メモリアクセスの特徴を活用した高速かつ正確なメモリアーキテクチャ・シミュレーション法

小野貴継[†] 井上弘士[††] 村上和彰[††]

本稿では，高速かつ正確なメモリアーキテクチャ・シミュレーション法を提案する．一般に，メモリアーキテクチャの評価には，メモリ参照のアドレス・トレースに基づいたシミュレーションを行う．しかしながら，評価対象の増加により，評価時間が長くなる傾向にある．トレースに基づくシミュレーションにおいて，1 回あたりのシミュレーション時間はアドレス・トレースの削減によって短縮できるが，精度が低下するという問題がある．そこで，本手法はメモリアクセスの特徴を活用して高い精度維持しつつトレース・サイズを削減し，シミュレーション時間の短縮を実現する．キャッシュ性能測定に基づく評価実験の結果，本手法はトレース・サイズを平均 98.8%削減し，そのときのキャッシュ・ミス率の予測誤差は平均 0.067 パーセンテージ・ポイントであった．

Fast, Accurate Memory Architecture Simulation Technique Using Memory Access Characteristics

Takatsugu Ono,[†] Koji Inoue[††] and Kazuaki Murakami[††]

This paper proposes a fast and accurate memory architecture simulation technique. To design memory architecture, the first steps commonly involve using trace-driven simulation. However, expanding the design space makes the evaluation time increase. A fast simulation is achieved by a trace size reduction, but it reduces the simulation accuracy. Our approach can reduce the simulation time while maintaining the accuracy of the simulation results. In order to evaluate validity of proposed technique, we measured the cache miss ratio. In our evaluation, the proposed technique reduces the trace size 98.8% and cache miss ratio differs from 0.067 percentage point on an average.

1. はじめに

メモリシステムが計算機性能に与える影響はきわめて大きい．したがって，高性能な計算機システムを構築するためには設計制約を満足する適切なメモリアーキテクチャを決定する必要がある．多くの場合，設計空間の探索を目的としてソフトウェア・シミュレーションによる性能評価が行われる．しかしながら，実機での実行に比べてシミュレーション速度は数桁遅いため，広大な設計空間を短期間で効率的に探索するのは難しいのが現状である．

メモリアーキテクチャ・シミュレーションの代表的な手法として，プログラム実行において発生したメモリアクセス情報を事前に採取し，それを入力とするトレース・ドリブン・シミュレーション法がある．より詳細なシミュレーションを目的とした実行ドリブン方式と比較して，高い抽象度での実行が可能なため高速に動作することができる．実際，現在でも CMP を評価対象としてトレース・ドリブン方式が広く利用されている[4),16)]．しかしながら，今後，メニーコアに代表されるより複雑かつ大規模なシステムを対象とした場合，トレース・ドリブン方式におけるさらなる高速シミュレーションが必要となる．

一般に，トレース・ドリブン方式シミュレーションの速度と精度は入力となるトレース・サイズに大きく依存しており，これらの間にはトレードオフ関係が存在する．たとえば，プログラム実行のある区間のみをトレース採取の対象とした場合，トレース・サイズの削減によりシミュレーション時間を短縮できる．しかしながら，プログラム実行に関するすべての振舞いを反映できないため，シミュレーション結果の精度が低下するといった問題が生じる．

† 九州大学大学院システム情報科学府
Graduate School of Information Science and Electrical Engineering, Kyushu University
†† 九州大学大学院システム情報科学研究院
Faculty of Information Science and Electrical Engineering, Kyushu University

そこで本稿では，メモリシステムの中でも設計選択肢が多く計算機性能に大きな影響を与えるキャッシュを対象とし，高速かつ正確なシミュレーションを可能とするメモリアーキテクチャ・シミュレーション法を提案する．また，アプリケーションとしてMPEG2およびSPEC2000ベンチマークを用いた定量的評価を行い有効性を明らかにする．本手法はまず，プログラムの実行開始から終了までを対象とした全アドレス・トレースを取得する．そして，それを小規模なアドレス・トレースに分割し，メモリアクセスの特徴を抽出する．得られた特徴の類似性に基づき，代表となるトレースを選択する．この代表サブ・トレースをもとに小規模なトレースを生成しシミュレーションすることで，精度を維持しつつ時間の短縮を実現する．さらに，この小規模なトレースは一度生成すると，異なるメモリアーキテクチャのシミュレーションにも使用することができる．したがって，評価時間を大幅に削減することが可能である．なお，このようにして生成した小規模なトレースによるシミュレーションはキャッシュに限らず，メモリアーキテクチャの評価に幅広く活用できる．

以下，2章でこれまでに行われてきた関連研究について述べる．3章では提案手法について述べ，4章でその有効性を調査する．5章でまとめと今後の課題について述べる．

2. 関連研究

シミュレーション時間は，シミュレータの実行速度とその入力に依存する．シミュレータの高速化技術として，シミュレーション過程を時間軸方向に分割して並列化し，その精度を低下させることなく高速に実行する手法が提案されている[15]．シミュレータの入力，つまりシミュレーションの対象区間を削減し高速化する手法も多く提案されている[18]．シミュレーション区間を削減することで比較的高速に実行できる一方精度が低下するため，いかに精度を維持するかが重要な課題となる．本稿では，より高速なシミュレーションを実現するため，後者の技術に着目する．精度を維持しつつシミュレーション時間を削減する手法は以下の3つに大別できる．

(1) プログラムの入力データを削減
(2) プログラムの特徴を維持しつつ小規模なプログラムを生成
(3) シミュレーションの対象区間をサンプリング

手法(1)に属するものとして文献5)があげられる．入力データの削減により実行時間は短縮するが，メモリアーキテクチャのシミュレーション精度が低いという問題がある．

(2)のアプローチをとっているものとして文献8), 9)がある．これらの先行研究では，評価対象システムにおいて1度プログラムを実行し，あるメモリ構成を前提とした小規模なベンチマーク・プログラムを生成している．そのため，異なるメモリ構成での評価を行う場合は，新たに小規模ベンチマークを作り直す必要がある．一方，本稿で提案する手法はメモリアーキテクチャに依存しないため，1度小規模なトレースを生成することで異なるメモリ構成においてもシミュレーション可能である．

(3)に分類される代表的な手法としてSMARTS[17]およびSimPoint[10),11)]があげられる．本稿で提案する手法も本分類に属する．SMARTSは，ある固定インターバルによって命令ストリームをサンプリングする．プログラムの振舞いに関係なく一定周期でサンプリングするため，フェーズの変化と周期が一致しない場合は精度が低下するといった問題が生じる．SimPointはプログラム全体の特徴を解析してサンプリングするため，このような問題が生じることはない．SimPointはプログラムの実行開始から終了までを一定命令数のインターバルで区切り，各インターバルにおける基本ブロックの実行回数によって，インターバルの特徴付けをしている．この特徴の類似性に基づいてインターバルをクラスタリングし，各クラスタから1つのインターバルをシミュレーションの対象として選出する．選出されたインターバルのみをシミュレーションし，結果に重み付けしている．SimPointを用いてアドレス・トレースを採取する方法も提案されている[6]．

本稿における提案手法は，各インターバルの特徴を解析して代表を選出するためSMARTSのような問題が生じることはない．また，各インターバルをメモリアクセスによって特徴付けしており，この点がSimPointと異なる．したがって，提案手法におけるメモリアーキテクチャ・シミュレーションの方がより高い精度を達成できる．実際4.2.5項での定量的評価において，提案手法の方がSimPointよりも高速かつ高精度であることを示している．

3. メモリアーキテクチャ・シミュレーション法

3.1 用語の定義

本稿で使用する用語について定義する．

- フル・トレース：アプリケーション・プログラムの実行開始から終了までのメモリアクセスの時刻とアドレスの集合．

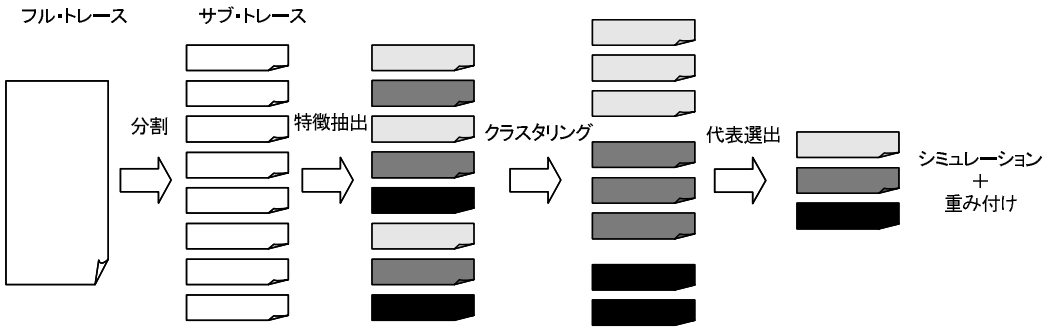


図 1 提案手法の概要
Fig. 1 Overview of proposed methodology.

- サブ・トレース：フル・トレースの部分集合であり，互いに素である．

3.2 提案手法の概要

本手法の概要を**図 1** に示す．まず，アプリケーション・プログラムを実行してフル・トレースを得る．これを一定クロック・サイクル間隔でサブ・トレースに分割する．次に，すべてのサブ・トレースを 3.3 節で説明する特徴の類似性に基づきクラスタリングし，各クラスタからシミュレーションの対象となる代表サブ・トレースを 1 つ選出する．なぜなら，サブ・トレースの特徴が同じであれば，それらのシミュレーション結果は非常に近い結果になると考えられるからである．選出されたサブ・トレースのシミュレーション結果に，各クラスタに属するサブ・トレース数に基づいて重み付けする．このように，サブ・トレースの特徴に基づいてシミュレーション対象とするサブ・トレースを決定することで，高速かつ正確なシミュレーションを実現する．

3.3 メモリアクセスの特徴抽出

メモリアクセスには時間局所性および空間局所性があることが知られている．これらの特徴はメモリシステムのシミュレーションにおいて重要である．また，単位時間あたりのメモリアクセス数や局所性は，プログラム実行におけるフェーズとともに変化すると予想される．したがって，フル・トレースを一定の間隔でサブ・トレースに分割し，その区間において特徴を抽出すると効果的であると考えられる．サブ・トレースの分割方法の選択肢として，命令単位とプログラム開始時刻からの経過クロック・サイクル数（時間）単位があげられる．命令単位で分割した場合，各命令間の時間はすべて一定と見なされることから，メモリアクセスがある時間に集中的に発生している場合や，逆に発生していない場合などの時間的特徴は抽出できない．一方，時間単位で分割すると時間的特徴は抽出可能である．メモリシステムのシミュレーションにおいて，この時間的な特徴は重要である．たとえば，CMP や SoC といった他コアや IP とバスを共有しているアーキテクチャにおいて，バスの性能をシミュレーションする状況を考える．ある時間にバスへのアクセスが集中すると競合が発生する可能性が高くなる．命令単位で分割した場合はメモリアクセスの時間的な特徴を考慮しないため，バスのシミュレーション精度は低下すると考えられる．時間単位で分割すると時間的な特徴を抽出できることから，精度の低下は前者ほど生じない．本稿では評価対象をキャッシュとしているが，バスの性能評価などへの適用を考え時間単位で分割する．さらに，空間局所性を抽出するために，サブ・トレースにおけるアドレス空間を**図 2** のように一定の間隔で分割する．以後この間隔のことをアドレス・インターバルと呼ぶ．図 2 の STEP1 で，各アドレス・インターバルにおけるメモリアクセスの数をカウントする．たとえば，最上位のアドレス・インターバルにおいてメモリアクセス数は点線の円で囲んだ 2 つであることから，表の最上位の要素に 2 を格納する．STEP2 では，STEP1 で作成したベクトルに特徴抽出精度の向上を目的として以下の 3 つの要素を追加する．

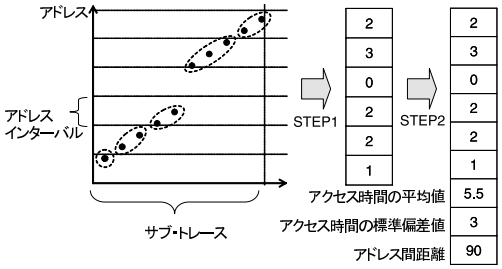
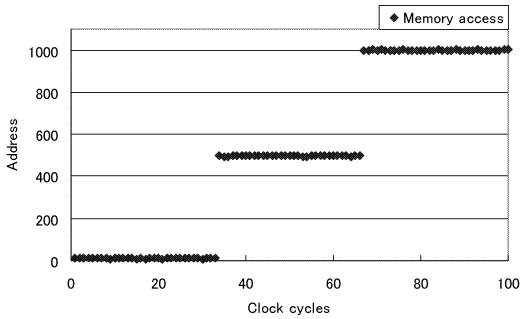
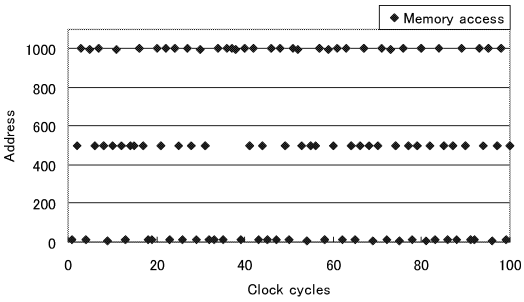


図 2 特徴抽出手順
Fig. 2 Feature vector for memory accesses.

(1) アクセス時間の平均値：各サブ・トレースの開始時刻を 0 として，各メモリアクセスの時間か



(a) 連続したメモリアクセス



(b) 離散的なメモリアクセス

図 3 アドレス間距離
Fig. 3 Address distance.

ら平均値を算出する．サブ・トレースにおけるメモリアクセス数を N，各メモリアクセスの時間を x_i とすると，平均値 $\bar{x}$ は式 (1) によって求められる．

$$\bar{x} = \frac{\sum x_i}{N} \tag{1}$$

(2) アクセス時間の標準偏差値：各サブ・トレースの開始時刻を 0 として，各メモリ・アクセスの時間から標準偏差値 σ を算出する．これは式 (2) によって求められる．

$$\sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{N}} \tag{2}$$

(3) アドレス間距離：時間的に連続したメモリアクセスにおけるアドレスの絶対差分の総和によって定義される．

アクセス時間の平均値および標準偏差値を導入することで，サブ・トレースにおけるメモリアクセスの時間的な偏りが分かる．アドレス間距離の必要性について，**図 3** の例を用いて説明する．図 3 の縦軸はアドレス，横軸は時間であり，各点はメモリアクセスの時刻とそのアドレスをもとにプロットしている．図 3 (a) はあるアドレス付近に時間的に連続してアクセスが生じており，図 3 (b) は離散的にアクセスが生じている．この違いがあるにもかかわらず，図 3 (a)，(b) ともにメモリアクセスの数，平均値ならびに標準偏差値は同じ値になる．そこで，これらの違いを表す指標としてアドレス間距離が必要となる．時間的に連続したメモリアクセスが生じた場合，アドレス間距離は小さい．一方，アクセスが分散した場合は大きな値を示す．つまり，図 3 (a) におけるアドレス間距離は図 3 (b) のそれよりも小さい．

このようにして得られたベクトルを以後特徴ベクトルと呼ぶ．すべてのサブ・トレースに対して特徴ベクトルを求める．

3.4 クラスタリングとサブ・トレースの選出

サブ・トレースの特徴が同じであれば，そのシミュレーション結果は近いと考えられる．そこで，3.3 節で述べた手法によって抽出された特徴ベクトルの類似性に基づいて，サブ・トレースをクラスタリグする．クラスタリング方法は既存の手法が適用できる．本手法において，ベクトルの数およびベクトルの次元が高いことから，たとえば K–平均法[1] といった比較的短時間で実行可能なアルゴリズムが適している．各クラスタには同じ特徴のサブ・トレースが属しているため，これらの中からどれをシミュレーションの対象としても結果は同じになることが予想される．したがって，各クラスタから 1 つのサブ・トレースをランダムに選出し，シミュレーションの対象として選出する．

3.5 代表サブ・トレースによるシミュレーション

代表サブ・トレースのみを用いたキャッシュ・シミュレーションでは，それ以前のトレースを実行していないためコールド・スタートの影響を受けるという問題が生じる．この問題を解決するために，各代表サブ・トレースの直前のトレースをウォームアップ・トレースとして用いる．以後，各クラスタの代表サブ・トレースと，それに対応するウォームアップ・トレースの集合を小規模ベンチマーク・トレースと呼ぶ．

各クラスタに属するサブ・トレースの数が多いほど，フル・トレースによるシミュレーション結果に大きな影響を与えていると考えられる．したがって，各代表サブ・トレースのシミュレーション結果に，クラスタに属するサブ・トレースの数を重みとして掛ける．重み付けした結果を合計することで，フル・トレースのシミュレーション結果を予測する．小規模ベンチマーク・トレースを用いた場合のキャッシュ・ミス率 $miss\_rate$ は式 (3) で求めることができる．

$$miss\_rate = \frac{\sum_{i=1}^{k} (miss_i \times weight_i)}{access} \times 100 \tag{3}$$

ここで，クラスタ i における代表サブ・トレースの総メモリアクセス数を $access$，ミス数を $miss_i$，クラスタに属するサブ・トレース数を $weight_i$ とする．k はクラスタ数である．

4. 評　　価

4.1 評価環境

3 章で生成した小規模ベンチマーク・トレースの有効性を評価する．ここでは L1 キャッシュを対象とし，命令およびデータキャッシュのミス率を測定する．フル・トレースと小規模ベンチマーク・トレースのシミュレーション結果を比較し，トレース・サイズ削減率およびキャッシュ・ミス率の予測精度を求める．予測精度を表す指標は，それぞれのキャッシュ・ミス率の差（パーセンテージ・ポイント）とする．また，先行研究の SimPoint との定量的比較を行い，本手法の有効性を明らかにする．

マイクロプロセッサ・シミュレータである SimpleScalar3.0[12] で MPEG2 のデコードプログラム[7] と SPEC2000 ベンチマーク[14] を実行し，フル・トレースを採取した．このときの命令キャッシュの構成は，キャッシュ・サイズ 16 KB，ブロック・サイズ 32 B，ウエイ数 1 とした．また，データキャッシュの構成は，キャッシュ・サイズ 16 KB，ブロック・サイズ 32 B，ウエイ数 4 とした．MPEG2 デコードプログラムでは akiyo，carphone ならびに foreman の 3 つの動画像を入力データとして使用した．それぞれの画像サイズは QCIF で，フレーム数は 150 である．SPEC2000 ベンチマークは 164.gzip，175.vpr，176.gcc ならびに 197.parser の 4 つの整数プログラムと，177.mesa および 183.equake の 2 つの浮動小数点プログラムを使用した．また，SPEC2000 の入力はいずれも test を用いた．

小規模ベンチマーク・トレースの生成において，各種パラメータを**表 1** のように設定した．サブ・トレース・サイズは，小さい程特徴抽出精度の向上が見込まれる．しかしながら，代表サブ・トレース数が増加することから，クラスタリングに長時間を要することになる．これらのトレードオフを考慮して，サブ・トレース・サイズを決定した．アドレス・インターバルの値をキャッシュのブロック・サイズに合わせることが考えられる．しかしながら，小規模ベンチマーク・トレースは異なるメモリアーキテクチャでも使用するため，それに依存する値に設定することはできない．アドレス・インターバルの値を変更してすべてのプログラムで実験を行った結果，精度に影響を与えることは明らかになったが，その差は小さいことが分かった．アドレス・インターバルの値を小さくすると，特徴ベクトルの次元が高くなりクラスタリング時間が長くなる．したがって，大幅な精度の向上は得られないにもかかわらず，クラスタリングに長時間を要することになることから，アドレス・インターバルを小さな値に設定しても利点が少ないと判断した．シミュレーション精度と，現実的な時間で実験を行うという点を考慮して，本評価においては表 1 の値を用いて議論する．アドレス・インターバルがシミュレーション精度に与える影響については，4.2.4 項で述べる．クラスタリング・アルゴリズムは，クラスタリング精度も高く比較的短時間で実行できる K–平均法[1] を用いた．K–平均法をサポートしているソフトウエア Cluster3.0[2] を使用して実験を行った．クラスタ数が少ない場合，特徴の異なるサブ・トレースが同じクラスタに属する可能性が高くなる．逆にクラスタ数を多くするとシミュレーション対象となるトレース・サイズが増加し，削減率が低下する．小規模ベンチマーク・トレースのメモリアクセス数は以下の式で見積もることができる．

表 1 パラメータなど
Table 1 Parameters.

サブ・トレース・サイズ（クロック・サイクル）		10,000
アドレス・インターバル	命令キャッシュ	1,024
	データキャッシュ	8,192
クラスタリング・アルゴリズム		K–平均法
クラスタ数		500

$$S = \sum_{i=1}^{k} (N_i + W) \tag{4}$$

ここで k はクラスタ数，N_i はクラスタ i における代表サブ・トレースのメモリアクセス数，W はウォームアップ・トレースのメモリアクセス数である．N_i はプログラムを実行した時点で決定するため，変更できるパラメータは k ならびに W である．k および W が大きいほどシミュレーション精度が高くなると予想される．しかしながら，小規模ベンチマーク・トレースのメモリアクセス数が増加するというトレードオフの関係にある．最適なクラスタ数はプログラムによって異なることから一意に決定できない．本稿における評価では，様々なクラスタ数を用いてすべてのプログラムに対して実験を行った結果，すべてのプログラムにおいて 97%以上の削減率の達成できるクラスタ数 500 を用いて議論を進める．クラスタ数と同様に，ウォームアップ・トレース・サイズもトレードオフの関係にある．目標のトレース・サイズ削減率を達成す

表 2 キャッシュ構成（キャッシュサイズ KB/ウエイ数 W）
Table 2 Simulated cache configurations.

Instruction	Data		
32 KB/1 W	32 KB/1 W	32 KB/2 W	32 KB/4 W
16 KB/1 W	16 KB/1 W	16 KB/2 W	16 KB/4 W
8 KB/1 W	8 KB/1 W	8 KB/2 W	8 KB/4 W
4 KB/1 W	4 KB/1 W	4 KB/2 W	4 KB/4 W

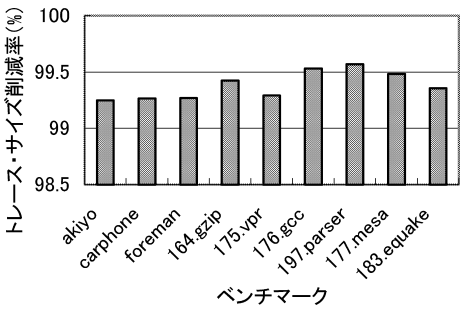


図 4 命令キャッシュにおけるトレース・サイズ削減率
Fig. 4 Trace size reduction for I-cache.

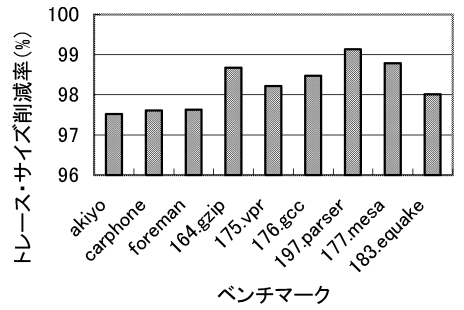


図 5 データキャッシュにおけるトレース・サイズ削減率
Fig. 5 Trace size reduction for D-cache.

るためにウォームアップ・トレース・サイズは 16,000 とした．代表サブ・トレースをランダムに選出するため，シミュレーション結果がばらつくことが予想される．したがって，1 つのベンチマークに対し，各クラスタから代表サブ・トレースをランダムに選出する試行を 10 回行い，10 個の異なる小規模ベンチマーク・トレースを生成した．4.2 節で示す結果はすべて 10 回のシミュレーションした結果の平均値である．また，提案手法がトレース採取時と異なるキャッシュ構成において有効であることを調査するため，**表 2** に示すキャッシュ構成について評価した．なお，これらのブロック・サイズはすべて 64 B である．

4.2 評価結果

4.2.1 トレース・サイズ削減率

命令，データキャッシュを対象としたサブ・トレースの削減率の平均をそれぞれ**図 4** および**図 5** に示す．縦軸がフル・トレースに対する小規模ベンチマーク・

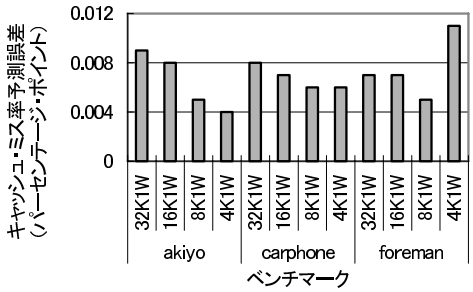


図 6 命令キャッシュにおける MPEG2 の
キャッシュ・ミス率予測誤差
Fig. 6 Simulation accuracy for I-cache (MPEG2).

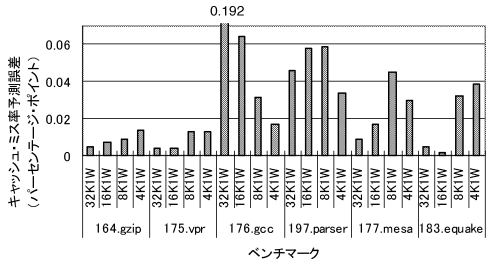


図 7 命令キャッシュにおける SPEC2000 の
キャッシュ・ミス率予測誤差
Fig. 7 Simulation accuracy for I-cache (SPEC2000).

トレースの削減率であり，横軸はベンチマークである．命令キャッシュの削減率は平均 99.38%，データキャッシュは平均 98.22%であった．また，クラスタリングに要した時間はクラスタ数や特徴ベクトルの次元数で異なるが，クラスタ数 500，データキャッシュのアドレス・インターバル 8,192 の場合において，フル・トレースによるシミュレーション時間の約 6 倍程度であった．1 章でも述べたように，小規模ベンチマーク・トレースの目的は，アーキテクチャに依存しないトレースを 1 度生成するだけで，異なるアーキテクチャでも再利用することである．したがって，クラスタリングに比較的長い時間が必要であっても，様々なキャッシュ構成でシミュレーションする場合さほど問題にならない．また，ランダム・プロジェクション[11] といった特徴ベクトルの次元を縮小する手法などの適用によりクラスタリング時間を大幅に短縮することは可能である．

4.2.2 キャッシュ・ミス率の予測誤差

MPEG2 における命令キャッシュのミス率の平均予測誤差を**図 6** に，SPEC2000 の場合を**図 7** に示す．図 6 および図 7 ともに，縦軸がフル・トレースによって得られたミス率と本手法によって得られたそれとのポイント差で，横軸はベンチマークである．図 6 における予測誤差の平均は 0.0069 パーセンテージ・ポイン

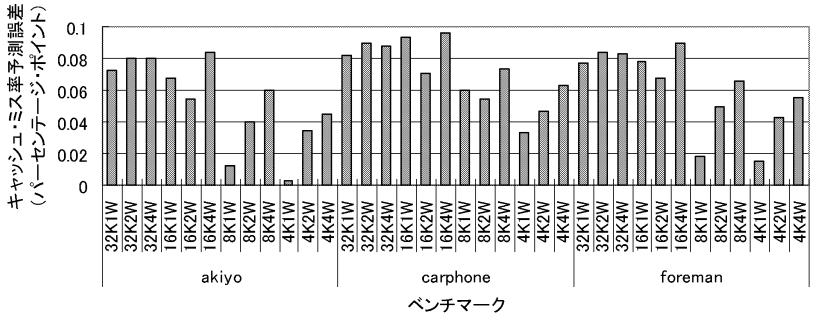


図 8 データキャッシュにおける MPEG2 のキャッシュ・ミス率予測誤差
Fig. 8 Simulation accuracy for D-cache (MPEG2).

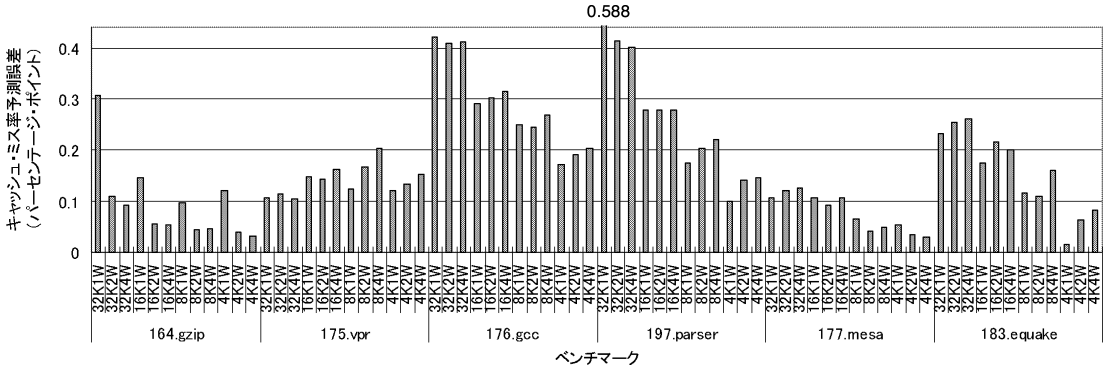


図 9 データキャッシュにおける SPEC2000 のキャッシュ・ミス率予測誤差
Fig. 9 Simulation accuracy for D-cache (SPEC2000).

トときわめて小さい．MPEG2 は同じ処理を一定周期で繰り返すことから，クラスタにおけるサブ・トレースの特徴の類似度が高いことが原因と考えられる．つまり，クラスタ内に特徴が異なるサブ・トレースが少ないため，予測精度が高い．図 7 の平均予測誤差は約 0.031 パーセンテージ・ポイントであった．多くのベンチマークにおいて 0.1 パーセンテージ・ポイント以下と MPEG2 より劣るものの，小さな値を示している．

次にデータキャッシュにおける，MPEG2 のミス率の予測誤差を**図 8** に，SPEC2000 の場合を**図 9** に示す．図 8 および図 9 ともに，縦軸がフル・トレースによって得られたミス率と本手法によって得られたそれとのポイント差で，横軸はベンチマークである．図 8 では，予測誤差の平均は約 0.061 パーセンテージ・ポイントと小さい値を示した．このことから，同一アプリケーションにおいて入力データが異なっても予測精度が高いことが分かる．図 9 において，予測誤差の平均は約 0.171 パーセンテージ・ポイントであった．

4.2.3 クラスタ数が削減率と精度に与える影響

4.1 節で述べたとおり，クラスタ数が多いほどシミュレーション精度が高くなるが，削減率が低下するという問題が生じる．クラスタ数を 100，500 および 1,000 として精度と削減率に与える影響を調査した．MPEG2 デコードプログラムの入力データ foreman を用い，アドレス・インターバルおよびウォームアップ・トレース・サイズは表 1 と同様である．

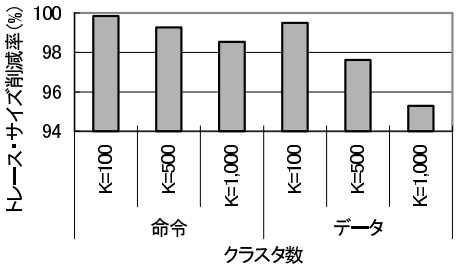


図 10 クラスタ数とトレース・サイズ削減率の関係.
Fig. 10 Impact of K on trace size.

トレース・サイズ削減率を**図 10** に示す．棒グラフは 4.1 節と同様に 10 回小規模ベンチマーク・トレースを生成しシミュレーションした結果の平均である．クラスタ数の増加とともにトレース・サイズ削減率が低下していることが確認できる．データキャッシュにお

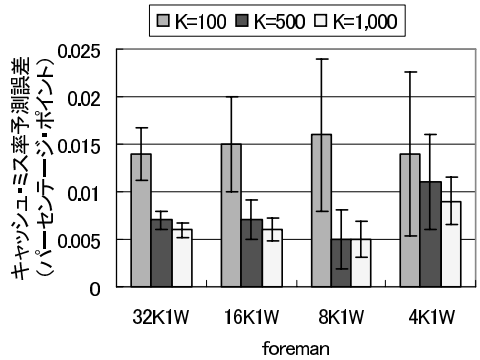


図 11 命令キャッシュにおけるクラスタ数と精度の関係
Fig. 11 Impact of K on accuracy (I-cache).

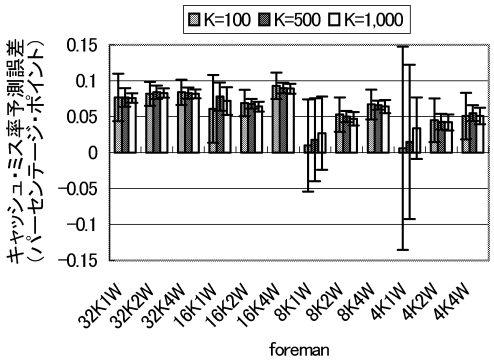


図 12 データキャッシュにおけるクラスタ数と精度の関係
Fig. 12 Impact of K on accuracy (D-cache).

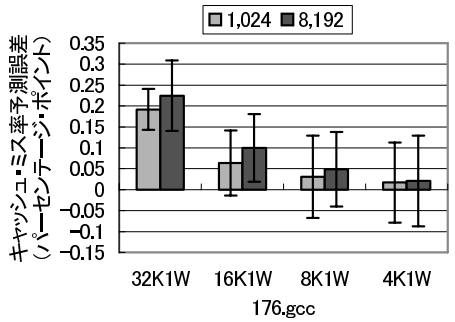


図 13 命令キャッシュにおけるアドレス・インターバルと精度の関係
Fig. 13 Impact of address-interval size on accuracy (I-cache).

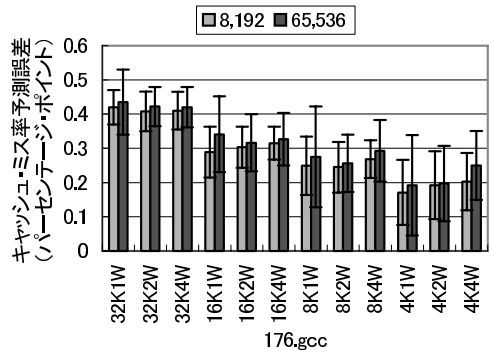


図 14 データキャッシュにおけるアドレス・インターバルと精度の関係
Fig. 14 Impact of address-interval size on accuracy (D-cache).

ける削減率の低下が大きいが，これはフル・トレース・サイズが比較的小さいためである．キャッシュミス率の予測誤差を命令キャッシュの場合を**図 11** に，データキャッシュの場合を**図 12** に示す．範囲グラフは 10 回シミュレーションした結果の標準偏差を表す．データキャッシュの一部でクラスタ数が小さい方が高い精度を示す場合があるものの，クラスタ数が大きい方が高精度という傾向にあることが分かる．標準偏差もクラスタ数の増加とともに小さくなる傾向にあり，クラスタ内に属するサブ・トレースの類似度が高くなっていることが確認できる．しかしながら，クラスタ数を増加させただけでは大幅な精度の向上は期待できないといえる．

4.2.4 アドレスインターバルが精度に与える影響

アドレス・インターバルを大きくすると，メモリアクセスの特徴抽出精度が低くなり，キャッシュ・ミス率の予測精度が低下すると考えられる．

SPEC2000 の 176.gcc において，命令キャッシュのアドレス・インターバルを 1,024 および 8,192 としたときの，キャッシュ・ミス率予測誤差を**図 13** に示す．データキャッシュのアドレス・インターバルを 8,192 および 65,536 とした場合のキャッシュ・ミス率の予測誤差を**図 14** に示す．命令，データキャッシュともにクラスタ数は 500 であり，ウォームアップ・トレース・サイズは 16,000 である．図 13 および図 14 から，アドレス・インターバルが大きな値の場合において，予測精度が低く標準偏差も大きな値を示していることが分かる．アドレス・インターバルを大きくすることで予測精度が低下することを確認したが，これらのアドレス・インターバルにおいては大幅な精度の低下はみられなかった．

4.2.5 関連研究との比較

本項では，2 章で述べた SimPoint [10),11)] との定量的比較を行い，本手法の有効性を明らかにする．SimPoint3.1 [3),13)] を用いて採取したアドレス・トレースと，本手法によるシミュレーション結果およびトレース・サイズ削減率の比較を行った．本手法の各種パラメータは 4.1 節で述べたものと同じである．キャッシュ構成は，命令キャッシュが，キャッシュ・サイズ

表 3 ウォームアップ・トレース・サイズとクラスタ数
Table 3 Warm-up trace size and the number of clusters.

ベンチマーク	対象キャッシュ	ウォームアップ・トレース・サイズ	クラスタ数
foreman	命令	64,000	1,200
	データ	64,000	2,500
176.gcc	命令	64,000	2,350
	データ	64,000	850
183.equake	命令	64,000	3,600
	データ	64,000	1,360

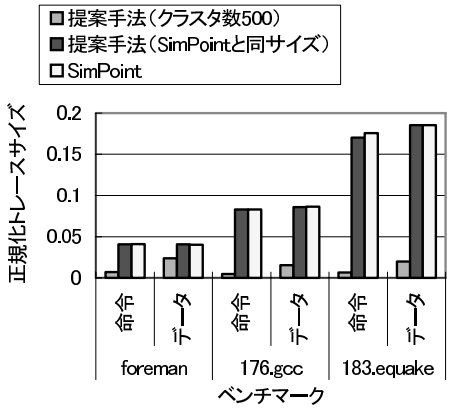


図 15 SimPoint とのトレース・サイズの比較
Fig. 15 Comparison with SimPoint (trace size).

4 KB，ブロック・サイズ 64 B，ウエイ数 1 とした．また，データキャッシュの構成は，キャッシュ・サイズ 4 KB，ブロック・サイズ 64 B，ウエイ数 4 とした．対象とするベンチマークとして，MPEG2 の foreman，SPEC2000 の 176.gcc ならびに 183.equake を選択した．SimPoint3.1 において，インターバルの値は高速かつ高精度な結果を示す 10 M 命令とした[3]．さらに，本手法によるトレース・サイズ削減率を SimPoint と同一にした際の予測誤差についても比較した．このときの本手法のウォームアップ・トレース・サイズおよびクラスタ数を**表 3** に示す．すべてにおいてウォームアップ・トレース・サイズを 64,000 とし，クラスタ数を調整することで SimPoint のトレース・サイズと同一にした．

図 15 に，フル・トレースのトレース・サイズを 1 として正規化した場合の，トレース・サイズを示す．左の棒グラフから順に，クラスタ数 500 の提案手法，SimPoint と同一トレース・サイズの提案手法，SimPoint の結果である．本手法によるトレース・サイズは SimPoint の約 10%から 60%であった．この結果から，SimPoint よりも高速に実行できることが分かる．**図 16** にキャッシュ・ミス率の予測誤差の比較結果を示す．縦軸はフル・トレースによって得られたキャッシュ・ミス率と本手法によって得られたそれとのポイント差であり，横軸はベンチマークである．foreman の命令キャッシュにおいては SimPoint よりも精度が低いが，その差は 0.004 パーセンテージ・ポイントときわめて小さい．それ以外では命令，データキャッシュともに SimPoint よりもシミュレーション精度が高いことが分かる．SimPoint と同トレース・サイズの場合，foreman や 176.gcc のデータキャッシュなどで精度の向上がみられる．クラスタ数 500 の提案手法と比較して 183.equake のデータキャッシュのように精度が低下する場合があることも分かった．

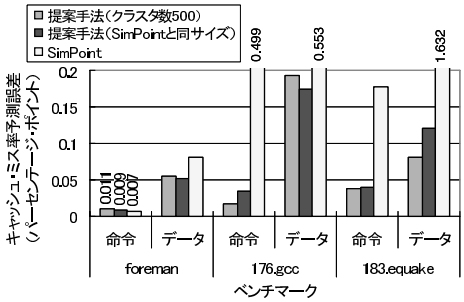


図 16 SimPoint とのキャッシュ・ミス率予測誤差の比較
Fig. 16 Comparison with SimPoint (accuracy).

これらすべてのベンチマークにおいて，提案手法は SimPoint よりもトレース・サイズが小さく，多くの場合において精度が高いことが分かる．SimPoint より精度が低い場合でも，その差はきわめて小さいことから，本手法は SimPoint よりも高速かつ高精度といえる．

5. おわりに

本稿では，高速かつ高精度なメモリアーキテクチャのシミュレーション手法を提案し，MPEG2 および SPEC2000 ベンチマークを用いた評価実験によって有効性を明らかにした．プログラムの実行時に得られるフル・トレースから，メモリアクセスの特徴に基づいて，より小規模なベンチマーク・トレースを生成した．その結果，トレース・サイズの削減率は平均 98.8%，キャッシュ・ミス率の予測誤差は平均 0.067 パーセンテージ・ポイントであった．したがって，本手法は短時間で高精度のメモリアーキテクチャ・シミュレーションが可能であるといえる．さらに，小規模ベンチマーク・トレースは生成時と異なるメモリアーキテクチャにおいてもシミュレーション精度が高く，有効であることを示した．また，関連研究との定量的比較を行い，本手法が高精度かつ高速であることを確認した．

今後は小規模ベンチマーク・トレースをバスの性能

評価などへ適用して，その有効性を評価する予定である．

謝辞 本稿をまとめるにあたり，ともにご討論いただいた九州大学の安浦・村上・松永・井上研究室の皆様に感謝いたします．なお，本研究は一部，科学研究費補助金（学術創成研究費：課題番号 14GS0218，若手研究 A：課題番号 17680005），21 世紀 COE プログラム「システム情報科学での情報基盤システム形成」ならびに松下電器産業株式会社との共同研究による．

参 考 文 献

1) 麻生英樹ほか：パターン認識と学習の統計学，岩波書店 (2006).
2) Cluster3.0. http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm#source
3) Hamerly, G., et al.: SimPoint 3.0: Faster and More Flexible Program Analysis, *Workshop on Modeling, Benchmarking and Simulation*, Wisconsin, Madison (2005).
4) Hsu, L., et al.: Exploring the Cache Design Space for Large Scale CMPs, *SIGARCH Comput. Archit. News*, Vol.33, No.4, pp.24–33 (2005).
5) KleinOsowski, A.J. and Lilja, D.J.: MinneSPEC: A New SPEC Benchmark Workload for Simulation-Based Computer Architecture Research., *Computer Architecture Letters*, Vol.1 (2002).
6) Laurenzano, M., et al.: Low Cost Trace-driven Memory Simulation Using SimPoint, *SIGARCH Comput. Archit. News*, Vol.33, No.5, pp.81–86 (2005).
7) MPEG Software Simulation Group. http://www.mpeg.org/MPEG/MSSG/
8) Bell, J.R.H. and John, L.K.: The Case for Automatic Synthesis of Miniature Benchmarks, *Workshop on Modeling, Benchmarking and Simulation*, Wisconsin, Madison, pp.88–97 (2005).
9) Bell, J.R.H. and John, L.K.: Improved Automatic Testcase Synthesis for Performance Model Validation, *Proc. 19th annual international conference on Supercomputing*, pp.111–120, ACM Press (2005).
10) Sherwood, T., et al.: Basic Block Distribution Analysis to Find Periodic Behavior and Simulation Points in Applications, *Proc. 2001 International Conference on Parallel Architectures and Compilation Techniques*, pp.3–14, IEEE Computer Society (2001).
11) Sherwood, T., et al.: Automatically Characterizing Large Scale Program Behavior, *Proc. 10th international conference on Architectural support for programming languages and operating systems*, pp.45–57, ACM Press (2002).
12) SimpleScalar Simulation Tools for Microprocessor and System Evaluation. http://www.simplescalar.org/
13) SimPoint3.1. http://www.cse.ucsd.edu/~calder/simpoint/
14) SPEC. http://www.specbench.org/
15) 高崎 透ほか：時間軸分割並列化による高速マイクロプロセッサシミュレーション，情報処理学会論文誌：コンピューティングシステム，Vol.46, No.12, pp.84–97 (2005).
16) Vera, J., et al.: A Novel Evaluation Methodology to Obtain Fair Measurements in Multithreaded Architectures, *Workshop on Modeling, Benchmarking and Simulation*, Boston, Massachusetts (2006).
17) Wunderlich, R.E., et al.: SMARTS: Accelerating Microarchitecture Simulation via Rigorous Statistical Sampling, *Proc. 30th International Symposium on Computer Architecture*, pp.84–95, IEEE Computer Society (2003).
18) Yi, J.J. and Lilja, D.J.: Simulation of Computer Architectures: Simulators, Benchmarks, Methodologies, and Recommendations., *IEEE Trans. Comput.*, Vol.55, No.3, pp.268–280 (2006).

(平成 19 年 1 月 22 日受付)
(平成 19 年 4 月 26 日採録)

小野 貴継（学生会員）

昭和 57 年生．平成 18 年福岡大学大学院工学研究科電子情報工学専攻修士課程修了．現在，九州大学大学院システム情報科学府情報理学専攻博士後期課程在学中．メモリアーキテクチャの評価に関する研究に従事．

井上 弘士（正会員）

昭和 46 年生．平成 8 年九州工業大学大学院情報工学研究科修士課程修了．同年横河電機（株）入社．平成 9 年より（財）九州システム情報技術研究所研究助手．平成 11 年の 1 年間 Halo LSI Design & Device Technology, Inc. にて訪問研究員としてフラッシュ・メモリの開発に従事．平成 13 年九州大学にて工学博士を取得．同年福岡大学工学部電子情報工学科助手．平成 16 年より九州大学大学院システム情報科学研究院助教授．平成 19 年 4 月より，同大学准教授．現在に至る．高性能/低消費電力キャッシュメモリ・アーキテクチャ，セキュアプロセッサ・アーキテクチャ，性能評価に関する研究に従事．電子情報通信学会，ACM，IEEE 各会員．

村上 和彰（正会員）

昭和 35 年生．昭和 59 年京都大学大学院工学研究科情報工学専攻修士課程修了．同年富士通（株）入社．汎用大型計算機の研究開発に従事．昭和 62 年九州大学助手．平成 6 年九州大学助教授．現在九州大学大学院システム情報科学研究院情報理学部門教授，情報基盤研究開発センター長，情報統括本部長．計算機アーキテクチャ，並列処理，システム LSI 設計技術等に関する研究に従事．工学博士．平成 3 年情報処理学会研究賞，平成 4 年情報処理学会論文賞，平成 9 年坂井記念特別賞，平成 12 年日経 BP 社 IP アワード，平成 12 年情報処理学会創立 40 周年記念論文賞，平成 14 年電子情報通信学会業績賞をそれぞれ受賞．