

Устройства хранения данных

Устройства хранения данных относятся к внешней памяти компьютера – они позволяют сохранять информацию для последующего ее использования независимо от состояния (включен или выключен) компьютера. В этих устройствах могут быть реализованы различные физические принципы хранения информации – магнитный, оптический, электронный – в любых их сочетаниях. Внешняя память принципиально отличается от внутренней (оперативной, постоянной и специальной) способом доступа к этой памяти процессора (исполняемой программы). Устройства внешней памяти оперируют *блоками* информации, но никак не байтами или словами, как, например, оперативная память. Процедуры обмена с устройствами внешней памяти привязаны к типу устройства, его контроллеру и способу подключения устройства к системе (интерфейсу).

Принцип действия и назначение устройств хранения

Устройства хранения, относящиеся к внешней памяти компьютера, обеспечивают энергонезависимое хранение блоков информации на каком-либо физическом носителе. Физические принципы энергонезависимого хранения и соответствующие им носители разнообразны. Наибольшее распространение получили следующие:

- Магнитный принцип основан на перемагничивании участков носителя в соответствии со значениями битов записываемой информации. Этот принцип реализуется в *устройствах с подвижным носителем* в виде диска или ленты, где запись и считывание производится на дорожку (трек). Головка записи вызывает изменение намагниченности участков трека в соответствии с записываемой битовой последовательностью. При считывании регистрируется изменение магнитного поля, связанное с прохождением под головкой участков трека, и из этих изменений извлекается ранее записанная информация. Существуют магнитные устройства хранения и с неподвижным носителем. В «древней» истории компьютеров применялись матрицы (кубы) памяти на магнитных сердечниках. В настоящее время используются (но пока еще не широко) микросхемы памяти FRAM (Ferroelectric Random Access Memory – ферроэлектрическая оперативная память). В магнитооптических устройствах принцип хранения – магнитный, оптика (лазер) используется лишь для разогрева перемагничиваемого участка при записи (это позволяет значительно уменьшить размер участка – повысить плотность записи) и считывании (свойства отраженного луча зависят от состояния магнитной «ячейки»).
- Оптический принцип основан на изменении оптических свойств участка носителя: степени прозрачности или коэффициента отражения. Способы, какими эти изменения достигаются, различны. В первых оптических устройствах использовался механический способ записи – пробивали отверстия в перфолентах и перфокартах. В современных оптических устройствах на дисках CD и DVD изменение оптических свойств достигается с помощью лазера, выжигающего лунки (необратимо, однократно) или изменяющего состояние

участка (обратимо, многократно). Выпуск массового тиража оптических носителей с информацией возможен и с помощью различных технологий печати.

- Электрический принцип основан на пороговых эффектах в полупроводниковых структурах. Этот принцип используется в *твердотельной памяти* – флэш-памяти и EEPROM. Здесь для изменения состояния хранящей ячейки требуется значительная энергия (довольно длительное воздействие сильного электрического поля), что и происходит в процессе записи, называемом программированием. Считывание требует значительно меньших затрат как энергии, так и времени.

Устройство хранения тем или иным способом подключается к *хосту* – компьютеру, в котором, как минимум, присутствуют процессор и оперативная память. Для хоста устройство хранения должно обеспечивать возможность записи блоков данных из внутренней памяти (как правило, ОЗУ) в устройство и считывание этих блоков из устройства в ОЗУ. Взаимодействие с устройством хранения выполняется по инициативе хоста (программы, выполняемой его процессором). В отличие от взаимодействия с внутренней памятью, с которой можно оперировать на уровне записи-чтения отдельных байтов, операции обмена с устройствами хранения всегда блочные. Блок может быть как фиксированного, так и произвольного размера. В настоящее время большее распространение получили устройства с фиксированным размером блока – это упрощает многие аспекты взаимодействия. Самый популярный размер блока – 512 байт, хотя в ряде устройств используются и иные размеры блока. Блок может быть переписан из внутренней памяти во внешнюю или обратно только целиком, и для выполнения любой операции обмена с внешней памятью требуется специальная процедура (подпрограмма).

Блоки в устройстве могут адресоваться различными способами. Наиболее простой и удобной является *линейная адресация логических блоков*, при которой каждый блок хранимых данных адресуется одномерным адресом (числом) *LBA* (Logical Block Address – адрес логического блока). Исторически сложилось использование и иных способов адресации; для дисковых устройств это *трехмерная адресация CHS* (Cylinder-Head-Sector – цилиндр-головка-сектор).

Основные характеристики и конструктивы устройств хранения

По *методу доступа* к информации устройства внешней памяти разделяются на устройства с прямым (или непосредственным) доступом и устройства с последовательным доступом. В *устройстве хранения с прямым доступом* (Direct Access Storage Device, DASD) есть возможность обращения к блокам по их адресам в произвольном порядке и, что важно, допускается произвольное чередование операций записи и чтения блоков. Традиционными устройствами с прямым доступом являются дисковые накопители, и часто в понятие «диск», или «дисковое устройство» (disk device), вкладывают значение «устройство внешней памяти прямого доступа». Так,

например, виртуальный диск в ОЗУ и электронный диск на флэш-памяти отнюдь не имеют круглых, а тем более вращающихся деталей.

В *устройствах последовательного доступа* произвольное чередование операций записи и чтения, относящихся к произвольным адресам блоков, либо невозможно, либо затруднительно (требует дополнительных внутренних операций, занимающих длительное время). Традиционными устройствами с последовательным доступом являются *накопители на магнитной ленте* (tape device), они же *стримеры*. Здесь для доступа к блокам информации с произвольными адресами приходится вхолостую считывать (или ускоренно перематывать) все блоки, находящиеся между ними. Необходимость последовательного сканирования блоков (вперед или назад) – неотъемлемое свойство устройств последовательного доступа с подвижным носителем. Несмотря на очевидный проигрыш во времени доступа к требуемым данным, ленточные устройства последовательного доступа в качестве внешней памяти находят применение для хранения очень больших массивов информации и эффективно используются для чтения-записи длинных последовательностей блоков.

Устройствами с последовательным доступом являются и оптические диски (CD, DVD). В этих устройствах информация записывается последовательно на один длинный спиральный трек. Конечно, устройство позиционирования головки позволяет ее довольно быстро (по сравнению с ленточными устройствами) перемещать на любой участок трека, обеспечивая произвольную адресацию. Однако по признаку невозможности произвольного чередования операций чтения-записи блоков (минимальная записываемая единица больше блока хранения) эти устройства являются последовательными. Программная эмуляция жесткого диска создает лишь иллюзию прямого доступа, скрывая от пользователя подробности непосредственной работы с устройством.

Главная характеристика устройства хранения – *емкость* (capacity), измеряемая в килобайтах, мегабайтах, гигабайтах и терабайтах (Кбайт, Мбайт, Гбайт, Тбайт, или в английской транскрипции KB, MB, GB, TB, или, еще короче – K, M, G, T). Здесь, как правило, приставки кило-, мега-, гига-, тера- имеют *десятичные* значения – 10^3 , 10^6 , 10^9 и 10^{12} соответственно. Емкость устройства в первую очередь определяется его носителем, однако она может ограничиваться и пределом возможности адресации блоков, свойственным тому или иному интерфейсу подключения.

Устройства внешней памяти могут иметь *сменные* или *фиксированные носители* информации. Применение сменных носителей (removable media) позволяет хранить неограниченный объем информации, а если носитель и формат записи стандартизованы, то они позволяют еще и обмениваться информацией между компьютерами. Существуют устройства с автоматической сменой носителя – ленточные карусели, дисковые устройства JukeBox. Эти довольно дорогие устройства применяют в крупных хранилищах данных. Для настольных машин имеются накопители CD-ROM с несколькими дисками (CD-changer), сменяемыми автоматически. Сменным может быть и целое устройство хранения.

Важнейшими общими параметрами устройств являются время доступа, скорость передачи данных и удельная стоимость хранения информации.

Время доступа (access time) определяется как усредненный интервал от получения устройством запроса на запись или чтение блока данных до фактического начала передачи данных. Дисковые устройства имеют время доступа от единиц до сотен миллисекунд. Для электронных устройств внешней памяти время доступа определяется быстродействием используемых микросхем памяти и при чтении составляет доли микросекунд, причем запись может продолжаться значительно дольше, что объясняется природой энергонезависимой электронной памяти. Для устройств с подвижными носителями основной расход времени имеет место в процессе *позиционирования* головок (seek time – время поиска) и *ожидания* подхода к ним требуемого участка носителей (latency – скрытый период). От того, что может делать система (хост и другие устройства) во время этой неизбежной задержки, предшествующей передаче запрашиваемых данных, зависит эффективность (общая производительность) компьютерной системы. В значительной степени эти возможности зависят от интерфейса устройства хранения.

Скорость записи и считывания определяется как отношение объема записываемых или считываемых данных ко времени, затрачиваемому на эту операцию. В затраты времени входят и время доступа, и время передачи данных. При этом оговаривается характер запросов – линейный или случайный, что сильно сказывается на величине скорости из-за влияния времени доступа. При определении скорости линейных запросов чтения-записи (linear transfer rate read/write) производится обращение к длинной цепочке блоков с последовательным нарастанием адреса. При определении скорости случайных запросов чтения-записи (random transfer rate read/write) соседние запросы направляются во все точки носителя. Для современных многозадачных ОС характерно чередующееся выполнение нескольких потоков запросов, и в каждом потоке высока вероятность последовательного нарастания адреса. Способность устройств хранения отрабатывать множество запросов, помещая их в свои внутренние очереди, существенно влияет на производительность системы в целом. Возможность эффективной работы с очередями существенно зависит от интерфейса устройства хранения.

Скорость передачи данных (Transfer Speed, Transfer Rate, или сокращенно XFER) определяется как производительность обмена данными, измеряемая после завершения поиска данных. Однако в способе измерения этого параметра возможны разночтения, поскольку современные устройства имеют в своем составе буферную память существенных размеров. Скорости обмена буферной памяти с собственно носителем (внутренняя скорость) и с внешним интерфейсом могут существенно различаться. Если *скорость работы внешнего интерфейса* ограничивается быстродействием электронных схем и достижимой частотой передаваемых сигналов, то *внутренняя скорость* более жестко ограничивается возможностями электромеханических устройств (скоростью движения носителя и плотностью записи). При измерениях скорости передачи на небольших объемах пересылок проявится ограничение внешнего интерфейса буферной

памяти, при средних объемах – ограничение внутренней скорости, а при больших объемах проявится еще и время поиска последующих блоков информации. Бывает, что в качестве скорости передачи данных указывают лишь максимальную скорость интерфейса, а о внутренней скорости можно судить по частоте вращения дисковых носителей и числу секторов на треке (об этих понятиях рассказывается далее).

Все устройства внешней памяти, применяемые в современных компьютерах, имеют *унифицированные конструктивные исполнения* (конечно, за исключением крупногабаритных устройств с автоматической сменой носителя). Их типоразмеры стандартизованы: наиболее жестко задаются ширина и высота устройств (поскольку их лицевые панели могут «выглядывать» из лицевой панели компьютера), глубина ограничена только максимально допустимым значением. Единообразное расположение резьбовых крепежных отверстий позволяет унифицировать конструкцию *отсеков* (bay) корпусов РС, предназначенных для установки накопителей. Первые накопители на гибких магнитных дисках диаметром 5,25", применяемые в РС, имели размеры лицевой панели 146,1 x 82,6 мм (5,75" x 3,25") и глубину около 203 мм (8"). Этот формат называется полновысотным пятидюймовым форматом (доли дюйма для краткости опускают) – 5" full-height form-factor. Такого же размера были и первые винчестеры. Вскоре высоту накопителя удалось уменьшить вдвое (half-height – половинная высота) – до 41,4 мм, и этот формат до сих пор используется как стандартный для многих типов устройств: НГМД с дискетами диаметром 5", приводов CD и DVD, стримеров, магнитооптики, накопителей на жестких дисках (старых и новых значительной емкости) и др. НГМД с дискетами диаметром 3,5" имеют ширину и высоту 101,6 x 25,4 мм (4" x 1") и длину (глубину) около 146 мм (5,74"). Этот формат широко используется и в современных трехдюймовых винчестерах, хотя среди них встречаются и «тонкие» модели высотой 20 мм (0,75"), и «толстые» – высотой 41 мм (1,6"). Для портативных компьютеров используют формат 2,75" (его называют и 2,5") с размерами 70 x 12,7 x 100 мм (2,76" x 0,5" x 3,95") и более тонкие – 70 x 9 x 100 мм. Накопители формата 1,8" имеют габариты 54 x 7 x 71 мм, а особо тонкие – 54 x 5 x 71 мм.

Стандартизованы также и *разъемы подключения питания* (рис. 1, один из вариантов). Миниатюрные разъемы используются только для питания трехдюймовых НГМД (дисковод для дискет), практически для всех остальных устройств трех- и пятидюймовых форматов применяются большие разъемы. Напряжение +5 В задействуют для питания электронных схем, напряжение +12 В – для питания двигателей, хотя в некоторых накопителях приводы могут питаться и от шины +5 В.

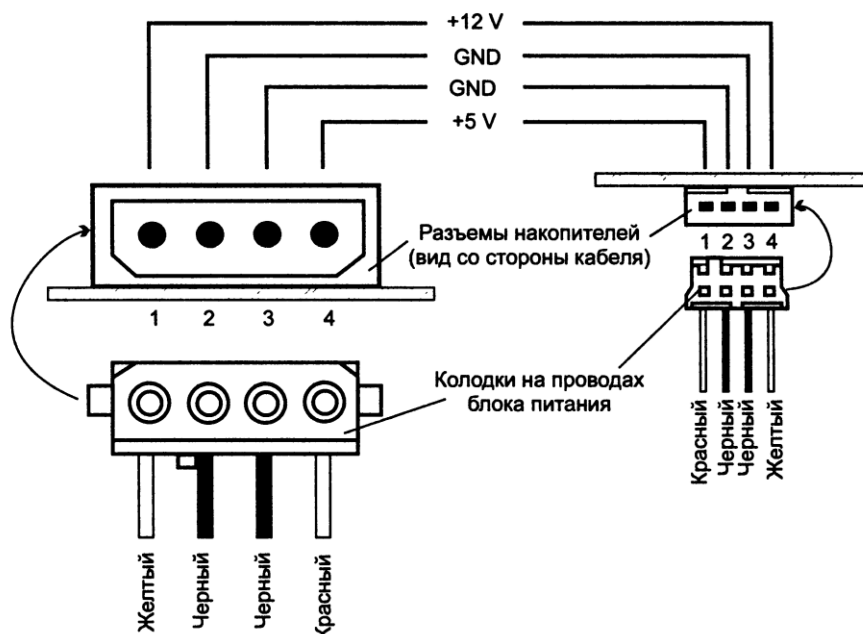


Рисунок 1 – Разъемы питания накопителей

В момент запуска двигателей ток потребления по цепи +12 В может превышать установившееся значение в несколько раз. В компьютерах и системах хранения с большим числом накопителей моменты их запуска стараются разнести во времени, что позволяет снизить пиковую нагрузку на источник питания. Возможность управления запуском двигателей зависит от интерфейса устройств хранения.

По отношению к корпусу компьютера устройства могут быть внутренними (internal) и внешними (external). *Внутренние устройства* помещаются в специальные трех- или пятидюймовые отсеки корпуса компьютера и питаются от его же блока питания. В описании корпусов компьютеров отсеки также подразделяются на внешние и внутренние, но они различаются лишь наличием или отсутствием выхода передней панели устройства, установленного в отсек, на лицевую панель корпуса. *Внешние устройства* помещают в отдельный корпус, а питаются они от собственного блока питания или от интерфейса (например, USB). Внешнее исполнение имеют как малогабаритные портативные устройства, так и особо крупные дисковые массивы. Сами приводы для внешних и внутренних устройств обычно имеют одинаковый конструктив одного из распространенных форматов.

Твердотельные устройства хранения на флэш-памяти выпускаются в разнообразных конструктивных исполнениях. Первые «статические диски» выполнялись в виде устройств формата 3,5" с интерфейсом ATA. Затем появились флэш-карты расширения с интерфейсом PC Card (PCMCIA), Card Bus, которые используются в блокнотных ПК, а также в ряде бытовых электронных устройств, например в цифровых фотокамерах. Современные малогабаритные карты (Compact Flash, SmartMedia, MMC, SD и др.) имеют разнообразные (собственные) конструктивы, и для их подключения к компьютеру требуются специальные устройства, оборудованные соответствующими слотами. Очень популярными стали твердотельные устройства хранения с

интерфейсом USB: устройства размером с брелок для ключей (или свисток) вставляются прямо в гнездо порта USB. Емкость такого «свистка» уже перевалила за десятки гигабайт.

Устройства хранения на магнитных дисках

Накопитель на магнитных дисках удобно использовать для иллюстрации принципов работы устройств хранения, принципов взаимодействия с такими устройствами, а также функций, выполняемых их контроллерами. Иные устройства с подвижными носителями можно рассматривать как вариации накопителя на магнитных дисках: возможны другие принципы функционирования и конструкции головок чтения-записи (оптические с лазерами и фотоприемниками), нюансы привода носителей (например, переменная скорость и стартстопный режим), особенности системы позиционирования.

Схематически устройство классического дискового накопителя представлено на рис. 2. Носителем информации является диск (один или несколько), на который нанесен слой вещества, способного намагничиваться (чаще всего ферромагнитный). Хранимую информацию представляет состояние намагниченности отдельных участков рабочей поверхности. Диски вращаются с помощью *двигателя шпинделя* (spindle motor), обеспечивающего требуемую частоту вращения в рабочем режиме. На диске имеется *индексный маркер*, который, проходя мимо специального датчика, отмечает начало каждого оборота диска. Информация на диске располагается на концентрических *треках* (дорожках), нумерация которых начинается с внешнего трека (track 00). Каждый трек разбит на *секторы* (sector) фиксированного размера. Сектор и является минимальным блоком информации, который может быть записан на диск или считан с него. Нумерация секторов начинается с единицы и привязывается к индексному маркеру. Если накопитель имеет несколько рабочих поверхностей (на шпинделе может быть размещен пакет дисков, а у каждого диска могут использоваться обе поверхности), то совокупность всех треков с одинаковыми номерами составляет *цилиндр* (cylinder). Для каждой рабочей поверхности в накопителе имеется своя *головка* (head), обеспечивающая запись и считывание информации. Головки нумеруются, начиная с нуля.

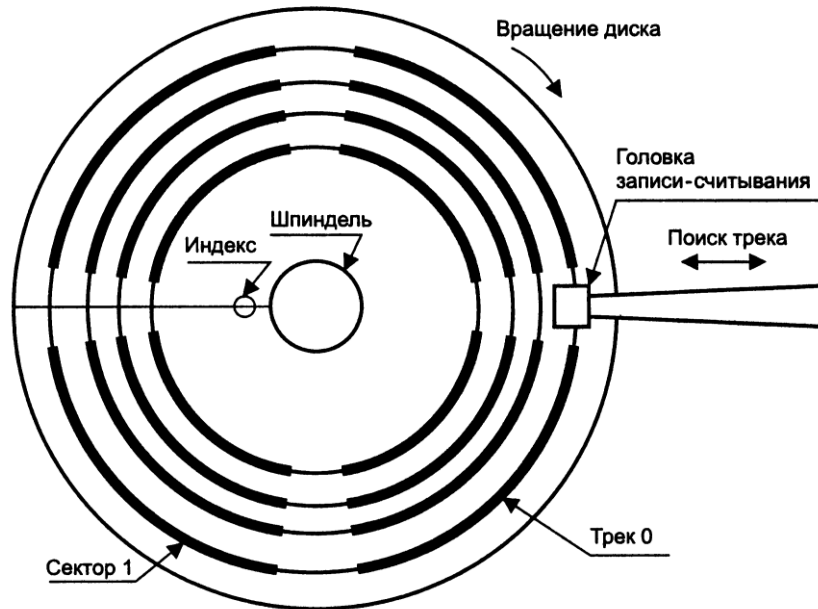


Рисунок 2 – Устройство дискового накопителя

Для того чтобы произвести элементарную операцию обмена – запись или чтение сектора, – шпиндель должен вращаться с заданной скоростью, блок головок должен быть подведен к требуемому цилиндру, и только когда нужный сектор подойдет к выбранной головке, начнется физическая операция обмена «полезными» данными между головкой и блоком электроники накопителя. Кроме того, головки считывают служебную информацию (адресную и сервисную), позволяющую определить и установить их текущее местоположение. Для записи информации на носитель используются различные *методы модуляции*, позволяющие кодировать двоичную информацию, намагничивая зоны магнитного слоя, проходящие под головкой. При считывании намагниченные зоны наводят в головке электрический сигнал, из которого декодируется ранее записанная информация. *Контроллер накопителя* выполняет сборку и разборку блоков информации (секторов или целых треков), включая формирование и проверку контрольных кодов, осуществляет модуляцию и демодуляцию сигналов головок и управляет всеми механизмами накопителя.

Несмотря на кажущуюся простоту конструкции, записать и потом достоверно считать информацию с диска не так-то просто. Для записи данных необходимо сформировать последовательный код, который должен быть самосинхронизирующимся: при последующем считывании из него должны извлекаться и данные, и синхросигнал, что позволяет восстановить записанную цепочку битов (этим занимается сепаратор данных – узел дискового контроллера). Кроме того, напомним, что индуктивные считывающие головки воспринимают только изменение намагниченности участков трека. Также учтем, что физическое исполнение – магнитные свойства носителя, конструкция головок, скорость движения, «высота полета» головок и т. п. – задает предельно достижимую плотность изменения состояния намагниченности (flux density), которую

хотелось бы использовать максимально эффективно. Эта плотность измеряется количеством зон с различным состоянием намагниченности на дюйм длины трека – *FCI* (Flux Changes per Inch – изменений потока на дюйм) и в современных накопителях достигает десятков тысяч. Для записи на диск применяют различные *схемы кодирования* (data encoding scheme), различающиеся сложностью реализации и эффективностью работы.

Из-за того что линейная скорость носителя относительно головки на внутренних цилиндрах меньше, чем на внешних, для нормальной записи при меньшей скорости приходится применять предварительную компенсацию записи. Для жестких дисков в CMOS Setup имеется параметр WPcom (Write Precompensation) – номер цилиндра, начиная с которого контроллер должен вырабатывать сигнал предварительной компенсации. Для накопителей со встроенным контроллером этот параметр игнорируется, поскольку они сами «знают», как работать со своими дисками. (Этот момент верен для старых жестких дисков)

Информация на дисках записывается и считывается посекторно, и каждый сектор имеет определенную *структуру* (формат). Не слишком вдаваясь в подробности, отметим, что в начале каждого сектора имеется заголовок, за которым следует поле данных и поле контрольного кода. В *заголовке* имеется поле идентификатора, включающее номер цилиндра, головки и собственно сектора. В этом же идентификаторе может содержаться и пометка о дефектности сектора, служащая указанием на невозможность его использования для хранения данных. Достоверность поля идентификатора проверяется с помощью контрольного кода заголовка. Заголовки секторов записываются только в ходе низкоуровневого форматирования, причем для всего трека сразу. При обращении к сектору по чтению или записи заголовок только считывается. *Поле данных* сектора отделено от заголовка небольшим зазором (gap), необходимым для того, чтобы при записи головка (точнее, обслуживающая ее схема) могла успеть переключиться из режима чтения (заголовка) в режим записи (данных). Сектор завершается *контрольным кодом поля данных* – CRC (Cyclic Redundancy Check – контроль с помощью циклического избыточного кода) или EC C (Error Checking and Correcting – обнаружение и коррекция ошибок). CRC-код позволяет только обнаруживать ошибки, а ECC-код – еще и исправлять ошибки небольшой кратности. В межсекторных промежутках может размещаться сервоинформация, служащая для точного наведения головки на трек.

Для того чтобы диск можно было использовать для записи и считывания информации, он должен быть *отформатирован*. Форматирование может разделяться на два уровня:

- ♦ Низкоуровневое форматирование (Low-Level Formatting, LLF) – формирование заголовков и пустых (размеченных заполнителем) полей данных всех секторов всех треков. При форматировании выполняется и верификация (проверка читабельности) каждого сектора, и в случае обнаружения неисправимых ошибок считывания в заголовке сектора делается пометка о его дефектности.

♦ Форматирование верхнего уровня заключается в формировании логической структуры диска (таблиц размещения файлов, корневого каталога и т. п.), соответствующей файловой подсистеме применяемой ОС. Эта процедура выполняется только после низкоуровневого форматирования.

Итак, структура трека – последовательность секторов – задается при его форматировании, а начало трека определяется контроллером по сигналу от индексного датчика или иным способом. Нумерация секторов, которая задается контроллеру при форматировании, может быть произвольной – важно лишь, чтобы все секторы трека имели уникальные номера в пределах допустимого диапазона. При обращении к сектору он ищется по идентификатору, а если за оборот диска (или за несколько оборотов) сектор с указанным номером не обнаруживается, контроллер фиксирует ошибку (Sector Not Found – сектор не найден). Забота о поиске сектора по его заголовку, помещении в его поле данных записываемой информации, снабженной контрольным кодом, а также считывании этой информации и ее проверке с помощью CRC- или ECC-кода лежит на контроллере накопителя. И, конечно же, контроллер управляет поиском затребованного цилиндра и коммутацией головок, выбирая нужный трек.

Современные жесткие диски внутренне могут быть организованы несколько иначе, чем в вышеописанной схеме. Индексные датчики теперь не используются – начало трека определяется по считываемому сигналу. Физическая разбивка на секторы (по 512 байтов данных, которым предшествует идентификатор) может отсутствовать – группа секторов трека представляет собой единый битовый поток, защищенный избыточным кодированием, из которого вычисляется блок данных, находящийся в требуемой позиции (так называемый ID-less format). Для обеспечения достоверности хранения данных (исправления ошибок) применяются избыточные коды Рида - Соломона (Reed - Solomon code), позволяющие большинство ошибок исправлять «на лету», не требуя повторного считывания блока данных (и дополнительного оборота диска). Заметим, что заметная вероятность искажения информации свойственна любым носителям информации, в том числе твердотельным.

Накопители на жестких магнитных дисках – винчестеры

Накопители на жестких магнитных дисках (НЖМД), они же HDD (Hard Disk Drive), являются главными устройствами дисковой памяти большинства компьютеров. По случайному совпадению цифр в названии первой модели НЖМД окрестили «винчестером» (просто игра слов), и это неофициальное название закрепилось в качестве синонима терминов HDD и НЖМД. Наряду с процессором и оперативной памятью винчестер определяет мощность компьютера. От него требуются большой объем хранимой информации (десятки и сотни гигабайт), малое время доступа (единицы миллисекунд), большая скорость передачи данных (десятки – сотни мегабайт в секунду), высокая надежность, умеренная стоимость и ряд других полезных свойств.

Прогресс в области производства винчестеров устойчивый и стремительный: от 10-мегабайтного диска ХТ уже пришли к сотням гигабайт, скорость передачи данных возросла на три порядка. «Модная» емкость винчестера ПК год от года увеличивается, при этом цена устройства постепенно снижается.

Конструкция НЖМД

Принципиально конструкция НЖМД соответствует рис. 2 и рис. 3. Вся электромеханическая часть накопителя – пакет дисков со шпиндельным двигателем и блок головок с приводом – находится в *гермоблоке*. Англоязычное сокращенное название этой сборки – *HDA* (Head Disk Assembly – диск с головками в сборке). На корпусе гермоблока размещается и плата электроники накопителя.

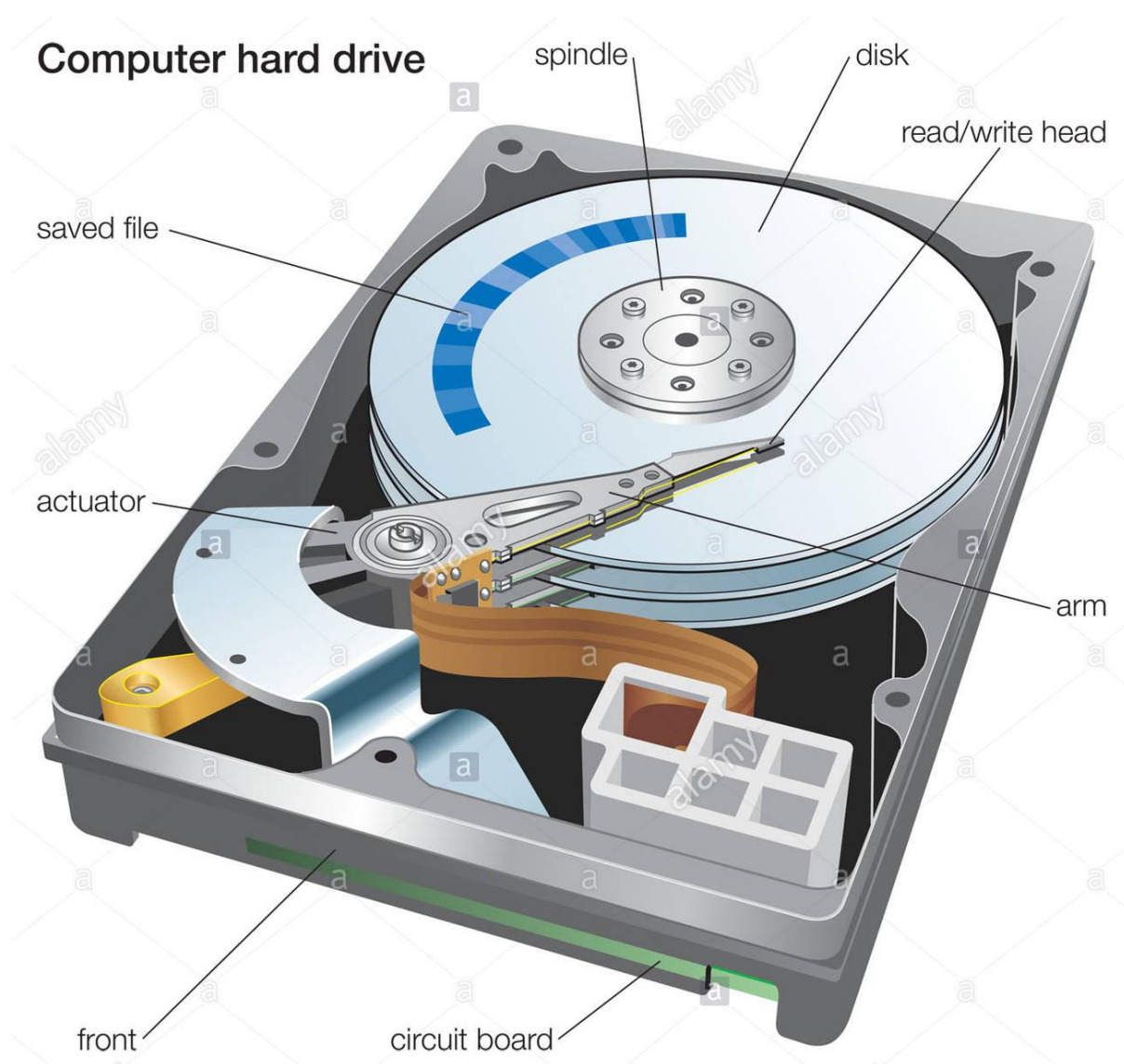


Рисунок 3 – Конструкция стандартного HDD

В качестве *привода шпинделя* используют, как правило, трехфазные синхронные двигатели. Схема управления двигателем обеспечивает пуск и останов шпинделя, а также поддерживает требуемую скорость с довольно высокой точностью. Шпиндельный двигатель является основным потребителем мощности по шине +12 В. В качестве датчика скорости вращения у старых винчестеров использовались сигналы индексного датчика, в современных винчестерах для точного управления шпиндельным двигателем служит информация сервометок. *Скорость вращения* долгие годы держалась на уровне 3600 об./мин, на текущий момент встречается скорость 5400, 5900 и 7200 об./мин. Самые быстрые винчестеры имеют скорость 10 000 и даже 15 000 об./мин. Чем выше скорость вращения, тем больше скорость обмена информацией с диском. Однако высокие скорости вращения порождают проблемы, связанные с балансировкой, тепловыделением, гироскопическим эффектом и аэродинамикой головок. Трение воздуха о пластины при скорости вращения 10 000 об./мин и выше является основной причиной нагрева винчестеров, заставляя заботиться об их охлаждении. Из-за гироскопического эффекта не рекомендуется перемещение (точнее, смена ориентации оси шпинделя) включенных накопителей с вращающимся шпинделем. Накопители для портативных компьютеров разрабатываются с учетом этих эффектов.

Пластины (platter) жестких дисков обычно изготавливают из алюминиевых сплавов, иногда из керамики или стекла. *Рабочий магнитный слой* основан на оксиде железа или оксиде хрома (более прочный). Поверхности пластин должны быть максимально плоскими, иначе высота «полета» головок будет отклоняться от номинальной, что ухудшает условия записи и считывания. Количество пластин у большинства современных винчестеров невелико (1-4), слишком большие пакеты пластин громоздки и конструктивно неудобны (для блока головок). Однако есть модели с десятком пластин в пакете. Емкость одной пластины винчестера формата 3,5" уже достигла 1,4 Тбайт и выше.

Традиционно для записи и считывания информации используются *магнитные головки*, представляющие собой миниатюрные катушки индуктивности, намотанные на магнитном сердечнике с зазором. Требования к оптимальной конструкции головок у процессов записи и считывания различаются, так что универсальная головка представляет собой некоторый компромисс. Первые индуктивные головки содержали проволочные обмотки, их сменили головки, выполненные по *тонкопленочной* (Think Film, TF) технологии.

Для магнитных головок весьма критично расстояние от головки до поверхности магнитного слоя носителя. Непосредственный контакт головки с поверхностью допустим лишь при малых скоростях движения носителя (как в НГМД). Головки винчестеров поддерживаются на микроскопическом расстоянии от рабочей поверхности аэродинамической подъемной силой. Высота «полета» головки должна выдерживаться довольно строго, иначе магнитные поля головок будут «промахиваться» мимо рабочего слоя. Высота определяется тем положением, при

котором подъемная сила, определяемая скоростью вращения, формой «крыла» головки и плотностью воздуха, уравнивает давление прижимающей головку пружины.

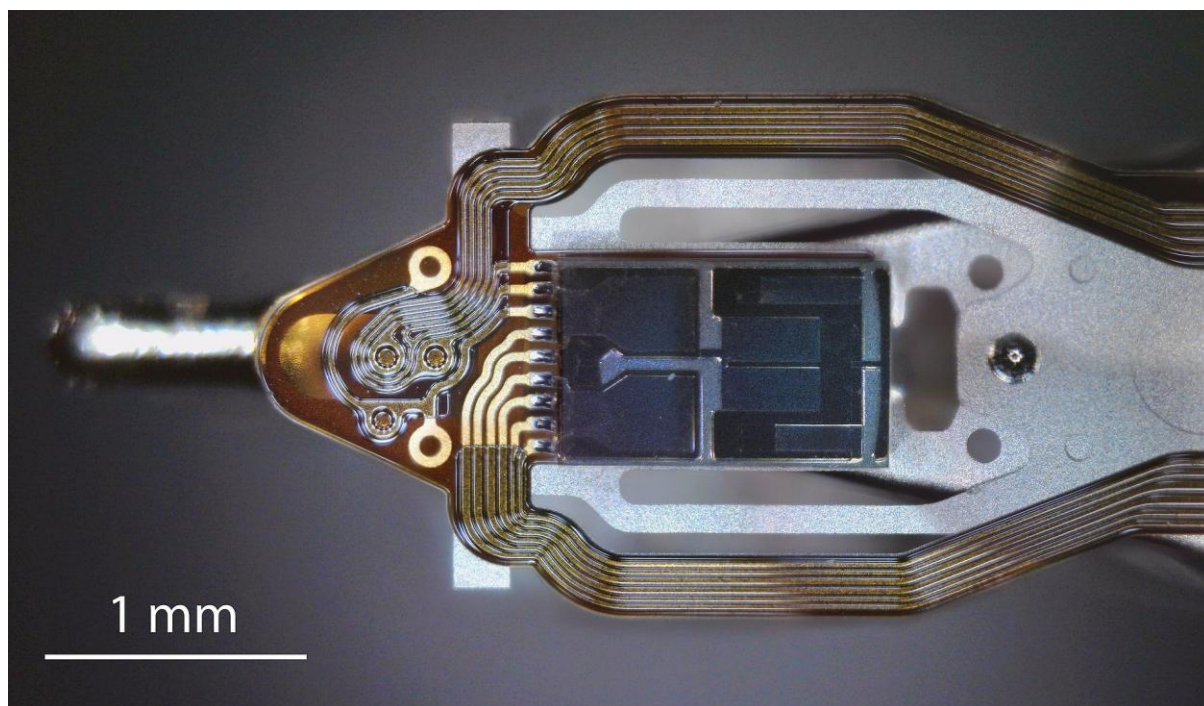


Рисунок 4 – Одиночная головка чтения/записи, сторона, обращённая к пластине накопителя

В современных накопителях для считывания часто применяют *магниторезистивные головки* (Magnetoresistance Head, MRH), основанные на эффекте анизотропии сопротивления полупроводников в магнитном поле (Anisotropic Magnetoresistance, AMR). В них через магниторезистивный датчик пропускают измерительный ток, и величина падения напряжения пропорциональна намагниченности находящегося под головкой участка магнитной поверхности. Сигнал с магниторезистивной головки повторяет форму записанного сигнала, а не является его производной (как у индуктивной головки). Магниторезистивная головка считывания хорошо «уживается» с индуктивной головкой записи, что позволяет достигать высокой плотности записи информации на магнитный носитель. Однако по технологии изготовления она сложнее тонкопленочной индуктивной, поскольку в ней сочетаются разнородные компоненты. От каждой комбинированной головки отходит четыре проводника: одна пара от электромагнитной головки записи (сопротивление постоянному току 8–10 Ом), вторая – от магниторезистивной головки чтения (около 30 Ом). Головки AMR позволяют считывать состояние намагниченности очень маленького пятна поверхности, с ними достигается плотность записи до 3 Гбит на квадратный дюйм. Изменение сопротивления при считывании составляет около 3 %. Следующий шаг – применение головок со сверхвысоким магниторезистивным эффектом (Giant Magnetoresistive, GMR). Они позволяют добиваться плотности порядка 10 Гбит на квадратный дюйм, а изменение сопротивления достигает 7-8 %. Чем больше расстояние, головки считывания до поверхности, тем слабее поле и тем больше должно быть пятно

магнитной поверхности, с которой головка может считать информацию, сохраняя приемлемый уровень ошибок (при одной и той же чувствительности). Правда, при этом повышается надежность (снижается вероятность повреждения поверхности головкой). Благодаря высокой чувствительности и узкой направленности головки GMR позволяют увеличить «высоту полета» без роста вероятности ошибок считывания.

При «падении» головки на рабочую поверхность, которое произойдет, если диск остановится, можно повредить как головку, так и поверхность диска. На современных дисках чистота обработки поверхности в рабочей зоне настолько высока, что если головки лягут на неподвижную поверхность, то из-за эффекта молекулярного притяжения они просто прилипнут к ней. Чтобы этого не происходило, в нерабочем положении головки *паркуются* (park) – отводятся в нерабочую зону, где допустимо их «приземление». Поверхность парковочной зоны имеет повышенную шероховатость, благодаря чему прилипания не происходит и диск может свободно набрать скорость до «взлета» головок.

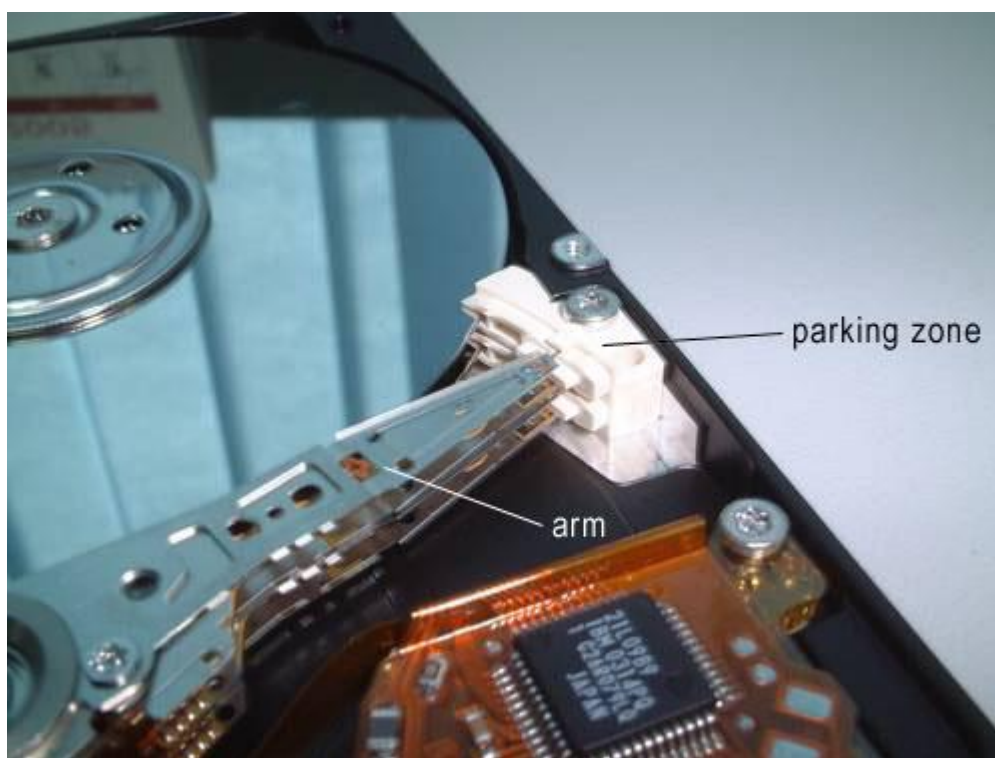


Рисунок 5 – Один из вариантов организации парковки головок

В старых винчестерах парковка обеспечивалась программно. Для ее выполнения в параметрах жестких дисков, хранимых в CMOS, присутствовал номер цилиндра для парковки (Landing Zone, или LZone). Парковка выполнялась запуском утилиты PARK или других утилит. В современных накопителях парковка осуществляется автоматически, когда напряжение питания или скорость вращения шпинделя падают ниже предельно допустимого значения. Для таких накопителей указанное в CMOS Setup значение параметра LZone игнорируется. Контроллеры современных дисков к

тому же не выпускают головки из зоны парковки, пока шпиндель не наберет заданных оборотов.

Для *позиционирования головок* на требуемый цилиндр в старых винчестерах применялись шаговые двигатели с червячной передачей, зубчатой рейкой или ленточной передачей. У шагового привода есть масса недостатков: ему свойственно большое время поиска – при произвольных обращениях нужно «прошагать» в среднем половину общего числа цилиндров; с ним не добиться высокой плотности записи, поскольку повышение плотности повышает требования к точности позиционирования. Система позиционирования получается разомкнутой (без обратной связи), и все отклонения размеров (температурные, связанные со старением и т. п.) приводят к уходу головок от исходных дорожек. Правда, позже появились винчестеры с шаговым приводом, в которых положение головки подстраивается на наилучшее считывание служебной информации.

В современных накопителях для головок применяют привод с *подвижной катушкой* (voice coil actuator), работающий по принципу звуковой катушки динамика. Этот тип привода называют еще *соленоидным*. В таком приводе блок головок связан с катушкой индуктивности, помещенной в магнитное поле постоянного магнита. При протекании тока через катушку на нее начинает действовать сила, пропорциональная силе тока, которая вызывает перемещение катушки, а следовательно, и блока головок. Привод может быть линейным или поворотным. В накопителе с линейным приводом катушка с блоком головок перемещается строго по радиусу дисков. Такой привод применялся в накопителях больших машин. В накопителе с поворотным приводом блок головок с катушкой размещен на поворотной рамке (рис. 6 и 7), и траектория головок отличается от радиальной. При этом азимут головки относительно трека меняется при перемещении головки, и эта азимутальная погрешность нежелательна для работы головок. Тем не менее, с этой неприятностью, отсутствующей у линейного привода, мирятся из-за относительной простоты исполнения, меньших габаритов, а, следовательно, и меньшей инерционности поворотного привода. В большинстве современных накопителей на жестких дисках применяется поворотный привод. Управляя направлением и силой тока, можно быстро перевести блок головок в любое положение – произвольное, а не по фиксированным шагам. Но в такой системе позиционирования необходима обратная связь – информация о текущем положении головок, по которой контроллер может управлять приводом. Привод, обеспечивающий точное позиционирование по сигналу обратной связи, называется *сервоприводом*. С точки зрения теории автоматического управления, соленоидный привод является *замкнутой системой*. Управление сервоприводом может быть оптимизировано по времени установления головок на требуемую позицию: когда отклонение от заданного положения велико, можно подавать большой ток, вызывающий большее ускорение блока головок. По мере приближения ток уменьшается, а для компенсации инерции в конце позиционирования ток может и поменять направление (активное торможение). Такая система привода позволяет сократить время доступа до единиц миллисекунд против сотен миллисекунд, характерных для шагового привода. На требуемом

цилиндре головки удерживаются следящей системой точного наведения. Остается только решить вопрос об источнике сигнала обратной связи для сервопривода.

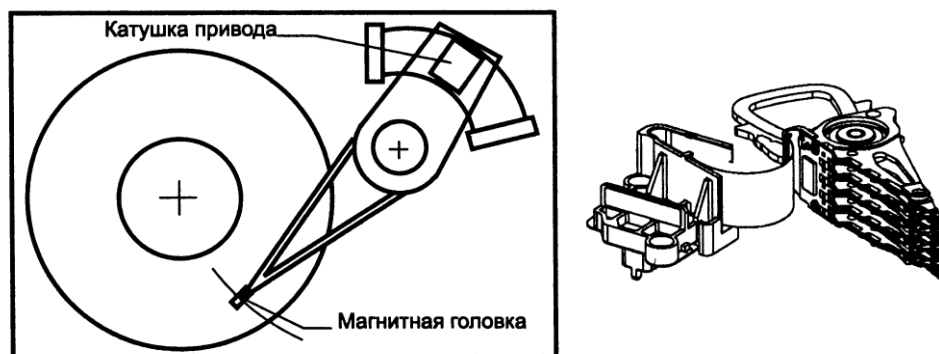


Рисунок 6 – Блок головок с поворотным приводом

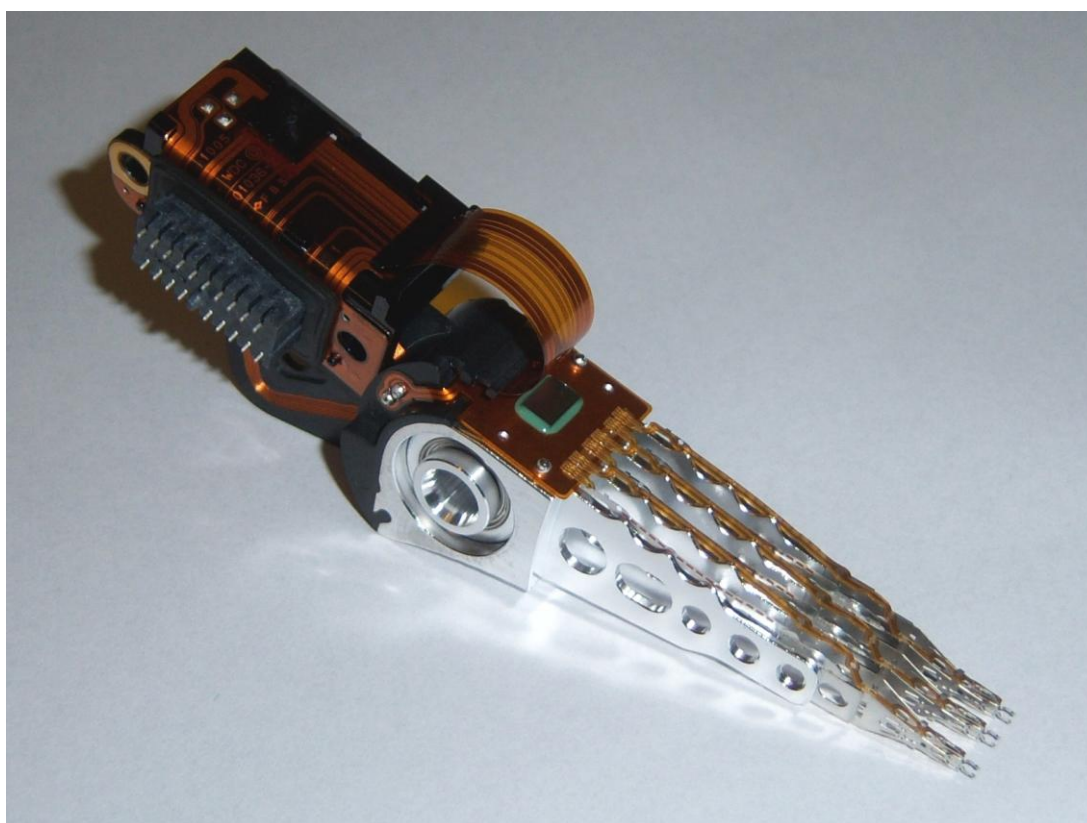


Рисунок 7 – Блок головок с поворотным приводом

В первых накопителях с линейным приводом использовались специальная зубчатая рейка и магнитный датчик, по сигналу которого отсчитывался номер трека. Однако по отношению к диску такая система привода все равно оставалась разомкнутой – привод позиционировал головки, руководствуясь собственными соображениями о координатах. Замкнуть систему управления позволило размещение прямо на диске *сервометок* – вспомогательной информации для «системы наведения». Благодаря сервометкам изменение размеров диска и привода под действием температуры и других факторов перестает существенно влиять на точность позиционирования,

поскольку сервометки располагаются рядом с теми же искомыми треками. Сервометки записываются в расположенных между треками областях при сборке накопителя, когда для позиционирования используется внешний «прицел» специального технологического оборудования. Четные и нечетные серводорожки различаются, хотя у них имеются совпадающие фрагменты синхронизации. Когда головка считывания находится точно посередине между парой соседних серводорожек, сигналы, наводимые от четной и нечетной серводорожек, имеют одинаковую амплитуду (фрагменты синхронизации синфазны и дают суммарный сигнал синхронизации сервометок). Если головка отклоняется от трека в сторону, например, четной серводорожки, амплитуда ее сигнала оказывается больше, что распознается детектором, и по сигналу рассогласования вырабатывается токовое воздействие на катушку, стремящееся вернуть головку на середину, то есть точно на трек. Этим обеспечивается хороший сигнал считывания, в составе которого имеются и адресные маркеры, содержащие номер цилиндра. По ним контроллер определяет, туда ли пришли головки, и, если необходимо, формирует импульсные сигналы позиционирования. После них снова вступает в дело система точного наведения по сервометкам, и в конце концов поиск успешно завершается. В процессе эксплуатации сервометки только считываются. По месту размещения сервометок различают накопители с *выделенной сервоповерхностью* (dedicated servo) и со *встроенными сервометками* (embedded servo). В первом случае в пакете дисков выделяется одна поверхность, используемая исключительно для хранения сервометок, и соответствующая ей головка является сервоголовкой. Ошибка позиционирования в такой системе может возникать вследствие изменения взаимного положения (перекоса) головок в блоке. Сервоголовка для следящей системы дает информацию практически непрерывно (на одном треке могут быть записаны сотни и тысячи сервометок), что повышает скорость и улучшает качество поиска и слежения. У накопителей с выделенной сервоповерхностью, как правило, число «полезных» головок нечетное.

В накопителях со встроенными сервометками информация для сервопривода записывается на рабочих поверхностях между секторами с данными. Она может размещаться в начале каждого трека – при этом на диске появляется клиновидная область сервометок. Однако из-за такого размещения сервоинформация (сигнал обратной связи) доступна только дискретно с периодичностью в один оборот диска, что при скорости 5400 об./мин составляет 11,1 мс. До точного позиционирования приходится выждать несколько оборотов диска. Более быстродействующий вариант – размещение сервометок перед каждым сектором – позволяет выйти на заданный трек даже за доли оборота шпинделя. Преимущество встроенных сервометок в том, что они обеспечивают компенсацию любых изменений в геометрии, поскольку система наводит головки именно по тому треку, к которому выполняется доступ. При этом снижаются требования к жесткости блока головок, что позволяет его облегчить и снизить инерционность. Однако даже переключение головок (без смены номера цилиндра) требует некоторого времени поиска – следящая система должна перестроиться на сервометки с другой головки. Существуют накопители с гибридной

сервосистемой, где помимо выделенной сервоповерхности используются сервометки, размещенные на рабочих поверхностях.

При выполнении операций записи и форматирования сигнал записи на время прохождения сервометок должен обязательно блокироваться. Если из-за неисправности устройства сервометки на какой-либо дорожке «накроет» сигнал записи, эта дорожка (а возможно, и соседние с ней) навсегда станет сбойной.

Для накопителей с соленоидным приводом проблема автопарковки решается легко, поскольку энергии для перемещения поворотного привода требуется немного. В парковочном положении головки удерживаются магнитной защелкой или механическим фиксатором.

Совершенно очевидно, что в гермоблоке воздух должен быть чистым – мелкая частичка, попавшая под головку, под которой пролетает носитель со скоростью несколько десятков километров в час, может поцарапать и головку, и диск. Вопреки названию гермоблок может и не быть герметичным – в нем имеется отверстие, закрытое фильтром и обеспечивающее выравнивание давления внутри сборки с атмосферным. Помимо этого фильтра, называемого барометрическим, имеется еще и внутренний рециркуляционный фильтр. Этот фильтр устанавливается на пути потока воздуха, увлекаемого вращающимся пакетом дисков. Он улавливает частички, которые могут выбиваться из поверхности дисков при «взлете» и «посадке» головок. Существуют, конечно, и специальные накопители для работы в особых климатических условиях. Они могут иметь герметический корпус, призванный выдерживать разность внутреннего и наружного давлений. В РС, как правило, применяют накопители обычного исполнения.

Помимо блока механики дисковый накопитель должен иметь и *блок электроники*, управляющий приводами шпинделя и головок, а также обслуживающий сигналы рабочих головок записи-считывания. Обычно в гермоблоке на маленькой плате устанавливают предварительные усилители считывания, коммутаторы и формирователи сигналов записи. К этой плате подключаются провода, идущие к головкам, а также кабель, связывающий ее с контроллером. *Контроллером накопителя* называют электронное устройство, на одной (интерфейсной) стороне которого идет обмен байтами команд, состояния и, конечно же, записываемой и считываемой информации, а другая его сторона связывается с гермоблоком. В современных накопителях на жестких дисках контроллер расположен на плате электроники, смонтированной вместе с гермоблоком. В старинных накопителях на жестких дисках с интерфейсами ST506/412 и ESDI контроллер был внешним и располагался на специальной карте расширения.

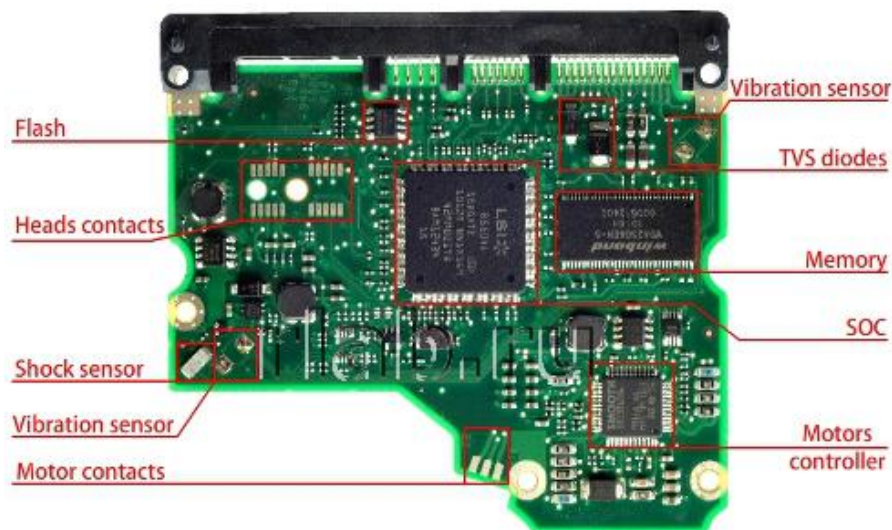


Рисунок 8 – Плата электроники жесткого диска

Объединение контроллера с гермоблоком позволило решить многие проблемы оптимизации накопителей, о которых речь пойдет далее. Контроллер современного винчестера состоит из нескольких основных блоков:

- Управляющий микроконтроллер (однокристальный, SOC) обеспечивает взаимодействие всех блоков накопителя и связь с внешним интерфейсом (ATA, SCSI, USB...). Здесь, как правило, используется 8- или 16-разрядный микроконтроллер общего назначения. Его программное обеспечение может храниться во внутреннем ПЗУ контроллера или в отдельной микросхеме памяти (EPROM или флэш). Часть ПО может загружаться во внутреннее ОЗУ, доступное микроконтроллеру, со служебной области диска.
- Внутреннее ОЗУ (буферная память накопителя, Memory) используется для считывания и записи секторов и локального кэширования. Объем внутреннего ОЗУ может быть от десятков килобайт до единиц мегабайт. Этот объем в сочетании с эффективностью алгоритмов кэширования заметно влияет на производительность винчестера.
- Блок управления шпиндельным двигателем (Motor controller) обеспечивает запуск и останов шпинделя по команде от микроконтроллера и поддерживает заданную скорость вращения по сигналам от датчиков индекса, специальных датчиков вращения или/и сервометок. Блок сигнализирует о достижении минимальной скорости, на которой можно выпускать головки, и номинальной.
- Блок управления позиционированием формирует импульсы управления соленоидом для перехода с цилиндра на цилиндр по команде микроконтроллера и следит за положением головки на треке по принятым сервосигналам.
- Коммутатор головок, совмещенный с предусилителем считывания и формирователем тока записи, – микросхема, смонтированная в непосредственной близости от головок (внутри гермоблока). Такое размещение позволяет улучшить отношение сигнал/шум при считывании.

- Канал чтения-записи представляет собой цепи, выделяющие из сигнала, принятого от предусилителя, импульсы синхронизации и данных и формирующие сигналы записи. В современных винчестерах в этом узле широко применяют сигнальные процессоры, реализующие метод PRML. Раньше здесь использовались аналоговые методы обработки сигналов.
- Детектор сервометок выделяет их из потока сигналов, принимаемых с головок считывания.
- Контроллер НЖМД (Hard Disk Controller, HDC) – специализированная микросхема, выполняющая основные функции, связанные с записью и считыванием данных. Она декодирует приходящий поток считываемых данных, выделяет служебные области, находит требуемые секторы, проверяет целостность данных (по алгоритму CRC или ECC), преобразует поток битов в байты и записывает их в буферную память. При записи микросхема в нужный момент (когда подходит поле данных требуемого сектора) формирует поток сигналов, требуемых для кодирования информации, преобразуя байты данных в поток битов и вычисляя контрольные последовательности. При форматировании она формирует заказанную структуру трека.

Разработчики винчестеров стремятся к сокращению числа микросхем, применяемых в блоке электроники (это с технико-экономической точки зрения почти всегда выгодно), и распределение функциональных узлов по микросхемам может варьироваться. Так, контроллер НЖМД может объединяться с управляющим процессором и некоторыми другими узлами в одной специализированной заказной ИС.

Производительность и оптимизация дисков

Задача, стоящая перед винчестером, выглядит просто – как можно быстрее отработать запрос на чтение или запись данных. Время, затрачиваемое на обмен данными одного сектора, складывается из *времени поиска* (seek time) цилиндра, *времени ожидания* (latency time) подхода сектора к головке, *времени обмена данными* сектора между накопителем и контроллером и между контроллером и памятью компьютера, являющейся источником и пунктом назначения для хранимых данных. Конечно, основные факторы, определяющие эти затраты времени, – механические: достижимая скорость позиционирования и скорость вращения диска. Последний параметр определяет и время ожидания сектора (в среднем это половина периода оборота диска), и время собственно передачи данных сектора (оно примерно равно времени оборота, деленному на количество секторов на треке). Однако одиночная передача сектора встречается редко, поскольку его размер невелик. Интерес представляет оптимизация передачи блока данных, расположенных в соседних секторах. Вполне естественно, что если блок данных не умещается в одном секторе, его продолжение логично разместить на том же треке в секторе со следующим номером, поскольку для доступа к нему нужно только дождаться его подхода к головке. Когда емкость трека оказывается исчерпанной, логично перейти на следующую поверхность того же цилиндра. Для

этого достаточно переключить головки, что выполняется электронным путем довольно быстро. Правда, у современных дисков с сервоинформацией, размещенной на рабочих поверхностях, требуется перенастройка системы точного позиционирования на сигнал от другой головки, а это уже требует заметного времени. Используя все секторы цилиндра, переходят к следующему цилиндру, для чего уже необходимо переместить головки, то есть затратить заметное время на операцию поиска. Дефрагментирующие программы как раз и занимаются тем, что размещают все блоки файлов в последовательных кластерах, а следовательно, и в секторах, упорядоченных по данному способу счета. Они справедливо рассчитывают на то, что для дискового накопителя оптимален именно такой порядок запроса секторов.

Для оптимизации производительности старых винчестеров, контроллеры которых не успевали обрабатывать и передавать хосту поток данных «на лету», применяли *чередование секторов* (interleaving). В современных накопителях чередование секторов не применяют. Для учета затрат времени на перенастройку сервосистемы при переходе на другую головку используют *послойное смещение секторов*, при котором начало следующего трека цилиндра чуть отстает от предыдущего. Для учета этой задержки при перемещении головок на соседний цилиндр вводят *радиальное смещение* (radial skew) секторов соседних цилиндров. Чередование секторов задавалось программно указанием последовательности номеров секторов в команде форматирования трека; для задания послойного и радиального смещений общепринятый программный интерфейс отсутствует. Смещение применяют только для накопителей со встроенными контроллерами, каковыми и являются все современные винчестеры.

Теперь несколько слов о скоростях вращения, плотности записи и количестве секторов на треке. Допустимое количество секторов определяется допустимой плотностью изменения магнитного потока, длиной трека и схемой кодирования данных. Длина трека определяется его диаметром, и, естественно, внешние треки дисков длиннее внутренних.

Зная количество секторов на треке, размер сектора и частоту вращения диска, можно определить максимальную скорость передачи данных между накопителем и контроллером.

Снизить неравномерность линейной плотности информации на внутренних и внешних цилиндрах можно форматированием треков с различным количеством секторов. Метод форматирования, называемый *зонной записью* (Zone Bit Recording, ZBR), позволяет существенно увеличить объем хранимых данных по сравнению с фиксированным числом секторов при тех же характеристиках носителя. Суть метода в том, что с учетом различия в длине треков цилиндры разбиваются на зоны, для которых принимается одинаковое число секторов на трек. Для внешних цилиндров число секторов на треке выбирается большим, чем для внутренних (примерно в 2 раза). Естественно, при зонной записи скорость передачи информации на внешних треках выше, чем на внутренних (за один оборот передается большее количество секторов). При этом и частота синхронизации схем записи и считывания для разных зон оказывается различной.

Контроллер, встроенный в накопитель, скрывает от системы переменное физическое число секторов на треке, а для общения с накопителем используется фиктивная *внешняя геометрия диска*. Когда говорят о собственно дисковых устройствах, эту геометрию часто называют логической. Однако в контексте с функциями BIOS ее уже называют физической. Остановимся на названии «внешняя геометрия», которая может быть либо *трехмерной* (с фиктивным числом цилиндров, головок и секторов на треке), либо *линейной*, где номер сектора задается одним числом.

Заметим также, что для накопителей с зонной записью вычислить скорость передачи по числу секторов на треке и скорости вращения шпинделя затруднительно, поскольку указываемое в паспорте число секторов обычно относится лишь к логической геометрии, а к физической отношения не имеет. Иногда в документации на винчестеры с зонной записью можно найти максимальное и минимальное количества секторов на трек, по которым можно оценить диапазон скоростей передачи. Повышение скорости вращения диска и увеличение числа секторов на треке обеспечивают повышение скорости обмена и ускорение позиционирования.

Производительность винчестера определяется не только возможностями гермоблока, но и свойствами его контроллера.

Интерфейс накопителя определяет возможную скорость обмена между буферной памятью винчестера и хостом. Основные интерфейсы винчестеров обеспечивают более высокую скорость передачи, чем внутренняя скорость обмена с пластинами, и благодаря этому они не тормозят обмен с дисками. Для интерфейса ATA (см. п. 19.2) эта скорость лежит в диапазоне от 3,3 (PIO Mode 0) до 66-133 Мбайт/с (Ultra DMA); Serial ATA – 150 Мбайт/с. Для SCSI – от 5 (обычный «узкий») до 320 Мбайт/с (Ultra4 Wide); для последовательного интерфейса SAS – 150 или 300 Мбайт/с. Интерфейс Fibre Channel обеспечивает скорость 100-800 Мбайт/с. Для внешнего подключения сменных устройств шину USB – до 300 МБ/с (в версии USB3.0).

Производительность винчестера в значительной степени зависит от объема его *буферной памяти*, который у современных винчестеров составляет до 64 Мбайт (есть тенденция роста для новых технологий хранения и в связи с ростом плотности записи). Часть этой памяти (до нескольких сотен килобайт) может использоваться для встроенного ПО. Естественно, чем больше объем памяти, тем лучше. Часто винчестеры с одной и той же механикой выпускаются с разными объемами памяти (это отражается в элементах названия модели), и более дорогие модели (с большей памятью) имеют заметный выигрыш в производительности. Возможности использования буферной памяти зависят от сложности контроллера. У старых дисков буфер вмещал всего один сектор и был однопортовым – не допускал одновременного обмена данных с диском и внешним интерфейсом. Более сложные и эффективные контроллеры имеют двухпортовый буфер на множество секторов, допускающий одновременное выполнение этих операций. В современных дисках контроллер способен использовать буфер и для *кэширования*. Общепринятой технологией кэширования диска является *упреждающее считывание* (read ahead). Суть его проста: если контроллер получает запрос на чтение сектора, то помимо запрошенного сектора он автоматически

считывает в буфер и следующие за ним секторы. В результате весьма вероятный запрос к следующему сектору будет обслужен из буфера без задержки, которая возможна из-за естественной асинхронности действий операционной системы и приложений с вращением диска. Более «ловкие» контроллеры идут дальше: они считывают в буфер весь трек, как только выполняется команда позиционирования, а когда приходит следующая за ней команда чтения, данные уже находятся в буфере. Такую нулевую задержку (zero latency) обеспечивает система команд интерфейсов ATA и SCSI. Кэширование применяют и для записи, но более сдержанно – здесь есть риск потери данных (например, при внезапном отключении питания).

Как и для всякой кэш-памяти, для эффективности встроенного кэша накопителя существенным фактором является алгоритм выделения памяти и замещения старых записей. Как обычно, замене подлежат самые старые записи. Вопрос о размере областей, выделяемых для упреждающего чтения, может решаться исходя из текущей статистики обращений. Контроллер с *адаптивным кэшированием*, заметив, что последние запросы чтения являются одиночными, перестает выделять большие области под упреждающее чтение. Если характер запросов изменяется, адаптивный контроллер принимает соответствующие решения. Кроме того, отпечаток на алгоритм кэширования накладывает и многозадачный характер современных операционных систем и их дисковых запросов. Таким образом, многозадачность проникает и во встроенные контроллеры дисков.

У современных дисков возможен выбор режима работы: для серверного применения (режим Server) можно использовать динамическое распределение буфера между потоками (переменное число и размер буферов), для настольного применения (режим Desktop) – статическое распределение (например, на 32 фиксированных сегмента).

Параметры винчестеров

После рассмотрения устройства и работы дисковых накопителей должен быть понятен смысл их параметров. Далее перечислены *общие параметры диска*.

- *Форматированная емкость* (formatted capacity), гигабайт (мегабайт), представляет собой объем хранимой полезной информации, то есть сумму полей данных всех доступных секторов. *Неформатированная емкость* (unformatted capacity) представляет собой максимальное количество битов, записываемых на всех треках диска, включая служебную информацию (заголовки секторов, контрольные коды полей данных). Соотношение форматированной и неформатированной емкостей определяется форматом трека (размером сектора), но поскольку для рядового пользователя свободы выбора формата нет, практический интерес представляет только форматированная емкость диска, которая указывается для стандартного размера сектора (512 байт). Напомним, что мегабайт и гигабайт здесь обычно означают 10^6 и 10^9 байт. Иногда указывается число доступных секторов.
- *Скорость вращения шпинделя* (spindle speed), измеряемая в оборотах в минуту (Revolutions Per Minute, RPM), позволяет косвенно судить о производи-

тельности (внутренней скорости). Сейчас обычной скоростью считается 5400 и 5900, а 7200 – более высокой. Там, где производительность особо критична, ранее использовались диски со скоростью 10 000 и 15 000 об./мин. (На текущий момент они практически полностью вытеснены накопителями на твердотельной памяти)

- *Интерфейс* (interface) определяет способ подключения накопителя. Для накопителей со встроенным контроллером распространены интерфейсы ATA и SCSI (параллельные и последовательные), для устройств внешнего исполнения применяют шины USB.
- Объем буферной памяти, возможности кэширования (чтение, запись, много-сегментность, адаптивность).

Следующая группа параметров – *параметры внутренней организации*'.

- *Количество физических дисков* (disks), или рабочих поверхностей (data surfaces), используемых для хранения данных. Современные накопители с небольшой высотой имеют малое (1-3) количество дисков для облегчения блока головок. Большое число дисков (и большая высота) характерно для старых накопителей и современных накопителей большой емкости.
- *Количество физических головок чтения-записи* (read/write heads), естественно, совпадающее с числом рабочих поверхностей. Заметим, что число головок (и рабочих поверхностей) может быть и меньше удвоенного числа дисков – обычно в каждом семействе есть такого рода модели. Это делается для утилизации дисков, у которых одна из поверхностей оказывается с производственным браком, или исходя из других технологических соображений.
- *Физическое количество цилиндров* (cylinders) от нескольких сотен, характерных для первых винчестеров, возросло до десятков тысяч.
- *Размер сектора* (bytes per sector) обычно составляет 512 байт.
- *Количество зон и количество секторов на треке* (sectors per track) в крайних зонах.
- *Расположение сервометок*. Они могут располагаться на выделенной поверхности (dedicated servo), встраиваться в рабочие поверхности (embedded servo) или иметь гибридный вариант расположения (hybrid servo).
- *Методом кодирования* (recording method, или data encoding scheme) и *декодирования* данных может быть метод MFM (FM почти и не применяли), RLL (ARLL) или PRML. Последний является наиболее прогрессивным.

Следующими описываются *параметры внешней* («логической») *геометрии* для устройств ATA:

- Поддержка режима линейной адресации LBA (может отсутствовать у старых дисков небольшого объема).
- Поддержка трансляции CHS (для дисков большого размера согласно ATA/ATAPI необязательна, но на практике почти всегда имеется).

- Количество цилиндров, головок и секторов на трек (при трансляции CHS) по умолчанию и возможности задания геометрии.

Быстродействие и производительность характеризуются следующими параметрами:

- *Время перехода на соседний трек (track-to-track seek)*, измеряемое в миллисекундах, характеризует быстродействие системы позиционирования. Для современных жестких дисков характерно время перехода 0,5-2 мс, причем для записи оно несколько больше, чем для считывания (записывать лучше при более точном позиционировании).
- *Среднее время поиска (average seek time)* определяется по серии обращений к случайным цилиндрам. Для большинства современных дисков оно составляет около 8-10 мс, в самых быстрых его удастся снизить до 4-5 мс. Для винчестеров со скоростью 10000 об./мин (обозначаются как ЮК) время поиска в зависимости от дальности перехода составляет от 0,3 мс (на соседний трек) до 12 мс (от края до края). Чем больше объем накопителя, тем сложнее достичь малого времени поиска: большее число головок труднее быстро перемещать; большее число цилиндров либо увеличивает длину перемещения головок, либо повышает требования к точности позиционирования.
- *Максимальное, или полное, время поиска (maximum seek time, full seek time)* относится к самым дальним переходам между крайними цилиндрами. Оно примерно в два раза превышает среднее время поиска.
- *Среднее время ожидания (average latency)* сектора при одиночном обращении обычно составляет половину времени полного оборота (7200 об./мин – 4 мс, 15 000 об./мин – 2 мс).
- *Внутренняя скорость передачи данных (internal transfer rate)* между носителем и буферной памятью контроллера задает физический предел производительности накопителя. Эта скорость выражается в разных величинах: если указывается в мегабитах в секунду (Mb/s), то сюда кроме пользовательских данных входят и «накладные расходы» – биты служебных полей. При выражении скорости в мегабайтах в секунду (MB/s) подразумеваются только байты пользовательских данных, поэтому пересчет в мегабиты в секунду простым умножением на 8 (число битов в байте) неправилен. У современных винчестеров с частотой вращения 5400 об./мин скорость составляет 80-180 Мбайт/с, с частотой вращения 7200 об./мин – до 200 Мбайт/с. Для каждой модели обычно указывают минимальное и максимальное значения скорости, соответствующие внутренним и внешним трекам (вспомним о зонном формате).
- *Внешняя скорость передачи данных (external transfer rate)*, измеряемая в килобайтах (мегабайтах) полезных данных в секунду, передаваемых по шине внешнего интерфейса, зависит от быстродействия электроники контроллера, типа интерфейсной шины и режима обмена. Для интерфейса ATA в режиме обмена PIO Mode 0 скорость составляет 3,3 Мбайт/с, в режиме PIO Mode 4 – 16,6 Мбайт/с, в режиме Ultra-DMA – 33, 66, 100 и 133 Мбайт/с. Для шин SATA

ограничения скорости в зависимости от версии интерфейса составляют 150, 300, 600 Мбайт/с.

- *Длительная производительность* (sustained throughput) определяется при последовательном чтении большого количества секторов. На нее влияют все составляющие: внутренняя и внешняя скорости, время позиционирования, задержка подхода сектора, количество ошибок позиционирования и чтения. Минимальное гарантированное значение этой скорости определяет возможность применения накопителя для мультимедийных приложений (записи и чтения аудио- и видеоданных) реального времени. Современные винчестеры с частотой вращения 5400 об./мин выдерживают потоки до 150 Мбайт/с, с частотой вращения 7200 об./мин – до 200 Мбайт/с. Эта скорость всегда ниже максимального значения внутренней скорости и предела внешней скорости. Для мультимедийных приложений в новых дисках имеются специальные варианты команд считывания и записи (*потоковое расширение ATA/ATAPI-7*), отличающиеся особой обработкой ошибок и политикой кэширования.

Надежность (reliability) устройства и *достоверность* хранения данных (data integrity) характеризуются следующими параметрами:

- *Ожидаемое время до отказа* (Mean Time Before Failure, MTBF), измеряемое в сотнях тысяч часов, является, естественно, среднестатистическим показателем для данного изделия. Реально столько часов (100 000 ч – это более 10 лет) испытания проводить, естественно, невозможно. На самом деле производится выборка из большой группы устройств, какая-то часть которых за вполне обозримое время испытаний выходит из строя. По зафиксированному потоку отказов теория вероятностей позволяет вычислить это условное ожидаемое время безотказной работы (хотя термин «поток» в данном контексте выглядит угрожающе). Значение MTBF, равное 100 000 ч, считается малым; 200 000-400 000 ч – нормальным, а 1 000 000 ч – высоким показателем надежности. Но, как говорится, «столько не живут», а возможный отказ винчестера в течение года (всего-то 8760 часов) вполне уложится в статистический показатель (если у вас за год не отказала значительная партия устройств). Иногда указывают и *ожидаемое время наработки на отказ* (Power On Hours, POH), в котором учитывается только время работы устройства (в MTBF не учитывается, включено устройство или нет).
- Более ценным для пользователя является *гарантийный срок* (limited warranty), в течение которого изготовитель (или поставщик) обеспечивает ремонт или замену отказавшего устройства. Примечательно, что даже при MTBF = - 800 000 ч (91 год) изготовитель дает гарантию всего 3-5 лет.
- *Вероятность неисправимых ошибок чтения* (nonrecoverable read errors per bits read) для современных винчестеров имеет порядок одной ошибки на 10^{14} считанных битов. Оценить, много это или мало, можно следующим образом. Пусть винчестер постоянно находится в работе, и к нему непрерывно идут обращения со средней производительностью чтения, которую «на глазок»

можно оценить в 1 Мбайт/с (это соответствует, например, умеренно загруженному диску сервера). Тогда простая арифметика показывает, что раз в 115 дней будут возникать ошибки, не восстанавливаемые (но обнаруженные!) схемами ЕСС-контроля. Вполне вероятно, что повторное считывание сектора пройдет без ошибок.

- *Вероятность исправимых ошибок чтения* (recoverable read errors per bits read) имеет порядок единицы на 10^{11} считанных битов. Если бы не было ЕСС-контроля (или при неисправной схеме контроля, с чем автору доводилось сталкиваться на практике), этот поток ошибок сделал бы работу с накопителем просто невыносимой (в том же примере ошибки будут появляться чаще, чем раз в три часа).
- *Вероятность ошибок поиска* (seek errors per seek) характеризует качество сервосистемы. Для современных винчестеров характерна вероятность одной ошибки на 10^8 операций поиска. Эти ошибки (при малом их числе) вполне безобидны, поскольку наличие номера цилиндра в заголовке каждого сектора не позволяет «промахнуться» при выполнении операций чтения или записи. Повторение операции поиска только слегка снижает среднее время доступа.

Уровень акустического шума характеризуется звуковой мощностью (sound power), излучаемой винчестером. На холостом ходу для винчестеров обычного применения (скорость вращения 5400 об./мин) желателен уровень до 30 дБ, при позиционировании желательно, чтобы он возрастал не более чем на 3-4 дБ. Для высокопроизводительных винчестеров (7200 об./мин), которые, естественно, шумят больше, желателен уровень до 35 дБ на холостом ходу. Для винчестеров, предназначенных для работы в устройствах бытовой электроники, желателен уровень до 25 дБ. Часто уровень шума указывают в беллах, эти цифры выглядят скромнее (25 db и 2,5 b – это одно и то же). Шум винчестера сильно зависит от корпуса компьютера, в который его устанавливают, и от способа крепления.

Потребляемая мощность определяется номинальными и пиковыми токами, потребляемыми по цепям +5 В и +12, В. Пик потребления по цепи +12 В возникает при раскрутке шпиндельного двигателя. Если блок питания компьютера не выдерживает этого пика (например, при одновременном запуске нескольких винчестеров) и напряжение «проседает», то шпиндель за требуемое время не наберет номинальную скорость и контроллер его остановит. Попытки «завести» мотор могут повторяться, что слышно по характерному звуку. Для высокоскоростных винчестеров приходится учитывать тепловыделение – для них может потребоваться специальный вентилятор.

Физические параметры включают ширину (width), высоту (height) и глубину (depth) корпуса накопителя, измеряемые в дюймах (inches) или миллиметрах, и Массу (weight), измеряемую в фунтах (lb) или килограммах.

Условия эксплуатации и хранения определяют возможные диапазоны температур, атмосферного давления, влажности и силы допустимых ударов. Вполне понятно, что условия эксплуатации (operating conditions) несколько жестче, чем условия хранения (non-operating conditions).