Departamento de Arquitectura de Computadores
Universidad de Málaga

TESIS DOCTORAL

Factorización LU de matrices dispersas en multiprocesadores

Rafael Asenjo Plaza

Málaga, 15 de Diciembre de 1997

Dr. Emilio López Zapata
Catedrático del Departamento de
Arquitectura de Computadores de
la Universidad de Málaga

CERTIFICA:

Que la memoria titulada *"Factorización LU de matrices dispersas en multiprocesadores"*, ha sido realizada por D. Rafael Asenjo Plaza bajo mi dirección en el Departamento de Arquitectura de Computadores de la Universidad de Málaga y concluye la Tesis que presenta para optar al grado de Doctor en Ingeniería de Telecomunicaciones.

Málaga, 31 de Octubre de 1997

Fdo: Dr. Emilio López Zapata
Director de la Tesis Doctoral
Director del Departamento de
Arquitectura de Computadores

A Mª Ángeles.

Muchos años después, [...] Aureliano Buendía había de recordar aquella tarde remota en que su padre lo llevó a conocer el hielo.

Cien años de soledad.
Gabriel García Márquez.

"Cualquier tecnología suficientemente avanzada es indistinguible de la magia."

Tercera ley de Arthur C. Clarke.

Agradecimientos

Muchas veces me he planteado la opción de comenzar y acabar esta sección únicamente con la palabra “Gracias.”... y nada más. Sencillamente por evitar el resumir en una sola página la mención de tantas personas a las que debo mucho:

A mi director de Tesis, Emilio L. Zapata, fuente de motivación y al que considero mi maestro y ejemplo de dedicación. Muchas gracias por, como tú dices, “enseñarnos a pescar”.

A David Padua por la atención, dedicación en largas horas de reuniones, y su espléndida hospitalidad. También a Iain Duff por sus consejos y supervisión. A Juan Touriño, Oscar Plata, Eladio Gutiérrez, Guillermo Pérez, Gerardo Bandera y Manuel Ujaldón, les debo su productiva colaboración.

A mis padres, siempre preocupados por mi felicidad, alentadores y reconfortantes en los momentos bajos y a los que es imposible pagar su cariño como se merecen, también debo este trabajo. A mi hermana, cómplice y amiga, tanto en la infancia como en la madurez.

A la memoria de Gabriel y Adelaida, y a María Teresa, que gracias a su esfuerzo y lucha son, seguramente en un mayor porcentaje de lo que incluso yo creo, responsables de la gran cantidad de oportunidades que he tenido.

A los amigos y compañeros de departamento, Nico, Mario, Juan, Oswaldo, Manuel, Julio, Felipe, Pablo, Julián, Andrés, Francisco, Mª Antonia, Patricia, Ezequiel y Mª Carmen, maestros del trabajo bien hecho y a todos los que mantienen día a día el departamento como un lugar de camaradería, alegre, distendido y bullicioso de trabajo.

A Nacho, Javi, Pedro, Juan, Miguel, Jesús, José, Pepe y Ester, y demás amigos pese a la distancia y los años.

A los miembros del departamento de Computer Science de la Univ. de Illinois, especialmente a Yuan Lin, Bill Pottenger y Sheila Clark. También a Jacko Koster y los demás compañeros del Parallel Algorithm Team de CERFACS (Francia).

A K.I.M. McKinnon y su grupo del Depto. de Matemáticas de la Univ. de Edimburgo, así como a los miembros del EPCC de la misma universidad, por su soporte e introducción en la programación paralela del Cray T3D.

También agradecemos al CIEMAT (Madrid), CINECA (Italia), EPFL (Suiza) y Cray Research Inc. (Minnesota, USA), por habernos brindado acceso a sus plataformas Cray T3D y T3E.

Agradecemos el soporte y la financiación de este trabajo y de las estancias realizadas en el extranjero a los proyectos TIC96-1125-C03 de la CICYT, BRITE-EURAM III BE95-1564 de la Unión Europea, ERB4050P1921660 del programa de Movilidad y Capital Humano y al programa TRACS (Training and Research on Advanced Computing Systems) del Centro de Computación Paralela de Edimburgo (EPCC).

Aunque en último lugar, con mayor énfasis quiero agradecer a Mª Ángeles todo el cariño, apoyo incondicional, alegrías, colaboración y comprensión que ella me da, y con la cual vivo cada día con una sonrisa en la cara.

Índice General

Prefacio

A menudo nos encontramos frente a compromisos del tipo prestaciones/coste. Esta dualidad que se manifiesta en todos los ámbitos, aparece, como no, en muchos de los párrafos de esta memoria. Ya en la propia tarea de realizar una tesis, aflora un primer compromiso: ante un problema, bien definido por el director de tesis, al estudiante se le presentan diversos caminos sin explorar. Cuando compensa seguir por un camino sin saber lo que hay al final, o volver a atrás para tomar otra dirección es una cuestión de intuición y presenta un ineludible compromiso. Con ingenio y perseverancia se pueden encontrar distintas soluciones al final de sus correspondientes caminos. Si se ha explorado una buena parte del mapa y se han topografiado las regiones hasta entonces vírgenes, el relato del viaje y la descripción de las soluciones, constituyen una metodología apropiada para realizar la tesis. Es claramente una tarea de investigación, pero no podemos negar que en el proceso también hay una gran componente de ingeniería, donde aparece el compromiso entre el tiempo invertido y el número de caminos explorados.

En la presente tesis, el problema estaba bien definido: encontrar soluciones paralelas eficientes para algoritmos de factorización dispersa LU y aplicar las técnicas encontradas en el camino a otros problemas irregulares. Nosotros hemos explorado tres rutas: paralelización manual, semi-automática y por último, automática. En este punto, aparece de nuevo el compromiso entre las prestaciones del código generado y el coste en términos de tiempo de desarrollo. Es decir, la paralelización manual consigue las mejores prestaciones a cambio de un gran esfuerzo en el desarrollo y depuración de los códigos. La situación al paralelizar usando herramientas automáticas es recíproca, situándose la paralelización semi-automática en un punto intermedio entre los otros dos paradigmas.

La motivación que soporta el esfuerzo que realizamos en resolver problemas irregulares es definitiva: se estima que al menos el cincuenta por ciento de todos los programas científicos son irregulares y que más del cincuenta por ciento de todos los ciclos de reloj consumidos en supercomputadores son consumidos por este tipo de códigos.

Dentro de los problemas irregulares hemos elegido como caso de estudio la resolución de sistemas de ecuaciones dispersos. Esta elección se justifica en que este problema abarca varios de los paradigmas presentes en otros códigos irregulares. De esta forma, las técnicas encontradas son útiles a otros problemas con similares características al primero: principalmente algoritmos que presentan llenado o movimiento de datos dentro de una estructura irregular.

De cualquier modo, la resolución paralela de sistemas de ecuaciones dispersos es un problema relevante por sí sólo. Así como es una realidad que los grandes problemas necesitan grandes computadores para ser resueltos, es igualmente cierto que en el corazón de la mayoría de los problemas computacionales a gran escala está la solución de grandes sistemas de ecuaciones dispersos.

Sin embargo, durante la realización de esta tesis, hemos querido ir más lejos, dirigiendo nuestra búsqueda con el objetivo de encontrar soluciones generales para la paralelización de códigos irregulares, y no el de resolver únicamente el problema particular de la factorización de matrices dispersas. En este punto es donde entran en juego las técnicas de compilación ya sean automáticas o semi-automáticas.

De este modo, la memoria se organiza en dos partes bien diferenciadas: la primera aborda directamente el problema de la paralelización manual de códigos dispersos de resolución de ecuaciones; la segunda parte trata de aplicar los conocimientos adquiridos en la primera, para extraer técnicas genéricas que puedan ser aplicadas automática o semi-automáticamente mediante un compilador.

La primera parte tiene tres capítulos. En el primero, formulamos el problema a resolver, enunciamos las distintas alternativas para abordarlo y profundizamos en una de ellas. Además describimos las características relevantes y las estructuras de datos apropiadas al trabajar con matrices dispersas. En el capítulo 2 introducimos los conceptos básicos de paralelismo (arquitecturas y programación paralela) utilizados a lo largo de esta memoria de forma que podamos seguir sin problemas su lectura. Finalmente en el capítulo 3 desarrollamos un eficiente programa paralelo para la resolución de sistemas de ecuaciones dispersos. Este algoritmo es evaluado experimentalmente y comparado con un algoritmo estándar de resolución, comprobándose las buenas prestaciones que proporciona tanto en secuencial como en paralelo. Además de este resultado, el manejo con este algoritmo nos ha permitido familiriazarnos con el problema, adquiriendo una visión global que nos facilita ahora abordar la segunda parte de la tesis. En esta primera parte nos ha sido de gran ayuda la colaboración con Iain Duff durante nuestra estancia en CERFACS, así como el entrenamiento en la programación del Cray T3D que nos brindaron en el EPCC de la Universidad de Edimburgo. Los trabajos publicados de esta primera parte de la tesis se encuentran en [17, 21].

La segunda parte la organizamos en dos capítulos. Abordamos primero la paralelización semi-automática de códigos dispersos en el capítulo 4. Aquí se determinan las limitaciones de los compiladores actuales de paralelismo de datos al intentar expresar códigos dispersos de factorización. Para salvar esas limitaciones proponemos una extensión sencilla y eficiente al lenguaje HPF-2 que resulta válida para otros problemas dispersos o irregulares. En el capítulo 5 se da un paso más en complejidad, tratando de delegar completamente el problema de la paralelización en el compilador. La tecnología actual de compiladores paralelos no es capaz de manejar eficientemente problemas irregulares, por lo que presentamos en este capítulo algunas técnicas que permiten solventar este inconveniente. Para ello hemos partido de un algoritmo secuencial de factorización dispersa, identificando las técnicas necesarias para su paralelización automática. Para justificar la implementación de estas técnicas en un compilador paralelo hemos comprobado que también son aplicables a una gran variedad de códigos numéricos dispersos. En esta segunda parte hemos contado con el apoyo de David Padua durante nuestras estancias en la Universidad de Illinois. En [19, 18, 14, 145, 15, 16] encontramos más detalladamente la evolución del trabajo desarrollado en esta línea.

Finalmente, en el capítulo de conclusiones, sintetizamos las principales aportaciones de esta tesis, así como enumeramos las principales líneas en las que seguiremos trabajando en el futuro.

Recientemente celebraron en la Universidad de Illinois el nacimiento simbólico de HAL 9000, que en la novela de Arthur C. Clarke "2001: A Space Odyssey" despierta en Enero de 1997 en el Digital Computer Laboratory de la misma Universidad. Es cierto que aún no existe ninguna máquina tan inteligente ni versátil como HAL, pero siempre he sido optimista al imaginarme el futuro y la evolución de la ciencia... Como dijo David J. Kuck con motivo de esta celebración: "La computación paralela práctica es algo escurridizo por el momento. HAL tendría que ser una máquina paralela, y hasta que el problema del paralelismo práctico sea resuelto, no seremos capaces de construir HAL".