

ДАЛЬНЕВОСТОЧНЫЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ТИХООКЕАНСКИЙ ИНСТИТУТ
ДИСТАНЦИОННОГО ОБРАЗОВАНИЯ И ТЕХНОЛОГИЙ



И. В. Соппа

ВВЕДЕНИЕ В АРХИТЕКТУРУ ПЕРСОНАЛЬНОГО КОМПЬЮТЕРА

© Издательство Дальневосточного университета 2001

ВЛАДИВОСТОК
2001 г.

Оглавление

Введение	3
Глава 1. Интерфейсы	4
1.1. Основные функции и принципы организации интерфейсов	4
1.2. Принцип магистральности интерфейсов	5
1.3. Принцип селекции или арбитража	7
1.4. Принцип синхронизации и среда интерфейса	11
1.5. Системные магистрали персональных компьютеров	13
Глава 2. Системы ввода/вывода персональных компьютеров	22
2.1. Базовая система ввода/вывода персональных компьютеров	22
2.2. Спецификация Plug&Play	24
2.3. Клавиатура	25
2.4. Обслуживание клавиатуры по прерываниям BIOS и DOS	29
2.5. Манипуляторы курсора	31
2.6. Классификация и принцип работы видеосистем ПК	33
2.6. Текстовый режим работы видеосистем	40
2.7. Организация дисплейных страниц и управление курсором	44
2.8. Графический режим работы видеосистем	46
2.10. Особенности графического программирования	49
2.11. Принцип работы и классификация печатающих устройств	52
2.11. Матричные принтеры	55
Глава 3. Дисковая система персональных компьютеров	58
3.1. Технология магнитного диска	58
3.2. Методы кодирования информации на магнитных дисках	59
3.3. Форматирование накопителей на гибких магнитных дисках	61
3.4. Логическая организация данных на дискете	63
3.5. Общая характеристика накопителя на жестких дисках. Фактор чередования	65
3.6. Логическая организация данных на жестком диске	67
3.7. Форматирование жесткого диска	70
3.8. Контроллеры накопителей на гибких магнитных дисках	70
3.9. Интерфейсы HDD	73
3.10. Оптические диски	77
3.11. Стримеры	83
Глава 4. Устройства расширения функциональных возможностей ПК	85
4.1. КЭШ-память	85
4.2. Общая характеристика систем передачи данных	89
4.3. Модемы	93
4.4. Организация сетей	96
4.5. Протоколы компьютерных сетей	101

Введение

Не вызывает сомнения тот факт, что появление в 1981 году первого персонального компьютера произвело настоящую информационную революцию. Открытая (модульная) архитектура IBM PC в совокупности с операционной системой MS-DOS версия 1.0 сделала применение персональных компьютеров (ПК) в быту, на производстве, в коммерческой деятельности действительно массовым. Заложённая при создании первого ПК возможность постоянного его обновления и совершенствования была обеспечена исключительно совместимостью данного семейства ПК как на аппаратном, так и программном уровнях. За прошедшее с 1981 года время, естественно, многое в вычислительной технике изменилось. Производительность компьютеров возросла в тысячи раз, их архитектура постоянно совершенствовалась. Однако до сих пор миллионы и миллионы пользователей продолжают работать именно на IBM-совместимых персональных компьютерах с присущей им модульной архитектурой, которая практически не претерпела существенных изменений за последние 20 лет.

Настоящее учебное пособие - это конспект лекций для студентов ВУЗов, изучающих дисциплины «Архитектура ПК», «Технические средства компьютерных систем» и смежные им. В пособии рассматривается устройство основных модулей IBM-совместимых компьютеров, в основном, на аппаратном уровне, и излагаются вопросы организации взаимодействия между этими модулями. При этом основное внимание уделяется не столько техническим особенностям изготовления модулей, сколько физическим принципам и идеологии их работы в составе ПК. Именно поэтому многие разделы пособия изложены достаточно подробно, а некоторые - носят исключительно обзорный характер.

Изложение материала построено исходя из того, что стремительные изменения в элементной базе и схемотехнике ПК не дают возможности рассмотрения конкретных технических решений. Здесь важно понять принципы построения тех или иных устройств и механизм взаимодействия между ними.

В пособии также изложен материал о компьютерах, построенных в соответствии со спецификацией Plug and Play. Разработка данной спецификации была предпринята с целью решения проблемы конфигурации аппаратных средств вследствие модернизации ПК. После ее опубликования в 1993 году поддержка стандарта Plug and Play со стороны производителей ПК, программного и аппаратного обеспечения к ним стала набирать силу и широко внедряться на практике. Однако появление спецификации Plug and Play не изменяет общей идеологии построения персонального компьютера, а предоставляет пользователям дополнительный сервис.

Глава 1. Интерфейсы

1.1. Основные функции и принципы организации интерфейсов

Под *интерфейсом* понимают совокупность аппаратных и программных средств, обеспечивающих непосредственное взаимодействие между составными частями вычислительной системы на основе информационной, электрической и конструктивной совместимости. Основные функции интерфейса заключаются в непосредственной организации этой совместимости на базе интерфейсных БИС, дискретных элементов и физических элементов связи.

Под *информационной совместимостью* подразумевают унифицированный набор и структуру шин, принятую разрядность и формат кодов данных, адреса или команд, а также совместимость временных соотношений между управляющими сигналами.

Под *электрической совместимостью* подразумевают согласованность статических и динамических характеристик электрических сигналов, генерируемых составными частями системы.

Конструктивная совместимость подразумевает согласованность конструктивных элементов интерфейсов, а именно: типоразмеров печатных плат, каркасов, типов разъемов и карт их распайки.

В зависимости от назначения все интерфейсы подразделяются на *системные* и *измерительные*. Системные интерфейсы предназначены для организации взаимодействия между составными частями персонального компьютера или сервера. Физически они расположены на системных платах компьютеров и не допускают изменения их конфигурации на аппаратном уровне. На базе измерительных интерфейсов строят автоматизированные модульные измерительные системы и робототехнические комплексы (автоматизированные технологические линии) на производстве. Измерительные интерфейсы допускают их замены или переналадки в процессе эксплуатации.

Независимо от типа интерфейса они имеют следующую обобщенную структурную блок-схему, которая изображена на рис. 1.1.

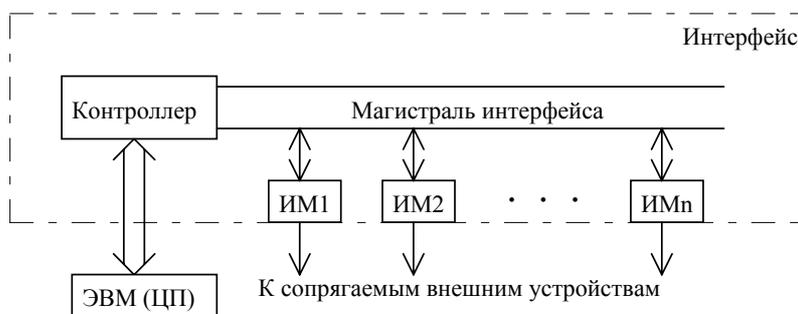


Рис. 1.1. Структурная блок-схема интерфейса

В его состав входят контроллер, магистраль интерфейса и интерфейсные модули. Функции контроллера заключаются в обеспечении взаимодействия между центральным процессором (ЦП) или внешней ЭВМ (ПК) и интерфейсными модулями (ИМ), а также в обеспечении межмодульных связей между ИМ внутри интерфейса. При этом, если интерфейс является *системным*, то контроллер (здесь чаще его называют chipset) непосредственно встроен в системную магистраль компьютера. В измерительных интерфейсах контроллер связан с внешним персональным компьютером.

Физическими элементами связи между контроллером (chipset) и интерфейсными модулями служат линии интерфейса – набор параллельных проводников, по которым в строго определенном формате, с соблюдением необходимой последовательности передаются данные, коды команд, состояния или адреса, а также требуемые сигналы управления. Совокупность всех этих линий называют *магистралью интерфейса*. По режиму работы магистраль интерфейса подразделяется

на одно- и двунаправленную. В случае двунаправленной магистрали предполагается передача информации в любом из направлений по единым линиям связи, в то время как для однонаправленной магистрали существуют различные линии связи для каждого из направлений передачи информации.

В качестве интерфейсных модулей в системных интерфейсах выступают компоненты персонального компьютера, такие как, дисковая подсистема, видеосистема, контроллер клавиатуры, порты ввода/вывода и т.п. Эти компоненты могут быть интегрированы непосредственно на системной плате компьютера, а могут быть размещены на отдельных (интерфейсных) платах, которые устанавливаются в специальные разъемы, называемые слотами расширения. В измерительных интерфейсах применяют набор интерфейсных модулей, обеспечивающих связь стандартного внешнего (сопрягаемого) оборудования с магистралью интерфейса. В качестве внешнего оборудования могут выступать: электроизмерительные приборы – частотомеры, вольтметры, генераторы и т.п. или технологическое оборудование – станки, роботы, конвейеры, манипуляторы и т.п. Возможно совместное использование как тех, так и других.

Наличие единой магистрали интерфейса неизбежно приводит к ситуациям, когда по единому информационному каналу требуется организовать взаимодействие одновременно между тремя и более интерфейсными модулями. Если это допустить, то на магистрали возникнет столкновение двоичной информации от разных источников, что приведет к ее искажению и появлению чрезмерных токовых нагрузок. Понятно, что в такой ситуации достоверная передача информации невозможна. Кроме того, некоторые БИС не выдерживают больших токовых нагрузок и выходят из строя. Следовательно, необходимо выделение только одного из источников информации и монопольное предоставление в его распоряжение магистрали интерфейса. Для решения этой задачи служат *схемы селекции или арбитража*, месторасположение которых, в общем случае, строго не регламентировано и зависит от типа используемого оборудования.

Искажение передаваемой информации также возникает в случаях, когда смены событий в различных частях интерфейса не согласованы. Например, один ИМ начинает передачу, а другой, являющийся приемником, не готов в данный момент к ее приему. В этом случае информация будет просто потеряна. Для устранения таких ситуаций необходимо засинхронизовать смены событий в приемнике и передатчике.

Таким образом, принятая структурная блок-схема интерфейса требует соблюдения определенных принципов, которые являются основными при их проектировании и построении. К ним относят следующие: *принцип магистральности, принцип селекции или арбитража и принцип синхронизации*. Перейдем к рассмотрению этих принципов.

1.2. Принцип магистральности интерфейсов

Общепринято магистраль интерфейса делить на две группы – информационную и управляющую. В этом случае информационная группа линий интерфейса образует *информационную магистраль*, а управляющая – *магистраль управления*. Составными частями каждой из магистралей служат *шины*. Шина представляет собой набор проводников, по которым одновременно передается логически связанная совокупность сигналов.

В интерфейсах информационная магистраль может состоять из следующего набора шин: *шины команд* (ШК), *шины данных* (ШД), *шины адреса* (ША) и *шины состояния* (ШС). То есть данная магистраль предназначена для передачи по ней кодов команд, данных, адреса и кодов состояния.

Коды команд используются для программирования микропроцессора, контроллера и интерфейсных БИС с целью обеспечения необходимого взаимодействия или выполнения ими определенных микроинструкций. В большинстве измерительных интерфейсов регламентируется минимальный набор кодов команд типа "Чтение/запись в регистр", "Конец передачи", "Начальная установка" и т. п. В системных интерфейсах набор кодов команд существенно расширен, поскольку включает в себя систему команд установленного в компьютере микропроцессора.

Коды данных используются для передачи данных. При этом, для измерительных интерфейсов, эти коды в основном соответствуют двоичному представлению символов в стандартах КОИ-7 (КОИ-8) или ASCII-кодировке. В системных интерфейсах к ним следует добавить любой двоичный код, воспринимаемый микропроцессором или интерфейсами модулями как данные. Основной характеристикой, как шины команд, так и шины данных является их разрядность.

Коды адреса используются для идентификации отдельных составных частей системы (ячеек памяти, регистров, портов ввода/вывода) с целью организации обращения к ним. Коды адреса могут передаваться либо в двоичной форме, либо с помощью позиционного кода, когда активизируется одна из линий ША. В системных интерфейсах используют только двоичную форму кодов адресов.

Коды состояния используются системой для передачи сообщений о состоянии устройства сопряжения. Как правило, эти коды формируются в ответ на коды команд чтения регистра состояния или просто информируют ЦП или контроллер о текущем функционировании устройства.

В последнее время наметилась тенденция к сокращению количества интерфейсных линий. В связи с этим большинство интерфейсов имеют только две информационных шины: шина адреса и шина данных. В этом случае коды команд и коды состояния передаются по шине данных, а назначение ША остается без изменений. Кроме того, наиболее широко используемая в современных компьютерах магистраль PCI использует для передачи кодов команд, данных, состояния и адреса единый набор проводников. Такой режим передачи называют мультиплексным.

В свою очередь, магистраль управления также подразделяется на ряд шин: шину управления обменом, шину передачи управления, шину прерываний и шину специальных управляющих сигналов.

Шина управления обменом служит для синхронизации передачи информации и, в зависимости от режима обмена (*синхронный* или *асинхронный*), может активизировать от одной до трех интерфейсных линий в процессе обмена. Общее количество интерфейсных линий данной шины не регламентировано. Так, например, при асинхронном обмене следует использовать три линии: линию готовности к приему (ГОТ), линию синхронизации (СИ) и линию подтверждения приема данных (ППД). Приемник информирует источник по линии ГОТ о готовности к приему, а источник выставляет данные на информационную магистраль и выдает импульс сопровождения по линии СИ. Приемник, получив данные, снимает сигнал ГОТ и информирует источник о приеме данных по линии ППД. По окончании обработки принятых данных, приемник снимает сигнал ППД и устанавливает сигнал ГОТ. Таким образом, текущий цикл обмена информацией завершён и приемник готов к следующему циклу обмена (рис. 1.2). Как правило асинхронный режим реализуется в системах, где имеется медленно работающая одна из составных ее частей. При синхронном обмене достаточно иметь только линию синхронизации СИ.



Рис. 1.2. Асинхронный режим передачи данных (фрагмент)

Шина передачи управления обеспечивает операции приоритетного захвата линий информационной магистрали. Наличие этой шины необходимо для организации взаимодействия только между двумя составными частями системы по единому для всех ее блоков информационному каналу. В качестве основных линий шины передачи управления, как правило, выступают: линия запроса информационных шин на организацию взаимодействия (сигнал BREQ),

линия занятости информационного канала (BUSY), совокупность входных (BPRN) и выходных (BPRO) линий подтверждения разрешения захвата шин отдельными интерфейсными модулями. В современных системных интерфейсах линии запроса захвата и подтверждения запроса захвата чаще обозначают соответственно DRQ_i и DACK_i.

Шина прерываний предназначена для организации запросов на обслуживание прерываний и выдаче сигналов подтверждения обслуживания прерываний. Напомним, что под прерыванием понимают приостановку выполнения текущей программы для перехода к выполнению другой, более приоритетной.

Существует два варианта реализации прерывания: *невекторное* интерфейсное и *векторное*. При передаче не векторного прерывания используется одна из линий запроса прерываний (IRQ). В этом случае контроллер прерываний распознает конкретную линию запроса IRQ_i, и, в зависимости от значения *i* выполняется соответствующая подпрограмма обслуживания прерываний. При векторном прерывании, после запроса на обслуживание прерывания, по информационной шине передается код вектора прерывания. Как правило, код вектора прерывания содержит начальный адрес подпрограммы обслуживания данного прерывания, которая должна быть заранее записана в ОЗУ или ПЗУ системы. В ряде интерфейсов шина прерываний содержит линию подтверждения запроса на прерывание INTA.

Шина специальных управляющих сигналов включает в себя линии, не вошедшие в предыдущие шины. К ним относят: линии питания системы, линии контроля за состоянием источника питания, линии начальной установки, общего сброса системы, линии тактовых импульсов и т.п.

1.3. Принцип селекции или арбитража

При выполнении нескольких параллельных процессов возникают ситуации, когда требуется осуществить передачу информации одновременно между несколькими интерфейсными модулями системы. В этом случае модули осуществляют запрос информационной магистрали на организацию взаимодействия. Поскольку эти модули физически связаны между собой единой информационной магистралью, то результаты выполнения одновременной передачи непредсказуемы, а в ряде случаев такая передача просто недопустима из-за возникающих больших токовых нагрузок. Для исключения подобных ситуаций необходимо обеспечить селекцию или арбитраж интерфейсных модулей, т.е. выделить только два из них: источник и приемник информации.

Существует два основных типа селекции или арбитража – *централизованный* и *децентрализованный*. Централизованный тип селекции предполагает размещение схем селекции или арбитража в одном из интересных модулей. Его отличительная особенность – наличие разомкнутых линий шин передачи управления. В децентрализованной структуре соответствующие линии, как правило, двунаправленные, а схемы арбитража симметрично распределены по интерфейсным модулям.

Централизованный тип селекции может быть организован с помощью одного из следующих вариантов:

а) *Временная селекция* (рис. 1.3).

В данном варианте контроллер связан с интерфейсными модулями (ИМ_i) только двумя линиями: *занято* (BUSY) и *такты*. В каждом из интерфейсных модулей находится счетчик, "настроенный" на анализ строго определенного количества входных импульсов, генерируемых контроллером по линии такты. По каждому тактовому импульсу последовательно активизируется только один ИМ по мере возрастания *i*. При этом, если ИМ_i необходимо организовать взаимодействие с контроллером, то он ожидает появления "своего" тактового импульса и по его поступлении выставляет сигнал BUSY. В ответ контроллер временно прекращает формирование тактовых импульсов и переходит в режим обмена информацией между контроллером и интерфейсным модулем. Если ИМ затребует организации взаимодействия в момент появления "своего" тактового импульса, то это взаимодействие будет ему предоставлено только в пределах следующего цикла опроса ИМ. Данный вариант селекции является наиболее простым и

обеспечивает случайную дисциплину обслуживания, включая тип: первый запросил – последний обслуживаешься. Временная селекция уравнивает все интерфейсные модули по уровню приоритета. В этом ее основной недостаток.

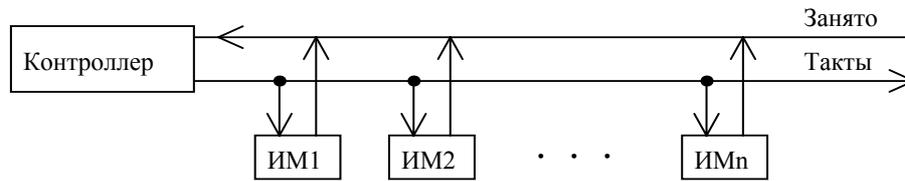


Рис. 1.3. Временная селекция

б) *Пространственная селекция* (рис. 1.4).

Для реализации данного варианта селекции вместе с линией BUSY контроллер связан с интерфейсными модулями линией запрос (BREQ) и шиной адреса информационной магистрали. Каждый модуль имеет свой уникальный адрес в системе. Если модуль запрашивает информационную магистраль на взаимодействие (захват), то он выставляет сигнал BREQ на линию запроса. Получив этот сигнал, контроллер выставляет на ША адрес первого из интерфейсных модулей в системе. Если этот адрес не совпадает с адресом ИМ, запросившего обслуживание, то контроллер выставляет адрес следующего ИМ и т.д. Перебор адресов будет продолжен до тех пор, пока адрес, установленный на ША не совпадет с адресом ИМ, установившего сигнал BREQ. При их совпадении данный ИМ выставляет сигнал BUSY, по поступлении которого как контроллер, так и ИМ переходят в режим обмена информацией. По окончании обмена оба сигнала должны быть сняты интерфейсным модулем, и система возвращается в исходное состояние. Данный вариант селекции принципиально отличается от временного возможностью изменения уровней приоритета ИМ в процессе работы. Для этого достаточно запрограммировать очередность выдачи адресов контроллером в соответствии с уровнем приоритета интерфейсных модулей.

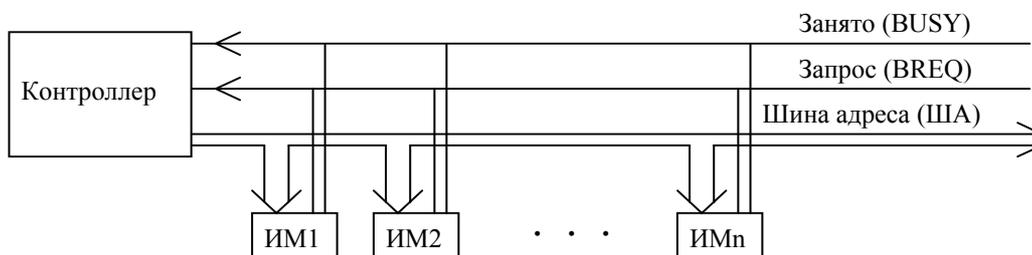


Рис. 1.4. Пространственная селекция

в) *Цепочечная селекция* (рис. 1.5).

Для данного варианта селекции вместо шины адреса используется линия подтверждения (BPR). При активизации сигнала BREQ одним или несколькими интерфейсными модулями, контроллер начинает поиск источника запроса выдачей сигнала BPR на линию подтверждения. Если первый в цепочке ИМ сигнал BREQ не выставлял, то он обязан пропустить сигнал BPR к следующему за ним ИМ без искажений, и так до тех пор, пока не будет обнаружен источник запроса. Как только источник запроса получает сигнал подтверждения, он выставляет сигнал BUSY и запрещает прохождение сигнала BPR к следующему ИМ. Таким образом, цепочка прохождения сигнала BPR "обрывается" на ИМ, запросившем взаимодействие, и поиск прекращается. Для нормального функционирования системы также необходимо обеспечить запрет на активизацию линии BUSY тем ИМ, на которые не поступил сигнал подтверждения BPR. Дальнейший протокол обмена аналогичен протоколу обмена пространственной селекции. Очевидно, что в этом случае наивысшим приоритетом будет обладать ближайший к контроллеру (схеме селекции) интерфейсный модуль.

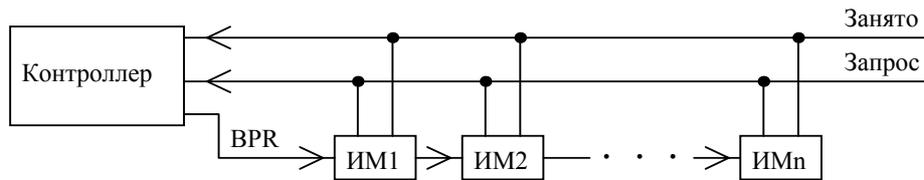


Рис. 1.5. Цепочечная селекция

Цепочечная селекция широко использовалась в персональных компьютерах и микро-ЭВМ, построенных на базе системных интерфейсов И41. К ним относятся СМ1800, СМ1810, СМ1820, "Искра-1030", "Нейрон И9.66" и др.

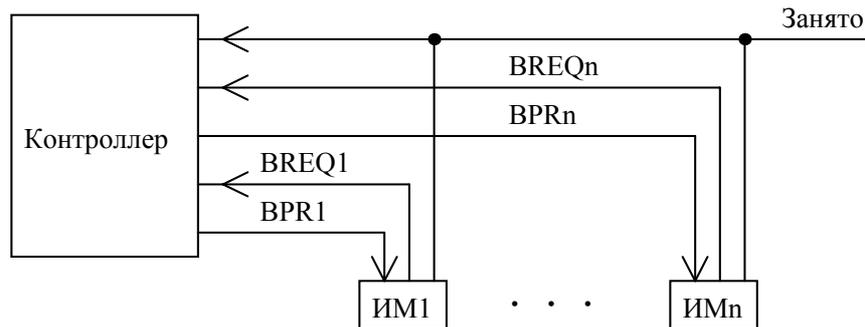


Рис. 1.6. Селекция по выделенным направлениям

г) Селекция по выделенным направлениям (рис. 1.6).

Данный вариант селекции требует наличия индивидуальных линий BREQ(DRQi) и BPR (DACKi) для каждого из существующих в системе ИМ. Для организации взаимодействия ИМ выставляет сигнал запроса BREQ и переходит в режим ожидания сигнала подтверждения. Получив запрос, контроллер активизирует линию подтверждения BPR того ИМ, который имеет наивысший приоритет из всех ИМ, затребовавших взаимодействия. Получив подтверждение, источник запроса выставляет сигнал BUSY на линии занято, захватывает информационную магистраль и система переходит в режим обмена информацией. По окончании взаимодействия, сигналы подтверждения и занято снимаются и система возвращается в исходное состояние.

Данный вариант селекции характеризуется минимальным временем поиска источника запроса, а также гибкостью установления дисциплины обслуживания, поскольку контроллер может быть запрограммирован или перепрограммирован в процессе работы на произвольный приоритет и порядок опроса. Его широко используют IBM-совместимые компьютеры и в настоящее время.

Характерным признаком децентрализованной селекции является наличие замкнутых линий запроса и подтверждения. Здесь различают три основных варианта ее реализации:

а) Цепочечная структура (рис. 1.7).

Данный вариант селекции может быть получен из централизованной путем исключения линии BUSY и замыкания линий запроса и подтверждения. В этом случае необходимым условием установления сигнала BREQ любым устройством является отсутствие сигнала BPR на его входе. В противном случае он должен быть снят. При выдаче запроса этот сигнал дизъюнктивно (например, с помощью схем исключающее-ИЛИ) формируется на линии запрос, трансформируется в сигнал BPR и поступает на вход первого из интерфейсных модулей, связанных в цепочку. Каждый из интерфейсных модулей, за исключением источника запроса, обязан без искажений пропустить его через себя. Таким образом, прохождение сигнала BPR "оборвется" на ИМ, который выставил сигнал BREQ на линии запроса. Схема селекции анализирует состояние линии запроса и линии подтверждения на выходе ближайшего к схеме селекции интерфейсного модуля и, в случае обнаружения сигнала BREQ и исчезновения сигнала BPR, перейдет в режим обмена информацией с активным ИМ.

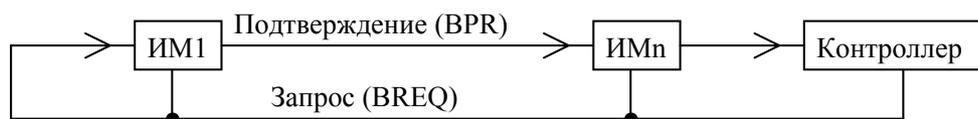


Рис. 1.7. Децентрализованная цепочечная структура

Очевидно, что наивысшим приоритетом в данном случае будет обладать интерфейсный модуль, который максимально удален от, в данном случае, контроллера в цепочке.

б) Кольцевая структура (рис. 1.8).

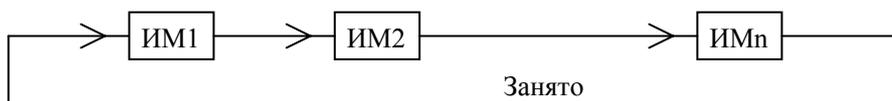


Рис. 1.8 Кольцевая структура селекции

В данном варианте используется только одна линия занятости информационного канала в системе. В исходном состоянии по ней циркулируют маркерные импульсы (код), генерируемые одним из модулей, а остальные модули системы постоянно контролируют их прохождение. Условием организации взаимодействия является либо прекращение прохождения маркерных импульсов (кода) тем ИМ, который затребовал это взаимодействие, либо изменение им исходного кода маркерных импульсов. По окончании взаимодействия, как правило, этот модуль становится генератором маркерных импульсов (кода) и система возвращается в исходное состояние. Кольцевая структура селекции в основном распространена в локальных сетях с большим количеством абонентов и характеризуется не очень высокой помехозащищенностью. Но, тем не менее, достаточно широко распространена в настоящее время.

в) Децентрализованное кодовое управление (рис. 1.9).

Достаточно широко распространенный способ селекции в магистрально модульных измерительных системах. Сущность селекции заключается в параллельном выделении приоритетного источника запроса путем поразрядного сравнения кодов приоритета на шине адреса в асинхронном режиме одновременно во всех модулях, выставивших запросы.

В начале цикла арбитража те интерфейсные модули, которые запросили взаимодействия, выставляют на шине адреса код своего вектора прерывания (уровень приоритета). Далее модули приступают к процедуре сравнения этого кода с кодом, дизъюнктивно установленным на шине адреса. В результате такого сравнения модули с меньшими уровнями приоритета снимают свои коды с ША, оставляя тем самым только модуль с наивысшим приоритетом. Дальнейший протокол обмена аналогичен протоколу обмена пространственного варианта селекции централизованных структур. Для обеспечения асинхронности сравнения кодов каждый интерфейсный модуль имеет дополнительные сигналы начала (Z1) и окончания (Z0) сравнения, включенные в схему последовательно.



Рис. 1.9. Децентрализованное кодовое управление

Таким образом, независимо от возможных вариантов реализации перечисленных выше вариантов селекции можно выделить следующие операции, выполняемые схемами арбитража при организации взаимодействия: выдача, хранение и восприятие запроса; выделение приоритетного

интерфейсного модуля; захват интерфейсных шин и организация обмена информацией. По окончании взаимодействия – возвращение в исходное состояние.

1.4. Принцип синхронизации и среда интерфейса

Принцип синхронизации интерфейсов определяет согласование в смене событий для различных составных частей системы. Здесь различают *асинхронный* и *синхронный* тип согласования. Два и более процесса являются *синхронными*, если смены состояний (событий) этих процессов взаимонезависимы и выполняются через строго определенные фиксированные моменты времени. Два и более процесса называются *асинхронными*, если смены состояний этих процессов взаимозависимы, а время изменения состояния одного процесса определяется моментом времени изменения состояния другого.

Схемотехническая реализация *асинхронного* процесса основывается на принципе петель обратной связи, которая может быть представлена одной или двумя интерфейсными линиями. Процесс приема передаваемой информации синхронизируется специальными стробирующими сигналами от приемника или передатчика (сигналы СИ, ГОТ и ППД), которые устанавливаются на шине управления обменом с фиксированной задержкой относительно времени установления кодов команд, данных или адреса на информационной магистрали. По характеру сигналов синхронизации данные схемы подразделяются на импульсные, когда распознаются перепады импульсов синхронизации, и потенциальные, когда фиксируется уровень синхроимпульсов. Один из вариантов реализации асинхронного процесса был достаточно подробно рассмотрен в пункте 1.2.

Схемы синхронизации без обратной связи (*синхронный* тип обмена) строятся на основе одного общего или двух (в передатчике и в приемнике) исходно независимых, но близких по частоте кварцевых генераторов. Для последней из перечисленных схем, синхронизация обоих генераторов выполняется в начале цикла обмена и поддерживается на протяжении всего времени передачи информации. В этом случае, для достоверной ее передачи, необходимо обеспечить превышение скорости обработки принимаемой информации приемником относительно скорости передачи информации передатчиком. Данный вариант синхронизации практически реализуется в последовательном интерфейсе RS-232C.

Следует отметить, что применительно к локальным компьютерным сетям в асинхронный и синхронный типы обмена информацией вкладывается несколько иной смысл. Поэтому при указании типа обмена необходимо дополнительно пояснять к компьютеру или локальным сетям он относится.

В тех случаях, когда приемник и передатчик находятся на значительном расстоянии друг от друга необходимо учитывать еще один фактор, называемый средой интерфейса. Под *средой интерфейса* понимают совокупность *физических принципов* передачи информации (электрический, электромагнитный или оптический), *типов* используемой кабельной системы (коаксиальный, плоский, витая пара и т.п.) или световода (градиентный или ступенчатый) и *характеристик* самих приемо-передатчиков. Среда интерфейса определяет зависимость предельной скорости передачи информации ($V_{\text{инф}}$ [бит/сек]) от расстояния между приемником и передатчиком (рис.1.10).

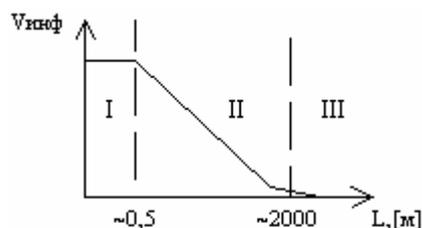


Рис. 1.10.

Независимо от среды интерфейса данная зависимость характеризуется тремя участками. Для первого участка скорость передачи информации практически не зависит от длины кабеля и в основном определяется быстродействием (временными задержками) приемо-передатчиков. На втором участке наблюдается падение предельной скорости передачи информации практически пропорционально логарифму длины линии. Это падение вызвано увеличением емкости кабеля, ростом амплитуды помех и естественным затуханием амплитуды полезного сигнала с ростом параметра L . Третий участок определяется существованием некоторой критической длины линии, при достижении которой предельная скорость передачи информации стремится к нулю и передача невозможна. Это обусловлено тем фактом, что при превышении критической длины линии уровень помех становится соизмеримым с уровнем сигнала, независимо от скорости передачи информации. Каждая среда интерфейса определяется конкретным значением $V_{инф}$ на первом участке зависимости и значением наклона на втором.

Передача информации по информационному каналу может осуществляться как по одно-, так и по двухпроводной линиям. Для реализации первой из них достаточно соединить приемник и передатчик единственным проводником (рис. 1.11).

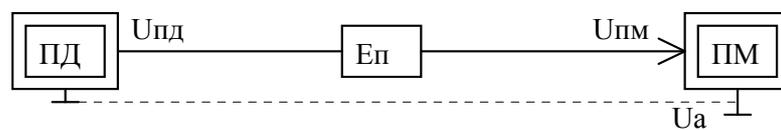


Рис. 1.11. Схема однопроводной передачи бита информации

В этом случае, совокупное влияние соседних сигналов и помех на данную линию представим эквивалентным генератором E_p . Учтем также тот факт, что при значительно протяженных линиях и применении различных источников питания для приемника (ПМ) и передатчика (ПД), неизбежно возникает некоторая разность потенциалов (U_a) между их общими шинами.

В этом случае для правильного выделения приемником полезного сигнала на линии, в самом неблагоприятном случае, должны выполняться следующие соотношения:

$$\begin{aligned} U_{пм}(1) &= U_{пд}(1) - (E_p + U_a) \\ U_{пм}(0) &= U_{пд}(0) + (E_p + U_a) \\ U_{пм}(1) - U_{пм}(0) &> U_{доп}, \end{aligned} \quad [1.1]$$

где: $U_{пд}(1)$ и $U_{пд}(0)$ – соответственно напряжение логической единицы и логического нуля на выходе передатчика;

$U_{пм}(1)$ и $U_{пм}(0)$ – то же на входе приемника;

$U_{доп}$ – допустимая разность потенциалов зоны перехода логических уровней на входе приемника. То есть выполнение последнего из соотношений [1.1] гарантирует достоверность передаваемых сообщений. Однако выполнение этого неравенства напрямую зависит от суммы $(E_p + U_a)$.

Из уравнения [1.1] можно легко исключить параметр U_a , если перейти от однопроводной линии к двухпроводной, имеющей общую точку заземления на выходе передатчика (рис. 1.12). Действительно, если принять во внимание синфазность изменения уровней помех E_p на каждой из линий и положить $U_a = 0$, то тогда:

$$\begin{aligned} U_{пм}(+) &= E_p, \\ U_{пм}(-) &= U_{пд} + E_p \end{aligned} \quad [1.2]$$

Следовательно, сигнал на входе приемника будет определяться из выражения:

$$U_{пм} = U_{пм}(-) - U_{пм}(+) = U_{пд} \quad [1.3]$$

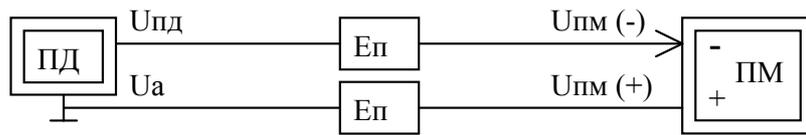


Рис. 1.12. Схема двухпроводной передачи бита информации

Таким образом, в данном случае, сигнал на входе приемника равен сигналу на выходе передатчика, что позволяет существенно увеличить скорость передачи информации и расстояние между приемником и передатчиком по отношению к однопроводной линии. В качестве физических элементов связи между составными частями системы в двухпроводных линиях используют либо витые пары, либо коаксиальный кабель. Единственный недостаток двухпроводной линии заключается в необходимости выполнения следующего условия:

$$U_a < U_{вх.макс}, \quad [1.4],$$

где: $U_{вх.макс}$ – максимально допустимое входное напряжение дифференциального каскада приемника.

От этого недостатка освобождена двухпроводная линия, в которой в качестве передатчика используется балансный усилитель. По этой схеме организовано взаимодействие в широко распространенном последовательном интерфейсе RS-232C (отечественный аналог – стык С2). При передаче логической единицы потенциал линии $U_a = -U_{пд}$, а линии $U_b = +U_{пд}$, в то время как при передаче логического нуля потенциалы обеих линий взаимно изменяются (рис. 1.13). Потенциалы $-U_{пд}$ и $+U_{пд}$ определяются питанием схем передатчика и могут принимать значения соответственно в диапазонах минус (12В – 3В) и плюс (3В – 12В). Кроме согласованности уровней напряжений в линии, здесь есть еще одно преимущество – низкие сопротивления каждой из линий, что существенно снижает влияние помех на уровень полезного сигнала.

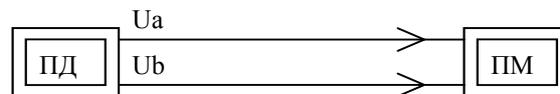


Рис. 1.13. Последовательный интерфейс RS-232C (фрагмент)

1.5. Системные магистрали персональных компьютеров

Как отмечалось выше, системная магистраль предназначена для организации взаимодействия между составными частями компьютера. Разнообразие существующих типов шин расширения для ПК является результатом эволюции – исходный стандарт первого персонального компьютера постепенно развивался, и дополнительно вводились новые. В зависимости от фирмы-изготовителя компьютер поддерживает один или несколько из следующих стандартов.

Стандарт XT-bus

Шины XT-bus использовались в системе микропроцессоров i8086 и i8088. Количество линий шины данных XT-bus ограничивалось восемью (SD0-SD7). Кроме того, адресная шина включала 20 линий (SA0-SA19). Эти линии давали информацию об адресе, который в данный момент используется. В реальном режиме они позволяли адресовать весь доступный в нем 1 Мбайт памяти. Для работы с внешними устройствами в этой шине были предусмотрены также 4 линии аппаратных прерываний (IRQ) и 4 линии для требования внешними устройствами прямого доступа в память (DMA). Шина адреса однонаправленная, а шины данных и управления двунаправленные. Конструктивно магистраль объединяла до восьми 62-контактных двухрядных разъема. Системная магистраль и микропроцессор синхронизировались от одного тактового генератора с частотой 4,77 МГц. Таким образом, теоретически скорость передачи данных могла достигать более 4,5 Мбайт/с.

Стандарт AT-bus

Стандарт AT-bus появился в 1984 году вместе с выпуском семейства компьютеров на базе микропроцессора i80286. В стандартном PC/AT компьютере часть системной шины, полностью

совместимая с шиной компьютера PC/XT, образует восемь разъемов по 62 вывода в каждом, расположенных в два ряда и пронумерованных как A1-A31, B1-B31. Кроме того, к системной шине также принадлежат шесть разъемов-расширений по 36 выводов с двухрядным расположением и пронумерованных как C1-C18, D1-D18. Эти шесть разъемов расположены как продолжение шести из восьми XT-разъемов.

Наличие у AT-bus дополнительного 36-контактного слота расширения приводит к увеличению количества адресных линий на четыре, а данных – на восемь. Здесь уже становится возможным передавать параллельно 16 разрядов данных, а благодаря 24 адресным линиям напрямую обращаться к 16 Мбайтам оперативной памяти. Полная блок-схема магистралей PC/AT представлена на рис. 1.16.

Магистраль L-типа объединяет выводы микропроцессора i8086, i80286 или i80386 и является вспомогательной по отношению ко всем остальным частям компьютера. Основной шиной, связывающей компьютер в единое целое, является системная магистраль S-типа, которая непосредственно выведена на все слоты расширения. Локальные шины X- и M-типа предназначены для сопряжения микросхем на системной плате и отделены от системной магистрали буферами (шинными формирователями).

Системная магистраль поддерживает порты ввода/вывода в диапазоне адресов 100H-3FFH. При обращении к памяти шина адреса SA0-SA19 активизируется в течении действия строба VALE, защелкивание адреса происходит по падающему фронту этого сигнала. Источником сигнала SA0-SA19 со стороны системной платы могут являться либо буферные регистры, относящиеся к процессору, либо буферные регистры, относящиеся к контроллера прямого доступа к памяти (ПДП).

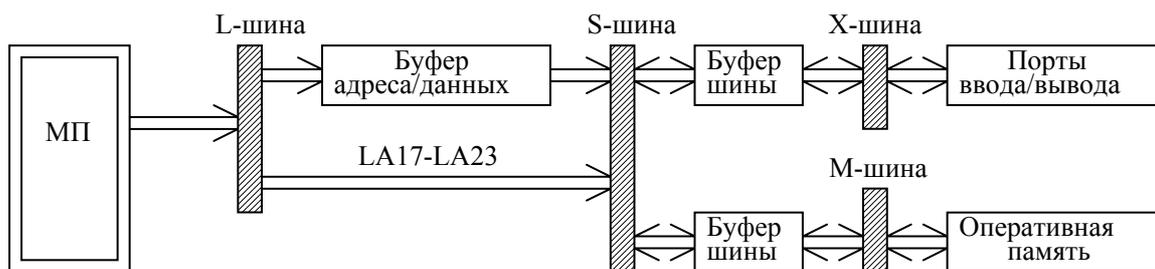


Рис. 1.16. Блок схема магистралей PC/AT

Линии LA17-LA23 дают старшую часть адреса при работе процессора в защищенном режиме, когда доступ возможен ко всем 16 Мбайтам оперативной памяти. Эти старшие разряды удерживаются в течение всего цикла обращения к памяти и не требуют запоминания.

Сигналы IRQ3-IRQ7, IRQ9-IRQ12, IRQ14, IRQ15 формируются платами периферийных адаптеров для запроса процессора на обслуживание прерывания от периферийного устройства. Эти сигналы обрабатываются контроллерами прерываний. Приоритеты запросов следуют в порядке IRQ7-IRQ3, IRQ15, IRQ14, IRQ12-IRQ9 возрастания приоритетов. Запрос на прерывание принимается по возрастающему фронту сигнала IRQ. Этот сигнал сохраняется высоким, пока процессор не даст подтверждение запроса на прерывание. Запросы IRQ0-IRQ2, IRQ8 и IRQ13 используются только на системной плате, на магистраль не выводятся, т.е. недоступны для периферийных адаптеров.

Сигналы запросов на прямой доступ к памяти DRQ0-DRQ3 и DRQ5-DRQ7 выставляются периферийными устройствами при необходимости обмена данными между этими устройствами и памятью. Обмен ведется по системной шине в режиме ПДП минуя микропроцессор. Запросы DRQ идут на соответствующие входы 2-х контроллеров ПДП. Запрос на прямой доступ генерируется переводом соответствующей DRQ-линии в состояние с высоким уровнем сигнала. Затем состояние высокого потенциала поддерживается до прихода сигнала DACK (DACK0-DACK3 и DACK5-DACK7) подтверждения запросов на ПДП (на что указывает низкий уровень этого сигнала). Сигнал DACK формируется контроллерами ПДП. Для стандартного исполнения данной магистрали линии DRQ0 и DACK0 имеют наивысший приоритет и используются для организации циклов регенерации ОЗУ. Применение этих линий для других целей нежелательно, т.к. может

привести к полной потере информации в ОЗУ. Далее в порядке возрастания номеров запросов идет уменьшение их приоритетов. Линии DRQ0-DRQ3 организуют байтовый обмен информацией, DRQ5-DRQ7 – двухбайтовый обмен. Запрос DRQ4 используется только на системной плате.

Сигнал IOR сообщает периферийным устройствам о чтении содержимого порта одного из них. Он выставляется либо шинным контроллером при чтении порта периферийного устройства, либо контроллером ПДП при организации переноса данных напрямую из периферийного устройства в оперативную память.

Сигнал IOW информирует периферийные устройства о записи в порт одного из них. Адрес порта находится на линиях SA0-SA15. Этот сигнал формируется либо шинным контроллером, либо ПДП-контроллером при проведении переноса данных по шине из оперативной памяти в порт периферийного устройства напрямую, минуя процессор.

Сигнал MEMR информирует схему управления памятью о чтении слова или байта из оперативной памяти и на шине формируется шинным контроллером, а во время циклов ПДП этот сигнал выставляется контроллером ПДП.

Сигнал MEMW сообщает о записи слова или байта в оперативную память и напрямую формируется шинным контроллером, а во время ПДП сигнал MEMW формируется ПДП-контроллером.

Отличительной особенностью шины специальных управляющих сигналов, является наличие двух линий тактовых импульсов CLK и OSC, по которым передаются синхроимпульсы скваженностью 2 с частотой 6-10 МГц (CLK) и фиксированной частотой 14,31818 МГц (OSC). Частота работы микропроцессора совпадает с частотой работы системной магистрали.

Стандарт ISA.

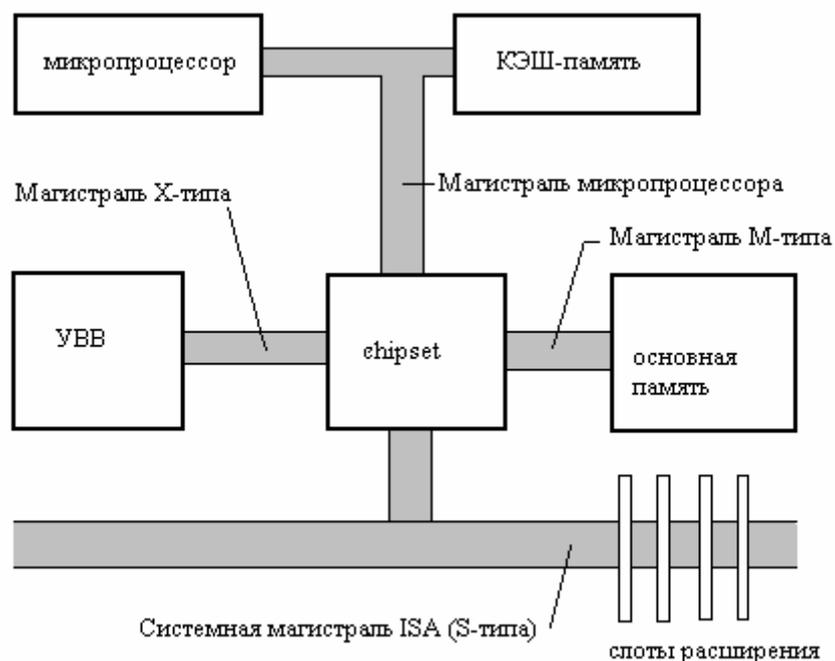


Рис. 1.17. Архитектура магистрали ISA

Стандарт ISA (Industry Standard Architecture) - промышленная стандартная архитектура, в том виде, в котором его используют сейчас, является дальнейшим развитием магистрали AT-bus и конструктивно полностью ее повторяет. Однако архитектура построения магистрали ISA принципиально другая (рис 1.17). Ядром магистрали, которое объединяет все компоненты системы в единое целое, является *chipset* – специализированная БИС, поддерживающая взаимодействие между различными типами магистралей на системной плате. В ISA архитектуре

конструктивно по другому решен вопрос установки основной памяти ПК: для этих целей используют специальные гнезда, в которые устанавливаются модули SIMM или DIMM.

В компьютерах PC/AT, использующих микропроцессор i80286 впервые стала применяться шина ISA, полностью реализующая возможности этого микропроцессора. В ее составе два типа ISA-шинных разъемов (слотов расширения): два 8-разрядных для стандартного PC/XT (XT-bus) и шесть 16-разрядных для PC/AT (AT-bus).

Не следует отождествлять частоту работы системной магистрали с частотой работы микропроцессора. В современных ПК они не совпадают.

Системные платы с шиной ISA допускают возможность синхронизации работы самой шины и микропроцессора разными тактовыми частотами, что позволяет устройствам, выполненным на платах расширения, работать медленнее, чем базовый микропроцессор. Это стало особенно актуальным, когда тактовая частота процессоров превысила 10-12 МГц. Системная шина ISA работает асинхронно с процессором на частоте 8,33 МГц. Таким образом, максимальная скорость передачи теоретически может достигать более 16 Мбайт/с. Шины X- и M- типов также могут иметь разные частоты.

Шина ISA определяет некоторые стандартные требования к коммуникационным средствам ПК и предназначена для того, чтобы исключить конфликты между различными платами расширения. Однако в данном случае эта задача выполняется не совсем полностью, т.к. понятие стандарта здесь трактуется достаточно свободно. Именно поэтому приходится на интерфейсных платах устанавливать DIP-переключатели и перемычки в **достаточном количестве**.

Стандарт MCA.

MCA (Micro Channel Architecture) – Микроканальная архитектура была разработана фирмой IBM в 1987 году вместе с выпуском семейства персональных компьютеров PS/2.

Оригинальная 16-разрядная (а позже и 32-разрядная) системная шина MCA была использована фирмой IBM в качестве принципиально нового способа реализации системной шины. Отметим, что это семейство основано на архитектуре из пяти процессоров фирмы Intel, три из которых используют магистраль Micro Channel, логически совместимую с магистралью AT-bus. Шина MCA была специально разработана и запатентована фирмой IBM для защиты от воспроизведения данной шины конкурентами. Конструктивно она полностью не совместима с магистралью ISA.

Основой магистрали является 58-контактный двухрядный соединитель, который разделен на три секции: 8-разрядную секцию с шиной адреса (A00-A23), шиной данных (D00-D07) и большинством сигналов управления; секцию 16-разрядного расширения со старшим байтом данных (D08-D15), дополнительными сигналами прерывания и секцию видеорасширения, предназначенную для обслуживания видеосистемы. Последняя из перечисленных секций конструктивно выполнена в виде отдельного соединителя и имеет одно установочное место на системной плате. Кроме того, в более поздних моделях ПК PS/2 используется дополнительная секция 32-разрядного расширения для микропроцессора i80386, работающего с тактовой частотой 20 МГц, и согласованной с ним памяти. Эта линия содержит восемь адресных (A24-A31) линий, 16 линий данных (D16-D31) и семь управляющих линий.

В результате быстродействие системы возросло до 3-5MIPS и стало возможным непосредственное обращение к 4 Гбайт физического адресного пространства или 64Тбайт (46 двоичных разрядов) виртуального адресного пространства памяти.

Архитектура системной магистрали MCA имеет встроенный арбитраж, который позволяет реализовать многопроцессорный режим работы. Так, при необходимости занятия магистрали устройство вырабатывает 4-разрядный код уровня арбитража (ARB0-ARB3) и активизирует линию запроса магистрали Preempt. Арбитр, расположенный на системной плате, получив сигнал Preempt, возбуждает линию Arb/Gnt, сообщая о начале цикла арбитража. Устройство, имеющее наивысший приоритет, подключается к магистрали. Арбитр снимает сигнал с линии Arb/Gnt, сообщая, что данное устройство может управлять магистралью.

В тех случаях, когда осуществляется групповая (пакетная) передача данных, вырабатывается сигнал Burst, сообщающий о захвате системной магистрали периферийным устройством. Однако при возбуждении линии Preempt, данное устройство должно участвовать в

новом цикле арбитража. Если в течение 7,8 мкс после формирования сигнала Preempt устройство не отключается от магистрали, специальный таймер фиксирует ошибку.

На системной магистрали возможны три типа цикла: согласованной памяти (без тактов ожидания МП), основной передачи, расширенной синхронной и асинхронной передач. Первые три являются специальными синхронными циклами.

Цикл согласованной памяти поддерживается только на ПК типа PS/2 модели 80. Он обеспечивает быструю передачу данных между микропроцессором и магистралью. Для реализации этого цикла используется секция согласованной памяти, находящаяся на системной магистрали и формирующая три управляющих сигнала. Если адресуемое устройство вырабатывает сигнал запроса MMCR, то выполняется цикл согласованной памяти, если нет – то цикл основной передачи. Сигнал MMC вырабатывается системной платой, выполняющей быстрый цикл передачи. Сигнал MMC CMD указывает на установку данных.

Выполнение цикла основной передачи начинается с возбуждения шины адреса и сигналов MADE 24 и M/IO. Затем устанавливаются сигналы состояния (S0 и S1) и возбуждается сигнал строга адреса ADL. Адресуемое устройство формирует сигнал CD Sfdbk и сигнал CD DS16 при 16-разрядной передаче или два сигнала CD DS16 и CD DS32 при 32-разрядной передаче. Если цикл основной передачи должен быть расширен, то снимается сигнал CD Ch RDy. Сигнал CMD формируется при установке шины данных.

При завершении цикла основной передачи снимаются сигналы с шины адреса и линий ADL, S0, S1. В ответ на изменение адресных шин адресуемое устройство снимает сигналы с возбуждаемых им линий.

В расширенном синхронном цикле передача данных завершается после формирования устройством сигнала готовности CD Ch RDy. Цикл отличается от асинхронного формированием синхронного сигнала готовности и завершается немного быстрее.

Конструктивно магистраль MCA объединяет до семи слотов расширения. Для модели 80 ПК PS/2 это четыре 16-разрядных разъема и три 32-разрядных. Следует также отметить, что в данной магистрали впервые выполнена поддержка автоконфигурирования интерфейсного оборудования, известная в настоящее время как спецификация Plug&Play. На системной и интерфейсных платах полностью отсутствуют переключки и DIP-переключатели. В компьютерах с шиной MCA могут устанавливаться слоты расширения шести типов:

- 16-разрядные;
- 16-разрядные с дополнительной секцией для плат памяти;
- 16-разрядные с дополнительной секцией для видеоадаптеров;
- 32-разрядные;
- 32-разрядные с дополнительной секцией для плат памяти;
- 32-разрядные с дополнительной секцией для видеоадаптеров;

Стандарт EISA.

EISA (Extended Industry Standard Architecture) – Расширенная промышленная стандартная архитектура была предложена в 1989 году как альтернатива магистрали MCA.. Эта шина была разработана консорциумом девяти ведущих фирм (AST, Compaq, Epson, HP, NEC, Olivetti, Tandy, Wyse и Zenith), изготавливающих высококачественные IBM-совместимые персональные компьютеры, в качестве 32-разрядной системной шины, работающей на тактовой частоте 10 МГц. Таким образом скорость передачи данных по шине возросла до 20 Мбайт/с.

В отличие от MCA шина EISA является открытой шинной архитектурой, которую могут повторять и использовать все желающие изготовители компьютеров и их компонентов. Эта системная шина обеспечивает возможный объем адресуемой памяти до 4 Гбайт (для i80386/486), 32-разрядную передачу данных, в том числе и в режиме ПДП, а также аппаратную поддержку операций арбитража, улучшенную систему прерываний по сравнению со стандартом ISA и автоматическую конфигурацию системы.

Кроме того, шина EISA совместима со всеми платами расширения, рассчитанными на шину ISA. Это обеспечивается простым, но поистине гениальным конструктивным решением. EISA-разъемы имеют два вертикальных ряда контактов, один из которых (верхний) использует сигналы шины ISA, а второй (нижний) – соответственно EISA. Контакты в слотах расширения расположены так, что рядом с каждым сигнальным контактом находится контакт "Общий".

Благодаря этому сводится к минимуму вероятность генерации электромагнитных помех и уменьшается восприимчивость к ним.

Доступ к адресному пространству может иметь не только центральный процессор, но и платы устройств, способных управлять передачей данных по шине, а также устройства, имеющие возможность организовать режим ПДП.

Теоретически максимальная скорость передачи данных по шине EISA в пакетном режиме может достигать 33 Мбайт/с.

Как и для шины ISA, в системе EISA имеется семь каналов ПДП. Контроллеры ПДП имеют возможность поддерживать 8-, 16- и 32-разрядные режимы передачи данных.

Поскольку запросы прерываний шины ISA передаются в виде перепадов уровней напряжения (фронтов сигналов), т.е. сильно подвержены импульсным помехам, то дополнительно в системе EISA предусмотрены также сигналы прерываний активные по уровню.

Каждый изготовитель плат расширения для компьютеров с шиной EISA предоставляет вместе с этими платами специальные файлы, обеспечивающие автоматическое конфигурирование системы. Для плат адаптеров, предназначенных для шины ISA, пользователь должен сам подобрать правильное положение DIP-переключателей и перемычек, однако, сервисная программа на компьютерах, использующих магистраль EISA, позволяет отображать установленные положения соответствующих переключателей на экран монитора и дает некоторые рекомендации по правильной их установке. Каждому разъему расширения отводится адресный диапазон 4 Кбайта, что позволяет избежать конфликтов между отдельными платами расширения EISA.

Конструктивно системная магистраль EISA может объединять до 15 слотов расширения. Однако компьютеры, использующие системные платы с шиной EISA, достаточно дорогие. К тому же скорость передачи данных увеличивается, в основном, благодаря увеличению разрядности шины данных. Поэтому данная системная магистраль не так часто встречается в ПК, и впоследствии была полностью вытеснена магистралью PCI.

Локальные шины.

Компьютеры на базе микропроцессоров i80486, обладали большой вычислительной производительностью и позволяли обеспечить достаточно быструю (комфортную) работу графических видеосистем, реализующих большое количество оттенков цветов при высоком разрешении. В связи с этим, а также для ускорения обмена данными с контроллерами жестких дисков и интерфейсными сетевыми платами, возникла необходимость в использовании специализированных высокоскоростных шин. Выход был найден в применении рядом фирм внутренних шин называемых *локальными*, которые напрямую связывают микропроцессор с контроллерами периферийных устройств более широким и быстрым каналом передачи информации по сравнению с традиционными системными шинами.

Первые IBM-совместимые компьютеры с локальными магистралями не имели определенного стандарта на эти шины. Одна из ведущих изготовителей персональных компьютеров, впервые связавшая видеосистему с микропроцессором через локальную шину была компания NEC Technologies. Еще в 1991 году эта фирма представила свою оригинальную разработку Image Video, позволяющую существенно увеличить производительность видеосистем. В результате последующих разработок других фирм появилось несколько вариантов локальных шин, среди которых наибольшее распространение получили две из них, признанные промышленными: VL-bus (VESA Local Bus), предложенная ассоциацией VESA (Video Electronics Standards Association), и PCI (Peripheral Component Interconnect), разработанная фирмой Intel (1993 год).

VL-bus – это 32-х разрядная локальная шина, работающая на тактовой частоте процессора. Эта шина непосредственно управляется процессором i80486 и поддерживает одновременно до трех слотов расширения. Некоторые изготовители, впрочем, убеждены, что добиться устойчивой работы трех устройств на высоких частотах практически невозможно, и устанавливают на свои системные платы не более двух слотов. Ограничение на количество слотов связано с тем, что электрическая нагрузочная способность на сигнальные линии любого микропроцессора весьма невелика.

В качестве устройств, подключаемых в VL-bus, в настоящее время выступают контроллеры накопителей, видеоадаптеры и сетевые платы. Конструктивно VL-bus выглядит как короткий соединитель типа MCA (112 контактов), установленный, например, рядом с разъемами

расширения ISA или EISA. При этом 32 линии используются для передачи данных и 30 – для передачи адреса. Максимальная скорость передачи информации по шине теоретически может составлять около 130 Мбайт/с. На VL-bus отсутствует арбитр шин, т.к. большинство подключаемых к ней устройств являются "пассивными", т.е. сами не инициируют передачу данных. Тем не менее, во избежание возможных конфликтов между подключенными к шине устройствами в спецификации выделяются "управляющие" (master) и "управляемые" (slave) адаптеры. Для "управляющих" устройств на системных платах обычно определены свои "управляющие" слоты расширения. По замыслу разработчиков, подобные "управляющие" устройства могут осуществлять арбитраж на данной шине.

В настоящее время VL-bus является сравнительно недорогим, простым, но обеспечивающим высокую скорость передачи данных дополнением для компьютеров на базе 486-х процессоров с шиной ISA, причем с обеспечением обратной совместимости.

После появления процессора Pentium ассоциация VESA приступила к работе над новым стандартом VL-bus (версии 2), который предусматривает, в частности 64-х разрядную шину данных и увеличение количества слотов расширения предположительно до трех разъемов на 40 МГц и двух – на 50 МГц. Ожидаемая скорость передачи информации теоретически должна возрасти до 400 Мбайт/с.

Современные ПК на 486-х процессорах работают на частоте от 33 до 100 МГц. А большинство процессоров Pentium работают на внешней частоте 60 МГц и выше, поэтому интерфейсные платы для VL-bus должны функционировать на слишком высокой частоте – выше той, которую способны поддерживать большинство периферийных устройств.

Шина PCI.

В 1993 году фирма Intel предложила новую чрезвычайно перспективную шину PCI. В отличие от VL-bus, шина PCI работает на фиксированной тактовой частоте 33 МГц, т.е. по отношению к микропроцессору не является локальной шиной. Однако, благодаря дополнительным функциям и буферизации, PCI-платы работают быстрее, чем устройства VL-bus. Специальный контроллер отвечает за разделение управляющих сигналов локальной шины процессора и PCI-шины, и кроме того, осуществляет арбитраж на шине PCI. Именно поэтому данная шина может использоваться на различных компьютерных платформах. Следует отметить, что гибкость и быстрдействие этой шины предполагают и большие аппаратные затраты, чем для VL-bus. Тем не менее шина PCI практически стала стандартом для систем на базе Pentium и не менее успешно используется в 486-х компьютерах. Таким образом шину PCI, в отличие от EISA и VL-bus следует считать дальнейшим, совершенно новым развитием системных магистралей для персональных компьютеров.

В соответствии со спецификацией PCI к шине могут подключаться до 10 периферийных устройств. Причем каждая плата расширения PCI может разделяться между двумя периферийными устройствами, поэтому уменьшается общее число устанавливаемых разъемов. Шина PCI предусматривает напряжение питания для интерфейсных плат как 5 В, так и 3,3 В, а также обеспечивает режим их автоконфигурации (спецификация plug and play - "включай и работай"). Интерфейсные платы, рассчитанные на различное напряжение, должны устанавливаться в разные слоты расширения. Впрочем, такое деление существовало недолго. Выпуск так называемых универсальных PCI-адаптеров, которые способны работать в любом из слотов, в настоящее время полностью вытеснил первые модели.

Шина PCI использует 124-контактный (32-разрядная передача данных) или 188-контактный разъем (64-разрядная передача данных), при этом теоретически возможная скорость обмена составляет соответственно 132 и 264 Мбайт/с. Спецификация PCI 2.0 в расчете на микропроцессор Pentium (100 МГц и выше) определяет работу шины с частотой 33-66 МГц и скоростью обмена до 520 Мбайт/с. В стандартной конфигурации ПК на системных платах устанавливается обычно не более трех-четырёх слотов расширения PCI.

В качестве первого удачного примера широкого практического использования шины PCI можно привести многофункциональную рабочую станцию AcerPower/M (рис. 1.18). Различают два варианта исполнения: Acer ChipUp (для компьютеров на базе i486 микропроцессора) и Acer ModuFlex (для компьютеров на базе процессоров Pentium).

В настоящее время шина PCI стала основной магистралью для систем на базе микропроцессоров Pentium и де-факто считается стандартом среди шин общего назначения. Она надежно «закрепилась» в компьютерном мире и многие фирмы-изготовители компьютеров (не IBM) интенсивно ее используют в своих разработках, например такие, как DEC и Apple. Успех шины связан, в первую очередь, со следующими ее основными характеристиками:

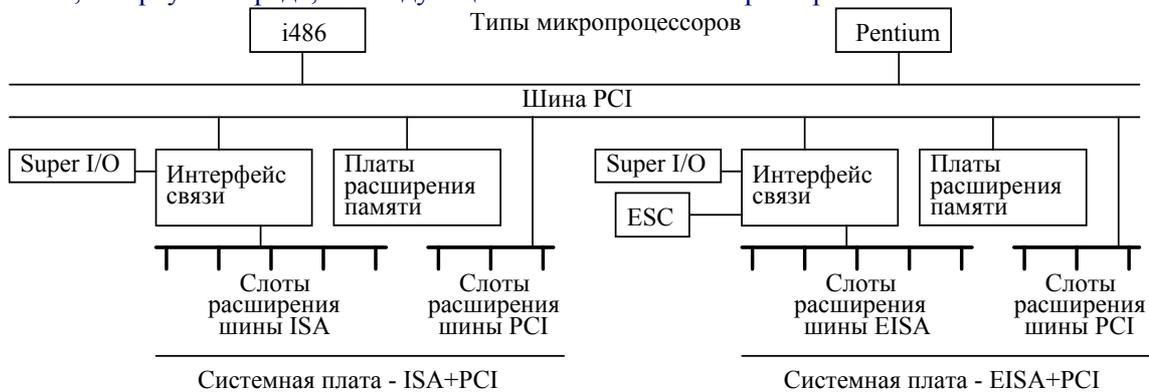


Рис. 1.18. Архитектура AcerPower/M

В шине используется совершенно иной способ организации взаимодействия между приемником и передатчиком. Когда передающее устройство готово к передаче, оно устанавливает данные на магистраль и сопровождает их синхроимпульсом. Приемник, получив данные, сохраняет их в регистре и выдает подтверждение. При этом все сигналы синхронизированы относительно тактовой частоты работы шины.

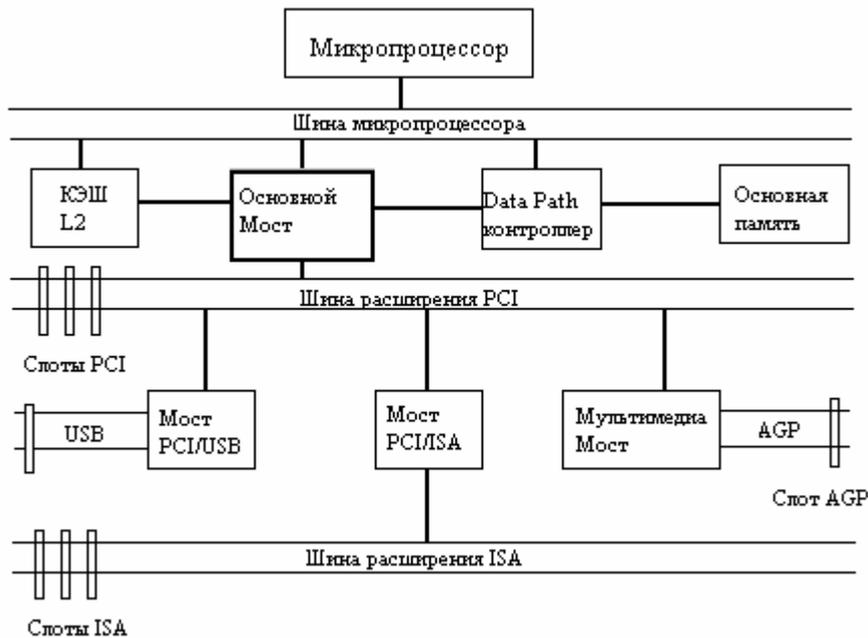
Пакетной передачей данных управляет не микропроцессор а мост, включенный между шиной и микропроцессором. Это дает возможность продолжить работу микропроцессора одновременно с обменом данными между другими составными частями системы.

Шина PCI использует принцип временного мультиплексирования, то есть когда передача кодов адреса и данных осуществляется по общим интерфейсным линиям.

В шине PCI полностью реализована спецификация Plug&Play, позволяющая автоматически распознавать установленное в ней оборудование, анализировать конфигурацию системы и выполнять их конфигурирование программными средствами.

Рис. 1.19. Архитектура шины PCI

Современные компьютеры, построенные на базе шины PCI имеют антресольную



архитектуру системной магистрали, чаще называемую архитектурой шины PCI, в которой задействованы как набор локальных шин, так и хорошо зарекомендовавшая себя шина ISA (рис. 1.19). Верхняя антресоль представляет собой высокоскоростную шину микропроцессора, по которой осуществляются обращения к памяти. Связь данной шины с шиной расширения PCI осуществляется с помощью главного (основного) моста – специализированной БИС, которую чаще отождествляют с chipset. Наличие моста позволяет использовать в системе несколько типов микропроцессоров путем простой их замены в рамках определенного перечня.

Нижняя антресоль является вспомогательной и представлена традиционной шиной ISA. Следует заметить, что работа данной шины эмулируется с помощью моста PCI/ISA. Мост содержит набор устройств, позволяющий полностью поддерживать стандарт ISA, а именно: два контроллера DMA, контроллер клавиатуры, таймер, контроллеры прерываний и др.

Средний уровень представлен семейством шин: основная шина PCI и дополнительные шины USB, AGP. USB шина является универсальной последовательной шиной, предназначенной для подключения таких устройств, как клавиатура, мышь, джойстик, сканер, принтер и др. Устройства подключаются к компьютеру через специальный USB разъем. Если к шине подключить USB концентратор, то возможное количество одновременно подключаемых к компьютеру устройств возрастает до 127. Тем самым высвобождаются основные слоты расширения. Шина поддерживает скорость обмена информацией 12Мбит/с.

Современное программное обеспечение насыщено графической информацией, которая циркулирует по шинам компьютера. В результате трафик работы шины становится перегруженным и производительность системы резко падает. Для разделения потоков графической и не графической информации было предложено вместе с основной PCI шиной использовать дополнительную параллельную AGP шину, которая по логике работы, тактовой частоте, разрядности и т.п. полностью дублирует шину PCI. Однако эта шина предназначена исключительно для потоков мультимедиа - информации: видео, звука, графики, мультимедиа. Как правило шина имеет единственный слот расширения, предназначенный для установки в него видеоадаптера. Остальные компоненты мультимедиа интегрированы на системной плате и подключены к AGP шине непосредственно. AGP шина связана с основной PCI шиной через специальный мультимедиа мост, который синхронизирует работу обеих шин. Установка AGP видеоадаптера в слот расширения позволяет видеосистеме напрямую обращаться к основной памяти компьютера, минуя микропроцессор, и «забирать» памяти столько, сколько это необходимо для данного режима работы видеосистемы.

Глава 2. Системы ввода/вывода персональных компьютеров

2.1. Базовая система ввода/вывода персональных компьютеров

Базовая система ввода/вывода (BIOS) является неотъемлемой составной частью современного ПК. Физически она размещается в нескольких микросхемах ПЗУ на системной плате компьютера и содержит от 64 до 128 Кбайт информации. Эта информация представлена в виде подпрограмм, которые логически объединены в следующие блоки: POST – процедуры, собственно BIOS, SETUP и внесистемный загрузчик.

Тест самоконтроля (POST) начинает автоматически выполняться при включении питания компьютера или при нажатии клавиши RESET. Это так называемая “холодная” перезагрузка компьютера. POST представляет собой совокупность подпрограмм, предназначенных для тестирования наиболее важных составных частей компьютера, таких как клавиатура, оперативная память, видеосистема и ряд других. При выявлении ошибки компьютер либо выдает комбинацию длинных или коротких гудков, либо высвечивает на экране монитора номер соответствующей ошибки.

При включении компьютера в течение примерно 0,3-0,5 секунд выполняется самотестирование блока питания. В случае, если все уровни напряжений питания находятся в допустимых пределах, на системную плату поступает сигнал Power Good, вырабатываемый блоком питания. Поступление этого сигнала эквивалентно сигналу Reset, получив который микропроцессор переходит к процедуре POST, записанной в системном ROM BIOS.

Действительная последовательность тестирования отличается у разных производителей BIOS, так как содержимое ПЗУ является уникальным для каждой модели компьютера. В общем случае сначала производится тестирование наиболее важных элементов, а затем – наименее важных. Рассмотрим некоторый усредненный вариант.

После получения сигнала Reset микропроцессор начинает выполнять команды с инструкции, записанной по адресу xxxxFFFF:0000h. Для IBM-совместимых компьютеров первой такой инструкцией будет команда “длинного” перехода (FAR JUMP) на начало процедуры POST, располагаемой в сегменте xxxxF000h. В первую очередь тестируется сам микропроцессор. Тесты включают выполнение базовых команд, работу с флагами, проверку АЛУ и регистров общего назначения.

После проверки микропроцессора процедура POST проверяет правильно ли считываются инструкции из системного ROM BIOS. Далее POST тестирует микросхему CMOS RAM, которая обеспечивает сохранность записанной информации.

Далее следует проверка таймера, нулевой канал которого связан с процессором через программируемый контроллер прерываний, генерирующего прерывания примерно через каждые 18,2 с. Первый канал таймера используется для запуска контроллера ПДП. Второй канал обеспечивает генерацию звука через динамики или измерение временных интервалов. Сначала POST тестирует первый канал таймера и инициализирует его на требуемую частоту регенерации памяти, а затем инициализирует нулевой канал. Таким образом становятся возможными обращения к оперативной памяти.

Далее POST начинает выполнять тесты записи/чтения регистров контроллера ПДП и, если не возникает ошибок, приступает к периодической регенерации первых 64 Кбайт оперативной памяти (системной памяти). В этой же фазе может тестироваться контроллер клавиатуры, поскольку он имеет собственную подпрограмму самотестирования.

Обязательно должны быть протестированы на системной плате контроллер прерываний и КЭШ-контроллер (при наличии КЭШ-памяти). Поскольку видеосистема еще не инициализирована, то все сообщения о возможных ошибках выдаются последовательностью звуковых сигналов. В таблице 2.1 приведены наиболее часто используемые основные звуковые сигналы POST.

Таблица 2.1. Звуковые сигналы ошибок, обнаруженных POST

Место ошибки	Звуковая последовательность.
Источник питания или системная плата.	Повторяющиеся короткие или непрерывные звуковые сигналы.
Системная плата, плата дисплея или источник питания.	Один длинный и один короткий звуковой сигнал.
Системная плата или источник питания.	Один длинный и два коротких звуковых сигнала.
То же.	Один короткий звуковой сигнал и пустой дисплей.

На завершающей стадии POST осуществляется проверка и инициализация таких периферийных устройств, как видеосистема, клавиатура, диски, последовательные и параллельные порты и т.п.

Последняя фаза POST начинается, как правило, с проверки данных, сохраняемых в CMOS RAM. Важнейшей операцией POST на текущем этапе является инициализация видеосистемы. После этого POST может выдавать уже не только звуковые сигналы, но и сообщения об ошибках на экран. В таблице 2.2 приведены номера некоторых таких ошибок.

Таблица 2.2. Некоторые коды ошибок, обнаруженных POST

Номер кода ошибки	Место ошибки
101-109	Системная плата
161	Батарея
162	Ошибка конфигурации
163	Ошибка даты и времени
165	Системные опции не установлены
201-203	Микросхема памяти
301-304	Клавиатура отключена, заблокирована или неисправна
401, 501	Монитор или видеоадаптер
601	Дисковод или контроллер
602	На диске A: отсутствует Boot-сектор
701	Математический сопроцессор
1100-1199 или 1200-1299	Последовательный порт
1401	Параллельный порт
1780, 1790	Первый жесткий диск
1781, 1791	Второй жесткий диск
1782	Контроллер жесткого диска
2400-2499	Системная видеоплата

После инициализации видеосистемы POST тестирует оставшуюся оперативную память компьютера. Обычно эта работа сопровождается “щелчками” из динамика и высвечиванием работоспособного количества проверенных килобайт. В этот период POST записывает данные в каждую ячейку ОЗУ, вычисляет правильный бит четности и сравнивает его с битом четности, полученным в процессе чтения. Если при этом установлено отличие, то ячейка памяти считается дефектной и в левом верхнем углу монитора POST выдает сообщение об адресе дефектной ячейки.

После тестирования памяти процедура POST проверяет работоспособность клавиатуры, микросхем портов и контроллеров дисков.

Нормальное завершение процедуры POST, т.е. если фатальных ошибок при тестировании не обнаружено, сопровождается одним коротким гудком динамика. Затем осуществляется пуск подпрограммы, которая активизирует дисководы компьютера под управлением BIOS.

Основные функции собственно BIOS заключаются в организации взаимодействия между электронными узлами ПК, а также между ними и внешним программным обеспечением. BIOS

представляет собой набор подпрограмм, выполнение которых активизирует конкретные модули и узлы ПК. С помощью этих подпрограмм осуществляется ввод данных с клавиатуры, формируется отображение информации на мониторе, выполняются функции чтения/записи файлов на диск и т.д. То есть BIOS является промежуточной средой, через которую любая прикладная программа или операционная система осуществляет управление периферийными по отношению к микропроцессору устройствами ПК.

Наличие двухступенчатого обращения к устройствам ПК существенно упрощает работу программистов, поскольку им нет необходимости знать всю совокупность адресов портов или регистров каждого из устройств в данном конкретном исполнении ПК, программисту достаточно знать соответствующие вектора прерываний BIOS. Для всех ПК, работающих в среде микропроцессоров INTEL 8086, 8088, 80286, 80386, 486 и Pentium вектора прерываний BIOS строго затабулированы и полное их описание достаточно широко представлено в специальной литературе. Именно поэтому BIOS рассматривают как одну из составных частей любой операционной системы, где под операционной системой понимают совокупность программ, резидентно загружаемую в память компьютера после его включения.

Кроме POST и BIOS в микросхемах ПЗУ также размещается специальная программа SETUP, которую можно вызвать сразу после включения компьютера удерживанием клавиши DEL или F1 клавиатуры. Данная программа позволяет интерактивно конфигурировать компьютер в зависимости от установленного в нем периферийного оборудования, выполнять процедуры форматирования жесткого диска, установки системных даты и времени и проводить ряд других важных установок. Как правило, пользователи редко обращаются к SETUP, но тем не менее ознакомиться с ее возможностями весьма полезно. Обычно обращения к SETUP осуществляются при смене жесткого диска, поскольку в этом случае, необходимо указать его характеристики.

Внесистемный загрузчик обеспечивает поиск на различных типах запоминающей среды, загрузку в память компьютера и передачу управления Master boot sector – программному коду, который всегда расположен в первом физическом секторе дисковых подсистем. Этот программный код обеспечивает загрузку операционной системы.

2.2. Спецификация Plug&Play

Спецификация Plug&Play (“подключай и работай”) была разработана фирмой Intel, при непосредственном участии фирм IBM и Microsoft, в 1993 году. Она позволяет выполнять автоматическое конфигурирование компьютера при подключении к нему дополнительного периферийного оборудования или его замены на новое. Спецификация Plug&Play специально разработана для распознавания и согласования всех изменений в конфигурации компьютера без вмешательства пользователя. Суть работы спецификации заключается в том, чтобы компьютер мог самостоятельно распознать установленное в нем оборудование и соответствующим образом выделить для него необходимые ресурсы. Пользователю нет необходимости задумываться об адресах ввода/вывода, каналах прямого доступа к памяти, линиях запроса на прерывание и т.п.

Практическая реализация спецификации связана не только со схемотехническими решениями, но, и в равной степени, с программным обеспечением. Это возможно, если в реализации спецификации принимают участие следующие три компонента:

Аппаратные средства, поддерживающие Plug&Play;

BIOS;

Операционная система.

Под аппаратными средствами, поддерживающими спецификацию Plug&Play следует понимать как системные платы, так и платы адаптеров, устанавливаемые в слоты расширения. В BIOS, поддерживающей эту спецификацию, дополнительно включено полтора десятка функций, используемых в дальнейшем операционной системой. В процессе работы аппаратные средства информируют BIOS и операционную систему о том, что это за устройства, и какие ресурсы им необходимы для нормальной работы. Операционная система выполняет загрузку соответствующих драйверов и, при необходимости, позволяет вручную настроить их параметры. Впервые спецификация Plug&Play была включена, как составляющая часть в операционные системы Windows 95 и Windows NT.

Автоматическое конфигурирование системы осуществляется во время выполнения расширенных POST процедур. BIOS идентифицирует устройства, определяет в каких слотах они установлены и, по-возможности, выделяет им ресурсы компьютера. Этот процесс осуществляется в несколько этапов.

Сначала отключаются все самонастраиваемые устройства как на системной плате, так и в адаптерах. Затем осуществляется поиск всех ISA, а далее PCI –устройств типа Plug&Play и создается исходная карта распределения их ресурсов: портов, прерываний, каналов прямого доступа памяти и собственно памяти. Следующий этап связан с активизацией устройств ввода/вывода и конфигурацией устройств начальной загрузки операционной системы. Этим устройствам передается информация о выделенным им ресурсам, что переводит их в рабочее состояние. На заключительном этапе запускается внесистемный загрузчик BIOS и осуществляется загрузка операционной системы.

Основная задача операционной системы, поддерживающей спецификацию Plug&Play – выдать сообщения о конфликтах, которые не были устранены BIOS. Так, например, те устройства, которые BIOS не смогла распознать или самонастроить могут быть сконфигурированы вручную с помощью операционной системы.

2.3. Клавиатура

Клавиатура является устройством оперативного ручного ввода информации в персональный компьютер. Ее основной функцией является преобразование позиционного кода нажатой клавиши в стандартный алфавитно-цифровой или управляющий двоичный код. Этот код позволяет кодировать до 256 символов и строится по весовому принципу, согласно которому двоичный код символа последовательно инкрементируется при переходе от цифры 0 к цифре 9 и при переходе от буквы к букве в алфавитном порядке. Такое построение кода существенно упрощает программирование и анализ наиболее часто встречаемых операций обработки текстов, связанных с процедурами сортировки или поиска.

В настоящее время наибольшее распространение получили коды КОИ-7, КОИ-8 и ASCII, последний из которых принят в качестве американского национального стандарта и используется во всех моделях компьютеров фирмы IBM, а также большинством зарубежных фирм-изготовителей вычислительной техники. Это обусловлено тем, что ASCII-коды обеспечивают байтовое представление 256 символов, включая символы национальных стандартов и символы псевдографики. Символы, кодируемые ASCII-кодами от 0 до 127 (шестнадцатеричные 00H – 7FH), составляют совокупность полного набора текстовой информации, некоторые математические символы и знаки препинания. Кроме того, в данной области находятся коды, соответствующие ряду клавиш электрических пишущих машинок, таких как "пробел", "возврат каретки", "перевод строки", "верхний регистр" и др.

В области кодов от 128 до 255 (шестнадцатеричные 80H – FFH), как правило, закодирован набор символов псевдографики, применение которых полезно при формировании всевозможных меню, запросов и специальных значков с экранными атрибутами, а также набор символов прописных букв латинского алфавита. Именно эта область наиболее часто используется прикладными программами кодировки символов национального алфавита, в частности русского шрифта (кириллицы).

Все компьютеры семейства IBM или совместимые с ними используют либо малогабаритную (84 клавиши), либо усовершенствованную (101 клавиша) клавиатуру. Независимо от типа клавиатуры, в ней содержится полный набор клавиш, используемый в обычной пишущей машинке и дополнительный набор специальных клавиш. Усовершенствованная клавиатура отличается от малогабаритной наличием малой цифровой панели, которая достаточно удобна для быстрого ввода цифровых данных в компьютер благодаря повторению структуры и месторасположения цифровых клавиш бытовых калькуляторов.

Каждая клавиатура располагает тремя основными типами клавиш: символьными, функциональными и управляющими. Символьные или алфавитно-цифровые клавиши предназначены для кодирования букв, цифр, знаков препинания, символов %, \$ и т. д. Данная группа клавиш практически полностью повторяет клавиши обычной пишущей машинки как по количественному составу, так и по их месторасположению. Функциональные клавиши подразделяются на программно-функциональные клавиши (F1 – F12) и клавиши управления курсором и редактирования. Нажатие функциональной клавиши клавиатуры не вызывает формирования какого-либо символа на мониторе, а используется для инициирования запроса на какие-либо действия, заложенные в текущей программе. В этом случае задача программиста заключается в идентификации нажатой клавиши и программного присвоения ей определенных функций. С помощью управляющих клавиш Caps Lock, Num Lock и Scroll Lock осуществляется фиксация режимов работы клавиатуры, а управляющие клавиши Alt, Ctrl и Shift предназначены для совместной работы одновременно с другими клавишами и их практическое применение будет рассмотрено ниже.

Аппаратурная реализация клавиатуры включает в себя совокупность ключей, замыкаемых при нажатии соответствующих клавиш, а также схем управления и кодировки нажатой клавиши. В основе механического действия клавиши лежит либо изменение электрического сопротивления (короткое замыкание), либо изменение емкости, либо возникновение контактной разности потенциалов. В настоящее время наиболее часто используется первый из перечисленных методов. Независимо от конструктивной реализации, данное механическое действие приводит к переходу клавиши из одного логического состояния в другое. На практике такой переход связан либо с достаточно плавными, либо многократно повторяющимися изменениями событий в процессе нажатия или отпускания клавиши, что приводит к появлению эффекта, называемого "дребезгом контактов". Схемы управления обязаны устранять влияние дребезга контактов с целью недопущения формирования произвольной последовательности одинаковых кодов при однократном нажатии клавиши пользователем. В этом заключается вспомогательная функция схем управления.

Основная же функция схем управления связана с установлением факта нажатия клавиши, ее кодировкой и передачей этого кода в последовательном или параллельном формате в компьютер. Простейшим вариантом реализации этой функции может служить клавиатура, схема которой представлена на рис. 2.1. При этом предполагается, что нажатие клавиши приводит к состоянию логического "0". Ее основу составляют элементы И-НЕ, четыре из которых являются весовыми и предназначены для формирования двоичного четырехразрядного кода D0 – D3, а пятый – для формирования стробирующего импульса сопровождения этого кода. Методика подключения цифровых клавиш к весовым логическим элементам заключается в подсоединении клавиши к отдельным входам тех элементов, на выходе которых при нажатии клавиши должен активизироваться уровень логической единицы. Например, клавиша "7", должна быть соединена с входами элементов D0, D1, D2, а клавиша "5" – с входами элементов D0, D2 и т. д. Кроме того, все клавиши клавиатуры объединяются функцией И-НЕ (нижний по схеме элемент) для формирования импульса стробирования. Несмотря на простоту схемотехнической реализации данной клавиатуры, подобные схемы на практике используются крайне редко, так как их применение оправдано при построении клавиатур с небольшим набором клавиш.

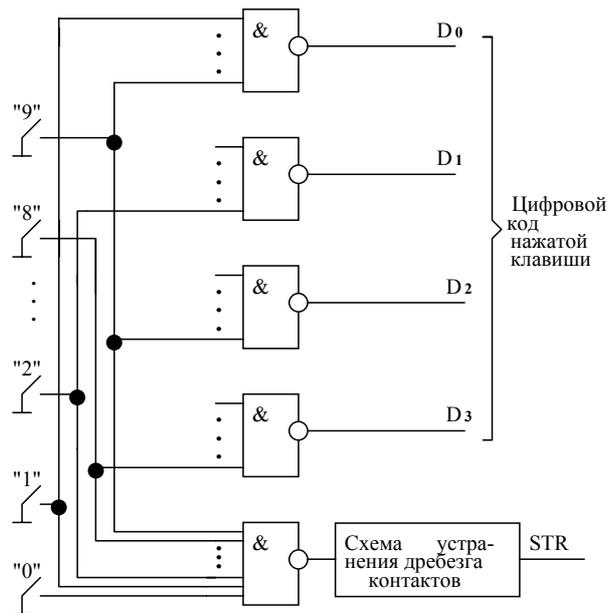


Рис. 2.1. Блок-схема обслуживания цифровых клавиш клавиатуры

В тех случаях, когда необходимо реализовать 84-х или 101-клавишную клавиатуру, предпочтение отдается схемам сканирования клавиш. Рассмотрим работу такой клавиатуры на примере обобщенной блок-схемы, представленной на рис. 2.2.

Генератор тактовых импульсов (ГТИ) работает в непрерывном режиме и непосредственно подключен к тактовому входу двоичного счетчика D1. В результате дешифратор D2 последовательно опрашивает X-столбцы матрицы ключей. Сами клавиши расположены на пересечении X-столбцов и Y-строк. При нажатии любой из клавиш, сигнал с выхода DC через замкнутый контакт поступает на соответствующую горизонтальную шину Y-строки и далее через селектор С на вход ПЗУ. Совокупность сигналов с выходов дешифратора и селектора образуют адресный вход ПЗУ, в ячейках которого, как правило, записаны младшие разряды кода символа нажатой клавиши. В этом случае дополнительные старшие разряды кода символа формируются специальным регистром RG, состояние которого обычно изменяется нажатием управляющих клавиш (например, Alt, Ctrl, или Shift). Большинство клавиатур позволяет формировать до семи комбинаций кодов старших разрядов, соответствующих полному набору ASCII-кодов. Для исключения влияния дребезга контактов выдача кода символа из регистра RG задерживается с помощью одновибратора на время завершения переходного процесса.

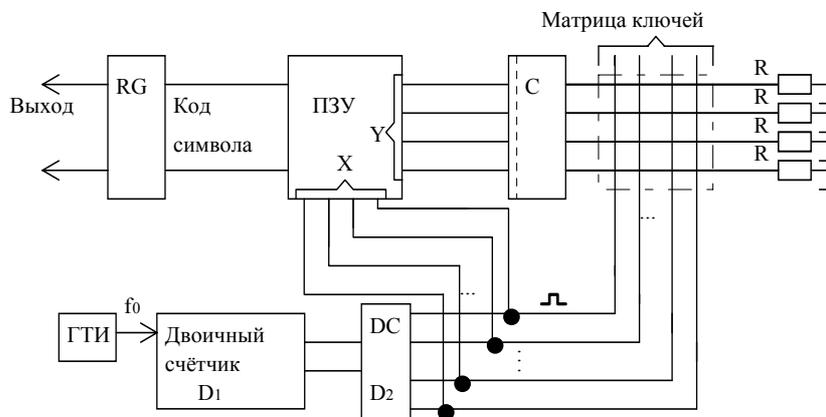


Рис. 2.2. Обобщенная блок-схема сканирующей клавиатуры

В современных клавиатурах, для сканирования клавиши и управления работой клавиатуры в целом, используют либо однокристальную микропроцессорную БИС, либо 8-разрядный микропроцессор. Это позволяет существенно упростить схемотехнику клавиатуры, увеличить ее функциональные возможности, а также усовершенствовать алгоритм устранения влияния дребезга контактов. Для данных схем вертикальные и горизонтальные шины матрицы-клавиатуры подключаются непосредственно к выходным и входным портам ввода/вывода БИС (рис 2.3), которая специально запрограммирована на выполнение тех же функций, что и схема, изображенная на рис. 2.2.

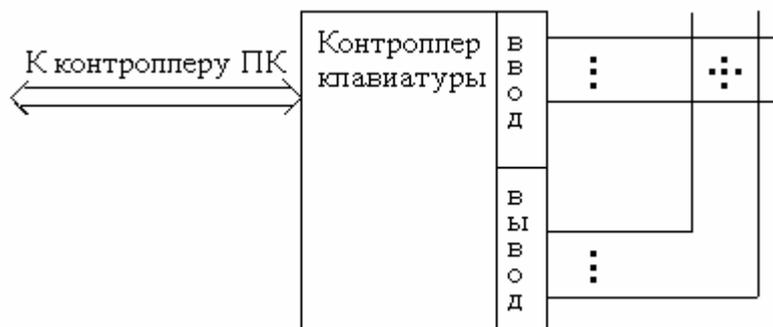


Рис. 2.3. Схема контроллера клавиатуры, построенная на базе однокристальной ЭВМ

Алгоритм работы такой клавиатуры заключается в следующем. В исходном состоянии счетчик повторений контроллера обнуляется, а код X-столбца на линии Вывод устанавливается равным 00...01. Каждое последующее значение кода X формируется путем циклического сдвига логической единицы в сторону старших разрядов. В результате на горизонтальных шинах матрицы устанавливается одна из комбинаций символов, программно считываемая контроллером по линии Ввод. Если эта комбинация содержит все нули, т. е. не нажата ни одна из клавиш, находящихся в вертикальном столбце по активной X линии, то микроконтроллер производит очередной сдвиг кода X и продолжает поиск нажатой клавиши. В результате этого поиска вместе с уже рассмотренной возможны ситуации, когда на линии Ввод устанавливается несколько единиц, т.е. одновременно нажато несколько клавиш на активной X-линии, или когда на линии Ввод присутствует только одна логическая единица. Первая из перечисленных ситуаций воспринимается микроконтроллером как ошибочная, возвращает его в исходное состояние и может вызвать включение звукового сигнала. При второй ситуации по текущему значению кода опроса столбца X и принятому из линии Ввод значению кода строки Y программно формируется весовой код нажатой клавиши, т.е. замкнутого ключа. Для этого коды опроса преобразуются к двоичному или двоично-десятичному виду с помощью программных счетчиков, затем определяется весовое значение нажатой клавиши. Данный алгоритм работы клавиатуры является своего рода стандартом для моделей ПК фирмы IBM и программно реализован в ПЗУ однокристального микроконтроллера типа INTEL 8048. С его помощью осуществляется обработка нажатой клавиши клавиатуры, устраняется дребезг ее контактов, а также выполняются диагностические тесты работоспособности после включения питания системного блока.

При однократном нажатии клавиши, или когда имеет место режим повторения за счет удержания клавиши в нажатом состоянии, результат каждого действия запоминается в 20-символьном буферном регистре клавиатуры. Появление в нем хотя бы одного символа инициирует прерывание в системный блок компьютера для последующей обработки нажатой клавиши. В ответ на прерывание BIOS считывает в последовательном формате состояние буферного регистра и посылает необходимые инструкции обратно в клавиатуру.

2.4. Обслуживание клавиатуры по прерываниям BIOS и DOS

Как отмечалось ранее, клавиатура располагает символьными и функциональными клавишами. Каждая клавиша имеет собственный скэн-код, посредством которого программа может определить нажатие любой клавиши на клавиатуре, независимо от изображенного на ней символа. Например, скэн-код символа "*" над цифрой 8 равен 09H, а на клавише Print Screen – 29H. При запросе ввода одного символа с клавиатуры возможны два типа ответа в зависимости от того, символьная или функциональная клавиша была нажата.

При нажатии символьной клавиши клавиатура посылает в микропроцессор два несовпадающих байта информации; соответственно ASCII и скэн-код нажатой клавиши. При нажатии функциональной клавиши клавиатура также посылает в микропроцессор два байта информации: код 00H и скэн-код нажатой клавиши. Таким образом, равенство нулю первого байта указывает на нажатие функциональной клавиши.

Данный алгоритм взаимодействия между ПК и клавиатурой имел место в первых моделях персональных компьютеров и в настоящее время практически не используются. Это связано с тем, что в ПК достаточно передать скэн-код нажатой клавиши, а обработку этого кода поручить BIOS.

Ввод данных с клавиатуры в компьютер может быть осуществлен как по прерыванию стандартной версии DOS (прерывание INT 21H), так и по прерыванию BIOS (INT 16H). Для ввода, использующего прерывания DOS, перед выполнением команды INT 21H должен быть определен полный список параметров для области ввода. В его составе: указатель начального адреса области ввода в ОЗУ, который заносится в регистр DX микропроцессора, однобайтовый параметр, определяющий максимально возможную длину вводимого текста, а также помещенный в регистр AH номер выполняемой функции (в данном случае 0AH).

Ввод текста осуществляется выполнением команды DOS INT 21H. В этом случае прерывания DOS ожидает, пока пользователь вводит с клавиатуры текст, проверяя при этом, чтобы количество введенных символов не превышало однобайтовый указатель максимально возможной длины. Если набран слишком длинный текст, то DOS предупреждает пользователя включением звукового сигнала, и система ожидает только ввода кода ENTER. Таким образом, символы, превышающие максимальную длину, игнорируются. Для указания окончания ввода пользователь нажимает клавишу ENTER, код 0DH которой также заносится в область ввода ОЗУ. Если, например, пользователь ввел символьную последовательность MILTON, предварительно указав адрес (DX) и максимальную длину в 10 символов, то список параметров и текст в области ввода будет выглядеть следующим образом.

Адрес памяти	(DX)	(DX)+1	(DX)+2	(DX)+3	(DX)+4	(DX)+5	(DX)+6	(DX)+7	(DX)+8	(DX)+9
Информация	0A	08	4D	49	4C	54	4F	4E	20	0D
Последовательность			M	I	L	T	O	N		←

Первый адрес области ввода содержит информацию о максимально установленной длине символов (0AH), следующий за ним (07H) – количество реально введенных символов за исключением кода 0DH – ENTER, а далее непосредственно занесены коды символьной последовательности. Поскольку установленная максимальная длина вводимых символов включает код 0DH, то действительная максимальная длина текста на единицу меньше.

Введенную символьную последовательность можно использовать для различных целей, например, для печати сообщений, вывода текста на монитор, сохранения в таблице, записи на диск и т. п. При этом, часто возникает необходимость замены кода символа ENTER на другой код. Это достаточно легко сделать, используя содержимое ячейки (DX)+1, поскольку код 0DH всегда находится по адресу, равному адресу этой ячейки плюс ее содержимое плюс единица. Определив адрес кода 0DH, его заменяют на необходимый.

При последующем вводе информации вводимые символы заменяют предыдущее содержимое области ввода на новое и остаются там, пока другие символы не заменят их. В связи с этим часто возникает необходимость в ее принудительной очистке до ввода следующей символьной последовательности. Данную процедуру можно выполнить двумя способами: либо

принудительным заполнением области ввода кодом символа пробел (20H) до очередного ввода текста, либо после введенного текста, но только ту ее часть, которая следует за кодом 0DH.

Расширенные версии DOS, т. е. версии DOS 3.3 и выше, позволяют существенно увеличить возможности обслуживания клавиатуры. Расширенные возможности включают файловый номер, который заносится в регистр BX микропроцессора и для ввода с клавиатуры имеет значение BX = 00H. Кроме того, изменяется функция прямого ввода в регистре AH с 0AH на 3FH, а максимальное количество вводимых символов заносится в регистр CX.

В расширенной версии DOS, команда INT ожидает, пока пользователь не введет символы с клавиатуры, но не проверяет, превышает ли число введенных символов значение регистра CX. Нажатие клавиши ENTER указывает на завершение ввода. В этом случае после введенного текста, т. е. после кодов текста, следует код символа возврата каретки 0DH и код символа возврата строки 0AH. Вследствие этой особенности расширенной версии DOS, максимальное количество символов и размер области ввода должны предусматривать место для двух дополнительных символов 0DH и 0AH.

Если вводить текст, который превышает содержимое регистра CX, то будут приниматься все без исключения символы, а фактическая длина введенной последовательности запишется в регистр AX. Однако в область ввода запишутся только первые CX кодов символов. В результате следующая команда INT, будет принимать данные не с клавиатуры, а из собственного внутреннего буфера, поскольку в нем еще сохранились предыдущие данные. Таким образом, содержимое регистра AX позволяет определить количество введенных с клавиатуры символов и установить факт ввода законченного текста.

Прерывание INT 21H DOS опирается на команду BIOS INT 16H, которая выполняет одну из трех основных функций ввода с клавиатуры. Номер функции ввода определяется содержимым регистра AH микропроцессора в момент обращения к BIOS.

1. Чтение символа: AH = 00H. При выполнении данной функции ASCII-код нажатой символьной клавиши помещается в регистр AL, а ее скэн-код в регистр AH. Если на клавиатуре нажата одна из функциональных клавиш, то в регистр AL заносится код 00H. Вывод символа на экран монитора в этом случае не происходит.

2. Определение наличия введенного символа: AH = 01H. Данная функция определяет наличие символа для чтения с клавиатуры, т. е. очередной код символа и скэн-код помещен в регистр AX.

3. Определение текущего состояния клавиатуры: AH = 02H. При выполнении этой функции BIOS возвращает в регистр AH текущее состояние клавиатуры из ячейки памяти 417H. В этом случае биты регистра состояния AH имеют следующее функциональное назначение:

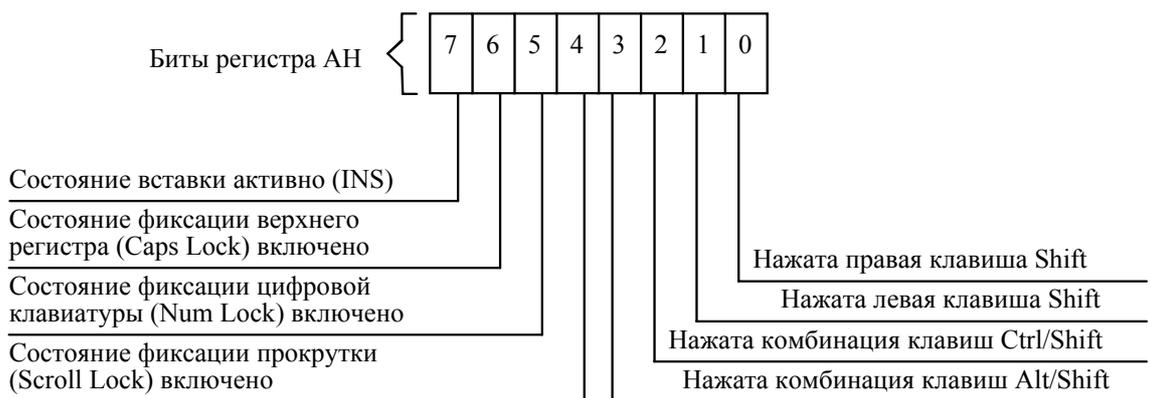


Рис. 2.4. Функциональное назначение битов регистра состояния AH

Вернемся к расширенной версии DOS по прерыванию INT 21H, для функции 0AH. В данной версии выполнение этой функции имеет более мощные средства по отношению к базовой версии DOS и их выбор для использования в прикладных программах является более предпочтительным. В этом случае возможна дополнительная реализация одной из следующих функций, в зависимости от содержимого регистра AH:

1. Ввод с клавиатуры с эхоотображением: АН = 01Н. Данная функция выводит на экран символ, ASCII-код которого находится в регистре AL, после нажатия символической клавиши клавиатуры. Если была нажата функциональная клавиша, то значение регистра AL = 00 и соответственно отсутствует эхо на экран. Для выделения скэн-кода необходимо повторить вызов данной функции. Функция реагирует на запрос Ctrl/Break.

2. Прямой ввод с клавиатуры: АН = 07Н. Выполнение для функции аналогично для АН = 01Н, с тем отличием, что введенный символ не отражается на экране и функция не реагирует на запрос Ctrl/Break.

3. Ввод с клавиатуры: АН = 08Н. Данная функция имеет только одно отличие от функции АН = 01Н, которое заключается в отсутствии эха на экран.

4. Проверка состояния клавиатуры: АН = 0ВН. Данная функция записывает код FF в регистр AL, если ввод с клавиатуры возможен или 00 в противном случае.

Следует отметить, что выполнение рассмотренных выше функций напрямую связано с проверкой состояния клавиатуры, поскольку функции 01Н, 07Н и 08Н не ожидают ввода с клавиатуры.

2.5. Манипуляторы курсора

К манипуляторным устройствам оперативного ввода информации в компьютер можно отнести мыши, трекболы и дигитайзеры. Они, как правило, подключаются к последовательному порту, с помощью кабеля, и при передвижении манипулятора по столу (или планшету) курсор синхронно перемещается по экрану. Однако существуют и беспроводные манипуляторы курсора, использующие инфракрасное излучение для передачи кодов в приемник, подключенный к последовательному порту на расстоянии 1,5-2 м. Связь манипуляторов с ПК осуществляется с помощью драйвера устройства, который не входит ни в состав BIOS, ни в состав DOS. Как правило, драйверы поставляются вместе с манипулятором и описанием к нему.

Среди манипуляторов типа мышь наибольшее распространение получили мыши таких фирм, как Microsoft, Logitech, Genius и трехкнопочная мышь Mouse System.

Мыши от Microsoft и им подобные имеют последовательный интерфейс и используют для передачи процессору трехбайтовый формат, содержащий информацию о перемещении мышки (два байта) и о состоянии кнопок. Все подобные мыши передают данные со скоростью 1200 бит/с и используют 7 бит данных без контроля четности и один стоповый бит. Некоторые мыши фирмы Logitech работают со скоростью 2400 бит/с. Трехкнопочная мышь Mouse System и совместимые с ней передают данные в пятибайтовом формате.

Существуют также так называемые шинные мыши (bus-mouse), использующие вместо стандартного последовательного порта RS-232C собственную интерфейсную карту, вставляемую в разъем расширения. Эти мыши в настоящее время применяются редко и их использование оправдано, если последовательные порты заняты.

Принцип работы манипулятора мышь заключается в следующем: механическое перемещение по поверхности стола и нажатия кнопок мыши преобразуются в двоичную информацию, передаваемую в ПК, где программа управления мышью (драйвер мыши) преобразует наборы кодов в конкретные действия.

Первая мышь появилась в 1963 году. Она перемещалась с помощью двух роликов, которые были связаны с осями переменных резисторов. Перемещение ее было прямо пропорционально изменению сопротивления переменных резисторов. В дальнейшем конструкция мыши претерпела ряд изменений: соприкосновение с поверхностью стало осуществляться с помощью резинового шарика, ролики были перенесены внутрь корпуса и установлены на оси с датчиками, с помощью которых определялось направление (X или Y) и скорость перемещения мыши.

В качестве датчиков перемещения использовались непроводящие диски с нанесенными печатным способом контактами, которые могли соприкоснуться с одним из неподвижных контактов. Движения роликов преобразовывались в электронные импульсы, которые после обработки электроникой самой мыши передавались в ПК.

Как известно, полностью механическая мышь имеет существенный недостаток, она крайне ненадежна. Поэтому впоследствии стали использовать оптико-механический принцип кодирования перемещения, т.е. смену механическому шифратору пришли оптопары типа светодиод-фотодиод, светодиод-фоторезистор или светодиод-фототранзистор. Порядок, в котором освещаются фоточувствительные элементы, определяют направление перемещения мыши, а частота проходящих от них импульсов – скорость. Перемещаться мышь в данном случае может и без помощи шарика, как, например, Honeywell. Уникальность этой мыши состоит в том, что вместо шарика она имеет две "ножки", являющиеся частью X-Y-механизма оптико-механического шифратора.

Существуют и полностью оптические мыши. Такие устройства работают только на специальном планшете (коврике). Поверхность планшета покрыта сеткой перпендикулярных линий, нанесенных на отражающую свет поверхность. Один из двух светодиодов испускает красные лучи, другой – инфракрасные. Эти лучи поглощаются вертикальными или горизонтальными линиями, а поверхность между линиями отражает их. Фотоэлементы улавливают отраженные лучи. По мере перемещения мыши по планшету фотоэлементы поочередно обнаруживают или не обнаруживают соответствующие им источники света. Следовательно, сигнал из чувствительного к красному фотоэлемента представляет перемещение в направлении оси X, а сигнал из чувствительного к инфракрасному фотоэлемента – в направлении оси Y. Эти сигналы передаются в ПК, где драйвер мыши преобразует их в движение курсора на экране. К недостаткам оптической мыши можно отнести ее высокую стоимость и сложное устройство. Но наряду с недостатками такая мышь обладает и существенными достоинствами, такими как отсутствие движущихся частей и более точное управление курсором.

В последнее время появились манипуляторы курсора, позволяющие эмулировать шесть степеней свободы. Они применяются при работе с трехмерной графикой и внешне выглядят как гибрид мыши с джойстиком.

Строго говоря, джойстик не является манипулятором курсора, но это устройство для ввода информации в компьютер, поэтому целесообразно рассмотреть его принцип действия в этом параграфе.

Современные модели джойстиков являются аналоговыми координатными устройствами. Технически джойстик это два реостатных датчика, для питания которых используется напряжение +5 В. Рукоятка джойстика связана с двумя переменными резисторами, изменяющими свое сопротивление при ее перемещении. Один резистор определяет перемещение по координате X, а другой – по Y. В задачу адаптера джойстика входит преобразование изменения параметра сопротивления в соответствующий цифровой код.

Для описания трекбола (trackball) можно провести аналогию с механической мышью. Грубо говоря, трекбол представляет собой перевернутую вверх шариком мышь. Перемещение курсора осуществляется за счет вращения шарика рукой или пальцем (для мини-трекболов, встроенных в портативные компьютеры).

Хотя дигитайзер, как и мышь, является последовательным устройством, но принцип его работы отличается от мыши. При движении мыши по столу или коврику курсор перемещается вслед за мышью, а при помощи дигитайзера указывается конкретная позиция курсора на экране.

Дигитайзеры предпочтительнее, чем мыши для программ с большим количеством команд и для графических режимов, когда нужно получить точную копию.

Дигитайзер состоит из двух частей: планшета и наводчика (или пера).

Планшет представляет собой плоский прямоугольник, являющийся областью координат (область позиционирования). Планшет подключается к последовательному порту и чаще всего имеет отдельный от него блок питания. Верхняя поверхность планшета обычно состоит из активной области и рамки. Если указать позицию пером или наводчиком в активной области, то с помощью электронных схем эта позиция фиксируется и передается в ПК. В активную область помещается шаблон, сделанный из картона или пластика. На нем печатается набор команд и соответствующая область экрана компьютера, в зависимости от используемой прикладной программы.

Перо – устройство управления позицией курсора подключается к планшету. Перо похоже на шариковую авторучку с кнопками. Перо дигитайзера имеет на конце переключатель. Если выбрана точка на планшете, то касание наконечником пера этой точки приводит к активизации

переключателя. Для многих прикладных программ активизация переключателя идентична нажатию клавиши Enter на клавиатуре. На корпусе пера также имеются переключатели. Корпусные переключатели обычно являются программируемыми (как правило, для прерывания выполнения или отмены команды).

Наводчик, как и перо, является устройством управления позицией курсора и тоже подключается к планшету. Он напоминает по форме мышшь, представляя собой небольшую коробку с набором кнопок. Но в отличие от мыши имеет перекрестную мишень для указания выбранной позиции на планшете.

Существуют три вида дигитайзеров:

1. *Электромагнитные дигитайзеры.* У этого типа дигитайзеров активная область планшета содержит проволочную сетку. Перо имеет проволочную катушку в корпусе, а наводчик - вокруг мишени. Многие электромагнитные дигитайзеры используют наводчик в качестве передатчика, а планшет – как антенну или приемник. При нажатии соответствующей кнопки короткий скачок электронного импульса проходит по катушке. Эти импульсы передаются к сетке на поверхности планшета. Электронные элементы планшета сканируют сетку на обнаружение импульсов. Самый сильный сигнал находится на горизонтальных и вертикальных линиях, ближайших к выбранной точке. И так устанавливается позиция точки. Затем позиция уточняется с помощью временной задержки между переданными и принятыми сигналами. Другой тип электромагнитных дигитайзеров использует сетку как приемник. Но в этом случае помехи и шум оказывают слишком большое влияние.

2. *Дигитайзеры резистивного типа.* Планшет содержит тонкий слой пленки. При выборе точки в активной области, планшет определяет ее координаты путем изменения напряжения. Каждая сторона планшета периодически испускает электронный импульс, а сигнал, индуцируемый в перо, изменяется. При приближении пера к источнику, сигнал возрастает. Электронная схема выполняет коррекцию ошибок и использует 4 значения для определения позиции курсора.

3. *Акустические дигитайзеры.* Эти дигитайзеры не имеют планшета. Их принцип действия основан на явлении эхо-локации. При выборе точки световое перо издает щелчок, а пары приемников регистрируют эти щелчки. Для каждого пространственного измерения предназначена своя пара приемников. Расстояние между приемниками известно, а щелчок достигает их с небольшой временной разницей. На основании триангуляции и значения скорости звука можно вычислить пространственные координаты светового пера для каждого измерения с достаточно большой точностью.

Первый тип дигитайзеров довольно широко распространен, а два последних не получили такого широкого распространения как первый.

2.6. Классификация и принцип работы видеосистем ПК

Видеосистемы являются устройствами оперативной визуальной связи пользователя с персональным компьютером и служат для отображения команд, инструкций и другой текстовой или графической информации, передаваемой от клавиатуры или процессора на монитор. Ее основу составляют специализированные схемы для генерирования электрических сигналов, управляющих монитором. Видеосистемы состоят из двух частей: видеоадаптера и монитора. В стандартной конфигурации монитор (дисплей) подключается 9-ти или 15-ти жильным кабелем к адаптеру, т.е. к печатной плате, которая устанавливается в один из слотов на системной плате. В состав адаптера входят ПЗУ знакогенератора, программируемый контроллер дисплея, видеобuffer и ряд вспомогательных схем. Схемы адаптера формируют необходимую совокупность сигналов, с помощью которых осуществляется управление выводимой на дисплей информацией.

В подавляющем большинстве ПК в качестве средства отображения информации выступают дисплеи, созданные на базе электронно-лучевой трубки (ЭЛТ). Поэтому большинство программируемых адаптеров (контроллеров) – это контроллеры ЭЛТ (CRTC). Их основная функция заключается в преобразовании данных из буфера в те сигналы, с помощью которых

собственно и формируется изображение. Videобуфер представляет собой область ОЗУ, предназначенную только для хранения информации, выводимой на экран.

Обобщенная блок-схема видеосистемы представлена на рис. 2.5.

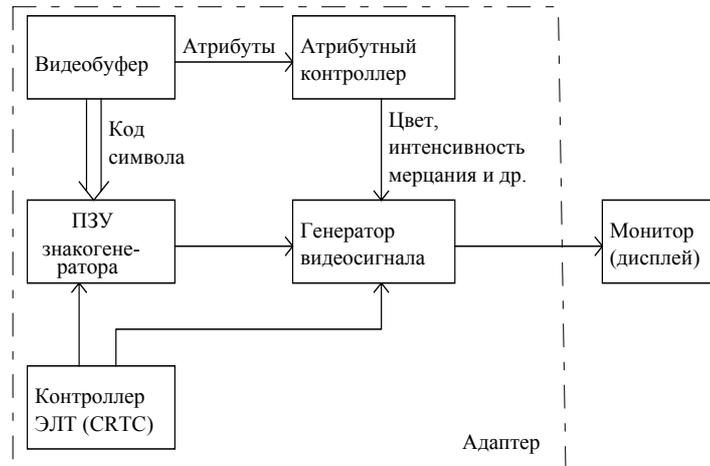


Рис. 2.5. Обобщенная блок-схема видеосистемы

По принципу формирования изображения на экране видеосистемы подразделяются на растровые и векторные. К векторным относятся те из них, в которых формирование символа или элемента графики на экране осуществляется с помощью светящихся отрезков (векторов). Данный способ напоминает формирование рисунка на планшете двухкоординатного самописца или плоттера, если вместо пера использовать электронный луч, а вместо бумаги – люминофор ЭЛТ. Несмотря на наглядность и идеологическую простоту формирования изображения векторным способом, его схемотехническая реализация достаточно сложна и требует прецизионных аналоговых схем. Еще более сложная задача реализовать приемлемое по четкости изображение в цвете. Данные обстоятельства явились основной причиной весьма ограниченного применения подобных видеосистем в вычислительной технике.

В растровом дисплее текст или рисунок формируется электронным лучом в процессе периодического сканирования экрана построчно слева направо и сверху вниз. В результате этого сканирования весь экран или растр оказывается разбитым на стандартное количество близко расположенных горизонтальных линий развертки. По мере движения луча по растру, схема управления лучом изменяет его яркость, образуя светящиеся точки или пиксели. Совокупность светящихся пикселей собственно и формирует конфигурацию символа в пределах знакоместа, а совокупность знакомест – в пределах всего экрана. Здесь и далее под знакоместом понимается область экрана размером $A \cdot B$ пикселей, с помощью которых можно изобразить любой символ в текстовом режиме работы видеосистемы.

Для получения немерцающего изображения необходимо повторить формирование растра электронным лучом 50 – 70 раз/с. В каждом цикле сканирования луч движется по зигзагообразной траектории, с левого края верхней горизонтальной строки растра к правому краю нижней горизонтальной строки. По окончании движения в пределах строки луч быстро отклоняется вниз и влево на начало новой строки и т.д. Когда луч сформирует последнюю строку растра, он возвратится в верхний левый угол экрана и процесс сканирования возобновляется.

В качестве мониторов в современных ПК наибольшее распространение получили мониторы на основе ЭЛТ. Однако такие мониторы громоздки, потребляют достаточно много энергии и не безвредны для зрения. Поэтому многими фирмами ведутся разработки, внедрение и усовершенствование новых типов мониторов. Так, наибольшее распространение в портативных компьютерах notebook получили монохроматические и цветные жидкокристаллические LCD-дисплеи. Действие таких LCD-мониторов основано на эффекте возникновения сильного поглощения света жидкими кристаллами при подаче к ним разности потенциалов. Если всю плоскость экрана изготовить в виде матрицы жидких кристаллов (пикселей), то в результате создается видеосистема с плоским экраном. LCD-мониторы не создают вредного для здоровья

излучения, портативны и наиболее экономичны в потреблении энергии. Однако для управления пикселями необходимы специальные видеоадаптеры и такие мониторы обладают большой инерционностью. Тем не менее, технология LCD-дисплеев быстро прогрессирует и их возможности еще не до конца исчерпаны.

Важнейшим и наиболее перспективным достижением в области разработки плоских мониторов сегодня является цветной TFT-дисплей или активная матрица. Активно-матричные тонкопленочные транзисторные дисплеи принципиально отличаются от обычных LCD-дисплеев, использующих пассивно-матричную технологию.

Каждый пиксель TFT-дисплея содержит отдельный транзистор, управляющий группой из трех цветных точек. Это так "называемый логический пиксель", состоящий из трех жидкокристаллических элементов, видимых сквозь три основных цветовых фильтра – красный, синий и зеленый. Все пиксели подсвечиваются флуоресцентным светом. В выключенном состоянии через жидкокристаллический элемент из-за поляризующих фильтров свет не проходит. Но при подачи напряжения на жидкокристаллический элемент, плоскость поляризации света поворачивается на 90° и свет становится видимым. Таким образом, комбинируя пропускание света через красный, синий и зеленый фильтры, можно в каждом логическом пикселе создать практически любой оттенок с высокой яркостью и насыщенностью цветов и с чрезвычайно высокой контрастностью.

Картинка на экране TFT-дисплея обновляется 80 раз/с. Для сравнения, в обычном VGA-мониторе с кинескопом обновление происходит 70 раз/с, а в пассивно-матричных LCD-дисплеях картинка обновляется примерно 10 раз/с. Именно поэтому применение мышки совместно с обычным LCD-дисплеем оказывается затруднительным. А TFT-мониторы позволяют достичь качества изображения выше, чем на ЭЛТ.

Число управляющих транзисторов в активно-матричном VGA-мониторе с диагональю 10,4" приближается к одному миллиону. Поэтому цена TFT-монитора существенно выше. Так что основной его недостаток - их дороговизна. Однако, также как и в случае с LCD-мониторами, совершенствование технологии изготовления снижает их стоимость.

Пересмотреть отношение к технологии изготовления мониторов на основе "светоизлучающего пластика" (Light Emission Plastics, или LEP), разработанной компанией Cambridge Display Technology (CDT), заставляет заявление компаний Seiko-Epson о начале совместной программы разработки LEP-дисплеев. В течение последних 30 лет внимание многих ученых было приковано к полимерным материалам (пластикам), обладающим свойствами проводников и/или полупроводников. Компания CDT одна из первых обнаружила такие материалы, которые обладают рядом преимуществ по сравнению с традиционными. Главными из них являются: простота и дешевизна производства, а также возможность синтеза новых материалов с заданными свойствами. Главными недостатками - непродолжительный срок службы и низкая мобильность зарядов вследствие аморфной структуры пластика. Однако, в последнее время недостатки постепенно удается преодолеть, в частности, за счет применения многослойных материалов.

Наиболее интересным применением пластиковых полупроводников на данный момент является создание разного рода устройств отображения информации на их базе. О том, что полупроводящий пластик под действием электрического тока может испускать фотоны известно давно. Однако крайне низкая (порядка 0.01%) квантовая эффективность этого процесса делала практическое применение пластика практически невозможным. Тем не менее, за последние 5 лет совершен прорыв в данном направлении, что позволило довести квантовую эффективность двухслойного пластика до 5%, правда, при излучении желтого света. А это уже сравнимо с эффективностью современных неорганических светодиодов (LED). Кроме того, удалось несколько расширить спектр излучения, а именно, в диапазоне от синего до ближнего инфракрасного света пластик может излучать с эффективностью порядка 1%.

Естественным применением столь перспективных материалов – их использование для отображения информации в видеосистемах компьютеров. По крайней мере, на сегодняшний день компания CDT может представить монохромные (желтого свечения) LEP-мониторы, приближающиеся по эффективности к жидкокристаллическим LCD –мониторам. К сожалению, LEP-мониторы уступают им по сроку службы, но имеют ряд существенных преимуществ, а именно:

- Многие стадии процесса изготовления LEP- мониторов технологически совпадают с аналогичными стадиями производства LCD, следовательно, существующее производство легко адаптировать к массовому выпуску LEP структур. Кроме того, технология LEP позволяет наносить пластик на гибкую подложку большой площади, что невозможно для жидких кристаллов
- Поскольку пластик сам излучает свет, то не нужна внешняя подсветка для получения, в том числе цветного изображения на LCD-мониторе. При этом LEP-монитор обеспечивает почти 180-градусный угол обзора.
- Поскольку устройство дисплея предельно просто: совокупность вертикальных электродов с одной стороны пластика, и горизонтальных - с другой, то изменением количества электродов на единицу длины по горизонтали или вертикали можно легко добиваться любого необходимого разрешения, а, при необходимости, и различной формы пикселей.
- LEP-дисплеи работают при достаточно низком напряжении питания (менее 3V) и имеет малый вес, следовательно, он прекрасно подходит для использования в портативных компьютерах, питающихся от батарей. Если при этом учесть малое время переключения (порядка 1мкс), то у LEP-технологии достаточно радужные перспективы.

При работе в составе персонального компьютера, на качество изображения влияет не только тип используемого монитора, но и потребительские свойства видеоадаптера. Совокупность этих свойств определяет стандарт или тип видеосистемы. Наиболее распространенными видеосистемами, на которые рассчитана основная часть программного обеспечения, являются следующие: MDA - монохроматический дисплейный адаптер, CGA - цветной графический адаптер, EGA - усовершенствованный графический адаптер, MCGA - многоцветный графический адаптер, VGA и SVGA. Основные (краткие) характеристики перечисленных видеосистем представлены в таблице 2.3.

Наиболее распространенными видеосистемами, на которые рассчитана основная часть программного обеспечения являются MDA – монохроматический дисплейный адаптер, CGA – цветной графический адаптер, EGA – усовершенствованный графический адаптер, MCGA – многоцветный графический адаптер, VGA и SVGA. Основные характеристики перечисленных адаптеров представлены в таблице 2.3.

Таблица 2.3. Основные характеристики адаптеров MDA, CGA, EGA, MCGA, VGA и SVGA.

Тип адаптера	Текстовый режим			Графический режим	Количество цветов
	Количество строк	Количество символов	Размер символа	Размер растра	
MDA	25	80	9 * 14	Нет	Нет
CGA	25	80	8 * 8	640 * 320	16
EGA	—	—	8 * 14	640 * 350	16 (64)
MCGA	—	—	8 * 16	640 * 480	256 (256К)
VGA	—	—	9 * 16	640 * 480	256 (256К)
SVGA	60	132	9 * 16	1188 * 480 1024 * 768	256 (256К)

Представленное в четвертом столбце таблицы значение А*В указывает на то, что горизонтальный размер знакоместа на экране монитора состоит из А пикселей, а для его полного

формирования требуется В строк растра. Пятый столбец таблицы определяет количество пикселей в строке, умноженном на количество формируемых строк при работе адаптера в графическом режиме. В адаптерах MDA, CGA и EGA применяется ТТЛ-кодирование всех сигналов, которые поступают на монитор через 9-контактный D-разъем. Адаптеры MCGA, VGA и SVGA рассчитаны на формирование аналоговых сигналов и требуют применения специальных аналоговых мониторов. Для стыковки адаптера с монитором в этом случае применяется 15-контактный разъем D-типа. Кроме того, адаптеры VGA и SVGA могут работать в режимах, имеющих 350, 400 или 480 строк развертки, что позволяет обеспечить его логическую совместимость с адаптерами MDA, EGA и CGA.

Наибольшее распространение в современных ПК получили видеосистемы VGA и SVGA. При этом принято относить к видеосистемам SVGA любую систему, поддерживающую все режимы VGA, но имеющую собственные, улучшенные режимы. Например, некоторыми отличительными особенностями видеосистем SVGA принято считать следующие: использование 16- и 256-цветных режимов с разрешением 1024*786 пикселей; применение новых цифро-аналоговых преобразователей RAMDAC (8 разрядов на цвет вместо 6); от 512 Кбайт до 4 Мбайт видеобuffers; поддержка аппаратного курсора; 16-разрядная шина данных (у VGA обычно 8); 16-разрядный видеоBIOS (у VGA обычно 8) и т.д.

Первой разработкой SVGA-видеосистемы можно считать разработку фирмы NEC, предложившую VGA-совместимый адаптер, который имел дополнительный режим с разрешением 800*600 пикселей при одновременном отображении до 256 из 256К возможных цветов. В последнее время все производители видеосистем, вносили свои, дополнительные режимы с улучшенной разрешающей способностью и большим количеством воспроизводимых цветов. При этом приведенный в таблице 2.4 стандарт VGA новыми видеосистемами полностью поддерживался.

Таблица 2.4. Моды стандарта VGA

Мода	Тип режима	Число отображаемых цветов	Формат	Начальный адрес видеобuffers	Разрешение	Размер экрана
0	текстовый	16/256К	40*25	B8000	8*8	320*200
0*	текстовый	16/256К	40*25	B8000	8*14	320*350
0**	текстовый	16/256К	40*25	B8000	9*16	360*400
1	текстовый	16/256К	40*25	B8000	8*8	320*200
1*	текстовый	16/256К	40*25	B8000	8*14	320*350
1**	текстовый	16/256К	40*25	B8000	9*16	360*400
2	текстовый	16/256К	80*25	B8000	8*8	640*200
2*	текстовый	16/256К	80*25	B8000	8*14	640*350
2**	текстовый	16/256К	80*25	B8000	9*16	720*400
3	текстовый	16/256К	80*25	B8000	8*8	640*200
3*	текстовый	16/256К	80*25	B8000	8*14	640*350
3**	текстовый	16/256К	80*25	B8000	9*16	720*400
4	графический	4	40*25	B8000	8*8	320*200
5	графический	4	40*25	B8000	8*8	320*200
6	графический	2/256К	80*25	B8000	8*8	640*200
7	текстовый	монохром	80*25	B0000	9*14	720*350
7*	текстовый	монохром	80*25	B0000	9*16	720*400
D	графический	16/256К	40*25	A0000	8*8	320*200
E	графический	16/256К	80*25	A0000	8*8	640*200
F	графический	монохром	80*25	A0000	8*14	640*350
10	графический	16/256К	80*25	A0000	8*14	640*350
11	графический	2/256К	80*30	A0000	8*16	640*480
12	графический	16/256К	80*30	A0000	8*16	640*480

13	графический	256/256К	40*25	A0000	8*8	320*200
----	-------------	----------	-------	-------	-----	---------

* – относится к текстовому EGA-режиму при 350 линиях вертикальной развертки;

** – относится к текстовому VGA-режиму при 400 линиях вертикальной развертки.

Чтобы устранить возможную несовместимость SVGA-видеосистем различных фирм в октябре 1989 года ассоциация VESA предложила свой стандарт на новые видеосистемы – VESA BIOS Extension, который в настоящее время поддерживается большинством производителей. Сначала VESA рекомендовала использовать режим с разрешением 800*600 пикселей и поддержкой 16 цветов как стандартный. Затем последовали 256-цветные режимы с разрешением 640*480, 800*600 и 1024*786 пикселей, а также 16-цветный режим с разрешением 1024*786 пикселей и т.д. Сегодня спецификация VESA включает в себя режимы вплоть до 1280*1024 пикселей с поддержкой 16,7 миллиона цветов.

Любая видеосистема требует определенный объем памяти для хранения выводимой на монитор информации. Эта память или видеобуфер в общем случае допускает операции записи/считывания и может быть расположена как в адаптере компьютера, так и в отдельном от него устройстве – видеотерминале. Если буфер находится в составе видеотерминала, то для вывода информации на экран необходимо выполнить последовательную передачу данных из компьютера в терминал. Терминал способен воспроизводить новые данные со скоростью около 1000 байт/с, что существенно ограничивает скорость обновления экрана. Кроме того, односторонняя передача данных не позволяет компьютеру отслеживать состояние видеобуфера и, следовательно, выполнять операции выборочного обновления содержимого ячеек памяти.

Отмеченные недостатки устраняются в видеосистемах, в которых видеобуфер является составной частью адресного пространства ОЗУ центрального процессора и расположен непосредственно в адаптере. Такое простое изменение местоположения видеобуфера существенно повышает быстродействие в обновлении экрана и дает возможности процессору "просматривать" экран монитора. В этом случае операция записи в видеобуфер эквивалентна операции "записи на экран", а операция чтения из видеобуфера – операции "считывания с экрана". Процессор способен обновлять содержимое экрана со скоростью нескольких тысяч байт в секунду, что, с учетом частоты регенерации раstra 50 – 70 Гц, обеспечивает практически незаметное изменение изображения на мониторе.

Расположение видеобуфера в адресном пространстве ПК требует его более сложной схемотехнической реализации, поскольку он должен быть реализован как двухпортовая память и содержать дополнительные схемы для синхронизации обращений процессора с операциями адаптера. Поэтому в нескольких разработках видеосистем для построения видеобуфера применяются специальные микросхемы памяти (VRAM). Они имеют два канала, один из которых допускает операции только считывания и используется адаптером для регенерации изображения синхронного с разверткой раstra. Второй канал допускает операции считывания и записи и предназначен для связи с центральным процессором. Оба канала действуют параллельно и независимо, что исключает необходимость синхронизации операций адаптера и процессора.

В текстовом режиме вместе с кодами символов в видеобуфере хранится информация, указывающая как должно формироваться изображение символа. Эта информация называется атрибутом символа и служит для организации таких режимов, как повышенная яркость, мерцание, подчеркивание и т.п., а также определяет цвет символа и окружающего его фона.

Емкость видеобуфера в различных системах варьируется от 4 Кбайт (MDA) до 4 Мбайт (SVGA). Очевидно, что увеличение объема памяти позволяет формировать на экране монитора все более и более сложные изображения. Кроме того, увеличение емкости видеобуфера позволяет хранить больше данных, чем требуется для заполнения экрана, т.е. организовать страничный режим вывода информации на экран. Данное обстоятельство позволяет существенно расширить возможности адаптера и реализовать множество интересных и удобных приемов управления выводом на монитор.

В текстовом режиме для преобразования записанных в видеобуфер кодов символов в их точечное изображение на экране монитора используют знакогенератор. Он представляет собой ПЗУ, в котором записано построчное разложение конфигурации символа в пределах одного знакоместа. Данная информация носит название символьной матрицы и представлена в знакогенераторе в двоичном формате. Знакогенератор воспринимает 7-битный ASCII-код символа

и двоичный код номера строки символьной матрицы и формирует на выходе соответствующий номеру строки параллельный код отображаемого символа. Так, например, для формирования символа "0" на экране монитора при размерах символьной матрицы 5*7, информация в ПЗУ должна быть представлена следующим образом:

	ПЗУ					Код строки
Строка 1	0	1	1	1	0	0 0 0
Строка 2	1	0	0	0	1	0 0 1
Строка 3	1	0	0	0	1	0 1 0
Строка 4	1	0	0	0	1	0 1 1
Строка 5	1	0	0	0	1	1 0 0
Строка 6	1	0	0	0	1	1 0 1
Строка 7	0	1	1	1	0	1 1 1

При поступлении на вход знакогенератора ASCII-кода символа "0", в ПЗУ выбирается соответствующая область расположения символьной матрицы. Дальнейшее изменение входной информации будет связано с перебором кодов строк синхронно с увеличением номера строки развертки раstra. В результате для нашего примера при коде строки 000, на выходе знакогенератора появится двоичный код 01110, а при коде строки 001 – код 10001 и т. д.

Контроллер ЭЛТ (CRTС-контроллер) предназначен для генерирования сигналов горизонтальной (HSYNC) и вертикальной (VSYNC) синхронизации, формирования на экране курсора и инкремента счетчика адреса видеобуфера синхронно с разверткой раstra; в этом заключается основная функция контроллера. В качестве дополнительной функции отметим определение формы и позиции курсора, выбор выводимой на экран части видеобуфера, а также восприятие сигналов светового пера.

Независимо от схемотехнической реализации контроллера ЭЛТ во всех видеосистемах более поздних моделей предусмотрено задание их временных характеристик.

Контроллер атрибутов предназначен для организации дополнительных режимов работы видеосистем, связанных с выполнением основных функций по кодированию цвета символа и окружающего его фона. Для кодировки цвета используется композиция трех основных цветов – красный (R), синий (B) и зеленый (G). В простейшем случае для кодирования каждого из них достаточно иметь по одному биту (1 = цвет включен, 0 = цвет выключен). В результате из совокупности трех битов возможно получение 8 цветов. Если ввести еще один бит, который управляет яркостью, то получится 4-битная комбинация, позволяющая получить на экране монитора до 16 цветов. Данный вариант кодировки цветов используется в видеосистемах CGA и более поздних текстовых режимах.

В адаптере EGA имеются режимы, в которых для кодирования каждого из цветов используется не один, а два бита (00 = цвет выключен, 01 = слабый цвет, 10 = обычный, 11 = яркий цвет). В результате полная цветовая гамма расширяется до 64, тем не менее, одновременно на экране можно наблюдать только 16, так как в видеобуфере цвет пикселя кодируется 4-битным значением. Кодировка же недостающих цветов осуществляется программно в контроллере атрибутов.

Аналогичная ситуация для адаптеров MCGA, VGA, SVGA, в которых для кодирования каждого из основных цветов отведено по 6 бит, т.е. полная цветовая гамма увеличена до 256K цветов. Однако 8-битная кодировка в видеобуфере ограничивает максимально возможную цветность экрана монитора до 256.

В качестве мониторов в персональных компьютерах используют дисплеи трех типов: монохроматический монитор прямого управления, композитные мониторы и цветные RGB-мониторы.

Принципы работы монохроматического монитора с прямым управлением и RGB-монитора достаточно близки. Каждый из них использует два сигнала горизонтальной HSYNC и вертикальной VSYNC синхронизации и, в зависимости от того цветной или черно-белый монитор в системе, соответственно четыре (I, R, G, B) или одну (I – интенсивность) линии передачи видеосигнала. Композитные мониторы используют композитный сигнал, формируемый

адаптером. Особенность этого сигнала заключается в объединении собственно видеосигнала с сигналом синхронизации. Подключение монитора к адаптеру осуществляется с помощью коаксиального кабеля. В этом случае цвет может либо подавляться на выходе композитного сигнала, но он всегда сохранен на выходах RGB, либо закодирован непосредственно в композитный сигнал. Последний из перечисленных вариантов в последнее время практически не используется из-за невысокого качества формируемого на экране изображения.

2.6. Текстовый режим работы видеосистем

Все видеосистемы ПК, за исключением адаптера MDA, способны работать в двух основных режимах – в текстовом и графическом. Основное отличие каждого из режимов заключается в способе представления информации в видеобуфере, и, как следствие этого, в некоторых особенностях вывода информации на экран.

В текстовом режиме экран, а следовательно, видеобуфер разделяется на отдельные символьные позиции, в каждой из которых располагается только один символ. Символьные позиции экрана определяются двумя координатами, где Y-координата соответствует номеру строки, а X-координата – номеру текстового столбца. Начало координат (позиция 0,0) находится в верхнем левом углу раstra. Изображение (конфигурация) символа в символьной позиции формируется с помощью точечной матрицы, размер которой определяется типом используемого адаптера и номером его режима. Точки (пиксели), образующие изображение символа, называются передним планом, а остальные пиксели знакоместа – фоном.

В видеобуфере каждой символьной позиции соответствуют два байта. Байт с четным адресом содержит ASCII-код символа, а байт со следующим нечетным адресом – его атрибуты. Адаптер считывает их с помощью знакогенератора и преобразует в видеосигнал. Область видеобуфера, необходимую для заполнения экрана, называют страницей. В зависимости от количества символов стандартная емкость страницы может составлять 2 Кбайта (размер 25*40) или 4 Кбайта (размер 25*80). При этом "лишние" 48 или 96 байт не используются.

Если емкость видеобуфера превышает размер одной страницы, то в нем можно организовать несколько страниц. В любой момент времени на экран выводится содержимое только одной страницы, которая называется активной или текущей. Наличие в видеобуфере нескольких страниц позволяет оперативно переходить от одной страницы к другой, изменять формат текста или выполнять другие процедуры. Кроме того, неактивные страницы всегда доступны процессору, что позволяет при отображении одной из них на экране, выполнять операции программного обновления с других страниц. Напомним, что для вывода символа на экран, необходимо поместить его код и атрибуты в определенную ячейку видеобуфера. Благодаря линейному отображению относительный адрес или смещение видеобуфера на экране можно получить из выражения:

$$\text{Смещение} = ((\text{Строка} * \text{Ширина}) + \text{Столбец}) * 2, \quad [2.1]$$

где под шириной понимают количество символов в каждой текстовой строке, а под строкой и столбцом, соответственно, их номер. Значение строки и столбца отсчитывается от верхнего левого угла экрана, начиная с нуля. Следовательно, если, например, необходимо отобразить символ в первой позиции четвертой строки, то относительный адрес видеобуфера будет равен:

$$((4 * 80) + 2) = 640_{10} = 0280_{16}$$

Это смещение используется прикладными программами для операций записи/считывания в конкретную область экрана.

Независимо от вида видеосистемы четные байты видеобуфера содержат ASCII-коды отображаемых символов, в то время как функциональное назначение битов атрибутного байта зависит от конкретного типа видеосистемы. В общем случае формат атрибутного байта условно

разделен на две тетрады (рис. 2.6), в котором младшая тетрада отвечает за атрибуты переднего плана, а старшая – за атрибуты фона. При этом старший бит атрибутного байта в большинстве видеосистем выполняет две функции: либо отвечает за включение режима мерцания, либо управляет интенсивностью фона. Выбор режима определяется состоянием 5-го бита регистра, управляющего режимами в атрибутном контроллере. Содержимое атрибутного байта используется адаптером для формирования сигналов управления монитором. Так, например, в адаптере CGA биты тетрады соответствуют символам цветов и интенсивности переднего плана и фона символьной позиции.



Рис. 2.6. Формат атрибутного байта адаптера CGA

Что касается адаптера MDA, то биты атрибутного байта позволяют компенсировать отсутствие цвета управлением дополнительных способов выделения информации на экране, такими как, например, подчеркивание и негативное изображение. Символьные позиции адаптеров MDA и CGA соответственно составляют 9*14 и 8*8, а собственно изображение символа занимает матрицу соответственно 7*9 и 7*7 пикселей. Несовпадение размеров символьной матрицы и матрицы самого символа связано с необходимостью разграничения последнего как по строкам, так и по столбцам.

Адаптер EGA, в зависимости от заданного режима работы позволяет реализовать символьную позицию размерностью 9*16, 8*8 или 8*14 пикселей. В 16-ти цветных текстовых режимах данный адаптер использует такой же формат атрибутного байта, как и адаптер CGA. Однако в адаптере EGA введены дополнительные преобразования 4-битных значений цветов переднего плана и фона. Алгоритм этих преобразований заключается в следующем: каждое четырехбитное значение маскируется младшими битами регистра разделения цветовой плоскости в атрибутном контроллере, и результат этого маскирования выбирает один из 16 регистров палитры (рис. 2.7).

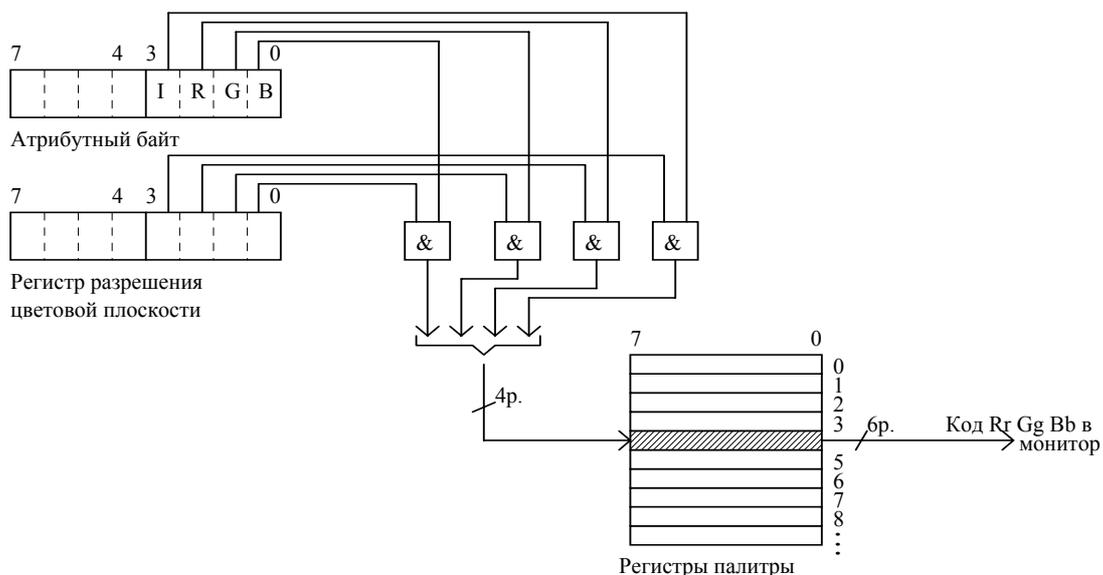


Рис 2.7. Формирование цвета пикселя в адаптере EGA

Таким образом, четырехбитный результат маскирования является адресом (индексом) регистра палитры, в каждом из которых содержится 6-битная информация для схем управления электронными пучками ЭЛТ. Напомним, что для кодировки каждого из цветов в адаптере EGA

используются не один, а два бита: 00 = цвет выключен; 01 = слабый цвет; 10 = обычный цвет; 11 = яркий цвет. Формирование сигналов цветов через регистры палитры обеспечивает адаптеру EGA существенно большую гибкость в окраске символов и фона по сравнению с адаптером CGA. В нем сохраняется возможность задания цветов переднего плана и фона независимо для каждого символа и дополнительно появляется возможность "полноэкранного" изменения цветов при обновлении содержимого одного или нескольких растров палитры. Для модифицирования содержимого данных регистров можно использовать либо команды прямой загрузки в регистр, либо процедуры BIOS. При использовании последней BIOS предоставляет возможность одновременной загрузки всех 16 регистров палитры из таблицы значений цветов.

При работе в монохроматическом текстовом режиме адаптер EGA полностью эмулирует монохроматические атрибуты адаптера MDA. В этом случае процедура BIOS инициализирует регистры палитры на следующие значения: 00h - черный; 08h - обычная яркость; 10h - серый; 18h - яркий, т. е. бит 3 определяет состояние включено/выключено пикселя, а бит 4 управляет яркостью.

Адаптеры MCGA, VGA и SVGA принципиально отличаются от описанных выше использованием аналогового способа формирования управляющих напряжений электронных пушек. Для этих целей в состав адаптера входят три 6-разрядных цифроаналоговых преобразователя (ЦАП), соответствующих цветам RGB. Преобразование атрибутов в сигналы цвета осуществляется по цепочке: атрибутный байт → форматор → регистр цвета ЦАП → ЦАП → монитор. В функции форматора входит дешифрация атрибутного байта в 8-битный код выборки регистра цвета ЦАП. Этот код выполняет функции, аналогичные функциям 4-битного маскированного кода адаптера EGA, т.е. является индексом одного из 256 регистров цвета ЦАП. Регистры цвета ЦАП имеют 18-битную организацию и выполняют функции, аналогичные функциям регистров палитры адаптера EGA. Процесс формирования цвета пикселя от атрибутного байта до уровней управляющих напряжений монитора изображен на рис. 2.8.

В текстовых режимах четыре младших бита на выходе форматора образуются из атрибутного байта, а четыре старших всегда нулевые. Следовательно, для генерирования выходных сигналов цвета используются только 16 первых регистров цвета ЦАП, а остальные недоступны. В исходном состоянии адаптера MCGA процедура BIOS инициализирует именно эти регистры цвета ЦАП на формирование тех же цветов, что и в адаптере CGA.

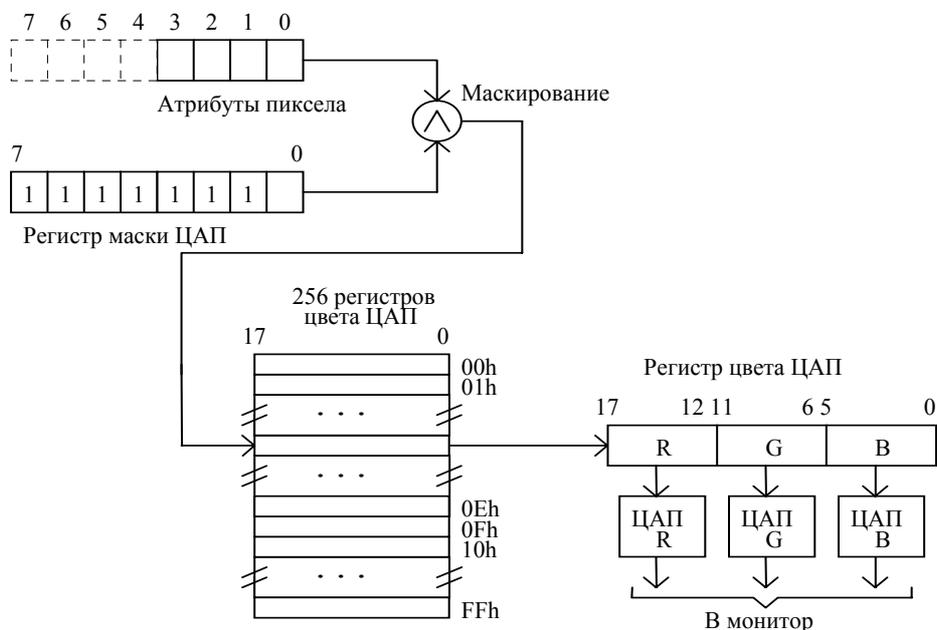


Рис. 2.8. Формирование цвета пикселя в адаптере MCGA

При работе с монохроматическим монитором адаптер MCGA не эмулирует режимы работы адаптера MDA. Вместо этого в первые 16 регистров ЦАП загружаются значения, которые образуют четыре группы по градациям уровня серого. Каждая последующая группа обеспечивает

более яркое изображение, чем предыдущая, а внутри группы интенсивность возрастает в соответствии с ростом значений атрибута (четыре младших бита в каждом 6-битном поле RGB). Следовательно, значения атрибутов 00h – 03h, 04h – 07h, 08h – 0Bh, 0Ch – 0Fh образуют диапазон все более и более ярких уровней серого.

Адаптер VGA (SVGA) можно считать объединенным вариантом адаптеров EGA и MCGA. В этом адаптере эмулируется дешифратор атрибутов адаптера EGA через 16 регистров палитры в совокупности с полной архитектурой адаптера MCGA (рис. 2.9). В связи с этим содержимое регистра палитры используется схемами последующего расширения до 8 бит и в операции маскирования, результат которых выступает в качестве индекса регистров цвета ЦАП для определения окончательного цвета пикселя.

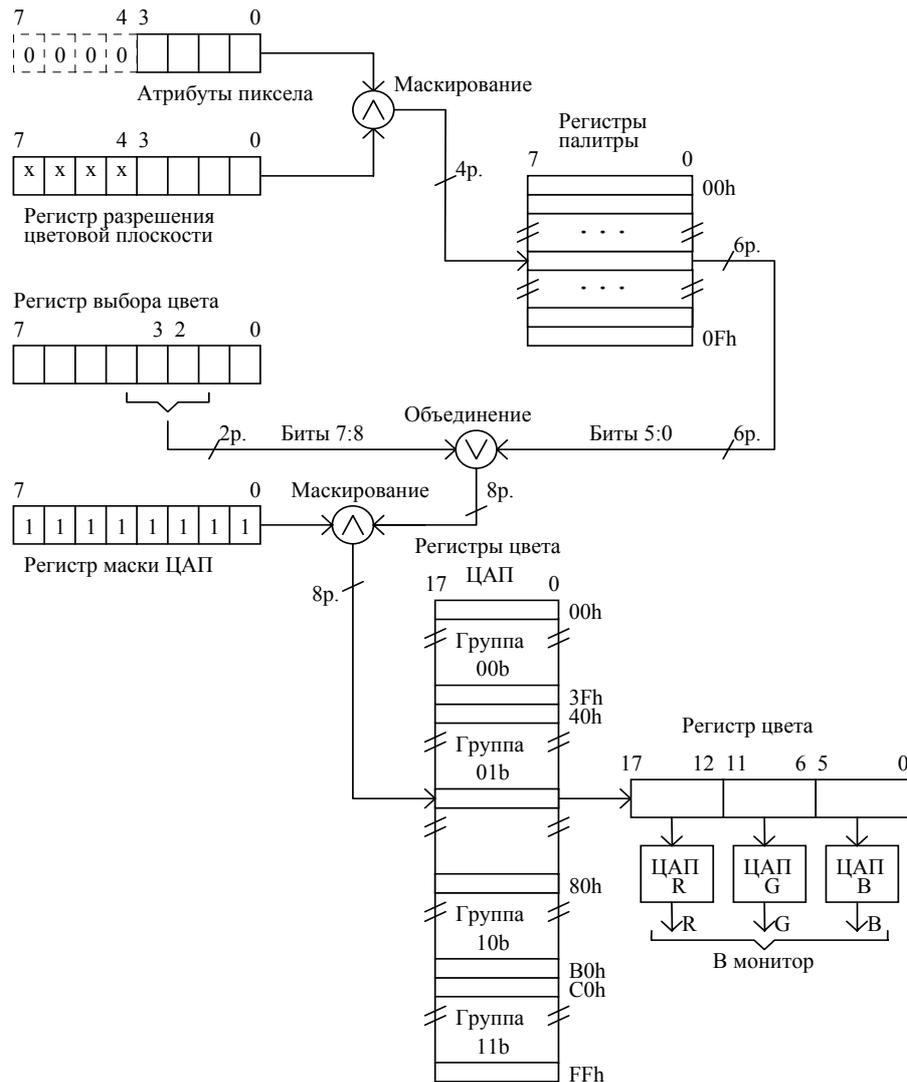


Рис. 2.9. Формирование цвета пикселя в адаптере VGA (бит 7 регистра управления режимом содержит 0)

В зависимости от состояния регистра управления режимом атрибутного контроллера рассматривается два способа выбора регистра ЦАП. Основное отличие этих режимов заключается в алгоритме объединения младших битов регистра палитры с младшими битами регистра выбора цвета. В первом алгоритме объединяются биты 0 – 5 регистра палитры с битами 2, 3 регистра выбора цвета, во втором биты 0 – 3 каждого из регистров (рис. 2.10).

Существующая двухуровневость преобразования атрибута пикселя, обеспечиваемая совместным использованием регистров палитры и регистров цвета, позволяет достаточно просто изменять палитру экрана. Для выбора любого из 256 регистров ЦАП достаточно изменить

содержимое битов регистра выбора цвета, а не перезагружать регистры палитры, как это делается в адаптере EGA. Такой прием позволяет организовать режим мерцания символов одновременно на значительной площади экрана. Для этого достаточно загрузить в регистры цвета ЦАП 16 палитр с постепенно увеличивающейся интенсивностью.

BIOS для адаптеров MCGA, VGA и SVGA позволяет преобразовывать значения красного, синего и зеленого цветов из регистров ЦАП в соответствующие значения оттенков серого. Для получения эквивалентного значения уровня серого производится вычисление суммы:

$$Y = 0.3R + 0.59G + 0.11B, \quad [2.2]$$

где Y – уровень серого, а R, G и B – равные интенсивности чистых цветов красного, зеленого и синего. BIOS считывает содержимое группы регистров цвета ЦАП, вычисляет для них слагаемые в выражении [2.2] и загружает результат вычислений в исходные регистры. Данный алгоритм автоматически реализуется при подключении к адаптеру монохроматического монитора и исключается при наличии цветного.

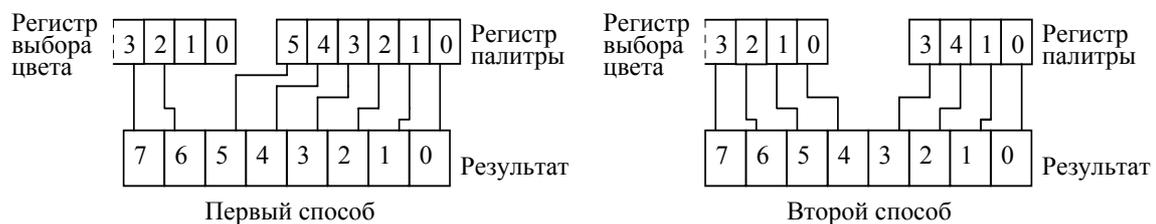


Рис. 2.10. Иллюстрация формирования результата объединения в зависимости от состояния регистра управления атрибутного контроллера адаптера VGA

2.7. Организация дисплейных страниц и управление курсором

Как отмечалось выше, большинство адаптеров имеют емкость видеобуфера большую, чем требуется для заполнения всего экрана текстовыми данными. В связи с этим предусмотрена возможность программирования контроллера CRTС для вывода на экран любых смежных байт видеобуфера. Для текстовых режимов с форматом 80*25 на экран выводится 4 Кбайт видеобуфера, а с форматом 40*25 – 2 Кбайт. Напомним, что та часть видеобуфера, которая непосредственно выводится на экран, называется окном или активной страницей. Следовательно, при общей емкости видеобуфера 16 Кбайт, в нем может быть организовано четыре или восемь страниц в зависимости от текстового режима работы (рис. 2.11).

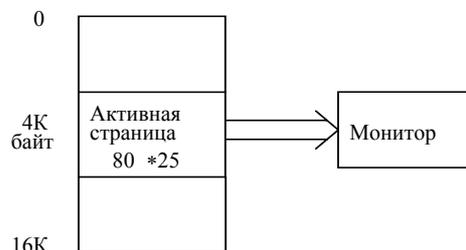


Рис. 2.11. Вывод на экран любой области буфера

Для программирования адаптера на вывод различных областей видеобуфера необходимо модифицировать регистры начального адреса контроллера CRTС – регистр 0Ch (старший байт адреса) и регистр 0Dh (младший байт). При загрузке в регистр начального адреса смещения (см. выражение [2.1]) можно активизировать любую часть видеобуфера, присвоив ей функции активной страницы. Загрузка этих регистров осуществляется гораздо быстрее пересылок кодов

символа и их атрибутов в видеобуфере, что приводит к практически мгновенному обновлению содержимого экрана.

При желании можно считать весь буфер "виртуальным экраном", содержащим, например, для емкости видеобуфера 16 Кбайт 102 текстовые строки. Когда видеобуфер полностью заполнен текстом, можно почти мгновенно вывести на экран любые 25 смежных строк.

В адаптере MDA видеобуфер занимает адресное пространство от B000:0000h до B000:0FFFh и его нельзя изменить. Видеобуфер адаптера CGA занимает адресное пространство от B800:0000h до B800:3FFFh и также не допускает изменения адреса. Видеобуферы адаптеров EGS и VGA допускают отображение на четыре области адресного пространства в зависимости от состояния вспомогательного регистра контроллера: A000:0000h – B000:FFFFh; A000:0000h – A000:FFFFh; B000:0000h – B000:7FFFh и B800:0000h – 7FFFh. Видеобуфер 64 Кбайт адаптера MCGA занимает область адресов от A000:0000h до A000:FFFFh. Второй видеобуфер емкостью 32 Кбайт по начальному адресу отображается на адресный диапазон B800:0000h – B800:7FFFh видеобуфера адаптера CGA.

При работе с постраничной организацией видеобуфера необходимо помнить о существовании в области данных дисплея по адресу 0040:0062h переменной ACT_PAGE, значение которой является номером текущей активной страницы, а также по адресу 0040:004Eh переменной CRT_START, значение которой задает смещение начала выводимой области видеобуфера в байтах. Поэтому при изменении содержимого регистров начального адреса видеобуфера (регистры 0Ch и 00h), необходимо одновременно модифицировать обе переменные ACT_PAGE и CRT_START.

Адресное пространство рассматриваемых видеобуферов является составной частью адресного пространства компьютера. Поэтому к видеобуферу постоянно обращаются как схемы регенерации адаптера в режиме считывания, так и прикладные программы (т.е. процессор) в режиме записи/считывания. Очевидно, что при обращении к видеобуферу одновременно в режимах записи и считывания изображение на экране монитора подвергается кратковременным искажениям в виде "снега". Чтобы избежать появления снега, процессор должен обращаться к видеобуферу в те интервалы времени, когда схемы регенерации не выполняют операций считывания, т.е. когда электронный луч либо находится за кадром, либо выполняет операции обратного хода.

Фиксация этих временных интервалов осуществляется схемами CRTС и отображается на содержимом битов разрешения ввода/вывода программно доступного регистра состояния контроллера. В конце каждой строки развертки ввод/вывод разрешен в течение 7 мкс, а по окончании последней строки развертки в кадре – в течение около 4 мкс. Этого времени достаточно для пересылки примерно 1200 байт информации в пределах одного кадра при частоте синхронизации 4.77 мгц, что используется на практике для подавления "снега".

Вместе с уже рассмотренной функцией контроллера CRTС по организации многостраничного режима работы видеосистем и подавления помех типа "снега", в текстовых режимах данный контроллер обеспечивает управление формой и позицией аппаратного курсора на экране. Форму курсора определяет содержимое двух регистров контроллера, задающее верхнюю и нижнюю строки развертки в символьной позиции, нумерация которых выполняется с нуля сверху вниз. При достижении нижней строки происходит "наворачивание курсора" на верхнюю строку. Позиция курсора задается также с помощью двух регистров контроллера в которых находится смещение относительно базового видеобуфера (см. выражение [2.1]).

Во всех видеосистемах фирмы IBM формой курсора управляют регистры 0Ah и 0Bh контроллера CRTС, в которых младшие биты определяют верхнюю и нижнюю строки развертки формируемого курсора (рис. 2.12 б)). Таким образом, для задания формы курсора или его изменения достаточно запрограммировать оба регистра. На рис. 2.12 а) приведены примеры возможных форм курсора на символьной матрице размерами 8*8.

На рис. 2.12 а) приведены примеры возможных форм курсора на символьной матрице 8*8, а на рис. 2.12 б) – формат этих регистров.

Содержимое битов 5, 6 регистра 0Ah служит для управления курсором и имеет следующий смысл:

- 00 – режим немерцающего курсора;
- 01 – вывод курсора запрещен;

Ix – изменение частоты мерцания курсора.



Рис. 2.12. а) Формы курсора; б) Формат регистров формы курсора

Регистры 0Eh и 0Fh, формат которых аналогичен приведенному на рис. 2.12 б), определяют символьную позицию в видеобуфере, в которой находится курсор.

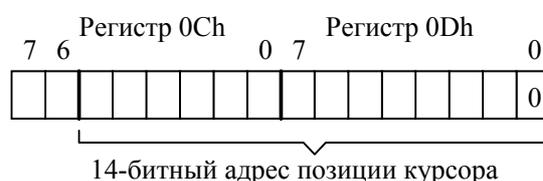


Рис. 2.13. Регистры начального адреса видеобуфера

Позиция курсора задается как смещение от начального адреса видеобуфера с добавлением младшего нулевого бита. Оба регистра допускают операции записи и считывания, что позволяет помещать курсор в любую символьную позицию экрана или определять его текущую позицию. Данный алгоритм формирования формы курсора применяется в адаптерах MDA, CGA и MCGA. В адаптере EGA значение, загружаемое в регистр конца курсора, должно быть на единицу больше номера нижней строки развертки курсора. В адаптере VGA нет режима “заворачивания” курсора на верх символьной матрицы, если номер строки в регистре 0Ah больше номера нижней символьной строки матрицы. Кроме того, в данном адаптере биты 6–5 регистра конечной строки развертки курсора (0Bh) определяют сдвиг курсора вправо относительно текущего его месторасположения. Если в них не содержится комбинация 00b, то десятичный эквивалент битов 6–5 определяет количество сдвигов вправо.

Для удаления курсора из области экрана (невидимый курсор) следует запрограммировать CRTS таким образом, чтобы либо позиция курсора оказалась за пределами выводимой части видеобуфера, либо значения содержимого регистров начала и конца курсора были больше количества строк в символьной матрице.

2.8. Графический режим работы видеосистем

В отличие от текстового режима работы, в графических режимах любая прикладная программа должна управлять цветом каждого пикселя, выводимого на экран. В связи с этим данные графических режимов хранятся в видеобуфере в виде двоичных полей. Максимальная длина двоичного поля ограничена восемью битами, поэтому в каждом байте видеобуфера содержится информация о представлении одного или нескольких пикселей на экране. Формат двоичной карты видеобуфера определяется количеством бит, отводимых для “описания” каждого пикселя, а нумерация битов и пикселей в пределах байта представлена в противоположном направлении. Например, при формате двоичной карты равном четырем бит, левый пиксель кодируется битами 4 – 7, а правый – битами 0 – 3, т.е. выполняется линейное отображение видеобуфера на экран. На рис. 2.14 представлены варианты кодировки пикселей в видеобуфере адаптера CGA.

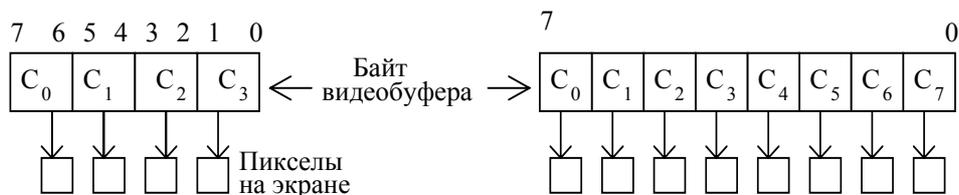


Рис. 2.14. Представление данных в видеобufferе адаптера CGA

Адаптер CGA поддерживает два графических режима: 4-цветный режим 320*200 и 2-цветный режим 640*320. Для первого из них каждый пиксель представлен в видеобufferе двумя битами, для второго – одним (рис. 2.12). Независимо от режима данные в видеобufferе хранятся в виде двух блоков: информация 100 четных строк развертки занимает 8 Кбайт с начальным адресом B800:0000h, а 100 нечетных строк развертки (тоже 8 Кбайт) с начальным адресом B800:2000h. Контроллер CRTС программируется на вывод 100 строк "символов" высотой в две строки развертки, при этом верхняя (четная) строка берется из нижнего блока, а нижняя (нечетная) строка – из второго.

В двухцветном режиме 640*200 каждый пиксель представлен одним битом. Если бит равен 0, то пиксель выводится на экран как черный, в противном случае цвет пикселя определяют четыре бита регистра выбора цвета. Процедура BIOS поддерживает в области данных дисплея по адресу 0040:0066h переменную CRT_PAL, значение которой определяет содержимое регистра выбора цвета. Следовательно, для задания цвета необходимо модифицировать данную переменную.

В 4-х цветном режиме 320*200 каждый пиксель представлен значениями 00b, 01b, 10b и 11b. При значении 00b цвет пикселя совпадает с цветом окаймления экрана и задается содержимым регистра выбора цвета. Цвет пикселя с ненулевыми значениями задается с помощью трех аппаратных палитр согласно таблице 2.5.

Таблица 2.5. Определение аппаратных палитр в адаптере CGA

Режим	0	0	1
Выбор цвета	0	0	1
01b	Зелёный	Голубой	Голубой
10b	Красный	Вишнёвый	Красный
11b	Коричневый	Белый	Белый
Значение пикселя	Цвет пикселя		

Термин "аппаратный" связан с тем, что программными средствами изменить состав входящих в них цветов не представляется возможным. В графе "режим" таблицы 2.5 указано содержимое бита 2 регистра управления режимом, а в графе "выбор цвета" – бита 5 одноименного регистра контроллера CRTС. Модификация этих регистров может быть осуществлена прямым программированием, либо командой INT 10h.

Адаптер EGA позволяет полностью эмулировать режимы работы адаптера CGA в графическом режиме. Однако в собственном графическом режиме (16-цветный 200 или 350 строчный) каждый байт видеобufferа содержит информацию о восьми пикселях. Это связано с тем, что видеобuffer адаптера емкостью 256 Кбайт состоит из четырех параллельных банков ОЗУ в том смысле, что они занимают одно и тоже адресное пространство со стороны процессора. Значения пикселя определяют состояния бит с одинаковым смещением байта в каждом банке и с одинаковыми номерами бит в пределах байта. Следовательно, цвет каждого пикселя определяют четыре бита, взятые одновременно из каждого банка (рис. 2.15).

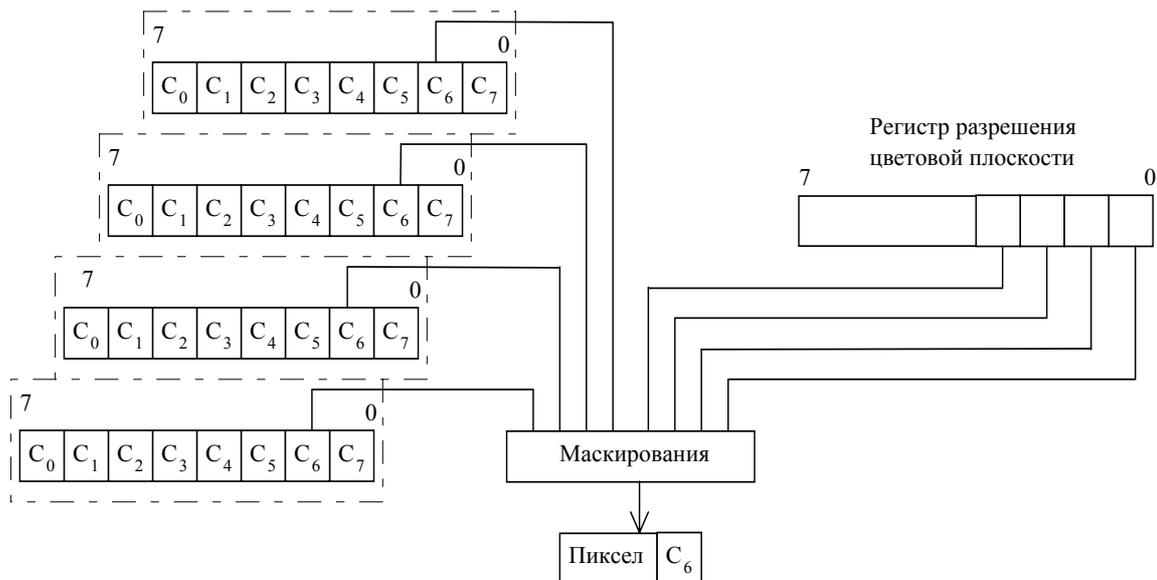


Рис. 2.15. Схема определения цвета пикселя

Полученное из видеобуфера 4-х битное значение пикселя объединяется по логическому И (маскируется) с содержимым младшей тетрады регистра разрешения цветовой плоскости, а результирующий код является индексом для 16 регистров палитры. Дальнейший алгоритм формирования цвета аналогичен формированию цвета для текстового режима работы адаптера EGA.

В собственном режиме работы в данном адаптере не применяется чересстрочная организация видеобуфера как в адаптере CGA. Вместо этого строки пикселей отображаются линейно аналогично линейному отображению текста в текстовых режимах (см. п. 2.6).

Адаптеры MCGA и VGA имеют три графических режима работы, два из которых напоминают собственные режимы адаптера CGA, но в случае построчной организации видеобуфера линейно отображаются на экран. Для третьего из них каждый байт видеобуфера соответствует одному пикселю на экране. В этом случае считываемый из видеобуфера байт объединяется по И с содержимым регистра маскирования, и результирующее значение выбирает один из 256 регистров цвета ЦАП.

При двухцветном графическом режиме управление цветами пикселей осуществляется следующим образом. Пиксели, значения которых равны 0, всегда отображаются через регистр цвета 0 ЦАП. Пиксели с ненулевыми значениями выбирают только один определенный регистр цвета ЦАП (обычно 07h). Цвет фона в данных адаптерах задается аналогично цвету пикселя с помощью регистра цвета ЦАП. Таким образом, для цветов фона и переднего плана можно установить любые два цвета из возможных 256К цветов.

В четырехцветном режиме 320*200 байт описывает цвет четырех пикселей. В результате два бита байта видеобуфера и два бита регистра выбора цвета (биты 5 и 4) образуют четырехбитное значение индекса одного из первых 16 регистров ЦАП. Таким образом можно выбрать любые четыре цвета из общей палитры 16 цветов. Для изменения палитры необходимо модифицировать регистр выбора цвета.

Аналогичным образом осуществляется управление цветом пикселя в случае его байтового представления в видеобуфере. Данный режим позволяет формировать 256-ти цветное изображение 320*200 пикселей из палитры 256К цветов. В этом случае процедура BIOS программирует регистры цвета ЦАП на следующие значения:

- регистры 00h - 0Fh содержат значения цветов, совместимые с цветами адаптера CGA;
- регистры 10h - 1Fh содержат различные уровни серого с постепенно увеличивающейся яркостью;

- регистры 20h -F7h содержат по три группы из 72 цветов с постепенно увеличивающейся яркостью;
- регистры F8h - FFh - значения черного цвета, т. е. обнулены.

Следует отметить, что в адаптере VGA атрибутный контроллер дешифрует значения пикселей с использованием регистров палитры, как это делается в текстовом режиме (см. п. 2.6), а сам видеобuffer организован в виде двоичных плоскостей, т. е. аналогично видеобufferу EGA.

В современных видеосистемах SVGA существуют дополнительные графические режимы: High Color, True Color и Real Color, в которых данные из видеобufferа непосредственно поступают на входы цифроаналоговых преобразователей. То есть в этих режимах приведенные выше схемы обработки двоичных полей не участвуют. Режимы High Color и Real Color близки, отличия заключаются в разрядности двоичных полей. Так первый из них использует 15-разрядное двоичное поле (5 бит на цвет), а второй – 16-разрядное. Разделение информации по основным цветам при 16-ти битном кодировании не одинаково и учитывает спектральную чувствительность глаза. Максимум разрядности приходится на зеленый цвет, а минимум – на синий. В настоящее время 16-ти битное кодирование обозначают как режим High Color.

Режим True Color предлагает самые высокие возможности создания реалистичных изображений. Он обеспечивает палитру цветов не менее 16,7 млн оттенков. Разумеется, что для графики глубиной в 24 разряда требуется большой объем памяти видеобufferа. Для оценки требуемого объема можно воспользоваться следующим выражением:

$$[\text{объем видеобufferа}] = (A * B * C) / 8 \text{ [байт]} \quad [2.3]$$

где: А – количество пикселей в строке;

В - количество строк пикселей;

С – размер двоичного поля.

При работе с Windows, операционная система самостоятельно оценивает возможности установленной в компьютере видеосистемы и предлагает пользователю выбрать один из возможных графических режимов.

2.10. Особенности графического программирования

Большинство графических программ опирается на элементарные процедуры записи/считывания значения пикселя между процессором и видеобufferом. В адаптерах CGA и MCGA эти процедуры позволяют непосредственно обращаться к ячейкам памяти видеобufferа, в то время как в адаптерах EGA и VGA доступом к значениям пикселей, хранящихся в двоичных плоскостях (банках), управляют специальные схемы, входящие в состав адаптера. Поскольку двоичные плоскости адресуются параллельно, то в процессе обращения по конкретному адресу происходит одновременное обращение к четырем ячейкам видеобufferа, расположенным в разных плоскостях. В этом случае для операций записи/считывания используются регистры-зашелки, посредством которых осуществляется обмен информацией между видеобufferом и процессором ПК.

При выполнении процессором операций считывания, данные из двоичных плоскостей одновременно переписываются в четыре регистра-зашелки, каждый из которых соответствует одной из плоскостей (рис. 2.16). Для выбора конкретного регистра-зашелки с целью передачи байта информации в регистр ЦП используется дополнительный регистр выбора считываемого банка, в котором содержимое двух младших бит определяет номер двоичной плоскости, начиная с нуля. Отметим, что в этом режиме (его называют режимом считывания 0) для получения значения пикселя используются четыре считывания, поскольку пиксель программируется одним битом в каждой из четырех плоскостей.

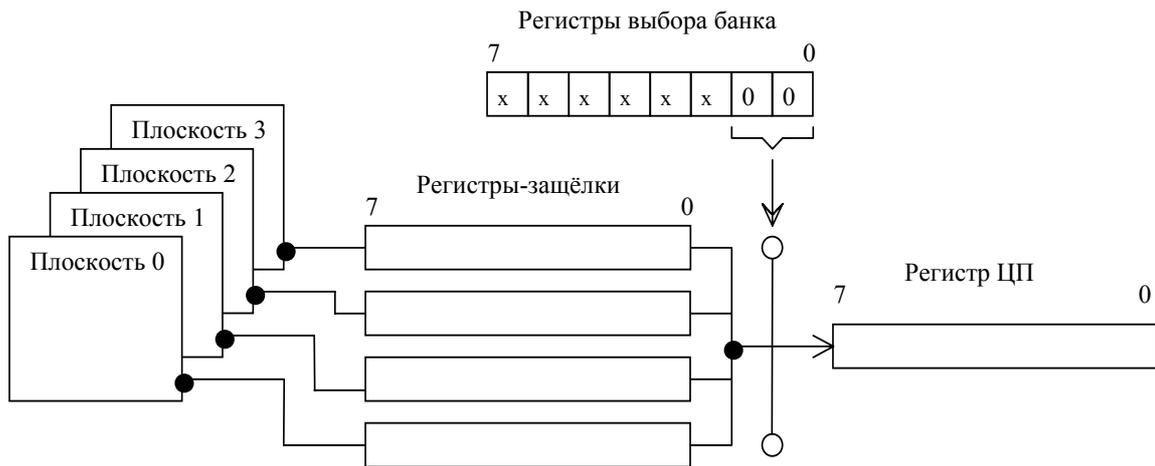


Рис. 2.16. Режим считывания 0

Кроме режима считывания 0 существует еще один режим (режим считывания 1), в котором значения восьми пикселей в регистрах-защелках сравниваются с содержимым младшей тетрады регистра сравнения цвета (рис. 2.17). Когда значение пикселя совпадает с содержимым тетрады регистра сравнения цвета, то в соответствующий бит регистра ЦП будет считана 1, в противном случае считается 0. В результате такого сканирования можно достаточно быстро находить пиксели, имеющие конкретное значение цвета, а также оперативно находить пиксели, имеющие цвет фона. Однако оба режима не позволяют также оперативно распознать текущее значение цвета пикселя, поскольку может потребоваться до 16 считываний, прежде чем в каждом из битов регистра ЦП будет считана единица.

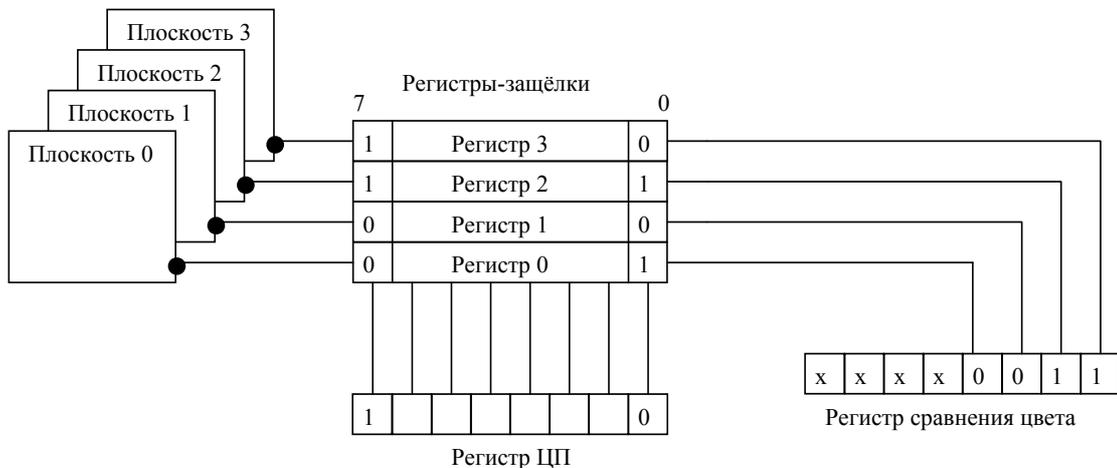


Рис. 2.17. Режим считывания 1

При выполнении процессором операций записи возникает обратная ситуация: данные из регистра ЦП должны быть переписаны в двоичные плоскости. В этом случае простейшим примером является режим записи 1, при выполнении которого данные из процессора последовательно переписываются в регистры-защелки и далее в двоичные плоскости видеобuffers. Остальные режимы используют достаточно сложный алгоритм передачи информации из регистров-защелок в видеобuffer с привлечением дополнительных регистров установки/сброса, разрешения установки/сброса, сдвига данных/выбора функции и двоичной маски. В качестве примера рассмотрим режим, в котором используется только один дополнительный регистр – регистр установки/сброса (рис. 2.18.). С помощью этого регистра можно выполнять различные логические процедуры как с содержимым регистров-защелок, так и с

данными, поступающими от процессора. Результаты логических процедур являются исходными данными для записи в двоичные плоскости.

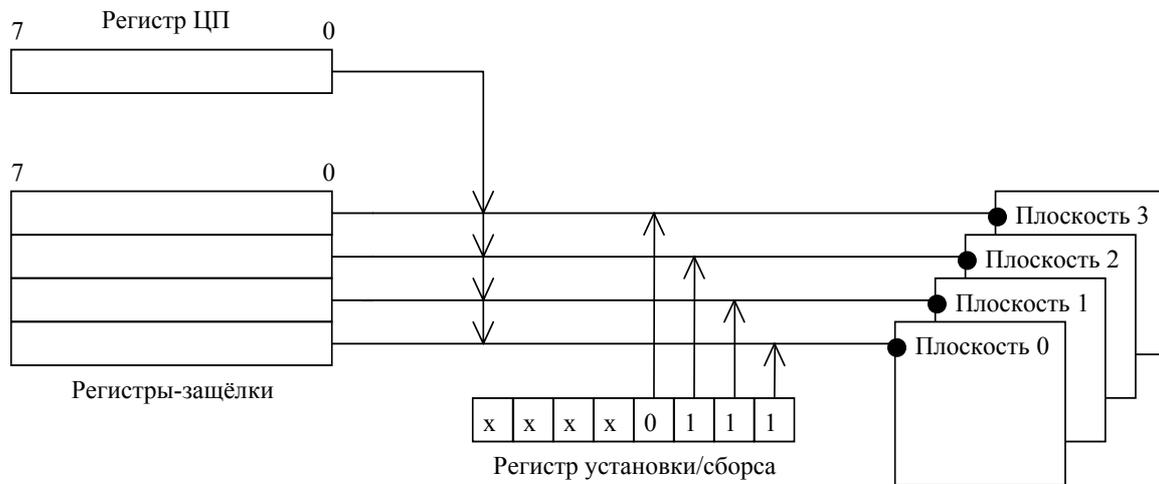


Рис. 2.18. Операции записи процессором в видеобуфер

Передачей данных между процессором, регистрами-защелками и видеобуфером управляет графический контроллер. Графический контроллер обычно содержит 9 регистров, содержимое которых определяет как обрабатывать данные в операциях записи/считывания. По существу графический контроллер обеспечивает двухмерную обработку находящихся в регистрах-защелках данных о восьми пикселях. Некоторые операции ориентированы на обработку байтов, т.е. в них участвуют одноименные биты всех пикселей отдельных регистров-защелок. Другие операции ориентированы на пиксели и в них данные из регистров-защелок представляют совокупность восьми значений пикселей.

Таким образом, графический контроллер позволяет выполнить три байтных операции над данными в регистрах-защелок: просто передавать данные из видеобуфера в регистры-защелки и обратно; передавать содержимое из одного регистра-защелок в регистр процессора; может объединять по И байт данных от процессора с байтом любого из регистров-защелок. Каждая операция представляет собой последовательность байтных и пиксельных процедур, на которые программируется графический контроллер.

При работе в графическом режиме вместе с элементами графики практически все прикладные программы позволяют выводить на экран монитора и текстовую информацию. В этом случае в видеосистеме необходимо организовать область памяти для пиксельного представления текстовых символов. Данную информацию удобнее всего хранить в виде регулярной таблицы, называемой фонтом, в которой каждый символ занимает определенную символьную матрицу пикселей. Совокупность символьных матриц представляет собственно таблицу шрифтов. При использовании символьной матрицы одного размера и при упорядочении расположения символов как это делается в таблице ASCII-кодов преобразование кода символа в смещение (индекс) таблицы шрифтов производится достаточно просто. Использование таблицы шрифтов в текстовых режимах позволяет оперативно изменять конфигурацию символов, выводить сложные математические символы и является необходимым атрибутом большинства прикладных программ.

Все IBM-совместимые персональные компьютеры имеют встроенную в BIOS таблицу шрифтов. При включении компьютера данная таблица может автоматически загружаться в область памяти видеоадаптера и является аналогом ПЗУ знакогенератора. В тех версиях BIOS, в которых это не предусмотрено, после включения компьютера таблица шрифтов находится в ПЗУ на системной плате, а загрузка шрифтов выполняется утилитой GRAFTABL операционной системы или драйвером дисплея. В этом случае загружаемая таблица шрифтов относится ко второй половине ASCII-кодов, т.е. к кодам 80h – Ffh. В качестве примера можно привести широко используемый в настоящее время драйвер uniser.com, который включается в файл autoexec.bat.

2.11. Принцип работы и классификация печатающих устройств

Печатающие устройства или принтеры предназначены для регистрации графической или текстовой информации на бумажном носителе и включают в себя многочисленную группу таких устройств. В качестве носителя информации в принтерах используется обычная или специальная бумага. Все принтеры принято классифицировать по следующим признакам:

1. По способу регистрации, т. е. по совокупности физико-химических процессов, используемых в устройстве для получения видимого изображения на бумаге.
2. По способу формирования изображения, полнопрофильное (печать символа целиком) или знаковосинтезирующее (изображение символа формируется из более мелких элементов).
3. По количеству символов, формируемых на бумаге за один такт работы принтера.

Способы регистрации подразделяются на ударные, струйные, термические и электрофотографические. *Ударный способ регистрации* основан на перенесении части красящего вещества на бумагу вследствие контакта, возникающего при ударе. Данный способ регистрации обеспечивает достаточно высокое качество печати с возможностью одновременного получения нескольких копий. К недостаткам следует отнести сравнительно малое быстродействие, высокий уровень шумов и невозможность получения цветных изображений.

Струйный способ регистрации основан на избирательном окрашивании участков бумаги с помощью струй жидкого красителя, которые выбрасываются из сопла под воздействием статического или импульсного давления. Возможны две разновидности струйного способа: непрерывный и импульсный. Для первого из них непрерывная струя распадается на отдельные капли, каждая из которых приобретает электрический заряд и посредством электрического поля формирует контур изображения, как это делает ЭЛТ. Во второй разновидности струйного способа выброс струи осуществляется в тот момент, когда сопло точно направлено в нужную точку изображения. Основным недостатком струйного способа связан с достаточно частым загрязнением сопел, что приводит к необходимости применения специальных систем очистки, а, следовательно, к удорожанию принтера в целом.

Существует несколько механизмов (методов) создания избыточного давления в соплах. Наибольшее распространение среди них получили следующие:

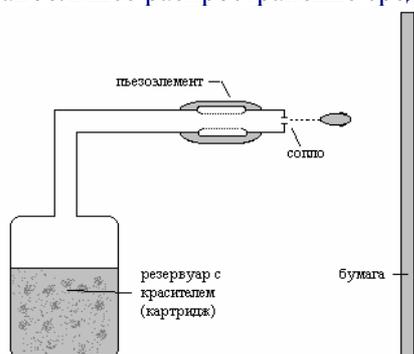


Рис. 2.19. Пьезоэлектрический метод

- Пьезоэлектрический метод, широко используемый компанией Epson (рис. 2.19). При реализации данного метода в трубку каждого сопла устанавливается пьезокристалл. Как известно, при подаче на него внешнего напряжения он деформируется. В результате трубка продавливается внутрь и капля красителя «выстреливается» на бумагу. Одновременно с этим избыток красителя выдавливается обратно в резервуар.

- Метод газовых пузырей, широко используемый компанией Canon (рис. 2.20). Данный механизм создания избыточного давления по природе является термическим и больше известен под названием инжектируемые пузырьки. Каждое сопло имеет встроенный нагревательный элемент, при пропускании тока через который, он быстро нагревается до порядка 500°C . Возникающие в процессе нагревания газовые пузыри «выстреливают» каплю красителя на бумагу,

а его избыток - обратно в резервуар. При отключении тока газовый пузырь уменьшается, и очередная порция красителя поступает в область нагревательного элемента.

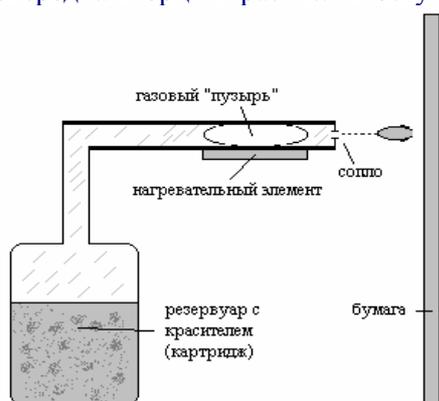


Рис. 2.20 Метод газовых пузырей

- Метод drop-on-demand разработан фирмой Hewlett-Packard и аналогично предыдущему является термическим. Однако в методе по-другому осуществляется подача красителя на нагревательный элемент (рис. 2.21). Если в методе газовых пузырей краситель втягивается из резервуара в область сопло за счет избыточного давления, то в данном методе транспортировкой красителя занимается специальное устройство. Как видно из рисунка 21 механизм печати заключается в испарении капли красителя с помощью нагревательного элемента.

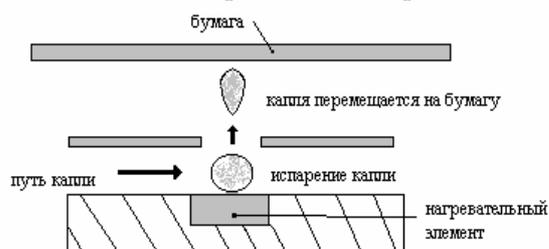


Рис. 2.21. Принцип действия метода drop-on-demand

Струйные принтеры используют печатающую головку, состоящую из совокупности множества сопел, расположенных в виде матрицы. Количество сопел зависит от модели принтера и может быть в интервале от 16 до 64. Некоторые последние модели имеют гораздо большее количество сопел. Например, печатающая головка принтера HP DeskJet1600 имеет 300 сопел для черно-белой печати и 416 для цветной.

Термический способ регистрации основан на способности ряда веществ изменять свои оптические свойства под воздействием температуры. Для формирования изображения применяется специальная печатающая термоголовка, соприкасающаяся с термобумагой. В последнее время появились принтеры, в которых термоголовка служит для нагревания красящей ленты (пленки), а в качестве регистрирующего элемента применяется обычная бумага. Такие принтеры получили название *сублимационных*. В них осуществляется нагрев красителя до жидкой или газообразной фазы и перенос его на бумагу. Независимо от этих особенностей в качестве термоголовки используют керамическую пластину, на поверхность которой нанесены нагревательные элементы - терморезисторы. Их расположение может быть как горизонтальным, так и вертикальным, или в виде матрицы.

В устройствах термического способа регистрации практически отсутствуют механически подвижные элементы, что обеспечивает хорошую устойчивость к ударным нагрузкам, высокую надежность и бесшумность работы. Однако качество печати и быстродействие ниже, чем в случае ударного или струйного способов регистрации. Это объясняется инерционностью процессов охлаждения и нагрева нагревательных элементов, а также отсутствием резкой границы между нагретыми и не нагретыми участками носителя.

Электрофотографический способ регистрации основан на явлении местного разрушения электростатического заряда под действием света и нашел широкое применение в лазерных

принтерах и копировальных аппаратах. Суть способа заключается в следующем (рис. 2.22): на заряженный коронирующий провод фотобарабан, имеющий на своей поверхности положительные ионы, попадает сканирующий лазерный луч (аналог электронного луча в ЭЛТ).

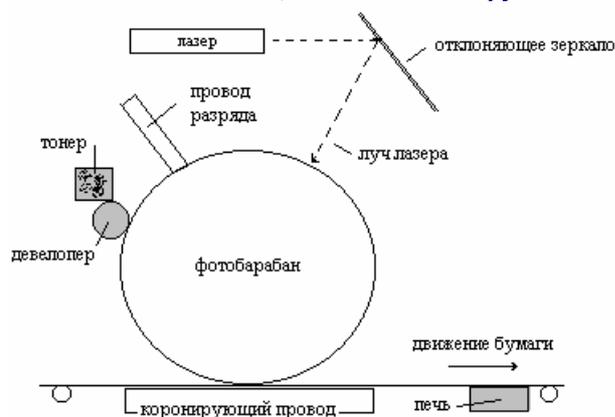


Рис. 2.22 Функциональная схема лазерного принтера

В результате в области попадания луча слабо связанные ионы рассеиваются. По окончании формирования раstra на фотобарабане образуется скрытое электростатическое изображение в виде отдельных точек, которое "проявляется" путем осаждения на положительно заряженные участки отрицательно заряженных ионов красителя. Затем краситель переносится на бумагу и фиксируется путем термосилового закрепления с помощью печи. По окончании печати поверхность барабана очищается, ионизируется и готова к формированию следующей страницы изображения. Девелопер выступает в качестве дозатора тонера, который равномерно покрывает его тонким слоем. Данный способ регистрации обеспечивает наилучшее качество печати, бесшумен и обладает высоким быстродействием. Однако лазерные принтеры достаточно сложны в изготовлении, требуют высококвалифицированного обслуживания и характеризуются высокой стоимостью.

В настоящее время появились более дешевые аналоги лазерных принтеров –LCD-принтеры. Принцип их действия полностью совпадает с функциональной схемой рис.2.22 с тем отличием, что лазер и отклоняющее зеркало заменены на светодиодную линейку. Такой принтер обладает меньшей контрастностью изображений, вследствие меньшей степени рассеяния связанных ионов. Однако он дешевле и позволяет получать изображения по качеству не хуже, чем струйные принтеры.

Контур напечатанного символа может быть непрерывным или состоящим из отдельных точек. Символы с непрерывным контуром принято называть полнопрофильными, а состоящие из отдельных элементов матричными. Для формирования полнопрофильного изображения используется специальный шрифтоноситель, с помощью которого путем выбора необходимого литеры (т. е. символа) осуществляется перенос его изображения на бумагу. Наиболее широко распространенными видами шрифтоносителей являются цилиндрические и сферические головки, лепестковые носители (ромашки), барабаны и ленты. Независимо от вида все они обеспечивают статическую печать, при которой отсутствует боковое смещение. Для каждой позиции печати на строке срабатывание молоточка происходит в тот момент, когда литер шрифтоносителя, соответствующий ASCII-коду данного символа, находится напротив молоточка. Основным недостатком полнопрофильных принтеров - невозможность получения графических изображений и ограниченность в выборе шрифтов.

Матричное изображение символа представляет собой аналог его пиксельного представления в символьной матрице ПЗУ знакогенератора видеосистем. Полное изображение символа формируется в процессе печати по строкам и напоминает формирование символа на экране растровых мониторов. Отличие заключается в одновременном сканировании как по строкам, так и по столбцам, т. е. за один проход печатающей головки по строке одновременно могут быть сформированы 9, 18 или 24 строки символьной матрицы. В результате такого прохода формируется изображение полной строки символов на бумаге.

По количеству символов, формируемых за один такт различают последовательные и параллельные печатающие устройства. Для первых из них за один такт работы на бумаге формируется изображение лишь одного символа в очередной позиции строки. Большинство параллельных устройств относится к классу полнопрофильных и их быстродействие достаточно велико. Тем не менее, наибольшее распространение получили последовательные матричные печатающие устройства.

2.12. Матричные принтеры

Как отмечалось ранее, в матричном принтере для формирования символа используется печатающая головка, состоящая из 9, 18 или 24-игольной матрицы. Микропроцессор в принтере управляет каждой из иглонок, которые, ударяя по красящей ленте, образуют отпечаток символа с небольшими расстояниями между отдельными точками. Данные принтеры имеют возможность работать как в текстовом, так и в графическом режимах. Обобщенная блок-схема матричного принтера изображена на рис. 2.19.

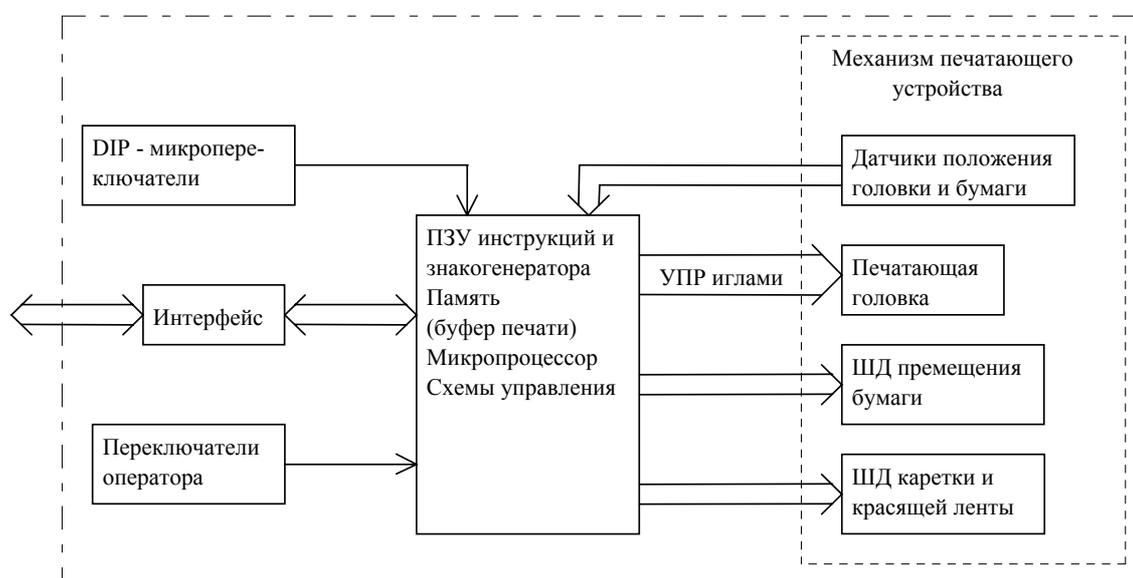


Рис. 2.19. Обобщенная блок-схема матричного принтера

Управление работой механизма печатающего устройства обычно выполняет микропроцессор, принадлежащий к семейству однокристальных микро-ЭВМ MCS-48 фирмы INTEL. Все выполняемые принтером операции, осуществляются микропроцессором в форме инструкций по определенной программе, размещенной в ПЗУ. Там же размещается соответствующий набор символов в форме таблицы шрифтов, количество которых может достигать 13 (например, международный набор символов EPSON).

Большинство принтеров позволяют работать с внешними устройствами синхронно или асинхронно. В синхронном режиме принятый ASCII-код после необходимых преобразований непосредственно вызывает печать соответствующего ему символа. В случае асинхронной работы принимаемые коды записываются в буфер печати, и после его заполнения начинается распечатка текста или графической информации. Емкость буфера печати как правило не менее 4 Кбайт.

Шаговые электродвигатели (ШД) предназначены для позиционирования печатающей головки в строке и прогонки бумаги. Для этих же целей служат датчики положения головки и бумаги, с помощью которых обеспечивается обратная связь для точного позиционирования головки.

Входной интерфейс предназначен для связи принтера с персональным компьютером и, как правило, имеет два режима работы: последовательный и параллельный. Последовательный режим работы аналогичен режимам работы интерфейса RC232C или интерфейсной БИС INTEL 8251-A. В параллельном режиме данные от ПК одновременно передаются по восьми линиям в байтовом формате.

Клавиши оператора предназначены для оперативного вмешательства в режим работы принтера и находятся на его лицевой панели. Общепринятый минимальный набор этих клавиш составляет три кнопки, одна из которых управляет режимом выбора качественного шрифта (NLQ), другая – режимом автоматического или неавтоматического состояния принтера (ON/OFF LINE), а третья – режимом протяжки бумаги (постраничная или построчная). Кроме того, практически все матричные принтеры имеют DIP-переключатели для выполнения некоторых стандартных установок печатающего устройства. Опрос состояния этих переключателей выполняется микропроцессором после включения принтера. Положение DIP-переключателей и выполняемые ими функции можно найти в инструкции по эксплуатации для подключенного к ПК принтера.

Связь между компьютером и печатающим устройством осуществляется с помощью кодов, выражающих символы и команды. Символы передаются в формате ASCII-кодов букв, знаков препинания, цифр и математических символов. Коды команд (33 кода из всей совокупности ASCII-кодов) применяются для таких целей как возврат каретки, перевод строки, включение звуковой сигнализации и т.п. Поскольку такого количества команд недостаточно для управления всеми функциями принтера, то большинство команд является последовательностью нескольких управляющих кодов, первый из которых всегда является кодом автогенератора 2 (код клавиши ESC). В этом случае о командах говорят как об ESC-последовательностях. С помощью ESC-последовательностей прикладные программы переводят печатающие устройства в текстовые или графические режимы и выполняют целый набор вспомогательных функций внутри каждого из них.

В текстовых режимах многие прикладные программы требуют максимума от печатающих устройств. При выводе текста на печать можно использовать несколько стилей печатания и различные шрифты, можно добавить заголовки и сноски, использовать шрифт жирного начертания, курсив, подчеркивание и другие эффекты. Кроме того с помощью управляющих кодов можно изменять печатные размеры текста и ширину символа, количество проходов печатающей головки по строке, а также установить режим печати справа-налево.

После включения питания принтер по умолчанию переходит в текстовый режим со стандартным для данного устройства и состояния DIP-микрпереключателей размером символов и поля печати на бумаге. Перевод принтера в графический режим осуществляется командой ESC C, n1, n2, где C – является буквой из набора K, L, Y, Z; n1 и n2 – определяют количество колонок, резервируемых для графики в пределах одной строки. При этом n1 представляет младший байт информации, а n2 – старший. Получив данную команду, печатающее устройство будет воспринимать все последующие данные как графическую информацию, т.е. как комбинацию срабатывания игл, и их количество должно точно соответствовать значению [n2 n1]. Если этих данных не хватает, то принтер будет останавливаться и ждать недостающей информации (это дает впечатление блокировки принтера). С другой стороны, если дают больше чем [n2 n1] количество графических данных, то избыточная часть представляемых кодов воспринимается как ASCII-коды и печатается как текст.

В графических режимах необходимо использовать определенный метод приведения каждой иглы в ударное действие в каждой строке. Обычно, этот метод основан на прямой пересылке байта данных из ПК, биты которого непосредственно связаны с 8-ю иглами печатающей головки. При этом старший бит управляет верхней из игл, в случае их вертикального расположения.

Другая графическая команда позволяет присвоить другой восьмиугольный графический режим, связанный с быстрым изменением плотности и пропорции распечаток. Эти изменения связаны с изменением ширины печати при неизменной высоте. Командой данного режима является команда ESC C m, где C – буква из набора K, L, Y, Z; m – номер режима согласно инструкции для конкретного печатающего устройства.

Для оперативной пересылки кодов команд в печатающее устройство можно использовать такие языки программирования как БЕЙСИК или ПАСКАЛЬ. Применение БЕЙСИКа более

предпочтительно, поскольку большинство персональных компьютеров комплектуется данным языком, а ряд наиболее часто используемых его операторов уже находится в ПЗУ BIOS. В этом случае вывод команд удобнее выполнять с помощью оператора LPRINT в совокупности с функцией CHR\$(), где в круглых скобках указывается десятичное представление кода ESC или других командных кодов.

Если используется БЕЙСИК другой версии или другой язык программирования, то необходимо ознакомиться с инструкцией программиста для нахождения правильного формата этих команд. Обычно управляющая последовательность команд оформляется в виде командного файла, т.е. файла с расширением .bat. Его можно запустить из командной строки DOS, задав тем самым определенный режим работы принтера.

Глава 3 Дисковая система персональных компьютеров

3.1. Технология магнитного диска

Практически все современные компьютеры в качестве внешних устройств хранения информации используют накопители на магнитных дисках. Их количество и тип, в принципе, не регламентируются и определяются фирмой-изготовителем или конфигурацией ПК. По типу используемого носителя информации различают жесткие (HDD) и гибкие (FDD) магнитные диски. Жесткий диск или "винчестер", является стационарной частью системы и предназначен для долговременного хранения информации в ПК, в то время как гибкие диски или дискеты специально предназначены для переноса информации с одного компьютера на другой или для создания архивных копий.

Современные дискеты представляют собой магнитный диск из эластичного пружинящего материала или алюминиевой пластины размером соответственно 5,25 или 3,5 дюйма. Обе стороны диска покрыты тонким слоем ферромагнитного материала, а сам диск размещается в специальном защитном конверте. В накопителе на жестком диске корпус полностью герметичен и защищает поверхность диска от любого проникновения на нее даже мельчайших частичек пыли. Все механизмы дисководов – диски, головки и двигатели также находятся внутри этого корпуса. Диск представляет собой круглую алюминиевую пластину с идеально ровной поверхностью. Обе стороны пластины покрыты тонким слоем магнитного носителя. Существуют две разновидности покрытия: оксидное и металлизированное. Оксидное – представляет собой ровный слой оксида железа, который в виде ферролака наносится на рабочую поверхность диска. Наличие связывающего вещества (лака) делает такое покрытие достаточно толстым с низкой упаковкой ферромагнитного материала.

Металлизированное покрытие формируют нанесением чистого на алюминиевую основу методом напыления. В современных моделях накопителей применяют диски покрытые слоем кобальта толщиной около десяти микрон. Благодаря металлизированным поверхностям диски с напыленным покрытием более долговечны, обладают большей информационной емкостью и обеспечивают более надежное хранение информации. Накопители с оксидным покрытием, напротив, имеют тенденцию к выкрашиванию ферромагнитных частичек в процессе работы.

Дисководы жестких дисков имеют скорость вращения круглых пластин около 3600 об/мин, в результате этого головки парят над его поверхностью благодаря воздушной подушке, образованной быстрым вращением диска. Головки медленно набирают высоту при включении дисководов и совершают мягкую посадку при его выключении.

Для увеличения информационной емкости дисководов на жестких дисках имеют не один, а два и более дисков (пластин), которые смонтированы на единой оси (шпинделе) и вращаются совместно. Каждая сторона пластины имеет свою магнитную головку, которые смонтированы на единой каретке в форме гребешка и передвигаются синхронно. Пластины, стороны и головки нумеруются в порядке возрастания, начиная с нуля.

Физически ровная поверхность диска (как жесткого, так и гибкого) логически разбивается на сектора и дорожки. Каждая из концентрических окружностей представляет собой дорожку, а каждая дорожка делится на равные сегменты, называемых секторами. Совокупность одноименных дорожек на всех пластинах дисководов образует цилиндр.

При первом после включения компьютера обращении к диску головки записи/чтения движутся от внешнего края диска к центру, т.е. в порядке возрастания номера дорожки. Этим движением управляют позиционеры с шаговым двигателем или соленоидальные позиционеры с линейным двигателем. Позиционеры с шаговым двигателем преобразуют угловые шаги вращения двигателя в линейные перемещения головки с помощью шаговой ленты. Каждый шаг механизма перемещает головку на одну дорожку.

Соленоидальные позиционеры с линейным двигателем используют соленоид (электромагнит с движущимся якорем) для "втягивания" головок к центру диска. Головки смонтированы на пружинном шарнире, с помощью которого осуществляется возвращение головок в исходное положение. Позиционирование головок на поверхности диска осуществляется

изменением величины магнитного поля внутри соленоида. Соленоидальная система обеспечивает "доставку" головок к дорожке с любым номером за один такт работы. Это свойство позволяет увеличить скорость позиционирования по сравнению с шаговым двигателем в два и более раз.

Дисководы с соленоидальным позиционированием сложнее и дороже дисководов с шаговым двигателем. Температурные нарушения в процессе позиционирования в шаговых системах существенно меньше. Соленоидальные системы, напротив, имеют тенденцию к перелету или к недолету и нуждаются в постоянных подстройках. Другими словами, если для шаговых систем нет необходимости в петле обратной связи, то для соленоидальной системы обратная связь – необходимый ее атрибут.

3.2. Методы кодирования информации на магнитных дисках

К настоящему времени для записи информации на магнитный диск наиболее широкое применение получили четыре метода кодирования информации: FM, MFM, M²FM и RLL – методы. Все они взаимнонесовместимы по кодировке, но полностью совпадают по методу записи, который называется без возвращения к нулю с инверсией (NRZ-запись). Запись по методу NRZ осуществляется путем изменения направления тока (магнитного потока) в обмотке записывающей головки. Изменения направления магнитного потока в сердечнике головки приводит к перемагничиванию областей ферромагнитного покрытия диска в процессе его вращения и перемещения магнитной головки от дорожки к дорожке.

При считывании эти намагниченные участки вызывают перемены направления магнитного потока в считывающей головке, что соответствует логической "1", а отсутствие такой перемены на протяжении фиксированного временного интервала рассматривается как логический "0". В этом случае направление перемены полярности магнитного потока не принципиально, важен сам факт изменения полярности.

Перечисленные выше методы кодирования информации не связаны с направлением изменения полярности магнитного потока, а только задают их распределение во времени в процессах записи или считывания. Именно эти временные интервалы и являются определяющим для каждого из методов кодирования. Одним из первых методов кодирования информации является частотная модуляция (FM-кодирование), которую относят к кодированию с единичной плотностью. В начале каждого битового элемента на поверхности диска записывается бит синхронизации, а в промежутках между ними биты данных. Где под битовым элементом понимают минимальный временной интервал между соседними битами данных. В случае гибкого магнитного диска диаметром 5,25 дюйма битовый элемент имеет длительность 8 мкс, следовательно, бит данных должен быть записан через 4 мкс после бита синхронизации. Постоянство частоты следования синхроимпульсов существенно упрощает процедуры кодирования и декодирования информации (рис. 3.1).

Простота схем кодирования/декодирования одно из основных достоинств FM-метода, а наличие битов синхронизации – основной его недостаток. Это связано с малоэффективным использованием поверхности диска, поскольку половина его рабочей поверхности отводится под биты синхронизации. Тем не менее метод FM-кодирования достаточно широко применялся в ранних моделях ПК и позволял обеспечивать максимальную информационную емкость гибкого диска до 500 Кбайт.

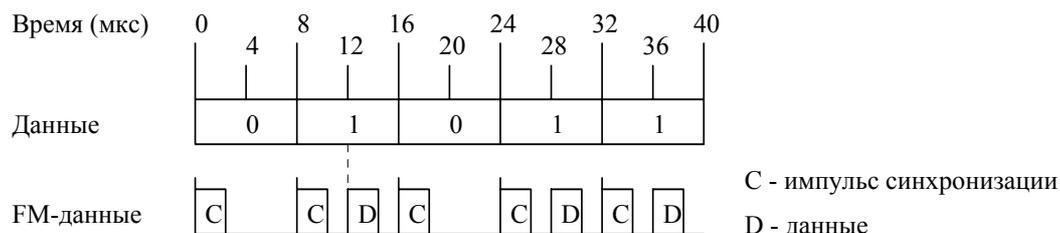


Рис. 3.1. Частотная модуляция

Модифицированная частотная модуляция (MFM-кодирование) позволяет увеличить продольную плотность записи вдвое, поскольку в местах расположения битов синхронизации находятся биты данных. Тем не менее, биты синхронизации все-таки существуют, но записываются в тех случаях, когда подряд следуют два и более логических "0". Бит синхронизации располагается вначале битового элемента (его длительность 4 мкс), а бит данных – в его середине, т.е. спустя 2 мкс (рис. 3.2). Таким образом, одновременно происходит как удвоение информационной емкости диска, так и удвоение скорости передачи данных. Это связано с тем фактом, что в битовом элементе невозможно одновременное размещение битов данных и синхронизации.

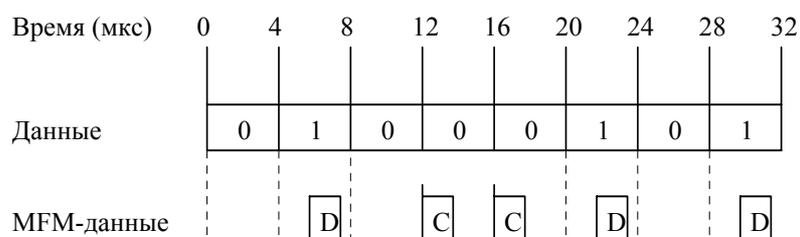


Рис. 3.2. Модифицированная частотная модуляция

Аналогичным образом кодируется информация методом миллеровской модифицированной частотной модуляции (M²FM-кодирование). Отличие заключается в том, что при M²FM-кодировании синхроимпульс записывается лишь в том случае, если в предыдущем битовом элементе не было изменения полярности магнитного потока и в текущем элементе не предполагается записывать бит данных.

При MFM и M²FM методах кодирования на схемы контроллера накладываются достаточно жесткие условия на синхронизацию битовых элементов. Необходимо непрерывно анализировать поток данных и с помощью фазового контура автоподстройки частоты обеспечивать сопряжение синхрогенератора контроллера с потоком данных. Данное обстоятельство и явилось одной из причин практически полного вытеснения M²FM-метода, в котором на синхронизацию накладывались самые жесткие условия. MFM-метод кодирования оказался более предпочтительным и нашел достаточно широкое практическое применение в накопителях на гибких магнитных дисках.

Одним из способов дальнейшего повышения максимальной информационной емкости диска явилась разработка метода записи с групповым кодированием (RLL-кодирование). Этот метод был заимствован из метода цифровой записи на магнитную ленту и основан на следующем принципе: каждый байт информации разделяется на два полубайта. Каждый полубайт кодируется специальным 5-разрядным двоичным кодом. Особенность этого кода заключается в том, что любое двоичное 5-разрядное число не должно иметь больше двух рядом стоящих нулей. Из 32 кодовых комбинаций этому условию отвечает 16, что является необходимым и достаточным условием кодировки полубайта (таблица 3.1).

Применение RLL-кодирования позволяет снизить длительность битового элемента с 4 мкс до 2,6 мкс, т.е. в худшем случае, когда рядом стоят два нуля, в любом временном интервале длительностью 7,8 мкс будет по крайней мере одно изменение направления магнитного потока. Поскольку этот максимальный интервал предсказуем, импульсы данных можно использовать для самосинхронизации системы. Именно поэтому системы, использующие RLL-кодирование, называют самосинхронизируемыми и самотазируемыми.

В последние годы в ряде накопителей на магнитных дисках используют ARLL-метод кодирования, который логически совместим с RLL-методом, но отличается от него увеличенной в 1,5 раза физической частотой обмена данными.

Все приведенные выше временные соотношения относятся к накопителям на гибких магнитных дисках (FDD). Что касается накопителей на жестких магнитных дисках (HDD), то единственное их отличие от FDD заключается в увеличении плотности данных на носителе, а следовательно, и в увеличении частоты обмена информацией примерно в 10 раз. Также отметим, что FM-кодирование в HDD не используется из-за малой линейной плотности записи.

Таблица 3.1. Один из вариантов RLL-кодирования

Шестнадцатиричное число	Двоичный полубайт	Кодовая группа RLL
0	0000	11001
1	0001	11011
2	0010	10010
3	0011	10011
4	0100	11101
5	0101	10101
6	0110	10110
7	0111	10111
8	1000	11010
9	1001	01001
A	1010	01010
B	1011	01011
C	1100	11110
D	1101	01101
E	1110	01110
F	1111	01111

3.3. Форматирование накопителей на гибких магнитных дисках

В настоящее время в персональных компьютерах наибольшее распространение получили 5- и 3-дюймовые накопители на гибких дисках. Для каждого из них информация хранится на съемных магнитных дисках – дискетах, соответственно имеющих размеры 5,25 и 3,5 дюйма. Конструктивно дискета состоит из магнитного диска и защитного конверта в форме квадрата. Центральное отверстие диска предназначено для его посадки на шпиндель механизма вращения, а отверстие, расположенное недалеко от центра, называется индексным. Индексное отверстие предназначено для форматирования сигнала начала оборота диска, который анализируется контроллером.

Напомним, что информация на диске хранится побитно на концентрических окружностях, называемых дорожками. Дорожки нумеруются от края к центру диска, начиная с нуля. Отметим, что накопители на гибких дисках работают только с одним диском и применительно к ним термин "цилиндр" практически не используется. Стандартное количество дорожек на 5-дюймовой дискете 40 или 80, что обеспечивает максимальную информационную емкость соответственно 360 Кбайт или 1,2 Мбайта. Непропорциональность в изменении информационной емкости связана с существующей переменной плотности записи на диск. Трехдюймовые дискеты всегда имеют 80 дорожек, что определяет их информационную емкость до 1.6 Мбайт.

Каждая дорожка на дискете разбита на равные участки, называемые секторами. Сектора внутри дорожки нумеруются не с нуля, а с единицы, начиная от индексного отверстия. Стандартная информационная емкость сектора – 512 байт. Разбивка на сектора и дорожки выполняется в процессе форматирования дискеты командой FORMAT. Процедура форматирования заключается в том, что в определенные места каждой дорожки записываются служебные последовательности символов. Это необходимо для однозначного определения позиции головки на дорожке в циклах записи/чтения или при выполнении операций поиска. Записывать на диск номер дорожки в принципе нет необходимости, поскольку управление шаговым двигателем осуществляет аппаратура адаптера дисководов, т.е. контроллер сам устанавливает головки на заданную дорожку.

Все операции записи данных на диск сопровождаются накоплением и записью в конце поля данных двухбайтового числа, называемого контрольной суммой. Алгоритм его определения, в принципе, строго не регламентирован, но, как правило, связан с расчетом полинома по известным весовым коэффициентам.

Все перечисленные выше процедуры по форматированию дискеты, организации ее логической структуры, а также записи/считывания информации выполняет схема адаптера FDD. На рисунке 3.3 приведена его обобщенная блок-схема. В ее состав входят регистры: управления (РУ), основного состояния (РОС) и данных (РД), которые программно доступны процессору.

Регистр РОС доступен процессору только для операции считывания и предназначен для организации взаимодействия между адаптером НГМД и ЦП. С его помощью адаптер информирует процессор о текущем состоянии накопителя и направлении передачи информации. Регистр управления выполняет функции включения/выключения моторов, выбирает один из четырех FDD, осуществляет общий сброс адаптера и доступен процессору только для записи. Регистр данных состоит из нескольких стековых регистров и предназначен для обмена информацией между адаптером и процессором. С его помощью запоминаются данные, команды, параметры и передается информация о состоянии FDD. Данные считываются или записываются в РД последовательно для программирования адаптера или для получения результата после выполнения отдельной команды. Результаты выполнения команд находятся в четырех регистрах состояния ST0 – ST3. В цикле обмена коды команд и коды результатов состояния должны быть расположены в строго определенном порядке без каких-либо сокращений. После получения последнего в цепочке команд кода адаптер FDD переходит в режим ее выполнения, после которого формируются байты результата. Стандартное количество команд для адаптера FDD равно пятнадцати.

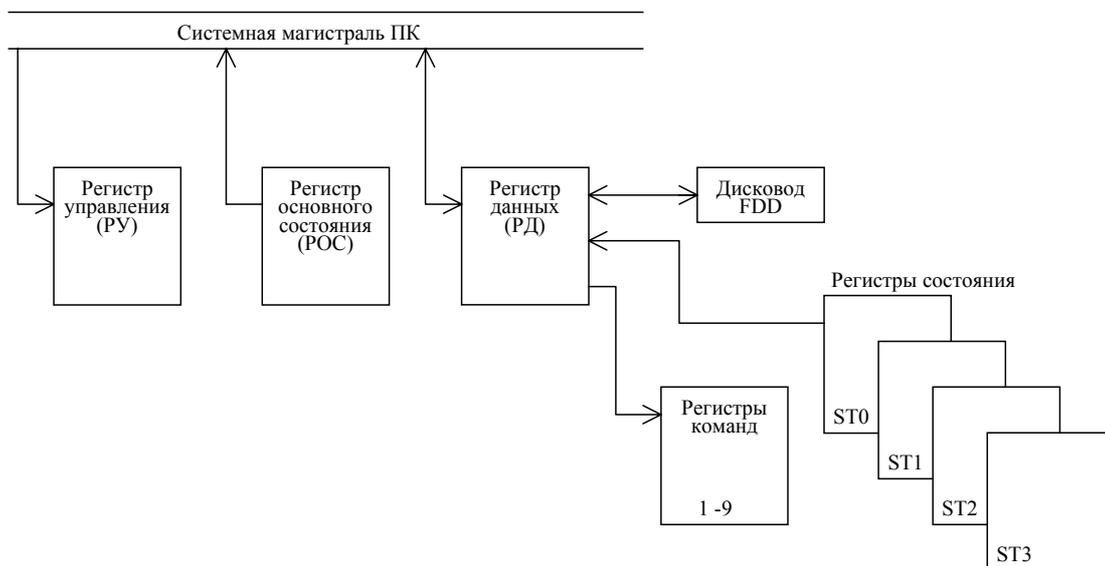


Рис. 3.3. Логическая блок-схема адаптера FDD

Как отмечалось ранее, перед первым использованием дискеты она должна быть отформатирована. Формат – это служебная информация, которая недоступна большинству прикладных программ, и используется исключительно адаптером для размещения, поиска и контроля информации на дискете. На рисунке 3.4 (а, б) представлены формат дорожки и структура сектора соответственно.

Индекс	Зазор индекса	Идентификатор сектора	1-ый блок данных	3		N-ый блок данных	3	Зазор дорожки	Индекс
--------	---------------	-----------------------	------------------	---	--	------------------	---	---------------	--------

а)

Идентификатор сектора		Зазор идентификатора	Блок данных		
Маркер идентификатора	Идентификатор адреса		Маркер данных	Поле данных	Код циклического контроля

б)

Рис. 3.4. Формат дорожки (а) и структура сектора (б)

Начало и конец каждой форматируемой дорожки определяется индексом. За ним через специальный зазор, равный 32 байтам кода 4EH, следует идентификатор сектора. Идентификатор сектора состоит из двух областей, первая из которых, называемая маркером идентификатора, содержит 12 байт кода 00H, три байта кода A1H и байт кода FEH. Идентификатор адреса состоит из шести байтов: байта номера дорожки; байта стороны дискеты; байта номера сектора; байта длины сектора и двух байт кода циклического контроля. Зазор идентификатора содержит 22 байта кода 4EH. Маркер данных по структуре аналогичен маркеру идентификатора и отличается от него кодом FBH (или F8H) в последнем байте. Код FBH означает, что данные верны и все поле данных может быть прочитано. Код F8H указывает, что данные стерты. Поле данных для операционной системы MS-DOS содержит 2 байта кода циклического контроля. Зазор блока данных содержит 80 байтов кода 4EH. Приведенные величины зазоров соответствуют форматам S-8 и D-8, для других же форматов они могут отличаться от указанных.

3.4. Логическая организация данных на дискете

После форматирования, т.е. придания диску его окончательной структуры на диске формируется несколько логических областей, количество и формат которых определяет операционная система компьютера. Так, например, в наиболее широко распространенной системе MS-DOS предусмотрены четыре логические области:

- 1.) загрузочный сектор;
- 2.) таблица размещения файлов;
- 3.) каталог;
- 4.) область данных.

Первая из перечисленных областей содержит версию DOS и короткую программу начальной загрузки операционной системы в компьютер. Независимо от формата записи эта программа всегда должна занимать первый сектор на нулевой стороне нулевой дорожки. В следующем за ним секторах находится таблица размещения файлов (FAT-таблица), в которой отражено расположение записанных на дискету файлов. В связи с важностью сохранности FAT-таблицы на дискете всегда существуют две ее копии. Непосредственно за FAT-таблицей находятся каталог и данные.

Размер области данных, каталога и FAT-таблицы зависит от количества секторов на дискете и используемой операционной системы. В MS-DOS длина сектора 512 байт, а количество их определяется выбором формата из следующего набора: формат S-8 (320 секторов на дискете); D-8 (640 секторов); S-9 (360 секторов); D-9 (720 секторов); DD-9 (1440 секторов); QD-15 (2400 секторов) и QD-18 (2880 секторов). В таблице 3.2 приведено размещение логических областей дискеты на примере формата D-9.

Таблица 3.2.

Номер стороны	Номер дорожки	Номер сектора	Логическая область дискеты
0	0	1	Загрузочный сектор
0	0	2-3	FAT (первая копия)
0	0	4-5	FAT (вторая копия)
0 1	0 0	6-9 1-3	Каталог
1 0.1	0 1-39	4-9 1-9	Область данных

Минимальная часть диска, резервируемая для одного файла, называется кластером. Его длина всегда кратна размерности сектора. Если размерность кластера равна 1 Кбайт, то под файл размерностью даже один байт будет отведено 2 сектора по 512 байт. Каждому кластеру присваивается последующий номер, начиная с двух в порядке возрастания секторов. В результате

номера секторов и кластеров на диске жестко связаны. Номер последнего кластера зависит от размерности дискеты, но, как правило, всегда превышает 256.

При записи на диск файла, превышающего размер одного кластера, фрагменты файла могут быть "разбросаны" по всему диску. Тем не менее, в процессе считывания, особых проблем не возникает. Это вызвано тем фактом, что файл как единое целое собирается последовательно с помощью FAT-таблицы и каталога. Элемент FAT-таблицы имеет 12-битовый размер, в связи с чем анализировать его состояние необходимо одновременно в трех соседних байтах. Все элементы FAT-таблицы нумеруются в порядке возрастания, начиная с нуля. Нулевой и первый элементы являются идентификаторами формата, а остальные определяют статус кластера. Элементы FAT-таблицы, определяющие статус кластера, могут принимать следующие значения:

000 – кластер свободен;

002-FF0 – кластер занят;

FF1-FF7 – кластер поврежден;

FFF – кластер занят и является последним кластером файла.

На основе FAT-таблицы определяется размещение файла на диске. Номер стартового элемента FAT-таблицы для искомого файла определяется содержимым каталога. Содержимое этого элемента определяет размещение второй части файла, т.е. номер следующего элемента FAT-таблицы, в котором указано расположение третьей части файла и т.д. Окончание файла соответствует коду FFF FAT-таблицы. В таблице 3.3 приведен пример определения расположения файла на диске с помощью FAT-таблицы.

Таблица 3.3.

№ элемента FAT	Содержимое	№ кластера на диске	Очередность частей файла
2	008	3	1
3	007	4	5
4	FFF	5	7
5	006	6	3
6	003	7	4
7	004	8	6
8	005	9	2

Таким образом, для определения расположения файла на диске необходимо иметь номер элемента FAT начала файла, который находится в каталоге. Рассмотрим структуру каталога MS-DOS. Каждый файл каталога описывается с помощью строки, состоящей из 32 битов. Структура единицы каталога, т.е. одной его строки представлена в таблице 3.4.

Таблица 3.4.

Количество байт	Описание
8	Имя файла
3	Тип файла
1	Атрибут
10	Резервное поле
2	Время создания файла
2	Дата создания файла
2	Номер начального кластера файла
4	Длина файла

Имя файла и его тип (или расширение) записывается в ASCII-кодах. При этом под имя файла отводится 8 байт, а под расширение – 3 байта. При записи имени диска эти байты объединяются, образуя последовательность из 11 символов. Неиспользованные байты

заполняются символом пробела. При стирании файла в первый байт его имени заносится код E5H. Байт атрибута файла может принимать следующие значения:

- 01 – файл только для чтения;
- 02 – скрытый файл;
- 04 – системный файл;
- 08 – имя диска;
- 10 – подкаталог;
- 20 – архив.

Атрибут "файл только для чтения" позволяет открывать и читать содержимое файлов, но полностью запрещает его модификацию. Скрытые или системные файлы недоступны для вводимых с клавиатуры команд операционной системы (например, для команды DIR). Подкаталоги функционируют на правах обычных файлов. Атрибут архива используется при создании резервных копий (например, файлов с расширением .BAK). DOS устанавливает этот атрибут при записи в файл. Программы создания резервных копий в процессе своей работы находят файлы с установленным атрибутом "архив", делают копии и сбрасывают атрибут. Таким образом, при создании резервной копии файл копируется, если он был модифицирован в промежутках между копированиями.

Каждый из двух байт, описывающих дату и время создания файла, разбивается на три поля и имеет формат, представленный на рисунке 3.5.

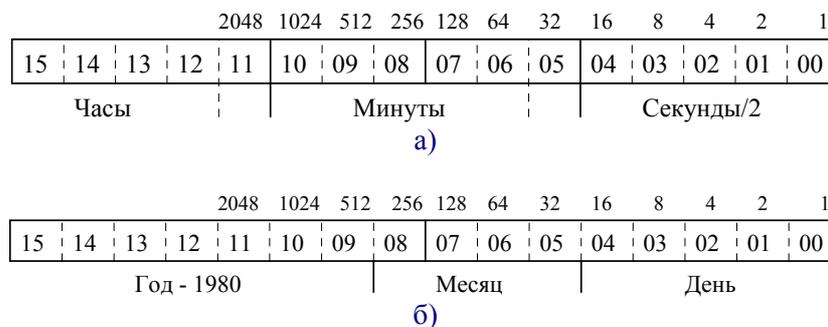


Рис. 3.5. Формат байтов, описывающих: а) время создания файла; б) дату создания файла

Таким образом, слова, описывающие время и дату создания файла могут быть рассчитаны по следующим формулам:

$$\begin{aligned} \text{Время} &= \text{Час} * 2048 + \text{Минуты} * 32 + \text{Секунды}, \\ &[3.1] \\ \text{Дата} &= (\text{Год} - 1980) * 512 + \text{Месяц} * 32 + \text{День}. \end{aligned}$$

Длина файла записывается в байтах, что ограничивает его объем количеством 2^{32} . В реальных системах DOS не позволяет создавать файлы размером более 32Мбайт. Поскольку каталог на дискете занимает 8 секторов (для формата D-9), то максимальное количество строк каталога может составлять 128. Информационная емкость области данных на дискете определяется типом формата и, в незначительной степени, заполненностью области каталога.

3.5. Общая характеристика накопителя на жестких дисках. Фактор чередования

В отличие от FDD накопителя на жестких магнитных дисках (HDD) имеют не один, а несколько магнитных дисков. Поэтому группу дорожек, находящихся на разных дисках и на разных сторонах одного диска, имеющих один и тот же номер, принято называть цилиндром. Цилиндры также как и дорожки нумеруются от края диска к его центру, начиная с нуля. Предел в 1024 цилиндра встроен в операционную систему на всех уровнях. Цилиндры с номером больше 1023 просто не могут обрабатываться ни DOS, ни BIOS, ни платой контроллера.

Разбивка дорожек по секторам, аналогична разбивке на сектора в FDD. Каждый сектор состоит из двух частей: поле идентификатора и поле данных. При этом информация, находящаяся в идентификаторе, записывается на диск, при форматировании низкого уровня.

Процесс форматирования жесткого диска состоит из двух стадий: форматирование низкого уровня и форматирование высокого уровня. При форматировании низкого уровня осуществляется запись последовательности кодов для определения начала каждого сектора. При форматировании высокого уровня на диск записывается рекорд загрузчика DOS, корневой каталог, таблицы размещения файлов (FAT) и системные файлы. Информационная емкость сектора такая же, как и у FDD и равна 512 байт.

Количество магнитных дисков и тип формата определяют максимальную информационную емкость HDD. При этом различают форматированную и неформатированную емкости. Неформатированная емкость определяет максимальное количество размещаемых на диске байт информации, независимо от их назначения. Форматированная емкость – это суммарное количество байт, отведенных только для записи данных. Форматированная емкость численно равна произведению размера сектора в байтах, количеству секторов на дорожке, количеству рабочих поверхностей (количеству магнитных головок), количеству цилиндров. При этом необходимо учитывать, что последний внутренний цилиндр используется исключительно для целей диагностики и для пользовательских программ недоступен.

Большинство современных контроллеров HDD имеют собственный буфер – область ОЗУ для временного хранения данных. Их емкость порядка 10 Кбайт, что позволяет хранить информацию одной дорожки магнитного диска. При наличии запроса на чтение любого сектора одной дорожки контроллер считывает всю дорожку, начиная с сектора, ближайшего к месторасположению магнитной головки. Это позволяет существенно ускорить обработку данных в пределах дорожки.

Обмен данными между ЦП и буфером накопителя может быть реализован двумя способами. В ПК IBM AT и PS/2 всю работу выполняет процессор, а в ПК IBM PC/XT специальные микросхемы прямого доступа к памяти (ПДП). Применение микросхем ПДП обусловлено низким быстродействием микропроцессора в ПК PC/XT. При работе с HDD операционная система резервирует в ОЗУ специальную область, называемую буфером DOS. В каждом буфере размещается один сектор диска, а количество буферов может быть определено пользователем. Как правило, компьютеры, имеющие жесткий диск, работают с 20 буферами. По мере чтения файла его сектора заполняют буфер и когда буфер полностью заполнен, то очередной сектор записывается в ту область буфера, к которой программа обращалась первой и содержимое которой уже обработано прикладной программой.

При записи данных на диск происходит обратное: прикладная программа помещает данные в буфер контроллера жесткого диска; затем контроллер начинает запись, указывая сектор дорожку, номер и сторону магнитного диска.

Скорость вращения жесткого диска составляет 3600 об/мин, что является достаточно высокой для процессов записи/считывания. В связи с этим скорость, с которой данные проходят под магнитной головкой, не обязательно совпадает со скоростью, с которой компьютер способен считывать или записывать данные на диск. Как показывают расчеты, под головкой за секунду проходит 5 млн. бит информации. Именно поэтому большинству дисководов приписывается скорость передачи данных в 5 Мбит.

Большинство компьютеров на базе микропроцессоров 8086, 8088 и 80288 не в состоянии обрабатывать информацию с такой скоростью. В результате ряд секторов “пролетает” мимо магнитной головки непрочитанными. Так, например, в IBM XT над головкой пройдут 5 секторов, пока контроллер будет готов читать (записывать) данные дальше. Что касается ПК IBM AT и PS/2, то для них контроллер готов к обработке следующего сектора через один или два сектора. Таким образом, если данные одного файла занимают рядом расположенные сектора, то чтение второго из них будет возможно только при следующем обороте диска. В результате будет потеряно около 11 мс времени. Для дисков с четырьмя и более дорожками на цилиндр эти потери могут достигать нескольких долей секунды на цилиндр.

Для устранения этого недостатка введен специальный параметр, называемый фактором чередования на диске. С помощью него задается не физическая, а логическая непрерывность секторов. Так, например, стандартный компьютер PC XT имеет фактор чередования 6:1 или

просто 6. Это означает, что логическая нумерация секторов будет продолжена в порядке возрастания через 5 физических секторов, т.е. на каждом шестом физическом секторе (рис. 3.6). Следовательно, для последовательного считывания всех секторов дорожки (цилиндра) необходимо шесть оборотов диска. В компьютерах IBM AT фактор чередования уменьшен до трех, а в диске, где этот фактор равен единице, все данные с дорожки будут считаны за один оборот диска. Очевидно, что чередование 1:1 является наиболее оптимальным и применяется во всех FDD.

√ Индекс		Физический сектор														
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	4	7	10	13	16	2	5	8	11	14	17	3	6	9	12	15
		Логический сектор														

Рис. 3.6. Соответствие логических номеров секторов физическим секторам для фактора чередования, равного 6

Фактор чередования устанавливается в процессе форматирования низкого уровня, путем записи на диск соответствующих идентифицирующих чисел. Чередование может быть изменено в процессе реформатирования с другим фактором, при котором секторы будут пронумерованы по-другому.

3.6. Логическая организация данных на жестком диске

Логическая организация диска отличается от аналогичной организации дискет. Первое отличие связано с совершенно другим представлением служебной информации, которая на дорожке строго детерминирована и называется форматом. Поскольку жесткие диски не предназначены для обмена между компьютерами, то форматы записи у различных ПК могут быть несовместимы. Здесь важно обеспечить полное "понимание" адаптером той информации, которая находится на магнитном диске.

Один из широко распространенных форматов приведен в таблице 3.5.

Таблица 3.5.

Мнемоника	Содержание	Количество байт
АМ	Адресный маркер	4
ИНТ 1	Интервал, заполняемый 00H	9 - 12
СИНХ 1	Синхробайт (01H)	1
ИНТ 2	Интервал (00H)	2
СРВ	Байт для сравнения (С2H)	1
ЦИЛС	Старший байт номера цилиндра	1
ЦИЛМ	Младший байт номера цилиндра	1
ГЛВ	Номер головки	1
СЕК	Номер сектора	1
ФЛАГ	Байт флажков	1
НУЛ	Байт нулей	1
ЕСС	Код обнаружения и коррекции ошибок	4
ИНТ 3	Интервал, заполняемый нулями	16
СИНХ 2	Синхробайт (01H)	1
ИНТ 4	Интервал, заполняемый нулями	2
ДААННЫЕ	Поле данных	512
ЕСС	Код обнаружения и коррекции ошибок	4
ИНТ 5	Межсекторный интервал	43

Начало каждого сектора обозначается адресным маркером, который представляет собой кодовую комбинацию, не входящую во множество кодовых комбинаций для используемого метода кодирования информации в HDD. Обычно это байт А1Н с соответствующим ему синхробайтом 0АН.

Поле идентификатора (И-поле) содержит служебную информацию, необходимую для безошибочного доступа к данным, а именно, номер цилиндра, головки, сектора и т.д. За И-полем следует поле данных (Д-поле). Каждое из полей имеет собственные контрольные суммы ЕСС-кода (Error Correction Code). Используемый для коррекции образующий полином может иметь разные корректирующие возможности, и, соответственно, у разных фирм виды полиномов отличаются друг от друга.

Интервалы ИНТ *i* обеспечивают время для подготовки адаптера к следующим процедурам:

ИНТ 1 – запись/чтение данного сектора после обнаружения адресного маркера (АМ);

ИНТ 2 – чтение информации И-поля;

ИНТ 3 – анализ правильности считывания И-поля и запись/считывание Д-поля;

ИНТ 4 – запись/считывание информации Д-поля;

ИНТ 5 – межсекторный интервал, подготовка адаптера к обнаружению следующего адресного маркера.

Байты СИНХ 1 и СИНХ 2 определяют начало соответственно И- и Д-полей и служат для синхронизации счетчиков битов в адаптере.

Байт СРВ имеет постоянное значение С2Н для всех секторов и используется для проверки правильного чтения И-поля.

Совокупность байтов ЦИЛС, ЦИЛМ, ГЛВ и СЕК (см. таблицу 3.5) определяют адрес сектора.

Биты байта ФЛАГ содержат следующую служебную информацию: бит 4 определяет последний физический сектор на дорожке (D4 = 1) или все остальные сектора (D4 = 0); бит 2 является меткой дефективного сектора (D2 = 1).

В ряде случаев после интервала ИНТ 3 также следует адресный маркер, формат которого аналогичен формату адресного маркера сектора. В этом случае после АМ следует байт F8H-маркер поля данных.

Описанная выше структура дорожки формируется в процессе форматирования низкого уровня или физического форматирования. Однако этого недостаточно для нормальной работы HDD. Необходимо еще создать загрузочный сектор, две копии FAT-таблицы и каталог. Данные логические области создаются в процессе форматирования высокого уровня.

В результате форматирования высокого уровня самый первый сектор HDD содержит главную корневую запись, которая после включения компьютера загружается в память под управлением BIOS и выполняется. В последней части этого сектора расположена 4-элементная таблица разделов, каждый элемент которой содержит 16-байтовую запись (элемент раздела).

Во время загрузки BIOS загружает в память главную корневую запись и передает ей управление. Она считывает таблицу разделов с целью нахождения активного, считывает корректный загрузочный сектор и выполняется. Структура главной корневой записи и элемента раздела приведена на рисунке 3.7 а) и б) соответственно.

Длина	Содержимое
1ВЕН	Код загрузки и выполнение загрузочного сектора активного раздела
10Н	Элемент раздела 1
10Н	Элемент раздела 2
10Н	Элемент раздела 3
10Н	Элемент раздела 4
2Н	55ААН-сигнатура

а)

Длина	Содержимое
1	Флаг загрузки: 0 - не активен; 80Н - активен
1	Номер головки начала раздела
2	Номер сектора и цилиндра загрузочного сектора начала раздела
1	Код операционной системы
1	Конец раздела: номер головки
2	Конец раздела: номер сектора и цилиндра последнего сектора раздела
4	Относительный номер начального сектора
4	Количество секторов

б)

Рис. 3.7. Структура: а) главной корневой записи; б) элемента раздела

Код операционной системы может принимать следующие значения:

- 00H – операционная система не определена;
- 01H – базовая версия DOS с 12-разрядной FAT;
- 04H – базовая версия DOS с 16-разрядной FAT;
- 05H – расширенная версия DOS.

Значения цилиндра и сектора занимают 10 и 6 бит соответственно. Их размещение в соответствующей ячейке элемента раздела показано на рисунке 3.8.

Значение относительного номера начала сектора в каждом разделе эквивалентно головке, сектору и цилиндру начального адреса раздела. При этом относительный номер сектора 0 совпадает с цилиндром 0, головкой 0 и сектором 0 на магнитном диске. Относительный номер сектора инкриминируется сначала по каждому сектору на головке, затем по каждой головке и далее по каждому цилиндру. Т.е. для его нахождения вполне приемлема формула:

$$\text{Относительный сектор} = (N_{\text{ц}} * C / \text{Ц} * \text{МГ}) + (N_{\text{г}} * C / \text{Ц}) + (N_{\text{с}} - 1), \quad [3.2]$$

где: $N_{\text{ц}}$ – номер цилиндра; $N_{\text{г}}$ – номер головки; $N_{\text{с}}$ – номер сектора; C – количество секторов на дорожке; Ц – количество цилиндров; МГ – количество магнитных головок.

Байт 2								Байт 1							Байты
15	14	13	12	11	10	09	08	07	06	05	04	03	02	01	
Цилиндр								Сектор							Содержимое

Рис. 3.8. Размещение номеров цилиндра и сектора в двухбайтовой ячейке элемента раздела

Дальнейшая логическая организация диска аналогична логической организации дискеты, т.е. содержит загрузочный сектор, две копии FAT-таблицы, каталог и данные. Отличие заключается в размерности элементов FAT-таблицы, поскольку для жесткого диска их размерность составляет 16 бит. При этом многие считают, что 16-битовая FAT-таблица не позволяет DOS работать с логическими дисками емкостью более 32 Мбайт. На самом деле эти ограничения вызваны тем, что прерывания INT 25/26H не позволяют работать с секторами, номера которых превышают 65535. Однако, если адаптер HDD и соответствующий ему драйвер позволяет поддерживать емкость сектора более 512 байт, то в этом случае ограничения в 32 Мбайт автоматически устраняются.

При 16-битовой организации FAT ее первый байт называют дескриптором носителя или FAT ID байтом. Следующие 7 байтов содержат код 0FFH, за которым следуют элементы определения статуса кластера. Заполнение этих элементов аналогично построению FAT-таблицы для дискет.

Кроме основного раздела DOS, на диске могут быть размещены так называемые расширенные разделы DOS. Их создание оправдано во всех случаях, когда емкость HDD больше 20 Мбайт. Пример размещения расширенных разделов DOS на диске приведен на рисунке 3.9.

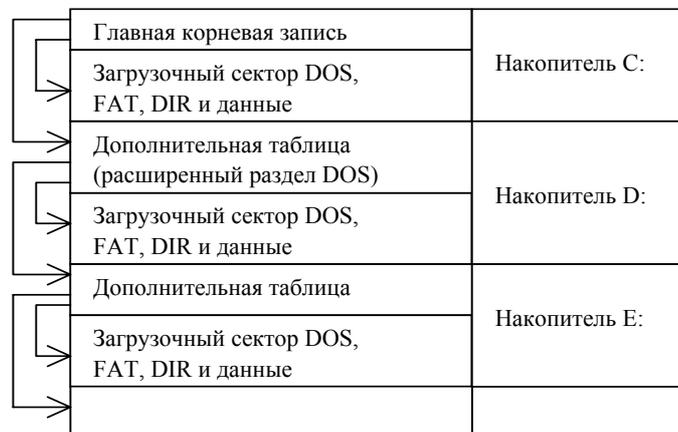


Рис. 3.9. Размещение на диске дополнительных разделов DOS

В случае размещения расширенных разделов DOS на магнитном диске создаются встроенные дисковые области с дополнительными загрузочными секторами DOS, таблицами FAT, каталога и данными. Т.е. на диске создаются логические накопители на жестких дисках, обозначаемые заглавными буквами латинского алфавита, начиная с буквы С. В этом случае каждая дополнительная таблица разделов содержит код системы 05Н.

Структура элементов дополнительной таблицы разделов подобна структуре элементов разделов главной корневой записи за исключением того, что загрузочный сектор не обязан содержать код загрузчика.

3.7. Форматирование жесткого диска

Форматирование низкого уровня следует выполнять в крайних случаях, когда появились ошибки на нулевом секторе или когда необходимо установить совершенно новую операционную систему. При логическом форматировании высокого уровня происходит разбиение диска на разделы, т.е. на несколько логических дисков. Под логическим диском понимают некоторую непрерывную область магнитного диска, которая работает как отдельный накопитель.

Существующие программы форматирования низкого уровня запрашивают фактор чередования. Практика показывает, что ПК класса XT работают с чередованием 6:1, класса AT – с чередованием 2:1, а ПК, имеющие в своем составе микропроцессор INTEL 80386 или INTEL 486 – с чередованием 1:1. Современные программы форматирования позволяют выбрать оптимальный фактор чередования для данной вычислительной системы и позволяют указать начальный и конечный цилиндр области форматирования. Примером может служить программа HFORMAT фирмы Kolod Reserch Inc. После форматирования низкого уровня выполняются процедуры тестирования отформатированных областей с целью проверки целостности созданных секторов.

Форматирование высокого уровня можно выполнить командой FDISK, которая включена в операционную систему MS-DOS. С ее помощью происходит разбиение диска на разделы и устанавливается определенная версия DOS на жесткий диск. Программа FDISK позволяет выполнять следующие функции: создание главной корневой записи в первом секторе диска; деление диска на разделы; создание и смена активного раздела. При назначении какого-либо раздела активным программа FDISK помечает его как "самозагруженный", поместив значение 80Н во флаг загрузки соответствующего раздела и одновременно сбросив флаг во всех остальных разделах. Однако необходимо помнить, что активный раздел должен содержать загрузочный сектор и копию операционной системы.

Расширенный раздел DOS идентифицируется байтом кода системы 05Н в таблице разделов главной корневой записи. Этот раздел не является загруженным и не может быть активным.

3.8. Контроллеры накопителей на гибких магнитных дисках

Существующие контроллеры FDD по ряду причин являются достаточно консервативными устройствами и не претерпели радикальных изменений в своей структуре за последнее время. Радикальные изменения коснулись лишь элементной базы и степени интеграции используемых БИС. Логическая структура адаптера (контроллера) FDD представлена на рисунке 3.10.

Все представленные на рисунке 3.10 регистры доступны процессору и обслуживаются по командам прерывания BIOS INT 13Н. Методика вызова конкретных функций стандартна, т.е. перед командой INT 13Н номер функции загружается в регистр АН, а необходимые входные параметры – в регистры AL, CX, DX, ES:BX.

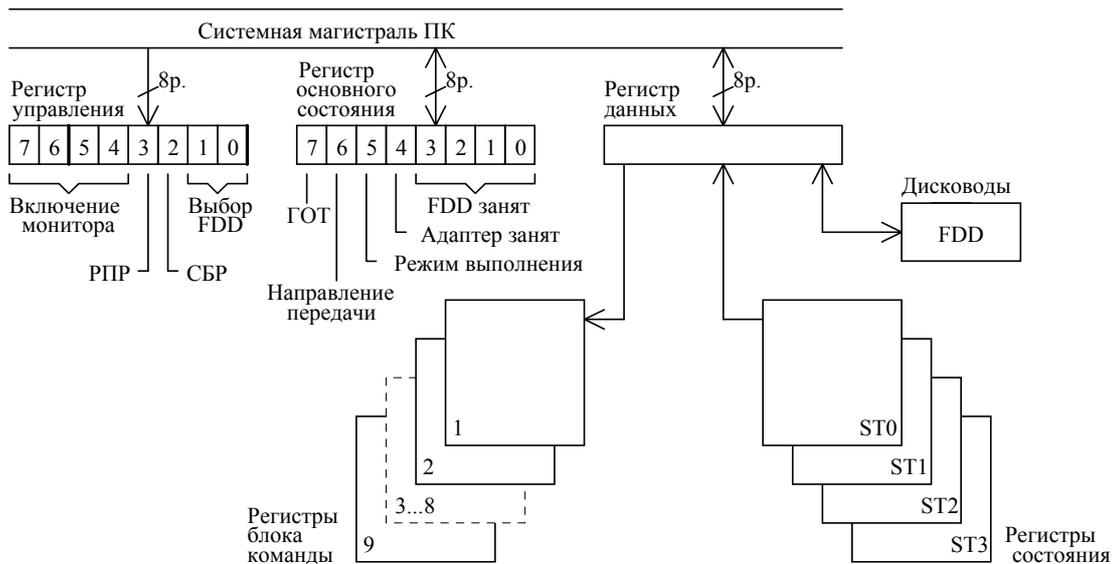


Рис. 3.10. Структурная схема адаптера FDD

Регистр основного состояния (РОС) работает только в режиме считывания и предназначен для организации обмена информацией между адаптером FDD и процессором. При этом установка любого из разрядов 0 – 3 данного регистра запрещает адаптеру FDD выполнять процедуры записи/считывания. Регистр управления выполняет функции включения/выключения моторов вращения дискетки, выбирает один из четырех FDD, осуществляет общий сброс адаптера и доступен процессору только для записи. Регистр данных выполняет функции буферного регистра для обмена информацией между FDD, регистрами блока команд и регистрами состояния. Через этот регистр осуществляется стековая передача байтов команды, состояния завершения команды и данных записи/чтения. Обмен информацией осуществляется мультибайтовой передачей между процессором и адаптером в виде последовательности трех фаз. В первой фазе (фазе команды) адаптер FDD получает от процессора необходимое число кодов команды и выполняет ее во второй фазе (фазе выполнения). Результаты выполнения команды находятся в регистрах состояния ST0 – ST3, которые доступны процессору в третьей фазе (фазе результата). В цикле обмена коды команд и коды результата должны быть расположены в строго определенном порядке без каких-либо сокращений. Адаптер FDD обеспечивает выполнение следующих 15 команд.

1. Чтение данных.
2. Чтение поля идентификатора сектора.
3. Чтение стертых данных.
4. Чтение дорожки.
5. Сканирование на "равно".
6. Сканирование на "меньше и равно".
7. Сканирование на "больше и равно".
8. Запись данных.
9. Запись стертых данных.
10. Форматировать дорожку.
11. Поиск.
12. Рекалибровка.
13. Указать.
14. Снять состояние прерывания.
15. Снять состояние FDD.

При выполнении команд 1, 3, 5 – 9 фаза команд состоит из последовательности 9 байтов, которые заполняются регистрами блока команды (РБК) и выполняются. Для выполнения остальных команд фаза команд состоит из последовательности 1 – 7 байтов, после получения которых адаптер FDD переходит в фазы выполнения и результата. Фаза результата, как правило, начинается с передачи информации от регистров ST0 – ST3 в микропроцессор через

двунаправленный регистр данных. При этом адаптер FDD формирует сигнал "запроса прерывания" в следующих случаях: в фазе результата при выполнении команд 1 – 12, а также при изменении состояния сигнала "готовность FDD". Возникающие в системе прерывания могут быть однозначно определены с помощью команды 14 – "снять состояние прерывания".

При работе под управлением BIOS драйвер FDD поддерживает шесть функций, пронумерованных от нуля до пяти, которые классифицируются следующим образом:

- 0 – сброс системы FDD;
- 1 – прочитать состояние;
- 2 – прочитать сектор;
- 3 – записать сектор;
- 4 – проверить сектор;
- 5 – разметка дорожки.

Все функции выполняются под контролем стандартной 11-байтовой базовой дисковой таблицы (таблица 3.6), адрес-указатель на которую находится в векторе прерываний IEN.

Таблица 3.6.

Байт	Содержание
0H	
1H	
2H	Байт указателя времени, в течении которого двигатель вращения дискеты должен быть выключён. Его длительность определяется содержимым байта делённым на 18.2с
3H	Байт длины поля данных сектора: 0 - 128; 1 - 256; 2 - 512; 3 - 1024
4H	Байт номера последнего сектора на дорожке
5H	Байт длины межсекторного промежутка (по умолчанию длина промежутка равна (42 байтам)
6H	Байт длины передаваемых данных при неопределённой длине сектора (по умолчанию - 0FF)
7H	Байт длины межсекторного промежутка во время форматирования (по умолчанию равна 80)
8H	Байт символа-заполнителя (его обычная конфигурация 0F6H)
9H	Байт времени позиционирования головки на дорожке (мс)
0AH	Байт времени разгона двигателя до нормальной скорости, которое надо умножить на 0.125с

Таким образом, с помощью данной таблицы производится установка необходимых параметров для нормальной работы механических частей дисководов при выполнении перечисленных выше функций. Сами функции предназначены для следующего:

а) **Функция 0** (сброс системы FDD) – приведение в исходное состояние контроллера FDD и дисководов. Данная функция не оказывает никакого влияния на дискеты, а служит для сброса дисковой системы и рекалибровки головки, если произошел сбой ее привода;

б) **Функция 1** (прочитать состояние) – подробное расшифрование ошибок схем контроллера, возникающих после сбоя. В этом случае необходимо прочитать байт состояния, в котором заложена расшифровка 11 ошибок;

в) **Функция 2** (прочитать сектор) – считывание информации в оперативную память из поля данных одного или нескольких секторов. Дискета должна быть предварительно отформатирована, а список параметров представлен следующими регистрами процессора:

- AH = 2 – номер функции;
- DL = номеру дисковода;
- DH = номеру головки;

CH = номеру дорожки;
 CL = номеру сектора;
 AL = количеству записываемых секторов;
 ES:BX = адрес буфера памяти;

г) Функция 3 (записать сектор) – полностью симметрична предыдущей функции, с той лишь разницей, что AH = 3;

д) Функция 4 (проверить сектор) – проверка возможности поиска нужного сектора на дорожке, а также считывания информации из поля данных с целью накопления контрольной суммы и сравнения ее с записанной;

е) Функция 5 (разметка дорожки) – для форматирования одной дорожки дискеты. Входными параметрами для этой функции выступают регистры процессора, назначение которых определено при рассмотрении функции 2. В этом случае значение ES:BX указывает на заранее подготовленный в памяти список дескрипторов секторов. В процессе выполнения за один оборот дискеты размечается целиком полная дорожка. Поэтому изменение параметров секторов в пределах этой дорожки невозможно.

В процессе своего развития в интерфейс FDD новых типов был введен сигнал, позволяющий изменять скорость вращения дискеты, что позволило поддерживать режимы с разной плотностью записи. С этой целью в адаптер FDD был введен новый порт, имеющий два бита (разряды 1, 0) для управления скоростью передачи данных. Разряд 7 этого регистра устанавливается в состояние 1 при смене дискеты и отбрасывается после выполнения дисководом операций поиска.

3.9. Интерфейсы HDD

Аппаратурная реализация HDD основана на применении специальных БИС, которые осуществляют логические функции управления HDD. Несмотря на достаточно большой выбор этих БИС, все они, как правило, поддерживают одни и те же форматы записи и содержат высокостабильный кварцевый генератор адреса программнодоступных портов ввода/вывода. Обмен данными с памятью ПК осуществляется в байтовом формате прямого доступа к памяти при помощи сигналов DRQ2, DACK2, IOW и IOR системной магистрали ПК (рис. 3.11). Последний цикл прямого доступа к памяти сопровождается сигналом TC. Для организации обмена адаптер HDD генерирует сигнал запроса на захват системной магистрали (сигнал DRQ2), и при получении ответа (сигнал DACK2) данные по шине данных передаются в адаптер HDD. Обмен данными между адаптером и накопителем на жестком диске производится по однопроводной линии в последовательном формате.

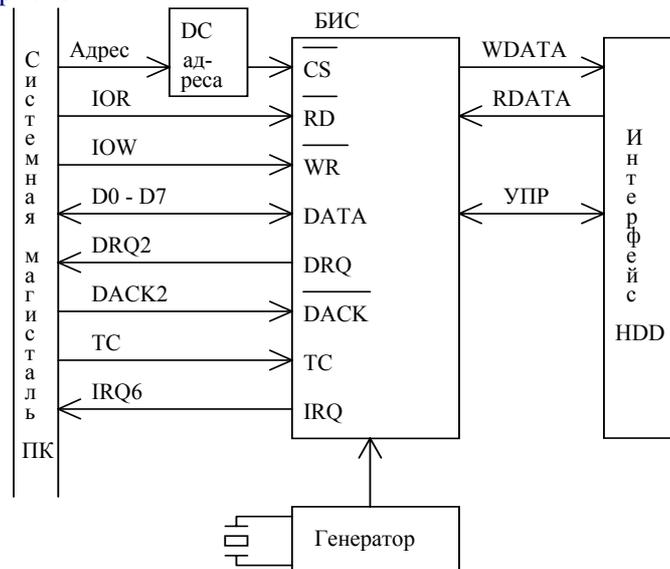


Рис. 3.11. Обобщенная блок-схема адаптера HDD

В настоящее время в качестве интерфейсов выступают модели от самого медленного ST506 до самого скоростного и дорогого IPI.

Интерфейс ST506 был разработан фирмой Seagate Technology в начале 80-х годов, он обслуживал HDD емкостью 5 Мбайт со скоростью передачи данных у накопителя 5 Мбайт/с. В дальнейшем он был усовершенствован в части управления перемещением магнитных головок и получил название ST412. С его помощью передача данных по каналу HDD – ПК достигала одного Мбайта/с, что существенно превышало скорость их переработки ранними моделями ПК.

Интерфейс ST506/412 позволяет обслуживать до четырех накопителей, схема сопряжения которых представлена на рисунке 3.12.

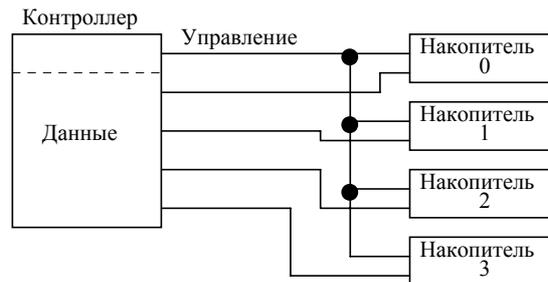


Рис. 3.12. Схема сопряжения контроллера с накопителями в интерфейсе ST412

Сигналы управления и состояния передаются по общему для всех накопителей 34-жильному кабелю, а обмен данными осуществляется по отдельному для каждого накопителя 20-жильному кабелю. Большинство линий 20-жильного кабеля являются резервными или общими. Для обмена данными посредством парафазных сигналов используется только четыре линии: две – WDATA и две – RDATA. Сигналы кабеля управления и кабелей данных требуют специального согласования с помощью блока резисторов 220/330 Ом и 100 Ом, соответственно, которые устанавливаются в специальную розетку.

Интерфейс IDE был предложен пользователям IBM-совместимых компьютеров кланов XT и AT в 1988 году и по функциональным характеристикам не отличается от аналогичных характеристик интерфейса ST412. Основные отличия коснулись конструктивного исполнения (контроллер реализован в накопителе) и количества обслуживаемых накопителей (максимум – два). Данный интерфейс подразделяется на два вида – XT и AT исполнения, в каждом из которых HDD подключается к адаптеру единственным 40-жильным кабелем (рис. 3.13). Отличительной особенностью интерфейса IDE (AT) от IDE (XT) является повышенная до 16 разрядов шина данных, которая активизируется в полном объеме сигналом IOCS16 (сигнал признака обращения к 16-разрядному порту).

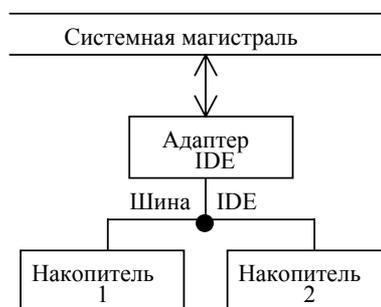


Рис. 3.13. Схемы сопряжения адаптера с накопителями в интерфейсе IDE

Интерфейс EIDE. Фирма Western Digital предложила интерфейс, получивший спецификацию Enhanced IDE (EIDE). Этот интерфейс имеет четыре основные особенности: использование IDE-накопителей свыше 504 Мбайт; более производительные режимы обмена

данными; подключение к одному адаптеру до четырех устройств и поддержка периферийных устройств, отличных от жестких дисков.

Увеличение емкости было достигнуто за счет реализации режима LBA (Logical Block Address). В этом случае информация транслируется в 28-разрядный логический адрес. Работа с такими логическими блоками позволяет преодолеть ограничение в 504 Мбайта.

Более производительный обмен данными обеспечивается реализацией режима PIO (Processor Input-Output) Mode3. Суть которого заключается в организации управления линией IORDY (I/O Channel Ready) шины IDE. Кроме того, обмен данными обеспечивает режим ПДП. Скорость обмена информацией в этом режиме теоретически составляет 13,3 Мбайта/с.

Расширение количества подключаемых устройств до четырех достигается за счет использования вторичного контроллера. Спецификация EIDE также позволяет использовать различные типы периферийных устройств, например, приводы CD-ROM или стримеры.

Интерфейс ESDI. Инициатором разработки данного интерфейса стала фирма Maxtor, в котором функции разделителя данных были перенесены из контроллера непосредственно в накопитель. В результате, во-первых, существенно уменьшились помехи в канале передачи данных по кабелю; во-вторых, появилась возможность настройки разделителя данных к характеристикам дисководов и носителя. Схемы фазовой автоподстройки частоты и генератора стали располагаться в HDD, и отпала необходимость в передаче по кабелю необработанных управляющих сигналов. Данные передаются в коде NRZ, и каждый бит данных сопровождается стробирующим импульсом. Все это позволило увеличить скорость передачи до 10 Мбит/с при теоретической производительности до 24 Мбит/с.

Стандартное подключение внешних устройств к контроллеру ESDI показано на рисунке 3.14.



Рис. 3.14. Схема сопряжения контроллера с внешними устройствами в интерфейсе ESDI

Интерфейс контроллера ESDI сопрягает шину данных ПК с конкретными внешними устройствами, в качестве которых могут выступать накопители на магнитной ленте, накопители на магнитных дисках, накопители на оптических дисках (CD-ROM), устройства памяти, HDD, а также сетевое оборудование. Контроллер ESDI устанавливается либо непосредственно на системной плате, либо представляет собой отдельную плату, которая устанавливается в слот системной магистрали ПК. При максимальной конфигурации интерфейс ESDI позволяет осуществлять управление до семи HDD, однако на практике их, как правило, два. Соединение контроллера с HDD осуществляется аналогично соединению в интерфейсе ST412: 20-жильным радиальным кабелем для шины данных и 34-жильным кабелем для шины управления. Последний из них подключается последовательно от накопителя к накопителю при максимальной длине до 3 метров.

Обмен управляющей информацией осуществляется двумя способами: либо контроллер активизирует соответствующие линии шины управления, либо информация передается по последовательному каналу. Для второго из способов передаваемые (принимаемые) команды (состояния) имеют размерность 16 бит плюс один бит контроля на четность. Информация по шине данных передается также по последовательному каналу со скоростью, равной скорости тактовой частоты процессов записи/считывания. Следовательно, реальная скорость передачи байтов информации равна около 1 Мбайт/с (в среднем, на один байт передаваемой информации расходуется 10 бит последовательного канала). Интересно, что количество линий шины

управления, отводимых для передачи компонент адреса позволяет теоретически обслуживать HDD емкостью до 1 Тбайт.

Интерфейсы SCSI. Данный интерфейс был предложен фирмой Shugart Associate и позволял ПК выдавать команды и передавать данные по параллельному каналу. Интерфейс SCSI выполняет функции сопряжения с несколькими внешними устройствами, каждое из которых имеет собственный контроллер (рис. 3.15).

Стандарт SCSI представляет совокупность аппаратных и технических требований к шине, предназначенной для сопряжения нескольких внешних устройств, а также систему команд и протоколы для управления ими. Шина состоит из 50-жильного кабеля, с помощью которого последовательно соединяются друг с другом до восьми устройств (контроллеров) при максимальной их удаленности до шести метров.

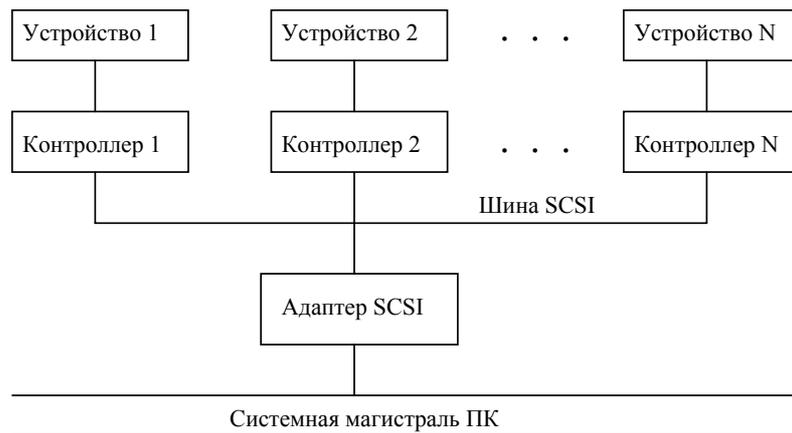


Рис. 3.15. Сопряжение внешних устройств в интерфейсе SCSI

Следует отметить, что подробное описание интерфейса SCSI занимает два тома, а его возможности существенно шире, чем у любого из описанных выше интерфейсов. Он способен обеспечивать управление не только накопителями на магнитных дисках и печатающих устройств, но также работу сопроцессоров в составе ПК.

В каждый момент времени только два устройства могут выполнять обмен информацией. Если одновременно шина требуется большему количеству устройств, то в работу включаются схемы арбитража и управление передается устройству с наивысшим приоритетом. По умолчанию принимается синхронный протокол обмена информацией. Однако фирма-изготовитель оборудования может задать и синхронные протоколы обмена. Независимо от типа обмена логический стандарт SCSI оперирует с данными, кодами команд и состояния. Для каждого вида устройств команды делятся на четыре категории: обязательные команды, команды расширения, необязательные и уникальные команды. Обязательные команды осуществляют управление накопителей на магнитных дисках. Расширенный набор команд поддерживается драйвером и позволяет использовать устройства с более широкими возможностями, чем, например, у HDD со стандартной емкостью. Необязательные команды еще более расширяют функциональные возможности системы, однако они не являются обязательными для интерфейса SCSI. Решение вопроса об использовании необязательных команд полностью зависит от фирмы-изготовителя внешних устройств.

В настоящее время появилась модификация интерфейса SCSI под названием SCSI-2. Он сохранил аппаратную и программную совместимость с интерфейсом SCSI, но вместе с тем позволяет обеспечивать передачу данных по 16- и 32-разрядной шине с максимальной скоростью передачи данных до 10 Мбайт/с. Для расширения функциональных возможностей SCSI-2 в нем предусмотрена возможность использования дополнительного 68-контактного кабеля. Стандарт SCSI-2 имеет режим создания очереди команд, что позволяет одновременно посылать одному устройству несколько команд, а также имеет набор команд для обслуживания устройств тип CD-ROM.

Успех широкого применения интерфейсов SCSI обусловлен тем, что он позволяет соединить с компьютером любое периферийное оборудование стандарта SCSI, включая его в конфигурацию системы или исключая из системы простым программированием. В результате у пользователя появляется возможность не только объединения по шине SCSI до семи внешних устройств, но и обеспечения непосредственного сопряжения между HDD различных ПК. Однако, в этом случае следует принимать во внимание существование двух разновидностей шин SCSI – однопроводная или дифференциальная, а также свободу выбора фирмой-изготовителем необязательных и уникальных команд.

50-жильный кабель интерфейса SCSI имеет следующие основные сигналы:

DB (0) – DB (7) – шина данных (D0 – D7);

DB (P) – бит паритета шины данных;

DIFFSENT – сигнал разрешения дифференциальных приемопередатчиков шины;

ANT – сигнал внимания, формируется инициатором для указания на возникшее условие внимания;

BISY – занято, сигнал занятости шины, объединенный по монтажному ИЛИ;

ASC – сигнал подтверждения приема информации;

RST – сигнал сброса, формируется любым из внешних устройств и объединяется по монтажному ИЛИ;

MSG – сигнал сообщения, формируется устройством в фазе сообщения;

SEL – сигнал выбора устройства;

C/D – сигнал команда/данные, определяет тип информации на шине данных;

REQ – сигнал запроса на передачу данных в режиме ответа с квитированием (REQ/ASC);

I/O – сигнал ввода/вывода для указания направления потока ввода/вывода.

Как и в интерфейсе ST412, данный интерфейс требует электрического согласования кабеля с помощью резистивной матрицы.

3.10. Оптические диски

Оптические диски в настоящее время представляют собой наиболее перспективные носители больших объемов информации. Главными недостатками по сравнению с магнитными носителями являются пока еще достаточно высокая стоимость и большое время доступа.

Все существующие оптические диски объединяет использование луча лазера для записи и чтения информации в цифровом виде. В процессе записи модулированный цифровым сигналом лазерный луч оставляет на активном слое оптического носителя след, который затем можно прочитать, направив на него луч меньшей интенсивности и проанализировав изменение характеристик отраженного луча. По характеру режима записи/чтения и свойствам носителя оптические запоминающие устройства можно подразделить на три категории: постоянные запоминающие устройства, устройства с однократной записью и многократным чтением и перезаписывающие устройства.

Главное преимущество оптической памяти перед магнитной - высокая плотность хранения информации. Дорожки оптического диска можно расположить значительно ближе друг к другу, чем на магнитном диске, благодаря точной фокусировке луча лазера. Для записи оптического бита требуется значительно меньше места, чем для записи бита на магнитном носителе.

Еще одно преимущество оптического диска состоит в том, что оптические носители значительно менее чувствительны к загрязнению по сравнению с магнитными. Активный записывающий слой оптического носителя не вступает в контакт ни с чем, кроме луча света. Благодаря наличию прозрачной оболочки пыль и грязь не попадают на фокальную плоскость. Оптические диски надежнее и прочнее гибких и жестких

магнитных дисков. С целью повышения надежности оптические дисководы снабжаются специальным защитным устройством - "кэдди" - для загрузки дисков.

Носители типа CD-ROM (компакт-диски - только для чтения). Основным стандартом, который определяет логический и файловый форматы записи компакт-дисков, является международная спецификация ISO 9660. Конструкция CD-диска описывается в стандарте Red-Book («Красная книга»), который появился в 1980 году, как плод совместных усилий двух компаний Sony & Philips. Конструктивно диск состоит из трех слоев: основного или подложки, отражающего и защитного. Основной слой изготавливают из прозрачного поликарбоната, на котором и формируется информационный рельеф. Информационный рельеф представляет собой непрерывную спиральную дорожку, состоящую из чередующихся углублений – питов и участков поверхности между ними -lands. На рельефную сторону подложки напыляется металлический отражающий слой толщиной 0,1 микрона. Сверху рельеф покрывается защитным изолирующим слоем толщиной около 200 микрон. В результате общая толщина CD-диска составляет 1,2 мм, а его диаметр – 120 мм.

Процесс тиражирования CD-ROM во многом похож на процесс тиражирования грампластинок. Такие диски получают путем литья в формы под давлением, при этом метки формата и данных представляются в виде бороздок, либо выступов на одной из внутренних поверхностей. Запись данных осуществляется с большой точностью (для этого используется аргоновый лазер) на поверхности слоя фоторезистора. Так получают оригинал или мастер-диск. После проявления мастер-диска изготавливают негативный диск методом гальванопластики. Этот диск потом используют в качестве матрицы для получения позитивного диска, который в свою очередь используется для изготовления негативных штампов, применяемых в процессе литья под давлением. В качестве материала для изготовления оптических дисков данного типа выбран поликарбонат с учетом его прочности и износостойчивости.

Принцип считывания информации основан на явлении поглощения или отражения луча лазера от поверхности диска. Для этого лазерный излучатель работает в диапазоне длин волн от 770 до 830 нм, а глубина пит выбирается равной 150 нм. То есть с учетом среды защитного слоя в области пита возникает эффект четвертьволнового согласования и луч лазера полностью поглощается. Метод кодирования информации – NRZ, то есть традиционный для дисковых подсистем. Переход от пита к промежутку и наоборот означает запись логической «1», а отсутствие перехода – запись логического «0».

Плотность размещения информации вдоль дорожки остается постоянной как в центре диска, так и на внешних радиусах. В связи с этим шпиндельный двигатель, который служит для вращения диска, работает в режиме постоянной линейной скорости.

Носители типа WORM. Эти носители, так же как и компакт-диски, могут быть записаны только один раз. Технология WORM - шаг вперед по сравнению с компакт дисками.

Дисководы технологии WORM имеют стандарты 5,25; 8; 12 и 14 дюймов. WORM диски формата 5,25 имеют емкость от 300 до 600 Мбайт на одной стороне. На одной стороне формата 14 дюймов можно записать 3Г байта информации (данные 1990 года). По заявлению фирмы Sony; ее 12-дюймовые WORM диски моделей WDM-3DLO и WDM-3DAO имеют время жизни 100 лет - в 5 раз больше, чем у других моделей.

В целях оптимизации времени доступа и информационной емкости диска в оптических дисководах используют различные схемы вращения и организации дорожек. Наиболее распространенными схемами вращения диска являются схема CAV (постоянная угловая скорость) и схема CLV (постоянная линейная скорость). В устройствах, работающих по схеме CAV, так же как в магнитных дисководах, поддерживается постоянная угловая скорость вращения диска. Поскольку скорость передачи данных

головкой постоянна, на всех дорожках может быть записано одинаковое количество информации. В результате на внешних дорожках плотность записи будет меньше, что приводит к неэкономичному использованию поверхности диска.

В дисководах, работающих по схеме CLV, максимизируется емкость диска, однако эти устройства имеют более сложную конструкцию и работают медленнее устройств CAV. Для поддержания постоянной скорости движения поверхности диска относительно головки частота вращения диска меняется в зависимости от номера дорожки, над которой находится головка. Таким образом, достигается максимальная плотность заполнения.

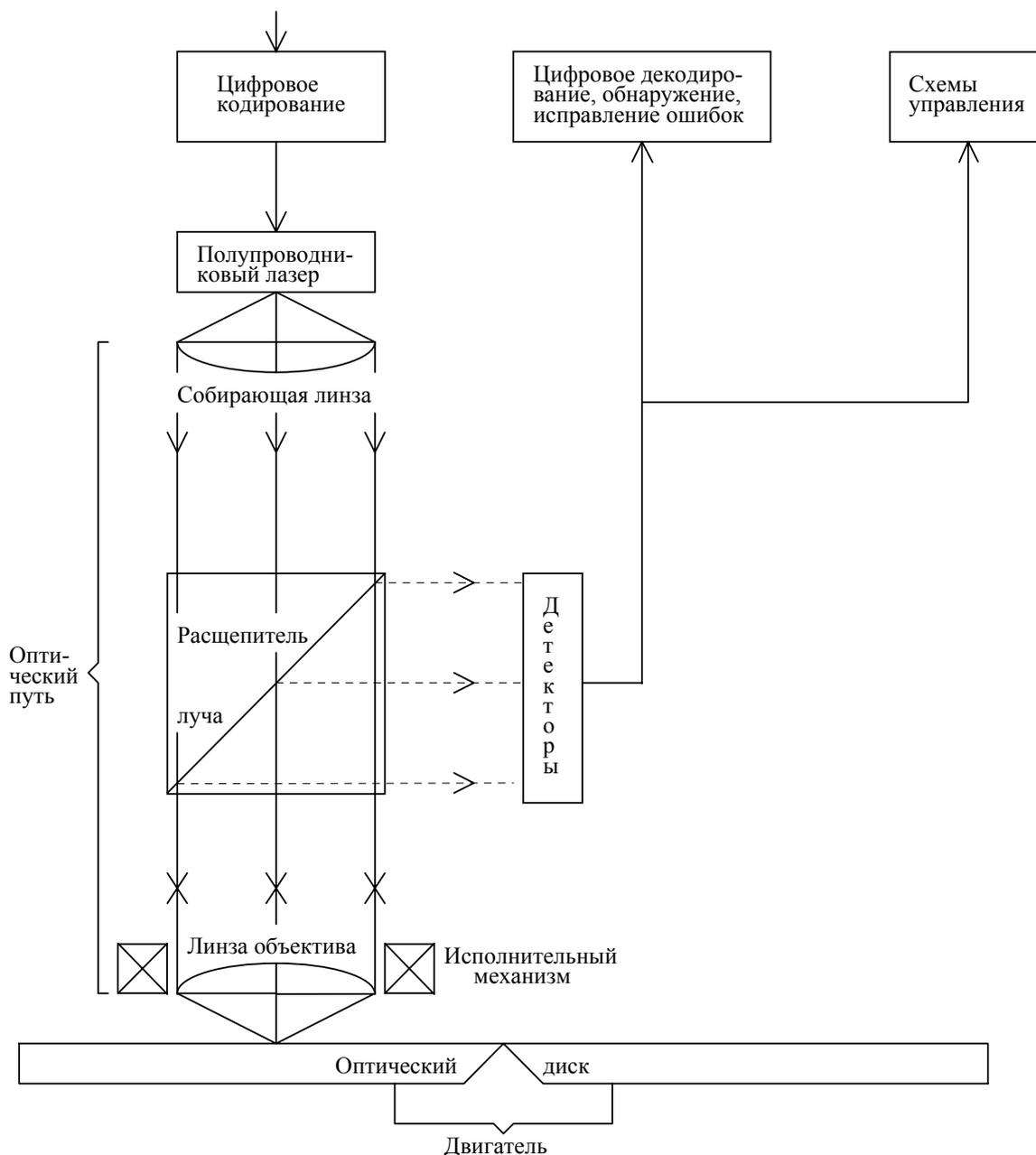


Рис. 3.16. Система записи/чтения WORM-дисковода

Существует множество схем функционирования WORM дисководов. В одних схемах поглощение мощного импульса лазерного излучения вызывает локальное испарение носителя информации, изготовленного из содержащего краситель полимера (органические пленки с поглощающим красящим веществом). Это приводит к образованию пузырька в тонком

поглощающем слое, расположенном над активным слоем. В других устройствах луч лазера создает микроуглубления в активном слое, под которым находится металлическая пленка с высокой отражающей способностью. По изменению коэффициента поглощения устанавливается отсутствие или наличие микроуглубления или пузырька.

Другим стандартным активным материалом, используемым в WORM-дисках, является полуметаллический теллур. Воздействие мощного импульса приводит к плавлению теллура. Силы поверхностного натяжения создают в пленке отверстия или углубления, которые остаются при затвердении материала.

На рисунке 3.16 представлены основные элементы системы записи/чтения WORM-дисководов. Головка записи/чтения содержит полупроводниковый лазер. Чтобы обеспечить работу лазера в импульсном режиме при записи данных, его ток модулируется электронными схемами.

Генерируемый лазером луч проходит через коллиматорную линзу, поляризующий расщепитель луча, четвертьволновую пластину и фокусирующую линзу объектива. Расщепитель луча пропускает свет, поляризованный в одном направлении, и отражает свет, поляризованный в перпендикулярном направлении. Четвертьволновая пластина изменяет поляризацию луча. Двукратное прохождение луча через эту пластину приводит к повороту его плоскости поляризации на 90° .

Линза объектива фокусирует лазерный луч на тонком записывающем слое (100 нм), расположенном под прозрачной пластиковой подложкой. Генерируемые системой автоматического регулирования сигналы управляют механическим исполнительным механизмом, который перемещая линзы объектива, фокусирует луч и держит его на подложке. Отраженный свет фиксируется фотодетекторами, преобразуется в электрические сигналы, которые передаются системе автоматического регулирования для коррекции управления.

Сигнал, генерируемые четырьмя фотодетекторами, декодируются блоком регистрации и исправления ошибок и схемами модуляции-декодирования. Отраженный луч несет информацию о двоичных нулях и единицах в зависимости от наличия или отсутствия микроуглублений, пузырьков или обратной намагниченности в определенном участке оптического диска. Большинство дисков имеет тонкие концентрические канавки (ширина 0,8 мкм) нанесенные на расстоянии 1,6 мкм друг от друга и предназначенные для слежения за дорожкой записи. Запись данных производится в промежутке между канавками.

Носители с возможностью перезаписи. Стираемые диски обычно имеют подложку из поликарбоната. Канавки формируют в процессе литья под давлением одновременно с созданием подложки.

Над канавками расположены три-четыре дополнительных слоя. Среди них - активный слой, используемый для записи данных. Активный слой обычно изготавливается из сплавов редкоземельных и переходных металлов, например, из сплава тербий-железо-кобальт (TbFeCo). Этот слой расположен между двумя прозрачными слоями диэлектрика, обеспечивающими как оптическое усиление поворота плоскости поляризации, так и защиту сплава от взаимодействия агрессивных сред.

Активный записывающий слой стираемого оптического диска подвергается намагничиванию. Поворот плоскости поляризации луча света, отраженного таким магнитным слоем, известен как магнитный эффект Керра. На рисунке 3.17 показано, как данный эффект используется в магнитооптической памяти. Чистый диск имеет однородную пространственную намагниченность, показанную на рисунке стрелками, направленными вверх. В информационном смысле такой диск содержит только двоичные нули.

Чтобы записать двоичную единицу в конкретную область диска, нужно сфокусировать на активном слое импульс лазерного излучения длительностью 100 нс. Одновременно, небольшой магнит, расположенный с противоположной стороны диска, создает внешнее магнитное поле, направленное в сторону, противоположную исходному вектору намагниченности активного слоя. Участок активного слоя, поглощая энергию лазерного импульса, разогревается до температуры в сотни градусов. При этом направление его намагниченности изменяется на противоположное и остается таким в результате быстрого охлаждения участка.

Чтобы стереть информацию в записанной области, изменяется направление внешнего магнитного поля и область вторично подогревается лазерным импульсом. При этом область

приобретает исходное направление намагниченности. Следует отметить, что перед повторной записью предварительное стирание обязательно.

В процессе чтения линейно поляризованный лазерный луч меньшей интенсивности фокусируется на активном слое. При отражении луча его вектор поляризации слегка меняет направление (приблизительно на один градус). Поворот вектора происходит по часовой стрелке либо против, в зависимости от направления намагниченности участка активного слоя. Знак поворота указывает, что записано в данном участке - 0 или 1.

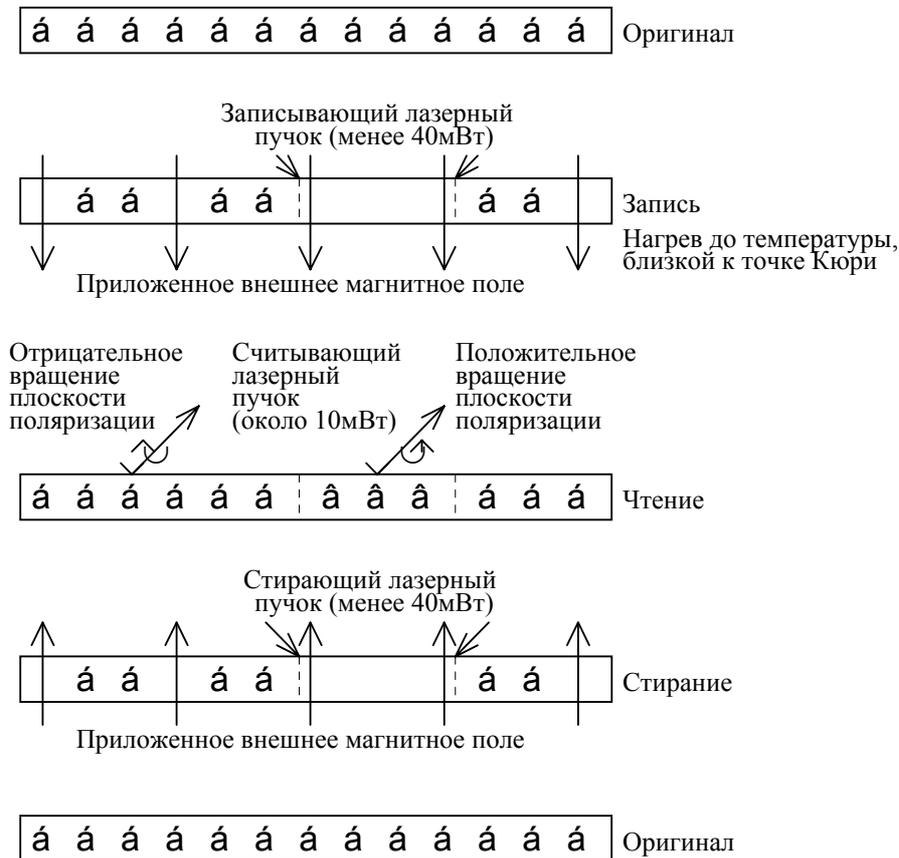


Рис. 3.17. Магнитооптическая запись

Продолжается поиск материалов, обеспечивающих большой угол поворота вектора поляризации. Среди них гранатовые пленки с примесью висмута и многослойные металлические структуры (или сверхструктуры). Среди других перспективных материалов для перезаписываемых оптических носителей - аморфные сплавы теллура, селена и других элементов, обеспечивающих изменение фазы отраженного луча.

Недостаток носителей с изменением фазы в том, что они допускают ограниченное число циклов записи/стирание, после чего характеристики ухудшаются до неприемлемого уровня.

Наряду с уже рассмотренным *методом магнитооптической комбинации*, существуют и другие методы, положенные в основу создания дисков со стиранием/перезаписью: *метод изменения фазового состояния материала* и *метод пластической деформации*.

Изменение фазового состояния материала (переход из аморфного состояния в кристаллическое). Определенные сплавы на основе теллура или селена в виде тонких металлических пленок могут находиться при комнатной температуре в двух фазовых состояниях - аморфном и кристаллическом, переходя под действием лазерного излучения из одного состояния в другое. Обычное аморфное состояние соответствует единице, а кристаллическое - нулю. Скорость записи зависит главным образом от времени кристаллизации и составляет от 200 нс до 1 мкс. Процесс чтения оказывает разрушающее действие на среду, это надо учитывать при работе.

Пластическая деформация. Согласно этому методу используются два соединенных между собой тонких слоя органических красящих полимеров с разными тепловыми, оптическими и механическими свойствами. При временном нагревании пятно на диске меняет свою форму, которая сохраняется неизменной до следующего цикла нагревания. Это свойство позволяет по желанию создавать и ликвидировать локальные механические деформации.

Эти два метода не нашли такого широкого применения как метод изменения фазового состояния материала и являются менее жизнеспособными, чем последний.

Полупроводниковые лазеры играют основную роль в развитии оптической записи. Эти крошечные устройства 200-300 микрон, работая в магнитооптических дисководах, создают излучение мощностью до 40 мВт, с длиной волны 830 нм (780 нм). Размер пятна на поверхности активного слоя зависит от длины волны лазерного излучения и от оптической силы объектива. В современных системах размер пятна немного больше длины волны.

Ведутся разработки по созданию устройства для одновременной записи и чтения нескольких дорожек, в котором используются матрицы из 4-х и 5-ти лазеров.

С оптическим каналом связаны проблемы, не встречающиеся в области магнитных дисков, например, ошибки, связанные с загрязнением диска в процессе его изготовления. Поскольку метки расположены на дорожке на расстоянии длины волны друг от друга, а дорожки стоят на расстоянии 1,6 мкм друг от друга, частица пыли на активном слое приведет к потере большого числа смежных битов. Для решения этой проблемы созданы сложные коды модуляции, ведутся исследования по поиску лучших кодов.

Логическая организация данных на CD-дисках изначально разрабатывалась исключительно для записи звука, поэтому первый формат получил название CD-DA (Compact Disk – Digital Audio). Перед записью на диск исходный стереофонический сигнал оцифровывается по методу импульсно – кодовой модуляции. Частота дискретизации 44,1 кГц при разрядности 16 бит. Каждые шесть отчетов левого и правого каналов группируются в микрокадры размером в 24 байта. К каждому микрокадру добавляется 64-битовый код коррекции ошибок (CIRC кодирование). Таким образом общий размер микрокадра равен 256 бит (24 x 8 + 64). Далее к каждому микрокадру добавляется еще 8 бит, которые называют *субкодами*, и обозначают символами P, Q, R, S, T, U, V и W. Совокупность одноименных бит, но в разных микрокадрах называют каналом (например, Q-канал, S-канал и т.д.). Субкодовые биты используются для хранения служебной информации: например, Q-канал содержит информацию о формате треков и кадров, а также временных меток, по которым отслеживается время воспроизведения. Таким образом каждый микрокадр с субкодами суммарно занимает 264 бита или 33 байта.

При работе со звуком минимально обрабатываемой единицей является кадр – объединение 98 микрокадров. Число 98 выбирается с таким расчетом, чтобы звучание одного кадра равнялось 1/75 секунды. Наконец, поток из кадров еще раз кодируется с использованием избыточного EFM-кода: каждый байт преобразуется в 14-ти битное слово, а он, в свою очередь, разбивается на участки по 11 бит. К этому участку добавляется по три бита нулей, что удобно для считывания по принятой в CD-технологии (кодировании).

Таким образом, после всех логических преобразований один кадр занимает уже 57624 бита или 7203 байта. Если учесть, что это только 1/75 секунды, то результирующий битовый поток в процессе считывания (или записи) равен 4321800 бит/с или 527,6 Кбайт/с. При этом плотность потока только информационных (звуковых) данных составляет 32,65% или 172,3 Кбайт/с. Такой режим работы CD- дисков называют «x1».

Все записанные на диск данные размещаются в трех основных областях: Lead-In Area, Program Area и Lead-Out Area. Первая из них расположена в начале диска и содержит служебную информацию: о формате диска, названии музыкальных произведений и таблицу содержимого (TOC). Далее следует Program Area, в которой хранятся собственно данные и которая отделена от Lead-In Area участком из 150 пустых кадров. Это обеспечивает 2-х секундную паузу перед проигрыванием музыки. Область Lead-Out Area является последней и расположена на краю диска. Общая длительность записи музыкальных произведений на компакт-диске не превышает 74 минут.

Информационная дорожка, проходящая через все три области диска разбивается на последовательно расположенные участки – треки. Треки нумеруются шестнадцатеричными двухразрядными числами, начиная с нуля. При этом нулевой трек всегда занимает область Lead-In

Area целиком. Последний трек на диске занимает область Lead-Out Area, которая нумеруется как “AA”.

Информация о треках области Program Area хранится в ТОС диска, а именно в Q-канале Lead-In Area. В ТОС содержится список треков с указанием их типа (звуковой или двоичные данные) и физическим адресом его начала. Физический адрес указан в формате MSF – «минуты, секунды, кадры», это привычное для аудио-CD, например, 20:30:40. Чтобы найти физический адрес трека, необходимо отсчитать от начала Program Area количество кадров, «умещающихся» в 20 минутах, 30 секундах плюс 40. Для нашего примера это $75 \times (20 \times 60 + 30) + 40 = 11290$ кадров.

Следующим после CD-DA формата записи данных на компакт диск был формат CD-ROM, разработанный в 1984 году компаниями Sony и Philips. Формат описан в стандарте Yellow Book и является первым форматом, специально предназначенным для записи двоичных данных. Его отличительная особенность от формата CD-DA – изменение формата кадра в той его части, где содержатся данные. Напомним, что это 2352 байта, совокупность которых в формате CD-ROM называют *сектором*. Стандарт описывает три моды формата записи на диск: Mode 0, Mode 1 и Mode 2. Отличие между ними в структуре логических полей в секторе (рис. 3.18).

Sync	Заголовок сектора		Данные	E DC	Intermediate	P-Parity	Q-Parity
	Адрес сектора	Mode					
12 байт	3 байта	1 байт	2048 байт	4 байта	8 байт	172 байта	104 байта

Рис. 3.18а Структура сектора формата CD-ROM Mode 1

Sync	Заголовок сектора		Данные
	Адрес сектора	Mode	
12 байт	3 байта	1 байт	2336 байт

Рис. 3.18б Структура сектора формата CD-ROM Mode 0 и Mode 2

Поля EDC и ECC являются кодами обнаружения и исправления ошибок и в совокупности занимают 288 байт (поля P-Parity и Q-Parity входят в код ECC). Mode 0 отличается от Mode 2 тем, что блок данных из 2336 байт заполняется кодом 00h.

Дальнейшим развитием стандартов CD-ROM являются стандарты CD-ROM/XA, High Sierra, CD-I, CD-E/CD-R и DVD. Каждый из них имеет свои конструктивные особенности и различные форматы представления данных. Это сделано в связи с тем, что стандарты предназначены для конкретных целей. Так, например, стандарт DVD предназначен для записи изображения и звука, емкость DVD-дисков увеличена в 7 раз за счет использования двухслойной записи. По сути это два диска, плотно соединенных друг с другом и т.д.

3.11. Стримеры

Носители на магнитной ленте являются цифровыми устройствами для обеспечения надежного хранения информации и применяются с начала 50-х годов.

Под стримером понимается лентопротяжный механизм, работающий в инерционном режиме. Инерционный режим лентопротяжного механизма был предложен фирмой IBM в 1978 году. Суть этого режима состоит в том, что длина отрезка магнитной ленты, проходящего мимо головки при остановке или перезапуске, превышает длину между блоками информации, записанными на ленте. Поэтому после остановки ленту необходимо вернуть назад, т.е. перепозиционировать. Так что инерционный режим может использовать очень короткие промежутки между блоками информации. Основным недостатком этого режима является большое время позиционирования (0,1 – 2с).

Стримеры выполняются в виде отдельного автономного блока, установленного рядом с компьютером, а чаще – в виде устройства, встроенного внутрь системного блока компьютера, которое по своим размерам аналогично накопителю для гибких дисков.

Чаще всего в настоящее время применяются накопители, использующие полудюймовые ленты, намотанные на бобины или катушки и более современные накопители, использующие для намотки ленты специальные кассеты – картриджи или компакт-кассеты.

Информация на полудюймовой ленте записывается по 9 дорожкам (8 бит данных плюс контрольный бит) с плотностью 800, 1600, 3200 или 6250 бит/дюйм. Преимуществом полудюймовой ленты является их практически стопроцентная совместимость и переносимость.

Кассеты, во многом определяя, по сути, различные форматы ленты, различаются как по внутреннему устройству, так и по ширине ленты, носят следующие названия: четвертьдюймовые картриджи (QIC-картриджи), 8-миллиметровые картриджи, 4-миллиметровые картриджи DAT и цифровые компакт-кассеты.

Для картриджей применяется специальная высококачественная магнитная лента, в которой используются особо чистые мелкозернистые магнитные частицы, а поверхность ленты тщательно полируется. Это гарантирует надежность записи и длительную сохранность записанных данных. Для разных типов картриджей применяются различные методы кодирования.

Цифровая компакт-кассета (аналог аудиокассеты) является разработкой фирмы Philips. Устройства для цифровых компакт-кассет не находят в настоящее время серьезного применения в составе персональных компьютеров.

Емкость картриджей и компакт-кассет колеблется от десятков мегабайт до нескольких гигабайт.

Главное достоинство стримеров – возможность быстрого и надежного сохранения больших объемов ценных данных или программ. Поэтому стримеры используются преимущественно для резервного архивирования информации в промышленности, в издательствах, в банках, в деловом и научном мире.

Выпускают стримеры различные фирмы. Хорошей репутацией пользуются изделия фирм Everex, Archive, Mountain Computer, Sony и Colorado Memory Systems. Лидером в этой области является фирма Wangtec, выпускающая широкий спектр моделей стримеров емкостью от 40 Мбайт до 1,3 Гбайт. Стримеры обычно комплектуются платой контроллера и драйвером.

К недостаткам стримера можно отнести его относительно высокую стоимость и затрудненный поиск информации, т.е., как и в магнитофоне, для поиска нужного места на пленке порой приходится перематывать всю кассету.

Глава 4. Устройства расширения функциональных возможностей ПК

4.1. КЭШ-память

Производительность высокоскоростных систем на базе микропроцессоров INTEL 80386 и выше в основном ограничивается быстродействием подсистем памяти. Этим и обусловлено использование КЭШ-памяти.

Для микропроцессоров, синхронизируемых, например, частотой 33 Мгц, длительность такого импульса составляет, приблизительно, 30 нс, а обычные микросхемы динамических ОЗУ имеют время доступа к памяти от 60 до 100 нс. Поэтому ЦП вынужден простаивать 2-3 периода тактовой частоты, т.е. пока информация из соответствующих микросхем памяти установится на системной шине данных компьютера. Этих циклов ожидания можно избежать, применяя микросхемы более скоростной статической памяти. Однако, их применение для построения памяти большого объема оказывается достаточно дорогостоящим и не всегда оправдано. Целесообразнее использовать сочетание небольшого по объему, но построенной на динамических ОЗУ памяти и более медленной, построенной на динамических ОЗУ основной памяти ПК большой емкости.

Хотя большинство современных систем на базе микропроцессора 80386 и выше работает в конвейерном режиме, т.е. когда микропроцессор выдает следующий адрес обращения к ОЗУ раньше, чем завершится цикл обращения к памяти, этого все равно не достаточно, чтобы превзойти производительность статических ОЗУ.

Одним из путей улучшения пропускной способности памяти является использование систем КЭШ-памяти (рис. 4.1). В данных системах используются статические ОЗУ небольшого объема с непосредственной выборкой, а основной объем памяти отводится под динамические ОЗУ в качестве вторичной памяти. В результате имитируется система с высокоскоростным ОЗУ большой емкости. В функции контроллера входит отслеживание направления передачи данных: какой блок данных должен быть направлен в КЭШ-блок, а какой – в главную память ПК. Каждый раз, когда микропроцессору требуется информация, отсутствующая в КЭШе, он вынужден обращаться через системную шину к основной оперативной памяти. После этого решается, должна ли происходить замена строки в КЭШ-памяти, и какая именно. В подавляющем большинстве случаев об этом заботится встроенный в контроллер LRU-алгоритм (Lost Recently Used), который обновляет именно ту строку КЭШа, которая используется менее интенсивно.

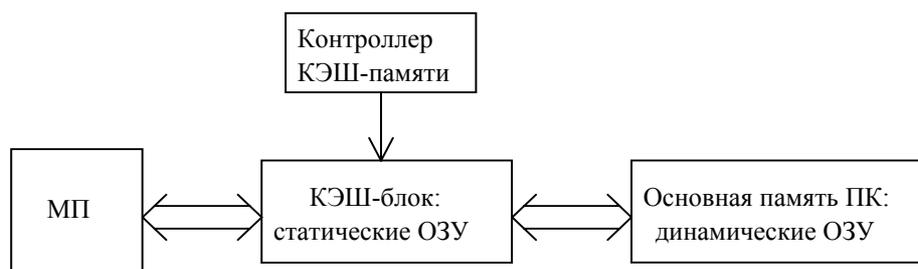


Рис. 4.1. Система КЭШ-памяти в ПК

Конструктивно КЭШ-память может представлять собой внешний автономный узел на системной плате ПК или быть встроенной внутрь кристалла микропроцессора. В этом случае объем КЭШа составляет до 256 Кбайт. При работе микропроцессора с памятью вначале проверяется наличие требуемых данных в КЭШ-блоке, и, если информация не найдена, то происходит обращение к главной памяти. Очевидно, что чем больше емкость КЭШ-блока, тем вероятнее, что в нем будут найдены требуемые данные.

Обнаружение искомой информации в КЭШ-блоке принято считать как удачное обращение к КЭШ-памяти. Так, например, если КЭШ-блок емкостью 256 Кбайт встроен в кристалл микропроцессора, то коэффициент удачных обращений обычно равен 0,75. Работа внешней КЭШ-

памяти в системах на базе МП 80386 позволяет увеличить коэффициент удачных обращений до 0,95.

Обращение к памяти в ПК, как правило, носит не случайный характер, что позволяет построить работу КЭШ-памяти по принципу предсказания следующего адреса обращения. Это вполне возможно, поскольку программы чаще всего обращаются к последовательным ячейкам памяти. Если спроектировать контроллер КЭШ-памяти таким образом, чтобы вместо одиночных выборок выбирать данные из памяти блоками, то вероятность удачных обращений неизбежно возрастает. Действительно, при разделении главной памяти на блоки и возможности пересылки таких блоков из главной памяти в КЭШ и обратно, в следующем обращении микропроцессора к памяти будет ссылка на тот же самый блок, что и в текущем обращении. Если запрашиваемая информация не обнаружена в КЭШ-памяти, то контроллер обновит весь блок целиком и т.д.

Размер блока существенно влияет на коэффициент удачных обращений. В целом, существует такая зависимость, согласно которой чем больше размер блока, тем выше коэффициент. Однако следует отметить, что уменьшение количества блоков в КЭШ-памяти снижает этот коэффициент.

Архитектура КЭШ-памяти определяется тем, каким образом основная память большая по объему отображается на сравнительно небольшой КЭШ-блок. Существуют три типа организации КЭШ-памяти. Самой простой организацией обладает КЭШ-память с *прямым отображением*. В этом случае адрес памяти полностью определяет используемую строку КЭШа. Таким образом, один или несколько блоков динамической памяти строго соответствуют одной строке КЭШа, однако, поскольку занимать ее в одно и тоже время может только один из них, то для каждой строки используется специальный признак – тег (tag). Более подробно работа КЭШ-памяти с прямым отображением рассмотрена на схеме, которая изображена на рисунке 4.2.

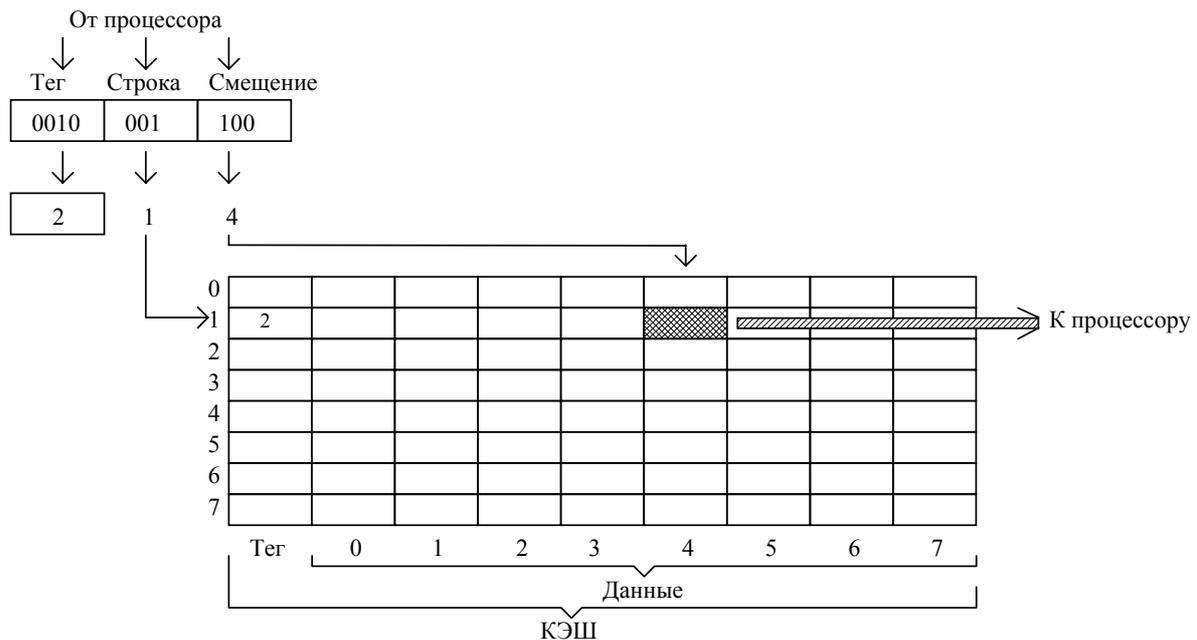


Рис. 4.2. Организация КЭШа с прямым отображением

Допустим, что гипотетический микропроцессор использует 10-разрядный адрес, размер КЭШа составляет 8 строк, а длина каждой строки – 8 байт. В какой-то момент времени микропроцессор хочет прочитать два байта информации, хранимой в оперативной памяти, с адреса 0010001100. Для КЭШ-памяти этот адрес подразделяется на три части. В данном случае три младших разряда называются смещением и полностью определяют положение каждого из восьми байт в строке. Средние три разряда позволяют однозначно выбрать одну из восьми строк КЭШа. Оставшиеся старшие разряды несут информацию о теге. И только в том случае, если тег строки КЭШа совпадает со значением, определяемым старшими разрядами адреса, байты информации считываются из КЭШ-блока. В противном случае чтение должно идти из основной

памяти, или информация в строке должна быть заменена. Разумеется, в случае ее замены меняется и значение ее тега.

Преимуществом реализации такого типа архитектуры являются довольно низкие затраты, поскольку, по сути, требуется всего лишь одна операция сравнения (для тегов). Недостатки ее, впрочем, также очевидны. Например, если два блока данных основной памяти, используемые одинаково часто, претендуют на одну и ту же строку в КЭШе. К недостаткам также можно отнести большой объем пересылок данных в случае частых обращений к различным частям главной памяти. В среднем производительность систем на базе КЭШа с прямым отображением имеет коэффициент удачных обращений от 0,41 для емкости КЭШа 1 Кбайт, до 0,89 для 128 Кбайт, при условии, что разрядность КЭШа равна 4 Кбайт. Внешняя КЭШ-память с прямым отображением используется, например, вместе с 80386 процессорами (КЭШ-контроллер i82385), а внутренняя – в микропроцессоре DEC Alpha 21064.

Другим типом организации КЭШ-памяти является *полностью ассоциативная*. В этом случае между блоками данных не существует взаимосвязи, т.е. любой блок памяти может занимать любую строку КЭШа (рис. 4.3).

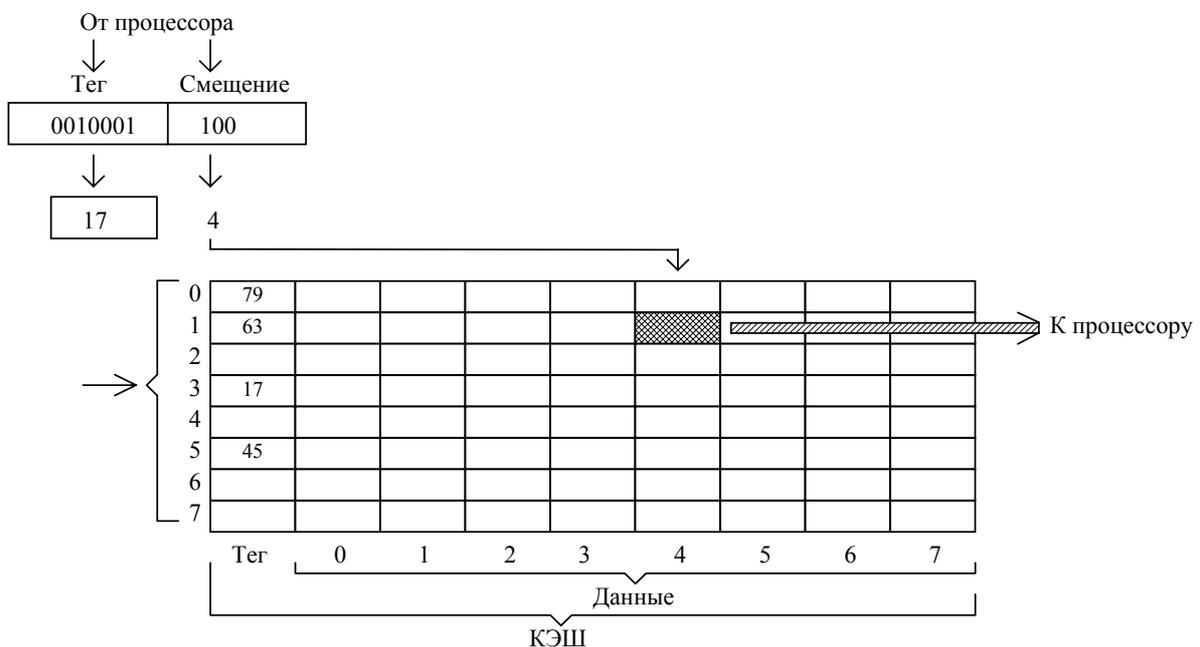


Рис. 4.3. Организация полностью ассоциативного КЭШа

Полный адрес памяти делится только на две части: младшие разряды – смещение в строке и старшие разряды – информация о теге. В этой архитектуре решена проблема конфликтов адресов, однако сама КЭШ-память требует для своей реализации больших аппаратных затрат, поскольку значения тегов должны уже сравниваться для всех линий КЭШа. Данная архитектура не нашла широкого применения. Тем не менее, процессор 6*86 фирмы Сугіх имеет вторичную встроенную 256-байтную КЭШ-память для команд, которая имеет полностью ассоциативную организацию.

Третий тип организации КЭШа – это *множественный ассоциативный КЭШ*, который является компромиссом между первыми двумя типами КЭШа. В этом случае несколько линий (две, четыре, пять, восемь) объединяются в наборы, и средние биты адреса памяти определяют уже не конкретную линию (как в прямом отображении), набор (рис. 4.4). Сравнение тегов со значением старших разрядов производится только для линий, входящих в набор. Данный метод увеличивает коэффициент удачных обращений, и как следствие, эффективность системы. Множественный ассоциативный КЭШ имеет подавляющее число процессоров, например, 486DX, 486DX2, 486DX4, Pentium и т.д.

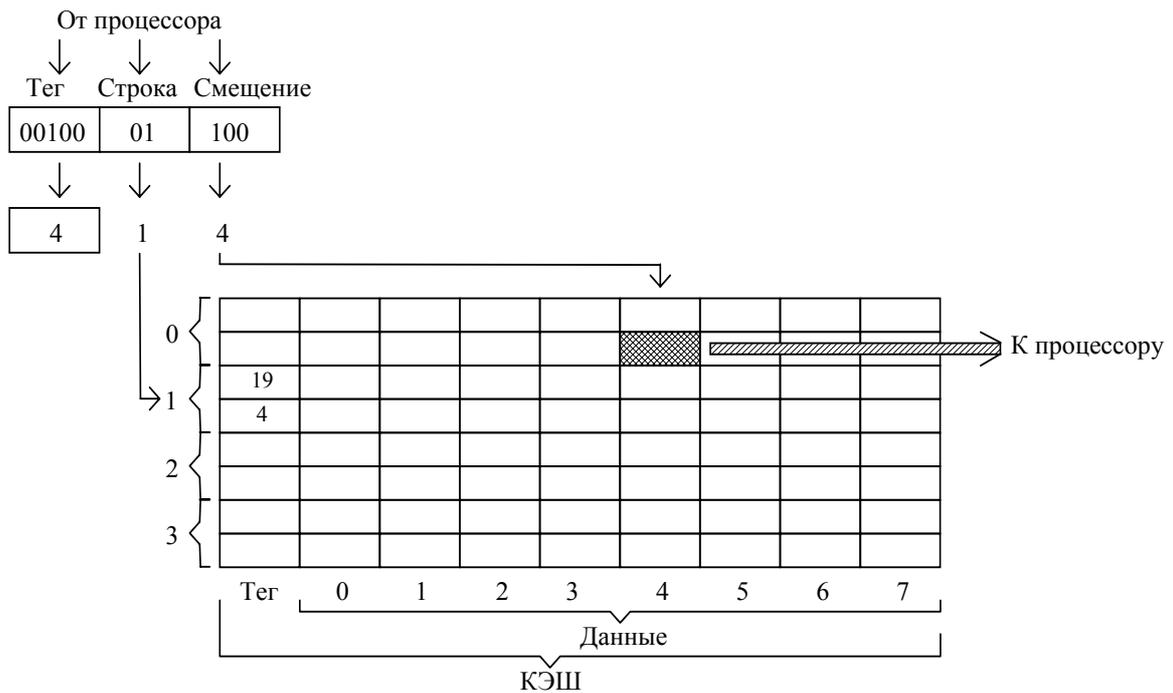


Рис. 4.4. Организация двухвходового наборно-ассоциативного КЭШа

По количеству линий КЭШа, входящих в набор, подобная архитектура может называться 2-входовой, 4-входовой и т.д.

В случаях, когда появляются операции записи в КЭШ, данные в динамических ОЗУ не корректируются при изменении содержимого статического ОЗУ. Если не принимать специальных методов управления, то может нарушиться целостность хранимых данных. Существует несколько способов обновления информации в главной памяти, гарантирующих, что любое обращение к данным закончится возвратом в главную память нового текущего значения данных. Один из способов носит название *сквозной записи* (Write Through) и предполагает наличие двух копий данных – одной в главной памяти, а другой – в КЭШ-памяти. Каждый цикл записи процессора в память идет через КЭШ. Это обуславливает высокую загрузку системной шины, т.к. на каждую операцию модификации данных приходится две операции записи. Поэтому каждое обновление содержимого КЭШ-памяти ощутимо сказывается на работе шины. С другой стороны, микропроцессор вынужден ожидать окончания процесса записи в основную память. Хотя следует отметить, что в главной памяти в этом случае всегда содержатся правильные данные. Такой способ записи используют, например, микропроцессоры 486 (SX, SX2, DX, DX2, DX4), за исключением изделий Sx486DX и Sx486DX2 фирмы Сугіх.

Способ сквозной записи может быть усовершенствован введением использования операции с буферизацией. Данный метод носит название *метод с буферизированной сквозной записью* (Buffered Write Through). В этом случае микропроцессор начинает обрабатывать новый тактовый цикл прежде, чем завершится операция записи в главную память. Данный метод результативен, когда любой запрос записи, за которым следует запрос чтения, заканчивается удачным обращением к КЭШу, иначе микропроцессор какое-то время должен ждать.

Существует еще один метод, который носит название *метода обратной записи* (Write Back). При использовании этого метода цикл записи микропроцессора происходит сначала в КЭШ-память, если там присутствует адрес приемника. Если адреса приемника в КЭШ-памяти нет, то информация записывается непосредственно в память. Содержимое основной памяти обновляется только тогда, когда из КЭШ-памяти в нее записывается полный блок данных, т.е. строка КЭШа. Таким образом, при обращении за данными контроллер КЭШ-памяти проверяет бит признака, если этот бит установлен (т.е. содержимое блока было изменено), то данные могут быть переписаны в главную память. Преимуществом данного метода является намного меньшая загрузка системной шины, потому что на ней выполняется меньше циклов записи. Описанный

метод записи, как правило, обеспечивает большее быстродействие, чем два предыдущих, поскольку число изменений блоков меньше, чем число операций записи данных. Однако для его организации требуются и более высокие затраты, особенно для того, чтобы обеспечить целостность данных. Такой метод записи используется, например, микропроцессорами Pentium фирмы Intel и Cx486DX и Cx486DX2 фирмы Сугіх.

Конструктивно КЭШ-память также можно разделить на *смешанный* (принстонская архитектура) и *раздельный* (гарвардская архитектура) КЭШ.

Раздельный КЭШ для команд и данных позволяет оптимизировать размер каждого КЭШа и его линии, а также количество входов независимо от другого. Кроме того, выборка команд и данных может в этом случае производиться одновременно. Тем не менее, имеются и недостатки. Они проявляются, например, если вдруг одна из выполняемых программ потребует КЭШ-памяти для команд чуть большего размера, а для данных – чуть меньшего, или наоборот.

Смешанный КЭШ используется как для чтения команд, так и для переноса данных. Смешанную КЭШ-память имеют процессоры Alpha фирмы DEC и 6*86 фирмы Сугіх. Более известный микропроцессор Pentium фирмы Intel использует раздельные 2-входные множественно ассоциативные КЭШи для команд и данных размером по 8 Кбайт.

Особое значение имеет Кэш-память для микропроцессоров, которые работают с умножением внешней тактовой частоты, когда скорость работы внутренних блоков микропроцессора в два и более раз выше скорости остальной части системы. Внутренние функциональные узлы подобных микропроцессоров – математический сопроцессор (если есть), КЭШ, устройства управления памятью, арифметико-логическое устройство – используют умноженную тактовую частоту, в то время как остальные элементы системной платы – системная и внешняя КЭШ-память, вспомогательные микросхемы – работают с обычной скоростью. Это позволяет увеличить производительность всей системы, как правило, за счет хранения части данных и вспомогательных кодов программ, именно, во внутренней КЭШ-памяти.

Хотя процессоры уровня 486-х и выше имеют внутреннюю Кэш-память, так называемую КЭШ-память первого уровня (Level 1, L1), все системные платы предусматривают установку и внешней (второго уровня, L2) КЭШ-памяти до 256 Кбайт для 486-х систем и от 512 Кбайт для систем на основе Pentium. В качестве такой Кэш-памяти обычно используются микросхемы статической ОЗУ (SRAM) со временем доступа 15-20 нс. Благодаря наличию внешней Кэш-памяти работа динамической памяти с 1-2 циклами ожидания приводит к заметному понижению производительности, и, практически, во всех платах можно применять модули со временем выборки 70 нс.

Системные платы на основе процессоров Pentium могут использовать синхронную пакетную (synchronous burst) или конвейерную пакетную (pipelined burst) КЭШ-память, выполненную на элементах статической памяти. Пакетная синхронная статическая память отличается от обычной асинхронной тем, что для нее достаточно только первого адреса строки данных, т.к. все остальные получаются автоувеличением базового адреса по тактовым сигналам. В конвейерной же памяти плюс к этому для хранения пакета выбранных данных используются регистры-”защелки”, за счет чего основная матрица памяти микросхемы может быть менее быстродействующей, нежели в случае стандартной пакетной статической памяти. Это обуславливает и стоимость самой памяти. При использовании синхронной памяти быстродействие компьютера может возрасти примерно на 10-15%.

4.2. Общая характеристика систем передачи данных

Любую систему передачи данных можно представить как совокупность трех составных частей: передатчик, приемник и канал, по которому непосредственно осуществляется передача информации между ними.

Для соединения персональных компьютеров с другими компьютерами используются три основных типа канала. Первый из них, называемый *выделенным* (dedicated) каналом, представляет собой линию прямого соединения, которая соединяет передатчик с приемником, расположенные в одном здании, конторе, промышленном предприятии и т.п. Передача осуществляется по одному и

тому же пути и, при необходимости, можно легко увеличить производительность канала. Выделенным каналом может быть проводная или локальная шина.

Второй тип канала, называемый *арендованным* (частным), организуется компанией по обеспечению связи, которая устанавливает связь между двумя учреждениями, расположенными в различных зданиях.

Последний тип канала носит название *коммутируемый* канал. Его принцип работы основан на использовании телефонной сети общего пользования. Если пользователь персонального компьютера хочет получить доступ к другому компьютеру, он набирает номер телефонной линии, которая подсоединена к этому компьютеру, затем переключающие центры телефонной компании осуществляют соединение, и устройства на концах линии начинают передачу данных. После завершения передачи связь, установленная при соединении, разрывается, и все линии связи восстанавливаются таким образом, что они становятся доступными для других пользователей.

Скорость, стоимость и качество передачи являются основными факторами при выборе между арендованным или коммутируемым каналами. Раньше в коммутируемых каналах скорость передачи не превышала 9600 бит/с, сейчас же с помощью современных электронных переключателей скорость передачи может быть доведена до 28,8 Кбит/с (стандарт V.34), хотя уже существуют экспериментальные действующие линии, имеющие скорость передачи до 2 Мбит/с при протяженности линии, по крайней мере, до одного километра. В общем случае, если передачи данных носят случайный характер и длятся непродолжительное время, то, обычно, использование коммутируемых каналов дешевле. В противном случае, организуется арендованный канал. Если он имеет фиксированное расположение в системе, то это можно использовать для уменьшения ошибок при передаче, а также для локализации тех мест, где возможно их появление.

Некоторые ограничивающие факторы, существенные для определения типа канала при связи между устройствами, представлены в таблице 4.1.

Таблица 4.1. Рекомендации по выбору канала

Тип канала	Расстояние между передатчиком и приемником	Скорость передачи	Рекомендуемое использование
Выделенный (прямое соединение)	Ограничено территорией района	Ограничена используемыми проводами	Произвольная длительность передачи
Арендованный (частный)	Ограничено возможностями телефонной компании	Ограничена типом среды передачи	Вызовы с произвольной длительностью передачи
Коммутируемый (телефонная линия)	Ограничено возможностью телефонного доступа	Не выше 28,8Кбит/с (стандарт V.34)	Небольшая длительность передачи

Передачу и прием информации по всем трем типам каналов осуществляют сервисные и передающие устройства. К ним относятся цифровые устройства, модемы, акустические соединители, аналоговые устройства, системы телефонного обслуживания.

Цифровые устройства используют для передачи данных однополярные цифровые сигналы (рис. 4.5), т.е. такие сигналы, когда двоичные единицы представлены положительным напряжением, а двоичные нули – отсутствием напряжения. С увеличением расстояния между передатчиком и приемником форма цифровых сигналов сильно искажается, поскольку кабель, используемый в качестве среды передачи, имеет сопротивление и емкость. При некотором расстоянии между двумя устройствами форма сигналов настолько искажается, что они становятся нераспознаваемыми приемником (рис. 4.5,б).

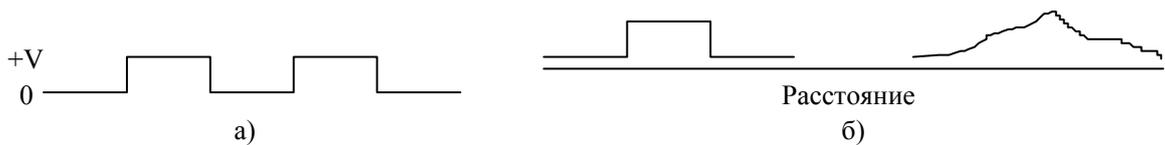


Рис. 4.5. Цифровой сигнал (а) и его искажения (б)

Для передачи цифровых данных на большие расстояния вдоль линии через определенные промежутки устанавливаются *повторители*, которые восстанавливают форму цифровых сигналов до первоначальной. Повторителем является такое устройство, которое при появлении импульса в линии регенерирует импульс первоначальной формы (рис. 4.6).

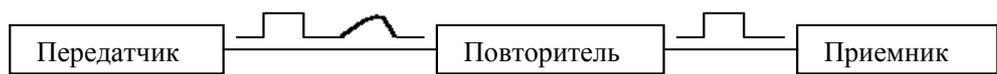


Рис. 4.6. Передача данных в цифровой форме

Для передачи однополярных импульсов на большие расстояния требуется увеличение постоянного напряжения. Чтобы избежать этого, в цифровых сетях осуществляет преобразование однополярных сигналов в модифицированные биполярные. Это требует установки на каждом конце линии специальных устройств, носящих название *блоков цифрового обслуживания* (digital service unit – DSU). На рисунке 4.7 показано применение блоков DSU для передачи данных в цифровых сетях.



Рис. 4.7. Передача данных в цифровой сети

Для передачи информации на большие расстояния, как уже отмечалось ранее, могут использоваться обычные телефонные сети. Единственной проблемой в этом случае является преобразование цифровых сигналов в аналоговые, т.к. большинство сетей аналоговые, хотя в настоящее время существуют и цифровые телефонные сети. Такое преобразование осуществляет специальное устройство – *модем*. Модемом называется электронное устройство, которое преобразует цифровые сигналы, вырабатываемые компьютером или терминальным устройством, в аналоговые, которые передаются по телефонной сети с помощью аналоговых аппаратных средств (рис. 4.8). Модем – это сокращение от составного термина МОдулятор-ДЕМОдулятор.

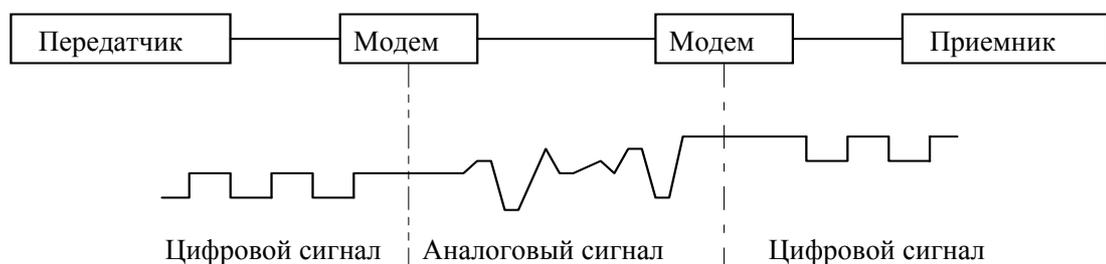


Рис. 4.8. Преобразование сигналов, выполняемое модемом

Очень популярные в начале 80-х годов у пользователей терминалами данных и персональными компьютерами *акустические соединители* в настоящее время достаточно редко применяются для передачи данных из-за невысокой скорости и низкого качества передачи.

Акустический соединитель – это модем, который подсоединяется к телефонной сети путем подключения к нему телефонного наушника (telephone headset). Основное достоинство акустического соединителя заключается в том, что он не требует жесткого проводного соединения с коммутируемой телефонной сетью, при этом обеспечивается мобильность терминалов и персональных компьютеров относительно передачи данных.

Акустический соединитель преобразует сигналы, вырабатываемые персональным компьютером, в последовательность звуковых тонов, которые через телефон передаются в коммутируемую телефонную сеть. Информация на другом конце линии связи, преобразуется обратно в звуковые тона и далее акустическим соединителем в соответствующие электрические сигналы, распознаваемые подключенным компьютером.

Взаимосвязь персональных компьютеров, акустических соединителей, модемов и среды передачи показана на рисунке 4.9.

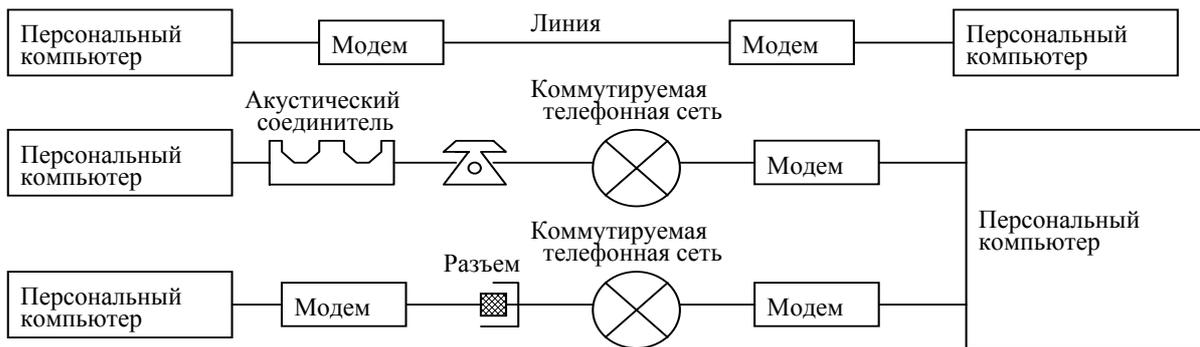


Рис. 4.9. Взаимодействие персональных компьютеров, модемов, акустических преобразователей и среды при аналоговой передаче

Изготовители связного оборудования предлагают несколько типов аналоговых устройств коммутации, отличающихся характеристиками и структурой. Аналоговые коммутируемые устройства характеризуются возможностью реализовать прямой телефонный вызов, телефонное обслуживание на большой территории и осуществлять связь с телефонными станциями других государств. Примером таких устройств может служить *международный коммутатор* (foreign exchange – FX), который обеспечивает передачу данных от группы терминальных устройств, удаленных от принимающего устройства, при стоимости, меньшей, чем при непосредственной связи по телефону. Канал FX можно рассматривать как сочетание аналогового коммутируемого и арендованного каналов.

При использовании этого канала необходимо получить разрешение на сеанс связи у местной телефонной станции. Из коммутатора FX информация передается далее через выделенный канал в коммутирующий центр, ближайший к устройству, с которым и запрашивалась связь (рис. 4.10). Линия местного коммутирующего центра, который заканчивается у пользователя, обслуживается системой FX.



Рис. 4.10. Обслуживание через международный коммутатор

Основное отличие канала FX от арендованного канала заключается в том, что любой персональный компьютер или терминал, обращающийся по каналу FX, обеспечивается вторым

модемом для передачи данных по каналу, в то время как арендованный канал, используемый обычно для передачи данных, должен иметь фиксированные модемы, подключенные к обоим концам канала.

В последнее время многие коммуникационные фирмы предлагают в дополнение к аналоговому также и цифровое обслуживание, при котором данные от источника до приемника передаются в цифровой форме по аналоговым каналам.

4.3. Модемы

В настоящее время выпускается огромное количество всевозможных модемов, начиная от простейших, обеспечивающих скорость передачи около 300 бит/с, до сложных факс-модемных плат. Но независимо от типа модема можно выделить некоторые основные этапы его работы. На первом этапе модем осуществляет прием данных, поступающих из компьютера, после чего разделяет их на исполняемые команды (командный режим) и информацию, которую необходимо передать в линию (режим обмена данными).

При включении питания, модем находится в командном режиме. Далее происходит переключение в режим передачи данных только в следующих случаях:

- при удавшейся попытке установления связи с другим модемом;
- при выполнении модемом процедур самотестирования.

Модем переходит из режима передачи данных в командный режим в следующих случаях:

- после неудачной попытки связаться с удаленным модемом;
- при потере несущей в течение передачи данных;
- при поступлении модему от компьютера специальной Escape-последовательности.

Режим работы, когда передача данных может осуществляться только в одном направлении, называется *полудуплексным* (half duplex). Но, как правило, оба компьютера могут одновременно обмениваться информацией в обоих направлениях. Этот режим работы называется *полным дуплексом* (full duplex) или просто дуплексом.

Для осуществления связи один из модемов вызывает другой по номеру телефона. Если вызов удачный, то начинается обмен сигналами между модемами.

подавляющее большинство современных модемов поддерживают разработанный фирмой Hayes Microcomputer Product набор AT-команд управления модемами. Такие модемы называются Hayes-совместимыми.

Модемы также могут отличаться друг от друга методами модуляции цифрового сигнала. Наиболее применяемыми являются три метода модуляции: FSK (Frequency Shift Keying), PSK (Phase Shift Keying) и QAM (Quadrature Amplitude Modulation). FSK является разновидностью частотной модуляции, а PSK – фазовой. В методе квадратурной амплитудной модуляции QAM одновременно изменяются фаза и амплитуда сигнала, что позволяет передавать большее количество информации, чем при применении двух первых методов модуляции. В современных модемах чаще всего используется модуляция с решетчатым кодированием информации TCQAM (Trellis Coded QAM).

Кроме собственно модуляции и демодуляции сигналов модемы могут выполнять сжатие и декомпрессию пересылаемой информации, а также заниматься поиском и исправлением ошибок, которые могут возникнуть в процессе передачи данных по линиям связи.

Аппаратно модемы выполнены либо как отдельная плата, вставляемая в слот расширения (внутренние модемы), либо в виде отдельного корпуса с блоком питания, который подключается к последовательному асинхронному порту компьютера (внешние модемы). К телефонной линии модем подключается либо непосредственно (через RJ11-с), либо при помощи микрофона и динамика с присосками к обычной телефонной трубке (акустические модемы). Последний способ подключения модема телефонной линии в настоящее время практически не используется из-за невысокого качества передачи и малой скорости (порядка 300 бит/с).

Типичный модем содержит следующие компоненты: специализированный микропроцессор, управляющий работой модема; оперативную память, хранящую значения регистров модема и буферизующую входную/выходную информацию; постоянную память;

динамик, позволяющий выполнять звуковой контроль связи, а также другие вспомогательные элементы (трансформатор, резисторы, конденсаторы, разъемы). Внутренние модемы также имеют ряд DIP-переключателей и перемычек, которые позволяют установить номер асинхронного последовательного порта ввода/вывода и линию прерываний IRQ, используемую данным модемом. В настоящее время большинство модемов содержит электрически перепрограммируемую постоянную память, в которой сохраняется конфигурация модема.

Существует множество коммуникационных стандартов модема. Стандарт в данном случае включает в себя общепринятый протокол или набор протоколов. Под протоколом понимается совокупность правил, регламентирующих формат и процедуры обмена информацией. В 1964 году крупнейшие производители модемов договорились установить соответствующих стандартов международной организации ССИТТ (сокращение французского названия, которое можно перевести как международный консультативный комитет по телеграфии и телефонии – МККТТ). Сегодня эта организация переименована в Международный союз электросвязи (International Telecommunications Union – ITU). Практически все стандарты, касающиеся модемов, установлены именно этой организацией (таблица 4.2).

Таблица 4.2. Стандарты модемов

Стандарт	Организация	Скорость передачи, бит/с
Bell 103	AT&T	300
Bell 212A	AT&T	12000
V.21	ССИТТ	300
V.22	ССИТТ	1200
V.22 bis	ССИТТ	2400, 1200
V.32	ССИТТ	9600, 4800
V.32 bis	ССИТТ	14400, 12000, 9000, 7200, 4800
V.34	ССИТТ	28800

До принятого в 1964 году соглашения телефонная компания AT&T разработала два модема (Bell 103 и Bell 212A), которые фактически стали стандартами для модемной связи при скорости передачи 300 и 1200 бит/с. В Bell 103 применялась FSK модуляция, а в Bell 212A – DFSK (дифференциальная частотная модуляция).

Стандарты V.21 и V.22 аналогичны стандартам Bell 103 и Bell 212A, но они распространены только за пределами США и Канады. А стандарт V.22 bis уже стал международным. Он специфицировал использование QAM, причем каждая модуляция (бод) соответствовала одному из 16 возможных состояний сигнала (по фазе и амплитуде), для кодирования которых необходимы 4 бита.

Стандарт V.32 был основан на модуляционном методе TCQAM, который заключается в добавлении еще одного контрольного разряда к пакету из четырех бит данных, что позволило выполнять немедленную коррекцию ошибок. Кроме того, в стандарт был добавлен метод эхо-компенсации, позволяющий использовать полную полосу частотного сигнала. В 1996 году стандарт V.32 bis стал расширением V.32.

Помимо модемов, производители которых придерживаются международных стандартов, существуют и нестандартные модемы, которые для больших скоростей передачи используют свои собственные (proprietary) протоколы. Это связано с тем, что до недавнего времени максимальная скорость передачи в соответствии с V.32 bis ограничивалась значением 14,4 Кбит/с. Примером таких нестандартных модемов могут служить достаточно популярные в нашей стране скоростные модемы фирмы ZyXEL, которые могут работать с использованием собственного протокола на скоростях 16,8 и 19,2 Кбит/с.

Летом 1994 года союз ITU принял стандарт V.34, который описывает протоколы передачи данных со скоростью 28,8 Кбит/с. Это практически предельная граница скорости на существующих телефонных линиях. Новый стандарт закрепляет ряд новых методов коррекции, кодирования и управления уровнем сигнала. Модемы, соответствующие V.34, поддерживают гораздо больше методов модуляции со скоростью модуляции 2400 бод (число модуляций в

секунду), чем предыдущие стандарты. Одним из нововведений стало 4-мерное решетчатое кодирование сигналов. Чтобы избежать потерь во времени, возникающих из-за большого количества поддерживаемых методов модуляции, при соединении модемов сначала определяется список функциональных возможностей, которые являются для этих модемов общими, затем с помощью специального сигнала тестируются характеристики используемой телефонной линии, после чего приемо-передающие узлы модемов начинают увеличивать скорость передачи до максимально возможной. Модемы стандарта V.34 создаются уже с учетом цифровых телефонных сетей нового поколения.

В течение многих лет протокол MNP (Microcom Network Protocol) фактически являлся стандартом на сжатие данных и коррекцию ошибок. Этот протокол был предложен и реализован фирмой Microcom. Данный протокол дополнял стандарты CCITT, а иногда вообще заменял их.

Этот протокол делится на двенадцать классов, первые четыре из которых определяют методы передачи данных и коррекцию ошибок. Принцип работы модема с использованием MNP-протокола (класс 4) заключается в передаче блоков информации переменной длины. В соответствии с этим протоколом модем принимает от компьютера данные и собирает их в пакет, который затем передается по линии другому модему. При сборке пакета вычисляется контрольная сумма, которая передается в самом конце пакета. Когда принимающий модем получает пакет, он заносит его в буфер данных и снова вычисляет контрольную сумму. Если эти две суммы совпадают, то пакет пересылается в компьютер и начинается передача нового пакета, если нет, то повторяется передача пакета, в котором была допущена ошибка.

MNP уровня 5, который предполагает наличие MNP уровня 4, обеспечивает возможность сжатия информации. Но его существенным недостатком является то, что он не проверяет, сжаты ли уже посылаемые данные передающим компьютером.

Такого недостатка лишен появившийся два года назад, но практически уже общепризнанный стандарт V.42 bis. Технологии, предложенные в стандарте V.42 bis, позволяют производить сжатие информации почти в четыре раза, тогда как MNP 5 – только в два.

Обычно при передаче текста, таких коммуникационных параметров, как число битов данных, битов синхронизации и четности, достаточно. Данные посылаются и принимаются посимвольно. Но иногда модемы используются для передачи файлов, а не отдельных символов.

Наиболее фундаментальным протоколом передачи данных является ASCII-протокол. Этот протокол поддерживается практически всеми компьютерами, но он применяется только для обмена текстом, но не программами. Этот протокол работает без коррекции ошибок. В результате при передаче файлов по телефонным линиям из-за больших уровней шума принятый файл может сильно отличаться от передаваемого.

В 1977 году Вардом Кристенсенем был разработан протокол XMODEM. Также как и предыдущий, этот протокол является полудуплексным. Файл разделяется на блоки длиной по 128 байт, и к каждому блоку добавляется контрольная сумма. Принимающий компьютер вычисляет свою контрольную сумму и, если эти суммы совпадают, передает сигнал подтверждения ACK исходному компьютеру, а тот в свою очередь начинает передачу следующего блока. В противном случае исходный компьютер получает сигнал переспроса NACK и затем повторно отправляет блок и его контрольную сумму. С последним блоком передающий компьютер посылает код EOT, который означает конец текста. Этот метод обеспечивает достаточно хорошую защиту от ошибок, поэтому не целесообразно применять его с протоколами коррекции ошибок, например, с MNP-протоколами.

К настоящему времени разработаны несколько вариантов основного протокола XMODEM. Наиболее удачным из них является протокол ZMODEM. Этот протокол обеспечивает высокую скорость передачи данных, используя окна. ZMODEM осуществляет передачу данных пакетами по несколько штук в окне. При этом принимающий данные компьютер не передает сигнал подтверждения или сигнал переспроса неправильного пакета, пока не получит все пакеты в окне. Если при передаче файла на линии произошел сбой, то в следующий раз при передаче этого же файла он автоматически начинается передаваться с того места, где произошел обрыв связи.

Наряду с текстовой информацией очень часто приходится передавать графические изображения и рукописный текст, особенно это актуально в Японии, где большинство документов пишутся от руки. Для этих целей используется система, получившая название *телефакса*, которая обеспечивает электронную передачу обычного текста, чертежей, фотографий и схем, она

обеспечивает сканирование документа на передающей стороне, осуществляет преобразование в аналоговые сигналы и обратно, а также формирует на бумажном носителе дубликат (факсимиле) исходного документа на приемной стороне. То есть в состав любого телефакса входят: сканер для считывания документа, модем, передающий и принимающая информация, а также принтер, печатающий принимаемое сообщение на бумаге. Для телефаксов также существует набор стандартов на модуляцию, сжатие, кодирование и коррекцию ошибок, принятый CCITT, под общим названием Group (G1-G4).

В качестве примера применения модемов можно также привести *электронные доски объявлений* (BBS). BBS обычно представляют собой компьютеры, которые конфигурируются с определенным программным обеспечением и модемом и откликаются электронным способом при наборе их номера. Большинство BBS содержат сообщения и файлы, которые можно копировать на ПК, подключенный к BBS. Процесс копирования обычно называется выгрузкой. Можно также передавать свои сообщения BBS.

Очевидно, что для нормальной работы перечисленных выше примеров необходима правильная организация сетей, т.е. физических элементов связи между передающими и принимающими устройствами.

4.4. Организация сетей

Все компьютерные сети подразделяются на локальные и глобальные вычислительные сети.

Упрощенно *локальную вычислительную сеть* (Local Area Network – LAN) можно рассматривать как электронную магистраль для передачи информации между различными устройствами, подключенными к этой сети. Внутри здания или в пределах небольшой территории LAN позволяет соединить между собой группу ПК для совместного использования информации. К преимуществам локальных сетей можно также отнести возможность совместного использования периферийных устройств, совместимость и рассредоточение оборудования. На основе этого можно сделать вывод, что в определенных случаях гораздо выгоднее организовать сеть, чем создать полную конфигурацию персональной компьютерной системы для каждого пользователя.

LAN – это совокупность компьютеров, кабелей, сетевых адаптеров, работающих под управлением сетевой операционной системы (NOS – Network Operating System) и прикладного программного обеспечения. Компьютеры, входящие в LAN, делятся на два типа: *рабочие станции* (PC), предназначенные для пользователей, и *файловые серверы* (ФС), как правило, недоступные для обычных пользователей. Каждая рабочая станция и каждый файловый сервер содержат карты адаптеров, которые с помощью сетевых кабелей соединяются друг с другом.

В качестве рабочих станций в большинстве случаев используются ПК среднего класса типа IBM PC/AT на базе процессоров 80286 или 80386. Они оснащены ОЗУ емкостью от одного до четырех Мбайт. Обычно рабочие станции имеют также недорогой HDD небольшого объема. Хотя некоторые рабочие станции, называемые *бездисковыми рабочими станциями*, вообще не имеют дисководов, и для них доступ к файлам полностью определяется локальной сетью. На каждой рабочей станции в дополнение к локальной операционной системе (обычно DOS) активизируется сетевое программное обеспечение, позволяющее станциям взаимодействовать с файловым сервером для чтения и записи файлов. Поэтому при включении на экране во время загрузки операционной системы появляются дополнительные сообщения, которые указывают на то, что сетевая операционная система загружается в рабочую станцию. Затем для входа в систему необходимо сообщить сетевому программному обеспечению идентификационный номер (ID) и пароль (password).

Файловый сервер – это компьютер, который обслуживает все рабочие станции. Он осуществляет совместное использование файлов, размещенных на его дисках. Файловые серверы – это, обычно, быстродействующие компьютеры на базе микропроцессоров 486 и Pentium с высокой тактовой частотой, с объемом ОЗУ более 16 Мбайт. Файловые серверы часто оснащены только монохромным монитором и недорогой клавиатурой, но почти всегда содержат один и более быстродействующих накопителей большой емкости. Файловый сервер может использовать иную, нежели рабочие станции, операционную систему. В последнее время при небольшом количестве

рабочих станций в качестве операционной системы на сервере используют Windows NT, а на самих рабочих станциях Windows 95. Иногда при небольшом количестве рабочих станций в сети файловый сервер также используется как и рабочая станция, но при увеличении числа пользователей обслуживание всей компьютерной сети не оставляет ресурсов для использования файлового сервера в качестве рабочей станции. К тому же, если пользователь такой рабочей станции выключит ее, то вся сеть также будет выключена.

Средой передачи, используемой для построения локальных вычислительных сетей, могут быть тонкие коаксиальные кабели (ThinNet или CheaperNet), толстые коаксиальные кабели (ThickNet), экранированные витые пары (STR – Shielded Twisted Pair) и неэкранированные витые пары (UTR – Unshielded Twisted Pair). Применяются даже оптоволоконные кабели, которые могут работать на больших расстояниях и с большей скоростью, чем другие типы кабелей. Однако их прокладка и сетевые адаптеры для них довольно дороги. Тип используемого кабеля в значительной степени зависит от карты сетевого адаптера, выбранного вами, а также от основных характеристик различных видов кабелей – стоимости и простоты установки, ширины полосы пропускания кабеля, определяющей число одновременных сеансов передачи, максимально допустимой скорости передачи и географического размещения сети.

Геометрическое расположение кабеля для объединения станций и серверов в сети, называемое *топологией*, не может быть произвольным, как в случае сетей передачи данных, которые конфигурируются самым разнообразным образом путем добавления аппаратных и программных средств. Для построения локальных сетей используют три основных типа топологий. Первый из которых называется *магистралью* или *дейзи-цепью* (рис. 4.11). При такой топологии единый кабель проходит от узла к узлу, последовательно соединяя все рабочие станции и файловый сервер. Каждая станция может связываться по общему кабелю с любой другой станцией, поэтому необходимы определенные протоколы для управления такими станциями в случаях когда, например, две станции одновременно пытаются получить доступ к основной магистрали данных. Иногда кабель, являющийся главной магистралью, имеет форму кольца. Таким образом, кольцо можно рассматривать как замкнутую магистраль.

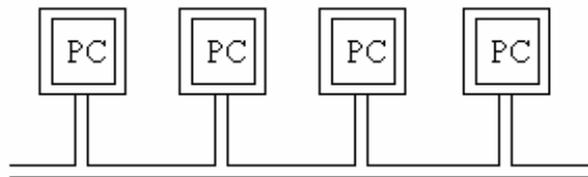


Рис. 4.11. Топология линейной магистрали, в которой все рабочие станции подсоединены к общему кабелю

Второй тип топологии называется *звездой* (рис. 4.12). В этом случае к каждой рабочей станции подходит собственный кабель из центрального узла, например, файлового сервера. Доступ от одной станции к другой может осуществляться только через центральное устройство.

Третий тип топологии называется *звездообразным деревом*. В такой сети кабели периодически разветвляются в узловых точках. В каждой точке ветвления необходимо использовать специальные соединители для расходящихся кабелей.

Некоторые сети имеют смешанную топологию. Например, сеть IBM Token Ring Network можно рассматривать как “звездообразное кольцо”, поскольку сначала до восьми персональных компьютеров подсоединяются к общему устройству, которое носит название *Блок многостанционного доступа* (Multistation Access Unit – MAU), а затем эти MAU соединяются друг с другом в соответствии с кольцевой топологией.

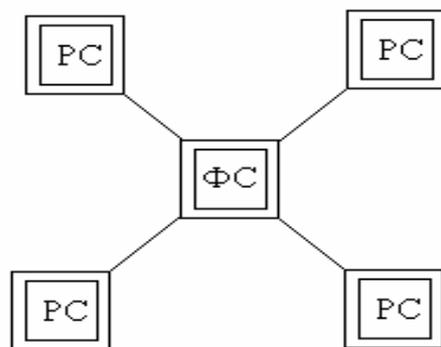


Рис. 4.12. Топология звезды, в которой все рабочие станции и сетевые устройства присоединены с помощью кабелей, радиально расходящихся от файлового сервера

Карты сетевых адаптеров, также как и карты адаптеров дисплеев, устанавливаются в каждой рабочей станции и в файловом сервере. Рабочая станция отправляет запрос через сетевой адаптер к файловому серверу и получает ответ через сетевой адаптер, который готов передать очередную порцию файла. Такие запросы и ответы аналогичны чтению и записи файлов на диск в обычном персональном компьютере.

Одновременно два сетевых адаптера могут обмениваться информацией друг с другом, в то время как другие рабочие станции должны ждать своей очереди. Такие задержки обычно незаметны, поэтому при работе в LAN создается впечатление, что доступ всех пользователей к файловому серверу осуществляется одновременно.

На картах адаптеров LANtastic имеются два разъема, один из которых предназначен для входного кабеля, а другой для выходного. На картах адаптеров EtherNet имеются: либо T-образный разъем, либо 15-контактный разъем, по форме напоминающий букву D, либо разъем типа телефонной розетки, а иногда – комбинация всех трех перечисленных разъемов. На картах адаптеров Token Ring имеется 9-контактный разъем и в ряде случаев еще дополнительный телефонный разъем.

Карты с двумя или более типами разъемов предоставляют возможность выбора из широкого ассортимента сетевых кабелей. Например, карта адаптера Token Ring с двумя разъемами допускает использование в качестве сетевого кабеля экранированной витой пары или неэкранированной витой пары (или телефонного провода).

Карта адаптера принимает все сообщения, которые передаются по кабелю, и отбирает из них только те, которые непосредственно адресованы данной рабочей станции. Полученное сообщение задерживается адаптером до тех пор, пока рабочая станция не будет готова принять это сообщение. В случае, когда рабочая станция собирается послать запрос на файловый сервер, адаптер дожидается перерыва в потоке данных (*трафике*) сети и выставляет свое сообщение в образовавшееся окно. Адаптер также автоматически производит проверку того, что сообщение было передано без ошибок и если это не так, то передача повторяется.

Для измерения быстродействия сети в технике введена специальная величина – *мегабит в секунду*. Например, при передаче содержимого 3,5 дюймовой дискеты емкостью 1,44 Мбайт максимальное быстродействие сети равно четырем мегабитам в секунду. Это сравнимо со скоростью передачи среднего HDD.

На практике максимальное быстродействие сети никогда не реализуется. Фактически LAN не может работать быстрее, чем самая медленная ее компонента. Запросы других пользователей сети, работающих одновременно с вами, будут пересекаться с вашими запросами, и суммарное время передачи будет еще больше. Таким образом, рабочая станция не может передавать данные по сети быстрее, чем она считывает их с дискет или с диска.

Адаптеры EtherNet

LAN, использующие адаптеры EtherNet, могут связывать широкий спектр систем, включая UNIX-компьютеры, компьютеры Macintosh фирмы Apple, ПК фирмы IBM и совместимые с ними,

производятся множеством конкурирующих между собой компаний и имеют три модификации (ThinkNet, UTR и ThickNet), в зависимости от толщины используемого кабеля. Адаптеры модификации ThinkNet могут работать на большие расстояния, чем другие, но они значительно дороже. Скорость работы сети с EtherNet равна 10 Мбит/с.

Между пересылками данных (запросов к серверу и ответов от него) LAN с адаптерами EtherNet находится в состоянии покоя. После передачи рабочей станцией сообщения по сетевому кабелю он снова освобождается.

Если одна из рабочих станций пытается запросить что-то из сервера, в то время как сервер посылает ответ другой рабочей станции, то возникает конфликт (только два компьютера могут обмениваться информацией через кабель). В этой ситуации оба компьютера – сервер и станция – прекратят обмен и попытаются повторить передачу. Для устранения конфликтов сетевые адаптеры EtherNet используют механизм, называемый CSMA/CD (Carrier Sense, Multiple Access/Collision Detection – Множественный доступ с обнаружением несущей и устранением конфликтов). Его суть заключается в том, что сервер и станция переходят в режим ожидания в течение случайного промежутка времени, после чего повторяют попытки обмена. Этот метод очень эффективен при не слишком большом трафике ЛВС и позволяет одному из компьютеров опередить остальные. С увеличением трафика сети частота конфликтов увеличивается, и время реакции на запросы ухудшается. Фактически сети на базе EtherNet начинают тратить больше времени на разрешение конфликтов, чем на передачу данных. Фирмы IBM и Texas Instruments, имея в виду эту проблему, разработали для ее преодоления LAN с адаптерами Token Ring.

Адаптер Token Ring

LAN с адаптерами Token Ring являются самыми дорогостоящими, за исключением лишь сетей с оптоволоконными кабелями/адаптерами. Token Ring может использовать кабели с экранированными и неэкранированными витыми парами. Затраты на сеть Token Ring оправданы, если она имеет напряженный трафик работы с большим количеством рабочих станций, особенно если она подключена к большой ЭВМ, например в крупных компаниях с большой ЛВС. ЛВС Token Ring работает при скорости 4 или 16 или 100 Мбит/с.

В сети Token Ring даже в отсутствии трафика все рабочие станции передают друг другу специальный сигнал, называемый *маркером* (token). Маркер – это очень короткое сообщение, являющееся признаком того, что сеть свободна.

Если рабочей станции нечего передать во время приема маркера, то она просто передает маркер следующей по очереди станции. Рабочая станция может передать свое сообщение в сеть только в том случае, если она получила маркер. Это сообщение будет циркулировать через другие рабочие станции и серверы, пока не вернется к его отправителю. Тогда отправитель посылает в сеть маркер, для того чтобы показать, что сеть освободилась снова. Во время циркуляции сообщения одна из рабочих станций (или файловый сервер) распознает, что сообщение адресовано ей и начинает его обрабатывать. Это занимает не так много времени, как могло показаться из описания процедуры, т.к. маркеру для прохождения по сети требуется очень мало времени, даже если сеть состоит из очень большого числа рабочих станций (порядка 200). Кроме того, имеется возможность присвоить приоритеты определенным рабочим станциям и файловым серверам таким образом, чтобы они получали более частый доступ к LAN. В результате схема с передачей маркера гораздо более устойчива к высокому уровню трафика, чем схема с обнаружением конфликтов типа EtherNet.

Иногда станции пропускают маркер или теряют его. Поэтому они следят друг за другом и используют сложную процедуру для регенерации потерянного маркера. Token Ring несколько сложнее, чем EtherNet, и поэтому карты сетевых адаптеров для нее соответственно дороже.

Адаптер LANtastic

Фирма Artisoft наряду с адаптерами EtherNet производит также адаптеры собственной разработки. Собственная модель адаптера фирмы Artisoft называется LANtastic, что вносит некоторую путаницу, т.к. фирма Artisoft производит сетевую операционную систему с таким же названием. Адаптер LANtastic работает со скоростью 2 Мбит/с и использует четырехпроводный

кабель, который поочередно соединяет рабочие станции в сети друг с другом. Процедура установки такой сети достаточно проста, если нет необходимости протягивать кабель сквозь стены или потолок. Заметим, что адаптеры типа EtherNet или Token Ring соответствуют общепринятым промышленным стандартам, в то время как адаптеры LANtastic являются оригинальной разработкой, поэтому большинство пользователей при создании новой локальной сети выбирают именно EtherNet или Token Ring.

Адаптеры ARCnet

Адаптер ARCnet является одним из старейших представителей сетевых аппаратных средств. Вначале он был собственным фирменным продуктом компании Datapoint Corporation, но в настоящее время многие фирмы выпускают ARCnet-совместимые карты. Адаптеры ARCnet имеют невысокое быстродействие, но зато они не критичны к небольшим ошибкам при установке. Они известны своей высокой надежностью, а возникшие проблемы с кабелем и адаптерами легко поддаются диагностике. Адаптеры ARCnet дешевле, чем EtherNet, и работают примерно по такому же принципу, как адаптеры Token Ring, но при гораздо меньшей скорости, равной 2,5 Мбит/с.

Хотя адаптеры ARCnet и Token Ring несовместимы друг с другом, тем не менее ARCnet использует похожую схему передачи маркера для управления доступом рабочей станции или сервера к сети.

Первый шаг на пути создания локальной вычислительной сети – объединение группы компьютеров на постоянной основе. На этой ступени не нужно отказываться от обычной работы на ПК, который будет служить файловым сервером. Фактически любой компьютер может быть одновременно и файловым сервером, и рабочей станцией. Такая конфигурация называется *одноранговой локальной вычислительной сетью*.

В одноранговой ЛВС дисковое пространство и файлы на каждом компьютере становятся общими. Главным преимуществом одноранговых сетей по сравнению с компьютерами, не объединенными в сеть, является то, что отпадает необходимость копировать файл для того, чтобы им могли пользоваться другие. В зависимости от установленной защиты данных, другие пользователи смогут пользоваться этими данными сразу же после их создания.

В противоположность одноранговому LAN сети на базе файл-серверов (*локальные вычислительные сети с выделенным сервером*) имеют лучшие характеристики и повышенную надежность. Здесь один из компьютеров с той же самой сетевой операционной системой можно использовать только в качестве файл-сервера. Отметим, что файловый сервер владеет главными ресурсами сети, к которым обращаются остальные ПК (рабочие станции).

При работе в LAN с выделенным сервером отпадает необходимость разделять дисковый накопитель на отдельно взятой рабочей станции с остальными пользователями. При разделении файла необходимо скопировать его на файл-сервер (или создать его там). Преимуществом данной организации сети является возможность перезагружать каждую рабочую станцию, не беспокоясь о том, что будут испорчены чьи-либо файлы. Организация дискового пространства на локальном дисковом накопителе целиком находится в руках пользователя.

В определенных случаях вообще нет необходимости устанавливать сеть. Например, когда нужно разделять между несколькими компьютерами принтер, но нет необходимости в обмене файлами. Для этой задачи вполне достаточно переключающего устройства для принтера. Наиболее простое из таких устройств называется А/В переключатель, т.к. для получения доступа к принтеру пользователь должен переключить его из положения А в положение В. Переключатель имеет два или более входных разъемов для подключения к параллельным и последовательным портам компьютера и один разъем для подключения к принтеру.

Как разновидность данной схемы существуют устройства, которые могут определять, какой из компьютеров пытается связаться с принтером и автоматически производить коммутацию. Когда два компьютера одновременно пытаются что-нибудь напечатать, то такие автоматические устройства не всегда правильно распознают управляющие сигналы. Однако, подобные устройства могут быть очень полезны, когда ПК, подключенные к одному принтеру, находятся в разных комнатах. Некоторые производители лазерных принтеров не рекомендуют применять подобные переключатели для работы с их изделиями.

В одноранговой LAN каждый компьютер может выполнять одновременно функции рабочей станции, файлового сервера и сервера печати.

При средней и высокой интенсивности использования принтера в LAN необходим отдельный сервер печати. Это особенно оправдано, если в сети имеется более одного принтера или же работает высокоскоростной лазерный принтер (такой, как Hewlett-Packard Ipsi). Персональный компьютер может иметь до трех принтерных портов (LPT1, LPT2 и LP3). Если в сети используется не один принтер, то разные номера портов связываются с разными принтерами, а пользователь просто сообщает прикладной программе, какой принтерный порт использовать для печати.

При использовании локальной вычислительной сети нередко возникает необходимость связи с удаленными компьютерами (например, с BBS). Существуют три варианта работы сети с внешними линиями связи. В первом случае, если связь с другими компьютерными системами носит единичный характер, то достаточно простого переключающего устройства последовательного порта RS-232. Это устройство работает так же, как переключатель принтера. Последовательные порты каждой рабочей станции через кабели соединяются с переключателем, а он в свою очередь с помощью кабеля подсоединяется к модему.

В случае, когда пользователям необходимо часто выходить на модем, можно установить набор модемов и коммуникационный сервер. Это устройство позволяет разделять модемы и линии связи пользователям сети.

Другая возможность работы сети с внешними линиями связи – это использование отдельного компьютера в качестве *шлюза*. Компьютерный шлюз подключается к головному компьютеру и обеспечивает одновременно множество сеансов связи с ним.

4.5. Протоколы компьютерных сетей

В любой локальной сети передачу и прием сообщений с помощью кабелей осуществляют сетевые адаптеры в виде пакетов данных. В связи с этим существует множество протоколов обмена информацией. В настоящее время существуют несколько уровней организации протоколов.

На самом низком уровне компьютеры в сети обмениваются пакетами сообщений. Сетевой адаптер осуществляет прием и передачу пакетов под управлением соответствующего программного обеспечения. Пакеты адресуются рабочим станциям, каждая из которых имеет свой адрес в LAN.

Пакеты несут различную информацию, которую можно условно разделить на несколько этапов:

- начало сеанса обмена данными;
- передача данных другому ПК;
- подтверждение приема пакета данных;
- передача широковещательного сообщения всем адаптерам;
- конец сеанса обмена данными.

На рисунке 4.13 показано, как выглядит типичный пакет. В различных системах компьютерных сетей пакеты определяются по-разному, но элементы, являющиеся общими для всех, следующие:

- уникальный адрес отправителя;
- уникальный адрес получателя;
- признак, определяющий содержимое пакета;
- данные или сообщение;
- контрольная сумма (CRC) для обнаружения ошибок при передаче.

Адрес отправителя	Адрес получателя	Тип пакета	Данные/Сообщение	CRC
-------------------	------------------	------------	------------------	-----

Рис. 4.13. Базовая схема пакета сообщений

В настоящее время получил широкое распространение принцип наложения протоколов (рис. 4.14). Согласно этому принципу протокол низкого уровня, управляя адаптером, выполняет функции передачи сообщения, а протокол высокого уровня обслуживает файл-сервер и перенаправляет файлы.

Протокол наивысшего уровня в процессе перенаправления файлов в другой ПК передает сообщение протоколу среднего уровня, протокол среднего уровня создает оболочку-конверт (на рисунке 4.14 каждая оболочка обозначена двумя словами: заголовок и окончание) вокруг этого сообщения и передает его протоколу наинизшего уровня, поддерживающему функционирование адаптеров, который в свою очередь создает оболочку, включающую сообщение, полученное с предыдущего уровня. При приеме сообщений на каждом уровне внешняя оболочка удаляется программным обеспечением, после чего производится передача сообщения на более высокий уровень.

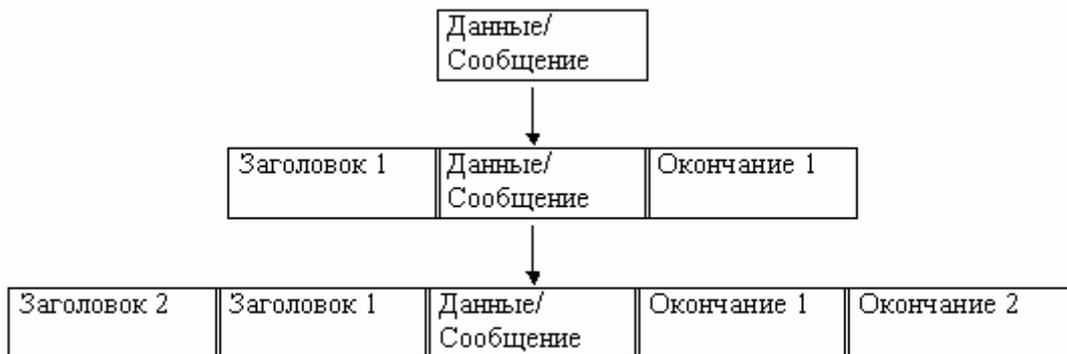


Рис. 4.14. Несколько уровней пакетов сообщений

Различные изготовители компьютерных сетей по-разному распределяют функции обслуживания передачи файлов между компьютерами сети. Поэтому, чтобы исключить возможные конфликты, организация ISO (International Standards Organization – Международная организация по стандартизации) опубликовала модель архитектуры вычислительной сети, названной OSI (Open System Interconnection – Связь открытых систем). Большинство производителей стремятся придерживаться модели OSI, но до сих пор нет изделий, полностью ей удовлетворяющих. Модель OSI разделяет коммуникационные функции в LAN на семь уровней. Однако большинство производителей сетевых операционных систем применяют три или четыре уровня протоколов.

Связь между компонентами сети осуществляется по двум основным принципам – обнаружение конфликтов и передача маркера. В компьютерной сети, работающей по схеме обнаружения конфликтов (CSMA/CD – Множественный доступ с проверкой несущей и обнаружением конфликтов), адаптеры непрерывно находятся в состоянии прослушивания сети. При возникновении конфликта (например, две рабочие станции одновременно начали передачу сообщений файловому серверу) адаптеры самостоятельно повторяют передачу. Время, затрачиваемое на обнаружение и обработку конфликтов, обычно не превышает микросекунды. Примером систем на базе принципа обнаружения конфликтов является сеть Ethernet.

Впервые экспериментальная LAN Ethernet, послужившая основой для данного стандарта, была создана в исследовательском центре Пало Альто фирмы Хегох в 1975 году. Современная версия Ethernet 2.0 была принята в 1982 году.

В LAN Ethernet узлы передают сообщения со скоростью 10 Мбит/с. Все узлы принимают каждое сообщение, но только адресат посылает в LAN подтверждение о его приеме.

Институт инженеров по электротехнике и электронике (IEEE) определил и опубликовал в качестве стандарта наборы физических характеристик, которыми должны удовлетворять сети с обнаружением конфликтов. Он получил название IEEE 802.3, как первая версия стандарта, и была опубликована в 1985 году. Различия между двумя стандартами находятся в области архитектуры LAN и форматов пакетов. Данные различия могут приводить к несовместимости этих стандартов.

Другой принцип (передача маркера), как уже отмечалось ранее, использует стандарт Token Ring. Организация IEEE также определила в качестве стандарта аналог стандарта Token Ring под названием IEEE 802.5. Эти два стандарта являются полностью совместимыми.

Более современным протоколом, чем Ethernet и Token Ring, является оптоволоконный интерфейс распределения данных (FDDI – Fider Distributed Data Interface). Разработанный специалистами Американского национального института стандартов (ANSI) протокол FDDI по схеме передачи маркера в логическом кольце с оптоволоконными кабелями и имеет производительность 100 Мбит/с. Этот протокол практически соответствует стандарту IEEE 802.5 Token Ring. Различия имеются только там, где это необходимо для реализации большей скорости обмена данными и способности перекрытия больших расстояний передачи. Для этого в FDDI используется схема 4B/5B, кодирующая 4 бита данных в пятибитовые комбинации световых импульсов, уменьшая таким образом число сигналов, необходимых для передачи одного байта информации. В Token Ring для передачи одного бита информации требуется чередование двух оптических сигналов – импульса света и паузы.

Существует два основных отличия в работе с маркером в протоколах FDDI и IEEE 802.5 Token Ring. В IEEE 802.5 новый маркер передается рабочей станцией в LAN только после возвращения к ней сообщения, которое она передала в сеть. В FDDI новый маркер будет передан непосредственно после передачи пакета сообщения. Кроме того, FDDI не использует областей приоритета и резервирования, которые применяются в IEEE 802.5 для выделения ресурсов системы. Вместо этого каждая рабочая станция в FDDI классифицируется как *асинхронная* (т.е. не предъявляющая жестких требований к интервалам времени, через которые возможен доступ к сети) и *синхронная* (которая предъявляет очень жесткие требования по отношению к интервалам времени между передачами). FDDI использует сложный алгоритм для предоставления доступа к сети этим двум классам устройств.

Все ранее рассмотренные протоколы относятся к протоколам низкого уровня, т.к. они имеют дело только с передачей сообщений (пакетов) между компьютерами. Протоколы высокого уровня, предназначенные для выполнения функций перенаправления файлов, используют протоколы среднего уровня для передачи пакетов сообщений между рабочими станциями и файловым сервером. Независимо от внутренней конструкции каждого конкретного протокола среднего уровня все они имеют определенные общие функции: инициализация связи, отправка и получение данных и завершение обмена.

Существуют два типа межкомпьютерного обмена данными – *датаграммы* и *сеансы*. Датаграмма – это сообщение, которое не требует подтверждения о приеме от принимающей стороны. Если такое подтверждение необходимо, то адресат должен сам послать специальное сообщение. Для осуществления обмена данным способом принимающая и передающая стороны должны придерживаться определенного протокола во избежание конфликтов. Если в сети находится несколько датаграмм, то порядок их доставки может нарушаться.

В противоположность датаграммам, в сеансе предполагается создание логической связи для обмена сообщениями между рабочими станциями и гарантируется получение сообщений. В то время как датаграммы могут передаваться в произвольные моменты времени, сеанс должен быть установлен, только после этого происходит обмен сообщениями и после окончания обмена данными сеанс должен быть закрыт. В большинстве локальных вычислительных сетей скорость передачи датаграмм значительно выше, чем сообщений в сеансах.

Данные, циркулирующие по сети, фактически поступают от прикладных программ или от прикладного программного обеспечения для перенаправления файлов, которые переносят запросы операционной системы на файловый сервер и обратно. Протоколы среднего уровня NetBIOS, IPX/SPX и TCP/IP позволяют компьютерам в LAN легко обмениваться данными друг с другом.

Протокол NetBIOS (Network Basic Input/Output System – Сетевая базовая система ввода/вывода), являющийся продуктом фирмы IBM, накладывает требования на обмен сообщениями, поступающими от модулей сетевой операционной системы. Каждая из рабочих станций в сети имеет одно или несколько имен. Эти имена хранятся NetBIOS в таблице, в первом столбце которой записывается уникальное и неизменное имя сетевого адаптера. Дополнительные имена пользователей могут при необходимости быть добавлены в таблицу. Также NetBIOS осуществляет управление процессами установки связи, обмена и окончания связи между любыми двумя компьютерами в LAN. Согласно этому протоколу сначала одна станция начинает

прослушивать сеть в ожидании вызова. Если вызов получен, то устанавливается сеанс, в течение которого оба компьютера на равных правах посылают и принимают сообщения. Далее следует закрытие сеанса. Сообщения могут передаваться для какого-то имени, группе имен или всем именам в LAN. Диалога между двумя машинами при этом не устанавливается и прием сообщений не гарантируется. С помощью NetBIOS прикладная программа имеет возможность получить информацию о локальном сетевом адаптере, о других адаптерах в сети и обо всех текущих сеансах связи.

Фирма Novell в сетевой операционной системе NetWare применяет протокол IPX для обмена датаграммами и протокол SPX для обмена в сеансах. Протокол IPX (Internetwork Packet Exchange – Межсетевой обмен пакетами) выполняет функции адресации, маршрутизации и подключения в процессе передачи пакетов сообщений. Протокол IPX является более быстродействующим, чем протокол SPX, но сообщения доставляются безошибочно только примерно в 95 процентах. В соответствии с данным протоколом реакция сервера на служебный запрос (например, записать данные в файл) последует только после соответствующего ответного подтверждения от сервера.

Протокол SPX (Sequenced Packet Exchange – Последовательный обмен пакетами) предназначен для установления диалога и применяется в течение сеанса. Для обмена данными по этому протоколу надо, прежде всего, установить логическую связь между двумя участниками диалога. После этого сообщения могут циркулировать в обоих направлениях с гарантией их безошибочной доставки. Протокол SPX также гарантирует, что сообщения будут передаваться в правильной последовательности (в случае одновременной передачи нескольких пакетов).

В операционной системе NetWare протокол IPX используется для обмена пакетами, предназначенными для обслуживания файлов, а протокол SPX используется для доступа к внутренним функциям управления и диагностики сети. Протокол SPX находится над IPX и фактически использует IPX для обмена пакетами сообщений. Он не применялся в ранних версиях NetWare.

Протокол TCP/IP (Transmission Control Protocol /Internet Protocol) дублирует протоколы NetBIOS, IPX и SPX во многих отношениях. Он был разработан Министерством Обороны США для глобальной сети ARPANET. Протокол TCP/IP является набором протоколов. Часть этого набора IP обеспечивает обмен датаграммами между узлами сети (подобно IPX). Протокол TCP (подобно NetBIOS) обеспечивает связь между двумя узлами с гарантированной доставкой сообщений. Кроме того, существует ряд стандартных утилит для передачи файлов (FTP), для дистанционного запуска программ (TELNET) и для обмена электронной почтой (SMTP). Эти утилиты не производят перенаправление файлов. Для этого существуют отдельные программные продукты, типа NFS (Network File System) фирмы Sun Microsystems.

Модули перенаправления файлов в различных сетевых операционных системах используют свои собственные высокоуровневые протоколы. Фирмы IBM и Microsoft в программных продуктах LAN Manager и LAN Server применяют протокол SMB (Server Message Block – Блоки сообщений сервера). Фирма Novell в операционной системе NetWare использует протокол NCP (Core Protocol – протокол ядра NetWare). В одноранговых LAN иногда используется протокол SMB, но чаще применяют свои собственные протоколы.

Сетевые операционные системы LAN Manager и LAN Server перехватывают обращение прикладных программ на рабочих станциях к функциям ввода/вывода и направляют их в файл-сервер. Для этого программные модули на рабочих станциях, использующие протокол SMB и называемые DLR (DOS LAN Requester), открывают сеанс связи NetBIOS с программным обеспечением LAN Manager или LAN Server, работающим на сервере. После этого DLR и сервер обмениваются SMB-блоками. Фирма IBM определила четыре категории SMB-блоков: управление сеансом, доступ к файлам, обслуживание принтера и сообщения.

В локальных вычислительных сетях NetWare, программные модули на рабочих станциях, выполняющие функции перенаправления файлов, называются *оболочкой* (shell). До недавнего времени для каждой версии DOS требовалась своя оболочка. В настоящее время создана не зависящая от версии DOS программа, которая называется NETX.COM. Все версии NETX.COM используют протокол NCP для обмена пакетами сообщений при работе с файлами.

Как уже отмечалось ранее, локальные вычислительные сети удобны на относительно небольших территориях. Когда этого недостаточно, то автономные LAN объединяют в *глобальные вычислительные сети* (Global Area Network – GAN).

Для объединения различных локальных сетей в одну необходимо, во-первых, наличие физической связи между ними, а во-вторых, наличие условий для обмена информацией между ними (общий протокол). Связанные LAN можно представлять в виде различных уровней, связанных общими протоколами. Самый нижний уровень составляют отдельные LAN со своими рабочими станциями и файловыми серверами. Общий протокол связывает нижние слои с помощью специальных устройств связи, которые не изменяют структуры локальных вычислительных сетей, а служат исключительно для передачи информации между ними. Аппаратные средства GAN состоят из маршрутизаторов, мостов и шлюзов, которые позволяют использовать различные протоколы и топологии в единой неоднородной системе.

Список литературы к изданию «Введение в архитектуру ПК»

1. Букчин Л.В., Безрукий Ю.Л. Дисковая подсистема IBM-совместимых ПК. -М.: 1993.
2. Григорьев В.Л. Видеосистемы ПК фирмы IBM. -М.: Радио и связь, 1993.
3. Ларионов А.М., Горнец Н.Н. Периферийные устройства в вычислительных системах. - М.: Высш.шк., 1991.
4. Левкин Г.Н., Левкина В.Е. Введение в схемотехнику ПЭВМ IBM PC/AT. -М.: Изд-во МПИ, 1991.
5. Мячев Г.Н., Степанов В.Н. Персональные ЭВМ и микро ЭВМ: основы организации. - М.: Радио и связь, 1991.
6. Паппас К., Марри У. Микропроцессор 80386. -М.: Радио и связь, 1993.
7. Хелд Г. IBM PS/2. Справочник пользователя. -М.: Радио и связь.
8. Скотт Мюллер Модернизация и ремонт персональных компьютеров. М.: Бином, 1998.
9. Айден К., Колесниченко О., и др. Аппаратные средства PC. С-Пб., ВНУ, 1998.
10. Соппа И.В. Введение в архитектуру персонального компьютера - Владивосток: Из-во ДВГУ, 1997.
11. Колесников О., Шарыгин М., Шишигин И. Лазерные принтеры. - СПб.:ВНУ, 1997.
12. Мячев А.А., Иванов В.В. Интерфейсы вычислительных систем на базе мини- и микроЭВМ. М.: Радио и связь, 1986, 248с.
13. Морс С.П., Альберт Д.Д. Архитектура микропроцессора 80286. . М.: Радио и связь, 1990.
14. Нанс Б. Компьютерные сети. М.: Бином,1996,440 с.