

並列階層キャッシュの性能とキャッシュ・ミスの分類

村谷博文

(株) 東芝 総合研究所

1.はじめに

並列推論マシンPIM/k[1]が採用している並列階層キャッシュは、バス共有の密結合並列計算機のス・ボトルネックの解消を目的とするアーキテクチャである[2]。PIM/kやその上のKL1処理系開発のため、シミュレーションによる性能評価を行なっている。PIM/k上のKL1処理系の評価に先立ち、並列階層キャッシュの特性の一般的傾向を理解するための評価を行なった。本稿は、キャッシュ・ミスを要因により分類し、それらがシステム性能に与える影響を調べたシミュレーション結果に関するものである。なお、本研究は、通産省第5世代コンピュータプロジェクトの一環である。

2.評価方法

2.1.評価モデル

シミュレーション・モデルを図1に示す。並列推論マシンPIM/kを参考にした。コンシステンシ・プロトコルは、2階層ともcopy-back、4状態キャッシュ・ブロック、Multi-level inclusion propertyである[1]。システム・パラメータを表1に示す。1-level cacheとの比較により評価する。

2.2.アドレ스트レース

トイ・プログラムによりアドレス・トレースを生成した。メモリ・アクセスの構成は命令フェッチ:データリード:データライト=7:2:1である。トレース長は約10K~80Kである。

2.3.キャッシュ・ミスの分類

キャッシュ・ミスを分類することは、キャッシュ

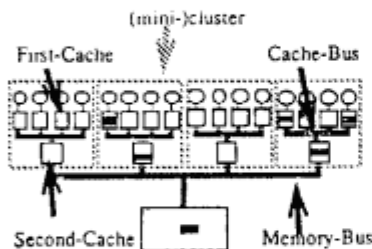


図1: シミュレーション・モデル

First-Cache(2-level)	
replacement algorithm	direct mapping
block size	4words
Second-Cache(2-level)	
replacement algorithm	set associative
block size	4words
associativity	8
Turn-around(2-level)	
First-Cache(hit)	1clock
Second-Cache(hit)	13clocks
Main-memory	50clocks
Cache(1-level)	
replacement algorithm	direct mapping
block size	4words
Turn-around(1-level)	
Cache(hit)	1clock
Main-memory	38clocks
Turn-around to Bus request	3clocks
Bus latency	1clock

表1: システム・パラメータの設定

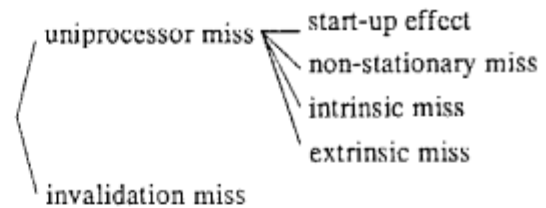


表2: キャッシュ・ミスの分類

性能の理解を助ける。Agarwal et al.[3]、Egger and Katz[4]に基づいた分類を表2に示す。intrinsic missは衝突により追い出されたデータをアクセスした際に生ずるミスである。invalidation missはprivate cacheを持つマルチプロセッサで生ずるミスである。invalidation missの影響を評価することが並列階層キャッシュでは重要である。

2.4.potentially colliding ratioとsharing ratio

intrinsic missの発生する頻度やinvalidation missの発生する頻度を表すために次のような量を定義した。

potentially colliding blockとはメモリ・アクセスのシーケンスをキャッシュ・ブロックにマップした

Performance of a hierarchical cache and classes of cache misses

Hirofumi Muratani

Toshiba R&D Center

際に衝突が起こるブロックのことである[3]。メモリ・アクセスがpotentially colliding blockへのアクセスである確率をpotentially colliding ratio(PCR)と定義する。以下ではこのPCRを近似的にintrinsic missの発生する頻度と見なすことにする。

メモリ・アクセスが他のプロセッサとの共有領域へのアクセスである確率をsharing ratio(SR)と定義し、近似的にinvalidation missの発生する頻度と見なすことにする。

3. シミュレーション結果

以下の二通りのシミュレーションを行った。プロセッサ台数はいずれの場合も16である。

(1) intrinsic missの多いプログラムの場合

SR=0で、PCRを変化させた。図2がその結果である。PCRが大きいと2-level cacheの方が性能が良いが、PCRが小さくなると1-level cacheの方が性能が良くなっている。

キャッシュ・サイズが大きいほど性能が劣るのはstart-up effectによると考えられる。

(2) invalidation missの多いプログラムの場合

PCR=0で、SRを変化させた。図3がその結果である。この場合も、SRが大きいと2-level cacheの方が性能が良いが、SRが小さくなると1-level cacheの方が性能が良くなっている。ただし、1-level cacheと2-level cacheの性能の差は先の場合ほど大きくはない。2-level cacheでのinvalidationのコストが大きいためと考えられる。

4. まとめ

private cacheを持つマルチプロセッサにおけるキャッシュ・ミス进行分类した。この分類に基づいて1-level cacheと2-level cacheの性能比較を行なった。今回設定したパラメータの下ではintrinsic miss、invalidation missが多くなると2-level cacheの方が1-level cacheより性能がよくなる。階層化によるバス・トラフィックの軽減の効果と考えられる。一般のアドレス・トレースは両方のミスを含んでいるが、この傾向は保たれると予想される。今後は、他のミス要因のシステム性能への影響の評価やKL1処理系から得られるアドレス・トレースを用いた評価を検討している。

謝辞

日頃、御指導いただくICOT第1研究室の方々に感謝します。

参考文献

[1] 浅野滋博: 並列推論マシンPIM/k, 情報処理学会第39年全国

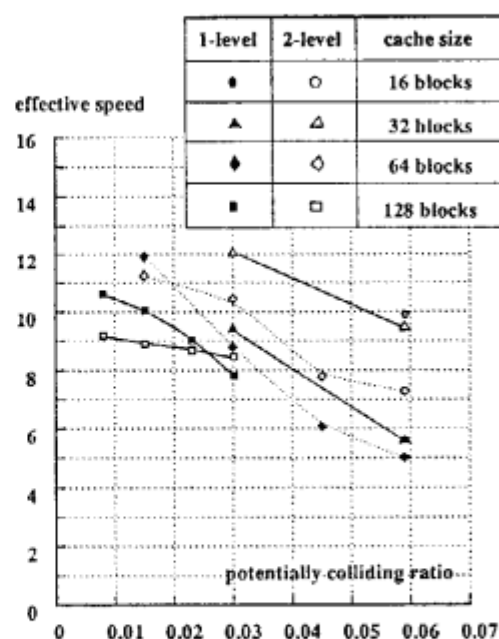


図2: potentially colliding ratio とシステム性能

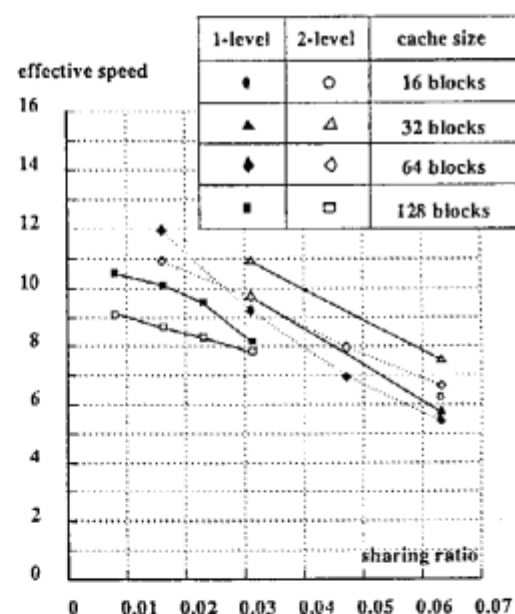


図3: sharing ratio とシステム性能

大会論文集(III), pp.1772-1773.

- [2] A.W.Wilson: Hierarchical cache/bus architecture for shared memory multiprocessors, Proc. of the 14th ISCA, Pittsburg, 1987, pp.244-252.
- [3] A.Agarwal et al: An analytic cache model, ACM Trans. on Computer Systems, Vol.7, No.2, 1989, pp. 184-215.
- [4] R.J.Eggar and R.H.Katz: The effect of sharing on the cache and bus performance of parallel programs, ASPLOS-III, 1989, pp.257-270.