



Contact:  
Magnus Jahre (952 22 309)

## TDT4260/DT8803 COMPUTER ARCHITECTURE EXAM

Monday 4. June  
Time: 09:00 – 13:00  
ENGLISH

Allowed Aids:

D.

No written or handwritten examination support materials are permitted.

A specified, simple calculator is permitted.

*Use the provided space to answer the problems. If you need more space, an extra answer box is available on the last page of the exam. The exam accounts for 80% of the final grade, and the provided points show the maximal number of points that can be achieved on each assignment. Read the problem texts thoroughly. You can answer the questions in English or Norwegian.*

**Candidate Number:**

**Problem 1 Multiple Choice (30 points)**

Answer by circling the answer alternative you believe is the correct answer. You are awarded 3 points for a correct answer and 0 points if you do not answer. If your answer is wrong or you circle more than one alternative, you will get -1 point.

a) (3 p) Which of the following trends are *not* correct?

1. The number of transistors on a chip increases by between 40% and 55% per year
2. Single processor performance improved by 52% per year between 1986 and 2000
3. Single processor performance is currently doubling every two years
4. Dynamic Random Access Memory (DRAM) latency improves by 7% per year

Answer:            1                            2                            3                            4

b) (3 p) Which of the following performance improvement techniques does *not* exploit parallelism?

1. Increasing the clock frequency by increasing the pipeline depth
2. Reducing the cache access latency by making the cache smaller
3. Increasing instruction throughput by adding more functional units
4. Improving system performance by adding an additional processor core

Answer:            1                            2                            3                            4

c) (3 p) Which of the following categories is *not* part of Flynn's taxonomy?

1. SISD
2. SIMD
3. SIMT
4. MISD

Answer:            1                            2                            3                            4

**Candidate Number:**

d) (3 p) Assume that 5% of the runtime of a certain application cannot be parallelized. What is the maximum possible speedup by parallelization for this application if we assume that the input set is fixed?

1. 2
2. 10
3. 20
4. 100

Answer:            1                            2                            3                            4

e) (3 p) Assume the same application as in Assignment 1 d). What happens to the speedup if we scale the data set and keep execution time fixed?

1. The speedup is significantly higher than with a fixed data set
2. The speedup is roughly the same as with a fixed data set
3. The speedup is significantly lower than with a fixed data set
4. Scaling the data set is impossible for most practical parallel applications

Answer:            1                            2                            3                            4

f) (3 p) Which of the following ratios is *not* part of the CPU performance equation (also known as the “Iron Law”)?

1. Instructions/Program
2. Program/Seconds
3. Clock Cycles/Instruction
4. Seconds/Clock Cycle

Answer:            1                            2                            3                            4

**Candidate Number:**

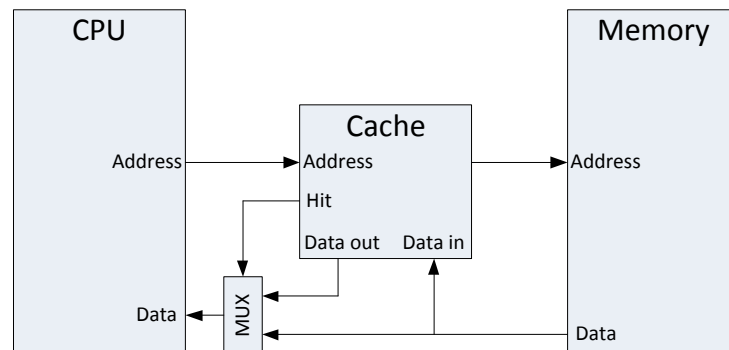


Figure 1: Cache System Block Diagram A

g) (3 p) Which of the following statements are *not* currently considered to be a challenge for future multi-core architectures?

1. The power budget of the chip is fixed and will remain so in the future
2. Architects need to control memory hierarchy power consumption to meet the power budget
3. Energy efficient parallel software must be developed
4. Architects need to find new aggressive ILP techniques to improve single thread performance

Answer:            1                            2                            3                            4

h) (3 p) Consider the system outlined in Figure 1. Assume a 1 clock cycle cache latency, a 20 clock cycle memory latency and that the cache access must be completed before memory is accessed. What is the average access time for an application with a 90% cache hit rate?

1. 1.0
2. 2.9
3. 3.0
4. 20.0

Answer:            1                            2                            3                            4

**Candidate Number:**

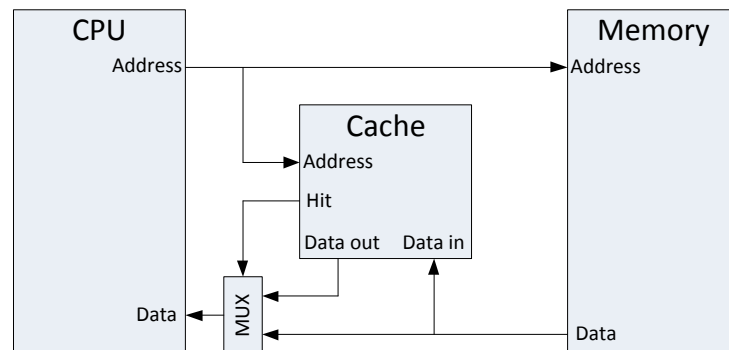


Figure 2: Cache System Block Diagram B

i) (3 p) Consider the system outlined in Figure 2. Assume a 1 clock cycle cache latency and a 20 clock cycle memory latency. What is the best average access time for an application with a 90% cache hit rate?

1. 1.0
2. 2.9
3. 3.0
4. 20.0

Answer:            1                            2                            3                            4

j) (3 p) Which statement about cache coherency protocols is *not* correct?

1. Write invalidate protocols are much more common than write update protocols.
2. Write update protocols consume more bandwidth than write invalidate protocols.
3. Snooping cache coherency protocols use a broadcast interconnect as the point of serialization
4. It is impossible to implement a directory-based cache coherency protocol in a system with a bus interconnect


Answer:            1                            2                            3                            4

**Candidate Number:**

**Problem 2    Memory Systems (10 points)**

- a) (5 p) Draw a block diagram of a 2-way set associative cache. Explain any necessary assumptions.

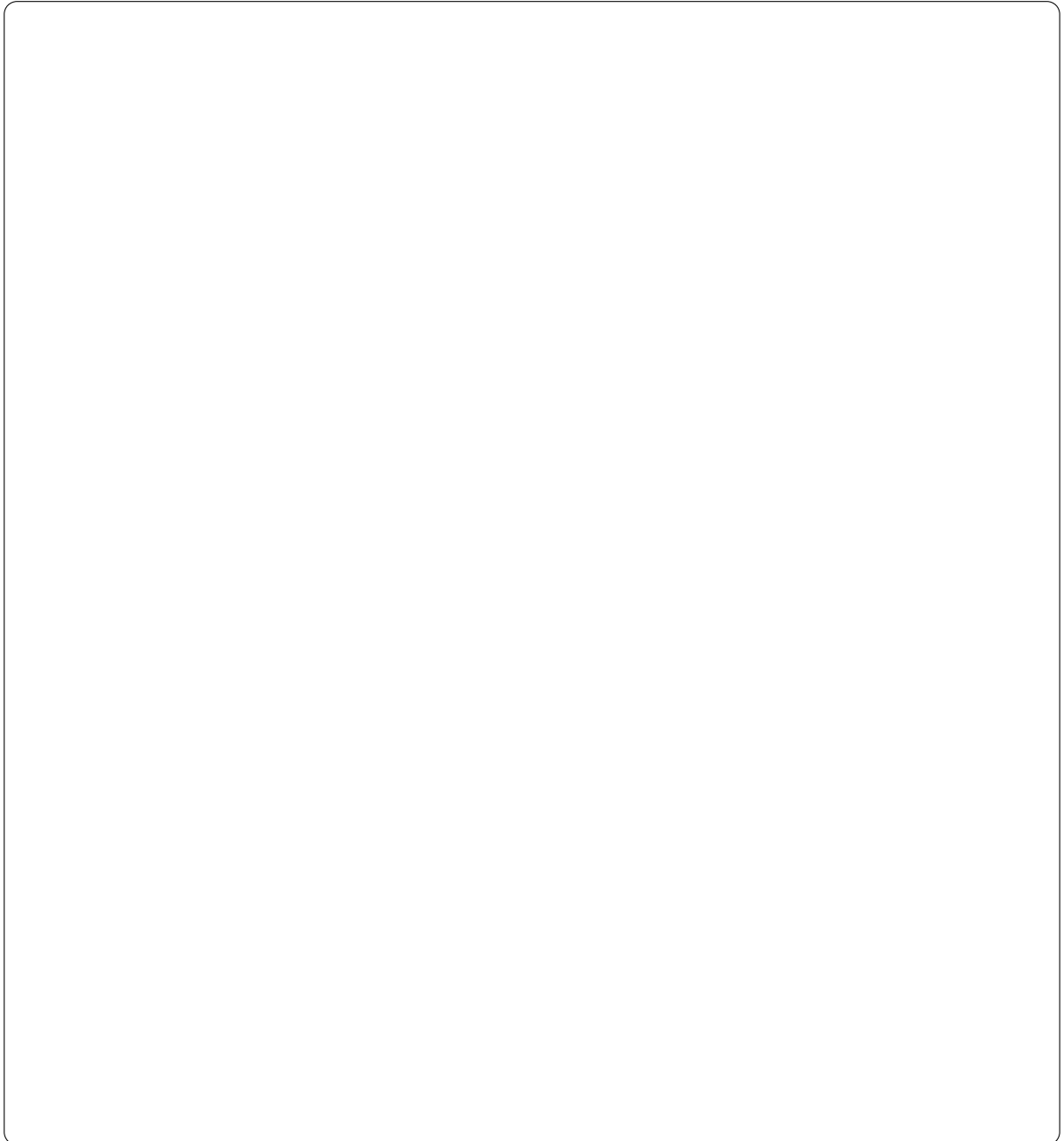
Answer:



**Candidate Number:**

- b) (5 p) Draw a block diagram of a system that implements the virtually indexed/physically tagged optimization. Explain any necessary assumptions.

Answer:



**Candidate Number:**

1	LW	R1, 0(R2)
2	SUB	R4, R1, R5
3	AND	R1, R4, R5

Figure 3: Assembly Code 1

**Problem 3 Instruction Level Parallelism (10 points)**

- a) (5 p) Find and name all dependencies in Figure 3. The first register in the operand list is output for all instructions.

Answer:

**Candidate Number:**



1	LW	R6, 32(R1)
2	LW	R2, 36(R1)
3	MUL	R0, R2, R4
4	SUB	R8, R2, R6
5	DIV	R10, R0, R6
6	ADD	R7, R8, R2

Figure 4: Assembly Code 2

Format 1:	LW/SW	LW/SW	ADD/SUB	ADD/SUB
Format 2:	MUL/DIV	MUL/DIV	ADD/SUB	ADD/SUB

Table 1: VLIW Instruction Format

- b) (5 p) A VLIW instruction format is given in Table 1. Each slot can be filled with an instruction of the given type or a no-operation (NOP). Find a schedule of the code in Figure 4 that minimizes execution time. State any necessary assumptions.

Answer:

**Candidate Number:**

1	LV	V1, Rx	;load vector x
2	MULVS.D	V2, V1, F0	;vector-scalar multiply
3	LV	V3, Ry	;load vector y
4	ADDVV.D	V4, V2, V3	;add x and y
5	SV	V4, Ry	;store the sum

Figure 5: Vector Assembly Code for DAXPY

**Problem 4 Data-Level Parallelism (10 points)**

Assume a vector computer with 1 vector load/store unit, 1 vector add/subtract unit, 1 vector multiply/divide unit and a vector length of 64. The vector computer implements *chaining*, and accessing off-chip memory takes 20 clock cycles. You can ignore effects resulting from start-up latencies in this assignment.

- a) (5 p) What is lowest execution time in clock cycles of the code Figure 5? Make any necessary assumptions.

Answer:

**Candidate Number:**

- b) (5 p)** How many memory banks are needed to sustain maximum performance for this machine?  
Explain your reasoning.

Answer:



**Candidate Number:**

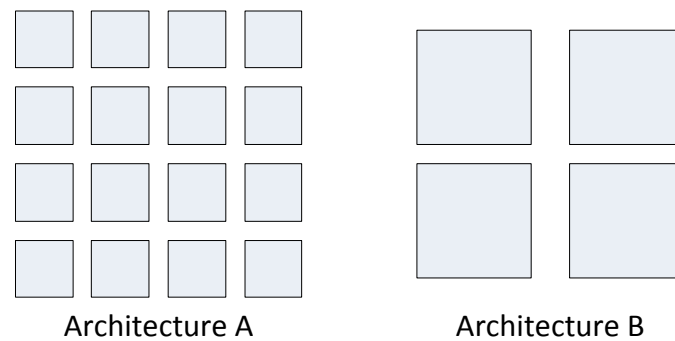


Figure 6: Multi-core Processors

**Problem 5 Thread-Level Parallelism (10 points)**

Figure 6 shows two multi-core processor architectures (Architecture A and Architecture B) that have the same area and power budget. For simplicity, assume that the performance of each of the small cores in Architecture A is 1 performance unit and that it consumes 1 power unit. A processor with unit 1 performance is able to execute 1 work unit every time unit.

- a) (5 p) What is the expected performance of a processor core in Architecture B? Make any necessary assumptions and explain your reasoning.

Answer:

**Candidate Number:**

- b) (5 p) Assume a scalable parallel program with 32 work units. Which architecture is most suitable for this program in terms of performance and power efficiency? Explain your reasoning and make any necessary assumptions.

Answer:



**Candidate Number:**

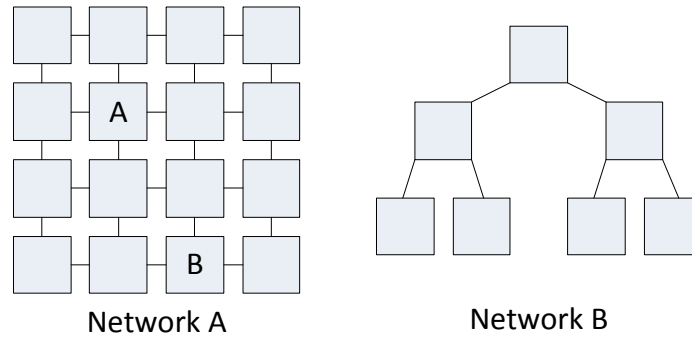


Figure 7: Interconnection Network Examples

**Problem 6    Interconnection Networks (10 points)**

- a) (5 p) Assume Network A in Figure 7, the information in Table 2 (on page 15) and a message size of 1024 bytes. What is the lowest transfer latency in clock cycles between points A and B? Make any necessary assumptions, and explain your reasoning. Control and status information is transmitted over separate lines and can be ignored in this assignment.

Answer:


**Candidate Number:**

Data	Value
Link latency	1 cycle
Router latency	1 cycle
Link data width	64 byte
Clock Frequency	2 GHz

Table 2: Network Properties

- b) (5 p) Find the bisection bandwidth of networks A and B in Figure 7 using the information in Table 2. Make any necessary assumptions.

Answer:



**Candidate Number:**

**Additional Answer Space**

Answer:



**Candidate Number:**