



הטכניון - מכון טכנולוגי לישראל

## מבנה מחשבים ספרתיים (234267)

מבחן מסכם מועד א'

2 פברואר 2017

**מרצים:** ליהוא רפופורט, עדי יועז.

**מתרגלים:** פרנק סלה, איתי רביד.

שם :	_____
מס. ת.ז. :	_____

- משך הבחינה: שלוש שעות.
- מותר כל חומר עזר.
- יש לכתוב את התשובות בטופס הבחינה בלבד ובמקום המיועד לתשובה.
- יש לכתוב בקיצור ככל האפשר, אך יש לנמק כל תשובה.
- בדקו שבטופס שבידכם יש 13 עמודים כולל עמוד זה.
- המבחן כולל 5 שאלות, יש לענות על כולן.

שאלה 1	/ 25
שאלה 2	/ 24
שאלה 3	/ 28
שאלה 4	/ 15
שאלה 5	/ 8
ציון סופי	/100

**בהצלחה !**

## שאלה 1 – זיכרון וירטואלי (25 נק')

נתון מעבד דמוי 86x העובד במוד של 64 ביט ומבנה הכתובת הבא:

63	44	43	40	39	36	35	32	31	0
Sign Ext	PML3	PML2	PTE	offset					

- גודל כניסה בטבלאות הדפים היא 4 בתים
  - במעבד קיים TLB גדול. במקרה של TLB miss פונים ל-PMH.
  - ב-PMH ישנם translation caches גדולים עבור כל אחת מרמות התרגום PML2-3 שהגישה אליהם מתבצעת במקביל.
  - במעבד קיים data cache בגודל 32KB, 4 way set associative, וגודל שורה של 32 בתים. הניחו שלאחר ששורה מובאת ל-cache, היא לא נזרקת ממנו במהלך סידרת הפניות.
- א. (8 נק') נתונה סידרת פניות לכתובות וירטואליות (בבסיס 16). עבור כל אחת מהפניות יש לפרט:
- עבור כל אחד מה-translation caches וה-TLB האם הניב hit או miss, או שלא ניגשו אליו.
  - מספר הגישות ל-Data cache שהסתיימו ב-hit, ומספר הגישות שהסתיימו ב-miss (עבור התרגום בלבד, לא כולל הפניות להבאת הנתון עצמו).
- הניחו כי בתחילת הסידרה כל ה-caches ריקים.

D\$ num misses	D\$ Num hits	PML3 hit/miss/ n.a.	PML2 hit/miss/ n.a.	TLB hit/ miss	כתובת
					FFFF FBA9 8765 4321
					FFFF FBA <u>A</u> 8765 4321
					FFFF FB <u>B</u> 9 8765 4321
					FFFF F <u>C</u> A9 8765 4321
					FFFF FB <u>C</u> 9 8765 4321

ב. (4 נק') מהו הגודל בבתים של טבלאות הדפים בעץ (המינימלי) שפורש את סידרת הכתובות שבסעיף ב' ? הסבירו

---

---

---

---

---

---

---

---

---

---

ג. (4 נק') בגישה לנתון, האם ניתן לפנות ל-set של ה-cache במקביל לתרגום הכתובת מווירטואליות לפיזיות? הסבירו

---

---

---

---

---

---

---

---

---

---

ד. (5 נק') מה הם השינויים הנדרשים על מנת להגדיל את הכתובת הפיזית הנתמכת ל-50 סיביות ?  
במידה ונדרש גם שינוי במבנה הכתובת, הוסיפו ציור של מבנה הכתובת החדש.

---

---

---

---

---

---

ה. (4 נק') ביחס לשאלה המקורית (ללא התחשבות בסעיף ד'), מה הם השינויים הנדרשים על מנת להגדיל את המרחב הוירטואלי ל-52 סיביות, מבלי לשנות את גודל הדפים ?  
במידה ונדרש גם שינוי במבנה הכתובת, הוסיפו ציור של מבנה הכתובת החדש.

---

---

---

---

---

---

## שאלה 2 – זיכרון Cache (24 נק')

נתון מעבד בעל הירארכית הזיכרון הבאה:

- (I\$) L1 I-Cache: Direct Mapped, 64 בתים בשורה, גודל 8KBytes. מוכל ב-L2.
- (D\$) L1 D-Cache: 2-way set associative, 64 בתים בשורה, גודל 16Kbytes, מדיניות החלפה LRU. מוכל ב-L2.
- L2 cache: 2-Way set associative, 64 בתים בשורה, גודל 512KBytes, מדיניות החלפה LRU. ה-LRU מעודכן בכל פעם שיש פנייה ל-L2 cache כתוצאה מ-miss באחד מה-L1 caches.

נתון קטע הקוד הבא: (הקבועים הם בבסיס 10)

```
for (int k=1; i≤10; i++)  
  for (int i=0; i< 100; i++)      // rows  
    for (int j=0; j< 200; j++)    // columns  
      S += A[i][j]^k;
```

המערך  $A[100][200]$  הוא מערך דו-מימדי שבו כל איבר הוא בגודל 8 בתים.  $A$  מתחיל בכתובת  $0x800000$  (הקידומת  $0x$  מציינת מספר בבסיס 16).

$S, i, j$  מאוחסנים כל אחד ברגיסטר.

הקומפיילר מייצר 16 פקודות מכונה למימוש קטע הקוד. כל פקודה היא באורך קבוע של 4 בתים והקוד מתחיל בכתובת  $0x4000000$ .

נתון כי בתחילת ביצוע התכנית כל זיכרונות המטמון ריקים.

א. (2 נק') המערך פרוש בקטע רציף בזכרון לפי שורות: תחילה כל האיברים של שורה  $i=0$ , אחריה כל האיברים של שורה  $i=1$ , וכו'. תנו נוסחה לכתובת של איבר  $A[i][j]$  בזיכרון.

---

---

---

ב. (3 נק') עבור ביצוע קטע הקוד כולו, מהו ה- hit rate ב-D\$? הסבירו

---

---

---

---

ג. (3 נק') עבור ביצוע קטע הקוד כולו, מה יהיה ה- hit rate ב-D\$ אם נהפוך את סדר הלולאות של i ו-j ? הסבירו

ד. (6 נק') עבור ביצוע קטע הקוד כולו, מהו ה- hit rate המדויק ב-\$1? הסבירו לשם פשטות נניח שהמעבד אינו pipelined, ושלפני כל פניה ל-\$D יש פניה בודדת לשורה ב-\$1, שבה מביאים את כל הקוד הנדרש באיטרציה הנוכחית, כולל הבאת הקוד מהלולאות החיצוניות (במידה ונדרש). מכאן גם שכמות הפניות הכוללת ל-\$1 שווה לכמות הפניות הכוללת ל-\$D.

ה. (4 נק') עבור ביצוע קטע הקוד כולו, מהו ה- hit rate ב-L2 cache ?  
לשם פשטות יש להזניח את ההשפעה של פניות להבאת קוד מה-L2 cache .

---

---

---

---

---

---

---

---

ו. (6 נק') כיצד ניתן לשנות את קטע הקוד, כך שה- hit rate ב-D\$ ישתפר משמעותית ?  
מהו ה- hit rate לאחר השינוי ? הסבירו

---

---

---

---

---

---

---

---

---

---

---

---

### שאלה 3 – Out-Of-Order Execution (28 נק')

א. (23 נק') יש למלא את הטבלה שבהמשך. לכל פקודה יש לרשום:

- |                  |   |  |
|------------------|---|--|
| <b>כבר מולאו</b> | { | • R3, R2, R1 – ערכי הרגיסטרים הארכיטקטוניים לאחר commit של הפקודה. |
|                  |   | • addr – כתובת הגישה לזיכרון – עבור פקודות load ו-store בלבד.      |
|                  |   | • data – ערך זיכרון שנקרא או נכתב – עבור פקודות load ו-store בלבד. |

- T alloc: הזמן בו מבוצעת אלוקציה: **עד 3 פקודות בכל מחזור**, החל מ- $t=1$ . ניתן לבצע אלוקציה רק כאשר **לכל הפקודות** שסיפק ה-frontend במחזור יש מקום הן ב-ROB והן ב-RS.
- ב-ROB יש **30 כניסות**, וב-RS יש **4 כניסות**. Store תופסת מקום אחד ב-ROB וב-RS.
- src2, src1: מספרי הרגיסטרים המשמשים כ-sources לפקודה:  $P_i$  עבור רגיסטר פיזי, ו- $R_i$  במידה וקוראים ישירות את הרגיסטר הארכיטקטוני.
  - עבור store: src1 – הרגיסטר המשמש לחישוב הכתובת. src2 – הרגיסטר המכיל את הנתון.
  - עבור פקודת JZ (jump if zero): פקודת קפיצה מותנית: בצע קפיצה אם Zero flag=1.
  - פקודת ALU (Add, Sub) שתוצאת החישוב שלה היא 0 מעדכנת את הדגל ל-1 ואחרת ל-0.
  - הדגל מהווה dst נוסף של פקודות ALU, ומשמש כ-src עבור פקודת JZ.
  - הדגל עובר renaming, בדומה לרגיסטרים: PZ<sub>i</sub> אם נכתב ע"י פקודה  $i$ , ו-Z אם ארכיטקטוני.
- T src2 ready, T src1 ready: הזמן בו מוכן כל אחד ערכי ה-sources לפקודה.
  - אם ה-src כבר מוכן בזמן האלוקציה, אז זמן זה יהיה שווה לזמן האלוקציה.
  - אחרת, זמן זה שווה ל-T data ready של הפקודה שמחשבת את הערך של ה-src.
- T exe: הזמן בו הפקודה נשלחת לביצוע. הניחו כי ישנן אינסוף יחידות ביצוע.
  - פקודה יכולה להיכנס לביצוע לכל המוקדם במחזור שלאחר האלוקציה.
  - פקודה נכנסת לביצוע במחזור השעון שלאחר המחזור בו כל ה-src-ים מוכנים. פקודת store נכנסת לביצוע במחזור השעון שלאחר המחזור בו src1 (המשמש לחישוב הכתובת) מוכן.
  - פקודת מוצאת מה-RS במחזור שלאחר הביצוע שלה ( $T_{exe}+1$ ), וכבר במחזור זה פקודה חדשה יכולה לבצע alloc ולהשתמש במקום שהתפנה.
- Load block code: עבור load שנשלח לביצוע בזמן  $t=T_{exe}$ , או שהוסר עבורו תנאי חסימה קודם בזמן  $t$ , תנאי החסימה נבדקים בזמן  $t+1$  לפי הסדר (רישמו את כל תנאי החסימה לפי הסדר):
  - 1 – חסימה כתוצאה מ-unknown store address
  - 2 – חסימה כתוצאה מ-waiting for store data
- T data ready:
  - עבור פקודות ALU:  **$T_{exe}+1$** .
  - עבור load שהוסרו עבורו כל תנאי החסימה בזמן  $t$  או שבוצע בזמן  $t$  ולא נחסם:
    - במידה וה-load פוגע ב-cache, או שיש store to load forwarding: בזמן  $t+3$ .
    - אחרת, במידה ובוצע load אחר לאותה שורה ב-cache בזמן  $t' < t$ : בזמן  $\max(t'+9, t+3)$ .
    - אחרת, בזמן  $t+9$ .



- עבור store: מחזור השעון בו הן ה-data לכתובה לזיכרון והן הכתובת מוכנים.
- כלומר  $T \text{ data ready} = \max(\text{Texe}+1, T \text{ src2 ready})$ .
- הכתובת של ה-store ידועה בזמן  $t = \text{Texe}+1$ . בזמן זה מוסר תנאי החסימה של load שנחסם ע"י ה-store על unknown store address. בפרט, load שמבוצע בזמן  $t$ , לא נחסם על unknown store address ע"י ה-store, אך load שמבוצע בזמן  $t < \text{Texe}+1$  נחסם.
- עבור פקודת Jump עם חיזוי שגוי, מבוצע flush בזמן  $\text{Texe}+1$ , והפקודות מהמסלול הנכון מבצעות אלוקציה החל מזמן **Texe+4** (במידה ואין סיבה אחרת שמעכבת את האלוקציה).
- T commit: הזמן בו הפקודה מבצעת commit. ניתן לבצע commit לעד **3 פקודות** בכל מחזור.
  - פקודה יכולה לבצע commit החל מזמן  $T \text{ data ready}+1$ .
  - פקודת מוצאת מה-ROB במחזור שלאחר commit ( $T \text{ commit}+1$ ), וכבר במחזור זה פקודה חדשה יכולה לבצע alloc ולהשתמש במקום שהתפנה.
- הנחות:
  - הכתובות הן פיזיות (אין צורך בתרגום). כל הערכים המספריים הם בבסיס 16.
  - L1 data cache: גודל שורה  $32_{10}B = 20_{16}B$ . write no allocate. ריק בתחילת הביצוע.
  - זמן האלוקציה של פקודה 2 הוא 2 וזאת מכיוון שה-frontend לא יכול היה לספק פקודה בזמן 1.
  - בטבלה רשומות אך ורק הפקודות מהמסלול הנכון.
- ב. (5 נק') עבור ריצת התוכנית שמצאתם בסעיף א', מלאו לכל מחזור שעון שבטבלה שבסעיף זה את נתוני ה-Top level breakdown.
  - שימו-לב שהן מחזורי השעון והן הנתונים הממולאים מתייחסים לזמן האלוקציה של הפקודות.
  - עמודות שמעבר לזמן האלוקציה האחרון בתוכנית יש להשאיר ריקות.
  - הניחו כי החל מהמחזור השני ה-frontend יכול לספק 3 פקודות למחזור.
  - הניחו כי במהלך המסלול השגוי אין חסימה של אלוקציה כתוצאה מ-RS או ROB מלאים.

Cycle	1	2	3	4	5	6	7	8	9	10	11
Back-end Stall											
Issue Slot 0											
Issue Slot 1											
Issue Slot 2											
Frontend Bound											
Backend Bound											
Retiring											
Bad Speculation											

Pdst	instruction	R1	R2	R3	addr	data	src1	src2	T alloc	T src1 ready	T src2 ready	T exe	block code	T data ready	T commit
0	load R3 $\leftarrow$ m[R2+10]	10	10	50	20	50									
1	add R2 $\leftarrow$ R2 + R1	10	20	50											
2	store m[R2+20] $\leftarrow$ R3	10	20	50	40	50			<b>2</b>						
3	sub R1 $\leftarrow$ R2 – 20	0	20	50											
4	load R2 $\leftarrow$ m[R1+40]	0	50	50	40	50									
5	store m[R3+10] $\leftarrow$ R1	0	50	50	60	0									
6	JZ 1000 wrongly predicted	0	50	50											
7	load R3 $\leftarrow$ m[R1+30]	0	50	40	30	40									
8	add R1 $\leftarrow$ R2 + R3	90	50	40											

## שאלה 4 – Power/Performance (15 נק')

נתון מעבד 1 בעל הנתונים הבאים:

- קיבול הדינאמי כפונקציה של ה-IPC של האפליקציה המורצת:  $C_{dyn} = IPC \times 700pF$ .
- שטח:  $4mm^2$

נתון מעבד 2 בעל הנתונים הבאים:

- קיבול הדינאמי כפונקציה של ה-IPC של האפליקציה המורצת:  $C_{dyn} = IPC \times 800pF$ .
- שטח:  $5mm^2$
- עבור אפליקציה נתונה, IPC גבוה ב-20% מזה של מעבד 1

עבור שני המעבדים:

- ההספק סטטי (Leakage Power) ליחידת שטח:  $0.3W/mm^2$ .
- נקודות מתח ותדר אפשריות: (בטבלה עמודות נוספות לשימושכם בפיתרון)

מתח ב Volt's	תדר ב Ghz			
0.60	1.0			
0.65	1.1			
0.70	1.3			
0.75	1.5			
0.80	1.8			
0.85	2.3			
0.90	2.8			
1	3.4			
1.1	4.0			

נתונה מערכת שההספק המוקצה בה עבור המעבד הוא  $5W$ .  
המערכת מריצה אפליקציה שעבורה מעבד 1 משיג IPC של 2.

נתון מדד הביצועים הבא:  $IPS (instruction per second) = IPC \times Frequency$ .

א. (5 נק') מהו ה-IPS של מעבד 1 במערכת הנתונה? יש לפרט את החישוב ולהסביר

---



---



---

ב. (5 נק') מהו ה-IPS של מעבד 2 במערכת הנתונה ? יש לפרט את החישוב ולהסביר

ג. (5 נק') נתון שהמערכת תומכת באופציית טורבו, כך שבמידה וקיים עודף תרמי, המערכת מקצה למעבד הספק של 10W למספר שניות. באיזה מהמעבדים תבחרו בהינתן פונקציית המטרה הבאה:  
 $IPS_{5W \times 0.8} + IPS_{10W \times 0.2}$  ? יש להסביר

## שאלה 5 – חיזוי קפיצות (8 נק')

נתונה סידרת הקפיצות המחזורית הבאה:  $11001001 \ 11001001 \dots$   
יהי  $N$  אורך ההיסטוריה המינימלי המאפשר חיזוי מושלם של הסידרה הנתונה במצב היציב.

א. (4 נק') מהו הוא אחוז החיזוי הנכון (במצב היציב) של חזאי בעל היסטוריה באורך  $N-1$  עבור הסידרה הנתונה? יש להסביר

---

---

---

---

---

---

---

ב. (4 נק') הראו שעבור אורך היסטוריה  $N$  יש חיזוי מושלם של הסידרה הנתונה במצב היציב.

---

---

---

---

---

---

---