



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2022년06월02일

(11) 등록번호 10-2404643

(24) 등록일자 2022년05월27일

(51) 국제특허분류(Int. Cl.)  
G06F 13/40 (2006.01) G06F 12/02 (2018.01)

(52) CPC특허분류  
G06F 13/40 (2013.01)  
G06F 12/0292 (2013.01)

(21) 출원번호 10-2017-0013952

(22) 출원일자 2017년01월31일

심사청구일자 2022년01월27일

(65) 공개번호 10-2018-0012180

(43) 공개일자 2018년02월05일

(30) 우선권주장

62/367,062 2016년07월26일 미국(US)

15/272,339 2016년09월21일 미국(US)

(56) 선행기술조사문헌

US06594728 B1

US08885380 B2

JP2015503160 A

JP2013016244 A

(73) 특허권자

삼성전자주식회사

경기도 수원시 영통구 삼성로 129 (매탄동)

(72) 발명자

스톡스테일, 타일러

미국 노스캐롤라이나주 27607 롤리 크래프톤 웨이 1619

창, 무-티엔

미국 캘리포니아주 95051 산타클라라 토리노 비아 2920

정, 홍종

미국 캘리포니아주 95032 로스 가토스 칼튼 애비뉴 120 6호

(74) 대리인

특허법인 고려

전체 청구항 수 : 총 20 항

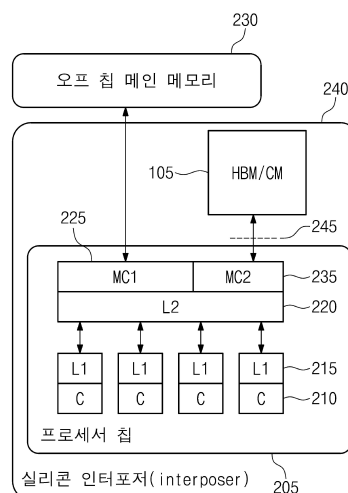
심사관 : 김세영

(54) 발명의 명칭 인-메모리 캐쉬 관리자를 구비한 HBM

### (57) 요약

높은 대역폭 메모리를 캐쉬 메모리로 사용하기 위한 시스템 및 방법이 제공된다. 높은 대역폭 메모리는 로직 다이 및 로직 다이 상에 스택된 복수의 DRAM 메모리 다이들을 포함한다. 로직 다이는 캐쉬 관리자를 포함하고, 캐쉬 관리자는 JESD235A 표준에 부합하는 외부 인터페이스를 통해한 외부 시스템들에 대한 인터페이스일 수 있고, 어드레스 변환기, 명령 변환기, 및 태그 비교기를 포함할 수 있다. 어드레스 변환기는 외부 인터페이스를 통해서 수신된 각각의 물리적 어드레스를 태그 값, 메모리 다이들 스택 내의 태그 어드레스, 및 메모리 다이들 스택 내의 데이터 어드레스로 변환한다. 태그 비교기는 어드레스 변환기에 의해서 생성되는 태그 값이 태그 어드레스에 저장된 태그값과 매칭되는지 여부에 따라 캐쉬 적중 및 캐쉬 미스가 발생할지를 결정한다.

대표도 - 도2a



## 명세서

### 청구범위

#### 청구항 1

복수의 메모리 다이들을 포함하는 메모리 스택; 및

로직 다이를 포함하고,

상기 메모리 다이들은 상기 로직 다이 상에 스택되고, 상기 로직 다이에 연결되고,

상기 로직 다이는:

호스트 프로세서에 대한 외부 인터페이스를 포함하고,

명령의 일부로서, 상기 외부 인터페이스를 통해서 수신된 어드레스를 변환하여 제 1 태그값과 태그 어드레스를 생성하고, 상기 태그 어드레스에 대응하는 상기 메모리 스택의 위치로부터 제 2 태그값을 폐치하고, 상기 제 1 태그값과 상기 제 2 태그값을 비교하고, 그리고 상기 제 1 태그값과 상기 제 2 태그값을 비교하는 것에 기반하여 상기 명령을 실행하는 캐쉬 관리자를 포함하는 메모리 시스템.

#### 청구항 2

제 1 항에 있어서,

상기 캐쉬 관리자는 상기 외부 인터페이스를 통해서 수신된 상기 어드레스를 변환하여 상기 제 1 태그값, 상기 메모리 스택의 데이터 어드레스, 및 상기 메모리 스택의 상기 태그 어드레스를 생성하는 어드레스 변환기를 포함하는 메모리 시스템.

#### 청구항 3

제 2 항에 있어서,

상기 캐쉬 관리자는 상기 외부 인터페이스를 통해서 수신된 읽기 명령에 응답하여 상기 제2 태그값을 폐치하는 제 1 명령 및 데이터 워드를 폐치하는 제 2 명령을 생성하는 명령 변환기를 포함하는 메모리 시스템.

#### 청구항 4

제 3 항에 있어서,

상기 캐쉬 관리자는 캐쉬 적중 신호를 생성하는 태그 비교기를 포함하고,

상기 캐쉬 적중 신호는 상기 제 1 태그값이 상기 제 2 태그값과 같은 경우 참 값을 가지고, 그리고 상기 제 1 태그값이 상기 제 2 태그값과 같지 않은 경우 거짓 값을 가지는 메모리 시스템.

#### 청구항 5

제 4 항에 있어서,

상기 태그 비교기는 상기 외부 인터페이스의 제 1 핀을 통해서 상기 캐쉬 적중 신호를 전송하는 메모리 시스템.

#### 청구항 6

제 5 항에 있어서,

상기 캐쉬 관리자는 상기 외부 인터페이스의 제 2 핀을 통해서 데이터 비트의 값 및/또는 유효 비트의 값을 전송하는 메모리 시스템.

#### 청구항 7

제 5 항에 있어서,

상기 캐쉬 관리자는 제 1 시간 구간 동안 상기 제 1 편을 통해서 상기 캐쉬 적중 신호를 전송하고, 제 2 시간 구간 동안 상기 제 1 편을 통해서 더티 비트의 값을 전송하는 메모리 시스템.

#### 청구항 8

제 3 항에 있어서,

상기 캐쉬 관리자는 의사 채널(pseudo channel)을 통해서 상기 제 1 명령을 실행하는 메모리 시스템.

#### 청구항 9

제 3 항에 있어서,

상기 캐쉬 관리자는 병렬 모드 동작이 선택되는 것 또는 직렬 모드 동작이 선택되는 것을 지시하는 모드 선택터를 포함하고,

상기 캐쉬 관리자는 상기 모드 선택터가 병렬 모드 동작이 선택되는 것을 지시하는 경우 상기 제 2 명령과 병렬로 상기 제 1 명령을 실행하고, 그리고 상기 모드 선택터가 직렬 모드 동작이 선택되는 것을 지시하는 경우 상기 제 2 명령을 실행하기 전에 상기 제 1 명령을 실행하는 메모리 시스템.

#### 청구항 10

제 9 항에 있어서,

상기 모드 선택터는 상기 외부 인터페이스를 통해서 제어되는 메모리 시스템.

#### 청구항 11

제 1 항에 있어서,

상기 메모리 다이들의 제 1 뱅크 내에 저장되고, 서로 다른 의사 채널들(pseudo channels)을 통해서 액세스 가능한 임의의 두 개의 데이터 워드들에 대해서, 두 개의 대응하는 태그들이 상기 메모리 스택의 서로 다른 서브어레이들에 저장되는 메모리 시스템.

#### 청구항 12

제 1 항에 있어서,

상기 외부 인터페이스는 JESD235A(Joint Electron Device Engineering Council standard)를 준수하여 동작하는 메모리 시스템.

#### 청구항 13

호스트 프로세서;

상기 호스트 프로세서에 연결된 제 1 메모리 시스템; 및

상기 호스트 프로세서에 연결된 제 2 메모리 시스템을 포함하고,

상기 제 1 메모리 시스템은

복수의 메모리 다이들을 포함하는 메모리 스택; 및

로직 다이를 포함하고,

상기 메모리 다이들은 상기 로직 다이 상에 스택되고, 상기 로직 다이에 연결되고,

상기 로직 다이는:

상기 호스트 프로세서에 대한 외부 인터페이스를 포함하고,

명령의 일부로서, 상기 외부 인터페이스를 통해서 수신된 어드레스를 변환하여 제 1 태그값과 태그 어드레스를 생성하고, 상기 태그 어드레스에 대응하는 상기 메모리 스택의 위치로부터 제 2 태그값을 폐치하고, 상기 제 1 태그값과 상기 제 2 태그값을 비교하고, 그리고 상기 제 1 태그값과 상기 제 2 태그값을 비교하는 것에 기반하

여 상기 명령을 실행하는 캐쉬 관리자를 포함하고,

상기 제 2 메모리 시스템은 상기 제 1 메모리 시스템에 대한 보조 기억 장치로서 동작하는 프로세싱 시스템.

#### 청구항 14

제 13 항에 있어서,

상기 캐쉬 관리자는 상기 외부 인터페이스를 통해서 수신된 상기 어드레스를 변환하여 상기 제 1 태그값, 상기 메모리 스택의 데이터 어드레스, 및 상기 메모리 스택의 상기 태그 어드레스를 생성하는 어드레스 변환기를 포함하는 프로세싱 시스템.

#### 청구항 15

제 14 항에 있어서,

상기 캐쉬 관리자는 상기 외부 인터페이스를 통해서 수신된 읽기 명령에 응답하여 상기 제2 태그값을 폐치하는 제 1 명령 및 데이터 워드를 폐치하는 제 2 명령을 생성하는 명령 변환기를 포함하는 프로세싱 시스템.

#### 청구항 16

제 15 항에 있어서,

상기 캐쉬 관리자는 캐쉬 적중 신호를 생성하는 태그 비교기를 포함하고,

상기 캐쉬 적중 신호는 상기 제 1 태그값이 상기 제 2 태그값과 같은 경우 참 값을 가지고, 그리고 상기 제 1 태그값이 상기 제 2 태그값과 같지 않은 경우 거짓 값을 가지는 프로세싱 시스템.

#### 청구항 17

제 13 항에 있어서,

상기 외부 인터페이스는 JESD235A(Joint Electron Device Engineering Council standard)를 준수하여 동작하는 프로세싱 시스템.

#### 청구항 18

복수의 메모리 다이들 및 로직 다이를 포함하는 메모리 스택을 동작시키는 방법에 있어서,

상기 메모리 다이들은 상기 로직 다이 상에 스택되고, 상기 로직 다이에 연결되고,

상기 로직 다이는 호스트 프로세서에 대한 외부 인터페이스를 포함하고,

상기 방법은:

상기 로직 다이에 의해, 명령의 일부로서, 상기 외부 인터페이스를 통해서 수신된 어드레스를 변환하여 제 1 태그값, 상기 메모리 스택의 데이터 어드레스, 및 상기 메모리 스택의 태그 어드레스를 생성하는 단계;

상기 태그 어드레스에 대응하는 상기 메모리 스택의 위치로부터 제 2 태그값을 폐치하는 단계;

상기 제 1 태그값과 상기 제 2 태그값을 비교하는 단계; 및

상기 제 1 태그값과 상기 제 2 태그값을 비교하는 것에 기반하여 상기 명령을 실행하는 단계를 포함하는 방법.

#### 청구항 19

제 18 항에 있어서,

상기 로직 다이에 의해서, 상기 외부 인터페이스를 통해서 수신된 읽기 명령에 응답하여 상기 제2 태그값을 폐치하는 제 1 명령 및 데이터 워드를 폐치하는 제 2 명령을 생성하는 단계를 더 포함하는 방법.

#### 청구항 20

제 19 항에 있어서,

상기 로직 다이에 의해서, 캐쉬 적중 신호를 생성하는 단계를 더 포함하고,

상기 캐쉬 적중 신호는 상기 제 1 태그값이 상기 제 2 태그값과 같은 경우 참 값을 가지고, 그리고 상기 제 1 태그값이 상기 제 2 태그값과 같지 않은 경우 거짓 값을 가지는 방법.

## 발명의 설명

### 기술 분야

[0001] 본 발명의 일 실시예에 따른 실시예의 하나 이상의 측면은 높은 대역폭 메모리에 관련되고, 좀 더 상세하게는 높은 대역폭 메모리를 캐쉬 메모리로 사용하기 위한 시스템 및 방법에 관련된다.

### 배경 기술

[0003] HBM(High Bandwidth Memory)은 3차원 스택형 동적 RAM(DRAM)을 위한 고성능 (RAM) 인터페이스이다. HBM을 캐쉬 메모리로 사용하는 관련 기술 시스템들은 호스트 상에 캐쉬 관리자를 가지면서 캐쉬 관리 기능들을 수행할 수 있다. 이러한 배열은 호스트와 호스트 및 HBM 사이의 인터페이스에 부담이 될 수 있다.

[0004] 그러므로, HBM을 캐쉬 메모리를 사용하는 데에 있어서 개선된 시스템 및 방법이 필요하다.

### 발명의 내용

#### 해결하려는 과제

[0006] 본 발명의 일 실시예에 따른 실시예의 하나 이상의 측면은 높은 대역폭 메모리에 관련되고, 좀 더 상세하게는 높은 대역폭 메모리를 캐쉬 메모리로 사용하기 위한 시스템 및 방법에 관련된다.

#### 과제의 해결 수단

[0008] 본 발명의 일 실시예에 따른 실시예의 하나 이상의 측면은 높은 대역폭 메모리를 캐쉬 메모리로 사용하기 위한 시스템 및 방법에 대한 것이다. 높은 대역폭 메모리는 로직 다이 및 로직 다이 상에 스택된, 복수의 동적 랜덤 액세스 메모리 다이들을 포함한다. 로직 다이는 캐쉬 관리자를 포함하고, 캐쉬 관리자는 JESD235A 표준에 부합하는 외부 인터페이스를 통해 외부 시스템들과 인터페이스할 수 있고, 어드레스 변환기, 명령 변환기, 및 태그 비교기를 포함할 수 있다. 어드레스 변환기는 외부 인터페이스를 통해서 수신된 각각의 물리적 어드레스를 태그 값, 메모리 다이들 스택 내의 태그 어드레스, 및 메모리 다이들 스택 내의 데이터 어드레스로 변환한다. 태그 비교기는 어드레스 변환기에 의해서 생성되는 태그 값이 태그 어드레스에 저장된 태그 값과 매칭되는지 여부에 따라 캐쉬 적중 및 캐쉬 미스가 발생했는지를 판별한다. 명령 생성기는 명령들을 생성시킬 수 있다. 예를 들어, 외부 인터페이스를 통해서 쓰기 명령을 수신함에 따라, 명령 생성기는 먼저 명령을 생성하여 태그 값을 폐치하여, 캐쉬 적중이 발생했는지 여부를 판단할 수 있고, 그리고 캐쉬 적중이 발생 했으면 명령 생성기는 쓰기 명령을 생성시킨다.

[0009] 본 발명의 일 실시예에 따라 제공되는 메모리 시스템은 복수의 메모리 다이들을 포함하는 메모리 스택, 로직 다이를 포함하고,

[0010] 상기 메모리 다이들은 상기 로직 다이 상에 스택킹되고, 상기 로직 다이에 연결되고, 상기 로직 다이는 상기 메모리 시스템에 대한 외부 인터페이스를 구비하고, 상기 로직 다이는 캐쉬 관리자를 포함한다.

[0011] 일 실시예에서, 상기 캐쉬 관리자는 상기 외부 인터페이스를 통해서 수신된 어드레스를 변환하여 제1 태그값, 상기 메모리 스택의 데이터 어드레스, 및 상기 메모리 스택의 태그 어드레스를 생성하는 어드레스 변환기(translator)를 포함한다.

[0012] 일 실시예에서, 상기 캐쉬 관리자는 상기 외부 인터페이스를 통해서 수신되는 읽기 명령에 응답하여 태그를 폐치하는 제1 명령 및 데이터 워드를 폐치하는 제2 명령을 생성하는 명령 변환기를 포함한다.

[0013] 일 실시예에서, 상기 캐쉬 관리자는 캐쉬 적중 신호를 생성시키는 태그 비교기를 포함하고, 상기 캐쉬 적중 신호는 상기 제1 태그값이 상기 제1 명령의 실행으로부터의 결과값과 동일한 경우 참 값을 가지고 상기 제1 태그값이 상기 제1 명령의 실행으로부터 결과값과 동일하지 않은 경우 거짓 값을 가진다.

- [0014] 일 실시예에서, 상기 태그 비교기는 상기 외부 인터페이스의 제1 핀을 통해서 상기 캐쉬 적중 신호를 전송한다.
- [0015] 일 실시예에서, 상기 캐쉬 관리자는 상기 외부 인터페이스의 제2 핀을 통해서 더티 비트의 값 및/또는 유효 비트의 값을 전송한다.
- [0016] 일 실시예에서, 상기 캐쉬 관리자는 제1 구간동안 상기 제1 핀을 통해서 상기 캐쉬 적중 신호를 전송하고, 제2 구간동안 상기 제1 핀을 통해서 더티 비트 값을 전송한다.
- [0017] 일 실시예에서, 상기 캐쉬 관리자는 의사 채널을 통해서 상기 제1 명령을 실행한다.
- [0018] 일 실시예에서, 상기 캐쉬 관리자는 병렬 모드 동작이 선택되거나 직렬 모드 동작이 선택되는 지를 지시하는 모드 선택터를 포함하고, 상기 캐쉬 관리자는 상기 모드 선택터가 병렬 모드 동작이 선택되는 것을 지시하는 경우 상기 제2 명령에 병렬적인 상기 제1 명령을 실행하고, 상기 모드 선택터가 직렬 모드 동작이 선택되는 것을 지시하는 경우, 상기 제2 명령을 실행하기 전에 상기 제1 명령을 실행한다.
- [0019] 일 실시예에서 상기 모드 선택터는 상기 외부 인터페이스를 통해서 제어되는 메모리 시스템.
- [0020] 일 실시예에서, 상기 메모리 다이들 내의 제1 뱅크에 저장되고, 서로 다른 의사 채널들을 통해 액세스가능한 임의의 두개의 데이터 워드들 대해서, 태그들에 대응하는 두 개는 상기 스택의 서로 다른 서브어레이들에 저장된다.
- [0021] 일 실시예에서, 상기 외부 인터페이스는 JESD235A(Joint Electron Device Engineering Council standard)를 준수하여 동작한다.
- [0022] 본 발명의 일 실시예에 따라 제공되는 프로세싱 시스템은 호스트 프로세서, 상기 호스트 프로세서에 연결된 제1 메모리 시스템, 및 상기 호스트 프로세서에 연결되는 제2 메모리 시스템을 포함하고, 제1 메모리 시스템은 복수의 메모리 다이들을 포함하는 메모리 스택 및 로직 다이를 포함하고, 상기 메모리 다이들은 상기 로직 다이 상에 스택킹되고, 상기 로직 다이에 연결되고, 상기 로직 다이는 상기 메모리 시스템에 대한 외부 인터페이스를 구비하고, 상기 로직 다이는 캐쉬 관리자를 포함하고, 상기 제2 메모리 시스템은 상기 제1 메모리 시스템에 대한 보조 기억 장치로써 동작한다.
- [0023] 일 실시예에서, 상기 캐쉬 관리자는 상기 외부 인터페이스를 통해서 수신된 어드레스를 제1 태그값, 상기 메모리 스택의 데이터 어드레스, 및 상기 메모리 스택의 태그 어드레스로 변환하는 어드레스 변환기(translator)를 포함한다.
- [0024] 일 실시예에서, 상기 캐쉬 관리자는 상기 외부 인터페이스를 통해서 수신되는 읽기 명령에 응답하여 태그를 폐치하는 제1 명령 및 데이터 워드를 폐치하는 제2 명령을 생성시키는 명령 변환기를 포함한다.
- [0025] 일 실시예에서, 상기 캐쉬 관리자는 캐쉬 적중 신호를 생성시키는 태그 비교기를 포함하고, 상기 캐쉬 적중 신호는 상기 제1 태그값이 상기 제1 명령의 실행으로부터의 결과값과 동일한 경우 참 값 및 상기 제1 태그 값이 상기 제1 명령의 실행으로부터 결과값과 동일하지 않은 경우 거짓 값을 가진다.
- [0026] 일 실시예에서, 상기 외부 인터페이스는 JESD235A(Joint Electron Device Engineering Council standard)를 준수하여 동작한다.
- [0027] 본 발명의 일 실시예에 따라 제공되는 복수의 메모리 다이들과 로직 다이를 포함하는 메모리 스택을 동작시키는 방법에 따르면, 상기 메모리 다이들은 상기 로직 다이 상에 스택킹되고, 상기 로직 다이에 연결되고, 상기 로직 다이는 상기 메모리 시스템에 대한 외부 인터페이스를 구비하고, 상기 방법은 상기 로직 다이에 의해서, 상기 외부 인터페이스를 통해서 수신된 어드레스를 변환하여 제1 태그값, 상기 메모리 스택의 데이터 어드레스, 및 상기 메모리 스택의 태그 어드레스를 생성하는 단계를 포함한다.
- [0028] 일 실시예에서, 상기 방법은 상기 로직 다이에 의해서 상기 외부 인터페이스를 통해서 수신되는 읽기 명령에 응답하여 태그를 폐치하는 제1 명령 및 데이터 워드를 폐치하는 제2 명령을 생성하는 단계를 포함한다.
- [0029] 일 실시예에서, 상기 방법은 상기 로직 다이에 의해서 캐쉬 적중 신호를 생성시키는 단계를 포함하고, 상기 캐쉬 적중 신호는 상기 제1 태그값이 상기 제1 명령의 실행으로부터의 결과값과 동일한 경우 참 값을 가지고 상기 제1 태그 값이 상기 제1 명령의 실행으로부터 결과값과 동일하지 않은 경우 거짓 값을 가진다.

## 발명의 효과

[0031] 로직 다이에서 캐쉬 관리자를 포함하는 것은 두번째 명령(쓰기 명령)이 호스트 프로세서가 아닌 오히려 로직 다이에서 생성하게 함으로써, 효율성을 개선할 수 있다.

### 도면의 간단한 설명

[0033] 본 발명의 특정 실시예의 상기한 내용 및 다른 내용들, 특징들, 및 이점들은 첨부 도면을 참조하여, 후술할 상세한 설명에서 더 명확하게 설명될 것이다. 상기 첨부 도면에서,

도 1은 본 발명의 일 실시예에 따른 고 대역폭 메모리 스택의 사시도이다.

도 2A는 본 발명의 일 실시예에 따른 고 대역폭 메모리 스택을 레벨 3 캐쉬로 채택하는 프로세싱 시스템의 블록도이다.

도 2B는 본 발명의 일 실시예에 따른 고 대역폭 메모리 스택의 블록도이다.

도 3은 본 발명의 일 실시예에 따른 고 대역폭 메모리 스택의 블록도이다.

도 4A는 본 발명의 일 실시예에 따른 저장도, 및

도 4B는 본 발명의 일 실시예에 따른 저장도이다.

### 발명을 실시하기 위한 구체적인 내용

[0034] 이하 첨부된 도면과 함께 설명되는 상세한 설명은 본 발명에 따라 제공되는 인-메모리 캐쉬 관리자를 구비한 고 대역폭 메모리의 예시적인 실시 예들의 설명이고, 본 발명이 구성되고 이용되는 하나의 유형만을 나타내는 것은 아니다. 설명은 설명되는 실시예들에 연결되는 본 발명의 특징들을 설명한다. 그러나 동일한 또는 균등한 기능 및 구조들이 본 발명의 범위 및 사상을 포함하는 서로 다른 실시예들에 의해서 수행도리 수 있음은 자명할 것이다. 본 명세서의 다른 부분에서 언급되는 바와 같이 동일한 요소 번호는 동일한 요소 또는 특징을 나타내기 위한 것이다.

[0035] 고 대역폭 메모리(HBM)는 고성능 3차원의 스택형 동적 RAM(DRAM)이다. 2세대 HBM은 스택마다 최대 8개의 다이를 포함할 수 있고, 최대 2GT/s(giga transfers per second)의 핀 전송 속도를 제공할 수 있다. 인터페이스는 총 1024 비트 폭의 액세스를 위해서 각각이 128 비트 폭의 8개의 채널을 포함할 수 있다. 2세대 HBM은 패키지 당 256GB/s의 메모리 대역폭에 도달할 수 있고, 패키지 당 최대 8GB의 저장 능력을 가질 수 있다. 2세대 HBM의 인터페이스는 표준 SESD235A로써 JEDEC(Joint Electron Device Engineering Council)에서 승인된 표준을 준수할 수 있다.

[0036] 도 1을 참조하면, HBM 스택(105)의 물리적 구성은 로직 다이(110) 및 로직 다이(110)의 최상층에 스택된 복수의 DRAM 다이들(즉 8개의 이러한 다이들)을 포함하는 3차원 DRAM 또는 “DRAM 스택(115)”을 포함할 수 있다. 상호 연결들은 TSVs(through-silicon via)를 통해서 스택 내에 형성된다. HBM 스택 관련 기술은 로직 다이에서 연결들, 신호 조절 회로를 포함하고, 그 결과로, DRAM 채널 인터페이스들이, 실질적으로 변경되지 않고, HBM의 외부 인터페이스에서 호스트 프로세서로 제공된다.

[0037] 도 2A를 참조하면, HBM 스택(105)은 호스트 프로세서(205, 즉 CPU 또는 GPU)에 연결될 수 있다. 호스트 프로세서(205)는 각각이 각자의 레벨 1 캐쉬(215)를 가지는 복수의 코어들(210)을 포함할 수 있다. 레벨 2 캐쉬(220)는 레벨 1 캐쉬들(215)에 연결될 수 있고, 제1 메모리 제어기(225)는 인터페이스를 오프 칩 메인 메모리(230)에 제공할 수 있다. 제2 메모리 제어기(235)는 인터페이스를 HBM 스택(105)에 제공할 수 있다. HBM 스택(105)은 HBM 스택(105)의 로직 다이에서 캐쉬 관리자(CM)를 포함할 수 있다. HBM 스택(105)는 그것의 집적 캐쉬 관리자와 함께 호스트 프로세서(205)에 의해서 레벨 3 캐쉬(또는 레벨 3 캐쉬를 또한 가지는 시스템에서 레벨 4 캐쉬로)로 채택될 수 있다. HBM 인터페이스(245)는 JESD235A에 부합하는 인터페이스일 수 있고, 즉, 이는 JESD235A 표준에 의해서 지정된 컨덕터들(conductors) 및 시그널링 프로토콜들(signaling protocols)을 제공할 수 있다.

[0038] 도 2B를 참조하면, 일부 실시예들에서, HBM 스택(105)은 로직 다이(110)를 포함할 수 있고, 로직 다이(110)는 도 2B에서 ch0 내지 ch7로 나타내어진 채널들로 칭해지는 8개의 내부 인터페이스들을 통해서 DRAM 스택(115) 내의 DRAM에 연결될 수 있다.

[0039] 도 3을 참조하면, 일 실시예에서, HBM 스택(105)은 전술한 바와 같이 DRAM 스택(115) 및 로직 다이(110)를 포함할 수 있고, 로직 다이(110)는 캐쉬 관리자를 구현하는 복수의 구성요소들을 포함할 수 있다. HBM 인터페이스(245)에서 명령 및 어드레스 라인들(305)은 HBM 스택(105)의 로직 다이(110)에서 명령 변환기(310) 및 어드레스



변환기(315)에 연결될 수 있다. 명령 및 어드레스 라인들(305)는 예를 들어, HBM 인터페이스(245)의 8개의 채널들 각각을 위해서 6개의 행의 명령/어드레스 라인들 및 8개의 열의 명령/어드레스 라인들을 포함할 수 있다.

[0040] 동작에서, 어드레스 변환기(315)는 명령(즉 읽기 명령 또는 쓰기 명령)이 수행되기 위한 물리적 메모리 어드레스를 주기적으로 수신할 수 있다. 이후, 어드레스 변환기(315)는 어드레스를 태그 값, 태그 어드레스, 및 데이터 어드레스로 변환한다. 태그 값은 “캐쉬 적중”이 발생하였는지를 결정하는데 이용될 수 있다. (즉 캐쉬 내의 어드레스가 HBM 인터페이스(245)를 통해 수신된 어드레스에 현재 할당되어 있는지 여부를 결정하는데 이용될 수 있다.) 예를 들어, 캐쉬 관리자는 태그 어드레스에서의 태그 값을 읽을 수 있고(혹은 폐치), 그것을 어드레스 트랜스레이터(315)에 의해서 생성된 태그 값과 비교할 수 있다. (즉, 아래에서 상세히 설명되는 태그 비교기(320)를 통해서) 만약 수신된 물리적 어드레스로부터 형성된 태그 값이(어드레스 변환기(315)에 의해서) DRAM 스택(115)에서 태그 어드레스에 저장된 태그 값에 매칭되는 경우, 캐쉬 적중은 발생한다. 즉 캐쉬 내의 어드레스는 프로세서의 물리적 메모리 공간에서의 수신된 어드레스에 현재 할당되어 있다. 수신된 물리적 어드레스로부터 형성된 태그 값이 DRAM 스택(115)의 태그 어드레스에 저장된 태그 값에 매칭되지 않는 경우(본 상태를 “캐쉬 미스”라고 본 명세서에서는 칭해진다.), 캐쉬 내 어드레스는 프로세서의 물리적 메모리 공간에서의 수신된 어드레스에 현재 할당되어 있지 않다.

[0041] 태그 비교기(320)는 비교하는 것에 사용될 수 있다. 즉, 상기 비교는 수신된 물리적 어드레스로부터 형성된 태그 값과 태그 어드레스에 저장된 태그 값을 비교하는 것이다. 태그 비교기(320)의 출력은 캐쉬 적중 신호로 지칭되는 신호일 수 있고, 캐쉬 적중 신호는 캐쉬 적중이 발생하는 경우 참 값(즉 이진 “1” 값) 및 캐쉬 미스가 발생하는 경우 거짓 값(즉 이진 “0” 값)을 가질 수 있다.

[0042] 명령 변환기(310)는 HBM 인터페이스(245)를 통해서 수신된 명령들에 응답하여 명령들을 생성하여 DRAM 스택(115)에서 실행될 수 있다. 예를 들어, HBM 인터페이스(245)를 통해서 수신된 명령이 읽기 명령이면, 명령 변환기(310)는 명령을 생성하여 데이터 어드레스에 저장된 데이터 워드를 읽고, 명령을 생성하여 태그 어드레스에 저장된 태그 값을 읽을 수 있다. 상기 명령들 각각은 복수의 마이크로 동작들을 포함(혹은 구성되는)하고, 즉 읽기 명령은 읽기 동작이 후속하는 활성 동작을 포함할 수 있다.

[0043] HBM(245)를 통해서 수신되는 명령이 쓰기 명령인 경우, 명령 변환기(310)는 먼저 명령을 생성시켜 태그 어드레스에 저장된 태그를 읽고, 태그 값이 어드레스 변환기(315)에 의해서 생성된 태그 값과 매칭되는 경우, 쓰기 명령을 생성시켜, DRAM 스택(115)에 데이터 쓰기 동작을 한다. 상기한 바와 같이, 호스트 프로세서(205)가 캐쉬 관리자를 구현하는 것을 불필요하게 하며, 나아가 로직 다이(110)에서 캐쉬 관리자를 포함하는 것은 두 번째 명령(쓰기 명령)이 호스트 프로세서(205)가 아닌 오히려 로직 다이(110)에서 생성되게 함으로써, 효율성을 개선할 수 있다.

[0044] 태그 값들 및 데이터를 폐치하는 것은 병렬적 또는 직렬적으로 수행될 수 있다. 예를 들어, 캐쉬 관리자가 병렬 모드에서 동작하는 경우, 태그 값 및 데이터는 병렬적으로 폐치되고, 폐치된 태그 값은 어드레스 변환기(315)에 의해서 생성된 태그 값과 태그 비교기(320)에 의해서 비교될 수 있고, 그리고, 상기 두 개의 태그 값들이 매칭되는 경우, HBM 인터페이스(245)를 통해서 데이터 읽기는 반환될 수 있다. 그렇지 않은 경우에는 캐쉬 미스는 아래에 후술할 바와 같이 HBM 인터페이스(245)를 통해서 호스트 프로세서(205)에 시그널링될 수 있다. 캐쉬 관리자가 직렬 모드로 동작하는 경우, 태그 값이 먼저 폐치되고, 폐치된 태그 값은 어드레스 변환기(315)에 의해서 생성된 태그 값과 태그 비교기(320)에 의해서 비교될 수 있고, 그리고, 두 개의 태그 값이 매칭되는 경우, 데이터는 폐치되고, HBM 인터페이스(245)를 통해 반환될 수 있다. 그렇지 않으면 캐쉬 미스는 HBM 인터페이스(245)를 통해 호스트 프로세서(205)에 시그널링될 수 있다. 직렬 모드에서의 동작은 병렬 모드에서의 동작보다 더 전력 효율적일 수 있다. 왜냐하면, 직렬 모드에서 데이터 폐치 동작들은 오직 캐쉬 적중의 경우에만 수행되기 때문이다. 병렬 모드에서 동작들은 직렬 모드에서 동작보다 빠를 수 있다. 왜냐하면 병렬 모드에서 태그 값 폐치 및 데이터 폐치는 동시해 수행될 수 있다. 캐쉬 관리자는 모드 선택기(즉 제어 레지스터의 비트)를 포함할 수 있고, 모드 선택기는 캐쉬 관리자가 병렬 모드에서 동작하는지 직렬 모드에서 동작하는지를 제어할 수 있다. 모드 선택기는 HBM 인터페이스(245)를 통해 (호스트 프로세서(205)에 의해서) 제어될 수 있다. (즉 새로운 값을 제어 레지스터에 기록하는 명령에 의해서)

[0045] 캐쉬 관리자는 각각의 태그와 함께, 2 비트의 메타데이터를 저장할 수 있다. 상기 2개 비트들은 (i) 대응하는 데이터 워드가 유효인지 무효인지를 지시하는 비트(“유효 비트”) (ii) 대응하는 데이터 워드가 클린(clean) 또는 더티(dirty)한지를 지시하는 비트(“더티 비트”)이다. 보조 기억 저장소내의 대응하는 데이터가 업데이트 되는 것(다시 말하면 클린한(clean))없이 캐쉬 내의 데이터가 업데이트 되는 경우에, 캐쉬 내 데이터는 더티로



간주되고, 그리고, 캐쉬내 대응하는 데이터가 업데이트되는 것(다시 말하면 유효한)없이 보조 기억 저장소내의 데이터가 업데이트되는 경우, 캐쉬 내의 데이터는 무효한 것으로 간주된다.

[0046] 게다가 HBM 인터페이스(245)를 통해서 수신된 명령들은 전술한 바와 같이, 캐쉬 적중 신호의 참 또는 거짓 값(캐쉬 적중 또는 캐쉬 미스에 각각 대응하는)의 결과를 가져온다. HBM 인터페이스(245)를 통해서 HBM 스택(105)에 의해서 수신된 임의의 명령이 완료되는 경우, 캐쉬 관리자는 3가지 값을 생성시키는데, 하나는 캐쉬 적중 신호를 위한 것이고, 하나는 더티 비트를 위한 것이고, 하나는 유효 비트를 위한 것이다. 상기 값들은 다른 기능들을 위해서 사용되지 않는 HBM 인터페이스(245)의 하나 이상의 핀들을 사용하여 HBM 인터페이스(245)를 통해서 호스트 프로세서(205)와 통신될 수 있다. 상기 핀들은 (i) 8개 채널들 각각의 212개 핀들중 임의의 핀이 아니고, 또한 (ii) RESET, TEMP[2:0], 또는 CATTRIP 핀들 중 임의의 핀도 아니다. 후에 사용되기 위해서 예약된 핀(RFU 핀들)으로써, JESD235A 표준에 의해서 정의된 핀들이 사용될 수 있다. 예를 들어, 일 실시예에서, RFU 핀은 명령이 실행된 후 데이터 버스트(data burst)의 제1 데이터 사이클동안 캐쉬 적중 신호를 전송하는데 사용되고, 데이터 버스트의 다음 데이터 사이클동안 더티 비트의 값을 전송하는 데 사용된다. 캐쉬 적중 신호 및 더티 비트의 전송은 데이터 버스트의 데이터의 전송에 동기화될 수 있다. 일부 실시예들에서, 복수의 RFU 핀들은 캐쉬 적중 신호, 더티 비트, 및/또는 유효 비트를 전송하는 데에 사용된다.

[0047] 캐쉬 내에서 데이터를 무효화 하기 위해서, 호스트 프로세서(205)는 “무효화” 신호를 무효화될 데이터의 어드레스와 함께 HBM 인터페이스(245)를 통해 로직 다이(110)에 전송할 수 있다. 무효화 신호는 캐쉬 적중 신호를 호스트 프로세서(205)에 전송하는 데에 사용되는 HBM 인터페이스(245)의 동일한 핀(즉 RFU 핀)을 통해서 전송될 수 있다. 어드레스는 HBM 인터페이스(245)의 CA 버스를 통해 전송될 수 있다. 그리고 상기 정보를 통해서, 로직 다이(110)는 DRAM 스택(115)에 저장된 대응하는 유효 비트를 업데이트할 수 있다.

[0048] 일부 실시예들에서, RFU 핀들은 또한 캐쉬 일관성(coherency)을 유지하는 데에 사용된다. 즉 도 2A에서와 같이 멀티 코어 시스템에서 보조 기억 저장소(즉, 오프 칩 메인 메모리(230))와 캐쉬 사이에서의 일관성을 유지시키는 것을 말한다. 상기 멀티 코어 시스템에서, 레벨 3 캐쉬(HBM 스택(105)에서 구현됨) 및 보조 기억 저장소는 코어들(210) 각각에 의해서 읽혀지거나 수정될 수 있다.

[0049] 로직 다이(110)의 스케줄러(325)는 명령 변환기(310) 및 어드레스 변환기(315) 각각으로부터 명령들 및 어드레스들을 수신할 수 있고, DRAM 스택(115) 상에서 상기 명령들의 실행을 예정할 수 있다. 로직 다이(110)의 데이터 버퍼(330)는 데이터가 HBM 인터페이스(245)를 통해서 수신된 후 및/또는 데이터가 DRAM 스택(115)으로부터 읽혀진 후에 일시적으로 데이터를 저장하는데 사용될 수 있다. 스케줄러(325) 및 데이터 버퍼(330)는 어떤 비율로의 변화들을 수용하는 것을 도와주고 상기 변화되는 비율은 다음과 같다. (i) HBM 인터페이스(245)를 통해서 명령들이 수신되고, (ii) DRAM 스택(115) 상에서 명령이 실행되고, (iii) HBM 인터페이스(245)를 통해서 데이터가 전송되거나 수신되고, 및 (iv) DRAM 스택(115)으로부터 데이터가 읽혀지거나 DRAM 스택(115)에 데이터가 쓰여진다.

[0050] JESD235A 표준은 의사(pseudo) 채널 모드에서 동작을 제공하고, 상기 의사 채널 모드에서, 8개의 128 비트 채널들 각각은 두 개의 반 독립적인(semi-independent) 의사 채널들으로써 동작한다. 이러한 모드에서는, 두 개의 의사 채널들이 개별적으로 명령들을 디코딩하고 실행함에도 불구하고, 의사 채널들 각각의 쌍은 채널의 행 및 열 명령 버스를 공유할 뿐만 아니라 CK 및 CKE 입력들도 공유한다. 일부 실시예들에서, 상기 모드는 태그 값들을 저장하는 데에 사용된다. 채널의 전체 128 비트 폭 데이터 버스(DQ 버스)가 DRAM 스택(115)으로부터 태그값을 읽거나 태그값을 쓰는데 사용되는 경우, 각각의 태그 값은 매우 효율적이지 못한(즉 25% 정도의 효율성) 결과를 가져오는 32 비트 워드일 수 있다. 의사 채널 모드에서 상기 버스의 폭의 단지 절반(즉 64비트)만이 태그 값을 읽거나 쓰는데 사용될 수 있고, 더 높은 효율성을 가져온다.(즉 50% 정도)

[0051] DRAM의 뱅크 각각은 16 서브 어레이들을 포함(또는 구성)할 수 있다. 도 4A를 참조하면, 일부 실시예들에서, 각각의 데이터 워드는 64 바이트(512 비트) 길이이고, 각각의 태그 값은 4바이트(32비트) 길이일 수 있다. 즉 길이들의 비율은 16: 1이다. 태그 값들 및 데이터는 서로 다른 채널들을 통해서 액세스될 수 있다. 예를 들어, 데이터는 채널 1 내지 15를 통해서 액세스될 수 있고, 태그 값들은 채널 0을 통해서 액세스될 수 있다. 상기 실시예에서, 병렬적으로 데이터를 액세스하는 것이 가능하지만, 태그 액세스들은 도시된 점선 타원에 보이는 것과 같이, 뱅크 충돌들(bank conflicts)을 겪게 된다. 따라서, 도 4B를 참조하면, 일 실시예에서, 태그들은 서로 다른 서브 어레이들에 저장되고, 상기 서브 어레이들은 본 명세서에서 SALP(subarray level parallelism)로 칭해지는 방법을 사용하여 병렬로 액세스될 수 있다. 상기 실시예에서, 태그 액세스들이 동시에 처리되는 것이 가능할 수 있고, 이를 통해서 태그 액세스들이 동일한 뱅크에 행해지는 경우에도, 태그 액세스들을 위한 충돌들을

피할 수 있다.

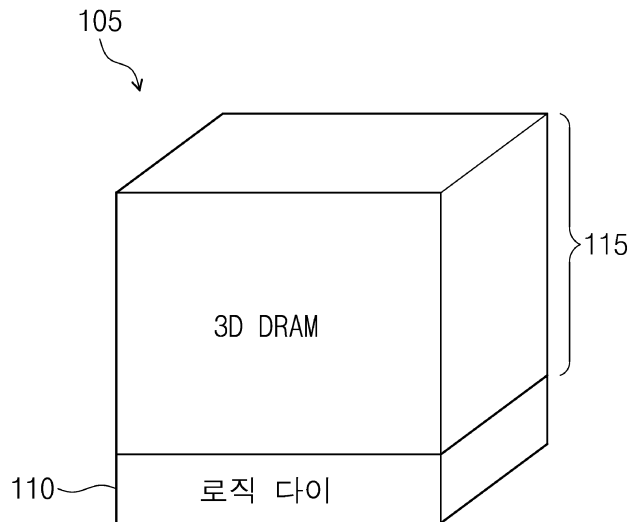
- [0052] 전술한 관점에서, 로직 다이(110)에서 캐쉬 관리자를 포함하는 HBM 스택(105)은 캐쉬 관리자가 없는 HBM 스택 관련 기술 대비 많은 장점을 가질 수 있다. 캐쉬 관리자를 포함하는 HBM 스택(105)의 사용은 호스트 프로세서의 크기, 가격, 및 전력 소비를 줄이면서, 호스트 프로세서(205)내에 캐쉬 관리자를 포함할 필요가 없도록 하고, 동일한 리소스를 호스트 프로세서(205)에서 다른 용도로 사용할 수 있게 한다. 게다가 조건부 실행은 이러한 조건부 실행이 호스트 프로세서(205)를 수반하는 경우 보다 HBM 스택(105) 내에서 전적으로 수행 될 경우 더 빠를 수 있다. 예를 들어, 쓰기 명령은 캐쉬 적중 결정이 호스트 프로세서(205)의 캐쉬 관리자 내에서 행해지는 경우 보다, HBM 스택(105) 내에서 행해지는 캐쉬 적중의 경우에 더 빨리 실행될 수 있다.
- [0053] “제1”, “제2”, “제3” 등과 같은 용어는 본 명세서에서, 다양한 요소들, 구성요소들, 영역들, 층들, 및/또는 부분들을 설명하기 위해 사용됨에데 불구하고, 상기 요소들, 구성요소들, 영역들, 층들, 및/또는 부분들은 상기 용어들에 한정되지 않음은 자명할 것이다. 상기 용어들은 하나의 요소, 구성요소, 영역, 층, 또는 부분을 또 다른 요소, 구성요소, 영역, 층, 또는 부분과 구분하기 위해서 단지 사용될 수 있다. 그러므로, 이하 논의될 제1 요소, 구성요소, 영역, 층, 또는 부분은 발명의 개념의 사상 및 범위를 벗어나지 않으면서, 제2 요소, 구성요소, 영역, 층, 또는 부분으로 지정될 수 있다.
- [0054] “아래” 및 “위” 등과 같은 공간적으로 관련된 용어는 하나의 요소 및 특징과 하나의 도면에서 설명되는 또 다른 요소(들) 또는 특징(들)과의 관계를 설명하는 것을 쉽게 하기 위해서 본 명세서에서 사용될 수 있다. 상기 공간적으로 관련된 용어들은 도면에서 설명되는 방향에 추가하여, 사용되는 장치 및 동작의 서로 다른 방향을 포함할 수 있다. 예를 들어, 도면에서 장치가 뒤집히는 경우에는 다른 요소들 또는 특징들의 “아래”와 같은 용어으로써 설명되는 요소는 다른 요소들 또는 특징들의 “위”로 배향될 수 있다. 그러므로, 예시적 용어 “아래”는 아래와 위를 모두 포함할 수 있다. 장치는 다른 방향으로 배향될 수 있고, (예를 들어, 90도 또는 다른 방향으로 회전될 수 있음), 본 명세서에서 사용되는 공간적으로 관련된 용어는 이에 따라 해석되어야 한다. 또한 하나의 층이 2개의 층 사이에 있는 것을 언급하는 경우, 2개의 층 사이에 유일한 층일 수 있거나, 하나 또는 그 이상의 개재된 층이 존재할 수 있음은 자명할 것이다.
- [0055] 본 명세서에서 사용되는 용어는 특정 실시예만을 설명하기 위한 것이며, 본 발명의 개념을 제한하고자하는 것은 아니다. 본 명세서에서 사용된 용어 “실질적으로”, “약” 및 이와 유사한 용어는 근사의 용어으로써 사용되며, 학술적인 용어로서 사용되지 않고, 이들에 의해서 인식되는 측정 또는 계산 값은 고유의 편차를 설명하기 위한 것이고 이는 당업자에게 자명할 것이다. 본 명세서에서 사용된 바와 같이 “주요 성분”이란 용어는 조성물의 적어도 절반을 구성하는 성분을 의미하고, “주요 부분”은 복수의 항목에 적용될 때, 항목의 적어도 절반을 의미한다.
- [0056] 본 명세서에서 사용된 단수 형태는 문맥에 달리 명시되어있지 않는 한 복수 형태를 포함하는 것으로 의도된다. 본 명세서에서 사용하는 “포함하는”이라는 용어는 명시되니 특징, 정수, 단계, 동작, 구성요소, 및/또는 구성요소의 존재를 나타내지만, 존재를 배제하지 않는다는 것으로 해석하는 것이 자명할 것이다. 또는 하나 이상의 다른 특징, 정수, 단계, 동작, 요소, 구성 요소, 및/또는 그룹의 추가를 포함할 수 있다. 본 명세서에서 사용된 바와 같이 “및/또는”이라는 용어는 하나 이상의 관련되어 열거된 항목의 임의 및 모든 조합을 포함한다. “적어도 하나”와 같은 표현은 요소 목록 앞에 있는 경우에 요소의 전체 목록을 수정하고, 목록의 개별 요소를 수정하지 않는다. 또한 본 발명의 개념의 실시예를 기술할 때 “할 수 있다”를 사용하는 것은 “본 발명의 하나 이상의 실시예”를 의미한다. 또한 “예시적인”이라는 용어는 예 또는 설명을 의미한다. 본 명세서에서 사용된 바와 같이 용어 “사용”, “사용하는” 및 “사용된”은 각각 “이용”, “이용하는” 및 “이용된”과 동의어로 간주될 수 있다.
- [0057] 요소 또는 층이 또다른 요소 또는 층의 “위”에 있는 것, “연결된”, “결합된” 또는 인접한 것으로 언급되는 경우, 하나 이상의 개재 요소 또는 층이 존재할 수 있다. 대조적으로 하나의 요소 또는 층이 다른 요소 또는 층에 “직접적으로”, “직접적으로 연결된”, “직접적으로 결합된” 또는 “바로 인접한”으로 언급될 때, 개재하는 요소 또는 층이 존재하지 않는다.
- [0058] 본 명세서에서 열거된 임의의 수치 범위는 열거된 범위 내에 포함된 동일한 수치 정밀도의 모든 하위 범위를 포함하도록 의도된다. 예를 들어, 1.0~10.0의 범위에는 지정된 1.0의 값과 10.0의 값의 사이의 모든 부분 범위가 포함된다. (최소의 값이 1.0 이상). 예를 들어, 2.4 내지 7.6과 같은 값이 있을 수 있다. 본 명세서에서 인용된 최대 수치 제한은 여기에 포함되는 모든 낮은 수치 제한을 포함하도록 의도되며, 본 명세서에 인용된 임의의 최소 수치 제한은 여기에 포함되는 모든 더 높은 수치 제한을 포함하도록 의도된다.

[0059] 메모리 내 캐쉬 관리자가 갖는 고 대역폭 메모리의 예시적인 실시예가 본 명세서에서 구체적으로 설명되고 예시되었지만, 많은 변형 및 변경이 당업자에게 자명할 것이다. 따라서 본 발명의 원리에 따라 구성된 메모리 내 캐쉬 관리자를 갖는 고 대역폭 메모리는 본 명세서에서 구체적으로 설명된 것 이외에서 구현될 수 있다. 본 발명은 또한 다음의 특허 청구 범위 및 그 균등물에서 정의된다.

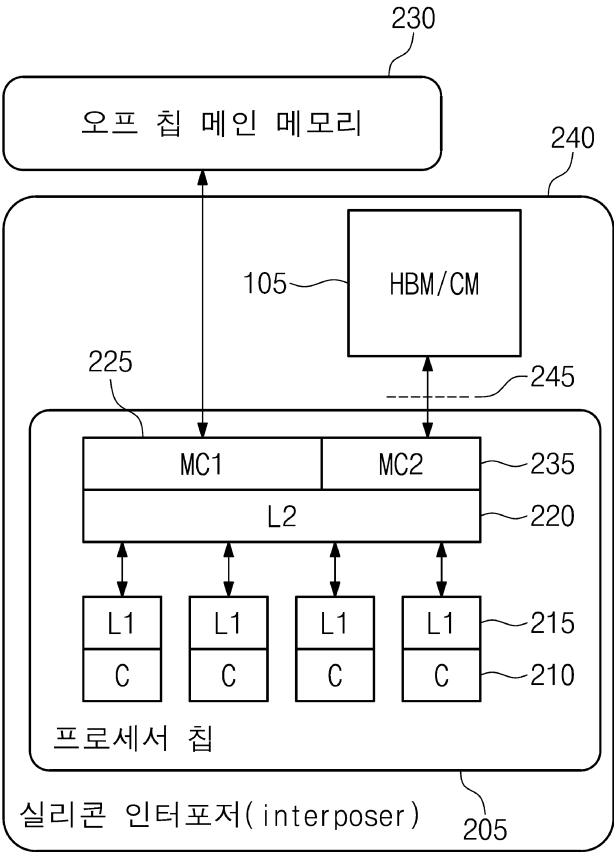
# 부호의 설명

## 도면

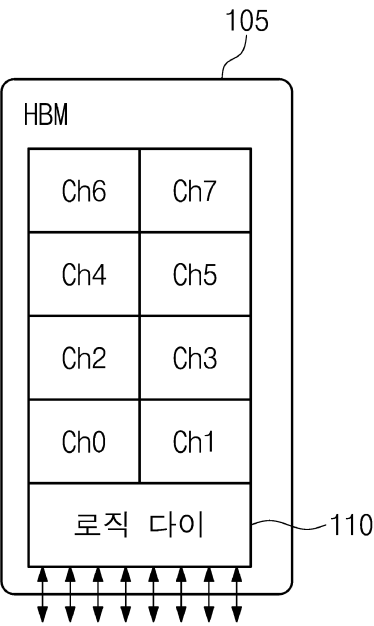
### 도면1



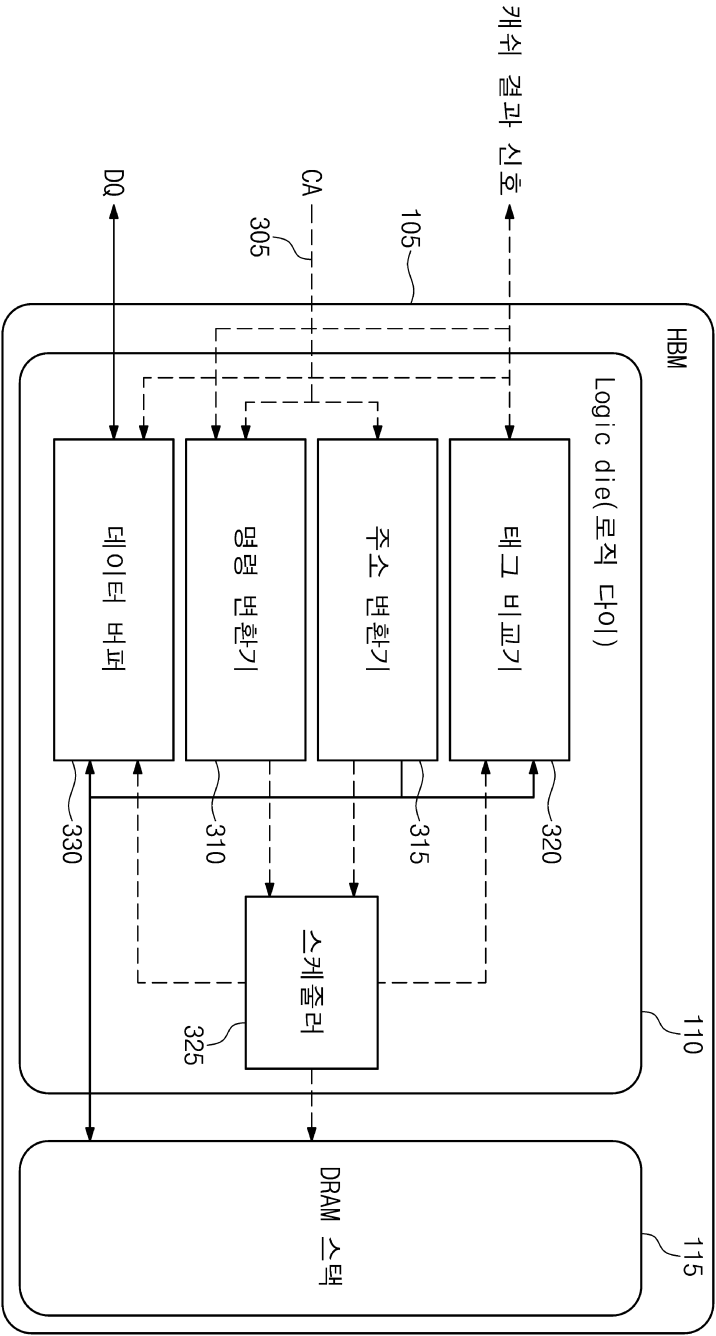
도면2a



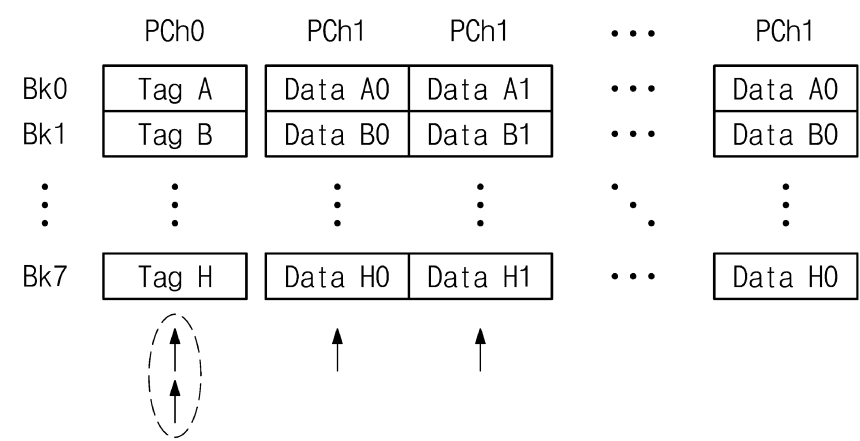
도면2b



도면3



도면4a



도면4b

