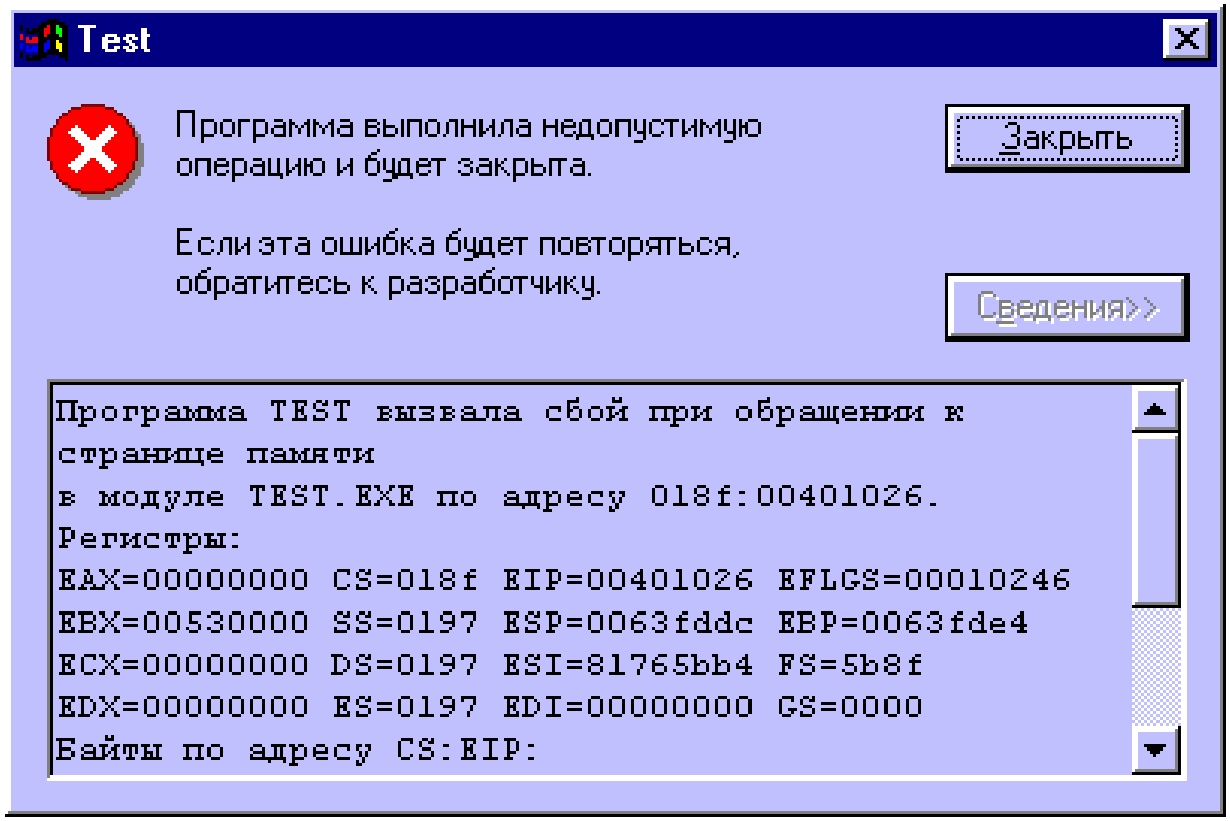
Виртуальная память

«Операционные системы»

Иртегов Д.В.

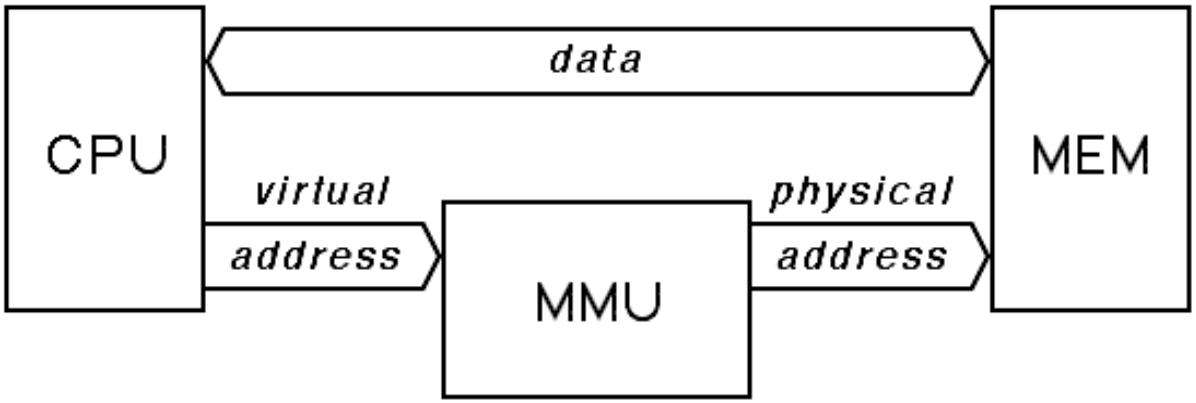
ФИТ НГУ

Защита памяти



Диспетчер памяти

- Memory Management Unit (MMU)
- Устройство управления памятью (УУП)



Базовая адресация

- DEC PDP6/PDP-10
- ICL 1900, «Одренок»

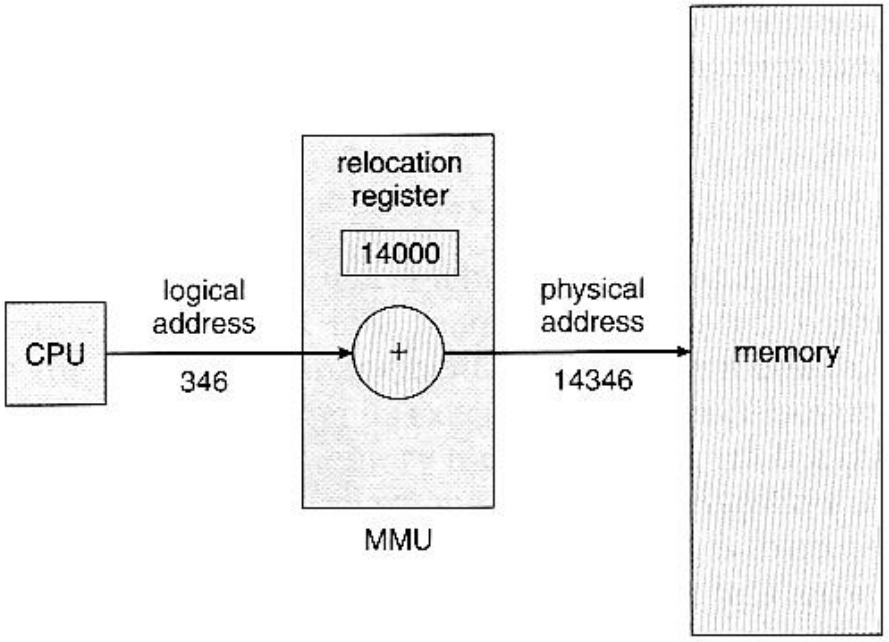
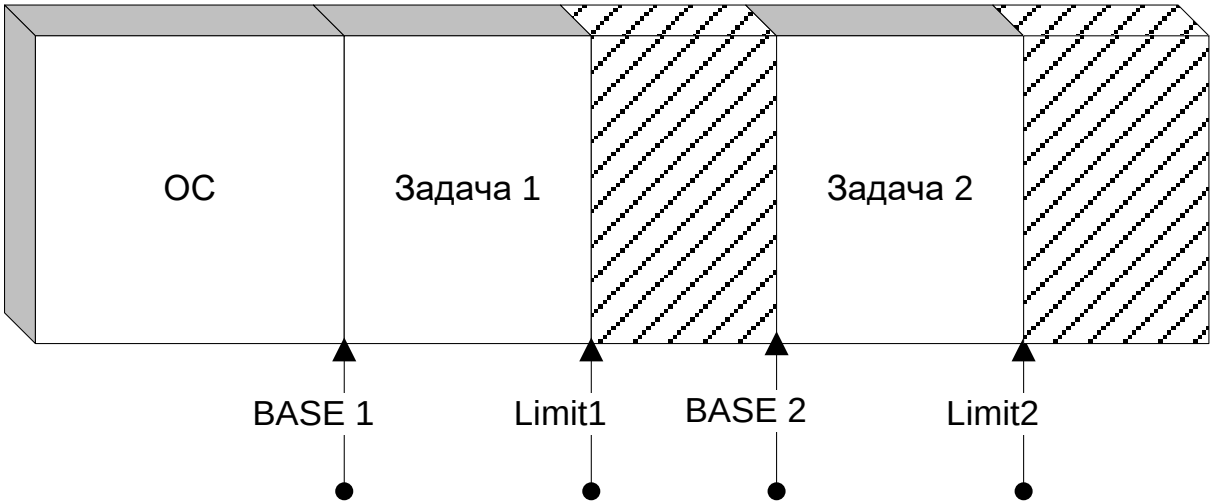


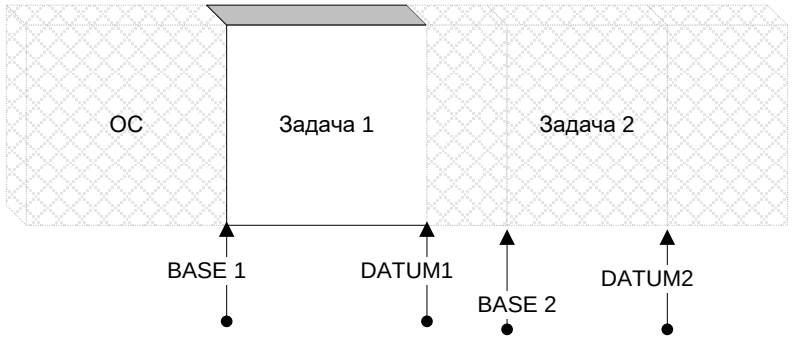
Figure 8.4 Dynamic relocation using a relocation register.

Базовая адресация (несколько задач)

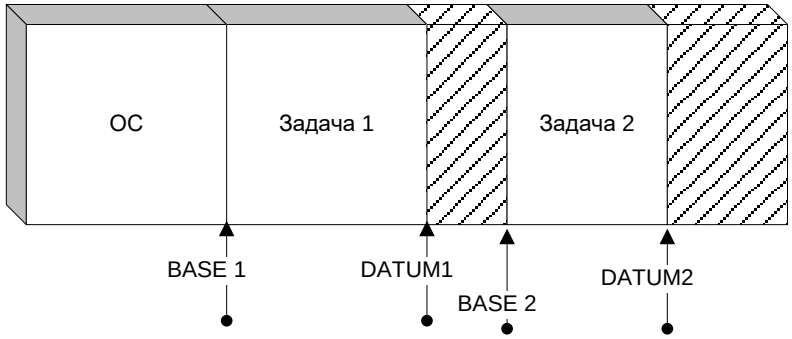


Системные вызовы

Пользовательский режим



Системный режим



Системный и пользовательский режимы

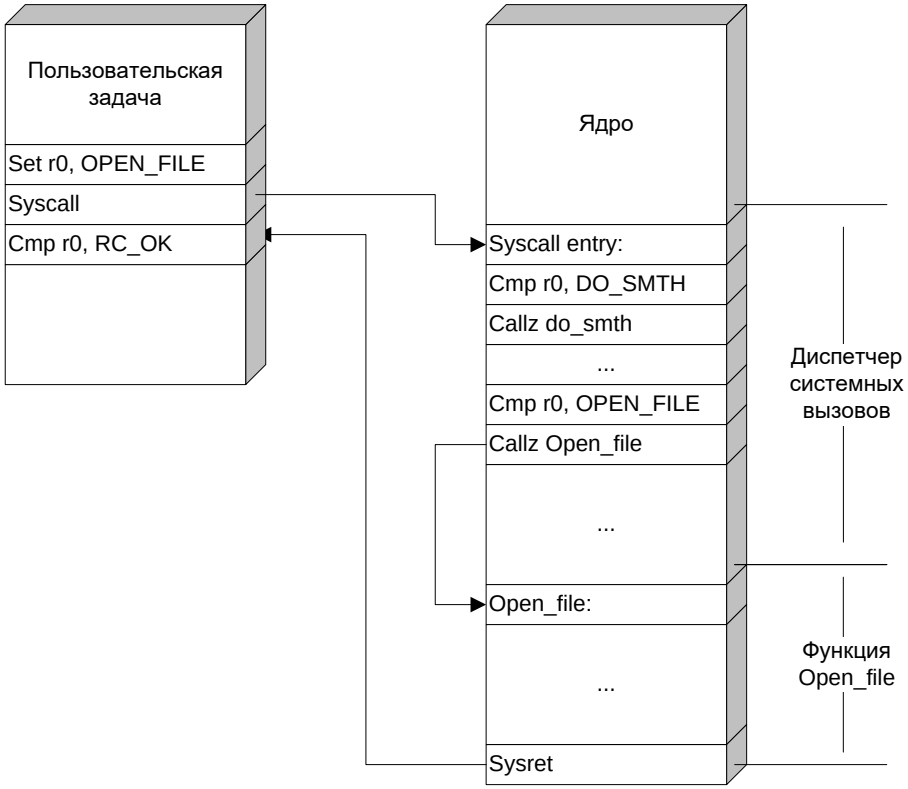
Система

- Может делать ввод/вывод
- Может изменять регистры MMU
- Иногда может вообще выключить MMU
- Имеет доступ к пользовательской памяти
- Может произвольно переходить в пользовательский режим

Пользователь

- Не может делать ввод/вывод
- Не может менять регистры MMU
- Не может выключить MMU
- Не имеет доступа к системной памяти
- Может перейти в системный режим только командой syscall

Системные вызовы



Чем плоха базовая адресация

- Образ задачи должен занимать непрерывную область памяти
  - Никаких разделяемых библиотек
  - Внешняя фрагментация (хотя можно проводить дефрагментацию)
  - Подкачка только задач целиком (task swapping)
  - Никакого отображения файлов на память

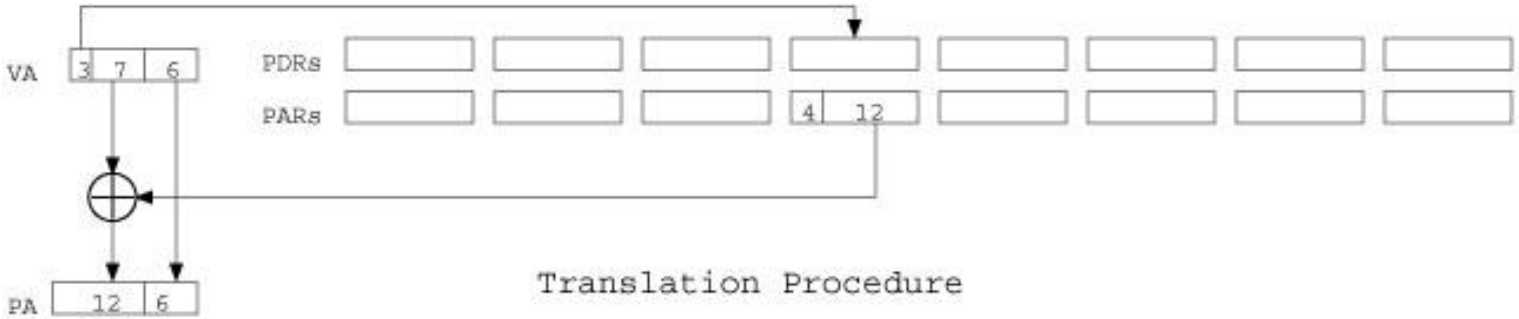
Развитие идеи

Base	Limit

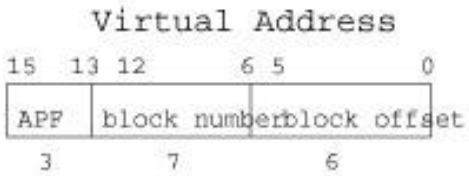
Base0	Limit0
Base1	Limit1
Base2	Limit2
Base3	Limit3
Base4	Limit4
Base5	Limit5
Base6	Limit6
Base7	Limit6

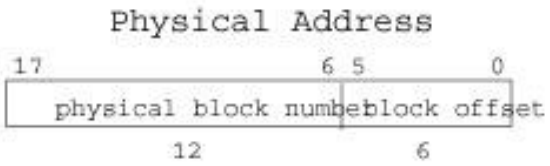
Адрес=Base[селектор]+смещение

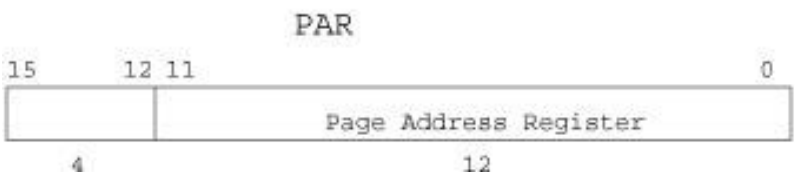
Диспетчер памяти PDP-11/40

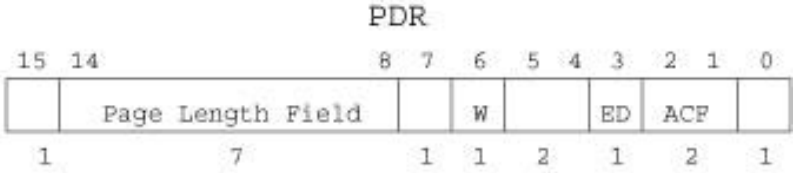


Translation Procedure

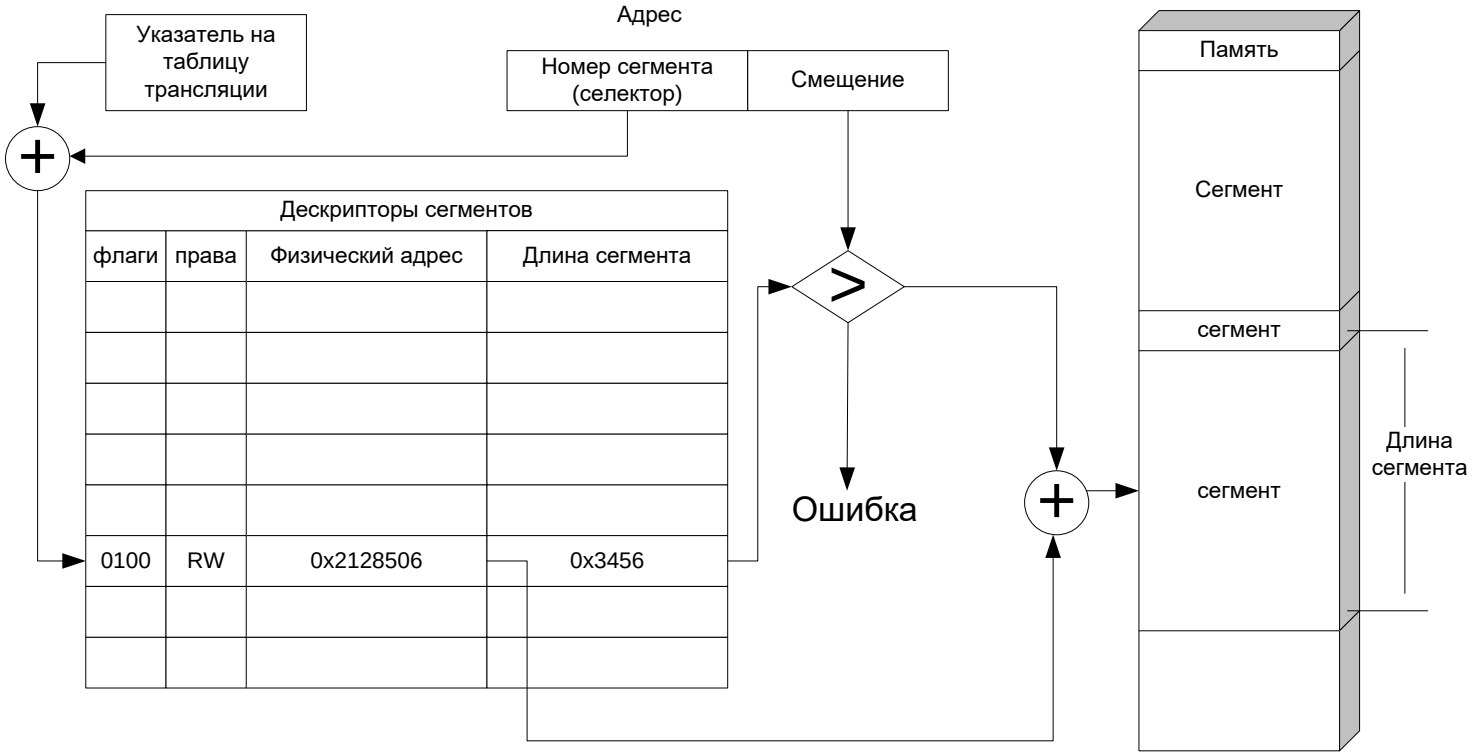




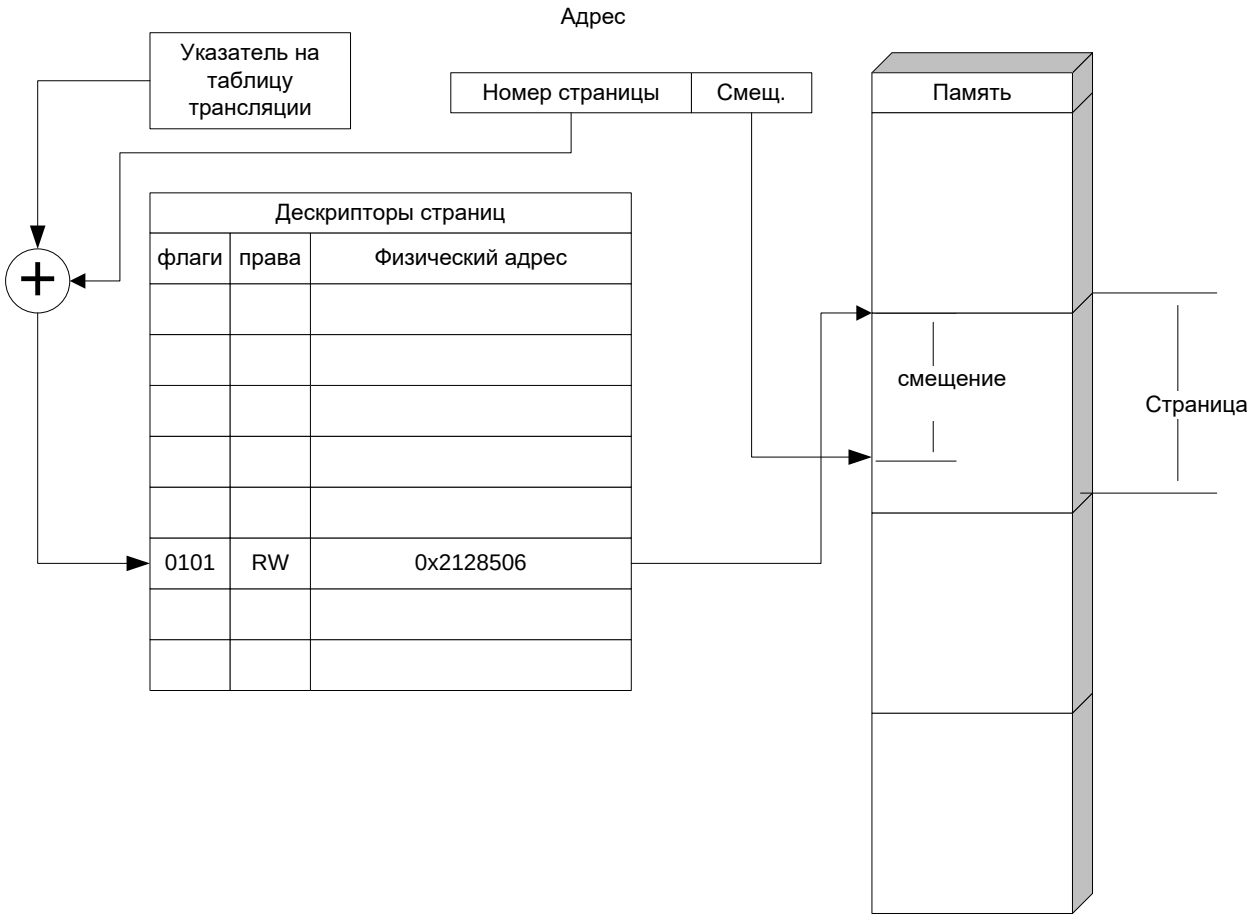




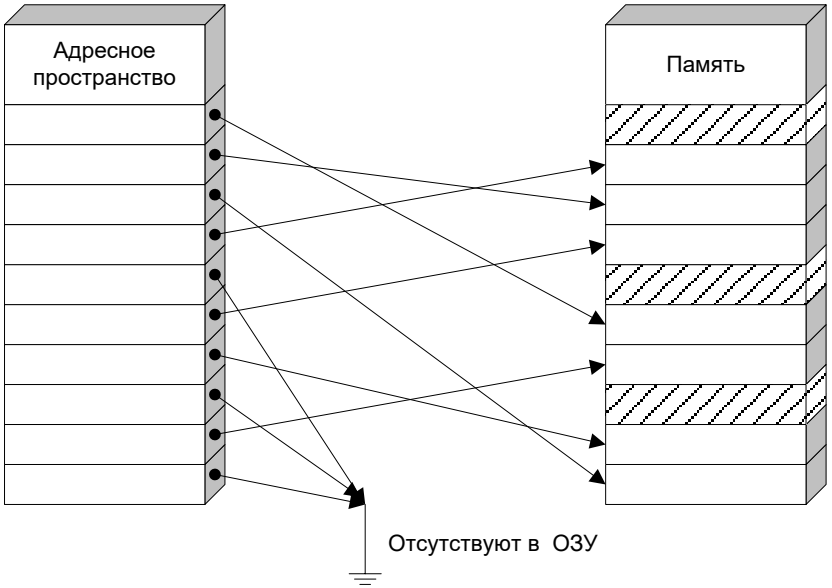
Восемь сегментов маловато



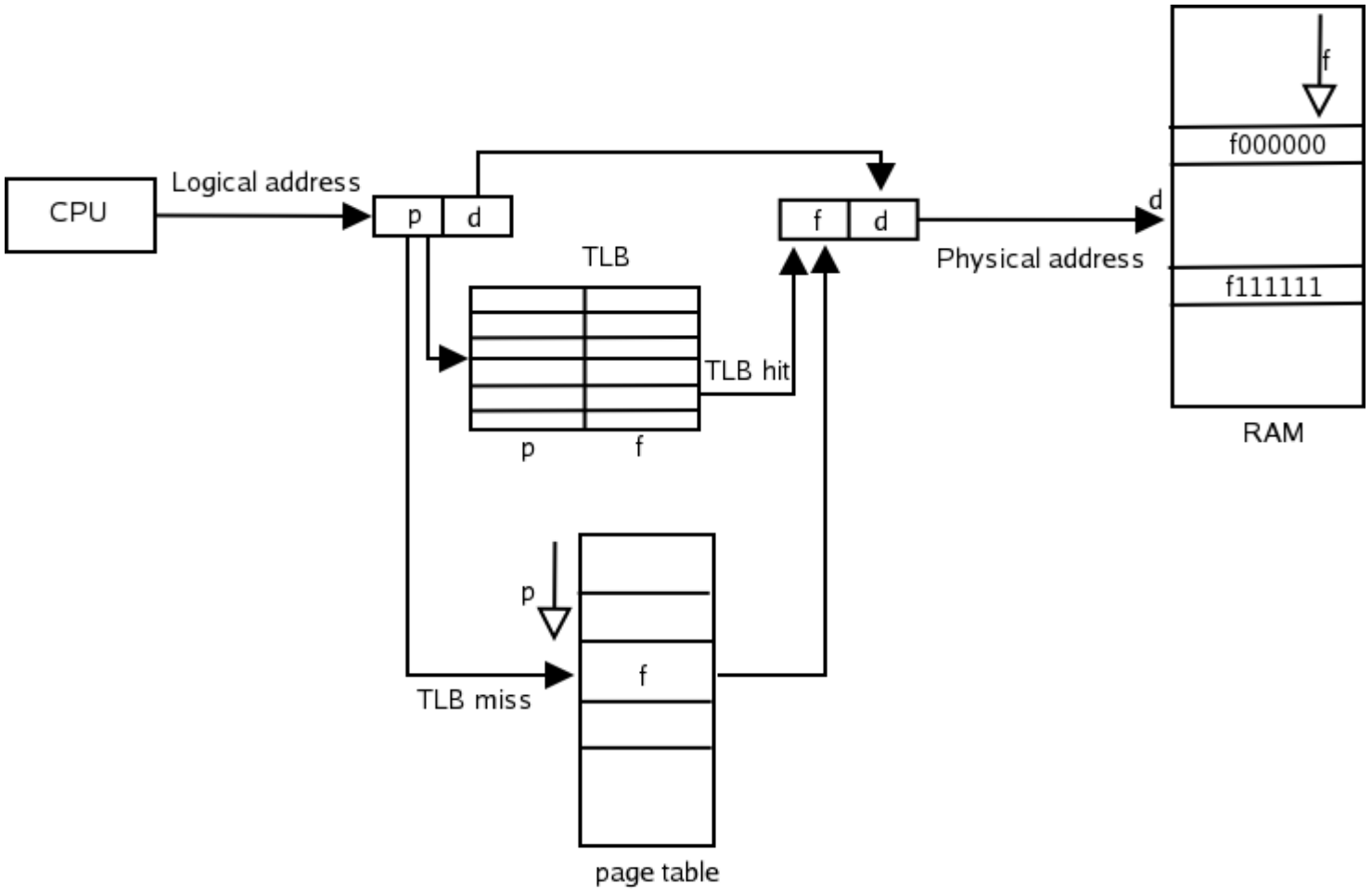
Страничный диспетчер



Отображение адресов



TLB
(Translation Lookaside Buffers)



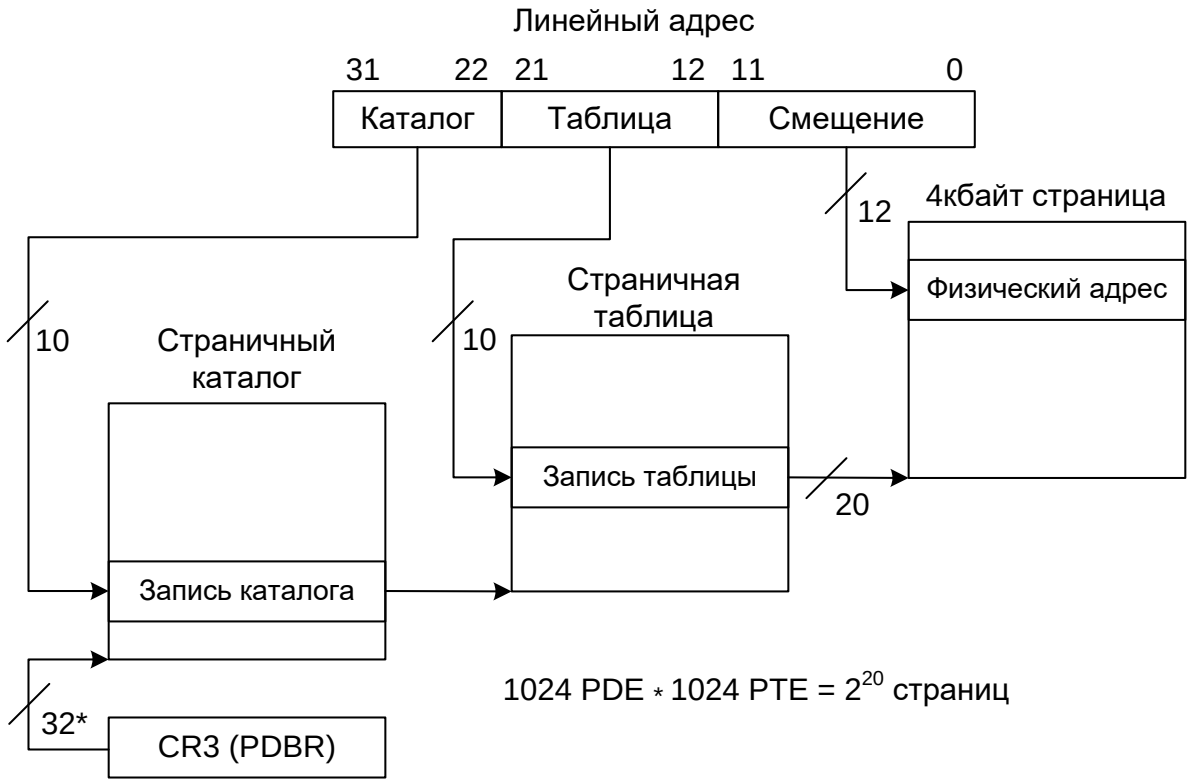
Что еще умеет диспетчер памяти?

- Защита чтения/записи
- Защита от исполнения (у x86 — начиная с P IV)
- Бит супервизора
- Бит присутствия
- Исключения
  - Ошибка защиты памяти
  - Страничный отказ

Дескриптор страницы 80386

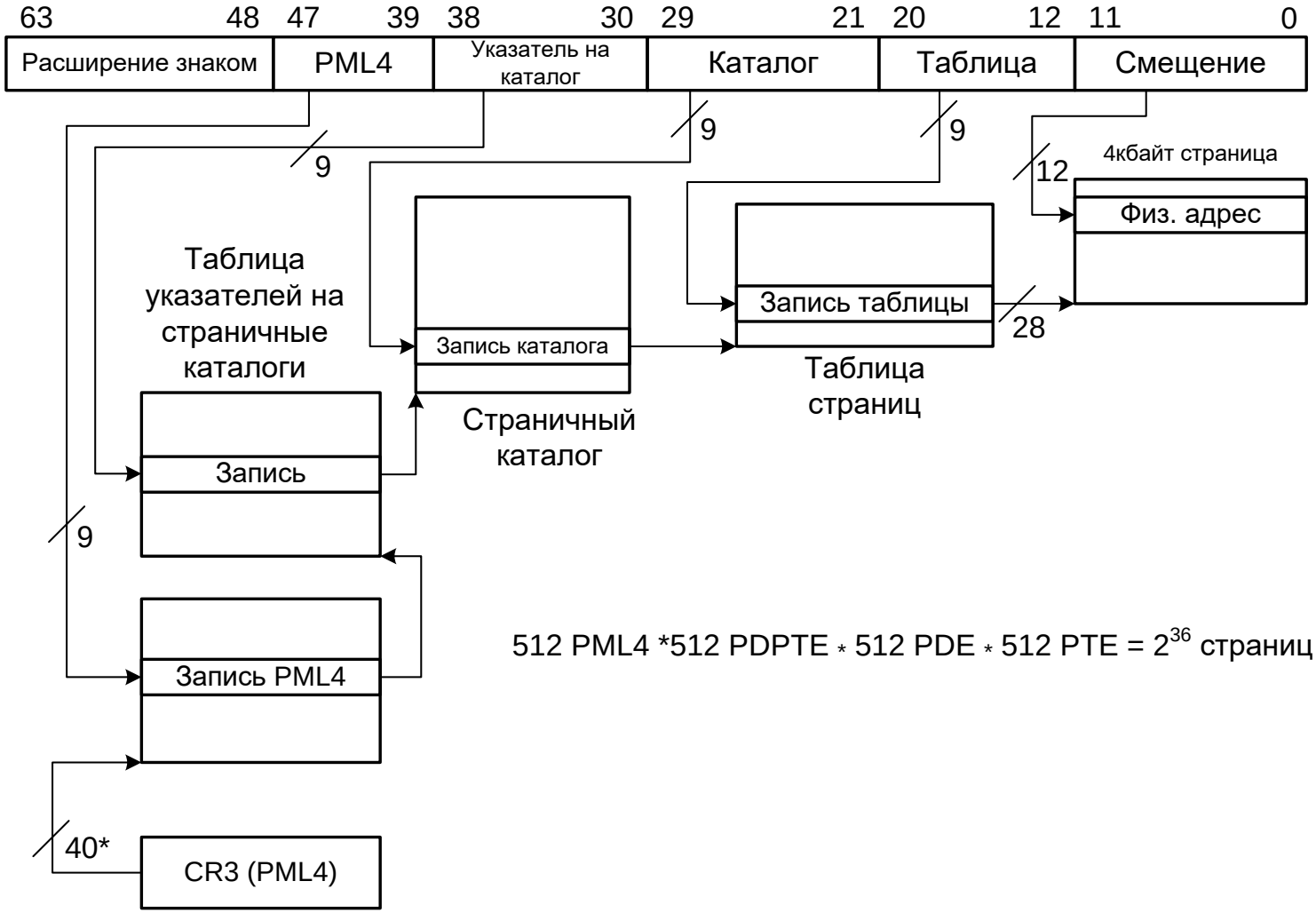
Запись таблицы страниц (4кбайт страницы)

Многоуровневая трансляция



*32 бита выровненные на 4кбайта

Многоуровневая трансляция x64



512 PML4 *512 PDPTE * 512 PDE * 512 PTE = 2^{36} страниц

* 40 бит выровненные на 4 килобайта

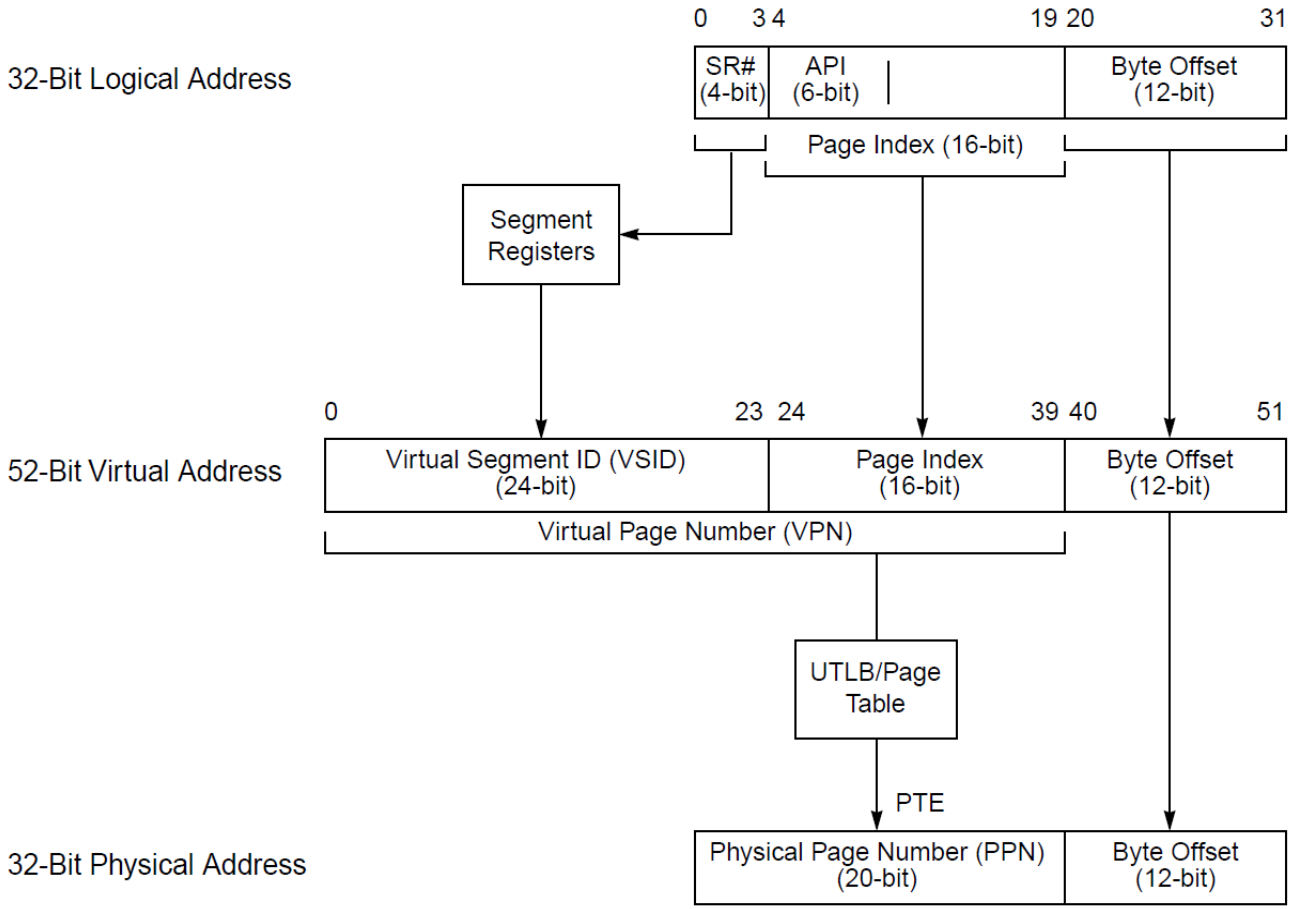
Переключение процессов

- Одна таблица трансляции (x86/x64, VAX)
  - При переключении задач надо переключить CR3 и сбросить TLB
  - Ядро должно отображаться в адресные пространства всех задач (защищено битом супервизора)

Переключение процессов

- Переключаемые таблицы трансляции (SPARC, PowerPC)
  - При переключении процессов достаточно переключить контекст (номер таблицы)
  - Записи TLB должны помнить номер таблицы
  - Зато TLB сбрасывать не надо!
  - Ядро может жить в отдельном адресном пространстве

Идентификатор контекста (на примере PowerPC 601)



Страничная подкачка (page swapping)

- В дескрипторе страницы есть бит отсутствия (или присутствия, зависит от машины)
- При обращении к такой странице, диспетчер генерирует исключение Page fault (страничный отказ)
- ОС ловит это исключение и думает, где взять страницу

Где взять страницу?

- Страница отображена на файл: прочитать из файла
- Страница анонимная (/dev/zero): выделить память и залить нулями
- Страница модифицированная приватная: прочитать из своп-раздела
- Страница никуда не отображена: Segmentation violation

Поиск жертвы

- Когда ОС не хватает памяти, она пытается отобрать редко используемые страницы у процессов
- Вместо статистики, «редкость» определяется грубыми эвристическими алгоритмами, например Clock algorithm
- Приватные модифицированные страницы сохраняются в своп-файл
- Разделяемые модифицированные страницы - в тот файл, из которого отображены

Mmap(2) и malloc(3C)

- Mmap просит память у системы и расширяет адресное пространство процесса
- Вызовы brk(2)/setbrk(2), увеличивают длину сегмента данных
  - В современных юниксах это обертка над mmap
- Malloc поддерживает собственный список свободных блоков в пользовательском адресном пространстве
- Когда свободных блоков нет, malloc(3C) делает mmap(2) или setbrk(2)