Introduction to Computing
Section 2 – The Information Layer

Binary Values and Number Systems

Data Representation

MAINLY DISCUSS ABOUT DATA TYPES

PART 2

Data Representation

Data and Computers

SCSE

Data basic values or facts

Information data that has been organized or processed in a useful manner

Data and Computers

Computers are **multimedia** devices,dealing with a vast array of information categories

Computers store, present, and help us modify

- Numbers
- Text
- Audio
- Images and graphics
- Video

Data and Computers

SCSE

Data compression

Reduction in the amount of space needed to store a piece of data.

Bandwidth

The number of bits or bytes that can be transmitted from one place to another in a fixed amount of time.

Data and Computers

SCSE

Compression ratio

The size of the compressed data divided by the size of the original data

A data compression techniques can be

- **lossless**, which means the data can be retrieved without any loss of the original information
- **lossy**, which means some information may be lost in the process of compaction

Analog and Digital Information

Computers are finite!

How do we represent an infinite world?

We represent enough of the world to satisfy our computational needs and our senses of sight and sound

Analog and Digital Information

Data can be represented in one of two ways: **analog** or **digital**

Analog data

A continuous representation, analogous to the actual information it represents

Digital data

A discrete representation, breaking the information up into separate elements

Analog and Digital Information

Computers cannot work well with analog data, so we digitize the data

Digitize

Breaking data into pieces and representing those pieces separately

Why do we use binary to represent digitized data?

Electronic Signals

Important facts about electronic signals

- An analog signal continually fluctuates in voltage up and down
- A digital signal has only a high or low state, corresponding to the two binary digits
- All electronic signals (both analog and digital) degrade as they move down a line
- The voltage of the signal fluctuates due to environmental effects

Electronic Signals *(Cont'd)*

Periodically, a digital signal is **reclocked** to regain its original shape

Figure 3.2
An analog and a digital signal

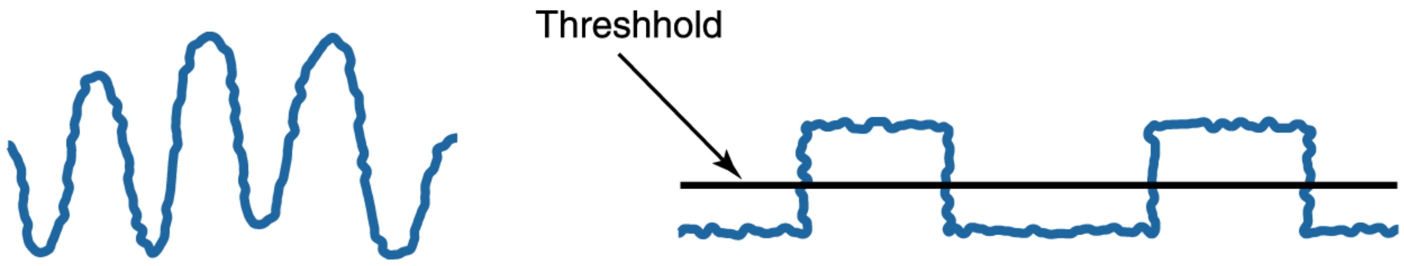


Figure 3.3
Degradation of analog and digital signals

Binary Representations

One bit can be either 0 or 1

One bit can represent two things *(Why?)*

Two bits can represent four things *(Why?)*

How many things can three bits represent?

How many things can four bits represent?

How many things can eight bits represent?

Binary Representations

1 Bit	2 Bits	3 Bits	4 Bits	5 Bits
0	00	000	0000	00000
1	01	001	0001	00001
	10	010	0010	00010
	11	011	0011	00011
		100	0100	00100
		101	0101	00101
		110	0110	00110
		111	0111	00111
			1000	01000
			1001	01001
			1010	01010
			1011	01011
			1100	01100
			1101	01101
			1110	01110
			1111	01111
				10000
				10001
				10010
				10011
				10100
				10101
				10110
				10111
				11000
				11001
				11010
				11011
				11100
				11101
				11110
				11111

SCSE

Counting with binary bits

Figure 3.4

Binary Representations

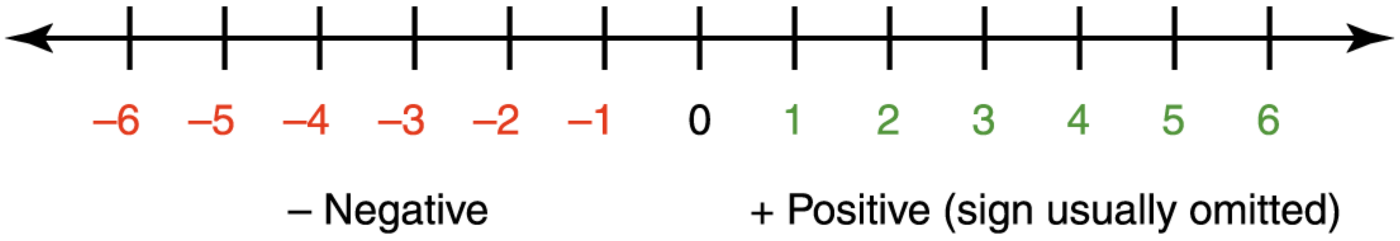
How many things can n bits *represent?*

Why?

What happens every time you increase the number of bits by one?

Representing Negative Values

Signed-magnitude number representation



Signed-magnitude representation

Number representation in which the sign represents the ordering of the number (negative and positive) and the value represents the magnitude

Representing Negative Values

There is a problem with the sign-magnitude representation: *Can you guess why?*

There is a plus zero and minus zero, which causes unnecessary complexity

Solution:

Keep all numbers as integer values, with half of them representing negative numbers

Representing Negative Values

Using two decimal digits,

let 1 through 49 represent 1 through 49

let 50 through 99 represent -50 through -1

Representing Negative Values

To perform addition, add the numbers and discard any carry

Signed-Magnitude	New Scheme
5 + − 6 − 1	5 + 94 99
− 4 + 6 2	96 + 6 2
− 2 + − 4 − 6	98 + 96 94

Now you try it

48 (signed-magnitude)
\- 1
47

How does it work in the new scheme?

Representing Negative Values

A-B=A+(-B)

Add the negative of the second to the first

Signed-Magnitude	New Scheme	Add Negative
–5	95	95
– 3	– 3	+ 97
–8		92

Try

4	*- 4*	*-4*
- 3	*+3*	*+ -3*

Representing Negative Values

Formula to compute the negative representation of a number

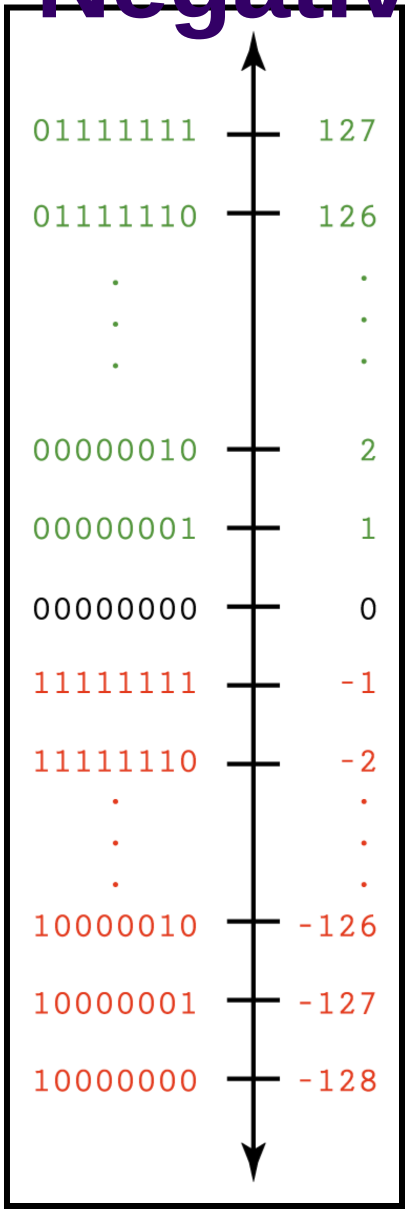
$$\text{Negative}(I) = 10^k - I, \text{ where } k \text{ is the number of digits}$$

This representation is called the ten's complement

Representing Negative Values

Two's Complement

(Vertical line is easier to read)



Representing Negative Values

Addition and subtraction are the same as in 10's complement arithmetic

-127	10000001
+ 1	00000001
-126	10000010

Do you notice something interesting about the left-most bit?

Number Overflow

What happen if the computed value won't fit?

Overflow

If each value is stored using eight bits, adding 127 to 3 overflows

```
   1111111
+  0000011
  10000010
```

Problems occur when mapping an infinite world onto a finite machine!

Representing Real Numbers

Real numbers

A number with a whole part and a fractional part

104.32, 0.999999, 357.0, and 3.14159

Positions to the right of the decimal point are the tenths position: 10^{-1}, 10^{-2} , 10^{-3} …

Representing Real Numbers

SCSE

Same rules apply in binary as in decimal

Decimal point is actually the **radix point**

Positions to the right of the radix point in binary are

2^{-1} (one half),

2^{-2} (one quarter),

2^{-3} (one eighth)

…

Representing Real Numbers

SCSE

A real value in base 10 can be defined by the following formula

$$\text{sign} * \text{mantissa} * 10^{\text{exp}}$$

The representation is called **floating point** because the number of digits is fixed but the radix point floats

Representing Real Numbers

Real Value	Floating-Point Value
12001.00	$12001 * 10^{0}$
–120.01	$-12001 * 10^{-2}$
0.12000	$12000 * 10^{-5}$
–123.10	$-12310 * 10^{-2}$
155555000.00	$15555 * 10^{3}$

A binary floating-point value is defined by the formula

$$\text{sign} * \text{mantissa} * 2^{\text{exp}}$$

Representing Real Numbers

Scientific notation

A form of floating-point representation in which the decimal point is kept to the right of the leftmost digit

12001.32708 is 1.200132708E+4 in scientific notation

What is 123.332 in scientific notation?

What is 0.0034 in scientific notation?

Representing Text

What must be provided to represent text?

There are finite number of characters to represent, so list them all and assign each a binary string

Character set

A list of characters and the codes used to represent each one

Computer manufacturers agreed to standardize

The ASCII Character Set

ASCII stands for American Standard Code for Information Interchange

ASCII originally used seven bits to represent each character, allowing for 128 unique characters

Later extended ASCII evolved so that all eight bits were used

How many characters could be represented?

The ASCII Character Set

Right	*ASCII*									
Left Digit(s) / *Digit*	*0*	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT
1	LF	VT	FF	CR	SO	SI	DLE	DC1	DC2	DC3
2	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS
3	RS	US	□	!	“	#	$	%	&	’
4	(	)	*	+	,	−	.	/	0	1
5	2	3	4	5	6	7	8	9	:	;
6	<	=	>	?	@	A	B	C	D	E
7	F	G	H	I	J	K	L	M	N	O
8	P	Q	R	S	T	U	V	W	X	Y
9	Z	[	\	]	^	_	`	a	b	c
10	d	e	f	g	h	i	j	k	l	m
11	n	o	p	q	r	s	t	u	v	w
12	x	y	z	{	|	}	~	DEL		

The ASCII Character Set

The first 32 characters in the ASCII character chart do not have a simple character representation to print to the screen

What do you think they are used for?

The Unicode Character Set

Extended ASCII is not enough for international use

Unicode uses 16 bits per character

How many characters can UNICODE represent?

Unicode is a superset of ASCII

The first 256 characters correspond exactly to the extended ASCII character set

The Unicode Character Set

Code (Hex)	Character	Source
0041	A	English (Latin)
042F	Я	Russian (Cyrillic)
0E09	ฉ	Thai
13EA	Ꮺ	Cherokee
211E	℞	Letterlike Symbols
21CC	⇌	Arrows
282F	⠯	Braille
345F	㑟	Chinese/Japanese/Korean (Common)

Figure 3.6 A few characters in the Unicode character set

Text Compression

Assigning 16 bits to each character in a document uses too much file space

We need ways to store and transmit text efficiently

Text compression techniques

- keyword encoding
- run-length encoding
- Huffman encoding

Keyword Encoding

Replace frequently used words with a single character

Word	Symbol
as	^
the	~
and	+
that	$
must	&
well	%
these	#

Keyword Encoding

Given the following paragraph,

> We hold these truths to be self-evident, that all men are created equal, that they are endowed by their Creator with certain unalienable Rights, that among these are Life, Liberty and the pursuit of Happiness. ㄙ That to secure these rights, Governments are instituted among Men, deriving their just powers from the consent of the governed, ㄙ That whenever any Form of Government becomes destructive of these ends, it is the Right of the People to alter or to abolish it, and to institute new Government, laying its foundation on such principles and organizing its powers in such form, as to them shall seem most likely to effect their Safety and Happiness.

Keyword Encoding

SCSE

The encoded paragraph is

We hold # truths to be self-evident, $ all men are created equal, $ ~y are endowed by ~ir Creator with certain unalienable Rights, $ among # are Life, Liberty + ~ pursuit of Happiness. — $ to secure # rights, Governments are instituted among Men, deriving ~ir just powers from ~ consent of ~ governed, — $ whenever any Form of Government becomes destructive of # ends, it is ~ Right of ~ People to alter or to abolish it, + to institute new Government, laying its foundation on such principles + organizing its powers in such form, ^ to ~m shall seem most likely to effect ~ir Safety + Happiness.

Keyword Encoding

What did we save?

Original paragraph

656 characters

Encoded paragraph

596 characters

Characters saved

60 characters

Compression ratio

596/656 = 0.9085

Could we use this substitution chart for all text?

Run-Length Encoding

A single character may be repeated over and over again in a long sequence

Replace a repeated sequence with

- a flag character
- repeated character
- number of repetitions

*n8

- * is the flag character
- n is the repeated character
- 8 is the number of times n is repeated

Run-Length Encoding

SCSE

Original text

bbbbbbbbjjjkllqqqqqq+++++

Encoded text

*b8jjjkll*q6*+5 *(Why isn't l encoded? J?)*

The compression ratio is 15/25 or .6

Encoded text

*x4*p4l*k7

Original text

xxxxpppplkkkkkkk

This type of repetition doesn't occur in English text; can you think of a situation where it might occur?

Huffman Encoding

Why should the character “X" and "z" take up the same number of bits as "e" or " "?

Huffman codes use variable-length bit strings to represent each character

More frequently used letters have shorter strings to represent them

Huffman Encoding

ballboard would be
1010001001001010110001111011

compression ratio

28/56

Encode roadbed

Huffman Code	Character
00	A
01	E
100	L
110	O
111	R
1010	B
1011	D

Huffman Encoding

In Huffman encoding no character's bit string is the prefix of any other character's bit string

To decode

- look for match left to right, bit by bit
- record letter when a match is found
- begin where you left off,going left to right

Huffman Encoding

Try it!

Huffman Code	Character
00	A
01	E
100	L
110	O
111	R
1010	B
1011	D

Decode

1011111001010

Huffman Encoding

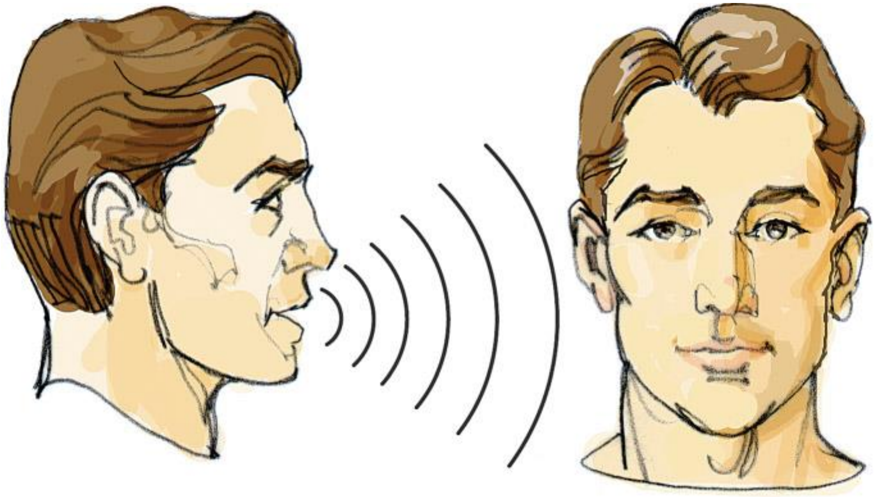
Technique for determining codes guarantees the prefix property of the codes

Two types of codes

- general, based on use of letters in English (Spanish, ….)
- specialized, based on text itself or specific types of text

Representing Audio Information

SCSE



We perceive sound when a series of air compressions vibrate a membrane in our ear, which sends signals to our brain

Representing Audio Information

SCSE

A stereo sends an electrical signal to a speaker to produce sound

This signal is an analog representation of the sound wave

The voltage in the signal varies in direct proportion to the sound wave

Representing Audio Information

Digitize the signal by sampling

- periodically measure the voltage
- record the numeric value

How often should we sample?

A sampling rate of about 40,000 times per second is enough to create a reasonable sound reproduction

Representing Audio Information

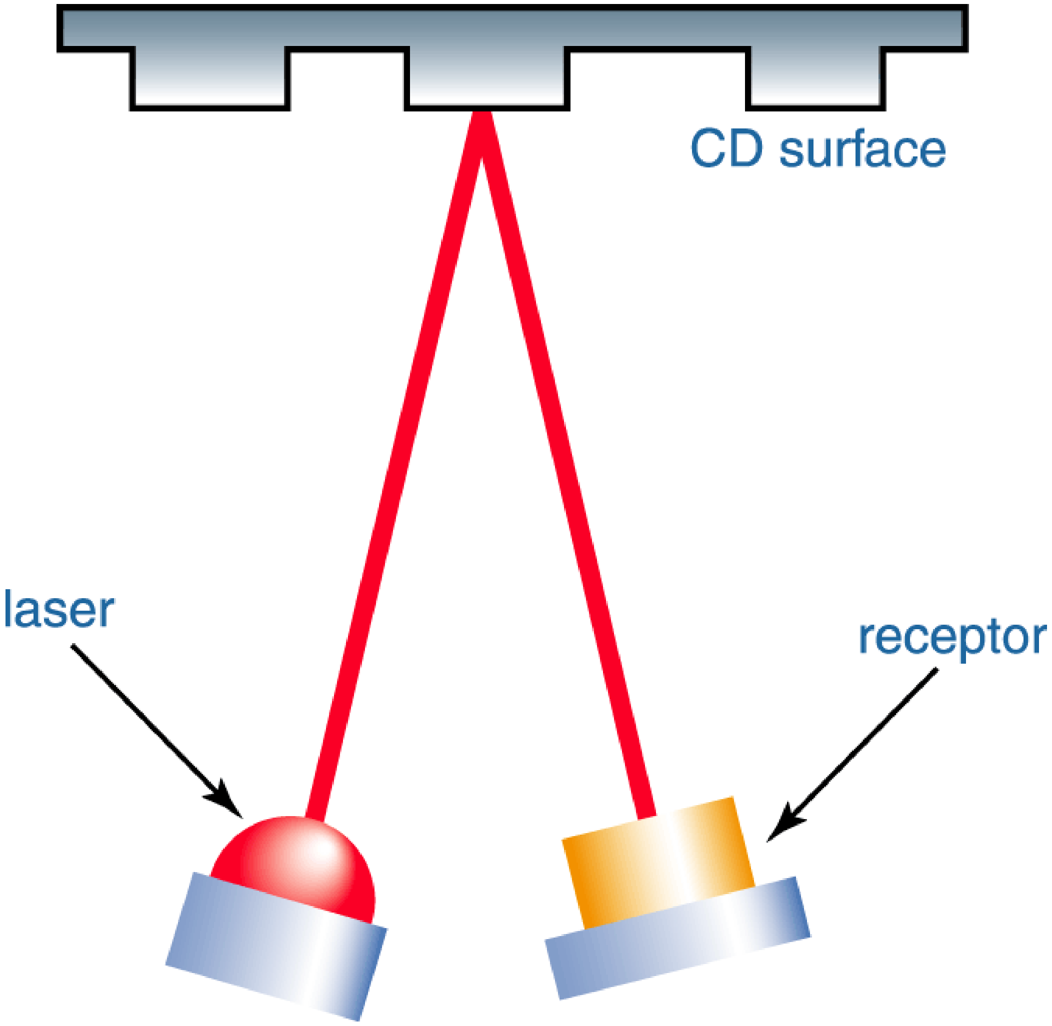
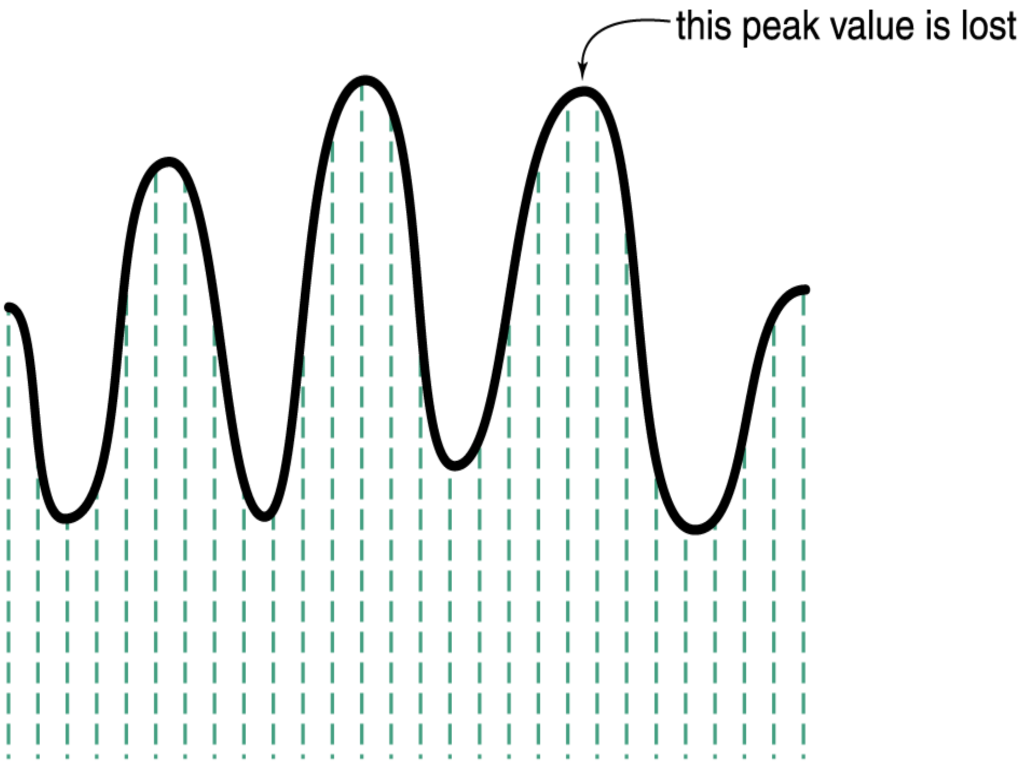


Figure 3.9
A CD player reading binary information

Representing Audio Information



Some data is lost, but a reasonable sound is reproduced

Figure 3.8 Sampling an audio signal

Representing Audio Information

CDs store audio information digitally

On the surface of the CD are microscopic pits that represent binary digits

A low intensity laser is pointed as the disc The laser light reflects

strongly if the surface is smooth and poorly if the surface is pitted

Audio Formats

Audio Formats

- WAV, AU, AIFF, VQF, and MP3

MP3 (MPEG-2, audio layer 3 file) is dominant

- analyzes the frequency spread and discards information that can't be heard by humans
- bit stream is compressed using a form of Huffman encoding to achieve additional compression

Is this a lossy or lossless compression (or both)?

Representing Images and Graphics

SCSE

Color

Perception of the frequencies of light that reach the retinas of our eyes

Retinas have three types of color photoreceptor cone cells that correspond to the colors of red, green, and blue

Representing Images and Graphics

SCSE

Color is expressed as an RGB (red-green-blue) value--three numbers that indicate the relative contribution of each of these three primary colors

An RGB value of (255, 255, 0) maximizes the contribution of red and green, and minimizes the contribution of blue, which results in a bright yellow

Representing Images and Graphics

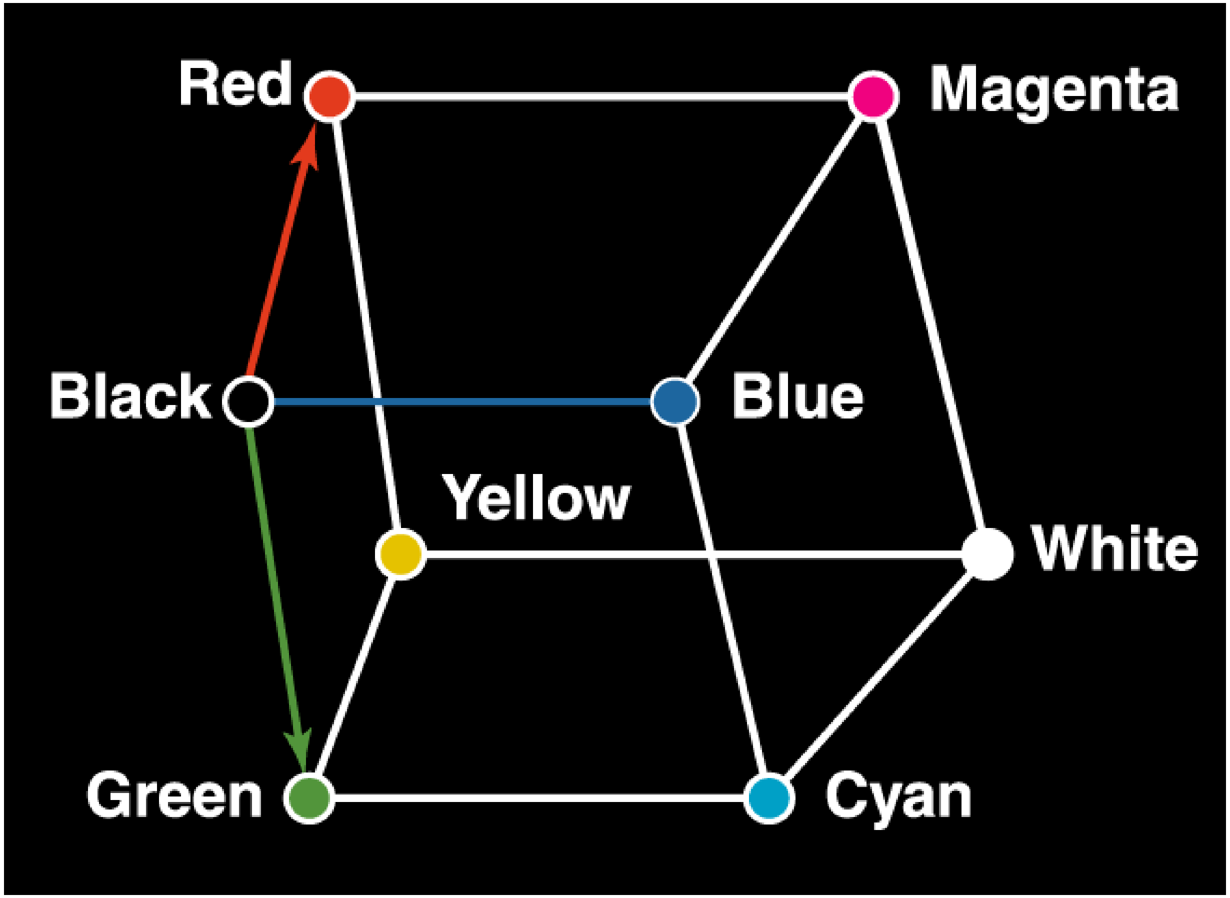


Figure 3.10 Three-dimensional color space

Representing Images and Graphics

color depth

The amount of data that is used to represent a color

HiColor

A 16-bit color depth: five bits used for each number in an RGB value with the extra bit sometimes used to represent transparency

TrueColor

A 24-bit color depth: eight bits used for each number in an RGB value

Representing Images and Graphics

RGB Value			Color
Red	Green	Blue	
0	0	0	black
255	255	255	white
255	255	0	yellow
255	130	255	pink
146	81	0	brown
157	95	82	purple
140	0	0	maroon

A few TrueColor RGB values and the colors they represent

Indexed Color

A browser may support only a certain number of specific colors, creating a palette from which to choose

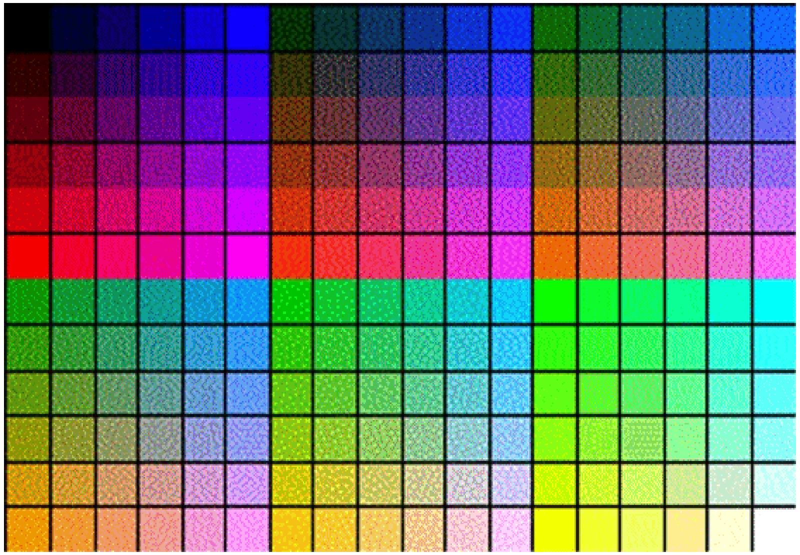


Figure 3.11
The Netscape color palette

Digitized Images and Graphics

Digitizing a picture

Representing it as a collection of individual dots called pixels

Resolution

The number of pixels used to represent a picture

Raster Graphics

Storage of data on a pixel-by-pixel basis

Bitmap (BMP), GIF, JPEG, and PNG are raster-grahics formats

Digitized Images and Graphics

SCSE

Bitmap format

Contains the pixel color values of the image from left to right and from top to bottom

GIF format (indexed color)

Each image is made up of only 256 colors

JPEG format

Averages color hues over short distances

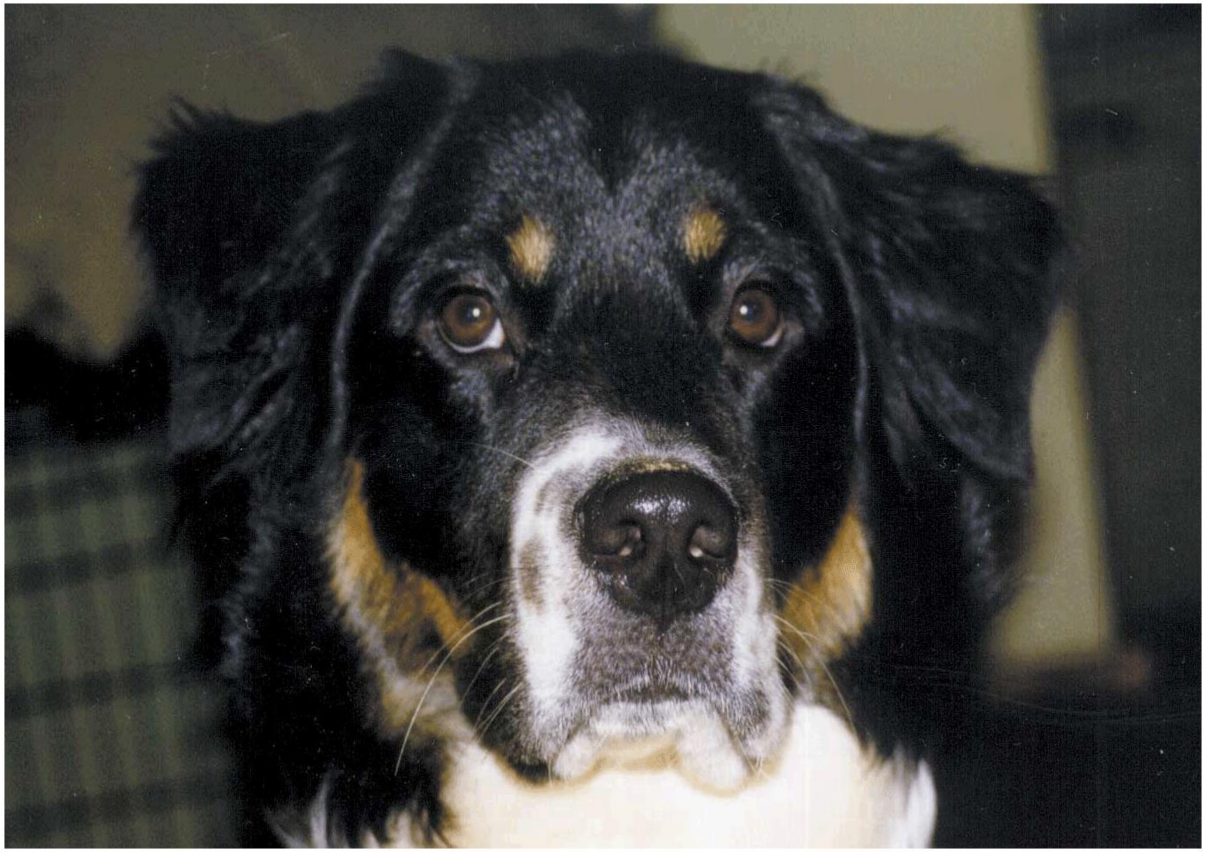
PNG format

Like GIF but achieves greater compression with wider range of color depths

Which is better for line drawings? Pictures?

Digitized Images and Graphics

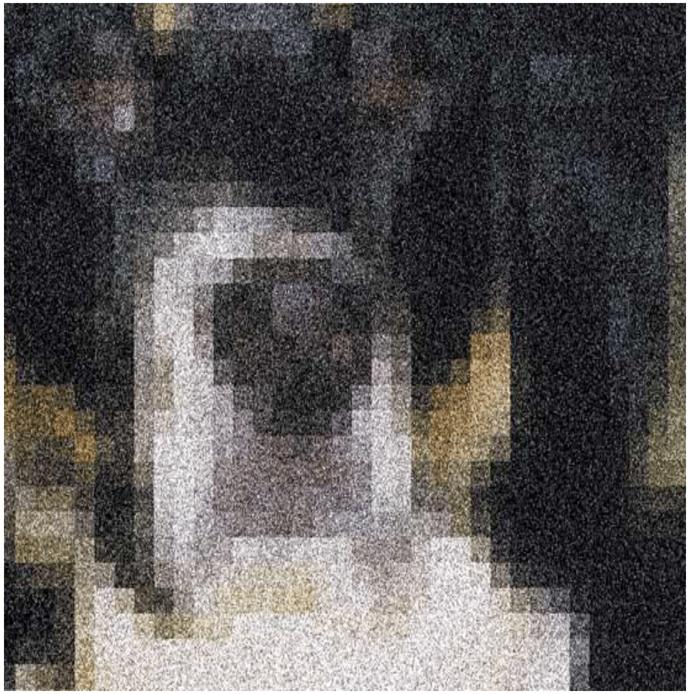
SCSE



Whole picture

Figure 3.12 A digitized picture composed of many individual pixels

Digitized Images and Graphics



Magnified portion of the picture

See the pixels?

Figure 3.12 A digitized picture composed of many individual pixels

Vector Graphics

Vector graphics

A format that describes an image in terms of lines and geometric shapes

A vector graphic is a series of commands that describe a line’s direction, thickness, and color

The file sizes tend to be smaller because not every pixel is described

Vector Graphics

The good side and the bad side…

Vector graphics can be resized mathematically and changes can be calculated dynamically as needed

Vector graphics are *not* good for representing real-world images

Representing Video

SCSE

Video codec COmpressor/DECompressor

Methods used to shrink the size of a movie to allow it to be played on a computer or over a network

Almost all video codecs use lossy compressions to minimize the huge amounts of data associated with video

Representing Video

Temporal compression

A technique based on differences between consecutive frames: If most of an image in two frames hasn't changed, why should we waste space to duplicate all of the similar information?

Spatial compression

A technique based on removing redundant information within a frame: This problem is essentially the same as that faced when compressing still images