Journal of Instruction-Level Parallelism 13 (2011) 1-2 Submitted 1/11; published 1/11

Papers from the First Data Prefetching Championship

Alaa R. Alameldeen ALAA.R.ALAMELDEEN@INTEL.COM
Intel Labs
Intel Corporation
Hillsboro, OR 97124 USA

1. Introduction

This special issue consists of extended papers from the First Data Prefetching Championship (DPC-1). Earlier versions of these papers were published at a workshop held in conjunction with the 15th Annual International Conference on High-Performance Computer Architecture (HPCA-15) in Raleigh, NC, in February 2009. I invited authors from the workshop to extend their papers for this special issue of JILP.

2. About the Competition

The tremendous increase in processor frequency over the last three decades has made memory latency a dominant factor influencing system performance. Even large caches incur many cold, capacity, and conflict misses and thereby cannot completely bridge the gap between processor and memory speeds. To improve performance, data prefetching techniques have emerged to significantly reduce the average memory access latency. Prefetchers can help performance by fetching blocks into the cache before they are requested. On the other hand, prefetchers can hurt performance by replacing useful cache blocks, or by saturating cache and memory bandwidth. In recent years, there has been great advancements in prefetching techniques. We are optimistic that even more innovative prefetchers are on the horizon.

The purpose of the Data Prefetching Championship (DPC-1) was to provide a dedicated venue for computer architecture researchers to put their latest prefetching ideas and implementations to the test. The spirit of competition is intended to provide incentive for students, faculty, and industry to push the envelope of prefetching. Contestants competed for the top honor of best-performing prefetcher in the context of a common simulation infrastructure, common workloads, a fixed storage budget, and diverse cache and bandwidth scenarios.

Contestants submitted source code and a paper describing their prefetcher. Their prefetchers were tested by the DPC-1 organizers on the common simulation infrastructure and ranked by a summary performance metric. Separately and without knowledge of the performance ranking, the papers were reviewed by the program committee to ensure a quality program and understand the contestants' prefetchers. Each contestant was allowed a maximum of three submissions, which can be totally different techniques, or can be variations on the same prefetcher.

The competition has been a great success. We received a total of thirteen papers and twenty code submissions. In the workshop with HPCA-15, we invited contestants in the top 8 of overall performance. All source code, papers, and detailed results on are publicly available on the DPC-1 website (http://www.jilp.org/dpc/), to benefit the computer architecture research community. We are continuing the tradition of holding periodic competitions with the establishment of the JILP Workshop on Architecture Competitions (JWAC), to be held annually with ISCA.

3. Contents

The first, second, and seventh papers combine prefetchers from previous proposals and optimize them to improve performance. The first paper, "Combining Local and Global History for High Performance Data Prefetching," exploits global and local strides and identify other data localities to issue accurate prefetches and eliminate redundant ones. The second paper, "Storage Efficient Hardware Prefetching Using Delta-Correlating Prediction Tables," combines local stride prefetching and AC/DC prefetching while reducing storage and complexity requirements. The seventh paper, "Efficient Prefetching with Hybrid Schemes and Use of Program Feedback to Adjust Prefetcher Aggressiveness," combines stride, stream, and AC/DC prefetchers, and collects runtime information to adjust the hybrid prefetcher's aggressiveness.

The third and eighth papers target spatially-correlated cache accesses in a certain memory region. The third paper, "Access Map Pattern Matching for High Performance Data Cache Prefetch," uses a stream prefetcher in the L1 cache and a novel region-based prefetcher for the L2 cache where prefetches are issued based on which pattern they fit best. The eighth paper, "Spatial Memory Streaming," identifies code-correlated spatial access patterns, and prefetches predicted streams to the L1 cache.

The fourth paper, "Enhancements for Accurate and Timely Streaming Prefetcher," proposes several enhancements to stream prefetchers to establish streams more efficiently and eliminate the noise that could affect establishing streams.

The fifth paper, "Multi-level Adaptive Prefetching Based on Performance Gradient Tracking" uses an adaptive L2 prefetching policy that tracks the performance gradient of workload phases and adjusts the prefetching degree accordingly.

The sixth paper, "Data Prefetching by Exploiting Global and Local Access Patterns," augment Nesbit and Smith's global history buffers with local buffers that tracks memory accesses of specific program counters.

Acknowledgments

I thank the JILP Editor-in-Chief, Eric Rotenberg, for allowing me to serve as editor for this special issue, and for his tremendous help in organizing the competition, serving as program committee chair for the workshop, and organizing the workshop with HPCA.

I thank organizers of the data prefetching championship. Aamer Jaleel, Daniel Luchi, and Zeshan Chishti spent a lot of effort in putting together the framework used by the participants. Chris Wilkerson provided tremendous help in setting up the competition rules and guiding the organization based on his experience from the branch prediction championships. I would also like to thank program committee members Yuan Chou, David Kaeli, Alvin Lebeck, Kyle Nesbit, and Suleyman Sair for taking time to help review papers and guide the paper selection process. I would like to thank all contestants who participated in the competition, even if their submissions were not accepted.

Finally, I thank all authors of this special issue for their effort, and for their patience while I was producing it. Other responsibilities delayed this special issue more than I intended, and I would like to acknowledge the authors' patience and graciousness during that time.