

HIJERARHIJSKA ORGANIZACIJA MEMORIJSKOG SUSTAVA RAČUNALA

NASTUPNO PREDAVANJE
Čakovec, 28. lipanj 2004.

Tedo Vrbanec, dipl. inf.
Visoka učiteljska škola u Čakovcu
Dr. Ante Starčevića 55, 40000 Čakovec
e-mail: tedo.vrbanec@vus-ck.hr

1. SAŽETAK

Memorijski podsustav današnjih snažnih i brzih računala neminovno ima svoju hijerarhijsku strukturu koja se temelji na priručnim memorijama. Priručne memorije predstavljaju bilo koje dijelove memorijskog sustava koji su brži od drugih, znatno većih i sporijih, a omogućavaju ubrzanje procesa dohvata podataka iz sporijih dijelova memorijskog sustava. Ako znamo da je cijena brzih memorija znatno veća od sporijih, ova koncepcija omogućuje zadržavanje niske cijene memorijskog podsustava, pa tako i cijelog računalnog sustava.

Priručna memorija ima svoju maksimalnu veličinu koja je znatno manja od većeg spremišta kojeg ubrzava. Moguće je koristiti više razina priručnih memorija.

U računalu postoje mnoga mjesta gdje je moguće znatno unaprijediti performanse, ukoliko dodamo priručne memorije koje predstavljaju "memorijske ubrzivače". Tako npr. procesor se može efikasnije iskoristiti ukoliko je zahvaćanje podataka ubrzano priručnom memorijom prve razine, koja ubrzava priručnu memoriju druge razine, a ona ubrzava radnu memoriju. Radna memorija dalje ubrzava tvrdi disk, a on "ubrzava" npr. Internet vezu.

2. UVOD

Memorijski sustav, njegova struktura i logika djelovanja vrlo su važan čimbenik rada svih računala. Iz mog iskustva u radu sa studentima, s obzirom da velika većina njih nema tehničku naobrazbu, većina studenata ne može reproducirati koncepciju rada memorijskog sustava i tek ih zanemari broj uspijeva suvislo objasniti problematiku. Ostali uspijevaju bez razumijevanja tek parcijalno nabrojiti elemente memorijskog sustava, a često se dešava i nehvaćanje pitanja navodeći vrste memorija.

Ovo predavanje **neće govoriti** o vrstama memorija, o tehnologijama njihove izrade ili usporedbi vrsta memorija, već o koncepciji hijerarhijske organizacije memorijskog sustava računala.

Iako se pojam memorije tehnički odnosi na bilo koju vrstu elektroničkog spremišta podataka, on se najčešće upotrebljava odnoseći se na brze ali i privremene oblike spremišta. Ako računalo stalno treba pristupati (sporom) tvrdom disku kako bi preuzelo svaki potrební odsječak podataka, ono će raditi jako sporo. Ako se nasuprot tome, podaci čuvaju u memoriji, procesor im može pristupati znatno brže. Namjena je većine vrsta spremišta privremena pohrana podataka i njihov dohvat na najbrži mogući način. Hijerarhijska organizacija memorijskog sustava način je za ostvarenje te namjene.

3. HIJERARHIJSKA ORGANIZACIJA MEMORIJSKOG SUSTAVA RAČUNALA

Memorijski sustav sastavljen je od više komponenti: registara, višerazinske priručne memorije, radne memorije i vanjskih spremišta. Zašto je to tako?

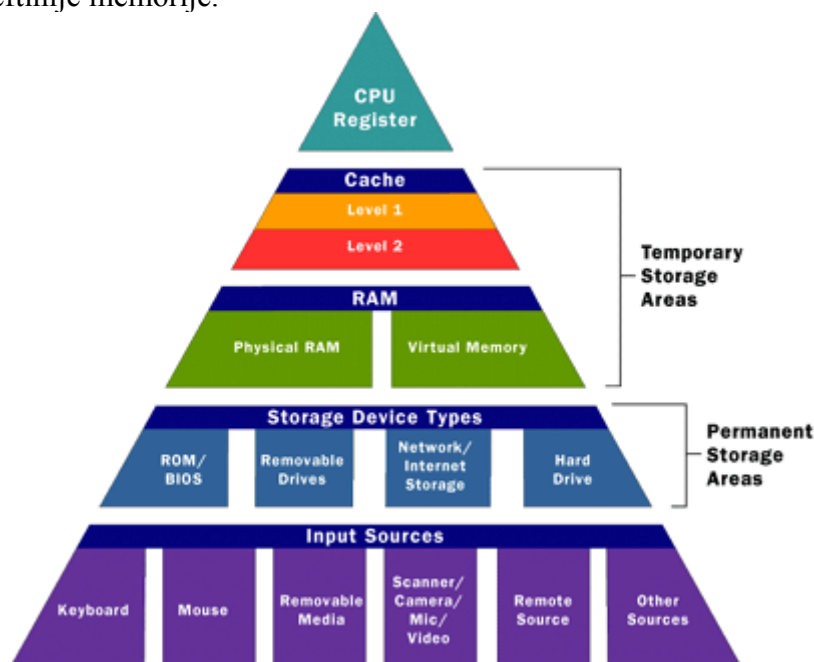
Da bi današnji brzi i snažni procesori maksimizirali svoje performanse potreban im je brz i lak pristup ogromnoj količini podataka. Ako procesor ne može dobiti potrebne podatke on doslovce stane i čeka na njih.

Mjera brzine rada procesora su Hz, tj. broj ciklusa u sekundi. Procesori koji rade brzinama koje se mjere u GHz mogu za svoj rad trošiti milijarde bajtova u sekundi. Tako 32-bitni procesor od 2 GHz (današnji minimum za novo računalo) može potencijalno obraditi dvije milijarde puta u sekundi po četiri bajta ili osam milijardi bajtova u sekundi ($\approx 7,45$ GB/sek.). Cilj je memorijskog sustava zadovoljenje ovakve velike potrebe za podacima ili instrukcijama iz memorijskog podsustava.

Memorije koje mogu pratiti tu brzinu rada procesora iznimno su skupe; toliko skupe da si ih praktički nitko ne može priuštiti u potrebnim količinama.

Problem vremenom ne nestaje već se potencira. *Moore-ov*¹ zakon govori da se brzina rada procesora udvostručava svakih godinu i pol, a brzina rada memorija svakih deset godina. Tako ispada da je prosječno godišnje povećanje raskoraka u brzinama rada procesora i memorije oko 50% u korist procesora.

Dizajneri računala riješili su problem visoke cijene brze memorije tehničkim inovacijama, povezujući memorije korištenjem manje količine skupe i brze koju podupire veća količina sporije i jeftinije memorije.



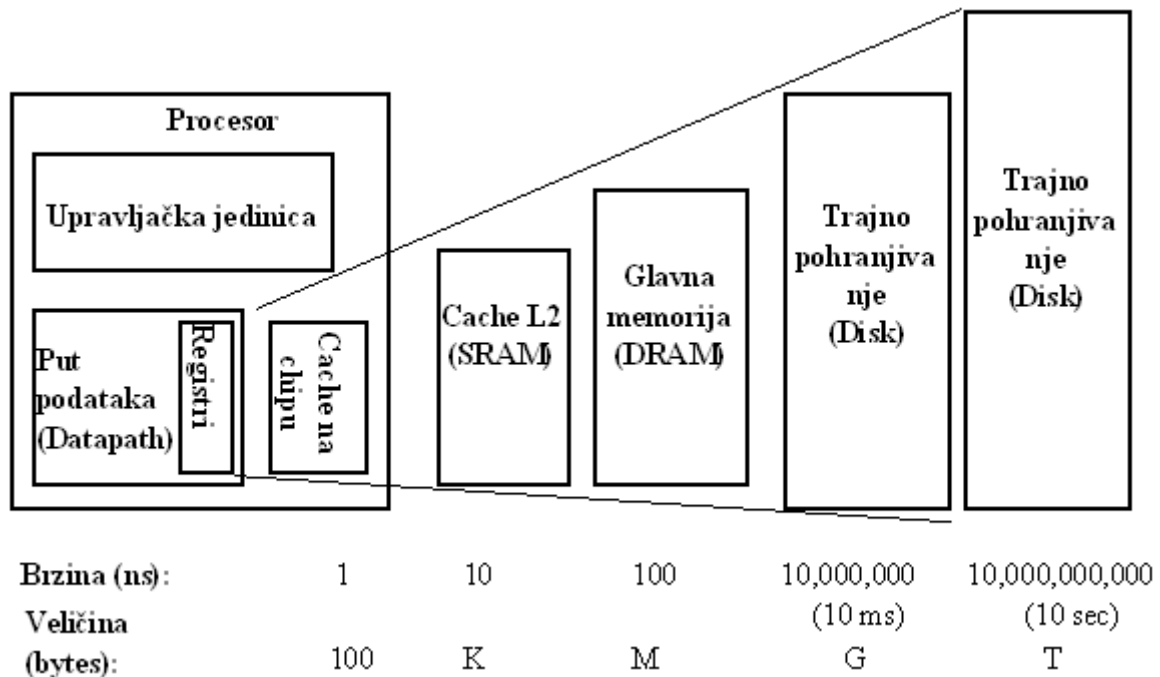
Sl. 1. Hijerarhijska organizacija memorijskog sustava današnjih računala (1)

Izvor: Jeff, Tyson, *How Computer Memory Works*,

<http://computer.howstuffworks.com/computer-memory7.htm>, (27.06.2004.)

¹ Tzv. *Mooreov zakon* postavio je 1965. godine Gordon Moore, suosnivač Intela, prognozirajući da će se broj tranzistora u čipu udvostručavati svakih 18 mjeseci. Ujedno je izjavio da će takav razvoj stati kod granice od 0.25 mikrona (milijuntinka) milimetra, jer tehnologija neće moći više baratati manjim veličinama. Ipak, današnji (lipanj 2004.) procesori izrađeni su u 0.09 mikronske tehnologiji (90 nm). Štoviše, Intelovi znanstvenici prognoziraju da će navedeno tehnološko ograničenje biti na samo 0.045 mikrona ili 45 nm.

Kao što se može vidjeti na slikama 1 i 2, procesor pristupa memoriji prema točno određenom hijerarhijskom slijedu. Većina podataka prvo ulazi u radnu memoriju, bez obzira na to iz kojeg izvora – vanjskih spremišta poput tvrdog diska ili ulaznih jedinica poput tipkovnice. Procesor zatim uzima dijelove podataka iz radne memorije u priručnu memoriju.



Sl. 2. Hijerarhijska organizacija memorijskog sustava današnjih računala (2)

Izvor: FESB, Sveučilište u Splitu, *Uvod u računalstvo- Računala i programiranje*, <<http://laris.fesb.hr/index.html>>, (27.06.2004.)

3.1. MOTIVACIJA KROZ PRIMJER

Na primjeru knjižnice u kojoj se korisnici za knjigu neposredno i isključivo obraćaju knjižničaru, lakše je shvatiti potrebu za priručnom memorijom.

Slučaj 1. – knjižničar bez priručnog spremišta

Knjižničar se nalazi iza radnog stola. Pristupa mu prvi korisnik i traži knjigu *Priče iz davnine*. Knjižničar ode u spremište i s police skine i donese zatraženu knjigu. Kasnije vraća knjigu, a knjižničar odlazi u spremište i vraća knjigu na njeno mjesto. Recimo da sljedeći korisnik zatraži istu knjigu. Knjižničar mora ponovo po nju pri čemu se troši mnogo vremena.

Slučaj 2. – knjižničar s priručnim spremištem

Sve je isto, ali knjižničar ima priručno spremište na koje može staviti 10 knjiga. Početna je situacija ista, a priručno spremište je prazno. Prvi korisnik traži knjigu, knjižničar odlazi po nju, izdaje ju, a korisnik je kasnije vraća. Knjižničar sada ostavlja knjigu u priručno spremište, ukoliko ono već nije puno. Dolazi sljedeći korisnik, traži istu knjigu, knjižničar pregleda knjige u priručnom spremištu, nađe ju i izdaje bez odlaska u pravo spremište. Korisnik je uslužen mnogo brže.

Ukoliko korisnik traži knjigu koja nije u priručnom spremištu, knjižničar je čak manje efikasan jer najprije mora pregledati priručno spremište. Ipak, to je dodatno kašnjenje zanamarivo jer je priručno spremište na dohvat ruke, a broj knjiga u njemu mali.

Slučaj 3. – knjižničar s višestrukim priručnim spremištima

U ovom slučaju, pored postojećeg priručnog spremišta u kome se nalazi 10 uglavnom najtraženijih knjiga, uvodimo i veće ali još uvijek lako dohvatljivo spremište kapaciteta 100

knjiga. Ukoliko knjige koju traži korisnik nema u prvom, manjem priručnom spremištu, velika je vjerojatnost da će je knjižničar naći u drugom, što opet u odnosu na odlazak u glavno spremište umanjuje potrebno vrijeme izdavanja knjige.

3.2. REGISTRI

U svome radu procesori se služe malim, vrlo brzim memorijama koje su dio strukture procesora, (posebno aritmetičko-logičke jedinice) i pod neposrednom su kontrolom procesora. Registri služe za privremeno pohranjivanje podataka. Napravljeni od bistabilnih poluvodičkih elemenata i njihova je brzina rada jednaka brzini rada procesora. Mogu biti namijenjeni za izvršavanje određenih zadataka (namjenski registri) ili raznih operacija (opći registri).

Današnje procesore nazivamo i registarskim strojevima, jer posjeduju (prilično) velik broj registra preko kojih rade najveći broj operacija. Naime, pristup registrima je mnogo brži od pristupa memoriji, a same instrukcije su jednostavnije, kraće i zauzimaju manje memorije. Procesor s većim brojem registara dopušta određivanje jednog ili više operandata tijekom izvođenja same instrukcije. Operand je u tom slučaju pohranjen u glavnoj memoriji, u jednom od registara, ili pak je sadržan u samoj instrukciji.

3.3. PRIRUČNA MEMORIJA PROCESORA

Podaci se iz radne memorije do procesora prenose podatkovnom sabirnicom. Brzina RAM-a ovisi o širini i brzini podatkovne sabirnice te od vremena pristupa. Širina određuje broj bitova koji se mogu istovremeno prenijeti, a brzina određuje koliko puta se taj prijenos u sekundi može izvršiti. Tako je primjerice preko 400 MHz 32-bitne sabirnice teorijski moguće poslati 4 bajta 400 milijuna puta u sekundi. U stvarnosti, brzina je znatno manja zbog latencije memorije. Naime, za početno čitanje ili pisanje iz ili u memoriju potrebno je vrijeme koje se potroši za 2-5 normalna ciklusa.

U današnjim računalnim sustavima najčešće postoje dvije razine priručne memorije: primarna i sekundarna. Ipak, za poslužiteljska računala ponekad se uvodi treća razina s od 2 do 256 MB dodatne priručne memorije smještene na posebnom čipu.

Priručna se memorija u računalima koristi kako bi se ubrzao pristup često korištenim podacima. Podaci mogu stići iz različitih izvora: iz memorije, tvrdog diska ili Interneta. Svaki izvor ima drugu brzinu. Na primjer, mnogo je brže pristupiti podacima na tvrdom disku nego ih skinuti s Interneta. Slično, mnogo je brže pristupiti podacima u radnoj memoriji nego podacima na tvrdom disku.

Priručna je memorija u suštini, komad vrlo brze memorije u kome su smješteni nedavno korišteni podaci. Kada procesor treba neki podatak, on prvo provjerava nalazi li se on u priručnoj memoriji. Ukoliko jeste, nema potrebe zahvaćati u sporije dijelove memorijskog sustava.

Svako spremište podataka može se koristiti za ubrzanje ukoliko je brže od spremišta koje ubrzava. Radna memorija služi za ubrzanje pristupa podacima kojima se često pristupa na tvrdom disku. Procesor ima vrlo brzu priručnu memoriju kojom ubrzava pristup često korištenim podacima iz radne memorije.

3.3.1. PRIRUČNA MEMORIJA PRVE RAZINE (PRIMARNA PRIRUČNA MEMORIJA)

Koliko god podatkovna sabirnica bila široka i brza, ipak je potrebno mnogo više vremena da podaci dođu iz radne memorije do procesora, nego što je potrebno procesoru da ih obradi. Priručna memorija smanjuje taj jaz tako što podatke što se često koriste čini dostupnim

praktično odmah. Dakle, u sam se procesor ugrađuje priručna memorija prve razine (*Level 1 cache*). Ona je obično vrlo mala, od 2 do 64 KB.

Potrebno je napomenuti da se primarna priručna memorija dijeli na dva dijela: na podatkovnu i na instrukcijsku.

3.3.2. PRIRUČNA MEMORIJA DRUGE RAZINE (SEKUNDARNA PRIRUČNA MEMORIJA)

Priručna memorija druge razine (*Level 2 cache*) nekada se smještala na memorijskoj kartici smještenoj u neposrednoj blizini procesora. Ima neposrednu vezu s procesorom. Namjenski upravljački čip na matičnoj ploči zadužen je za upravljanje vezom priručne memorije druge razine s procesorom. Ovisno o procesoru, ova razina priručne memorije ima kapacitet između 256 KB i 2 MB.

Prema [1], u većini računalnih sustava, podaci potrebni procesoru dostupni su u priručnoj memoriji u 95% slučajeva, značajno reducirajući vrijeme čekanja procesora za dohvat potrebnih podataka iz glavne (radne) memorije.

Neki jeftini računalni sustavi potpuno izostavljaju sekundarnu priručnu memoriju. Nasuprot tome, većina današnjih snažnih procesora ima sekundarnu priručnu memoriju kao i primarnu integriranu na sam procesorski čip. Veličina sekundarne priručne memorije i njezin smještaj na samom procesorskom čipu među glavnim su faktorima koji utječu na performanse procesora.

3.4. RADNA (GLAVNA) MEMORIJA

Podaci se u računalu čuvaju unutar brzo dostupnog memorijskog prostora i to u obliku binarnih brojeva, dakle kombinacija nula i jedinica. Svaka binarna znamenka predstavlja jedan bit podataka. Osam bitova čini jedan bajt (B). 1024^1 B čini 1 KB. 1024 KB čini 1 MB itd.

Iako ima mnogo različitih tipova memorija, obično ih dijelimo na dinamičke (DRAM) i statičke (SRAM). Dinamičke danas uvelike pretežu nad statičkim, jer su jeftinije iako sporije. Kod njih bit podataka čuva jedan tranzistor i pridružen mu kondenzator koji zajedno čine memorijsku ćeliju. Problem kod ove vrste memorije je što se podatak spremljen u takvoj memorijskoj ćeliji gubi ukoliko se ne osvježi svakih nekoliko milisekundi. To je znatno usporava. Za razliku od dinamičkih, statičke memorije zadržavaju podatak koji čuvaju sve dok se ne promijeni sadržaj ili dok ne nestane napajanje. Dakle, ne postoji potreba osvježavanja podataka, ali zato je za svaku memorijsku ćeliju potrebno 4 do 6 tranzistora. Nepostojanje osvježavanja čini ju znatno bržom, ali i znatno skupljom. Kao posljedica, SRAM se koristi kao brza priručna memorija, a DRAM kao sporija glavna memorija.

3.5. VANJSKA SPREMIŠTA PODATAKA

Zadatak je vanjskih spremišta podataka (vanjskih memorija) da trajno sačuvaju podatke koji se u radu računala nalaze u memoriji.

Danas najjeftiniji i naveliko rasprostranjeni oblik memorije sposobne za čitanje i pisanje su tvrdi diskovi koji predstavljaju veliko, jeftino² i trajno spremište podataka. Stoga tvrdi disk predstavlja posljednju instancu hijerarhijske organizacije memorije koju nazivamo virtualna memorija.

¹ $1024=2^{10}$

² Cijena 1MB spremišta manja je od 1 lipe.

Pored tvrdog diska, spomenimo i druge danas najraširenije vrste vanjskih spremišta: razne vrste optičkih diskova, flash memorije, magnetske vrpce, a postoje i mnoge druge manje zastupljene poput zip diskete, LS120, razne klasične diskete i slično.

4. RASPRAVA I ZAKLJUČAK

Logično je zapitati se: "Zašto ne bismo cijelu memoriju imali s brzinom istovjetnom brzini procesora, tj. primarne priručne memorije?" To bi bilo moguće, ali i nevjerovatno skupo! Hijerarhijska organizacija memorijskog sustava kompromis je između cijene i brzine rada memorijskog podsustava računalnog sustava. Takva koncepcija omogućuje da za prihvatljiv novčani iznos dobijemo performanse koje prihvatljivo malo zaostaju za maksimalnim mogućim.

Kako je to sve uopće moguće? Zbog *lokalnosti referenciranja*, a to znači da se u nekom programu samo manje količine koda uopće koriste u nekom vremenu. Čak i kada se izvršavaju veliki programi ili zahvaćaju velike količine podataka, samo se manji dio u nekom vremenskom odsječku koristi, a učestalost ponavljanja nekog istog dijela koda ili istog dijela podataka je velika.

Koristeći priručne memorije, moramo provjeravati da li se traženo već nalazi u prvoj priručnoj memorijskoj razini. Ako se nalazi, riječ je o pogotku, a ako se ne nalazi, govorimo o promašaju u priručnoj memoriji. U drugom slučaju, računalni sustav mora po podatak ili instrukciju u sljedeću, sporiju razinu memorijskog sustava. Ipak, ovakav parcijalno sporiji zahvat u prosjeku se višestruko isplati, jer će se već u neposredno sljedećim potrebama za podacima ili instrukcijama iz memorije, po principu lokalnosti referenciranja vrlo vjerojatno naći u tom zahvatu dohvaćen podatak ili instrukcija.

5. LITERATURA

1. -, *How Stuff Works*, <<http://computer.howstuffworks.com>>, (27.06.2004.)
2. FESB, Sveučilište u Splitu, *Uvod u računalstvo- Računala i programiranje*, <<http://laris.fesb.hr/index.html>>, (27.06.2004.)
3. Wikipedia – The Free Encyclopedia, <http://en.wikipedia.org/wiki/Cache_memory>, (27.06.2004.)
4. Waterwheel, *Memory/Cache*, <http://www.waterwheel.com/Guides/how_to/memory/memory.htm>, (27.06.2004.)
5. Radić, Drago, *Informatička abeceda*, <<http://www.ktf-split.hr/informatika/>>, (27.06.2004.)
6. Bong, Duane, *Cache Memory*, <http://www.visionengineer.com/comp/cache_memory.shtml#top>, (27.06.2004.)
7. Arunović, Denis, *Brzi i žestoki*, časopis "BUG", broj 139.(lipanj 2004.), str. 41.-43.
8. Bobinac, Mladen, *Procesori*, časopis "BUG", broj 107. (listopad 2001.), str. 91.-110.
9. *U središtu procesora*, časopis "Enter", broj 1. (listopad 2002.), umetak (slika)
10. Bobinac, Mladen, *AMD vs. Intel*, Dvoboj na tržištu procesora, časopis "BUG", broj 112. (ožujak 2002.), str. 79.-91.