

АППАРАТНЫЕ СРЕДСТВА ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ. КУРС ЛЕКЦИЙ

Учебное пособие

МИНИСТЕРСТВО ЦИФРОВОГО РАЗВИТИЯ,
СВЯЗИ И МАССОВЫХ КОММУНИКАЦИЙ РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ
БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ТЕЛЕКОММУНИКАЦИЙ
им. проф. М. А. БОНЧ-БРУЕВИЧА»
(СПбГУТ)

АППАРАТНЫЕ СРЕДСТВА ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ. КУРС ЛЕКЦИЙ

Учебное пособие

Электронное издание
локального распространения

Санкт-Петербург
Наукоемкие технологии
2026

© Авторский коллектив, 2026

© Санкт-Петербургский государственный университет
телекоммуникаций им. проф. М. А. Бонч-Бруевича, 2026
ISBN 978-5-00271-192-5

УДК 371.388:[654+004.3]

ББК 32.973-04я73

A76

Авторы:

Пачин А. В., Вивчарь Р. М., Резников Б. К., Исаенко Д. И., Киричек Р. В.

Рецензенты

И. В. Карташевский — доктор технических наук, доцент,
заведующий научно-исследовательской лабораторией «Инновационные проекты»
Поволжского государственного университета телекоммуникаций и информатики;

А. А. Прасолов — кандидат технических наук, доцент кафедры телевидения
и метрологии Санкт-Петербургского государственного университета телекоммуникаций
им. проф. М. А. Бонч-Бруевича

A76 Аппаратные средства вычислительной техники. Курс лекций [Электронный ресурс]: учебное пособие / А. В. Пачин, Р. М. Вивчарь, Б. К. Резников, Д. И. Исаенко, Р. В. Киричек; СПбГУТ. – Электрон, текстовые дан. (1,8 Мб). – СПб.: Научные технологии, 2026. – 81 с. – 1 электрон., опт. диск (CD-ROM).

ISBN 978-5-00271-192-5

Посвящено основам вычислительной техники и архитектуре персональных компьютеров. Охватывает широкий круг вопросов, начиная с истории создания и развития вычислительных систем, и заканчивая современными технологиями увеличения производительности процессоров и особенностями специализированных микропроцессоров. Подробно рассматриваются составные части персонального компьютера, включая системный блок, устройства хранения данных, элементы ввода-вывода, а также принципы их взаимодействия. Особое внимание уделено архитектуре и функционированию центрального процессора, включая различные подходы к выполнению команд (конвейеризация, мультитрединг), а также структуре и типам памяти. Материал изложен в логической последовательности и может быть использован как для изучения теоретических основ, так и для практического ознакомления с современными аппаратными средствами компьютерных систем.

Пособие предназначено для студентов технических специальностей, преподавателей, а также всех интересующихся устройством и принципами работы персональных компьютеров.

Утверждено редакционно-издательским советом СПбГУТ в качестве учебного пособия

© Авторский коллектив, 2026

© Санкт-Петербургский государственный университет

телекоммуникаций им. проф. М. А. Бонч-Бруевича, 2026

ISBN 978-5-00271-192-5

Учебное издание

Пачин Андрей Владимирович
Вивчарь Роман Михайлович
Резников Богдан Константинович
Исаенко Дмитрий Игоревич
Киричек Руслан Валентинович

Аппаратные средства вычислительной техники. Курс лекций

Учебное пособие

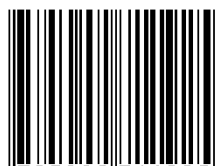
Электронное издание
локального распространения

Издается в авторской редакции
Верстка *Ю. Н. Сафонкина*

Издательство «Наукоемкие технологии»
ООО «Корпорация «Интел Групп»
<https://publishing.intelgr.com>
E-mail: publishing@intelgr.com
Тел.: +7 (812) 945-50-63

Подписано к использованию 26.10.2026 г.
Объем издания — 1,8 Мб.
Комплектация издания — 1 CD.
Тираж 100 CD.

ISBN 978-5-00271-192-5



9 785002 711925 >

ОГЛАВЛЕНИЕ

1. ИСТОРИЯ СОЗДАНИЯ И РАЗВИТИЯ ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ	5
2. ОСНОВНЫЕ ЭЛЕМЕНТЫ ПЕРСОНАЛЬНОГО КОМПЬЮТЕРА	9
Системный блок. Элементы, размещаемые на системной плате.....	9
Источник электропитания персонального компьютера.....	12
Внешние устройства хранения данных. Устройства ввода/вывода.....	13
3. ОБЕСПЕЧЕНИЕ ЭФФЕКТИВНОГО ФУНКЦИОНИРОВАНИЯ КОМПЬЮТЕРОВ	15
Показатели качества и эффективности функционирования компьютеров.....	15
Методы повышения надежности компьютеров.....	17
4. ЦЕНТРАЛЬНЫЙ ПРОЦЕССОР. АРХИТЕКТУРА ПРОЦЕССОРА	22
Классификация и основные характеристики микропроцессоров	22
Особенности процессоров CISC, RISC и VLIW	26
5. ЦЕНТРАЛЬНЫЙ ПРОЦЕССОР. ПРИНЦИПЫ ПОСТРОЕНИЯ И ФУНКЦИОНИРОВАНИЯ МИКРОПРОЦЕССОРОВ.....	33
Структура микропроцессоров общего назначения	33
Цикл тракта данных микропроцессора	36
6. УВЕЛИЧЕНИЕ БЫСТРОДЕЙСТВИЯ ПРОЦЕССОРОВ. ТЕХНОЛОГИИ ВЫПОЛНЕНИЯ КОМАНД В МИКРОПРОЦЕССОРЕ	38
Конвейеризация выполнения команд	38
Динамическое и мультитредовое выполнение команд.....	41
7. СПЕЦИАЛИЗИРОВАННЫЕ МИКРОПРОЦЕССОРЫ. ОСОБЕННОСТИ АРХИТЕКТУРЫ И СТРУКТУРЫ СПЕЦИАЛИЗИРОВАННЫХ МИКРОПРОЦЕССОРОВ.....	44
Процессоры цифровой обработки сигналов и сетевые процессоры	44
Графические процессоры	47
8. УСТРОЙСТВА ХРАНЕНИЯ ДАННЫХ В КОМПЬЮТЕРНЫХ СИСТЕМАХ	51
Пятиуровневая схема организации памяти компьютерных систем	51
Оперативная и кэш-память компьютеров	55
Энергонезависимые накопители данных	58
9. УСТРОЙСТВА ВВОДА И ВЫВОДА ДАННЫХ КОМПЬЮТЕРНЫХ СИСТЕМ ...	63
Шины и разъемы для подключения I/O устройств	63
Устройства ввода данных компьютерных систем	65
Устройства вывода данных компьютерных систем.....	68
Twisted Nematic (TN) технология.....	69
In-Phase Switching (IPS) технология. Дальнейшее развитие идеи TN ...	69
10. ЦЕНТРЫ ОБРАБОТКИ ДАННЫХ. ВИРТУАЛИЗАЦИЯ ВЫЧИСЛИТЕЛЬНЫХ РЕСУРСОВ	73
Список рекомендуемой литературы	81

1. ИСТОРИЯ СОЗДАНИЯ И РАЗВИТИЯ ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ

Деятельность человека по изучению и изменению окружающего мира, формированию своей среды, с развитием цивилизации очень быстро столкнулась с необходимостью выполнения математических вычислений. Наиболее активными драйверами развития математики были экономика и военное дело.

Первые устройства, при помощи которых выполнялись расчеты, были самые элементарны: палочки, камешки, бусы.

Следующим этапом развития таких устройств является применение счетных досок. Известны свидетельства применения их в древней Индии, Китае, Греции и Египте.

Последним из разработанных устройств для ручного счета является логарифмическая линейка, практически до конца прошлого столетия служившая инженерам и ученым.

Однако с наступлением промышленной революции сложность и объемы вычислений возрастали, и многие изобретатели задумывались на автоматизацией процесса вычислений. Развитие механики позволило создать ряд механических вычислителей. Попытку создать такое устройство предпринимал Леонардо да Винчи.

Первая механическая машина была описана в 1623 г. профессором математики Тюбингенского университета Вильгельмом Шиккардом, реализована в единственном экземпляре и предназначалась для выполнения четырех арифметических операций над 6-разрядными числами.

Действующая модель счетной суммирующей машины была создана в 1642 г. знаменитым французским ученым Блезом Паскалем. Для выполнения арифметических операций Паскаль заменил поступательное перемещение костяшек в абаконидных инструментах на вращательное движение оси (колеса) так, что в его машине сложению чисел соответствовало сложение пропорциональных им углов.

В начале XIX века Жозеф Мари Жаккар создал ткацкий станок, управляемый с помощью перфокарт, которые на долгие годы займут место устройств хранения данных.

В 1822 г. Чарльз Бэббидж приступил к осуществлению проекта так называемой разностной машины, предназначенной для расчета навигационных и астрономических таблиц. Машину эту строили десять лет, но так и не закончили. Финансовые трудности усугублялись тем, что

изобретатель постоянно пересматривал конструкцию и вносил в нее бесчисленные усовершенствования. Позднее Беббидж пытался создать универсальную счетную машину, предложив концепцию разделения блоков, предназначенных для хранения данных, выполнения вычислений и управлений устройством. Все эти составляющие характерны для современных процессорных систем.

Новые технологии со временем позволили совершенствовать вычислители. Появление электричества и изобретение реле позволили строить электромеханические устройства.

Первый такой комплекс был создан в США Германом Холлеритом в 1887 г. и состоял из ручного перфоратора, сортировочной машины и табулятора. Он предназначался для обработки результатов переписи населения. Управление механическими счетчиками и сортировкой осуществлялось электрическими импульсами, возникающими при замыкании электрической цепи при наличии отверстия в перфокарте. Импульсы использовались и для ввода чисел, и для управления работой машины. Поэтому машина Германа Холлерита была признана первой электромеханической счетной машиной с программным управлением.

Z3 — первая работоспособная полнофункциональная программно управляемая и свободно программируемая вычислительная машина, обладающая возможностью вычислений в двоичном коде с плавающей точкой и всеми свойствами современного компьютера. Создана немецким инженером Конрадом Цузе и представлена вниманию научной общественности 12 мая 1941 года.

Компьютер Марк 1 содержал около 765 тыс. деталей (электромеханических реле, переключателей и т. п.), достигал в длину почти 17 м (машина занимала в Гарвардском университете площадь в несколько десятков квадратных метров), в высоту — более 2,5 м и весил около 4,5 т. Общая протяжённость соединительных проводов составляла почти 800 км. Основные вычислительные модули синхронизировались механически при помощи 15-метрового вала, приводившегося в движение электрическим двигателем мощностью в 5 л. с. (4 кВт).

Компьютер оперировал 72 числами, состоящими из 23 десятичных разрядов, делая по 3 операции сложения или вычитания в секунду. Умножение выполнялось в течение 6 с, деление — 15,3 с, на операции вычисления логарифмов и выполнение тригонометрических функций требовалось больше минуты.

Создание двух машин, использующих новую базу — электронные лампы, определило начало новой эпохи вычислителей — эпохи компьютеров.

Представителем первого поколения является Electronic Numerical Integrator and Computer (Электронный числовой интегратор и вычислитель) или ENIAC. ENIAC создавался по заказу армии США для расчета баллистических таблиц. Изначально, подобные расчеты производились людьми и их скорость не могла соотноситься с масштабом военных действий. Построен компьютер был лишь к осени 1945 г.

Colossus в отличие от ENIAC был очень узконаправленной машиной. Он создавался исключительно с одной целью — декодирование немецких сообщений, зашифрованных с помощью Lorenz SZ. Эта машина была схожа с немецкой Enigma, но состояла из большего числа роторов. Для декодирования этих сообщений было решено создать Colossus. Он включал в себя 1500 электронных ламп, потреблял 8,5 КВт и обладал тактовой частотой в 5,8 МГц. Такое значение частоты достигалось за счет того, что Colossus был создан для решения только одной задачи и применяться в других областях не мог. К концу войны на вооружении Британии стояло 10 таких машин.

В послевоенное время активная разработка ЭВМ началась и в нашей стране. Представителями первого поколения являются машины серий МЭСМ, Минск, Урал.

В середине 50-х гг. 20 века, когда машины первого поколения по своим характеристикам в принципе достигли <<насыщения>>, ряд фирм объявил о ведущихся работах по созданию транзисторных ЭВМ. Некоторые авторитеты в области вычислительной техники, считали, что переход к машинам второго поколения будет медленным процессом из-за трудностей производства транзисторной элементной базы и из-за совершенствования характеристик электронных ламп. Однако этого не случилось. Преимущества транзистора как схемного элемента машин стали очевидными по мере совершенствования полупроводниковой технологии, и процесс его проникновения в ЭВМ стал необратимым и быстрым. и дальше скорость изменения технологий только возрастала!

Переход к третьему поколению произошел уже в конце 60-х гг. XX в. Компьютеры создавались на интегральных схемах, содержащих сотни, а потом и тысячи полупроводниковых элементов.

Первой машиной этого поколения считается компьютер IBM 360. В СССР в 70-х гг. XX в выпускались машины серии ЕС ЭВМ. Третьему

поколению машин соответствовали значительные изменения. Так, появилась возможность мультипрограммного режима. Данные стали храниться на магнитных лентах и дисках, заменивших перфокарты. Распространение получают дисплеи и принтеры.

В 1971 г. компания Intel представила свою новую разработку микропроцессор I 4004. Так машины перешли в четвертое поколение.

В 1981 г. IBM выбрала Intel в качестве производителя микропроцессора для своей новой машины IBM-PC. Это был микропроцессор Intel 8086.

Появился новый класс компьютеров — персональные компьютеры. Эти машины используют многозадачные операционные системы с графическим интерфейсом, видеокарты с поддержкой 3D графики, магнитные накопители, оптические и твердотельные. Имеют интерфейсы для быстрых проводных и беспроводных сетей. Области применения ПЭВМ перечислить невозможно. Они буквально везде. В последнее время мы являемся свидетелями взрывного развития таких подклассов компьютеров как смартфоны и ноутбуки. Они характеризуются еще большими объемами памяти, возрастающим быстродействием и широким спектром интерфейсов. Вместе с тем, помимо ПЭВМ огромную роль в современном мире играют и сверхмощные вычислительные комплексы — суперкомпьютеры, и компьютеры промышленного назначения — серверы, решающие большой спектр задач по всему земному шару и даже в космическом пространстве. Скорости эволюции вычислительных машин не снижается.

В заключение лекции необходимо сказать о перспективе 5 поколения компьютеров. В настоящее время нет четкого понимания, каким оно будет. Возможно к переходу в него приведет развитие технологий квантовых вычислений, сверхпроводников или искусственного интеллекта. Увидим в ближайшее время.

2. ОСНОВНЫЕ ЭЛЕМЕНТЫ ПЕРСОНАЛЬНОГО КОМПЬЮТЕРА

В данном разделе будут рассмотрены устройство и принцип функционирования персонального компьютера. Предметом исследования являются основные элементы, составляющие его архитектуру.

Изучение начнется с рассмотрения системного блока как базовой структурной единицы, вмещающей центральные компоненты. Будут проанализированы элементы, размещаемые на системной плате, такие как центральный процессор, оперативная память и графический адаптер, а также их взаимосвязь.

Далее будет рассмотрен источник электропитания как ключевой компонент, обеспечивающий преобразование и распределение электрической энергии, необходимой для работы всех систем.

Завершающим этапом станет анализ внешних устройств, классифицируемых по функциям: устройства хранения данных, предназначенные для постоянного и временного хранения информации, а также устройства ввода/вывода, обеспечивающие интерфейс между пользователем и системой.

Системный блок. Элементы, размещаемые на системной плате

Основным инструментом, позволяющим IT-специалисту реализовать свои профессиональные навыки, несомненно является компьютер. В подавляющем большинстве профессий в настоящее время достижение успеха без применения вычислительных машин просто немыслимо. Это обуславливает существенную актуальность темы занятия, необходимость изучения принципов построения и функционирования этого устройства.

Современная промышленность (а отрасли, связанные с разработкой компьютеров, уже десятилетиями имеют самые высокие темпы развития) производит несколько разновидностей компьютеров, начиная от крошечных <<одноразовых>> компьютеров для подарочных открыток и заканчивая огромными мощнейшими мейнфреймами, моделирующими термоядерные реакции и создающими прогнозы погоды. Ноутбуки, планшеты и смартфоны производятся миллионами штук ежегодно.

Однако устройством, наиболее широко применяемым для работы и научной деятельности, является персональный компьютер. Этот прибор включает несколько элементов, ряд из которых необходим для работы, а также различные модули расширения функциональности. Наше занятие будет посвящено изучению основных устройств.

Системный блок — корпус с установленными в него внутренними элементами ПК, такими как системная плата, устройства хранения данных, устройства охлаждения, электроснабжения, а также иными вспомогательными устройствами.

Корпус защищает внутренние компоненты компьютера от внешнего воздействия и механических повреждений. Кроме того, корпус поддерживает необходимый для нормального функционирования температурный режим, а также экранирует создаваемое внутренними компонентами электромагнитное излучение, которое может создавать помехи для работы других элементов ПК или иных электронных устройств.

Выпускаются корпуса для горизонтального (Desktop FootPrint SlimLine) или вертикального размещения (MiniTower, MidiTower, Full Tower).

Корпус системного блока, выполненный в распространенном формате MidiTower изготавливается из алюминиевого сплава и пластика. в нем предусмотрены:

- место установки источника электропитания;
- решетка системы вентиляции;
- заглушки гнезд плат расширения;
- место монтажа системной платы;
- отсек устройств хранения данных.

Одним из важнейших элементов ПК, устанавливаемых в системный блок, является системная плата — многослойная печатная плата, предназначенная для установки, соединения, питания и координации функционирования других элементов ПК.

На системной плате стандарта ATX предусматривается ряд элементов.

- Разъем установки центрального процессора (CPU socket).
- Разъемы установки плат ОЗУ (DIMM, DDR, RAM).
- Южный мост чипсета (south bridge). Чипсет — набор микросхем (контроллеров) системной платы, обеспечивающих взаимодействие устанавливаемых устройств (центральный процессор — оперативная память — постоянная память — видеоадаптер — устройства ввода-вывода).
- Разъемы подключения устройств хранения данных eSATA (ЖМД, твердотельных накопителей SSD, приводов CD/DVD-ROM).
- Разъемы подключения плат расширения PCI (видеоадаптер, звуковые и сетевые устройства).

- Задняя панель для подключения устройств ввода/вывода. Разъемы PS/2 предназначены для подключения клавиатуры и мыши, также подобные устройства могут подключаться в разъемы USB. Существует большое количество устройств, использующих подобный интерфейс: съемные носители, принтеры, телефоны и т. д. Интерфейс стандарта RJ-45 используется для подключения кабеля сети стандарта Ethernet. Разъемы HDMI, DVI, VGA используются для подключения устройств отображения информации (мониторов, телевизоров, проекторов, устройств AR). Разъем COM для прямого доступа к консоли управления различных специализированных устройств: мультиплексоров, роутеров, микроконтроллеров.

- Разъем подключения источника электропитания.

Самым важным элементом ПК является центральный процессор, устанавливаемый в специальное гнездо — сокет. Тип сокета является важнейшей характеристикой системной платы, определяя список совместимых процессоров. Сокеты отличаются числом контактов, которое обычно растет вместе с мощностью и сложностью процессоров. Часть контактов используется для питания процессора, а часть — для работы самого процессора, шины PCI Express, ОЗУ и т. д. Для каждого сокета существует уникальная распиновка контактов.

Оперативная память (RAM) применяется для хранения данных и команд, оперативной (быстрой) выдачи их в процессор. Скорость обращения к этой памяти и скорость передачи данных из нее в процессор значительно превышает скорость обмена с устройствами постоянной памяти.

Современные системные платы имеют несколько разъемов, позволяющих устанавливать различные платы расширения, которые соединяются с чипсетом, обеспечивающим взаимодействие устройств, с помощью шины стандарта PCI Express. В эти разъемы можно устанавливать платы видеозахвата, звуковые платы, сетевые карты. Одним из видов плат, позволяющих выполнять обработку данных и преобразование в вид, подходящий для вывода на устройство отображения (монитор) являются платы графического адаптера (графического процессора). Благодаря современным видеокартам возможна обработка видео высокого разрешения, моделирование 3D, и конечно, видеоигры с сценами высочайшей реалистичности.

Все перечисленные элементы ПК построены на современной элементной базе и им требуется электрическое питание. Это определяет высокие требования к следующему важнейшему элементу ПК.

Источник электропитания персонального компьютера

Встроенный источник электропитания (блок питания) — элемент ПК, обеспечивающий преобразование переменного тока бытовой сети в стабилизированные постоянные токи, необходимые для работы других элементов, а также защиту от перепадов напряжения и фильтрацию помех. Кроме того, БП должен быть компактным, не создавать дополнительных проблем для температурного режима, при этом эффективно и, что весьма существенно, безопасно преобразовывать ток.

В современном импульсном блоке питания ток в ходе преобразования из переменного тока напряжением 220 В трансформируется в целый набор постоянных токов различного номинала.

В состав блока питания входят:

- входной выпрямитель преобразует переменное напряжение в пульсирующее;
- конденсаторный фильтр сглаживает пульсации выпрямленного напряжения;
- импульсный высокочастотный трансформатор формирует необходимые номиналы напряжения, а также используется для гальванической развязки цепей (входных от выходных, а также, при необходимости, выходных друг от друга);
- дроссель выходной групповой стабилизации сглаживает импульсы и перераспределяет энергии между цепями выходных напряжений;
- выходные фильтрующие конденсаторы интегрируют импульсы для обеспечения необходимых значений напряжений.

Требуемые значения напряжений и допустимые отклонения представлены в табл. 1.

Таблица 1. Требуемые значения напряжений и допустимые отклонения

Выход	Параметр			
	Допуск, %	Минимум, В	Номинальное, В	Максимум, В
+12V1DC	± 5	+ 11,40	+ 12,00	+ 12,60
+12V2DC	± 5	+ 11,40	+ 12,00	+ 12,60
+5 VDC	± 5	+ 4,75	+ 5,00	+ 5,25

Выход		Параметр		
+3.3 VDC	± 5	+ 3,14	+ 3,30	+ 3,47
−12 VDC	± 10	− 10,80	− 12,00	− 13,20
+5 VSB	± 5	+ 4,75	+ 5,00	+ 5,25

Питание на системную плату подается через 24-контактный разъем, в котором 6 контактов для напряжения 5 В, один -5 В, два контакта 12 В, один −12 В, 4 контакта 3,3 В, контакты <<земли>> и служебные контакты PS_ON для команды включения БП от процессора, PW_OK сигнал готовности БП для процессора.

Кроме того, отдельно для питания процессора предусматривается 4-контактный разъем, два разъема 12 В и два общих.

Современные блоки питания имеют выходные кабели для питания устройств хранения данных и устройств охлаждения. Отдельного питания требуют многие современные видеоадаптеры.

Стоит отметить, что другие виды компьютеров, такие как ноутбуки, планшеты, смартфоны используют блоки питания, построенные по схожим принципам, но зачастую в ином исполнении. При этом блоки внешние, а питание подается через разъемы на корпусе.

Также необходимо упомянуть и средства гарантированного электроснабжения, позволяющие избежать проблем, связанных с внезапным отключением внешней электросети, что может привести как к потере важных данных, так и к выходу из строя элементов ПК.

Для избежания подобных проблем, применяются Источники бесперебойного питания (Uninterruptible Power Supply (UPS)) — устройства, обеспечивающие электропитание ПК от резервных источников (аккумуляторов, химических источников тока и т. п.) при пропадании или недопустимом ухудшении параметров напряжения основного источника.

Современные ИБП строятся по двум основным схемам, при этом первая схема, Back UPS, самая простая, надежная и дешевая, однако вторая схема линейно-интерактивная, позволяет помимо резервирования осуществлять и стабилизацию питания. Поэтому устройства с такой схемой получают все большую долю в современном парке компьютерной техники.

Внешние устройства хранения данных. Устройства ввода/вывода

Помимо энергозависимой оперативной памяти, в состав компьютера включаются и различные внешние устройства хранения данных.

Внешние устройства хранения данных — элементы ПК, размещаемые в системном блоке, либо подключаемые через разъёмы плат расширения, предназначенные для долговременного хранения данных, при этом не требующие постоянного наличия электропитания. К ним относятся:

- распаяваемая на системной плате микросхема памяти, содержащая программу базовой системы ввода/вывода (BIOS);
- различные твердотельные устройства, использующие технологию flash памяти — SSD накопители, USB накопители, карты microSD;
- накопители на жестких магнитных дисках и на магнитных лентах, могут размещаться как внутри системного блока, так и снаружи, подключаясь к портам задней панели системной платы.

Еще одним классом устройств, без которых невозможно использование компьютера оператором, являются устройства ввода/вывода данных.

Устройства ввода-вывода данных — элементы ПК, предоставляющие функции для взаимодействия ПК с пользователями и другими устройствами (системами). Устройства ввода данных — элементы, преобразующие неэлектрические величины (давление, температуру, освещенность, положение в пространстве) в цифровые электрические сигналы для их последующей обработки. Это такие устройства, как манипуляторы <<мышь>> или трэкбол, джойстики, клавиатуры, микрофоны, сенсоры экранов, считыватели штрих- и QR-кодов.

Устройства вывода данных — преобразователи цифровых электрических сигналов, выдаваемых из ПК, в сигналы иной физической сущности (световые, тепловые, механические) для передачи пользователю или другому устройству.

Это устройства визуального отображения, мониторы, проекторы, устройства печати, принтеры, плоттеры, наушники и устройства альтернативной/дополненной реальности.

Рассмотренные элементы ПК являются самыми важными, обеспечивающими работу и выполнение ключевых задач ПК — выполнение программ и обработку, и хранение данных, взаимодействие с пользователем.

3. ОБЕСПЕЧЕНИЕ ЭФФЕКТИВНОГО ФУНКЦИОНИРОВАНИЯ КОМПЬЮТЕРОВ

В данной теме будут рассмотрены следующие вопросы: показатели эффективности функционирования компьютеров и методы повышения надежности компьютеров.

Компьютеры — это сложные технические системы, включающие в себя множество разнотипных элементов, начиная от процессора и заканчивая периферийными устройствами. Полезные свойства этих систем реализуются в ходе процесса их функционирования, то есть выполнения задач, для которых создана система. Современные компьютерные системы позволяют решать сложнейшие вычислительные задачи, обрабатывая огромные объемы данных. При этом, с точки зрения практики использования компьютерных систем (КС) весьма актуальной задачей является определение соответствия системы поставленным целям.

Показатели качества и эффективности функционирования компьютеров

К компьютеру в полной мере может быть применено определение системы — множество элементов, находящихся в отношениях (связях) друг с другом, образующее определённую целостность и обладающее свойствами, не присущими отдельным элементам. В современной науке пригодность системы для использования по назначению описывается понятием качества, т. е. совокупности свойств.

В свою очередь, под свойством понимается объективная особенность (характерная отличительная черта) системы, которая может проявляться при ее создании, эксплуатации и утилизации.

Основными свойствами КС являются:

- надежность;
- производительность;
- экономичность;
- экологичность;
- транспортабельность.

Для того чтобы мы могли описывать проявление свойства системы, а также сравнивать системы, обладающие одинаковым набором свойств, вводятся понятия характеристики системы (в основном для качественного описания, например, надежная, безопасная, недорогая) и показателя свойства (для количественного описания).

Среди наиболее часто применяемых показателей качества для компьютерных систем выделяют:

- надежность:
 - вероятность безотказной работы (ВБР);
 - коэффициент оперативной готовности (КОГ);
- производительность:
 - количество операций, выполняемых в единицу времени;
 - частота тактового генератора;
 - пропускная способность шин;
- экономичность — разница стоимости приобретения относительно других однотипных КС;
- экологичность — риск (вероятность и последствия) нанесения ущерба окружающей среды при эксплуатации и утилизации;
- транспортабельность:
 - габаритные размеры ($Ш \times В \times Г$);
 - масса.

Отдельно необходимо отметить комплексное свойство системы, реализующееся в процессе ее функционирования и характеризующее приспособленность системы выполнять поставленные задачи (т. е. в какой степени могут быть достигнуты цели). Этим свойством является эффективность функционирования КС.

При исследовании этого свойства под процессом мы будем понимать последовательную смену состояний системы, действий (операций) для достижения какого-либо результата.

Для процесса функционирования показатели эффективности могут быть разделены на четыре основные группы:

- результативности;
- оперативности;
- ресурсоемкости;
- безопасности.

Наиболее часто используемыми являются показатели компьютерных систем, описывающие:

- результативность — вероятность того, что система выполнит требуемые задачи за заданное время (цель будет достигнута), РДЦ;
- оперативность — время, требуемое на выполнение требуемых (типовых) задач, T , с;
- ресурсоемкость — стоимость одного часа функционирования КС C , у. е./ч;

- безопасность — вероятность травмирования персонала при использовании КС.

Основным инструментом получения значений параметров КС являются различные программные продукты (тесты, бенчмарки). Однако необходимо иметь в виду, что результаты, полученные во время подобного тестирования, могут существенно отличаться от результатов реальной эксплуатации. Таким образом, данные полученные при тестировании, применяются для ориентировочного оценивания качества КС. Реальные же значения показателей могут быть получены только после обработки статистики по итогам нескольких итераций процесса функционирования.

Одной из современных концепций, позволяющих сочетать методологии синтетического тестирования, имитационного моделирования и анализа получаемой в режиме реального времени телеметрии, являются цифровые двойники.

Возможно с их применением удастся повысить точность и качество решений, принимаемых при управлении эффективностью КС. Задача эта весьма сложная, так как зачастую в ряду показателей эффективности присутствуют антагонисты, к примеру желая получить систему, которая позволит за короткое время с высокой вероятностью решать поставленные задачи, необходимо учитывать, что при этом стоимость ее будет возрастать, не забывая учитывать необходимость обеспечения безопасности.

Методы повышения надежности компьютеров

Одним из важнейших комплексных свойств КС, определяющих ее возможность выполнения задач, и, как следствие, результативность ее функционирования является надежность. В соответствии с действующими документами по стандартизации под надежностью следует понимать свойство системы сохранять во времени в установленных пределах значения всех параметров, характеризующих способность выполнять требуемые функции в заданных режимах, условиях применения, стратегиях технического обслуживания, хранения и транспортирования.

Надежность как комплексное свойство включает безотказность, ремонтпригодность, сохраняемость и долговечность.

Безотказность — свойство объекта непрерывно сохранять работоспособное состояние в течение некоторого времени или наработки в заданных режимах и условиях применения.

Долговечность — свойство объекта сохранять работоспособное состояние до наступления предельного состояния при установленной системе технического обслуживания и ремонта.

Ремонтопригодность — свойство объекта, заключающееся в приспособленности к поддержанию и восстановлению работоспособного состояния путем технического обслуживания и ремонта.

Сохраняемость — свойство объекта сохранять в заданных пределах значения параметров, характеризующих способности выполнять требуемые функции в течение и после хранения и (или) транспортирования.

С точки зрения эффективности функционирования наиболее важной компонентой надежности является безотказность. Отказ оборудования, т. е. переход его из работоспособного в неработоспособное состояние, приводит к его простоя, необходимости проведения ремонта, а в худшем случае, к необходимости полной замены КС. при этом возрастают потери, как финансовые, так и репутационные.

Однако существуют несколько методов, которые позволят не избежать отказов полностью, но увеличить (в некоторых случаях весьма существенно) время нахождения системы в работоспособном состоянии.

Наиболее распространёнными являются:

- резервирование — метод повышения надёжности технических устройств путём введения в их состав дополнительных избыточных элементов (узлов, связей) по сравнению с минимально необходимыми для выполнения заданных функций;
- обеспечение требуемого температурного, влажностного и чистотного режима работы. Проведение профилактического технического обслуживания и ремонта — организационных мероприятий и технических операций, направленных на поддержание и восстановление работоспособности КС, предупреждение ее отказов;
- выбор надежных элементов;
- упрощение системы.

Резервирование применяется в компьютерных системах самых разных типов: в ПК, серверах, супер ЭВМ и контроллерах. Основная идея, как следует из определения, это включение в состав избыточных элементов, которые начнут функционировать в случае отказа основных.

При этом применяются одна (или сочетание) следующих стратегий использования резерва:

- нагруженное (<<горячее>>) резервирование (нагруженный резерв), при котором резервные элементы находятся в режиме основного элемента;
- облегченное (<<теплое>>) резервирование (облегченный резерв), при котором резервные элементы находятся в менее нагруженном режиме по сравнению с основным;
- ненагруженное (<<холодное>>) резервирование (ненагруженный резерв), при котором резервные элементы до начала выполнения ими функций основного элемента находятся в ненагруженном режиме.

Основной характеристикой резервирования является кратность — отношение числа резервных элементов к числу основных, выраженное несокращенной дробью. Резервирование одного основного элемента одним резервным (с кратностью 1:1) называется дублированием.

В <<больших>> промышленных системах, центрах обработки данных и отказоустойчивых кластерах применяют резервирование целых компьютеров (серверов).

Для центров обработки данных существует стандарт, определяющий допустимое среднее время простоя центра, и необходимую для его обеспечения схему резервирования (табл. 2).

Таблица 2. Стандарты ЦОД

Уровень безотказности ЦОД	Среднее ожидаемое время простоя за год, ч	Схема резервирования
Tier I	28,8	нет
Tier II	22	N + 1
Tier III	1,6	N + 1 + TO
Tier IV	0,4	2(N + 1)

Для Tier I резервирования не предусмотрено.

Для Tier II на каждые N основных серверов предусмотрен один резервный.

На Tier III избыточность не увеличивается, однако при этом предусматривается возможность проведения обслуживания и ремонтных работ без прекращения функционирования ЦОД.

Tier IV предусматривает полное дублирование основных систем.

Для систем, включающих в свой состав всего один компьютер, также возможно применение резервирования. в этом случае основными приемами являются:

- резервирование накопителей информации (RAID);
- резервирование внешних сетевых соединений;
- резервирование соединений внутренней сети передачи данных;
- избыточные схемы электропитания с ИБП и резервируемыми блоками питания.

Требуемый режим работы включает в себя:

- допустимую рабочую температуру. Для ее регулирования используются как внутренние системы охлаждения КС, так и внешние, обеспечивающие нужную температуру в месте установки компьютера;
- допустимую влажность воздуха (либо иных газов, составляющих окружающую среду) обеспечивают с помощью промышленных осушителей;
- чистота воздуха, удаление из нее примесей пыли и других веществ, обеспечивается системой фильтрации.

Несмотря на любые меры, принимаемые для повышения безотказности, избежать отказов полностью невозможно. Это явление определяется природой технических элементов, в том числе, расходом имеющегося ресурса, изнашиванием и устареванием оборудования.

Для замены элементов, выработавших свой ресурс, устаревших морально или физически, проводится техническое обслуживание компьютерных систем.

Для восстановления работоспособности отказавших КС выполняется их ремонт.

Для своевременного и правильного проведения этих мероприятий создается система технического обслуживания и ремонта, включающая документы, персонал, организационные мероприятия и технические операции, выполняемые в соответствии с выбранной стратегией ТО для поддержания или восстановления работоспособного состояния компьютерной системы.

Стратегии технического обслуживания бывают:

- периодическая (календарная) — периодичность и объем выполняемых операций технического обслуживания определяются в соответствии с периодом времени (наработкой), заданным относительно начала эксплуатации или предыдущего выполнения ТО.

- по состоянию — периодичность и объем выполняемых операций технического обслуживания определяются в соответствии с фактическим состоянием компьютерной системы.

Для замены изношенного (ТО) или отказавшего (ремонт) элемента компьютерной системы необходимо создание комплекта запасных частей и инструментов. Кроме того, целесообразно иметь запас расходных материалов, необходимых для обеспечения нормальной работы системы.

На этапе проектирования КС в обязательном порядке должны быть рассмотрены вопросы ее максимального упрощения, с целью сокращения числа отказывающих элементов. Чем система проще, тем она надежнее. Однако упростить систему абсолютно невозможно, в таком случае она утратит возможность функционировать. Элементы, без которых невозможно обойтись, необходимо подбирать на основе максимально достоверной информации об их надежности, в первую очередь ориентируясь на имеющуюся статистику по их использованию в других КС.

Одной из важнейших задач, возникающих при проектировании, построении и эксплуатации компьютерных систем является оценивание их качества, в том числе и эффективности функционирования. Полученные оценки используются для принятия решений, позволяющих улучшить эффективность КС, в том числе и путем повышения ее надежности. Для этого существует несколько методов, однако наибольший эффект может быть получен при обоснованном целесообразном их сочетании.

4. ЦЕНТРАЛЬНЫЙ ПРОЦЕССОР. АРХИТЕКТУРА ПРОЦЕССОРА

В данной теме рассмотрены следующие вопросы: классификация и основные характеристики микропроцессоров и особенности процессоров CISC, RISC и VLIW.

Электронные вычислительные машины, получившие распространение в мире с начала 50-х гг. XX в., существенно повлияли на развитие нашей цивилизации. Однако настоящим прорывом, началом отсчета новой эпохи в истории можно считать день 15 ноября 1971 г. В этот день небольшая компания из США, Intel, начала продажи первого коммерческого микропроцессора Intel-4004, разработанного инженером Эдвардом Хоффом. Всего за 50 лет микропроцессорная техника совершила взлет от небольшой схемы, выполняющей операции над 4 разрядными числами с тактовой частотой 750 кГц, до современных процессоров x64, производительность которых приближается к триллиону операций над числами с плавающей точкой в секунду. Успех современных специалистов в области информационных технологий не возможен без понимания принципов построения и функционирования микропроцессоров — основы современной техники.

Классификация и основные характеристики микропроцессоров

Микропроцессор (МП) — программно управляемое устройство, предназначенное для обработки цифровой информации и управления процессом обработки, выполненное в виде одной или нескольких сверхбольших интегральных схем.

Разработчики при создании МП в первую очередь руководствуются принципами универсальности, технологичности и высокой производительности.

Технологичность и высокая производительность достигаются применением модульного принципа построения МП на основе быстродействующих больших (БИС) и сверхбольших интегральных схем (СБИС).

Интегральная микросхема (чип) — электрическая схема произвольной сложности, выполненная на общей подложке-кристалле. СБИС содержит более 10 000 элементов (транзисторов, диодов, резисторов и т. п.).

Высокая универсальность и гибкость МП, достигаемая благодаря программному управлению, низкая стоимость, небольшие размеры,

повышенная надежность, возможность встраивания микропроцессорных средств в приборы, машины и технологические процессы обеспечивают микропроцессорам исключительно широкое применение в различных управляющих и обрабатывающих данные цифровых устройствах и системах. Основу современных процессоров составляют полупроводниковые технологии. Наиболее широкое распространение в настоящее время получило изготовление МП по технологии КМОП — комплементарная структура металл — оксид — полупроводник. Применение данного подхода позволяет создавать с применением метода фотолитографии и легирования кремния МП, содержащие миллиарды транзисторов на одной схеме. Самые совершенные процессоры сейчас содержат транзисторы, размер которых составляет 10 нанометров.

Основные этапы изготовления МП:

1. выращивание кристалла кремния, нарезка <<вафель>> толщиной 1 мм, диаметром 300 мм;
2. нанесение фоторезиста, УФ-фотолитография;
3. легирование подложки и нанесение диэлектрика (оксидной пленки);
4. нанесение слоя меди гальванизацией, шлифовка дорожек.

МП могут быть классифицированы следующим образом.

- По числу используемых в МП интегральных схем:

- однокристалльные;
- многокристалльные;
- многосекционные.

Однокристалльные микропроцессоры получают при реализации всех аппаратных средств процессора в виде одной БИС или СБИС (сверхбольшой интегральной схемы). По мере увеличения степени интеграции элементов в кристалле и числа выводов корпуса параметры однокристалльных микропроцессоров улучшаются. Однако возможности однокристалльных микропроцессоров ограничены аппаратными ресурсами кристалла и корпуса.

Для получения многокристалльного микропроцессора необходимо провести разбиение его логической структуры на функционально законченные части и реализовать их в виде БИС (СБИС). Функциональная законченность БИС многокристалльного микропроцессора означает, что его части выполняют заранее определенные функции и могут работать автономно.

Микропроцессорная секция — это БИС, предназначенная для обработки нескольких разрядов данных или выполнения определенных управляющих операций. Секционность интегральных схем МП определяет возможность <<наращивания>> разрядности обрабатываемых данных или усложнения устройств управления микропроцессора при <<параллельном>> включении большего числа секций.

- По назначению:
 - универсальные;
 - специализированные.

Универсальные микропроцессоры могут быть применены для решения широкого круга разнообразных задач. при этом их эффективная производительность слабо зависит от специфики решаемых задач.

Специализация МП, т. е. его ориентация на ускоренное выполнение определенных функций позволяет резко увеличить эффективную производительность при решении только отдельных задач.

- По типу входных/выходных сигналов:
 - цифровые;
 - аналоговые.

Микропроцессоры — цифровые устройства, однако они могут иметь встроенные аналого-цифровые и цифро-аналоговые преобразователи. Поэтому входные аналоговые сигналы передаются в МП через преобразователь в цифровой форме, обрабатываются и после обратного преобразования в аналоговую форму поступают на выход. Такие МП выполняют функции любой аналоговой схемы (например, производят генерацию колебаний, модуляцию, смещение, фильтрацию, кодирование и декодирование сигналов в реальном масштабе времени и т. д., заменяя сложные схемы, состоящие из операционных усилителей, катушек индуктивности, конденсаторов и т. д.).

- По характеру временной организации работы:
 - синхронные;
 - асинхронные.

Синхронные МП — МП, в которых начало и конец выполнения операций задаются устройством управления (время выполнения операций в этом случае не зависит от вида выполняемых команд и величин операндов).

Асинхронные МП позволяют начало выполнения каждой следующей операции определить по сигналу фактического окончания выполнения предыдущей операции.

- По организации структуры:
 - одномагистральные;
 - многомагистральные.

В одномагистральных МП все устройства имеют одинаковый интерфейс и подключены к единой информационной шине, по которой передаются коды данных, адресов и управляющих сигналов.

В многомагистральных МП устройства группами подключаются к своей информационной магистрали. Это позволяет осуществить одновременную передачу информационных сигналов по нескольким (или всем) шинам. Такая организация систем усложняет их конструкцию, однако увеличивает производительность.

- По количеству выполняемых программ:
 - однопрограммные;
 - многопрограммные.

В однопрограммных микропроцессорах выполняется только одна программа. Переход к выполнению другой программы происходит после завершения текущей программы.

В много-, или мультипрограммных, МП одновременно выполняется несколько (обычно несколько десятков) программ.

Свойства каждого МП могут быть описаны рядом характеристик и показателей:

- Тактовая частота. Определяет минимальное возможное время выполнения переключения элементов в МП.
- Разрядность, т. е. максимальное число одновременно обрабатываемых двоичных разрядов. Разрядность МП 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.

Для современных процессоров общего назначения, таких как Intel Core или AMD Ryzen $m = n = k = 64$.

- TDP (Thermal Design Power). Эта величина показывает максимальное количество тепла, которое должна рассеивать система охлаждения МП. Эти требования указываются для нормальных условий эксплуатации. На случай отказа системы теплоотведения в современных МП

предусмотрена функция принудительного отключения либо перехода в режим дросселирования (trottlng), когда процессор начинает сокращать число исполняемых тактов.

- Архитектура микропроцессора — логически связанный набор характеристик и параметров МП, отражающий:

- структуру, т. е. совокупность компонентов, составляющих МП, и связей между ними;

- способы представления и форматы данных;

- способы обращения ко всем доступным для пользователя (программно-доступным) элементам структуры (адресация регистров, ячеек оперативной и постоянной памяти, внешних устройств);

- характеристики управляющих слов и сигналов, вырабатываемых микропроцессором и поступающих в МП извне;

- реакцию на внешние сигналы (схема обработки прерываний и т. д.);

- система набора команд и подход к их исполнению. Наиболее распространёнными системами в настоящее время являются CISC, RISC, VLIW.

Особенности процессоров CISC, RISC и VLIW

Команда микропроцессора — это такое двоичное слово, которое, будучи <<прочитано>> МП, заставляет последний выполнять определенные действия. Другие, отличные от команд, слова подобных действий в МП вызывать не могут. Команды осуществляют пересылку или обработку данных, расположенных в памяти или одном из регистров МП. Некоторые команды предназначены для управления некоторыми вспомогательными функциями МП, поддерживающими необходимый режим его работы.

Когда говорят о системе (наборе) команд МП, подразумевают все его команды.

В общем случае система команд процессора включает в себя следующие четыре основные группы команд:

- пересылки данных;
- арифметические;
- логические;
- переходов.

Команды пересылки данных не требуют выполнения никаких операций над операндами. Операнды просто пересылаются (точнее,

копируются) из источника (Source) в приемник (Destination). Источником и приемником могут быть внутренние регистры процессора, ячейки памяти или устройства ввода/вывода. АЛУ в данном случае не используется.

Арифметические команды выполняют операции сложения, вычитания, умножения, деления, увеличения на единицу (инкрементирования), уменьшения на единицу (декрементирования) и т. д. Этим командам требуется один или два входных операнда. Формируют команды один выходной операнд.

Логические команды производят над операндами логические операции, например, логическое И, логическое ИЛИ, исключающее ИЛИ, очистку, инверсию, разнообразные сдвиги (вправо, влево, арифметический и циклический сдвиги). Этим командам, как и арифметическим, требуется один или два входных операнда, и формируют они один выходной операнд.

Наконец, команды переходов предназначены для изменения обычного порядка последовательного выполнения команд. С их помощью организуются переходы на подпрограммы и возвраты из них, всевозможные циклы, ветвления, пропуски фрагментов программ и т. д. Команды переходов всегда меняют содержимое счетчика команд. Переходы могут быть условными и безусловными. Именно эти команды позволяют строить сложные алгоритмы обработки информации.

Первоначально почти все производители первых микропроцессоров использовали архитектуру с расширенным набором команд — CISC (Complex Instruction Set Computer).

Причина этого в том, что разработчики пытались уменьшить так называемый семантический разрыв между тем, что компьютеры способны делать, и тем, что требуют языки программирования высокого уровня, пытаясь заменить одной инструкцией многочисленные машинные коды.

На заре компьютерной эры на рынке коммерческих вычислений доминировали мини-компьютеры PDP компании DEC и мейнфреймы компании IBM, которые использовали архитектуру с CISC.

Среди микропроцессоров типичными представителями архитектуры с CISC стали процессоры компании Intel. На начальном этапе развития микропроцессоров (70-е гг. XX в.) были и другие процессоры подобной архитектуры производства компаний Motorola, Zilog, MOS Technology и т. д.

Но именно благодаря коммерческой привлекательности микропроцессоров Intel эта архитектура стала самой популярной на текущий момент и практически единственной для персональных

компьютеров. Даже компания Apple в своих компьютерах Apple Macintosh в итоге перешла от процессоров PowerPC к процессорам Intel.

Особенности системы Intel:

- нефиксированное значение длины команды;
- арифметические действия кодируются в одной команде;
- двухадресная система команд;
- небольшое количество регистров, каждый из которых выполняет

строго определённую функцию.

Формально все МП семейств x86 (x32, IA32) и x64 (AMD64) принадлежат к системе CISC.

Одним из основных направлений развития МП всегда было повышение их производительности. Для достижения этой цели применяется комплексный подход: одновременно с повышением тактовой частоты процессора используют конвейерные вычисления и параллелизм. Более подробно суть этих методов будет рассмотрена в дальнейших разделах.

Повышение производительности CISC-микропроцессоров из-за особенностей архитектуры приводило к росту количества транзисторов, в результате чего кристаллы становились всё более сложными и дорогостоящими в производстве. Архитектура с CISC в первозданном своём виде достигла потолка производительности. Переход к применению конвейерных вычислений потребовал использования простых и быстрых команд, что привело к появлению архитектуры с RISC (Reduced Instruction Set Computer), что означает <<компьютер с сокращённым набором команд>>.

Архитектура была разработана в рамках проекта Berkeley RISC. в 1980 г. группа разработчиков исследовав работу процессора Motorola 68000 выяснила, что программы попросту не использовали подавляющее большинство инструкций, заложенных в процессор. Планировалось создать процессор, который бы содержал лишь самые необходимые инструкции. При этом не только уменьшилось общее количество процессорных инструкций, принципиальное отличие заключается в том, что любая инструкция платформы RISC является простой и выполняется за один такт (по крайней мере, должна выполняться), тогда как на выполнение CISC-инструкции могло уходить несколько десятков тактов. при этом длина команды является фиксированной.

RISC имеет ряд характерных особенностей:

- Сокращённый набор команд. Первый <<настоящий>> RISC-процессор имел всего 31 команду. в дальнейшем их количество постепенно росло и достигло 100–200 инструкций в зависимости от реализации процессора. Но это всё равно в несколько раз, а то и на порядок меньше инструкций CISC-процессоров.

- Большинство команд выполняется за один такт. Все команды выполняются непосредственно аппаратным обеспечением, то есть напрямую без интерпретации микрокомандами. Устранение уровня интерпретации повышает скорость выполнения команд. В компьютерах типа CISC более сложные команды разбиваются на несколько шагов, которые потом выполняются как последовательность микрокоманд.

- Большое количество регистров общего назначения. Доступ к памяти происходит относительно медленно. Если слово было загружено из памяти, оно может храниться в регистрах до тех пор, пока не потребуется. Возвращение слова из регистра в память весьма нежелательно, и лучший способ избежать лишних изменений — это наличие достаточного количества регистров.

- Наличие параллелизма на уровне команд (одновременный запуск) обеспечиваемого многоступенчатыми конвейерами.

- Все команды имеют простой формат, и используются немногие способы адресации. Команды легко декодируются, и к памяти обращаются только команды загрузки и сохранения.

- Наличие вемстительной раздельной кэш-памяти. Это необходимо для уменьшения обращений к памяти и тем самым обеспечения необходимого быстродействия для заполнения регистров и конвейеров.

- Использование оптимизирующих компиляторов, которые анализируют исходный код и частично меняют порядок следования команд.

Уже первые микропроцессоры RISC значительно опережали процессоры CISC по производительности. Учитывая это, можно было предположить, что они должны были занять доминирующее положение на рынке ПК, но этого не произошло, по крайней мере, по двум причинам. Во-первых, компьютеры RISC несовместимы с архитектурой Intel x86, а многие компании уже вложили значительные средства в программное обеспечение для продукции Intel. И во-вторых, компания Intel сумела

воплотить те же идеи в своей архитектуре, все современные процессоры семейств x64 содержат RISC ядра.

Дальнейшее развитие архитектуры МП, направленное на повышение производительности, привело к созданию архитектуры с VLIW (Very Long Instruction Word — сверхдлинное командное слово). Эта система основывается на явно выраженном параллелизме вычислений, заложенном в систему команд процессора.

Основной принцип — перенос нагрузки с периода исполнения в период компиляции.

Задачи переупорядочивания команд, замены регистров, распределения функциональных блоков конвейера и т. п. В системе VLIW заранее решает компилятор, который располагает полной и достоверной информацией о регистрах процессора и генерирует оптимальный код, в котором нет конфликтов между регистрами.

Кроме того, компилятор следит за загрузкой функциональных блоков и не запускает команды, в которых предполагается обращение к занятым функциональным блокам, планирует поток команд статически.

В результате на вход VLIW-процессора поступает последовательность больших команд, состоящих из нескольких простых операций, которые могут выполняться параллельно разными функциональными блоками VLIW-процессор имеет простой, а значит, и более быстрый исполнительный конвейер с минимальным количеством ступеней.

Примерами VLIW-процессора являются Intel Itanium архитектуры IA-64 (Intel Architecture 64-bit) и «Эльбрус» производства компании «МЦСТ», цифровой сигнальный процессор Qualcomm Hexagon, используемый в SoC Snapdragon.

Основные особенности систем реализации набора команд CISC, RISC, VLIW представлены в табл. 3.

Таблица 3. Основные особенности систем реализации набора команд

Характеристика		Архитектура	
		CISC	VLIW
1. Длина команды	Различная	Одинаковая	Одинаковая
2. Расположение полей в командах	Различное	Одинаковое	Одинаковое

Характеристика	Архитектура		
	CISC	RISC	VLIW
3. Количество регистров	Малое, регистры специализированные	Большое, универсальные регистры	Большое, универсальные регистры
4. Доступ к памяти	Кодируется в команде, выполняется по макрокоманде	Выполняется по специальной команде	Выполняется по специальной команде
5. Длительность выполнения команд	Различная	Одинаковая (для большинства команд)	Различная

Кроме этих различий необходимо отметить и иные особенности, так для CISC-системы преимуществами можно считать простоту и эффективность программирования (несколько команд могут быть заменены одной более сложной), а также большое историческое наследие в виде множества написанных для неё программ. Недостатками же являются сложность построения процессора и невысокая эффективность использования конвейера.

Эта система на текущий момент де-факто является основной для настольных и серверных систем.

Для системы RISC преимущества следующие: относительно низкая сложность построения процессора, высокая производительность, меньшее энергопотребление, упрощение работы компилятора. Недостаток — отсутствие совместимости с CISC-машинами.

Реализация этой системы от компании ARM в настоящее время доминирует в мобильных системах, в первую очередь в смартфонах и планшетах.

Преимуществами VLIW системы являются выполнение большего количества операций за такт, упрощение конструкции процессора и технологического процесса его производства, уменьшение тепловыделения.

Недостатки: сложный компилятор, разработка, поддержка и оптимизация которого довольно ресурсоёмки, несовместимость процессоров VLIW с классической CISC архитектурой.

Задача повышения производительности МП не теряет актуальности и в настоящее время. Каждая из существующих систем команд, реализованных в различных архитектурах, создавалась в надежде превзойти предшествующие, компенсировать их недостатки. При этом, в свою очередь, проявлялись их собственные, неизвестные ранее ограничения. Современные компьютерные системы зачастую сочетают в себе несколько устройств, использующих различные системы и архитектуры.

5. ЦЕНТРАЛЬНЫЙ ПРОЦЕССОР. ПРИНЦИПЫ ПОСТРОЕНИЯ И ФУНКЦИОНИРОВАНИЯ МИКРОПРОЦЕССОРОВ

В данной теме рассмотрены следующие вопросы: структура микропроцессоров общего назначения и цикл тракта данных микропроцессора.

Очень существенная доля на рынке современной микропроцессорной техники принадлежит микропроцессорам общего назначения, применяемым как в персональных компьютерах, так и в промышленных вычислительных системах. Кроме того ряд суперкомпьютеров построен также с использованием такого рода процессоров. Структура и алгоритмы функционирования всех современных микропроцессоров общего назначения опираются на общие исходные принципы, изучение и понимание которых является безусловно актуальным для специалистов информационных технологий и компьютерных наук.

Структура микропроцессоров общего назначения

В 1946 году Джон фон Нейман, Герман Голдстайн и Артур Беркс в своей совместной статье изложили новые принципы построения и функционирования ЭВМ. В последствие на основе этих принципов производились первые два поколения компьютеров. В более поздних поколениях происходили некоторые изменения, хотя принципы Неймана, лежащие в основе архитектуры МП, актуальны и сегодня. По сути, Нейману удалось обобщить научные разработки и открытия многих других ученых и сформулировать на их основе принципиально новое.

Перечислим принципы фон Неймана:

- Использование двоичной системы счисления в вычислительных машинах. Преимущество перед десятичной системой счисления заключается в том, что устройства можно делать достаточно простыми, арифметические и логические операции в двоичной системе счисления также выполняются достаточно просто.
- Программное управление ЭВМ. Работа ЭВМ контролируется программой, состоящей из набора команд. Команды выполняются последовательно друг за другом. Созданием машины с хранимой в памяти программой было положено начало тому, что мы сегодня называем программированием.
- Память компьютера используется не только для хранения данных, но и программ. при этом и команды программы, и данные кодируются

в двоичной системе счисления, т. е. их способ записи одинаков. Поэтому в определенных ситуациях над командами можно выполнять те же действия, что и над данными.

- Ячейки памяти ЭВМ имеют адреса, которые последовательно пронумерованы. В любой момент можно обратиться к любой ячейке памяти по ее адресу. Этот принцип открыл возможность использовать переменные в программировании.

- Возможность условного перехода в процессе выполнения программы. Несмотря на то что команды выполняются последовательно, в программах можно реализовать возможность перехода к любому участку кода.

Самым главным следствием этих принципов можно назвать то, что теперь программа уже не была постоянной частью машины (как например, у калькулятора). Программу стало возможно легко изменить. А вот аппаратура, конечно же, остается неизменной и очень простой. Для сравнения, программа компьютера ENIAC (где не было хранимой в памяти программы) определялась специальными переключками на панели. Чтобы перепрограммировать машину (установить переключки по-другому) мог потребоваться далеко не один день. и хотя программы для современных компьютеров могут писаться годы, однако они работают на миллионах компьютеров после нескольких минут установки на жесткий диск.

В современной ЭВМ, построенной по схеме фон Неймана, два основных элемента: устройство управления и тракт данных. Они объединены в составе центрального микропроцессора.

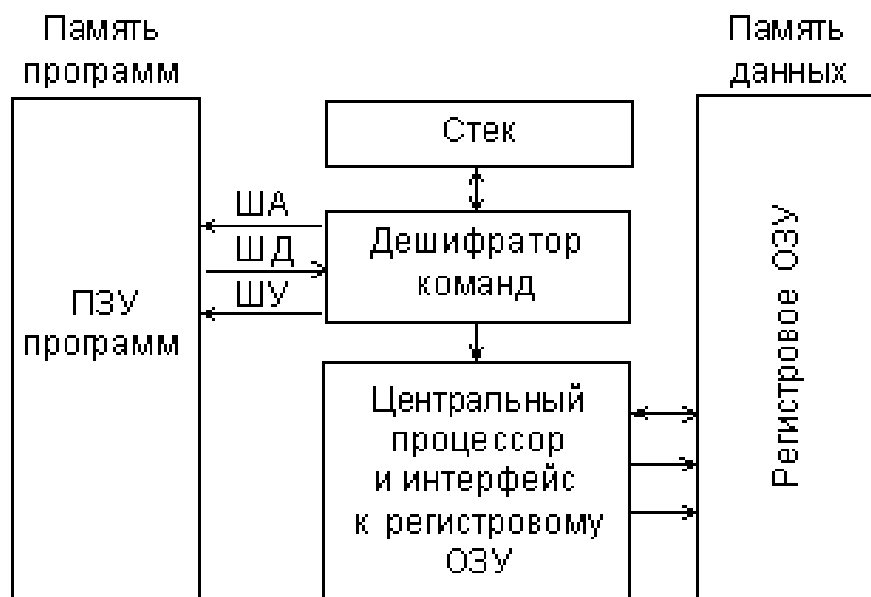


Рисунок 1.

Устройство управления МП представляет из себя микропрограммный конечный автомат.

При переходе из одного внутреннего состояния в следующее состояние конечного автомата УУ на основе полученного кода команды, адреса операндов, состояния регистра признаков результата выполнения предыдущей команды, состояния управляющих входов выдает на выход (линии управления) последовательности импульсов управления.

Тракт данных состоит из регистров, арифметико-логического устройства (АЛУ) и нескольких соединительных шин.

Регистр процессора — элемент энергозависимой памяти, размещаемый в ядре МП для хранения промежуточных результатов вычислений и служебной информации (инструкций, флагов).

Реализуются УУ и регистры на последовательностных цифровых устройствах — триггерах.

Шины это физические (электрические) соединительные линии для передачи данных между функциональными блоками микропроцессора, а также от внешних устройств и к ним.

В состав МП включаются:

- **кодовая шина данных (КШД)** для параллельной передачи кода всех разрядов операнда;
- **кодовая шина адреса (КША)** для параллельной передачи всех разрядов кода адреса ячейки основной памяти или порта ввода/вывода внешнего устройства;
- **кодовая шина инструкций (КШИ)** для передачи управляющих сигналов (импульсов) во все блоки МП;
- **шина электропитания.**

Арифметико-логическое устройство (АЛУ) — это функциональный блок процессора, предназначенный для выполнения логических и арифметических операций над данными — операндами. АЛУ реализуются на комбинационных цифровых устройствах.

Таким образом, на основе вышеизложенного и материала предыдущих занятий может быть сформулирована концепция, определяющая структуру всех современных микропроцессорных устройств. Она включает три основных уровня, погружаемых друг в друга.

Первым уровнем являются физические элементы, размещаемые на кристаллах кремния. В первую очередь это транзисторы, выполненные по КМОП-технологиям. Они объединяются в логические устройства, называемые ключами, или вентилями. Вентили выполняют элементарные

логические операции, преобразуя множество входных цифровых сигналов в выходной цифровой сигнал. Операции, выполняемые вентилями, могут быть записаны формализовано, либо в виде таблиц истинности.

Третий уровень — уровень микроархитектуры. на этом уровне из вентилях формируются цифровые комбинационные либо последовательностные устройства. Эти устройства выполняют гораздо более сложные операции шифрования и дешифрования, мультиплексирования, математические и логические операции. ПЦУ применяются для хранения данных и реализации устройств управления.

Эта концепция реализуется для построения современных МП общего назначения. Но при этом их структура с учетом возрастающего круга задач и требований к производительности постоянно совершенствуется.

Так, в процессорные ядра помещают дополнительные функциональные элементы для организации конвейерных вычислений. Число самих ядер постоянно увеличивается. Для этого в состав МП включаются средства управления и согласования их работы.

Обмен данными с памятью всегда был <<бутылочным горлом>> производительности. В структуру современных процессоров включают несколько контроллеров, обеспечивающих взаимодействие с оперативной памятью. Кроме того, каждое ядро располагает несколькими банками кэш-памяти, предназначенной для хранения наиболее часто используемых инструкций и данных.

В современных промышленных компьютерах широко используются многопроцессорные структуры. Это требует включения дополнительных устройств в состав МП для организации межпроцессорного взаимодействия.

Также необходимы элементы для организации ввода/вывода данных к внешним устройствам.

Цикл тракта данных микропроцессора

Функционирование современного микропроцессора общего назначения — сложный процесс, включающий в себя выполнение миллиардов операций в секунду. Но в его основе лежит последовательное повторяющееся выполнение ряда действий функциональными блоками ядра — цикл тракта данных. Задача МП — выполнение программ, находящихся в основной памяти, как последовательности команд (инструкций).

Основные типы инструкций:

- регистр-память — чтение данных из памяти в регистры, запись из регистров в память;
- регистр-регистр — действия над операндами с помощью АЛУ с сохранением результата в регистре.

Цикл тракта данных в довольно общем виде включает в себя следующие шаги:

1. Вызов следующей команды из памяти и перенос ее в регистр команд.
2. Изменение положения счетчика команд.
3. Определение типа вызванной команды (I v II).
4. Если необходимо, определение адреса данных в памяти.
5. Чтение данных из памяти в регистр.
6. Выполнение команды.
7. Переход к шагу 1.

При выполнении команд регистр-регистр устройством управления на основе полученного кода команды, адреса операндов, состояния регистра признаков результата выполнения предыдущей команды, состояния управляющих входов формируются сигналы управления, в соответствии с которыми в АЛУ выполняются действия над операндами.

Все операции цикла выполняются за время между фронтами двух импульсов тактового генератора. Крайне важно учитывать, что сигналы между элементами тракта распространяются не мгновенно, и необходима весьма точная синхронизация, а также точный расчет временных интервалов, необходимых для обеспечения правильной совместной работы.

Микропроцессоры общего назначения чрезвычайно сложные технические устройства, процесс функционирования которых с трудом поддается описанию и пониманию полностью. Однако основа их структуры и функционирования базируется на общих принципах, которые с дополнениями и изменениями продолжают реализовываться и в настоящее время.

6. УВЕЛИЧЕНИЕ БЫСТРОДЕЙСТВИЯ ПРОЦЕССОРОВ. ТЕХНОЛОГИИ ВЫПОЛНЕНИЯ КОМАНД В МИКРОПРОЦЕССОРЕ

В данной теме рассмотрены следующие вопросы: конвейеризация выполнения команд и динамическое и мультитредовое выполнение команд.

Растущая производительность компьютеров позволяет решать проблемы, к которым раньше нельзя было подступиться.

В промышленности и науке таких задач очень много: управление базами данных, молекулярно-динамические расчеты, моделирование сложных механизмов (автомобилей, реактивных двигателей) — все это требует возрастающей мощности вычислений. Даже на бытовом уровне есть задачи, требующие высокой производительности, например кодирование видео высокого разрешения. Производительность процессора определяется как количество инструкций, выполняемых за единицу времени. Таким образом, для повышения производительности необходимо либо увеличить тактовую частоту (повысить число тактов в единицу времени), либо увеличить число инструкций, выполняемых за один такт.

На начальных этапах развития компьютеров основной рост производительности обеспечивался с помощью уменьшения размеров элементов микропроцессоров. При этом падало энергопотребление и росли частоты работы, компьютеры становились все быстрее, сохраняя в общих чертах свою архитектуру.

Однако в настоящее время почти достигнут предел роста тактовой частоты — растут токи утечки, процессоры перегреваются и обойти это не получается. Можно, конечно, развивать системы охлаждения, применять водные радиаторы или системы на основе жидкого азота, но этот подход имеет свои сложности, он весьма затратный и при этом не обеспечивает существенного роста производительности.

Данные проблемы определили актуальность разработки нескольких методов, позволяющих повысить производительность процессоров.

Конвейеризация выполнения команд

С момента распространения процессоров, построенных на основе идей фон Неймана стало понятно, что главным препятствием для быстрого выполнения большого числа команд является необходимость их загрузки из памяти. Для разрешения этой проблемы можно вызывать команды из памяти заранее и хранить в специальном наборе регистров. Эта идея

использовалась еще в 1959 г. при разработке компьютера Stretch компании IBM, а набор регистров был назван буфером выборки с упреждением. Таким образом, когда требовалась определенная команда, она вызывалась прямо из буфера, а обращения к памяти не происходило.

В действительности при выборке с упреждением команда обрабатывается за два шага: сначала происходит выборка команды, а затем ее выполнение. Дальнейшим развитием этой стратегии стала концепция конвейера. при использовании конвейера команда обрабатывается уже не за два, а за большее количество шагов, каждый из которых реализуется определенным аппаратным компонентом, причем все эти компоненты могут работать независимо.

На рис. 2, а изображен конвейер из 5 блоков, которые называются ступенями.



Блоки пятиступенчатого конвейера

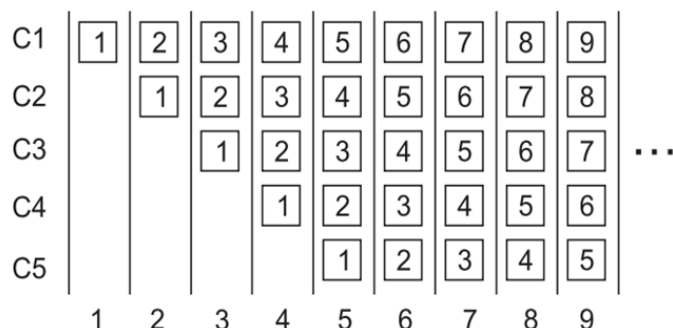


Рисунок 2. Продвижение команд по ступеням конвейера

Первая ступень, C1, по адресу, полученному из счетчика команд, считывание из памяти команды, запись ее в регистр;

C2 — определяет тип команды и операндов;

C3 — определение адресов операндов, чтение их из памяти, запись в регистры;

C4 — выполнение команды (в тракте данных);

C5 — запись результатов выполнения команды в регистры (в память).

Рис. 2 иллюстрирует, как действует конвейер во времени.

Во время цикла 1 блок C1 обрабатывает команду 1, вызывая ее из памяти.

Во время цикла 2 блок C2 декодирует команду 1, в то время как блок C1 вызывает из памяти команду 2.

Во время цикла 3 блок C3 вызывает операнды для команды 1, блок C2 декодирует команду 2, а блок C1 вызывает команду 3.

Во время цикла 4 блок C4 выполняет команду 1, C3 вызывает операнды для команды 2, C2 декодирует команду 3, а C1 вызывает команду 4.

Во время цикла 5 блок C5 записывает результат выполнения команды 1 обратно в регистр, тогда как другие ступени конвейера обрабатывают следующие команды.

Таким образом, применение конвейера позволяет получить прирост производительности по сравнению с простым последовательным выполнением инструкций.

Суммарное время, необходимое для выполнения N инструкций при известном времени выполнения каждой ступени T , может быть определено в соответствии с выражениями:

- последовательное выполнение инструкций

$$T_N = NTI;$$

- использование конвейера из I ступеней

$$T_N = IT + (N - 1) \cdot T.$$

Дальнейшим развитием идеи конвейерной обработки является применение двух конвейеров в сочетании с одной схемой выборки команд (рисунок 3).

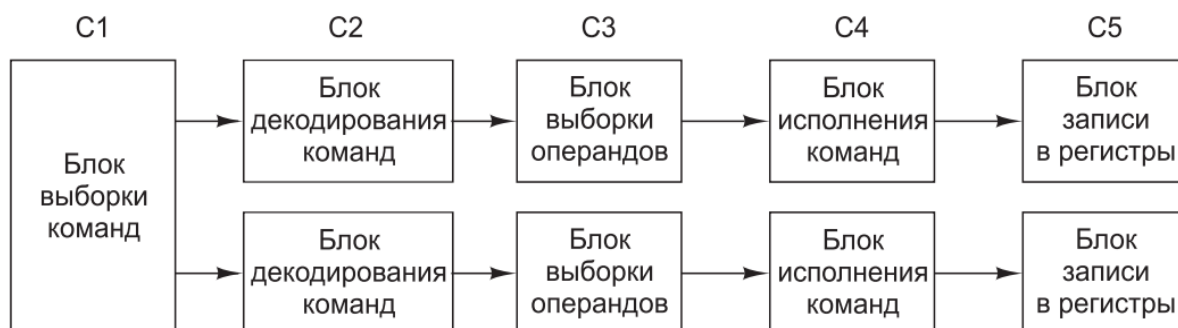


Рисунок 3. Параллельный пятиступенчатый конвейер

Здесь общий блок выборки команд вызывает из памяти сразу по две команды и помещает каждую из них в один из конвейеров. Каждый

конвейер содержит АЛУ для параллельных операций. Чтобы выполняться параллельно, две команды не должны конфликтовать из-за ресурсов (например, регистров) и ни одна из них не должна зависеть от результата выполнения другой. Как и в случае с одним конвейером, либо компилятор должен гарантировать отсутствие нештатных ситуаций (когда, например, аппаратура не обеспечивает проверку команд на несовместимость и при обработке таких команд выдает некорректный результат), либо конфликты должны выявляться и устраняться дополнительным оборудованием непосредственно в ходе выполнения.

Увеличение числа конвейеров возможно, но требует громоздкого аппаратного обеспечения. Вместо этого используется другой подход. Основная идея — один конвейер с большим количеством функциональных блоков. В 1987 г. для обозначения этого подхода был введен термин суперскалярная архитектура.

Основной принцип суперскалярности МП — применение структуры, включающей большое количество параллельных структур (конвейеров и функциональных блоков (АЛУ, FPU, умножителей, схем сдвига, блоков чтения/записи)) при сохранении последовательной программы. Загрузка параллельных структур обеспечивается автоматически аппаратурой МП либо компилятором (система VLIW).

Динамическое и мультитредовое выполнение команд

Использование суперскалярной технологии позволяет значительно (до 10 раз) повысить производительность МП.

Однако при этом возникает и ряд проблем. Одной из наиболее существенных является снижение производительности при выполнении программ, содержащих команды ветвления (условных переходов).

С целью снижения негативного влияния этой проблемы разработана технология динамического выполнения команд.

В состав технологии включены следующие основные методы: предсказание ветвлений, анализ потока команд, упреждающее выполнение.

Предсказание ветвлений (branch prediction). С использованием специального механизма процессор на основе статистики может предсказать переходы или ветвления в потоке команд и направляет поток выборки и декодирования по одной из ветвей конвейера. Это позволяет существенно ускорить выполнение программы, но при этом возникает

проблема — в случае, если предсказание окажется неверным, всё содержимое конвейера тоже окажется неверным, и его придется очищать.

Анализ потока команд (out-of-order execution). Это средство анализирует и планирует выполнение команд в оптимальной последовательности, независимо от их первоначального порядка в программе.

Упреждающее выполнение (data forwarding). Подразумевает начало исполнения инструкции до готовности всех операндов. При этом декодированная инструкция с одним операндом помещается в исполнительное устройство, где дожидается готовности второго операнда, выходящего из другого конвейера.

Применение данного подхода позволяет снизить простои МП, определенные ожиданием выполнения переходов и циклов.

Следующим шагом развития структуры МП, направленным на повышение производительности, является применение мультитредового исполнения команд.

Мультитредовое выполнение команд (multithreading) — метод повышения производительности микропроцессоров, основанный на применении нескольких устройств выборки команд и нескольких регистровых файлов.

Регистровый файл — несколько параллельных регистров с возможностью выборочного подключения на чтение/запись по сигналу управления.

Мультитредовая архитектура решает проблему борьбы с простоями функциональных устройств процессора, возникающими из-за невозможности выполнения следующей команды, путем переключения на другой регистровый файл и устройство выборки, тем самым процессор получает другой контекст для продолжения вычислений, переходя к выполнению другого треда (процесса).

Помимо методов повышения производительности, основанных на параллелизме уровня команд, широкое распространение получили подходы, использующие параллелизм процессоров.

Основными являются:

- **Матричные компьютеры (Single Instruction-stream Multiple Data-stream, SIMD).** Применение большого числа (сотни и тысячи) сходных микропроцессоров, которые выполняют одну и ту же последовательность команд применительно к разным наборам данных;

- **Многопроцессорные компьютеры (Multiple Instruction-stream Multiple Data-stream, MIMD).** Система из нескольких параллельных процессоров, имеющих общую память. Поскольку каждый процессор может записывать информацию в любую часть памяти и считывать информацию из любой части памяти, их работа должна согласовываться программным обеспечением.

- **Многокомпьютерные системы (суперкомпьютеры).** Масштабируемые системы, реализуемые на кластерах вычислительных узлов общего назначения, связанных специализированной высокоскоростной сетью и работающих под управлением единой операционной системы.

Постоянная потребность выполнения все более сложных и объемных вычислений определяет необходимость повышения производительности микропроцессоров. Применение различных существующих технологий, а также возможных их сочетаний, позволят повысить скорость вычислений в десятки и сотни раз.

7. СПЕЦИАЛИЗИРОВАННЫЕ МИКРОПРОЦЕССОРЫ. ОСОБЕННОСТИ АРХИТЕКТУРЫ И СТРУКТУРЫ СПЕЦИАЛИЗИРОВАННЫХ МИКРОПРОЦЕССОРОВ

В данной теме рассмотрены следующие вопросы: процессоры цифровой обработки сигналов и сетевые процессоры, графические процессоры.

Производительность современных микропроцессоров растет за счет применения различных аппаратных и алгоритмических решений. Однако для выполнения ряда задач, требующих существенных ресурсов, современные компьютеры оснащаются дополнительными устройствами, имеющими свои специализированные микропроцессоры.

Процессоры цифровой обработки сигналов и сетевые процессоры

Одной из очень важных и одновременно ресурсоемких задач, реализуемых в системах связи различного назначения, системах управления и измерения, в медицинской, космической, военной технике и системах обеспечения безопасности, является цифровая обработка сигналов (ЦОС).

Общей схемой ЦОС предусматриваются аналого-цифровое преобразование сигнала, обработка микропроцессором, цифро-аналоговое преобразование.

Для выполнения ЦОС в современной технике используются специализированные микропроцессоры.

Цифровые процессоры обработки сигналов (ЦПОС), или цифровые сигнальные процессоры (ЦСП) (англ. DSP (Digital Signal Processor) — микропроцессоры, реализующие алгоритмы цифровой обработки сигналов и систем управления в реальном времени.

ЦСП применяются для выполнения ряда задач:

- Цифровая фильтрация — это селекция по частоте, то есть выделение полезных составляющих спектра сигнала, подавление помех.
- Спектроскопия — это совокупность методов обработки, позволяющих найти все частотные составляющие сигнала — не выделяя и не искажая их.
- Идентификация сигналов — это выделение сигналов на фоне помех.
- Модуляция и демодуляция.

В соответствии с решаемыми задачами формируются требования, на основе которых разрабатывается архитектура ЦПОС:

- для качественной обработки сигналов на высоких частотах процессор должен иметь высокую производительность, что, в свою очередь, определяет необходимость применения параллелизма;
- основными операциями, выполняемыми при ЦОС являются сложение и умножение с накоплением результата, соответственно эти операции должны быть реализованы аппаратно и иметь минимальное время выполнения;
- ЦПОС применяются как дополнительные устройства в компьютерах, либо в других технических устройствах, зачастую в виде встраиваемых систем. Это определяет требование к минимальному энергопотреблению.

Эти требования и определили особенности архитектуры современных ЦПОС:

- применение векторных вычислений. при этом операнды функции это не одно число (скаляр) а упорядоченный набор (вектор). Для выполнения операций с такими операндами в АЛУ используют соответствующие цепочки функциональных блоков (например, каскады сумматоров);
- вместо кэш-памяти используют банки очень быстрой памяти SRAM, однако ее высокая стоимость накладывает ограничения;
- относительно простые программы, управляющие работой ЦПОС, позволяют широко использовать параллелизм системы команд VLIW;
- большинство современных ЦПОС используют трехшинную гарвардскую архитектуру.

Гарвардская архитектура микропроцессора (разработана Говардом Эйкеном в конце 1930-х гг. в Гарвардском университете) предусматривает разделение инструкций и данных на разные физические устройства памяти (рис. 4), что позволяет организовать к ним доступ разными шинами. Это, в свою очередь, позволяет вести операции с данными и командами одновременно и независимо друг от друга.

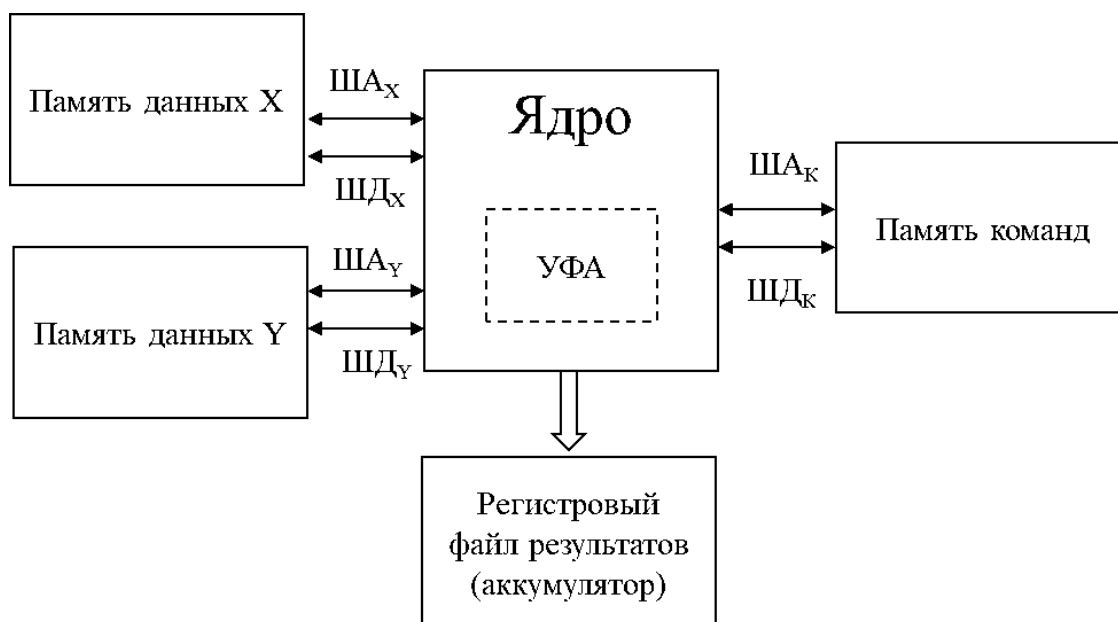


Рисунок 4. Трехшинная гарвардская архитектура

Архитектура ЦСП позволяет за один цикл выполнить:

1. выборку команды посредством шин адреса и данных программ;
2. выборку операндов посредством двух шин адреса и данных;
3. занесение операндов в регистры посредством двух шин данных;
4. выполнение команды в функциональном блоке (АЛУ);
5. запись результата в регистр-аккумулятор.

При этом обязательным элементом процессорного ядра является устройство формирования адресов.

Такая архитектура ЦПОС позволят получить им значительное преимущество над процессорами общего назначения при сравнении удельной производительности (GIPS/ядро/Вт). С учетом значительно более низкой стоимости ЦПОС эта разница становится еще более существенной.

Еще одним из видов микропроцессоров, применяющихся во всех современных компьютерных системах, являются сетевые процессоры.

Сетевые микропроцессоры — программно-управляемые микропроцессоры, архитектура которых оптимизирована для обеспечения обработки пакетов (packet processing) в сетевых устройствах (рис. 5).

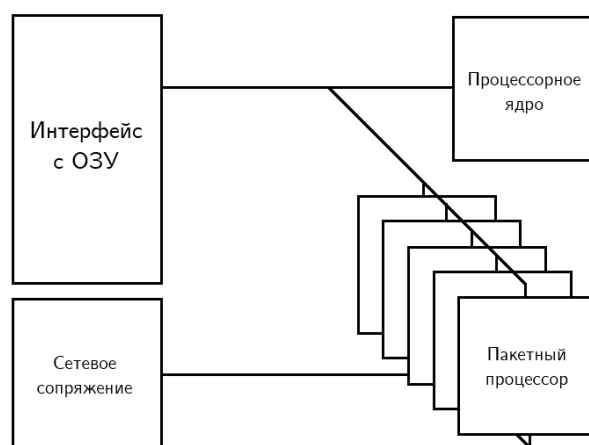


Рисунок 5. Структура сетевого процессора

Основные элементы сетевого процессора и их функции:

1. ядро координирует работу всех остальных блоков сетевого процессора, обрабатывает таблицы маршрутизации и качества обслуживания (QoS), а также обрабатывает пакеты, связанные с запросами по управлению сетью и обновлением маршрутных таблиц;

2. пакетный процессор (packet processor, PP), выполняет следующие функции:

- а. получает с сетевого интерфейса пакеты, ячейки или кадры и записывает их полностью либо частично в ОЗУ;
- б. определяет порядок обработки пакетов;
- с. обрабатывает пакеты (определение типа пакета, проверка политики (policing), маршрутизация, модификация и QoS);
- д. управляет перенаправлением или удержанием пакета;
- е. отправляет пакет;

3. блок высокоскоростного интерфейса физического уровня (их может быть несколько для поддержки сетей разных типов) обеспечивает связь сетью.

Ведущие компании в области пакетной коммутации, такие как Cisco или Alcatel, используют не только оригинальное оборудование — специализированные сетевые компьютеры (маршрутизаторы, коммутаторы), широко эксплуатирующее параллелизм уровня процессоров, но и собственные операционные среды для достижения максимальной эффективности своего оборудования.

Графические процессоры

Современные программы в подавляющем большинстве используют графический интерфейс, эффекты и видеоролики. Обработка трехмерной

видеографики является весьма ресурсоемкой задачей. Для того, чтобы разгрузить центральный процессор, обработку видео выполняют в отдельном графическом ядре, либо на отдельной плате расширения возможностей компьютера — видеокарте, основу которой составляет графический процессор.

Графический процессор (англ. graphics processing unit, GPU) — программно-управляемый микропроцессор, предназначенный для обработки графической информации.

Алгоритм работы графического процессора строится на использовании конвейерной обработки данных.

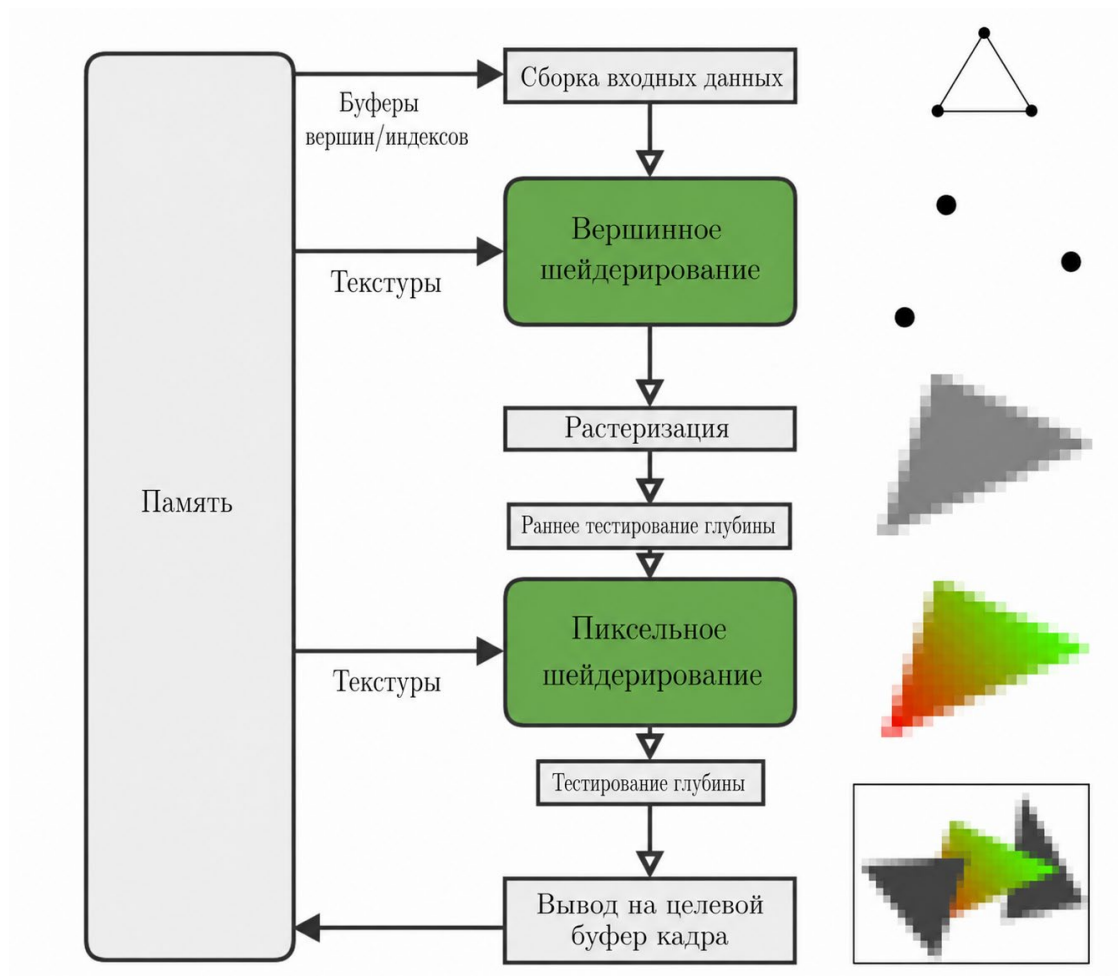


Рисунок 6. Конвейер графического процессора

На рис. 6 представлена схема работы графического конвейера.

Работа представленного на рисунке 6 конвейера строится по следующей схеме:

1. загрузка массивов вершин (VB, содержит список вершин со связанными свойствами) и индексов (IB, содержит список вершин, объединенных в треугольники) из памяти;

2. начальная сборка — формирование треугольников на основе данных VB и IB;

3. обработка (шейдинг) вершин — изменение параметров (текстура, координаты, цвет, нормаль) вершин с помощью вершинного шейдера (vertex shader); Шейдер — набор команд GPU, предназначенный для обработки элементов изображения.

4. растеризация — заполнение треугольников пикселями, интерполяция цвета пикселей по цвету вершин. Например: если одна вершина треугольника имеет чёрный цвет, а другая — белый, то пиксель, растеризированный посередине между ними получит интерполированный серый цвет;

5. тест глубины — для сокращения объема вычислений отбрасываются пиксели, находящиеся <<глубже>> тех, что уже имеются в выходном изображении;

6. шейдинг пикселей — изменение цвета, освещения и других параметров элементов изображения. Пикселей очень много (изображение с разрешением 1080p содержит больше двух миллионов), и каждый из них нужно обработать хотя бы раз, поэтому обычно видеопроцессор тратит на пиксельный шейдинг большую часть времени;

7. вывод полученного целевого изображения.

Таким образом очевидно, что графический процессор практически постоянно выполняет однотипные операции (в первую очередь, шейдинг пикселей). Для повышения производительности архитектура графических процессоров строится на основе параллелизма уровня процессоров SIMD. при этом большое число (сотни и тысячи) сходных микропроцессорных ядер выполняют одну и ту же последовательность команд применительно к разным наборам данных. При выполнении таких задач производительность GPU может достигать десятков GIPS.

Микропроцессорные ядра могут быть объединены в потоковые мультипроцессоры, имеющие устройство диспетчеризации команд. Размещаемые на плате видеокарты, графические процессоры обращаются в первую очередь к памяти карты, в том числе и кэш-памяти нескольких уровней. Работа видеокарты, и, как следствие, всех ее графических процессоров, выполняется под управлением центрального процессора, который на основе команд, включенных в исполняемую программу,

передает данные о том, что и каким образом необходимо отрисовать и вывести на экран.

Применение специализированных процессоров для выполнения специфических задач позволяет с одной стороны, разгрузить центральный процессор, тем самым повысить общую производительность компьютерной системы. с другой стороны, на основе специальных процессоров могут строиться системы, предназначенные для особых задач, например устройства управления сетевым трафиком-маршрутизаторы, при этом обеспечивающие достаточно высокую производительность при более низкой стоимости.

8. УСТРОЙСТВА ХРАНЕНИЯ ДАННЫХ В КОМПЬЮТЕРНЫХ СИСТЕМАХ

В данной теме рассмотрены следующие вопросы: пятиуровневая схема организации памяти компьютерных систем, оперативная и кэш-память компьютеров и энергонезависимые накопители данных.

Современные процессоры обладают существенной производительностью. Они способны обрабатывать огромные объемы данных за короткие промежутки времени. Однако обрабатываемые данные должны поступать в тракты процессора извне, и после обработки сохранить. Эти задачи выполняет подсистема хранения данных — память компьютера.

Пятиуровневая схема организации памяти компьютерных систем

Подсистема хранения данных (память) — подсистема компьютера, предназначенная для хранения информации, необходимой для его работы (адреса, инструкции, обрабатываемые данные). Память характеризуется двумя основными показателями: объемом данных, которые она может хранить, и временем доступа, т. е. временем, которое будет затрачено на чтение данных процессором из устройства памяти.

Основной единицей хранения данных в памяти является двоичный разряд, который называется битом. Бит может содержать 0 или 1. Это самая маленькая единица памяти. В системе с двоичным кодированием данных требуется различать всего два уровня физической величины (напряжения или тока), следовательно, это самый надежный метод кодирования цифровой информации.

Объем памяти, как и хранимых в ней данных, измеряется в следующих основных единицах.

Бит — двоичный разряд 0 или 1.

Байт — восемь бит, $bbbbbbbb$, где $b \in [0, 1]$.

Слово (строка) — 8 байт в x64 системах. Кроме того, используются различные производные единицы, например мегабит или гигабайт.

Все устройства хранения данных могут быть отнесены к двум классам:

- энергозависимые устройства, требующие для хранения данных наличие питающих токов. Это регистры процессора, оперативная и кэш-память.

- энергонезависимые устройства, способные сохранять записанные в них данные при отключении от источника питания. Это твердотельные и ленточные накопители, магнитные и оптические диски.

В современных компьютерных системах применяется память, разделенная на пять основных уровней (рисунок 7).

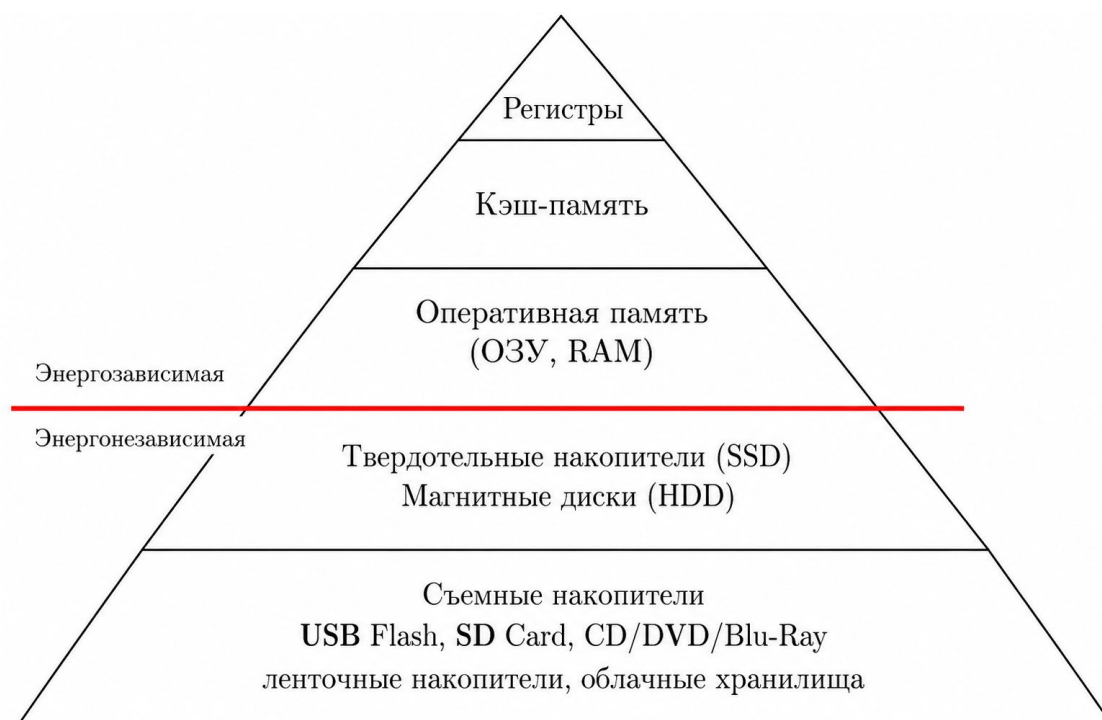


Рисунок 7. Пятиуровневая схема организации памяти

На вершине <<пирамиды>> находятся устройства памяти, размещаемые непосредственно в микросхеме процессора — регистры памяти. Они характеризуются минимальным временем доступа, так как располагаются непосредственно у тракта данных процессорного ядра, и значения хранимых данных постоянно выводятся на их выходы. Следовательно, задержка в чтении данных определяется только скоростью распространения сигналов в шинах процессора, и практически стремится к долям такта. Для процессора с тактовой частотой 1 ГГц это время составляет 1 нс. Однако объем данных, хранимых в регистрах современных процессоров исчисляется всего несколькими килобайтами.

Второй уровень занимает кэш-память, которая при времени доступа 2–5 нс может хранить уже десятки мегабайт данных.

Третий уровень — уровень оперативной памяти (ОЗУ). Ее объемы достигают сотен гигабайт при времени доступа 15–20 нс.

Устройства четвертого уровня позволяют хранить до 15 Тбайт, время доступа для SSD — до 200 нс, для HDD — от 1 до 20 мс.

Самый нижний, пятый уровень, занимают съемные (внешние) устройства, время доступа для которых может исчисляться секундами.

Таким образом, по мере продвижения по уровням схемы с ростом объема памяти увеличивается время доступа.

Память состоит из ячеек, каждая из которых может хранить некоторый объем данных. Каждая ячейка имеет номер, который называется адресом. Адресация ячеек строго последовательная и непрерывная. По адресу программы могут ссылаться на определенную ячейку.

Если память содержит n ячеек, они будут иметь адреса от 0 до $n - 1$. Все ячейки памяти содержат одинаковое число битов. Если ячейка состоит из k бит, она может содержать любую из 2^k комбинаций.

Память строится в виде двумерных матриц (рис. 8). Каждая строка матрицы — слово в памяти, а номер строки — ее адрес. Содержимое записывается и считывается из памяти целыми словами.

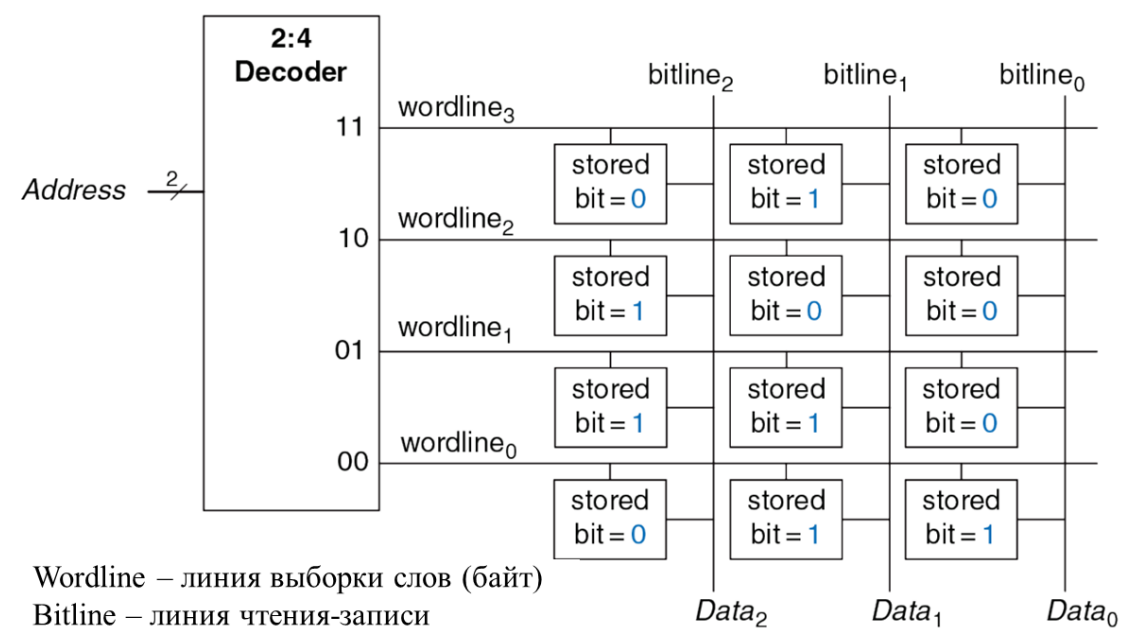


Рисунок 8. Матрица памяти 4 × 3 (глубина 4 слова, ширина 3 бит) с одним портом чтения данных

Матрицы памяти представляют собой набор запоминающих элементов, каждый из которых хранит один бит данных. Каждый запоминающий элемент соединен с wordline (линией выборки слов) и bitline (линией чтения-записи).

При любой комбинации адресных битов активируется только одна линия выборки слов и тем самым разрешается доступ к элементам

соответствующей строки. Когда линия выборки слов некоторой строки активна, элементы этой строки могут выдавать данные на линии чтения-записи или принимать данные с этих линий. В противном случае запоминающие элементы отсоединены от линии чтения-записи.

При чтении битов линия чтения-записи вначале находится в отключенном состоянии. Затем включается линия выборки слов, и запоминающие элементы выдают хранимое значение на линию чтения-записи.

При записи информации в запоминающий элемент сигнал на линию чтения-записи поступает со специального усилителя, имеющего небольшое выходное сопротивление. Затем включается линия выборки слов и линии чтения-записи соединяются с запоминающими элементами. Сигнал с линии чтения-записи перезаписывает содержимое запоминающего элемента, и в нее записывается новая информация.

Память всех типов имеет один или несколько портов (ports).

Порт памяти — набор шин и вентилях, предназначенных для записи/чтения данных по одному адресу.

В предыдущем примере была рассмотрена однопортовая память. Многопортовая память (рисунок 9) обеспечивает одновременный доступ к содержимому по нескольким адресам.

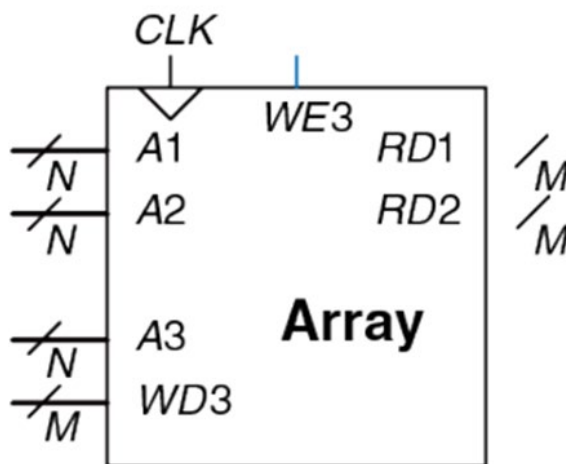


Рисунок 9. Матрица памяти с 3 портами

На рисунке показана трехпортовая память с двумя портами для чтения и одним для записи. Порт 1 считывает данные, которые хранятся по адресу A1, и выдает их на выход RD1. Порт 2 выдает информацию, хранимую по адресу A2, на выход RD2. Порт 3 позволяет записать данные, поданные на вход WD3, в элемент по адресу A3, запись информации

осуществляется по переднему фронту тактового импульса при активном сигнале WE3.

Оперативная и кэш-память компьютеров

Основными классами ОЗУ являются динамическое оперативное запоминающее устройство (динамическая память, DRAM) и статическое оперативное запоминающее устройство (статическая память, SRAM).

Динамическая память сохраняет данные в виде заряда конденсаторов, а статическая — в виде состояния бистабильной схемы, состоящей из двух перекрестно соединенных инверторов.

В динамическом ОЗУ (DRAM) битовым значениям соответствует наличие и отсутствие заряда конденсатора. N-канальный МОП-транзистор является ключом, который может подключить конденсатор к линии чтения-записи или отключить его. Когда линия выборки слов активна, транзистор включается и хранимые биты передаются на линию чтения-записи или наоборот, происходит запись новой информации в элемент.

Чтение уничтожает данные, которые хранились в конденсаторе, поэтому после каждого чтения данные должны быть восстановлены (перезаписаны). Даже если из динамического ОЗУ не нужно считывать данные, из-за саморазряда конденсаторов они должны регенерироваться (считываться и перезаписываться) каждые несколько миллисекунд.

Статическое ОЗУ (SRAM) называется так, поскольку в нем отсутствует необходимость регенерации хранимых данных. Данные хранятся в бистабильной схеме, состоящей из двух перекрестно соединенных инверторов. Когда линия выборки слов активна, оба *n*-канальных МОП-транзистора открываются и данные могут быть записаны в элемент или считаны из него.

Регистры, статические и динамические ОЗУ являются энергозависимыми запоминающими устройствами, но они различаются временными характеристиками и площадью кристалла, необходимой для хранения одного бита. Данные, хранимые в триггере, непосредственно доступны на его выходе. Но схема триггера состоит, по крайней мере, из 20-ти транзисторов. В общем случае, чем больше транзисторов используется в приборе, тем большую площадь он занимает, потребляет больше энергии и стоит дороже.

Так как в динамическом ОЗУ используется наименьшее количество транзисторов, то и цена его минимальная, что обеспечило таким микросхемам доминирующее положение на современном рынке.

Однако задержка чтения данных в динамическом ОЗУ больше, чем в статическом, так как определяется относительно медленной передачей заряда из конденсатора на линию считывания-записи. Из-за необходимости выполнения периодической регенерации и регенерации после чтения динамическое ОЗУ имеет меньшую пропускную способность, чем статическое.

Современная разновидность динамического ОЗУ, синхронное динамическое ОЗУ с удвоенной скоростью обмена (DDR SDRAM или коротко DDR), была разработана для преодоления этой проблемы. В синхронном динамическом ОЗУ используется тактовый сигнал для конвейеризации доступа к памяти. В синхронном динамическом ОЗУ с удвоенной скоростью обмена передача данных происходит как по переднему, так и по заднему фронту тактового импульса, что удваивает пропускную способность при заданной частоте тактового сигнала. В современном стандарте DDR5 используются частоты до 1300 МГц.

В современных компьютерных системах ОЗУ может быть распаяно непосредственно на системной плате, либо устанавливаться в ее специальные разъемы в виде отдельных модулей. Современные модули DDR4 выполняются в формате DIMM, с двусторонним расположением 288 контактов и функцией обнаружения и исправления ошибок.

Несмотря на то что ОЗУ способно отдавать данные в процессор с огромной скоростью, оно все же имеет время доступа, равное нескольким десяткам тактов. Для того чтобы сократить длительность простоев процессоров, разработчиками внедрен еще один уровень памяти — кэш память.

Кэш-память — это набор блоков SRAM, расположенных внутри микросхемы процессора, они предназначены для сокращения времени ожидания данных процессором за счет хранения наиболее часто используемых данных и/или последовательности команд. Схема кэш-памяти изображена на рис. 10.

Кэш-память имеет многоуровневую структуру, и включает:

L0 — кэш уровня 0, хранение микроопераций, непосредственно записываемых в регистры;

L1 Ins — кэш уровня 1, хранение cache-tag команд;

L1 D — кэш уровня 1, хранение cache-tag операндов;

L2 — хранение команд и операндов для одного ядра;

L3 — хранение команд и операндов, доступен всем ядрам.

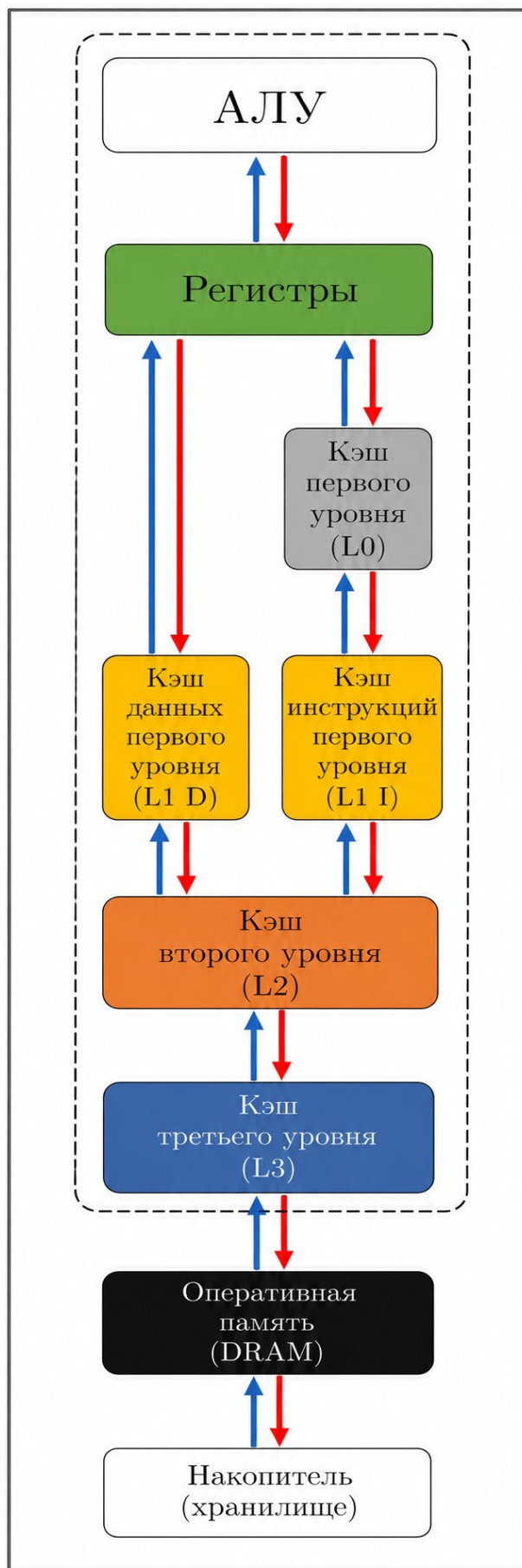


Рисунок 10. Кэш-память

Кэш повышает производительность, ускоряя передачу данных в ядра процессора и храня поблизости копию часто используемых инструкций и данных. Хранящаяся в кэше информация разделена на две части: сами данные и адрес места, где они изначально располагаются в системной памяти — такой адрес называется тег кэша (cache tag).

Когда процессор выполняет операцию чтения или записи данных из/в память, то он начинает с проверки тегов в кэше Level 1. Если нужные данные там есть (произошло кэш-попадание (cache hit)), то доступ к этим данным выполняется почти сразу же. Промах кэша (cache miss) возникает, если требуемый тег не найден на самом нижнем уровне кэша. В кэше L1 создаётся новый тег, и выполняется поиск в других уровнях кэша, а при необходимости и на остальных уровнях памяти, данных для этого тега.

Кэш-память позволяет повысить производительность процессора, но вместе с тем значительно усложняет его конструкцию и, как следствие, повышает стоимость. Тем не менее, в современных процессорах суммарный объем кэш-памяти достигает десятков мегабайт, хотя большая часть ее и приходится на L3.

Энергонезависимые накопители данных

Данные, хранящиеся в регистрах, ОЗУ и кэш-памяти, исчезают при отключении питания компьютера. Для долговременного хранения данных используются постоянные запоминающие устройства. На четвертом уровне находятся жесткие магнитные диски и твердотельные накопители, использующие технологию flash-памяти.

Накопитель на жестких магнитных дисках (НЖМД, hard disk drive (HDD)) — энергонезависимое перезаписываемое запоминающее устройство, использующее технологию магнитной записи.

Магнитная запись — изменение вектора намагниченности доменов в ферромагнетике (в НЖМД это диоксид хрома) под действием магнитного поля.

В НЖМД для хранения информации используют диски, покрытые ферромагнитным веществом, на которые с помощью головок записывается и считывается информация. Диски установлены на ось (шпиндель), которая вращается с помощью электродвигателя.

Диски современных накопителей позволяют записывать до 500 Гбайт данных на одну сторону. Применяются методы продольной и перпендикулярной записи. Для повышения плотности данных,

записываемых на единицу площади создана технология черепичной записи данных.

Основные характеристики НЖМД:

- интерфейс;
- емкость;
- форм-фактор (типоразмер);
- скорость вращения шпинделя.

Интерфейсом жесткого диска (как и других съемных устройств) называют набор шин, разъемов для их подключения на устройстве и системной плате, а также протокол обмена данными между этими устройствами. Основные интерфейсы, используемые в современных жестких дисках, и их характеристики приведены в табл. 4.

Таблица 4. Интерфейсы жестких дисков

Название	Скорость обмена, Гбит/с	Питание по кабелю	Количество дисков на канал
SCSI	2.5	Нет	16
SAS	2.4	Нет	16384
SATA	6	Нет	1
USB 3.2	4.8	Да	1

Емкость современных жестких дисков, как сказано выше, достигает нескольких ТБайт.

Форм-фактор диска, в первую очередь, зависит от компьютера, в котором планируется применять НЖМД. В настоящее время наибольшее распространение получили диски размеров 3,5 и 2,5 дюйма.

Частота вращения шпинделя — показатель, характеризующий производительность диска при последовательном чтении или записи информации. Чем чаще вращается мотор, тем быстрее пролетают под головкой секторы дисков. Основные популярные значения — 5400, 7200, 10000 и 15000 оборотов в минуту.

В современных компьютерах, особенно в ноутбуках и смартфонах, вместо НЖМД чаще всего применяют твердотельные накопители, карты памяти, использующие технологии flash-памяти.

Устройства с flash-памятью — класс твердотельных запоминающих устройств (SSD, USB, SD) на основе технологии полупроводниковых перепрограммируемых микросхем памяти (EEPROM).

В основе технологии лежит применение КМОП-транзисторов с двумя изолированными затворами: управляющим (control) и плавающим (floating). Важной особенностью последнего является способность удерживать электроны. При записи информации между истоком и стоком транзистора, вследствие воздействия положительного поля на управляющем затворе, создается канал — поток электронов. Некоторые из электронов, благодаря наличию большей энергии, преодолевают слой изолятора и попадают на плавающий затвор. На нем они могут храниться в течение нескольких лет. Определенный диапазон количества электронов (заряда) на плавающем затворе соответствует логической единице, а все, что больше его, — нулю. При чтении эти состояния распознаются путем измерения порогового напряжения транзистора. Для стирания информации на управляющий затвор подается высокое отрицательное напряжение, и электроны с плавающего затвора переходят (туннелируют) на исток.

Транзисторы могут подключаться по двум основным схемам NOR и NAND. В NOR возможно обращение к каждой ячейке. При этом увеличивается число подводимых проводников, сокращается число ячеек на чипе и возрастает стоимость чипа. Однако преимуществом является скорость чтения и записи. Поэтому чипы с такой организацией используются в качестве хранителей программного кода (BIOS, RAM смартфонов), иногда реализовываются в виде интегрированных решений (в системах на чипе).

В NAND схеме ячейки (транзисторы) группируются в небольшие блоки, и запись/чтение данных выполняется сразу блоком. Это дает такой схеме преимущества в скорости последовательной записи/чтения блоков, в плотности ячеек на чипе, что позволяет снизить ее стоимость. Однако с другой стороны NAND значительно проигрывает в операциях с произвольным доступом и не позволяет напрямую работать с байтами информации. Чипы с такой схемой используются в SSD/USB/SD накопителях.

В структуре flash-памяти для хранения 1 бита информации задействуется только один элемент (транзистор), в то время как в энергозависимых типах памяти для этого требуется несколько транзисторов и конденсатор. Это позволяет существенно уменьшить размеры выпускаемых микросхем, упростить технологический процесс, а следовательно, и снизить себестоимость. Современные технологии позволили создать плавающие затворы, на которых могут хранить

несколько уровней заряда. Таким образом, созданы ячейки, позволяющие хранить более 1 бит. Технология MLC - 2бит, TLC — 3 бит, QLC — 4 бит. Это позволяет создавать чипы с большим объемом памяти и снижать энергопотребление, а также стоимость такой памяти. Однако, с ростом плотности ячеек увеличивается время доступа к данным, и, что может быть весьма критичным, надежность таких чипов.

Помимо съемных накопителей, использующих flash-память, в составе компьютерных систем применяют накопители на оптических компакт-дисках.

Накопители на компакт-дисках (CD-ROM, CD-R, CD-RW, DVD-R/RW, Blu-Ray) — класс запоминающих устройств, использующих для хранения данных технологию записи <<холодным лазером>> на двуслойный материал.

В основе технологии лежит физическое свойство применяемого в качестве носителя данных сплава серебра, индия, сурьмы и теллура при нагреве выше температуры плавления переходить из кристаллического состояния в аморфное. при этом у нагретых участков меняется способность отражать свет. Данный переход обратим. с развитием технологии уменьшалась длина волны лазера, и, как следствие, росла плотность записи данных. Диски Blu-Ray позволяют хранить до 60 Гбайт на одной пластине, при этом для хранения данных не требуется энергии, а сроки «жизни» дисков составляют несколько сотен лет. Недостатком же является относительно большое время доступа к хранимым данным и подверженность поверхности диска механическому износу и повреждениям.

Еще одним классом съемных устройств для хранения данных являются ленточные накопители — класс запоминающих устройств, использующих для хранения данных магнитную ленту.

Несмотря на то что подобные устройства начали использоваться практически одновременно с созданием компьютеров, и в настоящее время они не применяются в ПК, однако в корпоративном секторе для хранения архивных данных лента применяются очень широко. Это обусловлено тем, что по плотности записи данных на один носитель ленточные картриджи не имеют равных. В современных системах емкость одного картриджа может достигать 580 Тбайт, при этом информация может храниться несколько десятков лет без потери качества. Однако, время доступа к данным на ленте с учетом времени, необходимого для поиска нужного участка ленты, может достигать десятков секунд.

Развитие сетей передачи данных и операционных систем компьютеров повлекло создание такого класса устройств хранения, как сетевые (облачные) хранилища. В основе лежат распределенные файловые системы, и технологии NAS (network attached storage) и SAN (storage area network). Эти технологии позволяют организовать хранение данных не на локальных машинах, а на специальных устройствах — системах хранения данных, обеспечивающих повышенную надежность (за счет RAID) и достаточно высокую производительность.

Современные компьютерные системы включают в себя целый ряд различных устройств памяти различного объема. Многоуровневая система организации памяти позволяет разрешать противоречие между требуемым большим объемом памяти для хранения данных с одной стороны и с желанием как можно быстрее получать доступ к этим данным для увеличения производительности компьютера с другой стороны.

9. УСТРОЙСТВА ВВОДА И ВЫВОДА ДАННЫХ КОМПЬЮТЕРНЫХ СИСТЕМ

В данной теме рассмотрены следующие вопросы: шины и разъемы для подключения I/O устройств, устройства ввода данных компьютерных систем и устройства вывода данных компьютерных систем.

Основной задачей компьютерных систем является обработка данных и выполнение вычислений. Несмотря на появление целого класса устройств, получающих и передающих данных на межмашинном уровне, основным источником информации для выполнения вычислений, а также потребителем результатов ее обработки продолжает оставаться оператор компьютера, человек. Для взаимодействия с компьютером оператор использует самые различные устройства ввода и вывода данных.

Шины и разъемы для подключения I/O устройств

Работой устройств ввода/вывода в компьютерных системах ЦП управляет не на прямую, а с применением дополнительных цифровых устройств — контроллеров ввода/вывода, реализующих логику взаимодействия процессора и устройства. Конструктивно контроллеры выполняются в составе схемы «южного моста» чипсета системной платы. Контроллеры соединяются с разъемами, к которым подключаются I/O устройства, шинами.

Шины системной платы — физические (электрические) соединительные линии для передачи данных между устройствами, размещенными на системной плате, а также внешними подключаемыми устройствами.

За все время развития компьютерной техники существовало большое количество различных типов шин и разъемов. В различные периоды применялись стандарты S100, PCI, AGP, IDE и другие.

В настоящее время де-факто во всех компьютерных системах применяются три основных стандарта шин и соответствующих им разъемов.

PCI-Express (PCIe) — стандарт двунаправленной шины с топологией «точка-точка». Данные передаются в дуплексном режиме, по паре проводов. Две пары составляют одну линию, в терминах стандарта это x1 link. Для организации параллельной передачи данных по нескольким линиям применяются коммутаторы. Предусмотрена возможность использования 1/2/4/8/16 линий.

Основные показатели различных версий шины данного стандарта представлены в табл. 5.

Таблица 5. Основные показатели различных версий шины стандарта PCIe

Версия стандарта	Производительность, ГБ/с	Кодирование	x1	x4	x16
PCIe 2.0	5	8b/10b	0.5	2	8
3.0	8	128b/130b	1	4	16
4.0	16	128b/130b	2	8	32
5.0	32	128b/130b	4	16	53

Шина использует разъемы для подключения внешних устройств, размещаемых на системной плате. Предусмотрены разъемы трех различных форм-факторов: с 1, 4 и 16 линиями.

USB — стандарт универсальной последовательной шины. Использует принцип активного дерева с обязательным наличием концентратора (хаба), выполняющего электрическое согласование, одним ведущим устройством (host) и до 127 подключаемых подчиненных устройств. Обеспечивается питание устройств и <<горячее>> подключение. USB-вилки показаны на рис. 11.

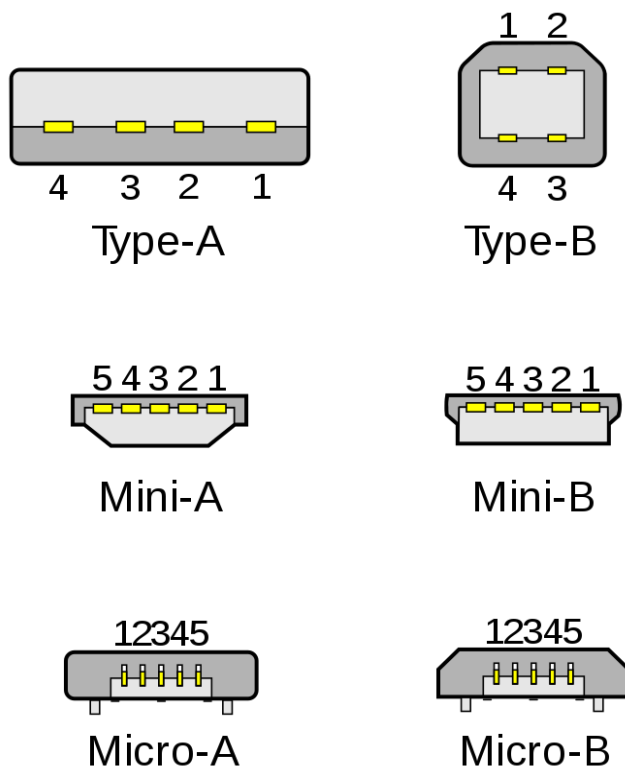


Рисунок 11. Стандартные, Mini- и Micro-USB вилки

Со времени появления стандарта сменились несколько спецификаций от Low-Speed (USB 1.0) до SuperSpeed++ 20Gbps (USB 3.2 Gen 2x2). Скорость передачи данных возростала от 1,5 Мбит/с до 20 Гбит/с. Применяются целый ряд разъемов, различных форм-факторов и распиновки.

SATA — стандарт последовательной шины передачи данных для устройств энергонезависимой памяти, таких как НЖМД, твердотельные накопители SSD, приводы CD/DVD. Применяется топология точка-точка, дуплексный обмен по двум витым парам. Возможно «горячее» подключение. Для устройства требуется дополнительное питание, подаваемое по отдельному кабелю.

Устройства ввода данных компьютерных систем

Устройства ввода данных предназначены для преобразования неэлектрических величин (давление, температуру, освещенность, положение в пространстве) в цифровые электрические сигналы для их последующей обработки процессором ПК.

Основным устройством ввода в современных компьютерных системах являются клавиатуры.

Клавиатура — клавишное устройство управления компьютером, позволяющее вводить алфавитно-цифровые (знаковые) данные, а также команды управления компьютерной системой. Представляет собой набор кнопок (более 100), в зависимости от назначения разделенных по нескольким группам (буквенно-цифровые, кнопки управления, кнопки-модификаторы, функциональные кнопки и т. д.) и расположенных в определенном порядке. Основные группы кнопок: текстовые, функциональные, управления курсором, дополнительные цифровые.

Наибольшее распространение в настоящее время получили клавиатуры, построенные с применением мембранной технологии. В таком устройстве пластиковая кнопка опирается на выступающий резиновый колпачок. Под клавишами размещены три слоя пластиковых мембран. на двух находятся электрические контакты, третья служит для обеспечения зазора. При нажатии на кнопку колпачок сминается и кнопка замыкает контакты, расположенные на пересечении токопроводящих линий, образуя матрицу. Замыкание контакта между линией строки и линией столбца позволяет однозначно идентифицировать нажатую клавишу, при этом контроллер клавиатуры в соответствии с полученным сигналом передает в процессор скан-код. Далее процессор в соответствии с логикой обработки прерываний выполняет соответствующие

инструкции. Конструкция мембранных клавиатур весьма проста и обеспечивает снижение уровня шума, при этом современные устройства достаточно долговечны и имеют относительно низкую стоимость.

Однако мембранные клавиатуры имеют и ряд недостатков, для компенсации которых можно применять клавиатуры с механической системой нажатия. В такой клавиатуре клавиша подпирается пружиной, контакты разомкнуты пластиковым размыкателем. При нажатии на кнопку механизм опускается, размыкатель смещается и происходит замыкание контактов. Срабатывание механизма происходит до полного нажатия, при этом ощущается отклик и слышен отчетливый звук. Это позволяет повысить эргономику, сокращает необходимые усилия при наборе больших объемов текстов или активной игре. Однако уровень шума может быть существенной проблемой, что определяет необходимость подбирать тип используемой клавиатуры для решения конкретных задач.

Появление и развитие программного обеспечения с графическим пользовательским интерфейсом стало возможно благодаря появлению такого устройства ввода как компьютерная мышь. Это манипуляторное (т. е. перемещаемое в пространстве) указательное устройство, предназначенное для взаимодействия с элементами графического интерфейса.

Первые манипуляторы для определения перемещения использовали различные механические устройства. Современные мыши оптические, используют отражение света от рабочей поверхности. При этом чувствительный оптический датчик фиксирует изменение освещенности участков поверхности под мышью и на основании обработки изображения формирует сигнал о перемещении устройства.

Наиболее значимым показателем, характеризующим качество мыши, является расстояние (количество пикселей), на которое сдвинется курсор на экране, если мышь физически передвинется на один дюйм — DPI.

Еще один тип устройств ввода, получивших широкое распространение, особенно в носимых и мобильных компьютерах — сенсорные устройства. Представляют из себя указательное устройство ввода, предназначенное для взаимодействия с элементами интерфейса компьютера путем нажатия одним или несколькими пальцами на его поверхность. в настоящее время применяются сенсорные панели, использующие две технологии: резистивную и емкостную.

Резистивный сенсор (рис. 12) представляет собой две пластины, на соприкасающиеся поверхности которых нанесен проводящий материал,

при этом без нажатия пластины не соприкасаются, между ними диэлектрик. По краям панели расположены электроды, на которые попеременно подается напряжение.

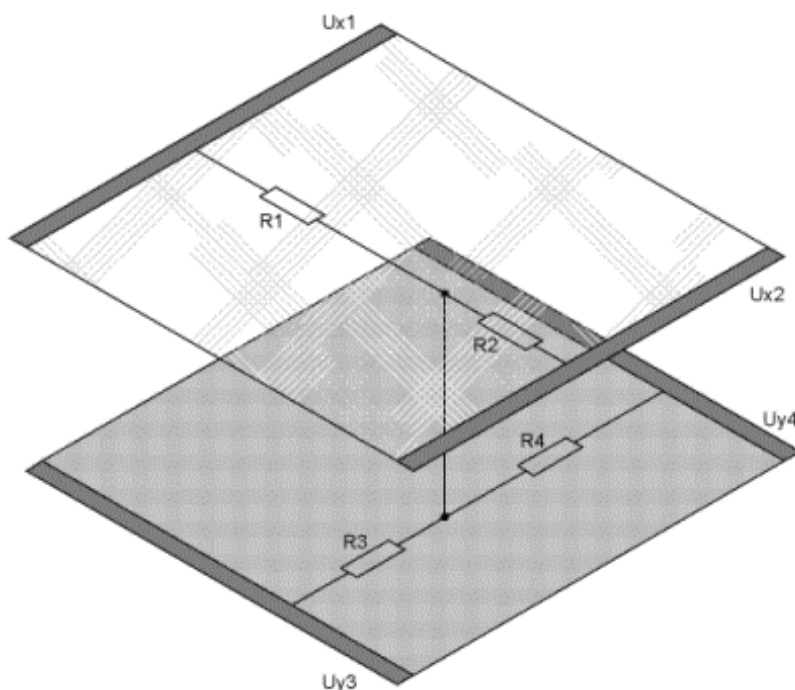


Рисунок 12. Резистивный сенсорный дисплей

В общих чертах алгоритм получения координат таков:

1. на верхний электрод подаётся напряжение +5В, нижний заземляется. Левый с правым соединяются накоротко и проверяется напряжение на них. Это напряжение соответствует Y-координате экрана.
2. Аналогично на левый и правый электрод подаётся +5В, и <<земля>> с верхнего и нижнего считывается как X-координата.

Цикл повторяется с частотой, обеспечивающей малую задержку между нажатием и откликом интерфейса устройства.

Основным преимуществом резистивных сенсорных устройств является их способность к работе при взаимодействии с любыми предметами и устойчивость к загрязнению поверхности. Однако они имеют и ряд недостатков, что привело к появлению емкостных сенсорных устройств.

Эти устройства для определения точки нажатия используют принцип измерения емкости конденсатора, образованного двумя пластинами. На внутреннюю сторону экрана нанесена сетка электродов, которые образуют конденсаторы. При прикосновении к конденсатору пальцем на него начинает утекать электрический заряд, при этом емкость

конденсатора изменятся. Контроллер сенсорной панели поочередно подает ток на контакты пластин, соединенных в линии строк и столбцов, измеряя напряжение. При нажатии это напряжение изменится, и будут определены координаты точки нажатия.

Такие сенсорные устройства по точности определения координат не уступают резистивным, но могут быть прозрачными (до 90 % и более светопропускание), долговечными, и могут определять нажатие в нескольких точках одновременно. Данные обстоятельства определили их широкое распространение в смартфонах и планшетах. Недостаток — невозможность для нажатия использования предметов, не проводящих ток.

Устройства вывода данных компьютерных систем

Устройства вывода данных предназначены для преобразования цифровых электрических сигналов, выдаваемых из ПК, в сигналы иной физической сущности (световые, тепловые, механические) для передачи пользователю или другому устройству.

Наиболее распространенными устройствами вывода являются компьютерные дисплеи. Дисплей — устройство вывода, предназначенное для отображения графической информации (растрового изображения), сформированного компьютером.

В современных дисплеях применяются технологии жидких кристаллов (TN или IPS) и органических светодиодов.

Жидкокристаллические дисплеи основаны на эффекте, открытом советским ученым Фредериксом. Он исследовал поведение жидких кристаллов — веществ, имеющих свойства жидкости и кристаллов. к примеру, молекулы такого вещества при приложении электрического поля ориентируются по направлению его силовых линий. Таким образом можно управлять светопропусканием ячейки, содержащей жидкие кристаллы.

Точка экрана, отвечающая за формирование изображения, называется пикселем, и состоит из:

- прозрачных электродов в количестве двух штук;
- прослойки молекул активного вещества между электродами (это и есть ЖК);
- поляризаторы, оптические оси которых перпендикулярны друг-другу (зависит от конструкции);
- цветные светофильтры.

Если между фильтрами не было бы ЖК, то свет от источника, проходя через первый фильтр и поляризуясь в одном направлении, полностью задерживался бы вторым, из-за того, что его оптическая ось перпендикулярна оси первого фильтра. Поэтому, как бы мы не светили на одну сторону матрицы, со второй стороны она остается черной.

Twisted Nematic (TN) технология

Изначальное расположение молекулы — в виде четверть оборотной спирали. То есть свет от первого фильтра преломляется так, что проходя вдоль кристалла он попадает на второй фильтр в соответствии с его оптической осью. Следовательно, в спокойном состоянии такая ячейка всегда прозрачна.

Воздействуя на электроды напряжением можно изменять угол поворота кристалла вплоть до его полного распрямления, при котором свет через кристалл пройдет без преломления. А так как он уже был поляризован первым фильтром, то второй его полностью задержит, и ячейка будет черной. Изменение величины напряжения изменяет угол поворота, а соответственно и степень прозрачности.

In-Phase Switching (IPS) технология. Дальнейшее развитие идеи TN

Если к пикселю IPS не приложено напряжение, молекулы жидких кристаллов не поворачиваются. Второй фильтр всегда повернут перпендикулярно первому, и свет через него не проходит. Поэтому отображение чёрного цвета близко к идеалу. При приложении напряжения молекулы жидких кристаллов поворачиваются перпендикулярно своему начальному положению и пропускают свет.

В качестве источников света в дисплеях TN и IPS служат люминесцентные лампы, либо светоизлучающие диоды. Последние также являются основой для создания другого типа дисплеев — LED, в результате развития технологии современные устройства построены на AMOLED — активных матрицах органических светоизлучающих диодов.

Каждая точка изображения формируется с помощью трех светодиодов, питанием которых управляет матрица транзисторов. при этом для формирования цветного изображения могут применяться различные схемы с использованием цветных излучателей или светофильтров. Созданные по такой технологии дисплеи обладают рядом преимуществ по сравнению с ЖК дисплеями:

- меньшая толщина дисплея;
- более высокая яркость;

- лучшая цветопередача;
- меньшее энергопотребление.

Однако наряду с преимуществами, имеются и недостатки, такие как меньший срок службы и большая стоимость.

Технология постоянно развивается, в том числе в направлении создания гибких или прозрачных дисплеев.

Несмотря на тенденцию повсеместного перехода к электронному документообороту, по-прежнему весьма широко используются принтеры. Это устройства вывода текстовой или графической информации на твердый носитель. Чаще всего применяется бумага, а также пленки, пластик или ткани.

Два наиболее распространенных принципа построения принтеров: лазерные принтеры и струйные принтеры (рис. [fig:printers]).

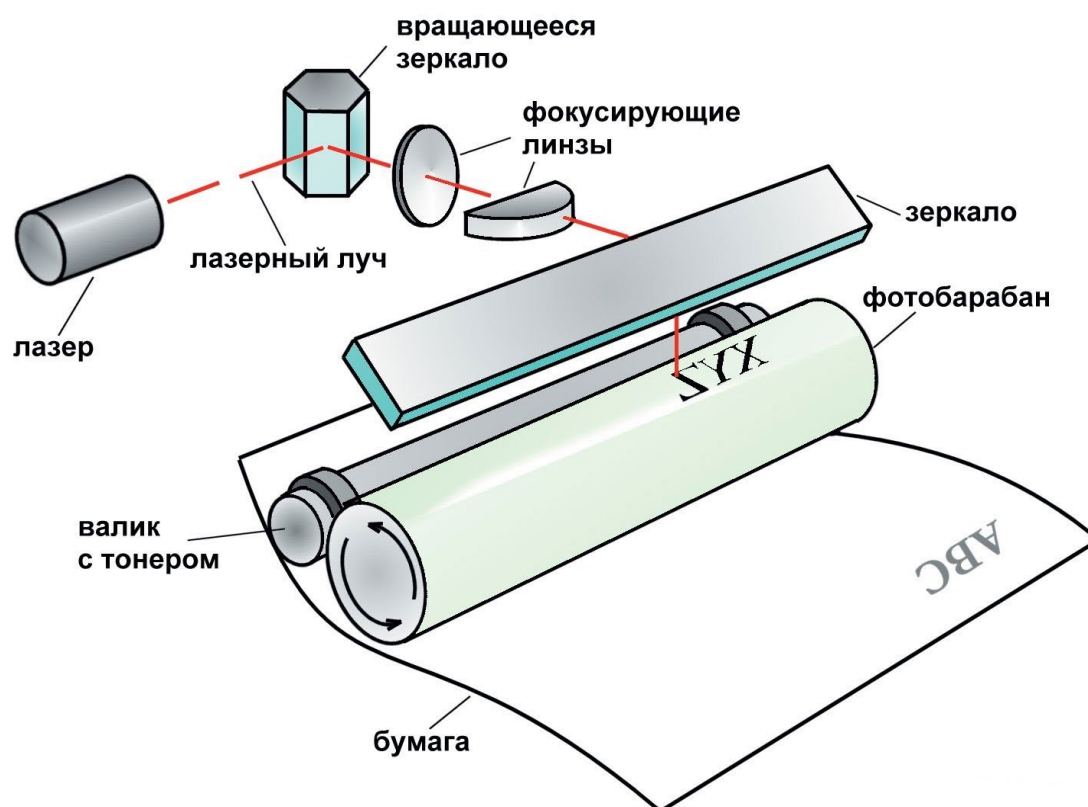


Рисунок 13. Лазерный принтер

Принцип работы лазерного принтера следующий:

1. нанесение равномерного слоя электрического заряда на поверхность фотобарабана;
2. поляризация молекул поверхности фотобарабана в необходимых точках с помощью лазера;

3. нанесение с помощью валика тонера, имеющего заряд, противоположный фотобарабану. при этом частицы тонера прилипают к обработанным лазером точкам на фотобарабане;

4. перенос тонера на бумагу и закрепление термомеханическим способом.

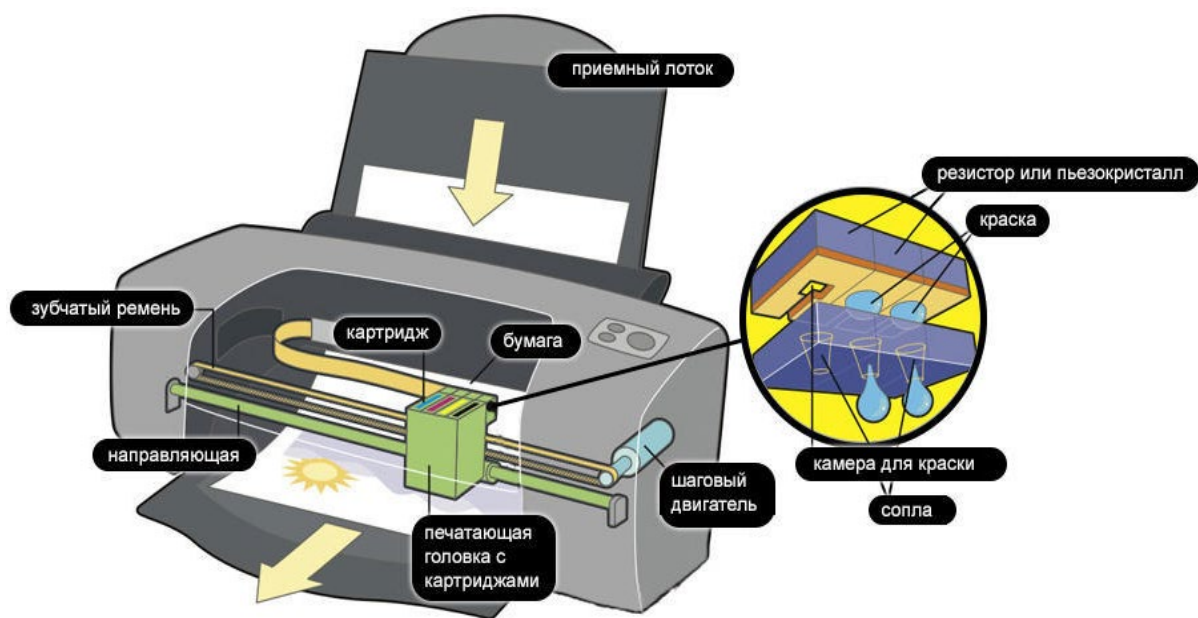


Рисунок 14. Струйный принтер

Струйные принтеры работают по следующему принципу:

1. печатающая головка с картриджами перемещается в нужную точку над листом бумаги с помощью шагового двигателя;

2. из камеры для краски с помощью нагревания (Canon) или деформации пьезокристалла (Epson) выбрасываются капли краски соответствующего цвета на поверхность бумаги.

Оба способа печати имеют ряд преимуществ и недостатков. Так, струйные устройства обладают лучшей цветопередачей, однако их производительность заметно ниже чем у лазерных. Стоимость для лазерных устройств выше струйных, однако они более надежные.

Помимо классических устройств ввода/вывода в современных компьютерных системах встречаются устройства, развивающие или сочетающие несколько принципов. Так, развитием клавиатур служат различные игровые устройства — джойстики, а сенсорные экраны сочетают в себе технологии сенсорных панелей и дисплеев, совмещая возможности ввода и вывода информации в одном устройстве. Развитие технологии струйной печати привело к созданию трехмерных принтеров.

Множество различных устройств ввода/вывода современных компьютерных систем позволяют максимально упростить и сделать комфортным взаимодействие пользователя с машиной. при этом применяется целый спектр технологий, от разработанных в середине прошлого века, до самых передовых.

10. ЦЕНТРЫ ОБРАБОТКИ ДАННЫХ. ВИРТУАЛИЗАЦИЯ ВЫЧИСЛИТЕЛЬНЫХ РЕСУРСОВ

В данной теме рассмотрены следующие вопросы: технологии построения современных центров обработки данных, виртуализация вычислительных ресурсов в компьютерных системах.

Стремительный рост объемов цифровой информации, развитие облачных вычислений, систем искусственного интеллекта, больших данных (Big Data) и интернета вещей (IoT) привели к тому, что одиночные серверы и локальные серверные комнаты перестали удовлетворять требованиям бизнеса, науки и государства по производительности, масштабируемости и отказоустойчивости. Современная информационно-телекоммуникационная инфраструктура строится на базе специализированных высокотехнологичных объектов — центров обработки данных (ЦОД).

В то же время эффективная утилизация колоссальных вычислительных мощностей современных ЦОД невозможна без применения технологий абстрагирования аппаратных ресурсов — виртуализации и контейнеризации, а также объединения географически распределенных вычислительных узлов в единые суперкомпьютерные и GRID-системы. Изучение принципов построения ЦОД и технологий виртуализации является необходимым элементом профессиональной подготовки современных ИТ-специалистов и инженеров вычислительной техники.

Технологии построения центров обработки данных (ЦОД)

Центр обработки данных (ЦОД, англ. Data Center) — это комплексное инженерно-техническое сооружение, предназначенное для обеспечения бесперебойного и безопасного функционирования серверного, сетевого оборудования и систем хранения данных (СХД). В физическом плане ЦОД включает специализированные помещения (машинные залы) для размещения телекоммуникационных стоек и шкафов, системы гарантированного и бесперебойного электроснабжения, прецизионного обеспечения температурно-влажностного режима, автоматического газового пожаротушения, контроля доступа и физической безопасности.

С точки зрения системного подхода ЦОД является сложной организационно-технической системой, которая объединяет в себе набор взаимосвязанных программных и аппаратных компонентов, регламентированных организационных процедур, специализированных

технологических площадок размещения оборудования и квалифицированного дежурно-эксплуатационного персонала для безопасной централизованной обработки, хранения и предоставления данных, информационных сервисов и бизнес-приложений.

В иерархической модели современного ЦОД, ориентированной на поддержку процессов основной деятельности организации через предоставление ИТ-услуг, выделяются три базовых горизонтальных уровня и два сквозных вертикальных контура:

Инженерная инфраструктура (нижний базовый уровень) — здания, специализированные помещения (машинные залы, кроссовые, щитовые, помещения ИБП и дизель-генераторных установок) и комплекс инженерных систем жизнеобеспечения. Этот уровень предоставляет услуги по физическому размещению и обеспечению непрерывной работоспособности средств автоматизации;

Системотехническая инфраструктура (средний уровень) — аппаратные платформы и комплексы технических средств, предоставляющие услуги по обработке, хранению и передаче данных. Она подразделяется на три функциональные подсистемы:

а) вычислительная инфраструктура — серверные платформы (стоечные, модульные блейд-системы, высокоплотные узлы с графическими и тензорными ускорителями) и системы обработки данных;

б) инфраструктура хранения данных — дисковые массивы, распределенные хранилища (SAN, NAS, объектные СХД) и ленточные библиотеки резервного копирования;

в) транспортная инфраструктура — структурированные кабельные системы (СКС), высокоскоростные коммутаторы, маршрутизаторы, оптические линии связи и узлы доступа;

Архитектура современных коммерческих, корпоративных и государственных дата-центров проектируется с учетом следующих фундаментальных принципов:

а) гибкость и масштабируемость: модульный принцип построения машинных залов и инженерных систем позволяет поэтапно наращивать вычислительную мощность и количество стоек без остановки работающего оборудования.

б) территориальная разнесенность кампусов: распределение вычислительных мощностей между несколькими географически удаленными площадками (основной и резервный ЦОД) обеспечивает

катастрофоустойчивость (Disaster Recovery) при природных или техногенных авариях.

г) гарантированное и бесперебойное энергоснабжение с высоким уровнем резервирования: использование нескольких независимых вводов от разных городских подстанций, промышленных источников бесперебойного питания (ИБП двойного преобразования) с аккумуляторными или литий-ионными массивами и дизель-генераторных установок (ДГУ) с запасом топлива на 24–72 часа автономной работы.

д) высокий уровень связности: наличие независимых оптических вводов от нескольких операторов связи, присутствие точек обмена трафиком (IX) и построение неблокируемых внутренних сетевых фабрик по топологии Spine-Leaf (позвоночник — лист).

е) контроль климата в машинных залах: применение прецизионных кондиционеров (CRAC/CRAH), чиллеров, систем фрикулинга (естественного охлаждения наружным воздухом), а также разделение воздушных потоков посредством изоляции «холодных» и «горячих» коридоров. Для высоконагруженных стоек искусственного интеллекта (мощностью свыше 30–100 кВт на стойку) все активнее внедряется прямое жидкостное (direct-to-chip) и иммерсионное (погружное) охлаждение.

ж) автоматическое пожаротушение: использование аспирационных систем сверхраннего обнаружения дыма (VESDA) и систем газового пожаротушения (хладоны, инертные газы, состав Noves 1230), которые ликвидируют возгорание без повреждения электроники и коррозии контактов.

з) удаленный круглосуточный доступ 24/7/365, охрана и видеонаблюдение: многорубежная физическая защита периметра, биометрические системы контроля и управления доступом (СКУД), шлюзовые тамбуры и дежурство смен инженеров в режиме 24/7/365.

Важнейшим показателем энергоэффективности инженерной инфраструктуры ЦОД является коэффициент PUE (Power Usage Effectiveness), определяемый как отношение общего энергопотребления всего объекта ЦОД к энергии, потребляемой непосредственно ИТ-оборудованием. Чем ближе значение PUE к 1,0, тем совершеннее инженерные системы охлаждения и электропитания (в современных ЦОД этот показатель составляет 1,1–1,35).

В соответствии с ГОСТ Р 27.102-2021 надежность представляет собой комплексное свойство системы сохранять во времени в установленных пределах значения всех параметров, характеризующих способность

выполнять требуемые функции. Основным методом обеспечения безотказности ЦОД является резервирование (нагруженное «горячее», облегченное «теплое» и ненагруженное «холодное») с различной кратностью избыточности.

В Российской Федерации проектирование, создание и эксплуатация ЦОД регламентируются комплексом национальных стандартов, разработанных Техническим комитетом ТК 120 «Центры обработки данных»:

ГОСТ Р 58811-2020 «Центры обработки данных. Инженерная инфраструктура. Стадии создания» — определяет упорядоченную совокупность стадий создания инженерной инфраструктуры ЦОД (ИИ ЦОД), этапы внутри каждой стадии и перечень работ. Стандарт устанавливает 9 последовательных стадий жизненного цикла создания объекта: 1) Предпроектное обследование (формирование требований со стороны ИТ-инфраструктуры, обследование площадки и оформление отчета); 2) Задание на разработку технической концепции; 3) Техническая концепция (предварительные технические решения); 4) Техническое задание (ТЗ); 5) Проектная документация; 6) Рабочая документация; 7) Реализация решений (поставка оборудования, монтажные и пусконаладочные работы); 8) Испытания (предварительные, опытная эксплуатация и комплексные приемочные испытания); 9) Эксплуатация (гарантийное, постгарантийное и сервисное обслуживание).

ГОСТ Р 58812-2020 «Центры обработки данных. Инженерная инфраструктура. Операционная модель эксплуатации. Спецификация» — формулирует требования к эксплуатирующей организации и процессам эксплуатации инженерных систем ЦОД для предоставления услуг надлежащего качества. Стандарт внедряет процессно-ориентированный подход: регламентирует документирование всех критичных процедур, структуру службы эксплуатации (включая матричную модель управления), ресурсную модель, управление инцидентами, планово-предупредительным техническим обслуживанием и ремонтом (ТОиР), контроль рисков и взаимодействие с внешними подрядчиками и поставщиками.

Перспективным направлением повышения надежности и оперативности управления современными ЦОД является внедрение интеллектуальных цифровых двойников (Digital Twins) и мультиагентных систем. Цифровой двойник объединяет BIM-модель машинного зала, термодинамическое CFD-моделирование воздушных потоков и потоковую

телеметрию от датчиков в реальном времени, позволяя прогнозировать отказы оборудования, оптимизировать охлаждение и сокращать время восстановления работоспособности системы.

Виртуализация вычислительных ресурсов

Виртуализация в компьютерных системах — это создание с помощью стандартных универсальных методов доступа дополнительного уровня абстракции над реальным аппаратным обеспечением. Иными словами, виртуализация позволяет отделить логическое представление вычислительных ресурсов (процессорного времени, оперативной памяти, дискового пространства, сетевых интерфейсов) от их физической реализации.

Исторически идеи виртуализации зародились еще в 1960–1970-х годах при создании мейнфреймов третьего поколения (таких как IBM System/360 и System/370) для организации мультипрограммного режима и разделения времени между пользователями. В 2000-х годах с появлением аппаратной поддержки виртуализации в процессорах архитектур x86/x64 (технологии Intel VT-x и AMD-V) и ARM виртуализация стала фундаментальной технологией всех современных ЦОД и облачных платформ.

Основные преимущества внедрения виртуализации:

а) консолидация серверов и повышение коэффициента утилизации оборудования (с типичных 10–15 % при физическом размещении одного приложения на сервере до 70–85 % в виртуализированном кластере);

б) строгая изоляция вычислительных сред: сбой или компрометация одной виртуальной среды не влияет на работу соседних виртуальных машин и хостовой системы;

в) гибкость, эластичность и быстрое развертывание новых серверных узлов из готовых шаблонов за считанные минуты;

д) высокая доступность и непрерывность бизнеса за счет технологий «живой» миграции (live migration) работающих виртуальных машин между физическими серверами без прерывания работы пользователей, создания мгновенных снимков состояния (snapshots) и централизованного резервного копирования.

В современных вычислительных системах выделяют три основных направления виртуализации: аппаратную виртуализацию, виртуализацию рабочих столов и виртуализацию на уровне операционной системы (контейнеризацию).

Аппаратная (платформенная) виртуализация

Аппаратная виртуализация представляет собой создание независимых и изолированных друг от друга виртуальных компьютеров с помощью программной и аппаратной имитации ресурсов (процессора, оперативной памяти, сети, дисковых контроллеров и др.) физического сервера.

При этом используются следующие базовые понятия:

а) хостовая машина (хост, host) — реальный физический сервер, предоставляющий свои аппаратные ресурсы;

б) виртуальная машина (ВМ, гостевая машина, guest) — программно эмулируемый виртуальный компьютер, внутри которого устанавливается собственная независимая гостевая операционная система и прикладное ПО;

г) гипервизор (монитор виртуальных машин, VMM) — специализированное программное обеспечение, которое создает виртуальные машины, распределяет между ними физические ресурсы хоста, обеспечивает их изоляцию и управляет их жизненным циклом.

По архитектурному исполнению гипервизоры делятся на два типа:

а) гипервизоры 1-го типа (автономные, bare-metal / native) — запускаются непосредственно на «голом железе» физического сервера без промежуточной хостовой ОС, фактически сами выполняя функции специализированного ядра ОС. Они обеспечивают максимальную производительность, минимальные накладные расходы и применяются в промышленных ЦОД. Примеры: VMware ESXi, Microsoft Hyper-V, KVM (Kernel-based Virtual Machine в составе ядра Linux), Xen, FreeBSD bhyve;

б) гипервизоры 2-го типа (хостовые, Hosted) — устанавливаются и работают как прикладная программа поверх обычной универсальной операционной системы (Windows, Linux, macOS). Они удобны для разработки, тестирования и обучения на рабочих станциях, но имеют большие накладные расходы. Примеры: Oracle VirtualBox, VMware Workstation, QEMU.

Виртуализация рабочих столов и приложений (VDI и RDP/RDS)

Данный вид виртуализации переносит пользовательскую рабочую среду с локальных персональных компьютеров в защищенный ЦОД, предоставляя сотрудникам удаленный доступ по сети (чаще всего через компактные «тонкие клиенты», ноутбуки или веб-браузеры). Существуют два основных подхода:

а) VDI (Virtual Desktop Infrastructure) — программно-аппаратный комплекс, позволяющий каждому пользователю иметь на сервере свой собственный изолированный образ клиентской операционной системы (персональную виртуальную машину), при этом к части аппаратных ресурсов сервера образы имеют общий разделяемый доступ. Обеспечивает полную персонализацию и высокий уровень информационной безопасности;

б) RDP / RDS (Remote Desktop Protocol / Remote Desktop Services, терминальный доступ) — рабочей средой является единая серверная операционная система, работающая в многопользовательском терминальном режиме. Пользователь получает удаленный сеансовый доступ к своему рабочему столу или конкретным опубликованным приложениям (RemoteApp). Этот подход требует меньше ресурсов на одного пользователя по сравнению с VDI, но обладает меньшей степенью изоляции между сеансами.

Виртуализация на уровне операционной системы (контейнеризация)

Виртуализация на уровне ОС позволяет запускать программное обеспечение в изолированных пользовательских пространствах (контейнерах), разделяющих единое общее ядро хостовой операционной системы.

Наиболее распространенной формой такого подхода являются контейнеры: Docker, containerd, LXC (Linux Containers), OpenVZ, а также механизм jails во FreeBSD. В основе контейнеризации в ОС Linux лежат два ключевых механизма ядра: пространства имен (namespaces), которые изолируют видимость процессов, сетевых интерфейсов, точек монтирования и идентификаторов пользователей, и контрольные группы (cgroups), которые жестко ограничивают и учитывают потребление аппаратных ресурсов (CPU, RAM, дискового и сетевого ввода-вывода) для каждого контейнера.

Контейнеры значительно более легковесны, чем классические виртуальные машины, так как они опираются на функционал уже работающего ядра ОС, не требуют эмуляции аппаратного обеспечения и запуска отдельной гостевой ОС. На практике контейнер представляет собой стандартизированную изолированную среду для запуска любого приложения со всеми его библиотеками, зависимостями и конфигурационными файлами. В современных ЦОД управление тысячами контейнеров автоматизируется с помощью систем оркестрации, стандартом де-факто среди которых является платформа Kubernetes.

Комплексное применение виртуализации ко всем подсистемам дата-центра привело к появлению концепции программно-определяемого ЦОД (Software-Defined Data Center, SDDC), в котором виртуализируются не только серверы, но и системы хранения данных (SDS — Software-Defined Storage), и сетевая инфраструктура (SDN — Software-Defined Networking). В архитектуре SDN плоскость управления сетью (control plane) отделяется от плоскости передачи трафика (data plane), что позволяет программно создавать виртуальные локальные сети (VLAN/VXLAN), маршрутизаторы и межсетевые экраны для каждой группы виртуальных машин.

Логическим развитием SDDC стала гиперконвергентная инфраструктура (Hyper-Converged Infrastructure, HCI) — архитектура, объединяющая вычислительные ресурсы (гипервизор), локальные накопители серверов (объединенные в единое распределенное программно-определяемое хранилище) и сетевую виртуализацию в стандартных модульных серверных узлах под единой системой управления.

Таким образом, современные центры обработки данных представляют собой сложные комплексные организационно-технические системы, объединяющие инженерную, системотехническую и программную инфраструктуры под управлением единых контуров диспетчеризации и информационной безопасности. Строгое соблюдение стандартов надежности (Tier I–IV) и требований национальных стандартов ГОСТ Р 58811-2020 и ГОСТ Р 58812-2020 на всех стадиях создания и эксплуатации ЦОД позволяет обеспечить непрерывность функционирования критически важных информационных сервисов в режиме 24/7/365.

Технологии аппаратной виртуализации, виртуализации рабочих столов и контейнеризации на уровне ОС кардинально изменили архитектуру вычислительных систем, позволив перейти от жесткой привязки программ к физическому «железу» к гибким программно-определяемым дата-центрам (SDDC).

Список рекомендуемой литературы

1. Харрис Д. М. Цифровая схемотехника и архитектура компьютера / Д. М. Харрис, С. Л. Харрис; пер. с англ. — 2-е изд., испр. и доп. — Москва: ДМК Пресс; Лондон: Morgan Kaufmann, 2013. — 1622 с. — Текст: непосредственный.
2. Угрюмов Е. П. Цифровая схемотехника: учебное пособие для вузов / Е. П. Угрюмов. — 2-е изд., перераб. и доп. — Санкт-Петербург: БХВ-Петербург, 2007. — 800 с.: ил. — Текст: непосредственный.
3. Микропроцессорные устройства. Проектирование и моделирование элементов микропроцессорных устройств на программируемых логических интегральных схемах: лабораторный практикум / А. Ю. Березин, Д. И. Исаенко, А. В. Пачин, Б. К. Резников, Г. В. Степаненков; Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича. — Санкт-Петербург: СПбГУТ, 2023. — 108 с. — Текст: непосредственный. *(Примечание: при наличии пяти и более авторов допускается также сокращенный вариант за косой чертой: / А. Ю. Березин, Д. И. Исаенко, А. В. Пачин [и др.]; СПбГУТ.)*
4. Таненбаум Э. Архитектура компьютера / Э. Таненбаум, Т. Остин; пер. с англ. Е. Матвеева. — 6-е изд. — Санкт-Петербург: Питер, 2013. — 816 с.: ил. — Текст: непосредственный.
5. Гуров В. В. Архитектура микропроцессоров: учебное пособие / В. В. Гуров. — 2-е изд. — Москва: ИНТУИТ, 2016. — 328 с. — Текст: непосредственный.
6. ГОСТ Р 27.102–2021. Надежность в технике. Надежность объекта. Термины и определения: национальный стандарт Российской Федерации: издание официальное: введен впервые: введен 2022-01-01 / разработан Обществом с ограниченной ответственностью «Институт надежности машин и технологий». — Москва: Российский институт стандартизации, 2021. — IV, 36 с. — Текст: непосредственный.
7. Неелова О. Л. Архитектура вычислительных систем: учебное пособие / О. Л. Неелова; Санкт-Петербургский государственный университет телекоммуникаций им. проф. М. А. Бонч-Бруевича. — Санкт-Петербург: СПбГУТ, 2021. — 72 с.: ил. — Текст: непосредственный.