

Lab 10

Cache Memory

Đinh Ngọc Khánh Huyền – 20225726

Assignment 1

How is the 32-bit address used to access a cache?

Địa chỉ sẽ được lưu theo cách tổ chức n-way set associative cache. Cache kết hợp n-way chia cache thành các sets, mỗi set chứa n dòng cache. Khi truy cập cache, địa chỉ được chia thành các phần để xác định vị trí cụ thể trong cache:

- Phần Tag: Được sử dụng để xác định xem dữ liệu được yêu cầu có tồn tại trong cache hay không. Phần này thường có độ dài bằng với số bit cần để xác định một dòng cache duy nhất trong mỗi bộ.
- Phần Index: Xác định bộ (set) trong cache mà dữ liệu được lưu trữ hoặc truy cập. Số bit cần cho phần index phụ thuộc vào số lượng các bộ trong cache.
- Phần Offset: Xác định vị trí cụ thể trong dòng cache mà dữ liệu được lưu trữ hoặc truy cập. Số bit cần cho phần offset thường phụ thuộc vào kích thước của mỗi dòng cache.

Khi một yêu cầu truy cập dữ liệu đến cache, hệ thống sẽ xác định set tương ứng bằng cách sử dụng phần index của địa chỉ. Sau đó, nó sẽ so sánh phần tag của yêu cầu truy cập với các tag của các dòng trong bộ đó để kiểm tra xem dữ liệu có tồn tại trong cache hay không. Nếu có một sự khớp, dữ liệu sẽ được trả về từ cache. Nếu không có khớp nào, dữ liệu sẽ phải được đọc từ bộ nhớ chính.

What happens when there is a cache miss/hit?

Đầu tiên, CPU sẽ gửi yêu cầu truy cập dữ liệu đến Cache. Cache sẽ kiểm tra xem có dữ liệu trong Cache hay không.

- Nếu có (hit), Cache sẽ trả về dữ liệu ngay lập tức cho CPU mà không cần phải truy cập bộ nhớ chính. CPU có thể tiếp tục thực hiện các thao tác xử lý tiếp.
- Nếu không (miss), CPU sẽ phải gửi yêu cầu truy cập dữ liệu đến bộ nhớ chính (RAM) để lấy dữ liệu và thao tác, dữ liệu sau đó sẽ được sao chép vào Cache để thuận tiện hơn cho các truy cập sau.

What is the block size?

Block size là kích thước hay dung lượng của 1 block, như trong Mips mặc định words là 4 bytes.

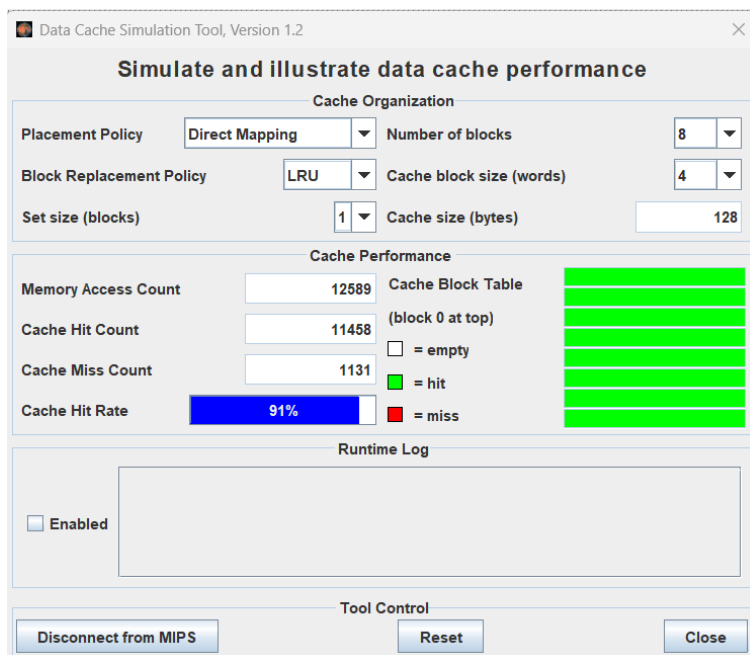
What is the function of the tag field in a referenced address?

Như đã nói ở trên, phần Tag được sử dụng để xác định xem dữ liệu được yêu cầu có tồn tại trong Cache hay không. Phần này thường có độ dài bằng với số bit cần để xác định một dòng cache duy nhất trong mỗi bộ.

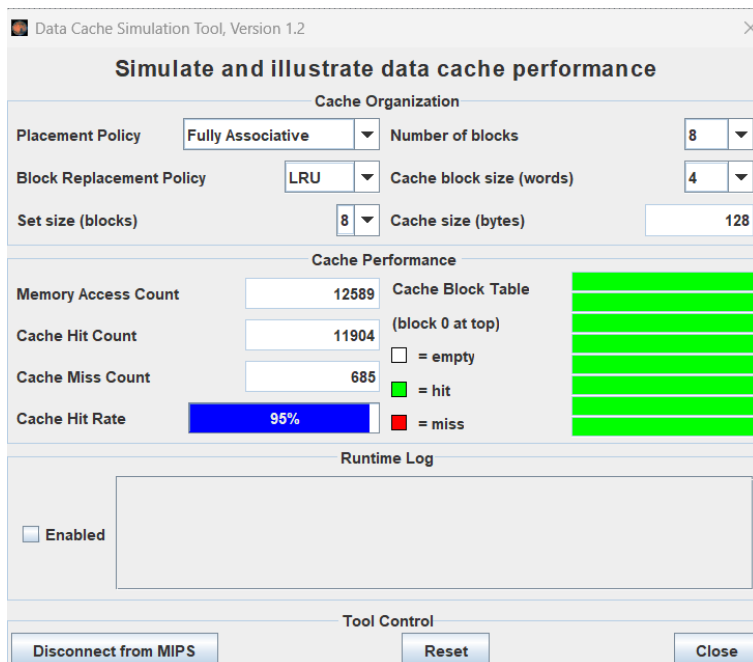
Assignment 3

- Placement Policy

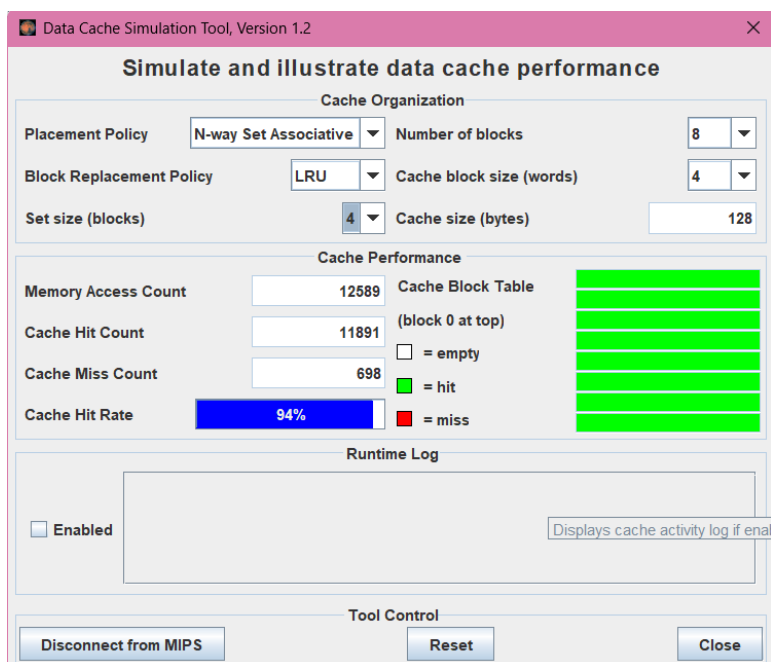
+ Direct Mapping : Hit rate có thể bị ảnh hưởng bởi hiện tượng xung đột khi nhiều khối dữ liệu chia sẻ cùng một dòng cache. Nếu các khối dữ liệu xung đột trong cùng một dòng cache, sẽ xảy ra cache miss khi cố gắng truy cập chúng đồng thời.



+ Fully-Associative: Hit rate trong cache toàn bộ liên kết thường cao hơn so với cache trực tiếp vì không có khả năng xảy ra lỗi xung đột. CPU có thể tìm kiếm dữ liệu ở bất kỳ vị trí nào trong cache.

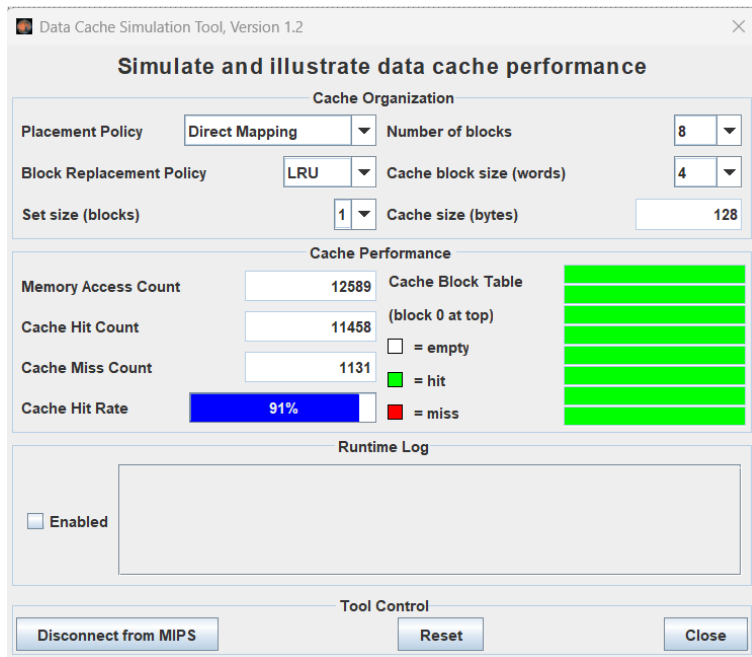


+ N-way Set Associative: Hit rate ở cache kết hợp n-way thường cao hơn so với cache trực tiếp nhưng thấp hơn so với cache toàn bộ liên kết. Điều này là do mỗi khối dữ liệu có một số lượng dòng hữu hạn để lưu trữ, giảm thiểu khả năng xảy ra lỗi xung đột.

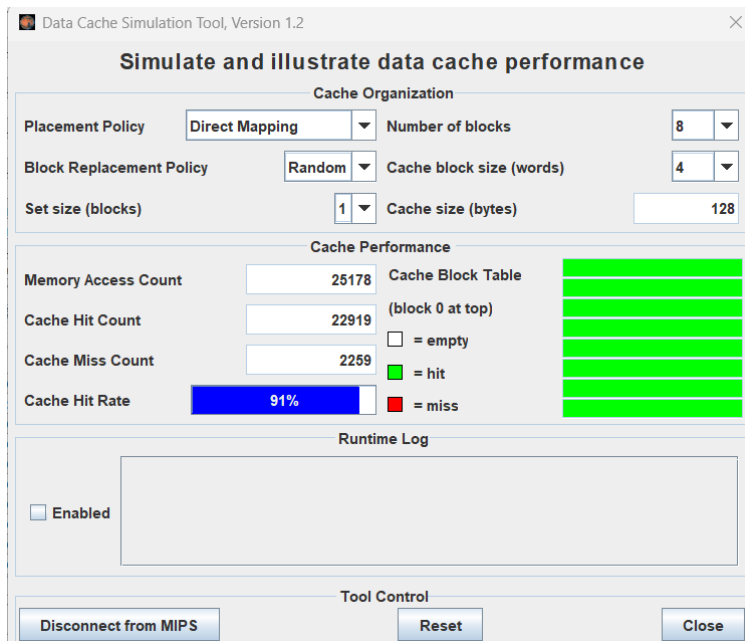


- Block Replacement Policy

+ LRU thường có thể cung cấp hit rate cao hơn so với chính sách ngẫu nhiên, đặc biệt trong các tình huống mà có sự tổ chức hoặc mô hình truy cập dữ liệu theo thời gian.

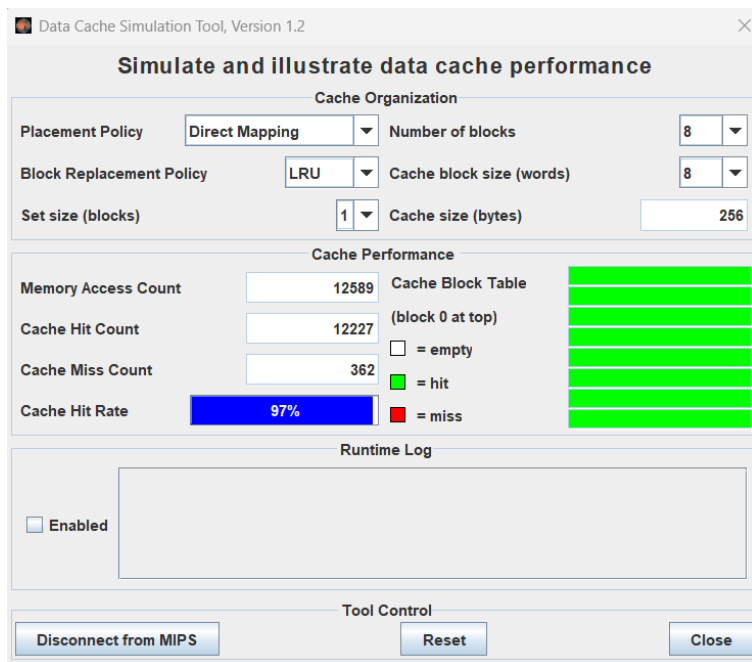
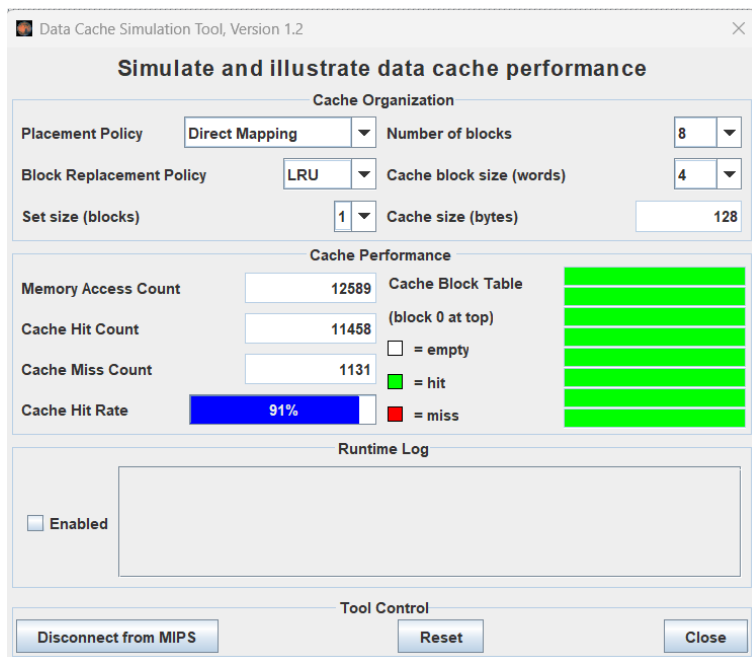


+ Tuy nhiên, trong một số trường hợp, chính sách ngẫu nhiên có thể hoạt động tốt nếu mô hình truy cập dữ liệu không có sự thứ tự đặc biệt hoặc không thay đổi theo thời gian.



- Cache Block Size

Khi tăng Cache Block Size thì hit rate sẽ tăng lên do nếu kích thước block lớn thì Cache có thể lưu trữ nhiều dữ liệu hơn, dẫn đến cơ hội trùng hit càng cao. Tuy nhiên, sẽ gây lãng phí bộ nhớ và tốn tài nguyên.



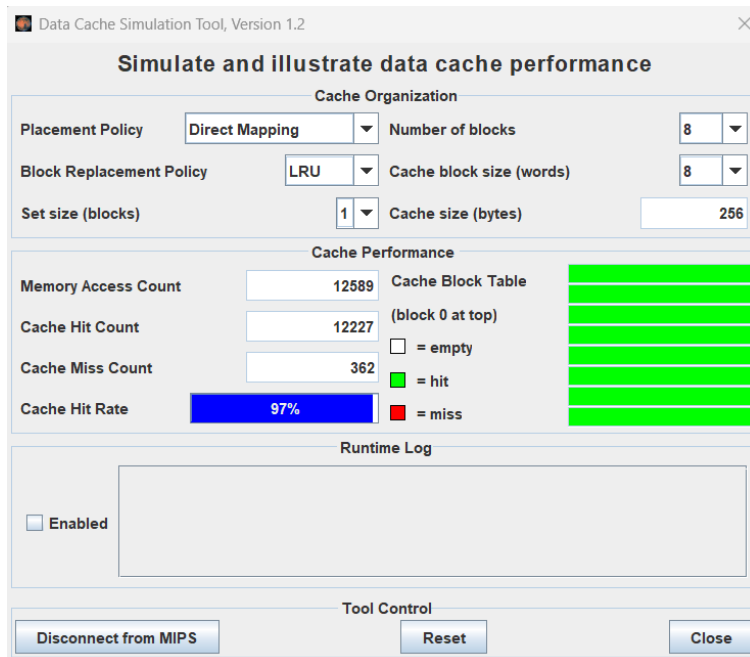
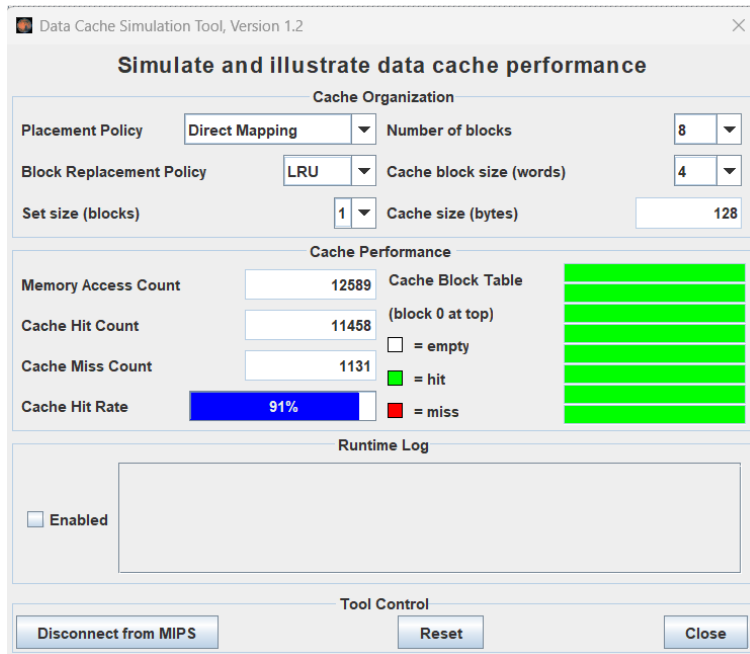
Khi tăng Cache Block Size từ 4 lên 8, hit rate đã tăng từ 91% lên 97%

- Set Size

Với cách tổ chức Cache N-way Set Associative, Set size càng lớn sẽ tăng hit rate càng cao. Do set size lớn sẽ giảm khả năng gặp các xung đột như đã giải thích ở phần Placement Policy. Tuy nhiên, Set Size nhỏ sẽ tiết kiệm thời gian tìm kiếm do thời gian tìm kiếm dữ liệu trong từng set sẽ giảm đi. (Có thể xem ảnh minh họa ở phần Placement Policy.)

- Cache Size

Khi tăng Cache Size thì hit rate sẽ tăng do Cache lớn hơn cung cấp nhiều dung lượng hơn để lưu trữ dữ liệu, giúp giữ lại các dữ liệu quan trọng hơn trong cache và giảm thiểu cache miss.



Khi tăng Cache Size từ 128 lên 256, hit rate đã tăng từ 91% lên 97%.