

## 事前実行によるキャッシュブロックの適正配置解析の提案

広山 貴之・請園 智玲・田中 清史 (北陸先端科学技術大学院大学)

## 1 はじめに

従来の階層型キャッシュには、Inclusion Property・Exclusion Property の二種類の配置方法が存在する。Inclusion Property は、下位キャッシュに上位キャッシュのコピーブロックを保持する。例えば、メニーコア環境において、コア毎に L1 キャッシュを持ち、共有 L2 キャッシュを持つキャッシュ構成では、常に L2 キャッシュに各コアの L1 キャッシュのコピーを持たなければならないため、L2 キャッシュの容量を圧迫する原因となる。その一方で、他のプロセッサからのキャッシュコヒーレンス維持のためのキャッシュ参照は L2 キャッシュに限定されるため、参照オーバーヘッドが小さい。Exclusion Property は、上位と下位キャッシュの間で排他的にブロックを保持する（上位キャッシュのコピーを下位キャッシュが持たなくとも良い）。このため、先の場合の場合の L2 容量オーバーヘッドを小さくすることができるが、コヒーレンス維持のための参照の範囲が各コアの L1 キャッシュに及ぶため、参照オーバーヘッドが大きくなる。

## 2 先行研究

このトレードオフ問題に取り組んだ先行研究 [1] がある。先行研究では、コヒーレンス維持のオーバーヘッド時間の削減とキャッシュ資源の高効率利用を同時に実現するために、キャッシュへ配置するブロックを次に述べる 5 つのカテゴリに分ける提案をしている。先行研究が想定するキャッシュメモリシステムの概要図を図 1 に示す。

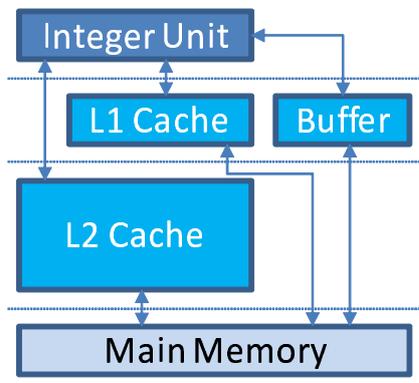


Figure 1: 先行研究が想定するキャッシュメモリシステム

それぞれのキャッシュは独立しており、互いのキャッシュ間でキャッシュブロック移動は行わない。キャッシュへのメモリブロック配置は事前実行により得られたメモリトレースを解析し下記の 5 種類に分類されるためブロックのロード時には配置されるべき位置が静的に決定されている。

1. 長期的に使用され、高い局所性を持つブロック  
L1 と L2 に存在すべきブロック
2. 長期的に使用されるが、低い局所性を持つブロック  
L2 のみに存在すべきブロック
3. 短期的に高い局所性を持つブロック  
L1 のみに存在すべきブロック
4. 時間的局所性の無いブロック  
小容量バッファに存在すべきブロック
5. プロセッサ間共有ブロック  
L2 に必ず配置すべきブロック

実際の実行で、これらにカテゴリ分けされたブロックが配置されるべき位置に適切フィルされれば、各キャッシュの容量オーバーヘッドと参照オーバーヘッドが最適な実行が可能になる。

## 3 研究の目的

本研究では、カテゴリ分け手法の潜在性能を示すことを目的とする。理想的なカテゴリ分けが行われたと仮定したときに、どの程度の最大性能が得られるのかを計測し、その性能向上要因を考察することによって、手法の価値のみならず、カテゴリ分けにどの程度の正確さが要求されるのかが知ることができ、本研究の発展形である現実的なカテゴリ分けのための解析の道筋を発見することができる。

## 4 提案手法

提案するカテゴリ分け手法は解析の為の材料として、アクセスしたブロックアドレスとアクセス時刻がセットになったデータアクセス系列のトレースを必要とする。このトレースは複数プロセッサ構成の CPU のシミュレーション上でベンチマークプログラムを実行することにより得る。

まず、カテゴリ (5) の一貫性維持のための参照オーバーヘッドの削減は容易に解析できる。並列プロセッサ構成を決め、各プロセッサ内のコアのプロセッサを跨いだ参照に関係するブロックを全て抽出することのみで対応できる。(1)~(4) のカテゴリは単一プロセッサ内に閉じた解析を行う必要がある。カテゴリ (4) は極端に参照の少ないブロックを抽出することで解析可能である。この抽出は比較的容易に行える。カテゴリ (1)~(3) は各プロセッサ内で与えられたキャッシュ容量を提案キャッシュ構成で有効に使うための最適な配置を探す解析が必要となる。

(1)~(3) の全てのブロックのカテゴリ分けを行う際、ブロックの使用期間に関する「長期」と「短期」の切り分け、参照局所性に関する「高い」と「低い」の切り分けは、それぞれ時間と参照回数に閾値を要求する。逆にいえば、各ブロックが (1)~(3) のどのカテゴリに属するかはその閾値を定めた時に決定される。本研究ではこの閾値を連続的に変えながら提案キャッシュメモリシステム上のシミュレーションを行い、各キャッシュのミス率を算出し、最も良い値を示した閾値を採用する。

## 5 おわりに

実際のカテゴリ分けの解析は膨大なメモリアクセストレースを対象に解析を行うため、膨大なトレース記憶容量と解析時間を要するが、カテゴリ分けしたブロックを静的に配置する提案がどの程度の潜在的な性能向上を持つのかを知るために、本研究では可能限りの解析時間と計算資源を投入しカテゴリ分けを行う。

実際の計算機環境に本研究の提案を適用する場合には、より現実的な解析の方法が必要となる。この問題は解析の実現容易性の向上と精度低下のトレードオフを見極める研究として発展させる予定である。

## References

- [1] 許 允碩, "階層型キャッシュシステムにおける高効率なブロック配置法", 北陸先端科学技術大学院大学修士論文, <https://dspace.jaist.ac.jp/dspace/bitstream/10119/9630/7/paper.pdf>, 2011.