

Memorija

Memorija je svaki uređaj koji je u stanju da podatke sačuva u formatu koji neka mašina može da prepozna
James A. Sean

Među delovima računarskog sistema, memorija računara ispoljava najširi spektar tipova, tehnologija koje se koriste za njenu izradu, organizacija, performansi i cena. Ne postoji jedna vrsta memorije koja optimalno zadovoljava sve memorijske zahteve računarskog sistema. Kao posledica toga, tipičan računarski sistem je opremljen hijerarhijom memorijskih podsistema, od kojih su neki unutrašnji (direktno dostupni procesoru), a neki spoljni (dostupni procesoru preko U/I kontrolera).

Karakteristike memorije

Podela memorije se može izvršiti na razne načine, zavisno od karakteristike memorije u odnosu na koju se podela vrši. Na slici 1. je dat pregled nekih karakteristika na osnovu kojih je moguće izvršiti podelu memorija.



Slika 1. Bitne karakteristike računarske memorije

Termin lokacija na slici 1 odnosi se na to da li je memorija unutrašnja (interna) ili spoljašnja (eksterna) u odnosu na centralni procesor. Interna memorija je često izjednačava sa glavnom memorijom. Ali postoje i drugi oblici unutrašnje memorije. Procesor zahteva sopstvenu lokalnu memoriju, u obliku registara. Dalje, kontrolna jedinica procesora može da traži svoju unutrašnju memoriju. Keš (*Cache*) je još jedan oblik interne memorije. Eksterna memorija se sastoji od perifernih uređaja, kao što su disk i trake, koji su dostupni na procesor preko U/I kontrolera.

Očigledna karakteristika memorije je i njen kapacitet. Za interne memorije, on se obično izražava u bajtova ili rečima. Uobičajene dužine reči su 8, 16, 32 i 64 bita. Kapacitet spoljne memorije se izražava u bajtovima. Srodan koncept je jedinica prenosa. Za interne memorije, jedinica prenosa može biti jednaka dužini reči, ali je često veća. Da bismo ovo razjasnili, razmotrićemo tri povezana koncepta interne memorije:

- **Reč:** “prirodna” jedinica organizacije memorije. Veličina reči je obično jednaka broju bitova koji se koriste za predstavljanje celog broja, ili dužini instrukcije koja se obrađuje u procesoru.
- **Adresibilna jedinica:** u nekim sistemima adresibilna jedinica je reč, dok neki sistemi dozvoljavaju adresiranje na nivou bajta. U svakom slučaju, ako je dužina adrese u bitovima jednaka m , onda se svim adresama koje se mogu napisati sa m bitova može adresirati 2^m adresibilnih jedinica.
- **Jedinica prenosa:** Za unutrašnju memoriju, ovo je broj bitova koji su pročitani iz ili su upisani u memoriju u datom trenutku. Jedinica transfera ne mora biti jednaka jednoj reči ili adresibilnoj jedinici. Kod eksternih memorija, podaci se često prebacuju u mnogo većim jedinicama od reči, a one se nazivaju blokovi.

Način pristupa memoriji je još jedna karakteristika po kojoj memorije delimo na:

- **Memorije sa sekvencijalnim pristupom:** memorija je organizovana u jedinice podataka, koji se nazivaju slogovi (*records*). Pristup slogovima se vrši sekvencijalno – jedan po jedan slog. Otuda vreme pristupa dosta varira, jer zavisi od trenutne pozicije i pozicije sloga koji treba pročitati. Primer su magnetne trake.
- **Memorije sa direktnim pristupom:** Slogovi su organizovani u blokove, pri čemu svaki blok ima jedinstvenu adresu i pristupa mu se direktno. Konkretnom slogu u okviru bloka se dalje pristupa sekvencijalno. Vreme pristupa je i dalje promenljivo, ali kraće nego kod sekvencijalnog pristupa. Primer su diskovi.
- **Memorije sa slučajnim pristupom:** svaka adresibilna jedinica memorije ima jedinstvenu adresu. Vreme pristupa željenoj lokaciji ne zavisi od prethodnih pristupa i konstantno je. Dakle, proizvoljno možemo odabrati neku lokaciju i pristupiti joj direktno preko njene adrese. Primer je glavna memorija.
- **Memorije sa asocijativnim pristupom:** podacima se pronalaze na osnovu sadržaja. Dakle, reč se preuzima na osnovu jednog dela njenog sadržaja koji se poklapa sa sadržajem koji tražimo, a ne na osnovu njene adrese. Ovo je jedna vrsta memorije sa slučajnim pristupom, pa i ovde svaka lokacija ima svoju adresu, i vreme pristupa je konstanta nezavisna od lokacije. Keš memorije mogu koristiti asocijativni pristup.

Sa aspekta korisnika, dve najbitnije karakteristike memorije su njen kapacitet i performanse. Za određivanje performansi memorije, koriste se tri parametra:

Vreme pristupa (latencija): Za memorije sa slučajnim pristupom, ovo je vreme potrebno da se izvrši operacija čitanja iz ili pisanja u memoriju, to jest vreme koje protekne od momenta kada se započne pristup nekoj adresi u memoriji, do trenutka kada je podatak sačuvan ili dostupan za upotrebu. Za memorije koje nemaju slučajni pristup, vreme pristupa je vreme potrebno da se mehanizam za čitanje i pisanje postavi na željenu memorijsku lokaciju.

Vreme memorijskog ciklusa je vezano prvenstveno na memorije sa slučajnim pristupom i izračunava se kao suma vremena pristupa i dodatno vremena koje protekne pre nego što se može izvršiti naredni pristup memoriji.

Brzina prenosa: Brzina kojom se podaci mogu preneti iz ili u memorijsku jedinicu. Za memorije sa slučajnim pristupom jednaka je $1/\text{vreme memorijskog ciklusa}$.

Što se tiče fizičkih tipova memorije, tu su poluprovodničke memorije, memorije sa magnetnom površinom, optički i magnetno-optičke memorije.

U zavisnosti od fizičkih karakteristika memorije mogu imati različitu stalnost zapisa. Kod privremenih memorija, podaci se gube kada je električna energija isključena. Kod trajne memorije, podaci koji su jednom sačuvani, ostaju nepromenjeni sve dok ih neko namerno ne promeni; električna energija nije potrebna da bi se podaci zadržali. Magnetne memorije su trajne. Poluprovodničke memorije mogu biti privremene ili trajne. Neizbrisiva memorija se ne može menjati, osim uništavanjem. Poluprovodnička memorija ovog tipa je poznata kao memorija samo za čitanje (ROM). Neizbrisiva memorija je svakako trajna.

Hijerarhija memorija

Ograničenja dizajna računarske memorije mogu se sumirati kroz tri pitanja: Kog je kapaciteta? Koliko je brza? Koliko košta?

Gornja granica za kapacitet ne postoji. Formiraju je zahtevi aplikacija. Što se tiče brzine, da bi memorija postigla najbolje performanse, ona morati biti u skladu sa procesorom. Cena memorije mora biti razumna i u skladu sa cenom ostalih komponenti računarskog sistema.

Između kapaciteta, brzine i cene važe sledeće relacije:

Što je brži pristup, veća je cena po bitu

Što je veći kapacitet, niža je cena po bitu

Što je veći kapacitet, sporiji je pristup

Sa jedne strane, dizajneri bi hteli da koriste tehnologije koje obezbeđuju memorije velikog kapaciteta, zato što je veliki kapacitet poželjan, a cena po bitu je niska. Ipak, da bi ispunio zahteve vezane za performanse memorije, dizajner mora da koristi skupe memorije, sa relativno malim kapacitetom, ali kratkim vremenom pristupa.

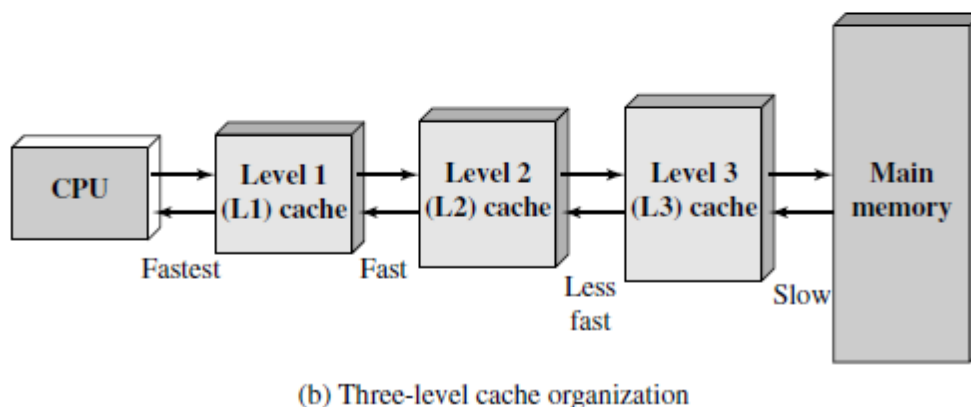
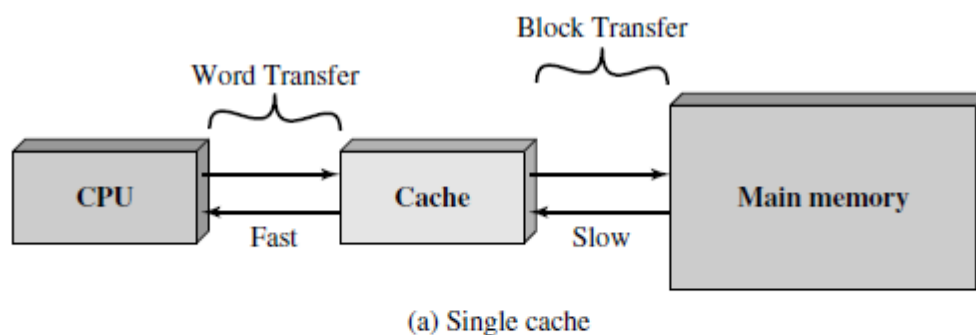
Da bi se pomirila ove dve suprotstavljene težnje, u računarski sistemi se ne oslanjaju na jednu memorijsku komponentu ili jednu tehnologiju za izradu memorija, već koriste različite komponente i tehnologije koje zajedno čine hijerarhiju memorija. Tipična hijerarhija je prikazana na slici 2. Ako se krećemo od vrha na dole, možemo uočiti sledeće:

- a. Smanjuje se cena po bitu
- b. Povećava se kapacitet
- c. Povećava se vreme pristupa
- d. Snižava se učestalost pristupa memoriji od strane procesora.

Otuda se, manje, skuplje i brže memorije nadopunjuju većim, jeftinijim i sporijim memorijama. Ključ uspeha ovakve organizacije leži u stavki d. smanjenju frekvence pristupa.

Najbrži, najmanji, a najskuplji tip memorije se sastoji od unutrašnjih registara procesora. Tipično, procesor će sadržati nekoliko desetina registara, iako neke mašine sadrže stotine registara. Glavna

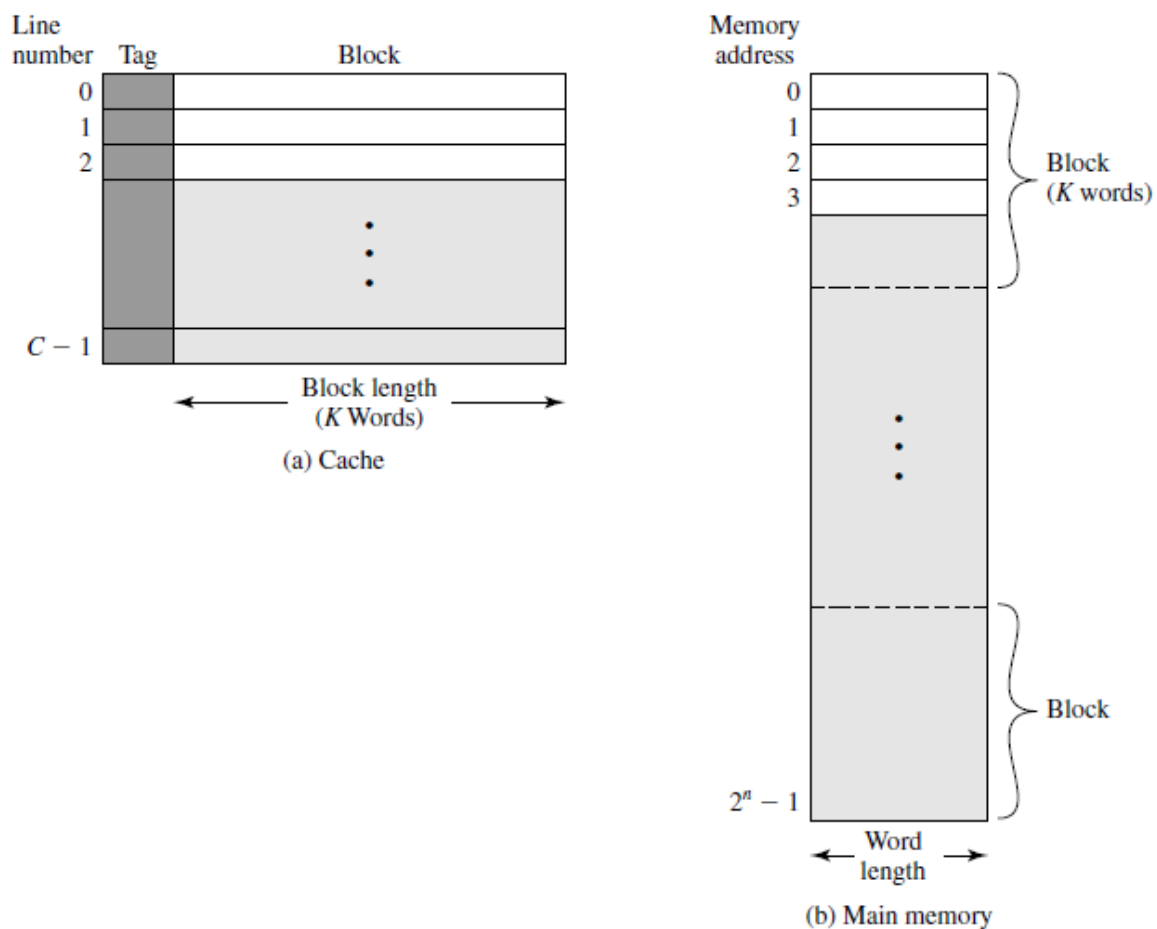
memorija je obično proširena sa manjom, ali bržom keš memorijom. Keš obično nije vidljiv za programera ili procesor. Njena uloga se ogleda u omogućavanju etapnog prenošenja podataka između glavne memorije i procesorskih registara radi poboljšanja performansi. Keš sadrži kopije delova glavne memorije. Kada procesor pokuša da pročita reč iz glavne memorije, proverava se da li ona već postoji u kešu, i ako je prisutna dostavlja se procesoru. Ako se tražena reč ne nalazi u kešu, blok glavne memorije koji između ostalog sadrži i traženu reč, biva prebačen u keš, a reč se dostavlja procesoru. Razlog za prenošenje čitavog bloka podataka leži u činjenici da se pretpostavlja da će se procesor i dalje referencirati na deo memorije u kome se nalazi reč koju je prethodno tražio. Procesor obično izvršava programe koji se sastoje od petlji i poziva podprograma. Kada jednom uđe u petlju ili započne podprogram, procesor se više puta za redom referencira na mali broj instrukcija, odnosno na jedan deo glavne memorije.



Slika 2. Organizacija keš memorije

Slika 2a) prikazuje upotrebu jednog, a slika 2b) upotrebu tri nivoa keš memorije. L2 keš je sporiji i obično veći od L1 keša, dok je L3 keš sporiji i veći od L2 keša.

Slika 3 prikazuje strukturu sistema keš/glavna memorija. Glavna memorija se sastoji od najviše 2^n adresibilnih reči., pri čemu svaka reč jedinstvenu n -bitnu adresu. Pretpostavimo da glavna memorija sadrži određeni broj blokova fiksne veličine od po K reči svaki. Dakle, u memoriji se nalazi $M=2^n/K$ blokova. Keš se sastoji od m blokova koji se nazivaju linije. Svaka linija sadrži K reči i dodatnu oznaku od nekoliko bita. Dodatno, reč sadrži i kontrolni bit koji govori o tome da li je linija bila izmenjena od kako je poslednji put učitana u keš. S obzirom na to da postoji više blokova memorije nego linija u kešu, pojedinačne linije ne mogu biti jedinstveno i trajno dodeljene određenim blokovima memorije. Otuda, svaka linija sadrži oznaku koja identifikuje blok memorije koji je trenutno smešten u liniju keša.



Slika 3. Struktura sistema keš/glavna memorija

Dinamička RAM-a (DRAM) je napravljen sa ćelijama koje skladište podatke kao naelektrisanje u kondenzatorima. Prisustvo ili odsustvo naelektrisanja u kondenzatoru se tumači kao binarno 1 ili 0. Zbog prirodne tendencije kondenzatora da se prazne, dinamički RAM zahteva periodično osvežavanje naelektrisanja za održavanje čuvanja podataka. Termin dinamičan se odnosi na ovu tendenciju curenja memorisanog naelektrisanja, čak i pri stalnom dotoku električne energije.

Kao što naziv sugeriše, memorija samo za čitanje (ROM) sadrži stalne podatke, koji se ne mogu menjati. Rom je trajna; to jest, nijedan izvor napajanja nije potreban za održavanje vrednosti bitova u memoriji. Dok je moguće da se pročita ROM, nije moguće pisati nove podatke u njega. ROM memorija se koristi za mikroprogramiranje.

Spoljašnje memorije

Hard disk

Hard diskovi su krute, kružne ploče, sastavljene od podloge i magnetnog medijuma. Podloga na ploči je osnovni materijal koji mora biti nemagnetni i takav da se može mašinski obrađivati dok ne postane gladak. Izrađuje se od legure aluminijuma ili mešavine stakla i keramike. Da bi se omogućilo skladištenje podataka, obe strane svake podloge su obložene su magnetnim medijumom - ranije magnet oksidom, ali sada, gotovo isključivo, slojem metala koji se zove tanki film. Ovaj skladišti podatke u vidu magnetnih obrazaca, na svakoj ploči može da sačuva milijarda i više bita po kvadratnom inču (bps) površine.

Kada hard disk prolazi kroz formatiranje niskog nivoa, on biva podeljen na staze i sektore. Ovaj format niskog nivoa dešava samo jednom u životu diska, i to pre nego što napusti fabriku proizvođača. Sva naredna formatiranja hard diskova polaze od ovog početnog formata niskog nivoa, koji sadrži informacije o stazama, raspored sektora, upravljanje kvarova i tako dalje.

Staze su koncentrični krugovi oko centralne osovinu na obe strane svake ploče. Svaka staza je iste širine kao glava. Susedne staze su razdvojene prazninom. Ovo sprečava, ili bar smanjuje, greške zbog nepravilno centrirane glave ili od mešanja magnetnih polja. Podaci se prenose na i sa diska u sektorima. Svaki sektor je obično veličine od 512 bajta. Susedni sektori su odvojeni međusektorske praznina.

Staze koje se fizički nalaze jedne iznad drugih na pločama, grupišu se u cilindre.

Solid state hard drive (SSD) technology guide

A solid state drive is a storage device that uses solid state memory to store data. While technically not a disk, a solid state drive will often be referred to as a solid state disk drive, or a solid state disk, in reference to the fact that, in some ways, it replaces the traditional hard disk drive.

Hard disk drives have been a faithful servant to computing for many years. But with heads, platters, magnetic surfaces, spindles and an array of other complex moving parts, they are most certainly fallible. They can be slow, too: disks have to start spinning if they're not already doing so, then they have to get the head to the correct position on the disk to read or write the data. Add this to the physical problems occurring when a piece of dust or other foreign object gets into the mechanism, or when an impact jolts the drive, and we have a distinctly imperfect system. Solid state drives address many of these timing and structural problems inherent in the hard disk drive.



Slika 4. Prikaz klasičnog tvrdog diska i diska čvrstog stanja

The principle behind solid state drives is that there should be no moving parts: no spinning platters, no moving heads. Data is split into word length pieces and stored in memory. It is then accessed almost instantaneously using unique system-wide addresses. This behaviour has been used in computer RAM for many years, but for a long time it was too expensive for manufacturers to consider using it as persistent storage in sufficient volumes to replace the hard disk drive.

Solid state disks use either NAND flash or SDRAM (non-volatile and volatile storage respectively). NAND flash is so-called because of the NAND-gate technology it uses and is common in USB flash drives and many types of memory card. NAND flash based drives are persistent and can therefore effectively mimic a hard disk drive. Synchronous dynamic random access memory (SDRAM) is volatile and requires a separate power source if it is to operate independently from a computer.

Solid state drives may be preferred over traditional disk drives for a number of reasons. The first advantage is found, as mentioned briefly above, in the speed of operation. Because hard disk drives need to be spinning for the head to read sectors of the platter, sometimes we have to wait for spin up time. Once the disk is spinning, the head must seek the correct place on the disk, and from there the disk must spin just enough so that the correct data is read. If data is spread over different parts of the disk (fragmented) then this operation is repeated until all the data has been read or written. While each individual operation only takes fractions of a second the sum of them may not. It is often the case that reads to and writes from the hard disk are the bottleneck in a system.

Because the information on solid state drives can be accessed immediately (technically at the speed of light) there is no latency experience when data is transferred. Because there is no relationship between spatial locality and retrieval speed, there is no degradation of performance when data is fragmented.

Consequences of the increased speed of writes for fragmented data include a much decreased application start up time: SanDisk, for instance, claim to have achieved Windows Vista start up times of around 30 seconds for a laptop with its SSD SATA 5000 2.5.

Solid state drives also enjoy greater stability over their disk counterparts. Because there are no moving parts there is less that can go wrong mechanically. Dust entering the device ceases to become a problem (and in any case solid state drives can be sealed air tight unlike disk drives which require a

certain air cushion to function properly), and dropping the drive is less likely to cause damage to the data. There are no heads so head crashes are a thing of the past.

This speed and stability comes at a price, of course, and in early models prices of even the most modest of solid state capacities greatly surpassed that of the largest hard disks.

Optičke memorije

In 1983, one of the most successful consumer products of all time was introduced: the compact disk (CD) digital audio system. The CD is a nonerasable disk that can store more than 60 minutes of audio information on one side. The huge commercial success of the CD enabled the development of low-cost optical-disk storage technology that has revolutionized computer data storage. A variety of optical-disk systems have been introduced. We briefly review each of these.

CD

Compact Disk. A nonerasable disk that stores digitized audio information. The standard system uses 12-cm disks and can record more than 60 minutes of uninterrupted playing time.

CD-ROM

Compact Disk Read-Only Memory. A nonerasable disk used for storing computer data. The standard system uses 12-cm disks and can hold more than 650 Mbytes.

CD-R

CD Recordable. Similar to a CD-ROM. The user can write to the disk only once.

CD-RW

CD Rewritable. Similar to a CD-ROM. The user can erase and rewrite to the disk multiple times.

DVD

Digital Versatile Disk. A technology for producing digitized, compressed representation of video information,

as well as large volumes of other digital data. Both 8 and 12 cm diameters are used, with a double-sided capacity of up to 17 Gbytes. The basic DVD is read-only (DVD-ROM).

DVD-R

DVD Recordable. Similar to a DVD-ROM. The user can write to the disk only once. Only one-sided disks can be used.

DVD-RW

DVD Rewritable. Similar to a DVD-ROM. The user can erase and rewrite to the disk multiple times. Only one-sided disks can be used.

Blu-Ray DVD

High definition video disk. Provides considerably greater data storage density than DVD, using a 405-nm

(blue-violet) laser. A single layer on a single side can store 25 Gbytes.