

Drowsy キャッシュにおけるモード切替アルゴリズムの評価

図子 純平[†] 富山 宏之[†] 高田 広章[†] 井上 弘士^{††}

[†]名古屋大学大学院情報科学研究科 〒464-8603 名古屋市千種区不老町

^{††}九州大学大学院システム情報科学研究院 〒819-0395 福岡市西区元岡 744 番地

E-mail: †{junpeiz,tomiyama,hiro}@ertl.jp, ††inoue@i.kyushu-u.ac.jp

概要 組み込み機器において、特にバッテリー駆動型のシステムでは消費エネルギーの削減が重要となる。近年、汎用プロセッサだけでなく組み込み向けプロセッサにもキャッシュメモリが搭載されるようになってきている。また、回路の微細化によりキャッシュメモリにおけるリークエネルギーは年々増加しており、リークエネルギーの削減が求められている。キャッシュのリークエネルギー削減手法のひとつに、Drowsy キャッシュがある。この手法では、キャッシュラインのモードを低リークモードに切り替えることで、リークによる消費エネルギーを削減する。しかし、低リークモードのキャッシュラインへアクセスが発生した場合、ラインを通常モードに切り替える必要があり、この切替には1~数サイクルの切替ペナルティとエネルギーオーバーヘッドが発生する。本論文では、これらの性能低下を最小限に抑えつつ、リークエネルギーを小さくするアルゴリズムとして時間的局所性を応用しモード切替にウェイ予測を用いたウェイ予測 Drowsy キャッシュを提案する。提案手法に対し、性能とリークエネルギーの削減に関しての評価を行う。

Evaluation of Algorithms to Change Cache Line Mode in Drowsy Caches

Junpei ZUSHI[†], Hiroyuki TOMIYAMA[†], Hiroaki TAKADA[†], and Koji INOUE^{††}

[†] Graduate School of Information Science, Nagoya University

Furou-cho, Chikusa-ku, Nagoya, Aichi, 464-8603 Japan

^{††} Department of Informatics, Kyushu University

744 Motoooka, Nishi-Ku, Fukuoka, Fukuoka, 819-0395 Japan

E-mail: †{junpeiz,tomiyama,hiro}@ertl.jp, ††inoue@i.kyushu-u.ac.jp

Abstract In the design of embedded systems, especially battery-powered systems, it is important to reduce energy consumption. In these days, cache memories are used not only in general-purpose processors but also in processors for embedded systems. Static energy (leakage energy) consumed in cache has been increasing with the decrease of the feature size. The Drowsy cache is one of the techniques to reduce leakage energy consumption of caches. The Drowsy cache reduces leakage energy by changing cache line mode into the low-leakage mode. In the Drowsy cache, when the cache line in the low-leakage mode is accessed, it has to be changed into the normal mode, and it takes one or more clock cycles. Thus, these penalty cycles may significantly degrade the cache performance. In this paper, we propose three kinds of *Way-Prediction Drowsy Cache* which achieve a high-energy reduction with the minimum performance overhead. Experimental results demonstrate the effectiveness of the proposed cache architectures.

1. はじめに

現在、大型計算機や家庭用コンピュータ、ノート型 PC などの汎用コンピュータだけでなく、携帯電話をはじめとしたさまざまな家電製品にプロセッサが組み込まれている。特にバッテリー駆動型の組み込みシステムにおいては、プロセッサの低消費電力化は重要な課題である。

昨今、組み込み向けプロセッサにおいて、キャッシュメモリが搭載されるようになってきている。これらのキャッシュの普及と半導体の微細化に伴い、キャッシュにおいて消費されるリークエネルギーは年々増加する傾向にある。

そこで、キャッシュメモリにおけるリークによるエネルギー消費を削減するため、様々な手法が研究されている。具体的には

は Drowsy キャッシュ [1]、DRI キャッシュ^(注1) [4]、Cache Decay [2]、キャッシュ階層を活用した方法 [3]、ウェイ予測キャッシュなどが提案されている。

本研究では Drowsy キャッシュに着目する。Drowsy キャッシュは、キャッシュラインを一定間隔毎に通常モードから低リークモードへ遷移させることにより、リーク電力を削減する。しかし、低リークモードのラインに対するアクセスが発生した場合、1~数サイクルのペナルティが生じ、キャッシュの性能が低下する。本研究では、Drowsy キャッシュをさらに改良したウェイ予測 Drowsy キャッシュを提案する。ウェイ予測により、リークエネルギー消費を抑えつつ、性能低下を抑制する。

本論文は、まず 2 章で Drowsy キャッシュの概要を述べる。3 章で、Drowsy キャッシュの改良であるウェイ予測 Drowsy キャッシュを提案し、4 章でモード切替アルゴリズムの比較を行う。5 章で実験手順について説明を行い、6 章では実験結果について考察を行う。最後に 7 章にて、本研究の結論を述べる。

2. Drowsy キャッシュ

本章では Drowsy キャッシュ [1] の概要と、モード切替アルゴリズムについて説明する。

2.1 概要

Drowsy キャッシュは、一定サイクル間隔ごとに、キャッシュのラインを低リークモードに切り替える。

低リークモードのキャッシュラインでは、キャッシュラインへ供給する電源電圧を低下させることによりリーク電力を削減する。低リークモードでは、SRAM セルの保持する情報が損なわれない程度にキャッシュラインへの供給電圧を低下させる。よって、低リークモードに遷移してもキャッシュミス率は変化しない。低リークモードのキャッシュラインへアクセスする場合には、当該ラインのモードを通常モードへと切り替えた後にアクセスする。低リークモードから通常モードへの切替には、1~数クロックのペナルティサイクル、ならびに、エネルギー消費のオーバーヘッドが発生する。

2.2 モード切替アルゴリズム

本章では、Drowsy キャッシュのキャッシュラインモード切替アルゴリズムについて説明を行う。文献 [1] では 2 種類のアルゴリズムが提案されており、以下、それぞれ Drowsy キャッシュ 1、Drowsy キャッシュ 2 と表記する。

2.2.1 Drowsy キャッシュ 1

文献 [1] で提案されている 2 つの切替アルゴリズムのうち、一定期間ごとにすべてのキャッシュラインのモードを無条件に低リークモードに切り替える方法がある。本研究ではこれを Drowsy キャッシュ 1 とする。

2.2.2 Drowsy キャッシュ 2

文献 [1] で提案されているもう 1 つの切替アルゴリズムは、一定期間ごとに、その期間にアクセスがなされなかったキャッシュラインのみを低リークモードに切り替える方法である。この切替アルゴリズムを Drowsy キャッシュ 2 とする。

3. ウェイ予測 Drowsy キャッシュ

Drowsy キャッシュは、キャッシュラインを低リークモードに切り替えることで、リークエネルギーを削減する。しかし低リークモードのキャッシュラインへのアクセスが発生した場合には、キャッシュラインのモードを通常モードへ切り替える必要がある。このとき、モードの切替には 1~数クロックの性能のペナルティ、および、モードを切り替えるためのエネルギーのオーバーヘッドを要する。より多くのキャッシュラインを低リークモードに切り替えることにより、キャッシュラインで消費されるリークエネルギーを削減できる。しかし同時に、キャッシュラインのモード切替回数が増加し、性能の低下を招く。

本研究は、低リークモードから通常モードへの切替回数が少なく、同時に不必要なキャッシュラインは極力低リークモードへと切り替えるアルゴリズムを探索する。本章では、ウェイ予測機構を Drowsy キャッシュに応用した、3 種類のウェイ予測 Drowsy キャッシュを提案する。以下、3 種類のウェイ予測 Drowsy キャッシュを、ウェイ予測 Drowsy キャッシュ 1、ウェイ予測 Drowsy キャッシュ 2、ウェイ予測 Drowsy キャッシュ 3 と表記する。ウェイ予測には、MRU (Most Recently Used) アルゴリズムを用いる。

3.1 ウェイ予測 Drowsy キャッシュ 1

ウェイ予測 Drowsy キャッシュ 1 は、一定サイクル間隔ごとに、キャッシュの全てのセットにおいて MRU ウェイ以外を低リークモードに切り替える。切り替え直後は、全てのキャッシュセットで MRU ウェイのみが通常モードに残される。

3.2 ウェイ予測 Drowsy キャッシュ 2

ウェイ予測 Drowsy キャッシュ 2 は、キャッシュの MRU ウェイのうち、一定間隔内にアクセスがなされたウェイのみを通常モードに残し、それ以外を低リークモードに切り替える。この切替方法では、通常モードに残されるキャッシュラインが、ウェイ予測 Drowsy キャッシュ 1、および、以下に述べるウェイ予測 Drowsy キャッシュ 3 に比べて少なくなる。このため、リークエネルギーの削減率は最も高くなると予想される。ただし、性能の低下も最も大きくなる。

3.3 ウェイ予測 Drowsy キャッシュ 3

MRU ウェイと、一定間隔以内にアクセスがあったライン全てを通常モードに残し、それ以外を低リークモードに切り替える。この切替方法では、3 種類のウェイ予測 Drowsy キャッシュの中で、低リークモードへ切り替えられるキャッシュラインの数が最も少なくなる。このため、性能の低下は最も低くなると予想される。ただし、リークエネルギーの削減率は最も低くなる。

4. 評価するモード切替アルゴリズム

本章では、2 章及び 3 章で説明した合計 5 種類のアルゴリズムを整理する。5 種類のモード切替アルゴリズムを表 1 にまとめる。

表において、「通常モードに残す条件」の「MRU」の列は、モード切替に MRU アルゴリズムを使用している場合は○、そうでない場合は×、MRU アルゴリズムを使っているが、MRU

(注1) : Dynamically Resizable cache: 動的にキャッシュサイズ変更が可能なキャッシュメモリ

表 1 各種モード切替アルゴリズムの比較

	通常モードに残す条件		モード切替に伴う Awake ライン数		
	MRU	周期内アクセス	最小	最大	モード切替直後の最大
Drowsy1	×	×	0	ウェイ数	0
Drowsy2	×	○	0	ウェイ数	0~ウェイ数
WPDrowsy1	○	×	1	ウェイ数	1~ウェイ数
WPDrowsy2	△	△	0	ウェイ数	0~ウェイ数
WPDrowsy3	○	○	1	ウェイ数	1~ウェイ数

ウェイもモード切り替えの対象となりうる場合は△と記されている。同様に、「周期内アクセス」の列では、ラインのモード切替の条件として、モード切替周期内のアクセスの有無を利用する場合は○、そうでない場合は×となっている。また、ウェイ予測 Drowsy キャッシュ2 では、切替周期内にアクセスがあったウェイでも、それが MRU でない場合には、低リークモードへと切り替えられる。そのため、△と記されている。

表において、「モード切替に伴う Awake ライン数」の「最小」の列は、通常モードとして残る最小のライン数を表している。一方、「最大」の列は、各セット上で通常モードとなりうる最大のライン数を表す。また、「モード切替直後の最大」の列は、モード切替直後に通常モードとなりうる最大のライン数を表す。

5. 実験手順

5.1 エネルギーと性能の評価式

提案したウェイ予測 Drowsy キャッシュの有効性を評価するため、5種類のモード切替アルゴリズムについて、リークエネルギーの削減率と性能の低下率を評価した。

リークエネルギー LE_{total} の評価は、文献 [5], [6] を参考に、以下の近似式を用いた。

$$LE_{total} = CC \times LE_{line} \times CSize$$

$$CC = CC_{conv} + CC_{extra}$$

$$LE_{line} = SR \times LE_{aline} + (1 - SR) \times LE_{aline}$$

ここで、 CC は実行サイクルの総数、 LE_{line} はビット当りのリークエネルギー、 $CSize$ はキャッシュサイズを表す。実行サイクル数 CC は、従来型の低リークモードを備えていないキャッシュを想定した場合のサイクル数 CC_{conv} と、低リークモードから通常モードへの遷移に要するサイクル数 CC_{extra} の和として表される。ライン当りのリークエネルギー LE_{line} は、キャッシュ全体のうち低リークモードとなっているラインの割合 SR と低リークモードのライン当りのリークエネルギー LE_{aline} の積に、通常モードのラインの割合 $(1 - SR)$ と通常モードのライン当りのリークエネルギー LE_{aline} の積を加えたものである。

また、キャッシュラインのモード切替によるエネルギー CE_{total} は、次式で評価する。

$$CE_{total} = CE_{line} \times MCTimes$$

ここで、 CE_{line} はキャッシュライン当りのモード切替エネルギーであり、 $MCTimes$ はモード切替回数の総和である。

モード切替に伴うエネルギーのペナルティとリークエネルギーとを足し合わせたエネルギー E_{total} は、

$$E_{total} = LE_{total} + CE_{total}$$

と定義される。本実験において、エネルギーの評価には E_{total} を用いる。

性能低下率 $PDown$ は次の式により評価する。

$$PDown = \frac{C_{incr}}{C_{conv}}$$

ここで、 C_{incr} はモード切替に起因するペナルティサイクルの総和、 C_{conv} はモード切替ペナルティ以外のサイクル総数である。 C_{conv} には、実行に必要なサイクル数と、キャッシュミス時のミスペナルティが含まれる。

キャッシュのエネルギーの評価値には、文献 [1] のエネルギーの値を参考に、ラインサイズ 32B の場合として表 2 の値を用いた。また、キャッシュラインを低リークモードから通常モードに切り替える際に要するペナルティサイクルに関しては、文献 [1] を参考に、1 サイクルと 3 サイクルの 2 つの場合について実験を行った。

表 2 評価に用いたエネルギー値 (単位: J)

LE_{aline}	LE_{aline}	CE_{line}
3.92E-13	6.63E-14	2.56E-11

5.2 実験環境

5種類のモード切替アルゴリズムを実装したキャッシュシミュレータを作成した。作成したキャッシュシミュレータは、データメモリへのアクセスのトレースを入力とし、プログラム実行時の低リークモードのキャッシュラインの割合、キャッシュラインの低リークモードから通常モードへの切替の総数、モード切替によるペナルティサイクルの総数など、エネルギー削減率や性能低下率の評価に必要な情報を出力する。

本実験では、SimpleScalar シミュレータ [9] を用いてデータメモリへのアクセスのトレースを取得した。その際、各ベンチマークプログラムの実行安定期における約 1000 万命令の実行についてデータメモリアクセスのトレースを取得した。

本研究では、高性能な汎用プロセッサではなく、組み込みプロセッサを想定し、プロセッサの CPI は 1 と仮定した。また、キャッシュは L1 データキャッシュのみを想定し、キャッシュミスペナルティは 20 サイクルとした。モード切替のペナルティサイクルは、先述の通り、1 サイクルまたは 3 サイクルとした。モード切替の周期は文献 [1] を参考に 4096 サイクルとした。また、キャッシュラインサイズは 32B とし、連想度は 2、4、8 の

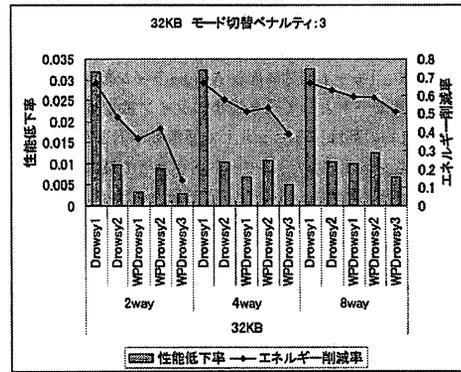
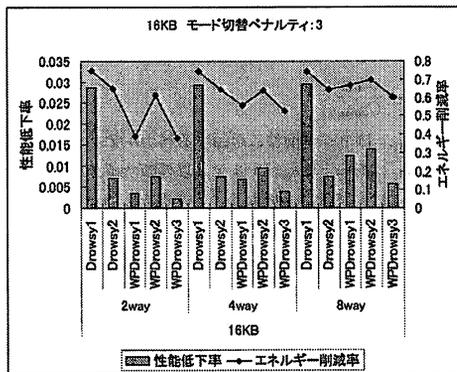
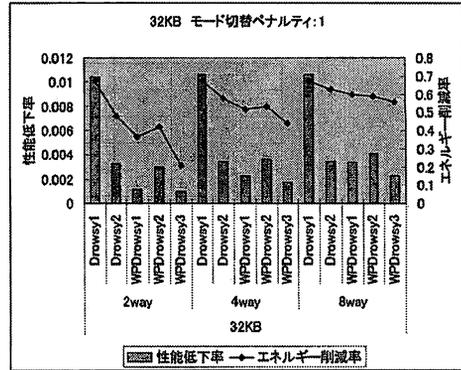
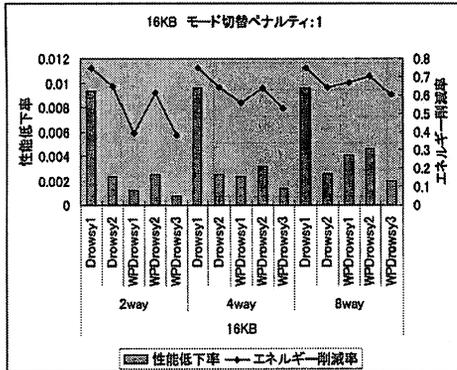


図1 実験結果 (キャッシュサイズ 16KB)

図2 実験結果 (キャッシュサイズ 32KB)

3通り、キャッシュサイズは16KB、32KBの2通りについて実験を行った。

ベンチマークプログラムとして、SPEC2000の ammp、art、bzip2、equake、gcc、gzip、mcf、mesaを用いた。

6. 実験結果

各種モード切替アルゴリズムの性能低下率とエネルギー削減率を図1と図2に示す。図1は、キャッシュサイズが16KBの場合の結果であり、上のグラフがモード切替ペナルティが1サイクルの場合、下のグラフは3サイクルとした場合を示している。図2はキャッシュサイズが32KBの場合の結果である。図において、Drowsy1は一定サイクル間隔ごとに全てのキャッシュラインを低リークモードに切り替える方法(2章におけるDrowsy キャッシュ1)、Drowsy2は一定サイクル間隔ごとに、一定期間内にアクセスがあったウェイトを通常モードに残し、それ以外を低リークモードに切り替える方法(Drowsy キャッシュ2)を表す。WPDrowsy1は、キャッシュの全てのセットにおいてMRU ウェイト以外を低リークモードに切り替える方法(3章におけるウェイト予測 Drowsy キャッシュ1)、WPDrowsy2は一定期間内にアクセスのあったMRU ウェイト以外のラインで低リークモードに切り替える方法(ウェイト予測 Drowsy キャッシュ2)、WPDrowsy3は一定期間内にアクセスのあったウェイトとMRU

ウェイトの両方を通常モードに残し、それ以外を低リークモードに切り替える方法(ウェイト予測 Drowsy キャッシュ3)を表している。図1と図2の結果は、実験で用いたすべてのベンチマークプログラムの平均である。

今回評価したエネルギーには、キャッシュのリークエネルギーと、キャッシュラインのモード切替に要するエネルギーが含まれている。Drowsy1は、5種類のモード切替アルゴリズムの中で、最も高いエネルギー削減率を達成している一方、性能低下率は最悪であった。

Drowsy キャッシュにおける性能低下の原因は、低リークモードのキャッシュラインにアクセスがあった場合に、キャッシュラインのモードを低リークモードから通常モードへ切り替える際に要する遅延(本研究では1サイクルまたは3サイクルと仮定)である。当然ながら、キャッシュラインが低リークモードとなっている割合が高くなるほど、ラインの切替回数が増加するため、性能も低下する。一定周期で全ラインを低リークモードに切り替える手法Drowsy キャッシュ1は、全てのキャッシュラインのモードを切り替えるため、近い将来再びアクセスされる可能性の高いラインも低リークモードへと切り替えられ、その結果、性能低下が大きくなる。本論文で提案するウェイト予測Drowsy キャッシュは、メモリ参照の時間的局所性を活用することにより、キャッシュラインの不要なモード切替の回数を削減

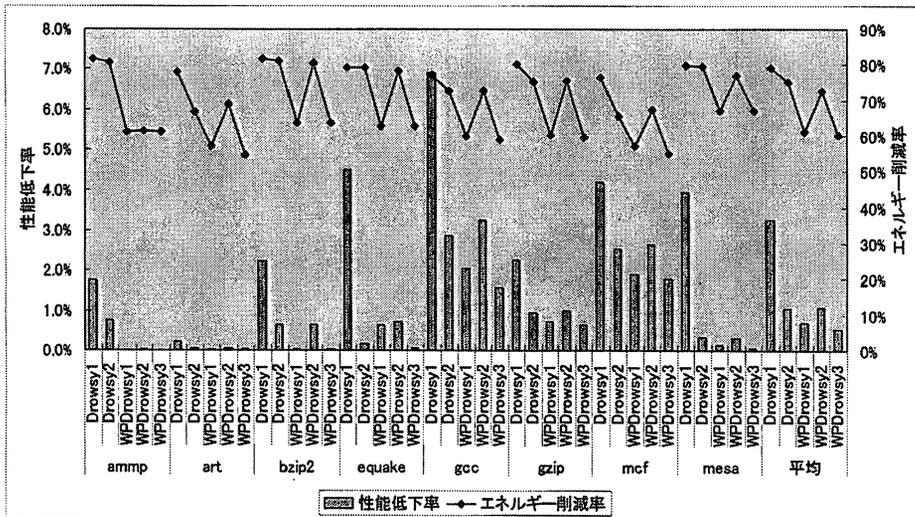


図3 ベンチマーク毎の実行結果 (キャッシュサイズ 32KB、連想度 4、モード切替ペナルティ 3 サイクル)

し、性能低下を抑制することを狙っている。実験の結果、ウェイ予測 Drowsy キャッシュでは、不要なモード切替の回数が削減され、性能低下が抑制されることが確認された。

キャッシュの連想度が大きくなると、ウェイ予測 Drowsy キャッシュの性能低下率が増加しているが、エネルギー削減率も高くなっている。

図3に、キャッシュサイズが 32KB、連想度が 4、モード切替ペナルティが 3 の場合における、各ベンチマークの実験結果を示す。プログラムにより性能低下率が大きく異なっている。

7. おわりに

半導体の微細化に伴い、キャッシュにおいて消費されるリークエネルギーは年々増加する傾向にある。文献 [1] で提案された Drowsy キャッシュは、高いリークエネルギーの削減率を達成する一方、キャッシュの性能が低下するという問題があった。本論文では、高いリーク削減率と性能低下の抑制の両立を目的として、Drowsy キャッシュにウェイ予測機構を導入したウェイ予測 Drowsy キャッシュを提案した。実験により、ウェイ予測 Drowsy キャッシュは、高いリーク削減率を達成しつつ、性能低下を抑制することを示した。

今回は 2 種類の Drowsy キャッシュ、3 種類のウェイ予測 Drowsy キャッシュのみの評価を行ったが、今後は、文献 [8] で提案されているアルゴリズムについても評価を行う必要がある。

謝 辞

本研究を進めるにあたり、多くの助言をいただいた名古屋大学情報科学研究科の本田晋也博士、九州大学 大学院システム情報科学研究科の小宮礼子氏に深く感謝致します。本研究の一部は、科学技術振興事業団 (JST) 戦略的創造研究推進事業

(CREST)「情報システムの超低消費電力化を目指した技術革新と統合化技術」の支援による。

文 献

- [1] K. Flautner, N. Kim, S. Martin, D. Blaauw, T. Mudge "Drowsy Caches: Simple Techniques for Reducing Leakage Power," *International Symposium on Computer Architecture*, May 2002.
- [2] S. Krkiras, Z. Hu, M. Martonosi "Cache Decay: Exploiting Generational Behavior to Reduce Cache Leakage Power," *International Symposium on Computer Architecture*, July 2001.
- [3] L. Li, I. Kadayif, N. Vijaykrishnan, M. Kandemir, M. J. Irwin, A. Sivasubramaniam, "Leakage Energy Management in Cache Hierarchies," *International Conference on Parallel Architectures and Compilation Techniques*, Sep. 2002.
- [4] S. Yang, M. D. Powell, B. Falsafi, K. Roy, T. N. Vijaykumar, "An Integrated Circuit/Architecture Approach to Reducing Leakage in Deep-Submicron High-Performance I-Caches." *International Symposium on High-Performance Computer Architecture*, Jan. 2001.
- [5] R. Komiya, K. Inoue, V. G. Moshnyaga, K. Murakami, "Quantitative Evaluation of Leakage Reduction Algorithm for L1 Data Caches," *IEICE Trans. Fundamentals*, vol. E88-A, no. 4 pp. 862-868, Apr. 2005.
- [6] 小宮礼子, 井上弘士, モニャガ・ワシリー, 村上和彰, "キャッシュ・リーク電力削減アルゴリズムに関する定量的評価," 第 17 回路とシステム軽井沢ワークショップ, 2004 年 4 月.
- [7] K. Inoue, T. Ishihara, and K. Murakami, "Way-Predicting Set-Associative Cache for High Performance and Low Energy Consumption," *International Symposium on Low Power Electronics and Design*, Aug. 1999.
- [8] S. Petit, J. Sahuquillo, J. Such, and D. Kaeli, "Exploiting Temporal Locality in Drowsy Cache Policies," *Computing Frontiers*, May. 2005.
- [9] D. Burger, T. M. Austin, "The SimpleScalar Tool Set Version 2.0," *UW Madison Computer Sciences Technical Report #1342*, June. 1997.