

Cevodona CPE:

Število ukazov, ki jih Cpe izvede v eni sekundi, merjeno v milijon operacijah na sekundo (MIPS)

$$\uparrow MIPS = \frac{\uparrow f_{cpe}}{\downarrow CPI \cdot 10^6}$$

Moorov Zakon: število tranzistorjev na čipu se bo podvojilo na vsaki dve leti.

Glavni pomnilnik:

Ločimo dva tipa RAM pomnilnika: SRAM (statični RAM) in DRAM (dinamični RAM). Statični ne potrebuje osvežitev, dinamični pa jih, če ne izgubimo podatke iz njega. Statični je tudi hitrejši zaradi tega od dinamičnega, je pa občutno dražji.

Glavne lastnosti pomnilnikov:

- Cena
- Hitrost dostopa
- Način dostopa
- Spremenljivost vsebine
- Obstojnost vsebine
- Zanesljivost

Način dostopa:

Navadni direktni pomnilniki: vsaka pomnilniška beseda ima svoj fiksni naslov, dostop do besede je z naslovom.

Asociativni pomnilnik: pomnilniške besede nimajo naslova, dostop do želene besede je preko vsebine ali dela vsebine te besede (primerjava vsebine besed in izbira glede na vsebino).

Hitrost dostopa (access rate)  $b_a$  je največje možno število prenesenih pomnilniških besed na sekundo:

$$b_a = \frac{1}{t_c} = \frac{1}{t_a + t_m}$$

Čas  $t_c$  je čas cikla (čas med dvema zaporednima dostopoma). Čas  $t_a$  je čas dostopa do pomnilnika, čas  $t_m$  pa je mrtvi čas, potreben za osvežitev pomnilnikovih kondenzatorjev in vsebine pomnilnika.

Spremenljivost vsebine: poznamo RAM in ROM pomnilnike: ločijo se predvsem po tem, da se da v RAM pomnilnik poljubno pisati, pri ROM pomnilnikih pa samo s posebnimi programatorji oz. vezji.

Obstojnost vsebine (pomembno predvsem pri DRAM pomnilnikih):

- Destruktivno branje (DRAM)
- Dinamično shranjevanje (DRAM)
- Odklop vira energije (DRAM in SRAM)

Pri DRAM pomnilnikih se pri branju kondenzatorji v celi vrstici spraznijo (imenujemo destruktivno branje). Zato preberemo celo vrstico in jo vpišemo nazaj (potreben dodaten čas imenovan mrtvi čas).

Kondenzatorji v DRAM pomnilniku se hitro praznijo in jih je treba periodično osveževati (potreben določen čas, med osveževanjem pomnilnik ni na voljo za branje in pisanje). To imenujemo dinamično shranjevanje, saj se pri tem vsebina vrstice prebere in zapiše nazaj.

Pri odklopu vira energije, se vsebina DRAM in SRAM pomnilnikov izgubi.

Zanesljivost:

Zanesljivost pomnilnika merimo z verjetnostjo za pojav napake: BER (Bit Error Rate) ali MTBF (Mean Time Between Failures).

Zanesljivost lahko povečamo z ECC (Error Correcting Codes – v glavnem v rabi v pomožnih pomnilnikih (diskih) in strežnikih).

#### Organizacija glavnega pomnilnika:

- Dolžina pomnilniške besede v bitih ( $n$ -bitov =  $2^n$  različnih možnih vsebin)
- Pomnilniški naslov v bitih ( $m$ -bitov = naslovimo lahko  $2^m$  pomnilniških besed)

Metabiti: (skupno ime ECC) so biti za korekcijo in detekcijo napak (paritetni biti). Metabiti so dodatni biti v vsaki pomnilniški besedi, lahko pa povedo kaj vsebuje beseda: ukaz, število (v fiksni ali plavajoči vejici), znak,...)

Zaščita glavnega pomnilnika: je mehanizem, ki omogoča zaščito enega dela pomnilnika (programa) pred neželenimi posegi drugega programa. Zato glavni pomnilnik razdelimo na bloke ali strani po 2048, 4096 ali 8192 besed, ki so zaščitene zase kot celota. Vsaka stran ima svoj zaščitni ključ za dostop (protection key). Ti ključi so shranjeni v posebnem pomnilniku in ni del glavnega pomnilnika, prav tako ni možen dostop do njih z običajnimi ukazi.

#### Pomnilniško prepletanje:

Obstajata dva načina pohitritve pomnilnika:

- Razširimo podatkovno pot do glavnega pomnilnika
- S pomnilniškim prepletanjem (memory interleaving)

Pri pomnilniškem prepletanju gre predvsem za to, da pomnilnik razdelimo na module (banke), kjer je modul samoostojen del pomnilnika, ki deluje neodvisno od drugih, kar pomeni, da če imamo  $m$  modulov imamo  $m$ -kratno prepletanje in lahko opravljamo  $m$ -dostopov hkrati. Pomnilniško prepletanje tako omogoča da se pri zaporednih dostopih, nasednji dostop prične preden se je predhodni dostop končal, možni pa so istočasni dostopi (če podatki niso v istem modulu), če ima računalnik več procesorjev ali RAM krmilnikov (tako imenovana Dual Channel tehnologija).

#### Predpomnilnik (cache):

Razlika med hitrostjo CPE in hitrostjo DRAM pomnilnika je zelo velika, lahko pa jo zmanjšamo z uporabo predpomnilnika in z uporabo naslednjih možnosti:

- Uporaba dveh predpomnilnikov (eden za ukaze in eden za operande – Hardvarvska arhitektura)
- Uporaba širših podatkovnih poti med predpomnilnikom in CPE in predpomnilnikom in DRAM-om
- Razdelitev glavnega pomnilnika na module (pomnilniško prepletanje)
- Drugi predpomnilniški nivo L2 izven CPE, ki je homogen (princetonska arhitektura), včasih tudi tretji nivo L3.
- S CPE povežemo predpomnilnik L2 s posebno prenosno potjo (TLB), v tem primeru lahko dostopamo do L2 predpomnilnika in glavnega pomnilnika hkrati in zmanjšamo potreben čas ob zgrešitvi
- Pri sosednjih besedah v bloku se uporabi page mode način prenosa

Vsebina predpomnilnika je podmnožica vsebine glavnega pomnilnika (predpomnilnik je vedno manjši kot glavni pomnilnik). Razlog je cena in velikost SRAM pomnilnika. Posledica pa je, da nimamo vedno v predpomnilniku željene vsebine. Razlog da bo iskana informacija pogosto v predpomnilniku je lokalnost pomnilniških dostopov. Predpomnilnik mora biti narejen tako, da vsebujedelovno množico informacij, ki se vedno prilagaja. Ker je v predpomnilniku samo del vsebine glavnega pomnilnika, mora biti v predpomnilniku še informacija kateri deli glavnega pomnilnika so v njem (naslovi). Pri dostopu CPE do informacije v predpomnilniku imamo dve možnosti:

- Zadetek (hit), naslov je v pred.
- Zgrešitev (miss), naslova ni v pred.

Uspešnost delovanja predpomnilnika merimo z verjetnostjo zadetka  $H$ :

$$H = \frac{N_p}{N} = \frac{N_p}{N_g + N_p}$$

$N$  =število vseh dostopov ( $N = N_g + N_p$ )

$N_p$  = število zadetkov

$N_g$  = število zgrešitev

Ali pa merimo z verjetnostjo zgrešitve  $1-H$

Verjetnost zadetka  $H$  je ponavadi večja od 0,9, večinoma celo večja od 0,95.

V primeru zgrešitve je potreben dostop do glavnega pomnilnika:

Čas dostopa do predpomnilnika označimo s  $t_{ap}$ , čas dostopa do glavnega pomnilnika pa s  $t_{ag}$ .

Povprečen čas dostopa  $t_a$  do predpomnilnika in glavnega pomnilnika skupaj je:

$$t_a = t_{ap} + (1 - H) \cdot t_{ag}$$

Čas v računalniku merimo z urinimi periodami. Čas  $t_{ap}$  je število urinih period za dostop do predpomnilnika na nivoju L1, in je pri večini računalnikov 1 urina perioda. Pri zgrešitvi je potreben dostop do glavnega pomnilnika, čas ki je zato potreben pa je zgrešitvena kazen.

Zgrešitvena kazen  $t_B$  je število urinih period, ki se ob zgrešitvi prištejejo k času dostopa. Je odvisna od velikosti bloka, širine podatkovne poti med pred. In DRAM-om. Če ima računalnik tudi pred. L2, je zgrešitvena kazen manjša. Povprečni čas dostopa je potem:

$$t_a = t_{ap} + (1 - H) \cdot t_B$$

Vsi časi so v urinih periodah, povprečni čas v sekundah lahko izračunamo, tako, da rezultat  $t_a$  v urinih periodah pomnožimo s časom trajanja ene urine periode  $t_{CPE}$ :

$$t_{avsek} = t_a \cdot t_{CPE}$$