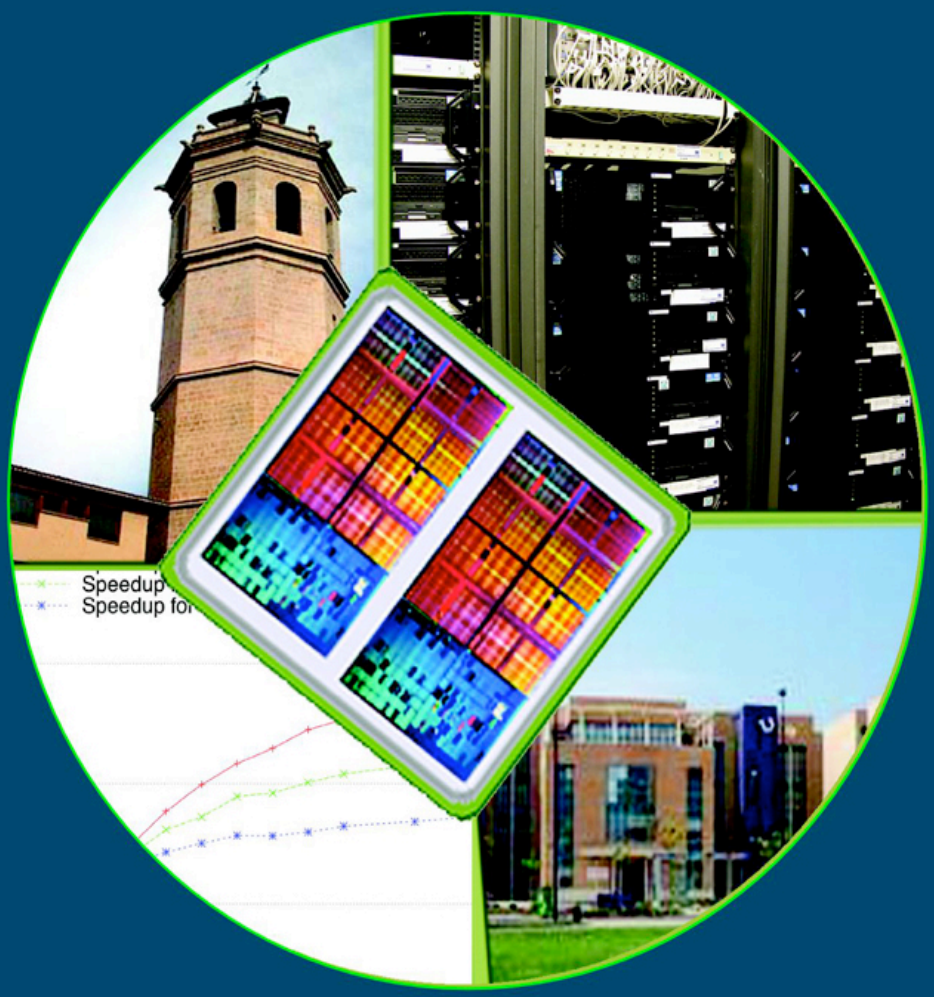
UNIVERSITAT JAUME·I

XIX JORNADAS DE PARALELISMO

José Manuel Badía Contelles, Francisco Daniel Igual Peña,
Enrique S. Quintana Ortí, Gregorio Quintana Ortí (eds.)



COMPACT disc

Col·lecció
e-Treballs d'Informàtica i Tecnologia
Núm. 8

BIBLIOTECA DE LA UNIVERSITAT JAUME I. Dades catalogràfiques

JORNADAS DE PARALELISMO
(19es. 2008. Castelló de la Plana)

XIX Jornadas de paralelismo [Recurs electrònic] : Castellón 17, 18 y 19 de septiembre de 2008 / José M. Badía … [et al.] (eds.) —Castelló de la Plana : Publicacions de la Universitat Jaume I, D.L. 2008

1 disc òptic (CD-ROM) + 1 llibret. —(e-Treballs d'informàtica I tecnologia ; 8)

ISBN 978-84-8021-676-0

1.Processament en paral·lel (Ordinadors) – Congressos. I. Badía Contelles, José Manuel, ed. II. Universitat Jaume I (Castelló). Publicacions de la Universitat Jaume I. III. Títol. IV. Sèrie.

681.324(063)

Edita: Publicacions de la Universitat Jaume I. Servei de Comunicació i Publicacions
Campus del Riu Sec. Edifici Rectorat i Serveis Centrals.12071 Castelló de la Plana
Fax 964 72 88 32
http://www.tenda.uji.es e-mail: publicacions@uji.es

ISBN: 978-84-8021-676-0

Dipòsit legal: V-3132-2008

Imprimeix: Digital Gràfic. Tel. 961 710 317
http://www.digitalgrafic.net c/e: info@digitalgrafic.net

Computación de altas prestaciones sobre arquitecturas paralelas heterogéneas (CAPAP-H)

Compilación para sistemas de altas prestaciones

Docencia en ATC

Evaluación de compiladores comerciales usando SPEC CPU2006

Sergio Aldea, Diego R. Llanos, Arturo González-Escribano[1]

Resumen— **La elección del compilador es una decisión de especial importancia a la hora de obtener el máximo rendimiento de una aplicación, no sólo para mejorar el tiempo de su ejecución secuencial sino también para explorar la posibilidad de paralelizar automáticamente parte de su código. El presente trabajo busca evaluar las capacidades de optimización de tres compiladores utilizando la suite SPEC CPU2006, analizando además las capacidades de paralelización automática que ofrecen al programador. Los compiladores analizados son el Intel C/Fortran Compiler 10.0, Sun Studio 12 y GCC 4.1.2. Se evaluará el rendimiento de dichos compiladores ejecutando la suite SPEC CPU2006 sobre diferentes arquitecturas.**

Palabras clave— **Compiladores, GCC, Intel, Sun, paralelismo, SPEC CPU2006, benchmark**

I. Introducción

ESTE trabajo busca comparar tres de los principales conjuntos de compiladores comerciales: Intel, Sun y GCC. El estudio se ha realizado sobre varias plataformas de 32 y 64 bits utilizando la suite SPEC CPU2006 como medio para comparar el rendimiento de los compiladores sobre los distintos benchmarks y configuraciones hardware. El objetivo de este trabajo es ofrecer unos resultados que puedan ayudar a elegir un compilador u otro, en función de los objetivos o presupuesto, y si realmente es necesario adquirir software comercial para obtener unos mejores resultados.

El trabajo se estructura como sigue. La sección 2 describe el conjunto de benchmarks SPEC CPU2006, con una breve descripción de su evolución y utilidad; la sección 3 define los SUT (System Under Test), es decir, las máquinas sobre las cuales se ha realizado los experimentos; las secciones 4 y 5 ofrecen los resultados más significativos y de mayor interés de entre los experimentos realizados con los dos conjuntos que forman el benchmark, el de números enteros, y el de punto flotante. Por último, en la sección 6 se realiza una valoración global y se describen las principales conclusiones extraidas de los resultados.

II. Descripción del benchmark SPEC CPU2006

SPEC (Standard Performance Evaluation Corporation) es un consorcio sin fines lucrativos en el que se engloban productos software, fabricantes de ordenadores, procesadores, etc. Fue fundado en 1988 por un grupo reducido de vendedores de ordenadores en busca de un test de rendimiento estándar y que ofreciera unos resultados realistas y comparables. Con el tiempo, SPEC ha crecido hasta convertirse en uno de los órganos de estandarización de medición de rendimiento de más exito, compuesto por más de 60 compañías.

Para conseguir estos objetivos, SPEC toma aplicaciones reales, utilizadas en diversos campos, y las utiliza como benchmarks con diversas cargas de trabajo, con el fin de obtener medidas de rendimiento. La primera suite fue lanzada en 1989[1], e incorporaba únicamente diez benchmarks. SPEC actualiza su suite periódicamente, debido a la evolución de los ordenadores, y el crecimiento y aumento de la complejidad de las aplicaciones. La última versión, SPEC CPU2006, fue lanzada el 24 de Agosto de 2006[2], incluyendo 29 benchmarks, que se engloban en dos grupos: los que trabajan con números enteros (**CINT2006**) y los que trabajan con números en punto flotante (**CFP2006**). SPEC CPU2006 utiliza como máquina de referencia una estación de trabajo Sun Ultra Enterprise 2 con un procesador UltraSPARC II a 296MHz de 2 núcleos y 2 GB de memoria RAM. Los valores de rendimiento medidos se expresan en unidades relativas a esta máquina de referencia.

En la actualidad, la utilización de esta suite esta bastante extendida, y es utilizada por las principales empresas fabricantes de procesadores y de ordenadores en general, publicando en sus portales web los resultados obtenidos, como pueden los casos de Intel[3], IBM[4], Sun Microsystems[5][6] o Fujitsu-Siemens[7].

III. Entornos de ejecución

Para este trabajo se han utilizado dos sistemas con Linux, instalando en uno de ellos las versiones de 32 y 64 bits del mismo sistema operativo. La configuración de los sistemas es la siguiente:

El primer sistema (**SUT1**) monta un procesador Intel®Core™2 CPU E6300 a una velocidad de 1.86 GHz, con un bus de 1066 MHz y 32 KB (datos) + 32 KB (instrucciones) de caché primaria y 2 MB de caché secundaria por núcleo. El sistema de archivos utilizado es ext3. La memoria alcanza los 3 GB con una frecuencia de 667 MHz. Como sistema operativo se instalaron las versiones de 32 y 64 bits de la distribución GNU/Linux Mandriva 2007.1, con versión del kernel 2.6.17-13. Las versiones de los compiladores utilizados fueron la 4.1.2 de GCC, la 10.0 Professional Edition de Intel, y el paquete Sun Studio 12, con versiones 5.9 de los compiladores de C y C++ y 8.3 de Fortran.

El segundo de los sistemas (**SUT2**) monta dos cpu Dual Core AMD Opteron®Processor 270, haciendo un total de 4 núcleos, a una velocidad de reloj de 1993 MHz, con un bus de 1066 MHz. La caché primaria

[1]Dpto. de Informática, Univ. de Valladolid, España. e-mail: saldeal@gmail.com, {diego,arturo}@infor.uva.es

TABLA I
Optimizaciones aplicadas

GCC	**Sin optimizaciones:** -O2 **Con optimizaciones:** -O3 -funroll-loops -fno-inline-functions ftree-vectorize
INTEL	**Sin optimizaciones:** -O2 -no-for-main (C y Fortran a la vez) **Con optimizaciones (32 bits):** -O3 -ipo -xT -axT -no-prec-div -funroll-all-loops -no-for-main (C y Fortran a la vez) **Con optimizaciones (64 bits):** -O3 -ipo -xW -axW -no-prec-div -funroll-all-loops -no-for-main (C y Fortran a la vez) -parallel (versión paralela)
SUN	**Sin optimizaciones:** -xO3 -library=stlport4 (Benchmarks C++ excepto 453.povray) **Con optimizaciones (32 bits):** -fast -xarch=sse3 -library=stlport4 (Benchmarks C++ excepto 453.povray) **Con optimizaciones (64 bits):** -fast -xarch=sse3 -m64 -library=stlport4 (Benchmarks C++ excepto 453.povray) -xautopar -xreduction (versión paralela)

es de 64 KB + 64 KB y la secundaria de 1 MB por núcleo. La memoria RAM es de 4 GB de capacidad. El sistema operativo es una versión de 64 bits de la distribución Gentoo 1.12.9, con la versión 2.6.19 del kernel. Las versiones de los compiladores instalados en este sistema son las mismas que en el anterior, a fin de poder comparar los resultados.

La forma de obtener los resultados es la proporcionada por SPEC CPU2006. La suite proporciona tres conjuntos de trabajo para cada uno de los benchmarks: *test*, *train* y *reference*, ordenados de menor a mayor carga de trabajo, de manera que sólo proporciona un informe válido si previamente ha ejecutado en tres ocasiones los conjuntos *test* y *train* para cada benchmark, y finalmente otras tres con el conjunto de referencia, que es el que proporciona realmente los valores finales. La obtención del valor definitivo para cada benchmark se obtiene a partir del valor medio de entre los tres tiempos de ejecución con el conjunto de referencia. Al final, el valor que ofrece el resultado de cada conjunto de benchmarks se obtiene como la media geométrica de los resultados normalizados de los benchmarks. Este valor normalizado para cada uno de ellos se obtiene a partir de su tiempo de ejecución en segundos y el tiempo de ejecución del mismo en la máquina de referencia de SPEC CPU2006, citada en la sección 2. En [8] puede verse un ejemplo de los resultados que se obtienen con el sistema de referencia. Esta máquina utiliza un procesador UltraSPARC II a 296 MHZ con dos núcleos, 16KB + 16KB de caché primaria y 2MB de caché secundaria, 2GB de memoria RAM y dos discos duros SCSI de 36GB, uno dedicado al sistema operativo y otro al código y conjunto de datos de SPEC2006. El sistema operativo es Solaris 10 3/05 y como compiladores, el conjunto Sun Studio 11.

IV. Rendimiento de CINT2006

En el conjunto de los benchmarks de números enteros se puede considerar como vencedor de la comparativa el conjunto de compiladores de Sun. Con excepción de la máquina con arquitectura de 32 bits, en el que Intel obtiene mejores resultados, Sun logra el mejor rendimiento en el resto de las comparativas. En esta máquina de 32 bits es, precisamente, donde se muestran las mayores diferencias entre los compiladores, ya que en el resto de comparativas la nota predominante es la igualdad, a nivel general, entre unos compiladores y otros, con pequeñas diferencias entre unos y otros.

En la Figura 1.a se resume el comportamiento para el SUT1 en configuración de 32 bits. Puede observarse que Intel es muy superior a los compiladores GCC, debido en parte al mecanismo de autovectorización de Intel, capaz de vectorizar hasta 5 bucles en el benchmark 462.libquantum, mientras que GCC sólo es capaz de hacerlo con uno. Por otro lado, Sun obtiene unos resultados cercanos a los conseguidos por Intel, el cual, individualmente, logra un rendimiento un escalón por encima de Sun en muchos de los benchmarks.

En la configuración de 64 bits de la misma máquina SUT1 (figura 1.b), es Sun quien obtiene los mejores resultados, aunque las diferencias entre compiladores son menores en líneas generales. En esta configuración merecen destacarse algunos resultados. Por ejemplo, en el caso del 456.hmmer, Sun se encuentra muy por encima de los otros dos, observándose un rendimiento muy pobre del compilador GCC. También es reseñable el bajo rendimiento del compilador C de Intel en el 462.libquantum, ya que en este mismo benchmark, en 32 bits, Intel obtiene el mejor resultado, como sí mantiene en el 464.h264ref.

Ya en la máquina SUT2, de 64 bits (figura 1.c), las diferencias numéricas son aún menores, así como el rendimiento general de la máquina, tal como se deduce de la comparación con la figura 1.b. Aún así, los compiladores de Sun vuelven a obtener los mejores resultados. Al igual que ocurría en los 64 bits de la anterior máquina, Intel obtiene un rendimiento muy bajo en el mismo benchmark, 462.libquantum, lo que da a entender que Intel no es capaz de optimizar correctamente esta aplicación bajo 64 bits. La mayor diferencia se produce, otra vez, en el 456.hmmer, con un resultado por parte de Sun muy superior a los otros dos compiladores. Por último, es destacable el rendimiento del compilador GCC en el benchmark 464.h264ref, mejorando un 22 % respecto de Intel.

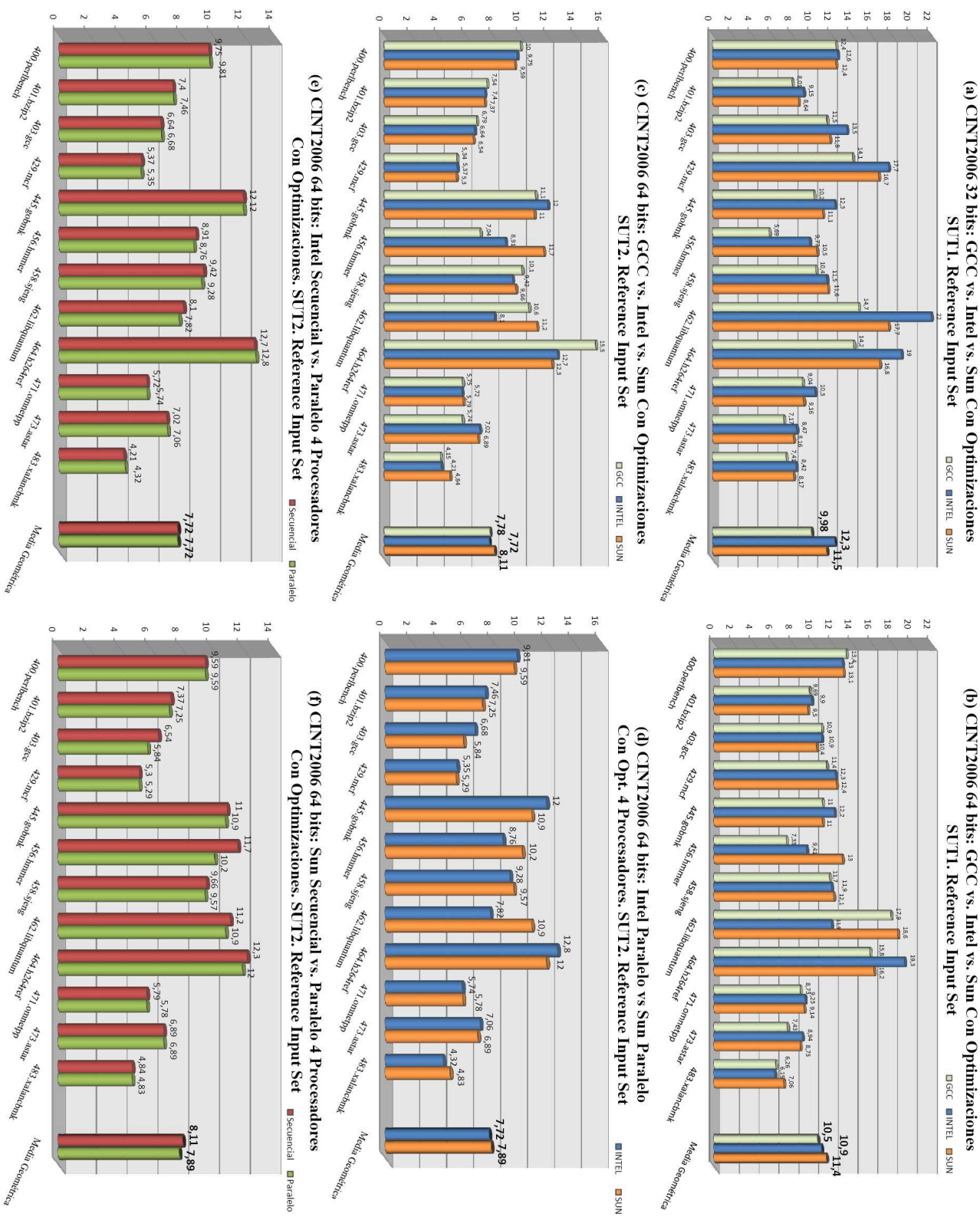


Fig. 1
Resultados del conjunto de benchmarks CINT2006

Los resultados con la opción de autoparalelización (figura 1.d) son muy similares y en algunos casos peores que los resultados secuenciales. Este hecho se debe a que los benchmarks de números enteros son difícilmente paralelizables, perdiéndose más tiempo creando y destruyendo hilos que realizando trabajo útil. Por tanto, al igual que ocurría en las ejecuciones secuenciales, las diferencias entre los compiladores de Intel y Sun son mínimas (GCC no tiene la opción de autoparalelización), repitiéndose, a nivel individual, los mismos patrones en las gráficas de 64 bits. Sun es superior en el 456.hmmer, mientras que Intel mantiene el bajo rendimiento en el 462.libquantum y obtiene un buen resultado en el 464.h264ref. En cuanto a la diferencia entre las versiones secuenciales y paralelas de los benchmarks (figuras 1.e y 1.f), como ya se ha comentado, son mínimas con la excepción de 456.hmmer bajo compilador Sun.

V. Rendimiento de CFP2006

A diferencia de lo que ocurre en la comparativa de los benchmarks de números enteros, en los que los resultados obtenidos por los diferentes compiladores son bastante similares, en las comparativas de los benchmarks de punto flotante las diferencias son mucho más grandes. No sólo estas diferencias son mayores, sino que el vencedor en todos los casos, y con una clara diferencia en muchos de ellos, es el compilador de Intel. Este compilador obtiene el mejor rendimiento con y sin optimizaciones, en 32 bits y 64 bits, y en secuencial y en paralelo, tanto en términos generales como en la mayor parte de los benchmarks individuales.

En lo que respecta al sistema SUT1 de 32 bits (figura 2.a), Intel obtiene los mejores resultados en 16 de los 17 benchmarks, estando en alguno de ellos muy por encima del resto, como es el caso de 410.bwaves, 435.gromacs, 436.cactusADM o 481.wrf. Curiosamente, es precisamente en estos benchmarks donde los compiladores de GCC obtienen los peores resultados, no sólo en comparación con Intel, sino también con Sun, ya que aunque los resultados de este último se encuentran casi siempre por encima de GCC, no es tanta la diferencia como en esos casos. Los compiladores de Intel presentan una mejora del rendimiento de un 26.5 % respecto a los de Sun y un 59.0 % respecto a GCC.

Los resultados en los 64 bits de la máquina SUT1 (figura 2.b) son similares al caso anterior. Las diferencias entre compiladores siguen siendo bastante grandes, con unos resultados superiores de Intel (mejoras de rendimiento del 18 % respecto a Sun y de casi el 40 % respecto a GCC), Sun en un punto intermedio, y GCC como el peor de los tres compiladores. Al igual que en 32 bits, Intel obtiene el mejor rendimiento en los mismos 16 benchmarks, repitiendo sus mayores diferencias en 410.bwaves, 434.zeusmp, 435.gromacs, 436.cactusADM y 481.wrf.

A diferencia de lo que ocurría con el conjunto de benchmarks de números enteros, en los que, sobre la misma máquina, se obtenían peores resultados en 64 bits que en 32 bits, aquí ocurre todo lo contrario. Como puede verse en la figura 2.b, todos los compiladores mejoran ostensiblemente su rendimiento sobre un sistema operativo de 64 bits que sobre uno de 32 bits, logrando mejoras de hasta el 24.6 %, como es el caso de los compiladores GCC. La menor diferencia se da en los compiladores de Intel, con una mejora del 9.24 %.

En la máquina SUT2 (figura 2.c), estas diferencias entre los compiladores se reducen bastante, aunque Intel sigue ofreciendo un mejor rendimiento que los otros compiladores. A nivel individual, sin embargo, la situación cambia ligeramente. Intel obtiene el mejor resultado en siete de los 17 benchmarks, de manera que su dominio ya no es tan claro. Aunque sigue obteniendo unos resultados por encima al resto en benchmarks como el 436.cactusADM, el 453.povray o el 481.wrf, en el resto de los benchmarks no obtiene tanta diferencia de rendimiento, quedando incluso por debajo de los compiladores de Sun en seis de los benchmarks, aunque Sun no obtiene un resultado que destaque por encima en ninguno de esos benchmarks. Y como viene siendo la tónica en todas las comparativas, salvo raras excepciones, los compiladores de GCC vuelven a quedarse por detrás de los compiladores de Intel y Sun, si bien las diferencias son menores que en anteriores: 21.3 % respecto a Intel y 16.3 % respecto a Sun. Y como ya ocurría antes, GCC mantiene su bajo rendimiento en benchmarks como 410.bwaves, 436.cactusADM o 454.povray.

En cuanto a la opción de autoparalelización de los compiladores de Intel y Sun (figura 2.d), es en el conjunto de los benchmarks en punto flotante donde realmente se notan las diferencias entre los mecanismos secuenciales y de paralelización automática. Si se considera únicamente el rendimiento global de todos los benchmarks se observa una gran similitud entre ambos conjuntos de compiladores, con una diferencia muy pequeña a favor de Intel. Sin embargo, analizando uno a uno los resultados de los benchmarks se pone de manifiesto que esta igualdad es sólo aparente, produciéndose grandes diferencias en aplicaciones individuales. Tal es el caso de los benchmarks 410.bwaves, 437.leslie3d o 459.GemsFTDT, donde se producen las mayores diferencias a favor de los compiladores de Sun (de hasta el 81.5 % en el caso del 437.leslie3d), o de los benchmarks 436.cactusADM, 453.povray o 481.wrf, donde son los compiladores de Intel los que obtiene el mejor rendimiento (hasta un 41.77 % mejor en el caso del 481.wrf).

Las diferencias de rendimiento entre las versiones secuenciales y paralelas de un mismo compilador (figuras 2.e y 2.f) sí que resultan, por tanto, significativas. Intel obtiene un beneficio del 17.5 %, mientras que Sun logra mejorar sus resultados algo más, un 21.5 %. En ambos conjuntos de compiladores se observa que los tres benchmarks que obtienen un mayor incremento de rendimiento al ser compilados con la opción de autoparalelización son el 410.bwaves (un 116 % en Intel y un 140 % en Sun), el 436.cactusADM (un 216 % en Intel y un 254 % en Sun) y el

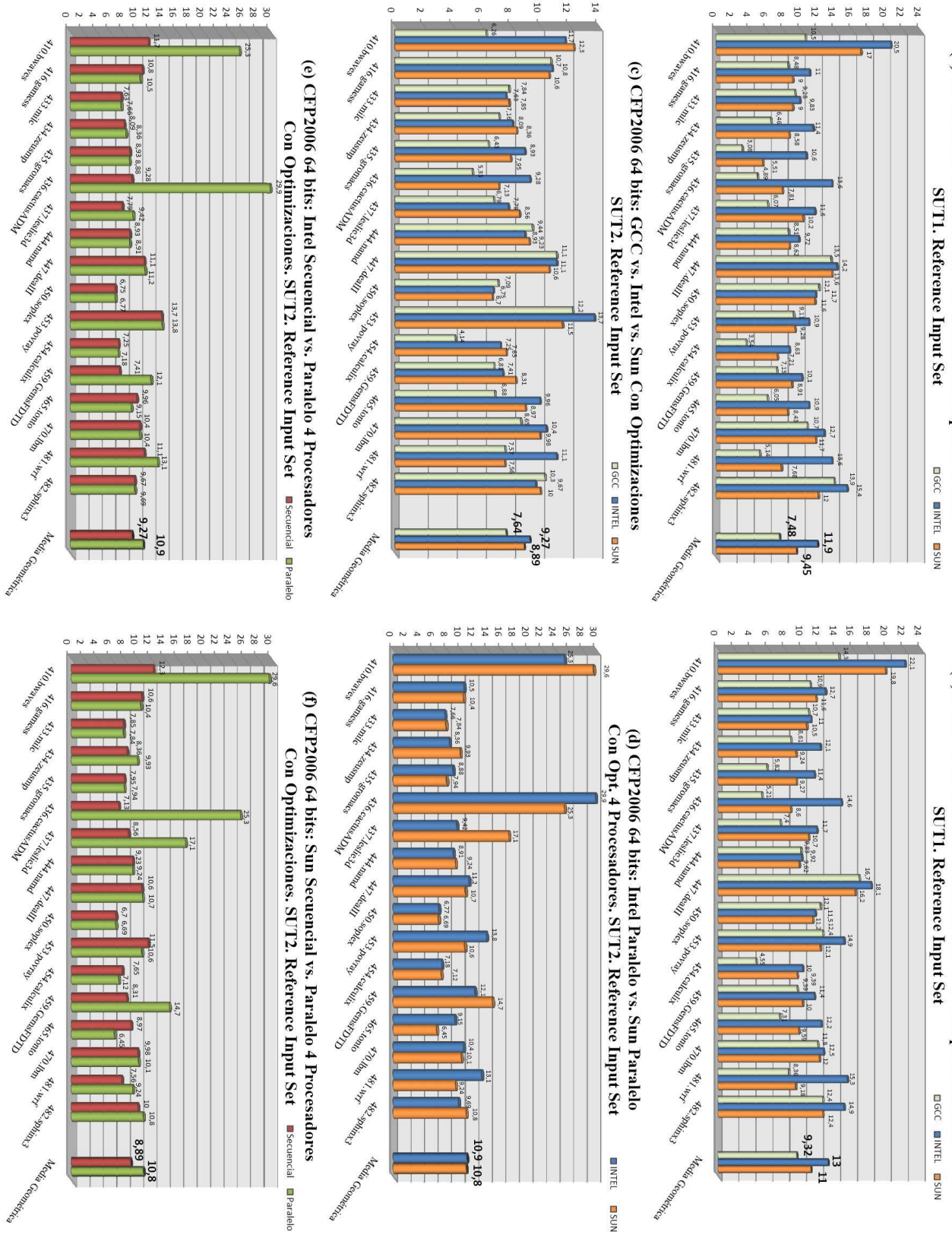


Fig. 2
Resultados del conjunto de benchmarks CFP2006

459.GemsFTDT (un 63 % en Intel y un 76.9 % en Sun). El resto de los benchmarks obtienen mejoras menores bajo los compiladores de Intel y de Sun. Algunos benchmarks incluso pierden rendimiento al ser paralelizados, siendo los principales perjudicados el 454.calculix y el 465.tonto en ambos casos. Sin embargo, los compiladores de Sun son capaces de paralelizar aún más alguno de los benchmarks restantes, como el caso el 437.leslie3d, donde el compilador Fortran de Sun es capaz de obtener una mejora de casi

el 100 % con cuatro procesadores.

VI. Conclusiones

Este trabajo busca ofrecer una visión más clara de las diferencias existentes entre los principales compiladores existentes en el mercado para arquitecturas Intel: GCC, Intel y Sun. El estudio tiene dos partes claras, correspondientes a los dos tipos de benchmarks, CINT2006 y CFP2006, no sólo por las diferencias entre ellos sino también por los resultados obtenidos.

Por un lado, en el conjunto de benchmarks de números enteros, el compilador que obtiene un mejor rendimiento de las aplicaciones es Sun Studio 12, obteniendo mejores resultados en la mayoría de los casos analizados, justo por delante de Intel y GCC, que queda rezagado al último lugar. Sin embargo, a pesar de encontrarse por encima, las diferencias entre estos compiladores y los de Intel o GCC son pequeñas como para resultar especialmente significativas, al igual que lo que ocurre al comparar las versiones secuenciales y paralelas de las distinstas aplicaciones. En el caso de CINT2006, como ha podido verse, no se obtiene beneficio al intentar paralelizar automáticamente su código.

En el conjunto de benchmarks de punto flotante, por el contrario, es donde se obtienen las diferencias más significativas de rendimiento entre las aplicaciones compiladas por un u otro compilador. Aquí, el que genera mejores resultados es Intel y su conjunto de compiladores, que obtienen rendimientos muy por encima en la mayoría de los análisis. Donde se iguala más la situación es al activar la opción de autoparalelización, quedando descartados los compiladores de GCC al no disponer de este mecanismo. La opción de autoparalelización permite obtener una mejora apreciable en la ejecución de los benchmarks, por lo que merece la pena utilizar esta opción si se dispone de la plataforma adecuada.

Además de los resultados generales, cabe también destacar el hecho de que las mejoras no son uniformes de una aplicación a otra, como puede verse en el bajo rendimiento del compilador Intel con el benchmark 462.libquantum de 64 bits, o el buen resultado de GCC con 464.h264ref sobre el SUT2. Por lo tanto, las conclusiones de este estudio no pueden generalizarse a todas las aplicaciones de un determinado lenguaje. Este hecho obliga al programador que desee elegir entre uno de estos compiladores a realizar mediciones sobre su propia aplicación y arquitectura.

Con todas estas puntualizaciones, es justo concluir que la mejor opción en la compilación del SPEC2006 son los compiladores de Intel, que mantiene un muy buen rendimiento en líneas generales. Por detrás se situaría Sun Studio 12, que es ligeramente mejor que Intel en CINT2006 aunque claramente peor en CFP2006, y con un rendimiento similar en las opciones de autoparalelización. Por último, es de destacar el mal resultado de los compiladores GCC, que no sólo se encuentran por detrás de los otros dos en todos los casos, sino que tampoco puede competir con ellos cuando se requiere paralelizar automáticamente el código al no disponer de dicha opción.

Agradecimientos

Este trabajo ha sido parcialmente financiado a través del Ministerio de Educación y Ciencia (TIN2007-62302), del Ministerio de Industria (FIT-350101-2006-46 y FIT-350101-2007-27) y de la Junta de Castilla y León.

Referencias

[1] Ken Polsson, "Chronology of workstation computers", http://www.islandnet.com/~kpolsson/workstat/work1987.htm.

[2] Standard Performance Evaluation Corporation, "SPEC releases CPU2006 benchmarks", http://www.spec.org/cpu2006/press/release.html.

[3] Intel Corporation, "Intel Xeon Processor performance summary", http://www.intel.com/performance/workstation/xeon/summary.htm.

[4] IBM Corporation, "IBM posts SPEC CPU2006 scores for next-generation System x scalable server", ftp://ftp.software.ibm.com/eserver/benchmarks/news/newsblurb_x3850M2_speccpu_090507.pdf.

[5] Sun Microsystems, "Sun Fire X4450 server performance", http://www.sun.com/servers/x64/x4450/benchmarks.jsp?display=1.

[6] Vijay Tatkar, "Build high performance applications on multicore systems using sun studio compilers and tools", http://developers.sun.com/events/techdays/presentations/locations-2008/india_techdays/solaris_track/td_hyd_sun_studio_tatkar.pdf.

[7] Fujitsu Siemens, "Performance and benchmarks: Performance reports for CELSIUS R640", http://www.fujitsu-siemens.com/products/deskbound/workstations/performance/celsius_r640.html.

[8] Robert Munafo, "Reference machine times for SPEC CPU2006", http://home.earthlink.net/~mrob/pub/benchmarks/spec.html#ref_06.