# FPGA による流体専用並列計算ハードウェアの試作実装と性能評価

## Implementation and performance evaluation of FPGA-based parallel computing machine for fluid dynamics

○ 上野 友也，東北大院，宮城県仙台市荒巻字青葉 6-6-01, tomoyaueno@caero.mech.tohoku.ac.jp
田中 大智，東北大院，宮城県仙台市荒巻字青葉 6-6-01, tanaka@caero.mech.tohoku.ac.jp
佐野 健太郎，東北大院，宮城県仙台市荒巻字青葉 6-6-01, kentah@caero.mech.tohoku.ac.jp
山本 悟，東北大院，宮城県仙台市荒巻字青葉 6-6-01, yamamoto@caero.mech.tohoku.ac.jp
Tomoya Ueno, Tohoku University, 6-6-01 Aramaki Aza Aoba, Aoba-ku, Sendai, 980-0865 Japan
Daichi Tanaka, Tohoku University, 6-6-01 Aramaki Aza Aoba, Aoba-ku, Sendai, 980-0865 Japan
Kentaro Sano, Tohoku University, 6-6-01 Aramaki Aza Aoba, Aoba-ku, Sendai, 980-0865 Japan
Satoru Yamamoto, Tohoku University, 6-6-01 Aramaki Aza Aoba, Aoba-ku, Sendai, 980-0865 Japan

**Abstract** In our research project, we are developing an FPGA-based custom computing machine for parallel fluid computation based on the building-cube method (BCM). In BCM, large-scale parallel computation is performed with cubes. The cubes have various sizes and an orthogonal grid with the same resolution of cells. Although BCM is expected to provide efficient load balancing with cubes, the sustained performance is limited in the case of an incompressible fluid simulation due to its low operational intensity. In this paper, we present prototype implementation and performance evaluation of a custom computing module, called a cube engine, which computes fluid simulation for cubes. As a result, we demonstrated this 4 cube engines for 128x128 on an FPGA performed 59.8 Gflops.

### 1. 緒言

数値流体力学 (Computational fluid dynamics, CFD) は，自動車や航空機およびプラントなどの基幹産業において，多くの実験を置き換え，コストを削減できる必要不可欠な技術となっている．高精度かつ大規模な計算が要求される今日の CFD に対しては，大規模計算機の導入コストのみならず，電力を含む運用コストの削減が課題となっている．さらに，流体シミュレーションの計算格子には，非構造格子が広く用いられているが，格子生成に多くの労力が必要となるうえ，その構造の不規則さのために並列計算において計算負荷の均等分散が困難で，大規模計算を行う場合に並列計算の利点が生かしきれないことが課題となる．

これら非構造格子による並列計算の問題を削減する次世代並列計算手法として，ビルディングキューブ法 (Building-cube method, BCM) [1]が提案されている(図 1)．BCM の大きな特徴は計算格子が簡単に生成できること，および立方体領域(キューブ)ごとの並列計算である．BCM は計算領域を，大小様々だがそれぞれ同じ数の格子点(セル)を持つため演算負荷の分散が容易である．

BCM における非圧縮性流体の大規模計算では負荷分散は優れているものの，大きく 2 つのボトルネックがある．1 つは読み出したデータ当たりの演算回数である演算密度の低さであり，計算機全体の性能がメモリ帯域幅に制限されてしまう．2 つ目はデータ交換のオーバーヘッドである．多数の計算ノード間においてデータ交換が頻発し，さらに大きさの異なる隣接キューブ間では，セルの物理量を補間する必要があり，複雑な処理が必要となる．特に汎用計算機ではこれらの問題のために処理時間短縮が困難であり，並列計算の規模に伴い計算効率を大幅に低下させる[2].

これに対し，専用計算機は対象となる計算問題に特化したデータパス，メモリシステムおよびネットワークを構築することで，汎用計算機のソフトウェア処理と比べ，メモリ帯域幅当たりの計算性能が高く，データ交換によるオーバーヘッドが小さい高効率な計算機を実現可能と期待されている[3]．本研究室では，少量多品種とならざるを得ない専用計算機を少ない初期導入費用のもと実現するために，回路再構成デバイスである FPGA を用いる．

これらの背景から，本研究において BCM に基づく高効率かつ高性能な大規模流体専用計算機を目的とし，FPGA を用いた専用計算機の実現を目指している．従来の研究では， キューブ幅 32 の 2 次元および 3 次元キューブ計算エンジンのハードウェア資源消費量およびピーク性能の評価を行った[4]．本稿では，より大きなキューブ計算エンジンの実機におけるハードウェア資源消費量および実際の流体シミュレーションにおける実効性能を評価する．

### 2. ビルディングキューブ法に基づく非圧縮性流体シミュレーション

BCM における各キューブは図 2 に示すように，等間隔直交格子により物体表面は階段状に表現される．物理量の変化が大きい物体表面近傍では格子密度が高く，誤差が小さく微細な計算が可能で，反対に物体から離れた計算領域は物理量の変化が比較的小さいため，密度の低い計算格子により計算負荷を低減できる．

フラクショナルステップ法 (Fractional -step method, FSM) [6]は，非圧縮性流体の数値計算手法である．BCM に基づく FSM においては，各セル中心に物理量として圧力と速度ベクトルを持ち，仮速度計算，圧力のポアソン方程式計算，次時間ステップ速度計算

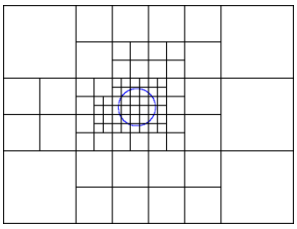

図1.BCM による計算領域の分割例

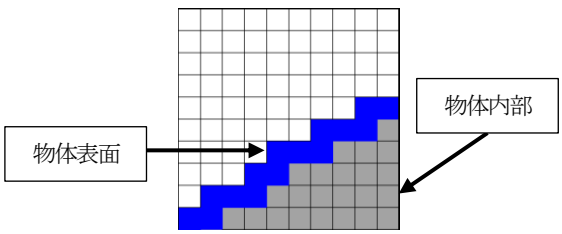


図2. キューブ内部とセルの様子

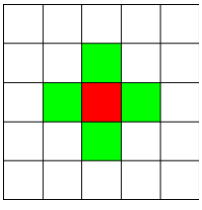

図3. 赤領域を求める場合の 3x3 スターステンシル

表 1. キューブ幅の異なるキューブエンジンのハードウェア資源消費量および実効性能

| | Logic utilization | % | Registers | % | BRAM [kbits] | % | 27-bit DSPs | % | 実効性能 $P$ [GFlops] | $R_{stall}$ [%] |
|---|---|---|---|---|---|---|---|---|---|---|
| Stratix V 5SGXEA7N2 | 234720 | 100.0 | 469440 | 100.0 | 51200 | 100.0 | 256 | 100.0 | - | - |
| Cube engine for 64x64 | 55780 | 23.8 | 59924 | 12.8 | 9360 | 18.3 | 50 | 19.5 | 12.75 | 0.18 |
| Cube engine for 128x128 | 56144 | 23.9 | 61068 | 13.0 | 9520 | 18.6 | 50 | 19.5 | 14.95 | 0.042 |
| Cube engine for 256x256 | 56938 | 24.3 | 62395 | 13.3 | 13640 | 26.6 | 50 | 19.5 | 15.42 | 0.011 |
| Cube engine for 512x512 | 57515 | 24.5 | 63454 | 13.5 | 21740 | 42.5 | 50 | 19.5 | 15.32 | 0.018 |

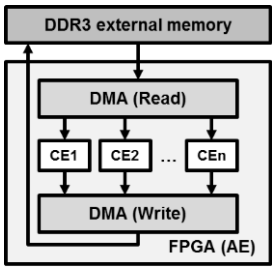


図4. BCM 専用計算機アーキテクチャ


図5. キューブ計算エンジンアーキテクチャ

の 3 ステップで行われる．このステップを全てのセルにおいて実行することで，各時間ステップにおける次時間ステップの圧力および速度を次々と求めることが出来る．また，2 次元 FSM 計算は図 3 に示すような上下左右の 4 点を使うステンシル計算となる．

## 3. BCM を対象とした非圧縮性流体計算専用計算機およびキューブ計算エンジンのハードウェア設計

これまでの研究において提案された BCM 流体専用計算機アーキテクチャを図 4 に示す．本専用計算機は主に DDR3 外部メモリと FPGA から構成され，FPGA 上に DMA およびキューブ計算エンジン(キューブエンジン，CE)を構成する．CE では，データ転送にかかるボトルネックを解消するため, ストリーム計算を用いる．ストリーム計算は，外部メモリから連続したデータを常に流し，同時に計算を行うことで，外部メモリの帯域幅を最大限使用することが出来る．2 章で述べたようにステンシル計算を用いる FSM は局所的な点を同時に参照する必要があり，本研究ではステンシルバッファ[4]を用いて実現する．

## 4. BCM キューブ計算エンジンの実装・評価

CE を，本研究室で開発しているストリーム計算コア高位合成コンパイラである SPGen[5]を用いて HDL を生成し，DE5-NET ボード上に実装した．本ボードは ALTERA Stratix V 5SGXEA7N2 FPGA を搭載している．また，このボードには 2 個の DDR3 PCI12800 SDRAM, PCI-Express (PCIe) 3.0 インターフェース，さらに 4 個の 10G イーサネット接続 SFP+ポートを備える．それぞれの DDR3 メモリは 12.8 [GB/s]でデータを転送する．

BCM において，各キューブの一辺のセル数 (キューブ幅) が小さいほど計算対象に適合できるが，その分データ交換量が増えオーバーヘッドとなる．対象的にキューブ幅が大きいとデータ交換のオーバーヘッドは減るが，適応格子計算の利点が損なわれるうえ，FPGA のハードウェア資源消費量が増える欠点もある．

専用計算機全体の性能向上のためには，単一 FPGA 上 CE の複数実装による並列性の向上が不可欠である．しかし，ハードウェア資源消費量は CE 実装数を制限するため，本研究では，64, 128, 256 および 512 の異なる幅を持つ 2 次元キューブのそれぞれに対し CE を設計し，ハードウェア資源消費量を評価する．

表 1. は，異なるキューブ幅に対する CE の，単一 FPGA におけるハードウェア資源消費量および実効性能を表す．表 1. から，CE を複数実装し並列計算を行う場合，キューブ幅によって単一 FPGA への CE 実装可能最大数を制限するハードウェア資源が異なることが分かる．キューブ幅 64 および 128 では論理ブロック，キューブ幅 256 および 512 は BRAM 消費量が支配的である．

また，今回実装した CE を用いて単一キューブにおける流体シミュレーションを行い，実効性能を次式により求めた．

$$P[\mathrm{Flops}] = F_{max} \times N_{FP} \times (1 - R_{stall})$$

ここで，$F_{max}$は動作周波数であり，150 [MHz]である．$N_{FP}$は CE 中の浮動小数点演算であり，これは CE の幅が異なっても流体の計算式に依存するため全て等しくなる．CE 全体では浮動小数点加算器が 67 個, 乗算器が 25 個, および定数乗算器が 12 個使われており，$N_{FP} = 104$となる．$R_{stall}$はストリーム計算に要した全サイクルにおいて，有効なデータが流れないサイクルの比率であり，計算機中のサイクルカウンタにより計測した．

表 1. の実効性能を見ると，キューブ幅が小さい場合に実効性能が低い．これはストリームのキックをソフトウェアによって行っており，一度のストリームにおけるデータが短く，CE の待機時間がソフトウェア制御のために増えたためだと考えられる．

## 5. 結言

本稿では，BCM に基づく大規模並列計算が可能な非圧縮性流体専用計算機を実装し，実機におけるハードウェア資源消費量と性能評価を行った．異なるキューブ幅における CE のハードウェア資源消費量から，CE 実装数を制限する資源は主に論理ブロックとブロックメモリであることが分かった. キューブ幅 128 の CE においては単一 FPGA 上に 4 個を実装可能でその場合の実効性能は単一 CE の実効性能から 59.8 [GFlops]であると見積もられる．

今後は，今回実装したアーキテクチャを基に複数の CE を単一 FPGA 上に実装し，実際のハードウェア資源使用量や性能評価，さらに電力評価も行う．

### 参考文献

[1] K. Nakahashi. Building-cube method for flow problems with broadband characteristic length. *Computational Fluid Dynamics*, 2002.

[2] 小林 広明, 中橋 和博, 新井 紀夫, 東田 学, 石井 克哉, 江川 隆輔. 次世代ペタスケール CFD のアルゴリズム研究. 学際大規模情報基盤共同利用・共同研究拠点, 平成 23 年度共同研究中間報告書, 2011.

[3] James P. Durbano and Femando E. Ortiz: FPGA- based acceleration of the 3d finite difference time- domain method, *Proceedings of the 12th Annual IEEE Symposium on Field-Programmable Custom Computing Machines*, 1 (2004) 156-163.

[4] Kentaro Sano, Ryotaro Chiba, Tomoya Ueno, Hayato Suzuki, Ryo Ito and Satoru Yamamoto: FPGA-based Custom Computing Architecture for Large-Scale Fluid Simulation with Building Cube Method, ACM SIGARCH Computer Architecture News - HEART '14, September 2014.

[5] 佐野 健太郎, 伊藤 涼, 菅原 啓介, 山本 悟. 階層的モジュール設計を可能とするストリーム計算コア高位合成コンパイラ, 信学技報, vol. 115, no. 109, RECONF2015-29, pp. 159-164, 2015 年 6 月.

[6] J. Kim and P. Moin. Application of a fractional-step method to incompressible Navier-Stokes equations. *Journal of Computational Physics*, Vol. 59, pp. 308-323, June 1985.