

Computación Paralela

Alain Paul Herrera Urtiaga

a.herrera@unap.edu.pe

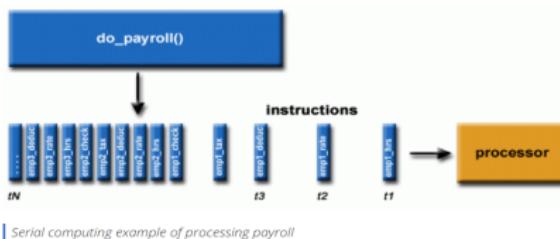
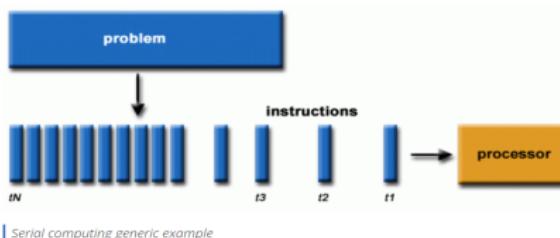
Escuela Profesional de Ingeniería Estadística e Informática

5 de mayo de 2025

¿Qué es la computación paralela?

- Tradicionalmente, el software se ha escrito para la **computación en serie**.

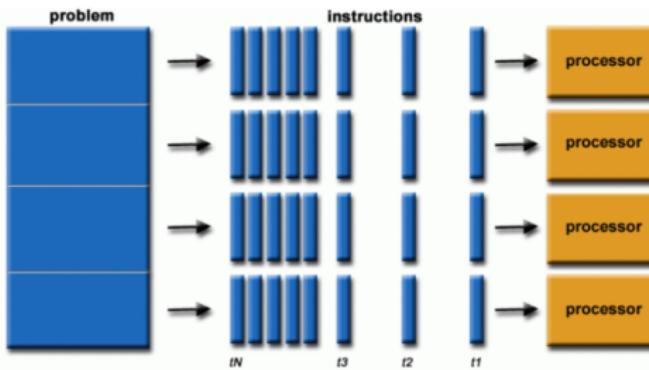
- Un problema se divide en una serie discreta de instrucciones.
 - Las instrucciones se ejecutan secuencialmente una tras otra.
 - Se ejecutan en un único procesador.
 - Solo una instrucción puede ejecutarse en cualquier momento dado.



Computación Paralela

En el sentido más simple, la computación paralela es el uso simultáneo de múltiples recursos informáticos para resolver un problema computacional:

- Un problema se divide en partes discretas que se pueden resolver simultáneamente
- Cada parte se divide a su vez en una serie de instrucciones
- Las instrucciones de cada parte se ejecutan simultáneamente en diferentes procesadores
- Se emplea un mecanismo general de control/coordinación



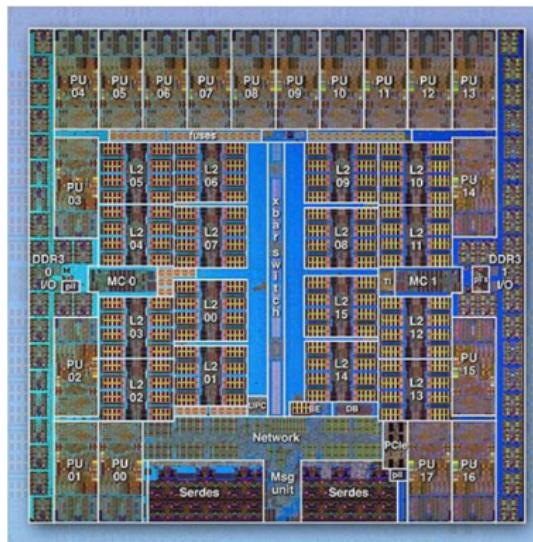
Parallel computing generic example

○○●○○○○○○○○

Ordenadores Paralelos

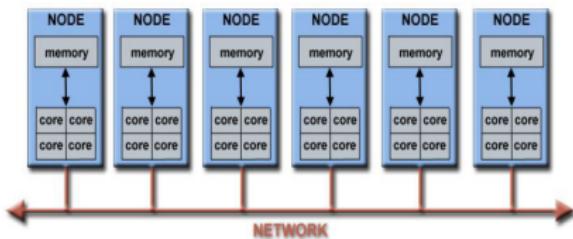
Prácticamente todos los ordenadores autónomos de hoy en día son paralelos desde una perspectiva de hardware:

- Múltiples unidades funcionales (caché L1, caché L2, branch, prefetch, decode, coma flotante, procesamiento de gráficos (GPU), entero, etc.)
 - Múltiples unidades/núcleos de ejecución
 - Múltiples hilos de hardware



Esta próxima imagen es de un procesador Intel Xeon con 6 núcleos y 6 unidades de caché L3, que es común en laptop y computadoras de escritorio:

- Cada nodo de computación es una computadora paralela multiprocesador en sí misma
- Varios nodos informáticos están conectados en red junto con una red Infiniband
- Los nodos de propósito especial, también multiprocesador, se utilizan para otros fines



¿Por qué utilizar la informática en paralelo? En el mundo natural, muchos eventos complejos e interrelacionados están sucediendo al mismo tiempo, pero dentro de una secuencia temporal. En comparación con la computación en serie, la computación en paralelo es mucho más adecuada para modelar, simular y comprender fenómenos complejos del mundo real, como por ejemplo:

- En el mundo natural, muchos eventos complejos e interrelacionados ocurren al mismo tiempo, pero dentro de una secuencia temporal.
- Por ejemplo, imagina modelar esto en serie:



Real world phenomena can be simulated with parallel computing



Real world phenomena can be simulated with parallel computing

■ AHORRAR TIEMPO Y/O DINERO

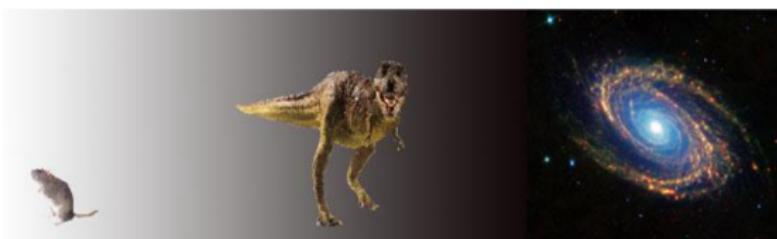
- En teoría, dedicar más recursos a una tarea acortará su tiempo de finalización, con el consiguiente ahorro de costes.
- Los ordenadores paralelos pueden construirse con componentes básicos y baratos.



Working in parallel shortens completion time

Resolver Problemas Más Grandes / Más Complejos

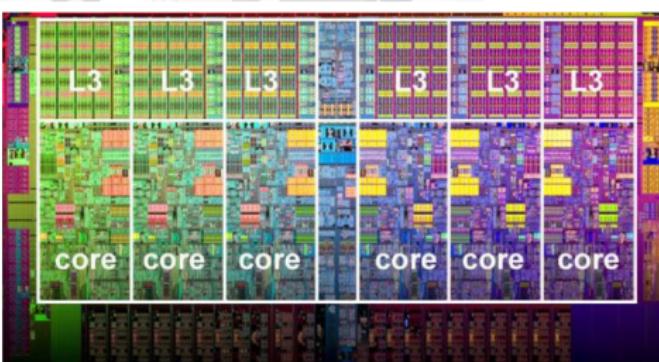
- Muchos problemas son tan grandes y/o complejos que resulta poco práctico o imposible resolverlos utilizando un programa secuencial, especialmente dada la limitada memoria del ordenador.
- **Ejemplo:** "Grand Challenge Problems" que requieren petaflops y petabytes de recursos informáticos.
- **Ejemplo:** Motores de búsqueda/bases de datos web que procesan millones de transacciones por segundo.



| Parallel computing can solve increasingly complex problems

Aprovechamiento del Hardware Paralelo Subyacente

- Los ordenadores modernos, incluso los portátiles, tienen una arquitectura paralela con múltiples procesadores/núcleos.
- El software paralelo está específicamente diseñado para hardware paralelo con múltiples núcleos, hilos, etc.
- En la mayoría de los casos, los programas en serie que se ejecutan en ordenadores modernos "desperdician" la potencia informática potencial.



| Parallel computing cores

El Futuro

- Durante los últimos más de 20 años, las tendencias indicadas por redes cada vez más rápidas, sistemas distribuidos y arquitecturas de ordenadores multiprocesador (incluso a nivel de escritorio) muestran claramente que el paralelismo es el futuro de la computación.
- En este mismo período de tiempo, ha habido un aumento de más de 500.000 veces en el rendimiento de las supercomputadoras, sin que se vislumbre un final actualmente.



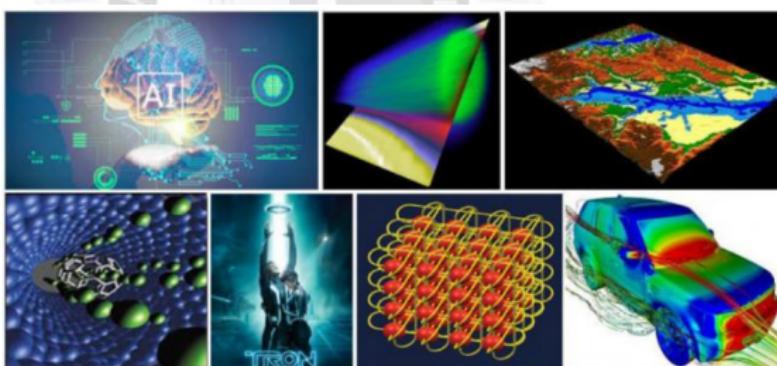
Source: Top500.org

¿Quién utiliza la computación paralela?

Ciencia e Ingeniería

Históricamente, la computación paralela ha sido considerada como “la gama alta de la computación” y se ha utilizado para modelar problemas difíciles en muchas áreas de la ciencia y la ingeniería:

- Atmósfera, Tierra, Medio Ambiente
- Física - aplicada, nuclear, de partículas, materia condensada, alta presión, fusión, fotónica

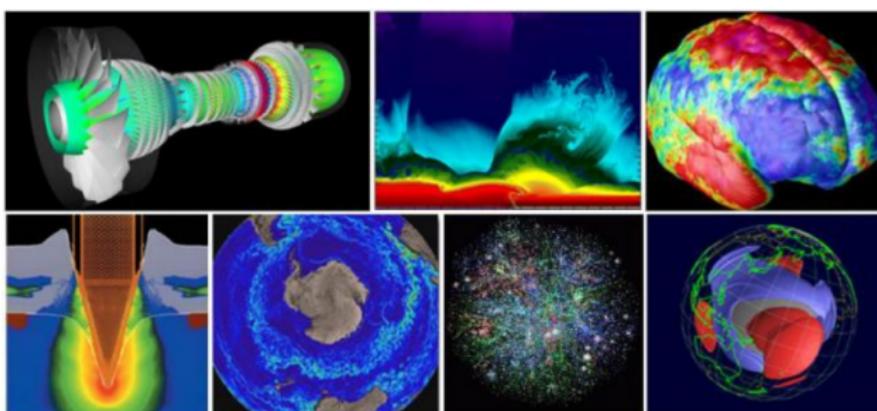


| Parallel computing is used in many commercial applications

Aplicaciones Industriales y Comerciales

Hoy en día, las aplicaciones comerciales proporcionan una fuerza impulsora igual o mayor en el desarrollo de ordenadores más rápidos. Estas aplicaciones requieren el procesamiento de grandes cantidades de datos de formas sofisticadas. Por ejemplo:

- "Big Data", bases de datos, minería de datos
- Inteligencia Artificial (IA)
- Motores de búsqueda web, servicios empresariales basados en la web



| Parallel computing is key to simulating a range of complex physical phenomena

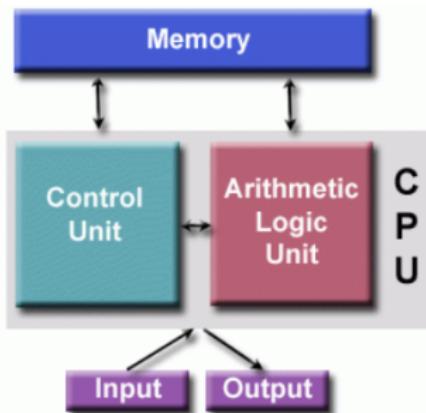
Arquitectura de Computadora von Neumann



Figura: John von Neumann circa 1940s (Fuente: Archivos de LANL)

- Nombrada en honor a John von Neumann (1945).
- "Computadora de programa almacenado": instrucciones y datos en memoria electrónica.
- Diseño básico seguido por la mayoría de las computadoras.

Componentes de la Arquitectura de von Neumann



Basic computing architecture

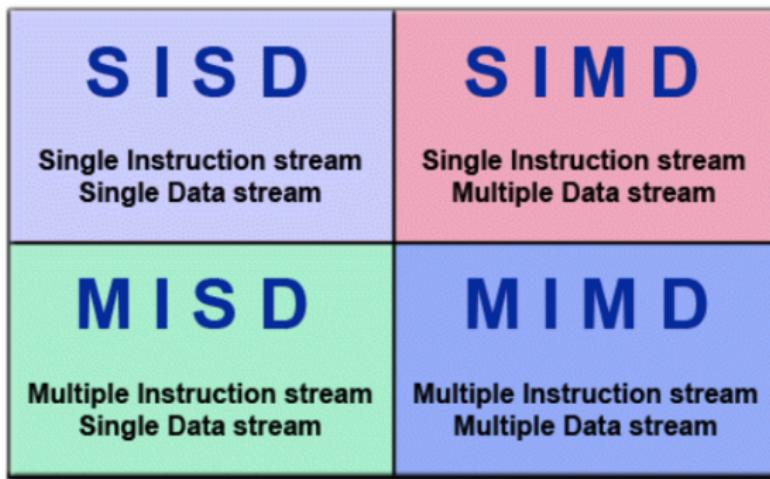
Figura: Arquitectura de computación básica

- Memoria (instrucciones y datos).
- Unidad de Control (obtiene, decodifica y coordina).
- Unidad Aritmético Lógica (operaciones básicas).
- Entrada/Salida (interfaz humano).

Las computadoras paralelas multiplican esta arquitectura.

Taxonomía Clásica de Flynn

- Clasificación de computadoras paralelas (desde 1966).
- Basada en el Flujo de Instrucciones y el Flujo de Datos (Simple o Múltiple).



| Flynn's taxonomy

SISD: Single Instruction, Single Data

- Computadora serial (no paralela).
- Un flujo de instrucciones, un flujo de datos.
- Ejecución determinista.
- Ejemplos: mainframes antiguos, minicomputadoras, PCs de un solo núcleo.

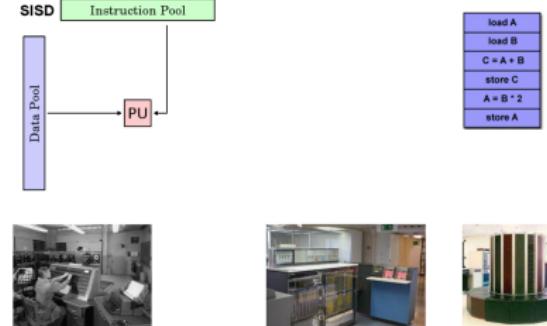


Figura: Ejemplo SISD



SIMD: Single Instruction, Multiple Data

- Computadora paralela.
- Una instrucción, múltiples datos.
- Ideal para problemas regulares (procesamiento de imágenes/gráficos).
- Ejecución síncrona y determinista.
- Variedades: Arreglos de Procesadores y Tuberías Vectoriales.
- GPUs modernas usan instrucciones SIMD.

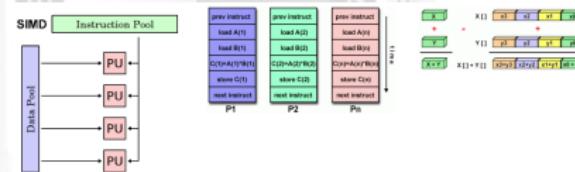


Figura: Ejemplo SIMD

MISD: Multiple Instruction, Single Data

- Computadora paralela.
- Múltiples instrucciones, un flujo de datos.
- Pocos ejemplos reales.
- Usos concebibles: múltiples filtros en una señal, criptografía múltiple en un mensaje.

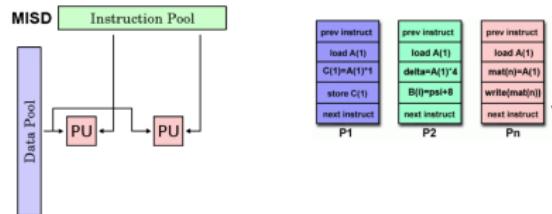


Figura: Ejemplo MISD

MIMD: Multiple Instruction, Multiple Data

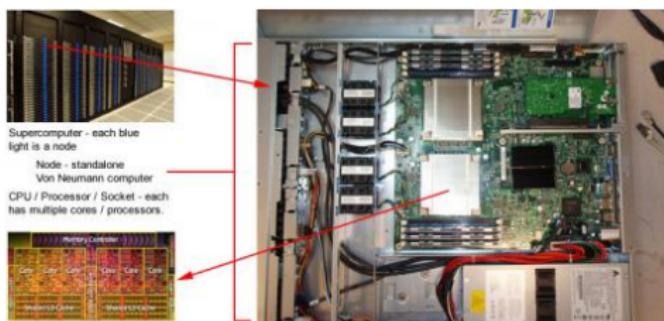
- Computadora paralela.
- Múltiples instrucciones, múltiples datos.
- Ejecución síncrona o asíncrona, determinista o no.
- Tipo más común (supercomputadoras, clústeres, PCs multinúcleo).
- Muchas arquitecturas MIMD incluyen SIMD.



Figura: Ejemplo MIMD

Terminología General: CPU y Nodo

- Uno o más **núcleos** (unidad de ejecución con su propio flujo de instrucciones).
- Núcleos en **sockets** (cada socket con su propia memoria).
- Sockets pueden compartir memoria.
- Computadora en una caja independiente.
- Múltiples CPUs/procesadores/núcleos, memoria, interfaces de red.
- Nodos conectados forman una supercomputadora.



Nodes in a supercomputer

Figura: Nodos en una supercomputadora

Terminología General: Tarea, Pipelining, Memoria

- Sección discreta de trabajo computacional.
- Programa o conjunto de instrucciones ejecutado por un procesador.
- Programa paralelo: múltiples tareas en múltiples procesadores.
- Dividir una tarea en pasos por diferentes unidades de procesamiento.
- Entradas fluyen como en una línea de ensamblaje.
- Todos los procesadores acceden directamente a la memoria física común.
- Tareas paralelas comparten la misma "imagen" de la memoria lógica.

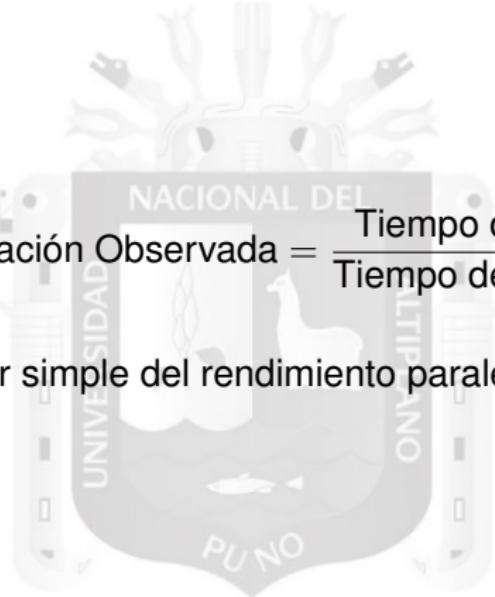
Terminología General: SMP, Memoria Distribuida, Comunicaciones

- Hardware de memoria compartida.
- Múltiples procesadores comparten un espacio de direcciones y recursos.
- Hardware: acceso a memoria no común a través de la red.
- Programación: tareas solo "ven"memoria local, usan comunicación para otras memorias.
- Intercambio de datos entre tareas paralelas (bus compartido, red).

Terminología General: Sincronización, Granularidad

- Coordinación de tareas paralelas en tiempo real.
- A menudo asociada con comunicaciones.
- Puede aumentar el tiempo de ejecución.
- Relación entre computación y comunicación.
- **Gruesa:** mucha computación entre comunicaciones.
- **Fina:** poca computación entre comunicaciones.

Terminología General: Aceleración Observada


$$\text{Aceleración Observada} = \frac{\text{Tiempo de ejecución serial}}{\text{Tiempo de ejecución paralela}}$$

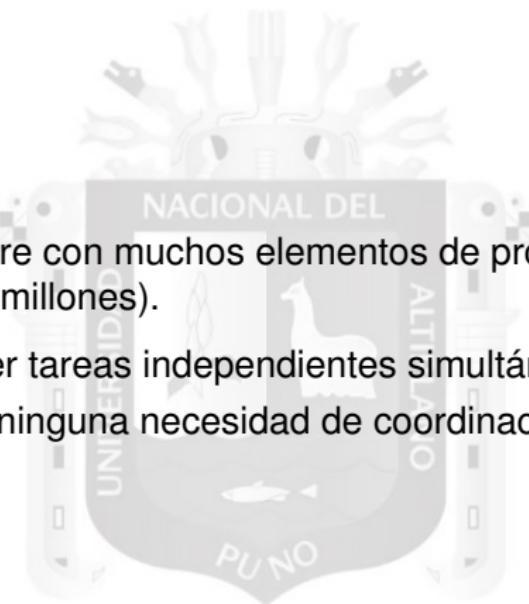
- Indicador simple del rendimiento paralelo.

Terminología General: Sobrecarga Paralela

- Tiempo de ejecución exclusivo de tareas paralelas (no para trabajo útil).
- Incluye:
 - Inicio de tareas
 - Sincronizaciones
 - Comunicación de datos
 - Sobrecarga de software
 - Finalización de tareas

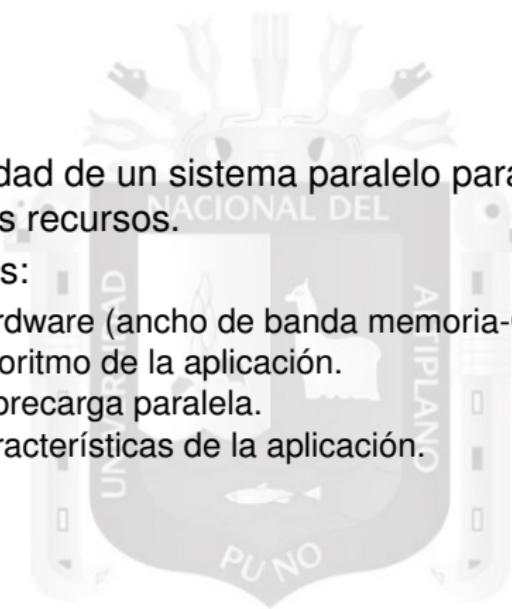


Terminología General: Masivamente Paralelo, Idealmente Paralelo

- 
- Hardware con muchos elementos de procesamiento (cientos de miles a millones).
 - Resolver tareas independientes simultáneamente.
 - Poca o ninguna necesidad de coordinación.

Terminología General: Escalabilidad

- Capacidad de un sistema paralelo para aumentar la aceleración con más recursos.
- Factores:
 - Hardware (ancho de banda memoria-CPU, red).
 - Algoritmo de la aplicación.
 - Sobrecarga paralela.
 - Características de la aplicación.



Gracias!

Contacto:

aherrera@unap.edu.pe

