

## 高速アクセスメモリ QRAM

大庭信之 新島秀人

日本アイ・ピー・エム 東京基礎研究所

マイクロプロセッサとメモリの速度ギャップを短縮するため、チップ内にアドレス比較器と、マイクロプロセッサと直接ハンドシェイクするメカニズムを搭載した新しいメモリ、QRAMを提案する。本稿では、まずQRAMの内部構造上の特長であるマルチアクティブアイランドと、QRAMを用いたシステムの構成法について述べる。次に、トレースデータを用いたシミュレーションによるQRAMの性能評価を、特にキャッシュとの比較を中心に報告する。コストと性能の点から、QRAMは比較的ローエンドのワークステーションに応用できると思われる。

### Quick Access Memory QRAM

Nobuyuki Ooba Hideto Niijima

IBM Reserch, Tokyo Research Laboratory, IBM Japan, Ltd.

This report discusses a new memory architecture QRAM, that is designed to bridge the speed gap between the microprocessor and the main memory with high cost-performance. For high performance and useability, QRAM has two features, a built-in handshake facility and a multiple active island structure. We describe the hardware configuration of the QRAM and its operation. We compared the QRAM's performance with standard caches' by trace-driven simulations. According to its cost and performance, it will be applicable to low-end workstations.

## 1. 初めに

近年、マイクロプロセッサの速度向上に伴い、メモリーシステムとの速度差が顕在化し、システム全体のパフォーマンスに大きな影響を及ぼすようになってきた。この速度差を回避する一つの方策として、キャッシュシステムが提案され、事実、現在の多くのシステムがこのキャッシュを搭載している。また、最新のマイクロプロセッサに於ては内部キャッシュを搭載するものも現れてきた。しかしながら外部キャッシュシステムは高速なSRAMとキャッシュコントローラとを必要とするため、システムコストの上昇は避けられず、また、内部キャッシュシステムでは現在のところ十分な容量を持つ事ができない。速度差を回避する他の方策としては高速なDRAM、もしくはページ、ニブルモードのような特殊なアクセスモードを用いてDRAMとのデータ転送を行うなど、主記憶装置自体の高速化を図る方法が考えられる。高速なDRAMを用いる事は本質的な解決策ではあるが、アクセスが50ナノを切る超高速DRAMは現在のところ非常に高価であり、コストとの兼合が問題となる。また、ページモード、ニブルモードなどを有効に使うためにはメモリーとデータの要求元(例えばCPU、DMAコントローラなど)の間にいわゆるブロック転送を可能にするための付加的なインターフェース回路を必要とする。即ち、フレームバッファなどのように、アドレスが1ずつ増加(もしくは減少)しながら繰り返し読み書きするような、メモリーアクセスが常に規則正しいパターンで繰り返される環境では有効な手段であるが、一般の主記憶・キャッシュのように命令とデータが混在し、メモリーに対するアクセスが規則正しくない場合にはこれらのモードを利用することは困難である。

このような問題を解決するため、一部のシステムでは疑似キャッシュシステム(Pseudo Cache System 図1)と呼ばれる方法を採用している。この方法は、アドレスラッチと比較器とから成り、アドレスラッチには前回アクセスされたアドレス情報が保持されている。ここで、次回アクセスされるデータのアドレスはアドレスラッチの内容と比較され、同じROWアドレスを持っている場合にはページアクセス、異なる場合には通常のアクセスを行う。しかしながら、この場合にはキャッシュシステムで言うエントリー数が高々であるため高いヒット率は望めない。これを拡張してアドレスラッチと比較器とを各々複数持つ方法も考えられるが、記憶容量の拡張に制限が出るなどの問題がある。

## 2. QRAM

QRAMは、従来のDRAMに、最小の外部付加回路で、高速マイクロプロセッサの主記憶として高い性能を発揮できるように開発されたメモリーである。即ち、チップ内部に、通常メモリーのアドレス回路にアドレスラッチとアドレス比較器を、さらにデータ授受のハンドシェイク用のMISS信号発生回路を付加する。また、高速なアクセスを実現するために、すべてのアドレスを一度に入力する必要があるが(アドレスの時分割をしない)、メモリー周辺回路が簡単になる。QRAMの特長を以下に示す。

### (1) 従来のメモリー製造技術で高速アクセス時間の実現

メモリーへのアクセスの局所性を利用し、同一ROWにアクセスが繰り返される場合にはページアクセスタイムで読み書きが可能。

### (2) MISS信号発生回路を内蔵

データ読みだし・書き込み終了のMISS信号をメモリー自身で発生するため、従来のメモリーシステムで必要であったREADY信号を発生させるための外付け回路が簡単になる。さらに、システムがメモリーのサイクルタイムを考慮する必要がなくなる。また、将来的にフル・セルフ・オートリフレッシュを容易に実現する手段を提供する。

### (3) 汎用のCPUに使用可能

CPU側からは、直前に行なわれたアクセスと現在のアクセスが同一ROWにあるかどうかをまったく意識する必要がないため、従来のCPUにそのまま使用できる。さらに、キャッシュシステムで問題となるデータコヒーレンシーについて

も、システムは関知しなくて良い。

#### (4) 高いヒット率を実現

RAMの内部構造を工夫する事により、実用上充分なヒット率を実現した。

### 3. 動作説明

メモリー外から読みだしアクセスのため、アドレスが与えられたとする。Q RAMはROWアドレスを直前のアクセスのROWアドレスと比較し、もし同じならばそのROWのデータは既にセンスアンプまたはラッチ回路に取り込まれているので、COLUMNアドレスデコーダで必要なデータを選択し、外部に出力する。逆に、ROWアドレスが直前のアクセスのROWアドレスと違う場合は、先ずMISS信号を出力し、外部に対しデータの遅れを知らせる。その後、通常のRAMのアクセスサイクルと同様にROWアドレスデコーダを通じて該当するROWをアクティブにして、さらにCOLUMNアドレスデコーダで要求されたアドレスのデータを返す。もし、プリチャージを必要とするメモリーではプリチャージが終了していなければプリチャージ終了を待って次のアクセスサイクルが始まる。ROWアドレスはラッチに格納され、さらに後から続くアクセスのアドレス比較に用いられる。書き込みアクセスの場合も同様に行なわれる。

図2にタイミングチャートを示す。ここでは、リードミス、リードヒット、ライトヒット、ライトミスとアクセスが続いた場合を示している。Q RAMは、CSが下がった時にアクティブとなり、アドレスを取り込む。リードミスが起こったサイクル1では、CSが下がってしばらくの後MISS信号が出力され、データは、このMISS信号が戻るのを待った後、出力される。サイクル2では、MISS信号が出力されず、直ぐにデータが出力される。データ書き込みの場合は、サイクル3、サイクル4に示すように書き込むべきデータをCSの下がるタイミングで与える。ミスが起こった場合は、サイクル4の様にCSを一定時間引き伸ばさなければならない。

### 4. Q RAMの内部構造

一個のQ RAMモジュールは複数のアイランドから構成され、それらのアイランドは同時にかつ、独立に動作することができる。それぞれのアイランドは、通常のDRAMの部品に加え、ROWアドレスのラッチと比較器、及びマイクロプロセッサとハンドシェイクを行うためのMISS発生器を装備する。

また、Q RAMはそのデータバッファリングの方法により、大きく二つに分けられ、センスアンプをバッファとして利用するシングルバッファタイプ(図3)とダブルバッファタイプ(図4)とがある。シングルバッファタイプの場合には、一つのセルに対してただ一つのバッファを持つ単純な構造を持ち、従来のDRAMをこのタイプのQ RAMに改造する事は比較的容易である。ダブルバッファタイプは、一つのセルに対して二つのバッファを持つため物理的構造はやや複雑であり、かつMISS時の対応に工夫を要するなど考慮すべき事項も多いが、シングルバッファタイプと比較してヒット率が高く、かつ安定している。

### 5. Q RAMの性能評価

Q RAMは、キャッシュシステムと同様に、マイクロプロセッサのメモリアクセスの局所性を利用してシステムの性能を向上させるという点から、RAMの中にキャッシュを組み込んだとも考えられる。即ち、Q RAM内のバッファをキャッシュデータメモリと対応付けると、非常にラインサイズの長いキャッシュであると考えられる。Q RAMを用いた主記憶の標準構成では、そのラインサイズは512~2Kバイトに達する。一般のキャッシュシステムではこのように長いラインサイズは採用しない。なぜなら、ヒット率があまり向上しない(むしろ下がることもある)上に、ミスしたときのメモリからキャッシュへの転送コストが大きくなるためである。そこで本節では、キャッシュとの比較をしながらQ RAMの性能を評価する。

評価基準として、ヒット率と実際のメモリアクセス性能を示す平均メモリアクセスサイクルを

用いた。キャッシュシステムの平均メモリアクセスサイクルは、キャッシュとメモリ間のデータ幅を16ビットと仮定し、次の式で表せるものとした。Hはヒット率、Lはラインサイズである。

$$\text{平均メモリアクセスサイクル} = H + (1-H) * (4+L/2)$$

QRAMを用いたメモリシステム、及びキャッシュを用いたメモリシステムの性能を、メモリアクセストレースをを用いたシミュレーションで評価した。ここで用いたメモリアクセストレースは、IBM PC/AT上で動作するスプレッドシートプログラムから得たものであり、合計数十万ステップから成る。

まず、1Mビット、入出力ピン4本、アイランド数4、ページ長512ビットのシングルバッファタイプQRAMを用いた、64Kバイトのメモリシステムの性能を評価した。この構成は、QRAM内のバッファ容量から考えて、キャッシュサイズ8Kバイト、ラインサイズ1Kバイトのキャッシュシステムにちょうど対応する。トレースデータを用いたシミュレーション結果を図5と図6に示す。キャッシュは2ウェイセットアソシアティブ構成である。図5から分かるように、QRAMのヒット率はキャッシュに比べ低い。これはキャッシュが2ウェイセット構成を採るためであり、キャッシュサイズが8Kバイト程度であればキャッシュのウェイ数はヒット率にかなり影響する。図6は平均メモリアクセスサイクルを評価したものである。QRAMは1ページ(ライン)をチップ内部で1サイクルでバッファにロードできるため、ヒット率が低くても、平均メモリアクセスサイクルは1.3程度である。

次に、ダブルバッファタイプのQRAMを評価した。バッファの容量が倍になるので、キャッシュとの比較で考えると、キャッシュサイズ16Kバイト、ラインサイズ1Kバイトのキャッシュに相当する。ここでキャッシュは4ウェイセットアソシアティブとした。図7にヒット率、図8に平均メモリアクセスサイクルを示す。ヒット率ではキャッシュに及ばないが、平均メモリアクセスサイクルではキャッシュより良い結果が得られている。

## 7. むすび

メモリの内部構造を十分活用して高速化を狙ったQRAMを提案、説明した。本稿ではDRAMを前提としていたが、QRAMの方式はDRAM以外にも、ROMやSRAMにも応用できよう。シングルバッファタイプのQRAMは、従来のDRAMとほとんど同じ構成で実現できるので、従来のものに比べ、面積で1割程度の拡大で済むと思われる。また、ダブルバッファタイプのQRAMは、バッファの分だけ大きくなるが、やはり2割程度の拡大で済むと見積もっている。コストに関しても同様であるが、今後、検討しなければならない。

ハードウェアコストの点から考えると、キャッシュシステムは高価なキャッシュコントローラや高速SRAMを必要とするが、QRAMは前述のような簡単なハードウェアで実現することができ、有利であると思われる。従って、比較的ローエンドのパーソナルコンピュータやワークステーションでは、キャッシュに代わって応用できるのではないかと考えている。コストよりも性能を重視する、ハイエンドマシンでは大容量、QRAMより高性能のキャッシュシステムを搭載したほうがよいであろう。

### 【参考文献】

- [1]新島、大庭：“高速アクセスメモリ QRAM,”平成元年、秋の情報処理学会全国大会。
- [2]Peelen and Goor: "Using the page mode of dynamic RAMs to obtain pseudo cache," Microprocessors and Microsystem, Vol.11, No.9, 1987.

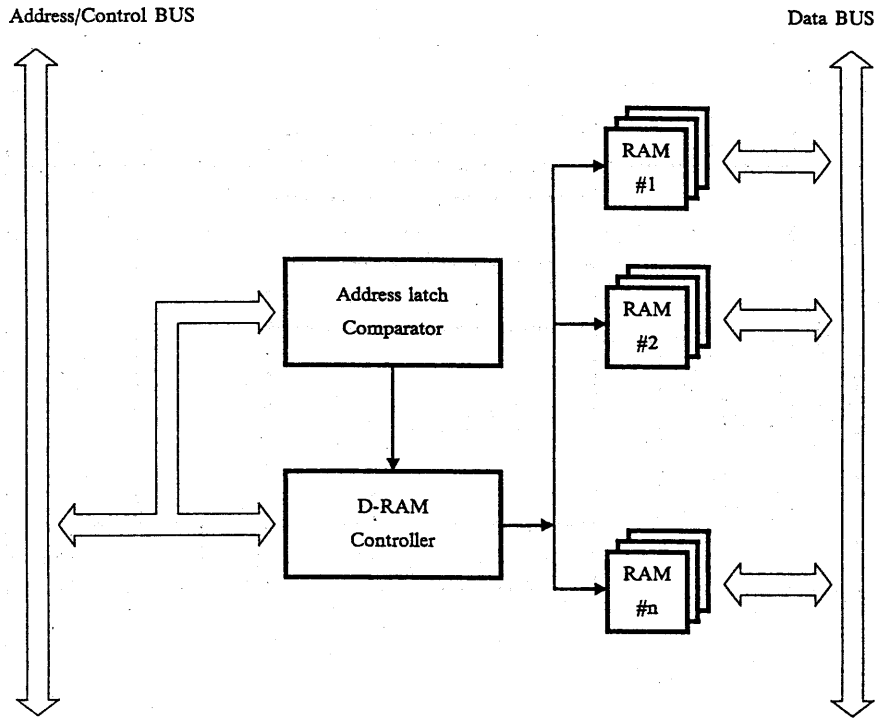


図1. 疑似キャッシュシステム

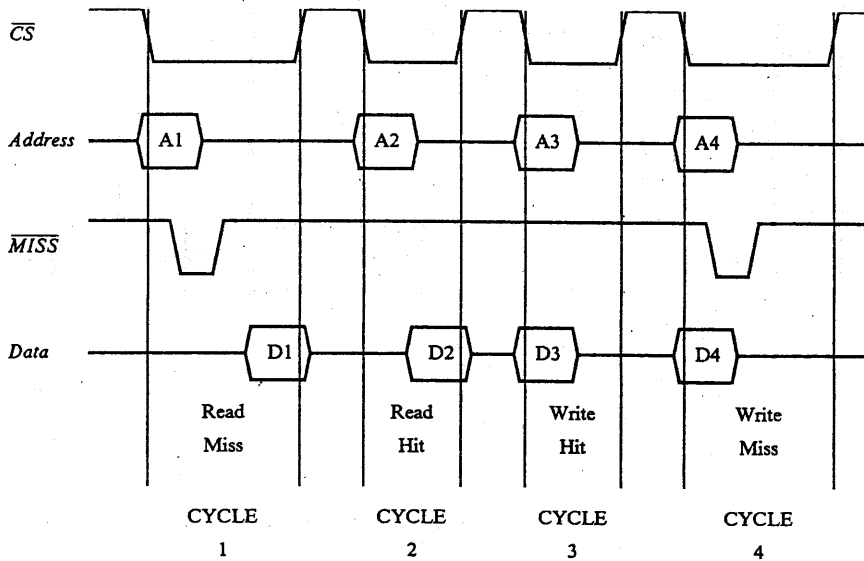


図2. タイミングチャート

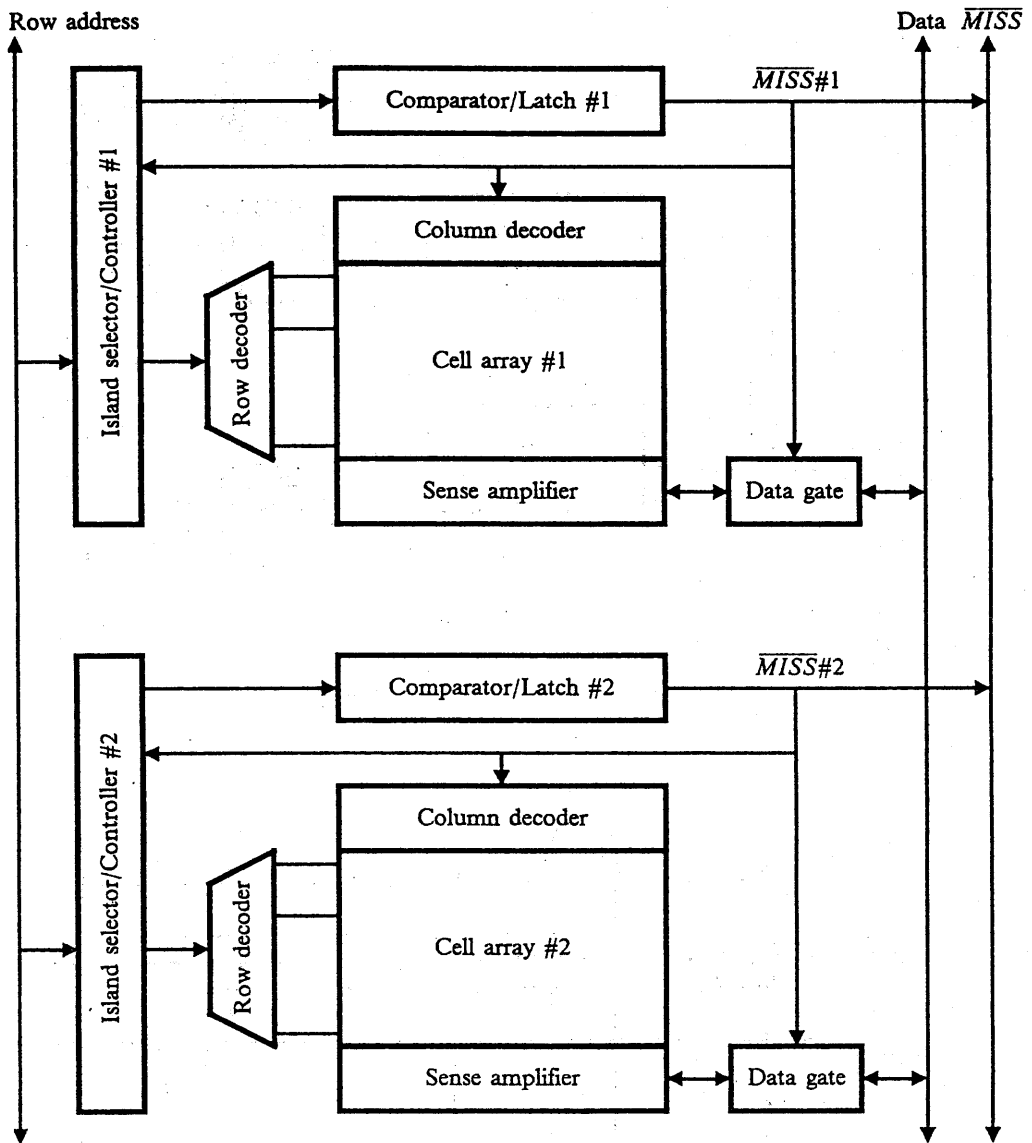


図3. シングルバッファタイプ

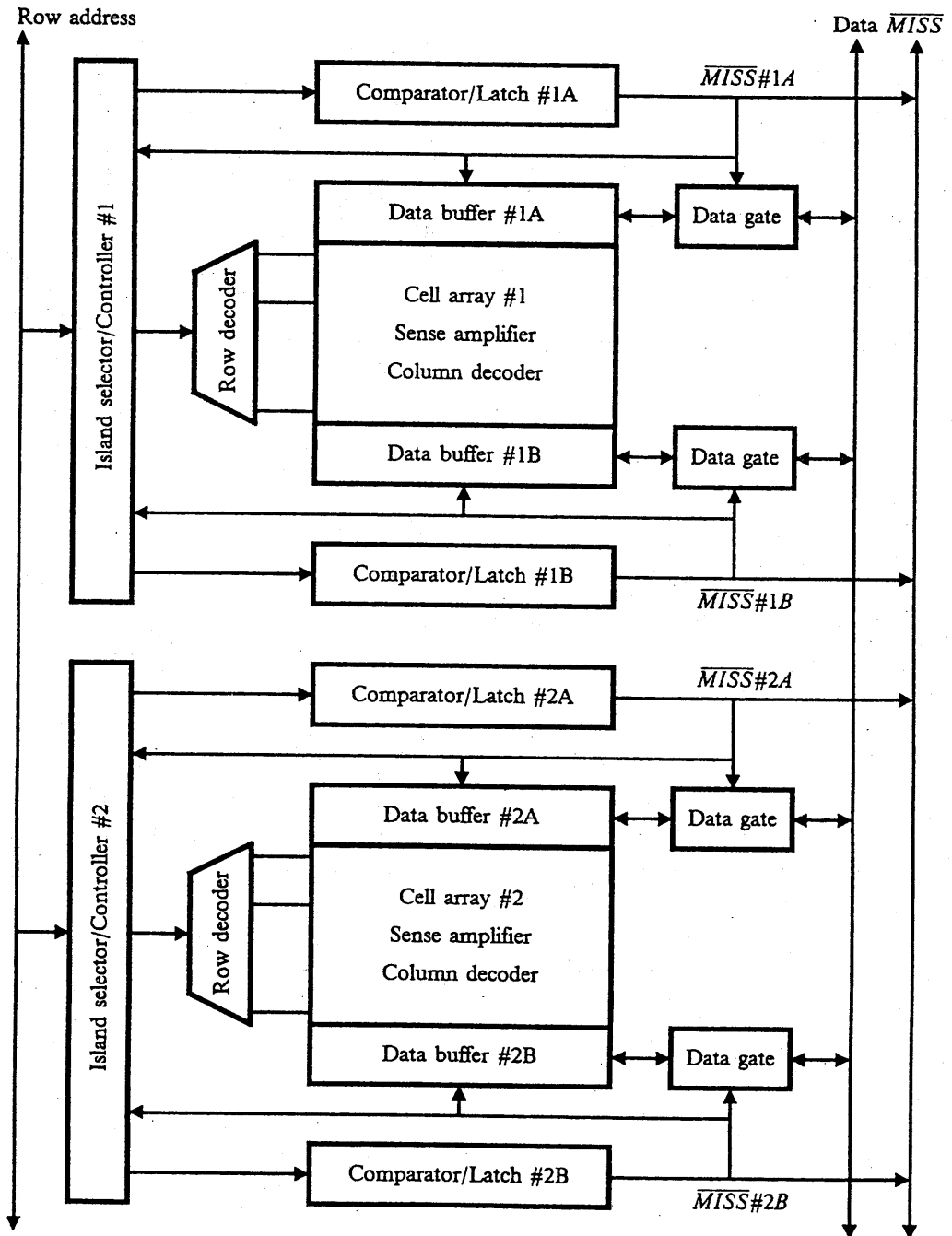


図4. ダブルバッファタイプ

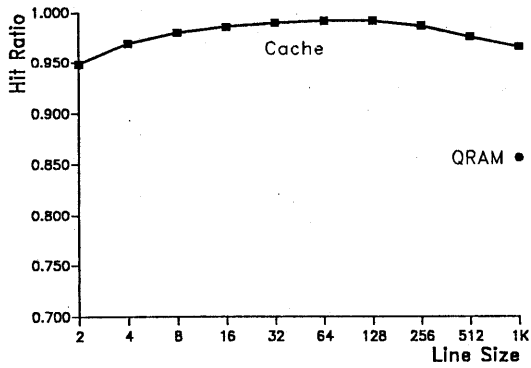


図5. キャッシュヒット率  
(シングルバッファタイプ)

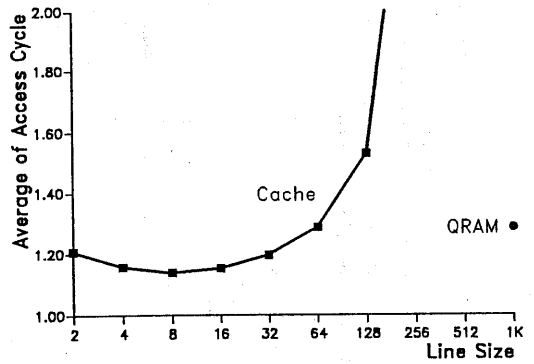


図6. 平均メモリアクセスサイクル  
(シングルバッファタイプ)

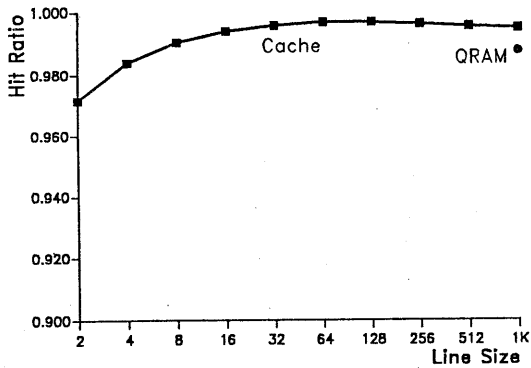


図7. キャッシュヒット率  
(ダブルバッファタイプ)

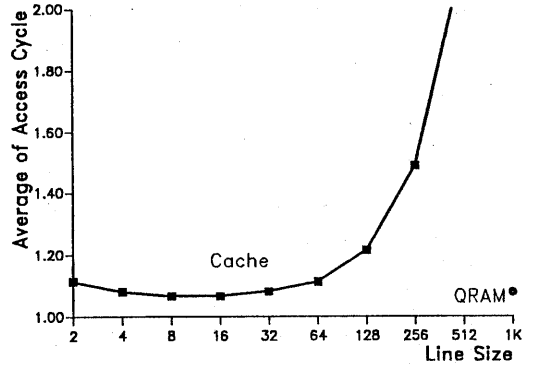


図8. 平均メモリアクセスサイクル  
(ダブルバッファタイプ)