



به موارد زیر توجه کنید:

- ۱- حتما نام و شماره دانشجویی خود را روی پاسخ نامه 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 ۶۰ نمره دارد که معادل ۰/۶ 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- (۵ نمره) در یک سیستم حافظه، یک لایه حافظه نهان با زمان دسترسی ht ثانیه و نرخ فقدان mt1 داریم. زمان دسترسی به حافظه اصلی t ثانیه است. برای کاهش زمان دسترسی متوسط (AMAT) یک لایه حافظه نهان دیگر با زمان دسترسی t/2 ثانیه اضافه می کنیم. نرخ فقدان این لایه جدید (mt2) حداکثر چقدر باشد که زمان دسترسی کاهش یابد؟

پاسخ:

$$AMAT_1 = ht + mt1 \times t$$

$$AMAT_2 = ht + mt1 \times \left(\frac{t}{2} + mt2 \times t\right)$$

$$AMAT_2 \leq AMAT_1 \Rightarrow ht + mt1 \times \frac{t}{2} + mt1 \times mt2 \times t \leq ht + mt1 \times t \Rightarrow mt2 \leq \frac{1}{2}$$

۲- (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 داده ها به ترتیب ۱۰٪ و ۲٪ است و هر بار دسترسی به حافظه اصلی ۱۰۰ چرخه ساعت زمان می برد. به طور متوسط ۴۰٪ از دستورات یک عملوند در حافظه دارند.

الف- مقدار واقعی CPI چقدر است؟

پاسخ:

$$CPI = 2 + 0.02 \times 0.40 \times 100 + 0.10 \times 100 = 2 + 0.8 + 10 = 12.8$$

ب- برای کاهش CPI یک لایه دیگر حافظه نهان فقط برای دسترسی به دستورات اضافه می کنیم که هر بار دسترسی به آن ۱۰ چرخه ساعت طول می کشد و نرخ سراسری فقدان دستورات (نیاز به دسترسی به حافظه اصلی) را به ۴٪ می رساند. مقدار واقعی CPI را در این حالت محاسبه کنید.

پاسخ:

$$CPI = 2 + 0.02 \times 0.40 \times 100 + 0.1 \times 10 + 0.04 \times 100 = 2 + 0.8 + 1 + 4 = 7.8$$

۳- (۱۵ نمره) یک حافظه نهان مجموعه انجمنی (set-associative) با حجم یک مگابایت داریم که هر بلوک آن ۱۶ بایت است. اگر آدرس پردازنده ۲۴ بیتی باشد، تعداد راه های هر مجموعه چقدر باشد تا طول برچسب ها ۱۰ بیتی شود؟ در این صورت، تعداد کل بیت های مورد نیاز در این حافظه نهان چقدر خواهد بود؟ اگر همین حافظه نهان کاملاً انجمنی باشد، تعداد کل بیت های مورد نیاز چند خواهد بود؟ اگر این حافظه نهان دارای نگاشت مستقیم باشد، چطور؟

پاسخ:

ظرفیت این حافظه نهان  $2^{17} / 8 = 2^{20}$  بایت است، بنابراین تعداد کل بلوک های این حافظه  $2^{17} / 16 = 2^{13} = 8192$  بلوک است. با توجه به این که هر بلوک ۱۶ بایت است، ۴ بیت برای بخش افست مورد نیاز است. اگر برچسب ها ۱۰ بیتی باشد، ۱۰ بیت برای اندیس دهی داریم که یعنی  $1024$  مجموعه خواهیم داشت. بنابراین در هر مجموعه  $8192 / 1024 = 8$  بلوک داریم که به معنای این است که تعداد راه های هر مجموعه ۸ است.

در این صورت تعداد کل بیت‌های موردنیاز این طور محاسبه می‌شود:

$$8192 \times (16 \times 8 + 10 + 1) = 8192 \times 139 = 1,138,688 \text{ bit}$$

اگر این حافظه کاملاً انجمنی باشد، هیچ بیتی برای اندیس نیاز نداریم، بنابراین طول برچسب‌ها ۲۰ بیت است و تعداد کل بیت‌های موردنیاز افزایش می‌یابد:

$$8192 \times (16 \times 8 + 20 + 1) = 8192 \times 149 = 1,220,608 \text{ bit}$$

اگر این حافظه دارای نگاشت مستقیم باشد، ۱۳ بیت برای اندیس داریم، بنابراین طول برچسب‌ها ۷ بیت است و تعداد کل بیت‌های موردنیاز افزایش می‌یابد:

$$8192 \times (16 \times 8 + 7 + 1) = 8192 \times 136 = 1,114,112 \text{ bit}$$

۴- (۲۰ نمره) می‌خواهیم اهمیت اصل تمرکز موضعی (Principle of Locality) را درک کنیم. فرض کنید در یک سیستم حافظه دسترسی به حافظه نهان یک چرخه، دسترسی به حافظه اصلی در صورت وجود حافظه نهان ۱۱۰ چرخه و دسترسی به حافظه اصلی در صورت غیرفعال بودن حافظه نهان ۱۰۵ چرخه طول بکشد. الف- اگر شما در حال اجرای یک برنامه با نرخ فقدان ۳٪ باشید، میانگین زمان دسترسی به حافظه چند چرخه خواهد بود؟ پاسخ:

$$AMAT = 0.97 \times 1 + 0.03 \times 110 = 4.27 \text{ cycle}$$

ب- حال برنامه‌ای را اجرا می‌کنیم که آدرس‌های تصادفی بدون تمرکز موضعی تولید می‌کند. به این شکل که یک آرایه یک گیگابایتی داریم که به طور تصادفی به عناصر آن دسترسی پیدا می‌کنیم. آدرس‌ها توسط یک مولد عدد تصادفی با توزیع یکنواخت تولید می‌شوند. اگر اندازه حافظه نهان ۶۴ کیلوبایت باشد، در بهترین حالت، میانگین زمان دسترسی به حافظه چقدر خواهد بود؟ پاسخ:

در بهترین حالت حافظه نهان کاملاً پر است، بنابراین احتمال وجود داده موردنظر در آن (نرخ برخورد) برابر است با:

$$hit_{time} = \frac{64 \times 1024}{1024 \times 1024 \times 1024} = 2^{16-230} = 2^{-14}$$

بنابراین، میانگین زمان دسترسی به حافظه برابر است با:

$$AMAT = 2^{-14} \times 1 + (1 - 2^{-14}) \times 110 \approx 110 \text{ cycle}$$

ج- اگر نتیجه بند ب را با زمان دسترسی به حافظه اصلی در حالتی که حافظه نهان غیرفعال است، مقایسه کنید، چه نتیجه‌ای در مورد نقش اصل تمرکز موضعی در توجیه استفاده از حافظه نهان می‌گیرید؟ پاسخ:

می‌بینیم در بند ب که تمرکز موضعی نداریم، نرخ برخورد بسیار کم شده و عملاً میانگین زمان دسترسی نسبت به اندازه miss penalty است که حتی از زمانی که اصلاً حافظه نهان نداریم هم بدتر است.

د- می‌بینیم که وجود داده در حافظه نهان (hit) به اندازه ۱۰۴ چرخه (یک در مقابل ۱۰۵) سود (Gain) دارد و در صورت فقدان داده در حافظه نهان (miss) به اندازه ۵ چرخه زیان (Loss) می‌کنیم. در حالت کلی، حداکثر نرخ فقدان را برحسب مقادیر سود (G) و زیان (L) طوری حساب کنید که استفاده از حافظه نهان به صرفه باشد. پاسخ:

نرخ فقدان باید طوری باشد که میانگین زمان دسترسی به حافظه با حافظه نهان کمتر از حالتی باشد که حافظه نهان غیرفعال است.

$$G = AMAT_{noCache} - ht \Rightarrow AMAT_{noCache} = G + ht$$

$$L = missPenalty - AMAT_{noCache} = missPenalty - (G + ht) \Rightarrow missPenalty = L + G + ht$$

$$AMAT_{withCache} = ht \times (1 - missrate) + (L + G + ht) \times missRate = ht + (G + L) \times missRate$$

$$AMAT_{withCache} \leq AMAT_{noCache} \Rightarrow ht + (G + L) \times missRate \leq G + ht \Rightarrow missRate \leq \frac{G}{G + L}$$

بنابراین، در این مسئله نرخ فقدان باید از  $\frac{104}{110} = 0.95$  کمتر باشد تا استفاده از حافظه نهان به صرفه شود.

در این سوال، اگر ۱۱۰ چرخه را شامل یک چرخه دسترسی به حافظه نهان در نظر نمی‌گیریم، به پاسخ‌های متفاوتی می‌رسیم که آنها هم درست محسوب می‌شود:

$$AMAT = 4.3 \text{ الف}$$

$$AMAT \approx 111 \text{ ب-}$$

$$missRate \leq \frac{G}{G+L+ht} = \frac{104}{110} = 0.945 \text{ د-}$$

۵- (۱۰ 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: (اولین دسترسی به آدرس ۳۲ است)

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

با فرض این که حافظه نهان در ابتدا خالی باشد و این که ساختار حافظه نهان از نوع انجمنی دو انتخابی (2-way set associative) باشد، 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 در کل ۶۴ بلوک داریم. چون ساختار حافظه انجمنی دو انتخابی است، در هر مجموعه دو بلوک داریم، بنابراین در کل ۳۲ مجموعه خواهیم داشت.

به این ترتیب، دو بیت کم‌ارزش برای افست به کار می‌روند و ۵ بیت بعدی اندیس و بقیه بیت‌ها برچسب هستند.

Byte#	Word#	Set	H/M	Miss Type	tag	Cache line	
32	8	0001000	8	Miss	Compulsory	0	M[32]
256	64	1000000	0	Miss	Compulsory	2	M[256]
384	96	1100000	0	Miss	Compulsory	3	M[256]
512	128	10000000	0	Miss	Compulsory	4	M[512]
384	96	1100000	0	Hit	-	3	M[512]
256	64	1000000	0	Miss	Contention	2	M[256]
1024	256	100000000	0	Miss	Compulsory	8	M[256]
768	192	110000000	0	Miss	Compulsory	6	M[768]
512	128	100000000	0	Miss	Contention	4	M[768]