです。

どのようなユースケースでも、システムパフォーマンスの向上は、固定されたベンチマークではなく、ユースケースやその要件によって左右される多面的な基準です。新しい機能やパフォーマンスの向上は、ゲームの持続性能、柔軟で拡張性の高い ML 機能、スマーフォンレベルの電力効率でのノート PC クラスのパフォーマンスなど、具体的なコンテキスト内で評価されます。以下に例を挙げます。

- 拡張性の高い最適なヘテロジニアスコンピューティング、スマートフォンの電力でノート PC クラスのパフォーマンス
- 画像処理の向上を通じたイマーシブで相互作用性の高いユーザー体験
- ゲームの持続性能
- 複数のコンピューティングクラスターにわたる高性能で柔軟な ML 機能

エッジコンピューティングに対するニーズが変化し、デバイスが複雑化するにつれ、セキュリティの要件も変わります。セキュアなデバイスを実現するには、セキュリティが二の次にならないようなハードウェアとソフトウェアの共生関係が必要です。この土台となるのが共同エコシステム内で充実しつつあるセキュア IP とソフトウェアのポートフォリオです。

5.2 セキュリティ

ハードウェアとソフトウェアでセキュアな基盤を構築

Arm は過去 20 年間、スマートデバイスのセキュリティにおいて最先端を走っています。Arm TrustZone テクノロジーは現在、世界で何十億台ものモバイルデバイスに搭載されています。スマートウォッチ、DTV、コネクテッドホームデバイス、そして次世代のノート PC にも広がっています。

このようなデバイスで最も便利なアプリの多くは個人データに依存しています。ユーザーを「認識」することで、次に何の言葉を入力するかを予測する、顔認識でセキュリティを確保する、予定を思い出させるなど、デバイスははるかに便利になります。しかし、ユーザーがデバイスに入れる個人情報が増えるにつれ、それを狙う者のチャンスは増えます。

Arm は Total Compute によってセキュリティとプライバシーの両方の問題に対処したいと考えています。メーカーやアプリケーション開発者が消費者の個人情報の問題に対処し、消費者の信頼を獲得、そして維持するには、適切なセキュリティアーキテクチャと堅牢な実装が必須です。

1. デバイスメーカーは、シリコンからファームウェアまでクラス最高のセキュリティ基盤を実装することで、攻撃者による脆弱性の悪用を防ぐことができます。したがって機密データ、財務情報、その他の一般的な個人情報が安全に保管されているという消費者の信頼を得られます。
2. 消費者は当然ながら自身のデータが無断で使用されることを心配しているため、個人データの処理をデバイス上で保護する必要があります。つまりセキュアな処理環境で保護し、データの収集とリークを最小限にするという意味です。
3. ビッグデータと AI が至るところで統合されている現在、新しいセキュリティ問題も検討する必要があります。

Total Compute は、個人のプライバシー保護権を損なうことなく、AI の大きなメリットをパートナー各社にさまざまな分野で活用していただけるよう、セキュリティの基盤構築を目指しています。パートナー各社がデジタルイマージョン体験においてユーザーのデータと ID を保護できるよう、導入の簡単なセキュリティ機能を提供することが Arm のソリューションのポイントです。

この結果、Arm は高価で多様なセキュリティソリューションを離れ、あらゆる市場区分にわたって問題やニーズの多様化に対処できる標準化された拡張可能なソリューションへとセキュリティアーキテクチャを見直しています。

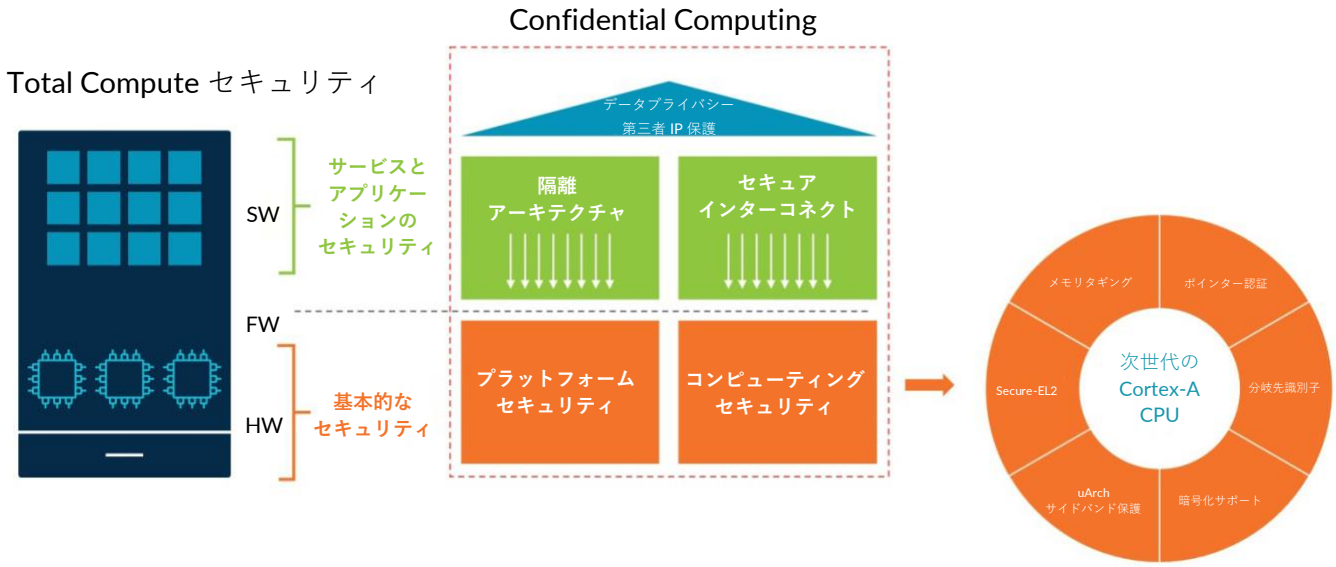
堅牢なデバイスセキュリティは Arm とそのパートナーの成功に不可欠です。消費者向けデバイスの中身は、ハードウェアコンポーネント、ハードウェアコンポーネントを実行するファームウェア、何百万行ものソフトウェアコードです。これらはすべて、さまざまなハッキングを受ける可能性があり、セキュリティを確保する必要があります。組み合わせや実装方法は無数にあり、場合によっては断片化しているため、短時間で幅広いセキュリティパッチを適用するのは困難です。セキュリティの穴を見つけ、塞ぐという終わりのない作業になります。

ユーザーに安心してさまざまなデバイスに個人データを預けてもらうには、セキュリティを二の次ではなく最優先事項とする必要があります。2020 年、Arm の委託で Northstar Research[1] が実施した調査によれば、もし使用しているデバイスが侵害を受け、個人データが詐取された場合、回答者の 3 分の 1 以上が永久にその種類のデバイスの使用を止めるとしています。さらに 3 分の 1 は別のブランドに乗り換えると回答しました。Arm は、消費者の期待に応えるだけでなく業界への責任を負っています。Arm は優れた統合型セキュリティの提供を促進するテクノロジーを構築し、IP レベルだけでなく、ファームウェア、プラットフォームソフトウェア、アプリケーションレベルでも実装を容易にするよう努めています。

Total Compute は、このセキュリティへの取り組みをさらに前進させ、新しいアプローチの基盤となります。ハードウェア、ソフトウェア、オペレーティングシステム、アプリケーション、サービスにわたる多層防御を実現し、エンドツーエンドのシステムレベルのセキュリティ保護を根本的に確保することによって、個人データおよびそれを扱うアプリケーションを保護するのです。あらゆるコンピューターの基本設計の基礎であるアーキテクチャからセキュリティに着手するという意味です。Arm の目的は、大量生産の前にデバイス内の脆弱性を緩和し、消費者向けデバイスの攻撃経路を大幅に削減することです。現実的に言えばセキュリティの 2 つの重要な側面に集中します。

- ハードウェア（HW）とファームウェア（FW）に対応する基本セキュリティ
- ソフトウェア（SW）に対応するアプリケーションとサービスのセキュリティ

次世代の消費者向けデバイスが複雑化するにつれ、セキュリティはデバイスの 1 つの側面だけを保護するスタンドアロンのソリューションでは終わらなくなります。デバイスハードウェアからデバイス上の多様な個人データを利用するクラウドサービスまで、システムが協調する多層型ソリューションが必要です。Total Compute は、個別の IP コンポーネントを保護するソリューションではなく、システム全体を保護する包括的なソリューションでセキュリティを確保します。ここで重要や役割を果たすのが Confidential Computing[2] です。これはセキュアコンピューティングを進化させ、エコシステムのあらゆる面で攻撃への脆弱性を軽減します。



「バグの 70%がメモリ安全性のエラーです。」

Arm にとって Confidential Computing とは、セキュリティアーキテクチャからアプリケーションソフトウェアや消費者が自分のデバイスで使用するサービスまですべてに対応し、多層型防御を実現するものです。

メモリタギング拡張（MTE）

ハードウェアレベルでは、Arm の革新的なセキュリティ機能、たとえばメモリタギング拡張（MTE）がメモリサブシステムに生じるセキュリティの脆弱性を軽減します。これが必要なのは、ソフトウェアベンダーが製品におけるセキュリティ問題の大半についてメモリ安全性の侵害による脆弱性を報告しているためです。事実、バグの 70%がメモリ安全性のエラーです。

MTE はエコシステム全体にわたってメモリ安全性に対する侵害を容易かつ効率的にします。まず MTE は、シリコンベンダーが SoC を製造受託会社に送る前にこのタイプのバグに対処します。次にデバイスの量産開始前に製造受託会社が他のメモリ安全性バグを検出するのを容易にします。ハードウェアが提供されなくても、HWASAN のような Android のツールはコードチェックをサポートします。製品が市場に出た後、OSV やアプリケーション開発者は MTE 対応デバイスを使用して自社コードにおけるバッファオーバーフローやヒープコラプション（メモリ破損）を把握できます。

Google はすでに将来のハードウェアを見越し、Android 11 以降での MTE サポートを表明しています。Google は Android Source のウェブサイトで、Android 11 以降での 64 ビット処理に関して、Arm Top-byte Ignore（TBI）をカーネルサポートするデバイスにおいて、すべてのヒープ割り当てでポインターの最上位バイトに実装定義のタグを設定すると公表しています。TBI とは、すべての Armv8 AArch64 ハードウェアの 64 ビットコードに対応する機能であり、ハードウェアがメモリにアクセスする際にポインターの上位バイトを無視することを意味します。これは MTE をサポートするハードウェアに必須です。

さらに、多くの開発者に MTE を利用してもらうため、Google は Android 13 にブートローダースイッチを実装しました。これで販売済みのデバイスでも MTE を利用できます。デバイスメーカーは MTE のハードウェアサポートを持つデバイスを出荷し、開発者や上級ユーザーは MTE がデフォルトでオンでなくてもオンに切り替えて利用できます。[5]をご覧ください。

相互運用可能なデバイスアテステーション規格

しかし、セキュリティにおける Total Compute の役割はハードウェアだけに終わりません。そこで社外の業界コンソーシアムとの協力が重要となります。たとえば Arm は現在、Confidential Compute Consortium とともに相互運用可能なデバイスアテステーション規格の策定に取り組んでいます。規格を通じてデバイスとファームウェアの信頼性と信用の証拠を取得し、検証することで、開発者はアプリケーションを実行する前にプラットフォームへの信用を確立できます。ソフトウェア開発者にとって、これはセキュリティのレベルアップであるだけでなく、ネットワーク経由（OTA）でのセキュアなアプリケーション更新を大幅に容易にし、エンジニアリングコストを削減します。

さらに Arm は、ユーザーの個人データやビジネス資産を保護することで開発者にセキュアで安全な自由スペースを提供するよう努めています。ユーザーに安心してさまざまなデバイスに個人データを預けてもらうには、セキュリティを二の次ではなく最優先事項とする必要があります。保存または転送されるデータのセキュリティ保護は、暗号化とプロトコル設計を通じて一般に理解され、広く実装されています。しかし、処理中または使用中のデータのセキュリティ保護についてはあまり理解されていないため、大きなリスクになる可能性があります。Total Compute はセキュアで開発者にとって使い勝手の良いプラットフォームであり、アプリとアプリケーションが使用する個人データの両方を保護します。

このようなアプローチにより Arm とそのパートナーは、消費者データの保護に関する未来の新しいニーズに対応し、なおかつデータの分析や理解も可能となるでしょう。デジタルイマージョンのユースケースでは、データや ML アルゴリズムを第三者に公開することなく、複数の当事者がセンシティブなデータを組み合わせ、分析し、学習に役立てる必要があります。Total Compute が対処するセキュリティ問題は一般に秘密計算、分散学習、プライバシー保護分析などと呼ばれますが、これはつまり、データ漏洩やデータ詐取のリスクを最小化しながら第三者と協力して演算を実行する Confidential Computing 環境を提供することです。これらは Total Compute をセキュリティの基盤として新しいイマーシブな体験を創出しようとする開発者の新しいニーズでもあります。

5.3 開発者の使い勝手

さらに高性能のソフトウェアやツールを開発者に使いやすく

世界には開発者が 2,300 万人[3]いて、全員が優れた体験の提供に力を尽くしています。すでに Arm は、複数の Arm テクノロジーを搭載したコンピューティングシステム全体で、すべてが順調に、スピーディーに、効率的に、一貫して、セキュアに実行されるよう、開発者を支援するソフトウェアとツールに大きく投資しています。これにより開発者は IP を検討、選択、確認し、パフォーマンスの最適化や分析を実行できます。Total Compute もこの取り組みをさらに進めるものです。

Total Compute の 3 本目の柱として注目されるのは開発者の使い勝手です。つまり、プラットフォームに組み込まれた各種機能を活用するためのツール、情報、開発者への支援がなければ、前述の 2 本の柱も実現しないということです。

たとえば Android のエコシステムを見てみましょう。Android エコシステムの多様性は最大の強みですが、一貫した優れた体験を幅広いリーチで提供したい開発者には大きな課題でもあります。2020 年半ばの時点で、Geekbench 5 スコアを公開している Android デバイスは 440 種類あり、開発者は多数の構成、形状、パフォーマンスを考慮する必要があります。パフォーマンスと効率に関して、Arm は 2 つの明確なメッセージを開発者から受け取っています。

1. できるだけ多くのデバイスで即座に高いパフォーマンスを使えるようにしてほしい。
2. 最大限のパフォーマンスを得るプロセスをできるだけシンプルで一貫したものにしてほしい。

Arm エコシステム内の開発者はこれまで、SoC のコンポーネントそれぞれに対応するツールについて学習する必要がありました。IP の組み合わせの異なる複数の SoC を開発する場合、これが時間とコストのかかる複雑な作業となります。さらに SoC に新しい IP コンポーネントと旧い IP コンポーネントが混在すると、コードの断片化によって多くの問題が生じます。

Total Compute は開発者にとって、SoC の包括的なプログラミング、デバッグ、分析の枠組みとなります。

ツールの改善はサポートの改善

Total Compute を通じて開発者の使い勝手を改善する方法は主に 2 つあります。
1 つは Arm ツールの改善です。もう 1 つは開発者向けツールチェーンにおける Arm 製品のサポート改善です。最終的な目的は、最新のセキュリティ機能を備えた複数のプラットフォームへの導入を容易にし、出荷前テスト、現場での機能向上、デバッグのスピードアップによってパフォーマンスを最大限に高めることです。

Arm はすでに開発者を支援するソフトウェアとツールに大きく投資しています。開発者による IP の検討、選択、確認、そしてパフォーマンスの最適化や分析を容易にするためです。Total Compute は、その取り組みをさらに一歩進め、Arm テクノロジーを使いやすくするとともに、開発者が各アーキテクチャ、あるいは使用する IP ブロックそれぞれに最適なツールを探し回る手間を省きます。IP の組み合わせの異なる複数の SoC を開発する場合、これが時間とコストのかかる複雑な作業となります。SoC に新しい IP ブロックと旧い IP ブロックが混在すれば、コードの断片化によって多くの問題が生じます。

断片化の解消

Trusted Firmware とは、対応の Arm 仕様に準拠する信頼性の高いリファレンスコードベースを SoC 開発者やデバイス製造企業に提供するオープンアプローチです。Total Compute を通じてできるだけ多くの企業にこれが採用され、ソフトウェアエコシステムの断片化が緩和されれば、オペレーティングシステム、ランタイム、アプリプラットフォームがすべて Arm で最適に動作するようになります。さらにパートナー各社の製品化期間が短縮され、システム全体のセキュリティとパフォーマンスが改善されることで、パートナー各社にとってのメリットも拡大します。

ツールとサポート

Arm のツールと他のサポートにより、開発者は異なる IP 上での各ワークロードの機能とシステム内でボトルネックが生じる場所だけでなく、ソリューションを効率的に実装する方法も理解することができます。Total Compute を通じて開発者は、生産性向上アプリからゲームや ML まで幅広いワークロードについてシステム全体のパフォーマンス分析を実行可能となります。ここで使用するのが Arm Development Studio と Arm Mobile Studio という 2 つのツールスイートです。これにより、アプリケーション開発時に使用するソフトウェアやツールの使いやすさや簡便さが高まり、シームレスでセキュアに、かつ高い信頼性での作業が可能となります。

「Google Play ストアで公開される新しいアプリにはすべて 64 ビットアーキテクチャのサポートが義務付けられました。」

Arm Development Studio は、ハードウェアカウンターを調査し、システムを画像ベースの複雑なアプリケーションに最適化するためのツールスイートです。パフォーマンス分析も開発者の使い勝手を大きく左右します。これにより開発者は各種のテクノロジー上でのワークロードの機能を理解し、システム内でボトルネックが生じる場所を特定できます。これをさらに助けるのが Arm Mobile Studio に含まれる Performance Advisor です。この Arm の新しいツールは、読みやすいパフォーマンス分析レポートを生成し、モバイルゲーム市場の幅広い開発者やアーティストにグラフィックス処理の問題やボトルネックを視覚的に表示します。レポートは、Arm Cortex CPU、Mali GPU、Ethos NPU を搭載したプラットフォームから収集した豊富な技術的パフォーマンスデータに基づいています。Mobile Studio のプロフェッショナルエディションには Continuous Integration（継続的統合）機能があり、自動パフォーマンス分析と複数デバイスにわたる毎時間の回帰テストにも対応します。

Arm は、複数のコンピューティングドメインにわたってパフォーマンスを改善するソフトウェアフレームワークとコンピューティングライブラリも提供しています。Arm NN は、すべての Arm IP で ML パフォーマンスを引き上げるそのようなソフトウェアフレームワークの代表例です。これは複数のアプリケーションで ML を改善したい開発者に広く利用されています。

エコシステム

Arm エコシステムは幅広さと膨大な規模を誇ります。このため Arm はエコシステム内のパートナーと協力し、シリコン設計のサポートからゲームの最適化まで幅広く取り組んでいます。開発者のコミュニティは、このエコシステムサポートの重要な部分を占めます。Total Compute は、開発サイクルのすべての段階で、Arm の IP がエコシステムに役立つクオリティの高いツールやコンプライアンススイートを提供することを保証します。たとえばシステム設計の早期段階では、パートナー各社が Arm Fast Models[4] を使用して仮想プロトタイピングを管理します。多くの場合は EDA ベリフィケーション/バリデーションツールも併用します。Fast Models は、シリコンの製造前からハードウェアとソフトウェアのシームレスな協調開発を促進するとともに、システムをサポートするソフトウェアの製品化期間を短縮し、速やかに市場に普及させます。

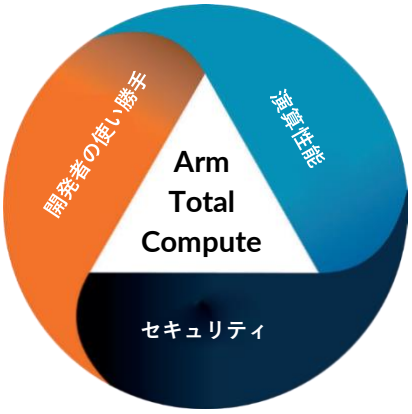
多くの場合、重要なツールはエコシステム内のさまざまなパートナーと共同で提供されます。その実際の例が Arm と Unity の提携です。これにより Arm のパフォーマンス分析機能と Unity ツールが統合され、ゲームの効率が上がるとともに、開発者はゲームエンジン上の各種ワークロードが使いやすくなります。Arm は Burst コンパイラに関しても Unity と協力し、Unity でプロジェクトを構築する開発者が Arm Neon 命令セットをトランスペアレントに利用できるようにしました。これにより Arm アーキテクチャがサポートする Unity プロジェクトのパフォーマンスが Android デバイス上で向上します。

世界中の Unity 開発者にとっては、効率的なアプリケーションを短期間で幅広い消費者向けに製品化できることがメリットです。消費者にとっては、高性能アプリケーションを利用し、複雑なデジタルイマージョン体験を楽しめることになります。

Arm は、エコシステム内のパートナーが提供する他のツールにも専門知識を提供しています。Unity との協力に加え、Arm は Google の新しい Android GPU Inspector（AGI）のローンチパートナーでもあります。
AGI も、ゲーム開発会社が GPU から最大限のパフォーマンスを引き出し、Android 上でイマーシブなモバイルゲームを実現するツールの 1 つです。
AGI により、グラフィックスプログラマーは GPU の使用状況を Arm Mali GPU を実行するモバイルデバイスで視覚化し、最も効果的な部分に力を注ぐことができます。

最後に、Arm は常に 64 ビット開発者との連携とサポートに取り組んでいます。2019 年 8 月 1 日現在、Google Play ストアで公開される新しいアプリにはすべて 64 ビットアーキテクチャのサポートが義務付けられました。

開発者にとって 64 ビットには大きなメリットがあります。アプリケーションを 64 ビットに移行することでパフォーマンスが向上（ワークロードによっては最大 20%）し、セキュリティも強化されます。このようなメリットはユーザー体験に直接的な影響を与えます。たとえば Arm は Unity との提携により、Unity 2018 の多様なコンテンツについて分析を実行し、64 ビットアプリケーションで総フレームレートが 9.5%から 16.7%に上昇することを確認しました。2023 年以降、Arm の CPU はすべて 64 ビット専用となります。



6 Arm の新しい方向性

Total Compute は、次世代のデバイスでセキュアな最先端のデジタルイマージョンを実現するテクノロジーの基盤です。Arm IP の開発、そしてパートナー各社への価値提供における新たな進化でもあります。Total Compute の 3 つの柱に沿って協調設計されたコンポーネントが、この一貫したソリューションを構成しています。

包括的なソリューションは、システム全体を考慮した SoC 設計に対するパートナー各社のニーズに一致しています。これにより現在、実際に存在するコンピューティング上の課題に対応できるだけでなく、未来のデジタルイマージョンへの対応も可能となります。

未来の消費者向けデバイスは、優れた効率で複雑なデジタルイマージョンワークロードを処理するため、現在より高いパフォーマンスを必要とします。しかも実装面積や発熱が制約される現在および未来のデバイス形状で、高いパフォーマンスを提供する必要があります。また、コンピューティングシステムの全側面にわたり、「多層防御」によるセキュリティの向上も必要です。最後に、このようなパフォーマンスとセキュリティの利点を多くの開発者が簡単に利用し、デジタルイマージョン体験を提供するアプリケーションを開発できることが大切です。

Total Compute は、開発するデバイスがスマートフォン、スマートホームデバイス、ノート PC、タブレット、XR ヘッドセット、あるいはまだ市場に出ていない斬新なデバイスなど何であれ、クラス最高の体験を目指して最適化された、堅牢、高性能、セキュアなフレームワークの上に構築されていると保証するものです。

コンピューティングの未来は夢にあふれています。すべての消費者向けデバイスにデジタルイマージョンが普及し、イマーシブな体験を楽しめるようになるでしょう。演算性能、セキュリティ、開発者の使い勝手を 3 本の柱とする Total Compute は、パートナー各社や消費者の求めるスピーディー、安全、シームレスで高度なコンピューティングの未来を実現します。

Arm Total Compute は、すべてのコンシューマー機器における最高の体験、そして Arm を基盤とするモバイルの未来の起爆剤なのです。

Total Compute がどのように革新と製品開発をスピードアップするのか、詳しくは www.arm.com/ja/markets/consumer-technologies をご覧ください。

注釈

1 Northstar Research による 2020 年の調査：https://www.arm.com/blogs/blueprint/read-arm-2020-global-ai-survey

2 Confidential Computing Consortium：https://confidentialcomputing.io/

3 出典：2022 年開発者全国調査

4 Fast Models は、Arm CPU およびシステム IP の機能的に正確なプログラマービューモデルであり、ハードウェアが実装される前でも最新の Arm IP をターゲットにしたソフトウェア開発を促進します。簡単に導入できて自動化可能なターゲットとして、統合や検証を繰り返すことができる点も評価されています。

5 詳細：https://source.android.com/docs/security/test/memory-safety/bootloader-support