HIGH PERFORMANCE INSTRUCTION FETCH USING SOFTWARE AND HARDWARE CO-DESIGN

Alex Ramirez

Department of Computer Architecture
Universitat Politècnica de Catalunya (UPC)
Barcelona, Spain. April 2002

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Computer Science

Thesis advisors
Josep L. Larriba-Pey and Mateo Valero

Author
Alex Ramirez

High performance instruction fetch using software and hardware co-design

Abstract

The design of higher performance processors has been following two major trends: increasing the pipeline depth to allow faster clock rates, and widening the pipeline to allow parallel execution of more instructions. Designing a higher-performance processor implies balancing all the pipeline stages to ensure that overall performance is not dominated by any of them. This means that a faster execution engine also requires a faster fetch engine, to ensure that it is possible to read and decode enough instructions to keep the pipeline full, and the functional units busy.

This thesis explores the challenges faced by the design of the instruction fetch engine from a dual perspective: design a better software for the existing fetch architectures, and design a better hardware for the newly constructed software.

Our approach to the design of a better software has been the proposal of a novel code layout algorithm which targets both the instruction cache performance and the effective fetch width. Based on the analysis of the behavior of these optimized codes, we also propose a modification to the trace cache mechanism to make a more efficient use of the available storage space, and a novel branch predictor which exploits the available profile information to obtain higher prediction accuracy.

Finally, we propose a novel fetch architecture designed to exploit the special characteristics of optimized codes. The proposed fetch engine has low cost and complexity, but provides very high fetch performance.

To my mother, who wanted me to get the best education.
I hope you can see this from where you are.

Acknowledgments

To Josep L. Larriba-Pey who took me into this PhD adventure without giving me a chance to think about other options, for which I thank him.

To Mateo Valero who provided energy and tons of technical discussions on endless subjects which could not be fully developed to be included here.

To Luiz Barroso, Kourosh Gharachorloo, Robert Cohn, Geoff Lawney, and the whole Western Research Lab team, for hosting me for two consecutive summers, for the chance to work with them, and for plenty of cakes.

To John Shen, Hong Wang, Ed Grochowsky, and the MRL teams in Santa Clara, Austin, and Oregon, for bringing me in for an excellent summer which provided me a broader view of this field, for their technical expertise, and for their friendship.

To Jesus Corbal, Dani Jimenez, Carlos Navarro, Daniel Ortega, Xavi Serrano, Josep Torrellas, and all my fellow PBCs, who have all contributed significantly to this thesis. I could not have gone this far without them.

To my sister Marta, and my girlfriend Alicia, who have endured this long path with me, providing support when it was most needed.

And to my advisors again, for this opportunity to work with such fine people.

Contents