Cache-Oblivious and Cache-Aware Algorithms

Michael A. Bender*

A recent direction in algorithmic design and analysis is to pay particular attention to the structure of the memory hierarchy. Because memory hierarchies are growing "steeper," the performance impact of the memory system is increasing dramatically. Many approaches and models have been developed for designing algorithms that interact well with the memory system, of which one of the more successful encompasses the body of work on *external-memory algorithms*, which effectively models two-level memory hierarchies (see, e.g., [AV88, Vit01]). Recently, the concept of *cache-oblivious algorithms* was proposed as a powerful theoretical approach with the potential for substantial practical impact [FLPR99, Pro99]. The idea is to hide any parameters of the memory hierarchy—such as block transfer sizes and the size of each memory level—from the algorithm. This simple idea has powerful ramifications:

1. Cache-oblivious algorithms automatically adapt to arbitrary memory hierarchies.
2. Cache-oblivious algorithms can be analyzed on a simple two-level memory hierarchy, and then automatically perform as well on a complex multilevel memory hierarchy with particular page replacement strategies, limited associativity, etc.

Motivated by these exciting consequences, an increasing number of researchers have started to develop algorithms and data structures in this model. In the past 3 years, there have been over fifteen papers developing efficient cache-oblivious algorithms and data structures. The field has been shown to have substantial depth and has led to new general techniques for maintaining data locality in a memory hierarchy. Yet the field of cache-oblivious algorithms is still in its infancy. Practical and theoretical issues remain unsolved and ripe for exploration, bridging the gap between the best algorithms known in the external-memory and cache-oblivious contexts.

This talk gives a overview of cache-oblivious techniques and algorithms. Particular focus will be given to cache-oblivious matrix multiplication [FLPR99, Pro99], sorting [FLPR99, Pro99, BF02a], and search-trees [BDIW02, BFJ02, BDFC00, RCR01, BCR02], as well to to some of the open questions in the area.

*Department of Computer Science, SUNY Stony Brook, Stony Brook, NY 11794-4400, USA; bender@cs.sunysb.edu.

References

[AV88] Alok Aggarwal and Jeffrey Scott Vitter. The input/output complexity of sorting and related problems. *Communications of the ACM*, 31(9):1116–1127, September 1988.

[BBF+03] Michael A. Bender, Gerth Stølting Brodal, Rolf Fagerberg, Dongdong Ge, Simai He, Haodong Hu, John Iacono, and Alejandro López-Ortiz. The cost of cache-oblivious searching. In *Proceedings of the 44st Annual Symposium on Foundations of Computer Science*, pages 271–282, Cambridge, Massachusetts, October 2003.

[BCR02] Michael A. Bender, Richard Cole, and Rajeev Raman. Exponential structures for efficient cache-oblivious algorithms. In *Proceedings of the 29th International Colloquium on Automata, Languages and Programming*, volume 2380 of *Lecture Notes in Computer Science*, pages 195–207, Málaga, Spain, July 2002.

[BDFC00] Michael A. Bender, Erik D. Demaine, and Martin Farach-Colton. Cache-oblivious B-trees. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pages 399–409, Redondo Beach, California, November 2000.

[BDIW02] Michael A. Bender, Ziyang Duan, John Iacono, and Jing Wu. A locality-preserving cache-oblivious dynamic dictionary. In *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 29–38, San Francisco, California, January 2002.

[BF02a] Gerth Stølting Brodal and Rolf Fagerberg. Cache oblivious distribution sweeping. In *Proceedings of the 29th International Colloquium on Automata, Languages, and Programming*, volume 2380 of *Lecture Notes in Computer Science*, pages 426–438, Malaga, Spain, July 2002.

[BFJ02] Gerth Stølting Brodal, Rolf Fagerberg, and Riko Jacob. Cache oblivious search trees via binary trees of small height. In *Proceedings of the 13th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 39–48, San Francisco, California, January 2002.

[FLPR99] Matteo Frigo, Charles E. Leiserson, Harald Prokop, and Sridhar Ramachandran. Cache-oblivious algorithms. In *Proceedings of the 40th Annual Symposium on Foundations of Computer Science*, pages 285–297, New York, October 1999.

[Pro99] Harald Prokop. Cache-oblivious algorithms. Master's thesis, Massachusetts Institute of Technology, Cambridge, MA, June 1999.

[RCR01] Naila Rahman, Richard Cole, and Rajeev Raman. Optimised predecessor data structures for internal memory. In *Proceedings of the 5th International Workshop on Algorithm Engineering*, volume 2141 of *Lecture Notes in Computer Science*, pages 67–78, Aarhus, Denmark, August 2001.

[Vit01] Jeffrey Scott Vitter. External memory algorithms and data structures: Dealing with massive data. *ACM Computing Surveys*, 33(2):209–271, June 2001.