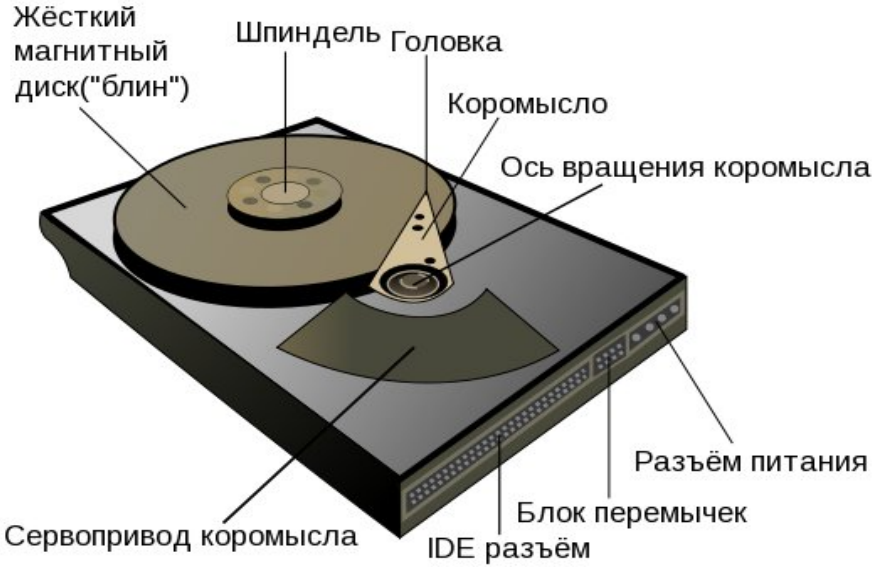
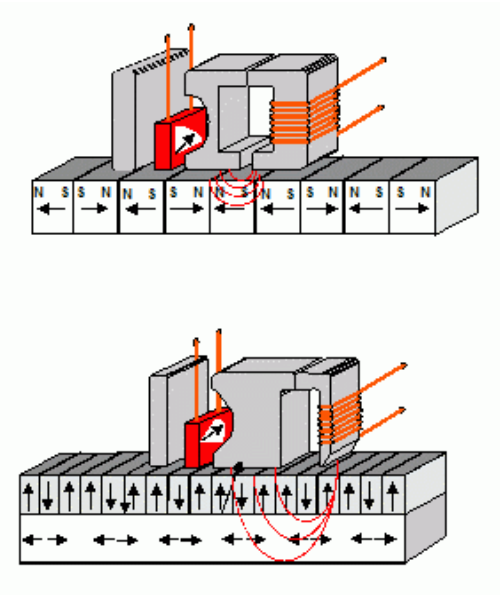
1 Аппаратная часть дисков.

1.1 Магнитные диски.

Более подробная информация - http://ru.wikipedia.org/wiki/Жёсткий_диск



Устройство жёсткого диска с IDE разъемом.

	
Головка HDD	Продольная и перпендикулярная запись информации

С 2005 по 2010 годы всеми производителями осуществлён переход с продольной на перпендикулярную запись данных на диск, что обеспечивает большую плотность записи данных.

С 2011 начат переход на "тепловую магнитную запись", место записи предварительно нагревается лазером, что уменьшает размер домена и повышает надёжность хранения. Предполагаемая емкость 50 ТБ.

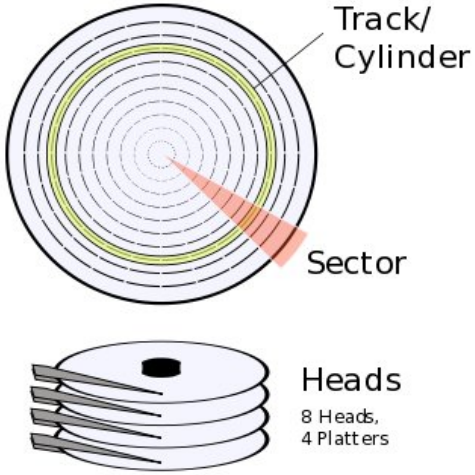
Основные понятия:

Головка (Head) - электромагнит, скользящий над поверхностью диска, для каждой поверхности используется своя головка. Нумерация начинается с 0.

Дорожка (Track) - концентрическая окружность, которое может прочитать головка в одной позиции. Нумерация дорожек начинается с внешней (первая имеет номер - 0).

Цилиндр (Cylinder) - совокупность всех дорожек с одинаковым номером на всех дисках, т.к. дисков может быть много и на каждом диске запись может быть с двух сторон.

Маркер - от него начинается нумерация дорожек, есть на каждом диске.

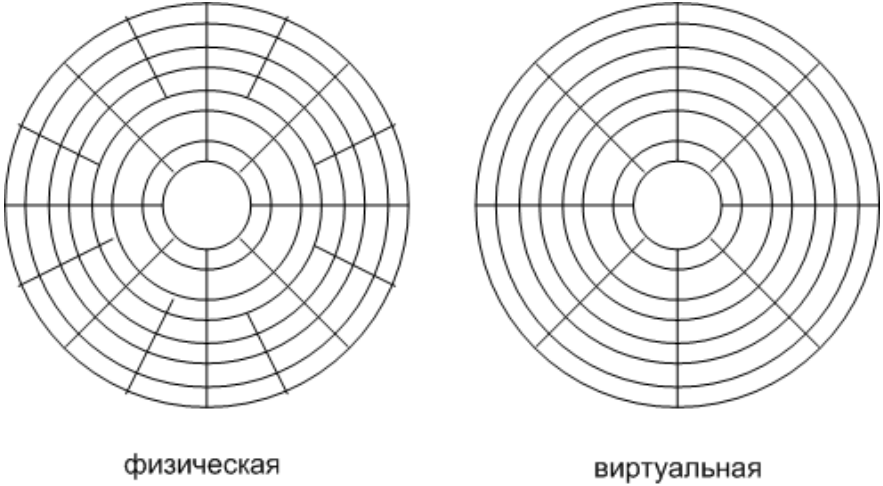


Сектор - на сектора разбивается каждая дорожка, сектор содержит минимальный блок информации. Нумерация секторов начинается от маркера.

Геометрия жесткого диска - набор параметров диска, количество головок, количество цилиндров и количество секторов.

У современных жестких дисков **контроллер** встроен в само устройство, и берёт на себя большую часть работы, прозрачно для ОС.

Например, контроллеры скрывают физическую геометрию диска, предоставляя виртуальную геометрии.



Физическая и виртуальная геометрия диска

На внешних дорожках число секторов делают больше, а на внутренних меньше. На реальных дисках таких зон может быть несколько десятков.

1.2. RAID.

Redundant Array of Independent Disk - массив независимых дисков с избыточностью. Более подробная информация - http://ru.wikipedia.org/wiki/RAID

Для увеличения производительности и/или надежности операций ввода-вывода с диском был разработан стандарт для распараллеливания и дублирования этих операций.

На практике, как правило, используют RAID 0, 1 и 5. Основные шесть уровней RAID (существуют и другие):

- **RAID 0 - чередующий набор**, соединение нескольких дисков в один большой логический диск, но логический диск разбит так, что запись и чтение происходит сразу с несколько дисков. Например, записываем блок 1, 2, 3, 4, 5, каждый блок будет записываться на свой диск.

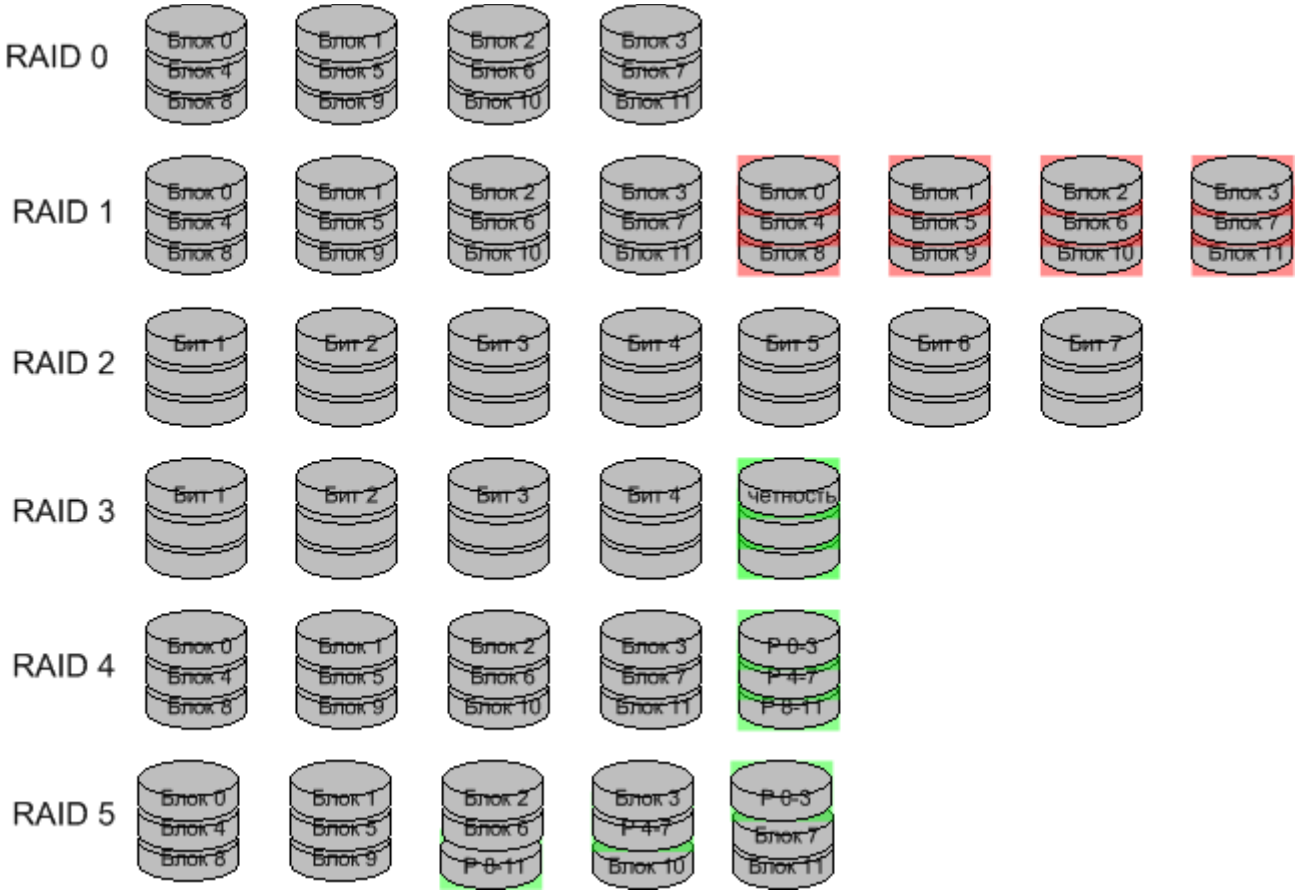
  Преимущества: удобство одного диска; увеличивает скорость записи и чтения.

  Недостатки: уменьшает надежность (в случае выхода одного диска, массив будет разрушен), т.к. избыточность не предусмотрена.

- **RAID 1 - зеркальный набор**, параллельная запись и чтение на несколько дисков с дублированием (избыточность).

  Преимущества: дублирование записей; увеличивает скорость чтения (но не записи).

Недостатки: требует в два раза больше дисковых накопителей.


RAID 0
Блок 0
Блок 4
Блок 8
Блок 1
Блок 5
Блок 9
Блок 2
Блок 6
Блок 10
Блок 3
Блок 7
Блок 11
RAID 1
RAID 2
Бит 1
Бит 2
Бит 3
Бит 4
Бит 5
Бит 6
Бит 7
RAID 3
Четность
RAID 4
P 0-3
P 4-7
P 8-11
RAID 5

- **RAID 2** - работает на уровне слов и даже байт. Например, берется полбайта (4 бита) и прибавляется 3 бита четности (1, 2, 4 - рассчитанные по Хэммингу), образуется 7-битовое слово. В случае семи дисков слово записывается побитно на каждый диск. Так как слово пишется сразу на все диски, они должны быть синхронизированы.

  Преимущества: надежность; увеличивает скорость записи и чтения (при потоке, но при отдельных запросах не увеличивает).

  Недостатки: нужна синхронизация дисков.

- **RAID 3** - упрощенная версия RAID 2, для каждого слова считается только один бит четности.

  Преимущества: надежность; увеличивает скорость записи и чтения (при потоке, но при отдельных запросах не увеличивает).

  Недостатки: нужна синхронизация дисков.

- **RAID 4** - аналогичен уровню RAID 0, но с добавлением диска четности. Если любой из дисков выйдет из строя, его можно восстановить с помощью диска четности.

  Преимущества: надежность; не нужна синхронизация дисков.

Недостатки: не дает увеличения производительности, узким местом становится диск четности при постоянных пересчетах контрольных сумм.

- **RAID 5** - аналогичен уровню RAID 4, но биты четности равномерно распределены по дискам.

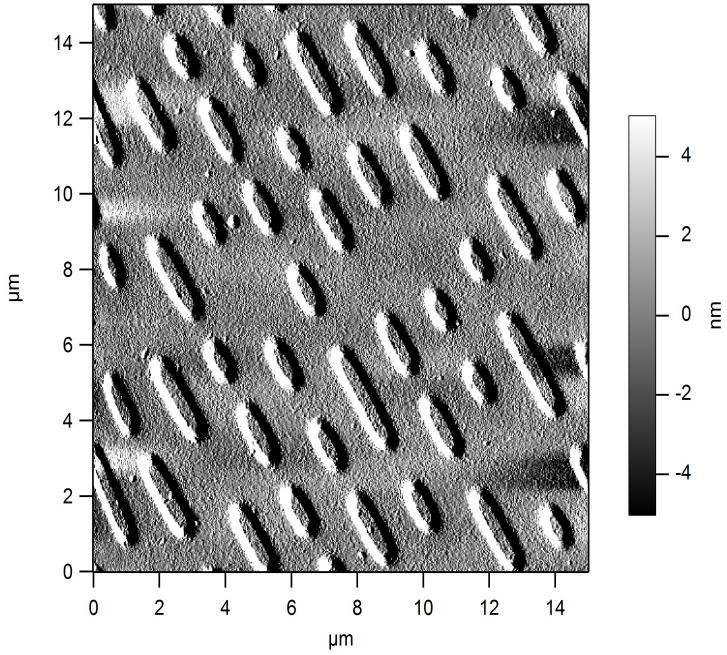
1.3 Компакт-диски.

Более подробная информация - http://ru.wikipedia.org/wiki/Компакт-диск

	
Фото диска	Фото устройства для работы с дисками

© Yuriy Shamshin

Запись на CD-ROM диски производится с помощью штамповки.



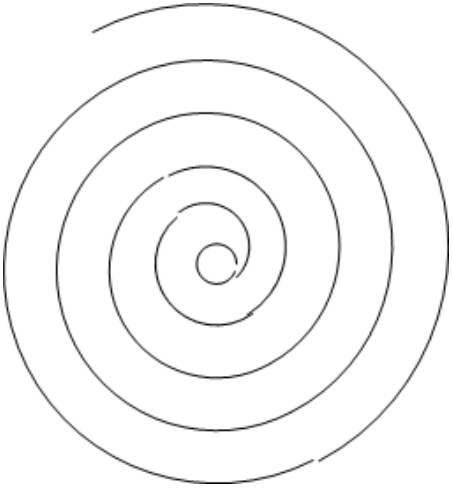
CD-ROM под электронным микроскопом. Длина пита варьируется от 850 нм до 3,5 мкм.

Пит - единица записи информации (впадина при штамповке, темное пятно, прожженное в слое краски в CD-R, область фазового перехода)

© Yuriy Shamshin

1.3.1. "Красная книга".

Сначала CD-диски использовались только для записи звука, стандарт которого был описан ISO 10149.

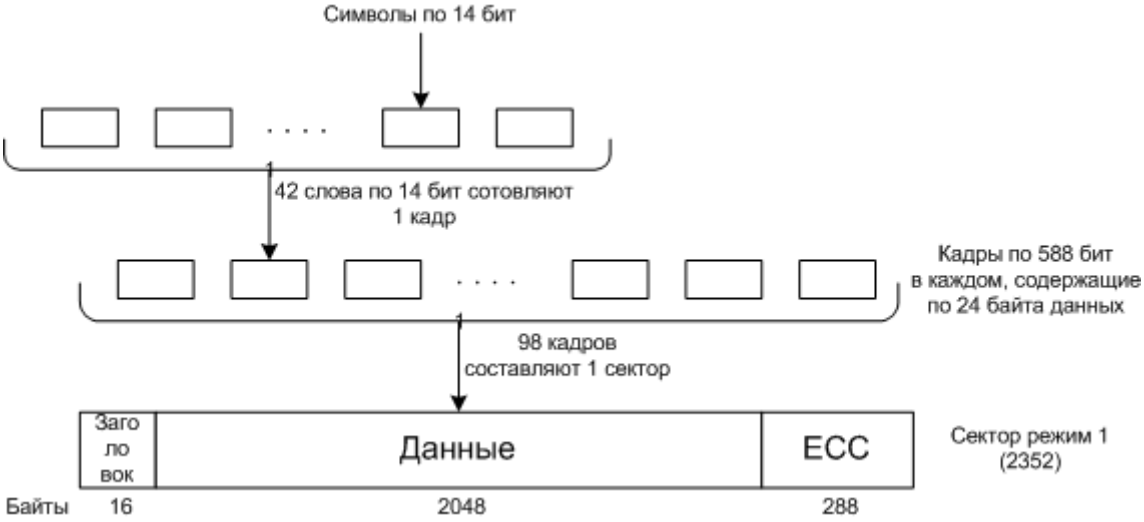


Запись на CD-ROM производится спирально

1.3.2. "Желтая книга".

В 1984 году был описан следующий стандарт и добавлена запись данных.

Для записи данных было необходимо повысить надежность, для этого каждый байт (8 бит) стали кодировать в 14 разрядное число (по размеру почти дублирование записи, но за счет кодирования эффективность может быть, как при тройной записи), чтобы можно было восстановить потерянные биты.



Логическое расположение данных на CD-ROM для режима 1

Заголовок содержит:

© Yuriy Shamshin

- Первые 12 байт 00FFFFFFFFFFFFFFFFFFFF00, чтобы считывающее устройство могло распознать начало сектора.
- Следующие три байта содержат номер сектора.
- Последний байт содержит код режима.

ECC (Error Correction Code) - код исправления ошибок в режиме 1 (используют для данных).

В режиме 2 поле данных объединено с полем ECC в 2336-байтное поле данных. Этот режим можно использовать, если не требуется коррекция ошибок, например, видео и аудио запись.

Коррекция ошибок осуществляется на трех уровнях: внутри слова, в кадре, в CD-ROM-секторе, поэтому 7203 байта содержат только 2048 байта полезной нагрузки, около 28%.

1.3.3. "Зеленая книга".

В 1986 году к стандарту была добавлена графика, и возможность совмещения в одном секторе аудио, видео и данных. Стандарт ISO 9660.

Файловая система для CD-ROM называется **High Sierra**. Файловая система имеет три уровня:

- 1 уровень - файлы имеют имена формата, схожего с MS-DOS - 8 символов имя файла плюс до трех символов расширения, файлы должны быть непрерывными. Глубина вложенности каталогов ограничена восемью. Этот уровень понимают почти все операционные системы.
- 2 уровень - имена файлов могут быть до 32 символов, файлы должны быть непрерывными.
- 3 уровень - позволяет использовать сегментированные файлы.

© Yuriy Shamshin

Для этого стандарта существуют расширения: Rock Ridge - позволяет использовать длинные файлы, а также UID, GID и символические ссылки.

© Yuriy Shamshin

1.3.4. Компакт-диски с возможностью записи CD-R.

Запись на CD-R диски производятся с помощью локального прожигания нанесенного слоя красителя.

Используются лазеры с двумя уровнями разной мощности, для записи 8-16 мВт, для чтения 0.5 мВт.

В 1989 году была выпущена "**Оранжевая книга**", это документ определяет формат CD-R, а также новый формат **CD-ROM XA** , который позволяет посекторно дописывать информацию на CD-R.

CD-R-дорожка - последовательно записанные за один раз секторы. Для каждой такой дорожки создается свой VTOC (Volume Table of Contents - таблица содержания тома), в котором перечисляются записанные файлы.

Каждая запись производится за одну непрерывную операцию, поэтому если у вас будет слишком загружен компьютер (мало памяти или медленный диск), то вы можете испортить диск, т.к. данные не будут поспевать поступать на CD-ROM.

1.3.5. Многократно перезаписываемые компакт-диски CD-RW.

Запись на CD-RW диски производятся локального перевода слоя из кристаллического в аморфное состояние.

© Yuriy Shamshin

Используются лазеры с тремя уровнями разной мощности. Эти диски можно отформатировать (UDF), использовать их в место дискет и дисков.

1.3.6. Универсальный цифровой диск DVD (Digital Versatile Disk).

Более подробная информация - http://ru.wikipedia.org/wiki/DVD

Были сделаны следующие изменения:

- Размер пита уменьшили в два раза (с 0.8 мкм до 0.4мкм)
- Более тугая спираль (0.74 мкм между дорожками, вместо 1.6 у компакт-дисков)
- Уменьшение длины волны лазера (650 нм вместо 780 нм)

Это позволило увеличить объем с 650 Мбайт до 4.7 Гбайт.

Определены четыре следующих формата:

1. Односторонний, одноуровневый (4.7 Гбайт)
2. Односторонний, двухуровневый (8.5 Гбайт), размеры пита второго уровня приходится делать больше, иначе не будут считаны, т.к. первый полуотражающий слой половину потока отразит и частично рассеет.
3. Двухсторонний, одноуровневый (9.4 Гбайт)
4. Двухсторонний, двухуровневый (17 Гбайт)

© Yuriy Shamshin

1.3.7. Универсальный цифровой диск Blu-ray (blue ray — синий).

Более подробная информация - http://ru.wikipedia.org/wiki/Blu-ray_Disc

Были сделаны следующие изменения:

- Размер пита уменьшили
- Более тугая спираль (0,32 мкм между дорожками, вместо 0.72 у DVD)
- Уменьшение длины волны лазера (405 нм вместо 650 нм в DVD), «синего» (технически сине-фиолетового) лазера, отсюда и название

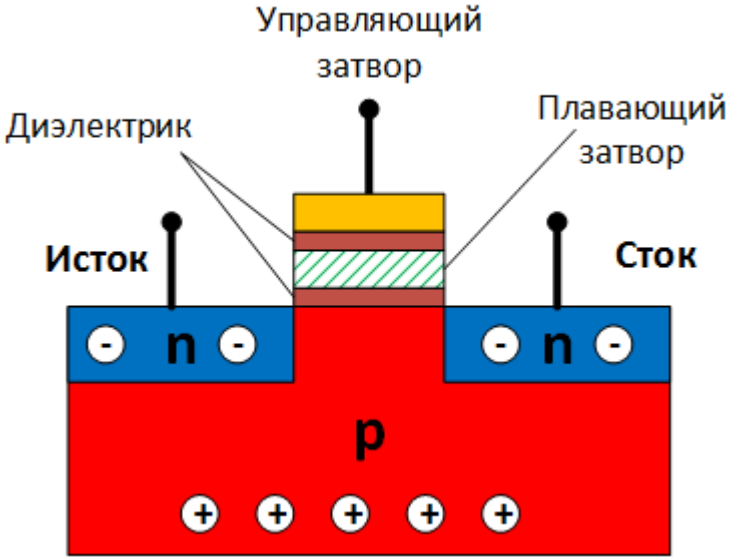
Определены следующие формата:

1. однослойный диск 23,3/25/27 или 33 Гб
2. двухслойный диск 46,6/50/54 или 66 Гб
3. четырёх слойный 100 Гб
4. восьми слойный 200 Гб

© Yuriy Shamshin

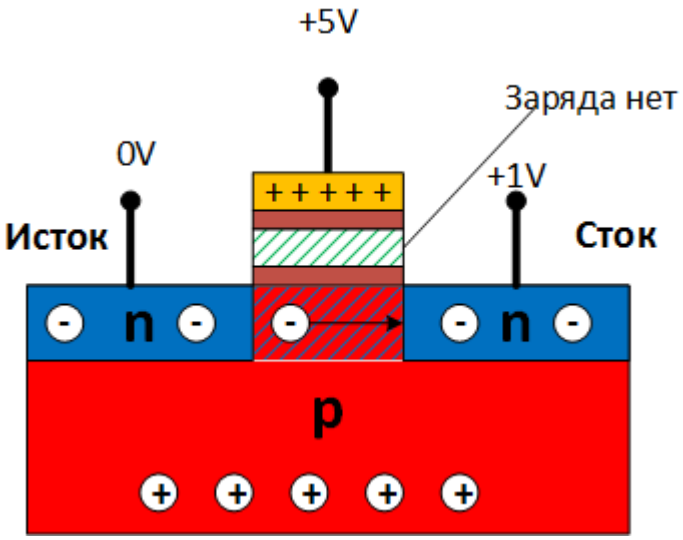
1.4 Твердотельные накопители (Flash, SSD, ...).

Устройство ячейки памяти: Используются полевые транзисторы с плавающим затвором.

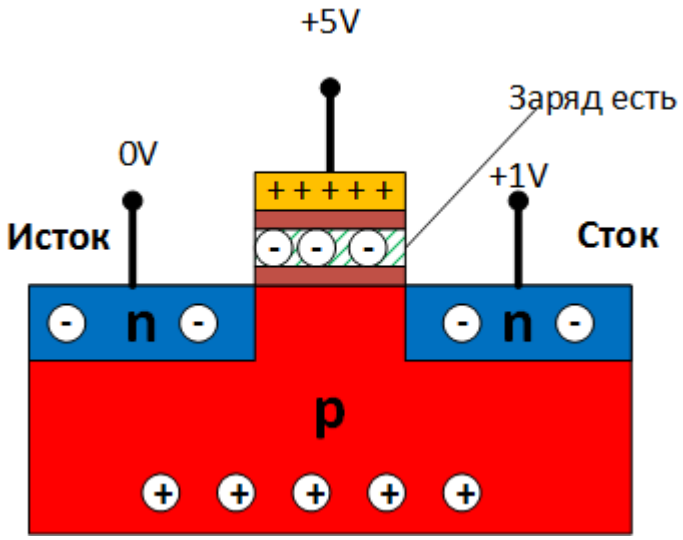


Устройство ячейки памяти

© Yuriy Shamshin



Считывание "0"

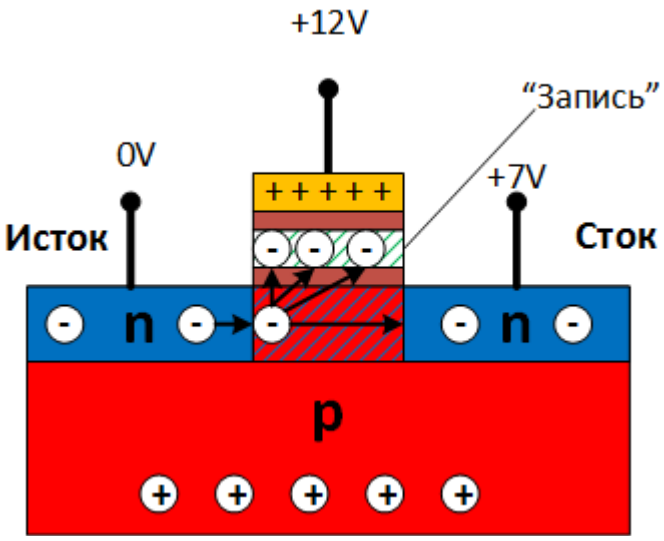


Считывание "1"

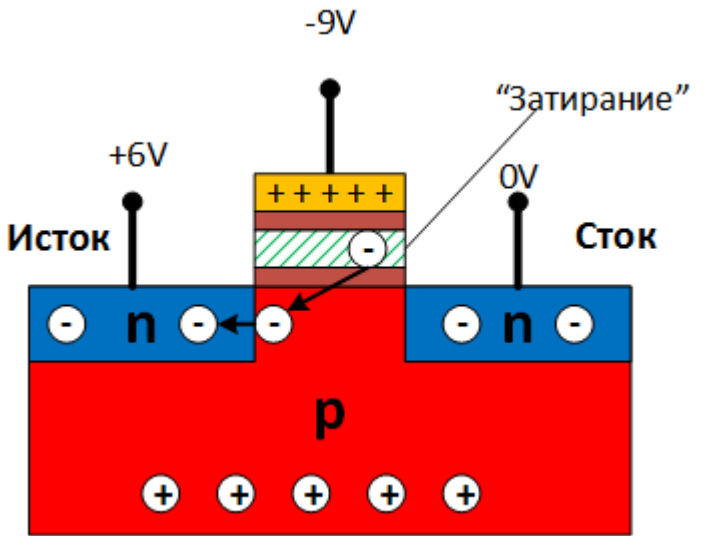
Считывание информации 0: Если ток через npn-переход идет, то "считывается 0". Ток идет за счет туннельного эффекта, который возникает под действием управляющего затвора, на который подается "+".

Считывание информации 1: Если ток через npn-переход не идет, то "считывается 1". Ток не идет за счет "экранирования" управляющего затвора плавающим затвором, на котором накоплен "-".

© Yuriy Shamshin



Запись информации



Затирание информации

Запись информации: "Запись" делается накоплением электронов в плавающем затворе, за счет повышенного напряжения на управляющем затворе и стоке.

© Yuriy Shamshin

Затирание информации: "Затирание" делается "изъятием" электронов из плавающего затворе, за счет положительного напряжения на истоке и отрицательного на управляющем затворе, на стоке 0В.

© Yuriy Shamshin

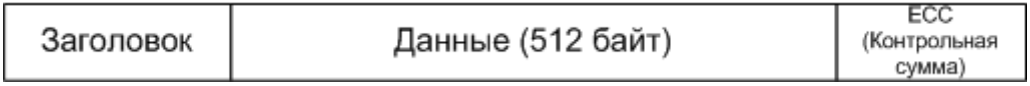
2. Программная часть дисков (Форматирование дисков).

2.1. Низкоуровневое форматирование.

Низкоуровневое форматирование - разбивка диска на сектора, производится производителями дисков. При низкоуровневом форматировании часть полезного объема диска, доступная пользователю, уменьшается, примерно до 80%.

Фактический размер сектора 571 байт. Каждый сектор состоит из:

- Заголовка (Prefix portion) - по которому определяется начало (последовательность определенных битов) сектора и его номер, и номер цилиндра.
- Область данных (размер блока, как правило, 512 байт, с2010 г. происходит переход на 4 Кб. При низкоуровневом (физическом) форматировании всем байтам данных присваивается некоторое значение, например F6h.
- Конец сектора (Suffix portion) - содержит контрольную сумму **ECC** (Error Correction Code - код корректировки ошибок). Позволяет обнаружить или даже исправить ошибки чтения. Размер зависит от производителя, и показывает, как производитель относится к надежности работы диска.



Сектор (блок) диска

© Yuriy Shamshin

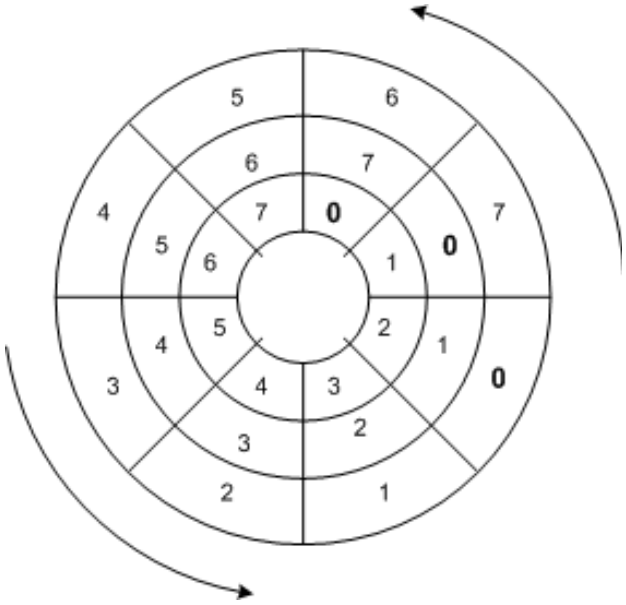
2.1.1. Промежутки между секторами.

Кроме того существуют **промежутки между секторами** на каждой дорожке и между самими дорожками. Промежутки нужны, например, для того, чтобы при переходе к следующему сектору завершился анализ контрольной суммы, см. ниже перекосы и чередование.

2.1.2. Перекос цилиндров.

Перекос цилиндров - сдвиг 0-го сектора каждой последующей дорожки, относительно предыдущей. служит для увеличения скорости. Головка тратит, какое то время на смену дорожки, и если 0-й сектор будет начинаться в том же месте, что и предыдущий, то головка уже проскочит его, и будет ждать целый круг.

Перекос цилиндров делают разным в зависимости скоростей вращения и перемещения головок.



2.1.3. Перекос головок.

Перекос головок - приходится применять, т.к. на переключение с головки на головку тратится время.

© Yuriy Shamshin

2.1.4. Чередование секторов.

Если, например, один сектор прочитан, а для второго нет в буфере места, пока данные копируются из буфера в память, второй сектор уже проскочит головку.

Чтобы этого не случилось, применяют чередование секторов. Если копирование очень медленное, может применяться двукратное чередование, или больше.

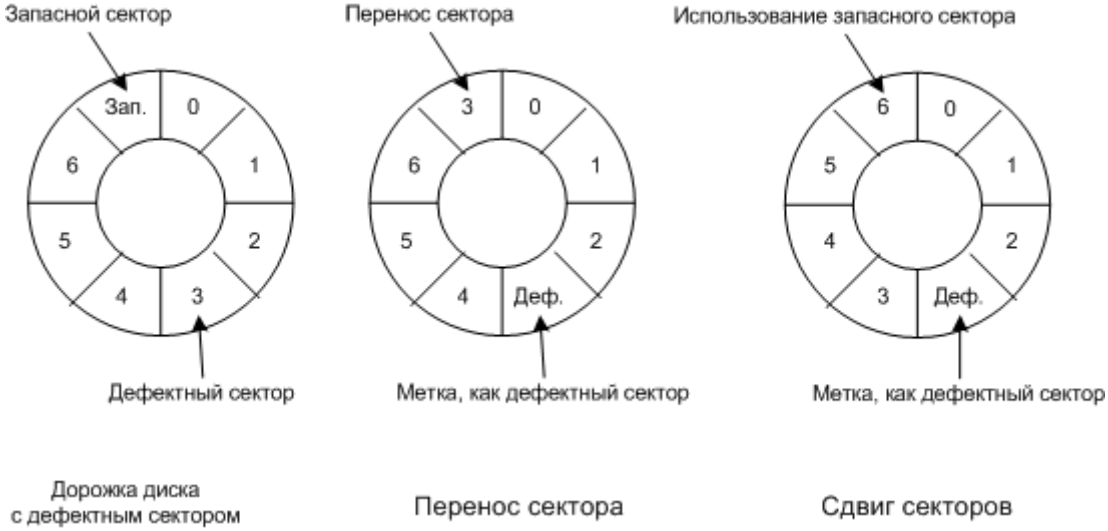
Чередование секторов

© Yuriy Shamshin

2.1.5. Запасные сектора и обработка ошибок дисков.

Т.к. создать диск без дефектов сложно, а во время использования появляются новые дефекты, то системе приходится контролировать и исправлять ошибки за счёт использования запасных секторов.

За счет этого обеспечивается одинаковая емкость на выходе производства и поддерживается надёжность диска при эксплуатации.



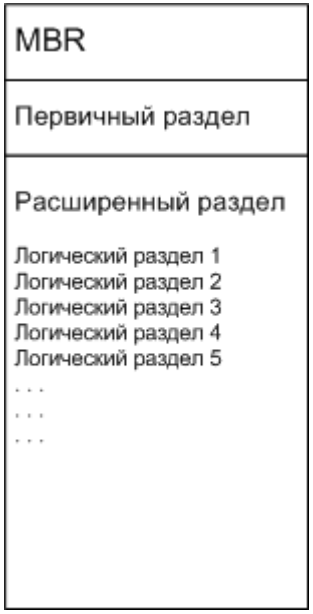
Способы замены дефектных кластеров

© Yuriy Shamshin

2.2. Разделы диска.

Более подробная информация - http://ru.wikipedia.org/wiki/Раздел_диска

После низкоуровневого форматирования диск разбивается на разделы, эти разделы воспринимаются ОС как отдельные диски.



Для чего можно использовать разделы:

- Отделить системные файлы от пользовательских (например, своп-файлы)
- Более эффективно использовать пространство (например, для администрирования).
- На разные разделы можно установить разные ОС.

Основные разделы диска:

- Первичный (Primary partition) - некоторые ОС могут загружаться только с первичного раздела. (В MBR под таблицу разделов выделено 64 байта. Каждая запись занимает 16 байт. Таким образом, всего на жестком диске может быть создано не более 4 разделов. Раньше это считалось достаточным.)
- Расширенный (Extended partition) - непосредственно данные не содержит, служит для создания логических дисков (создается, что бы обойти ограничение в 4-ре раздела).

© Yuriy Shamshin

- Логический (Logical partition) - может быть любое количество.

Информация о разделах записывается в 0-м секторе 0-го цилиндра, головка 0. Этот сектор называется главной загрузочной записью.

Главная загрузочная запись MBR (Master Boot Record) - содержит загрузочную программу и таблицу разделов.

Более подробная информация - http://ru.wikipedia.org/wiki/Главная_загрузочная_запись.

Таблица разделов (Partition Table) - содержит информацию о разделах, номер начальных секторов и размеры разделов. На Pentium-компьютерах в таблице есть место только для четырех записей, т.е. может быть только 4 раздела (к логическим это не относится, их может быть не ограниченное количество).

Активный раздел - раздел, с которого загружается ОС, может быть и логическим. В одном сеансе загрузки может быть только один активный раздел.

В Windows разделы будут называться (для пользователей) устройствами C:, D:, E: и т.д.

Т.к. MBR может работать только с разделами до 2.2 ТБ (2.2 ? 1012 байт), на смену MBR приходит **GPT**.

Таблица разделов GUID (GUID Partition Table — **GPT**) - позволяет создавать разделы диска размером до 9.4 ЗБ (9.4 ? 1021 байт).

© Yuriy Shamshin

Более подробная информация - http://ru.wikipedia.org/wiki/Таблица_разделов_GUID.

2.3. Высокоуровневое форматирование дисков.

Высокоуровневое форматирование (создание файловой системы) - проводится для каждого раздела в отдельности, и выполняет следующее:

- Создает загрузочный сектор (Boot Sector)
- Создает список свободных блоков (для UNIX) или таблицу (ы) размещения файлов (для FAT или NTFS)
- Создает корневой каталог
- Создает, пустую файловую систему
- Указывает, какая файловая система
- Помечает дефектные кластеры

Кластеры и блоки - единица хранения информации в файловых системах, файлы записываются на диск, разбитыми на блоки ли кластеры.

При загрузке системы, происходит следующее:

1. Запускается BIOS
2. BIOS считывает главную загрузочную запись, и передает ей управление
3. Загрузочная программа определяет, какой раздел активный
4. Из этого раздела считывается и запускается загрузочный сектор

5. Программа загрузочного сектора находит в корневом каталоге определенный файл (загрузочный файл)
6. Этот файл загружается в память и запускается (ОС начинает загрузку)

© Yuriy Shamshin

3. Алгоритмы планирования перемещения головок.

Факторы, влияющие на время считывания или записи на диск:

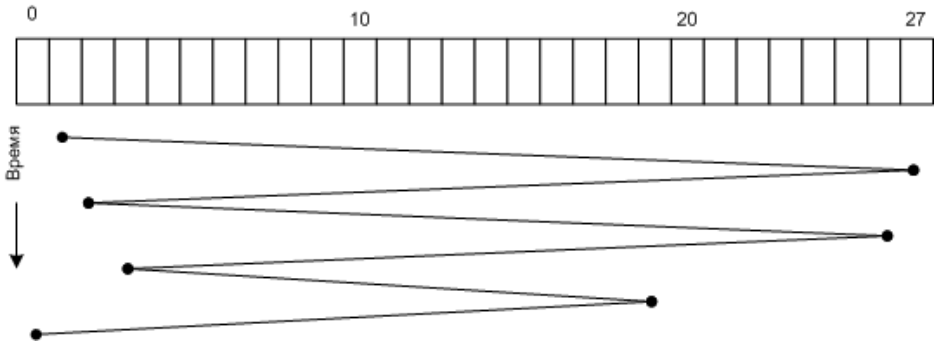
- Время поиска (время перемещения головки на нужный цилиндр)
- Время переключения головок
- Задержка вращения (время, требуемое для поворота нужного сектора под головку)
- Время передачи данных

Для большинства дисков самое большое, это время поиска. Поэтому, оптимизируя время поиска можно существенно повысить быстродействие. Алгоритмы могут быть реализованы в контроллере, в драйверах, в самой ОС.

3.1. Алгоритм "первый пришел - первым обслужен" FCFS (First Come, First Served).

Рассмотрим пример. Пусть у нас на диске из 28 цилиндров (от 0 до 27) есть следующая очередь запросов: 27, 2, 26, 3, 19, 0 и головки в начальный момент находятся на 1 цилиндре. Тогда положение головок будет меняться следующим образом:

© Yuriy Shamshin



Алгоритм FCFS

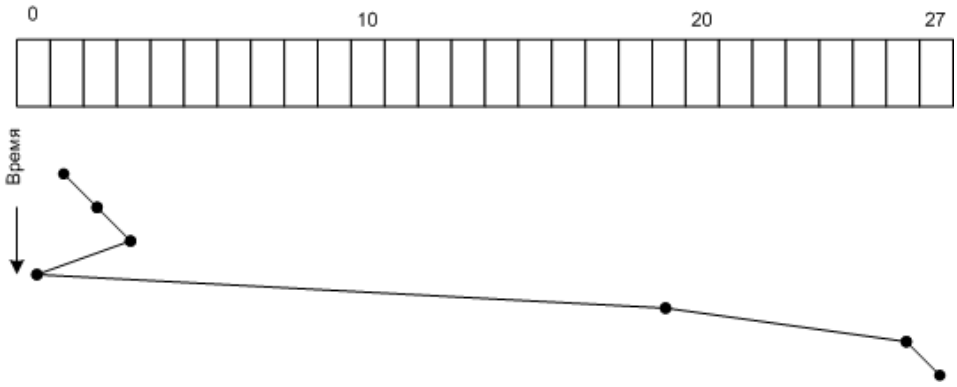
Как видно алгоритм не очень эффективный, но простой в реализации.

3.2. Алгоритм короткое время поиска первым (ближайший цилиндр) SSF (Shortest Seek First).

Этот алгоритм более эффективен. Но у него есть недостаток, если будут поступать постоянно новые запросы, то головка будет всегда находиться в локальном месте, вероятнее всего в средней части диска, а крайние цилиндры могут быть не обслужены вовсе.

Для предыдущего примера алгоритм даст следующую последовательность положений головок:

© Yuriy Shamshin



Алгоритм SSF.

3.3. Алгоритмы сканирования (SCAN, C-SCAN, LOOK, C-LOOK).

SCAN – головки постоянно перемещаются от одного края диска до его другого края, по ходу дела обслуживая все встречающиеся запросы. Просто, но не всегда эффективно.

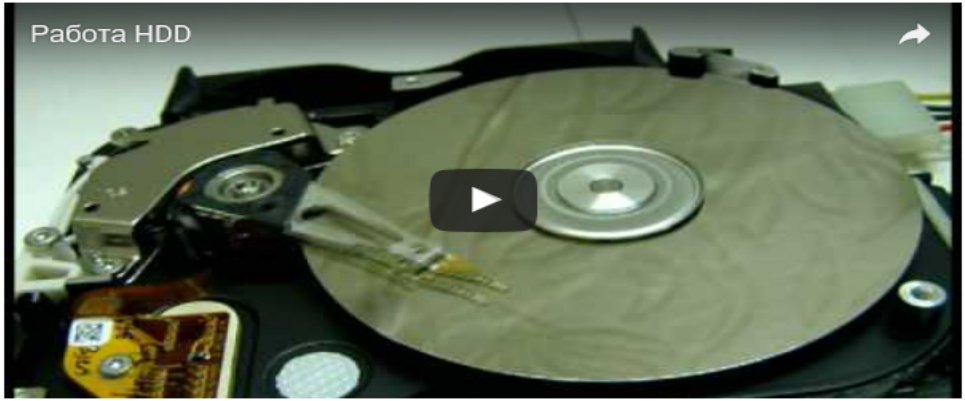
LOOK - если мы знаем, что обслужили последний попутный запрос в направлении движения головок, то мы можем не доходить до края диска, а сразу изменить направление на обратное

C-SCAN - циклическое сканирование. Когда головка достигает одного из краев диска, она без чтения попутных запросов перемещается на 0-й цилиндр, откуда вновь начинает свое движение.

C-LOOK - по аналогии с предыдущим.

© Yuriy Shamshin

Видео см. на https://youtu.be/y2iSTbKJy6E



Демонстрация работы жесткого диска

© Yuriy Shamshin