



# Computación Paralela y Distribuida

---

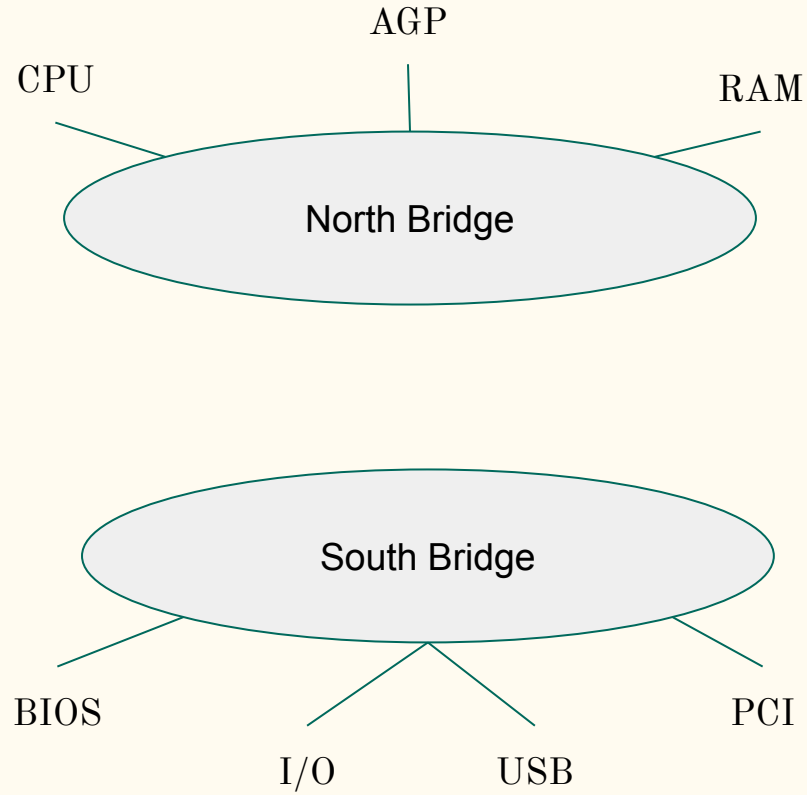
Latencia y Ancho de Banda

John Corredor, PhD  
ECEI - IS

# Objetivo:

- Conocer los conceptos de Latencia y Ancho de Banda
- Identificar las Arquitecturas Multicore

# Buses



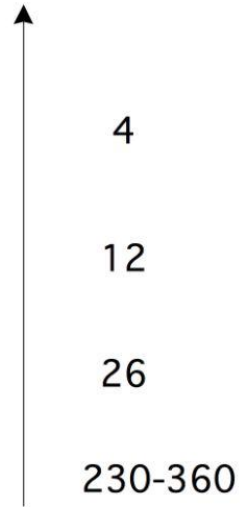
# Latencia y Ancho de Banda

La latencia es el retraso entre el procesador que emite una petición de un elemento de memoria y el elemento que llega realmente. Podemos distinguir entre varias latencias, como la transferencia de la memoria a la caché, la caché al registro, o resumirlas todas en la latencia entre la memoria y el procesador. La latencia se mide en (nanosegundos, o períodos de reloj).

El ancho de banda es la velocidad a la que los datos llegan a su destino, una vez superada la latencia inicial. El ancho de banda se mide en bytes (kilobytes, megabytes, gigabytes) por segundo o por ciclo de reloj. El ancho de banda entre dos niveles de memoria suele ser el producto de la velocidad de ciclo del canal (la velocidad del bus) y el ancho del bus: el número de bits que pueden enviarse simultáneamente en cada ciclo del reloj del bus.

Jerarquía de memoria de un Intel Sandy Bridge, caracterizada por la velocidad y el tamaño.

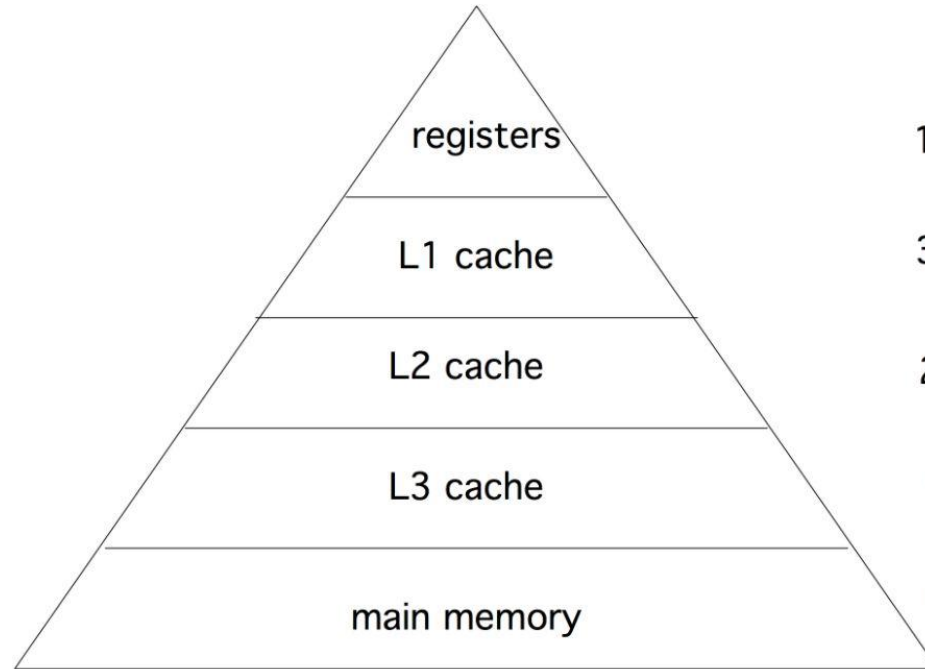
Latency from next  
level (cycles)



Size (bytes)

192  
32k  
256k  
2M  
2G

A downward-pointing arrow on the right side of the diagram, indicating that size increases as you move down the memory hierarchy from registers to main memory.



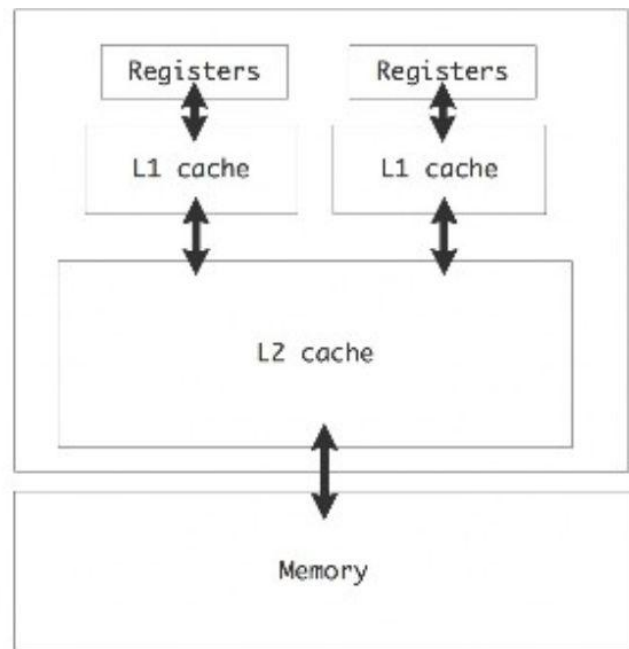
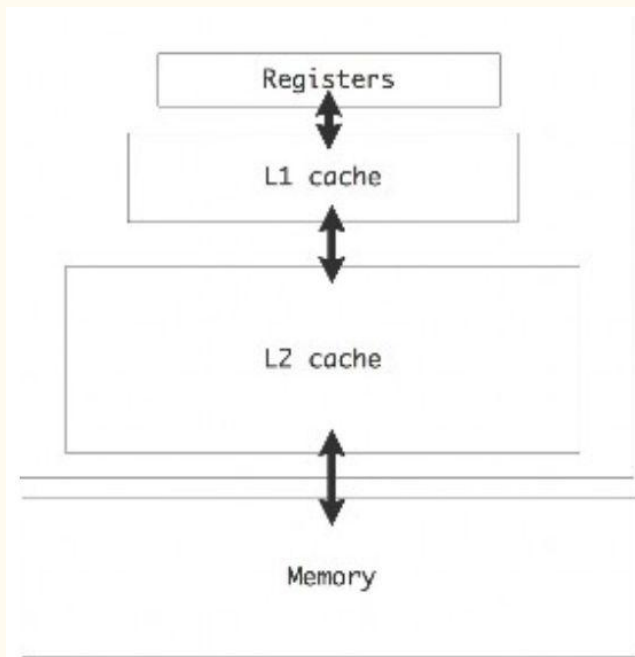
# Arquitecturas Multicore

La frecuencia de los relojes no puede aumentar más, ya que incrementa el consumo de energía, calentando demasiado los chips.

No es posible extraer más Paralelismo del Nivel de Instrucción (ILP) de los códigos, ya sea por las limitaciones del compilador, por la cantidad limitada de paralelismo intrínsecamente disponible, o porque la predicción de saltos lo hace imposible.

# Solución de problemas

- Dos núcleos a una frecuencia más baja pueden tener el mismo rendimiento que un solo procesador a una frecuencia más alta; por lo tanto, los núcleos múltiples son más eficientes desde el punto de vista energético.
- El ILP descubierto es ahora reemplazado por un paralelismo de tareas explícito, manejado por el programador.







Dual-core SCC Tile



1 Router



# Referencias

- ★ Hennessy J L and Patterson D. 2011. Computer Architecture, Fifth Edition: A Quantitative Approach 5th ed (San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.) ISBN 012383872X, 9780123838728.
- ★ David B. Kirk and Wen-mei W. Hwu. 2010. Programming Massively Parallel Processors: A Hands-on Approach, Third Edition. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA ©2010 ISBN:0123814723 9780123814722.
- ★ Michael J. Quinn. 2003. Parallel Programming in C with MPI and OpenMP. McGraw-Hill Education Group.
- ★ Victor Eijkhout. 2012. Introduction to High Performance Scientific Computing. Lulu.com.
- ★ Michael McCool, James Reinders, and Arch Robison. 2012. Structured Parallel Programming: Patterns for Efficient Computation (1st ed.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- ★ Tutorial online:  
[https://www.tutorialspoint.com/parallel\\_computer\\_architecture/index.htm](https://www.tutorialspoint.com/parallel_computer_architecture/index.htm)