Multiprocesorski sistemi

Interkonekcione mreže

Marko Mišić, Milo Tomašević

13S114MUPS, 13E114MUPS, 13M114MUPS

2024/2025.

Interkonekcione mreže

- Povezuju:
  - Procesore sa procesorima, keš memorijama i memorijama
  - Keš memorije sa keš memorijama
  - U/I uređaje
- Uticaj na skalabilnost
  - Veličina sistema?
  - Lakoća dodavanja novih čvorova (npr., procesora)
- Uticaj na performanse i energetsku efikasnost
  - Brzina komunikacije procesora, keš memorija i memorija?
  - Dužina latencije u pristupu memoriji?
  - Energija potrošena za komunikaciju?

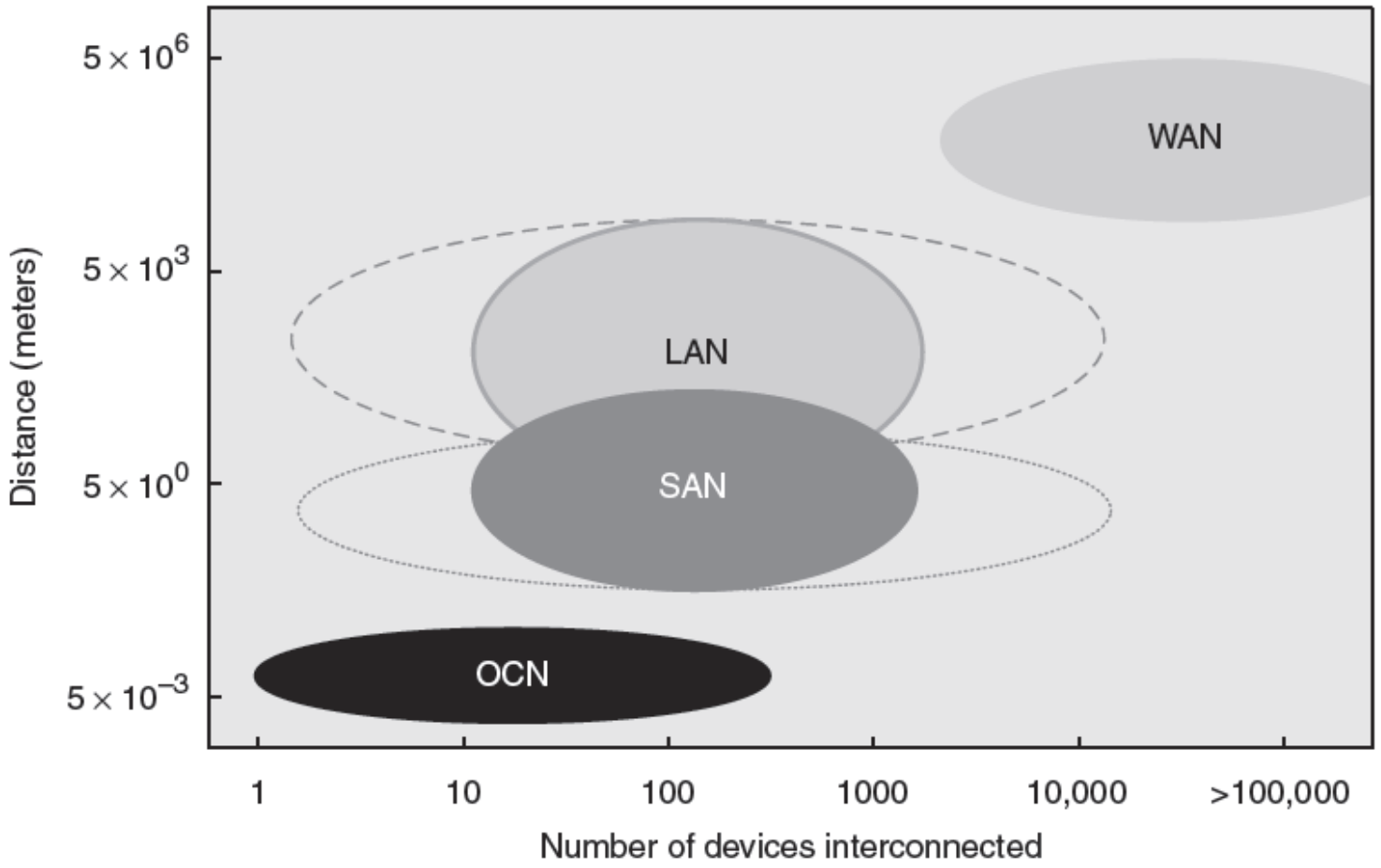
Projektne odluke

- Topologije
  - Kako su povezani čvorovi i ruteri/prekidački elementi
  - Obično prva odluka zbog uticaja na propusnost, latenciju, cenu/složenost implementacije, ...
- Rutiranje
  - Kako poruka stiže od izvora do odredišta?
  - Predefinisana putanja ili adaptivna u odnosu na uslove?
- Baferovanje i kontrola toka
  - Šta se čuva tokom prenosa (paketi, delovi paketa, ...)?
  - Tehnike rada sa baferima
  - Blisko vezano sa algoritmima rutiranja

Domeni mreža

- OCN (*On-Chip Networks*)
  - Povezivanje delova mikroarhitekture (PE, FU, $M, regs)
  - Obično posebno dizajnirane
- SAN (*System/Storage Area Networks*)
  - Unutar procesora, između procesora i memorije u MP, kao i za spoljne memorije i U/I u serverima i *data* centrima
  - npr., InfiniBand do 120 Gbps, na distanci od 300 m
- LAN (*Local Area Networks*)
  - Povezivanje relativno bliskih nezavisnih sistema (npr., PC)
  - Npr., Ethernet - 10 Gbps standard, može i do 40 km
- WAN (*Wide Area Networks*)
  - Ogroman broj sistema na velikim udaljenostima (npr., ATM)

Domeni mreža

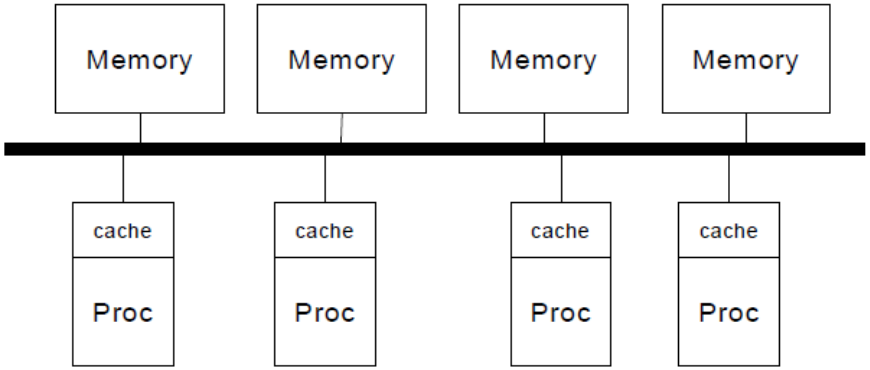


Vrste mreža

- Indirektne
  - Zasnovane na prekidačkim elementima (*switch-based*)
  - Prekidački elementi izvan terminalnog čvora
  - Prekidački elementi imaju ulazne i izlazne portove
  - Veze se dinamički uspostavljaju (dinamičke mreže)
  - Pogodne za različite šeme komunikacije
- Direktne
  - …

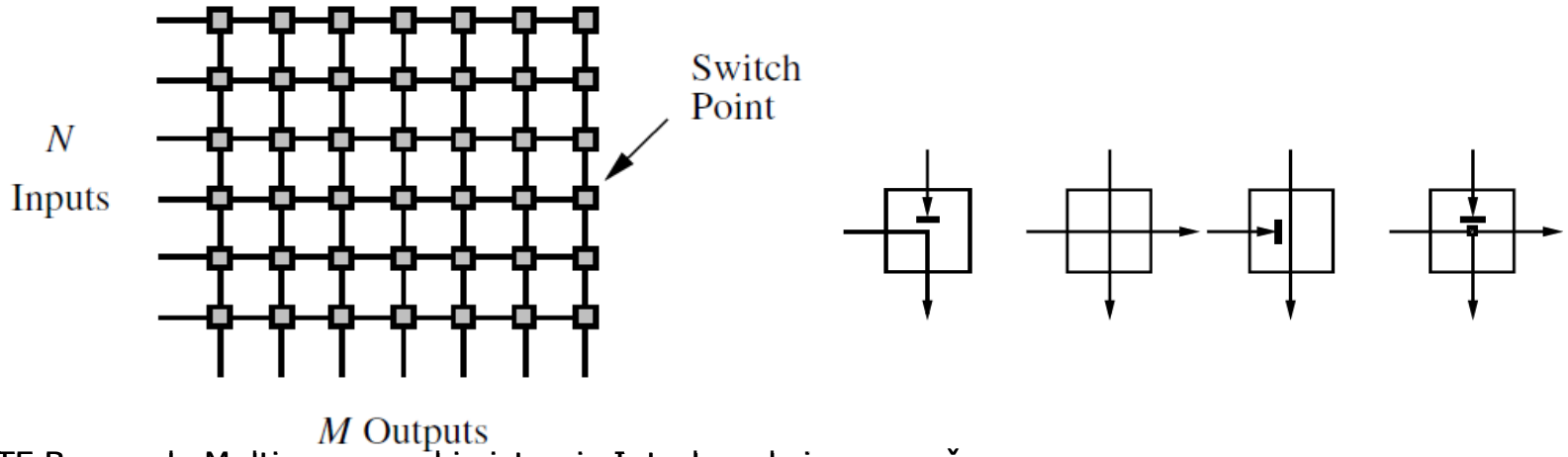
Magistrala

- Prednosti
  - Jednostavnost
  - Ekonomski isplativa za manje sisteme
  - Olakšava održavanje koherencije
- Mane
  - Neskalabilna (ograničen propusni opseg)
  - Ekskluzivno korišćenje (samo jedan istovremeni transfer)
  - Ograničenja električnih karaktersitika



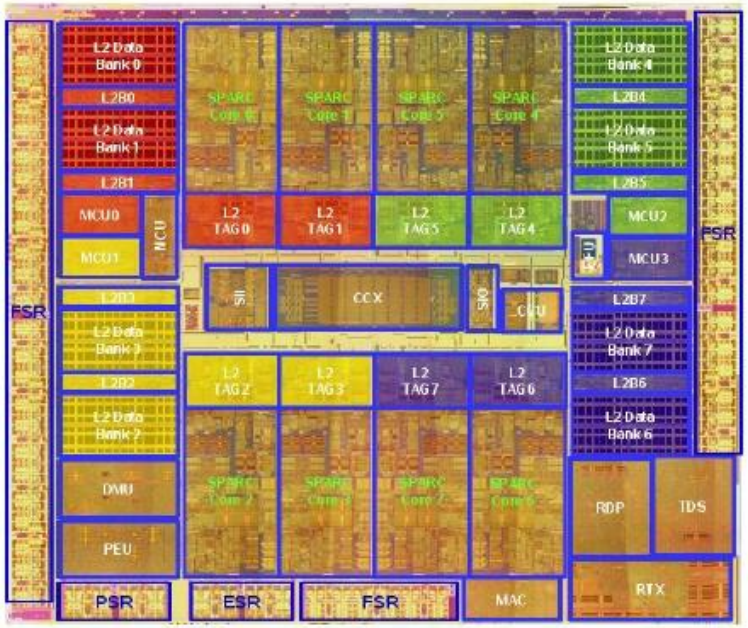
Krosbar

- Dozvoljava povezivanje svakog ulaza sa svakim izlazom, ako je izlazni port slobodan
- Prednosti
  - Niska latencija (1 hop), veliki propusni opseg, neblokirajuća
- Mane
  - Skupa, neskalabilna po ceni O(n^2), distribuirana arbitracija
- Implementacija većeg krosbara podelom na manje ($N \times N$ krosbar kao $n \times n$ krosbara dimenzija $N/n \times N/n$)



Krosbar

Primer - Krosbari za povezivanje procesorskih jezgara sa poslednjim nivoom keš hijerarhije

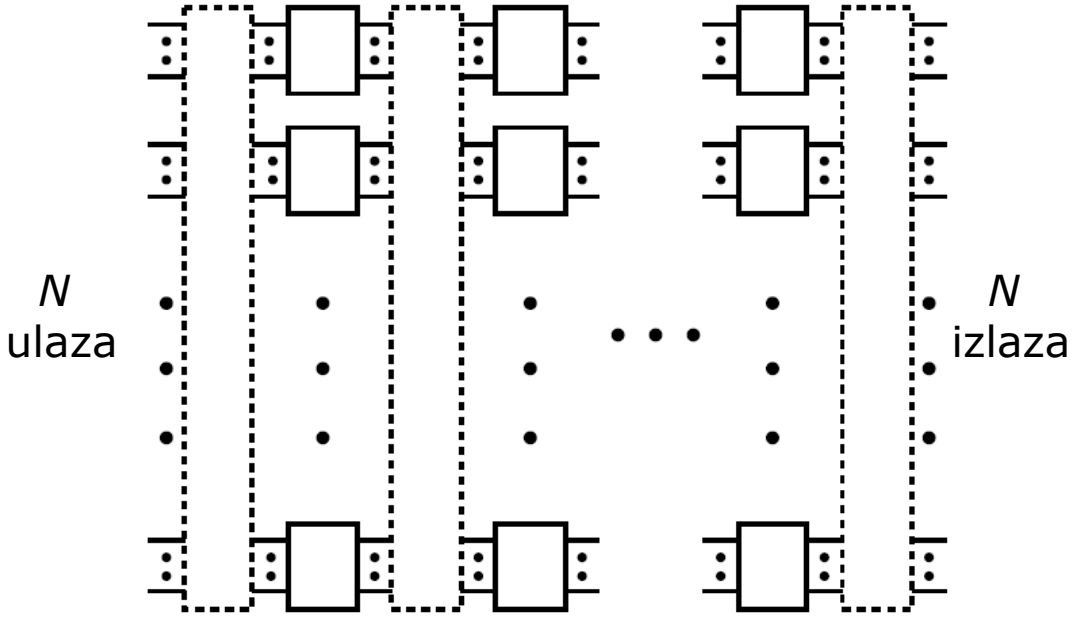


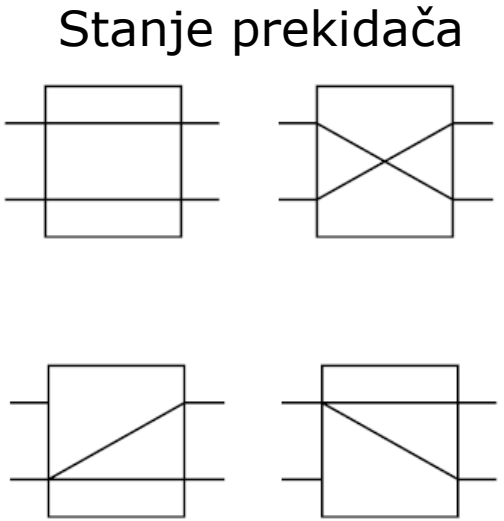
Sun SPARC T2 (8 cores, 8 L2 cache banks)

Oracle SPARC T5 (16 cores, 8 L3 cache banks)

MIN – Višestepene sprežne mreže

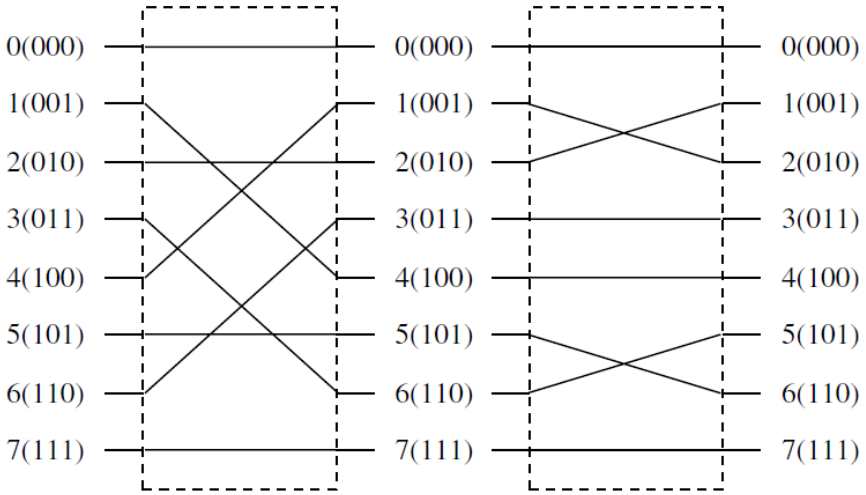
- Kompromis između magistrale i krosbara
- Više stepeni ($\log_k N$) prekidačkih elemenata k x k (N/k u svakom stepenu)
- Dinamičko rutiranje kroz distribiranu kontrolu
- Složenost O(N log N), latencija O(log N)

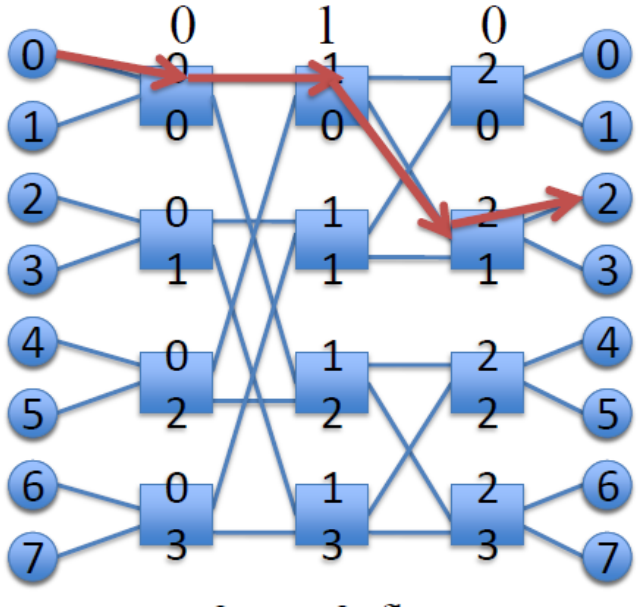




MIN – Višestepene sprežne mreže

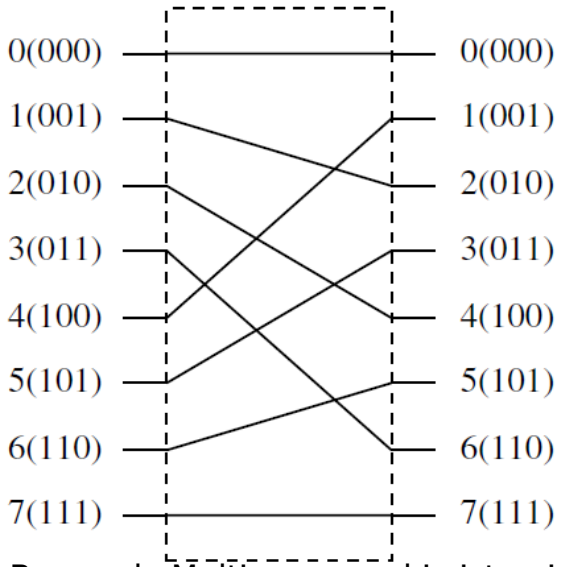
- Fiksna latencija (nema koristi od lokalnosti)
- Veze između stepeni - permutacije ulaza daju izlaze
- Na svakom stepenu jedan bit odredišne adrese određuje izlaz prekidačkog elementa
- Npr., Butterfly (menja nulti i drugi bit, pa nulti i prvi)

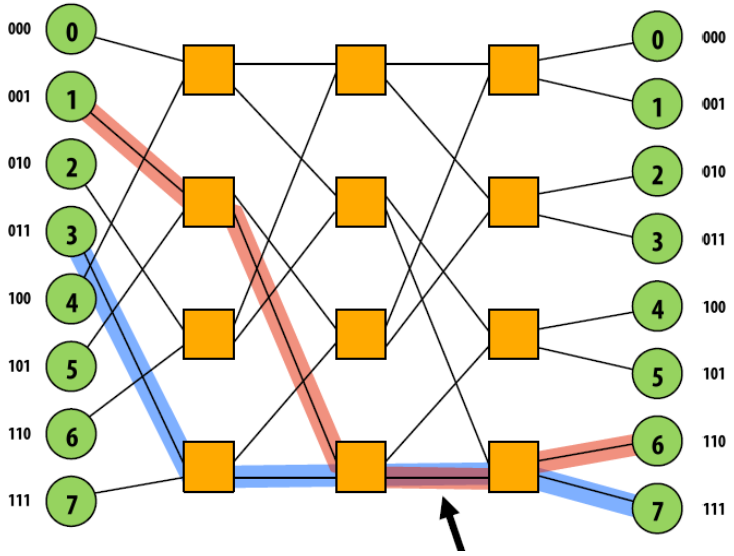




MIN – Višestepene sprežne mreže

- Blokirajuće mreže
  - obično jedan put od svakog ulaza do svakog izlaza
  - Npr., Omega (*perfect shuffle*, isti način povezivanja stepeni)
- Neblokirajuće mreže
  - Svaki ulaz može da se poveže na svaki slobodni izlaz
  - Dodatni stepeni i više puteva za par ulaz-izlaz (cena, latencija)





Vrste mreža

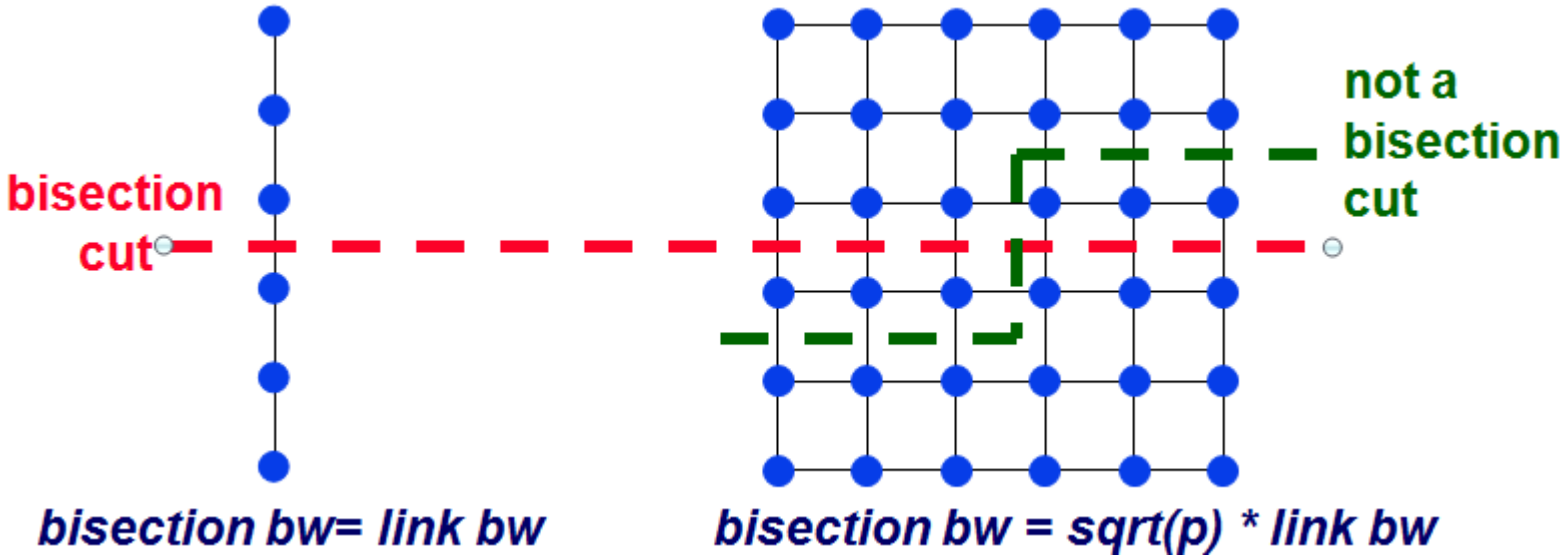
- Indirektne
  - …
- Direktne
  - Svaki čvor direktno vezan sa (obično manjim) brojem drugih čvorova (*point-to-point*)
  - Zasnovane na ruteru koji je deo čvora i povezan sa ruterima suseda kanalima
  - Statičke mreže – način vezivanja predefinisan
  - Bolje za komunikaciju suseda
  - OCN obično direktne

Parametri mreža

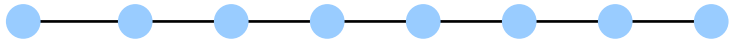
- Stepen čvora (d)
  - Broj linkova koji čvor vežu sa susedima
  - Poželjno da je konstantan (modularnost!) i manji (cena!)
  - Ako je konstantan, mreža je *regularna*
- Prečnik (D)
  - Maksimalno najkraće rastojanje između bilo koja dva čvora mereno brojem linkova
  - Poželjno da je što manji (latencija!)
- Simetrija
  - Mreža izgleda isto iz svakog čvora
  - Olakšava implementaciju
- Regularnost
  - Ako je stepen konstantan, mreža je *regularna*

Parametri mreža

- Propusni opseg bisekcije (*B*)
  - Bisekcija - mreža se preseče na dva jednaka dela
  - Propusnost preko najmanje bisekcije
  - Bitna za algoritme sa tipom komunikacije "*svi-sa-svima*"

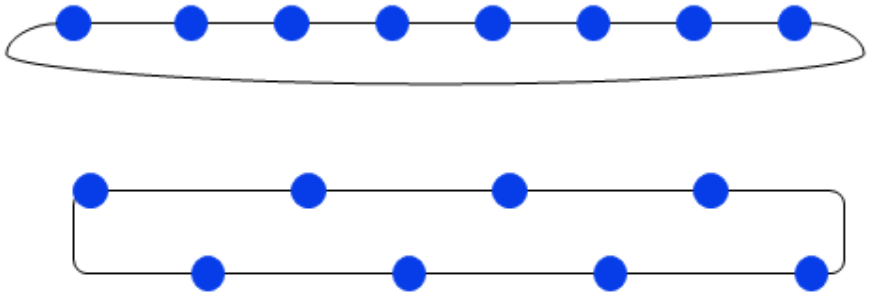


Linearni niz



- Osobine
  - $d = 2$
  - $D = n - 1$
  - $B = 1$
  - Prosečna distanca $\sim n/3$
  - Nesimetrična
  - 1D kao i magistrala,
    ali dozvoljava više istovremenih transfere

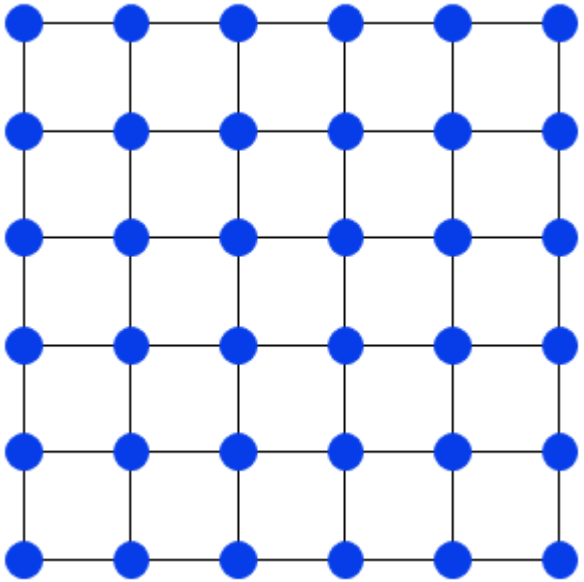
Prsten



- Osobine
  - $d = 2$
  - $D = n/2$
  - $B = 2$ (loše kad n raste)
  - Prosečna distanca $\sim n/4$
  - ... ali korišćenjem *pipeline* tehnike može biti brza
  - *Kordalni prstenovi* sa poprečnim vezama
  - Simetrična
  - Ima ga kod Intela (Corei7, Xeon Phi, Larrabee), IBM (Cell)

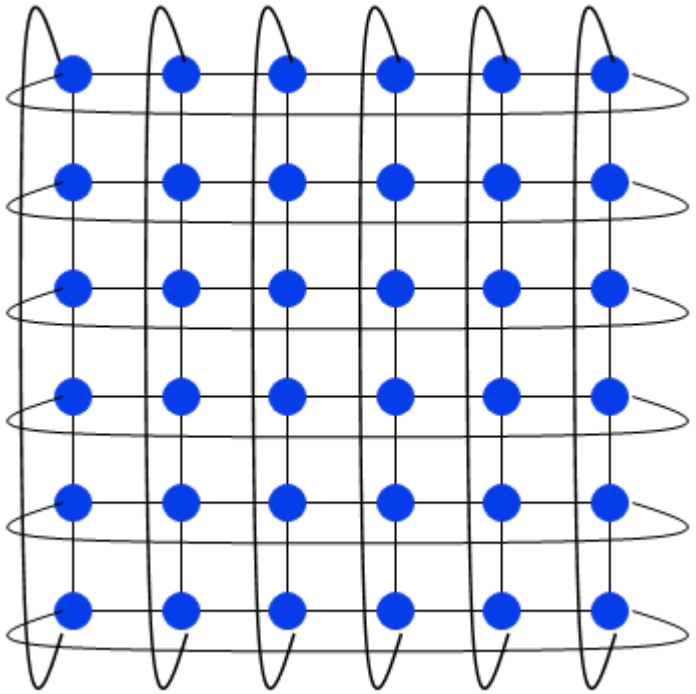
Mesh

- Osobine
  - $d = 4$ (ali i 3, 2)
  - $D = 2(\sqrt{n} - 1)$
  - $B = \sqrt{n}$
  - Nesimetrična
  - Pogodna za algoritme sa lokalnom komunikacijom tipa "*najbliži susedi*"
  - Dosta mogućih puteva
  - Lako izbegava *deadlock* (npr., prvo E-W, pa S-N)
  - Laka za implementaciju zbog regularnosti i kratkih veza
  - Ima je kod Tilera procesora i u nekim Intelovim čipovima

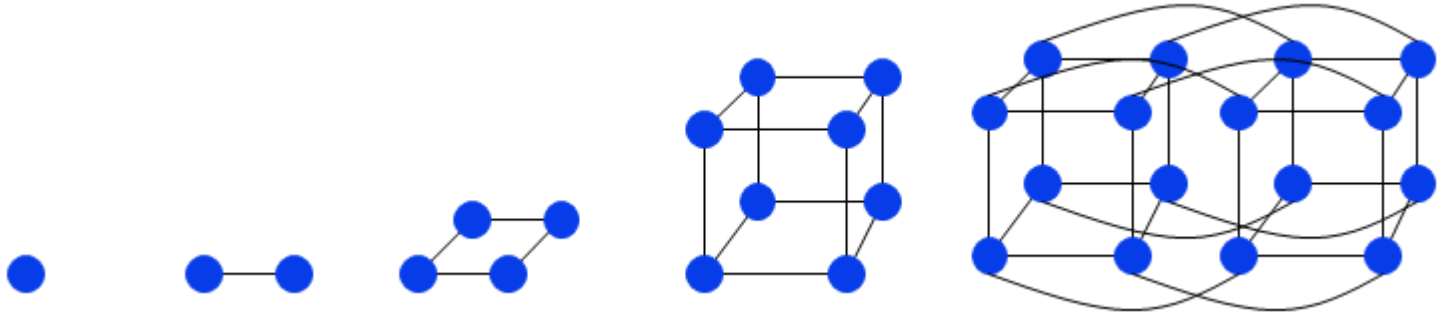


Torus

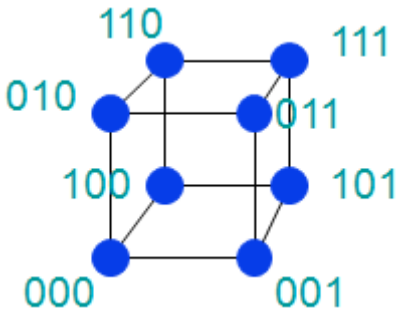
- Osobine
  - Okolne veze prave prstenove po svakoj dimenziji
  - $d = 4$ (za sve!)
  - $D = (\sqrt{n} - 1)$
  - $B = 2\sqrt{n}$
  - Simetrična
  - Pogodan za algoritme sa 2D nizovima
  - Teže za implementaciju na čipu, veze nejednake dužine
  - Cray XD koristi 3D torus



Hiperkocka

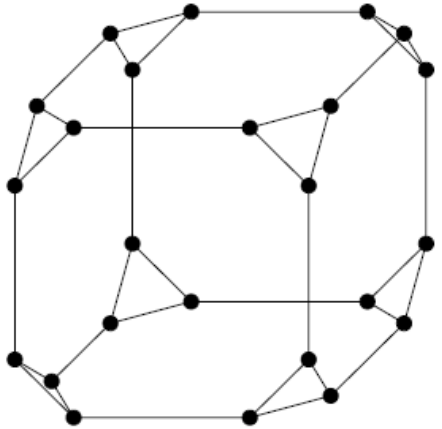


- Osobine
  - Binarna kocka dimenzije *d*
  - $n = 2^d$
  - $D = d$ - O(log n)
  - $B = n/2$
  - Teža za implementaciju za veće *d*, lošija skalabilnost
  - Lako rutiranje po bitovima, udaljenost jednaka broju različitih bitova
  - Intel iPSC, NCUBE, Caltech Cube

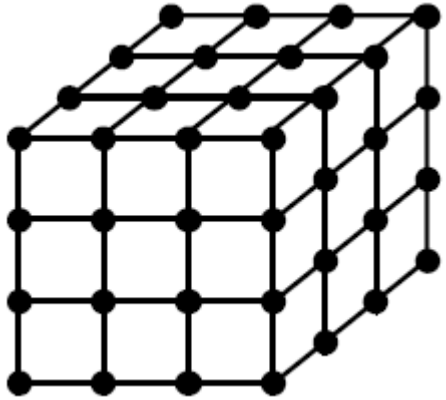


Hiperkocka

- CCC
  - Prstenovi u temenima
  - $n = d*2^d$
  - Rešava problem stepena čvora - 3
  - $D = 2d$

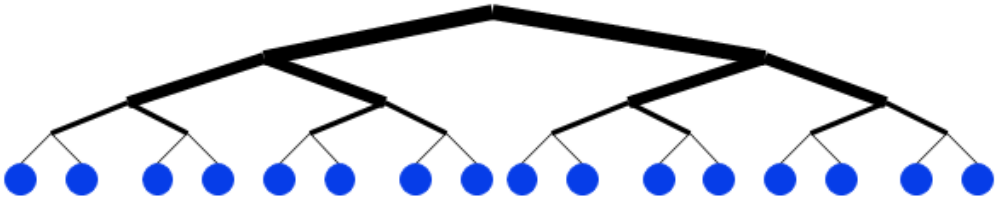


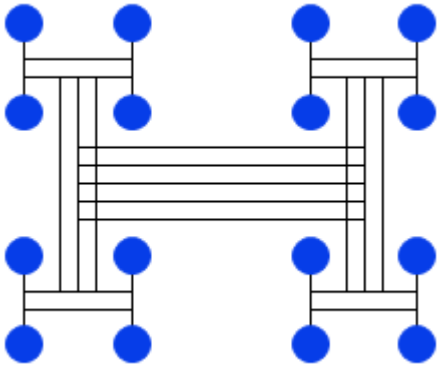
- *K*-arna *d*-kocka
  - Generalizacija hiperkocke sa *k* čvorova u jednoj dimenziji
  - $n = k^d$
  - Stepen čvora - $2d$
  - $D = dk/2$, $B = dk/4$
  - mesh ($d = 2$), prsten ($d = 1$)



k=4 *d*=3

Stabla





- Osobine
  - Planarna, hijerarhijska topologija
  - $n=2^k - 1$
  - $d = 3$
  - $D = 2(k\text{-}1)$ - O(log n)
  - Dobro za lokalizovan saobraćaj
  - Problem – usko grlo na višim nivoima
  - Rešenje - *fat tree* (npr., CM-5)

Topologije

older — newer

Cray XT3 and XT4	3D Torus (approx)
Blue Gene/L	3D Torus
SGI Altix	Fat tree
Cray X1	4D Hypercube*
Myricom (Millennium)	Arbitrary
Quadrics (in HP Alpha server clusters)	Fat tree
IBM SP	Fat tree (approx)
SGI Origin	Hypercube
Intel Paragon (old)	2D Mesh
BBN Butterfly (really old)	Butterfly

Ponekad hibridi i aproksimacije