
テクニカルノート

CC-NUMA 型並列計算機ノードの SOC 化に関する考察

古川 文人[†] 大津 金光^{††}
横田 隆史^{††} 馬場 敬信^{††}

将来の半導体技術のロードマップである ITRS と 2001 年現在の DRAM 製造技術を基に、2005 年時点のシステムオンチップ（以下、SOC）の現実的な製造技術を予測すると、最大集積トランジスタ数は約 5 億個、最大集積オンチップ DRAM 容量は約 460 MB と予測できる。これに従えば、2005 年には、2001 年時点の並列計算機ノードを部分的あるいは完全に SOC 化することが可能である。本稿では、2005 年における半導体技術の予測に基づき、CC-NUMA 型並列計算機ノードの SOC 化についてシミュレーションを行って考察した。その結果、2~64 台構成の規模において、1) SOC 化ノードへのオンチップ DRAM メモリの集積は、性能の面で有効であること、2) 大容量オンチップ DRAM メモリ集積による同チップ上のキャッシュの極端な小容量化は、並列計算機全体の性能を低下させる可能性があり、オンチップメモリ容量とキャッシュ容量との間にトレードオフ点が存在することの 2 点が判明した。

Consideration of System-on-chip Nodes for CC-NUMA Multiprocessors

FUMIHITO FURUKAWA,[†] KANEMITSU OOTSU,^{††} TAKASHI YOKOTA^{††}
and TAKANOBU BABA^{††}

We predict that it will be possible to integrate 500 million transistors or 460 MB DRAM onto a chip in 2005 from ITRS and the report about the manufacture of DRAM as of the year 2001; therefore, in 2005, the parallel computer node of the year 2001 can be implemented as a system-on-chip processor. Based on the quantitative prediction of the semiconductor technologies, this paper presents the consideration based on the simulation results of the system-on-chip node for CC-NUMA multiprocessors of the year 2005. When the total number of multiprocessors is 2 to 64, the results show that 1) the system-on-chip node integrated with on-chip DRAM memory is effective, 2) there is a trade-off point between the size of on-chip DRAM memory and the size of cache, because the performance of the parallel computer may degrade when the integration of large on-chip DRAM onto the system-on-chip makes the size of cache small.

1. はじめに

将来の半導体技術のロードマップである ITRS¹⁾と 2001 年現在の DRAM 製造技術²⁾を基に、2005 年時点のシステムオンチップ（以下、SOC）の現実的な製造技術を予測すると、最大集積トランジスタ数は約 5 億個、最大集積オンチップ DRAM 容量は約 460 MB と予測できる³⁾。これに従えば、2005 年には、2001 年時点の並列計算機ノードを部分的あるいは完全に SOC 化することが可能である。

このとき、ノード単体の高性能化の追求だけでは、システムの高性能化が達成できるとは限らない。並列プログラムは複数のノード間で通信をしながら処理を進めるため、ノードの SOC 化の際には、ノード単体の性能とネットワークの性能の両者がバランスを保って、有効に活用される設計が必要である。我々は、このような観点のもとに SOC 化ノード内の構成の初期的検討をメモリアクセスの単純なモデル化により行ってきた⁴⁾。その結果、PE とルータの同一チップ上への集積は、性能・コストの両面で有効であることが明らかにになっている。

本稿では、SOC 製造技術の予測に基づいて SOC 化ノードのモデルをさらに詳細化し、ベンチマークプログラムを用いたクロックレベルシミュレーションを行う

[†] 宇都宮大学大学院工学研究科

Graduate School of Engineering, Utsunomiya University

^{††} 宇都宮大学工学部

Faculty of Engineering, Utsunomiya University

表 1 SOC 化ノードの評価モデル

Table 1 Evaluation models of the node with SOC.

			Type A	Type B	Type C
プロセッサコア			4 命令同時発行 out-of-order スーパースカラ（2GHz 動作）		
オンチップ SRAM キャッシュ	ラインサイズ 命令キャッシュサイズ		64 Bytes 64 KBytes		
	データ キャッシュ	サイズ（連想度）	14.3 MBytes（4）	7.16 MBytes（4）	64 KBytes（4）
		書き込み方法	ライトバック・ライトアロケート		
		アクセスレイテンシ	3 サイクル	2 サイクル	1 サイクル
オンチップ DRAM メモリ	サイズ（バンク数）	-	225 MBytes（4）	449 MBytes（4）	
	アクセスレイテンシ	-		10 サイクル	
	バンド幅	-		64 Bytes/サイクル	
オフチップ DRAM メモリ	サイズ（バンク数）	無制限（4）	-		
	アクセスレイテンシ	40 サイクル	-		
	バンド幅	64 Bytes/2 サイクル	-		
フロー制御方式			ワームホール		
ネットワーク	レイテンシ /1hop	データなし（2hops 以降）	20 サイクル（+8 サイクル/1hop）		
		データ付き（2hops 以降）	42 サイクル（+8 サイクル/1hop）		

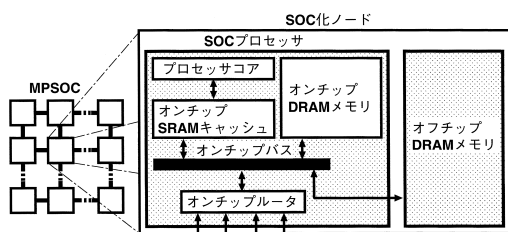


図 1 MPSOC と SOC 化ノード内構成

Fig. 1 Organization of the MPSOC and the node with SOC.

ことで、2005 年における 2~64 台構成の CC-NUMA 型並列計算機ノードの SOC 化について考察する⁵⁾。

2. 評価モデル

2.1 SOC 化ノードで構成する並列計算機

本稿でモデルとする SOC 化ノードで構成する並列計算機 MPSOC (MultiProcessors with System On a Chip) と SOC 化ノード内の構成を図 1 に示す。MPSOC は、スケーラビリティの面で有利な CC-NUMA 型並列計算機である。また、MPSOC のネットワークポロジは、2 次元メッシュである。SOC 化ノードは、1 つの SOC プロセッサと 0 バイト以上のオフチップ DRAM メモリから構成される。また、SOC プロセッサは 1 つのプロセッサコア、それぞれ 0 バイト以上のオンチップ SRAM キャッシュ (以下、キャッシュ) とオンチップ DRAM メモリ (以下、オンチップメモリ) および 2 次元メッシュの構成に必要な 4 本のネットワークリンクを持つオンチップルータ (以下、ルータ) を 1 チップ上に集積した DRAM 混載 SOC である。

2.2 SOC 化ノード内構成の評価モデル

プロセッサコアは 4 命令同時発行 out-of-order スーパースカラ、命令キャッシュ容量は 64 KB と仮定し、我々が予測した 2005 年の SOC 製造技術で実現可能な SOC 化ノードの 3 通りの評価モデルを表 1 に示す^{*}。それぞれの評価モデルの特徴を以下に示す。

• Type A

キャッシュ容量が 3 つのモデルの中で最大であるため、キャッシュアクセスレイテンシが 3 サイクルと 3 つのモデルの中で最大のものである。しかし、その大容量キャッシュによって、ノード内のローカルメモリアクセスおよびノード間のリモートメモリアクセス時のキャッシュヒット率が最も高い。一方、オンチップメモリを集積しないため、キャッシュミス時にはオフチップメモリへ低バンド幅・高レイテンシでアクセスしなければならず、キャッシュミスペナルティは 3 つのモデルの中で最大となる。

• Type B

キャッシュ容量が 3 つのモデルの中で平均的なものであり、キャッシュアクセスレイテンシも 2 サイクルと平均的なものである。また、オンチップメモリを集積するため、キャッシュミス時には、高バンド幅・低レイテンシでメモリアクセスが可能であり、キャッシュミスペナルティが小さい。また、他ノードからリモートメモリアクセスによるデータ要求があった場合、ルータ-オンチップメモリ間の高バンド幅・低レイテンシデータ転送により、要求元へ高

^{*} 本評価モデルの作成にあたっては、ITRS による予測¹⁾、2001 年現在での DRAM 製造技術²⁾に基づいてモデルを定義するとともに、半導体技術に造詣が深い複数の企業の研究者の方々のご意見をいただき、モデルの妥当性を確認している³⁾。

速にデータ転送を行うことが可能である。

● Type C

キャッシュ容量が3つのモデルの中で最小であるため、キャッシュアクセスレイテンシも1サイクルと最小のものである。しかし、その小容量キャッシュによって、ローカルメモリアクセスおよびリモートメモリアクセス時のキャッシュヒット率が最低となる。また、オンチップメモリを集積するため、Type Bと同じ利点がある。さらに、Type Bよりも大容量のオンチップメモリを備えるため、Type Bよりも大きなサイズのローカルデータをオンチップに配置できる。

3. 性能評価

CC-NUMA 型並列計算機シミュレータ RSIM⁽⁶⁾を機能拡張し、上記の3通りの評価モデルの動作をクロックレベルでシミュレートすることにより評価を行った。

3.1 評価条件

全評価モデルにおいて、命令キャッシュへのアクセスは完全ヒット、キャッシュコヒーレンスプロトコルは無効化方式であるMESI、メッセージヘッダのサイズは16バイトとした。また、評価アプリケーションはSPLASH-2のFFT、LU、Radixの3つのアプリケーションとし、問題サイズはそれぞれ 2^{18} 個の複素数データ、 512×512 の行列、 2^{21} 個の整数キーとした。

3.2 評価結果

各評価アプリケーションごとに、ノード数が1, 2, 4, 16, 64台の場合での各評価モデルの性能向上率を図2に示す。この性能向上率は、ノード数が1台の場合におけるType Aの実行速度を基準とした。また、各評価アプリケーションごとに、各評価モデルのキャッシュミス時のメモリアクセス回数と、その内訳(リモート/ローカル)を図3に示す。このメモリアクセス回数は、全ノードで発生したメモリアクセス回数の総和である。図2から得られる評価結果を以下にまとめる。

- Type Bがアプリケーション、ノード数によらず最高性能。
 - 4台以上でのFFTの実行において、Type CがType Aより高性能。
 - LUおよびRadixの実行において、Type Cがノード数によらず最も性能が低い。
- また、図3から得られる評価結果を以下にまとめる。
- 2台以下でのFFTおよびRadixの実行を除き、Type AとType Bのキャッシュミス回数の差が小さい。
 - Type Cのキャッシュミス回数がアプリケーショ

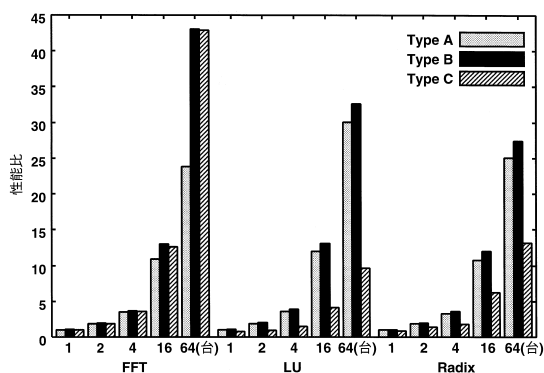


図2 Type A (1台)の実行速度を1としたときの性能向上率
Fig. 2 Performance comparison of evaluation models.

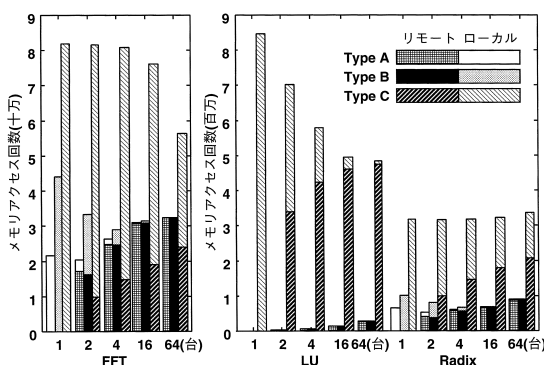


図3 キャッシュミス時のメモリアクセス回数の内訳
Fig. 3 Breakdown of the number of memory access after cache miss.

ン、ノード数によらず他のモデルと比べてきわめて多い。

- Type Cのリモートメモリアクセスの回数は、FFTの実行において最も少ないが、LUおよびRadixの実行においては最も多い。

3.3 考察

以下では、各評価モデルごとに、評価結果について考察する。

● Type A

アプリケーション、ノード数によらずType Bより性能が低い。この原因は、リモートメモリアクセスがType Bより高レイテンシ*であることによる。さらに、Type Bに対するType Aの大容量キャッシュの利点が現れなかったことも原因である。このことは、図3より、Type AとType Bの両モデ

* 表1のパラメータに従い計算を行うと、たとえば、Type A、Type Bの隣接ノード間でのレイテンシはそれぞれ108, 75サイクルであり、Type AのものがType Bのものに比べて1.44倍長い。

ルにおけるキャッシュミスの回数の差があまりないことから分かる。すなわち、本評価で用いたアプリケーションの範囲では、Type A の大容量キャッシュの優位性が現れなかった。

● Type B

アプリケーション、ノード数によらず最も性能が高い。Type A より高性能である原因は、リモートメモリアクセスが Type A より低レイテンシであることによる。Type C より高性能である原因は、キャッシュミスの回数が Type C よりきわめて少ないことによる。

● Type C

(1) 4 台以上での FFT の実行において、Type A より性能が高い。

この原因は、Type C の低レイテンシのリモートメモリアクセスが有利に働くことによる。また、総メモリアクセス回数は多いが、その中のリモートメモリアクセスの回数が Type A よりも少ないことも原因である。Type A よりも実行回数の少ない Type C のリモートアクセスは、任意のノードが共有データへの書き込みを行う際に他ノードがキャッシュに所持するコピーを無効化するものである。キャッシュ容量の最も小さい Type C では共有データのコピーが他の評価モデルよりも早期に破棄されるため、他ノードからの無効化はきわめて少ない。FFT においてリモートメモリアクセスが唯一実行される転置行列作成部では、共有データの参照の時間的局所性はなく、このようなアプリケーションでは Type C のように小さい容量のキャッシュが有利に働くことがある。

(2) LU および Radix の実行において、ノード数によらず最も性能が低い。

他の評価モデルよりもリモートメモリアクセスが頻繁に発生していることが原因である。上記の FFT の場合とは逆に、LU および Radix は共有データの参照の時間的局所性があるため、Type C の短所であるキャッシュヒット率の低さが性能低下に現れた。

以上をまとめると、ルーター-オンチップメモリ間の高バンド幅・低レイテンシデータ転送が低レイテンシのリモートメモリアクセスを実現するため、2~64 台構成の規模において、SOC 化ノードにオンチップメモリを集積する効果が現れる。ただし、アプリケーションによっては、大容量のオンチップメモリを集積してもキャッシュ容量が小さい場合は、リモートメモリアクセスの回数を増加させ、並列計算機全体の性能を低

下させる可能性がある。

4. ま と め

本稿では、2005 年時点での SOC 製造技術の予測に基づいて、CC-NUMA 型並列計算機ノードの SOC 化について考察した。その結果、2~64 台構成の規模において、1) SOC 化ノードへのオンチップ DRAM メモリを集積は、性能の面で有効であること、2) 大容量オンチップ DRAM メモリ集積による同チップ上のキャッシュの極端な小容量化は、並列計算機全体の性能を低下させる可能性があり、オンチップメモリ容量とキャッシュ容量との間にトレードオフ点が存在することの 2 点が判明した。

今後の課題として、1) 本稿で定義した全評価モデルに対してより大規模なアプリケーションでの評価、2) マルチプロセッサコアを考慮した評価があげられる。

謝辞 有益なご助言を賜りました富士通研究所の木村康則氏、安里彰氏、日立研究所の福永泰氏、堀田多加志氏、NEC インターネットシステム研究所の中田登志之氏、電気通信大学の吉永努氏に心より感謝いたします。本研究は、一部日本学術振興会科学研究費補助金基盤研究 (B) 14380135, (C) 14580362, 若手研究 (B) 14780186 の援助によるものである。

参 考 文 献

- 1) International Roadmap Committee and International Technology Working Groups: International Technology Roadmap for Semiconductors 2000 Update (2000). <http://public.itrs.net/Files/2000UpdateFinal/2kUdFinal.cfm>.
- 2) Kirihaata, T., et al.: A 390 mm² 16-bank 1 Gb DDR SDRAM with Hybrid Bitline Architecture, 1999 IEEE International Solid-State Circuits Conference (1999).
- 3) 堀田多加志, 安里 彰, 中田登志之, 福永 泰, 木村康則: private communication.
- 4) 古川文人ほか: 並列計算機ノードのシステムオンチップ化とその性能, 情報処理学会研究報告 99-ARC-134, pp.25-30 (1999).
- 5) 古川文人ほか: SOC ノードで構成する並列計算機の性能評価, 並列処理シンポジウム JSP2000 論文集, p.164 (2000).
- 6) Pai, V.S., et al.: RSIM: An Execution-Driven Simulator for ILP-Based Shared-Memory Multiprocessors and Uniprocessors, Proc. 3rd Workshop on Computer Architecture Education (1997).

(平成 14 年 2 月 1 日受付)

(平成 14 年 5 月 15 日採録)