

Luento 6 Tiedon esitysmuodot

Lukujärjestelmät
Kokonaisluvut
Liukuluvut
Merkit, merkkijonot
Totuusarvot
Kuvat, äänet, hajut(?)

6.4.2005

Copyright Teemu Kerola 2003

1

Tiedon tyypit ⁽³⁾

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

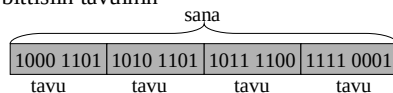
6.4.2005

Copyright Teemu Kerola 2003

2

Tiedon esitys laitteistossa ⁽³⁾

- Kaikki tieto koneessa on binääribitteinä (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)



6.4.2005

Copyright Teemu Kerola 2003

3

Tiedon esitys laitteistossa ⁽²⁾

- Tietoa siirretään muistiväylää pitkin sanoina
 - joskus useampi kuin yksi sana kerrallaan (lohko)
- Suorittimen rekisterit ovat yleensä yhden tai kahden sanan mittaisia
 - 1 sana: kokonaisluku, pieni liukuluku
 - 1 sana: 1 merkki tai 4 merkkiä
 - 2 sanaa: pitkä kokonaisluku, iso liukuluku

6.4.2005

Copyright Teemu Kerola 2003

4

Tiedon esitys ⁽³⁾

- 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
 - samalla tietotyyppille voi olla useita koodausmenetelmiä
 - kokonaisluvut, liukuluvut, merkit, merkkijonot, kuvat, ...
 - ongelma: ymmärtävätkö koneet toisiaan?
 - tiedon esitysmuotoa voidaan joutua muuttamaan, kun tietoa siirretään koneelta toiselle

6.4.2005

Copyright Teemu Kerola 2003

5

Suorittimen ymmärtämä tieto ⁽⁵⁾

- Kaikki tieto koneessa on koodattuna biteiksi
- Muistissa voidaan esittää kaikki tieto millä tahansa sovitulla esitystavalla (koodauksella)
- Suoritin osaa tehdä operaatioita joillakin esitystavoilla koodatuille tiedoille
 - kokonaisluvut ja liukuluvut (aina)
 - totuusarvot, merkit ja merkkijonot (joskus)
 - kuvat ja äänet (ei yleensä ellei erikoistunut suoritin)
 - näytönohjain, äänenohjain, multimediaohjain
 - hajut (ei vielä)
- Muiden tietojen käsittely tapahtuu ohjelmallisesti
 - esim. merkkejä (merkkien esitysmuotoa) voidaan käsitellä kokonaislukuoperaatioilla ja aliohjelmilla

TTK-91:
kokonaisluvut

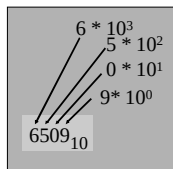
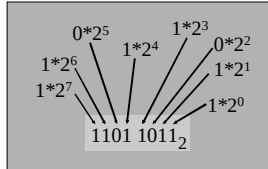
6.4.2005

Copyright Teemu Kerola 2003

6

Binäärijärjestelmä ⁽²⁾

- Kantaluku 2, numerot 0 ja 1
 - numeroiden painoarvot oikealta vasemmalle:
 $1=2^0$, $2=2^1$, $4=2^2$, $8=2^3$, $16=2^4$, $32=2^5$, ...
 - kymmenjärjestelmässä painoarvot ovat
 $1=10^0$, $10=10^1$, $100=10^2$, $1000=10^3$, ...

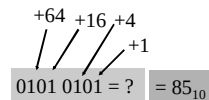
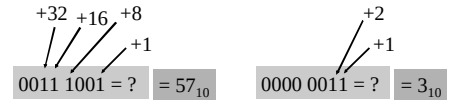


6.4.2005

Copyright Teemu Kerola 2003

7

Binäärilukuesimerkkejä ⁽⁹⁾



6.4.2005

Copyright Teemu Kerola 2003

8

Binäärilukujen laskutoimitukset ⁽³⁾

+	0	1
0	0	1
1	1	10

$$\begin{array}{r} 11 \\ 101101 \\ +1100 \\ \hline 111001 \end{array} \quad \begin{array}{r} 45 \\ +12 \\ =57 \end{array}$$

*	0	1
0	0	0
1	0	1

$$\begin{array}{r} 101 \\ *101 \\ \hline 101 \\ +101 \\ \hline 11001 \end{array} \quad \begin{array}{r} 5 \\ *5 \\ =25 \end{array}$$

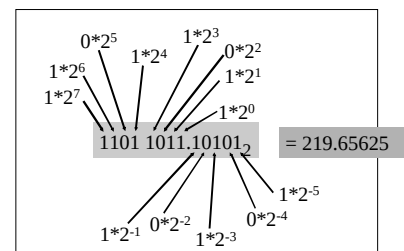
6.4.2005

Copyright Teemu Kerola 2003

9

Binääripiste ⁽³⁾

- Binääriluvuilla voi olla myös binääriosia (vrt. desimaaliosa)

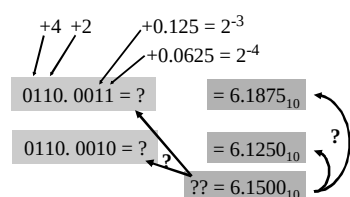
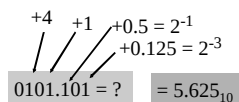


6.4.2005

Copyright Teemu Kerola 2003

10

Binääripiste-esimerkkejä ⁽¹⁰⁾



6.4.2005

Copyright Teemu Kerola 2003

11

Muunnokset lukujärjestelmien välillä ⁽⁵⁾

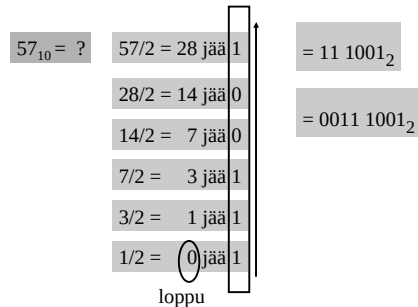
- 2-järjestelmä $\Rightarrow$ 10-järjestelmä
 - esitettiin jo edellä
- 10-järjestelmä $\Rightarrow$ 2-järjestelmä
 - kokonaisosa ja desimaaliosa erikseen
 - kokonaisosa:
 - jaa toistuvasti 2:lla, kunnes 0 jäljellä
 - ota jakojäännökset käännettyssä järjestyksessä

6.4.2005

Copyright Teemu Kerola 2003

12

10-järj $\Rightarrow$ 2-järj kokonaislukuesimerkki ⁽¹¹⁾



6.4.2005

Copyright Teemu Kerola 2003

13

10-järj $\Rightarrow$ 2-järj desimaaliosa $\Rightarrow$ binääriosia ⁽²⁾

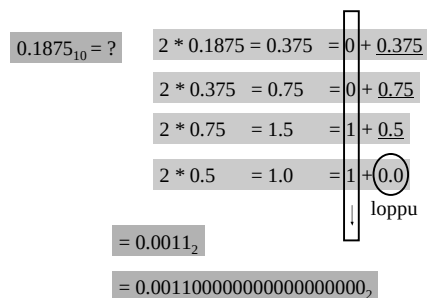
- Kerrotaan toistuvasti desimaaliluvun desimaaliosa 2:lla, kunnes
 - desimaaliosa = 0 (tarkka binääriesitys) **TAI**
 - tarpeeksi numeroita haluttuun tarkkuuteen
- Tulos saadaan ottamalla saatujen desimaalilukujen kokonaisosat (0 tai 1) lasketussa järjestyksessä

6.4.2005

Copyright Teemu Kerola 2003

14

10-järj $\Rightarrow$ 2-järj desimaaliosa $\Rightarrow$ binääriosia ⁽⁹⁾



6.4.2005

Copyright Teemu Kerola 2003

15

Heksadesimaaliesitys

- Binäärilukuja käyttö on tarpeellista, mutta niitä on ikävä kirjoittaa
 - liikaa numeroita
- Kirjoitetaan ne 16-järjestelmässä eli heksadesimaalijärjestelmässä
- 4 bittiä vastaa aina yhtä 16-järjestelmän numeroa
- Yksi 16-järjestelmän numero vastaa aina 4 bittiä
- 16-järjestelmän numerot ovat: 0,1,2,3,4,5,6,7,8,9, A, B, C, D, E ja F

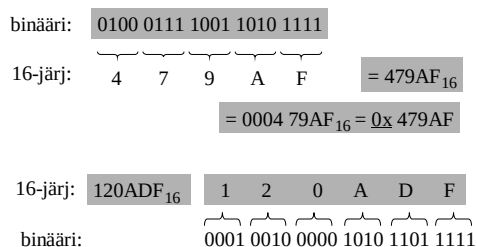
10	11	12	13	14	15
----	----	----	----	----	----

6.4.2005

Copyright Teemu Kerola 2003

16

Heksadesimaaliesimerkkejä ⁽⁷⁾



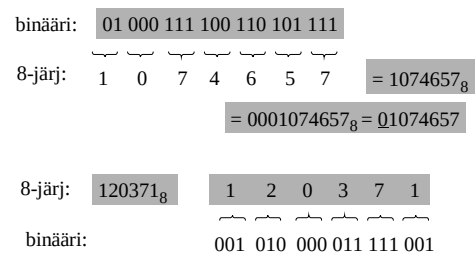
6.4.2005

Copyright Teemu Kerola 2003

17

Oktaaliesimerkkejä ⁽⁷⁾

Numerot: 0, 1, 2, 3, 4, 5, 6, 7



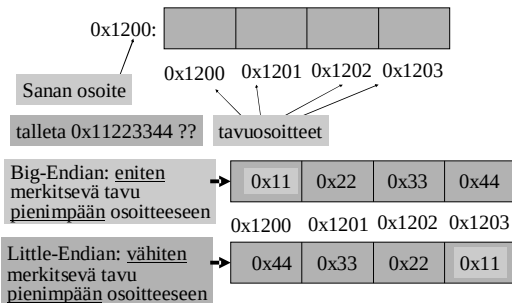
6.4.2005

Copyright Teemu Kerola 2003

18

Big vs. Little Endian (3)

- Miten monitavuiset arvot talletetaan?



6.4.2005

Copyright Teemu Kerola 2003

19

Big vs. Little Endian (5)

- Monitavuisen tiedon (sana-) osoite on sama molemmissa tapauksissa
- Tavujen järjestys on erilainen
- Suorittimen suunnittelija päättää
 - matematiikkapiirin tulee tietää miten luvut esitetty
 - täytyy ottaa huomioon siirrettäessä tietoa verkon yli
- Power-PC: bi-endian - molemmat moodit käytössä
 - voidaan valita ohjelmakohtaisesti
 - etuokeutetussa tilassa voidaan vielä valita erikseen
 - suoritin osaa laskea kummallakin tavalla talletetuilla luvuilla

TTK-91: big-endian

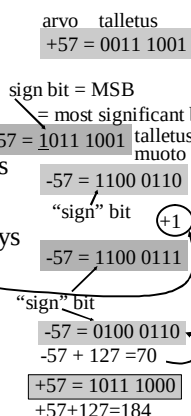
6.4.2005

Copyright Teemu Kerola 2003

20

Negatiiviset luvut (4)

- Etumerkkibitti erikseen
- Yhden komplementtiesitys
- Kahden komplementtiesitys**
- Vakiolisäys
 - Esim lisää 127 ($=2^8 - 1$)
 - yleensä: 2^{bittilkm} -1
 - Talleta etumerkittömänä



6.4.2005

Copyright Teemu Kerola 2003

21

Kahden komplementti (7)

- Useimmiten käytössä
- Etu: vain yksi nolla
 - yhden komplementtissa kaksi nollaa
 - +0 = 0000 0000 -0 = 1111 1111
- Helpot muunnokset: arvo ↔ esitysmuoto
 - miten arvo -57 esitetään?
 - 1100 0110 +1 = 1100 0111
 - mitä arvoa esitysmuoto 1100 0111 tarkoittaa?
 - (0011 1000 +1) = -0011 1001 = -57

6.4.2005

Copyright Teemu Kerola 2003

22

Liukuluvut (3)

- Tietokoneessa ei ole reaalilukuja tai rationaalilukuja (matemaattiset käsitteet)
- Aina rajallinen esityksen tarkkuus
 - lukuja π (pii), $\sqrt{2}$ (2), tai 1/3 ei voi esittää tarkasti
 - esim. luvut 1.000000000 ja luvut 1.000000001 ovat yhtäsuuria (joissakin esityksissä)
- Yleinen reaalilukuja vastaava esitysmuoto on liukukuesitysmuoto
 - float, double, real
 - 32 bittiä, noin 7-8 desimaalinumeron tarkkuus
 - 64 bittiä, noin 16-17 desimaalinumeron tarkkuus

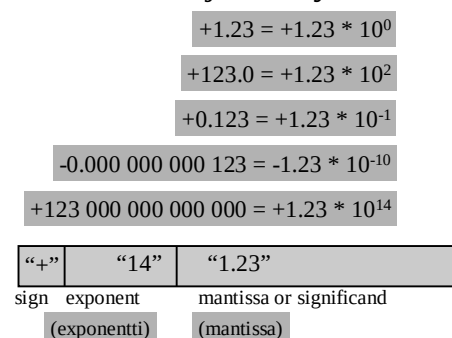
Ks. `accur.c` (out) ja `float.c` (out) vs. `double.c` (out)

6.4.2005

Copyright Teemu Kerola 2003

23

Liukulukujen esitys (5)

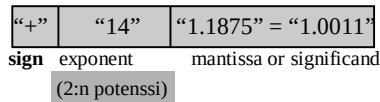


6.4.2005

Copyright Teemu Kerola 2003

24

IEEE 32-bit Floating Point Standard ⁽¹⁾

IEEE
Standard 754

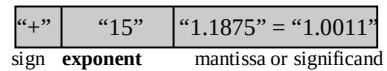
- Etumerkki
 - 1 bitti, ‘+’ koodattu 0:ksi, ‘-’ koodattu 1:ksi
 - etumerkkibitti $S \Rightarrow$ etumerkin arvo = $(-1)^S$
 - “+” merkin eli 0:n arvo on +1
 - “-” merkin eli 1:n arvo on -1

6.4.2005

Copyright Teemu Kerola 2003

25

IEEE 32-bit FP Standard ⁽²⁾



- 8 bittiä eksponentille, lisätynä 127:llä (biased form)



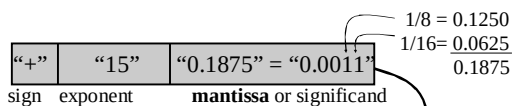
- esitysmuodot 0 ja 255 erikoistapauksia
 - laajennettu arvoalue: hyvin pienet luvut, NaN, $\pm\infty$
- talletettu arvoalue: **1-254** $\Rightarrow$ tod. arvoalue: **-126 ... +127**
(esitysmuoto) (arvoalue)

6.4.2005

Copyright Teemu Kerola 2003

26

IEEE 32-bit FP Standard ⁽⁷⁾

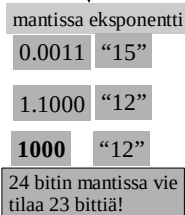


- 23 bittiä mantissalle, siten että ...

1) Binääripiste (.) on heti ensimmäisen bitin jälkeen

2) Mantissa on normalisoitu: vasemmanpuolimmainen bitti on 1

3) Vasemmanpuolimmaista (eniten merkitsevä) bittiä (1) ei talleteta (implied bit, piilobitti)



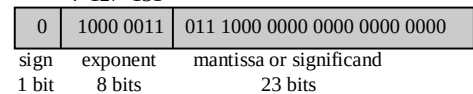
6.4.2005

Copyright Teemu Kerola 2003

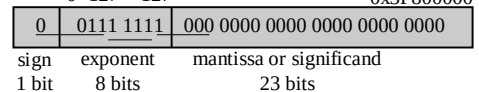
27

IEEE 32-bit FP Values ⁽⁹⁾

$$23.0 = +10111.0 * 2^0 = +1.0111 * 2^4 = ?$$



$$1.0 = +1.0000 * 2^0 = ?$$

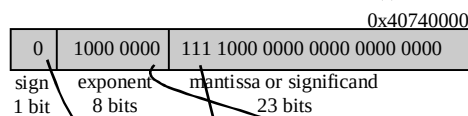


6.4.2005

Copyright Teemu Kerola 2003

28

IEEE 32-bit FP Values ⁽⁶⁾



$$\begin{aligned}
 X &= ? \\
 X &= (-1)^0 * 1.1111 * 2^{(128-127)} \\
 &= 1.1111_2 * 2 \\
 &= (1 + 1/2 + 1/4 + 1/8 + 1/16) * 2 \\
 &= (1 + 0.5 + 0.25 + 0.125 + 0.0625) * 2 \\
 &= 1.9375 * 2 = 3.875
 \end{aligned}$$

6.4.2005

Copyright Teemu Kerola 2003

29

Merkki ⁽⁵⁾

- Yleensä 1 tavu per merkki
- ASCII, 7 bittiä/merkki (+ tark. bitti?)

‘A’ = 0x41, ‘a’ = 0x61, LF = 0x0A

- EBCDIC, 8 bittiä/merkki
- ISO/IEC 8859-15 (‘Latin-9’),
 - 8-bittiä/merkki, 256 eri merkkiä käytössä
 - mukana myös ä, ö, š, €

Lisää tietoa: ks. Kolehmainen
(orig. <http://www.tieke.fi/edisty/edis699/stand699.htm>)

6.4.2005

Copyright Teemu Kerola 2003

30

UCS ja Unicode ⁽⁵⁾

- UCS - Universal Character Set
- Samat merkistöt, eri standardit
- 2 tavua eli 16 bittiä per merkki
 - 65536 merkkiä koko maailmassa käytössä oleville n. 200000 symbolille
- 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

6.4.2005

Copyright Teemu Kerola 2003

31

UCS koodistoja

- Merkit välillä 0x0000-00FF (16 bittiä) samassa järjestyksessä kuin Latin-9 merkistössä (8 bittiä)
 - 16-bittisen UCS:n "rivi 00" = 8-bittinen Latin-9
- Myös muut aakkoset:
 - I-zone = Kanji (0x4E00-9FFF, 20992 merkkiä)

	4E0	4E1	4E2	4E3	4E4	4E5	4E6
0	一	丐	北	丰	乚	乐	习
	4E00	4E10	4E20	4E30	4E40	4E50	4E60
1	丁	丑	兩	卯	乚	豕	乡
	4E01	4E11	4E21	4E31	4E41	4E51	4E61

- Ei omia konekäskyjä, manipulointi aliohjelmilla

6.4.2005

Copyright Teemu Kerola 2003

32

Muita UCS koodistoja

- Hebreä: 0x0590-05FF

	059	05A	05B	05C	05D	05E	05F
0	א	ב	ג	ד	ה	ו	ז
	0590	0591	0592	0593	0594	0595	0596
1	ח	ט	י	כ	ל	מ	נ
	0597	0598	0599	059A	059B	059C	059D

- Arabia: 0x0600-06FF

	060	061	062	063	064	065	066	067
0	ا	ب	ت	ث	ج	ح	خ	د
	0600	0601	0602	0603	0604	0605	0606	0607
1	ذ	ر	ز	س	ش	ص	ض	ط
	0608	0609	060A	060B	060C	060D	060E	060F

- Gujarati: 0x0A80-0AFF

	0A8	0A9	0AA	0AB	0AC	0AD
0	અ	બ	ગ	ઘ	ઙ	ચ
	0A80	0A81	0A82	0A83	0A84	0A85
1	છ	જ	ઝ	ઞ	ટ	થ
	0A86	0A87	0A88	0A89	0A8A	0A8B

6.4.2005

Copyright Teemu Kerola 2003

33

Merkkijonot ⁽⁵⁾

- Yleensä peräkkäin talletettu joukko tavuja
- Lisäksi tarvitsee jollain tavalla koodata merkkijonon pituus
 - laitetaan loppuun erikoismerkki, tai
 - C-kieli: '\0' = 0x00
 - toteutetaan tietueena

20 "Ei yleensä nyt enää!"

 pituus merkkijono
 - ei omia konekäskyjä, manipulointi aliohjelmilla
 - kokonaisluku- ja bittimanipulointikäskyt
 - joissakin koneissa "strcpy" ja "strcmp" konekäskyt

6.4.2005

Copyright Teemu Kerola 2003

34

Totuusarvot ⁽⁵⁾

- Boolean TRUE ja FALSE
- Yleensä koodattu TRUE=1, FALSE=0
 - muttei aina!
 - Totuusarvolauseke **A and B** = kokon.lukulauseke **A*B**
- Usein Boolean arvo per sana
 - loput 31 bittiä nollia
 - ohjelmointikielten Boolean muuttujat
- Joskus pakatussa muodossa 32 arvoa per sana
- Ei omia konekäskyjä, manipulointi aliohjelmilla
 - kokonaisluku- ja bittimanipulointikäskyt
 - haluttu konekäsky "**JTRUE ...**" voidaan toteuttaa konekäskynä "**JPOS ...**" (jos TRUE = 1)

6.4.2005

Copyright Teemu Kerola 2003

35

Kuvat

- Monta kuvastandardia
 - yleisyys, siirrettävyys, pakkaustiheys
 - näyttöä varten tarvittavan laskennan määrä
- Kuvatiedoston alussa otsake kertoo talletusformaatin
- Viiva- ja vektorikuvat
 - kuva koodattuna objekteina
 - ympyrä, monikulmio, käyrä, alueen väri
- Rasterikuvat
 - kuva koodattuna pisteinä
 - kunkin pisteen väri koodattu esim. 24 bitillä

6.4.2005

Copyright Teemu Kerola 2003

36

Kuvat

- Kuvat ovat yleensä pakattu mahdollisimman vähän tilaa vievää muotoon
 - optimoitu tilan, ei laskennan mukaan
 - purkaminen voi vaatia paljon laskentaa
- GIF, JPEG, TIFF, BMP, ...
- Ei omia konekäskyjä, manipulointi aliohjelmilla

6.4.2005

Copyright Teemu Kerola 2003

37

Videokuva

- Vie hyvin paljon muistitilaa
- Talletus kuva kerrallaan, esim. 25 kuvaa/sek
 - 1 sekunti hyvälaatuista videokuva pakkaamattomassa muodossa 20 MB
- Talletus ”incrementaalisesti”
 - kun seuraava kuva poikkeaa edellisestä vain vähän ...
 - talleta vain muutokset edelliseen

6.4.2005

Copyright Teemu Kerola 2003

38

Videostandardit

- MPEG (Moving Pictures Expert Group)
- AVI (Audio Visual Interleave)
- MOV, INDEO, FLI, GL, DVD, ...
- Ei omia konekäskyjä, manipulointi aliohjelmilla tai ...
- Erikoisprosessoreilla (GPU), joiden käskykanta suunniteltu (jonkin standardin mukaisten kuvien) kuvankäsittelyyn
 - grafiikkakorteilla

6.4.2005

Copyright Teemu Kerola 2003

39

Grafiikkakortit

- Esim. 4-64 MB VRAM (dual-port) muistia ...
 - 2 lukua/kirjoitusta samanaikaisesti
- ... tai ”tavallista”, mutta hyvin nopeaa RAMia
- Nopea väylä (ennen PCI, nyt AGP) suorittimelle
- Näytönohjaus monitoristandardien (VGA, DVI, XGA, RGB, ...) mukaisesti
- Oma suoritin (GPU)
 - lukee videodataa ja generoi näytettävän kuvan näyttöpuskuriin, josta monitori sen näyttää
- Voi olla integroitu emolevyn kanssa



6.4.2005

Copyright Teemu Kerola 2003

40

Äänet

- Täydellinen äänidata
 - 44100 näytettä/sek, 16 b/näyte, 88KB /sek
- Syntetisoitu ääni
 - MIDI-käskyjä
 - Music Instrument Digital Interface
 - ”Soita nuotti N voimakkuudella V”
- Ei omia konekäskyjä, manipulointi aliohjelmilla tai ...
- Erikoisprosessoreilla, joiden käskykanta suunniteltu äänen käsittelyyn
 - äänikortit

6.4.2005

Copyright Teemu Kerola 2003

41

Äänikortit

- Esim. 4-64 MB VRAM tai RAM muistia
- Nopea väylä (esim. PCI) suorittimelle
- Oma suoritin, joka lukee äänidataa ja generoi äänet kaiuttimille tai vahvistimeen
 - kaiuttimet tai vahvistin kiinni äänikortilla
- Voi olla integroitu emolevyn (tai grafiikkakortin) kanssa



6.4.2005

Copyright Teemu Kerola 2003

42

Maku, haju, tunto ja muu data ⁽³⁾

- Tähtien kirkkaus, hajut, ks. HS artikkeli 5.5.2000
veneen tyyppi, tunteen palo, ...
- Toteutus sovelluskohtaisesti, ei vielä yleisiä standardeja
 - kokonaisluvut (diskreetti data)
 - liukuluvut (jatkuva data)
- Ei omia konekäskyjä, manipulointi omilla aliohjelmilla

6.4.2005

Copyright Teemu Kerola 2003

43

-- Luennon 6 loppu --



6.4.2005

Copyright Teemu Kerola 2003

44