

MULTIPROCESADORES (MIMD)

6.1 - Generalidades de Multiprocesadores.

Pueden clasificarse en esta categoría muchos sistemas multiprocesadores y sistemas multicomputadores.

Un multiprocesador se define como una computadora que contiene dos o más unidades de procesamiento que trabajan sobre una memoria común bajo un control integrado.

Si el sistema de multiprocesamiento posee procesadores de aproximadamente igual capacidad, estamos en presencia de multiprocesamiento simétrico; en el otro caso hablamos de multiprocesamiento asimétrico.

Todos los procesadores deben poder acceder y usar la memoria principal. De acuerdo a esta definición se requiere que la memoria principal sea común y solamente existen pequeñas memorias locales en cada procesador.

Si cada procesador posee una gran memoria local se lo puede considerar un sistema de multicomputadoras, el cual puede ser centralizado o distribuido.

Todos los procesadores comparten el acceso a canales de E/S, unidades de control y dispositivos.

Para el sistema de multiprocesamiento debe existir un sistema operativo integrado, el cual controla el hardware y el software y debe asegurar la interacción entre los procesadores y sus programas al nivel elemental de dato, conjunto de datos y trabajos.

Una computadora MIMD intrínseca implica interacciones entre n procesadores debido a que todos los flujos de memoria se derivan del mismo espacio de datos compartido por todos los procesadores. Si los n flujos de datos provienen de subespacios disjuntos de memorias compartidas, entonces estamos en presencia del denominado operación SISD múltiple, que no es otra cosa que un conjunto de n monoprocesadores SISD.

Una MIMD intrínseca está fuertemente acoplada si el grado de interacción entre los procesadores es alto. De otra manera consideramos el sistema como débilmente acoplado. Muchos sistemas comerciales son débilmente acoplados, a saber, la IBM 370/168, Univac 1100/80, IBM 3081/3084, etc.

6.2 - MULTIPROCESADORES Y MULTICOMPUTADORES

Existen similitudes entre los sistemas multiprocesadores y multicomputadores debido a que ambos fueron pensados con un mismo objetivo: dar soporte a operaciones concurrentes en el sistema. Sin embargo, existen diferencias importantes basadas en el alcance de los recursos compartidos y la cooperación en la solución de un problema.

Un sistema multicomputador consiste de diversas computadoras autónomas que pueden o no comunicarse entre sí.

Un sistema multiprocesador está controlado por un sistema operativo que provee la interacción entre los procesadores y sus programas a nivel de dato, proceso y archivo.

6.3 - FORMAS DE ACOPLAMIENTO

Existen dos modelos arquitectónicos diferentes para los sistemas multiprocesadores: Fuertemente acoplado y Débilmente acoplado. Los sistemas fuertemente acoplados se comunican a través de una memoria común. De allí que el promedio de velocidad con la cual un procesador puede comunicarse con otro es del orden del bandwidth de la memoria.

Puede existir una pequeña memoria local o un buffer de alta velocidad (cache) en cada procesador.

Existe una completa conectividad entre los procesadores y la memoria. Esta conectividad puede alcanzarse insertando una red de interconexión entre los procesadores y la memoria; o mediante una memoria multipuertas.

Uno de los factores que limitan el crecimiento de los sistemas fuertemente acoplados es la degradación debido a la contención de memoria que ocurre cuando dos o más procesadores intentan acceder la misma unidad de memoria concurrentemente.

Puede reducirse el grado de conflictividad incrementando el grado de interleaving. Sin embargo, esto debe acompañarse de una cuidadosa asignación de los datos a los módulos de memoria.

Los sistemas multiprocesadores débilmente acoplados no tienen, en general, el grado de conflictos sobre la memoria de los fuertemente acoplados.

En este sistema cada procesador tiene un conjunto de dispositivos de E/S y una gran memoria local a donde accede para obtener la mayoría de sus datos e instrucciones.

Nos referiremos al procesador, sus dispositivos de E/S y su memoria local como al módulo computador.

Los procesos que se ejecutan en diferentes procesadores se comunican intercambiando mensajes a través de un sistema de transferencia de mensajes. El grado de acoplamiento en tales sistemas es realmente muy débil, de allí que se los conozca también como sistemas distribuidos.

El factor determinante del grado de acoplamiento es la topología de la comunicación del sistema de transferencia de mensajes asociado.

Los sistemas débilmente acoplados son eficientes cuando la interacción entre las tareas es mínima.

Los sistemas fuertemente acoplados pueden soportar una gran interacción entre las tareas sin un deterioro significativo de la performance.

La Fig. 6.1. muestra un ejemplo de un módulo computador de un multiprocesador débilmente acoplado no

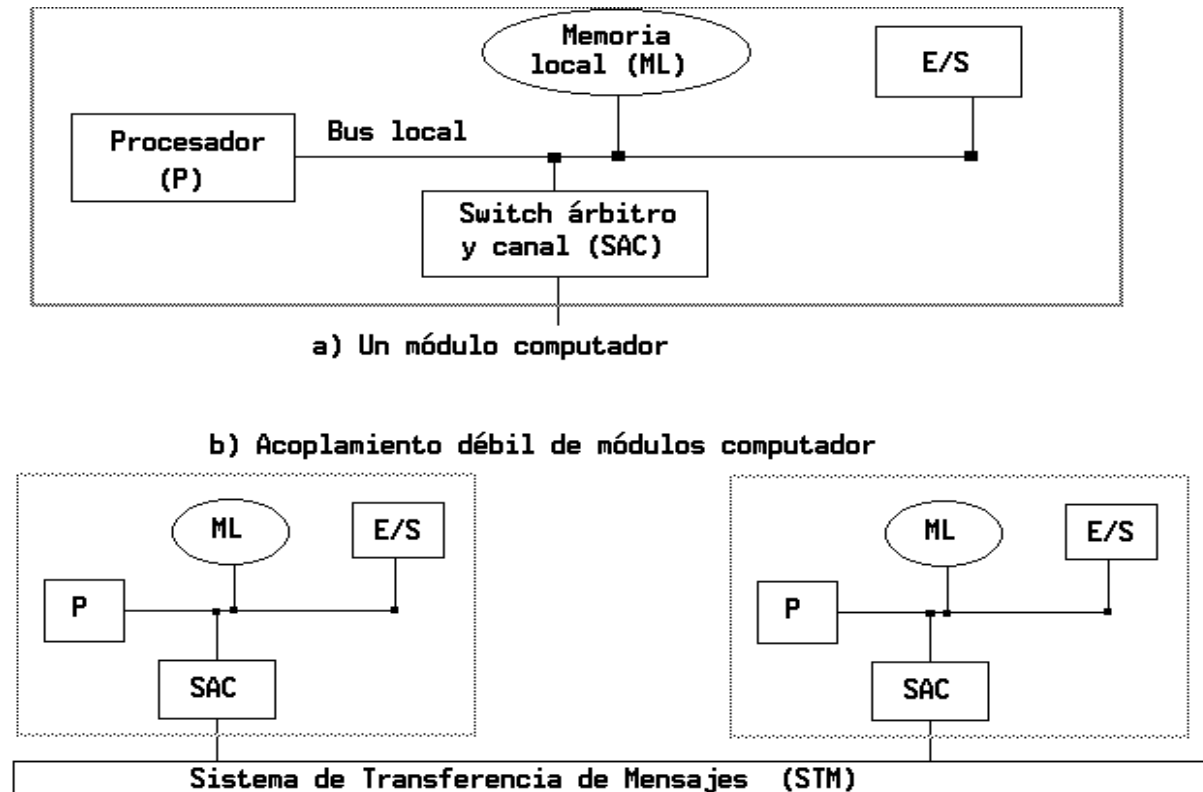


Fig. 6.1. - Sistema multiprocesador no jerárquico débilmente acoplado.

jerárquico. Consiste en un procesador, una memoria local, dispositivos de E/S locales y una interfase a otros módulos computadores.

La interfase puede contener un switch árbitro y un canal. La Fig. 6.1. muestra también la conexión entre los módulos computador y el sistema de transferencia de mensajes (STM).

Si los pedidos para dos o más módulos computador colisionan al acceder un segmento físico del STM, el árbitro es el responsable de elegir uno de los pedidos simultáneos de acuerdo a una determinada disciplina de servicio.

Es también responsable de hacer esperar los otros pedidos hasta que se complete la atención del pedido actual.

El canal que se encuentra dentro del SAC puede tener una memoria de comunicación de alta velocidad a efectos de bufferizar los bloques de transferencia de mensajes. La memoria de comunicación es accesible por todos los procesadores.

En la Fig. 6.2 podemos ver un esquema de multiprocesador fuertemente acoplado.

Este consiste de P procesadores, S módulos de memoria y D canales de Entrada/Salida. Estas unidades están conectadas mediante tres redes de interconexión, a saber, la red de interconexión entre los procesadores y los IOP, y la red de interconexión de interrupciones-señales.

Los conflictos de acceso a memoria por varios procesadores son resueltos por la red de interconexión procesador-memoria. Para evitar excesivos conflictos, la cantidad de módulos de memoria S es generalmente tan grande como P.

Otro método para reducir el grado de conflictos es asociar un área de almacenamiento reservada para cada procesador. Esta es la memoria local no mapeada que se usa para almacenar código Kernel y tablas del sistema operativo muy utilizadas por los procesos que se ejecutan en tal procesador.

Se puede agregar también a esta configuración una memoria cache propia de cada procesador a fin de disminuir las referencias a memoria principal.

La red de interconexión de interrupciones-señales permite que cada procesador envíe directamente una interrupción a otro procesador. La sincronización entre procesos se ve facilitada por esta red. Esta red puede actuar como un procesador de fallas ya que puede enviar una alarma originada por hardware a los procesadores que sí funcionen.

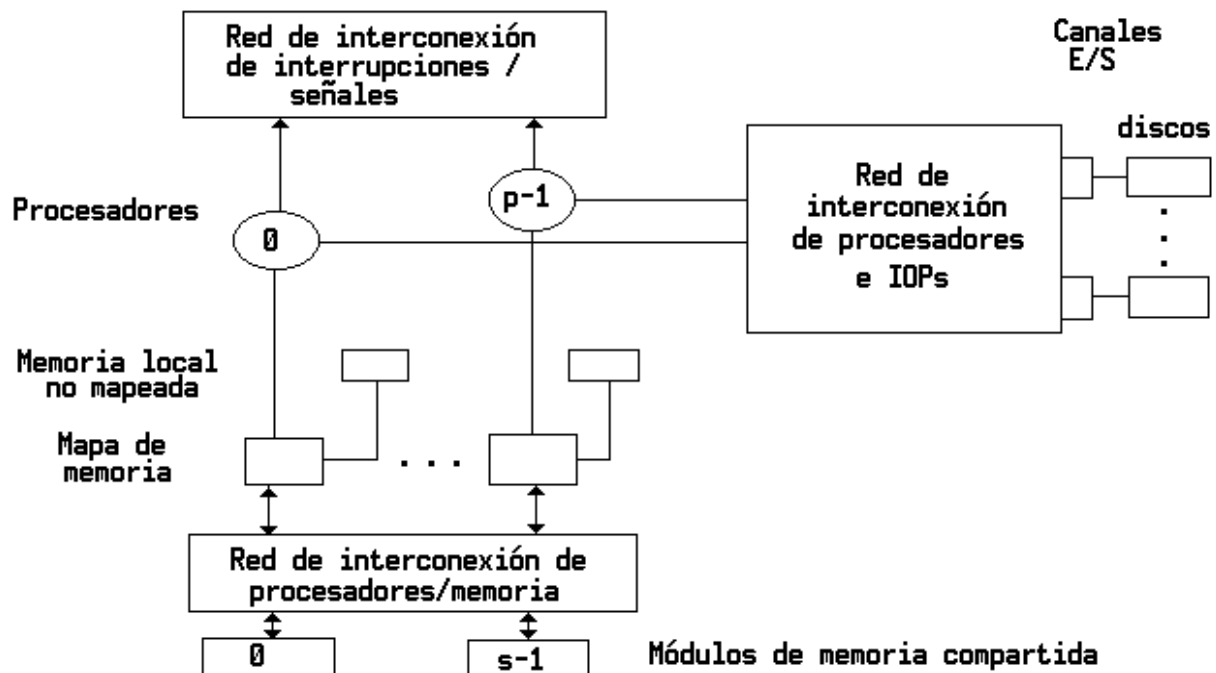


Fig. 6.2. - Configuración multiprocesador fuertemente acoplado.

El conjunto de procesadores puede ser homogéneo o heterogéneo. Es homogéneo si los procesadores son funcionalmente idénticos.

Pero aún siendo homogéneos pueden ser simétricos o asimétricos dependiendo de que dos unidades funcionalmente idénticas difieran en cuanto a dimensiones tales como accesibilidad de E/S, performance o confiabilidad. Esta última configuración de la Fig. 6.2 como procesador fuertemente acoplado es también conocida como sistema de multiprocesadores *diádicos*.

6.4. - DEFINICION DE ARQUITECTURAS PARALELAS

Problemas. Se han propuesto diversas definiciones para arquitecturas paralelas. La dificultad en definir con precisión el término está entrelazada con el problema de especificar una taxonomía de arquitecturas paralelas. El problema central para poder especificar una definición y consiguientemente la taxonomía para las modernas arquitecturas paralelas es lograr satisfacer el siguiente conjunto de imperativos :

- *) Excluir las arquitecturas que incorporan solamente mecanismos de paralelismo de bajo nivel y que se han popularizado tanto como característica típica de las modernas computadoras.
- *) Mantener los elementos útiles de la clasificación de Flynn tales como los flujos de datos e instrucciones.
- *) Incluir los procesadores vectoriales pipelinizados y otras arquitecturas que intuitivamente ameritan incluirse como arquitecturas paralelas, pero que no se ajustan fácilmente al esquema de Flynn.

Examinaremos cada una de estos imperativos así como obtendremos una definición que los satisfice totalmente y provee una base para una razonable taxonomía.

6.4.1. - Paralelismo de bajo nivel

Existen dos razones para excluir las máquinas que utilizan mecanismos de paralelismo de bajo nivel del conjunto de arquitecturas paralelas.

Primero, si no adoptamos un standard riguroso prácticamente la mayoría de las computadoras modernas serían "arquitecturas paralelas", anulando la utilidad del término en sí. Y segundo, las arquitecturas que solamente tienen las características que vamos a enunciar a renglón seguido no ofrecen un marco de referencia explícito y coherente para desarrollar soluciones paralelas de alto nivel :

Pipelining de instrucciones la descomposición de la ejecución de una instrucción en una serie lineal de etapas autónomas, permitiendo que cada etapa simultáneamente realice una porción del procesamiento de la instrucción (por. ej. decodificación, cálculo de la dirección efectiva, levantar operandos de memoria, ejecutar y almacenar).

Múltiples unidades funcionales en la CPU proveyendo unidades funcionales independientes para la ejecución concurrente de operaciones aritméticas y booleanas.

Procesadores separados para E/S y CPU liberando a la CPU del control sobre las entradas/salidas mediante el uso de procesadores dedicados, solución que abarca desde los controladores más sencillos de E/S hasta las complejas unidades de procesamiento periféricas.

A pesar de que estas características contribuyen significativamente a la performance, su presencia no hace que una computadora posea una arquitectura paralela.

6.4.2. - Taxonomía de Flynn

La taxonomía de Flynn clasifica las arquitecturas de las computadoras según la presencia de únicos o múltiples flujos de datos e instrucciones. Hemos visto ya en el capítulo 4 las cuatro categorías de esta clasificación.

SISD define las computadoras seriales.

MISD implica que muchos procesadores aplican diferentes instrucciones al mismo dato, esta posibilidad hipotética se considera generalmente impracticable.

SIMD implica que múltiples procesadores ejecutan simultáneamente la misma instrucción sobre diferentes datos.

MIMD implica que múltiples procesadores ejecutan autónomamente diversas instrucciones sobre diversos datos.

Si bien estas distinciones proveen elementos útiles para caracterizar arquitecturas, no bastan para clasificar varias de las computadoras modernas.

Por ejemplo, los procesadores vectoriales pipelineados merecen ser incluidos como arquitecturas paralelas, ya que muestran una concurrencia substancial en la ejecución aritmética y pueden manejar cientos de elementos de vectores en forma paralela, no obstante lo cual no se ajustan a los parámetros de la clasificación de Flynn, debido a que, si los consideramos SIMD carecen de procesadores que ejecutan la misma instrucción en pasos bien acotados, y si por otra parte los clasificaremos como MIMD les falta la autonomía asincrónica de la categoría.

6.4.3. - Definición y taxonomía

Un primer paso para proveer una taxonomía satisfactoria es articular una definición de arquitecturas paralelas. Esta definición debe incluir las computadoras que la clasificación de Flynn no puede manejar y debe excluir aquellas que incorporan el paralelismo de bajo nivel.

Por lo tanto, una arquitectura paralela provee un explícito marco de referencia de alto nivel para el desarrollo de soluciones de programación paralelas logrado mediante múltiples procesadores (simples o complejos) que cooperan para resolver problemas a través de ejecución concurrente.

La Fig. 6.3 muestra una taxonomía basada en las imperativas discutidas anteriormente y la definición propuesta.

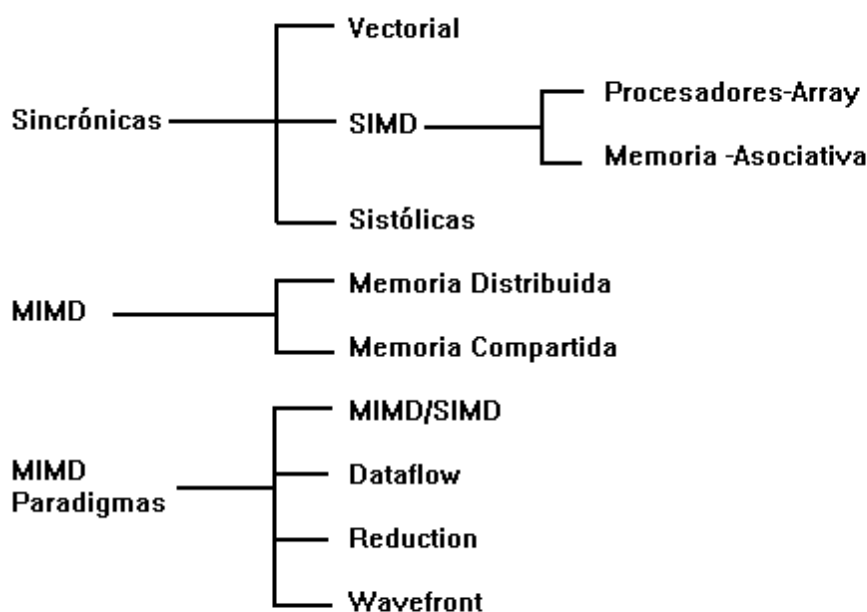


Fig. 6.3. - Una taxonomía de alto nivel para arquitecturas de computadores paralelos.

Esta taxonomía informal utiliza categorías de alto nivel para delinear los principales planteamientos sobre las arquitecturas de las computadoras paralelas y para mostrar que estos planteamientos definen un espectro coherente de alternativas arquitecturales. Las definiciones de cada categoría las definimos a continuación.

Esta taxonomía no intenta suplantarse a aquellas construidas con un mayor esfuerzo formal. Tales taxonomías proveen subcategorías que reflejan alteraciones de características arquitecturales y cubren las características de bajo nivel.

6.5. - **ARQUITECTURAS SINCRÓNICAS**

Las arquitecturas paralelas sincrónicas coordinan operaciones concurrentes en pasos acotados mediante relojes globales, unidades centrales de control o controladores de la unidad vectorial.

6.5.1. - **Procesadores Vectoriales Pipelinizados**

Los procesadores vectoriales se caracterizan por múltiples unidades funcionales pipelinizadas que operan concurrentemente e implementan operaciones aritméticas y booleanas tanto escalares como matriciales.

Debido a que tales arquitecturas proveen paralelismo a nivel de tareas, puede argumentarse con ciertas reservas que son MIMD, aún cuando las capacidades de procesamiento vectorial son el aspecto fundamental de sus diseños.

6.6. - **Arquitecturas SIMD y de Procesador Array**

Las arquitecturas SIMD y las de los procesadores array fueron ya vistas con mayor detalle en el capítulo 5.

6.7. - **ARQUITECTURAS MIMD**

Hemos ya discutido en este capítulo las características básicas de las arquitecturas MIMD. En el siguiente capítulo veremos los casos particulares según que la memoria sea compartida o distribuida, así como algunas de las diferentes topologías que se utilizan.

6.8. - **PARADIGMAS BASADOS EN ARQUITECTURAS MIMD.**

Las arquitecturas híbridas MIMD/SIMD, las máquinas de Reducción, las arquitecturas Dataflow y los Wavefront array son arquitecturas igualmente difíciles de acomodar ordenadamente en una clasificación de las arquitecturas paralelas.

Cada una de estas arquitecturas está basada en los principios de operación asincrónica y manejo concurrente de múltiples flujos de datos e instrucciones.

Sin embargo cada una de ellas se basa, a su vez, en algún principio muy distintivo que se suma a las características propias de ser MIMD y que, por lo tanto, merecen un tratamiento especial.

Por esta razón es que incluiremos estas arquitecturas en el capítulo 9 como "Arquitecturas Nuevas" para destacar, justamente, sus características particulares.

EJERCICIOS

- 1) Qué es una arquitectura MIMD y cómo funciona ? Grafíquelo.
- 2) Cuál es la diferencia entre un MIMD fuertemente acoplado y un MIMD débilmente acoplado ?
- 3) Falso o Verdadero : No mejora la performance de un sistema multiprocesador débilmente acoplado el contar con memoria interleaved. Justifique.
- 4) En qué caso son eficientes los sistemas multiprocesadores fuertemente acoplados ?
- 5) Qué término se utiliza para indicar cuando un sistema multiprocesador posee procesadores de aproximadamente igual capacidad ?
- 6) A qué se denomina paralelismo de bajo nivel y en qué situaciones es factible encontrarlo ?
- 7) Indique cuáles de los elementos que se enumeran constituyen formas de paralelismo de bajo nivel :
 - pipelining de instrucciones
 - procesadores diádicos
 - pipelines aritméticos
 - IOP's
 - pipeline vectorial
- 8) Cuál es la diferencia entre una arquitectura sincrónica y una no sincrónica ?
- 9) En base a qué se pueden clasificar las arquitecturas ?