

Exercise walk-through

6.4

Memory hierarchy

University Computer Science

Memory hierarchy: split the address, price the miss

- 1 **Split** 4 KiB, 16-byte lines: 256 lines, so 8 **index** bits, 4 offset, 20 **tag**; 0x4A8 lands at offset 8, line 74.
- 2 **AMAT** $1 + \text{miss ratio} \times \text{penalty}$: at 2 percent and 100 cycles, 3 cycles; halve the misses to reach 2.
- 3 **Layout** Row-major over 16-byte lines misses 1 in 4; stride-16 misses every access - layout is a performance feature.

Address decomposition and hierarchy

a 32-bit address split by the cache geometry



4 KiB cache, 16-byte lines: 256 lines - so 8 index bits, 4 offset bits, 20 tag bits

0x4A8: offset 8 (byte 8 of the line), index 74 (line 74 of 256)



a miss at one level is a request at the next: $AMAT = hit + miss\ ratio \times penalty$

The 32-bit address splits by the cache geometry into **tag**, **index** and offset.

Exercises 2.1-2.3

2.1 An 8 KiB direct-mapped cache with 16-byte lines: give the index and offset widths. [3]

2.2 Hit 1 cycle, miss ratio 5%, penalty 40 cycles: give the AMAT. [3]

2.3 Name the two localities a hierarchy exploits. [3]

Exercises 2.1-2.3

2.1 An 8 KiB direct-mapped cache with 16-byte lines: give the index and offset widths. [3]

2.1

- $2^{13}/2^4 = 512$ lines: index 9 bits, offset 4 bits ✓.
-

2.2 Hit 1 cycle, miss ratio 5%, penalty 40 cycles: give the AMAT. [3]

2.2

- $1 + 0.05 \times 40 = 3$ cycles ✓.
-

2.3 Name the two localities a hierarchy exploits. [3]

2.3

- Temporal (reuse) and spatial (adjacency) ✓.

Exercise 3.1

3.1 A direct-mapped cache holds 2 KiB in 32-byte lines on a 32-bit address space.

Compute the line count, the offset, index and tag widths, and the slot of address $0x7CE$ (offset, index). State what happens when two hot blocks collide in one slot, and name the miss class. Compute AMAT for hit 1 cycle, miss ratio 4%, penalty 50 cycles. [8]

Exercise 3.1

3.1 A direct-mapped cache holds 2 KiB in 32-byte lines on a 32-bit address space.

Compute the line count, the offset, index and tag widths, and the slot of address $0x7CE$ (offset, index). State what happens when two hot blocks collide in one slot, and name the miss class. Compute AMAT for hit 1 cycle, miss ratio 4%, penalty 50 cycles. [8]

3.1

- $2^{11}/32 = 64$ lines: offset 5 bits, index 6 bits, tag $32 - 11 = 21$ bits ✓.
- $0x7CE = 0111\ 1100\ 1110_2$ (low 11 bits): offset $01110_2 = 14$; index $111110_2 = 62$ – line 62 of 64 ✓.
- Two hot blocks in one slot evict each other turn by turn – a conflict miss ✓.
- $AMAT = 1 + 0.04 \times 50 = 3$ cycles ✓.

Exercise 3.2

3.2 Two scans of a 128×128 int array (16-byte lines, 4 ints per line, int = 4 bytes).

Scan A visits elements in row-major order; scan B visits column-major. Compute each scan's miss rate on compulsory grounds and state the AMAT multiplier between them (hit 1 cycle, penalty 40 cycles, miss ratio as computed). Explain why a bigger cache does not close the gap. [7]

Exercise 3.2

3.2 Two scans of a 128×128 int array (16-byte lines, 4 ints per line, int = 4 bytes).

Scan A visits elements in row-major order; scan B visits column-major. Compute each scan's miss rate on compulsory grounds and state the AMAT multiplier between them (hit 1 cycle, penalty 40 cycles, miss ratio as computed). Explain why a bigger cache does not close the gap. [7]

3.2

- A (row-major): one miss per line – 4 ints per line – miss rate $1/4 = 25\%$ (compulsory only) ✓.
- B (column-major): successive elements sit $128 \times 4 = 512$ bytes apart – one per line, miss rate 100% ✓.
- AMATs: $A = 1 + 0.25 \times 40 = 11$ cycles; $B = 1 + 1.0 \times 40 = 41$ cycles – a $3.7\times$ multiplier ✓.
- A bigger cache holds more lines but the stride still touches one int per line: bandwidth per line, not capacity, is the wall ✓.