

チップ間誘導結合 TCI を用いたヘテロジーニアスマルチコア Cube-2 の評価

天野英晴†

† 慶應義塾大学理工学部情報工学科

1 はじめに

科学研究費 S「ビルディングブロック型計算システムの研究」[1] では、チップ間無線通信用インタフェース (TCI) を用いて、レゴブロックのように様々な機能のチップを組合せて大規模システムを構築するビルディングブロック型計算システムの構築を目指している。様々な組み合わせのチップスタックを目的に応じて作成することで、目的に応じた機能、インタフェース、性能、消費電力を低コストで実現でき、IoT、Wearable Computing、自動運転、知能ロボットなどに広く利用されることを狙っている。

プロトタイプ 2 号機である Cube-2 は、最初のプロトタイプ Cube-1[2] の経験を基に開発中で、昨年 7 月にテープアウトを行い、現在チップの製造待ち状態である。Cube-2 に開発した 4 種類のチップは、統一されたインタフェースを持ち、それぞれのメモリがホスト CPU のメモリ領域にマップされ、DMA 機構によりストリーム処理に適したデータ転送を行う。

2 Cube-2

Cube-2 は、最初のプロトタイプ Cube-1[2] の実装経験を基にし、4 枚積層での安定稼働を目指すと共に、TCI を用いて容易に様々なシステムを構築するための技術を確認するのが目的である。Cube-1 と Cube-2 の仕様の相違を表 1 に示す。

Cube-1 で用いた e-shuttle に代って、プロセスには国家プロジェクト LEAP により開発された SOTB(Silicon On Thin Buried Oxide) を用いた。SOTB は、FD-SOI の一種であり、10nm 程度の極薄酸化膜層上にトランジスタを形成する。低電圧動作でもばらつきが少なく、高い性能を得ることができ

表 1: Cube-1 と Cube-2 の仕様

	Cube-1	Cube-2
Process	e-shuttle 65nm CMOS (12-Metal)	SOTB 65nm CMOS (7-Metal)
Area	2.1mm × 4.2mm	3mm × 6mm
TCI	3Gb/s 240 μ m	2.5Gb/s 240 μ m
Supply Voltage	0.5-1.2V	0.3-1.2V
Target Freq.	50-100MHz	50-100MHz
Chip Thickness	40-80 μ m	80 μ m

る。また、ボディバイアスによりトランジスタのスレッショルドレベルを制御することができ、リーク電流と性能のバランスを制御することができる。ビルディングブロック型計算システムは、様々なチップを自由に組合せることができるのが最大の特徴だが、チップのばらつき、特性差に対処するため、ビルディングブロックの電気的特性を接続時に調整できなければならない。SOTB 技術を利用し、バイアス制御を接続時に行うことで、この問題点に対処することが可能となる。アクセラレータを一種類に限定していた Cube-1 に対して Cube-2 の特徴は、異なった種類のアクセラレータを接続して一つのシステムとして利用できる点にある。積層の様子を図 1 に示す。

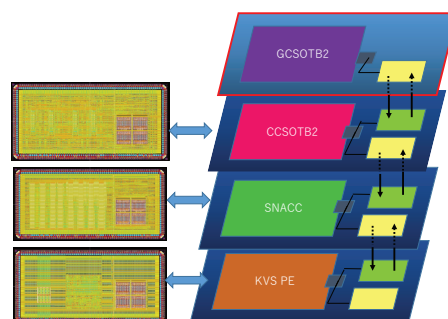


図 1: Cube-2 の積層構成

Evaluation of a heterogeneous multi-core Cube-2 using wireless inductive coupling Through Chip Interface

Hideharu Amano†

†Keio University

Cube-2 は、AI 画像認識システムを応用対象としており、ホストで受け取った画像を CC-SOTB2 で前処理し、SNACC で深層学習による画像認識を行い、KVS PE によりこれをデータベースに格納する操作を各チップで行う。このため、Cube-2 は以下のチップから構成される。それぞれのチップは独自の特徴を持っており、それぞれ個別に発表が行われており、参考文献はホームページ [1] を参照されたい。

GC-SOTB2: R3000 コンパチブルの CPU で、Cube-1 で用いらた Geyser と同一構成だが、細粒度パワーゲーティング機構は装備していない。代って、SOTB のボディバイアス技術を有効に利用できるメモリを装備している。Geyser 同様 Linux 他、組み込み用の OS が稼働する予定である。

CC-SOTB2: 低電力再構成可能アクセラレータ CMA(Cool Mega Array) の最新版で可変パイプライン機構を装備する。画像フィルタなどの処理を低電力で実行することができる。

SNACC: 畳みこみニューラルネットワーク用アクセラレータ。SIMD 型で専用命令を持つアクセラレータを 4 コア搭載している。

KVS: Key Value Store 型のデータベース処理用アクセラレータ。専用命令のプロセッサとメモリを搭載している。

Cube-2 は、Cube-1 の反省に基き、図 2 で示す直線状のネットワークを採用した。ホスト CPU の GC-SOTB2 は IP を 2 セット設け、他のアクセラレータは IP を 4 セット装備し、IP の間隔分シフトして積層する。これにより、積層された隣接チップ間に双方向リンクが形成され全体では、直線状のネットワークになる。ここではこれをエスカレータネットワークと呼ぶ。

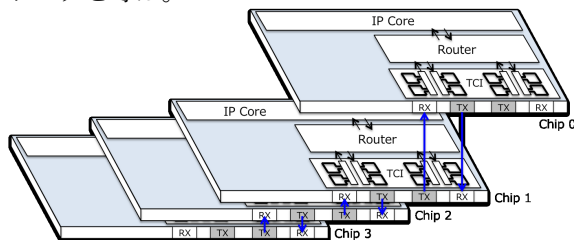


図 2: エスカレータネットワーク

エスカレータネットワークは、上りと下りが独立した経路になっており、最上層、最下層まで行くことなしに上下の方向を切り換えることができる。こ

の点で Cube-1 で用いた単方向リングよりも平均距離が短かくてすむ。

3 性能見積り

Cube-2 はチップ間でストリーム処理を想定しており、独立に処理できるストリームに対して図 3 に示すパイプライン処理を実行する。

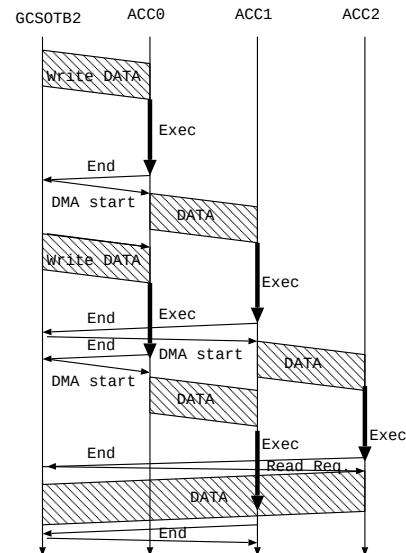


図 3: Cube-2 のパイプライン実行

GCSOTB2 は複数ストリームの処理を行うコマンドを適切に発行する必要がある。さらに、CMA、SNACC についてはデータメモリにダブルバッファが装備されており、データの送受信と演算を並行して行うことができる。見積りの結果、それぞれのチップの処理時間が 32bit データ当り 6 クロックを越えれば、転送時間の隠蔽が可能である。

4 おわりに

Cube-2 のそれぞれのチップは 1 月半ばに到着予定であり、まずはそれぞれを 1 枚ずつテストし、さらに接続してテストする予定である。予定通りにチップが到着すれば発表時には何らかの実機評価結果が得られる予定である。

参考文献

- [1] http://www.am.ics.keio.ac.jp/kaken_s. 科研費基盤研究 (S) ビルディングブロック型計算システムに関する研究ホームページ. 2015.
- [2] N. Miura and et al. A Scalable 3D Heterogeneous Multicore with an Inductive ThruChip Interface. In *IEEE Micro*, Vol.33, No.6, pages 6–15, 2013.