

ARQUITECTURA DE SISTEMAS PARALELOS II.

PROGRAMA (1998/99)

Facultad de Informática y Estadística. Cuarto Curso de Ingeniería Informática.

Profesores: D. Antón Civit Balcells, D. Fernando Díaz del Río y profesor a contratar.

Condicionantes.

En principio, el alumno de Segundo Ciclo debe formarse cara a dos posibilidades: el acceso al mundo laboral y la continuación de estudios en el Tercer Ciclo. Para este último caso, parece claro que la formación debe ser generalista, orientada a los conceptos y bases teóricas que permitan una visión académica y científica de la Arquitectura. Con respecto al mundo laboral, la formación de más alto nivel del Segundo Ciclo debe incluir todas las arquitecturas avanzadas, el diseño y demás cuestiones que no era posible estudiar con detalle en Primer Ciclo. Desde un punto de vista más práctico, en nuestro entorno siguen siendo necesarios Licenciados que conozcan con profundidad los sistemas más potentes, y que sean capaces de obtener el máximo rendimiento de ellos. Además, serán capaces de adaptarse con mayor facilidad a los avances que con toda probabilidad aparecerán en el futuro, a una velocidad muy superior a la de otras disciplinas.

Se supone una base sólida en los conceptos fundamentales de sistemas paralelizables, y en la interconexión de múltiples procesadores a otros elementos como sistemas de memoria o entrada/salida, adquirida en la asignatura Arquitectura de Sistemas Paralelos I.

Objetivos.

El objetivo general de esta asignatura es el estudio de las arquitecturas fuertemente paralelas y multiprocesadoras alternativas. En concreto, los sistemas paralelos y distribuidos se estudian con cierto detalle, incluyendo aspectos sobre sistemas operativos, compiladores, programación y algoritmos.

La asignatura se enfoca sobre todo desde el punto de vista del diseño, estimando las prestaciones que permitan elegir entre distintas alternativas. Esto es útil no sólo a la hora de diseñar nuevas máquinas, sino también cuando se trata de elegir la configuración más apropiada para una aplicación concreta. Al igual que en el primer ciclo, un objetivo fundamental es dotar de una visión unificada de los sistemas. Esto es más importante si cabe en arquitecturas que tratan de aprovechar el paralelismo, en las que el compilador ó las aplicaciones influyen decisivamente en las prestaciones obtenidas.

En consecuencia, al finalizar el curso, el alumno debe:

- 1 ☐ Tener una idea clara del funcionamiento de las distintas arquitecturas paralelas: procesadores superescalares, vectoriales, arquitecturas fuertemente acopladas, multiprocesadores,...
- 2 ☐ Poseer la base necesaria para la optimización de software y/o el desarrollo de compiladores o sistemas operativos para estas arquitecturas.

Temario de la asignatura Arquitectura de Sistemas Paralelos II.

Tema 0. Presentación del curso.

- 1 ☐ Profesorado de la asignatura.
- 2 ☐ Metodología y actividades docentes.
- 3 ☐ Criterios y sistema de evaluación.
- 4 ☐ Programación de la asignatura.

- 5□ Bibliografía recomendada.

Descripción y objetivos:

En este tema se exponen las normas que rigen la docencia de la asignatura, y se informa de horario y lugar de tutorías, documentación y modo de acceso, temario, etc.

Tema 1. Paralelismo en arquitecturas segmentadas avanzadas.

- 1□ Técnicas software: desenrollado de bucles.
- 2□ Paralelismo a nivel de instrucciones: conceptos y ejemplos.
- 3□ Planificación dinámica.
 - Algoritmo del marcador (*scoreboard*).
 - Algoritmo de Tomasulo.
- 4□ Predicción dinámica de saltos.
- 5□ Procesadores RISC superescalares.
 - Búsqueda de varias instrucciones.
 - Ejecución especulativa.
- 6□ Técnicas software.
 - Especulación.
 - Segmentación software.
 - Planificación de trazas.
- 7□ Paralelismo de instrucciones disponible.
- 8□ Procesadores CISC segmentados.
 - Técnicas clásicas: macro y microsegmentación.
 - Técnicas actuales: partición del conjunto de instrucciones.
- 9□ Procesadores CISC superescalares.

Descripción y objetivos:

Este tema comienza introduciendo el concepto de Paralelismo a Nivel de Instrucciones, que ya estaba implícito en la asignatura anterior. Así, las técnicas anteriores se observan desde la perspectiva de aprovechar en cierta medida este nivel de paralelismo, aunque limitándose a acercarnos al ideal $CPI=1$ de una máquina segmentada. En este tema se plantea superar este límite mediante las tres técnicas posibles:

- Reduciendo el número de ciclos por instrucción (CPI) mediante emisión múltiple de instrucciones (procesadores superescalares).
- Reduciendo el número de instrucciones necesario para realizar una tarea (procesadores VLIW). En realidad este también es el caso de los procesadores CISC, que se estudian más adelante en el tema.
- Reduciendo la “duración” del ciclo de reloj, fundamentalmente mediante unidades funcionales segmentadas (procesadores *superpipeline*).

Se dedica especial atención a los procesadores superescalares, que han dado lugar a un aumento espectacular en las prestaciones de los sistemas uniprocador en los últimos años. Se abordan las técnicas para ejecución especulativa, ilustrándolas con procesadores reales, sobre todo el PowerPC620. También se estudian las técnicas software para incrementar el paralelismo de instrucciones disponible (desenrollado de bucles, planificación de trazas, etc.). Se revisa el paralelismo disponible para comprender en qué medida puede seguirse esta línea en futuros diseños.

Por otro lado, se amplía todo lo anterior a los procesadores CISC, en los que, debido a la mayor complejidad de las instrucciones, las dependencias estructurales, de datos ó de control son más frecuentes y las soluciones más complicadas. Se consideran por un lado la aproximación clásica de segmentar o bien las

macroinstrucciones ó las microinstrucciones, y por otro la técnica más actual de dividir en dos los tipos de instrucciones mezclando técnicas RISC con el microcódigo. En este punto cobra especial importancia el ejemplo de la arquitectura i960, que dispone de versiones escalares (960KB) y superescalares (960CA y Hx), además de utilizar un núcleo RISC del mismo tipo que la mayoría de los actuales procesadores CISC segmentados por lo que se cubre prácticamente todo el tema. Téngase en cuenta que el uso de estas técnicas pueden utilizarse para mantener la compatibilidad en una familia CISC (caso del i486, Pentium y Pentium Pro, y del M68060) o bien para aumentar la densidad de código, caso típico de aplicaciones empotradas (por ejemplo, el i960). Además, hay material de prácticas disponible para cada una de estas versiones, por lo que las prácticas resultan un complemento a las clases de teoría, donde los alumnos tienen oportunidad de comprobar *in situ* lo aprendido.

En todo el tema, se incluyen otros ejemplos de máquinas reales (IBM RS6000, Digital Alpha AXP, PA-RISC, todos ellos accesibles para aquellos alumnos interesados), detallando las soluciones particulares. Para la parte correspondiente a los CISC se usan muy frecuentemente como ejemplo los procesadores de la familia x86, sobre todo el Pentium, dado que son los procesadores más abundantes en nuestro entorno.

Tema 2. Sistemas Multiprocesadores.

- 1 ☐ Aplicaciones.
- 2 ☐ Multiprocesadores con memoria compartida.
 - Sistemas basados en bus común.
 - Protocolos de coherencia de cachés.
 - Sistemas escalables.
- 3 ☐ Multicomputadores.
 - Utilización conjunta de estaciones de trabajo.
 - Otras arquitecturas.
- 4 ☐ Software de sistemas paralelos.
 - Lenguajes.
 - Compiladores.
 - Sistemas operativos.

Descripción y objetivos:

En este tema se presentan arquitecturas con varios procesadores, que suponen la manera natural de incrementar las prestaciones de las uniprosesadores, estudiadas anteriormente. En primer lugar se da cuenta de las aplicaciones en las que estas técnicas suponen un aumento de prestaciones, introduciendo el paralelismo de granularidad más gruesa y de datos. Se tratan con más detalle las arquitecturas multiprosesador con memoria compartida, con un número pequeño de procesadores, que son las que tienen un futuro más prometedor, estudiándose algunos ejemplos de máquinas reales. Se estudian los protocolos de coherencia de caches, y se presenta el concepto de escalabilidad, tratando los sistemas distribuidos con memoria compartida. Se utilizan técnicas simples de estimación de prestaciones para justificar los diseños y evaluar las diferentes alternativas, incluyendo los protocolos de coherencia de cachés.

Para el siguiente apartado, se estudian los aspectos generales de los sistemas basados en paso de mensajes como la coherencia de caches en multiprosesadores, y sus soluciones particulares (protocolos basados en directorios), se incluyen las configuraciones de varias estaciones funcionando como *clusters* y los métodos para escritura de software en estos sistemas (PVM). Los alumnos pueden acceder más fácilmente a estos sistemas. Brevemente se describen otras arquitecturas (Hiperubos, de flujo de datos).

Por último, aspectos como la programación y los sistemas operativos son discutidos. Se pretende que el alumno comprenda que el software de estos sistemas requiere la adopción de técnicas específicas, y se abordan temas como la sincronización, consistencia de memoria, etc.

Tema 3. Arquitecturas Vectoriales.

- 1 ☐ Arquitectura vectorial básica.
 - Arquitecturas de registro y de memoria.
 - Versión vectorial de DLX: DLXV.
- 2 ☐ Modelos de evaluación de prestaciones específicos.
- 3 ☐ Técnicas software de vectorización.
- 4 ☐ Procesadores vectoriales paralelos.

Descripción y objetivos:

En este último tema se tratan brevemente los procesadores vectoriales, especialmente diseñados para aprovechar el paralelismo de datos. Se vuelve al procesador ejemplo DLX, al cual se incluyen características vectoriales (DLXV). Se explican las distintas técnicas software de vectorización, así como modelos específicos de prestaciones para estos sistemas: conceptos como prestaciones pico frente a medias, inicialización, etc. Se incluyen ejemplos de máquinas reales, sobre todo del CRAY C-90. También se describen aplicaciones donde estos procesadores pueden ofrecer mejor rendimiento. Por último, se citan los sistemas vectoriales paralelos.

Sin embargo, parece que las últimas tendencias están dejando un poco obsoletas estas máquinas. Procesadores superescalares con ejecución especulativa y varias unidades punto flotante pueden ofrecer casi las mismas prestaciones a un precio órdenes de magnitud inferior. De hecho, el CONVEX C240 que estaba disponible en el Centro de Informática Científica de Andalucía (CICA) ha sido recientemente dado de baja. Por este motivo, se dedica proporcionalmente menos tiempo a este tema.

Prácticas de Arquitectura de Sistemas Paralelos II.

La asistencia al laboratorio y la entrega de las memorias correspondientes a cada prácticas son requisitos indispensables para aprobar la asignatura. Las prácticas se evaluarán como “apto” o “no apto”, pudiendo tener una influencia máxima en la nota final del 25%.

Todas las prácticas incluyen una explicación previa (aprox. 1 hora) de los conceptos necesarios y de la metodología a seguir. La duración completa, incluyendo la explicación previa, elaboración del software y realización de pruebas es de alrededor de cinco horas, aunque en función de la disponibilidad de laboratorios, las prácticas se podrán subdividir en varias sesiones.

Se impartirán las siguientes prácticas a lo largo del cuatrimestre:

Práctica 1. Algoritmos de Tomasulo y Marcador.

Práctica 2. Optimización de software para un procesador superescalar i960CA.

Práctica 3. Ejecución de software en sistemas multiprocesador.

Evaluación.

El examen de la asignatura constará de una parte teórica, que tendrá un valor en torno al 30% de la nota del examen, y otra de problemas (aproximadamente el 70% de la puntuación final), en la cual se permitirá el uso de libros y apuntes.

BIBLIOGRAFIA BASICA.

- [H&P96] J.L. Hennessy, D.A. Patterson "Computer Architecture. A Quantitative Approach". Morgan-Kaufmann (Second Edition), 1996.
- [STON90] H.S. Stone, "High-Performance Computer Architecture". Addison-Wesley (Second Edition) 1990.
- [HWAN93] K. Hwang "Advanced Computer Architecture: Parallelism, Scalability, Programmability". McGraw-Hill, 1993.
- [KOGG81] P. M. Kogge "The Architecture of Pipelined Computers". McGraw-Hill, 1981.
- [LILJ91] D.J. Lilja, de., "Architectural Alternatives for Exploiting Paralelism". IEEE Comp. Soc. Press, 1991.
- [JOHN91] M. Johnson, "Superscalar Microprocessor Design". Prentice-Hall, 1991.
- [ALMA94] G.S. Almansí, A.Gottlieb "Highly Parallel Computing". Benjamin/Cummings (Second Edition), 1994.

BIBLIOGRAFIA COMPLEMENTARIA.

- [TRED95] N. Tredennick, "Technology and Business: Forces Driving Microprocessor Evolution". *Proc. Of the IEEE*, vol. 83, no. 12, Diciembre 1995.
- [H&P95] D.A. Patterson, J.L. Hennessy "Organización y Diseño de Computadores. La Interfaz Hardware/Software". McGraw-Hill, 1995.
- [STAL96] W. Stallings, "Computer Organization and Architecture. Designing for Performance". Prentice Hall, 1996 (Fourth Edition).
- [ANTO96] J.L. Antonakos, "An Introduction to the Intel Family of Microprocessors". Prentice Hall, 1996 (Second Edition).
- [PATT83] D.A. Patterson, "Microprogramming". *Scientific American*, 248:3, pp. 36-43, 1983
- [JAIN91] R. Jain, "The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation and Modeling". John Wiley & Sons, 1991.
- [GILA95] R. Giladi, N. Ahituv "SPEC as a Performance Evaluation Measure". *IEEE Computer*, August 1995 (pp. 33-42).
- [MIRA92] S. Mirapuri, M. Woodacre, N. Vasseghi "The Mips R4000 Processor". *IEEE Micro*, April 1992 (pp. 10-22).
- [WANG93] C. Wang, F. Emmett "Implementing Precise Interruptions in Pipelined RISC Processors". *IEEE Micro*, August 1993 (pp. 36-43).
- [SMIT95] J.E. Smith, G.S. Sohi, "The Microarchitecture of Superscalar Processors". *Proc. of the IEEE*, vol. 83, no. 12, December 1995.
- [HWU95] W.W. Hwu *et al.* "Compiler Technology for Future Microprocessors". *Proc. of the IEEE*, vol. 83, no. 12, December 1995.
- [WALL93] D. W. Wall, "Limits of Instruction-Level Parallelism". WRL Research Report 93/6, DEC 1993.
- [ALPE93] D. Alpert, D. Avnon "Architecture of the Pentium Microprocessor". *IEEE Micro*, June 1993 (pp. 11-21).
- [MCMA95] S.C. McMahan *et al.* "6x86 : The Cyrix Solution to Executing x86 Binaries on a High Performance Microprocessor". *Proc. of the IEEE*, vol. 83, no. 12, December 1995.
- [MYER88] G.J. Myers, D.L.Budde "The 80960 Microprocessor Architecture". John Wiley & Sons, 1988.
- [LENO95] D.E. Lenoski, W. Weber "Scalable Shared-Memory Multiprocessing". Morgan-Kaufmann, 1995.
- [KAIN96] R.Y. Kain, "Advanced Computer Architecture: A Systems Design Approach". Prentice-Hall, 1996.
- [COSN95] M. Cosnard, D. Trystram "Parallel Algorithms and Architectures". Thomsom Computer Press, 1995.
- [GEIS94] A. Geist *et al.* "PVM: Parallel Virtual Machine - A Users' Guide and Tutorial for Networked Parallel Computing". The MIT Press, 1994.
- [VARM94] A. Varma, C.S. Raghavendra (Ed.) "Interconnection Networks for Multiprocessors and Multicomputers. Theory and Practice". IEEE Comp. Soc. Press, 1994.