2G-08

下位レベルキャッシュメモリへのアクセスフィルタによるタグ参照電力の削減

石田隆太† 請園智玲† 佐藤寿倫†
福岡大学工学部†

1 はじめに

近年のマイクロプロセッサは半導体集積技術の向上により得られた回路実装面積の多くをキャッシュメモリとして使用している．DRAM で構成される主記憶は GHz 帯のクロックで動作する CPU の動作周波数に比べ，10 の 2 乗のオーダで動作速度が遅い．この主記憶への参照遅延が演算のボトルネックとなり，CPU の高速な演算を妨げる一因となる．キャッシュメモリ（以降単にキャッシュと呼ぶ）は一時的にプロセッサ内部の SRAM に主記憶内のデータの一部を格納し，CPU がそのデータに対して高速に参照することが可能な機構を提供する事により，参照局所性のあるデータへのアクセス時間を削減し，同時に主記憶への参照回数を削減させる効果をもつ．

CPU の処理性能に対して影響力が大きいキャッシュの能力を向上させるためには，キャッシュが保持できるデータ量を向上させる必要がある．このため，近年のプロセッサは高集積化の恩恵をキャッシュの実装面積に優先的に割り当てている．しかしながら，キャッシュの増大にともない，キャッシュが消費する電力も無視できなくなる．並列のタグ一致検索処理はキャッシュの動的消費電力の大きな割合を占め，それを解決するための研究がなされている[1][2]．特に階層化キャッシュの下位（CPU からより遠い位置）では高連想度のメモリとしてキャッシュを構成することから，タグ一致検索の消費電力が問題となる．本研究はこの点に着目し，連想度が高い下位レベルのキャッシュの並列タグ検索の回数を減らすために，Bloom Filter を用いて明らかにミスが予測される場合の並列タグ検索を省略するマイクロアーキテクチャを提案し，キャッシュの低消費電力化を実現する．

2 提案手法

従来のタグ参照の概要を図 1 に，提案手法の概要を図 2 に示す．

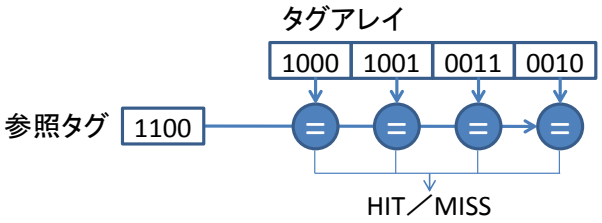

図 1 従来のタグ参照の概要．

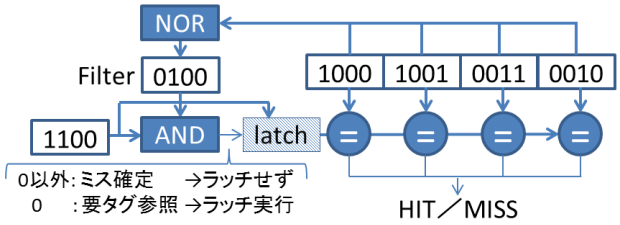

図 2 提案手法の概要．

図 1 は連想度 4 のキャッシュで 4 ビットのタグが保持される場合のタグ一致検索を示している．アドレスから切り出された参照タグはインデックスで指定されたキャッシュのセット内のタグの全て（タグアレイ）を並列に検索し，キャッシュのヒットまたはミスを判断する．この場合，タグ参照の結果がヒットまたはミスであるかに関わらず，全ての参照で必ず全てのタグが比較器で一致判定演算される．

図 2 は図 1 の状況で提案手法が導入されている場合の概要図である．参照タグは比較器の一方の入力として使用される前に，必ずラッチされる．提案手法は全てのクロックサイクルで参照タグをラッチしない．タグの一致比較が必要であると判断された場合にのみ，参照タグがラッチされる．もし，このラッチ出力の値が変化しない場合，トランジスタは入力値を変えないことから，比較器内でトランジスタスイッチが発生せず，タグ比較のための動的電力の消費は発生しない．図 2 で Filter と示される情報は，タグアレイ内の全てのタグを NOR 演算した値である．Filter の任意のビット位置に 1 が立つ場合，全てのタグの当該ビット位置の値が共通で 0 であることを示す．提案手法はこの Filter と参照タグの AND 演算を行う．AND は 2 つの入力値が

Reduction of Energy for Tag References by Filtering Accesses to Low-level Cache Memory
Ryuta ISHIDA†, Tomoaki UKEZONO†, Toshinori SATO†
†Dept. of Electronics Eng. & Computer Sci, Fukuoka Univ.
8-19-1 Nanakuma, Jonan-ku, Fukuoka 814-0180, JAPAN
{tl131260@cis, tukezo@, tsato@ }.fukuoka-u.ac.jp

ともに 1 である場合にのみ 1 を出力することから，この AND 演算結果の任意のビット位置で 1 が出力される場合（演算後の値が 0 でない場合）は，タグアレイ内の全てのタグにおいて共通で 0 であったビット位置に，参照タグでは 1 が立っていることを示す．この場合の Bloom Filter は陰性であり，参照を行う前からミスすることがわかる．前もってミスであることがわかっている場合は，タグ一致検索を実行する必要がない．この場合，提案手法は参照タグをラッチしないことで，タグ一致参照を省略する．AND の演算結果が 0 である場合，これは陰性を否定するものである．つまり，陽性または偽陽性である．この場合は，タグヒットする可能性が存在するため，提案手法は参照タグをラッチし，従来通りのタグ一致検索処理を行う．

3 評価

本研究はシミュレーションベースで提案手法の効果を評価する．評価シミュレーションは SPEC CPU 2006 [3]に収録されるベンチマークの一つである 465.tonto の入力データセット train を使用した実行時のデータアクセスのアドレスを Intel Pin 3.0 [4]を用いて取得する．また同時に，Pin に組み込んだ提案手法の機能シミュレータが，取得したデータアドレス系列を用いてオンタイムにキャッシュの動的な消費エネルギーを積算する．この際の電力指標は，CACTI5.3 [5]により得られたキャッシュのタグアレイ，データアレイの読み込み，書き込みに必要なエネルギーを用いた．

図 3 は図 1 で示した従来のタグ参照と図 2 で示した提案手法を電力シミュレーションし，比較したグラフである．シミュレーションに用いたキャッシュ構成パラメータを表 1 に示す．

図 3 は縦軸が消費電力(J)，横軸が L2 キャッシュの構成を示す．凡例の Filter が提案手法を，No Filter が従来のキャッシュのタグ参照の結果を示す．図 3 から L2 キャッシュでの消費電力は，提案手法は，従来の L2 キャッシュと比較して L2 キャッシュ全体の消費電力を約 30%から 40%削減した．

4 まとめ

本研究は，連想度の高い下位レベルキャッシュへのアクセスを Bloom Filter により制限することで消費電力の削減が可能であることを示した．

今後の課題として，様々な評価対象のベンチマークとキャッシュ構成を組み合わせて評価し，提案手法の適用性に関する評価を行う．加えて，

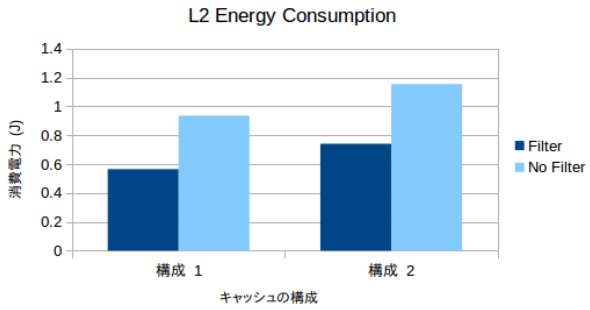

図 3 消費電力の比較.

表 1 キャッシュの構成.

		L1 Cache	L2 Cache
構成 1	容量(Byte)	32K	512K
	ブロックサイズ (Byte)	64	64
	連想度	1	32
構成 2	容量(Byte)	32K	1M
	ブロックサイズ (Byte)	64	64
	連想度	1	32

プログラム実行フェーズに応じて Filter の精度を動的に変化させることができるアーキテクチャレベルの提案を追加で行い，更なる低消費電力化を実現する．

謝辞

本研究は JSPS 科研費　基盤研究(C)2633007 の助成により行われた．

参考文献

[1] A. Ma, M. Zhang and L. Asanovic, “Way memoization to reduce fetch energy in instruction caches”, ISCA Workshop on Complexity Effective Design, July, 2001.
[2] A. Veidenbaum, and D. Nicolaescu “Low Energy, Highly- Associative Cache Design for Embedded Processors” Proc. of IEEE ICCD, pp. 332-335, 2004.
[3] Standard Performance Evaluation Corporation, SPEC CPU2006 Documentation, 2011. https://www.spec.org/cpu2006/Docs/ (参照 2017-1-12)
[4] Intel, Pin 3.0 User Guide, 2012. https://software.intel.com/sites/landingpage/pintool/docs/76991/Pin/html/ (参照 2017-1-12)
[5] Hewlett-Packard Development Company, CACTI5.1, 2008. http://www.hpl.hp.com/techreports/2008/HPL-2008-20.pdf (参照 2017-1-12)