

MICRO-41 TUTORIAL

COHERENCE AND MEMORY CONSISTENCY MODELS

Michel Dubois

University of Southern California

1. Background

The design of micro-architectures is deeply affected by technological trends. Current and future micro-architectures will be made of several multithreaded cores sharing memory.

The memory consistency model is a critical part of the ISA specification of an architecture. It affects programmers as well as designers of machines built for that ISA. In this course we will explore the impact of coherence and of the memory consistency model on machine design.

2. Intended audience

This is a tutorial for computer engineers and computer architects in industry, who will design chip multiprocessors or shared-memory systems. It is also intended for academic teachers and researchers both in computer engineering and science, who want to improve their teaching of these complex topics or need a clear understanding of these concepts to pursue their research. Some background in machine architectures is assumed, including the workings of individual components such as processors, caches, memory, interconnect, and input/output. Some familiarity with shared-memory multiprocessors is also recommended. In general, curious engineers or scientists with such background may also attend this tutorial to learn about the topic in a single day.

3. Content

We will use simple, abstracted models of hardware components such as processors, memory systems and interconnections. Then we use these models to understand the impact of coherence, store atomicity and memory consistency models on the design of hardware components and on their inclusion in systems. The goal is not to overview all the research on these topics, but rather to present simple examples to illuminate the relation between coherence, store atomicity and memory consistency models at the hardware level. There will be very little theory or formalism in this tutorial. This is a practical, engineering-oriented tutorial at the hardware/architecture level.

The following topics will be covered (possibly not in this order)

Shared-memory multiprocessor architecture fundamentals

Core architecture fundamentals; CMP architecture; systems built with CMPs.

Coherence

Store atomicity

Write synchronization

Sequential Consistency

Relaxed consistency models allowing store buffers (TSO, PC)

Use of fences

Relaxed consistency models based on synchronization (WO, RC)

Speculative violations of the memory consistency model in OoO processors

Impact of core multithreading

4. Biography

Michel Dubois is a Professor of Computer Engineering in the Department of Electrical Engineering at the University of Southern California. Before joining U.S.C. in 1984, he was a research engineer at the Central Research Laboratory of Thomson-CSF in Orsay, France. His area of expertise is in Computer Architecture and Parallel Processing.

He is well known for his work on cache coherence and memory consistency models. From 1993 to 2001 he led the RPM Project. RPM stands for “Rapid Prototyping engine for Multiprocessors”, an FPGA-based hardware platform to implement multiprocessor systems with different architectures. His current research interests are Chip Multiprocessors and the impact of technological trends on such micro-architectures.

Dubois holds a Ph.D. from Purdue University, an MS from the University of Minnesota, and an engineering degree from the Faculte Polytechnique de Mons in Belgium, all in Electrical Engineering. He is a fellow of the ACM and of the IEEE.