
Is Acyclic Directed Graph Partitioning Effective for Locality-Aware Scheduling?

M. Yusuf Özkaya¹, Anne Benoit^{1,2}, Ümit Çatalyurek¹

¹CSE, Georgia Institute of Technology, Atlanta, GA, USA

²LIP, ENS Lyon, France

{myozka, umit}@gatech.edu, anne.benoit@ens-lyon.fr

We investigate efficient execution of computations, modeled as Directed Acyclic Graphs (DAGs), on a single processor with a two-level memory hierarchy, where there is a limited fast memory and a larger slower memory. Our goal is to minimize execution time by minimizing redundant data movement between fast and slow memory. We utilize a DAG partitioner that finds localized, acyclic parts of the whole computation that can fit into fast memory, and minimizes the edge cut among the parts. We propose a new scheduler that executes each part one-by-one, obeying the dependency among parts, aiming at reducing redundant data movement needed by cut-edges. Extensive experimental evaluation shows that the proposed DAG-based scheduler significantly reduces redundant data movement.

Keywords: two-level memory hierarchy, cache misses minimization, DAG partitioner.