§8.2-8.4: Data Storage and Number Bases

- *announcements*

18 Oct 2006
CMPT14x
Dr. Sean Ho
Trinity Western University

What's on for today (§8.2-8.4)

- Number bases:
  - Binary, hexadecimal (0xBEEF), octal (0115)
- Bitwise operators: &, |, ^, <<, >>
- Units of measure of memory:
  - Bits, nibbles, bytes, words, pages
- Units of measure for hard disks:
  - C/H/S geometry
- SI units vs binary units, KB vs. Kb, etc.

Using bases in Python

- Python has special notation for expressing integer literals in hexadecimal and octal:
  - Hexadecimal: prefix "0x"

    hexNum = 0xBEEF # $11(16^3)+14(16^2)+14(16^1)+15 = 48879$
  - Octal: prefix "0"

    octNum = 0115 # $1(8^2) + 1(8^1) + 5(8^0) = 77$
- Convert into strings with hexadecimal/octal notation:

```
hexStr = hex(48879)     # '0xbeef'
octStr = oct(77)        # '0115'
```

Bitwise operators on Python ints

- Bitwise and: &

  6 & 3 == 2 # 110_2 & 011_2 == 010_2

- Bitwise or: |

  6 | 3 == 7 # 110_2 | 011_2 == 111_2

- Bitwise xor (exclusive or): ^

  6 ^ 3 == 5 # 110_2 ^ 011_2 == 101_2

- Bitwise right shift: >>, left shift: <<

  6 >> 2 == 1 # 110_2 >> 2 == 001_2

  6 << 2 == 24 # 110_2 << 2 == 11000_2

Bits, bytes, nibbles, words

- One hexadecimal digit can be represented by four bits: one nibble
- Two nibbles (eight bits) is called a byte
  - One byte can be used to store one CHAR
- A group of bytes can be used to represent one datum: this is called a word
  - Pentium CPUs generally use 4-byte words (32 bits)
  - Newer CPUs can use 8-byte words (64 bits)
  - Word is the smallest unit of data the machine can store or retrieve

Accessing memory

- A computer's main memory (generally, RAM) stores everything it needs to do its current tasks
- A location within memory is uniquely identified by its address
  - Most modern CPUs use 32-bit words to store memory addresses
  - This means there is a maximum of 2^{32} unique memory addresses (the address space)
  - If each location stores one byte of data, then there is 2^{32} bytes = 4GB of addressable memory

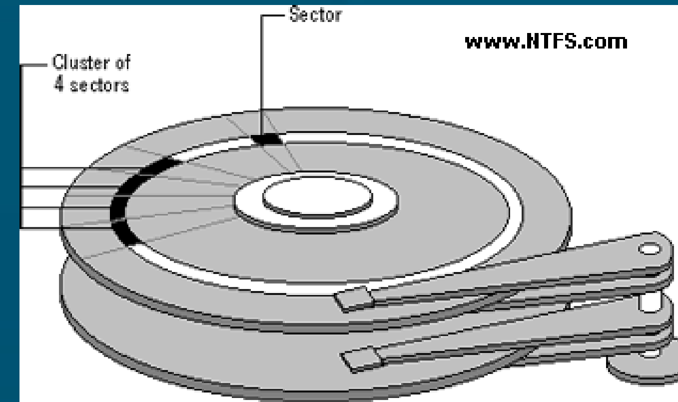
Units of measure

- SI abbreviations:
  - K = kilo = 1,000
  - M = mega = 1,000,000
  - G = giga = 1,000,000,000
- When working with binary data:
  - KB = kilobyte = 1,024 bytes = 2^{10} bytes
  - MB = megabyte = 1,024,576 = 2^{20} bytes
  - GB = gigabyte = 1,073,741,824 = 2^{30} bytes
  - But hard drive manufacturers use SI abbrevs

Units of measure, cont.

- Kilobytes vs. kilobits:
  - KB = kilobyte = 1,024 bytes = 8192 bits
  - Kb = kilobit = 1,024 bits
  - RAM chip manufacturers often use kilobits
- Also, in SI abbreviations,
  - M = mega = 10^6: e.g., megawatt = 10^6 watt
  - m = milli = 10^{-3}: e.g., milliwatt = 10^{-3} watt
- But not everyone is consistent, so be careful

Storage



- A page of memory is generally 256 bytes
- A sector is a unit of disk storage, also commonly 256 bytes (but sometimes 512 bytes)
- A block of disk storage is usually 512 bytes
- Hard disks are made up of platters, accessed by magnetic heads on movable arms
- The platters have concentric tracks that (across all heads) make up cylinders
- Hard drive geometry is often expressed in C/H/S: # cylinders / # heads / # sectors per track

Summary of today (§8.2-8.4)

- Number bases:
  - Binary, hexadecimal (0xBEEF), octal (0115)
- Bitwise operators: &, |, ^, <<, >>
- Units of measure of memory:
  - Bits, nibbles, bytes, words, pages
- Units of measure for hard disks:
  - C/H/S geometry
- SI units vs binary units, KB vs. Kb, etc.

TODO items

- HW07 due Fri: Py ch9 #5
  - Also, write your own pseudorandom number generator, and
  - Create a histogram using your own pseudorandom, another histogram using the built-in random(), and compare
- Quiz06 (ch7-8) on Fri
- CMPT140 Final next week: W-Th 25-26Oct