

Кеширане на паметта

Публикувано от **legolas** на **07.03.2010**

Кеширане на паметта

Това е една от най-популярните технологии за съвместяване по скорост с паметта.

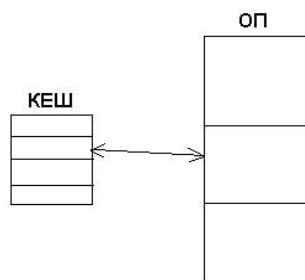
Кеша представлява блок бърза памет - типично високоскоростен статичен RAM , включен между микропроцесора и оперативната памет.

В него се съхранява съдържанието на онези данни и инструкции използвани най-скоро от микропроцесора или тези от които той най-вероятно ще се нуждае. "Кеш-контролер" се грижи за състоянието на кеша. Ако търсената от микропроцесора информация се намира в кеша, тя може да се използва без състояние на изчакване. Тази бърза операция се нарича **"кеш-успех"**. Когато тази информация не се намира в кеша, тя се извлича от обикновения RAM. Операцията се нарича **"кеш-пропуск"**. Кеш паметите се различават по размер, логическа организация, разположение и опериране. Що се отнася до големината на кеша тя може да бъде от 1В до няколко МВ или цялата оперативна памет. Очевидно е, че 1В е неефективен а цялата оперативна памет- много скъпо. Най-често използвания размер на кеша е 512 К. Логическата конфигурация на кеша се определя от това как е организирана паметта в него, как е адресирана и как микропроцесора определя дали нужната информация е на разположение в кеша.

Използват се **3 основни логически конфигурации - директно адресиран кеш, пълно асоциативен и наборно асоциативен.**

Класификация според логическата конфигурация на КЕШ.

1. **Кеш с директно адресиране** - бързата кеш памет се разделя на малки единици наречени линии (редове). Всяка от тези единици се идентифицира от индексен бит.



фиг.1

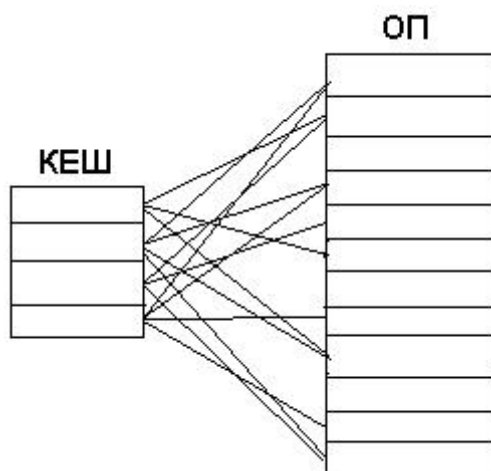
Главната памет е разделена на блокове с големината на кеша /фиг.1/. На линиите на кеша отговарят съответни редове от тези блокове. Всяка линия на кеша може да бъде заредена от различен блок памет но само от този ред на блока, който и съответства. Идентификатор определя от кой блок е извадена линията. Кеш-контролерът проверява дали идентификаторът съвпада с дадената индексна стойност.

Когато дадена програма работи често с адреси с едни и същи индекси в различни блокове от паметта се налага непрекъснато опресняване на кеша. Този проблем се появява при многозадачните системи и води до забавяне на директно адресирания кеш.

2. Пълно асоциативен кеш

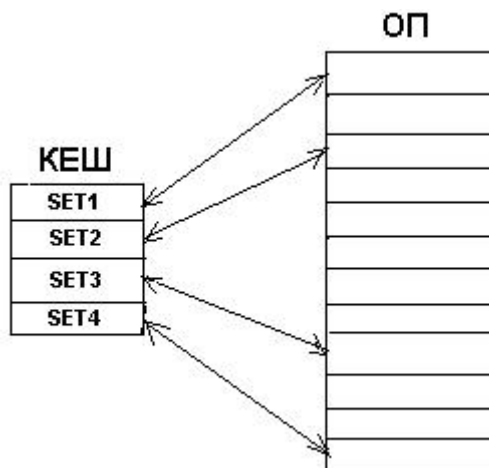
Напълно противоположен конструкторски подход. Тук всеки ред може да кореспондира с всяка част от главната памет. Редовете от различни места от главната памет могат да бъдат вмъкнати един до друг в кеша.

Главен недостатък на тази организация е, че кеш - контролерът трябва да проверява адресите на всеки ред в кеша за да провери дали заявката за памет от микропроцесора е "кеш- успех" или "кеш-пропуск". Това може да направи кеша дори по-бавен от главната памет.



фиг.2

3. Наборно асоциативен кеш Той е компромисен вариант между предишните две решения.



фиг.3

При него цялата кеш памет се разделя на няколко зони които се индексират директно. Кешът се описва с броя на зоните (каналите), на които е разделен. Тази конструкция преодолява проблема за движението м/у блокове с еднакъв индекс. Има по - голям потенциал за производителност, но е по-сложна и скъпа. Освен това ако броя на каналите е по-голям това ще доведе до забавяне на кеша.

Класификация според разположението на КЕШ.

Според разположението си кешовете могат да бъдат вътрешни и външни за микропроцесора.

4. **Вътрешния кеш** се нарича първичен и е вграден в чипа на микропроцесора. Ще го срещнете под името **кеш от първо ниво (Level1, L1 cache)**. Този вид памет е по-бърз от цялата останала, тъй като е свързан директно с вътрешните шини на микропроцесора, но има проблем с капацитета. Тя съхранява съвсем наскоро използваните инструкции, които се намират в малки и къси програмни цикли. Разширяването на вътрешния кеш води до увеличаване размерите на чипа на микропроцесора.
5. **Външния кеш** (вторичен) (**Level2, L2 cache**) използва външен кеш контролер и чипове с памет по-бавна от L1 и по-бърза от оперативната памет.

Понякога вътрешните кешове са разделени функционално в отделни “кешове за инструкции” “кешове за данни”.

Кешът за инструкциите съдържа само микропроцесорни инструкции, кешът за данни съдържа само данни. Това разделяне се използва от Pentium на Intel, някои процесори на Motorola и повечето RISC чипове и е вид подобрене на кеша.

Класификация според начина на извършване на записа в Кеш.

Кешовете се различават и по начина по който се осъществява записа в паметта.

-

6. КЕШ с директен запис - При повечето кешове се извършва **директен запис**- не се прави опит за ускоряване на записа. Вместо това командата за запис се изпълнява веднага, като едновременно се записва и в кеш паметта и в оперативната памет. Директният запис през кеша гарантира постоянна еквивалентност между централната памет и кеш паметта. Повечето процесори на Intel до последния модел на Pentium използват тази технология.

7. КЕШ с отложен запис - По бърза алтернатива на директния запис са кешовете с отложен запис. Този метод позволява на микропроцесора да прави промени в кешовата памет, които контролерът на кеша записва веднага или след време в оперативната памет в зависимост от времето което му се отпуска. Тук може да възникне проблем защото е възможно да се получи кеша и оперативната памет да имат различно съдържание на едни и същи адреси. Контролерът на кеша трябва постоянно да проверява промените направени в оперативната памет и да гарантира, че съдържанието на кеша следва тези промени. Това го прави по -сложен и по-скъп.

Местоположение на паметта

Освен от изброените до тук фактори, бързодействието на паметта зависи от системната архитектура. От местоположението на паметта и начина на свързване към системата зависи драстично бързодействието на микропроцесора и неговата производителност. Обикновено паметта е свързана с микропроцесорната шина, което означава че работи със скорост определена от микропроцесора и е свързана чрез шина с големината на микропроцесорната шина за данни. Повечето дънни платки нямат достатъчно пространство за увеличаване обема на паметта. Използват се разширителни модули за допълнителна памет. Такива модули могат да се свържат към дънната платка или чрез специални връзки, които работят с пълната скорост на микропроцесора и широчината на шината, или чрез стандартните външни шини като ISA, EISA или Micro Channel. Първите разширителни модули наречени пряко свързани не налагат ограничение върху бързодействието на паметта. Другият вид модули памет добавяни чрез разширителен слот довеждат до намаляване на бързодействието на всички системи по мощни от 8 MHz AT съвместимите. Слотовете обикновено работят на 8 MHz, което не съответства на голямата скорост на микропроцесора. Достъпът до добавената чрез разширителни слотове памет изисква намаляване честотите на системата до 8MHz, което намалява ефективността ѝ. За постигане на добра производителност не може паметта да се поставя на обикновените разширителни слотове, на какъвтои да е по-бърз компютър.

Ограничения за размера на паметта

Най-важните фактори за ограничаване обема на паметта са:

- архитектурни ограничения;
- адресен лимит на микропроцесора;
- изискване за непрекъсваемост на паметта.

8. Адресни възможности на микропроцесора -

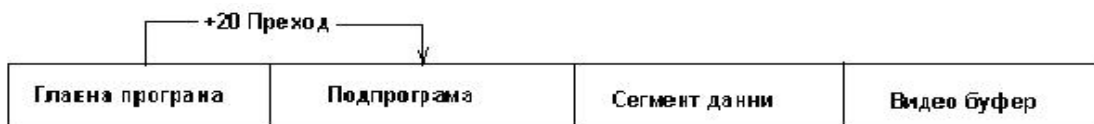
Размерът на паметта която микропроцесорът може да адресира се ограничава от броя на адресните линии и неговата вътрешна архитектура. Той не може да адресира повече от 2^n , където n е броят на адресните линии. При 8088/8086 $n=20$ – 1 MB При 80286 $n=26$ – 2^{24} B =16MB. С въвеждането на 32 разрядни адресни линии, пряко адресируемата оперативна памет става 2^{32} B, което означава практически неограничена. Осигурен е достъп до 4GB памет, но не всички компютри поемат такава памет, колкото техния микропроцесор адресира.

- 9. Капацитет на системата** – ограничение на шината. PC на базата на 386 използват 24 адресни линии (AT шина) и по този начин налагат ограничение на пряко адресираната оперативна памет до 16 MB. След като се премахне ограничението на шината ($n=32$) капацитетът на оперативната памет може да нарасне до тази на микропроцесора. Не всяка ISA шина може да осигури тази опция. Някои от тях налагат ограничението до 16 защото проектирането и производството на тези PC е по евтино. Второто поколение разширителни шини EISA и Micro Channel разширяват адресната шина до пълната 32 битова, но не е препоръчително да се използват за разширяване на паметта. Днешните VL bus и PCI имат същия размер от 32 битово ограничение. Адресирането при тях не е проблем. Най-сигурният начин да се идентифицира адресния лимит на една машина е да се провери листа ѝ със спецификация.

Проблем за непрекъснатост на паметта

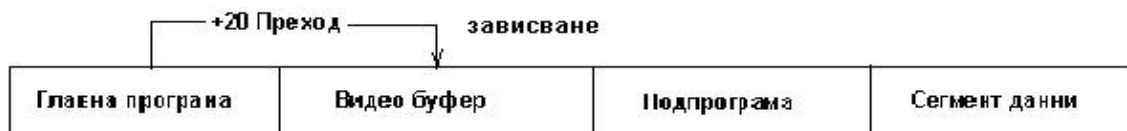
Почти всички приложения и операционни системи предполагат, че системната памет е непрекъсната (няма дупки в пространството от началото до края). Това е една от причините за относителните преходи в програмите. Това са инструкции които казват на микропроцесора да скочи от една на друга точка

на паметта. Вместо точен адрес в паметта, размерът на този скок се определя от брой байтове между старата и новата точка в инструкцията. Относителният преход казва на програмата да търси следващата инструкция на определен брой байтове от последната инструкция. Програмата може да се зареди на произволно място в RAM без никакъв проблем. Тази гъвкавост позволява едни и същи програми да се изпълняват в системи с различен капацитет на паметта и с различно зареден софтуер. Ако адресите в RAM не са последователни - има дупка в паметта - е възможно един относителен преход да пренасочи управлението на програмата някъде в средата, където няма инструкции и обикновено няма памет изобщо. Резултатът е блокиране на системата. Проблеми с непрекъснатостта на паметта може да възникнат, когато се използва сенчеста ROM и запазени адреси.. Когато програмата извиква подпрограма, се прави относителен преход на определен брой байтове, като се очаква управлението да се поеме от подпрограмата.



Фиг.1

Ако интервал от специализирана памет разкъса последователността на програмния код, може да се окаже че преходът е извършен не където трябва. Това ще доведе до зависване на системата.



Фиг. 2

Дублирането на паметта ускорява PC, но може да има неприятен ефект върху паметта. Допълнителната памет, която дублира пренасочения ROM се взема обикновено от най-горните адреси на допълнителната памет (около 256 K). Някои стари PC фиксират адресите за сенчестата памет точно до границата на 16MB . Ако в такъв компютър се инсталира повече от 16 MB памет, за засенчване се използват необходимия брой адреси от същата тази област, при което в адресното пространство ще се появи дупка. Операционната система и програмите ще се

сблъскват с нея и няма да могат да продължат нататък независимо от това колко памет сте инсталирали. Отстраняването на този проблем става чрез изключване на сенчестата памет посредством BIOS процедурите за настройка. Запазените адреси са области които периферните устройства използват за вход/изход и за управление. Устройствата вземат от там информация и я обработват. Видеоконтролерът с директно адресируема памет работи със запазени адреси.

Кеширане на паметта

Powered by

Bukvar.bg

Image not found

© 2010-2024