論 文 の 内 容 の 要 旨

論文題目　Cache Design Optimization for Energy-Efficient Processors
（高電力効率プロセッサのためのキャッシュの設計最適化）

氏名　有間　英志

It is critical to improve the energy-efficiency of microprocessors for many computing systems from HPC (High Performance Computing) systems to embedded devices. In a microprocessor, caches are major contributors to performance and also major power consumers, especially in mobile devices, thus it is critical to optimize cache designs for both performance and power.

To improve the energy-efficiency of microprocessors through the optimization of cache design, the following two approaches are regarded as the most promising. One is to optimize the power control to conventional SRAM based caches in order to reduce their energy while suppressing performance degradation. The other is to utilize non-volatile memories like STT-MRAM for caches because they do not require any power supply for data retention unlike the conventional SRAM based ones, thus their leakage power is quite small.

To maximize the energy-efficiency of caches, we need to choose the best approach from above in consideration of various factors such as the characteristics of typical workloads utilized on a certain system, the performance/cost constraint of the system and the parameters of each memory, thus the best choice highly depends on all of them. To maximize the energy-efficiency of caches for various systems like mobile systems this thesis proposes following three proposals.

First, this thesis proposes an energy-efficiency improvement technique called Lost Data Prefetch for SRAM caches during executing I/O bound workloads. In a conventional microprocessor, turning off an SRAM cache can save the leakage power significantly but introduces data loss which causes significant performance degradation

because of additional main memory accesses. To overcome this problem, the proposed technique fetches the necessary lost data from the main memory before they are accessed. The evaluation result shows that our methodology can save large part of the total leakage energy of an SRAM cache while suppressing performance degradation.

Second, we proposed an energy reduction technique for STT-MRAM caches. In an STT-MRAM cache, large and leaky transistors are required for its peripheral circuits to drive large current to its memory cells, thus the peripheral circuits consume large leakage power. To cope with this problem, this thesis proposes a novel power management scheme called Immediate Sleep which immediately turns off a subarray that contains leaky local peripheral circuits when the next access will be non-performance-critical. By doing so, the leakage energy can be significantly reduced with minimal performance impact.

Third, we clarified the parameter requirement of STT-MRAM caches for various systems depending on the typical load state. For systems on which only I/O bound applications are usually executed, the cache access performance/energy are less critical to the overall system performance/energy because caches are less frequently accessed. As the cost of STT-MRAM caches highly depends on their parameters, we have to know the parameter requirement to optimize the design. Therefore, this thesis evaluated the parameter requirement of STT-MRAM in consideration of various factors like the average load state of a system, the performance constraint of the system and the required energy reduction.

From above results, it is turned out that the energy-efficiency of microprocessors can be significantly improved by choosing the optimal memory technology for caches and applying the proposed techniques to them when taking various factors into consideration. This contribution is quite important to optimize the cache design which determines the energy-efficiency of various computing systems through the future.