

Bài thực hành số 2

Bộ nhớ Cache

Phần mềm SMPCache:

<https://drive.google.com/file/d/1nGdKZzwNZzBMcd8FpU4R8FHftLVFenei/view?usp=sharing>

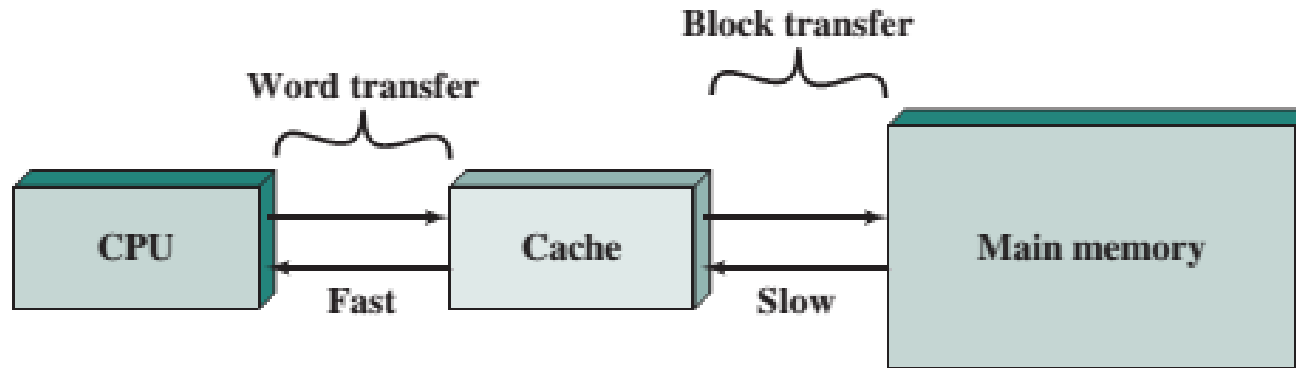
Mục đích

- Hiểu nguyên lý bộ nhớ Cache
- Đánh giá sự ảnh hưởng của các yếu tố đến hiệu suất hệ thống
 - Kiểm tra tính cục bộ (lân cận) của các chương trình
 - Khảo sát ảnh hưởng của kích thước Cache đến hiệu suất hệ thống
 - Khảo sát ảnh hưởng của kích thước Block đến hiệu suất hệ thống

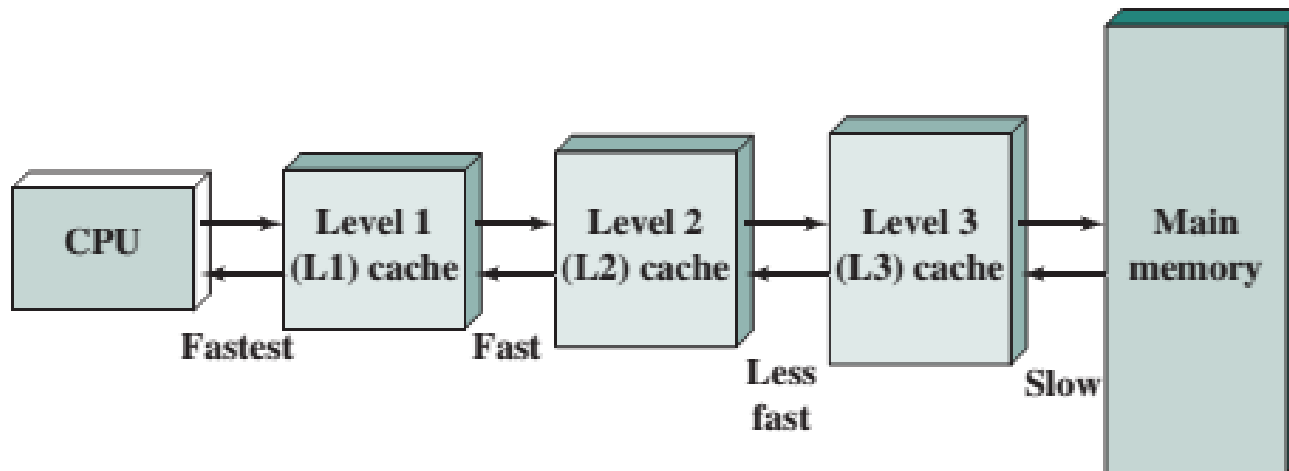
Nội dung

1. Kiến thức cơ bản về Cache
2. Làm quen với SMPCache
3. Quy trình mô phỏng trên SMPCache
4. Thực hành

1. Kiến thức cơ bản về Cache

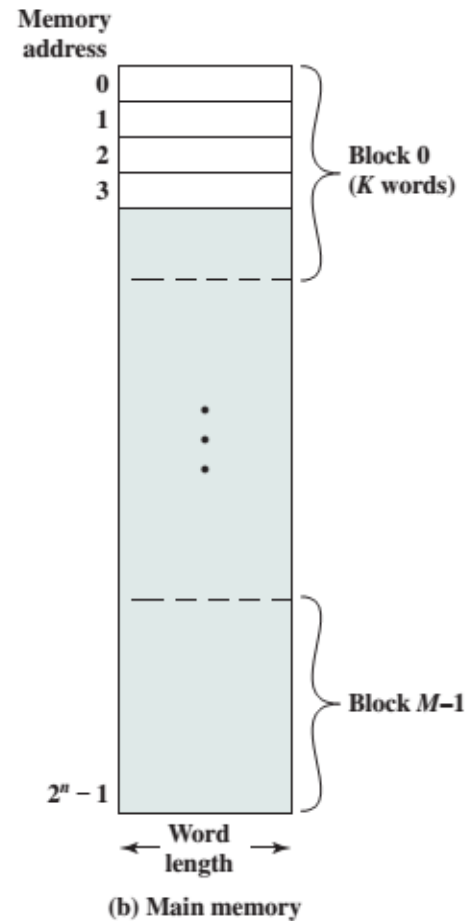
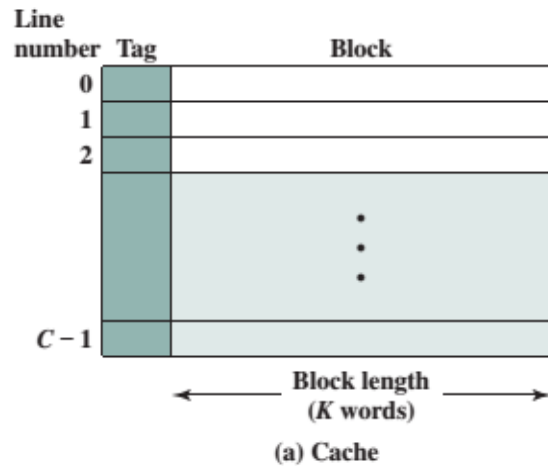


(a) Single cache



(b) Three-level cache organization

1. Kiến thức cơ bản về Cache



- Phương pháp ánh xạ (mapping):
 - Trực tiếp (**Direct**)
 - Kết hợp hoàn toàn (**Full-Associative**)
 - Tập kết hợp (**Set-Associative**)

1. Kiến thức cơ bản về Cache

- Các chính sách thay thế (áp dụng cho phương pháp ánh xạ kết hợp hoàn toàn và tập kết hợp)
 - **LRU** (Least Recently Used)
 - Block nằm trong cache lâu nhất mà không có tham chiếu đến nó sẽ bị thay thế
 - **LFU** (Least Frequently Used)
 - Block trong cache ít được truy xuất nhất sẽ bị thay thế
 - **FIFO** (First-In, First-Out)
 - Block trong Cache được đưa vào lâu nhất sẽ bị thay thế
 - **Random**
 - Chọn ngẫu nhiên một block trong Cache để thay thế

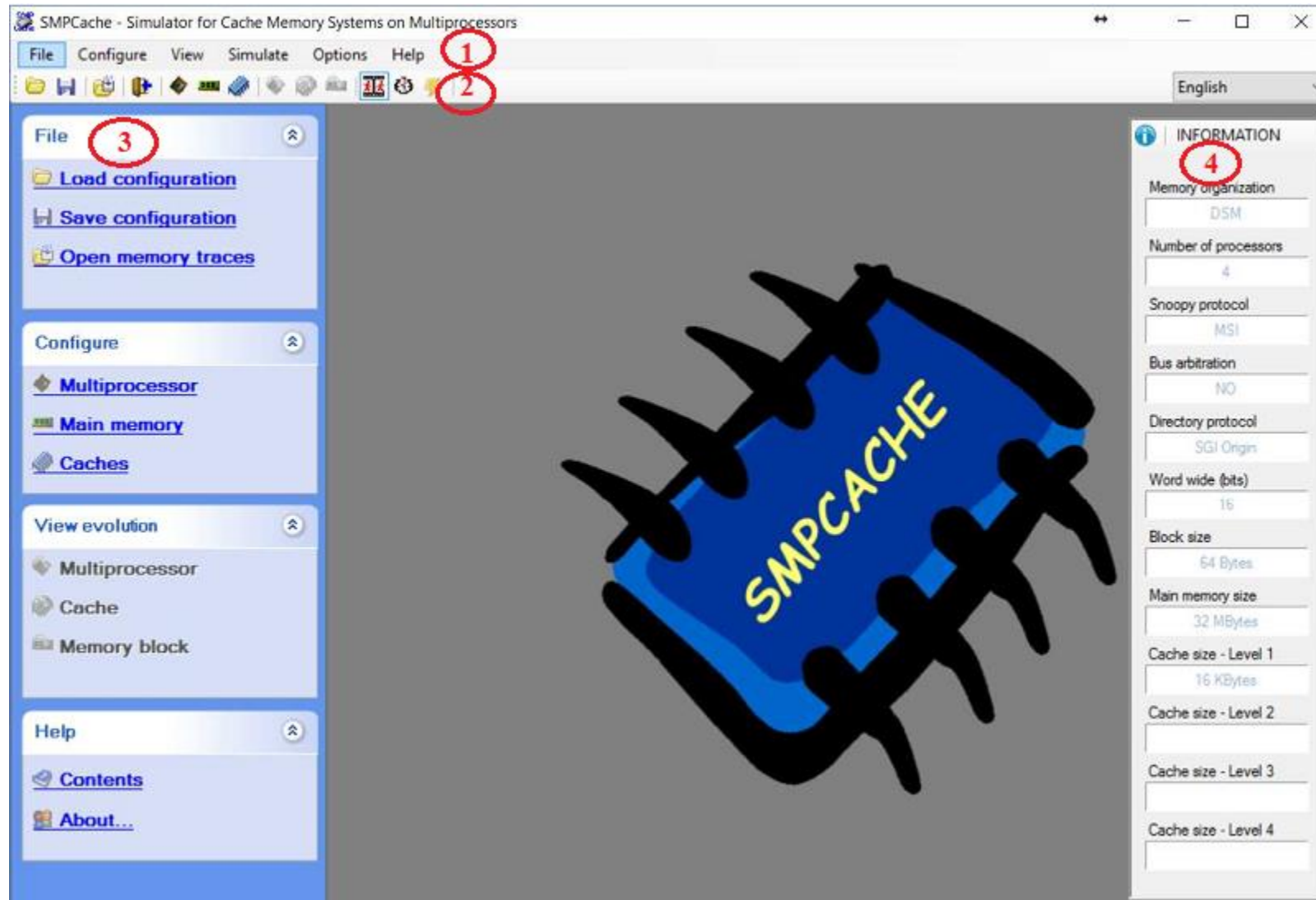
1. Kiến thức cơ bản về Cache

- Chiến lược ghi (write policy): khi nội dung trong cache bị thay đổi (write-hit)
 - **Write-through**: Ngay sau khi block trong cache được cập nhật, thì cập nhật luôn block tương ứng trong main memory
 - **Write-back**: Chỉ cập nhật block trong cache. Block tương ứng trong main memory chỉ được cập nhật khi block đó trong cache bị thay thế

2. Làm quen với SMPCache Simulator

- SMPCache is a simulator for teaching and analysis of cache memory systems on scalable multiprocessors with distributed shared memory (DSM-Distributed Shared Memory) or symmetric multiprocessors (SMP-Symmetric MultiProcessors) using a bus based shared memory.

2. Làm quen với SMPCache Simulator



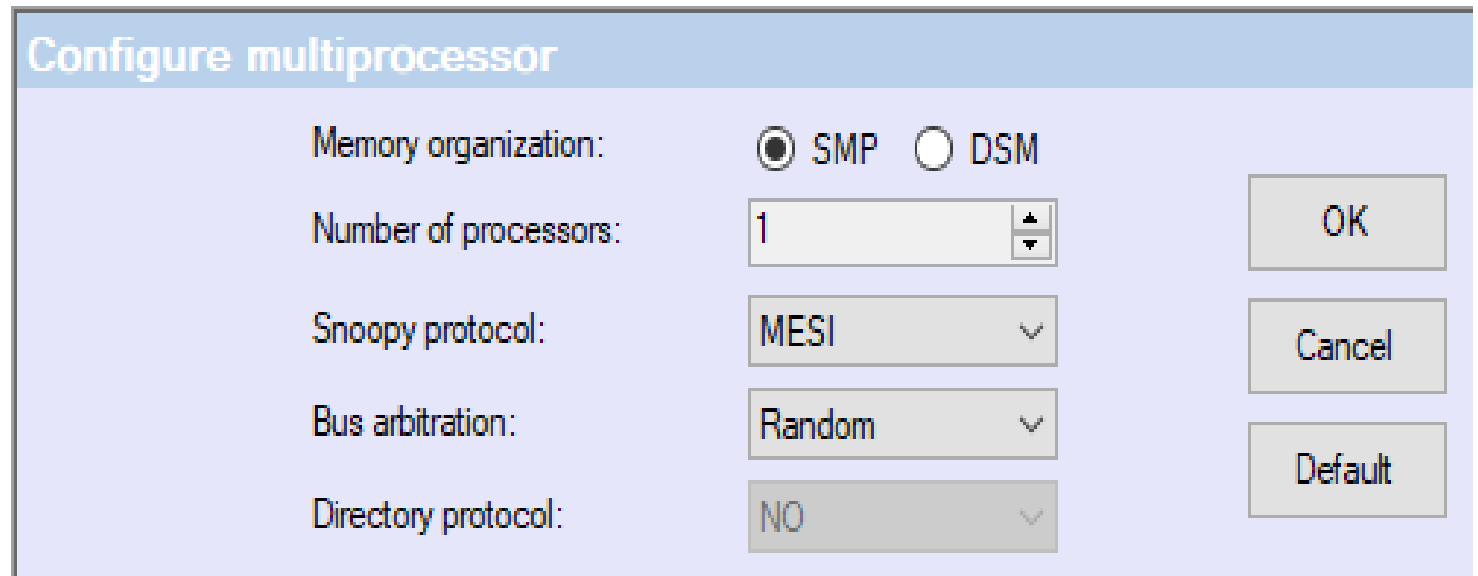
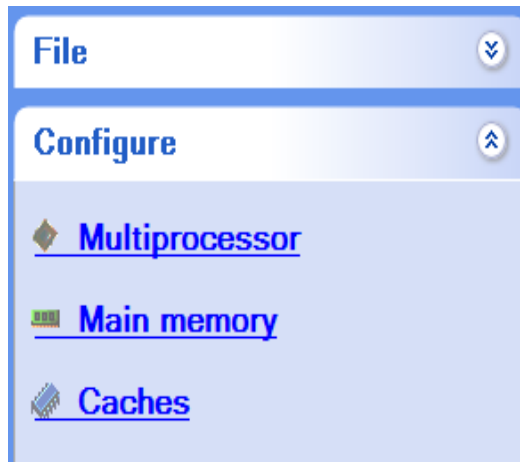
1. Main menu
2. Toolbar
3. Navigator Bar
4. Information Panel

3. Qui trình mô phỏng trên SMPCache

- Bước 1: Thiết lập cấu hình hệ thống
- Bước 2: Nạp file memory trace vào processors
- Bước 3: Lựa chọn chế độ hiển thị và bắt đầu mô phỏng
- Bước 4: Lựa chọn chế độ mô phỏng và thực hiện mô phỏng

Bước 1: Thiết lập cấu hình hệ thống

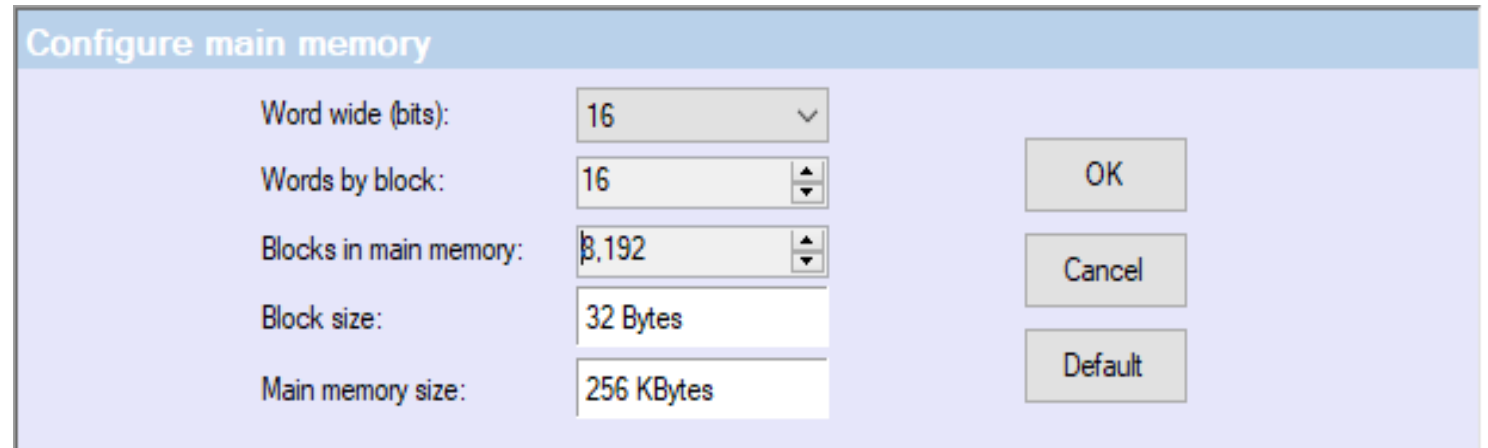
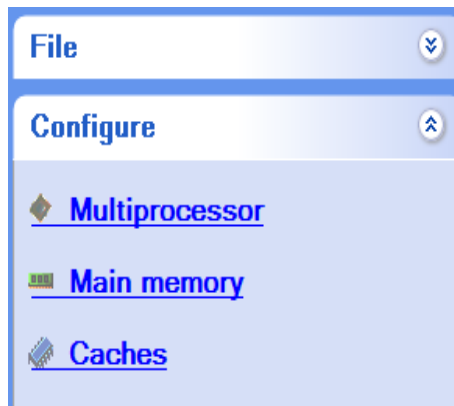
- Click vào **Multiprocessor**



- Memory organization: SMP (Symmetric Multi-Processors) và DSM (Distributed Shared Memory) là hai kiểu tổ chức hệ thống nhớ chia sẻ trong các hệ thống đa vi xử lý (multi-processor system)
- Snoopy protocol (Giao thức liên kết cache): MSI, MESI, DRAGON
- Bus arbitration (Phân xử bus): LRU, LFU, Random

Bước 1: Thiết lập cấu hình hệ thống

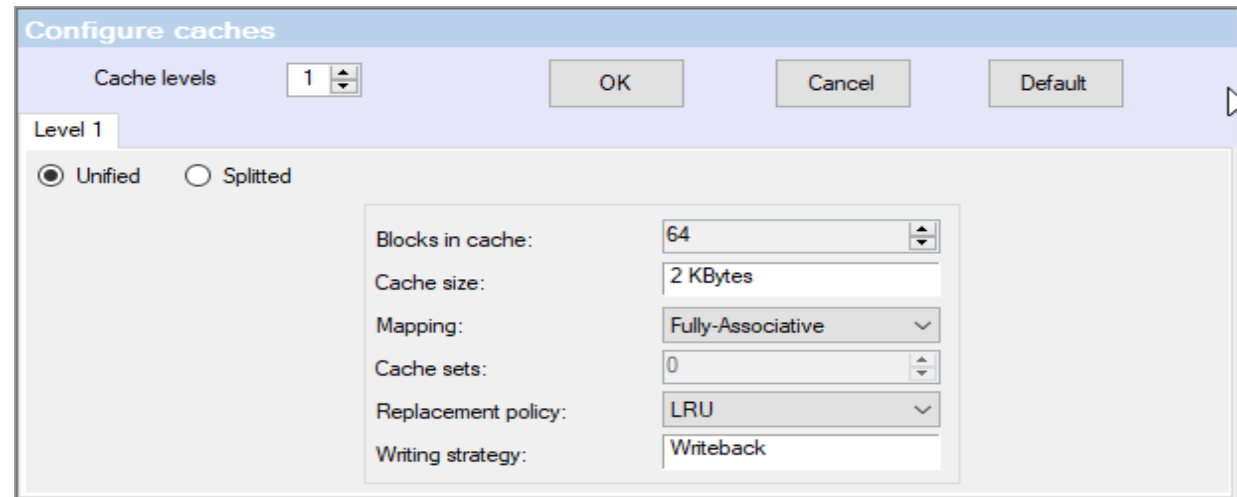
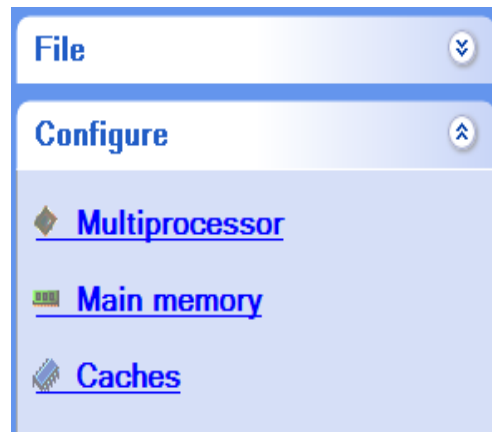
- Click vào **Main memory**



- Word wide (bits): Số bit của một word → Số bytes
- Words by block: Số word trong một block
- Blocks in main memory: Số block trong bộ nhớ chính
- Block size = Word wide (byte) x Words by block (Bytes)
- Main memory size = Blocks in main memory x Block size (Bytes)

Bước 1: Thiết lập cấu hình hệ thống

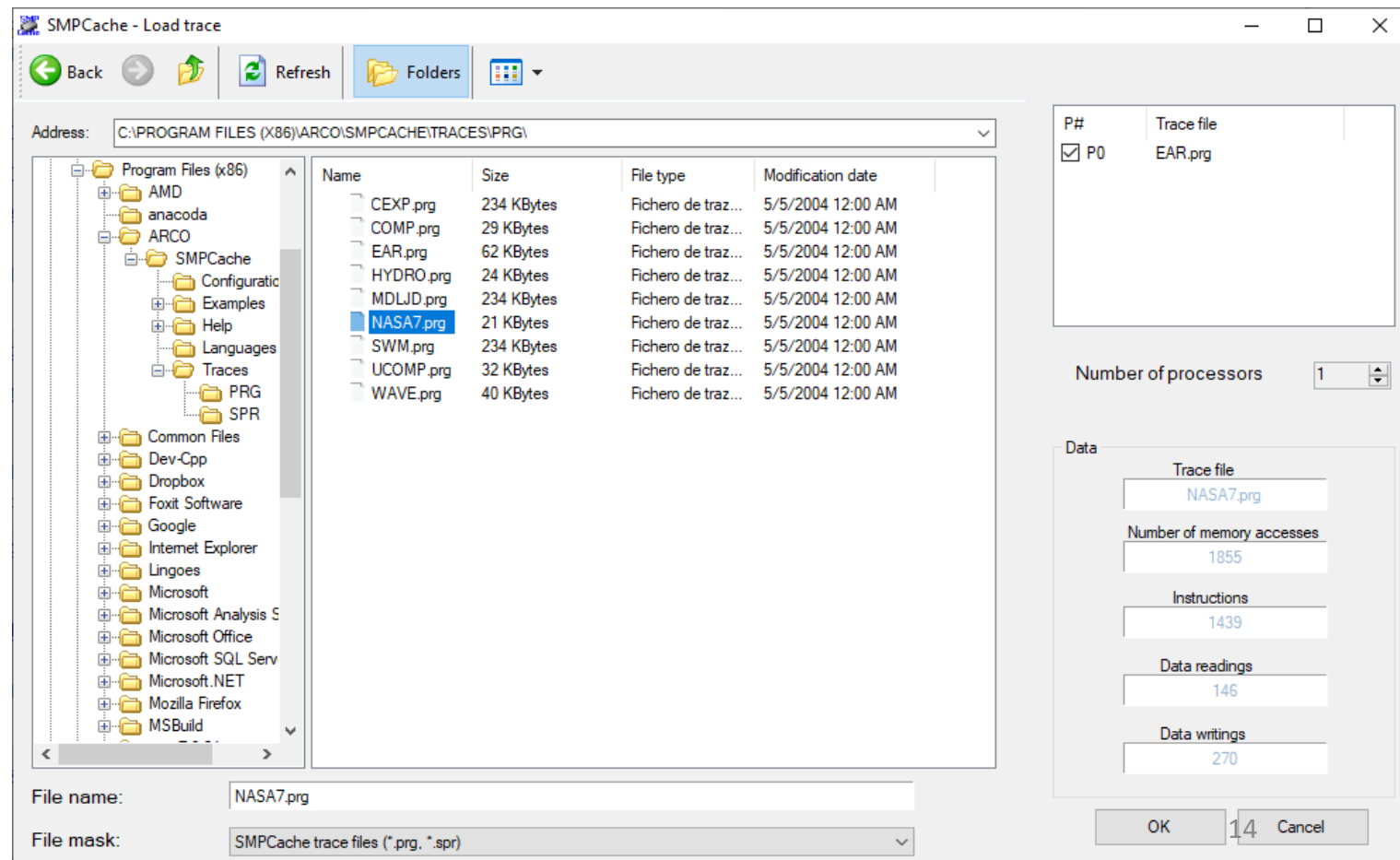
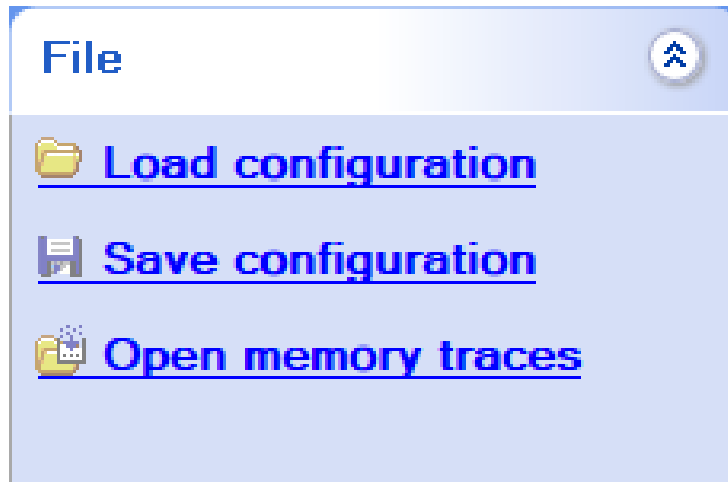
- Click vào **Cache**



- Blocks in cache: Số blocks mà cache có thể chứa (Số lines trong cache)
- Cache size (kích thước cache) = Blocks in cache x Block size → Số Byte
- Mapping (Chế độ ánh xạ): Direct, Fully-Associative, Set-Associative
- Cache sets (Chỉ có thể chọn khi chọn Mapping là Set-Associative): Số line trong một set (k-way)
- Replacement policy (Không áp dụng với chế độ ánh xạ là Mapping): Chính sách thay thế
- Writing strategy: chính sách ghi

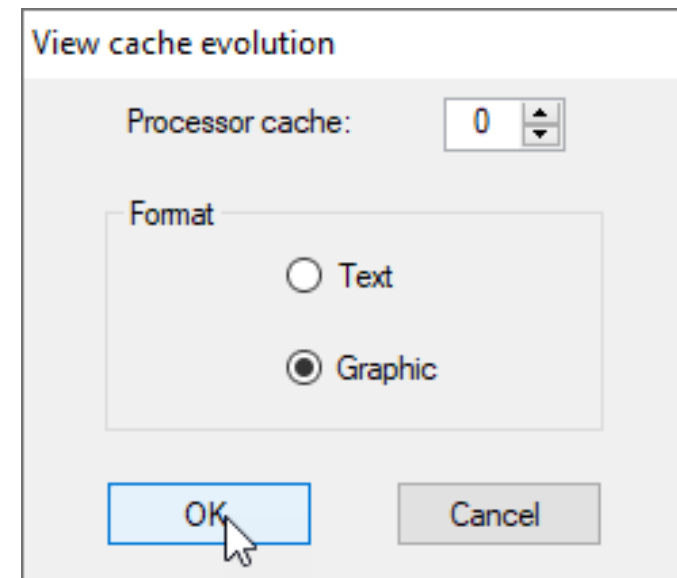
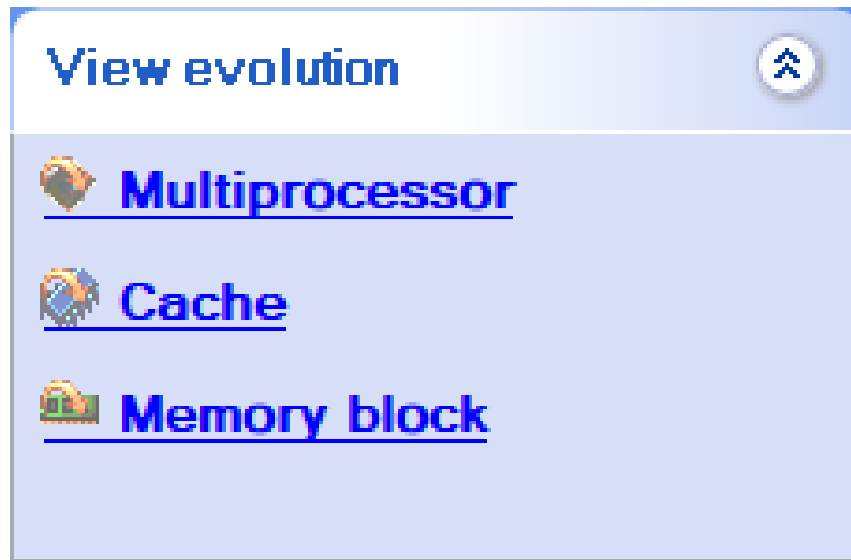
Bước 2: Nạp file memory traces vào processor

- Click vào **Open memory traces**

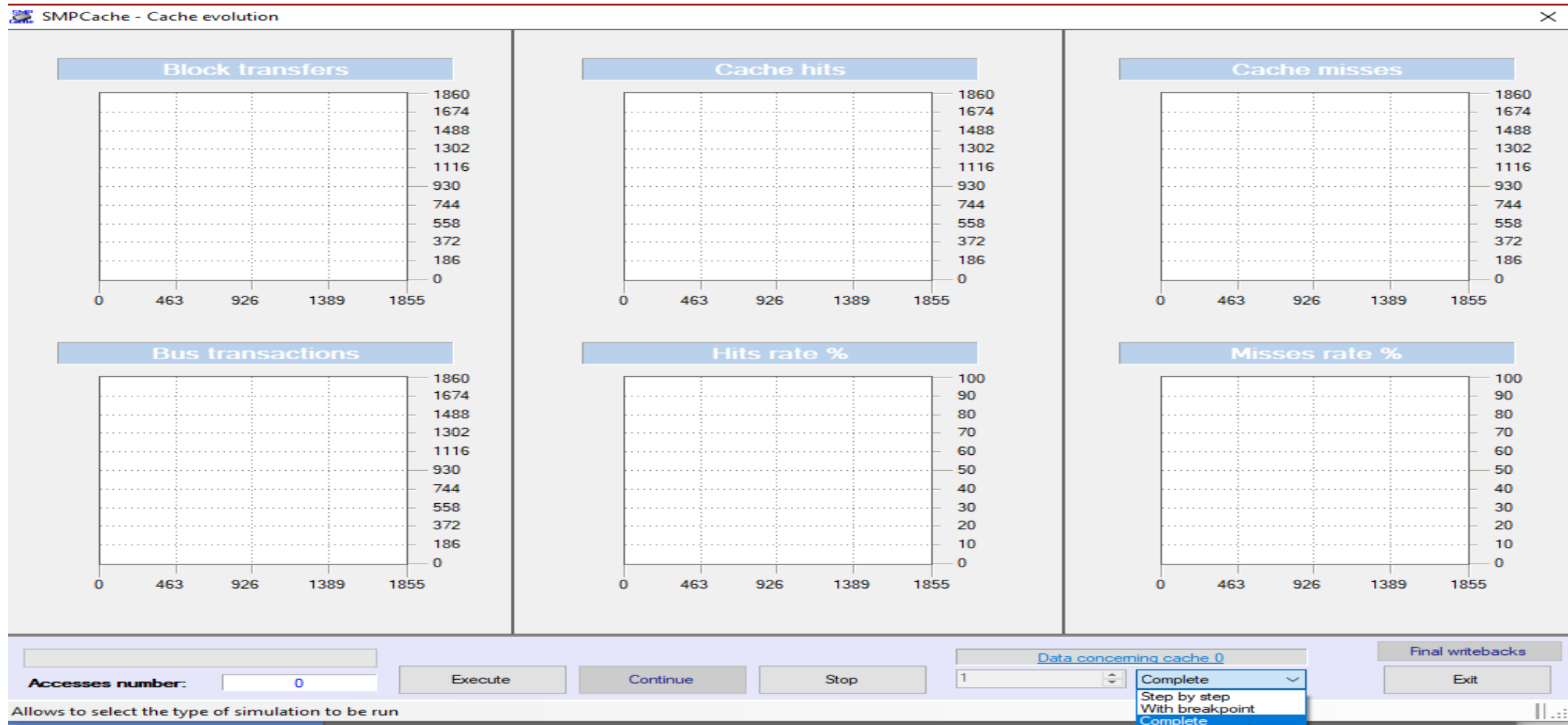


Bước 3: Lựa chọn chế độ hiển thị và bắt đầu mô phỏng

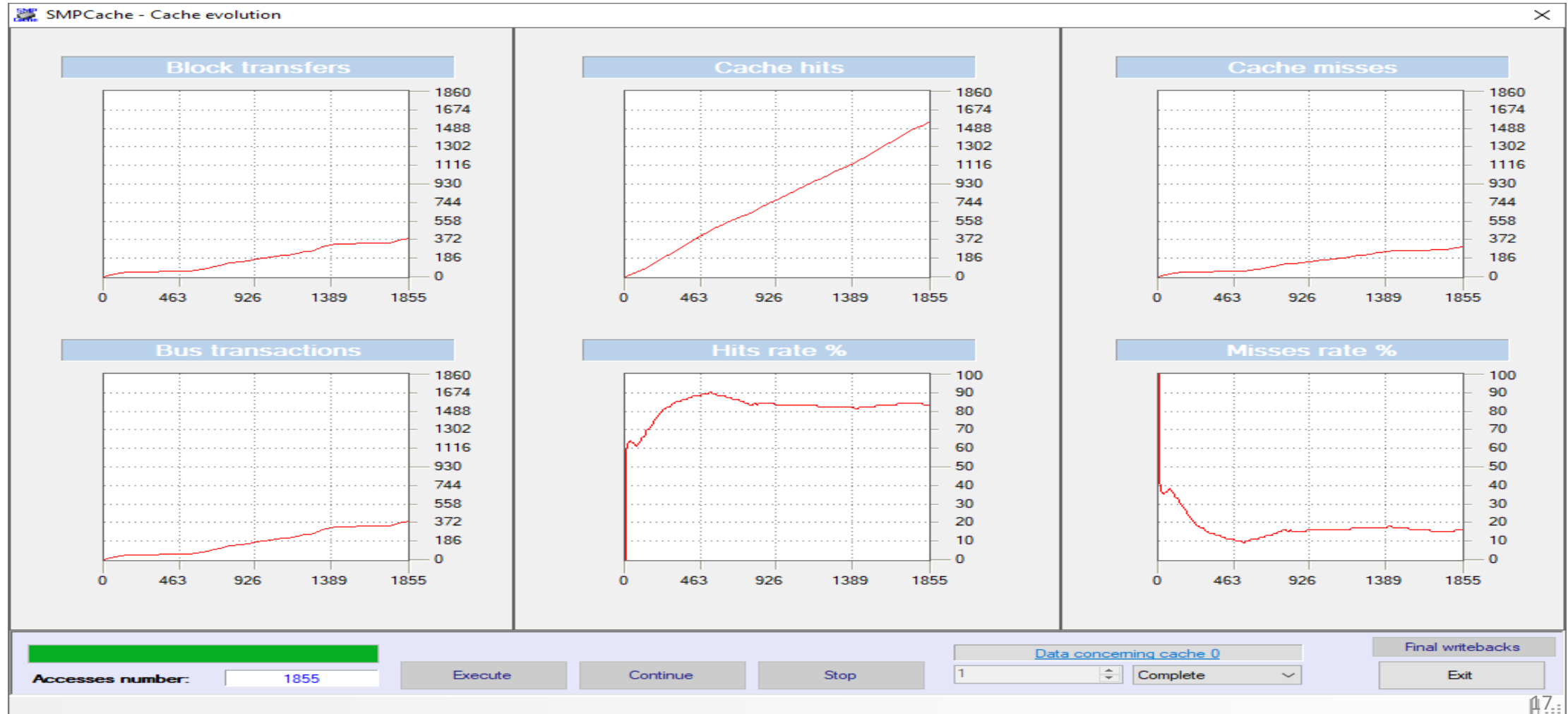
- Click vào **Cache** → Chọn chế độ hiển thị **Text** hoặc **Graphic**
 - Text**: hiển thị thông tin dạng bảng thống kê
 - Graphic**: hiển thị thông tin dạng biểu đồ



Bước 4: Lựa chọn chế độ mô phỏng và thực hiện mô phỏng



Kết quả mô phỏng dạng Graphic



Kết quả mô phỏng dạng Text

SMPCache - Cache evolution

Information View cache P0-P0

Number of blocks: 64 Cache of node: P0
Mapping: Fully-Associative Cache level: 1
Replacement requ: LRU Cache type: Unified Cache address: 0
Cache sets: 0

Address	State	Block	Last Access
0000 (0)	Exclusive (E)	2007	1853
0001 (1)	Exclusive (E)	2031	1692
0002 (2)	Exclusive (E)	1462	1849
0003 (3)	Exclusive (E)	3109	1387
0004 (4)	Exclusive (E)	1463	1854
0005 (5)	Exclusive (E)	3104	1360
0006 (6)	Exclusive (E)	3108	1384
0007 (7)	Exclusive (E)	2020	1428

Actual access

Access number: 1855
Access type: Data reading
Address: 000003d5 (Hex)
Block: 61
Word: 5

Events visor Node inspector Hits-Misses Bus States Configuration

Hits-Misses

Global readings

	Hits	Misses
Number	1376	209
Rate	86.814%	13.186%

Global writings

	Hits	Misses
Number	181	89
Rate	67.037%	32.963%

Global hits and misses

	Hits	Misses
Number	1557	298
Rate	83.935%	16.065%

Total replacements: 234

Local readings

	Hits	Misses
Number	1376	209
Rate	86.814%	13.186%

Local writings

	Hits	Misses
Number	181	89
Rate	67.037%	32.963%

Local hits and misses

	Hits	Misses
Number	1557	298
Rate	83.935%	16.065%

Cache level: 1

Data concerning cache 0

	Actual	Total
Accesses	1855	1855
Instructions	1439	1439
Data readings	146	146
Data writings	270	270

Final writebacks

Execute Continue Stop

1 Complete

Exit

Simulation steps: 1855

18

4. Thực hành

- Lab 2.1: Kiểm tra tính cục bộ (lân cận) của các chương trình
- Lab 2.2: Khảo sát ảnh hưởng của kích thước Cache đến hiệu suất hệ thống
- Lab 2.3: Khảo sát ảnh hưởng của kích thước Block đến hiệu suất hệ thống

Lab 2.1: Kiểm tra tính cục bộ (lân cận) của các chương trình khác nhau

1. Cấu hình một hệ thống với các đặc tính kiến trúc như sau:

- Số lượng bộ xử lý - Processors in SMP = 1.
- Giao thức liên kết cache - Cache coherence protocol = MESI.
- Phương pháp phân xử Bus - Scheme for bus arbitration = Random.
- Kích thước word - Word wide (bits) = 16.
- Số word trong một block - Words by block = 16 (block size = 32 bytes).
- Số block trong bộ nhớ chính - Blocks in main memory = 8192. **Kích thước bộ nhớ chính bằng bao nhiêu?**
- **Số blocks trong cache – Blocks in cache: 32**
- Ánh xạ - Mapping = Fully-Associative.
- Chính sách thay thế - Replacement policy = LRU

Lab 2.1: Kiểm tra tính cục bộ (lân cận) của các chương trình khác nhau

2. Quan sát miss rate (tỷ lệ truy xuất cache không thành công) với các memory traces: Hydro, Nasa7, Cexp, Mdljd, Ear, Comp, Wave, Swm và Ucomp và ghi vào file MS Excel.
3. Theo số liệu quan sát được, trả lời các câu hỏi:
 - 1) Liệu các chương trình có cùng mức độ cục bộ? Chương trình nào có tính cục bộ tốt nhất? Chương trình nào có tính cục bộ kém nhất?
 - 2) Liệu cấu hình trên có khai thác được tính cục bộ của các chương trình này và tăng hiệu suất hệ thống không? Tại sao?
 - 3) Trong quá trình chương trình thực hiện, nếu quan sát dưới dạng đồ thị, tỷ lệ miss rate giảm dần, nguyên nhân tại sao?

Memory Traces	Miss Rate (%)
Hydro	18.195
Nasa7	
Cexp	
Ear	
Comp	
Sum	
Ucomp	

Lab 2.1: Kiểm tra tính cục bộ (lân cận) của các chương trình khác nhau

- 1) Các chương trình không có cùng mức cục bộ
 - a. Cexp có mức độ cục bộ tốt nhất
 - b. Swm có mức độ cục bộ kém nhất
- 2) Cấu hình trên đã khai thác được mức độ cục bộ của các chương trình, tuy nhiên hiệu suất chưa cao (miss rate > 10% với hầu hết các chương trình)
- 3) Trong quá trình một chương trình chạy, miss rate sẽ giảm dần vì:

Memory Traces	Miss Rate (%)
Hydro	18.195
Nasa7	17.035
Cexp	0.77
Ear	16.993
Comp	18.661
Swm	21.985
Ucomp	17.892

- a. Khi bắt đầu chạy chương trình, bộ nhớ cache chưa có dữ liệu. Các truy xuất bộ nhớ của CPU phải truy xuất đến RAM (cache miss) → Miss rate lớn
- b. Sau khi các khối được ánh xạ nhiều trong bộ nhớ → Do tính cục bộ của chương trình, các câu lệnh và dữ liệu có thể được truy xuất trực tiếp từ Cache (cache hit) → Miss rate giảm dần.

Lab 2.2: Khảo sát ảnh hưởng của kích thước Cache đến hiệu suất hệ thống

1. Thiết lập cấu hình của hệ thống với các đặc tính kiến trúc sau:

- Số lượng bộ xử lý - Processors in SMP = 1.
- Giao thức liên kết cache - Cache coherence protocol = MESI.
- Phương pháp phân xử Bus - Scheme for bus arbitration = Random.
- Kích thước word - Word wide (bits) = 16.
- Số word trong một block - Words by block = 16 (block size = 32 bytes).
- Số block trong bộ nhớ chính - Blocks in main memory = 8192.
- Ánh xạ - Mapping = Fully-Associative.
- Chính sách thay thế - Replacement policy = LRU.
- Thiết lập cache với kích thước như sau (số block trong cache): 1, 2, 4, 8, 16, 32, 64, 128, 256 và 512.

Lab 2.2: Khảo sát ảnh hưởng của kích thước Cache đến hiệu suất hệ thống

2. Với mỗi thông số trên, thiết lập các memory traces: *Hydro*, *Nasa7*, *Cexp*, *Mdljd*, *Ear*, *Comp*, *Wave*, *Swm* và *Ucomp*, chạy mô phỏng và ghi lại các miss rate vào file MS Excel.
3. Phân tích kết quả
Vẽ biểu đồ thể hiện mối quan hệ giữa miss rate và cache size cho các memory traces và trả lời câu hỏi sau:
 - Miss rate như thế nào khi kích thước cache tăng?
 - Liệu việc tăng kích thước cache có giúp cải thiện hiệu suất của hệ thống không?

	Cache size (số block trong cache)											
	1	2	4	8	16	32	64	128	256	512		
Hydro	66.197%	42.360%	32.111%									
Nasa7												
Cexp												
Mdljd												
Ear												
Comp												
Wave												
Swm												
Ucomp												

Lab 2.3: Khảo sát ảnh hưởng của kích thước Block đến hiệu suất hệ thống

1. Thiết lập cấu hình của hệ thống với các đặc tính kiến trúc sau:

- Số lượng bộ xử lý - Processors in SMP = 1.
- Giao thức liên kết cache - Cache coherence protocol = MESI.
- Phương pháp phân xử Bus - Scheme for bus arbitration = Random.
- Kích thước word - Word wide (bits) = 16.
- Kích thước bộ nhớ RAM = 256 KB (số lượng block trong bộ nhớ sẽ thay đổi)
- Kích thước cache = 4 KB (số line –block trong cace thay đổi)
- Ánh xạ - Mapping = Fully-Associative.
- Chính sách thay thế - Replacement policy = LRU.
- Cấu hình kích thước Block (số word/1 block) như sau: 4, 8, 16, 32, 64, 128, 256, 512, 1024.

Lab 2.3: Khảo sát ảnh hưởng của kích thước Block đến hiệu suất hệ thống

2. Với mỗi cấu hình trên, gọi các memory traces: *Hydro*, *Nasa7*, *Cexp*, *Mdljd*, *Ear*, *Comp*, *Wave*, *Swm* và *Ucomp*, xác định miss rate của chúng và ghi kết quả vào file MS Excel.
3. Phân tích kết quả

Vẽ biểu đồ thể hiện mối quan hệ giữa miss rate và block size cho các memory traces và trả lời câu hỏi sau:

- Nếu kích thước block tăng thì miss rate tăng hay giảm? Tại sao?
- Kích thước Block nào cho hiệu suất hệ thống cao nhất đối với các chương trình trên.

	Block size (số word/1 block)								
	4	8	16	32	64	128	256	512	1024
Hydro	43.112	24.307	14.81						
Nasa7									
Cexp									
Mdljd									
Ear									
Comp									
Wave									
Swm									
Ucomp									

Viết báo cáo

- Báo cáo các bài thực hành Lab 2.1, Lab 2.2 và Lab 2.3 theo mẫu file *Thuchanh_Cache_Phieu ghi ket qua.docx*