

Tile64 Many-Core

vs.

Intel Xeon Multi-Core

Comparación del Rendimiento en Bioinformática

Myriam Kurtz
Francisco J. Esteban
Pilar Hernández
Juan Antonio Caballero
Antonio Guevara
Gabriel Dorado
Sergio Gálvez

random|plasmid



Contenido

- **Tecnologías Many Core**
- **Otras Tecnologías**
- **Alineación de Secuencias en Bioinformática**
- **Algoritmo MC64-NW/SW**
- **Resultados de la Comparación de Rendimiento**
- **Conclusiones y Perspectivas Futuras**

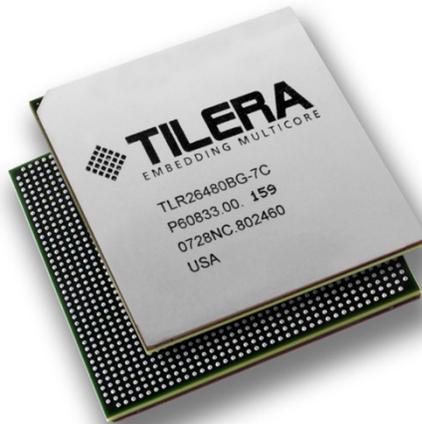
Tecnologías Many-Core

La ingeniería informática y la computación de alto rendimiento (HPC), están evolucionando muy rápido hacia nuevas arquitecturas de chips multiprocesador (CMP).

- **Las unidades de procesamiento gráfico de propósito general (GPGPU)**
- **Las unidades de procesamiento central de múltiples núcleos (CPU)**

Otras Tecnologías

Las unidades de procesamiento central de múltiples núcleos, que representan el marco de este trabajo, son:



Tilera Tile64



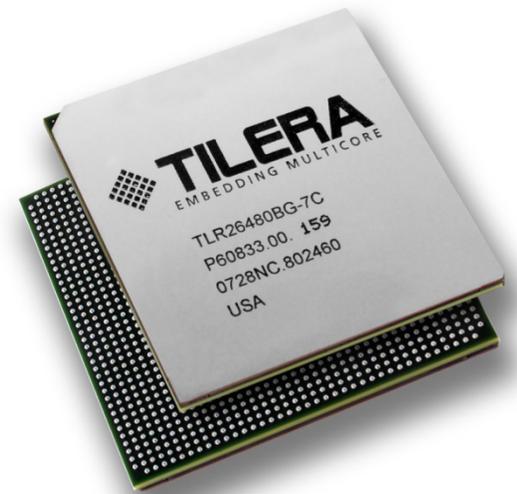
Intel Xeon

Existen otros enfoques, que van:

- Procesadores heterogéneos: Cell Broadband Engine (celular BE)
- Arquitecturas como la Epifanía

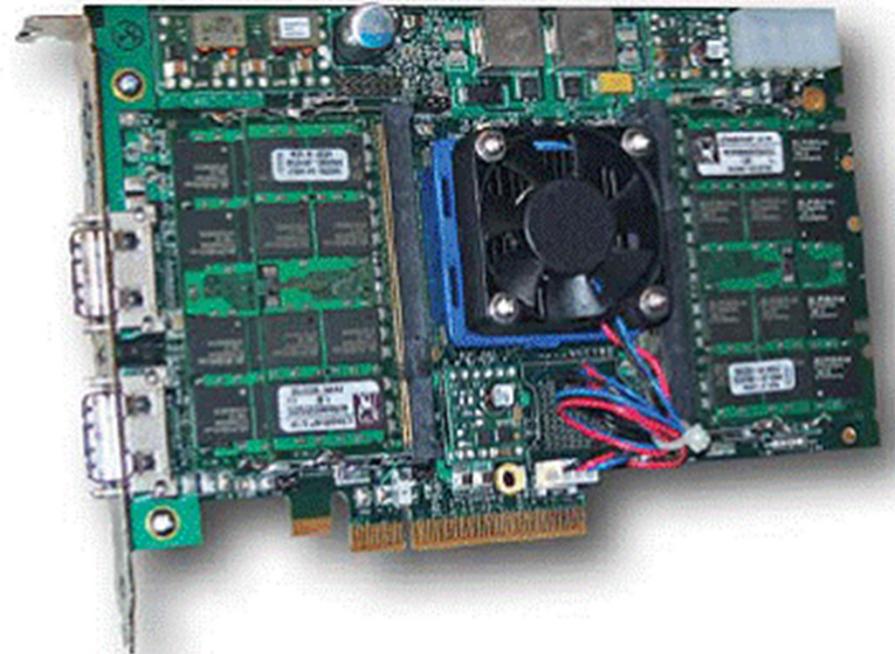
Procesador Tile64

- **Contiene 64 núcleos RISC**
- **Funciona a 866 MHz con tres niveles de memoria caché, interconectados a través de una red de banda ancha llamada inteligente Mesh (iMesh)**
- **Cada mosaico (tile) puede ejecutar un proceso independiente**
- **Pueden comunicarse entre sí a través de memoria compartida o mediante paso de mensajes y bibliotecas propietarias basadas en canal.**



Tarjeta Tiler TiExpress-20G

- Bus PCI Express (PCIe)
- 8 GB SO-DIMM
- DDR2
- Se controla desde un host, utilizando una interfaz de línea de comandos
- Se puede utilizar como memoria compartida y locales, así como disco de estado sólido (SSD)



Alineación de Secuencias en Bioinformática

Un alineamiento de secuencias en bioinformática es una forma de representar y comparar dos o más secuencias o cadenas de ADN (ácido nucleico), ARN (ácido ribonucleico), o estructuras primarias proteicas, para resaltar sus zonas de similitud.

Las secuencias alineadas, se escriben con las letras, que representan aminoácido o nucleótidos en filas de una **matriz**.

```
gi|160797|gb|AAA29796.1|
gi|9816|emb|CAA77743.1|
gi|56749856|sp|P68871|HBB_HUMA
gi|18015|emb|CAA37898.1|
```

```
MHSSIVLATVLFVAIASASKTRELCMKSLAHAKVGTTSKEAKQDGLDLYKH 50
MHSSIVLATVLFVAIASASKTRELCMKSLAHAKVGTTSKEAKQDGLDLYKH 50
-----
-----MSTL 4
::  ::  :  :::::  .  .  :  .  :  :::::.....
```

Algoritmo MC64-NW/SW

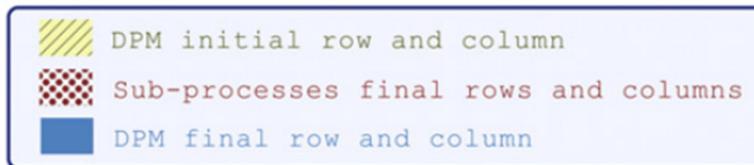
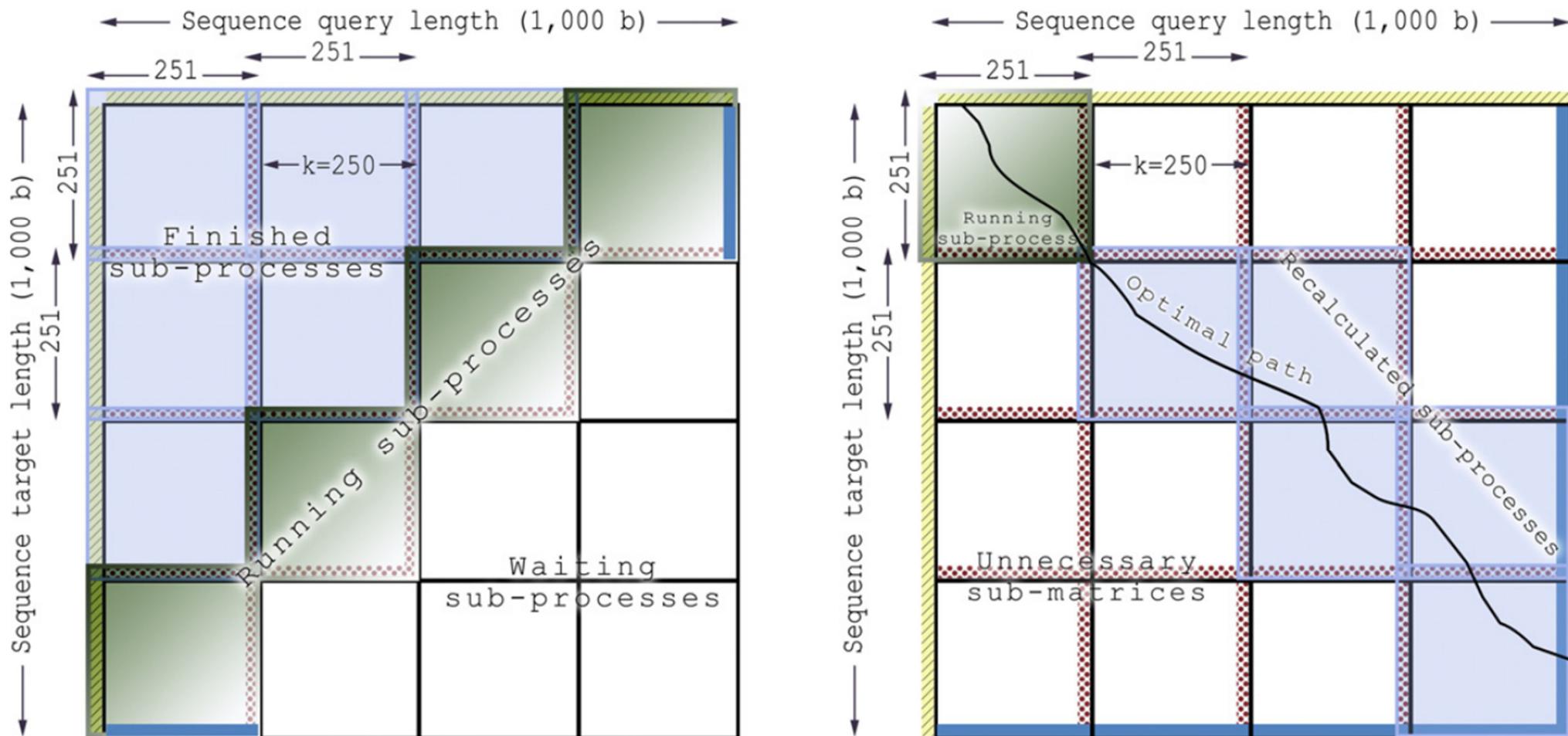
Versión paralela de los algoritmos *Needleman-Wunsch* (NW) y *Smith-Waterman* (SW) para probar el potencial del procesador Tile64.

Implementación (nos centramos en NW):

NW original, ejecuta dos fases:

- 1. Crea una cuenta matriz de programación dinámica (DPM), tamaño $n \times m$**
- 2. Atraviesa la DPM linealmente para conocer la alineación actual**

Algoritmo MC64-NW/SW



Algoritmo MC64-NW/SW

Conclusión primera fase:

El número de puestos de trabajo en paralelo, aumenta linealmente y la DPM se crea después de un frente de onda.

La potencia total del sistema, se consigue cuando el número de puestos de trabajo paralelos alcanza el número de mosaicos disponibles.

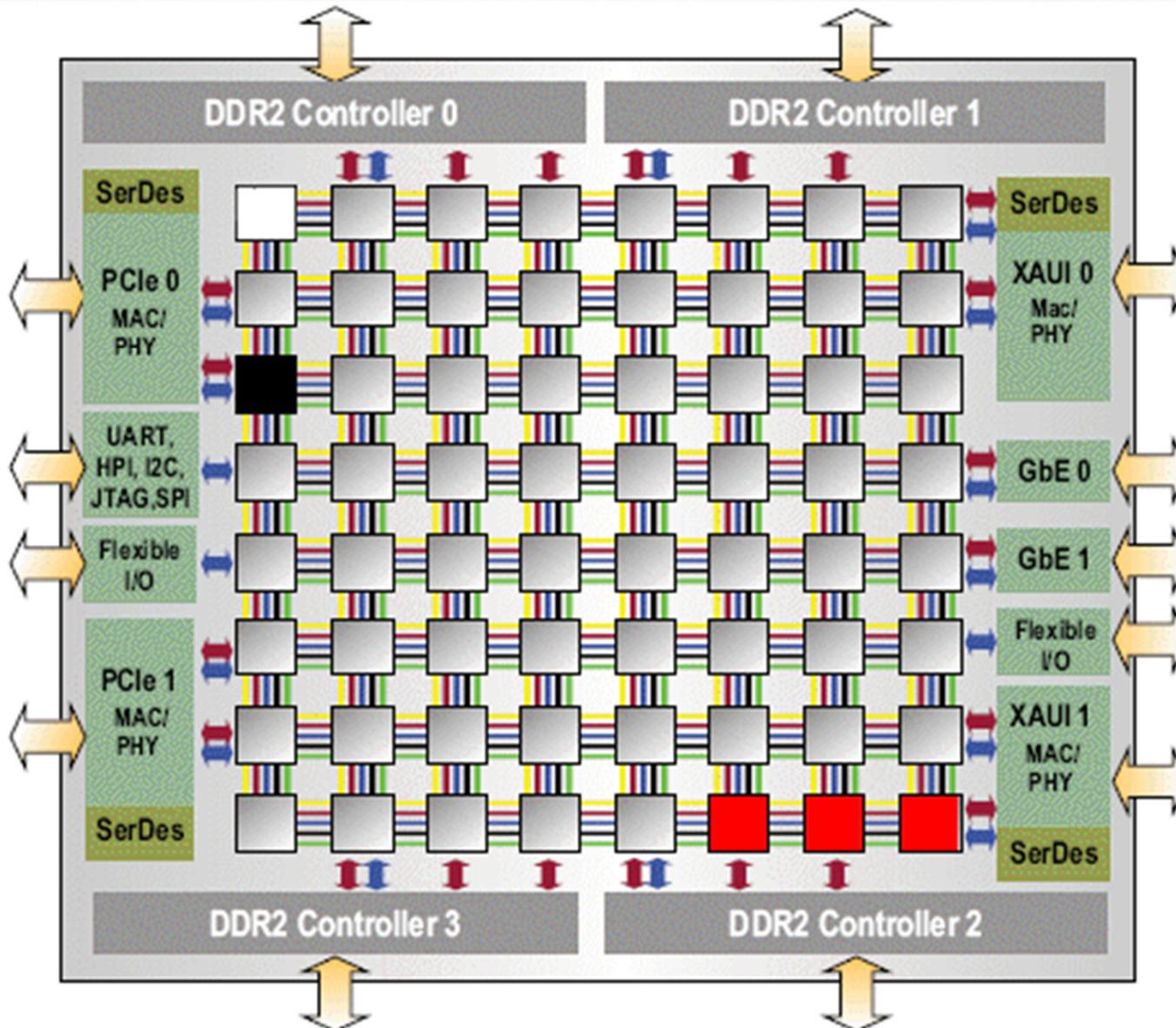
Algoritmo MC64-NW/SW

Conclusión segunda fase:

Se recalculan unas pocas submatrices: $\max(n, m)$ en el mejor de los casos, y $n + m - 1$ en el peor de los escenarios.

Este proceso es completamente lineal y se lleva a cabo por una sola celda.

Distribución de Trabajo y el Contenido de Cada Celda



Tile64

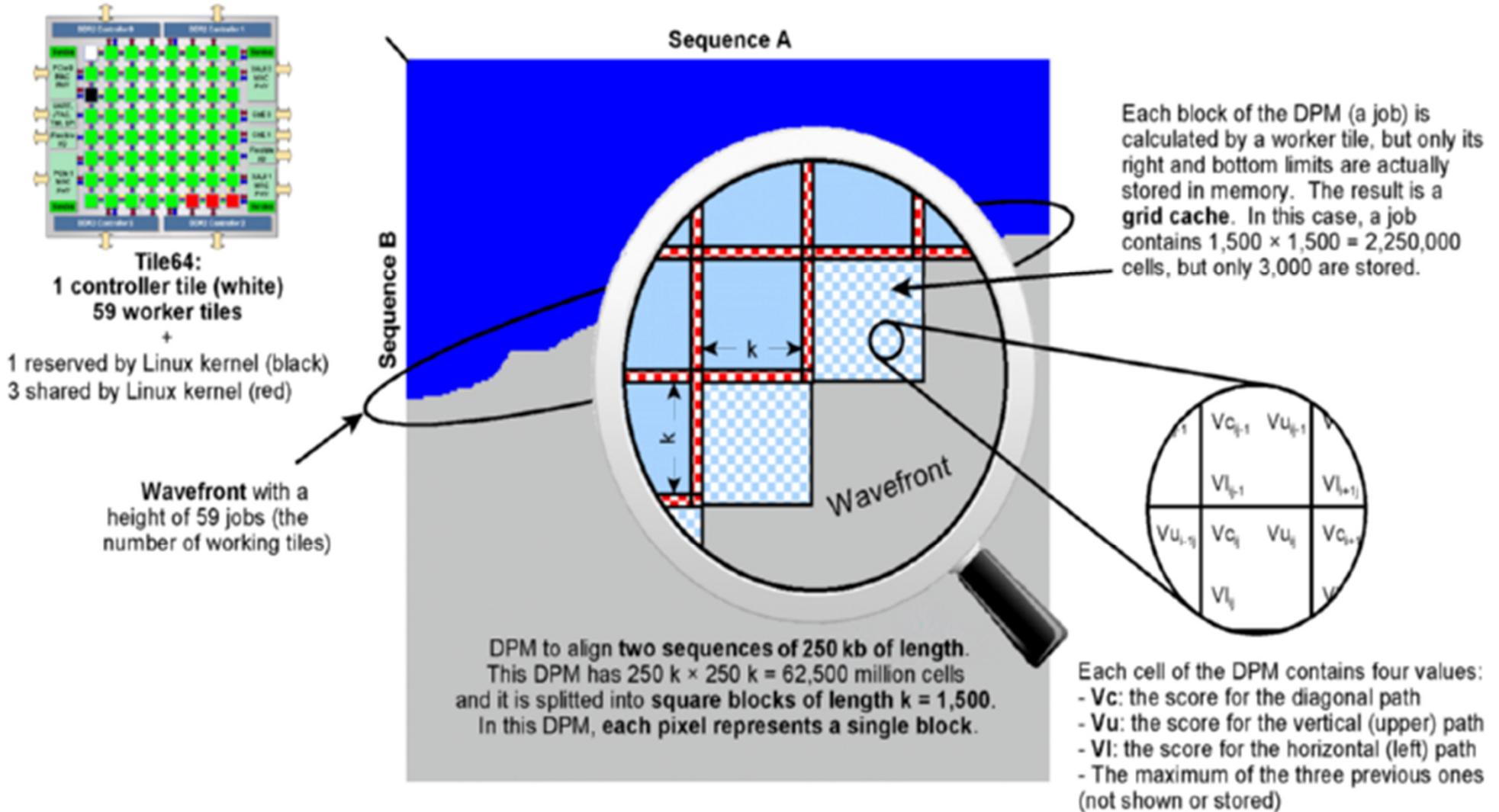
1 Mosaico controlador (blanco)

59 mosaicos de trabajo

1 reservado para el kernel de GNU/Linux (negro)

3 compartidos por el kernel de GNU/Linux (rojo)

Distribución de Trabajo y el Contenido de Cada Celda



MC64-NW/SW en una arquitectura multinúcleo x86

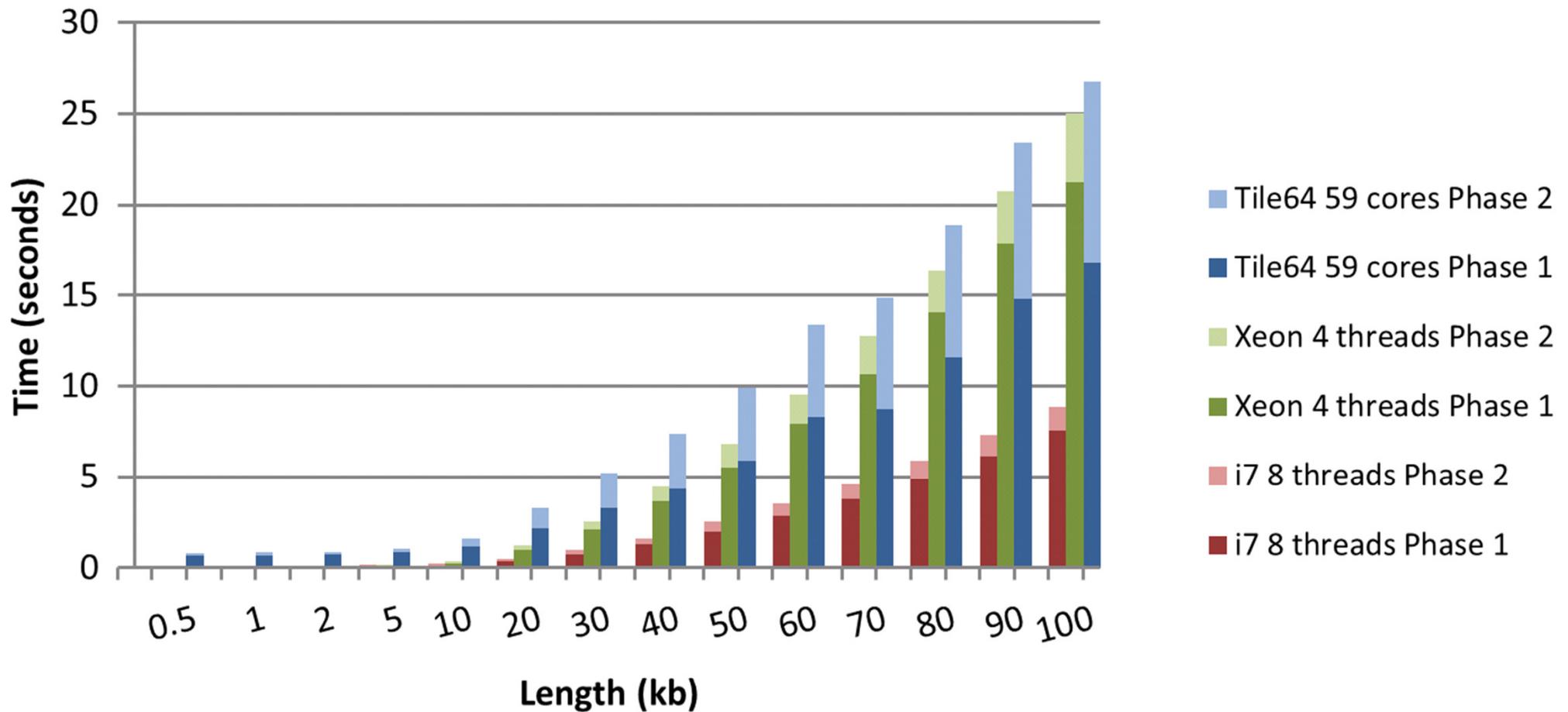
Migración del algoritmo **MC64-NW/SW** escrito en C a plataforma multi-core x86 = **MT-NW/SW**.

MT = multi thread

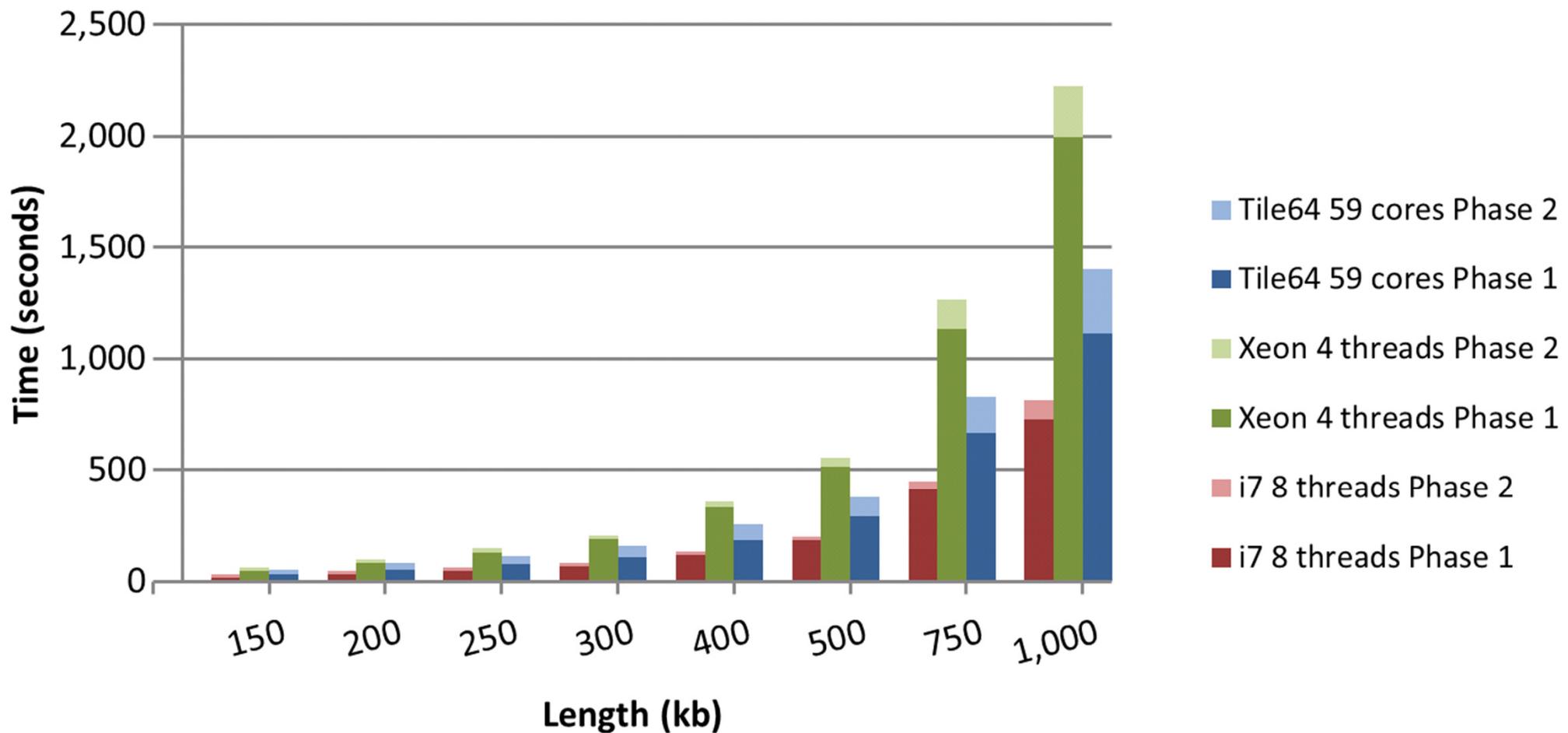
Tres Principales Cambios:

- El proceso ha sido sustituido por la creación del hilo, y la inicialización.
- Paso de mensajes, sustituida por acceso a la memoria compartida.
- Las variables globales se han encapsulado con una metodología de programación orientada a objetos.

Comparación de Rendimiento



Comparación de Rendimiento



Conclusiones y Perspectivas Futuras

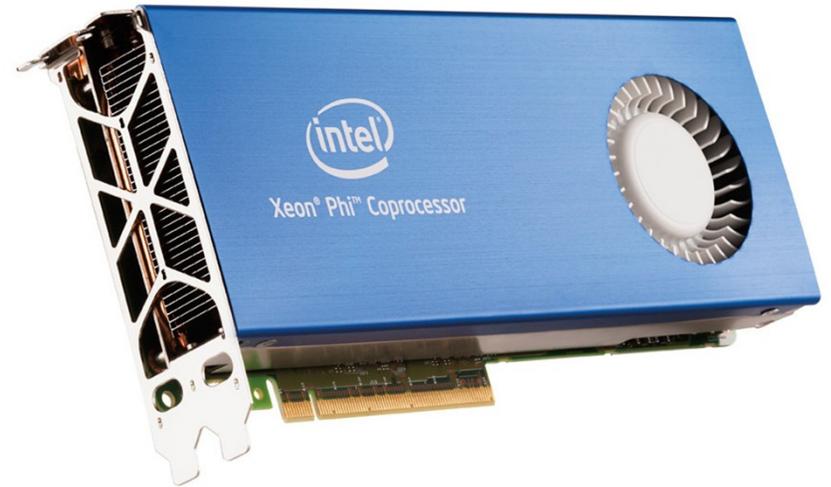
El análisis de las pruebas realizadas, demuestra:

- **Una clara ventaja de la nueva tecnología informática del conjunto de instrucciones complejas (CISC) sobre el conjunto de instrucciones reducidas (RISC).**
- **Por otro lado, la plataforma *Intel i7 3770* es de menor costo en comparación con la tarjeta *TiExpress-20G*.**

Conclusiones y Perspectivas Futuras

Líneas futuras de trabajo:

- Probar el algoritmo MT-NW/SW en una tarjeta coprocesadora PCIe *Intel Xeon Phi*,



- Integrar 60 núcleos con cuatro canales hyper-threading cada uno y 8 GB de GDDR5

Conclusiones y Perspectivas Futuras

Líneas futuras de trabajo:

En la actualidad, estamos desarrollando una variante para explotar al máximo las capacidades de vectorización del Xeon.

Al mismo tiempo, estamos aplicando un nuevo enfoque para poner en marcha inmediatamente los cuatro hilos de cada núcleo, al momento de iniciar un trabajo.

Conclusiones y Perspectivas Futuras

Líneas futuras de trabajo:

Dependiendo del rendimiento logrado, de las nuevas pruebas, se podrían desarrollar algoritmos más precisos, con el fin de obtener, por ejemplo, ejecuciones del algoritmo *BLAST* con valores de parámetros que permiten una búsqueda más a fondo, dendrogramas más precisos.

En una próxima publicación se verán otros resultados, teniendo en cuenta la utilidad desde el punto de vista biológico.

Agradecimientos

A  **TILERA**® por proporcionar las herramientas de hardware y software: *<http://www.tilera.com>*

Este trabajo fue apoyado por las siguientes Instituciones de España:

- "Ministerio de Economía y Competitividad" (MINECO subvenciones AGL2010-17316 y BIO2011-15237-E)
- "Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria" (MINECO y el INIA RF2012-00002-C02-02)
- "Consejería de Agricultura y Pesca" (041/C/2007, 75/C/2009 y 56/C/2010)
- "Consejería de Economía, Innovación y Ciencia" (AGR-7322) de la "Junta de Andalucía"
- "Grupo PAI "(AGR-248)
- "Universidad de Córdoba "(" Ayuda de Grupos ")

Tile64 Many-Core vs. Intel Xeon Multi-Core

Comparación del Rendimiento en Bioinformática

M. Kurtz / F. Esteban / P. Hernández / J. Caballero / A. Guevara / G. Dorado / S. Gálvez