

ARQUITECTURA DE SISTEMAS PARALELOS I.

PROGRAMA (1998/99)

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

Profesores: D. José Luis Sevillano Ramos, D. Fernando Díaz del Río y profesor a contratar.

Condicionantes.

- Se trata de la primera parte de una asignatura global, de forma que además de estudiarse los primeros conceptos de paralelismo, ha de servir de introducción a la segunda parte de la asignatura. De hecho, los programas de ambas asignaturas se han pensado en común y coinciden algunos conceptos de su metodología y su desarrollo.

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 estructura de computadores (exo y endoarquitectura) y en la interconexión de procesadores a otros elementos como sistemas de memoria o entrada/salida. Sin embargo, se tiene en cuenta la presencia de alumnos provenientes del Plan de Estudios antiguo o de otras Universidades y cuya formación puede ser muy variada. Por tanto, siempre que sea posible, se repasan los conceptos fundamentales para seguir la asignatura, y en todo caso se facilita el material y la bibliografía precisa para adquirir la base necesaria.

Objetivos.

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

La asignatura se enfoca sobre todo desde el punto de vista del diseño, introduciendo técnicas de estimación de 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 o las aplicaciones influyen decisivamente en las prestaciones obtenidas.

En consecuencia, al finalizar la asignatura, el alumno debe:

- a ☐ Tener una idea clara del funcionamiento de las distintas arquitecturas que son actualmente la base de la explotación del paralelismo: RISC-CISC segmentados, superescalares, ...
- b ☐ Manejar técnicas de evaluación de prestaciones que le permitan determinar la mejor opción entre las distintas alternativas de diseño o configuración.
- c ☐ 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 I.

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: Prestaciones y Soporte al Paralelismo de Sistemas de Memoria y Entrada/Salida.

- 1 ☐ Parámetros cuantitativos para la evaluación de prestaciones.
- 2 ☐ Programas para la evaluación de prestaciones (*benchmarks*).
- 3 ☐ Análisis cuantitativo de las jerarquías de memoria y memorias compartidas.
 - Caches.
 - Mejoras en el porcentaje de fallos
 - Mejoras en la penalización.
 - Mejoras en el tiempo de acceso.
 - Memoria Principal.
 - Memoria Virtual.
 - Memoria centralizada frente a distribuida
- 4 ☐ Prestaciones de los sistemas E/S.
 - Dispositivos
 - Buses
 - Medidas de prestaciones
 - Diseño de sistemas de E/S.

Descripción y objetivos:

El objetivo fundamental del tema 1 es introducir los distintos parámetros usados a lo largo del curso para la evaluación de prestaciones, y su aplicación a los sistemas de memoria y entrada/salida paralelizables. Estos parámetros incluyen los ciclos por instrucción (CPI), el tiempo de ejecución, *benchmarks* (programas para evaluar el rendimiento), etc. Se supone que los alumnos poseen ya un conocimiento suficiente de los elementos de la jerarquía de memoria y de los distintos tipos de dispositivos E/S (discos, pantallas, buses, etc.), aunque se repasarán los conceptos fundamentales como introducción al nuevo curso y para adaptar la base de aquellos alumnos provenientes de otras Universidades o planes de estudio. En concreto, la organización a nivel de bloques de los caches, TLB y memoria virtual son repasados con cierto detalle. De este modo, se aplica el enfoque cuantitativo a sistemas con los que el alumno está familiarizado, lo que permite una más rápida asimilación. Además puede cuantificarse el efecto de varios mecanismos de mejora lo cual permite encontrar soluciones de compromiso. Finalmente, se pretende ofrecer una visión unificada de todo el sistema: jerarquía de memoria, discos, buses y sistema operativo, con ejemplos de varios sistemas reales (en particular los DEC Alpha 21064 y los Intel 80x86). Además se tratan las arquitecturas con memoria compartida. Se estudian los protocolos de coherencia de caches, 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.

Tema 2. Paralelismo en Arquitecturas Segmentadas (pipeline) Básicas.

- 1 ☐ Descripción del procesador ejemplo (DLX).
- 2 ☐ Segmentación.
- 3 ☐ Dependencias y bloqueos.
 - estructurales.
 - de datos.
 - tipos
 - caminos de desvío o "bypass"
 - planificación estática
 - de control
 - Reducción de los bloqueos. Saltos retrasados
 - Predicción estática
- 4 ☐ Excepciones en procesadores segmentados.
- 5 ☐ Instrucciones de duración variable.
- 6 ☐ Interrupciones precisas.

Descripción y objetivos:

En este tema se introducen los conceptos básicos de la segmentación como técnica de diseño de procesadores RISC. Se trata de un tema importante para entender la mayoría de las máquinas actuales de altas prestaciones, así que se le dedica más tiempo que a ningún otro. Se utiliza el procesador ejemplo DLX de [H&P96] para exponer las ideas fundamentales, sobre todo las cuestiones relativas a los conflictos por recursos y los bloqueos asociados. Se presentan técnicas especiales para minimizar los efectos de los bloqueos de control (saltos retrasados, predicción de saltos) y de las dependencias de datos: desvíos (*bypassing*), etc. Las interrupciones en estos sistemas también se estudian en detalle, explicando mecanismos para introducir interrupciones *precisas* (Fichero de Historia, de Futuro, etc.). Se hacen continuas referencias a procesadores reales, sobre todo los Intel 960 y el MIPS R4000, pero también de los SPARC y HP-PA.

Prácticas de Arquitectura de Sistemas Paralelos I.

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. Prestaciones de la memoria cache en distintas máquinas.

Práctica 2. Estudio de la segmentación en DLX.

Práctica 3. Optimización de software para un procesador i960KB.

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.
- [ALMA94] G.S. Almansi, 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. Emnett "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.