



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2024년09월04일

(11) 등록번호 10-2702141

(24) 등록일자 2024년08월29일

(51) 국제특허분류(Int. Cl.)
H10B 63/00 (2023.01) G06N 3/0464 (2023.01)
G06N 3/063 (2023.01) H10N 70/00 (2024.01)
(52) CPC특허분류
H10B 63/84 (2023.02)
G06N 3/0464 (2023.01)
(21) 출원번호 10-2023-0092525
(22) 출원일자 2023년07월17일
심사청구일자 2023년07월17일
(56) 선행기술조사문헌
KR1020160018761 A*
T. Park et al., Nanoscale, 2023, Vol. 15, pp.
6387-6395 (2023. 3. 2.)*
KR1020220130299 A
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
울산과학기술원
울산광역시 울주군 언양읍 유니스트길 50
(72) 발명자
서준기
울산광역시 울주군 언양읍 유니스트길 50
허남욱
울산광역시 울주군 언양읍 유니스트길 50
(74) 대리인
특허법인 무한

전체 청구항 수 : 총 11 항

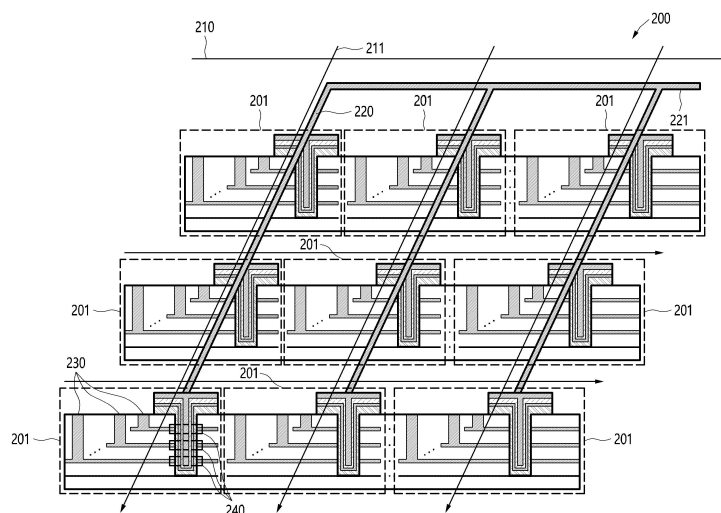
심사관 : 김진우

(54) 발명의 명칭 3차원 컨볼루션 신경망(3D Convolutional neural network, 3D CNN) 연산을 수행하는 메모리스터 기반의 어레이 장치 및 이를 통한 3D CNN 연산 방법

(57) 요약

일 실시예에 따른, 메모리스터 어레이 장치는, 복수의 절연층들의 각각 및 복수의 하부 전극층들의 각각이 교대로 적층된 제1 적층 구조체 및 제1 적층 구조체에 형성된 홀(hole) 내에 저항변화소자 물질층, 중간 전극층, 선택소자 물질층 및 상부 전극이 적층되고 복수의 하부 전극층들로부터 상부 전극으로 향하는 전기 경로들에서 형성되는 복수의 메모리스터 소자들을 포함하는 제2 적층 구조체를 가지는 수직형 메모리스터 기반 추론 소자를 포함할 수 있다.

대표도



(52) CPC특허분류

G06N 3/063 (2013.01)
H10B 63/24 (2023.02)
H10N 70/8265 (2024.05)
H10N 70/882 (2023.02)
H10N 70/883 (2023.02)

임동혁

울산광역시 울주군 언양읍 유니스트길 50

(72) 발명자

이동현

울산광역시 울주군 언양읍 유니스트길 50

이 발명을 지원한 국가연구개발사업

과제고유번호	1711188649
과제번호	2022M3I7A2079098
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	PIM인공지능반도체핵심기술개발(소자)
연구과제명	고성능 PRAM 기반 학습용 PIM 소자 및 연산 시스템개발
기 여 율	50/100
과제수행기관명	울산과학기술원
연구기간	2023.01.01 ~ 2023.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711187371
과제번호	2021R1C1C2012077
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)
연구과제명	물질의 자발성 및 불확실성을 이용한 나노소자 기반두뇌모사 인공신경망 시스템 구

현

기 여 율	25/100
과제수행기관명	울산과학기술원
연구기간	2023.03.01 ~ 2024.02.29

이 발명을 지원한 국가연구개발사업

과제고유번호	1711181452
과제번호	2021M3F3A2A01037844
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	차세대지능형반도체기술개발(소자)
연구과제명	저항 변화성 소재 및 구조의 정밀 제어를 통한 고성능고신뢰성 2단자 시냅스 소자

및 어레이 개발

기 여 율	25/100
과제수행기관명	울산과학기술원
연구기간	2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

멤리스터 어레이 장치에 있어서,

복수의 절연층들의 각각 및 복수의 하부 전극층들의 각각이 교대로 적층된 제1 적층 구조체; 및

상기 제1 적층 구조체에 형성된 홀(hole) 내에 저항변화소자 물질층, 중간 전극층, 선택소자 물질층, 및 상부 전극이 적층되고, 상기 복수의 하부 전극층들로부터 상기 상부 전극으로 향하는 전기 경로들에서 형성되는 복수의 멤리스터 소자들을 포함하는 제2 적층 구조체

를 가지는 수직형 멤리스터 기반 추론 소자

를 포함하고

입력 라인 및 출력 라인을 따라 어레이 구조로 배치된 복수의 멤리스터 기반 추론 소자들;

상기 복수의 멤리스터 기반 추론 소자들 중 동일한 출력 라인에 연결된 적어도 하나의 멤리스터 기반 추론 소자의 제2 적층 구조체들과 연결되는 연산 전극층; 및

상기 연산 전극층으로부터 전달된 컬럼 연산 결과와 다른 출력 라인의 다른 연산 전극층으로부터 전달된 컬럼 연산 결과에 기초한 출력 연산 결과를 출력하는 출력 전극층

을 포함하며,

상기 복수의 멤리스터 기반 추론 소자들 중 한 소자가 상기 수직형 멤리스터 기반 추론 소자이고,

상기 연산 전극층은,

상기 적어도 하나의 멤리스터 기반 추론 소자의 각각에서 복수의 하부 전극층들을 통해 수신되는 입력들과 복수의 멤리스터 소자들에 저장된 가중치들 간의 개별 연산 결과를 수신하며, 상기 적어도 하나의 멤리스터 기반 추론 소자로부터 수신된 상기 개별 연산 결과를 병합함으로써 생성된 컬럼 연산 결과를 상기 출력 전극층에게 전달하는

멤리스터 어레이 장치.

청구항 2

삭제

청구항 3

제1항에 있어서,

상기 멤리스터 어레이 장치는,

상기 복수의 멤리스터 기반 추론 소자들 중 같은 입력 라인에 연결된 하나 이상의 멤리스터 기반 추론 소자의 제2 적층 구조체가 같은 하부 전극층에 연결되는,

멤리스터 어레이 장치.

청구항 4

제1항에 있어서,

상기 멤리스터 어레이 장치는,

상기 연산 전극층의 제1 연산에 기초한 출력을 저장하는 제1 버퍼; 및
 상기 다른 연산 전극층의 제2 연산에 기초한 출력을 저장하는 제2 버퍼
 를 더 포함하고,
 상기 출력 전극층은,
 상기 제1 연산 및 상기 제2 연산이 개별적으로(individually) 수행되는 경우, 상기 제1 버퍼로부터 수신된 상기 제1 연산에 기초한 출력 및 상기 제2 버퍼로부터 수신된 상기 제2 연산에 기초한 출력을 통합한 결과를 출력하는,
 메모리스터 어레이 장치.

청구항 5

제1항에 있어서,
 상기 복수의 메모리스터 기반 추론 소자들 중 각 메모리스터 기반 추론 소자는,
 다른 메모리스터 기반 추론 소자들에 대해 병렬적으로, 입력 데이터로부터 미리 정해진 크기의 3차원 커널(kernel)에 대응하여 획득된 3차원 부분 데이터 중 해당하는 값들을 복수의 하부 전극 층들을 통해 수신하고,
 상기 다른 메모리스터 기반 추론 소자들에 대해 병렬적으로, 상기 해당하는 값들과 해당하는 메모리스터 기반 추론 소자에 형성된 복수의 메모리스터 소자들의 가중치들 간의 연산 결과를 생성하는,
 메모리스터 어레이 장치.

청구항 6

제1항에 있어서,
 상기 저항변화소자 물질층은,
 금속 산화물 또는 칼코제나이드계 물질 중 적어도 하나를 포함하고,
 상기 선택소자 물질층은,
 칼코제나이드계 물질을 포함하는,
 메모리스터 어레이 장치.

청구항 7

제1항에 있어서,
 상기 메모리스터 어레이 장치는,
 상기 복수의 하부 전극층들에 입력 데이터를 입력하는 주변 회로(peripheral circuit)에 대해 적층되는,
 메모리스터 어레이 장치.

청구항 8

메모리스터 어레이 장치에 기초한 3차원 컨볼루션 신경망(3D Convolutional neural network) 연산 방법에 있어서,
 입력 데이터로부터 미리 정해진 크기의 3차원 커널(kernel)에 대응하는 3차원 부분 데이터를 획득하는 단계;
 상기 획득된 부분 데이터의 값들을, 수직으로 이격되어 형성된 복수의 메모리스터 소자들에 개별적으로 연결된 복

수의 하부 전극층들에 입력하는 단계;

상기 복수의 하부 전극층에 입력된 값들과 상기 복수의 메모리스터 소자들에 저장된 가중치들 간의 개별 연산 결과들의 합산에 대응하는 컬럼 연산 결과를 생성하는 단계;

상기 생성된 컬럼 연산 결과 및 다른 출력 라인에 연결된 복수의 메모리스터 기반 추론 소자들의 컬럼 연산 결과를 출력 전극층에 전달하는 단계; 및

상기 전달된 컬럼 연산 결과들을 병합하여, 출력 연산 결과를 출력하는 단계

를 포함하는

3차원 컨볼루션 신경망 연산 방법.

청구항 9

제8항에 있어서,

상기 메모리스터 어레이 장치가, 입력 라인 및 출력 라인을 따라 어레이 구조로 배치된 복수의 메모리스터 기반 추론 소자들을 포함하고,

상기 입력하는 단계는,

상기 3차원 부분 데이터를 복수의 메모리스터 소자들 각각에 포함된 복수의 하부 전극 층들에 병렬적으로 입력하는 단계

를 포함하는

3차원 컨볼루션 신경망 연산 방법.

청구항 10

제8항에 있어서,

상기 메모리스터 어레이 장치가, 입력 라인 및 출력 라인을 따라 어레이 구조로 배치된 복수의 메모리스터 기반 추론 소자들을 포함하고, 상기 복수의 메모리스터 기반 추론 소자들 중 같은 입력 라인에 연결된 하나 이상의 메모리스터 기반 추론 소자의 제2 적층 구조체가 같은 하부 전극층에 연결되며,

상기 입력하는 단계는,

상기 3차원 부분 데이터를 슬라이스(slice)함으로써 복수의 2차원 데이터를 형성하는 단계; 및

상기 형성된 복수의 2차원 데이터 중 하나의 데이터에 해당하는 값을, 출력 라인들 중 해당 출력 라인에 연결된 복수의 메모리스터 기반 추론 소자들의 해당 복수의 하부 전극층들에 입력하는 단계;

를 포함하는

3차원 컨볼루션 신경망 연산 방법.

청구항 11

제8항에 있어서,

상기 생성된 컬럼 연산 결과를 버퍼에 저장하는 단계

를 더 포함하는

3차원 컨볼루션 신경망 연산 방법.

청구항 12

제8항에 있어서,

상기 메모리스트 어레이 장치는, 상기 복수의 하부 전극층들에 입력 데이터를 입력하는 주변 회로(peripheral circuit)에 대해 적층되고,

상기 입력하는 단계는,

상기 주변 회로가 상기 복수의 하부 전극층들에 데이터를 입력하는 단계

를 포함하는

3차원 컨볼루션 신경망 연산 방법.

발명의 설명

기술 분야

[0001] 본 개시는 3차원 컨볼루션 신경망(3D Convolutional neural network, 3D CNN) 연산을 수행하기 위한 장치 및 이를 이용한 3D CNN 연산 방법에 관한 것이다.

배경 기술

[0003] 3차원 컨볼루션 신경망(3D Convolutional neural network, 3D CNN)은 2차원 컨볼루션 신경망(2D Convolutional neural network, 2D CNN)에서 사용하는 2D 형태의 가중치(weight)가 아닌, 3D 형태의 가중치를 사용하여 컨볼루션 연산을 한다. 이에, 3D CNN 연산은, 3D 형태의 가중치를 포함하는 커널(kernel)을 입력 데이터에 대하여 3방향으로 슬라이드(slide) 함으로써 컨볼루션 연산을 수행하고, 이를 통해 하나의 결과 값을 도출한다. 이때, 3D CNN 학습에는 3D 물체 혹은 시간축 정보가 포함된 비디오 데이터가 사용될 수 있다.

[0004] 일반적인 인공 신경망을 활용한 뉴로모픽 기반 학습(예를 들어, 2D CNN, Artificial neural network(ANN), Deep neural network(DNN) 등)에서는 입력 데이터들을 1차원상의 부분 집합으로 나누어 학습을 진행해도 무방하기 때문에, 기존의 2차원 메모리스트 어레이 구조의 소자를 구현하여 이를 통해 연산 하는 것이 용이하다.

[0005] 그러나, 3D CNN 학습은 최종적으로 3차원 데이터를 처리해야 하므로, 학습 진행 시 입력 데이터들을 1차원 상의 부분 집합으로 나누면 그 수가 매우 많아진다. 구체적인 예시로서, 입력 데이터의 크기가 N^3 경우, 1차원 상의

부분 집합으로서 $1 \times N$ 크기의 데이터로 분할하여 연산하면, 데이터 처리(연산)를 N^2 번 진행해야 한다. 따라서, 기존의 2차원 메모리스트 어레이 구조로는 해당 학습을 위해 용이하지 않으므로, 이를 해결하고자, 새로운 개념의 추론 소자 구조 및 이를 활용한 연산 방법을 제시한다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0007] 일 실시예에 따른 메모리스트 어레이 장치는, 복수의 절연층들의 각각 및 복수의 하부 전극층들의 각각이 교대로 적층된 제1 적층 구조체, 및 상기 제1 적층 구조체에 형성된 홀(hole) 내에 저항변화소자 물질층, 중간 전극층, 선택소자 물질층, 및 상부 전극이 적층되고, 상기 복수의 하부 전극층들로부터 상기 상부 전극으로 향하는 전기 경로들에서 형성되는 복수의 메모리스트 소자들을 포함하는 제2 적층 구조체를 가지는, 수직형 메모리스트 기반 추론 소자를 포함할 수 있다.

[0008] 일 실시예에 따른 메모리스트 어레이 장치는, 입력 라인 및 출력 라인을 따라 어레이 구조로 배치된 복수의 메모리스트 기반 추론 소자들, 상기 복수의 메모리스트 기반 추론 소자들 중 동일한 출력 라인에 연결된 적어도 하나의

멤리스터 기반 추론 소자의 제2 적층 구조체들과 연결되는 연산 전극층, 상기 연산 전극층으로부터 전달된 컬럼 연산 결과와 다른 출력 라인의 다른 연산 전극층으로부터 전달된 컬럼 연산 결과에 기초한 출력 연산 결과를 출력하는 출력 전극층을 더 포함하고, 상기 복수의 멤리스터 기반 추론 소자들 중 한 소자가 상기 수직형 멤리스터 기반 추론 소자이며, 상기 연산 전극층은, 상기 적어도 하나의 멤리스터 기반 추론 소자의 각각에서 복수의 하부 전극층들을 통해 수신되는 입력들과 복수의 멤리스터 소자들에 저장된 가중치들 간의 개별 연산 결과를 수신하고, 상기 적어도 하나의 멤리스터 기반 추론 소자로부터 수신된 상기 개별 연산 결과를 병합함으로써 생성된 컬럼 연산 결과를 상기 출력 전극층에게 전달할 수 있다.

[0009] 일 실시예에 따른 멤리스터 어레이 장치는, 상기 복수의 멤리스터 기반 추론 소자들 중 같은 입력 라인에 연결된 하나 이상의 멤리스터 기반 추론 소자의 제2 적층 구조체가 같은 하부 전극층에 연결되는 멤리스터 어레이 장치일 수 있다.

[0010] 일 실시예에 따른 멤리스터 어레이 장치는, 상기 연산 전극층의 제1 연산에 기초한 출력을 저장하는 제1 버퍼, 및 상기 다른 연산 전극층의 제2 연산에 기초한 출력을 저장하는 제2 버퍼를 더 포함하고, 상기 출력 전극층은, 상기 제1 연산 및 상기 제2 연산이 개별적으로(individually) 수행되는 경우, 상기 제1 버퍼로부터 수신된 상기 제1 연산에 기초한 출력 및 상기 제2 버퍼로부터 수신된 상기 제2 연산에 기초한 출력을 통합한 결과를 출력하는 멤리스터 어레이 장치일 수 있다.

[0011] 일 실시예에 따른 멤리스터 어레이 장치에 있어서, 상기 복수의 멤리스터 기반 추론 소자들 중 각 멤리스터 기반 추론 소자는, 다른 멤리스터 기반 추론 소자들에 대해 병렬적으로, 입력 데이터로부터 미리 정해진 크기의 3차원 커널(kernel)에 대응하여 획득된 3차원 부분 데이터 중 해당하는 값들을 복수의 하부 전극층들을 통해 수신하고, 상기 다른 멤리스터 기반 추론 소자들에 대해 병렬적으로, 상기 해당하는 값들과 해당하는 멤리스터 기반 추론 소자에 형성된 복수의 멤리스터 소자들의 가중치들 간의 연산 결과를 생성할 수 있다.

[0012] 일 실시예에 따른 멤리스터 어레이 장치에 있어서, 상기 저항변화소자 물질층은, 금속 산화물 또는 칼코제나이드계 물질 중 적어도 하나를 포함하고, 상기 선택소자 물질층은, 칼코제나이드계 물질을 포함할 수 있다.

[0013] 일 실시예에 따른 멤리스터 어레이 장치는 상기 복수의 하부 전극층들에 입력 데이터를 입력하는 주변 회로(peripheral circuit)에 대해 적층될 수 있다.

[0014] 일 실시예에 따른 멤리스터 어레이 장치에 기초한 3차원 컨볼루션 신경망(3D Convolutional neural network) 연산 방법은, 입력 데이터로부터 미리 정해진 크기의 3차원 커널(kernel)에 대응하는 3차원 부분 데이터를 획득하는 단계, 상기 획득된 부분 데이터의 값들을, 수직으로 이격되어 형성된 복수의 멤리스터 소자들에 개별적으로 연결된 복수의 하부 전극층들에 입력하는 단계, 상기 복수의 하부 전극층에 입력된 값들과 상기 복수의 멤리스터 소자들에 저장된 가중치들 간의 개별 연산 결과들의 합산에 대응하는 컬럼 연산 결과를 생성하는 단계, 상기 생성된 컬럼 연산 결과 및 다른 출력 라인에 연결된 복수의 멤리스터 기반 추론 소자들의 컬럼 연산 결과를 출력 전극층에 전달하는 단계, 및 상기 전달된 컬럼 연산 결과들을 병합하여, 출력 연산 결과를 출력하는 단계를 포함할 수 있다.

[0015] 일 실시예에 따른 3차원 컨볼루션 신경망 연산 방법에 있어서, 상기 멤리스터 어레이 장치가, 입력 라인 및 출력 라인을 따라 어레이 구조로 배치된 복수의 멤리스터 기반 추론 소자들을 포함하고, 상기 입력하는 단계는, 상기 3차원 부분 데이터를 복수의 멤리스터 소자들 각각에 포함된 복수의 하부 전극층들에 병렬적으로 입력하는 단계를 포함할 수 있다.

[0016] 일 실시예에 따른 3차원 컨볼루션 신경망 연산 방법에 있어서, 상기 멤리스터 어레이 장치가, 입력 라인 및 출력 라인을 따라 어레이 구조로 배치된 복수의 멤리스터 기반 추론 소자들을 포함하고, 상기 복수의 멤리스터 기반 추론 소자들 중 같은 입력 라인에 연결된 하나 이상의 멤리스터 기반 추론 소자의 제2 적층 구조체가 같은 하부 전극층에 연결되며, 상기 입력하는 단계는, 상기 3차원 부분 데이터를 슬라이스(slice)함으로써 복수의 2차원 데이터를 형성하는 단계, 및 상기 형성된 복수의 2차원 데이터 중 하나의 데이터에 해당하는 값을, 출력 라인들 중 해당 출력 라인에 연결된 복수의 멤리스터 기반 추론 소자들의 해당 복수의 하부 전극층들에 입력하는 단계를 포함할 수 있다.

[0017] 일 실시예에 따른 3차원 컨볼루션 신경망 연산 방법은, 상기 생성된 컬럼 연산 결과를 버퍼에 저장하는 단계를 더 포함할 수 있다.

[0018] 일 실시예에 따른 3차원 컨볼루션 신경망 연산 방법에 있어서, 상기 멤리스터 어레이 장치는, 상기 복수의 하부 전극층들에 입력 데이터를 입력하는 주변 회로(peripheral circuit)에 대해 적층되고, 상기 입력하는 단계는,

상기 주변회로가 상기 복수의 하부 전극층들에 데이터를 입력하는 단계를 포함할 수 있다.

도면의 간단한 설명

- [0019] 도 1a는 일 실시예에 따른 수직형 메모리스터 기반 추론 소자(100)의 단면도를 도시한다.
- 도 1b 및 도 1c는 전기적 펄스가 가해진 경우의 펄스 응답 특성(101) 및 메모리스터 소자의 시간에 따른 저항 변화 특성(102)을 도시한다.
- 도 2는 일 실시예에 따른 메모리스터 어레이 장치(200)를 도시한다.
- 도 3은 일 실시예에 따른 메모리스터 어레이 장치(300)를 도시한다.
- 도 4는 일 실시예에 따른 메모리스터 어레이 장치(400)를 도시한다.
- 도5는 일 실시예에 따른 메모리스터 어레이 장치(500)를 도시한다.
- 도 6은 주변 회로(peripheral circuit)와 메모리스터 어레이 장치의 위치 관계를 비교 실시예와 비교하는 도면이다.
- 도 7a 내지 도 7b는 본 개시에 따른 메모리스터 어레이 장치에 기초한 3차원 컨볼루션 신경망 연산 방법에 대한 흐름도(701) 및 개략도(700)를 나타낸다.
- 도 8a 내지 도 8c는 메모리스터 어레이 장치에 n^2 의 2차원 데이터가 순차적으로 입력되는 경우, 3D CNN 연산을 위한 메모리스터 어레이 장치 작동 방법을 도시한다.

발명을 실시하기 위한 구체적인 내용

- [0020] 실시예들에 대한 특정한 구조적 또는 기능적 설명들은 단지 예시를 위한 목적으로 개시된 것으로서, 다양한 형태로 변경되어 구현될 수 있다. 따라서, 실제 구현되는 형태는 개시된 특정 실시예로만 한정되는 것이 아니며, 본 명세서의 범위는 실시예들로 설명한 기술적 사상에 포함되는 변경, 균등물, 또는 대체물을 포함한다.
- [0021] 제1 또는 제2 등의 용어를 다양한 구성요소들을 설명하는데 사용될 수 있지만, 이런 용어들은 하나의 구성요소를 다른 구성요소로부터 구별하는 목적으로만 해석되어야 한다. 예를 들어, 제1 구성요소는 제2 구성요소로 명명될 수 있고, 유사하게 제2 구성요소는 제1 구성요소로도 명명될 수 있다.
- [0022] 어떤 구성요소가 다른 구성요소에 "연결되어" 있다고 언급된 때에는, 그 다른 구성요소에 직접적으로 연결되어 있거나 또는 접속되어 있을 수도 있지만, 중간에 다른 구성요소가 존재할 수도 있다고 이해되어야 할 것이다.
- [0023] 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다. 본 명세서에서, "포함하다" 또는 "가지다" 등의 용어는 설명된 특징, 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것이 존재함으로써 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부분품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0024] 다르게 정의되지 않는 한, 기술적이거나 과학적인 용어를 포함해서 여기서 사용되는 모든 용어들은 해당 기술 분야에서 통상의 지식을 가진 자에 의해 일반적으로 이해되는 것과 동일한 의미를 가진다. 일반적으로 사용되는 사전에 정의되어 있는 것과 같은 용어들은 관련 기술의 문맥상 가지는 의미와 일치하는 의미를 갖는 것으로 해석되어야 하며, 본 명세서에서 명백하게 정의하지 않는 한, 이상적이거나 과도하게 형식적인 의미로 해석되지 않는다.
- [0025] 이하, 실시예들을 첨부된 도면들을 참조하여 상세하게 설명한다. 첨부 도면을 참조하여 설명함에 있어, 도면 부호에 관계없이 동일한 구성 요소는 동일한 참조 부호를 부여하고, 이에 대한 중복되는 설명은 생략하기로 한다.
- [0026] 도 1a는 일 실시예에 따른 수직형 메모리스터 기반 추론 소자(100)의 단면도를 도시한다.
- [0027] 수직형 메모리스터 기반 추론 소자(100)는 본 개시에 따른 메모리스터 어레이 장치를 구성하는 단일 소자 중 하나일 수 있다. 수직형 메모리스터 기반 추론 소자(100)는 제1 적층 구조체(110) 및 제1 적층 구조체(110)의 홀(113)에 형성된 제2 적층 구조체(114)를 포함할 수 있다.
- [0028] 제1 적층 구조체(110)는 복수의 절연층들(111)의 각각 및 복수의 하부 전극층들(112)의 각각이 교대로 적층된

구조일 수 있다. 복수의 하부 전극층들(112)은, 복수의 하부 전극층들(112)의 일부분이 외부로부터 입력 데이터 값을 수신할 수 있도록 제1 적층 구조체(110)의 외부로 노출된 구조일 수 있다. 예를 들어, 복수의 하부 전극층(112)은 제1 적층 구조체(110)의 상면을 향해 이격되어 노출된 구조일 수 있으나, 복수의 하부 전극층(112)의 구조는 이에 한정되지 않는다. 복수의 절연층들(111)은 복수의 하부 전극층들(112) 각각 간의 전기적 간섭, 상호작용, 또는 전하의 유출을 방지할 수 있다.

[0029] 제2 적층 구조체(114)는 제1 적층 구조체(112)에 형성된 홀(hole)(113) 내에 저항변화소자 물질층(115), 중간 전극층, 선택소자 물질층(116), 및 상부 전극(117)이 적층된 구조일 수 있다. 제1 적층 구조체(112)에 형성된 홀(113)은 에칭(etching) 프로세스를 통해 형성될 수 있다. 예를 들어, 에칭 프로세스는 건식 식각(dry etching) 또는 습식 식각(wet etching) 프로세스일 수 있다. 구체적으로, 건식 식각 프로세스는 유도성 결합 플라즈마 장치(inductively coupled plasma reactive ion etching, ICP-RIE) 또는 용량성 결합 플라즈마 장치(capacitively coupled plasma reactive ion etching, CCP-RIE)를 활용하여 진행될 수 있다. 에칭 프로세스에서 플라즈마 형성을 위해 Ar 가스 및/또는 화학적 반응을 위한 CF_4 , CHF_3 , O_2 , SF_6 , Cl_2 또는 HBr 가스 등이 활용될 수 있다. 이상 에칭 프로세스에 대한 구체적인 예시는 일 실시예 일뿐, 홀(hole) 형성을 위한 에칭 프로세스는 이에 제한되지 않는다. 일 실시예에 따르면, 저항변화소자 물질층(115)은, 홀(113) 내부 표면에 박막 형태로 증착 또는 도포 될 수 있으나, 제조 공정은 이에 제한되지 않는다. 저항변화소자 물질층(115)은, 금속 산화물 또는 칼코제나이드계 물질 중 적어도 하나를 포함할 수 있다. 예를 들어, 저항변화소자 물질층(115)은 산화물 기반의 저항성 램(Resistance Random Access Memory, ReRAM) 및 칼코제나이드계 물질 기반의 상변화 램(Phase-Change Random Access Memory, PCRAM)의 물질군들 중 적어도 하나를 포함할 수 있다.

[0030] 저항변화소자 물질층(115)이 홀(113) 내부에 형성된 후, 중간 전극층 및 선택소자 물질층(116)이 적층 형성될 수 있다. 선택소자 물질층(116)은 칼코제나이드계 물질을 포함할 수 있다. 예를 들어, 선택소자 물질층(116)은 칼코제나이드계 물질 기반의 오보닉 임계값 스위칭 선택기(Ovonic-threshold switching selector, OTS)의 물질군들을 포함할 수 있다. 선택소자 물질층(116)에 미리 설정된 임계치 이상의 입력이 인가되는 경우, 제2 적층 구조체(114)에 포함된 복수의 메모리스터 소자들(118)이 동작될 수 있다. 예를 들어, 선택소자 물질층(116)에 임계치 이상의 전압이 인가됨으로써, 제2 적층 구조체(114)에 포함된 메모리스터 소자들(118)이 동작될 수 있다. 보다 구체적으로, 선택소자 물질층(116)에 임계치 이상의 특정 전압이 인가됨으로써, 메모리스터 소자들(118)의 동작이 선택적으로 활성화 또는 비활성화 될 수 있다. 또한, 선택소자 물질층(116)에 임계치 이상의 전압이 인가됨으로써, 메모리스터 소자들(118) 각각에 저장된 가중치에 대응되는 서로 다른 저항 값이 쉽게 읽힐(reading) 수 있다.

[0031] 도 1a에 도시된 바와 같이 선택소자 물질층(116)은 메모리스터 소자(118)에 포함될 수 있으나, 별도의 소자로서 메모리스터 소자(118)와 연결될 수 있다.

[0032] 제2 적층 구조체(114)는 복수의 하부 전극층들(111)로부터 상부 전극(117)으로 향하는 전기 경로들에서 형성되는 복수의 메모리스터 소자들(118)을 포함할 수 있다. 예를 들어, 복수의 하부 전극층들(111) 각각 및 상부 전극(117) 사이에, 저항변화소자 물질층(115), 중간 전극층, 및 선택소자 물질층(116)이 순차적으로 형성될 수 있고, 복수의 하부 전극층들(111)에서 저항변화소자 물질층(115), 중간 전극층, 및 선택소자 물질층(116)을 거쳐 상부 전극(117)으로 향하는 전기 경로들 상에 복수의 메모리스터 소자들(118)이 형성될 수 있다. 보다 구체적으로, 복수의 하부 전극층들(111) 각각과 상부 전극(117) 사이에, $4F^2$ 면적의 복수의 메모리스터 소자들(118)이 형성될 수 있다. 복수의 메모리스터 소자들(118)은 전기적 펄스에 의한 저항변화소자 물질층(115)의 상변화를 야기하거나 전기적 통로(conductive path)의 생성을 야기함으로써 저항 값 또는 임피던스 값을 형성할 수 있다. 복수의 메모리스터 소자들(118)의 저항 값 또는 임피던스 값은 3D CNN 연산을 위해 설정된 가중치(weight)에 대응하는 값일 수 있다. 구체적인 메모리스터 소자(118)에 의한 가중치 설정 방법은 이하, 도 1b 및 도 1c의 전기적 펄스 응답 특성(101) 및 저항 변화 특성(102)에서 자세히 설명한다.

[0033] 수직형 메모리스터 기반 추론 소자(100)는 복수의 하부 전극층(112)을 통해 입력 데이터 값을 수신할 수 있다. 입력 데이터 값은 복수의 메모리스터 소자들(118)에 설정된 가중치들에 기초하여 연산될 수 있고, 복수의 메모리스터 소자들(118)에 의해 연산된 결과 값은 상부 전극(117)에서 병합될 수 있다.

[0034] 도 1b 및 도 1c는 전기적 펄스가 가해진 경우의 펄스 응답 특성(101) 및 메모리스터 소자의 시간에 따른 저항 변화 특성(102)을 도시한다.

[0035] 전기적 펄스 응답 특성(101)에 따르면, 전기적 펄스의 수가 증가함에 따라, 정규화된 컨덕턴스(normalized

conductance)의 값이 감소하는 특성을 나타낼 수 있다. 그러나, 전기적 펄스의 조건 변화에 따른 정규화된 컨덕턴스의 값의 변화 특성 방향은 이에 한정되지 않으며, 증가 혹은 그 외의 점진적 변화 특성을 나타낼 수 있다.

[0036] 저항 변화 특성(102)에 따르면, 메모리스트 소자는 전기적 펄스의 세기에 따라 서로 다른 저항 값(R)을 나타낼 수 있다. 예를 들어, 저항 변화 특성(102)에서 서로 다른 저항 값은 x축인 시간축에서 x값이 0일 때, y축인 저항 축 상의 서로 다른 8개의 y값에 대응될 수 있다. 저항 변화 특성(102)에 따르면, 메모리스트 소자(118)는 전기적

펄스에 따라 형성된 저항 값을 시간이 지남에 따라 일정 임계 범위 내에서 유지할 수 있다. 예를 들어, I_{set} 는

시간에 따른 저항의 변화 정도를 나타내는 상수 값에 대응될 수 있다. 시간에 따른 저항의 변화 정도(I_{set})가 10^4 초의 시간동안 최대 0.0890에 불과하므로, 서로 다른 저항 값들은 시간의 경과에도 서로 중복되지 않는 범위 내에서 유지될 수 있다. 저항 값들이 시간이 지남에 따라 서로 다른 일정 임계 범위 내에서 유지됨으로써, 메모리스트 소자(118) 간 저항 값이 서로 다른 범위로 구별될 수 있다. 서로 구별되는 복수의 저항 값은, 메모리스트 소자(118) 각각의 서로 다른 가중치 값에 대응될 수 있다. 따라서, 복수의 메모리스트 소자(118)에 서로 다른 세기의 전기적 펄스를 가하여, 각 소자 별 지정된 가중치(weight) 값에 대응하는 저항 값을 개별적으로 설정할 수 있다.

[0037] 도 2는 일 실시예에 따른 메모리스트 어레이 장치(200)를 도시한다.

[0038] 메모리스트 어레이 장치(200)는 입력 라인(210) 및 출력 라인(211)을 따라 어레이 구조로 배치된 복수의 메모리스트 기반 추론 소자들(201)을 포함할 수 있다. 메모리스트 어레이 장치(200)에 포함된 복수의 메모리스트 기반 추론 소자들(201) 중 한 소자가 도 1에 도시된 수직형 메모리스트 기반 추론 소자일 수 있다. 메모리스트 어레이 장치(200)는 입력 라인(210)의 수 및 출력 라인(211)의 수가 동일할 수 있고, 입력 라인(210) 및 출력 라인(211)을 따라 배치된 복수의 메모리스트 기반 추론 소자들(201)의 수가 동일할 수 있다. 다른 예를 들어, 입력 라인(210)의 수와 출력 라인(211)의 수가 서로 다를 수 있고, 입력 라인(210)을 따라 배치된 복수의 메모리스트 기반 추론 소자들의 수와 출력 라인(211)을 따라 배치된 복수의 메모리스트 기반 추론 소자들의 수가 서로 다를 수 있다. 예를 들어, 도 2에 도시된 메모리스트 어레이 장치는 입력 라인(210)을 따라 n개 및 출력 라인(211)을 따라 n개의 메모리스트 기반 추론 소자가 배치된 것을 나타낸다. 예를 들어, 메모리스트 기반 추론 소자들(201) 각각에 포함된 하부 전극층의 수가 n개인 경우, 메모리스트 어레이 장치(200)에 n^3 개의 복수의 메모리스트 소자들(240)이 형성될 수 있다. n^3 개의 복수의 메모리스트 소자들(240)로 인해 n^3 개의 가중치가 정의될 수 있다. 이하, 설명의 편의를 위해 입력 라인(210) 및 출력 라인(211)을 따라 배치된 복수의 메모리스트 기반 추론 소자들(201)의 수가 동일한 경우를 기준으로 설명한다.

[0039] 메모리스트 어레이 장치(200)는 복수의 메모리스트 기반 추론 소자들(201) 중 동일한 출력 라인(211)에 연결된 적어도 하나의 메모리스트 기반 추론 소자의 제2 적층 구조체들과 연결되는 연산 전극층(220)을 더 포함할 수 있다. 연산 전극층(220)은 적어도 하나의 메모리스트 기반 추론 소자의 각각에서 복수의 하부 전극층들을 통해 수신되는 입력들과 복수의 메모리스트 소자들에 저장된 가중치들 간의 개별 연산 결과를 수신하고, 적어도 하나의 메모리스트 기반 추론 소자로부터 수신된 개별 연산 결과를 병합함으로써 생성된 컬럼 연산 결과를 출력 전극층에게 전달할 수 있다. 예를 들어, 메모리스트 어레이 장치(200)에 포함된 복수의 메모리스트 기반 추론 소자들 중 하나의 소자는 복수의 하부 전극층들(230)을 통해 입력을 수신할 수 있다. 수신된 입력 각각과 복수의 메모리스트 소자들(240)에 저장된 가중치 각각에 대해 개별 연산이 수행될 수 있다. 예를 들어, 복수의 하부 전극층들(230)을 통해 전압이 입력된 경우, 메모리스트 소자들(240) 각각의 저항 값에 기초하여 전류 값이 출력될 수 있다. 동일한 출력 라인에 배치된 복수의 메모리스트 기반 추론 소자들에 의해 출력된 전류 값의 합산 결과가 컬럼 연산 결과에 대응될 수 있다.

[0040] 메모리스트 어레이 장치(200)는 연산 전극층(220)으로부터 전달된 컬럼 연산 결과와 다른 출력 라인의 다른 연산 전극층으로부터 전달된 컬럼 연산 결과에 기초한 출력 연산 결과를 출력하는 출력 전극층(221)을 더 포함할 수 있다. 출력 전극층(221)은 복수의 컬럼 연산 결과들의 합산 결과에 대응하는 출력 연산 결과를 출력할 수 있다. 예를 들어, 컬럼 연산 결과가 전류 값에 대응되는 경우, 출력 연산 결과는 모든 컬럼 연산 결과와 대응되는 전류 값을 병합한 전류 값에 대응될 수 있다.

[0041] 도 3은 일 실시예에 따른 메모리스트 어레이 장치(300)를 도시한다.

- [0042] 메모리스트 어레이 장치(300)는 복수의 메모리스트 기반 추론 소자들 중 같은 입력 라인(310)에 연결된 하나 이상의 메모리스트 기반 추론 소자의 제2 적층 구조체들(320)이 같은 하부 전극층(321)에 연결될 수 있다. 예를 들어, 메모리스트 어레이 장치(300)는 입력 라인(310)을 따라 하부 전극층(321)이 연장되고, 입력 라인(310)을 따라 이격 배치된 제2 적층 구조체들(320)이 연장된 하부 전극층(321)을 공유하는 구조일 수 있다. 보다 구체적으로, 메모리스트 어레이 장치(300)는 입력 라인(310) 각각에 대해 n 개의 하부 전극층(321)을 포함하는 제1 적층 구조체 n 개가 배치된 것을 나타낸다. 또한, 입력 라인(310)을 따라 n 개 및 출력 라인(311)을 따라 n 개의 제2 적층 구조체가 각각 배치된 것을 나타낸다. 제1 적층 구조체 각각에 포함된 하부 전극층의 수가 n 개이므로, 메모리스트 어레이 장치(300)에 n^2 크기의 데이터가 입력될 수 있다. 다른 예를 들어, 입력 데이터의 크기가 n^3 인 경우, 메모리스트 어레이 장치(300)에 n^2 크기의 데이터가 n 회 반복하여 순차적으로 입력될 수 있다. 또한, 제2 적층 구조체들(320) 각각과 n 개의 하부 전극층들에 의해 n^3 개의 복수의 메모리스트 소자들이 형성될 수 있다. n^3 개의 복수의 메모리스트 소자들로 인해 n^3 개의 가중치가 정의될 수 있다. 메모리스트 어레이 장치(300)에 n^2 크기의 데이터가 n 회 반복하여 순차적으로 입력되는 경우의 메모리스트 어레이 장치 동작 방법은 추후 도 8a 내지 8c에서 상세히 설명한다.
- [0043] 도 4는 일 실시예에 따른 메모리스트 어레이 장치(400)를 도시한다.
- [0044] 메모리스트 어레이 장치(400)는 n^2 개의 하부 전극층을 포함할 수 있다. 메모리스트 어레이 장치(400)에 포함된 n^2 개의 하부 전극층에 n^2 크기의 2차원 배열의 데이터가 입력될 수 있다. 입력 데이터의 크기가 n^3 인 경우, 메모리스트 어레이 장치(400)에 n^2 크기의 데이터가 n 회 반복하여 순차적으로 입력될 수 있다. 메모리스트 어레이 장치(400)는 n^2 크기의 입력 데이터 각각과 메모리스트 소자들에 저장된 가중치들 간의 개별 연산 결과들의 합산에 대응하는 컬럼 연산 결과를 생성하기 위해 연산이 수행될 출력 라인을 선택할 수 있다. n^2 크기의 데이터가 n 회 반복하여 메모리스트 어레이 장치(400)에 순차적으로 입력되는 경우, 입력 순서에 따른 연산 수행 시간 차가 발생할 수 있다. 메모리스트 어레이 장치(400)는 n^3 크기의 입력 데이터에 대한 하나의 출력 연산 결과를 생성함에 있어서, n^2 크기의 데이터의 순차적 입력으로 인한 연산 수행 시간 차를 상쇄하기 위해 각 출력 라인 상에 버퍼를 포함할 수 있다. 구체적으로, 메모리스트 어레이 장치(400)는 연산 수행 시간 차를 상쇄하기 위해 컬럼 연산 결과를 버퍼에 저장할 수 있다. 버퍼에 저장된 컬럼 연산 결과는 다른 컬럼 연산 결과가 모두 완료된 경우, 버퍼에서 출력 연산층으로 전달될 수 있다.
- [0045] 예를 들어, 메모리스트 어레이 장치(400)는 연산 전극층의 제1 연산에 기초한 출력을 저장하는 제1 버퍼(410) 및 다른 연산 전극층의 제2 연산에 기초한 출력을 저장하는 제2 버퍼(411)를 더 포함할 수 있다.
- [0046] 제1 버퍼(410)는 하나의 출력 라인 상에 배치된 연산 전극층의 제1 연산에 기초한 출력을 저장할 수 있다. 연산 전극층의 제1 연산은 같은 출력 라인 상에 배치된 메모리스트 기반 추론 소자들 각각의 복수의 하부 전극층들을 통해 수신된 입력 값과 메모리스트 소자들에 정의된 가중치에 기초한 가중합 연산에 대응될 수 있다.
- [0047] 제2버퍼(411)는 제1 버퍼(410)가 배치된 출력 라인과는 다른 출력 라인 상에 배치된 연산 전극층의 제2 연산에 기초한 출력을 저장할 수 있다. 제2 연산은 제1 연산을 출력하는 연산 전극층과는 다른 연산 전극층에서 수행될 수 있다.
- [0048] 도 4에 도시된 바와 같이, 메모리스트 어레이 장치(400)의 출력 라인이 n 개이고, 각 출력 라인에 메모리스트 기반 추론 소자가 배치된 경우, 메모리스트 어레이 장치(400)는 총 n 개의 버퍼를 포함할 수 있다. 제 n 버퍼(412)는 연산 전극층의 제 n 연산에 기초한 출력을 저장할 수 있다.
- [0049] 메모리스트 어레이 장치(400)의 출력 전극층(413)은, 제1 연산 및 제2 연산이 개별적으로(individually) 수행되는 경우, 제1 버퍼(410)로부터 수신된 제1 연산에 기초한 출력 및 제2 버퍼(411)로부터 수신된 제2 연산에 기초한 출력을 통합한 결과를 출력할 수 있다. 보다 구체적으로, 메모리스트 어레이 장치(400)의 출력 전극층(413)은 제1 연산, 제2 연산부터 제 n 연산까지 총 n 개의 연산이 개별적으로 수행되는 경우, 제1 연산, 제2 연산부터 제 n 연산까지 총 n 개의 연산에 기초한 출력을 통합한 결과를 출력할 수 있고, 통합된 결과는 전류 값에 대응될 수 있다.
- [0050] 도5는 일 실시예에 따른 메모리스트 어레이 장치(500)를 도시한다.

- [0051] 메모리스트 어레이 장치(500)에 포함된 복수의 메모리스트 기반 추론 소자들 중 각 메모리스트 기반 추론 소자는 다른 메모리스트 기반 추론 소자들에 대해 병렬적으로, 입력 데이터로부터 미리 정해진 크기의 3차원 커널(kernel)에 대응하여 획득된 3차원 부분 데이터 중 해당하는 값들을 복수의 하부 전극 층들(501, 502, 503)을 통해 수신할 수 있다. 예를 들어, 미리 정해진 크기의 3차원 커널의 크기는 출력 라인의 수를 가로로 길이로 하고, 입력 라인의 수를 세로 길이로 하며, 복수의 하부 전극층의 수를 높이로 하여, 가로 $\times$ 세로 $\times$ 높이의 크기일 수 있다. 3차원 부분 데이터는 3차원 커널에 대응하여 입력 데이터로부터 분할 또는 추출된 데이터에 대응될 수 있다. 예를 들어, 3차원 부분 데이터는 $n \times n \times n$ 의 크기일 수 있다. 또 다른 예를 들어, 3차원 부분 데이터 가로, 세로, 높이가 서로 다른 $n \times l \times k$ 의 크기일 수 있다. 메모리스트 기반 추론 소자들 각각에 포함된 복수의 하부 전극층의 수가 n 개이고, 전체 복수의 하부 전극층들(501, 502, 503)의 수가 n^3 개인 경우, 메모리스트 어레이 장치(200)에 n^3 개의 데이터가 병렬적으로 입력될 수 있다.
- [0052] 메모리스트 어레이 장치(500)는 병렬적으로 입력된 데이터에 기초하여, 각 메모리스트 기반 추론 소자들이 다른 메모리스트 기반 추론 소자들에 대해 병렬적으로, 해당하는 값들과 해당하는 메모리스트 기반 추론 소자에 형성된 복수의 메모리스트 소자들(510, 511, 512)의 가중치들 간의 연산 결과를 생성할 수 있다.
- [0053] 도 6은 주변 회로(peripheral circuit)와 메모리스트 어레이 장치의 위치 관계를 비교 실시예와 비교하는 도면이다.
- [0054] 주변 회로(peripheral circuit)는 메모리스트 어레이 장치에 포함될 수 있고, 메모리스트 어레이 장치와 별도의 회로로서, 메모리스트 어레이 장치에 입력 데이터에 해당하는 값을 공급할 수 있다.
- [0055] 비교 실시예(601)는 메모리스트 어레이 장치(610)에 입력 데이터에 해당하는 값을 공급하기 위한 별도의 주변 회로(611)를 포함한다. 예를 들어, 메모리스트 어레이 장치(610)가 도 2에 도시된 메모리스트 어레이 장치(200)에 대응되는 경우, 메모리스트 어레이 장치(610)는 총 n^3 개의 복수의 하부 전극 층들을 포함할 수 있다. 하부 전극 층들이 총 n^3 개인 메모리스트 어레이 장치(610)에 입력 데이터에 해당하는 전압을 인가하기 위해, 총 n^3 개의 전극을 포함하는 주변 회로(611)를 필요로 할 수 있다. 주변 회로(611)를 메모리스트 어레이 장치(610)와 별개의 구성으로 포함하면 주변 회로에 의한 부피 차이가 증가한다는 단점이 있다.
- [0056] 주변 회로 적층 구조(602)는 메모리스트 어레이 장치(620)가 복수의 하부 전극층들에 입력 데이터를 입력하는 주변 회로(621)에 대해 적층되는 구조를 도시한다. 예를 들어, 메모리스트 어레이 장치(620)가 도 2에 도시된 메모리스트 어레이 장치(200)에 대응되는 경우, 메모리스트 어레이 장치(610)는 총 n^3 개의 복수의 하부 전극 층들을 포함할 수 있다. 총 n^3 개의 복수의 하부 전극 층들을 포함하는 메모리스트 어레이 장치(610)에 입력 데이터에 해당하는 전압을 인가하기 위해, 주변 회로(621)는 메모리스트 어레이 장치(610)와 적층되어 주변 회로에(621)에 의한 부피 차이가 감소될 수 있다.
- [0057] 도 7a 내지 도 7b는 본 개시에 따른 메모리스트 어레이 장치에 기초한 3차원 컨볼루션 신경망 연산 방법에 대한 흐름도(701) 및 개략도(700)를 나타낸다.
- [0058] 단계(720)에서, 메모리스트 어레이 장치(713)는 입력 데이터(710)로부터 미리 정해진 크기의 3차원 커널(kernel)(711)에 대응하는 3차원 부분 데이터(712)를 획득할 수 있다. 예를 들어, 입력 데이터(710)의 크기는 m^3 , 3차원 커널(711)의 크기는 n^3 일 수 있고, $m > n$ 으로 정의될 수 있다. 또한, 3차원 커널(711)의 크기는 메모리스트 어레이 장치(713)에 포함된 메모리스트 소자의 수에 대응될 수 있다.
- [0059] 단계(721)에서, 메모리스트 어레이 장치(713)는 획득된 부분 데이터(712)의 값들을, 메모리스트 어레이 장치(713)에 포함된, 수직으로 이격되어 형성된 복수의 메모리스트 소자들에 개별적으로 연결된 복수의 하부 전극층들에 입력할 수 있다. 부분 데이터(712)의 입력 방법에 대하여 이하 도 8a 내지 도 8c에서 구체적으로 설명한다.
- [0060] 단계(722)에서, 메모리스트 어레이 장치(713)는 복수의 하부 전극층에 입력된 값들과 복수의 메모리스트 소자들에 저장된 가중치들 간의 개별 연산 결과들의 합산에 대응하는 컬럼 연산 결과를 생성할 수 있다. 예를 들어, 복수의 하부 전극층에 입력된 값들이 전압인 경우, 복수의 메모리스트 소자들에 저장된 가중치들은 메모리스트 소자들

의 저항 값에 대응될 수 있다. 따라서, 개별 연산 결과들 및 개별 연산 결과들의 합산에 대응하는 컬럼 연산 결과는 복수의 메모리스터 소자 저항 값들과 복수의 하부 전극층에 입력된 전압 값에 기초한 전류 값에 대응될 수 있다.

- [0061] 단계(723)에서, 메모리스터 어레이 장치(713)는 생성된 컬럼 연산 결과 및 다른 출력 라인에 연결된 복수의 메모리스터 기반 추론 소자들의 컬럼 연산 결과를 출력 전극층에 전달할 수 있다. 컬럼 연산 결과 및 다른 출력 라인에서 획득된 컬럼 연산 결과는 출력 전극층으로 동시에 또는 시계열 적으로 전달될 수 있다.
- [0062] 단계(724)에서, 메모리스터 어레이 장치(713)는, 전달된 컬럼 연산 결과들을 병합하여, 출력 연산 결과(714)를 출력할 수 있다. 출력 연산 결과(714)는 n^3 크기의 부분 데이터(712)에 대한 연산 결과에 대응하는 단일 전류 값에 해당할 수 있다.
- [0063] 입력 데이터(710)에 대하여, 3차원 커널(711)이 스위핑(sweeping)되고, 3차원 커널(711)에 의해 획득된 3차원 부분 데이터(712)가 메모리스터 어레이 장치(713)에 입력되면서 단계(720) 내지 단계(724)가 반복될 수 있다. 또한, 메모리스터 어레이 장치(713)는 단계(720) 내지 단계(740)을 반복함으로써, $(m-n+1)^3$ 개의 출력 연산 결과(714)를 출력하고, $(m-n+1)^3$ 크기의 출력 연산 결과들(715)을 생성할 수 있다.
- [0065] 도 8a 내지 도 8c는 메모리스터 어레이 장치에 n^2 의 2차원 데이터가 순차적으로 입력되는 경우, 3D CNN 연산을 위한 메모리스터 어레이 장치 작동 방법을 도시한다.
- [0066] 예를 들어, 메모리스터 어레이 장치가, 입력 라인 및 출력 라인을 따라 어레이 구조로 배치된 복수의 메모리스터 기반 추론 소자들을 포함하고, 복수의 메모리스터 기반 추론 소자들 중 같은 입력 라인에 연결된 하나 이상의 메모리스터 기반 추론 소자의 제2 적층 구조체가 같은 하부 전극층에 연결될 수 있다. 이 때, 메모리스터 어레이 장치는 3차원 부분 데이터(800)를 슬라이스(slice)함으로써 복수의 2차원 데이터를 형성할 수 있다. 예를 들어, 형성된 복수의 2차원 데이터는 n^2 크기일 수 있고, 제1 이차원 데이터부터 제n 이차원 데이터까지 총 n개의 슬라이스로 정의될 수 있다.
- [0067] 메모리스터 어레이 장치는, 형성된 복수의 2차원 데이터 중 하나의 데이터에 해당하는 값을, 출력 라인들 중 해당 출력 라인에 연결된 복수의 메모리스터 기반 추론 소자들의 해당 복수의 하부 전극층들에 입력할 수 있다.
- [0068] 예를 들어, 도 8a는 부분 데이터(800) 중 제1 이차원 데이터(801a)가 메모리스터 어레이 장치에 입력된 경우, 메모리스터 어레이 장치의 동작 방법을 도시한다. 제1 이차원 데이터(801a)는 복수의 하부 전극층(810)에 입력될 수 있다. 메모리스터 어레이 장치는, 제1 출력 라인(811a)에 배치된 제1 연산 전극층(812a)(또는 상부 전극)을 선택할 수 있다. 선택된 제1 연산 전극층(812a)은 제1 이차원 데이터(801a) 및 복수의 메모리스터 소자에 저장된 가중치들 간의 n^2 개의 개별 연산 결과를 병합하여 제1 컬럼 연산 결과로 생성할 수 있다. 제1 컬럼 연산 결과는 출력 전극층(820)으로 전달될 수 있다. 도 8a에는 도시되지 않았으나, 제1 출력 라인(811a)에 배치된 제1 연산 전극층(812a)과 출력 전극층(820) 사이에 버퍼가 포함될 수 있다. 제1 연산 전극층(812a)은 이하 다른 연산 전극층에서의 컬럼 연산 결과가 생성되기 전까지 제1 컬럼 연산 결과를 버퍼에 저장할 수 있다.
- [0069] 도 8b는 부분 데이터(800) 중 제1 이차원 데이터가 입력된 후, 제2 이차원 데이터(801b)가 메모리스터 어레이 장치에 입력되는 경우, 메모리스터 어레이 장치의 동작 방법을 도시한다. 제2 이차원 데이터(801b)는 복수의 하부 전극층(810)에 입력될 수 있다. 메모리스터 어레이 장치는, 제2 출력 라인(811b)에 배치된 제2 연산 전극층(812b)(또는, 상부 전극)을 선택할 수 있다. 선택된 제2 연산 전극층(812b)은 제2 이차원 데이터(801b) 및 복수의 메모리스터 소자에 저장된 가중치들 간의 n^2 개의 개별 연산 결과를 병합하여 제2 컬럼 연산 결과로 생성할 수 있다. 제2 컬럼 연산 결과는 출력 전극층(820)으로 전달될 수 있다. 도 8b에 도시되지 않았으나, 제2 출력 라인(811b)에 배치된 제2 연산 전극층(812b)과 출력 전극층(820) 사이에 버퍼가 포함될 수 있다. 제2 연산 전극층(812b)은 이하 다른 연산 전극층에서의 컬럼 연산 결과가 생성되기 전까지 제2 컬럼 연산 결과를 버퍼에 저장할 수 있다.
- [0070] 도 8c는 부분 데이터(800) 중 제1 및 제2 이차원 데이터가 입력된 후, 제n 이차원 데이터(801c)가 메모리스터 어레이 장치에 입력되는 경우, 메모리스터 어레이 장치의 동작 방법을 도시한다. 제n 이차원 데이터(801c)는 복수의 하부 전극층(810)에 입력될 수 있다. 메모리스터 어레이 장치는, 제n 출력 라인(811c)에 배치된 제n 연산 전극층(812c)(또는, 상부전극)을 선택할 수 있다. 선택된 제n 연산 전극층(812c)은 제n 이차원 데이터(801c) 및

복수의 맵리스터 소자에 저장된 가중치들 간의 n^2 개의 개별 연산 결과를 병합하여 제n 컬럼 연산 결과로 생성할 수 있다. 제n 컬럼 연산 결과는 출력 전극층(820)으로 전달될 수 있다. 도 8c에 도시되지 않았으나, 제n 출력 라인(811c)에 배치된 제n 연산 전극층(812c)과 출력 전극층(820) 사이에 버퍼가 포함될 수 있다. 제n 연산 전극층(812c)은 제n 컬럼 연산 결과를 버퍼에 저장할 수 있다.

[0071] 맵리스트 어레이 상치는 제1 부터 제n 컬럼 연산 결과가 각 버퍼(미도시) 모두에 저장된 경우, 버퍼에 저장된 컬럼 연산 결과들을 동시에 출력 전극층(820)으로 전달할 수 있다. 출력 전극층(820)은 전달 받은 복수의 컬럼 연산 결과들을 병합하여 출력 연산 결과를 출력할 수 있다.

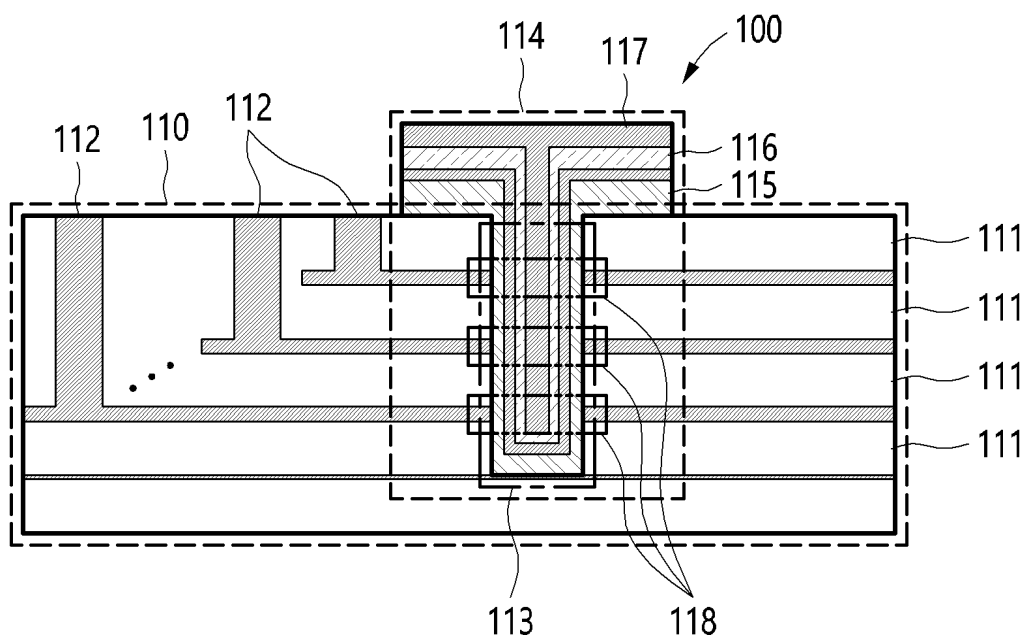
[0072] 본 문서에서, "A 또는 B", "A 및 B 중 적어도 하나", "A 또는 B 중 적어도 하나", "A, B 또는 C", "A, B 및 C 중 적어도 하나", 및 "A, B, 또는 C 중 적어도 하나"와 같은 문구들 각각은 그 문구들 중 해당하는 문구에 함께 나열된 항목들 중 어느 하나, 또는 그들의 모든 가능한 조합을 포함할 수 있다.

[0073] 이상과 같이 실시예들이 비록 한정된 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가진 자라면 이를 기초로 다양한 기술적 수정 및 변형을 적용할 수 있다. 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.

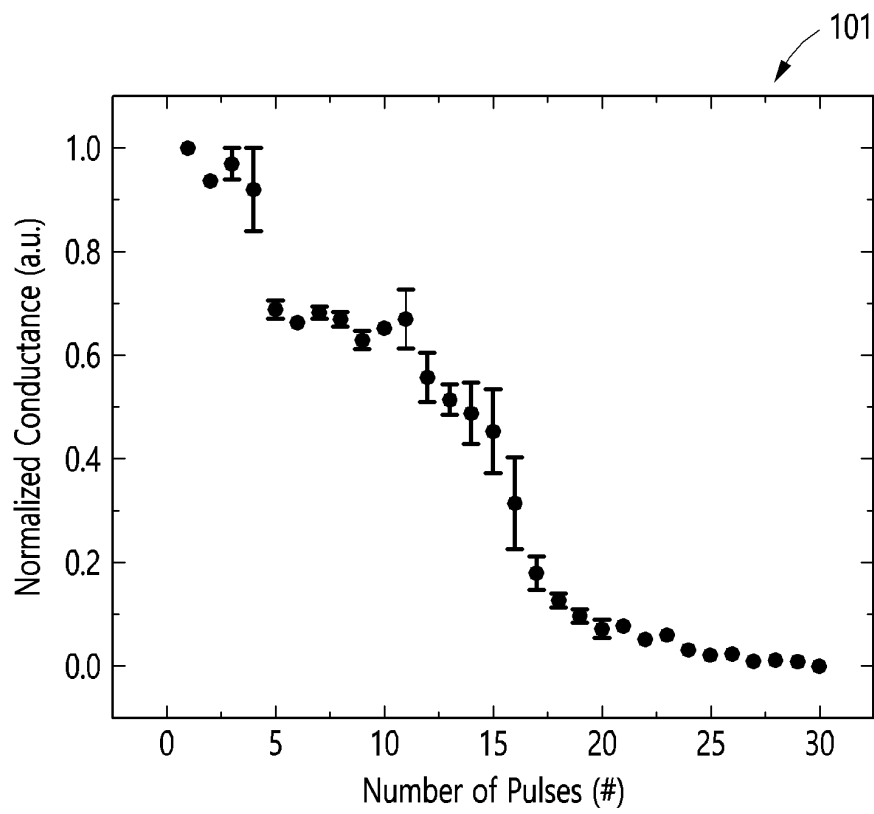
[0074] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

도면

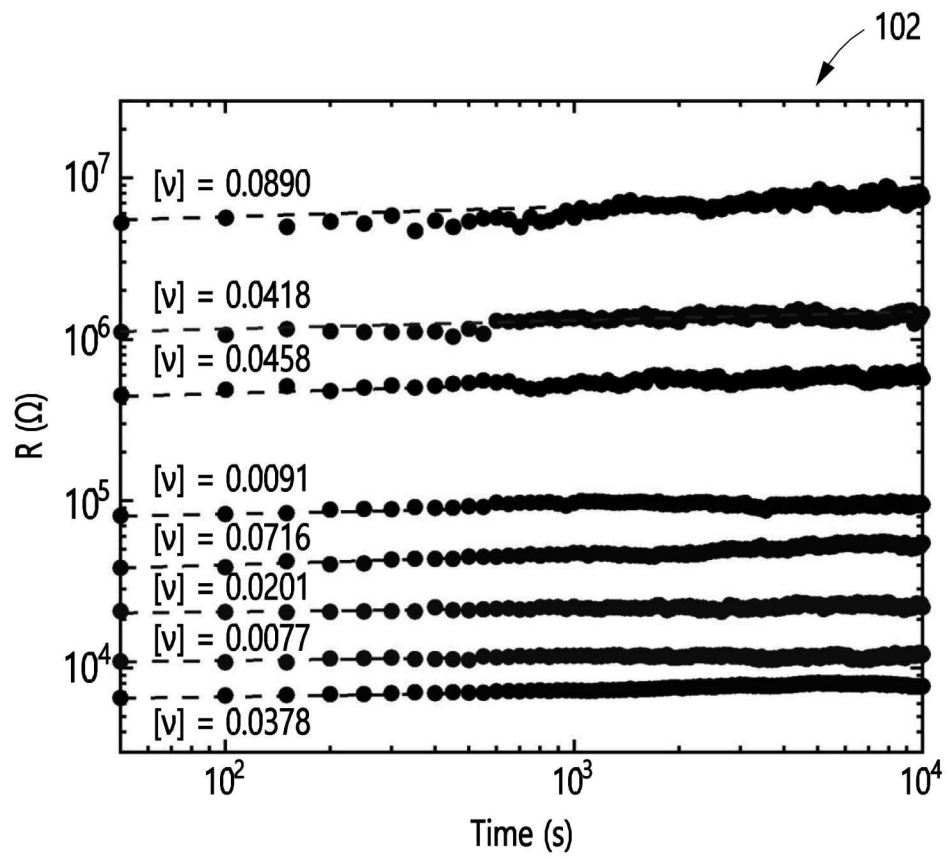
도면 1a



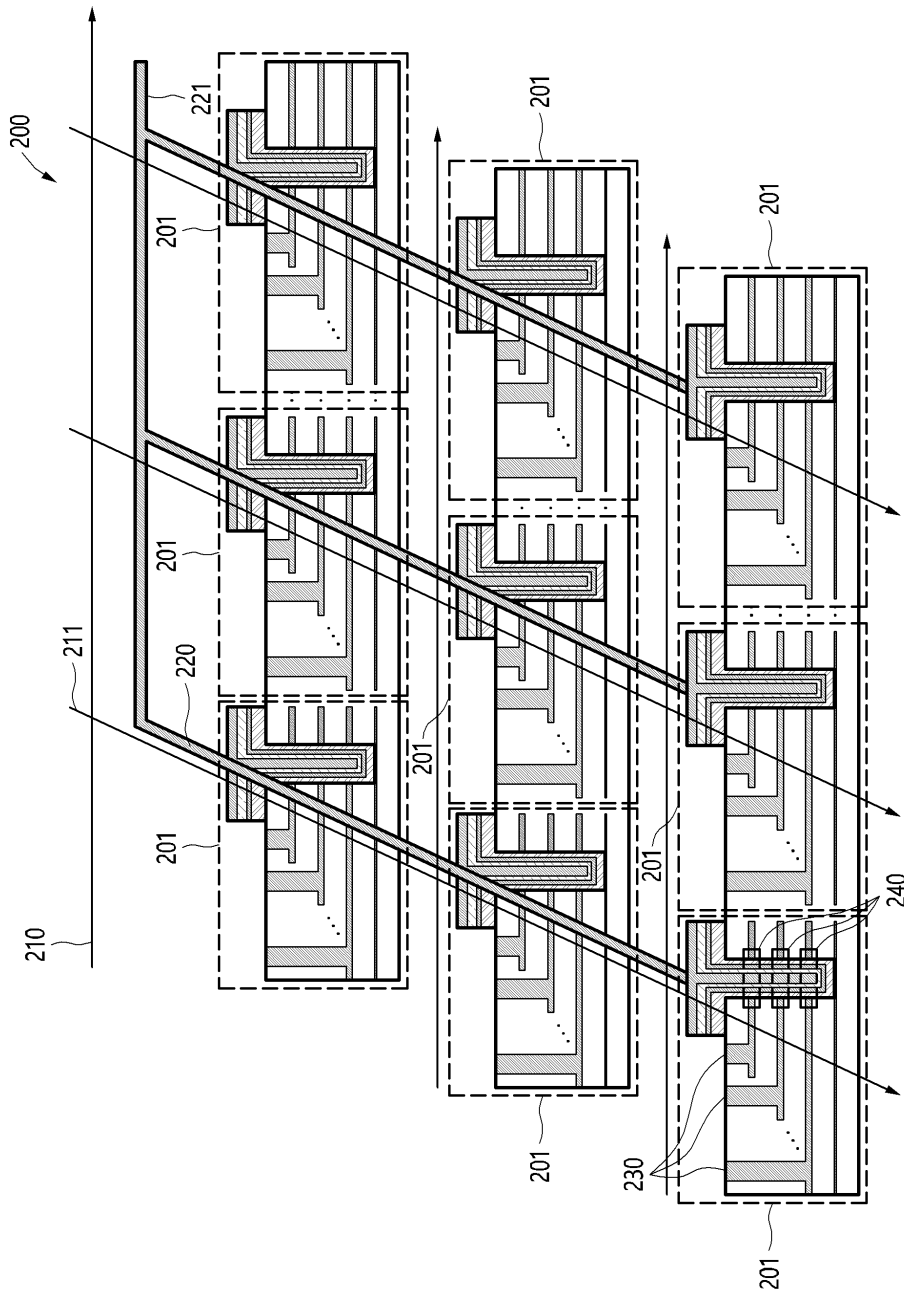
도면1b



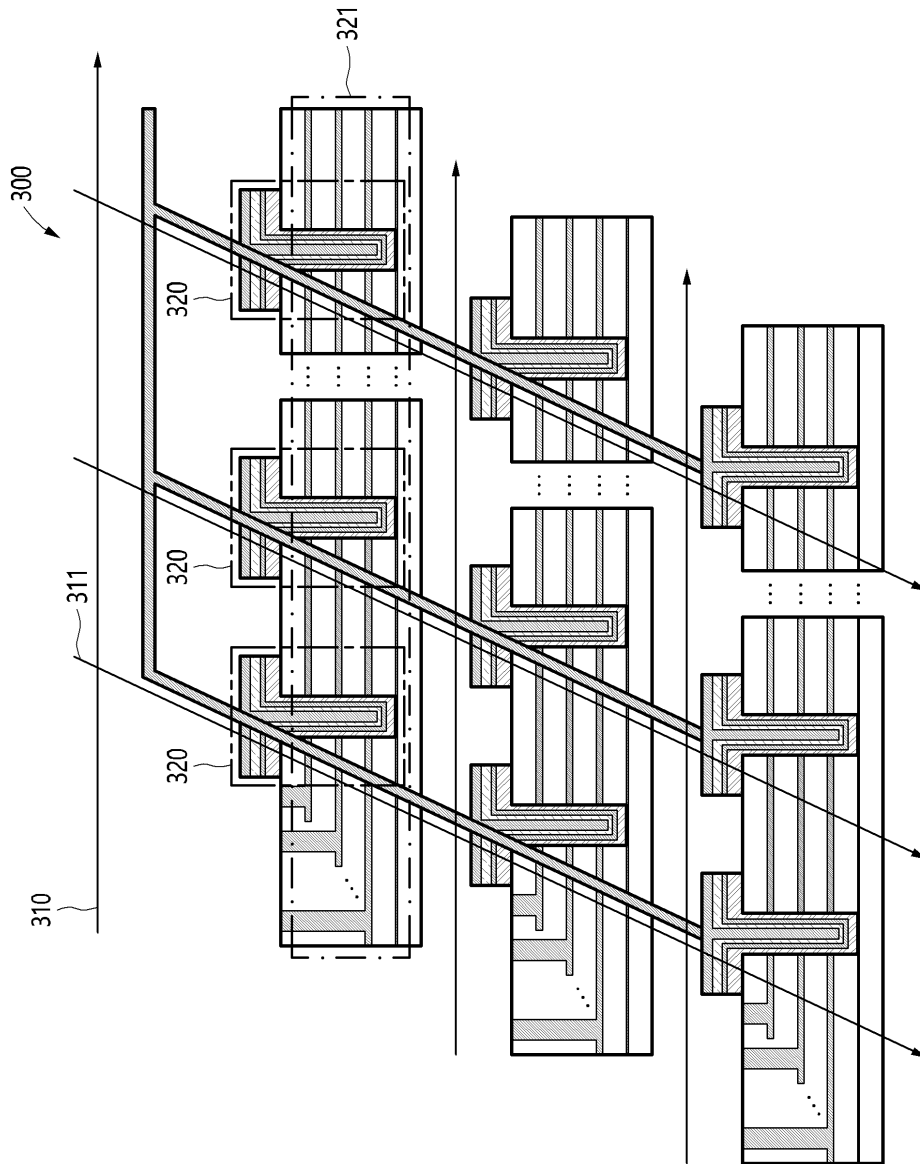
도면1c



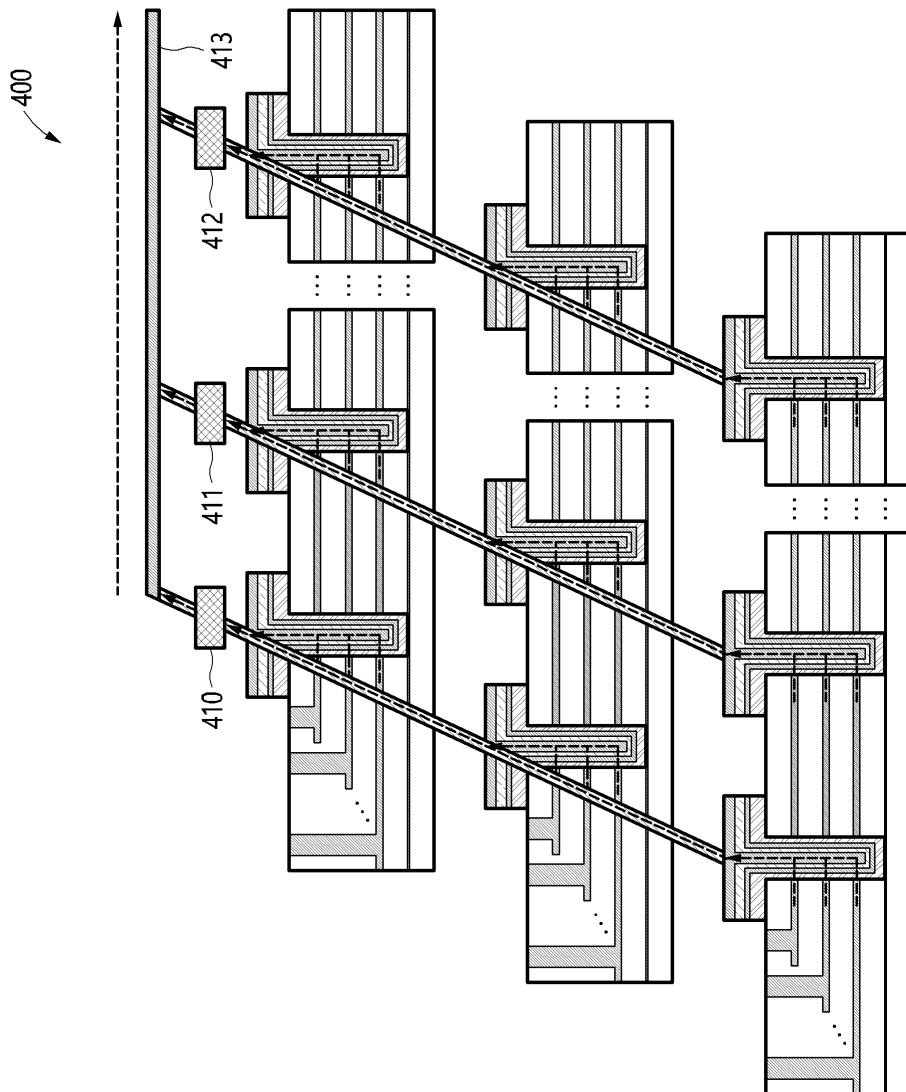
도면2



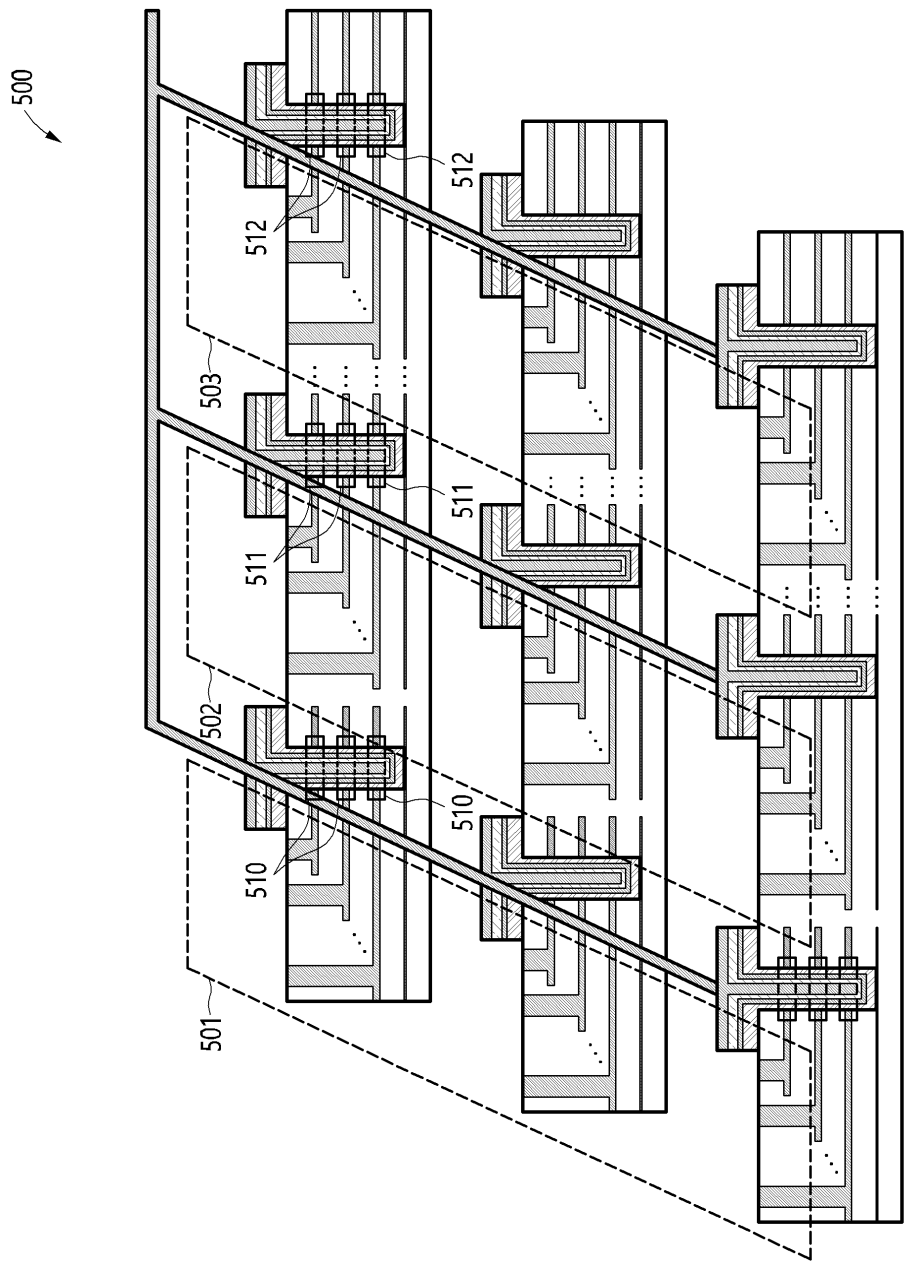
도면3



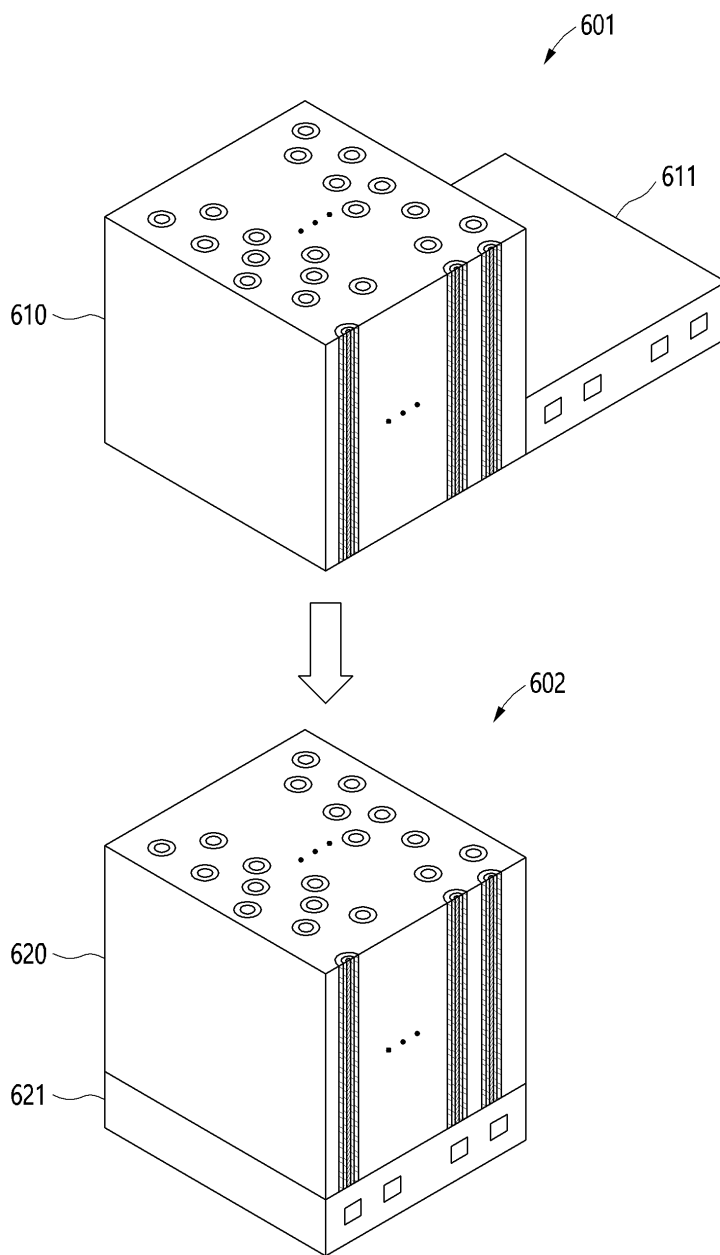
도면4



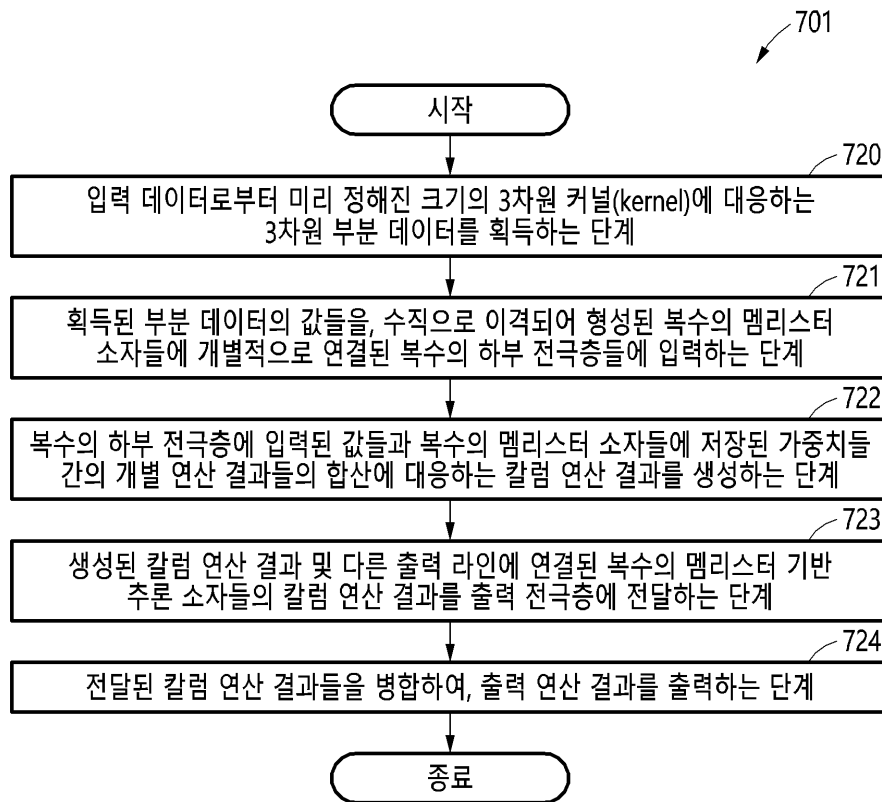
도면5



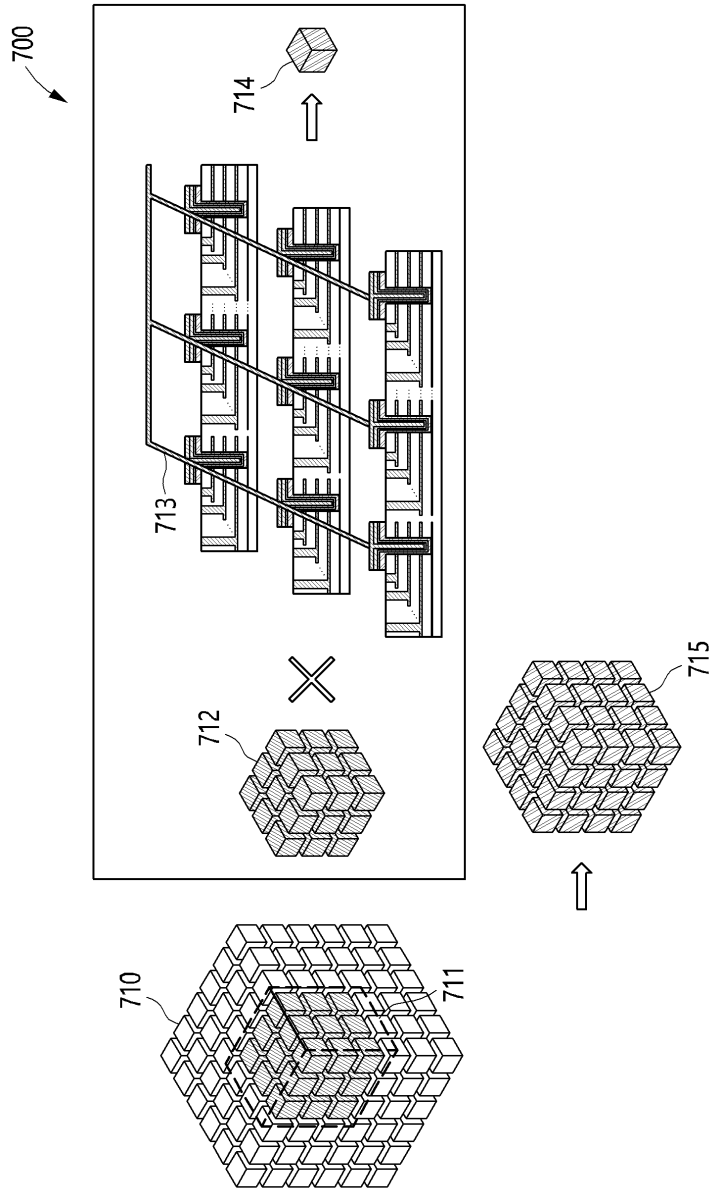
도면6



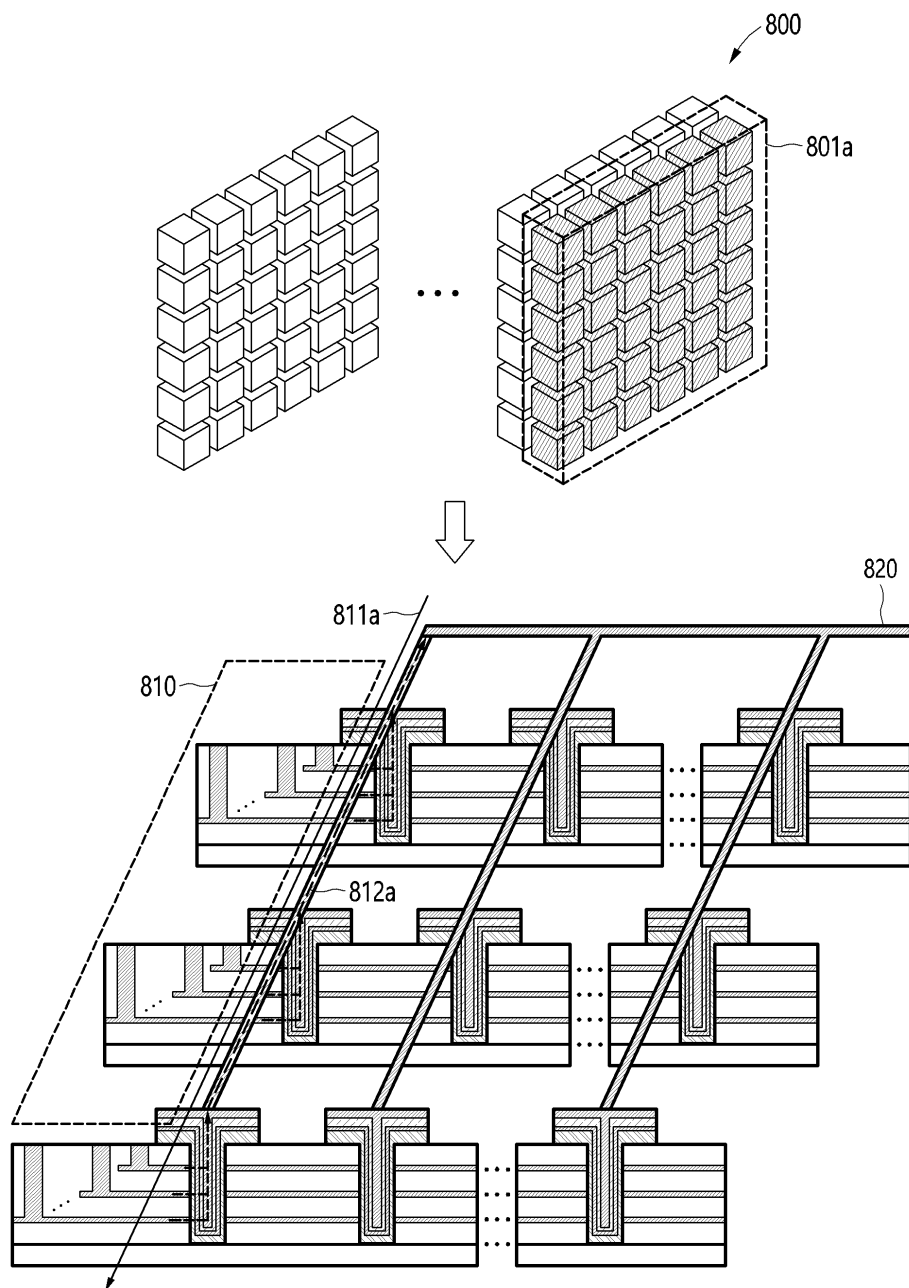
도면7a



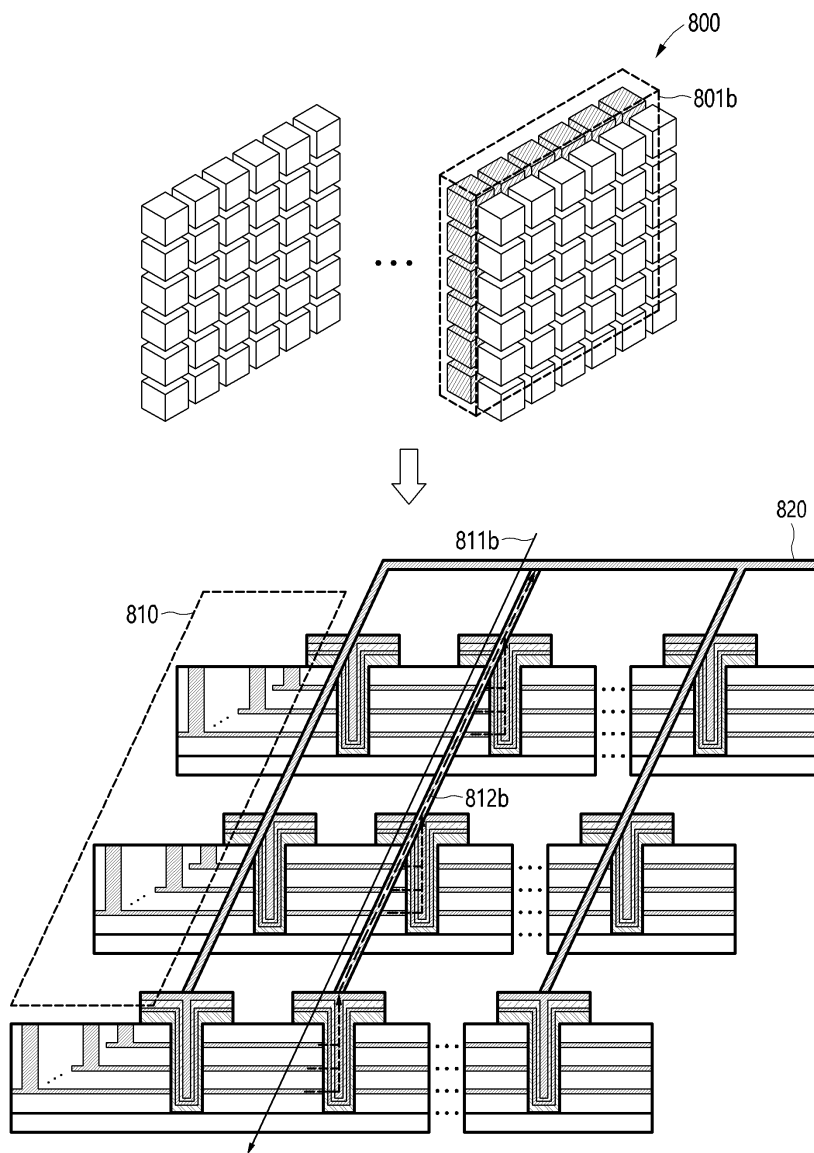
도면7b



도면8a



도면 8b



도면8c

