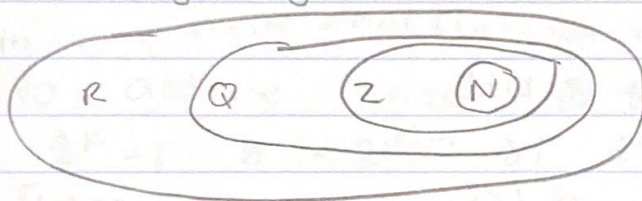


## on the nature of numbers

- representation of numbers
  - we use the hindu-arabic numeral system
  - base 10 because we have 10 fingers
  - babylonians use base 10
  - natural numbers ( $\mathbb{N}$ )
    - ↳ positive int
  - integers ( $\mathbb{Z}$ )
    - ↳ all  $\#$ s
  - rational nums ( $\mathbb{Q}$ )
    - ↳ ratio  $\frac{a}{b}$ ,  $a, b \in \mathbb{Z}$
  - irrational numbers ( $\mathbb{R}$ )
  - imaginary numbers / complex ( $\mathbb{C}$ )
- } all diff sets



Computers only work on a finite set of nums  
digit for a computer = a bit  
↳ a bit can be 0 or 1.

- signed: include negative
- unsigned: 0 - long number
- every num can be written as a polynomial



unsigned-char - 8 bits

unsigned-int  $\rightarrow$  short+long/long long

can be signed

• Char

• unsigned char  
int

Unsigned int

short int

Unsigned

Name

Size

char

usually 8 bits

short

usually 16

int

at least 16

long

at least 32

long long

at least 64

# include

<stdint.h>

signed

unsigned

Size

int8\_t

uint8\_t

8 bit

int16\_t

uint16\_t

16 bits

int32\_t

uint32\_t

32 bits

int64\_t

uint64\_t

64 bits

For loops that won't overflow = use int



- Outside of range with overflow
- remember 32 bits in a word!

## binary arithmetic

- Just like normal arithmetic

$$0 + 0 = 0$$

$$1 + 0 = 1$$

$$1 + 1 = 10, \text{ } 0 \text{ carry the } 1$$

$$0 \times 0 = 0$$

$$1 \times 0 = 0$$

$$1 \times 1 = 1$$

$$101 + 11 = 1000$$

$$101 \times 101 = 11001$$

in a finite field (fix set numbers)

do additive inverse to be its inverse

$$2^k - 1 \quad \& \quad -2^{k-1}$$

## Two's Complement Arithmetic

0000

0000

, flip the numbers, then

0001

1111

add one to the

0010

1110

result

0011

1101

0100

1100

0101

1011

0110

1010

0111

1001



## Real numbers

-R

- Continuous
- Uncountably infinite
- Same number of infinite numbers
- You cannot write down real numbers, only rational, golden ratio,  $\pi$ , or  $e$ .

floating point numbers are not real number  
↳ proper subset of real, rational, and int sets

- They're an approximation.
- a double is at least as precise as a float

float	single	IEEE 754 single	32
-------	--------	-----------------	----

double	double	IEEE 754 double	64
--------	--------	-----------------	----

long double	extended precision	IEEE 754 long double	128
-------------	--------------------	----------------------	-----

$$\frac{2}{5} = 0.4 = 0 \times 10^{-1}$$

$$\frac{1}{3} = 0.333 \dots$$

-> review binary fraction



## Single precision

128 power and -128 power

fraction 23 bits

first one = sign

$$1 + \sum_{i=1}^{23} b_{23-i} 2^{-i}$$

The square root of 2 is not rational

## Intel extended precision

1st bit = sign

15 bit = exponent

1 bit = integer part

64 bit = mantissa

40 or 112 bit

## Big endian, little endian

least significant bit is

little endian = low address byte

big endian = high address byte

big endian      little endian

1-8

7 6 5 4 3 2 1

## Random numbers

Computers cannot create random numbers

## Arithmetic Operators

x mult

/ divide

% modulo - integers only / ~~last~~ remainder

+ addition - subtraction



- follows pandas

- mixed types

- the lower type get promoted to a higher type

x %

→ research vers

+

left

< >

left

< < > >

left

= =

left

8

left

^

left

1

left

2 2

left

1 1

left

2 %

right

= + =

- =

\* =

/ = % =

right