Architecture Interaction with Databases (II)

Instructor: Josep Torrellas

CS533

Memory System Characterization of Commercial Workloads

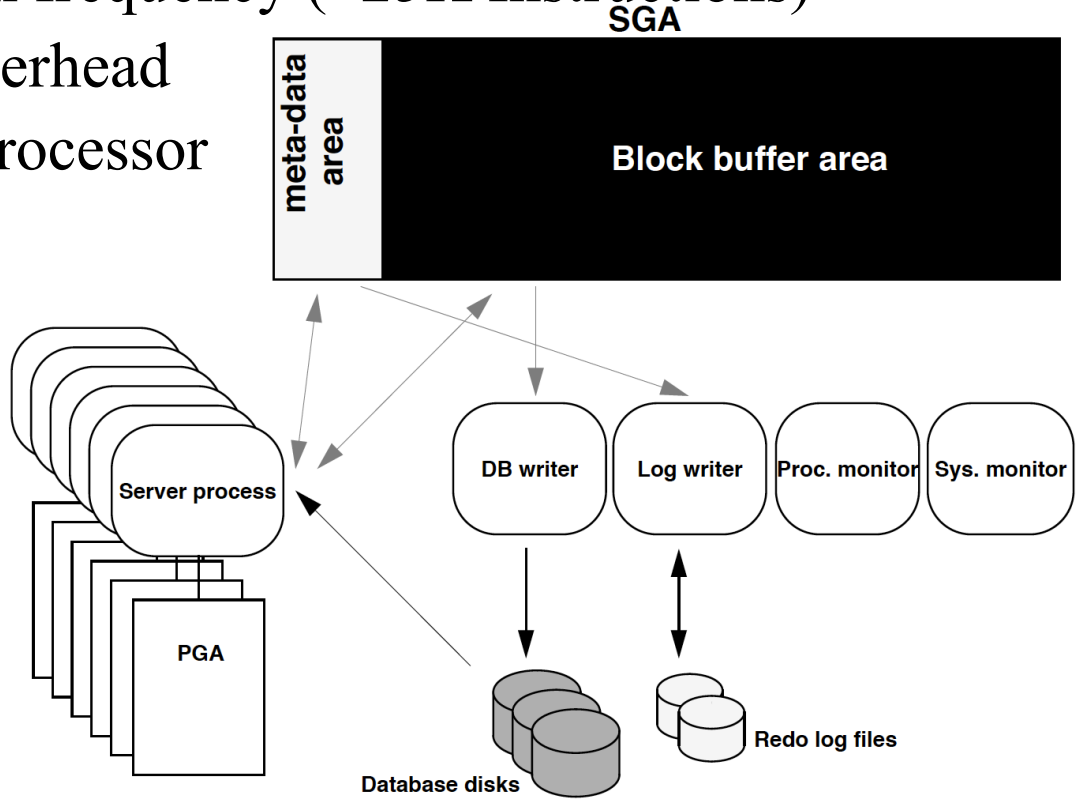
Barroso et al, ISCA-98

Workloads

- OLTP
  - banking system: transactions update balance in randomly-selected account
  - small computation
  - 4 tables updated per transaction --> I/O
- DSS
  - read-only queries
  - more complicated
  - queries can be parallelized
- Web index search
  - similar to DSS
  - read only

Oracle Database Engine

- Shared data in-memory with multiple server processes
- Leverage large process number to hide I/O
- High context switch frequency (~25K Instructions)
- Mask page-fault overhead
- 7-8 processes per processor



DSS Workload

- Decision Support Systems
  - Business analysis
  - Long running SQL queries
  - Spanning a large fraction of the DB
  - Read-only, amenable to intra-query parallelism
- SQL operations:
  - Select, Join, Sort, Aggregate
- Leverage multiple processes 6-7 per processor
  - Parallelize each query across each process
  - Hide IO latency
  - Low Kernel Util

	Tables Used	Selects	Joins	Sorts/ Aggr.	User	Kernel	Idle
Q1	1	1 FS	-	1	94%	2.5%	3.5%
Q4	2	1 FS, 1 I	-	1	86%	4.0%	10%
Q5	6	6 FS	5 HJ	2	84%	4.0%	12%
Q6	1	1 FS	-	1	89%	2.5%	8.5%
Q8	7	8 FS	7 HJ	1	82%	5.0%	13%
Q13	2	1 FS, 1 I	1 NL	1	87%	4.0%	9.0%

Alta Vista: Web Index Search

- Searching the internet before it was Cool! (Google)
- Large search index, 200GB
- The entire DB is memory mapped
- Leverage multiple threads to hide page fault latency
  - During search a thread often gives up the processor
  - Before its quantum due to high page fault rate

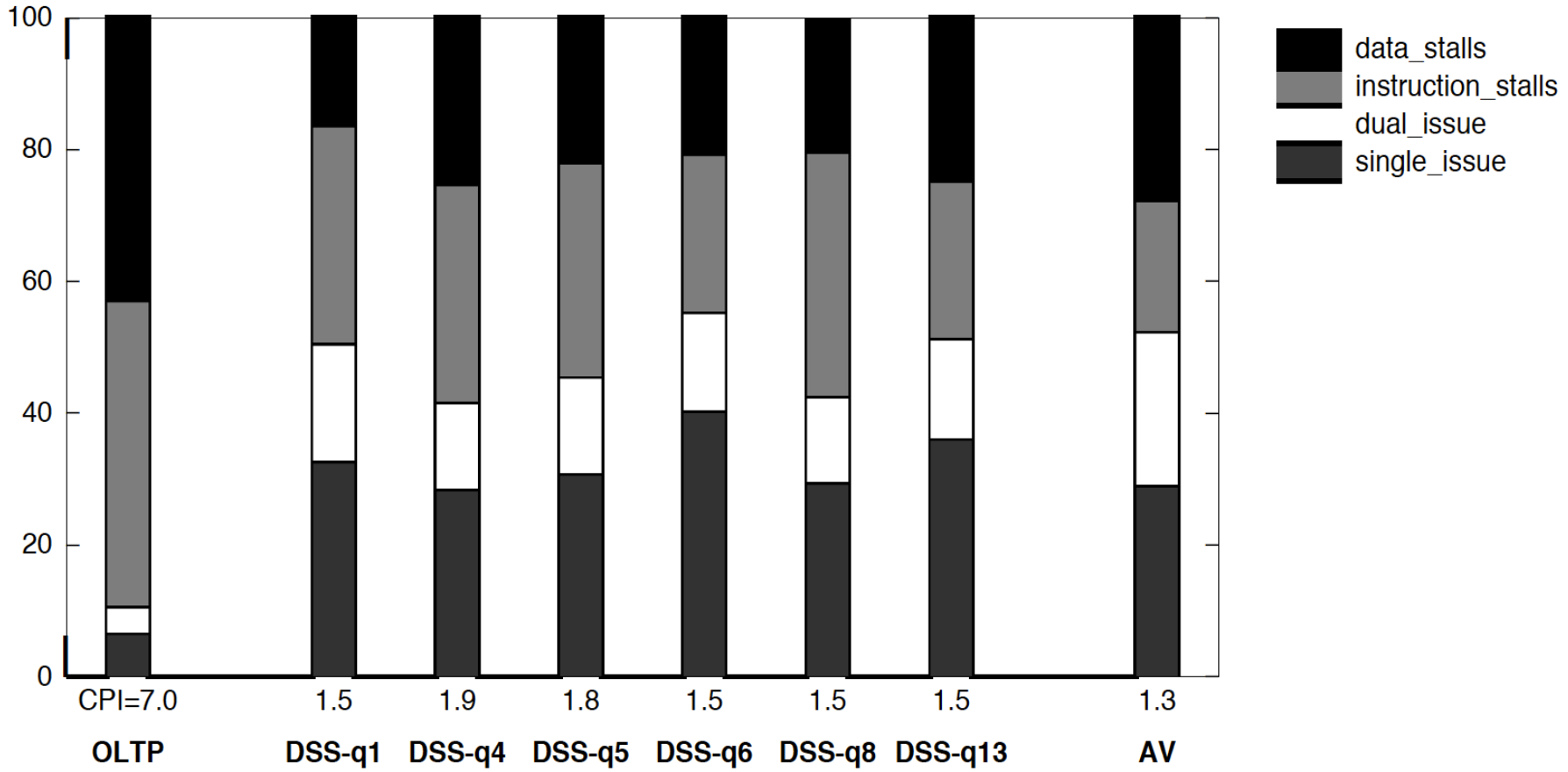
Experiments

- Hardware platform:
  - Server with 4 300-MHz processors
  - Caches: L1: 8K, L2: 96K, L3: 2MB
  - latency to mem: 260ns
  - hardware event counters
- Simulator platform:
  - SimOS: simulates app+os

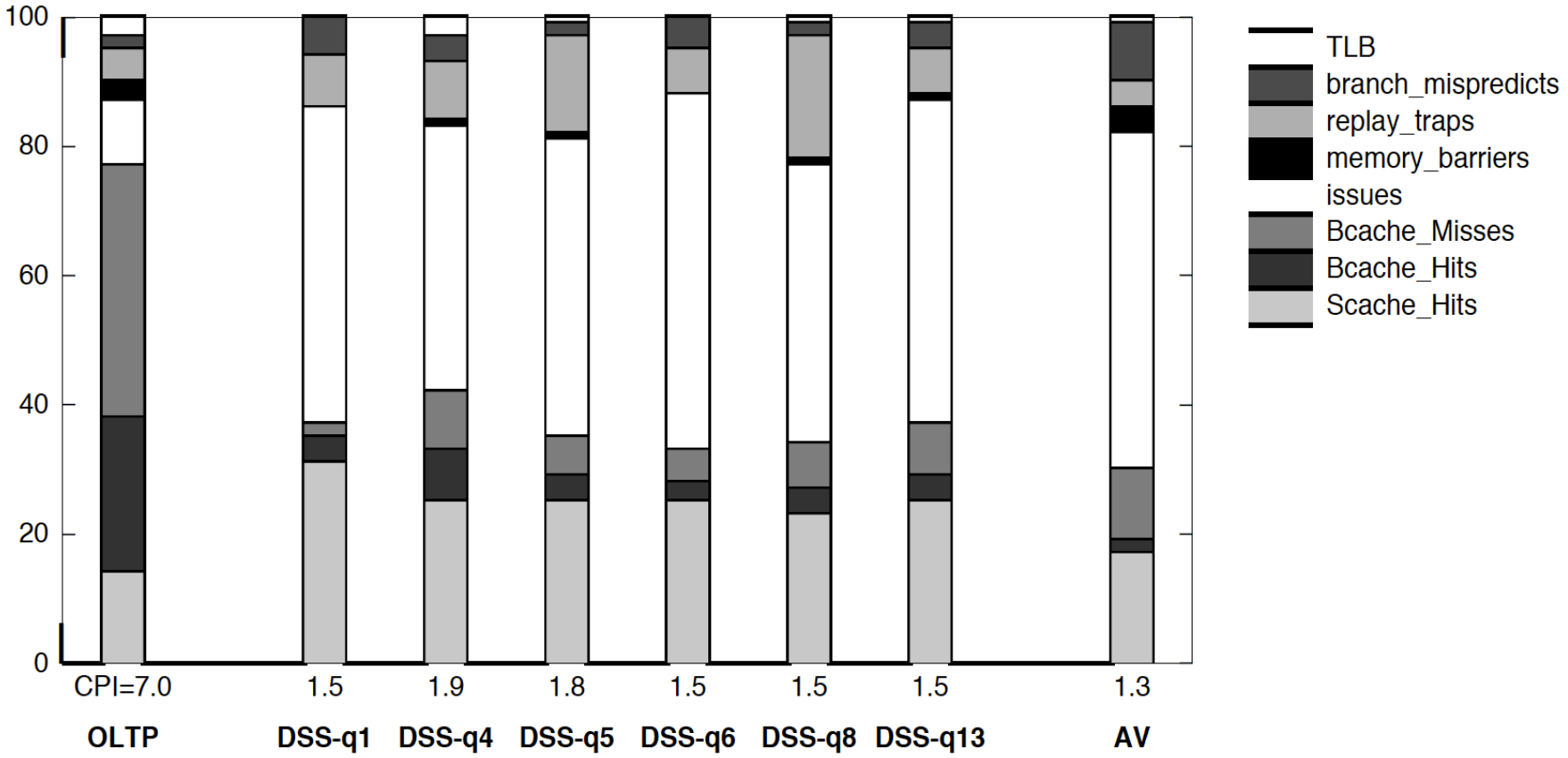
Monitoring Results

- Breakdown of execution time (Figs 3 and 4):
  - CPI of OLTP is very high --> poor performance
  - OLTP:
    - L3 misses (see table 2): workload overwhelms caches
    - Also important: L1 and L2 misses (that hit in L3)
    - dirty misses (data fetched from another cache): 15%, which are slower:
      - 417ns to get data from another cache vs 267ns to get data from memory
      - dirty misses increase with the number of processors and L3 size
  - DSS:
    - More efficient: lower CPI
    - L1 misses, since L1 not fully successful at keeping the working set
    - L2 can hold the state
    - no dirty misses
  - Altavista: best performance

Cycle Breakdown



Detailed Cycle Breakdown



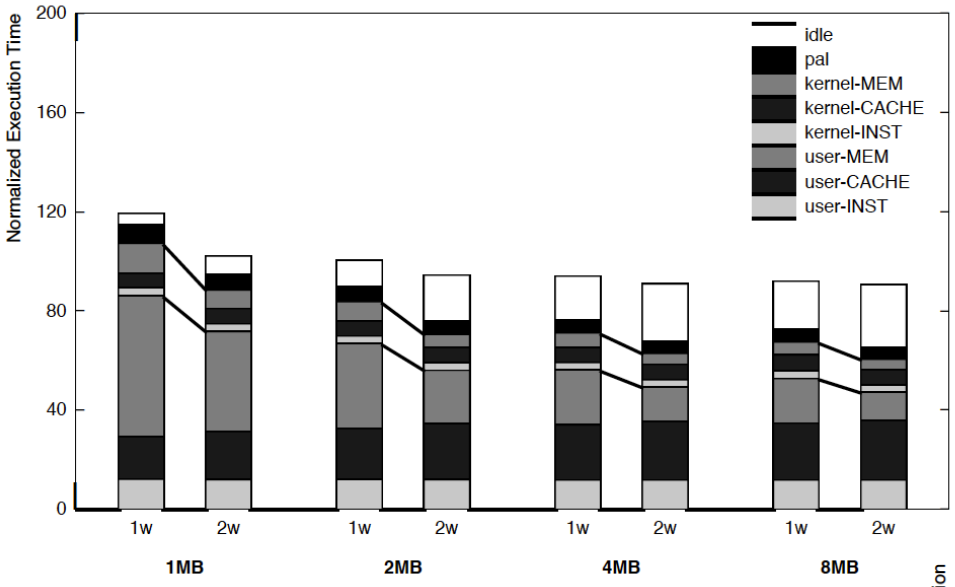
Cache Characterization

	OLTP	DSS-Q1	DSS-Q4	DSS-Q5	DSS-Q6	DSS-Q8	DSS-Q13	AltaVista
Icache (global)	19.9%	9.7%	8.5%	4.6%	5.9%	3.7%	6.7%	1.8%
Dcache (global)	42.5%	6.9%	22.9%	11.9%	11.3%	11.0%	12.4%	7.6%
Scache (global)	13.9%	0.8%	2.3%	1.0%	0.6%	1.0%	1.0%	0.7%
Bcache (global)	2.7%	0.1%	0.5%	0.2%	0.2%	0.3%	0.3%	0.3%
Scache (local)	40.8%	3.6%	10.7%	5.7%	3.9%	6.0%	6.1%	7.6%
Bcache (local)	19.1%	13.0%	21.3%	23.9%	30.7%	27.9%	31.3%	41.2%
Dirty miss fraction	15.5%	2.3%	2.2%	10.6%	1.7%	8.4%	3.3%	15.8%

Simulation Results

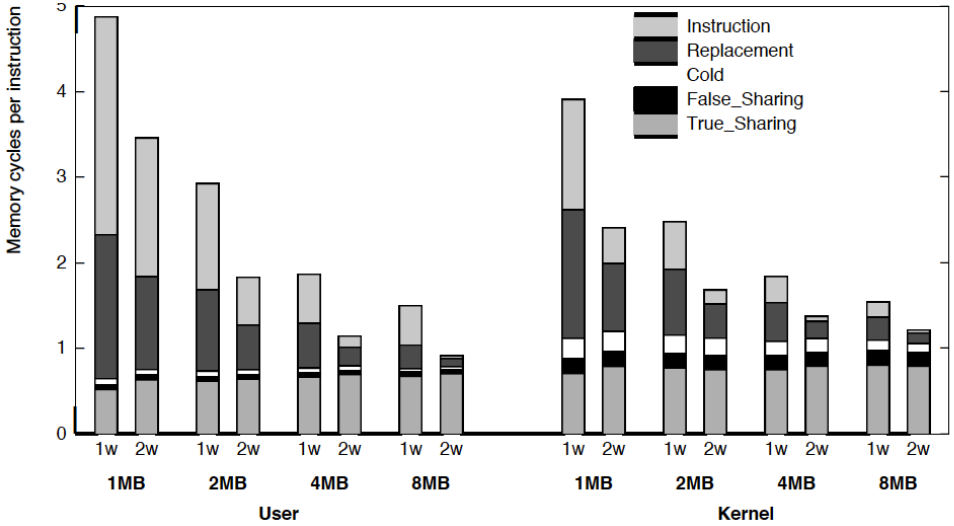
- L1-D and L1-I 32 KB caches
- unified 2 MB L2

Sharing Patterns



- Larger caches more idle time
- IO cannot be hidden
- What do we do with it?

- Communication misses dominate
- Small number of False sharing

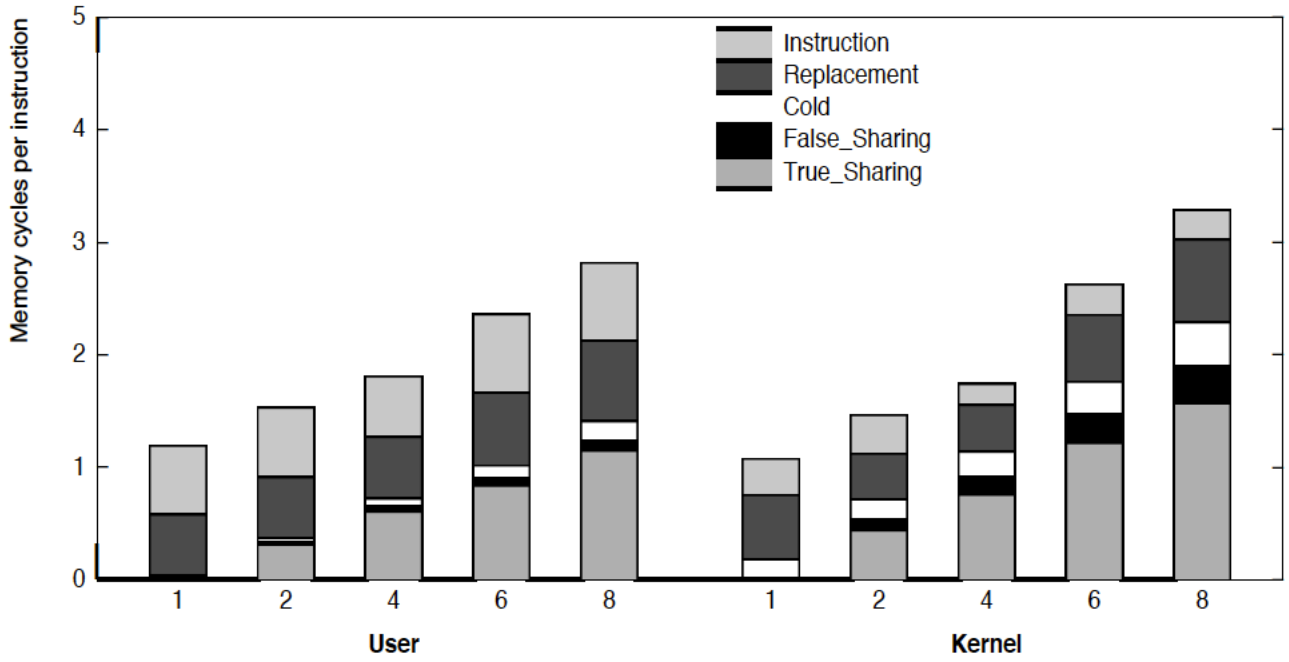


Simulation Results

- Sharing Patterns of OLTP (fig 5)
  - User time dominates
  - benefits from larger/more assoc caches
  - Cache and memory stall still important
  - L2 cache behavior for different cache sizes/assoc (Fig 5)
  - Most communication misses due to true (not F) sharing
  - Analyzing the number processors (Fig 6). IF P goes up:
    - ratio of true to false sharing does not change.
    - Overall: communication misses

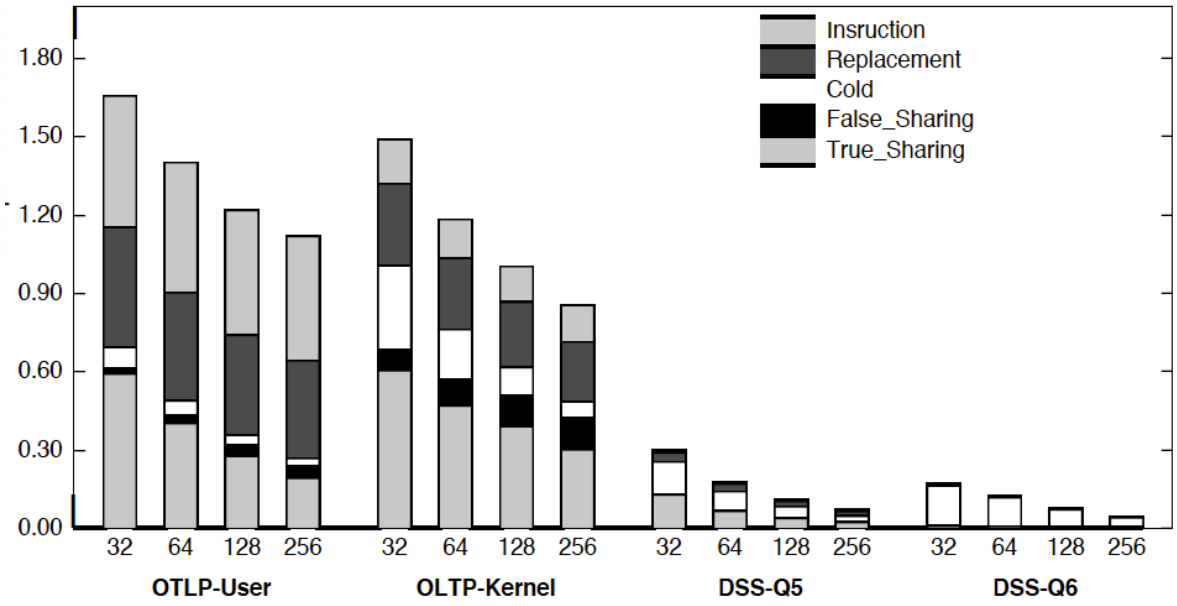
Cache Sizes

- Larger on-chip caches good for DSS and OLTP (Fig 6)
- Larger on-chip caches cache most misses in DSS
- OLTP has large footprint: continue to benefit with 4-8MB L



Sensitivity to Cache Line Size

- Changing line size of the L3 (Fig 8)
  - data communicated among processors (true sharing) has good spatial locality
  - Cold misses also spatial locality
  - No impact on replacement misses



Summary

- OS and I/O do not dominate tuned DB
- OLTP:
  - I and D locality only captured with large L3 caches
  - high communication miss rate (dirty misses)
- DSS and AltaVista:
  - Sensitive to the size and latency of L1
  - Less sensitive to off chip caches sizes/latencies
- Workloads require different server designs