A Hierarchical Model of Reference Affinity

Yutao Zhong, Xipeng Shen and Chen Ding

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

Abstract

To improve performance, data reorganization needs locality models to identify groups of data that have reference affinity. Much past work is based on access frequency and does not consider accessing time directly. In this paper, we propose a new model of reference affinity. This model considers the distance between data accesses in addition to the frequency. Affinity groups defined by this model are consistent and have a hierarchical structure. The former property ensures the profitability of data packing, while the latter supports data packing for storage units of different sizes. We then present a statistical clustering method that identifies affinity groups among structure fields and data arrays by analyzing training runs of a program. When used by structure splitting and array regrouping, the new method improves the performance of two test programs by up to 31%. The new data layout is significantly better than that produced by the programmer or by static compiler analysis.