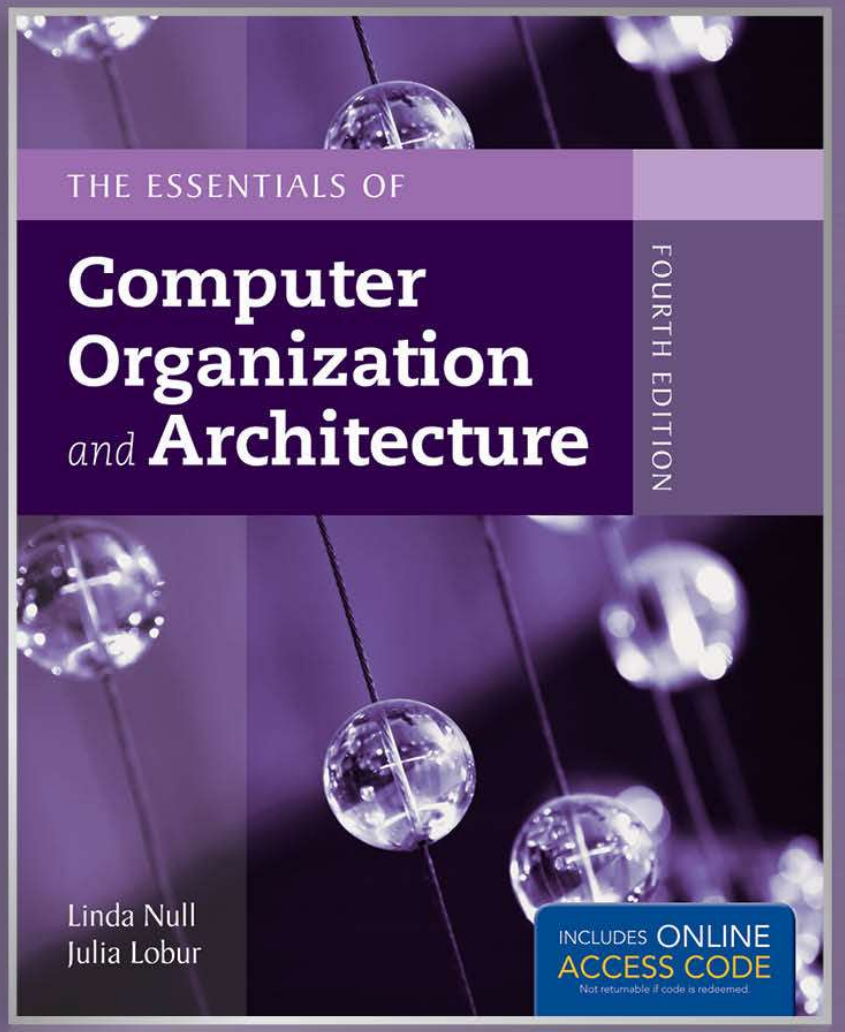
Chapter 6
Memory



Chapter 6
REVIEW OF ESSENTIAL TERMS AND CONCEPTS

6.What are the three forms of locality?

10.What are the three fields in a direct mapped cache address? How are they used to access a word located in cache?

15. What are the three fields in a set associative cache address and how are they used to access a location in cache?

19. What, exactly, is effective access time(EAT)?

22. What is a dirty block?

Chapter 6
EXERCISES

1. Suppose a computer using direct mapped cache has 1 048 576 words of main memory, and a cache of 32 blocks, where each cache block contains 16 words.

   a. How many blocks of main memory are there?

   b. What is the format of a memory address as seen by the cache, i.e., what are the sizes of the tag, block, and word fields?

   c. To which cache block will the memory reference $0DB63_{16}$ map?

Chapter 6
EXERCISES

4. Suppose a computer using fully associative cache has

65 536 words of main memory, and a cache of 64 blocks, where each cache block contains 32 words.

a. How many blocks of main memory are there?

b. What is the format of a memory address as seen by the cache, i.e., what are the sizes of the tag and word fields?

c. To which cache block will the memory reference 0xF8C9 map?

Chapter 6
EXERCISES

8. A 2-way set-associative cache consists of four sets. Main memory contains 2K blocks of eight words each.

a. Show the main memory address format that allows us to map addresses from main memory to cache. Be sure to include the fields as well as their sizes.

b. Compute the hit ratio for a program that loops 3 times from locations 8_{10} to 51_{10} in main memory. You may leave the hit ratio in terms of a fraction.

Chapter 6
EXERCISES

***11. Suppose we have a computer that uses a memory address word size of 8 bits. This computer has a 16-byte cache with 4 bytes per block. The computer accesses a number of memory locations throughout the course of running a program.**

Suppose this computer uses direct-mapped cache. The format of a memory address as seen by the cache is shown here:

The system accesses memory addresses in this exact order: 0x6E, 0xB9,0x17, 0xE0, 0x4E, 0x4F, 0x50, 0x91, 0xA8, 0xA9, 0xAB, 0xAD, 0x93,and 0x94. The memory addresses of the first four accesses have been loaded into the cache blocks as shown below. (The contents of the tag are shown in binary, and the cache "contents" are simply the address stored at that cache location.)

Chapter 6
EXERCISES

***11.**

	Tag Contents	Cache Contents (represented by address)
Block 0	1110	0xE0
		0xE1
		0xE2
		0xE3

	Tag Contents	Cache Contents (represented by address)
Block 1	0001	0x14
		0x15
		0x16
		0x17

	Tag Contents	Cache Contents (represented by address)
Block 2	1011	0xB8
		0xB9
		0xBA
		0xBB

	Tag Contents	Cache Contents (represented by address)
Block 3	0110	0x6C
		0x6D
		0x6E
		0x6F

a) What is the hit ratio for the entire memory reference sequence given above, assuming that we count the first four accesses as misses?

b) What memory blocks will be in the cache after the last address has been accessed?

Chapter 6
EXERCISES

12. Given a byte-addressable memory with 256 bytes, suppose a memory dump yields the results shown below. The address of each memory cell is determined by its row and column. For example, memory address 0x97 is in the 9^{th} row, $7t^{h}$ column, and contains the hexadecimal value 43. Memory location 0xA3 contains the hexadecimal value 58.

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	DC	D5	9C	77	C1	99	90	AC	33	D1	37	74	B5	82	38	E0
1	49	E2	23	FD	D0	A6	98	BB	DE	9A	9E	EB	04	AA	86	E5
2	3A	14	F3	59	5C	41	B2	6D	18	3C	9D	1F	2F	78	44	1E
3	4E	B7	29	E7	87	3D	B8	E1	EF	C5	CE	BF	93	CB	39	7F
4	6B	69	02	56	7E	DA	2A	76	89	20	85	88	72	92	E9	5B
5	B9	16	A8	FA	AE	68	21	25	34	24	B6	48	17	83	75	0A
6	40	2B	C4	1D	08	03	0E	0B	B4	C2	53	FB	E3	8C	0C	9B
7	31	AF	30	9F	A4	FE	09	60	4F	D7	D9	97	2E	6C	94	BC
8	CD	80	64	B3	8D	81	A7	DB	F1	BA	66	BE	11	1A	A1	D2
9	61	28	5D	D4	4A	10	A2	43	CC	07	7D	5A	C0	D3	CF	67
A	52	57	A3	58	55	0F	E8	F6	91	F0	C3	19	F9	BD	8B	47
B	26	51	1C	C6	3B	ED	7B	EE	95	12	7C	DF	B1	4D	EC	42
C	22	0D	F5	2C	62	B0	5E	DD	8E	96	A0	C8	27	3E	EA	01
D	50	35	A9	4C	6A	00	8A	D6	5F	7A	FF	71	13	F4	F8	46
E	1B	4B	70	84	6E	F7	63	3F	CA	45	65	73	79	C9	FC	A5
F	AB	E6	2D	54	E4	8F	36	6F	C7	05	D8	F2	AD	15	32	06

Chapter 6
EXERCISES

The system from which this memory dump was produced contains 4 blocks of cache, where each block consists of eight bytes. Assume the following sequence of memory addresses take place: 0x2C, 0x6D, 0x86, 0x29, 0xA5, 0x82, 0xA7, 0x68, 0x80, and 0x2B.

a) How many blocks of main memory are there?

b) Assuming a direct mapped cache:

i) Show the format for a main memory address assuming the system uses direct mapped cache. Specify field names and sizes.

ii) What does cache look like after the ten memory accesses have taken place? Draw the cache and show contents and tags.

iii) What is the hit rate for this cache on the given sequence of memory accesses?

Chapter 6
EXERCISES

c) Assuming a fully associative cache:

i) Show the format for a main memory address. Specify field names and sizes.

ii) Assuming all cache blocks are initially empty, blocks are loaded into the first available empty cache location, and cache uses a first-in-first-out replacement policy, what does cache look like after the ten memory accesses have taken place?

iii) What is the hit rate for this cache on the given sequences of memory accesses?

Chapter 6
EXERCISES

d) Assuming a 2-way set associative cache:

i) Show the format for a main memory address. Specify field names and sizes.

ii) What does cache look like after the ten memory accesses have taken place?

iii) What is the hit ratio for this cache on the given sequences of memory accesses?

iv) If a cache hit retrieves a value in 5ns, and retrieving a value from main memory requires 25ns, what is the average effective access time for this cache, assuming that all memory accesses exhibit the same hit rate as the sequence of 10 given, and assuming the system uses a non-overlapped (sequential) access strategy?

Chapter 6
EXERCISES

13. A direct-mapped cache consists of eight blocks. Main memory contains 4K blocks of eight words each. Access time for the cache is 22 ns and the time required to fill a cache slot from main memory is 300ns (this time will allow us to determine the block is missing and bring it into cache). Assume a request is always started in parallel to both cache and to main memory (so if it is not found in cache, we do not have to add this cache search time to the memory access). If a block is missing from cache, the entire block is brought into the cache and the access is restarted. Initially, the cache is empty.

Chapter 6
EXERCISES

13. a. Show the main memory address format that allows us to map addresses from main memory to cache. Be sure to include the fields as well as their sizes.

b. Compute the hit ratio for a program that loops 4 times from locations 0 to 67_{10} in memory.

c. Compute the effective access time for this program.

Chapter 6
EXERCISES

14. Consider a byte-addressable computer with 24-bit addresses, a cache capable of storing a total of 64K bytes of data and blocks of 32 bytes. Show the format of a 24-bit memory address for:

a. direct mapped

b. associative

c. 4-way set associative

Chapter 6
EXERCISES

18. Suppose a process page table contains the entries shown below. Using the format shown in Figure 6.22a, indicate where the process pages are located in memory.

Frame	Valid Bit
1	1
--	0
0	1
3	1
--	0
--	0
2	1
--	0

Chapter 6
EXERCISES

22. You have a virtual memory system with a two-entry TLB, a 2-way set associative cache and a page table for a process P. Assume cache blocks of 8 words and page size of 16 words. In the system below, main memory is divided up into blocks, where each block is represented by a letter. Two blocks equals one frame.

0	3
4	1

TLB

Set 0	tag	C	tag	I
Set 1	tag	D	tag	H

Cache

	Frame	Valid
0	3	1
1	0	1
2	-	0
3	2	1
4	1	1
5	-	0
6	-	0
7	-	0

Page Table

Frame		Block
0	C	0
	D	1
1	I	2
	J	3
2	G	4
	H	5
3	A	6
	B	7

Main Memory

Page		Block
0	A	0
	B	1
1	C	2
	D	3
2	E	4
	F	5
3	G	6
	H	7
4	I	8
	J	9
5	K	10
	L	11
6	M	12
	N	13
7	O	14
	P	15

Virtual Memory For Process P

Chapter 6
EXERCISES

22. Given the system state as depicted above, answer the following questions:

a. How many bits are in a virtual address for process P? Explain.

b. How many bits are in a physical address? Explain.

c. Show the address format for virtual address 18_{10} (specify field name and size) that would be used by the system to translate to a physical address and then translate this virtual address into the corresponding physical address. (Hint: convert 18 to its binary equivalent and divide it into the appropriate fields.) Explain how these fields are used to translate to the corresponding physical address.

Chapter 6
EXERCISES

22. d. Given virtual address 6_{10} converts to physical address 54_{10}. Show the format for a physical address (specify the field names and sizes) that is used to determine the cache location for this address. Explain how to use this format to determine where physical address 54 would be located in cache. (Hint: convert 54 to binary and divide it into the appropriate fields.)

e. Given virtual address 25_{10} is located on virtual page 1, offset 9. Indicate exactly how this address would be translated to its corresponding physical address and how the data would be accessed. Include in your explanation how the TLB, Page Table, Cache and Memory are used.

Chapter 6
EXERCISES

23. Given a virtual memory system with a TLB, a cache, and a page table. Assume the following:

- **A TLB hit requires 5ns**
- **A cache hit requires 12ns**
- **A memory reference requires 25ns**
- **A disk reference requires 200ms (this includes updating the page table, cache, and TLB)**
- **The TLB hit ratio is 90%**
- **The cache hit rate is 98%**
- **The page fault rate is 0.1%**

Chapter 6
EXERCISES

23. • On a TLB or cache miss, the time required for access includes a TLB and/or cache update, but the access is NOT restarted

• On a page fault, the page is fetched from disk, all updates are performed but the access IS restarted

• All references are sequential (no overlap, nothing done in parallel)

Chapter 6
EXERCISES

23. For each of the following, indicate whether or not it is possible. If so, specify the time required for accessing the requested data.

a. TLB hit, cache hit

b. TLB miss, page table hit, cache hit

c. TLB miss, page table hit, cache miss

d. TLB miss, page table miss, cache hit

e. TLB miss, page table miss

Write down the equation to calculate the effective access time.

Chapter 6
EXERCISES

25. A system implements a paged virtual address space for each process using a one-level page table. The maximum size of virtual address space is 16MB. The page table for the running process includes the following valid entries (the → notation indicates that a virtual page maps to the given page frame, that is, it is located in that frame):

Virtual page 2 → page frame 4

Virtual page 1 → page frame 2

Virtual page 0 → page frame 1

Virtual page 4 → page frame 9

Virtual page 3 → page frame 16

Chapter 6
EXERCISES

25. The page size is 1024 bytes and the maximum physical memory size of the machine is 2MB.

a. How many bits are required for each virtual address?

b. How many bits are required for each physical address?

c. What is the maximum number of entries in a page table?

d. To which physical address will the virtual address 0x5F4 translate?

e. Which virtual address will translate to physical address 0x400?