

Arhitectura Calculatoarelor

T.3 –Tipuri de arhitecturi ale calculatoarelor numerice

Tipuri de arhitecturi ale calculatoarelor numerice. Arhitectura von Neumann, Mașina Turing, Clasificarea Calculatoarelor, Taxonomia Flynn, Taxonomia lui Wang, Clasificare comercială.

Scopul Lecției: De a face cunoștință cu tipurile de arhitecturi a sistemelor de calcul și clasificarea sistemelor de calcul după diferitele elemente specifice.

Studentul trebuie să cunoască:

- § Arhitectura von Neumann, Mașina Turing;
- § Clasificarea calculatoarelor după Flynn, Wang, Clasificarea Comercială
- § Trendul în arhitectura calculatoarelor

Conf. Univ. Dr. Crețu Vasilii

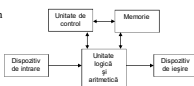
Arhitectura von Neumann

Într-un articol publicat în 1947, John von Neumann a expus niște principii care stau la baza calculatoarelor moderne. Acestea sunt:

- Existența unui **mediu de intrare** prin intermediul căruia să poată fi introdus un număr practic nelimitat de date și instrucțiuni.
- Existența unei **memorii** în care să fie depuși operanții și instrucțiunile și de unde să fie preluate rezultatele în ordinea dorită.
- Existența unei **scuții de calcul** care să fie capabilă să execute operații aritmetice și logice asupra datelor din memorie.
- Existența unui **mediu de ieșire** prin intermediul căruia să poată fi comunicat utilizatorului un număr nelimitat de instrucțiuni.
- Existența unei **unități de comandă** capabilă să interpreteze instrucțiunile citite în memorie și, pe baza informațiilor citite în memorie și a informațiilor furnizate de secțiunea de calcul, să fie capabilă să decidă între mai multe variante de desfășurare a operațiilor.

• Datele și instrucțiunile trebuie să fie stocate în memorie sub aceeași formă.

Programul stocat este cel mai important bloc al modelului von Neumann. Un program este stocat în memoria calculatorului împreună cu datele ce sunt procesate. Înainte de apariția calculatoarelor cu program stocat, programele erau stocate pe medii externe cum ar fi cartele perforate. În calculatorul cu program stocat acesta poate fi manipulat ca și cum ar reprezenta date. Aceasta a dus la apariția compilatoarelor și sistemelor de operare și face posibilă marea versatilitate a calculatoarelor moderne.



Ca urmare, caracterizarea arhitecturii von Neuman se face prin :

- utilizarea memoriei interne pentru a stoca secvențe de control pentru îndeplinirea unei anumite sarcini – secvențe de programe;
- datele, cât și instrucțiunile sunt reprezentate ca siruri de biți și sunt stocate într-o memorie readwrite;
- conținutul memoriei se poate accesa în funcție de locație (adresă), indiferent de tipul informației conținute;
- execuția unui set de instrucțiuni se efectuează secvențial, prin citirea de instrucțiuni consecutive din memorie.

Arhitectura Harvard și avantajele sale

Atunci când se implementează microprocesoare, se folosesc în mod tradițional două abordări pentru construirea unei arhitecturi:

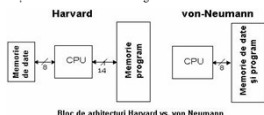
- Arhitectura lui Von Neumann
- Arhitectura Harvard

Arhitectura von Neumann utilizează o memorie omogenă a microprocesorului. Pot fi scrise diferite programe în această memorie. În acest caz, un program special de încărcare funcționează împreună cu ele ca date. Apoi, controlul poate fi transferat la aceste programe și ci încep deja să-și execute algoritmul. Cu această abordare a controlului cu microprocesor, este posibil să se obțină o flexibilitate maximă a sistemului microprocesor.

Dezavantajul arhitecturii von Neumann este posibilitatea unei încălziri neintenționată a performanței sistemului (erori de soft) și distrugerea intenționată a funcționării sale (atac viral). În arhitectura Harvard două tipuri de memorie microprocesor sunt fundamental diferite:

- Memorie program (pentru stocarea instrucțiunilor microprocesorului)
- Memorie de date (pentru stocarea temporară și prelucrarea variabilelor)

Arhitectura Harvard este fundamental imposibil de a efectua o operație de scriere în memoria de program, care elimină posibilitatea distrugerii accidentale a programului de control în cazul erorilor de program, atunci când se lucrează cu date, sau atacuri de către terțe părți. Mai mult, pentru o memorie de program și o memorie de date sunt organizate bus de schimb de date individuale (magistrala de sistem).



Bloc de arhitecturi Harvard vs. von Neumann

Aceste caracteristici au determinat domeniul arhitecturii Harvard. Arhitectura Harvard este folosită în microcontrolere și în procesoare de semnale, unde este necesară asigurarea unei fiabilități ridicate a echipamentului. Semnalul de procesare arhitectura de la Harvard completat folosind trei-bus unitate de microprocesor de operare. Arhitectura trei bus permite unității de operare pentru a combina operațiunile de citire două operații cu rezultate recori ale comenzii în memoria microprocesorului. Acest lucru sporște foarte mult performanța microprocesorului de semnal fără a crește frecvența ceasului.

Structura modificată Harvard este utilizată în jetoanele moderne ale procesoarelor de semnal. Chiar și în continuare pe calea reducerii costului cristallului prin reducerea zonei ocupate de magistralele de sistem a mers la un singur cristal producătorii de calculatoare - microcontrolere. Aceste cipuri folosesc un sistem bus pentru transmiterea de comenzi și date (arhitectura Harvard modificată) și în interiorul cristallului.

Procesoarele de semnalizare necesită deseori mai multe buse interne pentru a implementa algoritmi precum transformarea rapidă Fourier și filtrarea digitală. De obicei, sunt utilizate două magistrale de date, o magistrală de scriere a datelor și o magistrală de citire a instrucțiunilor. O astfel de structură a microprocesorului a fost numită arhitectura extinsă a lui Harvard.

Această abordare practică producătorii de semnal procesoare - firma Analog Devices (familie de procesoare de semnal Blackfin și Tiger Shark), Texas Instruments (familia C5000™ DSP-uri de procesoare de semnal și C6000™ DSP-uri), Freescale (familia DSP56K MSC8251 și procesoare de semnal).

Cum funcționează arhitectura Harvard?

Arhitectura Harvard are diferite zone de adresă de memorie pentru program și pentru date.

Acest lucru are ca rezultat capacitatea de a proiecta un circuit în așa fel încât o magistrală și un circuit de control pot fi utilizate pentru a gestiona fluxul de informații din memoria programului și unul separat pentru a gestiona fluxul de informații către memoria de date.

Utilizarea de magistrale separate înseamnă că este posibil ca un program să fie recuperat și executat fără a fi întrerupt de transferul ocazional de date în memoria de date.

De exemplu, într-o versiune simplă a acestei arhitecturi, unitatea de recuperare a programului ar putea fi ocupată cu recuperarea următoarelor instrucțiuni din secvența programului și în paralel efectuarea unei operații de transfer de date care ar fi putut face parte din instrucțiunea programului anterior.

La acest nivel, arhitectura Harvard are o limitare, deoarece în general nu este posibil să plasați codul programului în memoria de date și să îl executați de acolo.

Adăugări în arhitectură

Multe variante existente mai complicate pot fi adăugate la forma simplă a arhitecturii Harvard.

O adăugare obișnuită este adăugarea unui cache de instrucțiuni la magistrala de date a programului, care permite unității de execuție a instrucțiunilor un acces mai rapid la următorul pas din program, fără a fi nevoie să mergeti la o memorie mai lentă pentru a ajunge la pas, programului ori de câte ori este necesar.

Adrese de memorie

Un computer Harvard are zone de adresă de date și instrucțiuni diferite: adresa de instrucțiuni una nu este aceeași zonă cu adresa de date una.

Adresa de instrucțiuni l ar putea conține o valoare de douăzeci și patru de biți, în timp ce adresa de date una ar putea indica un octet de 8 biți, care nu face parte din acea valoare de douăzeci și patru de biți.

Sistem de memorie

Deoarece există o zonă de memorie separată pentru instrucțiuni și date, separând atât semnalele, cât și memoria de stocare a codului și a datelor, acest lucru face posibilă accesarea simultană a fiecărui sistem de memorie.

Avantaje

- Există mai puține șanse de corupție în transmisie, deoarece datele și instrucțiunile sunt transferate prin diferite autobuze.
- Datele și instrucțiunile sunt accesate în același mod.
- Permite diferite medii de stocare pentru instrucțiuni și date. De exemplu, puteți pune instrucțiunile într-un ROM ieftin și datele în RAM scump.
- Cele două amintiri pot utiliza dimensiuni de celule diferite, făcând astfel o utilizare eficientă a resurselor.
- Are o lățime de bandă de memorie mai mare, care este mai previzibilă având memorii separate pentru instrucțiuni și date.

Nivel de protecție

Pe sistemele care nu au o unitate de gestionare a memoriei, oferă un nivel suplimentar de protecție, deoarece datele nu pot fi pomite ca cod, expunând sistemul la multe probleme, cum ar fi depășirea bufferului.

De aceea este popular cu sistemele încorporate mici, cum ar fi un cupitor cu microunde sau un ceas.

Viteză mai mare

Arhitectura Harvard poate citi o instrucțiune și, de asemenea, accesa memoria de date simultan cu o viteză rapidă.

Oferi performanțe mai mari, deoarece permite obținerea simultană a datelor și instrucțiunilor pentru a fi stocate în amintiri separate și a călătorii prin diferite autobuze.

O arhitectură Harvard va ajuta, în general, un computer cu un anumit nivel de complexitate să funcționeze mai repede decât o arhitectură Von Neumann, alăta timp cât nu este necesar să partajați resursele între cod și memoriile de date.

Dacă limitările pinului sau alți factori forțază utilizarea unei singure magistrale pentru a accesa ambele spații de memorie, este posibil ca aceste beneficii să fie anulate în mare măsură.

Dezavantaje

Complexitate și costuri mai mari

Problema arhitecturii Harvard este complexitatea și costul său mare, deoarece în loc de o magistrală de date, acum sunt necesare două.

Producerea unui computer cu două autobuze este mult mai scumpă și consumă mult timp. Este nevoie de o unitate de control pentru două autobuze, care este mai complexă și consumă mult timp și este costisitoare de dezvoltat.

Acesta înseamnă o implementare mai complexă pentru producători. Necesită mai mulți pini pe CPU, o placă de bază mai complexă și trebuie să dublice cipurile RAM, precum și un design cache mai complex.

Utilizare redusă

Arhitectura Harvard nu este utilizată pe scară largă, ceea ce face mai dificilă implementarea. Acesta este motivul pentru care este rar folosit în afara procesorului.

Cu toate acestea, această arhitectură este uneori utilizată în CPU pentru a-și gestiona cache-urile.

Utilizarea greșită a spațiului de memorie

Când există spațiu liber în memoria de date, acesta nu poate fi folosit pentru a stoca instrucțiuni și invers.

Prin urmare, amintirile speciale dedicate fiecăruia dintre ele trebuie să fie atent echilibrate în fabricarea lor.

Mașina Turing

În 1936, matematicianul englez Allan Turing a creat un automat abstract care să opereze cu numere calculabile. Un număr calculabil este un număr a cărei parte zecimală poate fi determinată cu un număr finit de iterații.

Automatul a fost sintetizat pe baza următoarelor ipoteze:

- Automatul are un număr n finit de stări;
- Automatul se află în orice moment într-o stare $i, 1 \leq i \leq n$, urmând ca în momentul imediat următor să se afle în starea $j, 1 \leq j \leq n$.
- Fiecare din cele n stări se caracterizează prin:
 - valoarea caracteristică e_j , care este o valoare curentă a numărului ce se calculează;
 - funcția f_j , care aplicată stării i permite obținerea următoarei stări e_{j+1} ;
 - deplasamentul d_j , care va trebui aplicat numărului pentru a se realiza din starea i în starea j , adică $j=i+d_j$.

-Sistemul reprezentativ (SR) sau memoria mașinii este construit dintr-o bandă magnetică de lungime practic infinită, împărțită în segmente de lungime egală, fiecare segment putând stoca un număr finit de semne.

-Procesorul (P) este un circuit secvențial cu un număr finit de stări, care poate executa următoarele instrucțiuni, (setul de instrucțiuni al mașinii):

- schimbă segmentul de pe bandă de la poziția curentă;
- poziționează capul de citire (C) cu o poziție la dreapta;
- poziționează capul de citire cu o poziție la stânga;
- oprește sistemul.

Pentru a realiza un calcul cu această mașină, se înscriu datele într-un mod convenabil și se descompune algoritmul de calcul, în funcție de modul de reprezentare a datelor, într-o secvență de instrucțiuni ale mașinii. Ultima instrucțiune este cea de oprire a mașinii.

Modelul funcțional al mașinii Turing;
SR-sistem reprezentativ;
C/S-cap citire/scriere.

Mașina Turing este compusă din următoarele piese:

- o bandă infinită de hirtie cu pătrățele; în fiecare pătrățel se poate scrie exact un caracter din alfabetul nostru; banda este inițial plină cu "spații", mai puțin o parte (să-i spunem de început) unde este scris șirul cu datele de intrare;
- un cap de citire-scriere, care se poate mișca deasupra benzii, la stînga sau la dreapta;
- o unitate de control, care conține un număr finit de reguli care indică mașinii ce să facă la fiecare mișcare în funcție de litera curentă de pe bandă și starea în care mașina se află. Unitatea de control se află, în fiecare moment dat, într-o stare; stările posibile sunt fixate dinainte, și sunt în număr finit.

Fiecare regulă are forma următoare:

Dacă

- sunt în starea s_1 ;
- sub capul de citire este litera a ;

atunci:

- trece în starea s_2 ;
- scriu pe bandă litera b ;
- mut capul de citire/scriere în direcția D .

Și asta-i tot! Orice algoritm de calcul poate descrie de o astfel de mașină, prin toate stările posibile, și toate aceste reguli, numite reguli de tranziție, care indică cum se trece de la o stare la alta. În pofida simplității, mașina Turing poate deci calcula orice se poate calcula cu cele mai performante supercomputere

Clasificarea Calculatoarelor

Este destul de dificil să se clasifice tipurile de calculatoare din cauza multitudinii lor. Totuși, anumite taxonomii s-au impus. Vom prezenta trei clasificări după arhitectură și una strict comercială.

Taxonomia Flynn

A fost publicată în 1966 de către M. J. Flynn care avea în vedere existența într-un sistem de calcul a două fluxuri:

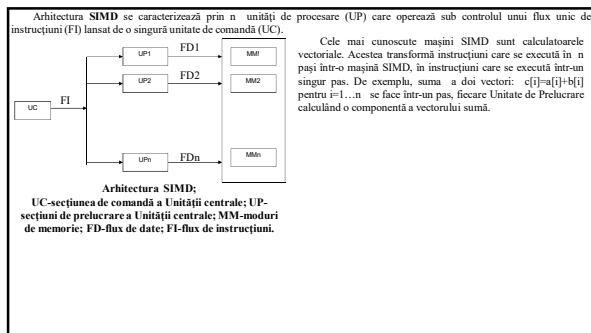
- fluxul de instrucțiuni, care reprezintă programul;
- fluxul de date, care reprezintă datele de intrare sau rezultatele parțiale.

Clasificarea lui Flynn ia în considerare gradul de multiplicitate ale celor două fluxuri și identifică patru tipuri de arhitecturi:

- SISD (Single Instruction Single Data)
- SIMD (Single Instruction Multiple Data)
- MISD (Multiple Instructions Single Data)
- MIMD (Multiple Instructions Multiple Data)

SISD realizează o execuție secvențială a instrucțiunilor. Principalul neajuns al acestei arhitecturi este viteza de execuție care, la un moment dat, este plafonată, situație denumită „gâtul sticlei lui Neumann” (Neumann Bottleneck). Spargerea acestei limitări este realizată prin arhitectura paralelă.

Arhitectura SISD; UC-secțiunea de comandă a Unității centrale; UP-secțiunea de prelucrare a Unității centrale; MM-modul de memorie; FD-flux de date; FI-flux de instrucțiuni.



MISD nu are nici un sens și de aceea nu este utilizată.

MIMD cuprinde două feluri de mașini:

- multiprocesoare și
- multicalculatoare.

Multiprocesoarele se caracterizează prin existența memoriei comune la care au acces n procesoare. Schimbul de informație dintre procesoare se face prin variabile partajate din memoria comună la care au acces toate procesoarele, însă accesul trebuie făcut prin excludere mutuală pentru a realiza ceea ce se numește consistența memoriei.

Multicalculatoarele se caracterizează prin existența unui număr foarte mare de calculatoare (de la ordinal sutelor în sus) care sunt legate printr-o rețea topologică. Fiecare procesor are memoria lui locală, văzută doar de el, iar comunicarea între procesoare se face prin mesaje.

Taxonomia lui Wang (Feng) (Цзе-юнь Фэн)

Această clasificare presupune o organizare matricială a datelor. O matrice de dimensiunea $m \times n$ are m cuvinte, fiecare cuvânt are o lungime de n biți. Criteriul de clasificare este gradul de paralelism în procesarea datelor organizate matricial. Conform acestui criteriu există patru tipuri de arhitecturi și anume:

- WSBS** (Word Serial-Bit Serial) în care se lucrează pe un singur cuvânt, fiecare cuvânt fiind prelucrat serial bit cu bit, $ns1, ms1, \dots$
- WSBP** (Word Serial-Bit Paralel) în care se lucrează pe un singur cuvânt , biții fiecărui cuvânt fiind prelucrați simultan, $n>1, ms1$.
- WPBS** (Word Paralel-Bit Serial) în care se lucrează pe un singur bit la toate cuvintele simultan, $ns1, m>1$.
- WPBP** (Word Paralel-Bit Paralel) în care se lucrează simultan pe toate cuvintele și pe toți biții, $n>1, m>1$.

WSBS nu are elemente de paralelism.
 WSPB și WPBS sunt parțial paralele, fiind orientate pe prelucrarea vectorilor.
 WPBP este complet paralel.

Taxonomia lui Shore

Spre deosebire de Flynn, Shore și-a bazat clasificarea pe modul în care este organizat calculatorul din părțile sale componente. Din acest punct de vedere, au fost identificate șase tipuri de mașini, fiecare atribuită o cifră romană.

Mașina I are o arhitectură convențională von Neumann, cu următoarea structură:

- unitate de comandă (CU)
- unitate de procesare (PU)
- memorie pentru instrucțiuni (IM)
- memorie pentru date (DM)

O cifră a DM produce toți biții unui cuvânt, care sunt prelucrați în paralel de PU, PU putând conține mai multe unități funcționale.

Această clasă include calculatoarele vectoriale, de exemplu CRAY 1.

Mașina II este similară mașinii I, cu deosebirea că, în timp ce mașina I citește slice-uri orizontale, mașina II citește un slice vertical.

Exemple de calculatoare de tip mașina II Shore : ICL, DAP, STARAN etc.

Mașina III este o combinație a mașinilor I și II. Un exemplu de astfel de mașină este calculatorul ortogonal Shooman (1970)

Mașina IV se obține prin multiplicarea unităților PU și DM din mașina I și prin trimiterea acestui ansamblu de la o singură unitate de control UC. Exemplu: PEPE.

Mașina V este exact mașina IV cu facultața suplimentară că unitățile PU sunt așezate pe o linie și se asigură conexiuni între vecinii cei mai apropiați; fiecare PU poate adresa informații din memoria sa dar și din cea a vecinilor săi imediați. Este un masiv conectat. Exemplu: ILIA CIV.

Mașina VI este denumită mașina cu logica în memorie. Este o abordare alternativă a distribuirii comenzii în memorie. Exemplu: calculatoare cu memorii asociative.


Clasificarea lui Tanenbaum

Clasificarea calculatoarelor paralele (Tanenbaum)


MPP = procesoare masiv paralele (*Massive Parallel Processors*)

COW = cluster de stații de lucru (*Cluster Of Workstations*)

Minicalcutoarele pot efectua sute de milioane de operații pe secundă, iar prețul lor nu depășește 200-300 de mii de dolari. Echipamentele periferice ale unui minicalculator includ câteva discuri magnetice, una sau două imprimante, mai multe console. Minicalcutoarele sînt mai ușor de utilizat și operat decît calculatoarele mari și se utilizează în proiectarea asisată de calculator, în automatizări industriale, pentru prelucrarea datelor în experimentele științifice etc. Dintre firmele producătoare de minicalcutoare vom remarca *IBM, Wang, Texas Instruments, Data General, DEC, Hewlett-Packard* etc.



Microcalculatoarele, denumite și calculatoare personale, sînt realizate la prețuri scăzute - între 100 și 15000 de dolari și asigură o viteză de calcul de ordinul milioanele de operații pe secundă. Echipamentele periferice ale unui microcalculator includ o unitate de disc rigid, una sau două unități de disc flexibil, o imprimantă și o consolă. Structura modulară și gruparea tuturor echipamentelor în jurul unei magistrale permite configurarea microcalculatorului în funcție de necesitățile individuale ale fiecărui utilizator. Corporații care produc microcalculatoare există în foarte multe țări, însă lideri mondiali, unanim recunoscuți, sînt firmele *IBM, DEC, Hewlett-Packard, Apple, Olivetti* etc.



TRENDUL ÎN ARHITECTURA CALCULATORILOR

Din punct de vedere tehnologic cele mai importante tendințe sînt:

- gradul de integrare al tranzistorilor pe cip crește cu cca. 55% pe an; tehnologia de integrare a microprocesoarelor a evoluat de la 10 micrometri (1971) la 0,18 micrometri 2001;
- frecvența ceasului crește și ea cu 50% pe an;
- pentru memoriile DRAM, densitatea de integrare crește cu cca 40-50% pe an, iar timpul de acces aferent scade cu 3 % pe an.

-tehnologia și performanțele rețelelor se îmbunătățesc semnificativ.

Se poate spune că aceste tendințe respectă legea lui Gordon Moore, cofondator împreună cu Obert Noyce a societății INTEL. Aceasta, în 1965, enunță celebra sa lege: *numărul de tranzistori din circuitele integrate se va dubla la fiecare doi ani*. Aceasta înseamnă că, la fiecare 10 ani se schimbă prefixul de măsurare, adică totul crește de 1000 de ori. Într-adevăr, dacă hard discurile din anii 90 aveau 100 MB, în 2000 ele au 100 GB. Frecvența ceasului era în 1990 de 8 MHz iar în 2000 era de 1 GHz etc.

Între cele mai evidente tendințe de evoluție în arhitectură amintim:

- exploatarea paralelismului la nivelul instrucțiunilor și firelor de execuție, atât prin tehnici statice (soft) cît și dinamice(hard);
- există și tehnici hibride cum ar fi cazul procesorului Intel Itanium IA-64;
- structuri tot mai performante de ierarhizare a sistemului de memorie prin utilizarea arhitecturilor evolute de memorie cache;
- reducerea latenței erifice de program, prin tehnici de predicție;
- utilizarea microprocesoarelor Shered memory în special în cadrul arhitecturii serverelor și stațiilor grafice.
