



به موارد زیر توجه کنید:

- ۱- حتما نام و شماره دانشجویی خود را روی پاسخنامه بنویسید.
- ۲- کل پاسخ تمرینات را در قالب یک فایل pdf با شماره دانشجویی خود نام گذاری کرده در سامانه CW بارگذاری کنید.
- ۳- این تمرین ۱۰۰ نمره دارد که معادل ۵، ۰ نمره از نمره کلی درس است.
- ۴- در صورت مشاهده هر گونه مشابهت نامتعارف هر دو (یا چند) نفر **کل نمره** این تمرین را از دست خواهند داد.

۱- (۱۰ نمره) سه طراحی از حافظه نهان در اختیار ما است که همه آنها جمعا هشت کلمه داده ظرفیت دارند: C1 بلوکهای تک کلمه‌ای، C2 بلوکهای دو کلمه‌ای و C3 بلوکهای چهار کلمه‌ای دارد. با معیار نرخ فقدان (miss rate) کدام یک از این سه طرح بهتر است؟ اگر نرخ برخورد (hit-rate) در این سه طراحی به ترتیب ۵۰٪، ۶۰٪ و ۶۵٪ باشد و اگر هر فقدان منجر به ۲۵ سیکل ساعت تعلیق شود و زمان دستیابی در C1 معادل ۲ سیکل، در C2 معادل ۳ سیکل و در C4 معادل ۵ سیکل ساعت طول بکشد، کدامیک از آنها بهترین طراحی حافظه نهان به شمار می‌رود؟

پاسخ: معمولا هر چه اندازه بلوک بزرگتر باشد نرخ فقدان کمتر خواهد بود، بنابراین اگر تنها معیار ما برای انتخاب نرخ فقدان باشد، به نظر می‌رسد C3 بهترین است.

اما اگر بخواهیم بهترین طرح را با معیار دقیق‌تری اندازه بگیریم، بهتر است AMAT را محاسبه کنیم:

$$AMAT_1 = hit_time_1 + miss_rate_1 \times miss_penalty = 2 + 0.5 \times 25 = 14.5 \text{ Clock Cycle}$$

$$AMAT_2 = hit_time_2 + miss_rate_2 \times miss_penalty = 3 + 0.4 \times 25 = 13 \text{ Clock Cycle}$$

$$AMAT_3 = hit_time_3 + miss_rate_3 \times miss_penalty = 5 + 0.35 \times 25 = 13.75 \text{ Clock Cycle}$$

بنابراین با توجه به زمان متوسط دسترسی به حافظه، C2 طرح بهتری است.

۲- (۲۰ نمره) ظرفیت بخش داده حافظه نهان یک پردازنده ۲۵۶ بایت و هر بلوک از آن یک کلمه ۳۲ بیتی است. (منظور این است که حجم داده‌ای که در این حافظه ذخیره می‌شود، بدون احتساب برچسب‌ها و بیت‌های valid ۲۵۶ بایت و میزان داده ذخیره شده در هر بلوک ۳۲ بیت است.) آدرس ۳۲ بیتی بایتهایی از حافظه که پردازنده باید به آنها مراجعه کند، به ترتیب زیر است:

۳۲، ۲۵۶، ۳۸۴، ۵۱۲، ۳۸۴، ۲۵۶، ۱۰۲۴، ۷۶۸، ۵۱۲

الف- با فرض این که حافظه نهان در ابتدا خالی باشد و داده‌ها با نگاشت مستقیم (direct map) در حافظه نهان قرار می‌گیرند، برای هر یک از مراجعات بالا مقادیر دودویی هر آدرس، مقدار برچسب (tag) و مقدار اندیس (index) را مشخص کنید. در کدام موارد miss رخ می‌دهد و در کدام موارد hit؟ در مواردی که miss رخ داده، نوع آن را هم مشخص کنید. محتوای نهایی حافظه نهان را رسم کنید.

پاسخ: حجم حافظه نهان ۲۵۶ بایت و طول هر بلوک ۴ بایت است، بنابراین کلا ۶۴ کلمه داریم.

| Byte# | Word# | index | H/M | Miss Type | tag |
|-------|-------|-------|------|------------|-----|
| 32 | 8 | 8 | Miss | Compulsory | 0 |
| 256 | 64 | 0 | Miss | Compulsory | 1 |
| 384 | 96 | 32 | Miss | Compulsory | 1 |
| 512 | 128 | 0 | Miss | Compulsory | 2 |
| 384 | 96 | 32 | Hit | - | 1 |
| 256 | 64 | 0 | Miss | Contention | 1 |
| 1024 | 256 | 0 | Miss | Compulsory | 4 |
| 768 | 192 | 0 | Miss | Compulsory | 3 |
| 512 | 128 | 0 | Miss | Contention | 2 |

محتوای نهایی حافظه نهان:

Cache[0]=M[512]

Cache[8]=M[32]

Cache[32]=M[384]

ب- با فرض این که ساختار حافظه نهان از نوع انجمنی دو انتخابی (2-way set associative) باشد، به سوالات بند الف دوباره پاسخ دهید.

پاسخ: این بار در هر مجموعه از حافظه نهان دو بلوک ۴ بایتی داریم، بنابراین کلا ۳۲ مجموعه خواهیم داشت.

| Byte# | Word# | Set | H/M | Miss Type | tag | Cache line | |
|-------|-------|-----|------|------------|-----|------------|---------|
| 32 | 8 | 8 | Miss | Compulsory | 0 | M[32] | - |
| 256 | 64 | 0 | Miss | Compulsory | 2 | M[256] | |
| 384 | 96 | 0 | Miss | Compulsory | 3 | M[256] | M[384] |
| 512 | 128 | 0 | Miss | Compulsory | 4 | M[512] | M[384] |
| 384 | 96 | 0 | Hit | - | 3 | M[512] | M[384] |
| 256 | 64 | 0 | Miss | Contention | 2 | M[256] | M[384] |
| 1024 | 256 | 0 | Miss | Compulsory | 4 | M[256] | M[1024] |
| 768 | 192 | 0 | Miss | Compulsory | 6 | M[768] | M[1024] |
| 512 | 128 | 0 | Miss | Contention | 4 | M[768] | M[512] |

ج- با فرض این که ساختار حافظه نهان از نوع انجمنی چهار انتخابی (4-way set associative) باشد، یک بار دیگر به سوالات بند الف پاسخ دهید.

در موارد ب و ج، روش جایگزینی بلوک‌ها را LRU فرض کنید.

پاسخ: این بار در هر مجموعه از حافظه نهان چهار بلوک ۴ بایتی داریم، بنابراین کلا ۱۶ مجموعه خواهیم داشت.

| Byte# | Word# | Set | H/M | Miss Type | tag | Cache line | | | |
|-------|-------|-----|------|------------|-----|------------|--------|--------|---------|
| 32 | 8 | 8 | Miss | Compulsory | 0 | M[32] | - | - | - |
| 256 | 64 | 0 | Miss | Compulsory | 4 | M[256] | - | - | |
| 384 | 96 | 0 | Miss | Compulsory | 6 | M[256] | M[384] | - | |
| 512 | 128 | 0 | Miss | Compulsory | 8 | M[256] | M[384] | M[512] | |
| 384 | 96 | 0 | Hit | - | 6 | M[256] | M[384] | M[512] | |
| 256 | 64 | 0 | Hit | - | 4 | M[256] | M[384] | M[512] | |
| 1024 | 256 | 0 | Miss | Compulsory | 16 | M[256] | M[384] | M[512] | M[1024] |
| 768 | 192 | 0 | Miss | Compulsory | 12 | M[256] | M[384] | M[768] | M[1024] |
| 512 | 128 | 0 | Miss | Contention | 8 | M[256] | M[512] | M[768] | M[1024] |

۳- (۳۰ نمره) آدرس‌های حافظه در یک پردازنده ۳۲ بیتی و حافظه byte-addressable است. یک حافظه نهان با حجم ۵۱۲ کیلوبایت برای داده داریم و اندازه هر بلوک داده هم ۳۲ بایت است.

الف- این حافظه نهان در مجموع چند بلوک داده دارد؟

ب- از حافظه نهان از نوع نگاشت مستقیم (direct map) باشد، هر یک از فیلدهای tag، index و block offset چند بیت دارند؟ حجم کل حافظه نهان (با احتساب برچسب‌ها و بیت‌های valid) چند بیت است؟

ج- اگر حافظه نهان از نوع انجمنی دو انتخابی (2-way set associative) باشد، چند مجموعه داریم؟ هر یک از فیلدهای tag، index و block offset چند بیت دارند؟ حجم کل حافظه نهان (با احتساب برچسب‌ها و بیت‌های valid) چند بیت است؟

د- اگر حافظه نهان از نوع انجمنی کامل (full associative) باشد، هر یک از فیلدهای tag، index و block offset چند بیت دارند؟ حجم کل حافظه نهان (با احتساب برچسب‌ها و بیت‌های valid) چند بیت است؟

a- $No\ of\ blocks = (512 \times 1024)/32 = 2^{14} = 16384$

b- $block_offset = 5\ bits$ $index = 14\ bits$ $tag = 32 - (14 + 5) = 13\ bits$
 $bits\ in\ a\ line = 1\ (valid) + 13\ (tag) + 32 \times 8\ (data) = 270\ bits$
 $total\ bits = 270 \times 16384 = 4423680\ bits$

c- $No\ of\ sets = (512 \times 1024)/(2 \times 32) = 2^{13} = 8192\ sets$
 $block_offset = 5\ bits$ $index = 13\ bits$ $tag = 32 - (13 + 5) = 14\ bits$
 $bits\ in\ a\ line = 2 \times [1\ (valid) + 14\ (tag) + 32 \times 8\ (data)] = 542\ bits$
 $total\ bits = 542 \times 8192 = 4440064\ bits$

d- $block_offset = 5\ bits$ $index = 0\ bits$ $tag = 32 - (5) = 27\ bits$
 $total\ bits = (1 + 27 + 32 \times 8) \times 16384 = 4653056\ bits$

۴- (۱۰ نمره) فرض کنید CPI پایه در یک پردازنده با یک لایه حافظه نهان ۲ چرخه ساعت باشد. نرخ فقدان (miss-rate) برای دستورات و داده‌ها به ترتیب ۱۲٪ و ۲٪ است و هر بار دسترسی به حافظه اصلی ۱۰۰ چرخه ساعت زمان می‌برد. به طور متوسط ۳۰٪ از دستورات یک عملوند در حافظه دارند.

الف- مقدار واقعی CPI چقدر است؟

$$actual_CPI = 2 + 0.12 \times 100 + 0.3 \times 0.02 \times 100 = 2 + 12 + 0.6 = 14.6 \text{ Clock Cycle}$$

ب- برای کاهش CPI یک لایه دیگر حافظه نهان فقط برای دسترسی به دستورات اضافه می‌کنیم که هر بار دسترسی به آن ۱۰ چرخه ساعت طول می‌کشد و نرخ فقدان دستورات (نیاز به دسترسی به حافظه اصلی) را به ۲٪ می‌رساند. مقدار CPI واقعی را در این حالت محاسبه کنید.

$$actual_CPI = 2.6 + (0.12 \times 10 + 0.02 \times 100) = 2.6 + 1.2 + 2 = 5.8 \text{ Clock Cycle}$$

۵- (۳۰ نمره) یک برنامه به زبان C برای محاسبه مجموع درایه‌های یک ماتریس به ابعاد ۶۴ در ۱۲۸ که از اعداد صحیح ۳۲ بیتی پر شده است به دو طریق زیر نوشته شده است:

حلقه یک:

حلقه دو:

```
for(i = 0; i < 128; i++)          for(j = 0; j < 64; j++)
    for(j = 0; j < 64; j++)        for(i = 0; i < 128; i++)
        Sum += A[i][j]              Sum += A[i][j]
```

می‌دانیم ماتریس A به صورت زیر در حافظه اصلی نوشته شده است. به عبارت دیگر A[i][j] در آدرس $(j+1 \times 64) \times 4$ قرار دارد.

| | | | | | | |
|---------|---------|-----|----------|---------------------|-----|------------|
| 0 | 4 | 252 | 256 | 4 * (64 * 127 + 63) | | |
| A[0][0] | A[0][1] | ... | A[0][63] | A[1][0] | ... | A[127][63] |

در سوالات زیر فرض کنید حافظه نهان در ابتدای کار خالی است و نیز فرض کنید تنها دسترسی به حافظه برای خواندن درایه‌های ماتریس A انجام می‌شود و همه متغیرهای موردنیاز دیگر در ثبات‌ها ذخیره شده‌اند. خود دستورات هم در حافظه نهان دیگری (instruction cache) قرار دارند.

الف- با در نظر داشتن یک حافظه نهان با نگاشت-مستقیم (direct-mapped) که ۴ کیلوبایت حجم و بلوک‌های ۸ کلمه‌ای (۳۲ بیتی) دارد، تعداد cache miss برای اجرای هر کدام از دو حلقه محاسبه کنید.

پاسخ: تعداد بلوک‌های حافظه نهان $128 = 4096 / 32$ است، بنابراین در ۵ بیت برای byte offset نیاز است و ۷ بیت برای index. اندازه ماتریس 64×128 کلمه است، یعنی ۳۲ کیلوبایت. هر درایه از ماتریس A در یک آدرس مشخص از حافظه نهان قرار می‌گیرد، اما چون اندازه هر بلوک ۸ کلمه است، با هر بار دسترسی به یکی از درایه‌ها، ۸ کلمه متوالی به حافظه نهان منتقل می‌شود و می‌دانیم که ماتریس به صورت سطری چیده شده، یعنی ستون‌های متوالی یک سطر در کنار هم قرار دارند.

در حلقه دو، دسترسی در حلقه داخلی به ستون‌های متوالی یک سطر انجام می‌شود، یعنی کلماتی که در حافظه هم در کنار هم قرار دارند. بنابراین هر بار که فقدان رخ می‌دهد، همزمان ۸ کلمه متوالی در حافظه نهان نوشته می‌شود و ۷ بار دسترسی بعدی بدون فقدان انجام می‌شود. بنابراین تعداد کل فقدان‌ها، $64 \times 128 / 8 = 1024$ بار خواهد بود.

در حلقه یک، دسترسی در حلقه داخلی به سطرها متوالی یک ستون انجام می‌شود که در حافظه کنار هم نیستند. بنابراین هر بار که حلقه داخلی تمام می‌شود و نوبت به دسترسی به سطر بعدی می‌رسد، باید یک بار دیگر ۸ کلمه متوالی دیگر که مربوط به سطر بعدی است وارد حافظه نهان شود. بنابراین برای هر دسترسی یک

فقدان رخ می‌دهد و تعداد فقدان‌ها ۸۰۹۶ بار خواهد بود. برای بررسی دقیق‌تر می‌توانیم نقشه حافظه نهان را بعد از ۳۲ بار اجرای دستور جمع به شکل زیر رسم کنیم. می‌بینیم که هر چند همه حافظه نهان پر نشده اما باز در دور بعدی فقط روی همین سطرها نوشته می‌شود، بنابراین به ازای هر دسترسی یک فقدان پیش می‌آید.

| | | | | | | | | | |
|-------|----------|---------|--|--|--|--|--|--|---------|
| index | | | | | | | | | |
| 0 | A[0][0] | A[0][1] | | | | | | | A[0][7] |
| ... | - | | | | | | | | |
| 8 | A[1][0] | A[1][1] | | | | | | | A[1][7] |
| ... | - | | | | | | | | |
| 16 | A[2][0] | A[1][1] | | | | | | | A[1][7] |
| ... | | | | | | | | | |
| 32 | A[3][0] | A[3][1] | | | | | | | A[3][7] |
| ... | | | | | | | | | |
| | | | | | | | | | |
| | | | | | | | | | |
| 120 | A[15][0] | | | | | | | | |
| ... | | | | | | | | | |

ب- فرض کنید می‌خواهیم یک حافظه نهان ۸ کلمه‌ای (۳۲ بایتی) بسازیم که در صورت اجرای حلقه یک هیچ فقدانی به جز فقدان‌های اجتناب‌ناپذیر (compulsory miss) رخ ندهد این حافظه نهان باید حداقل چند بلوک داشته باشد؟ همین محاسبه را برای اجرای حلقه دو و با همین شرایط انجام دهید.

چون به هر درایه ماتریس بیشتر از یک بار دسترسی نداریم، حافظه نهان یک بلوک هم داشته باشد کافی است که در اجرای حلقه دو فقط فقدان‌های اجتناب‌ناپذیر رخ دهد.
در حلقه یک، حافظه نهان باید ظرفیتش طوری باشد که بتواند یک ستون کامل را در خود جا دهد که این یعنی باید 8×128 بلوک داشته باشد.

ج- یک حافظه نهان ۴ کیلوبایتی از نوع انجمنی کامل (fully-associative) با بلوک‌های ۸ کلمه‌ای (۳۲ بایتی) در نظر بگیرید که برای جاگذاری بلوک‌ها از سیاست FIFO استفاده می‌کند. تعداد cache miss را با اجرای هر کدام از دو حلقه محاسبه کنید.
تعداد فقدان‌ها در اجرای حلقه دو تغییر نمی‌کند اما تعداد فقدان‌ها در اجرای حلقه یک $1/8$ می‌شود، یعنی ۱۰۲۴ بار، چون فضای خالی که در شکل بالا وجود داشت، دیگر وجود ندارد، چون هر آدرس حافظه را می‌توان در هر خط حافظه نهان نوشت.