(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2015-0100042
(43) 공개일자 2015년09월02일

(51) 국제특허분류(Int. Cl.)
G06F 12/00 (2006.01) *G06F 13/14* (2006.01)
G11C 5/02 (2006.01)
(21) 출원번호 10-2014-0021342
(22) 출원일자 2014년02월24일
심사청구일자 없음

(71) 출원인
한국전자통신연구원
대전광역시 유성구 가정로 218 (가정동)
국립대학법인 울산과학기술대학교 산학협력단
울산광역시 울주군 언양읍 유니스트길 50
(72) 발명자
김용주
대전광역시 유성구 가정로 270 2동 139호
이진용
서울특별시 관악구 관악로 1 서울대학교 신공학관 302동 516-1호
(뒷면에 계속)
(74) 대리인
특허법인 무한

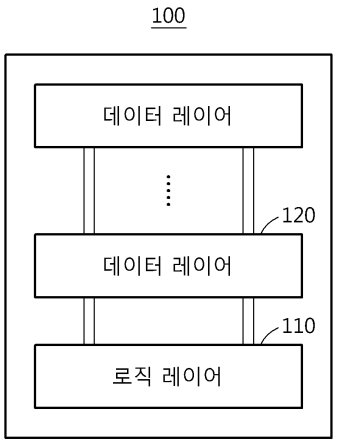
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 **3차원 다이 스택 디램에서의 가속 시스템**

(57) 요 약

메모리 장치는, 페리퍼럴(peripheral), 인터페이스(interface) 및 비스트 모듈(BIST module) 중 적어도 하나를 구현하고 재구성 가속기(reconfigurable accelerator: RA)를 포함하는 로직 레이어(logic layer) - 상기 재구성 가속기는 상기 로직 레이어의 빈공간에 위치하며 상기 메모리 장치가 처리하는 작업의 적어도 일부를 처리함 -및 데이터를 저장하는 적어도 하나의 데이터 레이어(data layer)를 포함할 수 있다.

대 표 도 - 도1



(72) 발명자

백윤흥

서울특별시 관악구 관악로 1 서울대학교 신공학관 302동 530호

이종은

울산광역시 울주군 언양읍 유니스트길 50 2 공학관(106) 501-5호

이 발명을 지원한 국가연구개발사업

과제고유번호 10044313

부처명 지식경제부

연구관리전문기관 한국산업기술평가관리원

연구사업명 산업원천기술개발사업

연구과제명 이종 멀티코어 클러스터 기반 스마트 디바이스용 하이퍼 커넥션 서비스 지원 SW-SoC 융합 플랫폼 핵심 기술

기 여 율 1/1

주관기관 한국전자통신연구원

연구기간 2013.03.01 ~ 2017.02.28

특허청구의 범위

청구항 1

페리퍼럴(peripheral), 인터페이스(interface) 및 비스트 모듈(Build-In-Self-Test module: BIST module)중 적어도 하나를 구현하고 재구성 가속기(reconfigurable accelerator: RA)를 포함하는 로직 레이어(logic layer) - 상기 재구성 가속기는 상기 로직 레이어의 빈공간에 위치하며 상기 메모리 장치가 처리하는 작업의 적어도 일부를 처리함 -; 및

데이터를 저장하는 적어도 하나의 데이터 레이어(data layer)

를 포함하는 메모리 장치.

청구항 2

제1항에 있어서,

상기 재구성 가속기는,

프로세싱 엘리먼트(processing elements:PE)를 포함하고, 상기 프로세싱 엘리먼트는 어레이(array)구조로 연결된 메모리 장치.

청구항 3

제2항에 있어서,

제1 프로세싱 엘리먼트는 상기 제1 프로세싱 엘리먼트에 인접한 제2 프로세싱 엘리먼트와 연결되고, 상기 데이터를 상기 제2 프로세싱 엘리먼트와 송수신하는 메모리 장치.

청구항 4

제2항에 있어서,

상기 프로세싱 엘리먼트는, 상기 데이터를 연산하는 기능부(functional unit: fu)를 포함하는 메모리 장치.

청구항 5

제1항에 있어서,

상기 로직 레이어는, 로컬 메모리(local memory)를 포함하는 메모리 장치.

청구항 6

제5항에 있어서,

상기 로컬 메모리는,

복수 개의 캐시(cache)로 구성된 제1 로컬 캐시;및

상기 데이터 레이어 및 상기 제1 로컬 캐시 사이를 연결하는 제2 로컬 캐시를 포함하는 메모리 장치.

청구항 7

제6항에 있어서,

상기 재구성 가속기가 동작하지 않는 경우, 상기 제2 로컬 캐시는 상기 데이터 레이어의 로우 버퍼(low buffer)로 사용되는 메모리 장치.

청구항 8

제1항에 있어서,

상기 로직 레이어 및 상기 적어도 하나의 데이터 레이어의 크기는 동일한 메모리 장치.

청구항 9

제1항에 있어서,

상기 로직 레이어에서 페리퍼럴(peripheral), 인터페이스(interface) 및 비스트 모듈(BIST module) 중 적어도 하나를 구현하는 제1 공간은, 상기 로직 레이어의 중심부에 위치하는 메모리 장치.

청구항 10

제1항에 있어서,

상기 로직 레이어는 상기 적어도 하나의 데이터 레이어의 하부에 배치되는 메모리 장치.

청구항 11

데이터를 저장하는 적어도 하나의 데이터 레이어; 및

상기 적어도 하나의 데이터 레이어의 하부에 배치되는 로직 레이어

를 포함하고,

상기 로직 레이어는, 페리퍼럴(peripheral), 인터페이스(interface) 및 비스트 모듈(BIST module) 중 적어도 하나를 구현하는 제1 영역 및 재구성 가속기를 포함하는 제2 영역을 포함하는 메모리 장치.

청구항 12

제11항에 있어서,

상기 제1 영역은 상기 로직 레이어의 중심부에 위치하고, 상기 제2 영역은 상기 제1 영역을 제외한 나머지 로직 레이어 영역인 메모리 장치.

청구항 13

제11항에 있어서,

상기 로직 레이어 및 상기 적어도 하나의 데이터 레이어는 쓰로우-실리콘-비아(Through-Silicon-Via:TSV)를 이용하여 적층된 메모리 장치.

청구항 14

제11항에 있어서,

상기 로직 레이어는, 로컬 메모리(local memory)를 포함하는 메모리 장치.

청구항 15

제14항에 있어서,

상기 로컬 메모리는,

복수 개의 캐시(cache)로 구성된 제1 로컬 캐시;및

상기 데이터 레이어 및 상기 제1 로컬 캐시 사이를 연결하는 제2 로컬 캐시를 포함하는 메모리 장치.

청구항 16

페리퍼럴(peripheral), 인터페이스(interface) 및 비스트 모듈(BIST module)을 구현하고 재구성 가속기(reconfigurable accelerator: RA)를 포함하는 로직 레이어(logic layer)를 형성하는 단계 - 상기 재구성 가속기는 상기 로직 레이어의 빈공간에 위치하며 상기 메모리 장치가 처리하는 작업의 적어도 일부를 처리함 -; 및

데이터를 저장하는 적어도 하나의 데이터 레이어(data layer)를 형성하는 단계

를 포함하는 메모리 장치의 제조 방법.

청구항 17

제16항에 있어서,

상기 로직 레이어 및 상기 적어도 하나의 데이터 레이어는 쓰로우-실리콘-비아(Through-Silicon-Via:TSV)를 이용하여 적층된 메모리 장치의 제조 방법.

청구항 18

제16항에 있어서,

상기 재구성 가속기에, 프로세싱 엘리먼트(processing elements:PE)를 형성하고, 상기 프로세싱 엘리먼트를 어레이(array)구조로 형성하는 단계를 더 포함하는 메모리 장치의 제조 방법.

청구항 19

제16항에 있어서,

상기 로직 레이어에, 로컬 메모리(local memory)를 형성하는 단계를 더 포함하는 메모리 장치의 제조 방법.

청구항 20

제19항에 있어서,

상기 로컬 메모리에,

복수 개의 캐시(cache)로 구성된 제1 로컬 캐시를 형성하는 단계;및

상기 데이터 레이어 및 상기 제1 로컬 캐시 사이를 연결하는 제2 로컬 캐시를 형성하는 단계를 더 포함하는 메모리 장치의 제조 방법.

명 세 서

기 술 분 야

[0001] SOC 분야나 임베디드 시스템 분야에서 메모리 아키텍쳐 시스템을 구성하는 방법에 연관되며, 구체적으로는 3차원 다이 스택 디램에서의 가속 시스템 및 방법에 연관된다.

배 경 기 술

[0002] 다이 스태킹(Die-stacking) 기술은 여러 개의 다이(die)를 한 공간에 여러층으로 쌓아 작은 공간에 많은 용량을 집적할 수 있고, 각 다이 사이에 빠른 상호 접속(interconnection)을 할 수 있도록 해준다.

[0003] 3차원 스태킹 방법 중, 이중 다이 간의 연결을 쓰로우-실리콘-비아(Through-Silicon-Via:TSV)를 이용하는 방법이 빠르면서도 높은 집약도를 가질 수 있어 많은 연구가 되고 있다.

[0004] 상기 쓰로우-실리콘-비아 방식을 사용한 스태킹을 하는 경우, 로직 레이어(logic layer)에 빈공간이 생길 수 있다.

[0005] 따라서, 이러한 공간은 이전까지는 낭비되는 공간으로 두었지만, 최근에는 이 공간을 활용하고자 하는 연구들이 시도되고 있다.

발명의 내용

과제의 해결 수단

[0006] 일측에 따르면, 페리퍼럴(peripheral), 인터페이스(interface) 및 비스트 모듈(Build-In-Self-Test module: BIST module) 중 적어도 하나를 구현하고 재구성 가속기(reconfigurable accelerator: RA)를 포함하는 로직 레이어(logic layer) - 상기 재구성 가속기는 상기 로직 레이어의 빈공간에 위치하며 상기 메모리 장치가 처리하는 작업의 적어도 일부를 처리함 -; 및 데이터를 저장하는 적어도 하나의 데이터 레이어(data layer)를 포함하는 메모리 장치가 제공된다.

[0007] 일실시예에 따르면, 상기 재구성 가속기는, 프로세싱 엘리먼트(processing elements:PE)를 포함하고, 상기 프로세싱 엘리먼트는 어레이(array)구조로 연결될 수 있다.

[0008] 다른 일실시예에 따르면, 제1 프로세싱 엘리먼트는 상기 제1 프로세싱 엘리먼트에 인접한 제2 프로세싱 엘리먼트와 연결되고, 상기 데이터를 상기 제2 프로세싱 엘리먼트와 송수신할 수 있다.

[0009] 한편, 상기 프로세싱 엘리먼트는, 상기 데이터를 연산하는 기능부(functional unit: fu)를 포함할 수 있다.

[0010] 또 다른 일실시예에 따르면, 상기 로직 레이어는, 로컬 메모리(local memory)를 포함할 수 있다.

[0011] 상기 로컬 메모리는, 복수 개의 캐시(cache)로 구성된 제1 로컬 캐시;및 상기 데이터 레이어 및 상기 제1 로컬 캐시 사이를 연결하는 제2 로컬 캐시를 포함할 수 있다.

[0012] 일실시예에 따르면, 상기 재구성 가속기가 동작하지 않는 경우, 제2 로컬 캐시는 상기 데이터 레이어의 로우 버퍼(low buffer)로 사용될 수 있다.

[0013] 다른 일실시예에 따르면, 상기 로직 레이어 및 상기 적어도 하나의 데이터 레이어의 크기는 동일할 수 있다.

[0014] 또 다른 일실시예에 따르면, 상기 로직 레이어에서 페리퍼럴(peripheral), 인터페이스(interface) 및 비스트 모듈(BIST module) 중 적어도 하나를 구현하는 제1 공간은, 상기 로직 레이어의 중심부에 위치할 수 있다.

[0015] 한편, 상기 로직 레이어는 상기 적어도 하나의 데이터 레이어의 하부에 배치될 수 있다.

[0016] 다른 일측에 따르면, 데이터를 저장하는 적어도 하나의 데이터 레이어; 및 상기 적어도 하나의 데이터 레이어의 하부에 배치되는 로직 레이어를 포함하고, 상기 로직 레이어는, 페리퍼럴(peripheral), 인터페이스(interface) 및 비스트 모듈(BIST module) 중 적어도 하나를 구현하는 제1 영역 및 재구성 가속기를 포함하는 제2 영역을 포함하는 메모리 장치가 제공된다.

[0017] 일실시예에 따르면, 상기 제1 영역은 상기 로직 레이어의 중심부에 위치하고, 상기 제2 영역은 상기 제1 영역을 제외한 나머지 로직 레이어 영역일 수 있다.

[0018] 다른 일실시예에 따르면, 상기 로직 레이어 및 상기 적어도 하나의 데이터 레이어는 쓰로우-실리콘-비아(Through-Silicon-Via:TSV)를 이용하여 적층될 수 있다.

[0019] 또 다른 일실시예에 따르면, 상기 로직 레이어는, 로컬 메모리(local memory)를 포함할 수 있다.

[0020] 상기 로컬 메모리는, 복수 개의 캐시(cache)로 구성된 제1 로컬 캐시;및 상기 데이터 레이어 및 상기 제1 로컬 캐시 사이를 연결하는 제2 로컬 캐시를 포함할 수 있다.

[0021] 또 다른 일측에 따르면, 페리퍼럴(peripheral), 인터페이스(interface) 및 비스트 모듈(BIST module)을 구현하고 재구성 가속기(reconfigurable accelerator: RA)를 포함하는 로직 레이어(logic layer)를 형성하는 단계 - 상기 재구성 가속기는 상기 로직 레이어의 빈공간에 위치하며 상기 메모리 장치가 처리하는 작업의 적어도 일부를 처리함 -; 및 데이터를 저장하는 적어도 하나의 데이터 레이어(data layer)를 형성하는 단계를 포함하는 메모리 장치의 제조 방법이 제공된다.

[0022] 일실시예에 따르면, 상기 로직 레이어 및 상기 적어도 하나의 데이터 레이어는 쓰로우-실리콘-비아(Through-Silicon-Via:TSV)를 이용하여 적층될 수 있다.

[0023] 다른 일실시예에 따르면, 상기 재구성 가속기에, 프로세싱 엘리먼트(processing elements:PE)를 형성하고, 상기 프로세싱 엘리먼트를 어레이(array)구조로 형성하는 단계를 더 포함할 수 있다.

[0024] 또 다른 일실시예에 따르면, 상기 로직 레이어에, 로컬 메모리(local memory)를 형성하는 단계를 더 포함할 수 있다.

[0025] 상기 로컬 메모리에, 복수 개의 캐시(cache)로 구성된 제1 로컬 캐시를 형성하는 단계;및 상기 데이터 레이어 및 상기 제1 로컬 캐시 사이를 연결하는 제2 로컬 캐시를 형성하는 단계를 더 포함할 수 있다.

도면의 간단한 설명

[0026] 도 1은 일실시예에 따른 메모리 장치의 구성을 나타낸 블록도이다.

도 2는 일실시예에 따른 로직 레이어의 구성을 나타낸 블록도이다.

도 3은 일실시예에 따라, 재구성 가속기의 구성을 나타낸 블록도 및 프로세싱 엘리먼트의 구성을 나타낸 블록도이다.

도 4는 다른 일실시예에 따른 메모리 장치의 구성을 나타낸 세부 블록도이다.

도 5는 또 다른 일실시예에 따른 메모리 장치의 구성을 나타낸 세부 블록도이다.

도 6은 일실시예에 따른 메모리 장치의 제조 방법의 순서를 나타낸 흐름도이다.

도 7은 일실시예에 따라, 로직 레이어의 형성 단계를 상세히 나타낸 흐름도이다.

도 8은 일실시예에 따라, 제2 영역 형성 단계를 상세히 나타낸 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0027] 이하에서, 일부 실시예들을, 첨부된 도면을 참조하여 상세하게 설명한다. 그러나, 이러한 실시예들에 의해 제한되거나 한정되는 것은 아니다. 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.

[0028] 아래 설명에서 사용되는 용어는 본 발명에서의 기능을 고려하면서 가능한 현재 널리 사용되는 일반적인 용어를 선택하였으나, 이는 당 분야에 종사하는 기술자의 의도 또는 관례, 새로운 기술의 출현 등에 따라 달라질 수 있다.

[0029] 또한 특정한 경우는 이해를 돕거나 및/또는 설명의 편의를 위해 출원인이 임의로 선정한 용어도 있으며, 이 경우 해당되는 설명 부분에서 상세한 그 의미를 기재할 것이다. 따라서 아래 설명에서 사용되는 용어는 단순한 용어의 명칭이 아닌 그 용어가 가지는 의미와 명세서 전반에 걸친 내용을 토대로 이해되어야 한다.

[0030] 도 1은 일실시예에 따른 메모리 장치의 구성을 나타낸 블록도이다.

[0031] 메모리 장치(100)는 데이터를 일시적으로 또는 영구히 보존하는 장치를 의미할 수 있다. 메모리 장치(100)는 저장 장치라고 불릴 수도 있다. 메모리는 종종 "기억장치"라는 용어와 혼용되기도 하지만, 대체로는 주기억 장치를 의미할 수도 있다. 램(RAM)을 가리킬 수도 있다.

[0032] 메모리 장치(100)는 휘발성 여부에 따라 비휘발성 메모리와 휘발성 메모리로 분류될 수 있다. 비휘발성 메모리(Non-volatile memory, NVM, NVRAM)는 전원이 공급되지 않아도 저장된 정보를 계속 유지하는 메모리이다. 반면에, 휘발성 메모리(volatile memory)는 지속적인 전력 공급을 요구하지 않는 비휘발성 메모리와는 달리 저장된 정보를 유지하기 위해 전기를 요구하는 메모리이다. 임시 메모리라고도 한다. 동적 램(DRAM), 정적 램(SRAM)을 포함한 랜덤 액세스 메모리(Random Access Memory: RAM)가 대표적인 휘발성 메모리이다.

[0033] 일실시예에 따르면, 메모리 장치(100)는 로직 레이어(110) 및 데이터 레이어(120)를 포함할 수 있다. 상기 데이터 레이어(120)에는 디램이 포함될 수도 있다.

[0034] 로직 레이어(110)는 페리페럴(peripheral), 인터페이스(interface), 및 비스트 모듈(BIST module) 중 적어도 하나를 구현하고, 재구성 가속기(reconfigurable accelerator: RA)를 포함할 수 있다. 여기에서, 페리페럴(peripheral), 인터페이스(interface), 및 비스트 모듈(Build-In-Self-Test module: BIST module) 중 적어도 하나를 구현하는 영역과 재구성 가속기가 포함되는 영역은 상이할 수 있다. 예를 들어, 페리페럴(peripheral), 인터페이스(interface), 및 비스트 모듈(BIST module) 중 적어도 하나는 로직 레이어(110)의 제 1 영역에 구현될 수 있으며, 재구성 가속기는 로직 레이어(110)의 제 2 영역에 구현될 수 있다. 일 실시 예에서, 제 1 영역은 로직 레이어(110)의 중심부에 위치할 수 있으며, 제 2 영역은 제 1 영역을 제외한 나머지 영역일 수 있다. 이에 대하여서는 도 2를 참조하여 더욱 상세하게 설명하도록 한다.

[0035] 메모리 장치(100)는 데이터를 저장하는 적어도 하나의 데이터 레이어(120)를 포함할 수 있다. 상기 데이터 레이어(120)는 디램 레이어(DRAM layer)일 수도 있다.

[0036] 예를 들면, 로직 레이어(110)에 페리페럴(peripheral), 인터페이스(interface), 및 비스트 모듈(BIST module) 중 적어도 하나를 구현한 후에는 빈 공간(unused space)이 있을 수 있다. 또한, 3D 스택 디램은 다른 여러 레이어들이 층층히 쌓여 있는 구조이기 때문에, 2D DRAM에 비해서 열을 배출하기 어려운 구조일 수도 있다.

[0037] 일실시예에 따르면, 상기 재구성 가속기는 프로세싱 엘리먼트(processing elements: PE)를 포함할 수 있고, 프로세싱 엘리먼트 어레이(array) 구조로 연결될 수도 있다. 또한, 재구성 가속기는 단순한 계산을 가속하기 위한 구조를 가진 가속기의 한 종류일 수 있다. 재구성 가속기는 구조가 단순하여 적은 공간을 차지하고 전력을 적게 소모할 수 있다.

[0038] 재구성 가속기는 적어도 하나의 프로세싱 엘리먼트를 포함할 수 있다. 제1 프로세싱 엘리먼트는 상기 제1 프로

세싱 엘리먼트에 인접한 제2 프로세싱 엘리먼트에 연결될 수 있다. 제1 프로세싱 엘리먼트 및 제2 프로세싱 엘리먼트는 서로 데이터를 송수신할 수 있다. 복수 개의 프로세싱 엘리먼트는 상대적으로 높은 퍼포먼스(performance)를 보일 수 있고, 데이터 집약적인 커널(data intensive kernel)을 가속하기에 적합할 수 있다.

[0039] 또한, 프로세싱 엘리먼트는 데이터를 연산하는 기능부를 포함할 수 있다. 다시 말하면, 각 프로세싱 엘리먼트는 ALU(arithmetic and logic unit)와 같은 기능부(functional unit: FU)를 포함할 수 있다. 제1 프로세싱 엘리먼트는 인접한 프로세싱 엘리먼트로부터 데이터를 전달받거나, 자신으로부터 전달받아 계산을 수행할 수 있다.

[0040] 재구성 가속기를 빈 공간(unused space)에 배치하면, 작은 영역(area)에서도 빈 공간을 활용할 수 있다. 또한, 재구성 가속기는 발열이 작고, 높은 데이터 밴드폭(data bandwidth)으로 풍부한 데이터를 지연(delay)없이 접근 가능할 수 있다.

[0041] 일실시예에 따르면, 로직 레이어(110)는 디램의 데이터 셀 바로 옆에 붙어 있어 내부적으로 데이터에 빠르게 접근할 수 있다. 로직 레이어(110)에서 데이터를 처리할 경우, 내부적인 연결을 통해 높은 데이터 밴드폭(data bandwidth)을 확보할 수 있고, 이를 이용해 빠르게 데이터에 접근할 수 있다.

[0042] 일실시예에 따르면, 로직 레이어(110)는 로컬 메모리(local memory)를 포함할 수 있다. 상기 로컬 메모리는, 제1 로컬 캐시 및 제2 로컬 캐시를 포함할 수 있다.

[0043] 상기 제1 로컬 캐시는 복수 개의 캐시로 구성될 수 있다. 또한, 상기 제2 로컬 캐시는 데이터 레이어(120) 및 제1 로컬 캐시 사이를 연결할 수 있다.

[0044] 일실시예에 따르면, 제1 로컬 캐시 및 제2 로컬 캐시를 포함하는 로컬 메모리는 재구성 가속기 및 데이터 레이어(120)의 수행을 효율적으로 보조할 수 있다.

[0045] 다른 일실시예에 따르면, 재구성 가속기가 동작하지 않는 경우, 제2 로컬 캐시는 데이터 레이어(120)의 로우 버퍼(low buffer)로 사용될 수 있다.

[0046] 한편, 로직 레이어(110) 및 적어도 하나의 데이터 레이어(120)의 크기는 동일할 수 있다. 쓰로우-실리콘-비아(Through-Silicon-Via:TSV)를 사용한 스태킹을 할 때에는 공정의 안정성을 위해서 데이터 레이어(120) 및 로직 레이어(110)의 크기를 동일하게 하여 쌓을 수 있다.

[0047] 일실시예에 따르면, 쓰로우-실리콘-비아의 방식은 여러 개의 디램 레이어(DRAM layer)를 합친 후, 디램 페리퍼럴 회로(DRAM peripheral circuitry), 인터페이스 회로(interface circuitry), 빌드-인-셀프-테스트(build-in-self-test: BIST) 모듈(module) 및 다른 컴포넌트(component)를 구현하기 위하여 하나의 CMOS 레이어와 합칠 수 있다.

[0048] 다른 일실시예에 따르면, 로직 레이어(110) 및 데이터 레이어(120)를 다른 공정에서 제작할 수 있다. 다른 공정에서 제작된 로직 레이어(110) 및 데이터 레이어(120)를 쓰로우-실리콘-비아의 방식으로 연결할 수 있다. 연결된 로직 레이어(110) 및 데이터 레이어(120)는 메모리 장치(100)를 구성할 수 있다. 메모리 장치(100)는 빠른 반응 속도를 보일 수 있고, 최적화된 공정은 많은 데이터를 집적할 수 있다.

[0049] 로직 레이어(110)에서 실제 필요한 회로(circuitry)를 구현하는 데에는 많은 영역이 필요하진 않지만, 데이터 레이어(120)와 로직 레이어(110)의 크기를 동일하게 맞추면 산출량을 증가시키는데 도움이 될 수 있다.

[0050] 따라서, 로직 레이어(110) 부분을 데이터 레이어(120)의 크기만큼 만들 경우, 사용되지 않은 공간이 로직 레이어(110)에 생길 수 있다.

[0051] 일실시예에 따르면, 로직 레이어(110)는 적어도 하나의 데이터 레이어(120)의 하부에 배치될 수 있다.

[0052] 도 2는 일실시예에 따른 로직 레이어의 구성을 나타낸 블록도이다.

[0053] 일실시예에 따르면, 로직 레이어(110)는, 제1 영역(210) 및 제2 영역(220)으로 구성될 수 있다. 제1 영역(210)은, 로직 레이어(110)에서 페리퍼럴(peripheral), 인터페이스(interface) 및 비스트 모듈(BIST module) 중 적어도 하나를 구현할 수 있다. 또한, 제1 영역(210)은 로직 레이어(110)의 중심부에 위치할 수도 있다. 제2 영역(220)은, 재구성 가속기를 포함할 수 있다. 제2 영역(220)은, 제1 영역(210)을 제외한 나머지 로직 레이어(110)의 영역일 수 있다.

[0054] 로직 레이어(110)는 로컬 메모리를 포함할 수 있다. 로컬 메모리는 복수 개의 캐시로 구성된 제1 로컬 캐시를

포함할 수 있다. 또한, 데이터 레이어 및 제1 로컬 캐시 사이를 연결하는 제2 로컬 캐시를 포함할 수 있다.

[0055] 로직 레이어(110)의 빈공간인 제2 영역(220)에 재구성 가속기를 포함시켜 추가적인 비용없이 작업의 성능을 향상시킬 수 있다. 또한, 저전력 재구성형 가속기(reconfigurable accelerator)는 디램의 환경에 적합할 수 있다. 디램의 레이어와의 밀접성으로 가속기의 주요 성능 제약 문제인 메모리 병목 현상(memory bottle-neck)을 해결할 수 있다.

[0056] 도 3은 일실시예에 따라, 재구성 가속기의 구성을 나타낸 블록도 및 프로세싱 엘리먼트의 구성을 나타낸 블록도이다.

[0057] 일실시예에 따르면, 도 310은, 재구성 가속기의 구조를 나타내고 있다. 재구성 가속기에는 작업을 수행할 수 있는 적어도 하나 이상의 프로세싱 엘리먼트(processing elements:PE)를 포함할 수 있다. 상기 작업은 데이터를 저장하거나, 연산하는 것일 수 있다. 각 프로세싱 엘리먼트는 인접한 다른 프로세싱 엘리먼트와 연결되어 있어 데이터를 주고 받으면서 작업을 수행할 수 있다. 예를 들면, 제1 프로세싱 엘리먼트(311)는 제2 프로세싱 엘리먼트(312)와 연결되고, 데이터를 상호 송수신할 수 있다.

[0058] 한편, 프로세싱 엘리먼트에서 수행할 작업들을 컨피겨레이션 (configuration)이라고 할 수 있다. 컨피겨레이션은 컨피겨레이션 메모리(configuration memory)에 담겨 있다가 각 프로세싱 엘리먼트에 전달될 수 있다. 또한, 컨피겨레이션을 바꾸면 다른 작업을 수행할 수 있기 때문에 리컨피겨러블(reconfigurable)하다고 할 수 있다.

[0059] 일실시예에 따르면, 도 320은, 프로세싱 엘리먼트의 구조를 나타내고 있다. 각 프로세싱 엘리먼트는 ALU(arithmetic and logic unit)와 같은 기능부(functional unit: FU)를 포함할 수 있다. 제1 프로세싱 엘리먼트의 포함된 기능부는 제2 프로세싱 엘리먼트로부터 데이터를 수신하여 연산을 수행할 수 있다. 기능부가 연산을 수행하면, 출력단으로 연산 결과를 전달할 수 있다. 출력단은 연산 결과를 다른 프로세싱 엘리먼트로 전달하거나 제1 프로세싱 엘리먼트로 다시 전달할 수 있다. 프로세싱 엘리먼트는 데이터를 주변 프로세싱 엘리먼트나 자기 자신으로부터 전달받아 계산을 수행할 수 있다. 또한, 프로세싱 엘리먼트는 브랜치부(branch unit) 등 일반적인 중앙처리장치(cpu)가 가지는 페리퍼럴(peripheral)이 생략되어 있어 사이즈가 작고 계산량에 비해서 전력 소모가 작을 수 있다.

[0060] 일실시예에 따르면, 레지스터 파일(RF: Register File)는 계산 도중 생기는 중간 값들을 저장하는 공간일 수 있다. 프로세싱 엘리먼트는 컨피겨레이션이 지정하는 내용을 수행할 수 있다. 상기 수행에 사용되는 데이터는 주변의 다른 프로세싱 엘리먼트에서 가져오거나 자신의 레지스터 파일에서 데이터를 가져와서 수행할 수 있다. 결과값은 아웃풋 레지스터에 전달하여 다른 프로세싱 엘리먼트로 전달하거나 레지스터 파일에 저장하여 나중에 사용할 수 있도록 할 수 있다.

[0061] 일실시예에 따르면, 디램의 가속기가 데이터 집약적인 커널(data intensive kernel)을 수행하는 경우에 디램의 가속기는 중앙처리장치(cpu)에서 작업을 오프로딩(offloading)하여 사용할 수 있다. 데이터 집약적인 커널을 디램의 가속기에서 수행하면 데이터를 버스(bus)를 통해서 옮기지 않아도 되어 오프-칩 데이터 트랜스퍼 오버헤드(off-chip data transfer overhead)를 없앨 수 있다. 따라서, 버스에서 생기는 파워 소모를 줄일 수 있고 성능을 향상시킬 수 있다. 또한, 일반적인 작업을 수행하는 경우에는 디램의 가속기는 사용하지 않고 중앙처리장치에서 작업을 수행할 수 있다.

[0062] 도 4는 다른 일실시예에 따른 메모리 장치의 구성을 나타낸 세부 블록도이다.

[0063] 메모리 장치(100)는 로직 레이어 및 데이터 레이어를 포함할 수 있다. 데이터 레이어는 디램 레이어(440)일 수 있다.

[0064] 일실시예에 따르면, 메모리 장치(100)의 재구성 가속기(410)는 로컬 메모리 용도로 캐시를 사용할 수 있다. 로컬 메모리는 2단계로 구성될 수 있다. 로컬 메모리는 제1 로컬 캐시(420) 및 제2 로컬 캐시(430)를 포함할 수 있다. 제1 로컬 캐시(420)는 여러 개의 작은 캐시로 구성될 수 있다. 재구성 가속기(410)에서 여러 개의 데이터를 동시에 메모리로 요청하는데, 동시에 발생하는 데이터 요청을 처리할 수 있도록 하기 위해 여러 개의 작은 캐시로 제1 로컬 캐시(420)를 구성할 수 있다.

[0065] 일실시예에 따르면, 로직 레이어는 크로스바(Crossbar)를 포함할 수 있다. 크로스바는 여러 개의 작은 캐시로 구성된 제1 로컬 캐시(420)의 다수의 입력값 및 다수의 출력값들이 충돌을 일으키지 않도록 분배하는 역할을 할 수 있다.

[0066] 또한, 디램 억세스는 억세스 타임(access time)이 길고 파워소모가 크기 때문에, 제1 로컬 캐시(420)와 디램 사

이에 제2 로컬 캐시(430)를 넣어서 파워 소모와 억세스 타임을 줄일 수 있다. 제2 로컬 캐시(430)는 재구성 가속기(410)가 동작하는 동안에는 일반적인 재구성 가속기(410)용 로컬 메모리로 사용될 수 있다.

[0067] 재구성 가속기(410)가 동작하지 않는 경우에는 제2 로컬 캐시(430)를 다른 칩에 있는 중앙처리장치들을 위해서 디램의 한 행(row)의 데이터를 임시로 저장하는 로우 버퍼 캐시 (row buffer cache)로 활용할 수 있다. 제2 로컬 캐시(430)의 블록 사이즈는 로우 버퍼의 일부나 전체를 담을 수 있도록 만들 수 있다.

[0068] 로컬 메모리를 캐시로 구성하였을 경우, SPM의 데이터 관리를 위해 필요하던 오프-칩 커뮤니케이션(off-chip communication)을 없앨 수 있어 버스에서 소모되던 파워와 추가적인 수행시간을 줄일 수 있다. 또한, SPM를 사용한 재구성 가속기(410)에서는 데이터를 미리 SPM에 배치하기 위해 규칙적인(regular) 억세스 패턴(access pattern)을 가지는 커넬 코드(kernel code)만을 수행할 수 있었는데, 캐시를 사용하는 경우에는 데이터를 수행 중에 가져오기 때문에 불규칙적인(irregular) 억세스 패턴을 가지는 커넬 코드도 수행할 수 있다.

[0069] 도 5는 또 다른 일실시예에 따른 메모리 장치(100)의 구성을 나타낸 세부 블록도이다.

[0070] 일실시예에 따르면, 메모리 장치(100)는 로직 레이어 및 데이터 레이어(440)를 포함할 수 있다.

[0071] 로직 레이어는 페피퍼럴, 인터페이스 및 비스트 모듈 중 적어도 하나를 구현하고 재구성 가속기(410)를 포함할 수 있다. 재구성 가속기(410)는 로직 레이어의 빈공간에 위치하며 메모리 장치(100)가 처리하는 작업의 적어도 일부를 처리할 수 있다.

[0072] 데이터 레이어(440)는 데이터를 저장할 수 있으며, 적어도 하나의 데이터 레이어(440)로 구성될 수 있다. 데이터 레이어(440)는 디램 레이어(440)로 구성될 수 있다.

[0073] 일실시예에 따르면, 로직 레이어는 로컬 메모리를 포함할 수 있다. 로컬 메모리는 복수 개의 캐시로 구성된 제1 로컬 캐시(420)를 포함할 수 있다. 또한, 데이터 레이어(440) 및 제1 로컬 캐시(420) 사이를 연결하는 제2 로컬 캐시(430)를 포함할 수 있다.

[0074] 도 6은 일실시예에 따른 메모리 장치의 제조 방법의 순서를 나타낸 흐름도이다.

[0075] 단계(610)은, 메모리 장치의 로직 레이어를 형성하는 단계이다. 로직 레이어는 페리퍼럴, 인터페이스 및 비스트 모듈을 구현할 수 있다. 또한, 재구성 가속기를 포함할 수 있다. 재구성 가속기는 로직 레이어의 빈공간에 위치하며 메모리 장치가 처리하는 작업의 적어도 일부를 처리할 수 있다.

[0076] 단계(620)은, 메모리 장치의 데이터 레이어를 형성하는 단계이다. 데이터 레이어는 복수 개로 구성될 수 있다.

[0077] 로직 레이어 및 데이터 레이어는 쓰로우-실리콘-비아(Through-Silicon-Via:TSV)를 이용하여 적층될 수 있다.

[0078] 도 7은 일실시예에 따라, 로직 레이어의 형성 단계를 상세히 나타낸 흐름도이다.

[0079] 일실시예에 따르면, 도 7은, 도 6에서의 로직 레이어를 형성하는 단계를 상세하게 나타낸 흐름도이다.

[0080] 단계(710)은, 로직 레이어에서 제1 영역을 형성하는 단계이다. 상기 제1 영역은, 페리퍼럴(peripheral), 인터페이스(interface) 및 비스트 모듈(BIST module) 중 적어도 하나를 구현할 수 있다. 제1 영역은 로직 레이어의 중심부에 위치할 수도 있다.

[0081] 단계(720)은, 로직 레이어에서 제2 영역을 형성하는 단계이다. 상기 제2 영역은, 재구성 가속기를 포함할 수 있다. 제2 영역은 로직 레이어에서 제1 영역을 제외한 나머지 영역에 해당할 수 있다.

[0082] 도 8은 일실시예에 따라, 제2 영역 형성 단계를 상세히 나타낸 흐름도이다.

[0083] 일실시예에 따르면, 도 8은, 도 7에서의 제2 영역을 형성하는 단계를 상세히 나타낸 흐름도이다.

[0084] 단계(810)은, 제2 영역에서 재구성 가속기를 형성하는 단계이다. 재구성 가속기는 로직 레이어의 빈공간에 위치할 수 있다. 재구성 가속기는 프로세싱 엘리먼트(processing elements:PE)를 포함할 수 있다. 프로세싱 엘리먼트는 어레이(array) 구조로 연결될 수 있다.

[0085] 일실시예에 따르면, 제1 프로세싱 엘리먼트는 상기 제1 프로세싱 엘리먼트에 인접한 제2 프로세싱 엘리먼트와 연결되고, 데이터를 제2 프로세싱 엘리먼트와 송수신할 수 있다. 또한, 프로세싱 엘리먼트는 데이터를 연산하는 기능부를 포함할 수 있다.

[0086] 단계(820)은, 제2 영역에서 로컬 메모리를 형성하는 단계이다. 로컬 메모리는 제1 로컬 캐시 및 제2 로컬 캐시

를 포함할 수 있다. 제2 영역에서 복수 개의 캐시로 구성된 제1 로컬 캐시를 형성할 수 있다. 또한, 제2 영역에서 데이터 레이어 및 제1 로컬 캐시 사이를 연결하는 제2 로컬 캐시를 형성할 수 있다.

[0087] 이상에서 설명된 장치는 하드웨어 구성요소, 소프트웨어 구성요소, 및/또는 하드웨어 구성요소 및 소프트웨어 구성요소의 조합으로 구현될 수 있다. 예를 들어, 실시예들에서 설명된 장치 및 구성요소는, 예를 들어, 프로세서, 콘트롤러, ALU(arithmetic logic unit), 디지털 신호 프로세서(digital signal processor), 마이크로컴퓨터, FPA(field programmable array), PLU(programmable logic unit), 마이크로프로세서, 또는 명령(instruction)을 실행하고 응답할 수 있는 다른 어떠한 장치와 같이, 하나 이상의 범용 컴퓨터 또는 특수 목적 컴퓨터를 이용하여 구현될 수 있다.

[0088] 처리 장치는 운영 체제(OS) 및 상기 운영 체제 상에서 수행되는 하나 이상의 소프트웨어 애플리케이션을 수행할 수 있다. 또한, 처리 장치는 소프트웨어의 실행에 응답하여, 데이터를 접근, 저장, 조작, 처리 및 생성할 수도 있다.

[0089] 이해의 편의를 위하여, 처리 장치는 하나가 사용되는 것으로 설명된 경우도 있지만, 해당 기술분야에서 통상의 지식을 가진 자는, 처리 장치가 복수 개의 처리 요소(processing element) 및/또는 복수 유형의 처리 요소를 포함할 수 있음을 알 수 있다. 예를 들어, 처리 장치는 복수 개의 프로세서 또는 하나의 프로세서 및 하나의 콘트롤러를 포함할 수 있다. 또한, 병렬 프로세서(parallel processor)와 같은, 다른 처리 구성(processing configuration)도 가능하다.

[0090] 소프트웨어는 컴퓨터 프로그램(computer program), 코드(code), 명령(instruction), 또는 이들 중 하나 이상의 조합을 포함할 수 있으며, 원하는 대로 동작하도록 처리 장치를 구성하거나 독립적으로 또는 결합적으로(collectively) 처리 장치를 명령할 수 있다.

[0091] 소프트웨어 및/또는 데이터는, 처리 장치에 의하여 해석되거나 처리 장치에 명령 또는 데이터를 제공하기 위하여, 어떤 유형의 기계, 구성요소(component), 물리적 장치, 가상 장치(virtual equipment), 컴퓨터 저장 매체 또는 장치, 또는 전송되는 신호 파(signal wave)에 영구적으로, 또는 일시적으로 구체화(embody)될 수 있다. 소프트웨어는 네트워크로 연결된 컴퓨터 시스템 상에 분산되어서, 분산된 방법으로 저장되거나 실행될 수도 있다. 소프트웨어 및 데이터는 하나 이상의 컴퓨터 판독 가능 기록 매체에 저장될 수 있다.

[0092] 실시예에 따른 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 실시예를 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다.

[0093] 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다.

[0094] 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다.

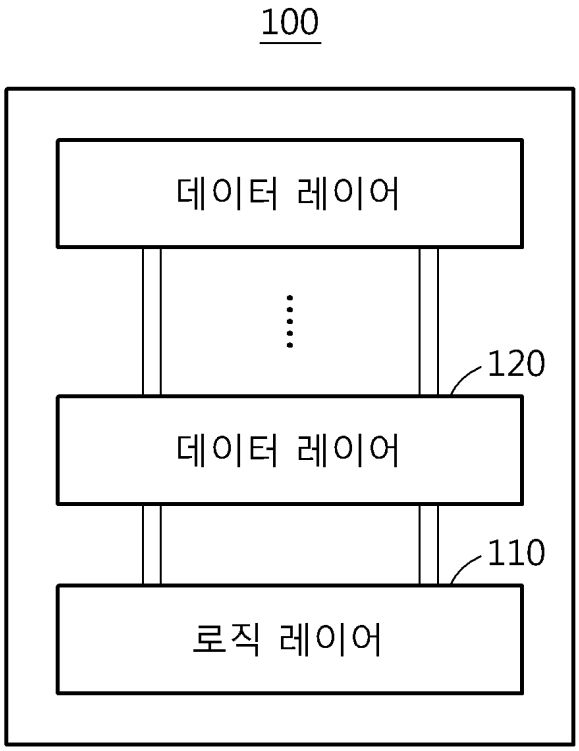
[0095] 상기된 하드웨어 장치는 실시예의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.

[0096] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

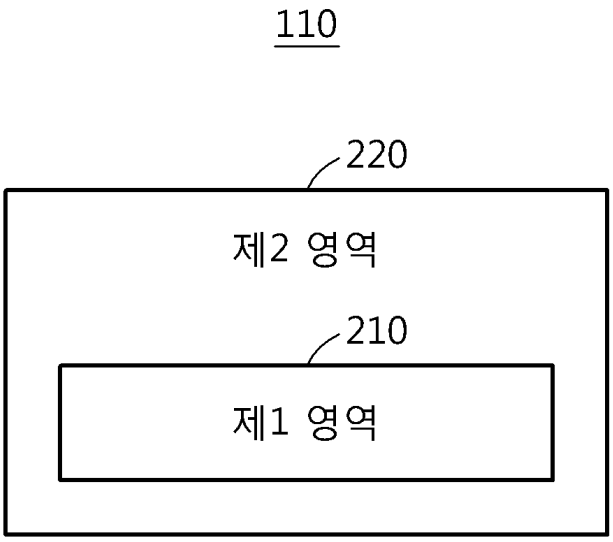
[0097] 그러므로, 본 발명의 범위는 설명된 실시예에 국한되어 정해져서는 아니 되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등한 것들에 의해 정해져야 한다.

도면

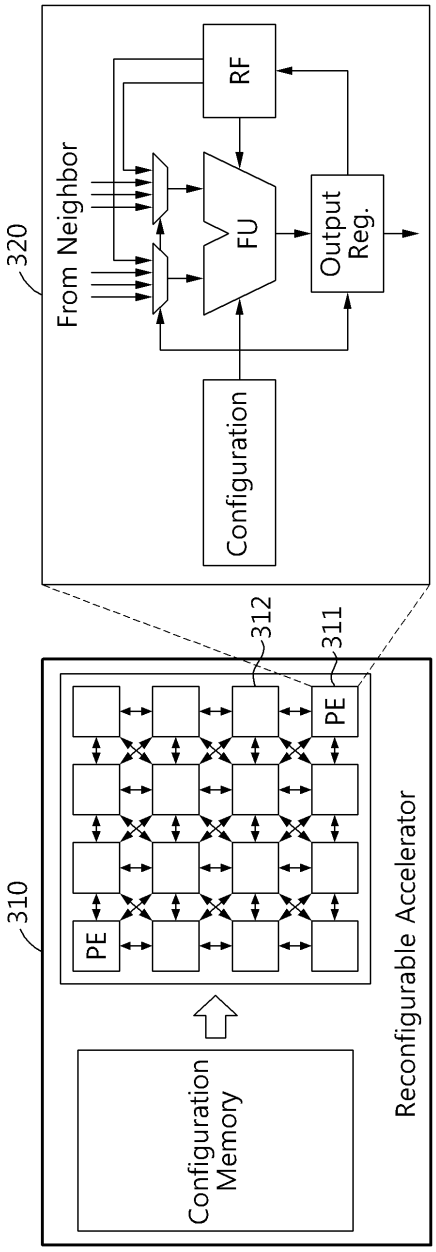
도면1



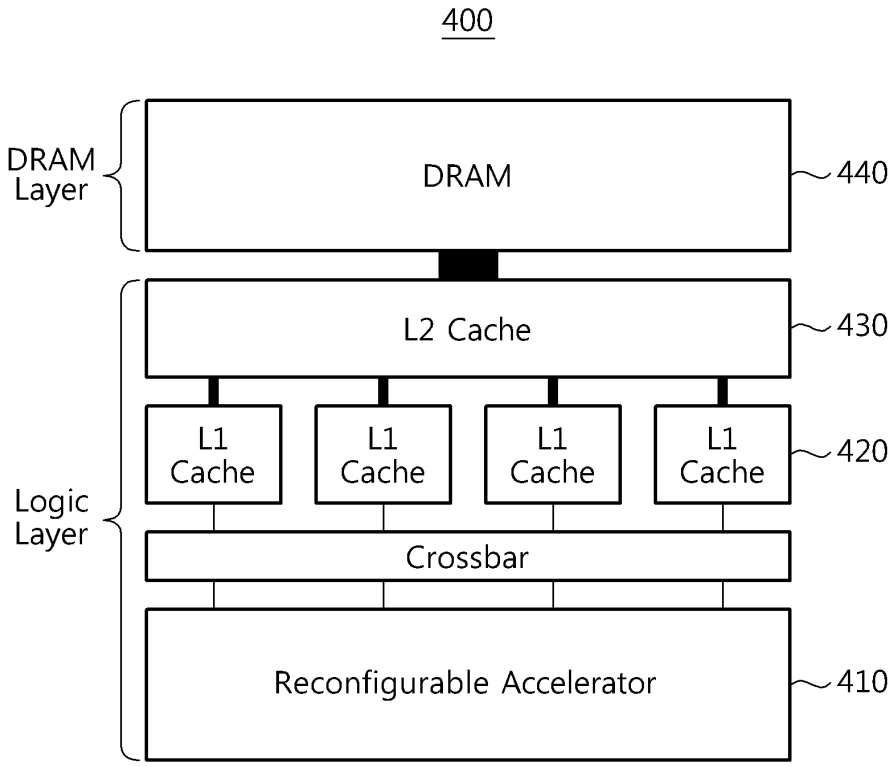
도면2



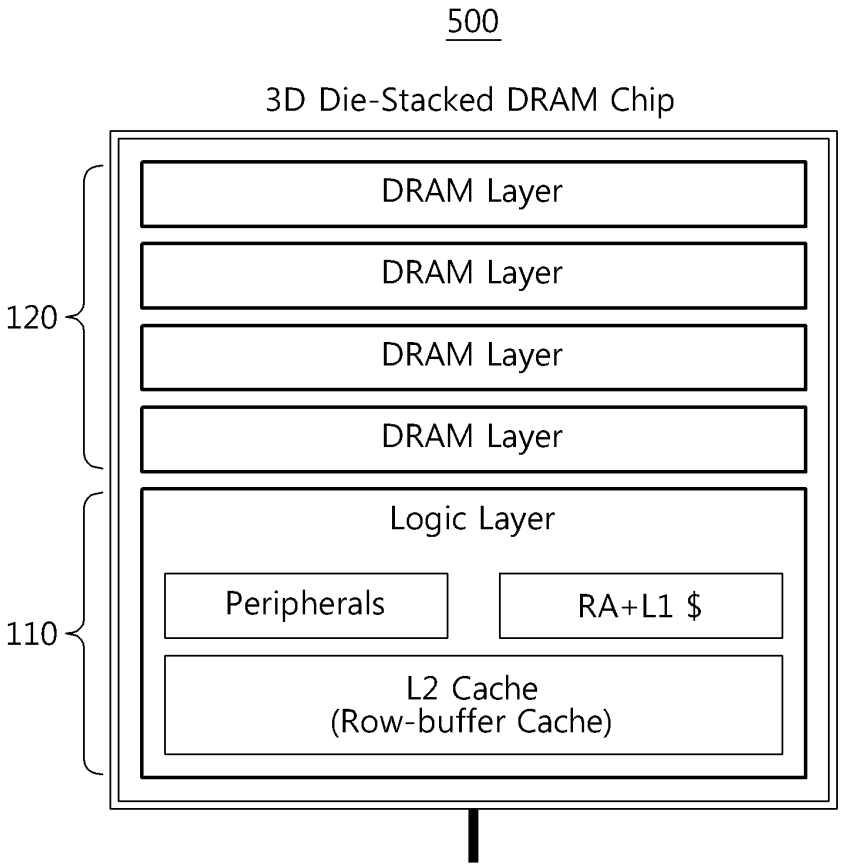
도면3


320
From Neighbor
RF
FU
Output Reg.
Configuration
312
311
PE
310
PE
Configuration Memory
Reconfigurable Accelerator

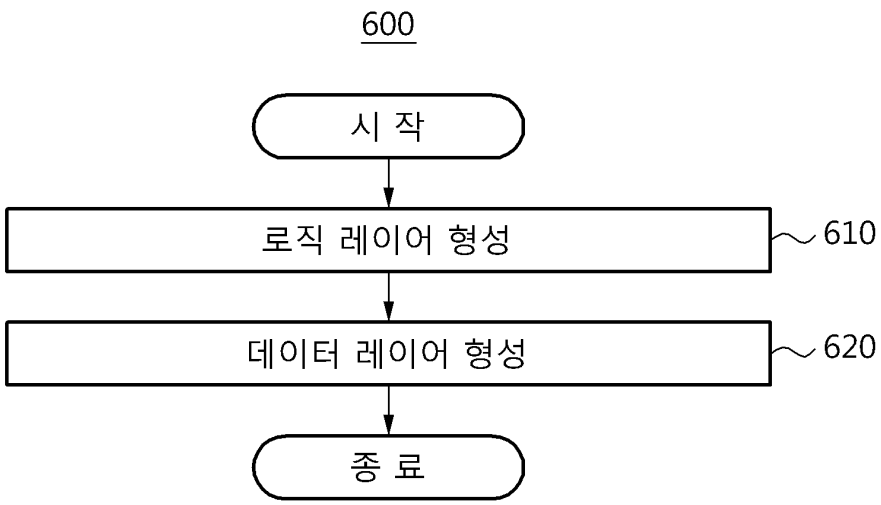
도면4


400
DRAM Layer
DRAM
440
L2 Cache
430
L1 Cache
L1 Cache
L1 Cache
L1 Cache
420
Logic Layer
Crossbar
Reconfigurable Accelerator
410

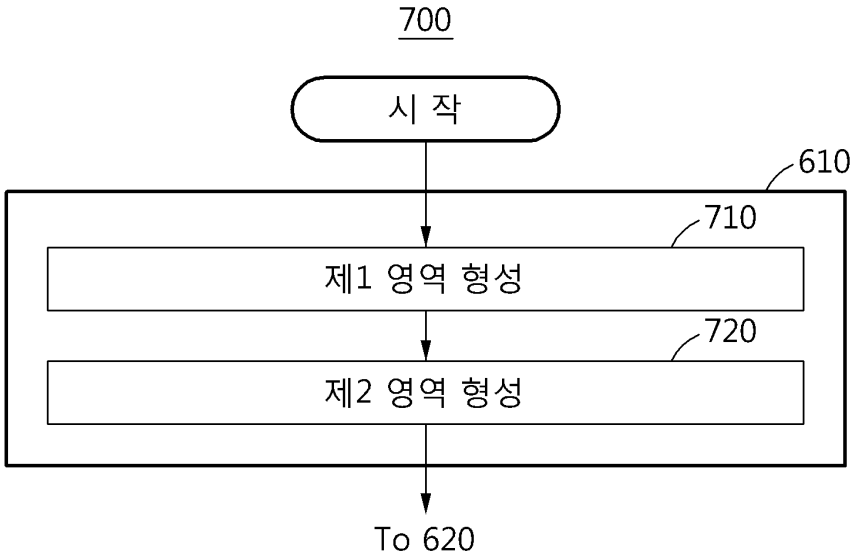
도면5


500
3D Die-Stacked DRAM Chip
DRAM Layer
DRAM Layer
DRAM Layer
DRAM Layer
120
Logic Layer
Peripherals
RA+L1 $
L2 Cache
(Row-buffer Cache)
110

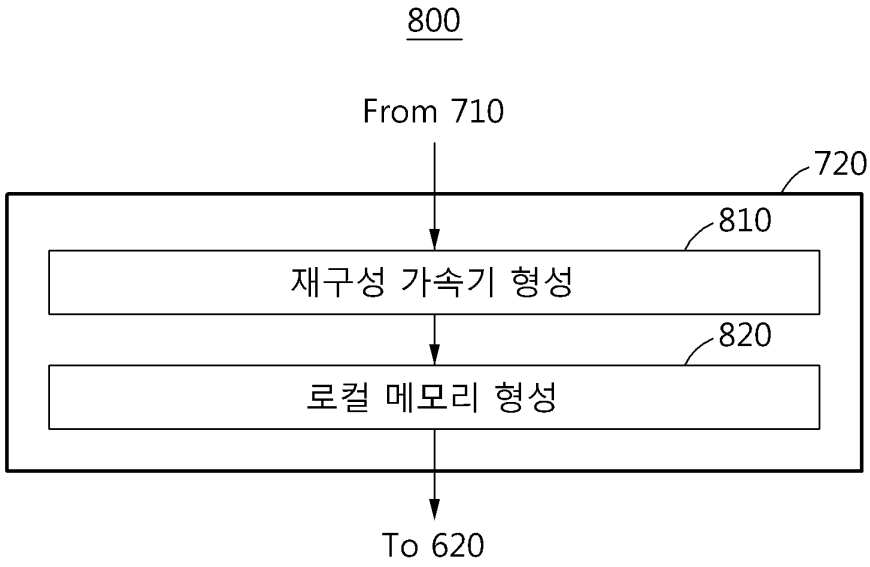
도면6


600
시 작
로직 레이어 형성
610
데이터 레이어 형성
620
종 료

도면7


700
시 작
610
710
제1 영역 형성
720
제2 영역 형성
To 620

도면8


800
From 710
720
810
재구성 가속기 형성
820
로컬 메모리 형성
To 620