

高速・低消費電力STT-MRAM キャッシュを用いた Run-time ノーマリオフプロセッサ

Runtime Normally-Off Processor with Nonvolatile Cache Memory Based on High-Speed and Low-Power STT-MRAMs

野村 久美子 安部 恵子 藤田 忍

■NOMURA Kumiko ■ABE Keiko ■FUJITA Shinobu

近年、CMOS（相補型金属酸化膜半導体）の微細化により、プロセッサの待機電力が大きくなってきている。このため、アプリケーションが停止している間は不要な回路ブロックの電源を遮断する、パワーゲーティングが用いられているが、これだけでは消費電力を十分に削減できない。

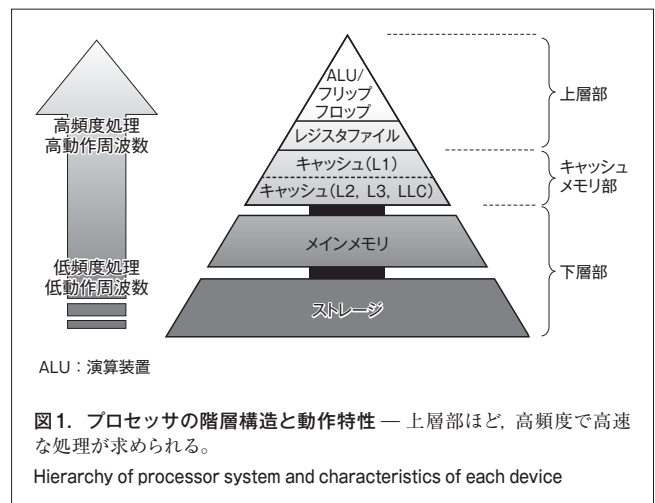
そこで東芝は、不揮発性キャッシュメモリを用いた新しい超低消費電力プロセッサの研究開発を進めている。このプロセッサは、アプリケーションが動作中（Run-time）でも、非常に短い時間間隔で電源の遮断と復帰を行うことで待機電力を削減できる。高速かつ超低消費電力で動作可能な垂直磁化型STT-MRAM（Spin Transfer Torque - Magnetoresistive RAM）を使ったキャッシュメモリによってこれが実現できることを、シミュレーションで実証した。

The total standby power of processor systems has been rapidly increasing with complementary metal-oxide semiconductor (CMOS) scaling. To reduce this standby power, processor systems use a power gating technique that regulates the power supply to each circuit block. The power supply of each circuit block is cut off when an application is suspended. However, this technique is unable to sufficiently reduce the total standby power consumption.

Toshiba has been engaged in research and development of a new ultra-low-power processor technique using nonvolatile magnetoresistive random access memories (MRAMs), which can reduce the standby power substantially by controlling the power supply even during application execution time. The results of simulations have confirmed that this new processor system can be realized by applying high-speed and ultra-low-power spin-transfer torque (STT)-MRAMs to the cache memory.

1 まえがき

CMOSの高集積化が進み、プロセッサの総消費電力に占める待機電力の割合が増加している。プロセッサ全体の低消費電力化を行うには、この待機電力の削減が重要な課題となる。高性能プロセッサやSoC（System on a Chip）では、待機状態の間、不要な回路ブロックの電源を遮断するパワーゲーティング技術が用いられているが、プロセッサ内のメモリが揮発性であるため、電源遮断を行えない場合が多い。これに対して、プロセッサ内のメモリを不揮発性メモリとして、待機している状態では完全に電源遮断を行う“ノーマリオフコンピュータ”という概念が提案されており、いくつかの関連技術が研究されている⁽¹⁾。東芝は、このノーマリオフ動作の概念を更に進めて、アイドル中の待機状態だけでなく、アプリケーションが動作中（Run-time）でも、非常に短い時間間隔で電源の遮断と復帰を行うことができる“Run-time ノーマリオフプロセッサ”を提案している⁽²⁾。より細かな時間間隔で電源の制御を行うことによって消費電力の更なる削減を実現ができる。更に、不揮発性メモリをキャッシュメモリに採用することで、電源遮断時の欠点であるキャッシュのデータの消失がなく、立上り及び立下り時のオーバヘッドを小さくできる。



ここでは、当社が開発したMTJ（磁気トンネル接合）素子で構成した超低消費電力垂直磁化型STT-MRAM（Ultra-Low-Power (ULP)-STT-MRAM⁽³⁾）を使用したキャッシュメモリ（以下、Magnetic キャッシュメモリと記す）から成るRun-time ノーマリオフプロセッサを想定し、アプリケーションを動作させたときの消費電力と処理性能のシミュレーションを行って、有効性を検証した結果について述べる。

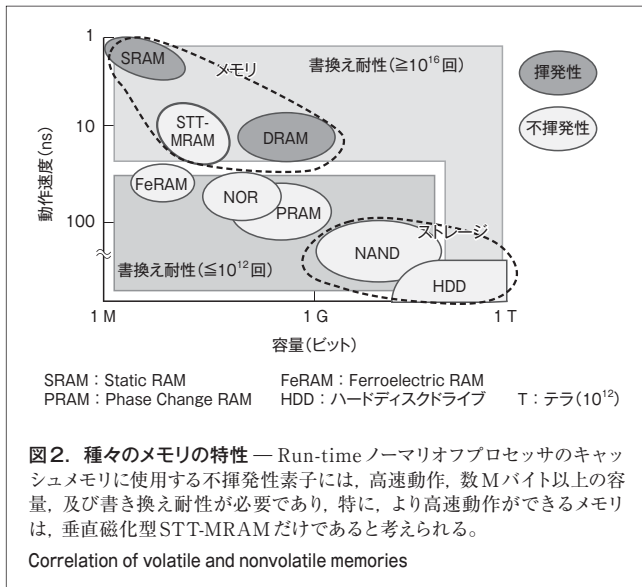
2 プロセッサの階層と不揮発性メモリの適用

より効果的に低消費電力化を行うためには、メモリ内で不揮発性メモリを適用する箇所が重要になる。

一般的なプロセッサの階層構造を図1に示す。下層部はメモリ部、上層部は論理演算部であり、その間にキャッシュメモリ部がある。上層部ほどシステム動作頻度が高く動作時の消費電力（以下、動作電力と略記）が支配的になり、また、従来の不揮発性メモリを用いた論理回路の動作電力は揮発性メモリを用いた論理回路に比べてはるかに大きい。このため、不揮発性メモリを適用しても、従来より消費電力が増大してしまう⁽⁴⁾。一方、プロセッサのキャッシュメモリは高速動作を要求されるが、動作頻度がそこまで大きくない。更に、トランジスタ数が上層部より多くリーク電流が大きくなることから、消費電力に対する待機電力の割合が大きくなる。このため当社は、プロセッサ内のキャッシュメモリに不揮発性メモリの使用が可能であろうと考えている。

プロセッサ内のキャッシュメモリには、上層部からL1, L2, L3, 及びLLC (Last Level Cache)がある。不揮発性メモリが高速かつ低消費電力で動作するほど、上の階層のキャッシュメモリに適用できる。また、上層のキャッシュメモリほど容量が小さく頻繁に動作するため、動作電力が支配的となる。このため当社は、動作頻度がある程度小さいL2キャッシュまでを不揮発性化できると考えている。更に、キャッシュメモリが不揮発性になることで、キャッシュメモリ回路部の電源遮断と復帰を瞬時に行うことができるようになり、アプリケーション動作中でもキャッシュメモリの待機電力を削減できると期待される。

Run-time ノーマリオフプロセッサのキャッシュメモリに使用する不揮発性素子には、高速動作、数Mバイト以上の容量、及び書換え耐性が必要である。様々なメモリについてこれら

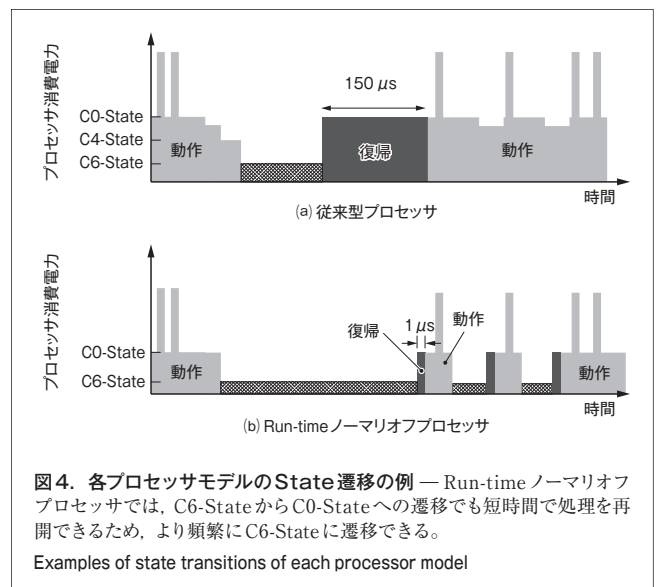
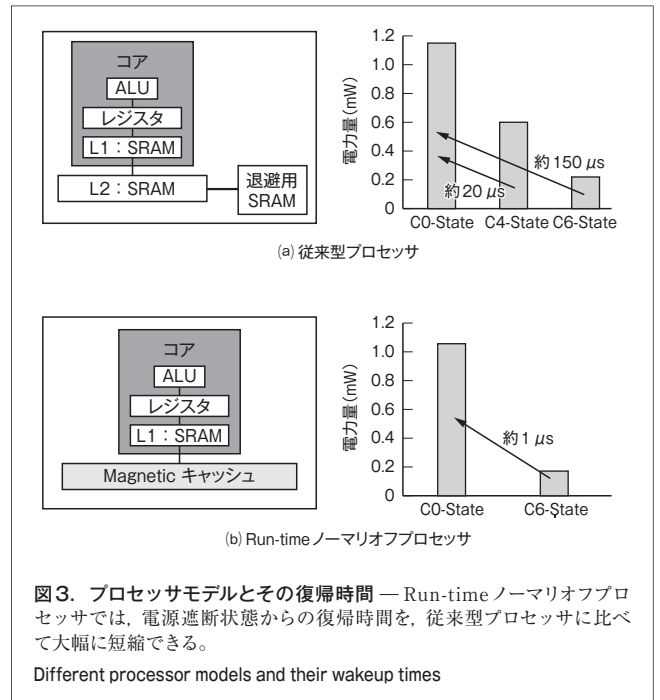


の特性の比較を図2に示すが、垂直磁化型STT-MRAMがもっとも可能性があると考えられる。

3 Run-time ノーマリオフプロセッサ

ここでは、Run-time ノーマリオフプロセッサと、既に製品化されているパワーゲーティング機能のある従来型プロセッサについて、それぞれの構造及びアプリケーション動作中の電力供給状況（以下、Stateと呼ぶ）を述べる。

従来型プロセッサと当社が提案するRun-time ノーマリオフプロセッサの比較を、図3及び図4に示す。従来型プロセッサ



サでは、アプリケーション動作中に電力供給を削減するために、大きく分けて次の三つのStateで動作するように設定されている。

- C0-State: 全てに電源供給している状態
- C4-State: L2キャッシュだけに電源供給した状態
- C6-State: 退避用SRAM (Static RAM)以外の全ての電源供給を遮断した状態

従来型プロセッサでは、アプリケーション実行中に電源が遮断されると、L2キャッシュのデータが消去され、復帰後にはこのデータを復元するために主記憶へのアクセス数が増加して処理時間が増大する。これを避けるため、図3(a)に示すように退避用SRAMを用意し、そこにプロセッサコアの動作情報を退避させ復帰時の処理時間増大を緩和している。しかし、復元のために主記憶から全てのデータを再度読み出すことになるため、C6-StateからC0-Stateへの復帰には長い時間が掛かる。したがって実際には、図4(a)に示すようにある程度の処理待ち時間が予測されるときでなければC6-Stateへの移行は発生することはない。そのため、C6-stateになることはめったにない。

Run-time ノーマリオフプロセッサでは、電源遮断時にキャッシュデータが消失しないため、次の二つのStateだけからなる。

- C0-State: 全てに電源供給している状態
- C6-State: 全ての電源供給を遮断した状態

図3(b)に示すようにC6-StateからC0-Stateへの遷移でも、従来型プロセッサのようなデータの復元がないため、短時間で処理を再開できる。そのため従来型プロセッサでは遷移不可能な短い待機状態でも、図4(b)に示すようにC6-Stateへ遷移でき、電力状態をより微細に変化させることができる。これにより、いっそう効果的な低消費電力化が期待できる。

4 消費電力と動作時間のシミュレーション結果

種々のプロセッサで演算動作を行った場合の総消費電力や性能は、実際に実行するアプリケーションによって大きく変化する。しかし、任意のアプリケーションが動作したときの内部状態を元にプロセッサシミュレーションを行った報告はまだない。これは、あるアプリケーションを実行しているときの内部状態を正確に測定することが一般に非常に難しく、それを元にしたシミュレーションを行えなかったからである。そこで当社は、米国Synopsys, Inc.の動作検証ソフトウェア上にプロセッサシステムを構築し、MPEG-4 (Moving Picture Experts Group-phase4) 動画再生及びビデオゲーム (以下、ゲームと略記) の2種類のアプリケーションを実行した場合のキャッシュアクセス状況を測定した。その結果をプロセッサシミュレータに取り入れることで、実際のアプリケーション動作を想定した総消費電力を計算した。

4.1 シミュレーションの条件

ここでは、次の三つの構成から成るプロセッサシステムについてシミュレーションを行った。

- (1) 全てのキャッシュが揮発性メモリ (SRAM) (図3(a))
- (2) L2キャッシュだけがMagneticキャッシュでL1はSRAM (図3(b))
- (3) L1及びL2キャッシュともMagneticキャッシュ

(1)はパワーゲーティング機能を持った従来型プロセッサシステムであり、(2)及び(3)はRun-time ノーマリオフプロセッサである。

このシミュレーションでは、各キャッシュメモリの動作速度及び消費電力は次のように決めた。SRAMキャッシュに関しては、キャッシュシミュレータCACTI6.5を使用して決定した。Magneticキャッシュに関しては、周辺回路はSRAMと同じであるととし、セル部分にULP-STT-MRAM⁽³⁾を使用したとしてCACTI6.5の結果を元に決定した。

4.2 シミュレーション結果

プロセッサシミュレータを用いて計算した、各プロセッサシステムの総消費電力及び総処理時間を、図5及び図6に示す。どちらも従来型プロセッサ(1)を想定して計算した結果を基準とし、プロセッサ(2)及び(3)の場合の結果を(1)に対する割合として示している。当社がターゲットとしているプロセッサは、スマートフォンなどの携帯端末向けを想定しているため、

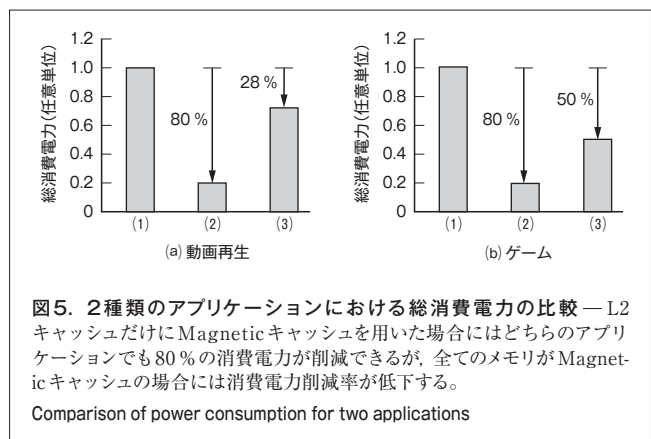


図5. 2種類のアプリケーションにおける総消費電力の比較 — L2キャッシュだけにMagneticキャッシュを用いた場合にはどちらのアプリケーションでも80%の消費電力が削減できるが、全てのメモリがMagneticキャッシュの場合には消費電力削減率が低下する。

Comparison of power consumption for two applications

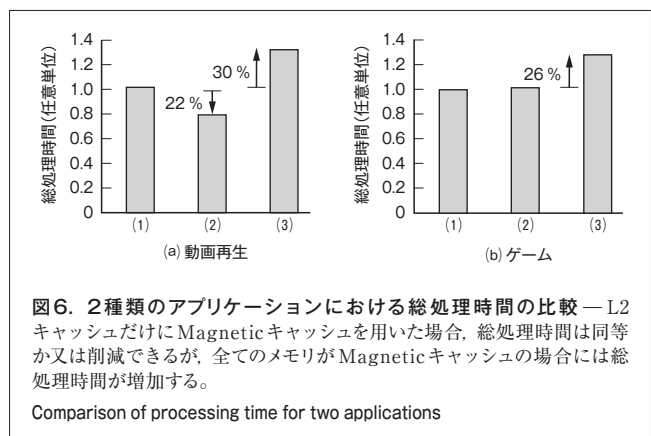


図6. 2種類のアプリケーションにおける総処理時間の比較 — L2キャッシュだけにMagneticキャッシュを用いた場合、総処理時間は同等か又は削減できるが、全てのメモリがMagneticキャッシュの場合には総処理時間が増加する。

Comparison of processing time for two applications

1 GHzで動作しているとして計算を行っている。

図5から、L2キャッシュにMagneticキャッシュを使用した(2)の場合は、どちらのアプリケーションでも80%の消費電力を削減できるが、全てのメモリがMagneticキャッシュの(3)の場合は、消費電力削減率が低下し、動画再生時で28%、ゲーム時でも50%しか消費電力の削減ができないことがわかる。しかも図6から、総処理時間の面でも、(2)の場合は総処理時間を削減できるか又は同等の時間で処理を行えるが、(3)の場合は総処理時間が20%以上増加することがわかる。

これらの結果から、全てのキャッシュにMagneticキャッシュを使用するより、L1を従来と同じSRAMキャッシュにし、L2にMagneticキャッシュを使用するほうが望ましいことがわかる。この結果は、アプリケーションの動作中にはL1キャッシュが頻繁に動作し、現状の不揮発性メモリの動作電力がSRAMと比べて大きいことから、パワーゲーティングを頻繁に行ったとしても動作電力の増大が大きく、消費電力削減効果が見込めない、という2章で述べた説明と一致している。これは、高頻度動作と高周波数動作が求められるCPUコア内の領域で既存の不揮発性メモリを用いたロジック回路を使うと、性能劣化と消費電力増大を招くことを意味する。

CPUコア内のSRAM若しくはレジスタを不揮発性化するためには、SRAMの処理性能と動作電力に相当する不揮発性素子を開発する必要がある。当社は、動作中はSRAMで、待機中はデータをMRAMに保存できる、不揮発性SRAM (Magnetic SRAM) も提案しているが、これをもってしても、L1キャッシュに適用するメリットはほとんどない。コア内部の回路を不揮発性にするには、不揮発性素子であるMTJをSRAM内のインバート回路並みの性能と消費電力に、もっと近づけなければならないからである。

一方、L2キャッシュだけにMagneticキャッシュを採用したノーマリオフプロセッサでは、動画再生及びゲームのどちらのアプリケーションを実行した場合でも、性能劣化を起こすことはなく、80%程度の総消費電力削減が見込めることが確認できた。これは、L2キャッシュの動作頻度が比較的少なく、また(1)のプロセッサがC6-Stateへ遷移できない場合でも、メモリ内のデータが消失せずに復帰時間が短いため、より頻繁にC6-Stateへの移行と復帰ができるためである。つまり、Run-timeでのノーマリオフ動作ができていることを意味している。更にこの結果は、不揮発性メモリをL2キャッシュだけに適用した場合には、動作電力が多少大きくてもL2キャッシュの動作数そのものが少なく、動作電力の増加量より待機電力の削減量が大きくなり、総消費電力を削減できる、という2章で述べた説明に一致している。

5 あとがき

従来の揮発性メモリだけから成るプロセッサと、Magnetic

キャッシュをL2だけ、及びL1とL2キャッシュの両方に使用した3種類のプロセッサについて、実際のアプリケーション動作を元に、ある一定の処理を実行した場合の総処理時間と総消費電力をシミュレーションによって検証した。その結果、現状で作成可能な高速MTJ素子でMagneticキャッシュを構成した場合、それをL2キャッシュに使用することで総消費電力を80%削減できることを確認した。これにより、当社が以前から提案しているメモリ階層構造の妥当性を、初めて実証することができた。同時に、現在もっとも高速であるとされるMTJ素子を使用しても、L1キャッシュやそれより上位のロジック回路へ使用すると、処理速度が低下し、消費電力が増大することも検証できた。以上から、MagneticキャッシュをL2キャッシュに使うことで、Run-time ノーマリオフプロセッサを実現できることが示された。

当社は今後も、超低消費電力Run-time ノーマリオフプロセッサの実用化を目指して、研究開発を進めていく。

この研究の一部は、独立行政法人 新エネルギー・産業技術総合開発機構 (NEDO) の「ノーマリオフコンピューティング基盤技術開発」プロジェクトにおいて実施したものである。

文献

- (1) Kaabouch, N.; Hu, W.-C. Energy-Aware Systems and Networking for Sustainable Initiatives. USA, IGI Global, 2012, 467p.
- (2) Nomura, K. et al. Ultra low power processor using perpendicular-STT-MRAM/SRAM based hybrid cache toward next generation normally-off computers. J.Appl.Phys. 111, 7, 2012, 07E329-1 - 07E329-3.
- (3) Kishi, T. et al. "Lower-current and Fast Switching of A Perpendicular TMR for High Speed and High Density Spin-Transfer-Torque MRAM". 2008 IEEE International Electron Devices Meeting (IEDM). San Francisco, CA, USA, 2008-12, IEEE. 2008, p.309 - 312.
- (4) Abe, K. et al. "Novel Nonvolatile Logic Circuits with Three-Dimensionally Stacked Nanoscale Memory Device". Technical Proceedings of the 2005 NSTI Nanotechnology Conference and Trade Show vol.3. Anaheim, CA, USA, 2005-05, Nano Science and Technology Institute. 2005, p.203 - 206.



野村 久美子 NOMURA Kumiko

研究開発センター LSI 基盤技術ラボラトリー研究主務。
次世代不揮発性メモリの新規システム応用の研究・開発に従事。電子情報通信学会会員。
Advanced LSI Technology Lab.



安部 恵子 ABE Keiko

研究開発センター LSI 基盤技術ラボラトリー研究主務。
次世代不揮発性メモリの回路・システム応用の研究・開発に従事。IEEE 会員。
Advanced LSI Technology Lab.



藤田 忍 FUJITA Shinobu

研究開発センター LSI 基盤技術ラボラトリー研究主幹。
不揮発性メモリ新領域開拓に向けた回路・システム応用の研究・開発に従事。IEEE 会員。
Advanced LSI Technology Lab.