

高性能な浮動小数点デザインの実現

Arria® 10 FPGA & SoC の高性能な浮動小数点処理能力を引き出す

Amulya Vishwanath

DSP Product Marketing Manager
Intel Programmable Solutions Group

はじめに

5G、機械学習、データセンター、高精度レーダーといった新世代の計算を多用するアプリケーションでは、浮動小数点処理を強化して数値精度を向上させ、消費電力を削減した FPGA & SoC が要求されます。Arria® 10 FPGA & SoC は、単精度浮動小数点 DSP ブロックモードや、ハード化した専用回路を使用する標準精度および高精度固定小数点計算をネイティブサポートする業界初の FPGA & SoC です。単精度浮動小数点 DSP ブロックモードは IEEE 754 に準拠しており¹、図 1 に示す IEEE 754 単精度浮動小数点加算器と IEEE 754 単精度浮動小数点乗算器で構成されます。新しい Arria® 10 単精度浮動小数点 DSP ブロックモードにより、固定小数点演算に匹敵する効率と消費電力の浮動小数点アルゴリズムを実装できます。Arria® 10 FPGA & SoC は、この DSP ブロック・アーキテクチャーによる生産性の利点²を備えているため、高性能コンピューティング・アプリケーションに使用するグラフィック処理ユニット (GPU) の魅力的な代替製品になります。

性能ベンチマーク

本ペーパーでは、以下の 2 つのデジタル信号処理 (DSP) アプリケーションを使用して Arria® 10 デバイスの単精度浮動小数点処理能力を実証します。

- 多相高速フーリエ変換 (FFT)
- 単精度一般行列乗算 (SGEMM)

浮動小数点演算回数 / 秒 (FLOPS) で測定した持続的な DSP 性能と FLOPS / ワット単位の電力効率を分析すると、Arria® 10 デバイスはこの 2 つのベンチマークで大幅な性能向上を示しました。

多相 FFT

FFT は、無線やレーダーなど、多くの DSP アプリケーションに共通のビルディング・ブロックです。高精度レーダーシステムには、長距離レーダー向けやステルス性ターゲット検出用の浮動小数点数値精度が要求されます。浮動小数点機能を搭載する Arria® 10 FPGA & SoC は、この高精度処理を可能にしてシステムのダイナミック・レンジの改善、信号損失の低減、および信号対ノイズ比の改善を実現します。このベンチマークでは、インテル®が開発した、クロックレートを超える速度でデータをサンプリングできる多相 FFT を使用します。多相 FFT ベンチマークは、DSP Builder for インテル® FPGA で利用できるプログラマブル FFT IP コアを使用した、MathWorks MATLAB*/Simulink* ソフトウェアに実装されるモデルベースのデザインです。³

目次

はじめに.....	1
性能ベンチマーク.....	1
多相 FFT.....	1
SGEMM.....	3
まとめ.....	4
参考文献.....	4
追加技術資料.....	4

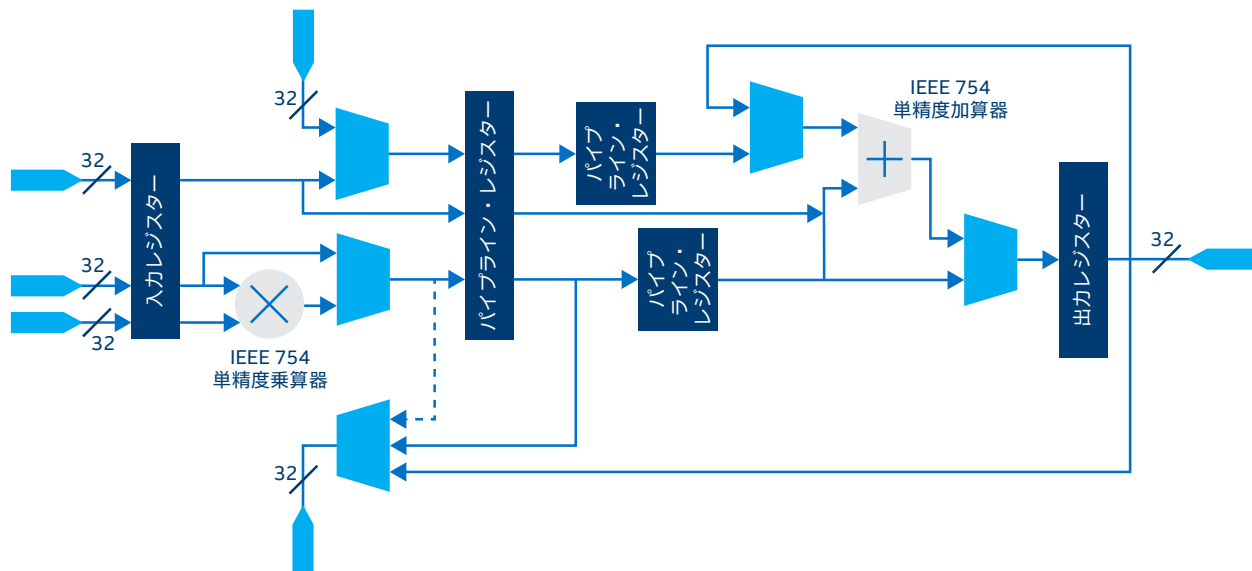


図 1. Arria® 10 デバイスの単精度浮動小数点 DSP ブロックモード

デザインのセットアップ

-1 スピードグレード (0.95 V) の 10AX066N2F40E1SG 量産デバイスに Arria® 10 SoC 開発キットを使用したハードウェアで、3つの多相 FFT 構成 (4K、16K、および 64K) の性能指標を測定しました。ソフトウェア・パッケージは、MathWorks MATLAB*/Simulink* ソフトウェア R2014a、DSP Builder for インテル® FPGA バージョン 16.0、およびインテル® Quartus® Prime プロ・エディション・ソフトウェア バージョン 16.0.2 で構成しました。

結果

多相 FFT ベンチマークは、図 2 に示す 4K FFT 構成で約 40 GFLOPS / ワットの電力効率を持つ Arria® 10 FPGA が、高精度レーダーなどの計算を多用するアプリケーションで 1 TFLOPS を超える持続的な浮動小数点 DSP 性能を発揮できることを示しました。表 1 に、性能指標と 3 つの多相 FFT 構成のリソース使用率を示します。

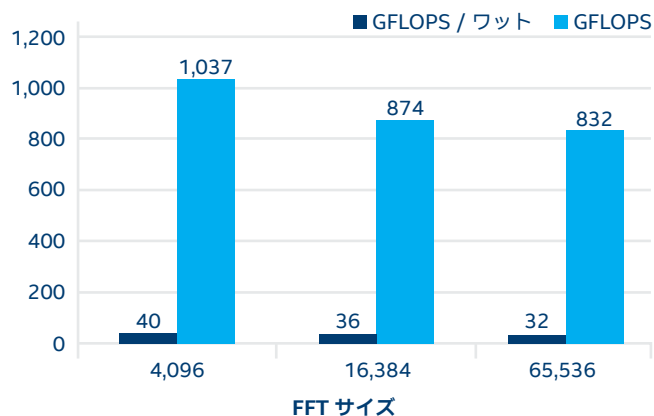


図 2. 4K、16K、および 64K 多相 FFT での電力効率 (GFLOPS / ワット) と持続的な DSP 性能 (GFLOPS)

項目	デザイン 1	デザイン 2	デザイン 3
FFT サイズ	4,096	16,384	65,536
並列相数	16	16	32
f _{MAX} (MHz) ⁵	360	390	325
インスタンス数	3	2	1
スループット (FFT 数 / 秒)	4,218,750	761,719	158,691
持続的な DSP 性能 (GFLOPS)	1,037	874	832
アダプティブ・ロジック・モジュール (ALM) (System in the Loop [SIL] を含む)	113,096 (45%)	89,602 (36%)	113,657 (45%)
DSP ブロック (SIL を含む)	1,687 (100%)	1,384 (82%)	1,616 (96%)
M20K ブロック (SIL を含む)	508 (24%)	617 (29%)	1,175 (55%)
接合部温度 (°C)	64	61	63
消費電力 (ワット)	26	24	26
電力効率 (GFLOPS / ワット)	40	36	32

⁵ OPTIMIZATION_MODE を「Aggressive Performance」に設定したデザイン・スペース・エクスプローラー (DSE) を使用しました。

表 1. 3 つの多相 FFT 構成でのリソース使用率と結果

SGEMM

SGEMM は、線形代数、ニューラル・ネットワーク、および機械学習アプリケーションに共通的に使用する演算です。インテル® FPGA SDK for OpenCL⁵ を使用して開発された SGEMM デザインは、データを効率的に移動させるコンピューティング・アーキテクチャーを実証します。配線をコンフィギュレーションできるため、結果キューの保存時間やデータ待ち時間がなくなります。図 3 に、2次元 (2D) 処理エレメント (PE) のアレイでの計算とデータ移動に対する配線の利点を示します。2つの1次元 (1D) フィーダーアレイがそれぞれ負荷 A と負荷 B から1ビットを取得し、1D ドレインアレイがフィーダーアレイのデータをドレイン C に送信します。コンピューティング・コアをフィーダーアレイから分離して、メモリアクセスを効率的に制御します。PE とホストは自律的に動作し、チャンネルがデータを効率的に移動させてファンアウトを最小限に抑えます。インテル® FPGA SDK for OpenCL⁴ は、GEMM アルゴリズムをリコンフィギュレーション可能なハードウェアに自動変換してドット積演算を実行します。図 4 に、累算による4ベクトルドット積を示します。

デザインのセットアップ

-1 スピードグレード (0.95 V) の 10AX115S2F45I1SG 量産デバイスと Arria® 10 SoC 開発キットを使用した 11R x 16C のパラメータ化可能アレイの性能指標を取得しました。平均消費電力は、70°C で 47 ワットです。浮動小数点機能をハード化していない以前のデバイスファミリーと性能を比較するために、-2 スピードグレード (0.95V) の 5SGSED8K2F40C2 量産デバイスを使用した Stratix® V SoC 開発キットに行列乗算用シングル N 次元カーネル・アーキテクチャーを実装しました。消費電力は 43.43 ワット、コア温度は 62°C でした。ソフトウェア・パッケージは、インテル® FPGA SDK for OpenCL⁴ バージョン 16.1、BSP バージョン 14.1、およびインテル® Quartus® Prime プロ・エディション・ソフトウェアバージョン 16.1 で構成しました。

結果

図 5 に示すように、Arria® 10 ベースの GEMM ファンクションは、Stratix® V デバイスにコンパイルした GEMM ファンクションより DSP 性能 (GFLOPS) と電力効率 (GFLOPS / ワット) が 4 倍近く改善されていました。この性能向上は、インテル® FPGA SDK for

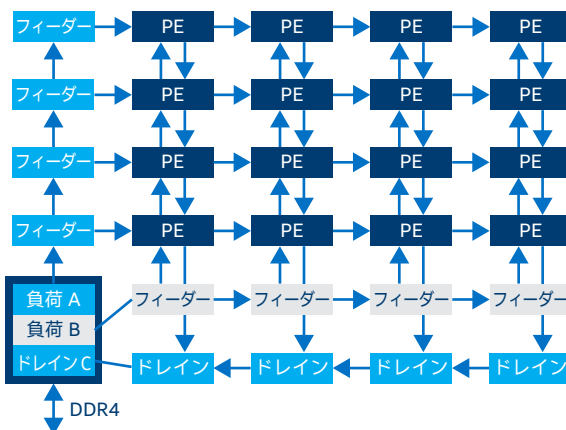


図 3. 2D 処理エレメントアレイ

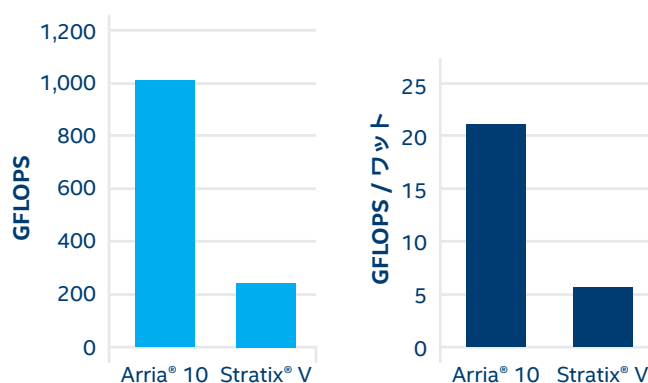


図 5. GEMM ファンクションでの持続的な DSP 性能 (GFLOPS) と電力効率 (GFLOPS / ワット) の比較

OpenCL⁴ と Arria® 10 デバイスを使用して実装した SGEMM ファンクションの効率的なデータ移動とコンピューティング構造によるものです。表 2 と 3 に、Arria® 10 と Stratix® V のリソース使用率と結果をそれぞれ示します。

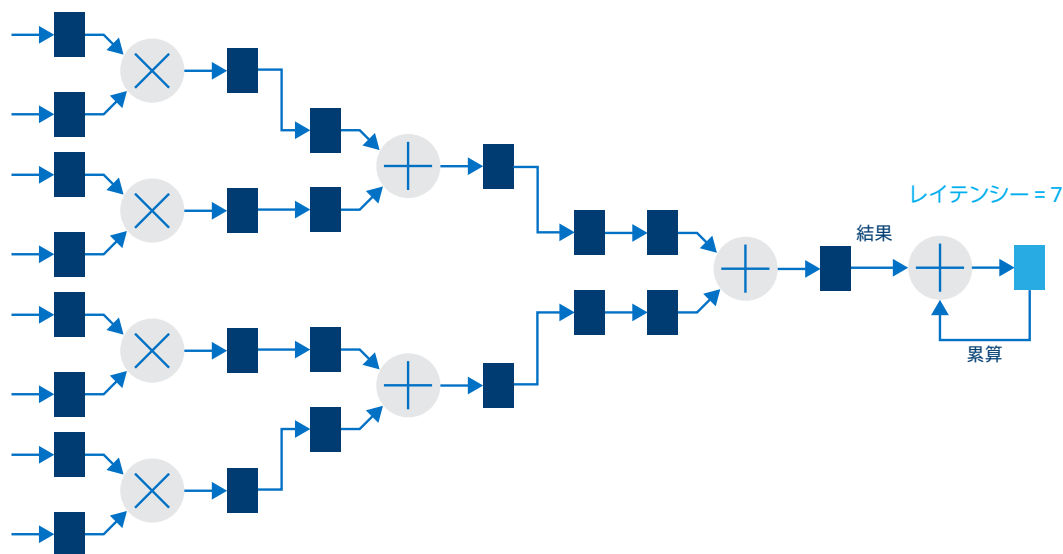


図 4. 累算による4ベクトルドット積

⁵ OpenCL と OpenCL ロゴは Apple Inc. の商標であり、Khronos の許可を得て使用しています。

カーネル f_{MAX} (MHz)	ALM	レジスター数	DSP ブロック数	M20K (RAM ブロック数)	持続的な DSP 性能 (GFLOPS)	電力効率 (GFLOPS / ワット)
365.36	234,420 (55%)	710,542	1,408 / 1,518 (93%)	2,176 / 2,713 (80%)	1,010	21.5

表 2. Arria® 10 SGEMM + BSP のリソース使用率と結果

カーネル f_{MAX} (MHz)	ALM	レジスター数	DSP ブロック数	M20K (RAM ブロック数)	持続的な DSP 性能 (GFLOPS)	電力効率 (GFLOPS / ワット)
173.13	219,953 (84%)	445,089	768 / 1,963 (39%)	1,334 / 2,567 (52%)	260.85	6.006

表 3. Stratix® V GEMM + BSP のリソース使用率と結果

まとめ

Arria® 10 FPGA & SoC は、前世代デバイスを使用した浮動小数点アプリケーションよりも高い性能とワットあたり性能を実証しました。

- Arria® 10 ベースの多相 FFT デザインでは、与えられた電力バジェットでの性能が大幅に向上し、32 ~ 40 GFLOPS / ワットの電力効率を持つ 4K、16K、および 64K 多相 FFT のそれぞれが 832 ~ 1,037 GFLOPS の持続的な DSP 性能を達成しました。
- Arria® 10 ベースの SGEMM デザインでは、11R x 16C のパラメーター化可能アレイで Stratix® V デバイスの 4 倍に向上した 1,010 GFLOPS という持続的な DSP 性能が達成されました。

Arria® 10 デバイスは、1.5 TFLOPS に達する持続的な単精度 DSP 性能を達成して次世代の計算集約型アプリケーションに対応できます。

注：これらのベンチマーク結果を得るには、オリジナルのインテル® デザインファイル、ソフトウェア、およびハードウェア・プラットフォームにアクセスする必要があります。

参考文献

- ¹ http://www.bogdan-pasca.org/resources/publications/2015_langhammer_pasca_fp_dsp_block_architecture_for_fpgas.pdf
- ² https://www.altera.com/en_US/pdfs/literature/po/bg-floating-point-fpga.pdf
- ³ <https://www.altera.com/jp/products/design-software/model---simulation/dsp-builder.html>
- ⁴ <https://www.altera.com/jp/products/design-software/embedded-software-developers/opencl/overview.html>

追加技術資料

インテル® および Arria® 10 FPGA の詳細については、<https://www.altera.co.jp/products/fpga/aria-series/aria-10/overview.html> をご覧ください。

Arria® 10 デバイスを使用した浮動小数点処理の詳細については、<https://www.altera.co.jp/products/fpga/features/dsp/aria10-dsp-block.html> をご覧ください。



© 2016 Intel Corporation. 無断での引用、転載を禁じます。Intel、インテル、Intel ロゴ、Altera、ARRIA、CYCLONE、ENPIRION、MAX、MEGACORE、NIOS、QUARTUS および STRATIX の名称およびロゴは、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。

インテル製品およびサービスは、予告なく変更される場合があります。

インテルが書面にて明示的に同意する場合を除き、インテルは記載されたアプリケーション、または、いかなる情報、製品、またはサービスの使用によって生じる一切の責任を負いません。インテル製品の顧客は、製品またはサービスを購入する前、および、公開済みの情報を信頼する前には、デバイスの仕様を最新のバージョンにしておくことをお勧めします。

* 第三者の社名、製品名などは、一般に各所有者の表示、商標または登録商標です。