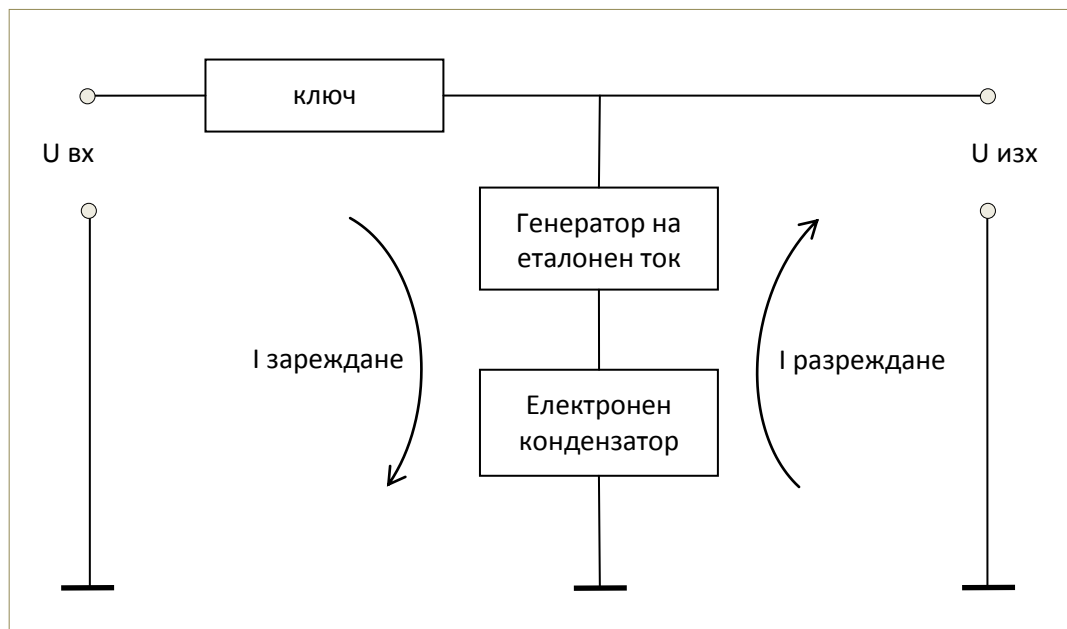


Видове DRAM памети - SDR SDRAM, DDR, DDR2, DDR3, DDR4, DDR5 и RAMBUS. Сравнителен анализ.

I. Въведение

При персоналните компютри се използва два вида RAM - динамична (DRAM) и статична (SRAM). Статичната памет, използвана при персоналните компютри, се нарича *кеш* памет. Времето за достъп до клетка от статичната памет е много по-малко отколкото времето за достъп при динамичната памет. Причината за това е начина на функциониране на запомнящите клетки при двата вида памет. Докато при SRAM запомнящият елемент е D тригер, то при DRAM запомнящият елемент е интегрален кондензатор (използва се капацитета между изводите дрейн и сорс на интегрален транзистор). На фиг. 1 е показана структурата на една запомняща клетка от DRAM.



Фиг. 1. Структура на запомняща клетка от DRAM

Запомнящият елемент е електронен кондензатор с капацитет около 0.1pF. Той се зарежда за определен интервал от време до стойността на входното напрежение $U_{вх}$ (лог. 0 или лог. 1). За да може времето за зареждане на кондензатора да е независимо от стойността на входното напрежение, токът на зареждане се получава от генератор за еталонен ток. Зареждането започва след затваряне на ключа (транзистор, включен като електронен ключ). След зареждане на кондензатора, ключът се отваря. От този момент нататък запомнящата клетка не консумира електроенергия, но кондензаторът започва да се разрежда, поради тока на утечка. Преди момента на прекомерно намаляване на изходното напрежение $U_{изх}$, ключът отново се затваря.

Този процес на зареждане-разреждане на електронните кондензатори продължава непрекъснато по време на работа на компютъра и се нарича опресняване на динамичната памет. Интервалът на опресняване на паметта при персоналните компютри е няколко десетки милисекунди (50-60 ms). Той се задава от един от таймерите на дънната платка.

Самото опресняване се реализира чрез специален цикъл на „четене“ на клетките на паметта. Той се реализира чрез канал 0 на контролера за директен достъп до паметта (DMA), който след включване на захранването се програмира от BIOS / UEFI да работи в режим „памет-памет“. По време на опресняване изходите за данни на паметта са изключени и следователно достъп до паметта с цел четене или запис е невъзможен. Стандартът JEDEC дефинира време за опресняване на ред от паметта всеки 64 ms или по-малко. Следователно интервалът между циклите на опресняване се получава по следната формулата

$$\text{интервал на опресняване} = \text{време за опресняване} / \text{брой на DRAM редовете}$$

Например, при DRAM с 8192 реда интервалът на опресняване трябва да е по-малък от 7.8185 μ s. Трябва да се има предвид, че съществуват памет, които могат да се тактуват от собствен тактов генератор. Този режим се използва, когато компютърната конфигурация се приведе в режим с намалена консумация. Една от най-новите технологии за опресняване на DRAM е на фирма Samsung и се нарича Partial Array Self Refresh (PASR). При тази технология целта е намаляване на енергопотреблението на модулите. За целта опресняването се реализира само в банките от паметта, които съдържат данни, но не се извършват операции четене или запис с тях.

Времето за превключването на тригерите (изградени са изцяло чрез интегрални транзистори) е в пъти по-малко от времето на презареждане на кондензаторите. Следователно, ако ни трябва бърза памет трябва да изберем SRAM. Динамичната памет има друго предимство, което обаче е съществено – интегралният кондензатор не се нуждае от захранване постоянно, за да помни логическото ниво до което е зареден. Следователно, при един и същи капацитет, динамичната памет ще консумира много по-малко от статичната памет. Освен това тя е и по-евтина, тъй като е изградена от по-малко на брой транзистори. Ако ви трябва памет, която е евтина и не консумира много, трябва да изберете DRAM. Ето защо оперативната памет при персоналните компютри е динамична. Тя представлява модули, които се поставят в специални слотове на дънната платка и може да има обем до няколко до стотици GiB, без да е необходимо специално охлаждане.

II. Достъп до DRAM

Една компютърна конфигурация, ако използва само оперативна памет (DRAM), няма да е ефективна. Причината за това е, че CPU ще е стотици пъти по-бърз от оперативната памет. Ако програмата, която се изпълнява от CPU, е в DRAM, то микропроцесорът трябва да изчаква всеки път, когато се налага да извлече операционния код на следващата инструкция. Решението е да се използва бърза SRAM (кеш) с която CPU да работи. В този случай някой трябва да зарежда кеша с инструкции и данни, когато е необходимо. Тъй като кеш паметта е скъпа, тя е с много по-малък размер от DRAM. Управлението на паметта при компютрите се реализира от чипсета – по-точно от северния мост при чипсети с мостова архитектура или Memory Controller Hub (MCH) при хъбова архитектура. Ако разликата в бързодействието на CPU и DRAM е много голяма може да се наложи да се използват няколко нива на кеширане. Кешът, който е най-близо до CPU се нарича кеш от ниво 1 (L1 кеш). Съвременните процесори са с 3 и повече нива на кеш. За да няма забавяне при комуникацията между CPU и кеш, кешът най-често се интегрира на една подложка с CPU. Съвременните процесори нямат физически достъп до оперативната

памет. Те адресират само L1 кеш. Синхронизирането на работата на цялата система CPU-SRAM-DRAM се реализира от чипсета.

Достъпът до кеш паметта е на ниво кеш ред. Един кеш ред е с размер от n на брой последователни байта от DRAM (най-често $n=16, 32, 64$). Следователно, достъпът до DRAM също трябва да е блоков. Най-малкото количество информация, която се чете или записва от/в DRAM е блок от данни, който е съобразен с размера на кеш реда. Блок-ориентираният обмен има съществено предимство – вместо да се налага синхронизиране при обмен на съдържанието на всеки 64 байта от DRAM, се реализира синхронизация само в началото на блока с данни. Така се намалява значително времето за обмен с динамичната памет.

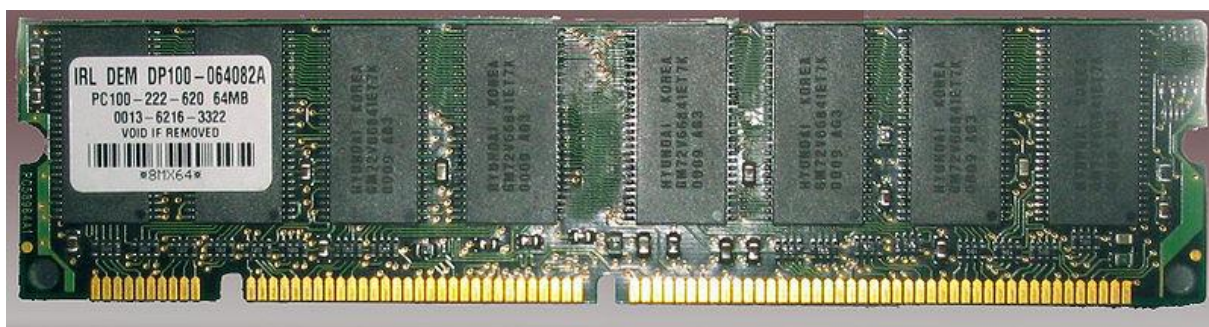
III. Видове DRAM

Във времето при персоналните компютри са се използвали различни видове DRAM. При първите персонални компютри оперативната памет се формира от група интегрални схеми DRAM, които се намират на дънната платка. Това решение не се използва отдавна, тъй като не позволява промяна на обема на инсталираната оперативна памет. За да е възможно това, започва да се произвеждат модули, които съдържат определен брой интегрални схеми DRAM. Тези модули се поставят в специални слотове на дънната платка. Така става възможно инсталирането на различен обем оперативна памет, в зависимост от предназначението на компютъра и от финансовите възможности на потребителите на компютърна техника. Минималното количество оперативна памет зависи от операционната система, която е инсталирана. Максималното количество зависи от приложенията, които ще използвате и финансовите ви възможности. Видът на DRAM и теоретичния максимален обем на оперативната памет, зависи от използвания чипсет.

Един от първите DRAM модули, използвани при персоналните компютри, е Fast Page Mode (FPM). Те се анонсира на пазара през 1995 година. Тези модули се захранват с 5V. Те са 32 битови с време за достъп 70 ns или 60 ns. Една страница (page) се формира от n последователни клетки от паметта – един ред от запомнящата матрица. Тази организация на достъпа позволява ускоряване на производителността при достъп до множество последователни адреси от DRAM. Наименованието на модулите е Single In-Line Memory Module (SIMM) и те имат 30 или 72 извода. Едно подобрение на FPM модулите са Extended Data Out (EDO) модулите, които започват да се използват през 1996 година. Те се захранват с 3.3 V. Модулите са 64-битови, а честотата на синхронизация е 66 MHz. Бързодействието при четене на EDO паметите е с 15% по-добро от FPM само благодарение на буфериране на данните на изхода. Това буфериране позволява докато чипсета чете текущо извлечените данни от буфера да се стартира нов цикъл за четене (конвейеризация на операция четене). Модули FPM и EDO няма да бъдат разглеждани, тъй като те отдавна не се използват.

3.1. Single Data Rate (SDR) Synchronous DRAM (SDRAM)

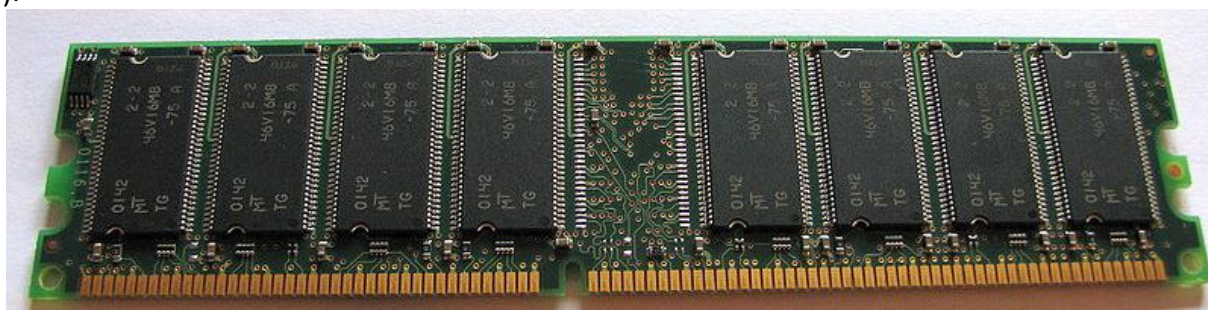
Модулите SDR SDRAM са синхронна динамична памет, обменът с която се синхронизира по преден фронт на синхроимпулсите. Тези модули са 64-битови, захранват се с 3.3V и имат 168 извода. Модулите се означават като PC-X, където числото X (66, 100 или 133) е честотата на такта на синхронизация в MHz. Следователно, времето за достъп до паметта е съответно 15 ns, 10 ns или 7.5 ns. На Фиг. 2 е показан модул памет PC-100 SDR SDRAM.



Фиг. 2. Модул SDR SDRAM (PC-100)

3.2. Double Data Rate (DDR) Synchronous DRAM (SDRAM)

Към 2000 година обемът (до 256 MiB) и бързодействието на SDR SDRAM станеха недостатъчни и тя бе заменена от DDR SDRAM. Тази памет е 64-битова и има по-ниска консумация на енергия, тъй като се захранва с 2.5V. Модулите са с 184 извода (виж Фиг. 3).



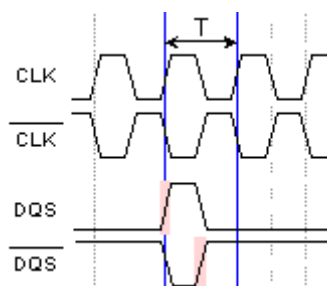
Фиг. 3. Модул DDR SDRAM

Наименованието, съгласно стандарта, на паметите е DDR-XXX, където числото XXX е ефективната честота на синхронизация. Тя е два пъти по-висока отколкото физическата тактова честота. Причината за това е, че DDR синхронизира както по преден, така и по заден фронт на синхроимпулсите. Наименованието на модулите е PC-YYYY, където YYYY е пиковата скорост на обмен в MiB/s. Параметрите на DDR паметите, които са били произвеждани, са описани в Табл. 1.

Табл. 1. Параметри на DDR SDRAM

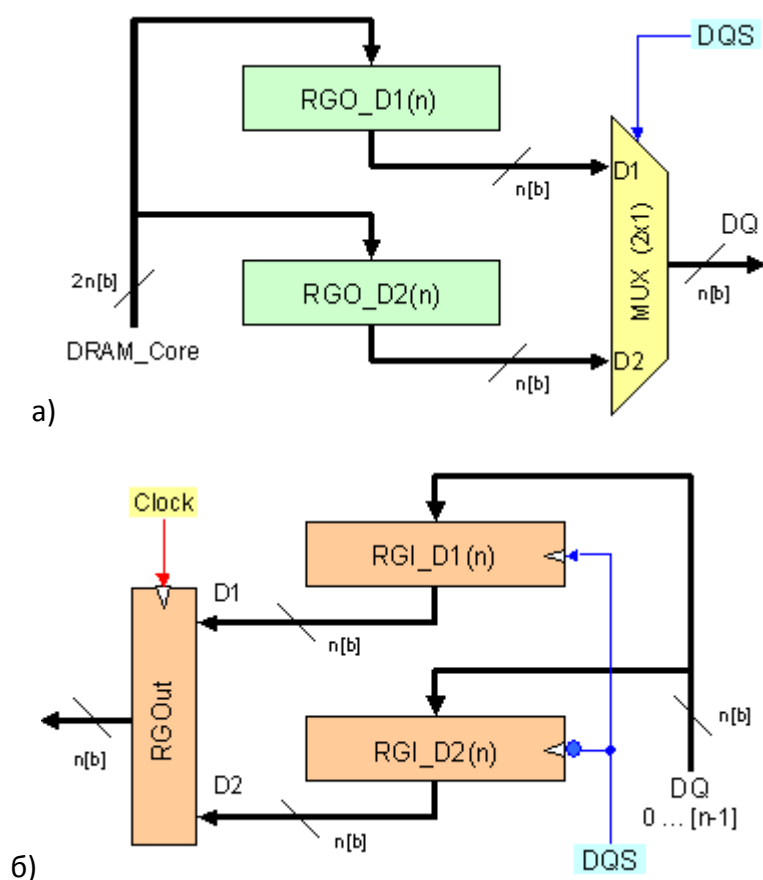
Наименование съгласно стандарта	Физическа Тактова честота, MHz	Време за достъп, ns	Ефективна тактова честота, MHz	Захранващо напрежение, V	Име на модула	Максимална скорост на обмен, MiB/s
DDR-200	100	10	200	2.5±0.2	PC-1600	1600
DDR-266	133	7.5	266	2.5±0.2	PC-2100	2100
DDR-333	166	6	333	2.5±0.2	PC-2700	2700
DDR-400	200	5	400	2.6±0.1	PC-3200	3200

На практика DDR не синхронизира по преден и заден фронт, тъй като тригерите в буферните регистри прехвърлят данни от входа към изхода само по преден фронт. Следователно, за да се удвои пропускателната способност на паметта е необходимо в рамките на един период на физическия такт да има два предни фронта. Това може да се постигне само ако тактовата последователност се използва парафазно. За целта се инвертира физическият тактов сигнал (CLK), както е показано на Фиг. 4. Получава се сигнал Data Queue Strobe (DQS).



Фиг. 4. *Формиране на два предни фронта за един период на сигнал CLK*

В този случай, въпреки, че паметта е 64 битова, за един период на сигнал CLK се извличат от паметта 128 бита данни. За да е възможно това се налага паметта да има два изходни буфера за реализация на операция четене, както и два входни буфера за реализация на операция запис (виж Фиг. 5). Технологията, която се базира на тази логика, е известна като 2n-prefetch (извличане на 2 пъти повече данни за един такт на CLK).



Фиг. 5. *Технология 2n-prefetch: а) операция четене; б) операция запис*

Сигнал DQS стробира извеждането на данните при четене и въвеждането им при запис. Например, при операция четене (фиг. 5а), при първия преден фронт на DQS двувходния мултиплексор (MUX) прехвърля към шината данни първите 64 бита, а при втория преден фронт – следващите 64 бита. Така за един период се прочитат 128 бита. Точното синхронизирането на фронтовете на сигнал DQS и сигнал CLK е много важно за операция запис и то се получава чрез използване на Delay Locked Loop (DLL).

3.3. DDR2 SDRAM

Следващата във времето SDRAM е DDR2. Паметта се анонсира на пазара през 2003 година. Тези памети са 64-битови и се захранват с 1.8V. Броят на изводите е 240 (виж Фиг. 6).



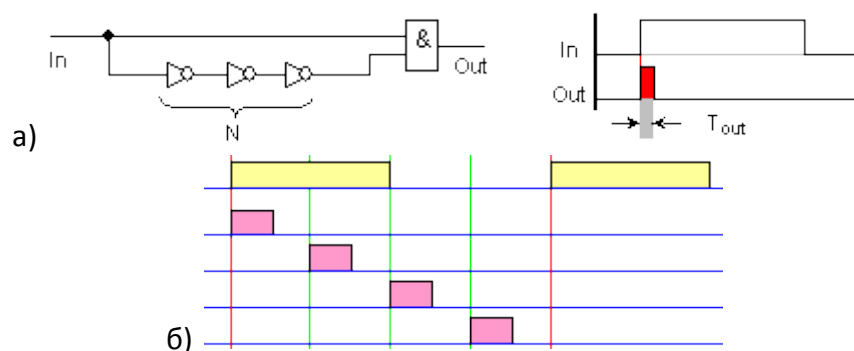
Фиг. 6. Модул DDR2 SDRAM (PC2-6400)

Видовете DDR2 памети и техните параметри са описани в Табл. 2.

Табл. 2. Параметри на DDR2 SDRAM

Наименование съгласно стандарта	Физическа тактова честота, MHz	Време за достъп, ns	Ефективна тактова честота, MHz	Име на модула	Максимална скорост на обмен, MiB/s
DDR2-400	100	10	400	PC-3200	3200
DDR2-533	133	7.5	533	PC-4200	4266
DDR2-667	166	6	667	PC-5300	5333
DDR2-800	200	5	800	PC-6400	6400
DDR2-1066	266	3.75	1066	PC-8500	8533

При DDR2 за един такт на физическия тактов сигнал се прочитат 2 пъти повече данни отколкото при DDR и 4 пъти повече данни отколкото при SDR. Следователно, за един такт се извличат 4 x 64 бита или 256 бита. За да е възможно това са необходими 4 предни фронта за стробиране на данните в рамките на един такт на CLK. Това се постига чрез технология 4n-prefetch. Получаването на един строб от 4n-prefetch е показано на Фиг. 7.



Фиг. 7. Получаване на стробиращи сигнали: а) схема за получаване на един сигнал; б) Четири стробиращи сигнала в рамките на един период на CLK

За целта входният сигнал In се подава на един от входовете на двувходов И логически елемент. На другия вход се подава инвертирано и закъсняло копие на сигнал In. Изходният сигнал Out е синхронен на In и е с продължителност T_{out} , която зависи от броя на инверторите N. За да се реализира 4n-prefetch, сигнал In е CLK. Необходими са 4 еднотипни схеми, както тази показана на Фиг. 7, за да се получат 4 стробиращи сигнала. Те трябва да са дефазирани спрямо CLK на 0, 45 ($\frac{1}{4}T$), 90 ($\frac{1}{2}T$) и на 135 ($\frac{3}{4}T$) градуса. Реализацията на 4n-prefetch изисква 4 изходни и 4 входни буфера, а мултиплексора при операция четене трябва да е 4-входов.

3.4. DDR3 SDRAM

През 2007 година на пазара се анонсира DDR3 SDRAM. Тази памет е 64-битова и се захранва с 1.5V. Максималният обем е 8 GiB. Видовете DDR3 памет и техните параметри са описани в Табл. 3.

Табл. 3. Параметри на DDR3 SDRAM

Наименование съгласно стандарта	Физическа тактова честота, MHz	Време за достъп, ns	Ефективна тактова честота, MHz	Име на модула	Максимална скорост на обмен, MiB/s
DDR3-800	100	10	800	PC3-6400	6400
DDR3-1066	133	7.5	1066	PC3-8500	8533
DDR3-1333	166	6	1333	PC3-10600	10667
DDR3-1600	200	5	1600	PC3-12800	12800
DDR3-1866	233	4.29	1866	PC3-14900	14933
DDR3-2133	266	3.76	2133	PC3-17064	17066

При DDR3 за един такт на физическия тактов сигнал се прочитат 2 пъти повече данни отколкото при DDR2, 4 пъти повече данни отколкото при DDR и 8 пъти повече данни отколкото при SDR. Следователно, за един такт се извличат 8 x 64 бита или 512 бита. За да е възможно това са необходими 8 предни фронта за стробиране на данните в рамките на

един такт на CLK. Това се постига чрез технология 8n-prefetch. Реализацията на 8n-prefetch изисква 8 изходни и 8 входни буфера, а мултиплексора при операция четене трябва да е 8-входов.

3.5. DDR4 SDRAM

През 2014 година на пазара се анонсират памети DDR4 SDRAM. Захранващото напрежение на тези памети е 1.2V. Броят на изводите е 288. Максималният обем DDR4 памет, който може да се инсталира е до 512 GiB. Видовете DDR4 памети и техните параметри са описани в Табл. 3.

Табл. 4. Параметри на DDR4 SDRAM

Наименование съгласно стандарта	Физическа тактова честота, MHz	Време за достъп, ns	Ефективна тактова честота, MHz	Име на модула	Максимална скорост на обмен, MiB/s
DDR4-1600	200	5	1600	PC4-12800	12800
DDR4-1866	233	4.29	1866	PC4-14900	14900
DDR4-2133	266	3.76	2133	PC4-17000	17000
DDR4-2400	300	3.33	2400	PC4-19200	19200
DDR4-2666	333	3	2666	PC4-21333	21333
DDR4-2933	366	2.73	2933	PC4-23466	23466
DDR4-2933	400	2.5	2933	PC4-25600	25600

При DDR4 за един такт на физическия тактов сигнал се прочитат 2 пъти повече данни отколкото при DDR3, 4 пъти повече данни отколкото при DDR2 и 8 пъти повече данни отколкото при DDR и 16 пъти повече данни отколкото при SDR. Следователно, за един такт се извличат 16 x 64 бита или 1024 бита. За да е възможно това са необходими 16 предни фронта за стробиране на данните в рамките на един такт на CLK. Това се постига чрез технология 16n-prefetch. Реализацията на 16n-prefetch изисква 16 изходни и 16 входни буфера, а мултиплексора при операция четене трябва да е 16-входов.

3.6. DDR5 SDRAM

През 2019 година JEDEC финализира стандарта DDR5 SDRAM. Тези модули се захранват с 1.1 V, като по този начин се намалява консумацията на енергия. Модулите DDR5 имат вградени регулатори на напрежение. Поради високата цена на паметите се очаква те да се използват само за сървърни платформи и 5G комуникационни платформи. DDR5 поддържа скорост до 51.2 GB/s на модул. Поддръжката на DDR5 за персонални компютри и лаптопи зависи основно от фирмите производители на чипсети.

3.7. RAMBUS SDRAM

Модулите SDRAM памети от тип DDRx имат един съществен недостатък – широката шина за данни (64 бита) води до генериране на електромагнитни смущения в цялото дъно. Това е основната причина, за да не се използват много високи тактови честоти (100-400 MHz).

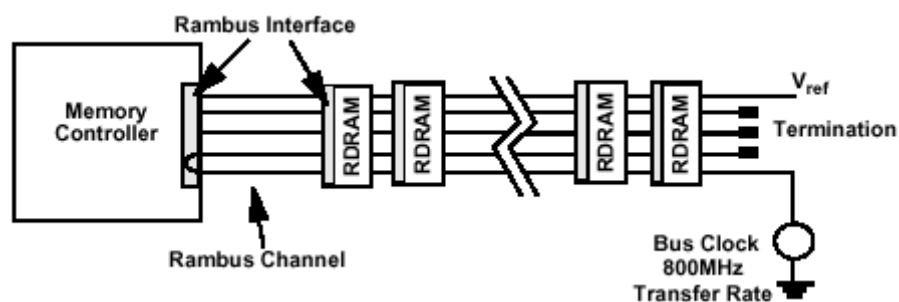
Увеличаването на производителността на една SDRAM е възможно в две направления:

1. Увеличаване на честотата на сигнал CLK.
2. Увеличаване на броя на линиите за данни.

Едновременно увеличаване на честотата на CLK и броя на линиите за данни е невъзможно, поради електромагнитна нестабилност. Увеличаването на броя на линиите за данни е неприемливо, тъй като това затруднява опроводяването на дънната платка, която има фиксиран по стандарт форм-фактор.

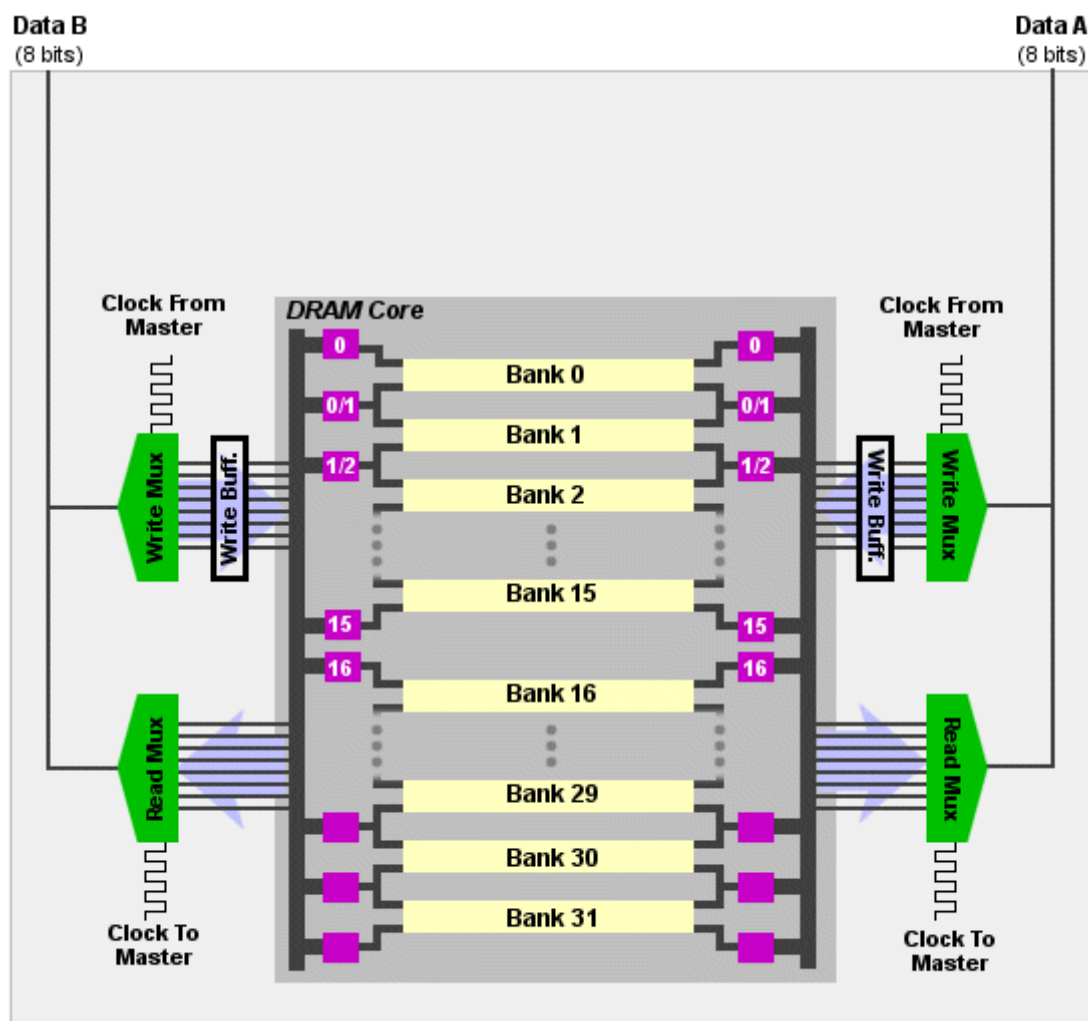
На този етап печелившата стратегия е увеличаване на честотата на CLK при намаляване на броя на линиите за данни. Типичен пример за използване на тази стратегия са RAMBUS паметите. Първоначално те са се използвали при игровите конзоли и значително по-късно започват да се използват и при персоналните компютри, благодарение на разработения за целта чипсет от фирма Intel.

На Фиг. 8 е показана блок-схема, която демонстрира как функционират RAMBUS паметите. За разлика от DDRx паметите при които всички сигнали се подават паралелно на всички DDR модули от дънната платка, при RAMBUS тези сигнали преминават последователно през всички RIMM модули. Освен това, сигнал CLK се получава от специален генератор на дънната платка, а не от чипсета. Сигнал CLK, който се получава от генератора се нарича Clock-To-Master. Този сигнал преминава през всички RIMM модули и достига до контролера на паметта. Той дефазира сигнал Clock-To-Master на $\frac{1}{2}T$ и така се получава нов тактов сигнал с име Clock-From-Master. Тези два CLK сигнала са в непосредствена близост един до друг и по този начин тъй като са дефазирани се компенсират електромагнитните смущения. Това позволява при тези памети да може да се използват много високи честоти за сигнал CLK.



Фиг. 8. Начин на функциониране на RAMBUS SDRAM

На Фиг. 9 е показана вътрешната структура на RAMBUS памет. Използват се две 8-битови шини за данни (Data A и Data B), които могат да се използват паралелно. Тъй като паметта вътрешно е 64-битова се налага прочитането на една клетка да се реализира за 4 периода на сигнал Clock-To-Master, а записът – за 4 периода на сигнал Clock-From-Master. Паметта е организирана на ниво банки, като едновременно има достъп до 2 банки памет.

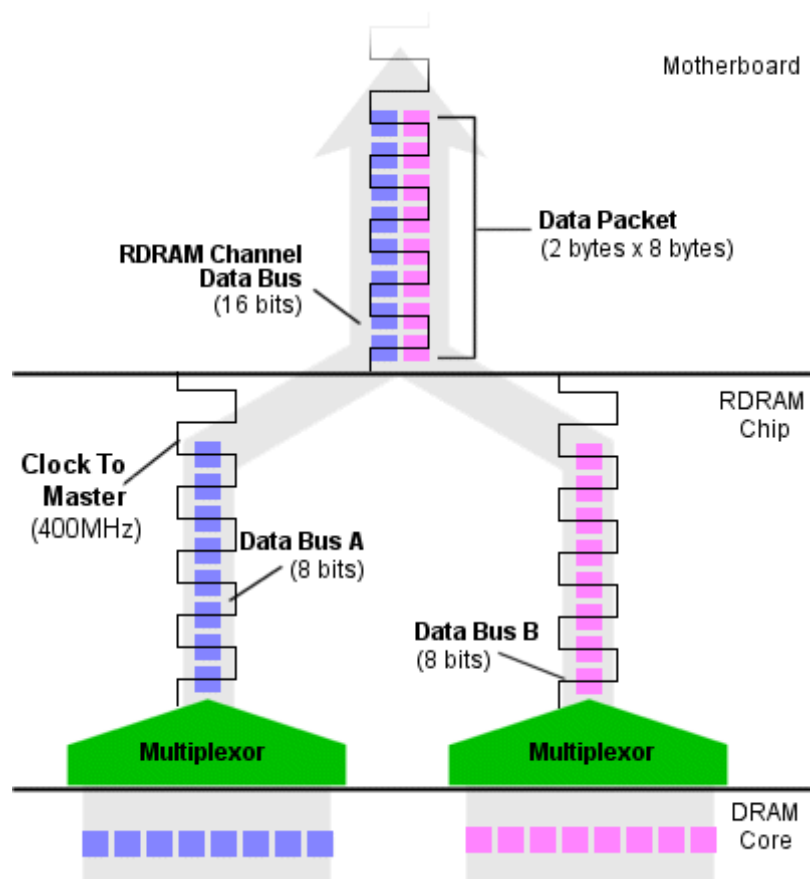


Фиг. 9. Вътрешна структура на RAMBUS SDRAM

На Фиг. 10 е показана връзката между RAMBUS памет и контролера на паметта при режим на четене. Един пакет с данни (16 байта) се получава за 4 такта на сигнал Clock-To-Master – 8 байта от канал Data A и 8 байта от канал Data B.

IV. Въпроси и задачи за самостоятелна работа

1. Каква е функцията на генератора на еталонен ток в запомнящите клетки на DRAM?
2. Защо при DDR-400 захранващото напрежение е 2.6V, а при останалите DDR памети е 2.5V?
3. Защо при някои модули DRAM има четен брой интегрални схеми, а при други – нечетен брой интегрални схеми? Какво представляват ECC паметите?
4. Защо при RAMBUS дъната не може да се остави празен RAMBUS слот?
5. Какво представляват Fully Buffered (FB) и Load Reduced (LR) модули памет?



Фиг. 10. Обмен с RAMBUS SDRAM при операция четене