

ОПЕРАТИВНАЯ ПАМЯТЬ



Оперативная память, или RAM-память (Random Access Memory),— это память с произвольным доступом. Используется центральным процессором для совместного хранения данных и исполняемого программного кода. Отличительной особенностью RAM является ее быстродействие, которое очень важно для современных процессоров.

Элементарная ячейка оперативной памяти представляет собой конденсатор, способный в течение короткого промежутка времени сохранять электрический заряд, наличие которого можно ассоциировать с информационным битом. Проще говоря, при записи логической единицы в ячейку памяти конденсатор заряжается, при записи нуля — разряжается.



ЭВОЛЮЦИЯ РАЗВИТИЯ

- ✓ Синхронная память SDRAM появилась в середине 1990-х годов. Именно тогда в массовое производство была внедрена синхронная память SDR SDRAM.
- ✓ В 2000 году на смену памяти SDR SDRAM пришла память DDR SDRAM, отличающаяся большей пропускной способностью при меньшем энергопотреблении.
- ✓ Вслед за памятью DDR была внедрена память DDR2, которая подразумевает в четыре раза большую скорость передачи.
- ✓ Летом 2007 года был принят стандарт DDR3.
- ✓ В январе 2011 года компания SAMSUNG официально представила новые модули, работающие в режиме DDR 4 - 2133 МГц, при напряжении в 1,2 В.

ЛОГИЧЕСКАЯ ОРГАНИЗАЦИЯ ПАМЯТИ

Микросхемы памяти организованы в виде матрицы, напоминающей лист бумаги в клетку, причем пересечение столбца и строки матрицы создает одну из элементарных ячеек. Понятно, что количество строк и столбцов в матрице памяти определяет ее объем в битах или байтах.

Кроме того, современные чипы памяти имеют несколько банков, каждый из которых можно рассматривать как отдельную матрицу со своими столбцами и строками.

При обращении к той или иной ячейке памяти следует задать адрес нужных строки и столбца.

С учетом того, что в модуле памяти используется несколько чипов памяти, а в каждом чипе — несколько банков памяти, прежде всего необходимо указать, в каком чипе и банке находится ячейка. Для этого применяются специальные сигналы: CS, BA0 и BA1.

Сигнал CS (Chip Select) позволяет выбрать требуемый чип памяти. Когда сигнал активен, возможен доступ к чипу памяти, то есть чип активируется.

Когда выбраны чип и банк памяти, можно получить доступ к требуемой ячейке памяти, задав адрес столбца и строки. То есть первоначально передается адрес строки, а затем, через некоторый промежуток времени, по той же шине передается адрес столбца.

Для считывания адреса строки на входы матрицы памяти подается специальный стробирующий импульс RAS (Row Address Strobe). Этот импульс представляет собой изменение уровня сигнала с высокого на низкий, то есть при переходе сигнала RAS с высокого уровня на низкий возможно считывание адреса строки. Аналогичным образом считывание адреса столбца происходит при изменении уровня сигнала (стробирующего импульса) CAS (Column Address Strobe) с высокого значения на низкое и синхронно с положительным фронтом тактирующего импульса.

Импульсы RAS# и CAS# подаются последовательно друг за другом, причем импульс CAS# всегда следует за импульсом RAS#, то есть сначала происходит выбор строки, а затем — выбор столбца.

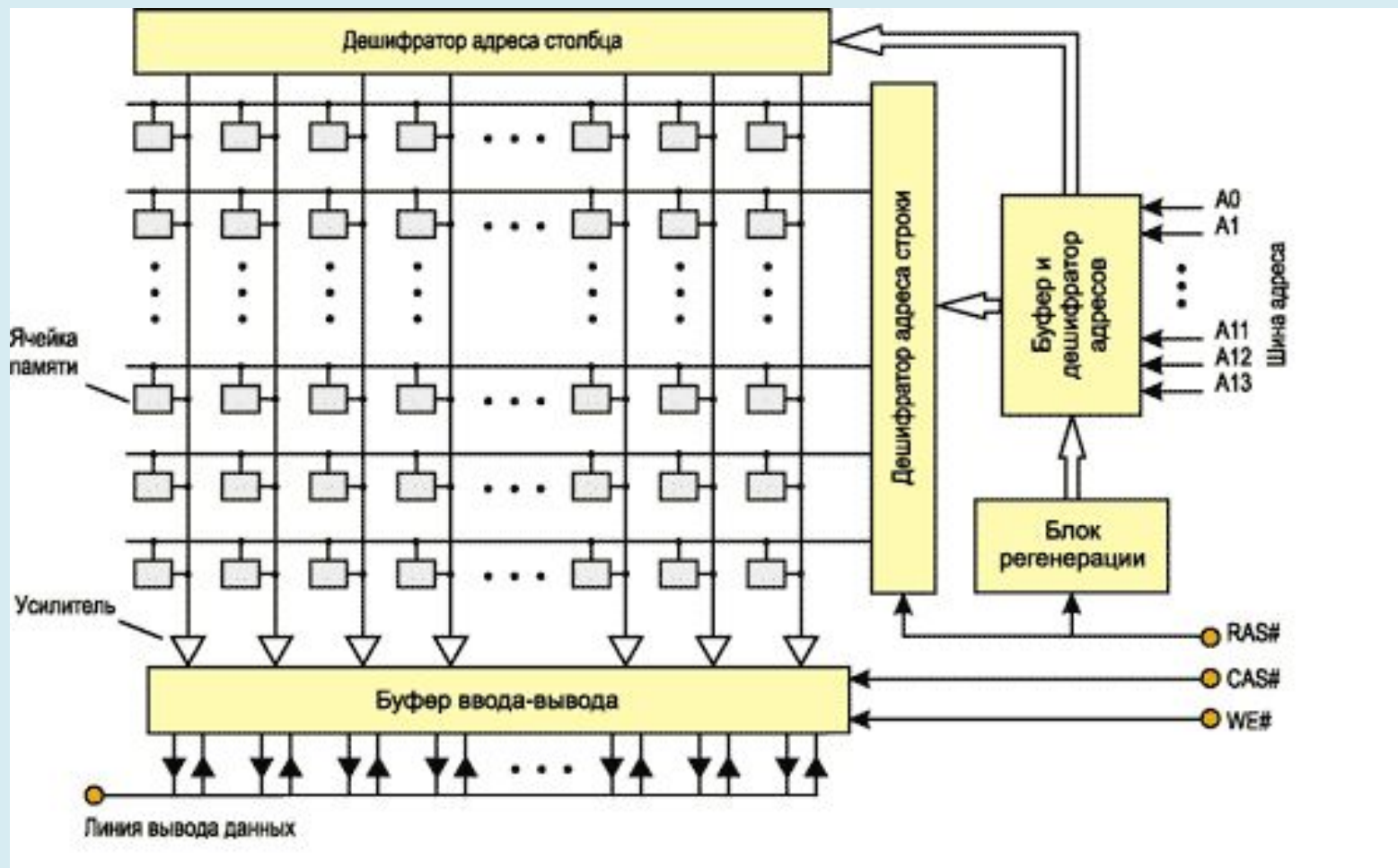


Рисунок 1 - Структура матрицы памяти

ПРИНЦИПЫ ФУНКЦИОНИРОВАНИЯ ПАМЯТИ

В современных компьютерах используются запоминающие устройства основных типов:

—*DRAM (Dynamic Random Access Memory)* - динамическое запоминающее устройство с произвольным порядком выборки.

—*SRAM (Static RAM)* - статическая оперативная память.

В настоящее время динамическая память DRAM используется в качестве оперативной памяти компьютера, а статическая память — для создания высокоскоростной кэш памяти процессора.

Различие между этими типами памяти заключается в принципе хранения информации.

ДИНАМИЧЕСКАЯ ПАМЯТЬ

В динамической памяти элементарная ячейка представляет собой конденсатор, который способен в течение некоторого промежутка времени сохранять электрический заряд, наличие которого можно ассоциировать с информационным битом.

При считывании данных каждый конденсатор разряжается, поэтому его необходимо зарядить до прежнего значения. Процесс считывания (обращения к ячейке) сочетается с подзарядкой конденсаторов, то есть с регенерацией заряда.

К тому же, если обращения к ячейке не происходит в течение длительного времени, то со временем за счет токов утечки конденсатор разряжается (неизбежный физический процесс) — и информация теряется. Вследствие этого память на основе массива конденсаторов требует постоянного периодического подзаряда конденсаторов.

К достоинствам динамической памяти относятся высокая удельная плотность размещения данных и низкое энергопотребление, а к недостаткам — низкое быстродействие по сравнению со статической памятью.

СТАТИЧЕСКАЯ ПАМЯТЬ

В статической памяти ячейки построены на различных вариантах триггеров – транзисторных схем с двумя устойчивыми состояниями. После записи бита в такую ячейку она может пребывать в одном из таких состояний и сохранять записанный бит сколь угодно долго – необходимо только наличие питания. Отсюда и название памяти – статическая, то есть пребывающая в неизменном состоянии. Достоинством статической памяти является ее быстродействие, а недостатком – высокое энергопотребление и низкая удельная плотность данных, поскольку одна триггерная ячейка состоит из нескольких транзисторов и, следовательно, занимает немало места на кристалле.

ПАМЯТЬ ТИПА DRAM

Динамическая оперативная память (Dynamic RAM — DRAM) используется в большинстве систем оперативной памяти современных персональных компьютеров.

Основное преимущество памяти этого типа состоит в том, что ее ячейки упакованы очень плотно, т.е. в небольшую микросхему можно упаковать много битов, а значит, на их основе можно построить память большой емкости.

По логике организации DRAM-память может быть *асинхронной* и *синхронной*.

Недостатком DRAM является низкое быстродействие. Поэтому существует множество различных типов организации DRAM, позволяющих улучшить эту характеристику.

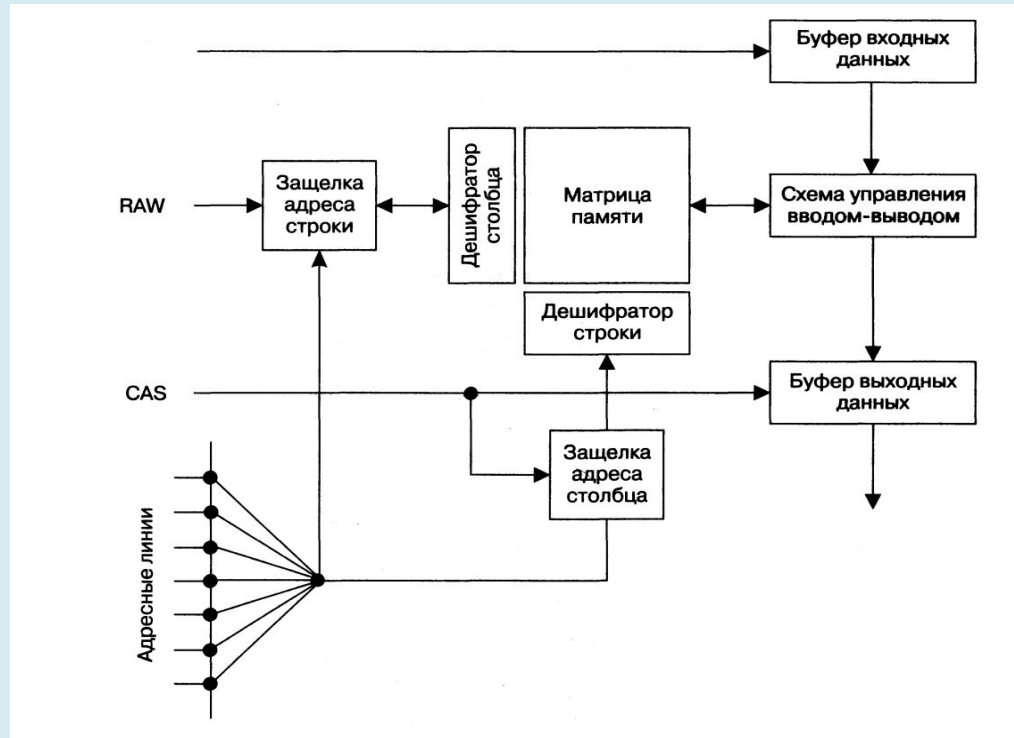


Рисунок 2 - Структурная схема динамической памяти

АСИНХРОННАЯ DRAM-ПАМЯТЬ

Работа асинхронной памяти не синхронизирована с тактовой частотой системной шины, то есть данные появляются на этой шине в произвольные моменты времени. С системной шины они считываются контроллером, который уже синхронизирован с тактовой частотой. Следовательно, если данные появляются в ближайший момент за фронтом тактового импульса, то они будут считаны только с началом следующего тактового импульса. Так происходит потому, что сигналом к считыванию данных является не уровень тактового импульса (высокий или низкий), а его фронт.

Поэтому при использовании асинхронной памяти часто возникает задержка с обработкой данных, что, в свою очередь, приводит к жестким ограничениям на частоту системной шины.

В персональных компьютерах использовалось несколько типов асинхронной динамической памяти: FRM (Fast Page Mode - память с произвольным доступом, поддерживающая быстрый страничный режим), EDO (Extended Data Out - динамическая память произвольного доступа с расширенным выводом данных), BEDO (Burst EDO - реализацией блочного, или пакетного, доступа).

СИНХРОННАЯ SDRAM ПАМЯТЬ

При синхронной работе с памятью SDRAM обеспечивается синхронизация всех входных и выходных сигналов с тактами системного генератора. Однако управление памятью усложняется, так как приходится вводить дополнительные регистры-защелки, которые хранят адреса, данные и управляющие сигналы, в то время как процессор, передав их в память, продолжает работать с другими устройствами. После определенного числа тактовых циклов, количество которых считает специальный счетчик, данные становятся доступными, и процессор может получить их с системной шины.

Помимо организации синхронного доступа к данным, память SDRAM имеет еще ряд принципиальных отличий от асинхронной памяти.

Весь массив памяти SDRAM модуля разделен на два независимых банка. Такое решение позволяет совмещать выборку данных из одного банка с установкой адреса в другом банке, то есть иметь одновременно две открытые страницы.

В SDRAM-памяти, как и в BEDO-памяти, организована конвейерная обработка данных, что позволяет производить обращение по новому адресу столбца ячейки памяти на каждом тактовом цикле. В микросхеме SDRAM имеется счетчик для наращивания адресов столбцов ячеек памяти с целью обеспечения быстрого доступа к ним.

ПАМЯТЬ DDR SDRAM

Аббревиатура DDR (Double Data Rate) в названии памяти означает удвоенную скорость передачи данных.

В обычной SDRAM-памяти ядро и буферы обмена работают в синхронном режиме на одной и той же частоте. Передача каждого бита из буфера происходит с каждым тактом работы ядра памяти. В DDR-памяти каждый буфер ввода-вывода передает два бита за один такт, то есть фактически работает на удвоенной тактовой частоте, оставаясь при этом полностью синхронизированным с ядром памяти. Однако чтобы такой режим работы стал возможным, необходимо, чтобы эти два бита были доступны буферу ввода-вывода на каждом такте работы памяти.

Для этого требуется, чтобы каждая команда чтения приводила к передаче из ядра памяти в буфер сразу двух бит. Для этого используются две независимые линии передачи от ядра памяти к буферам ввода-вывода шириной n бит каждая, откуда биты поступают на шину данных в требуемом порядке.

Поскольку при таком способе организации работы памяти происходит предвыборка $2n$ бит перед передачей их на шину данных, этот способ также называется $2n$ Prefetch (предвыборка $2n$ бит).

Все же остальные принципиальные моменты DDR-памяти не изменились: структура нескольких независимых банков позволяет совмещать выборку данных из одного банка с установкой адреса в другом банке.

Доступ к этим страницам чередуется (bank interleaving), что приводит к устранению задержек и обеспечивает создание непрерывного потока данных.

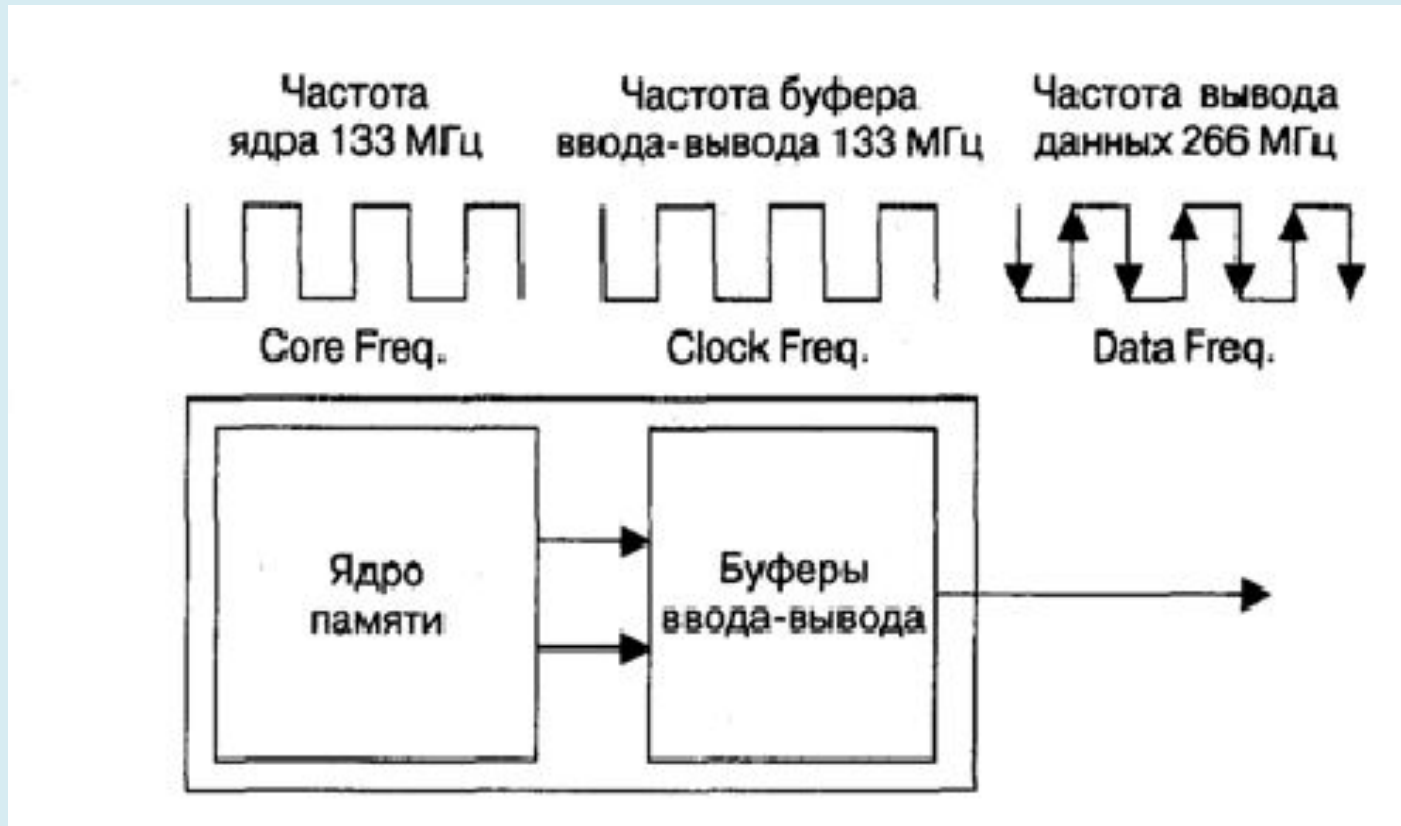


Рисунок 3 -Принцип функционирования памяти DDR

ПАМЯТЬ DDR2 SDRAM

DDR2 SDRAM (double-data-rate two synchronous dynamic random access memory — синхронная динамическая память с произвольным доступом и удвоенной скоростью передачи данных, второе поколение). Пришла на смену памяти DDR SDRAM.

Этот стандарт подразумевает в четыре раза большую скорость передачи. То есть в стандарте DDR2 при пакетном режиме доступа данные передаются четыре раза за один такт. Для организации данного режима работы памяти необходимо, чтобы буфер ввода-вывода работал на учетверенной частоте по сравнению с частотой ядра памяти.

Достигается это следующим образом: ядро памяти, как и прежде, синхронизируется по положительному фронту тактирующих импульсов и с приходом каждого такого положительного фронта по четырем независимым линиям передает в буфер ввода-вывода $4n$ бита информации (выборка $4n$ битов за такт, $4n$ -Prefetch). Сам буфер ввода-вывода тактируется на удвоенной частоте ядра памяти и синхронизируется как по положительному, так и по отрицательному фронту этой частоты. Это позволяет за каждый такт работы ядра памяти передавать четыре бита на шину данных, то есть вчетверо повысить пропускную способность памяти.

По сравнению с DDR, память DDR2 позволяет обеспечить ту же пропускную способность, но при вдвое меньшей частоте ядра.

Преимущества DDR2 по сравнению с DDR:

- Более высокая полоса пропускания
- Как правило, меньшее энергопотребление
- Улучшенная конструкция, способствующая охлаждению

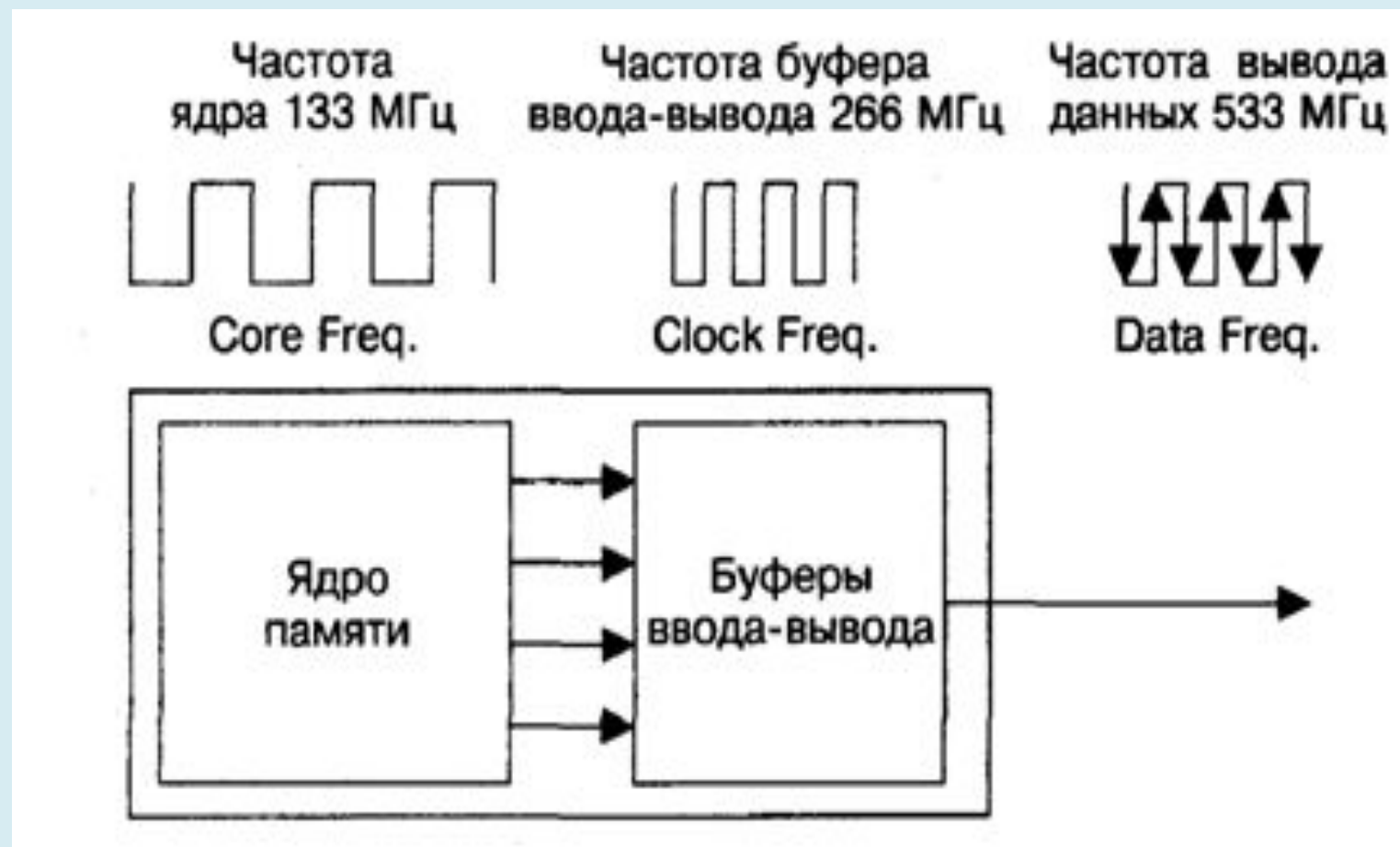


Рисунок 4-Принцип функционирования памяти DDR2

ПАМЯТЬ DDR3 SDRAM

Для памяти DDR3 реализована 8-банковая логическая структура. Принципиальное отличие памяти DDR3 от памяти DDR2 заключается в реализации механизма 8n-Prefetch вместо 4n-Prefetch. Для организации данного режима работы памяти необходимо, чтобы буфер ввода-вывода (мультиплексор) работал на частоте в 8 раз большей по сравнению с частотой ядра памяти. Достигается это следующим образом: ядро памяти, как и прежде, синхронизируется по положительному фронту тактирующих импульсов, а с приходом каждого положительного фронта по восьми независимым линиям в буфер ввода-вывода (мультиплексор) передаются 8n бита информации (выборка 8n битов за такт).

Сам буфер ввода-вывода тактируется на учетверенной частоте ядра памяти и синхронизируется как по положительному, так и по отрицательному фронту данной частоты. Это позволяет за каждый такт работы ядра памяти передавать восемь слов на шину данных, то есть в восемь раз повысить пропускную способность памяти.

Конечно, реализация механизма 8n-Prefetch вместо 4n-Prefetch — это не единственное различие между памятью DDR3 и DDR2. Другими нововведениями, реализованными в памяти DDR3, являются технология динамического терминирования сигналов (dynamic On-Die Termination, ODT) и новая технология калибровки сигналов.

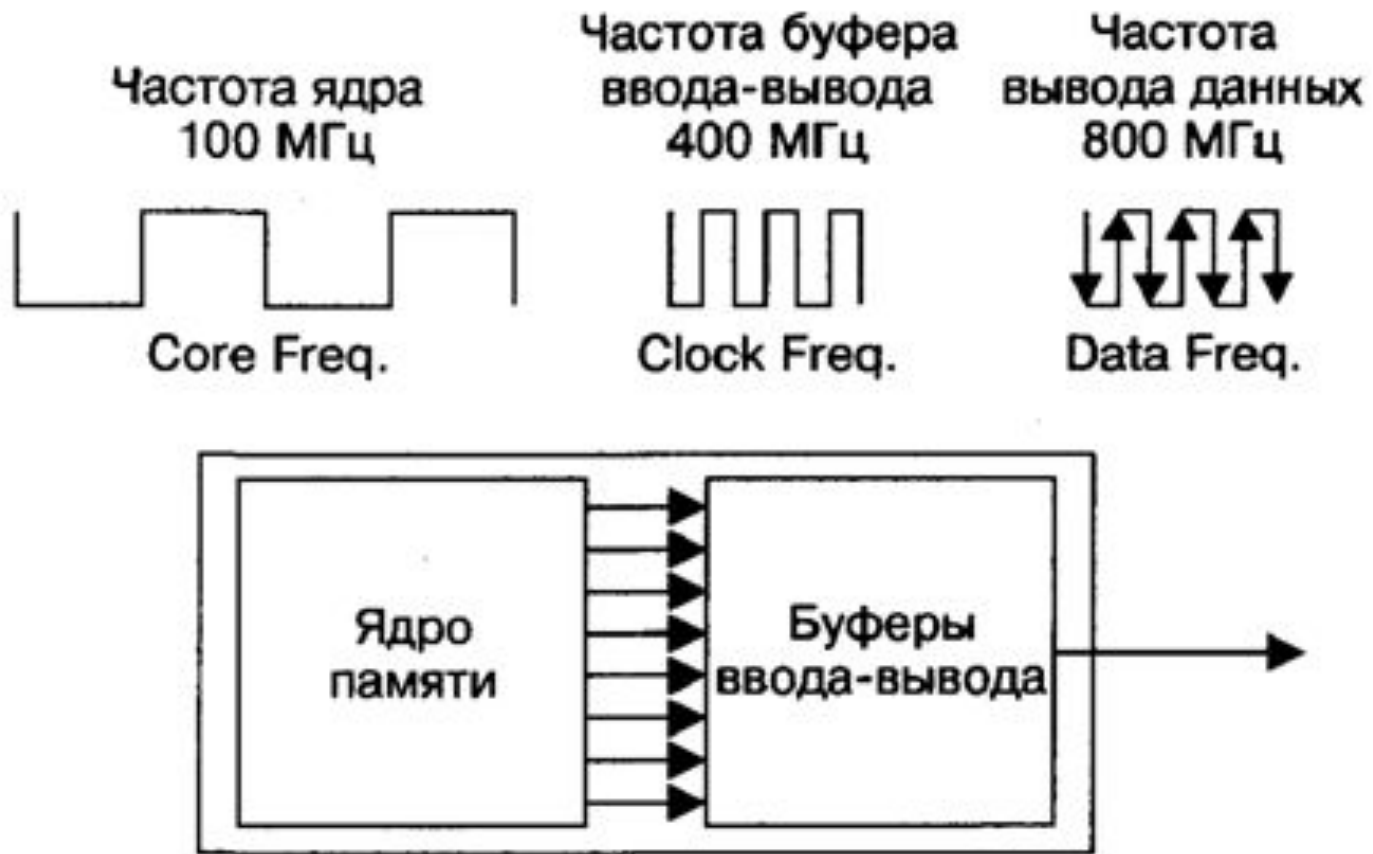


Рисунок 5 -Реализация метода 8n-Prefetch

Технология динамического терминирования сигналов позволяет гибко оптимизировать значения терминальных сопротивлений в зависимости от условий загрузки памяти.

Преимущества DDR3 по сравнению с DDR2:

- Более высокая пропускная способность (до 19200 МБ/с)
- Сниженное тепловыделение (результат уменьшения напряжения питания)
- Меньшее энергопотребление и улучшенное энергосбережение
- Немного уменьшенные задержки (в наносекундах)

Как и почему DDR3 тормозит появление DDR4

На сегодняшний день возможности архитектуры DDR3 вряд ли можно назвать исчерпавшими себя, так что пока есть смысл продолжать развитие этого стандарта и дальше.

Начало массового коммерческого производства DDR4 планируется на вторую половину 2012 года, а полномасштабный переход с DDR3 на DDR4 ожидается ближе к 2015 году.

ПАМЯТЬ DDR4 SDRAM

В январе 2011 года компания SAMSUNG официально представила новые модули, работающие в режиме DDR 4 - 2133 МГц, при напряжении в 1,2 В.

Повышение скоростных показателей памяти сохранится, и в скором времени разработчики освоят выпуск устройств, функционирующих на частоте 2667 МГц.

ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ПАМЯТИ

Пропускная способность — это максимальное количество данных, которое можно считать из памяти или записать в память в единицу времени. Именно эта характеристика прямо или косвенно отражается в названии типа памяти. Пропускная способность памяти зависит от ширины шины данных и частоты работы памяти. Ширина шины данных определяет количество бит, передаваемых за один такт, а частота работы памяти — количество тактов в единицу времени. Поэтому для того, чтобы определить пропускную способность памяти, нужно умножить частоту системной шины на ширину шины данных.

Тайминги памяти

Кроме максимальной пропускной способности, память характеризуется латентностью.

Причем во многих случаях латентность памяти оказывает большее влияние на производительность всей системы в целом, нежели тактовая частота работы памяти.

Под латентностью принято понимать задержку между поступлением команды и ее реализацией. Латентность памяти определяется ее таймингами, то есть задержками, измеряемыми в количествах тактов, между отдельными командами. Принято различать несколько разных таймингов памяти, соответствующих задержкам между различными командами. Для того чтобы разобраться с разными таймингами, рассмотрим последовательность команд при чтении или записи данных в память.

RAS-to-CAS Delay (tRCD)

Первоначально происходит активация нужной строки памяти (команда ACTIVE), для чего сигнал RAS переводится в низкий уровень. Далее следует команда записи (WRITE) или чтения (READ) данных, для чего сигнал CAS переводится в низкий уровень. При установке CAS на низкий уровень после прихода положительного фронта тактирующего импульса происходит выборка адреса столбца, наличествующего в данный момент на шине адреса, и открывается доступ к нужному столбцу матрицы памяти. Однако команда чтения или записи не может следовать непосредственно за командой активации — требуется, чтобы между этими командами, то есть между импульсами RAS и CAS, существовал некий промежуток времени — RAS-to-CAS Delay (задержка сигнала CAS относительно сигнала RAS).

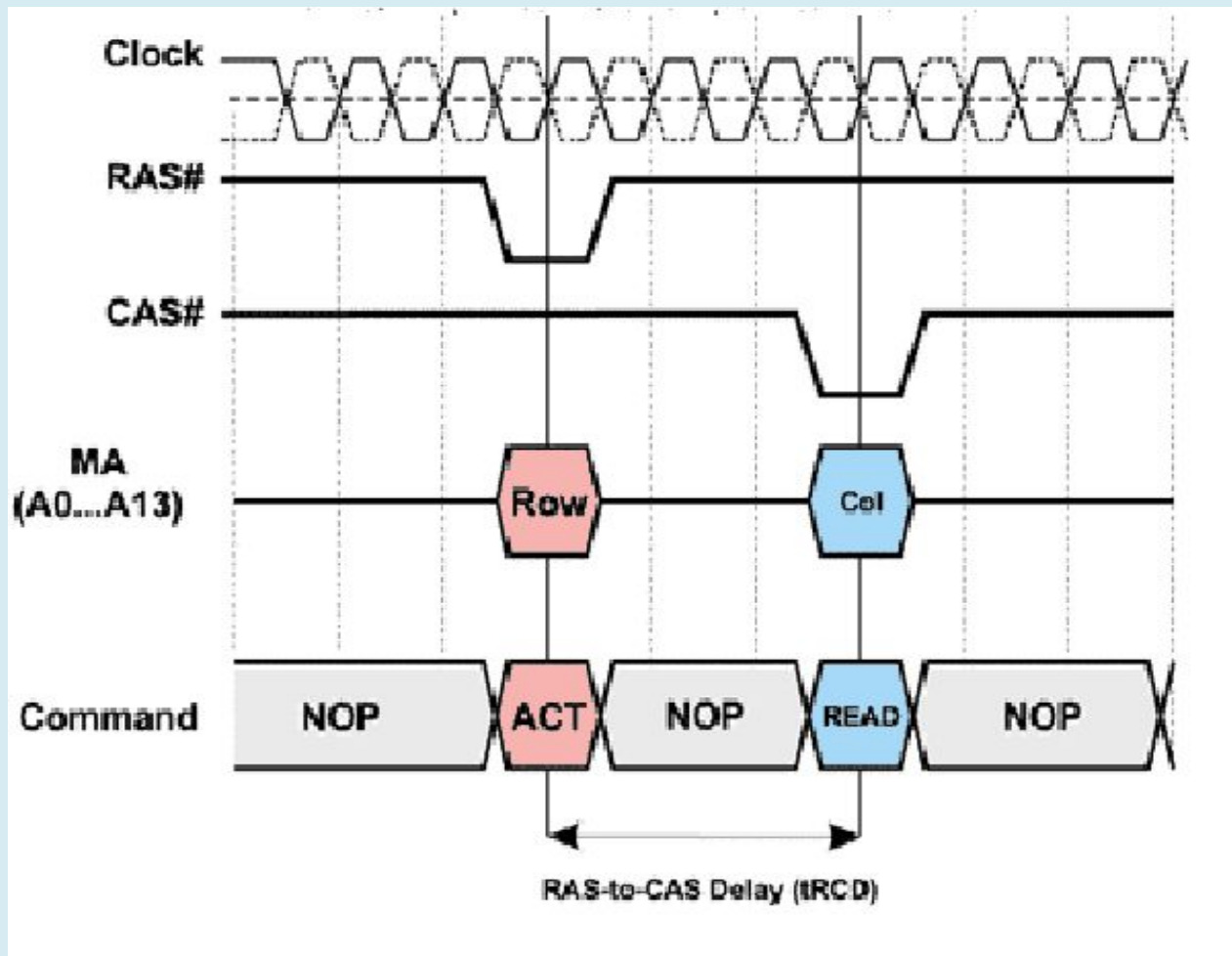


Рисунок 6 - Задержка RAS-to-CAS Delay (tRCD)

CAS Latency (tCL)

От команды чтения (записи) данных и до выдачи первого элемента данных на шину (записи данных в ячейку памяти) проходит промежуток времени, который называется CAS Latency. Эта задержка измеряется в тактах системной шины и обозначается tCL.

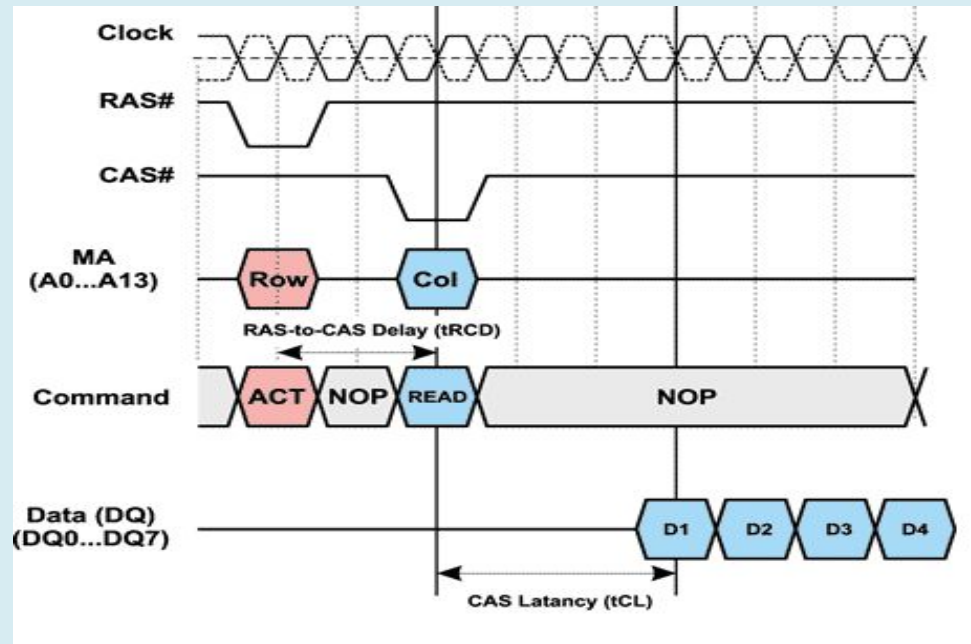


Рисунок 7 - Задержка CAS Latency (tCL)

Active-to-precharge delay (tRAS)

Еще один тип задержки, называемый Active-to-precharge delay, — это минимальный промежуток времени, который должен пройти с момента подачи команды активации строки до команды PRECHARGE. Эта задержка обозначается tRAS.

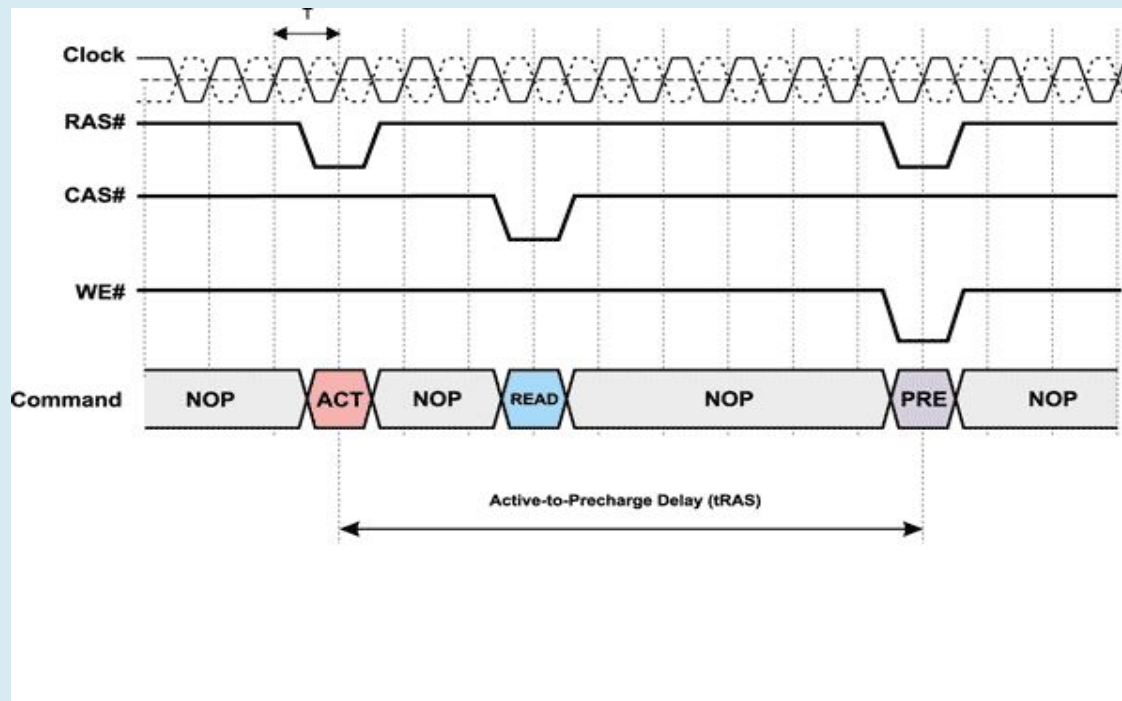


Рисунок 8 - Задержка Active-to-precharge (tRAS)

RAS Precharge (tRP)

Завершение цикла обращения к банку памяти осуществляется подачей команды PRECHARGE, приводящей к закрытию строки памяти. От команды PRECHARGE и до поступления новой команды активации строки памяти должен пройти промежуток времени (tRP), называемый RAS Precharge.

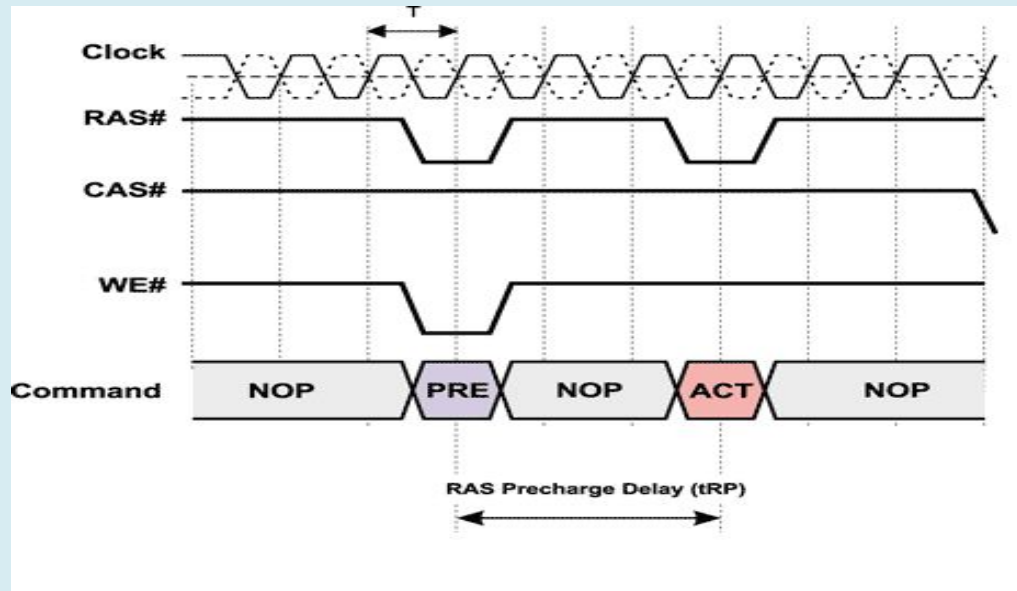


Рисунок 9 - Задержка RAS Precharge (tRP)

Command Rate

Следующий тип задержки, о котором необходимо сказать, — это скорость выполнения команд (Command Rate), представляющая собой задержку в тактах системной шины между командой CS# выбора чипа и командой активации строки.

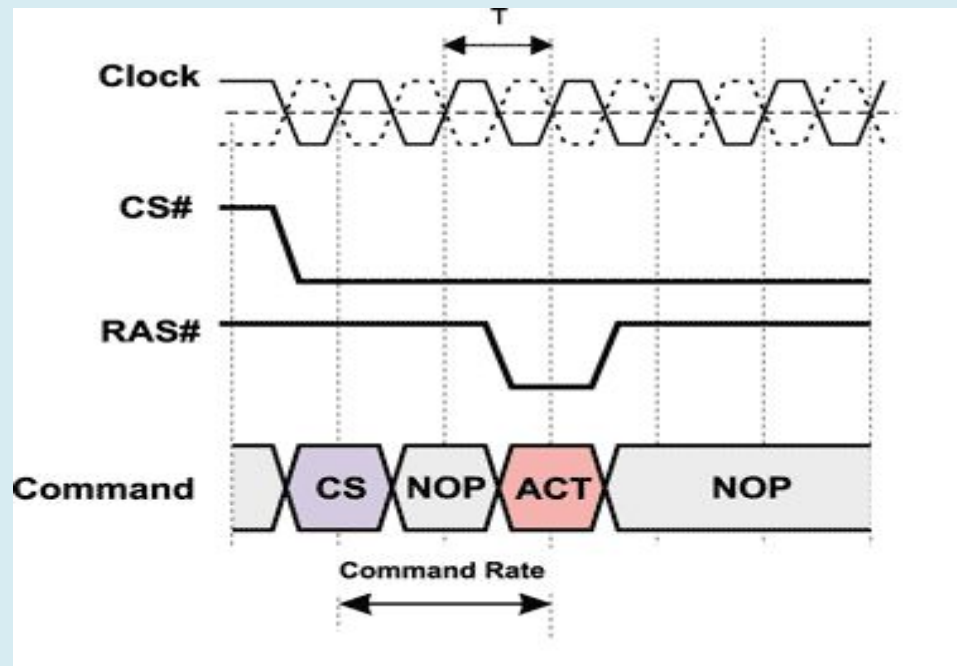


Рисунок 10 - Задержка Command Rate

Соотношения между таймингами памяти

Естественно, для каждого типа памяти значения различных задержек не могут быть произвольными и выбираются из допустимых значений.

Кроме того, между разными таймингам должны соблюдаться вполне определенные соотношения.

Временной промежуток между активацией строки и командой чтения определяется как t_{RCD} , а временной промежуток между командой чтения и появлением данных на шине — как t_{CL} . Временной промежуток между началом считывания данных и закрытием активной строки (t_{RAS}) зависит от длины передаваемого пакета, причем должно выполняться соотношение $t_{RAS} > t_{RCD} + t_{CL}$.

Минимальное значение t_{RAS} должно быть больше суммы t_{RCD} и t_{CL} на столько, на сколько велика длительность третьей операции, определяемая длиной передаваемого пакета.

Запись таймингов памяти

Описанные задержки — RAS-to-CAS Delay (t_{RCD}), CAS Latency (t_{CL}), RAS Precharge (t_{RP}), Active-to-precharge delay (t_{RAS}) и Command Rate — определяют тайминги памяти, обычно записываемые в виде последовательности t_{CL} - t_{RCD} - t_{RP} - t_{RAS} -Command Rate. К примеру, для модуля DDR400 (PC3200) тайминги могут быть следующими: 2-3-4-5-(1T). Это означает, что для данного модуля CAS Latency (t_{CL}) составляет 2 такта, RAS to CAS Delay (t_{RCD}) — 3 такта, RAS Precharge (t_{RP}) — 4 такта, ACTIVE-to-precharge delay (t_{RAS}) — 5 тактов и Command Rate — 1 такт.

GDDR

Как известно, оперативная память применяется для нужд не только центрального процессора, но и графического процессора. В современных графических видеокартах используется так называемая графическая память, микросхемы которой распаиваются на плате графической карты. Аналогично тому, что существуют различные типы оперативной памяти (SDR, DDR, DDR2 и DDR3), графическая память тоже бывает разной. Чтобы отличать оперативную память от графической, последнюю снабжают обозначением «G». Так, бывает память GDDR2, GDDR3, GDDR4 и GDDR5. Несмотря на схожие названия (GDDR2 и DDR2, GDDR3 и DDR3), графическая память существенно отличается от оперативной.

По принципу действия графическая память GDDR2 не имеет ничего общего с памятью DDR2 и в этом смысле более схожа с памятью DDR. В частности, в памяти GDDR2 не используется технология 4n-Prefetch, когда буфер ввода-вывода данных работает на удвоенной частоте. От обычной DDR-памяти GDDR2 отличается более высокими тактовыми частотами, требованиями к напряжению и способами терминирования сигналов.

Память GDDR3 также не имеет никакого отношения к DDR3 и по принципу действия более схожа с памятью DDR2, отличаясь от нее тактовыми частотами, требованиями к напряжению и способами терминирования сигналов. В памяти GDDR3, как и в DDR2, используется технология 4n-Prefetch.

Память GDDR4 является своеобразным аналогом DDR3 в том смысле, что в ней реализован механизм 8n-Prefetch.

GDDR5—5 поколение памяти DDR SDRAM, спроектированной для приложений, требующих высокой полосы пропускания. В отличие от его предшественника, GDDR4, GDDR5 основан на памяти DDR3, которая имеет удвоенные по сравнению с DDR2, DQ(DigitalQuest) каналы связи, но у GDDR5 также есть буферы предварительной выборки шириной 8 битов как у GDDR4.

Вывод: оперативная память — энергозависимая часть системы компьютерной памяти, в которой временно хранятся данные и команды, необходимые процессору для выполнения им операции. Обязательным условием является адресуемость (индивидуальный адрес) памяти.

Физически оперативная память в системе представляет собой набор микросхем или модулей, содержащих микросхемы, которые обычно подключаются к системной плате. Эти микросхемы или модули могут иметь различные характеристики и, чтобы функционировать правильно, должны быть совместимы с системой, в которую устанавливаются.

Производительность оперативной памяти гораздо выше дисковой, поэтому от ОЗУ зависит быстрдействие всей системы ПК.

Если оперативной памяти не стало хватать для нормальной работы, компьютер преобразует часть жесткого диска в оперативную память и продолжает работать, при этом компьютер начинает работать медленнее.

Поэтому необходимо увеличить ОЗУ. Для того чтобы ваш компьютер нормально функционировал, необходима оперативная память большого размера. Особенно, если вы играете, смотрите фильмы или работаете с программами обработки графики.

Оперативной памяти не бывает много, с каждым годом, выпускаемое новое программное обеспечение требует все более мощных характеристик устройств, установленных на вашем компьютере, в первую очередь это касается процессора и оперативной памяти.