***Resumen clase 9***

Procesamiento paralelo

Introducción

\*La demanda de máquinas de mayor rendimiento es una exigencia que surgió desde la aparición de las primeras computadoras, y continúa en forma permanente. ➢

\*Existen 2 caminos para aumentar la capacidad de procesamiento:

1. Mejorar el rendimiento de una máquina con un solo procesador.

2. Disponer de sistemas con varios procesadores.

1\_Mejorar el rendimiento de una máquina con un solo procesador requiere:

\*Explotar el paralelismo a nivel instrucción (ILP).

\*Optimizar la detección del paralelismo, a nivel de hardware (MPL).

2-\_Mejorar la capacidad de procesamiento con varios procesadores, requiere:

\*Explotar el paralelismo a nivel proceso.

\* Detección del paralelismo a nivel de sistema operativo, compilador, o programación.

Taxonomía de Computadoras

Taxonomía de Flynn

Existen varias formas distintas de clasificar los Sistemas de cómputo. Una de ellas se basa en determinar las cantidades de flujos de instrucciones y datos que se pueden transferir simultáneamente. En base a este concepto se tienen 4 tipos básicos de máquinas:

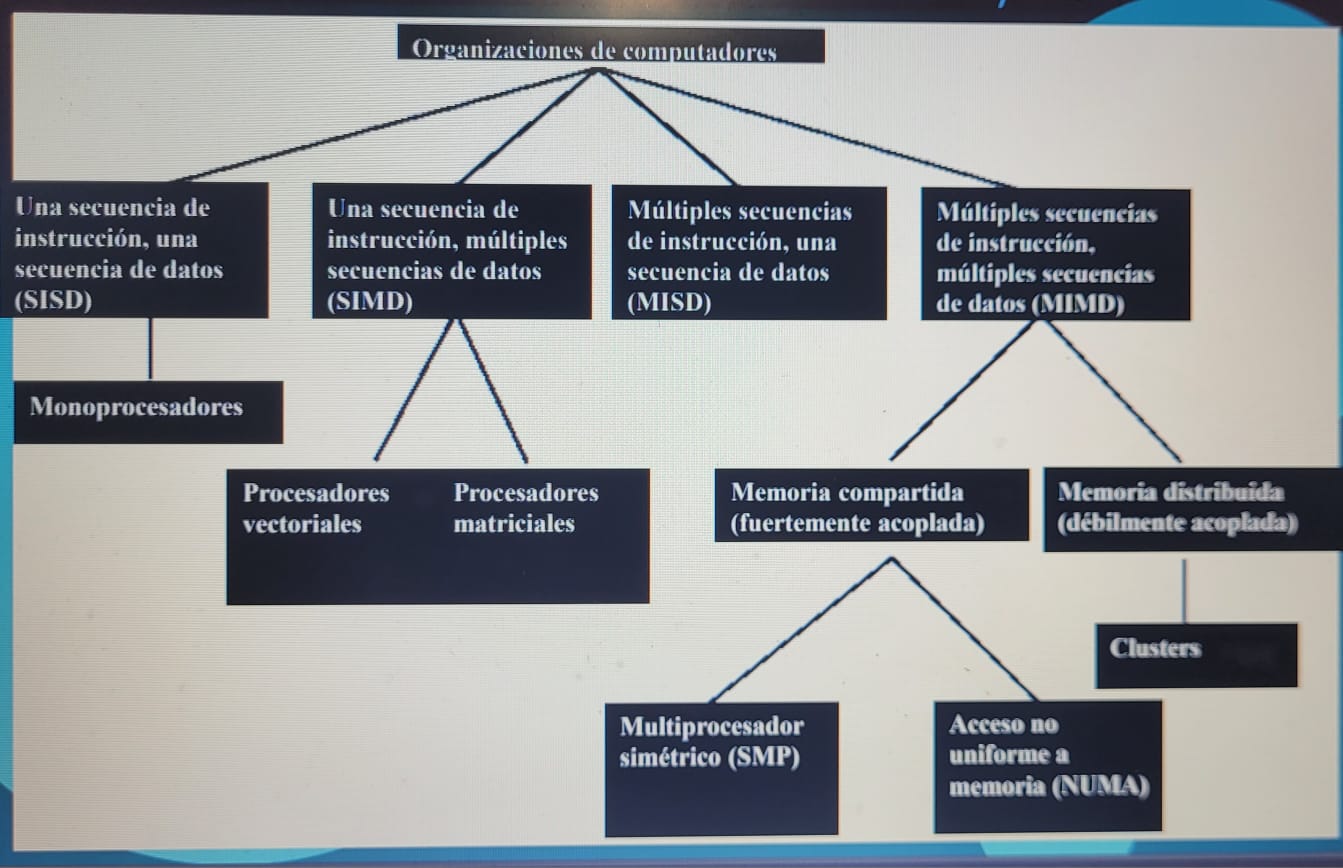
➢ SISD: una secuencia de instrucciones, una secuencia de datos.

➢ SIMD: una secuencia de instrucciones, múltiples secuencias de datos.

➢ MISD: múltiples secuencias de instrucciones, una secuencia de datos.

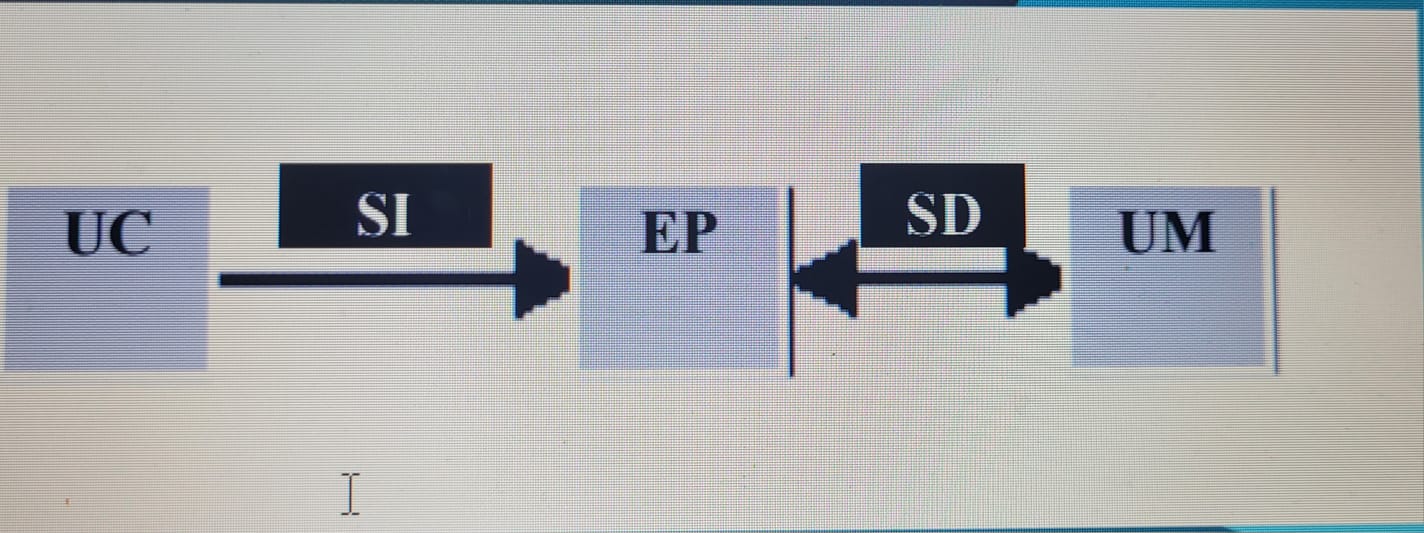
➢ MIMD: múltiples secuencias de instrucciones y múltiples secuencias de datos.

La imagen siguiente muestra los diferentes categorías de Computadoras de acuerdo a la taxonomía de Flynn.



Máquinas SISD

Una máquina tipo SISD (Single Instruction – Single Data), según Flynn, tiene la siguiente estructura funcional.



donde:

➢ UC: Unidad de control (captura de instrucciones)

➢ EP: Elemento de Proceso (ejecuta)

➢ UM: Unidad de Memoria (almacenamiento de datos)

➢ SI: secuencia de instrucciones

➢ SD: secuencia de datos

Es un sistema básicamente compuesto por:

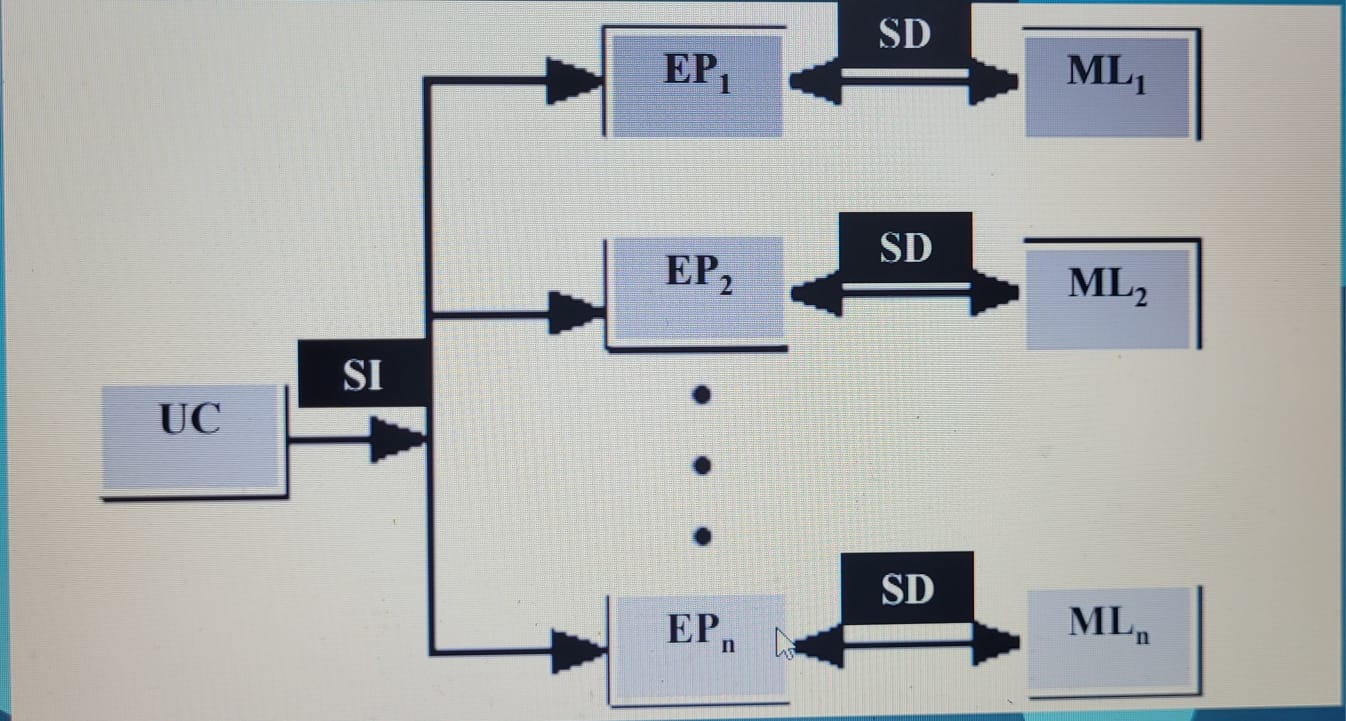
➢ una Unidad de Control (UC)

➢ un elemento de proceso (EP)

➢ una Unidad de Memoria (UM). La UC interpreta (lee) una única secuencia de instrucciones (SI), que el EP resuelve sobre una única secuencia de datos (SD) proveniente de la UM.

➢ Ejemplo máquina SISD: Computadoras monoprocesador.

Máquinas SIMD

******La imagen siguiente muestra el esquema de una máquina tipo SIMD (Single Instruction – Multiple Data).

SIMD

Es un sistema compuesto por:

➢ una Unidad de Control (UC).

➢ una matriz de elementos computacionales o elementos de proceso (EP).

➢ Una Unidad de memoria local (ML) por cada EP.

La UC interpreta una única secuencia de instrucciones (SI) que múltiples EP resuelven simultáneamente cada uno sobre una secuencia de datos (SD) proveniente de su propia ML.

Máquinas SIMD

Cada instrucción es ejecutada por varios elementos de proceso (EP), cada uno con sus propios datos (ML).

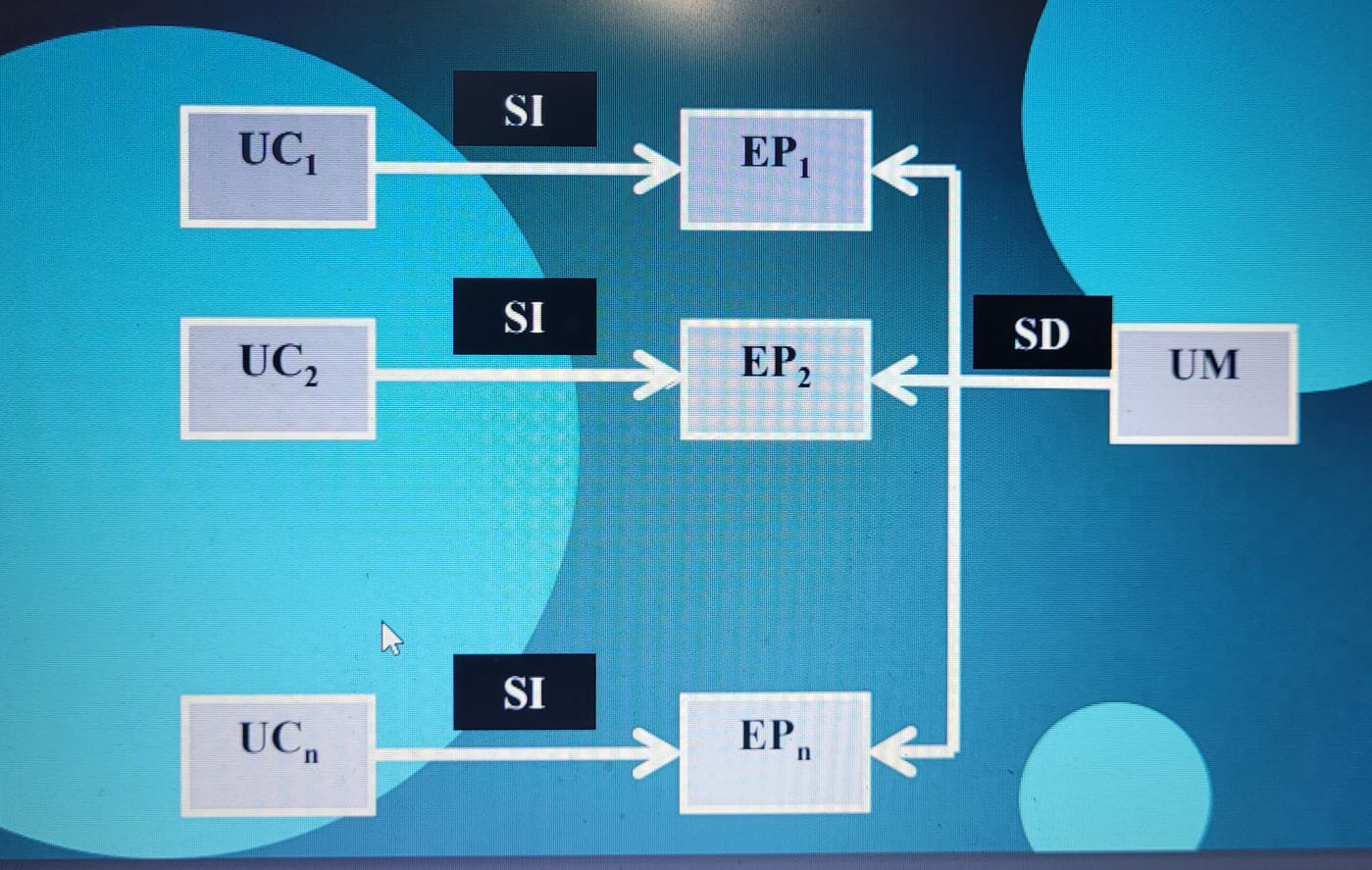
➢ Ejemplo máquina SIMD: Procesadores vectoriales y matriciales. En estas máquinas se requiere disponer de:

➢ Un conjunto ampliado de instrucciones (además de las “convencionales” o escalares) para permitir manejar operaciones vectoriales. Es decir, el repertorio de instrucciones incluye operaciones de suma, almacenamiento, multiplicación, etc., de vectores, para poder asignar a los EPs.

➢ Se deben agregar instrucciones para transferir datos entre EPs, que es un atributo típico de un “lenguaje paralelo”.

Máquinas MISD

La imagen siguiente muestra el esquema de una máquina tipo MISD (Multiple Instruction – Single Data).

******

Máquinas MISD

Es un sistema compuesto por:

➢ una matriz de elementos computacionales o elementos de proceso (EP) cada uno con su propia Unidad de Control (UC).

➢ Una Unidad de memoria (UM) para todos los EP.

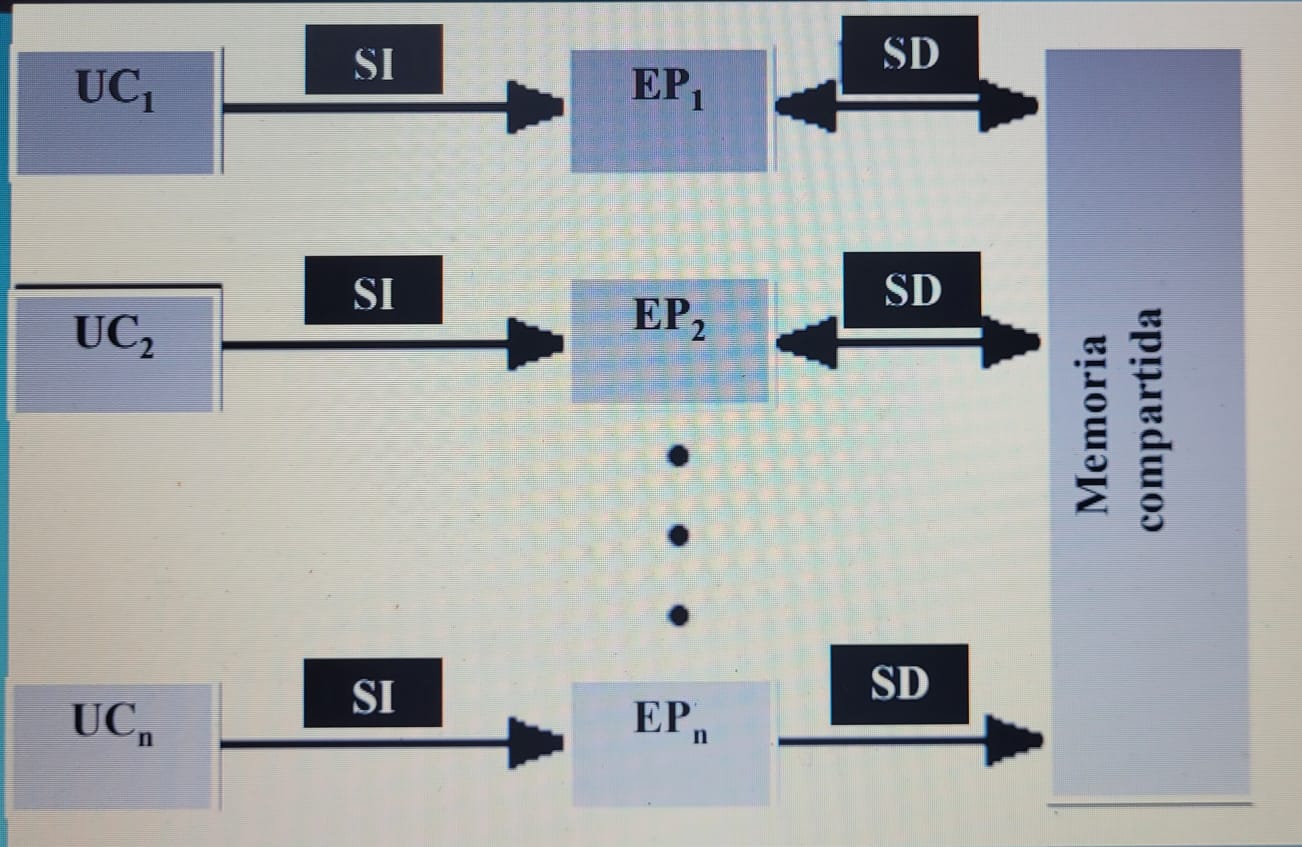
Cada UC interpreta una secuencia de instrucciones (SI) que cada EP resuelve simultáneamente sobre una única secuencia de datos (SD) proveniente de una sola memoria.

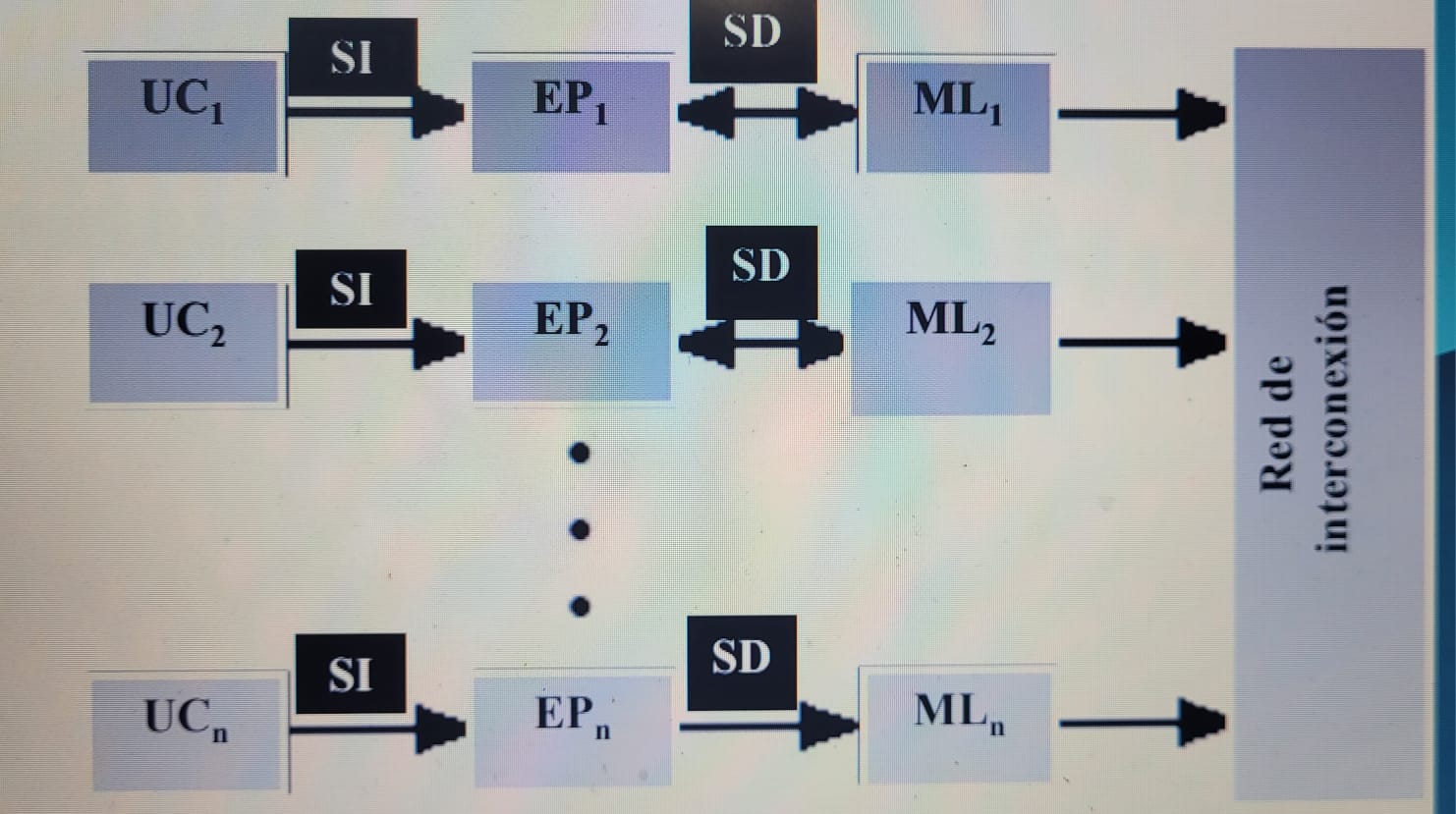
Se transmite una única secuencia de datos (SD) a un conjunto de procesadores cada uno ejecutando su propia secuencia de operaciones (MI). Cada procesador ejecuta una secuencia de instrucciones diferente.

➢ Ejemplo: es un modelo teórico, no hay registros de máquinas de este tipo (hay algunas de características similares pero no cumplen los requisitos para ser exactamente tipo SIMD).

Máquinas MIMD

Existen 2 modelos de Computadoras tipo MIMD (Multiple Instruction – Multiple Data). Un modelo es MIMD de Memoria compartida.

******

******El segundo tipo de Computadoras tipo MIMD es MIMD de Memoria distribuida.

Es un sistema compuesto por:

➢ Múltiples procesadores (UC), ejecutando

➢ Múltiples secuencias de instrucciones (MI), sobre

➢ Múltiples secuencias de datos (MD).

Se pueden dividir según la organización de la Memoria (que es también la forma de comunicarse) en:

➢ MIMD de Memoria compartida, con 2 variantes:

➢ SMP (multiprocesadores simétricos)

➢ Sistemas NUMA

➢ MIMD de Memoria distribuida, comúnmente llamados Clusters.

Multiprocesadores Simétricos SMP

Multiprocesadores SMP

El modelo de computadora tipo SMP se muestra en la figura siguiente.

Es un sistema con las siguientes características principales:

➢ Dos o más procesadores idénticos (o muy similares) de capacidades comparables (homogéneos).

➢ Memoria principal y E/S única (compartida) por todos los procesadores.

➢ Interconexión mediante un bus u otro tipo de medio similar (“fuertemente acoplados”).

➢ Igual tiempo de acceso a la memoria para todos los procesadores.

Por eso se los identifica como del tipo “UMA”: Uniform Memory Access.

➢ Todos los procesadores pueden desempeñar las mismas funciones.

➢ Sistema operativo integrado, que proporciona la interacción entre los procesadores y sus programas.

Multiprocesadores Simétricos SMP

Ventajas

➢ Mayores prestaciones: en general tienen buenos resultados, si las tareas pueden organizarse en paralelo.

➢ Buena disponibilidad: un fallo en un procesador no detiene la operación del sistema, dado que todos los procesadores pueden hacer las mismas tareas.

➢ Crecimiento: pueden aumentarse las prestaciones añadiendo más procesadores, pero hay restricciones a este mecanismo.

➢ Escalado: normalmente limitado, en función de la cantidad de procesadores.

Multiprocesadores Simétricos SMP

Desventajas

Los principales problemas se originan en los conflictos por el bus compartido, la coherencia y consistencia de los datos, y la sincronización de tareas entre procesadores.

➢ Las prestaciones están limitadas por el tiempo de ciclo del bus.

➢ Cada procesador está equipado con una memoria cache, que reduce los accesos a memoria y mejora las prestaciones. Pero eso trae un problema.

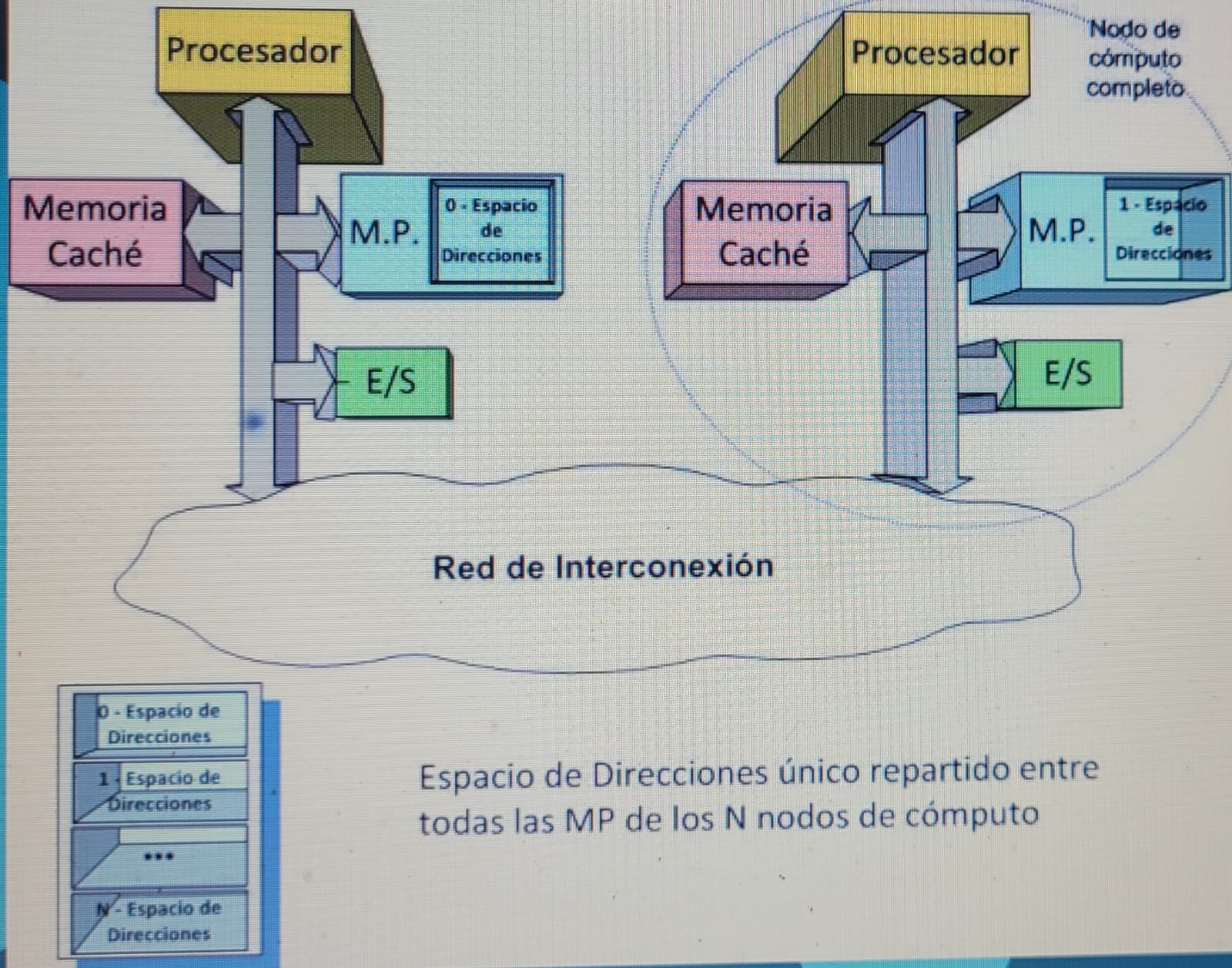
➢ Al disponer una caché en cada procesador, se pueden producir problemas de coherencia de cache (datos que pueden estar en más de una caché).

Por razones de velocidad, este problema debe ser resuelto por el hardware. Se requieren usar protocolos para la administración de los datos en las caché (protocolos de coherencia tipo sondeo o snoopy).

Multiprocesadores NUMA

Multiprocesadores con memoria compartida y distribuida – NUMA

El modelo de computadora tipo NUMA se muestra en la figura siguiente.



Es un sistema con las siguientes características principales:

➢ Dos o más procesadores idénticos (o muy similares) de capacidades comparables (homogéneos) formando “nodos”.

➢ Cada nodo es un procesador completo con su propia memoria local y E/S. Los nodos tienen características muy similares.

➢ Los nodos están interconectados mediante una red de interconexión.

➢ El espacio de direcciones (memoria lógica) es común a todos los procesadores, por eso es un sistema de memoria compartida.

Pero el tiempo de acceso al espacio de memoria es distinto para el acceso a la memoria local que para el acceso a las memorias de los otros nodos.

➢ Por eso se los llama NUMA: Non Uniform Memory Access.

➢ Todos los procesadores pueden desempeñar las mismas funciones.

➢ El Sistema operativo está integrado, y proporciona la interacción entre los procesadores y sus programas.

Cluster

Multiprocesadores de memoria distribuida - Cluster

El modelo de una arquitectura tipo Cluster se muestra en la figura siguiente

Es un sistema con las siguientes características principales:

➢ Compuesto por 2 o más nodos.

➢ Cada nodo es un procesador completo con su propia memoria local y E/S.

Los nodos pueden tener características similares (“homogéneos”) o distintas (“heterogéneos”).

➢ Interconectados mediante una red de interconexión. Debido al tipo de red y protocolo de comunicación se los considera del tipo “levemente acoplados”.

➢ Como los espacios de direcciones son independientes es un sistema de Memoria distribuida.

➢ Todos los procesadores pueden desempeñar las mismas funciones.

➢ El Sistema operativo está integrado, y proporciona la interacción entre los procesadores y sus programas.

➢ La comunicación entre procesos es en base a mensajes (a resolver por el programa).

➢ Los Cluster, básicamente, son computadoras completas, interconectadas, que trabajan conjuntamente como un único recurso.

➢ Es decir, para las tareas en ejecución se comportan como si fuera una única máquina.

➢ Como ya se dijo, cada computadora se denomina “nodo”.

➢ En general, presentan prestaciones y disponibilidad elevadas.

➢ Las aplicaciones son propias de un servidor, y constituyen una alternativa a los SMP.

Ventajas

➢ Escalabilidad absoluta: dependiente de la cantidad de nodos incorporados se puede disponer de mayores prestaciones.

➢ Escalabilidad incremental: posibilidad de agregar nuevos nodos fácilmente.

➢ Alta disponibilidad: capacidad de seguir operando con nodos en falla.

➢ Mejor relación precio/prestaciones: porque se usan equipos de cómputo estandar (y, posiblemente, de bajo costo).

Cluster vs SMP

SMP

➢ Permiten dar soporte a aplicaciones de alta demanda de recursos.

➢ Disponibles comercialmente (SMP es más antiguo).

➢ Más fácil de administrar y configurar.

➢ Cercano a los sistemas de un solo procesador.

➢ La planificación (scheduling) es la diferencia principal

➢ Menos espacio físico / Menor consumo de potencia

➢ SMP tiene límite práctico en su número de procesadores: entre 16 y 64 por degradación de prestaciones.

Cluster

➢ Permiten dar soporte a aplicaciones de alta demanda de recursos.

➢ Disponibles comercialmente (SMP es mas antiguo).

➢ Superior escalabilidad incremental y absoluta

➢ Superior disponibilidad.

➢ Mayor redundancia

➢ En Clusters cada nodo tiene su propia memoria principal. Así las aplicaciones no ‘ven’ la memoria global.

➢ La coherencia es mantenida por software y no por hardware.

➢ Alternativa a sistemas SMP para brindar multiprocesamiento a gran escala. Por ejemplo, el SGI Origin de Silicon Graphics es un NUMA con 1024 procesadores MIPS R10000.

TIPOS DE ACCESOS A MEMORIA: UMA, NUMA, CC-NUMA

Los sistemas multiprocesadores se pueden clasificar de acuerdo al tipo de acceso a memoria en:

1. UMA - Uniform memory access

➢ Igual tiempo de acceso a todas las regiones de memoria. ➢ Igual tiempo de acceso a memoria para los diferentes procesadores. Ej: SMP

1. NUMA - Non-uniform memory access

➢ El tiempo de acceso de un procesador difiere, dependiendo de la región de memoria que accede.

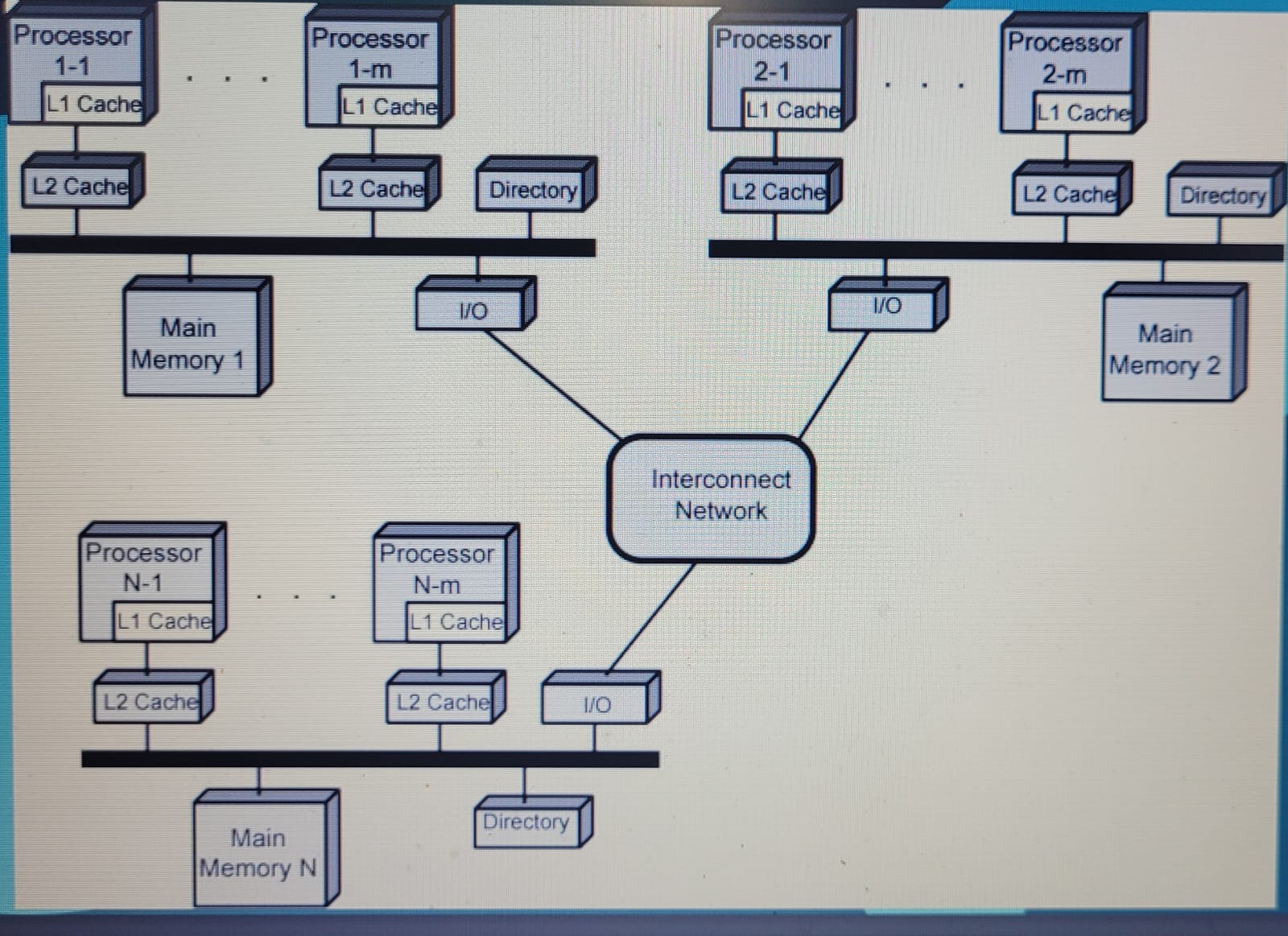
➢ Diferentes procesadores acceden a diferentes regiones de memoria a diferentes velocidades. Ej: Cluster

1. CC-NUMA - cache coherente NUMA

➢ Es un NUMA que mantiene coherencia de cache entre las cache de los distintos procesadores. Ej: Sistemas con memoria compartida distribuida.

Sistemas CC-NUMA

En la imagen siguiente se muestra el modelo de organización de un sistema NUMA con coherencia en las caché (CCNUMA).



Los sistemas CC-NUMA presentan las siguientes características:

➢ Cada nodo tiene 2 o más procesadores (por ejemplo un SMP), cierta cantidad de memoria (principal) y E/S.

➢ Cada procesador tiene su cache (típicamente L1 y L2). ➢ Los nodos están interconectados por algún tipo de red.

➢ Existe un espacio de direcciones de memoria único para todos los procesadores de todos los nodos.

➢ Orden de acceso a memoria: ➢ cache L1 (local al procesador).

➢ cache L2 (local al procesador).

➢ Memoria principal (local al nodo).

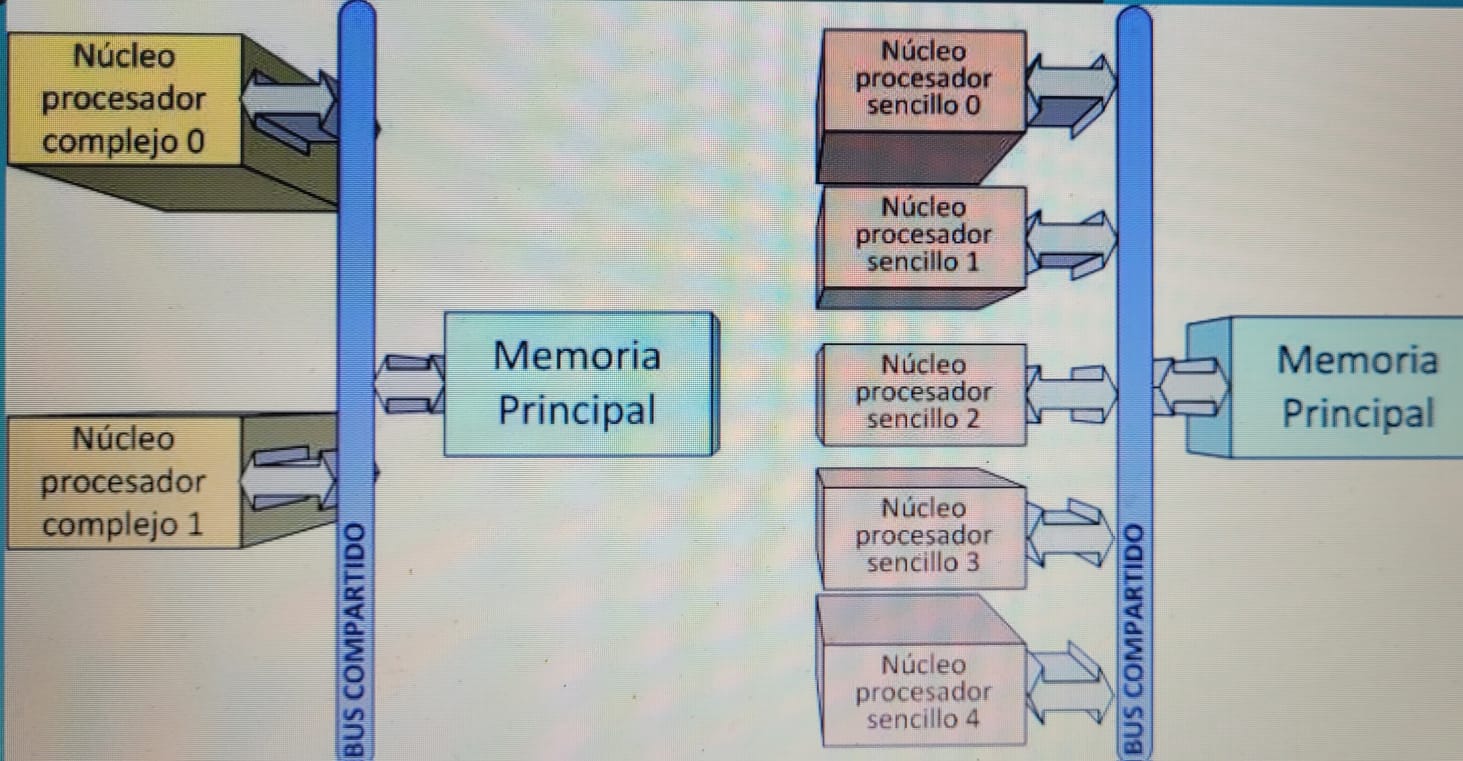
➢ Memoria remota (petición por red).

➢ La coherencia se mantiene en forma automática y transparente

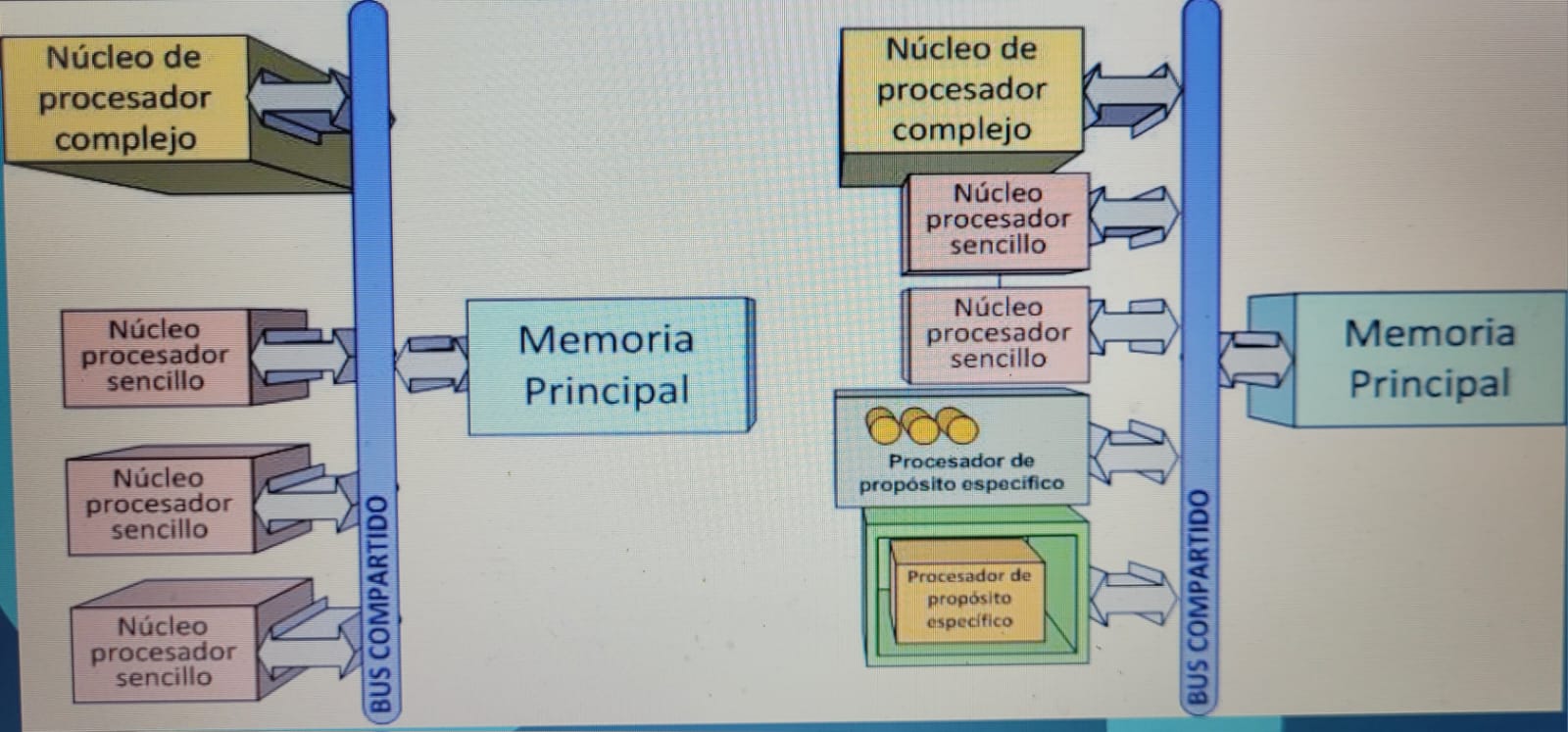
Arquitecturas on chip con Memoria compartida

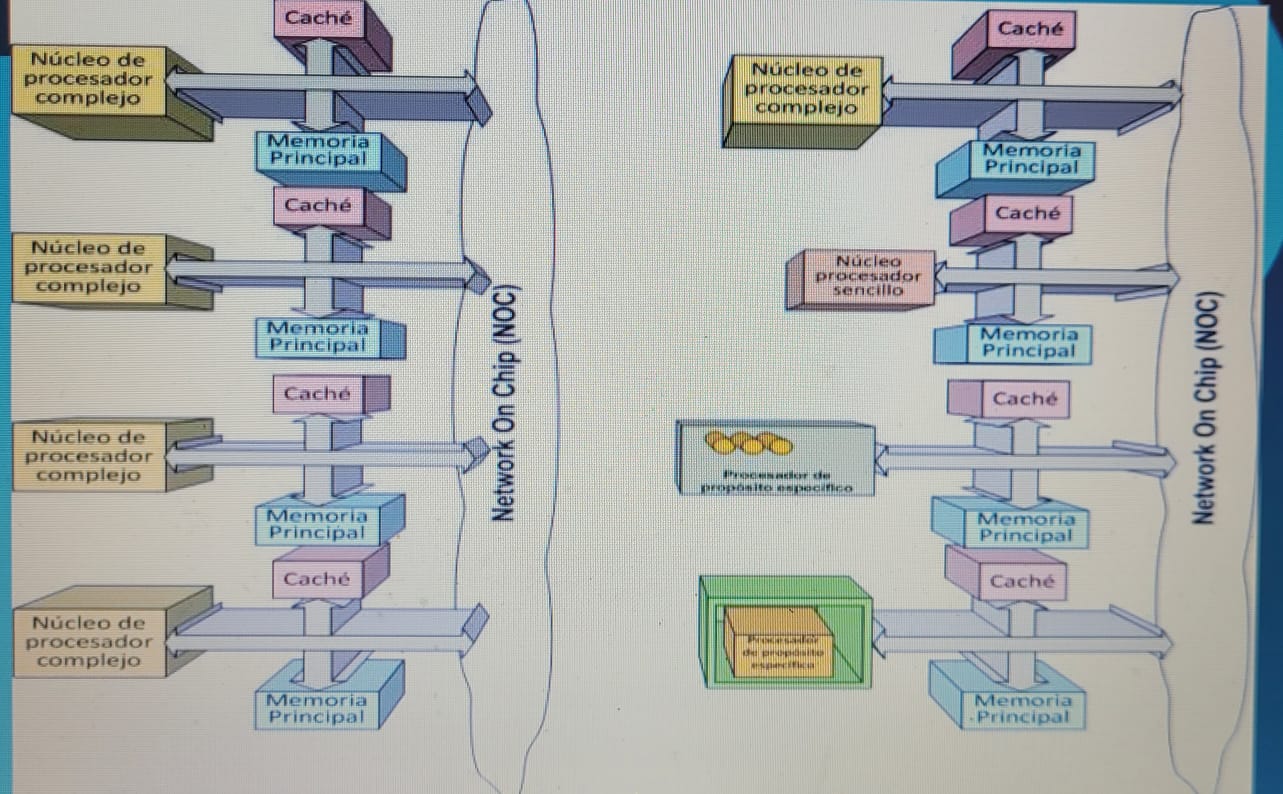
Las arquitecturas on-chip con memoria compartida pueden ser homogéneas o heterogéneas.

Son homogéneas si todos los procesadores son idénticos.



Son heterogéneas si los procesadores tienen distintas prestaciones





Arquitecturas on chip con Memoria distribuida

Las arquitecturas on chip con memoria distribuida también pueden ser homogéneas o heterogénas.