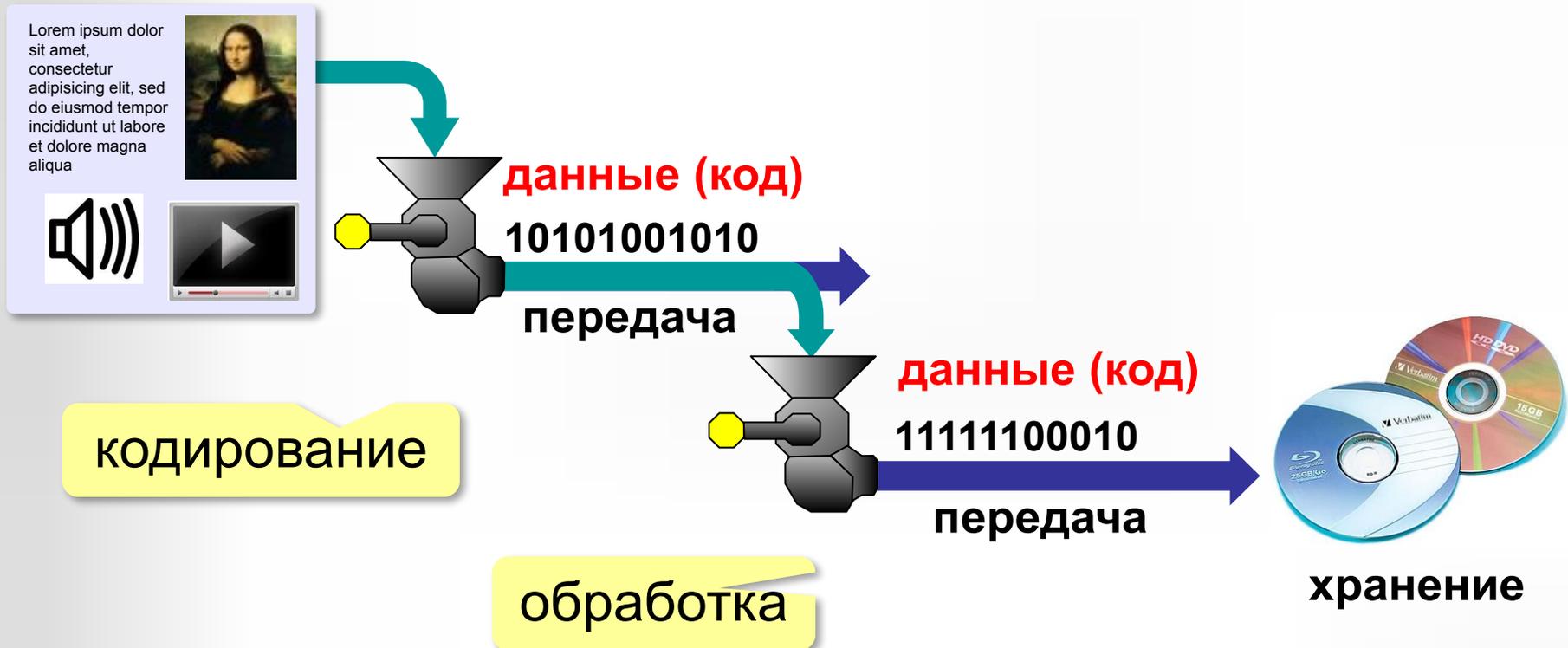


Кодирование информации

Зачем кодировать информацию?

Кодирование — это представление информации в форме, удобной для её хранения, передачи и обработки.

В компьютерах используется двоичный код:



Кодирование символов

Текстовый файл

- на экране (символы)
- в памяти — коды



1000001_2	1000010_2	1000011_2	1000100_2
65	66	67	68



В файле хранятся не изображения символов, а их числовые коды!

Файлы со шрифтами: ***.fon***, ***.ttf***, ***.otf***

Кодировка ASCII (7-битная)

ASCII = *American Standard Code for Information Interchange*

Коды 0-127:

0-31 **управляющие символы:**

7 – звонок, 10 – новая строка,

13 – возврат каретки, 27 – Esc.

32 пробел

знаки препинания: . , : ; ! ?

специальные знаки: + - * / () { } []

48-57 цифры **0..9**

65-90 заглавные латинские буквы **A-Z**

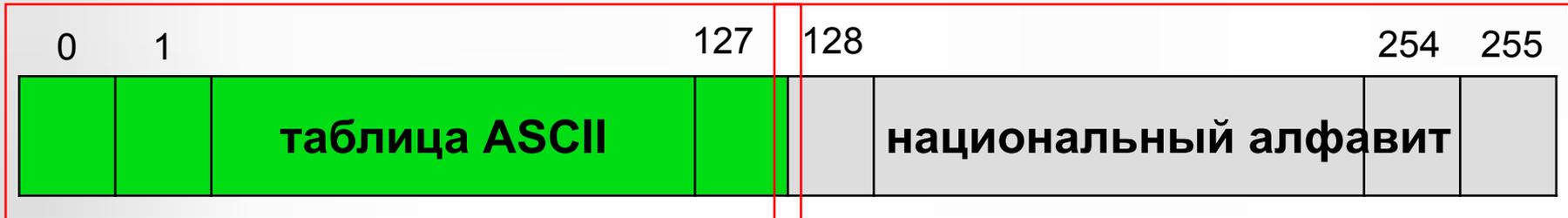
97-122 строчные латинские буквы **a-z**



Где русские буквы?

8-битные кодировки

Кодовые страницы (расширения ASCII):



Для русского языка:

CP-866 для *MS DOS*

CP-1251 для *Windows* (Интернет)

KOI8-R для *UNIX* (Интернет)

MacCyrillic для компьютеров *Apple*

Проблема:

Windows-1251

Привет, Вася!

рТЙЧЕФ,

чБУС!

KOI8-R

оПХБЕР,

бЮЯЪ!

Привет, Вася!

• Unicode

- для представления (кодирования) одного символа используется два байта (16 бит), и это позволяет включить в код символа информацию о том, какому языку принадлежит символ и как его нужно воспроизводить на экране монитора или на принтере.
- Два байта позволяют закодировать 65536 символов.

Стандарт UNICODE

1 112 064 знаков, используются около 100 000

Windows: **UTF-16**

16 битов на распространённые символы,
32 бита на редко встречающиеся

Linux: **UTF-8**

8 битов на символ для ASCII,
от 16 до 48 бита на остальные



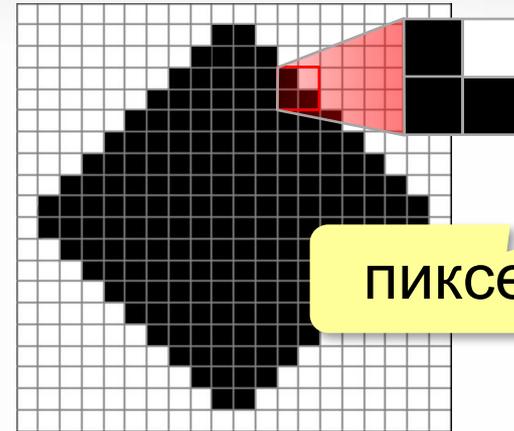
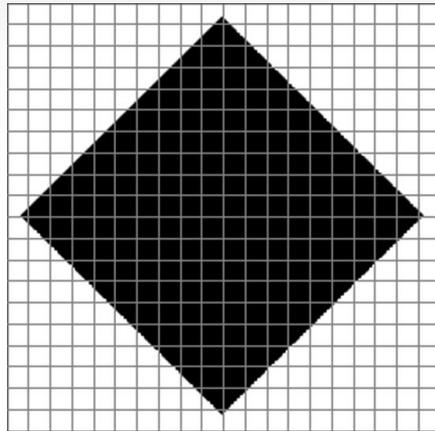
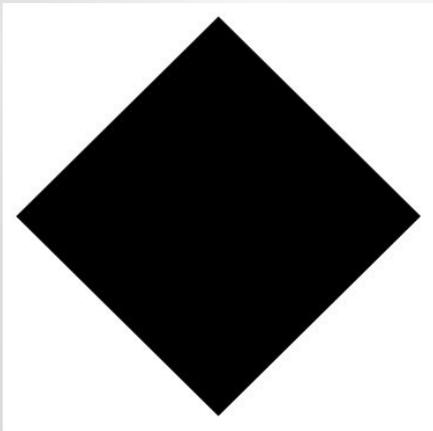
- совместимость с ASCII
- более экономична, чем UTF-16, если много символов ASCII



2010 г. – 50% сайтов использовали UTF-8!

Кодирование графической информации

Растровое кодирование



дискретизация

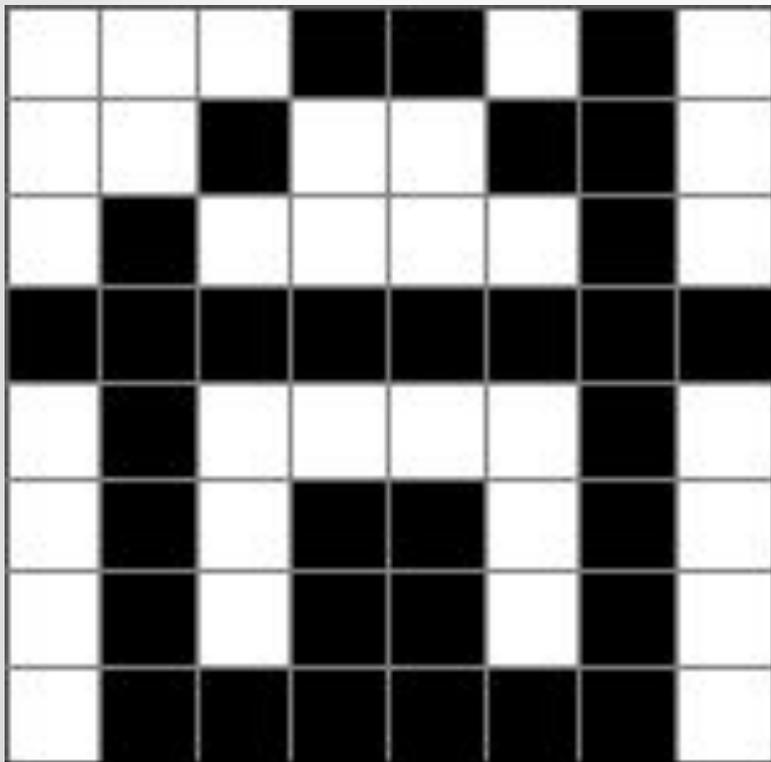


Рисунок искажается!

Пиксель – это наименьший элемент рисунка, для которого можно задать свой цвет.

Растровое изображение – это изображение, которое кодируется как множество пикселей.

Растровое кодирование



0	0	0	1	1	0	1	0	1A
0	0	1	0	0	1	1	0	26
0	1	0	0	0	0	1	0	42
1	1	1	1	1	1	1	1	FF
0	1	0	0	0	0	1	0	42
0	1	0	1	1	0	1	0	5A
0	1	0	1	1	0	1	0	5A
0	1	1	1	1	1	1	0	7E

1A2642FF425A5A7E₁₆

Разрешение

Разрешение – это количество пикселей, приходящихся на дюйм размера изображения.

ppi = *pixels per inch*, пикселей на дюйм

1 дюйм = 2,54 см



300 ppi

печать



96 ppi

экран

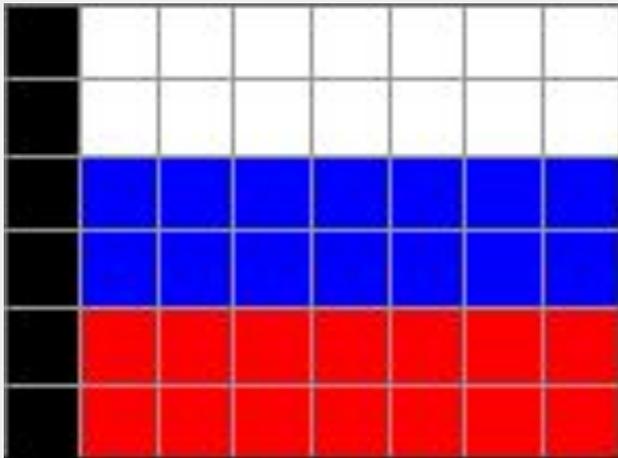


48 ppi



24 ppi

Кодирование цвета



00	11	11	11	11	11	11	11
00	11	11	11	11	11	11	11
00	01	01	01	01	01	01	01
00	01	01	01	01	01	01	01
00	10	10	10	10	10	10	10
00	10	10	10	10	10	10	10



Как выводить на монитор цвет с кодом 00?



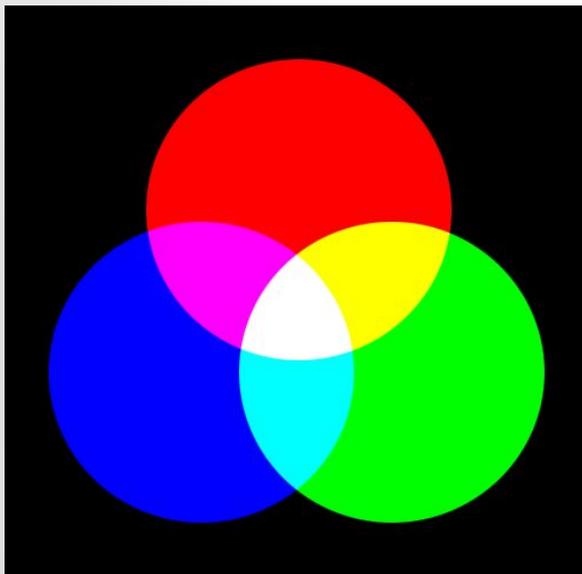
Как закодировать цвет в виде чисел?

Цветовая модель RGB

Д. Максвелл, 1860

цвет = (**R**, **G**, **B**)

red *green* *blue*
красный зеленый синий
0..255 0..255 0..255



■ (0, 0, 0)

□ (255, 255, 255)

■ (255, 0, 0)

■ (255, 150, 150)

■ (0, 255, 0)

■ (255, 255, 0)

■ (0, 0, 255)

■ (100, 0, 0)



Сколько разных цветов можно кодировать?

$256 \cdot 256 \cdot 256 = 16\ 777\ 216$ (*True Color*, «истинный цвет»)



RGB – цветовая модель для устройств, излучающих свет (мониторов)!

Цветовая модель RGB

(**255**, **255**, **0**) → **#FFFF00**

	RGB	Веб-страница
	(0, 0, 0)	#000000
	(255,255,255)	#FFFFFF
	(255, 0, 0)	#FF0000
	(0, 255, 0)	#00FF00
	(0, 0, 255)	#0000FF
	(255, 255, 0)	#FFFF00
	(204,204,204)	#CCCCCC

Глубина цвета

Глубина цвета — это количество битов, используемое для кодирования цвета пикселя.



Сколько памяти нужно для хранения цвета 1 пикселя в режиме *True Color*?

R (0..255) 256 = 2^8 вариантов 8 битов = 1 байт

R G B: 24 бита = 3 байта

True Color
(ИСТИННЫЙ ЦВЕТ)

Задача. Определите размер файла, в котором закодирован растровый рисунок размером **20×30 пикселей** в режиме истинного цвета (*True Color*)?

$20 \cdot 30 \cdot 3 \text{ байта} = \mathbf{1800}$

байт

Кодирование с палитрой



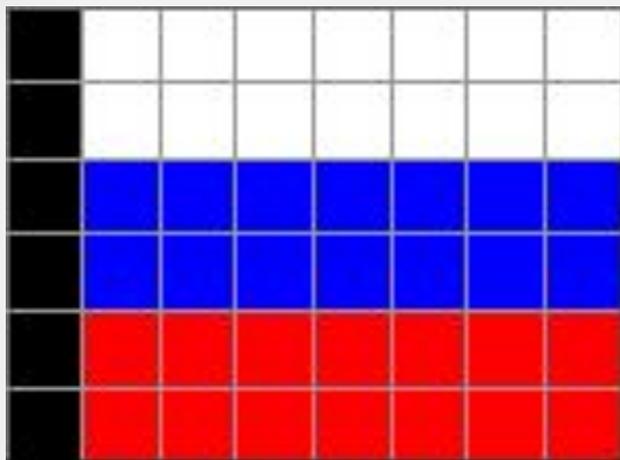
Как уменьшить размер файла?

- уменьшить разрешение
- уменьшить глубину цвета

снижается
качество

Цветовая палитра – это таблица, в которой каждому цвету, заданному в виде составляющих в модели RGB, сопоставляется числовой код.

Кодирование с палитрой



00	11	11	11	11	11	11	11
00	11	11	11	11	11	11	11
00	01	01	01	01	01	01	01
00	01	01	01	01	01	01	01
00	10	10	10	10	10	10	10
00	10	10	10	10	10	10	10

Палитра:

0	0	0	0	0	255	255	0	0	255	255	255
цвет 00 ₂			цвет 01 ₂			цвет 10 ₂			цвет 11 ₂		



Какая глубина цвета?

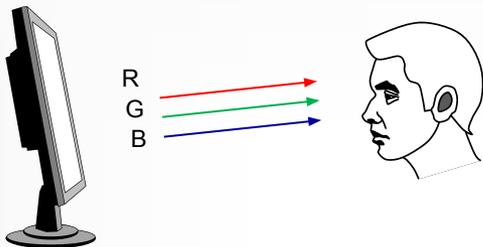
2 бита на пиксель



Сколько занимает палитра?

$3 \cdot 4 = 12$
байтов

Кодирование цвета при печати (СМУК)



Белый – красный

= голубой

C = Cyan

Белый – зелёный

= пурпурный

M = Magenta

Белый – синий

= желтый

Y = Yellow

C	M	Y
---	---	---

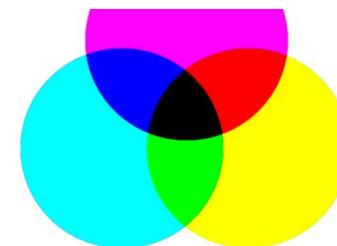
0	0	0
---	---	---

255	255	0
-----	-----	---

255	0	255
-----	---	-----

0	255	255
---	-----	-----

255	255	255
-----	-----	-----



Модель CMY

Модель СМУК: + **Key color**



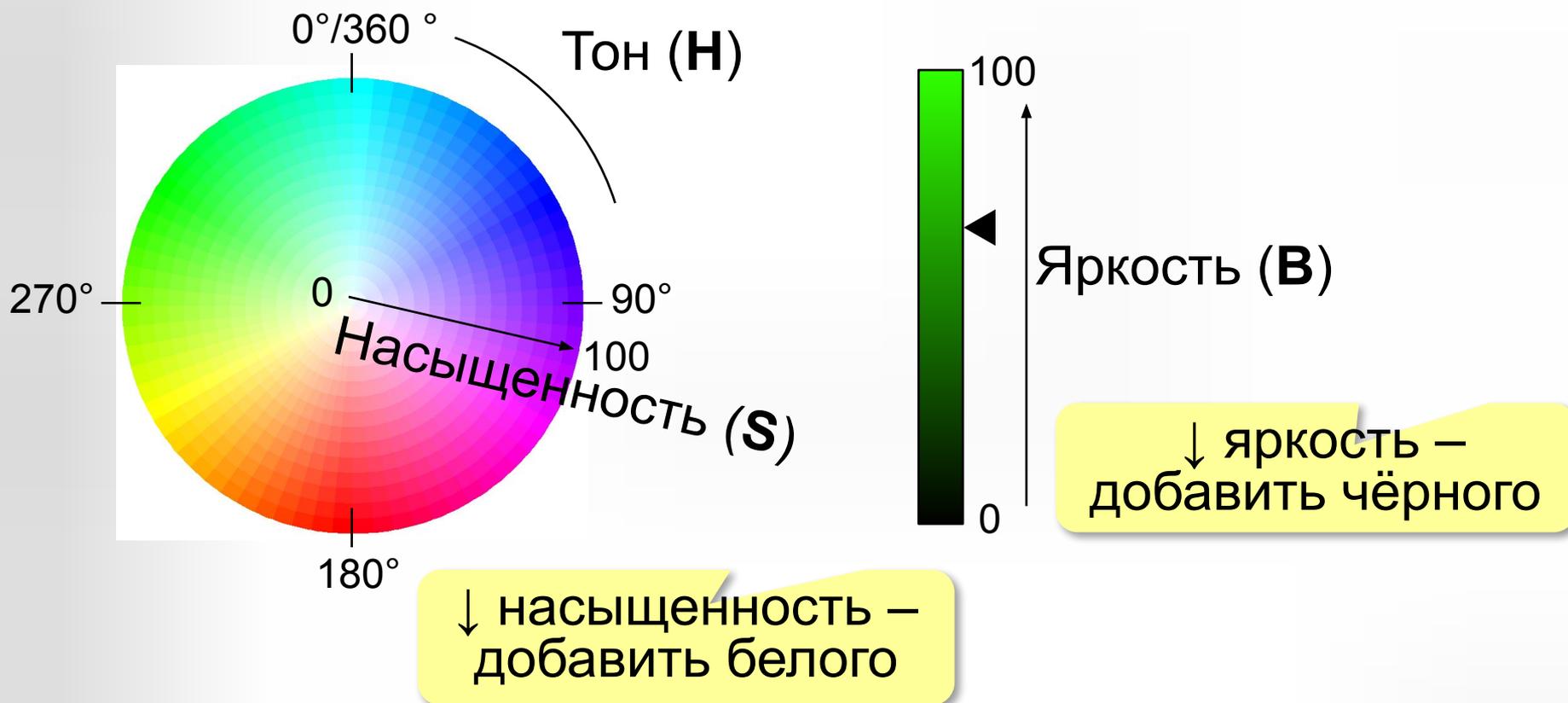
- меньший расход краски и лучшее качество для чёрного и серого цветов

Цветовая модель HSB (HSV)

HSB = *Hue* (тон, оттенок)

Saturation (насыщенность)

Brightness (яркость) или *Value* (величина)



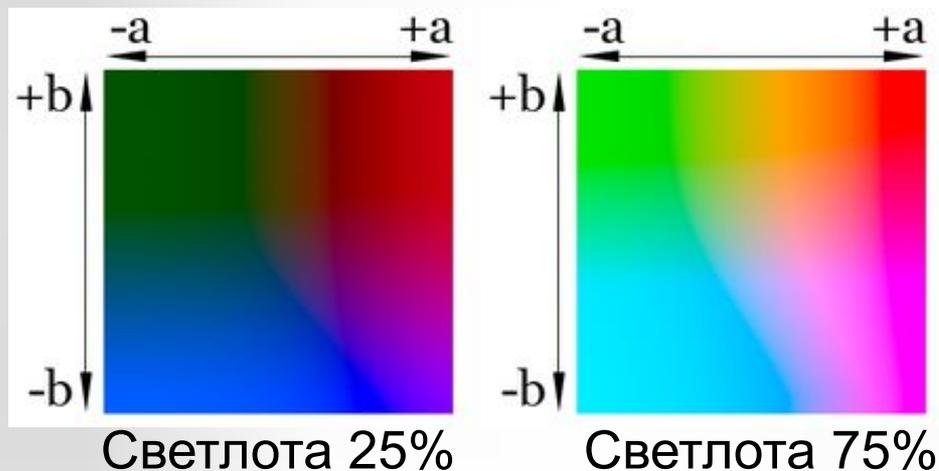
Цветовая модель Lab

Международный стандарт кодирования цвета, независимого от устройства (1976 г.)

Основана на модели восприятия цвета человеком.

Lab = *Lightness* (светлота)

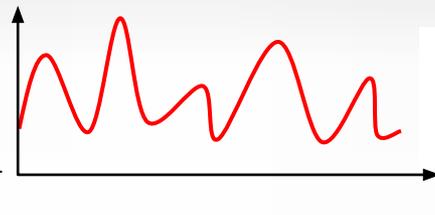
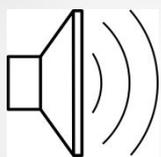
a, b (задают цветовой тон)



- для перевода между цветовыми моделями: RGB → Lab → CMYK
- для цветокоррекции фотографий

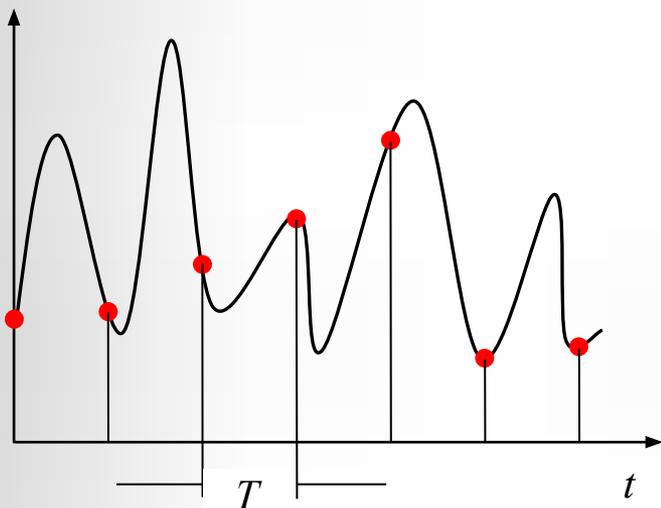
Кодирование звуковой и видеоинформации

Оцифровка звука



аналоговый
сигнал

Оцифровка – это преобразование аналогового сигнала в цифровой код (дискретизация).



$$T$$

– интервал дискретизации (с)

$$f = \frac{1}{T}$$

– частота дискретизации
(Гц, кГц)

8 кГц – минимальная частота для
распознавания речи

11 кГц, 22 кГц,

44,1 кГц – качество CD-дисков

48 кГц – фильмы на DVD

96 кГц, 192 кГц

Человек слышит

16 Гц ... 20 кГц

Оцифровка звука: квантование

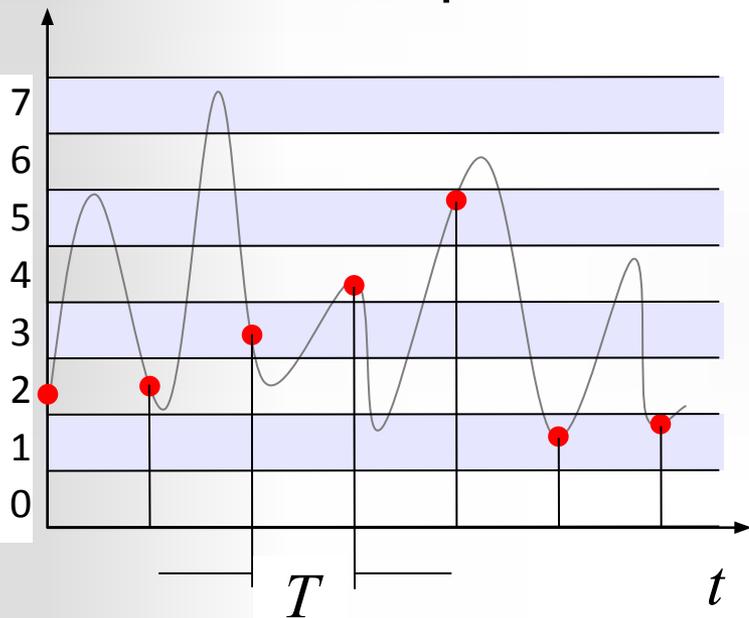


Сколько битов нужно, чтобы записать число 0,6?

Квантование (дискретизация по уровню) – это представление числа в виде цифрового кода конечной длины.

АЦП = Аналого-Цифровой Преобразователь

3-битное кодирование:



8 битов = 256 уровней

16 битов = 65536 уровней

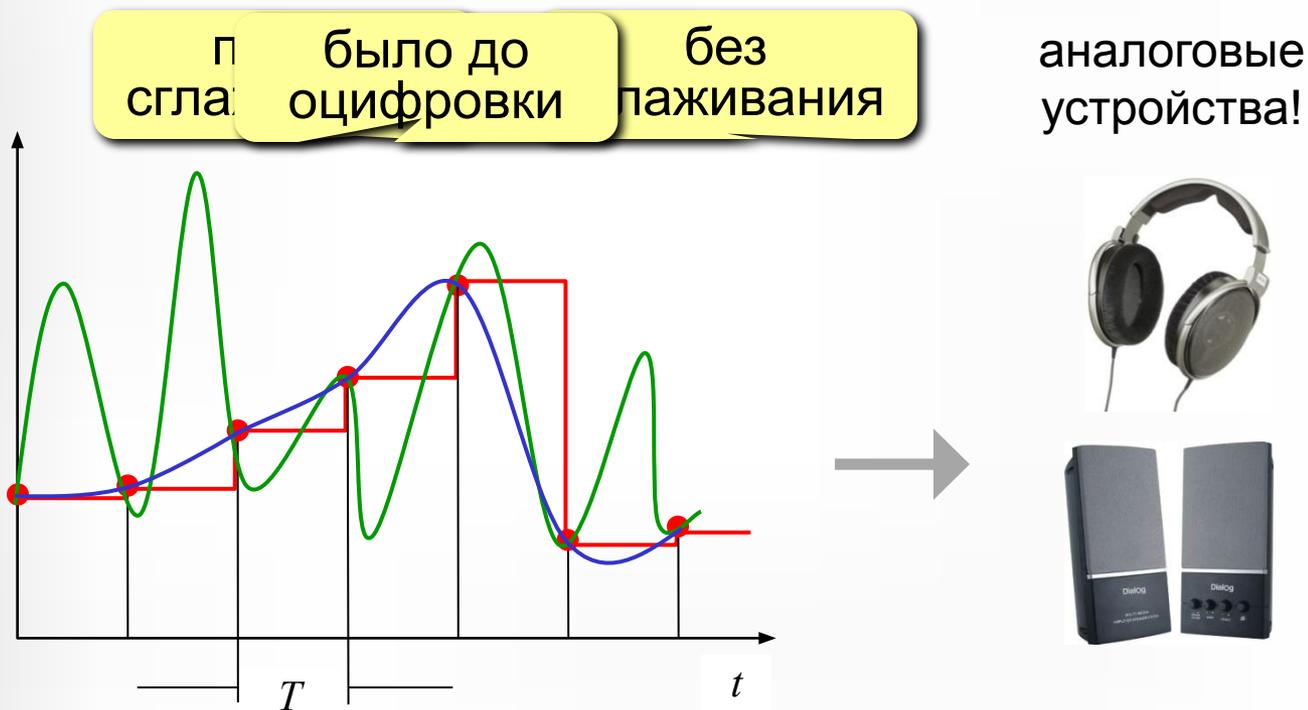
24 бита = 2^{24} уровней

Разрядность кодирования — это число битов, используемое для хранения одного отсчёта.

Оцифровка звука

Как восстановить сигнал?

ЦАП = Цифро-Аналоговый Преобразователь



? Как улучшить качество?

уменьшать T

? Что при этом ухудшится?

↑ размер файла

Оцифровка – итог

 можно закодировать **любой звук** (в т.ч. ГОЛОС, СВИСТ, шорох, ...)

 • есть **потеря информации**
• большой **объем файлов**

 Какие свойства оцифрованного звука определяют качество звучания?

Форматы файлов:

WAV (*Waveform audio format*), часто без сжатия (размер!)

MP3 (*MPEG-1 Audio Layer 3*, сжатие с учётом восприятия человеком)

AAC (*Advanced Audio Coding*, 48 каналов, сжатие)

WMA (*Windows Media Audio*, потоковый звук, сжатие)

OGG (*Ogg Vorbis*, открытый формат, сжатие)

Кодирование видео



Видео = изображения + звук

Синхронность!

изображения:

- ≥ 25 кадров в секунду
- **PAL**: 768×576, 24 бита
за 1 с: $768 \times 576 \times 25 \times 3$ байта ≈ 32 Мб
за 1 мин: 60×32 Мбайта $\approx 1,85$ Гб
- **HDTV**: 1280×720, 1920×1080.
- исходный кадр + изменения (10-15 с)
- сжатие (кодеки – алгоритмы сжатия)
- DivX, Xvid, H.264, WMV, Ogg Theora...

звук:

- 48 кГц, 16 бит
- сжатие (кодеки – алгоритмы сжатия)
- MP3, AAC, WMA, ...



Форматы видеофайлов

AVI – *Audio Video Interleave* – чередующиеся звук и видео; контейнер – могут использоваться разные *кодеки*

MPEG – *Motion Picture Expert Group*

WMV – *Windows Media Video*, формат фирмы *Microsoft*

MP4 – *MPEG-4*, сжатое видео и звук

MOV – *Quick Time Movie*, формат фирмы *Apple*

WebM – открытый формат, поддерживается браузерами