



# Procesamiento Paralelo

- La computación con **procesador en paralelo** lleva con nosotros bastantes años, concretamente desde las primeras implementaciones IBM 704 en 1954, fue hasta el 2005 cuando este paradigma de la informática llegó al público mediante el procesador multinúcleo,.

- Un procesador en paralelo es el método mediante el cual una serie de tareas e instrucciones se ejecutan de forma simultánea, se trata de dividir el trabajo en trozos más simples, que se llaman hilos, threads o subprocesos. Cada uno de los subprocesos se ejecutan en uno de los núcleos del procesador de forma simultánea para aminorar el tiempo de espera entre tarea y tarea.

Algunos ejemplos de procesadores en paralelo serían nuestros móviles cuentan con procesadores multinúcleo capaces de ejecutar varias tareas, así mismo como las computadoras e incluso los televisores inteligentes

# Ley de Moore, y el fin del escalado por frecuencia

- El procesamiento en paralelo no se trata simplemente de añadir núcleos a la CPU o de conectar y sincronizar procesadores de distintos nodos.
- Debido a la complejidad en la implementación, los procesadores de consumo general se basan en el aumento de frecuencia sobre un solo núcleo (en cada núcleo se lleva a cabo una tarea solamente) una vez terminada la tarea (tiempo de ejecución)
- ,esta variable es el resultado de multiplicar el número de instrucciones por el tiempo de ejecución promedio, lo único que modificara sería el aumento de frecuencia.
- La fórmula sería :

$$P = C \cdot V^2 \cdot F \text{ donde:}$$

*P es la potencia, C cambio de capacitancia por ciclo de reloj, V tensión y F la frecuencia*

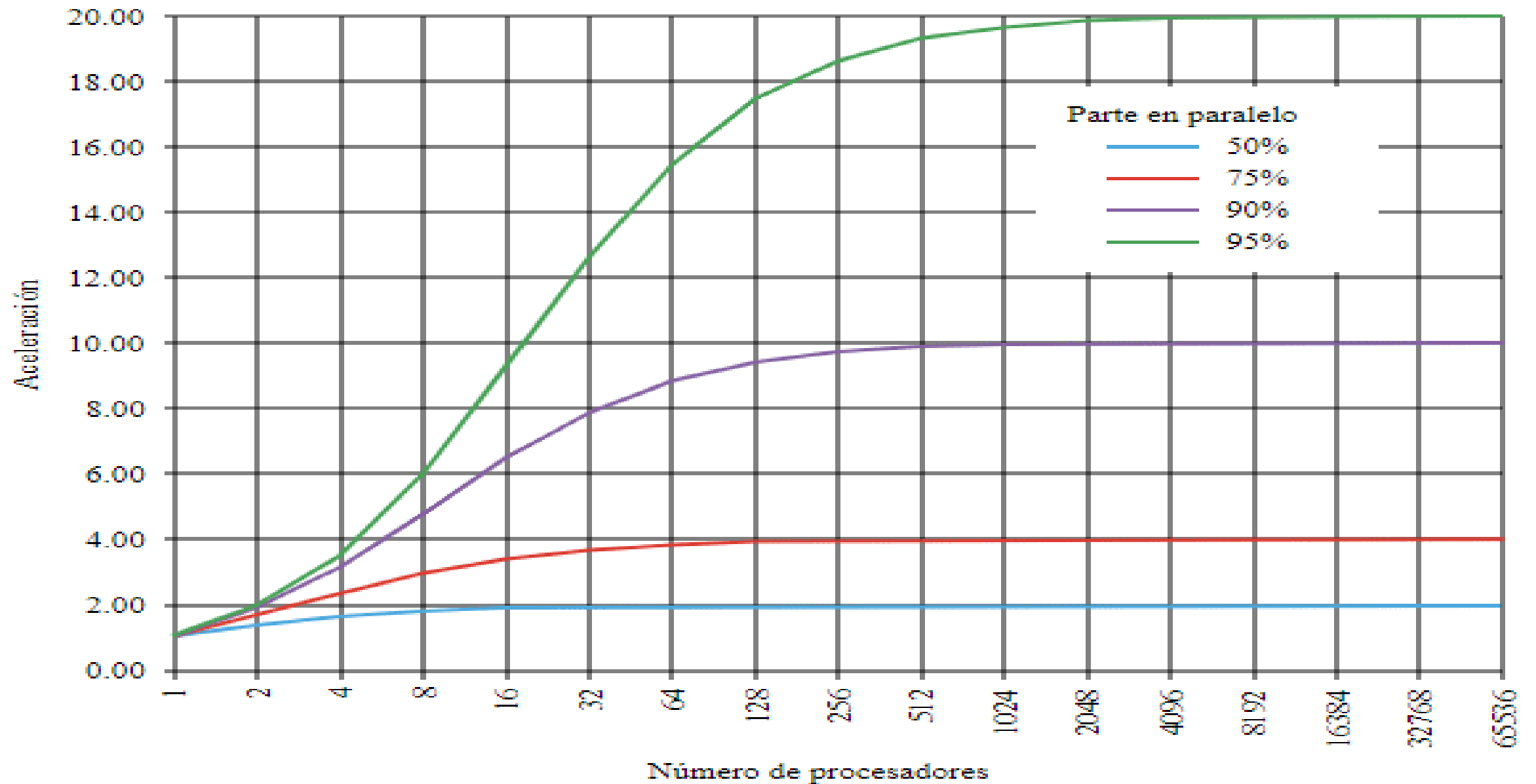
- un **nodo** es cada uno de los elementos de una lista enlazada, un árbol o un grafo en una estructura de datos. Cada **nodo** tiene sus propias características y cuenta con varios campos; al menos uno de éstos debe funcionar como punto de referencia para otro **nodo**.

- Si aumentamos la frecuencia se aumenta el consumo y el calor, la solución sería disminuir el voltaje que influye en la potencia. el voltaje se disminuye haciendo transistores mas pequeños que hagan lo mismo con menos energía, justo aquí es donde entra la ley de Moore, un señor que demostró que la densidad de transistores se duplica cada 18 o 24 meses en los procesadores debido a la disminución del proceso de fabricación, esto ha sido así ,al menos en AMD, ya que Intel lleva demasiado tiempo atascado en los 14nm.
- En 2004 los procesadores de un solo núcleo tocaron techo a unos 4,8 GHz de forma estable y es cuando el paradigma cambio por completo ,AMD fue el primero en lanzar su Athlon x2 para consumo general ,Intel haría lo propio con el Pentium-D. La ley de Moore sigue teniendo vigencia ,pero la cantidad extra de transistores se destinan ,no al aumento de frecuencia, sino a aumentar núcleos e implementar nuevas funciones, de esta forma es como hemos llegado a tener un procesador en paralelo dentro de un mismo silicio.

# Ley de Amdahl: mas es mejor, siempre que la programación sea optima

- Una ley fundamental esta sostiene que la mejora de velocidad en la ejecución de un programa debido a la paralelización de sus procesos estará limitada por el porcentaje de programa que no se pueda paralelizar.
- Es decir si un cierto porcentaje del programa en ejecución funciona solamente con tareas en serie.
- Ejemplo un 50%, la aceleración máxima seria de 2x y no importa cuantos núcleos añadamos, habrá una cierta cantidad en la que no obtengamos beneficio.
- La grafica muestra la ley de Amdahl:

## Ley de Amdahl



- Mientras mas paralelizado este el programa, mas veloz y mas susceptible de añadirle núcleos para aumentar su rendimiento será junto a la otra ley complementaria de Gustafson afirma que el tiempo de funcionamiento de la parte secuencial del programa (la no paralela) es independiente de la cantidad de procesadores.
- Para obtener el mejor rendimiento de un procesador en paralelo, el programa debe también ser lo mas paralelizado posible .

# Niveles de procesamiento en paralelo

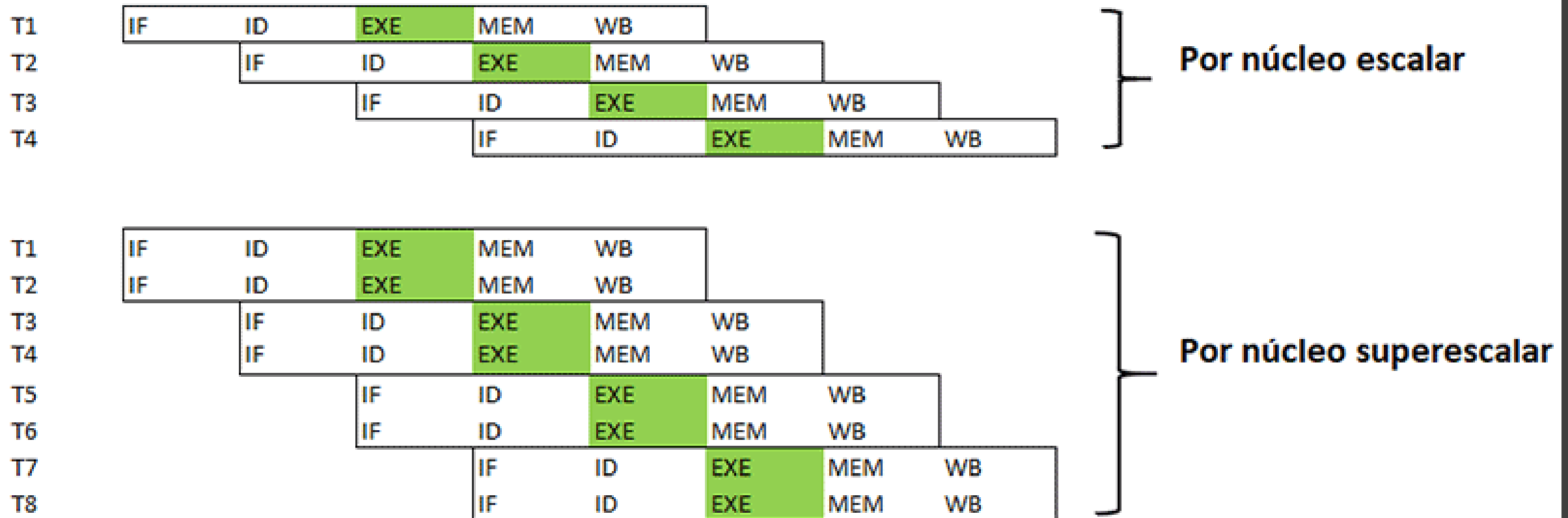
- El ordenador tiene un procesador en paralelo, de hecho tiene muchos mas como son la GPU o el chipset aunque se dediquen a otras tareas como los gráficos en el primer caos o periféricos en otro.
- Existen básicamente cuatro niveles de procesador en paralelo(en bit):





- La paralelización a nivel mas bajo es la que se realiza en las cadenas de bits. Toda instrucción está formada por palabras, que en definitiva son estas cadenas de bits que transportan la información. El contar con palabras más largas, permite añadir más contenido a la instrucción, y en consecuencia hacer más trabajo de una sola vez. En la actualidad todos los procesadores trabajan con palabras de 64 bits, mientras que antes permitían solo 32 bits.
- Paralelización por instrucción:

# 5 Pipeline



IF = Petición  
ID = Decodificación  
EXE = Ejecución  
MEM = Acceso a memoria  
WB = Escritura

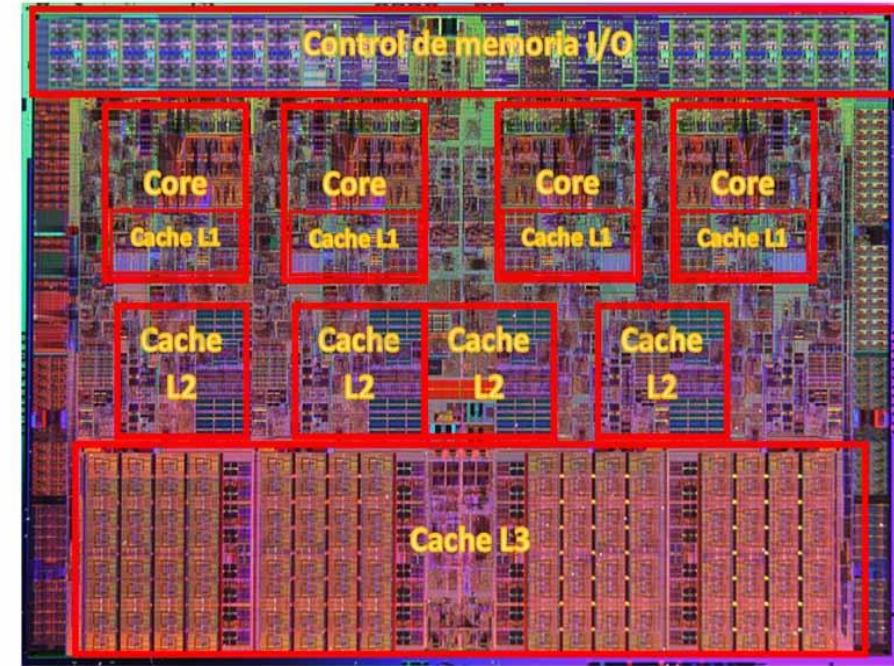
- El siguiente nivel de procesador en paralelo radica en **paralelizar instrucciones**. En instrucciones sin dependencias entre ellas se puede decir eso de “el orden de los factores no altera el producto”. Las instrucciones se **combinan en grupos para ser procesadas de forma paralela** en distintas etapas o “**pipeline**”.
- El **pipeline** es la capacidad del procesador de **trabajar con varias instrucciones a la vez**, cada una de ellas situadas en una **etapa distinta**. Cuando nos referimos a etapas son las típicas de la arquitectura RISC: **pedir instrucción -> decodificar -> ejecutar -> acceso a memoria -> escritura** que estudiamos en asignaturas de computación o informática.

Microarchitecture	Pipeline stages
P5 (Pentium)	5
P6 (Pentium 3)	10
P6 (Pentium Pro)	14
NetBurst (Willamette)	20
NetBurst (Northwood)	20
NetBurst (Prescott)	31
NetBurst (Cedar Mill)	31
Core	14
Bonnell	16
Sandy Bridge	14
Silvermont	14 to 17
Haswell	14
Skylake	14
Kabylake	14

- Los procesadores Intel Core suelen tener unos pipeline, los Ryzen de AMD tiene 19 ,son procesadores todos superescalares, son capaces de ejecutar mas de una instrucción a la vez gracias a sus núcleos y sus hilos de procesamiento.
- Paralelismo de datos: lo que pretende dividir es la entrada de datos de un programa, que trasladado a nivel de procesador consistiría en asignar un subconjunto de datos a cada procesador o núcleo para que ejecuten la misma secuencia de operaciones.
- Paralelismo de tareas: un programa capaz de entregar tareas al procesador que son totalmente distintas unas de otras para que se realicen en paralelo.(es un programa paralelizado).

# Tipos de procesador en paralelo

- Procesadores multinúcleo. inicio en 2004 es un procesador multinúcleo tenemos mas de una unidad de ejecución, lo que llamados núcleos o Cores.
- Un procesador con varios núcleos puede ejecutar tantas instrucciones de forma paralela con núcleos tenga. Forma secuencial y de distinta naturaleza.
- Se une la condición de procesador superescalar, donde cada núcleo puede ejecutar mas de una tarea en un solo ciclo de reloj, es un flujo de instrucciones mas simple que optimiza y disminuye los tiempos de espera de cada núcleo.
- Esta tecnología se le denomina Multithreading, y se aplica en forma de HyperThreading en Intel y SMT en AMD, creando procesadores con dos hilos o procesadores virtuales por cada núcleo físicos.



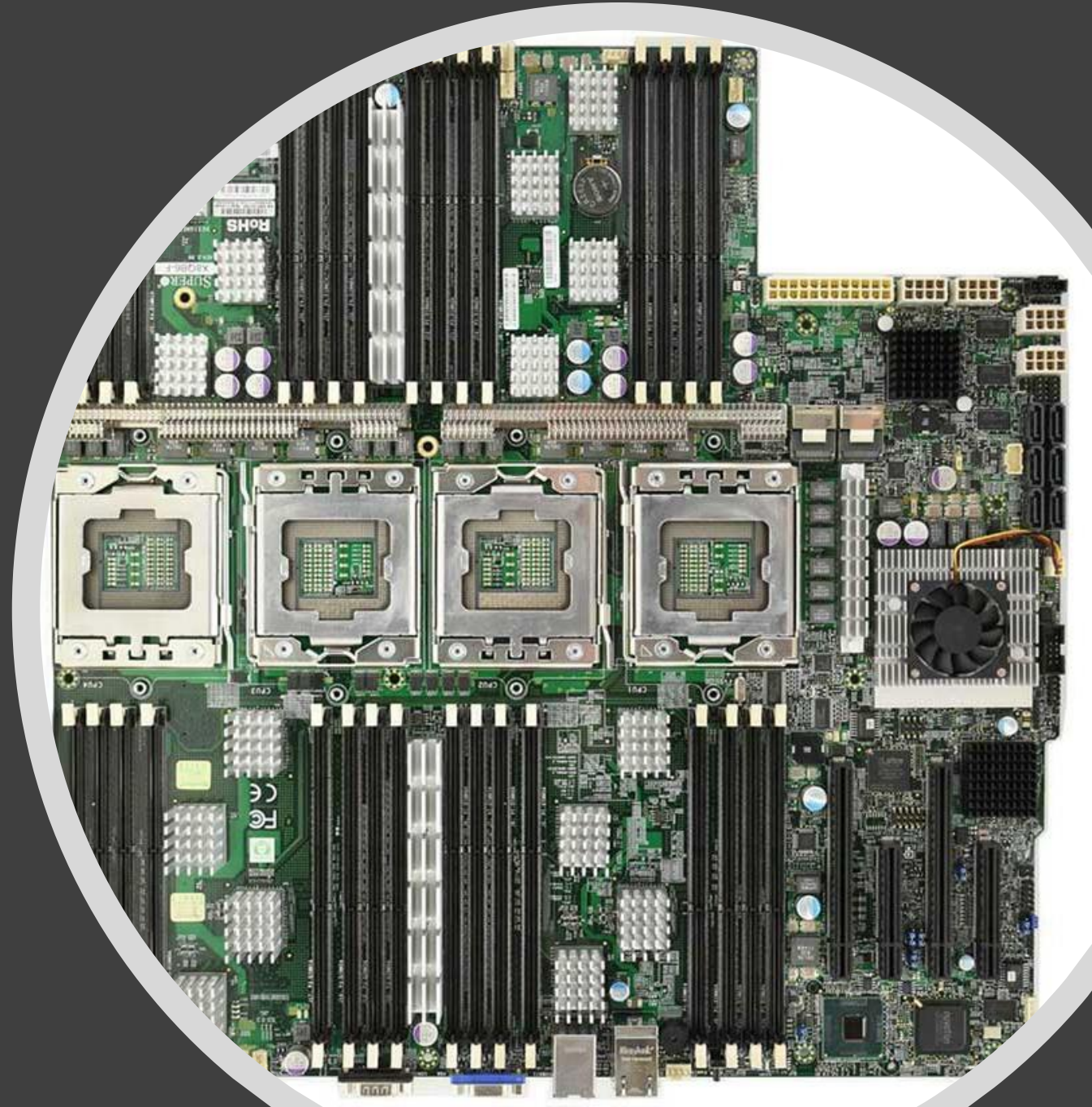


## Multiprocesadores simétricos(SMP)

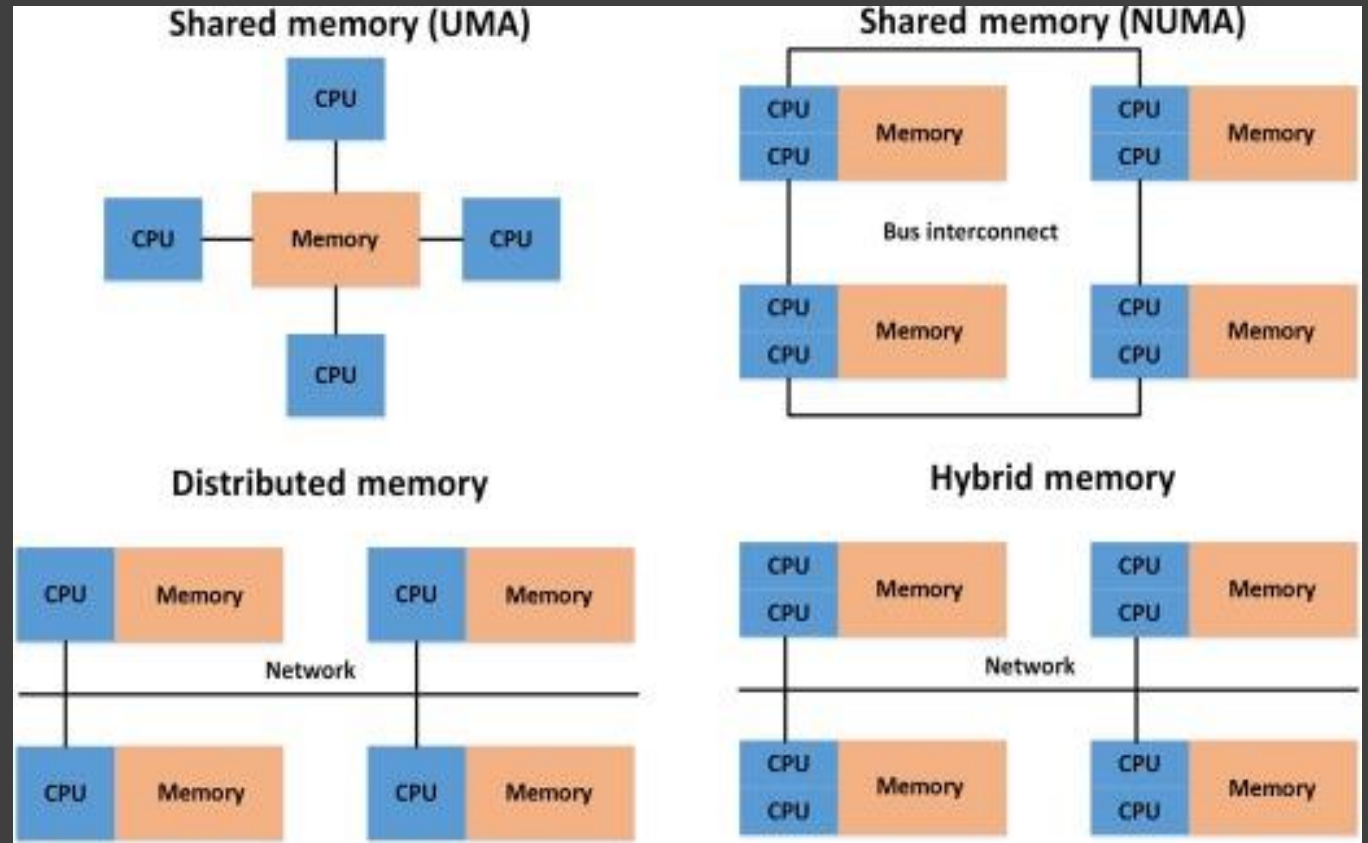
Se trata de tener una placa o PCB en donde dos o mas procesadores idénticos entre ellos trabajan con recursos compartidos como será una memoria principal y un bus de datos. A esto se le llama UMA(Uniform Memory Access).

Lo mejor de este método es su escalaridad la encontramos con placas que soportan hasta 32 procesadores, y de tenerlo ubicado en un solo sistema físico.

La desventaja seria tener que compartir bus de datos, el cual debe ser controlado y arbitrado para que este disponible para la CPU que solicite acceso generando tiempos de espera.



- Tiene una variante llamada NUMA(Non-Uniform Memory Access) en donde las CPU no necesariamente serán idénticas, y cada una contara con su propia porción de memoria.



# Procesadores en clúster

- Este es el que mas uso tiene en el sector de supercomputacion, los superordenadores mas potentes del mundo trabajan con este método de procesamiento en paralelo.
- Consiste en una agrupación de ordenadores a priori independientes que trabajan en colaboración. Esto se realiza gracias a un enlace entre ellos que normalmente será mediante una red LAN de alta velocidad, el llamado clúster Beowulf.
- Cuando hablamos de superordenadores que tienen miles de procesadores y decenas de terabytes de memoria RAM, casi siempre son clústeres de este tipo.

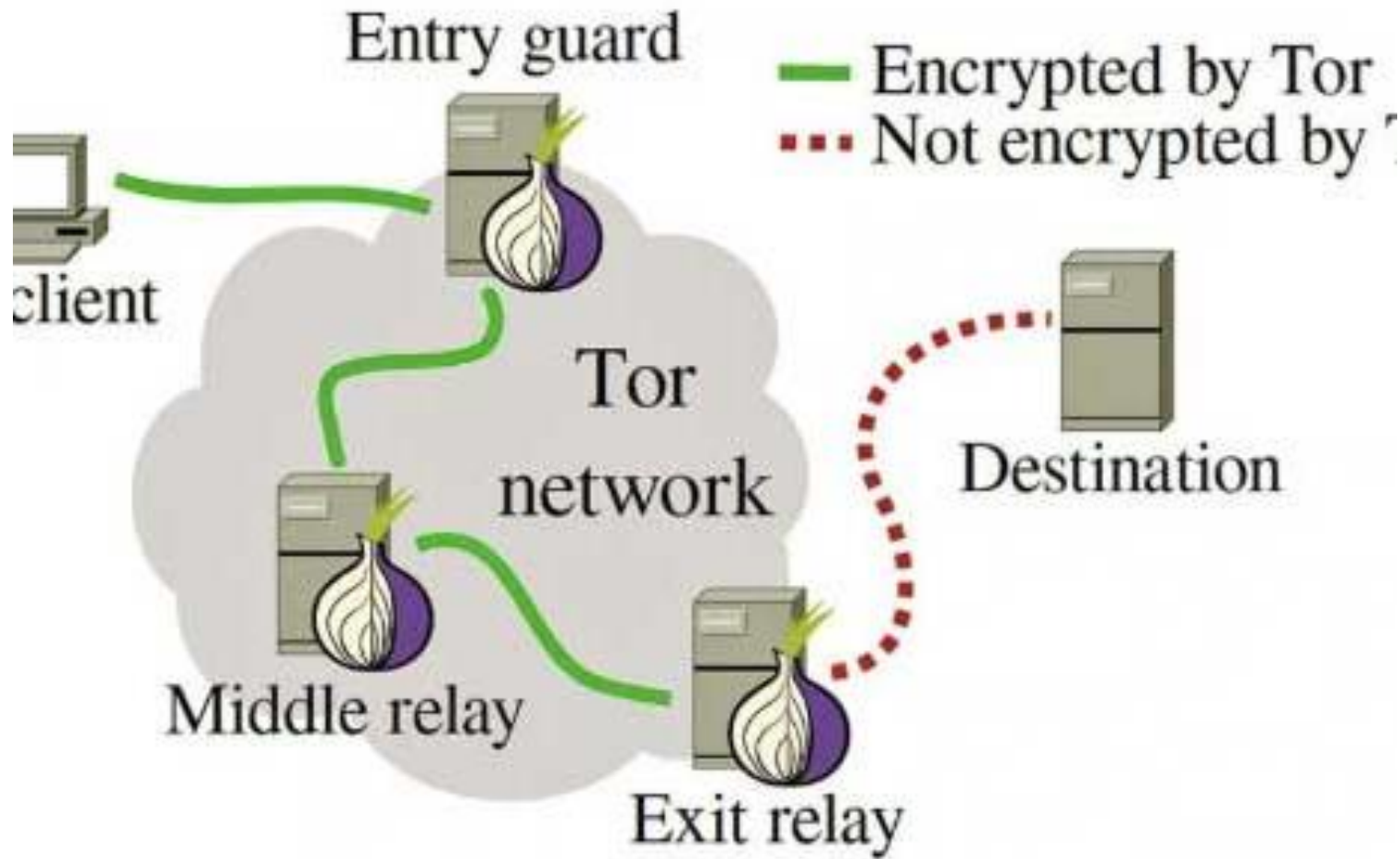




# MPP o Procesamiento paralelo masivo

- Un MPP será un solo equipo con multitud de procesadores interconectados en red, la interconexión es mas especializada que un clúster y no de consumo general, contando cada procesador con su propia memoria RAM y una copia del sistema operativo y aplicación para la ejecución en paralelo.





Procesamiento  
en paralelo  
distribuido

- se trata es de realizar una **computación en paralelo, pero con ordenadores que se comunican a través de Internet**, de forma remota y de distinta naturaleza y prestaciones. De esta forma se une la capacidad individual de cada equipo para ayudar a las tareas comunes de la red como es la búsqueda de resultados. Para ello es necesario que cada equipo cuente con la aplicación que se encarga de interconectar los equipos.
- El gran problema de este método es que **la red de Internet cuenta con mucha latencia** para el intercambio de datos, y además no todos los nodos cuentan con el mismo ancho de banda.

# GPGPU

- El GPGPU o Computo de propósito general con unidades de procesamiento gráfico es otra de las tendencias actuales que se están implementando sobre todo por la gran potencia de cálculo que tienen estos chips.
- una tarjeta gráfica está construida para trabajar principalmente con operaciones de coma flotante, vectores y matrices en donde la paralelización de procesos es muy alta. Una de estas aplicaciones es precisamente la del procesamiento y renderizado de gráficos 3D en movimiento, que a una CPU le costaría por tener “pocos núcleos”.
- La computación por tarjetas gráficas en paralelo es un método ideal de realizar simulaciones y cálculos científicos muy complejos, por ejemplo, resolver problemas matemáticos, trabajos en bases de datos, sintetizar elementos, modelar el comportamiento de fluidos.



- Ana Citlali García Chambasis
- 18052273