

Представление информации в компьютере

*Ганзеев П.В.
Преподаватель ГПОУ ТТТ*

Представление информации

- Компьютер может обрабатывать только информацию, представленную в **числовой форме**.
- Вся другая информация для обработки на компьютере должна быть преобразована в числовую форму.

Основы хранения информации в компьютере

- При компьютерной обработке информации приходится иметь дело
 - с числовой,
 - текстовой,
 - графической,
 - звуковой информацией.
- Для сведения всех видов информации в единую информационную основу они кодируются в последовательности нулей и единиц.
- Такое кодирование называется двоичным кодированием, а логические последовательности нулей и единиц образуют **машиночитаемый язык**.

Основы хранения информации в компьютере

- Элементарное устройство памяти компьютера, которое применяется для изображения одной двоичной цифры, называется **двоичным разрядом** или **битом**.
- Элемент памяти компьютера, состоящий из **восьми** битов, называется **байтом**.
- Каждый из восьми битов байта может содержать любую из двоичных цифр независимо от остальных.

Основы хранения информации в компьютере

- Последовательность из восьми нулей и единиц называют **двоичным числом**.
- Для хранения двоичных чисел в компьютере используется устройство, которое принято называть **ячейкой** памяти.

Кодирование информации

- **Кодирование** - это процесс установления взаимно однозначного соответствия элементам и словам одного алфавита элементов и слов другого алфавита.
- **Кодом** называется правило, по которому сопоставляются различные алфавиты и слова.

Кодирование информации

- Практически всегда основой кодирования чисел в современной ЭВМ является **двоичная система счисления**.
- **Системой счисления** называется способ записи чисел при помощи ограниченного числа символов (цифр).

Числовая информация

- Исторически первым видом данных, с которыми стали работать компьютеры, были числа.
- В соответствии с **принципами Джона Фон Неймана** ЭВМ выполняет расчёты в двоичной системе счисления.
- Числа в памяти ЭВМ хранятся в двух форматах: в **формате с фиксированной точкой** и в **формате с плавающей запятой**.
- Формат с фиксированной точкой используется для хранения в памяти целых чисел, и в этом случае число занимает 16 бит.

Чтобы получить код целого положительного десятичного числа следует:

- перевести число N из десятичной системы счисления в двоичную;
- полученный результат дополнить слева незначащими нулями до 16 разрядов.
- Например, $N = 1607_{10} = 11001000111_2$.
- Внутреннее представление этого числа в памяти компьютера будет следующим:

0000	0110	0100	0111
------	------	------	------

Числовая информация

- В сжатой 16-теричной форме этот код запишется так: 0647_{16} .
- Двоичные разряды в машинном слове нумеруются от нуля до 15 справа налево.
- Старший 15-й разряд в машинном представлении любого положительного числа равен нулю, поэтому максимальное целое число в такой форме равно:

$0111\ 1111\ 1111\ 1111_2$

Числовая информация

- Старший разряд в представлении любого **отрицательного** числа равен **1**. Следовательно, он указывает на знак числа и поэтому называется **знаковым разрядом**.
- Для перехода от десятичной системы счисления к двоичной и обратно – от двоичной к десятичной – применяются специальные правила, в подавляющем большинстве ситуаций эти переходы осуществляются компьютером автоматически.

Принципы Джона Фон Неймана

- В основу построения подавляющего большинства компьютеров положены следующие общие принципы, сформулированные в 1945 г. американским ученым Джоном фон Нейманом:
 - принцип программного управления,
 - принцип однородности памяти,
 - принцип адресности.

Символьная информация

- При хранении в компьютере любой **текст** рассматривается как линейная последовательность символов.
- Причем промежуток между отдельными символами – пробел, переход на следующую строку или страницу – также рассматриваются как специальные символы.

Символьная информация

- Для кодирования одного символа текстовой информации используется количество информации равное **1 байту** (8 бит).
- При вводе в память компьютера текстовой информации происходит ее двоичное кодирование, т. е. символ преобразуется в двоичный код.

Символьная информация

- При выводе символа на экран компьютера производится обратный процесс – **декодирование**, что означает преобразование двоичного кода символа в его изображение.
- Таким образом, различия в начертании символов преобразуются в различия между их кодами.

Символьная информация

- Множество символов, используемых на ЭВМ, для внешнего представления текста называется **символьным алфавитом** компьютера.
- Списки всех используемых при записи текстов символов и соответствующих им двоичных кодов образуют **кодовые таблицы**.

Символьная информация

- **Таблица кодировки** – это стандарт, ставящий в соответствие каждому символу алфавита свой порядковый номер: наименьший номер – 0, наибольший – 255.
- **Двоичный код символа** – это его порядковый номер в двоичной системе счисления (от 00000000 до 11111111).
- Присвоение символу конкретного кода – это вопрос соглашения, которое фиксируется кодовой таблицей (например, **ASCII**).

Таблицы кодировки

- ❑ Таблица, в которой устанавливается однозначное соответствие между символами и их порядковыми номерами, называется **таблицей кодировки**.
- ❑ Для разных типов ЭВМ используют различные таблицы кодировки:

ANSI - (American National Standards Institute)

ASCII - (American Standard Cod for Information Interchange)

Кодовые таблицы

- В практике программирования применяются различные кодовые таблицы. Наиболее часто используется кодовая таблица **ASCII** (American Standart Code for Information Interchange – стандартный американский код для обмена информацией), которая в настоящее время стала международным стандартом для персональных компьютеров.
- Международным стандартом является лишь первая (базовая) половина таблицы, т. е. **символы от 0 до 127.**

Таблица кодировки ASCII

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0																
1	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
2	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
3	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
4	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
5	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
6	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
7	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
8	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
9	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
A	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
B	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
C	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
D	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
E	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣
F	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣	␣

Код обмена информации ASCII

Первоначально – 7 бит

$N=2^7=128$ символов

0...31- всевозможные управляющие символы

32...127 – видимые на экране символы.

Сейчас – 8 бит

$N=2^8=256$ символов

128...255- национальные алфавиты,
псевдографика:

01000001 = буква A = 65

Системы кодирования

КОИ-7

Windows-1251

КОИ-8

ISO

Unicode

Символы от 0 до 127

- строчные и прописные буквы латинского алфавита,
- десятичные цифры,
- различные математические символы,
- знаки препинания,
- всевозможные скобки,
- коммерческие и другие символы и т. д.

Символы от 0 до 127

- Символы с номерами от **0** до **31** принято называть **управляющими** - управление процессом вывода текста на экран или печать, подача звукового сигнала, разметка текста и т. д.

Символы от 0 до 127

- Символ номера 32 – пробел, т. е. пустая позиция в тексте.
- Все остальные сопровождаются определенными знаками.

Кодовые таблицы

- Вторая половина кодовой таблицы может иметь различные варианты.
- В первую очередь она используется для размещения национальных алфавитов, отличных от латинского.

Кодовые таблицы

- Для кодировки **русского** алфавита – кириллицы, применяются разные варианты таблиц, поэтому возникают проблемы с переносом русского текста с одного компьютера на другой, из одной программной системы в другую.

Кодовые таблицы

- Например, согласно таблице ASCII машинный код текста, состоящего из одного слова «СИМВОЛ» следующий:
- 1001 0001 С
- 1000 1000 И
- 1000 1100 М
- 1000 0010 В
- 1000 1110 О
- 1000 1011 Л

КОИ-7

- Для представления букв русского языка (кириллицы) в рамках ASCII было предложено несколько версий.
- Первоначально был разработан **ГОСТ** под названием **КОИ-7**, оказавшийся по ряду причин крайне неудачным; ныне он практически не используется.

Unicode

- Проблема стандартизации символьного кодирования решается введением нового международного стандарта, который называется **Unicode**.
- Это 16-разрядная кодировка, т. е. в ней на каждый символ отводится 2 байта памяти. Конечно, при этом объем занимаемой памяти увеличивается в два раза. Но зато такая кодовая таблица допускает включение до 65 536 символов и в нее можно внести всевозможные национальные алфавиты.

Форматы текстовых файлов

- Существует много различных форматов – конкретных способов кодирования символов текста и фиксации элементов его оформления.
- Например, общепринятый, «понятный» большинству работающих с текстами программ формат **TXT** основывается на одной из кодовых таблиц для представления символов текста и практически не содержит никаких элементов его оформления. Именно поэтому с этим форматом могут работать очень многие программы.
- Значительно более сложным является формат **RTF** (Rich Text Format – богатый текстовый формат), который содержит очень много различных возможностей по оформлению текстов.

Кодирование графической информации

- Для обработки изображения на компьютере графическая информация должна быть представлена в цифровом формате (закодирована).
- Существует два способа кодирования графической информации – **растровый** и **векторный**. Различие между ними главным образом состоит в способе описания информации об изображении в графических файлах.



Графические
данные

растровые

векторные

В векторной графике

- все изображения описываются в виде математических объектов (примитивов): прямых линий, дуг, окружностей, эллипсов, прямоугольников, закрасок и пр.
- При таком способе компьютерная модель конкретного изображения – это математическое описание контуров, которым могут быть присвоены заливки и обводки.
- Каждый элемент представляет собой независимый объект, который можно перемещать, масштабировать и изменять.

Векторная графика

- Например, изображение древесного листа описывается точками, через которые проходит линия, создавая тем самым контур листа.
- Цвет листа задается цветом контура и области внутри этого контура.



Достоинства векторной графики:

- объекты легко трансформируются и ими просто манипулировать, что не оказывает никакого влияния на качество изображения;
- файлы, содержащие векторные изображения, невелики по размеру, т. к. хранится не само изображение, а только его основные данные, используя которые программа всякий раз воссоздает изображение заново;
- в программах векторной графики есть развитые средства интеграции изображений и текста, возможность создания конечного продукта.

Векторная графика

- используется там, где принципиальное значение имеет сохранение ясных и четких контуров.
- Сложность векторного принципа описания изображения не позволяет автоматизировать ввод графической информации.
- Кодирование зависти от прикладной среды.
- Векторный формат изображения создаётся в результате использования графических редакторов векторного типа, например, Corel Draw, Adobe Illustrator, Macromedia FreeHand.

Значительным недостатком векторной графики является

- **программная зависимость:** каждая программа сохраняет данные в своем собственном формате, поэтому изображение, созданное в одном векторном редакторе, как правило, не конвертируется в формат другой программы без погрешностей.

Растровый способ

- При использовании **растрового** способа изображение представляет собой прямоугольную матрицу точек (**пикселов**), имеющих свой цвет из заданного набора цветов (**палитры**).
- Растровое изображение чем-то напоминает мозаику, и любая графическая операция сводится к работе с отдельными элементами мозаики – пикселами.

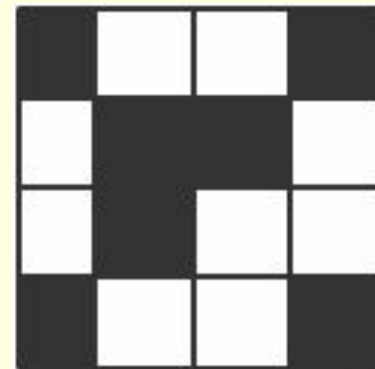
Растровая графика

- Растровая графика описывает изображения с использованием цветных точек, называемых **пикселями**, расположенных на сетке.
- Например, изображение древесного листа описывается конкретным расположением и цветом каждой точки сетки, что создает изображение примерно также, как в мозаике.

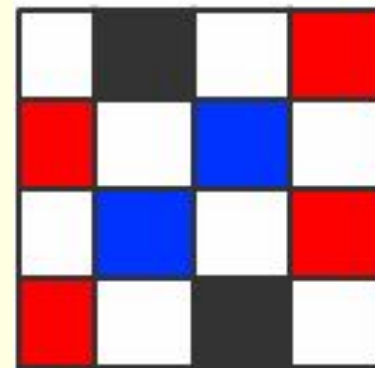


Кодирование растровых изображений

- Растровое изображение представляет собой совокупность точек (**пикселей**) разных цветов.
- Для черно-белого изображения информационный объем одной точки равен одному биту (либо черная, либо белая – либо 1, либо 0).



1 0 0 1
0 1 1 0
0 1 0 0
1 0 0 1



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

Кодирование растровых изображений

- Для представления изображения в растровом виде оно разбивается на определенное количество ячеек, и каждая ячейка обозначается единицей или нулем в зависимости от того, попало в нее изображение или нет.
- Цифры заносятся в таблицу такой же размерности. Полученная таблица называется **битовой картой**.

Разрешение

- Физический размер ячейки выражается через **разрешение** (количество пикселей на единицу длины, обычно дюйм).
- Чем разрешение больше, тем больше ячеек в битовой карте данного размера.
- При визуализации число ячеек (разрешение) исходного изображения может быть равно числу ячеек (разрешению) выводного устройства, например, монитора.

Разрешение

- Изображение в масштабе 1:1 отображается пиксел в пиксел.
- Если разрешение изображения меньше разрешения устройства вывода, то исходный пиксел отображается множеством пикселей устройства и получается размытое, ступенчатое изображение.

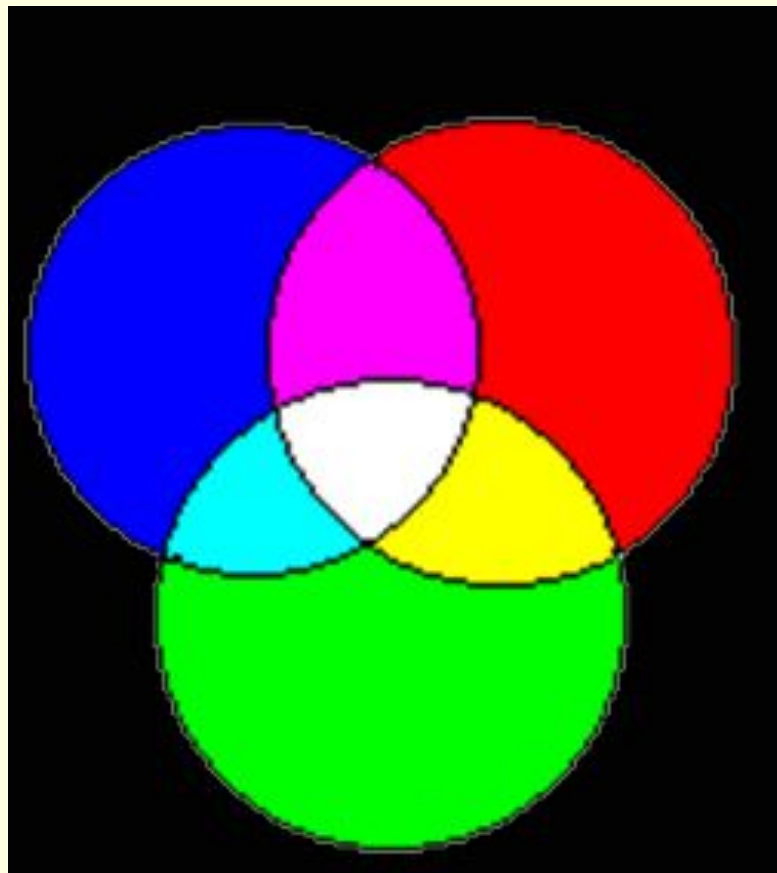
Кодирование растровых изображений

- Цветное изображение на экране монитора формируется за счет смешивания трех базовых цветов: красного, зеленого, синего (модель RGB).
- Для получения богатой палитры базовым цветам могут быть заданы различные интенсивности.

ЦВЕТОВЫЕ МОДЕЛИ

RGB (модель)

Красный (**Red**), Зеленый (**Green**) и Синий (**Blue**)



Кодирование растровых изображений

- Для четырех цветного – 2 бита.
- Для 8 цветов необходимо – 3 бита.
- Для 16 цветов – 4 бита.
- Для 256 цветов – 8 бит (1 байт).
- 4 294 967 296 цветов (True Color) – 32 бита (4 байта).
- Большое значение для растрового изображения имеет параметр **глубины цвета** – максимальное число цветов, которые могут быть в нем использованы.

Достоинства растровой графики:

- простота и, как следствие, техническая реализуемость автоматизации ввода (оцифровки) графической информации. Существует развитая система внешних устройств для ввода фотографий, слайдов, рисунков, акварелей и т. д. – сканеры, видеокамеры, цифровые фотокамеры;
- фотореалистичность: можно получить живописные эффекты используя нерезкость, размытость, цветовые переходы и т. д. Растровое изображение имеет полутона, что позволяет кодировать и отображать рисунки с фотографической точностью.

Недостатком растровых изображений является

- их пиксельный характер.
- При увеличении масштабов растрового изображения каждый пиксел увеличивается, и растр изображения становится виден, что мешает восприятию изображения.
- Любая трансформация (поворот, наклон) сопровождается искажениями.
- Хранение растровых изображений требует существенных затрат памяти компьютера и временных ресурсов для их обработки.

Графические файлы растровых типов

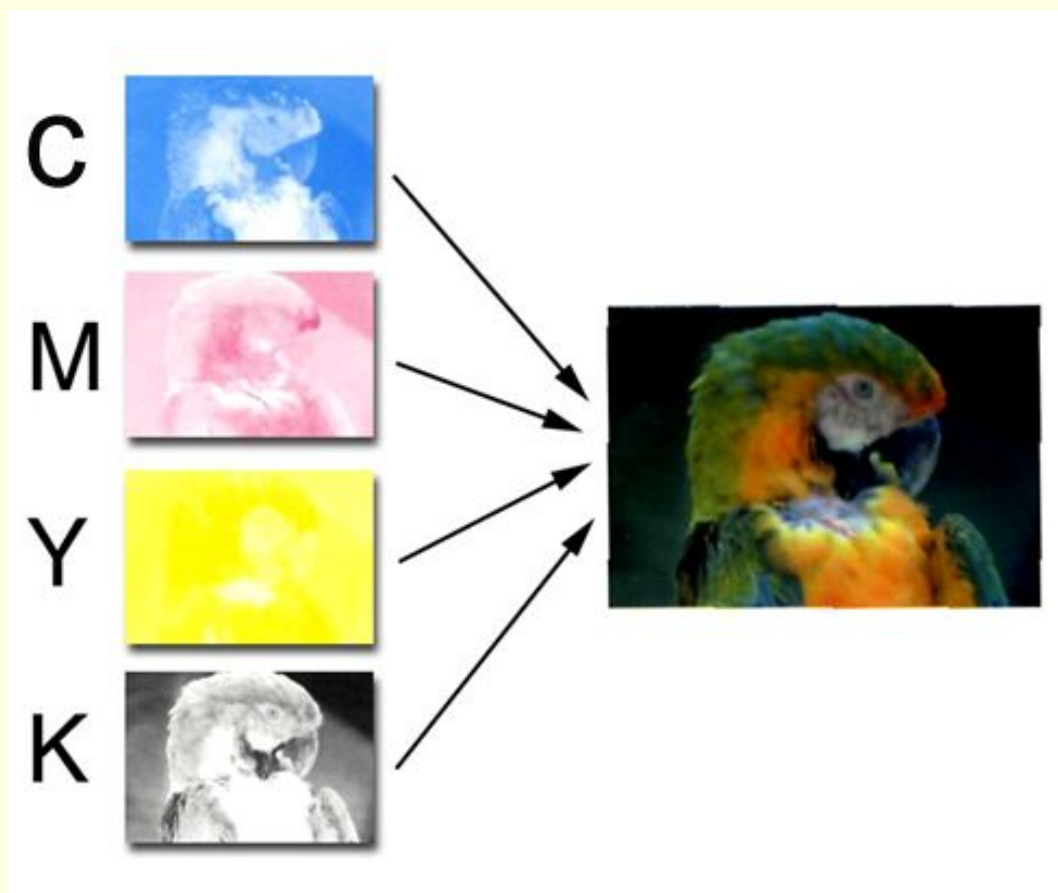
- получаются при работе с растровыми графическими редакторами, например, Paint, Adobe PhotoShop.
- Форматы файлов, предназначенные для сохранения точечных изображений, являются стандартными, поэтому не имеет решающего значения, в каком графическом редакторе создано то или иное изображение. Файл, сохраняющий точечное изображение, легко открывается и импортируется в редакторах точечной и векторной графики.

Форматы графических файлов

- Данные об изображении хранятся в графических файлах.
- Способ организации графических файлов называется *графическим форматом*.
- Размер графического файла сильно зависит от формата, выбранного для хранения изображения.
- Знание графических форматов и их возможностей является одним из ключевых факторов в допечатной подготовке изданий, подготовке изображений для Web и в компьютерной графике вообще.

CMYK

Голубой (**C**yan), Пурпурный (**M**agenta) Желтый (**Y**ellow) Черный (**K**Black)



ФОРМАТЫ ГРАФИЧЕСКИХ ФАЙЛОВ



Gif-глубина цвета = 8 бит,
 $2^8=256$



JPG, JPEG -глубина цвета = 24
бит,
 $2^{24}=16777216$



PNG-глубина цвета до 48 бит

ФОРМАТЫ ГРАФИЧЕСКИХ ФАЙЛОВ



TIFF-глубина цвета = 32 бит,
 $2^{32}=4294967296$



EPS-глубина цвета = 32 бит,
 $2^{32}=4294967296$



BMP



PSD, PDD

Сравнение векторной и растровой графики

	Основной элемент изображения	Цвет и форма	Основное применение
растровая графика	точка(в экранном изображении пиксел)	Цвет и форма неотделимы, но цвет первичен, а форма является производной от цвета (без <i>цвета</i> , форма не существует)	Создание фотореалистических изображений с тонкими цветовыми переходами - это портрет, пейзаж, живописный коллаж

Сравнение векторной и растровой графики

	Основной элемент изображения	Цвет и форма	Основное применение
векторная графика	линия(контур) прямая или кривая	Цвет и форма независимы друг от друга, и форма первична, а цвет - заполнитель формы (без <i>контура</i> , цвет не существует)	Используются для отображения объектов с чёткой границей и ясными деталями - это шрифт, логотип, графический знак, орнамент, декоративная композиция

Кодирование графической информации

Растровое изображение представляет собой однослойную сетку точек, называемых пикселами (pixel, от англ. picture element). Код пиксела содержит информации о его цвете.

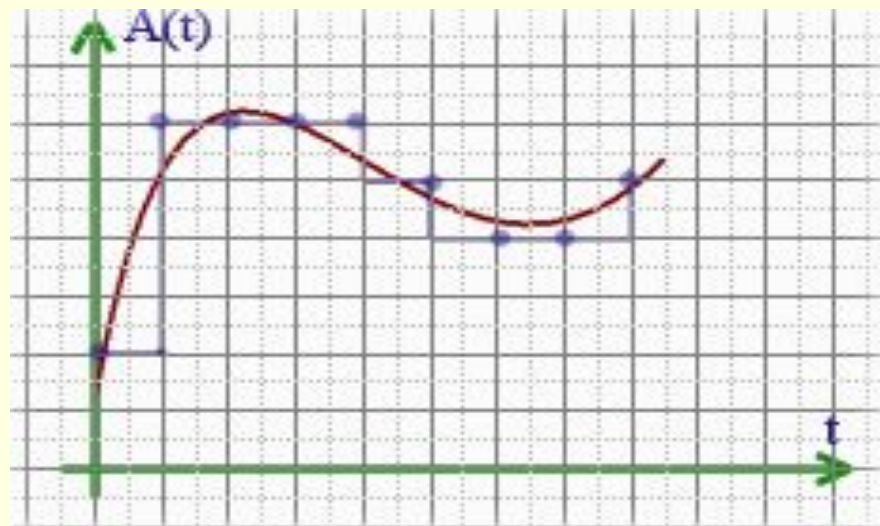
Векторное изображение многослойно. Каждый элемент векторного изображения - линия, прямоугольник, окружность или фрагмент текста - располагается в своем собственном слое, пиксеты которого устанавливаются независимо от других слоев.

Объем графического файла в битах определяется как произведение количества пикселей N на разрядность цвета M (битовую глубину)

$$N * M$$

Двоичное кодирование звука

- **Звук** – волна с непрерывно изменяющейся амплитудой и частотой.
- Чем больше амплитуда, тем он громче для человека, чем больше частота, тем выше тон.



Кодирование звука

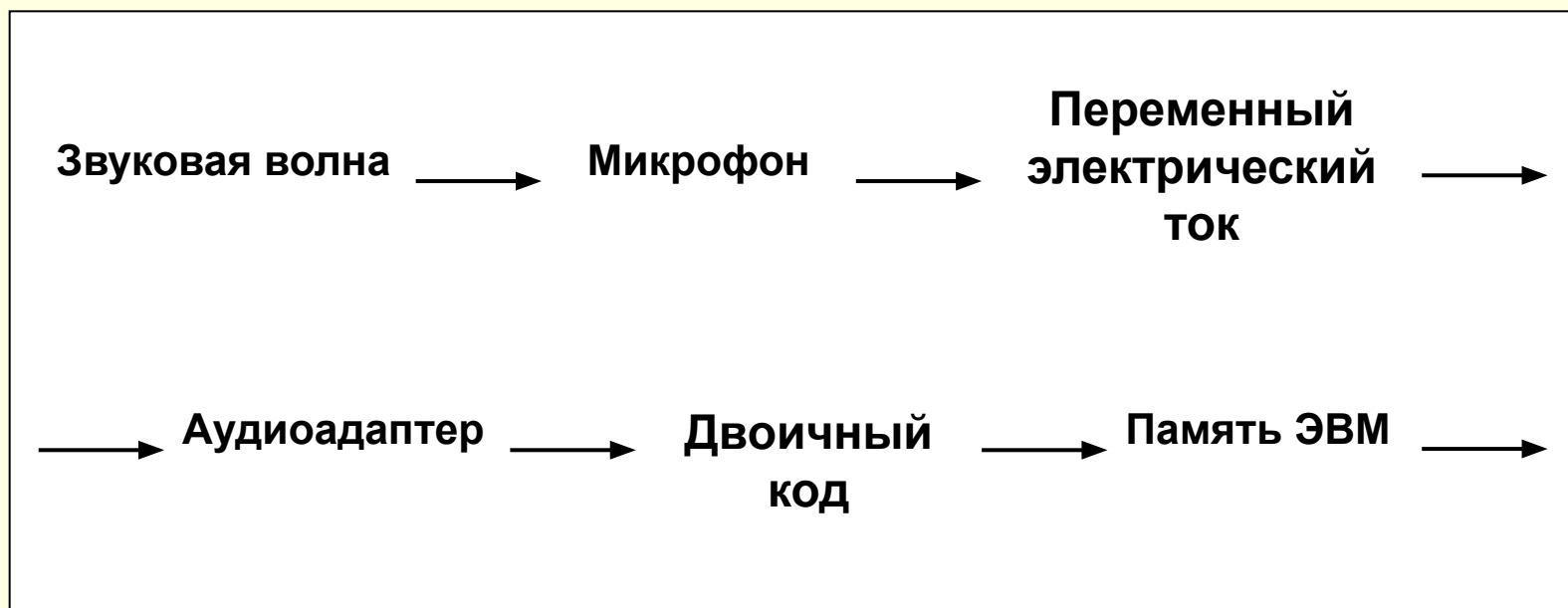
- Современные компьютеры «умеют» сохранять и воспроизводить звук (речь, музыку и пр.). Звук, как и любая другая информация, представляется в памяти ЭВМ в форме **двоичного** кода.
- Звук является обязательной компонентой мультимедийных продуктов. Кодирование аудиосигнала основано на временной дискретизации.

Кодирование звука

- В процессе кодирования звукового сигнала производится его временная **дискретизация** – непрерывная волна разбивается на отдельные маленькие временные участки.
- Качество двоичного кодирования звука определяется **глубиной кодирования** и **частотой дискретизации**.

Процесс преобразования звуковых волн в памяти компьютера

- Физическая природа звука – это колебания в определённом диапазоне частот, передаваемых звуковой волной, через воздух (или другую упругую среду).



Процесс воспроизведения звуковой информации, сохраненной в памяти компьютера:



Кодирование звука

- Результат кодирования и декодирования зависит от аудиоадаптера.
- **Аудиоадаптер** – звуковая плата, подключенная к компьютеру, предназначенная для преобразования электрических колебаний звуковой частоты в числовой двоичный код при вводе звука и для обратного преобразования при воспроизведении.

Кодирование звука

- В процессе записи звука аудиоадаптер с определённым периодом измеряет амплитуду электрического тока и заносит в регистр двоичный код полученной величины.
- Полученный код из регистра переписывается в оперативную память компьютера.

Кодирование звука

- Качество компьютерного звука определяется характеристиками аудиоадаптера: **частотой дискретизации и разрядностью.**
- **Частота дискретизации** – это количество измерений входного сигнала за 1 секунду. Частота измеряется в герцах (Гц). Характерные частоты – 11; 22; 44,1.

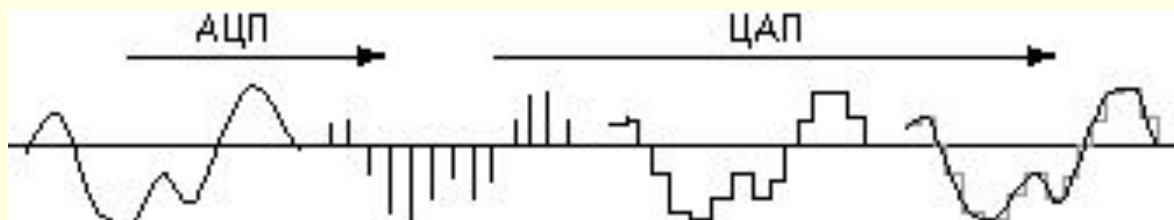
Кодирование звука

- **Разрядность регистра** – число бит в регистре аудиоадаптера. Разрядность определяет точность измерений входного сигнала. Чем больше разрядность, тем меньше погрешность каждого отдельного преобразования величины электрического сигнала в число и обратно. Если разрядность 8 (16), то при измерении входного сигнала может быть получено 28 либо 216 различных значений.
- **Звуковой файл** хранит звуковую информацию в числовой двоичной форме. Как правило, звуковая информация подвергается сжатию.

Кодирование звука*

Звук - это колебания воздуха

Процесс преобразования аналогового сигнала в последовательность двоичных чисел называется **дискретизацией (или оцифровкой)**, а устройство, выполняющее его - **аналого-цифровым преобразователем (АЦП)**.



Для того чтобы воспроизвести закодированный таким образом звук, нужно выполнить обратное преобразование (для него служит *цифро-аналоговый преобразователь* -- ЦАП), а затем сгладить получившийся ступенчатый сигнал.

Кодирование видеоинформации

Число кадров вычисляется как произведение длительности видеоклипа на скорость кадров, то есть их количество в 1 с

$$V = N * M * C * v * \Delta t$$

При разрешении 800*600 точек, разрядности цвета C=16, скорости кадров v=25 кадров/с, видеоклип длительностью 30 с будет иметь объем:

$$V=800*600*16*25*30=576*10^7(\text{бит})=72*10^7(\text{байт})=687(\text{Мбайт})$$