

## 2. РОЗПОДІЛЕНЕ СЕРЕДОВИЩЕ

### 2.1. Концепції апаратних рішень

Незважаючи на те, що всі розподілені системи містять по декілька процесорів, є різні способи їх об'єднання в систему. Особливо це стосується варіантів їх з'єднання й організації взаємного обміну.

Розглянемо коротко апаратне забезпечення розподілених систем, зокрема варіанти з'єднання машин між собою. За минулі роки було запропоновано різні класифікації комп'ютерних систем з декількома процесорами, але жодна з них не стала дійсно популярною й поширеною.

Розглянемо системи, до яких входять незалежні комп'ютери, які можна поділити на дві групи. Системи, у яких комп'ютери використовують *пам'ять спільно*, називають **мультипроцесорами (multiprocessors)**, а ті, що *працюють кожен зі своєю пам'яттю* – **мультикомп'ютерами (multicomputers)**. Основною відмінністю є те, що мультипроцесори мають *єдиний адресний простір*, який спільно використовують усі процесори. У мультикомп'ютерах кожна машина використовує свою власну пам'ять.

Кожну із цих категорій поділяють на додаткові підкатегорії за архітектурою мережі, яка їх з'єднує. Найчастіше використовують дві архітектури – *шинну (bus)* і *комутовану (switched)*. Під **шинною** розуміють *плату, шину, кабель або інше середовище*, що з'єднує всі машини між собою. **Комутовані системи**, на відміну від шинних, не мають єдиної магістралі, такої як, наприклад, у кабельного телебачення. Замість неї від машини до машини тягнуться окремі канали, виконані із застосуванням різних технологій зв'язку. Повідомлення передаються каналами з узгодженням рішення про комутацію з конкретним вихідним каналом для кожного з них.

Також розподілені комп'ютерні системи поділяють на *гомогенні (homogeneous)* і *гетерогенні (heterogeneous)*. Цей поділ стосується винятково мультикомп'ютерних систем. Для гомогенних мультикомп'ютерних систем характерна одна мережа, яка з'єднує комп'ютери та використовує єдину технологію. Однакові також усі процесори, які зазвичай мають доступ до одна-

кових обсягів власної пам'яті. Гомогенні мультикомп'ютерні системи нерідко використовують як паралельні (працюють з одним завданням), за аналогією до мультипроцесорних.

На відміну від гомогенних, гетерогенні мультикомп'ютерні системи можуть містити декілька незалежних комп'ютерів, з'єднаних різноманітними мережами.

**Приклад.** Розподілена комп'ютерна система може бути побудована з декількох локальних комп'ютерних мереж, з'єднаних комутованою магістраллю FDDI або ATM.

Мультипроцесорні системи мають одну характерну рису: всі процесори мають прямий доступ до загальної пам'яті. Мультипроцесорні системи шинної архітектури складаються з деякої кількості процесорів, приєднаних до загальної шини, а через неї – до модулів пам'яті. Найпростіша конфігурація містить плату із шиною або материнською платою, у яку вставляють процесори й модулі пам'яті. Проблема такої схеми полягає в тому, що у разі наявності 4-х або 5-ти процесорів шина стає перевантаженою і продуктивність різко знижується. Щоб розв'язати цю проблему, потрібно розмістити між процесором і шиною високошвидкісну кеш-пам'ять (cache memory).

На відміну від мультипроцесорів, побудувати гомогенну мультикомп'ютерну систему порівняно нескладно. Кожний процесор прямо зв'язаний зі своєю локальною пам'яттю, а єдиною проблемою є взаємодія процесорів між собою, за таких умов необхідна схема з'єднання, але оскільки у такому разі ключовим моментом є тільки наявність зв'язку між процесорами, то обсяг трафіку буде на кілька порядків нижчим, ніж у випадку використання мережі для підтримки трафіку між процесорами й пам'яттю.

Найбільшу кількість наявних нині розподілених систем побудовано за схемою гетерогенних мультикомп'ютерних систем. Це означає, що комп'ютери, які є частинами цієї системи, можуть відрізнятися, наприклад за типом процесора, розміром пам'яті й продуктивністю каналів введення-виведення. На практиці роль деяких із цих комп'ютерів можуть виконувати високопродуктивні паралельні системи, зокрема мультипроцесорні або гомогенні мультикомп'ютерні. Мережа, яка їх поєднує, може бути дуже неоднорідною. При-

кладом гетерогенності є створення великих мультикомп'ютерних систем, які використовують наявні мережі й канали.

**Приклад.** Нині є фактом існування кампусних університетських розподілених систем, що складаються з локальних мереж різних факультетів, з'єднаних між собою високошвидкісними каналами. У глобальних системах різні станції можуть, у свою чергу, з'єднуватися загальнодоступними мережами, наприклад мережними службами, пропонуваними комерційними операторами зв'язку, зокрема ATM або Frame relay.

## **2.2. Архітектура багатопроцесорних систем**

Серед усієї різноманітності багатопроцесорних систем розглянемо системи типу MIMD (множинні потоки команд і множинні потоки даних). Обчислювальні системи типу MIMD являють собою об'єднання процесорів, на кожному з яких може бути запущена своя, відмінна від інших, програма, що оперує власними даними. З огляду на це MIMD системи можна поділити на два класи: *зі спільною пам'яттю і з роздільною пам'яттю*.

### **2.2.1. Системи зі спільною пам'яттю**

У системах зі спільною пам'яттю програма поділена на процеси, які взаємодіють і автоматично розподіляються між доступними процесорами системи. Для систем цього класу є засоби автоматизованої побудови паралельних програм – векторизуючі компілятори. Достатньо широке коло послідовних алгоритмів може бути успішно адаптовано до функціонування на системах зі спільною або з розподільною пам'яттю (рис. 2.1). Система зі спільною пам'яттю зручна тому, що функціонує під єдиною копією операційної системи й не вимагає індивідуального налаштування кожного процесорного вузла. Такі комп'ютери, керовані Unix або подібними операційними системами, обслуговують користувачів у режимі колективного користування, автоматично розподіляючи потік завдань прикладного програмного забезпечення процесорами без участі адміністратора або оператора.

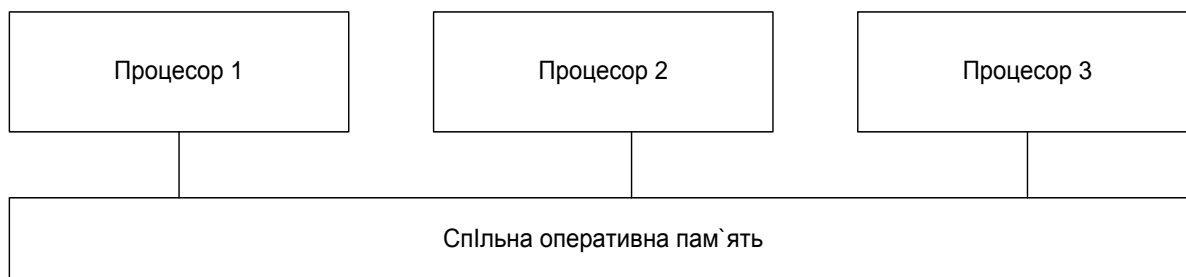


Рис. 2.1. Багатопроцесорна система зі спільною пам'яттю

Однак системам зі спільною пам'яттю властива низка істотних недоліків:

- порівняно невелика кількість процесорів;
- немає можливості збільшення кількості процесорів, тобто масштабувати систему;
- пікова продуктивність найбільших систем зі спільною пам'яттю нижча від пікової продуктивності найбільших систем з роздільною пам'яттю;
- висока вартість порівняно з аналогічними за продуктивністю системами з роздільною пам'яттю.

Ефективно об'єднати більше 20–30 процесорів на основі загальної пам'яті важко, оскільки кожний процесор повинен мати фізичний доступ до кожного з блоків оперативної пам'яті.

**Приклад.** У разі використання роздільних 32-розрядної адресної шини й 64-розрядної шини даних потрібною є як мінімум 96-розрядна високошвидкісна лінія доступу до пам'яті від кожного процесора, що становить технологічну проблему.

Необхідність забезпечення обміну зі швидкістю 500 Мбайт/с і вище обмежується фізичною відстанню від кожного процесора до кожного блока пам'яті. Зазвичай системи зі спільною пам'яттю виглядають досить компактно, розміщуючись в одному корпусі. Цей фактор є обмеженням під час збільшення загальної кількості процесорів через кількість блоків пам'яті, які можна розмістити в межах, заданих максимальною відстанню від блока пам'яті до процесора. Спільна пам'ять передбачає можливість усіх процесорів одночасно прочитати або записати різні дані з одного блока пам'яті, для цього кожний блок пам'яті повинен мати таку кількість точок входу, яка дорівнює кількості процесорів.

Як приклад можливого розв'язку цієї проблеми наведемо схему мережі типу «метелик» (рис. 2.2). Згідно із цією схемою вісім процесорних модулів

можуть отримати доступ до восьми блоків пам'яті через систему комутаторів, однак у кожен конкретний момент до кожного блока пам'яті може мати доступ тільки один із процесорів, що в цілому знижує продуктивність системи.

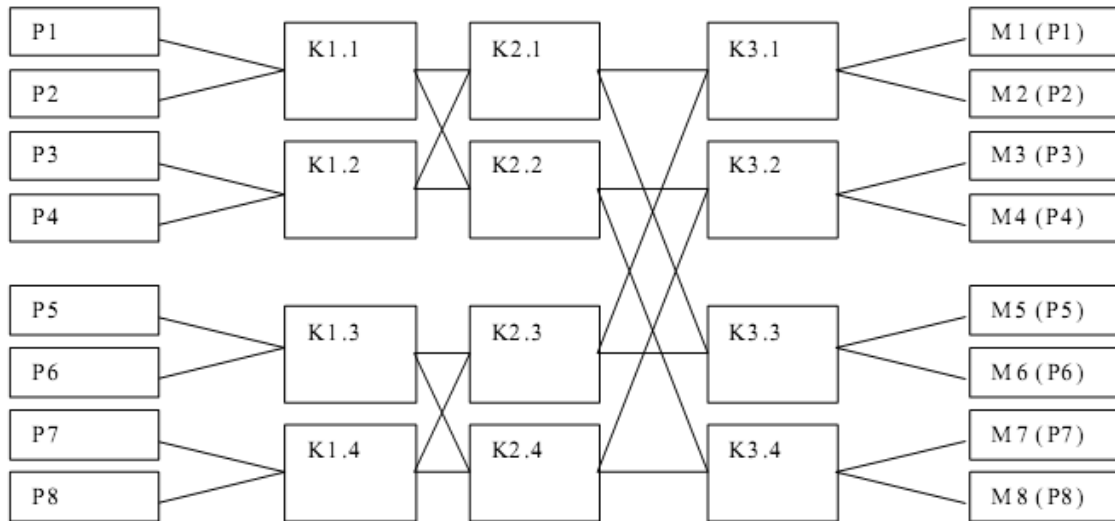


Рис. 2.2. Комунікаційна мережа типу «метелик»

Деякі сучасні системи побудовано за кластерним принципом, за яким процесори й оперативну пам'ять поділяють на кілька груп – **кластерів**. Усередині кластера процесори мають швидкий доступ до оперативної пам'яті. Доступ процесорів одного кластера до оперативної пам'яті, розташованої в іншому кластері, так само можливий, але час доступу при цьому значно зростає.

Конфлікти, що неминуче виникають під час запису різними процесорами тих самих даних, мають фундаментальний характер і знижують продуктивність системи за рахунок утрат на синхронізацію, незалежно від конкретної апаратної реалізації системи. У такий спосіб сконструйована система зазвичай не передбачає можливості суттєвого збільшення кількості процесорних вузлів і зумовлює високу вартість.

### **2.2.2. Системи з роздільною пам'яттю**

Масштабовані системи масового паралелізму конструюють на основі об'єднання каналами передачі даних процесорних вузлів, які мають свою локальну оперативну пам'ять, недоступну іншому процесору (рис. 2.3).

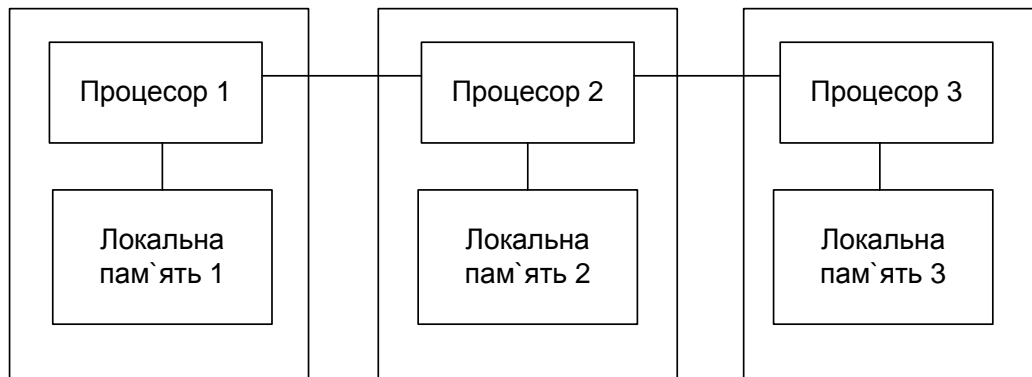


Рис. 2.3. Багатопроцесорна система з роздільною пам'яттю

Обмін даними між процесорами у разі такого підходу можливий лише за допомогою повідомлень, переданих каналами зв'язку. Така схема має низку переваг порівняно із системами, побудованими на основі загальної спільної пам'яті. Зазначимо основні з них:

- порівняно низька вартість – найменший показник серед показників «ціна/продуктивність»;
- масштабованість – можливість побудови систем необхідної продуктивності, зазвичай підвищення їх потужності за рахунок установки додаткових процесорів.

Системи з роздільною пам'яттю лідирують за показником пікової продуктивності, оскільки будь-які нові однопроцесорні (або багатопроцесорні на основі пам'яті) системи можуть бути легко об'єднані мережею й використані як багатопроцесорні комплекси з роздільною пам'яттю. Проте, на жаль, ефективне використання систем з розподіленою пам'яттю вимагає значних зусиль з боку розробників прикладного забезпечення й можливе далеко не для всіх типів завдань, оскільки не для всіх послідовних алгоритмів, що добре зарекомендували себе, вдається побудувати ефективні паралельні аналоги.

**Трансп'ютери.** Саме від трансп'ютерів почалося масове поширення багатопроцесорних систем. Типова трансп'ютерна система є фактично паралельним обчислювальним прискорювачем для певного комп'ютера загального призначення – хостом комп'ютера (HOST). Функції хост-систем виконують також робочі станції типу Sun, персональні комп'ютери IBM PC (рис. 2.4).

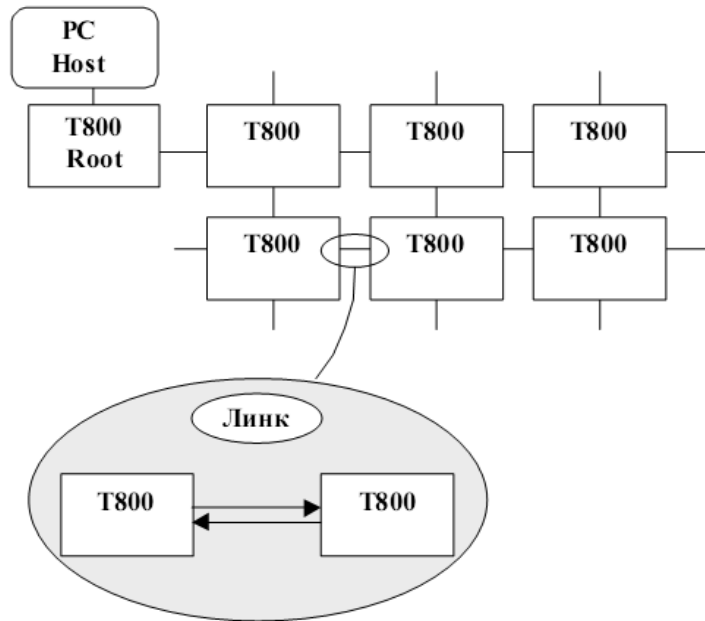


Рис. 2.4. Трансп'ютерна система й канали зв'язку

Кожний трансп'ютер виробництва фірми INMOS (Великобританія) є повнофункціональним процесором (рис. 2.5), який відрізняється від звичайного процесора тим, що містить чотири канали міжпроцесорного обміну даними – лінка (*link*).

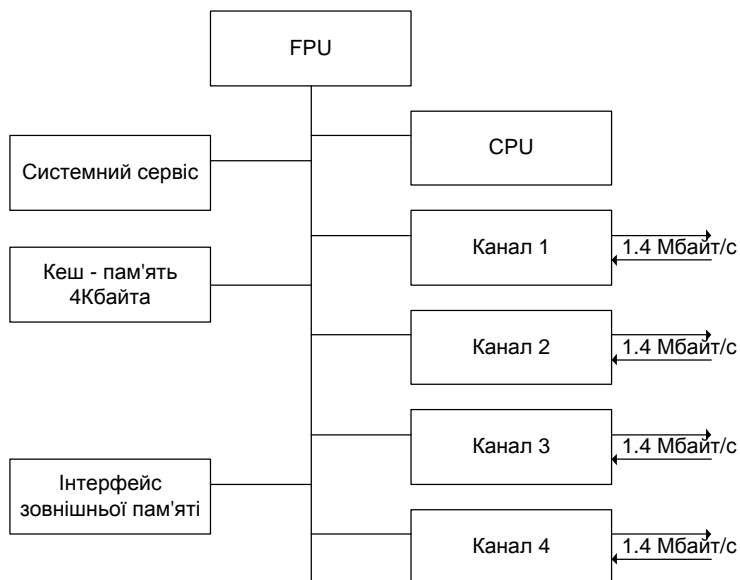


Рис. 2.5. Структура трансп'ютера типу Т-800

Кожний *link* – це обладнання синхронного небуферизованого послідовного обміну даними 4-провідною лінією зв'язку (порівняйте з 96-провідною

лінією доступу до оперативної пам'яті в системах зі спільною пам'яттю). Саме завдяки можливості з'єднувати між собою довільну кількість процесорів, керованих однією або декількома хост-системами, трансп'ютерні системи набули значного поширення.

Трансп'ютер типу T800 містить обладнання обробки цілих чисел (CPU), дійсних чисел і чотири двонаправлені канали передачі даних – лінка. Кожна з названих шістьох апаратних засобів може працювати одночасно з іншими й незалежно від них. Виконуючи запити на обробку або передачу даних, кожне з них звертається до змінних, розташованих у доступній усім спільній оперативній пам'яті, при цьому цілком можлива ситуація, коли кілька апаратних засобів одночасно звертаються до однієї й тієї ж змінної. Запити будуть оброблені коректно, коли дані тільки зчитуються, але якщо хоча б один з апаратних засобів записує дані, то результат виконання операцій читання/запису стає нез'ясованим. Якщо не визначено результату одночасного запису кількома апаратними засобами різних значень у ту саму змінну, то кожний трансп'ютер є системою зі спільною пам'яттю.

У перших трансп'ютерних системах трансп'ютери з'єднувалися лінками безпосередньо між собою й утворена в такий спосіб конфігурація, або **топологія**, фіксувалася на час розв'язування задачі. У разі потреби в отриманні іншої топології *link* доводилося з'єднувати вручну. Щоб спростити обслуговування й забезпечити можливість зміни топології системи безпосередньо у процесі обчислень було розроблено електронний комутатор С-004 (рис. 2.6), як електронно-конфігурований комутатор, що дозволяє задавати довільні парні зв'язки між 32 входами трансп'ютерних *link* і 32 виходами. На рис. 2.6 вісім трансп'ютерів жорстко з'єднано в лінійку (*pipe*) двома із чотирьох своїх *link*, а іншими під'єднано до комутатора. Між під'єднаними до комутатора *link* можна програмно визначити будь-які чотири зв'язки. Наприклад, показаний пунктиром зв'язок перетворює вихідну топологію «лінійка» в топологію «кільце».

Маючи порівняно невелику продуктивність, трансп'ютери швидко втратили своє значення як обчислювальні елементи, але довгий час зберігали його як комутаційні у складі гібридних обчислювальних систем.



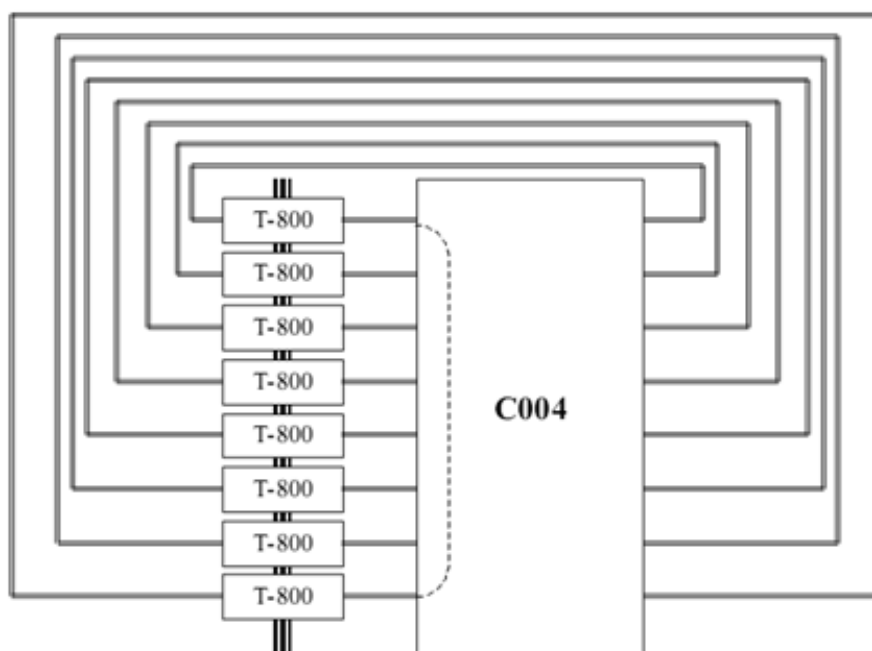


Рис. 2.6. Електронно-реконфігуроване з'єднання трансп'ютерів за допомогою комутатора C004

**Гібридні системи.** Розглянемо багатопроцесорну систему Parsytec PowerXplorer, яка складається з 12 процесорних вузлів (рис. 2.7), кожний з них використовує процесор Powerpc-601. Реальна продуктивність кожного процесора – близько 30 Mflops. Кожний процесор використовує 8 Мбайт локальної оперативної пам'яті, до якої має доступ трансп'ютер типу T805. Трансп'ютери за допомогою *link* з'єднані між собою топологією у вигляді прямокутної сітки, частина трансп'ютерів має вихід через один зі своїх каналів зв'язку на керувальну машину типу Sunsparc (рис. 2.8). Підключення здійснюється за допомогою SCSI-пристрою Scsi-Bridge, який підтримує чотири трансп'ютерних *link* для зв'язку із системою PowerXplorer. Реальна продуктивність кожного з чотирьох трансп'ютерних каналів зв'язку – від 1 Мбайт/с у разі передачі довгих повідомлень до 0,3 Мбайт/с у разі передачі повідомлень довжиною в декілька байтів.

Розробка програмного забезпечення для подібних систем можлива засобами системи програмування PARIX™, яка дозволяє використовувати будь-яку зручну віртуальну топологію (тор, решітку, зірку тощо), при цьому трансп'ютер прозорий для програміста. Фізична топологія жорстко задана й не змінюється.

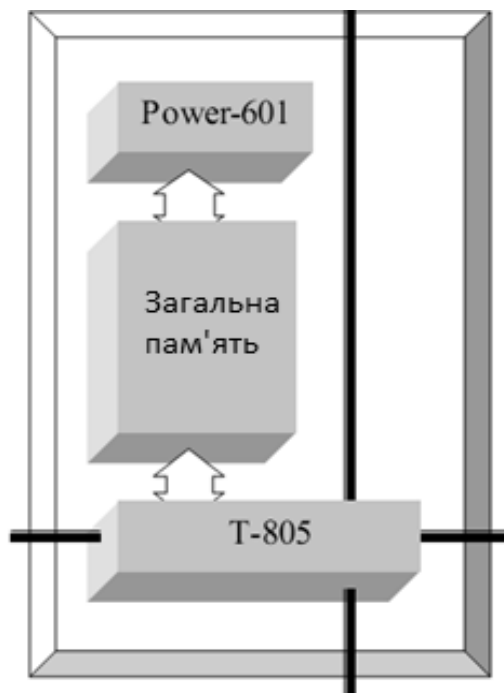


Рис. 2.7. Структура вузла PowerXplore

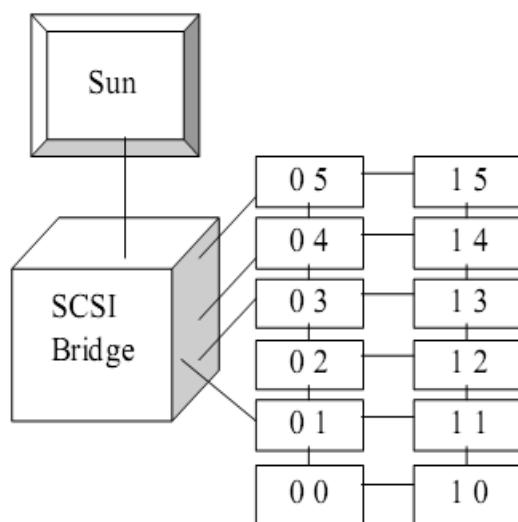


Рис. 2.8. Підключення 12-процесорної системи PowerXplorer до керувальної машини типу Sun

Гібридні системи, в основі яких лежать трансп'ютерні канали зв'язку, набули поширення, але нині вони втрачають своє значення через недостатню пропускну здатність *link*.

**Кластери робочих станцій.** З появою та розвитком локальних мереж з'явилися кластери робочих станцій, які являють собою об'єднання невеликої

кількості персональних комп'ютерів і/або робочих станцій. Хоч кластери робочих станцій є відносно дешевими, ці системи часто мають нижчу ефективність розв'язання прикладних задач порівняно зі спеціалізованими системами з таких основних причин:

- певні поширені локальні мережі (Ethernet, Token Ring) не підтримують одночасної передачі даних між різними парами комп'ютерів у межах одного сегмента мережі. Це означає, що дані між комп'ютерами  $c$  і  $d$  можуть бути передані тільки після передачі даних між комп'ютерами  $a$  і  $b$ , що зменшує не дуже високу наявну швидкість передачі даних у таких мережах;
- майже завжди на робочих станціях, поєднаних у кластер, продовжують виконуватися послідовні завдання користувачів, у результаті чого менш завантажені процесори змушені очікувати більш завантажених, що призводить до загального зниження продуктивності кластера під час розв'язування паралельної задачі до рівня, визначеного найбільш завантаженою машиною;
- наявність у мережі файлових серверів призводить до нерегулярної зміни обсягів даних, які передаються через локальну мережу, що може суттєво збільшувати час обміну повідомленнями між процесорами та інтервали їх простоювання.

**Системи на основі високошвидкісних мереж.** Найбільш перспективними вважаються багатопроцесорні системи, побудовані на основі спеціалізованих високошвидкісних мереж передачі даних, серед яких є подані на рис. 2.9, 2.10 системи Parsytec CC (Cognitive Compute – комп'ютер розумний).

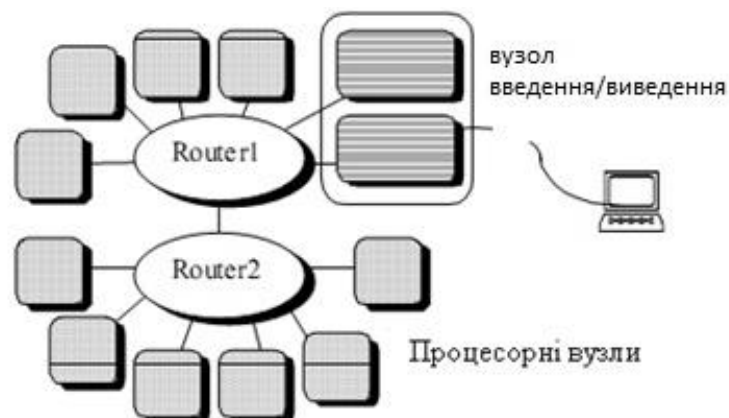


Рис. 2.9. Структура системи Parsytec CC-12

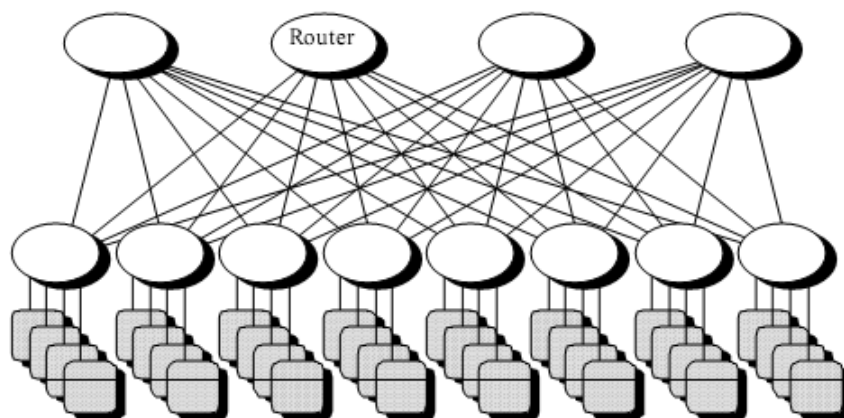


Рис. 2.10. Структура системи Parsytec CC-32

Кожний вузол систем цієї серії (рис. 2.11) являє собою повноцінний комп'ютер, керований Unix-подібною операційною системою AIX, система має вузли двох типів: обчислювальні й вузли введення/виведення. Процесорний вузол містить обчислювальний процесор Power PC-604, накопичувач на жорсткому диску, адаптер локальної мережі Ethernet, адаптер високошвидкісної мережі HS-Link. До складу вузла введення/виведення додатково входять відеоадаптер, адаптери клавіатури й маніпулятора типу «миша», жорсткий диск збільшеної місткості. Крім того, можна встановити другий адаптер мережі Ethernet для під'єднання системи до локальної або глобальної мережі. Показаний на рис. 2.11 послідовний інтерфейс RS-232 призначений для технічного обслуговування системи. Кожний концентратор високошвидкісної мережі (Router) може обслуговувати до восьми HS-Link каналів.

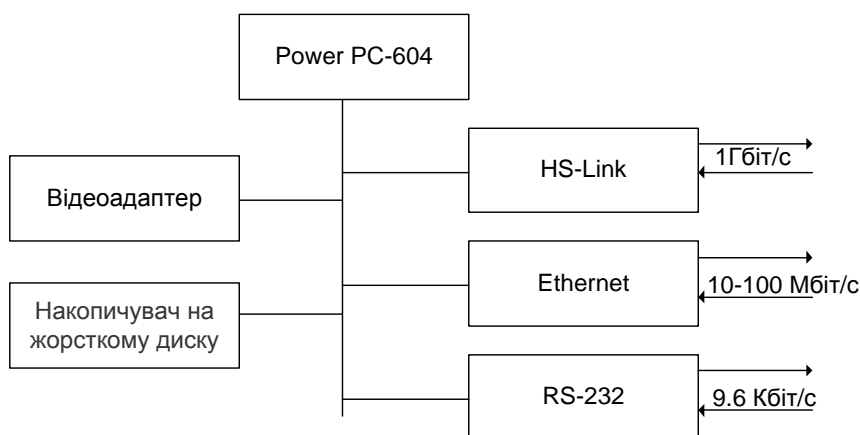


Рис. 2.11. Структура процесорного вузла системи Parsytec CC

Системи цієї серії мають такі особливості:

- висока швидкість передачі даних між процесорами за рахунок застосування спеціальних високошвидкісних комунікаційних модулів і лінків;
- кожний обчислювальний вузол обслуговує власна копія операційної системи, тобто кожний вузол є незалежною однопроцесорною обчислювальною системою зі своєю дисковою пам'яттю;
- для функціонування системи не потрібно наявності додаткової машини для керування, оскільки вона входить до локальної Ethernet-мережі, яка вже має визначені принципи керування під час свого функціонування;
- зв'язок між вузлами здійснюється як через стандартну внутрішню Ethernet-мережу, так і через спеціальну високошвидкісну мережу HS-Link.

Системи на основі розподіленої пам'яті є масштабованими. Для об'єднання великої кількості процесорів може використовуватися легко масштабована, необмежено нарощувана топологія, наведена на рис. 2.12.

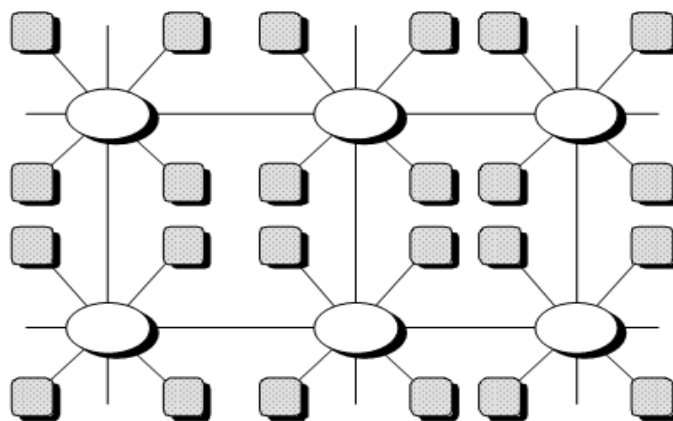


Рис. 2.12. Масштабований варіант топології системи Parsytec CC

Ще однією особливістю систем такого типу є можливість їх побудови без концентраторів високошвидкісної мережі (Routers). Для цього достатньо встановити на кожному із внутрішніх вузлів ланцюжка процесори (рис. 2.13) програмне забезпечення адаптера HS-Link.

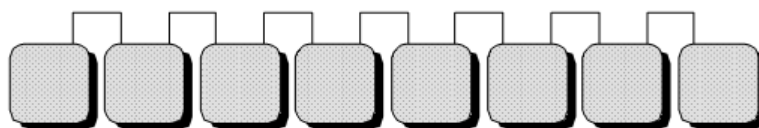
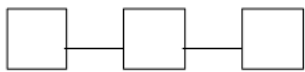


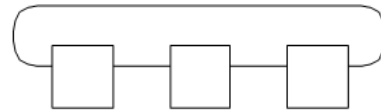
Рис. 2.13. Варіант топології системи Parsytec CC, що не вимагає використання високошвидкісних комутаторів

### 2.3. Топології багатопроцесорних систем

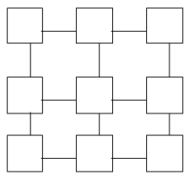
Під топологією багатопроцесорної системи розуміємо спосіб з'єднання процесорних вузлів між собою каналами передачі даних. Зручно подати топологію системи у вигляді графу, вершини якого відповідають процесорним вузлам, а ребра – каналам зв'язку відповідно. Умовно топології можна поділити на фіксовані, реконфігуровані, з одного боку, на регулярні й нерегулярні – з другого. Серед регулярних широко використовують топології таких типів:



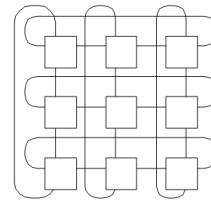
Топологія «лінійка»



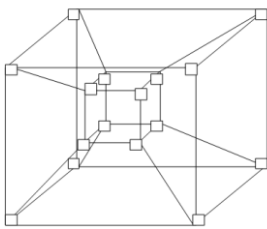
Топологія «кілеце»



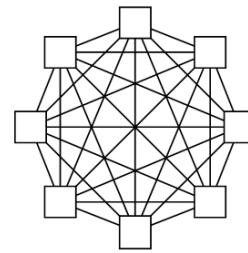
Топологія «решітка 3x3»



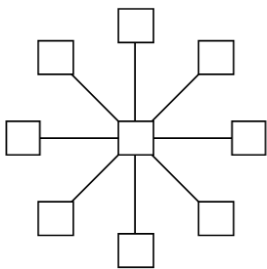
Топологія «тор 3x3»



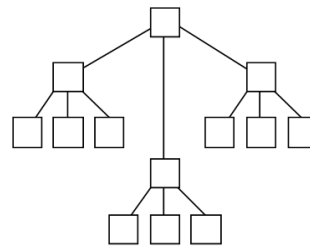
Топологія «гіперкуб ступеня 4»



Топологія «клику»



Топологія «зірка»



Топологія «трійкове дерево»