

高位設計と低位設計の違いとFPGA演算性能の関係について

横野 智也^{1,a)} 藤田 典久² 山口 佳樹^{1,2,b)} 小林 諒平^{2,1} 朴 泰祐^{2,1} 吉川 耕司^{2,3} 安部 牧人²
梅村 雅之^{2,3}

1. はじめに

近年、高速計算および電力対性能比の向上を目的にFPGAを利用する事例が増加している。この流れを加速した代表例としてMicrosoft社のBing検索[1]を挙げることができるが、FPGAの普及が進んでいる要因の一つはその回路規模が飛躍的に増加したと考えられている。XILINX社のFPGAを例にとると、2010年から現在まで、7-series, Ultrascale, Ultrascale+の3世代が登場している。この3世代の間に、ハイエンド製品であるVirtex FPGAにおいて、10倍以上の回路規模の増大が確認できる(図1)。この急激な変化は、FPGAの設計方法にも大きな変化をもたらした。

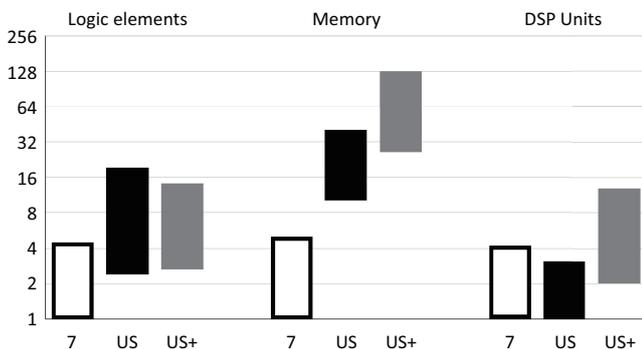


図1 Xilinx社VirtexFPGAの回路規模の変遷。7シリーズの最小VirtexFPGAを1として正規化し、各シリーズ(7, US, US+)の最大と最小の範囲を図示した。

FPGAの回路規模が数万システムゲートの場合、ハードウェア記述言語(HDL: Hardware description Language)によるRTL(Register Transfer Level)設計が妥当であった。これは、少ない回路資源を高効率に利用する観点から、細粒度を含む全ての演算について最適化可能な記述方法が求められたためである。しかし、FPGA1チップの回路規模が100万システムゲートを越えた現在、その全ての動作を把握し、RTL設計により完全な最適化を達成するのは困難になりつつある。そこで、高位記述言語によるHLS

(High Level Synthesis)設計に注目が集まっている。Intel社のIntel SDK for OpenCL, Xilinx社のVivado HLSおよびSDAccelなどHLS設計・開発環境は整いつつある。

そこで本論文では、HDL設計とHLS設計の現状を等距離から評価し議論することで、次世代のヘテロジニアス高性能計算およびそこにFPGAが存在する可能性について検討する。この評価において、初期天体に関する宇宙輻射輸送シミュレーションの一部であるART法をベンチマークとして選択した。

2. ART法

筑波大学計算機科学研究センターでは、宇宙輻射輸送シミュレーションコードARGOT(Accelerated Radiative transfer on Grids using Oct-Tree)[2]の開発を行っている。ARGOTは輻射輸送問題を解く際に、点光源からの輻射輸送計算と空間的に広がった光源からの輻射輸送計算の2つのスキームを組み合わせたものであり、後者の空間的に広がった点光源からの輻射輸送を計算するスキームがART(Authentic Radiation Transfer)[3]である。本研究ではこのART法の部分に注目しFPGAに実装する。

3. ART法のFPGA実装と評価環境

ART法で扱うレイは平行光のみであり、屈折や反射などは発生しない。この性質に注目し、RTL, HLS設計において最適化を行った。本研究の実装は[4]を基に、アクセラレータの並列度を増やすことができるよう、特にメモリアクセスの部分の改良を行った。性能評価において、単位時間あたりにどれだけのメッシュを処理したかを評価する。また指標としてCPUの実装を用い、これはFPGA実装の基となったコードでありC言語で記述されている。この実装はOpenMPを用いて並列化されており、レイ単位でスレッド並列化されている。またART法にはGPU実装もあるため、参考としてGPUも含めた比較評価を行う。評価に使用したデバイスを表1に示す。

表1 CPU/GPU/FPGA実装デバイス

CPU	Intel Xeon E5-2660 v4
GPU	NVIDIA Tesla P100
FPGA1	Intel Arria10 GX115
FPGA2	Xilinx Kintex XCKU115

¹ 筑波大学大学院システム情報工学研究科

² 筑波大学計算科学研究センター

³ 筑波大学数理物質科学研究科

a) yokono@hpcs.cs.tsukuba.ac.jp

b) yoshiki@cs.tsukuba.ac.jp

表 2 リソース使用量とその差分 (括弧内の%は使用 FPGA における回路利用率を意味する)

	LUTs/ALMs	Registers	On chip memory	DSP	Freq.[MHz]
HDL-based (Xilinx KCU1500)	148,010 (22.31%)	137,246 (10.34%)	14,336Kbit (20.74%)	904 (16.38%)	200.00
HDL-based (Intel GX115)	86,246 (20.19%)	142,762 (8.36%)	15,735Kbit (28.32%)	624 (41.11%)	200.00
OpenCL-based[5] (Intel GX115)	169,882 (39.77%)	344,447 (20.16%)	30,624Kbit (43.29%)	624 (41.11%)	173.80

表 3 CPU/GPU/FPGA 実装の性能比較

実装	性能 [M meshes/sec]	CPU 比
CPU(14C)	158.9	1
CPU(28C)	183.4	1.2
GPU(P100)	490.4	3.1
FPGA(HLS Arria10)	1165.2	7.3
FPGA(HDL Arria10)	1423.9	9.0
FPGA(HDL Kintex)	1642.4	10.3

4. 性能評価とリソース使用量

ART 法の問題サイズ $(N_x, N_y, N_z) = (32, 32, 32)$ における FPGA・GPU・CPU の各実装における性能を表 3 に示す。また、CPU の 14 コアでの実行を 14C とし、CPU2 基での実行を 28C とする。表 3 には CPU1 基を基準とした速度比も加えて示されている。ここでの、FPGA の性能は HLS/RTL 実装共に FPGA のリソースが許される限りの実装における性能となっている。

また、各種 FPGA の実装におけるリソース使用量を表 2 に示す。この表は HLS 実装における計算コア数に合わせて RTL 実装との比較になっている。

以上の点から、FPGA 実装では CPU 実装と比較し、設計手法によらず 7~10 倍と十分な性能を提供できる水準となっている。また、ALM 数、レジスタ数、メモリ量については実装の違いによるところが大きく、HLS 設計に対して十分に配慮した設計であれば、RTL 設計と HLS 設計の間で大きくリソース使用量が変わることはないと考えている。

5. おわりに

本研究では、宇宙輻射輸送シミュレーションコード ARGOT で用いられている ART 法を、Kintex UltraScale FPGA と Arria10 FPGA のそれぞれに対して実装し評価を行なった。FPGA 実装は CPU を基準とした時、Xilin/IntelFPGA でそれぞれ 7.26 倍、6.29 倍となった。

リソース使用量に関しては、アクセラレータの実装数は 8 となり、XilinxFPGA では LUT が使用量 22.31% と IntelFPGA では DSP の使用量 41.11% がボトルネックとなることを確認した。これは、問題サイズを大きくした場合、XilinxFPGA ではアクセラレータ数を最大 32、IntelFPGA ではアクセラレータ数を最大 16 まで実装可能であること

を意味し、理論上 XilinxFPGA は最大 4 倍、IntelFPGA では 2 倍の高速化を期待できる。

今後は、他アプリケーションおよび HLS 設計、マルチ FPGA 環境下における評価などを継続し、より応用性の高いデータおよび HLS 設計における知見などを集めて行く予定である。

謝辞 本研究の一部は、JSPS 科研費「JP17H01707」「JP18H03246」、文科省「次世代計算技術の開拓による学際計算科学連携拠点の構築」、文科省「次世代領域研究開発(高性能汎用計算機高度利用事業)」における「次世代演算通信融合型スーパーコンピュータの開発」による。また、Xilinx 社より「Xilinx University Program」を通じて開発ソフトウェアの支援を受けておりここに謝意を表する。

参考文献

- [1] Andrew Putnam, Adrian M.Caulfield, E. S. et al.: A Reconfigurable Fabric for Accelerating Large-Scale Datacenter Services, *ACM/IEEE 41st International Symposium on Computer Architecture (ISCA)*, pp. 13–24 (2014).
- [2] Okamoto, T., Yoshikawa, K. and Umemura, M.: ARGOT: accelerated radiative transfer on grids using oct-tree, *Monthly Notices of the Royal Astronomical Society*, Vol. 419, No. 4, pp. 2855–2866 (2012).
- [3] Tanaka, S., Yoshikawa, K., Okamoto, T. and Hasegawa, K.: A new ray-tracing scheme for 3D diffuse radiation transfer on highly parallel architectures, *Publications of the Astronomical Society of Japan*, Vol. 67, No. 4, pp. 62(1–16) (2015).
- [4] 横野智也, 藤田典久, 山口佳樹, 大島佑真, 小林諒平, 朴泰祐, 吉川耕司, 安部牧人, 梅村雅之ほか: FPGA による宇宙輻射輸送シミュレーションの演算加速, 電子情報通信学会技術研究報告= IEICE technical report: 信学技報 (2018).
- [5] Fujita, N., Kobayashi, R., Yamaguchi, Y., Oobata, Y., Boku, T., Abe, M., Yoshikawa, K. and Umemura, M.: Accelerating Space Radiative Transfer on FPGA using OpenCL, *Proceedings of the 9th International Symposium on Highly-Efficient Accelerators and Reconfigurable Technologies, HEART 2018, Toronto, ON, Canada, June 20-22, 2018*, pp. 6:1–6:7 (2018).