

캐시라인을 내부에 저장하는 멀티코어 캐시 일관성 관리 유닛

김은찬, 김봉진, 김태환, 박인철
한국과학기술원 전기 및 전자공학

e-mail : *engine@kaist.ac.kr*, *bjkim@kaist.ac.kr*, *thkim@kaist.ac.kr*, *icpark@ee.kaist.ac.kr*,

Coherence Management Unit Saving Modified Lines Internally for Multicore System

Eunchan Kim, Bongjin Kim, Taehwan Kim, Incheol Park
School of Electrical Engineering
KAIST

Abstract

Coherency issues are the key on the single chip multi-core processor system. In order to maintain the coherency, cache-to-cache transfer is required. However, cache-to-cache transfer can cause the performance degradation of other cores since requested caches from other cores have to stop their internal cache system to access internal storage such as tag and line memories. To relax this problem, this paper suggests coherence management unit having internal buffers to save some modified cache lines. This scheme can reduce extra traffics between coherence management unit and processors.

I. 서론

마이크로프로세서는 2000년대 초반까지 사이클을 증가시키는데 주력해 왔다. 60MHz에서 시작한 486에서 불과 십여 년 사이에 속도는 3GHz 수준으로 발전해 왔다. 그러나 3GHz 대역을 넘어서는 범용프로세서는 아직 나오지 않고 있다. 그 한계를 넘어서기 위해 코어의 개수를 늘리는 멀티코어 방향으로 연구가 진행되었다[1]. 한 칩에 여러개의 코어가 존재하는 CMP 시스템에서 각 코어에 있는 캐시의 동기화가 중요한 문제로

부각되었다.

II. 본론

2.1 MESI

MESI는 각 캐시라인이 Modified, Exclusive, Shared, 그리고 Invalid의 상태를 가지는 캐시 구조를 말한다. Modified는 그 캐시라인이 하나의 캐시에서만 존재하며 수정된 상태를 의미하고 Exclusive는 캐시라인이 하나의 캐시에만 존재하는 것을 의미한다. Shared는 여러 캐시라인에서 서로 데이터를 같이 가지고 있으며 Invalid는 캐시라인이 비어있음을 의미한다.

2.2 일관성 관리 유닛(CMU)

멀티코어에서 각 1차 캐시의 일관성을 유지하기 위해 일반적으로 일관성 관리 유닛을 두어 각 코어에서 일어나는 일을 주시하며 다른 코어와 동기화 시키는 작업을 한다. 그림 1은 일반적인 일관성 관리 유닛을 보여준다. 캐시는 필요할 때 내부 코어의 요청을 CMU에게 보내며 CMU는 이를 해석하여 다른 캐시에 데이터가 있을 경우 Cache-to-Cache 전송을 하여 메모리에서부터 읽어오는 시간을 줄일 수 있다.

2.3 문제점

MESI 규칙하에서 일반적인 Cache-to-Cache 전송은

타 캐시에게 자료를 요청하는 일이기 때문에 그 캐시는 해당 작업을 수행하기 위해 코어가 접근할 수 없는 상황이 발생한다. 이 경우 요청한 코어의 입장에서는 반응이 빨라졌지만 타 캐시에게는 그만큼의 성능저하가

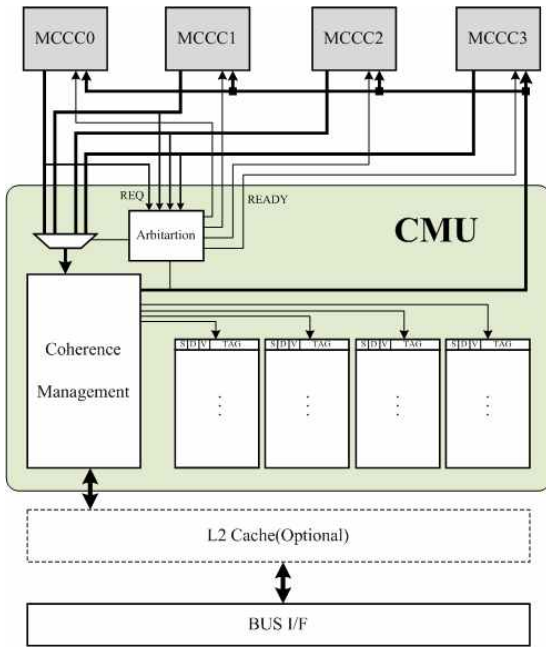


그림 1 일관성 관리 유닛

발생하게 된다. 이는 수정된 상태의 캐시라인에서 더 심각하게 된다. 수정된 캐시라인은 Cache-to-Cache 전송이 일어나면 수정된 상태를 다른 캐시로 옮기고 본래 그 라인을 가지고 있던 캐시는 초기화 된다. 이 경우 전송하는 부분과 초기화 하는 부분의 성능 저하가 더 심해지게 된다.

2.4 수정된 캐시라인을 저장하는 CMU

위에서 언급된 성능저하를 막기위해 본 논문에서는 일관성 관리 유닛 내에 수정된 캐시라인을 저장하는 공간을 두어 Cache-to-Cache 전송시 일관성 관리 유닛 내부에 복사된 태그 메모리를 보고 수정된 캐시라인이 내부에 있다면 타 캐시에 요청하지 않고 바로 내부에서 데이터를 전송하여 응답시간을 빠르게 하고 타 캐시의 성능 저하를 막을 수 있다.

III. 구현

일관성 관리 유닛 내부에 Modified 캐시라인을 위해 On-Chip SRAM을 붙여 넣어서 그림 3과 같은 시스템을 구축하였다.

일관성 관리 유닛은 캐시의 태그 사본을 보관하는데 이는 요청이 들어왔을 때, 타 캐시에 해당 캐시라인이

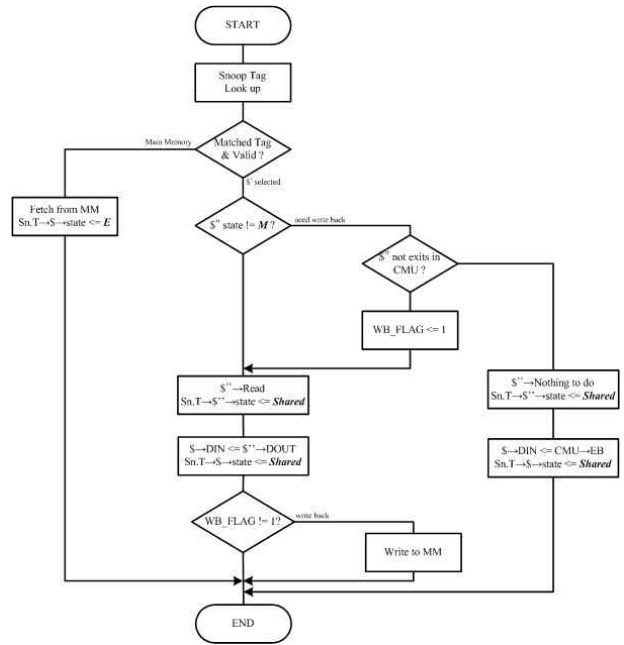


그림 2 수정된 일관성 관리 유닛 순서도

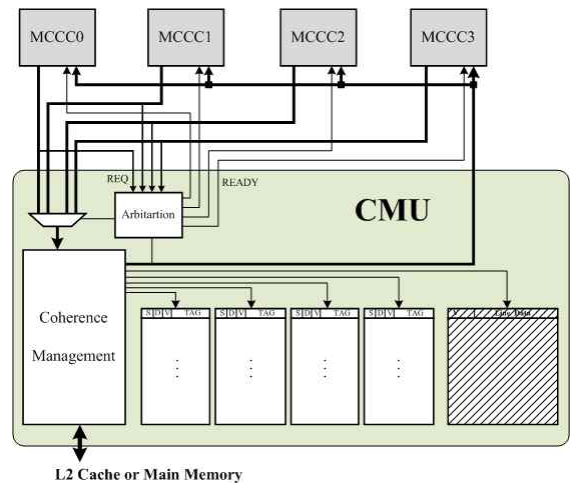


그림 3 수정된 일관성 관리 유닛 구조

있는지 검사하기 위해 타 캐시에게 요청하는 부분을 줄일 수 있기 때문에 효율적이다. 또한 이 태그는 기존 일관성 관리 유닛의 태그와는 다르게 Modified, Exclusive, Invalid 정보 외에 내부 Modified에 존재하는지를 알려주는 부가적인 비트가 들어있다. 일관성 관리 유닛은 이 정보를 보고 내부에서 바로 요청한 캐시로 전송이 일어나게 된다.

IV. 결론 및 향후 연구 방향

칩 내부의 마이크로프로세서의 멀티코어화에서 코어를 늘리면 그에 따라 고려해야 할 부분들이 있다. 그중 각 코어간의 데이터의 일관성을 유지하는 것이 큰 문제

이다. 이를 해결하기 위해 일관성 관리 유닛을 둔다. 본 논문에서는 이 일관성 관리 유닛 내부에 버퍼를 두어 성능 저하를 줄일 수 있었다. 그러나 지금의 상태에서도 캐시의 태그의 복사본을 일관성 관리 유닛이 가지고 있어야 함으로 인해 부가적인 캐시와 일관성 관리 유닛 사이의 데이터 전송이 발생하게 된다.

Acknowledgement

본 연구는 특허청 반도체 사무국의 핵심반도체 설계 재산권 창출 촉진사업의 지원과 반도체설계교육센터 (IDEC)의 도움을 받아 이루어 졌습니다.

참고문헌

- [1] D. Geer, Industry Trends: Chip Makers Turn to Multicore Processors, *Computer*, Vol. 38, no. 5, pp. 11-13, 2005
- [2] Per Stenström, A Survey of Cache Coherence Schemes for Multiprocessors, *Computer*, Vol. 23, no. 6, pp. 12-24, 1990