**Representing numbers in Base 10**

We are accustomed to base 10. What does 2510 mean?

* The 2 means
* The 5 means
* So 25 is really
* What about 00025?

In base 10, what symbols do we have?

**Representing values on a computer: Base 2**

Everything stored in a computer’s memory is really just

A number is NOT stored as text, i.e. 25 is NOT stored as '2' and '5’

Computer Data: stored using bits

i.e. in base 2 instead of base 10

* In base 2, there are \_\_\_\_\_ symbols:
* Base 2 is also called

Counting in binary:

So the values represent:

Let’s convert 2510 to binary:

Is there a mathematical technique to shorten the process?

2510 = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2  10010 = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_2

Convert the following numbers to base 10:

* 000001002
* 000100012
* 000110112

**Negative numbers in binary**

Signed binary: leftmost bit is sign

example with 8-bit integers: 3810

-3810

*Is there a biggest positive number we can store using signed 8 bits?*

*Overflow***:** an error that results from trying to store a number with insufficient

*Suppose we use 16 bits instead of 8?*

Why can our variables store the numbers they store?

|  |  |  |  |
| --- | --- | --- | --- |
| Type | Size in Bytes | Minimum Value | Maximum Value |
| byte | 1 | -128 | 127 |
| short | 2 | -32,768 | 32,767 |
| int | 4 | -2147483648 | 2147483647 |
| long | 8 | -9.223372e+18 | 9.223372e+18 |

Float/Double are stored differently in scientific notation – we aren’t going to worry about it

**Codes for Text**

Characters are stored in bits as well

How many bytes? Max value?

Unicode – 16 bits (2 bytes) so possible codes

* Why do we need so many?

**Information Theory -** How we store and share data

A. How many different messages can be encoded by a certain number of bits?

* 1 bit? Possible encodings:
* 2 bits? Possible encodings:
* 4 bits? Possible encodings:

If I have *x* bits, I have \_\_\_\_\_\_\_ possible encodings.

B. How many different bits do I need to store a certain number of messages?

* if I have 2 different messages, I need at least bit(s)
* if I have 4 different messages, I need at least bits
* if I have 7 different messages, I need at least bits

If I have *m* possible encodings, I need \_\_\_\_\_\_\_\_\_\_\_\_ bits

C. How much "information" is in a particular message? (Does the word "cat" contain more info than the word "the"?)

*Higher frequency of occurrence should mean information content.*

* ex: If the letter 'a' accounts for 25% of a text file then "a" has probability p = .25, and we say 'a' has

**bi bits of information** where

bi = Γ log 2 ( 1/ p ) ˥=

* **Remember:** *p is always a fraction between 0 and 1 inclusive*
* In the message “a cat has a head” how much information does
  + the letter ‘t’ contain?
  + the letter ‘a’ contain?