

Redes de Interconexión de Procesadores con Capacidad de Tolerancia a Fallos

S. Sánchez Solano, M. Valencia, J. L. Huertas, C. J. Jiménez

Dpto. de Diseño de Circuitos Analógicos,
Centro Nacional de Microelectrónica,
Edificio CICA, Avda. Reina Mercedes s/n, 41012-Sevilla

Tf: (95) 462 38 11

FAX: (95) 462 45 06

e-mail: santiago@cnm.us.es

1. Resumen

Se describe una familia de redes de interconexión para sistemas multiprocesadores. Estas redes utilizan buses múltiples estructurados en varios niveles e incorporan el concepto de tolerancia a fallos mediante redundancia en las vías de comunicación. Sus principales características son: buena relación coste/rendimiento, fiabilidad, modularidad y flexibilidad. La comunicación incluye asimismo una visión panorámica de los aspectos de diseño e implementación de una serie de prototipos de los elementos básicos de la red de interconexión.

2. Introducción

El concepto de "Computación Tolerante a Fallos" ha estado tradicionalmente asociado a aplicaciones críticas como el control aeroespacial y nuclear. Sin embargo, puesto que los sistemas digitales van participando más activamente en muchas actividades cotidianas, los conceptos de disponibilidad y fiabilidad de un sistema cobran cada vez mayor sentido, estimulando la utilización de técnicas de tolerancia a fallos [1]. Un ingrediente común a todas las técnicas de tolerancia a fallos es la redundancia. Redundancia significa aumentar la cantidad de información, tiempo o recursos necesarios para la operación normal del sistema. La replicación física de los componentes del sistema constituye una de las formas más usuales de proporcionar tolerancia a fallos [2].

Por otra parte, los avances en las tecnologías de fabricación de circuitos integrados han influenciado en gran medida el desarrollo de los sistemas digitales actuales, adoptándose, en numerosos casos, soluciones basadas en explotar el paralelismo subyacente en muchos problemas, mediante la utilización de procedimientos arquitecturales bien ajustados a las posibilidades y limitaciones de las tecnologías

VLSI. Dentro del amplio espectro reportado en la literatura [3], las arquitecturas MIMD (*Multiple Instruction, Multiple Data Stream*), especialmente las construidas con procesadores de coste relativamente bajo, ofrecen una solución rentable a la demanda constante de potencia de cálculo y velocidad. Una prueba de ello es que existen numerosos sistemas, tanto experimentales como comerciales, basados en arquitecturas masivamente paralelas compuestas por cientos (e incluso miles) de elementos de procesado.

La presencia de un número elevado de elementos idénticos y de bajo coste en un sistema multiprocesador, junto al hecho de que estos elementos puedan ser considerados como componentes básicos de diseño desde la perspectiva VLSI, permite utilizar este tipo de sistemas para cubrir aplicaciones que requieran una alta disponibilidad. En todo caso, una de las claves para explotar el paralelismo consiste en proveer al sistema de una adecuada red de interconexión que facilite el intercambio de información entre los distintos elementos (comunicación de datos entre procesadores y memoria en un sistema de memoria compartida, o entre elementos de procesado en un sistema de memoria distribuida) [4].

3. Redes de Interconexión Tolerantes a Fallos

Las redes de interconexión que proponemos intentan establecer un compromiso entre las soluciones de conexión total y de conexión mediante BUS en sistemas multiprocesador. Dichos sistemas se organizan en base a un conjunto de racimos de microprocesadores (que pueden actuar aisladamente o bajo configuraciones TMR o HMR) interconectados por buses múltiples estructurados en varios niveles. El principal objetivo consiste en tener disponibles caminos alternativos para la transferencia de datos, de forma que la velocidad de comunicación entre mó-

dulos se mantenga razonablemente alta, pueda aplicarse redundancia, y el sistema completo permita una adecuada degradación progresiva [5].

El nivel-0 de la arquitectura corresponde a un conjunto de N elementos de procesamiento fuertemente acoplados, organizados de forma que puedan trabajar bajo esquemas de redundancia masiva. El nivel-1 consiste en la interconexión de N racimos de procesadores siguiendo el esquema de la Figura 1. N buses conectan cada racimo con N módulos de comunicaciones (MC's) que realizan las funciones de control de tráfico y transferencia de paquetes de información entre los elementos del nivel-0. En general, el nivel n -ésimo de la arquitectura estará formado por N racimos de nivel $n-1$ y N MC's de nivel n organizados de manera similar al nivel-1.

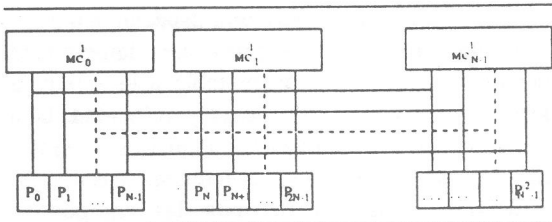


Figura 1: Nivel-1 de la red de interconexión.

Para minimizar la arbitración de los diferentes buses, cada racimo de módulos de comunicaciones actúa bajo un esquema de sincronización tal que en todo instante cada MC se encuentra en una fase de arbitración diferente, T_i ($i = 0, \dots, N-1$). En la fase T_i se establecen caminos de comunicación entre los buffers internos de los módulos de comunicaciones y los buses i -ésimos de comunicación con el nivel inferior. De esta forma cuando la red trabaja en modo "no redundante" pueden realizarse hasta N transferencias simultáneas en cada fase de arbitración. Este esquema de temporización es válido asimismo en el modo de operación "redundante" de la red de interconexión. En este último caso, tres (TMR) o cuatro (HMR) MC's transferirán simultáneamente, a través de sus correspondientes buses, las diferentes copias clónicas de un mensaje redundante.

Los racimos de procesadores que constituyen un sistema pueden estar asociados bajo esquemas de redundancia diferentes. Esto significa, por un lado, que en la red pueden coexistir paquetes con distintas configuraciones y, por otro, que pueden realizarse transferencias de mensajes en los que no coincida la configuración de los elementos fuente y destino. Por este motivo, los módulos de comunicaciones incluyen un "votador generalizado" capaz de realizar distintas operaciones de transferencia, votación y comparación de los paquetes que forman un mensaje.

4. Diseño e Implementación de Prototipos

Dada la complejidad del sistema, el desarrollo de los componentes de la red de interconexión debe abordarse de acuerdo con una metodología de diseño que facilite su realización. La metodología usada en nuestro caso contempla la partición del problema global de implementación en tres niveles de diseño (arquitectura, microarquitectura y circuitos e interconexiones) y propone la realización de tres fases de diseño (síntesis, verificación y análisis) en cada uno de los niveles anteriores.

En el desarrollo de la fase de síntesis a nivel de arquitectura aparecen una serie de opciones de diseño al considerar diferentes alternativas en cuanto al número de buffers que incluye cada MC, las características de estos buffers y los posibles esquemas de sincronización de los módulos de comunicaciones.

La utilización de lenguajes de descripción de hardware (VHDL) y simuladores lógicos (HILO) resulta imprescindible para garantizar la correspondencia entre las sucesivas representaciones del sistema en el nivel de diseño de la microarquitectura.

Por último, al efectuar el diseño a nivel de circuitos e interconexiones, se ha explorado la utilización de tres entornos de CAD diferentes. Estas herramientas se fundamentan, respectivamente, en los estilos de diseños basados en "gate-arrays", "gate-arrays más generadores de bloques parametrizables" y "standard-cells". El resultado de esta actividad, ha sido la implementación, en diferentes tecnologías, de tres circuitos integrados.

5. Referencias

- [1] A. D. Sing y S. Murugesan: "Guest Editors' Introduction: Fault-Tolerant Systems", IEEE Computer, Vol. 23, N.7, Julio 1990.
- [2] D. P. Siewiorek y R. S. Swarz: "The Theory and Practice of Reliable System Design", Digital Press, Bedford, Mass., 1982.
- [3] R. Duncan: "A Survey of Parallel Computer Architectures", IEEE Computer, Vol. 23, N.2, Febrero 1990.
- [4] T. Y. Feng: "A Survey of Interconnection Networks", IEEE Computer, Vol. 14, N.12, Diciembre 1981.
- [5] S. Sánchez Solano: "Diseño de Redes Flexibles de Interconexión de Procesadores", Tesis Doctoral, Univ. de Sevilla, 1990.