

Configuración y Optimización de Sistemas de Cómputo

Master Universitario en Ingeniería Informática
Depto. de Informática de Sistemas y Computadores (DISCA)
Universidad Politécnica de Valencia

Transparencias : Carles Hernández

Infraestructura del Sistema

- Procesadores
 - Características
 - Alta productividad
 - Alto rendimiento
 - Eficiencia Energética
 - Medidas de prestaciones
 - Consumo Energético
 - Aspectos de Diseño
 - Seguridad
 - Fiabilidad
 - Prestaciones
 - Performance/Watt

Procesadores

ESTADÍSTICAS							×
	Frecuencia núcleo	Caché	N.º de núcleos	N.º de procesos	Toma	Graphics Int.	
Intel® Core™ i5-11400F	2,6 GHz	12 MB L3	6	12	LGA 1200	✗	
Intel® Core™ i5-11400	2,6 GHz	12 MB L3	6	12	LGA 1200	✓	
Intel® Xeon® E-2314	2,8 GHz	8 MB	4 núcleos	4 hilos	1200	✗	
Intel® Core™ i7-11700	2,50 GHz	16 MB L3	8	16	LGA 1200	✓	
Intel® Xeon® E-2336	2,9 GHz	12 MB	6 núcleos	12 hilos	1200	✗	
Intel® Xeon® E-2374G	3,7 GHz	8 MB	4 núcleos	8 hilos	1200	✗	
Intel® Xeon® E-2378G	2,8 GHz	16MB		16 procesos	1200	✗	

<https://www.pcspecialist.es/servidores/intel-1200-1u/>

Rendimiento y Eficiencia

- Instrucciones/Ciclo
- Instrucciones/Ciclo/Watt
- Aumentar rendimiento
 - Aumentar el número de instrucciones por ciclo
 - Aumentar el numero de instrucciones (Procesador superscalar)
 - Aumentar el número de cores (Multicores)
 - Aumentar el número de hilos (Multithread)
 - Reducir ciclos
 - Mejorar la jerarquía de memoria y/o ancho de banda de memoria
 - Aumentar la frecuencia (↓ empeora Instrucciones/Ciclo/Watt)

Rendimiento

- Aumento instrucciones por ciclo
 - Procesadores segmentados básicos
 - Objetivo 1 instrucción/ciclo

Basic five-stage pipeline							
Instr. No. \ Clock cycle	1	2	3	4	5	6	7
1	IF	ID	EX	MEM	WB		
2		IF	ID	EX	MEM	WB	
3			IF	ID	EX	MEM	WB
4				IF	ID	EX	MEM
5					IF	ID	EX

(IF = Instruction Fetch, ID = Instruction Decode, EX = Execute, MEM = Memory access, WB = Register write back).

In the fourth clock cycle (the green column), the earliest instruction is in MEM stage, and the latest instruction has not yet entered the pipeline.

[illegible]

The diagram illustrates the architecture of the Intel® Xeon® Phi 7200 processor, showing the flow of instructions and data through various stages:

- Instruction Fetch Unit:** Receives instructions from the **128 Entry ITLB** and the **32 KB Instruction Cache (8 way)**. It feeds into the **32 Byte Pre-Decode, Fetch Buffer**.
- Instruction Queue:** The **32 Byte Pre-Decode, Fetch Buffer** feeds into the **18 Entry Instruction Queue**.
- Decoders:** The **18 Entry Instruction Queue** feeds into a set of decoders: **Micro-code**, **Complex Decoder**, **Simple Decoder**, **Simple Decoder**, and **Simple Decoder**. These decoders output **4 μops**, **1 μop**, **1 μop**, and **1 μop** respectively.
- μop Buffer:** The decoders feed into the **7+ Entry μop Buffer**.
- Register Alias Table and Allocator:** The **7+ Entry μop Buffer** feeds into the **Register Alias Table and Allocator**.
- Reorder Buffer (ROB):** The **Register Alias Table and Allocator** feeds into the **96 Entry Reorder Buffer (ROB)**.
- Retirement Register File:** The **96 Entry Reorder Buffer (ROB)** feeds into the **Retirement Register File (Program Visible State)**.
- Reservation Station:** The **96 Entry Reorder Buffer (ROB)** feeds into the **32 Entry Reservation Station**.
- Execution Units:** The **32 Entry Reservation Station** feeds into various execution units:
 - Port 0:** **ALU** and **SSE Shuffle ALU**. The **SSE Shuffle ALU** feeds into the **128 Bit FMUL FDIV** unit.
 - Port 1:** **ALU** and **SSE Shuffle MUL**. The **SSE Shuffle MUL** feeds into the **128 Bit FADD** unit.
 - Port 5:** **ALU Branch** and **SSE ALU**.
 - Port 3:** **Store Address**.
 - Port 4:** **Store Data**.
 - Port 2:** **Load Address**.
- Memory Ordering Buffer (MOB):** The **Store Address**, **Store Data**, and **Load Address** units feed into the **Memory Ordering Buffer (MOB)**.
- Internal Results Bus:** The **128 Bit FMUL FDIV** and **128 Bit FADD** units feed into the **Internal Results Bus**.
- Cache and DTLB:** The **Internal Results Bus** feeds into the **32 KB Dual Ported Data Cache (8 way)** and the **16 Entry DTLB**.
- Shared L2 Cache and DTLB:** The **32 KB Dual Ported Data Cache (8 way)** and the **16 Entry DTLB** feed into the **Shared L2 Cache (16 way)** and the **256 Entry L2 DTLB**.
- Shared Bus Interface Unit:** The **Shared L2 Cache (16 way)** and the **256 Entry L2 DTLB** feed into the **Shared Bus Interface Unit**.

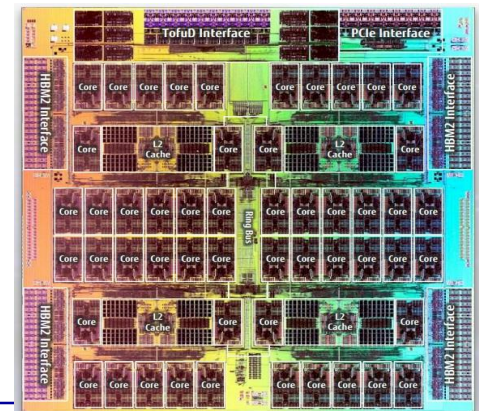


Procesadores Superescalares

- Aumentan el rendimiento ($> 1\text{inst/ciclo}$)
 - Replican unidades funcionales
 - Ejemplo = Pipeline de Enteros + Pipeline Coma Flotante
 - Ejemplo = Varias ALUs
 - Aumentan la complejidad interna
 - Generalmente son dinámicos
 - Ejecución fuera de orden y decisiones hardware (vs VLIW con decisiones SW)
 - Lógica para detectar/evitar dependencias (necesario para tener eficiencia)
 - Aumento de puertos de memorias y registros
 - Mayor ancho de banda en la búsqueda (fetch) y decodificación de instrucciones
 - Existen un limite en la mejora que pueden proporcionar
 - Los programas no puedes siempre explotar el paralelismo
 - Recursos infrautilizados??

Multi-Manycores

- Incluir más núcleos de procesamiento mejora:
 - La productividad global del chip (aplicaciones único hilo)
 - El rendimiento en aplicaciones paralelas
- Aumentar núcleos
 - Requiere de aumentar el ancho de banda dentro del chip
 - Sistemas basados en bus → Redes en el chip
 - Gestionar la coherencia de memoria de forma eficiente
 - Protocolos más complejos
 - Limitaciones Ancho de Banda de memoria
 - Número de pines limitado




Simultaneos Multithreading (SMT)

- Utiliza multiples hilos (threads) para procesar más instrucciones por ciclo
- SMT soporta varios contextos de ejecución de manera simultánea
- Complementan a los procesadores superescalares cuando no hay suficiente ILP por thread (hilo)
- **Aumento de la productividad global**
- **Mejora de la utilización de los recursos del sistema**

Simultaneous Multithreading (SMT)

Intel® Hyper-Threading can benefit performance


w/o SMT



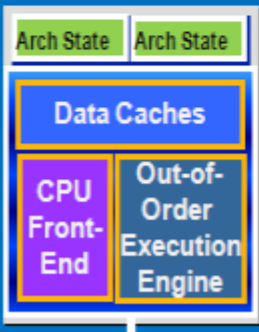
Time (proc. cycles)

Note: Each box represents a processor execution unit

SMT



- Also known as Simultaneous Multi-Threading (SMT)
 - Run 2 threads at the same time per core
- Shares Resources(Cache, Frontend, Execution Units)
- Improves Core CPI (Clockticks per Instruction)
- Potentially degrades Thread CPI

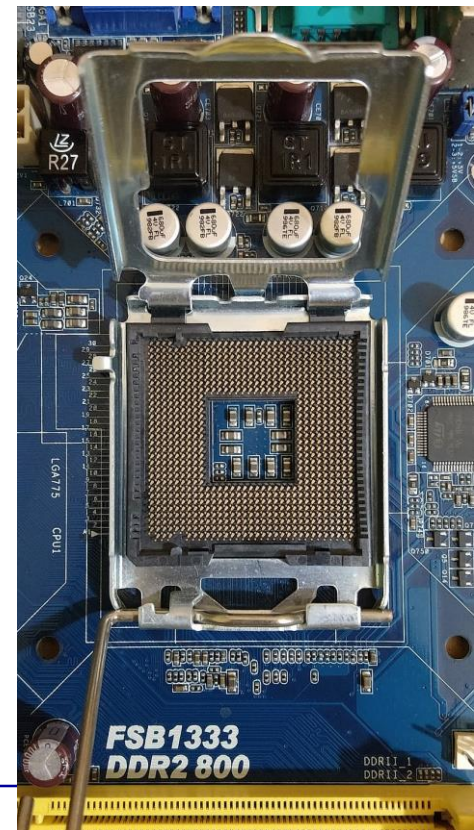


VISUAL ADRENALINE intel Software

18

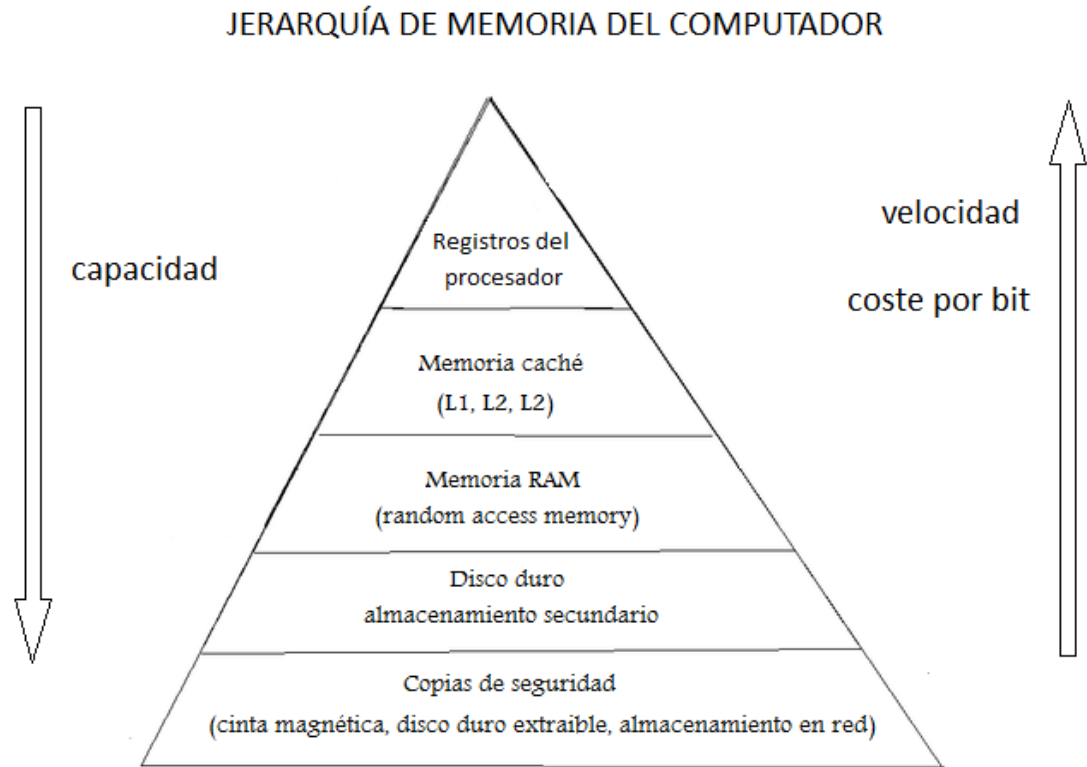
Multi-socket

- Un nodo de computación en un servidor puede incluir más de un socket (encapsulado)
- **Aumenta el número de núcleos de mi sistema**
- Comparten memoria / placa / espacio



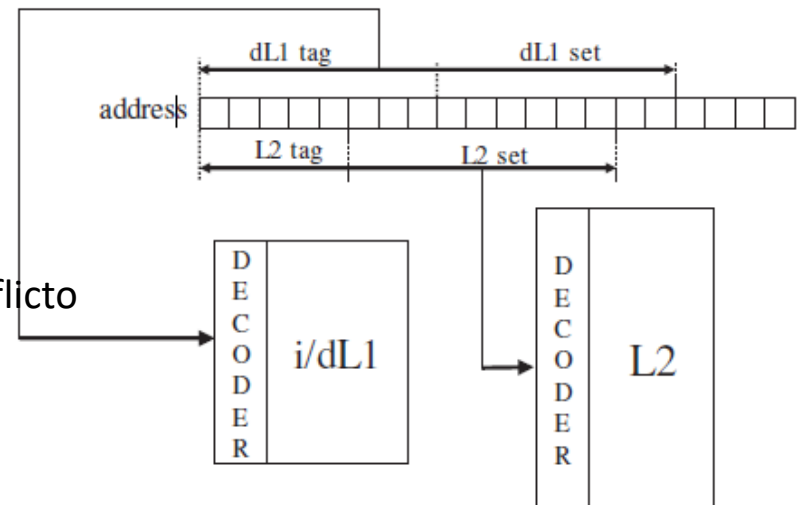
Jerarquía de Memoria

Una jerarquía adecuada permite reducir el número de ciclos por operación



Jerarquía de Memoria

- El objetivo es minimizar accesos a memoria (fuera del chip)
 - Elevados tiempos de acceso
 - Elevado consumo
- Los procesadores integran varios niveles de cache
 - Memorias en el chip de pequeño tamaño → rápidas
 - Albergan una parte del total de la memoria
 - Explotan localidad
 - Espacial
 - Temporal
 - Fallos / Aciertos
 - Fallos : Capacidad, calentamiento, conflicto

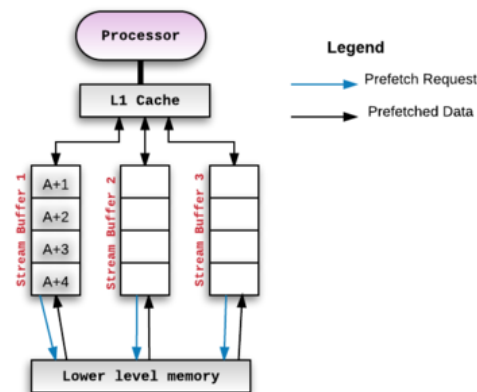


Jerarquía de Memoria

- Configuración más común en servidores
 - Cache de Datos y de Instrucciones separadas (L1)
 - Caches asociativas por conjuntos
 - El número de vías define la flexibilidad para tolerar fallos por conflictos
 - Diferentes configuraciones para cache de datos y de Instrucciones
 - Cache de segundo nivel (L2) privada
 - Mayor tamaño que L1
 - Mayor asociatividad (número de vías)
 - Cache de tercer nivel (L3) compartida
 - Mucho más grande que la L2 (ej. 22MB en Intel Xeon Gold)
 - Mayor número de vías
 - Permiten particionado → Performance isolation (QoS)

Jerarquía de Memoria

- La cache es útil para reducir los tiempos efectivos de acceso a memoria
 - Para datos que nunca he leído o que por capacidad he echado fuera no resulta efectivo
- El prefetcher es un elemento hardware adicional que se añade a la jerarquía para anticiparnos a esos potenciales accesos
 - Anticipar los datos de forma especulativa esperando que se utilicen en el futuro próximo



Configuraciones

CPU Platform	Graviton2	EPYC 7571	Xeon Platinum 8259CL
vCPUs	64		
Cores Per Socket	64	32	24 (16 instantiated)
SMT	-	2-way	2-way
CPU Sockets	1	1	2
Frequencies	2.5GHz	2.5-2.9GHz	2.9-3.2GHz
Architecture	Arm v8.2	x86-64 + AVX2	x86-64 + AVX512
µarchitecture	Neoverse N1	Zen	Cascade Lake
L1I Cache	64KB	64KB	32KB
L1D Cache	64KB	32KB	32KB
L2 Cache	1MB	512KB	1MB
L3 Cache	32MB shared	8MB shared per 4-core CCX	35.75MB shared per socket
Memory Channels	8x DDR4-3200	8x DDR-2666 (2x per NUMA-node)	6x DDR4-2933 per socket
NUMA Nodes	1	4	2
DRAM	256GB		
TDP	Estimated 80-110W?	180W	210W per socket

CPUs Virtuales

- La mayoría de proveedores de Cloud ofrecen a los clientes CPUs virtualizadas (vCPU)
- Asignación de vCPU
 - Utilizar núcleos físicos
 - Utilizar hilos de ejecución (SMT)
- Consideraciones:
 - Bajada de rendimiento por colisiones en recursos compartidos
 - Especialmente en SMT
 - Explotación de vulnerabilidades y “fallos” de seguridad
 - Virtualización es efectiva a nivel lógico pero no a nivel físico

CPUs Virtuales



- Seguridad
 - Ataques de canal lateral
 - 2 procesos en 2 máquinas virtuales diferentes comparten ciertos recursos del procesador (cache L3, interconexión, ...)
 - Se han demostrado que explotando las interferencias entre procesos es posible acceder a datos sensibles
 - Soluciones:
 - Tener el SW lo más actualizado posible (CVEs)
 - Limitar la compartición de recursos para aplicaciones sensible
 - Utilizar las técnicas de particionado de recursos que ofrecen los procesadores modernos

CPUs Virtuales: Seguridad

- Ejecución Especulativa
 - Mitiga el impacto en el rendimiento de las instrucciones de salto
 - Sobre todo superescalares fuera de orden
 - Permite acceder de forma “invisible” a zonas protegidas

```
void victim_function(size_t x) {  
    if (x < array1_size) {  
        temp &= array2[array1[x] * 512];  
    }  
}
```

Medidas Rendimiento

- Benchmarks para la caracterización del rendimiento del procesadores
 - Las cargas de trabajo no son iguales para todos los sistemas
 - Cloud, HPC, webserver,
 - Se pueden centrar en CPU, memoria, aceleradores, red, almacenamiento, etc.
- **Standard Performance Evaluation Corporation (SPEC)**
 - Organización que desarrolla benchmarks para caracterizar CPUs
 - Existen gran cantidad de benchmarks para cada tipo de aplicación

Medidas Rendimiento

- Tiempo de ejecución (s)
 - De un determinado servicio
 - De una aplicación
- Consumo medio (W)
 - De mi sistema cuando ejecuto una carga representativa
- Productividad ($\text{sum}(\text{Inst}/\text{ciclo})$)
 - Para ver la efectividad del sistema a nivel global
 - != ejecutar un hilo y multiplicar por el número total
- Contadores de rendimiento
 - Existentes en todas las CPUs
 - Útiles para identificar cuellos de botella y optimizar el sistema para nuestro tipo de carga
 - Ejemplo: Contadores de accesos, Fallos de cache (L1/L2/L3), instrucciones/tipo, accesos a memoria, fallos de página, ...

Medidas Rendimiento

- Los procesadores ofrecen muchas posibilidades de configuración
 - Políticas de remplazo de las caches, del prefetcher, del DVFS
 - Activar o desactivar políticas
 - Particionado de recursos
- Soporte Hardware → Performance Monitoring Unit
 - Los procesadores incluyen unidades de estadísticas
 - Ayudan a entender mejor el comportamiento
 - Fundamental para encontrar las configuraciones optimas

Fugaku A64FX (7nm)
N1 en el TOP500
(Junio 2020)
Basado en ARMv8
Implementa SVE
de 512bits

