

ארכיטקטורת מחשבים

תשע"ד סמס' ב' מועד ב'

83-248

- מרצה : פרופ' שמואל וימר
- מתרגל : מר יוסף לונדון
- **חומר עזר מותר:** מחברות/ דפים מהרצאות, תרגולים ותרגילי בית, ספר computer organization and design, מילון עברי צרפתי.
- משך המבחן שלוש שעות
- סך כל הנקודות הוא 110. הציון המרבי הינו 100.
- משקל השאלות השונות נתון בגוף השאלות.
- **יש לנמק את כל תשובותיכם.** אין צורך לפתח מחדש תוצאות שהוכחו בכיתה, אלא אם כן נאמר מפורשות לעשות כן.
- יש לשרטט דיאגרמות באופן בחר !
- הכתיבה בעט בלבד. כתיבה בעפרון לא תיבדק.

בהצלחה!

שאלה 1 (20 נקודות)

נתון מעבד MIPS מצונר עם תמיכה מלאה ב HAZARDS, כמופיע בציר המצורף. נתון BENCHMARK כדלהלן:

25% פקודות קריאה (LOADS), ב 40% מהן מופיעה מיד פקודה המשתמשת בנתון שנקרא.
10% פקודות כתיבה (STORES).
11% הסתעפויות (BRANCHES), ב 25% מתוכן ניחוש הקפיצה היה שגוי.
2% פקודות קפיצה. הנח שהפקודה שבאה מיד אחרי קפיצה מרוקנת (FLUSH) מהצינור.
52% פקודות אריתמטיות (R-TYPE).

חשבו את מספר מחזורי השעון הממוצע לפקודה (CPI) במעבד הנ"ל.

- יש לתאר באופן מפורט את החישוב.
- יש להסביר כל מספר שמשמש בחישוב.

Loads take one cycle when there is no dependency and two cycles when the processor must stall due to dependency. Therefore

$$CPI_{LOAD} = 0.6 \times 1 + 0.4 \times 2 = 1.4.$$

Since branch condition is checked at DECODE, miss prediction requires flushing the IF/ID register, so two cycles are charged, while branches take one cycle when they are properly predicted.

$$CPI_{BRANCH} = 0.75 \times 1 + 0.25 \times 2 = 1.25.$$

$$CPI_{JUMP} = 2.$$

$$CPI_{total} = 0.25 \times 1.4 + 0.1 \times 1 + 0.11 \times 1.25 + 0.02 \times 2 + 0.52 \times 1 = 1.1475.$$

שאלה 2 (45 נקודות)

עבור מעבד ה MIPS משאלה 1, נתונה טבלת ההשהיות הבאה של מרכיביו:

Element	Parameter	Delay (pSec)
register clock-to-Q	t_{pcq}	30
register setup	t_{setup}	20
multiplexer	t_{mux}	25
equality comparator	t_{eq}	40
AND gate	t_{AND}	15
ALU	t_{ALU}	200
memory read	$t_{MEMread}$	250
memory write	$t_{MEMwrite}$	220

register-file read	t_{RFread}	150
register-file write	$t_{RFwrite}$	100
register-file setup	$t_{RFsetup}$	20

א. חשבו מהו תדר השעון המרבי בו ניתן להפעיל את המעבד.

- תארו בפירוט מלא את חישוביכם.
 - רשמו ביטוי מפורש לזמן מחזור השעון במונחי הפרמטרים המופיעים בטבלה.
 - לאחר מכן השתמשו בערכיהם לחישובים מספריים.
- הנחיות:

1. היעזרו בתרשימים המצורפים וסמנו לכל דרגה ודרגה את מסלול החישוב הארוך ביותר. שימו לב שזהו תרשים של MIPS שנלמד בקורס, בתוספת MUX במוצא של ה REGISTER FILE, המשמשים לניתוב נתונים מקודמים (FORWARD) במקרה של HAZARD.

2. שימו לב שנתוני WRITE BACK נכתבים באוגר ה PIPELINE בצלע העולה של השעון, ונכתבים ל REGISTER FILE בצלע היורדת של השעון, ולכן שני הדברים קורים באותו מחזור שעון אבל בשני חצאים שונים. יש לקחת זאת בחשבון בחשוב תדר השעון המרבי האפשרי.

The clock cycle is dictated by the longest delay path of the processor.

Consider first the critical path at every pipeline stage. Such paths always start at the clock of a register and terminate at the data input of a register.

Instruction fetch stage:

$$t_{IF} = t_{pcq} + t_{MEMread} + t_{setup} = 30 + 250 + 20 = 300 \text{ pSec}$$

Decode stage:

Notice that the register file is written in the first half of the writeback cycle and is being read at the second half of the decode cycle. Therefore,

$$t_{ID} = 2 \times (t_{RFread} + t_{mux} + t_{eq} + t_{AND} + t_{mux} + t_{setup}) = 2(150 + 25 + 40 + 15 + 25 + 20) = 550 \text{ pSec}$$

Execution stage:

$$t_{EXE} = t_{pcq} + t_{mux} + t_{mux} + t_{ALU} + t_{setup} = 30 + 25 + 25 + 200 + 20 = 300 \text{ pSec}$$

Memory stage:

$$t_{MEM} = t_{pcq} + \max\{t_{MEMread}, t_{MEMwrite}\} + t_{setup} = 30 + 250 + 20 = 300 \text{ pSec}$$

Writeback stage:

Notice that the register file is written in the first half of the write-back cycle and is being read at the second half of the decode cycle. Therefore,

$$t_{WB} = 2 \times (t_{pcq} + t_{mux} + t_{RFwrite}) = 2(30 + 25 + 100) = 310 \text{ pSec}$$

$$T_{CLK} = \max\{t_{IF}, t_{ID}, t_{EXE}, t_{MEM}, t_{WB}\} = 550 \text{ pSec}.$$

ב. כמה זמן תימשך ריצת אוסף תכניות המכילות 100 ביליון פקודות מכונה המתפלגות ע"פ שאלה 1?

$$T_{\text{total}} = (100 \text{ billion instructions}) \times (1.15 \text{ cycles/instruction}) \times (550 \text{ pSec}) = 10^{11} \times 1.15 \times 550 \times 10^{-12} = 63.3 \text{ Sec}$$

ג. נמדוד את ביצועי המעבד ע"י מספר הפקודות לשנייה בתדר שעון מרבי. בניסיון לשפר ביצועים, ראובן, מהנדס מתחיל בחברה (בוגר בר-אילן) מציע להעביר את בדיקת תנאי השוויון של קפיצה מותנית משלב ה ID לשלב ה EXE, ואילו שמעון, מהנדס וותיק בחברה (בוגר אוניברסיטה אחרת) מציע להעביר את בדיקת תנאי השוויון לשלב ה MEM.

- האם כל אחת מההצעות הנ"ל תשפר, תקלקל או לא תשפיע על הביצועים? הסבר בפרוטרוט את תשובתך.
- במידה והביצועים משתנים, חשב באופן מספרי פי כמה.

The critical delay paths of the ID stage gets shorter by $t_{\text{eq}} + t_{\text{AND}} + t_{\text{mux}} = 80 \text{ pSec}$.

$$t_{\text{ID}} = 2 \times (t_{\text{RFread}} + t_{\text{mux}} + t_{\text{setup}}) = 2(150 + 25 + 20) = 390 \text{ pSec}$$

The critical path in the EXE stage gets longer by 80pSec at most in the first case and the critical paths in the MEM gets longer by 80pSec at most in the second case. Consequently, t_{ID} still dictates the clock cycle.

$$\text{Perf}_{\text{branch_ID}} = (1/(550 \times 10^{-12}))/1.1475 = 1.584 \times 10^9 \text{ instructions/Sec}$$

If equality check takes place at **EXE stage** two instructions are flushed in miss

$$\text{CPI_BRANCH} = 0.75 \times 1 + 0.25 \times 3 = 1.5.$$

$$\text{CPI} = 0.25 \times 1.4 + 0.1 \times 1 + 0.11 \times 1.5 + 0.02 \times 2 + 0.52 \times 1 = 1.175.$$

$$\text{Perf}_{\text{branch_EXE}} = (1/(390 \times 10^{-12}))/1.175 = 2.182 \times 10^9 \text{ instructions/Sec}$$

If equality check takes place at **MEM stage** three instructions are flushed in miss

$$\text{CPI_BRANCH} = 0.75 \times 1 + 0.25 \times 4 = 1.75.$$

$$\text{CPI} = 0.25 \times 1.4 + 0.1 \times 1 + 0.11 \times 1.75 + 0.02 \times 2 + 0.52 \times 1 = 1.2025.$$

$$\text{Perf}_{\text{branch_MEM}} = (1/(390 \times 10^{-12}))/1.2025 = 2.132 \times 10^9 \text{ instructions/Sec}$$

שאלה 3 (15 נקודות)

נתונים המאפיינים הבאים של מערכת זיכרון:

זיכרון מטמון בן 256 בתים,

זיכרון ראשי 1024 בתים,

גודל בלוק זיכרון המטמון 4 בתים.

תאר את תצורת הזיכרון עבור

א. זיכרון מטמון בעל מיפוי ישיר (Direct-mapped)

ב. 2-way set-associative

ג. 4-way set-associative

ד. 8-way set-associative

- ציינו את מספר הסיביות בשדות השונים tag, index, offset של כתובות הזיכרון בכל אחד מזיכרונות המטמון.
- לאילו כתובות בזיכרון המטמון ממופות הכתובות כתובת 0, כתובת 100, כתובת 256 בזיכרון הראשי בכל אחת מתצורות זיכרון המטמון הנ"ל?
- רשמו בפירוט ובמדויק כיצד חישבת את הכתובות בכל אחד מהמקרים.

Direct-mapped

main memory address 0

main memory block #: $0/4 = 0$

cache block #: $0 \% 64 = 0$

main memory address 100

main memory block #: $100/4 = 25$

cache block #: $25 \% 64 = 25$

main memory address 256

main memory block #: $256/4 = 64$

cache block #: $64 \% 64 = 0$

main memory address (10 bits)

tag	index	offset
x x	x x x x x x	x x

2-way set-associative

main memory address 0

main memory block #: $0/4 = 0$

cache set #: $0 \% 32 = 0$; cache block #: 0, 1

main memory address 100

main memory block #: $100/4 = 25$

cache set #: $25 \% 32 = 25$; cache block #: 50, 51

main memory address 256

main memory block number: $256/4 = 64$

cache set #: $64 \% 32 = 0$; cache block #: 0, 1

main memory address (10 bits)

tag	index	offset
x x x	x x x x x	x x

4-way set-associative

main memory address 0

main memory block #: $0/4 = 0$

cache set #: $0 \% 16 = 0$; cache block #: 0,1,2,3

main memory address 100

main memory block #: $100/4 = 25$

cache set #: $25 \% 16 = 9$; cache block #: 36, 37, 38, 39

main memory address 256

main memory block #: $256/4 = 64$

cache set #: $64 \% 16 = 0$; cache block #: 0,1,2,3

main memory address (10 bits)

tag	index	offset
x x x x	x x x x	x x

8-way set-associative

main memory address 0

main memory block #: $0/4 = 0$

cache set #0 % 8 = 0 ; cache block #: 0,1,2,3,4,5,6,7

main memory address 100

main memory block #: $100/4 = 25$

cache set #: $25 \% 8 = 1$; cache block #: 8,9,10,11,12,13,14,15

main memory address 256

main memory block #: $256/4 = 64$

cache set #: $64 \% 8 = 0$; cache block #: 0,1,2,3,4,5,6,7

main memory address (10 bits)

tag	index	offset
x x x x x	x x x	x x

שאלה 4 (30 נקודות)

נתונה מערכת זיכרון במעבד הפועל בתדר 500 MHz בעלת שתי רמות של זיכרון מטמון. L1-data cache הינו direct-mapped, write-through, בגודל כולל של 8KByte וגודל בלוק של 8Byte. מניחים שחוצץ הכתיבה שלו מושלם ואין אף פעם stalls. נתון שה miss-rate הינו 15%.

L1-instruction cache הינו direct-mapped, בגודל כולל של 4KByte וגודל בלוק של 8Byte. נתון שה miss-rate הינו 2%.

L2 הינו יחיד ומשותף, 2-way set associative, write-back, בגודל כולל של 2MByte וגודל בלוק של 32 Byte. נתון שה miss-rate הינו 10%. נתון שבממוצע 50% מהבלוקים ב L2 הינם "מלוכלכים", כלומר רשום בהם מידע שאיננו כרגע בזיכרון הראשי.

הנח ש 40% מהפקודות הינן פקודות גישה לזיכרון, 60% מתוכן קריאה (LOAD) ו 40% מתוכן כתיבה (STORE).

L1 hits אינם גורמים ל stalls. זמן גישה ל L2 הינו 20 ננו שניות. זמן גישה לזיכרון הראשי הינו 0.2 מיקרו שניה, ומרגע זה מספר מילים כרוחב ה memory bus נשלחות כל מחזור שעון. רוחב ה BUS המחובר בין L2 לזיכרון הראשי הינו 128 סיביות.

א. איזה אחוז מתוך גישות הנתונים לזיכרון מגיע לזיכרון הראשי?
- רשמו נוסחה מפורשת ולאחריה חישוב מספרי מדויק.

$$= (L1 \text{ miss rate}) \times (L2 \text{ miss rate}) = 0.15 \times 0.1 = 1.5\%$$

ב. חזרו על סעיף א', בהנחה שכל הפניות לכתיבה הינן L1 misses.

$$= (\% \text{ of data ref writes}) \times (L2 \text{ miss rate}) + (\% \text{ of data ref reads}) \times (L1 \text{ miss rate}) \times (L2 \text{ miss rate}) = (0.4) \times (0.1) + (0.6) \times (0.15) \times (0.1) = 4.9\%$$

ג. כמה סיביות בכל אחד מזיכרונות המטמון משמשות לאינדקס? רשמו חישוב מפורט.

$$L1 \text{ Data: } 8K\text{byte}/8\text{Byte} = 1024 \text{ blocks} \Rightarrow 10 \text{ bits}$$

$$L1 \text{ Instruction: } 4K\text{byte}/8\text{Byte} = 512 \text{ blocks} \Rightarrow 9 \text{ bits}$$

$$L2: 2M\text{Byte}/32\text{Bytes} = 64K \text{ blocks} = 32K \text{ sets} \Rightarrow 15 \text{ bits}$$

ד. מה מספר מחזורי השעון **המרב** שעשוי להידרש בעת גישה לזיכרון הראשי?
- ראשית הסבירו בדיוק את רצף האירועים המתרחש במצב קיצוני כזה.
- רשמו במדויק את חישוב מספר מחזורי השעון.

Maximum clock cycles occurs when L1 missed first, then L2 missed, then write-back takes place.

$$L2 \text{ access cycles: } (20 \text{ nano second}) / (2 \text{ nano second}) = 10 \text{ cycles}$$

$$\text{Main memory access cycles: } (0.2 \text{ micro second}) / (2 \text{ nano second}) = 100 \text{ cycles}$$

Since a block is 32 Bytes and memory bus is 128 bits (16 Bytes), two bus transactions of 16 Bytes each are required. The first 16 bytes takes 100 cycles, the next 16 bytes takes one cycle.

Notice that getting a new block from the memory may evict a block from L2, which is a write-back. In that case the evicted block must be written into the memory, requiring a total of L2-memory write-back $2 \times (100 + 1) = 202$ cycles. Summing all
 $L1 \text{ miss} + L2 \text{ miss} + \text{write-back} = 1 + 10 + 202 = 213 \text{ cycles}$

ה. מהו מספר מחזורי השעון הממוצע בגישה לזיכרון (AMAT) כולל פקודות ונתונים ?
 - רשמו בדיוק את המשוואות בהן אתם משתמשים.

The weight of instruction accesses to memory is $1/(1 + 0.4)$, while the weight of data accesses is $0.4/(1 + 0.4)$. Therefore
 $AMAT_{total} = 1/1.4 AMAT_{inst} + 0.4/1.4 AMAT_{data}$

For any 2-level cache system there is
 $AMAT = (L1 \text{ hit time}) + (L1 \text{ miss rate}) \times (L2 \text{ hit time}) + (L1 \text{ miss rate}) \times (L2 \text{ miss rate}) \times (\text{main memory transfer time})$.

Memory transfer time must account for the average percentage of L2 dirty blocks, which for the given L2 means that 50% of the blocks must be updated in main memory upon L2 miss, yielding a factor of 1.5 multiplying $(100 + 1)$.

$$AMAT_{inst} = 1 + 0.02 \times 10 + 0.02 \times 0.1 \times 1.5 \times (100 + 1) = 1.503$$

$$AMAT_{data} = 1 + 0.15 \times 10 + 0.15 \times 0.1 \times 1.5 \times (100 + 1) = 4.7725$$

$$AMAT_{total} = 1/1.4 \times 1.503 + 0.4/1.4 \times 4.7725 = 2.44$$

ו. חזרו על סעיף ה', בהנחה שכל הפניות לכתיבה הינן L1 misses.

$$AMAT_{total} = 1/1.4 AMAT_{inst} + (0.4 \times 0.6)/1.4 AMAT_{loads} + (0.4 \times 0.4)/1.4 AMAT_{stores}$$

$$AMAT_{inst} \text{ unchanged}$$

$$AMAT_{loads} \text{ unchanged}$$

$$AMAT_{stores} = 1 + 1 \times 10 + 1 \times 0.1 \times 1.5 \times (100 + 1) = 26.15$$

$$AMAT_{total} = 1/1.4 \times 1.503 + 0.24/1.4 \times 4.7725 + 0.16/1.4 \times 26.15 = 4.88$$

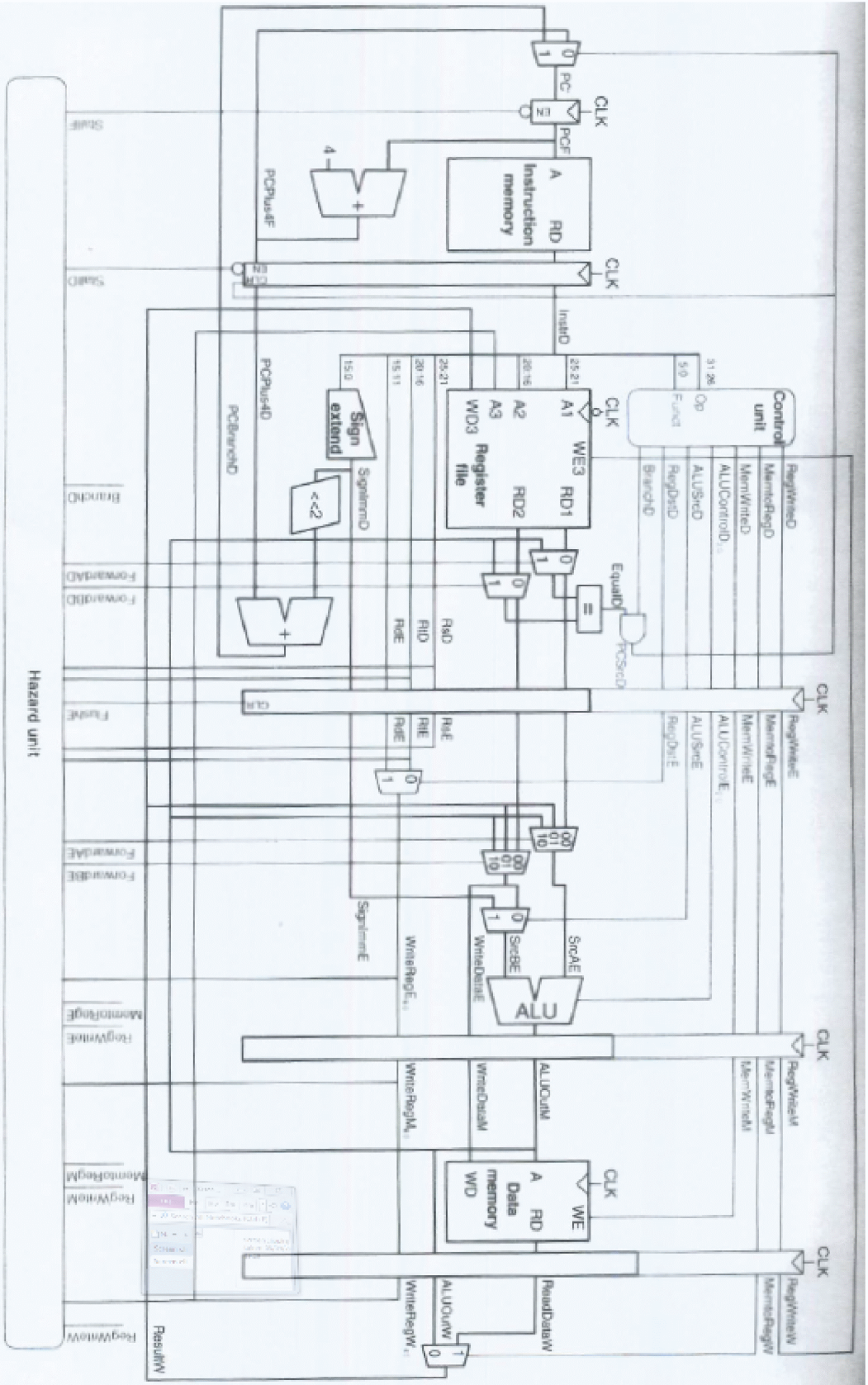


Figure 7.58 Pipelined processor with full hazard handling