



Effectiveness of Off-Chip Caches for Commercial Applications

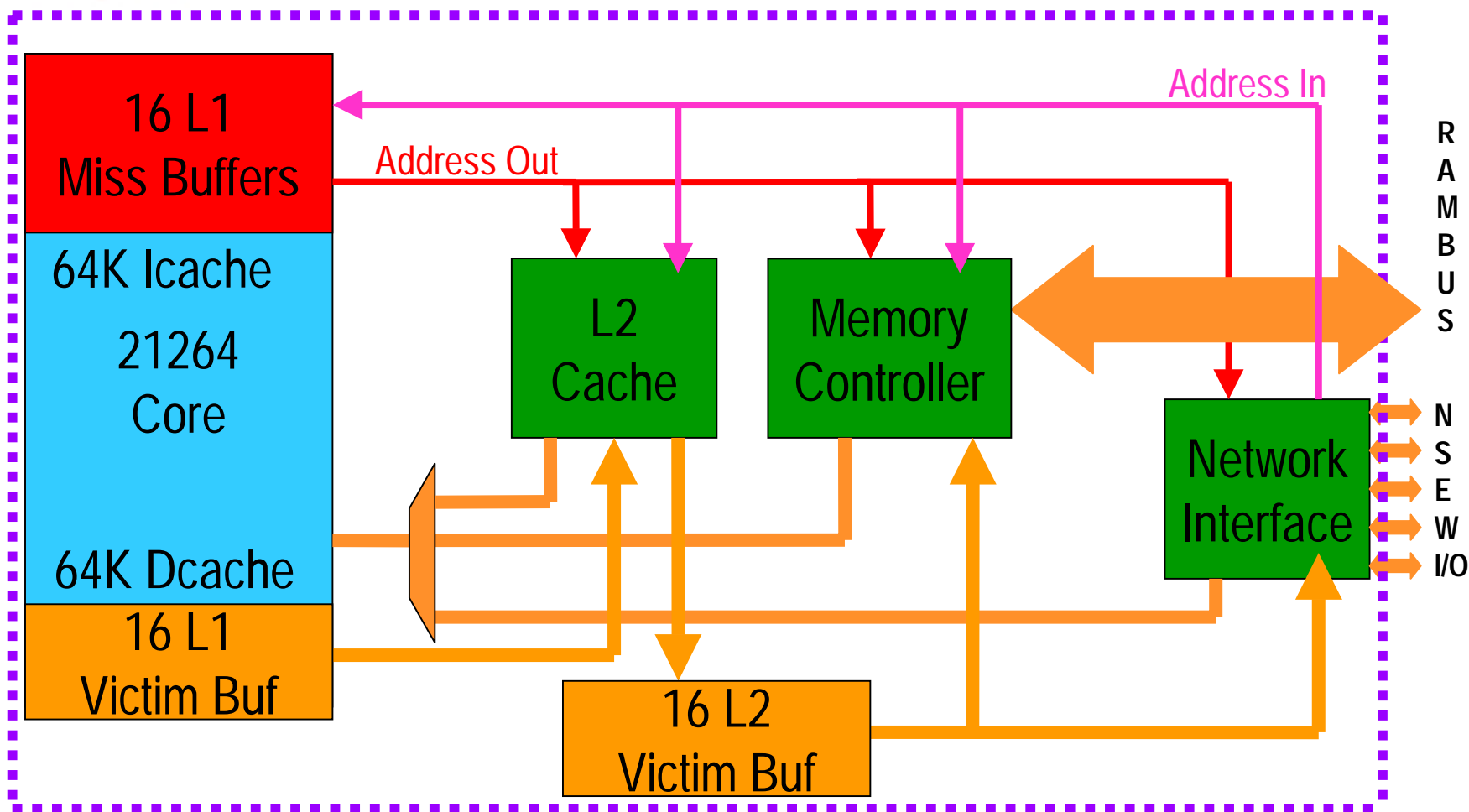
Ben Verghese, Luiz André Barroso,
and Kouros Gharachorloo

Western Research Lab
Compaq Computer Corporation

*Presented at the 1999 ISCA Workshop on Scalable Shared
Memory Multiprocessors
April 30, 1999*



Everything but the Kitchen Sink All on One Die

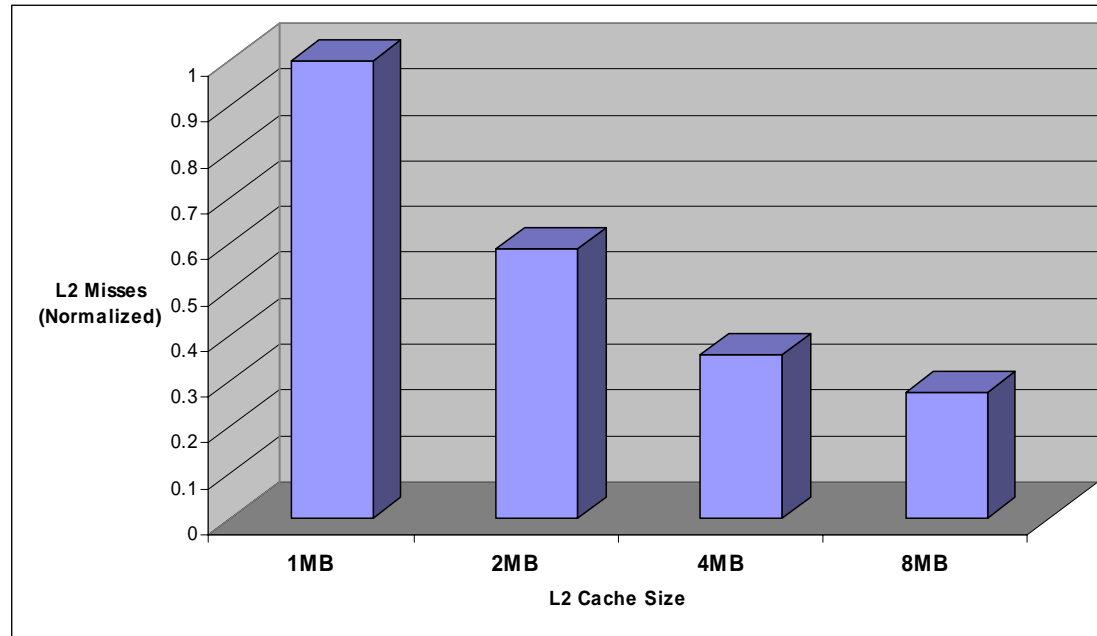


Modern integrated microprocessor → limited L2 cache size (<2MB)



But the Workload of Interest Is

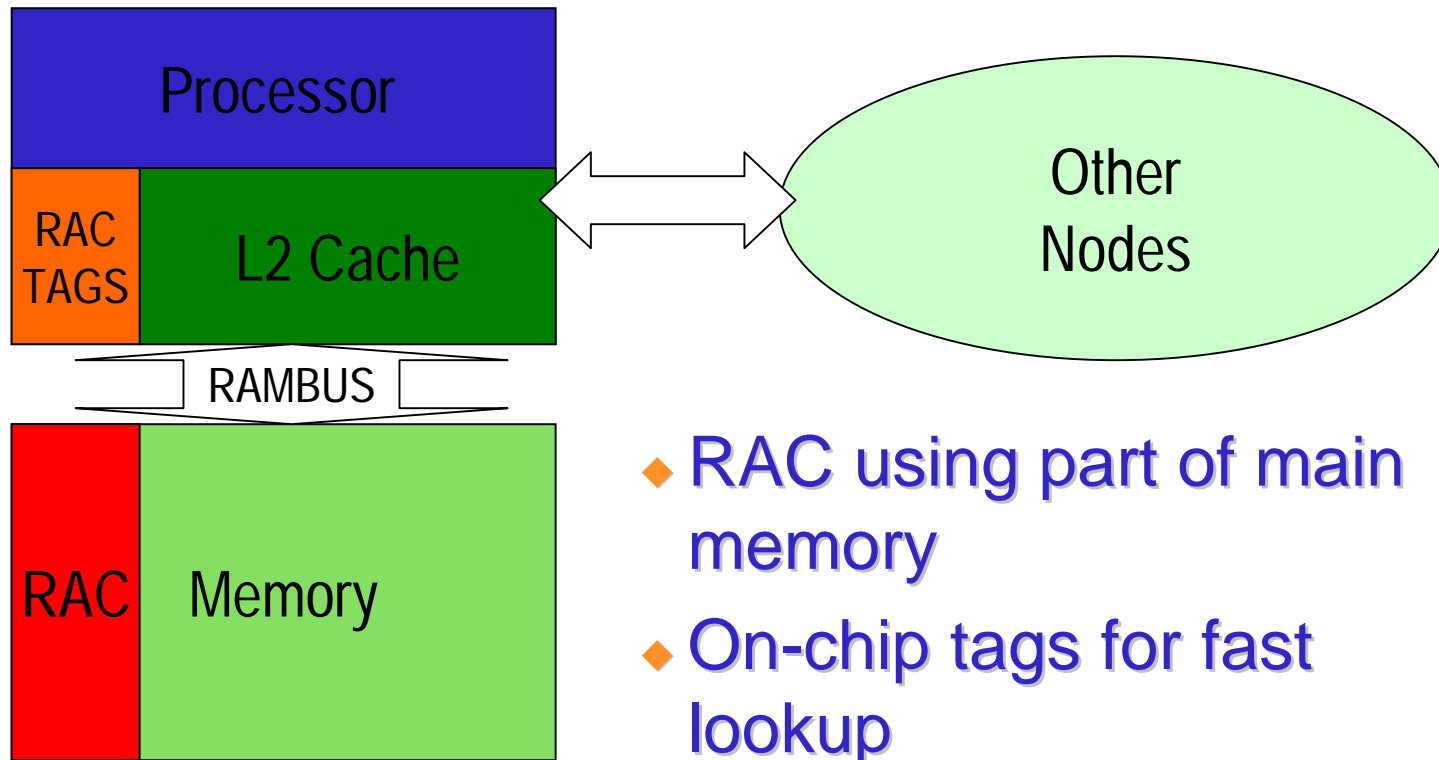
- ◆ Commercial Applications
 - OLTP workload
 - Database Scaled for realistic results



Seems to require large L2 caches (> 4MB)



What to do with the Misses



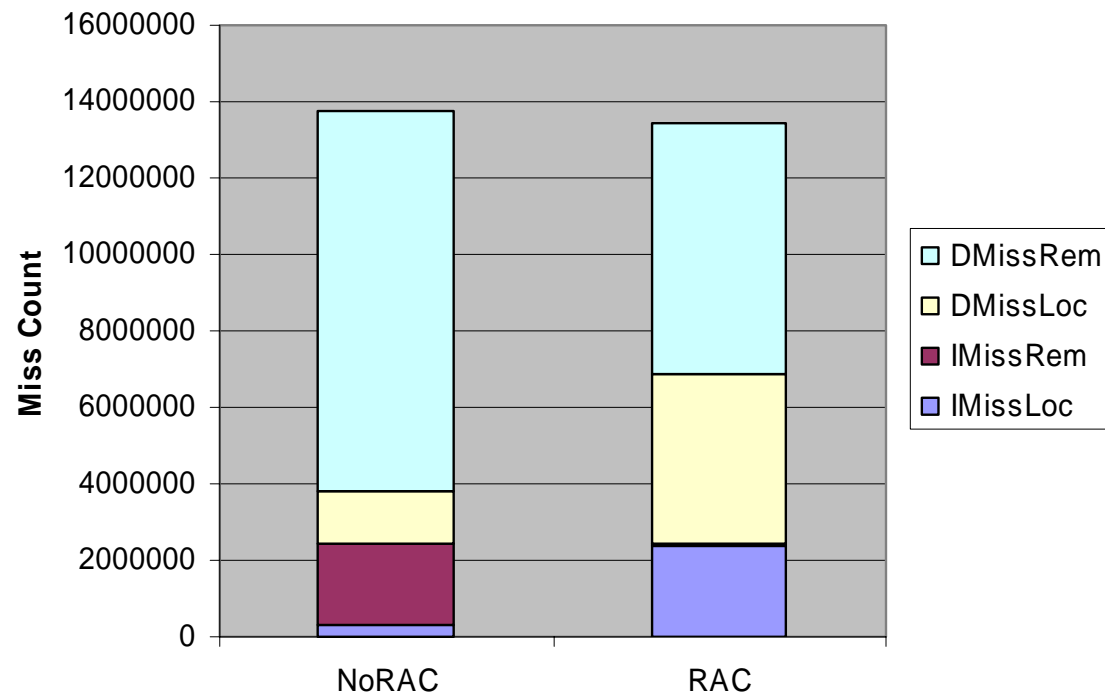


Lies, Damn Lies and Simulation

- ◆ Evaluated Using SimOS-ALPHA
- ◆ Machine parameters
 - 1GHz CPU/8 processor MP
 - 64K 2-way L1 I&D caches
 - 1M 4-way unified L2 cache
 - Associativity based on 256K banks
 - 15 cycle hit time
 - CC-NUMA memory system
 - 75:150:200
 - 8M 8-way RAC



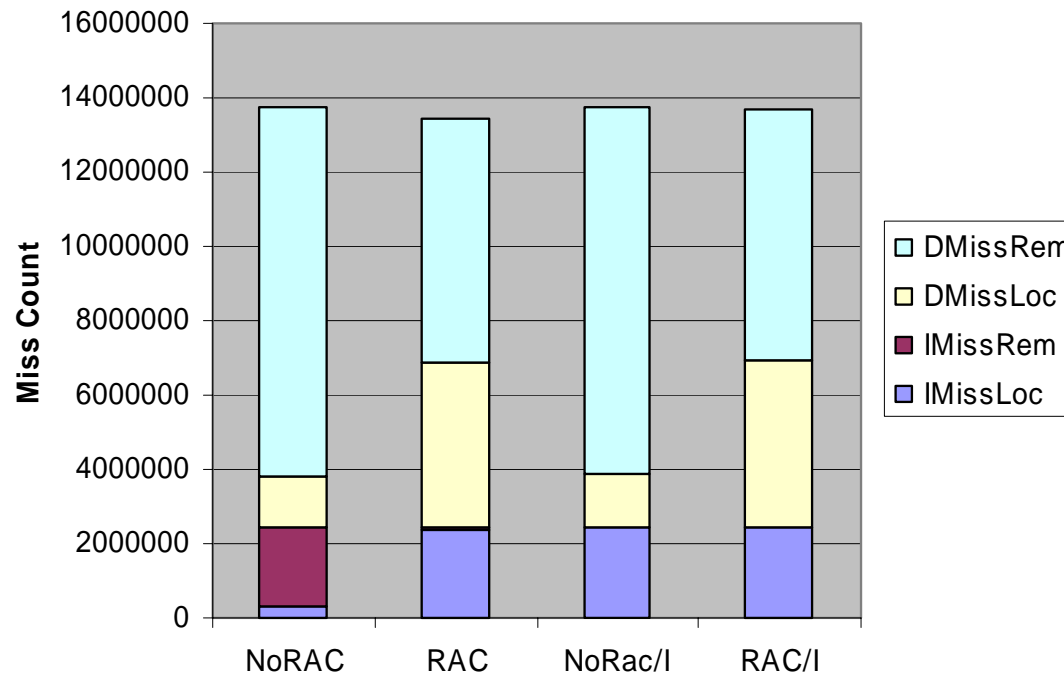
Why Add an Off-Chip Cache



- ◆ RAC hit rate 42.2%
- ◆ Seems to be quite effective



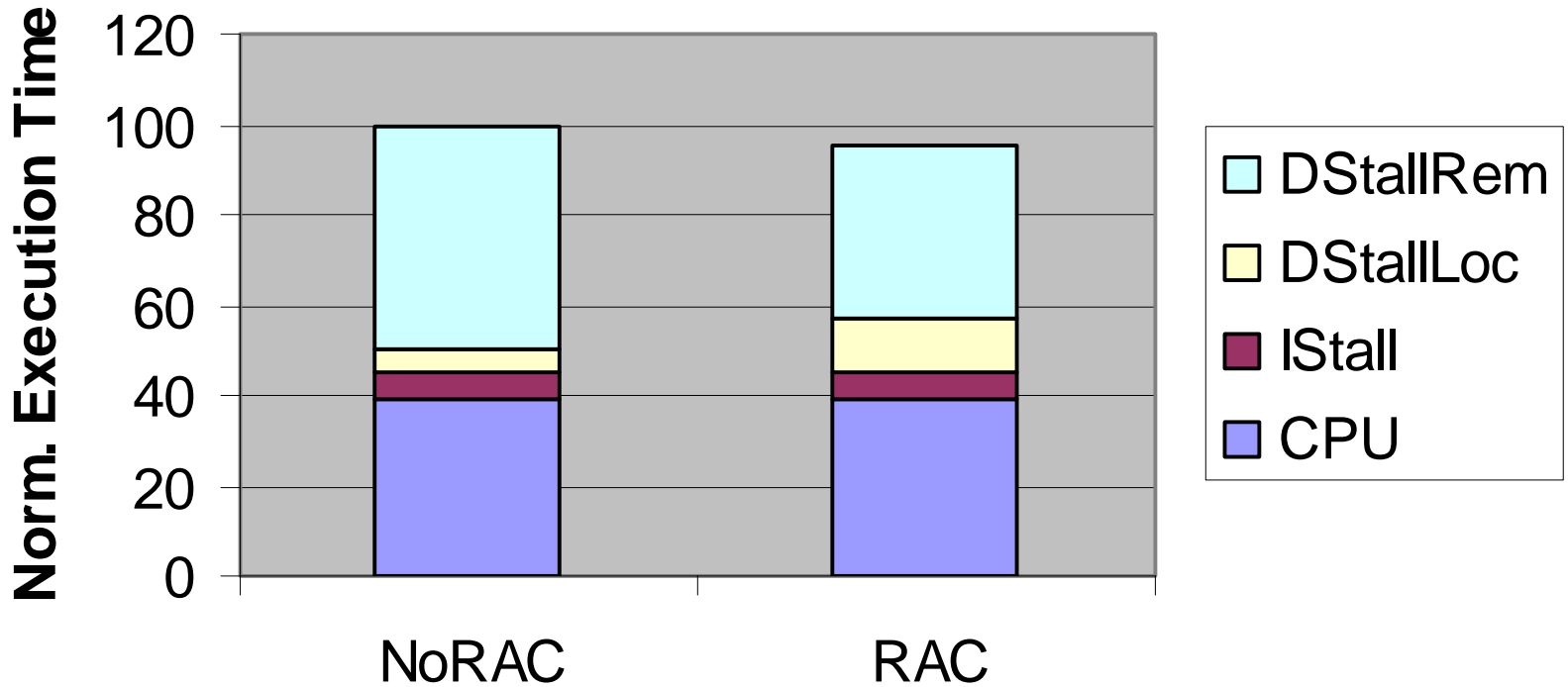
Instructions Can Be Made Local



- ◆ Replication of code makes instructions local
- ◆ RAC Hit Rate 29.9%



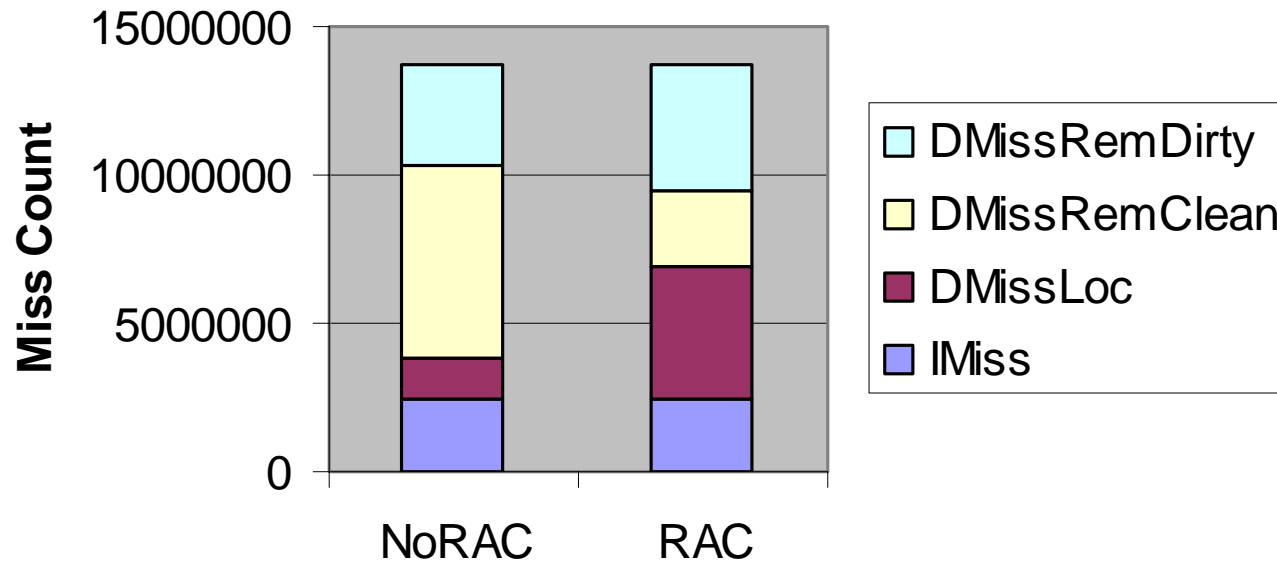
Where Did All That Time Go



◆ Overall benefit only 4.3%



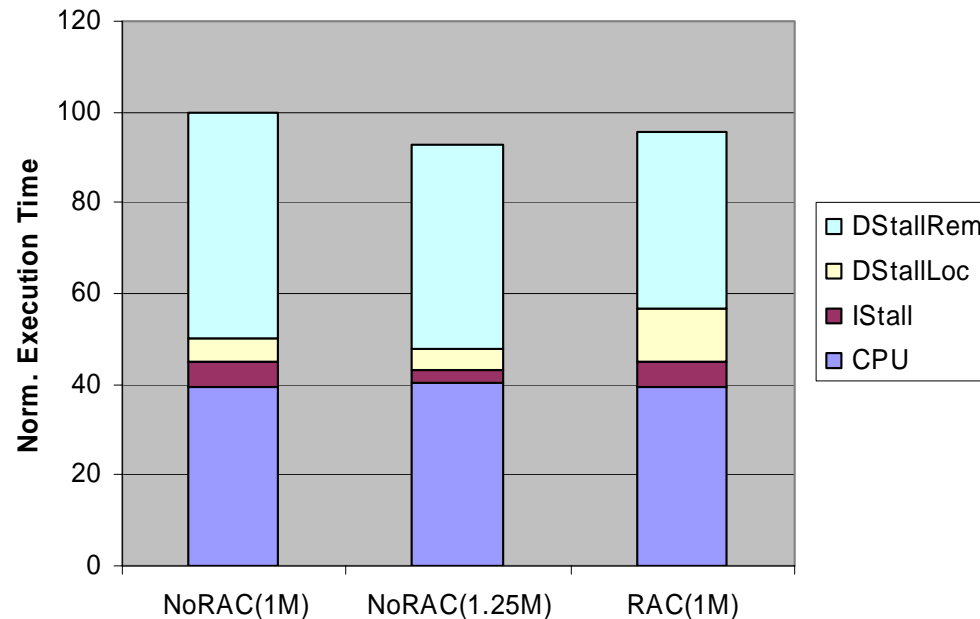
What Is Really Going on



- ◆ RAC reduces remote misses, but increases Dirty misses
- ◆ Dirty and Upgrades cannot be made local
- ◆ Dirty in the RAC costs more in latency (250ns)



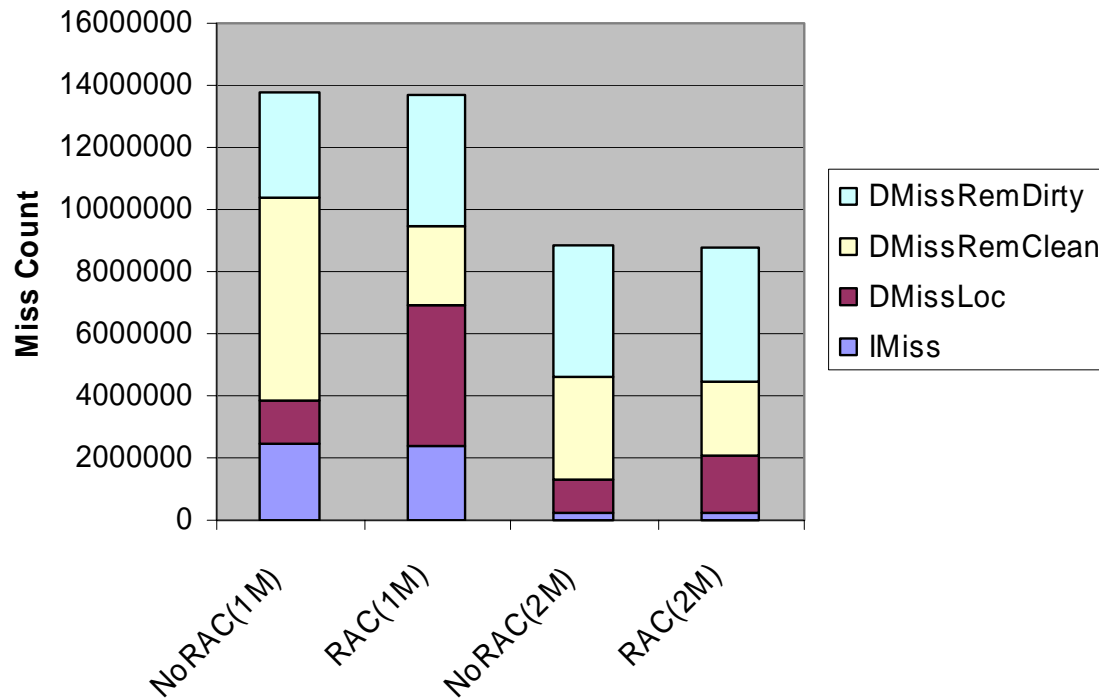
Apples To Apples



- ◆ RAC tags eat up some space on chip
- ◆ 1M L2 with RAC = 1.25M L2 without RAC
- ◆ Performance actually gets worse with a RAC



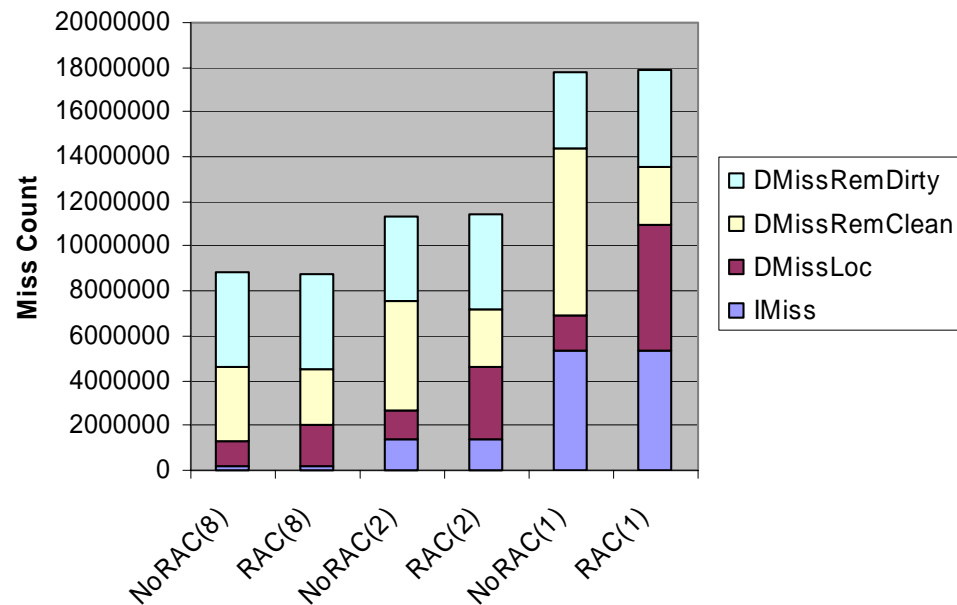
And If the L2 is Larger



- ◆ A larger L2 (2M) makes the RAC ineffective
- ◆ RAC hit rate 9.7%



An Ounce of Associativity is Worth a Pound of Capacity



- ◆ L2 miss rate increases with lower associativity
- ◆ The RAC is more effective. Hit rate:
9.7% (8-way), 21.5% (2-way), 42.18% (1-way)



And the Answer Is

- ◆ Processors are moving to high levels of integration with a 1-2MB on-chip L2 cache
- ◆ A RAC can be quite effective in reducing remote misses
- ◆ **But** the increase in dirty misses hurts performance
- ◆ Associativity is critical for making the small on-chip L2 cache effective



www.compaq.com