

Luento 5 (verkkoluento 6)

## Tiedon esitysmuodot

Lukujärjestelmät  
Kokonaisluvut, liukuluvut  
Merkit, merkijonot  
Äänet, kuvat, muu tieto

Ohjelman esitysmuoto  
Rakenteellinen tieto

6.11.2012 Copyright 2012 Teemu Kerola 1

## Tiedon tyypit

- Kommunikointi ihmisen kanssa
  - kuva, ääni, merkit, ...
- Laitteiston sisäinen talletus
  - kuvaformaattit, ääniformaattit, pakkausstandardit, ...
  - kokonaisluvut, liukuluvut, merkit, merkistöt
  - ohjelmat
- Suorittimen omia lajinaan ymmärtämät tyypit
  - on olemassa konekäskyjä tälle tietotyypille
  - kokonaisluvut
  - liukuluvut (useimmat suorittimet nykyään)
  - totuusarvot (jotkut suorittimet)
  - merkit (jotkut suorittimet)
  - konekäskyt

6.11.2012 Copyright 2012 Teemu Kerola 2

## Tiedon esitys

- Kysymys: miten esittää eri tyyppisiä tietoja?
- Vastaus: koodataan ne biteiksi
  - kaikki tieto on koneessa biteinä
- Kaikelle käsitellylle tiedolle on omat koodausmenetelmänsä
  - kaikkia koodausmenetelmiä ei ole standardoitu
  - samalle tietotyypille voi olla useita koodausmenetelmiä
    - kokonaisluvut, liukuluvut, merkit, merkijonot, kuvat, ...
  - ongelma: ymmärtävätkö koneet toisiaan?
    - tiedon esitysmuotoa voidaan joutua muuttamaan, kun tietoa siirretään koneelta toiselle

6.11.2012 Copyright 2012 Teemu Kerola 3

## Tiedon esitys laitteistossa

- Kaikki tieto koneessa on binääribiteinä (0 tai 1)
  - binäärijärjestelmän numerot: 0, 1
  - helppo toteuttaa piireillä
  - helppo suunnitella logiikkaa Boolean algebran avulla
- Muisti jaettu tasapituisiin sanoihin (word)
  - sana = word = 32 bittiä (16 bittiä, 64 bittiä, ...)
- Yleensä sana on jaettu tasapituisiin 8-bittisiin tavuihin (byte)
 

sana

1000 1101	1010 1101	1011 1100	1111 0001
tavu	tavu	tavu	tavu
8 D	A D	B C	F I

6.11.2012 Copyright 2012 Teemu Kerola 4

## Suorittimen ymmärtämä tieto

- Kaikki tieto koneessa on koodattuna biteiksi
- Muistissa voidaan esittää kaikki tieto millä tahansa sovitulla esitystavalla (koodauksella)
- Joillekin esitystavalle on omat konekäskyt (eli suoritin ymmärtää niitä)
  - Kokonaisluvut ja liukuluvut (lähes aina)
  - Totuusarvot, merkit ja merkijonot (joskus) TTK-91: kokonaisluvut
  - Kuvat ja äänet (ei yleensä ellei erikoistunut suoritin)
- Muiden tietojen käsittely tapahtuu ohjelmallisesti (eli usea käsky tai aliohjelma tai metodi)
  - Esim. merkkejä (merkien esitysmuotoa) voidaan käsitellä kokonaislukuoperaatioilla ja aliohjelmillä
  - Rationaaliluvut, 64-bittiset kokonaisluvut, isot taulukot, tietueet, oliot, sormenjäljet, äänet, kuvat, ...

6.11.2012 Copyright 2012 Teemu Kerola 5

## Numeromuunnokset

- Binääriluvuista kymmenjärjestelmään
 

10110011  $\Rightarrow$  179
- Kymmenjärjestelmästä binääriluvuiksi
 

179  $\Rightarrow$  10110011
- Binääriluvusta heksadesimaaliluvuksi
 

10110011  $\Rightarrow$  B3
- Heksadesimaaliluvusta binääriluvuksi
 

B3  $\Rightarrow$  10110011

6.11.2012 Copyright 2012 Teemu Kerola 6

### Big vs. Little Endian

- Miten monitavuiset arvot talletetaan?

0x1200: [ ] [ ] [ ] [ ]  
 Sanan osoite: 0x1200 0x1201 0x1202 0x1203  
 talleta 0x11223344 ?? tavuosoitteet

Big-Endian: eniten merkitsevä tavu pienimpään osoitteeseen  
 → 0x11 0x22 0x33 0x44  
 0x1200 0x1201 0x1202 0x1203

Little-Endian: vähiten merkitsevä tavu pienimpään osoitteeseen  
 → 0x44 0x33 0x22 0x11

6.11.2012 Copyright 2012 Teemu Kerola 7

### Negatiiviset kokonaisluvut

- Etumerkkibitti erikseen
- Yhden komplementtiesitys
- Kahden komplementtiesitys**
- Vakiolisäys

arvo talletus  
 +57 = 0011 1001  
 sign bit = MSB = most significant bit  
 luku -57 = 1011 1001 talletusmuoto  
 -57 = 1100 0110 "sign" bit  
 -57 = 1100 0111 "sign" bit +1  
 -57 = 0100 0110 "sign" bit  
 -57 + 127 = 70  
 +57 = 1011 1000  
 +57 + 127 = 184

6.11.2012 Copyright 2012 Teemu Kerola 8

### IEEE 32-bit FP Standard

sign exponent mantissa or significand  
 "15" "0.1875" = "0.0011"  
 $1/8 = 0.125$   
 $1/16 = 0.0625$   
 $0.1875$

- 23 bittiä mantissalle siten, että ...

Miksi käytetään piilobittia?

- Binääripiste (.) on heti ensimmäisen bitin jälkeen
- Mantissa on normalisoitu: vasemmanpuolimmainen bitti on 1
- Vasemmanpuolimmaista (eniten merkitsevä) bittiä (1) ei talleteta (implied bit, piilobitti)

mantissa eksponentti  
 0.0011 "15"  
 1.1000 "12"  
 1000 "12"  
 24 bitin mantissa!

6.11.2012 Copyright 2012 Teemu Kerola Keskustele 9

### UCS ja Unicode

- UCS - Universal Character Set
- Samat merkistöt, eri standardit
- 2 tavua eli 16 bittiä per merkki
  - 65536 merkkiä koko maailmassa käytössä oleville n. 200000 symbolille
- (32-bit UCS-4 sisältää myös kaikki kiinalaiset merkit)
- Kontrollimerkit
  - 0x0000-001F and 0x0080-009F
  - 0x007F = DELETE, 0x0020 = SPACE
- UCS:ssä myös 8-bittiset koodi "rivit"
  - eri alueille tai tarkoituksiin (zone) omat 8-bittiset koodinsa, esim. UTF-8

6.11.2012 Copyright 2012 Teemu Kerola 10

### Merkkijonot

- Yleensä peräkkäin talletettu joukko tavuja
- Lisäksi tarvitsee jollain tavalla koodata merkkijonon pituus:
  - laitetaan loppuun erikoismerkki
    - C-kieli: '\0' = 0x00
  - toteutetaan tietueena [20] "Ei yleensä nyt enää!"  
 pituus merkkijono
- ei (yleensä) omia konekäskyjä, manipulointi aliohjelmilla
  - kokonaisluku- ja bittimanipulointikäskyt
  - joissakin koneissa "strcpy" ja "strcmp" konekäskyt

6.11.2012 Copyright 2012 Teemu Kerola 11

### Konekäskyjen esitysmuoto

- Konekohtainen, jokaisella omansa
- Käskyt ovat 1 tai useamman tavun mittaisia
  - SPARC, kaikki käskyt: 1 sana eli 4 tavua
  - ARM, kaikki käskyt: 1 sana eli 4 tavua
  - Pentium II: 1-16 tavua, paljon variaatioita
- Käskyillä on yksi tai useampi muoto, kussakin tietty määrä erilaisia kenttiä
  - opcode, Ri, Rj, Rk, osoitusmoodi
  - pitkä tai lyhyt vakio

TTK-91, kaikki käskyt: 1 sana, 1 muoto

6.11.2012 Copyright 2012 Teemu Kerola Keskustele 12

## Taulukkojen esitysmuoto

- Peräkkäisrakenteena, kuten esimerkit aikaisemmin
  - Riveittäin tai sarakkeittain
  - Ei omia konekäskyjä, manipulointi aliohjelmilla tai toistorakenteilla
- Poikkeus: vektorisuorittimet, joilla
- vektorirekisterit (esim. 8 tai 64 liukulukua) tavallisten rekistereiden lisäksi
  - Multimediakäskyt: 64b = 8 \* 8b vektori (Intel MMX)
  - omia konekäskyjä vektorioperaatioita varten
- Indeksoitu tiedonosoitusmoodi tukee 1-ulotteisten taulukoiden käyttöä

6.11.2012

Copyright 2012 Teemu Kerola

Keskustele 13

## Tietueiden esitysmuoto

- Peräkkäisrakenteena
- Osoite on jonkin osoitinmuuttujan arvo
- Ei omia konekäskyjä, manipulointi aliohjelmilla tai kääntäjän generoimien vakioolisäysten avulla
- Indeksoitu tiedonosoitusmoodi tukee tietueiden käyttöä

6.11.2012

Copyright 2012 Teemu Kerola

Keskustele 14

## Olioiden esitysmuoto

- Kuten tietueet, yleensä varattu keosta (heap)
- Useat oliion kentistä sisältävät vuorostaan osoitteen keosta suoritusaikana varattuun toiseen oliioon
- Metodit ovat aliohjelmien osoitteita
- Ei omia konekäskyjä, manipulointi aliohjelmilla

6.11.2012

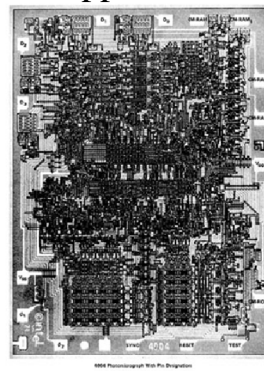
Copyright 2012 Teemu Kerola

15

## -- Luennon 5 loppu --

### Intel 4004, 1971

- Faggin, Hoff, Mazor
- Ens. suoritin lastulla 3x4 mm, \$200
- 2300 transistoria
- 4 bitin sana
- Laskinta varten
- Sama laskentateho kuin Eniacilla (18000 tyhjiöputkea)



6.11.2012

Copyright 2012 Teemu Kerola

16