



Preface

This book emerged from the course *Superscalar Processor Design*, which has been taught at Carnegie Mellon University since 1995. Superscalar Processor Design is a mezzanine course targeting seniors and first-year graduate students. Quite a few of the more aggressive juniors have taken the course in the spring semester of their junior year. The prerequisite to this course is the Introduction to Computer Architecture course. The objectives for the Superscalar Processor Design course include: (1) to teach modern processor design skills at the microarchitecture level of abstraction; (2) to cover current microarchitecture techniques for achieving high performance via the exploitation of instruction-level parallelism (ILP); and (3) to impart insights and hands-on experience for the effective design of contemporary high-performance microprocessors for mobile, desktop, and server markets. In addition to covering the contents of this book, the course contains a project component that involves the microarchitectural design of a future-generation superscalar microprocessor.

During the decade of the 1990s many microarchitectural techniques for increasing clock frequency and harvesting more ILP to achieve better processor performance have been proposed and implemented in real machines. This book is an attempt to codify this large body of knowledge in a systematic way. These techniques include deep pipelining, aggressive branch prediction, dynamic register renaming, multiple instruction dispatching and issuing, out-of-order execution, and speculative load/store processing. Hundreds of research papers have been published since the early 1990s, and many of the research ideas have become reality in commercial superscalar microprocessors. In this book, the numerous techniques are organized and presented within a clear framework that facilitates ease of comprehension. The foundational principles that underlie the plethora of techniques are highlighted.

While the contents of this book would generally be viewed as graduate-level material, the book is intentionally written in a way that would be very accessible to undergraduate students. Significant effort has been spent in making seemingly complex techniques to appear as quite straightforward through appropriate abstraction and hiding of details. The priority is to convey clearly the key concepts and fundamental principles, giving just enough details to ensure understanding of implementation issues without massive dumping of information and quantitative data.

The hope is that this body of knowledge can become widely possessed by not just microarchitects and processor designers but by most B.S. and M.S. students with interests in computer systems and microprocessor design.

Here is a brief summary of the chapters.

Chapter 1 Processor Design

This chapter introduces the art of processor design, the instruction set architecture (ISA) as the specification of the processor, and the microarchitecture as the implementation of the processor. The dynamic/static interface that separates compile-time software and run-time hardware is defined and discussed. The goal of this chapter is not to revisit in depth the traditional issues regarding ISA design, but to erect the proper framework for understanding modern processor design.

Chapter 2 Pipelined Processors

This chapter focuses on the concept of pipelining, discusses instruction pipeline design, and presents the performance benefits of pipelining. Pipelining is usually introduced in the first computer architecture course. Pipelining provides the foundation for modern superscalar techniques and is presented in this chapter in a fresh and unique way. We intentionally avoid the massive dumping of bar charts and graphs; instead, we focus on distilling the foundational principles of instruction pipelining.

Chapter 3 Memory and I/O Systems

This chapter provides a larger context for the remainder of the book by including a thorough grounding in the principles and mechanisms of modern memory and I/O systems. Topics covered include memory hierarchies, caching, main memory design, virtual memory architecture, common input/output devices, processor-I/O interaction, and bus design and organization.

Chapter 4 Superscalar Organization

This chapter introduces the main concepts and the overall organization of superscalar processors. It provides a “big picture” view for the reader that leads smoothly into the detailed discussions in the next chapters on specific superscalar techniques for achieving performance. This chapter highlights only the key features of superscalar processor organizations. Chapter 7 provides a detailed survey of features found in real machines.

Chapter 5

Superscalar Techniques

This chapter is the heart of this book and presents all the major microarchitecture techniques for designing contemporary superscalar processors for achieving high performance. It classifies and presents specific techniques for enhancing instruction flow, register data flow, and memory data flow. This chapter attempts to organize a plethora of techniques into a systematic framework that facilitates ease of comprehension.

Chapter 6

The PowerPC 620

This chapter presents a detailed analysis of the PowerPC 620 microarchitecture and uses it as a case study to examine many of the issues and design tradeoffs introduced in the previous chapters. This chapter contains extensive performance data of an aggressive out-of-order design.

Chapter 7

Intel's P6 Microarchitecture

This is a case study chapter on probably the most commercially successful contemporary superscalar microarchitecture. It is written by the Intel P6 design team led by Bob Colwell and presents in depth the P6 microarchitecture that facilitated the implementation of the Pentium Pro, Pentium II, and Pentium III microprocessors. This chapter offers the readers an opportunity to peek into the mindset of a top-notch design team.

Chapter 8

Survey of Superscalar Processors

This chapter, compiled by Prof. Mark Smotherman of Clemson University, provides a historical chronicle on the development of superscalar machines and a survey of existing superscalar microprocessors. The chapter was first completed in 1998 and has been continuously revised and updated since then. It contains fascinating information that can't be found elsewhere.

Chapter 9

Advanced Instruction Flow Techniques

This chapter provides a thorough overview of issues related to high-performance instruction fetching. The topics covered include historical, currently used, and proposed advanced future techniques for branch prediction, as well as high-bandwidth and high-frequency fetch architectures like trace caches. Though not all such techniques have yet been adopted in real machines, future designs are likely to incorporate at least some form of them.

Chapter 10

Advanced Register Data Flow Techniques

This chapter highlights emerging microarchitectural techniques for increasing performance by exploiting the program characteristic of *value locality*. This program characteristic was discovered recently, and techniques ranging from software memoization, instruction reuse, and various forms of value prediction are described in this chapter. Though such techniques have not yet been adopted in real machines, future designs are likely to incorporate at least some form of them.

Chapter 11

Executing Multiple Threads

This chapter provides an introduction to thread-level parallelism (TLP), and provides a basic introduction to multiprocessing, cache coherence, and high-performance implementations that guarantee either sequential or relaxed memory ordering across multiple processors. It discusses single-chip techniques like multithreading and on-chip multiprocessing that also exploit thread-level parallelism. Finally, it visits two emerging technologies—implicit multithreading and preexecution—that attempt to extract thread-level parallelism automatically from single-threaded programs.

In summary, Chapters 1 through 5 cover fundamental concepts and foundational techniques. Chapters 6 through 8 present case studies and an extensive survey of actual commercial superscalar processors. Chapter 9 provides a thorough overview of advanced instruction flow techniques, including recent developments in advanced branch predictors. Chapters 10 and 11 should be viewed as advanced topics chapters that highlight some emerging techniques and provide an introduction to multiprocessor systems.

This is the first edition of the book. An earlier beta edition was published in 2002 with the intent of collecting feedback to help shape and hone the contents and presentation of this first edition. Through the course of the development of the book, a large set of homework and exam problems have been created. A subset of these problems are included at the end of each chapter. Several problems suggest the use of the SimpleScalar simulation suite available from the SimpleScalar website at <http://www.simplescalar.com>. A companion website for the book contains additional support material for the instructor, including a complete set of lecture slides (www.mhhe.com/shen).