

# Control de lectura - Hardware eficiente apoyado en mecanismos de sincronización para muchos núcleos

Pedro J. Vargas B.

Maestría en Ciencias de la Información y las Comunicaciones

Ingeniero de sistemas

Universidad Distrital

Bogotá

Email: pjvargasb@correo.udistrital.edu.co

**Resumen**—El impacto de las necesidades de cómputo ha llevado a que se avance en el desarrollo de la nueva generación de procesadores, los manycore, que se caracterizan porque tienen una cantidad de núcleos amplia, además de que entregan un alto grado de paralelismo, debido a la gran cantidad de núcleos que poseen. En el capítulo se expresan las características de sincronización, de trabajo y las ventajas que proporciona este tipo de procesadores para los nuevos enfoques tecnológicos.

**Keywords**—*Hardware, Sincronización, núcleos.*

## I. INTRODUCCIÓN

Los centros de datos están evolucionando al acoger aplicaciones distribuidas y en paralelo tales como la computación en nube, la transmisión de vídeo o las redes sociales. Estas nuevas aplicaciones exigen no sólo más capacidad de almacenamiento o ancho de banda de red, sino también mayor rendimiento, por lo que requiere procesadores manycore como bloques de construcción de cada nodo o servidor computacional.

Como se están integrando cada vez más núcleos en el chip, arquitecturas manycore han surgido como la próxima generación de multicores. Los Manycores son sistemas especialmente adaptados a la explotación de rendimiento masivo mediante la incorporación de muchas unidades de computación simples y de baja frecuencia. Este cambio de paradigma conduce a las cargas de trabajo en paralelo con cada vez un mayor número de hilos que necesitan comunicarse y sincronizarse entre ellos, por lo general dependen de un solo dominio de memoria compartida por servidor.

En este contexto las implementaciones convencionales de las operaciones de sincronización, tales como barrier (barrera) y locks (bloqueos), hacen uso de las variables compartidas que se actualizan de forma atómica. En particular, al considerar barreras globales y locks contendió altos (es decir, una cantidad significativa de hilos que solicitan el lock al mismo tiempo), sin el apoyo adecuado de hardware, las implementaciones típicas basadas en software no pueden proporcionar buena escalabilidad como el número de núcleos aumenta.

## II. LAS LINEAS DE TECNOLOGIAS G

Las líneas G ya se han integrado con éxito en un sustrato de silicio con el fin de permitir las comunicaciones de velocidad de la luz punto a punto. Chang et al. y José et al. mostraron principios de circuitos de punto a punto que permiten la transmisión de línea, velocidad de onda similar por 10 mm de interconexión. No obstante, esta implementación inicial sufre

los gastos generales significativos en términos de disipación de energía. Un gran esfuerzo se ha dedicado a superar esta y otras limitaciones. Por ejemplo, se ha extendido las G-Lines para apoyar la difusión, multi-drop y transmisiones bidireccionales. Esta contribución permite tanto la baja latencia y capacidad multi-drop en una línea de transmisión con disipación de baja potencia. Sin embargo, sus resultados siguen mostrando varias cuestiones de integración.

## III. HARDWARE DE SINCRONIZACION BARRIER

Un Barrier o Barrera es una primitiva de sincronización que permite que varios procesos o hilos puedan esperar en un punto de ejecución especial, hasta que todos ellos han llegado a este antes de que cualquiera de ellos pueda continuar. Un ejemplo típico de su uso es la utilización de barreras para separar las diferentes fases que se encuentran comúnmente en aplicaciones paralelas. Al hacerlo, el programador se asegura de que la segunda fase no se inicia hasta que todos los procesos o hilos de la aplicación han completado la primera. En el contexto de los sistemas que implementan un modelo de programación de memoria compartida, con el advenimiento de las arquitecturas manycore, nuevos cambios están surgiendo para proporcionar una implementación Barrier eficiente. Esto se debe principalmente al hecho de que de manera diferente al multiprocesador, aplicaciones clásicas que tienen como objetivo el paralelismo coarse-grained, las aplicaciones Manycore tienden a explotar el paralelismo de fine-grained, y por lo tanto, pueden ser muy sensibles a rendimiento del Barrier.

## IV. EL MECANISMO DE SINCRONIZACION GBARRIER

Se presenta una propuesta para construir una infraestructura de hardware eficiente para la sincronización de Barrier en el contexto de los servidores manycore. Para ello, se comienza por la descripción de la arquitectura de la red en el chip dedicado. Por simplicidad, se da la explicación asumiendo la tecnología G-Lines con la técnica S-CSMA. Como caso de estudio, se elige un servidor con una interconexión de datos 2D de malla red con hileras R de C núcleos cada uno (para un total de  $N = R * \text{núcleos } C$ ).

### IV-A. Arquitectura de red dedicada On-Chip

El mecanismo GBarrier se basa en una red en el chip dedicado. Para simplificar, se concentran en una versión de la propuesta de red que proporciona apoyo a una de las

barreras. la infraestructura GBarrier se compone de dos tipos de componentes. G- Lines ( horizontal y vertical), que se utilizan para transmitir las señales requeridas por el protocolo de sincronización; y controladores ( M y S ) , que en realidad implementan el protocolo de sincronización.

Cada G-Line es un cable que permite la transmisión de un bit de información a través de una de las dimensiones del chip en un solo ciclo de reloj. la propuesta de red basada en G-Line emplea dos G- Líneas por barrera (Barrier) para cada fila y dos más para la primera columna. De esta manera, para cualquier diseño 2D de malla con hileras R y las columnas C , el número total de G- Líneas por la barrera que serían necesarios es igual a  $2 * ( R + 1 )$  (por ejemplo, 10 G -Lines para el servidor de 16 núcleos).

#### *IV-B. protocolo de sincronizacion*

El protocolo de sincronización implementado en la parte superior de la red basada en G-Line previamente descrito se basa en el intercambio de mensajes de 1 bit (señales) entre los controladores maestro y esclavo , y el uso de la técnica de S-CSMA en los controladores maestro para contar el número de señales transmitidas a través de cada G - Line. En la propuesta, cada sincronización barrera se lleva a cabo mediante el uso de un protocolo de dos fases : la fase de cuenta y la fase de liberación. La primera fase comienza cuando el primer hilo llega a la barrera y termina cuando el último alcanza la barrera. A continuación, la segunda fase , en la que todas las discusiones que participan en el Barrier, se les ordena reanudar ejecución , se inicia una interacción exacta entre los hilos , G -Lines y controladores.

#### *IV-C. Problemas de programabilidad*

El mecanismo GBarrier propuesto esta destinado a ser utilizado por los programadores en un medio transparente. Por esta razon, se propone proporcionar un metodo de barrera de nivel de biblioteca especial GL Barrier que encapsula la funcionalidad de GBarrier y que podrian utilizarse en aplicaciones paralelas para hacer frente a operaciones de barrera. Este metodo de barrera utiliza un registro de 1 bit especial, llamado bar reg, que notifica la llegada a la barrera estableciendo su valor a uno. El registro bar reg necesita tantos bits como el número de GBarriers proporcionado por el hardware (un bit por barrera). De esta manera, varias operaciones de barrera de diferentes conjuntos de núcleos (los hilos en cada juego se ejecuta una aplicación) podrían tener lugar simultáneamente. De este modo, el archivo de registro de cada núcleo debe estar aumentada con el registro bar reg y la interacción entre los controladores y éstos registros deben estar habilitados, el cambio en los controladores siempre que los registros bar reg están escritos, y la reposición de los registros y la desconexión de los controladores una vez todos los controladores han terminado la sincronización.

### **V. IMPLEMENTACION DE TECNOLOGIAS**

Había varias razones que llevaron a utilizar la tecnología G -Lines para desarrollar mecanismos de sincronización de barreras en los servidores manycore . En primer lugar, la conectividad patrón utilizado para desplegar la red de GBarrier dedicado es basado enlaces unidimensionales de 1 bit que se integran perfectamente en el concepto de G -Lines . En

segundo lugar, los resultados prometedores que podrían lograrse utilizando esta tecnología en términos de sobrecarga área marginal y la disipación de energía mínimo . Esto llevaria a que se se obtendrían con ello aún más bajas implicaciones para la zona On-Chip. esta area marginal de cabecera tendrá también un impacto insignificante sobre la disipación de energía.

### **VI. CONCLUSIONES**

Se analizo GBarrier para entender mejor su impacto potencial sobre la actuación en procesadores manycore. En primer lugar , se discutieron algunas consideraciones tomadas al usar ambos tecnologicas lines-G estándar para implementar una propuesta que aproveche mejor el poder de computo. Se presentaron implementaciones, mostrar sus posibles contribuciones a las prestaciones en términos de algunas estadísticas primas importantes como el área dedicada disipación de potencia en el chip , máxima velocidad de funcionamiento y latencias mínimas para completar una barrera operación.