

データセンター ワークロード向け 適応型アクセラレータ カード

クラウド/エッジ データセンター向けの 演算、ネットワーク、ストレージ アクセラレータ

ザイリンクスの Alveo™ U50 データセンター アクセラレータ カードは、金融コンピューティング、機械学習、計算用ストレージ、データ検索/分析などあらゆるワークロードに対応し、最適化されたアクセラレーション機能を提供します。ザイリンクスの UltraScale+™ アーキテクチャをベースに構築され、75 ワットと電力効率に優れたロープロファイル型のフォームファクターにパッケージングされた U50 は、8GB の HBM2、100GbE ネットワーク、PCI Express 4.0 を統合し、あらゆるサーバー形態で運用できるように設計されています。

強力なデータセンター アクセラレーション

高い性能と効率性

- 8GB メモリ (32 AXI チャンネル アクセス) と PCIe Gen4 インターコネクトで、より高速なアプリケーション性能を実現
- 4x 10GbE、4x 25GbE、1x 40GbE、1x 100GbE をサポートする 100G ネットワークで低レイテンシのネットワーク性能を実現

最も広範なワークロードに対応し、性能を最大限に引き上げる

- 計算、ネットワーク、ストレージのワークロードを高速化
- 再構成可能なファブリックを採用しているため、進化するワークロードやアルゴリズムに合わせてアプリケーション性能を最大限に引き上げることが可能 - 固定アーキテクチャでは不可

あらゆるサーバーで運用可能 — オンプレミスからクラウドまで

- スケールアウト アーキテクチャで、いかなるサーバー形態にも対応 - 75 ワットの消費電力とロープロファイル型フォームファクター

強力な開発者向けプラットフォーム

- ザイリンクスおよびパートナー企業が提供する [アプリケーション](#) を含む、拡大し続ける大規模ライブラリを利用可能
- ザイリンクスの [SDAccel™ 開発環境](#) および [Machine Learning Suite](#) を利用して、差別化したソリューションを構築可能



仕様

機能	ALVEO U50
アーキテクチャ	UltraScale+
フォーム ファクター	ハーフ ハイト、ハーフ レンダス、シングル スロット、ロープロファイル
ルックアップ テーブル (LUT)	872,000
HBM2 メモリ	8GB
HBM2 帯域幅	316GB/s ¹
ネットワーク インターフェイス	1 x QSFP28 (100GbE)*
クロック精度	IEEE 1588
PCI Express	PCIe Gen3 x 16、デュアル PCIe Gen4 x 8、CCIX
熱管理ソリューション	パッシブ
熱設計電力 (TDP)	75W

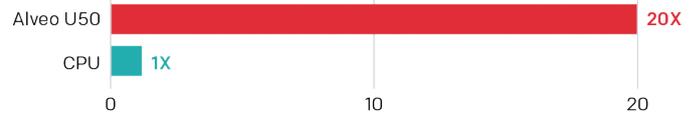
1. A-U50DD-P00G-ES3-G および A-U50-P00G-PQ-G の測定結果で HBM2 のピーク帯域幅は 316GB/s、公称帯域幅は 201GB/s です。
2. ES 版の U50 カードには SFP-DD ポートが 2 つあります。

広範なデータセンター アプリケーションを高速化

計算用ストレージの高速化

- Alveo U50 では、すばやく柔軟に圧縮/解凍の高速化が可能
- 低コスト - Alveo U50 で高速化した圧縮では 33% のコスト削減が可能 (10Gb/s スループット、2:1 圧縮に基づく)

GZIP 圧縮のスループット (GB/s)



Intel Skylake-SP 6152 @2.10GHz CPU (Ubuntu 16.04)
GB/s compression per CPU core = .0229. Alveo U50 = 10GB/s (estimate)

金融シミュレーション - グリッド コンピューティング

- 最短でインサイトを獲得
- 運用コストを削減し、電力効率を最大化
- 確定的なレイテンシによる安定した性能

モンテカルロ シミュレーション

性能と効率性 (パス/秒/W)



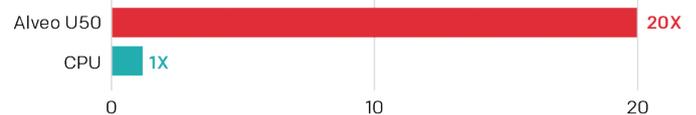
Intel Xeon E5-2697 v4 GCC 5.4.0 Nvidia Tesla V100 16GB PCIe CUDA 10.1 / GCC 5.4.0
Xilinx Alveo U50 SDAccel 2018.3 (estimate)

超低レイテンシ ネットワーク

- 低レイテンシ (20 分の 1)
- Alveo U50 は、レイテンシが 10u の CPU と比較して、500us 以下の取引時間を実現
- 確定的なスループット タイミング

取引時間の高速化

マーケット データから TCP メッセージ (高速化)



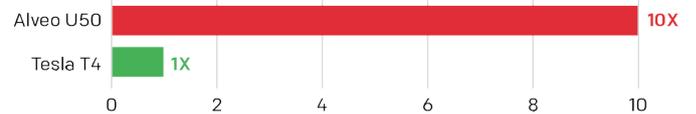
Alveo U50 latency is <0.5us, CPU latency is 10us. Measured from start of packet in on Tick (Market Data) to Start of Packet out on the order to Start of Packet Out on the Order (estimate)

深層学習推論の高速化

- スループット向上 (10 倍) - 1 秒間に変換されるシンボル数
- 低レイテンシ (25 分の 1)
- ノードあたりの電力効率が大幅に向上

音声翻訳のスループット

Transformer NMT (シンボル数/秒が高速化)



Performance of Alveo U50 – with both Alveo U50 and Tesla T4 running (B=2, L=8), Tesla T4 (B=8, L=8) (estimate)

データ分析の高速化

- CPU より優れたクエリ スループットと応答時間
- ノードあたりのコスト効率が良い
- 運用コストの削減

データベース クエリの高速化 (TPC-H Query 5)

(クエリ数/時間が高速化)



CPU Query time = 210ms, 34k query/hr. Alveo U50=24ms, 150k query/hr
INTEL® XEON® PLATINUM 8260 PROCESSOR 35.75M Cache, 2.40 GHz 24 core (estimate)

次のステップ

ザイリンクス Alveo U50 の詳細は、japan.xilinx.com/U50 をご覧ください。

本社
Xilinx, Inc.
2100 Logic Drive
San Jose, CA 95124
USA
Tel: 408-559-7778
www.xilinx.com

ヨーロッパ本社
One Logic Drive
Citywest Business Campus
Saggart, County Dublin
Ireland
Tel: +353-1-464-0311
www.xilinx.com

日本オフィス
ザイリンクス株式会社
〒141-0032
東京都品川区大崎 1-2-2
アートヴィレッジ大崎
セントラルタワー 4 階

アジア パシフィック本社
Xilinx, Asia Pacific
5 Changi Business Park
Singapore 486040
Tel: +65-6407-3000
www.xilinx.com

インド オフィス
Meenakshi Tech Park
Block A, B, C, 8th & 13th floors,
Meenakshi Tech Park, Survey No. 39
Gachibowli(V), Seri Lingampally (M),
Hyderabad -500 084
Tel: +91-40-6721-4747
www.xilinx.com