

Title	ウェーブパイプラインを用いたマルチスレッド型プロセッサアーキテクチャに関する研究
Author(s)	池田, 吉朗
Citation	
Issue Date	1999-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1288
Rights	
Description	Supervisor:日比野 靖, 情報科学研究科, 修士

Multi-threaded Processor Architecture Using Wave Pipelining

Yoshiro Ikeda

School of Information Science,
Japan Advanced Institute of Science and Technology

February 15, 1999

Keywords: wave pipelining, multi-thread.

Abstract

Many of high-performance computer architecture in recent days introduce parallelisms inside the program in some way. According to the type of parallelism used, those are called super scalar, VLIW, and a multiprocessor, etc. Almost all of those processors are accompanied with pipeline processing. The pipeline processing is indispensable for high-performance processor architecture.

This paper describes a multi-threaded processor adopting wave pipelining to operate at high rate.

A prototype wave pipelined multi-threaded processor using the following design techniques for realizing efficient wave pipelining operations and a low cost dies. In conclusion, performance of this processor is evaluated through precise timing simulation.

Multi-threaded processor architecture makes best use of pipelining technique, and can achieve high throughput only by a single processor.

Advantages introducing wave pipelining are discussed, and a design methodology to apply wave pipelining to multi-threaded processor is proposed.

First, a stage composition method for efficient wave pipelining operation of the multi-threaded processor is proposed, and then a design method for the combinational logic circuit inside stages are mentioned. Since those methods require model of logic gates and wire, procedures for modeling physical properties of those are also proposed. In addition, the cost considerations for keeping low are involved in the proposed design techniques.

1 Multi-threaded processor

Since a conventional pipelining processor executes instructions of only a single thread simultaneously, problems of pipeline hazards cannot be avoided, and the efficient pipeline processing is prevented by occurrence of pipeline stall.

Multi-threaded processor performs independent instructions from different threads in parallel, and fills up all pipeline stages with a operation from different threads. Since operations in different threads have no dependency each other, the pipeline of multi-threaded processor is free from pipeline hazards and does not cause the stall of the pipeline. In addition, multi-threaded processor allows long memory latency and processor can be kept always busy. Therefore, a multi-threaded processor derives high-performance pipeline processing.

2 Wave pipelining

It is difficult to improve the maximum operating frequency of multi-threaded processor with fine-grained pipeline because maximum propagation delay through the longest stage is need to be shortened. If wave pipelining (also called “maximum rate pipelining”) technique is available, however, some of difficulties can be overcome.

Wave pipelining is a timing methodology by increasing the number of effective pipeline stages without increasing the number of physical pipeline latches. The wave pipelining allows a combinational logic in pipeline stages to execute an operation overlapped with a predecessor or a successor, and allows synchronous operations to be clocked at higher rate than that can be achieved with conventional pipelining techniques, and also improve resource usage.

Pipeline structure of a multi-threaded processor and the multiple register sets of the number of threads are very convenient for wave pipelining.

3 The design for a wave pipelining

Since the same design techniques as conventional pipelined processor cannot be applied to wave pipelined processor, considerations for the property of wave pipelining is required to establish design methodology.

For preparation, a model of logic gates and wire is provided. Wave pipelining has severer timing constraint than conventional pipeline. Therefore, a proper modeling is required. It should be realistic to some extent, but should be abstracted by that a design may not become too complicated.

4 Pipeline stage composition

When pipeline stages are operated in wave pipelined fashion, not all stages need to be operated at the same timing. However, there may be mixed stages which cannot be wave

pipelined or stages which must synchronize each operating timing and stages operated in wave pipelined fashion. So, arrangement of the pipeline latches for efficient wave pipelining operation, the operating timing at which each stages operate, clock distribution strategy must be considered.

5 Circuit design inside stages

Maximum operating frequency of conventional pipelining systems is determined by the maximum propagation delay time of a pipeline stage. In the case of wave pipelining, on the other hand, the main factor determines the maximum operating frequency is the difference between the maximum propagation delay time and the minimum propagation delay time through the combinational logic. So, how to minimize delay variations inside a pipeline stage is important to achieve high-speed wave pipelined operation.

As a means to balance delays in combinational logic, only a simple technique is used in this paper, that is, inserting a buffer as a delay element into a signal path whose delay is smaller than the maximum path delay.

6 Performance and cost

This paper, takes the design approach of buffer insertion as a means of delay balancing inside pipeline stages. So if the number of buffers increases, the cost of a die will increase so much because it is proportional to exponential of the die size. A design technique that requires only minimum number of buffers is discussed.

7 Designing prototype processor

Those design techniques proposed above for wave pipelining is applied to the multi-threaded processor named MUP (Multi-threaded Ultra-pipelining Processor) which has been designed with ordinal pipeline technique[1].

The architecture of MUP is compatible with MIPS, that is, MIPS compatible instruction set, register set, exception processing, etc. Features of MUP which differ from that of MIPS are pipelined I/D-cache memory, un-supporting floating-point and shift operation, and the most notable, multi-threaded.

The wave pipelined version of MUP written in a behavior description language is expanded to gate level description, and is evaluated in both performance and cost. This evaluation will lead us to more sophisticated design of wave pipelined multi-threaded processor.

References

- [1] Eiji Itoh, “*A Multithreaded Processor Architecture for Executing Functional Programs*”, 1997