Chapter 18

Reducing Power and Energy Overhead in Instruction Prefetching for Embedded Processor Systems

Ji Gu
University of New South Wales, Australia

Hui Guo
University of New South Wales, Australia

ABSTRACT

Instruction prefetching is an effective way to improve performance of the pipelined processors. However, existing instruction prefetching schemes increase performance with a significant energy sacrifice, making them unsuitable for embedded and ubiquitous systems where high performance and low energy consumption are all demanded. This paper proposes reducing energy overhead in instruction prefetching by using a simple hardware/software design and an efficient prefetching operation scheme. Two approaches are investigated: Decoded Loop Instruction Cache-based Prefetching (DLICP) that is most effective for loop intensive applications, and the enhanced DLICP with the popular existing Next Line Prefetching (NLP) for applications of a moderate number of loops. The experimental results show that both DLICP and the enhanced DLICP deliver improved performance at a much reduced energy overhead.

INTRODUCTION

On-chip cache has been widely used in modern microprocessor systems to bridge the speed gap between the processor and main memory. Cache exploits the spatial and temporal locality of memory reference to avoid the long latency of memory access from the processor. The high cache hit ratio plays a vital role in the overall system performance. This is especially essential for the instruction cache (I-cache) because of frequent instruction fetch operations. An instruction cache miss will cause the processor stall, hence slowing down the system.
Plenty of techniques have been proposed to reduce I-cache misses for high system performance. Among them is the instruction prefetching (Smith, 1978; Hennessy & Patterson, 2003) -fetching instructions from memory into the cache before they are used so that cache misses can be avoided or reduced. However, existing instruction prefetching schemes mainly focus on improving cache performance, often suffering significant energy losses due to a large amount of wasteful over-prefetching operations and/or the complicated prefetching hardware components.

Nevertheless, low energy consumption is also one of the important design issues for embedded systems, especially for those used in battery driven mobile/handheld devices, where large and heavy batteries are not feasible. A low power/energy design is, therefore, mostly desired for a long battery life.

In this paper, we aim to reduce energy overhead in instruction prefetching by using a simple prefetching hardware/software design and an efficient prefetching operation scheme. We investigate two approaches: the decoded loop instruction cache-based prefetching (DLICP) and the enhanced DLICP.

The decoded loop instruction cache (DLIC) originates from the decoded instruction buffer (DIB) proposed in Bajwa et al. (1997). It is a small tag-less cache residing between the instruction decoder and the execution unit in the microprocessor to store decoded loop instructions so that fetching and decoding the same set of instructions for the following loop iterations can be avoided, hence reducing energy dissipation in the processor.

We extend this energy-saving technique to instruction prefetching by overlapping the execution of decoded loops with fetching instructions to the cache from memory so that most instructions are available in the cache when they are executed. This approach is effective for loop intensive applications. For applications with a small amount of loops, we enhance the design with the existing Next Line Prefetching (NLP) scheme, which has been proved efficient in cache miss reduction for applications with a dominant sequential instruction execution flow (Smith & Hsu, 1992).

INSTRUCTION PREFETCHING TECHNIQUES

The proliferation of handheld, mobile, and ubiquitous devices has led to the research boom on embedded systems. Various design issues and approaches have been proposed. Bisdikian et al. (1998) present an experimental platform to research technologies and applications that enable ubiquitous, environment-aware, and low-cost computing of handheld devices in the wireless personal access networks. Medvidovic et al. (2003) developed a software-architecture-based scheme to support distributed computation on handheld devices. Several power management schemes at the hardware level have also been proposed to reduce power consumption in different mobile system components such as the display (Min & Cha, 2007), the graphic processing unit (Nam, Lee, Kim, Lee, & Yoo, 2008).

Our proposal in this paper is an energy-efficient instruction prefetching scheme for embedded processors that can be used in the mobile/handheld devices for high system performance and low energy consumption. This section reviews some existing instruction prefetching methods for cache performance optimization.

Existing instruction prefetching techniques can be classified as software based prefetching and hardware based prefetching. Software prefetching schemes (Gornish, Granston, & Veidenbaum, 1990; Luk & Mowry, 1998; Cristal et al., 2005) rely on the compiler to insert prefetch instructions into the program code before the application is executed, which requires a known memory access behavior and a dedicated compiler.

The hardware-based prefetching is transparent to software and exploits the status of the program execution to dynamically prefetch instructions for