**Задание для группы 134 на 08.02.2022**

**Задание пересылать мне на электронную почту: param\_e@mail.ru**

**Задание** также можно найти на goole-диске по ссылке:

**https://drive.google.com/drive/folders/1so8k-nqGSs5huXCXUFHUo-Hus25OJGE7?hl=ru (если задания на сайте нет, то ищите на google-диске по расписанию)**

**Конспект лекции высылаете до 20-00 часов 09.02.2022 (конспект в тетради; высылаете фотографии конспекта со своей подписью на каждом листе).**

 **Без подписи не принимаю конспект.**

Ответы на вопросы в конце лекции, если они есть (ответы в тетради после конспекта).

Если конспект выслан позже указанного времени, то баллы снижаются за каждый час просрочки на 0,5.

**СРС2 «Технологии повышения производительности процессоров» - самостоятельно в Интернете найти и сделать конспект в тетради. Должны быть описаны следующие технологии: Конвейерная обработка, суперскалярная, Матричные и векторные процессоры, технология Hyper-Threading (HT). Срок выполнения задания – 2 недели (до 22.02.22). Буду проверять на занятиях в техникуме.**

**Лекция.** **Функции и характеристики подсистемы памяти.**

ЗУ, именуемые устройствами памяти, предназначены для хранения данных. Они включают в себя процессоры, схемы логики, матрицы памяти, схемы контроля данных, дешифраторы, буферы, регистры, электрические и механические компоненты.

Основные характеристики запоминающих устройств представлены в таблице.

Таблица 1. Функциональные характеристики запоминающих устройств

|  |  |
| --- | --- |
| Характеристика | Набор параметров/Возможное значение |
| Размещение | процессор;внутреннее;внешнее (вторичная память) |
| Емкость | размер слова;количество слов |
| Передаваемая порция | слово;блок |
| Метод доступа | последовательный;прямой;произвольный;ассоциативный |
| Производительность | время доступа; время цикла; скорость передачи |
| Физический тип | полупроводниковые микросхемы; магнитная среда;оптический;магнитооптический |
| Физические характеристики | энергозависимые/энергонезависимые;стираемые/нестираемые |

**Размещение.** Запоминающие устройства по месту размещения в системе делятся на *внутренние* и *внешние.* Очень часто ***внутренняя память*** отождествляется с оперативной памятью, хотя в эту группу входят и другие типы ЗУ. В составе процессора имеется небольшая *локальная* память в виде нескольких *регистров.* Кроме того, устройство управления, которое организационно входит в состав процессора, также имеет собственную внутреннюю память для хранения микропрограмм. Другим типом внутреннего ЗУ является *кэш-память*, которая стала непременным компонентом всех современных процессоров. ***Внешние ЗУ*** - периферийные устройства хранения информации: накопители, для доступа к которым процессору нужно обратиться к соответствующему модулю ввода-вывода.

**Емкость**- одна из основных естественных характеристик любого ЗУ. Общепринятой единицей измерения информационной емкости является 8-битовый байт. Емкость внутренних ЗУ измеряется в словах. Длина слова в современных компьютерах кратна 1 байт - 1, 2 или 4 байт (8, 16, 32 или 64 бит). Емкость внешней памяти всегда оценивается в байтах.

**Передаваемая порция.** При каждом обращении к внутреннему ЗУ передается (считывается из ЗУ или записывается в ЗУ) определенная порция данных. Размер порции соответствует длине слова ЗУ, но это не обязательно.

*Слово* - это «естественная» единица, отражающая организационную струк­туру ЗУ. В большинстве компьютеров размер слова оперативной памяти равен количеству разрядов, используемых для представления целого числа и машинной команды, хотя из этого правила существует масса исключений. (CRAY-1 использует ЗУ с длиной слова 64 бит, и целое число представляется 24 разрядами; семейства VAX при длине слова 32 бит, машинные команды имеют переменную длину, кратную 1 байту)

*Адресуемая единица.* В системах адресуемой единицей является слово, но иногда можно адресовать и отдельный байт, хотя слово имеет длину 2 байт или более. В любом случае между длиной (разрядностью) кода адреса *А* и количеством адресуемых единиц *N* соблюдается соотношение *2A=N.*

*Передаваемая порция.* Для оперативного ЗУ этот параметр характеризует количество битов информации, считываемых или записываемых за один цикл обращения, причем он не обязательно должен равняться длине слова, или адресуемой единицы. При обращении к внешним ЗУ, передаваемая порция имеет значительную длину, намного превышающую длину слова, и принято говорить, что обмен выполняется *блоками.*

**Метод доступа.** Различают следующие методы доступа к хранящейся в ЗУ информации.

*Последовательный доступ* *(sequential access).* Хранимая информация разделена на элементы, называемые *записями (records).* Помимо собственно данных хранится и дополнительная адресная информация, которая служит для разделения записей и используется в процессе поиска нужной записи. Для доступа используется совмещенный переключаемый механизм чтения/записи, который перемещается из текущего положения к нужной записи, проходя все промежуточные. В результате время доступа к нужной записи может изменяться в довольно широких пределах и зависит от исходного положения механизма чтения/записи и положения искомой записи. Последовательный доступ характерен для некоторых конструкций внешних ЗУ (магнитная лента).

*Прямой доступ (direct access).* При прямом используется совмещенный механизм чтения/записи. Но каждая запись имеет свой уникальный адрес, соответствующий ее физическому положению на носителе. Доступ осуществляется прямым обращением к зоне носителя, в которой находится адресуемая запись, и далее последовательно просматриваются записи внутри этой зоны, пока механизм чтения/записи не будет совмещен с искомой записью. Время доступа к произвольной записи переменно, хотя разброс значений и не так значителен, как при последо­вательном доступе. Прямой доступ реализован в конструкциях внешних ЗУ, которые используют в качестве носителя информации диски с концен­трическими дорожками (жесткие).

*Произвольный доступ* *(random access).* Каждая адресуемая единица имеет, свой встроенный механизм адресации. Время доступа к любой единице (ячейке) не зависит от пред истории и от адреса этой ячейки. Таким обра­зом, возможен непосредственный доступ к ячейкам в произвольном порядке с минимальными потерями времени. Именно такой способ доступа исполь­зуется в оперативных ЗУ и в некоторых конструкциях кэш-памяти.

*Ассоциативный доступ (associative access*). Существуют ЗУ с произ­вольным доступом, которые оснащены встроенным механизмом сравнения определенных битов в каждой ячейке с заданным образцом. Сравне­ние выполняется по всем ячейкам одновременно. Нужная ячейка отыскивается по ее содержимому. Время доступа не зависит от адреса ячейки или пред истории. Такой метод доступа распространен в конструкциях блоков кэш-памяти.

**Производительность ЗУ.** С точки зрения конструктора вычислительной системы наиболее важными ха­рактеристиками ЗУ являются информационная емкость и *производительность.*

Для оценки производительности используется три параметра.

*Время доступа или обращения (access time).* Для ЗУ с *произвольным* доступом этот пара­метр характеризует время выполнения операций чтения или записи, и из­меряется как длительность временного интервала между моментом, когда в устройство передан адрес, и моментом, когда данные зафиксированы в памяти (при выполнении записи) или когда ими может воспользоваться дру­гое устройство компьютера (при выполнении чтения). Для ЗУ с *последова­тельным* или *прямым* доступом этот параметр характеризует время, необ­ходимое для перевода механизма чтения/записи в нужную позицию по отношению к носителю информации.

*Длительность цикла обращения к памяти* *(memory cycle time).* Этот па­раметр характеризует ЗУ с произвольным доступом. Это мини­мальная длительность времени между последовательными сеансами доступа к памяти. Время цикла включает время доступа + время выполнения дополнительных операций, связанных с подготовкой устройства к следующему обращению. Эти операции зависят от типа конструкции устройства - в одних требуется не­которое время для сброса сигналов на линиях, а в других может понадо­биться восстановление информации в запоминающем элементе после ее считывания (так называемые ЗУ с разрушением информации при считывании).

*Скорость передачи (transfer rate).* Параметр характеризует интенсивность
информационного потока между ЗУ и теми устройствами, которые нуждаются в хранящейся в нем информации. Для ЗУ с произвольным доступом скорость передачи - величина, обратная *длительности цикла обращения.* Для ЗУ с последовательным или прямым доступом справедливо соотношение

$$T\_{N}=T\_{A}+\frac{N}{R}$$

*N* - объем блока данных (бит);

*R* - скорость передачи (бит/с);

*TN* - среднее время чтения или записи блока данных объемом и *N* бит;

*ТА* - среднее время доступа для данного ЗУ.

**Физические типы ЗУ.** За время истории электронной вычислительной техники были созданы ЗУ на основе самых разных физических явлений. В настоящее время из всего разнообразия конструкций на практике применяются четыре типа ЗУ (точнее, четыре типа физических носителей информации):

*полупроводниковые,* в которых хранение информации производится либо в электронных элементах с двумя устойчивыми состояниями, либо в запоминающих полупроводниковых емкостях;

*магнитные,* где для хранения используется эффект перемагничивания магнитного материала, нанесенного на какую-либо поверхность (ленту, диск, барабан и т.п.);

*оптические*;

*магнитооптические*.

Конструктора системы интересует характеристика физического типа носителей -возможность сохранения информации при отключении питания (этот параметр получил назван «энергонезависимость» ЗУ). Все магнитные, оптические и магнитооптические ЗУ - энергонезависимы. Большинство полупроводниковых ЗУ требует наличия электропитания для сохранения информации. Существуют специальные конструкции, сохраняющие информацию при отключении питания (постоянные ЗУ или - *ROM-память (read only memory*). Информация в ROM - память записывается на стадии изготовления компонента или включения ее в компьютерную систему, а затем может только считываться.

**Иерархия памяти в компьютерной системе.**

Повышение производительности вычислительных систем непо­средственно связано с увеличением быстродействия и емкости па­мяти. Емкость памяти наиболее крупных вычислительных систем возросла от 1000 байт до десятков терабайт, а время цикла умень­шилось с 20 мкс до 10 нс. Однако даже с учетом прогресса в техно­логии быстродействующие запоминающие устройства остаются бо­лее дорогими, чем медленные.

Чтобы система обладала максимальной производительностью возможной при данном типе процессора, быстродействие памяти должно быть таким, чтобы обеспечить его полную загрузку, т.е. процессор не должен простаивать ожидая, пока из памяти будут извлечены необходимые для его работы данные. Стоимость подсистемы памяти должна быть сопоставима со стоимостью остальных компонентов системы.

Все параметры подсистемы взаимосвязаны. Характерно такое отношение между параметрами:

чем выше быстродействие, тем выше относительная стоимость ЗУ (в пересчете на бит хранимой информации);

чем больше объем памяти, тем ниже относительная стоимость ЗУ;

чем больше объем памяти, тем ниже быстродействие.

Следовательно, с целью уменьшения стоимости ВС, при той же производительности, эффективнее иметь иерархию памяти с небольшим по емкости запоминающим устрой­ством, расположенным рядом с процессором и имеющим минимальное время доступа. Такая иерархия позволяет согласовать ха­рактеристики памяти и центрального процессора. Желательно включить в состав системы ЗУ большего объема, поскольку при этом снижается относительная стоимость хранения, и обеспечить необходимую производительность.

Эта задача решается созданием *иерархии памяти,* когда в состав системы включают несколько ЗУ, в которых используется разная технология хранения информации, (рис. Пирамида). По мере перехода от верхних уровней иерархии к ниж­ним наблюдается следующее изменение характеристик ЗУ:

- снижается относительная стоимость хранения информации;

- повышается емкость отдельного модуля;

- увеличивается время доступа;

- снижается частота обращения к памяти со стороны процессора.

Таким образом, менее емкие, но более скоростные ЗУ дополняются менее быстродействующими устройствами, но зато обладающими большей емкостью. Организуется поток информации в системе, обеспечивающий по мере перехода от устройств верхних уровней иерархии к устройствам нижних уровней, снижающий интенсивность обращений к ЗУ со стороны процессора.

**Пример.** Процессор в системе может обращаться к памяти двух уровней. Память 1-го уровня содержит 1000 слов и имеет время доступа 0,1 мкс, а память 2-го уровня - 100000 слов и время доступа 1 мкс. Предположим, что если нужное слово находится в памяти 1-го уровня, то процессор извлекает его непосредственно, а если в памяти 2-го уровня, то затребованное слово сначала записывается в память 1-го уровня, а уже оттуда его извлекает процессор. (Не будем учитывать время, необходимое процессору для того, чтобы выяснить, где именно находится искомое слово, - в памяти 1-го или 2-го уровня.) На рисунке с право показан вид типовой зависимости между средним временем доступа и распределением запросов между устройствами двух уровней. Параметр *H* характеризует долю запросов к памяти 1-го уровня (т.е. к кэш-памяти) во всем потоке запросов, параметр *T1* - время доступа к памяти 1-го уровня, а *Т2* - время доступа к памяти 2-го уровня. Чем выше значение *H*, т.е. чем большая часть всех обращений процессора за данными приходится на кэш-память, тем меньше среднее время обращения к интегрированной подсистеме памяти, тем ближе оно к значению *Т1*.

Предположим, что 95% всех обращений приходится на кэш-память. Тогда среднее время извлечения одного слова из подсистемы памяти определится следующим образом:

(0,95)\*(0,1 мкс) + (0,05)\*(0,1 мкс + 1 мкс) = 0,095+0,055 = 0,15 мкс.

Концепция иерархической организации памяти принесет успех лишь в том случае, если при конструировании системы будут приняты во внимание количественные характеристики ЗУ разных уровней.

Среди существующих на сегодняшний день типов ЗУ, созданных на основе разных технологий, не так уж сложно подобрать ряд, обладающий первыми тремя свойствами. Что касается четвертого свойства, снижения частоты обращения со стороны процессора к памяти нижнего уровня, то конструктор может реализовать его в системе, не прибегая к большим ухищрениям. Требуется лишь при структурном синтезе системы учитывать свойство *локализации ссылок*. В процессе выполнения фрагментов программы ссылки на ячейки памяти, в которых хранятся команды и данные, имеют тенденцию “скапливаться” в определенных подпространствах адресов - *кластерах.* В программе имеется множество итерационных процедур и подпрограмм. После входа в такую итерационную процедуру или подпрограмму начинает выполняться повторяющаяся последовательность команд, компактно размещенных в памяти. Обработка массивов или таблиц сопряжена с обращением к ячейкам памяти, компактно размещенным в общем адресуемом пространстве. Переход от одного кластера к другому происходит относительно редко, т.е. реже, чем обращение к ячейкам из одного и того же кластера.

Можно так распределить информацию между ЗУ разного уровня иерархии, что доля обращений к памяти нижних уремией иерархии будет меньше, чем обращений к памяти верхних уровней.

**Пример с двухуровневой памятью**. Предположим, что все команды и данные программы хранятся в ЗУ 2-го уровня. Временно скопируем текущие кластеры в ЗУ 1-го уровня. Время от времени один из скопированных (и, возможно, изме­ненных в процессе выполнения программы) кластеров нужно будет возвращать обратно в ЗУ 2-го уровня, а в образовавшееся "окно” копировать другой кластер. Нo, среднестатистически, большинство обращений в ходе выполнения программы придется на те кластеры, которые уже присутствуют в ЗУ 1-го уровня.

Этот же принцип можно применить и в системе, имеющей три или больше ЗУ разных уровней. Самой быстрой, но и самой маленькой по объему, а также самой дорогой (по отношению к объему хранимой информации) будет па­мять, состоящая из внутренних регистров процессора. Количество таких регистров ограничивается несколькими десятками, хотя существуют процессоры, включающие сотни регистров. Двумя уровнями ниже находится оперативная (основная) память компьютерной системы. Каждая ячейка опе­ративной памяти имеет уникальный атрибут - *адрес,* причем в машинных ко­мандах для идентификации подавляющего большинства обрабатываемых данных используется именно адрес соответствующего элемента в оперативной памяти. Выше оперативной памяти в иерархии размещается кэш-память, которая имеет значительно меньшую емкость, чем оперативная (и значительно большую, чем набор внутренних регистров процессора), но обладает на порядок большим быстродействием. Кэш-память обычно скрыта от програм­миста, т.е. он никак не управляет из программы размещением данных в кэш­ памяти или обращением к ней. Это промежуточное ЗУ “держит наготове” дан­ные и команды программы, которые, скорее всего, могут в ближайшее время понадобиться процессору, и таким образом сглаживает поток информации ме­**жду** регистрами процессора и оперативной памятью.

ЗУ, выполняющие в вычислительной системе роль памяти первых уровней иерархии, как правило, по своей конструкции являются полупроводниковыми энергозависимыми устройствами. Но каждое из них обычно изготавливается по своей технологии, которая обеспечивает оптимальное соотношение между емко­стью, скоростью и стоимостью для памяти данного уровня. Для более длитель­ного хранения информации используются внешние устройства памяти большого объема (в последнее время по отношению к таким устройствам очень часто ис­пользуется термин *массовая память* - *mass storage) -* чаще всего жесткие не­съемные магнитные диски, которые дополняются устройствами со съемными но­сителями - магнитными, оптическими и магнитооптическими. Во внешних ЗУ хранятся файлы выполняемых программ и обрабатываемых данных. Обращаются к этой ин­формации в терминах файлов или отдельных записей, а не в терминах отдель­ных байтов или слов. Диски часто используются и для организации виртуальной памяти в качестве расширения полупроводнико­вых оперативных ЗУ.

В больших машинах фирмы IBM имеется внутренняя память, которая в документации названа *расширенной памятью* (expanded storage). Для нее используются модули полупроводниковой памяти.

Иерархические уровни могут быть организованы программно. Часть основной внутренней памяти может быть использована операционной системой в качестве буфера при обмене данными между полупроводниковой оперативной и внешней дисковой памятью. Такая методика имеет специальный термин “дисковый кэш” (disk cache). Она способствуем повышению производительности системы. (1. Записи на дисках имеют четко выраженную кластерную структуру. Буферизация позволяет передавать данные большими порциями, примерно равными кластеру. 2. Некоторые данные могут неоднократно запрашиваться программой. Поэтому желательно как можно дольше хранить их в быстрой полупроводниковой памяти, а не считывать каждый раз с диска.)

Иерархия памяти существует главным образом для того, чтобы повысить экономическую эффективность системы путем оптималь­ного сочетания временных и стоимостных характеристик различных запоминающих устройств. Для того чтобы иерархия памяти была экономически эффективной, память должна иметь высокие эксплуатационные характеристики. Это означает, что технические па­раметры запоминающих устройств на разных уровнях иерархии должны существенно различаться. Невозможно достигнуть больших различий по эксплуатационным характеристикам и стоимости ма­лой и большой памяти, если и та и другая выполнены по одной тех­нологии и имеют близкие по значению времена доступа.

Полупостоянная память  в основном используется для хранения информации о конфигурации компьютера. Традиционная память конфигурации вместе с часами - календарем (CMOS Memory и CMOS RTS) имеет объем несколько десятков байт, ESCD (Extended Static Configuration Data) — область энергонезависимой памяти, используемая для конфигурирования устройств Plug and Play— имеет объем несколько килобайт. Сохранность данных CMOS при отключении питания компьютера обеспечивается маломощной внутренней батарейкой или аккумулятором. В качестве полупостоянной применяется и энергонезависимая память — NV RAM (Non Volatile RAM), которая хранит информацию и при отсутствии питания.

Буферная память различных адаптеров и контроллеров (коммуникационных, дисковых и пр.) обычно является разделяемой между процессором (точнее, абонентами системной шины) и контроллерами устройств. К этой памяти относятся и 16-байтные FIFO - буферы СОМ-портов, и 16-мегабайтные (и более) кэш-буферы высокопроизводительных SCSI-адаптеров. Специфическим типом буферной памяти является видеопамять дисплейного адаптера — к ней производятся интенсивные обращения со стороны центрального процессора и графического акселератора одновременно с непрерывным процессом регенерации изображения.