(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0167986
(43) 공개일자 2023년12월12일

(51) 국제특허분류(Int. Cl.)
*G06F 9/38* (2006.01) *G06F 12/06* (2006.01)
*G06F 17/16* (2006.01) *G06F 9/30* (2018.01)
(52) CPC특허분류
*G06F 9/3885* (2013.01)
*G06F 12/0607* (2013.01)
(21) 출원번호 10-2022-0068322
(22) 출원일자 2022년06월03일
심사청구일자 2022년06월03일

(71) 출원인
**연세대학교 산학협력단**
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
**정의영**
서울특별시 강남구 도곡로43길 20, 203동 604호
**김병진**
경기도 고양시 덕양구 화정로 27, 609동 802호
(74) 대리인
**특허법인(유한)아이시스**

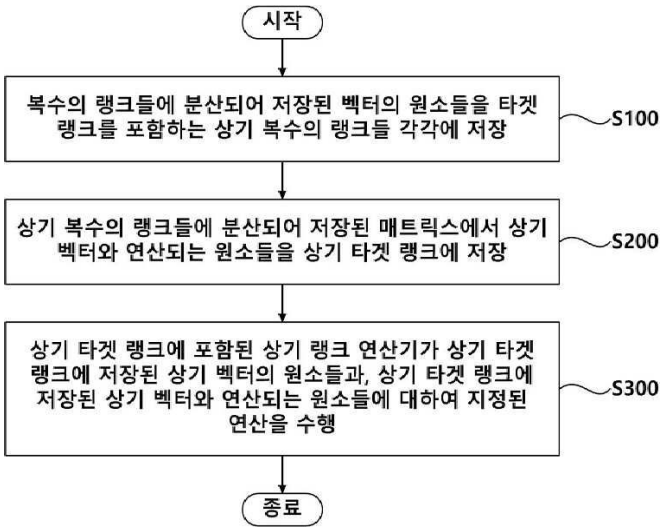
전체 청구항 수 : 총 14 항

(54) 발명의 명칭 **병렬 연산 방법, 브로드캐스터 및 분할 전송부**

## (57) 요 약

본 실시예는 데이터를 저장하는 복수개의 메모리 랭크(rank)와 상기 메모리 랭크 각각에 포함된 랭크 연산기(rank processor)를 포함하는 메모리에서 수행되는 병렬 연산 방법으로, 복수의 랭크들에 분산되어 저장된 벡터의 원소들을 타겟 랭크를 포함하는 상기 복수의 랭크들 각각에 저장하는 단계와, 상기 복수의 랭크들에 분산되어 저장된 매트릭스에서 상기 벡터와 연산되는 원소들을 상기 타겟 랭크에 저장하는 단계와, 상기 타겟 랭크에 포함된 상기 랭크 연산기가 상기 타겟 랭크에 저장된 상기 벡터의 원소들과, 상기 타겟 랭크에 저장된 상기 벡터와 연산되는 원소들에 대하여 지정된 연산을 수행하는 단계를 포함한다.

*대 표 도* - 도1



(52) CPC특허분류
*G06F 17/16* (2013.01)
*G06F 9/3004* (2013.01)

# 명 세 서

## 청구범위

### 청구항 1

데이터를 저장하는 복수개의 메모리 랭크(rank)와 상기 메모리 랭크 각각에 포함된 랭크 연산기(rank processor)를 포함하는 메모리에서 수행되는 병렬 연산 방법으로,

복수의 랭크들에 분산되어 저장된 벡터의 원소들을 타겟 랭크를 포함하는 상기 복수의 랭크들 각각에 저장하는 단계와,

상기 복수의 랭크들에 분산되어 저장된 매트릭스에서 상기 벡터와 연산되는 원소들을 상기 타겟 랭크에 저장하는 단계와,

상기 타겟 랭크에 포함된 상기 랭크 연산기가 상기 타겟 랭크에 저장된 상기 벡터의 원소들과, 상기 타겟 랭크에 저장된 상기 벡터와 연산되는 원소들에 대하여 지정된 연산을 수행하는 단계를 포함하는 병렬 연산 방법.

### 청구항 2

제1항에 있어서,

상기 벡터의 원소들을 상기 복수의 랭크들에 저장하는 단계는,

상기 복수의 랭크들에 분산되어 저장된 벡터의 원소들을 결합하여 상기 복수의 랭크들에 저장하여 수행하는 병렬 연산 방법.

### 청구항 3

제2항에 있어서,

결합되어 상기 복수의 랭크들에 저장된 상기 벡터의 원소들은 상기 매트릭스와 연산 가능하도록 결합되어 저장되는 병렬 연산 방법.

### 청구항 4

제2항에 있어서,

결합되어 상기 복수의 랭크들에 저장된 상기 벡터의 원소들은,

서로 동일한 병렬 연산 방법.

### 청구항 5

제1항에 있어서,

상기 벡터와 연산되는 원소들을 상기 타겟 랭크에 저장하는 단계는,

상기 매트릭스에서 상기 벡터와 연산되는 원소들만을 상기 타겟 랭크에 저장하여 수행하는 병렬 연산 방법.

### 청구항 6

제1항에 있어서,

상기 메모리 랭크 각각에 포함된 랭크 연산기 각각은,

지정된 연산을 병렬적으로 수행하는 병렬 연산 방법.

### 청구항 7

제1항에 있어서,

상기 병렬 연산 방법은

신경망의 연산시 사용되는 병렬 연산 방법.

**청구항 8**

데이터를 저장하는 복수개의 메모리 랭크(rank)와 상기 메모리 랭크 각각에 포함된 랭크 연산기(rank processor)를 포함하는 메모리에서 수행되는 병렬 연산을 위하여 데이터를 브로드캐스트하는 브로드캐스터로, 상기 브로드캐스터(broadcaster)는:

상기 복수개의 메모리 랭크로부터 각각 입력된 k 비트의 데이터를 입력받고 이들을 출력하는 복수개의 분배부 및

상기 복수 개의 분배부 각각이 제공한 데이터들을 제공받고, 이들을 결합하여 출력하는 복수개의 데이터 결합부를 포함하는 브로드캐스터.

**청구항 9**

제8항에서,

상기 분배부 각각은,

상기 복수개의 메모리 랭크(rank) 중 어느 하나로부터 k 비트의 데이터를 입력받고,

상기 k 비트의 데이터를 상기 데이터 결합부의 개수만큼 출력하는 브로드캐스터.

**청구항 10**

제8항에서,

상기 데이터 결합부는,

결합된 데이터를 각각의 상기 랭크 연산기에 출력하되,

상기 데이터 결합부가 상기 각각의 랭크 연산기에 출력하는 데이터는 서로 동일한 브로드캐스터.

**청구항 11**

제10항에서,

상기 데이터 결합부가 상기 각각의 랭크 연산기에 출력하는 데이터는,

상기 복수개의 메모리 랭크가 출력한 데이터가 순서대로 결합된 데이터인 브로드캐스터.

**청구항 12**

데이터를 저장하는 복수개의 메모리 랭크(rank)와 상기 메모리 랭크 각각에 포함된 랭크 연산기(rank processor)를 포함하는 메모리에서 수행되는 병렬 연산을 위하여 분할된 데이터를 전송하는 분할 전송부로, 상기 분할 전송부는:

각각의 상기 메모리 랭크로부터 데이터를 입력받고, 입력된 상기 데이터를 결합하여 출력하는 데이터 결합부와,

결합된 데이터를 상기 랭크 연산기에 출력하되, 랭크 선택 신호에 상응하는 상기 랭크 연산기에 출력하는 다중화기를 포함하는 분할 전송부.

**청구항 13**

제12항에 있어서,

상기 분할 전송부는,

상기 복수의 메모리 랭크에 분산되어 저장된 데이터들 중에서 상기 병렬 연산에 사용되는 데이터를 전송하는 분할 전송부.

**청구항 14**

제13항에 있어서,

상기 분할 전송부는,

상기 복수의 메모리 랭크에 분산되어 저장된 데이터들 중에서 상기 병렬 연산에 사용되는 데이터를 상기 병렬 연산에 사용되는 데이터로 상기 병렬 연산을 수행하는 상기 랭크 연산기에 출력하는 분할 전송부.

## 발명의 설명

### 기 술 분 야

[0001] 본 기술은 병렬 연산 방법, 브로드캐스터 및 분할 전송부와 관련되며, 메모리에서 수행되는 병렬 연산 방법, 메모리 내의 브로드캐스터 및 분할 전송부와 관련된다.

### 배 경 기 술

[0002] 메모리 패키지는 데이터를 저장하는 메모리 어레이가 형성된 칩들과 외부의 메모리 제어부(memory controller)와 통신하여 데이터 및 제어 신호를 수신/송신하는 버퍼부가 형성된 기판을 포함한다.

[0003] 메모리 패키지는 메모리 칩들이 DIMM(Dual Inline Memory Module), SIMM(Single Inline Memory Module) 형태로 배치되고 복수의 메모리 칩들은 하나의 랭크(rank)를 형성한다. 하나의 메모리 패키지 내에는 복수의 랭크들이 포함된다. 메모리 패키지 외부의 메모리 제어부는 동시에 여러 랭크에 접근할 수 없는 것이 일반적이다.

### 발명의 내용

#### 해결하려는 과제

[0004] 종래의 메모리 패키지와 달리, 랭크에 포함된 연산 장치를 이용하여 병렬 연산을 수행하기 위하여 랭크 인터리빙(rank interleaving) 기술이 개발되었으며, 이로부터 복수의 랭크들에 접근하여 효율적인 연산을 수행할 수 있다. 즉, 모든 랭크에 데이터가 고르게 분포하면 랭크 연산부가 각 랭크에 저장된 데이터에 랭크 레벨로 병렬 억세스할 수 있어 연산 속도가 향상된다.

[0005] 그러나, 랭크 인터리빙을 사용하는 경우에도 어느 하나의 랭크에서 연산한 데이터를 이용하여 전체 결과를 도출하는 경우가 있으며, 이러한 경우까지 연산 속도의 향상 효과를 얻을 수 없었다.

[0006] 본 실시예는 이러한 종래 기술의 난점을 해소하기 위한 것이다. 즉, 본 실시예로 해결하고자 하는 과제 중 하나는, 어느 하나의 랭크에서 연산한 데이터를 이용하여 전체 결과를 도출하는 경우에도 병렬연산을 수행하여 연산 속도를 향상시키기 위한 것이다.

#### 과제의 해결 수단

[0007] 본 실시예는 데이터를 저장하는 복수개의 메모리 랭크(rank)와 상기 메모리 랭크 각각에 포함된 랭크 연산기(rank processor)를 포함하는 메모리에서 수행되는 병렬 연산 방법으로, 복수의 랭크들에 분산되어 저장된 벡터의 원소들을 타겟 랭크를 포함하는 상기 복수의 랭크들 각각에 저장하는 단계와, 상기 복수의 랭크들에 분산되어 저장된 매트릭스에서 상기 벡터와 연산되는 원소들을 상기 타겟 랭크에 저장하는 단계와, 상기 타겟 랭크에 포함된 상기 랭크 연산기가 상기 타겟 랭크에 저장된 상기 벡터의 원소들과, 상기 타겟 랭크에 저장된 상기 벡터와 연산되는 원소들에 대하여 지정된 연산을 수행하는 단계를 포함한다.

[0008] 본 실시예의 어느 한 측면에 의하면 상기 벡터의 원소들을 상기 복수의 랭크들에 저장하는 단계는, 상기 복수의 랭크들에 분산되어 저장된 벡터의 원소들을 결합하여 상기 복수의 랭크들에 저장하여 수행한다.

[0009] 본 실시예의 어느 한 측면에 의하면 결합되어 상기 복수의 랭크들에 저장된 상기 벡터의 원소들은 상기 매트릭스와 연산 가능하도록 결합되어 저장된다.

[0010] 본 실시예의 어느 한 측면에 의하면 결합되어 상기 복수의 랭크들에 저장된 상기 벡터의 원소들은, 서로 동일한 병렬 연산 방법.

[0011] 본 실시예의 어느 한 측면에 의하면 상기 벡터와 연산되는 원소들을 상기 타겟 랭크에 저장하는 단계는, 상기 매트릭스에서 상기 벡터와 연산되는 원소들만을 상기 타겟 랭크에 저장하여 수행한다.

[0012] 본 실시예의 어느 한 측면에 의하면 상기 메모리 랭크 각각에 포함된 랭크 연산기 각각은, 지정된 연산을 병렬적으로 수행한다.

[0013] 본 실시예의 어느 한 측면에 의하면 상기 병렬 연산 방법은 신경망의 연산시 사용된다.

[0014] 본 실시예의 브로드캐스터는 데이터를 저장하는 복수개의 메모리 랭크(rank)와 상기 메모리 랭크 각각에 포함된 랭크 연산기(rank processor)를 포함하는 메모리에서 수행되는 병렬 연산을 위한 브로드캐스터로, 상기 브로드캐스터(broadcaster)는: 상기 복수개의 메모리 랭크로부터 각각 입력된 k 비트의 데이터를 입력받고 이들을 출력하는 복수개의 분배부 및 상기 복수 개의 분배부 각각이 제공한 데이터들을 제공받고, 이들을 결합하여 출력하는 복수개의 데이터 결합부를 포함한다.

[0015] 본 실시예의 어느 한 측면에 의하면 상기 분배부 각각은, 상기 복수개의 메모리 랭크(rank) 중 어느 하나로부터 k 비트의 데이터를 입력받고, 상기 k 비트의 데이터를 상기 데이터 결합부의 개수만큼 출력한다.

[0016] 본 실시예의 어느 한 측면에 의하면 상기 데이터 결합부는, 결합된 데이터를 각각의 상기 랭크 연산기에 출력하되, 상기 데이터 결합부가 상기 각각의 랭크 연산기에 출력하는 데이터는 서로 동일하다.

[0017] 본 실시예의 어느 한 측면에 의하면 상기 데이터 결합부가 상기 각각의 랭크 연산기에 출력하는 데이터는, 상기 복수개의 메모리 랭크가 출력한 데이터가 순서대로 결합된 데이터이다.

[0018] 본 실시예의 분할 전송부는 데이터를 저장하는 복수개의 메모리 랭크(rank)와 상기 메모리 랭크 각각에 포함된 랭크 연산기(rank processor)를 포함하는 메모리에서 수행되는 병렬 연산을 위하여 분할된 데이터를 전송하는 분할 전송부로, 상기 분할 전송부는: 각각의 상기 메모리 랭크로부터 데이터를 입력받고, 입력된 상기 데이터를 결합하여 출력하는 데이터 결합부와, 결합된 데이터를 상기 랭크 연산기에 출력하되, 랭크 선택 신호에 상응하는 상기 랭크 연산기에 출력하는 다중화기를 포함한다.

[0019] 본 실시예의 어느 한 측면에 의하면 상기 분할 전송부는, 상기 복수의 메모리 랭크에 분산되어 저장된 데이터들 중에서 상기 병렬 연산에 사용되는 데이터를 전송한다.

[0020] 본 실시예의 어느 한 측면에 의하면 상기 분할 전송부는, 상기 복수의 메모리 랭크에 분산되어 저장된 데이터들 중에서 상기 병렬 연산에 사용되는 데이터를 상기 병렬 연산에 사용되는 데이터로 상기 병렬 연산을 수행하는 상기 랭크 연산기에 출력한다.

## *발명의 효과*

[0021] 본 기술에 의하면 어느 하나의 랭크에서 연산한 데이터를 이용하여 전체 결과를 도출하는 경우에도 병렬연산을 수행하여 연산 속도를 향상시킬 수 있다는 장점이 제공된다.

## *도면의 간단한 설명*

[0022] 도 1은 본 실시예에 의한 메모리에서 수행되는 병렬 연산 방법의 개요를 도시한 순서도이다.

도 2는 메모리에 포함된 여러 랭크들에 데이터가 저장된 상태를 도시한 도면이다.

도 3은 메모리에 저장된 데이터를 이용하여 사용자가 지정한 연산을 수행하는 과정을 설명하기 위한 도면이다.

도 4는 브로드캐스터의 개요를 도시한 도면이다.

도 5는 벡터 A의 원소들이 결합되어 각각의 랭크들에 저장된 상태를 예시한 도면이다.

도 6은 본 실시예에 의한 분할 전송부의 개요를 도시한 도면이다.

도 7(a)는 $C_0$를 연산하기 위하여 필요한 매트릭스 B의 원소들이 타겟 랭크에 저장된 상태를 예시한 도면이고, 도 7(b)는 $C_9$를 연산하기 위하여 필요한 매트릭스 B의 원소들이 타겟 랭크에 저장된 상태를 예시한 도면이다.

## *발명을 실시하기 위한 구체적인 내용*

[0023] 이하에서는 첨부된 도면들을 참조하여 본 실시예를 설명한다. 도 1은 본 실시예에 의한 메모리에서 수행되는 병렬 연산 방법의 개요를 도시한 순서도이다. 도 1을 참조하면, 본 실시예의 병렬 연산 방법은 데이터를 저장하는 복수개의 메모리 랭크(rank)와 상기 메모리 랭크 각각에 포함된 랭크 연산기(rank processor)를 포함하는 메모리에서 수행되는 병렬 연산 방법으로, 복수의 랭크들에 분산되어 저장된 벡터의 원소들을 타겟 랭크를 포함하는

상기 복수의 랭크들 각각에 저장하는 단계(S100)와, 상기 복수의 랭크들에 분산되어 저장된 매트릭스에서 상기 벡터와 연산되는 원소들을 상기 타겟 랭크에 저장하는 단계(S200와, 상기 타겟 랭크에 포함된 상기 랭크 연산기가 상기 타겟 랭크에 저장된 상기 벡터의 원소들과, 상기 타겟 랭크에 저장된 상기 벡터와 연산되는 원소들에 대하여 지정된 연산을 수행하는 단계(S300)를 포함한다.

[0024] 도 2는 메모리에 포함된 여러 랭크들에 데이터가 저장된 상태를 도시한 도면이다. 도 1 및 도 2를 참조하면, 메모리(10)는 복수의 랭크들(100a, 100b, 100c, 100d)을 포함하고, 각 랭크는 데이터를 저장하는 메모리 어레이(110a, 110b, 110c, 110d)와 메모리 어레이(110a, 110b, 110c, 110d)에 저장된 데이터로부터 사용자가 지정한 연산을 수행하는 랭크 연산기(rank processor, 120a, 120b, 120c, 120d) 및 브로드캐스터(210)와 분할 전송기(220)를 포함하는 로직부(200)를 포함한다.

[0025] 도 3은 메모리에 저장된 데이터를 이용하여 사용자가 지정한 연산을 수행하는 과정을 설명하기 위한 도면이다. 일 실시예로, 사용자가 지정한 연산이 아래의 코드로 표현되는 연산으로, 1차원 벡터 A와 2차원 매트릭스 B의 원소의 곱을 합하여 1차원 벡터 어레이 C의 각 원소로 저장하는 연산인 경우를 예시한다. 다만, 이는 실시예일 따름이며, 매트릭스와 매트릭스의 원소의 곱을 합하는 연산의 경우에도 적용될 수 있다.

[0026] for(i=0; i<16; i++){

[0027] for(j=0; j<16; j++){

[0028] C[i] += A[j] * B[i][j];

[0029] }

[0030] }

[0031] 이하에서는 간결하고 명확하게 설명하기 위하여 벡터 A는 0에서 15 까지 16개의 정수형 원소(integer element)들을 가지고, 매트릭스 B는 0,0 에서 15, 15의 256개의 정수형 원소를 가지는 것을 예시한다. 또한, 각 랭크(100a, 100b, 100c, 100d)에 포함된 메모리 어레이(110a, 110b, 110c, 110d)는 하나의 행(row)에 16 바이트(16byte)의 저장용량을 가져 하나의 행에 네 개의 정수형 원소를 저장하는 것을 예시한다.

[0032] 다만 코드는 벡터 A의 j 번째 원소를 A[i]로 표기하고, 매트릭스 B의 i 행, j 열 원소를 B[i],[j]로 표기하나, 이하에서는 보다 간결한 설명 및 도시를 위하여 벡터 A의 j 번째 원소를 $A_j$로 표기하고, 매트릭스 B의 i 행, j 열 원소를 $B_{i,j}$로 표기하도록 한다.

[0033] 벡터 A, C와 매트릭스 B는 랭크 인터리빙되어 도 3으로 예시된 것과 같이 각 원소들이 메모리 어레이(110a, 110b, 110c, 110d)들에 걸쳐 고르게 저장된다. 원소들이 랭크 인터리빙 되어 저장됨에 따라 연산되는 데이터들이 동일한 랭크에 저장되는 경우에는 병렬 연산이 가능하므로 연산 속도가 향상된다는 장점이 제공된다.

[0034] 그러나, 예시된 코드와 같이 연산을 완료하기 위하여 모든 랭크에 저장된 데이터가 필요한 경우에는 각 랭크에서 해당 데이터를 읽고, 연산을 수행하여 연산 결과를 타겟 랭크에 저장하여야 하므로 병렬 연산이 불가능하다. 이하에서는 연산을 완료하기 위하여 모든 랭크에 저장된 데이터가 필요한 경우임에도 병렬 연산을 수행할 수 있는 예를 설명하도록 한다.

[0035] 사용자가 제공한 코드는 연산 결과를 도출하기 위하여 모든 벡터 A의 모든 원소들이 필요하다. 일 예로, 상기한 코드에서 $C_0$의 값을 연산하기 위하여 $A_0$ 내지 $A_{15}$의 값이 모두 필요하다. 마찬가지로, $C_8$의 값을 연산하기 위하여 $A_0$ 내지 $A_{15}$의 값이 모두 필요하다.

[0036] 도 4는 브로드캐스터(broadcaster, 210)의 개요를 도시한 도면이고, 도 5는 벡터 A의 원소들이 결합되어 각각의 랭크들(100a, 100b, 100c, 100d)에 저장된 상태를 예시한 도면이다. 도 4 및 도 5를 참조하면, 브로드캐스터(210)는 복수개의 분배부들(212a, 212b, 212c, 212d)과 복수개의 데이터 결합부(214a, 214b, 214c, 214d)를 포함한다. 일 실시예로, 각각의 분배부(212a, 212b, 212c, 212d)와 데이터 결합부(214a, 214b, 214c, 214d)는 각각의 랭크에 포함될 수 있으며, 분배부(212a, 212b, 212c, 212d)와 데이터 결합부(214a, 214b, 214c, 214d)를 포함하는 브로드캐스터(210)는 로직부(200)에 포함될 수 있다.

[0037] 분배부들(212a, 212b, 212c, 212d)은 각 랭크에 저장된 벡터 A의 원소들을 읽고 데이터 결합부(214a, 214b, 214c, 214d)에 제공한다. 도시된 실시예에서, 분배부(212a)는 랭크 100a에 저장된 벡터 A의 원소 $A_0$ ~ $A_3$을 읽

고 데이터 결합부(214a, 214b, 214c, 214d)에 출력하고, 분배부(212b)는 랭크 100b에 저장된 벡터 A의 원소 $A_4$ ~ $A_7$을 읽고 데이터 결합부(214a, 214b, 214c, 214d)에 출력한다. 분배부(212c)는 랭크 100c에 저장된 벡터 A의 원소 $A_8$ ~ $A_{11}$을 읽고 데이터 결합부(214a, 214b, 214c, 214d)에 출력하고, 분배부(212d)는 랭크 100d에 저장된 벡터 A의 원소 $A_8$ ~ $A_{11}$을 읽고 데이터 결합부(214a, 214b, 214c, 214d)에 출력한다. 따라서, 각각의 데이터 결합부(214a, 214b, 214c, 214d)에는 벡터 A의 모든 원소가 제공된다. 데이터 결합부(214a, 214b, 214c, 214d)는 연산대상인 매트릭스 B의 원소와 연산이 가능하도록 벡터 A의 원소들을 결합한다.

[0038] 데이터 결합부(214a, 214b, 214c, 214d) 각각은 데이터 분배부(212a, 212b, 212c, 212d)들이 제공한 벡터 A의 원소들을 결합하고 각각의 메모리 랭크(100a, 100b, 100c, 100d)에 저장한다.

[0039] 종래 기술에 의하면 벡터는 곱셈의 대상인 매트릭스의 열(Column)의 개수 만큼 연산에서 재사용되며, 랭크 인터리빙을 통해 벡터 원소와 매트릭스 원소를 각각 빠르게 접근할 수 있지만, 매트릭스의 열마다 같은 벡터를 반복적으로 접근해야 했다.

[0040] 그러나, 본 실시예의 브로드캐스터는 한 번의 메모리 접근으로 모든 랭크에 데이터를 제공한다. 즉, 단일한 데이터 획득을 통해, 랭크 수 X 랭크 수의 데이터 를 가져올 수 있어서 데이터 획득의 효율성을 향상시킬 수 있으며, 이로부터 연산 속도 향상의 효과가 제공된다.

[0041] 데이터 결합부(214a, 214b, 214c, 214d)들이 결합하여 출력한 벡터 A의 원소들은 도 5로 예시된 것과 같이 각각의 랭크들에 저장된다(S100).

[0042] 도 6은 본 실시예에 의한 분할 전송부(220)의 개요를 도시한 도면이다. 도 6을 참조하면, 분할 전송부(220)는, 데이터 결합부(222)와 다중화기(224)를 포함한다. 일 실시예에서, 사용자가 입력한 코드에서 $C_0$는 아래의 수학식 1과 같이 연산된다.

[0043] [수학식 1]

[0044] $$C_0 = A_0 \times B_{0,0} + A_1 \times B_{0,1} + A_2 \times B_{0,2} + \ldots + A_{15} \times B_{0,15}$$

[0045] 매트릭스 B의 원소 $B_{0,0}$, ..., $B_{0,3}$은 랭크 100a에 저장되어 있고, 원소 $B_{0,4}$, ... , $B_{0,7}$는 랭크 100b에 저장되어 있고, 원소 $B_{0,8}$, ..., $B_{0,11}$은 랭크 100c에 저장되어 있으며, 원소 $B_{0,12}$, ..., $B_{0,15}$는 랭크 100d에 저장되어 있다.

[0046] 데이터 결합부(222)는 각각의 랭크들(100a, 100b, 100c, 100d)에서 연산에 필요한 매트릭스의 원소들을 제공받고 결합하여 다중화기(224)에 출력한다. 즉, 데이터 결합부(222)는 랭크들(100a, 100b, 100c, 100d)로부터 연산에 필요한 행(혹은 열)의 원소만을 제공받고, 이들을 결합하여 다중화기(224)에 출력한다.

[0047] 다중화기(224)는 데이터 결합부(222)가 출력한 연산에 필요한 행(혹은 열)의 원소들을 제공받고, 랭크 선택 신호(rank_sel)에 따라 최종적으로 연산이 수행되어 결과가 저장되는 타겟 랭크에 출력한다. 위의 수학식 1로 예시된 연산에서 연산은 랭크 100a에 포함된 랭크 연산기 120a가 수행하고, 연산된 결과는 메모리 어레이 110a의 $C_0$에 저장된다. 따라서, 타겟 랭크는 100a이다. 이와 같이, $C_4$ ~ $C_7$ 값을 연산할 때의 타겟 랭크는 100b이고, $C_8$~ $C_{11}$ 을 연산할 때의 타겟 랭크는 100c 이며, $C_{12}$ ~ $C_5$ 값을 연산할 때 타겟 랭크는 100d 이다.

[0048] 도 7(a)는 $C_0$를 연산하기 위하여 필요한 매트릭스 B의 원소들인 $B_{0,0}$, $B_{0,1}$, ..., $B_{0,15}$들이 타겟 랭크 100a에 저장된 상태를 예시한 도면이고, 도 7(b)는 $C_9$를 연산하기 위하여 필요한 매트릭스 B의 원소들인 $B_{9,0}$, $B_{9,1}$, ..., $B_{9,15}$들이 타겟 랭크 100a에 저장된 상태를 예시한 도면이다. 도 6 및 도 7(a)를 참조하면, 다중화기(224)는 데이터 결합부(222)가 출력한 $C_0$를 연산시 요청되는 매트릭스 B의 원소 $B_{0,0}$, $B_{0,1}$, ..., $B_{0,15}$을 제공받고, 타겟 랭크인 100a에 제공하여 메모리 어레이 110a 저장한다(S200).

[0049] 또한 도 7(b)로 예시된 것과 같이 다중화기(224)는 $C_9$를 연산하기 위하여 데이터 결합부(222)가 출력한 필요한 매트릭스 B의 원소 $B_{9,0}$, $B_{9,1}$, ..., $B_{9,15}$을 제공받고, 타겟 랭크인 100c에 제공하여 메모리 어레이 110c에 저장한다. 도시되지 않았으나, 다중화기는 필요한 C의 원소값을 연산하기 위하여 데이터 결합부(222)가 출력한 필요한 매트릭스 B의 원소들을 제공 받고, 타겟 랭크에 제공하여 메모리 어레이에 저장한다. 본 실시예에 따른 분할

전송부에 의하면 연산 대상인 매트릭스의 값을 메모리에 할당하는 할당 과정의 복잡도를 낮출 수 있다는 장점이 제공된다.

[0050] 랭크들(100a, 100b, 100c, 100d)에 포함된 랭크 연산기(120a, 120b, 120c, 120d)는 각 랭크들(100a, 100b, 100c, 100d)에 저장된 벡터 A와 연산에 필요한 매트릭스 B의 원소들에 대한 연산을 병렬적으로 수행(S300)할 수 있으며, 이로부터 연산 속도 성능을 향상시킬 수 있다는 장점이 제공된다.

[0051] 위에서 설명된 실시예는 벡터와 매트릭스의 원소별 곱을 합하는 것으로 설명되었다. 그러나, 이는 수행할 수 있는 일 예로, 매트릭스와 매트릭스 사이의 원소별 곱의 합을 구하는 연산등에서도 당연히 수행될 수 있다. 나아가 일 예로 본 실시예는 입력된 벡터 원소와 웨이트 행렬 원소들 사이 곱의 합을 구하는 신경망의 연산 등에서 사용될 수 있을 것이다.

[0053] 본 발명에 대한 이해를 돕기 위하여 도면에 도시된 실시 예를 참고로 설명되었으나, 이는 실시를 위한 실시예로, 예시적인 것에 불과하며, 당해 분야에서 통상적 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시 예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호범위는 첨부된 특허청구범위에 의해 정해져야 할 것이다.

## *부호의 설명*

[0054] 10: 메모리

100a, 100b, 100c, 100d: 랭크

110a, 110b, 110c, 110d: 메모리 어레이

120a, 120b, 120c, 120d: 랭크 연산기

200: 로직부 210: 브로드캐스터

212a, 212b, 212c, 212d: 분배부

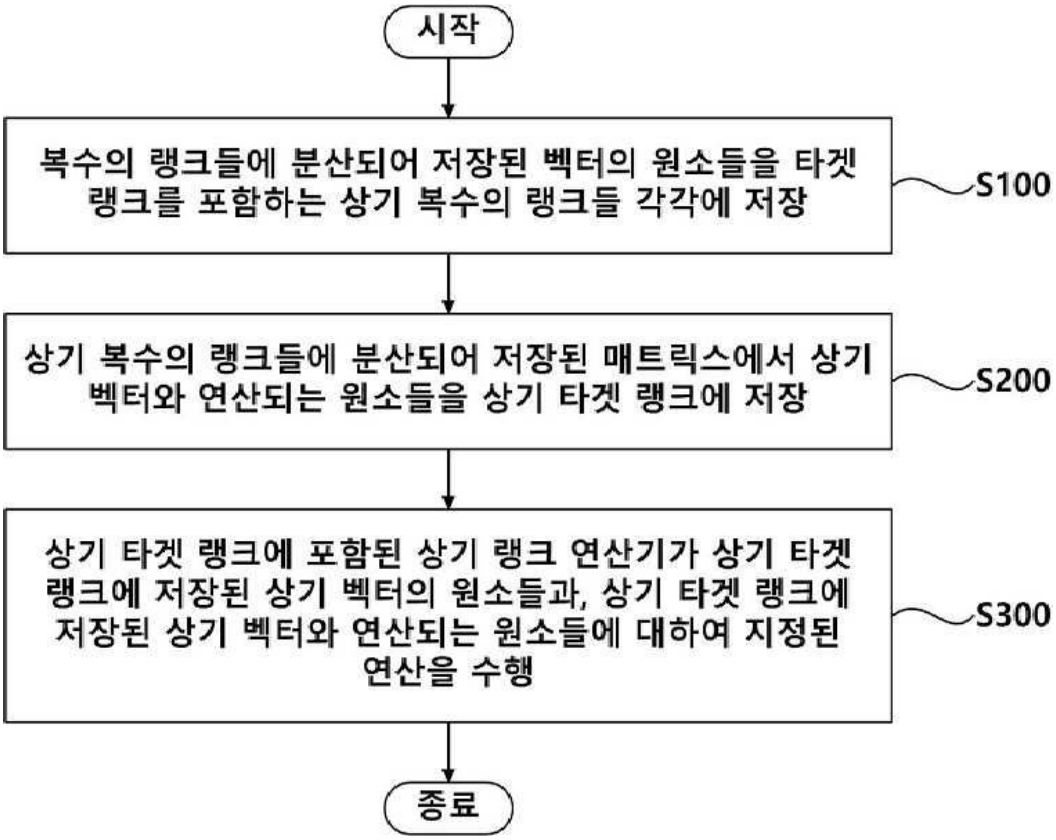
214a, 214b, 214c, 214d: 데이터 결합부

220: 분할 전송부 222: 데이터 결합부

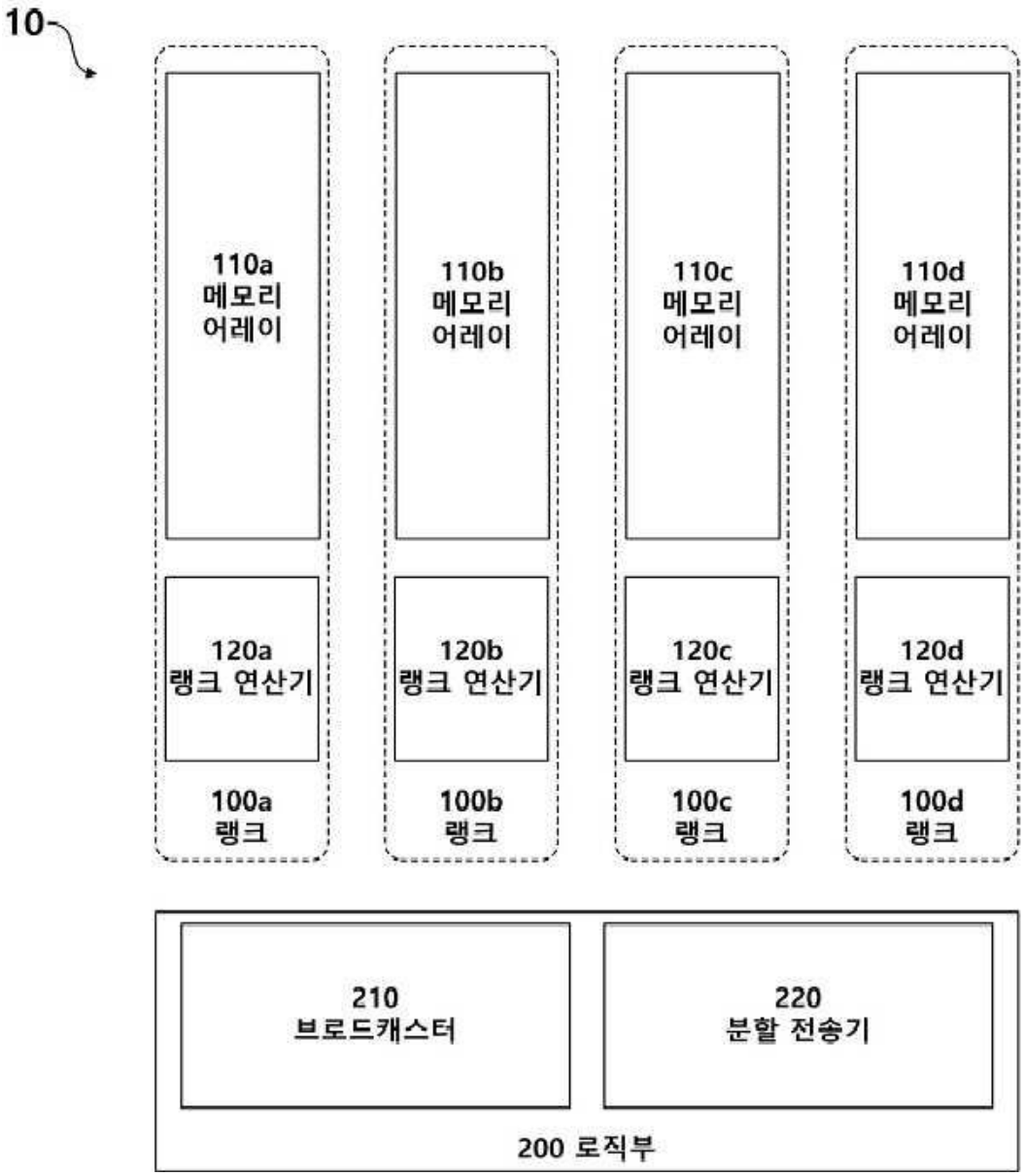
224: 다중화기

도면

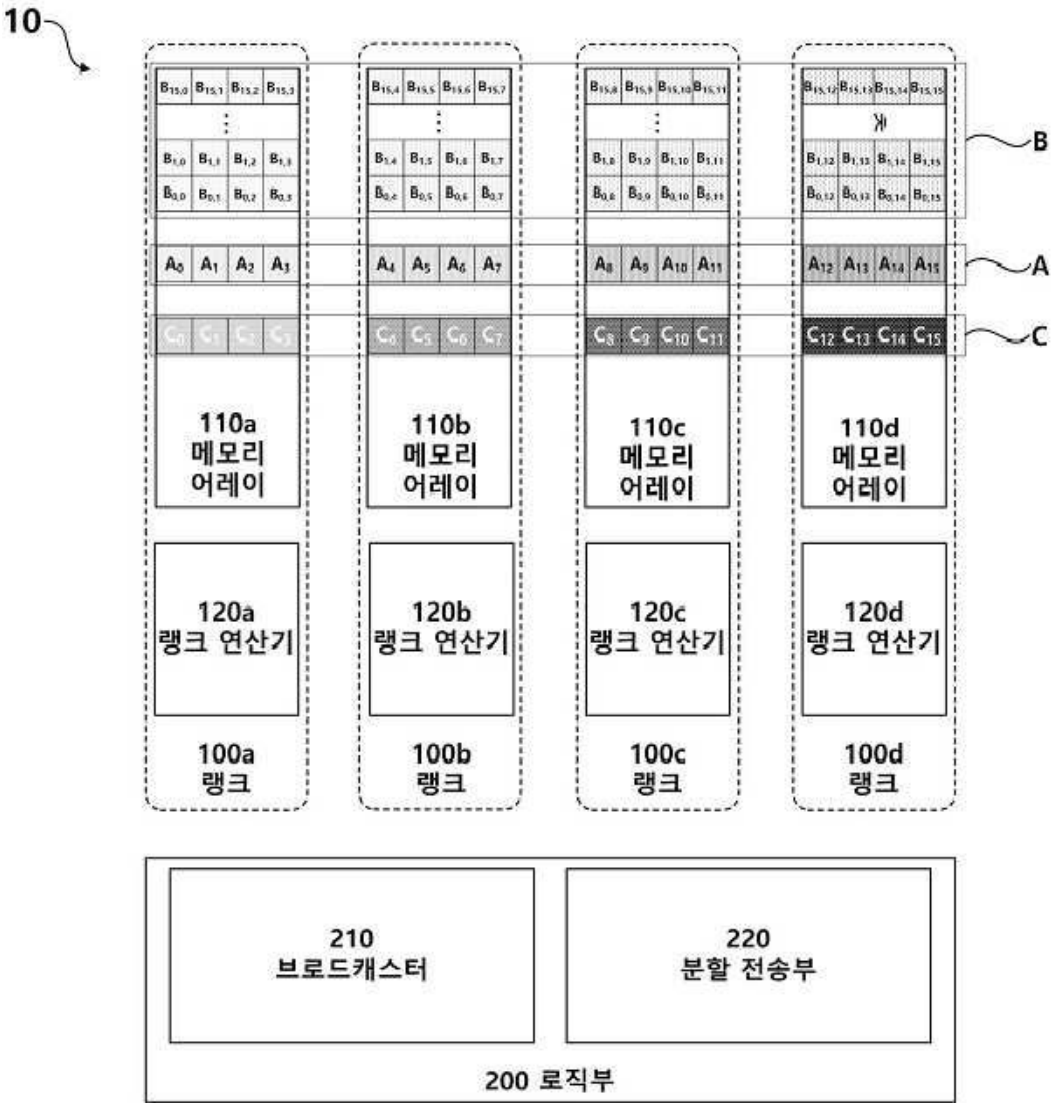
도면1


시작
복수의 랭크들에 분산되어 저장된 벡터의 원소들을 타겟 랭크를 포함하는 상기 복수의 랭크들 각각에 저장
S100
상기 복수의 랭크들에 분산되어 저장된 매트릭스에서 상기 벡터와 연산되는 원소들을 상기 타겟 랭크에 저장
S200
상기 타겟 랭크에 포함된 상기 랭크 연산기가 상기 타겟 랭크에 저장된 상기 벡터의 원소들과, 상기 타겟 랭크에 저장된 상기 벡터와 연산되는 원소들에 대하여 지정된 연산을 수행
S300
종료

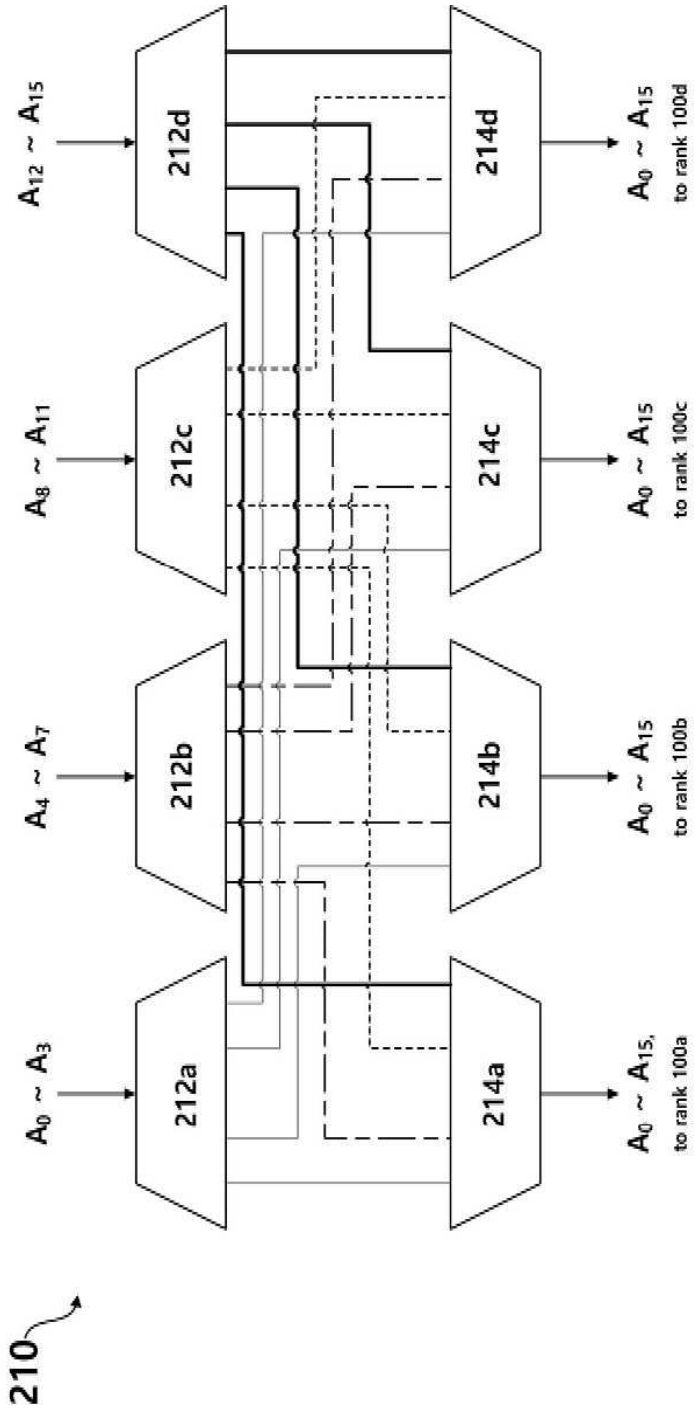
도면2


10
110a
메모리
어레이
110b
메모리
어레이
110c
메모리
어레이
110d
메모리
어레이
120a
랭크 연산기
120b
랭크 연산기
120c
랭크 연산기
120d
랭크 연산기
100a
랭크
100b
랭크
100c
랭크
100d
랭크
210
브로드캐스터
220
분할 전송기
200 로직부

도면3


10
B
A
C
110a
메모리
어레이
110b
메모리
어레이
110c
메모리
어레이
110d
메모리
어레이
120a
랭크 연산기
120b
랭크 연산기
120c
랭크 연산기
120d
랭크 연산기
100a
랭크
100b
랭크
100c
랭크
100d
랭크
210
브로드캐스터
220
분할 전송부
200 로직부

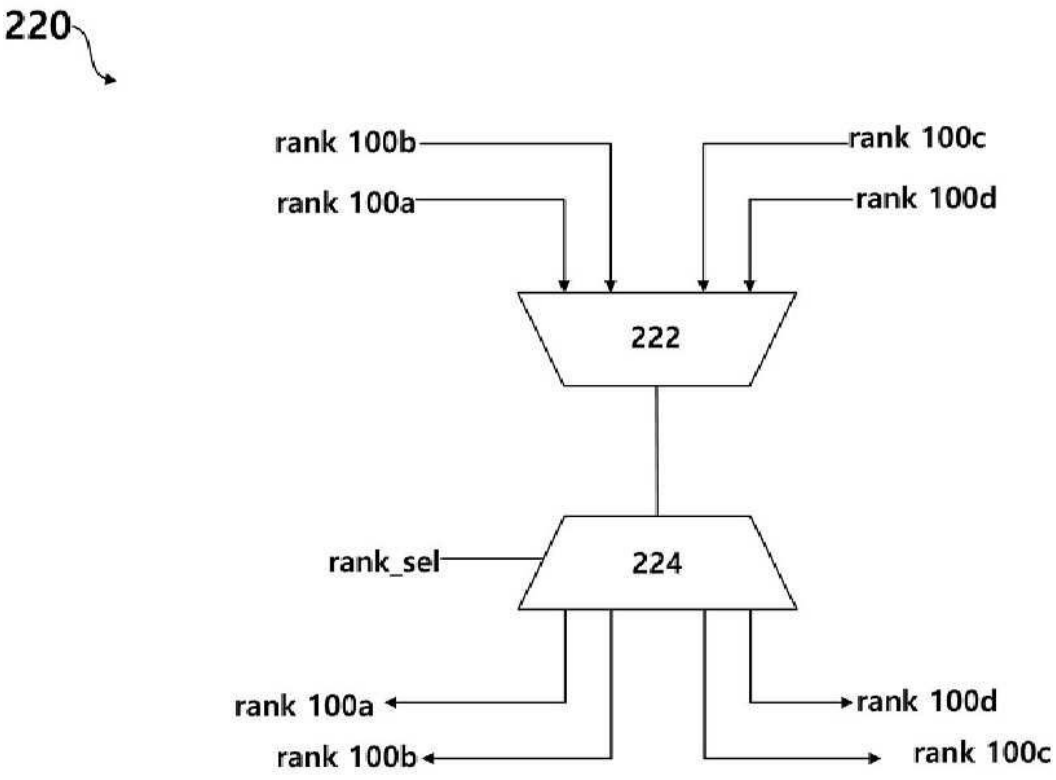
도면4


210
A12 ~ A15
212d
214d
A0 ~ A15
to rank 100d
A8 ~ A11
212c
214c
A0 ~ A15
to rank 100c
A4 ~ A7
212b
214b
A0 ~ A15
to rank 100b
A0 ~ A3
212a
214a
A0 ~ A15,
to rank 100a

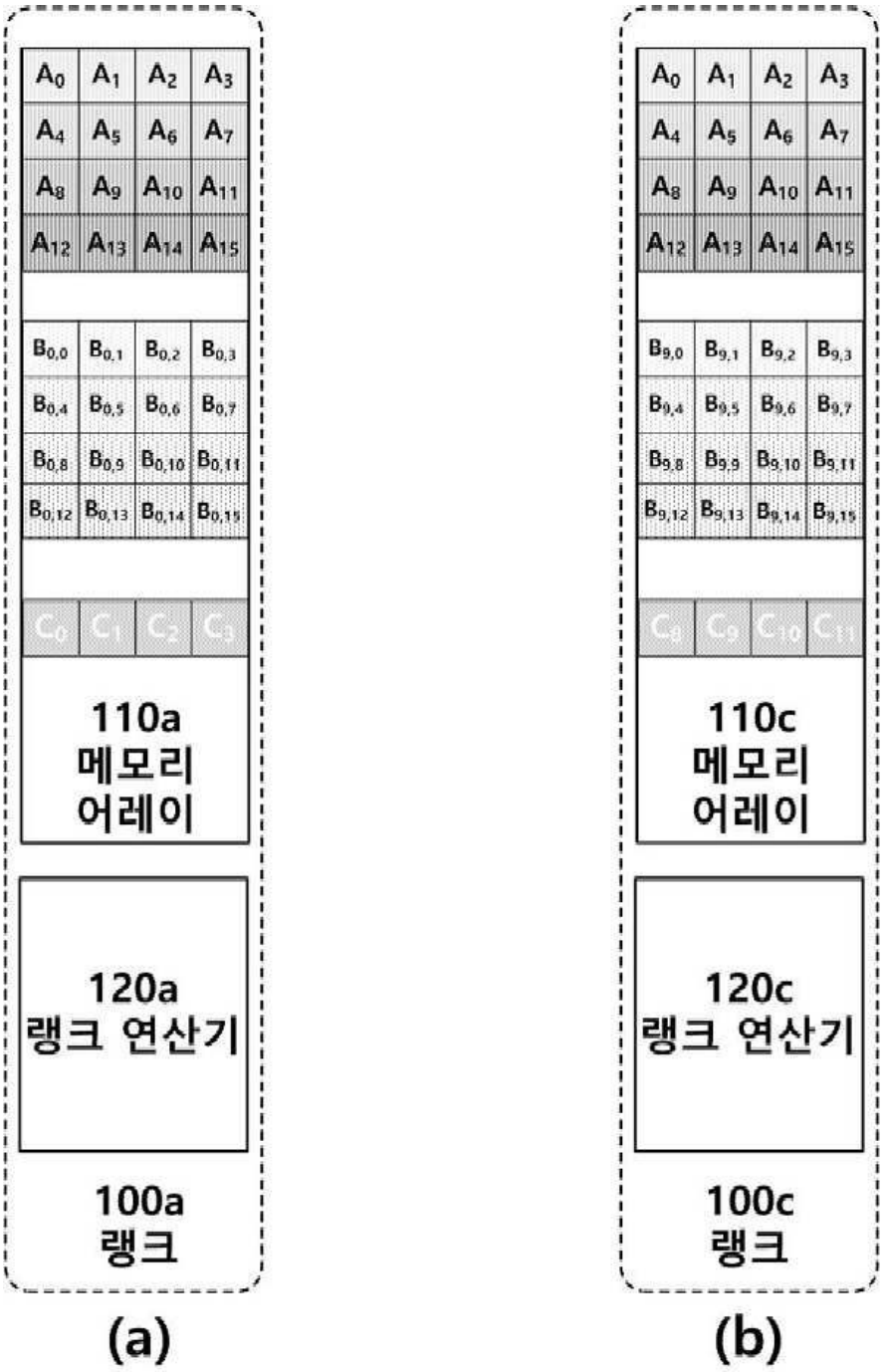
도면5


10
110a
메모리
어레이
110b
메모리
어레이
110c
메모리
어레이
110d
메모리
어레이
120a
랭크 연산기
120b
랭크 연산기
120c
랭크 연산기
120d
랭크 연산기
100a
랭크
100b
랭크
100c
랭크
100d
랭크
210
브로드캐스터
220
분할 전송부
200 로직부

도면6


220
rank 100b
rank 100a
rank 100c
rank 100d
222
rank_sel
224
rank 100a
rank 100b
rank 100d
rank 100c

도면7


A0 A1 A2 A3
A4 A5 A6 A7
A8 A9 A10 A11
A12 A13 A14 A15
B0,0 B0,1 B0,2 B0,3
B0,4 B0,5 B0,6 B0,7
B0,8 B0,9 B0,10 B0,11
B0,12 B0,13 B0,14 B0,15
C0 C1 C2 C3
110a 메모리 어레이
120a 랭크 연산기
100a 랭크
(a)
A0 A1 A2 A3
A4 A5 A6 A7
A8 A9 A10 A11
A12 A13 A14 A15
B9,0 B9,1 B9,2 B9,3
B9,4 B9,5 B9,6 B9,7
B9,8 B9,9 B9,10 B9,11
B9,12 B9,13 B9,14 B9,15
C8 C9 C10 C11
110c 메모리 어레이
120c 랭크 연산기
100c 랭크
(b)