

## Memory Constrained Data Locality Optimization for Tensor Contractions

Alina Bibireata Sandhya Krishnan Gerald Baumgartner Daniel Cociorva  
Chi-Chung Lam P. Sadayappan J. Ramanujam David E. Bernholdt Venkatesh  
Choppella

To appear at the *16th Workshop on Languages and Compilers for Parallel Computing (LCPC03)*, College Station, TX, 2-4 October 2003

### Abstract

The accurate modeling of the electronic structure of atoms and molecules involves computationally intensive tensor contractions over large multi-dimensional arrays. Efficient computation of these contractions usually requires the generation of temporary intermediate arrays. These intermediates could be extremely large, requiring their storage on disk. However, the intermediates can often be generated and used in batches through appropriate loop fusion transformations. To optimize the performance of such computations a combination of loop fusion and loop tiling is required, so that the cost of disk I/O is minimized. In this paper, we address the memory-constrained data-locality optimization problem in the context of this class of computations. We develop an optimization framework to search among a space of fusion and tiling choices to minimize the data movement overhead. The effectiveness of the developed optimization approach is demonstrated on a computation representative of a component used in quantum chemistry suites.