1.4 Representing Information as Bit Patterns

Representing Text

- **Each character (letter, punctuation, etc.) is assigned a unique bit pattern.**
  - ASCII: Uses patterns of 7-bits to represent most symbols used in written English text
  - Unicode: Uses bit patterns to represent the major symbols used in languages world side

Unicode编码了地球上大部分地区的文字

西文字符：ACSII码

(American Standard Code for Information Interchange)

用7位二进制编码，最高位0

问题：为什么用7位？

0~127共可表示128个字符

‘A’～ ‘Z’ 26

‘a’～’ z’ 26

‘0’～’ 9’ 10

其他键盘字符、控制键

<=128

0～32、127为非图形字符，其余94个图形字符

Printable ASCII Codes

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
2	space	!	"	#	$	%	&	'	(	)	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	DEL

- Examples:
  - ASCII code for space character = 20 (hex) = 32 (decimal)
  - ASCII code for 'L' = 4C (hex) = 76 (decimal)
  - ASCII code for 'a' = 61 (hex) = 97 (decimal)

The message “Hello.” in ASCII

01001000	01100101	01101100	01101100	01101111	00101110
H	**e**	**l**	**l**	**o**	**.**

Exercise

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
2	space	!	"	#	$	%	&	'	(	)	*	+	,	-	.	/
3	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
4	@	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
5	P	Q	R	S	T	U	V	W	X	Y	Z	[	\	]	^	_
6	`	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o
7	p	q	r	s	t	u	v	w	x	y	z	{	|	}	~	DEL

- 01000011 01001000 01001001 01001110 01000001

Unicode and UTF-8

- Unicode: Uses bit patterns to represent the major symbols used in languages world side
- UTF-8 : (Universal Character Set + Transformation Format—8-bit)

UNICODE 在存储,传输时可以选择具体的传输编码. 即UTF , 常见的有: UTF-8,UTF-16, UTF-32

UTF-8是针对Unicode的一种可变长度字符编码。它可以用来表示Unicode标准中的任何字符，而且其编码中的第一个字节仍与ASCII相容，它逐渐成为电子邮件、网页及其他存储或传送文字的应用中优先采用的编码。

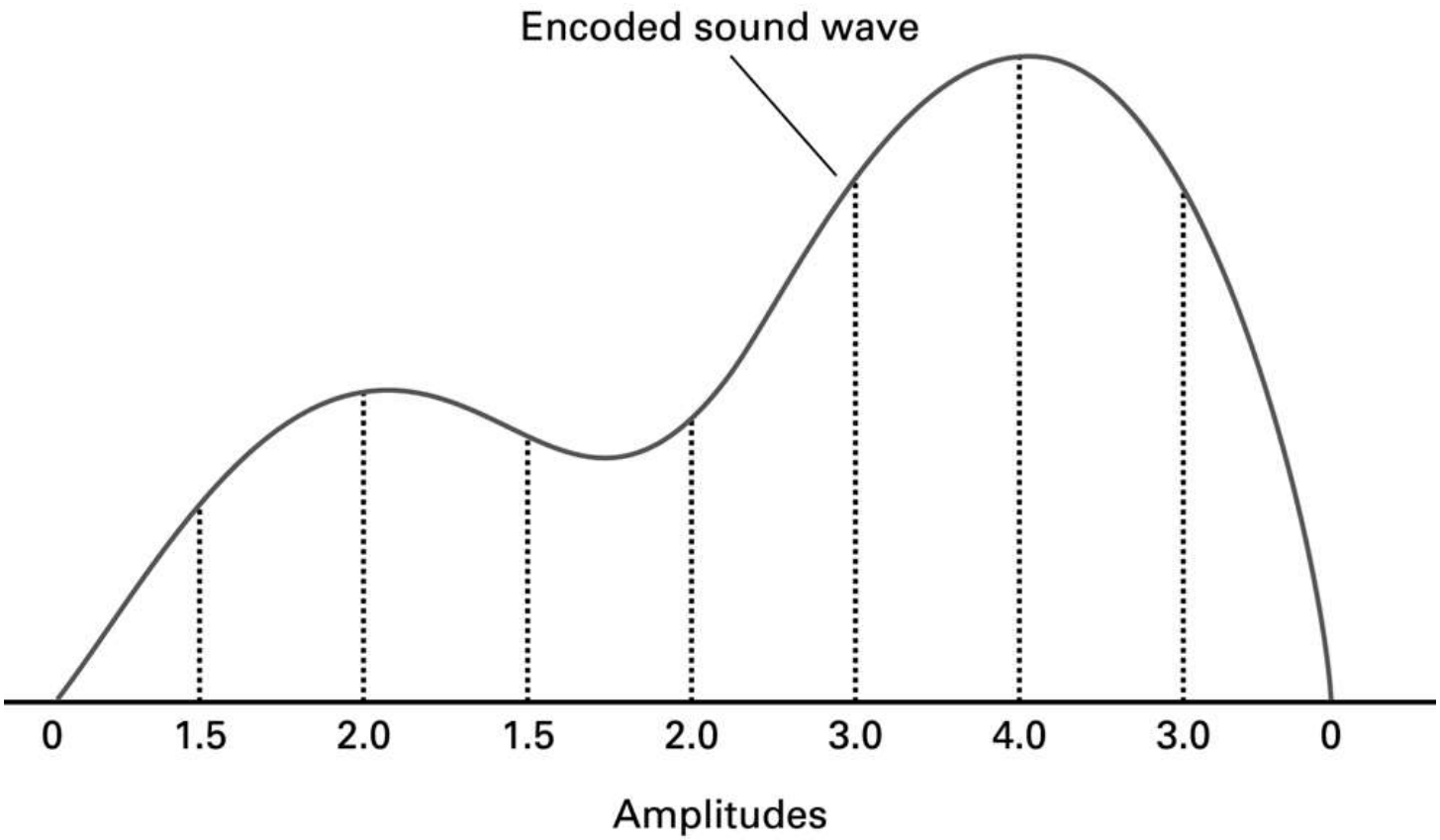
Representing Numeric Values

- **Binary notation**: Uses bits to represent a number in base two
  - All numeric values in a computer are stored in sequences of 0s and 1s
  - Counting from 0 to 8:
    - 0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111, 1000

Representing Sound

- Sampling techniques
  - Used for high quality recordings
  - Records actual audio
- MIDI
  - Used in music synthesizers
  - Records “musical score”

Figure 1.14 The sound wave represented by the sequence 0, 1.5, 2.0, 1.5, 2.0, 3.0, 4.0, 3.0, 0



存储声音

- 每秒存储声音容量的公式

采样频率Hz×采样精度bit / 8×声道数=每秒数据量Byte

- 例：用44.10KHz的采样频率，每个采样点用16位的进度存储，则录制1秒的立体声（双声道）节目，其文件所需存储量为：

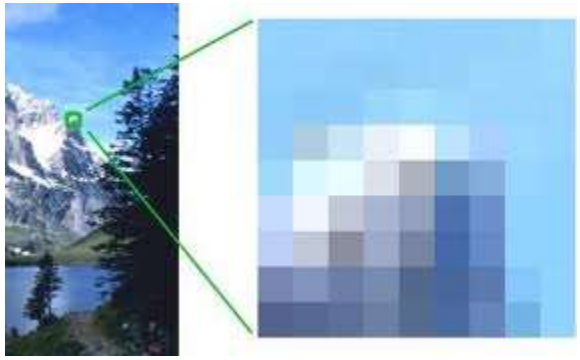
  44100×16/8×2=176.4KB

Representing Images

Representing Images

- Bit map techniques
  - Pixel像素
  - RGB
  - 存储图像：行×列×颜色深度=( )B
  - Problems with scaling up images

3或1字节
彩色、灰度



- Vector techniques
  - Scalable
  - Represent images with geometric structure
  - TrueType and PostScript

计算机中使用的一些字体

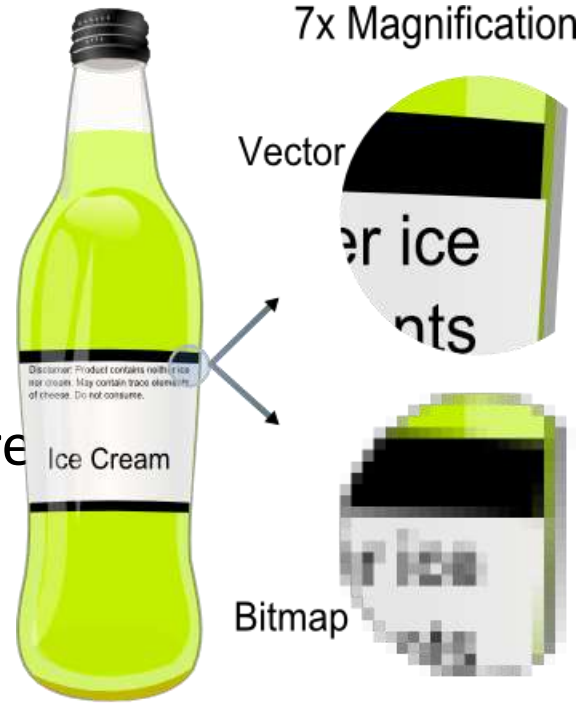
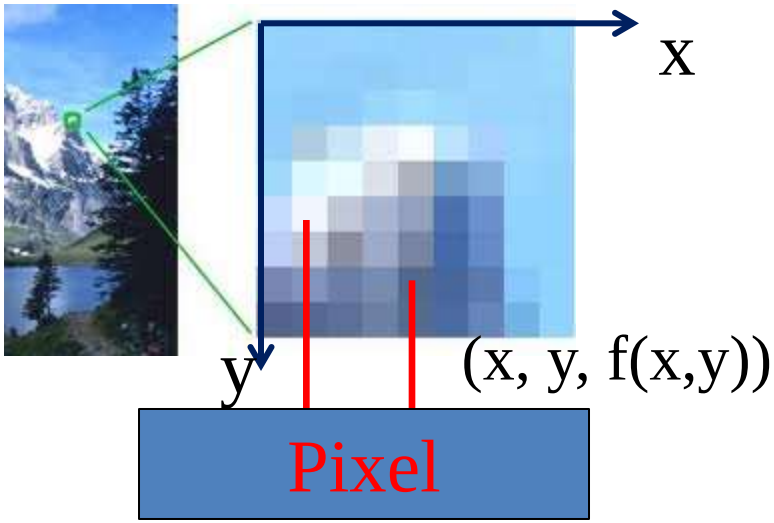


Image & pixel

Digital bitmap images are made up of pixels in a grid



Pixel:

- The smallest display element
- Associate with the position and color value

Color depth

Black & white

Gray scale

Color

Black & white

- Black & white images (binary images)

$$f(x, y) \in \{0, 1\}$$

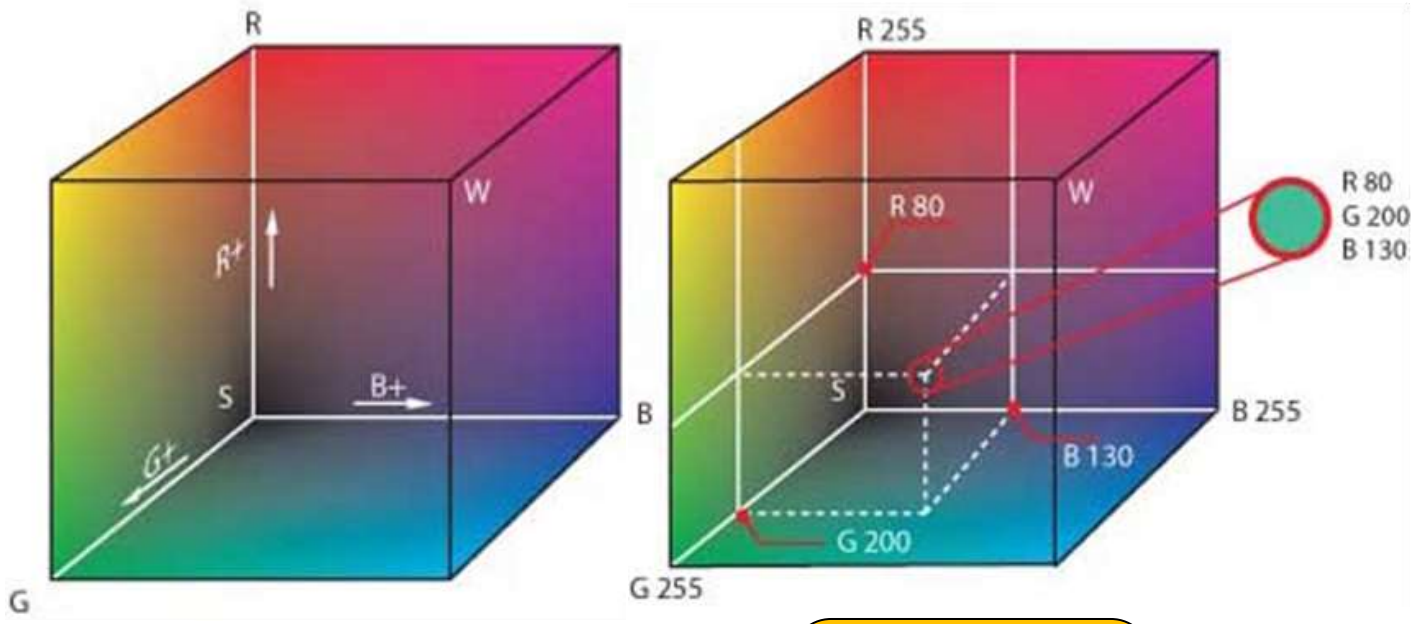
Gray scale

$$f(x, y) \in \{0, 1, \cdots, 255\}$$

- Many shades of gray in between

Color

- Typical color representation: RGB model



3 channels: red, green, blue

$R \in \{0,1,...,255\}$
$G \in \{0,1,...,255\}$
$B \in \{0,1,...,255\}$

Color

$$f(x,y) = \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad with \quad \begin{cases} R \in \{0,1,...,255\} \\ G \in \{0,1,...,255\} \\ B \in \{0,1,...,255\} \end{cases}$$

Required memory space

Color depth	Black & white	Gray	Color
Required bits	1	8	8*3=24

(Number of columns * number of rows * required bits) / 8 = required memory space (Byte)

Example:

For a 24-bit color image (1280 columns, 1024 rows), the required memory space is
1280*1024*24/8 = 4 MB

Video

- Video
  - Composed of a sequence of images, each image is called a frame
  - Played at a certain frame rate

Frame rate:
The number of images per unit of time of video

Required space for video

- If not compressed, the required memory space for a video (1 min)

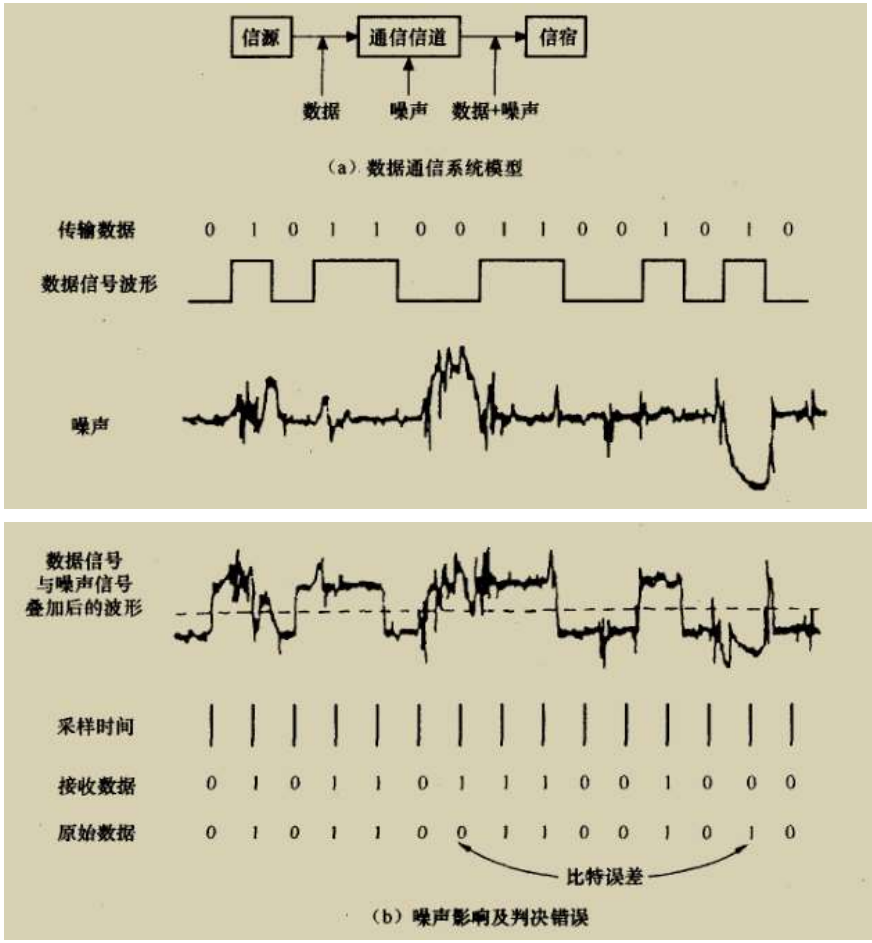
Required memory space for a video = required space for each image * frame rate*60

For a video (1280 columns, 1024 rows and 24bits for each frame) with a frame rate of 30, the required memory space is:

1280*1024*24/8*30*60=6.6GB

1.10 Communication Errors

- During transmission, error could happen
  - For example, bit 0 $\rightarrow$1 or bit 1 $\rightarrow$0



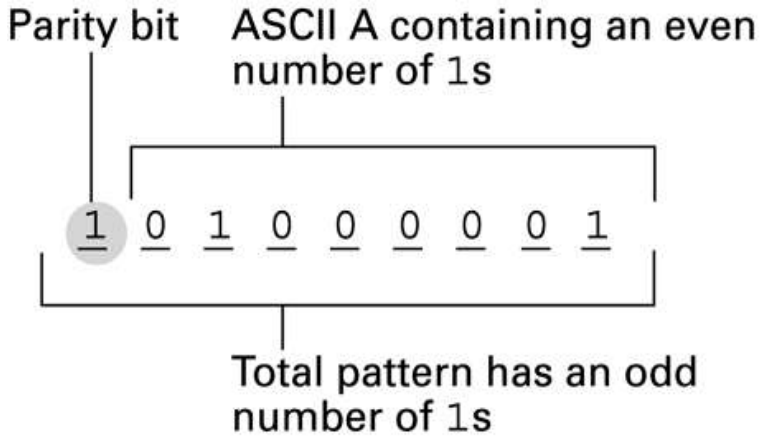
Communication Errors

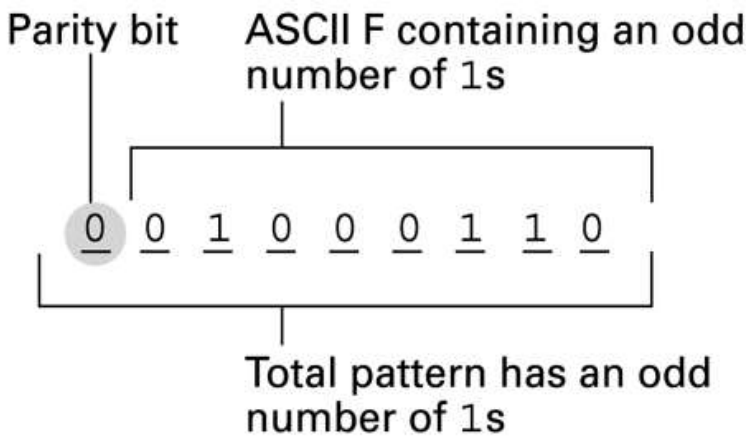
- Parity bits (even versus odd)奇偶校验位
- Checkbytes校验字节
- Error correcting codes纠错码

纠错编码方案：让传输的数据单元带有足够的冗余信息，以便在接收端发现并自动纠正传输错误。系统复杂，成本高。

检错编码方案：让传输的数据单元仅带有能使接收端发现错误的冗余信息，但不能确定错误位置，只能发现错误，不能纠正错误。简单，容易实现，编译码速度快，可通过重传使错误得以纠正。

The ASCII codes for the letters A and F adjusted for odd parity（奇校验）





奇偶校验的缺点？

奇偶校验: 可发现奇数个错误，无法纠正

An error-correcting code （纠错码）

- 汉明码
- 两个字码中不同位值的数目称为汉明距离 (Hamming distance)
- 任何一个模式要成为另一个模式，至少需要改变三个位

Hamming distance = 3

Symbol	Code
A	000000
B	001111
C	010011
D	011100
E	100110
F	101001
G	110101
H	111010

最多几位错时，利用这个编码可检测出错误？

Decoding the pattern 010100

每个模式之间的不同位的个数（称之为汉明距离）至少为3。任何一个模式要成为另一个模式，至少需要改变三个位。

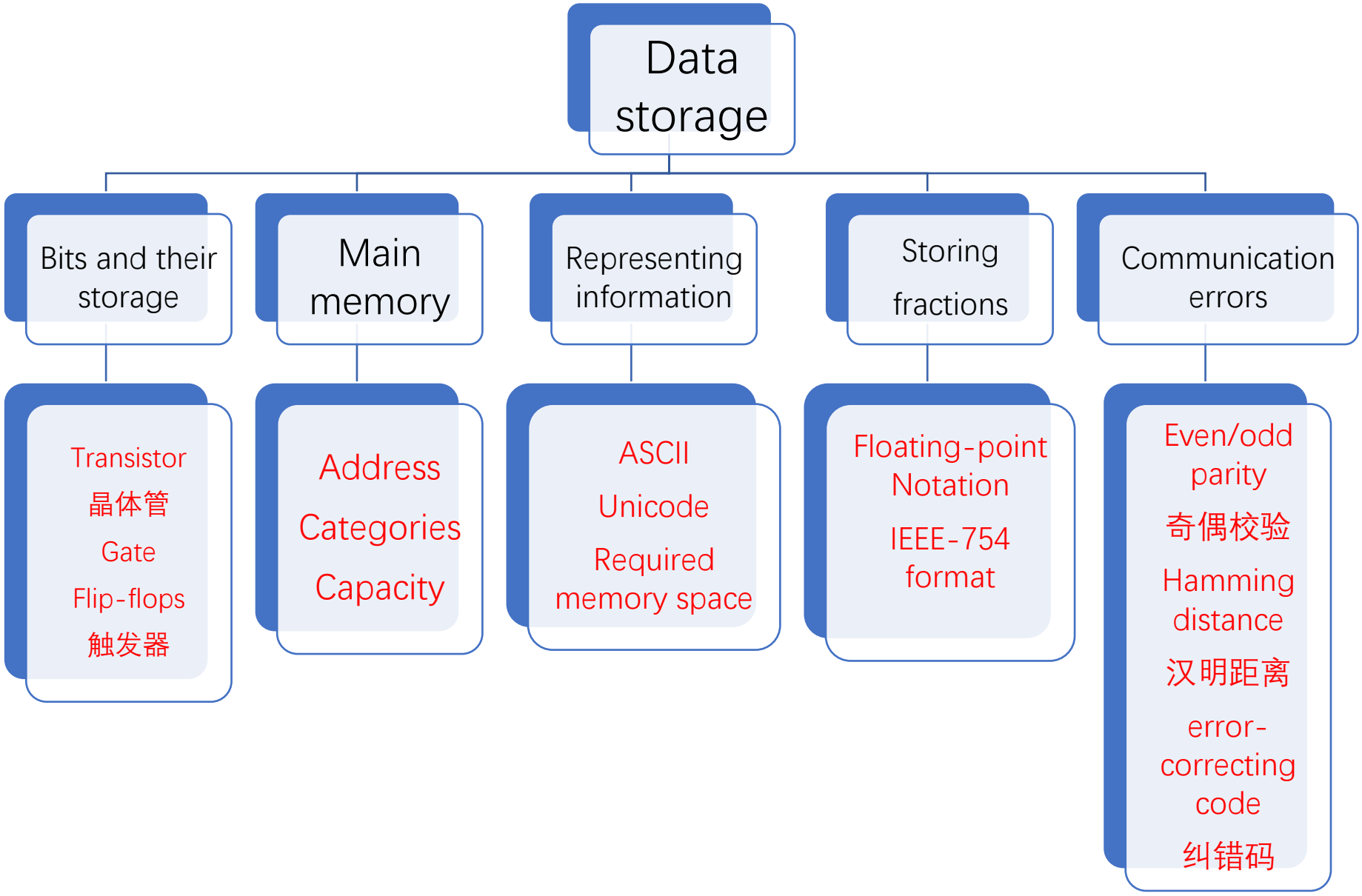
Character	Code	Pattern received	Distance between received pattern and code
A	0 0 0 0 0 0	0 **1** 0 **1** 0 0	2
B	0 0 1 1 1 1	0 **1** **0** 1 **0** **0**	4
C	0 1 0 0 1 1	0 1 0 **1** **0** **0**	3
D	0 1 1 1 0 0	0 1 **0** 1 0 0	**1** — Smallest distance
E	1 0 0 1 1 0	**0** **1** 0 1 **0** 0	3
F	1 0 1 0 0 1	**0** **1** **0** **1** 0 **0**	5
G	1 1 0 1 0 1	**0** 1 0 1 0 **0**	2
H	1 1 1 0 1 0	**0** 1 **0** **1** **0** 0	4

Hamming

发现最多2位错，纠正1位

Key points

1. 位及其存储（1.1 Bits and Their Storage）
2. 主存储（1.2 Main Memory）
3. ACSII码
4. 二进制系统（1.5 The Binary System）
5. 存储整数（1.6 Storing Integers）：补码表示（Two's Complement Notation）
6. 存储分数（1.7 Storing Fractions）：截断误差（Truncation Errors）、IEEE浮点数表示法
7. 通讯误差（1.9 Communication Errors）


Data storage
Bits and their storage
Main memory
Representing information
Storing fractions
Communication errors
Transistor
晶体管
Gate
Flip-flops
触发器
Address
Categories
Capacity
ASCII
Unicode
Required memory space
Floating-point Notation
IEEE-754 format
Even/odd parity
奇偶校验
Hamming distance
汉明距离
error-correcting code
纠错码