

Tecnología

Antonio González

Departamento de Arquitectura de Computadores,
Universitat Politècnica de Catalunya

<antonio@ac.upc.es>

Resumen: en este artículo se revisa la microarquitectura de los procesadores actuales. Seguidamente se presentan las principales expectativas en la evolución de la tecnología. A continuación se destacan las principales limitaciones de la arquitectura de los procesadores actuales para tecnologías futuras. Finalmente se comentan algunas de las alternativas que se están investigando para reducir o evitar el impacto de estas limitaciones.

1. Introducción

El primer computador electrónico de propósito general fue el ENIAC, desarrollado por J. Presper Eckert y John Mauchly en la Moore School de la Universidad de Pennsylvania, estuvo operativo durante la 2ª. Guerra Mundial. Su sucesor, el EDVAC incluyó una mejora significativa como es el concepto de programa almacenado, tal como describe el informe que en 1945 firmaba John von Neumann. Sin embargo, fue el EDSAC, diseñado por Maurice Wilkes en la Universidad de Cambridge en 1949, el primer computador operacional que lo implementó. Los computadores que hoy en día utilizamos siguen teniendo una estructura que a grandes rasgos es similar a la del EDVAC y EDSAC. Tal como muestra la **figura 1**, el computador consta de tres partes principales: la memoria, que almacena tanto el programa como los datos, el procesador que ejecuta las instrucciones de un programa y manipula los datos, y los dispositivos de entrada/salida que sirven para comunicarse con el exterior.

En este artículo nos centramos en una de estas tres partes: el procesador. En primer lugar describiremos como es la microarquitectura de los procesadores actuales. Seguidamente comentaremos las expectativas de evolución de la tecnología de fabricación de procesadores, enfatizando su impacto en la microarquitectura del procesador. A continuación comentaremos las principales limitaciones de las

Tendencias en la Microarquitectura de los Procesadores

actuales microarquitecturas para hacer un uso efectivo de las crecientes posibilidades que ofrece la tecnología. Finalmente describiremos algunas de las alternativas arquitectónicas que pueden tener un impacto mayor en un futuro venidero.

2. Procesadores actuales

La mayoría de los procesadores actuales son procesadores *segmentados* con una organización *superescalar*. Un procesador segmentado es aquel que divide la ejecución de una instrucción en diversas etapas, de forma que tan pronto como una instrucción finaliza una etapa n , a la vez que esta realizando la etapa $n+1$, la siguiente instrucción puede llevar a cabo la etapa n (este mismo concepto es el que rige el funcionamiento de una cadena de montaje). De esta manera se consigue ejecutar varias instrucciones a la vez, lo que se denomina *paralelismo a nivel de instrucción*.

Un procesador superescalar es aquel que es capaz de procesar más de una instrucción simultáneamente en cada una de las etapas (el símil sería tener varias cadenas de montaje). De esta manera pueden aumentar el paralelismo a nivel de instrucción. La mayoría de los procesadores superescalares actuales disponen de un mecanismo de *planificación dinámica de instrucciones*. Es decir, las instrucciones no se ejecutan en el orden en que aparecen en el programa sino que el hardware decide en que orden se ejecutan para un mayor rendimiento. Por eso, a estas máquinas también se les denomina procesadores con *ejecución fuera de orden*. La **figura 2** muestra una organización típica de un procesador superescalar. El procesador dispone de una memoria cache de instrucciones, donde en cada ciclo, la unidad de búsqueda va a traer varias instrucciones. A continuación las instrucciones son decodificadas y a la vez los operandos registro son renombrados. El renombramiento de registros tiene por

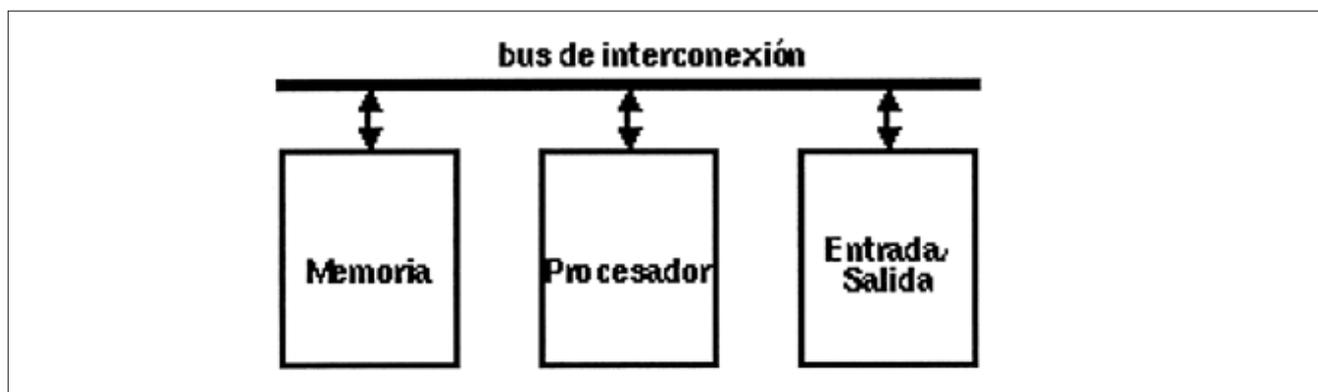


Figura 1: Estructura de un computador

objetivo que los operandos destino de todas las instrucciones cuya ejecución ha empezado pero no finalizado tengan un identificador diferente. De esta forma se eliminan las llamadas *dependencias de nombre*. Al reducir el número de dependencias, aumenta la cantidad de paralelismo que se puede explotar.

Tras la etapa de decodificación las instrucciones son *despachadas* (*dispatched*) a la *cola de instrucciones* y al *buffer de reordenación*. Las instrucciones permanecen en la cola de instrucciones hasta que pueden iniciar su ejecución, mientras que en el buffer de reordenación están hasta que el procesador está seguro de que su ejecución es correcta y no debe deshacerse. Los procesadores actuales realizan muchas actividades de forma *especulativa*, es decir, sin estar completamente seguros de si son correctas. Esto se hace para adelantar dichas actividades pero, en caso de fallo de especulación, éstas deben deshacerse.

La cola de instrucciones es inspeccionada en cada ciclo por un hardware que se encarga de determinar que instrucciones pueden iniciar la ejecución, o *emitirse*. Para ello, una instrucción debe de tener todos sus operandos, y los recursos que necesita deben de estar libres.

Tras emitirse, las instrucciones son ejecutadas y al finalizar escriben el resultado. Tras ello la instrucción es retenida en el procesador hasta que se está seguro de que su ejecución no debe deshacerse. Para ello, las instrucciones abandonan el procesador (se *gradúan*) en orden secuencial.

La **tabla 1** muestra algunas de las características principales de los procesadores actuales de uso más frecuente. La primera columna muestra el tamaño mínimo que es capaz de manejar la tecnología de fabricación, en nanómetros, lo que da una idea del factor de integración. La segunda columna corresponde a la frecuencia de reloj. La tercera columna muestra cuantas instrucciones puede llegar a ejecutar por ciclo. La cuarta columna indica si el procesador soporta planificación dinámica de instrucciones. Las últimas dos columnas son una medida de sus prestaciones. Los SPECint95 representa el rendimiento para programas no numéricos que utilizan básicamente aritmética de Enteros mientras que los SPECfp95 es para programas numéricos con abundante aritmética de Reales. La medida en cuestión muestra la velocidad relativa a una SparcStation 10/40 con un procesador SuperSparc a 40 MHz.

3. Evolución de la tecnología

La constante mejora de la tecnología ha sido sin duda uno de los principales motores de la evolución de los procesadores. Quizás el dato más significativo sea el constante aumento en la densidad de componentes de los chips. Esta tendencia se ha bautizado con el nombre de *ley de (Gordon) Moore*, que dice que el número de transistores por chip se duplica aproximadamente cada 18 meses. Esto ha hecho que de los 2.300 transistores que tenía el Intel 4004 en 1971 o los 29.000 del Intel 8086 en el 1978, se ha pasado en la actualidad a que la mayoría de los microprocesadores disponen de unas pocas decenas de millones de transistores, tal como muestra la **tabla**. Ello ha permitido incorporar técnicas cada vez más complejas así como una mayor cantidad de memoria dentro del mismo chip que el procesador (la mayoría de procesadores actuales disponen de un primer nivel de memoria *cache* de alrededor de 64 Kilobytes y otro tanto para datos, integradas en el mismo chip). Como consecuencia, el rendimiento de los procesadores ha ido aumentando desde mediados de los 80 a razón de un 50% anual.

La tecnología de fabricación de procesadores actual basada en transistores de campo MOS (*metal-oxide-semiconductor*) va a seguir vigente en las próximas décadas, a pesar de los esfuerzos y esperanzadores resultados que se están obteniendo en tecnologías alternativas como es la computación cuántica. La evolución de la tecnología MOS va a seguir un ritmo parecido al observado en el pasado, por lo menos durante los próximos 12 años. La **tabla 2** resume algunas de las previsiones por lo que se refiere a microprocesadores. Puede observarse que cada dos años el proceso de fabricación va disminuir el factor de integración, aunque a partir del 2003, el periodo entre generaciones va a extenderse a tres años. Esto se debe a que la técnica de litografía óptica que actualmente se utiliza para implantar el circuito en el cristal de silicio no va a ser apropiada por debajo de las 0.1 micras. Tecnologías basadas en rayos X y EBDW (*electron beam direct-write*) están siendo investigadas, entre otras, como prometedoras alternativas que pueden sustituir la actual. Esto va a suponer un cambio importante de infraestructura que va a ralentizar la introducción de nuevos procesos.

La frecuencia de los procesadores irá en aumento, pero sin lugar a dudas el aumento más espectacular será en número de transistores, que superará los mil millones en el 2012, lo que representa sobre 100 veces más que en la actualidad. La

Procesador (nm)	Tecnología (MHz)	Frecuencia orden	Anchura (millones)	Fuera de Transistors	SPECint95	SPECfp95	
Alpha 21264	250	700	4	si	15.2	35	55
AMD Athlon	250	750	3 (x86)	si	22	32	24
Intel PIII Xeon	180	733	3 (x86)	si	24	36	31
MIPS R12000	250	300	4	si	7.2	18	30
HP PA-8500	250	440	4	si	130	31	49
IBM Power3	250	222	4	si	15	12	28
PowerPC 7400 (G4)	220	450	3	si	10.5	21	20
Sun Ultra-2	290	450	4	no	3.8	16	24
Hal Sparc64-III	250	296	4	si	17.6	15	28

Table 1: Características de los procesadores actuales. Fuente: Microprocessor Report, December 27, 1999

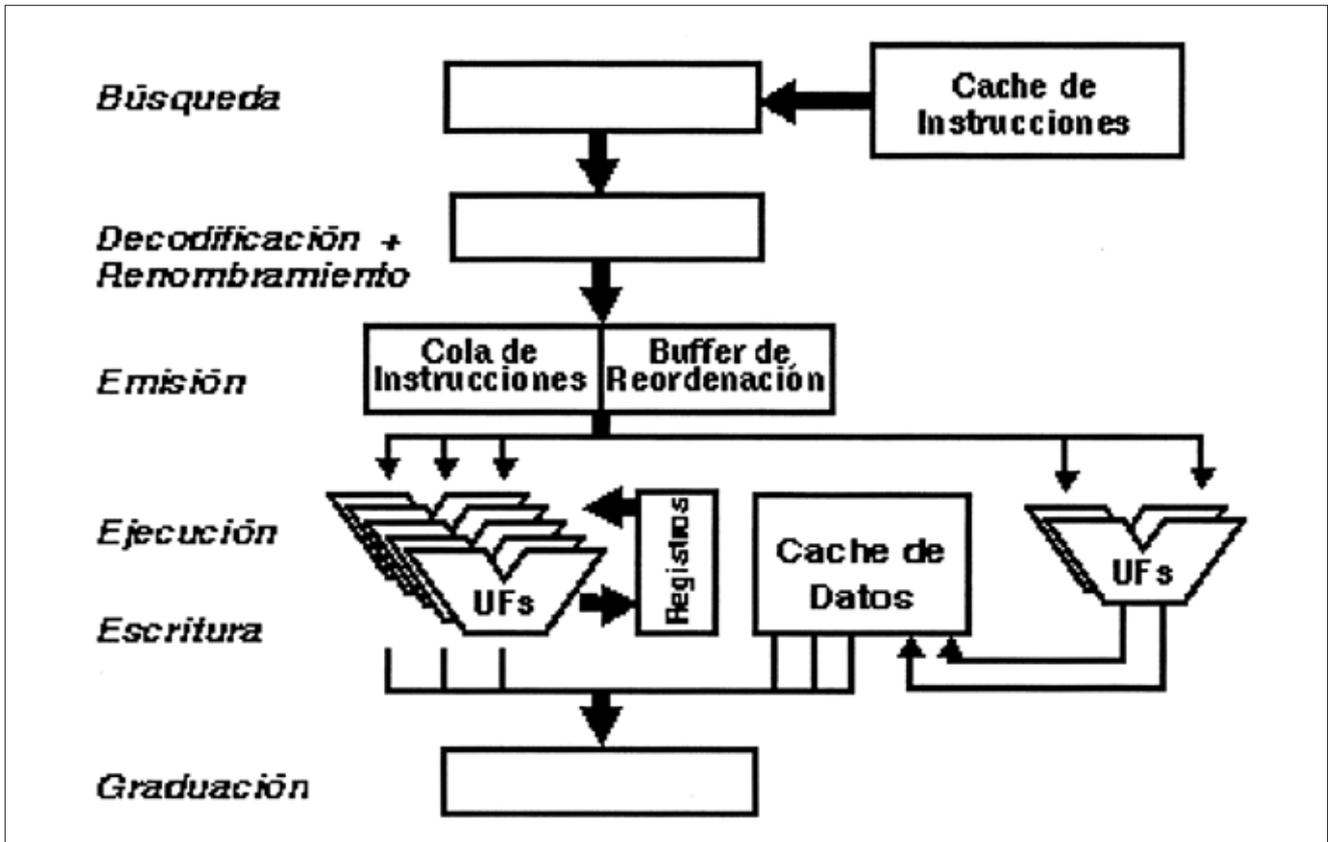


Figura 2: Procesador superescalares

tensión de alimentación se ira reduciendo, para así aliviar el aumento de consumo de energía que supone el mayor número y densidad de componentes. El consumo, y sobre todo la disipación de calor, son uno de los principales problemas que deberán afrontar los procesadores. Para ello, serán necesarias nuevas técnicas no sólo en el diseño de circuitos, sino también nuevas microarquitecturas que reduzcan el consumo de energía.

4. Limitaciones de los procesadores superescalares

La microarquitectura de los procesadores actuales empieza a tener problemas para hacer un uso efectivo del cada vez mayor número de transistores disponibles.

Las principales barreras a la escalabilidad de estas microarquitecturas son las siguientes:

- La memoria. El acceso a memoria es cada vez más lento en relación a la velocidad del procesador.
- Las dependencias. Las dependencias entre instrucciones restringen el grado de paralelismo que puede explotar el procesador. Por ejemplo, si una instrucción *j* utiliza un

dato que produce la instrucción *i*, la primera no puede empezar a ejecutarse hasta que acabe la segunda.

- Los retardos en las interconexiones. Al mejorar el factor de integración de la tecnología, disminuye el retardo de la lógica. Sin embargo, el retardo de la interconexiones entre diversos bloques de un mismo chip apenas se reduce, por lo que estos retardos tienen un peso cada vez mayor.
- Consumo/disipación de energía. Las progresivas reducciones de la tensión de alimentación no son suficientes para paliar este problema. Aportaciones desde otras áreas tales como la microarquitectura empiezan a ser vitales.
- Tiempo de verificación. La creciente complejidad de los procesadores hace que el tiempo necesario para validar los nuevos diseños sea cada vez mayor.

5. Organizaciones alternativas

Para reducir las limitaciones descritas en la anterior sección, diversas organizaciones se están investigando como alternativas a la organización superescalares.

Entre ellas podemos destacar la arquitectura EPIC (*Explicitly Parallel Instruction Computing*) definida por Intel y HP, y

	1999	2001	2003	2006	2009	2012
Tecnología (nm)	180	150	130	100	70	50
Frecuencia (MHz)	750	1200	1400	1600	2000	2500
Transistores (millones)	21	40	76	200	520	1400
Tensión de alimentación (V)	1.5-1.8	1.2-1.5	1.2-1.5	0.9-1.2	0.6-0.9	0.5-0.6

Tabla 2: Previsiones para los próximos 12 años. Fuente: The National Technology Roadmap for Semiconductors

que será adoptada por los futuros procesadores de Intel en su nuevo lenguaje máquina IA-64. La arquitectura EPIC es una evolución de las arquitecturas VLIW (*Very Long Instruction Word*). La principal característica de estas arquitecturas es que el lenguaje máquina incorpora mecanismos para que el compilador pase información al hardware sobre determinadas características del programa. Por ejemplo, el compilador puede identificar que instrucciones son independientes. De esta manera el hardware puede ser mucho más sencillo, lo que mejora sustancialmente los problemas de tiempo de verificación y consumo de energía. Otra organización que será realidad en pocos años es la denominada *multithreading*. Algunos fabricantes como Compaq ya han anunciado que alguno de sus futuros microprocesadores (Alpha 21464) tendrá esta organización. Un procesador *multithreading* es aquel que ofrece soporte hardware para mantener el estado de múltiples *threads* (o tareas) y pueda ejecutar en paralelo instrucciones de todos ellos. De esta manera se combina la explotación de paralelismo a nivel de instrucción con la de un paralelismo de grano más grueso para así reducir los problemas causados por las dependencias entre instrucciones.

Las microarquitecturas *cluster* tendrán sin duda una presencia cada vez mayor. La idea de una microarquitectura *cluster* es dividir el procesador en diversos *clusters*, de forma que cada cluster este formado por las unidades que más frecuentemente se comunican entre ellas, y las comunicaciones entre *clusters* sean infrecuentes. Así se puede conseguir que el impacto del retardo de las interconexiones se reduzca sustancialmente, ya que las interconexiones que se usan mas frecuentemente son las más cortas y por lo tanto la más rápidas. En otras palabras, podríamos decir que las microarquitecturas del futuro deben de explotar la localidad en las comunicaciones. Además, un diseño *cluster* es más sencillo y conlleva ventajas importantes en cuanto a consumo de energía.

Para el problema de la memoria una posible solución es integrar toda la memoria (o gran parte de ella) y el procesador en un solo chip. Esta organización se conoce como IRAM (*Intelligent RAM*). De esta manera el tiempo de acceso es mucho menor y, además, en cada operación se puede acceder a cantidades grandes de información. Para aumentar la cantidad de memoria que puede integrarse junto con el procesador se propone utilizar memoria RAM dinámica (DRAM) en lugar de estática (SRAM). Hasta la fecha, la tecnología para fabricar DRAM es diferente a la utilizada para fabricar SRAM y circuitería. Sin embargo, recientemente se han desarrollado diversos procesos tecnológicos capaces de integrar ambos componentes de una manera bastante eficiente.

Finalmente, existen diversos trabajos de investigación que proponen microarquitecturas que hacen un uso extensivo de técnicas de especulación. En la actualidad casi todos los procesadores especulan en las instrucciones de salto. Es decir, cuando encuentran un salto, en lugar de esperar a que se conozca su resultado, lo predicen y ejecutan especulativamente las instrucciones de la rama predicha. Si luego resulta que la predicción era errónea, el procesador es capaz de deshacer la ejecución de las instrucciones especu-

ladas. Sin embargo, las dependencias de datos (i.e. una instrucción produce un dato que consume otra) siguen limitando el paralelismo. Algunas propuestas como la arquitectura *multithreading especulativa* proponen no sólo especular en el control de flujo (predicción de saltos), sino también en los datos. La idea es que hay muchos datos cuyo valor es predecible y se han propuesto diversos esquemas hardware que pueden llevar a cabo dicha predicción. Si el procesador es capaz de predecir el valor de los operandos de una instrucción, no es preciso esperar a que estos estén disponibles para poder ejecutarla de forma especulativa y de esta manera se puede aumentar significativamente el paralelismo (instrucciones dependientes pueden ejecutarse en paralelo, e incluso puede ejecutarse la consumidora antes que la productora). Evidentemente, si la predicción resulta errónea, las instrucciones ejecutadas especulativamente deben deshacerse y re-ejecutarse de nuevo.

6. Conclusiones

Los avances tecnológicos van a seguir siendo un motor importante de la evolución de los procesadores en el futuro. La microarquitectura de los procesadores debe de adaptarse para hacer un uso efectivo de estos avances tecnológicos. La microarquitectura predominante en la actualidad, superescalar, está llegando a un punto donde su escalabilidad para ofrecer mayores rendimientos empieza a ser cada vez menor. Por ello, nuevas organizaciones microarquitectónicas son necesarias. Algunas de las alternativas que hemos descrito están en un proceso avanzado de desarrollo y tendrán una presencia importante en el mercado en un futuro inmediato, mientras que otras con un objetivo más ambicioso y a más largo plazo requieren todavía importantes esfuerzos de investigación. Para que las prestaciones de los procesadores mantengan el ritmo de crecimiento que han experimentado en el pasado son necesarias nuevas aportaciones de la microarquitectura en un futuro inmediato.

7. Referencias

- [1] R. Canal, J.M. Parcerisa and A. González, «Dynamic Cluster Assignment Mechanisms», *Proc. of 6th. Int. Symp. on High-Performance Computer Architecture (HPCA-6)*, Toulouse (France), pp. 133-142, Enero 2000.
- [2] J.A. Fisher, «Very Long Instruction Word Architectures and the ELI-512» *Proc. of 10th. Int. Symp. on Computer Architecture*, Junio 1983.
- [3] L. Gwennap, «Intel, HP Make EPIC Disclosure», *Microprocessor Report*, vol. 11, no. 14, Oct. 1997.
- [4] J.L. Hennessy and D. Patterson, *Computer Architecture. A Quantitative Approach*. Morgan Kaufmann (second edition), 1996.
- [5] C.E. Kozyrakis, «Scalable Processors in the Billion-Transistor Era: IRAM», *IEEE Computer*, Sept. 1997, pp. 75-78.
- [6] P. Marcuello and A. González, «Clustered Speculative Multithreaded Processors», *Proc. of the International Conference on Supercomputing*, Rhodes (Greece), pp. 365-372, Junio 1999.
- [7] Semiconductor Industry Association, *The National Technology Roadmap for Semiconductors*, 1997 Edition.
- [8] D.M. Tullsen, S.J. Eggers and H.M. Levy, «Simultaneous Multithreading: Maximizing On-Chip Parallelism», *Proc. of the Int. Symp. on Computer Architecture*, pp. 392-403, 1995.
- [9] J. Von Neumann, *First Draft of a Report on the EDVAC*, Moore School, University of Pennsylvania, 1945.