

ARHITEKTURA RAČUNALA 2

Nositelji:

Slobodan Ribarić

Siniša Šegvić

Tomislav Hrkać

Asistenti:

Petra Bevandić

Josip Šarić

Iva Sović

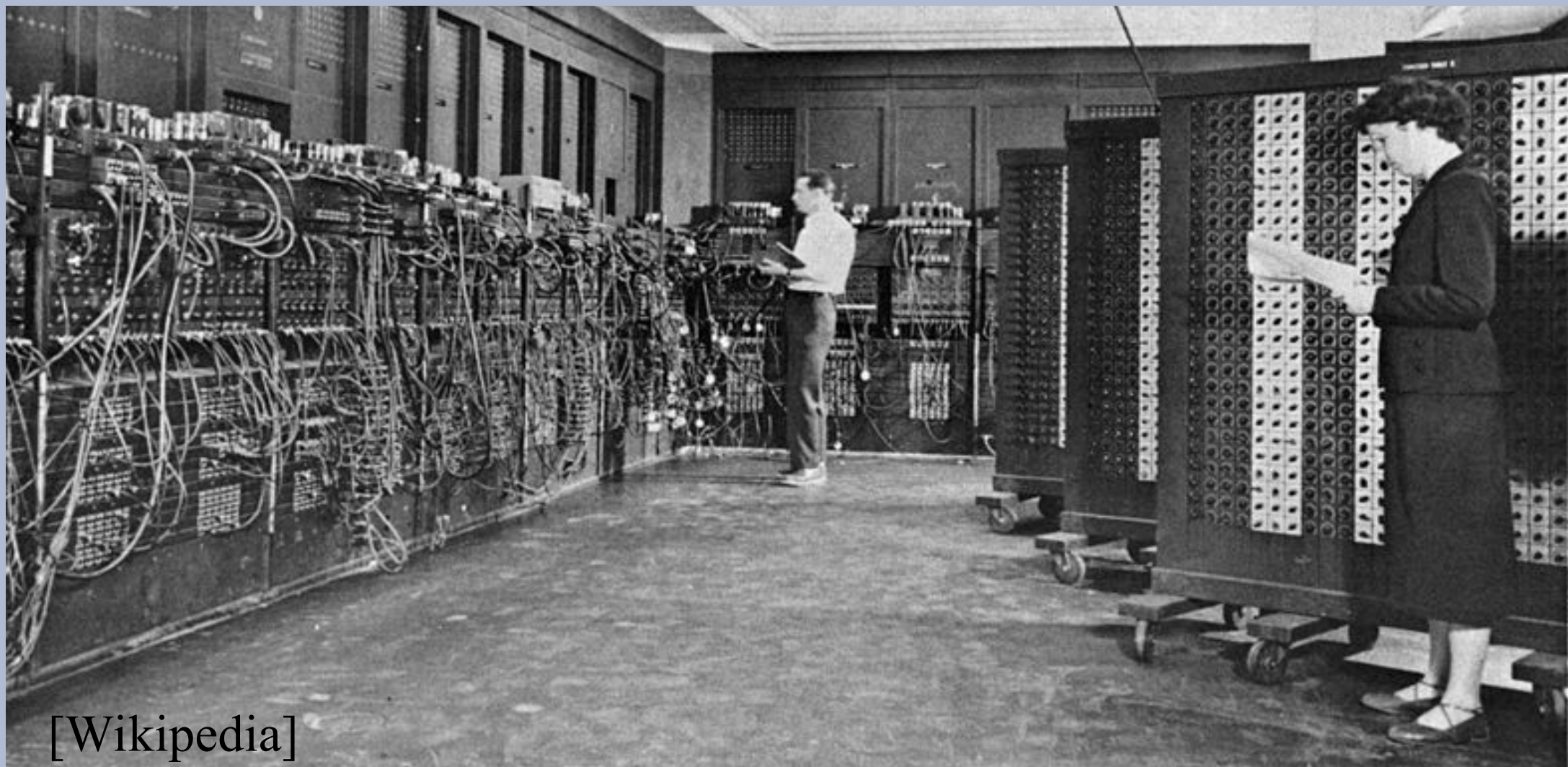
Zavod za elektroniku, mikroelektroniku, računalne
i inteligentne sustave (ZEMRIS), 3. kat D zgrade

Uvodno predavanje

- **Sadržaj:** što ćemo proučavati?
- **Motivacija:** zašto je to korisno znati?
- **Kontekst:** aktualni tehnološki trendovi
- **O predmetu:** način održavanja nastave, razdioba bodova, literatura, teme

Zašto koristimo računala?

Računalni sustav nekad (Eniac, 1943-1946, \$5e5, 2e4 cijevi):

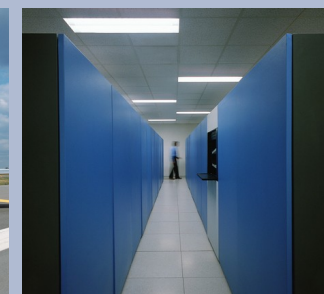
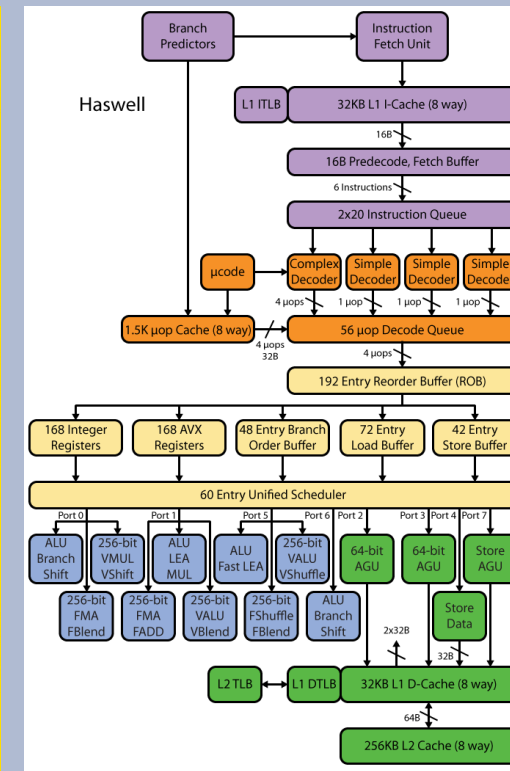
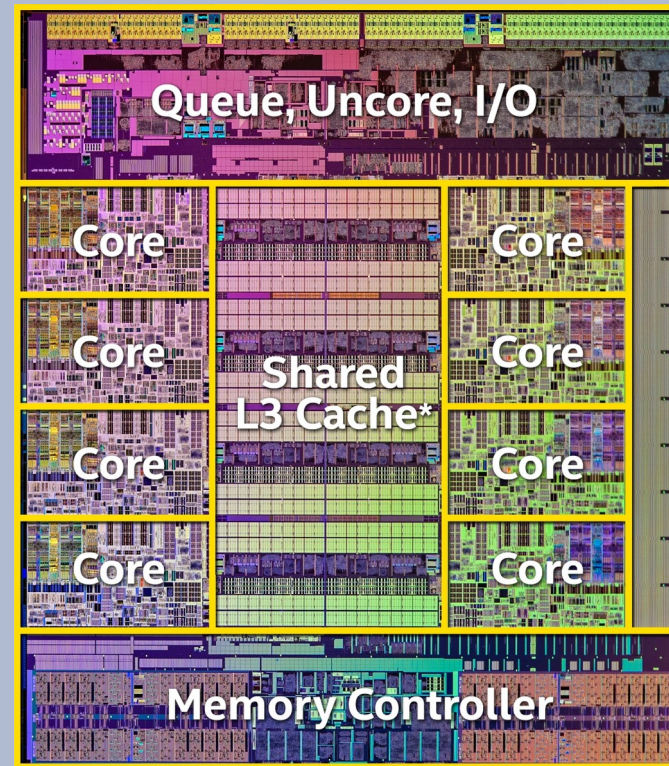
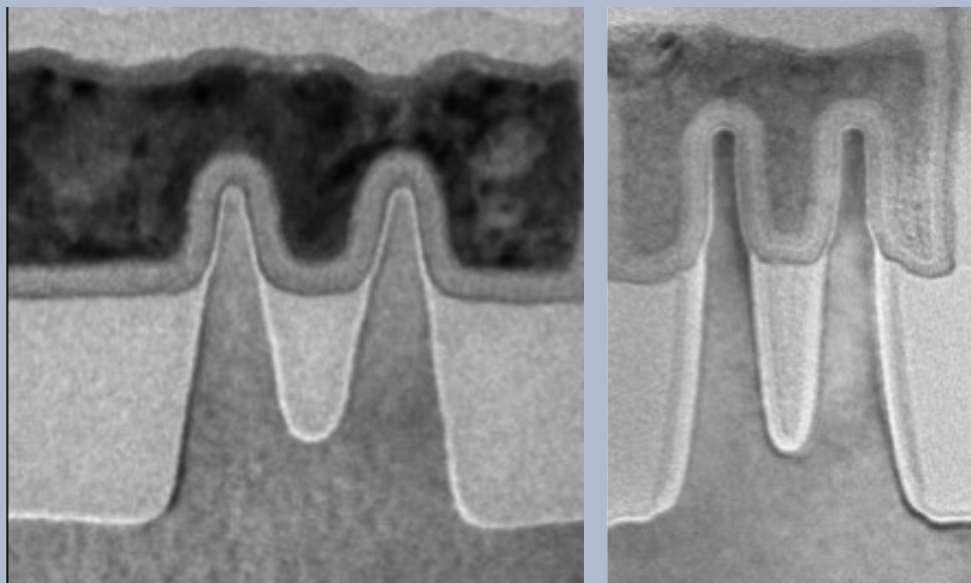


[Wikipedia]

Zašto koristimo računala? (2)

Moderno računalo: i7-5960x, 2014:

\$1000, 8x4 instrukcija/takt,
3.5cm², 2.6e9 tranzistora



Što je arhitektura računala?

- Veza između korisnika i **logičkih komponenata** (bistabila, zbrajala, ...) grozno komplicirana
- Moderni sustavi tu vezu izvode u više **slojeva**
- **Arhitektura računala** (IBM 360, 1964):
 - proučava hijerarhijsko organiziranje sklopovlja za **obradu podataka**
 - **cilj**: postići **optimalna** svojstva primjenom **dostupnih** tehnologija
 - performansa (brzina izvođenja, fokus na višim programskim jezicima)
 - cijena
 - utrošak energije
 - pouzdanost, raspoloživost, ...
- Arhitektura računala se prilagođava svojstvima elektroničkih sklopova
 - ponekad ćemo se morati podsjetiti kako rade tranzistori!

Razine apstrakcije u računalnim sustavima

Fokus kolegija!!

Aplikacijski program

Programski jezik

Operacijski sustav

Instrukcijski skup

Mikroarhitektura

(organizacija računala)

Oblikovanje podsustava

(sabitnice, memorija, registri, ...)

Sklopovi

Elementi

Fizika materijala



arhitektura
računala



fizičko
oblikovanje

Područja glavnog interesa kolegija

- Specifičnosti računala opće namjene
 - poslužitelji, radne stanice, prijenosnici
 - uključujući GPU-ove, moderne telefone i tablete
 - ARM Cortex-A76: 8x4x2 množenja/takt
 - Tesla A100: 6912 množenja/takt
 - velika većina izvornog koda pisana u višem programskom jeziku
- Detalji organizacije
 - organizacija (arhitektura) procesora
 - memorijska organizacija
- Iskorištavanje paralelizma
 - instrukcijski paralelizam (ILP)
 - vektorske instrukcije (SIMD)
 - višeprocorska računala (MIMD)

Motivacija: zašto bi se studenti računarstva trebali upoznati s konceptima arhitekture računala?

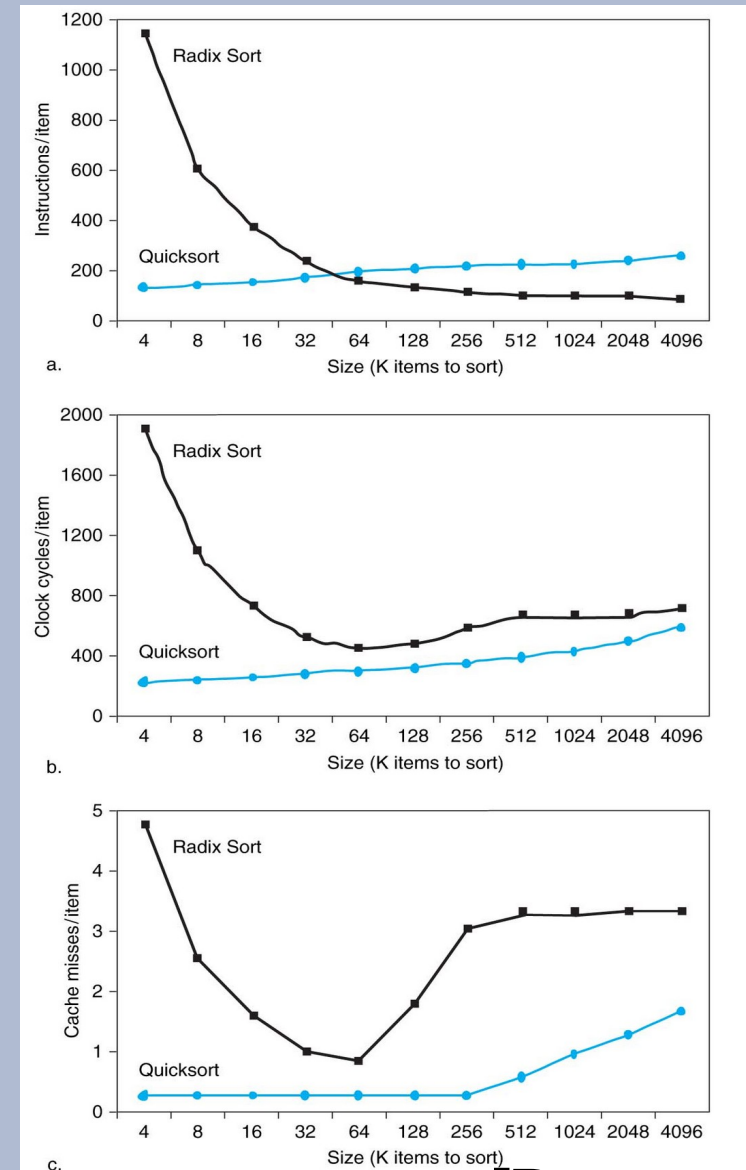
- Steći **osjećaj** što definira svojstva računala
 - moći odabrati optimalnu konfiguraciju s obzirom na zahtjeve (performansa, cijena, potrošnja, dostupnost, pouzdanost)
- Biti sposoban optimirati program na **računalu opće namjene**
 - tržište stalno traži sve sofisticiraniju funkcionalnost
 - npr. u našem laboratoriju izvođenje programa često traje po tjedan dana
 - razumijevanje organizacije računala kritično za optimiranje performanse
 - optimalnost programa postaje važnije zbog usporenog porasta brzine integriranih sklopova
- Projektiranje složenih digitalnih sustava kao **profesija** (mnogi FER-ovci sudjelovali u stvaranju vrhunskog sklopovlja)

Što ćete na ovom kolegiju naučiti?

1. razlikovati uloge glavnih komponenata računala uključujući procesor, memoriju, sabirnice i ulazno-izlazne uređaje
2. predvidjeti sabirničku aktivnost jednostavnog procesora tijekom izvođenja zadanog odsječka strojnog kôda
3. pokazati izvedbu jednostavnih strojnih instrukcija na razini logičkih vrata
4. razumjeti načela instrukcijskih arhitektura RISC i x86
5. riješiti manje programske zadatke dopunjavajući C assemblerom
6. razumjeti funkciju i organizaciju priručnih memorija
7. razumjeti organizaciju procesora s dinamičkim raspoređivanjem
8. ilustrirati korake pri generaciji fizičke adrese u prisutnosti priručne i virtualne memorije
9. razumjeti implementaciju grubo zrnatog paralelizma na višejezgrenim i višeprocorskim računalima

Motivacijski primjer (1)

- Quicksort, $O(n \log(n))$, vs Radix sort, $O(k n)$
- teorijska složenost (gore) i vrijeme izvođenja (sredina)
- složeniji algoritam brži zbog veće lokalnosti radnog skupa (dolje: učestalost promašaja PM)!
- lijepa ilustracija razlike između matematike i računarstva

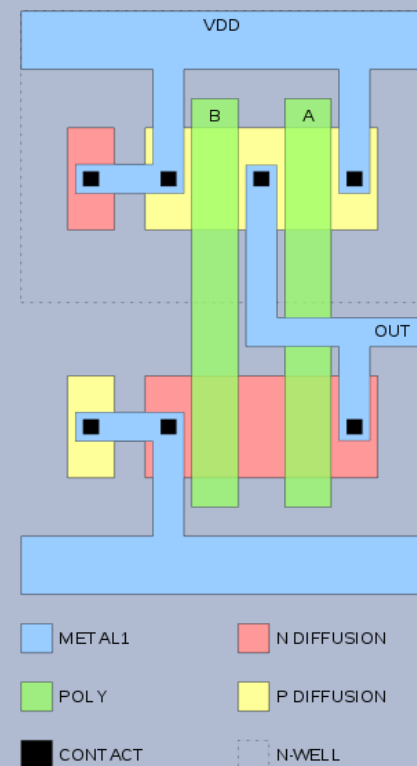
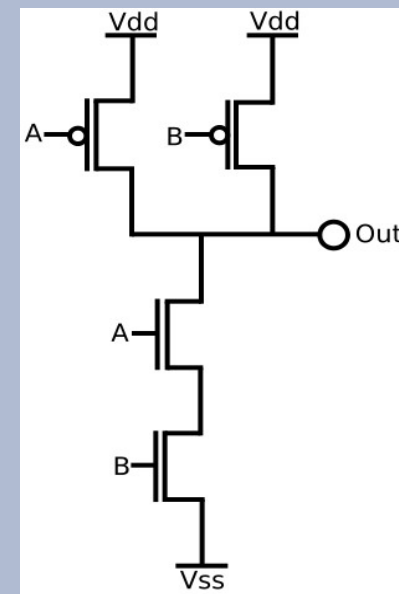
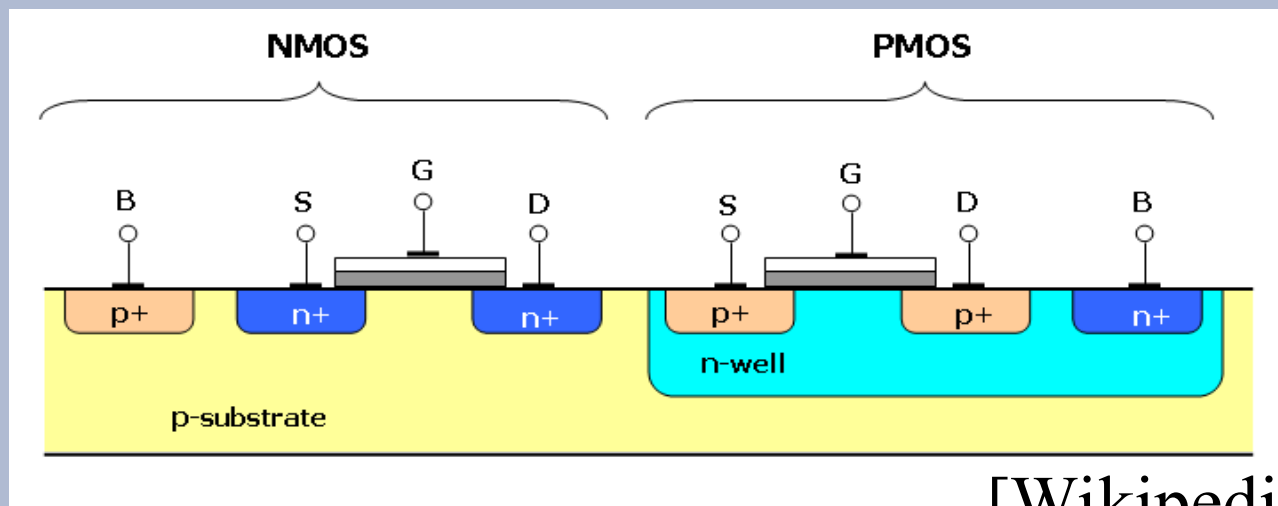


Motivacijski primjer (2)

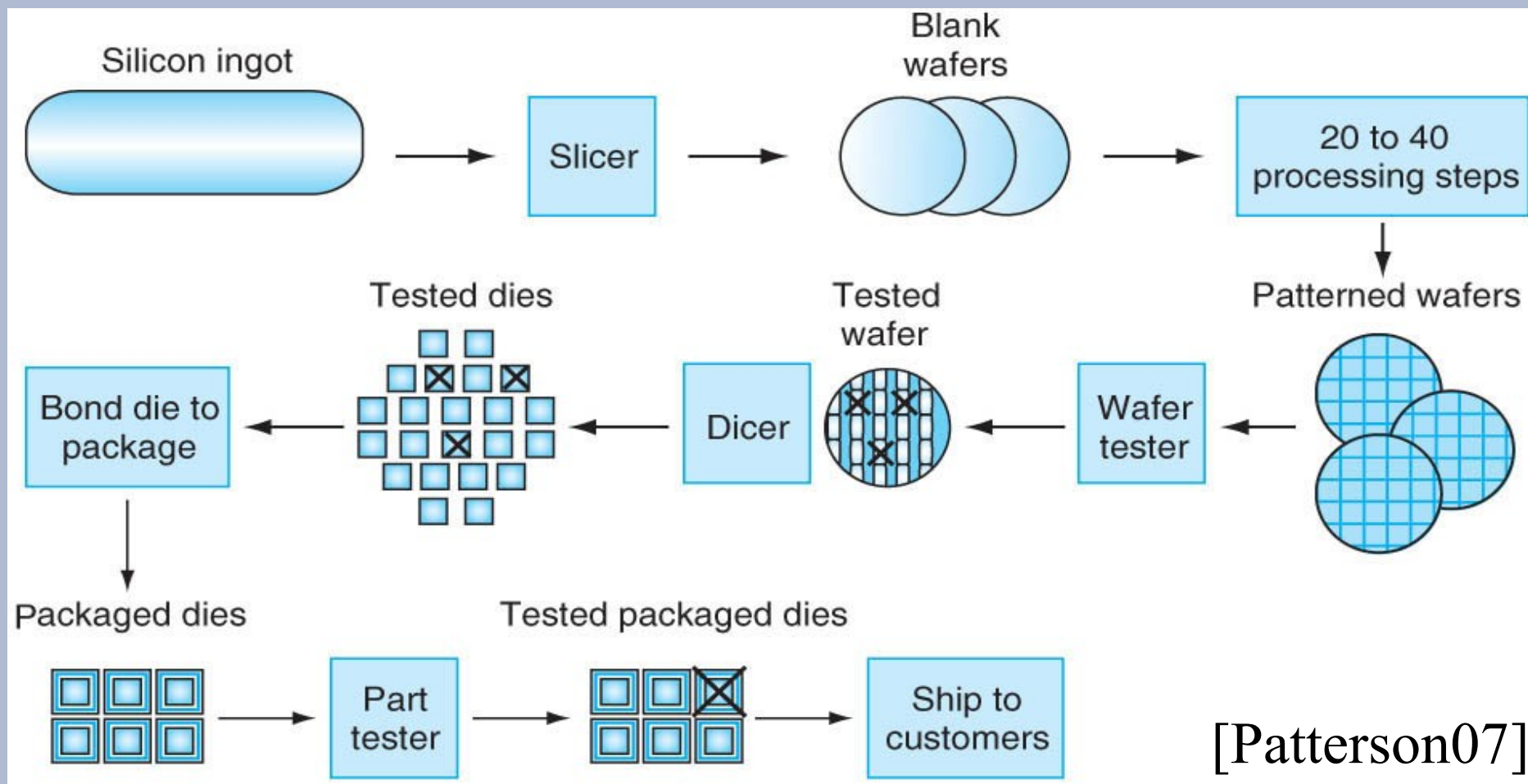
- ARM Cortex A9 vs Sandy Bridge i7 na testu SPEC INT 2011:
 - brzina izvođenja: i7 8 puta brži
 - broj ciklusa (brzina na istom taktu): i7 2.5 puta brži
- Koji je razlog?
 - utjecaj instrukcijske arhitekture zanemariv
 - i7 dinamički prevodi x86 instrukcije u RISC-like
 - bolje predviđanje grananja
 - manje promašaja priručne memorije
 - bogatiji instrukcijski skup
 - bolje maskiranje promašaja priručne memorije
 - uspješnije izvođenje izvan redoslijeda
- [blem13hpca] Power Struggles: Revisiting the RISC vs. CISC Debate on Contemporary ARM and x86 Architectures. Blem et al. HPCA 2013.

Tehnološki kontekst

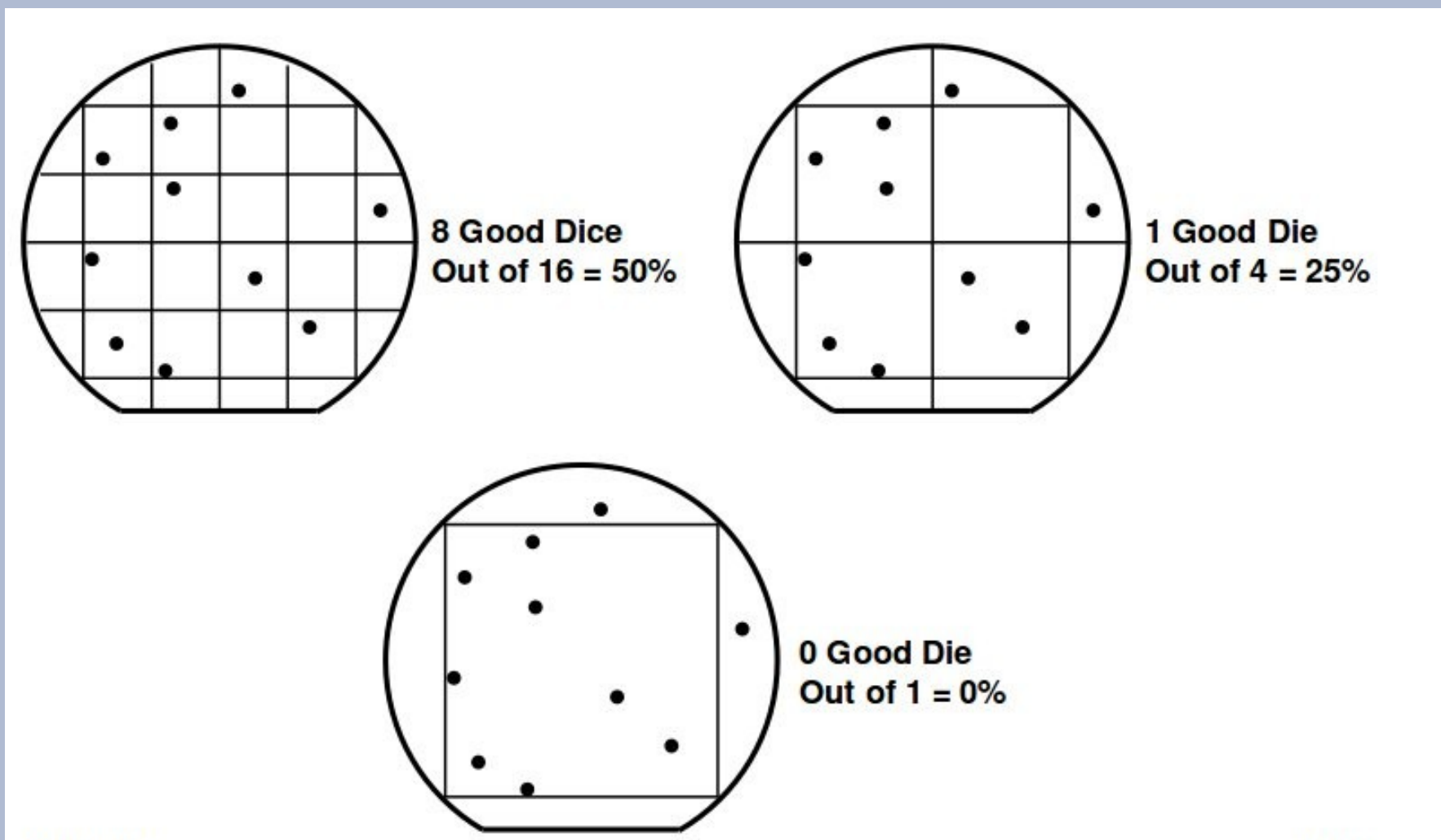
- računala se izvode na površini silicija (dominantno CMOS)
- logika se ostvaruje **upravljivim sklopkama** (nMOS, pMOS tranzistorii)
- brzina računala ovisi o trajanju prijelazne pojave pri prebacivanju sklopke (npr. Out: 0 → 1)
 - izlazna elektroda ima kapacitet → punjenje RC člana
 - komunikacija preko vanjskih izvoda **spora** → trebamo što veću **površinu** integriranog sklopa
 - komunikacija to **brža** što su sklopovi manji i bliži → trebamo što **sitnije** elemente



Proizvodni proces (1)

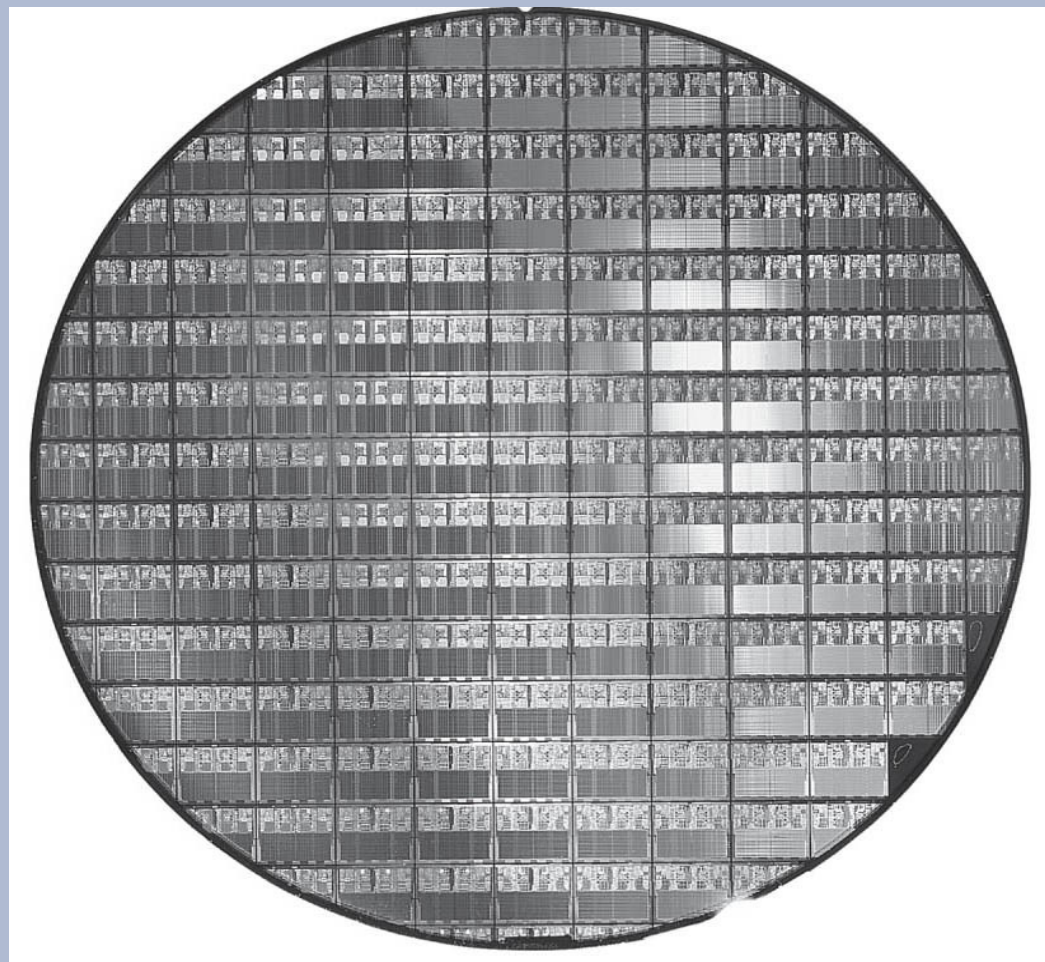


Utjecaj veličine sklopa na prinos (die size vs yield)



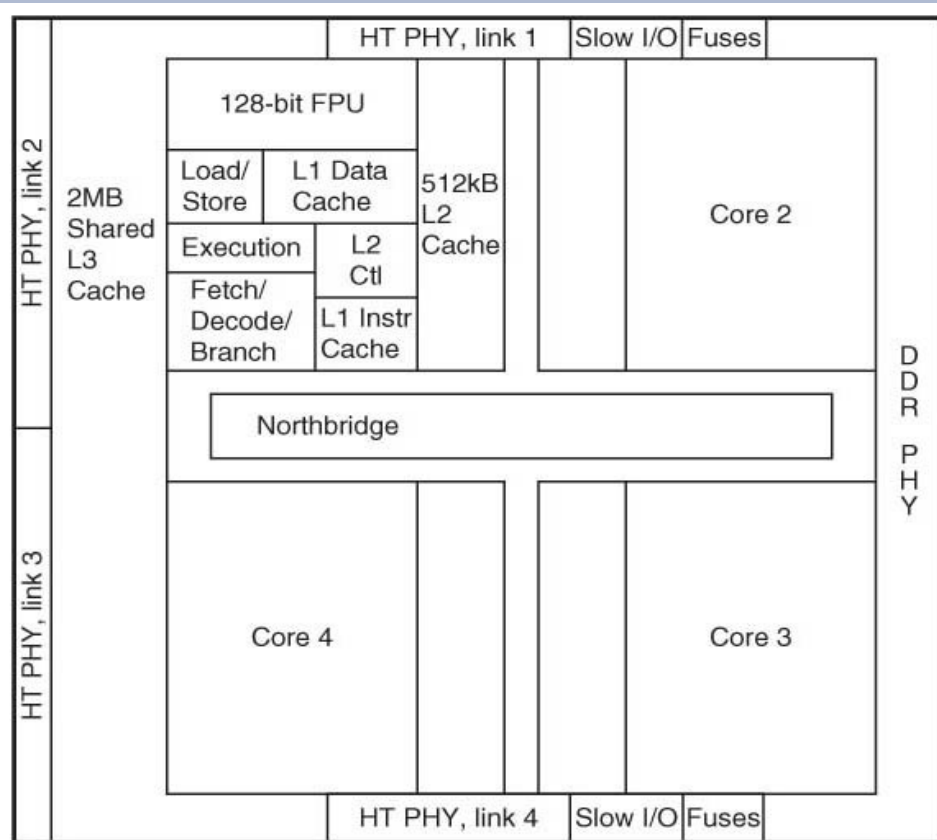
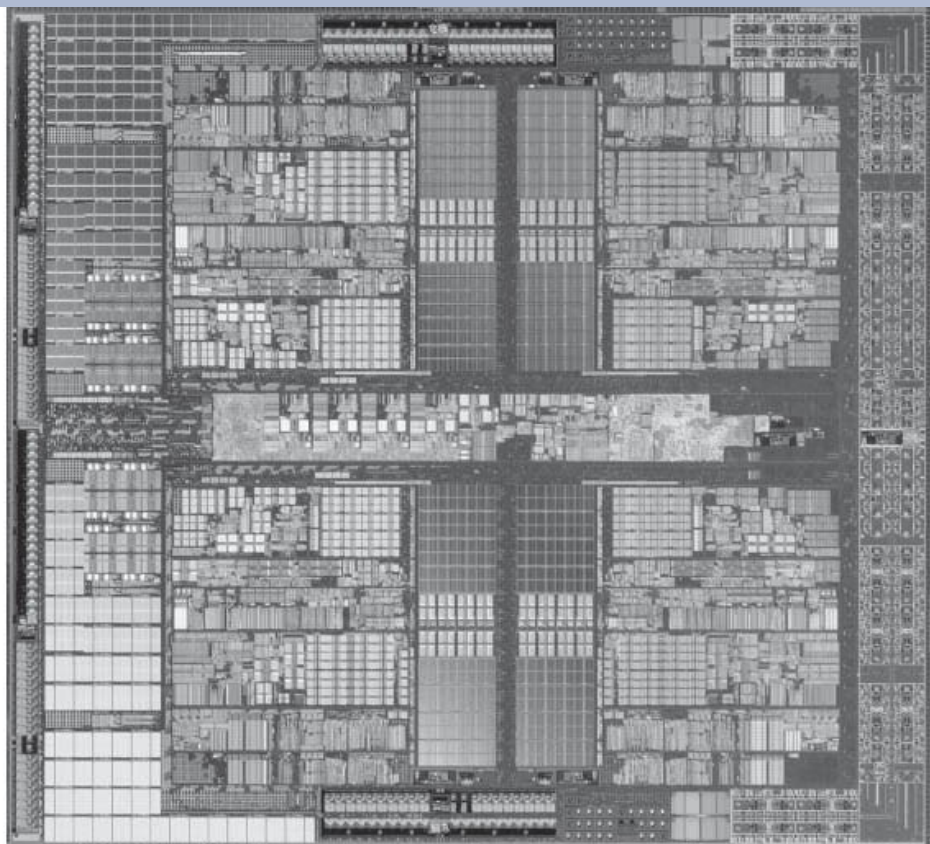
Proizvodni proces (2)

- silicijski wafer $d=20\text{cm}$
- uz 100% prinos \rightarrow 117 čipova AMD Opteron X2
- konkretni proces u praksi postiže prinos od oko 85%
- prinos utječe na cijenu i tako ograničava veličinu integriranog sklopa
- \rightarrow površina čipa je **kompromis** između performanse i cijene!



[Patterson07]

Proizvodni proces (3)

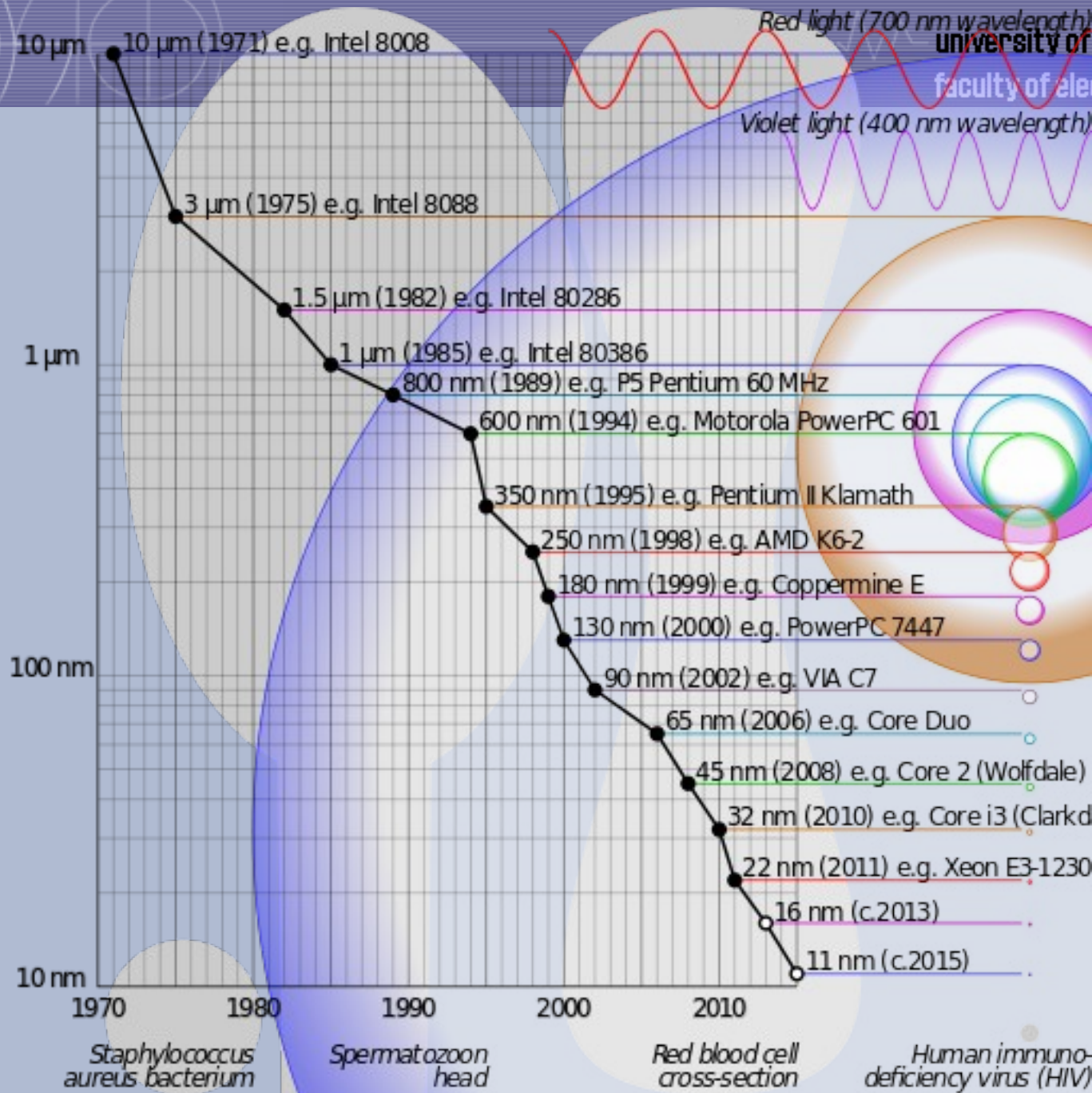


[Patterson07]

- AMD Opteron Barcelona X4, fizički raspored, 285 mm²

Tehnološki trendovi (1970-2000):

- Gustoća tranzistora: 35% godišnje
- Površina sklopa: 10%-20% godišnje
- Broj tranzistora na sklopu: 40%-55%
Moore: udvostručenje svake 2 godine
- Brzina tranzistora raste s korijenom gustoće (linearno s rezolucijom tehnologije)
- Gustoća energije ostaje konstantna (Dennardovo skaliranje)
- Kašnjenje signalnih linija slabo pada: važna lokalnost obrade u digitalnom sklopu
- “Višak” tranzistora stvara prilike za arhitekta: protočnost, cachevi, predviđanje grananja, šire sabirnice (8, 14, 32, 64), više jezgri, ...



[Wikipedia]

Napredak gustoće poluvodičke tehnologije

Number of transistors on an integrated circuit

1,000,000,000

100,000,000

10,000,000

1,000,000

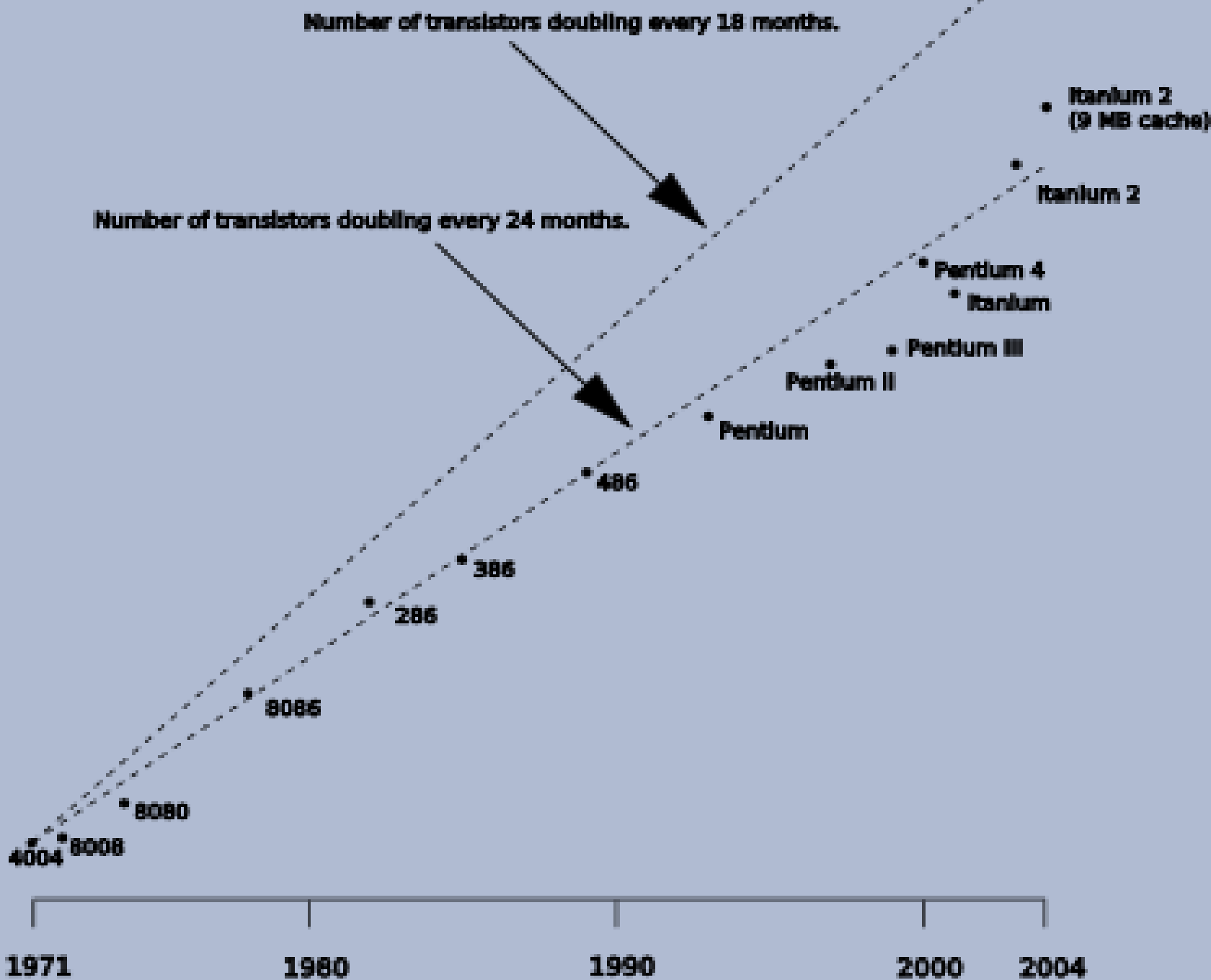
100,000

10,000

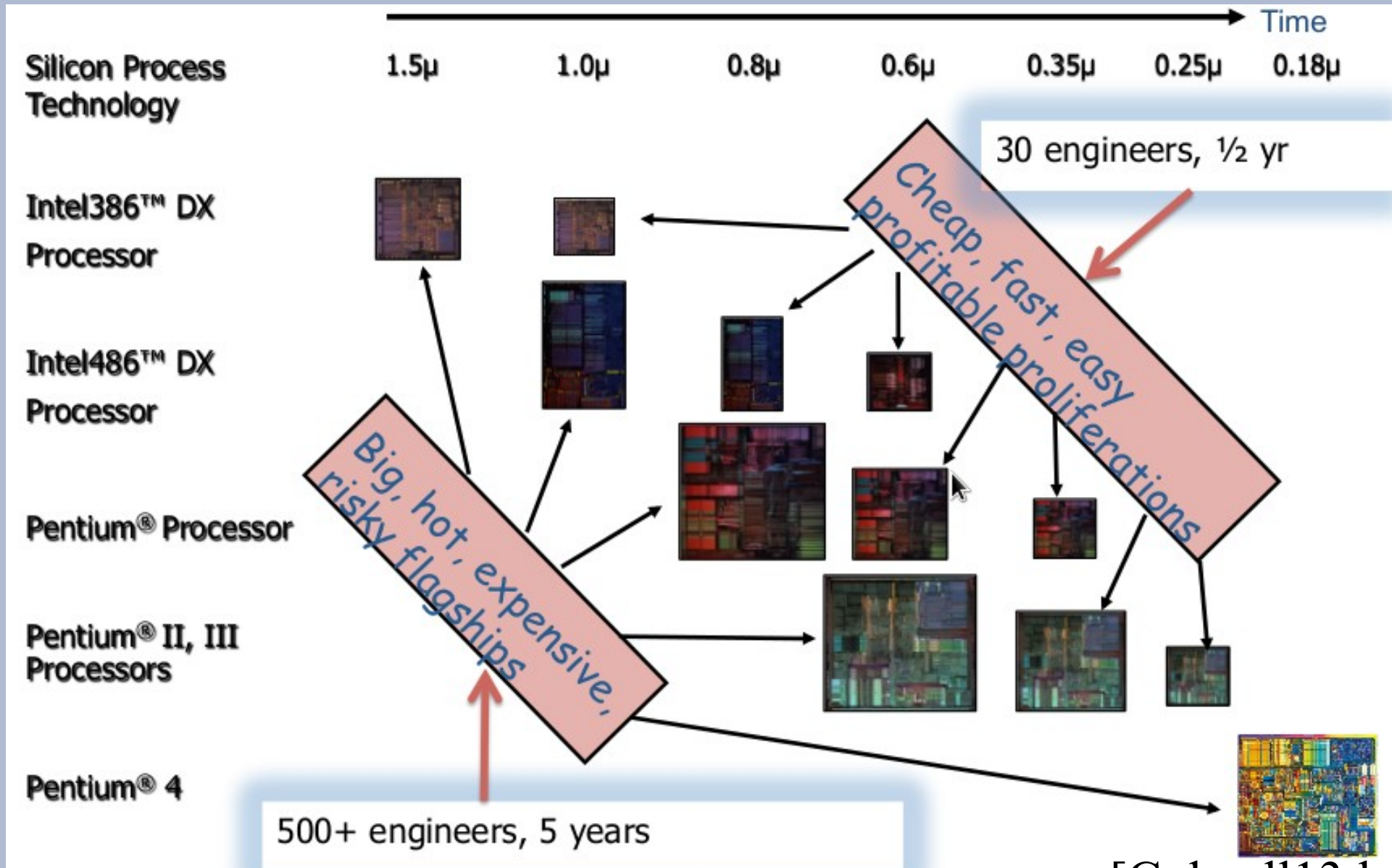
2,300

Number of transistors doubling every 18 months.

Number of transistors doubling every 24 months.



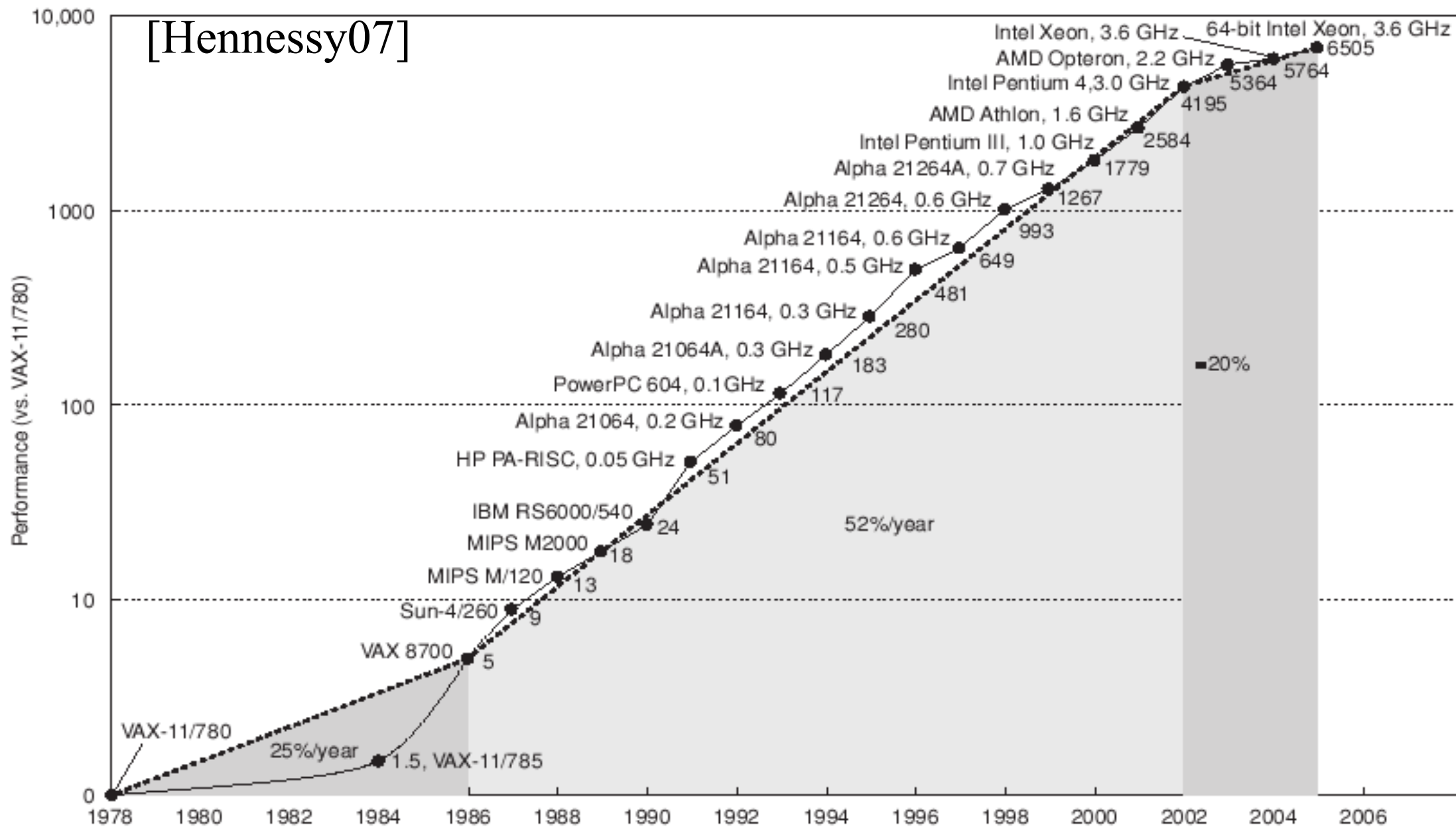
Intelova proizvodna strategija (tick-tock)



Performansa uniprocссора

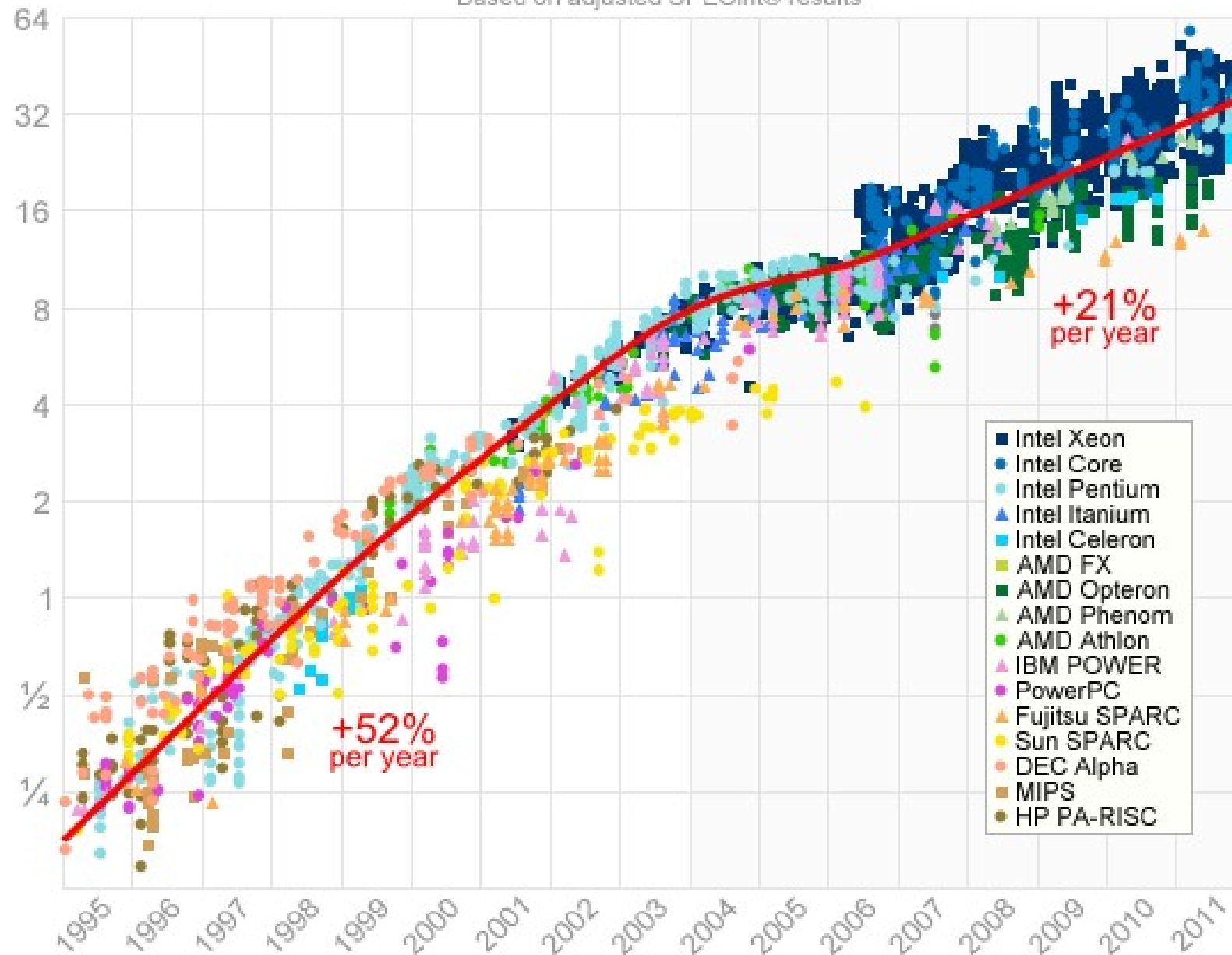
- zahvaljujući povoljnom tehnološkom okruženju, 1986-2002 zabilježen rast performanse od oko 50% godišnje
 - povećanje broja tranzistora na sklopu
 - povećanje radnog takta
- dostupnost performanse potiče stvaranje novih aplikacija (npr. autonomni automobil!)
 - a nove aplikacije traže još performanse
 - performanse neće nikad biti dovoljno!
- međutim, u 21 stoljeću imamo porast performanse od samo 20% godišnje, iako se Mooreov zakon još drži
- što se dogodilo?

[Hennessy07]



Single-Threaded Integer Performance

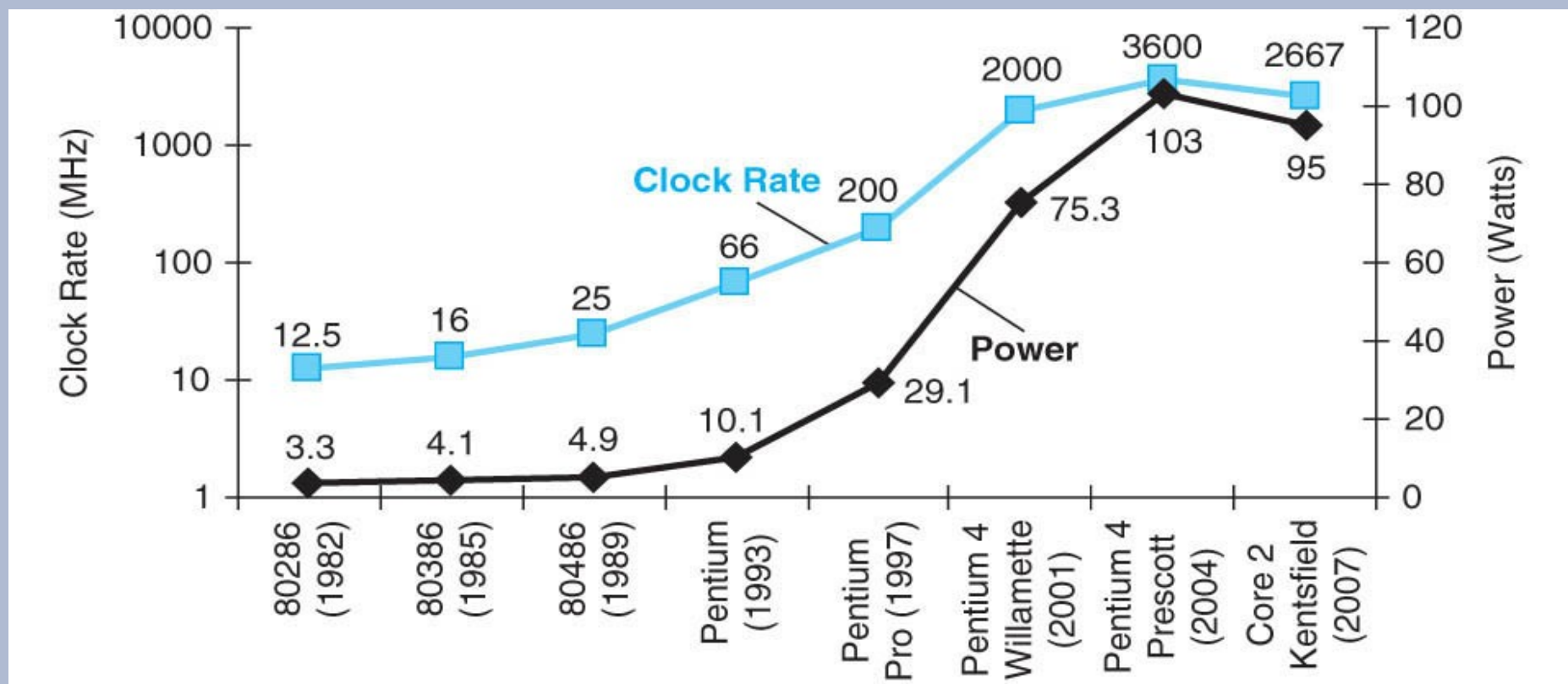
Based on adjusted SPECint® results



Aktualni tehnološki trend (2000-2030):

- Tehnologija je dosegla graničnu gustoću disipacije
 - Dennardov zakon više ne vrijedi: parazitske struje ne dozvoljavaju da smanjivanje tranzistora pratimo smanjivanjem napona (stali smo na 1V)
 - **kraj porasta radnog takta** (veći takt \Rightarrow veća disipacija, 0.1W \rightarrow 135W)
 - **"Power wall"**
- Procesna moć i propusnost rastu (puno) brže od latencije
 - pristup glavnoj memoriji 25 \times sporiji od FP množenja
 - **"Memory wall"**
- Usložnjavanje arhitekture donosi sve manje koristi
 - instrukcijski paralelizam (ILP) donosi sve manje prinose
 - protočnost, superskalarnost, izvođenje izvan redosljeda ...
 - tranzistora imamo više nego što ih možemo iskoristiti u uniprocessoru!
 - **"ILP wall"**
 - i dalje svake godine imamo 40% tranzistora nego lani, ali višak tranzistora sve je teže pretočiti u performansu

Kretanje disipacije procesora



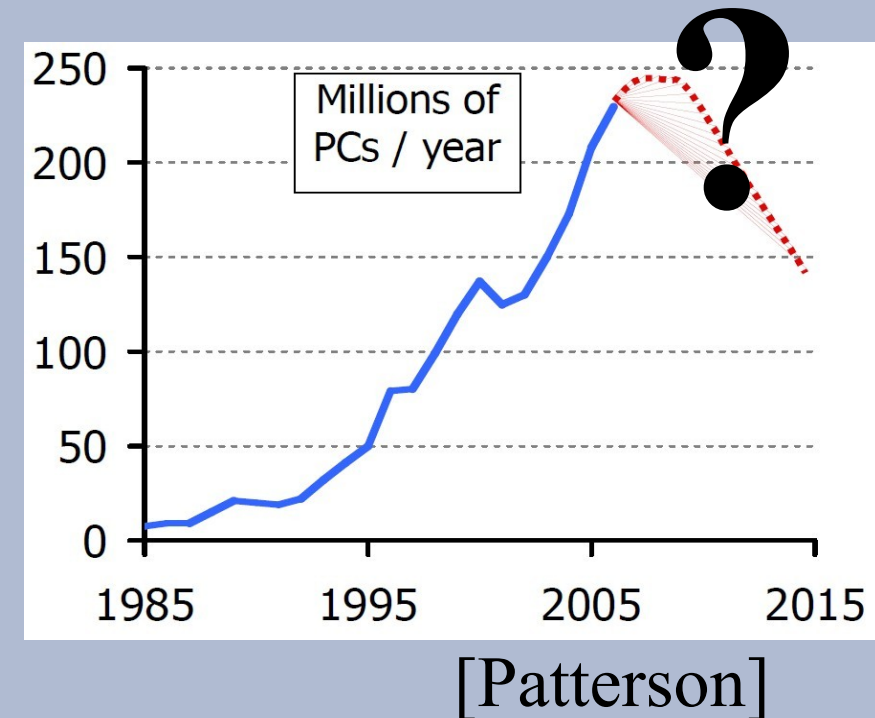
[Patterson07]

Zašto performansa slabije raste?

- Promjena tehnološkog konteksta!
 - disipacija ograničava gustoću elemenata
 - instrukcijski paralelizam iscrpljen
 - sporo poboljšanje latencije memorije
 - power wall + ILP wall + memory wall = **brick wall**
- Kraj ere uniprocссора!
 - nitko nije proizveo uniprocссора opće namjene već godinama
- Inovacije: SIMD, SMP, **multicore**, MPP, **GPGPU**, ...

Aktualni tehnološki kontekst

- svijet se mijenja: industrija navikla na ogromne prihode u velikom je strahu!
 - crni scenario (slika!): što ako ljudi počnu kupovati računala tek kad se pokvare?
 - sve nade usmjerene prema paralelizmu (teži problem od svih dosadašnjih!)
- uspješnost tvrtki koje su se prije 2005 drznule proizvoditi paralelna računala: **0%!**



- ambiciozni cilj industrije i istraživača:
 - odustati od uniprocessorske performanse (!)
 - rast postići udvostručenjem jezgara
 - stvoriti povoljno ozračje prikladnim inovacijama (podrška za prirodno izražavanje paralelnih programa, ...)

Informacije o predmetu

- Službene stranice predmeta:
 - obavijesti, forum,
 - raspored i zadatci za laboratorij,
 - repozitorij datoteka
 - <http://www.fer.hr/predmet/arhrac2>
- Neslužbene stranice predmeta:
 - studentski projekti
 - zanimljivi članci
 - <http://www.zemris.fer.hr/~ssegvic/ar2/>

Kakvo predznanje očkujemo?

- Digitalna logika
- Arhitektura računala 1
- Programiranje i programsko inženjerstvo
- (Elektronika)

Predavanja

- Prema rasporedu
- ~~Povremene kratke provjere znanja~~
- Materijali dostupni na stranicama predmeta nakon predavanja
- Nazočnost ne provjeravamo

Laboratorij

- Tijekom semestra održavaju se četiri grupe laboratorijskih vježbi
 - mikroprogramiranje,
 - programiranje u zbirnom jeziku IA32 (x86),
 - utjecaj arhitektonskih značajki na izvođenje programa u C-u
 - programiranje masivno paralelnih računala (GPU)
- Vježbe se pripremaju kod kuće te predaju u Ferku
 - rok za predaju je dan prije izvođenja vježbi
 - temeljem programa napisanih kod kuće ne dobivaju se bodovi
 - za pristup laboratoriju potrebno je predati barem 50% vježbe
 - moći će se nadoknaditi jedna vježba
 - uz ispričnicu ili najavu barem tjedan dana prije vježbe
 - termin će biti pravovremeno oglašen krajem semestra

Laboratorij

- Protokol provjere laboratorijskih vježbi:
 - rješavanje problemskog zadatka na računalu ili na papiru
(1/2 bodova)
(upotreba mobitela i interneta je na žalost zabranjena)
 - rješavanje blica
(10-ak zadataka, abcd pitalice, 1/2 bodova)
- Zadatci će biti zamrznuti tri tjedna prije termina odgovarajućih vježbi

Razdioba bodova

- ~~Periodičke provjere znanja~~
- Izlazni testovi laboratorijskih vježbi
 - **20** bodova
- Međuispit i završni ispit
 - **40** i **40** bodova
 - **40%** teorija (abcd), **60%** problemski zadatci
- Projektni zadatci (bonus)

Detalji

- Ponovljeni međuispit: **nema**
- Ponovljeni završni ispit: **nema**
- Preduvjeti za izlazak na završni ispit: **nema** 40% laboratorija
- Klasični ispitni rok
 - ◊ preduvjet za izlazak je 40% "laboratorijskih" bodova
 - ◊ bodovi s kontinuiranih provjera ne uzimaju se u obzir
- Usmeni dio ispita
 - ◊ prag za izlazak je 50% bodova do tada
 - ◊ kontinuirana provjera: mi + zi + lab
 - klasični ispit: pismeni ispit
- Pragovi ocjena su: 50 (2), 63 (3), 76 (4), 89 (5)

Literatura

- S. Ribarić, Zbirka riješenih zadataka iz Građe računala, arhitekture i organizacije računarskih sustava, Merkur A.B.D, Zagreb, 2017.
- S. Ribarić, *Građa računala - arhitektura i organizacija računarskih sustava*, Algebra, Zagreb, 2011
- S. Ribarić, *Arhitektura računala RISC i CISC*, Školska knjiga, Zagreb, 1996.
- J. L. Hennessy, D. A. Patterson, *Computer Architecture, A Quantitative Approach*, Morgan Kaufmann, 6th ed, 2019
- D. A. Patterson, J. L. Hennessy, *Computer Organization & Design, The Hardware/Software Interface*, Morgan Kaufmann, 4th ed, 2009.