重點五：快取效能的量測

- CPU time可以被分為CPU用在執行程式的時脈週期與CPU花在等待記憶體的時脈週期。一般而言，我們假設存取快取命中時所花的時間視為一般CPU執行的時脈週期中的一部份。因此：

  CPU time = (CPU execution cycles + Mem-stall cycles) × Cycle time

- 記憶體暫停所花的時脈週期可定義成程式存取記憶體的次數 × 失誤率 × 失誤處罰時間來表示：

$$\text{Mem-stall cycles/prog.} = \frac{\text{Memory access}}{\text{Program}} \times \text{Miss rate} \times \text{Miss penalty}$$

- 若要計算平均一個指令的執行因存取記憶要停多久可由以下公式表示：

$$\text{Mem-stall cycles/instr.} = \frac{\text{Memory access}}{\text{instruction}} \times \text{Miss rate} \times \text{Miss penalty}$$

- 以下列舉不同快取架構存取次數及有效CPI的計算。

Suppose there are100 instr. in a program among which 30% of load and store instr.	Separate cache		Combined cache
	Instruction Cache	Data Cache	Cache
Number of access/program	100	30	130
Number of access/instr.	1	0.3	1.3

For separate cache:

$$\begin{aligned} CPI_{effective} &= CPI_{base} + \text{Memory stall per instruction} \\ &= CPI_{base} + \text{I-cache stall per instruction} \\ &\quad + \text{D-cache stall per instruction} \\ &= CPI_{base} + \text{I-cache access per instr} \times \text{Miss rate} \times \text{Miss penalty} \\ &\quad + \text{D-cache access per instr} \times \text{Miss rate} \times \text{Miss penalty} \end{aligned}$$

For combined cache:

$$\begin{aligned} CPI_{effective} &= CPI_{base} + \text{Memory stall per instruction} \\ &= CPI_{base} + \text{Cache access per instr} \times \text{Miss rate} \times \text{Miss penalty} \end{aligned}$$

I-cache miss rate : 2%

D-cache miss rate : 4%

miss penalty = 100 cycles

練習

Assume an instruction cache miss rate for a program is 2% and a data cache miss rate is 4%. If a processor has a CPI of 2 without any memory stalls and the miss penalty is 100 cycles for all misses, determine how much faster a processor would run with a perfect cache that never missed. Use the instruction frequencies for SPECint2000 (load/store: 36%)

Answer effective CPI = base CPI + I-cache stall per instruction + D-cache stall per instruction

= 2 + 1 × 100 × 2% + 0.36 × 100 × 4%

Instruction cache stall per instruction = 1 × 2% ×100 = 2

Data cache stall per instruction = 0.36 × 0.04 × 100 = 1.44 = 2 + 2 + 1.44

Mem. stall per instr. = 2 + 1.44 = 3.44 = 5.44 ✗

Effective CPI = Base CPI + Memory stall per instr. = 2 + 3.44 = 5.44

Ration of the execution time = 5.44/2 = 2.72

假設：我們將前面範例的電腦速度加快，在不改變時脈速度的狀況下使其 CPI 從 2 減少到 1。當有快取失誤時，這系統的 CPI 將會變成 1 + 3.44 = 4.44。而有完美快取的系統會變成 4.44/1 = 4.44 倍快。執行時間中，花在記憶體暫停的比例將會從 3.44/5.44 = 63% 提高到 3.44/4.44 = 77%。

練習

Suppose we increase the performance of the computer in the previous example by doubling its clock rate. Since the main memory speed is unlikely to change, assume that the absolute time to handle a cache miss does not change. How much faster will the computer be with the faster clock, assuming the same miss rate as the previous example?

Answer

Measured in faster clock cycle, the new miss penalty will be 200 cycles

Mem. stall per instruction = 2% × 200 + 36% × (4% × 200) = 6.88

Faster system with cache miss, CPI = 2 + 6.88 = 8.88

Slower system with cache miss, CPI = 5.44

The faster clock system will be

Execution time of slow clock/Execution time of faster clock

= I × CPI_slow × Cycle time / (I × CPI_fast × ½ × Cycle time)

= 5.44 / (8.88 × ½) = 1.23 times faster

註：失誤率與失誤處罰時間是影響快取效能的兩個主要因素。以下**重點六**說明使用**集合關聯式快取(set associative cache)** — 藉由減少兩個不同的記憶體區塊競爭快取中同一個位置的機率，以降低失誤率。接著**重點七**說明如何使用**多層快取(multilevel cache)** —在階層中加入額外一層快取，來減小失誤處罰時間。 d-cache 8% miss rate 2, 124, 0.3

I-cache 10% 50, 1

練習

The data cache has a 92% hit rate and a 2-cycle hit latency, and the cache miss penalty is 124 cycles. 30% of instructions are loads and stores. The instruction cache has a hit rate of 90% with a miss penalty of 50 cycles. Assume the base CPI using a perfect memory system is 1.0. Calculate the CPI of the pipeline, assuming everything else is working perfectly. Assume the load never stalls a dependent instruction and assume the processor must wait for stores to finish when they miss the cache. Finally, assume that instruction cache misses and data cache misses never occur at the same time. Show your work.

1. Calculate the additional CPI due to the instruction cache stalls.
2. Calculate the additional CPI due to the data cache stalls.
3. Calculate the overall CPI for the machine.

Answer

(1) The additional CPI due to instruction cache stalls = $1 \times 0.1 \times 50 = 5$

(2) The additional CPI due to data cache stalls = $0.3 \times 0.08 \times 124 = 2.976$

(3) The overall CPI = $1 + 5 + 2.976 = 8.976$

重點六：集合關聯式快取

- 到目前為止，我們放置一個記憶體區塊到快取中，採用的是一種極端的放置方式就是規定一個記憶體區塊只能放置到快取的一個位置。這種放置方式稱為直接對映(direct mapped)。
- 如下圖所示，假設記憶體區塊0和4經常使用，若採用直接對映的方式則這兩個記憶體區塊都會被對應至快取區塊0，因此，快取區塊0便會經常被置換，縱使其餘的快取區塊都是閒置的，記憶體區塊0和4也無法佔用，因此miss rate便會上升。