Stavros Harizopoulos

MIT Computer Science and
Artificial Intelligence Laboratory
G930, The Stata Center
32 Vassar Street
Cambridge, MA 02139, USA

Office: +1 (617) 253-8236
Fax: +1 (617) 258-8607
Mobile: +1 (412) 726-7219
stavros@csail.mit.edu
http://nms.csail.mit.edu/~stavros

Research Interests

Design, implementation, and performance of software systems with an emphasis in database management systems; storage systems; scheduling algorithms; performance modeling; hardware/software co-design; thread-parallel architectures and distributed systems.

Education

Carnegie Mellon University, *Pittsburgh, PA*
Ph.D. in Computer Science (December 2005)
Thesis: Staged Database Systems
Advisor: Anastassia Ailamaki
Committee: Panos Chrysanthis, Christos Faloutsos, Todd Mowry, and Michael Stonebraker
M.S. in Computer Science (December 2000)

Technical University of Crete, *Chania, Greece*
5-year Diploma in Electronic and Computer Engineering (August 1998)
Thesis: Hierarchical Caching and Prefetching for Improving Continuous-Media Server Performance
Advisor: Peter Triantafillou

Work Experience

Massachusetts Institute of Technology, *Cambridge, MA*
October 2005 - Present
Post-doctoral Research Associate (CSAIL Visiting Scientist)

IBM Almaden Research Center, *San Jose, CA*
Summer 2000
Research Internship in Storage Systems

Publications

Conference Proceedings (acceptance rate < 15%)

[16] Stavros Harizopoulos, Velen Liang, Daniel Abadi, and Samuel Madden. "Performance Tradeoffs in Read-Optimized Databases." In *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB)*, Seoul, Korea, September 2006.

[15] Stavros Harizopoulos, Vladislav Shkapenyuk, and Anastassia Ailamaki. "QPipe: A Simultaneously Pipelined Relational Query Engine." In *Proceedings of the 24th ACM SIGMOD International Conference on Management of Data*, Baltimore, Maryland, June 2005.

[14] Stavros Harizopoulos and Anastassia Ailamaki. "STEPS Towards Cache-Resident Transaction Processing." In *Proceedings of the 30th International Conference on Very Large Data Bases (VLDB)*, Toronto, Canada, September 2004.

Conference Proceedings (acceptance rate > 15%)

[13] Michael Stonebraker, Chuck Bear, Ugur Çetintemel, Mitch Cherniack, Tingjian Ge, Nabil Hachem, Stavros Harizopoulos, John Lifter, Jennie Rogers, and Stan Zdonik. "One Size Fits All? - Part 2: Benchmarking Results." In *Proceedings of the Third International Conference on Innovative Data Systems Research (CIDR)*, Asilomar, CA, January 2007.

[12] Stavros Harizopoulos and Anastassia Ailamaki. "A Case for Staged Database Systems." In *Proceedings of the First International Conference on Innovative Data Systems Research (CIDR)*, Asilomar, CA, January 2003.

[11] Peter Triantafillou and Stavros Harizopoulos. "Prefetching into Smart-Disk Caches for High Performance Media Servers." In *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, Florence, Italy, June 1999.

Journal Articles

[10] Stavros Harizopoulos and Anastassia Ailamaki. "Improving Instruction Cache Performance in OLTP." In *ACM Transactions on Database Systems (TODS),* Vol. 31(3): 887-920, September 2006.

[9] Stavros Harizopoulos and Anastassia Ailamaki. "StagedDB: Designing Database Servers for Modern Hardware." In *IEEE Data Engineering Bulletin*, Vol. 28(2): 11-16, June 2005 (invited article).

[8] Stavros Harizopoulos, Costas Harizakis, and Peter Triantafillou. "Hierarchical Caching and Prefetching for Continuous Media Servers with Smart Disks." In *IEEE Concurrency*, Vol. 8(3): 16-22, July / September 2000.

Tutorials and Demonstrations

[7] Anastassia Ailamaki, Naga K. Govindaraju, Stavros Harizopoulos, and Dinesh Manocha. "Query Co-Processing on Commodity Processors." Tutorial, in *Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB)*, Seoul, Korea, September 2006.

[6] Kun Gao, Stavros Harizopoulos, Ippokratis Pandis, Vladislav Shkapenyuk, and Anastassia Ailamaki. "Simultaneous Pipelining in QPipe: Exploiting Work Sharing Opportunities Across Queries." Demonstration, *ICDE'06: Proceedings of the 22nd International Conference on Data Engineering,* Atlanta, GA, April 2006. **Best demo award**.

Technical Reports and Theses

[5] Nikos Hardavellas, Ippokratis Pandis, Ryan Johnson, Naju G. Mancheril, Stavros Harizopoulos, Anastassia Ailamaki, and Babak Falsafi. "Database Servers on Chip Multiprocessors: Limitations and Opportunities." *Carnegie Mellon University, CMU-CS-06-153*, September 2006.

[4] Stavros Harizopoulos. "Staged Database Systems." Ph.D. Thesis, *Carnegie Mellon University, CMU-CS-05-194*, December 2005.

[3] Vladislav Shkapenyuk, Ryan Williams, Stavros Harizopoulos, Anastassia Ailamaki. "Deadlock Resolution in Pipelined Query Graphs." *Carnegie Mellon University, CMU-CS-05-122*, March 2005.

[2] Stavros Harizopoulos and Anastassia Ailamaki. "Affinity scheduling in staged server architectures." *Carnegie Mellon University, CMU-CS-02-113*, March 2002.

[1] Stavros Harizopoulos and Garth A. Gibson. "PASTENSE: a Fast Start-up Algorithm for Scalable Video Libraries." *Carnegie Mellon University, CMU-CS-01-105*, March 2001.

RESEARCH EXPERIENCE

Massachusetts Institute of Technology

At MIT, I work primarily on the C-Store research project. C-Store is a read-optimized relational database system that contrasts sharply with most current database systems, which are write-optimized. Among the many differences in its design are: storage of data by column rather than by row, careful coding and packing of objects into storage including main memory during query processing, and storing an overlapping collection of column-oriented projections, rather than a set of tables and indexes. The goal of my work has been to understand the key differences between column-based and row-based databases. To that end, I conducted a performance study between both types of systems. To ensure a meaningful and unbiased comparison, I built both database engines from scratch and isolated the parameters that significantly affect performance. The description of the design and implementation of both systems, along with the methodology and results of the study, and an analytical model that quantifies the performance differences between the two architectures are published in VLDB'06 [16].

Currently, I am investigating database system designs for efficiently accommodating online inserts in data warehouse environments employing a read-optimized database. This is a significant problem, as users of data warehouses are demanding warehouse queries over increasingly fresh data. Since a read-optimized system cannot constantly reorganize data on disk, I am investigating possible solutions that include a main-memory staging area to temporarily store incoming data. My approach is the following: identify and categorize all possible designs, provide an analytical model to pick the most promising ones, and implement all basic system components necessary to produce a quantitative performance analysis of all tradeoffs in the proposed designs. I am also part of an effort from three Universities (MIT, Brown, and Brandeis) to build a novel software system that natively supports large multi-dimensional arrays, suitable for high-performance scientific and engineering applications (see publication [13] for a preliminary system evaluation).

2001 - 2005 *Thesis Research, Carnegie Mellon University*

My Ph.D. thesis, titled "Staged Database Systems," [4] introduced ways to re-engineer the design of database management systems to improve utilization of all levels of the memory hierarchy in modern processor architectures, both for instructions and data. Rather than rewriting the entire code of the database system, my thesis work provided the support to organize system components into "stages" and to change the request execution sequence to optimally perform group-processing at each stage. A paper describing the system design was published in CIDR'03 [12] (see also publications [2, 5, 9]).

At the microarchitectural level, my thesis addressed instruction-stream optimizations in transaction processing with a technique called *STEPS*. Research has shown that instruction-related delays in the memory subsystem account for 25% to 40% of the execution time in Online Transaction Processing (OLTP) workloads [7]. STEPS (Synchronized Transactions through Explicit Processor Scheduling) minimizes instruction cache misses by multiplexing concurrent transactions and exploiting common code paths. STEPS yields up to a 96% reduction in instruction cache misses. The results of evaluating STEPS inside the Shore storage manager, including a full-system evaluation on real hardware using TPC-C, the industry-standard transactional benchmark, were published in VLDB'04 [14] (additional techniques and results appear in [10]).

At the lower levels of the memory hierarchy, my thesis work enhanced data locality for concurrent database queries with *QPipe*, a staged, operator-centric query execution engine. QPipe proactively coordinates same-operator execution among concurrent queries, thereby exploiting common accesses to memory and disks, as well as common intermediate result computations. Evaluation of QPipe on top of BerkeleyDB showed that QPipe achieved a 2x speedup over a commercial DBMS when running a workload consisting of TPC-H queries (the industry-standard decision-support benchmark). The results were published in SIGMOD'05 [15] (see also [3]) and the system was demonstrated in ICDE'06 [6] (best demonstration award).

Summer 2000 *IBM Almaden Research Center*

While being an intern at IBM, I worked on the following problem: "Given an inexpensive collection of commodity IDE disk drives of various sizes, capacities, and rotational speeds, can we design a storage subsystem that offers better value per IOPS (I/O per sec) than a high performance system that uses SCSI drives, if we are willing to sacrifice capacity for disk throughput?" We examined inter- and intra-disk data replication techniques for increasing the system's throughput, and constructed benchmarks to characterize IDE drive behavior.

1998 - 2000 *Parallel Data Lab, Carnegie Mellon University*

With Professor Garth A. Gibson we worked on the design and implementation of a fast start-up algorithm for a low-cost, scalable video server built on Network Attached Secure Disks. The NASD video server stripes data over many disks to balance load with less data replication. Striped video algorithms can have high start-up latency, if the load is high. Our technique uses aggressive prefetching to exploit disk idle time, and available RAM to dynamically optimize the video delivery schedule. The algorithm does not require changes in data placement and achieves up to 50-75% faster start-up times [1].

1997 - 1998 *Technical University of Crete*

With Professor Peter Triantafillou we explored caching and prefetching techniques for improving the throughput of disk-based video servers. Based on the predictability of continuous media requests, we proposed video retrieval algorithms that utilize drive-level caches to improve system throughput [8, 11]. We evaluated the proposed techniques using the DiskSim disk simulator.

TEACHING EXPERIENCE

Fall 2003, Carnegie Mellon University

Teaching Assistant, **15-213 Introduction to Computer Systems** (2nd-year, class size: 120)

- Prepared and delivered weekly recitations to a group of 25 students (recitation assessment: "excellent").
- Designed, administered, and graded a lab assignment.
- Assisted in preparing and grading two exams.
- Held weekly office hours and answered student questions on class newsgroup.

Fall 2001, Carnegie Mellon University

Teaching Assistant, **15-721 Database Management Systems** (graduate course, class size: 25)

- Assisted in preparing and grading a semester-long class project which involved implementing the basic functionality of a database management system.
- Assisted in preparing and grading two exams.
- Held weekly office hours and answered student questions on class newsgroup.

Spring 2000, Carnegie Mellon University

Teaching Assistant, **15-415 Database Applications** (4th-year, class size: 70)
- Assisted in preparing and grading assignments, class project, and two exams.
- Created and delivered a lecture on indexing methods.
- Held weekly office hours and answered student questions on class newsgroup.

ACADEMIC SERVICE

Program Committees and Conference Proceedings Reviews
HDMS 2006 Program Committee member
SIGMOD 2005 Demonstrations Program Committee and Judging Committee member
VLDB'05, SIGMOD'05, DaMoN'05, CIDR'05, VLDB'04, CIDR'03, SIGMOD'03 (Reviewer)

Organizer
Weekly Database Seminar, Carnegie Mellon University. Fall 2003, Spring 2003, Summer 2003

HONORS

Best Demonstration Award, International Conference on Data Engineering (ICDE), 2006
IBM Ph.D. Fellowship, 2002-2003
Lilian Voudouri Foundation Ph.D. Fellowship, 1999-2002
Technical Chamber of Greece Merit Scholarship and Best Polytechnic Student Award, 1998
Greek National Scholarship Foundation (I.K.Y.), 1993,1997
Technical University of Crete Merit Scholarship, 1993, 1996, 1997
Panhellenic University Entrance Exams: 1st in Math and Physics (among 40,000), 1993
Panhellenic Olympiad for Mathematics: Bronze medal (third place), 1990, 1993

REFERENCES

Anastassia Ailamaki
Computer Science Department
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213, USA
+1 (412) 268-7848
natassa@cs.cmu.edu

Christos Faloutsos
Computer Science Department
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213, USA
+1 (412) 268-1457
christos@cs.cmu.edu

Samuel Madden
MIT Computer Science and
Artificial Intelligence Laboratory
32 Vassar Street
Cambridge, MA 02139, USA
+1 (617) 258-6643
madden@csail.mit.edu

Michael Stonebraker
MIT Computer Science and
Artificial Intelligence Laboratory
32 Vassar Street
Cambridge, MA 02139, USA
+1 (617) 253-3538
stonebraker@csail.mit.edu