

Bộ nhớ cache

Khi tốc độ làm việc của bộ vi xử lý ngày càng vượt xa tốc độ truy nhập bộ nhớ chính (được tính theo ns, DRAM làm việc nhanh nhất chỉ là 60ns – nanogiây) có nghĩa là bộ vi xử lý phải mất thêm vài chu kỳ đợi bộ nhớ hoàn thành quá trình đọc/ghi. Điều này làm giảm hiệu suất làm việc của bộ vi xử lý. Một giải pháp hữu hiệu là sử dụng thêm bộ nhớ đệm *cache* với tốc độ truy nhập chỉ vài ns đến 10ns.

Bộ nhớ cache còn được gọi là bộ nhớ truy cập nhanh. Nó nằm giữa bộ vi xử lý và bộ nhớ chính với dung lượng không lớn (cỡ KB đến 1 hoặc 2MB) tùy theo loại cache. Cache sẽ tiết kiệm thời gian truy xuất bộ nhớ của CPU bằng cách dự đoán trước lệnh kế tiếp mà CPU sẽ cần và nạp nó vào trong cache trước khi CPU thực sự cần đến nó. Nếu lệnh cần thiết đã có sẵn trong cache thì CPU sẽ truy xuất dữ liệu từ cache, nếu không, CPU mới truy xuất lên bộ nhớ chính.

Cache được phân thành 2 loại: Cache L1 và cache L2.

○ **Cache L1:**

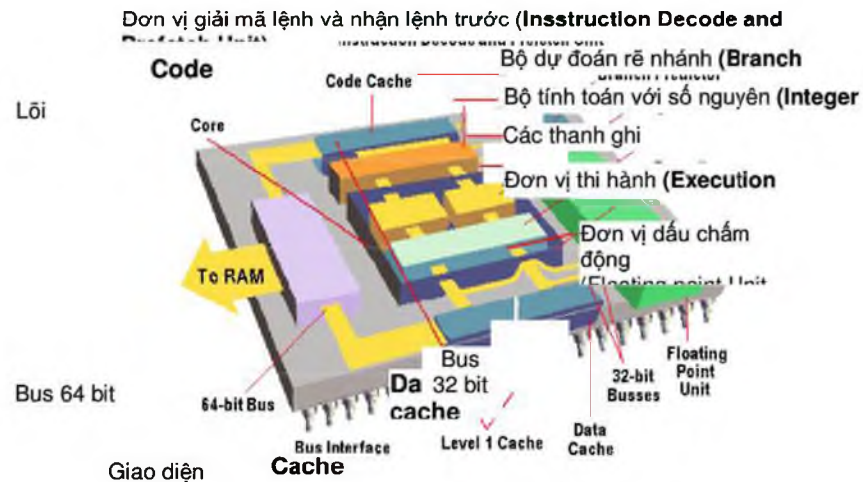
Bộ nhớ được tích hợp trong chính bản thân CPU được gọi là cache nội (internal cache) hay cache sơ cấp (Primary cache), cache L1 (level 1 cache). Tốc độ truy nhập cache xấp xỉ bằng tốc độ làm việc của CPU, nhưng dung lượng khá nhỏ.

Năm 1988, lần đầu tiên cache L1 được thiết kế cho CPU đời 80386SX, sau đó là 80486DX, 80486DX2. Cuối năm 1994, Intel đã giới thiệu sản phẩm CPU 80486DX4 với tốc độ 75-120MHz, lúc này cache L1 được phân thành 2 bộ nhớ với hai chức năng khác nhau đó là :

Data cache : để lưu trữ dữ liệu, với dung lượng là 8KB

Code cache : để lưu trữ mã lệnh, với dung lượng là 8KB

Tiếp theo 80486, Pentium và đời Pentium tiếp theo tuy bổ sung các tính năng kỹ thuật mới nhằm nâng cao tốc độ đến GHz, nhưng vẫn kế thừa việc sử dụng cache L1 với dung lượng : 8 + 8KB, 16+16 KB. Hình 3.3. là sơ đồ cấu trúc cơ bản của bộ vi xử lý Pentium.



Hình 3.3. Cấu trúc cơ bản của bộ vi xử lý Pentium

o **Cache L2:**

Một cache nằm bên ngoài CPU gọi là external cache, cache thứ cấp (Secondary cache), cache mức 2 - L2. Cache L2 thường có kích thước 256 KB hoặc 512KB. Trước kia tất cả các cache L2 đều được gắn lên mainboard, nhưng bắt đầu từ các CPU Pentium, cache L2 đã được đưa vào trong ***cùng một vỏ bọc*** với CPU – chứ không nằm ngay bên trong CPU như cache L1.

Để nối CPU tới cache L2 bắt buộc phải sử dụng Bus. Bus này được gọi là ***Bus tuyến sau - Back Side Bus***, vì bạn không thể thấy được bus do nó nằm kín trong vỏ bọc CPU. Trái lại bus nối CPU với bộ nhớ nằm ngoài vỏ bọc

gọi là *Bus tuyến trước* - **Front Side Bus (FSB)** hay **Bus bộ nhớ** và ta có thể thấy được trên mainboard.

Cache L2 có 2 loại :

- **Cache L2 rời** được gắn trên vi mạch nhỏ và nối với CPU bởi Back Side Bus có độ rộng là 64 bit. Kiến trúc này làm cho thời gian truy xuất cache L2 nhanh hơn so với khi nó được gắn trên bo mạch hệ thống. Nhưng cũng chỉ bằng một nửa tốc độ làm việc của CPU có nghĩa là tốc độ truy xuất của CPU tới cache L2 không thể nhanh bằng truy xuất lên cache L1.
- **Cache L2 ATC** được tích hợp trên bề mặt tại cùng một vị trí của lõi CPU (vị trí lõi CPU được chỉ ra trên hình 3.3) và nối với CPU bằng bus (Back Side Bus) có độ rộng là 256bit. Tốc độ truy xuất dữ liệu bằng với tốc độ làm việc của bộ vi xử lý. Do đó cache này còn được gọi là Advanced Transfer cache (ATC). Tuy vậy, ATC cache vẫn không được xem là cache L1 vì nó không nằm trong một vi chip với CPU, mà chỉ nằm trong cùng một vỏ bọc với CPU mà thôi.