

Учреждение образования  
«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ  
ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

**М. С. Шмаков**

# **ЭЛЕКТРОННЫЕ ВЫЧИСЛИТЕЛЬНЫЕ МАШИНЫ И ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ**

**Курс лекций  
для студентов специальностей 1-40 01 02-03  
«Информационные системы и технологии (издательско-  
полиграфический комплекс)», 1-36 06 01 «Полиграфическое  
оборудование и системы обработки информации»**

Минск 2009

УДК 004.3 (042.4. 034.4)

ББК 32.97 я 73

Ш71

Рассмотрен и рекомендован к изданию редакционно-издательским советом университета

Рецензенты:

проректор Минского государственного высшего радиотехнического колледжа, доцент, кандидат технических наук

*И. Л. Свито;*

доцент кафедры «Программное обеспечение вычислительной техники и автоматизированных систем» Белорусского национального технического университета, кандидат технических наук *О.В. Бугай*

**Шмаков М.С.**

Ш71      Электронные вычислительные машины и вычислительные системы: Курс лекций для студентов специальностей 1-40 01 02-03 «Информационные системы и технологии (издательско-полиграфический комплекс)», 1-36 06 01 «Полиграфическое оборудование и системы обработки информации»/ М. С. Шмаков – Минск: БГТУ, 2009. – 384 с.

В электронном курсе лекций дано систематизированное описание аппаратной части персональных компьютеров и вычислительных систем. Подробно рассматриваются электронные подсистемы компьютеров – системные платы, процессоры, память и др. Для лучшего понимания работы вычислительных устройств, приводится большое количество структурных и функциональных схем, временных диаграмм.

**УДК 004.3 (042.4. 034.4)**

**ББК 32.97 я 73**

© УО «Белорусский государственный технологический университет», 2009

© Шмаков М.С., 2009

## СОДЕРЖАНИЕ

1. ВВЕДЕНИЕ В КОМПЬЮТЕРНУЮ ТЕХНИКУ .....	5
1.1. Этапы развития компьютеров. Возникновение IBM PC.....	5
1.2. Основные характеристики ЭВМ.....	13
1.3. Режимы работы ЭВМ.....	17
1.4. Функциональный состав компьютера.....	19
2. СИСТЕМНЫЕ ПЛАТЫ .....	31
3. ПРОЦЕССОРЫ .....	43
3.1. Структура и принципы работы процессоров.....	43
3.2. Многоядерные технологии INTEL.....	55
4. ОПЕРАТИВНОЕ ЗАПОМИНАЮЩЕЕ УСТРОЙСТВО (ОЗУ) 62	
4.1. Основные понятия, типы оперативной памяти.....	62
4.2. Память типа ROM.....	63
4.3. Память типа DRAM.....	64
4.4. Память типа SDRAM.....	66
4.5. Память типа DDR SDRAM.....	67
4.6. Память стандарта DDR2. Сравнительная характеристика с DDR.....	68
4.7. Модули памяти.....	73
4.8. КЭШ-память – SRAM.....	75
4.9. Быстродействие ОЗУ.....	77
5. БАЗОВАЯ СИСТЕМА ВВОДА-ВЫВОДА BIOS.....	81
5.1. BIOS в структуре программного обеспечения компьютера .....	81
5.2. Системный модуль ROM BIOS.....	83
5.3. Начальный запуск и самотестирование. Программа POST.....	85
5.4. Начальная загрузка.....	87
5.5. Флэш-BIOS.....	91
6. ПИТАНИЕ КОМПЬЮТЕРОВ.....	92
6.1. Блок питания компьютера.....	92
6.2. Питание блокнотных ПК.....	95
6.3. Средства улучшения качества электропитания.....	96
6.4. Управление энергопотреблением.....	100
7. ВНЕШНЯЯ ПАМЯТЬ.....	103
7.1. Жесткий диск.....	103
7.2. Оптические диски – CD, DVD.....	115
8. ВИДЕОСИСТЕМА .....	126
8.1. Принципы вывода изображений.....	126
8.2. Видеоадаптер.....	132

8.3. Мониторы.....	133
9. АУДИОСИСТЕМА .....	152
9.1. Элементы прикладной звукотехники.....	152
9.2. Оцифровка звука.....	155
9.3. Использование ПК для обработки «цифрового» звука....	159
9.4. Компрессия звука.....	161
9.5. Звуковые платы.....	170
10. ПЕРИФЕРИЙНЫЕ УСТРОЙСТВА .....	178
10.1. Клавиатуры.....	178
10.2. Мыши.....	179
10.3. Сканеры.....	181
10.4. Видеопроекторы.....	184
10.5. Принтеры.....	186
11. ИНТЕРФЕЙСЫ КОМПЬЮТЕРА .....	194
11.1. Последовательный интерфейс.....	194
11.2. Интерфейс PS/2.....	195
11.3. Интерфейс LPT.....	195
11.4. Шины расширения.....	196
11.5. Шины PCI и PCI-X.....	198
11.6. Порт графического акселератора – AGP.....	202
11.7. PCI Express.....	203
11.8. Шина USB.....	206
11.9. Шина IEEE 1394 – FireWire.....	208
11.10. Интерфейс IDE – ATA/ATAPI и SATA.....	209
11.11. Шина SCSI.....	212
11.12. Инфракрасный интерфейс IrDA.....	213
11.13. Радиоинтерфейс BLUETOOTH.....	214
12. ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ .....	215
12.1. Классификация вычислительных систем.....	215
12.2. Многомашинные вычислительные системы.....	221
12.2. Многопроцессорные вычислительные системы.....	222
13. КОММУНИКАЦИОННЫЕ УСТРОЙСТВА .....	230
13.1. Модемы.....	230
13.2.Технология ADSL.....	239
13.3. Высокоскоростные радиосистемы передачи данных.....	252
13.4. IP-телефония.....	257
ЛИТЕРАТУРА.....	265

# 1. ВВЕДЕНИЕ В КОМПЬЮТЕРНУЮ ТЕХНИКУ

## 1.1. Этапы развития компьютеров. Возникновение IBM PC

**ЭВМ (компьютер)** представляет собой устройство, способное исполнять четко определенную последовательность операций, предписанную программой. *Персональный компьютер (ПК)* стал воплощением множества открытий и изобретений. Прежде чем обсуждать его устройство и возможности, скажем несколько слов об основных этапах развития компьютерной технологии. Первые компьютеры были очень похожи на простейшие калькуляторы. Они прошли путь от простых механических до сложных цифровых электронных устройств. В 1833 году английский ученый, профессор Кембриджского университета Чарльз Бэббидж разработал проект «аналитической машины» – вычислительного устройства, по своей схеме и принципам работы очень близкого к современным компьютерам. Но, как это часто бывает с опередившими свое время идеями, проект Бэббиджа оказался в то время технически нереализуемым. Даже материалы об этой машине были опубликованы лишь в 1888 году, уже после смерти автора. Тем не менее, Чарльз Бэббидж вошел в историю как пророк современной киберцивилизации.

Прошло более ста лет, и грянувшая Вторая мировая война — война машин и новых технологий — сделала востребованными высокопроизводительные вычислительные устройства. Без них создание новых типов вооружений становилось просто невозможным, так как на выполнение многих математических расчетов требовалось время, превышающее продолжительность человеческой жизни. Новейшая история вычислительных машин началась в США.

### Рождение американского компьютера

В 1940-х годах артиллерия все еще оставалась «богом войны», и, неудивительно, что именно артиллеристы были первыми, кто требовал новые вычислительные технологии. Орудия полевые, зенитные, корабельные, авиационные, дальнобойные — каждый тип пушек требовал своих, специфических расчетов баллистики. Без ученых тут было никак не обойтись. Именно поэтому в 1941 году сотрудники Лаборатории баллистических исследований Абердинского артиллерий-

ского полигона в США обратились в техническую школу при Пенсильванском университете с просьбой помочь им составить специальные таблицы для стрельб. Выполнить просьбу военных поручили сотруднику школы Джону Моусли (1907–1986). Однако, исследовав проблему, Моусли понял, что просто таблицами тут не отделаешься — уже не XVIII век на дворе. Он предложил создать не разрозненные таблицы, а сразу единое универсальное электронное счетное устройство.

Артиллеристы с готовностью поддержали эту инициативу. Моусли занялся теоретическими разработками, а военные добивались правительственного финансирования. К этому моменту Соединенные Штаты вступили в Мировую войну, так что требуемые деньги нашлись без особого труда. Причем сразу 400 тыс. долларов — колоссальная по тем временам сумма. Для разработки машины в 1943 году была создана группа более чем из двухсот человек. Самым активным помощником, «правой рукой» Моусли стал инженер-электронщик Джон Эккерт (1919–1995). Азарт молодых исследователей подстегивался и действиями конкурентов — как раз в те годы в Америке другой группой исследователей строилась вычислительная машина MARK-I.

Работа в группе Моусли кипела вовсю. Многие теоретические выкладки приходилось пересматривать после первых же попыток их практической реализации. Деньги быстро закончились, пришлось экономить и просить дополнительное финансирование. Артиллеристы торопили. Тем временем в 1944 году MARK-I, наконец, заработал. Однако исследователи из Пенсильванского университета понимали, что у этого компьютера на самом деле нет будущего. Дело в том, что MARK-I был электрическим вычислителем, работал на множестве электромагнитных реле. Его создание стало возможным благодаря накопленному опыту эксплуатации телефонной аппаратуры, счетно-аналитических и счетно-перфорационных машин. А Моусли создавал свое детище на основе принципиально других компонентов — электронных ламп. По ходу работы пришлось преодолевать множество технологических сложностей. Больше всего хлопот доставляли электронные лампы — всего их использовалось 20 тыс. штук, и при том уровне качества обеспечить работу такого массива на протяжении более чем нескольких минут оказывалось невозможно: какая-нибудь лампа обязательно сгорала. Выход нашли — понизить напряжение накала. Так или иначе, но к концу 1945 года машина была построена и

получила название ENIAC (Electronic Numerical Integrator and Computer). Она осуществляла 5000 операций сложения и 300 операций умножения в секунду, имела 26 метров в длину, 6 метров в высоту и весила 35 тонн. Для сравнения, это как четыре троллейбуса, собранных вместе. Энергопотребление было соответствующим. Но по тем временам подобный «арифмометр» казался вершиной технического прогресса.

Война, правда, уже закончилась, так что артиллеристы к вычислительной «игрушке» заметно охладели. Зато появились новые заинтересованные лица — сначала атомщики, а потом военно-воздушные силы. Первым вычислительные мощности требовались для разработки вначале ядерной, а затем термоядерной бомбы. Вторые же занялись созданием реактивных самолетов, что также требовало колоссальных расчетов. Затем подключились ракетчики, радиолокаторщики, подводники... Словом, спустя 10 лет в США уже работало порядка 7 тыс. вычислительных машин типа ENIAC и его усовершенствованных потомков.

### **Рождение советского компьютера.**

Первый советский компьютер был создан под руководством Сергея Алексеевича Лебедева (1902 – 1974). Правда, в СССР необходимость создания ЭВМ высокие начальники осознали позже американцев, так что соответствующие работы начались только с осени 1948 года. Инициаторами проекта выступили ученые-ядерщики — в те годы буквально вся страна работала над атомным проектом, который курировал лично Лаврентий Берия. Первым делом советские учёные приступили к разработке Малой электронной счетной машины (МЭСМ).

Советская история создания ЭВМ не менее драматична, чем американская. Для разработки отечественной ЭВМ Лебедеву и его сотрудникам отвели целое крыло двухэтажного здания тайной лаборатории, которая скрывалась в лесных дубравах в местечке Феофания под Киевом. По воспоминаниям участников тех событий, работали все члены коллектива без сна и отдыха. Прямо в лаборатории пили чай, а под утро, для снятия напряжения — спирт, предназначенный для протирки контактов, — другого спиртного в тех условиях достать было попросту невозможно.

Только к концу 1949 определилась принципиальная схема блоков машины. Далее начались чисто технические сложности — те самые, с которыми за несколько лет до этого столкнулись американцы. Но к концу 1950 года вычислительная машина была все-таки построена. После отладки, в конце 1951-го, МЭСМ прошла испытания и была принята в эксплуатацию Комиссией АН СССР во главе с академиком Мстиславом Келдышем. С 1952 г. на запущенных в масштабное производство МЭСМах решались важнейшие научно-технические задачи из области термоядерных процессов, космических полетов и ракетной техники, дальних линий электропередачи, механики, статистического контроля качества, сверхзвуковой авиации.

А коллектив разработчиков под руководством Сергея Лебедева тем временем продолжал работу — теперь уже над Большой электронной счетной машиной (БЭСМ). Первый ее экземпляр заработал в 1953 г. Эта разработка оказалась счастливой — именно на основе БЭСМ в последующие годы были созданы такие известные советские ЭВМ, как «Стрела», «Минск», «Урал», «Днепр», «Мир», М-20, М-220.

Впервые в отечественной практике именно в М-20 Лебедевым с целью повышения производительности были реализованы автоматическая модификация адреса, совмещение работы арифметического устройства и выборки команд из памяти, введение буферной памяти для массивов данных, выдаваемых на печать, совмещение ввода и вывода данных со счетом, использование полностью синхронной передачи сигналов в логических цепях. Но еще до этого, в 1956 году, доклад Сергея Лебедева о БЭСМ на международной конференции в Дармштадте произвел сенсацию — БЭСМ была на уровне лучших американских машин тех лет и самой быстродействующей ЭВМ в Европе.

В начале 50-х годов суровые реалии наступившей «холодной войны» потребовали создания специализированных ЭВМ для нужд оборонного комплекса. Делались они на основе именно лебедевских разработок. Созданные под руководством Сергея Лебедева и его последователей вычислительные машины для системы противоракетной обороны стали основой достижения стратегического паритета СССР и США в период ядерного противостояния. В 1952-1955 годах учеником Лебедева Виктором Бурцевым были разработаны специализированные ЭВМ «Диана-1» и «Диана-2» (по имени древнегреческой богини охоты) для автоматического съема данных с радиолокатора и автоматического слежения за целями. Затем для первой системы противора-



кетной обороны, генеральным конструктором которой был Григорий Кисунько, в 1958 г. была предложена ламповая ЭВМ М-40, а немного позже — М-50 (с плавающей точкой). В начале 60-х г. к строительству советских ЭВМ подключились спецслужбы. Сегодня уже не секрет, что советские и восточногерманские разведчики десятилетиями добывали на Западе множество новейших научно-технических разработок, в том числе и в области электронно-вычислительной техники. В результате в 1967 г. появился компьютер БЭСМ-6, в котором было смонтировано и одновременно работало 60 тысяч транзисторов и 200 тысяч полупроводниковых диодов. Этот, по тем временам «монстр» выполнял 1 миллион операций в секунду. На основе БЭСМ-6 были созданы вычислительные центры коллективного пользования для научных организаций, системы автоматизации научных исследований в ядерной физике и других областях науки, информационно-вычислительные системы обработки информации в реальном времени, востребованные в первую очередь космической индустрией. Она использовалась для моделирования сложнейших физических процессов и процессов управления, в системах проектирования программного обеспечения для новых ЭВМ. БЭСМ-6 выпускалась долгие 17 лет — абсолютный мировой рекорд для такого рода техники.

Сергей Лебедев занимался компьютерными разработками до последних дней своей насыщенной жизни. Последним его детищем стал «Эльбрус» — знаменитый многопроцессорный комплекс, положивший начало целому новому классу вычислительных систем. Но это уже тема раздела «Вычислительные системы».

### **Рождение персонального компьютера**

В 1973 году были разработаны первые микропроцессорные комплекты на основе микропроцессора 8008. Правда, они годились разве что для демонстрации своих возможностей и включения индикаторов. В конце 1973 г. Intel выпустила микропроцессор 8080, быстродействие которого было в 10 раз выше, чем у 8008, и который мог адресовать память объемом до 64 Кб. Это стало толчком к промышленному производству ПК.

В 1975 году фотография комплекта Altair компании MITS была помещена на обложку январского номера журнала *Popular Electronic*. Этот комплект, который можно считать первым ПК, состоял из процессора 8080, блока питания, лицевой панели с множеством индика-

торов и запоминающего устройства емкостью 256 байт (не килобайт!). Стоимость комплекта составляла 395 долларов, и покупатель должен был сам собрать компьютер. Этот ПК был построен по схеме с открытой шиной (разъемами), что позволяло другим фирмам разрабатывать дополнительные платы и периферийное оборудование. Появление нового процессора стимулировало разработку различного программного обеспечения, включая операционную систему CP/M (Control Program for Microprocessors) и первый язык программирования BASIC (Beginners All-purpose Symbolic Instruction Code) компании Microsoft.

В 1975 году IBM впервые выпустила устройство, которое можно было бы назвать персональным компьютером. Модель 5100 имела память емкостью 16 Кб, встроенный дисплей на 16 строк по 64 символа, интерпретатор языка BASIC и кассетный накопитель DC-300. Однако стоимость компьютера (9 000 долларов) для рядового покупателя оказалась слишком высокой, особенно если учесть, что множество любителей (названных позже хакерами) предлагали собственные комплекты всего за 500 долларов. Очевидно, что ПК компании IBM не могли выдержать такой конкуренции на рынке и продавались очень плохо.

До появления известного сейчас IBM PC (модель 5150) были разработаны модели 5110 и 5120. Хотя эти компьютеры и предшествовали IBM PC, они не имели с ним ничего общего. IBM PC был больше похож на выпущенную в 1980 году для применения в офисах модель System/23 DataMaster.

В 1976 году новая компания Apple Computer вышла на рынок с компьютером Apple I стоимостью 666 долларов. Его системная плата была привинчена к куску фанеры, а корпуса и блока питания не было вообще. Было выпущено всего несколько экземпляров этого компьютера, которые впоследствии продавались коллекционерам за 20 тыс. долларов. Но появившийся в 1977 году компьютер Apple II стал прообразом большинства последующих моделей, включая и IBM PC.

К 1980 году на рынке микрокомпьютеров доминировали две базовые модели компьютерных систем. Это был Apple II, имевший множество преданных пользователей и гигантское количество программ, и несколько других моделей, происходивших от комплекта Altair. Эти компьютеры были совместимы один с другим, имели одну операционную систему (CP/M) и стандартные разъемы расширения с шиной S-100 (по 100 контактов на разъем). Все они собирались различными компаниями и продавались под разными названиями. Но в

большинстве случаев ими использовались одинаковые программные и аппаратные части.

Новый конкурент, появившийся на горизонте, дал возможность определить факторы будущего успеха персонального компьютера: открытая архитектура, слоты расширения, сборная конструкция, а также поддержка аппаратного и программного обеспечения различных компаний. Конкурентом, как ни удивительно, оказался компьютер компании IBM, до сих пор занимавшейся только мощными производственными мэйнфреймами. Этот компьютер, по существу, напоминал раннюю версию Apple, в то время как системы Apple приобретали черты, более подходящие для IBM. Открытая архитектура IBM PC и закрытая архитектура компьютеров Macintosh произвели настоящий переворот в компьютерной индустрии.

### **Персональный компьютер компании IBM**

В конце 1980 года IBM решила выйти на стремительно развивающийся рынок дешевых ПК. Для разработки нового компьютера она основала в городе Бока-Ратон (шт. Флорида) свое отделение Entry Systems Division. Небольшую группу из 12 человек возглавил Дон Эстридж (Don Estridge), а главным конструктором был Льюис Эггебрехт (Lewis Eggebrecht). Именно эта группа и разработала первый настоящий IBM PC. Модель 5100, разработанную в 1975 году, IBM считала интеллектуальным программируемым терминалом, а не подлинным компьютером, хотя, в сущности, это был настоящий компьютер. Почти все инженеры группы ранее работали над проектом компьютера System/23 DataMaster, поэтому он фактически оказался прообразом IBM PC.

Многое в конструкции IBM PC было заимствовано от DataMaster. Так, например, раскладка и электрическая схема клавиатуры были скопированы с DataMaster, правда, в IBM PC дисплей и клавиатура были автономны, в отличие от DataMaster, где они объединялись в одно устройство (что было неудобно).

Были заимствованы и некоторые другие компоненты, включая системную шину (разъемы ввода-вывода), причем использовались не только те же самые 62-контактные разъемы, но и разводка контактов. В IBM PC применялись те же контроллеры прерываний и прямого доступа к памяти, что и в DataMaster. При этом платы расширения, разработанные для DataMaster, можно было использовать и в IBM PC.

Однако в DataMaster применялся процессор 8085 компании Intel, который мог адресовать всего 64 Кбайт памяти и имел 8-разрядные внутреннюю и внешнюю шины данных. Из-за этих ограничений в IBM PC использовался процессор 8088, который имел адресное пространство 1 Мбайт, 16-разрядную внутреннюю шину данных, но внешняя шина данных была 8-разрядной. Благодаря 8-разрядной внешней шине данных и аналогичной системе команд можно было использовать устройства, разработанные ранее для DataMaster.

Компания IBM создала компьютер менее чем за год, максимально внедрив в него имевшиеся разработки и компоненты других производителей. Группе Entry Systems Division была предоставлена большая независимость, чем другим подразделениям: им было разрешено использовать услуги и продукцию других фирм в обход бюрократического правила, предписывающего использовать в разработках только изделия IBM. Например, языки программирования и операционную систему для IBM PC разрабатывала Microsoft.

### **Закон Мура**

В 1965 г. Гордон Мур при подготовке доклада о перспективах развития компьютерной памяти обнаружил интересную особенность: емкость каждой новой микросхемы памяти удваивается по сравнению с ее предшественницей, а сама новая микросхема появляется каждые 18-24 месяца. Построив линию тренда, Мур отметил, что производительность компьютеров будет увеличиваться экспоненциально по времени.

Эту зависимость стали называть *законом Мура*. Кстати, этот закон описывает не только рост емкости оперативной памяти, он часто используется для определения степени роста быстродействия процессоров и емкости жестких дисков. За 26 лет количество транзисторов процессора увеличилось в 18 тыс. раз, от 2,3 тыс. в процессоре 4004 до 140 млн. в Pentium III Xeon.

Что нас ждет в будущем? Для рынка ПК можно с уверенностью сказать лишь одно: компьютеры будут быстрее, меньше и дешевле. Согласно закону Мура, купленный вами сегодня компьютер будет работать в два раза медленнее и хранить в два раза меньше данных по сравнению с системой, купленной через 2 года. Но самое удивительное в том, что прогресс на этом не остановится: уже сейчас срок морального старения компьютера вплотную приблизился к отметке в один год.

## 1.2 Основные характеристики ЭВМ

Появление любого нового направления в вычислительной технике определяется требованиями компьютерного рынка. Поэтому перед разработчиками компьютеров стоят много задач, одна из которых повышение основных характеристик:

- отношение стоимость/производительность;
- надежность и отказоустойчивость;
- масштабируемость;
- совместимость и мобильность программного обеспечения.

Большая универсальная вычислительная машина (мейнфрейм) или суперкомпьютер стоят дорого. Для достижения поставленных целей при проектировании высокопроизводительных конструкций приходится игнорировать стоимостные характеристики. Суперкомпьютеры фирмы Cray Research и высокопроизводительные мейнфреймы компании IBM относятся именно к этой категории компьютеров. Другим крайним примером может служить низкостоимостная конструкция, где производительность принесена в жертву для достижения низкой стоимости. К этому направлению относятся персональные компьютеры различных клонов IBM PC. Между этими двумя крайними направлениями находятся конструкции, основанные на отношении стоимость – производительность, в которых разработчики находят баланс между стоимостными параметрами и производительностью. Типичными примерами такого рода компьютеров являются миникомпьютеры и рабочие станции.

Для сравнения различных компьютеров между собой обычно используются стандартные методики измерения производительности. Эти методики позволяют разработчикам и пользователям использовать полученные в результате испытаний количественные показатели для оценки тех или иных технических решений, и в конце концов именно производительность и стоимость дают пользователю рациональную основу для решения вопроса: какой компьютер выбрать.

Важнейшей характеристикой вычислительных систем является **надежность**. Повышение надежности основано на принципе предотвращения неисправностей путем снижения интенсивности отказов и сбоев за счет применения электронных схем и компонентов с высокой и сверхвысокой степенью интеграции, снижения уровня помех, облегченных режимов работы схем, обеспечение тепловых режимов

их работы, а также за счет совершенствования методов сборки аппаратуры.

**Отказоустойчивость** – это такое свойство вычислительной системы, которое обеспечивает ей, как логической машине, возможность продолжения действий, заданных программой, после возникновения неисправностей. Введение отказоустойчивости требует избыточного аппаратного и программного обеспечения. Направления, связанные с предотвращением неисправностей и с отказоустойчивостью, основные в проблеме надежности. Концепции параллельности и отказоустойчивости вычислительных систем естественным образом связаны между собой, поскольку в обоих случаях требуются дополнительные функциональные компоненты. Поэтому на параллельных вычислительных системах достигается как наиболее высокая производительность, так и, во многих случаях, очень высокая надежность. Имеющиеся ресурсы избыточности в параллельных системах могут гибко использоваться как для повышения производительности, так и для повышения надежности. Структура многопроцессорных и многомашинных систем приспособлена к автоматической реконфигурации и обеспечивает возможность продолжения работы системы после возникновения неисправностей.

Следует помнить, что понятие надежности включает не только аппаратные средства, но и программное обеспечение. Главной целью повышения надежности систем является целостность хранимых в них данных.

**Масштабируемость** представляет собой возможность наращивания числа и мощности процессоров, объемов оперативной и внешней памяти и других ресурсов вычислительной системы. Масштабируемость должна обеспечиваться архитектурой и конструкцией компьютера, а также соответствующими средствами программного обеспечения.

Добавление каждого нового процессора в действительно масштабируемой системе должно давать прогнозируемое увеличение производительности и пропускной способности при приемлемых затратах. Одной из основных задач при построении масштабируемых систем является минимизация стоимости расширения компьютера и упрощение планирования. В идеале добавление процессоров к системе должно приводить к линейному росту ее производительности. Однако это не всегда так. Потери производительности могут возникать, например, при недостаточной пропускной способности шин из-за воз-

растания трафика между процессорами и основной памятью, а также между памятью и устройствами ввода – вывода. В действительности реальное увеличение производительности трудно оценить заранее, поскольку оно в значительной степени зависит от динамики поведения прикладных задач.

Возможность масштабирования системы определяется не только архитектурой аппаратных средств, но зависит от заложенных свойств программного обеспечения. Масштабируемость программного обеспечения затрагивает все его уровни от простых механизмов передачи сообщений до работы с такими сложными объектами как мониторы транзакций и вся среда прикладной системы. В частности, программное обеспечение должно минимизировать трафик межпроцессорного обмена, который может препятствовать линейному росту производительности системы. Аппаратные средства (процессоры, шины и устройства ввода – вывода) являются только частью масштабируемой архитектуры, на которой программное обеспечение может обеспечить предсказуемый рост производительности. Важно понимать, что простой переход, например, на более мощный процессор может привести к перегрузке других компонентов системы. Это означает, что действительно масштабируемая система должна быть сбалансирована по всем параметрам.

**Концепция программной совместимости** впервые в широких масштабах была применена разработчиками системы IBM/360. Основная задача при проектировании всего ряда моделей этой системы заключалась в создании такой архитектуры, которая была бы одинаковой с точки зрения пользователя для всех моделей системы независимо от цены и производительности каждой из них. Огромные преимущества такого подхода, позволяющего сохранять существующий задел программного обеспечения при переходе на новые (как правило, более производительные) модели были быстро оценены как производителями компьютеров, так и пользователями и начиная с этого времени практически все фирмы-поставщики компьютерного оборудования взяли на вооружение эти принципы, поставляя серии совместимых компьютеров. Следует заметить, что со временем даже самая передовая архитектура неизбежно устаревает и возникает потребность внесения радикальных изменений в архитектуру и способы организации вычислительных систем.

В настоящее время одним из наиболее важных факторов, определяющих современные тенденции в развитии информационных тех-

нологий, является ориентация компаний-поставщиков компьютерного оборудования на рынок прикладных программных средств. Это объясняется, прежде всего, тем, что для конечного пользователя в конце концов важно программное обеспечение, позволяющее решить его задачи, а не выбор той или иной аппаратной платформы. Переход от однородных сетей программно совместимых компьютеров к построению неоднородных сетей, включающих компьютеры разных фирм-производителей, в корне изменил и точку зрения на саму сеть: из сравнительно простого средства обмена информацией она превратилась в средство интеграции отдельных ресурсов – мощную распределенную вычислительную систему, каждый элемент которой (сервер или рабочая станция) лучше всего соответствует требованиям конкретной прикладной задачи.

Этот переход выдвинул ряд новых требований. Прежде всего, такая вычислительная среда должна позволять гибко менять количество и состав аппаратных средств и программного обеспечения в соответствии с меняющимися требованиями решаемых задач. Она должна обеспечивать возможность запуска одних и тех же программных систем на различных аппаратных платформах, т.е. обеспечивать мобильность программного обеспечения. Эта среда должна гарантировать возможность применения одних и тех же человеко-машинных интерфейсов на всех компьютерах, входящих в неоднородную сеть. В условиях жесткой конкуренции производителей аппаратных платформ и программного обеспечения сформировалась концепция открытых систем, представляющая собой совокупность стандартов на различные компоненты вычислительной среды, предназначенных для обеспечения мобильности программных средств в рамках неоднородной, распределенной вычислительной системы.

Одним из вариантов моделей открытой среды является модель OSE (Open System Environment), предложенная комитетом IEEE POSIX. На основе этой модели национальный институт стандартов и технологии США выпустил документ «Application Portability Profile (APP). The U.S. Government's Open System Environment Profile OSE/1 Version 2.0», который определяет рекомендуемые для федеральных учреждений США спецификации в области информационных технологий, обеспечивающие мобильность системного и прикладного программного обеспечения. Все ведущие производители компьютеров и программного обеспечения в США в настоящее время придерживаются требований этого документа.



### 1.3 Режимы работы ЭВМ

В современных ЭВМ можно выделить следующие режимы работы:

- однопрограммный;
- мультипрограммный;
- пакетной обработки;
- разделения во времени;
- диалоговый;
- режим реального времени.

**Однопрограммный режим работы** – режим, при котором выполняется не более одной независимой программы. При таком режиме работы ЭВМ решение задачи начинается с загрузки программы в ОП, после чего ЭВМ последовательно выполняет команды программы. При этом в каждый момент времени работает одно ее устройство, в то время как остальные простаивают в ожидании окончания ранее начатого действия. Значительные потери рабочего времени ЭВМ связаны с медленной работой устройства ввода-вывода по сравнению с работой быстродействующих устройств (АЛУ, ЦУУ, ОЗУ и т.д.).

**Мультипрограммный режим работы** – режим, при котором в памяти ЭВМ хранится несколько программ и выполнение одной программы может быть прервано для перехода к выполнению другой с последующим возвратом к прерванной программе. При совместном выполнении нескольких программ простои оборудования уменьшаются, поскольку увеличивается вероятность того, что среди находящихся в ЭВМ программ имеется одна, готовая к использованию освободившегося оборудования. Для уменьшения простоев оборудования ЭВМ широко применяют метод организации параллельной работы устройства ЭВМ за счет совмещения различных операций при работе ЭВМ. В целях более эффективного использования ЭВМ организуют мультипрограммную обработку информации на ЭВМ так, чтобы ею параллельно выполнялись команды, относящиеся к различным и независимым программам.

Мультипрограммный режим повышает производительность ЭВМ за счет увеличения числа задач, решаемых ЭВМ в течение некоторого промежутка времени. При этом время решения отдельной задачи увеличивается по сравнению с временем решения ее в однопрограммном режиме.

### **Режим пакетной обработки**

Для обеспечения мультипрограммной обработки информации необходимо наличие нескольких задач, ожидающих обработки. Для эффективной загрузки ЭВМ используется **режим пакетной обработки** данных. В этом режиме задачи (программы и данные), подготовляемые многими пользователями ЭВМ, собираются в пачки-пакеты. Пакет состоит из заданий (не более 15), относящимся ко многим задачам, обработка которых занимает не менее часа машинного времени.

Различают два режима пакетной обработки. В первом число задач, выполняемых одновременно, фиксируется, а во втором не фиксируется, но в процессе обработки пакета ЭВМ оно может изменяться динамически. Пакет, предварительно записанный на том или ином носителе информации, вводится в ОЗУ ЭВМ. Когда пакет загружен, ЭВМ выбирает на обработку несколько задач и начинает выполнять их в мультипрограммном режиме. Когда решение одной группы задач пакета закончено, из него выбирается для обработки следующая группа, это продолжается до тех пор, пока не будет обработана последняя группа задач пакета. После этого в ЭВМ вводится новый пакет задач.

Пакетная обработка данных позволяет увеличить производительность ЭВМ и уменьшить стоимость машинной обработки информации.

**Режим разделения времени**, обеспечивает непосредственный и одновременный доступ к ЭВМ некоторому количеству пользователей чаще всего с дистанционно удаленных пунктов (терминалов). Терминал – периферийное устройство, предназначенное для обслуживания одного человека, решающего задачи на ЭВМ.

Пользователи с помощью терминалов вводят в ЭВМ исходные данные и программы и получают результаты вычислений. ЭВМ предоставляет каждому активному терминалу квант времени, равный секундам и долям секунды. По истечении этого времени ЭВМ переходит к обслуживанию следующего пользователя. За некоторый период времени ЭВМ обслуживает всех пользователей. При достаточно высоком быстродействии ЭВМ у отдельного пользователя создается иллюзия непрерывного контакта с ЭВМ.

Разделение времени позволяет устранить потери машинного времени, связанные с вмешательством оператора в работу ЭВМ из-за его сравнительно низкой скорости реакции, необходимости выполнения им определенных действий вне ЭВМ и медленного ввода информации с пульта оператора. При мультипрограммной работе ЭВМ

в промежуточные паузы работы одного оператора к ЭВМ имеют доступ другие, что позволяет обеспечить полную загрузку внутренних устройств ЭВМ и тем самым поднять эффективность ее работы.

Режим разделения времени совместим с режимом пакетной обработки данных, которая предусматривается в ЭВМ для решения задач в отдельные периоды времени, когда пользователи не загружают ЭВМ полностью.

**Диалоговый режим работы** – режим (режим “запрос-ответ”), при котором все программы пользователей постоянно хранятся в памяти ЭВМ и пользователи имеют непосредственный доступ к ЭВМ. От пользователей в ЭВМ поступают входные данные и запросы с пультовых пишущих машинок или дисплеев. Ответ формируется по программе, соответствующей определенному запросу. Выбор допустимых запросов ограничен емкостью памяти. Каждый запрос имеет соответствующий приоритет и временные ограничения на срок обслуживания.

**Режим работы в реальном масштабе времени** – режим, при котором ЭВМ управляет работой какого-либо объекта или технологического процесса. Особенностью работы в реальном масштабе времени является то, что, помимо арифметической и логической обработки, выполняется слежение за работой объекта или прохождение некоторого процесса. Реализация этого режима привела к усложнению устройств и программного обеспечения ЭВМ.

#### 1.4 Функциональный состав компьютера

Любой IBM PC-совместимый компьютер представляет собой реализацию так называемой фон-неймановской архитектуры вычислительных машин. Эта архитектура была представлена Джорджем фон Нейманом (George von Neumann) еще в 1945 г. и имеет следующие основные признаки. Машина состоит из **блока управления, арифметико-логического устройства (АЛУ), памяти и устройств ввода-вывода**. В ней реализуется **концепция хранимой программы**: программы и данные хранятся в одной и той же памяти. Выполняемые действия определяются блоком управления и АЛУ, которые вместе являются основой центрального процессора. Центральный процессор выбирает и исполняет команды из памяти последовательно, адрес очередной команды задается «счетчиком адреса» в блоке управления. Этот принцип исполнения называется *последовательной передачей*

*управления*. Данные, с которыми работает программа, могут включать *переменные* — именованные области памяти, в которых сохраняются значения с целью дальнейшего использования в программе. Фон-неймановская архитектура не единственный вариант построения ЭВМ, есть и другие, которые не соответствуют указанным принципам (например, потоковые машины). Однако подавляющее большинство современных компьютеров основаны именно на указанных принципах, включая и сложные многопроцессорные комплексы, которые можно рассматривать как объединение фон-неймановских машин. Конечно же, за более чем полувековую историю ЭВМ классическая архитектура прошла длинный путь развития. Тем не менее ПК можно «разложить по полочкам» следующим образом.

**Центральный процессор** (АЛУ с блоком управления) реализуется микропроцессором семейства x86, от 8086/88 до новейших процессоров Pentium, Athlon и Opteron (и это не конец истории). При всей внутренней суперскалярности, суперконвейеризованности и спекулятивности (см. главу 7) современного процессора внешне он соблюдает вышеупомянутый принцип последовательной передачи управления. Набор арифметических, логических и прочих инструкций насчитывает несколько сотен, а для потоковой обработки придуман принцип SIMD (Single Instruction Multiple Data — множество комплектов данных, обрабатываемых одной инструкцией), по которому работают расширения MMX, 3DNow!, SSE. Процессор имеет набор регистров, часть которых доступна для хранения операндов, выполнения действий над ними и формирования адреса инструкций и операндов в памяти. Другая часть регистров используется процессором для служебных (системных) целей, доступ к ним может быть ограничен (есть даже программноневидимые регистры). Все компоненты компьютера представляются для процессора в виде наборов *ячеек памяти* или *портов ввода-вывода*, в которые процессор может записывать и/или из которых может считывать содержимое.

**Память** входит в состав различных компонентов компьютера. *Оперативная память* (ОЗУ) — самый большой массив ячеек памяти со смежными адресами — реализуется, как правило, на модулях (микросхемах) динамической памяти. Для повышения производительности обмена данными (включая и считывание команд) оперативная память кэшируется сверхоперативной памятью. Два уровня кэширования территориально располагаются в микропроцессоре. Оперативная память вместе с кэшем всех уровней (в настоящее время — до трех)

представляет собой единый массив памяти, непосредственно доступный процессору для записи и чтения данных, а также считывания программного кода. Помимо оперативной память включает также **постоянную (ПЗУ)**, из которой можно только считывать команды и данные, и некоторые виды специальной памяти (например, видеопамять графического адаптера). Вся эта память (вместе с оперативной) располагается в едином пространстве с линейной адресацией. В любом компьютере обязательно есть **энергонезависимая память**, в которой хранится программа начального запуска компьютера и минимально необходимый набор сервисов (ROM BIOS).

Процессор (один или несколько), память и необходимые элементы, связывающие их между собой и с другими устройствами, называют **центральной частью** или **ядром**, компьютера (или просто **центром**). То, что в фон-неймановском компьютере называлось устройствами ввода-вывода (УВВ), удобнее называть периферийными устройствами.

**Периферийные устройства (ПУ)** – это все программно-доступные компоненты компьютера, не попавшие в его центральную часть. Их можно разделить по назначению на несколько классов:

- **устройства хранения данных** (устройства внешней памяти) — дисковые (магнитные, оптические, магнитооптические), ленточные (стримеры), твердотельные (карты, модули и USB-устройства на флэш-памяти). Эти устройства используются для сохранения информации, находящейся в памяти, на энергонезависимых носителях и загрузки этой информации в оперативную память;

- **устройства ввода-вывода** служат для преобразования информации из внутреннего представления компьютера (биты и байты) в форму, понятную человеку и различным техническим устройствам и обратно;

- **коммуникационные устройства** служат для передачи информации между компьютерами и/или их частями. Сюда относят модемы (проводные, радио, оптические, инфракрасные и др.), адаптеры локальных и глобальных сетей. В данном случае преобразование формы представления информации требуется только для передачи ее на расстояние.

Процессор, память и периферийные устройства взаимодействуют между собой с помощью шин и интерфейсов, аппаратных и программных; стандартизация интерфейсов делает архитектуру компьютеров открытой.

## Биты, байты, слова, параграфы

Компьютер работает в двоичной системе исчисления – минимальным информационным элементом является *бит*, который может принимать значение 0 или 1. Этим значениям соответствуют различные физические состояния ячейки, чаще всего — уровень напряжения (низкий или высокий). Биты организуются в более крупные образования — ячейки памяти и регистры. Каждая ячейка памяти (и каждый регистр) имеет свой адрес, однозначно ее идентифицирующий в определенной системе координат. Минимальной адресуемой (пересылаемой между компонентами компьютера) единицей информации является *байт*, состоящий, как правило, из 8 бит. Два байта со смежными адресами образуют *слово* (word) разрядностью 16 бит, два смежных слова — *двойное слово* (double word) разрядностью 32 бита, два смежных двойных слова — *учетверенное слово* (quad word) разрядностью 64 бита. Байт (8 бит) делится на пару *тетрад* (nibble), старшую тетраду — биты [7:4] и младшую тетраду — биты [3:0].

В двухбайтном слове принят *LN-порядок следования байтов*: адрес слова указывает на младший байт L (Low), а старший байт H (High) размещается по адресу, большему на единицу. В двойном слове порядок аналогичен — адрес указывает на самый младший байт, после которого размещены следующие по старшинству. Этот порядок, называемый форматом *Little Endian*, и естественный для процессоров Intel и применяется не во всех микропроцессорных семействах. В формате *Big Endian* адрес указывает на самый старший байт (двойного, учетверенного) слова, остальные байты размещаются по нарастающим адресам. С несовпадением форматов представления придется сталкиваться, например, при передаче информации между различными подсистемами (формат Big Endian используется в сетевых протоколах и шине Fire Wire).

В технической документации, электрических схемах и текстах программ могут применяться разные ***способы представления чисел***:

- *двоичные (binary) числа* — каждая цифра отражает значение одного бита (0 или 1), старший бит всегда пишется слева, после числа ставится буква «b». Для удобства восприятия тетрады могут быть разделены пробелами, например, 1010 0101b;

- *шестнадцатеричные (hexadecimal) числа* — каждая тетрада представляется одним символом 0...9, A, B, ..., F. Обозначаться такое представление может по-разному, в данной работе используется толь-

ко символ «h» после последней шестнадцатеричной цифры, например, A5h. В текстах программ это же число может обозначаться и как 0xA5, и как 0A5h, в зависимости от синтаксиса языка программирования. Незначащий ноль (0) добавляется слева от старшей шестнадцатеричной цифры, изображаемой буквой, чтобы различать числа и символические имена;

– *десятичные (decimal) числа* — каждый байт (слово, двойное слово) представляется обычным числом, а признак десятичного представления (букву «d») обычно опускают. Байт из предыдущих примеров имеет десятичное значение 165. В отличие от двоичной и шестнадцатеричной форм записи, по десятичной трудно в уме определить значение каждого бита, что иногда приходится делать;

– *восьмеричные (octal) числа* — каждая тройка битов (разделение начинается с младшего) записывается в виде цифры из интервала 0–7, в конце ставится признак «o». То же самое число записывается как 245o. Восьмеричная система неудобна тем, что байт не разделить поровну, но зато все цифры — привычные. В «про-интеловских» системах это представление непопулярно.

Чтобы перевести любое 8-битное число в десятичное, нужно десятичный эквивалент старшей тетрады умножить на 16 и сложить с эквивалентом младшей тетрады.

Для нашего примера  $A5h = 10 \cdot 16 + 5 = 165$ . Обратный перевод тоже несложен: десятичное число делится на 16, целая часть даст значение старшей тетрады, остаток — младшей.

В «наследство» от процессоров 8086/88 достался своеобразный способ задания адреса ячейки памяти в виде *указателя* «seg:offset», состоящего из двух слов: *сегмента* (seg — segment) и *смещения* (offset). Такая запись предполагает вычисление полного адреса по формуле  $addr = 16 \cdot seg + offset$ . Такое представление 20-битного адреса двумя 16-битными числами в процессорах 8086/88 поддерживается и в реальном режиме всех последующих процессоров x86 (подробнее об адресации памяти см. в 7.3). Здесь сегмент указывает адрес *параграфа* — 16-байтной области памяти. Выравнивание адреса по границе параграфа означает, что он кратен 16 (4 младших бита нулевые). Нетрудно увидеть, что один и тот же адрес можно задавать разными сочетаниями этих двух компонентов. Так, например, адрес начала области данных BIOS (BIOS Data Area) 00400h представляют и как 0000:0400, и как 0040:0000 (шестнадцатеричное представление подразумевается). Возможны и другие варианты, но их не ис-

пользуют. Здесь в основном будем пользоваться первым способом, причем нулевое значение сегмента будем представлять кратко, то есть 0:0400. К счастью, в 32-разрядном (и 64-разрядном) режиме работы процессоров в современных ОС и приложениях сегментация не применяется, а адрес выражается одним (32- или 64-битным) числом.

Обозначение и порядок битов и байтов шин адреса и данных, принятое в аппаратуре PC, пришло от процессоров Intel 8086/88 (и даже от 8080). Самый младший бит (Least Significant Bit, LSB) имеет номер 0, самый старший (Most Significant Bit, MSB) бит байта — 7, слова — 15, двойного слова — 31. На рисунках принято старший бит изображать слева, а младший — справа.

### **Ячейки памяти, порты и регистры**

Поясним разницу между ячейками памяти, портами и регистрами. *Ячейки памяти* служат лишь для хранения информации — сначала ее записывают в ячейку, а потом могут прочитать, а также записать иную информацию. *Порты ввода-вывода*, как правило, служат для преобразования двоичной информации в какие-либо физические сигналы, и обратно. Например, порт данных параллельного интерфейса формирует электрические сигналы на разъеме, к которому обычно подключают принтер. Электрические сигналы, поступающие от принтера, порт состояния того же интерфейса отображает в виде набора битов, который может быть считан процессором. *Регистр* — довольно широкое понятие, которое зачастую используется как синоним порта. Регистры могут служить для управления устройствами (и их контроллерами) и для чтения их состояния. Регистры (как и порты) могут образовывать каналы.

– *Каналы ввода-вывода данных.* Пример: регистр данных СОМ-порта, байты, записываемые друг за другом в этот регистр, в том же порядке будут передаваться по последовательному интерфейсу, то есть поступать в *канал вывода*. Если этот интерфейс подключить к СОМ-порту другого компьютера и выполнять программные чтения его регистра данных, мы получим байт за байтом переданные данные. Таким образом, здесь регистр играет роль *канала ввода*.

– *Каналы управления.* Если запись в регистр определенных данных (битовых комбинаций) изменяет состояние некоего устройства (сигнал светофора, положение какого-то механизма), то регистр образует канал управления.



– *Каналы состояния*. Пример: регистр игрового порта (game-порт), к которому подключен джойстик. Чтение регистра дает информацию о состоянии кнопок джойстика (нажаты или нет).

Канал отличается от ячейки памяти рядом свойств. Если в *ячейку памяти* записывать раз за разом информацию, то последующее считывание возвращает результат последней записи, а все предшествующие записи оказываются бесполезными. Если ячейку памяти считывать раз за разом, не выполняя запись в нее, то результат считывания каждый раз будет одним и тем же (при исправной памяти). «Лишнее» спекулятивное чтение ячейки памяти не приведет ни к каким побочным эффектам. На этих свойствах «настоящей» памяти основаны методы ускорения работы с ней: кэширование и спекулятивное чтение. С *регистрами, образующими каналы*, такие вольности недопустимы. Здесь все обращения приводят к каким-либо изменениям. Кэширование и спекулятивное чтение недопустимы. Например, лишнее чтение регистра данных СОМ-порта «выдернет» байт из принимаемого потока. Операция чтения регистра состояния может быть неявным подтверждением сброса какого-либо признака (например, запроса прерывания), и она изменяет состояние устройства. Записи в канал данных (и управления) также нельзя опускать (для «ускорения»).

Каждый байт (ячейка памяти, порт, регистр) имеет собственный уникальный *физический адрес*. Этот адрес устанавливается на системной шине процессором, когда он инициирует обращение к данным ячейке или порту. По этому же адресу к этой ячейке (порту, регистру) могут обращаться и другие активные компоненты системы, так называемые *мастера шины*.

В семействе x86 и РС-совместимых компьютерах пространства адресов ячеек памяти и портов ввода-вывода разделены. Это предусмотрено с обеих сторон: процессоры позволяют, а компьютеры используют данное разделение. Нынешние 32-битные процессоры имеют разрядность физического адреса памяти 32 и даже 36 бит, что позволяет адресовать до 4 и 64 Гбайт соответственно. Пространство ввода-вывода использует только младшие 16 бит адреса, что позволяет адресовать до 65 384 однобайтных регистров. Адреса «исторических» системных устройств РС не изменились с самого рождения — это дань совместимости, которая без разделения пространств вряд ли бы обеспечивалась столько лет. Пространства памяти и портов ввода-

вывода неравнозначны не только по объему, но и по способам обращения. Способов адресации к ячейке памяти в x86 великое множество, в то время как для адресации ввода-вывода их существует только два. К памяти возможна (и широко используется) виртуальная адресация, при которой для программиста, программы и даже пользователя создается иллюзия оперативной памяти гигантского размера. К портам ввода-вывода обращаются только по реальным адресам, правда, и здесь возможна виртуализация, но уже чисто программными средствами операционной системы. И, наконец, самое существенное различие пространств памяти и портов ввода-вывода: процессор может считывать инструкции для исполнения только из пространства памяти. Конечно, через порт ввода можно считать фрагмент программного кода (что и происходит, например, при считывании данных с диска), но для того, чтобы этот код исполнить, его необходимо записать в память.

Регистры различных устройств могут быть приписаны как к пространству портов ввода-вывода, так и к пространству памяти. Под портом устройства, как правило, подразумевают регистр, связанный с этим устройством и приписанный к пространству портов ввода-вывода. Точность приведенной терминологии, конечно же, относительна. Так, к примеру, ячейки видеопамати (тоже память!) служат в основном не для хранения информации, а для управления свечением элементов экрана. Понятие Memory Mapped I/O означает регистры периферийных устройств, отображенные на пространство памяти (т. е. занимающие адреса именно в этом пространстве, а не в пространстве ввода-вывода).

Разделение пространств памяти и ввода-вывода было вынужденной мерой в условиях дефицита адресуемого пространства 16-битных процессоров и сохранилось во всех процессорах x86. В процессорах ряда других семейств такого разделения нет, и для нужд ввода-вывода используется выделенная область единого адресного пространства. Тенденция изживания пространства ввода-вывода наблюдается в современных спецификациях устройств и интерфейсов для PC.

## Диски и файловые системы

Дисковая память в компьютере используется для хранения *файлов* — цепочек байтов, несущих любую информацию. Для данного обсуждения содержимое файлов несущественно, в них могут быть исполняемые программные модули, данные (числовые, текстовые, мультимедийные) и их смесь. Способ размещения файлов на диске и возможности манипуляций с файлами определяются *файловой системой*. Каждый файл имеет набор атрибутов, состав которого зависит от используемой файловой системы. «Прожиточный минимум» — это имя файла, его длина и дата последней модификации. В простейшей файловой системе MS-DOS к этому минимуму добавляются атрибуты системного, скрытого, архивированного и только читаемого файла. Имя файла может задаваться как в классической форме «8.3» (8 символов на имя плюс 3 символа на тип), понимаемой всеми операционными системами, так и в почти произвольной длинной форме, характерной для более сложных ОС (OS/2, Windows 9x/NT/200x/XP, Unix). Файловая система включает в себя систему каталогов и системы размещения файлов на диске, простейшей из которых можно считать FAT в MS-DOS. Эти системы определяют возможности и эффективность манипулирования файлами — создания, записи, чтения, поиска, модификации, удаления, восстановления удаленных файлов, а также средства восстановления файловой системы после сбоев, вызванных неисправностями и некорректными действиями пользователей или программ.

Для каждой операционной системы характерны свои файловые системы (одна или несколько), которые она «понимает». Компьютер с незагруженной ОС «не понимает» ни одной файловой системы (в этом его универсальность). Изначально были приняты соглашения, позволяющие «голому» компьютеру загрузить ОС с так называемого *системного диска*, причем самая первая фаза загрузки выполняется без какой-либо файловой системы.

## Устройства ввода-вывода и коммуникаций

*Устройства ввода-вывода* связывают компьютер с внешним миром, без них он был бы «вещью в себе». Список устройств, делающих компьютер «вещью для нас», практически не ограничен. К ним относятся дисплеи (устройства отображения, то есть вывода), клавиа-

тура и мышь (устройства ввода), принтеры и сканеры, плоттеры и дигитайзеры, джойстики, акустические системы и микрофоны, телевизоры и видеокамеры и прочие устройства в великом множестве их разновидностей. Любопытно, что в этих парах обычно лидируют устройства вывода, появившиеся в компьютерах раньше соответствующих устройств ввода. Благодаря фантазии и техническому прогрессу появляются все новые и новые устройства. Например, шлем виртуальной реальности из области фантастики перешел в производственно-коммерческую область. К компьютеру можно подключать датчики и исполнительные устройства технологического оборудования, различные приборы, в общем, все, что в итоге может вырабатывать электрические сигналы и/или ими управляться.

*Консолью* компьютера называют его «выступающую часть», обращенную к пользователю. В РС стандартной консолью являются клавиатура (устройство ввода) и дисплей, позволяющий отображать символьную информацию (устройство вывода). На консоль выводятся все системные сообщения, с консоли можно управлять компьютером – запускать и принудительно завершать приложения, выполнять перезагрузку ОС и управлять процессом загрузки. Вместо стандартных устройств роль консоли могут играть и иные устройства, способные выводить и вводить символы и имеющие необходимую программную поддержку.

Коммуникационные устройства связывают компьютеры (и другие устройства) в сложные системы, составные части которых могут находиться довольно далеко друг от друга. Коммуникационные устройства обеспечивают передачу информации самого разного назначения. К этим устройствам относятся модемы, адаптеры локальных и глобальных сетей. Соответствующий набор устройств ввода-вывода и коммуникаций позволяет превратить персональный компьютер, например, в факс-машину, аппарат IP-телефонии (голосовой) или видео-конференцсвязи.

### **Адаптеры, контроллеры; иерархия подключений периферийных устройств**

Компоненты компьютера соединяются друг с другом иерархией средств подключения, наверху которой стоят *интерфейсы системного уровня подключения*. Для этой группы интерфейсов характерно то, что в их транзакциях фигурируют *физические адреса пространства*

памяти и (если есть) пространства ввода-вывода. Группа связанных между собой интерфейсов системного уровня образует логическую *системную шину компьютера*. Системную шину составляют следующие физические интерфейсы:

- *шина подключения центрального процессора* (или нескольких процессоров в сложных системах) — FSB (Front Side Bus — фасадная шина);

- *шина подключения контроллеров памяти*, оперативной и постоянной; собственно шина памяти (memory bus) системной уже не является, поскольку в ней фигурируют не системные адреса, а адреса физических банков памяти;

- *шины ввода-вывода*, обеспечивающие связь между центральной частью компьютера и периферийными устройствами.

Типичные представители шин ввода-вывода в IBM PC — шина ISA (отмирающая), а также шины PCI (развивающаяся в PCI-X) и PCI-E (PCI Express). Через шины ввода-вывода проходят все обращения центрального процессора (ЦП) к периферии. К шинам ввода-вывода подключаются *контроллеры* и *адаптеры* периферийных устройств или их интерфейсов.

*Адаптер* является средством сопряжения какого-либо устройства с какой-либо шиной или интерфейсом компьютера. *Контроллер* служит тем же целям сопряжения, но при этом подразумевается его некоторая активность — способность к самостоятельным действиям после получения команд от обслуживающей его программы. Сложный контроллер может иметь в своем составе и собственный процессор. На эти тонкости терминологии не всегда обращают внимание, и понятия «адаптер» и «контроллер» считают почти синонимами. Для взаимодействия с периферийными устройствами процессор обращается к регистрам контроллера (адаптера), «представляющего интересы» подключенных к нему устройств.

Часть периферийных устройств (ПУ) совмещена со своими контроллерами (адаптерами), как, например, сетевой адаптер Ethernet, подключенный к шине PCI. Другие же ПУ подключаются к своим контроллерам через промежуточные *периферийные интерфейсы*, находящиеся на нижнем уровне иерархии подключений. Периферийные интерфейсы самые разнообразные из всех аппаратных интерфейсов. К периферии, подключаемой через промежуточные интерфейсы, относится большинство устройств хранения (дисковые, ленточные), уст-

роЙств ввода-вывода (дисплеи, клавиатуры, мыши, принтеры, плоттеры), ряд коммуникационных устройств (внешние модемы).

Для взаимодействия с программой (с помощью процессора или сопроцессоров) адаптеры и контроллеры обычно имеют *регистры* ввода-вывода, управления и состояния, которые могут располагаться либо в адресном пространстве памяти, либо в пространстве портов ввода-вывода. Кроме того, используются механизмы аппаратных прерываний для сигнализации программе о событиях, происходящих в периферийных устройствах. Для обмена информацией с устройствами применяют также механизмы прямого доступа к памяти (Direct Memory Acces, DMA) и прямого управления шиной. Контроллер, который способен инициировать транзакции на системной шине, является активным компонентом компьютера. С помощью транзакций он может обращаться к другим устройствам (точнее, их контроллерам или адаптерам), обеспечивая равноправное взаимодействие. Чаще всего ограничиваются взаимодействием контроллера с системной памятью (это проще).

## 2. СИСТЕМНЫЕ ПЛАТЫ

Важнейшим узлом компьютера является *системная плата* (*system board*), иногда называемая *материнской* (*motherboard*), *основной* или *главной платой* (*main board*).

Материнская или системная плата – это тот фундамент, на котором построен любой современный компьютер, будь то настольный ПК, ноутбук или встраиваемая система. Именно *материнская плата* (МП) объединяет такие различные по своей сути и функциональности комплектующие, как процессор, оперативная память, платы расширения и всевозможные накопители. Именно благодаря материнской плате к компьютеру можно подключать периферийные устройства, ведь даже если набор системной логики (чипсет) поддерживает разнообразные шины и интерфейсы, к обычной микросхеме напрямую подключить, к примеру, принтер, вряд ли у кого получится.

Существует несколько наиболее распространенных формфакторов, учитываемых при разработке системных плат. *Формфактор* (*form factor*) представляет собой физические параметры платы и определяет тип корпуса, в котором она может быть установлена. Формфакторы системных плат могут быть стандартными (т. е. взаимозаменяемыми) или нестандартными. Нестандартные формфакторы, к сожалению, являются препятствием для модернизации компьютера. Наиболее известные формфакторы системных плат перечислены ниже:

*устаревшие:*

- Baby\_AT;
- полноразмерная плата AT;
- LPX;
- WTX (больше не производится);

*современные:*

- ATX;
- micro\_ATX;
- flex\_ATX;
- Mini\_ITX (разновидность flex\_ATX);
- NLX;

*Другие:*

- независимые конструкции (разработки компаний Compaq, Packard Bell, Hewlett Packard, портативные, мобильные системы и т. д.).

Основное назначение системной платы — соединение всех узлов компьютера в одно устройство, так что, по большому счету, она — это всего лишь набор проводов между контактами процессора и контактами модулей памяти и периферийных устройств. Все остальные расположенные на ней элементы носят второстепенные функции, служат только для развязки и согласования сигналов.

По сути, все материнские платы играют роль посредника между всеми компонентами, за счет которых происходит работа с информацией.

Материнская плата по существу очень важный компонент в вашей системе и, надо заметить, что общая производительность системы часто зависти от МП не менее, чем от процессора и ОЗУ.

### **PCB [Printed Circuit Board] – Печатная плата**

Печатная плата это и есть плата, на которой размещаются все компоненты. Конструктивно системная плата выполняется в виде многослойной текстолитовой печатной платы. Количество слоев может достигать 12, но чаще всего используют 8. Между каждым слоем располагаются печатные проводники, выполненные из металлической фольги, которые соединяют ножки микросхем, резисторов, конденсаторов и разъемов между собой. Так как современные процессоры работают с внешними устройствами на частоте в несколько сотен мегагерц, то длина и расположение печатных проводников теперь рассчитывается по тем же принципам, что и для СВЧ-устройств, когда каждый лишний сантиметр проводника играет огромную роль. В этом можно убедиться, осмотрев системную плату — некоторые проводники, например, проложены не по прямой линии, а с изгибами в виде пружины, что увеличивает их длину.

PCB обычно состоит из нескольких слоев, состоящих из плоских ламинированных пластин, между которыми находятся элементы цепи — соединительные линии, которые называются дорожки. Обычная PCB имеет четыре таких слоя, два слоя, которые находятся сверху и снизу являются сигнальными слоями. Два слоя, которые находятся посередине используются в качестве заземления и пластины питания (см. рис. 2.1). Путем помещения пластин питания и заземления в центр, была достигнута наивысшая коррекция и защита сигнала.





Рис.2.1. Четырёхслойная печатная плата

Для некоторых МП нужно шесть слоев, такими МП являются МП, разработанные для двухпроцессорных систем или же когда количество контактов процессора превышает 425 контакта. Это потому, что сигнальные дорожки должны быть расположены вдали друг от друга, чтобы предотвратить перекрестные помехи и дополнительные слои решают эту проблему. Платы, состоящие из шести слоев, могут иметь три или четыре сигнальных слоя, одну пластину заземления и одну или две пластины питания. В новых чипсетах три-четыре сигнальных слоя и две пластины питания вполне достаточно для того, чтобы выдерживать дополнительные потребности в энергопотреблении (см. рис. 2.2).



Рис. 2.2. Шестислойная печатная плата

Разметка и длина дорожек очень важна для нормальной работы системы. Основная задача – снизить любое искажение сигнала из-за пересечения дорожек. Чем длинней дорожка и/или выше скорость передачи сигнала, тем больше перекрестных помех, откуда следует, что нужно увеличить расстояние между дорожками. Некоторые дорожки должны быть максимальной длины для сохранения непрерывности сигнала, например, которые подходят напрямую к процессору. Все дорожки, подключенные к различным компонентам, должны быть не больше установленной допустимой длины.

Фактически **материнская плата** — это большая «коллекция разъемов», предназначенных для установки тех или иных комплектующих. Назовем лишь несколько основных типов.

**«Гнездо» для установки процессора.** На системных платах, предназначенных для установки процессоров Pentium и Pentium MMX, а также процессоров AMD и Cyrix M2, «гнездо» — квадратной формы, с многочисленными дырочками под «ножки» процессора по краю рамки. Такой тип разъема носит название Socket 7. На платах, предназначенных для установки процессора Pentium II – Pentium IV (Slot 370, Slot A, Slot 487), а также процессора фирмы AMD под названием K7 установлено не квадратное гнездо, а длинный «щелевой» разъем — слот (тип Slot 1). На большинстве материнских плат, предназначенных для построения домашних систем, сокет, или слот, имеется в единственном экземпляре.

**Слоты для подключения питания:** Бывают двух типов AT и ATX. Все зависит от класса материнской платы. По размеру они одинаковы, но конструктивные отличия не дают их перепутать между собой. В основном сейчас используются ATX питание для материнских плат.

**Разъемы-слоты стандарта PCI** слоты для подключения 32 разрядных устройств. Как правило, их четыре. Разъемы PCI обычно самые короткие на плате, белого цвета, разделенные своеобразной «перемычкой» на две неравные части. Устройствами, устанавливаемыми в PCI слот, могут быть видеокарта, звуковые карты типа PCI, внутренние модемы, платы контроллера (SCSI), платы подключения USB – составляющих, сетевые карты.

**Разъем AGP** — специальный, более быстрый с точки зрения пропускной способности слот, предназначенный для установки видеокарт формата AGP. Еще совсем недавно стандарт AGP, позволяющий видеокартам работать с немыслимой доселе скоростью и использовать не только свою оперативную память, но и оперативную память компьютера, был изюминкой и исключительной особенностью материнских плат Pentium II – IV. Однако появились снабженные этим разъемом платы старого формата Socket 7.

**Разъемы-слоты типа ISA** — слоты для подключения 16 разрядных устройств. Гораздо более слабые в отношении пропускной способности, чем слоты PCI, слоты ISA остались еще со времен компьютеров типа 386. Но именно эти слоты, а не их более быстрые собратья PCI, являются наиболее дефицитными. Именно к ним до сих пор под-

ключается громадное количество дополнительных карт: звуковые платы, внутренние модемы, сетевые карты, специализированные платы сканеров и т. д.. Их на плате бывает от 2 до 4 под Socet 7, и от 1 до 2 под Slot 1. По внешнему виду слоты ISA напоминают слоты PCI, только они почти в полтора раза длиннее и цвет их не белый, а черный. В наше время почти прекращен выпуск устройств под ISA слоты.

**Слоты для установки оперативной памяти** от слотов для установки плат отличаются наличием специальных замочков-«защелок». На платах Pentium MMX, как правило, предусмотрена установка двух типов памяти — более старого формата SIMM (72 контакта) и быстрой памяти типа DIMM (168 контактов). Слоты DIMM значительно короче слотов SIMM. Количество слотов этих типов может варьироваться от 2 до 4. Бывает, что на материнской плате установлены разъемы под один тип памяти. В конце 20 века выпуск памяти SIMM был полностью прекращен, что позволило перейти на более быструю и в данный момент дешевую память DIMM. Для компьютеров типа Pentium IV используют новый тип разъемов RIMM.

**Контроллеры портов.** Под «портами» понимаются разъемы на задней стенке компьютера, предназначенные для подключения таких внешних устройств, как принтер, мобильный дисковод большой емкости (для этого предусмотрен так называемый параллельный порт, или LPT), внешний модем, манипулятор типа «мышь» (через два последовательных порта — COM1 и COM2 с разъемами 9 и 25 штырьков). Параллельный порт (LPT) всегда один, число же портов COM может варьироваться от 2 до 4. Во многих платах Pentium II и выше, соответствующих стандарту ATX, имеются еще и специальные разъемы для подключения мыши и клавиатуры — круглые разъемы типа PS/2. В настоящее время широко применяется последовательный порт USB, позволяющий подключить к компьютеру множество внешних устройств, соединенных в своеобразную цепочку (первое устройство включается в компьютер, второе подключается к первому и т.д.). Устройств в формате USB с каждым днем становится все больше и больше: модемы, принтеры, сканеры, манипуляторы мышь и т. д.

**Интерфейс IDE**, или ATA - простой и недорогой интерфейс для PC. Все функции по управлению накопителем обеспечивает встроенный контроллер, а 40-проводной соединительный кабель является фактически упрощенным сегментом 16-разрядной магистрали AT-Bus (ISA). Простейший адаптер IDE содержит только адресный дешифратор – все остальные сигналы заводятся прямо на разъем ISA.

Адаптеры IDE обычно не содержат собственного BIOS – все функции поддержки IDE встроены в системный BIOS PC. Однако интеллектуальные или кэширующие контроллеры могут иметь собственный BIOS, подменяющий часть или все функции системного. Основным режимом работы устройств IDE – программный обмен под управлением центрального процессора, однако все современные винчестеры EIDE поддерживают обмен в режиме DMA, а большинство контроллеров – режим Bus Mastering.

### Чипсеты

Отдельный и самый важный компонент МП – это чипсет.

Для того, чтобы процессор в персональном компьютере мог работать в полную силу, ему требуется помощь специализированных микросхем, которые берут на себя рутинную работу с оперативной памятью и периферийными устройствами. Набор таких микросхем называется чипсетом. В комплект чипсета может входить различное количество микросхем, но в последнее время наиболее популярно решение из 2 – 3 микросхем.

Как говорилось ранее, чипсет определяет, какой процессор поддерживается, какая память может быть использована и набор других характеристик. Обычно это 2-3 микросхемы.

На рис. 2.3 показан в упрощенном виде принцип построения электроники системной платы. На рисунке, в центре, между процессором, модулями оперативной памяти и внешними устройствами расположен **чипсет** — набор микросхем, которые выполняют служебные функции по распределению сигналов между всеми блоками. При подаче напряжения питания чипсет вырабатывает определенную последовательность команд, которая активизирует процессор. Процессор, в свою очередь, по программе BIOS тестирует и активизирует остальные устройства, установленные и подключенные к системной плате. Если старт компьютера прошел успешно, то микросхемы чипсета связывают процессор, память и периферийные устройства в единое целое — вычислительное устройство, готовое выполнить команды пользователя или определенным образом реагировать на появление сигналов в интерфейсных линиях.



Рис. 2.3. Принцип работы системной платы

### *Архитектура чипсета*

Не для кого не секрет, что часто производители процессоров влияют на формирование дизайна материнских плат. Это делается, прежде всего, для обеспечения лучшей работы процессора, обеспечения его стабильности и надежности. К примеру, корпорация Intel начала разрабатывать свои собственные чипсеты с выходом процессора 486. AMD также была вынуждена выпустить свой собственный чипсет для поддержания раннего процессора Athlon, архитектура которого впервые значительно отличалась от привычной архитектуры процессоров Intel.

### **Чипсеты на Юге и Севере**

На рис.2.4 схематически изображено устройство архитектуры любой материнской платы. Как видно из этого рисунка, чипсет материнской платы состоит из двух компонентов (которые, как правило, представляют собой независимые чипсеты, связанные друг с другом). Называются эти компоненты Северный и Южный мост. Названия Северный и Южный – исторические. Они означают расположение чипсета моста относительно шины PCI: Северный находится выше, а Южный – ниже. Почему мост? Это название дали чипсетам по выполняемыми ими функциями, они служат для связи различных шин и интерфейсов. Для проектировщика особой сложностью является Северный мост, т.к. он работает с самыми скоростными устройствами, поэтому сам должен работать очень быстро, обеспечивая быструю и надежную связь процессора, памяти, шины AGP и Южного моста. Южный мост работает с медленными устройствами, такими как жесткие диски, шина USB, PCI, ISA и т. п.

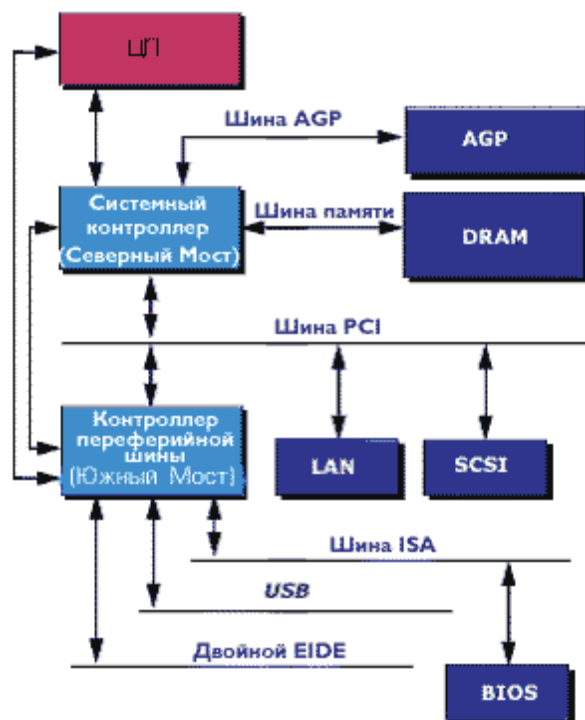


Рис. 2.4. Архитектура системы ПК высшего уровня

### Зачем нужны два моста

Почему производители разделили чипсет на два моста? Первая и, наверное, самая главная причина заключается в выполняемых чипами функциях. Северный мост должен работать гораздо быстрее, чем Южный. Разработка же обоих мостов на одном чипе значительно усложняет разработку и производство такого чипсета. Кроме того, обновление стандартов периферии происходит очень часто. При использовании двух чипсетов производителям материнских плат нет необходимости полностью менять весь набор логики – достаточно поменять Южный мост. Размер самого ядра чипсета намного меньше кремниевой подложки, на которой он находится и это необходимо для того, чтобы грамотно развести проводники от ядра процессора к его ножкам-выходам. Таким образом, в чипсете остается довольно много места, которое исчезает при использовании вместо двух чипсетов одного. В северный мост, как раз из-за неиспользованного пространства, некоторые производители встраивают наборы графики и кэш-память третьего уровня.

## **Intel назвал их хабами**

С недавнего времени (с выпуском корпорацией Intel чипсета i815) корпорация Intel отказалась от использования архитектуры мостов, и перешла к похожей архитектуре, в которой используются *хабы*. На первый взгляд, все тоже самое: два чипсета, один был раньше Северным мостом, а теперь называется «Хаб контроля памяти» («Memory Controller Hub»), другой же напоминает Южный мост и называется «Хаб контроля за вводом/выводом» (I/O Controller Hub), но не все так просто. Функции чипсетов (хабов) не менялись, просто в них добавлено несколько конструктивных изменений: в частности, чипсеты стали более независимы, а интерфейс связи друг с другом представляет собой связь «один-к-одному» (point-to-point). Такой подход оказался лучше, чем классический Южный мост, который можно сказать все устройства «сажает» на шину PCI и по ней же передает данные в Северный мост.

## **Модернизация чипсетов**

При изучении возможностей чипсетов пользователям следует обратить внимание, что если ранее разработка нового чипсета сопровождалась значительным увеличением производительности компьютера и появлением новых функций, то ныне разработчики исповедуют идеологию «ползучей» модернизации, когда следующий тип чипсета мало отличается от предшественника Т. е., в новом чипсете совершенствуют какую-то одну функцию или добавляют поддержку того или иного стандарта, например, работу с определённого вида памятью. Кроме того, имеет место разработка в рамках одного типа чипсета целого набора микросхем (несколько вариантов южного и северного мостов), которые производители системных плат могут произвольно комбинировать. В частности, в качестве южного моста могут применяться микросхемы, разработанные для предыдущего типа чипсета.

## **Пример конструкции системной платы.**

На фотографии (рис.2.5) изображена типичная материнская плата P5GDC-V Deluxe производства известной тайваньской компании Asus.

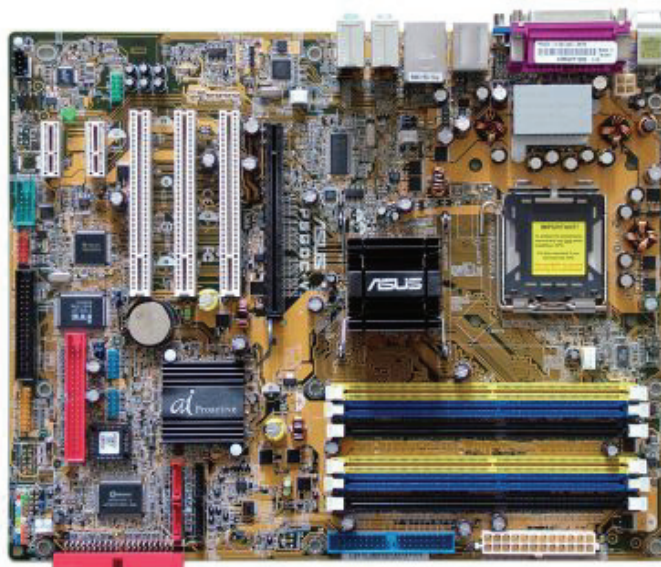


Рис. 2.5. Материнская (системная) плата

Эта плата на основе набора системной логики Intel 915G рассчитана на процессоры Intel Pentium 4 в корпусировке LGA 775 и поддерживает почти все технологии, встречающиеся в современных настольных компьютерах. Краткие характеристики этой модели:

- чипсет – 915G со встроенным графическим ускорителем («северный мост») + ICH6R («южный мост»);
- поддержка процессоров Pentium 4 или Celeron D в корпусировке LGA 775;
- поддержка оперативной памяти DDR и DDR2 533 объемом до 4 Гбайт;
- поддержка шины PCI Express x16 и x1, PCI ;
- поддержка скоростных интерфейсов USB 2.0 и IEEE 1394 (FireWire);
- контроллеры IDE и Serial ATA, гигабитный сетевой контроллер, восьмиканальный (7.1) звуковой контроллер;
- Форм-фактор – ATX (размеры - 305 x 244 мм).

Рассмотрим, где же на плате расположены эти компоненты.

Сразу в глаза бросается разъем для процессора, на который наклеена защитная желтая наклейка, одновременно призывающая к осторожности при установке чипа.

Левее расположен массивный радиатор с надписью ASUS, под ним скрывается «северный мост» - микросхема 915G, ответственная,



прежде всего, за связь процессора и оперативной памяти. Кроме того, в данном случае в нее встроен графический контроллер Intel Graphics Media Accelerator 900. Для того, чтобы собранный на основе этой платы компьютер заработал, не требуется установка видеокарты. Впрочем ничто не мешает поставить и более мощный графический ускоритель.

Левее и чуть ниже находится еще один массивный радиатор с фирменным названием семейства материнских плат Asus. Этим радиатором накрыт «южный мост» ICH6R, отвечающий за системы ввода – вывода и включающий в себя контроллер USB 2.0, а также RAID-контроллер.

Ниже процессорного гнезда и «северного моста» расположены слоты для оперативной памяти. Рассматриваемая нами плата отличается тем, что в ней могут использоваться как привычные модули памяти DDR, так и более дорогие, но потенциально более производительные модули памяти DDR2. Как видим, слотов шесть и они выделены цветом в пары, это значит, что плата рассчитана на оперативную память, работающую в двухканальном режиме, т. е. в данном случае модули памяти одинакового типа и объема должны устанавливаться попарно в слоты одного цвета. При этом в плату можно установить либо четыре модуля памяти DDR, либо два модуля типа DDR2, одновременная работа памяти различных типов не поддерживается (как, впрочем, и в любых других подобных системных платах).

Непосредственно под слотами для оперативной памяти размещены разъем IDE для накопителей (слева) и разъем для подвода к материнской плате электротока от блока питания.

Поскольку для современных процессоров Pentium 4 требуется дополнительное электропитание, в правом верхнем углу платы предусмотрен соответствующий разъем для подключения к блоку питания.

Под микросхемой «южного моста» можно увидеть четыре разъема Serial ATA, к которым подключаются жесткие диски. Благодаря чипу ICH6R из этих винчестеров можно создавать RAID – массивы, ускоряющие работу с дисковой подсистемой, либо повышающие надежность хранения данных.

Слева от «южного моста» расположен еще один разъем IDE для подключения винчестеров или оптических накопителей по традиционной параллельной шине ATA.

Разъем для подключения флоппи-дисководов расположен у левого края платы. Хотя сегодня дискетами уже мало кто пользуется, мно-

гие производители системных плат не торопятся отказываться от этого рудимента прошлого, по всей вероятности, во имя совместимости. Впрочем, в компьютерах Apple флоппи-дисководы не устанавливаются уже несколько лет, и этот факт никого не возмущает.

В левой верхней части системной платы расположены слоты для установки плат расширения. Два крайних слева маленьких слота – это разъемы для плат с интерфейсом PCI Express x1 (теоретическая полоса пропускания – до 250 Мбайт/с). Далее следуют три слота PCI, а черным цветом выделен слот PCI Express x16 (теоретическая полоса пропускания - до 4000 Мбайт/с) для установки видеокарты.

В правой верхней части распаяны различные разъемы, которые после установки платы в системный блок будут выведены на его заднюю стенку. Рассмотрим их подробнее. Слева, один над другим, размещены два порта PS/2 для подключения мыши и клавиатуры. Затем идут параллельный порт, цифровой аудиointерфейс S/PDIF, аналоговый видеовыход VGA (15-контактный разъем D-Sub) для подключения монитора, порт IEEE 1394 или FireWire (вверху) и два порта USB 2.0, сетевой порт и еще два порта USB 2.0, и, наконец, шесть аудиопортов от встроенного звукового контроллера.

### **Производители системных плат**

Производство системных плат сопряжено с меньшими техническими трудностями и требует меньших финансовых возможностей, чем производство микросхем, поэтому фирм, которые разрабатывают и производят системные платы на порядок больше, чем производителей процессоров и чипсетов. На основе одной и той же системной логики можно выпускать самые разнообразные системные платы, которые будут отличаться своими функциональными возможностями и надежностью. Соответственно, однотипные системные платы (для конкретного типа процессора и на определенном типе чипсета) от разных производителей могут иметь весьма различные потребительские качества. Причем, даже у именитых фирм могут быть неудачные системные платы, а малоизвестный производитель может разработать и выпустить плату, которая будет иметь технические характеристики, превосходящие образцы ведущих компаний.

При выборе новой системной платы следует руководствоваться информацией на сайтах производителей, а также отзывами пользователей, которые можно найти в Интернете.

### 3. ПРОЦЕССОРЫ

#### 3.1 Структура и принципы работы процессоров

Микропроцессор, иначе, центральный процессор (ЦП) — Central Processing Unit (CPU) — функционально законченное программно-управляемое устройство обработки информации, выполненное в виде одной или нескольких больших (БИС) или сверхбольших (СБИС) интегральных схем (БИС или СБИС).

В общем случае в состав микропроцессора входят: арифметико-логическое устройство (АЛУ), блок управления и синхронизации, ЗУ, регистры и другие блоки, необходимые для выполнения операций вычислительного процесса. Как БИС микропроцессор характеризуется степенью интеграции, потребляемой мощностью, помехоустойчивостью, нагрузочной способностью активных выводов (определяющей возможность подключения к данному микропроцессору других БИС), технологией изготовления, типом корпуса, техническим ресурсом, устойчивостью к механическим, климатическим и радиационным воздействиям. Как вычислительное устройство микропроцессор характеризуется производительностью, разрядностью обрабатываемых данных и выполняемых команд, возможностью увеличения разрядности, числом команд (микрокоманд), количеством внутренних регистров, возможностью обеспечения режима прерывания (уровней приоритета), способностью к обработке десятичных кодов, объемом адресной памяти, наличием канала прямого доступа к памяти, типом и числом входных и выходных шин и их разрядностью, наличием и видом программного обеспечения, способом управления.

Микропроцессоры, используемые в средствах вычислительной техники различного назначения (для решения широкого круга разнотипных задач), называются *универсальными*. Микропроцессоры, предназначенные для построения какого-либо одного типа вычислительных устройств, называются *специализированными*. По способу управления различают микропроцессоры со *схемным* и *микропрограммным управлением*. По структуре микропроцессоры подразделяются на *секционированные* (как правило, с микропрограммным управлением) и *однокристалльные* (с фиксированной разрядностью и постоянным набором команд).

Микропроцессоры получили широкое применение в системах управления технологическим и контрольно-испытательным оборудованием, транспортными средствами, космическими аппаратами, бытовыми приборами и т. д. Малые размеры, масса и энергоемкость микропроцессора позволяют встраивать его непосредственно в объект управления. На базе микропроцессора создаются различные типы микро-ЭВМ, контроллеров, программаторов и другие устройства автоматики и вычислительной техники.

Для МП на БИС или СБИС характерны:

- простота производства (по единой технологии);
- низкая стоимость (при массовом производстве);
- малые габариты (пластина площадью несколько квадратных сантиметров или кубик со стороной несколько миллиметров);
- высокая надежность;
- малое потребление энергии.

Микропроцессор выполняет следующие функции:

- чтение и дешифрацию команд из основной памяти;
- чтение данных из ОП и регистров адаптеров внешних устройств;
- прием и обработку запросов и команд от адаптеров на обслуживание ВУ;
- обработку данных и их запись в ОП и регистры адаптеров ВУ;
- выработку управляющих сигналов для всех прочих узлов и блоков ПК.

### **Архитектура микропроцессора**

**Архитектура микропроцессора (Architecture)** — принцип его внутренней организации, общая структура, конкретная логическая структура отдельных устройств. Понятие архитектуры микропроцессора включает в себя систему команд и способы адресации, возможность совмещения выполнения команд во времени, наличие дополнительных устройств в составе микропроцессора, принципы и режимы его работы. Выделяют понятия микроархитектуры и макроархитектуры.

*Микроархитектура микропроцессора* — это аппаратная организация и логическая структура микропроцессора, регистры, управляющие схемы, арифметико-логические устройства, запоминающие устройства и связывающие их информационные магистрали. *Макроархитектура микропроцессора* — это система команд, типы обрабатываемых данных, режимы адресации и принципы работы микропроцес-

сора. В общем случае под архитектурой ЭВМ понимается абстрактное представление машины в терминах основных функциональных модулей, языка ЭВМ, структуры данных.

В соответствии с архитектурными особенностями, определяющими свойства системы команд, различают:

– **микропроцессоры с CISC – архитектурой.**

**CISC** (Complex Instruction Set Computer) — компьютер со сложной системой команд. Исторически они первые и включают большое количество команд. Все микропроцессоры корпораций Intel и AMD относятся к категории CISC;

– **микропроцессоры с RISC – архитектурой.**

**RISC** (Reduced Instruction Set Computer) — компьютер с сокращенной системой команд. Упрощена система команд и сокращена до такой степени, что каждая инструкция выполняется за единственный такт. Вследствие этого упростилась структура микропроцессора, и увеличилось его быстродействие.

– **Микропроцессоры с MISC – архитектурой.**

**MISC** (Minimum Instruction Set Computer) — компьютер с минимальной системой команд. Последовательность простых инструкций объединяется в пакет, таким образом программа преобразуется в небольшое количество длинных команд.

*Разрядность* — максимальное количество разрядов двоичного кода, которые могут обрабатываться или передаваться одновременно.

Разрядность микропроцессора обозначается  $m/n/k/$  и включает:

$m$  — разрядность внутренних регистров, определяет принадлежность к тому или иному классу процессоров;

$n$  — разрядность шины данных, определяет скорость передачи информации;

$k$  — разрядность шины адреса, определяет размер адресного пространства.

*Объем адресуемой памяти* — максимальный объем памяти, который может обслужить микропроцессор.

32-х разрядный микропроцессор может обслужить 64 Гб ( $4 \times 10^9$  байт) памяти, а 64-х разрядный микропроцессор может обслужить 64 Тб ( $64 \times 10^{12}$  байт) памяти.

*Набор дополнительных инструкций* (Instruction Set) применяется в современных CISC-микропроцессорах и способен значительно ускорить их работу, при условии поддержки данных наборов со стороны приложения.

*Тактовая частота* — это количество электрических импульсов в секунду. Тактовая частота задается кварцевым генератором — одним из блоков, расположенных на материнской плате. Тактовая частота кварцевого генератора выдерживается с очень высокой точностью.

*CPU Clock* — внутренняя частота процессора, на которой работает его вычислительное ядро. Эта частота на системной плате не присутствует, но на плате могут быть средства задания коэффициента умножения. Коэффициент умножения большинства современных процессоров фиксирован изготовителем; раньше его задавали перемычками на системной плате, заземляющими определенные выводы процессора. Новые процессоры способны динамически изменять коэффициент умножения (для управления потреблением).

Микропроцессор, работающий на частоте 2 ГГц, выполняет 2 миллиарда рабочих тактов в секунду. В зависимости от сложности обрабатываемой команды процессору для выполнения задачи необходимы сотни и тысячи тактов. Но для выполнения простых операций бывает достаточно одного такта. Чем выше тактовая частота ядра, тем выше скорость обработки данных.

### **Процессорная шина.**

Любой процессор обязательно оснащен *процессорной шиной*, которую в среде x86 CPU принято называть FSB (Front Side Bus). Эта шина служит каналом связи между процессором и всеми остальными устройствами в компьютере: памятью, видеокартой, жестким диском, и т. д. Между собственно памятью и процессором находится *контроллер памяти*. Соответственно процессор посредством FSB связывается с контроллером памяти, а уже тот, в свою очередь, по специальной шине — с модулями ОЗУ на плате. Однако, повторимся: поскольку «внешняя» шина у классического x86 CPU всего одна, она используется не только для работы с памятью, но и для общения процессора со всеми остальными устройствами.

Микропроцессор имеет две частоты — тактовая частота ядра и частота системной шины.

Чем выше частота системной шины, тем выше скорость передачи данных между микропроцессором и остальными устройствами.

### **Контроллер памяти**

В большинстве современных систем процессор как устройство к памяти обращаться не может, так как не имеет в своем составе со-

ответствующих узлов. Поэтому он обращается к «промежуточному» специализированному устройству, называемому *контроллером памяти*, а уже тот, в свою очередь — к микросхемам ОЗУ, размещенным на модулях памяти. Роль контроллера ОЗУ, таким образом, проста: он служит своего рода «мостом» между памятью и использующими ее устройствами. Как правило, контроллер памяти входит в состав чипсета — набора микросхем, являющегося основой системной платы. От быстродействия контроллера во многом зависит скорость обмена данными между процессором и памятью.

### **Кэш - память**

Кэш-память — быстрая память малой емкости, используемая процессором для ускорения операций, требующих обращения к памяти. Кэш — промежуточное звено между микропроцессором и оперативной памятью.

В один какой-то момент программа работает не со всей памятью целиком, а, как правило, с относительно маленьким её фрагментом. Какое решение напрашивается? Загрузить этот фрагмент в «быструю» память, обработать его там, а потом уже записать обратно в «медленную» (или просто удалить из кэша, если данные не изменялись). В общем случае, именно так и работает процессорный кэш. Любая считываемая из памяти информация попадает не только в процессор, но и в кэш. И если эта же информация (тот же адрес в памяти) нужна снова, сначала процессор проверяет: а нет ли её в кэше? Если есть — информация берется оттуда, и обращения к памяти не происходит вовсе. Аналогично с записью: информация, если её объем влезает в кэш — пишется именно туда, и только потом, когда процессор закончил операцию записи, и занялся выполнением других команд, данные, записанные в кэш, параллельно с работой процессорного ядра «потихоньку выгружаются» в ОЗУ.

Разумеется, объем данных, прочитанных и записанных за все время работы программы, намного больше объема кэша. Поэтому некоторые из них приходится время от времени удалять, чтобы в кэш могли поместиться новые, более актуальные. Самый простой из известных механизмов обеспечения данного процесса — отслеживание времени последнего обращения к данным, находящимся в кэше. Так, если нам необходимо поместить новые данные в кэш, а он уже «забит под завязку», контроллер, управляющий кэшем, смотрит: к какому фрагменту кэша не происходило обращения дольше всего? Именно

этот фрагмент и является первым кандидатом на «вылет», а на его место записываются новые данные, с которыми нужно работать сейчас. Вот так, в общих чертах, работает механизм кэширования в процессорах.

А для того чтобы было понятно, насколько важен кэш, приведем простой пример: скорость обмена данными процессора Pentium 4 со своим кэшем более чем в 10 раз (!) превосходит скорость его работы с памятью. Фактически, в полную силу современные процессоры способны работать только с КЭШем. Опять-таки, простой пример: выполнение простейшей инструкции процессором происходит за один такт, то есть за секунду он может выполнить такое количество простых инструкций, какова его частота (на самом деле еще больше). А вот время ожидания данных из памяти может в худшем случае составить более 200 тактов!

### **Дешифратор (декодер)**

На самом деле, исполнительные блоки всех современных x86-процессоров вовсе не работают с кодом в стандарте x86. У каждого процессора есть своя, «внутренняя» система команд, не имеющая ничего общего с теми командами (тем самым «кодом»), которые поступают извне. В общем случае, команды, исполняемые ядром — намного проще, «примитивнее», чем команды стандарта x86. Именно для того, чтобы процессор «внешне выглядел» как x86 CPU, и существует такой блок как декодер; он отвечает за преобразование «внешнего» x86-кода во «внутренние» команды, исполняемые ядром (при этом достаточно часто одна команда x86-кода преобразуется в несколько более простых «внутренних»). Декодер является очень важной частью современного процессора: от его быстродействия зависит то, насколько постоянным будет поток команд, поступающих на исполняющие блоки. Ведь они неспособны работать с кодом x86, поэтому, будут они что-то делать, или простаивать во многом зависит от скорости работы декодера.

### **Исполняющие (функциональные) устройства**

Пройдя через все уровни кэша и декодер, команды попадают на исполняющие устройства. Именно исполняющие устройства и являются единственно необходимым элементом процессора. Можно обойтись без кэша — скорость снизится, но программы работать будут. Можно обойтись без декодера — исполняющие устройства станут сложнее, но работать процессор будет. В конце концов, ранние процессоры архитектуры x86 (i8086, i80186, 286, 386, 486, Am5x86) как-



то без декодера обходились. Без исполняющих устройств обойтись невозможно, ибо именно они исполняют код программы. В самом первом приближении они традиционно делятся на две больших группы: арифметико-логические устройства (ALU) и блок вычислений с плавающей точкой (FPU).

### **Арифметико-логические устройства**

ALU традиционно отвечают за два типа операций: арифметические действия (сложение, вычитание, умножение, деление) с целыми числами, логические операции с целыми числами (логическое «и», логическое «или», «исключающее или», и т. п.). Что, собственно, и следует из их названия. Блоков ALU в современных процессорах, как правило, несколько. Понятно, что ALU может исполнить только те команды, которые предназначены для него. Распределением команд, поступающих с декодера, по различным исполняющим устройствам, занимается специальный блок.

### **Блок вычислений с плавающей запятой**

FPU занимается выполнением команд, работающих с числами с плавающей запятой, кроме того, традиционно он выполняет дополнительные наборы команд (MMX, 3DNow!, SSE, SSE2, SSE3...) — независимо от того, работают они с числами с плавающей запятой, или с целыми. Как и в случае с ALU, отдельных блоков в FPU может быть несколько, и они способны работать параллельно.

### **Регистры процессора**

*Регистр* — совокупность устройств, используемых для хранения информации, и обеспечения быстрого доступа к ней.

Регистры — по сути, те же ячейки памяти, но «территориально» они расположены прямо в процессорном ядре. Разумеется, скорость работы с регистрами во много раз превосходит как скорость работы с ячейками памяти, расположенными в основном ОЗУ, так и с кэшами любого уровня. Поэтому большинство команд архитектуры x86 предусматривает осуществление действий именно над содержимым регистров, а не над содержимым памяти. Однако общий объем регистров процессора, как правило, очень мал — он не сравним даже с объемом кэшей первого уровня. Код программы (не на языке высокого уровня, а именно бинарный, «машинный») часто содержит следующую последовательность операций: загрузить в один из регистров процессора

информацию из ОЗУ, загрузить в другой регистр другую информацию (тоже из ОЗУ), произвести некое действие над содержимым этих регистров, поместив результат в третий, потом снова выгрузить результат из регистра в основную память.

### **Сегментные регистры**

*Сегментом* называется область, которая начинается на границе параграфа, т.е. по любому адресу кратному 16. Хотя сегмент может располагаться в любом месте памяти и иметь размер до 64 Кбайт, он требует столько памяти, сколько необходимо для выполнения программы.

*Сегмент кодов CS* содержит машинные команды, которые будут выполняться. Обычно первая выполняемая команда находится в начале этого сегмента и операционная система передает управление по адресу данного сегмента для выполнения программы.

*Сегмент данных DS* содержит определенные данные, константы и рабочие области, необходимые программе.

*Сегмент стека SS* содержит адреса возврата как для программы при возврате в операционную систему, так и для вызовов подпрограмм при возврате в главную программу.

Еще один сегментный регистр — *дополнительный регистр сегмента ES* — предназначен для специального использования.

### **Регистры общего назначения**

*Регистр AX* является основным сумматором и применяется для всех операций ввода-вывода, некоторых операций над строками и некоторых арифметических операций.

*Регистр BX* является базовым регистром. Это единственный регистр общего назначения, который может использоваться в качестве индекса для расширенной адресации.

*Регистр CX* является счетчиком. Он необходим для управления числом повторений циклов и для операций сдвига влево или вправо. Регистр (CX) используется также для вычислений.

*Регистр DX* является регистром данных. Он применяется для некоторых операций ввода-вывода и операций умножения и деления над большими числами.

*Регистровые указатели SP и BP* обеспечивают системе доступ к данным в сегменте стека. Регистр SP обеспечивает использование стека в памяти, позволяет временно хранить адреса и иногда данные.

Этот регистр связан с регистром SS для адресации стека. Регистр BP облегчает доступ к параметрам (данным и адресам, переданным через стек).

*Индексные регистры SI и DI* применяются для расширенной адресации и для использования в операциях сложения и вычитания. Регистр SI является индексом источника и применяется для некоторых операций над строками. Регистр DI является индексом назначения и применяется также для некоторых операций над строками.

*Регистр командного указателя IP* содержит логический адрес текущей инструкции.

*Флаговый регистр* определяет текущее состояние машины и результаты выполнения операций (проверка четности, переполнения, переносов, знака).

### **Блок FPU**

*Блок FPU* содержит стек из восьми 80-битных регистров и несколько вспомогательных регистров. FPU присутствует во всех современных процессорах; его не было в 486SX, а в предыдущих процессорах он как *математический сопроцессор* устанавливался дополнительно. Блок FPU предназначен для расширения вычислительных возможностей центрального процессора — выполнения арифметических операций, вычисления основных математических функций (тригонометрических, экспоненциальных, логарифмических) и т. д. В разных поколениях процессоров он назывался по-разному — FPU (Floating Point Unit — блок чисел с плавающей точкой) или NPX (Numeric Processor extension — числовое расширение процессора). Сопроцессор поддерживает семь типов данных: 16-, 32-, 64-битные целые числа; 32-, 64-, 80-битные числа с плавающей точкой (FP-форматы) и 18-значные числа в двоично-десятичном (BCD) формате. Применение сопроцессора повышает производительность вычислений в сотни раз.

### **Блок MMX и расширение 3DNow!**

*Технология MMX* и расширение 3DNow! (введена фирмой AMD) ориентирована на приложения мультимедиа, 2D/3D-графику и коммуникации. Основная идея MMX заключается в одновременной обработке нескольких элементов данных за одну инструкцию, так называемая технология *SIMD* (Single Instruction — Multiple Data). Расширение MMX использует новые типы целочисленных данных, упа-

кованных в 64-битных регистрах: восемь байтов, четыре слова, два двойных слова или одно учетверенное слово. Эти типы данных обрабатываются в 64-битных регистрах MMX0–MMX7, представляющих собой младшие 64 бита 80-битных регистров FPU. Блок MMX присутствует практически на всех современных процессорах.

Каждая инструкция MMX выполняет действие сразу над всем комплектом операндов (8, 4, 2 или 1), размещенных в адресуемых регистрах. Блок MMX (и поддержка соответствующих инструкций) имеется на всех современных процессорах, начиная с 6-го поколения. В расширении SSE набор целочисленных инструкций для блока MMX расширен.

Расширение 3DNow! дает заметный результат при обработке графики, хотя не претендует на вытеснение графических ускорителей, а призвано служить мощным дополнением. Инструкции сигнальных процессоров (DSP) позволяют повысить производительность таких приложений, как программные модемы (включая ADSL), MP3 и процессоры объемного звучания (Dolby Digital surround sound).

В процессорах Intel расширение 3DNow! не используется.

### **Блок XMM и расширение SSE**

*Расширение SSE* (Streaming SIMD Extensions — потоковые SIMD-расширения) предназначено для ускорения обработки больших потоков данных в формате с плавающей точкой. Потоковое расширение реализуется на аппаратном **блоке XMM**, в котором первоначально было восемь 128-битных регистров XMM0 – XMM7. В 64-битном расширении число регистров XMM увеличили до 16. Регистры XMM позволяют работать с различными операндами:

- четырьмя вещественными числами одинарной точности (32-битные);
- парой вещественных чисел двойной точности (64-битные, только в SSE2);
- упакованными целыми числами: 16 байт, 8 слов, 4 двойных слова или пара учетверенных (по 64 бита) слов (только в SSE2).

Блок позволяет выполнять векторные (они же пакетные) и скалярные инструкции. **Векторные инструкции** реализуют операции сразу над всеми комплектами операндов. **Скалярные инструкции** работают с одним комплектом операндов — младшим словом. Помимо инструкций с новым блоком XMM в расширение SSE входят и дополнительные целочисленные инструкции с регистрами MMX, а также

инструкции управления кэшированием. В процессоре Pentium 4 набор инструкций получил очередные расширения **SSE2**, а позже и **SSE3**, касающиеся добавления новых типов 128-битных операндов для блока XMM. Использование блока XMM позволяет смешивать выполнение в режиме SIMD инструкций с целочисленными данными и с плавающей точкой. В процессорах Pentium III появилось расширение SSE, а в Pentium 4 расширение SSE2, предназначенное для 3D-графики, кодирования, декодирования видео, а также шифрования данных. Блок XMM и расширение SSE используются и в современных процессорах AMD.

### **Конвейерные суперскалярные процессоры**

Большинство современных микропроцессоров относятся к классу конвейерных суперскалярных процессоров с внеочередным исполнением операций.

Поясним основные понятия, относящиеся к конвейеризации и распараллеливанию выполнения инструкций (точнее, микроопераций).

**Конвейеризация** (pipelining) предполагает, что каждая инструкция обрабатывается за несколько этапов, причем каждый этап выполняется на своей ступени конвейера процессора. При выполнении инструкция продвигается по конвейеру по мере освобождения последующих ступеней. Таким образом, на конвейере одновременно может обрабатываться несколько последовательных инструкций.

В современных процессорах параллельно могут работать несколько конвейеров, так что производительность процессора можно оценивать темпом выхода выполненных инструкций со всех его конвейеров. Для достижения максимальной производительности процессора — обеспечения полной загрузки конвейеров — программа должна составляться с учетом микроархитектурных особенностей процессора. Конечно, и код, сгенерированный обычным способом, будет исполняться на более новых процессорах достаточно быстро (за счет более высокой тактовой частоты). Однако ряд программ на процессорах Pentium III работает быстрее, чем на Pentium 4, при одинаковых тактовых частотах: сверхдлинный конвейер Pentium 4 на «поворотах» непредсказуемо ветвящихся программ «заносит», что приводит к его простоям. Конвейер «классического» процессора Pentium имеет пять ступеней. Конвейеры процессоров P6 с **суперконвейерной архитектурой** (superpipelined) имеют большее число ступеней (10–12), что позволяет упростить каждую из них и, следовательно, сократить вре-

мя пребывания в них инструкций. *Гиперконвейер* Pentium 4 имеет уже 20 ступеней, а если считать его полную длину (начиная с декодирования), то наберется около 30 ступеней.

**Скалярным** называют процессор с единственным конвейером, к этому типу относятся все процессоры Intel до класса 486 включительно. **Суперскалярный** (superscalar) процессор имеет более одного конвейера; эти конвейеры способны обрабатывать инструкции параллельно. Суперскалярность наряду с тактовой частотой является важнейшим показателем пропускной способности процессора. Уровень суперскалярности («ширина обработки», гарантированно обеспеченная на всех этапах) в процессорах варьируется от 3 до 4–5.

Инструкции переходов и особенно ветвлений нарушают непрерывность работы начальных ступеней конвейера, поскольку они должны начинать выборку и декодирование инструкций с нового, заранее неизвестного адреса. **Предсказание переходов** (branch prediction) позволяет продолжать выборку и декодирование потока инструкций после выборки инструкции ветвления (условного перехода), не дожидаясь проверки самого условия. В процессорах прежних поколений инструкция перехода приостанавливала конвейер (выборку инструкций) до исполнения собственно перехода, на чем, естественно, терялась производительность. Предсказание переходов направляет поток выборки и декодирования по одной из ветвей, при этом используется ряд методов предсказания:

- при **статическом предсказании** (схема заложена в процессор) переходы по одним условиям, вероятнее всего, произойдут, а по другим — нет. Переходы назад скорее произойдут (это типичный цикл), вперед — нет (типично для обработки ошибок);

- **динамическое предсказание** опирается на предысторию вычислительного процесса. Для каждой конкретной команды перехода (ее адреса в памяти) накапливается статистика поведения, на основе которой предсказывается переход. Для динамического предсказания в процессор вводят таблицу *BTB* (Branch Table Buffer — буфер таблицы переходов), напоминающую кэш с ассоциативным поиском;

- **программные «намеки»** (hints) — новые префиксы инструкций, перекрывающие статическое предсказание. Намеки закладываются в программный код на этапе компиляции.

Внеочередное исполнение операций означает, что операции не обязаны выполняться в функциональных устройствах строго в том порядке, который определен в программном коде. Более поздние (по

коду) операции могут выполняться перед более ранними, если не зависят от порождаемых ими результатов. Процессор должен лишь гарантировать, чтобы результаты «внеочередного» выполнения программы совпадали с результатами «правильного» последовательного выполнения.

Механизм внеочередного исполнения позволяет в значительной степени сгладить эффект от ожидания считывания данных из кэшей верхних уровней и из оперативной памяти, что может занимать десятки и сотни тактов.

Отличия работы процессоров x86 от IA-64 приведены на рис.3.1



Рис.3.1. Процедура выполнения программы.

### 3.2 Многоядерные технологии INTEL

В 1965 г. соучредитель корпорации Intel Гордон Мур обратил внимание, что количество транзисторов на интегральной микросхеме растет экспоненциально, и предположил, что эта тенденция сохранится. Это предсказание сегодня известно как Закон Мура. Благодаря непрекращающимся научным исследованиям и конструкторским разработкам Intel удвоение количества транзисторов каждые несколько лет продолжается уже на протяжении 40 лет.

Большую часть этого времени закон Мура позволял относительно просто повышать производительность. Большинство среднестатистических пользователей просто применяли вертикальное масштаби-

рование, когда производительность приложений повышается за счет роста мощности и скорости работы одного процессора. Пользователи, которым нужно было выполнять более ресурсоемкие задачи, применяли горизонтальное масштабирование, когда производительность приложений повышается за счет распределения вычислительных задач между двумя или более процессорами. Однако в настоящее время вертикальное масштабирование стало достаточно трудной задачей, т. к. уже действуют ограничения, определяемые законами физики.

Во-первых, скорость доступа к памяти растет не так быстро, как скорость работы вычислительных устройств. Эти потерянные тактовые импульсы могут свести на нет преимущества от повышения частоты процессора.

Во-вторых, чем меньше становятся транзисторы и выше плотность их размещения, тем больше становится общая длина проводников, которыми они соединены между собой. Длина этих проводников в одном процессоре выросла с сотен до тысяч метров. Это ведет к появлению задержек распространения сигналов, которые могут свести на нет преимущества от повышения скорости работы транзисторов.

И, наконец, может быть самое главное препятствие – неконтролируемый рост плотности рассеиваемой мощности. За последние несколько лет количество транзисторов на одном кристалле выросло во много раз, и возникли технические проблемы, связанные с тем, что все транзисторы являются электрическими устройствами, которые потребляют электроэнергию и выделяют тепло.

В 1993 г., когда наступила эра процессоров Intel Pentium, они содержали около 3 миллионов транзисторов, а современный процессор содержит почти миллиард транзисторов. Если эта тенденция будет сохраняться, процессоры Intel скоро будут выделять больше тепла на квадратный сантиметр, чем поверхность Солнца – именно поэтому проблема отвода тепла уже сейчас устанавливает строгие ограничения на повышение тактовой частоты.

Однако непреложные законы физики необязательно ведут к жестким ограничениям для пользователей компьютеров. Новые архитектуры кристаллов, построенные по принципу горизонтального, а не вертикального масштабирования, позволят повысить производительность, сократить энергопотребление и одновременно более эффективно выполнять несколько задач.

В двухъядерных и многоядерных процессорах Intel используется именно такой подход к повышению производительности – горизон-



тальное масштабирование. Эта архитектура по существу отражает стратегию «разделяй и властвуй». Вычислительная работа, которая в традиционном процессоре выполняется одним ядром, в многоядерном процессоре Pentium распределяется между несколькими вычислительными ядрами, поэтому такой процессор может выполнить больше работы за заданный период тактовых импульсов.

Корпорация Intel впервые представила двухъядерную технологию в мае 2005 г., выпустив двухъядерные процессоры первого поколения — процессоры Intel Pentium Extreme Edition и Intel Pentium D. В начале января 2006 г. на рынок вышел двухъядерный процессор Intel Core Duo, предназначенный для мобильных ПК и созданный на базе новой микроархитектуры Intel Core, а в июле того же года — двухъядерный процессор второго поколения Intel Core 2 Duo на базе микроархитектуры Intel Core.

Наконец, в ноябре 2006 г. корпорация Intel представила первый в индустрии ПК четырехъядерный процессор Intel Core 2 Extreme, а в 2007 года выпустила на рынок четырехъядерный процессор для массового сегмента рынка Intel Core 2 Quad.

Многоядерность стала возможна благодаря существенному уменьшению размеров электронных элементов, интегрируемых на кристалле микропроцессора. Корпорация Intel уже смогла добиться уменьшения проектной нормы технологии производства транзисторов до 32 нанометров. Сегодня уже нельзя надеяться на большой скачок в развитии одноядерных процессоров, которые уже сейчас выпускаются на пределе производственных возможностей. В течение еще нескольких лет эти процессорные ядра будут оснащаться кэш-памятью большего размера, чем современные, немного повысится их тактовая частота. Но из-за упомянутых физических и технологических ограничений осуществлять эти усовершенствования будет все труднее, и они не будут успевать за ростом требований к повышению производительности. Для выполнения этих запросов неминуемо потребуются горизонтальное масштабирование — подход, при котором вычислительные задачи распределяются между множеством вычислительных ядер.

Но добавление даже одного дополнительного процессора не всегда реально для большинства пользователей ПК. Из-за необходимости в дополнительном разъеме, специальных микросхемах и электрических цепях двухпроцессорные системные платы гораздо дороже своих однопроцессорных аналогов. К тому же в большинстве ПК про-

цессор часто является самым дорогим компонентом, поэтому попытка достичь повышения производительности путем увеличения количества процессоров существенно удорожит систему.

Другой путь горизонтального масштабирования, который избрала корпорация Intel, начав разработку многоядерных архитектур — добавить в каждый процессор дополнительные «мозги». Проще говоря, проектировщики интегральных микросхем помещают в один процессор два или больше «исполнительных ядер», или вычислительных устройств. Такой многоядерный процессор занимает один процессорный разъем, но операционная система воспринимает каждое ядро как отдельный процессор со всеми необходимыми вычислительными ресурсами.

Многоядерные процессоры могут выполнить больше работы за заданный период тактовых импульсов, поэтому они могут работать на меньших тактовых частотах, чем их одноядерные собратья. Из-за того что энергопотребление растет пропорционально частоте, многоядерная архитектура позволяет инженерам решить проблему постоянно растущих требований к электропитанию и охлаждению.

Приложения, которые работают на одноядерном процессоре Intel, будут работать и на двухъядерном процессоре Intel. Однако для того чтобы сегодня получить все преимущества от выполнения приложения на таком процессоре, ПО, работающее на такой платформе, должно быть разработано так, чтобы вычислительная нагрузка распределялась между множеством вычислительных ядер. Такая функциональная возможность называется параллелизм на уровне потоков или «многопоточность», а приложения и ОС (такие как Microsoft Windows\* XP, Windows Server и версии Linux\*), которые поддерживают этот режим работы, называются «поточными» или «многопоточными».

Из-за того что энергопотребление растет пропорционально частоте, многоядерная архитектура позволяет инженерам решить проблему постоянно растущих требований к электропитанию и охлаждению. В результате существенно повышается общая продуктивность работы пользователей и появляется возможность расширить действие закона Мура на будущее.

И так, несколько лет назад компания Intel отказалась от «гонки мегагерцев» и взяла курс на разработку эффективных процессорных микроархитектур с экономным энергопотреблением.

Максимальная эффективность работы процессора напрямую зависит не столько от тактовой частоты, сколько от количества инструкций, выполняемых за один такт. Иными словами, тактовая частота процессора — лишь один из множителей в простой формуле:

[Производительность] = [Тактовая частота] x [Количество инструкций, выполняемых за один такт]

Другим весьма эффективным способом повышения одного из множителей выше приведенной формулы расчета производительности можно назвать метод снижения количества инструкций, необходимых для исполнения той или иной задачи, иными словами, оптимизация потока команд. Яркий пример такой технологии — так называемая технология микрослияния команд (microfusion), в результате чего несколько внутренних микроопераций (micro-ops) процессора могут быть скомбинированы в одну микрооперацию, чем также значительно сокращается общее количество микроопераций для выполнения конкретной задачи.

Рассмотрим многоядерную технологию на примере архитектуры Intel Core (рис. 3.2).

Новые чипы на базе архитектуры Intel Core обещают значительный прирост производительности — от 40% для Conroe до 80% для Woodsrest, при одновременном снижении энергопотребления на 35–40%.

Новая процессорная архитектура наследует философию эффективного энергопотребления. Ключевая роль отводится инновациям, впервые реализованным в новом поколении процессорной архитектуры Intel:

- Технология **Intel Wide Dynamic Execution** призвана обеспечить выполнение большего количества команд за каждый такт, повышая эффективность выполнения приложений и сокращая энергопотребление. Каждое ядро процессора, поддерживающего эту технологию, теперь может выполнять до четырех инструкций одновременно с помощью 14-стадийного конвейера.

- Технология **Intel Wide Dynamic Execution** призвана обеспечить выполнение большего количества команд за каждый такт, повышая эффективность выполнения приложений и сокращая энергопотребление. Каждое ядро процессора, поддерживающего эту технологию, теперь может выполнять до четырех инструкций одновременно с помощью 14-стадийного конвейера.

# Conroe Processor Block Diagram

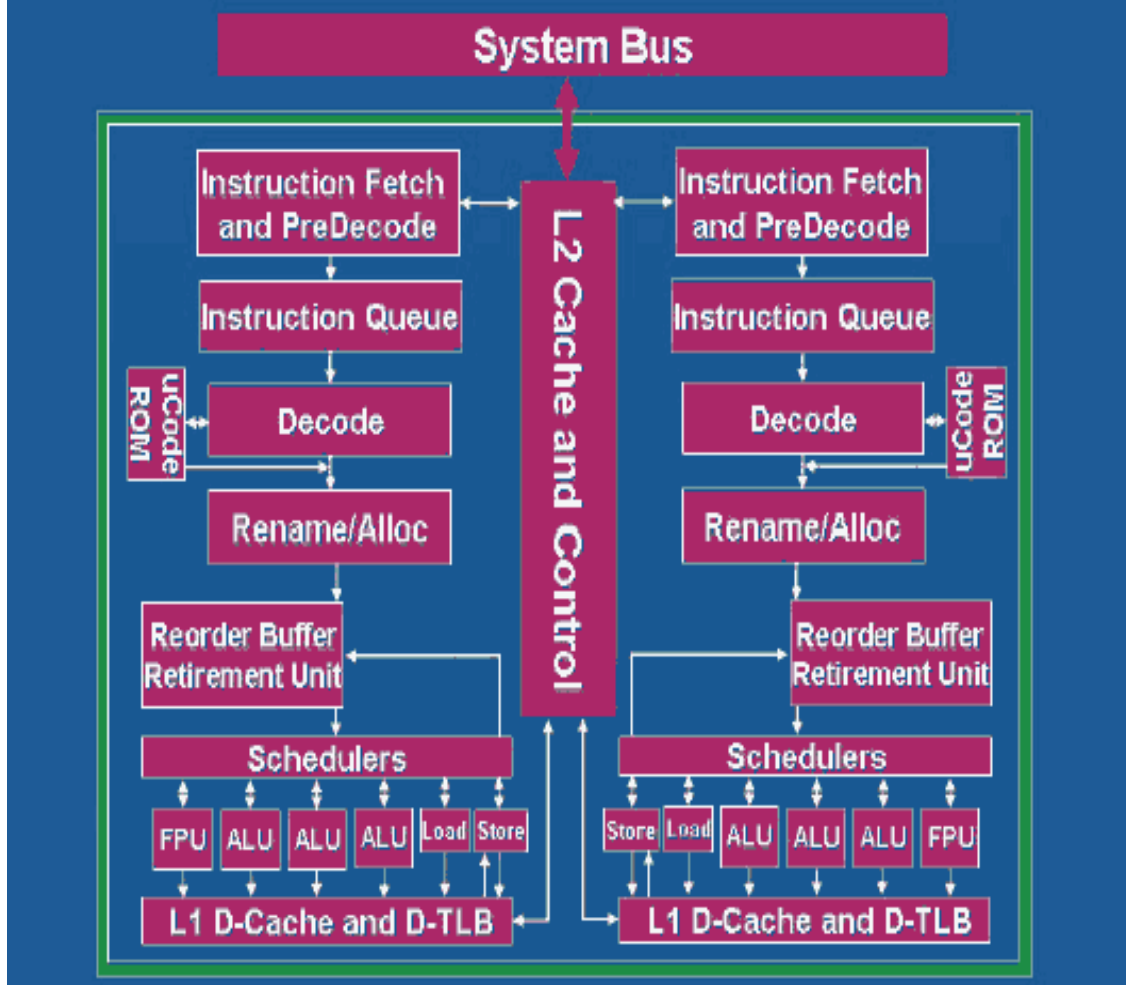


Рис. 3.2. Основные особенности архитектуры Intel Core

– Технология **Intel Intelligent Power Capability**, активируя отдельные узлы чипа только по мере необходимости, значительно снижает энергопотребление системы в целом.

– Технология **Intel Advanced Smart Cache** подразумевает наличие общей для всех ядер кэш-памяти L2, совместное использование которой снижает энергопотребление и повышает производительность. При этом, по мере необходимости, одно из ядер процессора может использовать весь объём кэш-памяти при динамическом отключении другого ядра.

– Технология **Intel Smart Memory Access** повышает производительность системы, сокращая время отклика памяти и оптимизируя, таким образом, использование пропускной способности подсистемы памяти.

– Технология **Intel Advanced Digital Media Boost** позволяет обрабатывать все 128-разрядные команды SSE, SSE2 и SSE3, широко используемые в мультимедийных и графических приложениях, за один такт, что увеличивает скорость их выполнения.

## 4. ОПЕРАТИВНОЕ ЗАПОМИНАЮЩЕЕ УСТРОЙСТВО

### 4.1. Основные понятия, типы оперативной памяти.

**Оперативная память** — это рабочая область для процессора компьютера. В ней во время работы хранятся программы и данные. Оперативная память часто рассматривается как временное хранилище, потому что данные и программы в ней сохраняются только при включенном компьютере или до нажатия кнопки сброса (reset). Перед выключением или нажатием кнопки сброса все данные, подвергнутые изменениям во время работы, необходимо сохранить на запоминающем устройстве, которое может хранить информацию постоянно (обычно это жесткий диск). При новом включении питания сохраненная информация вновь может быть загружена в память.

Устройства оперативной памяти иногда называют запоминающими устройствами с произвольным доступом. Это означает, что обращение к данным, хранящимся в оперативной памяти, не зависит от порядка их расположения в ней. Когда говорят о памяти компьютера, обычно подразумевают оперативную память, прежде всего, микросхемы памяти или модули, в которых хранятся активные программы и данные, используемые процессором.

Термин оперативная память часто обозначает не только микросхемы, которые составляют устройства памяти в системе, но включает и такие понятия, как логическое отображение и размещение. *Логическое отображение* — это способ представления адресов памяти на фактически установленных микросхемах. *Размещение* — это расположение информации (данных и команд) определенного типа по конкретным адресам памяти системы. Поскольку невозможно постоянно хранить файлы в оперативной памяти, все измененные после загрузки в память файлы должны быть вновь сохранены на жестком диске перед выключением компьютера. Если измененный файл не будет сохранен, то первоначальная копия файла на жестком диске останется неизменной.

Во время выполнения программы в оперативной памяти хранятся ее данные. Микросхемы оперативной памяти (RAM) иногда называют энергозависимой памятью: после выключения компьютера данные, хранимые в них, будут потеряны, если они предварительно не

были сохранены на диске или другом устройстве внешней памяти. Чтобы избежать этого, некоторые приложения автоматически делают резервные копии данных (например, Microsoft Word). Файлы компьютерной программы при ее запуске загружаются в оперативную память, в которой хранятся во время работы с указанной программой. Процессор выполняет программно-реализованные команды, содержащиеся в памяти, и сохраняет их результаты. Оперативная память хранит коды нажатых клавиш при работе с текстовым редактором, а также величины математических операций. При выполнении команды *Сохранить* (*Save*) содержимое оперативной памяти сохраняется в виде файла на жестком диске.

Физически оперативная память в системе представляет собой набор микросхем или модулей, содержащих микросхемы, которые подключаются к системной плате. Эти микросхемы или модули могут иметь различные характеристики и, чтобы функционировать правильно, должны быть совместимы с системой, в которую устанавливаются.

### **Типы запоминающих устройств памяти**

В современных компьютерах используются запоминающие устройства трех основных типов:

**ROM** (Read Only Memory) – постоянное запоминающее устройство (ПЗУ), не способное выполнять операцию записи данных.

**DRAM** (Dynamic Random Access Memory) – динамическое запоминающее устройство с произвольным порядком выборки.

**SRAM** (Static RAM) – статическая оперативная память.

**SDRAM** – синхронная динамическая память

## **4.2. Память типа ROM**

В памяти типа ROM (Read Only Memory), или ПЗУ (постоянное запоминающее устройство), данные можно только хранить, изменять их нельзя. Именно поэтому такая память используется только для чтения данных. ROM также часто называется энергонезависимой памятью, потому что любые данные, записанные в нее, сохраняются при выключении питания. Поэтому в ROM помещаются команды запуска персонального компьютера, т. е. программное обеспечение, которое загружает систему. В современных компьютерах типовое значение ПЗУ от 128 Кбайт до 2 Мбайт.

Заметьте, что ROM и оперативная память — не противоположные понятия. На самом деле ROM представляет собой часть оперативной памяти системы. Другими словами, часть адресного пространства оперативной памяти отводится для ROM. Это необходимо для хранения программного обеспечения, которое позволяет загрузить операционную систему.

Основной код BIOS содержится в микросхеме ROM на системной плате, но на платах адаптеров также имеются аналогичные микросхемы. Они содержат вспомогательные подпрограммы базовой системы ввода-вывода и драйверы, необходимые для конкретной платы, особенно для тех плат, которые должны быть активизированы на раннем этапе начальной загрузки, например видеоадаптер. Платы, не нуждающиеся в драйверах на раннем этапе начальной загрузки, обычно не имеют ROM, потому что их драйверы могут быть загружены с жесткого диска позже — в процессе начальной загрузки.

В настоящее время в большинстве систем используется одна из форм Flash-памяти, которая называется электронно-перепрограммируемой постоянной памятью (Electrically Erasable Programmable Read-only Memory — EEPROM). Flash-память является по настоящему энергонезависимой и перезаписываемой, она позволяет пользователям легко модифицировать ROM, программно-аппаратные средства системных плат и других компонентов (таких, как видеоадаптеры, платы SCSI, периферийные устройства и т. п.).

### **4.3. Память типа DRAM**

Динамическая оперативная память (Dynamic RAM — DRAM) используется в большинстве систем оперативной памяти современных персональных компьютеров. Основное преимущество памяти этого типа состоит в том, что ее ячейки упакованы очень плотно, т. е. в небольшую микросхему можно упаковать много битов, а значит, на их основе можно построить память большой емкости.

Ячейки памяти в микросхеме DRAM — это крошечные конденсаторы, которые удерживают заряды. Именно так (наличием или отсутствием зарядов) и кодируются биты. Проблемы, связанные с памятью этого типа, вызваны тем, что она динамическая, т. е. должна постоянно регенерироваться, так как в противном случае электрические заряды в конденсаторах памяти будут «стекают» и данные будут потеряны. Регенерация происходит, когда контроллер памяти системы бе-



рет крошечный перерыв и обращается ко всем строкам данных в микросхемах памяти. Большинство систем имеют контроллер памяти (обычно встраиваемый в набор микросхем системной платы), который настроен на соответствующую промышленным стандартам частоту регенерации, равную 15 мкс. Ко всем строкам данных обращение осуществляется по прохождении 128 специальных циклов регенерации. Это означает, что каждые 1,92 мс ( $128 \cdot 15$  мкс) прочитываются все строки в памяти для обеспечения регенерации данных.

Регенерация памяти, к сожалению, отнимает время у процессора: каждый цикл регенерации по длительности занимает несколько циклов центрального процессора. В старых компьютерах циклы регенерации могли занимать до 10% (или больше) процессорного времени, но в современных системах, работающих на частотах, равных сотням мегагерц, расходы на регенерацию составляют 1% (или меньше) процессорного времени. Некоторые системы позволяют изменить параметры регенерации с помощью программы установки параметров CMOS, но увеличение времени между циклами регенерации может привести к тому, что в некоторых ячейках памяти заряд «стечет», а это вызовет сбой памяти. В большинстве случаев надежнее придерживаться рекомендуемой или заданной по умолчанию частоты регенерации. Поскольку затраты на регенерацию в современных компьютерах составляют менее 1%, изменение частоты регенерации оказывает незначительное влияние на характеристики компьютера.

В устройствах DRAM для хранения одного бита используется только один транзистор и пара конденсаторов, поэтому они более вместительны, чем микросхемы других типов памяти. В настоящее время имеются микросхемы динамической оперативной памяти емкостью 512 Мбайт и больше. Это означает, что подобные микросхемы содержат более 256 млн транзисторов! А ведь в Pentium 4 только 42 млн транзисторов. Откуда такая разница? Дело в том, что в микросхеме памяти все транзисторы и конденсаторы размещаются последовательно, обычно в узлах квадратной решетки, в виде очень простых, периодически повторяющихся структур, в отличие от процессора, представляющего собой более сложную схему различных структур, не имеющую четкой организации.

Транзистор для каждого одноразрядного регистра DRAM используется для чтения состояния смежного конденсатора. Если конденсатор заряжен, в ячейке записана 1; если заряда нет — записан 0. Заряды в крошечных конденсаторах все время «стекают», вот почему

память должна постоянно регенерироваться. Даже мгновенное прерывание подачи питания или какой-нибудь сбой в циклах регенерации приведет к потере заряда в ячейке DRAM, а следовательно, и к потере данных. В работающей системе подобное приводит к появлению «синего» экрана, глобальным отказам системы защиты, повреждению файлов или к полному отказу системы.

Динамическая оперативная память используется в персональных компьютерах, поскольку она недорогая, микросхемы могут быть плотно упакованы, а это означает, что запоминающее устройство большой емкости может занимать небольшое пространство. К сожалению, память этого типа не отличается высоким быстродействием, обычно она намного «медленнее» процессора. Поэтому существует множество различных типов организации DRAM, позволяющих улучшить эту характеристику.

#### **4.4. Память типа SDRAM**

Это тип динамической оперативной памяти DRAM, работа которой синхронизируется с шиной памяти. SDRAM передает информацию в высокоскоростных пакетах, использующих высокоскоростной синхронизированный интерфейс. SDRAM позволяет избежать использования большинства циклов ожидания, необходимых при работе асинхронной DRAM, поскольку сигналы, по которым работает память такого типа, синхронизированы с тактовым генератором системной платы.

Этот тип памяти стал самым популярным в новых системах, выпущенных в 2000 и 2001 годах. Эффективность SDRAM значительно выше по сравнению с оперативной памятью FPM или EDO. Поскольку SDRAM — это тип динамической оперативной памяти, ее начальное время ожидания такое же, как у памяти FPM или EDO, но общее время цикла намного короче. Схема синхронизации пакетного доступа SDRAM выглядит так: 5–1–1–1, т. е. четыре операции чтения завершаются всего лишь за восемь циклов системной шины (сравните с 11 циклами для EDO и 14 для FPM).

Кроме этого, память SDRAM может работать на частоте 133 МГц (7,5 нс) и выше, что стало новым стандартом для системного быстродействия начиная с 1998 г. Память SDRAM поставляется в виде модулей DIMM и, как правило, ее быстродействие оценивается в мегагерцах, а не в наносекундах

## 4.5. Память типа DDR SDRAM

Память DDR (Double Data Rate) – двойная скорость передачи данных) – это еще более усовершенствованный стандарт SDRAM, при использовании которого скорость передачи данных удваивается. Это достигается не за счет удвоения тактовой частоты, а за счет передачи данных дважды за один цикл: первый раз в начале цикла, а второй — в конце. Именно благодаря этому и удваивается скорость передачи (причем используются те же самые частоты и синхронизирующие сигналы).

В основном память DDR SDRAM используется в системах, оснащенных процессорами AMD и Cyrix, она выпускается в виде 184-контактных модулей DIMM (см. рис. 4.1).

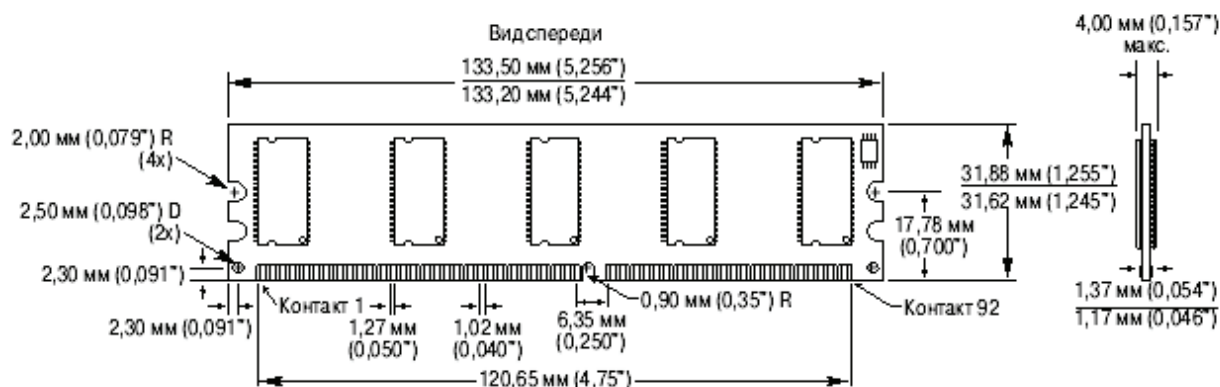


Рис. 4.1 184-контактный модуль DIMM

Поставляемые модули DIMM памяти DDR SDRAM отличаются своим быстродействием, пропускной способностью и обычно работают при напряжении 2,5 В. Они представляют собой, в сущности, расширение стандарта SDRAM DIMM, предназначенное для поддержки удвоенной синхронизации, при которой передача данных, в отличие от стандарта SDRAM, происходит при каждом тактовом переходе, т. е. дважды за каждый цикл. Для того чтобы избежать путаницы, обычную память SDRAM часто называют памятью с одинарной скоростью передачи данных (Single Data Rate — SDR).

Как работает DDR? Число, следующее за «DDR», указывает на скорость передачи данных. Например, у DDR 400 скорость передачи

400 МГц. При этом использовать термин «МГц» некорректно. Правильно указывать скорость в «миллионах передач в секунду через один вывод данных». Память DDR 400 работает на частоте 200 МГц или на частоте в 2 раза меньше скорости передачи данных (вернее, скорость передачи данных в 2 раза больше тактовой частоты). Все управляющие сигналы синхронизируются частотой 200 МГц. Внутри чипа все работает классически по переднему фронту сигналов тактового генератора с частотой 200 МГц. Официальная частота DDR333 равна 167,0 МГц.

Чтобы обеспечить передачу данных дважды за такт, используется специальная архитектура «2n Prefetch». Внутренняя шина данных имеет ширину в два раза больше внешней. При передаче данных сначала передается первая половина шины данных по переднему фронту тактового сигнала, а затем вторая половина шины данных по заднему фронту (рис.4.2).

Для возможности работы на высоких частотах вместо одного тактового сигнала используется два (Differential Clock). Дополнительный тактовый сигнал инвертирован относительно основного. Поэтому на самом деле синхронизация происходит не по заднему фронту. В документации написано, что синхронизация происходит при пересечении этих двух тактовых сигналов. Однако, вместо пересечения просто используется передний фронт дополнительного тактового сигнала.

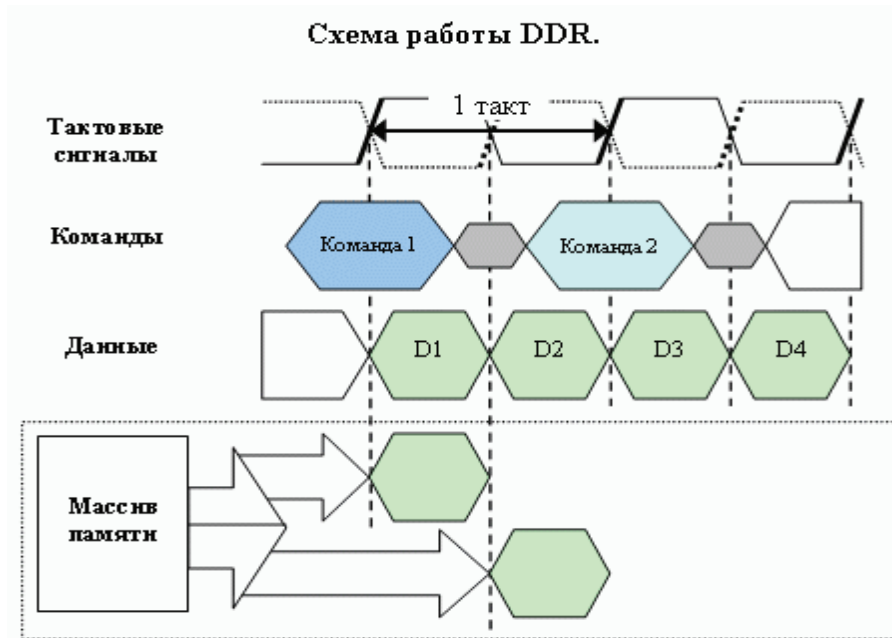


Рис. 4.2. Схема работы DDR

Кроме передачи двух данных за такт, DDR SDRAM имеет несколько других принципиальных отличий от простой памяти SDRAM. В основном они являются технологическими. Например, был добавлен сигнал QDS, который располагается на печатной плате вместе с линиями данных. По нему происходит синхронизация при передаче данных. Если используется два модуля памяти, то данные от них приходят к контроллеру памяти с небольшой разницей из-за разного расстояния. Возникает проблема в выборе синхросигнала для их считывания. Использование QDS успешно это решает.

#### **4.6. Память стандарта DDR2. Сравнительная характеристика с DDR**

DDR2 — новый стандарт памяти, утвержденный JEDEC (Joint Electronic Device Engineering Council), в состав которого входят многие производители модулей памяти и чипсетов. Ранние версии стандарта были опубликованы в марте 2003 г., окончательно он был утвержден лишь в январе 2004 г. и получил наименование DDR2 SDRAM SPECIFICATION, JESD79-2, ревизия А (JESD79-2A). DDR2 основывается на хорошо известной и зарекомендовавшей себя технологии DDR (Double Data Rate). Можно даже сказать так: «DDR2 начинается там, где заканчивается DDR». Другими словами, DDR2 работает на частотах, являющихся пределом для памяти DDR-400 (стандарт PC3200, тактовая частота 200 МГц. Первым поколением памяти DDR2 являются ее разновидности DDR2-400 и DDR2-533, работающие на частотах 200 МГц и 266 МГц, соответственно. Также появились модули нового поколения DDR2-667 и DDR2-800 (рис.4.3).

##### **DDR2 изнутри**

Стандарт DDR2 включает в себя несколько важных изменений спецификации DDR, связанных с передачей данных, которые позволяют достигать более высоких частот при более низкой потребляемой мощности.



Рис. 4.3. Модуль памяти DDR2

### Выборка данных

Главным изменением в DDR2 является возможность выборки сразу 4 бит данных за такт (4n-prefetch), в противоположность 2-битной выборке (2n - prefetch), реализованной в DDR. Это значит, что на каждом такте шины памяти DDR2 осуществляет пересылку 4 бит информации из логических (внутренних) банков микросхемы памяти в буферы ввода-вывода по одной линии интерфейса данных, тогда как обычная DDR способна переслать лишь 2 бита за такт на линию. Довольно закономерно возникает вопрос — если это так, то почему же тогда эффективная пропускная способность DDR2-400 оказывается такой же, как и у обычной DDR-400 (3.2 ГБ/с), а не удвоенной?

Для ответа на этот вопрос сначала вспомним, как работает обычная память типа DDR-400. В этом случае, как ядро памяти, так и буферы ввода-вывода функционируют на частоте 200 МГц, а «эффективная» частота внешней шины данных, благодаря технологии DDR, равна 400 МГц. По правилу 2n-prefetch, на каждом такте памяти (200 МГц) по каждой линии интерфейса данных в буфер ввода-вывода поступает 2 бита информации. Задачей этого буфера является мультиплексирование/демультиплексирование (MUX/DEMUX) потока дан-

ных — по-простому, «перегонка» узкого высокоскоростного потока в широкий низкоскоростной, и наоборот. Поскольку в микросхеме памяти типа DDR SDRAM логические банки имеют ширину шины данных, соединяющую их и усилитель уровня, в два раза шире, чем от защелок на чтение до внешнего интерфейса, буфер данных включает в себя мультиплексор типа 2-1. В общем случае, поскольку микросхемы памяти, в отличие от модулей, могут иметь разную ширину шины данных — обычно это  $x4/x8/x16/x32$ , применение такой схемы MUX/DEMUX (2-1), реализованной в DDR, означает, что внутренний поток данных шириной  $X$  и частотой передачи  $Y$  от массива преобразуется во внешний поток шириной  $X/2$  и частотой  $2Y$ . Это называется балансом по пиковой пропускной способности (рис. 4.4).

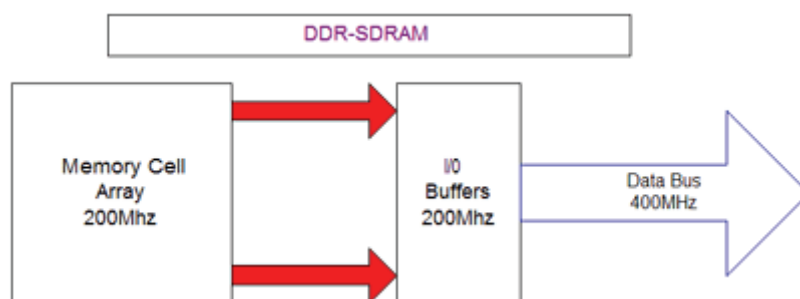


Рис. 4.4. Схема функционирования устройства микросхемы памяти типа DDR – SDRAM

Рассмотрим теперь схему функционирования устройства микросхемы памяти типа DDR2 SDRAM, равночастотной и «равноширокой» (т. е. одинаковой ширины шины данных) относительно микросхемы DDR модуля памяти DDR2-400 (рис. 4.5). Прежде всего, отметим, что ширина внешней шины данных осталась абсолютно такой же — 1 бит/линию, как и ее эффективная частота (в рассматриваемом примере — 400 МГц). Очевидно, что использование мультиплексора типа 2-1, применяемого в DDR SDRAM, в случае DDR2 SDRAM, осуществляющей выборку данных по правилу  $4n$ -prefetch, уже не подходит. Вместо этого требуется введение более сложной схемы с дополнительной ступенью преобразования — мультиплексора типа 4-1. Это означает, что выход ядра стал шире в четыре раза внешнего интерфейса микросхемы и во столько же раз ниже по частоте функционирования. То есть, по аналогии с рассмотренным выше примером, в

общем случае схема MUX/DEMUX 4-1 осуществляет преобразование внутреннего потока данных шириной  $X$  и частотой передачи  $Y$  от массива во внешний поток шириной  $X/4$  и частотой  $4Y$ .

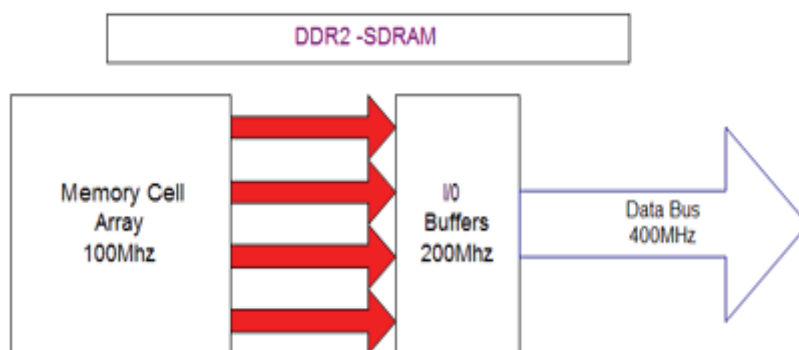


Рис. 4.5. Схема функционирования устройства микросхемы памяти типа DDR2 – SDRAM

Поскольку в этом случае ядро микросхем памяти синхронизируется на частоте, вдвое меньшей по отношению к внешней (100 МГц), тогда как в DDR синхронизация внутреннего и внешнего потока данных происходят на одной частоте (200 МГц), среди преимуществ такого подхода отмечается увеличение процента выхода годных чипов и снижение энергопотребления модулей. Кстати, это также позволяет объяснить, почему стандарт DDR2 предполагает существование модулей памяти с «эффективной» частотой 800 МГц — что вдвое выше, чем у текущего поколения памяти типа DDR. Ведь именно такой «эффективной» частоты DDR2 можно достичь уже сейчас, имея чипы памяти DDR-400, работающие на собственной частоте 200 МГц, если осуществлять выборку данных по правилу 4n-prefetch по схеме, рассмотренной выше.

Таким образом, DDR2 означает отказ от экстенсивного пути развития чипов памяти, простого дальнейшего увеличения их частоты, которое существенно затрудняет производство стабильно работающих модулей памяти в большом количестве. На смену ему выдвигается интенсивный путь развития, связанный с расширением внутренней шины данных (что является обязательным и неизбежным решением при использовании более сложного мультиплексирования).



### **Внутричиповое терминирование**

Стандарт DDR2 включает в себя и ряд других усовершенствований, улучшающих различные характеристики нового типа памяти, в том числе — электрические. Одним из таких новшеств является внутричиповое терминирование сигнала. Суть его заключается в том, что для устранения излишнего электрического шума (вследствие возможного отражения сигнала от конца линии) на шине памяти для нагрузки линии используются резисторы не на материнской плате (как это было с предыдущими поколениями памяти), а внутри самих чипов. Эти резисторы деактивируются, когда чип находится в работе и, наоборот, активируются, как только микросхема входит в состояние ожидания. Поскольку гашение сигнала теперь осуществляется намного ближе к его источнику, это позволяет устранить электрические помехи внутри чипа памяти при передаче данных.

## **4.7. Модули памяти**

### **Модули DIMM**

Существует два типа модулей DIMM. Модули памяти DIMM обычно содержат стандартные микросхемы SDRAM или DDR-SDRAM и отличаются друг от друга физическими характеристиками. Стандартный модуль DIMM имеет 168 выводов, по одному радиусному пазу с каждой стороны и два паза в области контакта. Модули DDR DIMM, в свою очередь, имеют 184 вывода, по два паза с каждой стороны и только один паз в области контакта. Тракт данных модулей DIMM может быть равен 64 бит (без контроля по четности) или 72 бит (с контролем по четности или поддержкой кода коррекции ошибок ECC). На каждой стороне платы DIMM расположены различные выводы сигнала.

На рис. 4.12 показан 168-контактный модуль SDRAM DIMM. Выводы модулей памяти пронумерованы слева направо. Выводы модулей DIMM, расположенные на разных сторонах платы, различны. Обратите внимание, что размеры приведены в миллиметрах и в дюймах.

Именно поэтому они называются модулями памяти с двухрядным расположением выводов. Эти модули примерно на один дюйм (25 мм) длиннее модулей SIMM, но благодаря своим свойствам содержат гораздо больше выводов.

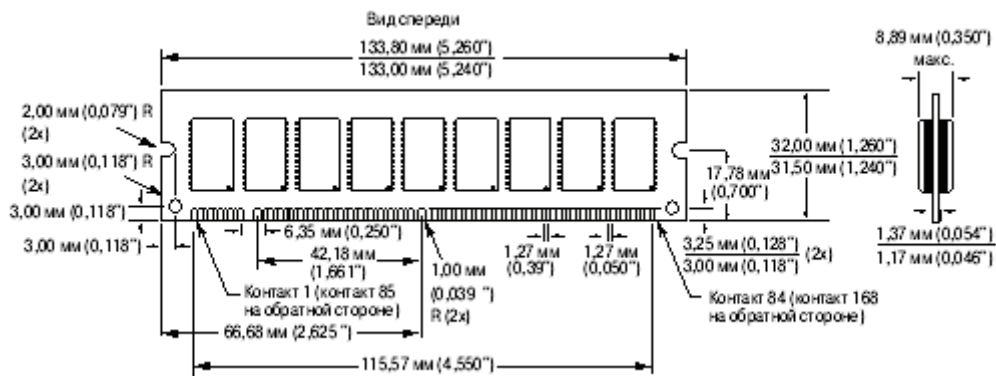


Рис. 4.12. Обычный 168-контактный (72-разрядный) модуль DIMM

### Модули RIMM

Сигнальные выводы, расположенные на разных сторонах платы RIMM, также различны. Существует три физических типа модулей RIMM: 16/18-разрядная версия со 184 выводами, 32/36-разрядная версия, имеющая 232 вывода, и 64/72-разрядная версия, содержащая 326 выводов. Размеры разъемов, используемых для установки модулей памяти, одинаковы, но расположение пазов в разъемах и платах RIMM различны, что позволяет избежать установки несоответствующих модулей. Данная плата поддерживает только один тип модулей памяти. В настоящее время наиболее распространенным типом является 64/72-разрядная версия; 32-разрядная версия модулей памяти была представлена в конце 2002 г., а 64-разрядная — 2004 г.

Стандартный 64/72-разрядный модуль RIMM имеет 326 выводов, по одному пазу с каждой стороны и два симметрично расположенных паза в области контакта.

### Конструкция и организация микросхем и модулей памяти

В системных платах используется несколько типов микросхем памяти. Большинство из них одноразрядные, но емкость их различна. Как правило, емкость модулей памяти кратна четырем, поскольку матрица, на основе которой создают микросхемы памяти, является квадратной. Четырехкратное увеличение емкости подразумевает увеличение количества транзисторов в четыре раза. Емкость современных модулей DIMM составляет 256 Мбайт – 4 Гбайт.

## 4.8. КЭШ-память — SRAM

Существует тип памяти, совершенно отличный от других, — *статическая оперативная память* (Static RAM — SRAM). Она названа так потому, что, в отличие от динамической оперативной памяти (DRAM), для сохранения ее содержимого не требуется периодической регенерации. Но это не единственное ее преимущество. SRAM имеет более высокое быстродействие, чем динамическая оперативная память, и может работать на той же частоте, что и современные процессоры.

Время доступа SRAM не более 2 нс; это означает, что такая память может работать синхронно с процессорами на частоте 500 МГц или выше. Однако для хранения каждого бита в конструкции SRAM используется кластер из шести транзисторов. Использование транзисторов без каких-либо конденсаторов означает, что нет необходимости в регенерации. Пока подается питание, SRAM будет помнить то, что сохранено. Почему же тогда микросхемы SRAM не используются для всей системной памяти? Ответ можно найти в таблице 4.1:

Таблица 4.1

Сравнительные характеристики типов оперативной памяти

Тип	Быстродействие	Плотность	Стоимость
DRAM	Низкое	Высокая	Низкая
SRAM	Высокое	Низкая	Высокая

По сравнению с динамической оперативной памятью быстродействие SRAM намного выше, но плотность ее гораздо ниже, а цена довольно высока. Более низкая плотность означает, что микросхемы SRAM имеют большие габариты, хотя их информационная емкость намного меньше. Большое число транзисторов и кластеризованное их размещение не только увеличивает габариты микросхем SRAM, но и значительно повышает стоимость технологического процесса по сравнению с аналогичными параметрами для микросхем DRAM. Таким образом, габариты SRAM в среднем в 30 раз превышают размер динамической оперативной памяти, то же самое можно сказать и о стоимости. Все это не позволяет использовать память типа SRAM в качестве оперативной памяти в персональных компьютерах.

Несмотря на это, разработчики все-таки применяют память типа SRAM для повышения эффективности ПК. Но во избежание значи-

тельного увеличения стоимости устанавливается только небольшой объем высокоскоростной памяти SRAM, которая используется в качестве **кэш-памяти**. Кэш-память работает на тактовых частотах, близких, или даже равных, тактовым частотам процессора, причем обычно именно эта память непосредственно используется процессором при чтении и записи. Во время операций чтения данные в высокоскоростную кэш-память предварительно записываются из оперативной памяти с низким быстродействием.

Эффективность кэш-памяти выражается *коэффициентом совпадения*, или *коэффициентом успеха*. Коэффициент совпадения равен отношению количества удачных обращений в кэш к общему количеству обращений. Попадание — это событие, состоящее в том, что необходимые процессору данные предварительно считываются в кэш из оперативной памяти, иначе говоря, в случае попадания процессор может считывать данные из кэш-памяти. Неудачным обращением в кэш считается такое, при котором контроллер кэша не предусмотрел потребности в данных, находящихся по указанному абсолютному адресу. В таком случае необходимые данные не были предварительно считаны в кэш-память, поэтому процессор должен отыскать их в более медленной оперативной памяти, а не в быстродействующем кэше.

Когда процессор считывает данные из оперативной памяти, ему приходится какое-то время «ждать», поскольку тактовая частота оперативной памяти значительно ниже, чем процессора. Если процессор со встроенной в кристалл кэш-памятью работает на частоте 2 ГГц, то продолжительность цикла процессора и интегральной кэш-памяти в этом случае достигнет 0,5 нс, в то время как продолжительность цикла оперативной памяти будет в шесть раз больше, т. е. примерно 3 или 6 нс для памяти с удвоенной скоростью передачи данных (Double Data Rate — DDR). Следовательно, в том случае, когда процессор с тактовой частотой 2 ГГц считывает данные из оперативной памяти, его рабочая частота уменьшается в шесть раз, что и составляет 333 МГц. Замедление обусловлено периодом ожидания (wait state). Если процессор находится в состоянии ожидания, то на протяжении всего цикла (такта) никакие операции не выполняются. Процессор «ждет», пока необходимые данные поступят из более медленной оперативной памяти. Поэтому именно кэш-память позволяет сократить количество «простоев» и повысить быстродействие компьютера в целом.

Чтобы минимизировать время ожидания при считывании процессором данных из медленной оперативной памяти, в современных

персональных компьютерах обычно предусмотрены два типа кэш-памяти: кэш-память первого уровня (L1) и кэш-память второго уровня (L2). Кэш-память первого уровня также называется встроенным, или внутренним, кэшем; он непосредственно встроен в процессор и фактически является частью микросхемы процессора. Во всех процессорах 486 и выше кэш-память первого уровня интегрирована в микросхему процессора. Кэш-память второго уровня называется вторичным, или внешним, кэшем; он устанавливается вне микросхемы процессора. Первоначально она устанавливалась на системной плате. (Так было во всех компьютерах на основе процессоров 386, 486 и Pentium.) Если кэш-память второго уровня установлена на системной плате, то она работает на ее частоте. В этом случае кэш-память второго уровня обычно находится рядом с разъемом процессора.

Для повышения эффективности в более поздних компьютерах на основе процессоров Pentium Pro, Pentium II/III и Athlon кэш-память второго уровня является частью процессора. Конечно же, он внешний по отношению к кристаллу центрального процессора, просто эта отдельная микросхема устанавливается внутри корпуса (картриджа) процессора. Поэтому на системных платах для процессоров Pentium Pro или Pentium II нет никакого кэша. В моделях процессоров Pentium III и Athlon кэш-память второго уровня является частью микросхемы процессора (подобно кэш-памяти первого уровня) и работает на более высоких частотах (на частоте процессора, половинной или трети). В современных процессах процессорах для увеличения производительности используется три уровня кэш-памяти.

Кэш-память работает синхронно с шиной процессора, что повышает ее быстродействие и эффективность. В настоящее время вопрос об ограничении объема кэшируемой памяти не стоит, так как этот параметр превышает максимальный объем оперативной памяти, которая может поддерживаться тем или другим набором микросхем.

#### **4.10. Быстродействие ОЗУ**

*Быстродействие памяти* определяется временем выполнения операций записи и считывания данных. Основными параметрами любых элементов памяти является минимальное время доступа и длительность цикла обращения. *Время доступа* (access time) определяется как задержка появления действительных данных на выходе памяти относительно начала цикла чтения, *длительность цикла* — как мини-

мальный период следующих друг за другом обращений к памяти, причем циклы чтения и записи могут требовать различных затрат времени. В цикл обращения, помимо активной фазы самого доступа, входит и фаза восстановления (возврата памяти к исходному состоянию), которая соизмерима по времени с активной фазой. Временные характеристики самих запоминающих элементов определяются их принципом действия и технологией изготовления.

*Производительность памяти* можно характеризовать как скорость потока записываемых или считываемых данных и измерять в мегабайтах в секунду. Производительность подсистемы памяти наравне с производительностью процессора существенным образом определяет производительность компьютера. Выполняя определенный фрагмент программы, процессору придется, во-первых, загрузить из памяти соответствующий программный код, а во-вторых, произвести требуемые обмены данными, и чем меньше времени потребуется подсистеме памяти на обслуживание этих операций, тем лучше.

Производительность памяти, как основной, так и кэша второго уровня, обычно характеризуют *длительностью пакетных циклов чтения* (memory burst read cycle). Пакетный режим обращения является основным для процессоров, использующих кэш (класса 486 и выше); циклы чтения выполняются гораздо чаще, чем циклы записи (хотя бы потому, что процессору приходится все время считывать инструкции из памяти). *Эта длительность выражается в числе тактов системной шины, требуемом для передачи очередной порции данных в пакете. Обозначение вида 5-3-3-3 для диаграммы пакетного цикла чтения соответствует пяти тактам на считывание первого элемента в цикле и трем тактам на считывание каждого из трех последующих элементов. Первое число характеризует латентность (latency) памяти — время ожидания данных, последующие — скорость передачи. При этом, конечно же, оговаривается и частота системной шины.*

Производительность подсистемы памяти зависит от *типа и быстродействия* применяемых запоминающих элементов, *разрядности* шины памяти и некоторых «хитростей» архитектуры. Современные типы памяти обеспечивают высокую скорость передачи внутри пакета, используя двойную и даже четырехкратную синхронизацию. При этом параметром шины, по которой передаются данные, может быть как *частота тактового сигнала*, так и *частота передачи данных*. Последняя может в 2 (DDR SDRAM) или в 4 (DDR2 SDRAM, шина Pentium 4) раза превышать тактовую частоту. Задержка получения

данных чтения процессорным ядром в современных компьютерах может составлять от нескольких десятков до нескольких сотен наносекунд в зависимости от способа подключения памяти.

Производительность микросхем, или модулей памяти, повышают применением различных вариантов конвейеризации, о чем подробнее рассказывается далее.

*Разрядность шины памяти* — это количество байтов (или битов), с которыми операция чтения или записи может быть выполнена одновременно. Именно с целью повышения производительности у 32-битных (по внутренним регистрам) процессоров класса Pentium и выше внешняя шина, связывающая процессор с памятью, имеет разрядность 64 бита. У современных процессоров пропускная способность системной шины превышает пропускную способность шины памяти. Это подталкивает к использованию двухканальной памяти.

Важнейшая характеристика памяти, от которой зависит производительность — это пропускная способность, которая выражается как произведение частоты системной шины на объем данных, передаваемых за каждый такт. В случае с памятью SDRAM мы имеем шину шириной 64 бита или 8 байт, следовательно, к примеру, пропускная способность памяти типа DDR333 составляет  $333 \text{ МГц} \cdot 8 \text{ Байт} = 2,7 \text{ Гбайт/с}$  или  $2700 \text{ Мбайт/с}$ . Отсюда, кстати, и другое название памяти — PC2700, по ее пропускной способности в мегабайтах в секунду. В последнее время часто используете двухканальное подключение памяти, при котором теоретическая пропускная способность удваивается. Т. е., в случае с двумя модулями DDR333 мы получим максимальную возможную скорость обмена данных —  $5,4 \text{ Гбайт/с}$ .

Тем не менее, частота работы памяти и, следовательно, ее теоретическая пропускная способность не являются единственными параметрами, отвечающими за производительность. В действительности не менее важную роль играют и латентность памяти, т. е. значения задержек между подачей команды и ее выполнением. Эти значения принято называть *таймингами*, которые выражаются в тактах, прошедших между моментом поступлением какой-либо команды и ее реальным исполнением.

Общее правило такое: чем меньше тайминги при одной тактовой частоте, тем быстрее память. Более того, в целом ряде случаев быстрее оказывается память с меньшими таймингами, работающая даже на более низкой тактовой частоте. Все дело в том, что, как мы уже упоминали, оперативная память работает синхронно с системной

шиной, поэтому память с частотой не кратной частоте системной шины и с пропускной способностью, превышающей пропускную способность системной шины, никаких преимуществ перед более дешевой не имеет.

К примеру, системная шина современных процессоров Pentium 4 работает на частоте 800 МГц, что при ширине шины 64 бит обеспечивает максимальную пропускную способность в 6,4 Гбайта/с. Оптимальным выбором в этом случае является двухканальная память DDR2 400 с аналогичной пропускной способностью в те же 6,4 Гбайта/с. Использование в двухканальном режиме более дорогих модулей типа DDR2 533/677 реальной прибавки в производительности вряд ли даст. Более того, в иных случаях есть смысл снизить рабочую частоту таких модулей, но добиться более низких таймингов. Это положительно скажется на производительности. Чтобы убедиться в этом, достаточно «прогнать» различные тестовые программы.

Быстродействие процессора выражается в мегагерцах (МГц), а быстродействие запоминающего устройства и его эффективность — в наносекундах (нс).

Для успешного взаимодействия процессора с более медленной основной памятью обычно требуется несколько уровней высокоскоростной кэш-памяти.

Как правило, компьютер работает гораздо быстрее, если пропускная способность шины памяти соответствует пропускной способности шины процессора. Тип памяти, пропускная способность которой соответствует скорости передачи данных процессора, является наиболее приемлемым вариантом для систем, использующих соответствующий процессор.

Производителей самих микросхем памяти не так уж много, гораздо больше фирм занимается выпуском модулей, напаявая чипы на платы собственной разработки. Впрочем, от фирмы, выпускающей модули, зависит не меньше, чем от производителя чипов: для изготовления качественного модуля недостаточно разработать хорошую конструкцию, нужен еще и современный техпроцесс. Кроме того, известные фирмы много внимания уделяют тестированию как самих чипов, так и готовых модулей, при этом чрезвычайно редки ситуации, когда изделия таких компаний не отвечают заявленным параметрам. Среди хорошо себя зарекомендовавших производителей модулей следует назвать фирмы Crucial, Hynix, Kingmax, Kingston, OCZ, Samsung (SEC) и Transcend.

Если вам просто нужен надежный производительный компьютер, работающий с номинальными параметрами, память стоит выбирать исходя из пропускной способности шины.



## 5. БАЗОВАЯ СИСТЕМА ВВОДА–ВЫВОДА (BIOS)

### 5.1. BIOS в структуре программного обеспечения компьютера

Неотъемлемой от компьютера частью программного обеспечения является *базовая система ввода–вывода* (Basic Input-Output System, BIOS), которая хранится в постоянной (энергонезависимой) памяти ROM BIOS (ПЗУ базовой системы ввода–вывода).

Зачем нужна BIOS? Дело в том, что ПЗУ является связующим звеном между операционной системой и «железом». Не будь ROM BIOS, то операционная система была бы слишком привязана к аппаратным средствам и полностью бы от них зависела. Пришлось бы «подгонять» операционную систему под каждую конфигурацию аппаратных средств. Любая система укомплектована своей ROM BIOS, а поскольку операционные системы имеют единый интерфейс для работы с различной аппаратурой, то проблем в несовместимости аппаратной и программной частей, как правило, нет, так как между ними как раз и стоит BIOS. Все это может выглядеть примерно так, как изображено на рис. 5.1:

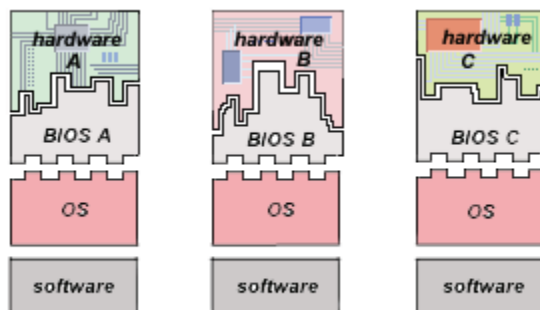


Рис. 5.1. Связь аппаратной и программных составляющих компьютера.

В ROM BIOS находится *программа инициализации*, называемая POST (PowerOn Self Test — самотестирование по включению), которая обеспечивает тестирование и запуск компьютера при включении, а также загрузку операционной системы. В ROM BIOS содержатся процедуры для работы со стандартными устройствами, реализующие

связь операционной системы и прикладных программ с аппаратными средствами компьютера. BIOS предоставляет такие сервисы, как ввод символа с клавиатуры, вывод на экран или принтер, чтение-запись сектора на диске и ряд других. *BIOS* находится на самом нижнем уровне ПО, который обеспечивает изоляцию вышестоящих уровней от подробностей реализации аппаратных средств компьютера. В ROM BIOS имеется также утилита CMOS Setup, обеспечивающая настройку аппаратных средств компьютера.

Следующий уровень — *операционная система* (ОС), основным назначением которой является загрузка прикладных программ и предоставление им некоторых сервисов. Сервисы ОС функционируют на более высоком уровне. Если BIOS работает с физическими устройствами, то ОС предоставляет возможность работы на логическом уровне. Сервисы ОС, обслуживающие стандартные устройства, могут обращаться к соответствующим сервисам BIOS; они расширяют функциональность сервисов BIOS, а также выполняют обработку ошибок физических устройств. ОС может работать с системными устройствами и в обход BIOS, через собственные драйверы. В MS-DOS все сервисы ОС обслуживались через BIOS, что обеспечивало высокий уровень совместимости и переносимости ПО с машины на машину, но ценой невысокой эффективности. Более современные ОС работают в обход BIOS — это эффективнее с точки зрения производительности, но осложняет переносимость ПО. ОС осуществляет распределение системных ресурсов.

И наконец, верхний уровень иерархии ПО — *прикладное программное обеспечение*. Прикладные программы могут пользоваться сервисами ОС, BIOS, а также обращаться к аппаратным средствам компьютера напрямую, адресуясь к портам и ячейкам памяти. Чем ближе прикладная программа к аппаратной части, тем эффективнее она может с ней работать (вызовы сервисов BIOS, а тем более ОС вносят дополнительные накладные расходы на организацию программных интерфейсов). Однако использование сервисов высокого уровня (BIOS, а тем более ОС) страхует от возможных проблем совместимости программного обеспечения с аппаратными средствами компьютера. Прикладные программы, как и ОС, как правило, загружаются с устройств внешней памяти. Именно возможность загрузки любой прикладной программы в совокупности с неограниченным ассортиментом подключаемых устройств и позволяет считать персо-

нальный компьютер универсальным инструментом с неограниченными возможностями.

Важными компонентами программного обеспечения являются *драйверы* (driver — буквально, водитель) — программные модули, содержащие процедуры работы с устройствами. Необходимость выделения драйверов в отдельные модули вполне очевидна: устройство определенного назначения (к примеру, дисплейный адаптер) может иметь самые разные реализации — от MDA до современных видеокарт с трехмерными акселераторами. Для прикладных программ или операционной системы драйвер представляет *набор сервисов* с интерфейсом, понятным «потребителю»-устройству.

## 5.2. Системный модуль ROM BIOS

ROM BIOS состоит из четырех основных компонентов.

**POST (Power-On Self Test)** – Процедура самотестирования при включении питания. При проверке тестируются центральный процессор, само ПЗУ, элементы материнской платы, ОЗУ и основные периферийные устройства. Грубо говоря, POST – это набор маленьких подпрограмм. Проверка не является тщательной, но серьезные неисправности все же обнаруживаются. Информация об ошибках выводится тремя путями: вывод информации в виде текста или шестнадцатеричных кодов ошибок на экран и набор звуковых сигналов. В случае успешного прохождения теста встроенный динамик выдает один короткий сигнал.

**BIOS (Basic Input/Output System)** – Это набор драйверов (драйвер – программа управления устройством), обеспечивающих работу системы при запуске компьютера или при загрузке в безопасном режиме. Дело в том, что когда включается компьютер, то еще до загрузки операционной системы можно управлять им с клавиатуры, видеть все действия на его мониторе. Кроме этого, если загружать его в безопасном режиме, то происходит отказ от драйверов операционной системы и в работе остаются только драйвера BIOS.

**Программа начальной загрузки** – Эта программа после включения компьютера ищет операционную систему, загружает ее и отдает ей все управление компьютером.

**CMOS (Complementary Metal Oxide Semiconductor)** – точнее процедура установки параметров конфигурации системы в CMOS. Эту часть ROM BIOS можно просмотреть и при необходимости внести туда изменения, если сразу после включения компьютера нажать клавишу [Delete]. В основном все параметры касаются материнской платы, но так или иначе они связаны и остальными компонентами компьютера.

Базовая система ввода – вывода является ключевым элементом системной платы.

*Системный модуль ROM BIOS (System ROM BIOS)* обеспечивает программную поддержку стандартных устройств PC, конфигурирование аппаратных средств, их диагностику и вызов загрузчика операционной системы. Системный модуль ROM BIOS в значительной степени привязан к конкретной реализации системной платы, поскольку именно ему приходится программировать все микросхемы чипсета системной платы. Функции BIOS разделяются на следующие группы:

- инициализация и начальное тестирование аппаратных средств, POST;
- настройка и конфигурирование аппаратных средств и системных ресурсов, CMOS Setup;
- автоматическое распределение системных ресурсов, PnP BIOS;
- идентификация и конфигурирование устройств PCI, PCI BIOS;
- начальная загрузка (первый этап загрузки операционной системы), Bootstrap Loader;
- обслуживание аппаратных прерываний от системных устройств (таймера, клавиатуры, дисков), BIOS Hardware Interrupts;
- отработка базовых функций программных обращений (сервисов) к системным устройствам, ROM BIOS Services;
- поддержка управляемости конфигурированием, DMI BIOS;
- поддержка управления энергопотреблением и автоматического конфигурирования, APM и ACPI BIOS.

Все функции или их часть исполняет системный модуль BIOS, хранящийся в микросхеме ПЗУ или флэш-памяти на системной плате. Большинство сервисных функций выполняется в 16-битном режиме, хотя некоторые новые функции могут иметь и альтернативные вызовы для 32-битного исполнения.

Системный модуль BIOS должен обслуживать по вышеуказанным функциям все компоненты, установленные на системной плате: процессор, контроллер памяти (ОЗУ и кэш), стандартные архитектурные компоненты (контроллеры прерываний и DMA, системный таймер, системный порт, CMOS RTC), контроллер клавиатуры, а также набор стандартных периферийных контроллеров и адаптеров, даже если они и не установлены на системной плате. В этот набор входят графические адаптеры, порты COM и LPT, контроллер НГМД, диски АТА. Если на системной плате установлены дополнительные компоненты, например контроллер SCSI, графический адаптер SVGA, адаптер локальной сети, то их поддержка тоже должна быть в системном модуле BIOS.

### **5.3. Начальный запуск и самотестирование. Программа POST**

При *включении* питания, аппаратном сбросе от кнопки Reset или нажатии комбинации клавиш Ctrl+Alt+Del процессор переходит к исполнению кода начального самотестирования *POST* (PowerOn Self Test — самотестирование при включении), хранящегося в микросхеме BIOS. POST выполняет тестирование процессора, памяти и системных средств ввода-вывода, а также конфигурирование всех программно-управляемых аппаратных средств системной платы. Часть процедуры конфигурирования выполняется однозначно, часть управляется «джамперами» системной платы, но ряд параметров позволяет или даже требует конфигурирования по желанию пользователя. Для этих целей служит утилита *Setup*, встроенная в код BIOS. После тестирования и конфигурирования (включающего настройку устройств PnP) POST инициирует загрузку операционной системы.

Во время выполнения POST может выдавать диагностические сообщения в виде последовательности коротких и длинных звуковых сигналов, а после успешной инициализации графического адаптера — в виде небольших текстовых сообщений на экране монитора.

#### ***Примечание***

На новых системных платах реализуется и речевая звуковая диагностика (voice diagnostics) и через динамик пользователю предлагают проверить установку модулей памяти, видеоадаптера, подключения кабеля винчестера и т. п. Язык сообщений можно выбрать в

CMOS Setup (когда эту процедуру удастся запустить), правда, русского языка пока не встречается.

Итак, при включении блок питания вырабатывает сигнал аппаратного сброса, приводящего все узлы в исходное состояние. Этот же сигнал вырабатывается и при нажатии кнопки Reset. Во время действия сигнала аппаратного сброса процессор пассивен — он не управляет системной шиной. Процессор подготавливается к работе, принимая со своих выводов сигналы, задающие его конфигурацию (коэффициент умножения, роль в многопроцессорных системах и некоторые другие параметры). Внутренний кэш очищается, регистры (не все) приводятся в определенное состояние. Сигнал сброса поступает на все устройства (контроллеры и адаптеры), нуждающиеся в переводе в исходное состояние и находящиеся на системной плате или подключаемые к шинам расширения. После окончания сигнала процессор по определенному адресу выбирает из памяти и исполняет первую инструкцию — управление передается на точку входа в программу инициализации компьютера (POST). Первым делом необходимо выполнить инициализацию процессора — установить значения некоторых регистров. После сброса процессор всегда начинает работу в реальном режиме (совместимом с 8086/88). Далее происходят проверка работоспособности и инициализация подсистем компьютера. Эта «раскрутка» выполняется в несколько этапов, причем постепенно в работу вовлекаются протестированные подсистемы. Поначалу программа может смело пользоваться только регистрами процессора и постоянной памятью (которые тоже желательно проверить, например, подсчитать контрольную сумму). Если ПЗУ исправно, можно двигаться дальше, в противном случае — лучше остановиться. Пока неизвестна работоспособность ОЗУ, оперативной памятью пользоваться нельзя и, следовательно, недоступны вызовы процедур и обработка прерываний (вызвать процедуру можно, а вот возврат не гарантируется, поскольку адрес возврата берется из стека, то есть из ОЗУ). Далее инициализируется ОЗУ (программируются регистры чипсета, заведующие настройкой контроллера памяти и регенерацией) и выполняется тестирование небольшого блока в начале ОЗУ. Если тест проходит успешно, то для дальнейшей работы уже можно пользоваться и вызовами, и прерываниями (не забыв проинициализировать таблицу прерываний), и задействовать память для хранения переменных — в таком окружении работать гораздо удобнее. Затем можно проинициализировать и протестировать дисплейный адаптер, и дальнейшая «раскрутка» пой-

дет уже с «ожившим» экраном. Далее тестируют ОЗУ в полном объеме, определяют наличие контроллеров и адаптеров, инициализируют их и тестируют. После этого программа POST узнает конфигурацию компьютера и готова к загрузке операционной системы. Векторы прерываний, за которые отвечает BIOS, проинициализированы — ими можно пользоваться. В этот момент можно войти в меню встроенной утилиты конфигурирования CMOS Setup и изменить настройку различных подсистем компьютера. После окончания работы этой утилиты тест POST приходится выполнять снова, т.к. конфигурация может стать уже иной. Программа POST завершается вызовом процедуры начальной загрузки.

К моменту окончания теста POST все стандартные **устройства** (клавиатура, дисплей, диски, порты) приводятся в состояние готовности к работе в стандартном режиме по умолчанию, часть настроек может выполняться в соответствии с выбранными установками CMOS Setup. Список обнаруженных устройств и их основные параметры могут отображаться в таблице на экране, но возможна и такая настройка CMOS Setup, при которой экран будет пустым до появления логотипа (или текстового сообщения) загружаемой ОС.

### **Загрузка ОС и прикладных программ**

Последним шагом программы POST является выполнение *процедуры начальной загрузки* (bootstrap loader), которая вызывается как программное прерывание Int 19h BIOS. Ее задача довольно скромная — с выбранного устройства загрузить в память один блок (сектор) данных, и если он похож на загрузчик передать ему управление. Напомним, что «передать управление» означает выполнить инструкцию перехода на адрес точки входа в программу, загруженную в оперативную память. На этом деятельность POST заканчивается, и компьютер управляется загружаемым ПО.

## **5.4. Начальная загрузка**

Стандартная процедура начальной загрузки (bootstrap loader), вызываемая по прерыванию Int 19h BIOS в конце теста POST, выбирает устройство начальной загрузки (Initial Program Loader, IPL) — блочное устройство, поддерживающее функцию чтения секторов. С этого устройства процедура пытается загрузить в ОЗУ самый первый сектор, и если у него в конце имеется сигнатура загрузчика, ему пере-

дается управление. До выполнения начальной загрузки должны быть инициализированы перечисленные ниже *загрузочные устройства* (boot device), которыми пользуется сама эта процедура и загружаемые ею модули:

- *устройство ввода* (input device) — как правило, клавиатура, с которой можно управлять загрузкой, отвечая на запросы. Это устройство должно поддерживать посимвольный ввод — сервис Int 09h BIOS;

- *устройство вывода* (output device) — как правило, дисплей, на который выводятся сообщения загрузчика. Это устройство должно поддерживать посимвольный вывод — сервис Int 10h BIOS;

- *устройство начальной загрузки* (IPL) — как правило, НГМД, НЖМД и другие устройства, поддерживающие функции блочного чтения — сервис Int 13h (02 или 42h) BIOS. Для краткости в дальнейшем его будем называть просто устройством загрузки.

Процедура загрузки определяет первое готовое устройство из списка разрешенных и доступных загрузочных устройств. Загрузочным устройством может быть НГМД, жесткий диск, CD/DVD, флэш-память, сетевой адаптер. Список устройств, разрешенных для загрузки, а также порядок их опроса задаются параметрами CMOS Setup. При традиционном порядке опроса (А, С) сначала анализируется готовность НГМД попыткой прочесть загрузочную запись дискеты — первый сектор нулевой дорожки. Если дискета отсутствует (при этом дисковод не может прочитать никаких данных), то процедура переходит к попытке чтения главной загрузочной записи жесткого диска С. Если сектор с дискеты прочесть удалось, то по определенному признаку в его содержимом процедура определяет, имеется ли и нем загрузчик ОС. Если загрузчик есть, то ему передается управление для загрузки операционной системы. Если в считанном секторе загрузчик не обнаружен, то компьютер останавливается с сообщением Non-system disk or disk error (несистемный диск или ошибка чтения), после которого остается лишь установить системную дискету или, наоборот, извлечь дискету и нажать комбинацию трех клавиш Ctrl+Alt+Del. Это приведет к повторному запуску теста POST, но в сокращенном варианте, который завершится вызовом прерывания Int 19h — следующей попыткой загрузки. Такой способ перезапуска (перезагрузки) называется «теплым» (warm). Повторный запуск POST можно выполнить и «холодным» (cold) способом, нажав кнопку Reset, но при этом



POST будет выполняться полностью, как после включения питания, что несколько дольше.

Загрузка с жесткого диска (в нашем случае — если не обнаружена готовность НГМД) тоже начинается с попытки чтения первого сектора нулевой головки нулевого цилиндра, но здесь ищется главный загрузчик, или главная загрузочная запись (Master Boot Record, MBR), с таблицей разделов диска. Этот загрузчик должен найти описатель активного раздела, загрузить в память его первый сектор и, убедившись в том, что в нем находится загрузчик ОС, передать ему управление. Загрузчик ОС и должен загрузить операционную систему из выбранного (активного) раздела в память и передать ей управление. Главный загрузчик может являться и средством для выбора загружаемой ОС (boot manager) по желанию пользователя (из меню).

Как уже отмечалось, последовательность опроса устройств, с которых делается попытка загрузки операционной системы, задается параметрами CMOS Setup. «Классической» является последовательность А, С, и в старых компьютерах не было способа ее изменения. Позже появилась возможность ее изменения на обратную (С, А), а затем в нее стали включать и CD-ROM, и флэш-память, и даже обнаруженные устройства удаленной загрузки (Remote Program Loader, RPL), позволяющие загружать ОС по сети. Если первое по порядку загрузочное устройство не готово (не установлен сменный диск, на жестком диске нет активного раздела), то процедура обращается к следующему по списку устройству. Таким образом, современные компьютеры могут загружаться не только с первого жесткого диска (С), но и с другого, если на нем имеется активный раздел. Правда, активизировать раздел с помощью утилиты FDISK, как это обычно и делается, можно только на первом физическом жестком диске (таково ее ограничение), так что для этого приходится временно переставлять устройства.

Главный загрузчик инвариантен по отношению к загружаемым операционным системам и дискам, его программный код, как и таблица разделов, записывается утилитой FDISK при конфигурировании жесткого диска.

Помимо вышеописанного штатного способа загрузки с диска, выполняемой традиционным системным модулем BIOS, имеется возможность загрузки с любого устройства, с которого можно загрузить в память требуемый блок данных. Так, возможна загрузка с CD-ROM, но для этого необходимо выполнение специальной процедуры, включающей эмуляцию диском CD-ROM дискеты или жесткого диска.

Возможна загрузка и с нестандартного устройства, подключаемого через карту расширения. Однако системный модуль BIOS сам этого не умеет, и такое загрузочное устройство должно иметь ПЗУ расширения BIOS с собственной процедурой загрузки. Для регулярной работы попытка загрузки с гибкого диска вредна по ряду причин, самой безобидной из которых является то, что на опрос готовности накопителя затрачивается излишнее время. Хуже то, что случайно оставленная в накопителе дискета может содержать вирус в загрузочном секторе, и попытка загрузки с такой дискеты, даже если на ней нет операционной системы, передаст управление коду этого вируса. При незагруженной операционной системе вирусу «залезть» в файловую систему затруднительно, но «поселиться» в загрузочном секторе жесткого диска легко. Дальнейшее развитие событий зависит от уровня опасности вируса и опытности пользователя (пока операционную систему не загрузили, этот вирус из загрузчика не распространится). Запрет попытки загрузки с дискеты служит и средством защиты от несанкционированного доступа к данным. Если загрузка операционной системы закрыта паролем, а разрешена загрузка с дискеты, то, загрузив операционную систему, взломщик получит программный доступ к жесткому диску — по крайней мере, на физическом уровне. При заданной последовательности C, A загрузчик активного раздела жесткого диска не допустит передачи управления загрузчику с дискеты. Конечно, если изменение последовательности производится в целях защиты от несанкционированного доступа, вход в CMOS Setup также нужно закрыть паролем.

Загруженная *операционная система* выполняет инициализацию подведомственных ей программных и аппаратных средств. Она добавляет новые сервисы, вызываемые, как правило, тоже через программные прерывания, и расширяет некоторые сервисы BIOS. Под управлением операционной системы загружаются и исполняются *пользовательские приложения*, вспомогательные процессы и утилиты. Операционная система осуществляет распределение всех ресурсов компьютера — памяти (как оперативной, так и пространства на устройствах хранения данных), процессорного времени (в многозадачных системах), периферийных и коммуникационных устройств. Она же предоставляет пользователям интерфейс, с помощью которого пользователи запускают приложения, настраивают параметры ОС и выполняют иные действия.

Приложения и утилиты располагаются в виде файлов на устройствах внешней памяти — локальных и сетевых дисках или других устройствах хранения. Пользователь запускает интересные его приложения, указывая имя требуемой программы (или командного файла) с клавиатуры в ответ на приглашение (prompt) ОС или же «щелкая» на представляющем его значке.

## 5.5. Флэш-BIOS

Флэш-память широко применяется в качестве носителя BIOS в современных компьютерах. В принципе, это позволяет даже конечному пользователю обновлять версию BIOS, не вызывая высокооплачиваемых специалистов и оперативно получая необходимые файлы через Интернет. Наиболее эффективно применение флэш-памяти с выделенным блоком загрузчика. Блок загрузчика после программирования может быть аппаратно защищен от перезаписи и работать в режиме ROM. Это позволяет его использовать как неизменяемую часть BIOS, обеспечивающую минимальные условия для загрузки утилиты программирования основного блока. Основной блок хранит главную часть BIOS, которая при необходимости может заменяться новыми версиями. В случае некорректности новой запрограммированной версии всегда есть путь к отступлению, обеспечиваемый неизменяемым блоком загрузчика.

*Об обновлении ROM BIOS.* Это ответственная процедура и поэтому определитесь — действительно ли Вам необходимо обновлять версию BIOS. Вот несколько советов для тех, кто самостоятельно обновляет BIOS

1. Помните всегда — нет никаких гарантий, что обновление пройдет гладко, безболезненно и безошибочно.
2. Не факт, что если система нормально работает с данной версией BIOS, то будет еще лучше работать с обновленной.
3. Обязательно сохраните текущую версию BIOS перед заменой.
4. Прежде чем заняться обновлением, соберите как можно больше информации о процессе смены версий BIOS конкретно для Вашей материнской платы.
5. Если нет опыта работы с DOS, следует потренироваться в работе с этой операционной системой.
6. И самое главное, помните, что если во время перезаписи пропадет питание или случится еще какой-то сбой, то последствия непредсказуемы. Тогда, возможно, останется один вариант — помощь специалиста.

## 6. ПИТАНИЕ КОМПЬЮТЕРОВ

### 6.1. Блок питания компьютера

Главное назначение блоков питания — *преобразование электрической* энергии, поступающей из сети переменного тока, в энергию, пригодную для питания узлов компьютера. Блок питания преобразует сетевое переменное напряжение 220 В, 50 Гц (120 В, 60 Гц) в постоянные напряжения +3,3, +5 и +12 В. Как правило, для питания цифровых схем (системной платы, плат адаптеров и дисковых накопителей) используется напряжение +3,3 или +5 В, а для двигателей (дисководов и различных вентиляторов) — +12 В. Компьютер работает надежно только в том случае, если значения напряжения в этих цепях не выходят за установленные пределы.

В персональных компьютерах используются импульсные, а не линейные блоки питания. В линейном блоке применяется большой встроенный трансформатор для формирования напряжений питания разной величины, а в импульсном — генератор высокой частоты для формирования различных напряжений питания. Импульсный блок питания имеет меньшие размеры, меньший вес и более низкое энергопотребление. Линейные блоки питания имеют по меньшей мере три очевидных недостатка. Во-первых, выходное напряжение трансформатора линейно следует входному напряжению (отсюда и название линейный), поэтому любые скачки переменного тока отражаются на выходном напряжении. Во-вторых, потребность ПК в большой мощности требует использования проводов большого сечения для трансформатора. И, наконец, в-третьих, переменный ток с частотой 50 Гц трудно фильтровать внутри блока питания, т. е. необходимы большие и дорогие конденсаторы фильтра, а также стабилизаторы. Импульсный блок питания, в свою очередь, характеризуется импульсной схемой, принимающей входящую энергию на относительно высокой частоте. Это позволяет использовать более легкие и дешевые высокочастотные трансформаторы. Кроме того, высокие частоты выходного напряжения гораздо проще фильтровать, а входное напряжение часто нестабильно. Изменение входного напряжения от 90 до 135 В все равно приводит к подаче нужного выходного напряжения, а многие импульсные блоки питания автоматически переключаются на входное напряжение 220 В. Особенность импульсных блоков питания заклю-

чается в том, что они не работают без *нагрузки*, т. е. к источникам +5 В (+12 В) должны быть подключены какие-либо потребители энергии. Если поставить блок питания на стол, ничего к нему не подсоединив, и включить в сеть, то либо внутренняя схема защиты его отключит, либо он перегорит. Как правило, блоки питания защищены от работы без нагрузки и отключаются, но в некоторых дешевых моделях схема защиты отсутствует, и на холостом ходу они моментально выходят из строя.

Блок питания PC обеспечивает напряжениями постоянного тока системный блок со всеми его сложными и часто привередливыми устройствами. С самых первых моделей PC здесь применяется двухтактная схема преобразователя с бестрансформаторным входом, без революционных изменений эта схема дошла и до наших дней (ее упрощенный вариант приведен на рис. 6.1).

Входное напряжение после высокочастотного фильтра выпрямляется и поступает на накопительные конденсаторы (C1 и C2), являющиеся главными хранителями энергии на случай кратковременного провала питающего напряжения. Мощные высоковольтные транзисторы T1 и T2 и конденсаторы C1 и C2 образуют полумостовую схему генератора-преобразователя, нагрузкой которого является высокочастотный импульсный силовой трансформатор Tr2. Этот трансформатор обеспечивает и гальваническую развязку выходных и входных цепей. Преобразователь является регулирующим элементом стабилизатора напряжения основного источника: +3,3 В для ATX (и более новых конструктивов) или +5 В (AT). Остальные напряжения могут быть стабилизированы дополнительными выходными стабилизаторами, но чаще их оставляют нестабилизированными. При этом чем больше нагрузка блока по основной (стабилизированной) цепи, тем выше напряжения на остальных шинах. Убедиться в этом просто — понаблюдайте за вентилятором блока питания, который питается от цепи +12 В, изменяя нагрузку по основной цепи, например подключая и отключая системную плату. При подключении нагрузки скорость вращения вентилятора повышается. Это происходит потому, что с повышением тока нагрузки преобразователь вырабатывает более широкие импульсы, а выходное напряжение нестабилизированных выпрямителей при постоянной нагрузке пропорционально ширине этих импульсов. По этой причине уровни напряжения на не основных выходах большинства блоков питания соответствуют номиналам лишь при номинальной и сбалансированной нагрузке. Однако, как правило, потребители этих напряже-

ний не требуют особой точности напряжения, а стабильность обеспечивается относительным постоянством нагрузки основной цепи.

Для мощных блоков питания обеспечить работу в широком диапазоне питающих напряжений довольно сложно, и на них устанавливают переключатель входного напряжения:

- 230 В — напряжение в диапазоне 180-265 В;
- 115 В — напряжение в диапазоне 90-135 В.

При включении в сеть 220 В блока, предназначенного для работы при напряжении 110 В, часто выходят из строя ключевые транзисторы или диоды. Блоки, у которых указано свойство Autoswitching Power Supply, работают в диапазоне 110-230 В без переключателя. В них применяют силовые компоненты с большим запасом по допустимым напряжению и току.

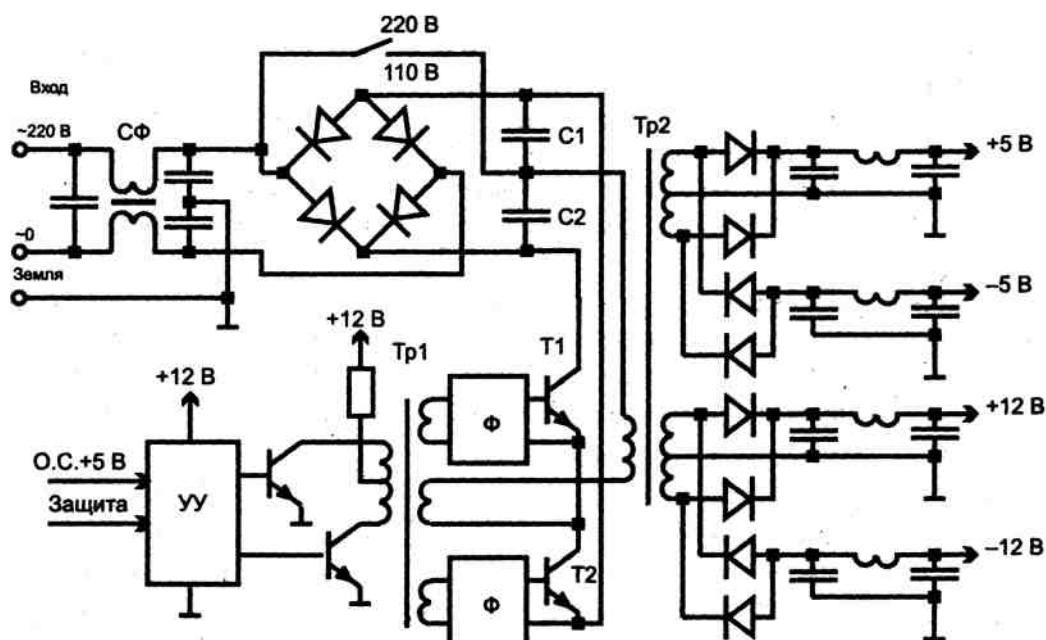


Рис. 6.1. Двухтактный блок питания: СФ – сетевой фильтр, УУ – устройство управления, Ф – формирователи импульсов, Тр1 – трансформатор развязки цепей управления, Тр2 – силовой трансформатор

### Сигнал Power\_Good

Блок питания не только вырабатывает необходимое для работы узлов компьютера напряжение, но и приостанавливает функционирование системы до тех пор, пока величина этого напряжения не достигнет значения, достаточного для нормальной работы. Иными сло-

вами, блок питания не позволит компьютеру работать при «нештатном» уровне напряжения питания. В каждом блоке питания перед получением разрешения на запуск системы выполняется внутренняя проверка и тестирование выходного напряжения. После этого на системную плату посылается специальный сигнал *Power\_Good* (питание в норме). Если такой сигнал не поступил, компьютер работать не будет. Уровень напряжения сигнала *Power\_Good* – около +5 В (нормальной считается величина от +3 до +6 В). Он вырабатывается блоком питания после выполнения внутренних проверок и выхода на номинальный режим и обычно появляется через 0,1–0,5 с после включения компьютера. Сигнал подается на системную плату, где микросхемой тактового генератора формируется сигнал начальной установки процессора.

При отсутствии сигнала *Power\_Good* микросхема тактового генератора постоянно подает на процессор сигнал сброса, не позволяя компьютеру работать при «нештатном» или нестабильном напряжении питания. Если выходные напряжения блока питания не соответствуют номинальным (например, при снижении напряжения в сети), сигнал *Power\_Good* отключается и процессор автоматически перезапускается. При восстановлении выходных напряжений снова формируется сигнал *Power\_Good* и компьютер начинает работать так, будто его только что включили. Благодаря быстрому отключению сигнала *Power\_Good* компьютер «не замечает» неполадок в системе питания, поскольку останавливает работу раньше, чем могут появиться ошибки четности и другие проблемы, связанные с неустойчивостью напряжения питания.

## **6.2. Питание блокнотных ПК**

Питание блокнотных ПК значительно отличается от питания настольных — у блокнотных ПК имеется встроенный аккумулятор, обеспечивающим автономную работу в течение нескольких часов. От аккумулятора (напряжение 12–20 В) питается импульсный преобразователь, поддерживающий необходимые уровни питающих напряжений. Внешний блок питания (адаптер питания) дает напряжение 12–20 В (зависит от аккумулятора ПК) при питании от сети переменного тока или бортовой сети автомобиля (12 В). В зависимости от модели компьютера от блока требуется мощность в диапазоне 20–60 Вт. Внешний блок обеспечивает зарядку аккумулятора и работу ПК при

наличии питающей сети. Заметим, что такая схема питания от сети (при исправном аккумуляторе) избавляет от необходимости использования источников бесперебойного питания в тех местах, где питающая сеть ненадежна. Возможна работа компьютера и без аккумулятора (но и без защиты от перебоев и провалов напряжения). Внешний блок обеспечивает гальваническую развязку компьютера от питающей сети.

Аккумуляторы для блокнотных и других малогабаритных ПК довольно сложные и дорогостоящие компоненты: от них требуется большая емкость при ограниченной массе и габаритах. Аккумуляторы имеют внутренние средства контроля уровня заряда, что позволяет оптимизировать процесс их зарядки и прогнозировать момент отключения компьютера из-за разрядки батарей при отсутствии внешнего питания.

Для оптимизации работы блокнотного ПК на нем должна быть установлена программная *система управления потреблением* (например, PowerGear), отслеживающая состояние внешнего питания и уровень заряда аккумулятора. Эта система управляет потреблением основных подсистем ПК: процессора, памяти, графического адаптера, винчестера, дисплея, а также информирует пользователя о перспективах отключения. У этих подсистем есть возможность работать быстро (с максимальным потреблением) или экономно (с минимальной производительностью), возможны также промежуточные варианты. Политику управления питанием выбирает и настраивает пользователь.

### **6.3 Средства улучшения качества электропитания**

Электронное оборудование, питающееся от сети переменного тока, подвергается различным негативным воздействиям со стороны этой питающей сети. Стандартным требованием к питающей сети является напряжение питания 220 В с допустимыми отклонениями от -15 до +10 % от номинала (187-242 В) при частоте 50+1 Гц. Возмущения со стороны сети могут приводить к сбоям (импульсным помехам и провалам питающего напряжения), самопроизвольному отключению или перезапуску устройств и даже к выходу их из строя под воздействием импульсных напряжений или длительных перенапряжений. Поскольку большинство блоков питания имеют импульсный преобразователь с бестрансформаторным входом, к отклонениям частоты или формы напряжения они обычно почти нечувствительны. Однако последствия сбоев питания могут быть весьма тяжелыми,



вплоть до потери данных на диске мощного и ответственного сервера (не считая выхода из строя аппаратуры).

Для защиты от воздействия сетевых возмущений применяется целый комплекс мер.

*Сетевой LC-фильтр* задерживает высокочастотные помехи из сети и в сеть от импульсных блоков питания. Этот фильтр входит в состав практически любого блока питания, а также в сетевые колодки питания типа «Pilot» и им подобные.

*Ограничитель перенапряжений* (surge protector) подавляет высоковольтные выбросы, как относительно длинные коммутационные (до 10 мс), возникающие при переключениях мощных цепей, так и короткие, грозовые. Энергия импульсов перенапряжений поглощается полупроводниковым *варистором*. При хорошем подборе параметров варистор может спасти также от длительных (и значительных) повышений напряжения сети, например, из-за перекоса фаз. В этом случае варистор ограничивает напряжение, выделяя значительную мощность, что приводит к его пробое на короткое замыкание и отключению питания предохранителями токовой защиты (если они есть и рассчитаны на соответствующий ток).

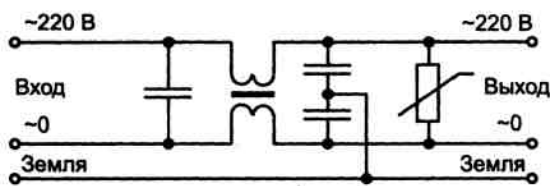


Рис. 6.2. Фильтр-ограничитель с варистором

На рис. 6.2 приведена схема фильтра, комбинированная с ограничителем перенапряжений.

От внезапного пропадания напряжения сети предохраняют **источники бесперебойного питания** — ИБП (Uninterruptible Power System, UPS). В их состав обязательно входят аккумуляторные батареи, выпрямитель входного напряжения и инвертор, обеспечивающий нагрузку напряжением переменного тока.

Источники бесперебойного питания различают по классам (режимам работы). Существуют блоки Off-Line (Stand-By), Line-Interactive и On-line; их «полезность» (и цена) растут в порядке этого перечисления.

От класса, мощности устройства и емкости батарей, определяющей время автономной работы при максимальной нагрузке, существенно зависит цена ИБП.

При пропадании сетевого напряжения ИБП переключается на резервное питание и обычно подает звуковой сигнал. Для защиты данных компьютера устройство ИБП должно иметь возможность передать сигнал о грядущем отключении питания. Сигнал может подаваться аппаратным прерыванием через специальную плату сопряжения с PC или разъем PS/2 Mouse (как варианты у Smart UPS), через СОМ-порт или встроенный в ИБП адаптер ЛВС. Два последних варианта более универсальны и обеспечивают двунаправленный обмен развернутой управляющей и диагностической информацией. При восстановлении питания происходит обратное переключение, и батареи подзаряжаются. Если питание не восстановилось за время работы батарей, ИБП отключается, а его повторное включение после подачи напряжения может быть ручным или автоматическим.

Современные модели ИБП имеют в своем составе микроконтроллер, который в совокупности со специализированным ПО серверов и станций, поставляемым для конкретных моделей, может предоставлять широкий спектр услуг в зависимости от интерфейса связи ИБП с системой.

– *Телеметрия.* Информация о состоянии питающей сети, батареи и других узлов, температуре внутри ИБП, величине нагрузки и т. д., передается в систему сбора, обработки и отображения информации. Система может прогнозировать время работы от батарей и соответственно корректировать задержку закрытия сервера.

– *Телеуправление.* Двунаправленный интерфейс с ИБП обеспечивает подачу управляющих команд — отключение, запуск диагностических тестов и т. д.

– *Планирование включения и выключения.* Администратор может задать график работы сервера, указывая время включения и отключения питания на каждый день недели. Программа при наступлении времени отключения посылает предупреждение всем клиентам, через некоторое время инициирует закрытие сервера и программирует ИБП на отключение питания через определенный интервал времени и повторное включение в заданное время. После отключения по команде ИБП переходит в режим ожидания и своим внутренним таймером отсчитывает время до включения. В заданное время ИБП включает питание нагрузки, сервер автоматически загружается, и следующее за-

планированное отключение произойдет по инициативе программы, работающей на сервере.

Возможности взаимодействия по сети оператора с ИБП определяются ПО этого устройства. Популярные пакеты PowerChute («парашют») для Smart UPS фирмы APC, OnliNet Basic для ИБП фирмы EXIDE обеспечивают вышеперечисленные функции для различных ОС, они вполне удовлетворительны для систем с одним устройством ИБП. В системах с более сложным питанием желательно использовать сетевые варианты ПО, предоставляющие централизованное управление сетями ИБП. Для ИБП фирмы EXIDE это OnliNet Network, OnliNet NVX и др.

Простейшая программная поддержка ИБП должна обеспечивать оповещение о пропадании сетевого напряжения и принудительное завершение работы приложений и операционной системы, когда остается небольшой ресурс времени автономной работы (от аккумулятора). Сигнал о пропадании сетевого напряжения от ИБП к защищаемому компьютеру должен подаваться обязательно, он инициирует оповещение. Принудительное завершение может выполняться по дополнительному сигналу ИБП, когда устройство «чувствует», что батарей хватит только на определенное время. Возможна настройка ПО на работу только от одного сигнала — принудительное завершение инициируется, если сигнал пропадания напряжения удерживается дольше заданного времени. Штатная служба UPS в Windows NT/9x позволяет использовать для сигнализации управляющие сигналы COM-порта: линия CTS — для сигнализации о пропаже Питающего напряжения (power fail), DCD — для сигнализации о малом ресурсе батарей (battery low). Интерфейс настройки сервиса позволяет выбрать полярность сигналов, а также использовать только первый сигнал и инициировать завершение по тайм-ауту. Некоторые модели ИБП указанные сигналы в двухполярном представлении, воспринимаемом COM-портом, не генерируют, а имеют интерфейс «сухой контакт». Событие отражается замыканием или размыканием этого контакта, гальванически не связанного ни с какими цепями.

Более точное представление о состоянии ИБП, а также планирование включения-выключения питания возможны только при полноценной двусторонней связи ИБП со специальным модулем ПО, функционирующим на компьютере. Наиболее широко распространенный вариант связи — через COM-порт. Многие модели ИБП имеют разъем DB9, который обычно и используется интерфейсом RS-232. Однако

зачастую назначение его контактов в значительной степени отличается от стандартного.

#### 6.4. Управление энергопотреблением

В работе персонального компьютера после загрузки ОС и запуска приложения часто возникают довольно длительные паузы (когда пользователь отвлекается на свои дела). В это время вхолостую расходуются электроэнергия (выделяется тепло), а также жизненные ресурсы некоторых устройств (например, зря выгорает люминофор монитора). Чтобы сократить эти напрасные потери, уже давно в компьютер ввели средства *управления энергопотреблением* (power management), а затем и улучшили их, обеспечив *расширенное управление энергопотреблением APM* (Advanced Power Management). Цель APM — переводить в «спящее» состояние (sleep) устройства, не требующиеся для работы в данный момент, и «будить» их (wake up) по первому требованию, по возможности незаметно для приложений и пользователей. Менеджер APM представляет собой часть системного ПО (BIOS и ОС), для работы ему требуется поддержка со стороны устройств. В «спячку» разные устройства можно вводить по-разному: у монитора можно погасить лучи и даже остановить генераторы развертки, у винчестера — остановить шпиндель, у процессора — остановить внутреннее тактирование или понизить эффективную тактовую частоту, память можно перевести в режим саморегенерации. С самого начала подобие «спячки» применяют для НГМД — мотор останавливается, если в течение 2 с нет обращений к диску.

Заснувшее устройство должно реагировать на свои интерфейсные сигналы, чтобы проснуться по сигналу от APM или по команде обычного обращения. В спящий режим устройства могут переходить по таймерам, отсчитывающим время от последнего обращения, прекращения активности пользователя (отсутствия сигналов от устройств ввода), а также по команде пользователя.

Пробуждение системы выполняется по какому-либо внешнему событию — нажатию клавиши на клавиатуре (или специальной кнопки на системном блоке), движению мыши, специальному обращению по локальной сети (wake on LAN) или телефонному вызову модема. Конечно же, пробуждение устройства занимает некоторое время, требующееся на разгон двигателя, нагрев трубки монитора и подобные операции, и первое обращение к заснувшему устройству выполняется

относительно долго. Пользователи, не желающие возиться с настройкой АРМ (установкой параметров CMOS Setup и ОС), часто просто запрещают работу АРМ, в результате их компьютер всегда готов к работе, но, возможно, потребляет больше энергии, чем необходимо.

Вопрос энергосбережения стоит особо остро для мобильных (блокнотных) ПК при питании от аккумуляторов.

Помимо энергопотребления, АРМ снижает шум работающего компьютера: шумят вентиляторы блока питания, процессора и видеокарт; шумят винчестеры и приводы CD и DVD, особенно высокоскоростные. Если охлаждаемые устройства переходят в энергосберегающий режим, то можно снизить и скорость вращения вентиляторов, а следовательно, и их шум. Современные винчестеры позволяют регулировать уровень шума: появляется выбор между быстрой, но шумной работой и тихой, но менее производительной.

В плане управления потреблением различают следующие состояния устройств (в порядке «углубления сна»):

- *on* — активная (нормальная) работа, полное потребление, максимальная производительность;
- *standby* — отключение питания некоторых узлов, с возможностью быстрого (порядка секунды) перехода в активное состояние;
- *suspend* — более глубокое отключение (например, строчной развертки и накала трубки монитора), выход из которого (*resume*) требует единиц – десятков секунд;
- *off* — отключение питания всех узлов, кроме цепей, обеспечивающих последующее включение по команде.

Для ряда устройств (в том числе процессоров) применяют и иные названия состояния, например *sleep* (сон) и *deep sleep* (глубокий сон). Для других устройств применимо понятие *уровня активности (APM level)*, который выражается численно: 0 — минимальное потребление, 255 — максимальная активность. Кроме того, имеется состояние механического отключения, когда устройство обесточено механическим выключателем и никакой менеджер АРМ его уже не включит.

Конечно, самый тихий и холодный компьютер — выключенный, но его «пробуждение» (включение) требует загрузки ОС и приложений, которые, по мере технического прогресса, требуют все больше времени на разгрузку. В принципе, можно процесс загрузки обойти, для чего достаточно сохранить все содержимое ОЗУ, а также содержимое всех регистров процессора и внутренних регистров (буферов памяти) всех устройств, например, на жестком диске. После этого можно обесточить компьютер, а по включении быстро восстановить запомненное состояние и продолжить работу с точки останова. Такой способ «усыпления» называется «зимней спячкой» (*hibernate*), при

этом сохранение и восстановление состояния компьютера занимает всего десятки секунд. Приложения, работа которых была приостановлена, продолжают работу с того места, на котором были остановлены. В принципе, пользоваться таким способом выключения компьютера можно сколь угодно большое число раз, но на практике периодически приходится перезагружать ОС и приложения в обычном режиме. К этому вынуждают и сбои работы ОС и приложений, и накапливающиеся «отходы» памяти, которые не всегда могут быть использованы до перезагрузки. Для подстраховки перед усыплением компьютера рекомендуется все-таки явно сохранять пользовательские файлы на диске — риск невозможности восстановления состояния после включения хоть и невелик, но есть. Менее радикальный способ «усыпления», называемый *Standby*, сводится к остановке винчестеров, выключению дисплея, максимальному торможению процессора и всех остальных устройств. При этом состояние подсистем компьютера сохраняется, но на своих местах (данные остаются в памяти, регистрах процессора и всех устройств). Переход в нормальный режим происходит гораздо быстрее (время уходит только на раскрутку диска и на прогрев монитора, и то если он традиционный электронно-лучевой). Однако в таком состоянии компьютер все-таки потребляет заметную мощность, вентилятор настольного ПК продолжает работать. Надежность сохранения состояния получается ниже: провал (скачок) питания, толчки, способные нарушить контакт в модулях памяти или карт расширения, и прочие возмущения могут привести к потере состояния (впрочем, как и в нормальном рабочем режиме).

Система управления потреблением настраивается параметрами CMOS Setup (раздел Power Management), а также средствами современных ОС, включая Windows 9x/200x/XP. Основной аппаратной базой управления являются системная плата и BIOS, поддерживающие спецификацию ACPI (а прежде, APM). Возможность программного включения обеспечивает «дежурный» (standby) источник блока питания ATX; также важна поддержка APM и ACPI периферийными устройствами. Для быстрого запуска со стороны BIOS предусматривается сокращение времени выполнения начального теста POST — подробное тестирование может выполняться лишь при обнаружении проблем во время предыдущей загрузки. Время от включения до начала загрузки ОС стремятся сократить до нескольких секунд; правда, процесс может затянуться медленно запускаемыми дисками или необходимостью инициализации всего объема ОЗУ при обнаружении и коррекции ошибок (ECC). В BIOS должна быть возможность отключить визуализацию теста POST — тогда после включения ПК, если все в порядке, пользователь увидит заставку загружаемой им ОС.

## 7. ВНЕШНЯЯ ПАМЯТЬ

### 7.1. Жесткий диск

Жесткий диск предназначен для долговременного хранения больших объемов информации.

#### **Общие положения**

*Время доступа* (Access time) – период времени, необходимый накопителю на жестком диске для поиска и передачи данных в память или из памяти. Быстродействие накопителей на жестких магнитных дисках часто определяется временем доступа (выборки).

*Кластер* (Cluster) – наименьшая единица пространства, с которой работает ДОС в таблице расположения файлов. Обычно кластер состоит из одного или более секторов. Количество секторов зависит от типа диска. Многие жесткие диски имеют кластеры из четырех секторов или 2048 байтов. Поиск кластеров вместо отдельных секторов сокращает издержки ДОС по времени. Крупные кластеры обеспечивают более быструю работу накопителя, поскольку количество кластеров в таком случае меньше, но при этом хуже используется пространство (место) на диске, так как многие файлы могут оказаться меньше кластера и оставшиеся байты кластера не используются.

*Контроллер* (УУ) (Controller) – схемы, обычно расположенные на плате расширения, обеспечивающие управление работой накопителя на жестком диске, включая перемещение головки и считывание, запись данных.

*Головка накопителя* (Drive head) – механизм, который перемещается по поверхности жесткого диска и обеспечивает электромагнитную запись или считывание данных.

*Таблица размещения файлов* (FAT) (File Allocation Table (FAT)) – запись, формируемая ДОС, которая отслеживает размещение каждого файла на диске и то, какие сектора использованы, а какие – свободны для записи в них новых данных.

*Зазор магнитной головки* (Head gap) – расстояние между головкой накопителя и поверхностью диска.

*Чередование* (Interleave) – отношение между скоростью вращения диска и организацией секторов на диске. Обычно скорость вращения диска превышает способность компьютера получать данные с диска. К тому моменту, когда контроллер производит считы-

вание данных, следующий последовательный сектор уже проходит головку. Поэтому данные записываются на диск через один или два сектора. С помощью специального программного обеспечения при форматировании диска можно изменить порядок чередования.

*Логический диск* (Logical drive) – определенные части рабочей поверхности жесткого диска, которые рассматривают как отдельные накопители. Некоторые логические диски могут быть использованы для других операционных систем, таких как, например, UNIX.

*Парковка* (Park) – перемещение головок накопителя в определенную точку и фиксация их в неподвижном состоянии над неиспользуемыми частями диска, для того, чтобы свести к минимуму повреждения при сотрясении накопителя, когда головки ударяются о поверхность диска.

*Разбивка* (Partitioning) – операция разбивки жесткого диска на логические диски. Разбиваются все диски, хотя небольшие диски могут иметь только один раздел.

*Диск* (Platter) – металлический диск, покрытый магнитным материалом, на который записываются данные. Накопитель на жестких дисках имеет, как правило, более одного диска.

*RLL* (Run-length-limited) – кодирующая схема, используемая некоторыми контроллерами для увеличения количества секторов на дорожку для размещения большего количества данных.

*Время позиционирования* (Seek time) – время, необходимое головке для перемещения с дорожки, на которой она установлена, на какую-либо другую нужную дорожку.

*Дорожка* (Track) – концентрическое деление диска. Дорожки похожи на дорожки на пластинке. В отличие от дорожек пластинки, которые представляют собой непрерывную спираль, дорожки на диске имеют форму окружности. Дорожки в свою очередь делятся на кластеры и сектора.

*Сектор* (Sector) – деление дисковых дорожек, представляющее собой основную единицу размера, используемую накопителем. Секторы ДОС обычно содержат по 512 байтов.

*Цилиндр* (Cylinder) – дорожки, расположенные напротив друг друга на всех сторонах всех дисков.

*Время перехода с дорожки на дорожку* (Track-to-track seek time) – время, необходимое для перехода головки накопителя на соседнюю дорожку.



*Скорость передачи данных* (Transfer rate) – объем информации, передаваемый между диском и ЭВМ в единицу времени. В него входит и время поиска дорожки.

### **Устройство жесткого диска**

В основе функционирования винчестера лежит принцип магнитной записи/считывания сигналов на/с диск, покрытый магниточувствительным рабочим слоем. Каждая сторона диска, покрытая рабочим слоем, называется рабочей поверхностью.

При записи цифровые данные преобразуются в аналоговые электрические сигналы, создающие с помощью головки записи участки с различной намагниченностью, расположенные вдоль окружности по всей рабочей поверхности вращающегося диска (так называемые треки или дорожки). Размеры участков и расстояние между соседними дорожками определяют поверхностную плотность записи данных.

При чтении участки диска движутся под магнитной головкой и индуцируют в ней электрические сигналы, которые преобразуются в цифровые данные.

Жесткий диск (рис. 7.1) можно условно разделить на две составные части: механическую и электронную. Современные винчестеры устроены очень сложно. До 90% стоимости устройства составляет прецизионная механика.

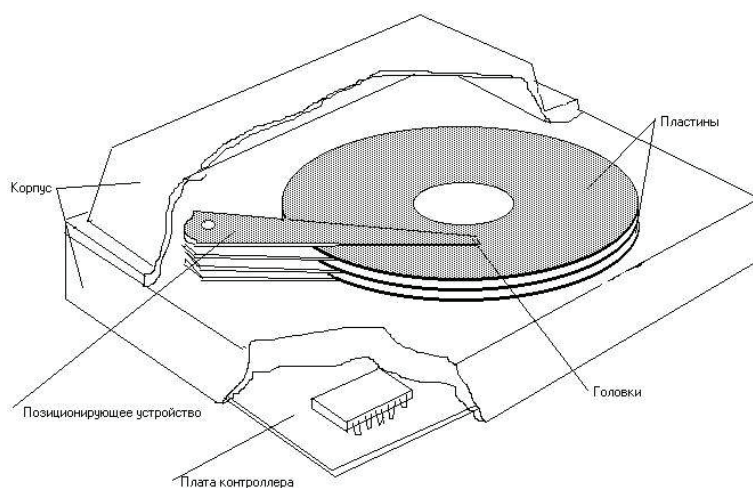


Рис. 7.1. Устройство НЖМД

Механическая, основная, часть выполняет самую низкоуровневую работу и состоит, прежде всего, из корпуса, шпинделя, электродвигателя, носителей информации – дисков, позиционера и блока магнитных головок.

Электронная часть представляет собой обычную плату с напаянными на ней

элементами, где, помимо резисторов, диодов и конденсаторов, находятся память, процессоры, «инженерный цилиндр» и т. д.

Привод выполняет важнейшую роль: он, будучи подключенным к шпинделю, на котором крепятся носители информации, раскручивает его и, соответственно, носители. К двигателю подключается источник тока с напряжением в 5 или 12 В, что позволяет современным моделям раскручивать шпиндель до 15 тыс. оборотов в минуту! Впрочем, этот показатель ограничен не возможностями двигателя – тут масса других проблем: шумовые, тепловые и, прежде всего, ограничения, связанные со считыванием информации с магнитных носителей. Шпиндель крепится к жесткому диску не напрямую, а через подшипники, которые сводят к минимуму трение в осях шпинделя, вследствие чего минимизируются шум и выделяющееся в результате трения тепло. В современных моделях используются гидравлические подшипники.



Рис. 7.2. Винчестер без защитного кожуха

В качестве хранителей данных выступают тонкие пластинки толщиной примерно 2 мм, на которые с обеих сторон нанесено магнитное покрытие, обладающее способностью менять свои свойства под воздействием электромагнитных излучений. Раньше они выполнялись из стекла, сейчас используются преимущественно пластики – это «дешево, надежно и практично», но притом еще и эстетично. В качестве магнитного покрытия сейчас, в основном, используют окись хрома – вещество, которое, в отличие от ранних материалов – окиси железа, сохраняет свои магнитные свойства значительно более длительный срок. Магнитное покрытие чрезвычайно чувствительно к физическим на него воздействиям, поэтому на каждой пластине есть так называемая «парковочная зона» – область пластины у шпинделя, где не содержится никакой информации и куда опускается после завершения работы считывающая и записывающая информацию магнитная головка. В современных жестких дисках, с целью повысить их вместительность, нередко используется несколько пластин.

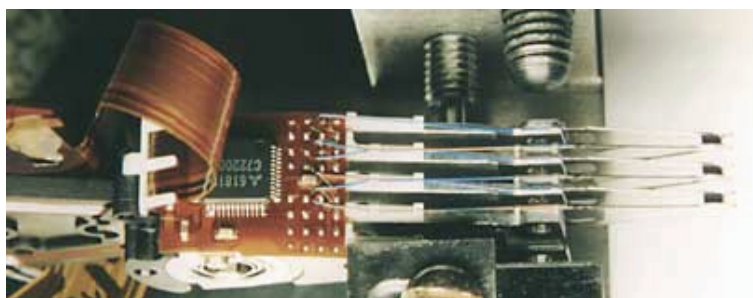


Рис. 7.3. Внешний вид блока головок

Головки (рис. 7.3) предназначены для чтения и записи информации на пластины и представляют собой совокупность замкнутых контуров, выполненных из тонкой проволоки в пластике, и излучателей электромагнитных колебаний.

Все это крепится на упругом крылышке, так называемом позиционере – алюминиевой «руке», способной перемещаться между краем пластины и шпинделем. Оно обеспечивает перемещение головок к нужному месту на поверхности пластин. Головки расположены с обеих сторон блина и подпружинены таким образом, что в свободном состоянии головка ложится на поверхность носителя. Шпиндель с блинами очень быстро вращаются вокруг своей оси, и вследствие этого вращающиеся пластины увлекают за собой потоки воздуха, благодаря

которым головка «летает» над магнитным покрытием блина на расстоянии 0,13 мкм от него. Следует пояснить, что магнитное вещество намагничивается неравномерно и при быстром вращении блинов относительно неподвижной головки в замкнутом контуре возникают электромагнитные колебания. Эти колебания, собственно, и представляют собой прочитанную информацию, они поступают в его электронную часть, где, после обработки, идут по шлейфу на материнскую плату.



Рис. 7.4. Головка в зоне «парковки»

Запись информации выполняется аналогичным путем, но в обратном порядке. Процессор устройства генерирует электромагнитные импульсы, поступающие на головку, производящую запись информации. Эти колебания излучаются в сторону магнитного покрытия, вследствие чего оно меняет свои магнитные свойства. При снижении скорости вращения ниже номинальной величины или выключения питания, процессор винчестера отводит головки ближе к шпинделю, в так называемую «парковочную зону» (рис. 7.4), где головки ложатся на поверхность пластины. «Парковка» необходима для сохранности головок и информации, находящейся на поверхности пластин. В случае если головка «упадет» на ту часть поверхности пластины, где содержится информация, могут запросто повредиться и данные, и сама головка. Именно поэтому резкое обесточивание диска в момент интенсивной работы может вызвать его повреждение. Головка попросту не успевает достичь «парковочной зоны» и падает на данные, вызывая их повреждение.

Пакет дисков с двигателем и блок головок размещаются в специальном герметичном металлическом корпусе со съемной крышкой,

который называется гермоблоком или камерой. Ее внутренний объем не изолирован от внешней среды – обязательно предусматривается возможность перетока воздуха снаружи в камеру и наоборот. Это необходимо для выравнивания давления внутри камеры с внешним давлением для предотвращения деформаций корпуса.

Остальная электроника винчестера менее уязвима и находится на отдельной плате за пределами гермоблока. По своей структуре она очень напоминает отдельный компьютер, т. к. среди основных компонент значатся: центральный процессор, ОЗУ (буфер диска), ПЗУ с программой управления, а также DSP (Digital Signal Processor), служащий для обработки считанных сигналов и подготовки записываемых.

Вся эта сложная электроника обеспечивает управление приводами головок и дисков. В современных моделях есть устройство для отключения винчестера при отсутствии запросов к нему.

Самый распространенный формфактор ширины диска, конечно же, 3,5 дюйма, но можно встретить 1,8 или 5,25 дюймовые модели.

### **Геометрия жесткого диска**

Пластины накопителя изготавливаются из металла или стекла и имеют с одной или обеих сторон магнитный слой, на который и происходит запись информации. Сторона пластины с нанесенным магнитным слоем называется рабочей поверхностью. Поверхности пластин тщательно отполированы и покрыты ферромагнитным слоем. Материал покрытия и количество слоев (магнитный слой может состоять из нескольких слоев разных материалов) может быть различным для разных накопителей. На каждую рабочую поверхность приходится по одной головке (на самом деле в современных накопителях для увеличения плотности записи применяются отдельные головки записи и чтения, изготовленные по различным технологиям). Поверхность пластины разбивается на тонкие концентрические кольцевые зоны, называемые дорожками (рис. 7.5). А каждая дорожка, в свою очередь, делится на несколько участков, получивших названия секторов.

Дорожки – это концентрические окружности (т. е. у всех у них центр находится в одной точке), вдоль которых располагается записанная на диске информация. Каждая дорожка имеет свой собственный номер. Нумерация производится по порядку, начиная с нулевой

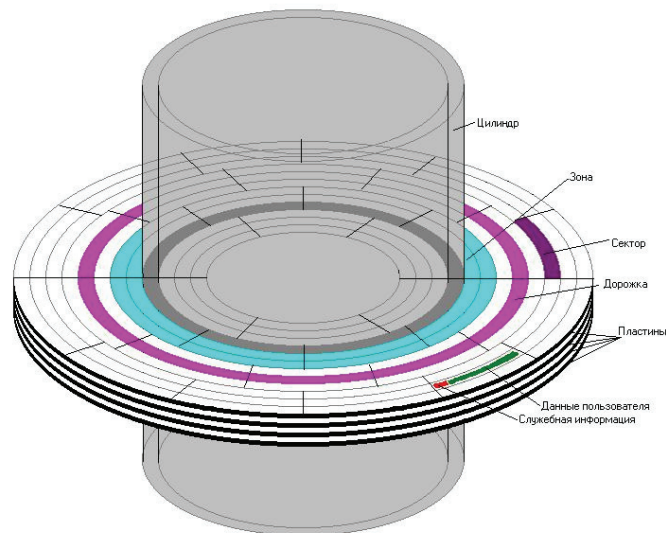


Рис.7.5. Логическая структура жесткого диска

дорожки по направлению от края пластины к шпинделю винчестера. Цилиндром называют совокупность двух дорожек с одинаковыми номерами, расположенных на противоположных сторонах пластины.

Сектор можно условно разделить на две области: *область данных* и *область служебной информации*. Служебная информация записывается на пластину один раз на заводе-изготовителе и в дальнейшем не подлежит изменению. Служебная область включает уникальный адрес сектора в накопителе, по которому его опознает контроллер при записи или считывании информации.

Область данных содержит полезную информацию, записываемую на накопитель. Эта область может быть многократно изменена в период эксплуатации. Объем области данных несколько превосходит информационную емкость сектора за счет дополнительной информации – для верификации и, возможно, исправления ошибок. Область данных сектора может быть обновлена только целиком. Т. е. на накопитель нельзя записать один или десять байт – только сектор целиком. Все головки перемещаются синхронно, и этот процесс занимает некоторое время. Совокупность дорожек на разных пластинах доступных одновременно при неизменном положении головок называется *цилиндром*. С точки зрения производительности дисковой системы целесообразно последовательные данные располагать в пределах одного цилиндра.



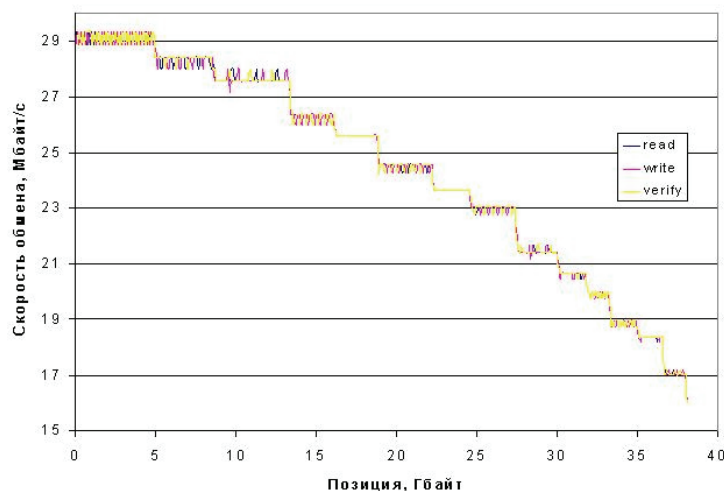


Рис. 7.6. Профили чтения, записи и верификации при зонной записи

Не менее 16% суммарной рабочей поверхности дисков отводится под служебную информацию, которая обеспечивает нормальную работу винчестера. В первую очередь, это *инженерная зона* (секторы конфигурации, таблицы дефектов, рабочие программы винчестера).

Оставшееся дисковое пространство делится на зоны (для большинства винчестеров — от 8 до 20) с различным числом секторов в каждой зоне. Не все секторы используются в качестве рабочих. Часть секторов являются запасными. При первоначальной разметке дисков на заводе-изготовителе производится проверка поверхности диска и информация об обнаруженных дефектных участках записывается в таблицу дефектов, которая размещается в инженерной зоне.

В процессе функционирования винчестера эта таблица используется для переназначения (переадресации) обращения к дефектным участкам (секторам) на обращение к хорошим секторам, которые как раз и размещаются на запасных дорожках. Ввиду важности служебной информации инженерная зона различных моделей накопителей может содержать от 2 до 6 копий.

### Логический диск

Область данных диска разбита на *кластеры* (cluster) — группы смежных секторов, называемые также единицами распределения про-

странства (allocation unit). Каждый кластер имеет свой номер; размер кластера (число секторов) выбирается кратным степени двойки в зависимости от объема диска и размера FAT. Файл на диске занимает целое число кластеров, от одного до всех кластеров, входящих в область данных. Если файл занимает более одного кластера, то все занятые кластеры организуются в *цепочку кластеров* (cluster chain). Количество файлов на диске не может превышать количества кластеров (элементов FAT). Обращения к файлам, занимающим цепочку смежных кластеров, выполняются гораздо быстрее, чем к файлам, у которых кластеры раскиданы по всему диску, — меньше времени тратится на позиционирование головок. *Файлы*, которые располагаются в цепочках из несмежных кластеров, называются *фрагментированными*. Соответственно, процедура наведения порядка на диске, повышающая производительность файловой системы ОС и компьютера в целом называется *дефрагментацией* диска. Специальные утилиты дефрагментации (SPEDDISK, DEFRAG) занимаются тем, что разрозненные фрагменты файлов собирают в единую, по возможности непрерывную цепь смежных кластеров.

Помимо свободных, занятых и плохих кластеров на дисках могут образовываться *потерянные кластеры* (lost clusters). Это отдельные кластеры или даже цепочки, помеченные как занятые, но не принадлежащие ни одному из файлов (на них нет ссылок из элементов каталога). Их происхождение легко объяснимо: если при записи нового файла происходит внезапное отключение питания или аппаратный сбой, может оказаться, что в FAT уже внесены изменения (элементы уже заняты), а в каталог новый элемент со ссылкой на начало цепочки не внесен. Эти «бесхозные» кластеры уже не могут использоваться ОС, они просто «съедают» доступное дисковое пространство. Поиском потерянных кластеров занимаются специальные утилиты, например NDD, SCANDISK. Найденные «бесхозные» цепочки они предлагают либо пометить как свободные кластеры, либо преобразовать в файлы (дать на них ссылку из корневого каталога). Эти файлы иногда содержат ценную информацию, которую можно использовать для восстановления пропавших данных, но чаще их просто удаляют, освобождая место на диске. В системе FAT могут встречаться и иные ошибки, например, пересечения цепочек кластеров. Ошибки в файловой системе обычно происходят из-за неисправностей в любом из компонентов тракта «память — контроллер — диск» или в связывающих их шинах. «Лечение» этих ошибок выполняют специальные ути-



литы (опять же NDD, SCANDISK), которые пользуются информацией копий FAT (на диске их, как правило, не менее двух) и элементов каталогов.

Есть и другие файловые системы, более сложные, защищенные и эффективные. К ним можно отнести, например, NTFS . Они построены иначе, используют другие механизмы распределения дискового пространства.

### **Скоростные характеристики жестких дисков**

Кроме объема представляют интерес и скоростные характеристики накопителей. Из них можно выделить две основные: среднее время доступа и скорость линейной передачи данных.

*Время доступа* (access time) – время с момента запроса небольшой порции данных (обращения к прерыванию BIOS) до момента их получения (возвращения из прерывания). Временем выполнения программы здесь можно пренебречь, в этом случае время доступа состоит из времени позиционирования (seek time), то есть нахождения нужной дорожки и среднего времени ожидания данных (latent time), т. е. времени на поворот диска, чтобы нужный сектор оказался под головкой чтения. Очевидно, что среднее время ожидания равно половине периода обращения диска: 5,56 мс при частоте вращения 5400 об/мин и 4,17 мс - при 7200 об/мин. Время позиционирования же складывается из времени необходимого на перемещение головки и времени для успокоения ее колебаний после перемещения. Так как требования к допустимой амплитуде колебаний при чтении и при записи отличаются, могут отличаться также средние времена доступа и позиционирования. К сожалению, не существует единой методики измерения этой величины, поэтому каждая фирма-производитель применяет собственную методику, выставляющую ее продукцию в наилучшем свете. Кроме того, для обозначения одинаковых величин разные фирмы зачастую используют различные термины. Чаще в спецификациях приводится время позиционирования, т. к. оно меньше.

Уменьшение времени доступа возможно за счет уменьшения диаметра пластин, что делает возможным как увеличение скорости вращения, так и уменьшения времени позиционирования. Однако этот подход ведет к радикальному уменьшению емкости накопителя.

Наиболее эффективным средством повышения производительности дисковой системы является кэширование, т. е. хранение в оперативной памяти наиболее часто используемых данных с жесткого

диска. Ведь для доступа к определенному байту, расположенному на диске, требуется около 15 мс, а к расположенному в оперативной памяти – порядка 0,1 мкс. Даже однокластерное (буфер 4 Кбайта) кэширование при построчном чтении текстового файла с длиной строки 80 символов приведет к уменьшению времени считывания в 50 раз. Увеличение объема буфера еще более ускорит этот процесс, поэтому, во-первых, сами накопители содержат буфер объемом, как правило, от 2 до 8 Мбайт, а во-вторых, кэширование производится на уровне ОС.

### **Интерфейс**

В течение ряда лет широко применяется параллельный интерфейс IDE.

В настоящее время разработан и используется последовательный вариант интерфейса IDE – Serial ATA. К контроллеру каждый накопитель подключается своим 7-жильным кабелем. Первый рубеж скорости – 150 Мбайт/с, второй – 300 Мбайт/с. Эти интерфейсы, несмотря на существенное аппаратное различие программно совместимы с существующим сейчас параллельным IDE.

### **RAID системы**

Первоначальное предназначение RAID – создание на базе нескольких винчестеров диска большого объема с увеличенной скоростью доступа. Но затем к двум основным целям добавилась третья – сохранение данных в случае отказа части оборудования. Именно эти три кита сделали RAID-массивы столь востребованными бизнесом и военными.

Сейчас достаточно хорошие диски имеют разумную стоимость, и RAID становится основным элементом современного сервера любого уровня (а часто и не только сервера). Тем не менее, когда мы говорим о RAID, то о дешевизне лучше сразу забыть.

### **Надежность винчестера**

Сложная конструкция винчестера имеет много «узких» мест, которые могут стать причиной возникновения различных неполадок. Они связаны с естественным износом механических узлов и старением электронных компонентов накопителей в силу их ограниченной надежности. Обычно декларируемое производителями среднее время безотказной работы от 0,5 до 1 млн часов (непрерывная работа винче-

стера от 57 до 114 лет) является характеристикой скорее теоретической, чем практической.

С течением времени намагниченность рабочего слоя постепенно ослабевает. Качество записи файлов пользователя практически не ухудшается, т. к. они обычно периодически перезаписываются, но служебные метки секторов, данные в инженерной зоне записываются изготовителем один раз и навсегда. Поэтому спустя несколько лет могут возникнуть проблемы с доступом к данным, вызванные магнитным затуханием служебной информации.

### **Технология S.M.A.R.T.**

S.M.A.R.T. (Self-Monitoring, Analysis and Reporting Technology «Технология Самодиагностики, Анализа и Отчета») позволяет отслеживать и, самое главное, предсказывать возникновение ошибок, связанных с функционированием жесткого диска, отсюда появляется возможность вовремя сделать резервную копию данных, тем самым избежать морального и материального ущерба от потери информации, ограничившись лишь покупкой нового диска.

## **7.2. Оптические диски - CD,DVD**

### **CD – компакт-диски**

Компакт-диски изготавливаются из поликарбоната толщиной 1,2 мм, покрытого тончайшим слоем алюминия (ранее использовалось золото) с защитным слоем из лака, на котором обычно печатается этикетка. Поэтому, вопреки распространенному мнению, компакт-диск никогда не следует класть вверх ногами этикеткой вниз т. к., отражающий алюминиевый слой, на котором и хранятся данные, снизу более защищён, как было сказано выше, 1,2- миллиметровым слоем поликарбоната.

Обычно компакт-диски имеют в диаметре 12 см и вмещают до 650 Мбайт информации (или 74 мин. аудио). Есть предположение, что разработчики рассчитывали объём так, чтобы на диске полностью поместилась девятая симфония Бетховена, длящаяся именно 74 минуты. Однако большое распространение получили диски объёмом 700 мегабайт (80 минут аудио). Встречаются и носители объёмом 800 мегабайт (90 минут) и даже больше, однако они могут не читаться на некоторых приводах компакт-дисков. Бывают также мини-CD,

диаметром 8 см, на которые вмещается около 140 Мб данных или 21 минута аудио.

Формат хранения данных на диске, известный как «красная Книга» («Red Book»), был разработан компанией Philips. В соответствии с ним на компакт-диск можно записывать звук в два канала с 16-битной импульсно-кодовой модуляцией (PCM) и частотой дискретизации 44,1 кГц. Благодаря коррекции ошибок с помощью кода Рида-Соломона, небольшие царапины не влияют на читаемость диска. Philips также владеет всеми правами на знак «Compact disk digital audio», который проставляется на дисках.

Информация на диске записывается в виде спиральной дорожки, так называемых питов (углублений), выдавленных на поликарбонатном слое (рис. 7.7.). Каждый пит имеет примерно 125 нм в глубину и

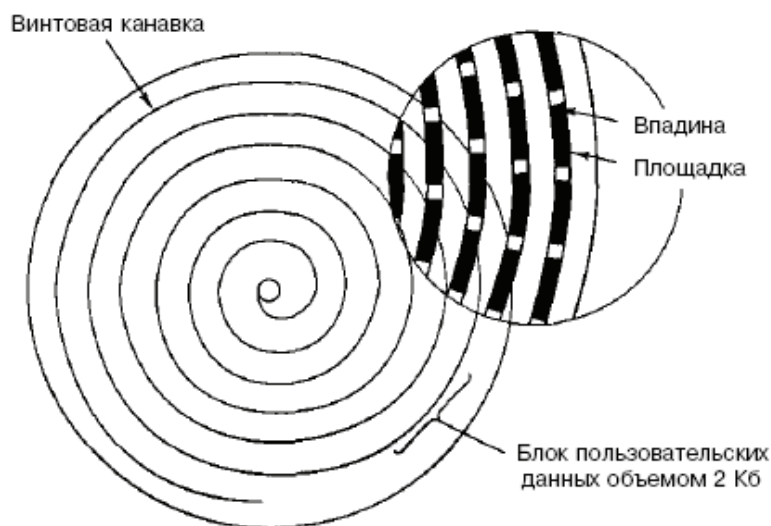


Рис. 7.7. Размещение информации на диске

500 нм в ширину. Длина пита варьируется от 850 нм до 3,5 мкм. Расстояние между соседними дорожками спирали 1,5 мкм. Данные с диска читаются при помощи лазерного луча с длиной волны 780 нм, который просвечивает поликарбонатный слой, отражается от алюминиевого и считывается фотодиодом. Луч лазера образует на отражающем слое пятно диаметром примерно 1,5 мкм. Так как диск читается с нижней стороны, каждый пит выглядит для лазера как возвышение. Места, где такие возвышения отсутствуют, называются площадками.

Свет от лазера, попадающий на площадку, отражается и улавливается фотодиодом. Если же свет попадает на возвышение, он испы-

тывает интерференцию со светом, отраженным от площадки вокруг возвышения и не отражается. Так происходит, потому что высота каждого возвышения равна четверти длины волны света лазера, что приводит к разнице в фазах в половину длины волны между светом, отраженным от площадки и светом, отраженным от возвышения.

Компакт-диски бывают штампованные на заводе, для однократной записи (CD-R), для многократной записи (CD-RW). Диски последних двух типов предназначены для записи в домашних условиях на специальных пишущих приводах для компакт-дисков.

Скорость чтения/записи CD указывается кратной 150 кБ/с, т. е. 48-скоростной привод обеспечивает максимальную скорость чтения/записи дисков равную  $48 \times 150 = 7200$  кБ/с (7,03 МБ/с).

Спецификация компакт-дисков не предусматривает никакого механизма защиты от копирования – диски можно свободно размножать и воспроизводить. Однако, начиная с 2002 года, различные западные звукозаписывающие компании начали предпринимать попытки создать компакт-диски, защищенные от копирования. Суть почти всех методов сводится к намеренному внесению ошибок в данные, записываемые на диск, так, чтобы на бытовом CD-плеере или музыкальном центре диск воспроизводился, а на компьютере – нет.

Обычные компакт-диски штампуются на заводах при помощи стеклянной матрицы с вытравленным на ней рисунком дорожек, которой прессуется металлический слой диска.

Процесс записи называется «прожигом» или «нарезкой» диска.

CD-RW является дальнейшим развитием записываемого лазерного компакт-диска CD-R, однако, в отличие от него, позволяет не только записывать информацию, но и многократно стирать уже записанные данные. CD-RW во многом похож на своего предшественника CD-R, но его записывающий слой изготавливается из специального сплава, который можно нагреванием приводить в два различных устойчивых агрегатных состояния – аморфное и кристаллическое. Этот сплав обычно изготавливается из серебра (Ag), индия (In), сурьмы (Sb) и теллура (Te). При записи/стирании луч лазера нагревает участок дорожки и переводит его в одно из устойчивых агрегатных состояний, которые характеризуются различной степенью прозрачности. Читающий луч лазера имеет меньшую мощность и не изменяет состояние записывающего слоя, а чередующиеся участки с различной прозрачностью формируют картину аналогичную питам и площадкам обычных штампованных CD.

Диски CD-RW позволяют перезаписывать информацию порядка 1000 раз. За исключением возможности стирать записанную информацию для пользователя работа с дисками CD-RW очень похожа на работу с однократно записываемыми CD-R. Данные записываются сессиями, можно дописывать новые файлы. С каждой новой сессией свободное место на диске уменьшается, и когда оно закончится, можно будет полностью стереть информацию со всего диска или ее часть, после чего он вновь будет доступен для записи новой.

### **Устройство CD-дисковод.**

Внешний вид дисководов изображен на рис. 7. 8. Типичный привод состоит из платы электроники, шпиндельного двигателя, системы оптической считывающей головки и системы загрузки диска. На плате электроники размещены все управляющие схемы привода, интерфейс с контроллером компьютера, разъемы интерфейса и выхода звукового сигнала. Большинство приводов использует одну плату электроники, однако в некоторых моделях отдельные схемы выносятся на вспомогательные небольшие платы.



Рис. 7. 8. Типичный CD-дисковод

Двигатель служит для приведения диска во вращение с постоянной или переменной линейной скоростью, сохранение постоянной линейной скорости требует изменения угловой скорости диска в зависимости от положения оптической головки. При поиске фрагментов диск может вращаться с большей скоростью, нежели при считывании, поэтому от шпиндельного двигателя требуется хорошая динамическая характеристика. Двигатель используется как для разгона, так и для торможения диска. На оси двигателя закреплена подставка, к которой после загрузки прижимается диск. Поверхность подставки обычно покрыта резиной или мягким пластиком для устранения проскальзывания диска. Прижим диска к подставке осуществляется при помощи шайбы, расположенной с другой стороны диска, подставка и шайба

содержат постоянные магниты, сила притяжения которых прижимает шайбу через диск к подставке.

Система оптической головки состоит из самой головки и системы ее перемещения. В головке размещены лазерный излучатель на основе инфракрасного лазерного светодиода, система фокусировки, фотоприемник и предварительный усилитель. Система фокусировки представляет собой подвижную линзу, приводимую в движение электромагнитной системой voice coil, сделанной по аналогии с подвижной системой громкоговорителя. Изменение напряженности магнитного поля вызывают перемещение линзы и перефокусировку лазерного луча. Благодаря малой инерционности, такая система эффективно отслеживает вертикальные биения диска даже при значительных скоростях вращения. Система перемещения головки имеет собственный приводной двигатель, приводящий в движение каретку с оптической головкой при помощи зубчатой либо червячной передачи. Для исключения люфта используется соединение с начальным напряжением: при червячной передаче – подпружиненные шарики, при зубчатой – подпружиненные в разные стороны пары шестерней.

Система загрузки диска выполняется в двух вариантах: с использованием специального футляра для диска (caddy), вставляемого в приемное отверстие привода, и с использованием выдвижного подноса (tray), на который кладется сам диск. В обоих случаях система содержит двигатель, приводящий в движение лоток или футляр, а также механизм перемещения рамы, на которой закреплена вся механическая система вместе со шпиндельным двигателем и приводом оптической головки. Все современные приводы используют второй способ загрузки. При использовании обычного лотка привод невозможно установить в иное положение, кроме горизонтального. В дисководы, допускающих монтаж в вертикальном положении, конструкция лотка предусматривает фиксаторы, удерживающие диск при выдвинутом лотке. Надо отметить, что такие фиксаторы имеют практически все дисководы, то есть в современных приводах предусмотрена возможность работы и в вертикальном положении. На передней панели дисковода располагаются кнопка Eject для загрузки/выгрузки диска, индикатор обращения к приводу и гнездо для подключения наушников с электронным или механическим регулятором громкости.

В ряде моделей добавлена кнопка Play/Next для запуска проигрывания звуковых дисков и перехода между звуковыми дорожками, а кнопка Eject при этом обычно используется для остановки проигры-

вании без выбрасывания диска. На некоторых моделях с механическим регулятором громкости, выполненным в виде ручки, проигрывание и переход осуществляются при нажатии на торец регулятора, но такое встречается редко. Большинство приводов также имеет на передней панели небольшое отверстие, предназначенное для аварийного извлечения диска в тех случаях, когда обычным способом это сделать невозможно, например, при выходе из строя привода лотка дисководов, при пропадании питания и т. п. В отверстие нужно вставить шпильку или распрямленную скрепку и аккуратно нажать, при этом снимается блокировка лотка и его можно выдвинуть вручную.

Скорость работы приводов измеряется относительно скорости вращения обычного музыкального CD-плеера (скорость обмена около 150 Кбит/с). Например, параметр 40х значит, что дисковод может вращать диск в 40 раз быстрее, чем это делает стандартный CD-плеер. Однако не стоит думать, что, скажем, 72-скоростной CD-ROM будет работать заметно быстрее, чем 40-скоростной. Из-за того, что физические параметры диска (неоднородность массы, эксцентриситет и тому подобное) стандартизированы для основной скорости вращения, на скоростях, больших 4х–6х, уже возникают значительные колебания диска, и надежность считывания, особенно для дисков низкого качества, может ухудшаться. Естественной на это реакцией дисковода является снижать скорости вращения диска (вплоть до 1х). При этом большинство дисководов после этого не могут возвращаться к максимальной скорости вплоть до смены диска, даже если некачественным был только небольшой его участок. На скоростях свыше 4000–5000 об/мин, надежное считывание становится практически невозможным, поэтому многие модели CD-ROM ограничивают верхний предел скорости вращения. При этом на внешних дорожках скорость передачи достигает номинальной скорости (например, 1800 Кбит/с), а по мере приближения к внутренним падает (в нашем примере приблизительно до 1200 – 1300 Кбит/с). В результате даже при существенно разной скорости вращения диска современные приводы достаточно высокого качества имеют в общем-то одинаковую скорость передачи данных.

Гораздо более важным параметром является время доступа. Оно у лучших моделей составляет около 80–60 мс, однако у дешевых может быть намного больше – вплоть до нескольких сотен мс. Но дешевые дисководы обладают такими особенностями, как плохое качество чтения диска (например, диски с дефектами, которые отлично чита-



ются на других приводах, могут не читаться вовсе или читаться на более низкой скорости), сильный шум, вибрация и прочее.

### **DVD-диск**

DVD (Digital Versatile Disc – цифровой многоцелевой диск или Digital Video Disk – цифровой видеодиск) – носитель информации в виде диска, внешне схожий с компакт-диском, однако имеющий возможность хранить больший объем информации за счет использования лазера с меньшей длиной волны, чем для обычных компакт дисков.

### **Техническая информация**

При создании стандарта DVD разработчики поставили цель существенно увеличить объем информации, вмещающейся на диск, при сохранении тех же геометрических размеров, что и у CD. Для этого был использован лазер с более короткой длиной волны (650 нм, что позволило разместить на одном слое 120-мм диска до 4,7 Гб данных), а число информационных слоев было увеличено до (максимально) четырех, причем считывание двух из них происходит с одной стороны диска, а оставшихся двух – с другой. Сами информационные слои, как и отражающий слой, располагаются в середине диска (на глубине 0,6 мм от каждой из сторон, т. е. технологически такой диск представляет собой два склеенных между собой диска толщиной 0,6 мм). Такая конструкция практически не оставляет места для нанесения на диск изображения, да и довольно дорога в производстве, поэтому на практике чаще всего встречаются диски с одним или двумя информационными слоями. Благодаря расположению информационных слоев в глубине диска DVD, по сравнению с CD, оказывается более устойчивым к поперечным повреждениям (царапинам), однако наличие технологической склейки делает его более уязвимым к продольным механическим воздействиям (изгибам).

DVD может иметь одну или две рабочие стороны и один или два рабочих слоя на каждой стороне. От их количества зависит вместительность диска:

- однослойные односторонние (DVD-5) вмещают 4,7 гб информации;
- двухслойные односторонние (DVD-9) вмещают 8,5 гб информации;
- однослойные двусторонние (DVD-10) вмещают 9,4 гб информации;

– двухслойные двусторонние (DVD-18) вмещают 17,4 гб информации.

Вместимость можно определить на глаз – нужно посмотреть, сколько рабочих (отражающих) сторон у диска и обратить внимание на их цвет. Двухслойные стороны обычно имеют золотой цвет, а однослойные – серебряный, как компакт-диск.

В DVD всегда используется файловая система UDF.

Скорость чтения/записи DVD указывается кратной 1350 Кб/с, т. е. 16-скоростной привод обеспечивает чтение/запись дисков со скоростью  $16 \times 1350 = 21\,600$  Кб/с (21,09 Мб/с).

### **Стандарт DVD+RW**

Носители DVD+RW, называемые также DVD Phase Change Rewritable (перезаписываемые DVD с изменяющейся фазой), наименее дорогие, самые простые в использовании и наиболее совместимые с существующими форматами. Этот стандарт был разработан компаниями Philips, Sony, Yamaha, Hewlett-Packard и Mitsubishi Chemical Corporation, Verbatim и Thompson, входящими в группу промышленного стандарта, которая называется DVD+RW Alliance. Фактически более 19 независимых поставщиков программного и аппаратного обеспечения объединили свои усилия в поддержке DVD+RW, что делает его наиболее поддерживаемым среди существующих перезаписываемых форматов DVD.

После завершения записи все физические параметры соответствуют требованиям спецификации DVD-ROM.

Потребительские и компьютерные приложения для цифровых универсальных дисков постоянно разрабатываются и совершенствуются, DVD+RW является форматом перезаписываемых DVD, который полностью удовлетворяет требованиям потребительского рынка DVD-Video и компьютерного рынка DVD-ROM. Технология DVD+RW очень похожа на CD-RW, и накопители DVD+RW могут читать как диски DVD-ROM, так и все форматы компакт-дисков, включая CD-R и CD-RW.

### **Blu-ray диск**

Blu-ray Disc (рис. 7. 9) или сокращенно BD (от англ. blue ray – голубой луч и disc – диск) – это следующее поколение формата оптических дисков – используемый для хранения видео высокой чет-

кости (с разрешением 1920×1080 точек) и данных с повышенной плотностью.

Стандарт Blu-ray был совместно разработан группой компаний по производству бытовой электроники и компьютеров во главе с Sony, которые вошли в Ассоциацию Blu-ray дисков (BDA). По сравнению со своим основным конкурентом, форматом HD DVD, Blu-ray имеет большую информационную емкость на слой – 25 вместо 15 Гб, но в тоже время он более дорогой в использовании и поддержке.

Blu-ray получил своё название от коротковолнового 405 нм «синего» (технически сине-фиолетового) лазера, который позволяет записывать и считывать намного больше данных, чем на DVD, который имеет те же физические объёмы, но использует для записи и воспроизведения красный лазер большей длины волны (650 нм).

### **Вариации и размеры**

Односторонний диск Blu-ray (BD) может хранить 23,3; 25, или 27 Гб, этого объёма достаточно для записи приблизительно четырёх часов видео высокой чёткости со звуком. Двухслойный диск может вместить 46,6; 50, или 54 Гб, что достаточно для записи на него приблизительно восьми часов HD-видео. Также в разработке находятся диски вместимостью 100 Гб и 200 Гб с использованием, соответственно, четырёх и восьми слоёв.

Стандарт BD-RE (перезаписываемые BD) доступен наравне с BD-R (записываемые) и BD-ROM форматами.



Рис.7. 9. Диск Blu-ray от Panasonic, объём 50 Гб, в картридже.

### **Технические детали**

В технологии Blu-ray для чтения и записи используется синевioletовый лазер с длиной волны 405 нм. Обычные DVD и CD используют красный и инфракрасный лазеры с длиной волны 650 нм и 780 нм соответственно.

Такое уменьшение позволило сузить дорожку вдвое по сравнению с обычным DVD-диском, до 0,32 микрон, и увеличить плотность записи данных.

Более короткая длина волны синевioletового лазера позволяет хранить больше информации на 12 см дисках того же размера, что и у CD/DVD. Эффективный «размер пятна», на котором лазер может сфокусироваться ограничен дифракцией и зависит от длины волны света и числовой апертуры линзы используемой для его фокусировки. Уменьшение длины волны, использование большей числовой апертуры (0,85 в сравнении с 0,6 для DVD), высококачественной двухлинзовой системы, а также уменьшение толщины защитного слоя в шесть раз (0,1 мм вместо 0,6 мм) предоставило возможность проведения более качественно и корректно операций чтения/записи. Это позволило записывать информацию в меньшие точки на диске, а значит, хранить больше информации в физической области диска, а также увеличить скорость считывания до 36 Мбит/с. В дополнение к оптическим улучшениям диски Blu-ray также имеют улучшенную технологию кодирования данных, позволяющую хранить больший объем информации.

### **Терабайтный оптический диск**

Сотрудники Лондонского имперского колледжа (Великобритания) разработали новую технологию записи информации на оптические носители, которая, вполне возможно, позволит создавать однослойные диски емкостью до 250 Гб.

В настоящее время информация записывается на оптические носители в виде микроскопических углублений – питов, каждый из которых соответствует одному биту. Луч лазера отражается от различных сегментов диска по-разному, что соответствует двоичным единице или нулю. Для повышения емкости производители используют не-

сколько слоев, а также сокращают расстояние между питами, увеличивая тем самым плотность записи.

Британские ученые пошли по несколько иному пути. Исследователи создали носитель, питы которого имеют наклоненные под десятью различными углами выступы. Отражаясь от этих выступов, луч лазера меняет поляризацию, что и регистрируется считывающим устройством. В результате, каждый пит может хранить не один, как в случае с традиционными дисками, а сразу десять бит данных. Естественно, при этом в десять раз повышается емкость носителя без увеличения его диаметра.

По заявлениям ученых, на однослойный диск, выполненный по новой технологии, которая получила название MODS (Multiplexed Optical Data Storage), теоретически можно записать до 250 Гб информации (примерно 118 часов видео). Соответственно, четырехслойный носитель будет иметь емкость около одного терабайта.

## 8. ВИДЕОСИСТЕМА.

### 8.1. Принципы вывода изображений.

Видеосистема РС ориентирована на растровый метод вывода изображения.

Растровый метод подразумевает, что некий рисующий инструмент, способный оставлять видимый след, сканирует всю поверхность, на которую выводится изображение. Траектория движения инструмента постоянна и не зависит от выводимого изображения, но инструмент может рисовать, а может и не рисовать отдельные точки траектории. Видимое изображение образуется оставляемыми им точками. В случае видеомонитора инструментом является модулированный луч (или три луча базисных цветов), построчно сканирующий экран и вызывающий свечение люминофора, нанесенного на внутреннюю поверхность экрана. Каждая строка раstra разбивается на некоторое количество точек — пикселов (pixel — сокращение от picture element — элемент изображения), засветкой каждой из которых по отдельности может управлять устройство, формирующее изображение (например, графическая карта).

Видеомонитор является растровым устройством вывода динамически изменяемых изображений. Его луч сканирует экран с частотой, которая не должна позволять глазу видеть мерцание изображения. Матричные дисплеи также относятся к растровым устройствам. Растровыми устройствами вывода статических изображений являются принтеры, в которых сканирование листа производится однократно (хотя возможны и многократные проходы).

Альтернатива растровым устройствам — векторные устройства вывода изображений. В этих устройствах инструмент прорисовывает только изображаемые фигуры и его траектория движения определяется выводимым изображением. Изображение состоит из графических примитивов, которыми могут быть отрезки прямых — векторы (откуда и название метода вывода), дуги, окружности. К векторным устройствам вывода статических изображений относятся перьевые плоттеры. Существовали (а может, где-то и сейчас используются) и векторные мониторы, однако ввиду сложности построения системы управления лучом, обеспечивающей быстрое и точное движение луча по сложной траектории, не получили распространения.

Рассмотрим растровую систему вывода изображений, подразумеваемая в качестве оконечного устройства монитор с электронно-лучевой трубкой – ЭЛТ (CRT, Cathode Ray Terminal, дословно – монитор на катодно-лучевой трубке). Сканирование экрана модулированным лучом обеспечивается генераторами горизонтальной и вертикальной разверток монитора. Луч оставляет след только во время прямого хода по строке (слева направо). Строка разбивается на некоторое количество точек, каждая из которых может иметь независимое от других состояние (яркость и цвет). Для ЭЛТ-монитора это разбиение условно, в матричных дисплеях пикселы являются физическими. На обратном ходе по строке луч принудительно гасится. Следующая строка прорисовывается параллельно предыдущей, но с некоторым вертикальным смещением (вниз), и так происходит сканирование до окончания кадра — достижения правого нижнего угла экрана. Во время обратного хода луча по вертикали, за время которого генератор горизонтальной развертки успеет сделать несколько строчных циклов, луч также принудительно гасится. В следующем кадре сканирование может производиться по-разному. В системах с прогрессивной (progressive), или не чересстрочной (Non-Interlaced), разверткой луч идет по тем же самым строкам (рис. 8.1, а). В системах с чересстрочной разверткой (Interlaced) луч идет по строкам, смещенным по вертикали на половину шага строки (рис. 8.1, б). Таким образом, всю поверхность экрана луч проходит за два цикла кадровой развертки, называемых полукадрами. Чересстрочная развертка позволяет почти вдвое снизить частоту горизонтальной (строчной) развертки, а следовательно, и темп вывода точек изображения.

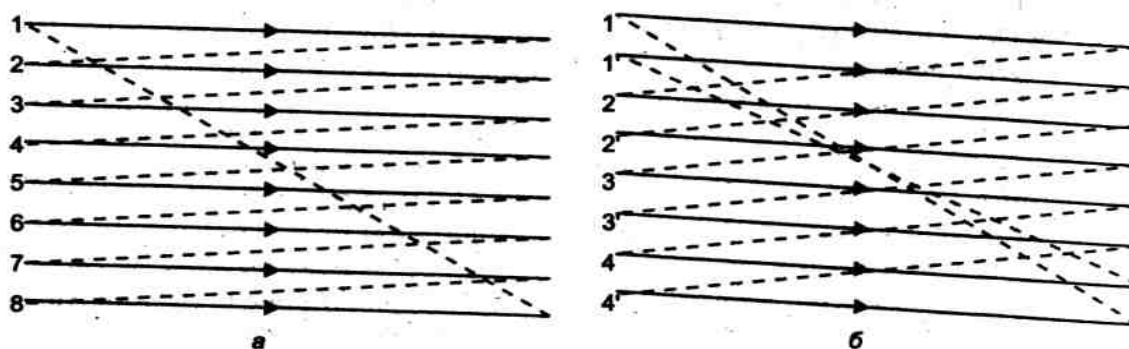


Рис. 8.1. Сканирование экрана:  
а — прогрессивная развертка; б — чересстрочная развертка

Как известно, глаз является инерционным органом зрения, он воспринимает изменение яркости или освещенности только до какой-то определенной частоты. Существует понятие критической частоты слияния мельканий (КЧСМ), которую измеряют так: человек смотрит неподвижно на некоторый безынерционный источник света (например, светодиод), который вспыхивает и гаснет с плавно возрастающей частотой. Сначала человек воспринимает вспышки по отдельности, с повышением частоты он видит уже только мерцание, а начиная с некоторой частоты мерцания для него сливаются в ровный свет. Эта частота и называется критической, и у разных людей она может находиться в пределах примерно 40–60 Гц. Наблюдение мерцающих объектов раздражает и утомляет зрительную систему, поэтому частота кадров (прорисовки экрана) должна быть по крайней мере, не ниже значения КЧСМ. Таким образом, мы получили ориентировочное значение минимальной частоты кадров, равное 50 Гц (эта компромиссная частота применяется во многих телевизионных системах). Вполне очевидно, что для получения качественного изображения экран должен иметь как можно больше точек матрицы разложения, т. е. строк в кадре и точек на строке. Возьмем режим  $800 \cdot 600$  (600 строк по 800 точек). За один период прогрессивной кадровой развертки луч должен успеть прочертить 600 видимых строк, да еще некоторое количество невидимых строк (50) он прочертит на обратном ходу. Получается, что частота строк должна составить  $50 \text{ Гц} \cdot (600 + 50) = 32,5 \text{ кГц}$ : (вроде и не так уж много). Этой частоте соответствует период около 30 мкс ( $1/32,5$ ), из которого на прямой ход по строке остается около 25 мкс. За это время необходимо вывести 800 точек строки, так что на каждую точку отводится  $25/800 = 0,03 \text{ мкс}$ . Это соответствует частоте вывода точек 30 МГц, а для электронных схем такая частота считается высокой. Поскольку соседние точки выводимого изображения в принципе друг с другом не связаны, то полоса частот сигнала, модулирующего интенсивность луча, должна быть несколько выше этого значения (примерно на 25 %). Такую широкую полосу пропускания должен обеспечивать весь видеотракт: видеоусилители модуляторов лучей, сигнальные линии интерфейсного кабеля, и наконец (вернее, сначала), такой широкополосный сигнал должен сформировать графический адаптер. На всех этих стадиях высокие частоты порождают технические проблемы, если реальная полоса пропускания в этом тракте окажется уже, четкого изображения получить не удастся — переходы будут размыты. Если же частотная характеристика тракта в



требуемом диапазоне будет неравномерной, появятся специфические искажения цветов или яркости около границ отображаемых объектов (справа от граничной линии). Понятно, что с технической точки зрения есть стимулы снижать требуемую верхнюю границу полосы частот видеотракта. При чересстрочной развертке за каждый полукадр сканируется только половина строк разложения (четные в одном полукадре и нечетные в другом), следовательно, строчная частота уменьшается, а длительность прохода видимой части строки увеличивается примерно вдвое. Таким образом, при заданных условиях (разрешении экрана и ограничении минимальной кадровой частоты) чересстрочная развертка позволяет снизить требуемую полосу пропускания вдвое.

Как видно из приведенных выкладок, частотные параметры видеосистемы определяются исходя из желаемой частоты кадров, разрешения экрана и режима развертки. Заботясь о зрении пользователя, частоту кадров стремятся повышать. При низкой частоте экран начинает мерцать, что особо заметно на больших белых полях изображения. Разрешение экрана стремятся увеличивать — чем оно выше, тем больше информации можно уместить на экране. Поскольку размер экрана постоянно увеличивается, потребность в разрешении, скажем, 1600 x 1200 вполне реальна. Но для этого, как явствует из сказанного ранее, уже требуется полоса 120 МГц (а кадровая частота 50 Гц — это отнюдь не идеал). В популярном для 17-дюймовых мониторов режиме 1280 · 1024 при частоте 85 Гц частота пикселей составляет 152 МГц (частота строк — 91,1 кГц).

Применение чересстрочной развертки годится лишь как вынужденная мера, поскольку имеет свои специфические неприятные «видеоэффекты»: если выводится тонкая (в одну строку точек) горизонтальная линия, она заметно мерцает. Это и понятно, ведь прорисовывается она только в одном из полукадров — следовательно, с половинной кадровой частотой. Если изображение потолще (один и тот же элемент имеет точки в соседних строках), его мерцание оказывается почти незаметным. Итак, лучше когда, частота кадров выше, разрешение (по вертикали и горизонтали) больше, развертка не чересстрочная.

Забегая немного вперед, заметим, что чем выше частота развертки, тем ниже производительность графической системы при построении изображений. С точки зрения пользователя привлекательность чересстрочного режима развертки заключается в цене устройств

— поскольку для прогрессивной развертки требуется более высокое качество компонентов всего видеотракта, построочная развертка с высокой частотой кадров в режимах высокого разрешения может оказаться дорогим удовольствием. Но для профессиональной работы с текстом, графического дизайна и других видов деятельности ухудшение зрения обойдется дороже. Чересстрочная развертка широко применяется в телевидении, видеосигнал там проходит через радиоканал, с шириной полосы которого всегда имеются проблемы. В современных мониторах и графических адаптерах, применяемых в РС, используются оба режима развертки с различными значениями частоты кадров. Естественно, что работать они должны в согласованных режимах.

Рассмотрев работу оконечного устройства (монитора), обсудим способы формирования изображения в графическом адаптере. Итак, у нас имеется матрица точек экрана, образованная горизонтальными строками раstra (номер строки — вертикальная координата матрицы) и точками разложения строки (номер точки в строке — горизонтальная координата матрицы). Матрица сканируется построочным или чересстрочным образом, и во время прямого хода луча по видимым строкам графический адаптер должен формировать сигналы управления яркостью базисных цветов монитора (или одного сигнала яркости в монохромном варианте). За это время последовательно (и синхронно с ходом луча) должна выводиться информация о яркости и цвете всех точек данной строки. Синхронизация обеспечивается формированием горизонтальных и вертикальных синхроимпульсов. Таким образом, графический адаптер является задающим устройством, а монитор со своими генераторами разверток должен вписаться в заданные параметры синхронизации.

Существует два основных *режима вывода информации* — графический и символьный (текстовый). Первые дисплейные адаптеры из-за технических ограничений на доступный объем памяти адаптера работали в символьном режиме. Современные адаптеры в основном работают в графическом режиме, текстовый режим используется только до загрузки ОС.

В графическом режиме имеется возможность индивидуального управления свечением каждой точки экрана монитора независимо от состояния остальных. В графическом режиме каждой точке экрана — пикселу — соответствует ячейка специальной памяти, которая сканируется схемами адаптера синхронно с движением луча монитора. Точнее, физически движение луча вторично, так как монитор можно и

не подключать, а графический адаптер все равно будет сканировать память, но логически вся конструкция строится исходя именно из поведения монитора. Эта постоянно циклически сканируемая (с кадровой частотой) память называется видеопамятью (video memory). Процесс постоянного сканирования видеопамяти называется регенерацией изображения.

В символьном, или текстовом, режиме формирование изображения происходит несколько иначе. Если в графическом режиме каждой точке экрана соответствует своя ячейка видеопамяти, то в текстовом режиме ячейка видеопамяти хранит информацию о символе, занимающем на экране знакоместо определенного формата. Знакоместо представляет собой матрицу точек, в которой может быть отображен один из символов определенного набора. Здесь умышленно применяется слово «точка», а не «пиксел», поскольку пиксел является сознательно используемым элементом изображения, в то время как точки разложения символа в общем случае программиста не интересуют. В ячейке видеопамяти хранятся код символа, определяющий его индекс в таблице символов, и атрибуты символа, определяющие способ его отображения. К атрибутам относятся цвет фона, цвет символа, инверсия, мигание и подчеркивание символа.

В текстовом режиме экран организуется в виде матрицы знакомест, образованной горизонтальными линиями (Line, LIN) и вертикальными колонками (Column, COL). Этой матрице соответствует аналогичным образом организованная видеопамять. Адаптер, работающий в текстовом режиме, имеет дополнительный блок — знакогенератор. Во время сканирования экрана выборка данных из очередной ячейки видеопамяти происходит при подходе к соответствующему знакоместу, причем одна и та же ячейка видеопамяти выбирается при проходе по всем строкам раstra, образующим линию знакомест. Считанные данные попадают в знакогенератор, который вырабатывает построчную развертку соответствующего символа — его изображение на экране. Знакогенератор представляет собой запоминающее устройство — ОЗУ или ПЗУ.

## 8.2. Видеоадаптер

Схема видеоадаптера изображена на рис.8.2.

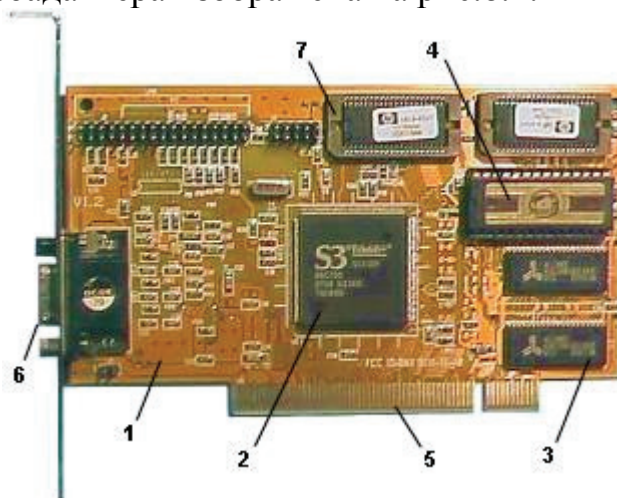


Рис. 8.2. Схема видеоадаптера

1 — монтажная печатная плата с элементами крепления; 2 — графический контроллер; 3 — видеопамять; 4 — система ввода/вывода устройства (BIOS); 5 — разъем для подключения адаптера к системной шине; 6 — разъем подключения монитора; 7 — разъемы расширения видеопамяти

Рассмотрим подробнее основные элементы.

*Графический контроллер* — устройство, которое отвечает за обмен данными между CPU и видеопамятью, регенерацию ее содержимого, и обработку запросов центрального процессора. Для исключения конфликтов при обращении к памяти со стороны видеоконтроллера и центрального процессора первый имеет отдельный буфер, который в свободное от обращений ЦП время заполняется данными из видеопамяти. Если конфликта избежать не удастся — видеоконтроллеру придется задерживать обращение ЦП к видеопамяти, что снижает производительность системы. Для исключения подобных конфликтов в ряде карт применялась так называемая двухпортовая память, допускающая одновременные обращения со стороны двух устройств.

*Последовательный преобразователь* выбирает данные из памяти и преобразует их в поток битов.

*Синхронизатор* — обеспечивает синхронную работу всех узлов адаптера, задает временные параметры и управляет доступом CPU к видеопамяти.

*Цифро-аналоговый преобразователь* служит для преобразования результирующего потока данных, формируемого видеоконтроллером, в уровни интенсивности цвета, подаваемые на монитор. Все современные мониторы используют аналоговый видеосигнал, поэтому возможный диапазон цветности изображения ограничен только соотношениями целесообразности. Большинство ЦАП имеют разрядность по 8 бит на каждый из трех каналов основных цветов (красный, синий, зеленый — RGB), по 256 уровней яркости на каждый цвет, что в сумме дает 16,7 млн. цветов. Обычно ЦАП совмещен на одном кристалле с видеоконтроллером.

*Видеопамять* используется как буфер видеоконтроллера для промежуточного хранения и модификации изображения.

Связь между графическим адаптером и ядром компьютера осуществляется посредством шины PCI-Express 16x.

Одним из решений для ускорения работы графической системы стало применение технологии MMX (Multi Media Extension), разработанной фирмой Intel. Дальнейшим развитием расширения возможностей графики, интегрированной в центральный процессор, явилась разработка фирмой AMD технологии 3Dnow!, а затем разработка фирмой Intel технологии SSE (Streaming SIMD Extensions).

Для ускорения разработки 3D-приложений используются специализированные прикладные программные (графические) библиотеки. Эти библиотеки могут быть как стандартными (разработанными лидерами 3D-индустрии), так и фирменными (разработанными производителями 3D-ускорителей). Функции библиотеки доступны через соответствующий API (Application programming interface) – программный интерфейс разработчика. Можно приближенно сказать, что API – язык описания трехмерной графики. Соответственно, каждое 3D-приложение написано с использованием некоторого API и соответственно будет работать в системе только в том случае, если видеочип поддерживает соответствующий API. От самого API во многом зависит качество и производительность работы видеоадаптера.

### **8.3. Мониторы.**

#### **CRT мониторы**

В основе CRT (Cathode Ray Tube) мониторов лежит электронно-лучевая трубка (ЭЛТ). Используемая в этом типе мониторов технология изобретена много лет назад и первоначально создавалась в качест-

ве специального инструментария для измерения переменного тока, проще говоря, для осциллографа. На рис. 8.3 показан разрез типичного электронно-лучевого монитора.

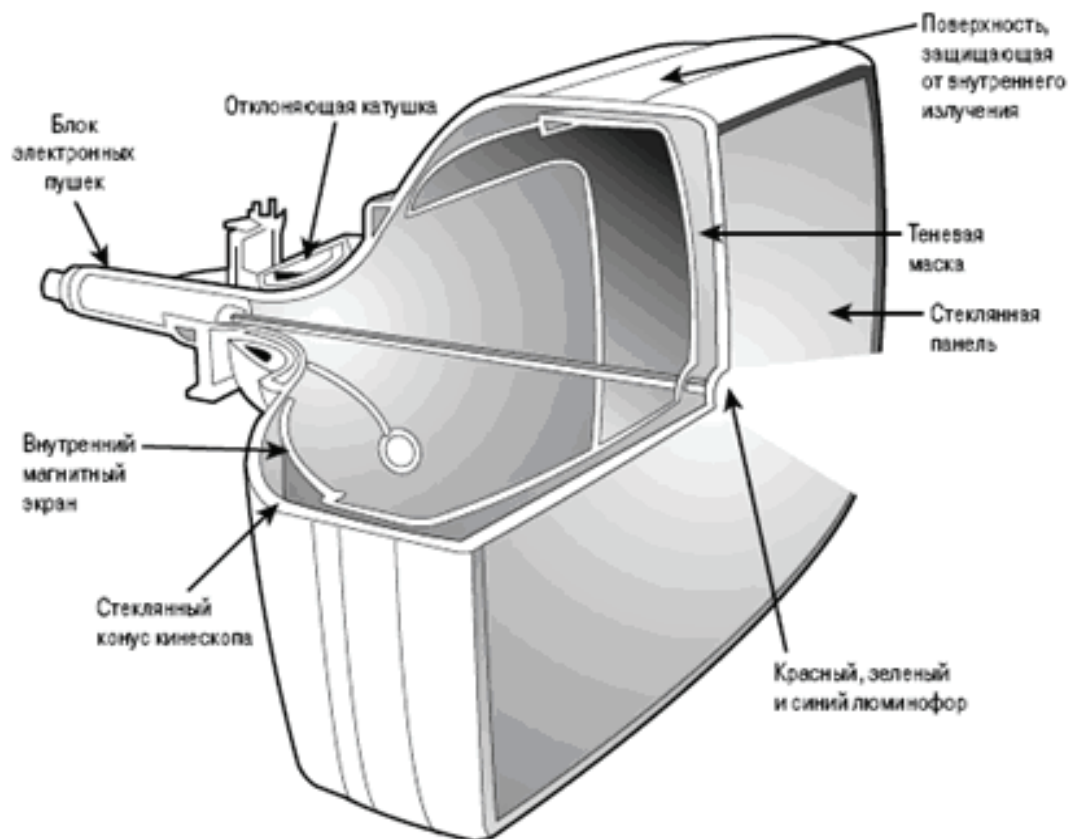


Рис. 8.3. Типичный электронно-лучевой монитор

Рассмотрим принципы работы CRT мониторов. CRT или ЭЛТ монитор имеет стеклянную трубку, внутри которой находится вакуум. С фронтальной стороны внутренняя часть стекла трубки покрыта люминофором (Luminofor). В качестве люминофоров для цветных ЭЛТ используются довольно сложные составы на основе редкоземельных металлов иттрия, эрбия и т. п. Люминофор – это вещество, которое испускает свет при бомбардировке его заряженными частицами. Для создания изображения в CRT мониторе используется электронная пушка, которая испускает поток электронов сквозь металлическую маску или решетку на внутреннюю поверхность стеклянного экрана монитора, которая покрыта разноцветными люминофорными точками. Поток электронов на пути к фронтальной части трубки проходит

через модулятор интенсивности и ускоряющую систему, работающие по принципу разности потенциалов. В результате электроны приобретают большую энергию, часть из которой расходуется на свечение люминофора. Электроны попадают на люминофорный слой, после чего энергия электронов преобразуется в свет, т. е. поток электронов заставляет точки люминофора светиться. Эти светящиеся точки люминофора формируют изображение на мониторе. Как правило, в цветном CRT мониторе используется три электронные пушки, в отличие от одной пушки, применяемой в монохромных мониторах.

Люминофорный слой, покрывающий фронтальную часть электронно-лучевой трубки, состоит из очень маленьких элементов (на столько маленьких, что человеческий глаз их не всегда может различить). Эти люминофорные элементы воспроизводят основные цвета, фактически имеются три типа разноцветных частиц, чьи цвета соответствуют основным цветам RGB.

Люминофор начинает светиться под воздействием ускоренных электронов, которые создаются тремя электронными пушками. Каждая из трех пушек соответствует одному из основных цветов и посылает пучок электронов на различные частицы люминофора, чье свечение основными цветами с различной интенсивностью комбинируется и в результате формируется изображение с требуемым цветом. Например, если активировать красную, зеленую и синюю люминофорные частицы, то их комбинация сформирует белый цвет.

Для управления электронно-лучевой трубкой необходима управляющая электроника, качество которой во многом определяет и качество монитора.

### **Различные типы CRT – трубок**

#### **Теневая маска (Shadow Mask)**

В данной технологии цветной элемент состоит из трех зерен, расположенных в вершинах правильного треугольника. Резкость изображения определяется расстоянием между геометрическими центрами соседних элементов.

Основное достоинство дельтовидной технологии – точная четкая картинка, диагональные линии не имеют зазубрин. Основным же недостатком считается большое расстояние между зернами, отчего такие трубки имеют не слишком насыщенный цвет.

*Теневая маска (shadow mask)* – это самый распространенный тип масок для CRT-мониторов (рис. 8.4). Теневая маска состоит из ме-

таллической сетки перед частью стеклянной трубки с люминофорным слоем. Как правило, большинство современных теневых масок изготавливают из инвара (invar, сплав железа и никеля). Отверстия в металлической сетке работают, как прицел (хотя и не точный), именно этим обеспечивается то, что электронный луч попадает только на требуемые люминофорные элементы, и только в определенных областях.

Минимальное расстояние между люминофорными элементами одинакового цвета называется dot pitch (шаг точки) и является важным параметром качества изображения. Шаг точки обычно измеряется мм. Чем меньше значение шага точки, тем выше качество воспроизводимого на мониторе изображения.

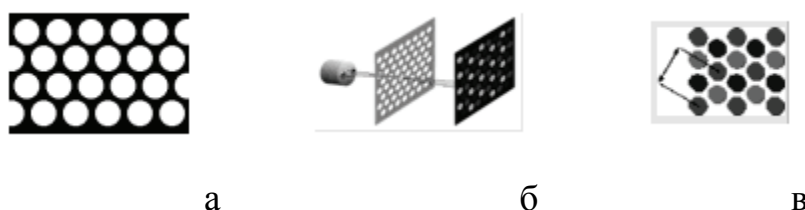


Рис.8.4. Теневая маска (Shadow Mask):

а – маска; б – прохождение луча через маску; в – расположение пикселей на экране монитора

### **Щелевая маска (slot mask)**

*Щелевая маска* изображена на рис.8.5. В данном случае люминофорные элементы расположены в вертикальных эллиптических ячейках, а маска сделана из вертикальных линий. Преимущества щелевой маски в передаче более насыщенного цвета, нежели при использовании теневой маски, за счет большей площади светящегося люминофора, однако такая маска несколько проигрывает в четкости наклонных линий.



Рис. 8.5. Щелевая маска (slot mask):

а – маска; б – прохождение луча через маску



### Апертурная решетка(Aperture Grill)

Есть и еще один вид трубок, в которых используется «Aperture Grill» (апертурная, или теневая решетка).

Апертурная решетка не включает в себя металлическую решетку с отверстиями, как в случае с теневой маской, а имеет решетку из вертикальных линий. Вместо точек с люминофорными элементами трех основных цветов апертурная решетка содержит серию нитей, состоящих из люминофорных элементов, выстроенных в виде вертикальных полос трех основных цветов.



Рис.8.6. Апертурная решетка (показан размер шага)

Такая система обеспечивает высокую контрастность изображения и хорошую насыщенность цветов, что вместе обеспечивает высокое качество мониторов с трубками на основе этой технологии.

Как у любой другой технологии, у апертурной решетки есть свои достоинства и недостатки. Основное достоинство этой технологии – великолепные по насыщенности цвета, кроме того, экран таких мониторов гораздо менее выпуклый, чем при применении теневых масок и щелевых масок, так как в принципе не имеет причин быть искривлен по вертикали. Основной же недостаток – ступенчатость и нечеткость диагональных линий, отчего проектировщики не любят мониторы с применением апертурной решетки.

Нельзя напрямую сравнивать размер шага для трубок разных типов: шаг точек (или триад) трубки с теневой маской измеряется по диагонали, в то время как шаг апертурной решетки, иначе называемый горизонтальным шагом точек, - по горизонтали. Поэтому при одинаковом шаге точек трубка с теневой маской имеет большую плотность точек, чем трубка с апертурной решеткой. Для примера: 0.25 мм strip pitch (для апертурной решетки) приблизительно эквивалентно 0.27 мм dot pitch.

Оба типа трубок имеют свои преимущества и своих сторонников. Трубки с теневой маской дают более точное и детализированное изображение, поскольку свет проходит через отверстия в маске с четкими краями. Поэтому мониторы с такими CRT хорошо использовать

при интенсивной и длительной работе с текстами и мелкими элементами графики. Трубки с апертурной решеткой имеют более ажурную маску, она меньше заслоняет экран, и позволяет получить более яркое, контрастное изображение в насыщенных цветах. Мониторы с такими трубками хорошо подходят для настольных издательских систем и других приложений, ориентированных на работу с цветными изображениями.

Кроме электронно-лучевой трубки внутри монитора есть еще и управляющая электроника, которая обрабатывает сигнал, поступающий напрямую от видеокарты РС. Эта электроника должна оптимизировать усиление сигнала и управлять работой электронных пушек, которые инициируют свечение люминофора, создающего изображение на экране.

Выводимое на экране монитора изображение выглядит стабильным, хотя на самом деле таковым не является. Изображение на экране воспроизводится в результате процесса, в ходе которого свечение люминофорных элементов инициируется электронным лучом, проходящим последовательно по строкам слева направо и сверху вниз на экране монитора. Этот процесс происходит очень быстро, поэтому нам кажется, что экран светится постоянно. В сетчатке глаз изображение хранится около  $1/20$  секунды. Это означает, что если электронный луч будет двигаться по экрану медленно, можно видеть это движение как отдельную движущуюся яркую точку, но когда луч начинает двигаться, быстро прочерчивая на экране строку хотя бы 20 раз в секунду, наши глаза не увидят движущейся точки, а увидят лишь равномерную линию на экране. Если теперь заставить луч последовательно пробежать по многим горизонтальным линиям сверху вниз за время меньшее  $1/25$  секунды, мы увидим равномерно освещенный экран с небольшим мерцанием. Движение самого луча будет происходить настолько быстро, что глаз не будет в состоянии его заметить. Чем быстрее электронный луч проходит по всему экрану, тем меньше будет заметно и мерцание картинки. Считается, что такое мерцание становится практически незаметным при частоте повторения кадров (проходов луча по всем элементам изображения) примерно 75 в секунду. Однако эта величина в некоторой степени зависит от размера монитора. Дело в том, что периферийные области сетчатки глаза содержат светочувствительные элементы с меньшей инерционностью. Поэтому мерцание мониторов с большими углами обзора становится заметным при больших частотах кадров.

Способность управляющей электроники формировать на экране мелкие элементы изображения зависит от ширины полосы пропускания (bandwidth). Ширина полосы пропускания монитора пропорциональна числу пикселей, из которых формирует изображение видеокарта.

**Основные параметры монитора** – размер, разрешение и частота обновления.

В случае с мониторами, *размер* – один из ключевых параметров. Под размером обычно понимают размер диагонали монитора. Монитор требует пространства для своей установки, а пользователь хочет комфортно работать с требуемым разрешением. Кроме этого, необходимо, чтобы монитор поддерживал приемлемую частоту регенерации или обновления экрана (refresh rate). При этом все три параметра – размер (size), разрешение (resolution) и частота регенерации (refresh rate) – должны всегда рассматриваться вместе, если вы хотите убедиться в качестве монитора, потому что все эти параметры жестко связаны между собой, и их значения должны соответствовать друг другу.

*Разрешение* монитора (или разрешающая способность) связана с размером отображаемого изображения и выражается в количестве точек по ширине (по горизонтали) и высоте (по вертикали) отображаемого изображения. Например, если говорят, что монитор имеет разрешение 640·480, это означает, что изображение состоит из  $640 \cdot 480 = 307200$  точек в прямоугольнике, чьи стороны соответствуют 640 точкам по ширине и 480 точкам по высоте. Это объясняет, почему более высокое разрешение соответствует отображению более содержательного (детального) изображения на экране. Понятно, что разрешение должно соответствовать размеру монитора, иначе изображение будет слишком маленьким, чтобы его разглядеть. Возможность использования конкретного разрешения зависит от различных факторов, среди которых возможности самого монитора, возможности видеокарты и объем доступной видеопамати, которая ограничивает число отображаемых цветов.

*Максимальная разрешающая способность* – одна из основных характеристик монитора, которую указывает каждый изготовитель. Однако реальную максимальную разрешающую способность монитора можно определить самостоятельно. Для этого надо иметь три числа: шаг точки (шаг триад для трубок с теневой маской или горизон-

тальный шаг полосок для трубок с апертурной решеткой) и размеры используемой области экрана в мм. Последние можно узнать из описания устройства либо измерить самостоятельно.

Так, для 17-дюймового монитора с трубкой, использующей апертурную решетку, и шагом полосок 0,25 мм по горизонтали и размером используемой области экрана 320·240 мм получается максимальная действительная разрешающая способность 1280·600 точек. Апертурная решетка не имеет шага по вертикали, и разрешающая способность по вертикали такой трубки ограничена только фокусировкой луча.

*Частота регенерации*, или *обновления* (кадровой развертки для CRT мониторов) экрана, — это параметр, определяющий как часто изображение на экране заново перерисовывается. Частота регенерации измеряется в Герцах, (Гц), где один Гц соответствует одному циклу в секунду. Например, частота регенерации монитора в 100 Гц означает, что изображение обновляется 100 раз в секунду. В случае с традиционными CRT-мониторами время свечения люминофорных элементов очень мало, поэтому электронный луч должен проходить через каждый элемент люминофорного слоя достаточно часто, чтобы не было заметно мерцания изображения. Если частота такого обхода экрана становится меньше 70 Гц, то инерционности зрительного восприятия будет недостаточно для того, чтобы изображение не мерцало. Чем выше частота регенерации, тем более устойчивым выглядит изображение на экране. Мерцание изображения (flicker) приводит к утомлению глаз, головным болям и даже к ухудшению зрения. Чем больше экран монитора, тем более заметно мерцание, особенно периферийным (боковым) зрением, так как угол обзора изображения увеличивается. Значение частоты регенерации зависит от используемого разрешения, от электрических параметров монитора и от возможностей видеоадаптера. Минимально безопасной частотой кадров считается 75 Гц, при этом существуют стандарты, определяющие значение минимально допустимой частоты регенерации. Считается, что чем выше значение частоты регенерации, тем лучше, однако исследования показали, что при частоте вертикальной развертки выше 100 Гц глаз человека уже не может заметить никакого мерцания. Частота 100 Гц является рекомендуемой для любых разрешений, однако порогом безопасности считается частота 85 Гц.

## **LCD Мониторы**

Экраны LCD (Liquid Crystal Display, жидкокристаллические (ЖК) мониторы) сделаны из вещества (цианофенил), которое находится в жидком состоянии, но при этом обладает некоторыми свойствами, присущими кристаллическим телам. Фактически это жидкости, обладающие анизотропией свойств (в частности, оптических), связанных с упорядоченностью в ориентации молекул.

Работа ЖК–монитора основана на явлении поляризации светового потока. Известно, что так называемые кристаллы-поляроиды способны пропускать только ту составляющую света, вектор электромагнитной индукции которой лежит в плоскости, параллельной оптической плоскости поляроида. Для оставшейся части светового потока поляроид будет непрозрачным. Таким образом поляроид как бы «просеивает» свет. Этот эффект называется поляризацией света. Когда были изучены жидкие вещества, длинные молекулы которых чувствительны к электростатическому и электромагнитному полю и способны поляризовать свет, появилась возможность управлять поляризацией. Эти аморфные вещества за их схожесть с кристаллическими веществами по электрооптическим свойствам, а также за способность принимать форму сосуда, называли жидкими кристаллами.

Основываясь на этом открытии и в результате дальнейших исследований стало возможным обнаружить связь между повышением электрического напряжения и изменением ориентации молекул кристаллов для обеспечения создания изображения. Первое свое применение жидкие кристаллы нашли в дисплеях для калькуляторов и в электронных часах, а затем их стали использовать в мониторах для портативных компьютеров. Сегодня, в результате прогресса в этой области, начинают получать все большее распространение LCD для настольных компьютеров.

Экран LCD представляет собой массив маленьких сегментов, называемых пикселями, которыми можно манипулировать для отображения информации. LCD имеет несколько слоев, где ключевую роль играют две панели, сделанные из свободного от натрия и очень чистого стеклянного материала, называемого субстрат или подложка. Слои и содержат тонкий слой жидких кристаллов между собой. На панелях имеются бороздки, которые направляют кристаллы, сообщая им специальную ориентацию. Бороздки расположены таким образом, что они параллельны на каждой панели, но перпендикулярны между двумя панелями. Продольные бороздки получаются в результате раз-

мещения на стеклянной поверхности тонких пленок из прозрачного пластика, который затем специальным образом обрабатывается. Соприкасаясь с бороздками, молекулы в жидких кристаллах ориентируются одинаково во всех ячейках. Молекулы одной из разновидностей жидких кристаллов (нематиков) при отсутствии напряжения поворачивают вектор электрического (и магнитного) поля в световой волне на некоторый угол в плоскости, перпендикулярной оси распространения пучка. Нанесение бороздок на поверхность стекла позволяет обеспечить одинаковый угол поворота плоскости поляризации для всех ячеек. Две панели расположены очень близко друг к другу.

Жидкокристаллическая панель освещается источником света (в зависимости от того, где он расположен, жидкокристаллические панели работают на отражение или на прохождение света). Плоскость поляризации светового луча поворачивается на  $90^\circ$  при прохождении одной панели. При появлении электрического поля, молекулы жидких кристаллов частично выстраиваются вертикально вдоль поля, угол поворота плоскости поляризации света становится отличным от  $90^\circ$ , и свет беспрепятственно проходит через жидкие кристаллы.

Поворот плоскости поляризации светового луча незаметен для глаза, поэтому возникла необходимость добавить к стеклянным панелям еще два других слоя, представляющих собой поляризационные фильтры. Эти фильтры пропускают только ту компоненту светового пучка, у которой ось поляризации соответствует заданному. Поэтому при прохождении поляризатора пучок света будет ослаблен в зависимости от угла между его плоскостью поляризации и осью поляризатора. При отсутствии напряжения ячейка прозрачна, т. к. первый поляризатор пропускает только свет с соответствующим вектором поляризации. Благодаря жидким кристаллам вектор поляризации света поворачивается, и к моменту прохождения пучка ко второму поляризатору он уже повернут так, что легко проходит через второй поляризатор.

В присутствии электрического поля поворот вектора поляризации происходит на меньший угол, тем самым второй поляризатор становится только частично прозрачным для излучения. Если разность потенциалов будет такой, что поворота плоскости поляризации в жидких кристаллах не произойдет совсем, то световой луч будет полностью поглощен вторым поляризатором, и экран при освещении сзади будет спереди казаться черным (лучи подсветки поглощаются в экране полностью). Если расположить большое число электродов, которые создают разные электрические поля в отдельных местах экрана (ячей-

ки), то появится возможность при правильном управлении потенциалами этих электродов отображать на экране буквы и другие элементы изображения. Электроды помещаются в прозрачный пластик и могут принимать любую форму. Технологические новшества позволили ограничить их размеры величиной маленькой точки, соответственно на одной и той же площади экрана можно расположить большее число электродов, что увеличивает разрешение LCD-монитора, и позволяет отображать даже сложные изображения в цвете. Для вывода цветного изображения необходима подсветка монитора сзади, чтобы свет исходил из задней части LCD. Это необходимо для того, чтобы можно было наблюдать изображение с хорошим качеством, даже если окружающая среда не является светлой.

Цвет получается в результате использования трех фильтров, которые выделяют из излучения источника белого света три основные компоненты. Комбинируя три основных цвета для каждой точки или пикселя экрана, появляется возможность воспроизвести любой цвет. В случае с цветом несколько возможностей: можно сделать несколько фильтров друг за другом (приводит к малой доле проходящего излучения), можно воспользоваться свойством жидкокристаллической ячейки — при изменении напряженности электрического поля угол поворота плоскости поляризации излучения изменяется по-разному для компонент света с разной длиной волны. Эту особенность можно использовать для того, чтобы отражать (или поглощать) излучение заданной длины волны (проблема состоит в необходимости точно и быстро изменять напряжение). Какой именно механизм используется, зависит от конкретного производителя.

### **Основные параметры LCD-мониторов.**

#### **Время отклика**

Состояние пикселя в ЖК-панели меняется за счет изменения угла поворота жидких кристаллов под действием приложенного к ним электрического поля. Однако жидкие кристаллы — вещество сравнительно вязкое, поэтому поворот происходит не мгновенно, а за достаточно большое время, порядка единиц или даже десятков миллисекунд.

Традиционно производители матриц и мониторов измеряют время отклика как суммарное время переключения пикселя с черного на белый и обратно, причем измеряется время изменения яркости пикселя от 10% до 90% (такое определение, является необходимо-

стью — дело в том, что момент точного начала загорания пикселя и момент точного достижения им яркости 100% принципиально невозможно определить из-за наличия шумов и конечной точности измерительного оборудования, а потому имеет смысл говорить лишь о вхождении яркости пикселя в некоторый интервал, который в данном случае определяется как 10%):

Время переключения с черного на белый зависит от установленной на мониторе контрастности и в некоторых случаях яркости. Вообще, яркость конкретного пикселя  $L$  определяется как  $L = B + x \cdot C$ , где  $B$  — величина, напрямую зависящая от положения регулятора «Brightness» (яркость) монитора,  $C$  — величина, зависящая от положения регулятора «Contrast» (контрастность), а  $x$  — сигнал, подаваемый на данный пиксел с компьютера ( $x=0$  соответствует черному цвету, а максимальное значение  $x$  — белому). Регулировка контрастности осуществляется достаточно просто. Приходящий с видеокарты сигнал  $x$  не подается напрямую на матрицу, а сначала умножается на коэффициент  $C$ , после чего уже подается на матрицу. Очевидно, что белый цвет, соответствующий максимальному углу поворота кристаллов, на самом деле достигается только при максимальной контрастности; если же она ниже максимума, то кристаллы поворачиваются на меньший угол, а потому, время этого поворота больше заявленного производителем. Говоря коротко, снижение контрастности всегда ведет к увеличению времени отклика монитора.

Регулировка «Brightness» в большинстве мониторов реализована изменением яркости ламп подсветки, а потому не связана с матрицей и никак не влияет на время отклика. Также стоит обратить внимание на несимметричность времени отклика, иначе говоря, на разницу между временем зажигания и временем гашения пикселя.

Можно попытаться оценить время отклика по специальным тестам, в которых обычно используется бегающий на черном фоне белый квадратик (например, Passmark Monitor Test). По смазыванию этого квадратика можно определить только время переключения с черного на белый и обратно. Если перенести на ЖК-мониторы опыт тестирования ЭЛТ-мониторов, оценивая время отклика по тянущемуся за бегающим квадратиком шлейфу, то в ЭЛТ-мониторах в силу их специфики (практически мгновенное зажигание пикселя и экспоненциальный график при его гашении) бегающий квадратик будет иметь четкие края и слабо светящийся сравнительно длинный шлейф («хвост» экспоненциальной функции, описывающей гашение пикселя), а на



ЖК-мониторах такой шлейф будет далеко не всегда, ибо на многих матрицах график имеет совершенно иной вид, без длинного «хвоста» — но нельзя делать из этого вывод, что современные ЖК-мониторы уже превзошли ЭЛТ. Обращать же внимание надо на размытие передней и задней граней квадрата — именно они демонстрируют время отклика матрицы. Ниже на рис. 8.7 показана типичная картина для белого квадрата, движущегося по черному фону слева направо: сверху изображена картинка с ЭЛТ-монитора (четкие края, но длинный слабо светящийся шлейф), а внизу — с типичного ЖК-монитора (отсутствие заметного шлейфа, но сильно размытые края):

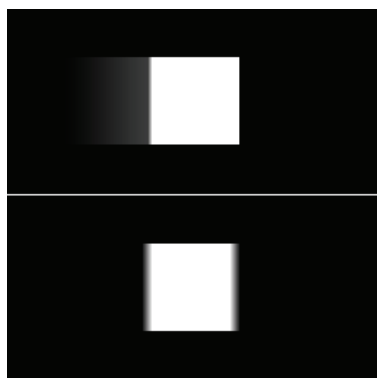


Рис. 8.7. Тест с черным квадратов для ЭЛТ-монитора(сверху) и ЖК-монитора(снизу).

### **Углы обзора**

Другой проблемой ЖК-мониторов являются углы обзора — если изображение на ЭЛТ практически не страдает даже при взгляде почти параллельно плоскости экрана, то на многих ЖК-матрицах даже небольшое отклонение от перпендикуляра приводит к заметному падению контрастности и искажению цветопередачи.

Согласно текущим стандартам, производители матриц определяют угол обзора как угол относительно перпендикуляра к центру матрицы, при наблюдении под которым контрастность изображения в центре матрицы падает до 10:1.

На данный момент реализованы большие углы обзора — у большинства моделей мониторов они составляют не менее 170 градусов как по вертикали, так и по горизонтали.

### **Яркость и контрастность**

Под яркостью понимается яркость белого цвета (т. е. на матрицу подается максимальный сигнал) в центре экрана, под контрастностью – отношение уровня белого цвета к уровню черного, также в центре экрана.

Проблема с контрастностью принципиальна для ЖК-матриц в силу самого их принципа действия. В отличие от абсолютного большинства электронных устройств отображения информации (ЭЛТ, электролюминесцентные и светодиодные табло, OLED и т. д.) по отношению к свету матрица является не активным, а пассивным элементом, она не способна излучать свет, а может лишь модулировать проходящий через нее. Поэтому позади ЖК-матрицы всегда размещается модуль подсветки, а матрица только управляет своей прозрачностью, ослабляя свет от модуля подсветки в заданное количество раз. Регулировка прозрачности осуществляется за счет поворота плоскости поляризации.

Очевидно, что в силу принципа действия ЖК-матриц увеличить яркость можно только увеличением интенсивности подсветки. Пользователь может сам регулировать яркость и контрастность, что влияет на многие параметры изображения.

Регулировкой «Contrast» пользователь меняет яркость белого цвета (а точнее говоря, и всех оттенков серого, но вот черный цвет остается неизменным), а регулировкой «Brightness» – яркость как черного, так и белого одновременно.

В большинстве мониторов регулировка «Brightness» реализована изменением яркости ламп подсветки. Используемые в мониторах лампы дневного света с холодным катодом (CCFL – Cold Cathode Fluorescent Lamp) позволяют это делать двумя способами: либо регулируя ток разряда в лампе, либо с помощью широтно-импульсной модуляции питания лампы на сравнительно небольшой частоте (сравнительно – потому что она достаточно мала с точки зрения физики разряда в лампе, но при этом достаточно велика, чтобы глаз не замечал мерцания подсветки; на практике частота составляет обычно от 200 до 500 Гц). Широтно-импульсная модуляция – очень распространенный способ регулировки напряжений и токов, заключающийся в том, что в зависимости от нужного напряжения регулируется ширина подаваемых импульсов при их неизменной частоте и амплитуде – и среднее напряжение как раз оказывается пропорционально этой ширине.

Также встречается регулировка яркости с помощью матрицы: при увеличении яркости пользователем монитор добавляет к подаваемому на матрицу сигналу постоянную составляющую. При таком способе регулировки заметно страдает контрастность, ведь лампы подсветки всегда работают на мощности, необходимой для обеспечения максимально возможной для монитора яркости, а потому при работе на небольшой яркости, даже если добавляемая к сигналу постоянная составляющая будет уже равняться нулю, такой монитор покажет заведомо более высокий уровень черного, чем модель с регулировкой яркости с помощью ламп.

В современных ЖК мониторах яркость достигает 400 кд/кв. м (кандела/квадратный метр) и выше, контрастность достигает 20000:1 и выше.

### **Цветопередача**

С точки зрения цветопередачи производители обычно указывают одну цифру – количество цветов, которое традиционно равняется 16,2 млн. или 16,7 млн. (что соответствует 24 битам, или по 8 бит на каждый из трех базовых цветов).

## **Другие типы мониторов**

### **Плазменная технология**

Цветные плоские плазменные панели PDP (Plasma Display Panel) появились на нашем рынке несколько лет назад и вызвали огромный интерес и специалистов, и широкой публики. Плазменная технология известна довольно давно (начиная с 80-х годов), разрабатывать ее начала фирма JVC. Принцип действия плазменной панели основан на свечении специальных люминофоров при воздействии на них ультрафиолетового излучения. В свою очередь, это излучение возникает при электрическом разряде в среде сильно разреженного газа. При таком разряде между электродами с управляющим напряжением образуется проводящий шнур, состоящий из ионизированных молекул газа (плазмы). Поэтому газоразрядные дисплеи, работающие на этом принципе, и получили название газоразрядных, или плазменных панелей. Подавая управляющие сигналы на вертикальные и горизонтальные проводники, нанесенные на внутренние поверхности стекол панели, схема управления PDP осуществляет соответственно строчную и кадровую развертку. При этом яркость каждого элемента изображе-

ния определяется временем свечения соответствующей ячейки плазменной панели: самые яркие элементы горят постоянно, а в наиболее темных местах они вовсе не поджигаются. Светлые участки изображения на PDP светятся ровным светом, и поэтому изображение абсолютно не мерцает, что выгодно отличает плазменные дисплеи от традиционных кинескопов.

Плазменные панели, как и LCD-панели, не создают вредных магнитных и электрических полей, так как в них отсутствуют устройства развертки и высоковольтный источник анодного напряжения кинескопа. Хотя и современные CRT-мониторы тоже практически не имеют вредных излучений.

Для плазмы по сравнению с LCD характерно отсутствие серьезных ограничений на размер экрана (уже повсюду производятся экраны с диагональю 64", 72" и т. д.).

### **Дисплей OLED**

Технология OLED является технологией следующего поколения в ряду FDP (flat panel displays). Приборы OLED — это светоизлучающие полноцветные приборы, которые обеспечивают высокую яркость, малую потребляемую мощность, широкий угол обзора, хорошую контрастность изображения. Кроме того, они компактные и легкие, выдерживают значительные механические нагрузки, обладают широким диапазоном рабочих температур и имеют достаточный срок службы. Область применения таких дисплеев довольно широкая: от сотовых телефонов и автомагнитол до наשלемых индикаторов, дисплеев на лобовом стекле транспортных средств и осветительных приборов. При последующем развитии фосфоресцентных материалов, приборы OLED могут стать не только эффективным средством отображения, но и тонкопленочным источником света, заменяя многочисленные дискретные лампы накаливания и дорогие большие неорганические светодиоды. Не исключено, что через пару лет TFT LCD дисплеи будут сменяться мониторами на базе OLED.

### **Устройство и основные параметры OLED**

Органический электролюминесцентный дисплей OLED (рис.8.8) представляет собой монолитный тонкопленочный полупроводниковый прибор, который излучает свет, когда к нему приложено напряжение. OLED состоит из ряда тонких органических пленок, которые

заклучены между двумя тонкопленочными проводниками. Рабочее напряжение OLED – всего лишь 3-10 В.

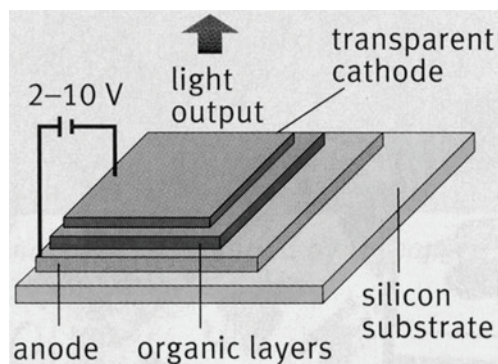


Рис. 8.8. Дисплей OLED

Цвет, эффективность и интенсивность излучения приборов OLED зависят от использованных органических материалов, которыми определяется многообразие воспроизводимых дисплеем цветов. Основное внимание разработчиков приборов OLED направлено на создание материалов для полноцветных приборов OLED (широкий цветовой охват, высокая точность и постоянство цветопередачи позволяют мониторам OLED по области применения обогнать LCD TFT мониторы).

В приборах OLED используются два класса органических материалов. Это микромолекулы и полимеры. С 1987 г. (который принято считать началом исследования OLED) работы по обоим направлениям создания OLED велись параллельно. В то время как пользователь не смог бы отличить полимерный прибор OLED (PLED) от прибора OLED с микромолекулами (sm-OLED), эти две системы имеют несколько различий. Сегодня приборы sm-OLED опережают приборы PLED по эффективности и сроку службы. Тем не менее, для обоих приборов достигнут огромный прогресс. Недавно эффективность лучших приборов sm-OLED и PLED более чем удвоилась.

Итак, OLED — это не что иное, как тонкопленочное устройство со светоизлучающей поверхностью. Поверхность эта образована множеством одновременно излучающих свет ячеек на одной подложке. Причем эти ячейки могут быть изготовлены либо методом напыления, либо методом струйной печати; для создания дисплея с произвольным структурированием можно применить обычную литографию. Другими словами, OLED имеют значительные преимущества в технологии формирования структуры.

Дисплеи OLED имеют очень широкий угол обзора (более 175°) и время запаздывания — 8 микросекунд.

Приборы OLED равномерно и без мелькания излучают яркость от нескольких кд/кв.м (для ночной работы) до очень высоких яркостей — свыше 100 000 кд/кв.м, причем их яркость может регулироваться в очень широком динамическом диапазоне.

Т. к. срок службы OLED обратно пропорционален яркости, для приборов рекомендуется работа при более умеренных уровнях яркости.

Чтобы обеспечить хорошую читаемость информации, при прямом солнечном свете следует увеличивать *контраст*, а не яркость приборов. Чтобы добиться этого, отражающая способность дисплея должна быть управляемой. Приборы OLED имеют очень хорошую контрастность.

Повысить контраст удастся в просветных приборах OLED (т. н. приборы с верхним излучением — TOLED). Приборы TOLED, которые имеют коэффициент пропускания 70 – 85 % в выключенном состоянии, могут быть применены для воспроизведения информации на лобовом стекле.

Возможность работы в широком диапазоне температур – актуальный момент для транспортных дисплеев, которые должны работать от температур ниже нуля до температур, которые превышают 80°C. В то время как для ЖК дисплеев воздействие низких температур оказывается неблагоприятным, и обычно требуется использование подогрева подложки, приборы OLED хорошо работают даже при температуре минус 40°C. При высоких температурах приборы OLED имеют допустимую рабочую температуру порядка 70°C.

Оценивая габариты дисплея OLED, следует отметить, что они тонкие и легкие. Используя стекло толщиной 0,7 мм, дисплей OLED будет иметь толщину порядка или чуть больше 1,4 мм.

### **Направление развития**

Основные усилия разработчиков OLED сегодня направлены на уменьшение дифференциального старения, повышение чистоты цвета и увеличение срока службы полноцветных приборов.

Отдельная проблема — получение эффективного белого цвета либо путем разработки новых материалов, либо методом смешения цветов.

Немаловажно, что для OLED требуется совершенная герметизация, потому что органические флуоресцентные материалы чрезвычайно чувствительны к влажности.

Сегодня возможно практическое применение двух вариантов построения схем управления OLED дисплеев — с пассивными и активными матрицами. OLED дисплеи с пассивными схемами управления имеют потенциальную возможность отображать полноцветные подвижные изображения с превосходным качеством. Это демонстрировалось на многих прототипах среднего размера, но при создании маленьких цветных панелей с высоким разрешением и при использовании пассивной схемы управления возникают значительные трудности. Альтернативный вариант, который решает возникающие проблемы: использование активно-матричной технологии для управления тонкопленочными транзисторами на низкотемпературном поликристаллическом кремнии. Дальнейшее развитие активно-матричной технологии OLED позволит расширить область применения OLED дисплеев.

Дисплеи OLED на гибкой подложке (FOLED) еще более прочные, чем обычные дисплеи OLED. Кроме того, они допускают изгиб с радиусом кривизны  $< 1$  см. Использование дисплея FOLED открывает совершенно новые возможности при проектировании панели транспортного средства. Экран может быть установлен в пространство, раньше недоступное для дисплеев на плоских панелях, например, интегрировать дисплей FOLED в изогнутую приборную доску, на внутреннюю облицовку боковых панелей и потолка и, возможно, на ветровое стекло, что обеспечит одновременно удобство и безопасность для оператора.

## 9. АУДИОСИСТЕМА

### 9.1. Элементы прикладной звукотехники

#### Носители звуковой информации

Первичным носителем звуковой информации является акустическая среда, способная передавать звуковые волны – колебания давления воздуха, воды и другой среды. Источником звука являются механические колебания (колебания голосовых связок, струны и др.). Воздействуя на слуховой аппарат человека, колебания воздуха преобразуются в слуховые ощущения. При этом частота колебаний определяет высоту звука, а амплитуда – громкость (уровень). Человеческое ухо воспринимает колебания с частотой 20 Гц – 20 кГц (Гц – 1 колебание в секунду, герц). «Полезный» звук всегда сопровождается некоторым шумом. По отношению уровня полезного звукового сигнала к уровню шума человеческое ухо настроено на, так называемый, максимальный динамический диапазон в 90 дБ (децибел). Превышение этого диапазона сопровождается чрезмерным увеличением громкости и вызывает болезненные ощущения.

Важнейшим вторичным носителем звуковой информации является *электромагнитное поле*. Электромагнитные волны возникают при воздействии звуковых волн на мембрану микрофона, распространяются по проводам или в радиозфире, а в акустических системах (звуковых колонках, наушниках) преобразуются снова в звук. При этом частота и амплитуда звуковых колебаний определенным образом связаны с частотой и амплитудой электромагнитных колебаний. Такое электромагнитное представление звуковой информации называется аналоговым звуковым сигналом. Дальность передачи аналогового сигнала значительно превышает дальность передачи звуковых волн.

Акустический сигнал научились записывать и воспроизводить в 1877 году. Сначала для этого использовался механический способ записи звука на восковые валики (фонограф Т. Эдисона), затем на пластмассовые граммофонные пластинки (граммофон, патефон) — механическим и электромеханическим способом. В кино используется фотоэлектрический способ записи звука на киноленту. В отличие от «акустического» звука для его аналогового представления существуют весьма совершенные способы сохранения звуковой информации на магнитных лентах. Однако, при передаче или перезаписи акустического или ана-



логового сигнала он искажается, звуковая информация теряется и качество звука ухудшается.

И в «акустическом» звуковом сигнале, и в аналоговом амплитуда колебаний является непрерывной функцией времени и пространства. Разложение этой функции в ряд Фурье (по функциям  $\sin(nt)$  и  $\cos(nt)$ ) имеет естественную интерпретацию. Каждая такая функция называется гармоникой и соответствует звучанию одного тона (ноты), соответствующего звуковой волне с частотой  $n$ . Величина коэффициента при этой гармонике определяет громкость ее звучания. При извлечении некоторого тона с помощью музыкального инструмента этот основной тон сопровождается обертонами — более высокими и слабыми тонами, определяющими тембр инструмента. Важной характеристикой является спектр звуковой волны — совокупность коэффициентов ряда Фурье для всех гармоник. Анализ спектра используется для борьбы с излишними шумами.

Итак, слышимые звуки представляют собой механические колебания, достигающие ушей слушателя обычно по воздуху. Диапазон частот, воспринимаемых человеческим ухом, простирается от 20 Гц до 20 кГц, причем наибольшая чувствительность приходится на частоты 2-5 кГц. В этой области ухо воспринимает сигналы в динамическом диапазоне около 140 дБ (отношение звукового давления болевого порога к порогу слышимости  $10^7$ ). На краях частотного диапазона динамический диапазон сужается до 50 дБ (чувствительность уха существенно снижается, а давление болевого порога уменьшается). Разговорная речь в спектре занимает область примерно 200 Гц – 4 кГц при динамическом диапазоне около 40 дБ. Музыка может занимать практически весь слышимый диапазон частот и требовать динамического диапазона 70 – 90 дБ. Важной особенностью слуха является способность к локализации источника звука, обеспечиваемая его бинауральным восприятием. Дело в том, что звуковые волны воспринимаются обоими ушами, которые пространственно разнесены. Колебания от одного источника достигают ушей с разными амплитудой и фазой, что позволяет мозгу оценить направление (азимут) источника звука. Сигналы с частотами ниже 300 Гц локализуются плохо, поскольку длина волны относительно размера головы велика. Наибольшее значение для локализации имеют частоты от 1 до 3,2 кГц. Бинауральное восприятие позволяет не только локализовывать, но и выделять отдельные источники (например, отдельные инструменты в оркестре).

Для передачи, хранения, воспроизведения и синтеза звуков традиционно акустические колебания преобразуют в электрические (микрофон) и обратно (динамик). Первоначально вся промежуточная обработка (усиление, преобразования) сигналов производилась в аналоговой форме, естественной для конечных электромеханических преобразователей. Хранение, опять-таки в аналоговой форме, выполнялось на механических (грампластинки) или магнитных (магнитофонные ленты) носителях. Для повышения достоверности звукопередачи, включая пространственное расположение источников звука, применяется двухканальная передача и хранение — стереофония. Упрощенно идея заключается в разделении трактов сигналов, предназначенных для левого и правого ушей слушателя. Такая система позволяет создать иллюзию звуковой панорамы — кажущиеся источники звука (КИЗ) располагаются на воображаемой сцене перед слушателем. Однако пара колонок не позволяет добиться большой ширины зоны стереоэффекта. Прослушивание через головные телефоны не всегда удобно и тоже не дает полной иллюзии присутствия — поворот головы в сторону КИЗ ведет к его уходу в ту же сторону. В более сложных системах используют большее число каналов, например, 4 в квадрофонии и 6 в системе АС – 3. Здесь колонки располагаются вокруг (спереди и сзади) слушателя, что позволяет получить эффект присутствия внутри некоторого озвученного объема. Есть и промежуточные варианты между дорогой квадрофонией и фактически стандартной стереофонией — квази- и псевдоквадрофония.

Аналоговое представление сигналов для обработки (фильтрации, создания различных эффектов) и хранения имеет массу недостатков. Во-первых, все устройства в той или иной степени обладают нелинейными передаточными характеристиками: проходящий через них гармонический (синусоидальный) сигнал «обрастает» гармониками — составляющими с частотами, кратными основной. Мерой искажений, вносимых нелинейностью, является коэффициент гармоник, он же коэффициент нелинейных искажений (КНИ), который определяется как отношение мощности гармоник выходного сигнала к мощности основного тона. Эти искажения вносят все элементы тракта, так что их всюду стремятся минимизировать. Для современных высококачественных усилителей считается хорошим значение КНИ в десятые и сотые доли процента, для электромеханических преобразователей (особенно динамиков) значения гораздо выше.

Следующая проблема — шумы и помехи, характерные для любой аналоговой техники. Они сужают динамический диапазон устройства. Отношение сигнал/шум порядка 90 – 100 дБ для аналоговых устройств удалось получить сравнительно недавно.

Что касается хранения информации, то и здесь аналоговая форма наиболее уязвима — магнитные ленты осыпаются и размагничиваются, в результате ранее записанный сигнал при воспроизведении сильно искажается. Потери происходят и при тиражировании — каждая перезапись вносит свою долю искажений.

С развитием электроники появилась возможность большую часть «путешествия» электрического сигнала производить в цифровой форме. Теперь входной сигнал (от микрофона) после предварительного усиления оцифровывается. В цифровой форме он может передаваться, храниться (долго и без накопления ошибок), подвергаться различным искусственным преобразованиям. При воспроизведении выполняются обратное преобразование в аналоговую форму, окончательное усиление и преобразование в акустические колебания. Для цифрового хранения акустической информации стали применять лазерные компакт-диски (Audio – CD, DVD) и магнитные ленты для цифровой звукозаписи (Digital Audio Tape, DAT), которые долгое время считались эталонами качества. По мере развития средств вычислительной техники возможности обычных РС доросли до того, чтобы пропускать через себя поток цифровых аудиоданных (или создавать собственный).

## 9.2. Оцифровка звука

*Цифровой звук* — это способ представления электрического сигнала посредством дискретных численных значений его амплитуды. Допустим, есть аналоговая звуковая дорожка хорошего качества (говоря «хорошее качество» предполагается нешумная запись, содержащая спектральные составляющие из всего слышимого диапазона частот — приблизительно от 20 Гц до 20 КГц) и необходимо «ввести» ее в компьютер (то есть оцифровать) без потери качества. Как этого добиться и как происходит оцифровка?

*Звуковая волна* — это некая сложная функция, зависимость амплитуды звуковой волны от времени. Казалось бы, что раз это функция, то можно записать ее в компьютер «как есть», то есть описать математический вид функции и сохранить в памяти компьютера. Од-

нако практически это невозможно, поскольку звуковые колебания нельзя представить аналитической формулой (как  $y=2x$ , например). Остается один путь – описать функцию путем хранения ее дискретных значений в определенных точках. Если преобразовать звук в электрический сигнал (например, с помощью микрофона), получится плавно изменяющееся с течением времени напряжение. Для компьютерной обработки такой аналоговый сигнал нужно каким-то образом преобразовать в последовательность двоичных чисел.

Следует поступить таким образом: измерять напряжение через равные промежутки времени и записывать полученные значения в память компьютера. Они называются выборками. Этот процесс называется дискретизацией или — сэмплингом, а устройство, выполняющее его — аналого-цифровым преобразователем (АЦП). На рис. 9.1 представлена фонограмма первых 10 секунд 40-й симфонии Моцарта, содержащая 2 205 000 выборок. На рис. 9.2 показан фрагмент этой записи, содержащий только 50 выборок, начиная с места, отмеченного на рис. 9.1 вертикальной чертой.



Рис. 9.1. Выборки напряжения на отрезке времени равном 10 секунд

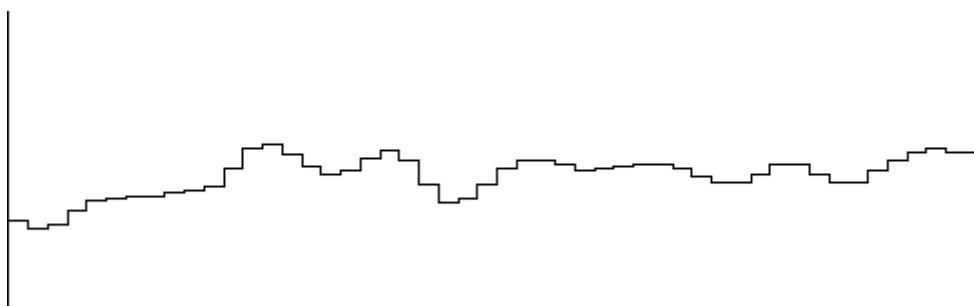


Рис.9.2. 50 выборок напряжения

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

Звуковая информация представляется в цифровом виде в двух принципиально разных формах — WAVE и MIDI. WAVE-форма используется для цифровых аудиодисков, а MIDI — в электронных музыкальных инструментах.

### **WAVE-форма звука**

WAVE-форма цифрового звука характеризуется пятью параметрами: — частотой дискретизации;

- разрядностью выборок;
- числом каналов или звуковых дорожек;
- алгоритмом компрессии/декомпрессии — кодеком;
- форматом хранения.

### **Частота дискретизации**

Количество выборок в секунду называется частотой дискретизации и измеряется в герцах и килогерцах ( $1 \text{ кГц} = 1000 \text{ выборок в сек.}$ ).

Но при оцифровке звука дискретизируются и высокочастотные гармоники с частотой, превышающей половину частоты дискретизации (это высокочастотный, неслышимый шум). При этом появляются новые нежелательные гармоники из слышимого диапазона (низкочастотный шум, см. рис. 9.3).

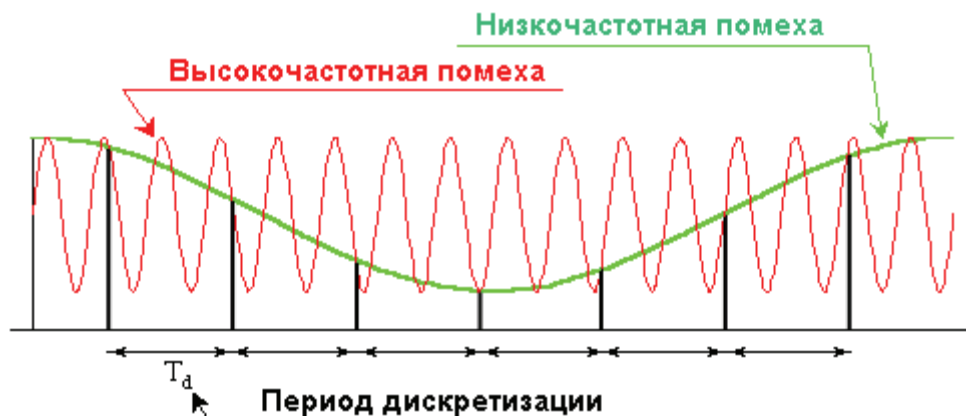


Рис.9.3. Помехи при оцифровке звука

Чем выше частота дискретизации (т. е. количество отсчетов за секунду) и чем больше разрядов отводится для каждого отсчета, тем точнее будет представлен звук. Но, естественно, увеличивается и размер звукового файла. Поэтому в зависимости от характера звука, требований, предъявляемых к его качеству и объему занимаемой памяти, выбирают некоторые компромиссные значения. Теоретически для правильного восстановления аналогового сигнала по его цифровой записи достаточно, чтобы частота дискретизации более чем в два раза превосходила максимальную частоту звука (теорема Котельникова-Найквиста). Таким образом, для качественного воспроизведения самого высокого слышимого звука 20 кГц необходима частота дискретизации не менее 40 кГц. Стандарт CD цифровых аудиодисков требует частоты дискретизации 44,1 кГц (так называемое CD-качество). Используются также частоты 22,05 кГц (радио-качество) и 11,025 и 8 кГц (телефонное качество), во многих случаях дающие удовлетворительные результаты (например, для записи речи).

Поэтому для получения качественного цифрового звука процесс дискретизации строится по следующей схеме :

- частичное подавление высокочастотных помех в аналоговом сигнале с помощью аналогового фильтра;
- оверсэмплинг – дискретизация с частотой, значительно превышающей требуемую (при этом образующийся шум лежит все еще в неслышимом диапазоне);
- подавление высокочастотного шума в цифровом звуке с помощью цифрового фильтра;
- преобразование к требуемой частоте дискретизации.

Оверсэмплинг применяется и при обратном преобразовании в аналоговый сигнал. При этом используются различные методы линейной и нелинейной интерполяции.

Описанный способ кодирования звуковой информации универсален, он позволяет представить любой звук, преобразовывать его самыми разными способами. Но в этом методе есть свои недостатки, так как значения амплитуды сигнала мы не можем записывать с бесконечной точностью, и вынуждены их округлять. Говоря иначе, мы будем приближать эту функцию по двум координатным осям – амплитудной и временной (приближать в точках — значит, говоря простым языком, брать значения функции в точках и записывать их с конечной точностью).

Таким образом, оцифровка сигнала включает в себя два процесса – процесс дискретизации (осуществление выборки) и процесс квантования. *Процесс дискретизации* – это процесс получения значений величин преобразуемого сигнала в определенные промежутки времени. *Квантование* – процесс замены реальных значений сигнала приближенными с определенной точностью. Измеренная амплитуда (выборка) преобразуется в целое число с некоторой погрешностью, определяемой разрядностью этого числа. Погрешность при квантовании вносит шум тем больший, чем меньше разрядность. Теоретически, при  $n$ -разрядном квантовании отношение сигнал/шум (динамический диапазон) будет составлять  $6n$  дБ. Таким образом, при 8-разрядной оцифровке динамический диапазон будет 48 дБ (качество УКВ-радио, достаточное для речи), а при 16-разрядной оцифровке он равен 96 дБ и практически покрывает весь нормально слышимый диапазон. Другими словами, при 8-разрядном кодировании различаются 256 уровней громкости, а при 16-разрядном – 65536, что дает более точную запись. На CD применяется 16-разрядное квантование. Звуковые платы компьютеров обычно используют 8- и 16-разрядное квантование. Применение оверсэмплинга в АЦП и ЦАП позволяет значительно уменьшить шум, добавляемый при квантовании.

Таким образом, оцифровка – это фиксация амплитуды сигнала через определенные промежутки времени и регистрация полученных значений амплитуды в виде округленных цифровых значений (так как значения амплитуды являются величиной непрерывной, нет возможности конечным числом записать точное значение амплитуды сигнала, именно поэтому прибегают к округлению). Записанные значения амплитуды сигнала также называются отсчетами. Очевидно, что чем чаще делать замеры амплитуды (чем выше частота дискретизации) и чем меньше округлять полученные значения (чем больше уровней квантования), тем более точное представление сигнала в цифровой форме будет получено. Оцифрованный сигнал в виде набора последовательных значений амплитуды можно сохранить.

### **9.3. Использование ПК для обработки «цифрового» звука**

Аппаратные средства цифровой записи и воспроизведения для РС достаточно просты — АЦП, ЦАП и средства доставки потока данных на диск или с диска. Этими средствами, называемыми аудиокодеком (кодер-декодер), обладает любая звуковая карта. На современных

картах PCI доставку осуществляет ее собственный контроллер путем прямого управления шиной. Канал DMA справляется с любым аудиопотоком, вопрос лишь в размещении данных — звукозапись с высоким качеством довольно расточительно расходует дисковое пространство. Нетрудно сосчитать, что на дискету 1,44 Мбайт можно уместить около 5 мин. речи с низким качеством. А 1 мин. стереомузыки с качеством CD занимает около 10 Мбайт (на CD вмещается до 74 мин.), так что есть повод задуматься о *компрессии*. В задачу компрессоров/декомпрессоров входит сокращение потока данных в канале передачи (хранения) относительно потока на выходе АЦП и входе ЦАП. Современные аудиокодеки производят компрессию аудиопотока. При этом снижается качество звука. Но, например, схема Dolby обеспечивает сжатие в два раза без потерь.

Цифровое хранение обеспечивает богатейшие возможности нелинейного монтажа аудиозаписей. Под нелинейностью подразумевается возможность произвольного доступа к любым фрагментам записи. Даже простейшие средства «фонографа» из стандартных утилит Windows позволяют вырезать или вставить любой фрагмент записи, состыковать (один за другим) фрагменты или наложить один на другой. Программная реализация монтажа является «бескровной» и позволяет легко отказаться от проведенных действий. Поиск начала и конца требуемого фрагмента делается теперь с помощью точного указателя-курсора.

При цифровом хранении легко реализуются многие эффекты, которые ранее требовали громоздких электромеханических или электроакустических устройств или сложной аналоговой электроники. Прежде всего, это искусственная реверберация и эхо. Известно, что в закрытом помещении (например, зале) от источника до слушателя доходит не только прямой, но и многократно отраженный от различных поверхностей (стен, колонн и т. п.) звук. Отраженные сигналы приходят с различными задержками и затуханием относительно прямого. Это явление называется реверберацией. Акустика зала (реверберационные характеристики) должна соответствовать его назначению — в помещении с малой реверберацией музыка будет звучать «сухо», а в помещении с длительной реверберацией речь окажется неразборчивой. В студиях звукозаписи естественную реверберацию часто подавляют, а в запись добавляют искусственную реверберацию «по вкусу».

В доцифровую эпоху использовали ревербераторы на электромеханических линиях задержки, акустические камеры с микрофонами



или магнитофонные ревербераторы. Все это было дорого, громоздко, некачественно и трудноуправляемо. При цифровой записи или воспроизведении реверберация достается практически даром. Для этого достаточно на вход ЦАП подавать не просто очередную выборку, а ее сумму с одной или несколькими предыдущими выборками, которые еще присутствуют в памяти. Конечно, эти «отстающие» выборки должны быть снижены по уровню — в цифровой форме эти операции выполнить нетрудно. Задержку (отставание выборки) и уровень каждого эхо-сигнала легко задать программно, количество составляющих ограничивается только производительностью процессора, обрабатывающего эти сигналы. Таким образом, появляется возможность имитации воспроизведения в каком-либо знаменитом зале — для этого достаточно снять его характеристики и заложить их в программу (конечно, на практике это не так просто, но возможно). Искусственное эхо технически отличается от реверберации лишь большим временем задержки, так что теперь и с ним нет проблем.

На основе смещения выборок можно делать и более сложные эффекты. В цифровой форме представления легко имитируется эффект Доплера — изменение частоты при быстром приближении источника звука к слушателю или удалении источника от слушателя. С этим эффектом сталкивались все: однотонный свисток приближающегося поезда звучит выше, а удаляющегося — ниже реального тона.

## **9.4. Компрессия звука**

### **Типы кодировки**

Теперь о практических проблемах. Во-первых, надо иметь в виду, что память компьютера не бесконечна, так что каждый раз при оцифровке необходимо находить какой-то компромисс между качеством (напрямую зависящим от использованных при оцифровке параметров) и занимаемым оцифрованным сигналом объемом.

Во-вторых, согласно теореме Котельникова частота дискретизации устанавливает верхнюю границу частот оцифрованного сигнала, а именно, максимальная частота спектральных составляющих равна половине частоты дискретизации сигнала. Попросту говоря, чтобы получить полную информацию о звуке в частотной полосе до 22050 Гц, необходима дискретизация с частотой не менее 44,1 КГц.

Если обозначить:

$W$  – объем памяти в байтах для хранения 1 секунды звука в WAVE-форме,

$w$  – скорость потока звуковых данных в WAVE-форме в бит/сек;

$H$  – частота дискретизации (число выборок в секунду);

$B$  – разрядность квантования (число разрядов на выборку);

$C$  – число каналов

тогда очевидно, что  $w = H \cdot B \cdot C$  и, если  $B$  кратно 8,  $W = w/8$ .

Следовательно, скорость потока данных CD-качества ( $H=44100$ ,  $B=16$ ,  $C=2$ ) составляет 1 411 200 бит/сек или 1378,125 Кбит/сек (такая скорость обеспечивается только CD-дисководом с не менее, чем 2-кратной скоростью), 1 час звука с качеством CD-DA потребует 605,6 Мбайт (поэтому на аудиодиске помещается около 70 мин. несжатого звука), скорость потока данных телефонного качества ( $H=8000$ ,  $B=8$ ,  $C=1$ ) составляет 64 000 бит/сек или 62,5 Кбит/сек (такая скорость обеспечивается далеко не каждым модемом, так что такой звук не может использоваться в Интернет-телефонии).

С целью уменьшения объема и потока звуковых данных в WAVE-форме используются различные специальные алгоритмы компрессии/декомпрессии (кодеки), т. к. обычные алгоритмы сжатия информации здесь не дают эффекта. Сжатие аудиоданных возможно лишь с некоторой потерей информации, но учет психофизиологических особенностей восприятия звука (например, не все частоты в слышимом диапазоне существенны для восприятия), позволяет в ряде случаев сделать эти потери практически незаметными. Следует, однако, учитывать очень высокую чувствительность человеческого слухового аппарата, особенно к временным характеристикам звука. Наиболее известными являются следующие типы кодировки, используемые в мультимедиа под Windows:

PCM (Pulse Code Modulation) — импульсно-кодовая модуляция (ИКМ) — сжатие может достигаться только за счет выбора меньших значений величин  $H$ ,  $B$  и  $C$  (фактически, это несжатый звук); квантование происходит по равномерной шкале из  $2B$  значений;

DPCM (Differential PCM) — дифференциальная ИКМ; сжимает в несколько раз;

ADPCM (Adaptive DPCM) — адаптивная ДИКМ (АДИКМ) — то же, что ДИКМ, только квантование происходит не по равномерной шкале, а с учетом динамики изменений амплитуды; сжимает в несколько раз;

RealAudio — метод, разработанный фирмой RealNetworks, сжимает в несколько десятков раз, но с невысоким качеством; используется в Интернете для проигрывания звуковых файлов в реальном времени.

Для цифровой телефонии используются, как правило, другие кодеки.

MPEG (Motion Picture Experts Group) — стандарты группы экспертов в области кино. Для сжатия звуковой информации используются стандарты MP2 и MP3, применяется психоакустическая компрессия, при которой удаляются звуки, не воспринимаемые человеческим ухом; сжимает звуковую информацию в несколько десятков раз при довольно высоком качестве. С точки зрения степени сжатия/качество — это наиболее интересный вариант сжатия звуковых файлов. Подобные техники называются адаптивным кодированием и позволяют экономить на наименее значимых с точки зрения восприятия человеком деталях звучания. Степень сжатия и, соответственно, объем дополнительного квантования определяются не форматом, а самим пользователем в момент задания параметров кодирования. Ширина потока (bitrate) при кодировании сигнала, аналогичного CD Audio (44,1kHz 16Bit Stereo) варьируется от наибольшего 320kbs (320 килобит в секунду, также пишут kbs, kbps или kb/s) до 96kbs и ниже. (см. 9.4., фото из меню-настройки программы «АудиоГрэйбер».)

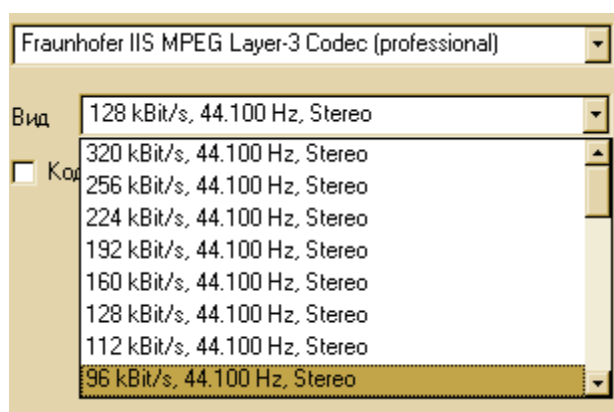


Рис. 9.4. Вариации битрейта.

Термин «битрейт» в общем случае обозначает общую величину потока, количество передаваемой за единицу времени информации, и поэтому не связан с внутренними тонкостями строения потока, его смысл не зависит от того, содержит ли поток моно или стерео, или пятиканальное аудио с текстом на разных языках, или что-либо еще.

## Описание процесса кодирования в МРЕГ

### Подготовка к кодированию. Фреймовая структура

Перед кодированием исходный сигнал разбивается на участки, называемые *фреймами*, каждый из которых кодируется отдельно и помещается в конечном файле независимо от других. Последовательность воспроизведения определяется порядком расположения фреймов. Каждый фрейм может кодироваться с разными параметрами. Информация о них содержится в заголовке фрейма. В «цифровом звуке» из-за дискретности информации об амплитуде оригинального сигнала появляются различные шумы и искажения (под фразой «в цифровом звуке есть такие-то частоты и шумы» подразумевается, что когда этот звук будет преобразован обратно из цифрового вида в аналоговый, то в его звучании будут присутствовать упомянутые частоты и шумы).

Так, например, *джиттер* (jitter) – шум, появляющийся в результате того, что осуществление выборки сигнала при дискретизации происходит не через абсолютно равные промежутки времени, а с отклонениями. Т. е., если дискретизация проводится с частотой 44.1 КГц, то отсчеты берутся не точно каждые  $1/44100$  секунды, а то немного раньше, то немного позднее. А т. к. входной сигнал постоянно меняется, то такая ошибка приводит к «захвату» не совсем верного уровня сигнала. В результате во время проигрывания оцифрованного сигнала может ощущаться некоторое дрожание и искажения. Появление джиттера является результатом не абсолютной стабильности аналогово-цифровых преобразователей. Для борьбы с этим явлением применяют высокостабильные тактовые генераторы. Еще одной неприятностью является шум дробления, т. к. при квантовании амплитуды сигнала происходит ее округление до ближайшего уровня. Такая погрешность вызывает ощущение «грязного» звучания (см. рис. 9.5.).

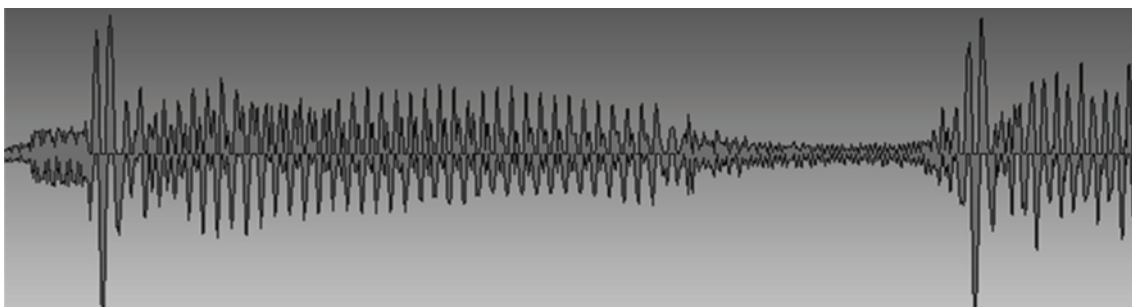


Рис. 9.5. Наглядное изображение округления сигнала и появления дрожания

### **Начало кодирования**

Кодирование начинается с того, что исходный сигнал с помощью фильтров разделяется на несколько, представляющих отдельные частотные диапазоны, сумма которых эквивалентна исходному сигналу.

### **Работа психоакустической модели**

Самой важной характеристикой кодера, от которой зависит качество кодирования, является психоакустическая модель, использованная в нем. Но следует заметить, что модель должна варьироваться для разных категорий битрейтов. Аналогично тому, как использование MS/IS стерео приводит к повышению качества на низких битрейтах, но на средних и высоких только понижает его, также и разные составляющие модели могут иметь максимальный битрейт, до которого они полезны, но от использования которых на больших битрейтах следует отказаться.

Психоакустическая модель работает с учетом особенностей человеческого слуха. В том числе используется эффект маскирования слабого сигнала одного диапазона частот более мощным сигналом соседнего диапазона, когда он имеет место, т. е., если он превышает мощность сигнала интересующего диапазона или мощность сигнала в нем оказывается ниже определенного опытным путем порога слышимости, то для данного фрейма данный диапазон сигнала не кодируется. Также учитывается неспособность большинства людей различать сигналы, по мощности лежащие ниже определенного уровня, разного для разных частотных диапазонов. Также при мощном сигнале текущего фрейма у человеческого слуха возникает временное понижение чувствительности к сигналу следующего и т. д.

Говоря точнее, механизм подобных кодеров выполняет анализ кодируемого сигнала, в процессе которого определяются участки сигнала, в определенных частотных областях которых имеются неслышимые человеческому уху нюансы (замаскированные или неслышимые частоты), после чего происходит их удаление из оригинального сигнала.

Таким образом, степень сжатия оригинального сигнала зависит от степени его «упрощения»; сильное сжатие достигается путем «агрессивного упрощения» (когда кодер «считает» ненужными множественные нюансы), такое сжатие, естественно, приводит к сильному ухудшению качества, поскольку удалению могут подлежать не только незаметные, но и значимые детали звучания. При этом учитывается, что потеря одного бита ведет к внесению шума квантования величиной порядка 6 dB.

После завершения работы психоакустической модели формируется итоговый поток, который дополнительно кодируется по Хаффману, на этом кодирование завершается.

### **Замечание**

На практике схема кодирования несколько сложнее. Например, необходимо согласовываться с требованиями битрейта. В зависимости от кодера это приводит при повышении битрейта к разного рода релаксациям при отборе сохраняемой части исходного сигнала, а при понижении – наоборот, к ужесточению критериев.

Небольшая справка. Стандартные параметры записи аудио компакт-дисков следующие: частота дискретизации - 44.1 КГц, уровень квантования – 16 бит. Такие параметры соответствуют 65536 (216) уровням квантования амплитуды при взятии ее значений 44100 раз в секунду.

### **Способы кодирования стерео сигнала MP3**

В рамках MP3 кодирование стереосигнала допустимо четырьмя различными методами.

Dual Channel — каждый канал получает ровно половину потока и кодируется отдельно как моносигнал. Рекомендуются, главным образом, в случаях, когда разные каналы содержат принципиально разный сигнал — скажем, текст на разных языках. Выставляется в некоторых кодерах по требованию.

Stereo — каждый канал кодируется отдельно, но кодер может принять решение отдать одному каналу больше места, чем другому. Это может быть полезно в том случае, когда после отброса части сигнала, лежащей ниже порога слышимости или полностью маскируемой, оказалось, что код не полностью заполняет выделенный для данного канала объем, и кодер имеет возможность использовать это место для кодирования другого канала. В документации к mp3 замечено, что этим, например, избегается кодирование «тишины» в одном канале, когда в другом есть сигнал.

Joint Stereo (MS Stereo) — стереосигнал раскладывается на средний между каналами и разностный, при этом второй кодируется с меньшим битрейтом. Это позволяет несколько увеличить качество кодирования в обычной ситуации, когда каналы по фазе совпадают, но приводит к резкому его ухудшению, если кодируются сигналы, по фазе не совпадающие. В частности, фазовый сдвиг практически всегда

присутствует в записях, оцифрованных с аудиокассет, но встречается и на CD, особенно если CD был записан в свое время с аудиоленты.

Joint Stereo (MS/IS Stereo) вводит еще один метод упрощения стереосигнала, повышающий качество кодирования на особо низких битрейтах. Состоит в том, что для некоторых частотных диапазонов оставляется уже не разностный сигнал, а только отношение мощностей сигнала в разных каналах. Понятно, для кодирования этой информации употребляется еще меньший битрейт.

В отличие от всех предыдущих, этот метод приводит к потере фазовой информации, но выгоды от экономии места в пользу среднего сигнала оказываются выше, если речь идет об очень низких битрейтах. Но, как уже говорилось, при применении данного режима происходит потеря фазовой информации, также теряется любой противofазный сигнал. Также следует отметить, что увеличение числа каналов ведет к пропорциональному увеличению объема памяти, занимаемой звуковой информацией.

### **Подробнее о форматах WAVE-файлов**

Для хранения цифрового звука существует много различных способов. Как было отмечено ранее, оцифрованный звук представляет собой набор значений амплитуды сигнала, взятых через определенные промежутки времени. Таким образом, во-первых, блок оцифрованной аудио информации можно записать в файл «как есть», т. е. последовательностью чисел (значений амплитуды). Существуют следующие варианты сжатия:

- импульсно-кодовая модуляция (ИКМ)
- дифференциальная ИКМ (ДИКМ)
- адаптивная ДИКМ (АДИКМ)

Во-вторых, можно сжать или упростить данные так, чтобы они занимали меньший объем памяти. Тут имеются два пути.

*Кодирование данных без потерь* (lossless coding) — это способ кодирования аудио, который позволяет осуществлять стопроцентное восстановление данных из сжатого потока. К такому способу уплотнения данных прибегают в тех случаях, когда сохранение оригинального качества данных критично. Например, после сведения звука в студии звукозаписи, данные необходимо сохранить в архиве в оригинальном качестве для возможного последующего использования. Существующие сегодня алгоритмы кодирования без потерь (например,

Monkeys Audio) позволяют сократить занимаемый данными объем на 20–50%, но при этом обеспечить стопроцентное восстановление оригинальных данных из полученных после сжатия. Подобные кодеры – это своего рода *архиваторы данных* (как ZIP, RAR и др.), только предназначенные для сжатия именно аудио.

Имеется и второй путь кодирования — *кодирование данных с потерями* (lossy coding). Цель такого кодирования — любыми способами добиться схожести звучания восстановленного сигнала с оригиналом при как можно меньшем объеме упакованных данных. Это достигается путем использования различных алгоритмов, «упрощающих» оригинальный сигнал, выкидывая из него «ненужные» слабослышимые детали (данная психоакустическая модель была разобрана выше), что приводит к тому, что декодированный сигнал фактически перестает быть идентичным оригиналу, а лишь похоже звучит.

Методов сжатия, а также программ, реализующих эти методы, существует много. Наиболее известными являются MPEG-1 Layer I, II, III (последним является всем известный MP3), MPEG-2 AAC (advanced audio coding), Ogg Vorbis, Windows Media Audio (WMA), TwinVQ (VQF), MPEGPlus, TAC и пр. В среднем, коэффициент сжатия, обеспечиваемый такими кодерами, находится в пределах 10–14 (раз).

Часто кодек определяет и формат аудиофайла, и, соответственно, его расширение:

MPEG – «.mpa», «.mp3»,

RealAudio – «.ra», «.rm». и т.д.

### **MIDI-форма звука**

MIDI-форма звука была разработана для электронных музыкальных инструментов еще в 1982 г. и предназначалась для передачи цифровых сигналов – так называемых MIDI-последовательностей или MIDI-посылок – от одних инструментов к другим, например, от клавиатуры к синтезатору. Слово MIDI – сокращение от англ. Musical Instruments Digital Interface, т. е. цифровой интерфейс музыкальных инструментов.

По существу, MIDI-последовательность состоит из цифровых команд, посылаемых музыкальным инструментом. Они сообщают о произошедших событиях – нажатии или отпуске клавиши или педали, переключении регистра или инструмента и т. п. В этом смысле MIDI-последовательность сродни нотной записи. Устройство, которое



по этим командам может воспроизвести звук, называется MIDI-синтезатором. На звуковой плате компьютера имеется такой синтезатор, но можно подключить еще и внешний синтезатор.

Аппаратные синтезаторы различаются по методу синтеза. В настоящее время используются методы FM- и WT-синтезы. FM-синтез (от англ. Frequency Modulation – частотная модуляция) основан на использовании нескольких генераторов сигнала (операторов), обычно синусоидального, со взаимной модуляцией. Тембр звука получается искусственный. WT-синтез (от англ. Wave Table – таблица волн) основан на воспроизведении сэмплов – заранее записанных в WAVE-форме образцов звучания реальных инструментов (обычно одной ноты), которые перед воспроизведением должным образом преобразуются. Этим достигается большая реалистичность звучания классических инструментов, но нужна память для хранения сэмплов. Кроме того, компьютер может по MIDI-последовательности синтезировать аналоговый аудиосигнал программой-синтезатором (но для его воспроизведения все равно нужна звуковая плата).

Звуковая плата позволяет также вводить MIDI-последовательность от любого подключённого к ней музыкального инструмента и записывать ее в виде MIDI-файла (расширения .mid, .rmi), который потом можно проигрывать уже без всяких внешних инструментов и включать в любые мультимедиа-программы. MIDI-файлы в сотни раз компактнее, чем аналогичные WAVE-файлы, поэтому их часто используют для создания музыкального фона в мультимедиа-программах и Web-страничках.

Любой MIDI-синтезатор может воспроизводить звучание более 150 различных музыкальных инструментов, имеющих свои номера (по англ. patch) и названия. Проигрывание идет одновременно на 16 каналах (треках) – 16 инструментах (или, точнее, в 16 оркестровых группах, состоящих из одинаковых инструментов). На каждом канале одновременно может звучать несколько голосов – нот, извлекаемых на этом инструменте (их может быть до 32). Получается целый оркестр, в котором одновременно может звучать 512 голосов! Число голосов на одном канале характеризует полифонию синтезатора.

Создание MIDI-музыки на компьютере очень популярно среди музыкантов-профессионалов и любителей. Они используют специальные программы – музыкальные редакторы (или программы-секвенсоры), которые достигли высокого совершенства. В этих программах обычно

нотная запись преобразуется в MIDI-последовательность (по англ. — sequence) и наоборот, чем и объясняется название программ.

## 9.5. Звуковые платы

В современном компьютере звук воспроизводится и записывается с помощью звуковой карты, подключаемой либо встроенной в материнскую плату компьютера. Задача звуковой карты в компьютере — ввод и вывод аудио. Практически это означает, что звуковая карта является тем преобразователем, который переводит аналоговый звук в цифровой и обратно.

Типовая звуковая карта имеет в своем составе цифровой звуковой канал записи- воспроизведения, микшер, синтезатор и MIDI-порт. Цифровой аудиоканал, он же аудиокодек, обеспечивает возможность моно- и стереофонической записи и воспроизведения аудиофайлов с уровнем качества, начиная от уровня кассетного магнитофона и заканчивая уровнем аудио-CD и даже выше. Запись производится оцифровкой (аналогово-цифровым преобразованием) выборки мгновенного значения сигнала; современные карты позволяют принимать и цифровые аудиоданные. Оцифрованный звук может храниться в файлах, для которых обычно используется расширение .WAV1 (сокращенно от wave — волна). Размер файла зависит от длительности записи, разрядности преобразования, частоты квантования и количества каналов (моно- или стереозапись). Эти «волновые файлы» могут редактироваться программными средствами, которые обычно позволяют вывести на экран подобие осциллограмм записанных сигналов. При воспроизведении (playback) поток цифровых данных выводится на внешний интерфейс, аналоговый (линейный выход или выход усилителей на колонки или наушники) или цифровой.

Микшер с программным управлением обеспечивает регулировку входных и выходных сигналов, позволяя смешивать входные сигналы от нескольких источников (микрофона, CD, внешнего входа и синтезатора). В стереокарте (а монокарты уже давно не используют) каждый источник должен иметь отдельные регуляторы уровня для каждого канала. Внешне (в графической оболочке программного интерфейса) это может выглядеть и как общий регулятор уровня и регулятор баланса. Для монофонических источников (например, микрофона) помимо регулятора уровня есть регулятор панорамы, позволяющий балансировать уровни сигналов, посылаемых от данного ис-

точника в левый и правый стереоканалы. Физически это опять-таки могут быть просто два регулятора уровня для одного и того же сигнала. Дополнительно к микшеру карта обычно допускает регулировку тембра по низким и высоким частотам или даже имеет эквалайзер — многополосный регулятор тембра.

Синтезатор обеспечивает имитацию звучания музыкальных инструментов и воспроизведение различных звуков. Из множества *методов синтеза в звуковых картах* в основном используют два — *частотный и волновой*:

FM Music Synthesizer — синтезатор с частотной модуляцией (аббревиатура FM означает Frequency Modulation — частотная модуляция) — обеспечивает невысокое качество синтеза.

WTMusic Synthesizer — синтезатор с табличным синтезом (аббревиатура WT означает Wave Table — волновая таблица), хранит в своей памяти образцы сигналов натуральных инструментов. Волновые синтезаторы обеспечивают высокое качество синтеза, но поначалу они были заметно дороже.

Встроенный усилитель имеет мощность до 4 Вт на канал, хотя многие адаптеры обеспечивают мощность, достаточную только для наушников.

Колонки (speakers) для PC несколько отличаются от обычных бытовых акустических систем. Они, как правило, малогабаритные, поскольку предназначены для установки на столе по бокам от монитора. Малые габариты, конечно же, отражаются на качестве и выходной мощности. Хорошие колонки имеют специальный магнитный экран или улучшенную конструкцию магнитной системы динамиков, чтобы предотвратить воздействие магнитного поля на ЭЛТ-монитор. Сильное магнитное поле нарушает линейность развертки и сведение лучей на экране монитора. Ряд моделей «мультимедийных» мониторов оборудован встроенными акустическими системами. Активные колонки (active speakers) имеют встроенный усилитель, требующий внешнего (или батарейного) питания. Они могут иметь регуляторы громкости и тембра. Пассивные колонки встроенного усилителя не имеют, их мощность невелика. Есть модели колонок, режим работы которых (активный или пассивный) выбирается переключателем. Полоса частот и мощность обычных малогабаритных колонок недостаточны для воспроизведения в режиме Hi-Fi (High Fidelity — высокая достоверность звуковоспроизведения). Более качественные системы

имеют две колонки для средних и высоких частот и одну (большую) для низких. Для высококачественного воспроизведения лучше использовать внешний стереоусилитель с собственными акустическими системами или стереонаушники.

Наушники или усилитель можно подключать и к аудиовыходу привода CD/ DVD, что позволит прослушивать аудио-CD (но не CD-ROM с файлами .MP3) независимо от наличия звуковой карты. Регулятор уровня этого выхода (диск потенциометра или кнопки) расположен на лицевой панели привода. Там же в ряде моделей приводов имеются кнопки воспроизведения и выбора трека, позволяющие управлять проигрыванием без привлечения каких-либо программных средств.

Существуют звуковые устройства для шин USB и FireWire — колонки, микрофоны и другие приемники и источники сигналов. Они передают аудиопоток в цифровом виде (изохронная передача) и к обычным звуковым картам непосредственно не подключаются. Доставка аудиоданных к ним и от них осуществляется программно через контроллер соответствующей шины, имеющийся на современных системных платах. При наличии достаточно мощного процессора такие устройства позволяют обходиться и без звуковой карты, реализуя все перечисленные функции программными средствами. Однако применение звуковой карты расширяет возможности аудиосистемы и снижает загрузку процессора.

Для подключения электромusикальных инструментов звуковые карты имеют порт MIDI. Устройством ввода могут служить специальные MIDI-клавиатуры (как на клавишных музыкальных инструментах), устройством вывода — синтезатор звуковой карты или внешний синтезатор, подключаемый к порту MIDI. Компьютер в такой системе используется как мощное средство создания, редактирования и хранения музыкальных произведений. MIDI-интерфейс имеют многие профессиональные и полупрофессиональные клавишные синтезаторы.

На рис. 9.6 показана принципиальная схема звуковой платы. На схеме видны основные потоки звуковой информации, ее входы и выходы, а также модули, осуществляющие те или иные преобразования звуковых данных.

Звуковые данные циркулируют либо в аналоговой форме, либо в WAVE-форме, либо в MIDI-форме.

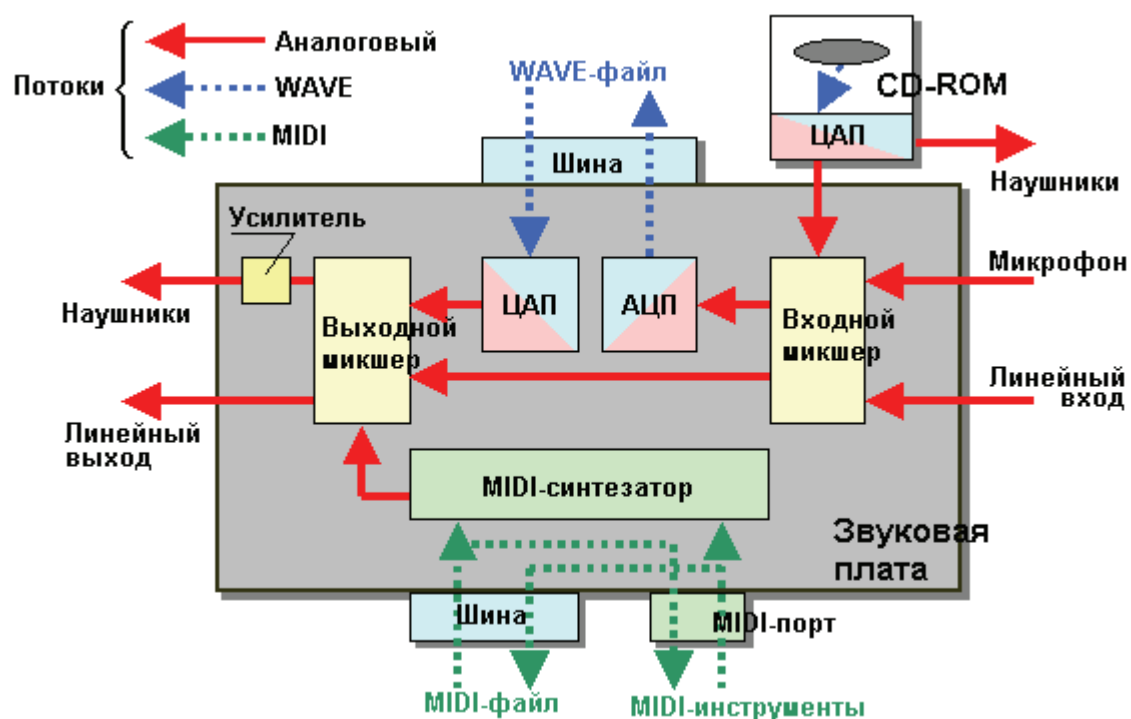


Рис. 9.6. Принципиальная схема звуковой платы.

Внешний (по отношению к звуковой плате) аналоговый сигнал поступает по микрофонному входу – от микрофона, линейному входу – от линейного аудиовыхода любого теле-радио устройства, внутреннему аналоговому аудиокабелю, идущему от аналогового аудиовыхода дисковода CD-ROM (как показано на схеме, такой дисковод имеет собственный ЦАП, позволяющий ему автономно, т. е. без участия звуковой платы проигрывать цифровые аудиодиски).

Аналоговых выхода на плате два:

- *наушниковый* — на наушники или пассивные акустические колонки с предварительным усилением с помощью внутреннего усилителя (не очень качественного);

- *линейный* (без усиления, но с сохранением качества) — на линейный аудиовход любого воспроизводящего или записывающего теле-радио устройства (активные акустические колонки, магнитофон и т.п.).

Шина данных компьютера, к которой подключается звуковая плата, передает на плату и принимает от неё цифровые аудиоданные в WAVE- и MIDI-форме.

MIDI-порт служит для ввода и вывода MIDI-последовательностей с помощью подключаемых к нему внешних источников (например,

MIDI-клавиатуры) и воспроизводящих устройств (MIDI-синтезаторов). Входящая извне через MIDI-порт MIDI-последовательность может быть по шине передана некоторой программе для записи в MIDI-файл, а также преобразуется внутренним синтезатором в аналоговый сигнал, который может быть тут же озвучен. MIDI-последовательность, поступающая на плату по шине из MIDI-файла, может быть через MIDI-порт передана внешнему синтезатору для воспроизведения, а также преобразована в аналоговый сигнал внутренним синтезатором платы.

Входной микшер соединяет (микширует) все поступающие к нему аналоговые потоки в один, который передается на оцифровку и запись в WAVE-файл. Выходной микшер соединяет все поступающие к нему аналоговые потоки в один, который передается на аналоговые выходы. Оба микшера могут управлять уровнем каждого канала каждого своего входного потока и отключать ненужные потоки. Работают они одновременно и, как правило, независимо друг от друга.

Если описывать упрощенно, то работа звуковой карты может быть пояснена следующим образом. Если на вход звуковой карты подан аналоговый сигнал и карта включена (программно) в режим, то сначала входной аналоговый сигнал попадает в аналоговый микшер, который занимается смешением сигналов и регулировкой громкости и баланса. Микшер необходим, в частности, для предоставления возможности пользователю управлять уровнями. Затем отрегулированный и сбалансированный сигнал попадает в аналогово-цифровой преобразователь, где сигнал дискретизируется и квантуется, в результате чего в компьютер по шине данных направляется бит-поток, который и представляет собой оцифрованный аудио сигнал.

Вывод аудио информации почти аналогичен вводу, только происходит в обратную сторону. Поток данных, направленный в звуковую карту, преодолевает цифро-аналоговый преобразователь, который образует из чисел, описывающих амплитуду сигнала, электрический сигнал. Полученный аналоговый сигнал может быть пропущен через любые аналоговые тракты для дальнейших преобразований, в том числе и для воспроизведения. Надо отметить, что если звуковая карта оборудована интерфейсом для обмена цифровыми данными, то при работе с цифровым аудио никакие аналоговые блоки карты не задействуются.

### Аудиокодек AC'97

Архитектура звуковой подсистемы аудиокодек AC'97 состоит из звукового кодека и цифрового контроллера. Обе части соединяются вместе с помощью цифрового интерфейса AC-link.

Аудиокодек AC'97 представляет собой довольно универсальное решение для звуковых карт и модемов, предложенное фирмой Intel в 1997 г. Версия 2.2, в которой введена поддержка интерфейса S/PDIF, опубликована в 2000 году. За несколько лет использования аудиокодек AC'97 практически вытеснил дешевые звуковые карты, обеспечивая качество звука, удовлетворяющее большинство пользователей. Собственно кодек AC'97 представляет собой микросхему в 48-выводном корпусе с унифицированным назначением выводов, которая выполняет аналого-цифровые и цифро-аналоговые преобразования, микширование, аналоговый ввод и вывод сигналов для аудиосистемы и модема. Кодек обеспечивает необходимые преобразования частот дискретизации. Состав функций, реализуемых кодеком, в зависимости от версии микросхемы может варьироваться — от минимально необходимых до расширенных. Кодек подключается к компьютеру по специальному цифровому интерфейсу AC-Link, предоставляемому специализированным *контроллером интерфейса* (AC'97 Digital Controller). Контроллер подключается к шине PCI (как ведущее устройство) или встраивается в чипсет системной платы. В обязательные (минимальные) задачи контроллера входит доставка цифровых аудиоданных в *формате PCM* аудиокодеку из памяти и в память от кодека. Для этого, как правило, контроллер должен быть многоканальным контроллером шины PCI (bus master). В память аудиоданные для кодека поступают под управлением центрального процессора от различных источников: из файлов, с внешних цифровых интерфейсов, а также генерируемых программно в реальном времени (например, от игр, от программных декодеров MPEG и т. п.). Из памяти данные от кодека различным потребителям (в файлы, на внешние интерфейсы) доставляет также программа центрального процессора. Более мощный контроллер помимо самой доставки обеспечивает и аппаратное ускорение генерации PCM-данных, например, при декодировании MPEG, формировании 3D-звука, синтезе звуков, а также кодирование PCM-данных в более плотные форматы.

Кодек AC'97 по отношению к контроллеру является подчиненным устройством. Разделение аудиосистемы на две части позволяет собирать ее в соответствии с предъявляемыми требованиями и при необходимости без особых проблем модернизировать.

### **Стереофоническое и объемное воспроизведение**

Для обычной стереофонии достаточно двух колонок, расположенных перед слушателем, подавляющее большинство звуковых карт имеют стереофонический аудиовыход. Некоторыми ухищрениями перекрестного смешивания сигналов удалось расширить зону стереоэффекта, но добиться объемности звучания таким путем не удавалось.

В системе объемного, или обволакивающего (surround), звучания Dolby Surround Pro Logic, применяемой в «домашнем кинотеатре» с аналоговой записью звука, используются 4 воспроизводящих канала усилителей. Здесь слушателя окружают колонками со всех сторон: перед ним располагают три колонки (слева, справа и по центру), а за ним еще две тыловые (слева и справа). Для каждой из фронтальных колонок задействуется собственный широкополосный канал, а обе тыловые колонки используют сигнал одного канала с ограниченной полосой пропускания (100–7000 Гц). Все 4 канала «упакованы» в стереосигнал с обычными параметрами каналов. Этот стереосигнал может храниться и передаваться по любому стереотракту — компакт-дисков, радио FM, стерео в телевидении и видеокассетах. На обычных моно- и стереосистемах этот сигнал воспроизводится естественным для них способом, но с помощью специального декодера Dolby Surround Pro Logic он раскладывается на вышеуказанные 4 канала.

Для цифровых систем фирма Dolby разработала систему Dolby Digital, она же AC-3 и 5.1, в которой передается (хранится) в сжатом виде информация шести каналов — пяти широкополосных и одного низкочастотного. Здесь слушателя также окружают колонками со всех сторон: три колонки спереди (слева, справа и по центру), две тыловые (слева и справа) и еще одна колонка, низкочастотная, называемая сабвуфером (subwoofer), располагается за спиной (хотя ее положение относительно произвольно). Самая мощная колонка в этой системе — сабвуфер — центральная колонка несколько мощнее (и «широкополоснее») боковых. Тыловые колонки (rear speakers) могут иметь сравнительно небольшую мощность и размеры. В отличие от предыдущей системы здесь сигнал для каждой колонки передается собственным каналом, и не в аналоговом, а в цифровом виде. Как и положено фирме Dolby, в системе приняты эффективные способы шумоподавления. Эта полная схема поддерживается цифровым форматом и называется «5.1». Такая система воспроизведения устанавливается в современных кинотеатрах с «цифровым» звуком.



Название AC-3 означает «аудиокодек-3». Кодер Dolby Digital упаковывает 5 каналов с полосой 20–20 000 Гц ( $\pm 0,5$  дБ, завал -3 дБ на частотах 3 и 20 300 Гц) и 1 канал с полосой 20–120 Гц. Входные сигналы могут иметь разрядность 20 бит и более и частоту дискретизации 32; 44,1 или 48 кГц. В зависимости от требований к качеству и числу каналов поток данных на выходе кодера имеет скорость от 32 Кбит/с (моно) до 640 Кбит/с.

### **Трехмерный звук**

В компьютерной аудиотехнике, как и в графике, мощным двигателем прогресса являются игры. Первые игры довольствовались заранее запрограммированными звуками, издаваемыми программно-управляемым динамиком. Потом игры стали озвучиваться звуковыми картами, воспроизводящими цифровые записи и синтезирующими звук (в основном FM). Параметрами звуков уже стали управлять динамически в зависимости от сюжета. Здесь уже появилась стереофония с ее возможностью панорамирования источников звуков. Но с ростом возможностей цифровой обработки растут и потребности, в результате чего появилось несколько разработок трехмерного звука (3D Sound, или 3D Audio). Их целью является создание у слушателя впечатления локализации источников в окружающем его трехмерном пространстве — не просто линейной стереопанорамы, а по трем координатам: лево-право, перед-зад, верх-низ.

Разработчики 3D-звука исходили из того, что сферическую локализацию человек ухитряется выполнять по сигналам от пары своих ушей — значит, можно найти способ формирования этих сигналов даже через пару выходных акустических устройств. Алгоритмы обработки цифрового аудиопотока с целью создания иллюзии трехмерной локализации при выводе через 2-, 4-, 6- и даже 8-канальные акустические системы строятся на основе психоакустических моделей восприятия звука человеком. Конечно же, как и в случае вывода трехмерной графики через плоский экран монитора, это все-таки иллюзия, но ее убедительность совершенствуется по мере усложнения алгоритмов. В значительной степени эта иллюзия подкрепляется озвучиваемым изображением (большую часть информации человек воспринимает через зрение).

## 10. ПЕРИФЕРИЙНЫЕ УСТРОЙСТВА.

### 10.1. Клавиатуры.

#### Типы клавиатур:

- типы клавиатур;
- мембранные;
- полумеханические;
- механические.

Название *мембранных клавиатур* происходит оттого, что при нажатии клавиши замыкаются две мембраны. Возврат клавиши осуществляется резиновым куполом (с шахтой в центре). Мембраны обычно выглядят как диски на пластиковой пленке, выполненной печатным способом. Для разделения мембран служит промежуточная пленка с отверстиями. Поэтому в предложениях часто пишут «пленка» или «пленочная клавиатура». Так как мембраны находятся на внутренних сторонах пленок, то конструкция хорошо защищена.

В *полумеханических клавиатурах* используются более долговечные (до 50 – 100 млн. нажатий) и не протирающиеся металлические контакты, в дорогих моделях они могут быть позолоченными. Все это размещается на печатной плате. Клавиша возвращается резиновым куполом.

В *механических клавиатурах* клавиша возвращается пружиной. Минусы такого механизма заключаются в отсутствии герметичности и дороговизне. Плюсом является долговечность и надежность, особенно когда контакты позолочены, а также отсутствие усталости (то есть сопротивление нажатию практически не зависит от количества нажатий, чего нельзя сказать о механических и полумеханических клавиатурах).

Клавиатуры бывают с кликом или без. В буквальном переводе клик (click) – щелчок. Реализуется дугообразной тонкой пластиной под клавишей, которая рывком прогибается. Клик позволяет точно чувствовать, что клавиша нажата, и не пропускать буквы при быстром наборе.

Есть еще одна разновидность клавиатур — *сенсорные клавиатуры*. Принцип ее действия основан на усилении разности потенциалов, приложенной к одному элементу. Количество этих элементов соответствует числу клавиш. В качестве чувствительных элементов ис-

пользуются токопроводящие контакты, выполненные в виде двух пластин, разделенных небольшим зазором. В момент касания пальцем контактных площадок статистический потенциал усиливается соответствующей схемой, на выходе которой формируется сигнал, аналогичный сигналу обычной клавиатуры. Сенсорные клавиатуры самые долговечные, потому что в них нет механических элементов, но за счет этого они заметно дороже, чем обычные. Так как электроника покрыта слоем полимерной пленки (на ней также нанесены символы), то сенсорная клавиатура почти полностью защищена от внешних воздействий.

### **Дополнительные возможности**

До появления Windows все клавиатуры имели 101 клавишу. Затем были добавлены еще три клавиши: две клавиши вызова меню «Пуск» и одна клавиша вызова контекстного меню. Бурное развитие Интернета привело к появлению соответствующих кнопок. Обычно они выполняют следующие функции: подключиться к Интернету, выйти на заданный сайт. Число таких кнопок различно: от нуля до двух десятков.

Мультимедийные кнопки. Эти кнопки позволяют управлять проигрывателями: увеличить или уменьшить громкость, выключить звук, перейти на следующую/предыдущую дорожку, начать проигрывание, сделать паузу, закончить проигрывание, выдвинуть диск.

Обычно дополнительные кнопки реализуются физически, что увеличивает размеры и цену клавиатуры. Но есть оригинальное решение, когда вводится только одна кнопка – переключатель MF (Multimedia Function), которая переводит в мультимедийные 12 функциональных клавиш. Такие клавиатуры есть, например, у компании BTC. У Genius кнопка носит название EasyKey, но смысл такой же. Есть еще клавиатуры с различными дополнительными устройствами ввода.

## **10.2. Мыши**

Наряду с клавиатурой, мышь тоже является важнейшим устройством ввода.

Принцип действия мышей очень простой: два датчика отслеживают перемещение мыши соответственно по горизонтали и по верти-

кали, и на основе их данных драйвер мыши формирует движение курсора на экране.

### Типы мышей

Существует три типа мышей.

– *Механическая*, в ней основным элементом являются датчики, отслеживающие движение шарика. Датчики механические, отсюда и название мыши.

– *Оптомеханическая*. Аналогична механической, но движение шарика отслеживаются оптическими датчиками. Такая мышь довольно надежна, универсальна (может работать на любой ровной поверхности, дешева, поэтому получила очень большое распространение, и обычно используются именно она.

– *Оптическая* мышь посылает луч на непрозрачную поверхность, а после отражения луч поступает обратно в мышь и там анализируется электроникой, которая в зависимости от характеристик полученного сигнала и отслеживает два направления движения мыши, основываясь либо на углах падения, либо на каких-либо других признаках. Преимущество такой мыши — очень высокая надежность, достоверность.

Разрешение мыши измеряется в dpi и показывает, сколько отсчетов (импульсов, на которые разбивается пройденное расстояние) совершает мышь при прохождении одного дюйма. Естественно, чем больше, тем лучше, так как большее разрешение позволяет более точно позиционировать курсор, а движение курсора становится более плавным. Нормальное разрешение мыши составляет 200 – 900 dpi. При этом точность, естественно, зависит от разрешения экрана монитора и выбранной скорости движения.

Оптические мыши могут иметь разрешающую способность более 2000 dpi, поэтому при их использовании курсор можно очень точно позиционировать даже при больших разрешениях экрана и максимальной скорости его перемещения. Другое дело, сколько отсчетов мышь может сделать за секунду. Это уже зависит не от мыши, а от интерфейса. У мышей с интерфейсами COM и PS/2 это число равно 40, а у USB больше 100. Этот параметр указывает, насколько плавно указатель мыши может двигаться при быстрых ее перемещениях (именно быстрых перемещениях самой мыши).

### 10.3. Сканеры

Сканер является устройством, которое позволяет оцифровывать изображения фотографий, рисунков, книг, монет, печатных плат и иных более или менее плоских объектов.

#### **Способы формирования изображения**

Способов формирования изображения три. Они соответственно осуществляются прибором с зарядовой связью (ПЗС), фотоэлектронным умножителем (ФЭУ) или так называемым контактным датчиком. ПЗС — это твердотельный электронный компонент, состоящий из множества миниатюрных датчиков, преобразующих падающий на них свет в пропорциональный его интенсивности электрический заряд. В основу ПЗС положена чувствительность проводимости  $p-n$  — перехода обыкновенного полупроводникового диода к степени его освещенности. На  $p-n$  — переходе создается заряд, который уменьшается со скоростью, зависящей от освещенности. Чем меньше заряд, тем больше ток через диод.

В зависимости от типа сканера ПЗС могут иметь различную конфигурацию. При линейном способе считывания микродатчики ПЗС размещаются на кристалле в одну линию (для трехпроходного сканирования) или три линии (для однопроходного сканирования). Дешевые сканеры обрабатывают оригинал за три прохода, более дорогие и качественные используют второй способ, который лучше, т. е. более производительный. Способ относительно недорог и позволяет сканировать с высоким качеством, поэтому ПЗС – сканеры являются наиболее распространенными. ПЗС также широко используется в цифровых камерах.

В барабанных сканерах в качестве светочувствительных приборов применяются фотоэлектронные умножители. Источником света служит ксеноновая или вольфрамогалогенная лампа. Ее излучение с помощью системы конденсаторных линз и волоконной оптики фокусируется на чрезвычайно небольшой области сканируемого объекта. ФЭУ осуществляют электронное усиление отраженного от оригинала света (чем больше интенсивность света, тем больше электрический сигнал). Попадая на катод ФЭУ, свет выбивает из него электроны, которые, проходя через пластины динодов, вызывают вторичную электронную эмиссию. Коэффициент усиления зависит от свойств материала и количества динодов. Напряжение, прямо пропорциональное

освещенности катода ФЭУ, снимается с анода и затем преобразуется в цифровой код. Фотоэлектронные умножители дороги, поэтому большой популярности не получили.

Еще одна, недавно появившаяся технология, называется CIS (Contact Image Sensor). Сканирующая головка, выполненная по CIS – технологии, имеет 3 основных компонента — источник света, специальную цилиндрическую линзу (или набор линз), а также приемный элемент с электронной начинкой для формирования выходного аналогового сигнала и синхронизации с другими компонентами сканера, выполненными на единой печатной плате.

В качестве источника света у большинства сканеров с контактным датчиком используются светодиоды, излучение которых отражается от сканируемого изображения, и, пройдя через линзу, фокусируется на датчике изображения, который представлен фототранзисторами, выполненными по технологии MOS. На выходе получается аналоговый сигнал, который усиливается и подается на вход аналого-цифрового преобразователя. В том случае, если необходимо сканировать цветное изображение, источником света служат светодиоды трех основных цветов (красный, зеленый, синий).

### **Типы сканеров**

Для пользователя наиболее важен тип самого сканера. Сканеры бывают:

- ручные;
- планшетные;
- барабанные;
- роликовые;
- проекционные.

*Ручные сканеры* самые простые и поэтому дешевые. В небольшом корпусе находится лишь датчик и источник света, а перемещение этого агрегата относительно объекта осуществляется вручную. Естественно, такая технология не является хорошей ни в одном отношении, потому что само по себе качество сканирования невысоко, и оно еще больше ухудшается неравномерностью перемещения сканера вдоль оригинала. Если нужно отсканировать большой оригинал, то приходится делать это по частям, а затем сшивать полученные кусочки, что, естественно, очень неудобно.

К категории *планшетных сканеров* относятся сканеры, в которых сканирующая головка перемещается вдоль оригинала с помощью

шагового двигателя. Как и у копировальных аппаратов, у таких сканеров есть крышка с зеркалом (у сканеров с контактным датчиком внутренняя поверхность, наоборот, темного цвета), что позволяет сканировать самые разнообразные объекты. Обычно планшетный сканер предназначен для сканирования листов формата А4, но есть сканеры и больших размеров.

В *барабанных сканерах* оригинал закрепляется на поверхности из прозрачного материала цилиндрической формы (барабане), которая укреплена на массивном основании, которое обеспечивает необходимую устойчивость. Барабан вращается с высокой скоростью (от 300 до 1350 об/мин), а находящийся рядом с ним сканирующий датчик через крошечную конусообразную апертуру пиксель за пикселем считывает изображение. В барабанных сканерах используются датчики на основе ФЭУ. Недостаток барабанных сканеров в том, что невозможно сканировать объекты, отличные от гнущегося листа, а также в том, что стоимость их намного превышает стоимость аналогичных планшетных сканеров. Зато у них самое высокое качество полученного изображения.

Еще один довольно редко встречающийся вид — *роликовые сканеры*. В них оригинал пропускается с помощью системы роликов (как в принтерах) и считывается обычной ПЗС — матрицей. Аналогом этих устройств является факс.

*Проекционные сканеры* работают приблизительно так же, как фотоаппарат (кстати, цифровую фото — или видеокамеру тоже можно рассматривать как сканер). Оригиналы располагаются на подставке под сканирующей головкой на расстоянии около 30 см. Внешнего освещения вполне достаточно, поэтому собственная подсветка необязательна. Механизм поворота датчика внутри головки направляет его последовательно на каждую линию объекта. Значительной популярностью проекционные сканеры не получили.

Для любого сканера независимо от его типа важно разрешение, которое он поддерживает. Оно может колебаться от 100 – 150 dpi до нескольких тысяч dpi.

Часто разрешение по горизонтали и вертикали неодинаково. Меньшая цифра в таком случае, как правило, обозначает шаг двигателя, а большая — разрешение самого сканирующего элемента. Например, 1200x600 dpi означает, что датчик способен отсканировать оригинал с разрешением 1200 dpi, а шаг двигателя, который перемещает этот датчик, ограничен 1/600 дюйма.

Стоит различать истинное разрешение и разрешение с учетом интерполяционных возможностей устройства. Истинное разрешение — разрешение, с которым сканер может действительно обработать оригинал. Но с помощью интерполяции (смешивании нескольких отдельных элементов с целью получить новый, но более объективный) его можно повышать. Например, если два соседних пикселя имеют соответственно цветовое насыщение 36 и 88, то предполагается, что пиксель, стоящий между ними, имеет насыщенность 62. Таким образом, разрешение повышается вдвое.

#### 10.4. Видеопроекторы

Видеопроекторы являются средством вывода графической информации.

Первые появившиеся проекторы имели в своем составе обычный кинескоп, изображение с которого с помощью несложной оптики проецировалось на экран. Затем появились проекторы, основанные на *технологии LCD*, которые работали как обычный LCD – монитор, изображение с которого тоже проецировалось на экран.

Интерес представляют проекторы, в основу которых заложена *технология DLP* (Digital Light Processing), изобретенная компанией Texas Instruments. Благодаря ей оказалось возможным создавать очень небольшие, очень легкие (около 3 кг) и, тем не менее, достаточно мощные (более 1000 ANSI Lm) мультимедиапроекторы.

В 1987 г. Dr. Larry J. Hornbeck изобрел *цифровое мультизеркальное устройство* (Digital Micromirror Device, или DMD). Суть изобретения состояла в отказе от гибких зеркал в пользу матрицы жестких зеркал, имеющих всего два устойчивых положения. Именно эта матрица является ключевым элементом проекторов, созданных по технологии DLP. Матрица состоит из микроскопических зеркал (DMD – элементов) из алюминиевого сплава, обладающего очень высоким коэффициентом отражения. Каждое зеркало крепится к жесткой подложке, которая через подвижные пластины соединяется с основанием матрицы. Под противоположными углами зеркал размещены электроды, соединенные с ячейками памяти. Под действием электрического поля подложка с зеркалом принимает одно из двух положений, отличающихся точно на 20° благодаря ограничителям, расположенным на основании матрицы.



Два этих положения соответствуют отражению поступающего на матрицу светового потока, соответственно, в объектив или в эффективный светопоглотитель, обеспечивающий надежный отвод тепла и минимальное отражение света. Матрица зеркал вместе с элементами SRAM и составляют DMD – кристалл – основу технологии DLP. Его размеры невелики. Площадь каждого зеркала матрицы составляет 16 микрон и менее, а расстояние между зеркалами около 1 микрона. Количество зеркал в матрице соответствует максимальному ее разрешению (а следовательно, и разрешающей способности проектора). Это разрешение обычно составляет от 800·600 до 1280·1024 точек для матриц разного типа и стоимости.

Изображение получается путем освещения матрицы. Как уже говорилось, зеркало может принимать два положения: положение, при котором свет полностью отражается и фокусируется с помощью оптической системы на экран, и положение, при котором свет поглощается светопоглотителем. Полутона обеспечиваются колебанием зеркала в этих двух положениях. Например, если зеркало одинаковое количество времени будет проводить в обоих положениях, то получится уменьшение яркости вдвое. Но картинка еще не цветная. Если поставить на пути луча диск с тремя светофильтрами, который будет вращаться с заданной частотой, и изменять колебания зеркал в соответствии с положением диска, то получится обычное трехкомпонентное цветное изображение. Все это возможно только благодаря одному из недостатков нашего зрения, а именно инерционности зрительных ощущений. Такой способ применяется в *одноматричных проекторах*, самых простых и распространенных.

Между источником света и матрицей помещается вращающийся диск с цветными светофильтрами — синим, зеленым и красным. Частота вращения диска определяет частоту кадров.

Другим, более сложным типом проекторов является *трехматричный проектор*, где свет расщепляется на три цветовых потока и отражается сразу от трех матриц. Такой проектор имеет самый чистый цвет и частоту кадров, не ограниченную скоростью вращения диска, как у одноматричных проекторов.

Есть еще промежуточный тип — *двухматричный проектор*. В данном случае свет расщепляется на два потока: красный отражается от одной DMD – матрицы, а синий и зеленый – от другой. Светофильтр, соответственно, удаляет из спектра синюю либо зеленую составляющие поочередно. Двухматричный проектор обеспечивает

промежуточное качество изображения по сравнению с одноматричным и трехматричным.

DLP — технология является на сегодня самой передовой в области «проекторостроения». Миниатюрные размеры DMD — матрицы позволяют создавать очень небольшие проекторы

## 10.5. Принтеры

*Ударные принтеры* являются прародителями всех сегодняшних принтеров, и некоторые их типы до сих пор успешно используются.

### Типовой принтер

Основным элементом является типовый диск в виде ромашки, на конце лепестков которой выштампованы символы. Диск вращается вокруг своей оси перпендикулярно бумаге; между ним и бумагой находится красящая лента. В соответствующий момент ударный механизм бьет по концу лепестка, и в результате мы получаем на бумаге оттиск нужного символа. Меняя диски, можно изменять печатаемые символы и шрифт.

Достоинство этого принтера в том, что он дает очень чистое и качественное изображение символов — теоретически больше никакой ныне существующий принтер не сможет лучше напечатать текст, на практике же качество вполне сопоставимо с хорошими лазерными принтерами. Но зато ничего кроме символов, жестко заданных типовым диском, напечатать не удастся. Также ничего не получится, если выделить часть текста каким-либо другим шрифтом (только если установить принтер и заменить диск, что, конечно же, очень не удобно). Кроме того, типовые принтеры обеспечивают очень низкую скорость печати.

### Игольчатые принтеры

Дальнейшим развитием ударной технологии стал *игольчатый принтер*. Принцип его действия похож на принцип действия типового принтера, только здесь уже используется головка с иглами, перемещающаяся по горизонтальной штанге. Из точек, оставляемых иглами с помощью красящей ленты, и формируется желаемое изображение. В то время как струйные принтеры были еще очень несовершенны, а лазерная печать была слишком дорога, игольчатые принтеры де-факто стали стандартом для принтеров, используемых в IBM PC.

Так как головка по сути представляет собой матрицу иглоков, то игольчатые принтеры обычно называют матричными.

### **Струйные принтеры**

Первой фирмой, изготовившей струйный принтер, является Hewlett-Packard (HP). Долгое время струйные принтеры оставляли желать лучшего как в плане качества, так и цены, но со временем технология развилась и сейчас это, пожалуй, самый дешевый и распространенный тип принтеров. Принцип их работы схож с матричными принтерами но с тем отличием, что вместо иглоков и красящей ленты используются тонкие сопла, которые «выплывают» чернила. Струйные принтеры отличаются главным образом методом переноса чернил на бумагу. Их всего три.

– Пьезоэлектрический метод. В каждое сопло установлен пьезокристалл, связанный с диафрагмой. Под воздействием электрического поля происходит деформация пьезоэлемента. Пьезоэлемент, сжимая и разжимая трубку, наполняет сопло чернилами, при этом часть их попадает обратно в резервуар. Та же краска, которая выдавилась наружу, оставляет на бумаге точку. Такой способ активно используется в принтерах Epson.

– Метод газовых пузырей. Известен также под названием *Bubblejet*. В этом случае каждое сопло оборудовано электрическим нагревательным элементом, который при пропускании через него тока за несколько микросекунд нагревается до температуры около 500°C. Чернила закипают и образуют пузырьки, которые стремятся вытолкнуть часть краски наружу. Затем нагревательный элемент отключается и остывает, газовый пузырь уменьшается в объеме и через отверстие из резервуара поступает новая порция чернил. Т. к. здесь отсутствуют какие-либо механические элементы, то принтеры, сделанные по этой технологии, самые долговечные и имеют высокое разрешение. Небольшим недостатком является то, что полутонные изображения получаются несколько расплывчатыми. Способ используется фирмой Canon.

– HP разработала собственный метод, называемый методом *drop-on-demand*. Как и в предыдущем методе, для подачи чернил в сопла используется нагревательный элемент. Однако вместе с ним дополнительно работает еще и специальный механизм. Такой способ обеспечивает более быстрое выпрыскивание чернил, что повышает скорость печати, а также позволяет несколько повысить контраст-

ность изображения, что положительно сказывается на качестве печати полутоновых изображений.

*Способа подачи чернил в головку* два: либо головка объединена с резервуаром для чернил (замена картриджа происходит одновременно с заменой головки), либо используется отдельная емкость, а краска передается в головку посредством системы капилляров. Последний способ, возможно, лучше, т. к. тогда картридж стоит существенно дешевле. Кстати, он и используется в большинстве струйных принтеров.

Большинство принтеров могут печатать цветное изображение. Цветное изображение формируется наложением друг на друга нескольких основных цветов: *Cyan, Magenta, Yellow* и *Black*(CMYK). Добавление черного цвета продиктовано тем, что первые три при сложении все же не дают чистого черного цвета, как это должно быть. В хороших струйных принтерах применяется большее количество цветов (до шести), что позволяет повысить качество цветопередачи (тогда в названии принтера, как правило, присутствует слово «Photo»). Лазерные принтеры все еще дороги, поэтому дешевый метод смешения разноцветных чернил, который тем не менее может обеспечить приемлемое для многих целей качество, получил достаточно широкое распространение. Однако качество цветной печати у струйных принтеров все равно хуже, поэтому как полноценную альтернативу лазерным или специальным принтерам для цветной печати их рассматривать нельзя.

Струйные принтеры придирчивы к качеству бумаги, поэтому по возможности следует использовать более качественную бумагу. Хорошее качество цветных изображений достижимо только на специальной фотобумаге, предлагаемой фирмой-производителем принтера (иногда хороший результат дает и бумага другой фирмы). Стоит фотобумага довольно дорого, так что вряд ли струйный принтер подходит для серьезной цветной печати. А вот черно-белое изображение на уровне, и качество текста или одноцветных графических объектов сравнимо с качеством лазерных принтеров.

Максимальное разрешение у струйных принтеров достаточно высокое. Обычно оно не менее 300·300 dpi (у самых дешевых моделей), у более качественных 600·600 dpi. Лучшие принтеры могут иметь разрешение порядка 1440·720 dpi или даже больше. Принтеры с более высоким разрешением обладают улучшенным качеством печати, особенно черно-белых изображений. Следует заметить, что максимальное разрешение при цветной печати, как правило, раза в два меньше. Причем оно (впрочем, это касается всех принтеров) дости-

жимо только при самом качественном режиме, что замедляет скорость. Последняя у струйных принтеров больше, чем у игольчатых, и составляет (в качественном режиме) около 3 стр. в минуту. Высокопроизводительные устройства могут иметь скорость 8-9 страниц в минуту, но и стоят они, естественно, подороже. Цветная печать длится значительно дольше, и лист формата А4 принтер может печатать до нескольких минут. Это еще один аргумент в пользу того, что струйные принтеры малопригодны для цветной печати.

Струйные принтеры могут работать с нестандартным форматами бумаги (например, с конвертами), а также печатать на пленке, ткани и т. д. Последние материалы, правда, хорошо поддерживаются только достаточно дорогим моделями. Небольшое количество принтеров может печатать на рулонной бумаге. Практически все принтеры рассчитаны на формат А4. Есть также широкоформатные принтеры (как цветные, так и черные), предназначенные для печати чертежей, графиков и прочего. Пожалуй, это единственная область, где струйным принтерам не нашлось достойного конкурента. Графопостроители (плоттеры), которые использовались раньше, сегодня встречаются очень редко из-за их высокой стоимости и ограниченности в применении.

### **Лазерные принтеры**

Лазерный принтер является ближайшим «родственником» ксерокса, в нем используется тот же самый принцип. Этот принцип позволяет достичь довольно высокого качества печати за небольшие деньги, что позволило лазерным принтерам достичь на рынке принтеров практически доминирующего положения. Кроме более высокого качества печати (особенно цветной) по сравнению со струйными принтерами, лазерные также выгодно отличаются хорошей производительностью — скорость печати раза в три выше по сравнению со струйными, а также себестоимостью отпечатка: стоимость черно-белого изображения ниже где-то в два раза, а цветного чуть ли не в 5 раз или даже больше. Шумят лазерные принтеры не больше, чем струйные, а в режиме off-line их характерное жужжание еще меньше. Пожалуй, единственным их серьезным недостатком являются большая масса (до сотни килограммов) и внушительные габариты. Хотя этими особенностями обладают только цветные принтеры, а черные по этим параметрам вполне сопоставимы со струйными. Правда, и стоят эти устройства куда дороже струйных принтеров, но, если

учесть стоимость их эксплуатации и качественные характеристики, то затраты на лазерный принтер в большинстве случаев вполне оправданы.

Важнейшим конструктивным элементом принтера является вращающийся барабан, который и отвечает за перенос изображения на бумагу. Барабан выполнен в виде металлического цилиндра, покрытого тонкой пленкой фотопроводящего полупроводника (обычно используется оксид цинка). Во время печати по поверхности барабана с помощью тонкой проволоки или сетки, называемой коронирующим проводом (когда на этот провод подается высокое напряжение, вокруг него возникает светящаяся ионизированная область — корона), распределяется статический заряд. Лазер, управляемый соответствующей электроникой, генерирует световой луч, который, отражаясь от вращающегося зеркала, попадает на барабан и изменяет его статистический заряд в точке прикосновения. У некоторых принтеров разность потенциалов на поверхности барабана может составлять несколько сотен вольт. Лазер попадает на барабан только в тех точках, в которых должно быть изображение, таким образом на барабане получается скрытая копия отпечатка.

На следующем шаге на барабан наносится тонер (мельчайшая красящая пыль). Проходя мимо тонера, барабан притягивает его в точках, подвергшихся облучению лазером. Затем на поверхность барабана накладывается бумага, и тонер переходит на нее. Чтобы краска лучше держалась, перед самым барабаном ей тоже сообщается заряд, а в самом конце бумага проходит между двумя роликами с температурой 180°.

Альтернативной является светодиодный принтер, где вместо лазера и зеркала используется неподвижная строка с диодами LED (Light Emitting Diode), которая засвечивает барабан не по одной точке, а сразу по целой линии. На этом принципе работают принтеры фирмы OKI.

Другая категория лазерных принтеров — цветные лазерные принтеры. Они стоят значительно дороже и имеют большие размеры, т. к. состоят из двух — четырех обычных. Изображение формируется последовательно для каждого основного цвета, и лист печатается за четыре прохода (как и в струйных принтерах, используется СМΥК — представление цвета), что, естественно, сказывается на скорости печати. В принтере, соответственно, имеются четыре емкости с тонером каждого цвета. Качество цветной печати заметно лучше, чем у струйных принтеров, однако оно все — таки не фотографическое. Так что

для подделки документов и иных криминальных применений, а также просто для качественной печати фотографий лучше все же воспользоваться специальными принтерами для цветной печати, о которых будет рассказано ниже. Зато черно – белое изображение на высоте (однако не следует думать, что качественно печатать в цвете на лазерных принтерах невозможно – если на изображении отсутствуют ярко выраженные полутона, то качество просто отличное, да и фотографии тоже получаются довольно неплохо).

Для лазерного принтера очень важно, сколько памяти в нем установлено. Дело в том, что лазерный принтер сначала формирует изображение целиком, а только потом начинает его печатать. Понятно, что цветное изображение с разрешением порядка 600·600 dpi занимает не так уж мало, поэтому для обеспечения приемлемого быстродействия принтер должен иметь достаточное количество памяти. Цветные принтеры обычно имеют около 32 – 128 Мб памяти. Кроме оперативной памяти, в принтере обычно имеется (хотя далеко и не обязательно) и собственная Flash ROM, в которой хранятся различные шрифты, шаблоны, микропрограммы и тому подобное, что отменяет необходимость загрузки этих аксессуаров из РС, и, естественно, ускоряет работу. Иногда в принтере присутствует еще и маленький жесткий диск.

Важно и разрешение принтера. Оно, как и у струйных, может не быть симметричным. В таком случае большее значение определяет шаг лазерного луча, а меньшее — шаг барабана. Для лазерных принтеров стандартом является разрешение 600·600 dpi, однако бывает и меньше. У хороших принтеров (как правило, только черных, хотя иногда встречаются и цветные) оно может быть 1200·1200 или больше. Для передачи полутонов изображение разбивается на ячейки из нескольких десятков точек, и цветовая насыщенность обеспечивается числом пикселей в этой ячейке. Естественно, чем больше их в ячейке, тем выше качество.

В лазерных принтерах для повышения качества применяется технология повышения разрешения, называемая RET (Resolution Enhancement Technology). Она реализована с помощью специального чипа, который управляет интенсивностью лазера и таким образом позволяет варьировать энергией заряда барабана. При использовании этого метода заряд каждой точки определяется несколькими градациями. RET — технология позволяет существенно повысить качество изображения, особенно полутонового. Хотя изобретение было сделано

фирмой Hewlett – Packard, используется оно и другими производителями.

Еще одной особенностью описываемых устройств является еще то, что они работают через *стандартный программный интерфейс*, или *язык принтера*. Стандартный язык для лазерных принтеров разработала HP, назвав его PCL, или *Printer Control Language*. Он имеет несколько версий, совместимых сверху вниз. Все принтеры должны понимать PCL. Кстати, этот язык не только ответственен за управление принтером, в него также входят некоторые графические функции, описывающие, например, геометрические фигуры, поворот шрифтов и т. п. PCL иногда применяется и в струйных принтерах HP.

Другой, более «продвинутый» язык называется *PostScript*. Это стандартизированный язык описания страниц, который также предусматривает работу с цветом. PostScript разрабатывался фирмой Evansj Suther Land, создающей языки программирования для графической анимации, затем был дополнен Xerox, и, наконец, усовершенствован Adobe. Шрифты посылаются на принтер не в виде растрового изображения (типа bitmap), а используется их векторное представление. Так как принтер работает все же с растровым изображением, то шрифты должны быть снова преобразованы в растровую форму. Этим занимается RIP (Raster Image Processor) — специальный процессор, который, пользуясь интерпретатором языка, создает битовый образ символа на основе его размера, очертаний, расположения на странице и т. д. Таким же образом обрабатываются и геометрические изображения. В связи с этим PostScript обладает рядом существенных преимуществ перед PCL. PostScript экономит память принтера, т. к. записанные в векторной форме изображения занимают намного меньше места, предоставляет большие возможности в графическом плане — масштабирование, поворот, зеркальное отображение и прочее, причем все это с абсолютной точностью, т. к. данные представляются в математической форме. Файлы языка не зависят от аппаратного обеспечения — они могут восприниматься любым устройством (фотонаборным аппаратом или другим лазерным принтером), поддерживающим язык PostScript.

Наконец, третий язык. Он носит название HP – GL (Hewlett - Packard Gtaphics Language) и представляет собой практически индустриальную норму для чертежных принтеров или плоттеров. Он является расширением PCL. Туда входят инструкции, с помощью которых можно поворачивать изображение на любой угол или отображать зер-



кально. Геометрические фигуры генерируются с помощью HP – GL существенно быстрее, чем с помощью стандартного PCL. Этот язык поддерживают только широкоформатные принтеры или графопостроители, обычные лазерные принтеры такой поддержки не имеют.

### **Термические принтеры для цветной печати**

Как бы ни были хороши лазерные принтеры, они не позволяют все же достичь качества, при котором отпечаток будет схож с фотографией. Для этого предназначены специальные цветные принтеры, в которых используется термоперенос красителя.

Существует несколько технологий термической печати.

- Струйный перенос расплавленного красителя (термопластичная печать).
- Контактный перенос расплавленного красителя (термовосковая печать).
- Термоперенос красителя (сублимационная печать).

В последних двух случаях краситель нагревается и переносится на бумагу (или иной материал) в жидкой или газообразной форме. За несколько проходов формируется полноцветное изображение. Принтеры, в которых используется струйный перенос, называются еще восковыми твердокрасочными принтерами. При печати кусочки цветного воска плавятся и выбрызгиваются на носитель, при этом получаются очень насыщенные и реалистичные цвета, и изображение выглядит даже лучше некоторых фотографий. Отпечаток выходит слегка зернистым, поэтому такие принтеры не годятся для изготовления диапозитивов, так как капли воска после высыхания имеют сферическую форму. Существуют также устройства, совмещающие в себе как сублимационную, так и термовосковую технологии (они различаются тем, что при термовосковой печати, как правило, используется специальная бумага с восковым покрытием, а при сублимационной она не нужна).

Как и лазерные принтеры, термические принтеры тоже являются страничными. Скорость печати у них из-за инертности тепловых процессов намного ниже, чем у лазерных, и составляет для сублимационных 0.1 – 0.8 стр. в мин., а для термовосковых — от 0,5 до 2 страниц формата А4 в минуту.

Кроме того, стоимость подобных устройств достаточно велика, поэтому используются они в основном только там, где действительно есть острая необходимость в высококачественной печати цветных изображений.

## 11. ИНТЕРФЕЙСЫ КОМПЬЮТЕРА

### 11.1. Последовательный интерфейс

Разъемы COM – интерфейса, как и параллельного, появились на задней стенке РС чуть ли не с самого момента создания самого РС. На рис. 11.1 изображен наиболее часто используемый 9-контактный разъем Sub-D (папа).

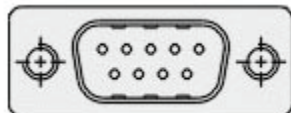


Рис. 11. 1. разъем Sub-D

COM — интерфейс очень часто называют последовательным. Это означает, что за один момент времени передается только один бит, а не несколько, как в случае параллельной передачи. В результате число линий, передающих данные, равно двум (одна на передачу, другая на прием), из – за чего последовательный кабель заметно тоньше, чем параллельный.

Уровень напряжения при передаче изменяется от  $-12$  до  $+12$  В, благодаря чему обеспечивается относительно большая помехоустойчивость, и длина кабеля может достигать 50 м и более. Как и в случае с параллельным портом, название «последовательный» не совсем верно — существуют и другие интерфейсы, в которых используется последовательная передача (USB, IEEE1394 и др.).

Стандартное обозначение последовательного порта RS – 232. Так называется протокол передачи данных, на основе которого действует COM – интерфейс. RS означает Recommended Standard.

Главный элемент последовательного интерфейса — микросхема UART (Universal Asynchron Recceiver Transmitter).

Данные при последовательной передаче разделяются служебными посылками, которые называются стартовый бит (Start bit) и стоп – бит (Stop bit). Они указывают соответственно на начало и конец передачи последовательности данных (Data bit). Данный метод позволяет осуществить синхронизацию между приемной и передающей сторонами, а также выровнять скорость обмена данными. Иногда

для идентификации ошибок используют еще бит контроля четности. Существует два варианта бита контроля четности: бит контроля четный (Even Parity) и, соответственно, нечетный (Odd Parity). Значение определяется двоичной суммой всех передаваемых битов данных. В большинстве случаев проверка четности не осуществляется. В панели управления Windows все эти значения можно изменять, но обычно это никому не нужно, т. к. в большинстве случаев коммуникационные (отсюда, кстати, и COM) порты сами настраиваются оптимальным образом.

Существует довольно большое количество устройств, подключаемых к последовательному порту. Однако быстродействие его невысокое, составляет 115 кбит/сек. Поэтому в последних спецификациях PC говорится, что вся периферия должна подключаться к USB, подключение же к другим портам хотя и возможно, но нежелательно. Вполне возможно, что уже очень скоро будут производиться системные платы без устаревших разъемов COM, PS/2 и LPT.

### **11.2. Интерфейс PS/2**

В настоящее время наиболее распространенный способ подключения клавиатуры и мыши – подключение к 6 – контактному порту PS/2. На материнских платах имеется два разъема, один из которых предназначен для мыши, другой — для клавиатуры. Сделать наоборот не получится.

Контроллер, обрабатывающий сигналы, поступающие с клавиатуры или мыши, очень чувствителен к перепадам напряжения. Поэтому не следует вставлять в порт PS/2 ничего до выключения компьютера — иначе контроллер может выйти из строя, и придется тогда менять материнскую плату.

Для того, чтобы не перепутать разъемы, предназначенные для мыши и клавиатуры, стандартом предусмотрена различная цветовая раскраска штекеров: фиолетовый для клавиатуры и зеленый для мыши. Хотя не всегда, надо заметить, производители придерживаются этого правила.

### **11.3. Интерфейс LPT**

Интерфейс LPT также часто называют параллельным (имеется ввиду параллельный порт). Из его названия следует, что обмен дан-

ными происходит параллельно. Это означает, что биты передаются не один за другим, как это делается в последовательных интерфейсах, а несколько бит передаются одновременно (параллельно), или, точнее, рядом, друг возле друга. То число бит, которое может быть передано за один такт, определяет разрядность интерфейса. Интерфейс LPT является 8 – разрядным. Существует также множество других параллельных интерфейсов (например, SCSI, PCI и др.).

Сегодня параллельный порт есть в каждом компьютере. Первоначально он предназначался исключительно для подключения принтера (LPT означает Line PrinTer), но впоследствии стали появляться и другие устройства: сканеры, мобильные дисководы, цифровые фотоаппараты, так что сейчас работа параллельного интерфейса не ограничивается только принтером, хотя в большинстве случаев это именно так и есть. LPT также часто называют Centronics в честь соответствующей фирмы, ставшей основным разработчиком параллельного порта. Соответственно, и кабель для подключения принтера к PC тоже называется Centronics.

Передача данных по кабелю может вестись только в одном направлении. Но некоторые устройства (современные принтеры, дисководы ZIP и т. д.) позволяют осуществлять и обратную связь. Для этого нужен другой кабель, называемый Bitronics.

BIOS компьютера поддерживает до трех параллельных портов (которые на практике редко кому требуются).

#### **11.4. Шины расширения**

Шины расширения (expansion bus) ввода-вывода являются средствами подключения системного уровня: они позволяют адаптерам и контроллерам периферийных устройств непосредственно использовать системные ресурсы компьютера — пространство адресов памяти и ввода-вывода, прерывания, прямой доступ к памяти. Устройства, подключенные к шинам расширения, могут и сами управлять этими шинами, получая доступ к остальным ресурсам компьютера. Шины расширения механически реализуются в виде слотов (щелевых разъемов) или штырьковых разъемов. Для них характерна малая длина проводников, то есть они сугубо локальны, что позволяет достигать высоких скоростей работы. Эти шины могут и не выводиться на разъемы, а использоваться для подключения устройств в интегрированных системных платах. В истории шин расширения ПК насчитывается уже 3 поколения.

К первому поколению относится ISA — асинхронная параллельная шина с низкой пропускной способностью, не имеющая средств обеспечения надежности обмена и автоконфигурирования.

Второе поколение началось с шины EISA (а также MCA), за которой последовали шина PCI и ее расширение PCI-X. Это поколение параллельных синхронных надежных шин со средствами автоконфигурирования. Скорость передачи достигает единиц Гб/сек. Для подключения большого числа устройств применяется иерархическое объединение шин с помощью мостов в древовидную структуру.

Для третьего поколения (шина PCI Express, она же 3GIO, Hyper Transport, Advanced Switching и InfiniBand) характерен переход от шин к двухточечным соединениям с последовательным интерфейсом.

В современных компьютерах основной шиной расширения пока является шина PCI и ее расширение PCI-X; ее дополняет порт AGP. Сейчас осуществляется переход на PCI Express — это средство подключения графического адаптера вытесняет AGP. В блокнотных компьютерах широко применяются слоты PCMCIA с шинами PC Card и Card Bus, появилась и Express Card. Сравнительные характеристики шин приведены в таблице 11.1.

Таблица 11.1.

### Сравнительные характеристики шин расширения

Шина	Пиковая пропускная способность, Мбайт/с	Каналы DMA	Bus-Master	ACFG <sup>1</sup>	Разрядность данных	Разрядность адреса	Частота, МГц
ISA-8	4	3	-	-	8	20	8
ISA-16	8	7	+	-	16	24	8
LPC	6,7	7	+	-	8/16/32	32	33
EISA	33,3	7	+	+	32	32	8,33
MCA-16	16	-	+	+	16	24	10
MCA-32	20	-	+	+	32	32	10
VLB	132				32/64	32	33-50(66)
PCI	133-533		+	+	32/64	32/64	33/66
PCI-X	533-4256		+	+	16/32/64	32/64	66-133
PCI Express	496-15872		+	+	1/2/4/8/12/16/32	32/64	2,5 ГГц
AGP 1x/2x/4x/8x	266/533/1066/2132		+	+	32	32/64	66
PCMCIA	10/20	+	-	+	8/16	26	10
Card Bus	132		+	+	32	32	33

## 11.5. Шины PCI и PCI-X

PCI и PCI-X — синхронные параллельные шины расширения ввода-вывода, обеспечивающие надежный высокопроизводительный обмен и автоматическое конфигурирование устройств. Шины PCI и PCI-X являются ближайшими «родственниками» с полной взаимной совместимостью устройств. Большинство положений, относящихся к PCI, относится и к PCI-X, так что в дальнейшем описании термин «PCI» в основном относится к обоим вариантам (различия подчеркиваются особо).

Шина PCI позволяет объединять одноранговые устройства. Любое устройство шины может выступать как в роли *инициатора транзакций* (задатчика), так и в роли *целевого устройства*. Целевое устройство отвечает на транзакции, адресованные к его ресурсам (областям памяти и портам ввода-вывода). Ядро компьютера (центральный процессор и память) для шины PCI также представляется устройством — *главным мостом* (host bridge). В транзакциях к устройствам PCI, инициированных центральным процессором, главный мост является задатчиком. В транзакциях от устройств PCI, обращающихся к ядру (к системной памяти), главный мост является целевым устройством. Право на управление шиной в любой момент времени дается лишь одному устройству данной шины, арбитраж запросов на управление шиной осуществляется централизованным способом.

Наличие активных устройств (помимо ЦП) позволяет в компьютере выполнять параллельно несколько операций обмена, одновременно с обращениями процессора могут выполняться транзакции от мастеров шины PCI. В системе с несколькими шинами PCI возможно параллельное функционирование устройств-мастеров на разных шинах PCI Peer Consistency. Однако если они обращаются к одному ресурсу (системной памяти), то какие-то фазы этих обменов все-таки приходится выполнять последовательно.

Каждая *физическая шина PCI* позволяет объединять лишь небольшое число устройств (обычно не более шести). Для увеличения числа подключаемых устройств применяют *мосты PCI* (PCI-to-PCI Bridge) — устройства PCI с парой интерфейсов, которыми шины объединяются в древовидную структуру. В корне этой структуры находится *хост* — «хозяин шины», в обязанности которого входит конфигурирование всех устройств, включая мосты. В роли хоста, как правило, вы-

ступают центральный процессор с главным мостом. Мосты позволяют объединять шины PCI и PCI-X с разными характеристиками, а также подключать к PCI/PCI-X иные шины: шины блокнотных ПК, PCI Express, Hyper Transport и др.

Важной частью шины PCI является *система автоматического конфигурирования*; конфигурирование выполняется каждый раз при включении питания и инициализации системы. Специальное конфигурационное ПО позволяет обнаружить и идентифицировать все установленные устройства, а также выяснить их потребности в ресурсах (областях памяти, адресах ввода-вывода, прерываниях). Одно и то же функциональное устройство может быть сконфигурировано по-разному, отображая свои операционные регистры либо на пространство памяти, либо на пространство адресов ввода-вывода.

### **Взаимодействие устройств**

С устройством PCI, когда оно является целевым, можно взаимодействовать несколькими способами:

- командами *обращения к памяти и портам ввода-вывода*, которые адресуются к областям, выделенным устройству при конфигурировании;

- командами *обращения к конфигурационным регистрам*, которые адресуются по *идентификатору* — номеру шины, устройства и функции (компонентам многофункционального устройства PCI).

Для обращений к пространству памяти используется 32- или 64-битная адресация, причем разрядность адресации не зависит от разрядности шины. Таким образом, шина позволяет адресовать до  $2^{32}$  (4 Гбайт) или  $2^{64}$  ячеек памяти. На шине PCI фигурирует физический адрес памяти.

*Конфигурационные регистры* устройств PCI расположены в обособленном пространстве адресов (отдельном от пространства адресов памяти и ввода-вывода). В конфигурационных регистрах указываются адреса, отведенные устройству (как целевому), — через них разрешается работа в роли инициатора и целевого устройства, кроме того, через них конфигурируются **прерывания**. Конфигурационные регистры обеспечивают возможность автоматической настройки всех устройств шины PCI. К этим регистрам система обращается на этапе конфигурирования — переучета обнаруженных устройств, выделения им неперекрывающихся ресурсов (областей памяти и пространства ввода-вывода) и назначения номеров аппаратных прерываний. При дальнейшей регулярной работе

взаимодействие прикладного ПО с устройствами осуществляется преимущественно путем обращений по назначенным в процессе конфигурирования адресам памяти и ввода-вывода. Конфигурационные же регистры в регулярной работе используются для системных целей: настройки параметров, описывающих поведение устройства на шине, обработки ошибок, идентификации источника прерываний.

Обращения к регистрам и памяти устройств PCI выполняются командами шины PCI. Команды может подавать любой инициатор, как хост (главный мост) по командам центрального процессора, так и рядовое устройство PCI. Возможность распространения ряда команд зависит от взаимного расположения инициатора и целевого устройства на ветвях дерева шин PCI. Однако хост может безусловно подать любую команду любому устройству PCI. Только хост всегда имеет доступ к конфигурационным регистрам всех устройств (и мостов), поэтому он и должен заниматься конфигурированием. После конфигурирования любое устройство PCI может безусловно обратиться к системной памяти, то есть реализовать *прямой доступ к памяти* (DMA).

Устройства PCI могут вырабатывать *запросы аппаратных прерываний*:

- обычные маскируемые — для сигнализации событий в устройстве; эти прерывания могут сигнализироваться как традиционным способом, по специальным сигнальным линиям, так и передачей сообщений (MSI);

- немаскируемые — для сигнализации о серьезных ошибках;

- прерывания системного управления (System Management Interrupt, SMI) — для сигнализации о событиях в системе управления энергопотреблением

Наиболее эффективно возможности шины PCI используются при применении *активных устройств — мастеров шины (PCI Bus Master)*. Только эти устройства могут обеспечить скорость передачи данных, приближающуюся к декларируемой пиковой пропускной способности. Максимальная производительность обменов по шине PCI достигается только в пакетных транзакциях значительной длины. Транзакции по инициативе программы, исполняемой на ЦП, проводимые главным мостом, как правило, являются одиночными (или очень короткими пакетными). По этой причине программно-управляемый обмен данными с устройствами PCI по производительности значительно уступает обмену, выполняемому устройством-мастером. Таким образом, применение активных устройств дает двойной эффект: разгружает централь-



ный процессор и обеспечивает лучшее использование пропускной способности шины.

### **Шины, устройства, функции и хост**

Каждое устройство PCI при установке в конкретную систему получает *идентификатор*, однозначно определяющий его положение на дереве шин PCI данного компьютера. Идентификатор имеет иерархическую структуру и состоит из номеров *шины* (bus), *устройства* (device) и *функции* (function). Идентификатор задает положение блока конфигурационных регистров заданной функции выбранного устройства в общем конфигурационном пространстве системы. Идентификаторы фигурируют при обращениях к регистрам конфигурационного пространства, а также при обмене сообщениями между устройствами.

*Шина PCI* представляет собой набор сигнальных линий, непосредственно соединяющих интерфейсные выводы группы устройств (слотов, микросхем на плате). В системе может присутствовать несколько шин PCI, соединенных *мостами PCI*. Мосты электрически отделяют интерфейсные сигналы одной шины от другой, соединяя шины логически. Главный мост соединяет главную шину PCI с хостом (процессором и памятью). Каждая шина имеет свой *номер шины* (PCI bus number). Шины нумеруются последовательно, начиная от хоста, шина PCI, подключенная к главному мосту, имеет нулевой номер.

*Устройством PCI* называется микросхема, или карта расширения, подключенная к одной из шин PCI. Устройство может быть многофункциональным, т. е. состоять из 1–8 так называемых *функций*. Каждой функции отводится конфигурационное пространство в 256 байт, в PCI-X оно расширено до 4096 байт. Многофункциональные устройства должны отзываться только на конфигурационные циклы с номерами функций, для которых имеется конфигурационное пространство.

### **Пропускная способность шин PCI и PCI-X**

Декларируемая высокая пропускная способность шины достигается только в длинных пакетных циклах, однако пакетные циклы выполняются далеко не всегда.

Стремиться к пакетной передаче транзакций (операций обмена) записи стоит только в том случае, если устройство PCI поддерживает пакетные передачи в ведомом (target) режиме. Если это не так, попытка пакетной передачи приведет даже к небольшой потере производительности.

При одиночных транзакциях на стандартной шине PCI достижима максимальная скорость чтения 33 Мбайт/с, скорость записи может достигать 66 Мбайт/с. Скорость, соизмеримую с максимальной пиковой, можно получить только при пакетных передачах. При длине пакета 16 байт (4 фазы данных) достижима скорость чтения 76 Мбайт/с и скорость записи 106,6 Мбайт/с. При шестнадцати фазах данных скорость чтения может достигать 112 Мбайт/с, а записи — 125 Мбайт/с.

Итак, для выхода на максимальную производительность обмена устройства PCI сами должны быть ведущими устройствами шины, причем способными генерировать пакетные циклы. Радикально повысить пропускную способность позволяет переход на частоту 66 МГц и разрядность 64 бита, что обходится недешево. Для того, чтобы на шине могли нормально работать устройства, критичные ко времени доставки данных (сетевые адаптеры, устройства, участвующие и записи и воспроизведении аудио- и видеоданных, и др.), не следует пытаться «выжать» из шины ее декларируемую полосу пропускания полностью. Перегрузка шины может привести, например, к потере пакетов из-за несвоевременности доставки данных.

### **Транзакции шин PCI и PCI-X**

Обмен информацией по шинам PCI и PCI-X организован в виде *транзакций* — логически завершенных операций обмена. В каждой транзакции выполняется одна *команда* — как правило, чтение или запись данных по указанному адресу. Транзакция начинается с *фазы адреса*, в которой инициатор задаст команду и целевой адрес. Далее могут следовать *фазы данных*, в которых одно устройство (источник данных) помещает данные на шину, а другое (приемник) их считывает. Транзакции, в которых присутствует множество фаз данных, называются *пакетными*. Есть и одиночные транзакции (с одной фазой данных). Транзакция может завершиться и без фаз данных, если целевое устройство (или инициатор) не готово к обмену. В шине PCI-X добавлена *фаза атрибутов*, в которой передается дополнительная информация о транзакции.

## **11.6. Порт графического акселератора — AGP**

Порт AGP (Accelerated Graphic Port — порт ускоренной графики) был введен для подключения *графических адаптеров с 3D-акселераторами*. Такой адаптер содержит: *акселератор* — специали-

зированный графический процессор; *локальную память*, используемую и как видеопамять, и как локальное ОЗУ графического процессора; *управляющие и конфигурационные регистры*, доступные как локальному, так и центральному процессорам. Акселератор может обращаться и к локальной памяти, и к системному ОЗУ, в котором для него могут храниться наборы данных, не уместяющиеся в локальной памяти (как правило, текстуры большого объема). Основная идея порта AGP заключается в предоставлении акселератору максимально быстрого доступа к системной памяти (локальная ему и так близка), более приоритетного, чем доступ к ОЗУ со стороны других устройств.

Порт AGP представляет собой 32-разрядный параллельный синхронный интерфейс с тактовой частотой 66 МГц; большая часть сигналов позаимствована с шины PCI. Однако, в отличие от PCI, интерфейс порта AGP двухточечный, соединяющий графический акселератор с памятью и системной шиной процессора каналами данных чипсета системной платы, не пересекаясь с «узким местом» — шиной PCI. Обмен через порт может происходить как по протоколу PCI, так и по протоколу AGP. Отличительные особенности порта AGP:

- конвейеризация обращений к памяти;
- умноженная относительно тактовой частоты порта частота передачи данных (2х/4х/8х);
- «внеполосная» подача команд, обеспеченная демультиплексированием шин адреса и данных.

Порт AGP позволяет акселератору работать в двух режимах — DMA и DIME (Direct Memory Execute). В *режиме DMA* акселератор при вычислениях рассматривает локальную память как первичную, а когда ее недостаточно, подкачивает в нее данные из основной памяти. В режиме DIME, он же режим исполнения (executive mode), локальная и основная память для акселератора логически равнозначны и располагаются в едином адресном пространстве.

## 11.7. PCI Express

### Зачем нужно последовательное соединение

Новым направлением в развитии высокоскоростных периферийных шин является использование последовательной архитектуры. Для передачи информации в параллельной архитектуре, где биты передаются одновременно, необходимы линии, имеющие 8, 16 и более проводов. Можно предположить, что за одно и то же время через па-

параллельный канал передается больше данных, чем через последовательный, однако на самом деле увеличить пропускную способность последовательного соединения намного легче, чем параллельного.

Параллельное соединение обладает рядом недостатков, одним из которых является фазовый сдвиг сигнала, из-за чего длина параллельных каналов, например SCSI, ограничена (не должна превышать 3 м). Проблема в том, что, хотя 8- и 16-разрядные данные одновременно пересылаются передатчиком, из-за задержек одни биты прибывают в приемник раньше других. Следовательно, чем длиннее кабель, тем больше время задержки между первым и последним прибывшими битами на приемном конце. Последовательная шина позволяет одновременно передавать 1 бит данных. Благодаря отсутствию задержек при передаче данных значительно увеличивается тактовая частота.

Еще одно преимущество последовательного способа передачи данных — возможность использования только одно- или двухпроводного канала, поэтому помехи, возникающие при передаче, очень малы, чего нельзя сказать о параллельном соединении. Стоимость параллельных кабелей довольно высока, поскольку провода, предназначенные для параллельной передачи, не только используются в большом количестве, но и специальным образом укладываются, чтобы предотвратить возникновение помех, а это весьма трудоемкий и дорогостоящий процесс. Кабели для последовательной передачи данных, напротив, очень дешевые, так как состоят из нескольких проводов и требования к их экранированию намного ниже, чем у используемых для параллельных соединений.

### **Организация шины PCI Express**

Разработкой шины ввода/вывода третьего поколения, называемой PCI Express, а также известной как Arapahoe и 3GIO (3rd generation I/O), занимался консорциум Arapahoe Working Group, в который наряду с другими входят компании Intel, Compaq, Dell, HP, IBM и Microsoft. В начале 2002 г. разработка была закончена, а версия 1.0 передана на рассмотрение отраслевой группе PCI-SIG (PCI Special Interest Group), занимающейся утверждением спецификаций шины PCI и всего, что с ней связано. Позднее шина PCI Express была утверждена в качестве открытого стандарта. Что же представляет собой PCI Express.

В топологию шины добавлен новый элемент — коммутатор (switch), который может находиться на материнской плате как отдель-

ное устройство, а может быть интегрирован в чипсет. Коммутатор призван заменить традиционную шину с множеством подключений коммутируемой технологией и обеспечить одноранговую связь с устройствами. Это помимо всего прочего позволит избавиться от некоторого бесполезного трафика, поступающего на мост ввода/вывода, ведь данные не будут отправляться чипсету, если они затрагивают только конечные устройства и не нуждаются в дополнительной обработке или доступе к памяти. PCI Express – шина не параллельная, как ее предшественница, а последовательная, что в нынешних условиях предполагает более высокую производительность. Последовательная шина использует прямые соединения между устройствами с малым количеством служебного трафика и низкими задержками, а также обеспечивает гибкую масштабируемость производительности при помощи изменения тактовой частоты и добавления линий (lane). Каждая линия может передавать 250 Мб/с в каждую сторону, что почти вдвое больше пропускной способности обычной способности обычной PCI. Всего же линий может быть до 32, т. е. суммарная пропускная способность PCI Express может составлять 16 Гб/с, и это без учета увеличения тактовой частоты. Понятно, что не вся периферия требует такой огромной пропускной способности, поэтому на материнских платах будут присутствовать разные PCI Express слоты – X1 и X2 для более медленных устройств (GbE, RAID и т.п.) и X16 для видеокарт. Учитывая темпы развития трехмерной графики, через некоторое время можно также ожидать появление X32-слотов. Архитектура PCI Express состоит из нескольких уровней. Самый нижний, физический уровень (Physical Layer), состоящий из проводников, передает пакеты между канальными уровнями двух подключенных устройств. Физический уровень поддерживает ширину шины X1, X2, X4, X8, X16 и X32 линий, а сами линии могут состоять из металлических или оптических проводников. Канальный уровень (Data Link Layer) отвечает за целостность пакетов, добавляя уникальный идентификатор и контрольную сумму к каждому пакету. В соответствии с применяемым протоколом управления потоком разрешение на передачу пакета дается только в случае готовности буфера на принимающей стороне, что позволяет избежать повторных передач и использовать канал более эффективно. Уровень транзакций (Transaction Layer), получающий запросы на чтение/запись от программного уровня, создает пакеты для передачи канальному уровню. Формат пакетов поддерживает 32-битную и расширенную 64-битную адресацию памяти. Каждый пакет также

может иметь атрибуты, такие как no-snoop, relaxed-ordering и priority, позволяющие оптимизировать маршрутизацию.

Программный уровень (Soft-ware Layer), отвечающий за программную совместимость, включает в себя два подуровня – загрузку и выполнение, которые, по сравнению с PCI, остались неизменными. При загрузке операционная система находит подключенные устройства и распределяет ресурсы между ними. Выполнение же производится по модели PCI «загрузка-сохранение», что также поддерживается архитектурой PCI Express. Таким образом, существующие операционные системы, драйверы и приложения, способные работать с PCI, будут работать и с PCI Express, без каких-либо изменений, а новые версии ПО смогут использовать новые функции шины.

К преимуществам последовательного соединения можно добавить возможность управления питанием, горячего подключения и горячей замены устройств – для этого будут использоваться специальные горячие отсеки. Одним из условий при разработке PCI Express было соответствие ее стоимости уровню PCI. Последовательное соединение предусматривает меньшее число проводников и позволяет более эффективно использовать освободившееся на печатной плате место для размещения других компонентов. Это в некоторой степени сократит расходы производителей на разработку и дизайн самой платы, которые, естественно, включены в стоимость конечного продукта. В отличие от PCI, которая разрабатывалась только для домашних ПК, PCI Express позиционируется как универсальная шина для разных сегментов рынка, включающих помимо десктопов серверы, мобильные платформы, устройства связи и межкомпонентные соединения. С большой долей вероятности можно предположить, что слоты PCI и PCI Express будут мирно сосуществовать на протяжении как минимум нескольких лет, однако постепенное вытеснение PCI теперь неизбежно. PCI Express заменила собой AGP-порт.

## **11.8. Шина USB**

Сколько устройств можно подключить к двум COM-портам и одному LPT? Правильно, три. До появления разъема PS/2 один COM-порт занимала мышь. Остается два порта. Этого мало. Да и с подключением могут быть проблемы – прерывания, драйверы и все прочее.

Скорости COM и LPT – портов постепенно стало не хватать для нужд потребителей. Поэтому, проявляя заботу о пользователях и

удовлетворяя всеобщую потребность в унификации, компании Intel, IBM, NEC, Nozther Telecom и Compaq объединились в консорциум и попытались решить проблему. Нужен был новый интерфейс подключения к компьютеру периферийных устройств, который, помимо всего прочего, отвечал бы трем основным требованиям: простота подключения, высокая скорость обмена данными, возможность подсоединения множества устройств. Результатом проведенной работы (1995–1996 г.) стало создание шины USB – Universal Serial Bus, что переводится как универсальная последовательная шина. Несмотря на многочисленные достоинства нового интерфейса и его преимущества перед старыми портами (а их много, например, высокая скорость обмена данными – 12 Мбит/с для USB 1.0 и 480 Мбит/с для USB 2.0, поддержка до 127 устройств, возможность «горячего» подключения, электропитание устройств непосредственно через шину USB и пр.), сразу после выхода USB в свет не отмечалось ее особой популярности и всплеска производства USB – устройств. Причина проста – отсутствие поддержки в операционной системе. Ну а уж после появления полноценной программной поддержки в популярной ОС, производители периферии не стали тормозить – на рынок обрушился поток USB – устройств. Как абсолютно новых моделей, так и более старых, переделанных под стандарт USB. А еще компаниям-производителям оборудования понравилось то, что через USB можно питать устройство.

### **Работа USB**

Помимо общего вопроса работы шины, интересны и некоторые частные моменты. Например, как реализовано «горячее подключение», почему можно подключить целых 127 устройств, почему не нужно возиться с установкой драйверов и т. д. Нужно разобраться.

Такая большая цепочка устройств – это уже самая настоящая сеть. Сеть с топологией «многоуровневая звезда». В системную плату компьютера встроены два устройства с верхней ступени иерархии, которая носит название roottier (главный уровень, верхний слой): хост-контроллер (может быть только один) и корневой разветвитель (root hub). Хост-контроллер (или USB – контроллер) отвечает за следующие действия: обнаружение, присоединение и удаление устройств (по терминологии USB – спецификации), управление командным потоком между хостом и устройствами USB, управление потоком данных между хостом и устройствами USB, сбор статистики о состоянии и активности устройств, обеспечение подачи ограниченного количества

мощности на присоединение устройства USB. Корневой хаб дает нам возможность подключать к шине более одного устройства. Root hub соединен непосредственно с хост-контроллером.

Все устройства находящиеся в цепи, делятся на два типа: хабы, которые позволяют подключить к одному порту несколько конечных устройств и являются расширителями сети, и конечные устройства (функции), Хаб – это устройство, которое занимает один порт, но при этом имеет несколько портов USB для подключения других устройств. Функция – это фотоаппарат, принтер, внешний CD-ROM и т. д. Хаб также может быть встроен и в устройство, чаще всего в монитор.

Одна из самых интересных возможностей USB -это «горячее» подключение. Как оно происходит? Все просто. При подключении нового устройства происходит изменение напряжения в цепи (100мА), которое улавливается хостом. Тот сразу начинает опрашивать новое устройство о том, кто оно, что и зачем, то есть выясняет его класс, нужную ему полосу пропускания и прочую служебную информацию. После получения ответов новичку присваивается уникальный номер, а в ОС загружается нужный драйвер. Никаких требований ввести IRQ, DMA или что-то еще нет, потому что программы, работающие с устройством, общаются не с ним, а с программным интерфейсом, через который USB взаимодействует с окружающей средой. Всю работу выполняет он и хост-контроллер. По USB – кабелям не только передаются данные, но и осуществляется питание для устройств. Как и все в мире, оно конечное, и хватает его не всем. Мышке, клавиатуре или еще какому-нибудь небольшому устройству вполне достаточно. А вот джойстику или рулю с обратной связью обязательно понадобится отдельный электрический кабель.

## **11.9 Шина IEEE – 1394 – FireWire**

Высокопроизводительная последовательная шина (high performance serial bus) IEEE 1394 – FireWire создавалась как более дешевая и удобная альтернатива параллельным шинам (SCSI) для соединения равноранговых устройств. Шина без дополнительной аппаратуры (хабов) обеспечивает связь до 63 устройств. Устройства бытовой электроники — цифровые камкордеры (записывающие видеокамеры), камеры для видеоконференций, фотокамеры, приемники кабельного и спутникового телевидения, цифровые видеоплееры (CD и DVD), акустические



системы, цифровые музыкальные инструменты, а также периферийные устройства компьютеров (принтеры, сканеры, устройства дисковой памяти) — и сами компьютеры могут объединяться в единую сеть. Шина не требует управления со стороны компьютера. Она поддерживает *динамическое реконфигурирование* — возможность «горячего» подключения и отключения устройств. События подключения-отключения вызывают сброс и повторную инициализацию: определение структуры шины (дерева), назначение физических адресов всем узлам и, если требуется, избрание ведущего устройства (мастера) циклов, диспетчера изохронных ресурсов и контроллера шины. Менее чем через секунду после сброса все ресурсы становятся доступными для последующего использования, и каждое устройство имеет полное представление обо всех подключенных устройствах и их возможностях. Благодаря наличию линий питания интерфейсная часть устройства может оставаться подключенной к шине даже при отключении питания функциональной части устройства.

По инициативе VESA шина позиционируется как основа «домашней сети», объединяющей всю бытовую и компьютерную технику в комплекс. Эта сеть является одноранговой (peer-to-peer), чем существенно отличается от USB.

Стандарт IEEE – 1394 имеет ряд совместимых реализаций под разными названиями: FireWire, iLink, Digital Link, MultiMedia Connection. Наиболее часто используется разработанная фирмой Apple шина FireWire, на основе которой и появился стандарт.

### **11.10. Интерфейс IDE – ATA/ATAPI и SATA**

Интерфейс IDE предназначен для подключения устройств хранения данных, обладающих собственным контроллером. В настоящее время интерфейс ATA/ ATAPI является самым массовым интерфейсом устройств хранения данных, причем не только в мире PC-совместимых компьютеров. Пока что наибольшее распространение получил его «классический» параллельный вариант, ему на смену идут последовательные интерфейсы Serial ATA (SATA) и Serial ATA-II (SATA-II). Теперь параллельный интерфейс ATA/ATAPI стали называть PATA (Parallel ATA — параллельный интерфейс ATA).

*Параллельный интерфейс ATA* (Advanced Technology Attachment) был введен в конце 1980-х г. как интерфейс для подключения накопителей на жестких магнитных дисках к компьютерам IBM PC AT с шиной

ISA. Интерфейс появился в результате переноса стандартного (для PC/AT) контроллера накопителя на жестком диске (Hard Disc Controller, HDC) ближе к накопителю, то есть создания *устройств со встроенным контроллером* (Integrated Drive Electronics, IDE). Для связи устройства с системной шиной использовали ленточный кабель с *параллельным шинным интерфейсом*, получившим названия *ATA* и *IDE*, которые фактически, являются синонимами.

Поскольку стандартный контроллер АТ позволял подключать до двух накопителей, эту возможность получил и интерфейс АТА. Однако теперь два накопителя стали означать и два контроллера, подключенных к одной интерфейсной шине.

Интерфейс АТА предназначен для обмена с устройствами хранения *блоками* фиксированного размера — *секторами* по 512 байт. Адресация данных внутри устройств АТА имеет «дисковые корни»: для накопителей изначально указывали адреса цилиндра (cylinder), головки (head) и сектора (sector) — так называемая *трехмерная адресация CHS*. Позже пришли к *линейной* адресации логических блоков (Logical Block Address, LBA), где адрес блока (сектора) определяется 28-битным числом. Трехмерная и 28-битная линейная адресация в АТА имеют предел емкости устройств в 136,9 и 137,4 Гбайт соответственно, что по нынешним меркам недостаточно. В современных версиях интерфейса линейную адресацию расширили до 48-битной, при этом предел адресации составляет  $2^{48}$  - 281 474 976 710 656 секторов, или около 144 Пбайт (петабайт), то есть 144 115 188 075 855 360 байт. Устройство может поддерживать различные форматы адресации, причем формат адреса может меняться даже в соседних командах.

Принятая система команд и регистров, являющаяся частью спецификации АТА, ориентирована на блочный обмен данными с жесткими магнитными дисками — устройствами хранения с непосредственным доступом. Позже спецификацию расширили для иных устройств хранения.

– Спецификация АТАPI позволяет передавать *пакет*, содержащий *командный блок* (откуда и часть названия PI — Package Interface). Структура командного блока заимствована из SCSI, его содержимое определяется типом подключенного устройства: ленточного, оптического (CD, DVD), магнитооптического и т. п. АТАPI позволяет расширить границы применения шины АТА, введя всего лишь одну новую команду передачи управляющего пакета.

– Набор дополнительных команд CFA (Compact Flash Association) введен для устройств хранения на флэш-памяти. От обычных устройств хранения (с непосредственным доступом) флэш-память отличается сравнительно длительным стиранием данных перед перезаписью. Группа дополнительных команд позволяет более эффективно работать с этими устройствами (хотя возможен доступ к ним и обычными, «дисковыми» командами ATA).

Параллельный интерфейс ATA исчерпал свои ресурсы пропускной способности, достигшей 133 Мбайт/с в режиме UltraDMA Mode 6. Для дальнейшего повышения пропускной способности интерфейса (но, увы, не самих устройств хранения, которые имеют гораздо меньшие внутренние скорости обмена с носителем) было принято решение о переходе от параллельной шины к последовательному двухточечному интерфейсу Serial ATA (SATA). Цель перехода – улучшение и удешевление кабелей и коннекторов, улучшение условий охлаждения устройств внутри системного блока (избавление от широкого шлейфа), обеспечение возможности разработки компактных устройств, облегчение конфигурирования устройств пользователем.

Интерфейс SATA позволяет сохранять (и развивать) сложившуюся систему команд ATA/ATAPI, что обеспечивает преемственность и программную совместимость со старым ПО. Поначалу интерфейс SATA отличался только способом транспортировки данных и команд между контроллером и устройствами. Главная революция в организации обмена с устройствами хранения произведена в спецификации SATA II, в которой описан эффективный механизм обслуживания очередей — NCQ на базе механизма FPDMA. Для SATA II появилась новая спецификация контроллера — AHCI, которая меняет и идеологию взаимодействия (сохраняя команды).

Физически интерфейс IDE реализован при помощи плоского 40-жильного кабеля, на котором могут быть разъемы для подключения одного или двух устройств. Общая длина кабеля не должна превышать 45 см, причем между разъемами должно быть расстояние не менее 15 см.

Но вряд ли можно сказать, что пропускная способность интерфейса Ultra ATA/100 не удовлетворяет нуждам имеющихся винчестеров. Безусловно, скорости винчестеров постоянно увеличиваются, но пропускная способность интерфейса никогда не становилась узким местом. Даже самые быстрые винчестеры IDE не могут перешагнуть

за черту 45 Мбайт/с (скорость чтения данных с пластин), и поэтому проблемным местом становится не интерфейс, а собственно винчестер.

### 11.11. Шина SCSI

SCSI (Small Computer System Interface) – интерфейс системного уровня, стандартизованный ANSI, в отличие от интерфейсных портов (COM, LPT, IR, MIDI), представляет собой шину: сигнальные выводы множества устройств-абонентов соединяются друг с другом «один в один».

Основным предназначением SCSI-шины во время разработки первой спецификации в 1985 г. было «обеспечение аппаратной независимости подключаемых к компьютеру устройств определенного класса».

В отличие от жестких шин расширения SCSI-шина реализуется в виде отдельного кабельного шлейфа, который допускает соединение нескольких устройств (спецификация SCSI-1) внутреннего и внешнего исполнения. Одно из них – *хост адаптер* (Host Adapter) связывает шину SCSI с системной шиной компьютера, другие свободны для периферии.

К шине могут подключаться:

- дисковые внутренние и внешние накопители (CD-ROM, винчестеры, сменные винчестеры, магнитооптические диски и др.);
- стримеры;
- сканеры;
- фото- и видеокамеры;
- другое оборудование, применяемое не только для IBM PC.

Каждое устройство, подключенное к шине, имеет свой *идентификатор* SCSI ID, который передается позиционным кодом по 8-битной шине данных (отсюда и ограничение на количество устройств на шине). Устройство (ID) может иметь до 8 подустройств со своими LUN (Logical Unit Number – логический номер устройства).

Любое устройство может инициировать обмен с другим *целевым устройством* (Target).

Режим обмена по SCSI-шине может быть асинхронным, или синхронным с согласованием скорости (Synchronous Negotiation), где передача данных контролируется по паритету.

*Спецификация SCSI-3* - дальнейшее развитие стандарта, направленное на увеличение количества подключаемых устройств, спецификацию дополнительных команд, поддержку Plug and Play.

SCSI-устройства соединяются кабелями в *цепочку* (Daisy Chain), на крайних устройствах подключаются *терминаторы*. Часто одним из крайних устройств является хост-адаптер. Он может иметь для каждого канала как внутренний разъем, так и внешний.

### 11.12. Инфракрасный интерфейс IrDA

Применение излучателей и приемников инфракрасного (ИК) диапазона позволяет осуществлять беспроводную связь между парой устройств, удаленных на расстояние нескольких метров. Инфракрасная связь — *IR (InfraRed) Connection* — безопасна для здоровья, не создает помех в радиочастотном диапазоне и обеспечивает конфиденциальность передачи. ИК-лучи не проходят через стены, поэтому зона приема ограничивается небольшим, легко контролируемым пространством. Инфракрасная технология привлекательна для связи портативных компьютеров с периферийными устройствами. Инфракрасный интерфейс имеют некоторые модели принтеров, им оснащают многие современные малогабаритные устройства: карманные компьютеры (PDA), мобильные телефоны, цифровые фотокамеры и т. п.

Различают инфракрасные системы низкой (до 115,2 Кбит/с), средней (1,152 Мбит/с) и высокой (4 Мбит/с) скорости. Низкоскоростные системы служат для обмена короткими сообщениями, высокоскоростные — для обмена файлами между компьютерами, подключения к компьютерной сети, вывода на принтер, проекционный аппарат и т. п. Ожидаются более высокие скорости обмена, которые позволят передавать «живое видео».

Излучателем для ИК-связи является светодиод с длиной волны 880 нм; светодиод дает конус излучения с углом около 30°. В качестве приемника используют PIN-диоды, эффективно принимающие ИК-лучи в конусе 15°. Помимо полезного сигнала на приемник воздействуют помехи, в том числе засветка от солнечного освещения или ламп накаливания, дающая постоянную составляющую оптической мощности, и засветка от люминесцентных ламп, дающая переменную (но низкочастотную) составляющую. Эти помехи приходится фильтровать. Для передачи сигналов используют двоичную модуляцию (есть свет — нет света) и различные схемы кодирования.

### 11.13. Радиointерфейс BLUETOOTH

В 1998 г. ряд компаний объединились для разработки и продвижения технологии беспроводной передачи данных, которая получила название Bluetooth. Новый интерфейс должен был позволить соединять друг с другом практически любые устройства — ноутбуки, принтеры, цифровые фотоаппараты, мобильные телефоны, а в перспективе, и бытовые приборы, оснащенные блоками «интеллекта», например холодильники, микроволновые печи и кондиционеры. То есть требовался способ соединять друг с другом любые устройства без проводов и сложной настройки.

Для беспроводного интерфейса Bluetooth отведен частотный диапазон от 2,4 до 2,48 ГГц. А поскольку радиоэфир полон помех естественного и искусственного происхождения, то для использования в интерфейсе Bluetooth был предложен принцип скачкообразной перестройки частоты в пределах отведенного диапазона по псевдослучайному алгоритму, например до 1600 изменений в секунду между 79 точками. Кроме передачи данных через интерфейс, можно организовать три голосовых канала.

Дальность надежного соединения для устройств с интерфейсом Bluetooth составляет 10 м (существуют варианты с дальностью до 100 м). Скорость передачи данных в асимметричном режиме до 721 Кбит/с, а в симметричном — 432,6 Кбит/с в обоих направлениях. Для целей безопасности в спецификации интерфейса предложено использовать аутентификацию и шифрование данных с ключом длиной от 8 до 128 битов.

Для приемопередатчиков интерфейса Bluetooth разработаны специальные маломощные чипы, которые можно встраивать в любые устройства, даже в наушники и микрофоны.

В настоящее время существуют самые разнообразные устройства, которые используют данную технологию. Но можно отметить, что наибольший интерес этот интерфейс вызвал у фирм, которым требовалось быстро и просто создавать небольшие локальные сети. Правда, в процессе эксплуатации нашлись и слабые стороны спецификации Bluetooth, которая разрешала пользователям не применять шифрование данных, что привело к новой разновидности хакерских атак на локальные сети.

## 12. ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ

### 12.1. Классификация вычислительных систем

Создание вычислительных систем (ВС) – наиболее реальный путь разрешения противоречия между непрерывно растущими потребностями в быстродействующих и надежных средствах вычислений и пределом технических возможностей ЭВМ на данном этапе развития.

**Вычислительная система** представляет собой сложный комплекс, состоящий из разнообразных технических средств соответствующего программного обеспечения. Как технические, так и программные средства имеют модульную структуру построения, позволяющую наращивать ее в зависимости от назначения и условий эксплуатации системы. Программная автоматизация управления вычислительным процессом осуществляется с помощью ОС.

Первыми ВС были однопроцессорные мультипрограммные ЭВМ, высокая производительность которых была достигнута за счет распределения во времени основных устройств системы между программами.

Дальнейшее повышение производительности ЭВМ было достигнуто за счет мультиобработки программ (задач), т. е. за счет разбиения программ на отдельные блоки и параллельной обработки этих блоков на нескольких обрабатывающих устройствах, входящих в состав ВС. Мультиобработка позволяет не только повысить производительность, но и сократить время выполнения отдельных программ, которые могут разбиваться на части и распределяться между различными обрабатывающими устройствами.

Первым типом ВС с мультиобработкой был *многомашинный комплекс* (МК) – многомашинная ВС. В состав МК объединялись различные ЭВМ с классической структурой, имеющие возможность обмениваться информацией. На рис. 12.1 – 12.5 представлены типовые структуры вычислительных систем.

На рис. 12.1 представлена структура двухмашинной ВС. Каждая ЭВМ имеет ОП, ВЗУ (внешнее запоминающее устройство), ПфУ (периферийное устройство), подключаемые к центральной части ЭВМ – процессору (ПР) с помощью каналов ввода-вывода (КВВ), и работает под управлением своей ОС. Обмен информацией между ЭВМ1 и

ЭВМ2 осуществляется через системные средства обмена (ССО) в результате взаимодействия ОС машин между собой.

Основной недостаток многомашинной ВС – не в полной мере эффективно используется оборудование комплекса. Достаточно в ВС в каждой ЭВМ выйти из строя по одному устройству (даже разных типов), как вся ВС становится неработоспособной.

Следующим шагом в направлении дальнейшего увеличения производительности ВС явилось создание многопроцессорных ВС с мультиобработкой, в составе которых содержится два или несколько процессоров (ПР), работающих с единой ОП, общий набор каналов ввода-вывода (КВВ) и ВЗУ (рис. 12.2). Наличие единой ОС делает возможным автоматическое распределение ресурсов системы на различных этапах ее работы. В результате достигается высокая «живучесть» ВС, позволяющая в случае отказа отдельных модулей перераспределить нагрузку между работоспособными, обеспечив тем самым выполнение наиболее важных для ВС функций.

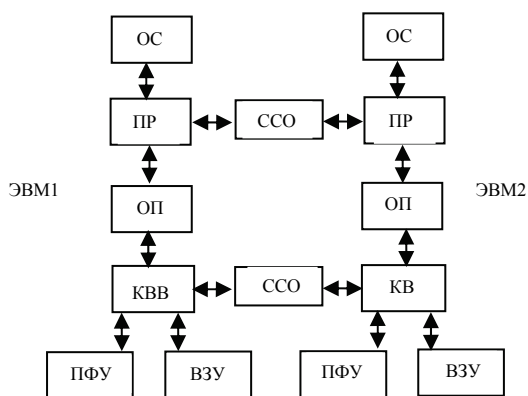


Рис. 12.1 Структура многомашинной ВС

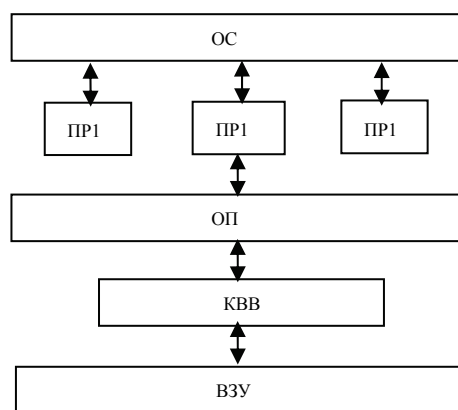


Рис. 12.2 Структура многопроцессорной ВС

К недостаткам многопроцессорных ВС относят трудности, возникающие при реализации общего поля ОП, ВЗУ, а также при разработке специальной ОС.

Дальнейшее развитие идей мультиобработки привело к созданию крупных многопроцессорных систем высокой производительности, получивших название высокопараллельных ВС. Такие ВС в зависимости от структуры могут одновременно обрабатывать множественный поток данных или команд. Под потоком команд понимается последовательность команд, выполняемых ВС, а потоком данных –



последовательность данных, обрабатываемых под управлением потока команд.

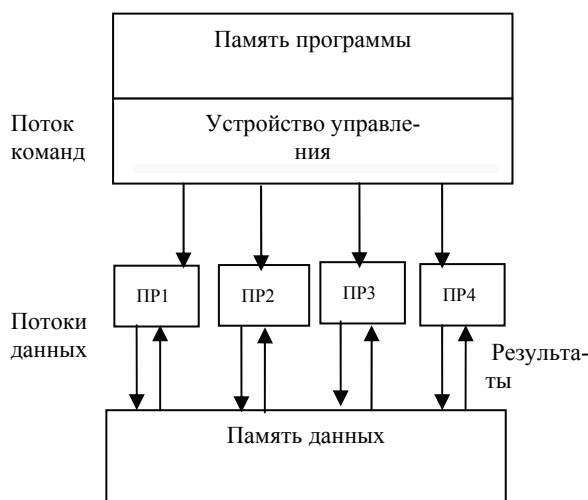


Рис. 12.3. Многопроцессорная ВС структуры типа ОКМД

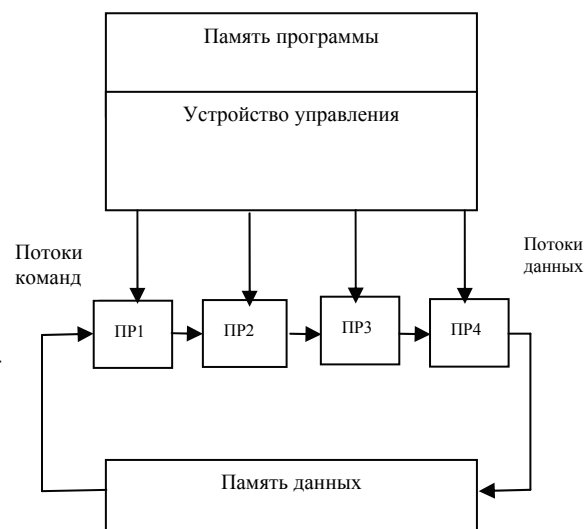


Рис. 12.4. Многопроцессорная ВС структуры типа МКОД

Высокопараллельные ВС структуры типа ОКМД (единочный поток команд и множественный поток данных) получили название *матричных* ВС (рис. 12.3). Они содержат некоторое количество одинаковых сравнительно простых быстродействующих процессоров (ПР), соединенных друг с другом так, что образуется сетка (матрица), в узлах которой размещаются ПР. Все ПР выполняют одну и ту же команду, но над разными операндами, доставляемыми процессорам из памяти несколькими потоками данных.

Высокопараллельные ВС структуры типа МКОД (множественный поток команд и одиночный поток данных) получили название *конвейерных*. ВС. Такие ВС (рис. 12.4) содержат цепочку последовательно соединенных ПР, так что информация на выходе одного ПР является входной информацией для другого ПР. Каждый ПР обрабатывает соответствующую часть задачи, передавая результаты соседнему ПР, который использует их в качестве исходных данных.

Так, например, операция сложения чисел с плавающей запятой может быть разделена на 4 этапа: сравнение порядков, выравнивание порядков, сложение мантисс, нормализация результата. В конвейерной ВС все эти этапы вычислений будут выполняться отдельными процессорами, образующими конвейер.

Высокопараллельные ВС по сравнению с многопроцессорными ВС обеспечивают более высокую производительность, надежность. Однако при этом усиливаются недостатки – усложнение управления системой, трудность программирования и малая загрузка системы.

Первые два недостатка компенсируются благодаря применению БИС и специальных языков программирования. Третий недостаток приводит к тому, что большинство высокопараллельных ВС ориентируется на специализированное применение.

Реализация идей мультипрограммной работы потребовала пересмотра способов взаимодействия оператора с программой. Прежде всего возникла необходимость в большом количестве абонентских пунктов, позволяющих нескольким пользователям одновременно взаимодействовать с ЭМВ по своим программам. Часто такие пульта находятся на значительном расстоянии от ЭВМ и связаны с ней телефонными или телеграфными каналами. В качестве оконечных устройств (терминалов) обычно используют дисплеи. Такая связь ЭВМ со многими дистанционно расположенными пультами получила название мпогопультной.

Многопультный режим работы, обеспечивающий одновременный доступ к ЭВМ многих пользователей, вызвал соответствующую организацию их обслуживания. Обычно такое обслуживание осуществляется в режиме разделения времени между пользователями.

Вычислительная система с разделением времени последовательно и циклически опрашивает терминалы всех пользователей, записывает информацию обратившихся абонентов и обслуживает их в той же последовательности. Каждому пользователю (абоненту) выделяется определенный, строго ограниченный квант машинного времени (максимум доли секунд). В течение таких циклически повторяющихся квантов времени осуществляется выполнение каждой программы. В идеальном случае интервал времени между включениями в работу одной и той же программы не должен превышать обычное время реакции человека. Тогда программист, управляющий своей программой, не будет ощущать прерывистого характера ее выполнения и у него создастся иллюзия индивидуального общения с ВС.

Вычислительные системы с разделением времени являются основой сети вычислительных центров коллективного пользования.

Чтобы удовлетворять предъявляемым требованиям к ВС, они должны иметь:

- развитую ОС, обеспечивающую одновременное выполнение различных программ и организующую доступ пользователя к стандартным программам;
- трансляторы с языков программирования, облегчающих работу программистов по подготовке программ;
- средства, обеспечивающие динамическое распределение памяти между программами, а также свободное перемещение программ в процессе вычислений;
- средства защиты памяти и программ от вмешательства других программ;
- датчик времени (таймер), позволяющий в соответствии с запросами пользователей выделять им необходимое время для работы, по истечении которого ВС автоматически переключается на выполнение других программ;
- как аппаратурные, так и программные средства с целью организации приоритетов для одновременно ждущих программ.

Многопроцессорные и многомашинные ВС классифицируют по различным признакам. Рассмотрим некоторые из них.

*По назначению* ВС делятся на универсальные и специализированные. *Универсальные* ВС предназначены для решения широкого круга задач, специализированные — для решения определенного круга задач. *Специализированные* ВС, как правило, должны иметь аппаратурные и программные средства, предназначенные специально для этой системы.

*По типу оборудования* ВС подразделяются на однородные и неоднородные. *Однородные* системы содержат несколько однотипных ЭВМ (или процессоров), *неоднородные* – разнотипные ЭВМ (или процессоры). Основным недостатком однородных ВС – неполная загруженность отдельных ЭВМ (процессоров) во время работы. В целях повышения эффективности использования ЭВМ (процессоров) используются неоднородные ВС. Например, более производительная центральная ЭВМ системы выполняет обработку информации, а менее производительные ЭВМ1, ЭВМ2 и ЭВМ3 осуществляют ввод и вывод информации, ее предварительную обработку и передачу в центральную ЭВМ. Коммутатор при передаче информации в центральную ЭВМ и выдаче из нее результатов настраивается на выбор соответствующей ЭВМ 2-го уровня (рис. 12.5).

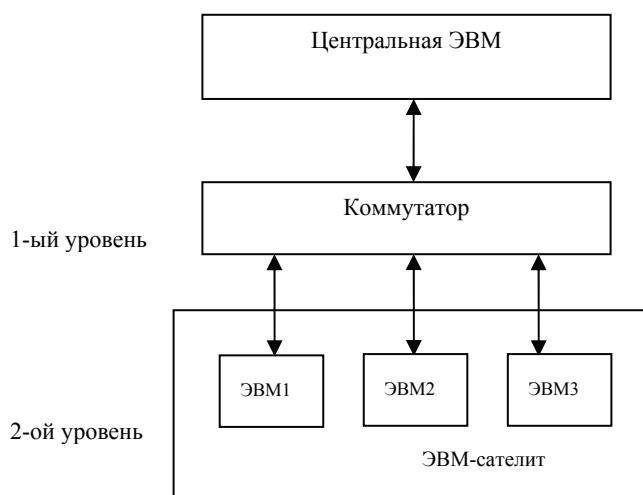


Рис. 12.5 Иерархическая структура многомашинной ВС

Вычислительные системы с иерархической структурой могут иметь и более двух уровней иерархии. ЭВМ, выполняющие предварительную обработку информации, часто, называют машинами-спутниками.

*По типу структуры* ВС разделяются с постоянной и переменной структурами. Под структурой ВС понимают состав системы и схемы функциональных и управляющих связей между ее элементами. В системах с *постоянной* структурой в процессе ее функционирования не изменяется состав функциональных и управляющих связей между ее элементами. *Переменную* структуру имеют адаптивные системы, т. е. такие, у которых структура изменяется на основе анализа текущей информации. Подобные системы позволяют достичь оптимального состояния в любых изменяющихся условиях функционирования.

*По степени централизации управления* ВС разделяются на централизованные, децентрализованные и со смешанным управлением.

В *централизованных* ВС все функции управления сосредоточены в одном элементе, в качестве которого используется одна из ЭВМ, называемая машиной-директором, или центральный процессор.

В *децентрализованной* ВС каждый процессор или ЭВМ действуют автономно, решая свои задачи.

В системе *со смешанным управлением* ВС разбивается на группы взаимодействующих ЭВМ (или процессоров), в каждой из которых осуществляется централизованное управление, а между группами – децентрализованное.

## 12.2. Многомашинные вычислительные системы

В настоящее время наиболее широко используют двухмашинные вычислительные комплексы, которые могут работать в одном из следующих режимов.

1. 100%-ное горячее резервирование. Обе ЭВМ в этом режиме исправны и работают параллельно, выполняя одни те же операции над одной и той же информацией (дуплексный режим). После выполнения каждой команды результаты преобразования сравниваются и при их совпадении процесс вычислений продолжается. При этом в памяти обеих ЭВМ в каждый момент находится одна и та же информация. При обнаружении несовпадения в результатах обработки неисправная ЭВМ выводится на ремонт, а исправная продолжает работать под контролем встроенной в ЭВМ системы автоматического контроля.

2. Одна исправная ЭВМ решает задачи без дублирования, а другая ЭВМ находится в режиме «Профилактика», в котором осуществляется прогон контролирующих тестов. Если основная ЭВМ не в состоянии выполнить задачу, то резервная может прекратить «Профилактику» и начать работу параллельно с основной.

3. Обе ЭВМ работают в автономном режиме со своим набором ПФУ по автономным рабочим программам.

Задание режимов работы вычислительного комплекса (ВК) возможно программным путем, или с помощью команд прямого управления или с пульта управления комплекса.

*По типу организации* многомашинные ВК можно разделить на две группы: незвязанные и связанные.

*Несвязанные* ВК разрабатывались с целью разгрузить центральный процессор от выполнения операций по вводу-выводу данных извне. Они состоят из центральной и периферийной ЭВМ, между которыми нет прямого физического соединения и отсутствуют какие-либо совместно исполняемые аппаратные средства. Целесообразность их применения определяется тем, что операции ввода-вывода информации и вычисления совмещаются во времени. Небольшая и недорогая ЭВМ выполняет медленные операции ввода-вывода информации (считывание с перфокарт, печать и т. п.), а центральная ЭВМ — высокоскоростные операции, обмениваясь в процессе вычислений с ВЗУ.

*Связанные* ВК включают несколько ЭВМ, которые совместно используют общие аппаратные средства, т. е. в этих ВС возможно электрическое сопряжение между процессорами. В таких ВК, обе ЭВМ могут выполнять, две различные программы автономно или во взаимодействии друг с другом.

Наличие нескольких тесно связанных ЭВМ в составе единой ВС позволяет существенно уменьшить время вычислений благодаря параллельному выполнению на отдельных ЭВМ различных подзадач (пакетов программ), входящих в общую задачу. Основное условие эффективного использования таких ВК – координация работы всех ЭВМ с помощью управляющей программы ОС, которая составляет список подзадач, подлежащих решению, и распределяет их между ЭВМ. Благодаря возможности передачи не только числовой, но и командной информации между ЭВМ можно передавать программы и части программ. Обмен программами значительно упрощает управление ВК и позволяет при загрузке какой-либо ЭВМ часть нагрузки передать другой ЭВМ, упрощает создание библиотеки стандартных программ, пригодных для любой машины ВК.

Для реализации межмашинной связи могут использоваться как средства, имеющиеся в составе ЭВМ, так и средства, предусмотренные специально для работы в составе данного ВК.

### **12.3. Многопроцессорные вычислительные системы**

В настоящее время особое внимание уделяется созданию многопроцессорных ВК. Основной целью при разработке таких ВК является повышение производительности систем за счет: обеспечения возможности параллельного выполнения независимых задач, повышения эффективности работы и улучшения распределения нагрузки в системе, обеспечения наиболее экономичного обслуживания экстренных заданий и заданий при пиковых нагрузках, достижения высокого коэффициента эффективного использования ресурсов для создания новых типов архитектуры комплекса.

В многопроцессорных ВС при решении задач с небольшими емкостями памяти возможно одновременное решение на разных процессорах. Если в какой-либо интервал времени требуется резкое увеличение емкости памяти, то вся память отдается для решения задачи.

Основные особенности построения многопроцессорных ВК заключаются в следующем:

- система включает в себя один или несколько процессоров;
- центральная память системы должна находиться в общем пользовании и к ней должен быть обеспечен доступ от всех процессоров системы;
- система должна иметь общий доступ ко всем устройствам ввода-вывода, включая каналы;
- система должна иметь единую ОС, управляющую всеми аппаратными и программными средствами;
- в системе должно быть предусмотрено взаимодействие элементов аппаратного и программного обеспечения на всех уровнях: на уровне системного программного обеспечения, на программном уровне при решении задач пользователей (возможность перераспределения заданий), на уровне обмена данными и др.

В многомашинных ВС связь может осуществляться только на информационном уровне.

Важнейшее значение для организации многопроцессорной ВС имеют способы соединения между собой различных функциональных блоков системы, т. к. эффективность такой системы определяется степенью параллельности или совмещения по времени работы всех устройств системы.

### **Мультипроцессорные и избыточные системы**

В современных ПК встречаются варианты установки нескольких (двух или более) процессоров на одной системной шине. При этом возможны конфигурации с *симметричной мультипроцессорной обработкой* (Symmetric Multi-Processing, SMP) и *избыточным контролем функционирования* (Functional Redundancy Checking, FRC).

В конфигурации с избыточным контролем функционирования два процессора (пара Master и Checker) выступают как один логический. Основной процессор (Master) работает в обычном однопроцессорном режиме. Проверочный процессор (Checker) выполняет все те же операции «про себя», не управляя шиной, и сравнивает выходные сигналы основного (проверяемого) процессора с теми сигналами, которые он генерирует сам, выполняя те же операции без выхода на шину. В случае расхождения вырабатывается сигнал ошибки, который может обрабатываться как прерывание. FRC-контроль применяют только в особо ответственных системах. Поддержка FRC появилась, начиная с процессоров Intel Pentium, но не во всех последующих моделях; она имеется и у процессоров фирмы AMD.

### **Симметричные мультипроцессорные системы**

Наиболее популярной целью объединения процессоров является *симметричная мультипроцессорная обработка (SMP)*. В системе SMP каждый процессор решает свою задачу, порученную ему операционной системой. В документе Intel «Multiprocessor Specification» (MPS) симметрия рассматривается в двух аспектах:

- симметрия памяти — все процессоры пользуются общей памятью, работают с одной копией ОС;
- симметрия ввода-вывода — все процессоры разделяют общие устройства ввода-вывода и общие контроллеры прерываний.

Система может быть симметричной по памяти, но асимметричной по прерываниям от ввода-вывода, если для обслуживания этих прерываний выделяется собственный процессор. В x86 симметрию по прерываниям обеспечивают контроллеры APIC. Аппаратная (физическая) реализация SMP может быть различной, здесь уже есть своя история:

- объединение нескольких физических процессоров на одной локальной шине — процессоры Pentium, P6, Pentium 4;
- подключение каждого процессора к системной плате (с общей памятью) выделенными шинами — процессоры Athlon;
- подключение к каждому процессору собственного ОЗУ и их объединение с периферийными устройствами через HyperTransport — процессоры Athlon-64, Opteron;
- размещение на одном кристалле нескольких логических процессоров с разделяемыми операционными блоками — «гиперпотоковые» (hyperthreading) модели Pentium 4;
- размещение на одном кристалле нескольких независимых процессорных ядер с разделяемым вторичным кэшем — мультиядерные модели Pentium 4.

Применение SMP требует поддержки со стороны BIOS, ОС и приложений (чтобы работать быстрее, они должны быть многопоточными). Поддержку SMP имеют такие ОС, как Novell NetWare, Microsoft Windows NT/2000x/XP и различные диалекты Unix. Цена мультипроцессорных версий ОС, как правило, значительно выше цены соответствующих однопроцессорных версий. Поначалу это было препятствием к применению гиперпотоковых (и мультиядерных) процессоров в обычных системных платах. Теперь достигнута договоренность о том, что число процессоров, на которое лицензируется ОС, соответствует числу физиче-



ских (отдельно покупаемых и устанавливаемых) процессоров. Это открывает возможности широкого распараллеливания на уровне процессоров (ускорение работы CISC-процессоров x86 за счет микроархитектуры обходится все дороже).

### **Объединение процессоров на локальной шине**

В первых мультипроцессорных системах на базе Pentium и P6 процессоры объединялись с помощью общей локальной («фасадной», FSB) шины, через которую они связывались с чипсетом системной платы. В таких системах, в принципе, могут использоваться процессоры одной модели. Частоты ядра у них должны совпадать (внешняя частота у них, естественно, едина). Процессоры, объединенные общей локальной шиной, разделяют ресурсы и этой шины, и компьютера (память и периферийные устройства). В каждый момент времени шиной может управлять только один процессор, по определенным правилам они меняются ролями.

Поскольку каждый из процессоров имеет свой внутренний первичный кэш, интерфейс шины обязан поддерживать *согласованность данных* (когерентность) во всех иерархических ступенях оперативной памяти — в первичном и вторичном кэшах, в основной памяти (в Pentium вторичный кэш у процессоров общий). Эта задача решается с помощью локальных циклов слежения, воспринимаемых процессором, даже не управляющим шиной в данный момент.

Симметричные системы имеют специальные механизмы *арбитража* доступа к локальной шине. Процессор — текущий владелец шины — отдает управление шиной другому процессору по его запросу только по завершении операции.

Интерфейс Pentium (начиная со второго поколения) позволяет на одной локальной системной шине устанавливать два процессора, при этом почти все их одноименные выходы просто объединяются. Роль конкретного процессора в системе фиксирована — она определяется внешними сигналами во время спада сигнала RESET. Один из процессоров назначается *первичным* (primary processor), или *загрузочным* (Bootstrap Processor, BSP), другой — *вторичным* (Dual Processor, DP). После сигнала RESET сразу начинает функционировать только первичный процессор (BSP), выполняя программный код инициализации. Вторичный процессор начинает функционирование только после приема соответствующего сообщения по шине APIC, посланного в ходе программы инициализации.

В процессорах P6 заложены более развитые возможности SMP. Системная шина P6, в отличие от локальной шины Pentium, изначально

ориентирована на разделяемое управление множеством симметричных (до четырех на шине) и несимметричных (до восьми) агентов. Сокет 8 (Pentium Pro) и слот 2 (Pentium II Хеоп) позволяют объединять до четырех процессоров, слот 1 (Pentium II) допускает объединение не более двух процессоров. Какой из процессоров станет первичным, определяется по загрузочному протоколу — здесь нет жесткой аппаратной привязки роли процессора к его «географическому» адресу. Это позволяет повысить надежность системы SMP, поскольку любой процессор может без механического вмешательства во время инициализации взять на себя роль BSP. Протокол мультипроцессорной инициализации работает на шине APIC, он позволяет управлять инициализацией до 15 процессоров. Процессоры могут пользоваться содержимым «чужого» кэша без его предварительной выгрузки в основную память.

### **Гиперпотоковые и мультиядерные процессоры**

В процессорах Pentium 4 (начиная с частоты 3,06 ГГц) применяется *гиперпотоковая* (hyperthreading) технология: один физический процессор одновременно может выполнять два потока инструкций x86. Для фон-неймановской машины это означает, что физический процессор (микросхема, устанавливаемая в сокет) имеет два комплекта архитектурных (прикладных и системных) регистров. В каждом комплекте имеется, естественно, свой указатель инструкций, «идущий» по своему потоку. Таким образом, речь идет о двух *логических процессорах*, физически расположенных на одном кристалле микросхемы. Эти логические процессоры совместно используют ряд общих микроархитектурных блоков физического процессора (вторичный кэш, исполнительные блоки арифметико-логического устройства). Такое разделение позволяет повысить эффективность функционирования исполнительных блоков: один даже «гиперконвейер» Pentium 4, исполняющий инструкции x86, плотно загрузить их работой не может. Конечно, логические процессоры не являются полностью независимыми — иногда приходится ожидать освобождения ресурса, занятого соседом.

Следующий шаг в этом направлении — *мультиядерные процессоры*, в которых на одном кристалле объединены общей шиной несколько функционально-законченных процессоров. Структурная схема мультиядерных процессоров Pentium 4 приведена на рис. 12.6. В них каждое ядро обладает собственным кэшем L2 (и, естественно, кэшем данных L1 и кэшем трасс). Размер кэша L2 каждого ядра может достигать 1–2 Мбайт. Интерфейс системной шины у двух ядер может быть как общим, так и

раздельным. Двухъядерный процессор, предназначенный для использования в относительно небольших системах (размером до четырех ядер), фактически является двумя отдельными процессорами, каждый со своим интерфейсным блоком (рис.12.6, а). Их интерфейсные сигналы объединяются в общую системную шину на системной плате. Процессоры для более крупных систем содержат общий блок интерфейса системной шины (рис. 12.6, б), поскольку объединение на одной шине интерфейсов более чем двух процессоров для высокоскоростных шин вызывает ряд трудностей.

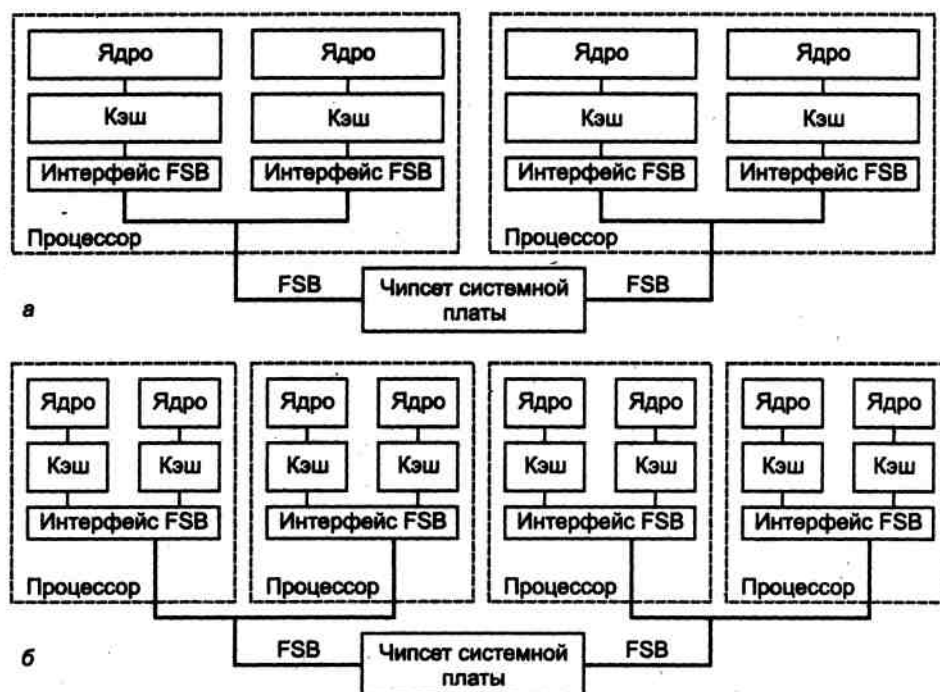


Рис. 12.6. Структура мультиядерных процессоров Pentium 4:  
а - с раздельными интерфейсами системной шины, б — с объединенным интерфейсом

Заметим, что мультипроцессорные кристаллы в RISC-архитектурах начали применять намного раньше, причем в более мощных вариантах (с интегрированным контроллером памяти).

### Мультипроцессорные системы Athlon и Opteron

Фирма AMD долгое время для своих процессоров x86 механизм SMP не поддерживала, ограничиваясь поддержкой FRC (избыточный контроль). Первым из процессоров AMD, поддерживающих SMP, стал процессор Athlon. Используют в нем шина EV6 является двухточечной, и для мультипроцессорных систем чипсет должен каждому процессору

предоставлять собственную шину в монопольное владение. Таким образом, средством объединения нескольких процессоров является северный хаб чипсета системной платы. Основным разделяемым ресурсом становится контроллер памяти, расположенный в чипсете.

В процессорах Athlon-64 и Opteron (ядро Hammer) подход к построению системы (даже однопроцессорной) иной: каждый процессор имеет собственный контроллер памяти (DDR SDRAM) с отдельной шиной, к которой непосредственно подключаются модули памяти. У Athlon-64 шина данных памяти 64-битная у Opteron — 128-битная (двухканальная память). Такое решение позволяет обеспечить минимальные задержки доступа к памяти — они могут быть порядка 45 нс. Для сравнения заметим, что при «классическом» подключении памяти к чипсету задержка доступа со стороны процессора составляет около 100 нс. Для связи с остальными компонентами (периферией и другими процессорами) используется высокопроизводительный интерфейс HyperTransport. У процессора Opteron, предназначенного для серверов (1–8-процессорных систем) и рабочих станций (1–4-процессорных систем), имеется три 16-битных интерфейса HyperTransport с суммарной пропускной способностью  $3 \cdot 6,4 = 19,2$  Гбайт/с. HyperTransport обеспечивает «прозрачный» доступ к любым компонентам из любой точки системы транзакциями чтения и записи по адресам памяти или ввода-вывода.

Некоторые варианты систем SMP на базе Opteron приведены на рис. 12.7. Здесь каждый процессор входит в систему со своей памятью, ему доступна память и других процессоров. Конечно, время доступа к «чужой» памяти несколько больше, чем к своей; оно зависит от числа *хопов* (hop) через интерфейс HyperTransport. Так, для 4-процессорной системы время доступа при локальном обращении (нулевое число хопов) составляет 100 нс, при обращении к памяти ближайшего соседа (один хоп) — 118 нс, к самой дальней памяти два хопа) — 136 нс. Обращение к локальной памяти по сравнению с однопроцессорным вариантом происходит несколько дольше — очевидно, это издержки, вносимые протоколом обеспечения когерентности кэш-памяти разных процессоров.

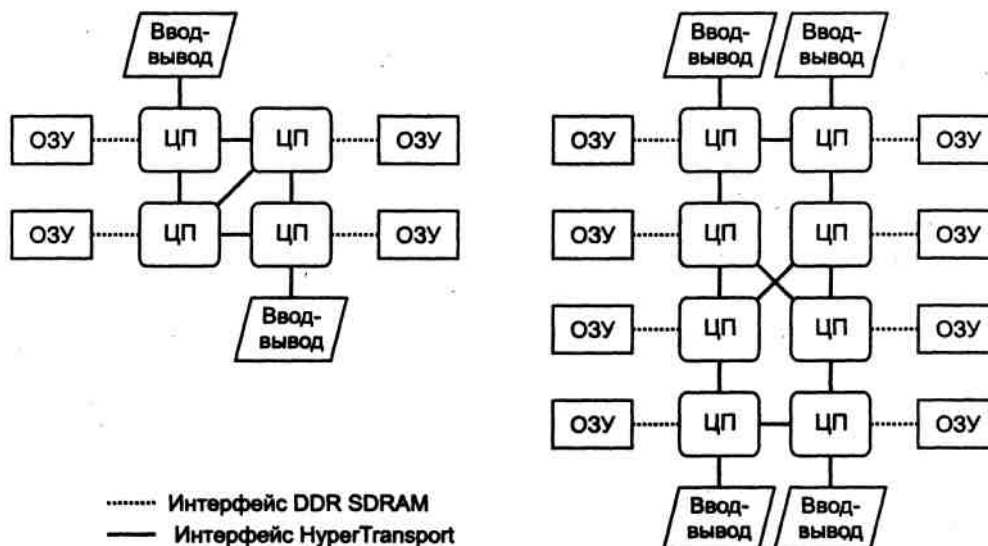


Рис.12.7. Структуры мультипроцессорных систем Opteron

Варианты использования вышеприведенных топологий соединения могут быть различными. Это может быть и симметричная по памяти система; правда, при распределении памяти и исполняемых потоков по процессорам желательно, чтобы ОС учитывала разницу во времени доступа к «близкой» и «дальней» памяти. Возможны построения и несимметричных (по памяти) систем. Если у объединяемых процессоров организовать раздельное адресное пространство, то получается уже *мультикомпьютерная система*. Каждый компьютер — это процессор, память и средства ввода-вывода, и при совместной параллельной работе они могут обмениваться лишь сообщениями (но не по сети, а через интерфейс HyperTransport).

## 13. КОММУНИКАЦИОННЫЕ УСТРОЙСТВА

### 13.1. Модемы

#### Введение

Современные издательства и типографии представляют собой высокотехнологичные программно-аппаратные комплексы, оснащенные самым современным компьютерным оборудованием и высокоскоростными каналами связи. Данные технологии используются в процессе подготовки к изданию газет, которая определяется необходимостью оперативной доставки их полос в региональные типографии. В 1964 г. ленинградский тираж газеты «Правда» был отпечатан с фотокопий, принятых по каналам связи из Москвы. Так в нашей стране началась практика децентрализованной печати ежедневных и периодических изданий. И на сегодня процесс доставки полос в типографии, находящиеся как в столице, так и за ее пределами, является основным «потребителем» телекоммуникаций в газетной индустрии.

В прежние времена создатель первичного оригинала приносил рукопись, затем получал на проверку первую корректуру, затем ему высылали для окончательной проверки последнюю корректуру перед печатанием, сводку. Типография полностью контролировала допечатный процесс. Однако с внедрением настольных издательских систем типография этот контроль утратила. Причиной этого стали закрытые для типографии компьютерные программы. Теперь дизайн и допечатную подготовку (вторичный оригинал) работы смог выполнять автор (издатель) и посылать буквально в любую типографию. В прошлом полиграфист проверял (печатную форму) или, позднее, пленку и фактически «держал в руках» выполняемую работу. Сегодня заказчик взял эту роль на себя. В таком стандартизированном виде заказы пересылались в типографию вначале с курьером или специальными круглосуточными службами курьерской доставки, а позднее по телекоммуникационным каналам.

Сегодня заказы создаются в электронном виде и пересылаются по скоростным линиям связи типа T-1, T-3, DSL, ISDN, через Интернет. Через несколько лет уже не будут использовать ни дискеты, ни компакт-диски. Все это из-за телекоммуникационного оборудования, а в частности модемов, с помощью которых стало возможным умень-

шить затраты на доставку оригиналов в типографии, увеличить скорость подготовки и издания всех типов печатной продукции.

### **Что такое модем**

*Модем* — самый важный после компьютера элемент, необходимый для установления связи с Интернетом. Без него компьютер, даже самый мощный, — «вещь в себе». Связь с внешним миром он поддерживает только за счет гибких дисков и других съемных носителей. Даже если имеется локальная сеть, но без модема — проблемы те же, что и у одиночного компьютера.

Название «модем» произошло от двух слов — модулятор-демодулятор (MOdulator-DEModulator) — и обозначает устройство для преобразования цифрового сигнала в аналоговый и обратно. Все происходит аналогично тому, как радиопередатчик преобразовывает человеческий голос в радиоволны, а радиоприемник восстанавливает голос уже в другой точке пространства.

*Факс-модем* отличается от простого модема тем, что может принимать и отправлять факсы. Факс — это графическое изображение документа. Если нужно срочно отправить документ, а времени для доставки по почте нет, пользуются факсом. Принципиальной разницы в работе обычного модема и факса нет. Практически все современные модемы являются факс-модемами.

Модемы бывают внешние (английское сокращение ext, от слова external) и внутренние (английское сокращение int, от слова internal).

Внешний модем представляет собой самостоятельный прибор, как радиоприемник или телевизор. Это небольшая прямоугольная коробка, у которой на лицевой панели расположены от одной до восьми индикаторов, мигающих во время приема и передачи информации. Внешний модем имеет свой блок питания, который является отдельным устройством и совмещен с сетевой вилкой. Модем получает от него по двум проводам напряжение питания 5 или 9 В. В семействе внешних модемов существуют модели, которые подключаются к USB-интерфейсу компьютера. В этом случае питание модем получает непосредственно от компьютера по USB-кабелю, правда, это относится только к тем моделям, которые не имеют другого интерфейса для связи с компьютером.

Внутренний модем отличается от внешнего тем, что у него нет своего блока питания и выполнен он в виде платы, которая вставляет-

ся в слот PCI компьютера точно так же, как и все остальные вспомогательные платы: видеоадаптер, звуковая карта и т. п.

Win-модем — карта PCI, выполняющая модуляцию/демодуляцию (с точки зрения ЦП — аппаратно). При этом протоколы модема (команды, коррекция, сжатие) выполняются программно на ЦП. Это вызывает ряд проблем: драйверы неустойчивы, могут оказаться доступными только под Windows (откуда и название). Модем вызывает дополнительную загрузку ЦП, работе модема мешают другие задачи (как и он задачам). Soft-модем (софт-модем) в этом плане еще хуже, поскольку и модуляция/демодуляция осуществляется программно (на плате модема остается только телефонная часть и АЦП/ЦАП). Win- и Soft-модемы могут быть картами для AMR или CNR (специальный слот на системной плате), связанными со средствами доставки (AC-Link) AC'97.

Софт-модемы имеют привлекательную для многих пользователей цену — в два-три раза ниже, чем у аппаратного модема с аналогичными функциями. Правда, при покупке софт-модема следует учитывать, что под этим термином скрывается огромное разнообразие конструкций, со своими достоинствами и недостатками. Поэтому нужно рассмотреть разницу между различными типами этих модемов.

В модеме выделяют несколько основных узлов, которые отвечают за выполнение определенных функций. В аппаратном модеме эти узлы выполнены на микросхемах или чипсете, а в софт-модеме либо все функции, либо только часть функций выполняются центральным процессором компьютера.

Основной элемент модема — это контроллер, который отвечает за взаимодействие всех узлов модема, обеспечивает связь с компьютером и обрабатывает AT-команды.

За качество работы модема с аналоговым сигналом отвечает сигнальный процессор (DSP — Digital Signal Processor, или Data Pump). Он отвечает за формирование и детектирование сигналов в соответствии с основными модемными протоколами, модуляцию и демодуляцию аналогового сигнала, взаимодействие с телефонной линией. Для хранения программного обеспечения и данных в модеме присутствует несколько типов постоянной и оперативной памяти. Постоянная память (ПЗУ) хранит ту информацию, которая не должна стираться после выключения питания. В ней находится библиотека инструкций, которые определяют работу модема. Программное обеспечение модема, содержащееся в ПЗУ, называют прошивкой. Если это



ПЗУ представляет собой флэш-память (FlashRom), то можно самому сменить прошивку, например для того, чтобы «научить» модем работать с новым протоколом или исправить ошибки, допущенные разработчиком. Кроме того, в модеме присутствует ПЗУ, в которое записываются текущие настройки модема. Такое ПЗУ называют NVRAM. В него можно также записать номера телефонов провайдера или друга, с которым вы часто связываетесь по модему.

Софт-модемы, в которых все перечисленные выше функции переложены на центральный процессор, излишне требовательны к качеству линий связи. Такие модемы могут хорошо работать только на современных компьютерах. Устанавливать их в компьютер на i486 просто бессмысленно, ведь даже процессор 800 Мгц загружается на 100%, когда подобный модем устанавливает связь. Софт-модемы, в которых функции обработки входного сигнала остались в ведении самого модема, например выполненные на чипсетах Rockwell, почти равноценны обычным (полным, аппаратным) модемам. Более того, в них часто имеются сервисные функции, которых нет в аппаратных модемах. Когда в софт-модеме отсутствует ПЗУ с прошивкой, программное обеспечение (прошивка) загружается в модем во время старта операционной системы. Можно считать это преимуществом, так как для того, чтобы установить, например, новый протокол связи, достаточно сменить программное обеспечение. Также к софт-модемам относятся модемы, интегрированные с материнской платой. К сожалению, разработчики плат идут по пути удешевления производства, поэтому чаще всего такие модемы мало подходят для работы с не цифровыми АТС (у них плохая защита от электрических перегрузок в линии).

В компьютере для управления модемом должен быть установлен соответствующий драйвер, работающий под управлением операционной системы и сопутствующего коммуникационного программного обеспечения.

Наиболее удобен в работе внешний модем. Он выполняется в виде отдельного прибора, который можно установить недалеко от компьютера. Его можно подключить к любому компьютеру, использовать дома и на работе. Визуальная индикация состояния телефонной линии позволяет отслеживать процесс работы модема. Недостатки: провода, соединяющие модем с компьютером и с отдельным блоком питания, занимают место и путаются. Кроме того, цена внешнего модема примерно на 20 – 30% больше, чем внутреннего такой же марки.

Внутренний модем выполняется в виде стандартной платы, устанавливаемой в слот компьютера. Выпускаются модемы для слотов типа ISA и типа PCI. Для установки внутреннего модема необходимы вскрытие корпуса компьютера и правильная установка модема в слот материнской платы, что требует определенных навыков. Один из недостатков внутреннего модема — можно не заметить, что не произошло отключение от провайдера, в результате чего придется платить деньги за неиспользованное время на линии. Другой недостаток — некоторые модемы для своего перезапуска требуют выключения питания компьютера.

Преимущества встроенных модемов — низкая цена и отсутствие дополнительных блоков на рабочем месте. Недостатки — необходимость вскрытия системного блока для установки модема и возможные сложности конфигурирования системных ресурсов, а иногда и отсутствие свободного слота. Следует отметить и низкую защищенность компьютера в случае попадания высокого электрического потенциала на телефонный вход модема (например, при ударе молнии в открытую телефонную линию). Правда, если линия не защищена ограничителем перенапряжений, то и внешнее подключение модема не станет надежной защитой компьютера.

Все упомянутые ранее модемы предназначены для линий связи, созданных для телефонных разговоров. Подключение к Интернету через такие линии называют соединением *по коммутируемой линии связи*. Сигнал от вашего телефона (модема) проходит через ряд механических или электронных коммутаторов, которые соединяют вас с абонентом (провайдером). Достигнуть хорошего качества и надежности связи в таких линиях весьма трудно. Для создания компьютерной сети или обеспечения надежной связью офиса (в том числе и доступом в Интернет) используют выделенную линию связи, которую арендуют у телефонной компании. Для выделенных линий разработано много различных способов передачи цифрового сигнала (стандартов), например ISDN, Frame Relay, ADSL и т. д. Соответственно для каждого стандарта производится разнообразный спектр оборудования, в том числе специализированные модемы.

### **Протоколы связи**

*Протоколы связи* (точнее, протоколы модуляции) являются правилами установления, поддержки и окончания сеанса связи между двумя модемами. Правила однозначно определяют метод преобразо-

вания цифрового сигнала в аналоговый (способ модуляции), скорость передачи данных, последовательность передачи служебных данных и полезной информации. Все этапы сеанса связи жестко регламентируются, потому что только так можно добиться безукоризненной работы канала связи. Протоколов достаточно много, рассмотрим самые распространенные из них. Они совместимы друг с другом сверху вниз. Это означает, что новые модемы при установлении связи со старыми автоматически переключаются на тот протокол, с которым те могут работать.

**V.22bis.** Это самый простой протокол из применяемых в Интернете. Обеспечивает скорость передачи данных до 2 400 бит/с. Долгое время V.22bis был самым распространенным стандартом. Устаревшие факс-модемы используют его, даже если работа с факсами идет на скорости 9 600 бит/с. Но, несмотря на низкую по современным понятиям скорость, эти модемы позволяют неплохо путешествовать по Интернету (без загрузки графики), а в условиях больших помех на телефонной линии обеспечивают устойчивую связь.

**V.32.** Обеспечивает скорость до 9 600 бит/с.

**V.32bis.** Доработанный вариант V.32, предусматривающий скорость до 14 400 бит/с. Массовый выпуск модемов с этим протоколом пришелся на начало взрывного роста Интернета. Позволяет передать 1 Мбайт данных примерно за 10 мин.

**V.34.** Принятый в 1994 г., стандарт рассчитан на скорость до 28 800 бит/с.

**V.34bis.** В Рекомендации V.34 была предусмотрена возможность передачи данных со скоростью 33 600 бит/с, но юридически она закреплена в виде поправки к стандарту в 1996 г. Практически скорость 33 600 бит/с является предельной для аналоговых телефонных линий.

**V.90.** В основе стандарта лежит технология K56Flex фирмы Rockwell, несовместимая с x2 фирмы US Robotics. Пропускная способность канала увеличивается до уровня 56 кбит/с, это значение сильно превышает теоретический предел пропускной способности телефонной линии. Достижение такой скорости стало возможным при появлении цифровых АТС, внутри которых есть скоростная магистраль, позволяющая с помощью специального оборудования передавать данные быстрее, чем обычно. Кроме того, скорость 56 кбит/с обеспечивается только в направлении от провайдера к пользователю,

тогда как в обратном направлении скорость ограничена прежним пределом 33 600 бит/с.

**V.92.** Первая редакция протокола утверждена 4 июля 2000 г. Кроме увеличения скорости передачи до 48 кбит/с за счет использования новой разновидности импульсно-кодовой модуляции, в рекомендации оговариваются новые сервисные функции для модемной связи (сервисные услуги АТС) и использование протокола V.44 для исправления и сжатия информации, что позволит получать web-страницы со скоростью до 300 кбит/с. Первые модели модемов с новым протоколом появились в 2001 г.

Протокол V.92 позволяет модему запомнить характеристики телефонной линии (процедура начальной тренировки). При последующем подключении время соединения с удаленным модемом уменьшается в два раза — с 25 – 30 до 12 секунд.

С появлением протокола V.92 вошло в практику использование услуги Call Waiting (ожидание входящего вызова, Modem-on-Hold). Ее суть состоит в том, что, когда вы говорите по телефону или работаете в Интернете, к вам можно, хоть ваша линия и занята, дозвониться. Такая ситуация определяется по коротким высоким сигналам в динамике трубки, которые продолжаются в течение 5-7 секунд. В этом случае можно кратковременно соединиться с новым абонентом, чтобы решить — с кем вы будете продолжать разговор.

Служба переадресации вызовов позволяет принимать звонки в другом месте. Если ваша АТС предлагает эту услугу, то надо в настройках вашего модема указать определенный код активации.

Некоторые телефонные компании поддерживают службы, позволяющие изменять вид вызывающих гудков в зависимости от того, откуда идет вызов. Например, для различных комбинаций коротких и длинных гудков можно настроить модем таким образом, что он самостоятельно будет решать — какой программе передать управление телефонной линией (это может быть факс-программа или автоответчик).

В документации на современные модемы вы найдете упоминание еще об одном протоколе — **V.80**. Этот протокол предназначен для проведения видеоконференций. Заметьте, что к Интернету данный протокол не имеет отношения, а предназначен для прямых соединений между компьютерами — через телефонную или выделенную линию.

## Скорость

Скорости передачи данных могут обозначаться тремя различными единицами измерения: боды (baud), биты в секунду (bps — bits per second), а также символы или байты в секунду (cps — characters per second). Они обозначают одно и то же — скорость передачи данных модемом. Казалось бы, что значения в разных единицах должны находиться в ясных соотношениях друг с другом. На самом деле численные величины находятся в непонятном соотношении. На модемах сегодня чаще всего пишут, например, 33 600 бит/с — максимальное количество битов, которое модем может передать за одну секунду. При приеме файлов используют величину байт в секунду (часто употребляют также килобайты в минуту). Если взять калькулятор и посчитать соотношение величин (при практическом перекачивании файлов), становится окончательно непонятно, как боды, биты и байты связаны друг с другом. Первая величина вошла в употребление еще во времена появления телеграфа. Количество посылок в секунду стали называть бодами, в честь французского изобретателя Бодо. Посылкой считается одно изменение сигнала в линии: например, если вы за секунду включите свет в своей комнате два раза, то вы достигли скорости четыре бода. Например, у модема со скоростью передачи 14 400 бит/с скорость в бодах — всего лишь 2 400, поэтому для современных модемов правильнее говорить о скорости модуляции, которая всегда выражается в бодах.

Телефонные линии рассчитаны на передачу голоса, частотный диапазон которого принимается равным 3 300 Гц. Поэтому модемы не могут «разговаривать» на частотах выше разрешенных — все равно никто не услышит. А ведь хочется передавать данные все быстрее и быстрее. Для этого ученые придумали очень простую вещь — в одну посылку стали упаковывать не одно сообщение (бит), а несколько. Так, для модема на 14 400 бит/с в одной посылке упаковано шесть бит.

Протокол **V.34** позволяет в одну посылку упаковать до девяти бит, а **V.90** еще больше. Весь фокус в способах модуляции сигнала. Если использовать одновременно несколько видов модуляции — амплитудную и фазовую, то можно в одну посылку упаковать несколько битов. Поэтому когда работает современный модем на 56 кбит/с, он «свистит» на частоте не 56 кГц, а на значительно более низкой частоте, укладывающейся в частотный диапазон телефонного канала. Кроме того, пользуясь не идеальностью амплитудно-частотной характеристики (АЧХ) линий связи, модемы также используют частоты 3 450, 3 600 и 3 750 Гц, но гарантий, что модем сможет их использовать, нет. Поэтому на практике

модемы не всегда могут работать на скорости 33,6 или 56 кбит/с. Например, более «популярными» у модемов являются скорости 31 200, 45 333 и 48 000 бит/с.

Разберемся со следующей величиной, которая используется при измерении скорости приема и передачи информации, — с байтами в секунду. Часто говорят также «символов в секунду», так как один символ (в принятой в Интернете системе) соответствует одному байту, или восьми битам. Теоретически модем на 28 800 бит/с может передавать информацию со скоростью  $28\,800/8 = 3600$  байт/с, или 216 Кбайт/мин. В реальности такого не происходит, скорость передачи даже в идеальном случае всегда ниже. Дело в том, что модем добавляет к передаваемому байту два служебных бита, стартовый и стоповый, чтобы другой модем мог разобраться, где начало передачи, а где конец. Возможная скорость передачи информации с учетом служебных сигналов. Она достигает для нашего модема  $28\,800/10 = 2880$  байт/с или 172 Кбайт/мин. Но иногда скорость передачи превышает теоретическую и составляет, например, 3 Кбайт/с, а то и выше. Это связано уже с режимом передачи информации, который может быть асинхронным и синхронным.

Асинхронный способ передачи данных в основном используется при установлении связи между модемами, когда служебные сигналы — стоповые и стартовые — нужны для правильного определения начала и конца полезной информации. Ведь модемы не могут знать, когда начинается полезная информация, а когда «шумит» линия связи. Поэтому они ждут стартовый бит для начала приема и стоповый — для его окончания. В общем, выходит десять бит на один 8-битовый знак: букву, цифру. Плюс неизбежное время на перерывы между передачами — для проверки правильности полученной информации и определения, что передавать дальше. Такое бывает и при плохой линии связи, когда модемы не могут перейти в синхронный режим передачи, а скорость передачи файлов в этом случае еле достигает половины возможной максимальной скорости модема или вообще падает до минимума.

Второй режим передачи — синхронный. В Интернете принято передавать информацию пакетами, в которых объединяется сразу много байтов, например 64. В этом случае асинхронный метод используется только для определения начала пакета и его конца. Внутри пакета информация передается в синхронном режиме, то есть не обрамляется служебными битами. Дополнительно пакет снабжается контрольным числом для проверки достоверности информации. Получив такой пакет, принимающий модем проверяет контрольное чис-

ло, и если оно правильное — делает запрос на передачу нового пакета данных. Если произошла ошибка, принимающий модем просит передающий модем повторить передачу пакета. При хорошей связи за счет пакетных передач в синхронном режиме для нашего модема можно достичь скорости приема/передачи примерно 200 Кбайт/мин.

### 13.2.Технология ADSL

Технология ADSL (асимметричная цифровая абонентская линия) была разработана в Северной Америке в середине 1990-х г. для предоставления таких услуг, которые требуют асимметричной передачи данных, например, видео по запросу, когда требуется передавать большой поток данных в сторону пользователя, а в сторону сети от пользователя передается гораздо меньший объем данных.

Требовалось очень высокое качество передачи (коэффициент битовых ошибок BER не менее  $1 \cdot 10^{-9}$ ), потому что была нужна технология передачи потоков видеоданных с кодировкой MPEG, характеризующейся очень высокой кодировкой и низкой избыточностью, когда даже единичные ошибки оказывают значительное влияние на качество изображения. Платой за это послужило увеличение времени ожидания. Именно поэтому ранние системы ADSL имели задержку в 20 мсек по сравнению с ISDN-BA или HDSL, которые не превышали предел в 1,25 мс.

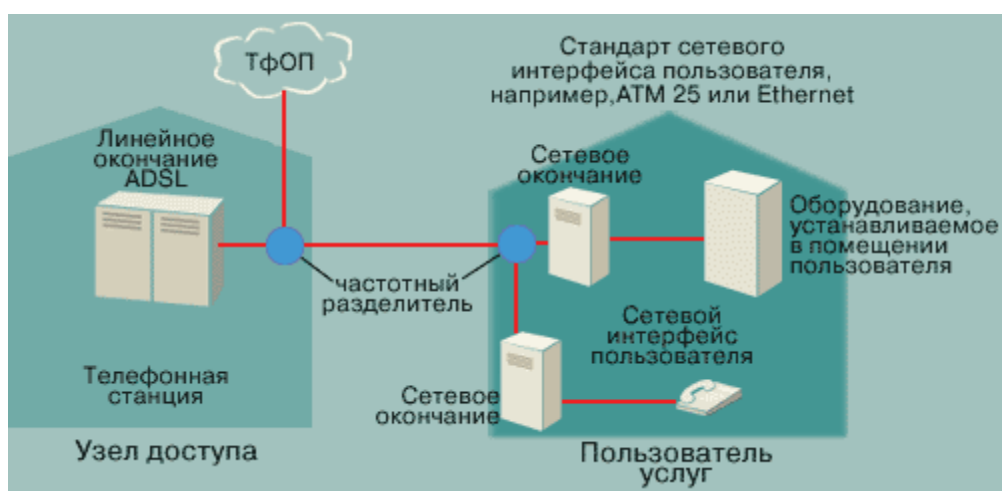


Рис. 13.1. Концепция асимметричной цифровой абонентской линии

Кроме того, что технология ADSL обеспечивает крайне асимметричную передачу данных, она также отличается от ISDN-BA/HDSL тем, что позволяет использовать ту же самую пару проводов для традиционной телефонной связи. Для этого используются специальные устройства разделения сигналов (сплиттеры) (рис. 13.1).

ADSL использует технологию FDD (частотное разделение для обеспечения дуплексной связи), которая позволяет выделить одну полосу частот для восходящего потока данных (направление от пользователя в сторону станции), а другую полосу частот — для нисходящего потока данных (от станции в сторону пользователя). Это позволяет расширить используемую полосу частот приблизительно до 1 МГц. В некоторых вариантах ADSL используется технология подавления эхосигналов, что позволяет еще лучше использовать доступный спектр частот, перекрывая часть диапазона, занятого нисходящим потоком данных, передачей данных в восходящем направлении. На рис. 13.2 показан пример использования технологии FDD для разделения восходящего и нисходящего потоков данных и сплиттера.

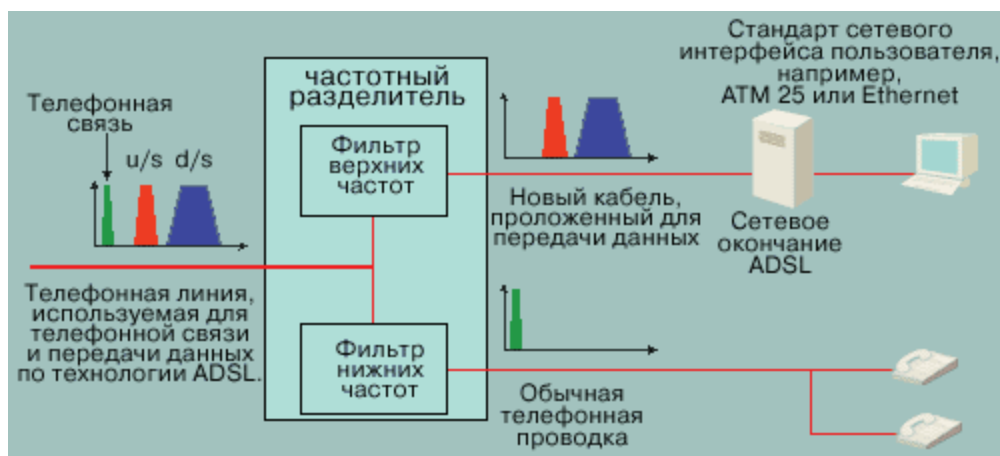


Рис. 13.2. Пример ADSL с частотным уплотнением и сплиттером.

Скорости нисходящего и восходящего потоков данных изменяются и зависят от длины абонентской телефонной линии и уровня шумов. В основном на ADSL оказывают влияние помехи на дальнем конце линии, в то время как ISDN-BA и HDSL обычно имеют ограничения из-за помех на ближнем конце линии. Именно то, что основные ограничения касаются помех на дальнем конце линии, позволяет достигнуть скорости передачи для нисходящего потока данных в



2 Мбит/с по большинству абонентских телефонных линий. Полоса частот, используемая для восходящего потока данных, по технологии значительно уже, поэтому обычно скорость передачи восходящего потока данных достигает нескольких сотен Кбит/с.

### **ADSL**

Выбор ADSL-технологии оправдан в тех случаях, когда требуется достичь высоких скоростей на больших расстояниях (до 5,5 км) от АТС до абонента и когда необходимо сохранить телефонную связь. Технология ADSL обеспечивает скорость «нисходящего» потока данных в пределах от 1,5 Мбит/с до 8 Мбит/с и скорость «восходящего» потока данных от 640 кбит/с до 1,5 Мбит/с. ADSL позволяет передавать данные со скоростью 1,54 Мбит/с на расстояние до 5,5 км по одной витой паре проводов. Скорость передачи 6–8 Мбит/с может быть достигнута при передаче данных на расстояние до 3,5 км по проводам диаметром 0,5 мм.

Обычно ADSL-технология используется в жилом секторе. Для одновременной передачи голосового трафика и данных по одной телефонной линии устанавливаются сплиттеры. Они могут быть как встроенными, так и внешними. Кроме «классической» технологии ADSL (включает в себя «облегченную» версию — ADSL Lite), современные концентраторы поддерживают улучшенные модификации технологии ADSL — ADSL2, ADSL2+.

Они разрабатывались с учетом возросших требований провайдеров и конечных пользователей к концентраторам. В ADSL2 увеличена скорость и дальность передачи информации, реализована функция адаптивного изменения скорости. Благодаря этим изменениям стала возможной поддержка большого количества новых приложений и дополнительных услуг. В ADSL2+ увеличена вдвое скорость приема информации на расстояниях до 1,5 км.

### **Типы интерфейсов к транспортной сети**

Выбор оборудования для сети абонентского доступа во многом зависит от того, какая технология используется на магистральной сети: ATM, IP/Ethernet, SDH. В соответствии с этим выбираются и интерфейсы концентратора.

## Ethernet

Современные тенденции перехода на протокол IP заставляют производителей оборудования широкополосного доступа DSL применять в своих решениях более перспективные технологии Ethernet. Прежде всего, это влияет на стоимость решения, существенно снижается его цена.

Современные транспортные сети Metro Ethernet достигли скоростей 10 Гбит/с и продолжают активно развиваться. Для доступа к таким высокопроизводительным сетям концентраторы DSLAM комплектуются интерфейсами Fast или Gigabit Ethernet в зависимости от плотности портов устройства. Однако при использовании транспортного интерфейса Ethernet в оборудовании широкополосного доступа встает задача обеспечения качества обслуживания абонентов. Для этого в оборудовании доступа DSLAM используются несколько механизмов. Прежде всего, это возможность создания виртуальных частных сетей VLAN в соответствии со стандартом IEEE802.1Q. Кроме того, для обеспечения параметров качества в оборудовании DSLAM должна быть реализована функция приоритезации трафика Ethernet в соответствии с IEEE802.1p.

## Частотные разделители для ADSL

Использование самых современных технологий не сводит к нулю потребность в традиционном общении по телефону. Технология ADSL позволяет одновременно и передавать данные, и говорить по телефону, используя одну и ту же телефонную линию. Доступная полоса пропускания медного телефонного кабеля разделяется на две неравные части — частоты до 4 кГц используются для традиционной телефонной связи, а вся остальная полоса используется для высокоскоростной передачи данных (рис. 13.3). Использование новых возможностей медного телефонного кабеля равноценно обнаружению новой золотой жилы на давно заброшенном руднике. Голос и данные объединяются на абонентской стороне линии и разделяются на станционной стороне той же линии (и наоборот) с помощью специальных разделяющих устройств — сплиттеров (splitter) или *частотных разделителей* (ЧР).

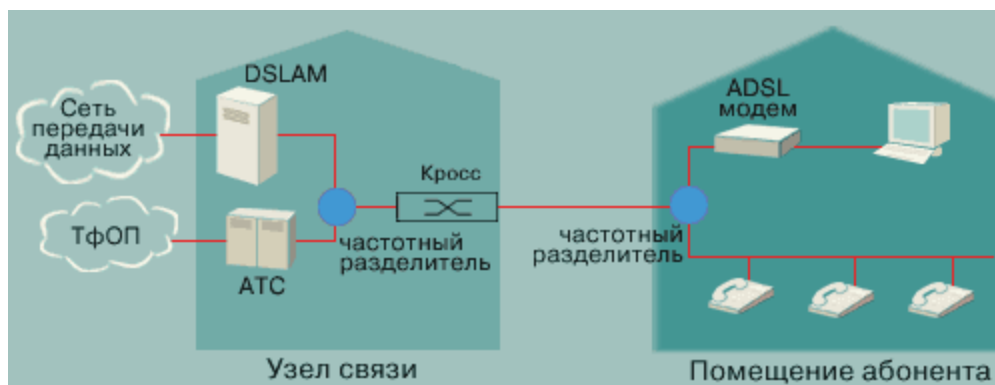


Рис. 13.3. Использование частотных разделителей на линиях ADSL

Одновременное использование абонентской телефонной линии для традиционной телефонной связи и для высокоскоростной передачи данных всегда выдвигалось и выдвигается в число самых главных преимуществ технологии ADSL. Одним из ключевых устройств этой технологии и является ЧР.

ЧР представляет собой обычное электронное устройство — фильтр, имеющий три порта. Некоторые производители предлагают *активные* ЧР иногда даже встроенные в модем. Однако при использовании такого ЧР авария электропитания или отказ модема приведут к потере телефонной связи. *Пассивный* же ЧР поддерживает функционирование телефонной линии даже тогда, когда неисправен модем, потому что традиционный телефонный аппарат получает питание постоянного тока по абонентской телефонной линии.

ЧР включает в себя фильтр нижних частот и фильтр верхних частот. Фильтр верхних частот представляет собой комбинацию устройств, находящихся на ЧР и модеме ADSL (обычно разделительные по постоянному току конденсаторы). Фильтр нижних частот, предназначенный для выделения голосового канала, как правило, находится на самом ЧР. Одной из основных функций ЧР является фильтрация импульсных помех, создаваемых телефонным аппаратом и коммутационным оборудованием телефонной станции и способных помешать нормальной работе модема. С другой стороны осуществляется фильтрация высокочастотного сигнала модема ADSL, который может снизить качество работы традиционной телефонной связи. Большинство используемых на практике ЧР — пассивные. Это объясняется их более высокой надежностью, которая крайне необходима для обеспечения надежной телефонной связи.

Рабочие параметры системы определяются следующими характеристиками ЧР (рис. 13.4):

- минимальным вносимым затуханием в диапазоне частот 0 – 4 кГц и максимальными обратными потерями в том же частотном диапазоне;
- особенностями характеристики в переходной полосе частот (4 кГц – 16 кГц;
- максимальной симметрией линии;
- плоской амплитудно-частотной характеристикой в диапазоне голосовых частот и диапазоне ADSL;
- согласованием импеданса портов во всем диапазоне используемых частот;
- максимальной изоляцией между портами;
- минимальной задержкой сигнала в рабочем диапазоне частот;
- сопротивлением по постоянному току каждого тракта передачи сигнала;
- использованием подходящих конденсаторов, включая компенсацию перенапряжения из-за наличия индуктивности между контактами разъема;
- граничными частотами, необходимыми для поддержания факсимильной связи и работы модема V.90.

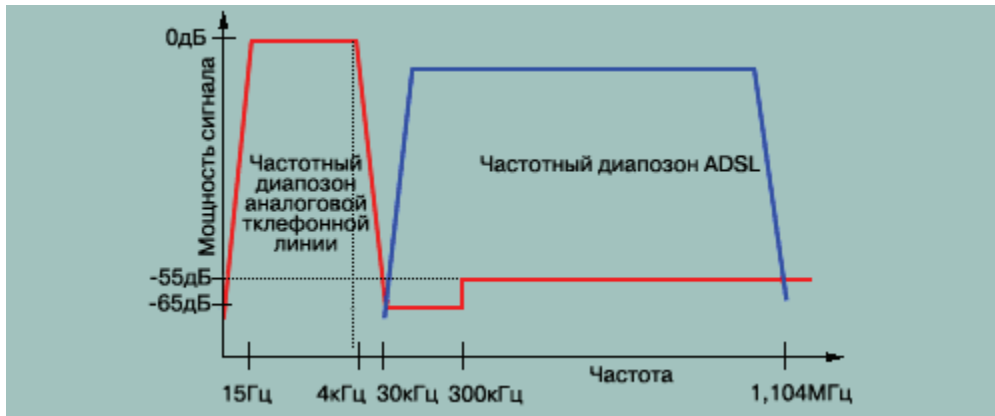


Рис. 13.4. Пример разделения спектра голосового сигнала и ADSL по ANSI T1.413

При выборе ЧР в обязательном порядке должны учитываться все технические стандарты, используемые в местной телефонной связи, а также технические характеристики кабельной сети.

Из всего вышеизложенного можно сделать вывод, что ЧР, используемый при установке систем ADSL, не является предметом массового потребления, а требует тщательно продуманного подхода в каждом случае установки. Кроме того, он оказывает существенное влияние на технические и организационные аспекты использования телефонных кабельных линий. При выборе «правильного ЧР» обязательно следует учитывать несколько важных моментов.

По мере внедрения систем высокоскоростной передачи данных на абонентской кабельной сети на станции будет устанавливаться все большее количество оборудования, поэтому через какое-то время возникнет вопрос нехватки места для его установки. Чем больше портов имеет ЧР, тем меньше места в сумме будет занимать оборудование; поэтому при выборе ЧР следует учитывать количество портов каждого устройства.

Использование оборудования должно соответствовать принятым стандартам. На практике часто приходится сталкиваться с кабельными линиями, проложенными десятки лет назад. Поэтому использование оборудования, соответствующего всем необходимым стандартам (а лучше превышающих эти стандарты), приносит гораздо меньше проблем.

Технология ADSL позволяет системе адаптироваться к условиям работы. Она реагирует на разнообразные отрицательные внешние воздействия, например, на шумы, снижением скорости передачи данных. Чем меньше воздействие ЧР оказывает на сигнал, тем более высокими будут характеристики системы. Более высокие характеристики означают увеличение рабочего расстояния, а значит получение возможности обслуживания удаленных абонентов, что расширяет рынок предоставления услуги высокоскоростной передачи данных.

### **Описание технологии ADSL**

Прежде всего, ADSL является технологией, позволяющей превратить витую пару телефонных проводов в тракт высокоскоростной передачи данных. Линия ADSL соединяет два модема ADSL, которые подключены к каждому концу витой пары телефонного кабеля (рис. 13.5). При этом организуются три информационных канала — «нисходящий» поток передачи данных, «восходящий» поток передачи данных и канал обычной телефонной связи (рис. 13.6). Канал телефонной связи выделяется с помощью фильтров, что гарантирует работу телефона даже при аварии соединения ADSL.

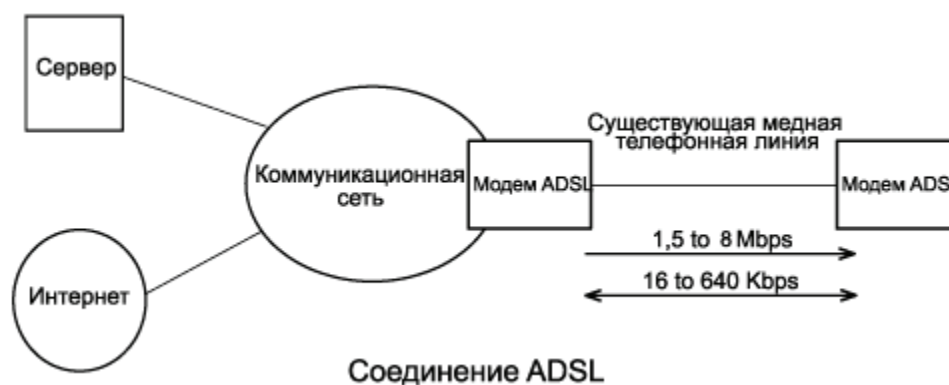


Рис. 13.5 Система связи с ADSL-модемом

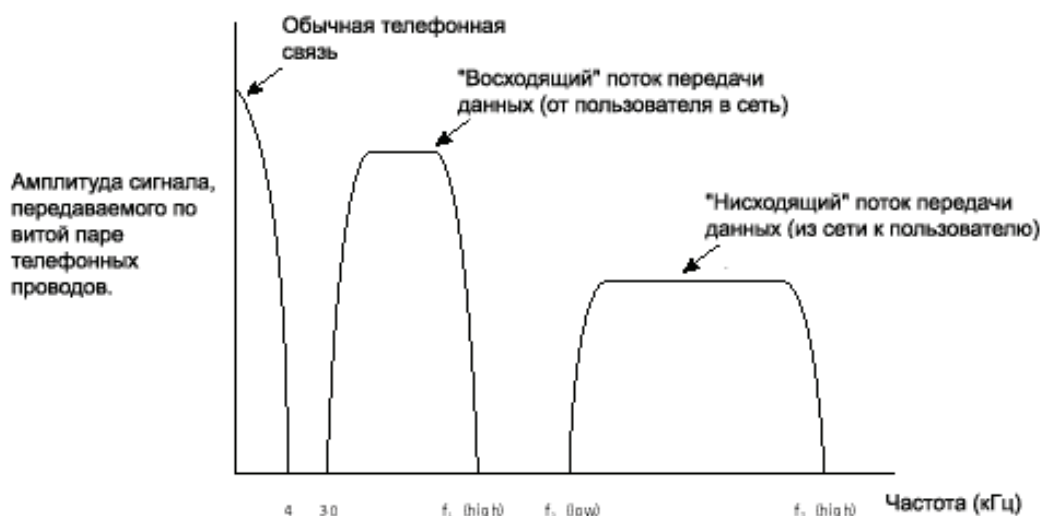


Рис. 13.6. Организация информационных каналов

ADSL является асимметричной технологией — скорость «нисходящего» потока данных (т. е. тех данных, которые передаются в сторону конечного пользователя) выше, чем скорость «восходящего» потока данных (в свою очередь передаваемого от пользователя в сторону сети). Сразу же следует сказать, что не следует искать здесь причину для беспокойства. Скорость передачи данных от пользователя (более «медленное» направление передачи данных) все равно значительно выше, чем при использовании аналогового модема. Фактически же она также значительно выше, чем ISDN (Integrated Services Digital Network — Интегральная цифровая сеть связи).

Для сжатия большого объема информации, передаваемой по витой паре телефонных проводов, в технологии ADSL используется

цифровая обработка сигнала и специально созданные алгоритмы, усовершенствованные аналоговые фильтры и аналого-цифровые преобразователи. Телефонные линии большой протяженности могут ослабить передаваемый высокочастотный сигнал (например, на частоте 1 МГц, что является обычной скоростью передачи для ADSL) на величину до 90 дБ. Это заставляет аналоговые системы модема ADSL работать с достаточно большой нагрузкой, позволяющей иметь большой динамический диапазон и низкий уровень шумов. На первый взгляд система ADSL достаточно проста — создаются каналы высокоскоростной передачи данных по обычному телефонному кабелю. Но, если детально разобраться в работе ADSL, можно понять, что данная система относится к достижениям современной технологии.

Технология ADSL использует метод разделения полосы пропускания медной телефонной линии на несколько частотных полос (также называемых несущими). Это позволяет одновременно передавать несколько сигналов по одной линии. Точно такой же принцип лежит в основе кабельного телевидения, когда каждый пользователь имеет специальный преобразователь, декодирующий сигнал и позволяющий видеть на экране телевизора футбольный матч или увлекательный фильм. При использовании ADSL разные несущие одновременно переносят различные части передаваемых данных. Этот процесс известен как частотное уплотнение линии связи (Frequency Division Multiplexing — FDM) (рис. 13.7, а). При FDM один диапазон выделяется для передачи «восходящего» потока данных, а другой диапазон для «нисходящего» потока данных. Диапазон «нисходящего» потока, в свою очередь, делится на один или несколько высокоскоростных каналов и один или несколько низкоскоростных каналов передачи данных. Диапазон «восходящего» потока также делится на один или несколько низкоскоростных каналов передачи данных. Кроме этого может применяться технология эхокомпенсации (Echo Cancellation), при использовании которой диапазоны «восходящего» и «нисходящего» потоков перекрываются (рис. 13.7, б) и разделяются средствами местной эхокомпенсации.

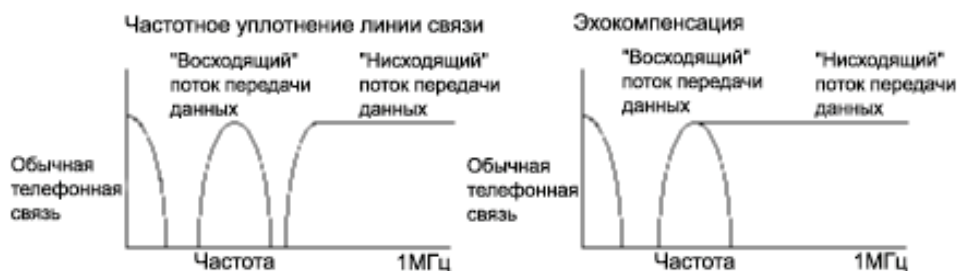


Рис. 13.7 Принцип ADSL-технологии:  
а – Частотное уплотнение линии связи; б – Эхокомпенсация

Именно таким образом ADSL может обеспечить, например, одновременную высокоскоростную передачу данных, видеосигнала и факса. И все это без прерывания обычной телефонной связи, для которой используется та же телефонная линия. Технология предусматривает резервирование определенной полосы частот для обычной телефонной связи (или POTS — Plain Old Telephone Service). Удивительно, как быстро телефонная связь превратилась не только в «простую» (Plain), но и в «старую» (Old); получилось что-то вроде «старой доброй телефонной связи». Однако, следует отдать должное разработчикам новых технологий, которые все же оставили телефонным абонентам узенькую полосу частот для живого общения. При этом телефонный разговор можно вести одновременно с высокоскоростной передачей данных, а не выбирать одно из двух. Более того, даже если у вас отключат электричество, обычная «старая добрая» телефонная связь будет работать по-прежнему и с вызовом электрика у вас никаких проблем не возникнет. Обеспечение такой возможности было одним из разделов оригинального плана разработки ADSL. Даже одна эта возможность дает системе ADSL значительное преимущество перед ISDN.

Одним из основных преимуществ ADSL над другими технологиями высокоскоростной передачи данных является использование самых обычных витых пар медных проводов телефонных кабелей. Совершенно очевидно, что таких пар проводов насчитывается гораздо больше, чем, например, кабелей, проложенных специально для кабельных модемов. ADSL образует, если можно так сказать, «наложенную сеть». При этом дорогостоящей и отнимающей много времени



модернизации коммутационного оборудования (как это необходимо для ISDN) не требуется.

ADSL является технологией высокоскоростной передачи данных, но насколько высокоскоростной? Учитывая, что буква «А» в названии ADSL означает «asymmetric» (асимметричная), можно сделать вывод, что передача данных в одну сторону осуществляется быстрее, чем в другую. Поэтому следует рассматривать две скорости передачи данных: «нисходящий» поток (передача данных от сети к вашему компьютеру) и «восходящий» поток (передача данных от вашего компьютера в сеть).

Факторами, влияющими на скорость передачи данных, являются состояние абонентской линии (т. е. диаметр проводов, наличие кабельных отводов и т. п.) и ее протяженность. Затухание сигнала в линии увеличивается при увеличении длины линии и возрастании частоты сигнала, и уменьшается с увеличением диаметра провода. Фактически функциональным пределом для ADSL является абонентская линия длиной 3,5 – 5,5 км при толщине проводов 0,5 мм. В настоящее время ADSL обеспечивает скорость «нисходящего» потока данных в пределах от 1,5 Мбит/с до 8 Мбит/с и скорость «восходящего» потока данных от 640 Кбит/с до 1,5 Мбит/с. Общая тенденция развития данной технологии обещает в будущем увеличение скорости передачи данных, особенно в «нисходящем» направлении.

### **Сравнение ADSL с другими технологиями**

Для того, чтобы оценить скорость передачи данных, обеспечиваемую технологией ADSL, необходимо сравнить ее с той скоростью, которая может быть доступна пользователям, использующим другие технологии. Аналоговые модемы позволяют передавать данные со скоростью от 14,4 до 56 Кбит/с. ISDN обеспечивает скорость передачи данных 64 Кбит/с на канал (обычно пользователь имеет доступ к двум каналам, что в сумме составляет 128 Кбит/с). Различные технологии DSL дают пользователю возможность передавать данные со скоростью 144 Кбит/с (IDSL), 1,544 и 2,048 Мбит/с (HDSL), «нисходящий» поток 1,5 – 8 Мбит/с и «восходящий» поток 640 – 1500 Кбит/с (ADSL), «нисходящий» поток 13 – 52 Мбит/с и «восходящий» поток 1,5 – 2,3 Мбит/с (VDSL). Кабельные модемы имеют скорость передачи данных от 500 Кбит/с до 10 Мбит/с (при этом следует учитывать, что полоса пропускания кабельных модемов делится между всеми пользователями, одновременно имеющими доступ к данной линии, поэтому число

одновременно работающих пользователей оказывает значительное влияние на реальную скорость передачи данных каждого из них). Цифровые линии E1 и E3 имеют скорость передачи данных соответственно 2,048 Мбит/с и 34 Мбит/с.

При использовании технологии ADSL полоса пропускания той линии, с помощью которой конечный пользователь связан с магистральной сетью, принадлежит этому пользователю всегда и целиком.

### **Преимущества технологии ADSL**

Прежде всего, скорость передачи данных. Цифры были указаны выше. Причем эти цифры не являются пределом. В стандарте ADSL 2 реализованы скорости 10 Мбит/с «нисходящего» и 1 Мбит/с «восходящего» потока при дальности до 3 км, а в технологии ADSL 2+, фигурируют скорости «нисходящего» потока в 20, 30 и 40 Мбит/с (соответственно по 2,3 и 4 парам).

Для того, чтобы подключиться к сети Интернет или к ЛВС, не нужно набирать телефонный номер. ADSL создает широкополосный канал передачи данных, используя уже существующую телефонную линию. После установки модемов ADSL пользователь получает постоянно установленное соединение. Высокоскоростной канал передачи данных всегда готов к работе — в любой момент, когда вам это требуется.

Полоса пропускания линии принадлежит пользователю целиком. В отличие от кабельных модемов, которые допускают деление полосы пропускания между всеми пользователями (что в значительной мере оказывает влияние на скорость передачи данных), технология ADSL предусматривает использование линии только одним пользователем.

Технология ADSL позволяет полностью использовать ресурсы линии. При обычной телефонной связи используется около одной сотой пропускной способности телефонной линии. Технология ADSL устраняет этот «недостаток» и использует оставшиеся 99% для высокоскоростной передачи данных. При этом для различных функций используются различные полосы частот. Для телефонной (голосовой) связи используется область самых низких частот всей полосы пропускания линии (приблизительно до 4 кГц), а вся остальная полоса используется для высокоскоростной передачи данных.

Многофункциональность данной системы является не самым последним аргументом в ее пользу. Т. к. для работы различных функ-

ций выделены различные частотные каналы полосы пропускания абонентской линии, ADSL позволяет одновременно передавать данные и говорить по телефону. Можно звонить по телефону и отвечать на звонки, передавать и принимать факсы, одновременно с этим находясь в сети Интернет или получая данные из корпоративной сети ЛВС. Все это по одной и той же телефонной линии.

ADSL открывает совершенно новые возможности в тех областях, в которых в режиме реального времени необходимо передавать качественный видеосигнал. К ним относятся, например, организация видеоконференций, обучение на расстоянии и видео по запросу. Технология ADSL позволяет провайдерам предоставлять своим пользователям услуги, скорость передачи данных которых более чем в 100 раз превышает скорость самого быстрого на данный момент аналогового модема (56 Кбит/с) и более чем в 70 раз превышает скорость передачи данных в ISDN (128 Кбит/с).

Технология ADSL позволяет телекоммуникационным компаниям предоставлять частный защищенный канал для обеспечения обмена информацией между пользователем и провайдером.

Не следует забывать и о затратах. Технология ADSL эффективна с экономической точки зрения хотя бы потому, что не требует прокладки специальных кабелей, а использует уже существующие двухпроводные медные телефонные линии. Т. е., если дома или в офисе есть подключенный телефонный аппарат, не нужно прокладывать дополнительные провода для использования ADSL (хотя компания, обеспечивающая вам возможность обычной телефонной связи, должна при этом предоставлять и услугу ADSL).

Для того, чтобы линия ADSL работала, необходимо не так уж много оборудования. На обоих концах линии устанавливаются модемы ADSL: один на стороне пользователя (дома или в офисе), а другой на стороне сети (у провайдера Интернет или на телефонной станции). Причем пользователю совсем не обязательно покупать свой модем, но достаточно взять его у провайдера в аренду. Кроме того, пользователю для того, чтобы модем ADSL работал, необходимо иметь компьютер и интерфейсную плату, например, Ethernet 10baseT.

По мере того, как телефонные компании постепенно вступают на еще неосвоенное поле передачи данных форматов видео и мультимедиа конечному пользователю, технология ADSL продолжает играть большую роль. Разумеется, через какое-то время широкополосная кабельная сеть охватит всех потенциальных пользователей. Но успех

этих новых систем будет зависеть от того, какое количество пользователей будет вовлечено в процесс использования новых технологий уже сейчас. Принося кинофильмы и телевидение, видеокаталоги и Интернет в дома и офисы, ADSL делает данный рынок жизнеспособным и прибыльным как для телефонных компаний, так и для других компаний, предоставляющих услуги в различных областях.

### **13.3. Высокоскоростные радиосистемы передачи данных**

#### **Общие сведения**

В последние годы тезис о том, что информационные технологии оказывают самое прямое влияние на состояние и развитие экономики, стал практически общепризнанным. Компьютерный мир еще несколько лет назад стал сетевым. Сетевая инфраструктура дает возможность оперативного обмена данными и доступа к информационным ресурсам, как на локальном уровне, так и в мировом масштабе. Во многих случаях использование проводных или оптоволоконных линий связи невозможно или экономически нецелесообразно. В этой ситуации одним из наиболее эффективных решений проблемы связи, а зачастую и единственно возможным, является использование радиосетей передачи данных.

*К отличительным свойствам беспроводных технологий передачи данных можно отнести:*

- мобильность. Невозможность подсоединения подвижных абонентов является принципиально непреодолимым ограничением кабельных сетей. Медсестры, врачи, рабочие на конвейере, маклеры на бирже и складские рабочие постоянно перемещаются с места на место. Для них беспроводная технология представляет несковывающий их перемещений канал в проводную сеть, открывая доступ ко всей имеющейся в этой сети информации;

- возможность организации сети там, где прокладка кабеля технически невозможна. Например, в зданиях, являющихся памятниками архитектуры;

- возможность объединить в сеть удаленных абонентов. Если абоненты разбросаны по обширной малонаселенной (или труднодоступной) территории, то во многих случаях протягивать кабель оказывается экономически нецелесообразно. Радиосети связывают населенные пункты, до которых просто не доходят телефонные линии. Если

все же доходят, то телефонные станции не торопятся предоставлять линии связи в аренду, да и качество связи низкое. Но главное даже в другом – пропускная способность телефонных каналов не оставляет никаких надежд на организацию эффективного обмена данными;

– срочность. Надежные коммуникации нужны сейчас, немедленно, а для прокладки кабельной сети требуются колоссальные инвестиции и длительное время. Радиооборудование позволяет развернуть сеть всего за несколько часов. Радиооборудование может также использоваться для организации временных сетей. Например, выставки, избирательная компания и т. д.

Рассмотрим радиооборудование, которое может быть использовано для создания радиосетей передачи данных, и задачи, которые позволяет решать тот или иной класс оборудования. Радиооборудование можно классифицировать по используемой частоте. От того, в каком диапазоне работает оборудование зависят такие показатели, как дальность связи, скорость передачи информации, зависимость от погодных условий, требование к обеспечению «прямой видимости».

1.6 – 30 МГц (коротковолновый диапазон). Системы работающие в диапазоне позволяют передавать данные и голосовые сообщения на расстояния до нескольких тысяч километров, что предоставляет уникальную возможность охвата значительных территорий, в том числе с гористым рельефом, что абсолютно невозможно для традиционных решений в диапазонах УКВ и СВЧ при соизмеримом вложении средств. Скорость передачи в КВ-системах относительно невысокая до 6 Кбит/с.

136 – 174 МГц – скорость передачи данных до 19.2 Кбит/с, дальность связи до 70 км, связь может осуществляться «из-за» угла и за горизонтом за счет искривления пути прохождения радиолуча у земли. Радиомодемы, работающие в этом диапазоне, используются для передачи файлов и электронной почты, позволяют организовать мобильный доступ в базы данных. Применяются в территориально распределенных сетях, в системах телеметрии и телеуправления, могут быть очень полезны для таких организаций, как ГАИ, служба скорой медицинской помощи и т.п. Интегральные радиомодемы, работающие в этом диапазоне частот, выпускаются такими фирмами, как Pacific Crest, Maxon, Young Design и др.

400 – 512 МГц – скорость передачи данных до 128 Кбит/с, дальность связи до 50 км. Желательно наличие прямой видимости, но возможна работа и на отраженных сигналах. В этом диапазоне могут ра-

ботать узкополосные синхронные радиомодемы RAN производства фирмы Wireless, Inc (ранее Multipoint Networks) (9.6, 19.2, 64, 128 Кбит/с).

Радиомодемы RAN 64/25, 128/50 используют модуляцию 16 QAM, что позволяет передавать данные со скоростью 64 Кбит/с в полосе 25 кГц или 128 Кбит/с в полосе 50 кГц. Радиомодемы данного типа применяются для построения высокоскоростных каналов точка-точка для мультиплексированной передачи данных, голоса, видео-изображений и другой информации. На их основе также возможна организация многоузловых территориально распределенных сетей. Радиомодемы RAN могут работать также и в диапазоне 820 – 960 МГц.

Выше 2 ГГц — возможна организация каналов передачи данных со скоростью более 2 Мбит/с, при этом обязательным является условие прямой видимости между антеннами. На этом участке радиочастотного спектра работает оборудование Radio-Ethernet (стандарт IEEE 802.11). Стандарт Radio-Ethernet имеет два основных применения. Первое из них — беспроводная локальная сеть в стенах одного здания или на территории предприятия, таким образом решается проблема «ограниченной мобильности» в пределах одного предприятия (сотрудник с портативным компьютером, переходящий из одной комнаты в другую отовсюду имеет доступ к сети). Второе применение стандарта Radio-Ethernet решает проблему подсоединения абонентов к большой сети передачи данных или, как говорят связисты, проблему последней мили.

В Radio-Ethernet может применяться технология шумоподобных сигналов или широкополосных сигналов (ШПС). Узкополосные устройства излучают в эфир сигнал с шириной спектра 12,5 – 200 кГц, причем ширина излучаемого спектра увеличивается с увеличением скорости передачи информации. Узкополосные системы обладают существенным недостатком: если в частотном диапазоне такой системы появляются помехи, то качество связи резко падает. Именно эта незащищенность от помех узкополосных систем привела к разработке, сначала для военных целей, ШПС- технологии.

Системы на основе шумоподобных сигналов обладают следующими преимуществами:

- помехозащищенность;
- не создаются помехи другим устройствам (низкая мощность сигнала);
- конфиденциальность передач;

- низкая стоимость при массовом производстве (низкая мощность сигнала – дешевые высокочастотные компоненты оборудования);
- шумоподобный сигнал обеспечивает возможность работы в диапазоне, уже занятыми другими системами радиопередач;
- высокая скорость передачи.

Идея технологии широкополосного сигнала состоит в том, что для передачи информации используется значительно более широкая полоса частот, чем это требуется при передаче в узкополосном канале. Стандарт 802.11 для получения шумоподобных сигналов предусматривает метод прямой последовательности (Direct Sequence Spread Spectrum-DSSS) и метод частотных скачков (Frequency Hopping Spread Spectrum-FHSS).

В методе со скачками по частоте (FHSS) весь диапазон от 2400 МГц до 2483,5 МГц разбит на 79 подканалов. Приемник и передатчик синхронно каждые несколько миллисекунд перестраиваются на различные несущие частоты в соответствии с алгоритмом, задаваемым псевдослучайной последовательностью. Лишь приемник, использующий ту же самую последовательность, может принимать сообщение. При этом предполагается, что другие системы, работающие в том же частотном диапазоне, используют иную последовательность и поэтому практически не мешают друг другу. Для тех случаев, когда два передатчика пытаются использовать ту же самую частоту одновременно, предусмотрен протокол разрешения столкновений, по которому передатчик делает попытку повторно послать данные на следующей в последовательности частоте.

Согласно методу с прямой последовательностью (DSSS) диапазон от 2400 МГц до 2483,5 МГц разбит на три широких подканала, которые могут использоваться независимо и одновременно на одной территории. Принцип работы DSSS систем состоит в следующем: в передаваемый радиосигнал вносится значительная избыточность путем передачи каждого бита информации одновременно в нескольких частотных каналах. Если на каком-либо из них (или сразу на нескольких) появляются помехи, система определяет правильный поток данных путем выбора наибольшего количества одинаковых потоков.

Наиболее крупными производителями оборудования Radio-Ethernet являются Proxim, BreezeCom, Aironet, Cylink, Lucent Technologies, Solectek, WaveAccess.

На одном и том же пространстве могут сосуществовать, не мешая друг другу, не более трех сетей DSSS. При попытке увеличить

число пользователей, такое неэкономное использование эфира может оборачиваться проблемами. FHSS позволяет определить для каждой сети свой набор и последовательность дискретных частот. Еще одна существенная особенность технологии «прыгающей частоты» состоит в том, что весь широкополосный диапазон разбивается на 79 отдельных подканалов. FHSS-оборудование (например, компании BreezeCom) позволяет использовать не все 79 каналов, а любое количество частот из этого набора, вплоть до одной частоты. В системах DSSS использование широкой полосы принципиально необходимо.

ШПС-технология, кроме оборудования Radio-Ethernet, применяется в высокоскоростных синхронных радиомодемах диапазонов 2,4 и 5,7 ГГц. Эти радиомодемы используются для организации дуплексных магистральных синхронных радиоканалов передачи данных со скоростями до 2 048 Кбит/с. Оборудование этого класса производят такие компании, как Wireless Inc (модели RAN64ss, RAN128ss, RAN2048ss), BreezeCom (серия BreezeLINK), Wave Wireless (SpeedCOM).

ШПС-технология используется еще в одном интересном и весьма полезном продукте фирмы Wireless, Inc – радиомаршрутизаторе WaveNet IP. В отличие от радио-Ethernet устройств это оборудование включает в свой состав маршрутизатор IP и специально предназначено для организации радиосетей городского и районного масштаба на расстоянии до 30–40 км от центральной станции. Кроме того, конструктивное исполнение WaveNet IP позволяет решить так называемую проблему длинного кабеля. Проблема заключается в том, что достаточно часто точка подключения к локальной сети и точка установки антенны на крыше находятся на достаточно большом расстоянии друг от друга. Оборудование Radio-Ethernet обычно имеет исполнение для использования внутри помещений и может применяться только в нормальных климатических условиях. Поскольку высокочастотный радиосигнал испытывает значительное затухание в кабеле, это накладывает серьезные ограничения на максимальную длину кабеля между устройством и антенной. WaveNet IP имеет внешнее погодозащитное исполнение и устанавливается в непосредственной близости от антенны, что позволяет без потерь сигнала размещать высокочастотный блок на расстоянии до 100 м от физической точки входа в сеть.



## 13.4. IP-телефония.

### Принципы IP-телефонии

IP-телефония – это технология, позволяющая использовать Интернет или любую другую IP-сеть в качестве средства организации ведения телефонных переговоров, передачи факсов, изображения и т. д. и т. п.

Вообще IP – это просто Internet Phone (Интернет-телефон).. Принцип IP-телефонии объяснить проще, чем теорему Пифагора: вы что-то говорите в трубку своему старому другу из Майами, ваш голос оцифровывается и пересылается через Интернет на другой конец света отдельными пакетами в режиме реального времени. При этом максимальная задержка звука составляет 300–400 миллисекунд в зависимости от того, сколько времени требуется аппаратному оборудованию, чтобы создать цифровой аудиосигнал. Человеческое ухо не воспринимает задержки менее 250 миллисекунд, но в настоящее время существуют технологии, позволяющие свести потери сигнала в сети к минимуму и избежать пропадания голоса. Плюс к этому вы заплатите за IP-разговор в 5–10 раз меньше, чем по обычной телефонной линии.

Из этого очевидные преимущества:

- возможность существенного снижения затрат на междугородные и международные телефонные переговоры;
- возможность передачи голосового трафика от головных офисов в филиалы в единой информационной IP магистральной практически без затрат независимо от расстояния.

Теперь вкратце о том, как это реализуется. IP-телефония основывается на двух базовых операциях: преобразовании двунаправленной аналоговой речи в цифровую форму внутри кодирующего/декодирующего устройства (кодека) и упаковке данных в пакеты для передачи по IP сети. Эти базовые функции IP-телефонии могут быть реализованы в широком спектре оборудования – специальных программ в обычном компьютере и настольных телефонов до высокоскоростных шлюзов операторов связи. Шлюзы IP-телефонии обрабатывают трафик, поступающий от других телефонных устройств и шлюзов. Они способны обслуживать как одно, так 500 и более телефонных устройств и могут быть установлены как у конечного пользователя, так и у сервис-провайдера. Другой тип автономных устройств представляют пограничные устройства, в которых шлюз объединен с удаленным доступом и пулом модемов. Кроме того, функции шлюза се-

годня доступны на уровне Ethernet-оборудования, установленного у конечных пользователей и поддерживающих H.323 или SIP — стандарт ITU на передачу мультимедийного трафика по IP. Таким образом, одноранговые бесшлюзовые системы IP-телефонии теоретически возможно создать посредством объединения частных локальных и глобальных сетей. Проблема на настоящий момент состоит в том, что лишь немногие производители поставляют телефоны Ethernet H.323/SIP, поэтому заинтересованным в таком подходе заказчикам приходится искать нестандартные решения.

В общих чертах передача голоса в IP-сети происходит следующим образом. Входящий звонок и сигнальная информация из телефонной сети передаются на пограничное сетевое устройство, называемое телефонным шлюзом, и обрабатываются специальной картой устройства голосового обслуживания. Шлюз, используя управляющие протоколы семейства H.323 или SIP, перенаправляет сигнальную информацию другому шлюзу, находящемуся на приемной стороне IP-сети. Приемный шлюз обеспечивает передачу сигнальной информации на приемное телефонное оборудование согласно плану номеров, гарантируя сквозное соединение. После установления соединения голос на входном сетевом устройстве оцифровывается (если он не был цифровым), кодируется в соответствии со стандартными алгоритмами ITU, такими как G.711 или G.729, сжимается, инкапсулируется в пакеты и отправляется по назначению на удаленное устройство с использованием стека протоколов TCP/IP.

Приходящие на приемный шлюз IP-пакеты преобразуются обратно в телефонный сигнал, и принимающий абонент получает вызов. Конечные потребители услуги могут даже не догадываться о том, как осуществляется этот звонок. Поскольку пакеты могут доставляться не в той последовательности, в которой были отправлены, то в первую очередь происходит их накопление и восстановление требуемой последовательности. Для восстановления исходного объема упорядоченных данных используются порядковые номера пакетов. Для приложений, где не важен порядок и интервал прихода пакетов, таких как e-mail, время задержек между отдельными пакетами не имеет решающего значения. IP-телефония является одной из областей передачи данных, где важна динамика передачи сигнала.

С целью минимизации потери сигнала, ведущие организации по стандартизации разрабатывают новые протоколы, а производители поставляют качественно новые, современные решения в области IP

телефонии, позволяющие избежать пропадания голоса. Однако звонок Телефон-Телефон является самым очевидным, но далеко не единственным сервисом, который может предоставлять оператор IP-телефонии. Используя IP-сеть, можно обмениваться цифровой информацией для пересылки голосовых или факсимильных сообщений между двумя компьютерами в режиме реального времени. Применение Internet позволит реализовать данную службу в глобальном масштабе. Для IP-телефонии чаще всего используются стандарты H.323 и SIP, определяющие передачу видео и аудио по сетям с негарантированным качеством услуг, таким как Ethernet и IP. H.323 описывают несколько элементов, в том числе аудио- и видеокодеки (кодееры/декодеры), коммуникационные протоколы и синхронизацию пакетов.

### **Виды IP-телефонии и преимущества**

*Голосовая связь через IP-сеть* может осуществляться различными способами:

*Компьютер и компьютер* – самый ранний и простой способ IP-телефонии. Для его реализации необходимы два компьютера со звуковыми картами, системами ввода-вывода звуковой информации (колонки, микрофон).

*Компьютер и телефон.* При этом абонентам необходимо иметь, с одной стороны, компьютер со звуковой картой и системой ввода-вывода звука, специальный IP-фон или телефонный IP-шлюз, а с другой, обычный телефон, находящийся в любой точке планеты.

*Телефон и телефон.* При этом абоненту не нужно никакого специального оборудования, только обычный телефон. Качество связи сопоставимо с качеством при обычном телефонном разговоре, т. к. компанией используются выделенные каналы связи

*«WEB – телефон».* Еще одна новая услуга, которую предоставляют провайдеры IP-телефонии – это звонок с Вэб-сайта или Surf&Call – решение в области веб-телефонии для коммерческих сайтов или службы поддержки, позволяющее осуществлять вызов, выбрав со страницы Интернет ссылку на имя вызываемого абонента. Это решение направлено, прежде всего, на расширение возможностей электронной коммерции. Surf&Call позволяет пользователям Интернет напрямую, поговорить, например, с торговым представителем либо со специалистом технической поддержки интересующей его фирмы. Установление телефонного соединения происходит при нажатии курсором на ссылку, представляющую собой, например, название компании, имя

вызываемого абонента и т. д. на странице Интернет. При этом пользователю не требуется вторая телефонная линия или прерывание работы в Интернет, необходимо лишь загрузить небольшое клиентское программное обеспечение, которое обычно можно найти на той же WEB-странице, и которое устанавливается автоматически. С другой стороны Surf&Call позволяет представителям компаний отвечать на вопросы, демонстрировать WEB-страницы, передавать необходимую информацию, улучшая тем самым качество предоставляемых услуг.

Какие же преимущества дает нам использование IP Телефонии?

Использование IP-телефонии позволяет достигнуть значительной экономии на услугах международной и междугородной связи из-за низких тарифов. При этом качество связи будет отличное.

Не потребуется выход на международную АТС.

Гарантированная конфиденциальность, т. к. весь сеанс разговора будет кодирован в пакеты и передача пакетов будет производится независимо друг от друга, то такую передачу просто невозможно перехватить.

Используя IP-телефонию, вы будете защищены от телефонных «пиратов», которые могут подключиться к вашему телефону и звонить за ваш счет без вашего ведома.

Вы можете звонить из дома, офиса от соседей и не заботиться о счетах за разговоры и их оплате.

Тарификация ведется посекундно, первые десять секунд бесплатны, чтобы Вы могли убедиться дома ли собеседник.

Полный контроль над расходованием средств на междугородные и международные переговоры.

В любой момент возможно узнать остаток средств на счете и полную детализацию звонков в своей статистике.

*Решение проблемы занятой линии.* Уже давно любители бороздить всемирную сеть сталкиваются с проблемой занятости телефонных линий во время сеанса связи. IP-телефония позволяет решить эту проблему. Единственное, что должен сделать абонент – это заказать на своей АТС переадресацию по сигналу занято на телефонный номер сервера IP-телефонии. При звонке на номер абонента во время Интернет-сессии вызов переадресуется на сервер IP-телефонии, который преобразовывает его в IP-пакеты и отправляет на компьютер абонента. На компьютере абонента появляется иконка Входящий звонок, кликнув на которую он может поговорить со звонящим.

*Интеграция филиалов* в единую информационную структуру. В последнее время с развитием информационных технологий и увеличением пропускной способности каналов, все для наиболее оперативного решения деловых задач, филиалы компании объединяются в одно целое, образуя интрасеть. Так как предлагаемая технология использует для передачи голоса как раз сети передачи данных, то появляется возможность объединять не только компьютерные сети, но и телефонные.

*Виртуальные частные сети (VPN)*. IP-телефония является идеальной технологией для построения виртуальных частных сетей предприятия. Главная черта технологии VPN – использование IP-сети в качестве магистрали для передачи корпоративного IP-трафика. Сети VPN решают задачи подключения корпоративного пользователя к удаленной сети и соединения нескольких удаленных ЛВС и АТС в единую корпоративную сеть передачи голоса и данных.

*Глобальный роуминг*. IP-телефония позволяет операторам связи очень просто и с минимальными затратами организовать роуминг услуг связи.

*Совмещенный доступ в Интернет*. Голосовые данные, факсимильные сообщения передаются с использованием IP – основного набора протоколов Интернет. Данное решение само собой подразумевает доступ к ресурсам Сети и очевидную экономию на аренде линий связи и оплату услуг.

### **Стандарты IP-телефонии**

Стандарты являются критическим фактором для мира IP-телефонии. Одна из наиболее важных областей стандартизации - протокол обмена сообщениями в IP-телефонии.

Ранние решения IP-телефонии использовали для связи друг с другом закрытые протоколы. Оба участника беседы должны были иметь аналогичные продукты. Intel и Microsoft возглавили направление на разработку стандартов на основе H.323, рекомендованного International Telecommunications Union (ITU). Этот стандарт формулирует технические требования для передачи аудио- и видеоданных по сетям передачи данных. H.323 включает в себя стандарты на видеокодер-декодеры; голосовые кодер-декодеры; общедоступные приложения; управление вызовами; управление системой.

При передаче в режиме реального времени до 30% пакетов могут потеряться или опоздать (что в режиме реального времени одно и

то же). Хорошее приложение IP-телефонии должно возместить нехватку пакетов, восстановив потерянные данные. Сам алгоритм кодировки также оказывает влияние на восстановление данных. Сложные алгоритмы увеличивают стоимость необходимого оборудования.

Еще одна особенность состоит в том, что системы IP-телефонии должны иметь возможность поддерживать разные кодеры и добавлять новые по необходимости.

Один из современных протоколов для IP-телефонии - это протокол SIP.

Он родился в недрах IETF (Internet Engineering Task Force) – организации, занимающейся утверждением стандартов интернета и, вообще, стандартов, имеющих отношение к протоколам TCP/IP. Здесь за основу были взяты протоколы, применяемые в самых популярных на сегодняшний день IP-сервисах, такие, как HTTP (Web) и SMTP (электронная почта). Идейно SIP основан на том же подходе, что HTTP: запрос – ответ (request – reply).

Существенно, что SIP, хотя и может использоваться в IP-телефонии, не является протоколом для передачи голосовых данных – он вообще не привязан к передаче данных какого-то определенного вида. Название SIP расшифровывается как Session Initiation Protocol – протокол инициирования сеанса. Это означает, что SIP обеспечивает инициирование, контроль и ликвидацию сеансов обмена информацией, а в качестве самой передаваемой информации может выступать что угодно: и речь (как в случае IP-телефонии), и музыка, и видео, и, например, текст (протокол позволяет организовывать сеансы коллективной работы над документами, какие поддерживаются в MS Exchange или Lotus Notes).

Тип данных определяется отдельным протоколом SDP (Session Description Protocol – протокол описания сеанса), который работает в паре с SIP и обладает способностью менять параметры сеанса по ходу обмена данными. Простейший пример: два собеседника разговаривают по IP-телефону, и один хочет показать другому фотографию – SDP позволит сделать это в рамках того же SIP-сеанса. Более того, ничто не мешает перейти по ходу дела на другой терминал (например, если разговор происходит по мобильному телефону, вы можете перейти на компьютер или на телефон с дисплеем и там просмотреть фотографию).

### **Коммуникации реального времени**

Очевидно, SIP существенно лучше, чем H.323, согласуется с пониманием IP-телефонии как массового глобального IP-сервиса, а ее несомненно следует рассматривать именно так. И поскольку SIP способен обслуживать не только телефонию, но и вообще любые коммуникации в реальном времени поверх протокола IP, ему, по-видимому, предстоит стать основой также для ряда других IP-сервисов, которые пока не существуют, но появятся в будущем.

### **Качество IP-телефонии**

Телефонная сеть была создана таким образом, чтобы гарантировать высокое качество услуги даже при больших нагрузках. IP-телефония, напротив, не гарантирует качества, причем при больших нагрузках оно значительно падает.

Стоимость любой услуги, как правило, и её качество прямо пропорциональны. Однако понятно и то, что в ряде случаев представляется неразумным сравнительно небольшое увеличение качества оплачивать относительно большим увеличением цены. Качество связи можно оценить следующими основными характеристиками:

- уровень искажения голоса;
- частота «пропадания» голосовых пакетов;
- время задержки (между произнесением фразы первым абонентом и моментом, когда она будет услышана вторым абонентом).

По первым двум характеристикам качество связи значительно улучшилось в сравнении с первыми версиями решений IP-телефонии, которые допускали искажение и прерывание речи. Улучшение кодирования голоса и восстановление потерянных пакетов позволило достичь уровня, когда речь понимается абонентами достаточно легко. Понятно, что задержки влияют на темп беседы. Известно, что для человека задержка до 250 миллисекунд практически незаметна. Существующие на сегодняшний день решения IP-телефонии превышают этот предел, так что разговор похож на связь по обычной телефонной сети через спутник, которую обычно оценивают как связь вполне удовлетворительного качества, требующую лишь некоторого привыкания, после которого задержки для пользователя становятся неощутимы. Отметим, что даже в таком виде связи решения IP-телефонии вполне подходят для многих приложений.

Задержки можно уменьшить благодаря следующим трем факторам:

во-первых, совершенствованием телефонных серверов (их работчики борются с задержками, улучшая алгоритмы работы).

во-вторых, развитием частных (корпоративных) сетей (их владельцы могут контролировать ширину полосы пропускания и, следовательно, величины задержки).

в-третьих, развитием самой сети Интернет. Современный Интернет не был рассчитан на коммуникации в режиме реального времени.

Оценить качество при использовании различных протоколов сжатия можно различными способами. Один из подходов для таких измерений - использование субъективных методов. В субъективных методах группа людей, обычно достаточно большая, оценивает качество связи по определенной стандартной процедуре. Самый известный субъективный метод - это метод общего мнения (Mean Opinion Score - MOS). В этом методе, качество связи оценивается большой группой разных людей, и затем их мнение усредняется.

Основным недостатком IP-телефонии на данном этапе развития технологии является более низкое качество связи, хотя это и не обязательно, т. к. качество связи напрямую зависит от качества (скорости) интернета. При этом причиной основного недостатка является основное преимущество – пакетная маршрутизация голосового трафика.

Представим, что слово разбито на пакеты, состоящие из одной буквы. В этом случае потеря одного пакета приведет к выпадению из слова одной буквы, потеря более одного пакета – к потере слога или слова целиком. Также немаловажна задержка в получении пакетов на удаленном сервере. Надо отметить, что повторная передача пакетов при наличии ошибок для IP-телефонии не приносит положительного эффекта.

Телефонная сеть была создана таким образом, чтобы гарантировать высокое качество услуги даже при больших нагрузках. IP-телефония, напротив, не гарантирует качества, причем при больших нагрузках оно значительно падает.

Конечно, ничей опыт не заменит собственного - лучше попробовать самому оценить качество этого вида связи, проведя разговор посредством IP-телефонии и сравнив его с традиционной телефонной связью.



## 14. ЛИТЕРАТУРА

1. Гук, М.Ю. Аппаратные средства IBM PC. Энциклопедия/ М. Ю. Гук.– 3-е изд. – СПб.: Питер, 2006. – 1072 с.
2. Мюллер, С. Модернизация и ремонт ПК: пер. с англ./ С. Мюллер.– 16-е изд.– М.: Издательский дом «Вильямс», 2006. – 1344 с.
3. Таненбаум, Э. Архитектура компьютера/ Э.таненбаум.– 5-е изд.– СПб.: Питер, 2007. — 844 с.

Учебное издание

**М. С. Шмаков**

# **ЭЛЕКТРОННЫЕ ВЫЧИСЛИТЕЛЬНЫЕ МАШИНЫ И ВЫЧИСЛИТЕЛЬНЫЕ СИСТЕМЫ**

**Курс лекций**

Редактор *М. А. Юрасова*  
Компьютерная верстка *М. С. Шмаков*

Подписано в печать 15.06.09. Формат 60×84<sup>1</sup>/<sub>16</sub>.  
Бумага офсетная. Гарнитура Таймс. Печать офсетная.  
Усл. печ. л. 7,4. Уч.-изд. л. 7,7.  
Тираж 200 экз. Заказ .

Учреждение образования  
«Белорусский государственный технологический университет».  
220006. Минск, Свердлова, 13а.  
ЛИ № 02330/0549423 от 08.04.2009.

Отпечатано в лаборатории полиграфии учреждения образования  
«Белорусский государственный технологический университет».  
220006. Минск, Свердлова, 13.  
ЛП № 02330/0150477 от 16.01.09.