Information coding

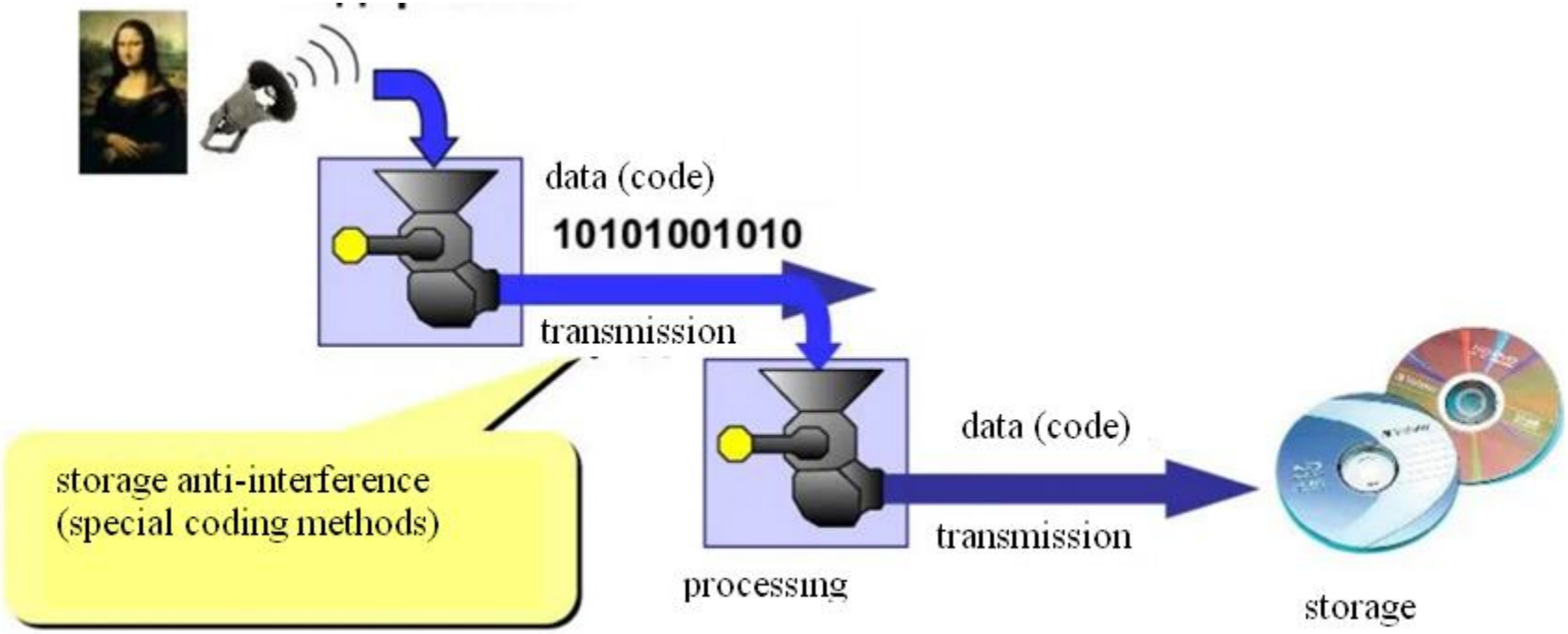
Encoding and decoding

A person uses natural languages to exchange information with other people. Along with natural languages, formal languages have been developed for their professional use in any field. The presentation of information using a language is often called coding. Code - a set of symbols (conventions) to represent information. Code is a system of conventional signs (symbols) for transmission, processing and storage of information (messages). Encoding is the process of representing information (messages) in the form of a code. The whole set of characters used for encoding is called the encoding alphabet. For example, in the computer's memory, any information is encoded using a binary alphabet containing only two characters: 0 and 1

What is coding?

Coding is the recording of information using some sign system (language).

Why do coding information?



Information is transmitted, processed and stored in the form of codes.

Information encoding methods

Different methods can be used to encode the same information; their choice depends on a number of circumstances: the purpose of coding, conditions, available means.

If you need to write down the text at the rate of speech, we use shorthand;

if you need to transfer text abroad, we use the English alphabet;

if it is necessary to present the text in a form understandable for a literate English man, we write it down according to the rules of the grammar of the English language. "Hello!"

What is coding?

Coding is the presentation of information in a form convenient for its storage, transmission and processing. The rule for this transformation is called code.

Text:

in English: Hello, Word!

Transfer abroad (transliteration): Хелло, Ворд!

Windows-1251: CFF0E8E2E52C20C2E0F1FF21

shorthand: 

encryption: Rsigzhu-! GBta "

Numbers:

for calculations: 25

in words: twenty five

Roman system: XXV

How is it encrypted? Why?

Information encoding methods

The choice of a method for encoding information may be associated with the intended method of processing it.
Let us show this using the example of the representation of numbers - quantitative information.
Using the English alphabet, you can write down the number “forty-seven.”
Using the alphabet of the Arabic decimal number system, we write “47”.
The second method is not only shorter than the first, but also more convenient for performing calculations. Which record is more convenient for performing calculations: “forty-seven multiply one hundred twenty-five "or" 47 x 125 "?
Obviously the second.

Message encryption

In some cases, there is a need to privacy the text of a message or document so that it cannot be read by those who are not supposed to. This is called tamper protection. In this case, the secret text is encrypted. In ancient times, encryption was called cryptography. Encryption is the process of converting plaintext into encrypted text,
decryption is the reverse conversion process

Encryption is also encoding, but with a secret method known only to the source and addressee. A science called cryptography deals with encryption methods.

Chapp's optical telegraph

Chappe created a special table of codes, where each letter of the alphabet corresponded to a specific figure formed by the Semaphore, depending on the positions of the cross bars relative to the support pole. Chappe's system allowed messages to be transmitted at two words per minute and quickly spread throughout Europe. In Sweden, a chain of optical telegraph stations operated until 1880.

First telegraph

The first technical means of transmitting information over a distance was the telegraph, invented in 1837 by the American Samuel Morse. A telegraph message is a sequence of electrical signals transmitted from one telegraph apparatus over wires to another telegraph apparatus. The inventor Samuel Morse invented an amazing code (Morse code, Morse code, "Morse code") that still serves mankind today.

First telegraph

The information is encoded with three "letters": a long signal (dash), a short signal (dot) and no signal (pause) to separate letters. Thus, encoding is reduced to using a set of characters arranged in a strictly defined order. The most famous telegraph message is the "SOS" distress signal (Save Our Souls). This is how it looks: "• • • - - - • • •"

First wireless telegraph (radio receiver)

On May 7, 1895, the Russian scientist Alexander Stepanovich Popov, at a meeting of the Russian Physicochemical Society, demonstrated a device, which he named "lightning detector", which was designed to register electromagnetic waves. This device is considered the world's first wireless telegraphy device, a radio receiver. In 1897, using wireless telegraphy devices, Popov received and transmitted messages between the coast and a military vessel. In 1899, Popov designed a modernized version of an electromagnetic wave receiver, where signals (in Morse code) were received by the operator's head phones.

First wireless telegraph (radio receiver)

In 1900, thanks to radio stations built on the island of Gogland and at the Russian naval base in Kotka under the leadership of Popov, rescue operations were successfully carried out aboard the General-Admiral Apraksin warship, which ran aground off the island of Gogland. As a result of the exchange of messages transmitted by wireless telegraphy, the crew of the Russian icebreaker Ermak was promptly and accurately transmitted information about the Finnish fishermen who were on the torn ice floe.

International Morse Code

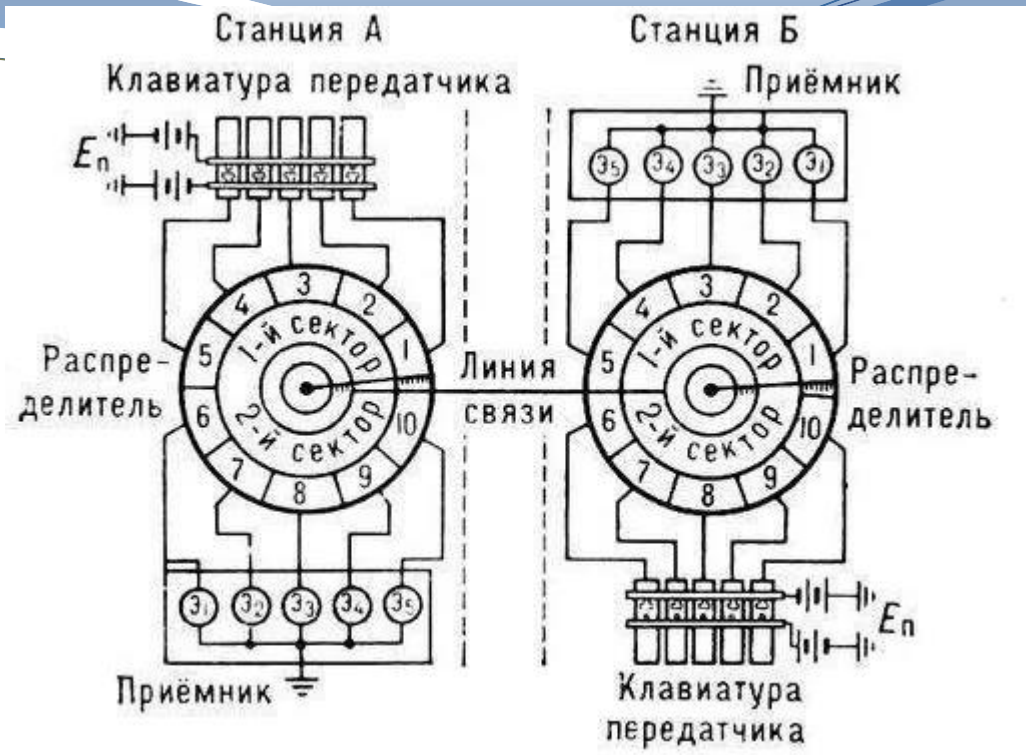
A	•–	N	–•	1	•––––
B	–•••	O	–––	2	••–––
C	–•–•	P	•––•	3	•••––
D	–••	Q	––•–	4	••••–
E	•	R	•–•	5	•••••
F	••–•	S	•••	6	–••••
G	––•	T	–	7	––•••
H	••••	U	••–	8	–––••
I	••	V	•••–	9	––––•
J	•–––	W	•––	0	–––––
K	–•–	X	–••–	.	•–•–•–
L	•–••	Y	–•––	,	––••––
M	––	Z	––••	?	••––••

The code is uneven, you need a separator!

- • - - • • • - - • • - A characteristic feature of Morse code is the variable length of the code of different letters, therefore Morse code is called an Uneven code. The letters that are more common in the text have a shorter code than the rare letters. This is done in order to shorten the length of the entire message. But due to the variable length of the letter code, there is a problem of separating letters from each other in the text. Therefore, you have to use a pause (skip) to separate. Consequently, the Morse telegraph alphabet is ternary, since it uses three characters: dot, dash, gap. Uneven code

Bodo's Telegraph

The uniform telegraph code was invented by the Frenchman Jean Maurice Baudot at the end of the 19th century. It used only two different types of signals. It doesn't matter what you call them: point and dash, plus and minus, zero and one. These are two distinct electrical signals.



The length of the code of all characters is the same and is equal to five. In this case, there is no problem of separating letters from each other: each five of the signals is a text sign. Therefore, a pass is not needed. A code is called uniform if the length of the code of all characters is equal. Baudot code is the first method of binary encoding of information in the history of technology. Thanks to this idea, it was possible to create a direct-printing telegraph apparatus that looks like a typewriter. Pressing a key with a specific letter generates a corresponding five-pulse signal, which is transmitted over the communication line. In honor of Bodo, the unit of information transmission speed was named - baud. Modern computers also use a uniform binary code to encode text.

Language

Language is a system of signs used to store, transmit and process information.

Egyptian Hieroglyphs:

Egyptian writing	
	hand
	house
	cobra
	lion
	water

China Hieroglyphs:

Hieroglyphs (China)	
日	sun
月	moon
雨	rain
山	mountain
马	horse

What languages are there?

Natural	Formal
Russian English Chinese Swedish	$y = 3\sin x + 1$ $2H_2 + O_2 = 2H_2O$  1. **e2-e4 e7-e5…**

A formal language is a language that uniquely defines the meaning of each word, the rules for constructing sentences and giving them meaning.

Natural and formal languages

Natural

- the result of the development of society
- for communication at home
- the meanings of words depend on the context
- there are synonyms
- there are homonyms
- no strict rules for the formation of sentences
- there are exceptions

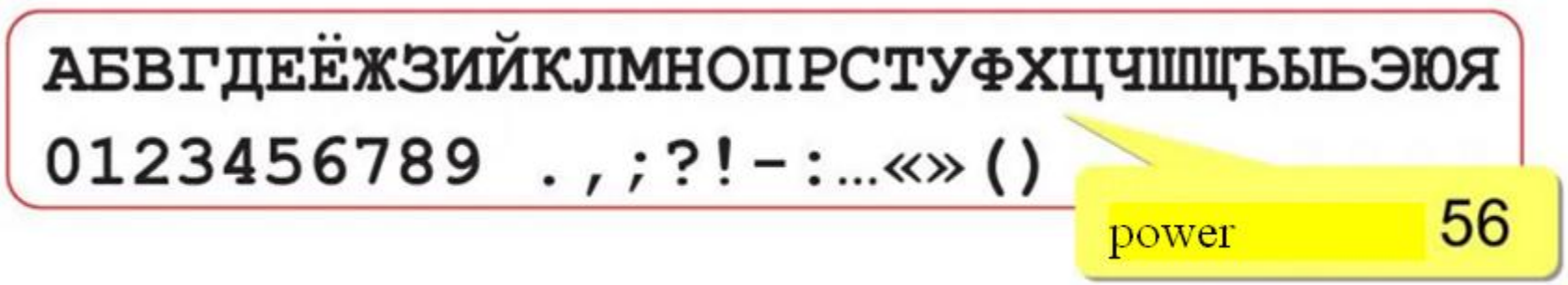
Natural and formal languages

Formal

- created by people in special fields of knowledge
- the meaning of words does not depend on the context
- without synonyms
- without homonyms
- without rule for the formation of sentences
- strictly defined
- without exceptions

Alphabetic writing

An alphabet is a set of characters used in a language. The power of the alphabet is the number of characters in the alphabet.



A word is a sequence of alphabet characters that is used as an independent unit and has a specific meaning.

Message

A message is any sequence of characters in an alphabet.

Example:

alphabet @ # $%.

Messages of length 1:

@ # $ %.

total 4

Messages of length 2:

@@	@#	@$	@%
#@	##	#$	#%
$@	$#	$$	$%
%@	%#	%$	%%

total 16

total 16

How many messages are of length L?

Information coding:

Binary encoding

Binary encoding is an encoding using two characters.

А	Б	В	Г
00	01	10	11

АБАВГБ → 000100101101

Number of messages of length I bits $N = 2^I$

Example.

You need to encode the athlete's number from 1 to 200.

How many bits will it take?

8 bits $2^7 < 200 \leq 2^8 = 256$

8 bits

Uniform code

Binary encoding in a computer

All information that the computer processes must be represented in binary code using two digits: 0 and 1. These two characters are usually called binary digits or bits. Any message can be encoded with two digits 0 and 1. This was the reason that two important processes must be organized in a computer: encoding and decoding. Binary coding is the transformation of input information into a form that is perceived by a computer, i.e. binary code.

Why binary encoding?

From the point of view of technical implementation, the use of a binary number system for encoding information turned out to be much simpler than the use of other methods. Indeed, it is convenient to encode information in the form of a sequence of zeros and ones, if these values are presented as two possible stable states of an electronic element: 0 - no electrical signal; 1 - presence of an electrical signal. Methods of encoding and decoding information in a computer, first of all, depend on the type of information, namely, what should be encoded: numbers, text, graphics or sound.

Encoding text information

- Traditional encodings use 8 bits to encode one character. It is easy to calculate by the formula that such an 8-bit code can encode 256 different characters.
- Assigning a specific numeric code to a character is a matter of convention. ASCII (American Standard Code for Information Interchange) code table has been adopted as an international standard.
- ASCII code was introduced in the USA in 1963, modified in 1973.19 1 Character = 1 Byte = 8 Bits

A – 01000001, B – 01000010,
C – 01000011, D – 01000100

Encoding text information

Characters from 0 to 127 - Latin letters, numbers and punctuation marks form a permanent (base) part of the table. (Codes from 0 to 32 are not assigned to symbols, but to function keys). Characters from 128 to 255 are reserved for the national standard - an extended table. The composition of these characters is determined by the code page.

Encoding text information

National standards of coding tables include the international part of the coding table without changes, and in the second half they contain codes of national alphabets, pseudo-graphic symbols and some mathematical symbols.

Basic ASCII table - of characters

ASCII Code Chart

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	NUL	SOH	STX	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
1	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
2		!	"	#	$	%	&	'	(	)	*	+	,	-	.	/
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

Characters

H e l l o

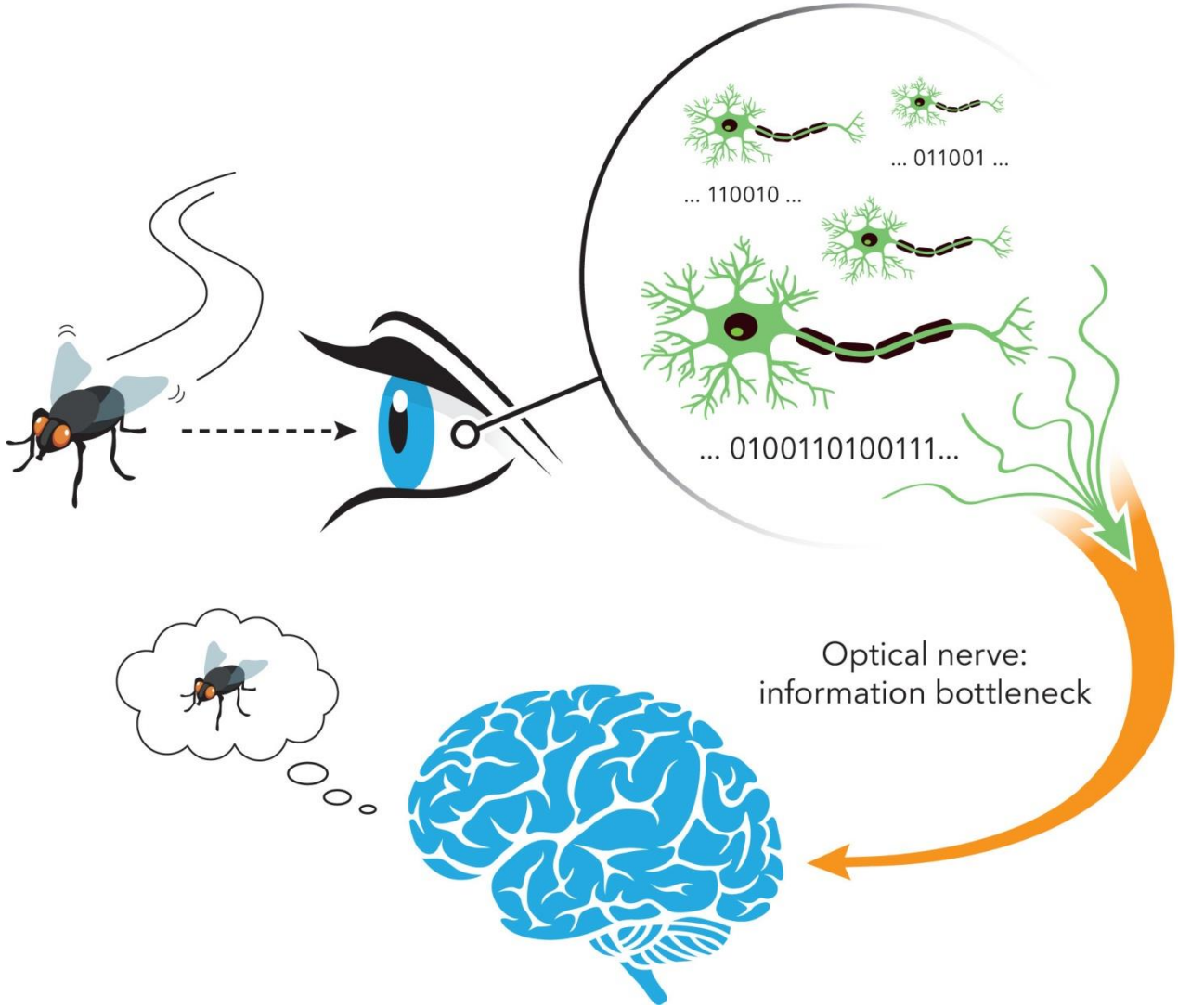
Unicode Code Points

U+0048 U+0065 U+006C U+006C U+006F

UTF-8 Encoding

01001000 01100101 01101100 01101100 01101111

Coding graphic information

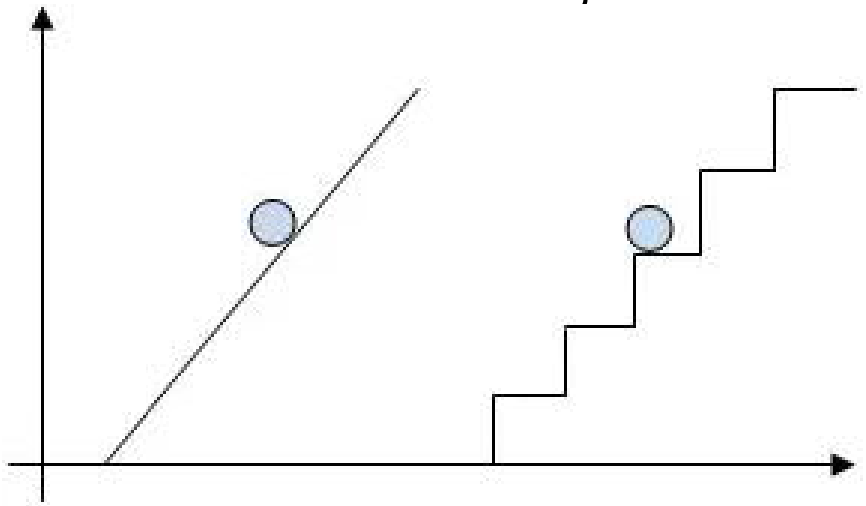


Analog and discrete information presentation

A person is able to perceive and store information in the form of images (visual, sound, tactile, gustatory and olfactory). Visual images can be saved as images (drawings, photographs, and so on), and sound images can be recorded on records, magnetic tapes, laser disks, and so on. Information, including graphic and sound, can be presented in analog or discrete form. With analog representation, a physical quantity takes on an infinite set of values, and its values change continuously. In a discrete representation, a physical quantity takes on a finite set of values, and its value changes abruptly

Analog and discrete information presentation

Let's give an example of analog and discrete information representation. The position of the body on an inclined plane and on a staircase is set by the values of the X and Y coordinates.When a body moves along an inclined plane, its coordinates can take on an infinite set of continuously changing values from a certain range, and when moving along a staircase, only a certain set of values, and changing abruptly.



Sampling

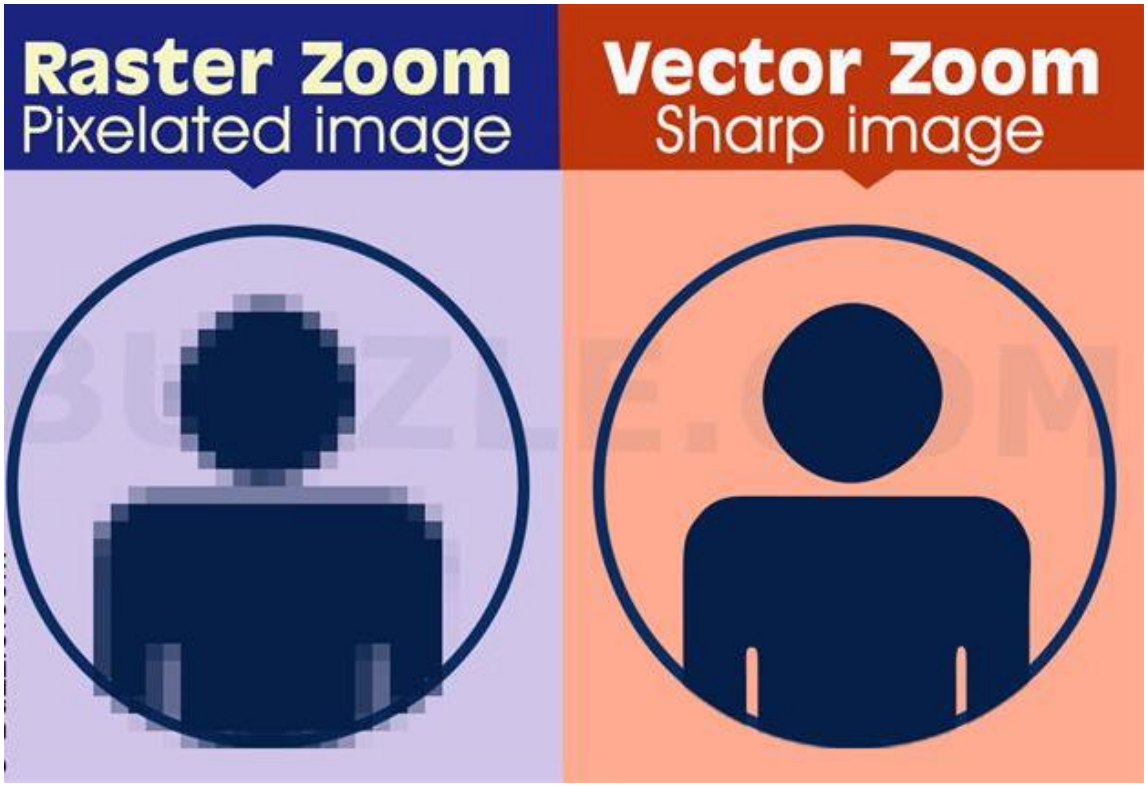
An example of an analog representation of graphic information is, for example, a painting canvas, the color of which changes continuously, and discrete - an image printed using an inkjet printer and consisting of separate dots of different colors. An example of analog storage of sound information is a vinyl record (the sound track changes its shape continuously), and discrete storage is an audio CD (the sound track of which contains areas with different reflectivity)

Sampling

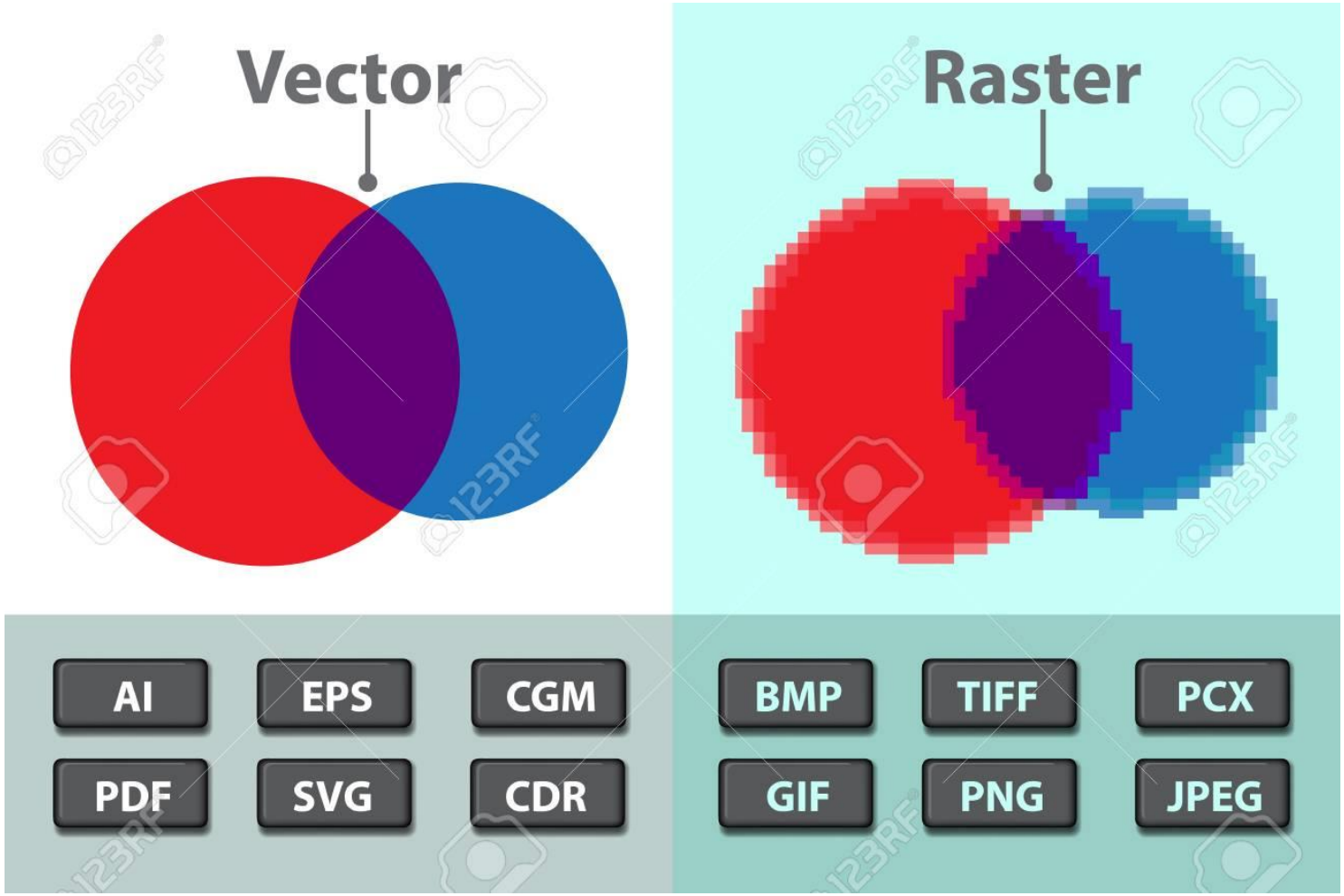
Conversion of graphic and audio information from analog to discrete form is performed by sampling, that is, splitting a continuous graphic image and a continuous (analog) audio signal into separate elements. In the process of sampling, coding is performed, that is, the assignment of each element to a specific value in the form of a code. Sampling is the transformation of continuous images and sound into a set of discrete values in the form of codes

Representation of images in a computer

There are two ways to create and store graphic objects in a computer - as a raster or as a vector image. Each type of image uses its own encoding method.



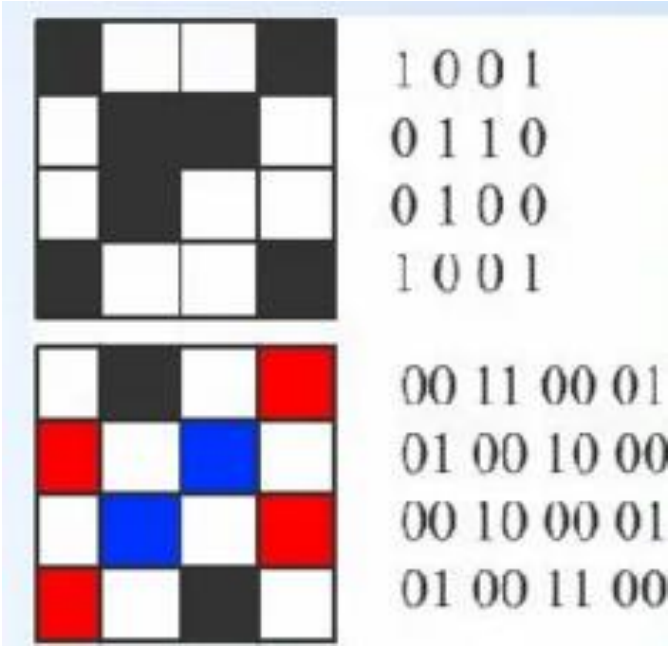
Representation of images in a computer



Vector	Raster
WEB USES	
SOURCE files for logos, charts, icons, or any hard-edged graphics	**OUTPUT files** for most web graphics displayed on the screen
PRINT USES	
SOURCE files to be sent to the printer	**HI-RES files** can be printed at 300dpi
FILE TYPES	
.ai .eps .pdf .svg	.jpg .gif .png .tif

Bitmap encoding

A raster image is a collection of dots (pixels) of different colors. Pixel is the smallest area of the image, the color of which can be set independently. In the process of image encoding, its spatial sampling is performed.



Bitmap encoding

Spatial sampling of an image can be compared to building an image from a mosaic (a large number of small multi-colored glasses). The image is divided into separate small fragments (dots), and each fragment is assigned a value for its color, that is, a color code (red, green, blue, and so on). The image quality depends on the number of dots (the smaller the dot size and, accordingly, the larger the number, the better the quality) and the number of colors used (the more colors, the better the image is encoded)

Bitmap encoding

Spatial sampling of an image can be compared to building an image from a mosaic (a large number of small multi-colored glasses). The image is divided into separate small fragments (dots), and each fragment is assigned a value for its color, that is, a color code (red, green, blue, and so on). The image quality depends on the number of dots (the smaller the dot size and, accordingly, the larger the number, the better the quality) and the number of colors used (the more colors, the better the image is encoded)

Basic characteristics of bitmaps

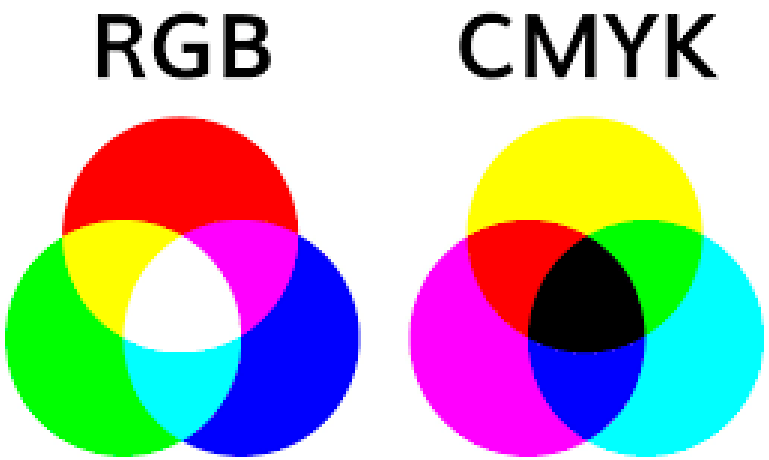
1) Resolution - the number of pixels in width and height, or the total number of pixels. 1024 × 768, 640 × 480 or 4 Megapixels;
2) Color depth - the number of colors used. N = 2 k, where N is the number of colors and k is the color depth;
3) Pixel size;
4) Color space (color model) RGB, CMYK, XYZ, YCbCr Basic characteristics of raster images

Color models

To represent color in the form of a numerical code, two inverse color models are used: RGB or CMYK. The RGB model is used in TVs, monitors, projectors, scanners, digital cameras ...

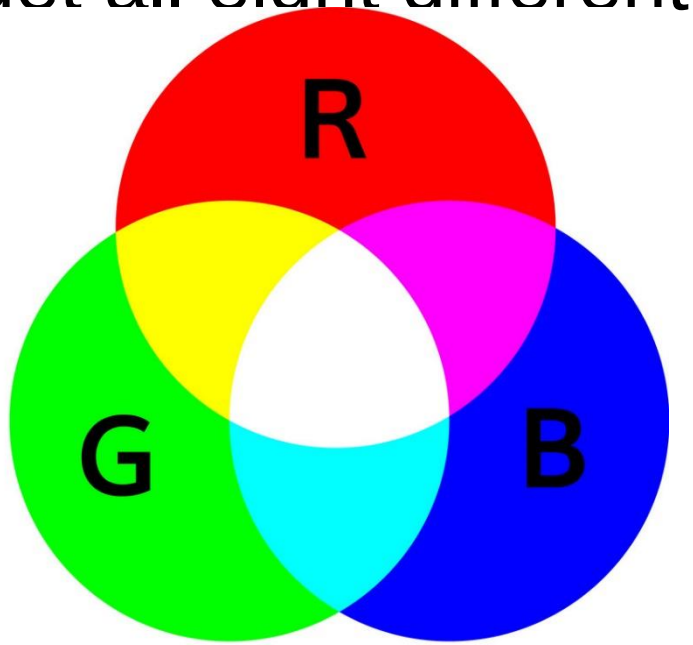
The main colors in this model are: red (R Red), green (G Green), blue (B Blue).

The CMYK color model is used in the printing industry to form images intended for printing on paper



RGB color model

Color images can have different color depths, which is defined by the number of bits used to encode the color of a point. If we encode the color of one point in the image with three bits (one bit for each RGB color), then we get all eight different colors.



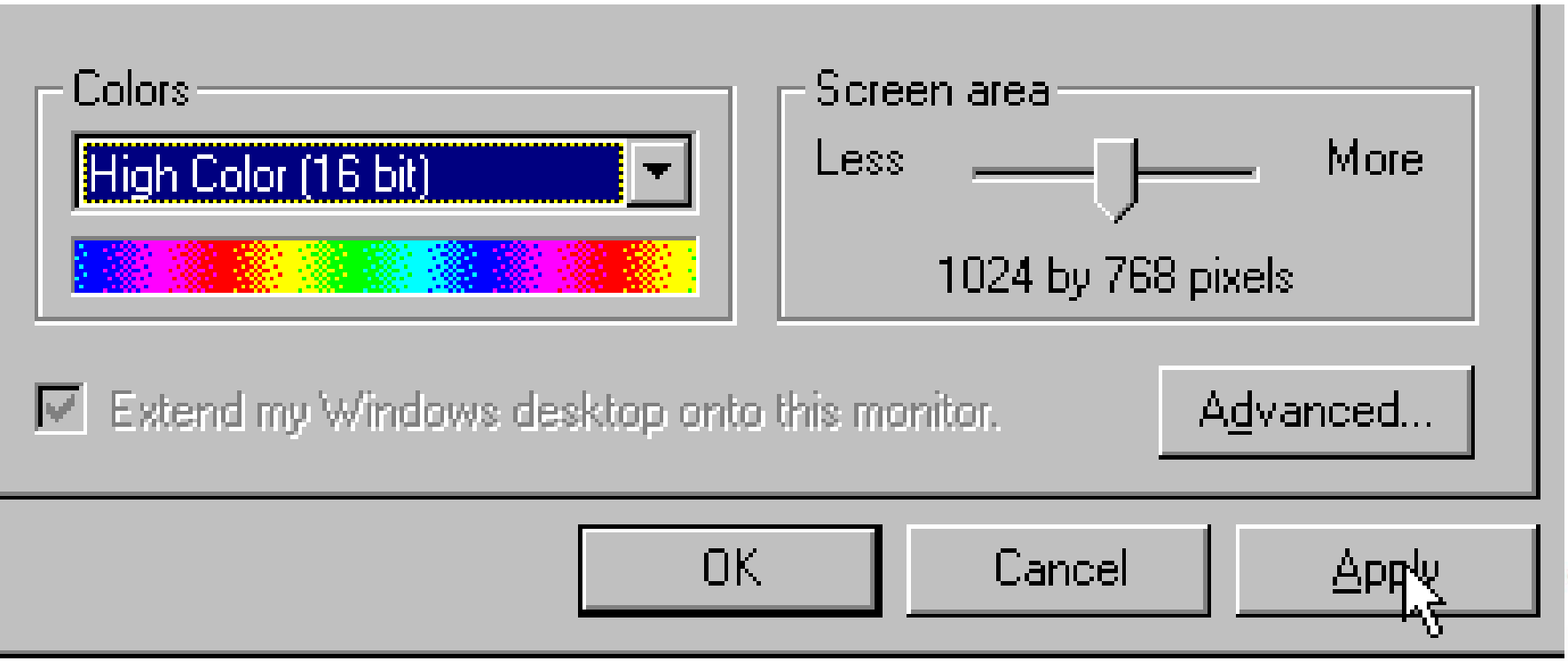
RGB color model

Color images can have different color depths, which is defined by the number of bits used to encode the color of a point. If we encode the color of one point in the image with three bits (one bit for each RGB color), then we get all eight different colors.

R	G	B	Color
1	1	1	White
1	1	0	Yellow
1	0	1	Perple
1	0	0	Red
0	1	1	Blue
0	1	0	Green
0	0	1	dark blue
0	0	0	Black

High Color

Two bytes (16 bits) define 65536 color shades. This mode is called High Color. In cases where the image color is encoded with two bytes (High Color mode), the screen can display 65 thousand colors (2^16 = 65536).



True Color

In practice, to store information about the color of each point of a color image in the RGB model, 3 bytes (24 bits) are usually allocated - 1 byte (8 bits) for the color value of each component. Thus, each RGB component can take a value in the range from 0 to 255 (total 2^8 = 256 values), and each point of the image, with such a coding system, can be colored in one of 2^24 = 16 777 216 colors. Such a set of colors is usually called True Color, because the human eye is still not able to distinguish more variety.

Advantages and of raster graphics

1. Allows you to create any picture
2. Prevalence
3. High processing speed
4. Raster image is natural for most input-output devices

Disadvantages of raster graphics

1. Large file size
2. Impossibility of perfect scaling
3. Inability to print to a plotter

Graphic file formats

Graphics file formats determine how information is stored in a file (raster or vector), as well as the form of information storage (compression algorithm used). Most popular bitmap formats:

- BMP
- GIF
- PCX
- PNG
- JPEG
- TIFF
- RAW

Bitmap formats

Bit MaP image (BMP) is a universal raster image file format used in the Windows operating system. This format is supported by many graphic editors, including Paint. Recommended for storing and exchanging data with other applications. Graphics Interchange Format (GIF) is a bitmap graphics file format supported by applications for various operating systems. Includes a lossless compression algorithm that allows you to reduce the file size by several times. Recommended for storing images created by software (diagrams, graphs, etc.) and drawings (such as applications) with a limited number of colors (up to 256). Used for posting graphic images on Web sites on the Internet.

Bitmap formats

PCX is a legacy format that can compress simple hand-drawn images well. Portable Network Graphic (PNG) is a bitmap graphics file format similar to GIF. Recommended for placing graphic images on Web sites on the Internet. Joint Photographic Expert Group (JPEG) is a bitmap graphics file format that implements an efficient compression algorithm (JPEG method) for scanned photographs and illustrations. The compression algorithm allows you to reduce the file size tenfold, but it leads to the irreversible loss of some information. Supported by applications for various operating systems. Used for posting graphic images on Web sites on the Internet.

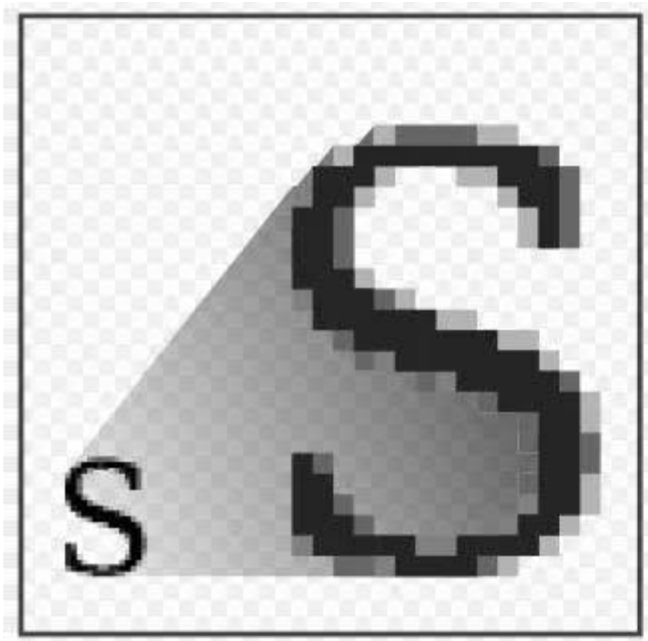
Bitmap formats

Tagged Image File Format (TIFF) is a raster image file format supported by all major graphics editors and computer platforms. Includes a lossless compression algorithm. Used to exchange documents between different programs. Recommended for use with publishing systems. The format supports a wide range of color depth changes, different color spaces, different compression settings (both lossy and non-lossy). RAW - stores information directly received from the matrix of a digital camera or similar device without applying any transformations to it, and also stores camera settings.

Vector image encoding

A vector image is a collection of graphic primitives (point, line, ellipse ...). Each primitive is described by mathematical formulas. The encoding depends on the application environment. The advantage of vector graphics is that the files that store vector graphics are relatively small. It is also important that vector graphics can be enlarged or reduced without loss of quality.

Raster and vector graphics



RASTER **VECTOR**

CONCLUSION

The computer works on machine language (“0” and “1” - binary code)

Good to see you next time!