

Fundamentos del diseño de computadores

J. Daniel García Sánchez (coordinador)
David Expósito Singh
Javier García Blas
Óscar Pérez Alonso
J. Manuel Pérez Lobato

Arquitectura de Computadores
Grupo ARCOS
Departamento de Informática
Universidad Carlos III de Madrid

1. Objetivos del curso

Este es el primer módulo del curso de *Arquitectura de Computadores*. El módulo pretende dar una visión general del curso completo. Comenzaremos por revisar el objetivo fundamental del curso.

Objetivo: que el estudiante conozca los conceptos básicos sobre la **arquitectura de un computador** y el impacto que estos tienen sobre el **rendimiento de las aplicaciones** y los sistemas informáticos.

Para alcanzar este objetivo, a lo largo del curso, se profundiza en la adquisición de tres competencias:

1. Capacidad de conocer, comprender y evaluar la **arquitectura de los computadores**, así como los **componentes básicos** que los conforman.
2. Conocimiento y aplicación de los principios fundamentales y técnicas básicas de la **programación paralela y concurrente**.
3. Capacidad de analizar y evaluar **arquitecturas de computadores**, incluyendo **plataformas paralelas**, así como desarrollar y **optimizar software** para las mismas.

Tenga en cuenta que para poder seguir con éxito este curso, se espera que el estudiante haya seguido previamente al menos un curso en las siguientes materias:

- Estructura de computadores.
- Sistemas operativos.
- Programación.

Con esta aproximación el curso se estructura en seis bloques temáticos:

1. Fundamentos del diseño de computadores.

2. Evaluación del rendimiento de sistemas informáticos.
3. Paralelismo a nivel de instrucción.
4. Jerarquía de memoria.
5. Introducción a los multiprocesadores.
6. Modelos de programación paralela y concurrente.

2. Estructura del módulo

Este módulo de introducción se estructura en una única lección que introduce el curso y ofrece una visión general del diseño de computadores.

La lección tiene la siguiente estructura general:

1. Introducción.
2. Perspectiva histórica.
3. Clasificación de computadores.
4. Paralelismo.
5. Arquitectura del computador.

3. Contenidos de la lección

3.1. Introducción

El término arquitectura de computadores tiene múltiples definiciones. En el material del curso se proporcionan dos posibles definiciones del término. Observe que las definiciones son complementarias. Además, el conocimiento de la arquitectura de computadores es esencial para comprender las tendencias en la evolución de los computadores así como las limitaciones de los mismos.

Un aspecto fundamental en la evolución pasada de los procesadores viene dado por la conocida *Ley de Moore*. Tenga en cuenta que la mencionada ley define la evolución del número de transistores por chip. Aunque ha habido momentos en que esto se ha traducido en incrementos en el rendimiento en los procesadores, esto no siempre ha sido así. De hecho, desde 2005, aunque la *Ley de Moore* ha seguido cumpliéndose se ha producido un estancamiento en la frecuencia de reloj de los procesadores. Para profundizar en este aspecto se recomienda que lea con atención el artículo “*The free lunch is over*” que se suministra como lectura complementaria.

En la evolución histórica de los procesadores ha habido dos puntos de inflexión relevantes. El primero ocurrió en la década de los 80 con la aparición de la tecnología RISC. Tras esto se produjo un incremento sostenido en el rendimiento de los procesadores del 52 % anual.

3.2. Perspectiva histórica

Desde un punto de vista histórico se pueden identificar tres cambios importantes que se pueden calificar de revoluciones.

La primera de ellas es la aparición del microprocesador a principios de los 70, que permitió integrar suficientes transistores en un único chip para incluir todos los componentes necesarios para un procesador de 16 bits. Esto dio lugar a la aparición de nuevos segmentos de mercado.

La segunda revolución consistió en la incorporación del paralelismo a nivel de instrucción (ILP - *Instruction Level Parallelism*) que permitió incrementar el rendimiento ocultando los detalles al programador que sigue teniendo esencialmente una vista secuencial del modelo de ejecución.

Finalmente, la tercera revolución consiste en la introducción del soporte explícito del paralelismo de datos así como del soporte de múltiples hilos de ejecución. Este modelo hace el paralelismo visible a los programadores. Este cambio fue debido a que el paralelismo a nivel de instrucción cada vez reportaba menores incrementos en el rendimiento.

Como tendencias arquitectónicas se identifican tres modelos de paralelismo:

- **DLP** (*Data Level Parallelism*): Paralelismo a nivel de datos, que permite aplicar una misma operación sobre múltiples datos en paralelo.
- **TLP** (*Thread Level Parallelism*): Paralelismo a nivel de hilo, que permite ejecutar múltiples hilos de ejecución simultáneamente.
- **RLP** (*Request Level Parallelism*): Paralelismo a nivel de peticiones, que permite procesar múltiples peticiones de forma simultánea.

3.3. Clasificación de computadores

En el material encontrará una clasificación de los computadores en diversas categorías: dispositivos móviles personales, computadores de escritorio, servidores, clústers y computadores empujados. Observe que cada una de estas categorías implica distintas características y distintos costes.

3.4. Paralelismo

En el material se realiza una estimación teórica de las características que serían necesarias para construir una máquina con una capacidad de procesamiento de 1 Teraflop y una capacidad de almacenamiento de 1 TB de datos. La conclusión es que se requeriría 3 Angstroms por byte, lo que hace esta aproximación secuencial inviable. Sin embargo, si se pueden alcanzar estas prestaciones mediante aproximaciones paralelas.

Desde el punto de vista de las aplicaciones se distinguen dos tipos de paralelismo: paralelismo de datos y paralelismo de tareas. Esto es una propiedad de las aplicaciones.

Desde el punto de vista del hardware se distinguen cuatro aproximaciones al paralelismo: ILP, arquitecturas vectoriales, TLP, RLP. Esto es una propiedad del hardware. No obstante, es cierto que existen equipos que combinan varios de estos modelos.

Una clasificación tradicional de los posibles tipos de arquitecturas paralelas es la de Flynn (1966). Esta clasificación se realiza atendiendo a dos criterios: el número de flujos de instrucciones simultáneos y el número de flujos de datos simultáneos. Estos dos criterios dan lugar a cuatro categorías: SISD, SIMD, MISD, y MIMD. No obstante, la clase MISD es teórica y no se conocen aplicaciones industriales. Por otra parte, la clase MIMD da lugar a dos subclases, dependiendo de si la arquitectura es altamente acoplada (TLP) o débilmente acoplada (RLP).

3.5. Arquitectura del computador

Existen diversas visiones sobre la arquitectura del computador. Cada una de estas visiones se centra en algún atributo importante de la arquitectura. Siga el material para identificar el espacio de diseño en cada una de estas dimensiones y las decisiones tomadas para algunas de las arquitecturas más conocidas.