The Representaion of Data

An Overview

Prof. David Bernstein
James Madison University

Computer Science Department

bernstdh@jmu.edu

A Quiz

What Is This?

```
01101001011010110100
10000101101001010011
00010110011010110010
11101001001100101001
10000101110100010110
```

The Point of the Quiz

- It could be many things

  Bits (i.e., binary values) can be and are used to represent a wide variety of things

- We need contextual information

  To interpret a bunch of bits we need to know the representation scheme being used

Why 0/1?

- Electronic/Magnetic Systems:

  Positive/Negative

  On/Off

  Clockwise/Counterclockwise

- Mechanical Systems:

  Up/Down

  Pits/Lands

  Hole/Solid

  Bump/Flat

An Easy First Example - The Counting Numbers

Base 10 (Decimal)				Base 2 (Binary)			
	10^2	10^1	10^0	2^3	2^2	2^1	2^0
	100	10	1	8	4	2	1
0	0	0	0	0	0	0	0
0	0	0	1	0	0	0	1
0	0	0	2	0	0	1	0
0	0	0	3	0	0	1	1
		.				.	
		.				.	
0	0	1	0	1	0	1	0
0	0	1	1	1	0	1	1
0	0	1	2	1	1	0	0
0	0	1	3	1	1	0	1

An Interesting Question and Answer

- The Question:

  How many bits do we need to represent all of the counting numbers less than N?

- Getting to the Answer:

  With B bits, we can represent all of the counting numbers less than 2^B

What About Negative Integers?

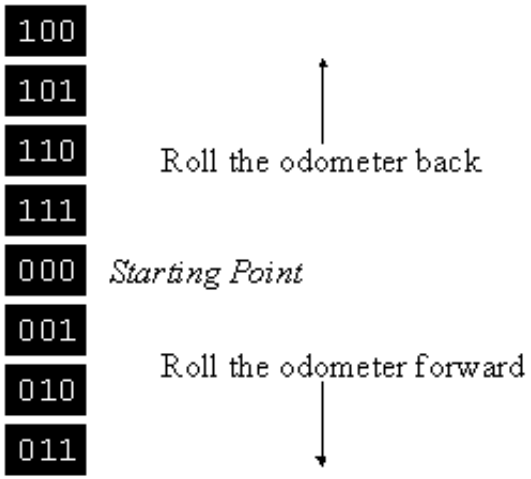
- An Obvious Place to Start:

  Since there are two signs, use one bit (e.g., the left-most) to represent the sign

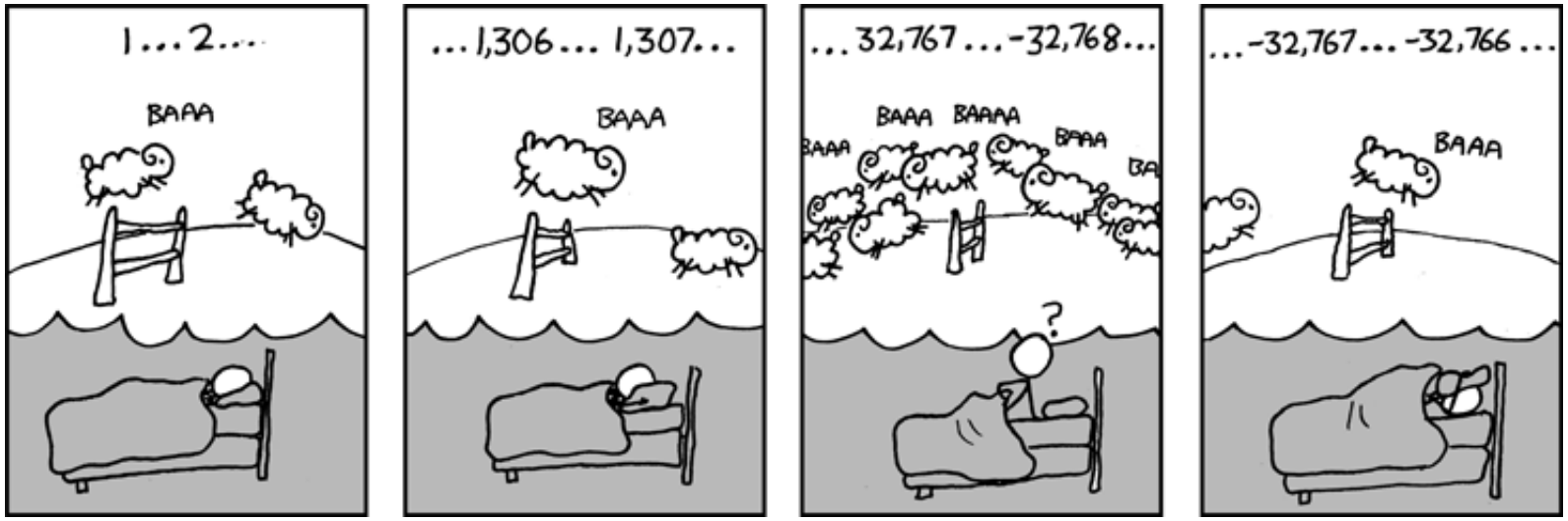
- A Shotcoming of this Approach:

  It results in both a +0 and a -0

Going Further: Negative Numbers (cont.)



Going Further: Negative Numbers (cont.)



(Courtesy of xkcd)

What About Real Numbers?

- Think About Base 10:

  The positions to the left of the decimal point are powers of 10 and the positions to the right of the decimal place are powers of 1/10

- An Obvious Place to Start in Binary:

  The positions to the left of the decimal point are powers of 2 and the positions to the right of the decimal place are powers of 1/2

Going Further: What About Real Numbers? (cont.)

- Terms:
  - Sign
  - Exponent
  - Mantissa
- Normalization:
  - One digit left of the decimal
- Example: $+1.101101 \times 2^3$
  - Sign: +
  - Exponent: 3
  - Mantissa: 1.101101

Going Further: What About Real Numbers? (cont.)

- IEEE Short Real (Single Precision):

  1 bit for the sign

  8 bits for the exponent

  23 bits for the mantissa

- IEEE Long Real (Double Precision):

  1 bit for the sign

  11 bits for the exponent

  52 bits for the mantissa

What About Characters?

- An Obvious Place to Start:

  Count the number of characters

  Determine the number of bits needed

  Assign a binary number to each character

- An Example:

  There are 26 letters in the alphabet

  8 bits can represent 2^5 (i.e., 32) different things

  Assign 00001 to A, 00010 to B, 00011 to C, ..., 11010 to Z

What About Characters? (cont.)Motivation

- The ASCII Encoding:

Character	Decimal	Binary
!	33	00100001
A	65	01000001
B	66	01000010
C	67	01000011
a	97	01100001
b	98	01100010
c	99	01100011

- Unicode:

  A mapping for every character in every language (including many dead languages)

What About Other Things?

- Discrete Sets:

  We can use the same approach as for characters

- Continuous Sets:

  We either have to sample the set (to create a discrete approximation) or describe the elements

What About Colors?

- A Sampling Scheme:

  Use the fact that we have red, green and blue cones to think of colors as having a red, green and blue component

  Think of each color as having a discrete number of levels (e.g., $2^8 = 256$)

  A paletteof $2^24 = 16,777,216$ colors

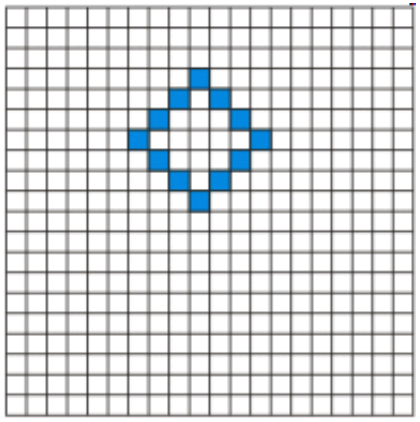
- A Description Scheme:

  Use the wavelength

What About Pictures?

- A Sampling Scheme:

  Create a finite grid (called a raster) with equal sized cells (called picture elements or pixels)



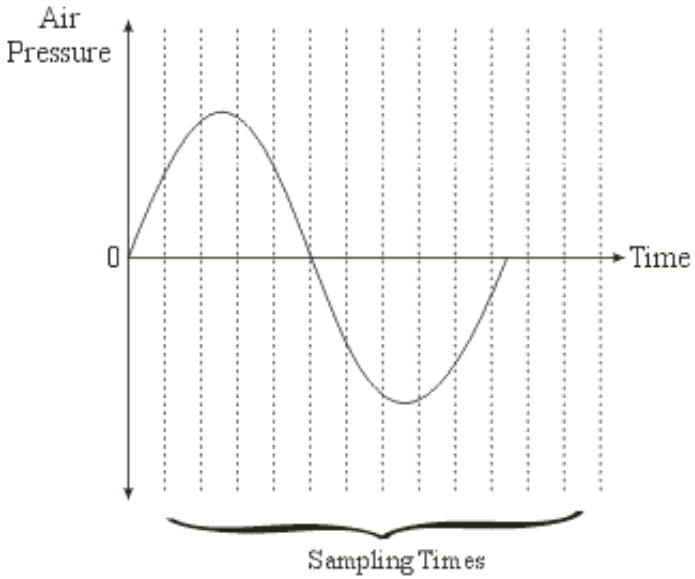
- A Description Scheme:

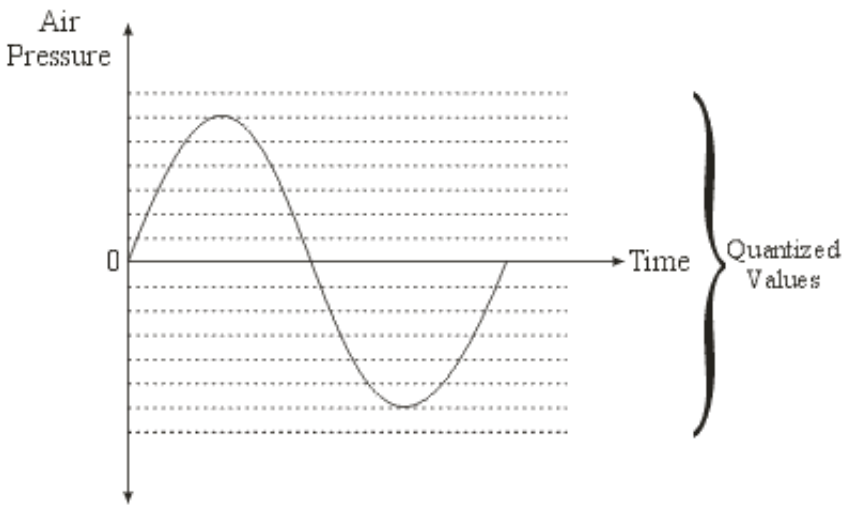
  Use geometric shapes (e.g., points, lines, curves, rectangles, polygons, ellipses)

What About Audio?

- A Sampling Scheme:

  Need to use both temporal sampling and amplitude sampling (called quantization)





- A Description Scheme:

  Use something like standard musical notation

What About Programs?

- Getting Started:

  Each processor is capable of executing a discrete set of operations

  Each operation is given a code

- The Next Step:

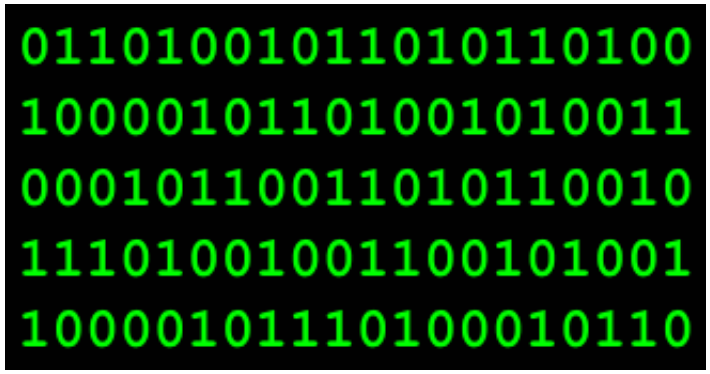
  Each operation has a discrete number of operands, each of which is represented in binary

- The Final Step:

  A program is just a sequence operation codes and operand values

The Quiz Revisited

- What Is This?



- It Could Be Anything!

  I could treat it as a number or bunch of numbers

  I could treat it as a color or bunch of colors

  I could treat it as a audio

  I could treat it as a program

...