Trabajo sobre "Interconnection networks" y también de la Ley de Amdahl y Gustafson

Iñigo Manuel Diez Canseco Fuentes Univesidad Católica San Pablo Arequipa, Perú Email: inigo.diezcanseco@ucsp.edu.pe

I. Introducción - Interconnection Networks

La interconexión juega un papel decisivo en el desempeño tanto de los sistemas distribuidos como de los sistemas de memoria compartida.

II. MEMORIA COMPARTIDA O Shared-memory interconnects

En este modelo, todos los procesadores de un sistema comparten un espacio de memoria común. La comunicación entre procesadores se realizan por operaciones de lectura/escritura en las celdas de memoria compartida a través de *interconnection networks*.

Existen arquitecturas para la organización de la estructura de procesadores paralelos que utilizan memoria compartida.

■ Arquitectura **UMA**: Significa *Uniform Memory Access*, trata de que todos los procesadores se conectan a la memoria a través de un bus[1]. Su propósito es que todos los procesadores compartan la memoria física uniformemente.

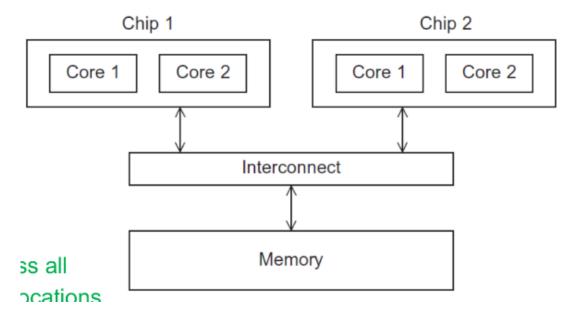


Figura 1. Uniform Memory Access, Referencia de las diapositivas.

■ Arquitectura **NUMA**: Significa *Non Uniform Access*, como la memoria compartida depende de la ubicación de cada elemento del proceso a través de una red y a la vez dentro de la memoria[2].

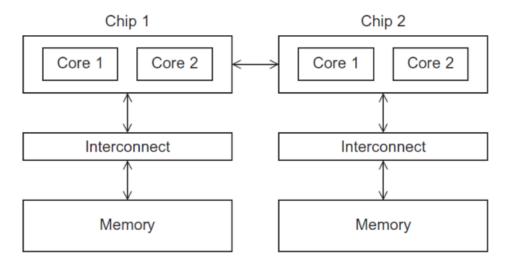


Figura 2. Non Uniform Access, Referencia de las diapositivas.

Dentro de la estructura de la memoria compartida existe la conexión por bus compartida, esta organización utiliza los buses como medios con los que se conectan todos los dispositivos o procesos relacionados con cada una de las arquitecturas mencionadas.

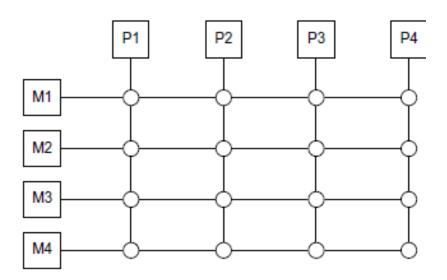


Figura 3. Un interruptor de barra transversal que conecta cuatro procesadores.

Se utiliza además switches para alcanzar un control utilizando además un crossbar.

Crossbars permiten simultáneamente la comunicación entre diferentes dispositivos, entonces ellos son mucho más rápidos que los buses[3].

III. MEMORIA DISTRIBUIDA O Distributed-memory interconnects

Las interconexiones a menudo se dividen en dos grupos.

III-A. Direct interconnects

En una directa interconexión, cada interruptor es directamente conectado con el procesador de la memoria de diferentes orientaciones, y los *switches* están conectados entre sí.

Sus características más importantes son:

- Topología de Red: Determina el patrón de conexión de nodos.
- Grafo de un Nodo: Número de enlaces o canales que inciden en el nodo.
- Escalabilidad: Posibilidad de expansión modular de la red.

Poseen distintas topologías tales como:

- Anillo.
- Las Mallas/Toroidal.
- Estrella.
- Panal.
- Árbol.
- Anillo Cordado.
- Etc.

Algunos de estas distintas topologías son como los *Anillos*, donde cada *switch* tiene 3 *links* o 2 *links* para cada procesador que se utiliza, o también la Malla Toroidal, donde cada switch tiene 5 *links* para cada procesador que se utiliza.

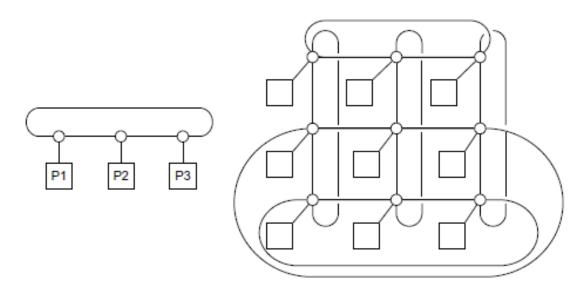


Figura 4. Un Anillo y una Malla toroidal.

III-B. Indirect interconnects

Provee una alternativa a las interconexiones directas, en una interconexión indirecta, los *switches* no pueden estar conectados directamente a un procesador[3]. Se muestra con enlaces unidireccionales y una colección de procesadores, cada uno de los cuales tiene un enlace saliente y uno entrante y una *switching network*.

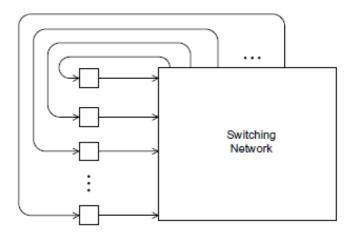


Figura 5. Una genérica red indirecta.

Se construyen usando *switches* y enlaces de comunicación. No poseen comunicación directa entre nodos, por lo que esta se realiza entre 1 o más *switches* y cada *switch* contiene un conjunto de puertos.

IV. LATENCY AND BANDWIDTH

La *latency*, es el tiempo que transcurre entre que la fuente comienza a transmitir la data y el destino comienza a recibir el primer *byte*.

El *bandwidth* es la velocidad a la que un enlace puede transmitir datos. Usualmente, dado en megabits o megabytes por segundo. Es la velocidad a la que el destino recibe datos después de haber comenzado a recibir el primer *byte*.

V. LEY DE AMDAHL

La ley formulada por Gene Amdahl[4], es utilizada para averiguar la mejora máxima de un sistema de información cuando solo una parte de este es mejorado.

La definición exacta de esta ley establece que: «La mejora obtenida en el rendimiento de un sistema debido a la alternación de uno de sus componentes está limitada por la fracción de tiempo que se utiliza dicho componente».

Según el Libro de Peter Pacheco[3], Gene Amdahl dice que a menos que prácticamente cada programa y tarea este en paralelo, entonces la mejora en los tiempos van a estar muy limitados; independientemente del número de núcleos que se tiene disponible.

Entonces, la fórmula de la ley de Amdahl y su descripción son los siguientes:

$$Tm = Ta * ((1 - Fm) + \frac{Fm}{Am}) \tag{1}$$

- Fm: Porcentaje de la mejora que el sistema está utilizando.
- *Am*: Factor de mejora específica del sistema.
- *Ta*: Tiempo actual del programa antes de la mejora.
- *Tm*: Tiempo máximo de mejora de ejecución.

Sirve para definir si introducir una mejora en su sistema, en otras palabras, si merece o no la pena.

Dado que el incremento de velocidad de un programa utilizando múltiples núcleos de un procesador da igual, ya que en computación distribuida está limitada por la fracción secuencial del programa, con la Ley de Amdahl

podrán ya no solo calcular, sino ver gráficamente si el mejorar el rendimiento de uno u otro componente merecerá o no la pena, y hasta qué punto.

VI. LEY DE GUSTAFSON

La ley de Gustafson[5], es una ley en ciencia de la computación que establece que cualquier problema suficientemente grande puede ser eficientemente paralelizado.

El impacto de la ley de Gustafson fue el cambio de dirección de los objetivos[6] de investigación hacia la selección o reformulación de problemas a fin de que fuera posible la solución de mayores problemas en el mismo intervalo de tiempo. Redefine la eficiencia como una necesidad para minimizar la parte secuencial de un programa, incluso si esto incrementa la cantidad total de cálculos.

Según el Libro de Peter Pacheco[3], se menciona que hay miles de programas utilizados que obtienen grandes aceleraciones en sistemas de memoria distribuida, pero la obtención de aceleración es siempre adecuado.

$$S(P) = P - a * (P - 1) \tag{2}$$

- S(P): Es la mejora obtenida a partir de la cantidad de núcleos o procesadores.
- a: Es la parte no paralelizable del proceso del programa que se quiere paralelizar.
- *P*: Es el número de procesadores del computador utilizado.

VII. CUADRO COMPARATIVO

A continuación se presentará un cuadro comparativo entre la Ley de Amdahl y la Ley de Gustafson:

	Ley de Amdahl	Ley de Gustafson
Propuesta	Averigua la mejora máxima de un sistema de información cuando solo una parte de este es mejorado.	Establece que cualquier problema suficientemente grande puede ser eficientemente paralelizado.
Busca	Obtener el tiempo máximo de mejora como límite de la mejora en el tiempo de ejecucion.	Obtener la mejora del tiempo de ejecucion y establecer que cualquier programa puede ser eficientemente paralelizado.
Fórmula	Considera el tiempo de ejecucion antiguo y el mejorado.	Considera los procesadores y la parte no paralelizable.
Límites	No considera la cantidad de procesadores.	Limitaciones en el hardware y la cantidad de procesadores.

Figura 6. Cuadro comparativo[7].

REFERENCIAS

- [1] Uniform memory access. [Online]. Available: https://en.wikipedia.org/wiki/Uniform_memory_access#:~:text=Uniform%20memory%20access%20(UMA)%20is,share%20the%20physical%20memory%20uniformly.
- [2] Non uniform access. [Online]. Available: https://es.wikipedia.org/wiki/NUMA
- [3] P. Pacheco, An introduction to parallel programming. Elsevier, 2011.
- [4] Ley de amdahl. [Online]. Available: https://es.wikipedia.org/wiki/Ley_de_Amdahl
- [5] Ley de gustafson. [Online]. Available: https://es.wikipedia.org/wiki/Ley_de_Gustafson
- [6] Especificaciones de la ley de gustafson. [Online]. Available: https://www.ecured.cu/Ley_de_Gustafson
- [7] Reevaluating amdahl's law. [Online]. Available: https://web.archive.org/web/20070927040654/http://www.scl.ameslab.gov/Publications/Gus/Amdahls.html