



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0123038
(43) 공개일자 2019년10월31일

(51) 국제특허분류(Int. Cl.)
G06F 11/07 (2006.01) G06F 11/00 (2017.01)
(52) CPC특허분류
G06F 11/073 (2013.01)
G06F 11/008 (2013.01)
(21) 출원번호 10-2018-0046707
(22) 출원일자 2018년04월23일
심사청구일자 없음

(71) 출원인
에스케이하이닉스 주식회사
경기도 이천시 부발읍 경충대로 2091
(72) 발명자
김선웅
경기도 화성시 동탄청계로 303-13(청계동, 신안인스빌 리베라 2차) 1112동 603
(74) 대리인
특허법인신성

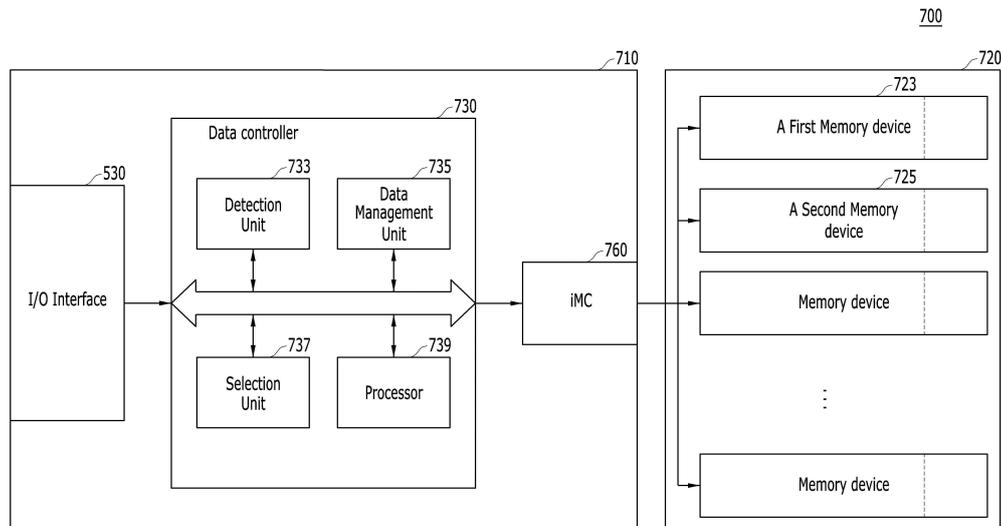
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 메모리 시스템 및 그것의 동작방법

(57) 요약

본 발명의 실시 예에 따른 메모리 시스템에 있어서, 유저 영역 및 오버-프로비저닝 영역을 각각 포함하는 복수의 메모리 장치들; 및 상기 복수의 메모리 장치들 각각을 제어하는 컨트롤러를 포함하고, 상기 컨트롤러는 상기 복수의 메모리 장치들 중 결합 메모리 장치를 검출하는 검출부, 상기 복수의 메모리 장치들 각각의 상기 오버-프로비저닝 영역의 가용 정보를 저장하는 관리부, 상기 가용 정보에 기초하여 상기 복수의 메모리 장치들 중 상기 결합 메모리 장치를 제외한 가용 메모리 장치를 선택하는 선택부, 및 상기 결합 메모리 장치에 저장된 타겟 데이터를 상기 가용 메모리 장치의 상기 오버-프로비저닝 영역에 이동시키는 프로세서를 포함할 수 있다.

대표도



명세서

청구범위

청구항 1

유저 영역 및 오버-프로비저닝 영역을 각각 포함하는 복수의 메모리 장치들; 및
상기 복수의 메모리 장치들 각각을 제어하는 컨트롤러
를 포함하고,
상기 컨트롤러는
상기 복수의 메모리 장치들 중 결함 메모리 장치를 검출하는 검출부,
상기 복수의 메모리 장치들 각각의 상기 오버-프로비저닝 영역의 가용 정보를 저장하는 관리부,
상기 가용 정보에 기초하여 상기 복수의 메모리 장치들 중에서 상기 결함 메모리 장치를 제외한 가용 메모리 장
치를 선택하는 선택부, 및
상기 결함 메모리 장치에 저장된 타겟 데이터를 상기 가용 메모리 장치의 상기 오버-프로비저닝 영역에 이동시
키는 프로세서
를 포함하는 메모리 시스템.

청구항 2

제 1 항에 있어서,
상기 검출부는
상기 복수의 메모리 장치들 각각에 대응하는 신뢰도를 모니터링하여, 사전 설정된 임계 값보다 낮은 신뢰도를
갖는 메모리 장치를 상기 결함 메모리 장치로써 검출하는
메모리 시스템.

청구항 3

제 1 항에 있어서,
상기 관리부는
상기 복수의 메모리 장치들 각각의 메모리 맵을 저장하는
메모리 시스템.

청구항 4

제 3 항에 있어서,
상기 관리부는
상기 타겟 데이터가 상기 오버-프로비저닝 영역으로 이동된 후, 상기 메모리 맵을 업데이트하는
메모리 시스템.

청구항 5

제 1 항에 있어서,
상기 가용 메모리 장치가 복수인 경우,
상기 선택부는
사전 설정된 기준에 따라 상기 가용 메모리 장치 중 적어도 하나의 메모리 장치를 선택하는
메모리 시스템.

청구항 6

제 1 항에 있어서,
상기 결합 메모리 장치가 복구된 경우,
상기 프로세서는 상기 타겟 데이터를 상기 복구된 메모리 장치로 이동시키는
메모리 시스템.

청구항 7

제 6 항에 있어서,
상기 검출부는
상기 복구된 메모리 장치에 대응하는 신뢰도를 업데이트하는
메모리 시스템.

청구항 8

제 6 항에 있어서,
상기 관리부는
상기 타겟 데이터에 대응하는 이동 동작이 종료된 후 상기 메모리 맵을 업데이트하는
메모리 시스템.

청구항 9

메모리 시스템의 동작방법에 있어서,
복수의 메모리 장치들 중 결합 메모리 장치를 검출하는 단계;
상기 복수의 메모리 장치들 각각의 오버-프로비저닝 영역의 가용 정보를 저장하는 단계;
상기 가용 정보에 기초하여 상기 복수의 메모리 장치들 중에서 상기 결합 메모리 장치를 제외한 가용 메모리 장
치를 선택하는 단계; 및
상기 결합 메모리 장치에 저장된 타겟 데이터를 상기 가용 메모리 장치의 상기 오버-프로비저닝 영역에 이동시
키는 단계
를 포함하는 메모리 시스템의 동작방법.

청구항 10

제 9 항에 있어서,

상기 검출하는 단계는

상기 복수의 메모리 장치들 각각에 대응하는 신뢰도를 모니터링하여, 사전 설정된 임계 값보다 낮은 신뢰도를 갖는 메모리 장치를 상기 결합 메모리 장치로써 검출하는

메모리 시스템의 동작방법.

청구항 11

제 9 항에 있어서,

상기 복수의 메모리 장치들 각각의 메모리 맵을 저장하는 단계

를 더 포함하는 메모리 시스템의 동작방법.

청구항 12

제 11 항에 있어서,

상기 타겟 데이터가 상기 오버-프로비저닝 영역으로 이동된 후, 상기 메모리 맵을 업데이트하는 단계

를 더 포함하는 메모리 시스템의 동작방법.

청구항 13

제 9 항에 있어서,

상기 가용 메모리 장치가 복수인 경우,

상기 가용 메모리 장치를 선택하는 단계는

사전 설정된 기준에 따라 상기 가용 메모리 장치 중 적어도 하나의 메모리 장치를 선택하는

메모리 시스템의 동작방법.

청구항 14

제 9 항에 있어서,

상기 결합 메모리 장치가 복구된 경우,

상기 타겟 데이터를 복구 메모리 장치로 이동시키는 단계

를 더 포함하는 메모리 시스템의 동작방법.

청구항 15

제 14 항에 있어서,

상기 복구 메모리 장치에 대응하는 신뢰도를 업데이트하는 단계

를 더 포함하는 메모리 시스템의 동작방법.

청구항 16

제 14 항에 있어서,
 상기 타겟 데이터에 대응하는 이동 동작이 종료된 후 상기 메모리 맵을 업데이트하는 단계를 더 포함하는 메모리 시스템의 동작방법.

청구항 17

복수의 메모리 시스템들; 및
 상기 복수의 메모리 시스템들 각각과 통신하는 메모리 시스템 관리부를 포함하고,
 상기 복수의 메모리 시스템들 각각은
 유저 영역 및 오버-프로비저닝 영역을 각각 포함하는 복수의 메모리 장치들 및 상기 복수의 메모리 장치들 각각을 제어하는 컨트롤러를 포함하며,
 상기 컨트롤러는
 상기 복수의 메모리 장치들 각각에 대응하는 신뢰도를 모니터링하여, 사전 설정된 임계 값보다 작은 신뢰도를 갖는 메모리 장치를 상기 결합 메모리 장치로써 검출하는 검출부,
 상기 복수의 메모리 장치들 각각의 상기 오버-프로비저닝 영역의 가용 정보를 저장하며, 상기 복수의 메모리 장치들 각각의 메모리 맵들을 저장하는 관리부,
 상기 가용 정보에 기초하여 상기 복수의 메모리 장치들 중 상기 결합 메모리 장치를 제외한 가용 메모리 장치를 선택하는 선택부, 및
 상기 결합 메모리 장치에 저장된 타겟 데이터를 상기 가용 메모리 장치의 상기 오버-프로비저닝 영역에 이동시키는 프로세서
 를 포함하는 컴퓨팅 시스템.

청구항 18

제 17 항에 있어서,
 상기 MMU는
 상기 복수의 메모리 장치들 각각에 대응하는 신뢰도, 상기 가용 정보 및 상기 메모리 맵을 상기 복수의 메모리 시스템들 각각으로부터 전달받아 저장하는
 컴퓨팅 시스템.

청구항 19

제 18 항에 있어서,
 제 1 메모리 시스템에 상기 결합 메모리 장치가 발생되고, 상기 가용 메모리 장치가 존재하지 않는 경우,
 상기 제 1 메모리 시스템은 타겟 데이터를 상기 MMU로 전달하고,
 상기 MMU는 상기 가용 정보에 기초하여 상기 타겟 데이터를 제 2 메모리 시스템으로 전달하며,
 상기 제 2 메모리 시스템은 상기 MMU로부터 전달받은 상기 타겟 데이터를 상기 가용 메모리 장치의 상기 오버-

프로비저닝 영역에 저장하는
컴퓨팅 시스템.

청구항 20

제 19 항에 있어서,
상기 결합 메모리 장치가 복구된 경우,
상기 제 2 메모리 시스템은 상기 타겟 데이터를 상기 오버-프로비저닝 영역에서 리드하여 상기 MMU에 전달하고,
상기 MMU는 상기 타겟 데이터를 상기 제 1 메모리 시스템으로 전달하며,
상기 제 1 메모리 시스템은 상기 타겟 데이터를 복구 메모리 장치에 저장하는
컴퓨팅 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 메모리 시스템에 관한 것으로, 보다 구체적으로는 효율적인 데이터 처리를 위한 메모리 시스템 및 그것의 동작방법에 관한 것이다.

배경 기술

[0003] 데이터는 4차 산업혁명 시대 기업들의 비즈니스에 가장 중요한 자산이 되고 있으며, 이에 대규모 데이터를 빠르게 전송 및 분석하도록 지원하는 최신 기술에 대한 수요가 점차 증가하고 있다. 예를 들어, 인공지능, 자율주행, 로봇, 헬스케어, 가상현실(virtual reality, VR)/증강현실(augmented reality, AR), 스마트홈 등이 확대됨에 따라 서버나 데이터 센터에 대한 수요의 증가가 예상되고 있다.

[0004] 레거시(legacy) 데이터 센터는 자원들(예; 컴퓨팅, 네트워킹, 스토리지)을 하나의 장비 내에 포함시키는 구조였다. 그러나 미래의 대용량 데이터 센터는 자원들을 각각 별개로 구성하고, 논리적으로 자원들을 재구성하는 구조를 가질 수 있다. 예를 들어, 대용량 데이터 센터는 자원들을 각각 랙(rack) 수준에서 모듈화하고, 용도에 따라 자원들을 재구성하여 공급할 수 있는 구조를 가질 수 있다. 따라서 미래의 대용량 데이터 센터에 사용하기 적합한 통합형 스토리지 또는 메모리 디바이스가 요구되고 있다.

발명의 내용

해결하려는 과제

[0006] 본 발명은 시스템의 가용성(availability)을 높게 유지하면서 결합이 발생된 메모리 장치를 복구할 수 있는 메모리 시스템, 컴퓨팅 시스템 및 그들 각각의 동작방법에 대하여 제안한다.

과제의 해결 수단

[0008] 본 발명의 실시 예들에 따른 메모리 시스템에 있어서, 유저 영역 및 오버-프로비저닝 영역을 각각 포함하는 복수의 메모리 장치들; 및 상기 복수의 메모리 장치들 각각을 제어하는 컨트롤러를 포함하고, 상기 컨트롤러는 상기 복수의 메모리 장치들 중 결합 메모리 장치를 검출하는 검출부, 상기 복수의 메모리 장치들 각각의 상기 오버-프로비저닝 영역의 가용 정보를 저장하는 관리부, 상기 가용 정보에 기초하여 상기 복수의 메모리 장치들 중 상기 결합 메모리 장치를 제외한 가용 메모리 장치를 선택하는 선택부, 및 상기 결합 메모리 장치에 저장된 타

겟 데이터를 상기 가용 메모리 장치의 상기 오버-프로비저닝 영역에 이동시키는 프로세서를 포함할 수 있다.

[0009] 본 발명의 실시 예들에 따른 메모리 시스템의 동작방법에 있어서, 복수의 메모리 장치들 중 결합 메모리 장치를 검출하는 단계; 상기 복수의 메모리 장치들 각각의 오버-프로비저닝 영역의 가용 정보를 저장하는 단계; 상기 가용 정보에 기초하여 상기 복수의 메모리 장치들 중 상기 결합 메모리 장치를 제외한 가용 메모리 장치를 선택하는 단계; 및 상기 결합 메모리 장치에 저장된 타겟 데이터를 상기 가용 메모리 장치의 상기 오버-프로비저닝 영역에 이동시키는 단계를 포함할 수 있다.

[0010] 본 발명의 실시 예에 따른 컴퓨팅 시스템에 있어서, 복수의 메모리 시스템들; 및 상기 복수의 메모리 시스템들 각각과 통신하는 메모리 시스템 관리부를 포함하고, 상기 복수의 메모리 시스템들 각각은 유저 영역 및 오버-프로비저닝 영역을 각각 포함하는 복수의 메모리 장치들 및 상기 복수의 메모리 장치들 각각을 제어하는 컨트롤러를 포함하며, 상기 컨트롤러는 상기 복수의 메모리 장치들 각각에 대응하는 신뢰도를 모니터링하여, 사전 설정된 임계 값보다 작은 신뢰도를 갖는 메모리 장치를 상기 결합 메모리 장치로써 검출하는 검출부, 상기 복수의 메모리 장치들 각각의 상기 오버-프로비저닝 영역의 가용 정보를 저장하며, 상기 복수의 메모리 장치들 각각의 메모리 맵들을 저장하는 관리부, 상기 가용 정보에 기초하여 상기 복수의 메모리 장치들 중 상기 결합 메모리 장치를 제외한 가용 메모리 장치를 선택하는 선택부, 및 상기 결합 메모리 장치에 저장된 타겟 데이터를 상기 가용 메모리 장치의 상기 오버-프로비저닝 영역에 이동시키는 프로세서를 포함할 수 있다.

발명의 효과

[0012] 본 발명의 실시 예에 따르면, 시스템의 가용성을 높게 유지하면서 결합이 발생된 메모리 장치를 복구할 수 있어 전체적인 시스템의 효율성을 증가시킬 수 있다.

도면의 간단한 설명

- [0014] 도 1은 데이터 처리 시스템을 도시하는 도면이다.
- 도 2 본 발명의 실시 예에 따른 컴퓨팅 랙의 구조를 개략적으로 도시하는 도면이다.
- 도 3은 본 발명의 실시 예에 따른 컴퓨팅 랙의 블록 구성을 도시하는 도면이다.
- 도 4는 본 발명의 실시 예에 따른 연산 보드의 구성을 도시하는 도면이다.
- 도 5는 본 발명의 실시 예에 따른 메모리 보드의 구성을 개략적으로 나타낸 도면들이다.
- 도 6은 본 발명의 실시 예에 따른 메모리 장치의 구조를 나타낸 도면이다.
- 도 7은 본 발명의 일 실시 예에 따른 메모리 시스템의 구조를 나타낸 도면이다.
- 도 8은 본 발명의 일 실시 예에 따른 데이터 컨트롤러의 동작을 나타낸 흐름도이다.
- 도 9는 본 발명의 일 실시 예에 따른 데이터 컨트롤러의 동작을 나타내는 흐름도이다.
- 도 10은 본 발명의 일 실시 예에 따른 컴퓨팅 시스템의 구조를 나타내는 도면이다.
- 도 11은 본 발명의 일 실시 예에 따른 컴퓨팅 시스템의 동작을 나타내는 흐름도이다.
- 도 12는 본 발명의 일 실시 예에 따른 컴퓨팅 시스템의 동작을 나타낸 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0015] 이하, 본 발명에 따른 바람직한 실시 예를 첨부한 도면을 참조하여 상세히 설명한다. 하기의 설명에서는 본 발명에 따른 동작을 이해하는데 필요한 부분만이 설명되며 그 이외 부분의 설명은 본 발명의 요지를 흐트리지 않도록 생략될 것이라는 것을 유의하여야 한다.

[0016] 도 1은 데이터 처리 시스템(10)을 도시하는 도면이다. 도 1을 참조하면, 데이터 처리 시스템(10)은 다수의 컴퓨팅 랙들(computing racks, 20)과 관리 인터페이스(management interface, 30), 그리고 이들 사이의 통신이 가능하도록 하는 네트워크(network, 40)를 포함할 수 있다. 이러한 랙스케일 구조(rack-scale architecture)를

가지는 데이터 처리 시스템(10)은 대용량 데이터 처리를 위한 데이터 센터 등에 사용될 수 있다.

- [0017] 다수의 컴퓨팅 랙(20)들은 하나의 컴퓨팅 시스템을 구현할 수 있다. 이러한 컴퓨팅 랙(20)들의 구체적인 구성 및 동작에 대한 설명은 후술될 것이다.
- [0018] 관리 인터페이스(30)은 사용자가 데이터 처리 시스템(10)을 조정, 운영 또는 관리할 수 있도록 하는 인터랙티브 인터페이스(interactive interface)를 제공할 수 있다. 관리 인터페이스(30)는 컴퓨터, 멀티프로세서 시스템, 서버, 랙마운트(rack-mount) 서버, 보드(board) 서버, 랩탑(lap-top) 컴퓨터, 노트북 컴퓨터, 태블릿 컴퓨터, 웨어러블 컴퓨팅 시스템, 네트워크 기기, 웹 기기, 분산 컴퓨팅 시스템, 프로세서 기반 시스템, 및/또는 소비자 전자 기기를 포함하는, 임의 유형의 연산 디바이스로서 구현될 수 있다.
- [0019] 일부 실시 예들에서, 관리 인터페이스(30)은 컴퓨팅 랙(20)들에 의해 수행될 수 있는 연산 기능들이나, 관리 인터페이스(30)에 의해 수행될 수 있는 사용자 인터페이스 기능들을 가지는 분산 시스템에 의해 구현될 수 있다. 다른 일부 실시 예들에서, 관리 인터페이스(30)는 네트워크(40)를 통하여 분산된 다중 컴퓨팅 시스템들에 의해 구성되고 클라우드(cloud)로서 동작하는 가상 서버(virtual server)에 의해 구현될 수 있다. 관리 인터페이스(30)은 프로세서, 입력/출력 서브시스템, 메모리, 데이터 스토리지 디바이스 및 통신 회로를 포함할 수 있다.
- [0020] 네트워크(40)은 컴퓨팅 랙들과 관리 인터페이스(30) 사이 및/또는 컴퓨팅 랙들 사이에서의 데이터를 송수신할 수 있다. 네트워크(40)은 적절한 수의 다양한 유선 및/또는 무선 네트워크들에 의해 구현될 수 있다. 예를 들어, 네트워크(40)은 유선 또는 무선 LAN(local area network), WAN(wide area network) 셀룰라 네트워크, 및/또는 인터넷과 같이 공개적으로 액세스 가능한 글로벌 네트워크(publicly-accessible, global network)에 의해 구현되거나 이를 포함할 수 있다. 추가적으로, 네트워크(40)은 보조적인 컴퓨터들, 라우터(router)들 및 스위치들과 같은 적절한 수의 보조적인 네트워크 디바이스들을 포함할 수 있다. 또한, 네트워크(40)는 CCIX(Cache Coherent Interconnect for accelerators) 및 GEN-Z와 같은 인터페이스 규격에 의해서 연결될 수 있다.
- [0021] 도 2는 본 발명의 실시 예에 따른 컴퓨팅 랙 구조를 개략적으로 도시하는 도면이다.
- [0022] 도 2를 참조하면, 컴퓨팅 랙(20)은 구성요소들의 구조, 형태 및 호칭 등에 제한되는 것은 아니지만, 다양한 형태의 구성요소들을 포함할 수 있다. 예를 들어, 컴퓨팅 랙(20)은 다수의 드로워(drawer)들(21 내지 29)를 포함할 수 있다. 다수의 드로워들(21 내지 29) 각각은 다수의 보드(board)들을 포함할 수 있다.
- [0023] 다양한 실시 예들에 있어서, 컴퓨팅 랙(20)은 적절한 수의 연산 보드(compute board), 메모리 보드(memory board) 및/또는 상호접속 보드(interconnect board)의 조합에 의해 구현될 수 있다. 여기서는 컴퓨팅 랙(20)이 다수의 보드들의 조합에 의해 구현되는 것으로 정의되고 있지만, 이를 대신하여, 드로워들, 모듈들, 트레이들, 보드들, 샷시들 또는 유닛들 등의 다양한 이름으로 구현되는 것으로 정의될 수 있음에 유의하여야 한다. 이러한 컴퓨팅 랙(20)의 구성요소들은 구현의 편의상 기능별로 분류 및 구별되는 구조를 가질 수 있다. 제한되는 것은 아니지만, 컴퓨팅 랙(20)은 상단부터 상호접속 보드, 연산 보드, 메모리 보드의 순서로 분류된 구조를 가질 수 있다. 이러한 컴퓨팅 랙(20) 및 이에 의해 구현되는 컴퓨팅 시스템은 "랙스케일 시스템(rack-scale system)" 또는 "분류 시스템(disaggregated system)"으로 명명될 수 있다.
- [0024] 다양한 실시 예들에 있어서, 컴퓨팅 시스템은 하나의 컴퓨팅 랙(20)에 의해 구현될 수 있다. 이를 대신하여, 컴퓨팅 시스템은 2개 이상의 컴퓨팅 랙들에 포함되는 모든 구성요소들에 의해 구현되거나, 2개 이상의 컴퓨팅 랙들에 포함되는 일부 구성요소들의 조합에 의해 구현되거나, 하나의 컴퓨팅 랙(20)에 포함되는 일부 구성요소들에 의해 구현될 수 있다.
- [0025] 다양한 실시 예들에 있어서, 컴퓨팅 시스템은 컴퓨팅 랙(20)에 포함되는 적절한 수의 연산 보드, 메모리 보드 및 상호접속 보드(interconnect board)의 조합에 의해 구현될 수 있다. 예를 들어, 컴퓨팅 시스템(20A)은 2개의 연산 보드들, 3개의 메모리 보드들 및 1개의 상호접속 보드에 의해 구현될 수 있다. 다른 예로, 컴퓨팅 시스템(20B)은 3개의 연산 보드들, 2개의 메모리 보드들 및 1개의 상호접속 보드에 의해 구현될 수 있다. 또 다른 예로, 컴퓨팅 시스템(20C)은 1개의 연산 보드들, 4개의 메모리 보드들 및 1개의 상호접속 보드에 의해 구현될 수 있다.
- [0026] 비록 도 2에서는 컴퓨팅 랙(20)이 적절한 수의 연산 보드, 메모리 보드 및/또는 상호접속 보드의 조합에 의해 구현되는 경우를 도시하고 있지만, 컴퓨팅 랙(20)은 통상의 서버 등에서 확인될 수 있는, 파워 시스템, 냉각 시스템, 입력/출력 디바이스들과 같은 추가적인 구성요소들을 포함할 수 있다.

- [0027] 도 3은 본 발명의 실시 예에 따른 컴퓨팅 랙(20)의 블록 구성을 도시하는 도면이다.
- [0028] 도 3을 참조하면, 컴퓨팅 랙(20)은 다수의 연산 보드들(compute boards, 200), 다수의 메모리 보드들(memory boards, 400) 및 상호접속 보드(interconnect board, 300)를 포함할 수 있다. 다수의 연산 보드들(200)은 풀드 연산 보드(pooled compute boards), 풀드 연산 시스템 등으로 불리울 수 있다. 유사하게, 다수의 메모리 보드들은 풀드 메모리 보드(pooled memory board), 풀드 메모리 시스템 등으로 불리울 수 있다. 여기서는 컴퓨팅 시스템이 다수의 보드들의 조합에 의해 구현되는 것으로 정의되고 있지만, 이를 대신하여 드로워들, 모듈들, 트레이들, 보드들, 샷시들 또는 유닛들 등의 다양한 이름으로 구현되는 것으로 정의될 수 있음에 유의하여야 한다.
- [0029] 다수의 연산 보드들(200) 각각은 하나 또는 그 이상의 프로세서, 프로세싱/컨트롤 회로, 또는 중앙처리장치(central processing unit, CPU)와 같은 프로세싱 요소를 포함할 수 있다.
- [0030] 다수의 메모리 보드들(400) 각각은 다수의 휘발성 메모리(volatile memory)들 및/또는 비휘발성 메모리(nonvolatile memory)들과 같은 다양한 형태의 메모리 장치들을 포함할 수 있다. 예를 들어, 다수의 메모리 보드들(400) 다수의 DRAM(Dynamic Random Access Memory)들, 플래시 메모리(flash memory)들, 메모리 카드들, 하드 디스크 드라이브(hard disk drive, HDD)들, 솔리드 스테이트 드라이브(solid state drive, SSD)들, 및/또는 이들의 조합들을 포함할 수 있다.
- [0031] 다수의 메모리 보드들(400) 각각은 연산 보드들(200) 각각에 포함되는 하나 이상의 프로세싱 요소에 의해 분할되거나, 할당되거나, 또는 지정되어 사용될 수 있다. 또한 다수의 메모리 보드들(400) 각각은 연산 보드들(200)에 의해 초기화 및/또는 수행될 수 있는 하나 이상의 오퍼레이팅 시스템(operating system, OS)들을 저장할 수 있다.
- [0032] 상호접속 보드(300)은 연산 보드들(200) 각각에 포함되는 하나 이상의 프로세싱 요소에 의해 분할, 할당 또는 지정되어 사용될 수 있는, 임의의 통신 회로, 디바이스, 또는 이들의 결합에 의해 구현될 수 있다. 예를 들어, 상호접속 보드(300)은 임의의 개수의 네트워크 인터페이스 포트들, 카드들, 또는 스위치들로서 구현될 수 있다. 상호접속 보드(300)은 통신을 실행시키기 위한, 하나 이상의 유선 또는 유선 통신 기술들과 관련된 프로토콜들을 사용할 수 있다. 예를 들어, 상호접속 보드(300)은 PCIe(Peripheral Component Interconnect Express), QPI(QuickPath Interconnect), 이더넷(Ethernet) 등과 같은 프로토콜들에 따라 연산 보드들(200)과 메모리 보드들(400) 사이의 통신을 지원할 수 있다. 뿐만 아니라, 상호접속 보드(300)는 CCIX(Cache Coherent Interconnect for accelerators) 및 GEN-Z와 같은 인터페이스 규격에 의해서 연산 보드들(200)과 연결될 수 있다.
- [0033] 도 4는 본 발명의 실시 예에 따른 연산(200)의 구성을 도시하는 도면이다.
- [0034] 도 4를 참조하면, 연산 보드(200)은 하나 이상의 중앙처리장치(central processing unit, CPU, 210), 하나 이상의 로컬 메모리(local memory, 220) 및 입출력(input/output, I/O) 인터페이스(interface, 230)를 포함할 수 있다.
- [0035] CPU(210)은 도 3에 도시된 다수의 메모리 보드들(400) 중에서 사용하기 위한 적어도 하나의 메모리 보드를 분할, 할당 또는 지정할 수 있다. 또한 CPU(210)은 분할, 할당 또는 지정된 적어도 하나의 메모리 보드를 초기화하고, 이들을 통해 데이터의 리드 동작, 라이트(또는 프로그램) 동작 등을 수행할 수 있다.
- [0036] 로컬 메모리(220)은 CPU(210)의 동작 수행 중에 필요한 데이터를 저장할 수 있다. 다양한 실시 예들에 있어서, 하나의 로컬 메모리(220)은 하나의 CPU(210)에 일대일 대응하는 구조를 가질 수 있다.
- [0037] I/O 인터페이스(230)은 도 3의 상호접속 보드(300)을 통한 CPU(210)과 메모리 보드들(400) 사이에서의 인터페이스를 지원할 수 있다. I/O 인터페이스(230)은 하나 이상의 유선 또는 유선 통신 기술들과 관련된 프로토콜들을 사용하여, CPU(210)으로부터 상호접속 보드(300)으로의 송신 데이터를 출력하고, 상호접속 보드(300)으로부터 CPU(210)으로의 수신 데이터를 입력할 수 있다. 예를 들어, I/O 인터페이스(230)는 PCIe(Peripheral Component Interconnect Express), QPI(QuickPath Interconnect), 이더넷(Ethernet) 등과 같은 프로토콜들에 따라 CPU(210)과 상호접속 보드(300) 사이의 통신을 지원할 수 있다. 뿐만 아니라, I/O 인터페이스(230)는 CCIX(Cache Coherent Interconnect for accelerators) 및 GEN-Z와 같은 인터페이스 규격에 따라 CPU(210)와 상호접속 보드(300) 사이의 통신을 지원할 수 있다.
- [0038] 도 5는 본 발명의 실시 예에 따른 메모리 보드(400)의 구성을 도시하는 도면들이다.

- [0039] 도 5를 참조하면, 메모리 보드(400)은 컨트롤러(410)와 다수의 메모리 장치들 (420)을 포함할 수 있다. 다수의 메모리 장치들(420)은 컨트롤러(410)의 제어에 따라 데이터를 저장(또는 라이트)하고, 저장된 데이터를 출력(또는 리드)할 수 있다. 다수의 메모리 장치들(420)은 제1 그룹의 메모리 장치들(420A), 제2 그룹의 메모리 장치들(420B) 및 제3 그룹의 메모리 장치들(420C)를 포함할 수 있다. 제1 그룹의 메모리 장치들(420A), 제2 그룹의 메모리 장치들(420B) 및 제3 그룹의 메모리 장치들(420C)은 서로 동일한 특성을 가질 수도 있고, 서로 다른 특성을 가질 수 있다. 다양한 실시 예들에 있어서, 제1 그룹의 메모리장치 들(420A), 제2 그룹의 메모리장치들(420B) 및 제3 그룹의 메모리 장치들(420C)은 저장 용량(capacity) 또는 레이턴시(latency) 측면에서 서로 다른 특성을 가지는 메모리 장치들일 수 있다.
- [0040] 컨트롤러(410)은 데이터 컨트롤러(data controller, 510), 메모리 컨트롤러들(memory controller, MC, 520A-520C), 및 I/O 인터페이스(530)을 포함할 수 있다.
- [0041] 데이터 컨트롤러(510)은 도 3의 연산 보드들(200)과 다수의 메모리 장치들(420) 사이에서 송수신되는 데이터를 제어할 수 있다. 예를 들어, 라이트 요청 또는 커맨드에 응답하여, 데이터 컨트롤러(510)은 연산 보드들(200)로부터의 라이트를 위한 데이터를 수신하고, 이 데이터를 다수의 메모리 장치들(420) 중에서 해당하는 메모리에 라이트하는 라이트 동작을 제어할 수 있다. 다른 예로, 리드 요청 또는 커맨드에 응답하여, 데이터 컨트롤러(510)은 연산 보드들(200)으로부터 다수의 메모리 장치들(420) 중에서 특정 메모리에 저장된 데이터를 리드하고, 리드 데이터를 연산 보드들(200) 중에서 해당하는 연산 보드로 출력하는 리드 동작을 제어할 수 있다.
- [0042] 메모리 컨트롤러들(520A-520C)는 데이터 컨트롤러(510)과 다수의 메모리 장치들 (420)의 사이에 위치하며, 이들 사이에서의 인터페이싱을 지원할 수 있다. 메모리 컨트롤러들(520)은 다수의 메모리 장치들(420)에 포함되는 제 1 그룹의 메모리 장치들(420A), 제2 그룹의 메모리 장치들(420B) 및 제3 그룹의 메모리 장치들(420C)를 각각에 대응하는 메모리 컨트롤러(iMC0, 520A), 메모리 컨트롤러(iMC1, 520B), 메모리 컨트롤러(iMC2, 520C)를 포함할 수 있다. 메모리 컨트롤러(iMC0, 520A)는 데이터 컨트롤러(510)와 제 1 그룹의 메모리 장치들(420A)의 사이에 위치하며, 이들 사이에서의 데이터 송수신을 지원할 수 있다. 메모리 컨트롤러(iMC1, 520B)는 데이터 컨트롤러(510)과 제2 그룹의 메모리 장치들(420B)의 사이에 위치하며, 이들 사이에서의 데이터 송수신을 지원할 수 있다. 메모리 컨트롤러(iMC2, 520C)는 데이터 컨트롤러(510)과 제3 그룹의 메모리 장치들(420C)의 사이에 위치하며, 이들 사이에서의 데이터 송수신을 지원할 수 있다. 예를 들면, 제 3 그룹의 메모리 장치들(420C)가 플래시 메모리인 경우, 메모리 컨트롤러(iMC2, 520C)는 플래시 컨트롤러(flash controller)일 수 있다. 제 1 그룹 내지 제 3 그룹의 메모리 장치들(420A 내지 420C)는 설명의 편의를 위한 예시일 뿐이며, 이에 제한되는 것은 아니다.
- [0043] I/O 인터페이스(530)은 도 3의 상호접속 보드(300)을 통한 데이터 컨트롤러(510)과 연산 보드들(200) 사이에서의 인터페이싱을 지원할 수 있다. I/O 인터페이스(530)는 하나 이상의 유선 또는 유선 통신 기술들과 관련된 프로토콜들을 사용하여, 데이터 컨트롤러(510)으로부터 상호접속 보드(300)으로의 송신 데이터를 출력하고, 상호 접속 보드(300)으로부터 데이터 컨트롤러(510)으로의 수신 데이터를 입력할 수 있다. 예를 들어, I/O 인터페이스(530)는 PCIe(Peripheral Component Interconnect Express), QPI(QuickPath Interconnect), 이더넷(Ethernet) 등과 같은 프로토콜들에 따라 데이터 컨트롤러(510)과 상호접속 보드(300) 사이의 통신을 지원할 수 있다. 뿐만 아니라, I/O 인터페이스(530)는 CCIX(Cache Coherent Interconnect for accelerators) 및 GEN-Z와 같은 인터페이스 규격에 따라 데이터 컨트롤러(510)와 상호접속 보드(300) 사이의 통신을 지원할 수 있다.
- [0044] 전술한 바와 같이, 미래의 데이터 센터와 같은 서버 시스템 또는 데이터 처리 시스템은 연산 보드들, 메모리 또는 스토리지 보드들 등과 같은 다수의 보드들이 단위 랙 내에서 구별되어 장착되는 구조를 가질 수 있다. 이때 하나의 메모리 보드에는 다양한 사용자 워크로드(user workload)를 충족시키기 위하여 서로 다른 특성을 가지는 다수의 메모리 장치들이 포함될 수 있다. 즉, 하나의 메모리 보드는 DRAM, PCRAM, MRAM, STT-RAM, 플래시 메모리와 같은 다수의 메모리 장치들이 통합된 형태를 가지는 통합 메모리 디바이스(convergence memory device)일 수 있다. 이러한 통합 메모리 디바이스는 각 메모리 장치들이 서로 다른 특성을 가지기 때문에 다양한 사용 모델(usage model)에 활용될 수 있다.
- [0045] 오늘날 대용량의 데이터를 처리하기 위하여 도 2 내지 도 5를 통하여 설명된 바와 같이 복수의 연산보드들(200) 및 복수의 메모리 보드들(400)을 필요로 한다. 나아가, 대용량의 데이터를 저장하기 위하여 복수의 메모리 보드들(400) 각각은 복수의 메모리 장치들(420)을 포함할 수 있다.
- [0046] 다만, 복수의 메모리 장치들 중 어느 하나에 결함이 발생된 경우, 결함이 발생된 메모리 장치(이하, 결함 메모리

리 장치)를 복구하기 위하여 전체 시스템이 중단될 수 있다. 뿐만 아니라, 결합 메모리 장치에 저장된 데이터는 보호받지 못할 수 있다. 본 발명은 복수의 메모리 장치들 중 결합 메모리 장치가 발생되더라도 전체 시스템의 가용성을 높게 유지하면서 결합 메모리 장치를 복구할 수 있는 메모리 시스템(700), 컴퓨팅 시스템(1000) 및 그들 각각의 동작방법을 제안한다.

- [0047] 도 6은 본 발명의 실시 예에 따른 메모리 장치(600)의 구조를 나타낸 도면이다. 도 6에 도시된 메모리 장치(600)는 도 5에 도시된 복수의 메모리 장치들(420) 각각과 대응될 수 있다.
- [0048] 메모리 장치(600)은 일반적으로 데이터가 저장되는 유저 영역(610)과 특별한 사정이 없는 한 사용되지 아니하는 오버-프로비저닝(over-provisioning) 영역(이하, OP 영역, 630)을 포함할 수 있다.
- [0049] OP 영역(630)은 메모리 장치의 수명을 증가시키기 위하여 구비될 수 있다. 메모리 장치(600)는 데이터가 저장되는 복수의 메모리 셀들을 포함할 수 있다. 나아가, 복수의 메모리 셀들 각각은 일정한 수명을 가질 수 있다. 예를 들면, 메모리 셀에 데이터가 10000번 라이트되면, 해당 메모리 셀의 수명은 종료될 수 있다. 따라서, 메모리 셀의 수명을 보존하기 위한 방법 중 하나로 메모리 셀에 데이터가 8000번 라이트되면, 해당 메모리 셀에 더 이상 라이트 동작이 수행되는 것을 막고, 리드 동작만이 해당 메모리 셀에 대하여 수행될 수 있다. 상기와 같은 방법으로 메모리 셀의 수명을 보존하기 위해서는 궁극적으로 복수의 메모리 셀들 각각은 가능한 고르게 사용되어야 한다. 즉, 특정 메모리 셀이 반복적으로 사용되는 경우, 해당 메모리 셀의 수명은 급속히 감소될 것이다. 다만, OS 동작을 수행하기 위한 데이터가 저장된 메모리 셀 혹은 자주 사용되는 데이터가 저장된 메모리 셀 같이 사용이 집중되는 메모리 셀이 존재하는 경우, 해당 메모리 셀의 과도한 사용을 방지하기 위하여 메모리 장치(600)의 여분 영역이 존재할 수 있다. 즉, 해당 메모리 셀의 과도한 사용을 방지하기 위하여 본래 해당 메모리 셀에 저장되어야 할 데이터가 여분 영역에 저장될 수 있다. 이러한 여분 영역이 OP 영역(630)이다.
- [0050] 도 7 내지 도 12을 통하여 설명되는 본 발명은 앞서 설명된 OP 영역(630)을 활용하여, 결합 메모리 장치 발생시, 무결정성을 유지하면서 결합 메모리 장치를 복구할 수 있는 메모리 시스템(700), 컴퓨팅 시스템(1000) 및 그들 각각의 동작방법을 제안한다.
- [0051] 도 7은 본 발명의 일 실시 예에 따른 메모리 시스템(700)의 구조를 나타낸 도면이다. 도 5를 참조하면, 메모리 시스템(700)은 메모리 보드(400)와 대응될 수 있다.
- [0052] 메모리 시스템(700)은 컨트롤러(710) 및 복수의 메모리 장치들(720)을 포함할 수 있다. 컨트롤러(710)는 도 5에 도시된 컨트롤러(410)와 대응될 수 있으며, 복수의 메모리 장치들(720)은 도 5에 도시된 복수의 메모리 장치들(420)과 대응될 수 있다. 도 7에 도시된 복수의 메모리 장치들(720)은 도 5에 도시된 복수의 메모리 장치(420)과는 다르게 동종인 복수의 메모리 장치들(720)로 도시된다. 다만, 이는 설명의 편의를 위한 것일 뿐이며, 이에 제한되는 것은 아니다.
- [0053] 컨트롤러(710)는 데이터 컨트롤러(730), I/O 인터페이스(750) 및 메모리 컨트롤러(760)를 포함할 수 있다. 도 5를 참조하면, 데이터 컨트롤러(730)는 도 5에 도시된 데이터 컨트롤러(510)와 대응되며, I/O 인터페이스(750)는 도 5에 도시된 I/O 인터페이스(530)와 대응될 수 있다. 그리고, 메모리 컨트롤러(760)는 도 5에 도시된 복수의 메모리 컨트롤러들(520A 내지 520C) 각각과 대응될 수 있다.
- [0054] 도 5를 참조하여, I/O 인터페이스(750)는 도 3의 상호접속 보드(300)을 통한 데이터 컨트롤러(730)와 연산 보드들(200) 사이에서의 인터페이싱을 지원할 수 있다. 또한, 메모리 컨트롤러(760)는 데이터 컨트롤러(730)와 복수의 메모리 장치들(720)의 사이에 위치하며, 이들 사이에서 각각의 인터페이싱을 지원할 수 있다. 앞서 설명된 바와 같이, 설명의 편의를 위하여 도 7에는 한 종류의 복수의 메모리 장치들(720)이 도시되어 있다. 따라서, 복수의 메모리 장치들(720) 각각의 인터페이싱을 지원할 수 있는 메모리 컨트롤러(760)는 역시 한 종류의 메모리 컨트롤러(760)로 컨트롤러(630)에 구비될 수 있다.
- [0055] 데이터 컨트롤러(730)는 검출부(733), 관리부(735), 선택부(737) 및 프로세서(739)를 포함할 수 있다.
- [0056] 검출부(733)는 복수의 메모리 장치들(720) 각각의 신뢰도를 모니터링할 수 있으며, 나아가 복수의 메모리 장치들(720) 중 결합 메모리 장치를 검출할 수 있다. 구체적으로, 데이터 컨트롤러(730)에 포함된 ECC부(미도시)는 복수의 메모리 장치들 각각으로부터 전달된 데이터에 대하여 ECC 보호 동작을 수행할 수 있으며, ECC 보호 동작을 통하여 전달된 데이터에 대한 에러 데이터를 판단할 수 있다. ECC부(미도시)는 판단된 에러 데이터에 대한 정보를 검출부(733)에 전달할 수 있으며, 검출부(733)는 복수의 메모리 장치에 대응하여 에러 데이터를 관리할 수 있다. 그리고, 검출부(733)는 에러 데이터의 수에 기초하여 복수의 메모리 장치들(720) 중 결합 메모리 장치를 검출할 수 있다. 결합 메모리 장치는 해당 메모리 장치에 대응하는 신뢰도가 사전 설정된 임계 값보다 낮

아진 상태로 판단되는 메모리 장치를 의미한다. 예를 들어, 신뢰도는 해당 메모리 장치에서 발생된 에러 데이터 수에 기초하여 판단될 수 있으며, 특정 메모리 장치에서 에러 데이터의 수가 사전 설정된 임계 값보다 높을 경우, 검출부(733)는 해당 메모리 장치를 결함 메모리로 판단할 수 있다. 다만, 이는 하나의 실시 예에 불과하며, 이에 제한되는 것은 아니다.

[0057] 나아가, 검출부(733)는 복수의 메모리 장치(620) 각각에 대응하는 신뢰도를 저장할 수 있다. 예를 들어, 제 1 메모리 장치(723)가 결함 메모리 장치로 검출된 경우, 검출부(733)는 제 1 메모리 장치(723)에 대응하는 신뢰도를 업데이트할 수 있다. 반면에, 검출부(733)는 정상 메모리 장치인 제 2 메모리 장치(725)에 대응하는 신뢰도는 별도로 업데이트하지 않을 수 있다. 추후에, 제 1 메모리 장치(723)가 정상 메모리 장치로 복구된 경우, 검출부(733)는 제 1 메모리 장치(723)에 대응하는 신뢰도를 다시 업데이트할 수 있다. 예를 들면, 검출부(733)는 결함 메모리 장치에 대하여 '1', 정상 메모리 장치에 대하여 '0'으로 표시하여 복수의 메모리 장치(620) 각각의 신뢰도를 관리할 수 있다. 이는 일 실시 예에 해당할 뿐이며, 이에 제한되는 것은 아니다.

[0058] 관리부(735)는 복수의 메모리 장치들(720) 각각의 OP 영역을 관리할 수 있다. 구체적으로, 관리부(735)는 복수의 메모리 장치들(720) 각각의 OP 영역 중 가용(available) OP 영역을 모니터링할 수 있으며, 복수의 메모리 장치들(720) 각각의 OP 영역에 대한 가용 정보(이하, 가용 OP 정보)를 저장할 수 있다. 예를 들어, 관리부(735)는 복수의 메모리 장치들(720) 각각의 가용 OP 정보를 '1', 불가용 OP 영역에 대응하는 가용 OP 정보를 '0'으로 표시하여, 가용 OP 정보를 해당 메모리 장치와 대응하도록 저장할 수 있다. 이는 일 실시 예에 해당할 뿐이며, 이에 제한되는 것은 아니다.

[0059] 또한, 관리부(735)는 복수의 메모리 장치들(720) 각각의 메모리 맵을 저장할 수 있다. 따라서, 복수의 메모리 장치들(720)간에 데이터 이동이 발생된 경우, 관리부(735)는 데이터 이동을 반영하기 위하여 메모리 맵을 업데이트할 수 있다.

[0060] 선택부(737)은 가용 OP 정보에 기초하여 결함 메모리 장치에 저장된 데이터(이하, 타겟 데이터)가 저장될 수 있는 OP 영역을 가진 메모리 장치(이하, 가용 메모리 장치)를 메모리 시스템(700) 내에서 검색할 수 있다. 나아가, 선택부(737)는 사전 설정된 기준에 따라 가용 메모리 장치를 선택할 수 있다. 만약, 메모리 시스템(700) 내에 가용 메모리 장치가 존재하며, 복수의 메모리 장치들(720) 각각에 순번에 존재한다면, 선택부(737)는 결함 메모리 장치의 후 순번을 갖는 메모리 장치 중 가용 메모리 장치를 선택할 수 있다. 예를 들면, 제 1 메모리 장치(723)가 결함 메모리 장치이고, 제 2 메모리 장치(725)가 가용 메모리 장치이면서 제 1 메모리 장치(723)의 후 순번을 갖는 메모리 장치라면, 선택부(737)는 가용 메모리 장치로 제 2 메모리 장치(725)를 선택할 수 있다. 또 다른 예로, 만약, 결함 메모리 장치에 저장된 데이터의 크기가 가용 메모리 장치의 OP 영역보다 큰 경우, 선택부(737)는 복수의 가용 메모리 장치들을 선택할 수 있다. 이는 일 실시 예에 해당할 뿐이며, 이에 제한되는 것은 아니다.

[0061] 반면에, 메모리 시스템(700) 내에 가용 메모리 장치가 존재하지 않는다면, 타겟 데이터를 다른 메모리 시스템으로 전달할 수 있다. 이에 대한 동작은 도 11에서 자세히 설명된다.

[0062] 프로세서(739)는 타겟 데이터를 결함 메모리 장치에서 선택부(737)에 의해 선택된 가용 메모리 장치의 OP 영역으로 이동시킬 수 있다. 구체적으로, 프로세서(739)는 타겟 데이터를 결함 메모리 장치로부터 리드할 수 있다. 리드된 타겟 데이터는 도면에 도시되진 아니하였으나, 데이터 컨트롤러(710)의 내부 메모리에 임시로 저장될 수 있다. 그리고, 프로세서(739)는 가용 메모리 장치의 OP 영역에 타겟 데이터를 저장할 수 있다. 타겟 데이터가 가용 메모리 장치의 OP 영역으로 이동된 후에, 관리부(735)는 타겟 데이터에 대응하는 주소 정보를 반영하기 위하여 메모리 맵을 업데이트할 수 있다. 나아가, 추후 타겟 데이터에 대한 리드 요청이 입력된 경우, 프로세서(739)는 메모리 맵에 기초하여 OP 영역에 저장된 타겟 데이터를 리드할 수 있다.

[0063] 그리고 나서, 결함 메모리 장치가 정상 메모리 장치(이하, 복구 메모리 장치)로 복구된 경우, 프로세서(739)는 OP 영역에 저장된 타겟 데이터를 리드하여 복구 메모리 장치에 타겟 데이터를 저장할 수 있다. 나아가, 검출부(733)는 복구 메모리 장치에 대한 신뢰도를 업데이트할 수 있으며, 관리부(735)는 타겟 데이터에 대응하는 주소 정보를 반영하기 위하여 메모리 맵을 업데이트할 수 있다.

[0064] 도 8은 본 발명의 일 실시 예에 따른 데이터 컨트롤러(710)의 동작을 나타낸 흐름도이다.

[0065] 먼저, 단계 S801에서, 검출부(733)은 복수의 메모리 장치들(720) 각각의 신뢰도를 모니터링할 수 있다.

[0066] 단계 S803에서, 검출부(733)는 모니터링된 신뢰도에 기초하여 복수의 메모리 장치들(720) 각각에 대응하는 신뢰

도가 사전 설정된 임계 값보다 크거나 같은 메모리 장치를 결합 메모리 장치로써 검출할 수 있다.

- [0067] 만약, 결합 메모리 장치가 검출되지 않은 경우(단계 S803에서, 'N'), 검출부(733)는 단계 S801에서 신뢰도를 계속 모니터링할 수 있다.
- [0068] 반면에, 결합 메모리 장치가 검출된 경우(단계 S803에서, 'Y'), 단계 S805에서, 프로세서(739)는 결합 메모리 장치에 저장된 타겟 데이터를 리드할 수 있다. 나아가, 프로세서(739)는 타겟 데이터를 데이터 컨트롤러(710) 내부 메모리에 임시로 저장할 수 있다.
- [0069] 그 후, 단계 S807에서, 선택부(737)는 관리부(735)에 저장된 복수의 메모리 장치들(720) 각각의 가용 OP 영역 정보에 기초하여 메모리 시스템(700) 내에서 가용 메모리 장치를 검색할 수 있다.
- [0070] 만약, 메모리 시스템(700) 내에 가용 메모리 장치가 존재하지 않는다면(단계 S807에서, 'N'), 단계 S809에서, 도 10에서 설명될 동작이 수행될 수 있다. 이에 대한 동작은 도 10을 통하여 설명된다.
- [0071] 반면에, 메모리 시스템(700) 내에 가용 메모리 장치가 존재한다면, 단계 S811에서, 선택부(737)는 사전 설정된 기준에 따라 메모리 장치를 선택할 수 있다.
- [0072] 그리고, 단계 S813에서, 프로세서(739)는 가용 메모리 장치의 OP 영역에 타겟 데이터를 저장할 수 있다.
- [0073] 뿐만 아니라, 단계 S815에서, 관리부(735)는 타겟 데이터에 대응하는 주소 정보를 업데이트할 수 있다.
- [0074] 도 9는 본 발명의 일 실시 예에 따른 데이터 컨트롤러(710)의 동작을 나타내는 흐름도이다. 도 9에 도시된 데이터 컨트롤러(710)의 동작은 도 8에서 설명된 데이터 컨트롤러(710)의 동작의 후속으로 진행될 수 있다.
- [0075] 단계 S901에서, 결합 메모리 장치는 복구 메모리 장치로 복구될 수 있다.
- [0076] 만약, 결합 메모리 장치가 복구되지 않았다면(단계 S901에서, 'N'), 단계 S903에서, 타겟 데이터에 대응하는 액세스 요청은 도 8에서 설명된 OP 영역에서 수행될 수 있다. 구체적으로, 타겟 데이터에 대응하는 리드 요청에 대하여 프로세서(739)는 타겟 데이터를 리드하도록 타겟 데이터가 저장된 가용 메모리 장치를 제어할 수 있다.
- [0077] 반면에, 결합 메모리 장치가 복구되었다면(단계 S901에서, 'Y'), 단계 S905에서, 관리부(735)는 복구 메모리 장치에 대응하는 신뢰도를 업데이트할 수 있다.
- [0078] 그리고, 단계 S907에서, 프로세서(739)는 현재 타겟 데이터가 저장된 OP 영역에서 타겟 데이터를 리드할 수 있다. 타겟 데이터는 데이터 컨트롤러(710) 내부의 메모리에 임시로 저장될 수 있다.
- [0079] 나아가, 단계 S909에서, 프로세서(739)는 타겟 데이터를 복구 메모리 장치에 저장할 수 있다.
- [0080] 마지막으로, 단계 S911에서, 관리부(735)는 타겟 데이터에 대응하는 주소 정보를 업데이트 할 수 있다. 구체적으로 관리부(735)는 복구 메모리 장치에 저장된 타겟 데이터의 주소 정보를 반영하기 위하여 메모리 맵을 업데이트할 수 있다.
- [0081] 도 7 내지 도 9에 설명된 본 발명은 복수의 메모리 장치들(720) 중 결합 메모리 장치가 발생되더라도 상기의 과정을 통하여 메모리 시스템(700)의 가용성을 높게 유지하면서 결합 메모리 장치를 복구할 수 있다.
- [0082] 앞서, 도 7 내지 도 9에서, 단일 메모리 시스템(700)에 결합 메모리 장치가 발생된 경우, 시스템의 가용성을 높게 유지하면서 결합 메모리 장치를 복구하는 본 발명의 동작 과정이 설명되었다. 다만, 단일 메모리 시스템(700)에 결합 메모리 장치에 저장된 데이터를 임시로 저장할 수 있는 가용 OP 영역이 존재하지 않을 때, 시스템의 가용성을 높게 유지하면서 결합 메모리 장치를 복구할 수 없는 문제가 발생할 수 있다. 이하에서는, 도 10 내지 도 12에서, 이러한 문제점을 해결할 수 있는 컴퓨팅 시스템(1000)이 설명된다.
- [0083] 도 10은 본 발명의 일 실시 예에 따른 컴퓨팅 시스템(1000)의 구조를 나타내는 도면이다.
- [0084] 컴퓨팅 시스템(1000)은 복수의 메모리 시스템(700A 및 700B) 및 메모리 시스템 관리부(이하, MMU, 1010)를 포함할 수 있다. 도 10에 컴퓨팅 시스템(1000)은 제 1 메모리 시스템(700A) 및 제 2 메모리 시스템(700B)만을 포함하고 있으나, 더 많은 메모리 시스템들을 포함할 수 있음은 당연하다.
- [0085] 복수의 메모리 시스템들(700A 및 700B) 각각은 도 7에 설명된 메모리 시스템(700)과 대응될 수 있다. 나아가, 메모리 시스템들(700A 및 700B) 내부에 포함된 컨트롤러(710A 및 710B) 및 복수의 메모리 장치들(720A 및 720B) 각각은 도 7에 도시된 컨트롤러(710) 및 복수의 메모리 장치들(720)과 각각 대응될 수 있다.

- [0086] MMU(1010)는 도 2 내지 도 4에서 설명된 연산보드(200) 및 상호접속보드(300)에 대응될 수 있다. 따라서, MMU(1010)는 복수의 메모리 시스템들(700A 및 700B) 각각을 관리할 수 있으며, 복수의 메모리 시스템들(700A 및 700B) 각각과 데이터 통신할 수 있다. 즉, MMU(1010)는 제 1 메모리 시스템(700A)으로부터 전달받은 데이터를 제 2 메모리 시스템(700B)으로 전달할 수 있다. MMU(1010)가 반대 방향으로 데이터를 전달할 수 있음은 당연하다.
- [0087] MMU(1010)는 복수의 메모리 시스템들(700A 및 700B) 각각에 포함된 복수의 메모리 장치들(720A 및 720B) 각각의 가용 OP 정보를 데이터 컨트롤러(730A 및 730B)로부터 I/O 인터페이스(750A 및 750B)를 통하여 전달받을 수 있다. 따라서, MMU(1010)는 모든 메모리 장치들 각각의 가용 OP 영역에 대한 정보를 알 수 있다. 나아가, 가용 OP 정보의 업데이트가 발생된 경우, MMU(1010)은 데이터 컨트롤러(730A 및 730B)로부터 업데이트된 정보를 전달받아 MMU(1010)에 저장된 가용 OP 정보를 업데이트할 수 있다.
- [0088] 또한, MMU(1010)는 복수의 메모리 장치들(720A 및 720B) 각각의 메모리 맵을 데이터 컨트롤러(730A 및 730B)로부터 I/O 인터페이스(750A 및 750B)를 통하여 전달받을 수 있다. 또한, MMU(1010)는 상기 메모리 맵들이 모두 반영된 글로벌 맵(global map)을 저장할 수 있다. 그리고, 데이터 이동이 발생된 경우, MMU(1010)은 데이터 컨트롤러(730A 및 730B)로부터 업데이트된 주소 정보를 전달받아 MMU(1010)에 저장된 각각의 메모리 맵들 및/또는 글로벌 맵을 업데이트할 수 있다.
- [0089] 설명의 편의를 위하여, 제 1 메모리 시스템(700A)에 포함된 제 1 복수의 메모리 장치들(720A) 중 결합 메모리 장치가 발생되었으며, 제 1 복수의 메모리 장치들(720A) 중 가용 메모리 장치가 존재하지 않는다고 가정한다. 또한, 제 2 복수의 메모리 장치들(720B) 중 가용 메모리 장치가 존재한다고 가정한다.
- [0090] 제 1 데이터 컨트롤러(730A)는 결합 메모리 장치에 저장된 타겟 데이터를 리드할 수 있다. 제 1 데이터 컨트롤러(730A)는 타겟 데이터를 임시로 저장할 가용 메모리 장치를 제 1 복수의 메모리 장치들(720A)에서 검색할 수 있다. 다만, 제 1 복수의 메모리 장치들(720A)에 가용 메모리 장치가 존재하지 않는다면, 제 1 데이터 컨트롤러(730A)는 제 1 I/O 인터페이스(750A)를 통하여 타겟 데이터를 MMU(1010)에 전달할 수 있다.
- [0091] MMU(1010)는 복수의 메모리 시스템들(700A 및 700B) 각각에게 전달받은 가용 OP 정보에 기초하여 제 1 메모리 시스템(700A)으로부터 전달받은 타겟 데이터를 임시로 저장할 수 있는 가용 메모리 장치가 포함된 제 2 메모리 시스템(700B)을 검색할 수 있다. 나아가, MMU(1010)는 타겟 데이터를 제 2 메모리 시스템(700B)에 전달할 수 있다.
- [0092] 제 2 데이터 컨트롤러(730B)는 제 2 I/O 인터페이스(750B)를 통하여 타겟 데이터를 전달받을 수 있다. 제 2 데이터 컨트롤러(730B)는 타겟 데이터를 임시로 저장할 가용 메모리 장치를 제 2 복수의 메모리 장치들(720B)에서 검색할 수 있다. 나아가, 제 2 데이터 컨트롤러(730B)는 타겟 데이터를 검색된 가용 메모리 장치의 OP 영역에 저장할 수 있다.
- [0093] 이때, 제 1 데이터 컨트롤러(730A) 및 제 2 데이터 컨트롤러(730B)는 타겟 데이터에 대응하는 주소 정보를 각각 업데이트할 수 있다. 구체적으로, 제 1 데이터 컨트롤러(730A)는 타겟 데이터가 제 1 복수의 메모리 장치들(720A)에 저장되어 있지 않다는 것을 나타낼 수 있도록, 제 2 데이터 컨트롤러(730B)는 타겟 데이터가 상기 OP 영역에 저장되어 있다는 것을 나타내도록 타겟 데이터에 대응하는 주소 정보를 각각 업데이트할 수 있다. 나아가, 제 1 데이터 컨트롤러(730A) 및 제 2 데이터 컨트롤러(730B)는 업데이트된 주소 정보를 각각 MMU(1010)에 전달할 수 있다. MMU(1010)는 전달받은 타겟 데이터의 각각의 주소 정보를 반영하기 위하여 제 1 메모리 시스템(700A) 및 제 2 메모리 시스템(700B)의 메모리 맵들 및/또는 글로벌 맵을 각각 업데이트할 수 있다.
- [0094] 그리고 나서, 제 1 복수의 메모리 장치(720A)에 포함된 결합 메모리 장치가 복구된 경우, 앞서 설명된 동작 과정의 역방향으로 타겟 데이터를 복구 메모리 장치에 저장하기 위한 동작이 진행될 수 있다.
- [0095] 제 1 데이터 컨트롤러(730A)는 결합 메모리 장치가 복구 메모리 장치로 복구된 정보를 MMU(1010)에 전달할 수 있다. 이때, MMU(1010)는 복구 메모리 장치의 신뢰도를 업데이트할 수 있다. 그리고 MMU(1010)는 제 2 메모리 시스템(700B)에게 타겟 데이터에 대한 리드 요청을 이슈할 수 있다. 제 2 데이터 컨트롤러(730B)는 타겟 데이터를 리드하고, MMU(1010)에 전달할 수 있다. MMU(1010)는 전달받은 타겟 데이터를 제 1 메모리 시스템(700A)에 전달할 수 있다. 제 1 데이터 컨트롤러(730A)는 타겟 데이터를 MMU(1010)로부터 전달받을 수 있으며, 타겟 데이터를 복구 메모리 장치에 저장할 수 있다.
- [0096] 이때, 제 1 데이터 컨트롤러(730A) 및 제 2 데이터 컨트롤러(730B)는 타겟 데이터에 대응하는 주소 정보를 각각

업데이트할 수 있다. 구체적으로, 제 2 데이터 컨트롤러(730B)는 타겟 데이터가 제 2 복수의 메모리 장치들(720B)에 저장되어 있지 않다는 것을 나타낼 수 있도록, 제 1 데이터 컨트롤러(730A)는 타겟 데이터가 복구 메모리 장치에 저장되어 있다는 것을 나타내도록 타겟 데이터에 대응하는 주소 정보를 각각 업데이트할 수 있다. 나아가, 제 1 데이터 컨트롤러(730A) 및 제 2 데이터 컨트롤러(730B)는 업데이트된 주소 정보를 각각 MMU(1010)에 전달할 수 있다. MMU(1010)는 전달받은 타겟 데이터의 각각의 주소 정보를 반영하기 위하여 제 1 메모리 시스템(700A) 및 제 2 메모리 시스템(700B)의 메모리 맵들 및/또는 글로벌 맵을 각각 업데이트할 수 있다.

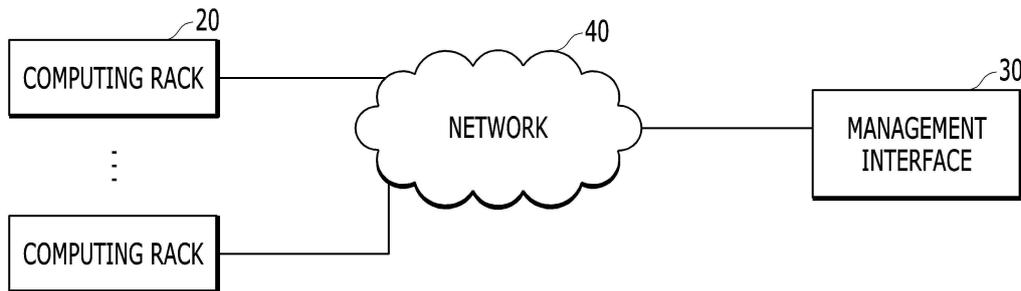
- [0097] 도 11은 본 발명의 일 실시 예에 따른 컴퓨팅 시스템(1000)의 동작을 나타내는 흐름도이다. 나아가, 도 11에 도시된 동작들은 도 8에 도시된 단계 S809에 대응될 수 있다. 즉, 단계 S1101 내지 단계 S1111의 동작들은 도 8에 도시된 단계 S807의 후속 동작일 수 있다.
- [0098] 단계 S1101에서, 제 1 메모리 시스템(700A)은 도 8에 도시된 단계 S801 내지 단계 S807에 대응하는 동작을 수행할 수 있다.
- [0099] 단계 S1103에서, 제 1 메모리 시스템(700A)은 MMU(1010)에 타겟 데이터를 전달할 수 있다. 구체적으로 제 1 메모리 시스템(700A)에 포함된 제 1 데이터 컨트롤러(730A)가 제 1 I/O 인터페이스(750A)를 통하여 타겟 데이터를 MMU(1010)으로 전달할 수 있다.
- [0100] 단계 S1105에서, MMU(1010)는 제 1 메모리 시스템(700A) 외의 메모리 시스템 중 가용 메모리 장치를 포함한 메모리 시스템을 검색할 수 있다.
- [0101] 만약, 다른 메모리 시스템에도 가용 메모리 장치가 존재하지 않는다면(단계 S1105에서, 'N'), MMU(1010)는, 단계 S1105에서, 다시 가용 메모리 장치를 포함하는 메모리 시스템을 검색할 수 있다.
- [0102] 반면에, 특정 메모리 시스템(이하, 제 2 메모리 시스템, 700B)에 가용 메모리 장치가 존재하는 경우(단계 S1107에서, 'Y'), 단계 S1109에서, MMU(1010)는 타겟 데이터를 제 2 메모리 시스템(700B)으로 전달할 수 있다.
- [0103] 나아가, 단계 S1111에서, 제 2 데이터 컨트롤러(730B)는 타겟 데이터를 가용 메모리 장치의 OP 영역에 저장할 수 있다.
- [0104] 그 후, 단계 S1113에서, 도 10에서 설명된 바와 같이, 제 1 데이터 컨트롤러(730A), 제 2 데이터 컨트롤러(730B) 및 MMU(1010)은 타겟 데이터에 대응하는 주소 정보를 업데이트할 수 있다.
- [0105] 도 12는 본 발명의 일 실시 예에 따른 컴퓨팅 시스템(1000)의 동작을 나타낸 흐름도이다. 도 12에 도시된 동작들은 도 11에 도시된 동작 이후에 수행될 수 있다. 구체적으로, 도 11에서 설명된 바와 같이, 제 1 메모리 시스템(700A)에서 발생한 결함 메모리 장치가 복구된 때, 컴퓨팅 시스템(1000)이 타겟 데이터를 복구 메모리 장치에 저장하는 동작이 도 12를 통하여 설명된다. 반면에, 도면이 도시되지 아니하였으나, 결함 메모리 장치가 복구되지 않은 경우, 타겟 데이터에 대한 액세스 요청은 현재 타겟 데이터가 저장된 제 2 메모리 시스템(700B)의 OP 영역에 대하여 수행될 수 있다. 예를 들면, 타겟 데이터에 대한 리드 요청이 이슈된 경우, 제 2 데이터 컨트롤러(730B)는 상기 OP 영역에 저장된 타겟 데이터를 리드할 수 있다.
- [0106] 단계 S1201에서, 결함 메모리 장치는 복구 메모리 장치로써 복구될 수 있다.
- [0107] 단계 S1203에서, 제 1 데이터 컨트롤러(730A)는 복구 메모리 장치에 대응하는 신뢰도를 업데이트할 수 있다. 예를 들면, 제 1 데이터 컨트롤러(730A)는 결함 메모리 장치에 대하여 '1'로 표시된 신뢰도를 결함 메모리 장치를 복구 메모리 장치로 복구 완료된 후에 복구 메모리 장치에 대하여 '0'으로 업데이트할 수 있다.
- [0108] 단계 S1205에서, 제 1 메모리 시스템(700A)은 MMU(1010)로 상기 업데이트된 신뢰도를 전달할 수 있다.
- [0109] 단계 S1207에서, MMU(1010)는 제 1 메모리 시스템(700A)으로부터 전달받은 신뢰도에 기초하여 제 1 메모리 시스템(700A)에 대응하는 신뢰도를 업데이트할 수 있다.
- [0110] 그 후, 타겟 데이터를 본래 저장 위치였던 제 1 메모리 시스템(700A)으로 이동시키기 위하여, 단계 S1209에서, MMU(1010)는 제 2 메모리 시스템(700B)에 타겟 데이터를 리드하도록 요청할 수 있다.
- [0111] 단계 S1211에서, 제 2 메모리 시스템(700B)은 상기 리드 요청에 따라 타겟 데이터를 리드할 수 있다. 구체적으로, 제 2 데이터 컨트롤러(730B)는 타겟 데이터를 리드하도록 타겟 데이터가 저장된 OP 영역을 갖는 메모리 장치를 제어할 수 있다.
- [0112] 그리고, 단계 S1213에서, 제 2 메모리 시스템(700B)은 타겟 데이터를 MMU(1213)로 출력할 수 있다.

- [0113] 단계 S1215에서, MMU(1010)는 제 1 메모리 시스템(700A)으로 타겟 데이터를 전달할 수 있다.
- [0114] 단계 S1217에서, 제 1 메모리 시스템(700A)은 MMU(1010)로부터 전달받은 타겟 데이터를 복구 메모리 장치에 저장할 수 있다. 구체적으로, 제 1 데이터 컨트롤러(730A)는 타겟 데이터를 라이트하도록 복구 메모리 장치를 제어할 수 있다.
- [0115] 그 후, 단계 S1219에서, 도 10에서 설명된 바와 같이, 제 1 데이터 컨트롤러(730A), 제 2 데이터 컨트롤러(730B) 및 MMU(1010)은 타겟 데이터에 대응하는 주소 정보를 업데이트할 수 있다.
- [0116] 도 10 내지 도 12에 설명된 바와 같이, 복수의 메모리 시스템으로 구성된 컴퓨팅 시스템은 단일 메모리 시스템(700)의 문제점을 극복할 수 있으며, 그 결과 전체적인 시스템의 가용성을 높게 유지하여 결함 메모리 장치를 복구할 수 있다.
- [0117] 한편, 본 발명의 상세한 설명에서는 구체적인 실시 예에 관해 설명하였으나, 본 발명의 범위에서 벗어나지 않는 한도 내에서 여러 가지 변형이 가능함은 물론이다. 그러므로, 본 발명의 범위는 설명된 실시 예에 국한되어 정해져서는 안되며 후술하는 특허청구의 범위뿐만 아니라 이 특허청구의 범위와 균등한 것들에 의해 정해져야 한다.

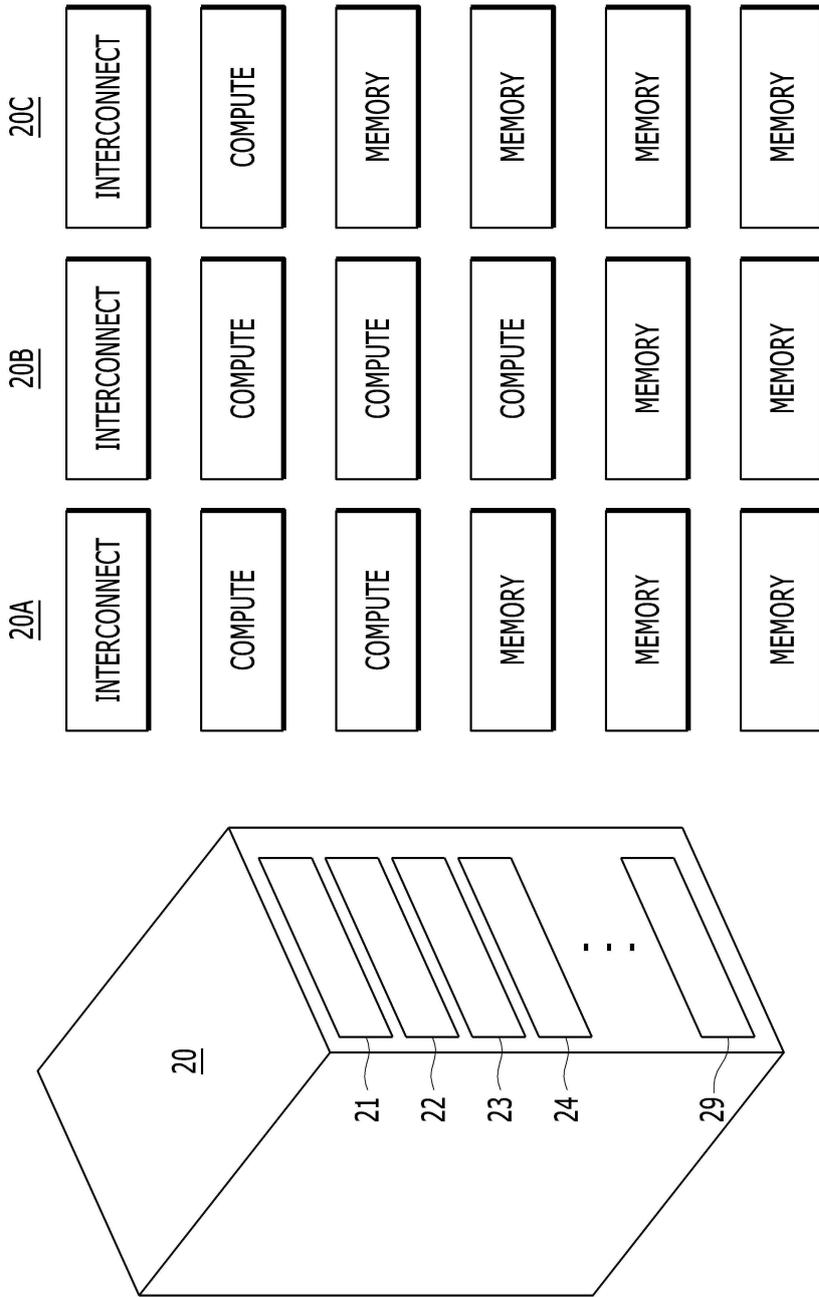
도면

도면1

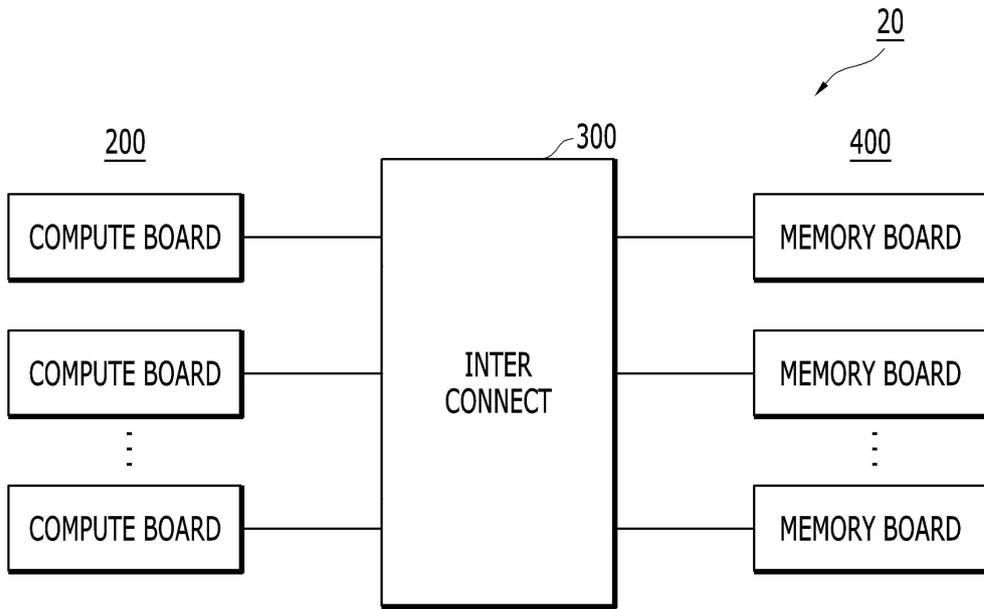
10



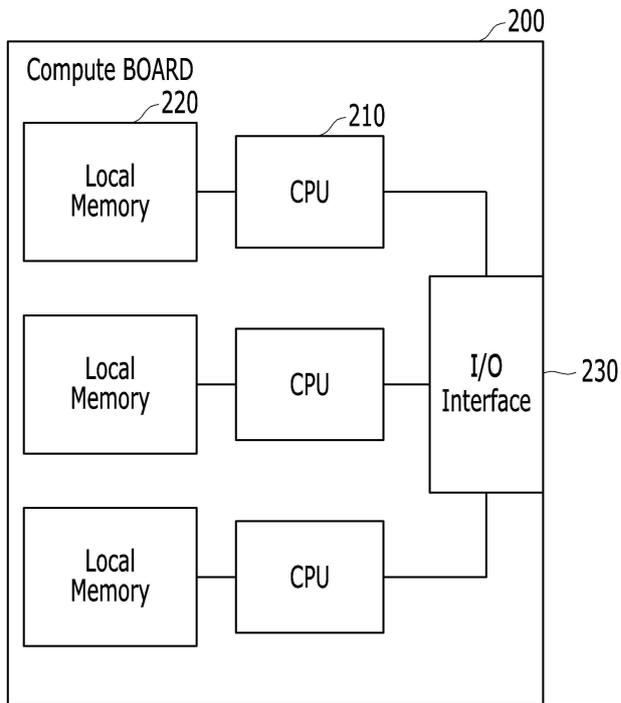
도면2



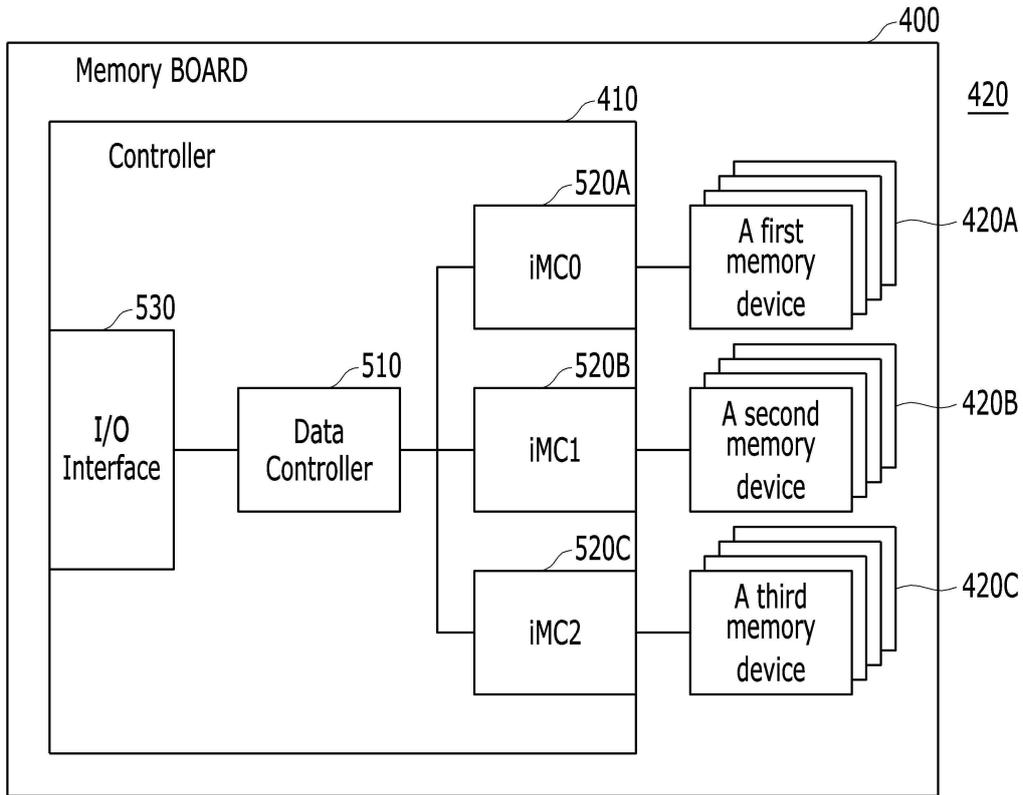
도면3



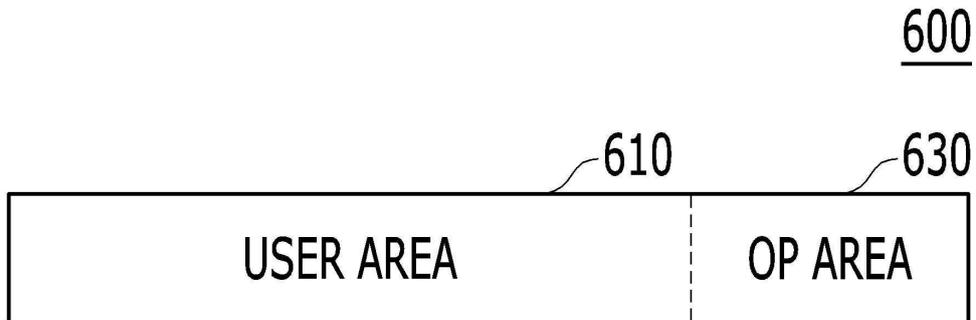
도면4



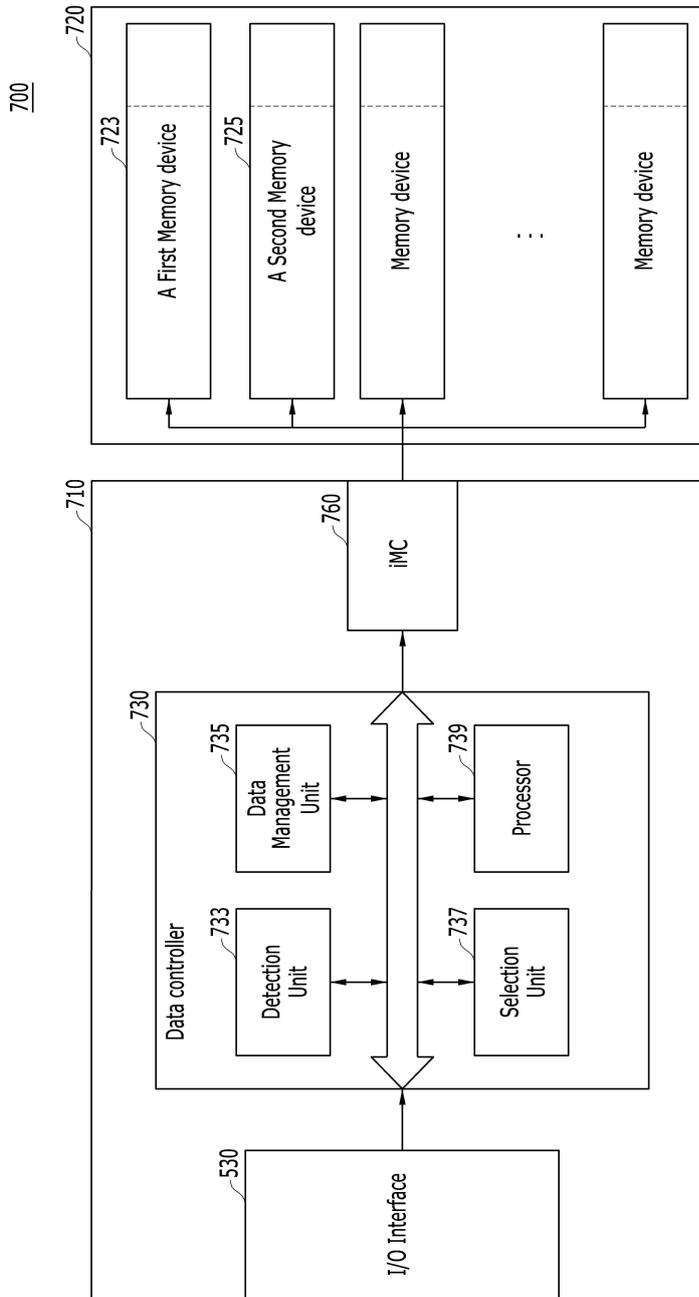
도면5



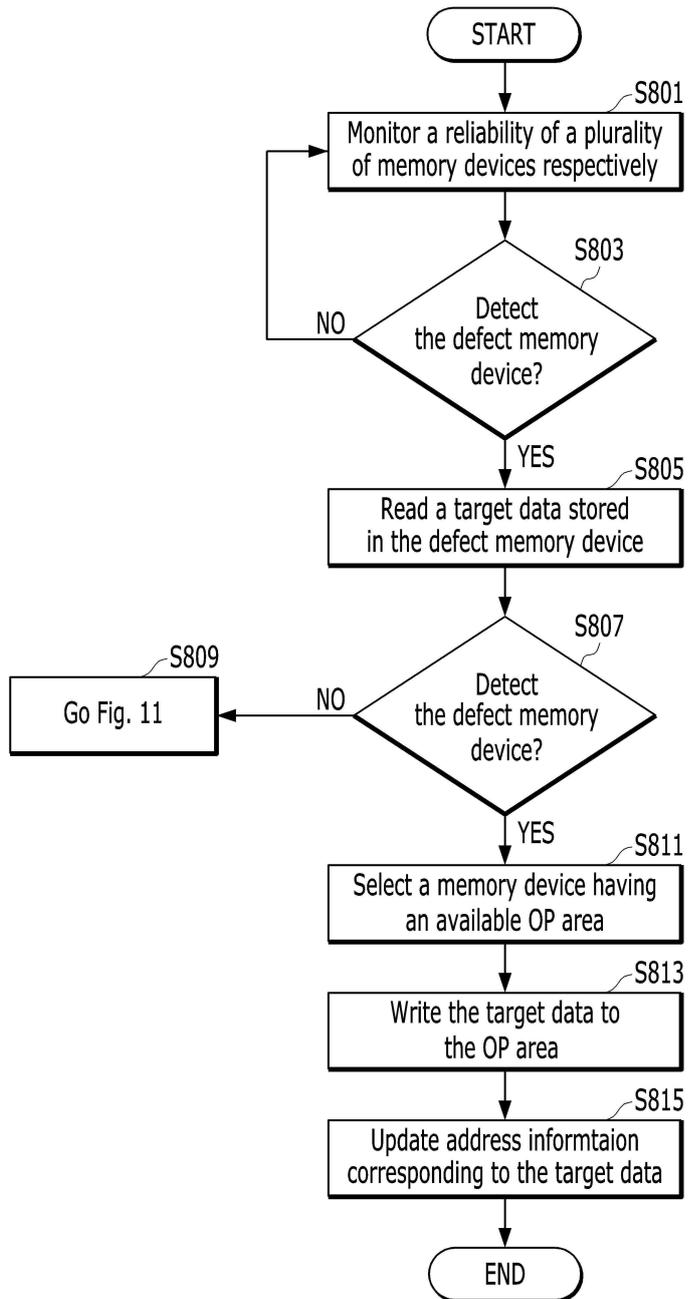
도면6



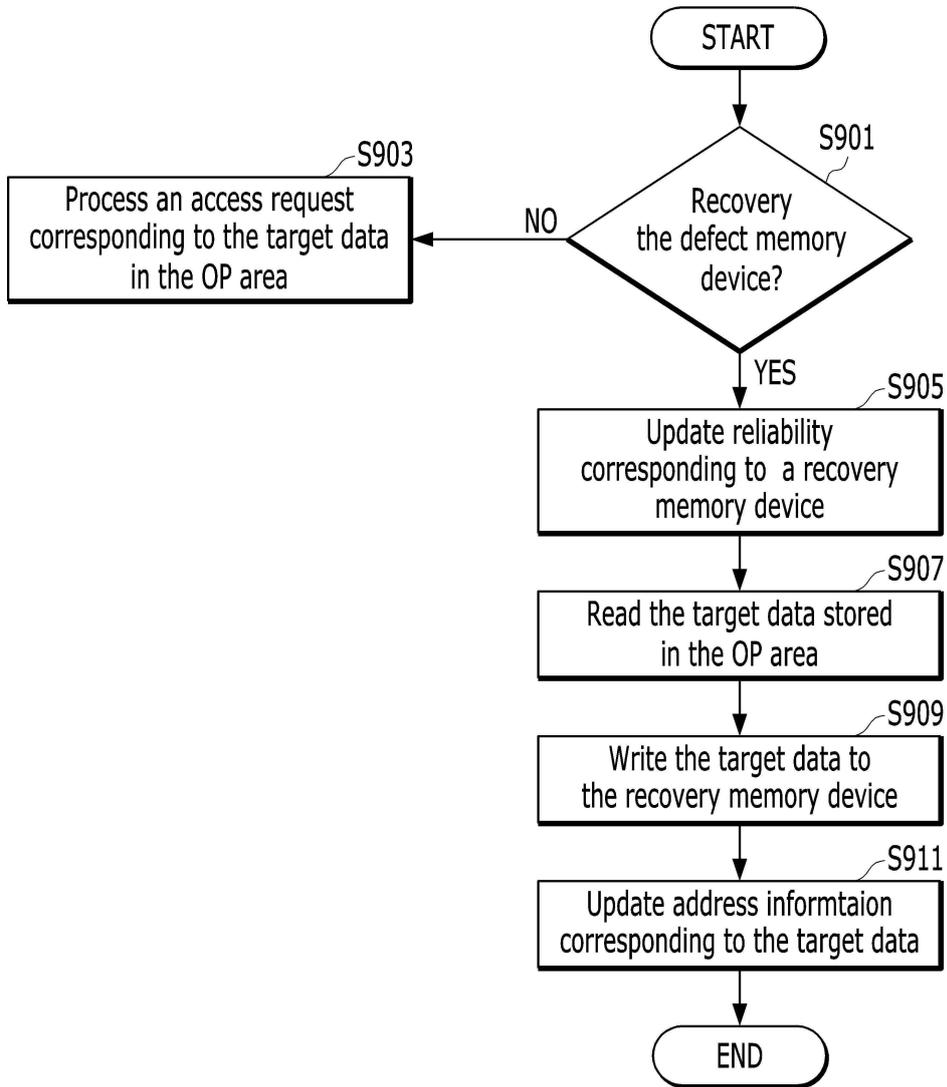
도면7



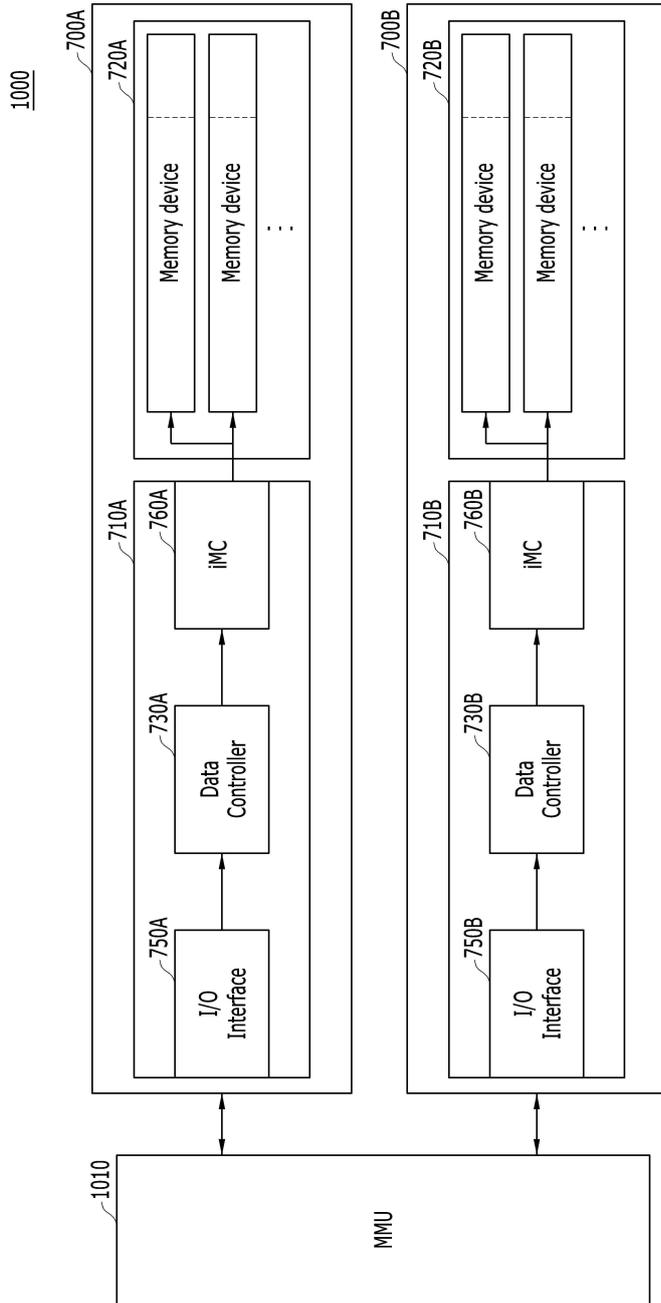
도면8



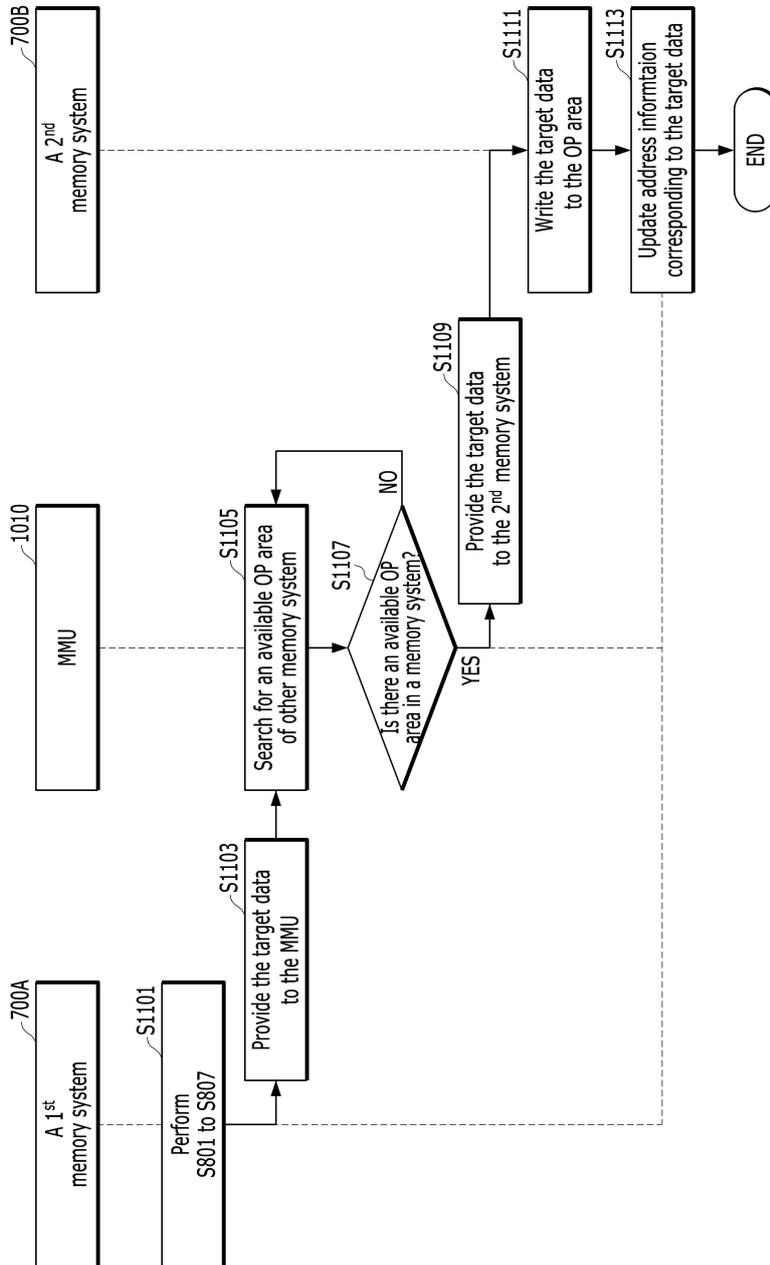
도면9



도면10



도면11



도면12

