



WP499 (v1.0) 2018 年 11 月 19 日

データセンターを刷新する適応型 Alveo アクセラレータ カード

Alveo データセンター アクセラレータ カードは幅広いワークロードにおいて卓越した演算性能を発揮し、総保有コスト (TCO) を最小限に抑えます。また、すぐに使用可能なアプリケーションも豊富に揃えられています。

概要

これまでデータセンターを支えてきた CPU は、かつての勢いを失っています。

次世代 CPU の性能向上が微増にとどまると見られる一方、演算リソースに対する要求は非常に多くのワークロードやアプリケーションで爆発的に増大しています。幅広い演算ニーズに応え、スケールアウトと運用コストを管理するため、適応性に優れた演算アクセラレータが求められています。

ザイリンクス Alveo™ データセンター アクセラレータ カードには、すぐに使用可能なアプリケーションが豊富に揃っており、高い適応性を備えたアクセラレータを迅速に量産運用できます。この結果、演算性能が飛躍的に向上すると共に、TCO を最大 65% 削減できます。

はじめに

ビッグデータ解析や機械学習など、強力なコンピューティング性能と膨大なデータセットをスマートに活用することが、今や業界を問わず競争力を高める上で必須となっています。この結果、より高性能で低コストなデータセンター向けコンピューティングソリューションへのニーズが高まっています。

CPU (マルチコア CPU を含む) だけでこうした要求に応えるのは、コスト、消費電力、スケーラビリティの面から現実的ではありません。しかも多くのワークロードでは、CPU ベースのサーバーの台数を増やしても、問題解決に十分な性能が得られなくなっています。ムーアの法則が停滞しているため、次世代 CPU でもこの状況はほとんど変わりそうにありません。幅広い演算ニーズに応え、スケールアウトと運用コストを管理するため、適応性に優れた演算アクセラレータが求められています。

高速、適応性、アクセシブル

CPU のみを使用したソリューションではデータセンターで必要な演算性能を達成できなくなっている現状を打破するため、ザイリンクスはすぐに量産対応できる適応性に優れたデータセンター アクセラレータ カードとして、PCIe® フォームファクターの Alveo ポートフォリオを開発しました。図 1 に、パッシブ冷却タイプの Alveo U200 および U250 アクセラレータ カードを示します。Alveo ポートフォリオの詳細は、ザイリンクスウェブサイトの [Alveo のページ](#) を参照してください。

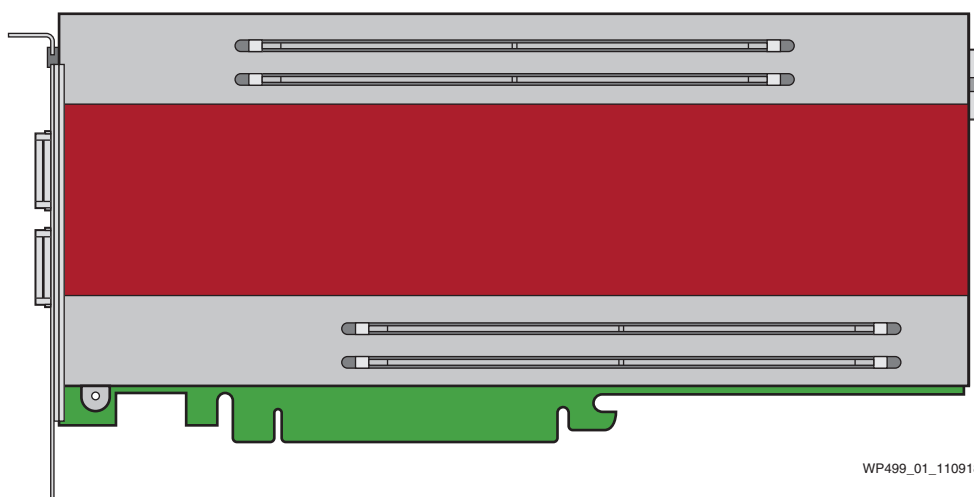


図 1: Alveo U200 および U250 データセンター アクセラレータ カード (パッシブ冷却タイプ)

Alveo データセンター アクセラレータ カードは、ザイリンクスデバイスの利点 (『ザイリンクス デバイス: 演算負荷の大きなシステムに最適な優れたプラットフォーム』(WP492: [英語版](#)、[日本語版](#)) 参照) を活かすことにより、幅広い重要なワークロードに対処します。このソリューションには、次の特長があります。

- 高速 - 最も高いスループットと最も低いレイテンシ
- 適応性 - 現在から将来にわたって多くのワークロードのニーズに対応
- アクセシブル - すぐに使用可能なアプリケーションが豊富に揃い、オンプレミスまたはクラウド ベースでの運用が可能

これらの利点により、Alveo カードは次の点においてこれまでにない価値を提供します。

- アプリケーションの幅 - 機械学習の推論、データベース検索および分析、ビデオ、金融コンピューティング、ゲノム解析など
- 優れた性能と電力効率 - 最大で CPU の 90 倍、GPU の 5 倍
- TCO の最小化 - TCO を最大 65% 削減
- 長期運用が可能なソリューション - ワークロードの変化に合わせてハードウェアが適応

Alveo U200/U250 アクセラレータ カードは現在発売中で、U280 アクセラレータ カードは近日発売予定です。今後登場する Alveo カードは、先に発表された Versal™ ACAP (Adaptive Compute Acceleration Platform) デバイスをベースにすることで、さらに高度な演算性能と電力効率を実現します。たとえば、Versal デバイス内蔵の AI エンジンを使用すると機械学習の推論性能が 20 倍に向上します。Versal の詳細は、サイリンクス ウェブサイトの [Versal ACAP のページ](#) および『Versal: 初の ACAP (Adaptive Compute Acceleration Platform)』(WP505: [英語版](#)、[日本語版](#)) を参照してください。

高速

Alveo アクセラレータ カードは、その特長として、幅広いデータセンター ワークロードで最大のスループットと最小のレイテンシを実現できます。CPU と GPU はアーキテクチャに柔軟性がなく、ソフトウェア主導のデータフローとなるため、一般的にレイテンシとスループットのトレードオフは避けられません。図 2 に、重要なデータセンター ワークロードで CPU を Alveo カードで置き換えた場合の高速化の例を示します。CPU および GPU に対する Alveo の高速化の利点については、この後で詳しく説明します。

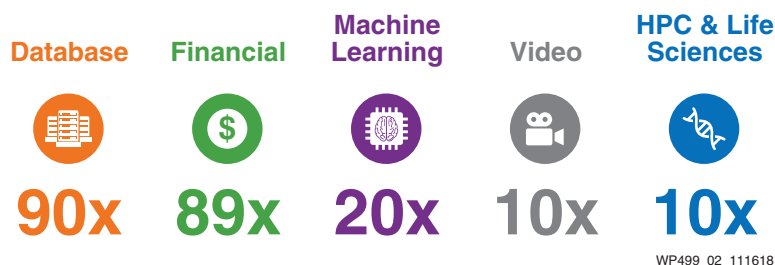


図 2: 高速 - Alveo データセンター アクセラレータ カードによる高速化の例

適応性

サイリンクスはこれまで一貫して、他社をはるかにしのぐ適応性を備えたデバイスを提供してきました。Alveo アクセラレータ カードもこの伝統を引き継いでいます。CPU や GPU とは異なり、Alveo カードはベースのハードウェアを何度でもコンフィギュレーションおよびリコンフィギュレーションして任意のワークロードに対応できるハードウェア適応性を備えています。これに対し、CPU や GPU はベースのハードウェアが固定されており、ソフトウェアドメインの柔軟性しかありません。したがって、ワークロードがハードウェアアーキテクチャとうまく適合しなければ、十分な性能は得られません。ハードウェア適応性がない場合の影響については、後ほどいくつかの例を挙げて説明します。Alveo アクセラレータ カードのハードウェア適応性には、次のような利点があります。

- データセンターのハードウェア使用率が最大限に向上 - より少ないハードウェアで幅広いアプリケーションのアクセラレーションが可能
- 長期運用が可能なハードウェア - ハードウェアを更新しなくてもワークロードの変化へ対応可能 (図 3 参照)。

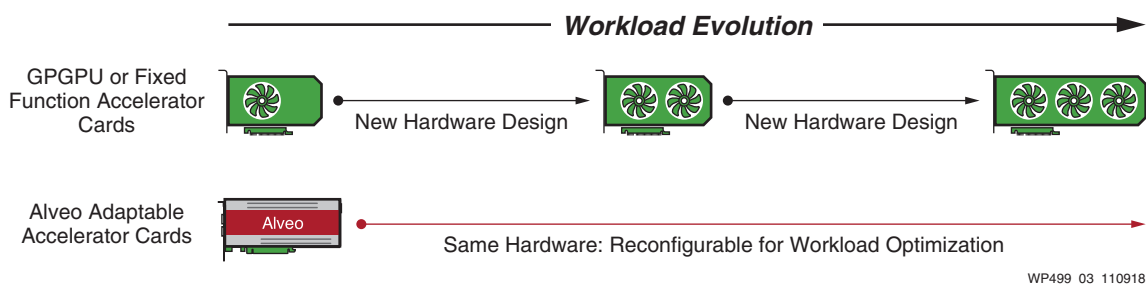


図 3: Alveo アクセラレータ カードの適応性により、ハードウェア更新の回数が減少

Alveo アクセラレータ カードをターゲットにしたアプリケーションを独自に開発、最適化する場合は、使用が簡単で包括的な開発環境としてサイリンクスの [SDAccel™ 開発環境](#) を利用できます。

アクセシブル

Alveo データセンター アクセラレータ カードによる高速化をいち早く体験していただけるよう、ザイリンクスとパートナー各社はすぐに使用可能なアプリケーションを多数開発しています。これらのアプリケーションは、機械学習の推論、ビデオトランスコーディング、データベースの検索および分析、ゲノム解析、金融コンピューティングなどの重要なワークロードを幅広くカバーしています。表 1 に、これらアプリケーションの例と、高速化の度合いを示します。現在提供されているアプリケーションの一覧は、ザイリンクス ウェブサイトの Alveo アクセラレータ カードのページを参照してください。

表 1: Alveo アクセラレータ カードで高速化が可能なアプリケーションの例

分野	アプリケーション	ベンダー	カード	高速化
データ分析	検索および分析ツールキット	BlackLynx	U200	90 倍
機械学習	推論用 ML (Machine Learning) Suite	ザイリンクス	U200 U250	20 倍
ビデオ	ABR (アダプティブビットレート) ビデオトランスコーディング アプリケーション	ザイリンクス、 NGCodec	U200	10 倍
ゲノム解析	高速化されたゲノム解析パイプライン	Falcon Computing	U200	10 倍
金融コンピューティング	リアルタイム リスク ダッシュボード	Maxeler	U200	100 倍

ザイリンクス デバイスによるアクセラレーションの効果に着目し、AWS、Nimbix、Alibaba、Baidu、Tencent、Huawei など多くのクラウド サービスプロバイダーが、ザイリンクス ベースのアクセラレータを FaaS (FPGA as a Service) として提供しています。これらのクラウド ソリューションを利用すると、上記のようなアプリケーションがどれだけ高速化するかを手軽に評価できます。このようにして、ザイリンクス ベースのアクセラレーションがもたらす効果を十分に検討してから Alveo アクセラレータ カードをオンプレミスで運用することが可能です。一般に、機密情報を扱うアプリケーションや、ワークロードのレイテンシ要求が厳しく最高水準の性能が求められるアプリケーションなどでは、オンプレミスでの運用が必要となります。

ザイリンクス デバイスをベースにした FaaS の詳細は、ザイリンクス ウェブサイトの[アクセラレーテッド クラウド サービスのページ](#)を参照してください。

データセンター ワークロードの例

ここでは、3 つの主要なワークロードを例に挙げて、高速、適応性、アクセシブルという Alveo データセンター アクセラレータ カードの利点を詳しく見ていきます。

- 機械学習の推論
- データベースの検索および分析
- ビデオ

Alveo による機械学習の推論の高速化

医療診断、ビデオ監視、衛星画像解析、ネットワークセキュリティ、自動運転車など、多くの業界が機械学習 (ML) テクノロジーを導入して成果を上げています。ML は非常に幅の広い分野で、活発な研究によって多くの種類の ML ネットワークが登場し、それぞれが常に進化しています。また、ML には大きく 2 つのステージが存在します。

- 学習 - 特定のタスクを実行するようにオフラインで機械学習ネットワークをトレーニングする
- 推論 - 学習済みのネットワークを運用してリアルタイムにタスク (翻訳など) を実行する

学習と推論はいずれも非常に多くの演算処理を必要とします。このため、CPU のみをベースにしたソリューションでスループットとレイテンシの要件を満たそうとするのは、コストと消費電力の面から現実的ではありません。

GPU アーキテクチャは ML の学習ステージのワークロードでは大きな効果を発揮しますが、これは主に浮動小数点演算性能が非常に強力で、データ転送能力よりも演算性能に比重を置いていることに理由があります。一方で GPU は、ML の推論ステージへの適用には向いていません。ML の学習ステージと推論ステージのワークロードでは、求められる特性が大きく異なります。

- 推論ではレイテンシが重要
- 推論では演算性能よりもデータ転送能力の比重が高い
- 推論ではそれほど高い精度のデータ型は要求されない

ML の推論では、低レイテンシが求められること、および演算性能よりもデータ転送能力の比重が高いことが GPU にとって大きな課題となっています。Nvidia 社が発表しているデータでも、同社のフラッグシップ製品 Tesla V100 は 900GB/s の HBM メモリ帯域幅を利用した場合でも GoogleNetv1 (Batch=1) で毎秒 870 枚の画像しか処理できておらず、V100 の理論上の GoogleNet 推論性能である毎秒約 4 万枚の 2% 程度にとどまっています⁽¹⁾。これに対し、Alveo アクセラレータ カードは GoogleNet v1 (Batch=1) の理論上の推論性能の約 50% を達成しています。

ML ネットワークが変化していくと、GPU アーキテクチャで対応するのはさらに困難になります。たとえば、GPU がサポートするデータ型は特定のものに限られます。スループットと効率を最大化するために、精度の低いデータ型やカスタム精度のデータ型が必要になった場合、新しい GPU カードが登場するまでこれらを実装できません。このため、「**適応性**」のセクションの図 3 でも説明したように、GPU アクセラレーションを利用したシステムではハードウェアの更新が頻繁に発生し、機器コストが高くなります。

Alveo データセンター アクセラレータ カードは ML に最適な選択肢です。特に次のような利点があります。

- 確定的な超低レイテンシ
- ML ネットワークのニーズに合わせて適応可能なハードウェア
- ML ワークロードが変化した場合でもハードウェアをリコンフィギュレーションできるため、新しいハードウェアの購入が不要
- 新規ハードウェアを購入しなくても任意の精度のデータ型をサポート可能
- ホスト CPU を介在しなくても直接ネットワークに接続可能

ML の推論において Alveo データセンター アクセラレータ カードの多くの利点をすぐに活用できるよう、ザイリンクスとパートナー企業によって開発されたアプリケーションが幅広く用意されています。表 2 に、これらアプリケーションの例と、CPU に対する高速化の度合いを示します。

表 2: Alveo カード向け ML 推論アプリケーションの例

ML アプリケーション	ベンダー	Alveo カード	高速化
推論用 ML (Machine Learning) Suite (TensorFlow、Caffe、MXNet)	ザイリンクス	U200 U250	20 倍
画像分類用ニューラル ネットワーク推論	Mipsology	U200 U250	100 ~ 500 倍
非構造化テキスト解析のアクセラレーション	SumUp Analytics	U200	100 倍

次のセクションではザイリンクス ML Suite を例に、Alveo データセンター アクセラレータ カードが ML 推論にどれだけの利点をもたらすかについて説明します。

1. Nvidia 社ホワイト ペーパー: 『NVIDIA AI INFERENCE PLATFORM』

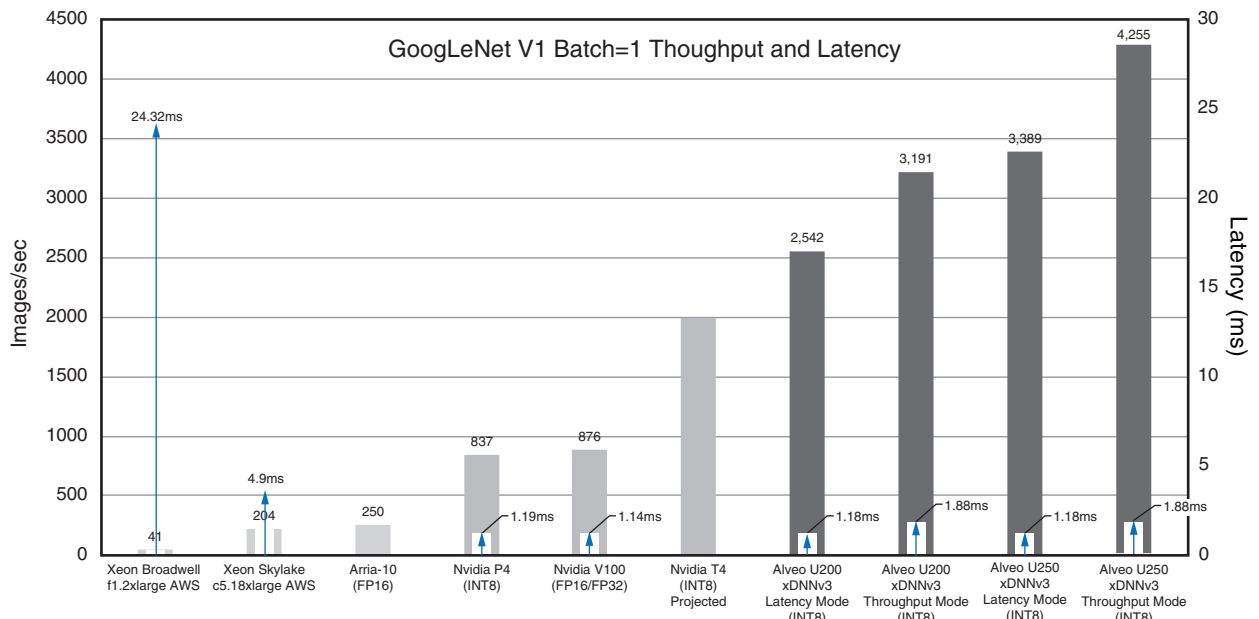
<https://www.nvidia.com/en-us/data-center/resources/inference-technical-overview/> (最終アクセス日: 2018 年 10 月 16 日)。

ザイリンクス ML Suite

ザイリンクス ML Suite は、たたみ込みニューラル ネットワーク (CNN、画像分類や物体検出に使用される ML ネットワークの一種) の推論アクセラレーションをザイリンクス デバイスおよびアクセラレータ カード上に短時間で導入できるようにすることを目的に開発されています。ML Suite はザイリンクス デバイスにインプリメントした xDNN プロセッシング エンジンと xFNN ソフトウェア ツールで構成されており、ユーザーは高位 ML フレームワークに接続して C++/Python API を使用できます。ザイリンクス ML Suite の機能など詳細は、ザイリンクス ウェブサイトの [機械学習のページ](#) を参照してください。

高速 - 最高のスループット/最小のレイテンシによる推論

ML Suite を使用すると、Alveo データセンター アクセラレータ カード用に最適化した CNN 推論アプリケーションを生成できます。『ザイリンクス Alveo アクセラレータ カードを使用した DNN の高速化』(WP504: [英語版](#)、[日本語版](#)) で説明しているように、ザイリンクス ML Suite は現在利用可能なアクセラレータの中で最小のレイテンシ、最大のスループット、および最高の電力効率を達成できます。図 4 は、Alveo U200 および U250 カードでザイリンクス ML Suite を実行した場合の性能と、CPU のみ、GPU、および競合 FPGA の性能と比較したものです。U250 カードは最小限のレイテンシ (1.8ms) と消費電力 (110W) で毎秒 4,100 枚の画像を処理できています。このスループットは CPU の 20 倍以上、Arria 10 FPGA の 15 倍以上、そして Nvidia のフラッグシップ製品 Tesla V100 のほぼ 5 倍に相当します。なお、Alveo カードと ML Suite の組み合わせは、AWS F1 を利用した場合に比べ 50% の改善を実現しています。



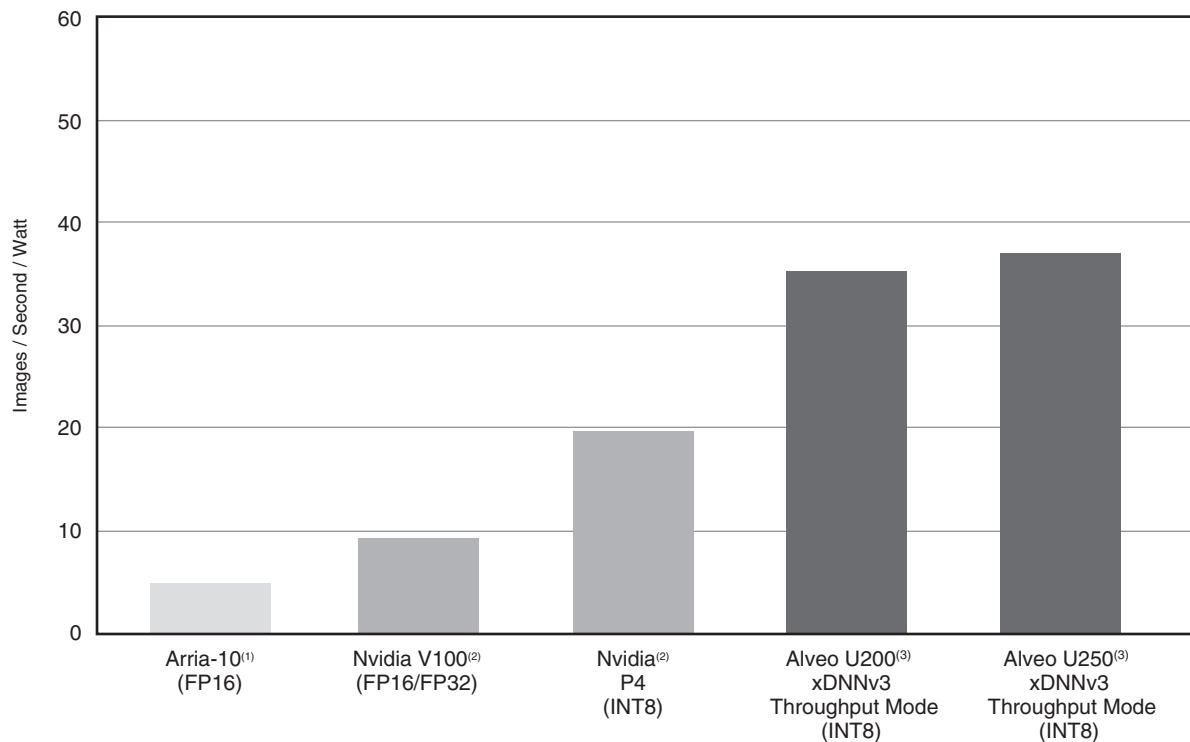
Notes:

1. Xeon E5-2696 v4 f1.2xlarge AWS instance, Ubuntu 16.04LTS, amd64 xenial image built on 2018-08-14, Intel Caffe (<https://github.com/intel/caffe>), Git Version: a3d5b02, run_benchmark.py w/ Batch=1 modification.
2. Xeon Platinum 8124 Skylake, c5.18xlarge AWS instance, Ubuntu 16.04LTS, amd64 xenial image built on 2018-08-14, Intel Caffe, Git Version: a3d5b02, run_benchmark.py w/ Batch=1 modification.
3. Arria-10 numbers taken Intel White Paper, "Accelerating Deep Learning with the OpenCL™ Platform and Intel Stratix 10 FPGAs." <https://builders.intel.com/docs/aibuilders/accelerating-deep-learning-with-the-opencl-platform-and-intel-stratix-10-fpgas.pdf>. Arria latency figures have not been published.
4. Nvidia P4 and V100 numbers taken from Nvidia Technical Overview, "Deep Learning Platform, Giant Leaps in Performance and Efficiency for AI Services, from the Data Center to the Network's Edge." <https://images.nvidia.com/content/pdf/inference-technical-overview.pdf>. Data retrieved on September 3, 2018.
5. Nvidia T4 projection based on current available published benchmark. GoogLeNet Batch=1 performance range between 1700-2000 images/sec based on early power efficiency benchmarks.
6. Alveo U200 numbers measured Intel Xeon CPU E5-2650v4 2.2GHz, 2400MHz DDR4, Ubuntu 16.04.2 LTS Instance running on OpenStack Pike, Centos 7.4, Pre-release Version of MLSuite, streaming_classify.py, synthetic data, MLSuite DSA Thin Shell, FC and SoftMax layers running on Xeon Host and operations not included in compute totals (0.06% of overall compute).
7. Alveo U250 numbers measured Intel Xeon Silver 4110 CPU @ 2.10GHz, CentOS Linux release 7.4.1708, Pre-release version of MLSuite, streaming_classify.py, synthetic data, DSA: ML Thin Shell, FC and SoftMax layers running on Xeon Host and Operations not included in compute totals (0.06% of overall compute).
8. Alveo accelerator cards throughput latency represents worst-case latency (2X latency of large systolic kernel delay), though measured latency might be less.

WP499_04_110718

図 4: GoogLeNet v1 Batch=1 のスループットとレイテンシ

もう1つの重要なパラメーターとなるのが、電力効率です。これによって、運用コスト (OpEx) の大きな要素である全体的な消費電力、および1つのラックに搭載できる演算能力の量が左右されます。図5に示すように、低レイテンシ/リアルタイムの推論において、U200 および U250 は競合する FPGA や GPU に比べ 80% 以上の電力効率を達成しています。



Notes:

1. Arria-10 numbers taken Intel White Paper, "Accelerating Deep Learning with the OpenCL™ Platform and Intel® Stratix® 10 FPGAs," <https://builders.intel.com/docs/aibuilders/accelerating-deep-learning-with-the-opencl-platform-and-intel-stratix-10-fpgas.pdf>.
2. Nvidia P4 and Nvidia V100 numbers taken from Nvidia Technical Overview, "Deep Learning Platform, Giant Leaps in Performance and Efficiency for AI Services, from the Data Center to the Network's Edge." <https://images.nvidia.com/content/pdf/inference-technical-overview.pdf>. Data retrieved on September 3, 2018.
3. Board power figures reported by Xilinx run-time software during benchmark execution.

WP499_05_110718

図5: GoogleNet v1 Batch=1 の電力効率

Alveo U200 および U250 カードは理論上の性能の 50% の性能を達成し、このことから Alveo アクセラレータ カードの適応性がもたらす効果の大きさがわかります。GPU とは異なり、下位のハードウェアをワークロードに合わせてコンフィギュレーションすることで、高い使用率と性能を達成しています。

ML の推論性能を最大 20 倍に高めたザイリンクスの Versal ACAP が登場すると、データセンターのリアルタイム推論における優位性はさらに揺るぎないものになります。今後、Alveo ポートフォリオには Versal デバイスを内蔵した製品が追加されます。

Alveo によるデータベース検索および分析の高速化

データ分析は現在、大企業にとって非常に重要なツールとなっています。基本的に、データ分析とは主に大規模データベースに格納されたデータを検索して関連性のあるものを取得し、これを処理、分析して洞察や知見を得るというものです。大規模な非構造化データセットをより高速に処理、分析することが、大企業の競争力強化につながります。データ分析が高度化する中、CPU のみのソリューションはコストが高く、拡張性に乏しいことが大きな課題となっています。

このため、一部のデータベース検索および分析を GPU で実行して高速化を図る企業も現れています。しかし GPU はそのアーキテクチャの性質上、高速化の効果に限界があります。データベース検索および分析ワークロードは、多くの面で現在の GPU アーキテクチャになじみません。

- 演算性能よりもデータ転送能力の比重が高い
- データフロー パターンの変動が大きい
- ネットワークへの直接アクセスが高速化の鍵

これに対し、Alveo データセンター アクセラレータ カードはベースのハードウェアを (たとえば個々のクエリごとに) 何度でもコンフィギュレーションおよびリコンフィギュレーションでき、ネットワークにも接続できるため、幅広いデータベース検索および分析ワークロードの高速化に理想的な選択肢であることが実証されています。このアクセラレーション効果により、少ないコスト (ハードウェアと消費電力) で高速化が可能です。

Alveo データセンター アクセラレータ カードには、すぐに使用可能なアプリケーションが幅広く用意されています。表 3 に、これらアプリケーションの例と、CPU に対する高速化の度合いを示します。電力効率も、これと同じ程度まで改善されます。幅広いアプリケーションで高速化が可能ということは、Alveo カードの適応性に高い価値があるということです。

表 3: データベースの検索および分析アプリケーションのアクセラレーション

データベースの検索および分析アプリケーション	ベンダー	Alveo カード	高速化
高速検索	BlackLynx	U200	90 倍
低レイテンシの KVS (Key-Value ストア)	Algologic	U200 U250	20 倍
ETL、ストリーミング分析、SQL 分析用のハイパーアクセラレーションレイヤー	Bigstream	U200	30 倍
Hyperion 10G RegEx ファイル スキャン	Titan IC	U200	N/A
Deepgreen DB - Greenplum Database から派生したスケーラブルな MPP データ ウェアハウス ソリューション	Vitesse Data	U200	86 倍
Spark-MLlib ランダム フォレスト アクセラレーション	Xelera	U200	25 倍
Apache Spark MLlib アクセラレーション	InAccel	U200	12 倍

Alveo によるビデオ アクセラレーション

ビデオ ストリーミングは 2021 年まで年平均成長率 (CAGR) 31% で増大し、そのトラフィック量は月間約 16K PB (ペタバイト) に達すると予想されています。これは、コンシューマー インターネット トラフィックの約 81% に相当します⁽¹⁾。この膨大なビデオ データが、既存のネットワークおよびストレージ インフラストラクチャへの大きな負担となっています。ビデオ ストリーミングの転送と保存に伴う負担とコストを軽減するには、画質を損なうことなくビデオ ストリーミングのサイズを削減できる最高品質のビデオ コーデックが必要です。

一部の ASIC や GPU が採用するビデオ コーデックなど、ハード ビデオ コーデックは品質の面で CPU 上で動作する最新のソフト コーデックにはかきません。このため、データセンターのビデオ トランスコーディングでは、CPU へのソフト実装が事実上の標準となっています。

ただしデータセンターで扱うビデオの量が爆発的に増大し、ビデオ コーデックが複雑化している現在、CPU のみを使用してこうしたニーズの増大に対応するのは、コスト、消費電力、拡張性の面から現実的ではなくなっています。また、CPU ではリアルタイムのビデオ トランスコーディングに対応できません。ライブ イベントのストリーミングや監視カメラなどでは、リアルタイムのビデオ トランスコーディングが必須です。リアルタイムのビデオ ストリーミングは 2021 年までに 15 倍の成長が見込まれ、インターネット ビデオ トラフィックの 13% を占めるようになると言われています。⁽¹⁾

Alveo データセンター アクセラレータ カードは、こうしたビデオの課題を解決します。

- 高い演算密度 - 1 つの Alveo カードで多くの CPU を置き換え可能
- 抜群の適応性 - ベースのハードウェアを進化させることで最先端のコーデックをサポート
- リアルタイム性能 - リアルタイムのビデオ トランスコーディングをサポート

これらの利点をすぐに活用できるよう、ザイリンクスとパートナー企業によって開発されたビデオ アプリケーションが幅広く用意されています。

表 4 に、これらビデオ アプリケーションの例と、CPU に対する高速化の度合いを示します。

表 4: ビデオ アプリケーションの高速化

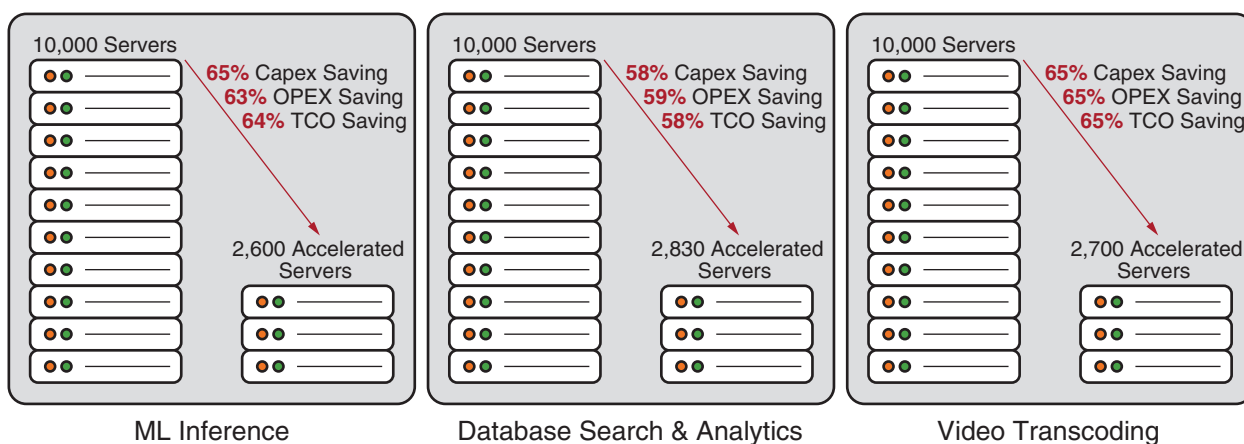
ビデオ アプリケーション	ベンダー	Alveo カード	高速化
ABR (アダプティブ ビット レート) ビデオ トランスコーディング アプリケーション	ザイリンクス、NGCodec、VYUSync	U200	10 倍
パーソナライズド ストリーミング ビデオ エンジン	Skreens	U200	5 倍
動的画像処理/変換	CTACCEL	U200	10 倍

1. Cisco 社ホワイト ペーパー: 『Cisco Visual Networking Index: Forecast and Methodology, 2016-2021』
<https://www.cisco.com/c/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.html>
 (日本語版: https://www.cisco.com/c/ja_jp/solutions/collateral/service-provider/visual-networking-index-vni/complete-white-paper-c11-481360.pdf)
 (最終アクセス日: 2018 年 10 月 25 日)

TCO 削減の例

Alveo データセンター アクセラレータ カードによるアクセラレーションと電力効率の利点は、CPU のみのソリューションに比べ設備投資 (CapEx) と運用コスト (OpEx) の大幅な削減につながります。Alveo アクセラレータ カードは電力効率に優れ、より少ない数で CPU と同じ性能が得られるため、ハードウェア コスト (CapEx) および消費電力と冷却コスト (OpEx) の両面でコストを削減できます。

こうした TCO の削減は、ここまで紹介した 3 つのワークロード例 (機械学習の推論、データベースの検索および分析、ビデオ) のいずれにおいても可能です。これら 3 つのワークロードにおいて、10,000 台の CPU ベース サーバーで構成したソリューションを Alveo アクセラレータ カードによるサーバーに置き換えると、サーバー台数をそれぞれ 2600、2800、2700 台まで減らすことができ、CapEx を 58 ~ 65% 削減できます。同様に、Alveo アクセラレータ カードの優れた電力効率によって消費電力と冷却コストは 59 ~ 65% 削減され、TCO 全体でも 58 ~ 65% の削減が可能です。これらの削減効果を図 6 にまとめます。



WP499_06_110918

図 6: Alveo アクセラレータ カードによる TCO 削減効果

この図には含めていませんが、Alveo カードには長期的に見た利点もあります。「[適応性](#)」のセクションの図 3 で説明したように、Alveo カードのハードウェアはリコンフィギュレーションによってワークロードの変化に適応させることができます。このため、Alveo データセンター アクセラレータ カードは頻繁な交換が不要で、機器/ハードウェア コストのさらなる削減につながります。

まとめ

Alveo データセンター アクセラレータ カードには次の特長があり、すぐに高速化が可能なアプリケーションも幅広く提供されています。

- 高速
- アクセシブル
- 適応性

これらの性質により、ソリューション アーキテクトや製品オーナーはデータセンターおよびそのワークロードとアプリケーションに新たな息吹を与えることができます。また、Alveo データセンター アクセラレータ カードにはハードウェアと消費電力/冷却のコスト削減効果もあり、TCO を最大 65% 削減できます。

アクセラレーションによるメリットの詳細は、ザイリンクス サイトの [Alveo データセンター アクセラレータ カードのページ](#) を参照してください。

ザイリンクス FPGA ベースのアクセラレーションを今すぐ体験するには、各種 FaaS 上でソリューションを評価するか、[Alveo アクセラレータ カードのページ](#) から製品をご注文ください。

改訂履歴

次の表に、この文書の改訂履歴を示します。

日付	バージョン	内容
2018年11月19日	1.0	初版

免責事項

本通知に基づいて貴殿または貴社（本通知の被通知者が個人の場合には「貴殿」、法人その他の団体の場合には「貴社」。以下同じ）に開示される情報（以下「本情報」といいます）は、ザイリンクスの製品を選択および使用することのためにのみ提供されます。適用される法律が許容する最大限の範囲で、(1) 本情報は「現状有姿」、およびすべて受領者の責任で (with all faults) という状態で提供され、ザイリンクスは、本通知をもって、明示、黙示、法定を問わず（商品性、非侵害、特定目的適合性の保証を含みますがこれらに限られません）、すべての保証および条件を負わない（否認する）ものとします。また、(2) ザイリンクスは、本情報（貴殿または貴社による本情報の使用を含む）に関係し、起因し、関連する、いかなる種類・性質の損失または損害についても、責任を負わない（契約上、不法行為上（過失の場合を含む）、その他のいかなる責任の法理によるかを問わない）ものとし、当該損失または損害には、直接、間接、特別、付随的、結果的な損失または損害（第三者が起こした行為の結果であった、データ、利益、業務上の信用の損失、その他あらゆる種類の損失や損害を含みます）が含まれるものとし、それは、たとえ当該損害や損失が合理的に予見可能であったり、ザイリンクスがそれらの可能性について助言を受けていた場合であったとしても同様です。ザイリンクスは、本情報に含まれるいかなる誤りも訂正する義務を負わず、本情報または製品仕様のアップデートを貴殿または貴社に知らせる義務も負いません。事前の書面による同意のない限り、貴殿または貴社は本情報を再生産、変更、頒布、または公に展示してはなりません。一定の製品は、ザイリンクスの限定的保証の諸条件に従うこととなるので、<https://japan.xilinx.com/legal.htm#tos> で見られるザイリンクスの販売条件を参照してください。IP コアは、ザイリンクスが貴殿または貴社に付与したライセンスに含まれる保証と補助的条件に従うこととなります。ザイリンクスの製品は、フェイルセーフとして、または、フェイルセーフの動作を要求するアプリケーションに使用するために、設計されたり意図されたりしていません。そのような重大なアプリケーションにザイリンクスの製品を使用する場合のリスクと責任は、貴殿または貴社が単独で負うものです。<https://japan.xilinx.com/legal.htm#tos> で見られるザイリンクスの販売条件を参照してください。

自動車用のアプリケーションの免責条項

オートモーティブ製品（製品番号に「XA」が含まれる）は、ISO 26262 自動車用機能安全規格に従った安全コンセプトまたは余剰性の機能（「セーフティ設計」）がない限り、エアバッグの展開における使用または車両の制御に影響するアプリケーション（「セーフティアプリケーション」）における使用は保証されていません。顧客は、製品を組み込むすべてのシステムについて、その使用前または提供前に安全を目的として十分なテストを行うものとします。セーフティ設計なしにセーフティアプリケーションで製品を使用するリスクはすべて顧客が負い、製品の責任の制限を規定する適用法令および規則にのみ従うものとします。

この資料に関するフィードバックおよびリンクなどの問題につきましては、jpn_trans_feedback@xilinx.com まで、または各ページの右下にある [フィードバック送信] ボタンをクリックすると表示されるフォームからお知らせください。いただきましたご意見を参考に早急に対応させていただきます。なお、このメール アドレスへのお問い合わせは受け付けておりません。あらかじめご了承ください。