**Стукало Виктор Александрович, к.т.н. 905-854-1294**

**stukalo.v@mail.ru**

**Аппаратные средства вычислительной техники**

Эндрю Таненбаум – «Архитектура компьютера».

**Процессоры**

**Архитектура системы команд. Классификация процессоров CISC и RISC.**

Термин архитектура системы употребляется в двух смыслах – узком и широком.

В узком смысле под архитектурой понимается совмещение с определенным набором команд и это наиболее частое употребление термина. Архитектура набора команд является границей между аппаратурой и программным обеспечением.

В широком смысле архитектура означает организацию вычислительной системы, включающую процессор, систему памяти, структуру системной шины, организацию ввода-вывода и т.п.

В мире существуют 2 основных архитектуры системы команд:

1. CISC (Complete Instruction Set Computer)
2. RISC (Reduced Instruction Set Computer)

Основоположником CISC архитектуры считается IBM с базовой архитектурой system 360. Ядро этой системы используется с 1960. Используется в современных операционных системах. Лидером в разработке микропроцессоров CISC считается компания Intel. Их линейка x86(x64) как раз относится к системам с полным набором команд (CISC). Для CISC процессоров характерно небольшое число регистров общего назначения (не больше 16).

**ДЗ: Повторить назначения регистров процессора фирмы Intel. Повторить систему команд.**

CISC процессоры характеризуются:

1. Большим количеством машинных команд (некоторые из них семантически нагружены аналогично операторам высокоуровневых языков программирования).
2. Большое количество методов адресации

**ДЗ: Повторить способы адресации микропроцессора Intel.**

1. Большое количество форматов команд

**ДЗ: Повторить форматы команд микропроцессора Intel.**

1. Преобладание двухадресного формата команд (поле команды и 2 поля адреса, один источника, другой приемника операции, результат заменяет первый адрес)
2. Наличие команд обработки типа регистров памяти

Альтернативой CISC является RISC архитектура. До недавнего времени считалось, что серверные процессоры должны строиться по RISC архитектуре. Современное понятие RISC сформировалось на базе трех исследовательских проектов. На интервале 2 года IBM выпустила процессор 801, университет Беркли выпустил процессор RISC, а Стэндфордский университет выпустил свой процессор. Машины имели много общего. В этой архитектуре команды обработки отделены от команд работы с памятью. Упор делался на конвейерную обработку команд. Система команд разрабатывалась таким образом, чтобы исполнение команды занимало один такт машинного времени. В CISC архитектуре огромное количество команд выполняется за несколько сотен тактов. С целью повышения производительности логика исполнения команд была аппаратной, а не микропрограммной. Для упрощения логики декодирования команд использовались команды фиксированных длинны и формата. Процессоры RISC архитектуры обладают большим регистровым файлов (набором регистров). От 32-х регистров. Это позволяет большему объему данных храниться в процессоре. Меньше обращаться к памяти. Упрощает работу компилятора по распределению регистров под переменные. Для обработки используется трехадресные команды. Что позволяет хранить большее число данных без перезагрузки.

Пример **A + B = C**

CISC

LD R1, A: [A] -> R1

ADD R1, B: [R1] + [B] -> R1

ST C, R1: [R1] -> C

RISC

LD R1, A: [A] -> R1

LD R2, B: [B] -> R2

ADD R3, R2, R1: [R1] + [R2] -> R3

ST C, R3: [R3] -> C

**Организация конвейера**

Ступени конвейера

1. IF – instruction fetch (выборка команды)
2. ID – instruction decode (декодирование команды)
3. Exe – executive (исполнение команды)
4. Mem – memory (обращение к памяти)
5. WB – write back (обратная запись)

Диаграмма конвейера (#1).

Отчет

1. Титульный лист
2. Постановка задачи
3. Описание алгоритма решения задачи (исходный текст с комментариями)
4. Таблица с результатами экспериментов
5. Заключение
   1. Констатация фактов (что наблюдал, какие данные получил)
   2. Почему получил именно такие данные, чем это вызвано, какими эффектами, работает или нет теория, изложенная в описании лабораторной работы

**Организация 8-bit процессора**

В команде нужно указать:

1. Код операции
2. Адрес операнда 1
3. Адрес операнда 2
4. Адрес результата
5. Адрес следующей команды

Сделан вывод, что адрес следующей команды можно убрать, потому как выполняется обычно следующая команда. Вместо него вводится (Program Counter) программный счетчик, который инкрементируется при выполнении каждой команды и указывает на следующую команду.

Add A, B, C, D – адрес следующей команды не нужен (используется PC)

Add A, B, C – адрес результата не нужен (используется первый операнд для вывода результата)

Add A, B – адрес первого операнда не нужен (используется аккумулятор)

Add B – получилась команда с одним операндом

Add Rn, B – получаем команду с операндом, указывающим на регистр общего назначения (с коротким адресом) и операндом со вторым значением

Вычисление по элементарной формуле:

B + C = D

Исходное положение процессора: PC {1000}

1 команда 1 фаза:

PC{1000/1001}

AR{1000}

MD{LD}

IR{LD}

1 команда 2 фаза:

PC{1001/1002}

AR{1001}

MD{LoB}

IR{LD||LoB}

1 команда 3 фаза:

PC{1002/1003}

AR{1002}

MD{HiB}

IR{LD|B}

1 команда 4 фаза:

PC{1003}

AR{B}

MD{100}

IR{LD|B}

1 команда 5 фаза:

MD{100}

A{100}

2 команда 1 фаза:

PC{1003/1004}

AR{1003}

MD{ADD}

LR{ADD}

2 команда 2 фаза:

PC{1004/1005}

AR{1004}

MD{LoC}

LR{ADD|LoC}

2 команда 3 фаза:

PC{1005/1006}

AR{1005}

MD{HiC}

LR{ADD|C}

2 команда 4 фаза:

PC{1006}

AR{B}

MD{200}

IR{ADD}

2 команда 5 фаза:

A{100/200}

3 команда 1 фаза:

PC{1006/1007}

AR{1006}

MD{ST}

IR{ST}

3 команда 2 фаза:

PC{1007/1008}

AR{1007}

MD{LoD}

IR{ST|LoD}

3 команда 3 фаза:

PC{1008/1009}

AR{1008}

MD{HiD}

IR{ST|D}

3 команда 4 фаза:

PC{1008}

AR{D}

MD{255}

{B+C}

255

**Системный таймер**

Раньше это была схема intel-8083 (PIT – programmable interval timer).

В составе таймера имеется 4 программно-доступных регистра, каждый имеет свой адрес:

1. 40h – регистр делитель
2. 41h – регистр делитель
3. 42h – регистр делитель звукового канала
4. 43h – регистр управляющего слова

61h – 8bit порт, многофункциональный. Младшие 2 бита этого порта используются, один для подачи сигнала запуска таймера, другой открывает электронный ключ. Остальные биты трогать не рекомендуется.

43h – описание битов:

1. binary coded decimal / binary.
2. Режим работы канала
3. Режим работы канала
4. Режим работы канала (011)
5. 1
6. 1
7. Канал
8. Канал (00, 01, 10)

B6 (управляющее слово для лабораторной)

Делитель = Входная величина/выходную величину(задается мелодией)

Разработчики архитектуры компьютеров издавна прибегали к методам проектирования известным под общим названием совмещение операций, при котором аппаратура компьютера в любой момент времени выполняет одновременно более одной операции. Этот общий метод включает 2 понятия:

1. Параллелизм
2. Конвейеризация

У них много общего, зачастую их трудно различить на практике, но эти термины отражают два совершенно различных подхода.

При параллелизме совмещение операций достигается путем воспроизведения в нескольких копиях аппаратной структуры. Высокая производительность достигается за счет одновременной работы всех элементов структур осуществляющих решение различных частей задачи.

Конвейеризация – в общем случае основана на разделении подлежащей исполнению функции на более мелкие части, ступени и выделение для каждой из них отдельного блока аппаратуры. Обработку любой машинной команды можно разделить на несколько этапов, организовав передачу данных от одного этапа данных к следующему. При этом конвейеризацию можно использовать для совмещения этапов выполнения разных команд. Производительность при этом возрастает благодаря тому, что одновременно на различных ступенях конвейера выполняется несколько команд. Конвейерная обработка такого рода широко применяется во всех современных процессорах.

Для иллюстрации основных принципов построения процессоров будем использовать простейшую пяти ступенчатую архитектуру, содержащую 32 регистра общего назначения от до , 32 регистра плавающей точки от до и программный счетчик. Набор команд процессора включает типичные арифметические и логические операции, операции плавающей точки, пересылки данных, управления и разного рода системные команды. В арифметических командах используется трех-адресный формат. Для обращения к памяти используются операции загрузки и записи содержимого регистров в память. Общее время выполнения команды на таком конвейере будет составлять 5 тактов. В каждом такте аппаратура выполняет в совмещенном режиме 5 различных команд.

При реализации конвейерной обработки возникает ситуация, препятствующая выполнению очередной команды из потока команд в предназначенном для неё такте. Такие ситуации называются конфликтами. Конфликты снижают реальную производительность конвейера, которая могла бы быть достигнута в идеальном случае. Существуют 3 класса конфликтов:

1. Структурные конфликты (которые возникают из-за конфликтов по ресурсов, когда аппаратные средства не могут поддерживать все возможные комбинации команд в режиме одновременного выполнения с совмещением)
2. Конфликты по данным (возникающие в случае, когда выполнение одной команды зависит от результата выполнения предыдущей команды)
3. Конфликты по управлению (которые возникают при конвейеризации команд переходов и других команд, изменяющих значение программного счетчика)

Конфликты приводят к необходимости приостановки выполнения команд. В простейших конвейерах, если останавливается какая-либо команда, то все следующие за ней также приостанавливаются. Предшествующие команды продолжают выполнение, но во время приостановки не выбирается ни одна новая команда.

**Способы минимизации конфликтов.**

Структурный конфликт.

Совмещенный режим выполнения команд требует конвейеризации функциональных устройств и дублирования ресурсов, для разрешения всех возможных комбинаций команд в конвейере. Если какая-либо комбинация команд не может быть принята из-за конфликта по ресурсам, то имеется структурный конфликт. Время работы такого устройства может составлять несколько тактов синхронизации конвейера. В этом случае команды, использующие данное функциональное устройство, не могут поступать на него в каждом такте. Допускается наличие структурных конфликтов по двум причинам:

1. Снижение стоимости
2. Уменьшение задержки устройства

Конвейеризация всех функциональных устройств может оказаться слишком дорогой. Машины допускающие 2 обращения к памяти в одном такте должны иметь удвоенную пропускную способность памяти. Однако, если структурные конфликты не будут возникать слишком часто, то может быть и не стоит платить за то, чтобы их обойти.

Конвейер может иметь один порт для записи в память, но если необходимо 2 записи – возникает структурный конфликт. При этом, например, исполнение уже выбранной команды продолжается, операнд записывается в память.

Для разрешения конфликтов по данным используются так называемые пути обхода.

Конфликты по управлению.

Смысл перехода с задержкой перехода длиной n:

Команда условного перехода

Команда 1

Команда 2

…

команда n

точка перехода

Задача перехода заключается в том, чтобы сделать команду следующей за командой перехода действительным и полезными.

Конфликты по данным не устраняемые с помощью аппаратуры обхода.

Современные компиляторы используют технику планирования команд для улучшения производительности конвейера. В простейшем алгоритме компилятор просто планирует распределение команд в одном и том же базовом блоке. Базовый блок представляет собой линейный участок последовательности программного кода, в котором отсутствуют команды перехода. За исключением начала и конца участка. Переходы внутрь участка, тоже должны отсутствовать. Планирование такой последовательности команд осуществляется достаточно просто. Компилятор строит граф зависимостей этих команд и упорядочивает их так, чтобы минимизировать остановки конвейера. Для простых конвейеров стратегия планирования на основе базовых блоков удовлетворительна. Однако при интенсивной конвейеризации и росте задержек конвейера требуются более сложные алгоритмы планирования. Эти методы получили общее название методов динамической оптимизации. Часто используется термин out of order execution, т.е. неупорядоченное выполнение. А также метод переименования регистров.

**Параллелизм уровня команд**

Зависимости и конфликты по данным

Потенциальный параллелизм заложенный в последовательности команд называется параллелизмом уровня команд или ILP (Instruction Level Parallelism). Степень параллелизма доступная внутри одного базового блока очень мала. Например частота перехода в целых программах составляет около 16%. Это означает, что между двумя переходами выполняется 5 команд, а каждая 6 – переход. Поскольку эти 5 команд возможно взаимозависимые, то степень перекрытия, которую можно использовать внутри базового блока, возможно, будет меньше чем 5. Чтобы получить существенное улучшение производительности необходимо использовать параллелизм уровня команд одновременно для нескольких базовых блоков. Самый простой способ увеличения степени параллелизма, доступного на уровне команд – использование параллелизма между итерациями цикла. Этот тип параллелизма часто называется параллелизмом уровня итеративного цикла. Например, цикл выполняющие сложение двух векторов является полностью параллельным. Каждая итерация цикла может перекрываться любой другой итерацией. Хотя внутри каждый итерации практическая возможность перекрытия небольшая. Имеется несколько методов для превращения такого параллелизма уровня цикла, в параллелизм уровня команд. Методы основаны на разворачивании цикла, статистически используя компилятор, либо динамически, с помощью аппаратуры.

Основы планирования загрузки конвейеров и разворачивания циклов

Для поддержания максимальной загрузки конвейера зависимые команды должны отделяться от исходных на расстояние в тактах равное задержкам конвейера для этих исходных команд. Способность компилятора выполнять подобное планирование определяется способностями функциональных устройств. Во всех дальнейших примерах будем предполагать задержки конвейера указанные в таблице.

|  |  |  |
| --- | --- | --- |
| Команда вырабатывающая результат | Команда использующая результат | Задержка в тактах |
| Операция АЛУ с плавающей точкой | Др. оп. АЛУ с ПТ | 3 |
| Оп. АЛУ с ПТ | Запись двойного слова | 2 |
| Загрузка двойного слова | Др. оп. АЛУ с ПТ | 1 |
| Загрузка двойного слова | Запись двойного слова | 0 |
| Оп. условного перехода | Др. оп. УП | 1 |

Предположим, что в регистре R1 вектор, А в регистре f2 скалярная величина, которая добавляется к каждому элементу вектора.

Loop:

LD F0, 0(R1) ; загрузка в F0 элемента вектора

1

ADDD F4, F0, F2 ; добавление скалярной величины к элементу вектора

2

SD 0(R1), F4 ; запись в R1 нового значения

0

SUBI R1, R1, #8 ; пересчет указателя элемента вектора

0

BNEZ R1, Loop ; переходим на метку loop, если R1 = 0

1

9 тактов на результат итерации

Loop:

LD F0, 0(R1)

1

ADDD F4, F0, F2

0

SUBI R1, R1, #8

0

BNEZ R1, Loop ; задержанный переход

0

SD 8(R1), F4 ; размещается в слоте задержанного перехода и выполняется перехода

0

6 тактов на результат итерации

В примере за одну итерацию выполняется запись одного элемента вектора на каждые 6 тактов (после оптимизации цикла). Действительная работа по обработке элемента вектора занимает только 3 такта (загрузка, сложение, запись), остальные 3 такта составляют накладные расходы на выполнение цикла (пересчет указателя и условный переход, приостановка). Для устранения этих трех тактов нужно иметь больше операций в цикле относительно команд создающих накладные расходы. Один из основных методов является метод разворачивания циклов. Разворачивание циклов выполняется путем многократной репликации тела цикла и коррекции соответствующего кода конца цикла. Разворачивание циклов может также использоваться для улучшения планирования. В этом случае можно устранить задержку вызванную загрузкой элемента вектора. Для использования нескольких итераций потребуется большее количество регистров.

Представляем цикл развёрнутым четырехкратно, предполагаем, что R1 первоначально кратен четырем. Любые излишние вычисления устраняем, повторно не используем никакие регистры.

Loop:

LD F0, 0(R1)

1

ADDD F4, F0, F2

2

SD 0(R1), F4

0

LD F6, -8(R1)

1

ADDD F8, F6, F2

2

SD -8(R1), F8

0

LD F10, -16(R1)

1

ADDD F12, F10, F2

2

SD -16(R1), F12

0

LD F14, -24(R1)

1

ADDD F16, F14, F2

2

SD -24(R1), F16

0

SUBI R1, R1, #32

0

BNEZ R1, Loop

1

14 команд + 13 тактов = 27

27/4=6.75 такта на результат

Loop:

LD F0, 0(R1)

LD F6, -8(R1)

LD F10, -16(R1)

LD F14, -24(R1)

ADDD F4, F0, F2

ADDD F8, F6, F2

ADDD F12, F10, F2

ADDD F16, F14, F2

SD 0(R1), F4

SD -8(R1), F6

SD -16(R1), F10

SUBI R1, R1 #32

BNEZ R1, Loop

SD 8(R1), F16

14 тактов на все

14/4=3.5 такта на 1 результат

9 т/рез

после оптимизации

6 т/рез

после разворачивания

6.75 т/рез

после оптимизации разворачивания

3.5 т/рез

Одновременная выдача нескольких команд для выполнения и динамическое планирование

Методы минимизации приостановок конвейера из-за наличия в программах логических зависимостей по данным и по управлению, нацеленные на достижение идеального коэффициента CPI (Cycle-per-instruction). Идеальный CPI равен единице. Чтобы еще больше повысить производительность необходимо сделать CPI меньше единицы. Этого нельзя сделать, если в одном такте на выполнение выдается только одна команда. Следовательно, необходима параллельная выдача команд в каждом такте. Существует 2 типа подобного вида машин.

1. Супер скалярные машины
2. Машины с очень длинным машинным словом

Супер скалярные машины могут выдавать на исполнение в каждом такте переменное число команд. Работа их конвейера может планироваться как статически, с помощью компилятора, так и динамически с помощью средств оптимизации.

В отличие от супер скалярных машин, машины с длинным машинным словом выдают на исполнение фиксированные количество команд, которые сформатированы либо как одна большая команда, либо как пакет команд фиксированного формата. Планирование работы таких машин всегда осуществляется компилятором.

Супер скалярные машины используют параллелизм уровня команд, путем посылки нескольких команд из обычного потока команд в несколько функциональных устройств. Чтобы снять ограничение последовательного выполнения команд в этих машинах используются механизмы внеочередной выдачи и внеочередного завершения команд. Прогнозирование переходов. Кеши целевых адресов переходов. И условное, т.е. по предположению, выполнение команд. В типичной, супер скалярной машине, аппаратура может выдавать до 8 команд в одном такте. Обычно это должны быть независимые команды, удовлетворяющие некоторым ограничениям. Например, в каждом такте может выдавать не более одной команды обращения к памяти, если какая либо команда в потоке является зависимой, логически или не удовлетворяет критериям выдачи, на выполнение будут выданы только команды предшествующие данной, поэтому скорость выдачи команд в супер скалярных машинах является переменной, это и отличает их от машин с очень длинным машинным словом, в которых пакет команд, выполняемых одновременно, формирует компилятор, а аппаратура не принимает решений, относительно параллельной выдачи нескольких команд.

Первые супер скалярные компьютеры выдавали на выполнение 2 команды в одном такте. Одной из таких команд была загрузка регистров из памяти, запись регистра в память, команда перехода, операция целочисленного АЛУ. Другой командой могла быть команда с плавающей точкой. Параллельная выдача команд одной целочисленной и одной с плавающей точкой намного проще, чем выдача двух произвольных команд. В процессорах Pentium, например, применялся именно такой подход. Выдача двух команд в каждом такте требует одновременной выборки и декодирования по крайней мере 64 бит. Для упрощения декодирования команды можно располагать парами и выравнивать их по 64 битовым границам, в противном случае необходимо анализировать команды в процессе выборки и менять их местами в момент пересылки в ЦЧ устройство и ПТ. При этом возникают дополнительные требования к схемам обнаружения конфликтов. В любом случае вторая команда может выдаваться, если только на исполнение может быть выдана первая. Аппаратура обеспечивает выдачу второй команды, если условия выполнения первой не соблюдаются.

Чтобы спланировать цикл добавления скалярной величины, к вектору памяти для работы без задержек, необходимо развернуть его пятикратно.

|  |  |  |
| --- | --- | --- |
| № такта | Целочисленная команда | Команда с плавающей точкой |
| 1  2  3  4  5  6  7  8  9  10  11  12 | Loop:  LD F0, 0(R1)  LD F6, -8(R1)  LD F10, -16(R1)  LD F14, -24(R1)  LD F18, -32(R1)  SD 0(R1), F4  SD -8(R1), F8  SD -16(R1), F12  SD -24(R1), F16  SUBI R1, R1, #40  BNEZ R1, Loop  SD 8(R1), F20 | ADDD F4, F0, F2  ADDD F8, F6, F2  ADDD F12, F10, F2  ADDD F16, F14, F2  ADDD F20, F18, F2 |

12/5 = 2.4

2.4 на такт текущая скорость, в старом примере 3.5.

Команд ПТ недостаточно для полной загрузки конвейера ПТ, первоначальный, неразвернутый цикл, выполнялся со скоростью 6 тактов на результат, мы получили ускорение в 2.5 раза. Больше половины которого произошло за счет разворачивания цикла. Чистое разворачивание дало улучшение в 1.5 раза.

**Архитектура машин с длинным командным словом (VLIW – Very Long Instruction Word)**

VLIW-архитектура позволяет сократить объем оборудования и потенциально, чем большее количество команд выдается параллельно, тем больше эта экономия.

Супер скалярная машина, обеспечивающая выдачу двух команд одновременно, требует реализацию параллельного выполнения 2-х команд, 6 регистров, а так же динамического анализа возможности выдачи одной или двух команд и распределения этих команд по функциональным устройствам. Требования по объему аппаратуры для параллельной выдачи двух команд считаются умеренными. Увеличением количества выдаваемых параллельно команд для выполнения приводит к вырастанию сложности реализации из-за определения порядка следования команд и существующих между ними зависимостей.

Архитектура VLIW базируется на множестве изолированных друг от друга устройств. Вместо того, чтобы выдавать в эти устройства независимые команды, в таких машинах несколько операций упаковывается в одну очень длинную команду. При этом ответственность за выбор параллельно выдаваемых на выполнение операций полностью ложится на компилятор, а аппаратные средства необходимые для реализации супер скалярной обработки просто отсутствуют.

Рассмотрим работу цикла инкрементирования элементов вектора на VLIW машине в предположении, что одновременно могут выдаваться две операции обращения к памяти, 2 операции с плавающей точкой, одна целочисленная, либо одна команда перехода. Для устранения всех возможных приостановок конвейера цикл разворачивается семикратно.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | Обращение 1 | Обращение 2 | Плавающая 1 | Плавающая 2 | Целое/переход |
| 1 | LD F0, 0(R1) | LD F6, -8(R1) |  |  |  |
| 2 | LD F10, -16(R1) | LD F14, -24(R1) |  |  |  |
| 3 | LD F18, -32(R1) | LD F22, -40(R1) | ADDD F4, F0, F2 | ADDD F8, F6, F2 |  |
| 4 | LD F26, -48(R1) |  | ADDD F12, F10, F2 | ADDD F16, F14, F2 |  |
| 5 |  |  | ADDD F20, F18, F2 | ADDD F24, F22, F2 |  |
| 6 | SD 0(R1), F4 | SD -8(R1), F8 | ADDD F28, F26, F2 |  |  |
| 7 | SD -16(R1), F12 | SD -24(R1), F16 |  |  | SUBI R1, R1, #56 |
| 8 | SD 24 (R1), F20 | SD 16 (R1), F24 |  |  | BNEZ R1, Loop |
| 9 | SD 8(R1), F28 |  |  |  |  |

Один проход по циклу осуществляется за 9 тактов и вырабатывает 7 результатов. На вычисление каждого результата расходуется 1.28.

**Д/З:** Концепция машин с явным параллелизмом (EPIC). Изучить микроархитектуру процессора Itanium 2 и концепцию EPIC.

Для машин с VLIW архитектурой был разработан новый метод планирования команд. При использовании этого метода из последовательности исходной программы генерируются длинные команды, путем просмотра программы за пределами базовых блоков. С точки зрения архитектурных идей, машину с длинным командным словом можно рассматривать как расширение RISC архитектуры. Как и в RISC архитектуре аппаратные ресурсы VLIW машины предоставлены компилятору и ресурсы планируются статически. К этим ресурсам относятся конвейерные функциональные устройства, шины, банки памяти. Для поддержки высокой пропускной способности используется большее количество регистров, предпочтение отдается простой логике управления. В отличие от традиционных машин регистры и шины не резервируются, а их использование определяется во время компиляции. Калькулятор с трассировочным планированием определяет участок программы без обратных дуг, т.е. без переходов назад, который становится кандидатом для составления расписания. Обратные дуги, обычно, имеются в программах с циклами. Для увеличения размера тела цикла, используется раскрутка цикла, что приводит к появлению больших фрагментов программы без обратных дуг. Если в программе имеются только переходы вперед, то компилятор выполняет эвристическое предсказание путей. Используется для оптимизации проводимой с учетом зависимостей между командами и ограничений аппаратуры. Во время планирования генерируется длинное командное слово. Все операции длинного командного слова выдаются одновременно и выполняются параллельно. Процесс продолжается до тех пока не будет оптимизирована вся программа. Ключевым условием достижения эффективной работы VLIW машины - является корректное предсказание переходов.

**Буферы прогнозирования условных переходов.**

Простейшей схемой динамического прогнозирования направления условных переходов является буфер прогнозирования или таблица истории условных переходов (branch prediction buffer). Буфер прогнозирования УП представляет собой небольшую память, адресуемую с помощью младших разрядов адреса команды перехода. Каждая ячейка этой памяти содержит один бит, который говорит о том, был ли предыдущий переход выполняемым или нет. В действительности не известно, является ли прогноз корректным. Т.к. значение буфера могла установить совсем другая команда перехода, которая имела тоже самое значение младших разрядов адреса, но это не имеет значения, прогноз – предположение, которое рассматривается как корректное и выборка команд начинается по прогнозируемому направлению. Если же предположение окажется неверным, бит (вид) прогноза инвертируется. Производительность буфера зависит от частоты его применения и точности предсказания.

**Пример:** команда условного перехода в цикле, которая являлась выполняемым переходом выполнялась 9 раз подряд, а затем не выполнилась. При однобитовой схеме прогноза направление перехода будет предсказано неверно дважды. На первой и последней итерации цикла. На последней итерации неправильный прогноз неизбежен, поскольку перед этим переход был 9 раз подряд выполняемым. Неправильный прогноз происходит из-за того, что бит прогноза инвертируется, поскольку на последней итерации переход был невыполняемым. Таким образом, точность прогноза, который выполнялся в 90% случаев составляет 80%. 2 некорректный и 8 корректных. В общем случае для команд условного перехода, переход является выполняемым много раз подряд, а затем является невыполняемым. Следовательно однобитовая схема предсказаний будет предсказывать неверно на двух итерациях, на первой и на последней.

Для повышения точности прогнозирования используется схема двух битового прогноза. В двух битовой схеме ошибка должна быть сделана дважды, прежде чем он изменится на противоположное значение. Диаграмма двух битовой схема предсказания.

выполняется

да

Не выполн.

нет

Не выполн.

Выполн.

Выполн.

Не выполняется

Двух битовая схема прогнозирования в действительности является частным случаем более общей схемы, которая в каждой строке буфера прогнозирования имеет n-битовый счетчик. Этот счетчик может принимать значение от нуля до , при этом схема прогноза будет следующей.

1. Если значение больше либо равно в точке на середине интервала , то переход объявляется как правильный, если переход выполняется – то к значению добавляется единица, если только это значение не достигло максимальной величины. Если прогноз был неверный, из счетчика вычитается единица.
2. Если значение счетчика меньше , то переход прогнозируется как невыполняемый, если направление перехода оказывается правильным, то от значения счетчика вычитается единица, если не правильным, то добавляется единица.

Исследование n-битовых схем прогнозирования показывают, что она работает почти так же хорошо, поэтому почти во всех случаях используют двух битовые схемы.

**Коррелированные схемы прогнозирования переходов.**

Двух битовая схема прогнозирования использует информацию о недавнем поведении команды условного перехода для прогноза будущего команды этой же команды. Вероятно, можно улучшить точность прогноза, если учитывать не только поведение того перехода, который пытаемся предсказать, но рассматривать также недавнее поведение других команд перехода.

**Рассмотрим пример:** фрагмент из тестового пакета SPEC92.

If (a==2) { a=0; }

If (b==2) { b=0; }

If (a!=b) {некие действия }

A и B размещены в регистрах r\_1 и r\_2, а r\_0 равен нулю.

SUBI R3, R1, #2 ; R3 = a-2

BNEZ R3, L1 ; переход в L2 (a != 2)

ADD R1, R0, R0 ; a = 0

L1: SUBI R3, R2, #2 ; R3 = b-2

BNEZ R3, L2; переход в L2 (a != 2)

ADD R2, R0, R0 ; b=0

L2: SUB R3, R1, R3 ; R3 = a-b

BEQZ R3, L3 ; переход в L3 (a==b)

…

…

L3: …

Схемы прогнозирования, которые для предсказания перехода используют поведение других команд перехода называются коррелированными. Схема прогнозирования называется прогнозом 1,1 если она используется поведение одного последнего перехода для выбора из пары однобитовых схем прогнозирования на каждый переход. Схема прогнозирования n,m использует поведение последних m переходов, для выбора из схем прогнозирования, каждая из которых представляет собой n-битовую схему прогнозирования для каждого отдельного перехода. Привлекательность коррелируемых схем прогнозирования переходов заключается в том, что они могут давать больший процент успешного прогнозирования, чем обычная двух битовая схема и требуют небольшого объема дополнительной аппаратуры. Глобальная история последних m переходов может быть записана в m-битовом сдвиговом регистре. Каждый разряд которого запоминает, был ли переход выполняемым или нет. Тогда буфер прогнозирования переходов может индексироваться конкатенацией младших разрядов адреса перехода с n-битовой глобальной историей.

**Д/З:** попытаться представить схему 2,2.

**Буфер прогнозирования коррелированной схемы прогнозирования 2,2.**

Рисунок #1.

На рисунке буфер типа 2,2 с общим числом строк, равным 64, где 4 младших разряда адреса команды перехода и 2 бита глобальной истории коррелированных переходов формируют 6-ти битовый индекс, который может использоваться для обращения к 64 счетчикам.

Сравнение простой двух битовой схемы прогнозирования с 4К строк и схемы прогнозирования 2,2 с 1к строк показывает, что схема 2,2 не только превосходит двух битовую схему с тем же количеством состояний, но часто превосходит даже двух битовую схему прогнозирования с бесконечным количеством строк.

Рассмотрим ситуацию при которой на стадии выборки находится команда перехода. На следующей стадии будет выполняться её дешифрация, для сокращения потерь необходимо знать по какому адресу необходимо выбирать следующую команду, для этого надо выяснить, что еще не дешифрованная команда является командой перехода, для этого нужно знать счетчик. Если знать все это, то потери на команду перехода могут быть сведены к нулю. Специальный аппаратный кэш прогнозирования переходов, который хранит прогнозируемый адрес следующей команды, называется буфером целевых адресов перехода (branch target buffer). Каждая строка этого буфера включает программный адрес команды перехода, прогнозируемый адрес следующей команды и предысторию команды перехода.

Рисунок #2.

Существуют вариации данного метода. Основная суть вариаций заключается в том, чтобы хранить в процессоре одну или несколько команд из прогнозируемой ветви перехода. Метод может применяться вместе с буфером целевых адресов перехода и без него. При этом появляется дополнительное преимущество. Буферизация самих целевых команд позволяет использовать метод свертывания переходов (branch folding). Свертывание переходов может использоваться для реализации нулевого времени выполнения для реализации самих команд перехода и в некоторых случаях нулевого времени выполнения условных переходов. Рассмотрим буфер команд перехода, который буферизует команды из прогнозируемой ветви. Пусть к нему выполняется обращение по адресу команды без условного перехода. Единственной задачей этой команды безусловного перехода является замена текущего значения программного счетчика. В этом случае, когда буфер адресов регистрирует попадание и показывает, что переход безусловный, конвейер просто может заменить команду, которая выбирается из кэш памяти, т.е. саму команду безусловного перехода на команду из буфера. Таким образом в некоторых случаях удается убрать потери для команд условного перехода, если код условия установлен заранее.

**Д/З:** представить такой случай.

Схемы прогнозирования условных переходов ограничены, как точностью прогноза, так и потерями в случае неправильного прогноза. Типичные схемы прогнозирования достигают 80-95% точности, в зависимости от типа программы и размера буфера. Кроме увеличения точности можно пытаться уменьшить потери при неверном прогнозе, обычно это делается с помощью выборки команд из обеих ветвей условия, т.е. по предсказанному и не предсказанному условию. Это требует двух портовой памяти, кэш памяти с расслоением или осуществления выборки в начале по одному направлению, а затем по другому. Хотя подобная организация увеличивает стоимость организации, возможно, это единственный способ снижения потерь на условные переходы ниже определенного уровня. Другие решения заключаются в кэшировании адресов и команд в другом буфере. В современных микропроцессорах используется около двух десятков различных алгоритмов.

**Кэш память**

В основе реализации иерархии памяти современных компьютеров лежат 2 основных принципа:

1. Принцип локальности обращений
2. Соотношение стоимость/производительность

Принцип локальности обращений говорит о том, что большинство программ не выполняют обращений ко всем своим командам и данным равновероятно, а оказывают предпочтения некоторой области адресного пространства. Иерархия памяти современных компьютеров строится на нескольких уровнях, причем более верхний уровень меньше по объему, быстрее и имеет большую стоимость в пересчете на один байт хранения, чем более низкий. Уровни иерархии взаимосвязаны. Данные некоторого уровня могут быть найдены и на более низком уровне. Сверху вниз до конца иерархии, от начала до конца. Иерархия памяти может состоять из многих уровней, но в каждый момент времени рассматриваются только 2 близ лежащих. Минимальная единица информации, которая может присутствовать, а может отсутствовать в двух уровневой реализации называется блоком. Размер блока может быть, как фиксированным, так и переменным. Если размер блока фиксированный – то объем памяти кратен размеру блока. Успешное или не успешное обращение к высшему уровню называется попаданием или промахом. Попадание – есть обращение к объекту в памяти, который найден на более высоком уровне. Промах означает, что он не найден на этом уровне. Коэффициент попаданий есть доля обращений, найденных на более высоком уровне. Может представляться в процентах. Доля промахов есть доля обращений, которые не найдены на более высоком уровне. Коэффициенты попаданий и промахов являются важными характеристиками. Чтобы описать некоторый уровень иерархии памяти надо ответить на следующие 4 вопроса:

1. Где может размещаться блок на верхнем уровне иерархии?
2. Как можно найти блок, когда он находится на верхнем уровне?
3. Какой блок должен быть замещен в случае промаха?
4. Что происходит во время записи?

**Организация кэш памяти**

Концепция кэш памяти возникла раньше, чем архитектура IBM 360. Сегодня кэш память имеется во всех классах компьютеров, и даже во множественном числе.

Типовые значения ключевых параметров для кэша памяти рабочих станций и серверов.

* Размер блока до 128 байт.
* Время попадания – один такт синхронизации.
* Потери при промахе – до 10-ти тактов синхронизации.
* Время пересылки – до 20-ти тактов.
* Доля промахов – от 1 до 20%.
* Размер кэш памяти – до 16МБ.

Детальная организация кэш памяти.

**Размещение блока в кэш памяти.**

По типу размещения определяют 3 основных типа организации:

Если каждый блок памяти имеет только одно фиксированное место, на котором он может появиться, то такая память называется кэшем с прямым отображением. Это наиболее простая организация кэша, при которой для отображения адресов блоков основной памяти на адреса кэш памяти используются младшие разряды адреса блока. Таким образом все блоки основной памяти, имеющие одинаковые младшие разряды в своем адресе попадают в один блок кэш памяти. **Д.З. проблема aliasing кэш памяти.**

Direct Mapped Cache (DMC)

Адрес проц.

|  |  |  |
| --- | --- | --- |
| Tag | Line | Offset |
| 0010 | 001 | 100 |
| 2 | 1 | 4 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Tag | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 0 |  |  |  |  |  |  |  |  |  |
| 1 | 2 |  |  |  |  | X |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |
| 3 |  |  |  |  |  |  |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |
| 6 |  |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  |

Процессор выберет заданную ячейку, если указаны нужные line и offset и tag равен значению, заданному в таблице.

**Следующий принцип организации кэш памяти.**

Если некоторый блок памяти может располагаться в любом месте кэш памяти – то такой кэш называется полностью ассоциативным.

Fully Associative Cache (FAC)

|  |  |
| --- | --- |
| Tag | Offset |
| 0010001 | 4 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Tag | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 0 |  |  |  |  |  |  |  |  |  |
| 1 | 2 |  |  |  |  |  |  |  |  |
| 2 |  |  |  |  |  |  |  |  |  |
| 3 | 17 |  |  |  |  | X |  |  |  |
| 4 |  |  |  |  |  |  |  |  |  |
| 5 |  |  |  |  |  |  |  |  |  |
| 6 | 45 |  |  |  |  |  |  |  |  |
| 7 |  |  |  |  |  |  |  |  |  |

**N-way Set Associative Cache (NSAC)**

|  |  |  |
| --- | --- | --- |
| Tag | Set | Offset |
| 00100 | 01 | 100 |
| 7 | 1 | 4 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Tag | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 0 0set |  |  |  |  |  |  |  |  |  |
| 1 0set | 2 |  |  |  |  |  |  |  |  |
| 2 1set | 4 |  |  |  |  | X |  |  |  |
| 3 1set | 17 |  |  |  |  |  |  |  |  |
| 4 2set |  |  |  |  |  |  |  |  |  |
| 5 2set |  |  |  |  |  |  |  |  |  |
| 6 3set | 45 |  |  |  |  |  |  |  |  |
| 7 3set |  |  |  |  |  |  |  |  |  |

Обычно множество представляет собой группу из двух или большего числа блоков в кэше. Если множество состоит из n блоков, то такое размещение называется множественно-ассоциативным с n каналами. Для размещения блока прежде всего необходимо определить множество. Множество определяется младшими разрядами адреса блока памяти.

Диапазон возможных организаций кэш-памяти очень широк: кэш-память с прямым отображением есть просто одноканальная множественно-ассоциативная кэш-память, а полностью ассоциативная кэш-память с m блоками может быть названа m-канальной множественно-ассоциативной. В современных процессорах как правило используется либо кэш-память с прямым отображением, либо n-канальная множественно-ассоциативная кэш-память.

У каждого блока в кэш-памяти имеется адресный тег, указывающий, какой блок в основной памяти данный блок кэш-памяти представляет. Эти теги обычно одновременно сравниваются с выработанным процессором адресом блока памяти.

Кроме того, необходим способ определения того, что блок кэш-памяти содержит достоверную или пригодную для использования информацию. Наиболее общим способом решения этой проблемы является добавление к тегу так называемого бита достоверности (valid bit).

Адресация множественно-ассоциативной кэш-памяти осуществляется путем деления адреса, поступающего из процессора, на три части: поле смещения используется для выбора байта внутри блока кэш-памяти, поле индекса определяет номер множества, а поле тега используется для сравнения. Если общий размер кэш-памяти зафиксировать, то увеличение степени ассоциативности приводит к увеличению количества блоков в множестве, при этом уменьшается размер индекса и увеличивается размер тега.

**Какой блок кэш памяти должен быть замещен при промахе?**

При возникновении промаха, контроллер кэш-памяти должен выбрать подлежащий замещению блок. Польза от использования организации с прямым отображением заключается в том, что аппаратные решения здесь наиболее простые. Выбирать просто нечего: на попадание проверяется только один блок и только этот блок может быть замещен. При полностью ассоциативной или множественно-ассоциативной организации кэш-памяти имеются несколько блоков, из которых надо выбрать кандидата в случае промаха. Как правило для замещения блоков применяются две основных стратегии: случайная и LRU.

В первом случае, чтобы иметь равномерное распределение, блоки-кандидаты выбираются случайно. В некоторых системах, чтобы получить воспроизводимое поведение, которое особенно полезно во время отладки аппаратуры, используют псевдослучайный алгоритм замещения.

Во втором случае, чтобы уменьшить вероятность выбрасывания информации, которая скоро может потребоваться, все обращения к блокам фиксируются. Заменяется тот блок, который не использовался дольше всех (LRU - Least-Recently Used).

Достоинство случайного способа заключается в том, что его проще реализовать в аппаратуре. Когда количество блоков для поддержания трассы увеличивается, алгоритм LRU становится все более дорогим и часто только приближенным.

**Что происходит во время записи?**

При обращениях к кэш-памяти на реальных программах преобладают обращения по чтению. Все обращения за командами являются обращениями по чтению и большинство команд не пишут в память. Обычно операции записи составляют менее 10% общего трафика памяти. Желание сделать общий случай более быстрым означает оптимизацию кэш-памяти для выполнения операций чтения, однако при реализации высокопроизводительной обработки данных нельзя пренебрегать и скоростью операций записи.

К счастью, общий случай является и более простым. Блок из кэш-памяти может быть прочитан в то же самое время, когда читается и сравнивается его тег. Таким образом, чтение блока начинается сразу, как только становится доступным адрес блока. Если чтение происходит с попаданием, то блок немедленно направляется в процессор. Если же происходит промах, то от заранее считанного блока нет никакой пользы, правда нет и никакого вреда.

Однако при выполнении операции записи ситуация коренным образом меняется. Именно процессор определяет размер записи (обычно от 1 до 8 байтов) и только эта часть блока может быть изменена. В общем случае это подразумевает выполнение над блоком последовательности операций чтение-модификация-запись: чтение оригинала блока, модификацию его части и запись нового значения блока. Более того, модификация блока не может начинаться до тех пор, пока проверяется тег, чтобы убедиться в том, что обращение является попаданием. Поскольку проверка тегов не может выполняться параллельно с другой работой, то операции записи отнимают больше времени, чем операции чтения.

Очень часто организация кэш-памяти в разных машинах отличается именно стратегией выполнения записи. Когда выполняется запись в кэш-память имеются две базовые возможности:

* сквозная запись (write through, store through) - информация записывается в два места: в блок кэш-памяти и в блок более низкого уровня памяти.
* запись с обратным копированием (write back, copy back, store in) - информация записывается только в блок кэш-памяти. Модифицированный блок кэш-памяти записывается в основную память только когда он замещается. Для сокращения частоты копирования блоков при замещении обычно с каждым блоком кэш-памяти связывается так называемый бит модификации (dirty bit). Этот бит состояния показывает был ли модифицирован блок, находящийся в кэш-памяти. Если он не модифицировался, то обратное копирование отменяется, поскольку более низкий уровень содержит ту же самую информацию, что и кэш-память.

**Д.З. сравнение подходов к организации записи.**

**Общая организация современного микропроцессора**

Поток команд записывается в специальную очередь – FIFO. Очередь размером 128 байт (ROB = re order buffer). RS – станции резервирования или планировщик очереди на исполнение.

MOB – буфер, в котором накапливаются данные, подлежащие пересылке в кэш.

**Микроархитектура Nehalem фирмы intel.**

Причины появления микроархитектуры.

Предыдущая архитектура Core не удовлетворяет требованиям построения многоядерных систем. Многоядерные системы в архитектуре Core собирались из двуядерных кристаллов, что приводило к затруднению взаимодействия между ними. Обмен данными между разрозненными ядрами происходил через системную память. Это могло вызывать большие задержки из-за ограниченной пропускной способности процессорной шины. Кроме того, увеличение количества ядер в многопроцессорной системе ограничивалось невысокой пропускной способностью шины памяти. Поэтому фирма intel в новой архитектура попыталась решить указанные структурные проблемы.

Ключевыми особенностями новой архитектуры стали: интегрированный контроллер памяти, новая шина с топологией точка-точка, которая может использоваться для создания многопроцессорной системы, а не только связывать процессор с чипсетом.

Самым важным нововведением является модульный дизайн процессора. Микроархитектура включает несколько стандартных строительных блоков, из которых на конечном этапе проектирования и производства могут быть собраны итоговые процессоры. Этот набор блоков включает в себя: процессорное ядро с кэшем второго уровня, общий кэш третьего уровня, контроллер шины QPI, контроллер памяти, графическое ядро, блок управления электропитанием и тактовой частотой.

Из этого можно построить процессор Bloomfield. Включает в себя: 4 ядра, кэш L3, контроллер памяти и контроллер шины.

Серверные процессоры включают 8 и более ядер, 4 или более контроллеров шины, кэш L3, контроллер памяти и так далее.

Технологические нововведения в микроархитектуре:

1. Появление технологии SMT – технология много поточности. Нужна для исполнения одновременно двух вычислительных потоков на одном ядре.
2. Поддержка новых команд, потоковых расширений, SSE 4.2.

В первую очередь в микроархитектуре модификация затронула декодеры. В архитектуре Nehalem имеется один комплексный декодер для дешифровки комплексных команд и три простых для дешифровки простых команд. Эти 4 декодера способны формировать в каждом такте до 5-ти команд. Некоторые пары команд, благодаря технологии macro fusion, сливаются в одну (загрузка/выполнение). Технология macro fusion стала работать в 64 битном режиме (ранее работала только в 32-х битном). Увеличилось число пар команд для слияния. Сюда вошли пары, образованные командой сравнения и условного перехода. В начальной стадии конвейера существовал loop stream detector – этот блок был перенесен и установлен после декодера. Это позволяет не загружать и повторно детектировать короткие циклы. Этот блок предназначен только для хранения коротких циклов.

Блок исполнительных устройств процессора остался без сильных изменений.

Унифицированная станция резервирования. Планирует операцию загрузки функциональных устройств. Имеет один планировщик на все функциональные устройства. Может быть использована для всех операций с плавающей точкой.

Исполняющие устройства могут выполнять до 6 операций за такт. 1 операция загрузки, 1 операция сохранения адреса, 1 операция сохранения данных и 3 вычислительных операции. Длина буфера ROB – 128 операций. Величина буфера планировщика расширена с 32 до 36 инструкций. Увеличена вместимость буфера для работы с данными. Ядра новых процессоров поддерживают технологию SMT и поддерживают работу по 2 вычислительным потокам и нуждающихся в разделении ресурсов друг друга. В результате выросла производительность без существенных энергетических затрат. В новой архитектуре технология SMT дает больший выигрыш, чем технология Hyper Threading. Возможность обрабатывать большее число инструкций одновременно, существенного увеличения сложности процессора не произошло. Остальные ресурсы при включении SMT либо динамически между потоками, как станции резервирования или кэш, либо жестко пополам, как ROB. Иногда активация SMT может приводить к снижению производительности. В новой архитектуре физические и логические ядра различимы и не полноправны. Программист, при необходимости, может самостоятельно решать вопрос о правильном использовании ресурсов.

**TLB и кэш.**

TLB – translation look aside buffer. Увеличение размера TLB позволяет повысить число страниц в памяти, которые могут быть одновременно использованы, без дополнительных дорогостоящих преобразований по таблицам трансляции адресов, находящимся в обычной памяти. Изменения сделаны с прицелом на серверные приложения. Дополнительные нововведения, поднимающие скорость работы, заключаются в ускорении работы инструкций.

**Кэш память**

L1 – разделенный, L2 – совмещенный, L3 - для разделения приложениями.

Intel не отказалась от дублирования данных в кэшах с 1-3 уровни. Все 3 уровня кэша инклюзивные. Инклюзивный разделяемый кэш обеспечивает в многоядерных процессорах более высокую скорость работы подсистемы памяти за счет избыточного дублирования содержимого кэшей первого и второго уровня всех ядер.

Если некоторое ядро модифицирует данные в кэше L3, изначально принадлежащие другим ядрам, то в этом случае обновляются кэши уровней 1 и 2 и этих ядер. Таким путем решается проблема с избыточным меж ядерным трафиком, направленным на поддержание когерентности эксклюзивной кэш памяти.

Новые SSE инструкции – добавлено 5 инструкций для ускорения обработки строк и текстов. Новые инструкции для аккумулирования hash функций.

**Интегрированный контроллер памяти**

Главное свойство контроллера памяти в микроархитектуре Nehalem – гибкость. Фирма intel нашла возможность варьировать число каналов и скорость памяти. Процессоры в четырех ядерном варианте чаще будут иметь трех канальный контроллер памяти с поддержкой DDR3 SDRAM. Основное преимущество переноса контроллера заключается в уменьшении латентности памяти. Теперь между процессорами памяти нет промежуточных устройств. Ранее за работу с памятью отвечал северный мост чипсета, который вносил собственные задержки из-за необходимости синхронизации шин памяти и процессора.

Шина QPI. Интеграция контроллера памяти в процессор разгружает процессорную шину, которая оказывается свободна от передачи данных между процессором и памятью. Это верно, но только для однопроцессорных систем. Архитектура Nehalem используется в серверных продуктах, поэтому спроектирована шина для много процессорных систем, обеспечивает высокую пропускную способность и масштабируемость. Quick Path Interconnect. С технической точки зрения шина QPI представляет из себя 2 20-ти битных соединения. 20 в одну и 20 в другую сторону. 16 – для передачи данных, еще 4 – вспомогательные. Используется протоколом и коррекцией ошибок. Шина работает на максимальной скорости 6.4 миллионов передач данных в секунду. Имеет, соответственно, пропускную способность 12.8 ГБ/с.

**Классификация систем параллельной обработки данных**

Любая вычислительная система достигает максимальной производительности благодаря использованию высокоскоростным элементов и параллельного выполнения большого числа операций. Именно возможность параллельной работы устройств является причиной ускорения вычислений. Наиболее распространена классификация Флинна. По Флинну все вычислительные системы разделяются на следующие классы:

1. Машины типа Single Instruction Multiply Data (Один поток команд применяется к множеству потоков данных).
2. Multiply Instruction Multiply Data (Множество команд применяется к множеству потоков данных).
3. SISD – одноядерные настольные компьютеры.
4. MISD – сугубо теоретический.

Классификация Флинна не делает различия по другим важным характеристикам, по уровню зернистости параллельных вычислений и по уровню синхронизации.

Одна и та же операция применяется ко всем элементам вектора, или элементам пары векторов. Для настройки конвейера на выполнение векторной операции может потребоваться некоторое установочное время, однако затем операнды могут поступать на конвейер с максимальной скоростью. Задержек, связанных с выбором новой команды или выбором ветви вычислений, здесь не происходит. Операциям обработки применяемых к блоку данных в программе соответствуют компактные или небольшие цифры. Коротко о конвейерных и векторных машинах.

Второй тип машин (SIMD) состоят из большого числа процессорных элементов, имеющих свою память. Все процессорные элементы выполняют одну и ту же программу. Машина, состоящая из большого количества процессов, обеспечивает высокую производительность только на тех задачах, в которых все процессоры могут выполнять одну и ту же работу. Модель вычислений для модели SIND очень похожа на модель для векторного процессора. Одиночная операция выполняется над большим блоком данных. Для большинства SIND машин синонимом является термин матричный процессор. Обрабатывающие элементы таких процессоров – это универсальные программируемые ЭВМ, так что задача, решаемая параллельно, может быть сложной и содержать ветвления.

Третий класс (MIMD). Термин мультипроцессор покрывает большинство машин типа MIMD и используется как синоним для машин этого типа. В такой системе каждый процессорный элемент исполняет свою программу достаточно независимо от других процессорных элементов. По способу связи между собой процессорных элементов классификация детализируется. Различают сильно связанные мультипроцессоры или мультипроцессоры с общей памятью. С общей памятью процессорные элементы связываются общей шиной или сетью обмена. В противоположность этому варианту в слабо связанных системах вся память делится между процессорными элементами, и весь блок памяти доступен только связанному с ним процессорному элементу, а сеть обмена связывает процессорные элементы друг с другом. Базовой моделью вычислений на MIMD модели, является совокупность процессоров, эпизодически обращающейся к памяти. У этой модели существует большое количество вариантов. На одном конце спектра находится модель распределенных вычислений. В этой модели программа делится на большое число параллельных задач, состоящих из множества подпрограмм. На другом конце спектра модель потоковых вычислений, в которых каждая операция может рассматриваться как отдельный процесс. Такая операция ждет своих операндов, которые должны быть переданы ей другими процессами. Результат операции передается тем процессам, которые в нем нуждаются. В потоковых моделях вычислений большого и среднего уровня гранулярности процессы содержат большое число операций, выполняются в потоковой манере.

Многопроцессорные машины с SIMD процессорами. Многие современные суперэвм представляют собой многопроцессорные системы, в которых в качестве процессоров используются векторные процессоры или процессоры типа SIMD. Такие машины могут быть классифицированы как MSIMD. Языки программирования для машин типа MSIMD обеспечивают способность описывать крупнозернистый параллелизм. В пределах каждый задачи компилятор векторизует задачи. Модели типа MSIMD представляет возможность использовать лучший из двух способов векторизации.

В архитектурах с локальной памятью непосредственное разделение памяти невозможно, вместо этого организуется доступ к совместно используемым данным посредством передачи сообщений по сети обмена. Существующие MIMD машины распадаются на 2 класса по количеству объединяемых процессоров. От их количества зависит многое.

1. Машины с общей или разделяемой памятью. Используется объединение до 32-х процессоров. Небольшое количество процессоров позволяет иметь централизованную общую память и объединять процессоры с памятью одной общей шиной. При наличии у процессоров кэш памяти требования к пропускной способности шины и памяти снижаются. Такие машины называются или относят к классу UMA (Uniform Memory Access).
2. Представляется крупномасштабными системами с распределенной памятью. При этом подходе необходимо реализовать связь процессоров между собой. Локальная память позволяет устранить недостатки, связанные с ограниченной полосой пропускания общей памяти, не позволяющей удовлетворить запросы от большого количества процессоров. Основная разница в архитектуре, которая выделяется в машинах с распределенной памятью, заключается в том, как осуществляется связь между процессорами и в логической модели памяти. Имеется 2 альтернативных способа адресации такого типа памяти и 2 альтернативных метода передачи данных между процессорами: физически активные устройства могут адресовать как логически единое адресное пространство. Можно выполнять обращения к любым ячейкам в памяти. Такие машины называются машинами с распределенной разделяемой памятью. Некоторые авторы классифицируют их как NUMA (non-UMA).

Механизмы обмена через общую память. Простота программирования.

Архитектуры параллельного действия

1. SISD
2. SIMD
   1. Векторные
   2. Матричные
3. MISD
   1. ?
4. MIMD
   1. Мультипроцессоры (SMP – symmetrical multi-processor)
      1. UMA
         1. С шиной
         2. С координатным коммутатором
      2. COMA
      3. NUMA
         1. CC-NUMA
         2. NC-NUMA (no cache)
   2. Мульти компьютеры
      1. MPP
         1. Решетка
         2. Куб
      2. COW

При оценке производительности системы часто используют только оценку процессора и пренебрегают оценкой системы ввода-вывода. Это противоречит здравому смыслу. Компьютер без устройств ввода-вывода, как автомобиль без колес. Одной из более правильных оценок производительности системы является время ответа. Т.е. время между моментом ввода задания и получением результата. Это учитывает все накладные расходы на выполнение задачи в системе. Кроме того, в настоящее время только по организации системы ввода-вывода можно как-то грубо различать классы вычислительных систем. Разница между мейнфреймом и миникомпьютером заключается в том, что первый поддерживает на много больше терминалов и дисков. Разница между миникомпьютером и рабочей станцией состоит в том, что рабочая станция оснащена клавиатурой, мышью и экраном. А между файловой системой и файл-сервером в том, что файл-сервер имеет огромное количество дисков, а экран, клавиатура и мышь отсутствуют. Разница между рабочей станцией и компьютером в том, что рабочие станции соединены в локальную сеть.

Различия в стоимости и производительности определяются организацией систем памяти и устройств ввода-вывода.

Необходим механизм их взаимодействия. Подсистемы должны эффективно обмениваться данными. Одним из простейших механизмов, позволяющих организовать такое взаимодействие. Доступ к шине разделяется между всеми подсистемами. Преимуществом такой системы являются: низкая стоимость и универсальность. Поскольку шина – место для подключения различных устройств, то новые устройства легко добавляются в систему и одинаковые устройства могут применяться в разных системах. Стоимость такой организации получается низкой, поскольку для передачи информации используются набор линии шины, разделяемых множеством устройств.

Главным недостатком системы с единственной шиной является то, что шина создает узкое горло, ограничивая максимальную производительность обмена. В коммерческих системах, где ввод-вывод осуществляется часто, а также в суперкомпьютерах, из-за высокой производительности процессора, одним из главных вопросов разработки является создание системы нескольких шин, способной удовлетворять все запросы. Трудности при разработке шин связаны с ограничениями накладываемыми физическими факторами. К этим факторам относят длину шины и количество подключаемых устройств. Требования быстродействия системы ввода-вывода и высокой пропускной способности являются и противоречивыми. В современных системах используется комплекс взаимосвязанных шин. Чрезмерно высокую пропускную способность и избыточность используют для компенсации отказоустойчивости и эффективности. Традиционно шины делятся на 2 типа: процессор с памятью, шины ввода-вывода.

1. Процессор с память – высокоскоростные, короткие и соответствуют организации системы памяти для обеспечения максимальной пропускной способности.
2. Ввода-вывода – имеют большую протяженность, много типов подключаемых устройств.

Персональные компьютеры в недавнем прошлом строились на основе одной системной шины ISA, EISA, MCA, PCI, PCI Express. Для сохранения производительности и баланса процессов используют 2-х уровневую организацию шин. Локальная шина объединяет процессор, память и контроллер.

Рассмотрим транзакции на шине:

1. Запись (ЦП посылает в память адрес и данные и не ожидает возврата)
2. Чтение (инициализация чтения, память отвечает, возвращая данные)

Разработка связаны с реализацией дополнительных возможностей. Выбор той или иной возможности связан со стоимостью и производительностью.

|  |  |  |
| --- | --- | --- |
| Возможность | Высокая производительность | Низкая стоимость |
| Общая разрядная шина | Отдельные линии для адресов и данных | Мультиплексирование адресов и данных |
| Ширина шины данных | Чем шире, тем быстрее | Чем уже, тем дешевле |
| Размер пересылки | Чем больше слов, тем меньше накладные расходы | Пересылка одного слова дешевле |
| Главное устройство шины | Несколько (требуется арбитраж) | Одно (арбитраж не нужен) |
| Расщепление транзакции | Да – пакеты (запрос, ответ) | Нет (задержка меньше) |
| Тип синхронизации | Синхронные шины | Асинхронные шины |
|  |  |  |

Синхронная шина включает сигналы синхронизации и фиксированный протокол определяющий положение адресов и данных, относительно синхронизирующего сигнала. Синхронные шины быстры и дешевы. Их недостатки – работа на одной и той же частоте всех устройств, не может быть длинной из-за проблем перекоса сигналов.

## Шина PCI (Peripheral component interconnect)

**Основные возможности шины**

1. Синхронный 32 или 64 битный обмен.
2. Поддержка пяти и трех вольтовой логики.
3. Полная поддержка нескольких активных устройств на шине.
4. Поддержка кэша с прямой и обратной записью.
5. Автоматическое конфигурирование карт расширения при включении питания. На одной карте расширения могут комбинироваться до 8 функций. Шина позволяет использовать 4 слота расширений. Для увеличения количества карт расширений используется специальный мост PCI2PCI. Все PCI устройства оборудованы таймером, использующимся для определения максимального промежутка времени, в течение которого устройство может владеть памятью. Шина поддерживает метод передачи времени. Метод предполагает, что пакет считывается или записывается одним куском, с автоматическим увеличением адреса для одного байта. Скорость передачи данных при этом увеличивается за счет увеличения количества передаваемых адресов.

Временная диаграмма



В каждой транзакции участвует 2 устройства: инициатор обмена и целевое устройство. Шина PCI трактует все транзакции как пакетные. Транзакция начинается фазой адреса, за которой могут следовать одна или несколько фаз данных. Адреса и данные передаются по мультиплексируемой шине AD. Линии C/BE для кодирования команд в фазе адреса и разрешения в фазе данных. В начале инициатор активирует сигнал FRAME по шине передает целевой адрес, а по линиям C/BE – команду или тип транзакции. Адресованное целевое устройства отзывается сигналом DEVSEL после чего инициатор может указать свою готовность к обмену сигналом IRDY, при готовности к обмену устройство выставляет сигнал TRDY. Данные по шине AD, только при наличии обоих сигналов. Если оба устройства вводят сигналы в конце фазы адреса и не снимают до конца обмена, то в каждом такте после фазы адреса передаются 32 бита данных, что соответствует предельной скорости обмена.

**Конфигурационное пространство шины PCI.**

Доступ к данному адресному пространству осуществляется с помощью специальных команд.

Конфигурационное пространство.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 31  16 | | 15  0 | | 00h |
| Device ID | | Vendor ID | | 04h |
| Status | | Code | | 08h |
| Class | | Code | Rev.ID | 0Ch |
| BIST | Head Type | Late Time | Cache Line Size |  |
| **Base Address Reg.** | | | | 10h-24h |
| Card bus CIS Pointer | | | |  |
| Subsystem ID | | Subsystem vendor ID | |  |
| Exp. | | ROM Base Addr. | |  |
| Res | | | | 34-38h |
| Dev | | Spec. Reg | | 40h |
| End | | | | FBh |

0CF8h – 0CFCh

формат регистра определяет координаты устройства на шине PCI.

География PCI шины: номер шины, номер устройства.

[

31(1),

30-24(Reserve),

23-16(Bus. number)

][

15-11(Dev. number),

10-8(Fun. number),

7-2(Reg. name),

2-0(0)

]

Под устройством понимается абонент шины. Физический абонент шины – плата, которая вставляется в коннектор на шине. На одной плате может быть реализовано до 8 устройств. Если на одной карте выполнены 2 устройства, то они воспринимаются как 2 устройства с разными функциями. Деление на устройства и функции чисто логическое. Основное устройство практически всегда соответствует функции ноль. Младшие 2 бита адреса регистра всегда нулевые, т.к. регистры рассматриваются как 4-х битные в конфигурационном пространстве. При обращении к несуществующему устройству из него считывается значение – все единицы (FFFFFFFF). Если считывается какая-то другая информация – то устройство существует.

Рассмотрим процедуру чтения из конфигурационного устройства PCI.

BL – номер функции, BH – номер устройства.

CL – задается функция, CH – регистр.

RD\_PCI PROC NEAR

mov dx, 0CF8h

xor eax, eax

mov al, bl

or ah, 80h

shl eax, 16

mov ah, bh

shl ah, 3

or ah, cl

mov al, ch

and al, 0FCh

out dx, eax

mov dx, 0CFCh

in eax, dx

ret

RD\_PCI ENDP

Процедура записи в регистр конфигурационного пространства.

**Написать программу определения конкретного устройства.**

**Общее описание конфигурационного пространства на PCI**

По спецификации каждое устройство PCI имеет конфигурационное пространство размером 256 байт в котором находится информация о самом устройстве и ресурсы, занимаемые самим устройством. Это пространство не приписано ни к пространству памяти, ни к пространству ввода-вывода. Доступ к нему осуществляется по специальным циклам шины, которые так и называются – чтение/запись конфигурации. После аппаратного сброса (или по включении питания) устройства PCI доступны только для операций конфигурационного чтения и записи. В этих операциях устройства выбираются по индивидуальным сигналам IDSEL и сообщают о потребностях в ресурсах и возможных вариантах конфигурирования. После распределения ресурсов, выполняемого программой конфигурирования (во время теста POST), в конфигурационные регистры устройства записываются параметры конфигурирования. Только после этого к устройству становится возможным доступ по командам обращения к памяти и портам ввода-вывода. Для того чтобы всегда можно было найти работоспособную конфигурацию, все ресурсы, занимаемые картой, должны быть перемещаемыми в своих пространствах. Для многофункциональных устройств каждая функция должна иметь свое конфигурационное пространство.



**Vendor\_ID, Device\_ID, Class\_Code**

Поля Vendor\_ID, Device\_ID и Class\_Code содержат код фирмы-изготовителя устройства, код устройства и код класса устройства. Классификация устройств и указание кода класса в его конфигурационном пространстве является важной частью спецификации PCI.

Код изготовителя, код устройства и код класса применяются в процессе поиска заданного устройства. Если необходимо найти конкретное устройство, то поиск выполняется по кодам устройства и его изготовителя; если необходимо найти все устройства определенного типа, то поиск выполняется по коду класса устройства. После того как устройство найдено, при помощи регистров базовых адресов можно определить выделенные ему области в адресном пространстве памяти и пространстве ввода-вывода (I/O).

**Header\_Type**

Поле Header\_Type определяет формат header-type области, а также является ли устройство многофункциональным. Идентификатором многофункционального устройства является бит 7 поля: если бит установлен в 1 – устройство поддерживает несколько функций, если сброшен в 0 – устройство выполняет только функцию.

Биты 0 – 6 определяют собственно формат header-type области: если эти биты обнулены (содержат код 0x00), то формат области header-type соответствует формату, представленому на рис. 1; значение 0x01 идентифицирует устройство как мост PCI-to-PCI, и формат header-type области описан в спецификации PCI-to-PCI Bridge Architecture Specification; значение 0x02 идентифицирует устройство как мост CardBus. Остальные значения зарезервированы.

**Command**

Командный регистр (поле Command) содержит средства управления устройством. Назначение отдельных бит этого регистра:

n  бит 0 – определяет реакцию устройства на обращение к нему через пространство портов I/O. Если бит сброшен в 0, устройство игнорирует попытки доступа к нему через порты I/O;

n  бит 1 – определяет реакцию устройства на обращение к нему через адресное пространство. Если бит установлен в 1, устройство отвечает на обращения к нему через адресное пространство; если бит сброшен в 0, то устройство на попытки доступа к нему не реагирует;

n  бит 2 – установленный в 1 бит разрешает устройству работать в режиме Bus Master.

**Base Address Registers**

Регистры базовых адресов (Base Address Registers) содержат выделенные устройству области в адресном пространстве и пространстве портов I/O. Бит 0 во всех регистрах базовых адресов определяет, куда будет отображен ресурс – на пространство портов I/O или на адресное пространство. Регистр базового адреса, отображаемый на пространство портов, всегда 32-разрядный, бит 0 установлен в 1. Регистр базового адреса, отображаемый на адресное пространство, может быть 32- и 64-разрядным, бит 0 сброшен в 0.

Поскольку конфигурационное пространство не имеет привязки к какой-либо определенной области адресного пространства, для доступа к нему применяется специальный механизм, названый в спецификации Configuration Mechanism #1. Для работы этого механизма в пространстве портов I/O зарезервированы два 32-разрядных порта, входящих в главный мост: CONFIG\_ADDRESS с адресом 0xCF8 и CONFIG\_DATA с адресом 0xCFC. Формат CONFIG\_ADDRESS представлен на рис. 2.

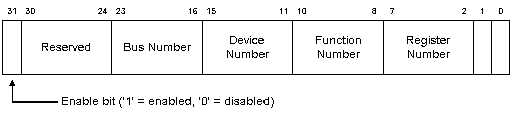


Рисунок 2. Формат регистра CONFIG\_ADDRESS

Установленный в 1 бит 31 разрешает обращение к конфигурационному пространству через порт CONFIG\_DATA, биты 30 – 24 зарезервированы (read-only), при чтении должны возвращать 0, биты 23 – 26 содержат номер шины, биты 15 – 11 – номер устройства, биты 10 – 8 – номер функции и биты 7 – 2 – номер регистра, к которому выполняется обращение (смещение в конфигурационном пространстве).

Порядок работы Configuration Mechanism #1 следующий – в порт CONFIG\_ADDRESS (0xCF8) заносится адрес, соответствующий формату, приведенному на рис. 2; обращением к порту CONFIG\_DATA (0xCFC) производится чтение или запись данных в требуемый регистр конфигурационного пространства.

**PCI BIOS**

Для взаимодействия с устройствами PCI имеются дополнительные функции BIOS, доступные как из реального, так и защищенного режима работы процессора. Эти функции предназначены для работы с конфигурационным пространством и генерации специальных циклов PCI.

Функции PCI BIOS для 16-битного реального режима вызываются через прерывание int 0x1A. Номер функции задается в регистре AX. Признаком нормального выполнения являются значения флага CF = 0 и ноль в регистре AH (AH = 0x00, SUCCESFUL). Если CF = 1, то регистр AH содержит код ошибки:

n  0x81 – неподдерживаемая функция (FUNC\_NOT\_SUP-PORTED);

n  0x83 – неправильный идентификатор производителя (BAD\_VENDOR\_ID);

n  0x86 – устройство не найдено (DEVICE\_NOT\_FOUND);

n  0x87 – неправильный номер регистра PCI (BAD\_REGIS-TER\_NUMBER), т.е. неправильно задано смещение в конфигурационном пространстве.

Перечислим некоторые функции PCI BIOS (полный перечень содержится в [6]):

n  0xB101 – проверка присутствия PCI BIOS;

n  0xB102 – поиск устройства по коду фирмы-изготовителя;

n  0xB103 – поиск устройства по коду класса;

n  0xB108 – чтение байта конфигурационного пространства устройства PCI;

n  0xB109 – чтение слова конфигурационного пространства устройства PCI;

n  0xB10A – чтение двойного слова конфигурационного пространства устройства PCI.

При чтении информации из конфигурационного пространства в регистры процессора заносятся следующие значения:

n  AX – номер функции;

n  BH – номер шины, к которой подключено устройство (от 0 до 255);

n  BL – номер устройства в старших 5 битах и номер функции в трех младших;

n  DI – смещение в конфигурационном пространстве.

После этого следует вызов прерывания int 0x1A, в результате которого в регистрах процессора будут размещены следующие значения:

n  ECX – считанная информация (байт/слово/двойное слово);

n  AH – код возврата (SUCCESFUL/BAD\_REGISTER\_NUM-BER);

n  CF – статус возврата (0 – функция успешно выполнена, 1 – ошибка).

mov AX, 0B103h

mov ECX, 030000h

mov SI, 0

int 1Ah

jc 1\_not, found

mov [PCL\_BUSNO], BH

xor, EAX, EAX

mov AL, BL

mov [FUNCNO], BAX

**ДЗ:**

unsigned int GetSataPort() {

byte interface,

}

**Шина PCI-Express.**

Общая характеристика шин.

При переходе от параллельных шин к последовательным удается не только скомпенсировать падение пропускной способности, но и даже значительно её поднять. Более того, отличная масштабируемость последовательных шин, относительно легко достигается путем как повышения частоты работы, так и добавлением нескольких последовательных линий к шине.

Технология PCI Express является открытым стандартом и разработана с расчетом на разнообразные применения — от полной замены шин PCI и PCI-X внутри настольных и серверных компьютеров, до использования в мобильных, встроенных и коммуникационных устройствах. Номинальной рабочей частотой шины PCI Express является 2,5 ГГц.



Рис. 2. Схема организации данных в архитектуре шины PCI Express.

Спуск еще на уровень по схеме — и мы попадаем в мир драйверов, обслуживающих конкретные устройства PCI Express. Здесь все также без изменений — разработчикам не придется изучать новую шину, совместимость с PCI стопроцентная. То есть получается, что на программном уровне отличия PCI Express от PCI очень малы — обеспечена полноценная совместимость старой и новой шин. Однако все последующие уровни уже относятся к «железной» реализации и здесь происходят кардинальные изменения.

Прежде всего, добавлено два новых уровня (Transaction Layer и Link Layer), которых иначе как TCP и IP не назовешь — выполняемые функции абсолютно те же, что и у «сетевых» аналогов. Transaction Layer получает запросы на чтение и запись от программного уровня и заведует первоначальной упаковкой данных, передачей их конкретному получателю и гарантиями корректной доставки сообщения. Проще говоря если какой-то пакет не дойдет до получателя либо получатель обнаружит в принятом пакете ошибку, то протокол транспортного уровня будет повторять его передачу до тех пор, пока пакет не будет получен — тем самым гарантируется, что передаваемый через PCI Express поток данных достигнет получателя в целостности и сохранности. Каждый пакет снабжен уникальным идентификатором. На этом уровне поддерживается 32-битная и расширенная 64-битная адресация памяти, а также четыре адресных пространства — новое «пространство сообщений» (Message Space) и три уже известных по шине PCI — память, пространство ввода/вывода и конфигурационное пространство. Пространство сообщений предназначено для упрощения формата передачи данных — замены сигналов боковой полосы (side-band) в спецификации PCI 2.2 и исключения «специальных циклов» старого формата (прерывания, запросы управления энергосбережением, сброс).

Link Layer заведует более приземленными делами - здесь указывается уникальный номер пакета (его маршрутизация осуществляется по заголовку, относящемуся к транспортному уровню), по которому контроллеры шины принимают решение о направлении пакета в конкретную физическую линию передачи данных, здесь же располагается код обнаружения и исправления ошибок в принятом пакете (CRC), номер пакета, позволяющий отличить один пакет от другого, и разная вспомогательная информация (например, удостоверяющая, что пакет не был искажен в ходе его передачи). Однако в отличие от TCP/IP, маршрутизация пакетов (принятие решений о том, на которую шину перенаправить пакет, какой из нескольких претендующих пакетов передать первым) осуществляется на транспортном уровне. Интересно также, что пакет передается только в том случае, когда поступил сигнал готовности от буфера приема. Как следствие, уменьшается число повторов пакета и шина используется более эффективно. Формат пакетов шины PCI Express показан на рис. 3 и детализирован во врезке.

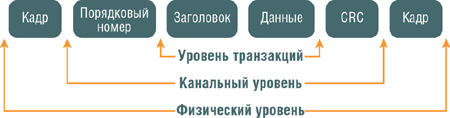
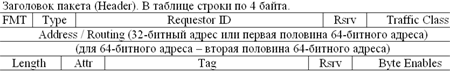


Рис. 3. Формат пакетов шины PCI Express.

Формат пакетов шины PCI-Express

http://www.bestreferat.ru/images/paper/59/97/9729759.gif

Frame — начальный и конечный фрейм пакета - его добавляет физический уровень для определения начала и окончания передачи пакета данных;  
Packet # — номер пакета, добавляется на сетевом уровне чтобы пакеты можно было отличить друг от друга;  
Header — заголовок пакета, описывает тип пакета, получателя, приоритет и другие свойства, это информация транспортного уровня;  
Data — собственно данные пакеты;  
CRC — контрольная сумма пакета.



Fmt — указание типа заголовка (12 или 16 байт) и признак наличия в пакете данных;  
Type — тип пакета (один из четырех основных типов - Memory, I/O, Config, Message и бит, определяющий запрос это или ответ на запрос);  
RequestorID — получатель пакета (шина, устройство, функция устройства);  
Reserved — зарезервированное поле;  
Traffic Class — используется для маршрутизации;  
Address/Routing — адрес в памяти, куда предназначается пакет (32- или 64-разрядный) или иная информация о маршрутизации пакета;  
Length — объем передаваемых в пакете данных;  
Attr — вспомогательные атрибуты пакета (Snoop, Ordering);  
Tag — идентификатор транзакции (Transaction Tag);  
Reserved — зарезервированное поле;  
Byte Enables — вспомогательная информация.

В самом низу этой пирамиды (рис. 2) размещается собственно физическая реализация шины передачи данных — это две независимые дифференциальные пары проводников с импедансом 50 Ом (первая пара работает на прием данных, вторая - на передачу), данные по которым передаются с использованием избыточного кодирования по схеме «8/10» с исправлением ошибок. Если говорить более простым языком, то каждый байт (8 бит) данных, по определенной схеме кодируется 10 битами передаваемых данных (10 бит для передачи этой информации многовато, поэтому его и называют избыточным). Избыточное кодирование позволяет исправлять многие простые ошибки, неизбежные на столь высоких частотах, без привлечения протоколов вышележащих уровней и без лишних повторных передач пакетов. Кроме того, это нужно для того, чтобы уменьшить долю «постоянных» составляющих в сигнале (не более 4 нулей или единиц подряд, см. рис. 4) — обеспечить баланс дифференциальной пары по постоянному току и позволить приемнику уверенно синхронизироваться по фронтам поступающего сигнала, поскольку никакого дополнительного («внешнего») синхронизирующего сигнала от тактового генератора в PCI Express не используется.

Как и в любой сети, передаваемые данные дополнительно нарезаются небольшими кусочками - фреймами. При тактовая частоте шины 2,5 ГГц без учета кодирования мы получим скорость передачи в 2,5 Гбит/с в каждом направлении. С учетом выбранной схемы «8/10» получается 250 Мбайт/с, однако многоуровневая сетевая иерархия не может не сказаться на скорости работы и реальная производительность шины оказывается значительно ниже — всего лишь чуть более 200 Мбайт/св каждую сторону(в пике до 220–230 согласно документации Intel). Впрочем, даже это на 50% больше, чем теоретическая пропускная способность шины PCI. Но это далеко не предел: пожалуй, единственная интересная особенность PCI Express — возможность объединения в одну шину нескольких независимых линий передачи данных. Стандартом предусмотрено использование 1, 2, 4, 8, 16 и 32 линий — передаваемые данные поровну распределяются по ним по схеме «первый байт на первую линию, второй — на вторую, …, n-й байт на n-ю линию, n+1-й снова на первую, n+2 снова на вторую» и так далее (см. рис. 6).

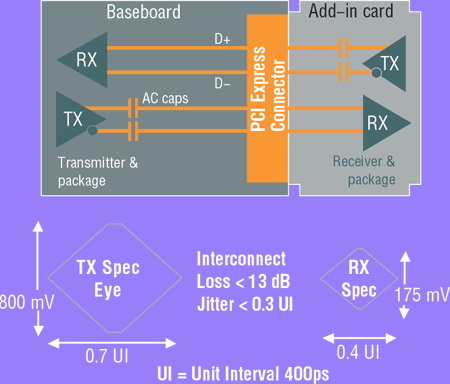


Рис. 5. Схема и электрика шины PCI Express x1.

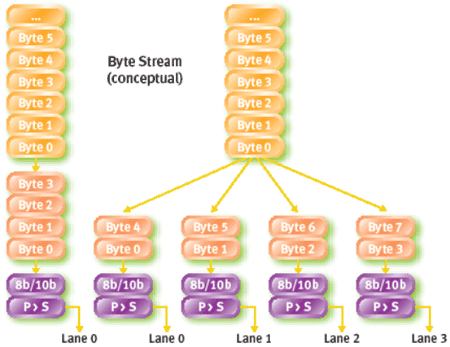


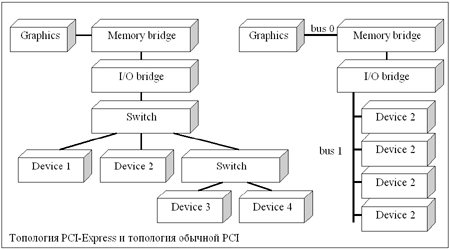
Рис. 6. Асинхронная передача данных по нескольким линиям шины PCI Express.

Это не параллельная передача данных и даже не увеличение разрядности шины (поскольку все передающиеся по линиям данные передаются абсолютно независимо и асинхронно) — это именно объединение нескольких независимых линий. Причем, передача по нескольким линиям никак не влияет на работу остальных слоев «пирамиды» и реализуется сугубо на «нижнем», физическом уровне (рис. 2). Именно этим достигается великолепная масштабируемость шины PCI Express — она позволяет организовывать шины с максимальной пропускной способностью до 32x200=6,4 Гбайт/с — как раз под стать пропускной способности лучших параллельных шин сегодняшнего дня (см. также таблицу 1).

|  |  |  |  |
| --- | --- | --- | --- |
| Таблица 1. Число контактов в разъемах и полоса пропускания шин PCI, PCI-X, AGP и PCI Express | | | |
| Тип слота | Число контактов в разъеме | Полоса пропускания, Мбайт/с | |
| Теоретическая | Эффективная |
| PCI (32 бит 33 МГц) | 120 | 133 | ~110 |
| PCI-X (64 бит 133 МГц) | 184 | 1064 | ~900 |
| PCI Express x1 | 36 | 250\* | ~220\* |
| PCI Express x4 | 64 | 1000\* | ~800\* |
| PCI Express x8 | 98 | 2000\* | ~1600\* |
| PCI Express x16 | 164 | 4000\* | ~3200\* |
| PCI Express x32 | 294 | 8000\* | ~6400\* |
| AGP 8x | 124 | 2133 | ~2000 |
| \* — в каждом направлении независимо. | | | |

**Дополнительные возможности PCI Express**

Стоит упомянуть и о других новых возможностях, появившихся в стандарте PCI Express (по сравнению с PCI) — поддержке виртуальных каналов, QoS (Quality of Service) и изохронной передаче данных. Начнём с рассмотрения механизма, обеспечивающего совместимость PCI Express с обычным PCI. Как мы уже заметили, сегодняшний физический уровень PCI Express обеспечивает лишь соединения «точка-точка», что вынуждает использовать для подключения множества устройств специальные свитчи, объединяя устройства в «звездную» сеть. Но «классическая» PCI — параллельная шина, к тому же использующая механизм прерываний, не поддерживаемый в PCI Express! Сравните: если в случае PCI у нас в ПК было, скажем, две шины — одна для графического адаптера (AGP), и другая — для всех остальных устройств, обращения к которым так и производились по адресу — «шина такая-то, устройство такое-то», то при переходе на PCI Express от былой топологии не остается и следа. В нашем примере (см. рисунок) появляется семь шин PCI Express (не считая шины, соединяющей северный и южный мосты чипсета), причем шесть из них относятся к единственной бывшей PCI (bus 1). К счастью, механизмы маршрутизации, заложенные в стандарт, позволяют особенно не задумываться над этим вопросом — при пересылке пакетов «свитчи» сами определят, на какую шину его необходимо передать. Отправителю достаточно указать устройство-получатель и пакет каким-то образом до него дойдет. То есть ПО теперь работает не непосредственно с аппаратурой, а с непонятно каким образом функционирующими виртуальными каналами данных (старая схема адресации «шина-устройство-функция устройства» при этом сохраняется, хотя такое разделение теперь достаточно условно). Для полной имитации «обычной» PCI-шины контроллер PCI Express даже имитирует прерывания этой шины при поступлении от устройства соответствующего сообщения (служебная информация вроде вызова прерывания также передается в виде пакетов). Впрочем, как уже говорилось, к механизму сообщений есть и прямой доступ, без использования этого режима совместимости.



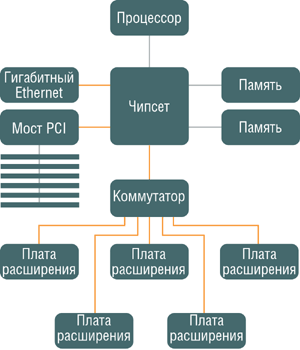
Итак, с устройствами можно продолжать работать, как с обычными PCI, но «виртуальность» этой шины позволяет обеспечить большую гибкость полученной системы. Каждый виртуальный канал до устройства (напомним, что их может быть по нескольку на каждое устройство — для этого и нужна последняя компонента PCI-адреса) никак не привязан к «физическому» носителю, а значит, его можно настроить произвольным образом. Например, стандарт позволяет указывать для виртуального канала его пропускную способность и максимально допустимую задержку передачи данных по нему. Физическая среда передачи данных, конечно, накладывает некоторые ограничения на допустимые здесь значения — больше 200 Мбайт/с через PCI Express 1x при всем желании пропустить невозможно. Виртуальные каналы создаются и изменяются «на лету» — например, плата видеозахвата может большую часть времени обходиться единственным каналом доставки сообщений и запрашивать дополнительный виртуальный канал для передачи данных лишь в момент подключения к ней внешнего устройства. Если у контроллера PCI Express для создания канала не хватит физических ресурсов, то он честно об этом сообщит, но если канал будет создан, то он будет в точности отвечать запрошенным параметрам и никакие «внешние» события — активизация других PCI-устройств, действия пользователя и т.п. на него не повлияют (QoS — запрошенный сервис обладает гарантированным качеством).

Помимо каналов с гарантированной пропускной способностью PCI Express поддерживает также создание изохронных каналов — информация по ним передается с гарантированной максимальной задержкой (это нужно для устройств, работающих в режиме реального времени - например, для устройств, передающих по сети человеческую речь). Впрочем, QoS устройство может отключить (собственно при работе в режиме совместимости с PCI так и происходит), тогда «виртуальным каналам» устройства будут отводиться все остающиеся после QoS-каналов ресурсы шины. «Физическая» реализация QoS и изохронности зависит от конкретной реализации контроллеров PCI Express и использующихся «свитчей», но в конечном итоге все это сводится лишь к тому, какие из пакетов, претендующих на одновременную передачу по одной и той же шине, контроллер пошлет в первую очередь, а какие — лишь по мере возможности. Возможный вариант: изохронные пакеты идут «вне очереди», остальное время пропорционально делится между устройствами, требующими некоторую заданную полосу пропускания и лишь все, что остается распределяется между «обычными», не приоритетными пакетами данных, которые передаются в порядке их поступления в контроллер.

В новой шине также поддерживаются режимы пониженного энергопотребления — в полном соответствии с «четырехуровневыми» стандартами ACPI. Линия PCI Express может «отключаться», если она не используется в данный момент для передачи данных — отключаются линии передачи тактового сигнала, линии приема и передачи данных (и вместе с ними могут отключаться и приемник и передатчик в PCI-Express контроллере), с устройства может быть снято питание — целиком (устройство «логически выключено») или частично (остается маломощное дежурное напряжение питания, функционирует «линия пробуждения» WAKE#, по которой передается сигнал на перевод устройства в нормальный рабочий режим). Если шина состоит из нескольких линий, то при небольшой загрузке шины можно отключать ненужные в данный момент линии (например, использовать PCI Express x4 как x1, а три линии выключить). Переключение в «энергосберегающий» режим при этом может потребовать как само устройство PCI Express, так и «система» в целом — скажем, при переходе в «спящий режим» (hibernate). В «десктопных» вариантах шины PCI Express энергосберегающие режимы являются необязательными (то есть могут быть реализованы, а могут и нет), но в мобильных описанные возможности являются обязательными.

Аппаратные конфигурации

В плане практической реализации шина PCI Express представляет собой целый аппаратный комплекс, затрагивающий северный и южный мосты чипсета, коммутатор и конечные устройства. Новым термином здесь является коммутатор (switch). Он заменяет одну шину с множественными подключениями коммутируемой технологией (рис. 7).



**Система прерываний IBM совместимого компьютера.**

Способы обмена

1. ПИО
2. По прерыванию
3. В режиме прямого доступа к памяти

Основные функции контроллера:

1. фиксация запросов на прерывания от восьми внешних источников;
2. программное маскирование поступающих запросов;
3. присвоение фиксированных или циклически изменяемых приоритетов входам контроллера, на которые поступают запросы;
4. инициация вызова процедуры обработки поступившего аппаратного прерывания.

В состав контроллера входят:

1. схема управления чтением/записью;
2. схема управления;
3. схема каскадирования;
4. регистр запросов на прерывания;
5. схема обработки приоритетов;
6. регистр состояния;
7. регистр маскирования запросов на прерывания.

ПКП может находиться в двух основных состояниях: настройки и обслуживания запросов на прерывания. В состоянии настройки контроллер принимает управляющие слова инициализации (Initialization Command Words, ICW), в состоянии обслуживания - операционные управляющие слова (Operation Control Words, OCW). Возможны несколько режимов обслуживания источников прерываний:

1. - режим фиксированных приоритетов по по уровням прерываний;
2. - два различных варианта циклического сдвига приоритетов;
3. - режим автоматического завершения обработки прерывания;
4. - режим специального маскирования;
5. - режим опроса устройств.

Схема управления чтением/записью (Read/Write Control Logic). Основной функцией этого блока является прием команд от микропроцессора и передача ему информации о состоянии ПКП. Обмен с микропроцессором осуществляется через специальный 8-разрядный буфер данных (Data Bus Buffer), являющийся интерфейсом между ПКП и шиной данных. В состав блока входят регистры управляющих слов ICW и OCW. Схема управляется входами CS, RD, WR и A0. Вход CS (Chip select) отвечает за выбор микросхемы. Низкий уровень сигнала на входе CS разрешает выполнение обмена с ПКП. Низкий уровень сигнала на входе WR (Write) разрешает микропроцессору выводить управляющие слова ICW и OCW для приема их ПКП. Низкий уровень сигнала на входе RD (Read) разрешает ПКП передать микропроцессору информацию о состоянии специальных регистров IRR, ISR и IMR, которые описаны ниже. Все управляющие слова ICW и OCW принимаются контроллером в виде 9-разрядных значений. Разряды 0 - 7 передаются через 8-разрядный буфер данных. Старший разряд (восьмой, считая с нуля) носит название А0 и устанавливается в 0 или 1 в зависимости от того, через какой из двух возможных портов ввода-вывода (четный или нечетный) было передано управляющее слово. Если для вывода значения использовался порт с четным адресом, А0 будет равен 0, если использовался порт с нечетным адресом на единицу большим, чем предыдущий, тогда А0 будет равен 1.  
Регистр запросов на прерывания (Interrupt Request Register, IRR) обслуживается через входы IR0 - IR7 контроллера. Сигнал на  
одном входов IR0 - IR7 - это запрос на прерывание соответствующего уровня (0 - 7). В соответствии с сигналом запроса на прерывание схемой управления устанавливается соответствующий бит в регистре IRR.  
Регистр состояния (регистр обрабатываемых запросов, In-Service Register, ISR) описывает в битах 0 - 7 прерывания каких  
уровней (0 - 7) в данный момент обрабатываются.  
Регистр маскирования запросов на прерывания (Interrupt Mask Register, IMR) описывает, прерывания каких уровней в настоящий  
момент замаскированы. Единичное значение бита в IMR указывает на то, что прерывание соответствующего уровня при появлении запроса в IRR блокируется. Схема обработки приоритетов (шифратор приоритетов, Priority Resolver) определяет, прерывание какого уровня в данный момент является наиболее приоритетным для выполнения. Схема управления ПКП формирует сигнал запроса на прерывaние, поступающий на вход INT (запрос на прерывание) микропроцессора. Если флаг IF регистра флагов процессора равен 1 (прерывания разрешены), процессор отвечает сигналом по линии INTA (подтверждение прерывания), после чего сбрасывается в 0 разряд IRR и устанавливается в 1 разряд ISR, соответствующие уровню обрабатываемого прерывания. После получения второго сигнала подтверждения от процессора по линии INTA, ПКП передает на шину данных 8-битовый номер прерывания. Данная последовательность работы схемы управления выполняется при подключении ПКП к системе с микропроцессорами 8088/8086. При работе с микропроцессорами 8080/8085 последовательность работы схемы управления несколько отличается от описанной выше. Основное отличие состоит в том, что процессору передается не только номер прерывания, но и код команды процессора INT (прерывание) - байт 0CDh.

Схема каскадирования отвечает за работу каскада из нескольких контроллеров. При подключении к ведущему контроллеру выход  
INT каждого ведомого подключается к одному из входов IR0 - IR7 ведущего. Далее этот сигнал передается ведущим на вход INT процессора. Когда процессор возвращает сигнал INTA, ведущий контроллер не только устанавливает бит в ISR и сбрасывает бит в IRR, но и выдает на свои выходы CAS0 - CAS2 номер уровня прерывания, к которому подключен ведомый, пославший запрос на прерывание. Сигналы по линии CAS0 - CAS2 принимаются всеми ведомыми, однако обрабатываются только тем, который подключен к линии IR с соответствующим номером.

**Программирование контроллера прерываний.**

Выводом в порт с четным адресом управляющего слова инициализации ICW1 начинается инициализация ПКП. В процессе инициализации контроллер последовательно принимает управляющие слова ICW1 —ICW4. При наличии в системе одного контроллера ICW3 не выводится. Наличие ICW4 определяется содержанием ICW1. При наличии каскада из нескольких ПКП каждый из них инициализируется отдельно.

Формат ICW1 следующий:

A0 D7 D6 D5 D4 D3 D2 D1 D0

0 - - - “1” LTIM ADI SNGL IC4

│ │ │ │

│ │ │ └─ 1– будет вывод ICW4

│ │ │ 0– не будет

│ │ └────── 1– один контроллер

│ │

│ └─────────── 0– каскад игнорируется

└────────────────── 0–запуск запросов фронтом

1–запуск запросов уровнем

Управляющее слово ICW2 задает номер вектора прерывания для прерываний уровня 0 (например 8 для IBM PC, у которых по уровню «0» происходят прерывания от таймера). Так как вектора аппаратных прерываний располагаются подряд друг за другом, вывод в ICW2 значения 8 не только задает восьмой вектор для таймера, но и девятый для прерываний уровня 1, десятый (0Ah) для прерываний уровня 2 и так далее. Управляющее слово ICW3 выводится только при наличии каскада и имеет разный формат для ведущего и ведомых контроллеров. ICW3 ведущего указывает, к каким входам IR0 —IR7 подключены ведомые контроллеры, при этом соответствующие биты устанавливаются в 1. Остальные биты при этом равны 0.

**Управляющее слово ICW3** выводится только при наличии каскада и имеет разный формат для ведущего и ведомых контроллеров. ICW3 ведущего указывает, к каким входам IR0 - IR7 подключены ведомые контроллеры, при этом соответствующие биты устанавливаются в 1. Остальные биты при этом равны 0. ICW3 следующего вида:

A0 7 6 5 4 3 2 1 0

-----T----T----T----T----T----T----T----T----¬

¦ 1 ¦ 0 ¦ 0 ¦ 0 ¦ 1 ¦ 0 ¦ 0 ¦ 1 ¦ 0 ¦

L----+----+----+----+----+----+----+----+-----

задает, что в каскаде имеется 2 ведомых контроллера, подключенных к входам IR1 и IR4. ICW3 ведомого (подчиненного, slave) ПКП в трех младших

битах задает номер уровня, на котором работает ведомый контроллер. Для ведомого контроллера, работающего на уровне 1 ICW3 будет выглядеть следующим образом:

A0 7 6 5 4 3 2 1 0

-----T----T----T----T----T----T----T----T----¬

¦ 1 ¦ 0 ¦ 0 ¦ 0 ¦ 0 ¦ 0 ¦ 0 ¦ 0 ¦ 1 ¦

L----+----+----+----+----+----+----+----+-----

Если ведомый контроллер работает на уровне 4, то его ICW3 будет таким:

A0 7 6 5 4 3 2 1 0

-----T----T----T----T----T----T----T----T----¬

¦ 1 ¦ 0 ¦ 0 ¦ 0 ¦ 0 ¦ 0 ¦ 1 ¦ 0 ¦ 0 ¦

L----+----+----+----+----+----+----+----+-----

Последнее слово инициализации (ICW4) имеет следующий формат:

A0 D7 D6 D5 D4 D3 D2 D1 D0

┌─╥─┬─┬─┬────┬───┬───┬────┬───┐

1 0 0 0 SFNM BUF M/S AEOI mPM

└─╨─┴─┴─┴────┴───┴───┴────┴───┘

│ │ │ │ │

│ │ │ │ └── 1– 8088/8086

│ │ │ │ 0– 8080/8085

│ │ │ └─ 1–AEOI

│ │ │ 0– EOI

│ │ └───── 1– ведущий

│ │ 0– ведомый (только при BUF=1)

│ │

│ └───────────1– режим буферизации

│

└────────────── 1– специальный вложенный режим

В процессе работы с ПКП можно без переинициализации:

— маскировать и размаскировать аппаратные прерывания;

— изменять приоритеты уровней;

— издавать команду завершения обработки аппаратного

прерывания;

— устанавливать/сбрасывать режим специальной маски;

— переводить контроллер в режим опроса и считывать состояние

регистров РОЗП и РЗП; для этого потребуется вывести в порты ПКП одно

из трех слов команд обслуживания прерывания OCW1 —OCW3.

A0 D7 D6 D5 D4 D3 D2 D1 D0

┌──╥───┬───┬───┬──┬──┬───┬───┬──┐

0 R SL EOI 0 0 L2 L1 L0

└──╨───┴───┴───┴──┴──┴───┴───┴──┘

│ │ │ │ │ └── Номер уровня прерывания

│ │ │ │ └───── в двоичном коде, если он

│ │ │ └──────── требуется в команде.

│ │ └─ 1 – Завершение обработки аппаратного прерывания

│ │

│ │ Используются вместе с EOI=1

┌────┐

0│ 0 Безличный EOI (сбрасывается бит РОЗП с макс. приор. обраб. прерывания)

├────┤

0│ 1 Персонифицированный EOI (сбрасывает бит, опред. кодом L0 —L2 в РОЗП)

├────┤

1│ 0 Циклический сдвиг приоритетов влево на одну позицию

├────┤

1│ 1 Назначение низшего приоритета уровню, определяемому кодом L0 — L2

└────┘

Как уже говорилось, процедура обработки аппаратного прерывания должна перед своим завершением очистить свой бит в ISR вы1водом команды завершения обработки прерывания (End Of Interrupt, EOI). Существует два варианта команды EOI: обычный и специфицированный EOI. Обычный EOI очищает бит в ISR, соответствующий прерыванию с максимальным приоритетом. Специфицированный EOI (R=0, SL=1, EOI=1, L0 - L2 равно номеру уровня прерывания) очищает в ISR бит, соответствующий прерыванию с номером, указанным в L0 - L2 независимо от его приоритета. Команды с битом R=1 позволяют изменить приоритеты уровней. Циклический сдвиг приоритетов сдвигает приоритеты влево на единицу, при этом, если после обычного распределения приоритетов, издать команду циклического сдвига, уровень 0 получит низший приоритет, уровень 1 - наивысший, уровень 2 - следующий за ним и т.д. Команда явного назначения низшего приоритета одному из уровней изменяет приоритеты остальных уровней циклически. Таким образом, если Вы зададите низший приоритет уровню 5, то уровень 6 получит наивысший.

**Третье слово рабочих приказов OCW3** позволяет установить и отменить режим специальной маски, перевести контроллер в режим опроса и прочитать содержимое IRR и ISR. Назначение битов OCW3 приведено на рисунке

А0 7 6 5 4 3 2 1 0

¦ 0 ¦ 0 ¦ESMM¦ SMM¦ 0 ¦ 1 ¦ P ¦ RR ¦ RIS¦

└──────── 1 – режим опроса

¦ 1 ¦ 0 ¦-Отменить режим ¦ 0 ¦ 1 ¦ 0 ¦ Чтение IRR

+---+---+ специальной маски +---+---+---+

¦ 1 ¦ 1 ¦-Установить режим ¦ 0 ¦ 1 ¦ 1 ¦ Чтение ISR

L---+---- специальной маски L---+---+----