



**Universidad
Europea de Madrid**

LAUREATE INTERNATIONAL UNIVERSITIES

MULTIPROCESADORES

TIPOS DE PARALELISMO

© Todos los derechos de propiedad intelectual de esta obra pertenecen en exclusiva a la Universidad Europea de Madrid, S.L.U. Queda terminantemente prohibida la reproducción, puesta a disposición del público y en general cualquier otra forma de explotación de toda o parte de la misma.

La utilización no autorizada de esta obra, así como los perjuicios ocasionados en los derechos de propiedad intelectual e industrial de la Universidad Europea de Madrid, S.L.U., darán lugar al ejercicio de las acciones que legalmente le correspondan y, en su caso, a las responsabilidades que de dicho ejercicio se deriven.

Índice

Presentación	4
Clasificación de Flynn	6
Tipos de MIMD	7
Multiprocesadores. Arquitecturas memoria compartida	9
Multiprocesadores. Arquitecturas con memoria distribuida	11
Tipos de memoria compartida	13
Modelo UMA	14
Modelo NUMA	15
Modelo COMA	16
Arquitecturas con memoria distribuida	17
Resumen	18

Presentación

Cuando hablamos de paralelismo o procesamiento paralelo, normalmente nos viene a la mente un sistema multiprocesador. En este tema veremos que hay distintos tipos de paralelismo y de arquitecturas de procesamiento paralelo, entre las que se encuentran, los sistemas multiprocesadores y sus variantes.

Los multiprocesadores, así como otras arquitecturas de procesamiento paralelo, surgieron de la necesidad de resolver problemas que se denominan *de procesamiento intenso*, tales como:

- El tratamiento de imágenes capturadas por satélites.
- Aplicaciones médicas que implican la generación de imágenes filtradas en tiempo real.
- La investigación en el campo de la astrología, donde las aplicaciones procesan grandes imágenes.
- El tratamiento de datos extraídos de Internet.

Este tipo de aplicaciones necesitan arquitecturas capaces de operar, del orden de **TFLOPS**:

1 Tera FLOP = 10^{12} FLOPS = 1 Billón de operaciones en coma flotante por segundo.



Esta necesidad de aumentar el rendimiento de los computadores a una mayor velocidad que la evolución de los procesadores, llevó a los arquitectos de computadores a realizar diseños creativos, como la construcción de multiprocesadores, que pudieran multiplicar el rendimiento de la máquina.

En este tema aprenderemos a clasificar distintos tipos de arquitecturas de procesamiento intenso, y seremos capaces de identificar cuál es la más adecuada según el tipo de aplicación que se le quiera asignar a la arquitectura.

Clasificación de Flynn

En 1966, Flynn estableció una clasificación del paralelismo que todavía hoy es utilizada para describir las arquitecturas de procesamiento intenso. La clasificación de Flynn se basa en dos características:

- **Número de instrucciones que ejecutan en paralelo:** una instrucción (*Single Instruction*) o múltiples (*Multiple Instructions*)
- **Número de fuentes de datos que se tratan en paralelo:** una entrada (*Single Data*) o múltiples (*Multiple Data*).

Sobre la base de estas características, la clasificación de Flynn establece cuatro tipos de arquitecturas según el paralelismo utilizado en instrucciones y datos.

TIPO	DESCRIPCIÓN	EJEMPLOS
SISD	Single Instruction, Single Data.	Monoprocesador convencional.
SIMD	Single Instruction, Multiple Data.	Procesador vectorial (trabaja con vectores, múltiples datos en paralelo), Procesador superescalar (múltiples ALU que operan en paralelo).
MISD	Multiple Instruction, Single Data.	No existen arquitecturas de este tipo.
MIMD	Multiple Instruction, Multiple Data.	Sistemas multinúcleo y multiprocesador.

En los últimos tiempos, se habla también del concepto de **SPMD** (*Single Program Multiple Data*), haciendo referencia al modelo de programación de las arquitecturas **MIMD**, donde un único programa se ejecuta en todos los procesadores. El programador puede incluir directivas condicionales de paralelismo en el código secuencial.

Tipos de MIMD

Dentro de las arquitecturas MIMD, existen dos tendencias de desarrollos arquitectónicos:

- Arquitecturas multiprocesador.
- Arquitecturas multicomputador.

La principal diferencia entre estos dos tipos de arquitecturas, radica en el grado de compactación de la arquitectura y en la forma en la que se comunican los componentes. Se dice que las arquitecturas multiprocesador están fuertemente acopladas, y las arquitecturas multicomputador, que están débilmente acopladas. En la realidad, existe una línea muy delgada para distinguir dónde terminan los multiprocesadores y dónde comienzan los multicomputadores.

Vamos a plantear dos ejemplos que reflejan los casos más extremos de ambas arquitecturas, donde se diferencia claramente la naturaleza de cada tendencia.

Arquitectura multiprocesador multinúcleo	Los procesadores se encuentran localizados en el mismo chip, normalmente comparten el mismo espacio de memoria para ejecutar programas en paralelo y resolver un mismo problema en común en mucho menor tiempo.
Arquitectura cliente-servidor	La misma aplicación se ejecuta en dos máquinas independientes, normalmente cada una tiene su propio procesador, su propia memoria principal, incluso su propio controlador de E/S con dispositivos periféricos independientes. Cada máquina ejecuta una parte de código distinto y solo se intercambian información a través de paso de mensajes. Si llevamos al extremo la definición de arquitectura multicomputador, podríamos incluir, en este tipo de computadores, una arquitectura cliente-servidor ejecutada en red donde los ordenadores están físicamente separados en lugares distintos.

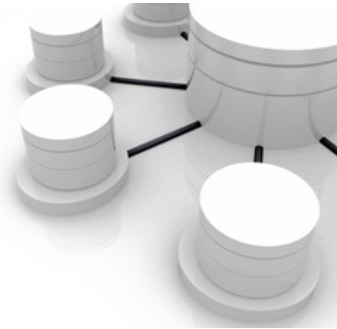


Multiprocesadores. Arquitecturas memoria compartida

Dentro de los multiprocesadores existen dos tipos de arquitecturas claramente diferenciadas:

- Arquitecturas con memoria compartida.
- Arquitecturas con memoria distribuida.

En este apartado vamos a tratar las arquitecturas con memoria compartida, en las que todos los procesadores comparten el mismo espacio de direcciones. Es decir, todo el direccionamiento de la memoria principal es único, y los procesadores pueden acceder a todas las posiciones de memoria.



Por su parte, la comunicación entre procesadores normalmente se hace a través de la memoria mediante el uso de variables compartidas. Si el número de componentes de la arquitectura es reducido, puede ser suficiente un bus de datos digital lo bastante ancho (suficientes hilos) para comunicar los componentes. Arquitecturas grandes utilizan como modelo de interconexión una red **crossbar** o una red **multietapa**.

Este tipo de multiprocesadores debe utilizar memorias cachés suficientemente grandes. Cada procesador tendrá su memoria caché independiente y privada para evitar, más de lo necesario, la búsqueda de datos en la memoria principal, y por tanto, impedir un cuello de botella cuando varios procesadores requieran acceder a datos de memoria al mismo tiempo.

Estas arquitecturas son las más utilizadas en la actualidad, puesto que ofrecen buenas prestaciones por tener varios procesadores trabajando en paralelo. Además, la gestión de varios procesadores se hace muy sencilla por estar muy compactos y apenas se consumen recursos para este fin.

Desde el punto de vista de consumo, las nuevas arquitecturas multinúcleo, al compartir un mismo chip, hacen que el consumo de potencia disminuya respecto a arquitecturas multiprocesadores más antiguas y con menor grado de acoplamiento.

Este tipo de arquitecturas suele tener un número de procesadores relativamente pequeño, entre 2 y 32 procesadores.

Multiprocesadores. Arquitecturas con memoria distribuida

En el caso de arquitecturas con memoria distribuida, todos los procesadores poseen una memoria privada. Es decir, que no todos pueden acceder a todas las posiciones de memoria, y hay varios módulos de memoria independientes que no comparten el mismo direccionamiento. Pueden trabajar con datos de memoria compartidos, pero la gestión de la memoria compartida es mucho más costosa que en el otro tipo de multiprocesadores, puesto que no comparten ningún módulo de memoria principal.

En este tipo de arquitecturas, la comunicación entre procesadores se hace a través de paso de mensajes. Como modelo de interconexión, normalmente se utiliza una red **crossbar** o una red **multietapa**.

Como ventaja principal de este tipo de arquitecturas encontramos, por un lado, la escalabilidad, ya que permite ampliar el número de procesadores sin grandes cambios en la arquitectura. Por otro lado, ofrecen una solución al problema del ancho de banda, ya que los procesadores no comparten la misma memoria principal y hay menos comunicaciones. Normalmente, el número de procesadores de este tipo de arquitecturas es alto (entre 512 y 1024 procesadores), por eso se necesita una red de interconexión que permita conectar muchos componentes.

Como principal desventaja, debemos mencionar el coste en la gestión de tantos procesadores, especialmente cuando trabajan en problemas comunes, y el envío de datos a través de la red es más lento debido a los protocolos de comunicación. La gestión de muchos procesos y tareas hace que la arquitectura pierda eficiencia.

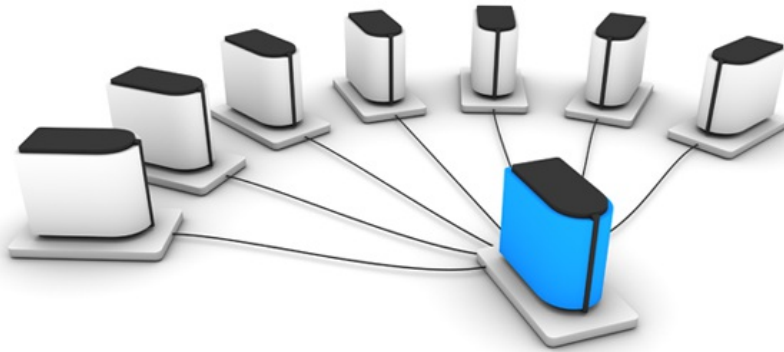
Desde el punto de vista de consumo, el consumo de potencia es mayor que en arquitecturas de memoria centralizada.



Tipos de memoria compartida

Dentro de las arquitecturas multiprocesador de memoria compartida, hay que distinguir tres modelos.

Tipo	Descripción
Modelo UMA	<i>Uniform Memory Architecture.</i> Memoria compartida uniformemente por todos los procesadores. Todos tardan el mismo tiempo en acceder a cualquier posición de memoria.
Modelo NUMA	<i>Non Uniform Memory Architecture.</i> Aunque el espacio de direcciones es único, cada procesador tiene su propio módulo de memoria principal. Cada procesador tiene prioridad de acceso a su módulo privado respecto al resto de procesadores.
Modelo COMA	<i>Cache Only Memory Architecture.</i> En este tipo de arquitecturas existen módulos de memoria privados y módulos de memoria compartida. Algunos módulos de memoria principal se utilizan como memoria caché privada.

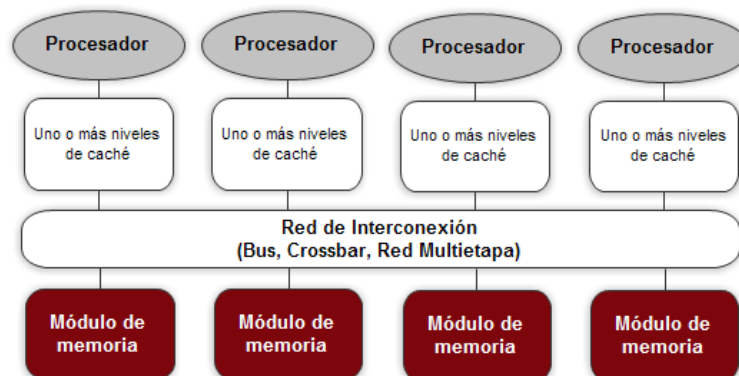


Modelo UMA

El modelo **UMA** es un tipo de arquitectura multiprocesador de memoria compartida. En esta arquitectura, todos los procesadores comparten el mismo espacio de memoria y tardan el mismo tiempo en acceder a cualquier posición de memoria, de ahí su nombre, UMA: arquitectura de acceso a memoria uniforme.

En la figura de pantalla, podemos observar un ejemplo gráfico de una arquitectura UMA. Normalmente, cada procesador emplea una caché privada (o jerarquía de cachés privadas). Además, todos los procesadores comparten un único controlador de entrada/salida, y por tanto, todos los periféricos, como el disco duro, los periféricos de entrada y los de salida, son comunes a todos los procesadores.

Normalmente, los procesadores se agrupan en *clústeres*, y dentro de cada clúster puede haber una nueva arquitectura organizada como UMA o NUMA. Actualmente, las arquitecturas multiprocesador han llegado a los PC de los hogares, y se habla de [arquitecturas multinúcleo](#).



Modelo UMA

Arquitecturas multinúcleo

Un multinúcleo es una agrupación de procesadores que están situados en el mismo chip (en la misma pastilla).

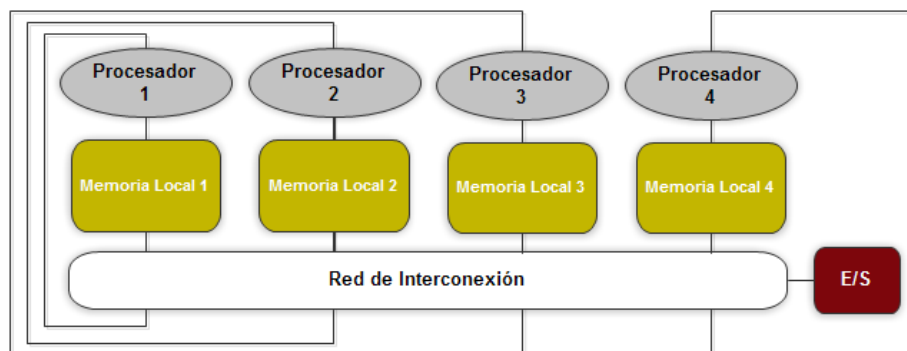
Modelo NUMA

El modelo **NUMA** es un tipo de arquitectura multiprocesador de memoria compartida. En esta arquitectura, todos los procesadores comparten el mismo espacio de direcciones de memoria, pero cada procesador tiene su propia memoria local.

Los procesadores pueden acceder a las memorias locales de otros procesadores, pero el procesador *dueño* de la memoria local, tendrá prioridad de acceso sobre todos los demás.

Cada procesador, por tanto, puede acceder a cualquier posición de memoria de todo el espacio, pero tardará menos tiempo en acceder a su memoria local que en acceder al resto de memorias. Por esta razón, se dice que en este modelo el tiempo de acceso no es uniforme, de ahí su nombre, NUMA: arquitectura de acceso a memoria no uniforme.

En la figura que aparece en pantalla, podemos ver un ejemplo gráfico de una arquitectura NUMA. Normalmente, cada procesador emplea una caché privada (o jerarquía de cachés privadas).



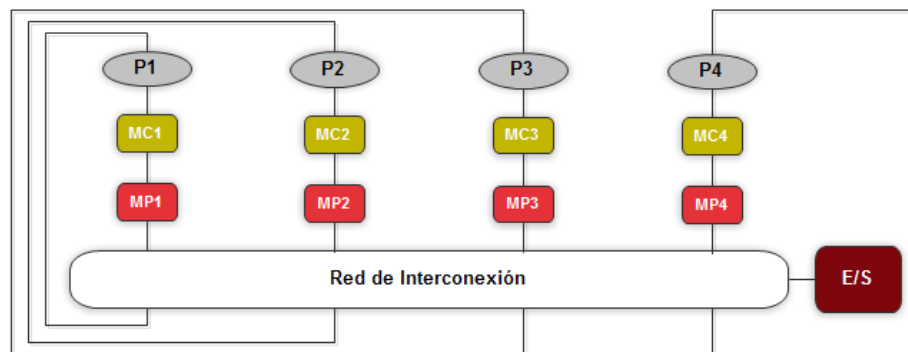
Modelo NUMA

Modelo COMA

El modelo **COMA** (arquitectura de acceso solo a memoria caché) es un caso particular dentro de los NUMA. El acceso a todas las posiciones de memoria no es uniforme, y se tarda menos tiempo en acceder a las memorias locales que a las otras memorias.

El modelo COMA tiene la particularidad de que cada módulo de memoria local se utiliza como memoria caché, y todas las cachés forman un espacio único de direccionamiento global.

En la imagen de pantalla, podemos ver un esquema de arquitectura COMA, donde los módulos de memoria caché son en realidad módulos de memoria principal utilizados como memoria caché (almacenan los datos a los que se accede con mayor frecuencia).

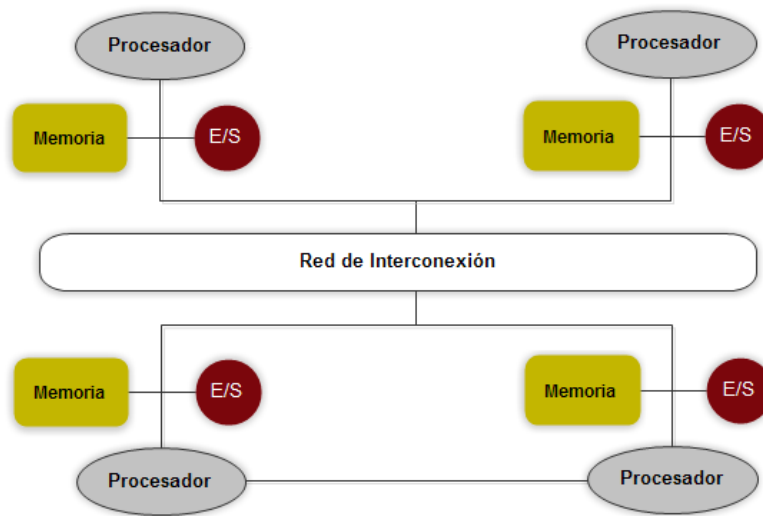


Modelo COMA

Arquitecturas con memoria distribuida

En las arquitecturas multiprocesador con memoria distribuida los procesadores se comunican a través de paso de mensajes. Tenemos que establecer protocolos de comunicación y utilizar rutinas de sincronización, envío de mensajes y recepción de los mismos.

En la imagen siguiente podemos observar una arquitectura multiprocesador de memoria distribuida.



Arquitectura multiprocesador de memoria distribuida

Resumen

Es importante conocer las características que tiene cada tipo de arquitectura. A partir de ahora, cuando se hable de arquitecturas avanzadas, no solo sabremos desenvolvernó con el concepto genérico de multiprocesador, también nos daremos cuenta que hemos asimilado muchos más conceptos y detalles al respecto.

En el siguiente esquema se resume gráficamente toda la clasificación de arquitecturas avanzadas que hemos estudiado en este tema, en base a la clasificación de Flynn y al grado de acoplamiento de los componentes de la arquitectura.

