

누수에너지 절약과 시스템 성능 향상을 위한

백업 캐시 제안

최병창^o 우장복 서호중
가톨릭대학교 컴퓨터공학과
{hellocbc^o, sofe4u, hjsuh}@catholic.ac.kr

A Backup-Cache for Leakage-Energy-Reduction and High Performance System

ByeongChang Choi^o, JangBok Woo, Hyo-Joong Suh
Dept of Computer Science and Engineering, The Catholic University of Korea

요 약

임베디드 시스템에서의 캐시 메모리는 시스템의 성능에 큰 영향을 줄뿐만 아니라 전체 에너지 소비 중 50% 정도를 소비하고 있어 캐시 메모리의 성능과 에너지 소비는 큰 관심거리 중 하나이다. 공정의 미세화로 캐시 메모리의 에너지 소비 중 누수 전류에 의한 에너지 소비의 비중이 더 커지고 있어, 정적 에너지 소비를 줄이기 위한 다양한 연구가 진행 중이다. 에너지 절약과 성능 향상은 손익 상쇄(Trade-off)관계에 있어 두 가지 목표를 동시에 달성하기는 힘들다. 본 논문에서는 성능 향상을 위하여 여러 가지 캐시 구조 중 접속 속도가 가장 빠른 직접 사상 캐시를 사용하고, 완전 연관 캐시를 사용하여 직접 사상 캐시의 단점을 보완 할 수 있는 백업 캐시 시스템을 제안한다. 시스템 성능을 향상 시키면서 백업 캐시의 누수 에너지를 절약하기 위해 직접 사상 캐시와 완전 연관 캐시를 서로 다른 한계 전압을 가지는 SRAM으로 구성한다. 직접 사상 캐시는 낮은 한계 전압의 SRAM로 구성되어 높은 성능을 내고, 완전 연관 캐시는 직접 사상 캐시에 비해 상대적으로 속도는 느리지만 누수 에너지가 적은 높은 한계 전압을 가지는 SRAM으로 구성되어 직접 사상 캐시를 보완하는 역할을 할 것이다.

1. 서 론

임베디드 시스템에서 캐시 메모리는 시스템의 성능에 큰 영향을 줄뿐만 아니라, 캐시 메모리에서 소비 되는 에너지가 전체 에너지 중 50% 정도를 차지하고 있어 캐시 메모리의 성능과 에너지 소비는 큰 관심거리 중 하나이다[1]. 기존의 CMOS의 장점은 낮은 전력 소모였는데, 반도체 공정 기술의 발달로 속도의 향상과 밀도가 증가 하면서 누수 전류에 의한 정적 에너지 소모가 급격히 증가하여, 칩의 총 에너지 소모 중 15~20%를 차지하게 되었다[2]. 따라서 정적 에너지 소모를 줄이기 위한 여러 가지 연구가 진행 되고 있다. 캐시 메모리에는 여러 가지 구조가 있는데, 직접 사상 캐시, 완전 연관 캐시, 집합 연관 캐시 등이 대표적이다. 직접 사상 캐시는 캐시 접속 속도가 가장 빠르지만 다루는 메모리의 크기가 커지거나 특별한 경우에 적중 실패율이 급격하게 상승하는 단점이 있고, 완전 연관 캐시는 연관성이 높아 적중 성공률이 높은 반면 태그 비교를 위하여 많은 에너지를 소비한다. 집합 연관 캐시는 높은 적중 성공률을 가지지만 속도가 느리다는 단점이 있다. 본 논문에서는 캐시 접속 속도가 가장 빠른 직접 사상 캐시를 높은 한계 전압을 가지는 SRAM으로 구성하여 시스템 성능을 향상시키고, 연관성이 가장 높은 완전 연관 캐시를 사용하여

특별한 경우 캐시 적중 실패율이 급격하게 증가하는 직접 사상 캐시의 단점을 보완하는 구조를 제안하고 백업 캐시라고 부르기로 한다. 백업 캐시는 직접 사상 캐시에서 적중 실패가 발생했을 경우 사용되며 자세한 설명은 3장에서 계속할 것이다. 백업 캐시는 누수 에너지를 절약하기 위하여 비록 속도는 느리지만 누수 에너지가 높은 한계 전압의 SRAM로 구성하고, 완전 연관 캐시를 n-뱅크로 나누어 태그 비교에 소비되는 에너지를 절약할 것이다.

2. 관련 연구

2.1 정적 에너지 소모

정적 에너지(Static Power: P_{off}) 소모란 컨덕터에서의 스위칭 활동이 없어도 소모되는 에너지로 누수 전류(Leakage Current: I_{off})와 공급 전압(Supply Voltage: V_{dd})의 곱으로 정의 된다.

$$P_{off} = I_{off} V_{dd} \quad (1)$$

전력 소모를 줄이기 위한 가장 효과적인 방법은 공급 전압(Supply Voltage)을 줄이는 것인데, 공급 전압을 줄이면 전송지연(Propagation Delay)이 증가하여 시스템의 성능이 떨어지게 된다.

$$t_{pd} \propto \frac{V_{dd}}{(V_{dd} - V_t)} \quad (2)$$

따라서 전력 소모를 줄이면서 동시에 시스템의 성능을 유지하기 위해서는 한계 전압(Threshold Voltage)도 같이 내려야 되는데, 한계 전압을 내리게 되면 성능은 높아지지만 누수 전류가 기하급수적으로 증가하게 된다 [4]. 공급 전압을 줄여서 정적 에너지를 줄이는 방법으로 [3]에서는 Gated- V_{dd} 를 사용하여 자주 사용하지 않는 캐시 메모리의 전원 공급을 끊어서 전력 소모를 줄이는 방법을 제안했다. 그러나 이 방법은 휴면 상태의 캐시 메모리 셀들이 다시 사용될 때, 전압을 끊어서 없어진 데이터를 다시 포워딩하기 위한 복잡한 추가 작업이 필요하다. 한계 전압을 조절해서 캐시의 데이터를 유지하면서 누수 에너지를 줄이는 연구로는 ABB-MTCMOS를 사용하는 방법이 있는데, ABB-MTCMOS는 휴면 모드에서 누수 에너지를 상대적으로 많이 소비하고 캐시의 상태 변환 속도가 느리다는 단점이 있다. 공급 전압을 조절해서 캐시의 데이터를 유지하면서 누수 에너지를 줄이는 연구로 DVS(Dynamic Vdd Scaling) 방법이 제안되었다[2]. DVS는 활성화 상태에는 정상 전압을 공급하고 휴면 모드 상태에 있을 때는 데이터를 유지할 수 있는 최소한의 전압을 공급한다. [4]에서는 서로 다른 한계 전압을 가지는 캐시 웨이로 구성된 구조를 제안하고 있다. 많이 사용되는 부분에는 낮은 한계 전압을 가지는 웨이를 할당하여 빠른 속도로 작동하게 하고 나머지 웨이는 높은 한계 전압을 가지게 하여 누수 에너지를 줄이고 있다.

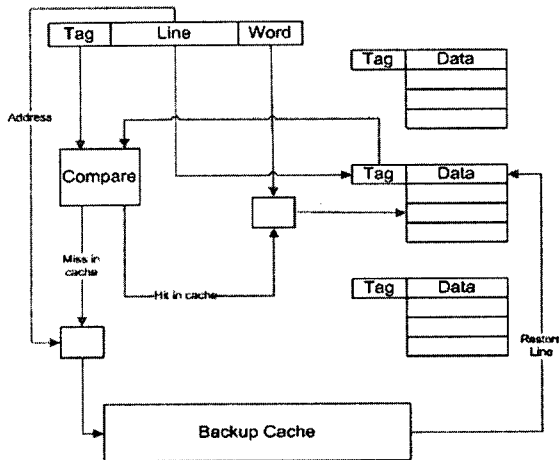


그림 1 캐시 구성도

3. 백업 캐시의 구조와 동작원리

3.1 캐시 구조

그림 1에서 직접 사상 캐시는 낮은 한계 전압의

SRAM으로 구성 되어 있어, 캐시 메모리에 적중 성공했을 경우 빠른 속도로 데이터를 읽을 수 있어 성능 향상이 가능하다. 그림 2의 백업 캐시는 완전 연관 캐시이며 4-뱅크 구조로 구성되어 있다. 뱅크는 DVS(Dynamic Voltage Scaling)를 사용하며, 선택된 뱅크는 정상 전압이 가해져 활성화 상태로 되고 나머지 뱅크들은 최소한의 전압이 가해지는 저전력 모드에 있게 된다.

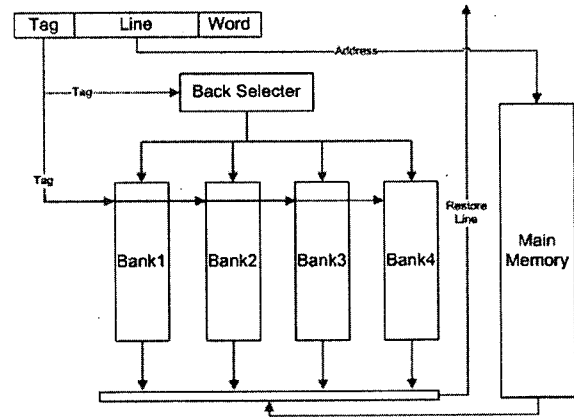


그림2 백업 캐시 구조

3.2 백업 캐시의 동작 원리

처음 태그를 비교하여 데이터를 가져 오는 단계에서는 그 동작 원리가 일반적인 직접 사상 캐시와 동일하다. 우선 태그를 비교하여 적중이 성공했을 경우 해당 라인을 찾아 워드 값을 이용하여 데이터를 찾는다. 만약 태그 비교에 실패했을 경우 메모리 주소 값을 백업 캐시로 보낸다. 백업 캐시는 4-뱅크로 구성된 완전 연관 캐시로 4개의 뱅크 모두 휴면 상태로 있다. 태그 값을 이용하여 뱅크를 선택하면 선택된 뱅크는 활성화 상태가 된다. 활성화 상태가 된 뱅크에서 태그를 비교하여 성공할 경우 해당 라인에 직접 사상 캐시의 해당 라인에 바뀐서 기록하고 실패했을 경우 메모리로부터 읽어와서 백업 캐시와 직접 사상 캐시에 기록한다. 백업 캐시가 가득 찼을 경우 가장 오랫동안 접근 하지 않은 라인을 지우고 그곳에 기록한다.

4. 성능 분석

4.1 수식

캐시 메모리에서 사용되는 에너지는 다음과 같다[1].

$$\begin{aligned}
 Energy_{mem} &= energy_{dynamic} + energy_{static} \\
 energy_{dynamic} &= cache_{hits} * energy_{hit} + \\
 &\quad cache_{misses} * energy_{miss} \quad (3) \\
 energy_{miss} &= energy_{offchip_access} + \\
 &\quad energy_{uP_stall} + energy_{cache_block_fill} \\
 energy_{static} &= cycles * energy_{static_per_cycle}
 \end{aligned}$$

전체 에너지는 동적 에너지와 정적 에너지의 합과 같다.

동적 에너지는 캐시가 적중률수와 적중했을 때 소비되는 에너지의 곱과 캐시 적중이 실패한 수와 실패했을 때 소비되는 에너지의 곱의 합으로 나타 낼 수 있고, 이 때 적중 실패 했을 때 소비 되는 에너지는 외부 메인 메모리에 접근할 때 소모되는 에너지와 메인 메모리에서 데이터를 가져오는 사이에 마이크로프로세서가 쉬게 되는데 이때 소모되는 에너지 그리고 캐시 블록에 데이터를 기록하는데 사용되는 에너지의 합으로 나타 낼 수 있다. 이때 이 세 가지 에너지는 메모리와 프로세서의 종류에 따라 달라 질 수 있는데 보통 적중 실패 때 소모되는 에너지는 적중 했을 때 소모되는 에너지의 50~200배 정도 되는 것으로 알려져 있다[1]. 본 논문에서는 평균 100배로 백업 캐시에서 가져 오는 경우 10배로, 백업 캐시에서의 적중률은 평균 90%로 가정한다. 정적 에너지는 프로그램 수행에 필요한 총 사이클 수와 한 사이클에 소비되는 정적 에너지의 곱으로 나타 낼 수 있다. 캐시 메모리에 적중 했을 경우에는 1사이클이 필요하고, 실패 했을 경우에는 메인 메모리에서 데이터를 가져오는데 20사이클이 필요하고, 백업 캐시에서 가져오는 경우에는 5사이클이 필요하다고 가정한다. 한 사이클에 사용되는 총 에너지 중 정적 에너지의 비율을 높은 한계전압의 캐시에서는 30%, 낮은 한계 전압의 캐시에서는 50%로 가정하고, 적중률이 100%일 경우 100사이클에 끝날 수 있는 작업으로, 적중했을 경우 1사이클에 소비되는 에너지는 상대적 수치인 1로 가정한다.

4.2 성능 평가

우선 직접 사상에서의 적중률에 따른 에너지 소비와 필요한 사이클 수를 구해보면 다음과 같다.

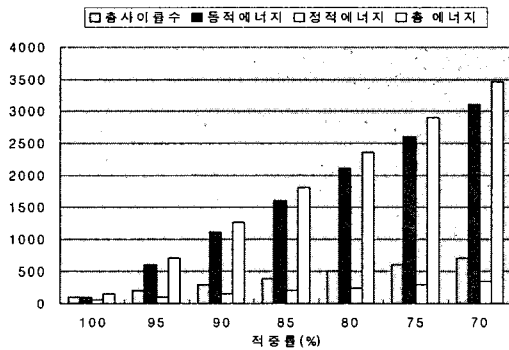


그림 3 직접 사상

적중률이 100%에 가까울 때는 직접 사상이 백업 캐시를 사용한 직접 사상 보다 에너지 소비와 성능 면에서 약간 좋은 모습을 보인다. 그러나 적중률이 95%이하로 내려 갈수록 총 에너지 소비와 사이클 수가 급격히 늘어나는 것을 볼 수 있다. 직접 사상의 경우 캐시 적중이 실패했을 경우 메인 메모리에서 데이터를 가져오기 위해 많은 시간과 에너지를 소비 하는 반면, 백업 캐시를 적용한 경우 비교적 짧은 시간 안에 데이터를 가져 올 수 있으며, 백업 캐시가 완전 연관 캐시로 구성되어 있어 높은 캐시 적중률을 보이기 때문에 메인 메모리에 접근할 확률이 낮다.

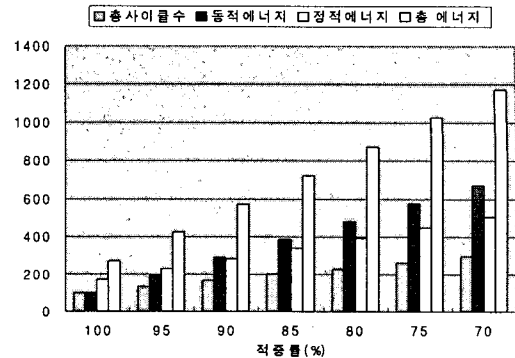


그림 4 백업 캐시를 적용한 직접 사상

적중률이 높을 때 백업 캐시를 적용한 직접 사상이 에너지 소비가 많은 이유는 직접 사상의 4배 크기인 완전 연관 캐시에서 정적 에너지를 소비하기 때문인데 적중률이 내려 갈수록 수행에 필요한 사이클 수가 줄어들어 정적 에너지와 동적 에너지 모두 줄어들게 된다.

5. 결론 및 차후 연구 과제

앞서서 본 바와 같이 백업 캐시를 적용한 직접 사상은 백업 캐시를 사용하여 직접 사상의 장점인 빠른 접근 속도를 취하고, 특정한 경우 캐시 적중 실패율이 급격히 증가하는 단점을 보완 할 수 있어, 에너지 소비와 성능 면에서 좋은 성과를 보여 줄 것으로 예상 된다. 차후 연구 과제로는 다른 여러 가지 캐시 메모리 구조에 성능 및 에너지 소비를 비교해 보고, 다양 경우에 백업 캐시를 적용을 해보는 것과, 캐시 적중 실패가 발생했을 경우에 백업 캐시에서의 백업 전략과 구현의 단순화에 대해서 연구 중이다.

6. 참고 문헌

[1] C. Zhang, F. Vahid, W. Najjar, A Highly Configurable Cache Architecture for Embedded Systems, Proceedings of the Int. Symp. on Computer Architecture, pp. 136 - 146, San Diego, CA June. 2003.
 [2] K. Flautner, N.S. Kim, S. Martin, D. Blaauw, T. Mudge, Drowsy Caches: Simple Techniques for Reducing Leakage Power, Proceedings of the Int. Symp. on Computer Architecture, 2002.
 [3] M. Powell, S.H. Yang, B. Falsafi, K. Roy, T.N. Vijaykumar, Gated-Vdd: A Circuit Technique to Reduce Leakage in Deep-Submicron Cache Memories, Int. Symp. on Low Power Electronics and Design, 2000.
 [4] A. Sakanaka, S. Fujii, T. Sato, A Leakage-Energy-Reduction Technique for Highly-Associative Caches in Embedded Systems, ACM SIGARCH Computer Architecture News, Vol 32, No 3, June 2004