

## 計算機システムにおける分岐予測外れパスの影響

大河原 英喜<sup>†</sup> 河場 基行<sup>†</sup> 安里 彰<sup>†</sup>

計算機システムの性能評価に用いられるシミュレータには、実行ドリブン型シミュレータ (EDS) とトレースドリブン型シミュレータ (TDS) の大きく 2 種類がある。TDS を用いた性能評価は、動的な振る舞いが評価できず、評価誤差が生じるという問題点が知られている。本論文では、TDS において分岐予測外れパスが評価されないことに着目し、分岐予測外れパスが計算機システムに如何に影響を与えるかについて調査した。評価の結果、極端に違いが現れる例では、分岐予測外れパスの命令をシミュレーションするか否かで、実行サイクル数に約 93% の違いが見られた。また、SPEC INT92 で評価を行った結果、fetch 命令数で 30%、実行サイクル数では 3% 程度の違いが見られた。

### The influence of branch miss-predicted path to computer system performance

HIDEKI OKAWARA,<sup>†</sup> MOTOYUKI KAWABA<sup>†</sup> and AKIRA ASATO<sup>†</sup>

To evaluate computer system performance, two types of simulators, Execution Driven Simulator (EDS) and Trace Driven Simulator (TDS), are commonly used. The drawback of TDS is its inability of simulating dynamic behavior, for example, speculative instructions fetching. In this paper, we focus on speculative instruction fetching and investigate the influence of branch miss-predicted path to the computer system performance. In an extreme case, the evaluation error of TDS is about 93% in execution time. In SPEC INT92 benchmark, the evaluation errors are 30% in the number of instructions fetched and 3% in the execution time.

#### 1. はじめに

プロセッサのマイクロアーキテクチャやメモリアーキテクチャなど、計算機システムの性能評価にはソフトウェアシミュレータが用いられることが多い。ソフトウェアシミュレータには、シミュレーション方式において、大きく下に示す 2 種類がある。

**実行ドリブン型シミュレータ (EDS)<sup>1)2)</sup>** 性能評価時に命令を動的にエミュレーションし、それと連動してシミュレーションを行う方式。動的に振る舞いに変化する場合にも正しく評価することができる。

**トレースドリブン型シミュレータ (TDS)** 予め採取された、ある環境下における実行時の命令列のトレースデータを基にシミュレーションを行う方式。動的に振る舞いに変化する場合は評価できない。

TDS で正しく評価できない振る舞いの具体的な例として、以下の様なものが挙げられ、これらを解決する TDS 手法の研究も行われている<sup>3)</sup>。

- マルチプロセッサシステムにおけるロック変数獲得の様な、処理タイミングによって命令フローが変化するアプリケーションの評価ができない。
- 分岐予測ミス時に誤って投機実行される、分岐予測

外れパスがシミュレーションに含まれず、評価誤差につながる。

分岐予測外れパスがシミュレーションに含まれないという問題点はよく知られているが、その影響に関する定量的な評価はあまりされていない。単体プロセッサにおいて分岐予測外れパスの影響を評価した研究<sup>4)</sup>では、分岐予測外れパスにおけるキャッシュミスがプリフェッチ効果を持つことを示し、EDS と TDS とでキャッシュの振る舞いに違いがあることを示した。この評価では、分岐予測外れパスが IPC に与える影響は小さいという結論が得られた。

しかし、分岐予測外れパスによってキャッシュミスが頻発する場合や、メモリアクセスが性能のボトルネックとなるマルチプロセッサシステムでは、分岐予測外れパスの影響はより大きいと考えられる。そこで、本論文では単体プロセッサに加えマルチプロセッサシステムにおける分岐予測外れパスの影響に関する評価を行う。

#### 2. 分岐予測ミス時の動作について

本章では、分岐予測ミス時の EDS と TDS の動作の違いについて、一例をとりあげて述べる。

分岐予測機構を備えたプロセッサにおいては、分岐予測ミス時にも投機実行が行われる。EDS では分岐予測外れパスもシミュレーションされるが、TDS では分岐

<sup>†</sup> (株) 富士通研究所

FUJITSU LABORATORIES LTD.

予測外れパスはシミュレーションされず、評価誤差につながる。分岐予測ミス時の実機、EDSでのパイプラインの動作例を図1に、TDSでの動作例を図2に示す。これらの例では、命令0の分岐命令が分岐予測ミスを起こすとする。

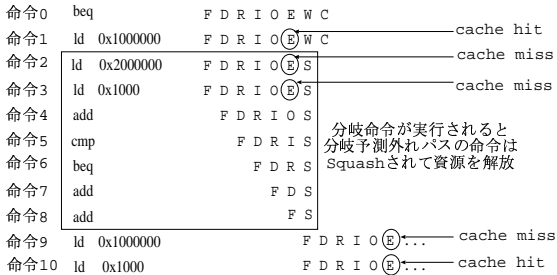


図1 分岐予測ミス時の動作例 (実機、EDS)

実機やEDSでは、分岐予測外れパスの命令(命令2~命令8)が投機的にfetchされパイプラインに投入される。分岐命令(命令0)が実行され、正しい分岐方向が確定したら、分岐予測外れパスの命令はsquashされ、Reservation Station(RS)やrenaming registerなど、確保していた資源の解放を行う。但し、一度、出されたメモリアクセス要求に関しては処理が続行される。

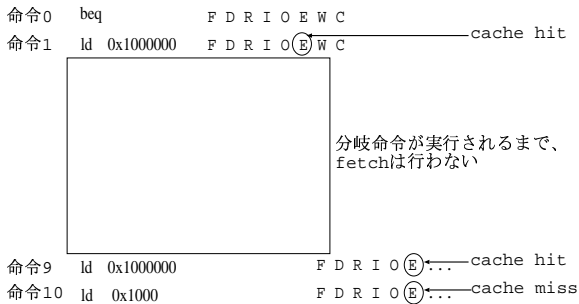


図2 分岐予測ミス時の動作例 (TDS)

それに対し、TDSでは分岐外れパスの命令がトレースデータに含まれないため、分岐命令(命令0)が実行されて正しい分岐方向が確定するまで、シミュレーションは行われぬ。

この様に、EDSとTDSとは、分岐予測外れパスの命令が実行されるか否かという違いがあり、EDSの方がより正確なシミュレーションを行っている。この、分岐予測外れパスが実行されるか否かの違いが計算機システムに与える影響として以下の点が挙げられる。

**キャッシュ状態の変化** 分岐予測外れパス中にあるメモリアクセス命令がシミュレーションされるか否かでキャッシュ状態の違いが現れる。図1、図2の例では、以下の2通りのケースを示している。

- 命令1の時点ではアドレス0x1000000のデー

タはキャッシュ上にあるとする。EDSでは、命令2が実行されるとアドレス0x1000000とアドレス0x2000000が干渉し、アドレス0x1000000のデータはキャッシュ上から追い出される。そのため、命令9ではアドレス0x1000000はキャッシュミスとなる。一方、TDSでは命令9においてアドレス0x1000000はキャッシュヒットとなる。

- EDSでは、命令3が実行されるとキャッシュミスを起こしてアドレス0x1000のデータを読み込む。そのため、命令10ではアドレス0x1000はキャッシュヒットとなる。一方、TDSでは命令10がアドレス0x1000への初アクセスであり、この時点でキャッシュミスを起こす。この様に、分岐予測外れパスにおけるメモリアクセスがプリフェッチ効果を持つ場合がある。

**分岐予測外れパスによる資源の枯渇** EDSでは、パイプラインに投入された分岐予測外れパスの命令がハードウェア資源の確保を行い、squashされて資源を解放するまでの間、資源が枯渇する可能性がある。それに対し、TDSでは分岐予測外れパスの命令はパイプラインに投入されず、分岐予測外れパスの命令による資源の枯渇は起きない。

近年の計算機システムの動向を考えると、プロセッサ内部の資源の枯渇よりも、メモリアクセスがより性能に影響を与えると考えられる。そこで本論文では、特に、分岐外れパスによるキャッシュ状態の変化に着目し、分岐予測外れパスが性能に与える影響について評価する。

### 3. 分岐予測外れパスの影響に関する評価

#### 3.1 評価環境

##### 3.1.1 SEEDS (SPARC Emulator and Execution Driven Simulator)

本節では、我々が開発した実行ドリブン型シミュレータ、SEEDSの概要を述べる。本論文における性能評価はこのSEEDSを用いて行った。

##### 3.1.1.1 SEEDSの構成

SEEDSは、SPARC V9命令セットに対応した実行ドリブン型シミュレータで、マルチプロセッサシステム、マルチスレッドアプリケーションのシミュレーションを行うことができる。SEEDSの構成を図3に示す。

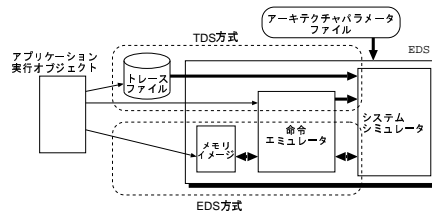


図3 SEEDSの構成

SEEDSは、SPARC V9対応の命令エミュレータ部、サイクルベースでパイプライン動作やキャッシュ動作などをシミュレーションするシステムシミュレータ部から構成される。SEEDSでは、入力ファイルや分岐予測ミス時の挙動に関して、以下の3通りの方式のシミュレーションを行うことができる。

**EDS方式** 実行オブジェクトを入力とし、動的に命令をエミュレーションしながらシミュレーションを行う。分岐予測外れパスのシミュレーションも行い、分岐方向が確定されたら分岐命令時の正しい状態を復帰し、正しいパスの命令から再開する。

**TDS-e方式** 実行オブジェクトを入力とし、動的に命令をエミュレーションする。分岐予測外れパスのシミュレーションは行わず、分岐方向が確定するまで命令 *fetch* は行わない。

**TDS-t方式** 従来の *TDS* と同様にトレースファイルを入力とし、分岐予測外れパスのシミュレーションは行わない。

3.2節以降では、EDS方式とTDS-e方式のシミュレーション結果の比較を行うことで、計算機システムにおける分岐予測外れパスの影響を評価した。どちらのシミュレーション方式においても、実行オブジェクトを基に動的に命令をエミュレーションしているため、両者の違いは、分岐予測外れパスのシミュレーションを行うか否かという点だけである。以後、TDS-e方式のことを単にTDS方式と記述する。

### 3.1.1.2 システムシミュレータ部の概要

システムシミュレータ部の概要を図4に示す。システムシミュレータ部は、パイプライン動作をシミュレーションするCPUモデルや、キャッシュ動作、メモリアクセス動作をシミュレーションするメモリモデルから構成されている。

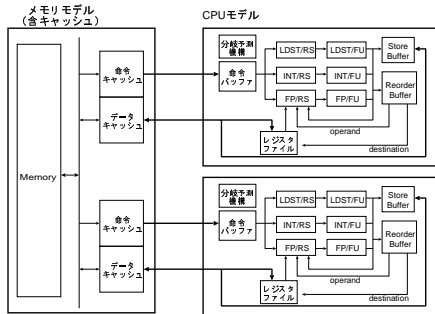


図4 システムシミュレータ部の概要

CPUモデルの主な特徴は以下の点である。

- 8段パイプラインのスーパースカラプロセッサ
- *out-of-order issue, in-order commit*
- 動的分岐予測機構 (*gshare*)

メモリモデルの主な特徴は以下の点である。

- マルチプロセッサに対応

- バスアーキテクチャ、キャッシュプロトコル (*MOESI*プロトコル) を考慮したシミュレーション
- キャッシュは3階層まで

### 3.2 サンプルプログラムによる評価

本節では、分岐予測外れパスの影響が大きく現れるサンプルプログラムを作成し、最悪なケースでの分岐予測外れパスの影響を調べる。

#### 3.2.1 サンプルプログラムの概要

分岐予測外れパスの影響が大きく現れる場合として、以下の様な場合が考えられる。

- 分岐予測ミス率が高く、分岐予測外れパスの命令が多く実行される場合
- 分岐予測外れパスでのキャッシュアクセスによりキャッシュミス数が増加する場合

これらのことを考慮し、予測外れパスでキャッシュミスが頻発する例として、以下に示すサンプルプログラムを用いて評価を行った。

```
int P[4] = {1, 0, 1, 0};
int sum, array[STRIDE/sizeof(int)+1];

main() {
    int i;
    for (i = 0; i < 10000; i++) {
        if (P[i%4]) {
            sum += *(array + STRIDE*(1-P[i%4])); /* 処理1 */
        } else {
            sum += *(array + STRIDE*P[i%4]); /* 処理2 */
        }
    }
}
```

このサンプルプログラムでは、表1に示す様に、*if*文での分岐予測がヒットすれば  $*(array)$  を、ミスすれば  $*(array+STRIDE)$  をアクセスする。*STRIDE* がキャッシュサイズと等しい場合には、 $*(array)$  と  $*(array+STRIDE)$  は同一キャッシュラインとなる。そのため、連想度が *direct-mapped* であると仮定すれば、分岐予測外れパスが実行されると以下の様に2回のキャッシュミスが発生すると考えられる。

- (1) 分岐予測外れパスにおいて  $*(array+STRIDE)$  がキャッシュミスを起こし、 $*(array)$  をキャッシュ上から追い出す。
- (2) その後、正しいパスに復帰した際に、 $*(array)$  へのアクセスがキャッシュミスを起こす。

表1 サンプルプログラムのメモリアクセスパターン)

P[i%4]	分岐予測	処理	アクセスアドレス
1	ヒット	処理1	$*(array + STRIDE*(1-1))$
1	ミス	処理2	$*(array + STRIDE*1)$
0	ヒット	処理2	$*(array + STRIDE*0)$
0	ミス	処理1	$*(array + STRIDE*(1-0))$

#### 3.2.2 評価条件

3.2.1に示したサンプルプログラムを用いて、最悪なケースでの分岐予測外れパスの影響を評価した。評価

は、UP(単一プロセッサ)、MP(マルチプロセッサ)について行った。MPとは、4CPU上でサンプルプログラムを1プロセスずつ、計4プロセスを同時に実行した場合である。評価に用いたCPUの主なパラメータを表2に示す。

fetch幅	8命令/cycle
演算器(int,fp,ldst,br)	(4,4,4,1)
RS size(int,fp,ldst,br)	(32,16,24,16)
レジスタ(int,fp,cc,y)	(32,32,32,1)
1次キャッシュ	I1 + D1(各64kB,direct) レーテンシ 17 cycle
外部キャッシュ	unify(2MB,direct) レーテンシ 100 cycle

ここではキャッシュは2階層とし、\*(array)と\*(array+STRIDE)がキャッシュ干渉を起こしやすい様、1次キャッシュ、外部キャッシュの連想度をdirect mappedとした。サンプルプログラムのSTRIDEの値を以下の3通りに変化させて評価を行った。

**32kB** \*(array)と\*(array+STRIDE)はキャッシュ干渉を起こさない

**64kB** \*(array)と\*(array+STRIDE)は1次キャッシュで干渉を起こす

**2MB** \*(array)と\*(array+STRIDE)は1次キャッシュ、外部キャッシュで干渉を起こす

### 3.2.3 性能評価(サンプルプログラム)

サンプルプログラムにおけるEDS方式とTDS方式のシミュレーション結果の比較を、UPの場合を図5、MPの場合を図6に示す。図中には、EDS方式のシミュレーション結果を基準(=1)とした場合の、TDS方式のシミュレーション結果の相対値を示している。

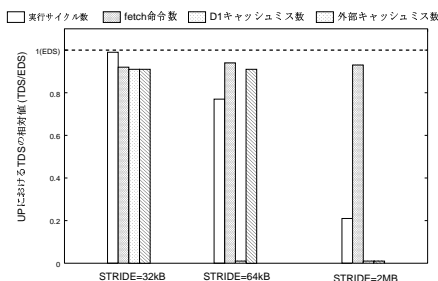


図5 サンプルプログラムにおけるTDSとEDSの違い(UP)

3種類のSTRIDEに関して比較すると、いずれの場合にもfetch命令数は約7%の差が見られる。キャッシュミス数は、STRIDEが64kBの場合には1次キャッシュ、STRIDEが2MBの場合には1次キャッシュと外部キャッシュにおいて、EDS方式とTDS方式に非常に大きな違いが見られる。実行サイクル数に関しては、STRIDEが32kBの場合にはEDS方式とTDS方式に

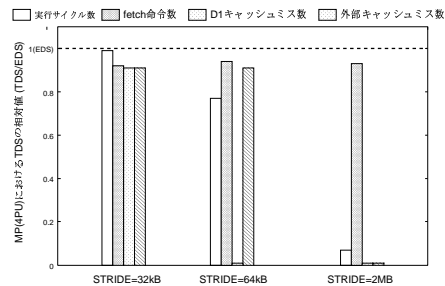


図6 サンプルプログラムにおけるTDSとEDSの違い(MP)

はほとんど違いがない。しかし、STRIDEが64kBの場合には約23%、STRIDEが2MBの場合には約79%もの違いが見られる。

MPの場合にも、UPの場合とほぼ同じ結果が得られたが、STRIDEが2MBの場合の実行サイクル数において、MPの方がEDS方式とTDS方式の違いが大きく、約93%の違いが見られる。

## 3.3 ベンチマークを用いた性能評価

### 3.3.1 評価条件

本節では、分岐予測外れパスの影響の、より一般的な評価として、ベンチマークを用いた評価を行った。ベンチマークとしては、シングルスレッドアプリケーションとして、分岐予測ミス率が大きいolden(mst,health)<sup>5)6)</sup>,SPEC INT92(espresso,gcc,li)と、キャッシュミス率が大きいCPU2000(art)を用いた。又、マルチスレッドアプリケーションとして、SPLASH2(fft,lu,radix,cholesky)<sup>7)</sup>も評価に用いた。

主な評価パラメータを表3に示す。3.2節の評価ではキャッシュミスを頻発させるためにキャッシュの連想度をdirect-mappedとしたが、ここではより一般的なパラメータとして、1次キャッシュの連想度を4way、外部キャッシュの連想度を2wayとした。

CPU数	1 CPU
CPU fetch幅	8命令/cycle
演算器(int,fp,ldst,br)	(4,4,4,1)
RS size(int,fp,ldst,br)	(32,16,24,16)
レジスタ(int,fp,cc,y)	(32,32,32,1)
1次キャッシュ	I1 + D1(各64kB,4way) レーテンシ 17 cycle
外部キャッシュ	unify(2MB,2way) レーテンシ 100 cycle

3.2節と同様に、UPとMP(4CPU)の2種類の計算機システムに対して評価を行った。MPモデルの評価に関しては、シングルスレッドアプリケーションでは同一ベンチマークを各CPU上で1プロセスずつ、計4プロセスを同時実行した。マルチスレッドアプリケーションであるSPLASH2では、主要部を4スレッドに分割し、4CPU上で1スレッドずつ処理を行った。

### 3.3.2 評価結果 (SPEC,olden)

olden, SPEC INT92, art を用いたシミュレーションにおける、EDS方式とTDS方式のシミュレーション結果の比較を、図7(UP)、図8(MP)に示す。

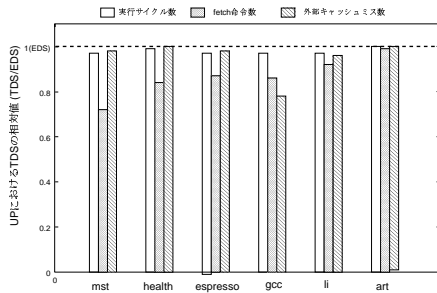


図7 SPEC,oldenにおけるEDSとTDSの違い(UP)

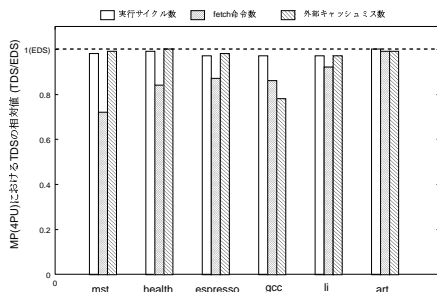


図8 SPEC,oldenにおけるEDSとTDSの違い(MP)

図7、図8を比較すると、UP,MPの結果に大きな違いは見られなかった。各ベンチマークの結果を比較すると、fetch命令数に関しては、最大で30%近くの違いが得られ、olden(mst,health)に最も大きな違いが見られた。しかし、実行サイクル数に関してはいずれのベンチマークでも大きな違いは見られず、EDS方式とTDS方式の実行サイクル数の違いは最大でも約3%であった。外部キャッシュミス数の違いはgccが最も大きく、約20%の違いが見られる。他のベンチマークに関しては、外部キャッシュミス数の違いは小さく、5%未満の違いしか見られなかった。

### 3.3.3 評価結果 (SPLASH2)

fft, lu, radix, cholesky の4種類について評価を行った結果、EDS方式とTDS方式の実行サイクル数の違いは最大1.5%であった。また、fetch命令数、外部キャッシュミス数の違いも小さく、fetch命令数で3.5%、外部キャッシュミス数で3%の違いしか見られなかった。

## 4. 考 察

### 4.1 サンプルプログラムによる性能評価

サンプルプログラムによる評価(図5、図6)では、STRIDEが64kB,2MBの場合にEDS方式とTDS方式の実行サイクル数に違いが見られた。これは、キャッシュミス数の違いが原因として考えられる。ここで、プ

ログラムの性質を明らかにするために、EDS方式、TDS方式における分岐予測ミス率、1次キャッシュミス率、外部キャッシュミス率を、図9に示す。各ミス率は、(ミス数)/(全commit命令数)として求めた値である。

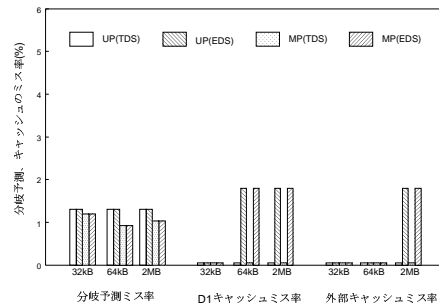


図9 サンプルプログラムの性質

分岐予測ミス率は、いずれの場合にも1%前後である。キャッシュミス率を見ると、STRIDE=32kBの場合には非常に小さいが、STRIDEが64kBの場合の1次キャッシュ、STRIDEが2MBの場合の1次キャッシュと外部キャッシュは、2%近くのキャッシュミス率となっている。この、キャッシュミス率にして2%近くの違いが実行サイクル数に大きく影響を与えている。

また、STRIDEが2MBの場合には、外部キャッシュミスによる、メモリやバスに対するアクセスの増加が実行サイクル数に影響を与えている。この様な場合、マルチプロセッサシステムでは単一プロセッサより分岐予測外れバスの影響がより顕著に現れている。

### 4.2 ベンチマークによる性能評価

ベンチマークによる評価(図7、図8)において、fetch命令数や外部キャッシュミス数のEDS方式とTDS方式の違いが、実行サイクル数にそれほど影響を与えない原因として、以下の様な点が考えられる。

- 分岐予測ミス率が小さく、分岐予測外れバスとして実行される命令数そのものが少ない
- キャッシュミス率が小さく、キャッシュミス率(キャッシュミス数/全commit数)の違いとしてはそれほど大きくない。

そこで、各ベンチマークにおける、分岐予測ミス率を図10に、外部キャッシュミス率を図11に示す。

分岐予測ミス率は、mstの3.2%が最も大きく、health, espressom, gcc, liは2%前後である。artの分岐予測ミス率は0.1%未満であった。

外部キャッシュミス率は、artの5.5%が最も大きく、次いでhealthの1.7%であった。espresso, gcc, liのキャッシュミス率は0.1%未満と非常に小さかった。

キャッシュミス率の高いartは分岐予測ミス率が低く、図7、図8において、EDS方式とTDS方式のfetch命令数に差が見られない。逆に、分岐予測ミス率が高いmst, espresso, gcc, liでは、fetch命令数、キャッシュミス数にEDS方式とTDS方式の差があるものの、キャッ

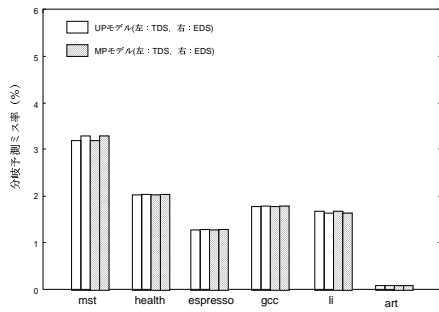


図 10 各ベンチマークの分岐予測ミス率

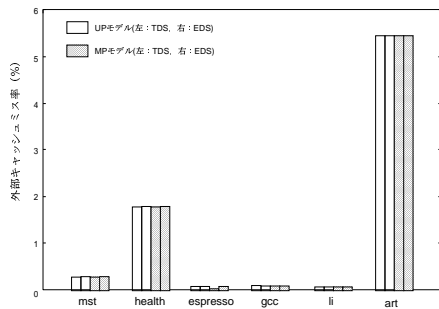


図 11 各ベンチマークの外部キャッシュミス率

シミュミス率が低く、キャッシュミス数の違いが実行サイクル数に与える影響は小さいと考えられる。

*health* に関しては、分岐予測ミス率、キャッシュミス率は、共に大きい、*EDS*方式と*TDS*方式のキャッシュミス数には違いが見られない。*health*では分岐予測外れパスでキャッシュミスが発生しないため、実行サイクル数に与える影響はそれほどないと考えられる。

*SPLASH2*において*EDS*方式と*TDS*方式の違いが見られない原因も、表 4に示す様に分岐予測ミス率、外部キャッシュミス率が小さいためであると考えられる。

表 4 *SPLASH2* の特性

	fft	lu	radix	cholesky
分岐予測ミス率	0.28%	0.65%	0.001%	1.6%
キャッシュミス率	0.02%	0.001%	0.001%	0.003%

この様に、分岐予測ミス率、キャッシュミス数が低く、分岐予測外れパスの影響が *CPU* 内で吸収される場合には、マルチプロセッサにおいても、単体 *CPU* と同様に実行サイクル数の差は見られない。

## 5. ま と め

本論文では、分岐予測外れパスがシミュレーションされるか否かという実行ドリブン型シミュレータ (*EDS*) とトレースドリブン型シミュレータ (*TDS*) の違いに注目し、分岐予測外れパスが計算機システムに及ぼす影響について評価を行った。分岐予測外れパスがシミュレーションされる *EDS*方式とシミュレーションされない *TDS*方式のシミュレーション結果の比較を行なった

ところ、分岐予測外れパスで外部キャッシュミスが頻発するサンプルプログラムにおいて、単一プロセッサシステムで 79%、マルチプロセッサシステムで 93%の実行サイクル数の違いが見られ、分岐予測外れパスが性能に大きく影響を与えることを示した。また、*SPEC INT*などのベンチマークを用いた評価の結果、分岐予測外れパスが実行サイクル数に与える影響はそれほど大きくなかった。これは、今回評価に用いたベンチマークが、分岐予測ミス率、外部キャッシュミス率が小さいためであると考えられる。この様に、分岐予測外れパスが実行サイクル数に影響を与えない場合もあるが、分岐予測ミス率、キャッシュミス率が大きい場合には、分岐予測外れパスが実行サイクル数に影響を及ぼす可能性がある。

今回は、分岐予測外れパスの影響として、キャッシュ状態が変化する点に注目したが、ハードウェア資源利用状況の変化も考えられる。資源利用の状況の違いについては、*multi threading* や *SMT (Simultaneous Multi Threading)*<sup>8)</sup> といった、複数スレッドでハードウェア資源を共有するアーキテクチャにおいて、より問題となると考えられる。今後は、本論文で開発した *SEEDS*シミュレータを用いて *SMT*アーキテクチャの評価を行なうと同時に、*SMT*アーキテクチャにおける分岐予測外れパスの影響についても調査していく予定である。

## 参 考 文 献

- 1) *SMTSIM*.  
<http://www-cse.ucsd.edu/users/tullsen/smtsim.html>.
- 2) *SimpleScalar*. <http://www.simplescalar.org/>.
- 3) L.K.John R.Bhargava and F.Matus. *Accurately modeling speculative instruction fetching in trace-driven simulation*. *IPCCC'99*, pp. 65-71, 1998.
- 4) J.D.Wellman M.Moudgill and J.Moreno. *An approach for quantifying the impact of not simulating mispredicted paths*. *Workshop on Performance Analysis and its Impact on Design, 1998*.
- 5) J.H.Reppy A. Rogers, M.C.Carlisle and L.J.Hendren. *Supporting dynamic data structures on distributed-memory machines*. *ACM Trans. on Programming Languages and Systems*, pp. 233-263, 1995.
- 6) A. Moshovos A.Roth and G.S.Sohi. *Dependence based prefetching for linked data structures*. *Proc. on ASPLOS-8, 1998*.
- 7) *Stanford Parallel Application for Shared Memory*.  
<http://www-flash.stanford.edu/apps/SPLASH>.
- 8) Susan J.Eggers Dean M.Tullsen and Henry M.Levy. *Simultaneous multithreading: Maximizing on-chip parallelism*. *ISCA'95, 1995*.