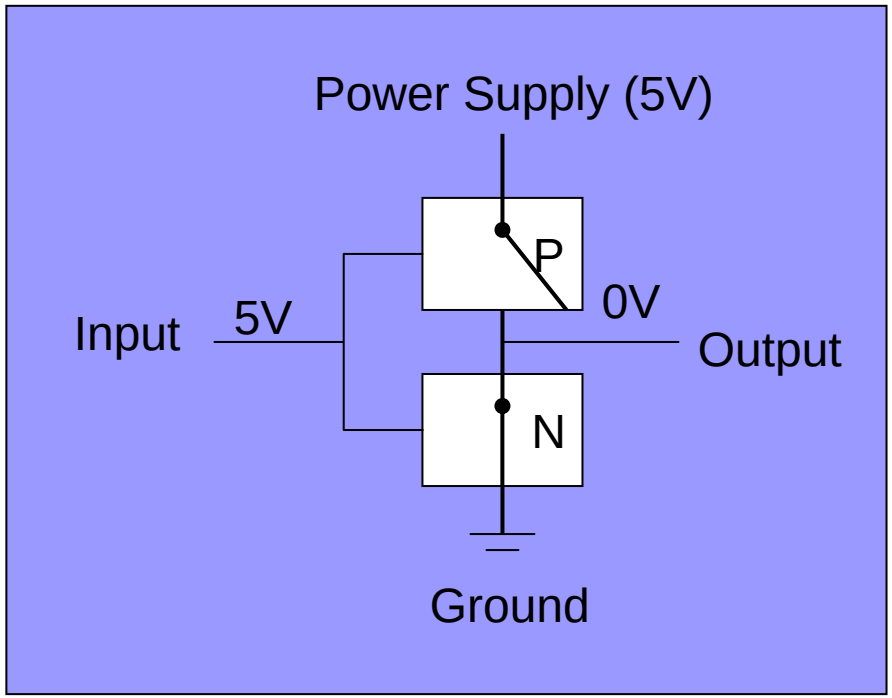
Lecture 4: Data representation, storage devices

Bit

A Not Gate Example



- A bit can only have two potential values: **0** or **1**.
- A bit refers to a digit in binary numerable system.
- A bit is the smallest discrete storage unit.
- Implementations:
  - A bit corresponds to the voltage level of the input or output: 5v (1), and 0v (0).
  - A bit corresponds to the state of a switch: on (1), and off (0).
  - etc.

Byte, KB, MB, GB

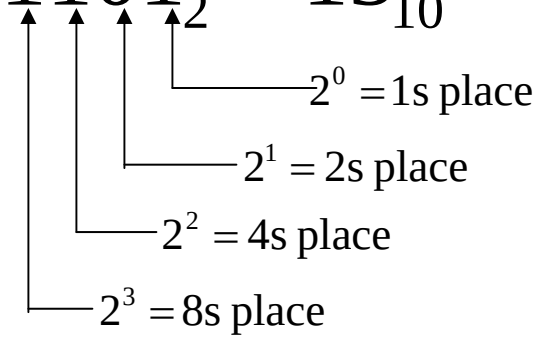
- Byte: a collection of 8 bits
  - E.g. 1101 1010

    Or 1010 0110
- KB (killobyte) = 1024 bytes
- MB (metabyte) = 1024 KBs
- GB (gigabyte) = 1024 MBs

Binary numbers

$$13_{10} = 1 \times 10^1 + 3 \times 10^0$$

$$1101_2 = 13_{10}$$



$$\begin{aligned} 1101_2 &= 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \\ &= (8 + 4 + 0 + 1)_{10} \\ &= 13_{10} \end{aligned}$$

- Decimal number
  - Base 10
  - Each digit corresponds to escalating powers of 10.
- A binary number
  - Base 2.
  - Each digit corresponds to escalating powers of 2.

Representing Integers as bit patterns

- Integers are stored as bit patterns
  - 19 as 10011
  - 116 as 1110100
- Fixed-width
  - 32-bit width: 19 as 0000 0000 0000 0000 0000 0000 0001 0011 (with many leading 0s)
- Number of representable integers = number of bit patterns
  - 32 bit width representation has 2^{32} =4,294,967,296 bit patterns, namely 4,294,967,296 integers

Two-complement notation of integers

- Sign bit (usually the highest bit of the representation)
  - 0 means that this is a positive integer
  - 1 means that this is a negative integer
- The representation
  - Positive integer s is represented as its corresponding binary number
  - Negative integer s is represented as the corresponding binary number of 2^32 - |s|
- Examples (32-bit width integer):
  - 3: **0**000 0000 0000 0000 0000 0000 0000 0011
  - -1: **1**111 1111 1111 1111 1111 1111 1111 1110 = 2^32 – 1
- Why two-complement notation?
  - Avoid to two 0s: positive zero and negative zero
  - Simplify the binary addition

Scientific notation and floating-point numbers

- Scientific notation of decimals
  - the fractional part: the digits surrounding the decimal place
  - The exponent: (the power of 10)
- Example: 1234.56 = 1.23456 * 10^3
- Floating-point numbers (fractional part, exponent)
  - E.g. 1234.56 is represented by a pair (1.23456, 3)

Binary representation of real number

- Binary-based Floating-point number s = (fractional part m, exponent e) , then the value of s = m * 2^e
  - m is the two-complement representation of the fractional part
  - e is the two-complement representation of the exponent part
- Fixed-width
  - Single precision (32-bit) : 1bit for sign, 8 bits for exponent, 23 bits for fractional part
  - Double precision (64-bit): 1bit for sign,11 bits for exponent, 52 bits for fractional part
- Example: 1110110.101 = 1.110110101 × 2^6
  - m will be 11011010100000000000000 (zeros are filled at the end to have 23 bits)
  - e will be 00000110 (6 in decimal), but in the practice it is 10000101 (6 + 127) which is out of the this course's scope.
- Round-off:
  - There are infinite number of real values but only a finite number of 32- or 64-bit patterns.
  - The real value will be round off to the pattern with closest value.
- Round-off examples
  - Single-precision floating-point number: roughly 7 decimal digits of precision
  - Double-precision floating-point number: roughly 16 decimal digits of precision

Representing Characters and Strings

- Character is encoded by binary bit patterns
- ASCII (American Standard Code for Information Interchange) codes
  - ASCII maps each character to a specific 8-bit pattern
- Unicode
  - Unicode maps each character to a specific 16-bit pattern
  - For non-English language with more than 256 characters

The ASCII Character Set

- Link: http://en.wikipedia.org/wiki/ASCII

Representing Images

- Bitmap
  - partitions an image into a grid of dots, called **pixels**
  - represents each pixel with bit pattern
- Pixel bit pattern
  - 1bit Black-and-white image
    - 1 for white
    - 0 for black
  - 24-bit color image (R, G, B)
    - R: 0-255 intensity of red
    - G: 0-255 intensity of green
    - B: 0-255 intensity of blue
- Resolution: The sharpness or clarity of an image
  - E.g. It can be measured with how many pixels per square inch in an image

RGB color maps

http://en.wikipedia.org/wiki/Web_colors

Image formats: BMP, GIF, JPEG, etc.

- BMP (bitmaps)
  - Simple
  - But require lot of memory
- GIF (Graphics Interchange Format) – a lossless format
  - Identify repetitive patterns in the image
  - Store those patterns only once with their markers
  - Replace the occurrence of such patterns with their markers
- JPEG (Joint Photographic Experts Group) – a lossy format
  - The compression is not fully reversible: e.g. compress several neighboring pixels by storing their average color value
- Usages
  - GIF format is used for line drawings and other images with discrete boundaries
  - JPEG format is mostly used for photographs

Representing Sound

- Each particular sound produces a pressure wave with
  - **Unique** amplitude (height, usually measured in pascals)
  - **Unique** frequency (duration over time)
- Digital sampling
  - Measure amplitudes of the sound wave at **regular intervals**
  - Store the amplitudes as sequence of discrete measurements in binary numbers
- CD-quality sound
  - 44,100 amplitude measurements per second
  - 16-bit representation of the amplitude

ASCII file (text file)

- A document that contains only plain text (such as notepad file, html file) is called **ASCII** file or a **text** file.
- The size of a file = number of bytes stored in the file
- Size of ASCII file = number of characters stored in the file

Binary file

- Files that contain data that is not plain text (e.g. image, sound, word document, etc.) are called binary files.
- The size of a file = number of bytes stored in the file
- Examples
  - An ASCII file with 1000 words is of size roughly 5,000bytes = **5KB** assuming average 5 characters per word.
  - Most graphics on web are over **30KB**.

Storage devices

- Floppy disks (1.44MB, 1.2MB, etc.): http://en.wikipedia.org/wiki/Floppy_disk
- Zip drives (100MB, 250MB, 750MB): http://en.wikipedia.org/wiki/Zip_drive
- Jaz drives (1GB, 2GB): http://en.wikipedia.org/wiki/Jaz_drive
- Hard drives (0 – 500GB, or more): http://en.wikipedia.org/wiki/Hard_disk
- CDs (650MB): http://en.wikipedia.org/wiki/Compact_disc
- DVDs(4.7GB): http://en.wikipedia.org/wiki/DVD

Speed of data transmission

- Bps – bits per second
- Modems
  - 28.8k – 28,800 bps
  - 56K -56,000 bps

Summary

- Bit
- Binary numbers
- Two-complement representation
- Representing integers, real numbers, characters, images, sounds
- File types: ASCII, Binary
- Size of data and file
- Storage devices