

Виртуальная память

Определение виртуальной
памяти

Виртуальная память

Опр. Виртуальная память (virtual memory) – концепция, позволяющая решить проблему ограниченной емкости оперативной памяти за счет предоставления каждому процессу виртуального адресного пространства (возможно, большего объема, чем объем оперативной памяти машины) для хранения данных и исполняемых инструкций.

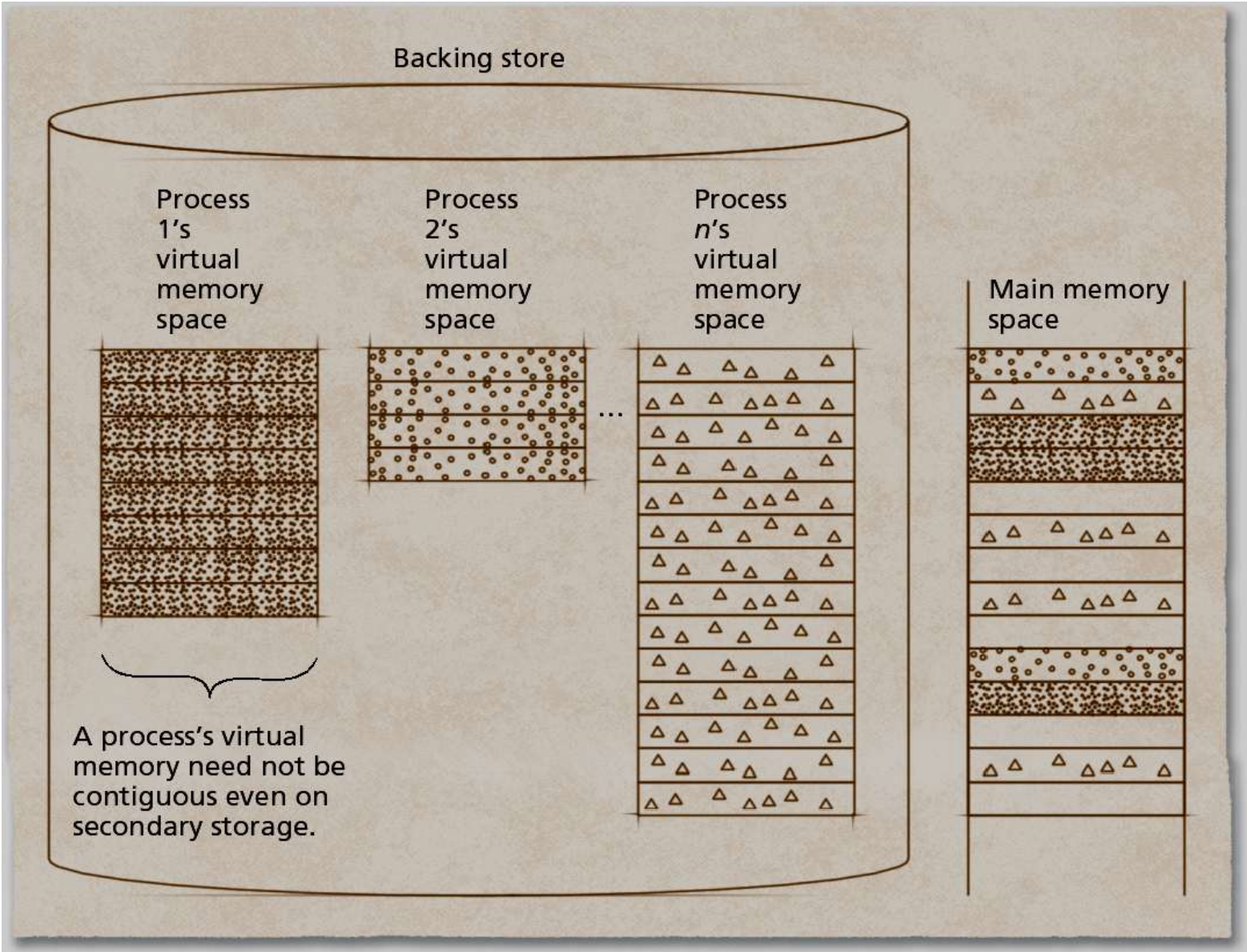
Виртуальный адрес

Опр. Виртуальный адрес (virtual address) – адрес, по которому процесс обращается к системе виртуальной памяти.

Виртуальное адресное пространство

Опр. Виртуальное адресное пространство (virtual address space) – множество виртуальных адресов, по которым может обращаться процесс.

Фрагменты адресных пространств в оперативной памяти и на вторичном устройстве хранения



Физический адрес

Опр. Физический адрес (physical address или real address) – адрес ячейки в оперативной памяти.

Физическое адресное пространство

Опр. Физическое адресное пространство (physical address space) – диапазон физических адресов, соответствующий объему оперативной памяти данного компьютера. Физическое адресное пространство может быть меньше, чем виртуальное адресное пространство.

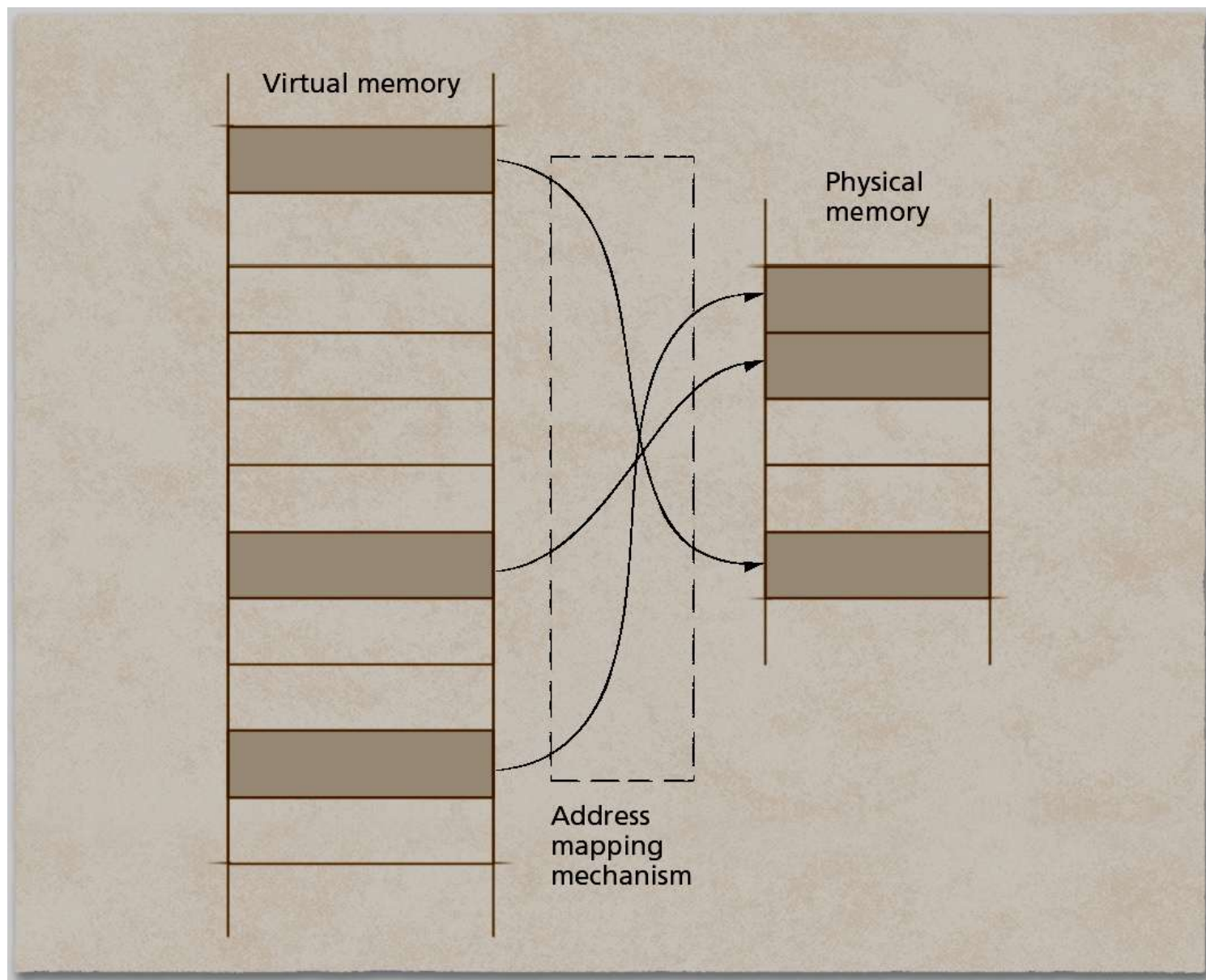
Динамическая трансляция адресов

Опр. Динамическая трансляция адресов (Dynamic Address Translation, DAT) – механизм, преобразующий виртуальные адреса в физические во время выполнения программы. Чтобы не замедлять выполнение, трансляция должна выполняться очень быстро.

Устройство управления памятью

Опр. Устройство управления памятью (Memory Management Unit, MMU) – специализированное аппаратное устройство, выполняющее, в частности, трансляцию виртуальных адресов в физические.

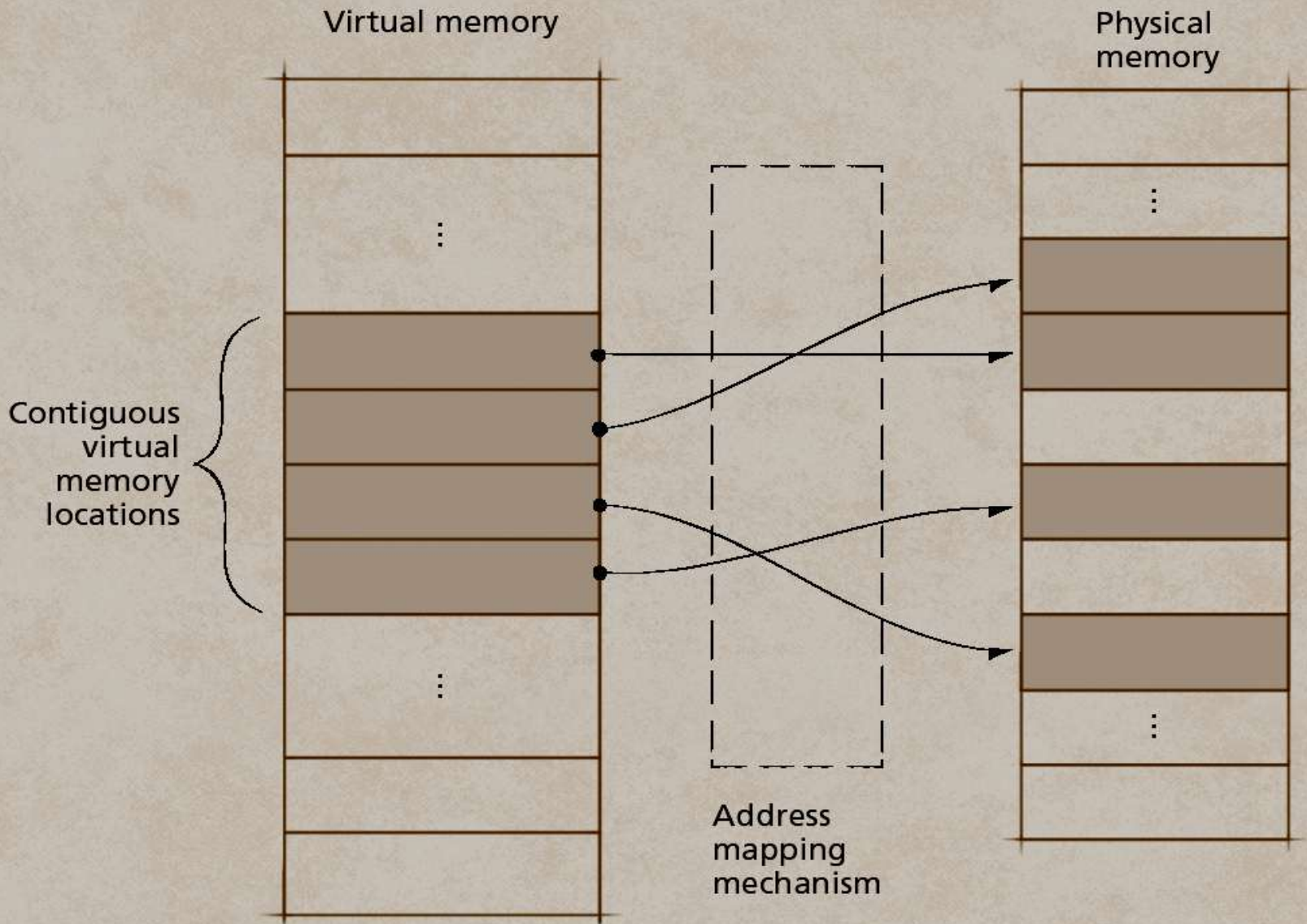
Преобразование виртуальных адресов в физические



Искусственная непрерывность

Опр. Искусственная непрерывность (artificial contiguity) – подход, используемый в системах виртуальной памяти, создающий для программ иллюзию хранения их инструкций и данных в непрерывных областях, хотя в реальности фрагменты кода и данных рассеяны по оперативной памяти. Искусственная непрерывность упрощает программирование.

Искусственная непрерывность



Вопрос для самопроверки

- Существуют ли программы, которые неэффективно загружать в память целиком перед исполнением? (Да/Нет)

Вопрос для самопроверки

- Существуют ли программы, которые неэффективно загружать в память целиком перед исполнением? (Да/Нет)
- Да. Например, во многих программах есть функции обработки ошибок, которые редко используются, если используются вообще. Загрузка этих функций в память уменьшает объем памяти, доступный процессам.

Вопрос для самопроверки

- Эффективно ли решение проблемы ограниченности оперативной памяти наращиванием ее объема? (Да/Нет)

Вопрос для самопроверки

- Эффективно ли решение проблемы ограниченности оперативной памяти наращиванием ее объема? (Да/Нет)
- Нет. Покупка дополнительной оперативной памяти не всегда экономически оправдана. Лучше создать иллюзию наличия в системе объема памяти, заведомо большего, чем может понадобиться программам.

Вопрос для самопроверки

- Совпадают ли виртуальное адресное пространство процесса и физическое адресное пространство системы? (Да/Нет)

Вопрос для самопроверки

- Совпадают ли виртуальное адресное пространство процесса и физическое адресное пространство системы? (Да/Нет)
- Нет. Виртуальное адресное пространство процесса – это множество адресов, по которым процесс может обращаться к памяти, работая под управлением системы виртуальной памяти. Процессы не используют физических адресов, соответствующих реальным ячейкам физической памяти.

Вопрос для самопроверки

- Искусственная непрерывность решает проблему ограниченности оперативной памяти? (Да/Нет)

Вопрос для самопроверки

- Искусственная непрерывность решает проблему ограниченности оперативной памяти? (Да/Нет)
- Нет. Искусственная непрерывность упрощает программирование, позволяя процессу работать с его памятью как с непрерывной областью, даже если данные и код рассеяны по оперативной памяти.

Виртуальная память

Размещение блоков

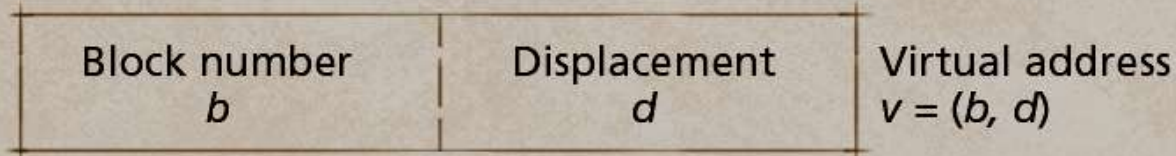
Блок

Опр. Блок (block) – область пространства памяти (реального или виртуального), представляющая собой диапазон смежных адресов.

Смещение

Опр. Смещение (displacement или offset) – разность между адресом элемента данных и адресом начала его блока.

Формат виртуального адреса в системе с блочным размещением



Номер блока Смещение

Таблица размещения блоков

Опр. Таблица размещения блоков (block map table) – таблица, содержащая записи о размещении виртуальных блоков процесса в блоках реальной памяти.

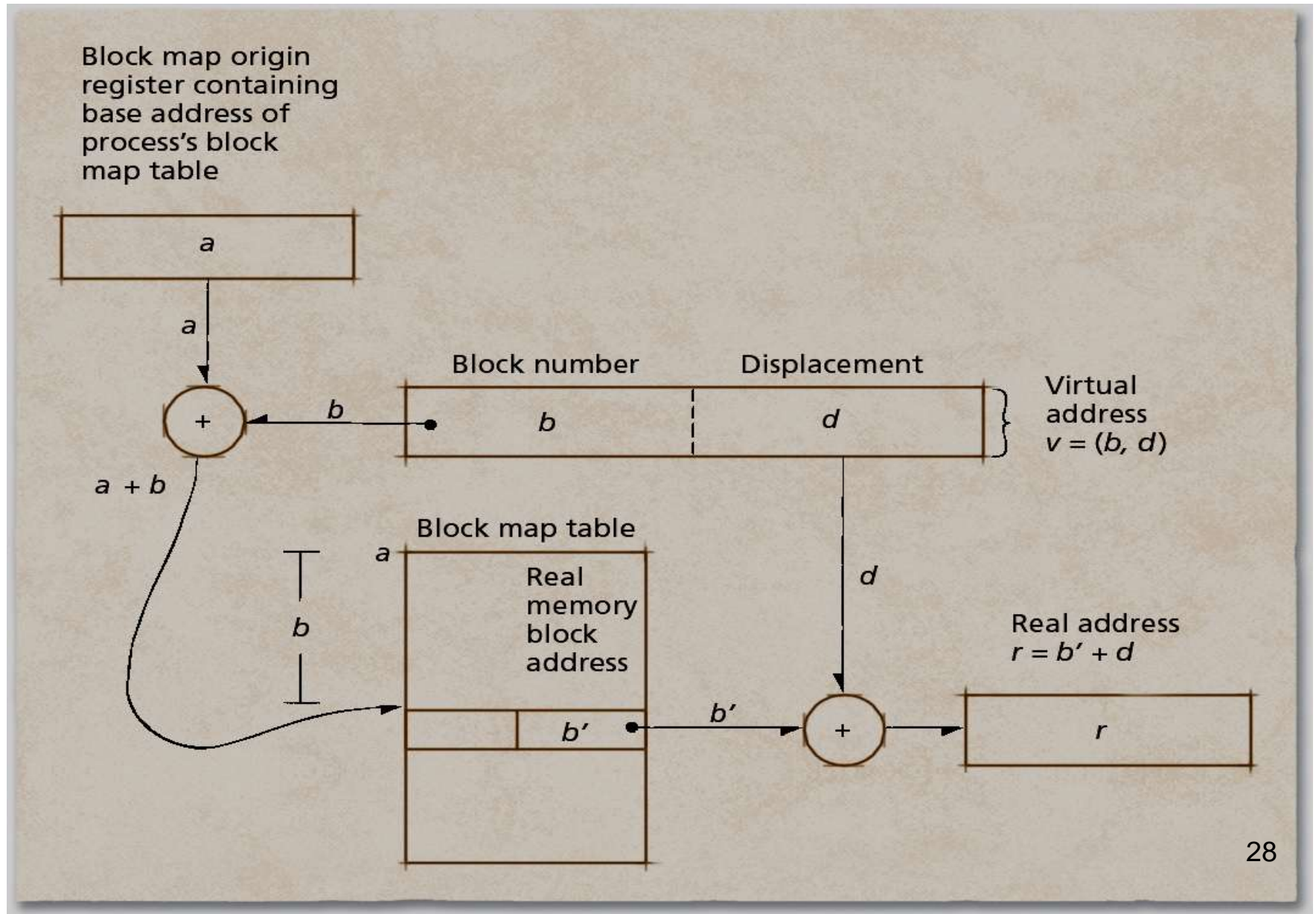
Таблица размещения блоков

- Находится в оперативной памяти, а ее записи обычно загружаются в кэш-память перед использованием
- Если бы размер блока совпадал с размером ячейки оперативной памяти, то таблица размещения блоков занимала бы больше места, чем доступно в оперативной памяти
- Блоки одного размера называются страницами, разного – сегментами

Регистр адреса таблицы размещения блоков

Опр. Регистр адреса таблицы размещения блоков (block map table origin register) – регистр, в котором хранится адрес, по которому в оперативной памяти находится таблица размещения блоков процесса. Этот быстродействующий регистр обеспечивает быструю трансляцию виртуальных адресов.

Преобразование виртуальных адресов в физические в системе с блочным размещением



Вопрос для самопроверки

- Значение регистра адреса таблицы размещения блоков изменяется при переключении контекста? (Да/Нет)

Вопрос для самопроверки

- Значение регистра адреса таблицы размещения блоков изменяется при переключении контекста? (Да/Нет)
- Да. Каждый процесс имеет свою таблицу размещения блоков. При переключении контекста система определяет реальный адрес, соответствующий адресу таблицы размещения блоков нового процесса в оперативной памяти.

Вопрос для самопроверки

- Большие блоки памяти лучше маленьких?
(Да/Нет)

Вопрос для самопроверки

- Большие блоки памяти лучше маленьких?
(Да/Нет)
- Нет. Чем больше средний размер блока, тем меньше нужно хранить информации о размещении блоков. Однако использование больших блоков может привести к значительной внутренней фрагментации, кроме того, такие блоки требуют много времени на перемещение между оперативной памятью и вторичными устройствами хранения.

Вопрос для самопроверки

- Должны ли все блоки памяти быть одинакового размера? (Да/Нет)

Вопрос для самопроверки

- Должны ли все блоки памяти быть одинакового размера? (Да/Нет)
- Нет. Если размер всех блоков одинаков, то блоки называют страницами. Если блоки могут быть произвольных размеров, они называются сегментами. В некоторых системах оба подхода совмещаются и сегменты состоят из страниц.

Виртуальная память

Страничные системы

Страница

Опр. Страница (page) – определенного размера участок виртуального адресного пространства процесса, который управляется как единое целое. В страницах содержатся порции данных и/или кода процесса.

Формат виртуального адреса в страничной системе

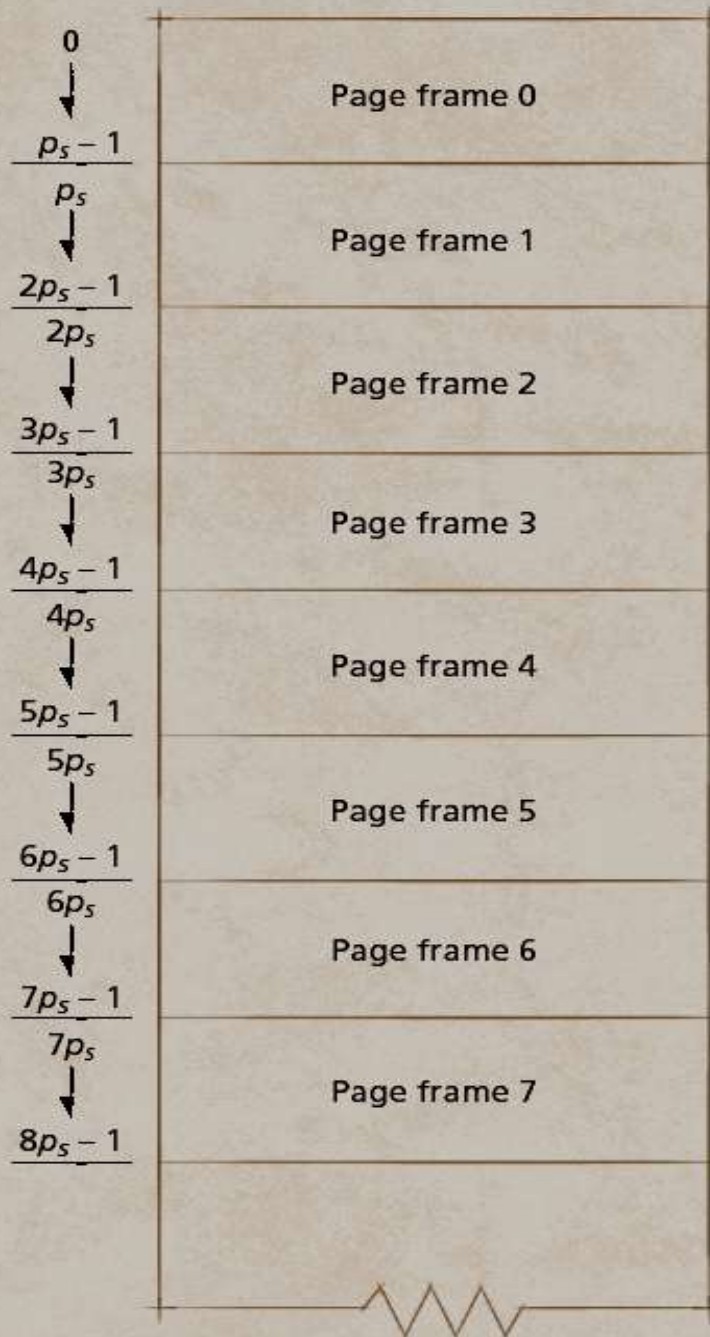


**Номер
страницы**

Смещение

Кадр страницы

Опр. Кадр страницы (page frame) – блок оперативной памяти, в котором может размещаться страница виртуальной памяти. Страницу можно поместить в любой доступный кадр.



Page frame number	Page frame size	Range of physical memory addresses
0	p_s	$0 \rightarrow p_s - 1$
1	p_s	$p_s \rightarrow 2p_s - 1$
2	p_s	$2p_s \rightarrow 3p_s - 1$
3	p_s	$3p_s \rightarrow 4p_s - 1$
4	p_s	$4p_s \rightarrow 5p_s - 1$
5	p_s	$5p_s \rightarrow 6p_s - 1$
6	p_s	$6p_s \rightarrow 7p_s - 1$
7	p_s	$7p_s \rightarrow 8p_s - 1$
\vdots		

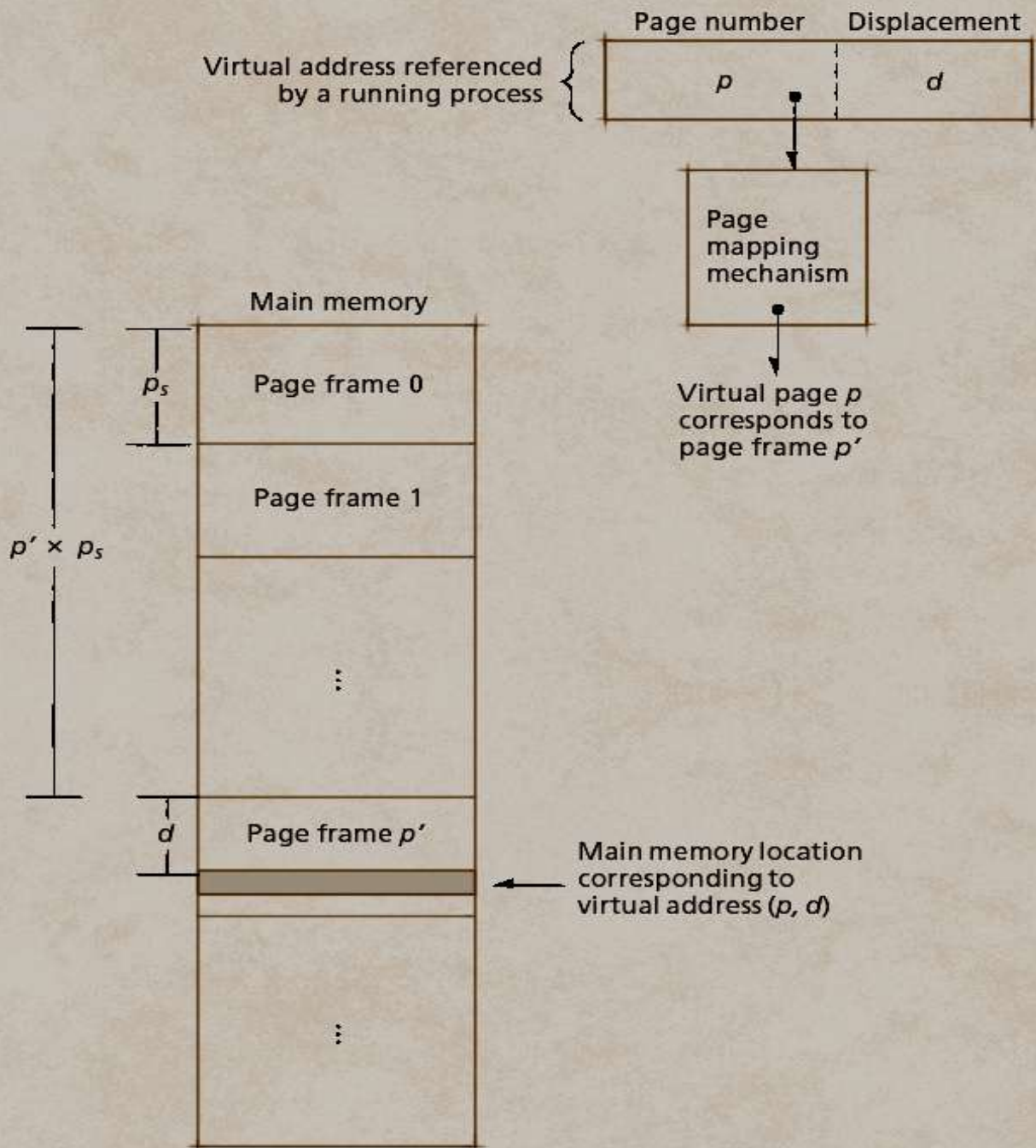
Оперативная память, разделенная на кадры страниц

Страничные системы

Опр. Страничные системы (paging systems) – системы виртуальной памяти, в которых она делится на фиксированного размера непрерывные блоки. В применении к виртуальному адресному пространству эти блоки называются страницами. В применении к реальному адресному пространству блоки называются кадрами страниц.

Страничные системы

- Страницы хранятся на вторичных устройствах хранения и при необходимости загружаются в кадры страниц в оперативной памяти
- Страничная организация упрощает решения о распределении памяти и не обладает внешней фрагментацией
- В страничных системах имеет место внутренняя фрагментация



**Соответствие
между
адресами
виртуальной
и физической
памяти в
страничной
системе**

Страничная таблица

Опр. Страничная таблица (page table или page map table) – таблица, в которой хранятся записи о номерах кадров, в которых размещаются страницы. В страничной таблице индексом является номер виртуальной страницы, и такая таблица содержит по одной записи для каждой страницы процесса.

Запись в страничной таблице

Опр. Запись в страничной таблице (Page Table Entry, PTE) – запись, в которой хранится номер кадра страницы, соответствующего странице виртуальной памяти. Кроме того, в этой записи хранится информация о том, находится ли эта страница в данный момент в памяти и разрешения на доступ к странице.

Запись в страничной таблице

Page
resident
bit

Secondary storage address
(if page is not in main
memory)

Page frame number
(if page is in
main memory)

r

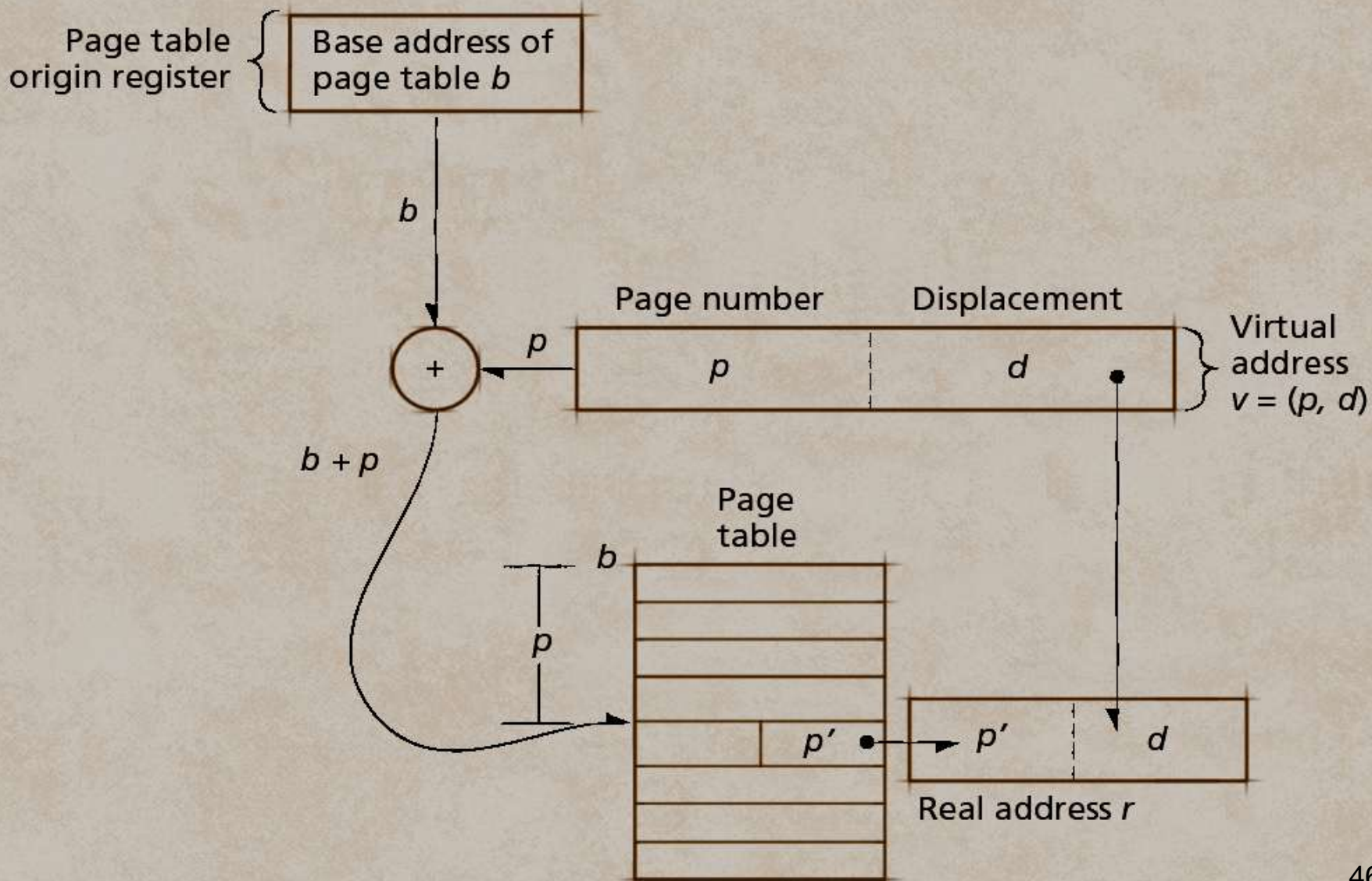
s

p'

$r = 0$ if page is not in main memory

$r = 1$ if page is in main memory

Трансляция адресов в страничных системах



Вопрос для самопроверки

- Требуется ли механизм отображения страниц, чтобы p' и s хранились в отдельных ячейках РТЕ? (Да/Нет)

Вопрос для самопроверки

- Требуется ли механизм отображения страниц, чтобы p' и s хранились в отдельных ячейках РТЕ? (Да/Нет)
- Нет. Чтобы уменьшить объем памяти, занимаемый РТЕ, многие системы используют только одно поле, значение которого воспринимается либо как номер кадра, либо как адрес на вторичном устройстве хранения в зависимости от значения бита резидентности.

Вопрос для самопроверки?

- Страница и кадр страницы – это одно и то же? (Да/Нет)

Вопрос для самопроверки?

- Страница и кадр страницы – это одно и то же? (Да/Нет)
- Нет. Хотя страницы и кадры страниц одинаковы по размеру. Страница – это постоянный по размеру блок виртуального адресного пространства процесса. Кадр страницы – это постоянный по размеру блок оперативной памяти.

Вопрос для самопроверки

- Записи в страничной таблице должны быть одинаковыми по размеру? (Да/Нет)

Вопрос для самопроверки

- Записи в страничной таблице должны быть одинаковыми по размеру? (Да/Нет)
- Да. Для реализации виртуальной памяти важно, чтобы трансляция адресов выполнялась как можно быстрее. Если записи в таблице одинаковы по размеру, процедура поиска нужной записи проста, и трансляция адресов будет выполняться быстро.

Вопрос для самопроверки

- Для трансляции адресов в страничных системах нужны специальные аппаратные устройства? (Да/Нет)

Вопрос для самопроверки

- Для трансляции адресов в страничных системах нужны специальные аппаратные устройства? (Да/Нет)
- Да. Для хранения базового адреса страничной таблицы нужен быстродействующий регистр в процессоре.

Виртуальная память

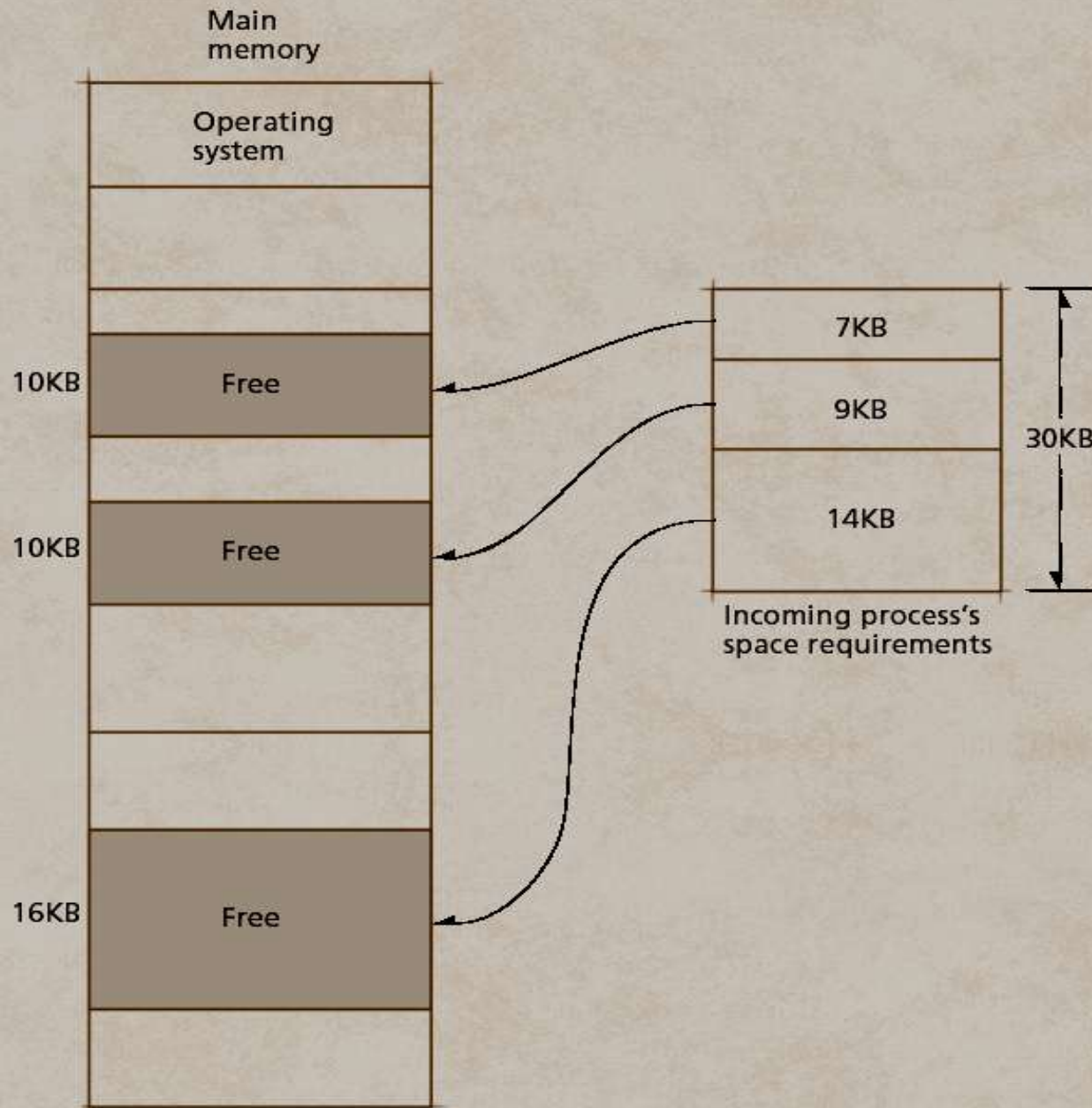
Сегментация

Сегмент

Опр. Сегмент (segment) – переменного размера непрерывная область виртуального адресного пространства, управляемая как единое целое. Размер сегмента обычно соответствует размеру части программы – например, процедуры или массива, что позволяет системе защищать такие части, задавая четкие правила доступа к ним.

Сегменты

- Сегменту данных обычно назначаются права доступа только для чтения или для чтения и записи, но не для исполнения
- Для сегмента, содержащего исполняемые инструкции, обычно назначается доступ для чтения и исполнения, но не для записи
- Использование сегментов обычно приводит к внешней фрагментации оперативной памяти, но не к внутренней



**Фрагменти-
рованное
выделение
памяти в
системе
реальной
памяти с
сегмента-
цией**

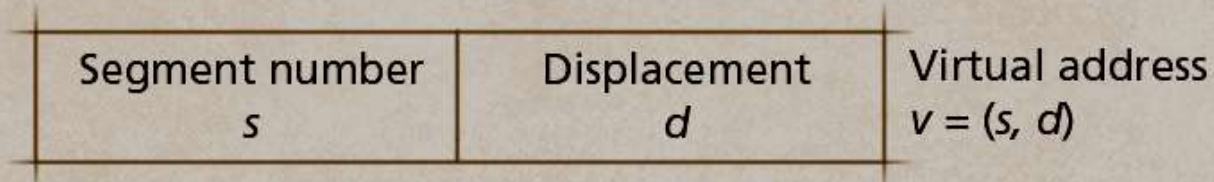
Системы виртуальной памяти с сегментацией

- В оперативной памяти хранятся только те сегменты, которые нужны программе для работы в данный момент
- Остальные сегменты могут находиться на вторичном устройстве хранения
- Если процесс обращается к памяти в сегменте, который находится на вторичном устройстве, система виртуальной памяти загружает его в основную память

Системы виртуальной памяти с сегментацией

- Сегменты переносятся из внешней памяти как неделимые единицы
- Все ячейки одного сегмента размещаются в непрерывной последовательности адресов в оперативной памяти
- Стратегии размещения сегментов аналогичны используемым в мультипрограммных системах с изменяемым распределением памяти

Формат виртуального адреса в сегментной системе



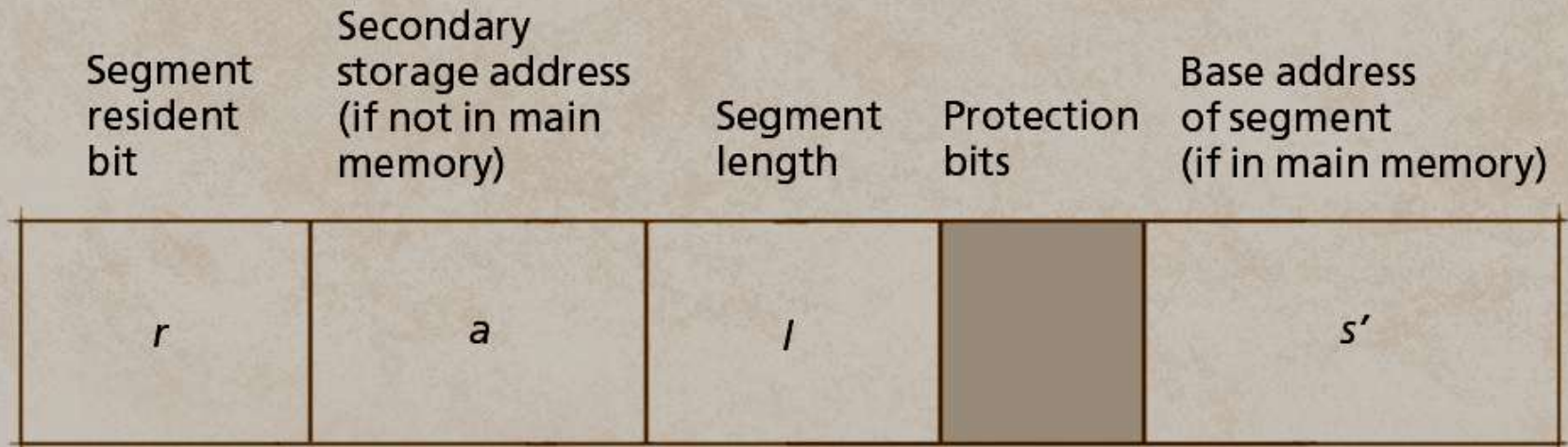
**Номер
сегмента**

Смещение

Биты защиты

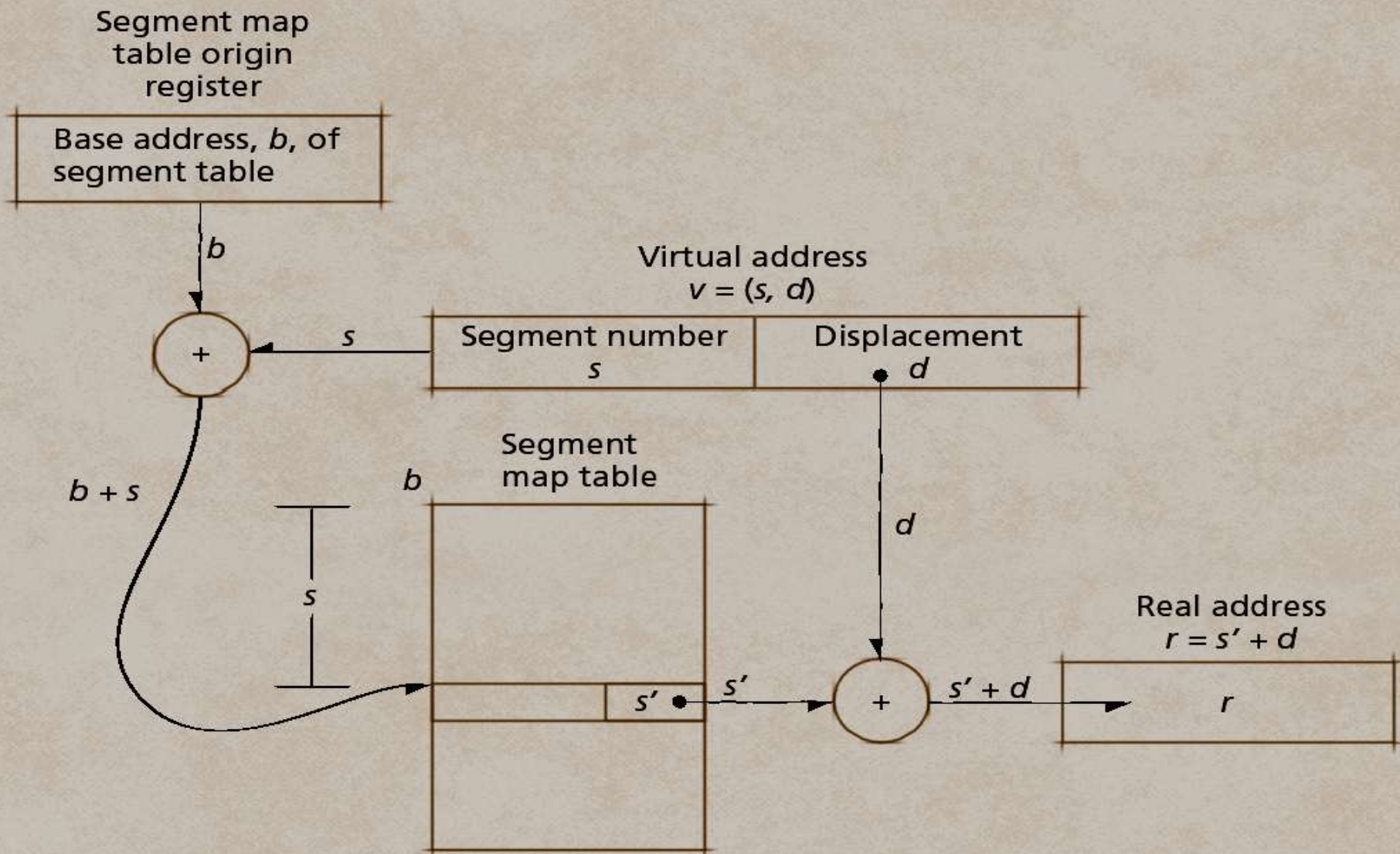
Опр. Биты защиты (protection bits) – биты в записи в сегментной таблице, значения которых ограничивают права доступа к сегменту.

Запись в сегментной таблице



↑
 $r = 0$ if segment is not in main memory
 $r = 1$ if segment is in main memory

Трансляция виртуального адреса в сегментной системе



Вопрос для самопроверки

- В сегментных системах виртуальной памяти отсутствует фрагментация?
(Да/Нет)

Вопрос для самопроверки

- В сегментных системах виртуальной памяти отсутствует фрагментация?
(Да/Нет)
- Нет. В сегментных системах виртуальной памяти может возникнуть внешняя фрагментация, точно так же, как в мультипрограммных системах с изменяемым распределением памяти.

Вопрос для самопроверки

- Сегментация отличается от мультипрограммности с изменяемым распределением памяти? (Да/Нет)

Вопрос для самопроверки

- Сегментация отличается от мультипрограммности с изменяемым распределением памяти? (Да/Нет)
- Да. В отличие от программ в мультипрограммных системах с изменяемым распределением памяти, программы в сегментных системах виртуальной памяти могут быть большего размера, чем оперативная память и находиться в нескольких областях памяти.

Вопрос для самопроверки

- Требуется ли механизм трансляции адресов в сегментной системе, чтобы s' и a хранились в отдельных ячейках записи в сегментной таблице? (Да/Нет)

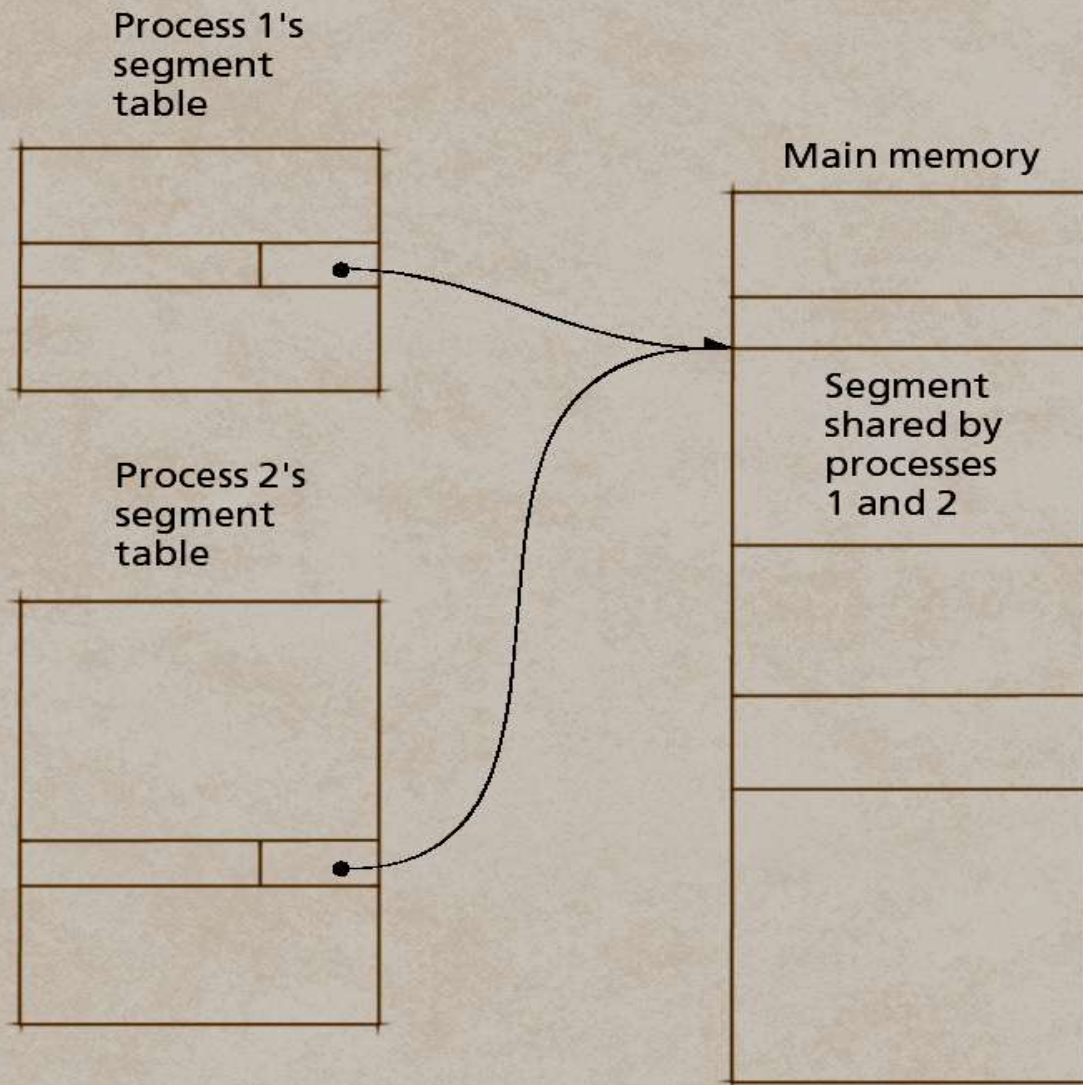
Вопрос для самопроверки

- Требуется ли механизм трансляции адресов в сегментной системе, чтобы s' и a хранились в отдельных ячейках записи в сегментной таблице? (Да/Нет)
- Нет. Чтобы уменьшить объем памяти, занимаемой сегментной таблицей (а она обычно хранится в кэше) эти параметры хранятся в одной ячейке содержимое которой интерпретируется в зависимости от значения бита резидентности.

Виртуальная память

Контроль доступа в сегментных
системах

Совместное использование ресурсов в сегментной системе



Право доступа

Опр. Право доступа (access right) – право, определяющее, к каким ресурсам может обращаться программа. В памяти права доступа определяют, к каким сегментам может обращаться процесс и каким образом.

Права доступа

- Для чтения (процесс может считывать данные из сегмента)
- Для записи (процесс может изменять содержимое сегмента)
- Для исполнения (процесс может исполнять инструкции, хранящиеся в сегменте)
- Для дополнения (процесс может дописывать в конец сегмента новые данные но не изменять записанные ранее)

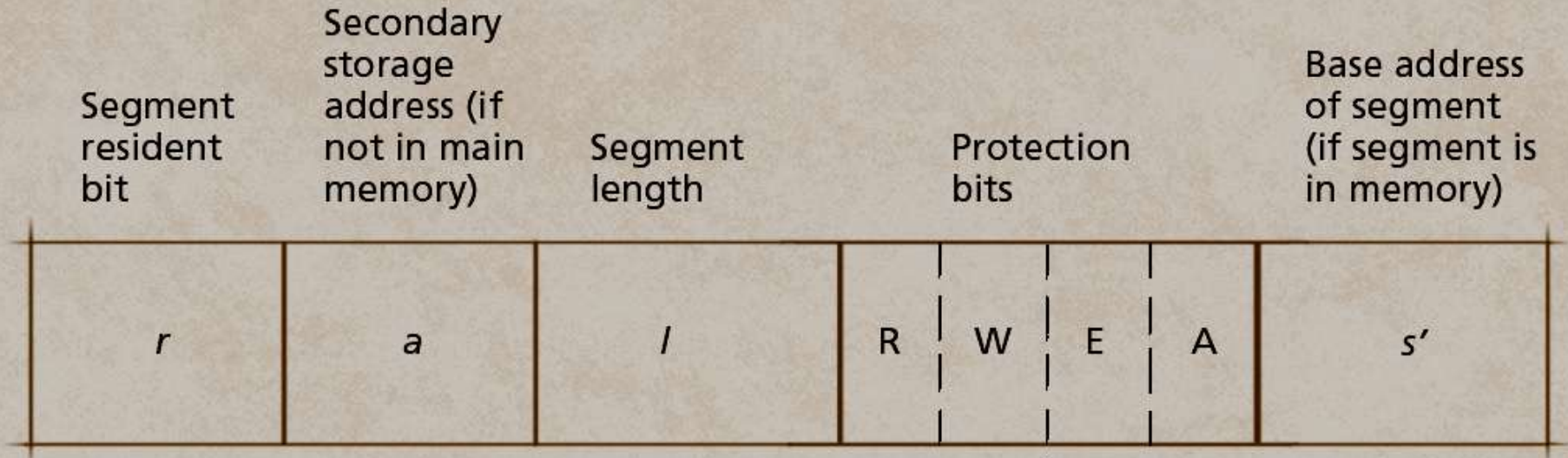
Права доступа

<i>Type of access</i>	<i>Abbreviation</i>	<i>Description</i>
Read	R	This segment may be read.
Write	W	This segment may be modified.
Execute	E	This segment may be executed.
Append	A	This segment may have information added to its end.

Режим контроля доступа

Опр. Режим контроля доступа (access control mode) – набор прав (на чтение, на запись, на выполнение, на дополнение), которые определяют, каким образом можно обращаться к сегменту памяти.

Запись в сегментной таблице с битами защиты



Protection bits: (1=yes, 0=no)

R–Read access

W–Write access

E–Execute access

A–Append access

Вопрос для самопроверки

- Может ли один сегмент памяти использоваться несколькими процессами? (Да/Нет)

Вопрос для самопроверки

- Может ли один сегмент памяти использоваться несколькими процессами? (Да/Нет)
- Да. Один сегмент памяти может совместно использоваться несколькими процессами. Для этого сегментные таблицы этих процессов могут ссылаться на один и тот же сегмент.

Вопрос для самопроверки

- Существует 16 полезных режимов контроля прав доступа процессов к сегментам памяти? (Да/Нет)

Вопрос для самопроверки

- Существует 16 полезных режимов контроля прав доступа процессов к сегментам памяти? (Да/Нет)
- Нет. По разному сочетая четыре права доступа можно создать 16 разных режимов контроля прав доступа. Некоторые из этих режимов весьма интересны, а другие – просто бессмысленны.

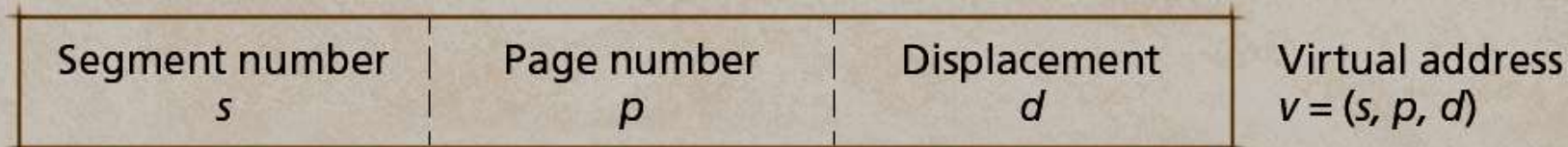
Виртуальная память

Сегментно-страничные системы

Сегментно-страничные системы

- Используются в большинстве современных систем
- Сочетают преимущества сегментных и страничных систем
- Сегменты состоят из множества страниц
- Не обязательно всем страницам сегмента находиться в оперативной памяти одновременно

Формат виртуального адреса в сегментно-страничной системе



**Номер
сегмента**

**Номер
страницы**

**Смещение
от начала
страницы**

Локальность

Опр. Локальность (locality) – эмпирическая закономерность, связывающая близкие во времени или в пространстве события.

Пространственная локальность

- В применении к последовательностям обращений к памяти пространственная локальность означает, что процесс, обращавшийся ранее по некоторому адресу, вероятно будет обращаться и по близким к нему адресам

Временная локальность

- Временная локальность означает, что процесс обращавшийся недавно по какому-то адресу, вероятно, скоро обратится к нему снова

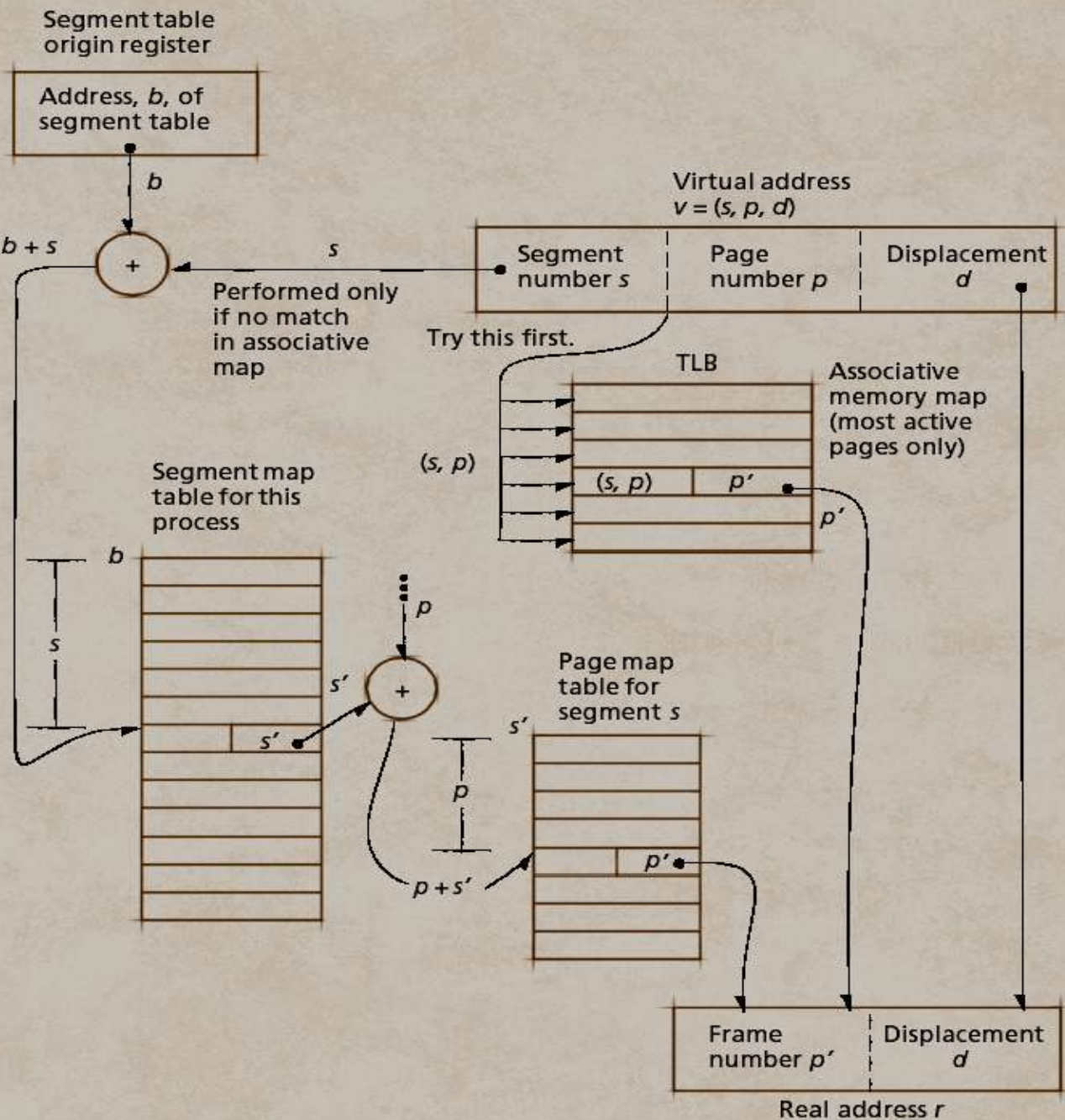
Буфер быстрой трансляции

Опр. Буфер быстрой трансляции (Translation Lookaside Buffer, TLB) – схема быстросдействующей ассоциативной памяти, в которой хранится небольшое количество записей о страницах виртуальной памяти и соответствующим им кадрам оперативной памяти.

Обычно в TLB хранятся данные о страницах, к которым недавно были обращения, что повышает производительность процессов, обладающих локальностью.

Ассоциативная память

Опр. Ассоциативная память (associative memory) – память, в которой поиск данных осуществляется по их значению, а не по адресу. Все записи в ассоциативной памяти просматриваются одновременно. Быстродействующая ассоциативная память помогает реализовать высокоскоростные механизмы динамической трансляции адресов.



Трансляция виртуальных адресов в сегментно-страничной системе

Вопрос для самопроверки

- Порождает ли циклическое обращение к массиву пространственную локальность?
(Да/Нет)

Вопрос для самопроверки

- Порождает ли циклическое обращение к массиву пространственную локальность? (Да/Нет)
- Да. Циклическое обращение к массиву порождает пространственную локальность, поскольку элементы массива обычно расположены в непрерывной области виртуальной памяти.

Вопрос для самопроверки

- Порождает ли циклическое обращение к массиву временную локальность?
(Да/Нет)

Вопрос для самопроверки

- Порождает ли циклическое обращение к массиву временную локальность?
(Да/Нет)
- Да. Временная локальность появляется потому, что элементы массива обычно намного меньше по размеру, чем страница. Поэтому обращения к двум соседним элементам массива, происходя за короткий промежуток времени, обычно направляются к одной и той же странице.

Вопрос для самопроверки

- Для сегментно-страничных систем необходимы специальные аппаратные устройства? (Да/Нет)

Вопрос для самопроверки

- Для сегментно-страничных систем необходимы специальные аппаратные устройства? (Да/Нет)
- Да. Для таких систем нужны: быстродействующие регистры для хранения базовых адресов сегментной таблицы и соответствующей страничной таблицы, а также схема ассоциативной памяти (TLB).

Вопрос для самопроверки

- Сегментно-страничным системам присуща внешняя фрагментация?
(Да/Нет)

Вопрос для самопроверки

- Сегментно-страничным системам присуща внешняя фрагментация?
(Да/Нет)
- Нет. В сегментно-страничных системах нет внешней фрагментации, поскольку, в конечном счете, вся память делится на кадры равного размера, в которых может помещаться любая страница.

Вопрос для самопроверки

- Сегментно-страничным системам присуща внутренняя фрагментация?
(Да/Нет)

Вопрос для самопроверки

- Сегментно-страничным системам присуща внутренняя фрагментация? (Да/Нет)
- Да. Это явление возникает тогда, когда объем сегмента меньше суммарного размера страниц, в которых он размещается. Чем меньше размер страниц, тем меньше памяти теряется на внутреннюю фрагментацию.