Drake Bridgewater & Ryan Phillips
May 11, 2014
CS 472 HW 5

(9.2)

Why do computer use cache memory?

(9.3)

What is the meaning of temporal locality and spatial locality?

(9.4)

From first principles, derive an expression for the speedup fatio of memory system with cache (assume the hit ration is h and the ratio of the main storage access time to cache access time to cache access time is k, where k¡1). Assume that the system is an ideal system and that you don't have to worry about the effect of clock cycle times.

(9.5)

For the following systems, calculate the speedup ratio S in the cas t_c is the access time of the cache memory, t_m is the access time of the main store, and h is the hit ratio.

a $t_m = 70ns,\ t_c = 7ns,\ h = 0.9$

b $t_m = 60ns,\ t_c = 3ns,\ h = 0.9$

c $t_m = 60ns,\ t_c = 3ns,\ h = 0.8$

d $t_m = 60ns,\ t_c = 3ns,\ h = 0.97$

(9.6)

For the following ideal systems, calculate the hit ratio h required to achieve the stated speedup ratio S.

a $t_m = 60ns,\ t_c = 3ns,\ S = 1.1$

b $t_m = 60ns,\ t_c = 3ns,\ S = 2.0$

c $t_m = 60ns,\ t_c = 3ns,\ S = 5.0$

d $t_m = 60ns,\ t_c = 3ns,\ S = 15.0$

(9.8)

For the following system that use a clocked microprocessor, caculate the maximum speedup ratio you could expect to see as h approaches 100%.

a $t_{cyc} = 20ns,\ t_m = 75ns,\ t_c = 15ns$

b $t_{cyc} = 20ns,\ t_m = 75ns,\ t_c = 25ns$

c $t_{cyc} = 10ns,\ t_m = 75ns,\ t_c = 15ns$

(9.11)

In a direct-mapped cache memory system, what is the meaning of the following terms.

a Word

b Line

c Set

(9.12)

How is data in main store mapped on to each of the following?

a A direct-mapped cache

b A fully associative cache

c A set associative cache

(9.17)

What is cache coherency?

(9.22)

Why is it harder to design a data cache than an instruction cache?

(9.23)

When a CPU writes to the cache, both the item in the cache and the corresponding item in the memory must be updated. If data is not in the cache, it must be fetched from memory and loaded in the cache. If t_1 is the time taken to reload the cache on a miss, show that the effective average access time of the memory system is given by

$t_{avg} = ht_c + (1 - h)t_m + (1 - h)t_l$.

(9.26)

A system has a level 1 cache and a level 2 cache. The hit rate of the level 1 cache is 90%, and the hit rate of the level 2 cache is 80%. An access to level 1 cache requires one cycle, an access to level 2 cache requires four cycles, and an access to main memory requires 50 cycles. What is the average access time?

(9.28)

In the context of multilevel caches, what is the difference bettwen a local miss rate and a global miss rate?

(9.35)

A 64-bit processor has a 8-MB, four-way set-associative cache with 32-byte lines. How is the address arranged in terms of set, line, and offset bits?

(9.41)

What are the fundamental differences between cache memory (as found in a CPU) and cache memory found in a hard disk drive?

(9.42)

What are the differences between write-back and write-through caches, and what are the implications for system performance?

(9.43)

A computer with 32-bit address architecture has a memory management system with single-level 4 KB page tables. How much memory space must be devoted to the page tables?

(9.45)

A computer runs an instruction set with ethe characteristics in the following table.

Arithmatic operations, 70%, 1 cycle
Conditional operations, 15%, 2 cycles
Load, 10%, 2 cycles
Store, 5%, 2 cycles
Hit rate, 95%
Cost of a cache miss (read), 10 cycles
Write-through time, 5 cycles (writes to memory are not buffered)
What is the average number of cycles per instruction?

(9.46)

Consider the following code that accesses three values in memory scalar integers x and s, and an integer vector y[i]. What is the memory latency in clock cycles for a trip round the loop (after the first iteration)? Assume that the array is not cached and each new access to the array results in a miss.

The system has both L1 and L2 caches. The access time of the L1 cache is two cycles, the acess time of the L2 cache is 6 cycles and main memory has an access time of 50 cycles. In this case all memory and cache memory accesses take place in parallel.

```
for (i = 0; i < 100; i++)
{
x = y[i];
s = s + x;
}
```

(9.57)

A computer with a 24-bit address bus has a main memory of size 16 MB and a cache size of 64 KB. The wordlength is two bytes.

a What is the address format for a direct-mapped cache with a line size of 32 words?

b What is the address format for a fully associative cache with a line size of 32 words?

c What is the adress format for a four-way set-associative cache with a line size of 16 words?