

**Efficient Execution of Multi-Query Data Analysis Batches Using
Compiler Optimization Strategies**

Henrique Andrade Suresh Aryangat Tahsin Kurc Joel Saltz Alan Sussman
To appear at the *16th Workshop on Languages and Compilers for Parallel
Computing (LCPC03)*, College Station, TX, 2-4 October 2003

Abstract

This work investigates the leverage that can be obtained from compiler optimization techniques for efficient execution of multi-query workloads in data analysis applications. Our approach is to address multi-query optimization at the algorithmic level, by transforming a declarative specification of scientific data analysis queries into a high-level imperative program that can be made more efficient by applying compiler optimization techniques. These techniques – including loop fusion, common subexpression elimination and dead code elimination – are employed to allow data and computation reuse across queries. We describe a preliminary experimental analysis on a real remote sensing application that analyzes very large quantities of satellite data. The results show our techniques achieve sizable reductions in the amount of computation and I/O necessary for executing query batches and in average execution times for the individual queries in a given batch.