EPFL-TTO

EPFL Innovation Park J
CH-1015 Lausanne
Switzerland
+41 21 693 70 23
http://tto.epfl.ch/

Research Contact

Prof. Babak Falsafi
+41 21 693 55 92
sabine.susstrunk@epfl.ch

TTO Contact

Dr. Mauro Lattuada
+41 21 693 35 63
mauro.lattuada@epfl.ch

Licensing Opportunity

TTO – Technology Transfer Office

Unified prefetching into instruction cache and branch target buffer

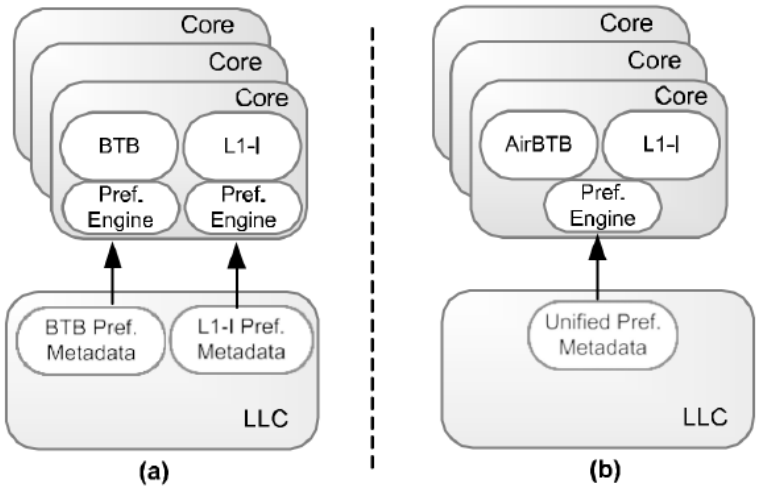

Figure: high-level organization of cores around (a) disparate BTB and L1-I prefetcher metadata (b) confluence with unified and shared prefetcher metadata;

Ref. Nr

TE 6.1468

Keywords

Servers, architecture, accelerator, low latency, branch target buffer, prefetching

Intellectual Property

US 9,996,358

Publications

"Confluence: unified instruction supply for scale-out servers" https://infoscience.epfl.ch/record/220653?ln=en

Date

22/12/2021

Description

This technology identifies the redundancy in the control flow metadata for both types of prefetchers and eliminates it by unifying the two histories. To that end, a single frontend design has been designed with a single prefetcher and a unified metadata feeding both the L1-I and the BTB. An important challenge solved by this technology is in managing the disparity in the granularity of control flow required by each of the prefetchers.

Whereas an I-cache prefetcher needs to track block-grain addresses, a BTB must reflect fine-grain information of the individual branches.

This technology overcomes this problem by exploiting a critical insight that a BTB only tracks branch targets, which do not depend on whether or not the branch is taken or even executed. Based on this insight, this technology maintains the unified control flow history at the block granularity and for each instruction block brought into the L1-I, it eagerly inserts the targets of all PC-relative branches contained in the block into the BTB.

Because the control flow exhibits spatial locality, the eager insertion policy provides high intra-block coverage without requiring fine-grain knowledge of the control flow. Finally, to overcome the exorbitant bandwidth required to insert 3-4 branches found in a typical cache block into the BTB, this technology employs a block-based BTB organization, which is also beneficial for reducing the tag overhead.

Advantages

The advantages of this technology are as follows:

- unified instruction supply architecture that maintains one set of metadata used by a single prefetcher for feeding both the L1-I and the BTB.
- light-weight block-based BTB design that takes advantage of a block-grain temporal stream and spatial locality within blocks to maintain only a small set of BTB targets.
- Elimination of 93% of the misses of a conventional BTB design with the same storage budget (around 10KB), and 85% of all L1-I misses, providing 83% of the speedup possible with a perfect L1-I and BTB.

Applications

- Scale-out servers, multi-cores processors