

Тема 7. Організація основної пам'яті комп'ютерних систем

Основна пам'ять (ОП) є єдиним видом пам'яті, до якої ЦП може звертатися безпосередньо (виняток становлять лише регістри центрального процесора). Інформація, що зберігається на зовнішніх ЗП, стає доступною процесору тільки після того, як буде переписана в основну пам'ять.

Основну пам'ять утворюють запам'ятовуючі пристрої з довільним доступом. Такі ЗП утворені як масив комірок, а «довільний доступ» означає, що звернення до будь-якої комірки займає один і той же час і може проводитися в довільній послідовності. Кожна комірка містить фіксоване число запам'ятовуючих елементів і має унікальну адресу, що дозволяє розрізнити комірки під час звернення до них для виконання операцій запису і зчитування.

Наслідком величезних успіхів у області напівпровідникових технологій стала зміна елементної бази основної пам'яті. На зміну ЗП на базі феромагнітних кілець прийшли напівпровідникові мікросхеми, використання яких у наші дні стало широко розповсюдженим.

Основна пам'ять може включати два типи пристроїв: оперативні запам'ятовуючі пристрої (ОЗП) і постійні запам'ятовуючі пристрої (ПЗП).

Переважну частку основної пам'яті утворює ОЗП, який називають оперативним, тому що він допускає як запис, так і зчитування інформації, причому обидві операції виконуються однотипно, практично з однією і тією ж швидкістю, і проводяться за допомогою електричних сигналів. У англійській літературі ОЗП відповідає абревіатура RAM – Random Access Memory, тобто «пам'ять з довільним доступом», що не зовсім коректно, оскільки пам'яттю з довільним доступом є також ПЗП і регістри процесора. Для більшості типів напівпровідникових ОЗП характерна енергозалежність – навіть під час короткочасного переривання живлення інформація, що зберігається, втрачається. Мікросхема ОЗП повинна бути постійно підключена до джерела живлення і тому може використовуватися тільки як тимчасова пам'ять.

Другу групу напівпровідникових ЗП основної пам'яті утворюють незалежні мікросхеми ПЗП (ROM – Read-Only Memory). ПЗП забезпечує зчитування інформації, але не допускає її зміни (у ряді випадків інформація в ПЗП може бути змінена, але цей процес сильно відрізняється від зчитування і вимагає значно більшого часу).

7.1 Блочна організація основної пам'яті

Ємність основної пам'яті сучасних комп'ютерів достатньо велика, щоб її можна було реалізувати на одній інтегральній мікросхемі (ІМС). Необхідність об'єднання декількох ІМС виникає також, коли розрядність комірок у мікросхемі запам'ятовуючого пристрою менше розрядності слів комп'ютера.

Збільшення розрядності ЗП реалізується за рахунок об'єднання адресних входів ІМС ЗП, які об'єднуються. Інформаційні входи і виходи мікросхем є входами і виходами модуля ЗП збільшеної розрядності (рис. 7.1). Таку об'єднану сукупність мікросхем називають модулем пам'яті. Модулем можна називати і одну мікросхему, якщо вона вже має необхідну розрядність. Один або декілька модулів утворюють банк пам'яті.

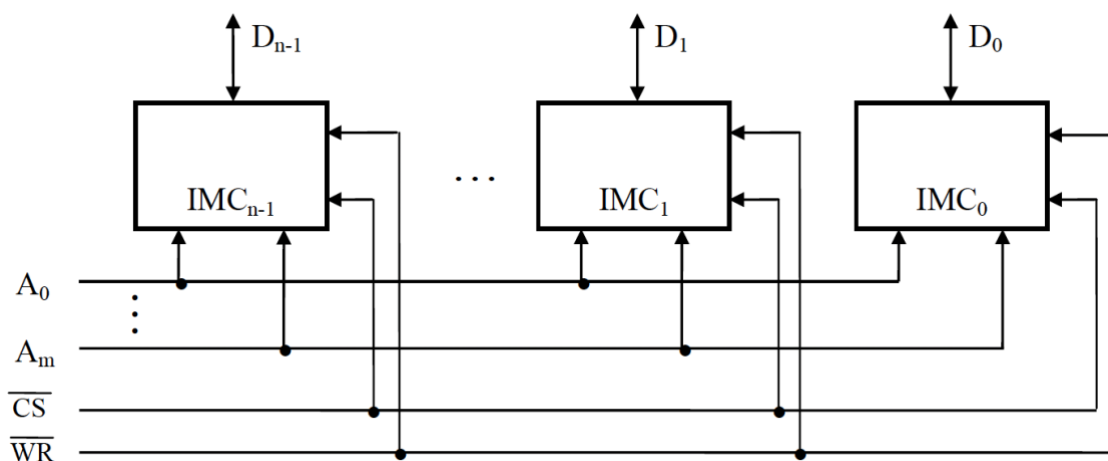


Рисунок 7.1 – Збільшення розрядності пам'яті

Для отримання потрібної ємності ЗП потрібно певним чином об'єднати декілька банків пам'яті меншої ємності. Взагалі основна пам'ять комп'ютера практично завжди має блочну структуру, тобто містить декілька банків.

Під час використання блочної пам'яті, яка складається з B банків, адреса комірки A перетворюється в пару (b, w) , де b - номер банку, w - адреса комірки всередині банку. Відомі три схеми розподілу розрядів адреси A між b і w : блочна; циклічна; блочно-циклічна.

Типова структура блочної пам'яті показана на рис. 7.2.

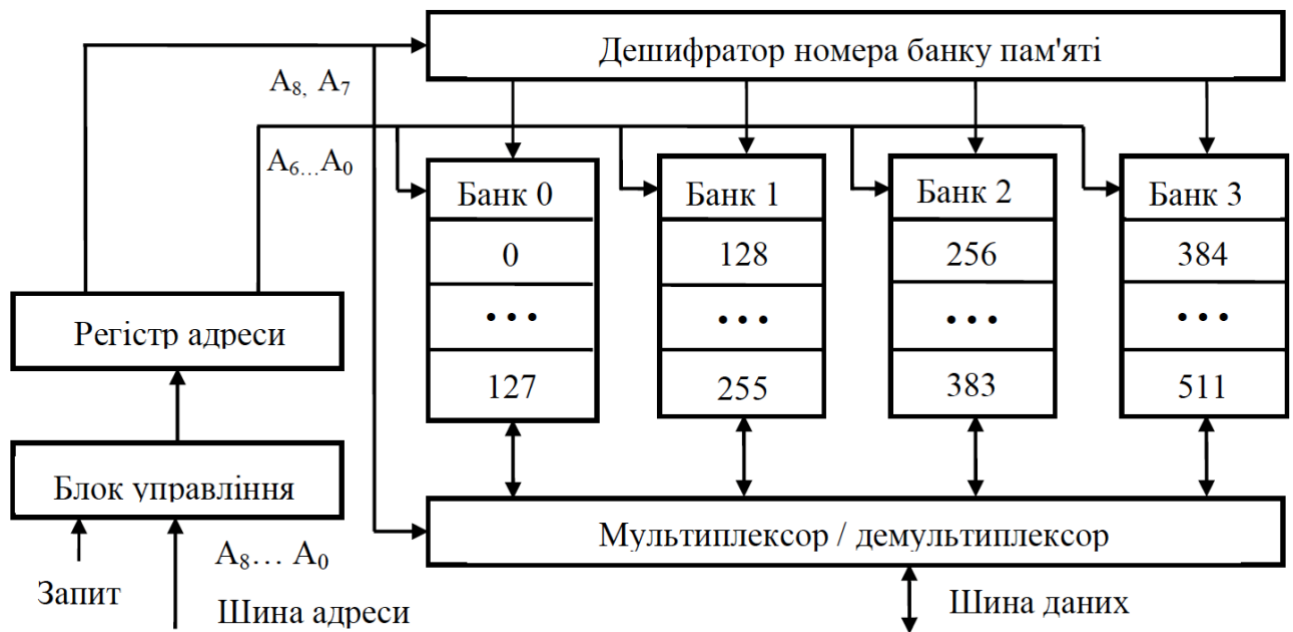


Рисунок 7.2 – Структура основної пам'яті блочного типу

У функціональному відношенні така пам'ять може розглядатись як єдиний ЗП, ємність якого дорівнює сумі ємностей складових, а швидкодія – швидкодії окремого банку.

Блочний принцип побудови оперативної пам'яті має ще одну перевагу – дозволяє зменшити час доступу до інформації. Це стає можливим завдяки потенціальному паралелізму, який є властивим для блочної організації.

Більшій швидкості доступу можна досягти за рахунок одночасного доступу до багатьох банків пам'яті. Одна з використовуваних для цього методик називається розшаруванням пам'яті. В її основу покладене так зване чергування адрес (address interleaving – інтерлів), яке полягає в зміні системи розподілу адрес між банками пам'яті. Причому чергування адрес базується на властивості локальності по зверненню, відповідно до якого послідовний доступ у пам'ять

звичайно виконується до комірок, які мають суміжні адреси. Іншими словами, якщо в даний момент виконується звернення до комірки з адресою 5, то наступне звернення найвірогідніше буде до комірки з адресою 6, а потім 7 і т.д.

Чергування адрес забезпечується за рахунок циклічного розбиття адреси (для вибору банку використовуються два розряди адреси, а інші розряди – для вибору комірки в банку). У пам'яті з інтерлівом кожен наступний рядок вибирається в новому банку, що дозволяє запам'ятовуваним елементам рядка з попереднього банку відновлювати свій стан (рис. 7.3).

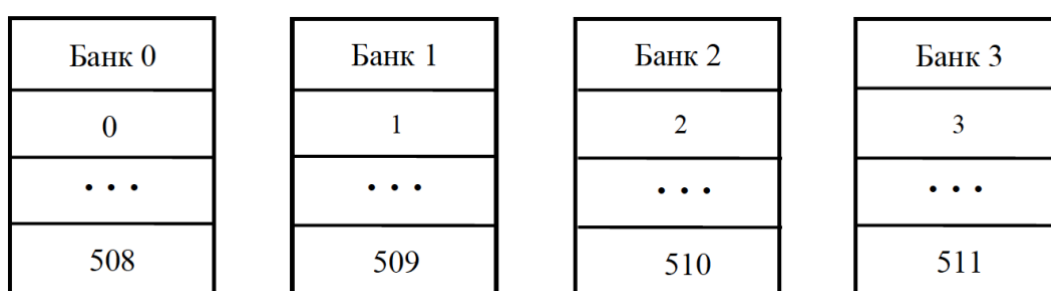


Рисунок 7.3 – Нумерація комірок у банках пам'яті з циклічною адресацією

У разі великої кількості банків середній час доступу до основної пам'яті скорочується в N разів (N - кількість банків), але за умови, що комірки, до яких проводиться звернення, відносяться до різних банків. Якщо ж запити до одного й того ж банку приходять один за одним, то кожний наступний запит повинен очікувати завершення обслуговування попереднього. Така ситуація називається конфліктом по доступу. У разі частого виникнення конфліктів по доступу метод стає неефективним.

У блочно-циклічній схемі розшарування пам'яті кожний банк складається з декількох модулів, що адресуються за круговою схемою (рис. 7.4).

Адреси між банками розподілені по блочній схемі. Таким чином, адреса комірки розбивається на три частини. Старші біти визначають номер банку, наступна група розрядів адреси вказує на комірку в модулі, а молодші біти адреси вибирають модуль у банку.

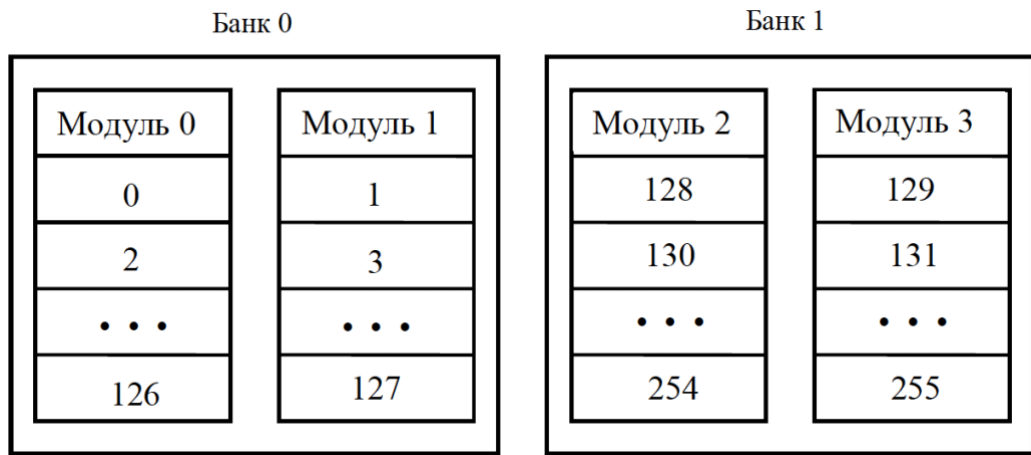


Рисунок 7.4 – Блочно-циклічна схема розшарування пам'яті

У багатопроцесорних системах з загальною пам'яттю, де запити на доступ до пам'яті достатньо незалежні, у систему включають декілька контролерів пам'яті, що дозволяє окремим банкам функціонувати абсолютно автономно.

7.2 Організація мікросхем пам'яті

Як оперативну пам'ять комп'ютера використовують динамічну пам'ять, яка здатна зберігати інформацію тільки протягом достатньо короткого проміжку часу, після якого інформацію необхідно відновлювати, інакше вона буде загублена. Аббревіатура динамічної пам'яті – DRAM (Dynamic Random Access Memory).

Динамічна пам'ять здатна запам'ятовувати біти інформації завдяки паразитній ємності (близько 10-15 Ф). Ця пам'ять сконструйована на базі структури компліментарної технології метал-оксид-напівпровідник (CMOS – Complimentary Metal Oxide Semiconductor). За CMOS-технологією, завдяки її безсумнівним технічним перевагам, будуються сучасні чіпи швидкодіючих електронних елементів з високою щільністю упакування.

Інтегральні мікросхеми (ІМС) пам'яті організовані у вигляді матриці комірок, кожна з яких, залежно від розрядності ІМС, складається з одного або більше запам'ятовуючих елементів (ЗЕ) і має власну адресу. Кожний ЗЕ здатний зберігати один біт інформації.

Популярна сьогодні ІМС пам'яті має об'єм 64 Мбіт – тобто 67108864 запам'ятовуючих елементів (ЗЕ). ЗЕ розміщується на перехрестях сітки з провідників. У цьому випадку кожний ЗЕ є під'єднаним до двох провідників – один з яких використовують для керування читанням (записом), інший для передачі (подачі) даних. Структура запам'ятовуючого елемента зображена на рис. 7.5.

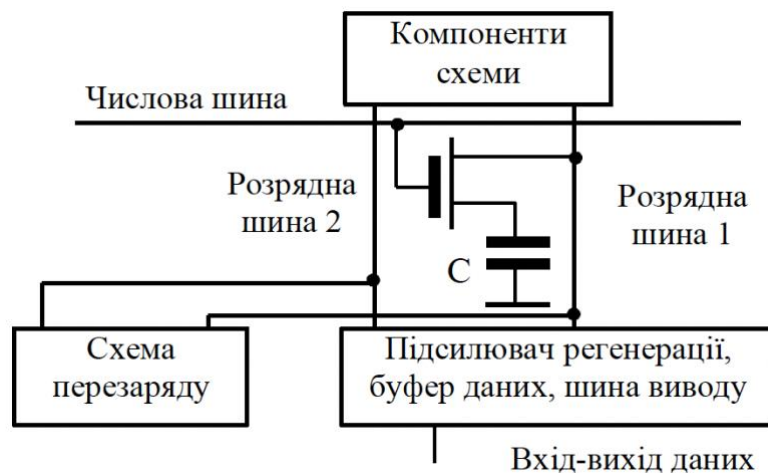


Рисунок 7.5 – Структура запам'ятовуючого елемента

Транзистор у динамічному ЗЕ працює як ключ, який керує передачею заряду. Під час запису в конденсатор біта інформації ключ відкривається, заряджаючи конденсатор до певної величини.

Для доступу до мікросхеми пам'яті з контролера ОЗП надходять сигнали керування, які переводять числову шину в активний стан. При цьому на числовій шині підвищується потенціал, транзистор відкривається і замикає ланцюг: корпус => розрядна шина 1. Якщо ємність заряджена, вона розряджається на розрядну шину, підвищуючи її потенціал.

Між розрядними шинами 1 і 2 виникає напруга. Струм, що при цьому циркулює, створює на вихідній шині заряд (одиниця). Якщо ємність не була зарядженою, то на виході формується струм протилежного напрямку й із шини даних знімається нуль. Процес запису зворотний читанню.

У DRAM кожен ЗЕ можна відшукати за його адресними координатами, що оформлені у рядки і стовпці (рис. 7.6).

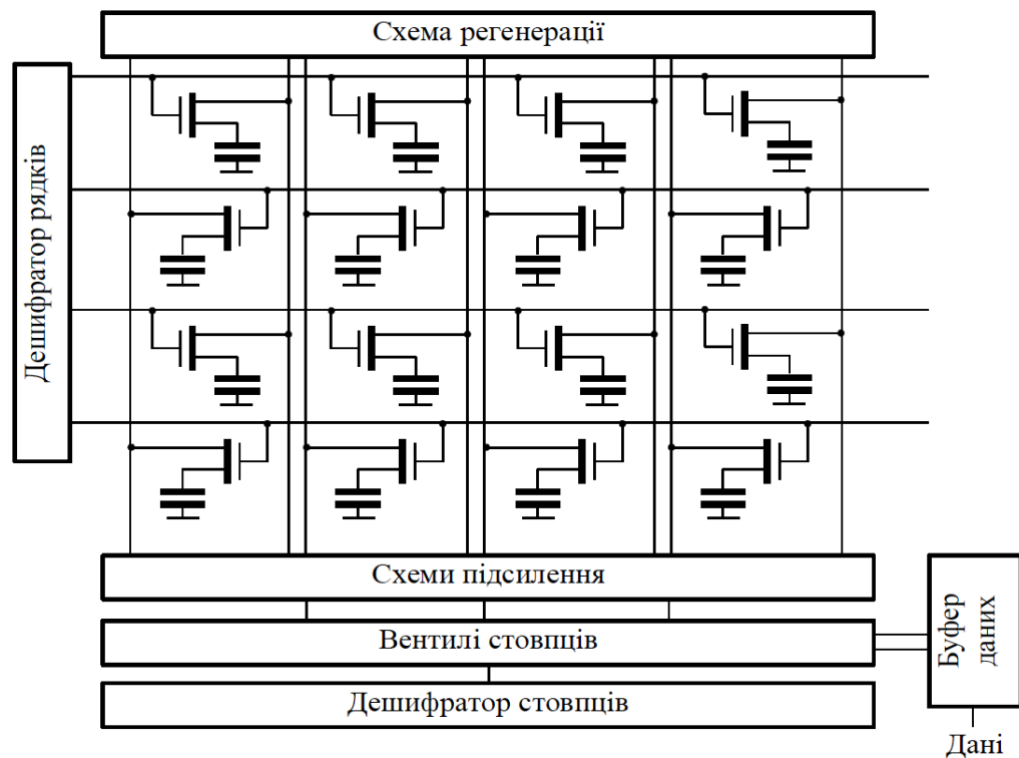


Рисунок 7.6 – Структура динамічного ОЗП

Запам'ятовуючі елементи, які об'єднані керуючим провідником, прийнято називати рядком, або row (розташовані в квадратній таблиці горизонтально). Запам'ятовуючі елементи, які об'єднані провідником, що передає значення, називають стовпцем, або column (розташовані по вертикалі). Вибір відповідної адреси рядка і стовпця дозволяє визначити місце ЗЕ.

Таким чином, під час вибору рядка читання здійснюється відразу на всіх ЗЕ, тобто на кожному із провідників стовпців, виникає напруга, обумовлена логічним значенням відповідного ЗЕ обраного рядка. Описану сукупність ЗЕ і логічні елементи, що їх обрамляють і які зв'язані з вибором рядків і стовпців, називають ядром ІМС.

Усі ЗЕ виводяться на загальну числову шину. Вміст декількох ЗЕ, об'єднаних на виході, утворює інформаційну групу – байт або слово і виводиться на шину даних пам'яті. Розрядність мікросхеми визначає кількість ЗЕ, які мають

одну і ту ж адресу (така сукупність запам'ятовуючих елементів називається коміркою). Розрядність зовнішньої шини даних пам'яті дозволяє підвищити її пропускну здатність.

7.3 Принцип дії динамічної пам'яті

Адреса пам'яті містить відомості для вибору: байта, банку, рядка і стовпця. Вона надходить в один із портів контролера ОЗП, трансформуються в дві адреси – рядка і стовпця, які по шині МА потрапляють у DRAM (рис. 7.7) з деяким проміжком часу (ΔT_1 на рис. 7.8).

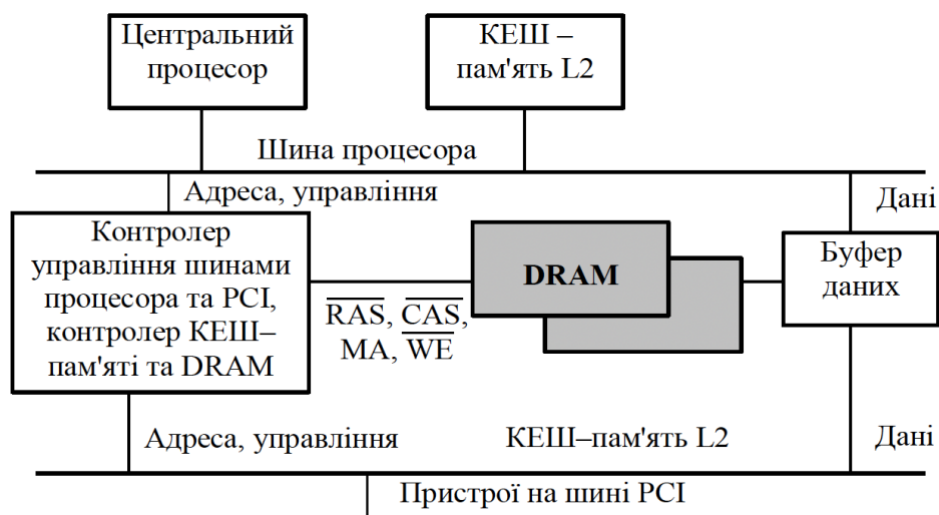


Рисунок 7.7 – Один з варіантів взаємодії з підсистемою пам'яті

Контролер пам'яті оснащений портом для обміну даними з процесором і ще одним портом – для обміну з пристроями вводу/виводу на системній шині.

Оскільки запитів для обміну багато, на вході підсистеми є арбітр. Він під'єднує до пам'яті пристрої відповідно до пріоритетів.

Шина між процесором і контролером ОЗП – FSB (Front Side Bus) тактується системними синхроімпульсами.

Кожний з елементів адресної групи стробується імпульсами сигналів керування RAS (Row Address Strobe) – адреса рядка і CAS (Column Address Strobe) – адреса стовпця (рис. 7.8).

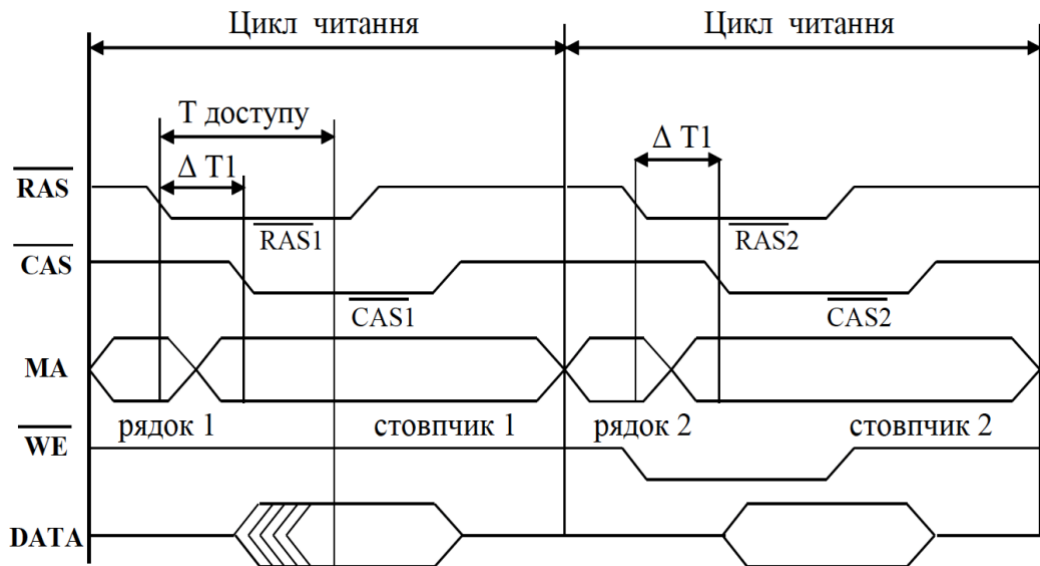


Рисунок 7.8 – Часова діаграма циклів доступу до ОЗП

Щоб стробування було надійним, ці сигнали подаються з затримкою, достатньою для завершення перехідних процесів на шині адреси і в адрес-них ланцюгах мікросхеми. Сигнал вибору мікросхеми CS (Crystal Select) дозволяє роботу ІМС і використовується для вибору певної мікросхеми в системах, які складаються з декількох ІМС. Вхід WE (Write Enable – дозвіл запису) визначає вид операції, яка виконується (зчитування або запис).

На якийсь час, поки ІМС пам'яті не використовує шину даних, інформаційні виходи мікросхеми переводяться в третій (високоімпедансний) стан. Управління перемиканням у третій стан забезпечується сигналом OE (Output Enable – дозвіл видачі вихідних сигналів). Цей сигнал активізується під час виконання операції читання.

Для більшості перерахованих вище управляючих сигналів активним зазвичай вважається їх низький рівень, що показано рискою над позначенням сигналу (рис. 7.8).

Типову процедуру доступу до пам'яті розглянемо на прикладі читання з ІМС з мультиплексуванням адрес рядків і стовпців. Спочатку на вході WE встановлюється рівень, відповідний операції читання, а на адресні контакти ІМС подається адреса рядка, що супроводжується сигналом RAS. По задньому фронту цього сигналу адреса запам'ятовується в регістрі адреси рядка

мікросхеми, після чого дешифрується. Після стабілізації процесів, викликаних сигналом RAS, вибраний рядок підключається до підсилювачів читання/запис (ПЧЗ). Далі на вхід ІМС подається адреса стовпця, яка по задньому фронту сигналу CAS заноситься в реєстр адреси стовпця. Одночасно готується вихідний реєстр даних, куди після стабілізації сигналу CAS завантажується інформація з вибраних ПЧЗ.

Управління операціями з основною пам'яттю здійснюється контролером пам'яті. Зазвичай цей контролер входить до складу центрального процесора або реалізується у вигляді зовнішнього по відношенню до пам'яті пристрою. У останніх типах ІМС пам'яті частина функцій контролера покладається на мікросхему пам'яті. Хоча робота ІМС пам'яті може бути організована як по синхронній, так і по асинхронній схемі, контролер пам'яті – пристрій синхронний, тобто він спрацьовує виключно по тактових імпульсах. З цієї причини операції з пам'яттю прийнято описувати з прив'язкою до тактів.

У разі відсутності даних у кеш, доступ до ОЗП можна уявити так: за час першого і другого тактів синхронізації із шини FSB у контролер ОЗП направляються керуючі й адресні сигнали. Сигнали аналізуються і керують логікою ОЗП.

Два-три (залежно від якості DRAM) синхроімпульси витрачаються на запуск схеми дешифрації і вибір відповідного рядка.

Під час доступу до шин рядків активізується числова шина і всі ЗЕ у даному рядку читаються. На розрядні шини надходять відповідні потенціали від конденсаторів.

На активізацію шин стовпців, під'єднання розрядних шин до буфера даних і витягнення із ЗЕ пам'яті даних також потрібно два-три такти синхронізації. Ще один такт іде на доставку даних у буфер даних DRAM. По такту витрачається на доставку даних у контролер ОЗП і далі – у процесор.

Таким чином, за один цикл звертання до пам'яті система генерує взагалі 9 – 11 тактів синхронізації.

Під час читання даних варто врахувати ще два такти, що витрачаються на відновлення заряду ЗЕ.

Динамічна пам'ять енергозалежна і вимагає періодичного поповнення енергії в паразитних ємностях, що реалізується стандартною процедурою регенерації. Ця апаратна процедура ініціюється інтервальним таймером кожні 15,6 мкс (рис. 7.9) і виконується через канал регенерації динамічної пам'яті (РДП).

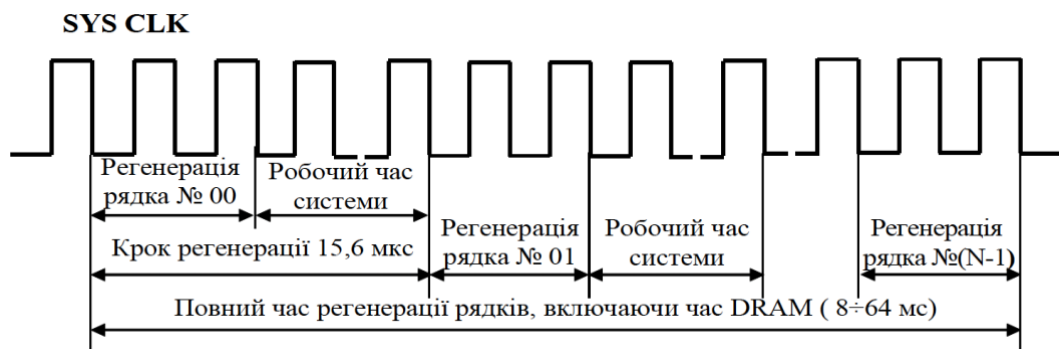


Рисунок 7.9 – Часова діаграма регенерації динамічної пам'яті

Для регенерації використовуються тільки строби RAS, а строби CAS у процесі не беруть участі. Протягом цього часу, який називається кроком регенерації, у DRAM перезаписується цілий рядок ЗЕ. Так, протягом 8...64 мс обновляються всі рядки пам'яті.

Для перезапису ЗЕ ОЗП досить перебирати рядок за рядком і виконувати «фіктивну» (без виведення даних на магістраль даних пам'яті) команду читання.

У цьому випадку кожен ЗЕ рядка перезапишеться через схему перезарядження, а дані не потраплять у буфери виводу даних.

Шина даних знаходиться у високоімпедансному стані. На модифікацію ЗЕ під час читання витрачається два такти синхронізації.

Очевидно, що процедура регенерації пам'яті (у класичному варіанті) «гальмує» роботу системи, оскільки в цей час обмін даними з ОЗП неможливий. Регенерація, заснована на звичайному перебиранні рядків (незалежно від послідовності), у сучасних типах DRAM не застосовується. Існує кілька

економічних варіантів цієї процедури – розширений, пакетний, розподілений і ін.

Тимчасових характеристик динамічної пам'яті дуже багато, але найважливіших – три: час перезарядження пам'яті – являє собою затримку, зв'язану з попереднім зарядженням розрядних шин опорною напругою; час доступу до пам'яті – активізація числової шини, у результаті чого на вихідну шину даних пам'яті викладається інформація; час циклу – складається з затримок часу перезарядження і доступу.

Час затримки виведення даних DRAM вимірюється величинами від десятків до сотень наносекунд. Систему пригальмовують не тільки затримки всередині пам'яті. Будь-яке звертання до ОЗП супроводжується передачею у контролер пам'яті великої групи сигналів, що ускладнюють схемотехніку і підвищують латентність підготовчого періоду циклу обміну даними.