

特別論説 情報処理最前線

最先端ダイナミックメモリ

Recent Topics on Dynamic Random Access Memories by Yasunao KATAYAMA (IBM Research, Tokyo Research Laboratory).

片山 泰尚¹¹ 日本アイ・ビー・エム(株)東京基礎研究所

1. はじめに

半導体技術の絶え間ない進歩は、RISC、パイプラインなどの設計技術の進歩と相俟って、CPUの性能を指数関数的に向上させてきた。CPUの動作周波数は500MHzを超えるものが発表され、研究開発では1GHzを目指している。しかしながら、DRAMのメモリアクセスは、高ファンアウトのワードライン・ビットラインの充放電にかかる時定数とパッシブなメモリセルに対する信号の読み書きにかかる時間により制限されるため高速化が容易ではない。このため、コンピュータシステムの設計において、メモリシステムの選択がシステム全体の性能を決める上できわめて重要になってきている。

このような状況の中で、図-1に示されるように、従来は汎用品によるコストダウンが最も重要といわれていたDRAMチップにおいても、さまざまな形での、高速・高機能な製品群が現れた。ここでは、これらの現状を踏まえた上で、今後の展開を模索したい。

2. 高速DRAMの現状

高速DRAMの研究は筆者らも参加したCMOSベースのIBMのグループで¹⁾、続いてBiCMOSを用いた日立のグループで²⁾80年代後半に始まった。これらのチップは当時としては画期的に高速な、20ns台前半のランダムアクセスと、10ns台前半のコラムアクセスを実現した。前者は、主にメモリアレイのアーキテクチャならびにセンス回路・データバス回路の高速化により、後者は、主にゲインの大きいバイポーラトランジスタをうまくCMOSトランジスタと組み合わせることにより高速化を実現した。しかしながら、当時は、CPUのスピードは、まだ、これらのDRAMを必要とするほど高速ではなく、また、プログラムならびにデータアクセスのローカリティも大きかったため、比較的小容量のキャッシュメモリを付加することによりその当時のシステムに対しては十分なメモリバンド幅が得られたことから、これらの高速DRAMは商業的な成功には至らなかった。

しかしながら、これらの高速DRAMの先駆的研究は、SDRAM(Synchronous DRAM)の開発により、形を変えて製品化されることになる。SDRAMにおいては、これまでのDRAMの非同期式インタフェースに代わって、外部クロックに同期したインタフェースを採用し、コラムアクセスでのデータレートをパイプライン化、プリフェッチなどの技術を用いて向上させることに重点がおかれた。ランダムアクセスの向上は、内部のメモリアレイを2バンク構成にしインタリーブアクセスを行うことにより実現された。200MHzを超える速度で動作するものも開発され³⁾、次世代の高速メモリの最右翼にある。しかしながら、モードレジスタによるアクセスモードの変更など、慣

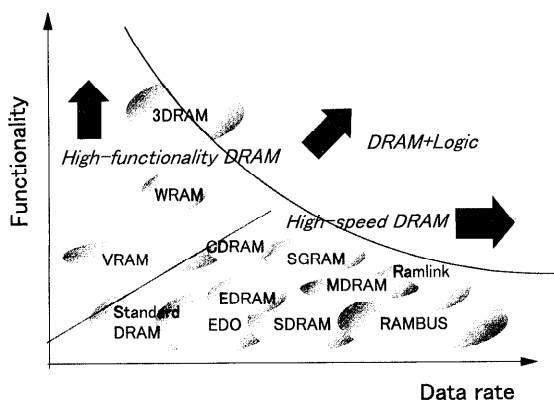


図-1 メモリ技術の現状と方向性

れないシステム設計者の間には戸惑いも多く、商業的成功は遅れて出てきた EDO (Extended Data Out) DRAM に先を越されることになる。EDO-DRAM は、同期構造をとらずにコラムアクセスに簡単なパイプラインを付加することによって汎用 DRAM に最小限の変更で、コラム転送速度を向上させたもので、メモリコントローラチップセットにほとんど変更を加える必要がなかったことからいち早く広まり、こちらは、現在メインメモリのマーケットの主要な一角を構成する。

また、DRAM チップ内にキャッシュメモリを付加することにより、DRAM の性能を SRAM に近づけようとすることも考えられた。これらの代表例としては、Distributed cache をもつ EDRAM (Extended DRAM)⁴⁾、Lumped cache をもつ、CDRAM (Cache DRAM)⁵⁾ などが共に製品化されている。これらのメモリは、キャッシュアドレスタグと比較回路を内蔵し、内部のキャッシュにヒットするかしないかによりデータが出てくるまでのアクセス時間が異なる。そのほか、DRAM 自身をキャッシュメモリとして使うことも提案され製品化されている。

さらに、新たなパケット方式のメモリインタフェースの採用により、1 チップで 500MB/s というデータ転送レートを実現する RDRAM (RAMBUS DRAM)⁶⁾ も、比較的早くから研究開発が進められており、最近ではゲーム機、グラフィックなどのメモリとして、着実に市場を固めている。クロックとしては 250MHz を使い、各クロックの立ち上がり、立ち下がり両方でデータを転送する。DRAM アレイ上のデータは、いったんラインバッファに転送され、そこからパラレル・シリアル変換により、高速同期転送される。オフチップのメモリバスは、反射がなく、スキューも最小になるように専用のツールを用いて設計される。さらなるデータ転送速度の向上 (600MB/s) とページミス時のレイテンシの改良を実現したコンカレント RAMBUS もある。高いバンド幅が要求される状況下 (1GB/s 程度以上) では、SDRAM より、コストや実装スペースの点で有利であるとの見方もあり、将来はデスクトップ PC のメインメモリの座を狙っている。

そのほか、内部メモリを多数のバンクに分け、それぞれのバンクでセンスアンプにデータをラッ

チした状態に保つことにより、実効的なアクセス時間をコラムアクセス時間まで近づける MDRAM (Multibank DRAM)、また、新たな高速標準インタフェースを目指す Ralink、またそのサブセットである SyncLink なども提案され、MDRAM はすでにグラフィックならびにキャッシュメモリとして製品化されている。

3. 高機能 DRAM の現状

高機能 DRAM として、最初に商業的成功を収めたのは、VRAM (VideoRAM) であろう。これは、1980 年中頃考案され、DRAM のアレイ内部に潜在的に存在する超並列データ構造による高データバンド幅を、ビデオメモリのスクリーンリフレッシュ用に効率よく利用する構造である。センスアンプに貯えられたデータは、その脇にあるシリアルレジスタにいったん貯えられた後に、通常のメモリアクセスと違うパスを通して (すなわち、通常のメモリアクセスを邪魔することなく) 出力される。

この後、高機能 DRAM の重要性は、認識されていたものの、商業的成功は、VRAM 以外は最近まで起こらなかった。しかし、スクリーン解像度の向上と、GUI (Graphic User Interface) ・ 3 次元グラフィックスなどのアプリケーションからの要求は VRAM に代わる新たなビデオメモリを出現させた。WRAM (WindowRAM)⁷⁾ は、Window GUI の高速化をねらって、アラインドビットブリットと、ビット演算機能、シリアルポートなどを集積したものである。また 3DRAM⁸⁾ は、3 次元グラフィックスのイメージレーヤコンポジションに的を絞って、Z-compare、アルファブレンディングの機能をフレームバッファに集積したものである。前者は主に PC 用のグラフィックボードに、後者は 3D グラフィックワークステーション用のメモリとして製品化されている。さらに、SGRAM はインダストリスタンダードなグラフィック専用メモリで、高速な同期式シングルポート I/O をもち、また、ブロックライト、マスクライトなどの最小限のグラフィック論理演算回路をもつ。

4. DRAM とロジックの混載

これらの高速・高機能 DRAM は、コンピュータの性能向上にともなう演算処理機構とメモリの間のフォンノイマンボトルネックを解消する手段として、研究開発が進められてきたものであるが、これらの技術は、さらなる集積度の向上、高バンド幅を実現するために、DRAM とロジックの混載技術という新たな展開をみせている。メモリと論理回路を混載することは、コンピュータアーキテクチャの観点からは、Stone らの研究⁹⁾にさかのぼるが、現実的な DRAM と論理回路の混載の研究の歴史は比較的新しい¹⁰⁾。DRAM とロジックの混載技術は、ある意味では、上に述べた、高機能 DRAM ですでに実用化されているものであるが(極論すれば、汎用 DRAM もアドレスデコーダなどの論理回路を含むわけでこの点から DRAM ロジック混載は、汎用 DRAM にすでに存在しているともいえる。)、最近は、この技術の一番の利点である、1GB/s を軽く超える一昔前なら夢のようなメモリバンド幅を利用した画像・グラフィック用途チップの研究開発が盛んである^{11)~13)}。さらに、CPU とメモリのワンチップ化¹⁴⁾、通信用途への応用¹⁵⁾も発表されている。

また、この技術を使って、新たなコンピュータシステムを構築しようとの試みも始まっている。基本的なアイデアの多くはメモリ技術をベースに演算回路とメモリを1つのユニットにして、それを、同一チップ上、または、別のチップとの間で、ネットワーク上に並列につなげたものである。通信プロトコルとメモリコピーレンシの取り方などにそれぞれの特徴がある。たとえば、EXECUBE^{16), 17)}は、4M ビットの DRAM テクノロジをベースに、SIMD(Single Instruction stream, Multiple Data stream)もしくは、MIMD(Multiple Instruction stream, Multiple Data stream)でコントロールされる複数のプロセッサを集積したもので、各論理演算回路がキュービクネットワークで接続されることからこの名前がついた。主に、パターン認識に用いられる。PPRAM(Parallel Processing RAM)¹⁸⁾は、256M ビット DRAM 程度の集積度を仮定して、複数個の32 ビット CPU と SRAM・DRAM を集積したアーキテクチャである。IRAM(Intelligent

DRAM)¹⁹⁾は、CPU と DRAM を1G ビット DRAM 程度の集積度を基に1チップに混載したものであり、行列計算のように、メモリアクセスに規則性がある場合にとくに高いバンド幅を実現する。

5. 今後の動向

これらの技術は近い将来どのような方向に向かうのであろうか? ここに僭越ながら我々の経験を紹介させていただきたい。我々が過去において、高速 DRAM の研究を始めたときにこの技術は高性能だが高価格という性質上、まずハイエンドで使われるであろう、そして、その後ローエンドに展開されていくであろうと予測した。しかしながら、現在、高速 DRAM の市場はローエンド PC にまで広がっており、逆に、ワークステーションなどでは、汎用 DRAM を用いながら、メモリバスの幅を256 ビット程度に広げ、さらに大容量の SRAM キャッシュメモリを搭載することによりバンド幅を確保しているものもある。これはどうしてであろうか? この答えは、グラニュラリティにある。たとえば、現在の PC の標準的なメモリ増設単位である、16MB のメモリを256 ビットバスで実現することを考えてみる。現在、主要な DRAM 容量は16Mb であるので、8個の16Mb DRAM により構成される。しかしながら、1つのチップより32 ビットの I/O をとる必要があり、これらの、多ピンの DRAM チップは、チップサイズ、パッケージサイズの増大により、コスト高になってしまう。すなわち、メモリグラニュラリティが小さいときは、高速 DRAM を使い、その分 I/O ピンの数を減らした方が、コストパフォーマンスがよくなる。この状況は、今後のメモリシステムの動向を考える上で、データバンド幅とともに、グラニュラリティが重要な指標になることを示している。一方、メモリ総量が大きく、グラニュラリティが大きくとれるときは、得てしてバスに対して多くメモリチップがぶら下がることになり、バス的高速化は難しくなる。

さらにグラニュラリティの重要性は、半導体の技術の進歩とともに、ますます顕著になる。これは、半導体の世界ではよく知られている Moore's law(1チップあたりのトランジスタの集積度は3年で4倍になる)と各世代ごとの単位面積あた

りの製造コストの上昇を考えると最も重要な結論である。簡単のために、もしユーザがメモリシステムに払う金額が一定で、各世代ごとの単位面積あたりの製造コストが2倍(ただし、ビットコストは半分)とすると²⁰⁾、メモリシステムを構成するチップの数は各世代ごとに2分の1になる。これは、グラニュラリティが今後ますます重要になることを示す。確かに、シリコンサイクルの変動はこの規則にパータベーションを与えるが、いかなるシステムの要求も経済原理は超えることができないという観点から、この帰結は正しいと思われる。

これらのことを踏まえて、筆者の考える将来のメモリシステムのフェイズダイアグラムは図-2のようである。汎用 DRAM、高速 DRAM、DRAM ロジック混載技術の間の技術の遷移は、ある、一定のメモリバンド幅、グラニュラリティの中でコスト、パワー、フォームファクタなどを考慮して決定される。すなわち、データバンド幅とグラニュラリティのトレードオフが汎用 DRAM1 チップあたりの I/O ピンの数を増加させると、高速 DRAM にとって代わられる、さらに高速 DRAM でも満たされないとき、DRAM ロジック混載技術が必要となる。DRAM ロジック混載技術は、最終的にはグラニュラリティの要求の厳しい組み込み用途向けのようなローエンドから、データ転送速度の要求の厳しい3次元グラフィックスのようなハイエンドの技術まで幅広い分野で使われると思われる。ただし、デスクトップ PC やワークステーションでのメインメモリは、拡張性、汎用性の観点から、主に高速 DRAM が使われることになろう。モバイルコンピュータの

ような特定のアプリケーションでは、DRAM ロジック混載によるスピード・パワーのトレードオフの改善やフォームファクタでの利点なども見逃せないため、混載技術がより多く使われるであろう。

では、近い将来の DRAM ロジック混載技術の商業的な成功はどのようなシナリオにより始まるであろうか？ この結論は、おそらく、アーキテクチャのみの観点からは決定されず、半導体のテクノロジーの問題が大きくかかわってくる。現在、最先端の半導体テクノロジーは、高速のデバイスと多層で配線ピッチの細かい金属配線を用いたロジック用のテクノロジーと、リーク電流を押さえた高耐圧のデバイスと2層程度の比較的配線ピッチの緩い金属配線からなる DRAM 用のテクノロジーに分化している。さらに、DRAM 用のテクノロジーは、メモリセルとして、シリコン基盤の上(スタック)もしくは内部(トレンチ)にキャパシタを形成する必要がある、DRAM ロジック混載技術を低コストで実現するには超えるべき課題は多い。現在、DRAM ロジック混載技術の方向性は、DRAM ベースの技術を基にデバイス的高速化、金属配線の改良を目指す動きと、ロジックベースの技術を基に DRAM のメモリアレイを作り込む動きがある²¹⁾。前者は、テクノロジー開発は比較的容易であるが、ロジック用のライブラリを新たに DRAM ベースの技術に変換する必要がある。後者は、ロジック用のライブラリをそのままに近い形で(タイミングルールなども含めて)使用できるが、製造技術開発に多額の資金がかかり、さらに製造工程数の増加により製造コストが上昇する。また、このほかに、ロジック用のトランジスタを3個用いてメモリセルを構成することも行われており、メモリ容量の点で劣るものの、ロジック用のテクノロジーがほぼそのまま使用できる利点がある。

この中で、現在は DRAM ベースの技術開発が比較的多数を占めており、これらの状況を考えるに、DRAM ロジック混載は、まず、高速 DRAM・高機能 DRAM の延長として、まず、メモリ部分に集積してコストパフォーマンスの観点で意味ある論理回路を混載する方向で進むものと思われる。それでは、どのような論理回路が経済的に見合うのであろうか？ この問いに答えるべ

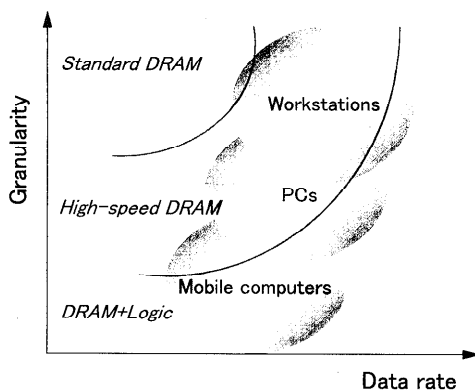


図-2 将来のメモリシステムの構成

く、3つのクライテリアを考えてみたい。

まず第1のクライテリアは、

$$T_m > T_a$$

である。ただし、 T_m はメモリのデータ転送時間、 T_a はメモリ内部の演算処理時間である。このような例としては、メモリ間のブロック転送、単純なビット演算、簡単な整数加算乗算などが考えられる。この場合は、演算回路を DRAM 側に置き、DRAM ロジック混載の高いバンド幅を生かして、性能を向上させることができる。反対に、浮動小数点演算のように、メモリバンド幅でなくて、計算時間自体で性能が決まる場合には、DRAM ベースのトランジスタによる、ロジック回路の性能の低下により、DRAM ロジック混在はかえって全体の性能を落としてしまう。

第2に、メモリ内部に混載されている論理回路には制限があるためにある仕事を効率よく行うには、外部の CPU との連携が重要であり

$$T_c < T_{ex}$$

であることが要求される。ただし、 T_c は外部からのコマンド転送時間、 T_{ex} はそのコマンドによる処理時間である。メモリ内部での演算と外部 CPU との間でどのようなコマンドを受け渡すかについては、

1) CPU がコプロセッサ命令体系をもっていればこれを使う。

2) メモリにメモリインタフェース回路を搭載し、ある特定のメモリにマッピングをされたレジスタにコマンドを書き込み、メモリ側で常にレジスタポーリングをすることによって、CPU からメモリ内部の演算回路の起動ができる。

3) メモリにマルチプロセッサインタフェースと CPU コアをもたせることにより、メインの CPU と混載チップとでマルチプロセッサシステムを構成し、仕事の割り振りは、OS を通して行う。

4) ハードワイヤを使ったコントロールをする。などが考えられる。1) は、連携の効率、ならびに、システムコンパチビリティの点で優れているが、メインの CPU がコプロセッサとの連携命令体系をサブセットでもっている必要がある。2) は、比較的簡単なバスインタフェースをメモリ内部にもつだけで CPU との連携がはかれるが、CPU からメモリへの一方的な制御に限られる。メモリ内部

の論理演算ユニットは、定型処理に対してはハードコードすることも可能だし、柔軟性が必要ならば、CPU からインストラクションストリームをそのつど与える、もしくは、CPU からはローカルメモリに貯えられたインストラクションの列の飛び先だけを与えるという手もある。3) は、OS がマルチプロセッサをサポートしていれば、既存のシステムに対しての互換性という点では優れているが、メモリ側に CPU コアを含めてかなりの論理回路をのせる必要がある。4) は、連携の効率の点では優れているが、特殊な制御信号が必要で一般性に欠ける欠点がある。

第3のクライテリアは、論理回路のチップ全体に占める割合である。先ほど述べたように、DRAM ベースの混載技術では、メタルの配線効率がコストとのトレードオフで決まるため、あまり大きな論理回路を乗せることは、いたずらに、チップ面積を大きくすることになり、コストが上昇する。さらに、混載した論理回路のいくつかが第1のクライテリアを満たさない場合は、性能も低下してしまう。

もし、ロジックベースの混載技術を使うならば、上に述べた制限はほとんどなくなる。すなわち、設計者は、既存のロジックライブラリにある SRAM ハードマクロと同じ感覚で DRAM マクロを扱うことができ、なおかつ、ほかのチップとの連携などを考慮することなしに、必要な論理回路のすべてを混載することが可能であり、設計者にとって理想的な環境を提供する。ただし、製造工程の複雑化による単位面積あたりのコストの上昇をどこまで押さえられるかは、今後のプロセス技術開発による。

もし、半導体技術が進歩して、高速の論理デバイスと、多層の金属配線が、大容量 DRAM メモリセル、さらに、フラッシュなどの不揮発性メモリが容易に混載できるようになったとき、CPU、メモリ、通信機能を含めたシステムオンチップは夢物語ではなくなるだろう。この方向は、1GbDRAM 以降の半導体開発を DRAM とロジックで再び共用化を進めようとの動きや、集積度の向上にともない、パワーデンシティが上昇し論理回路のスピードは、とくにモバイルコンピュータにおいて、トランジスタの性能ではなく消費電力や熱的要因により決定されることも合わせて商

業的にも現実味を帯びてくるだろう。そして、その時代には、人々は高速無線通信ネットワークで有機的につながれた高性能の1チップ超小型コンピュータを誰もがあたかもネックレスやブレスレットを身にまとうように複数個身につけて、町を歩いているに違いない。

参 考 文 献

- 1) Lu, N. C. -C., Katayama, Y. et al.: IEEE JSSC, Vol.24, pp.1198-1205 (Oct. 1989).
- 2) Kitsukawa, G. et al.: IEEE ISSC, Vol.25, pp.1102-1111 (Oct. 1990).
- 3) Takai, Y. et al.: Symp. on VLSI Circuit Dig. Tech. Pap., pp.59-60 (June 1993).
- 4) Burskey, D.: Electronic Design, pp.39-43 (Jan. 1992).
- 5) Dousaka, K. et al.: IEEE JSSC, Vol.27, pp.1534-1539 (Nov. 1992).
- 6) Kushiyama, N. et al.: Symp. on VLSI Circuit Dig. Tech. Pap., pp.66-67 (June 1992).
- 7) Burskey, D.: Electronic Design, pp.43-48 (Nov. 1993).
- 8) Deering, M. et al.: Proc. 21st Int. SIGGRAPH Conf., pp.167-174 (July 1994).
- 9) Stone, H. S.: IEEE Trans. Computers, pp.73-78 (Jan. 1970).
- 10) Sawada, K. et al.: Proc. IEEE CICC 20.3.1 (May 1988).
- 11) Miyano, S. et al.: ISSCC Dig. Tech. Pap., pp.300-301 (Feb. 1995).
- 12) Aimoto, Y. et al.: ISSCC Dig. Tech. Pap., pp.372-373 (Feb. 1996).
- 13) Watanabe, T. et al.: Symposium on VLSI Circuits Dig. Tech. Pap., pp.42-43 (June 1996).
- 14) Shimizu, T. et al.: ISSCC Dig. Tech. Pap., pp.216-217 (Feb. 1996).
- 15) Gillingham, P. et al.: ISSCC Dig. Tech. Pap., pp.262-263 (Feb. 1996).
- 16) Kogge, P. M. et al.: Proc. Int. Conf. on Parallel Processing, pp.77-84 (Aug. 1994).
- 17) Sunaga, T. et al.: IEEE JSSC, Vol.31, pp.1556-1559 (Oct. 1996).
- 18) Murakami, K. et al.: ISSCC Dig. Tech. Pap., pp.228-229 (Feb. 1997).
- 19) Patterson, D. et al.: ISSCC Dig. Tech. Pap., pp.224-225 (Feb. 1997).
- 20) Tarui, Y.: IEDM Tech. Dig., pp.1.2.1-1.2.10 (Dec. 1994).
- 21) Foss, R. C.: ISSCC Dig. Tech. Pap., pp.260-261 (Feb. 1996).

(平成9年1月14日受付)



片山 泰尚 (正会員)

1961年生。1984年東京大学理学部物理学科卒業。1986年同大学院修士課程修了。同年日本アイ・ピー・エム(株)入社。東京基礎研究所にてVLSIの研究に従事。1990年より、同社海外留学制度により、プリンストン大学電子工学科博士課程に入学、1994年工学博士授与。現在東京基礎研究所主任研究員。電子情報通信学会、IEEE、APS各会員。