A Study of Leakage Power Reduction Mechanisms on Functional Units and TLBs for Embedded Processors

Zhao LEI

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

School of Science for Open and Environmental Systems
Graduate School of Science and Technology
Keio University

2011

Preface

Power consumption has been widely recognized as a first-class design constraint for embedded processors, due to its impact on operation reliability, system density, and integration costs. While dynamic power has represented the predominant factor in CMOS circuits for many years, the leakage power, which is consumed by each transistor even when no active switching is taking place, is increasingly prominent with technology scale. Now, suppressing the leakage power of embedded processors, especially for battery-driven devices, is a critical challenge facing the embedded system community.

Leakage-efficient design requires an in-depth examination of each system component. In this thesis, we explore the leakage reduction mechanisms on functional units, instruction TLB, and data TLB, all of which take up a significant share of the total leakage consumption of embedded processors but have not been well studied due to their high access frequency and utilization.

As for functional units, we propose a framework to Power Gating (PG) each of the units at runtime by integrating circuit-level, architecture-level, and system software techniques. At circuit-level, we propose a fine-grained power gating technique, which has nano-second order wakeup latency and can be implemented at arbitrary granularity. At architecture-level, a PG control scheme, which keeps a functional unit active only when being used, has been applied. In addition, BET-aware (Break Even Time) PG control schemes, which are guided by the system software (compiler and operating system), have also been proposed to achieve maximum leakage reduction effects.

As for the instruction TLB (iTLB), we exploit the spatial locality of the page-based iTLB references. By inserting a small size storage component, which keeps the recent address-translation information, between the processor and the iTLB, a majority of address-translation requests can be satisfied with the small component without accessing the iTLB. Then, with integration of the Dual Voltage Supply (DVS) technique, the iTLB can be put into low-leakage mode (with the lower voltage supply) and restored to the active mode only when the iTLB look-up becomes necessary. Based on such a design philosophy, three different leakage control policies have also been proposed to achieve the best leakage reduction efficiency.

As for the data TLB (dTLB), we exploit the temporary locality dTLB references. By dividing the overall execution time into smaller time slices, we can observe the dTLB referencing in a finer time resolution, and the locality of dTLB references in and between adjacent slices can be utilized to recognize the contributive dTLB entries in each slice. Then, with integration of the DVS technique,

those non-contributive entries can be put into low leakage mode dynamically.

The proposed mechanisms are evaluated in terms of leakage power consumption by using real-chip/post-layout evaluation based on 65nm CMOS technology. Evaluation results show that the proposed mechanisms can reduce the leakage power consumption of the functional units, iTLB and dTLB by 20%, 50%, and 37% respectively.