Many Cores

A discussion of the challenges encountered in implementing a multi-core system.

Many Core Challenges

-As the number of cores increases, coherence traffic increases
-As the number of cores increases, the bus traffic increases to the point of bottleneck

Network on Chip

Instead of a bus, use a mesh connection. This will increase the throughput of the entire network.
With a mesh, each additional core increases the number of connections to the network.
Torus Networks are a wrapped mesh network.

Many Core Challenges 2

More cores means increased coherence traffic, requiring:

1. A scalable on-chip network
2. Directory Coherence

Off chip traffic increases while the number of pins on the chip increases but not at the same rate. To solve this problem:

The last level cache is shared and distributed among all the cores and its size increases with each core.

Distributed LLC

The distributed last level cache (LLC) is sliced up and controlled by each core.
How to slice the cache?

-Round robin - this method is not good for locality
-Round robin with page numbers - this method is better for locality

Many Core Challenges

The Coherence Directory becomes too large when there are many cores.

On-Chip Directory

So the on-chip directory is sliced and distributed to each core.
The directory is sliced the same as the LLC.

Partial Directory - a limited number of entries reserved for blocks that are in at least one cache.

On-Chip Directory 2

When a directory becomes full, use an LRU protocol to replace an entry. This type of miss is caused by invalidation due to a replacement.

Many Core Challenges 3
The power budget available for a chip must be split between cores, so the frequency and voltage must be reduced.

Multi-Core Power and Performance
The more cores, the slower each core can operate. To combat this, the cores that are operating get a boost of power and frequency.

Multi-Core Challenges 4
The final challenge for Multi-cores is Operating system confusion caused by multi-threading and cache sharing between cores.

SMT, Cores, Chips
The operating system needs to know on which core to run the thread to obtain the best performance.