

Intel Tera-Scale Computing

Alumno: Roberto Rodriguez Alcala

1. Introducción

Los procesadores con dos núcleos existen actualmente, y los procesadores de cuatro están insertándose en el mercado lentamente, pero el número de núcleos en un chip continuara creciendo, lanzando una nueva era de computadores tremendamente más potentes. Estas maquinas serán capaces de proveer desempeño mayor o igual a un teraflop (Tera instrucciones de punto flotante por segundo) por chip, con la capacidad de soportar las aplicaciones del futuro.

La tecnología Intel Tera-Scale es el programa de investigación de los futuros procesadores de la empresa, y su misión es lidiar frente a los desafíos de hardware y software utilizando docenas de núcleos, con una nueva jerarquía de memoria y un consumo energético eficiente.

Porque tantos núcleos? La razón es simple, es una doble evolución. Por un lado la capacidad que estos dispositivos tendrán y por otro la cantidad de innovación necesaria involucrada en el proyecto.

El termino Tera-Scale se refiere a la capacidad de manejar terabytes de información por segundo. Actualmente los procesadores manejan rendimientos de lo que sería Giga-Scale.

La idea de estos avances de gran salto es soportar aplicaciones de uso intensivo como: Data-Mining real-time de teraflops de información, Inteligencia Artificial para dispositivos, Realidad Virtual para modelado,

visualización, simulación de física y fines médicos, y algunas otras aplicaciones que todavía no son del todo posibles debido a su gran exigencia de procesamiento.

La computación Tera-Scale pretende poner al servicio de servidores, desktops y laptops, el poder computacional de una Supercomputadora de hoy en día.

2. Motivación

En los últimos 20 años los avances en tecnología de procesamiento se manifestaron como mejoras en la frecuencia de los procesadores, desde 5 Mhz hasta más de 3 Ghz. Pero recientemente los problemas de energía, calentamiento y limitaciones físicas fueron un problema para estas mejoras.

Con este nuevo modelo, y teniendo en cuenta la ley de More de que cada 18 a 24 meses se reduce al 70 % el tamaño de un transistor, es preferible construir chip de múltiples núcleos antes que chip mono-núcleo de mayor velocidad.

- **Desempeño:** Debido a que el aumento de frecuencia ya no es una posibilidad como antes, se considera como opción la utilización de procesamiento paralelizado en diseños de múltiples núcleos a fin de aprovechar el incremento de densidad de transistores por área.
- **Consumo:** Se podrán poseer algunos núcleos especializados para necesidades específicas a fin de optimizar el consumo. Y, los núcleos inactivos podrán ser desconectados a fin de ahorrar.

- **Ciclos de Diseño más rápidos:** Debido a la estandarización y separación de los diseños será posible reutilizar parte de los diseños actuales para las mejoras posteriores.

3. Oportunidades

Además de los desafíos, esta tecnología ofrece nuevas oportunidades a través de la integración de múltiples núcleos.

La comunicación entre núcleos es muchísimo más rápida que la conexión de chip a chip en equipos de múltiples procesadores. Así mismo superan a la cantidad actualmente posible de núcleos en servidores de alto desempeño actuales, y al estar integrados en un chip implican un menor costo.

También será posible construir chips especializados y de menor cantidad de núcleos para sistemas especializados, como por ejemplo un teléfono Voip o Celular.

4. Áreas de Investigación

4.1 Requisitos para la plataforma:

- *Programación:* Será necesario el desarrollo de aplicaciones altamente paralelizables a fin de aprovechar la tecnología, o bien el desarrollo de técnicas de predicción.
- *Adaptabilidad:* Deberá ser capaz de adaptarse a los cambios en distintas cargas de uso, entornos de hardware y cuestiones de temperatura.
- *Confiabilidad:* Al menos deberá mantener la confiabilidad actual de los procesadores existentes.
- *Escalabilidad:* La plataforma debe incrementar el desempeño a la misma razón de aumento de núcleos, es decir, debe ser real el hecho de a mas núcleos, mas rendimiento.

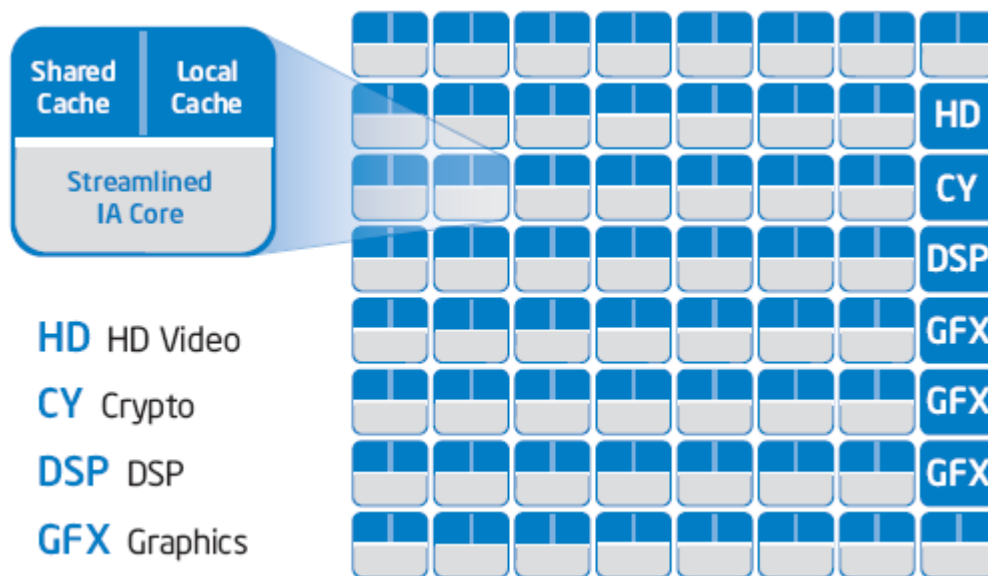
4.2 Diseño de Núcleos

Uno de los aspectos diferentes de esta arquitectura es el uso de micro arquitecturas optimizadas para el paralelismo, y cargas multifilares.

Los investigadores de Intel trabajan en determinar el diseño optimo de los núcleos así como en número de núcleos que pueden ser efectivamente utilizados dado un proceso de fabricación (65nm, 45nm, 32nm, 22nm).

Además se investiga las técnicas para lidiar con software primitivo y mono hilo, como técnicas de división de ejecución y predicción.

Como se aprecia abajo, variando la mezcla de componentes funcionales es posible optimizar el rendimiento en aplicaciones determinadas y son fácilmente reutilizables de procesador a procesador.



4.3 Unidades específicas

A fin de optimizar la utilización de energía y de proporcionar mejor rendimiento en tareas específicas en comparación a procesadores de uso general estas unidades se combinan dentro del chip según las necesidades de procesamiento.

4.4 Intercomunicación

Se hace evidente que tantos núcleos requieren una tecnología de intercomunicación avanzada, y para esto se implementan dos buses dentro del mismo procesador, uno de anillos de distintos tamaños, siguiendo el modelo de calle y autopista, y otro bus de tipo mesh que conecta todos los núcleos horizontal y verticalmente, permitiendo así comunicaciones de baja latencia y alta velocidad.

Ventajas:

- Estas redes son más baratas y numerosas que las posibles fuera del chip (sobre circuitos impresos).
- Debido a la menor distancia, las latencias tienen mejor respuesta y menor consumo energético.
- El ancho de banda de estos canales es mayor que cualquiera fuera del procesador en sí.

4.5 Administración de Energía

La tecnología implementa sistemas de balanceo de carga entre núcleos con la finalidad de reducir el calentamiento y por ende la resistencia de los conductores, y al mismo tiempo, la confiabilidad de los mismos.

También permite la activación de solamente los núcleos necesarios para soportar la carga actual de procesamiento a fin de ahorrar energía, administrando la cantidad de núcleos dinámicamente.

Además, el diseño de los circuitos de memoria y cache pretenden optimizar hasta tres veces la velocidad de acceso, reduciendo al mismo tiempo hasta 5 veces la consumo.

Por otro lado, también se implementan tecnologías de variación dinámica de frecuencia de los núcleos para cubrir las necesidades planteadas.

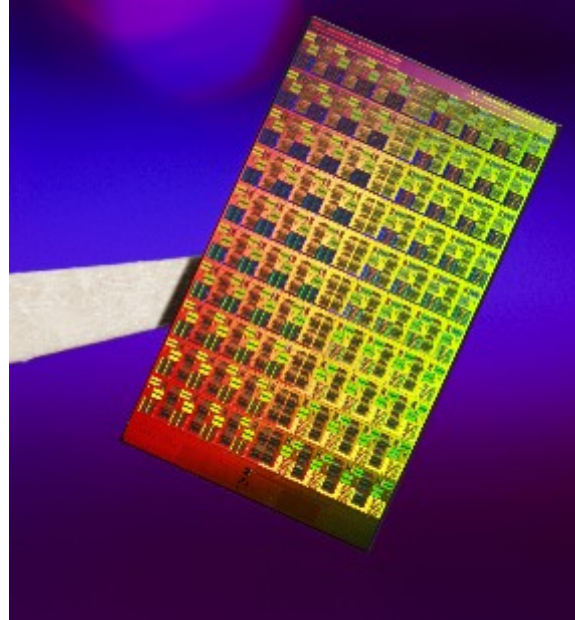
5. Investigaciones en la Plataforma

5.1 Cache y Memoria

La jerarquía del cache incluirá un cache privado por unidad de procesamiento (Cache L1), y varios caches compartidos ubicados físicamente en las unidades (Cache L2), sería una especie de cache L2 Distribuido.

En cuanto a la memoria, el cambio es importante, esta se encuentra ubicada físicamente debajo del procesador formando una interconexión 3-D a fin de permitir una gran capacidad de transferencia y baja latencia.

Esto se vuelve indispensable debido a la gran cantidad de unidades de procesamiento disponibles para consumir y producir información.



5.2 Entrada y Salida

En cuanto a la entrada, los buses actuales se reorganizaran en jerarquías según la capacidad de transferencia y latencia, esto impone la necesidad de una mejora en el rendimiento de los buses existentes y donde se hable inclusive hasta de buses basados en luz en vez de electricidad. Además de buses basados en radios para dispositivos remotos.

También será contemplada la virtualización por hardware, los espacios de direcciones de entrada y salida virtuales y otras tecnologías para permitir una vitalización completa del hardware. Debido a que el diseño del Procesador es ideal para este tipo de tareas.

5.3 Virtualizacion y Particionamiento

Además de lo mencionado anteriormente, se plantea también la inclusión de reglas de QoS dentro del mismo procesador y en el acceso a la entrada y salida, a fin de permitir la garantía en ciertos servicios además de otras seguridades. Con esto quedaría cubierto a nivel de hardware todo lo que se requiere hoy en día en servidores virtuales de alta calidad.

6. Desarrollo de Software

Esta área contiene a dos de los más grandes desafíos del proyecto en sí.

Primero, asegurarse de explotar al máximo la masiva capacidad de cómputo de esta tecnología, estudiando e identificando las cargas y algoritmos que serán utilizados a fin de proveer tecnologías y herramientas que faciliten el desarrollo y la escalabilidad.

Segundo, La depuración, la gran cantidad de unidades de procesamiento genera una marcada dificultad en el testeado y la depuración de aplicaciones, lo que implica una necesidad importante en el desarrollo de herramientas de programación en ambientes altamente multifilares.

7. Conclusión

Intel encuentra futuro en los dispositivos alrededor de la arquitectura tera-scale, que utilizaran de decenas a centenas de núcleos a fin de procesar gran caudal de información además de ser dispositivos escalables y adaptables a las necesidades.

En el mundo tera-scale existirán nuevas posibilidades de minería e interpretación de la gran cantidad de información existente hoy en el mundo, permitiendo así mayor eficiencia.

Los agentes inteligentes podrán tomar mejores decisiones en tiempo real, basándose en una mayor cantidad de observaciones, e inclusive en las casas se mejorara el procesamiento de videos y fotos que las personas acostumbran a tener en grandes cantidades dentro de sus discos duros. Se menciona también la posibilidad de implementar algoritmos complejos y avanzados de mejora de fotografías y algoritmos de búsqueda de personas dentro de filmaciones, además de generación y administración de otros metadatos dentro de estos medios.

También es importante la posibilidad de emparejamiento con las redes que hoy en día superan ampliamente a los anchos de banda disponibles para las computadoras (Ej, el ancho de banda de un procesador a su cache se encuentra alrededor de los 10 Gbps y las redes están manejando magnitudes de los 50 Gbps, donde al combinarse aumentan linealmente).

Debido a todo lo expuesto en este documento, se hace clara la posibilidad de que esta arquitectura sea la siguiente generación de procesadores planteando una mejora drástica en las distintas capas de almacenamiento y procesamiento de información.

Se espera que esta tecnología se encuentre en el mercado para el 2015, pero a medida que las investigaciones individuales vayan teniendo éxito, Intel menciona la incorporación de estas en sus procesadores actuales.

