2008－ARC－180 (4)
2008／10／20

プロセッサ性能に対する主記憶バンド幅の影響の評価

江口修平[†1] 塩谷亮太[†1]
五島正裕[†1] 坂井修一[†1]

プロセッサの処理速度の向上に伴い，主記憶のバンド幅を広げる必要性があると言われている．そのため近年，新たなメモリ・モジュールの規格が出る度に広いバンド幅を持つメモリ・モジュールが登場しており，また，複数のチャネル上のメモリ・モジュールに同時にアクセスすることによりバンド幅を広げる技術も存在する．しかし，主記憶のバンド幅の変化が，実際にどのような影響を与えるかについては，詳細な評価がなされて来なかった．そこで，今回我々は実機とシミュレーションを用いて，主記憶バンド幅がプロセッサ性能に与える影響を評価した．その結果，主記憶へのアクセスが多いプログラムでは，バンド幅の低下による大きな影響が見られたが，SPEC2000FP，INT では，バンド幅の低下による影響は大きくないという結果が得られた．

The effect of main memory bandwidth on processor performance

Shuhei Eguchi,[†1] Ryota Shioya,[†1] Masahiro Goshima[†1]
and Shuichi Sakai[†1]

It is now a popular understanding that greater bandwidth for main memory is required to fully exploit high-speed processor. The bandwidth of main memory is attracting more attentions recently, and the two following facts state this trend. One is that newer specifications of memory module always have greater bandwidth. The second is emergence of techniques such as one that widens the bandwidth by accessing multiple channels of memory modules at once. However, the effect of the bandwidth on the performance have never been precisely evaluated. In this paper, we evaluated the effect of the bandwidth on a real machine and a simulator. Our evaluation showed that the narrower bandwith lead to the decrease of performance in memory-intensive programs, but it had little effect on the SPEC 2000 INT and FP benchmarks.

1. はじめに

一般に，プロセッサの演算速度に比べて主記憶へのアクセスは遅く，主記憶のバンド幅が性能のボトルネックになる場合が多いと言われている．

主記憶のバンド幅を向上させるためには，モジュールそのもののバンド幅を広げる方法と，多数のメモリ・モジュールに同時にアクセスすることによりバンド幅を向上させる方法がある．

メモリ・モジュール，**図1**に示すように，新しい規格が出るに従って広いバンド幅のものが規格化されている[5)4)]．また，メモリ・モジュールを二つ用いることでバンド幅を倍増させる，デュアル・チャネルと呼ばれる技術もある．

しかし，プロセッサの演算速度が向上したからと言って，本当にその分主記憶へのアクセスが増加するかは疑問である．多くのロード命令が存在するプログラムでも，データがキャッシュに存在すれば主記憶へのアクセスは起こらない．

そこで，今回我々は，主記憶バンド幅がプロセッサの性能に与える影響を明らかにするために，SPEC2000[1)]ベンチマークを用いて，主記憶バンド幅を変化させた時の，プロセッサの性能の変化を評価した．

2. 評価環境

シミュレータと実機を用いて，主記憶のバンド幅がプロセッサに与える影響を測定した．

表1 主なメモリ・モジュールの規格
Table 1 Main standards of memory modules

名称	バンド幅
DDR533(PC4200)	4.2GB/s
DDR2-1066(PC2-8500)	8.5GB/s
DDR3-14400(DDR3-1800)	14.40GB/s

†1 東京大学
University of Tokyo

```
for(i = 0; i < size ; ++i){
    for(k = 0; k < size ; ++k){
        for(j = 0; j < size ; ++j){
            c[i][j] += a[i][k] * b[k][j];
        }
    }
}
```

図 1　行列積のプログラム

2.1　ベンチマーク

SPEC2000 のベンチマークの他に，**表 2** の自作のマイクロ・ベンチマーク 3 種を用いた．

主記憶リード及び主記憶コピーは，64KB のキャッシュ・ラインの一部だけにアクセスすることによって，プロセッサ側の処理を最小限にし，主記憶バンド幅が測定できるようにしている．

行列積のプログラムは，主記憶に連続アクセスをするよう，**図 1** のような中間積型の記述をしている。

実機での測定では，SPEC2000 についてはSPEC2000 に付属のツール runspec が出力したレポートに記載された実行時間を用いる．また，マイクロ・ベンチマークについては *getrusage()* 関数を用いて実行時間を測定した．

シミュレータでの測定では，全てのベンチマークで，先頭の 1G 命令をスキップしたのち後続の 100M 命令を実行した．

2.2　実　　機

実機として，AMD Athlon 64 X2[2] を搭載したマシンを使用した．Athlon X2 のアーキテクチャの概要を**図 4** に示す

主記憶は DDR2 SDRAM 667 の 1GB のモジュールを 2 枚搭載しており，メモリ・モジュール単体で 5.3GB/s のバンド幅を持つ．

さらに，これをデュアル・チャネルで動作させることにより，最大で 10.6GB/s のバンド幅となる．そして，メモリ・モジュールを差すスロットの位置を変更することにより，シングル・チャネルとデュアル・チャネルの動作を切り替えることができる．

Athlon 64 X2 の主記憶コントローラはプロセッサに内蔵しており，主記憶バスはチップセットなどを介さずプロセッサと直結している．主記憶及びキャッシュに関係するパラメータは**表 3** の通りである．その他のパラメータは**表 4** の通りである．ただし，Athlon 64 X2 はデュアルコアであり，表 4 の L1I，L1D，L2 キャッシュの組を各コアに独立して 1 組ずつ持つ．L2 キャッシュ及びメインメモリのレイテンシは実測値ある．

表 2　マイクロ・ベンチマーク
Table 2　Micro benchmarks

名前	内容
主記憶リード	64MB の int 配列の一部にアクセス
主記憶コピー	64MB の int 配列の一部をコピー
行列積	1024 × 1024 の double の行列同士の積

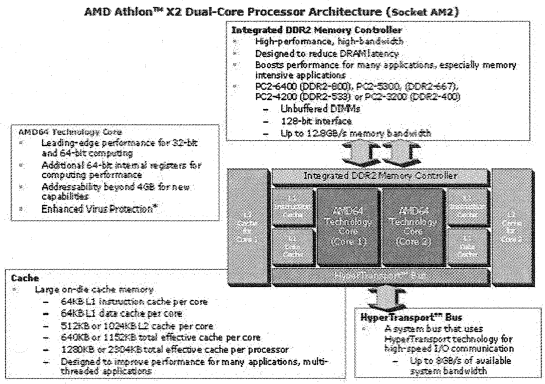

図 4　Athlon X2 のアーキテクチャ(6) の図を引用)

表 3　主記憶及びキャッシュに関係する評価パラメータ
Table 3　Parameters of main memory and caches

パラメータ	値
L1ICache	64Bytes line,2Way,64kB
	3cycles Access latency
L1DCache	64Bytes line,2Way,64kB
	3cycles Access latency
L2Cache	64Bytes line,16Way,1MB
	25cycles Access latency
主記憶	DDR2-5300 × 2 2GB,
	10.6GB/s(デュアル・チャネル)
	5.3GB/s(シングル・チャネル) (以上実機)
	200 cycles Access Latency
	64Bytes / 30cycles 又は 64Bytes / 15cycles (以上シミュレータ)

表 4　実機の評価パラメータ
Table 4　Parameters of Real Machine

パラメータ	値
プロセッサ	AMD Athlon X2 5200+
動作周波数	2.6GHz
コンパイラ	gcc3.4.4
OS	CentOS 4.7(64bit 版)

2.3　シミュレータ

シミュレーションには，本研究室で開発したシミュレータ「鬼斬弐」を用いた．

主記憶及びキャッシュに関係するパラメータは表 3 の通りであり，実機に合わせることができる部分はできるだけ実機に合わせてある．その他のパラメータは

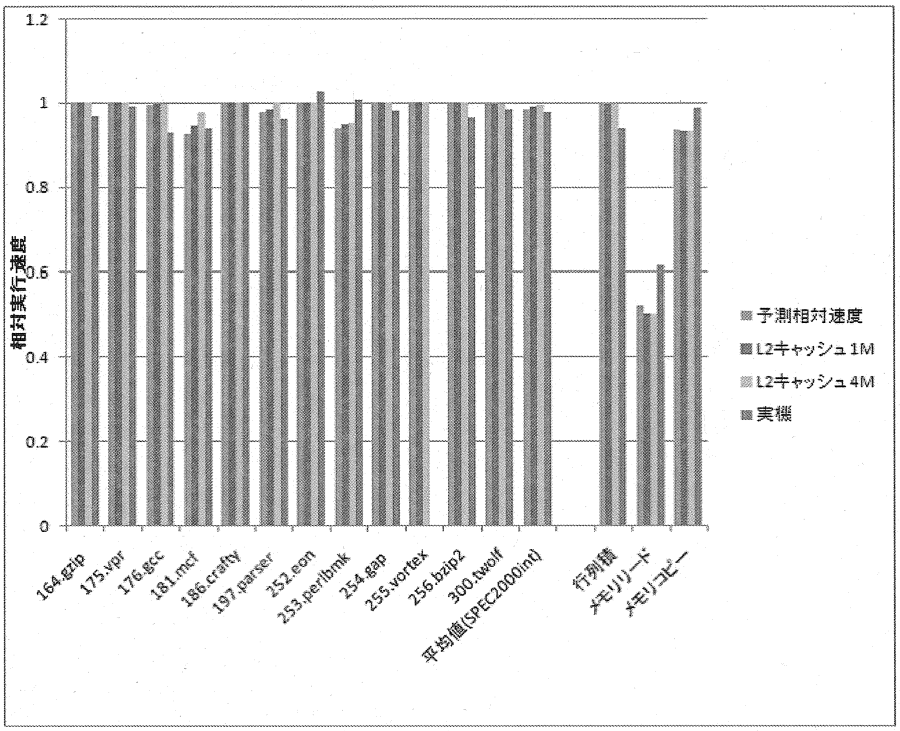

図 2 シミュレータと実機で主記憶バンド幅を半減させた時の実行速度の低下率と主記憶アクセス頻度 (SPEC2000INT 及び マイクロ・ベンチマーク)

表 5 シミュレータの評価パラメータ
Table 5 Parameters for Simulator

パラメータ	値
Simulator	鬼斬弐
ISA	alpha
Fetch Width	4
Issue Width	INT:4 FP:2 Mem:2
Integer Units	ALU:2 MUL:1,DIV:1
Floating Point Units	ADD:1 MUL:1 DIV :1
Register File	INT: 192 FP:128
Instruction Windows	INT:32 FP:32 Mem:32

表 6 の通りである.

実機で主記憶バンド幅を半減させたことを再現するために，シミュレータ上でも主記憶バンド幅を半減させた。

ただし，シミュレータ上では主記憶から L2 キャッシュにキャッシュ・ラインが転送することができる時間間隔を二倍に伸ばしただけで，デュアル・チャネルおよびシングル・チャネルでのアクセスのタイミングまで再現したわけではないことに注意が必要である．また，L2 キャッシュから主記憶にキャッシュ・ラインが追い出されることによってバンド幅が占有されることについては考慮していない.

3. 評　　価

3.1 相対実行速度の変化

図 2 及び**図 3** はシミュレータおよび実機で，バンド幅を半減させる前の実行速度を 1 として，主記憶のバンド幅を半減させた時の相対実行速度と，主記憶へのアクセスの間隔から予測される相対実行速度を表したものである。マイクロ・ベンチマークでの結果は図 2 に含めてある．さらに，キャッシュ・サイズが各ベンチマークに与える影響を調べるため，L2 キャッシュが 1MB でのシミュレーション結果の他に 4MB でのシミュレーション結果も掲載している．また，相対実行速度の低下率をまとめたものが表 6 である.

表 6 の通り，SPEC2000INT では，低下率は実機で 1.1%，シミュレータで 2.4%にとどまり，主記憶バンド幅が半減した影響はほとんど無いといえる．また，SPEC2000FP では，179.art など大きく低下したベンチマークもあったが，平均すると低下率は実機で 3.5%，シミュレータで 5.1%であり，INT よりは影響は大きいものの，それほど大きな影響はないといえる.

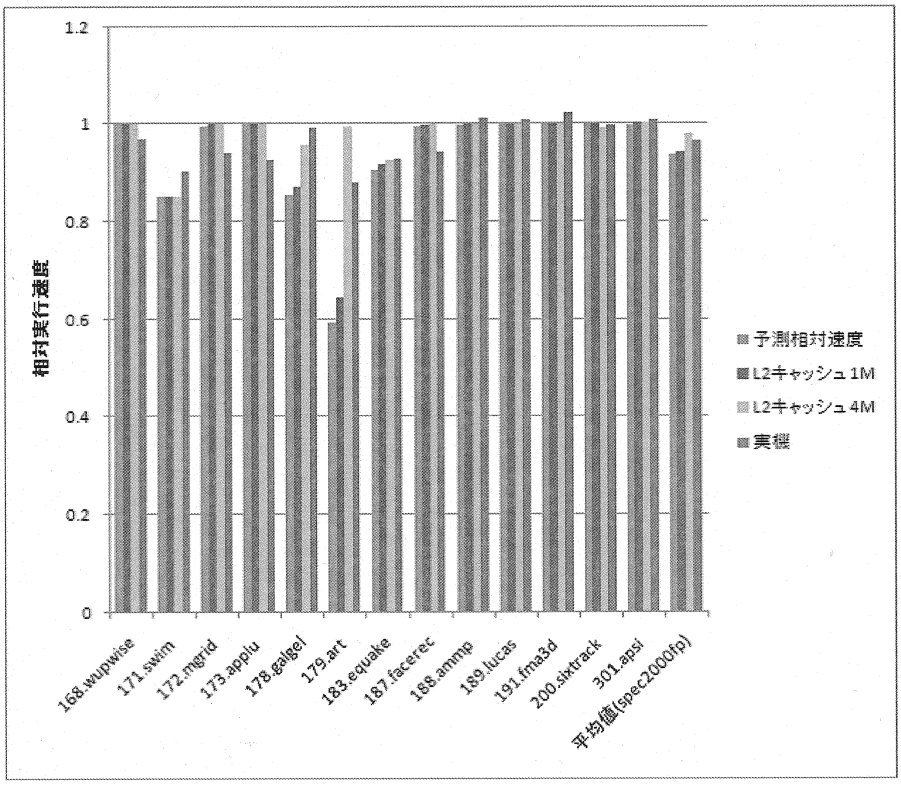

図 3 シミュレータと実機で主記憶バンド幅を半減させた時の実行速度の低下率と主記憶アクセス頻度 (SPEC2000FP)

表 6 実行速度の低下率
Table 6 Decresing rate of execution speed

ベンチマーク	シミュレータ	実機
SPEC2000int	1.1%	2.4%
SPEC2000fp	5.1%	3.5%
行列積	4.1%	6.1%
主記憶リード	50.0%	38.7%
主記憶コピー	7.9 %	1.3%

また，SPEC2000int での低下率が小さくそれに比べると FP の低下率が大きいという傾向は，実機とシミュレーションで同じであった．また，171.swim,179.art や主記憶リードなど，シミュレーションで IPC の低下率が大きいベンチマークは実行速度の低下も大きいということが分かった．

大きく影響を受けたベンチマークは，主記憶アクセス の間隔が小さいものが多いものと考えられるが，このことについては次の節で明らかにする．

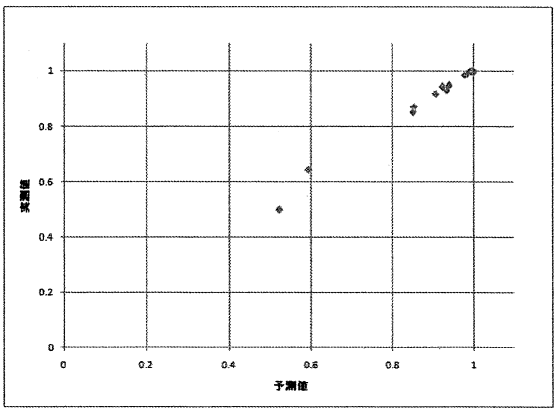

図 5 相対実行速度の予測値と実測値の相関

3.2 実行速度への影響が小さかった理由についての考察

シミュレータではバンド幅を表現するために 15 サイクルまたは 30 サイクルに 1 回しか主記憶から L2

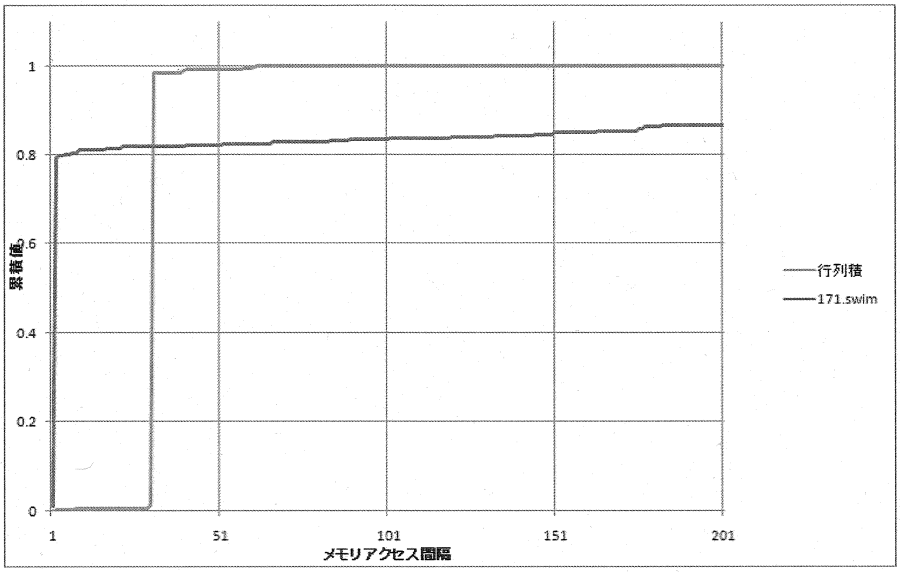

図 6 主記憶アクセス間隔の分布

キャッシュにデータを転送しないようにしている．このことから，主記憶へのアクセスの間隔の統計を取れば，バンド幅が制限されることにより，どれだけ実行サイクル数が増加するか予測することができ，これから相対実行速度の低下率も求めることができる．これを表したものが，図 2 と図 3 の中の,「予測相対速度」である．

短い間隔での主記憶アクセスが多ければ多いほど，バンド幅を減少させた時に，大きく実行速度が低下すると予測される．

179.art や主記憶リードでは予測相対速度の低下率が大きいことが示されているが，実際の実行速度の低下率も大きいということが分かる．**図 5** は予測値と相対実行速度の低下率の分布を表したものであるが，全ての点がほぼ一直線上に並んでいる．このことから，間隔が短い主記憶アクセスの多いベンチマークがバンド幅の影響を受けやすいという仮定は妥当であると思われる．

主記憶へのアクセス間隔の傾向を詳細に見るために，行列積と 171.swim のベンチマークについてアクセス間隔の統計を取ったグラフが**図 6** である．横軸に主記憶アクセス間隔を取り，縦軸にその間隔でのアクセスが全体のアクセス数に対してどれだけ起こったかの割合の累積値 を表す．アクセス頻度が高かったのに低下率が少なかった行列積と，アクセス頻度が低かったのに低下率が大きかった 171.swim を選んだ．

このグラフの行列積のベンチマークのグラフを見ると，主記憶アクセス間隔が 30 を超えたあたりで急に立ち上がっていることが分かる．

これは，主記憶へのアクセスが規則的に起こっていることを示している．今回，主記憶バンド幅を半減させた時は，30 サイクルに一回だけ主記憶から L2 キャッシュにデータを転送できることにしているので，間隔 30 サイクルを下回るアクセスがほとんど無ければ，バンド幅の影響を受けない．行列積では，30 サイクル未満でのアクセスは 1.3%に過ぎなかった．一方，171.swim のベンチマークを見ると，主記憶アクセス間隔が 1 のところで急に立ち上がっている．そして，アクセス間隔が 30 サイクル未満でのアクセスは 82%にも上る．

このように，主記憶アクセスの間隔の分布の偏りによって，バンド幅の影響の受けやすさに違いがあることが分かった．

4. おわりに

主記憶バンド幅がプロセッサに与える影響を明らかにするため，実機及びシミュレータで主記憶バンド幅を変化させて，ベンチマークの相対実行速度を測定した．その結果，主記憶バンド幅を半減させたにも関わらず，相対実行速度の低下率はシミュレータでは INT で 1.1%，FP で 5.1%，実機では INT で 2.4%，FP で 3.5% にとどまった．また，シミュレータと実機で

の相対実行速度の低下率には相関が見られることも分かった．

さらに，主記憶バンド幅からの影響の受けやすさは短い間隔での主記憶アクセスがどれだけあるかによって決まり，単純に主記憶アクセスの回数の多さだけでは決まらないということが分かった．そして，多くのベンチマークでは短い間隔での主記憶アクセスは少ないため，主記憶バンド幅の影響を受けるベンチマークが少なかったということが分かった．

謝　　辞

本論文の研究は，一部半導体理工学研究センター（STARC）及び文部科学省科学研究費補助金（特定領域研究）No. 19024020 による

参 考 文 献

1) Standard Performance Evaluation Corporation,http://www.spec.org
2) AMD Athlon X2 Dual-Core Processor Product Data Sheet, http:// www. amd. com/ us-en/assets/conten_type/white_papers_and_tech_docs/43042.pdf
3) BIOS and Kernel Developer's Guide for AMD NPT Family 0Fh Processors, http://www.amd.com/us-en/assets/content_type/white_papers_and_tech_docs/32559.pdf
4) JEDEC STANDARD DDR3 SDRAM Specification JESD79-3A,2007
5) JEDEC STANDARD DDR2 SDRAM Specification JESD79-2C,2006
6) AMD Athlon X2 デュアルコア・プロセッサ アーキテクチャの特長, http://www.amd.com/jp-ja/Processors/ProductInformation/0,,30_118_9485_13041%5E13043,00.html