

ТЕМА 3

ЗАГАЛЬНІ ПОНЯТТЯ ПРО АРХІТЕКТУРУ КОМП'ЮТЕРІВ

ПЛАН

- 1 ПОНЯТТЯ АРХІТЕКТУРИ КОМП'ЮТЕРА
- 2 КЛАСИЧНА (НЕЙМАНІВСЬКА) АРХІТЕКТУРА КОМП'ЮТЕРА
- 3 НЕНЕЙМАНІВСЬКІ АРХІТЕКТУРИ КОМП'ЮТЕРІВ

1 ПОНЯТТЯ АРХІТЕКТУРИ КОМП'ЮТЕРА

Обчислювальна машина (ОМ) - комплекс технічних і програмних засобів, призначений для автоматизації підготовки і рішення завдань користувачів.

Обчислювальна система (ОС) - сукупність взаємозв'язаних і взаємодіючих процесорів або обчислювальних машин, периферійного устаткування і програмного забезпечення, що призначена для підготовки і рішення завдань користувачів.

Таким чином, формальна відмінність ОС від ОМ виражається у кількості обчислювачів. Множинність обчислювачів дозволяє реалізувати у ОС паралельну обробку. З іншого боку, сучасні обчислювальні машини з одним процесором також мають певні засоби розпаралелювання обчислювального процесу. Іншими словами, грань між ОМ і ОС часто буває дуже розпливчатою, що доцільно, розглядати ОМ як одну з реалізацій ОС. І навпаки, обчислювальні системи часто будуються з традиційних ОМ і процесорів, тому багато з положень, що відносяться до ОМ, можуть бути поширені і на ОС.

1 ПОНЯТТЯ АРХІТЕКТУРИ КОМП'ЮТЕРА

Під архітектурою обчислювальної машини зазвичай розуміється логічна побудова ОМ, тобто те, якою машина представляється програмістові. Уперше термін «архітектура обчислювальної машини» (computer architecture) був ужитий фірмою ІВМ при розробці машин сімейства ІВМ 360 для опису тих засобів, якими може користуватися програміст, складаючи програму на рівні машинних команд. Подібне трактування називають «вузьким», і охоплює воно перелік і формат команд, форми представлення даних, механізми введення/виводу, способи адресації пам'яті і т. п. З розгляду випадають питання фізичної побудови обчислювальних засобів: склад пристроїв, число регістрів процесора, місткість пам'яті, наявність спеціального блоку для обробки дійсних чисел, тактова частота центрального процесора і т. д. Цей круг питань прийнято визначати поняттям організація або структурна організація.

Для вивчення нашої дисципліни таке ділення не дає яких-небудь переваг, то надалі користуватимемося терміном «архітектура», правда, в «широкому» його тлумаченні, що об'єднує як архітектуру у вузькому сенсі, так і організацію ОМ.

1 ПОНЯТТЯ АРХІТЕКТУРИ КОМП'ЮТЕРА

Обчислювальна машина як закінчений об'єкт являє собою плід зусиль фахівців в найрізноманітніших галузях людських знань. Кожен фахівець розглядає обчислювальну машину з позицій завдання, що стоїть перед ним, абстрагуючись від несуттєвих, на його думку, деталей. У таблиці 1 перераховані фахівці, що беруть участь в створенні ОМ, і круг питань, що входять в їх компетенцію.

Таблиця 1 - Розподіл функцій між розробниками обчислювальної машини

Фахівець	Коло питань
Виробник напівпровідникових матеріалів	Матеріал для інтегральних мікросхем (легований кремній, діоксид кремнію і т. п.)
Розробник електронних схем	Електронні схеми вузлів ОМ (розробка і аналіз)
Розробник інтегральних мікросхем	Надвеликі інтегральні мікросхеми (схеми електронних елементів, їх розміщення на кристалі)
Системний архітектор	Архітектура і організація обчислювальної машини (пристрої і вузли, система команд і т. п.)
Системний програміст	Операційна система, компілятори
Теоретик	Алгоритми, абстрактні структури даних

Коло питань, що розглядаються в цьому курсі, здебільшого відноситься до компетенції системного архітектора.

2 КЛАСИЧНА (НЕЙМАНІВСЬКА) АРХІТЕКТУРА КОМП'ЮТЕРА

У **1946** році працюючий у той час в Англії угорський математик **Джон фон Нейман** (із співавторами) описав в технічній доповіді архітектуру деякого абстрактного обчислювача, який зараз прийнято називати **машиною фон Неймана**. Ця машина є абстрактною моделлю ЕОМ, проте, ця абстракція відрізняється від абстрактних виконавців алгоритмів. Машина фон Неймана не піддається реалізації з іншої причини: багато деталей в архітектурі цього виконавця алгоритму не конкретизовано. Це було зроблено спеціально, щоб не сковувати творчого підходу до справи у інженерів-розробників нових ЕОМ. Можна сказати, що машина фон Неймана розглядається не на внутрішньому, а тільки на концептуальному рівні бачення архітектури.



2 КЛАСИЧНА (НЕЙМАНІВСЬКА) АРХІТЕКТУРА КОМП'ЮТЕРА

На рисунку 1 приведена схема машини фон Неймана. Товстими (подвійними) стрілками показані потоки команд і даних, а тонкими - передача між окремими пристроями комп'ютера управляючих та інформаційних сигналів. Як ми незабаром побачимо, виконання кожної команди призводить до вироблення послідовності управляючих сигналів, які примушують вузли комп'ютера здійснювати ті або інші дії. За допомогою інформаційних сигналів одні вузли комп'ютера повідомляють інші вузли про те, що вони виконали дії, наказані управляючими сигналами, або зафіксували помилку у своїй роботі.

Як видно з приведеного рисунка, машина фон Неймана складається з пам'яті, пристроїв вводу/виводу і центрального процесора (ЦП). Центральний процесор, у свою чергу, складається з пристроєм управління (ПУ) і арифметико-логічного пристроєм (АЛП).

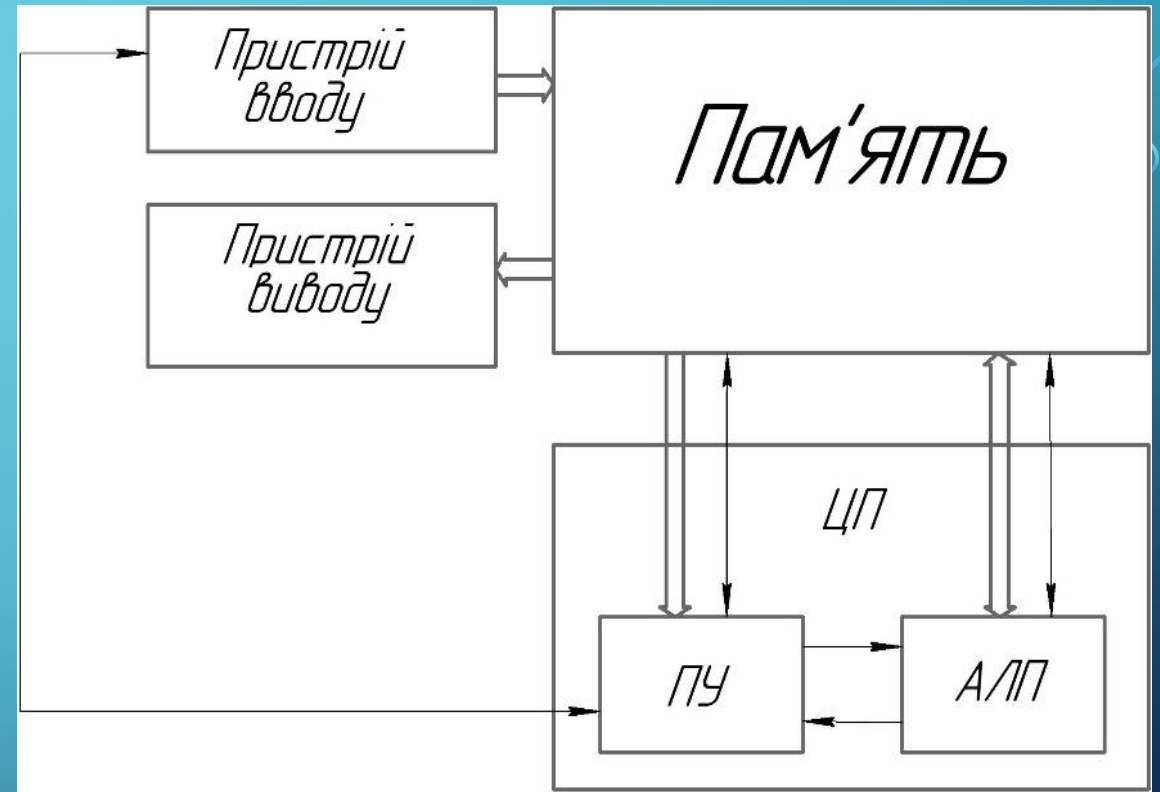


Рисунок 1 – Схема машини фон Неймана

2 КЛАСИЧНА (НЕЙМАНІВСЬКА) АРХІТЕКТУРА КОМП'ЮТЕРА

Основоположні властивості архітектури машини фон Неймана будуть сформульовані у вигляді **принципів фон Неймана**. Ці принципи багато років визначали основні риси архітектури декількох поколінь ЕОМ.

а) Принцип лінійності і однорідності пам'яті. Пам'ять машини фон Неймана - це лінійна (впорядкована) і однорідна послідовність деяких елементів, що називаються **комірками**. У будь-який елемент пам'яті інші пристрої машини (по товстих стрілках на схемі рисунка 1) можуть записувати і зчитувати інформацію, причому час читання з будь-якої комірки однаковий для усіх комірок. Час запису у будь-яку комірку теж однаковий (це і є принцип однорідності пам'яті). Зрозуміло, що час читання з елементу пам'яті може не співпадати з часом запису в неї. Така пам'ять в сучасних комп'ютерах називається **пам'яттю з довільним доступом (Random Access Memory - RAM)**. На практиці багато сучасних ЕОМ можуть мати ділянки пам'яті різних видів. Наприклад, деякі області пам'яті підтримують тільки читання інформації (по-англійськи ця пам'ять називається **Read Only Memory, ROM**), дані в таку пам'ять записуються один раз при виготовленні цієї пам'яті. Інші області пам'яті можуть допускати запис, але за значно більший час, чим в звичайну пам'ять (це так звана **напівпостійна пам'ять, такою пам'яттю комплектуються популярні нині карти флеш-пам'яті**) та ін.

2 КЛАСИЧНА (НЕЙМАНІВСЬКА) АРХІТЕКТУРА КОМП'ЮТЕРА

Елементи пам'яті в машині фон Неймана нумеруються від нуля до деякого додатного числа N (це і означає, що пам'ять лінійна).

Адресою комірки називається її номер.

Кожна комірка складається з більш дрібних частин, що іменуються розрядами і нумеруються також від нуля і до певного числа. Кількість розрядів в комірці визначає розрядність пам'яті. Кожен розряд може зберігати одну цифру в деякій системі числення. У більшості ЕОМ використовується двійкова система числення, оскільки це вигідніше з точки зору апаратної реалізації. В цьому випадку кожен розряд зберігає одну двійкову цифру або один біт інформації. Сам фон Нейман теж був прибічником використання двійкової системи числення, що дозволяло добре описувати архітектуру вузлів ЕОМ за допомогою логічних (булевих) виразів.

Вміст комірки називається машинним словом. З точки зору архітектури, машинне слово - це мінімальний об'єм даних, яким можуть обмінюватися між собою різні вузли машини по товстих стрілках на схемі. З кожного елементу пам'яті можна зчитати копію машинного слова і передати її в інший пристрій комп'ютера, при цьому оригінал не міняється. При записі в пам'ять старий вміст комірки зникає і замінюється новим машинним словом.

2 КЛАСИЧНА (НЕЙМАНІВСЬКА) АРХІТЕКТУРА КОМП'ЮТЕРА

Помітимо, що на практиці рішення задачі збереження початкового машинного слова при читанні з комірки для деяких видів пам'яті є нетривіальним і досить трудомістким. Річ у тому, що в цій пам'яті (вона називається **динамічною пам'яттю**) при читанні оригінал руйнується, і його доводиться кожного разу відновлювати після читання даних. Крім того, дані, що зберігаються в динамічній пам'яті, руйнуються і самі по собі з часом, (вони втрачають електричний заряд), тому доводиться часто (через кожні декілька мілісекунд) відновлювати вміст цієї пам'яті.

У комп'ютерах може використовуватися і інший вид пам'яті, яка називається **статичною пам'яттю**, при читанні з неї і при зберіганні дані не руйнуються (поки подається електрична напруга). Статична пам'ять, в порівнянні з динамічною, працює швидше, проте вона дорожча і вимагає при реалізації більше електронних схем.

2 КЛАСИЧНА (НЕЙМАНІВСЬКА) АРХІТЕКТУРА КОМП'ЮТЕРА

б) Принцип нероздільності команд і даних. З точки зору програміста машинне слово є або командою, або даним, що підлягає обробці (це число, символна інформація, елемент зображення і так далі). Скорочено надалі називатимемо таку інформацію «числами». **Цей принцип Фон Неймана полягає в тому, що числа і команди невідмітні одне від одного - в пам'яті і ті і інші представляються деяким набором розрядів, причому за зовнішнім виглядом машинного слова не можна визначити, що воно є - команда або число.**

З нероздільності команд і даних витікає очевидне слідство – **в) Принцип зберігання програми.** Цей принцип є дуже важливим, його суть полягає в тому, що програма зберігається в пам'яті разом з числами. Щоб зрозуміти важливість цього принципу розглянемо, а як програми взагалі з'являються в пам'яті машини. Зрозуміло, що, по-перше, команди програми можуть, нарівні з числами, вводиться в пам'ять за допомогою пристрою введення. А тепер згадаємо, що на вхід алгоритму можна в якості вхідних даних подавати запис деякого іншого алгоритму (зокрема, свій власний запис). Залишається зробити останній крок в цих міркуваннях і зрозуміти, що вихідними даними алгоритму теж може бути запис деякого іншого алгоритму. Таким чином, одна програма може в якості результату своєї роботи помістити в пам'ять комп'ютера іншу програму. Саме так і працюють

2 КЛАСИЧНА (НЕЙМАНІВСЬКА) АРХІТЕКТУРА КОМП'ЮТЕРА

б) Принцип нероздільності команд і даних. З точки зору програміста машинне слово є або командою, або даним, що підлягає обробці (це число, символна інформація, елемент зображення і так далі). Скорочено надалі називатимемо таку інформацію «числами». **Цей принцип Фон Неймана** полягає в тому, що числа і команди невідмітні одне від одного - в пам'яті і ті і інші представляються деяким набором розрядів, причому за зовнішнім виглядом машинного слова не можна визначити, що воно є - команда або число.

З нероздільності команд і даних витікає очевидне слідство – **в) Принцип зберігання програми.** Цей принцип є дуже важливим, його суть полягає в тому, що програма зберігається в пам'яті разом з числами.

Помітимо також, що, коли Джон фон Нейман (із співавторами) писав цю свою роботу, більшість тодішніх ЕОМ зберігали програму в пам'яті одного виду, а числа - в пам'яті другого виду, тому цей принцип був революційним. У сучасних ЕОМ і програми, і дані, як правило, зберігаються в одній і тій же пам'яті.

2 КЛАСИЧНА (НЕЙМАНІВСЬКА) АРХІТЕКТУРА КОМП'ЮТЕРА

г) Принцип автоматичної роботи або принцип програмного управління. Машина, виконуючи записану в її пам'яті програму, функціонує автоматично, без участі людини. Програма - набір записаних в пам'яті машинних команд, що описують кроки роботи алгоритму. Команди програми обробляють збережені в пам'яті комп'ютера дані. Таким чином, програма - це запис алгоритму на мові машини. Мова машини - набір усіх можливих її команд. Наприклад, машинна мова сучасних найбільш поширених комп'ютерів фірми Intel містить більше трьохсот команд.

д) Принцип послідовного виконання команд. Пристрій управління виконує деяку команду від початку до кінця, а потім за певним правилом вибирає наступну команду для виконання, потім наступну і так далі. При цьому кожна команда або сама явно вказує на команду, яка виконуватиметься наступною (такі команди називаються командами переходу), або наступною виконуватиметься команда з комірки, розташованої в пам'яті безпосередньо за тою коміркою, в якій зберігається тільки що виконана команда. Цей процес продовжується, поки не буде виконана спеціальна команда зупинки, або при виконанні чергової команди не виникне аварійна ситуація (наприклад, ділення на нуль).

3 НЕНЕЙМАНІВСЬКІ АРХІТЕКТУРИ КОМП'ЮТЕРІВ

До цього часу більшість універсальних комп'ютерів будують за принципами архітектури Джона фон Неймана. Але недоліки цієї архітектури, пов'язані з закладеним в ній послідовним характером організації обчислень, ставлять перепони в пошуку шляхів побудови швидких комп'ютерних систем. Тому крім найманівської архітектури за час існування комп'ютерної техніки було створено цілий ряд інших архітектур.

а) Гарвардська архітектура. Вперше була реалізована Ховардом Айкеном в комп'ютері Марк-1 в Гарварді. Вона передбачає розділення пам'яті на пам'ять даних і пам'ять команд. Тим самим розділяються шини передачі керуючої і оброблюваної інформації, рисунку 2. При цьому підвищується продуктивність комп'ютера за рахунок суміщення в часі пересилання та обробки даних і команд.

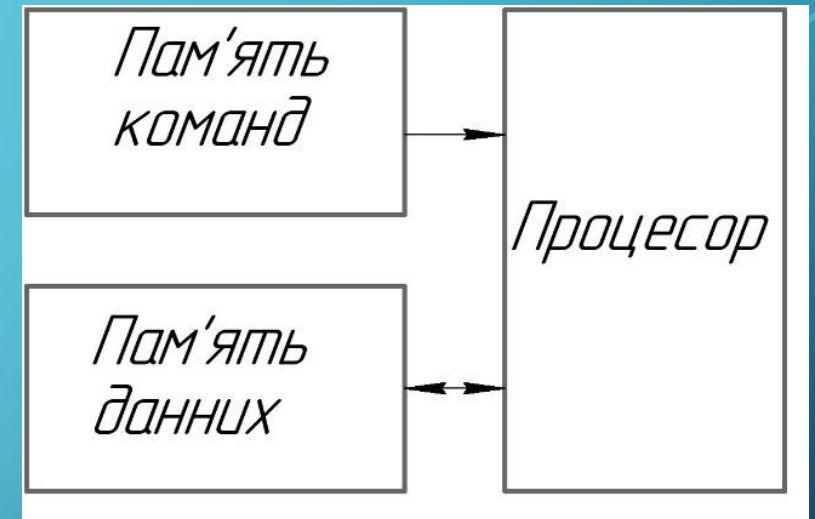


Рисунок 2 – Гарвардська архітектура

3 НЕНЕЙМАНІВСЬКІ АРХІТЕКТУРИ КОМП'ЮТЕРІВ

Необхідність використання гарвардської архітектури можна пояснити так. Ядро комп'ютера нейманівської архітектури складається з процесора та основної пам'яті. Бажано, аби обидві компоненти ядра не пригальмовували одна одну, тобто працювали із рівною швидкістю. На практиці вузол пам'яті є значно (на порядок) повільнішим від процесора і цей розрив у швидкодії з прогресом інтегральних технологій лише зростає. Зменшити розрив можна структурними методами, збільшуючи розрядність інформаційного слова пам'яті. Саме цей підхід реалізує гарвардська архітектура з двома запам'ятовувальними пристроями. Зрозуміло, що тут паралельно виконуються операції вибирання команд програми, з одного боку, а з другого — вибирання та запис кодів даних і результатів обчислень.

3 НЕНЕЙМАНІВСЬКІ АРХІТЕКТУРИ КОМП'ЮТЕРІВ

б) Дуальна неймано-гарвардська архітектура. Швидкі комп'ютери гарвардської архітектури є складнішими щодо програмування порівняно з комп'ютерами нейманської архітектури. Зрозуміло, що бажано створити комп'ютер з дуальною архітектурою, яка водночас запозичує нову якість — швидкодію від гарвардської архітектури та стандартну парадигму розробки програм від нейманської архітектури. Злиття двох архітектур виконують на рівні кеш пам'яті шляхом її поділу на кеш даних та кеш команд, рисунок 3.

Злиттям архітектур програмісту надано зручність програмних технологій нейманської архітектури, а з боку процесора реалізовано гарвардську архітектуру, в результаті чого він значно менше пригальмовується з боку основної пам'яті.

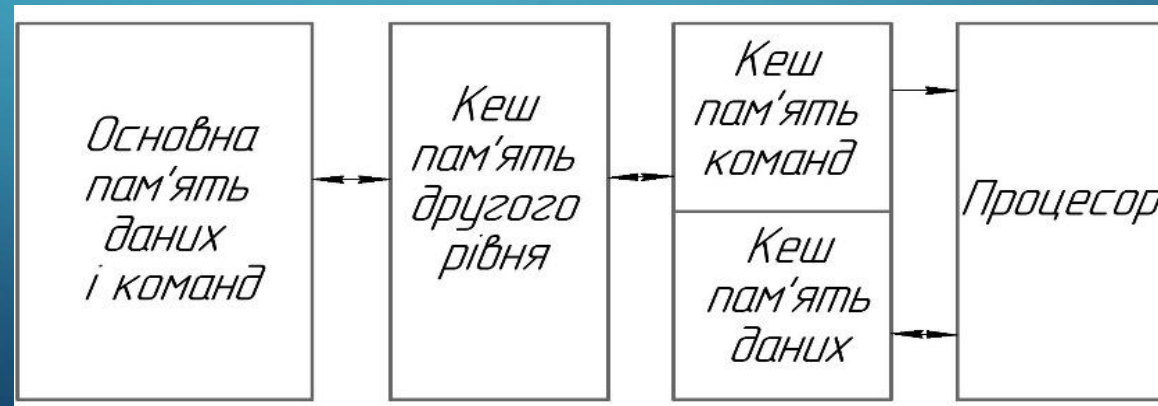


Рисунок 3 - Дуальна неймано-гарвардська архітектура

3 НЕНЕЙМАНІВСЬКІ АРХІТЕКТУРИ КОМП'ЮТЕРІВ

в) Паралельні комп'ютерні архітектури. Паралельні обчислювальні системи — комп'ютерні системи, що реалізують тим або іншим способом паралельну обробку даних на багатьох обчислювальних вузлах для підвищення загальної швидкості розрахунку. Ідея розпаралелювання обчислень базується на тому, що більшість завдань можуть бути розділені на набір менших завдань, які можуть бути вирішені одночасно. Зазвичай паралельні обчислення вимагають координації дій. Якщо при обчисленні не застосовуються циклічні (що повторюються) дії, то N обчислювальних модулів ніколи не виконають роботу в N разів швидше, ніж один єдиний обчислювальний модуль.

Наприклад, для швидкого сортування масиву на двопроцесорній машині можна розділити масив навпіл і сортувати кожну половину на окремому процесорі. Сортування кожної половини може зайняти різний час, тому необхідна синхронізація.

Однак і паралельні комп'ютерні системи мають обмеження. По-перше, зі збільшенням кількості процесорів ускладнюється задача розподілу завдань між процесорами. Для її вирішення використовуються додаткові процесори, кількість яких може значно перевищувати кількість процесорів, зайнятих безпосередньо виконанням алгоритму. По-друге, послідовна природа багатьох алгоритмів

УЗАГАЛЬНЕННЯ ВИВЧЕНОГО МАТЕРІАЛУ:

Усно дайте відповіді на питання.

- 1 Поясніть зміст поняття «Обчислювальна машина».
- 2 Поясніть зміст поняття «Обчислювальна система».
- 3 Назвіть рівні деталізації поняття «Обчислювальна машина».
- 4 Поясніть принципи будови машини фон Неймана.
- 5 В чому полягає принцип лінійності і однорідності пам'яті?
- 6 В чому полягає принцип нероздільності команд і даних?
- 7 В чому полягає принцип програмного управління?
- 8 В чому полягає принцип послідовного виконання команд?
- 9 Поясніть особливості побудови гарвардської архітектури.
- 10 Поясніть особливості побудови дуальної неймано-гарвардської архітектури.

ДОМАШНЄ ЗАВДАННЯ

Усно дати відповіді на контрольні запитання в розділі **Узагальнення вивченого матеріалу**
Пройти **Тест до теми 3 у Moodle.**