

キャッシュメモリに不揮発性メモリを用いたプロセッサの省電力化の研究

高性能コンピューティング学講座 本多・近藤研究室

1053010 坂田 英

主任指導教員：近藤正章

1 研究概要

近年、プロセッサの電源管理として、コアを一時的にサスペンド状態にし、電源を遮断して消費電力を削減する手法が広く用いられている。例えばインテルプロセッサでは、複数の低電力動作モードを搭載しており、プロセッサの状態に応じて各部電源を遮断し、消費電力を削減している。しかし、従来のシステムは、プロセッサ全体の電源を遮断するとキャッシュメモリのデータが失われる。そのため、プロセッサの電源遮断時にはデータの書き戻しのため、また、復帰時にはキャッシュミスが頻発し性能が低下するため、時間的オーバーヘッドが発生する。このオーバーヘッドから、プロセッサの電源を遮断する状況が限られてしまう。

本研究は、不揮発性メモリをキャッシュメモリに用いることで、前述のオーバーヘッドの課題を解決する。従来、不揮発性メモリはリーク電力を非常に少なくすることができるため、特に主記憶の待機電力を削減するデバイスとして多くの研究がなされている。本研究は、不揮発性メモリをキャッシュメモリに用いる新たなシステムを提案することで、プロセッサ全体の電源遮断の機会を増加させ、電力削減を狙う。

2 研究背景

2.1 キャッシュメモリ

キャッシュメモリはCPUと主記憶の性能差を埋めるために用いる高速で小容量のメモリである。通常、キャッシュメモリにはDRAMより非常に処理が高速なSRAMが用いられている。

SRAMはフリップフロップ等の順序回路を用いてデータを記憶する揮発性メモリである[1]。しかし、従来のSRAMキャッシュメモリは揮発性メモリのため、プロセッサ全体の電源を遮断するとキャッシュ内のデータが失われる。そのため、プロセッサの電源を遮断する前に、キャッシュメモリ内のデータを主記憶に書き戻す必要がある。また、プロセッサの電源復帰時にはキャッシュミスが頻発し、性能が低下するため、時間的オーバーヘッドが発生する。プロセッサが長い間待機状態になる場合は電源を遮断することができるが、待機状態が短い場合は、前述の書き戻しお

よびキャッシュミスのオーバーヘッドから、プロセッサの電源を遮断することが難しい。

2.2 不揮発性メモリ

不揮発性メモリは揮発性メモリであるDRAMやSRAMに比べ、常時電源を供給する必要がないため、非常にリーク電力を削減することができる。そのため、主記憶の待機時の消費電力を削減するデバイスとして、多く研究がされている[2]。

本研究は、キャッシュメモリとして、不揮発性メモリの1つである *Magneto resistive Random Access Memory*(以下、MRAM)を用いることを想定する。SRAMとMRAMの性能を表1に示す。

表1：SRAMとMRAMの性能

	SRAM	MRAM
不揮発性	×	○
Read latency(ns)	2~8	5~20
Write latency(ns)	2~8	5~50
セル面積(F ²)	50~100	10~20

MRAMはSRAMに比べ、ReadおよびWriteの処理速度は劣るが、不揮発性であることから、電力供給が無くてもデータを保持することができる。また、他の不揮発性メモリに比べて書き換え回数が長く、寿命が長い。

3 提案システム

本研究は短時間でもプロセッサの電源を遮断することで、消費電力を削減することを目的とし、キャッシュメモリに不揮発性メモリであるMRAMを用いるシステムを提案する。提案システムを図1に示す。

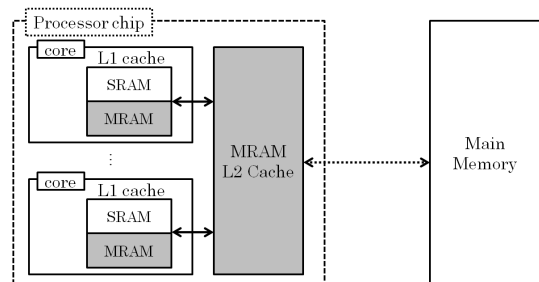


図1: キャッシュメモリにMRAMを用いた提案システム

MRAMを用いることで、プロセッサの電源を遮断してもキャッシュメモリ内のデータが失われない。そのため、電源を遮断する前に、キャッシュメモリ内のデータを主記

憶に書き戻すオーバーヘッドと、プロセッサの電源復帰時に発生するキャッシュミスのオーバーヘッドを改善できる。これにより、短時間でもプロセッサの電源遮断ができ、電源遮断の機会を増やすことで、消費電力を削減する。

3.1 提案システムの利点

キャッシュメモリにMRAMを用いる利点として、具体的に以下の2点がある。

- リーク電力の削減

SRAMはLSIの製造プロセスの微細化に伴い、リーク電力が増大している。一方、MRAMは2枚の磁性体のスピンの向きの違いによる抵抗差によってデータを保持していることから、リーク電力が非常に少ない。MRAMとSRAMのリーク電力を表2に示す[3]。

表2: SRAMとMRAMのリーク電力

	SRAM	MRAM
Leakage power(W)	2.089	0.255

- 電源遮断制御の細粒度化

従来のSRAMキャッシュメモリは、プロセッサの電源を遮断するとキャッシュ内のデータが失われる。そのため、プロセッサの電源復帰時にキャッシュミスが頻発に発生する課題がある。

本提案システムはキャッシュメモリにMRAMを用いることで、主記憶に書き戻す必要を無くし、キャッシュミスのオーバーヘッドを改善する。例えば、数ミリ秒間プロセッサがアイドル状態になった場合、SRAMでは電源遮断時に書き戻しやキャッシュミス等の時間的オーバーヘッドのため、かえって消費電力が増加してしまう可能性がある。しかし、MRAMは書き戻しやキャッシュミス等の時間的オーバーヘッドが発生しないことから、瞬時にプロセッサをサスペンド状態することが可能である。これにより、プロセッサの電源を遮断する機会を増やし、消費電力を削減する。

3.2 提案システムの課題

MRAMはSRAMに比べ、ReadおよびWrite処理速度が劣る課題がある。そのため、通常のキャッシュメモリの読み込みや書き込みにより、性能低下が考えられるので、今後検討する必要がある。

4 進捗状況

現在までに行った事として、Zesto X86 simulatorとSPEC2006ベンチマークを用いてSRAMとMRAMをL2キャッシュに用いた場合の実行時間を測定した。図2に実行時間の比較結果を示す。

図2はSRAM L2キャッシュとMRAM L2キャッシュの各ベンチマークの実行時間を比較したものである。シミュレーションの結果、MRAMはSRAMに比べ、Read

およびWriteの処理速度が劣ることから、実行時間が平均約1.2倍になった。特に、MRAMはWrite処理速度が非常に遅いため、401.bzip2や445.gobmk等のWriteアクセスの多いベンチマークでは約1.4倍になった。

これらから、Write処理速度が遅い課題を検討していく予定である。

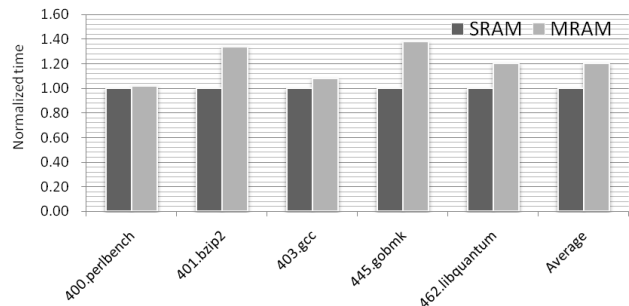


図2: SRAMおよびMRAM L2キャッシュの実行時間

5 今後の方針

MRAMをL2キャッシュに用いた場合、SRAM L2キャッシュに比べてレイテンシが大きいことがこれまでのシミュレーションの結果から分かっている。

本提案システムは、L1キャッシュにSRAMとMRAMを組み合わせた提案システムを実現していく予定である。MRAMをL1キャッシュに用いるには、さらなる工夫が必要であるため、今後検討していく。

6 まとめ

本研究はMRAMをキャッシュメモリに用いるシステムを提案し、プロセッサの電源遮断制御の細粒度化を行う。これにより、プロセッサの電源を遮断する機会を増やし、消費電力を削減する。

現在は提案システムを実現するための第一段階として、SRAMおよびMRAMキャッシュメモリの基礎的なデータを調査している。今後は提案システムの課題である、高いレイテンシを検討していく。また、L1キャッシュにSRAMとMRAMを組み合わせた提案システムを実現していく予定である。

参考文献

- [1] 高田雅史, 中山和也, 泉貴富, 新村達, 北川章夫, “相変化メモリを利用した不揮発性SRAMアーキテクチャの研究” IEICE集積回路研究会(ICD), Vol.106, No.2, pp.49-54, (2006)
- [2] Yongsoo Joo, et al. “Energy and Endurance-Aware Design of Phase Change Memory Caches” *Design, Automation and Test in Europe*, pp.136-141. 2010.
- [3] Guangyu Sun and Xiangyu Dong and Yuan Xie and Jian Li and Yiran Chen “A Novel Architecture of the MRAM L2 Cache for CMPs” *High Performance Computer Architecture*, pp. 239-249. 2009.