

Кэш память

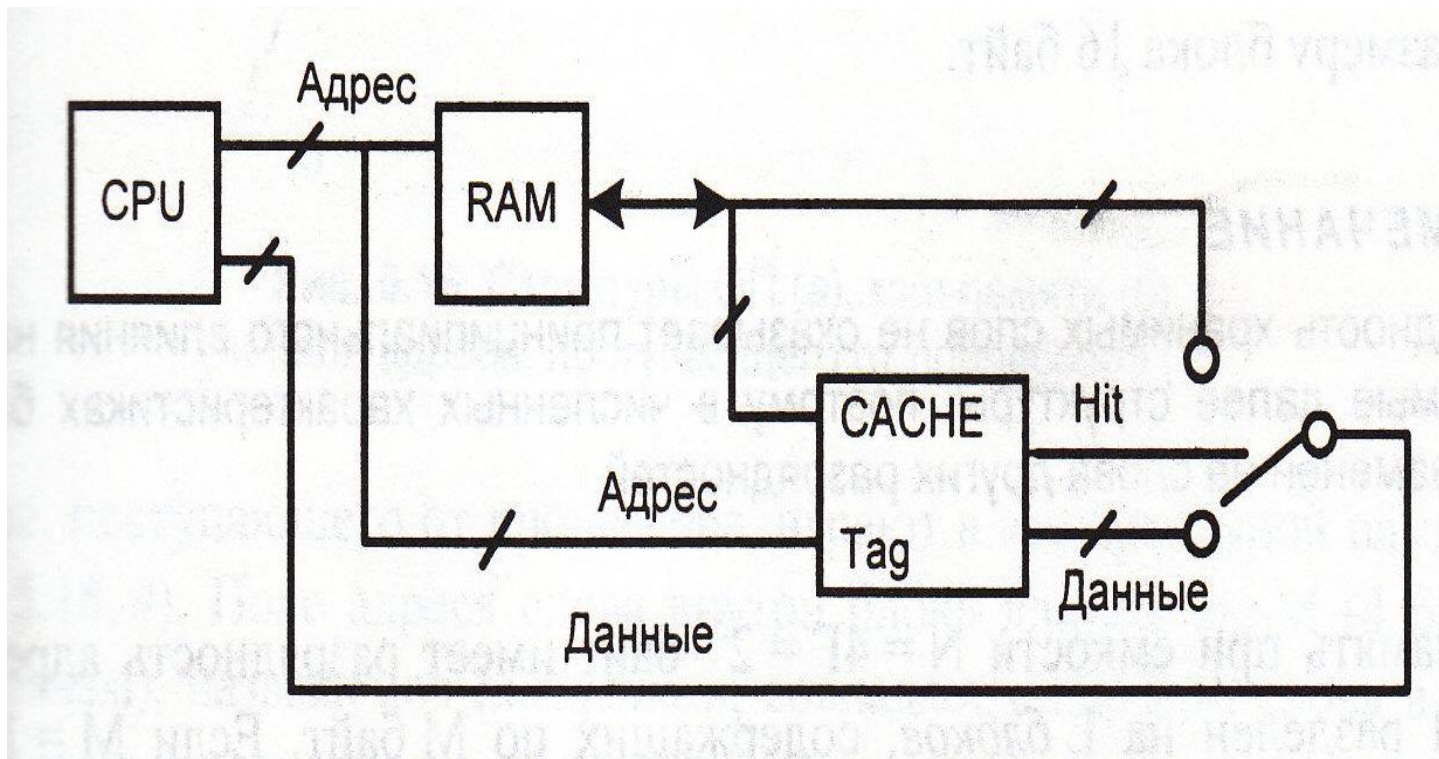
Цель лекции: рассмотреть назначение, принципы построения кэш-памяти, особенности ее использования в мультипроцессорных системах

Общие принципы функционирования КП

- **Кэш-память** – представляет собой организованную в виде ассоциативного запоминающего устройства, быстродействующую буферную память ограниченного объема, которая располагается между оперативной памятью и регистрами микропроцессора.
- В кэш памяти хранится информация чаще всего используемая в текущем цикле вычислений. Основа кэш памяти - принцип локальности ссылок.
- **Назначение** – повышение производительности системы за счет уменьшения количества циклов обращения к оперативной памяти.
- Кэш память скрыта от программиста и «невидима» для процессора.

Схема взаимодействия процессора ОЗУ и кэш-памяти

Процессор имея адрес нужной информации сначала обращается к кэш памяти. По тегу определяется Есть информация в кэш или нет. Если есть, то HIT – кэш попадание.



Оценка времени доступа к памяти при кэш попадании

$$T_{\text{ср}} = c + (1 - H) * m$$

$$H = (k-1) / k$$

С – время доступа к кэш памяти;
М – время доступа к основной памяти;
Н – коэффициент попаданий;

К – количество обращений к ячейке памяти;
(К – 1) количество обращений к кэш памяти;

При Н стремящемся к 1 все обращения идут к кэш памяти и $T_{\text{ср}} = c$.
При Н стремящемся к 0, $T_{\text{ср}} = c + m$.

Факторы влияющие на эффективность применения кэш памяти

- Емкость кэш-памяти.
- Формат строки.
- Способ отображения основной памяти на кэш-память.
- Алгоритм замещения информации в кэш-памяти.
- Алгоритм согласования содержимого кэш и оперативной памяти.
- Число уровней кэш-памяти.
- Схемотехническая реализация кэш-памяти.

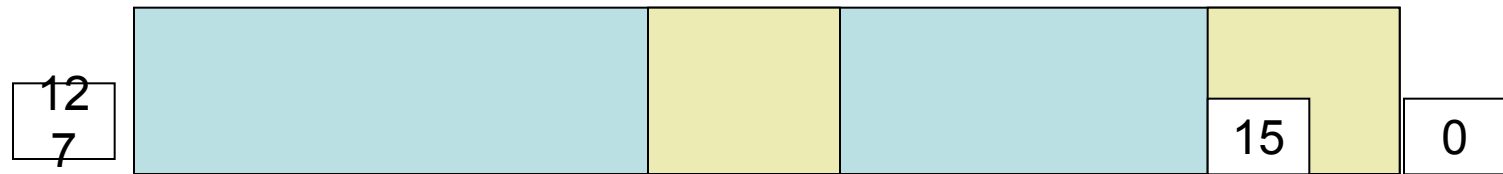
Термины

- Кэш промах.
- Кэш попадание.
- Размер кэш – памяти.
- Строка кэш памяти.



16 Кб = 1 К строк по 16 байт

Строка кэш-памяти
16-64 байт



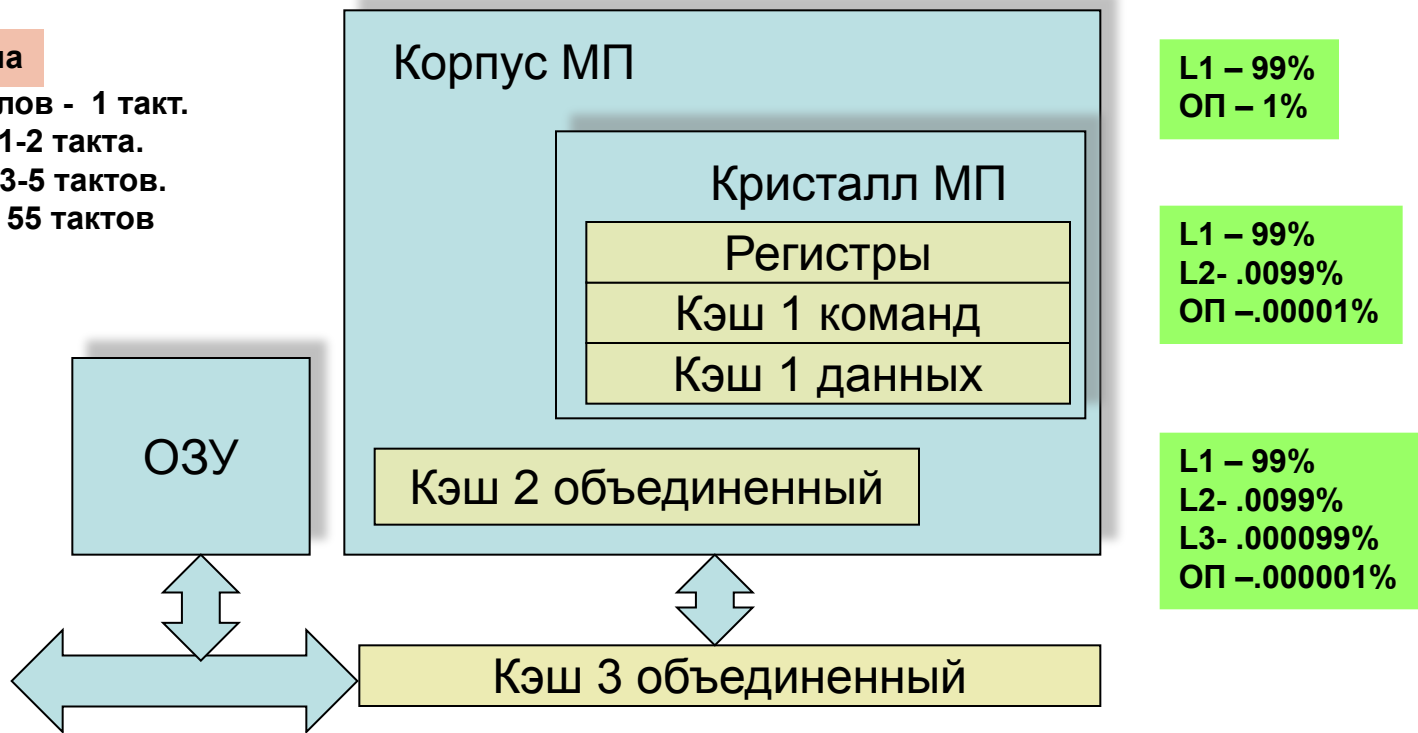
Стоимость !!!

Востребованная
информация

Расположение кэш памяти

Время доступа

Регистры 64-256 слов - 1 такт.
Кэш 1 32 Кбайт - 1-2 такта.
Кэш 2 256 Кбайт - 3-5 тактов.
ОЗУ 4 Гбайт - 12 - 55 тактов



Гарвардская архитектура – разделение на кэш команд и кэш данных.

Кэш – память статического типа.

Принципы построения многоуровневой кэш-памяти

- **Инклюзивный метод.** Вся информация из верхних уровней содержится в нижних уровнях. Быстродействие.
- **Эксклюзивный метод.** Блок удаляемый из кэш более высокого уровня не дублируется в кэш нижнего уровня. Эффективное использование емкости.

Классификация КЭШ по способу отображения основной памяти

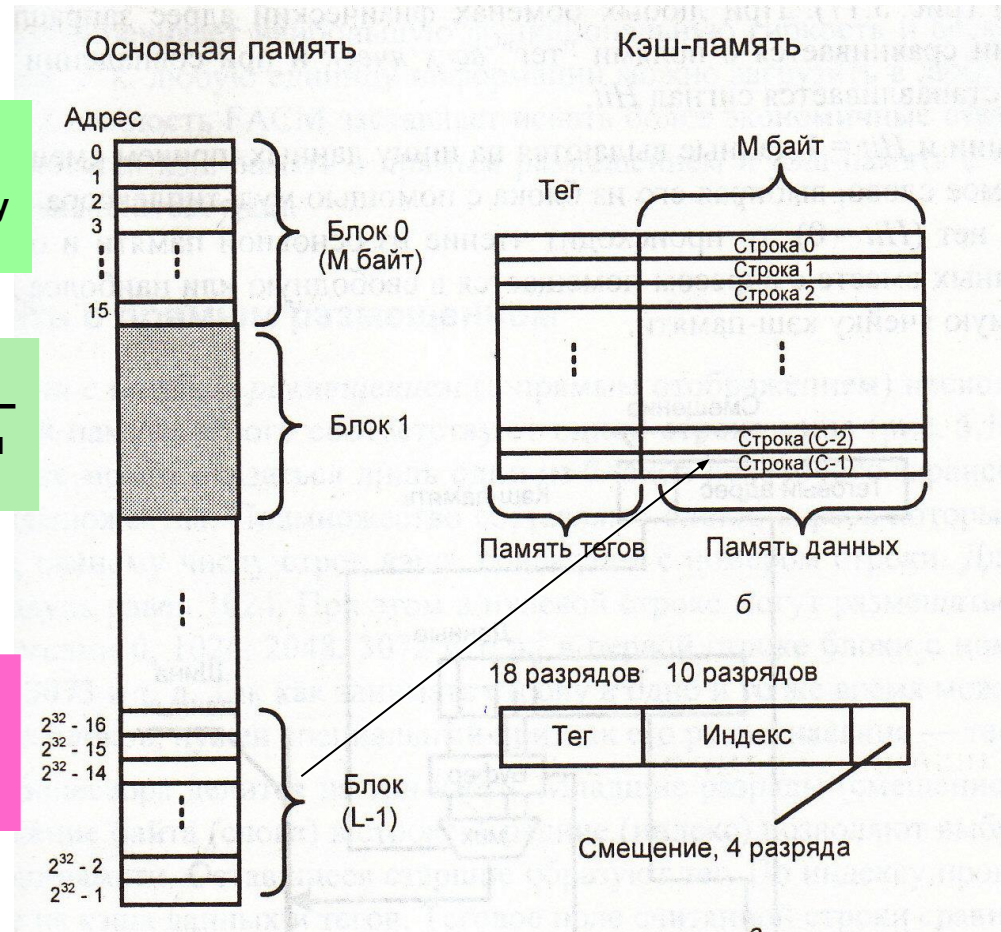
- Кэш память прямого отображения.
- Кэш память ассоциативная с множественным доступом или частично ассоциативная.
- Кэш память полностью ассоциативная.

Структура оперативной памяти и кэш-памяти

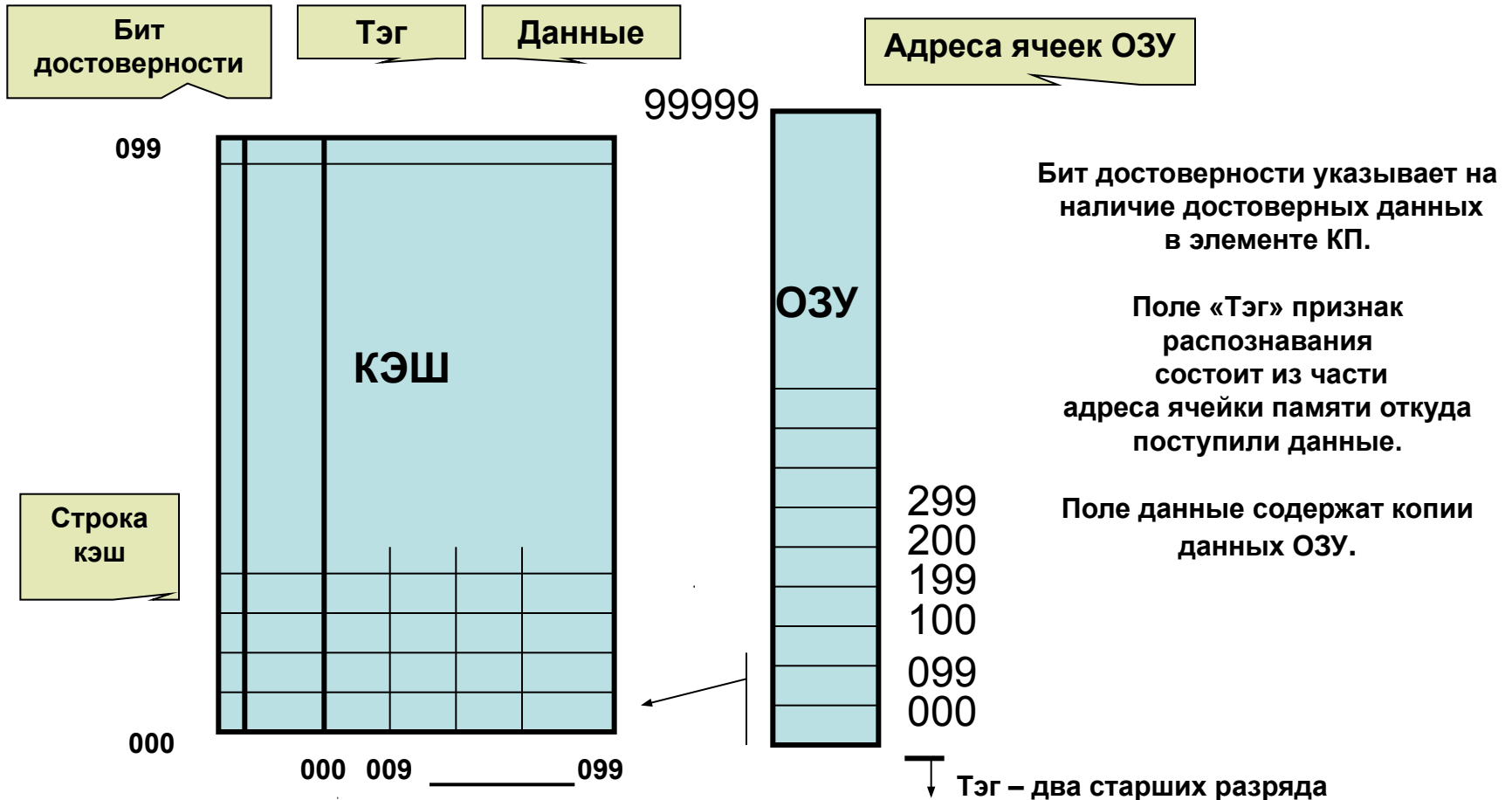
Основная память делится на блоки одинакового размера с кратными адресами ячеек. Обычно по размеру строки кэш.

Так как все блоки в кэш памяти не умещаются вводится понятие тэга – указателя, какой блок ОП находится в кэш.

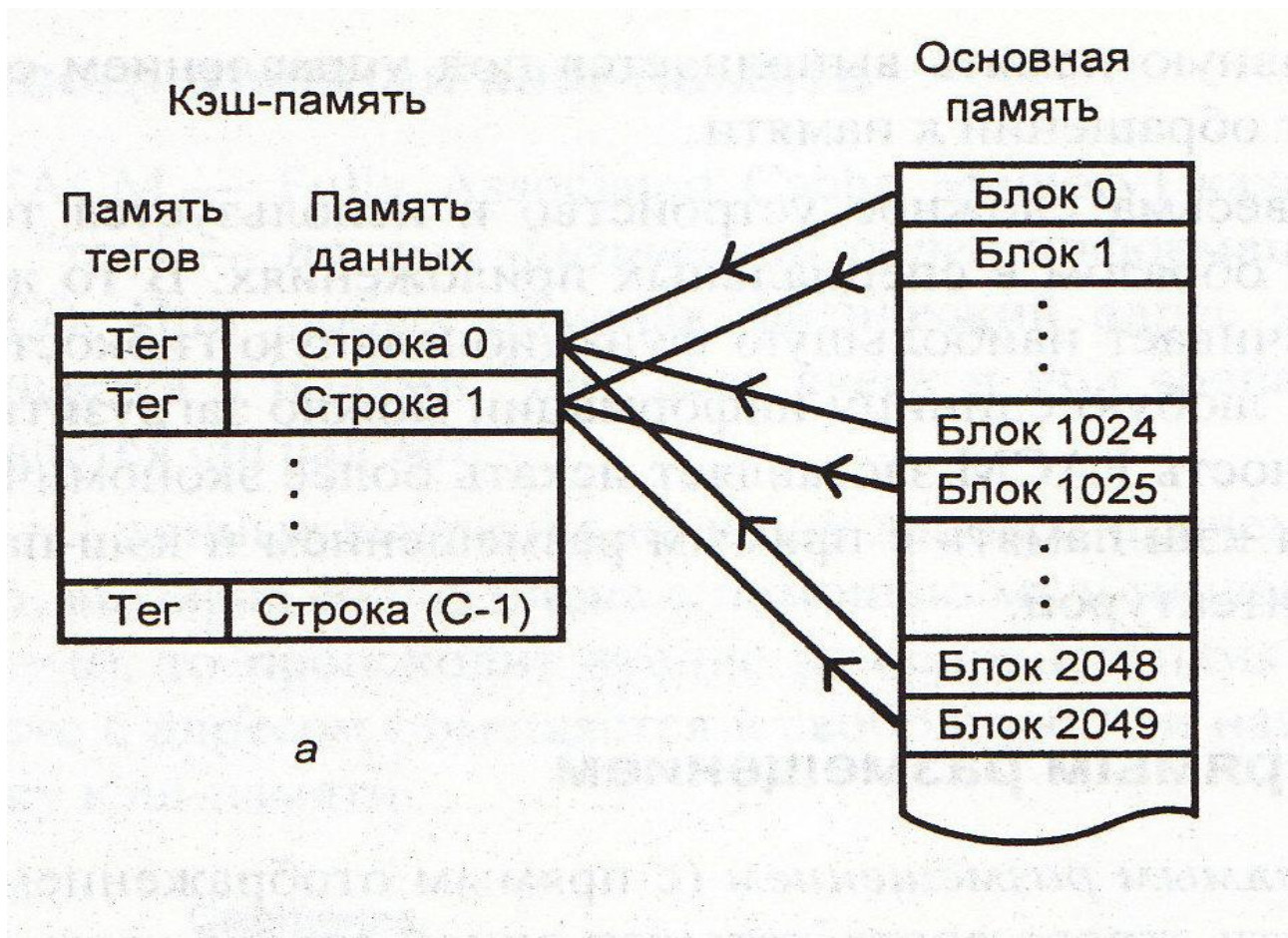
Пример: при ОП 4 ГГб и кэш 16Кбайт размер блока 16 байт. Обычно размер блока соответствует размеру строки.



Кэш память прямого отображения

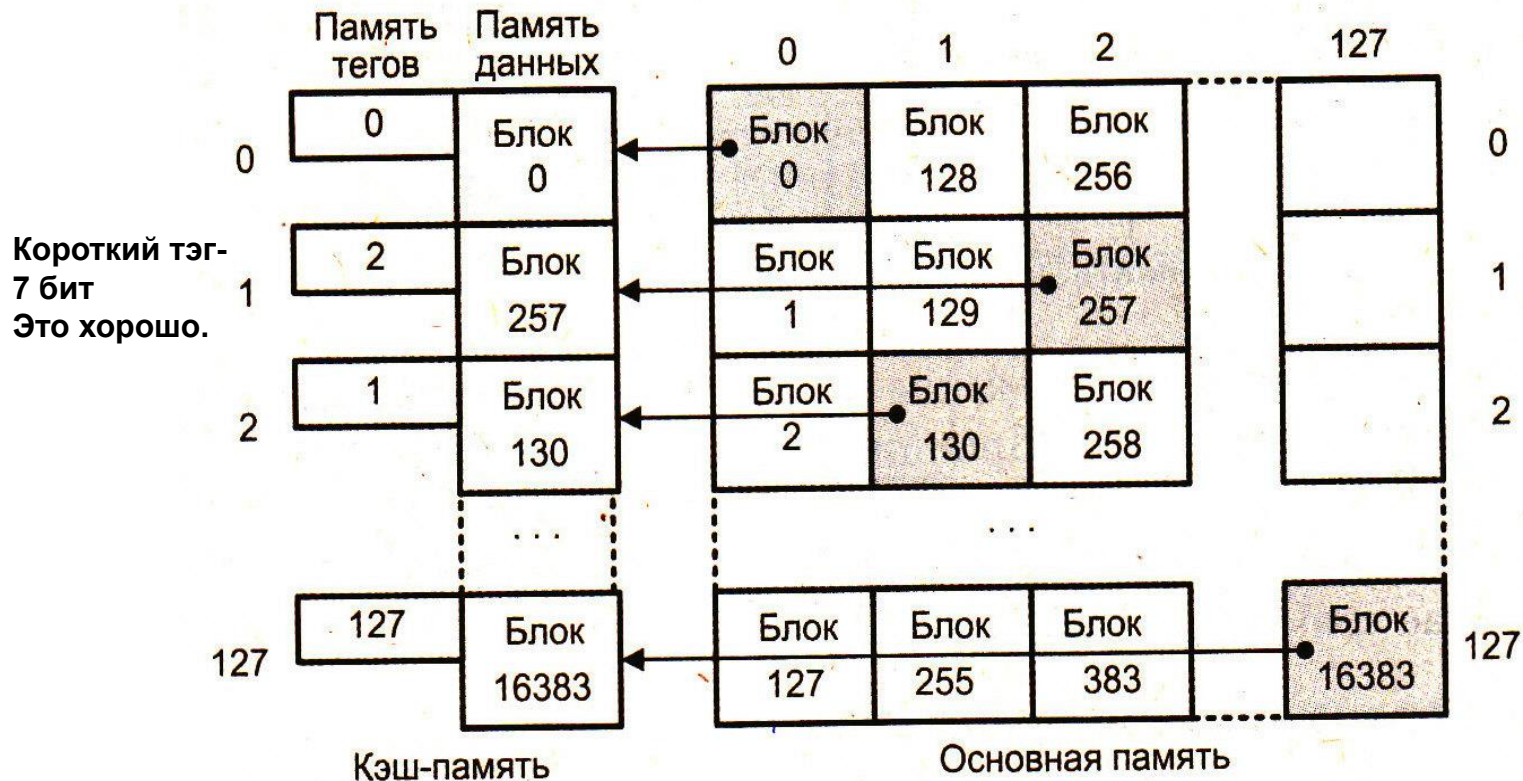


Соответствие адресов блоков ОП и строк кэш прямого отображения



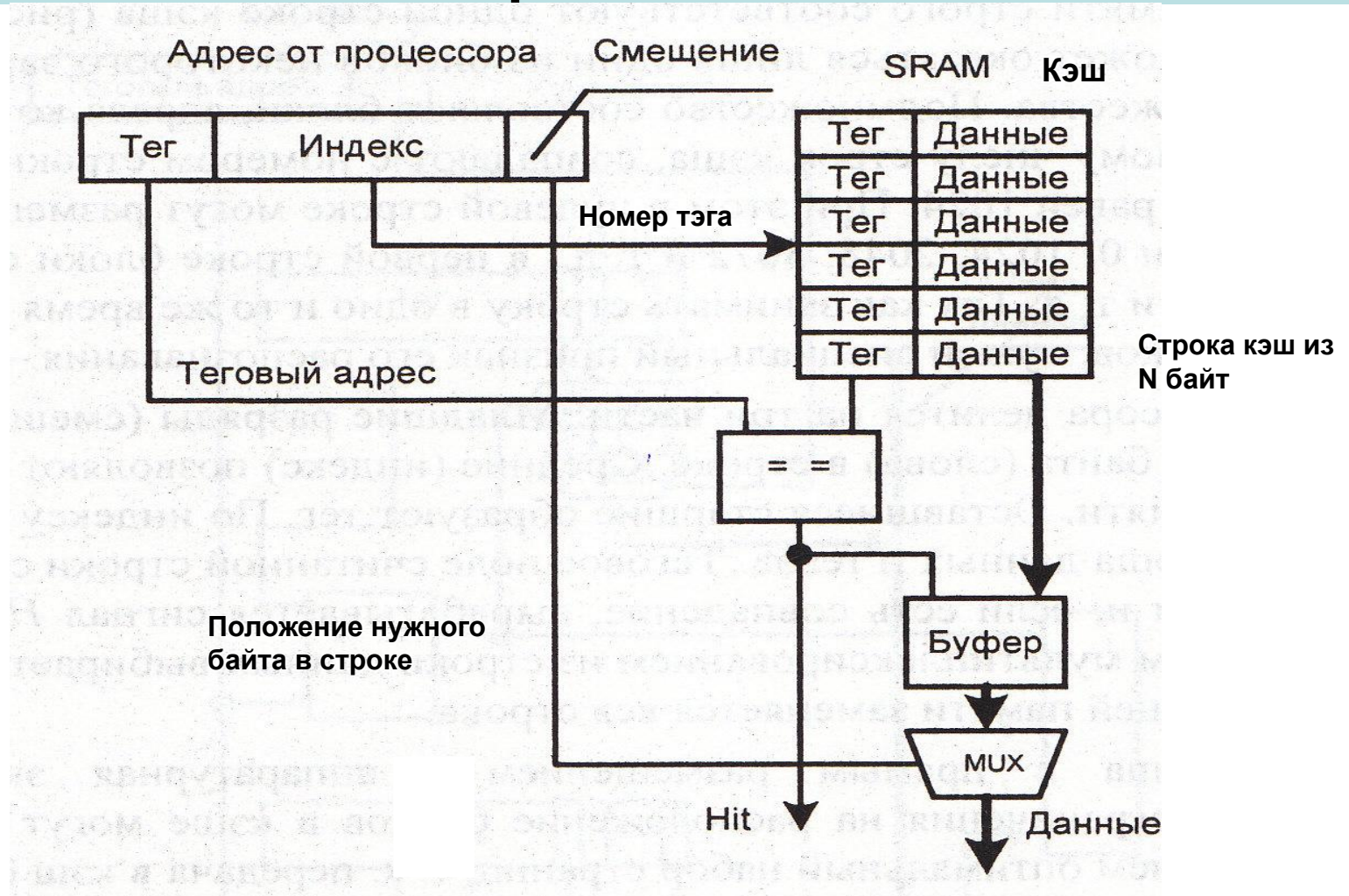
Кэш прямого отображения

Простой и недорогой способ реализации кэш. При проверке наличия в кэш нужной копии достаточно проверить память тегов.



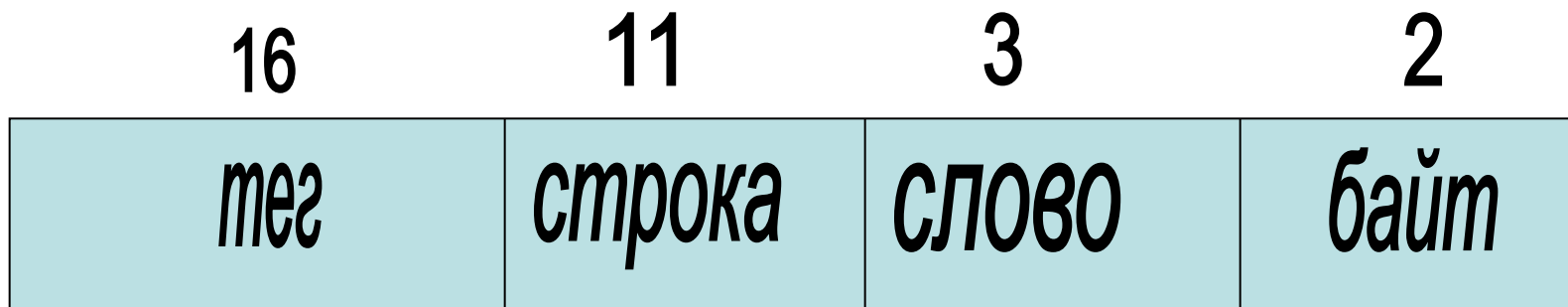
Недостаток – жесткое закрепление за определенными блоками ОП одного блока кэш. Если программа Поочередно обращается к двум разным блокам ОП, отображаемым на на одну строку кэш, то строка Кэш будет постоянно перезагружаться.

Организация кэш памяти прямого отображения



Кэш прямого отображения

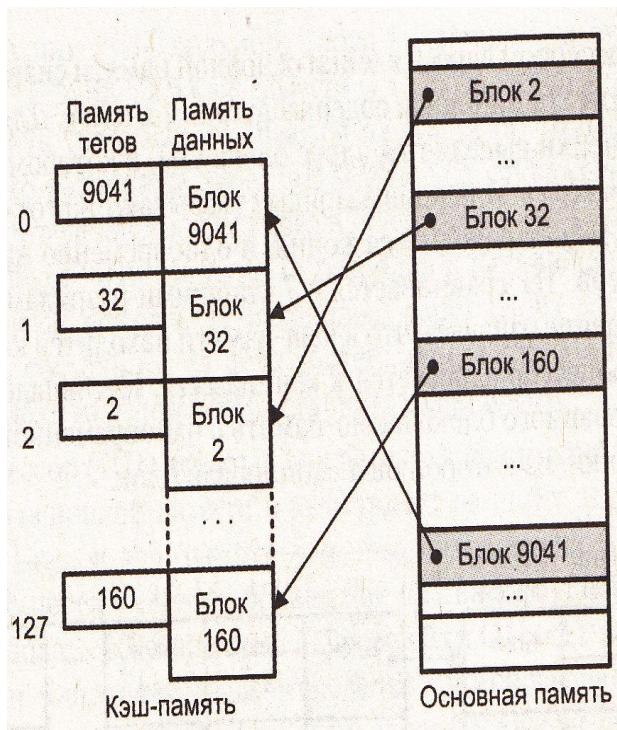
- **Пример:** 32 битный виртуальный адрес



- Поле "ТЭГ" соответствует полю кэш.
- Поле "строка" указывает какой элемент кэш памяти содержит соответствующие данные, если они есть в КП.
- Поле "Слово" указывает на какое слово в строке производится ссылка.
- Поле "Байт" определяет какой байт в слове требуется.

Полностью ассоциативная кэш память

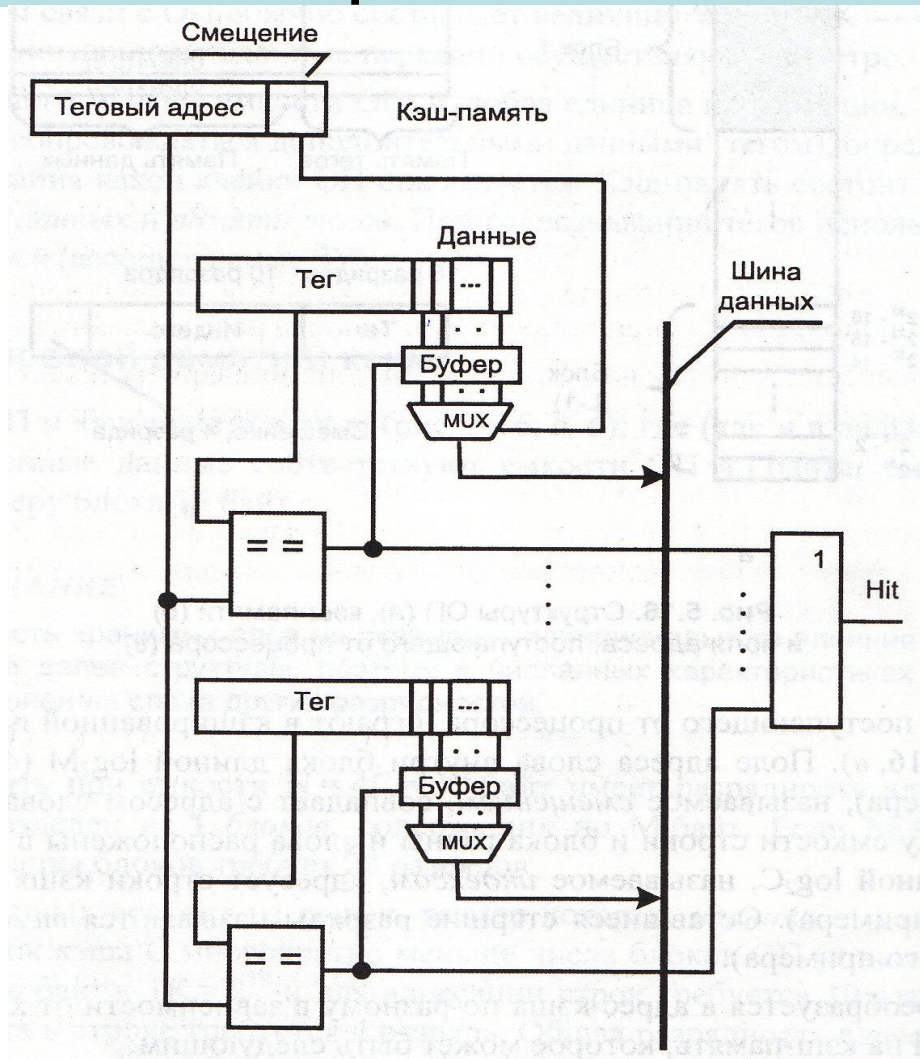
- Позволяет преодолеть недостаток прямого кэш, разрешая загрузку любого блока ОП в любой блок кэш.



Недостаток – длинный тэг, который надо сравнивать со старшими разрядами адреса, поступающего из процессора.

Такую память сложно организовать и дорого.

Организация полностью ассоциативной кэш памяти



Любая ячейка ОЗУ может находиться в любом месте КЭШа.

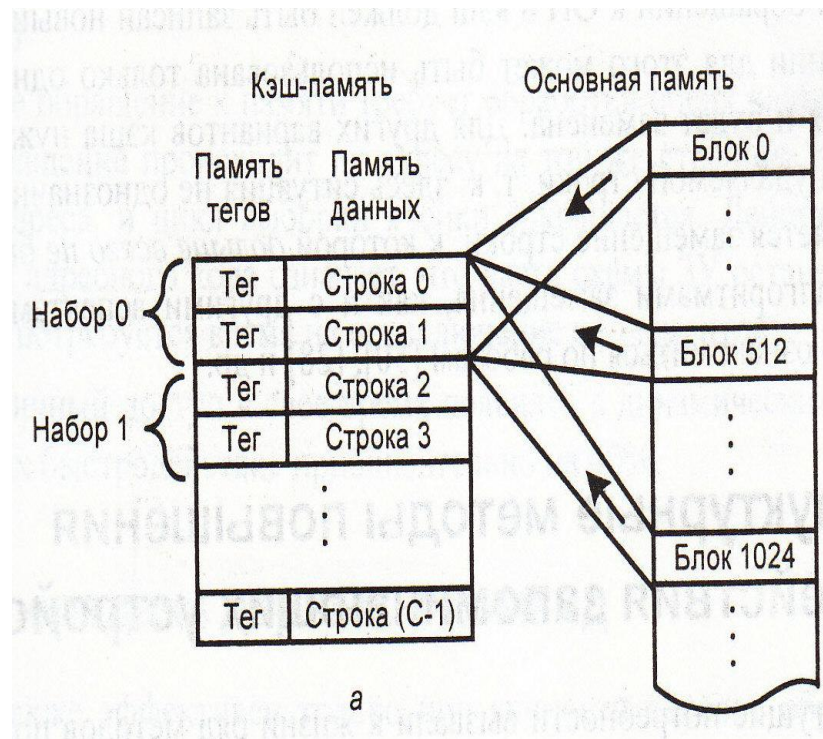
Сложная структура, дорогая. Используется в специальных устройствах с малым объемом памяти.

Кэш память с частично ассоциативным отображением

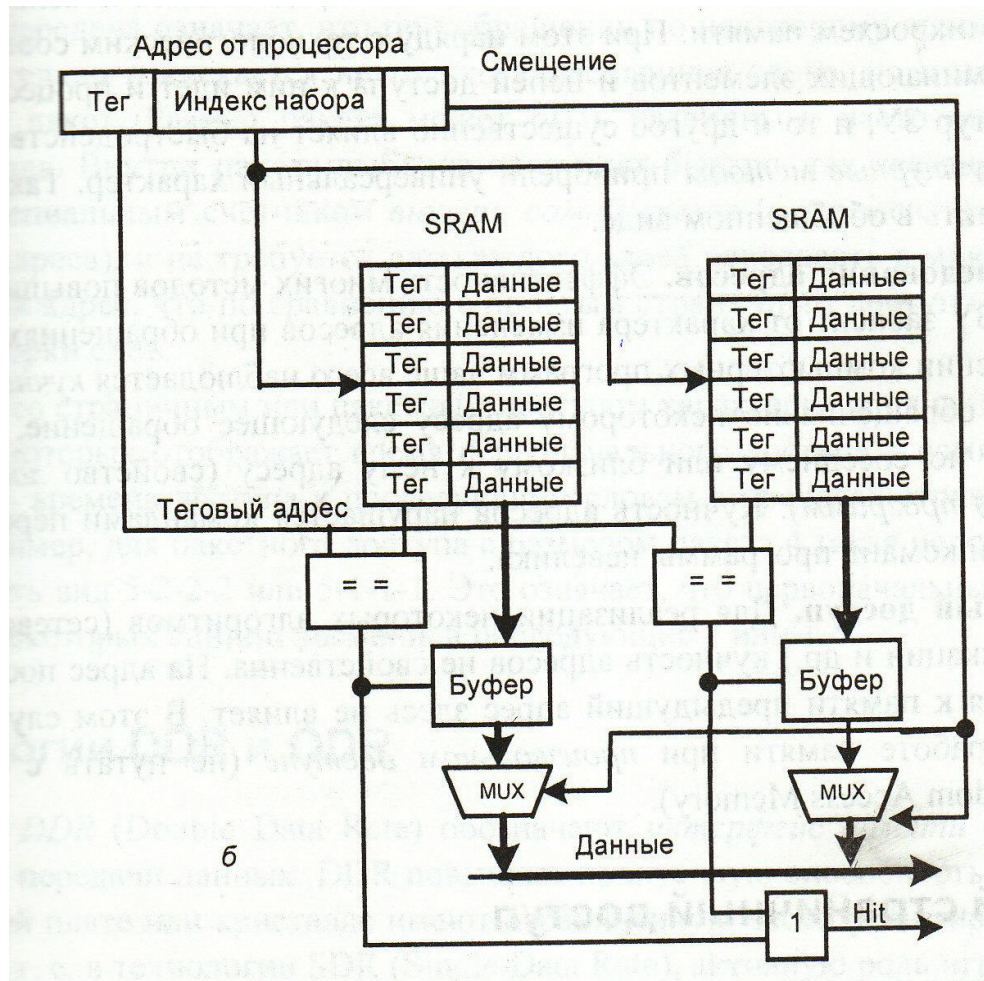
- Сочетает достоинства полностью ассоциативной кэш и кэш прямого отображения.



Соответствие блоков ОП и строк частично-ассоциативной кэш



Организация частично ассоциативной КЭШ – двухвходовой



Пример четырех входовой ассоциативной кэш-памяти



Каждая ячейка памяти может храниться в 4 строках кэш
Размеры строк 16,64,128,256 байт.

Возникает задача – замещения строк.

Стратегии замещения строк в кэш

- **LRU (Least Recently Used)** - замещается строка, к которой дольше всего не обращались. Не востребованная информация.
- **FIFO** – замещается строка самая давняя по пребыванию в кэш. Первый пришел – первый вышел.
- **Random** – замещение случайным образом.

Вероятность кэш промаха для различных стратегий замещения

Размер кэша, Кбайт	Организация кэш-памяти					
	2-канальная ассоциативная		4-канальная ассоциативная		8-канальная ассоциативная	
	LRU	Random	LRU	Random	LRU	Random
16	5,2	5,7	4,7	5,3	4,4	5,0
64	1,9	2,0	1,5	1,7	1,4	1,5
256	1,15	1,17	1,13	1,13	1,12	1,12

Согласование данных КЭШ и ОЗУ

- **Сквозная запись.**

Данные ОЗУ обновляются вместе с КЭШ.

Недостаток – снижение быстродействия.

Достоинство – простота реализации.

- **Обратная запись.**

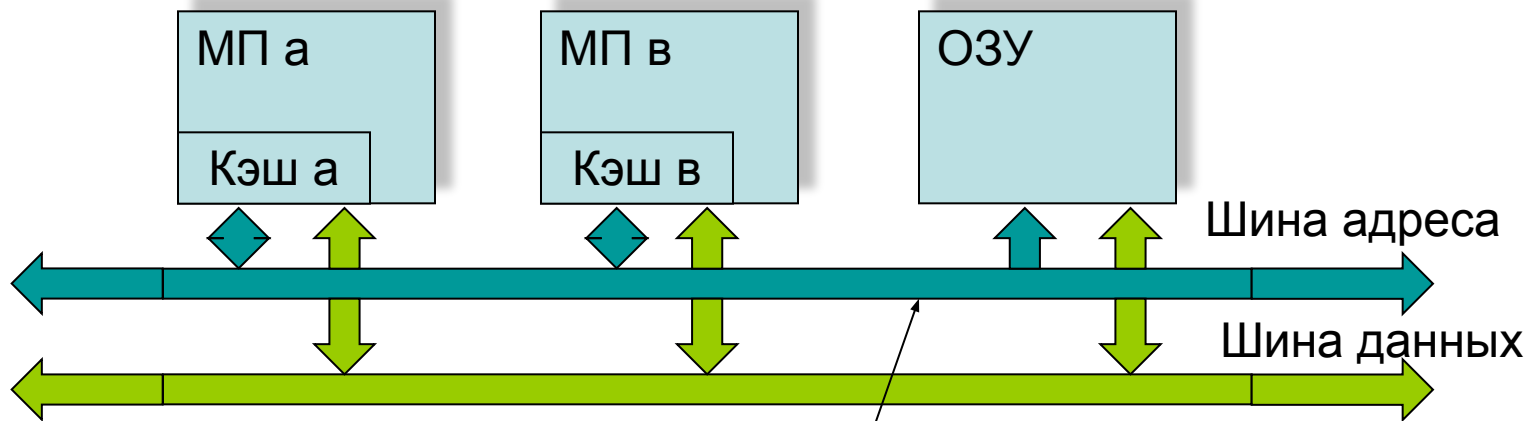
Данные ОЗУ обновляются при вытеснении строки из Кэш памяти.

Недостаток – временная неоднозначность данных в ОЗУ и Кэш.

Режим работы кэш

- Задается программной установкой разрядов CD (запрет кэширования) и NW (запрет сквозной записи) в управляющем регистре CR0.

Обеспечение согласованности кэш памяти МП в мультипроцессорных системах



Протоколы когерентности:

- на основе справочника;
- на основе протоколов наблюдения:
 - протокол записи с аннулированием ;
 - протокол записи с обновлением;

MESI – протокол согласования

Протокол наблюдения с аннулированием – при выполнении записи все копии стираются.

Состояние признака протокола		Состояние строки памяти		
		Кэш А	Кэш В	ОЗУ
Измененное	Modified	Д	НД	НД
Разделяемое	Shared	Д	Д	Д
исключительное	Exclusive	Д	НД	Д
Недостоверное	Invalid	НД	Х	Х

Д – информация достоверна;

НД – информация не достоверна;

Х – статус информации не имеет значения;

MESI диаграмма обеспечения когерентности

RH (Read Hit) – кэш попадание при чтении;

WH (Write Hit) – кэш попадание при записи;

RME (Read Miss Exclusive) – кэш промах при чтении;

RMS (Read Miss Shared) - кэш промах при чтении, но такой блок есть в памяти другого МП;

WM (Write Miss) - кэш промах при записи;

SHR (Snooper Hit Read) – обнаружение копии блока при прослушивании операции чтения другого кэш;

SHW (Snooper Hit Write) - обнаружение копии блока при прослушивании операции записи другого кэш;

