Cache Memory Design and Analysis Report

Mohit Jagini IMT2023528
Bojja Sunhith Reddy IMT2023113
P Venkata Rami Reddy IMT2023102

September 6, 2024

Contents

1   Introduction

This report presents the design and analysis of a 4-way set associative cache of size 1024 kilobytes with a block size of 4 bytes. The assignment aims to study the cache performance by varying parameters such as cache size, block size, and associativity.

2   Design and Implementation

2.1   Cache Design Details

The cache is structured as a **4-way set associative** cache, meaning that each set contains 4 cache lines (or ways). The cache size is 1024 KB, and each cache block (or cache line) is 4 bytes in size. The total number of sets in the cache is determined by dividing the total cache size by the product of block size and the number of ways.

- **Cache Size**: 1024 KB

- **Block Size**: 4 bytes

- **Associativity**: 4-way set associative

- **Number of Lines per Set**: 4

- **Number of Sets**: 65,536 sets

The number of sets is calculated using the formula:

$$\frac{\text{Cache Size}}{\text{Block Size} \times \text{Associativity}} = \frac{1024 \times 1024}{4 \times 4} = 65536 \, \text{sets}$$

Each memory address is 32 bits wide and divided into three parts:

- **Tag**: Identifies whether the requested block is present in the cache set. This is derived from the higher bits of the address.

- **Index**: Selects the specific cache set from which to check for the block. The number of bits needed for the index is determined by the number of sets (16 bits, since there are 65,536 sets).

In this design, the cache checks for a tag match and valid bit, as the assignment does not involve data fetching from the cache. The operation is simplified to reporting cache hits or misses based on the tag comparison within the selected set.

2.2   Implementation Steps

The implementation of the 4-way set associative cache involved the following steps:

1. **Initialization of Cache Structure:**

    - Defined the cache as a 2D list where each row represents a set, and each set contains 4 lines (or ways).
    - Each cache line stores a tag, valid bit, and other metadata required for hit/miss reporting.

2. **Address Breakdown:**

   - For each memory access, the 32-bit address is broken down into three parts:
     - **Tag**: The upper bits of the address are used to identify if the requested block is in the cache.
     - **Index**: The middle bits of the address used to select the cache set.

3. **Processing Input Trace:**

   - Read the input trace file containing a sequence of memory addresses.
   - For each address, the index is used to access the corresponding cache set, and the tag is compared with the tags in the set's cache lines.
   - If a matching tag with a valid bit is found, it results in a cache hit; otherwise, it is a cache miss.

4. **Cache Replacement Policy:**

   - Implemented the Least Recently Used (LRU) policy to handle cache line replacement when a miss occurs, ensuring that the least recently accessed block is evicted.

5. **Recording Hits and Misses:**

   - For each memory access, the result (hit or miss) was recorded and aggregated for further analysis.

6. **Varying Cache Parameters:**

   - The cache size, block size, and associativity were varied to conduct experiments and compare hit/miss rates across different configurations.
   - Output graphs were generated to visualize the performance impact of these changes.

7. **Final Observations:**

   - Analyzed the experimental data to derive insights into the performance of different cache configurations, focusing on the trade-offs between cache size, block size, and associativity.

3 Experiments and Results

This section discusses the experiments conducted to evaluate the performance of different cache configurations by varying cache size, block size, and associativity. Each experiment focuses on how these parameters influence the hit and miss rates, which are crucial for understanding cache efficiency.

3.1 Experiment A: Cache Size Variation

In this experiment, the cache size is varied while keeping the block size (4 bytes) and associativity (4-way set associative) constant. The goal is to observe how cache size impacts the cache performance, specifically the miss rate.

3.1.1 Experimental Setup

The following cache sizes were tested:

- 128 KB

- 256 KB

- 512 KB

- 1024 KB

- 2048 KB

- 4096 KB

For each cache size, the same memory access trace file was processed, recording the number of cache hits and misses.

3.1.2 Results

The results of the experiment are summarized in the excel

3.1.3 Discussion

The results indicate that increasing the cache size leads to a reduction in miss rate, as expected. However, after a certain point (1024 KB), the miss rate plateaus and shows no significant improvement with further increases in cache size. This suggests that the working set of the program fits well within a 1024 KB cache, and additional cache size does not contribute to performance gains.

3.1.4 Conclusion

This experiment demonstrates that while increasing cache size can improve performance by reducing the miss rate, there is a point of diminishing returns. A cache size of 1024 KB is sufficient for this workload, as larger sizes do not provide additional benefits. This highlights the importance of selecting an optimal cache size that balances performance improvements with hardware costs.

3.2 Experiment B: Block Size Variation

In this experiment, the block size is varied while keeping the cache size (1024 KB) and associativity (4-way set associative) constant. The block size influences how much data is fetched from memory in a single cache line, and larger block sizes can improve performance by reducing the number of memory accesses for consecutive addresses.

3.2.1 Experimental Setup

The following block sizes were tested:

- 4 bytes

- 8 bytes

- 16 bytes

- 32 bytes

- 64 bytes

- 128 bytes

Each configuration processes the same memory trace as in Experiment A, recording the hit and miss rates for different block sizes.

3.2.2 Results

As the block size increases, the miss rate generally decreases, as a larger block can bring in more consecutive data. However, very large block sizes may result in wasted cache space if the program exhibits low spatial locality.

3.2.3 Discussion

Increasing the block size improves the miss rate, but as the block size increases beyond a certain point (e.g., 16 bytes), the performance gains diminish. Larger blocks can increase the chances of loading unnecessary data, leading to inefficient cache usage, especially if the memory access pattern has low spatial locality.

3.2.4 Conclusion

This experiment demonstrates that while increasing block size generally improves the cache miss rate, excessively large blocks can lead to diminishing returns. An optimal block size should balance between reducing the miss rate and avoiding the unnecessary fetching of unrelated data.

3.3 Experiment C: Associativity Variation

This experiment focuses on varying the associativity of the cache while keeping the cache size (1024 KB) and block size (4 bytes) constant. Associativity determines how many cache lines can reside in a single set, affecting the flexibility of where a block can be stored in the cache.

3.3.1 Experimental Setup

The associativity was varied as follows:

- 1-way (Direct Mapped)

- 2-way set associative

- 4-way set associative

- 8-way set associative

- 16-way set associative

- 32-way set associative

- 64-way set associative

3.3.2 Results

As the associativity increases, the miss rate decreases due to fewer cache evictions. A higher associativity allows for more flexibility in placing data within the cache, which reduces the chance of conflict misses.

3.3.3 Discussion

Higher associativity generally leads to better cache performance, as fewer evictions occur due to the flexibility in placing blocks. However, increasing associativity beyond a certain point (e.g., 4-way) results in diminishing returns, as conflict misses become less common. Additionally, higher associativity increases the hardware complexity and cost.

3.3.4 Conclusion

Varying associativity shows that increasing it improves cache performance by reducing conflict misses. However, beyond a certain point, the performance improvement is minimal, and the trade-off between complexity and performance needs to be considered.

3.4 Graphs and Tables

The Excel sheet containing all experimental data and results can be found at the following link: Download Excel Sheet

3.5 Output Snapshots


```
----------------gcc.trace-------------------
Part (a) - Cache Size: 1024 KB, Block Size: 4 Bytes, Associativity: 4-way
Hits: 483894, Misses: 31789, Hit Rate: 93.83555401283347, Miss Rate: 6.164445987166534
Part (b) - Varying Cache Size
Cache Size: 128 KB, Hits: 483719, Misses: 31964, Hit Rate: 93.80161843613227, Miss Rate: 6.198381563867724
Cache Size: 256 KB, Hits: 483871, Misses: 31812, Hit Rate: 93.83109390846703, Miss Rate: 6.1689060915329765
Cache Size: 512 KB, Hits: 483893, Misses: 31790, Hit Rate: 93.83536009525231, Miss Rate: 6.164639904747684
Cache Size: 1024 KB, Hits: 483894, Misses: 31789, Hit Rate: 93.83555401283347, Miss Rate: 6.164445987166534
Cache Size: 2048 KB, Hits: 483894, Misses: 31789, Hit Rate: 93.83555401283347, Miss Rate: 6.164445987166534
Cache Size: 4096 KB, Hits: 483894, Misses: 31789, Hit Rate: 93.83555401283347, Miss Rate: 6.164445987166534
Part (c) - Varying Block Size
Block Size: 1 Bytes, Hits: 480611, Misses: 35072, Hit Rate: 93.19892259391914, Miss Rate: 6.801077406080867
Block Size: 2 Bytes, Hits: 482807, Misses: 32876, Hit Rate: 93.62476560212379, Miss Rate: 6.3752343978762145
Block Size: 4 Bytes, Hits: 483894, Misses: 31789, Hit Rate: 93.83555401283347, Miss Rate: 6.164445987166534
Block Size: 8 Bytes, Hits: 494677, Misses: 21006, Hit Rate: 95.92656729037024, Miss Rate: 4.073432709629754
Block Size: 16 Bytes, Hits: 504467, Misses: 11216, Hit Rate: 97.82502040982541, Miss Rate: 2.174979590174584
Block Size: 32 Bytes, Hits: 509644, Misses: 6039, Hit Rate: 98.82893172743721, Miss Rate: 1.1710682725627954
Block Size: 64 Bytes, Hits: 512310, Misses: 3373, Hit Rate: 99.3459159987822, Miss Rate: 0.6540840012178024
Block Size: 128 Bytes, Hits: 513728, Misses: 1955, Hit Rate: 99.62089112885242, Miss Rate: 0.37910887114758485
Part (d) - Varying Associativity
Associativity: 1-way, Hits: 483868, Misses: 31815, Hit Rate: 93.83051215572358, Miss Rate: 6.169487844276426
Associativity: 2-way, Hits: 483890, Misses: 31793, Hit Rate: 93.83477834250886, Miss Rate: 6.165221657491133
Associativity: 4-way, Hits: 483894, Misses: 31789, Hit Rate: 93.83555401283347, Miss Rate: 6.164445987166534
Associativity: 8-way, Hits: 483894, Misses: 31789, Hit Rate: 93.83555401283347, Miss Rate: 6.164445987166534
Associativity: 16-way, Hits: 483895, Misses: 31788, Hit Rate: 93.83574793041461, Miss Rate: 6.164252069585385
Associativity: 32-way, Hits: 483896, Misses: 31787, Hit Rate: 93.83594184799577, Miss Rate: 6.164058152004235
Associativity: 64-way, Hits: 483896, Misses: 31787, Hit Rate: 93.83594184799577, Miss Rate: 6.164058152004235
```

Figure 1: This is the output for gcc.trace

```
----------------swim.trace--------------------
Part (a) - Cache Size: 1024 KB, Block Size: 4 Bytes, Associativity: 4-way
Hits: 280825, Misses: 22368, Hit Rate: 92.62252096849201, Miss Rate: 7.37747903150798
Part (b) - Varying Cache Size
Cache Size: 128 KB, Hits: 280817, Misses: 22376, Hit Rate: 92.6198823851474, Miss Rate: 7.380117614852585
Cache Size: 256 KB, Hits: 280825, Misses: 22368, Hit Rate: 92.62252096849201, Miss Rate: 7.377479031507983
Cache Size: 512 KB, Hits: 280825, Misses: 22368, Hit Rate: 92.62252096849201, Miss Rate: 7.377479031507983
Cache Size: 1024 KB, Hits: 280825, Misses: 22368, Hit Rate: 92.62252096849201, Miss Rate: 7.377479031507983
Cache Size: 2048 KB, Hits: 280825, Misses: 22368, Hit Rate: 92.62252096849201, Miss Rate: 7.377479031507983
Cache Size: 4096 KB, Hits: 280825, Misses: 22368, Hit Rate: 92.62252096849201, Miss Rate: 7.377479031507983
Part (c) - Varying Block Size
Block Size: 1 Bytes, Hits: 280588, Misses: 22605, Hit Rate: 92.54435293690817, Miss Rate: 7.455647063091826
Block Size: 2 Bytes, Hits: 280737, Misses: 22456, Hit Rate: 92.59349655170139, Miss Rate: 7.406503448298609
Block Size: 4 Bytes, Hits: 280825, Misses: 22368, Hit Rate: 92.62252096849201, Miss Rate: 7.377479031507983
Block Size: 8 Bytes, Hits: 283377, Misses: 19816, Hit Rate: 93.46422905542015, Miss Rate: 6.535770944579855
Block Size: 16 Bytes, Hits: 291770, Misses: 11423, Hit Rate: 96.23243280682601, Miss Rate: 3.7675671931739845
Block Size: 32 Bytes, Hits: 296797, Misses: 6396, Hit Rate: 97.89045261599047, Miss Rate: 2.1095473840095256
Block Size: 64 Bytes, Hits: 299740, Misses: 3453, Hit Rate: 98.86112146388604, Miss Rate: 1.1388785361139604
Block Size: 128 Bytes, Hits: 301367, Misses: 1826, Hit Rate: 99.39774335159453, Miss Rate: 0.6022566484054711
Part (d) - Varying Associativity
Associativity: 1-way, Hits: 280819, Misses: 22374, Hit Rate: 92.62054203098357, Miss Rate: 7.3794579690164355
Associativity: 2-way, Hits: 280825, Misses: 22368, Hit Rate: 92.62252096849201, Miss Rate: 7.377479031507983
Associativity: 4-way, Hits: 280825, Misses: 22368, Hit Rate: 92.62252096849201, Miss Rate: 7.377479031507983
Associativity: 8-way, Hits: 280825, Misses: 22368, Hit Rate: 92.62252096849201, Miss Rate: 7.377479031507983
Associativity: 16-way, Hits: 280825, Misses: 22368, Hit Rate: 92.62252096849201, Miss Rate: 7.377479031507983
Associativity: 32-way, Hits: 280825, Misses: 22368, Hit Rate: 92.62252096849201, Miss Rate: 7.377479031507983
Associativity: 64-way, Hits: 280825, Misses: 22368, Hit Rate: 92.62252096849201, Miss Rate: 7.377479031507983
```

Figure 2: This is the output for swim.trace

```
----------------mcf.trace--------------------
Part (a) - Cache Size: 1024 KB, Block Size: 4 Bytes, Associativity: 4-way
Hits: 7508, Misses: 719722, Hit Rate: 1.03241065412593, Miss Rate: 98.96758934587407
Part (b) - Varying Cache Size
Cache Size: 128 KB, Hits: 7508, Misses: 719722, Hit Rate: 1.03241065412593, Miss Rate: 98.96758934587407
Cache Size: 256 KB, Hits: 7508, Misses: 719722, Hit Rate: 1.03241065412593, Miss Rate: 98.96758934587407
Cache Size: 512 KB, Hits: 7508, Misses: 719722, Hit Rate: 1.03241065412593, Miss Rate: 98.96758934587407
Cache Size: 1024 KB, Hits: 7508, Misses: 719722, Hit Rate: 1.03241065412593, Miss Rate: 98.96758934587407
Cache Size: 2048 KB, Hits: 7508, Misses: 719722, Hit Rate: 1.03241065412593, Miss Rate: 98.96758934587407
Cache Size: 4096 KB, Hits: 7603, Misses: 719627, Hit Rate: 1.0454739215928937, Miss Rate: 98.9545260784071
Part (c) - Varying Block Size
Block Size: 1 Bytes, Hits: 7451, Misses: 719779, Hit Rate: 1.0245726936457518, Miss Rate: 98.97542730635425
Block Size: 2 Bytes, Hits: 7481, Misses: 719749, Hit Rate: 1.0286979360037403, Miss Rate: 98.97130206399626
Block Size: 4 Bytes, Hits: 7508, Misses: 719722, Hit Rate: 1.03241065412593, Miss Rate: 98.96758934587407
Block Size: 8 Bytes, Hits: 7551, Misses: 719679, Hit Rate: 1.0383235015057135, Miss Rate: 98.96167649849428
Block Size: 16 Bytes, Hits: 367273, Misses: 359957, Hit Rate: 50.50300455151741, Miss Rate: 49.49699544848260
Block Size: 32 Bytes, Hits: 547152, Misses: 180078, Hit Rate: 75.23782022193804, Miss Rate: 24.762179778061
Block Size: 64 Bytes, Hits: 637112, Misses: 90118, Hit Rate: 87.60804697275965, Miss Rate: 12.3919530272403
Block Size: 128 Bytes, Hits: 682109, Misses: 45121, Hit Rate: 93.79549798550664, Miss Rate: 6.20450201449335
Part (d) - Varying Associativity
Associativity: 1-way, Hits: 7505, Misses: 719725, Hit Rate: 1.031998129890131, Miss Rate: 98.96800187010987
Associativity: 2-way, Hits: 7507, Misses: 719723, Hit Rate: 1.0322731460473304, Miss Rate: 98.9677268539526
Associativity: 4-way, Hits: 7508, Misses: 719722, Hit Rate: 1.03241065412593, Miss Rate: 98.96758934587407
Associativity: 8-way, Hits: 7508, Misses: 719722, Hit Rate: 1.03241065412593, Miss Rate: 98.96758934587407
Associativity: 16-way, Hits: 7508, Misses: 719722, Hit Rate: 1.03241065412593, Miss Rate: 98.96758934587407
Associativity: 32-way, Hits: 7508, Misses: 719722, Hit Rate: 1.03241065412593, Miss Rate: 98.96758934587407
Associativity: 64-way, Hits: 7508, Misses: 719722, Hit Rate: 1.03241065412593, Miss Rate: 98.96758934587407
```

Figure 3: This is the output for mcf.trace

```
----------------gzip.trace--------------------
Part (a) - Cache Size: 1024 KB, Block Size: 4 Bytes, Associativity: 4-way
Hits: 320883, Misses: 160161, Hit Rate: 66.70554044952229, Miss Rate: 33.29445955047771
Part (b) - Varying Cache Size
Cache Size: 128 KB, Hits: 320883, Misses: 160161, Hit Rate: 66.70554044952229, Miss Rate: 33.29445955047771
Cache Size: 256 KB, Hits: 320883, Misses: 160161, Hit Rate: 66.70554044952229, Miss Rate: 33.29445955047771
Cache Size: 512 KB, Hits: 320883, Misses: 160161, Hit Rate: 66.70554044952229, Miss Rate: 33.29445955047771
Cache Size: 1024 KB, Hits: 320883, Misses: 160161, Hit Rate: 66.70554044952229, Miss Rate: 33.29445955047771
Cache Size: 2048 KB, Hits: 320883, Misses: 160161, Hit Rate: 66.70554044952229, Miss Rate: 33.29445955047771
Cache Size: 4096 KB, Hits: 320883, Misses: 160161, Hit Rate: 66.70554044952229, Miss Rate: 33.29445955047771
Part (c) - Varying Block Size
Block Size: 1 Bytes, Hits: 320875, Misses: 160169, Hit Rate: 66.70387739998836, Miss Rate: 33.29612260001164
Block Size: 2 Bytes, Hits: 320876, Misses: 160168, Hit Rate: 66.7040852811801, Miss Rate: 33.2959147188199
Block Size: 4 Bytes, Hits: 320883, Misses: 160161, Hit Rate: 66.70554044952229, Miss Rate: 33.29445955047771
Block Size: 8 Bytes, Hits: 320891, Misses: 160153, Hit Rate: 66.70720349905622, Miss Rate: 33.29279650094378
Block Size: 16 Bytes, Hits: 321268, Misses: 159776, Hit Rate: 66.7855747083427, Miss Rate: 33.214425291657314
Block Size: 32 Bytes, Hits: 321459, Misses: 159585, Hit Rate: 66.82528001596528, Miss Rate: 33.17471998403472
Block Size: 64 Bytes, Hits: 321559, Misses: 159485, Hit Rate: 66.8460681351394, Miss Rate: 33.153931864860596
Block Size: 128 Bytes, Hits: 321609, Misses: 159435, Hit Rate: 66.85646219472648, Miss Rate: 33.14353780527353
Part (d) - Varying Associativity
Associativity: 1-way, Hits: 320883, Misses: 160161, Hit Rate: 66.70554044952229, Miss Rate: 33.29445955047771
Associativity: 2-way, Hits: 320883, Misses: 160161, Hit Rate: 66.70554044952229, Miss Rate: 33.29445955047771
Associativity: 4-way, Hits: 320883, Misses: 160161, Hit Rate: 66.70554044952229, Miss Rate: 33.29445955047771
Associativity: 8-way, Hits: 320883, Misses: 160161, Hit Rate: 66.70554044952229, Miss Rate: 33.29445955047771
Associativity: 16-way, Hits: 320883, Misses: 160161, Hit Rate: 66.70554044952229, Miss Rate: 33.29445955047771
Associativity: 32-way, Hits: 320883, Misses: 160161, Hit Rate: 66.70554044952229, Miss Rate: 33.29445955047771
Associativity: 64-way, Hits: 320883, Misses: 160161, Hit Rate: 66.70554044952229, Miss Rate: 33.29445955047771
```

Figure 4: This is the output for gzip.trace

```
-----------------twolf.trace--------------------
Part (a) - Cache Size: 1024 KB, Block Size: 4 Bytes, Associativity: 4-way
Hits: 476844, Misses: 5980, Hit Rate: 98.76145344887578, Miss Rate: 1.238546551124219
Part (b) - Varying Cache Size
Cache Size: 128 KB, Hits: 476843, Misses: 5981, Hit Rate: 98.76124633406789, Miss Rate: 1.2387536659320995
Cache Size: 256 KB, Hits: 476844, Misses: 5980, Hit Rate: 98.76145344887578, Miss Rate: 1.238546551124219
Cache Size: 512 KB, Hits: 476844, Misses: 5980, Hit Rate: 98.76145344887578, Miss Rate: 1.238546551124219
Cache Size: 1024 KB, Hits: 476844, Misses: 5980, Hit Rate: 98.76145344887578, Miss Rate: 1.238546551124219
Cache Size: 2048 KB, Hits: 476844, Misses: 5980, Hit Rate: 98.76145344887578, Miss Rate: 1.238546551124219
Cache Size: 4096 KB, Hits: 476844, Misses: 5980, Hit Rate: 98.76145344887578, Miss Rate: 1.238546551124219
Part (c) - Varying Block Size
Block Size: 1 Bytes, Hits: 475470, Misses: 7354, Hit Rate: 98.47687770284824, Miss Rate: 1.5231222971517573
Block Size: 2 Bytes, Hits: 476358, Misses: 6466, Hit Rate: 98.66079565224595, Miss Rate: 1.339204347754047
Block Size: 4 Bytes, Hits: 476844, Misses: 5980, Hit Rate: 98.76145344887578, Miss Rate: 1.238546551124219
Block Size: 8 Bytes, Hits: 477319, Misses: 5505, Hit Rate: 98.85983298261893, Miss Rate: 1.1401670173810747
Block Size: 16 Bytes, Hits: 479869, Misses: 2955, Hit Rate: 99.38797574271369, Miss Rate: 0.612024257286299
Block Size: 32 Bytes, Hits: 481182, Misses: 1642, Hit Rate: 99.65991748546054, Miss Rate: 0.3400825145394595
Block Size: 64 Bytes, Hits: 481870, Misses: 954, Hit Rate: 99.80241247328219, Miss Rate: 0.19758752671781024
Block Size: 128 Bytes, Hits: 482249, Misses: 575, Hit Rate: 99.88090898546884, Miss Rate: 0.11909101453117493
Part (d) - Varying Associativity
Associativity: 1-way, Hits: 476771, Misses: 6053, Hit Rate: 98.74633406790052, Miss Rate: 1.2536659320994814
Associativity: 2-way, Hits: 476841, Misses: 5983, Hit Rate: 98.76083210445215, Miss Rate: 1.23916789554786
Associativity: 4-way, Hits: 476844, Misses: 5980, Hit Rate: 98.76145344887578, Miss Rate: 1.238546551124219
Associativity: 8-way, Hits: 476844, Misses: 5980, Hit Rate: 98.76145344887578, Miss Rate: 1.238546551124219
Associativity: 16-way, Hits: 476844, Misses: 5980, Hit Rate: 98.76145344887578, Miss Rate: 1.238546551124219
Associativity: 32-way, Hits: 476844, Misses: 5980, Hit Rate: 98.76145344887578, Miss Rate: 1.238546551124219
Associativity: 64-way, Hits: 476844, Misses: 5980, Hit Rate: 98.76145344887578, Miss Rate: 1.238546551124219
```

Figure 5: This is the output for twolf.trace