

Memorijska hijerarhija i radna memorija

Literatura: Hennessy & Patterson, 2.1–2.2

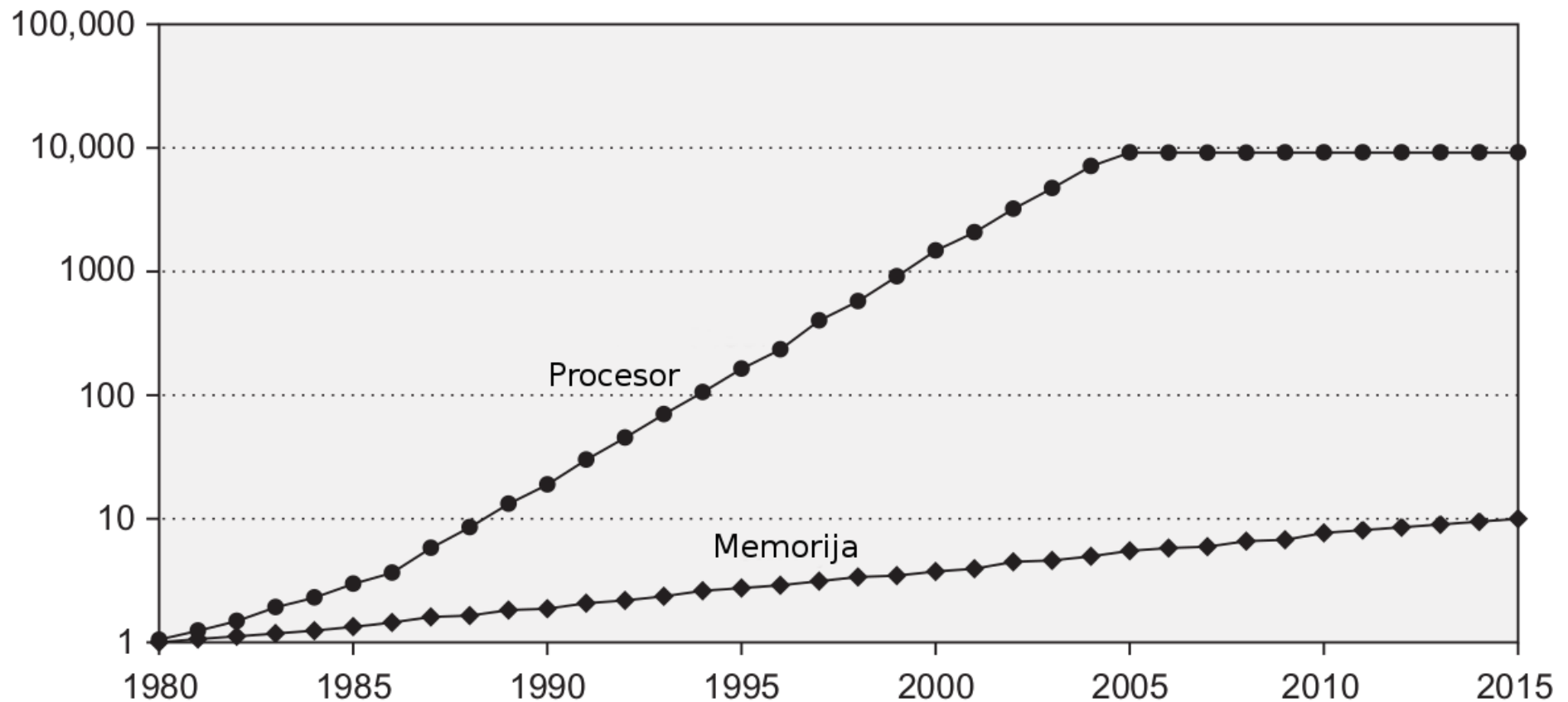
Ideally one would desire an indefinitely large memory capacity such that any particular... word would be immediately available... We are... forced to recognize the possibility of constructing a hierarchy of memories each of which has greater capacity than the preceding but which is less quickly accessible.

Burks, Goldstine, von Neumann (1946)

Memorijska hijerarhija

- Što je izvor podataka bliži procesoru, ima veće performanse, višu cenu po jedinici kapaciteta i manji kapacitet
- Hijerarhija postoji od početka razvoja računara, a nivoi su bili srazmerno stabilni tokom decenija
- Standardni nivoi:
 - Procesorski registri
 - Keš memorija
 - Radna memorija
 - [PCM, Flash] masovna memorija
 - Magnetni diskovi

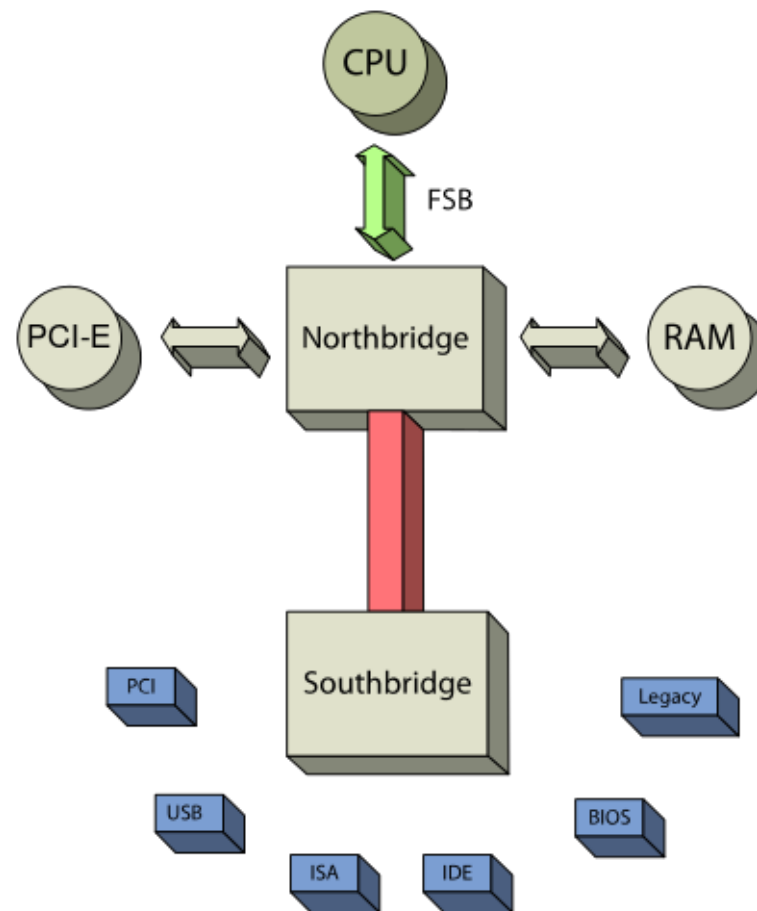
Osnovna disproporcija u performansama



Tipičan hardver računara opšte namene

- Jedan do dva fizička procesora (*socket*)
 - retko se ide na više fizičkih procesora zbog komplikovanja konstrukcije
- Prostor za 16 GB – 256 GB dinamičke memorije
 - velike razlike između serverskih i desktop varijanti
 - za veće servere su dostupni i memorijski kapaciteti reda TB
- Integrirani periferni uređaji
- Relativno slična osnovna arhitektura

Šema hardverske organizacije (do ~2011.)



Osnovni elementi organizacije

- Northbridge: direktno komunicira s procesorom
 - FSB = *Front Side Bus*
 - zadužen za upravljanje memorijom
 - dodatno, za brze periferne uređaje
- Southbridge: ostali periferni uređaji
 - sve sporije magistrale
 - memorija gde brzina pristupa nije kritična (BIOS)

Posledice organizacije (1)

- Komunikacija između fizičkih procesora koristi FSB
- Sva komunikacija s memorijom ide preko NB
 - memorija ima jednu pristupnu tačku (port)
 - višestruki pristup kod specijalizovanih uređaja
- Komunikacija sa perifernim uređajima zakačenim na SB ide preko NB

Posledice organizacije (2)

- Brzi periferni uređaji direktno pristupaju memoriji
 - DMA (*Direct Memory Access*)
 - Ovime se rasterećuje procesor...
- ... Ali stvara konkurenciju s procesorom za pristup memoriji preko NB
- Veza NB s memorijom i njeno iskorišćenje su kritični za performanse

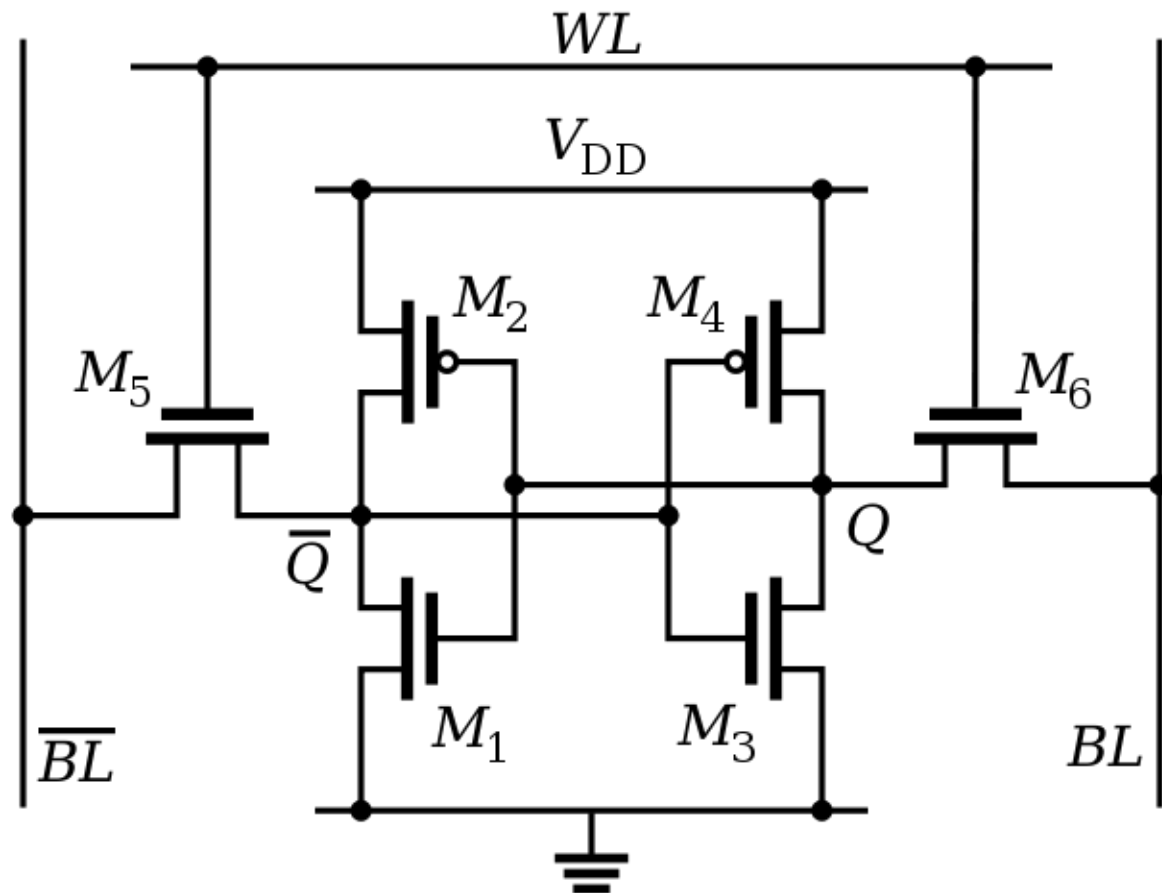
Kako povećati propusnu moć (1)

- NB ne mora sam da upravlja memorijom, već za to mogu postojati posebna kola
- Više memorijskih kontrolera na NB
 - veća količina podržane memorije
 - bolja propusna moć
- Ograničenje postaje interna propusna moć NB

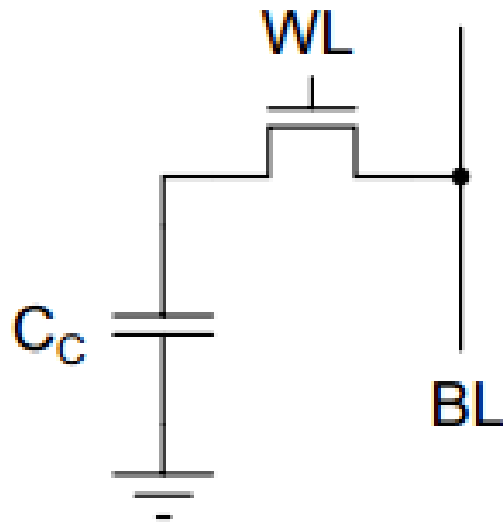
Kako povećati propusnu moć (2)

- Integrirani memorijski kontroler na fizičkom procesoru
 - današnja varijanta
- U ovoj organizaciji ne postoji poseban NB, SB se zove „Platform Controller Hub“ (kod Intela)
- Više procesora znači i više kontrolera, sa prednostima kao kod prethodne varijante
- Problem: kod više fizičkih procesora, pristup memoriji više nije uniforman
 - ekstremna varijanta: NUMA (*Non-Uniform Memory Access*)

Statički RAM (6 tranzistora)



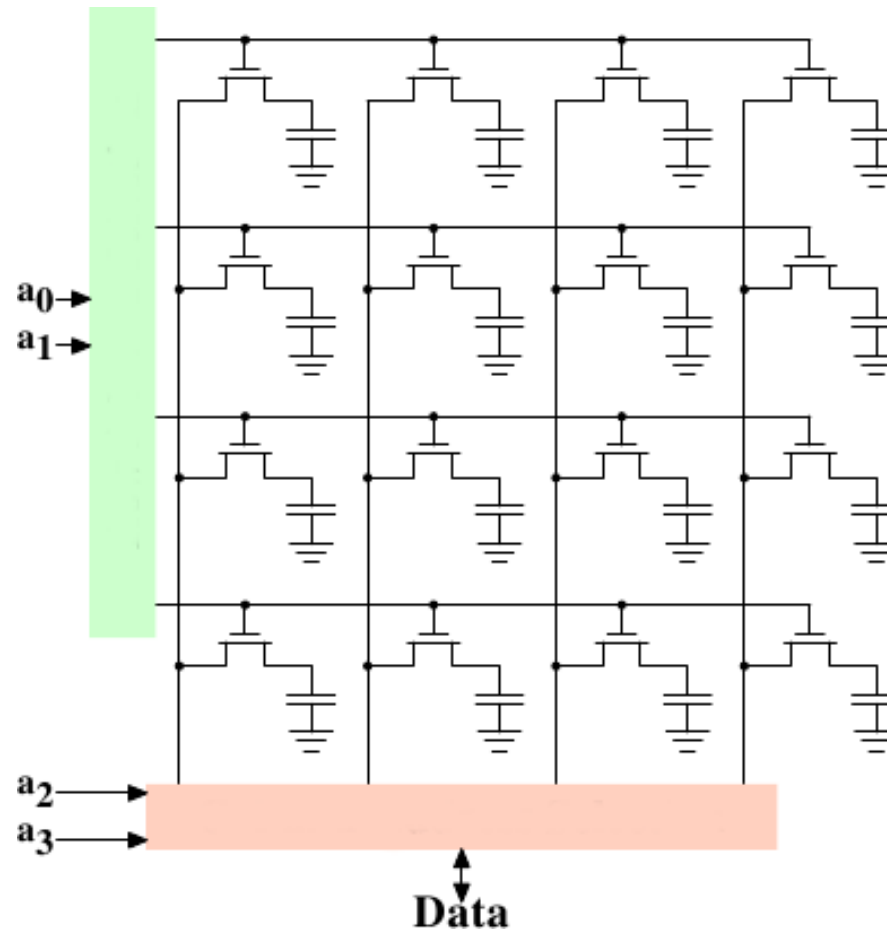
Dinamički RAM (1 tranzistor)



Opšte napomene

- Dinamički RAM je očigledno jednostavniji od statičkog
- Problemi:
 - čitanje sadržaja je posle izvesnog broja ciklusa destruktivno
 - kapacitet kondenzatora ne može biti veliki
 - mora se periodično osvežavati
- U toku osvežavanja pristup je nemoguć
- Direktno čitanje je nemoguće
- Moraju postojati pauze tokom čitanja/pisanja

Organizacija dinamičkog RAM-a (1)



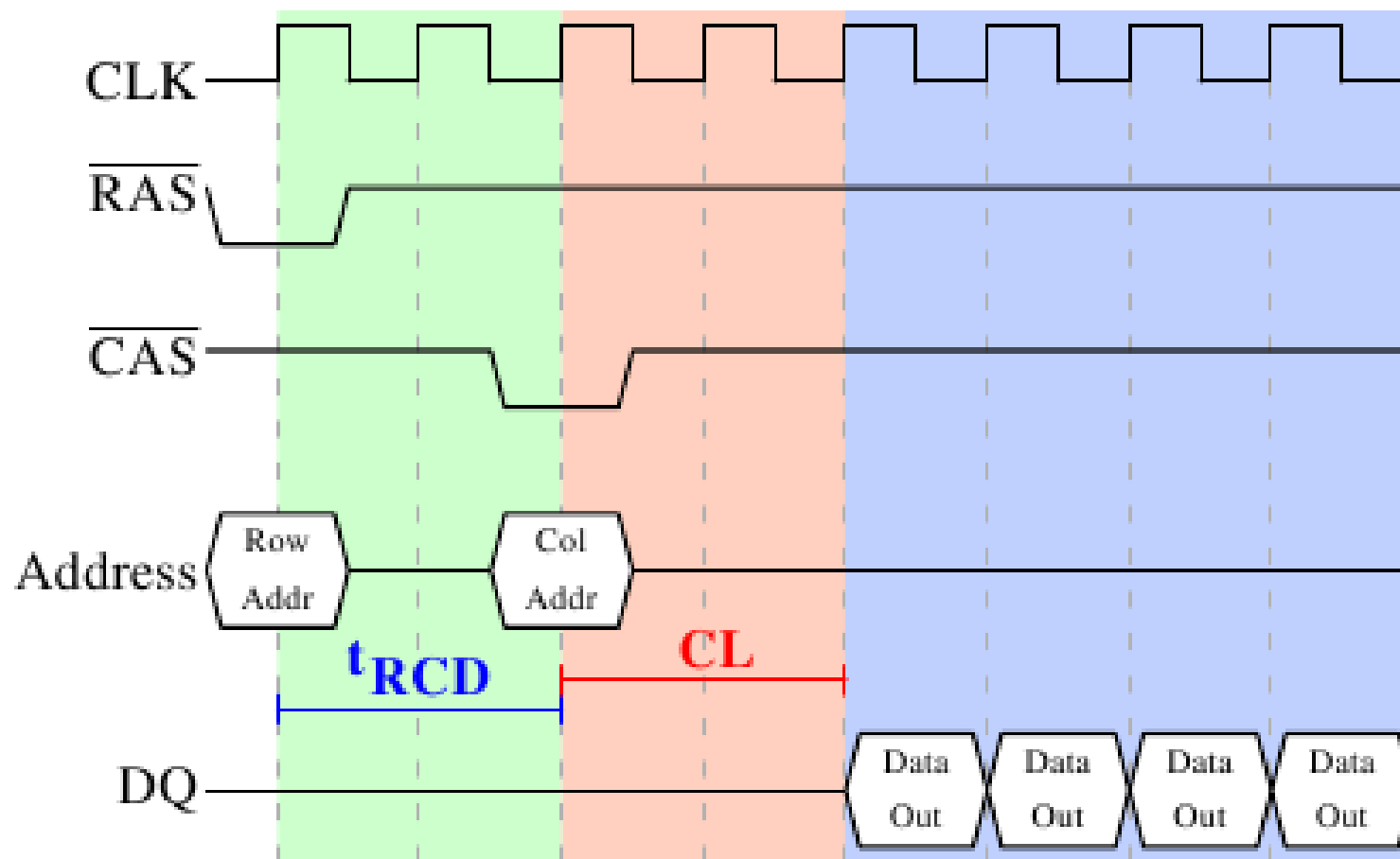
Organizacija dinamičkog RAM-a (2)

- Čelije su organizovane u matricu, koja treba da ima jednak broj vrsta i kolona ako je moguće
 - raskorak komplikuje hardver za (de)multipleksiranje na većoj strani
- Vrsta se bira signalom \overline{RAS} (*Row Address Selection*)
 - linija iznad znači da je signal invertovan
- Kolona se bira signalom \overline{CAS} (*Column Address Selection*)

Organizacija dinamičkog RAM-a (3)

- $\overline{\text{RAS}}$ se demultipleksira na osnovu dela adrese (tj. signal se usmerava na jednu od vrsta u zavisnosti od kombinacije **a0/a1**)
- Aktiviranje vrste će kao rezultat imati iščitavanje svih ćelija u okviru te vrste, čiji se signali sprovode do multipleksera kolona
- $\overline{\text{CAS}}$ aktivira multiplekser na osnovu ostatka adrese, kombinacije **a2/a3**, i na izlazu se pojavljuje očitavanje adresirane ćelije
- Sama adresa je često multipleksirana

Dijagram pristupa RAM-u



Posledice načina pristupa

- Pristup je mnogo sporiji u odnosu na kapacitet procesora
 - vremenski, 10–15:1
- Pristup proizvoljnoj lokaciji zahteva čekanje
- Prenos susednih lokacija je efikasan
- Moguće je raditi čitanje unapred da bi se smanjilo čekanje

Drugi korisnici memorije

- Grafički podsistem
 - integrisana grafika bez odvojene memorije
 - ako se zahtevaju bolje performanse ove komponente su izdvojene
- Disk kontroleri
- Mrežni kontroleri
 - uporedivi sa (magnetnim) diskovima po sirovoj brzini