Архитектура ЭВМ

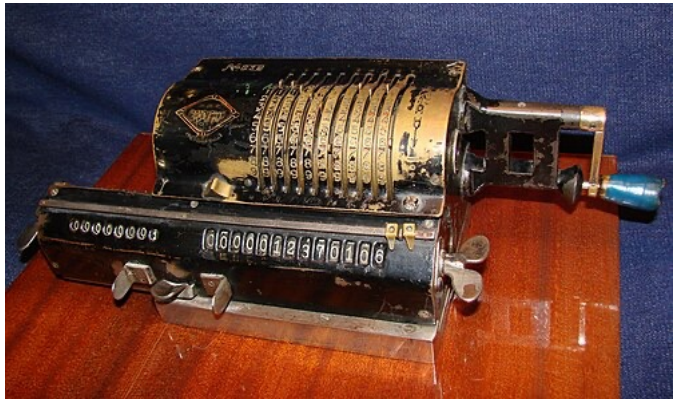
Иртегов Дмитрий Валентинович

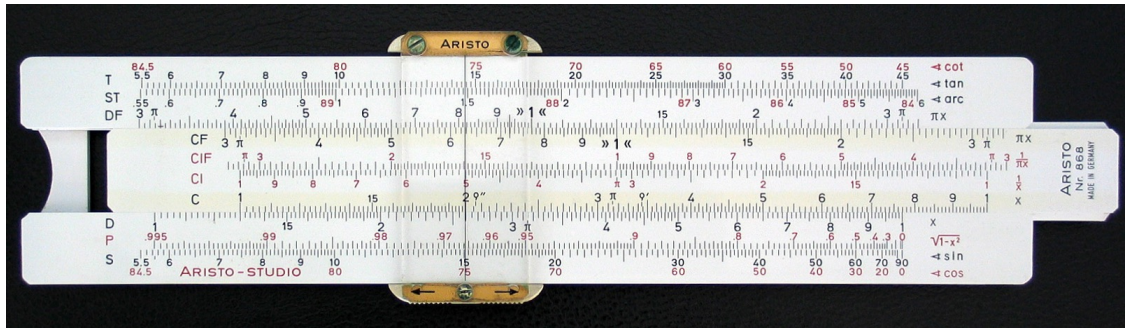
НГУ

2025

Что такое «компьютер»

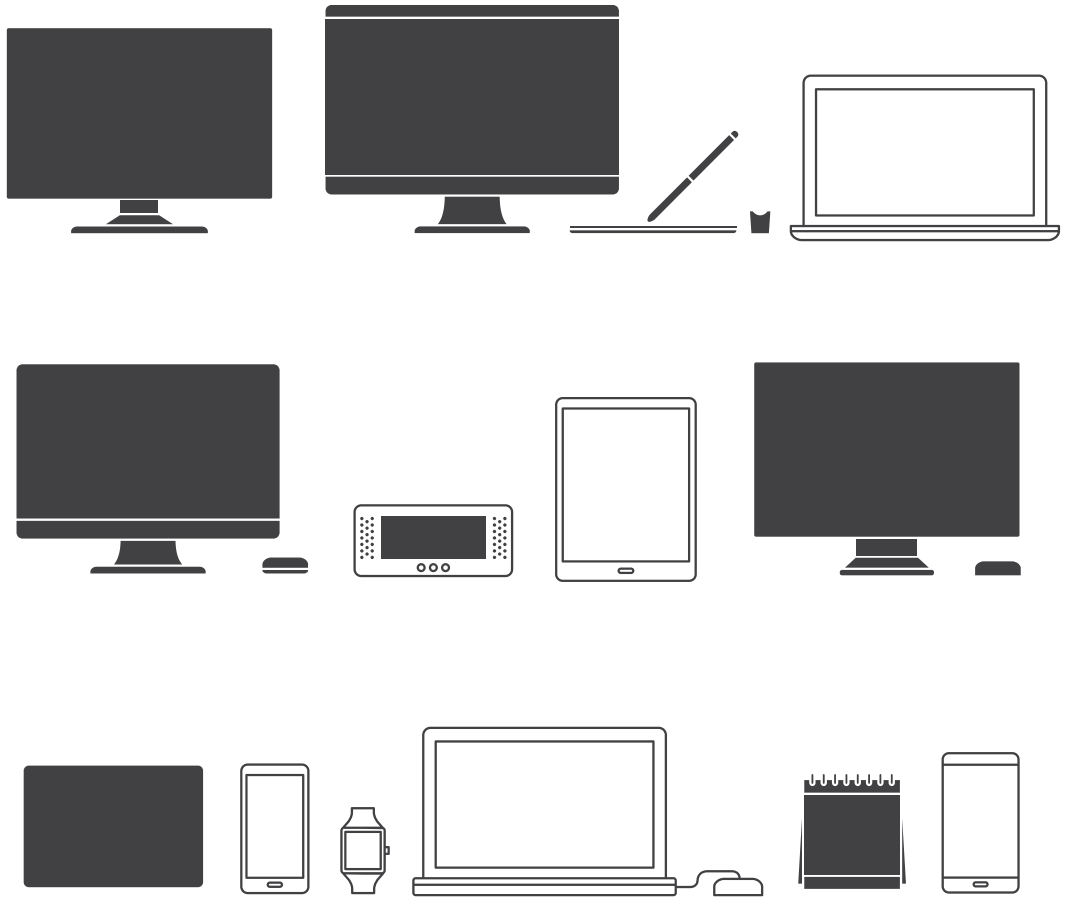
- По-русски это «вычислитель»





Что понимают под компьютером сейчас?

- Программируемое вычислительное устройство
- Большинство современных компьютеров - электронные цифровые
- Обычно подразумевается компьютер фон-Неймановской архитектуры (что это такое, поймем позже)

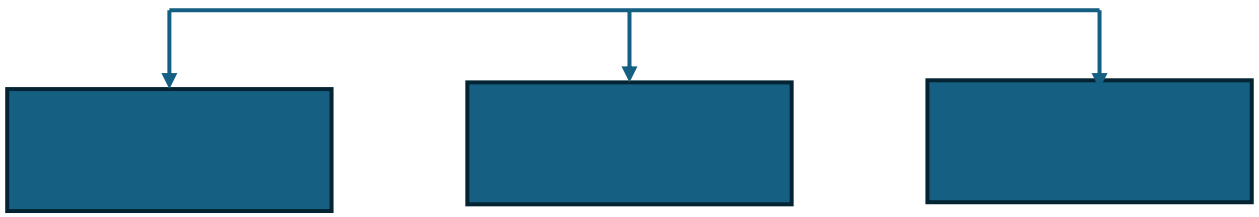


Типы компьютеров

- Персональные
  - Ноутбуки
  - Настольные («десктоп»)
- Серверы
- Планшеты, смартфоны, «умные часы», …
- «Не компьютеры», содержащие компьютер внутри
  - Фотоаппараты, телевизоры (Smart TV), игровые приставки, …
  - Маршрутизаторы, NAS, …
  - Банкоматы, кассовые аппараты, терминалы оплаты, …
- Встраиваемые компьютеры
  - «Интернет вещей»
  - Станки с ЧПУ (Числовым Программным Управлением)
  - Бортовые компьютеры автомобилей, самолетов, спутников
  - …

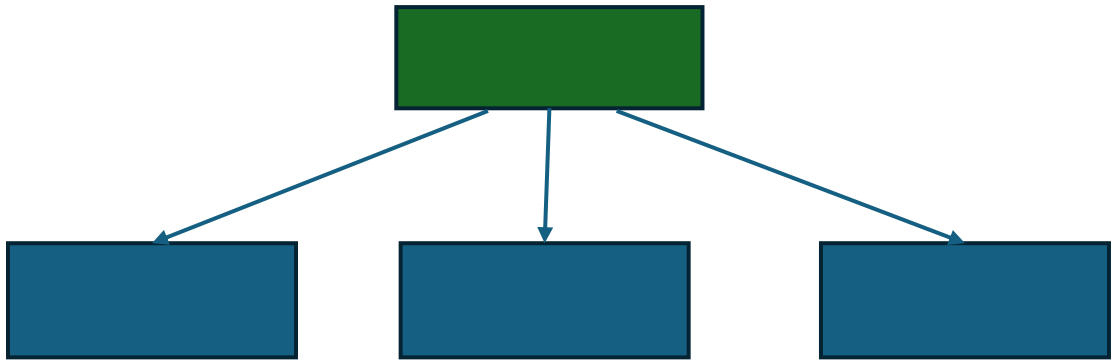
Немного терминологии: шина

- Слово «Шина» мы будем слышать много раз в этой лекции
- Полезно его понимать
- Формальное определение:
- Шина – это среда передачи (устройство или комплекс устройств) соединяющее несколько других устройств, например функциональных блоков компьютера
- В простейшем случае просто пучок проводов.



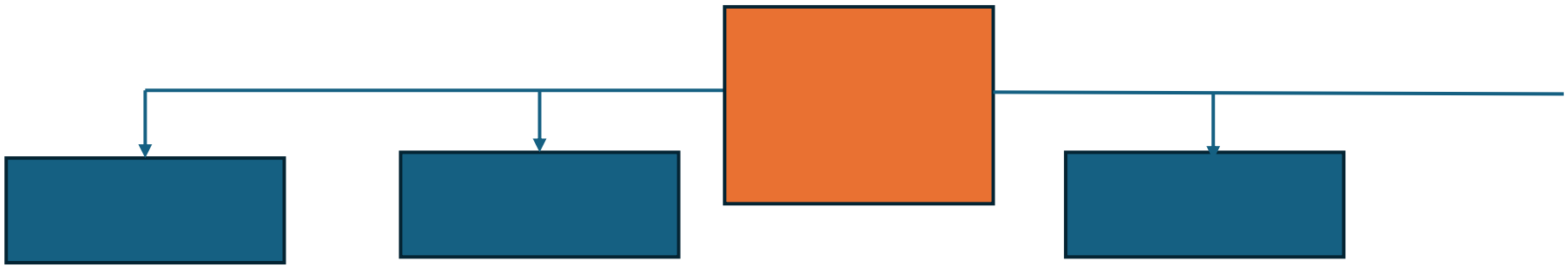
В простейшем случае?
А что в более сложных?

- Кроме проводов, шина может содержать другие устройства
- Повторитель – аналоговый усилитель, нужен при большой длине проводов
- Хаб (концентратор, многопортовый повторитель) - позволяет придать шине звездообразную или древовидную топологию



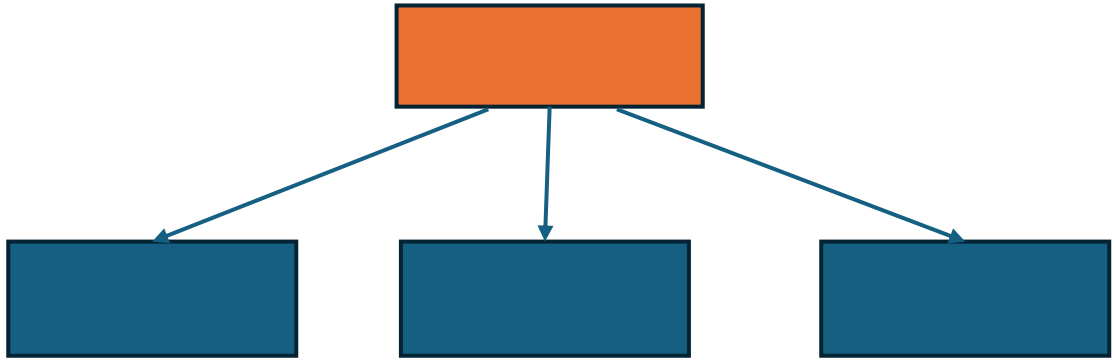
А в еще более сложных?

- Мост: устройство с внутренним буфером, которое принимает пакеты данных из одного сегмента шины и (возможно) передает их в другой
  - Разделение домена коллизий (устройства в разных сегментах могут разговаривать друг с другом одновременно)
  - Сопряжение сегментов с разной тактовой частотой



Коммутатор (свитч)

- Можно описать как многопортовый мост
- Часто довольно умные устройства (т.наз. неблокирующие коммутаторы), позволяющие парам устройств разговаривать друг с другом независимо
- Позволяют строить сети сложной топологии (нужны совсем умные коммутаторы)



Еще определений

- Задатчик шины (bus master): устройство, которое начинает коммуникацию
- Single-master bus – шина, в которой только одно устройство может быть задатчиком (компьютер с одним процессором)
- Multi-master bus – например, компьютер с несколькими процессорами или умными устройствами
- Арбитр шины - в multi-master bus, устройство, которое разрешает коллизии (определяет, кто из задатчиков сейчас будет говорить)
  - Бывают шины с разрешением коллизий без арбитра, например старые версии протокола Ethernet

Адресация

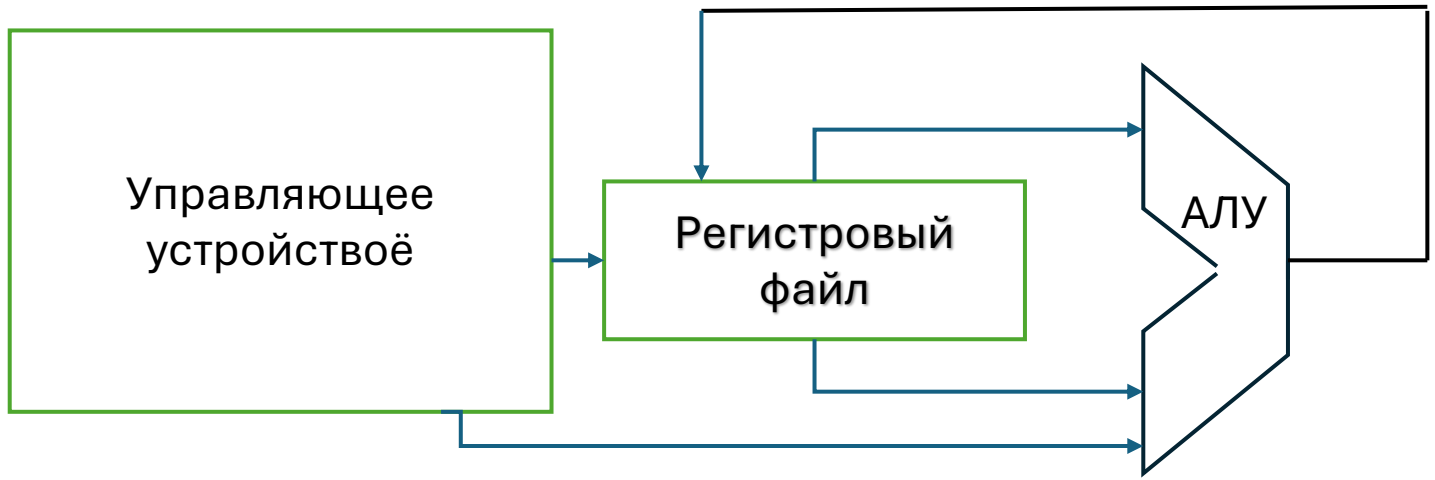
- К шине подключается много устройств
- Задатчик шины должен сказать, с кем он сейчас будет разговаривать. Это называется адресация.
- Обычно, адрес – это просто число (битовая строка)
- Некоторые устройства (оперативная память, видеоадаптер) занимают не один адрес, а диапазоны адресов
  - Это позволяет обращаться к отдельным байтам памяти
- Шина, в том числе сложной топологии, состоящая из сегментов, объединенных мостами, коммутаторами и т.д., но имеющая единую адресацию, называется *логической шиной*

И еще определение

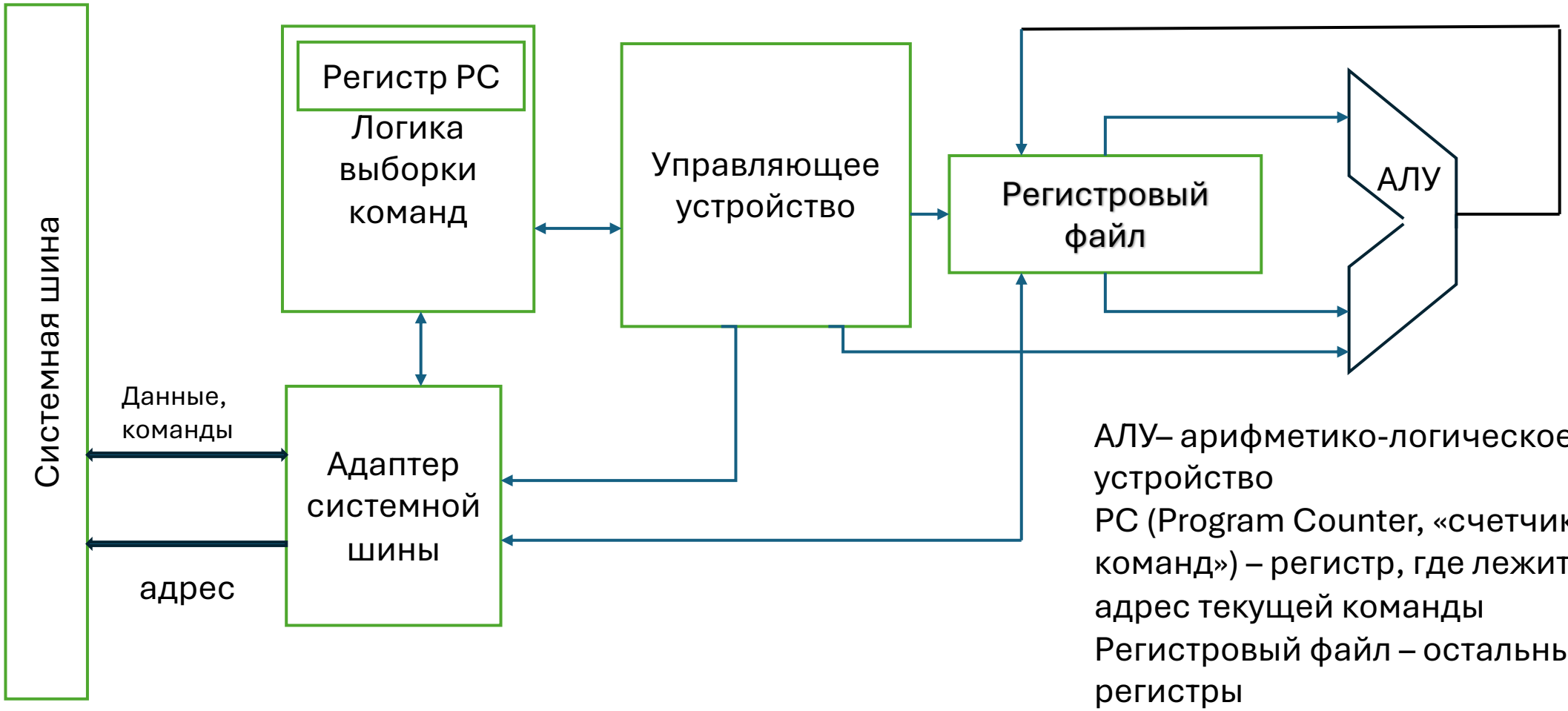
- Адаптер – устройство, соединяющее шины с разной адресацией (например, системную шину и шины USB, SCSI или Ethernet)
  - В сетях такие устройства называют шлюзами (gateway)
- Это не последнее определение в этой лекции, будут еще
- Но мы уже можем переходить к основной теме

Центральный процессор

- Не обязателен, но цифровые устройства без центрального процессора компьютерами обычно не называют
- В наше время и некоторые устройства с процессором компьютером не называют
  - контроллеры «интернета вещей», телефоны, маршрутизаторы - …



Центральный процессор чуть подробнее



АЛУ– арифметико-логическое устройство
PC (Program Counter, «счетчик команд») – регистр, где лежит адрес текущей команды
Регистровый файл – остальные регистры

Что делает центральный процессор

- Исполняет команды
- Команды лежат в памяти с байтовой адресацией (ОЗУ или ПЗУ)
- Команда – это битовая строка, разбитая на несколько полей
- У RISC-V команда имеет длину 32 бита
  * С опцией C он также поддерживает 16-битные команды, но сейчас это неважно
- После исполнения команды, регистр PC увеличивается на длину команды (4 байта) и процессор исполняет следующую команду

Какие бывают команды

- Обработки данных
  - сложение, вычитание, побитовые операции
- Передачи данных
  - Между регистрами (move)
  - Между регистром и памятью (load/store)
- Передачи управления (записи в регистр PC) – позволяют реализовать условные операторы, циклы, подпрограммы
  - Условные
  - Безусловные
  - Прямые (адрес берется из команды)
  - Вычислимые
  - Простые (делается только переход)
  - С сохранением адреса возврата

На самом деле, типов команд больше

- Изменение настроек процессора
- Переключение в привилегированный режим и обратно
- Обращение к сопроцессору
- Еще много чего
- Но мы не будем на них отвлекаться

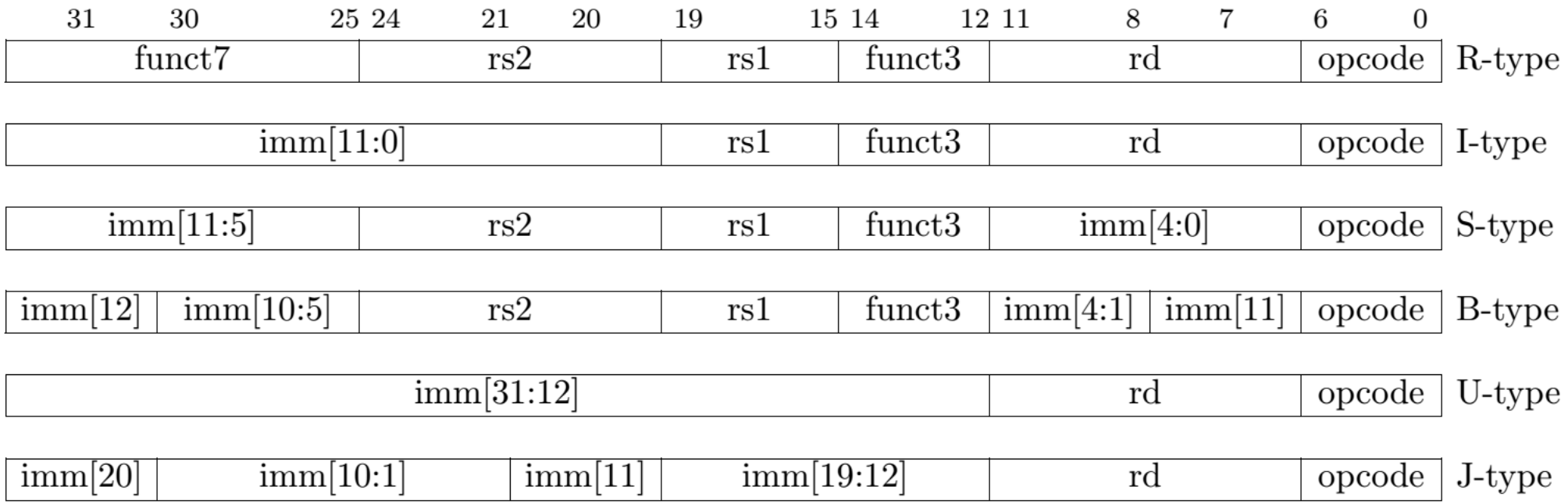
Определение фон-неймановской архитектуры

- Компьютер имеет программную память с произвольной адресацией
- Программа исполняется последовательно, но может содержать команды перехода (передачи управления по произвольному адресу)
- Варианты:
  - Манчестерская архитектура
    данные и программа хранятся в одной и той же памяти
  - Гарвардская архитектура
    данные и программа хранятся в разных банках памяти
  - Много книг и других источников, где фон-неймановской архитектурой называют только «манчестерскую»

А бывают не фон-неймановские компьютеры?

- Компьютер с несколькими процессорами строго говоря не является фон-неймановским (программа выполняется не последовательно)
- Векторные вычислители (Single Instruction Multiple Data):
  - Графические ускорители
  - Ускорители для «искусственного интеллекта»
  - Сопроцессоры для цифровой обработки сигналов
  - SIMD («векторные») и «мультимедийные» команды фон-неймановских процессоров

Форматы команд RISC-V

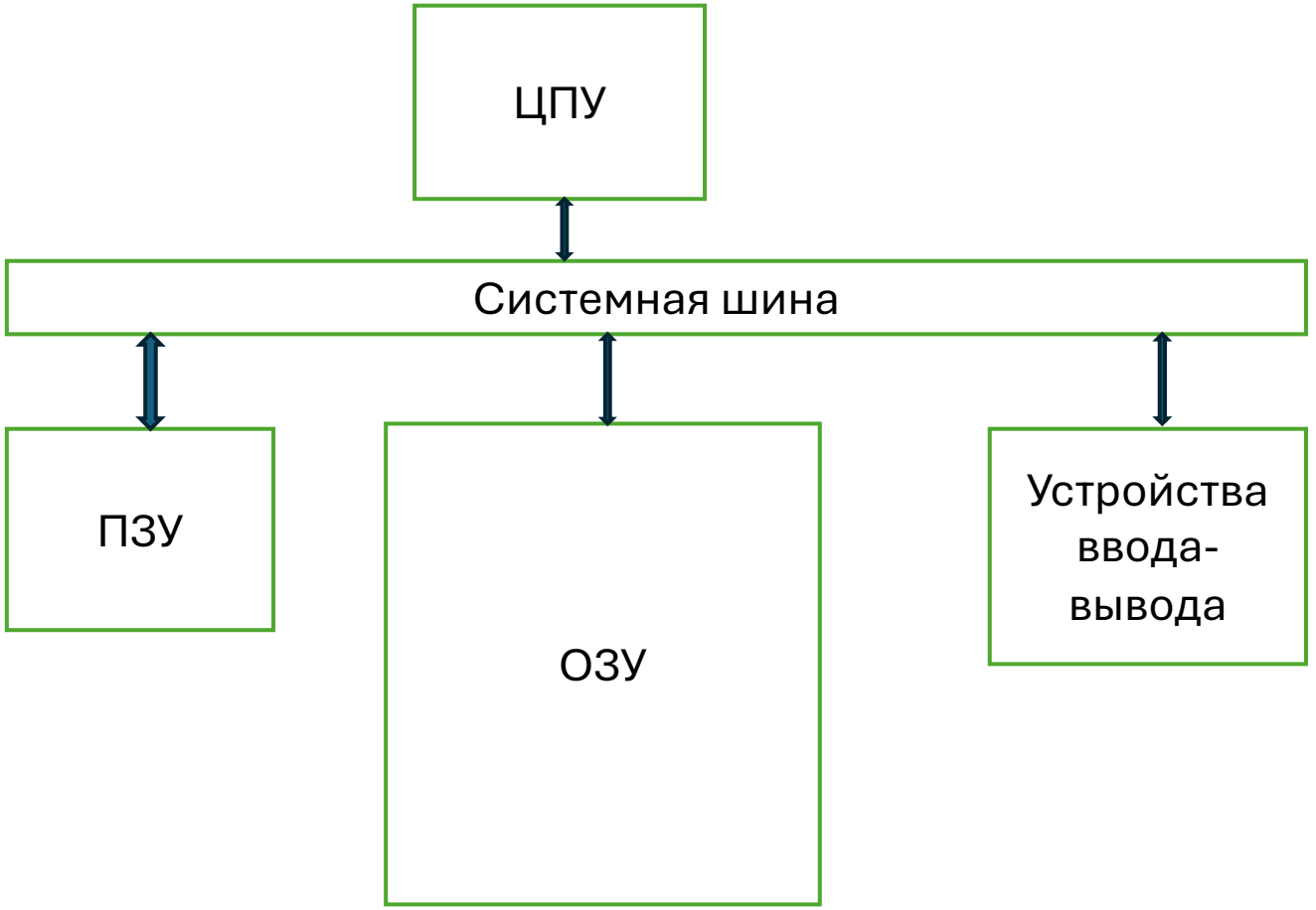


Еще про форматы команд RISC-V

- Команда может содержать
  - Номера трех регистров (R-type)
  - Номера двух регистров и 12-битное число (I-type, S-type, B-type)
  - Номер одного регистра и 20-битное число (U-type, J-type)
- Номера регистров 5-битные, то есть процессор может иметь 32 регистра
- Команды обработки данных работают только с регистрами (взять регистр rs1,rs2, сложить результат в rd) или двумя регистрами и константой
- Команды обращения к памяти (ld и st) – сложить два регистра или регистр и константу, получается адрес.

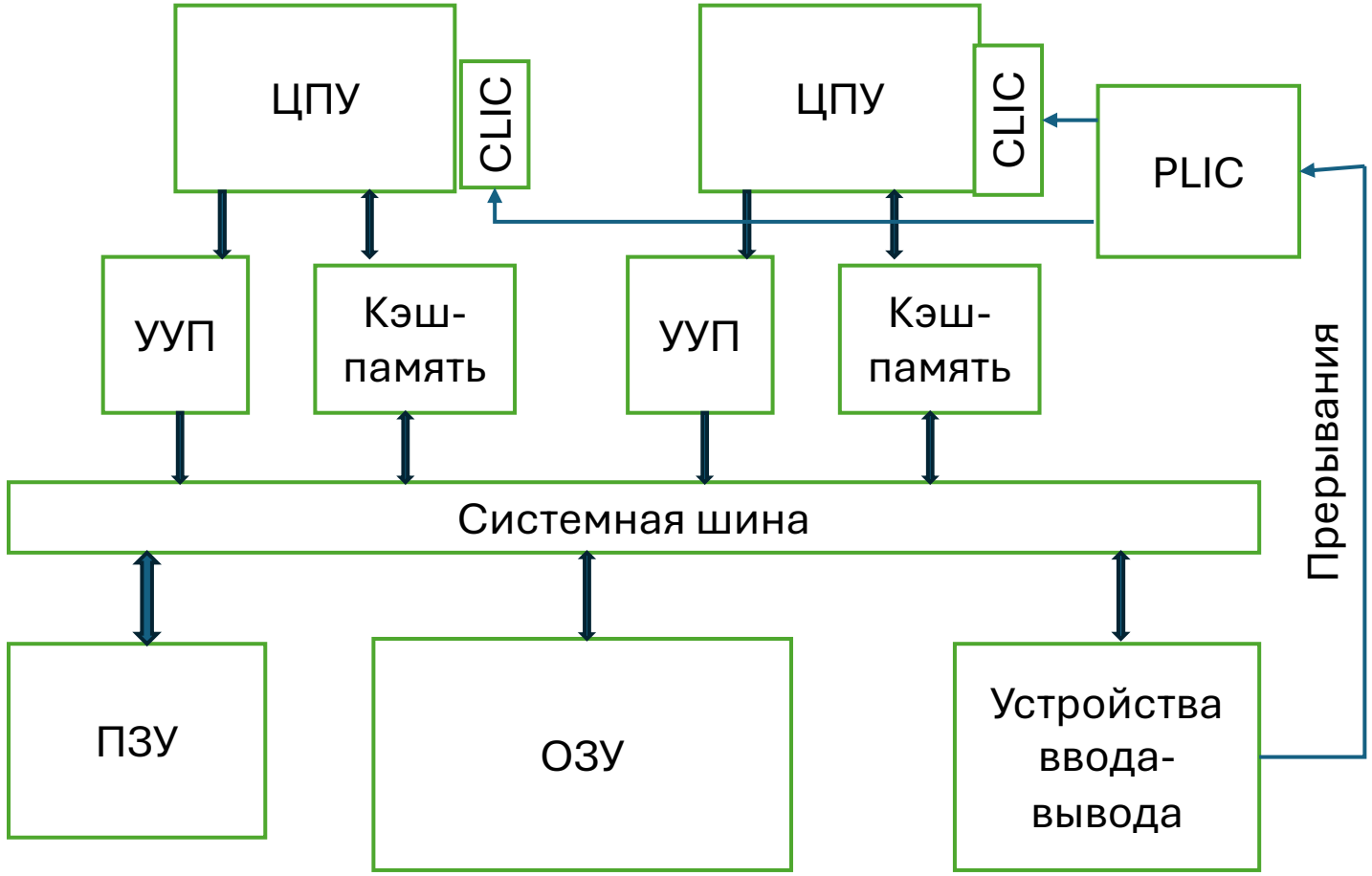
Теперь давайте рассмотрим компьютер в целом

- Крайние варианты
- Контроллер «интернета вещей»
  - Большое ПЗУ, в которое прошито все ПО
  - Маленькое ОЗУ
- Сервер или ПК
  - Маленькое ПЗУ (загрузочный монитор)
  - Большое ОЗУ
  - ОС и программы загружаются с внешних носителей в ОЗУ



Картинка чуть подробнее

- УУП, Устройство управления памятью – преобразует виртуальные адреса в физические, используется в ОС с виртуальной памятью (Linux, другие Unix-системы, Windows)
- Кэш-память – быстрая память с ассоциативным доступом, хранит команды или данные, к которым часто обращаются



О прерываниях

- Прерывание – сигнал, обычно от внешнего устройства, о каком-то событии (например, завершении операции)
- В ответ на этот сигнал, процессор останавливает текущую программу, сохраняет значение PC и еще некоторых регистров и запускает программу-обработчик
- CLIC – Core Level Interrupt Controller, в RISC-V – устройство, управляющее доставкой прерываний отдельному процессорному ядру (в x86 это называется LAPIC)
- PLIC – Platform-Level Interrupt Controller, в RISC-V – устройство, маршрутизующее сигналы прерываний межу несколькими процессорными ядрами (в x86 это называется IOAPIC)

Что такое «устройства ввода-вывода»?

- В первом приближении, все остальное, кроме ЦПУ и ОЗУ
- Еще есть термины «внешние» или «периферийные» устройства
  - Внешние запоминающие устройства с блочным доступом: диски, флэш память (SD-card, SSD, NVME)
  - Сетевые адаптеры
  - Видеоадаптеры
  - Клавиатуры
  - Мыши, планшеты, датчики сенсорных экранов
  - ЦАП/АЦП (звуковые адаптеры, цифровые осциллографы)
  - Датчики и актуаторы роботов и «интернета вещей»
  - Да тыщи их

Подключение устройств к шине

- Системная шина устроена просто
- Процессор выставляет адрес и сигнал, читает он или пишет данные
- Если он пишет данные, после адреса сразу выставляются данные
- Если он читает, он ждет, пока устройство выставит сигнал готовности и данные
- Ждать необходимо, так как память гораздо медленнее процессора, а в современных компьютерах процессоров несколько, а память часто одна
- Периферийные устройства могут откликаться еще медленнее, чем память

Подключение устройств к шине (продолжение)

- Устройство обязано иметь так называемый адресный дешифратор – логическое устройство, опознающее «его» адрес.
  - В простейшем случае это просто логический вентиль «И», у которого перед частью входов стоят инверторы
  - Это не позволяет изменять адрес, поэтому на практике часто используются более сложные конфигурируемые схемы
- Устройство также обязано иметь один или несколько регистров, обращение к которым триггерится адресным дешифратором
- Такие регистры называют «портами»
  - Слово «порт» имеет много других значений, в том числе в контексте устройств ввода-вывода
- Также, устройство может иметь банк памяти, отображаемый на диапазон адресов, как обычная память
  - (так обычно устроены видеоадаптеры)

Как задавать адреса портов

- «Зашить» при изготовлении
  - Невозможно подключить однотипные устройства
  - Возможны конфликты между устройствами разных производителей
  - Вполне приемлемо, когда большинство устройств на одной микросхеме (т.наз. System On a Chip)
- Конфигурировать перемычками/тумблерами
  - Было популярно в 70е-80е годы
  - Легко ошибиться и запутаться, особенно если много устройств
  - Возможно только когда устройство – это достаточно большая плата (DEC Q-Bus/Unibus, IBM PC ISA)
- Конфигурировать программно
  - Нужно подключать к шине еще одно устройство (через которое делается конфигурация)
  - Задача определения адреса сводится к предыдущей
- Географическая адресация – у каждого разъема свой диапазон адресов
  - Во многих отношениях неудобно
  - Но конфликты невозможны!

Еще немного определений - SoC

- System On a Chip
- Не имеет точного общепринятого определения
- Современные технологии позволяют разместить множество устройств на одной микросхеме.
- Термин SoC может обозначать:
  - Все функциональные модули компьютера (процессор[ы], ОЗУ, ПЗУ, все периферийные устройства) на одной микросхеме
    но это чаще называют «микроконтроллер»
  - Мост системной шины, контроллер ОЗУ, мост периферийной шины и большая часть периферии. Достаточно добавить процессор[ы], ОЗУ и накопители внешней памяти. Мост периферийной шины позволяет расширять систему.
    Типичная архитектура для современных ноутбуков, смартфонов, контроллеров «интернета вещей» и плат для разработки, таких, как Raspberry Pi и их аналогов

Преимущества и недостатки SoC

- Микросхемы, даже сложные, выгоднее всего производить большими партиями
- Процессор - универсальное устройство
- Для разных применений нужны разные наборы устройств
  - Например, в интернете вещей не нужен графический адаптер
- Если делать SoC, пригодную и для ноутбука, и для смартфона, и для интернета вещей, в ней будет много лишнего
- Если делать разные SoC для разных целей, каждую из них надо будет делать меньшими партиями
- Если делать «модульную» SoC, она будет состоять из нескольких микросхем, что обесценивает всю затею

Периферийная шина PCI

- Подключается к системной шине через мост
- В современной редакции стандарта (PCI Express) - коммутируемая последовательная шина. Коммутаторы по историческим причинам называются мостами.
- Это позволяет использовать одинаковые устройства с процессорами разных поколений или архитектуры
- Порты, память и ПЗУ устройств отображаются на адреса системной шины, в этом смысле системная шина и PCI остаются единой логической шиной
- Устройство обязано иметь интерфейс для автоконфигурации, адрес ему назначается географически (номер моста и номер порта на этом мосту)
- Через этот интерфейс устройство рассказывает, кто оно такое (идентификатор модели и производителя, тип устройства) и какие параметры ему можно сконфигурировать.
- Обычно эту конфигурацию проводит загрузочное ПЗУ, но часто ОС после загрузки переделывает эту конфигурацию «под себя»
- ОС использует эту информацию для поиска подходящего драйвера

Еще про шину PCI

- В настольных компьютерах, ноутбуках и серверах (не только с процессорами x86) ничего другого и не применяют с ~2005 года
- В Banana/Lichee Pi и многих других одноплатных компьютерах есть мост PCI и разъем m.2 (да, это вариант разъема PCI Express), но нет ни одного PCI устройства на плате

Рассмотрим UART 16550

- Universal Asynchronous Receiver Transmitter (универсальный асинхронный приемник-передатчик), совместимый с микросхемой NS16550
- Обеспечивает коммуникацию по протоколу RS-232
- Протокол придуман телеграфистами в 1960е годы, использовался для подключения к компьютерам телетайпов, терминалов, модемов, мышей, различных низкоскоростных устройств
- UART есть во многих микроконтроллерах, поэтому устройств с таким или похожим интерфейсом вокруг до сих пор много

Зачем UART/RS-232 нужен сейчас?

- UART есть во многих микроконтроллерах и платах Arduino, Raspberry Pi, Banana/Lichee Pi
- Если подключить к нему терминал, его можно использовать в качестве консоли для разговора с загрузочным ПЗУ
- Если ваша плата загрузится до Линукса, можно будет зайти в Linux терминальной сессией
- Полезно, если в вашей плате нету графического адаптера или вы почему-то не хотите его включать
- Незаменимо, если плата не грузится

Что такое терминал?

- Устройство, придуманное в 1970е годы для замены телетайпов
- Включает монитор, клавиатуру, контроллер на основе микропроцессора и порт RS-232
- Кнопки, которые вы нажимаете на клавиатуре, преобразуются в байты и отправляются в порт
- Байты, получаемые из порта, рисуются на экране в виде буковок

Где взять терминал в наше время?

- Ограбить музей
  - Не говорите что я вам это советовал
- В Linux/Unix и Windows есть специальные приложения, эмулирующие терминал, используя окно GUI вместо экрана
  - В Windows 10 и более старых это называлось «консоль»
- Чтобы соединить такое приложение с портом RS-232, нужна еще одна утилита, в Linux обычно используют minicom
- В современных PC и Mac портов RS-232 нету, но доступны адаптеры с интерфейсом USB
  - С ними есть одна плюшка при подключении к Banana/Lichee Pi
  - Стандарт RS-232 использует -12V для кодирования 1,
  - Порты Banana/Lichee/Raspberry используют +5V («TTL-совместимое»)
  - Порт вы вряд ли сожжете, но данные не пойдут
  - Ищите подходящий адаптер или конвертор напряжения

Еще про RS-232

- Посылает данные по одному байту
- Соединение «точка-точка», никакой адресации не предусмотрено
- Никакой нумерации пакетов и их пересылки в случае ошибок тоже
- «Пакет» состоит из стартового бита (всегда 1), от 5 до 8 бит данных, возможно бита четности и одного или двух стоповых битов (тоже всегда 1)
- Скорость передачи выставляется программно, от 300 бит/сек до 115200 бит/сек.
- Автосогласования частот и вариантов формата кадра нету
- Допустимые частоты получаются удвоением 300 бит/сек (600, 1200, 2400, 4800, 9600...)

UART 16550 для программиста (разработчика драйвера)

- 8 однобайтовых регистров («портов»), отображенных на последовательные адреса системной шины (адреса памяти)
- В компьютере может быть несколько UART, поэтому адрес первого регистра может различаться
- Как мы увидим далее, на некоторые адреса отображено два или три разных регистра
- Также имеет по 16 регистров буферов приема и передачи и сдвиговые регистры, которые последовательно передают и принимают биты, но они непосредственно программисту не доступны

Регистры UART 16550 (регистры 0-1)

- Регистр 0 (DLAB – это бит 7 в регистре 3)
  - Чтение (DLAB=0) – Receive Buffer Register (RBR), полученный байт
  - Запись (DLAB=0) – Transmit Holding Register (THR), байт который нужно передать
  - Чтение/запись (DLAB=1) - Divisor Latch LSB (DLL), младший байт делителя тактовой частоты – коэффициента, на который будет разделена базовая тактовая частота, 1.8432MHz в IBM PC
- Регистр 1
  - Чтение/запись (DLAB=0) - Interrupt Enable Register (IER), используется для управления прерываниями
  - Чтение/запись (DLAB=1) - Divisor Latch MSB (DLH), старший байт делителя тактовой частоты

Регистры UART 16550 (регистры 2-3)

- Регистр 2 (там много интересного, но в презентацию не влезет)
  - Запись - FIFO Control Register (FCR) – управление буферизацией
  - Чтение - Interrupt Identification Register – содержит описание причин прерывания (если UART его посылал) и состояние буфера
- Регистр 3 - Line Control Register (LCR)
  - Бит 7 - Divisor Latch Access Bit (DLAB), см. предыдущий слайд
  - Бит 6 – Set break
  - Биты 3-5 – биты управления битом четности.
  - Бит 2 – количество стоповых бит
  - Биты 0-1 – выбор количества бит данных, от 5 до 8 бит.

Регистры UART 16550 (регистры 4-5)

- Регистр 4 - Modem Control Register (MCR), управление так называемыми «модемными линиями». Подробности слишком ужасны, чтобы здесь их излагать. У Banana/Lichee Pi их все равно нету.
- Регистр 5 - Line Status Register (LSR)
  - Bit 7 → Error in Receiver FIFO
  - Bit 6 → Transmitter Empty (TEMT)
  - Bit 5 → Transmitter Holding Register Empty (THRE)
  - Bit 4 → Break Interrupt (BI)
  - Bit 3 → Framing Error (FE)
  - Bit 2 → Parity Error (PE)
  - Bit 1 → Overrun Error (OE)
  - Bit 0 → Data Ready (DR)

Регистры UART 16550 (регистры 6-7)

- Регистр 6 - Modem Status Register (MSR). Состояние «модемных линий»
- Регистр 7 – Scratch Register (SCR). Не влияет на работу UART. Хранит любые данные, которые в него записали

Регистры UART 16550
Interrupt Enable Register

- Биты 7-4 Зарезервированы, всегда 0
- Бит 3 → Enable Modem Status Interrupt (EDSSI). Разрешает прерывания по получению сигналов на «модемных линиях»
- Бит 2 → Enable Receiver Line Status Interrupt (ELSI) Разрешает прерывания по ошибкам или сигналу Break
- Бит 1 → Enable Transmitter Holding Register Empty Interrupt (ETBEI) Разрешает прерывание, когда Transmit Holding Register свободен (можно передать еще один байт)
- Bit 0 → Enable Received Data Available Interrupt (ERBFI) Разрешает прерывание по приходу байта или, если буферизация включена, по заполнению буфера до заданного уровня

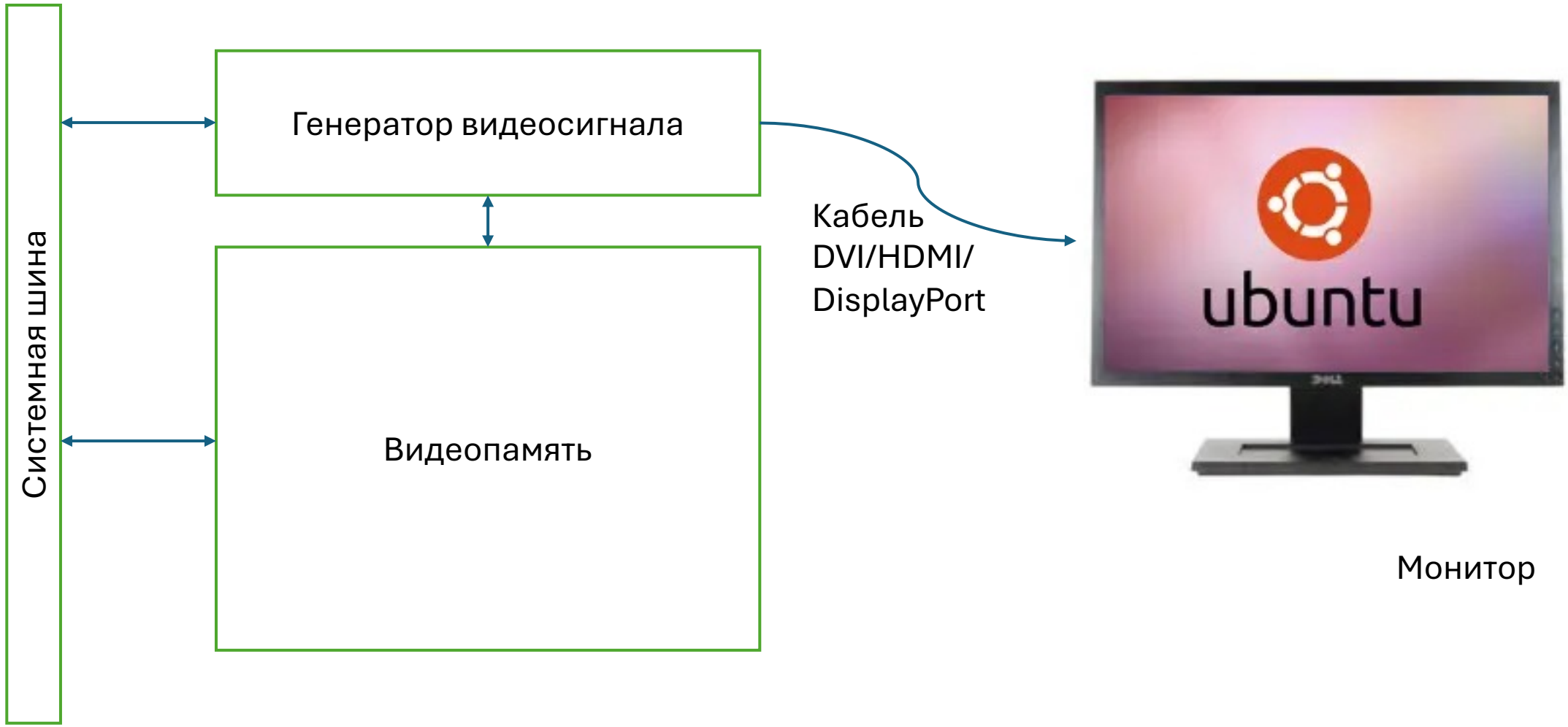
Что же со всем этим делать?

- Основные задачи – передать или получить байт – решаются просто
- Получить данные
  - Проверить бит Data Ready в Line Status Register
  - Пока Data Ready, прочитать байт из Receive Buffer Register
  - Если Data Ready = 0, настроить прерывание Received Data Available или сидеть в холостом цикле (обычно плохая идея)
- Отправить данные
  - Проверить бит Transmitter Holding Register Empty
  - Пока он 1, записать байт в Transmit Holding Register
  - Если он 0, настроить прерывание или сидеть в холостом цикле
- Много писанины требуют начальная настройка (особенно выбор частоты делителя) и обработка ситуаций, когда что-то идет не так

Какая из всего этого мораль?

- Даже простые устройства могут иметь много дополнительных функций и тонкостей
- Часто это связано с совместимостью со старыми устройствами и/или старыми диалектами протокола
- Тем не менее, по документации разобраться можно, а для популярных устройств можно найти примеры, например в коде драйверов Linux/BSD/U-boot

Более сложные устройства: видеоадаптер



Устройство видеоадаптера

- Видеопамять (видеобуфер) обычно отображается на адреса системной шины
- Дешевые интегрированные адаптеры используют часть системного ОЗУ в качестве видеобуфера
- Генератор видеосигнала (собственно видеоконтроллер) можно программно настраивать:
  - Спрашивать, какие к нему подключены мониторы
  - Спрашивать, сколько у него видеопамяти
  - Задавать видеорежим (разрешение, цветовую глубину, частоту кадра)
- Видеоданные последовательно, пиксел за пикселом, по строкам передаются на монитор или, в ноутбуках, прямо на матрицу дисплея
- Структура видеобуфера отличается большим разнообразием в зависимости не только от типа адаптера, но и от видеорежима

Структуры видеобуфера

- Текстовый режим: в видеобуфере лежат не пиксели, а буковки
  - В генераторе видеосигнала должен быть т.наз. знакогенератор с пиксельным шрифтом
  - Шрифт часто можно задавать программно
- Битовые плоскости
  - Монохромный дисплей без градаций яркости
    один байт кодирует 8 пикселов одной строки
  - Несколько битовых плоскостей (у дисплеев с 4 или 16 цветами)
- Байт на пиксел
  - Монохромный дисплей с градацией яркости
    (считается, что больше градаций человек на дисплее увидеть не может)
  - Indexed color - цветной дисплей с «палитрой»
    таблицей трансляции 256 значений в RGB
- Три и более байта на пиксел
  - True color – каждый пиксел описывается 3 байтами RGB

Размер видеобуфера

- 16-дюймовый дисплей Retina - 3456 × 2234 пикселов,
- 7.7 миллионов пикселов
- Если брать 4 байта на пиксел – около 30 мегабайт
- Даже адаптеры Banana/Lichee Pi имеют гораздо больше видеопамяти
- Зачем?

Графические ускорители

- В наше время, даже видеоадаптеры для телефонов содержат так называемый графический ускоритель
- Векторный процессор, главным образом предназначенный для синтеза 3D изображений из полигональных моделей
- Хранение этих моделей, текстур и промежуточных данных как раз и требует много памяти
- Может использоваться для других задач, сводимых к перемножению матриц: цифровой обработки сигналов, нейросетей, научных вычислений, майнинга криптовалют