

# Computer Architecture 2011 (CART'11)

## Exam

June 1<sup>st</sup>

### Student number<sup>1</sup>:

Please write your student number on every page of the exam. If you need to add additional pages, please indicate how many and write your student number on each of them. **You should answer on the exam sheets**, there is space for it.

## 1 Introduction

This exam is divided into 6 sections that give a total of 143 points plus 9 bonus points. The mapping to the 7-scale grade is as follows:

0 ... 24	→	-3
25 ... 41	→	00
42 ... 58	→	02
59 ... 84	→	4
85 ... 110	→	7
111 ... 127	→	10
128 ... 143	→	12

You will find in appendix (last page) the list of the first powers of 2 that may help you for the whole exam. The questions are weighted in points in function of their difficulty or importance. As a hint, questions requiring to write some code that give  $x$  points can be solved with  $x$  lines of code.

## 2 Representing and Manipulating Information (36+2)

**Exercise1:** Let us consider integers represented on 8 bits.

1. **(4 pt)** Give the hexadecimal and decimal representations of the following binary numbers on 8 bits. Integers are **signed**. Hint: for decimal, it helps to write the formula, e.g.,  $00100010 = 0x22 = 2 + 2 \cdot 16 = 34$ .

01110010 = 0x =

00110101 = 0x =

11001101 = 0x =

11111111 = 0x =

---

<sup>1</sup>No need to write your name.

2. (4 pt) Give the binary and hexadecimal representations of the following decimal numbers.

127 = 0x =

-128 = 0x =

-1 = 0x =

15 = 0x =

**Exercise2:** We want to store the integer 0xdeadbeef (32 bits) in memory. This will occupy 4 consecutive bytes.

1. (1 pts) Where in memory will each byte (in hexadecimal) be stored if a computer is using the little endian representation?

Address	0	1	2	3
Bytes				

2. (1 pts) Where in memory will each byte (in hexadecimal) be stored if a computer is using the big endian representation?

Address	0	1	2	3
Bytes				

**Exercise3:** Write C expressions, in terms of a variable `x`, for the following values. Your code should work for any word size  $w \geq 8$ . For reference, we show the result of evaluating the expressions for `x=0x87654321`, with  $w = 32$ . Use hexadecimal for the constants that you will need.

- (2 pts) The least significant byte of `x`, with all other bits set to 0. [0x00000021].
- (2 pts) The least significant byte set to all 1s, and all other bytes of `x` left unchanged. [0x876543FF].

**Exercise4:** Let us interpret an array of integer as an array of bits. Let us assume integers are 32 bits wide. We want to toggle bit `n` in that array, i.e., turn a 1 to a 0 and vice-versa. For example toggle bit 37 means toggle the 5<sup>th</sup> bit of the second integer.

```

1 void toggle_bit (unsigned int *bits, unsigned int n)
2 {
3
4
5
6
7
8
9
10
11
12
13 }
```

Listing 1: Function that toggles bit `n` in an array of bits.

1. **(5 pts)** Fill in the function in listing 1 without using any *if-statement*. Hint: use the xor operator  $\wedge$ . (Bonus +1 pt) Do not hard-code the assumption that integers are on 32 bits.

**Exercise5:** Assume we are running code on a 32-bit machine using two's-complement arithmetic for signed values. The variables are declared and initialized as in listing 2.

```
1 int x = foo(); /* Some arbitrary value. */  
2 int y = bar(); /* Some arbitrary value. */
```

Listing 2: Declarations for `x` and `y`.

For each of the following C expressions, either (1) argue that it is true (it evaluates to 1) for *all* values of `x` and `y`, or (2) give values of `x` and `y` for which it is false (evaluates to 0). You can use a power of two or the constants `INT_MIN` and `INT_MAX` in your answers.

1. **(2 pts)** `(x > 0) || (x-1 < 0)`
2. **(2 pts)** `(x < 0) || (-x <= 0)`
3. **(2 pts)** `(x > 0) || (-x >= 0)`

**Exercise6:** In the C language right shifts are performed arithmetically for signed values and logically for unsigned values.

1. **(1 pt)** What is the difference between arithmetic and logical right shifts? Be precise and say which shift does what.

**Exercise7:** In the Java language you have only signed integers.

1. **(Bonus +1 pt)** How to do you do arithmetic and logical right shifts in Java?

**Exercise8:** Assume variables `x`, `f`, and `d` are of type `int`, `float`, and `double`, respectively. Their values are arbitrary, except that neither `f` nor `d` equals  $+\infty$ ,  $-\infty$ , or *NaN*. For each of the following C expressions, either argue that it will always be true (evaluates to 1) or give a value for the variables such that it is not true (evaluates to 0). You can give the values in decimal or hexadecimal.

1. **(2 pts)** `x == (int)(double) x`
2. **(2 pts)** `x == (int)(float) x`
3. **(2 pts)** `f = (float)(double) f`
4. **(2 pts)** `(f+d)-f == d`

**Exercise9:** Commutativity and associativity are important properties on operators that matter for programmers and compilers.

1. (2 pt) Fill in the table to indicate if the + operator is commutative and associative for integers and floating point numbers.

+	int	float
commutative		
associative		

### 3 Program Encodings (28)

**Reminder** The general registers on IA32 are `%eax`, `%ebx`, `%ecx`, `%edx`, `%esi`, `%edi`, `%esp`, and `%ebp`. We adopt the same convention used by `gcc` for assembly, in particular for the order of the arguments. The syntax for the `mov` instruction is `mov src, dst`.

**Exercise10:** Assume the following values are stored at the indicated memory addresses and registers:

Address	Value	Register	Value
0x100	0xFF	<code>%eax</code>	0x100
0x104	0xAB	<code>%ecx</code>	0x1
0x108	0x13	<code>%edx</code>	0x3
0x10C	0x11		

1. (3 pts) Fill in the following table showing the byte values for the indicated operands. As a recall the general format for addressing is D(B,I,S) where D is the displacement (0 if missing), B the base address (0 if missing), I (0 if missing) the index, and S the size for the offset (1 if missing). The address is then  $D+B+I*S$ .

Operand	Value
<code>4(%eax)</code>	...
<code>9(%eax,%edx)</code>	...
<code>0xFC(,%ecx,4)</code>	...

**Exercise11:** A function that has the following prototype

```
void decode(int *xp, int *yp, int zp);
```

has its implementation compiled into assembly code. The code is given in listing 3. Assume we are in 32-bit mode on some Intel compatible CPU. As a reminder `movl X, Y` will copy the value of X to Y, where X or Y (not both) can be a memory reference (in the syntax given in the previous exercise) and the other argument a register (or a constant for X only).

```

1  pushl    %ebp
2  movl     %esp, %ebp
3  movl     8(%ebp), %edi ; xp
4  movl     12(%ebp), %edx ; yp
5  movl     16(%ebp), %ecx ; zp
6  movl     (%edx), %ebx
7  movl     (%ecx), %esi
8  movl     (%edi), %eax
9  movl     %eax, (%edx)
10 movl     %ebx, (%ecx)
11 movl     %esi, (%edi)
12 movl     %ebp, %esp
13 pop      %ebp
14 ret

```

Listing 3: Assembly code of `decode`.

Parameters `xp`, `yp`, and `zp` are stored at the memory locations with offsets 8, 12, and 16, respectively, relative to the address in the register `%ebp`.

1. (1 pt) What is the register `%ebp` used for?
2. (2 pt) What do the first `pushl` and `movl` instructions do?
3. (2 pt) What do the last `movl` and `pushl` instructions do?
4. (6 pts) Write C code for `decode` that will have an effect equivalent to the assembly code of listing 3.

```
1 void decode(int *xp, int *yp, int zp)
2 {
3
4
5
6
7
8
9
10
11
12
13 }
```

Listing 4: Your implementation of `decode`.

5. (4 pts) Suppose that the function is called with the values of `%eax`, `%esi`, and `%edi`, respectively, for `xp`, `yp`, and `zp`. The assembly code to call the function is

```
pushl %edi
pushl %esi
pushl %eax
call decode
```

Suppose that `%esp=0x120` before calling `decode`. Fill in the written values in the stack after having executed the instructions to call the function and up to line 2 of listing 3.

Notes: You do not have to use every line in the table. To write the value of, e.g., `%eax`, write `%eax` in the corresponding memory cell (of 32 bits). Be careful with `call`, you may want to write in plain English what you want to put in the stack.

Hint: You can cross-check with listing 3 to make sure you get it right.

Address	Memory
0x134	
0x130	
0x12C	
0x128	
0x124	
0x120	
0x11C	
0x118	
0x114	
0x110	
0x10C	

6. (1 pt) What is the new value of `%esp`?

**Exercise12:** Conditional jumps are very common in programs. If we consider the code in listing 5 that computes  $|x - y|$ , the return statement depends on a comparison between  $x$  and  $y$ . Note the C statement `cond ? p : q` evaluates to  $p$  if `cond` is true else  $q$ .

```

1 void absdiff (int x, int y)
2 {
3     return (x < y) ? y - x : x - y;
4 }

```

Listing 5: Implementation of `absdiff`.

```

1 push    %ebp
2 mov     %esp, %ebp
3 movl    8(%ebp), %edx ; Get x
4 movl    12(%ebp), %eax ; Get y
5
6
7
8
9
10
11
12
13
14
15
16
17

```

```

18
19
20 mov    %ebp, %esp
21 pop    %ebp
22 ret

```

Listing 6: Your assembly code for absdiff using conditional jumps.

1. **(6 pts)** Write the assembly code corresponding to this function using conditional jumps. You may want to use the instructions `cmpl`, `jge`, `jle`, `subl`, and `jmp` (although not necessarily all of them). The result of the function should be in `%eax` upon returning. You do not need to do memory access, though it is possible to get an acceptable solution that does it. Hint: this is may easier if you write pseudo-code or C using goto statements before you write your answer in assembly.

`cmpl Y, X` compares `X` and `Y` and sets flags accordingly. To simplify, we'll consider only the right conditional jumps.

`jge LABEL` jumps if `X >= Y` holds when comparing `X` and `Y`.

`jle LABEL` jumps if `X < Y` holds when comparing `X` and `Y`.

`subl X, Y` computes `Y = Y - X`.

`jmp LABEL` jumps without condition.

**Exercise13:** Let us consider the structure declared as

```

struct rec {
    int i;
    int j;
    int a[4];
};

```

1. **(3 pts)** Fill-in the assembly code in listing 7 that implements the function

```
int readrec(struct rec* p);
```

that should return `p->a[p->i+p->j]`. Hint: it may help to write down the offsets to access the data before trying to write the code. You will want to use the `addl` instruction. You can use it as `addl D(adr),X` for adding the number at address `adr` with some offset (or displacement) `D` to `X`.

```

1  push    %ebp
2  mov     %esp, %ebp
3  movl    8(%ebp), %edx ; Get p
4
5
6
7
8
9
10
11
12
13 mov     %ebp, %esp
14 pop     %ebp
15 ret

```

Listing 7: Your assembly code for returning `p->a[p->i+p->j]`.

Stage	OP rA,rB	mrmovl D(rB),rA
Fetch	icode:ifun $\leftarrow$ M <sub>1</sub> [PC] rA:rB $\leftarrow$ M <sub>1</sub> [PC+1]  valP $\leftarrow$ PC+2	icode:ifun $\leftarrow$ M <sub>1</sub> [PC] rA:rB $\leftarrow$ M <sub>1</sub> [PC+1] valC $\leftarrow$ M <sub>4</sub> [PC+2] valP $\leftarrow$ PC+6
Decode	valA $\leftarrow$ R[rA] valB $\leftarrow$ R[rB]	valB $\leftarrow$ R[rB]
Execute	valE $\leftarrow$ valB OP valA set CC	valE $\leftarrow$ valB + valC
Memory		valM $\leftarrow$ M <sub>4</sub> [valE]
Write back	R[rB] $\leftarrow$ valE	R[rA] $\leftarrow$ valM
PC update	PC $\leftarrow$ valP	PC $\leftarrow$ valP

Table 1: Computation stages for an arithmetic operator OP and the memory to register move mrmovl.

## 4 Y86 (28+6)

**Exercise14:** For these questions we will consider the simplified CPU model of the Y86. Its processing is separated into different stages as shown in table 1 for two instructions. The stages are using a restricted set of hardware registers (rA, rB, valA, ...) to do the computations.

1. **(6 pts)** What do each of the computation stages do? The table given as an example is only an example. The question is about the meaning of the stages. You should not refer to valA, valB, etc ...

Fetch:

Decode:

Execute:

Memory:

Write back:

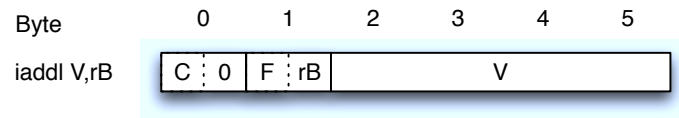
PC update:

2. **(2 pts)** In table 1 the PC update is the same for both instructions. Obviously this is not always the case. Give an instruction for which this is not the case and explain what this instruction does (no need to give the computation stages). You should point out the difference in the PC update. The solution to that question is not unique, it is enough to pick one solution.
3. **(Bonus +2 pts)** Give another solution and explain it.

**Exercise15:** It is common to add a constant value to a register. Unfortunately the instruction



set of the Y86 requires first using an `irmovl`<sup>2</sup> instruction to set a register to the constant, and then an `addl` instruction to add this value to the destination register. Suppose we want to add a new instruction `iaddl` with the following format:



This instruction adds the constant value `V` to register `rB`. The values of the constants, i.e., `C`, `0`, and `F`, in the encoding are irrelevant for the purpose of the exercise.

1. **(Bonus +1)** The value `F` is special. What is it used for?
  
2. **(8 pts)** Describe the computations performed to implement this instruction.

Stage	<code>iaddl V,rB</code>
Fetch	
Decode	
Execute	
Memory	
Write back	
PC update	

3. **(1 pt)** Why would a sequential implementation of our computation stages be inefficient?

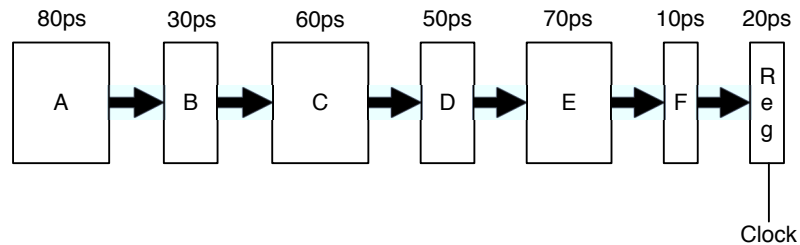
**Exercise16:** To improve performance, a *pipeline* architecture is used instead of a sequential one.

1. **(1 pts)** How does it improve performance?
  
2. **(2 pts)** Name two problems that need to be solved with pipelines.

---

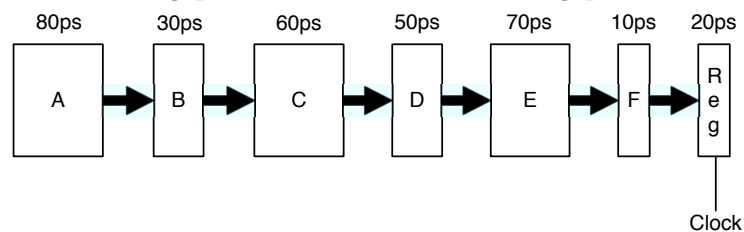
<sup>2</sup>Move immediate (constant) to register.

**Exercise17:** Suppose we analyze some combinational logic and determine that it can be separated into a sequence of six blocks, named A to F, having delays of 80, 30, 60, 50, 70, and 10 ps, respectively, illustrated as follows:



We can create pipelined versions of this design by inserting pipeline registers between pairs of these blocks. Different combinations of pipeline depth (how many stages) and maximum throughput arise, depending on where we insert the pipeline registers. Assume that a pipeline register has a delay of 20 ps.

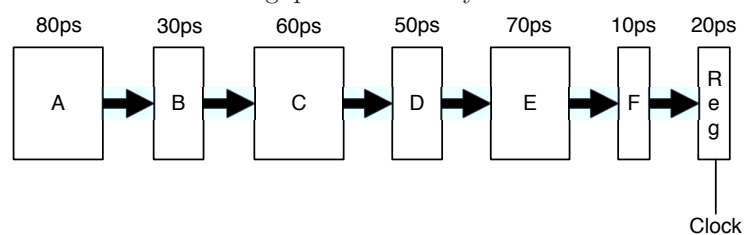
1. **(2 pt)** On the figure, you notice a clock connected to the output (hardware) register “Reg”. Explain.
2. **(3 pts)** Inserting a single register gives a two-stage pipeline. Where should the register be inserted to maximize throughput<sup>3</sup> and latency<sup>4</sup>?



Throughput:

Latency:

3. **(3 pts)** Where should two registers be inserted to maximize the throughput of a three-stage pipeline? What would be the throughput and latency?



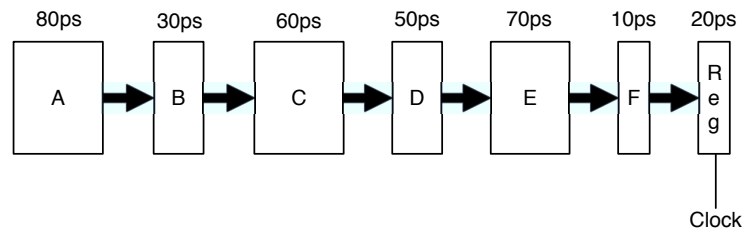
Throughput:

Latency:

4. **(Bonus +3 pts)** Where should three registers be inserted to maximize the throughput of a four-stage pipeline? What would be the throughput and latency?

<sup>3</sup>Give the expression for the throughput, e.g.,  $\frac{1}{320 \times 10^{-12}}$  instructions per seconds (IPS).

<sup>4</sup>Do not forget the (inserted registers) taking 20ps each.



Throughput:  
Latency:

## 5 RAM and Memory Hierarchy (15)

**Exercise18:** Different RAM technologies are used together in computers.

1. (1 pt) What is SRAM memory and where is it used?
2. (1 pt) What is DRAM memory and where is it used?
3. (3 pts) Give three differences between SRAM and DRAM.

**Exercise19:** The principle of locality is the fundamental reason why modern computers work in practice.

1. (2 pts) What is temporal locality?
2. (2 pts) What is spatial locality?

```

1  int sum1(int a[M][N]) {
2      int i, j, sum = 0;
3
4      for (i = 0; i < M; i++)
5          for (j = 0; j < N; j++)
6              sum += a[i][j];
7
8      return sum;
9  }
10
11 int sum2(int a[M][N]) {
12     int i, j, sum = 0;
13
14     for (j = 0; j < N; j++)
15         for (i = 0; i < M; i++)
16             sum += a[i][j];
17
18     return sum;
19 }
```

Listing 8: Different implementations of sums.

3. **(2 pts)** If we consider the implementations of listing 8 that compute the sums of all elements in a matrix in two different ways, which one is better and why?
4. **(4 pts)** Modern architectures implement a *memory hierarchy*. Illustrate it with a drawing<sup>5</sup>. Give two main characteristics of the type of memory when you go to the top or the bottom of the hierarchy?

## 6 Cache Memory (15)

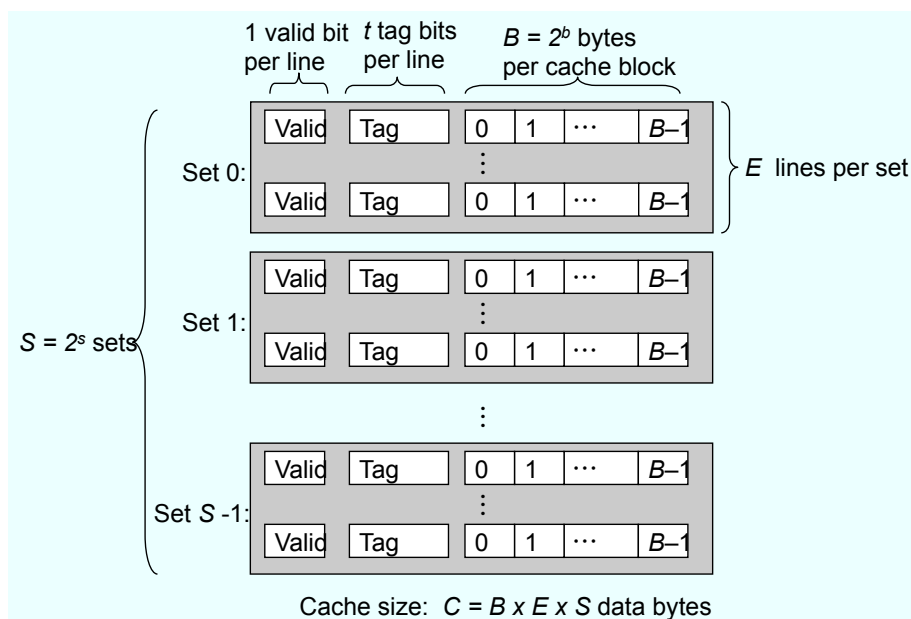


Figure 1: General organization of caches.

**Exercise20:** Caches are organized in general into sets of lines where each line stores a block of bytes as shown in Figure 1. A cache is determined by its parameters  $S$ ,  $E$ , and  $B$ .

1. **(1 pt)** What does it mean to have an  $N$ -way associative cache?

<sup>5</sup>Use at least 4 different types of levels in your illustration. You can use three levels of cache but they count as one.

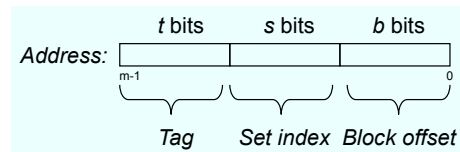


Figure 2: Physical addressing.

2. (1 pt) What is a fully associative cache?
3. (1 pt) What is a direct mapped cache?
4. (1 pt) What is the valid bit used for?
5. (1 pt) The addresses are split with a sequence of t, s, and b bits for the different field. Why not use the sequence t,b,s instead?
6. (3 pts) The following table gives the parameters for a number of different caches. For each cache, determine the number of cache sets (S), tab bits (t), set index bits (s), and block offset bits (b). In the table m is the number of physical address bits. Hint: rewrite C, B, and E as powers of two.

m	C	B	E	S	t	s	b
32	1024	4	1				
32	1024	8	4				
32	1024	32	32				

**Exercise21:** Assume that the memory is byte addressable with addresses on 13 bits, you can access bytes individually, and the cache is two-way associative ( $E=2$ ), with a 4-byte block size ( $B=4$ ) and eight sets ( $S=8$ ).

1. (2 pts) Indicate which bits would be used to determine the cache block offset (CO), the cache set index (CI), and the cache tag (CT).

**Exercise22:** To benchmark memory systems, an experiment known as the memory mountain can be done. Figure 3 shows the result obtained on a Core i7.

1. (2 pts) Describe the memory mountain experiment.
2. (2 pts) Explain the riges of temporal locality and the slopes of spacial locality.

3. (1 pt) At stride 1, the throughput is kept flat through L2 and L3, and stays high even when going through memory. Explain.

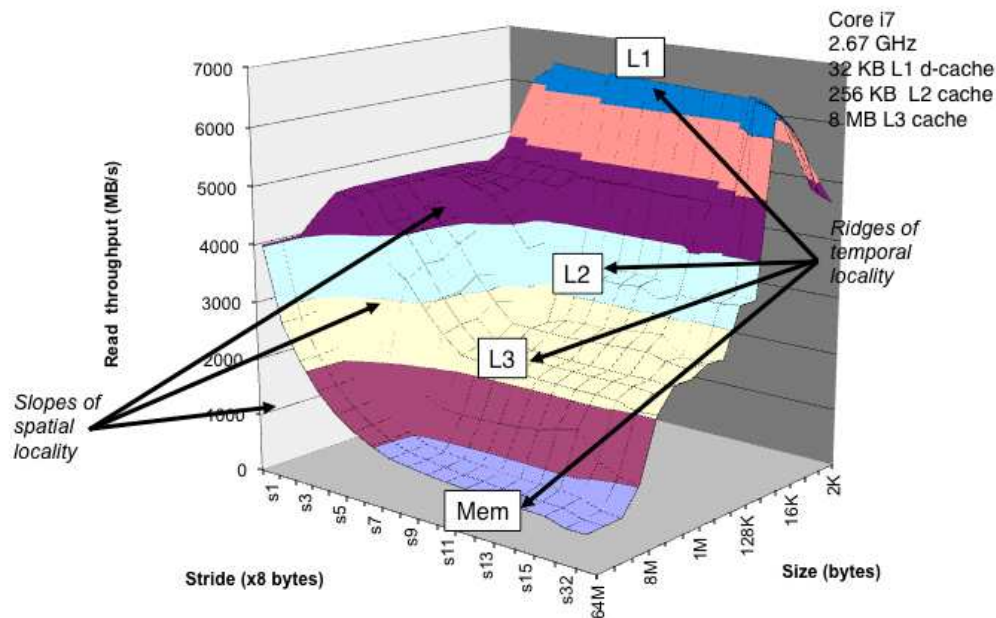


Figure 3: The memory mountain for the Core i7.

## 7 Virtual Memory (21+1)

**Exercise23:** Virtual addressing is used instead of physical addressing in running programs. Virtual addresses need to be translated at some point in time to physical addresses to access data.

- (2 pts) Name two benefits of using virtual addressing.
- (4 pts) Given a 32-bit virtual address space and a 24-bit physical address, determine the number of bits in the virtual page number (VPN), virtual page offset (VPO), physical page number (PPN), and physical page offset (PPO) for the following page sizes P. Hint: rewrite P as a power of 2.

P	No. VPN bits	No. VPO bits	No. PPN bits	No PPO bits
1 KB				
2 KB				
4 KB				
8 KB				

**Exercise24:** Figure 4 shows a simple memory system with a 4-way associative TLB with 16 entries and a direct mapped cache with 16 lines and 4 bytes per line (block size).

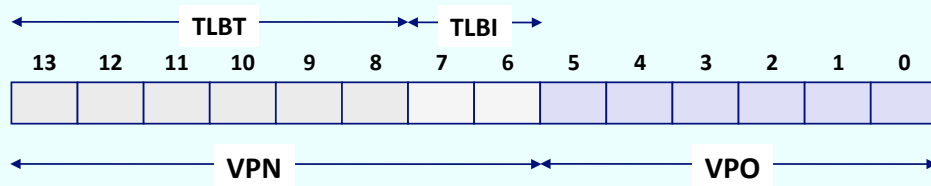
1. **(1 pt)** What does TLB stand for?
2. **(1 pt)** What is the TLB used for?
3. **(1 pt)** Why is important to have a TLB?
4. **(Bonus +1 pt)** As you notice, CI+CO matches PPO that matches VPO. This is no coincidence. How does the memory system exploit this?
5. **(3x4 pts)** Translate the virtual addresses 0x03D5, 0x0B8E, 0x0021 into their physical addresses according to Figure 4. Indicate if you have hits, misses, or page faults. If you cannot continue with the information in the table, indicate what the memory system will do. Use the physical address to access the corresponding byte in the cache. Indicate if you have a cache hit or not. Indicate what the memory system will do if you cannot read the data. Answer by clearly enumerating the different stages.





## Simple Memory System TLB

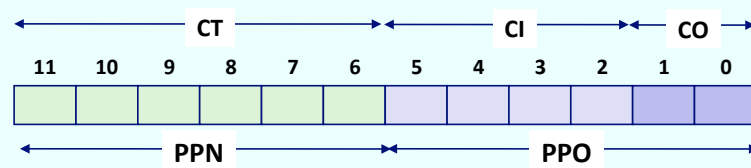
- 16 entries
- 4-way associative



Set	Tag	PPN	Valid	Tag	PPN	Valid	Tag	PPN	Valid	Tag	PPN	Valid
0	03	–	0	09	0D	1	00	–	0	07	02	1
1	03	2D	1	02	–	0	04	–	0	0A	–	0
2	02	–	0	08	–	0	06	–	0	03	–	0
3	07	–	0	03	0D	1	0A	34	1	02	–	0

## Simple Memory System Cache

- 16 lines, 4-byte block size
- Physically addressed
- Direct mapped



Idx	Tag	Valid	B0	B1	B2	B3
0	19	1	99	11	23	11
1	15	0	–	–	–	–
2	1B	1	00	02	04	08
3	36	0	–	–	–	–
4	32	1	43	6D	8F	09
5	0D	1	36	72	F0	1D
6	31	0	–	–	–	–
7	16	1	11	C2	DF	03

Idx	Tag	Valid	B0	B1	B2	B3
8	24	1	3A	00	51	89
9	2D	0	–	–	–	–
A	2D	1	93	15	DA	3B
B	0B	0	–	–	–	–
C	12	0	–	–	–	–
D	16	1	04	96	34	15
E	13	1	83	77	1B	D3
F	14	0	–	–	–	–

Figure 4: Simple memory system.

## A First Powers of 2

$2^0$	$= 1$
$2^1$	$= 2$
$2^2$	$= 4$
$2^3$	$= 8$
$2^4$	$= 16$
$2^5$	$= 32$
$2^6$	$= 64$
$2^7$	$= 128$
$2^8$	$= 256$
$2^9$	$= 512$
$2^{10}$	$= 1024 = 1k$
$2^{11}$	$= 2048 = 2k$
$2^{12}$	$= 4096 = 4k$