

パーソナルコンピュータ向けキャッシュシステム に関する一考察

4P-10

中田武男 大庭信之 清水茂則

日本アイ・ビー・エム株式会社 東京基礎研究所

1. はじめに

半導体技術の進歩に伴いマイクロプロセッサの内部動作速度は 100MHz を越えるものが登場してきている。それに対して主記憶となる DRAM は、マイクロプロセッサの持っている性能を十分に発揮させられるだけの能力を備えていない。高速に動作するプロセッサの性能を十分に発揮するためにキャッシュメモリは、必要不可欠なものである⁽¹⁾⁽²⁾。数年前までキャッシュメモリは、大型コンピュータや高性能ワークステーションなど一部の高価なコンピュータシステムにしか実装されていなかった。しかし、ハードウェアが低価格になるにつれ、最近では個人使用のパーソナルコンピュータにまでキャッシュメモリが実装されるようになってきた。本稿では、パーソナルコンピュータの中核となるプロセッサ、キャッシュ、メモリコントローラの3つの構成要素の結合方式の違いによるキャッシュの効果について論じる。

2. キャッシュシステム構成

キャッシュシステムの構成は、プロセッサ、メモリコントローラとキャッシュ（キャッシュコントローラ+キャッシュメモリ）の論理的な接続形態の違いによりシリアルキャッシュとパラレルキャッシュに大別される。図1に示される様にシリ

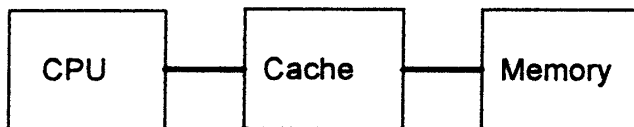


図1 シリアルキャッシュシステム

アルキャッシュシステムでは、キャッシュが、プロセッサとメモリコントローラの上に直列に置かれる。プロセッサのアクセスは全てキャッシュを通して行われるため、Look-through キャッシュとも呼ばれる。それに対して、パラレルキャッシュシステムでは、図2に示される様にプロセッサに対してキャッシュとメモリコントローラは並列に置かれ、プロセッサのアクセス要求はキャッシュとメモリコントローラの両方に同時に発行され、キャッシュはプロセッサとメモリコントローラの動作を横から監視するような構成になり、Look-aside キャッシュとも呼ばれる。

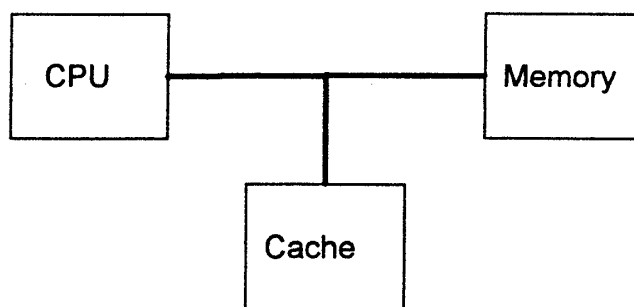


図2 パラレルキャッシュシステム

3. シリアルキャッシュ

シリアルキャッシュシステムでは、プロセッサとメモリコントローラの上にキャッシュが置かれる。従って、プロセッサが主記憶やI/Oにアクセスするためには必ずキャッシュを通して行う必要がある。

長所としては、主に次の2点が考えられる。
(1) プロセッサ側のバスとメモリ側のバスは、キャッシュにより分離されるため、プロセッサバスとメモリバスの速度を独立に設定することができる。例えば、プロセッサバスが 100MHz でメモリバスを 33MHz で使用することも可能になる。そのため、外部(メモリ)バスが 33MHz のシ

システムにもかかわらずキャッシュヒットの間は、外部バスが 100MHz のシステムとして動作させることができ、高性能のプロセッサの能力を最大限に発揮させることができる。(2) メモリバス側からのスヌープ要求は、1度キャッシュで判定し、必要なスヌープ要求だけをプロセッサ側に伝える方法がとれるため、プロセッサに対して不必要なスヌープのオーバーヘッドがかからない。

短所としては、主に次の2点があげられる。

(1) キャッシュミスサイクルでは、プロセッサがバスサイクルを発行した次のサイクルにメモリコントローラ側にバスサイクルが発行され、I/O サイクルなどの本来キャッシュに関係の無いサイクルに対しても1サイクル以上余計にかかり、バスサイクルのレイテンシーは、長くなってしまう。(2) データのコンシステンシを保つために、キャッシュラインをリプレースする際の方法には十分注意する必要がある⁽⁹⁾。例えば、あるデータがプロセッサ内のキャッシュには存在するにもかかわらず、外部キャッシュには存在しない、ということが起こらないように配慮しなければならない。

4. パラレルキャッシュ

パラレルキャッシュシステムでは、プロセッサのバスアクセスはキャッシュメモリとメモリコントローラの双方に対して発行される。データの読み出しの際にキャッシュメモリにヒットしたときは、キャッシュからデータが供給されメモリコントローラにはデータ読み出しサイクルの中断を要求する信号が出される。データの書き込みの際には、ライトバック方式をとった場合には、キャッシュにヒットしたときは、キャッシュの中にだけデータが書き込まれ主記憶への書き込みは行われない。

長所としては、次の3点が考えられる。(1) メモリコントローラとキャッシュが同時に動作を開始するためキャッシュが付いていてもI/O サイクルなどキャッシュに無関係なサイクルに対してのオーバーヘッドはない。同様に、メモリアクセスに対して、キャッシュミスしたときのペナ

ルティがない。(2) 外部キャッシュを簡単にオプション化できる。パラレルキャッシュでは、キャッシュとメモリコントローラの両方に対してプロセッサのアクセス要求が同時に発行される。そのため、キャッシュをオプションとして、システムアップグレードの際にキャッシュを組み込むような設計をとることが容易である。(3) スヌープ要求はキャッシュとプロセッサの両方に発行されるので、データのコンシステンシ上の問題がほとんど生じないため制御回路が容易になる。

短所としては、次の3点がある。(1) DMA 転送の際などに起こるスヌープ要求が全てキャッシュとプロセッサの両方に伝えられるため、プロセッサに対するスヌープのオーバーヘッドを削減できない。(2) 全てのプロセッサがメモリコントローラに直結されるため、マルチプロセッサ構成をとった場合に高い性能を引き出せない。なぜならば、マルチプロセッサ構成をとった場合、プロセッサバスに全てのアクセスが出てしまい、本来キャッシュとプロセッサの間で処理が続けられる場合にもプロセッサの動作が止められてしまうためである。(3) 1つのプロセッサバスに多くのデバイスがつながるためバスの負荷が重くなり、バスクロック周波数をあまり高くできない。

5. まとめ

マルチプロセッサのサーバーシステムでは、シリアルキャッシュでなければ性能が出せないのが必然的にシリアルキャッシュが採用される。しかし、シングルプロセッサのパーソナルコンピュータシステムではパラレルキャッシュの方が価格性能比が高いシステムを組むことができる。

参考文献

- (1) J. L. Hennessy and D. A. Patterson, *Computer Architecture: A Quantitative Approach*, Morgan Kaufmann Publishers, 1990.
- (2) A. J. Smith, *Cache Memories*, Computing Surveys, vol. 14, No. 3, pp. 473-530, 1982.
- (3) J. L. Baer and W. H. Wang, *On the Inclusion Properties for Multi-level Cache Hierarchies*, 15th Computer Architecture, pp.73-80, 1988.