

Evaluación de una red en anillo para un sistema distribuido de cómputo sobre FPGAs

N. Cardona & F. Guzmán

Grupo de investigación en Microelectrónica, Universidad Pontificia Bolivariana, Medellín, Antioquia, Colombia

RESUMEN: El proyecto consiste en implementar un sistema distribuido coprocesado, capaz de gestionar memoria eficientemente y establecer la comunicación entre procesadores, con el objetivo de evaluar el impacto de una red en anillo que enlaza ocho supernodos de procesamiento, cada uno de ellos con 4 procesadores Microblaze y 2 PowerPC405, conformando en total 48 procesadores RISC de 32bits que integran un sistema con mayor rendimiento que un computador personal, para realizar operaciones matemáticas. Este trabajo se desarrolla con el fin de brindar a la comunidad científica resultados acerca de la importancia de las redes de interprocesadores en los sistemas modernos de cómputo.

1 INTRODUCCIÓN

Las comunicaciones entre chips constituyen un campo de investigación que pretende proporcionar al sistema la capacidad de mantener los altos rendimientos de procesamiento en sistemas distribuidos. En la medida que el número de nodos crece, se requiere un esquema para administrar la comunicación de manera efectiva. Para este fin, es necesario establecer arquitecturas de procesamiento y de red eficientes para lograr una elevada capacidad de cómputo.

En la actualidad están surgiendo arquitecturas computacionales con una alta capacidad de procesamiento de algoritmos complejos y de grandes cantidades de datos, que requieren sistemas distribuidos óptimos en tamaño, costo y flexibilidad. Los últimos avances en tecnología han dado lugar a las FPGAs, dispositivos reconfigurables con arquitecturas eficientes, que hoy en día son usadas para implementar arquitecturas paralelas con un gran número de nodos computacionales interconectados.

Diversas ramas investigativas buscan utilizar todo tipo de recursos existentes y ofrecer soluciones en diferentes niveles para abordar la creciente demanda de sistemas de alto rendimiento, así que trabajan el paralelismo en diferentes tipos: a nivel de procesador, de procesadores (memoria compartida y memoria distribuida), y de computadores (clusters).

Este proyecto evalúa la comunicación de alta velocidad para un sistema distribuido coprocesado de 48 procesadores RISC de 32 bits implementados en 8 tarjetas XUPV2P, que contienen una FPGA VirtexII Pro cada una. La novedad de este proyecto radica en que busca diseñar de forma eficiente un sistema completo de red que permita velocidades entre FPGAs de 1.5 y 3.125Gbps implementando una red en anillo y un protocolo de comunicación. Para esto es necesario diseñar los módulos de control y establecer una estructura de comunicación con el fin de interconectar los supernodos de procesamiento. Cada tarjeta de desarrollo cuenta con una memoria DDRSDRAM de 512MB y una memoria Platform Flash de Xilinx. La FPGA tiene embebidos dos procesadores PowerPC405, además cuenta con 8 MGT que serán empleados como los módulos de comunicación serial de alta velocidad para integrar la red.

2 DESCRIPCIÓN DEL PROYECTO

Este sistema pretende implementar una arquitectura de red en anillo con el fin de lograr un alto desempeño permitiendo determinar cual es la máxima velocidad pico a la cual se puede transferir información entre procesadores.

El anillo está conformado por ocho supernodos de procesamiento, que implementan un protocolo de comunicación mediante software basado en estándares de comunicación serial. La gestión de red y el acceso a la memoria principal lo controla un procesador PowerPC405 que se encuentra en cada una de las FPGAs, y su organización se muestra en la Figura 1.

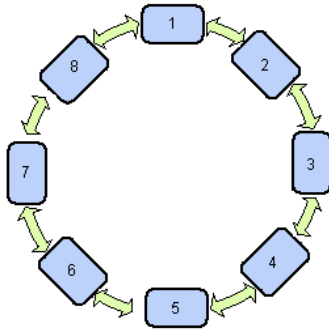


Figura 1. Organización de los 8 supernodos de procesamiento en anillo.

En cada uno de los supernodos de procesamiento esta configurado el siguiente sistema:

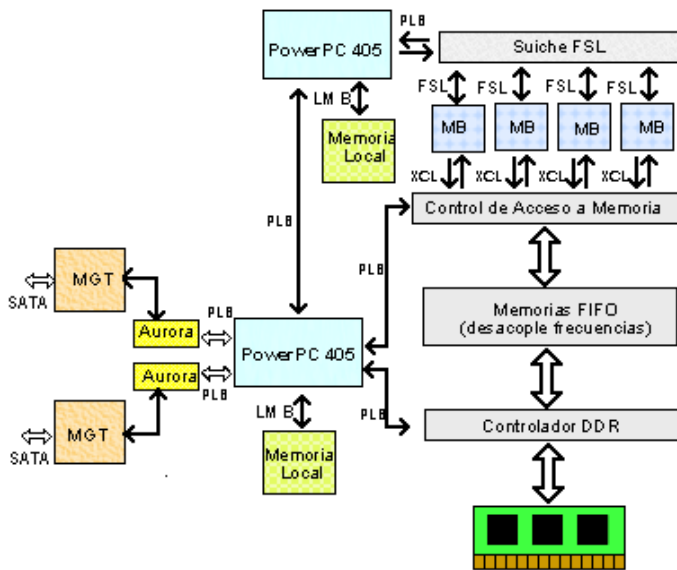


Figura 2. Arquitectura de un supernodo de procesamiento.

En este sistema se integran dos procesadores PowerPC405, cuatro procesadores Microblaze y dos módulos Aurora que son usados para controlar los componentes embebidos MGT (Multi Gigabit Transceivers) que operan a una velocidad de comunicación de 1.5Gbps y que están organizados como se muestra en la Figura 2. El sistema tiene una arquitectura de memoria compartida con un pico de rendimiento de 4.25Gbps y entre procesadores de 3.2Gbps.

Cada procesador Microblaze se conectan a un suiche (que es controlado por un procesador PowerPC405), mediante una interfaz Fast Simple Link (FSL). En el momento en que un procesador MB requiera compartir información con otro, el procesador PowerPC405 gestiona y controla la comunicación entre ellos.

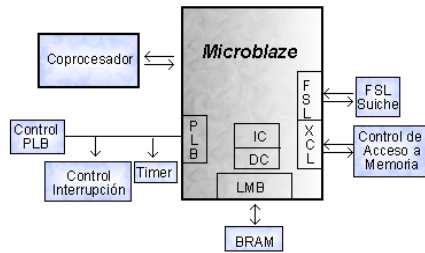


Figura 3. Detalle de la arquitectura para cada procesador Microblaze.

2.1 Manejo de memoria y comunicación

Se realiza una partición estática de memoria física disponible, de tal forma que los 4 procesadores Microblaze y el procesador PowerPC405 tengan acceso a memoria. En conjunto, cada procesador Microblaze tiene implementado una memoria caché de nivel uno para mejorar el desempeño. El bloque de control de acceso a la memoria se encarga de asignar un espacio de memoria para cada procesador, esta asignación es dinámica de acuerdo a las políticas controladas por el procesador PowerPC405.

Para la comunicación entre chips se usa el módulo Aurora, sobre el cual se implementa, por software y mediante el procesador PowerPC405 encargado de la comunicación y el acceso a memoria principal, un protocolo basado en el estándar IEEE 802.5. En la comunicación se usan enlaces seriales de alta velocidad que forman un anillo cuya conexión física se hace con cables SATA que interconectan los módulos MGT de cada tarjeta.

3 METODOLOGÍA

3.1 Análisis de requerimientos

Para el inicio del proyecto se hizo una evaluación para elegir la CPU, las interfases de entrada y salida, la topología de red, la plataforma de FPGA, la organización de memoria, el sistema operativo y el conjunto de herramientas de desarrollo, de tal forma que permitieran la implementación de las arquitecturas computacionales y comunicar los nodos a velocidades superiores a 1.5Gbps. En esta fase se definen las especificaciones del prototipo a implementar y se realiza el dimensionamiento del sistema distribuido y de la red.

3.2 Diseño de los módulos

En esta fase se establece la cantidad de subsistemas de cada módulo y se desarrolla una representación funcional mediante lenguajes de descripción de hardware para proceder a la verificación de los módulos como el de control de acceso a memoria y el suiche de interconexión FSL.

3.3 Implementación e Integración

Al estimar las características del dispositivo reconfigurable FPGA, se hace una revisión del diseño para integrar el sistema completo mediante la herramienta de desarrollo de sistemas embebidos de Xilinx EDK.

3.4 Evaluación del sistema

Se evalúan parámetros de importancia como los rendimientos y velocidades en la comunicación mediante un software de prueba.

3.5 Estrategia en caso de fallos

En el evento que un segmento del anillo falle, el protocolo mediante unos bits de estado notifica la caída del respectivo segmento y ejecuta una rutina de verificación para el estado de enlace produciendo que el sistema entre en un modo de diagnóstico que detiene la comunicación hasta que las causas del daño sean identificadas y solucionadas.

4 RESULTADOS

Con el fin de evaluar el sistema, se realizaron pruebas en donde se definen una serie de variables para determinar el rendimiento.

A continuación se muestra un análisis del número de recursos ocupados en la FPGA de Xilinx XC2VP30FF896 para dos implementaciones del procesador Microblaze, que varían de acuerdo al número de estados de segmentación (pipeline), una a 3 estados y otra a 5 estados de segmentación:

Tabla 1. Recursos y rendimiento del procesador Microblaze.

	Frecuencia	Número de LUTs	Dhrystone MIPS
MB 5 estados	100MHz	1912	115 DMIPS
MB 3 estados	100Mhz	1307	92 DMIPS

Los resultados obtenidos para la integración de todo el sistema (teniendo en cuenta los 4 procesadores Microblaze y toda la lógica implementada como el controlador de acceso a memoria, el control de las memorias FIFO, el controlador de memoria DDRSDRAM, el suiche de interconexión FSL, el sistema de buses y los periféricos de cada procesador Microblaze) se muestran en la Tabla 2.

Tabla 2. Porcentaje de utilización de recursos en general.

	Numero de Slices	numero de LUTs	Numero de FFs	Número de BlockRAMs
Implementación 1*	13559 (99%)	29275 (95%)	19105 (62%)	134 (99%)
Implementación 2**	9724 (71%)	19722 (64%)	15099 (49%)	164 (99%)

* Sistema Completo con 4 Microblaze implementados con 5 estados de segmentación.

** Sistema Completo con 4 Microblaze implementados con 3 estados de segmentación.

A continuación se muestra el rendimiento del sistema completo: 32 procesadores Microblaze + 16 Procesadores PowerPC405 en dos configuraciones de red en anillo.

Anillo1: Configurado a una frecuencia de 75MHz manejado por Aurora para obtener un rendimiento pico de 1.5Gbps.

Anillo2: Configurado a una frecuencia de 156.250MHz manejado por Aurora para obtener un rendimiento pico de 3.125Gbps.

La columna de rendimiento de la Tabla 3, se evaluó por medio de un software de aplicación en el sistema distribuido coprocesado, que hace una simulación de una ecuación electromagnética utilizando el método numérico de diferencias finitas. También con ayuda de un software monitor y la herramienta *profile* del depurador de microblaze, se obtuvo dicha estimación de rendimiento.

Tabla 3. Rendimiento del sistema para dos frecuencias configuradas en el anillo

	Frecuencia	Velocidad enlace	Velocidad efectiva	Rendimiento
Anillo 1	75Mhz	1.5Gbps	1.425Gbps	400MFlops
Anillo 2	156.250Mhz	3.125Gbps	2.8125Gbps	790MFlops

5 CONCLUSIONES

Los dispositivos reconfigurables permiten implementar de forma flexible diferentes diseños, posibilitando, de esta manera, cambios en la arquitectura que permitan evaluar los impactos en un sistema distribuido coprocesado.

La metodología de depuración del sistema entero debe ser segmentada, analizando cada su-pernodo de procesamiento y utilizando las herramientas de medida de rendimiento diseñadas para Microblaze, de esta forma se obtiene una estimación general de la capacidad de procesamiento de todo el sistema.

En el futuro se pretende crear un esquema que permita utilizar todo el sistema mediante herramientas que, autónomamente, repartan las tareas de forma distribuida y permitan que el sistema sea más dinámico. Actualmente la distribución de tareas se hace por una planeación estática regida a la aplicación final.

REFERENCIAS

- Chang, C., Wawrzynek, J. & Brodersen, R. 2005. *BEE2: A high-end reconfigurable computing system. IEEE Design and Test of Computers*, 22(2):114–125.
- Williams, J. “*MicroBlaze uClinux Project Home Page*,” <http://www.itee.uq.edu.au/~jwilliams/mblaze-uclinux/>
- Marathe, V. J., Scherer III, W. N. & Scott, M. L. 2005. *Adaptive Software Transactional Memory. In 19th Intl. Symp. On Distributed Computing.*
- Mc.CABE, James. *Network Analysis, Architecture and Design*. Second Edition. Ed. Morgan Kaufmann publishers. 2003. 536p
- Xilinx. 2007. *LogiCORE Aurora v2.8, Reference Manual*: 101p.
- XUPV2P. 2005. *Reference Manual, Xilinx University Program Virtex-II Pro Development System*: 138p.
- Xilinx. 2007. *User Manual RocketIO™ Transceiver*: 156p.
- Xilinx. 2007. *Virtex-II Pro and Virtex-II Pro X FPGA*: 559p.
- Xilinx. 2007. *Microblaze Processor Reference Guide*: 194p.
- Xilinx. 2007. *PowerPC 405 Processor Reference Guide*: 597p.