Optimizing Large Data Structures with an Unpredictable Access Pattern in the Intel KNL Processor

Jose M. Herruzo, Sonia Gonzalez-Navarro, and Oscar Plata

Department of Computer Architecture, University of Málaga, 29071 Málaga, Spain,
{jmherruzo,sgn,oplata}@uma.es

Abstract. The continuous increase of data-intensive workloads that we are experiencing in recent times has led to a great interest in the design of techniques for the efficient processing of huge amounts of data. From the data storage viewpoint, performance is boosted if the data is properly organized in main memory and the reference patterns exploit data access locality. However, handling a large amount of data becomes a hard challenge when unpredictable access patterns are present. Unfortunately, the large and deep cache hierarchies found in modern processors work inefficiently with such workloads, and frequently cause a high demand of memory bandwidth. This results in a severe performance drop. This paper tackles this challenge by re-organizing the data structures, specifically, re-arranging data layout and changing data codification, with the aim of minimizing memory bandwidth demand.

The FM-index data structure is used as a case study throughout the paper. FM-index combines compression of a text and indexing information in a single data structure, allowing fast and efficient pattern searching in large reference texts. However, due to compression and data layout, searching algorithms based on FM-index exhibit non-uniform memory access patterns that cause frequent cache misses. This paper presents a new organization of the FM-index data structure capable of drastically reducing memory bandwidth requirements for the searching process. As a result, the operational intensity of the search algorithm increases significantly, offering the opportunity to take better advantage of the available memory bandwidth.

The proposal has been implemented and evaluated in an Intel Xeon Phi 7210 processor (KNL) that includes 64 cores and 16 GB of ultra high-bandwidth stacked 3D MCDRAM integrated on package, able to provide a peak bandwidth of 400 GB/s. The experimental results show that performance of the searching algorithm on KNL reaches about 95% of that limited by the peak random access bandwidth of MCDRAM, outperforming other CPU and GPU solutions previously reported in the literature.