Lab 11

128 MiB MM
64 KiB cache
byte addressable
1 word = 32 bits

i) DM cache
blk size 4 words WT

ii) 4 way SA
blk 2 words WB, LRU

a) index, tag, offset
b) total cache size
c) 0x3, 0x84, 0xAC

Cache Performance

$$CPU_{time\ ideal} = IC \times CPI_{ideal} \times CCT$$

$$CPI_{real} = CPI_{ideal} + \underbrace{\text{Mem. Stall Cycles}}_{\text{(only misses)}}$$

Read Stall Cycles = Read Miss Penalty × MM access time
× Read Miss Rate ×
× Reads / program

Write Stall Cycles = —11— + write buffer stall cycles

Mem Stall Cycles = Mem. Acc × MR × MP

IC -> MR = 2%

DC -> MR = 4%

36% of instr are L/S

CPI ideal = 2

MP = 100 cc

a) By how much faster -> perfect memory?

b) —11— CPI ideal to 1

$$\frac{CPI_{real}}{CPI_{ideal}} = \frac{CPI_{ideal} + (0.36 \times 0.04 \times 100) + (1 \times 0.02 \times 100)}{CPI_{ideal}}$$

(0.36 × 0.04 × 100): IC

(1 × 0.02 × 100): IC

$$= 2,72$$

b) $CPI_{ideal} = 1$

$$\frac{1 + 1,44 + 2}{1} = 4,44$$

$$AMAT = \underbrace{\text{Hit Time}}_{\text{L1 access}} + MR \times MP$$

$CPI_{ideal} = 1$

$CR = 4\ GHz$

MM access time = 100 ns

$MR_{L1} = 2\%$

a) How much faster would the processor be if we added a L2 cache with access time 5 ns

$MR_{L2} = 0.5\%$

b) $AMAT_{L1}$, $AMAT_{L2}$ = ?

HW 5.12 RISC-V EA.

5.12 Multilevel caching is an important technique to overcome the limited amount of space that a first-level cache can provide while still maintaining its speed. Consider a processor with the following parameters:

Base CPI, No Memory Stalls	Processor Speed	Main Memory Access Time	First-Level Cache Miss Rate per Instruction**	Second-Level Cache, Direct-Mapped Speed	Miss Rate with Second-Level Cache, Direct-Mapped	Second-Level Cache, Eight-Way Set Associative Speed	Miss Rate with Second-Level Cache, Eight-Way Set Associative
1.5	2 GHz	100 ns	7%	12 cycles	3.5%	28 cycles	1.5%

**First Level Cache miss rate is per instruction. Assume the total number of L1 cache misses (instruction and data combined) is equal to 7% of the number of instructions.

Base CPI 1,5

CR 2 GHz

MM t.ac 100 ns

MM cc = $100 \cdot 10^{-9}$ s · 2 · 10^{9} · s^{-1}

MM c.c. = 200 cc

MR_{L_1} = 7%

L2 DM

L2 cc = 12 cc

MR_{L_2} = 3.5%

L2 $t_{ac} = \frac{12 \text{ cc}}{2 \text{ GHz}} = 6$ ns

L2 cc = 28 cc

L2 8SA

L2 $t_{ac} = \frac{28 \text{ cc}}{2 \text{ GHz}} =$ 19 ns

$MR_{L_{2\,8SA}}$ = 1,5%

2 GHz Base CPI = 1,5

MM tac = 100 ns MM cc = 200 cc

L_1 $MR_{L_1} = 7\%$

L_2 DM $MR_{L_2\,DM} = 3.5\%$

$L_{2\,DM}$ t. ac = 6 ns

$L_{2\,DM}$ cc = 12 cc

L_2 8-way SA $MR_{L_2\,8SA} = 1.5\%$

$L_{2\,8SA}$ t. ac = 19 ns

$L_{2\,8SA}$ cc = 28 cc

I CPI for only L1 cache

$CPI_{L_1} = CPI_{ideal} +$ Mem Stall Cycles

Mem Acc / Instr x MR x MP (cc)

$CPI\,L_1 = 1,5 + 0.07 \times 200\,cc = 1,5 + 14$

$= 15,5$

II L_2 DM

$CPI_{real} = CPI_{real} + MR_{L_1} \times MP_{L_1} + MR_{L_2} \times MP_{L_2}$

$= 1,5 + 0.07 \times 12 + 0.035 \times 200$

$= 1,5 + 0,84 + 7 = 9,34$

III L_2 8 way SA

$CPI_{real} = 1,5 + 0.07 \times 28 + 0.015 \times 200$

$= 6,46$

MM cc = 400 cc

I L_1 only $CPI_{real} = 1,5 + 0,07 \times 400 = 29,5$

II L_2 DM $CPI_{real} = 1,5 + 0,07 \times 12 + 0,035 \times 400$

$= 16,34$

III L_2 8 way SA $CPI_{real} = 1,5 + 0,07 \times 28 + 0,015 \times 400$

$= 9,46$

5.12.2 [10] <§5.4> It is possible to have an even greater cache hierarchy than two levels? Given the processor above with a second-level, direct-mapped cache, a designer wants to add a third-level cache that takes 50 cycles to access and will have a 13% miss rate. Would this provide better performance? In general, what are the advantages and disadvantages of adding a third-level cache?

L_2 DM L_3 cc = 50 $MR\ L_3 = 13\%$

$CPI_{L_3} = 1.5 + 0.07 \times 12 + 0.035 \times 50 + 0.13 \times 200$

$= 30.09$ too high miss rate for L_3

5.12.3 [20] <§5.4> In older processors, such as the Intel Pentium or Alpha 21264, the second level of cache was external (located on a different chip) from the main processor and the first-level cache. While this allowed for large second-level caches, the latency to access the cache was much higher, and the bandwidth was typically lower because the second-level cache ran at a lower frequency. Assume a 512 KiB off-chip second-level cache has a miss rate of 4%. If each additional 512 KiB of cache lowered miss rates by 0.7%, and the cache had a total access time of 50 cycles, how big would the cache have to be to match the performance of the second-level direct-mapped cache listed above?

L_2 size 512 KiB
MR L_2: 4%

each additional 512 KiB lowered miss rates by 0.7%
L_2 CC = 50 CC
match L_2 AM

(512 KiB, 4% MR)

CPI old = $1,5 \times 0,07 \times 12 + 0,035 \times 200 = 9,34$

$CPI_{new\ 512} = 1,5 + 0,07 \times 50 + 0,04 \times 200$ X

1 MB L_2 | $0,04 \rightarrow 0,033$

CPI : $1,5 + 0,07 \times 50 + 0,033 \times 200$ X

1.5 MB L_2 | $0,04 \rightarrow 0,026$

CPI = $1,5 + 0,07 \times 50 + 0,026 \times 200$
= 10,2

---

2 MB L_2 | $0,04 - 3 \times 0,007 = 0,019$

CPI = $1,5 + 0,07 \times 50 + 0,019 \times 200$
= 8,8 CC → 2 MB L_2 needed