論文の内容の要旨

# A STUDY ON HIGH PERFORMANCE PROCESSOR USING PREDICTION TECHNIQUE

(プロセッサ性能の予測技術を用いる高速化の研究)

氏名　石井　康雄

As the semiconductor process technology has been improved, the clock frequency of the processor chip has increased significantly. However, it also increases the power consumption of the processor chip. Recently, the power consumption of processor chip limits the improvement of the clock frequency. To increase the performance of processor chip without increasing the power consumption, the processor has to eliminate the pipeline stall time for exploiting Instruction-Level Parallelism (ILP). A speculative execution using hardware predictor is promising approach to eliminate the pipeline stall time due to data dependency and control dependency. To eliminate the pipeline stall time, many speculative execution techniques such as aggressive data prefetching and complicated branch target prediction have been proposed separately. However, these separated predictors often conflict with each other because the design of these predictors does not consider the effect of the other predictors. This often reduces the prediction accuracy and degrades the performance of the speculative execution.

To resolve this problem, we unify multiple separated predictors into one unified predictor. The unified predictor design has three advantages over existing separated predictor design.

First, the confliction among multiple speculation techniques can be effectively eliminated because each prediction algorithm can consider the action of the other speculative execution features. Second, the total cost of the predictor can be reduced because multiple features often profiles similar run-time information to predict future execution status in current design. However, such information can be shared among multiple features. Finally, new opportunities to eliminate the pipeline stall time are arise because the prediction results of the other predictors often useful for the other predictors.

As the unified predictor design, we propose two unified predictor, which we call Unified Memory Optimizing (UMO) architecture and Unified Branch Prediction (UBP) architecture. UMO architecture unifies data prefetching, advanced cache replacement policy and memory access scheduling to improve the performance of memory subsystem. UMO architecture eliminates pipeline stall time due to off-chip memory access. As the implementation of UMO architecture, we propose Map-based Unified Memory Subsystem Controller (MUMSC). MUMSC utilize memory access map, which is cost-effective data structure to track previous memory access, for predicting future memory access. On this data structure, we implement a unified predictor which has the responsibility to control the while memory subsystem. MUMSC improves the performance by 17.2% compared with current separated memory subsystem design. UBP architecture reduces the branch misprediction due to branch target prediction. UBP architecture unifies two branch target prediction algorithms, one is designed for single-target branches and the other is designed for multi-target branches. UBP architecture employs the unified prediction table to adapt the resource allocation for each prediction algorithms. On the unified prediction table, UBP architecture performs two-phase prediction to realize two different prediction algorithms. As the implementation of UBP architecture, we propose Bimode-Cascading Branch Target Buffer (BCBTB). BCBTB unifies the traditional BTB and the ITTAGE branch predictor. Compared with current separated branch target predictor, BCBTB increase the performance by 7.1% and reduce the energy by 4.5% for SPEC JVM2008 benchmarks. These unified predictors effectively improve performance, reduce implementation cost, and reduce energy consumption. This type of design approaches will be becoming important in future processor design.