

Performanse racunara

- Brzina mikroprocesora
- Ravnoteza performanse

Performanse racunara

- Cena racunara pada, a performanse rastu velikom brzinom iz godine u godinu.
- Procesori su jeftini da mogu da se bacaju umesto da se zamene, bezicni senzori, uredjaji za jednokratnu upotrebu...
- Performanse omogucavaju razvoj aplikacija velike slozenosti.

Performanse racunara

- Video konferencije
- Obrada slike
- Multimedija
- Modeliranje I simulacije
- Serveri ...

Performanse racunara

- Sto se tice arhitekture I organizacije racunara, gradivni blokovi su gotovo isti kao sto su bili kod racunara od pre 50 godina!

-

Performanse racunara

- Tehnike koje se koriste za postizanje maksimalnih performansi postale su izuzetno sofisticirane i stalno se unapredjuju.

Brzina mikroprocesora

- Razvoj procesora sledi Moorov zakon tako da se lansiraju nove generacije cipova svake tri godine sa cetiri puta vecim brojem tranzistora.
- Kod memorija, svake tri godine se duplira kapacitet dinamičkih memorija sa direktnim pristupom(DRAM), koja predstavlja osnovnu tehnologiju za glavnu memoriju racunara.

Brzine mikroprocesora

- Brzina mikroprocesora nije ključni faktor kod postizanja maksimalnih performansi.
- Ukoliko procesor veći deo vremena “spava” ili čeka podatke, bez obzira na brzinu, performanse kompletnog sistema su na nizem nivou.
- Procesor mora konstantno da se hrani podacima, instrukcijama da bi njegova brzina dosla do izražaja.

Brzine mikroprocesora

- Proizvodjaci cipova razvijaju tehnologiju da proizvedu cipove sto vece gustine.
- Zadatak projektanata procesora je da pronadju mogucnost da sto je moguće vise “zaposle” procesor.

Tehnike za obradu podataka ugradjene u savremene procesore

- Protocna obrada
- Predvidjanje grananja
- Analiza toka podataka
- Spekulativno izvrsavnje

Protocna obrada

- Kod protodne obrade procesor moze istovremeno da radi na vise instrukcija.
- Vremenski se preklapaju operacije tako da se svi stepeni protoka obraduju istovremeno.
- Jedna instrukcija se izvrsava, druga se dekodira!

Predvidjanje grananja

- Na osnovu koda instrukcije procesor predvidja koje grupe instrukcija ce sa najvecom verovatnocom da budu sledece za obradu.
- Na taj nacin moze unapred da uzme predvidjene, moguće sledeće instrukcije, da ih memorise, stavi u bafer.
- Savremeni procesori imaju mogućnost predvidjanja više od jedne grane unapred.

Analiza toka podataka

- Procesor analizira koje instrukcije zavise jedna od druge, na osnovu rezultata ili podataka.
- Na taj nacin optimizuje raspored instrukcija.
- Instrukcije se rasporedjuju prema redosledu izvorsavanja, a ne prema prvobitnom redosledu u programu.
- Na taj nacin sprecavaju se kasnjenja.

Spekulativno izvršavanje

- Koriscenjem predvidjanja grananja I analize toka podataka, procesori spekulativno izvršavaju instrukcije unapred, pre njihovog stvarnog pojavljivanja u izvršenju programa.
- Rezultati se cuvaju u privremenim lokacijama.

Ravnoteza performansi

- Brzina i snaga procesora rastu ogromnom brzinom, druge komponente računara ne drže korak.
- Potrebno je da se napravi ravnoteža performanse, da se podesi organizacija i arhitektura u cilju kompenzacije razlika u performansama.

Ravnoteza performansi

- Najveci problem-interfejs procesor-memorija
- Brzina procesora vrtoglavo raste, brzina prenosa podataka izmedju glavne memorije I procesora postala je kriticzna.
- To predstavlja najkriticniju putanju za kompletan racunarski sistem.
- Ukoliko memorija ili interfejs ne drze korak sa procesorom, on mora da ceka, da se zaustavi I na taj nacin gubi vreme za obradu.
- Sa aspekta arhitekture racunara, problem moze da se resi na sledece nacine:

Ravnoteza performansi

- Povećanje broja bitova koji se citaju iz DRAM memorije, što znači da memorije moraju da imaju širu magistralu, a manju dubinu.
- Uključiti keš memoriju u interfejs između DRAM memorije i procesora ili buferovanje na samom DRAM cipu.

Ravnoteza performansi

- Smanjiti frekvenciju pristupa glavnoj memoriji postavljanjem složenijih struktura keš memorije. Ugradjivanje jedne ili više keš memorija na samom procesoru kao i keša izvan cipa, blizu procesora.
- Povećavanje propusnog opsega veze između procesora i memorije, magistrale većih brzina, hijerarhija magistrala da bi se tok podataka baferovao i strukturirao.

Ravnoteza performansi

- Upravljanje UI uredjajima predstavlja drugi segment kod projektovanja arhitekture u cilju povecanja brzine.
- UI uredjaji upucuju velike zahteve za podacima i propusnom moci podataka.
- Procesor je u stanju da obradi podatke, ali ostaje problem prenosa podataka izmedju samih uredjaja, procesora i uredjaja.

Tipicne brzine UI uredjaja

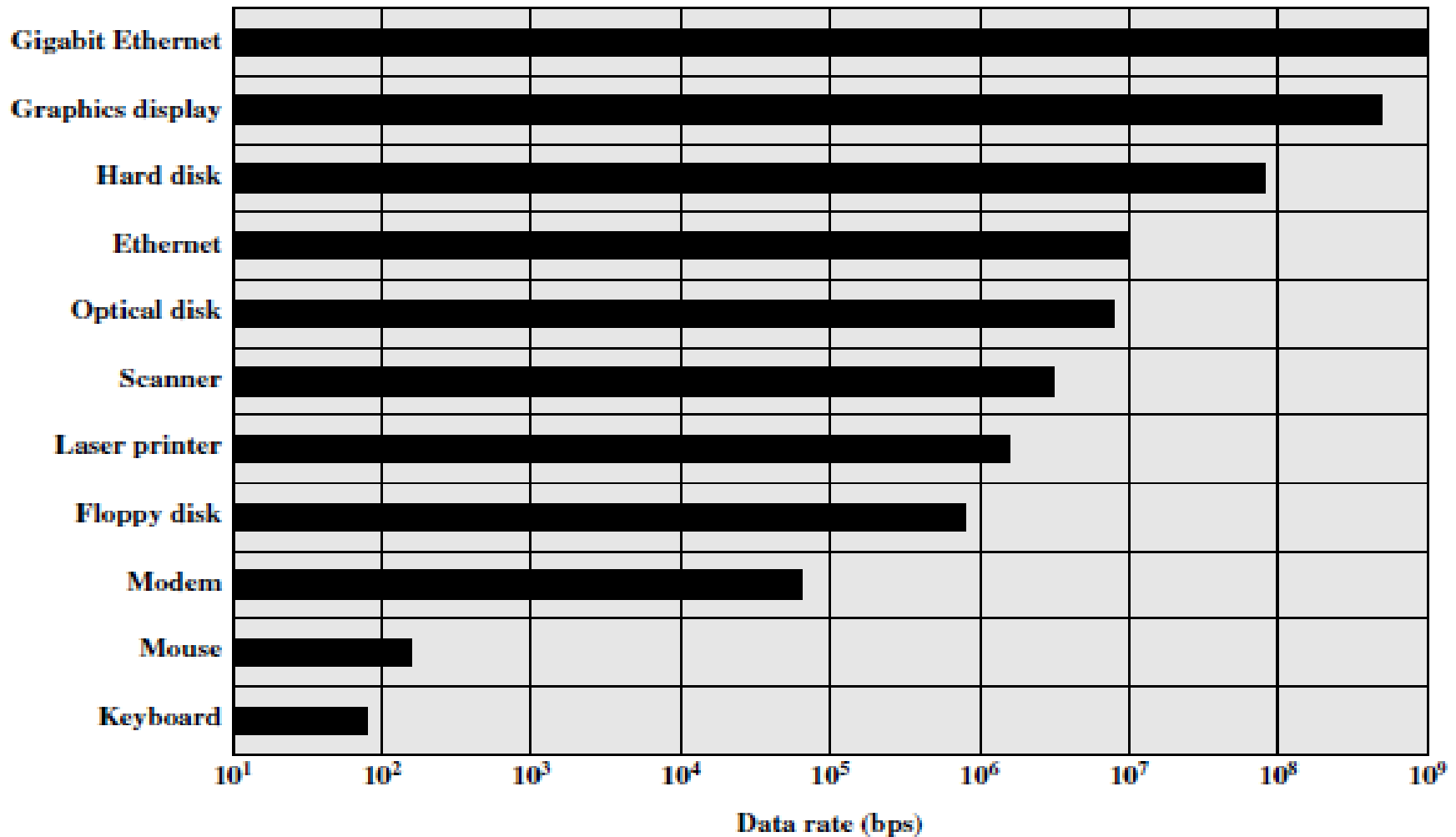


Figure 2.11 Typical I/O Device Data Rates

Ravnoteza performansi

- Ključ je ravnoteža propusne moći i zahteva za obradu procesora, glavne memorije, UI uređaja i međusobno povezanih struktura.
- Ti zahtevi se stalno menjaju u skladu sa brzinama kojima se menjaju performanse u različitim tehnološkim oblastima, procesor, magistrala, memorija, periferni uređaji...
- Nove aplikacije i novi periferni uređaji stalno menjaju prirodu zahteva u sistemu.

Poboljšanja u organizaciji i arhitekturi cipa

- Povecati hardversku brzinu precesora, povezano je sa cinjenicom da se vise logickih kola moze spakovati zajedno i povecati brzinu takta. Kada su logicka kola bliza jedno drugom, vreme prostiranja signala se znacajno smanjuje sto dovodi do ubrzanja procesora. Povecanje brzine generatora takta znaci da se pojedinačne operacije brze izvrsavaju.

Poboljšanja u organizaciji i arhitekturi čipa

- Povećanje veličine i brzine keš memorije umetnute između procesora i glavne memorije.
- Umetanje keš memorije na samom procesorskom čipu, značajno smanjuje vreme za pristup keš memoriji.

Poboljšanja u organizaciji i arhitekturi cipa

- Promena organizacije i arhitekture procesora tako da se poveća efektivna brzina izvršenja instrukcija.
- To znači korišćenje paralelizma!

Poboljšanja u organizaciji i arhitekturi cipa

- Povećanje brzine generatora takta i gustine logičkih kola nastaju sledeća ograničenja:

Poboljšanja u organizaciji i arhitekturi čipa

- Električno napajanje: povećanjem gustine logičkih kola i taktne brzine, gustina napajanja električnom energijom se povećava(W/cm^2).
- Javlja se problem hlađenja, odvodjenja toplote i to su konstruktivni problemi.

Poboljšanja u organizaciji i arhitekturi cipa

- RC kasnjenje
- Memorijsko kasnjenje, memorijske brzine kasne za procesorskim brzinama.
- Za poboljšanje performansi akcenat se stavlja na organizaciju i arhitekturu.

Poboljšanja u organizaciji i arhitekturi cipa

- Prvi pristup je je povećanje kapaciteta keš memorije.
- Sada postoje dva ili tri nivoa keša između procesora i glavne memorije.
- Kako se povećava gustina cipa, sve više keš memorije se ugradjivalo u procesorski cip.
- Na savremenim procesorima za keš je izdvojena polovina cipa.

Poboljšanja u organizaciji i arhitekturi cipa

- Drugi pristup je logika za izvršenje instrukcija unutar procesora postaje sve složenija da bi se omogućio paralelizam u izvršavanju instrukcija u samom procesoru.
- U tom smislu izdvajaju se ***protocna obrada i superskalarni pristup.***

Poboljšanja u organizaciji i arhitekturi cipa

- Protocna obrada omogućava da se različite faze izvršenja različitih instrukcija desavaju u isto vreme.
- Superskalarni pristup predstavlja više protocnih obrada unutar procesora, tako da instrukcije koje ne zavise jedna od druge mogu da se izvrsavaju nezavisno, paralelno.

Poboljšanja u organizaciji i arhitekturi cipa

- Za nastavak povećanja performansi računara, da bi se smanjili negativni efekti povećanja brzine takta, gustine, projektanti procesora umesto razvoja složenog procesora prelaze na princip projektovanja računarskog cipa sa više jezgara.

Poboljsanja u organizaciji i arhitekturi cipa

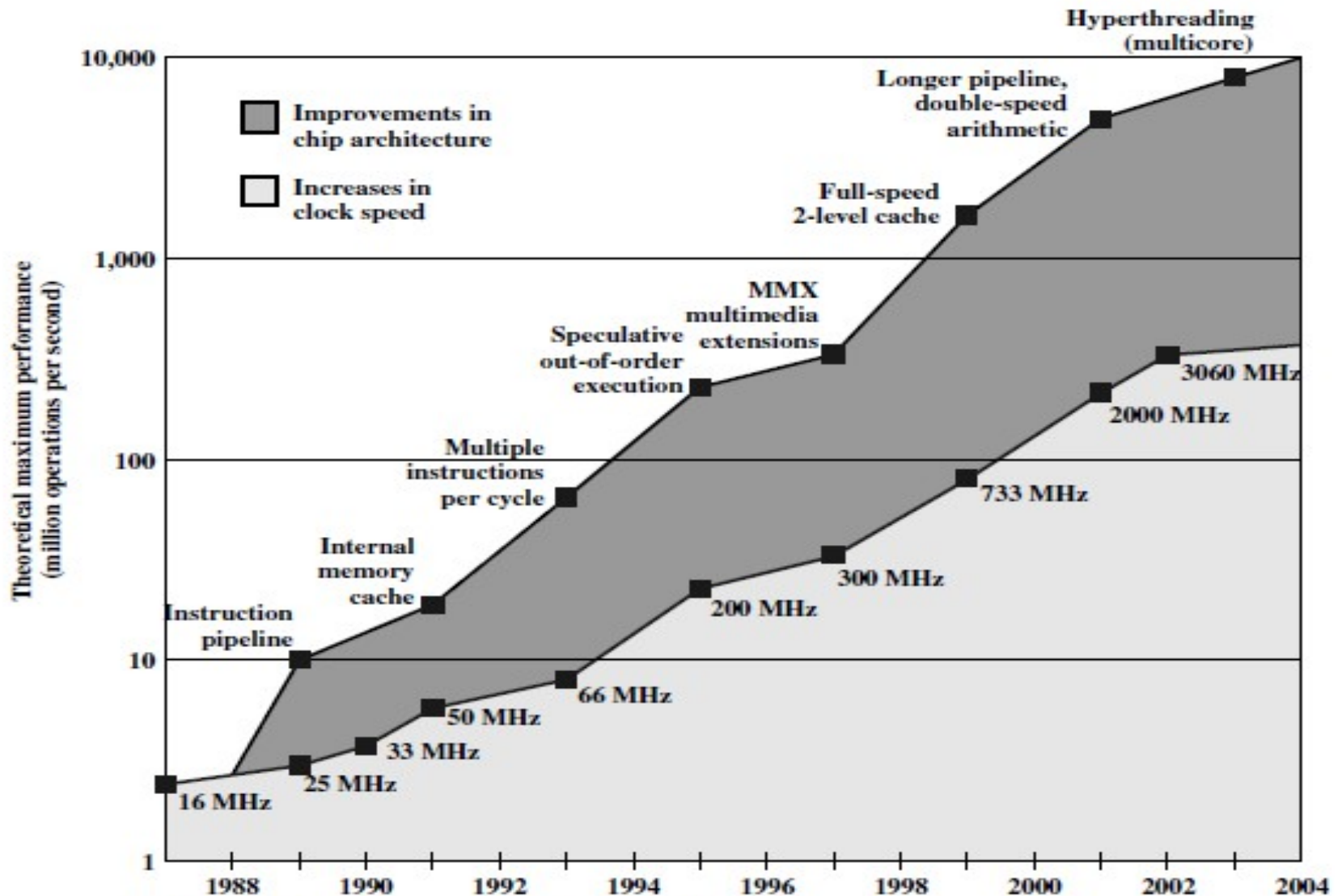


Figure 2.12 Intel Microprocessor Performance [GIBB04]

Vise jezgara, MIC i GPGPU

- Vise procesora na istom cipu, naziva se visestruka jezgra ili multicore.
- Multicore omogućava povećanje performansi bez povećanja brzine generatora takta.
- Istrazivanja pokazuju da je unutar procesora povećanje performansi približno proporcionalno kvadratnom korenu povećanja njegove složenosti.

Vise jezgara

- Dupliranje broja procesora skoro duplira performansu, ukoliko software moze da podrzi multicore.
- Strategija je da se koriste dva jednostavnija procesora na cipu, a ne jedan slozeni procesor.
- Dva procesora opravdavaju upotrebu vecih kes memorija na cipu.
- Potrosnja elektricne energije za memorijsku logiku na cipu je mnogo manja od one za procesorsku logiku.

Vise jezgara

- Kako raste gustina logickih kola na cipu, nastavlja se trend povecanja broja jezgara, kao i vecih kes memorija.
- Povecanje kapaciteta kes memorija, u cilju povecanja performansi, prave se dva i tri nivoa kes memorije.
- Prvi nivo namenjen je pojedinaenom procesoru, a nivoi dva i tri se dele izmedju svih procesora.

Vise jezgara

- Povećanje broja jezgara od početnih 8, 16 i tako dalje, do današnjih 50 po čipu, povećava performanse.
- Stavljaju se izazovi pred projektovanjem softvera u cilju efikasnog iskoriscenja tako velikog broja procesora.
- Ovako veliki broj procesora naziva se MIC (*many integrated core*), mnogo integrisanih jezgara.

GPU

- Povećanje broja jezgara u cilju projektovanja procesora opšte namene.
- Istovremeno se projektuju čipovi sa više procesora opšte namene, a uz njih i više grafickih procesorskih jedinica (GPU, *graphic processing unit*).
- GPU jezgro je projektovano da izvodi paralelne operacije nad grafickim podacima.
- Nalazi se na grafickim karticama i koristi za renderovanje 2D i 3D grafike, kao i za obradu videa.

GPU

- Kako GPU izvode paralelne operacije na vise skupova podataka, sve vise se koriste kao vektorski procesori za razne primene koje zahtevaju ponavljajuca racunanja.
- Na taj nacin brise se granica izmedju GPU I CPU I pomocu ovakvog procesora podrzava se sirok spektar primena.
- Ovo se naziva racunanje opste namene na GPU(*GPGPU general-purpose computing on GPUs*)

Evolucija INTEL x86

- **8080:** The world's first general-purpose microprocessor. This was an 8-bit machine, with an 8-bit data path to memory. The 8080 was used in the first personal computer, the Altair.
- **8086:** A far more powerful, 16-bit machine. In addition to a wider data path and larger registers, the 8086 sported an instruction cache, or queue, that prefetches a few instructions before they are executed. A variant of this processor, the 8088, was used in IBM's first personal computer, securing the success of Intel. The 8086 is the first appearance of the x86 architecture.
- **80286:** This extension of the 8086 enabled addressing a 16-MByte memory instead of just 1 MByte.
- **80386:** Intel's first 32-bit machine, and a major overhaul of the product. With a 32-bit architecture, the 80386 rivaled the complexity and power of minicomputers and mainframes introduced just a few years earlier. This was the first Intel processor to support multitasking, meaning it could run multiple programs at the same time.
- **80486:** The 80486 introduced the use of much more sophisticated and powerful cache technology and sophisticated instruction pipelining. The 80486 also offered a built-in math coprocessor, offloading complex math operations from the main CPU.
- **Pentium:** With the Pentium, Intel introduced the use of superscalar techniques, which allow multiple instructions to execute in parallel.
- **Pentium Pro:** The Pentium Pro continued the move into superscalar organization begun with the Pentium, with aggressive use of register renaming, branch prediction, data flow analysis, and speculative execution.
- **Pentium II:** The Pentium II incorporated Intel MMX technology, which is designed specifically to process video, audio, and graphics data efficiently.
- **Pentium III:** The Pentium III incorporates additional floating-point instructions to support 3D graphics software.
- **Pentium 4:** The Pentium 4 includes additional floating-point and other enhancements for multimedia.⁸
- **Core:** This is the first Intel x86 microprocessor with a dual core, referring to the implementation of two processors on a single chip.
- **Core 2:** The Core 2 extends the architecture to 64 bits. The Core 2 Quad provides four processors on a single chip.

Ugradjeni sistemi I ARM

- Za razliku od laptop I desk top racunara ugradjeni sistam podrazumeva:
- Kombinaciju racunarskog hardvera I softvera, kao I dodatnih mehanickih sklopova, projektavanih da izvedu namensku funkciju, nazivaju se ugradjeni sistemi.
- Cesto su ugradjeni sistemi deo veceg sistema ili proizvoda.

Ugradjeni sistemi I ARM

Table 2.7 Examples of Embedded Systems and Their Markets [NOER05]

Market	Embedded Device
Automotive	Ignition system Engine control Brake system
Consumer electronics	Digital and analog televisions Set-top boxes (DVDs, VCRs, Cable boxes) Personal digital assistants (PDAs) Kitchen appliances (refrigerators, toasters, microwave ovens) Automobiles Toys/games Telephones/cell phones/pagers Cameras Global positioning systems
Industrial control	Robotics and controls systems for manufacturing Sensors
Medical	Infusion pumps Dialysis machines Prosthetic devices Cardiac monitors
Office automation	Fax machine Photocopier Printers Monitors Scanners

Ugradjeni sistemi I ARM

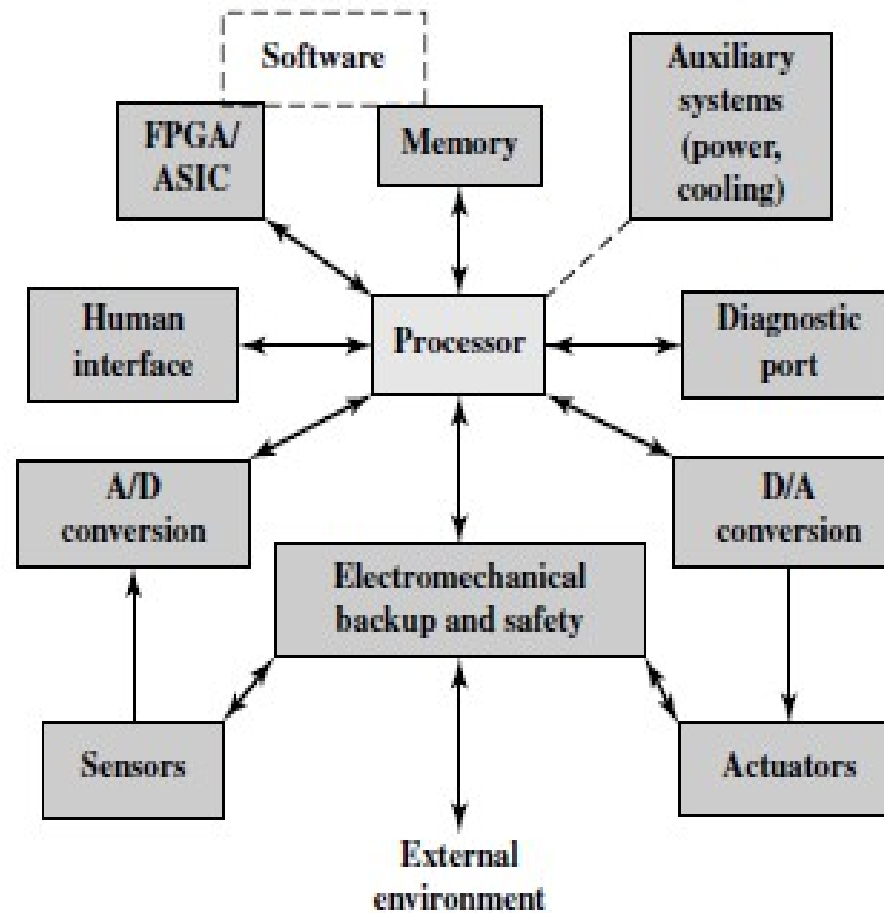


Figure 2.13 Possible Organization of an Embedded System

Evolucija ARM-a

Table 2.8 ARM Evolution

Family	Notable Features	Cache	Typical MIPS @ MHz
ARM1	32-bit RISC	None	
ARM2	Multiply and swap instructions; Integrated memory management unit, graphics and I/O processor	None	7 MIPS @ 12 MHz
ARM3	First use of processor cache	4 KB unified	12 MIPS @ 25 MHz
ARM6	First to support 32-bit addresses; floating-point unit	4 KB unified	28 MIPS @ 33 MHz
ARM7	Integrated SoC	8 KB unified	60 MIPS @ 60 MHz
ARM8	5-stage pipeline; static branch prediction	8 KB unified	84 MIPS @ 72 MHz
ARM9		16 KB/16 KB	300 MIPS @ 300 MHz
ARM9E	Enhanced DSP instructions	16 KB/16 KB	220 MIPS @ 200 MHz
ARM10E	6-stage pipeline	32 KB/32 KB	
ARM11	9-stage pipeline	Variable	740 MIPS @ 665 MHz
Cortex	13-stage superscalar pipeline	Variable	2000 MIPS @ 1 GHz
XScale	Applications processor; 7-stage pipeline	32 KB/32 KB L1 512 KB L2	1000 MIPS @ 1.25 GHz

DSP = digital signal processor

SoC = system on a chip