

キャッシュ内蔵 SDRAM のレイテンシを短縮できる メモリコントローラの提案

三浦誓士[†] 秋山悟[†]

[†]日立製作所 中央研究所 〒185-8601 東京都国分寺市東恋ヶ窪一丁目 280 番地
E-mail: [†]{smiura,a-satoru}@crl.hitachi.co.jp

あらまし キャッシュ内蔵 SDRAM のレイテンシを短縮するメモリコントローラを提案した。本コントローラはアドレスアライメント制御ブロックおよびダミーキャッシュ制御ブロックから構成される。ベンチマークプログラムによる性能評価を行った結果、標準 SDRAM のセンスアンプをキャッシュメモリとして利用するセンスアンプキャッシュ制御方式と比較し、レイテンシを最大 25% 短縮でき、プログラム実行時間を最大 13% 短縮できることを確認した。また、0.18- μ m CMOS テクノロジを想定し、本コントローラのゲート規模を評価した。その結果、電源電圧 1.8V 時に、動作周波数 133MHz を実現する本コントローラのゲート規模は 9.2k ゲートとなった。

キーワード^{*} キャッシュメモリ内蔵 SDRAM、メモリコントローラ、レイテンシ、ダミーキャッシュ制御

A Memory Controller that Reduces Latency of Cached SDRAM

Seiji MIURA[†] and Satoru AKIYAMA[†]

[†]Hitachi, Ltd. Central Research Laboratory 1-280 Higashi-koigakubo, Kokubunji-shi, Tokyo, 185-8601 Japan
E-mail: [†]{smiura,a-satoru}@crl.hitachi.co.jp

Abstract The proposed controller has two main control schemes, address-alignment control and dummy-cache control scheme. These two schemes cooperatively control cached SDRAM to reduce its latency. Testing of the controller using benchmark programs demonstrated that latency was reduced 25% and execution time was reduced 13% compare to those of a sense-amplifier cache controller for standard SDRAM. The proposed controller requires 9.3K gates at a supply voltage of 1.8V and an operating frequency of 133MHz.

Keyword cached SDRAM, memory controller, latency, dummy-cache control scheme

1. はじめに

CPU と DRAM との動作周波数の差は依然として大きく、DRAM のレイテンシがシステム性能に大きな影響を及ぼしている。このような中、システム性能を向上するために、DRAM のレイテンシを短縮するメモリコントローラが発表されている [1-2]。これらメモリコントローラは SDRAM のセンスアンプをキャッシュメモリとして利用している (センスアンプキャッシュ制御)。このセンスアンプキャッシュ制御では、ヒットの場合は、直接センスアンプからデータを読み出すことができる。

一方、メモリアーキテクチャからのアプローチで、キャッ

シュメモリを内蔵した SDRAM (キャッシュ内蔵 SDRAM) も開発されている [3]。このキャッシュ内蔵 SDRAM では、内蔵キャッシュメモリへヒットした場合は、低レイテンシでデータを読み出すことができる。しかし、ダーティミスの場合は、要求データを読み出す前に、ライトバック動作が必要なため、レイテンシは大幅に増加してしまう。

そこで、この問題を解決し、キャッシュ内蔵 SDRAM のレイテンシを短縮できる新しいメモリコントローラを提案した [4]。以下、第 2 章ではキャッシュ内蔵 SDRAM の読み出し方法について述べ、第 3 章では本メモリコントローラの詳細を説明する。第 4 章ではベンチマークプログラムによる本メモリコントローラと、従来センスアンプキャッシュ制御との性

能比較を行う。

2. キャッシュ内蔵 SDRAM のレイテンシ

キャッシュ内蔵SDRAMの読み出し動作を図1に示す。キャッシュ内蔵SDRAMは、16個のキャッシュブロック (C-block0～C-block15) からなるキャッシュメモリと、4バンクのメモリアレイから構成される。さらに、各バンクのメモリアレイは、4個のアレイブロック (A-block0～A-block3) から構成される。

まず、ヒット時の読み出し動作について説明する。CPUが要求したデータが、キャッシュメモリに存在する場合 (ヒット) は、リード命令 (Rd) とカラムアドレス (C0) によって、キャッシュメモリから直接、データを読み出すことができる (①)。この際の読み出しレイテンシは2サイクルとなる。

次に、クリーンミス時の読み出し動作について説明する。CPUが要求したデータがキャッシュメモリに存在せず、且つ、キャッシュメモリとメモリアレイ内データとのコヒーレンシが保たれている場合 (クリーンミス) は、まず、バンクアクティブ命令 (Ac) とロウアドレス (R0) によって、メモリアレイ内の1Kバイトデータをセンスアンプへ転送し、保持する (①)。次に、ロード命令 (Ld) とアレイブロックアドレス (Ab2) によって選択されたセンスアンプ内の256バイトデータを、キャッシュブロックアドレス (Cb15) によって選択されたキャッシュブロックへ転送する (②)。そして、リード命令 (Rd) とカラムアドレス (C0) によって、キャッシュメモリからデータを出力する (③)。最後にプリチャージ命令 (Pre) によって、センスアンプに保持しているデータを消去する (④)。このように、クリーンミスの場合は、レイテンシは7サイクルとなる。

最後に、ダーティミス時の読み出し動作について説明する。CPUが要求したデータがキャッシュメモリに存在せず、且つ、キャッシュメモリとメモリアレイ内データとのコヒーレンシが保たれていない場合 (ダーティミス) は、まず、ライトバック動作を行う。このライトバック動作では、最初にバンクアクティブ命令 (Ac) とロウアドレス (R1) によってメモリアレイを活性化させる (①)。次にライトバック命令 (Wb) とキャッシュブロックアドレス (Cb15)、アレイブロックアドレス (Ab3)、ロウアドレス (R1) によって、キャッシュメモリからメモリアレイへデータを書き戻す (②)。そして、プリチャージ命令 (Pre) によってセンスアンプに保持しているデータを消去する (③)。

ライトバック動作が終了した後に、CPUが要求したデータの読み出しを行う。この読み出し動作では、まず、メモリアレイ内のデータを、バンクアクティブ命令 (Ac) とロウアドレス (R0) によって、センスアンプへ転送し、保持する (④)。次に、ロード命令 (Ld) とアレイブロックアドレス (Ab1)、キャッシュブロックアドレス (Cb15) によってセンスアンプの

データをキャッシュメモリへ転送する (⑤)。次に、リード命令 (Rd) とカラムアドレス (C0) によって、キャッシュメモリからデータを読み出す (⑥)。最後にプリチャージ命令 (Pre) によって、センスアンプに保持しているデータを消去する (⑦)。このようにダーティミスの場合は、データを読み出す前にライトバック動作が必要のため、レイテンシは18サイクルとなる。

以上説明したように、ヒット時はデータ転送の高速化が可能となるが、ダーティミス時は、レイテンシが大幅に増加してしまう。そこで、キャッシュ内蔵 SDRAM のレイテンシを短縮するために、ヒット率を向上させ、ダーティミス時のレイテンシを短縮できるメモリコントローラを提案した。

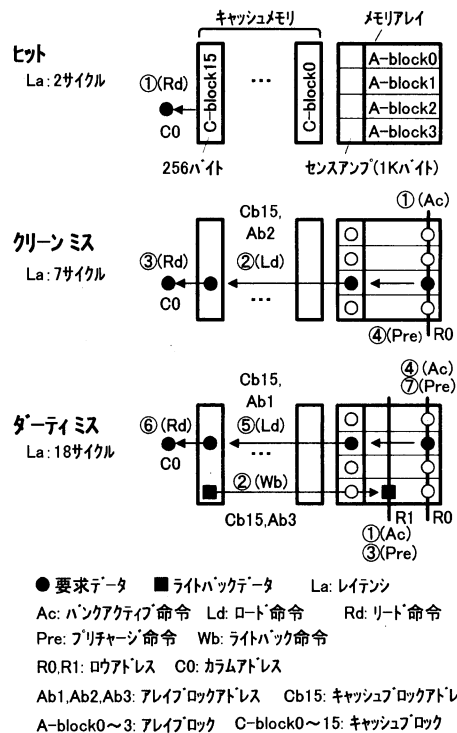
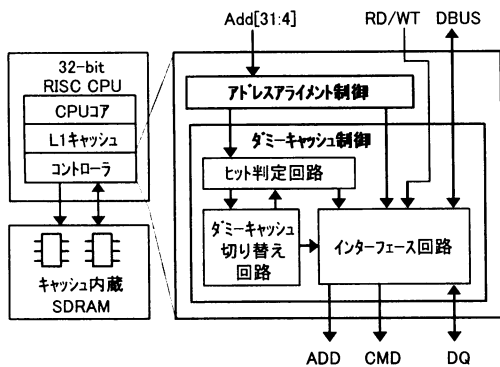


図1: キャッシュ内蔵SDRAMの読み出し動作

3. 新メモリコントローラの提案

図2に、我々が提案するキャッシュ内蔵 SDRAM を低レイテンシで動作させるメモリコントローラの構成図を示す。本メモリコントローラはCPUに内蔵され、アドレスアライメント制御ブロックと、ダミーキャッシュ制御ブロックから構成さ

れる。ダミーキャッシュ制御ブロックは、ヒット判定回路、ダミーキャッシュ切り替え回路およびインターフェース回路を含んでいる。



Add[31:4]: L1キャッシュアドレス ADD: DRAMアドレス
RD: リード命令 WT: ライト命令 CMD: コマンド
DBUS: データバス DQ: データバス

図2: 本コントローラの構成図

まず、アドレスアライメント制御ブロックについて説明する。アドレスアライメント制御ブロックは、キャッシュ内蔵SDRAMのヒット率を向上させるため、CPU内のL1キャッシュメモリのタグおよびインデックスアドレスと、キャッシュ内蔵SDRAMのバンク、ロウ、アレイブロックおよびカラムアドレスの対応付けを行う。さらに、本アドレスアライメント制御ブロックは、アレイブロックアドレスおよびバンクアドレスと、キャッシュブロックアドレスの対応付けも行う。本コントローラが設定できるアドレス対応付けは3種類(Align-A, Align-B, Align-C)あり、図3に示す。

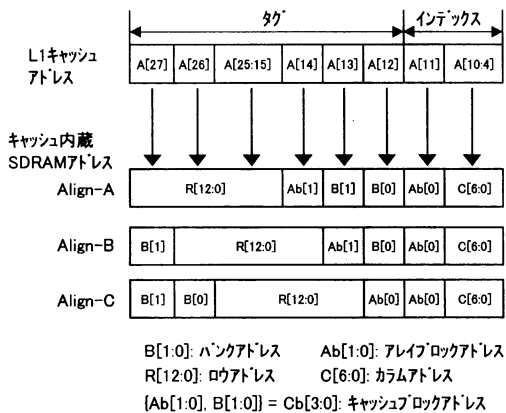


図3: アドレスアライメント制御

Align-Aでは、L1キャッシュメモリのインデックスアドレスA[11:4]へアレイブロックアドレスAb[0]およびカラムアドレスC[6:0]を割り当てている。タグアドレスA[13:12]へはバンクアドレスB[1:0]を、A[14]へはアレイブロックアドレスAb[1]を、A[27:15]へはロウアドレスR[12:0]を、割り当てている。

また、3種類のアドレス対応付けに共通して、アレイブロックアドレスAb[1:0]へは、キャッシュブロックアドレスCb[3:2]を、バンクアドレスB[1:0]へは、キャッシュブロックアドレスCb[1:0]を割り当てている。

次に、ダミーキャッシュ制御ブロックについて説明する。ダミーキャッシュ制御ブロックでは、16個のキャッシュブロックの内、15個のキャッシュブロックをヒット判定の対象とする通常キャッシュメモリとして機能させ、残りの1個のキャッシュブロックをヒット判定の対象とせず、バッファ(ダミーキャッシュメモリ)として機能させる。このダミーキャッシュメモリによりダーティミス時のレイテンシを大幅に短縮することができる。

ヒット判定回路では、キャッシュ内蔵SDRAMへ生じたアクセスが通常キャッシュメモリに対して、ヒットなのか、クリーンミスなのか、ダーティミスなのかを判定する。

ダミーキャッシュ切り替え回路はヒット判定回路の判定結果により、どのキャッシュブロックをダミーキャッシュメモリとして機能させるかを決定し、切り替え命令をヒット判定回路およびインターフェース回路へ伝える。ヒット判定回路では、この切り替え命令に従い、ダミーキャッシュメモリの切り替えを行う。インターフェース回路は、命令、アドレス、ヒット判定結果および切り替え命令に従い、キャッシュ内蔵SDRAMを制御する。

図4に、ダミーキャッシュ制御ブロックによる、ダーティミス時の詳細な読み出し動作を示す。

初期設定では、16個のキャッシュブロック(C-block0~C-block15)の内、キャッシュブロック0はダミーキャッシュメモリに、ブロック1からブロック15までは通常キャッシュメモリに設定される。

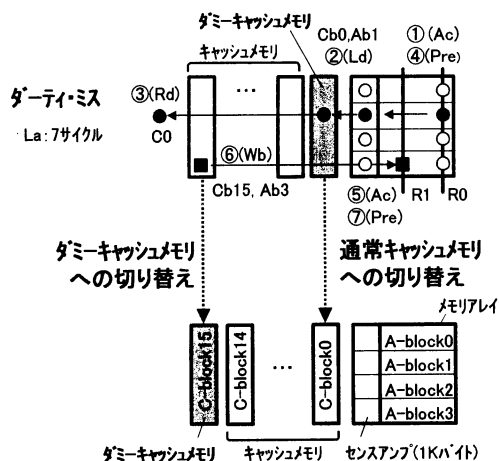
ダーティミスが生じた場合は、まず、CPUの要求データの読み出しを行う。この読み出し動作では、最初に、要求データをメモリアレイからセンスアンプを経由し、ダミーキャッシュへ転送する(①、②)。次に、ダミーキャッシュメモリから要求データを読み出す(③)。そして、プリチャージ命令によってセンスアンプ内のデータを消去する(④)。

要求データの読み出しが終了した後に、ライトバック動作を行う。このライトバック動作では、キャッシュブロック15に保持されているデータをメモリアレイへ書き戻す(⑤、⑥、⑦)。

最後に、ダミーキャッシュメモリであるキャッシュブロック0をヒット判定の対象とする通常キャッシュメモリへ切り替え、さらに、ライトバック動作が発生したキャッシュブ

ロック 15 をヒット判定の対象としないダミーキャッシュメモリへと切り替える。

以上説明したように、ダミーキャッシュメモリを設けることで、CPU の要求データの読み出し動作を最優先させ、ライトバック動作を後回しにできるため、レイテンシを 18 サイクルから 7 サイクルへ短縮することができる。



- 要求データ ■ ライトバックデータ La: レイテンシ
- Ac: バンクアクティブ命令 Ld: ロード命令 Rd: リード命令
- Pre: プリチャージ命令 Wb: ライトバック命令
- R0,R1: ロウアドレス C0: カラムアドレス
- Ab1,Ab2,Ab3: アレイブロックアドレス Cb0, Cb15: キャッシュブロックアドレス
- A-block0~3: アレイブロック C-block0~15: キャッシュブロック

図4: ダミーキャッシュ制御

4. 性能評価

本メモリコントローラによる平均レイテンシとプログラム実行時間の改善効果を明らかにするため CPU シミュレータ上でベンチマークプログラムを実行させた。表 1 にシミュレーションの前提としたシステム構成を示す。

4.1 ヒット率

図 5 にキャッシュ内蔵 SDRAM のヒット率を示す。アドレスアライメント制御によって、ヒット率が大きく変動することがわかる。アドレス対応付けを Align-A に設定した場合、全ベンチマークプログラムのヒット率が最も高くなることがわかった。Espresso, Linpack および Hydro2d を実行した際の、最高ヒット率は、それぞれ 92%、91%、57% となった。

表1: シミュレーション条件

CPUコア	32-bit RISC タイプ(1GHz)
L1キャッシュ	命令キャッシュ: 16Kバイト データキャッシュ: 16Kバイト ラインサイズ: 16バイト セットアソシアティブ: 4Way
キャッシュ内蔵 SDRAM	キャッシュメモリ: 16ブロック×256バイト 容量: 256Mビット×2チップ バンク数: 4バンク CASレイテンシ: 2、バースト長: 4 リフレッシュ: 64msec/8192サイクル 動作周波数: 133MHz バス幅: 32ビット
本メモリコントローラ	アドレスアライメント制御: 3種アドレス対応 ダミーキャッシュ制御 ダミーキャッシュメモリ: 17ブロック×256バイト 通常キャッシュメモリ: 15ブロック×256バイト セットアソシアティブ: ダイレクトマップ
SDRAM	容量: 256Mビット×2チップ バンク数: 4バンク CASレイテンシ: 2、バースト長: 4 リフレッシュ: 64msec/8192サイクル 動作周波数: 133MHz バス幅: 32ビット
従来SDRAM コントローラ	センスアンプキャッシュ制御 センスアンプキャッシュ容量: 4バンク×1Kバイト セットアソシアティブ: ダイレクトマップ
	非センスアンプキャッシュ制御
ベンチマークプログラム	Espresso: 論理圧縮 Linpack: 行列演算 Hydro2D: 流体力学

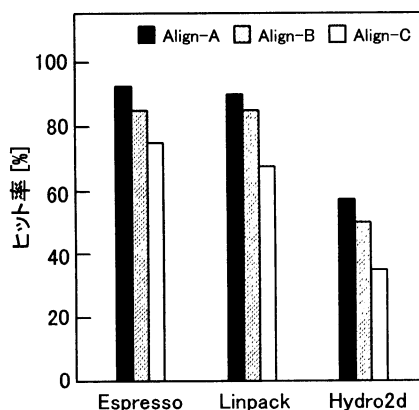


図5: ヒット率

4.2 レイテンシの評価

図6に、本メモリコントローラによる平均レイテンシの短縮効果を示す。

本メモリコントローラによって、従来のセンスアンプキャッシュ制御と比較し、レイテンシを、Espresso 実行時に 15%、Linpack 実行時に 16%、Hydro2d 実行時に 25%短縮できることがわかった。

4.3 プログラム実行時間の評価

図7に本メモリコントローラによるプログラム実行時間の短縮効果を示す。

本メモリコントローラによって、従来のセンスアンプキャッシュ制御と比較し、プログラム実行時間を、Espresso 実行時に 0.1%、Linpack 実行時に 7%、Hydro2d 実行時に 13%短縮できることがわかった。

5. 本メモリコントローラの回路規模

0.18- μm CMOS テクノロジ想定し、本メモリコントローラを実現するための回路規模を評価した。表2にゲート規模を示す。電源電圧 1.8V 時に、動作周波数 133MHz を実現する本コントローラのゲート規模は 9.2k ゲートとなった。

6. まとめ

キャッシュ内蔵 SDRAM のレイテンシとプログラム実行時間を短縮できるメモリコントローラを提案した。

ベンチマークプログラムを用いて性能評価を行った結果、従来のセンスアンプキャッシュ制御方式と比較し、レイテンシを最大 25%短縮でき、プログラム実行時間を最大 13%短縮できることを確認した。また、本メモリコントローラの回路規模は 9.2k ゲートとなった。

参考文献

- [1] T.Watanabe, et al., "AccessOptimizer to Overcome the Future Walls of Embedded DRAMs in the Era of Systems on Silicon," 1999 ISSCC Digest of Technical Papers, pp370-371.
- [2] Y.Kim, et al., "A Memory Access System for Merged Memory with Logic LSIs," 1999 AP-ASIC, pp384-387.
- [3] Y.Matsui, et al., "64Mbit Virtual Channel Synchronous DRAM," NEC Research & Development Report, July 1999, pp282-286.
- [4] S.Miura, et al., "A Memory Controller that Reduces Latency of Cached SDRAM," March 2005 ISCAS, pp5250-5253.

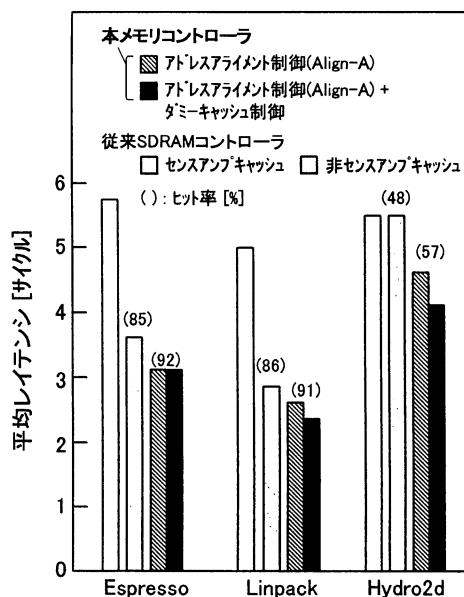


図6: 平均レイテンシ

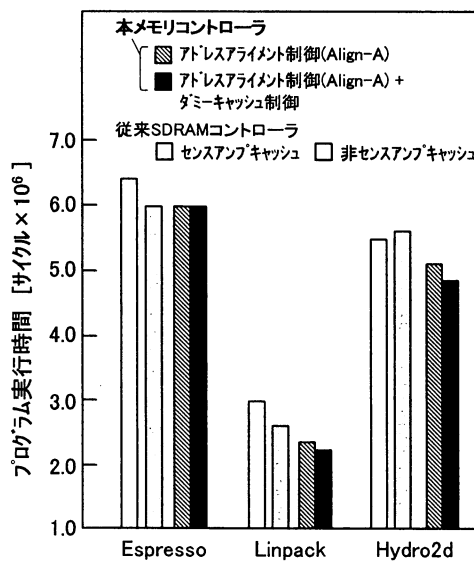


図7: プログラム実行時間

表2: 本メモリコントローラの回路規模

アドレスアライメント制御ブロック	0.5k ゲート
ダミークャッシュ制御ブロック	
ヒット判定回路	3.8k ゲート
ダミークャッシュ切り替え回路	0.6k ゲート
インターフェース回路	4.3k ゲート
合計	9.2k ゲート