



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0046247
(43) 공개일자 2020년05월07일

(51) 국제특허분류(Int. Cl.)
H01L 27/11556 (2017.01) G06N 3/063 (2006.01)
H01L 21/768 (2006.01) H01L 27/11524 (2017.01)
H01L 27/1157 (2017.01) H01L 27/11582 (2017.01)
H01L 49/02 (2006.01)
(52) CPC특허분류
H01L 27/11556 (2013.01)
G06N 3/063 (2013.01)
(21) 출원번호 10-2018-0127101
(22) 출원일자 2018년10월24일
심사청구일자 2018년10월24일

(71) 출원인
포항공과대학교 산학협력단
경상북도 포항시 남구 청암로 77 (지곡동)
(72) 발명자
유인경
경기도 용인시 수지구 신봉2로 26 LG신봉자이1차 아파트 113동 802호
곽명훈
경상북도 포항시 남구 청암로 77 포항공과대학교 기숙사 12동 114호
황현상
대구광역시 수성구 들안로 360 수성태영데시앙아파트 106동906호
(74) 대리인
장인석

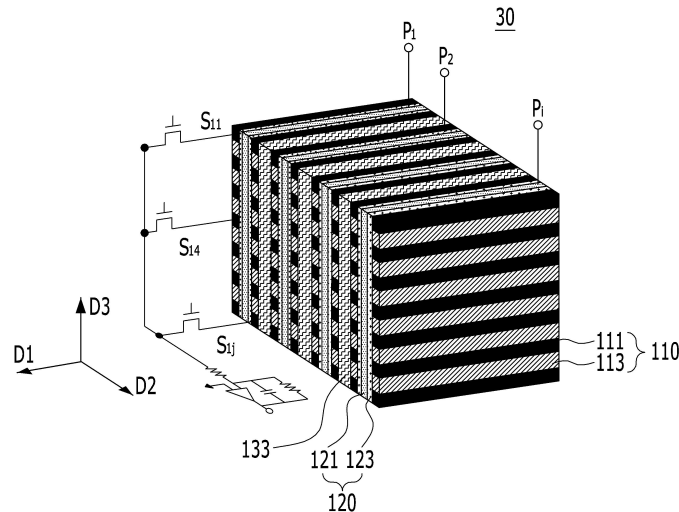
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 크로스-포인트 커패시터 기반의 가중치 소자 및 이를 이용한 뉴럴 네트워크

(57) 요약

본 발명의 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자는, 제1 방향으로 연장된 수평 도전 라인들, 및 수평 도전 라인들 사이에 제3 방향으로 수평 도전 라인들과 교대로 위치하는 수평 절연 라인층들을 포함하는 단위 수평 적층 구조체와, 제2 방향과 직교를 이루며 교대로 위치하는 수직 도전층 및 수직 유전체층을 포함하는 단위 수직 적층 구조체를 포함하고, 수평 도전 라인들, 수직 도전층, 및 수직 유전체층으로 형성된 커패시터를 시냅틱 가중치(synaptic weight)로 활용한다.

대표도 - 도2



(52) CPC특허분류

- H01L 21/76832* (2013.01)
- H01L 21/76846* (2013.01)
- H01L 27/11524* (2013.01)
- H01L 27/1157* (2013.01)
- H01L 27/11582* (2013.01)
- H01L 28/75* (2013.01)
- H01L 28/91* (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711070564
부처명	과학기술정보통신부
연구관리전문기관	한국산업기술평가관리원
연구사업명	산업기술혁신사업
연구과제명	시냅스 소자 기반 패턴인식 하드웨어 시스템의 시뮬레이션을 통한 소자/회로/아키텍처 플랫폼 개발
기여율	1/1
주관기관	포항공과대학교산학협력단
연구기간	2018.01.01 ~ 2018.12.31

명세서

청구범위

청구항 1

제1 방향으로 연장된 수평 도선 라인들, 및 상기 수평 도선 라인들 사이에 제3 방향으로 상기 수평 도선 라인들과 교대로 위치하는 수평 절연 라인층들을 포함하는 단위 수평 적층 구조체; 및

제2 방향과 직교를 이루며 교대로 위치하는 수직 도전층 및 수직 유전체층을 포함하는 단위 수직 적층 구조체를 포함하고,

상기 수평 도선 라인들, 상기 수직 도전층, 및 상기 수직 유전체층으로 형성된 커패시터를 시냅틱 가중치 (synaptic weight)로 활용하는 크로스-포인트 커패시터 기반의 가중치 소자.

청구항 2

제 1항에 있어서, 상기 커패시터에 충전된 후 방전되는 전하를 상기 시냅틱 가중치로 활용하는 크로스-포인트 커패시터 기반의 가중치 소자.

청구항 3

제 2항에 있어서, 상기 커패시터에 충전된 후 방전되는 전하는 입력 전압 펄스(input voltage pulse)에 비례하는 크로스-포인트 커패시터 기반의 가중치 소자.

청구항 4

제 2항에 있어서, 상기 커패시터에 충전된 후 방전되는 전하는 상기 수평 도선 라인들에 대응하는 워드 라인 (word line)이 선택된 개수와 비례하는 크로스-포인트 커패시터 기반의 가중치 소자.

청구항 5

제 4항에 있어서,

상기 워드 라인 선택 정보를 저장하기 위한 선택 트랜지스터를 더 포함하는 크로스-포인트 커패시터 기반의 가중치 소자.

청구항 6

제 5항에 있어서,

상기 선택 트랜지스터는 플로팅 게이트(floating gate) 트랜지스터를 포함하는 크로스-포인트 커패시터 기반의 가중치 소자.

청구항 7

제 1항에 있어서, 상기 단위 수직 적층 구조체는 상기 복수 개의 단위 수평 적층 구조체 사이에 위치하는 크로스-포인트 커패시터 기반의 가중치 소자.

청구항 8

제 1항에 있어서, 상기 수직 도전층은 수직 도선 라인 및 수직 절연 라인을 포함하는 크로스-포인트 커패시터 기반의 가중치 소자.

청구항 9

제 8항에 있어서, 상기 커패시터는 상기 수평 도선 라인들, 상기 수직 도선 라인, 및 상기 수직 유전체층으로 형성되는 크로스-포인트 커패시터 기반의 가중치 소자.

청구항 10

제 1항에 있어서, 상기 수평 도전 라인 및 상기 수직 도전층은 폴리 실리콘을 포함하는 크로스-포인트 커패시터 기반의 가중치 소자.

청구항 11

제 1항에 있어서, 상기 수평 절연 라인은 SiO₂를 포함하는 크로스-포인트 커패시터 기반의 가중치 소자.

청구항 12

제 1항에 있어서, 상기 수직 유전체층은 SiO₂, HfO₂, ZrO₂, Si₃N₄ 및 Al₂O₃ 중 적어도 하나를 포함하는 크로스-포인트 커패시터 기반의 가중치 소자.

청구항 13

제1 방향으로 연장된 수평 도전 라인들, 및 상기 수평 도전 라인들 사이에 제3 방향으로 상기 수평 도전 라인들과 교대로 위치하는 수평 절연 라인들을 포함하는 제1 수평 적층 구조체와 제2 수평 적층 구조체, 및 제2 방향과 직교를 이루며 교대로 위치하는 수직 도전층 및 수직 유전체층을 포함하고, 상기 제1 수평 적층 구조체와 상기 제2 수평 적층 구조체 사이에 위치하는 단위 수직 적층 구조체를 포함하는 가중치 그룹; 및

상기 복수 개의 가중치 그룹 사이에 위치하는 수직 절연층을 포함하고,

상기 수평 도전 라인들, 상기 수직 도전층, 및 상기 수직 유전체층으로 형성된 커패시터를 시냅틱 가중치 (synaptic weight)로 활용하는 크로스-포인트 커패시터 기반의 가중치 소자.

청구항 14

제 13항에 있어서, 상기 수직 절연층은 SiO₂를 포함하는 크로스-포인트 커패시터 기반의 가중치 소자.

청구항 15

제1 방향으로 연장된 수평 도전 라인들, 및 상기 수평 도전 라인들 사이에 제3 방향으로 상기 수평 도전 라인들과 교대로 위치하는 수평 절연 라인들을 포함하는 단위 수평 적층 구조체; 및 제2 방향과 직교를 이루며 교대로 위치하는 수직 도전층 및 수직 유전체층을 포함하는 단위 수직 적층 구조체를 포함하고, 상기 수평 도전 라인들과 상기 수직 도전층으로 형성된 커패시터에 충전된 후 방전되는 전하를 시냅틱 가중치(synaptic weight)로 활용하는 가중치 소자를 포함하는 뉴럴 네트워크.

청구항 16

제 15항에 있어서, 상기 커패시터에 충전된 후 방전되는 전하는 입력 전압 펄스에 비례하는 뉴럴 네트워크.

청구항 17

제 15항에 있어서, 상기 커패시터에 충전된 후 방전되는 전하는 상기 수평 도전 라인들에 대응하는 워드 라인 (word line)이 선택된 개수와 비례하는 뉴럴 네트워크.

청구항 18

제 17항에 있어서, 상기 가중치 소자는,

상기 워드 라인 선택 정보를 저장하기 위한 선택 트랜지스터를 더 포함하는 뉴럴 네트워크.

청구항 19

제 18항에 있어서, 상기 선택 트랜지스터는 플로팅 게이트(floating gate) 트랜지스터를 포함하는 뉴럴 네트워크.

청구항 20

제 17항에 있어서,

상기 워드 라인 선택 정보를 저장하기 위한 저장장치를 더 포함하는 뉴럴 네트워크.

발명의 설명

기술 분야

[0001] 본 발명은 크로스-포인트 커패시터 기반의 가중치 소자 및 이를 이용한 뉴럴 네트워크에 관한 것이다.

배경 기술

[0002] 최근 인간의 신경계를 닮은 뉴로모픽 회로(neuromorphic circuit)에 관한 관심이 증대되고 있다. 인간의 신경계에 존재하는 뉴런(neuron)과 시냅스(synapse)에 각각 대응되는, 뉴런 회로와 시냅스 회로를 설계하여, 뉴로모픽 회로를 구현하고자 하는 연구가 활발히 진행되고 있다.

[0003] 뉴로모픽 회로는 불특정한 환경에 스스로 적응할 수 있는 지능화된 시스템을 구현하는 데에 효과적으로 활용될 수 있다. 이 기술이 발전하면 음성인식, 위험 인지, 실시간 고속 신호처리, 등 인지 및 추정 등을 수행하는 컴퓨터, 로봇, 가전기기, 소형 이동 기기, 보안 및 감시, 지능형 차량 안전, 자율 주행 등으로 발전할 수 있다.

[0004] 위 기재된 내용은 오직 본 발명의 기술적 사상들에 대한 배경 기술의 이해를 돕기 위한 것이며, 따라서 그것은 본 발명의 기술 분야의 당업자에게 알려진 선행 기술에 해당하는 내용으로 이해될 수 없다.

발명의 내용

해결하려는 과제

[0005] 뉴럴 네트워크를 하드웨어(hardware)화 하는데 있어서 핵심 소자는 시냅틱 가중치(synaptic weight)와 뉴런이다. 시냅틱 가중치는 불휘발성 멀티레벨(multi level)이면서 선형성이 있어야 하고 뉴런은 활성화(activation) 기능이 있어야 한다. 이를 위하여 제작이 용이한 CMOS 기술을 사용하는 것이 바람직하나, CMOS 기술을 사용하는 경우 회로가 증가하면서 칩 사이즈가 커지는 문제가 있다. 이에 대해 멀티레벨 메모리 소재를 시냅틱 가중치로 사용하고 임계 스위칭(threshold switching) 소재를 뉴런으로 사용하려고 하는 노력이 있으나 CMOS 기술 정도의 성능과 제조성이 없다.

[0006] 본 발명의 실시예는 수직 적층형 크로스-포인트 커패시터 셀을 이용하여 선형적인 다중 시냅틱 가중치를 가지는 크로스-포인트 커패시터 기반의 가중치 소자의 구조 및 이를 이용한 뉴럴 네트워크를 제공함을 목적으로 한다.

과제의 해결 수단

[0007] 상기 목적을 달성하기 위하여 본 발명의 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자는 제1 방향으로 연장된 수평 도전 라인들, 및 상기 수평 도전 라인들 사이에 제3 방향으로 상기 수평 도전 라인들과 교대로 위치하는 수평 절연 라인층들을 포함하는 단위 수평 적층 구조체와; 제2 방향과 직교를 이루며 교대로 위치하는 수직 도전층 및 수직 유전체층을 포함하는 단위 수직 적층 구조체를 포함하고, 상기 수평 도전 라인들, 상기 수직 도전층, 및 상기 수직 유전체층으로 형성된 커패시터를 시냅틱 가중치(synaptic weight)로 활용한다.

[0008] 상기 커패시터에 충전된 후 방전되는 전하를 상기 시냅틱 가중치로 활용할 수 있다.

[0009] 상기 커패시터에 충전된 후 방전되는 전하는 입력 전압 펄스(input voltage pulse)에 비례할 수 있다.

[0010] 상기 커패시터에 충전된 후 방전되는 전하는 상기 수평 도전 라인들에 대응하는 워드 라인(word line)이 선택된 개수와 비례할 수 있다.

[0011] 상기 가중치 소자는 상기 워드 라인 선택 정보를 저장하기 위한 선택 트랜지스터를 더 포함할 수 있다.

[0012] 상기 선택 트랜지스터는 플로팅 게이트(floating gate) 트랜지스터를 포함할 수 있다.

[0013] 상기 단위 수직 적층 구조체는 상기 복수 개의 단위 수평 적층 구조체 사이에 위치할 수 있다.

[0014] 상기 수직 도전층은 수직 도전 라인 및 수직 절연 라인을 포함할 수 있다.

[0015] 상기 커패시터는 상기 수평 도전 라인들, 상기 수직 도전 라인, 및 상기 수직 유전체층으로 형성될 수 있다.

- [0016] 상기 수평 도전 라인 및 상기 수직 도전층은 폴리 실리콘을 포함할 수 있다.
- [0017] 상기 수평 절연 라인은 SiO₂를 포함할 수 있다.
- [0018] 상기 수직 유전체층은 SiO₂, HfO₂, ZrO₂, Si₃N₄ 및 Al₂O₃ 중 적어도 하나를 포함할 수 있다.
- [0019] 본 발명의 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자는 제1 방향으로 연장된 수평 도전 라인들, 및 상기 수평 도전 라인들 사이에 제3 방향으로 상기 수평 도전 라인들과 교대로 위치하는 수평 절연 라인들을 포함하는 제1 수평 적층 구조체와 제2 수평 적층 구조체, 및 제2 방향과 직교를 이루며 교대로 위치하는 수직 도전층 및 수직 유전체층을 포함하고, 상기 제1 수평 적층 구조체와 상기 제2 수평 적층 구조체 사이에 위치하는 단위 수직 적층 구조체를 포함하는 가중치 그룹과; 상기 복수 개의 가중치 그룹 사이에 위치하는 수직 절연층을 포함하고, 상기 수평 도전 라인들, 상기 수직 도전층, 및 상기 수직 유전체층으로 형성된 커패시터를 시냅틱 가중치(synaptic weight)로 활용한다.
- [0020] 상기 수직 절연층은 SiO₂를 포함할 수 있다.
- [0021] 본 발명의 실시예에 따른 뉴럴 네트워크는 제1 방향으로 연장된 수평 도전 라인들, 및 상기 수평 도전 라인들 사이에 제3 방향으로 상기 수평 도전 라인들과 교대로 위치하는 수평 절연 라인들을 포함하는 단위 수평 적층 구조체; 및 제2 방향과 직교를 이루며 교대로 위치하는 수직 도전층 및 수직 유전체층을 포함하는 단위 수직 적층 구조체를 포함하고, 상기 수평 도전 라인들과 상기 수직 도전층으로 형성된 커패시터에 충전된 후 방전되는 전하를 시냅틱 가중치(synaptic weight)로 활용하는 가중치 소자를 포함한다.
- [0022] 상기 커패시터에 충전된 후 방전되는 전하는 입력 전압 펄스에 비례할 수 있다.
- [0023] 상기 커패시터에 충전된 후 방전되는 전하는 상기 수평 도전 라인들에 대응하는 워드 라인(word line)이 선택된 개수와 비례할 수 있다.
- [0024] 상기 가중치 소자는, 상기 워드 라인 선택 정보를 저장하기 위한 선택 트랜지스터를 더 포함할 수 있다.
- [0025] 상기 선택 트랜지스터는 플로팅 게이트(floating gate) 트랜지스터를 포함할 수 있다.
- [0026] 상기 뉴럴 네트워크는 상기 워드 라인 선택 정보를 저장하기 위한 저장장치를 더 포함할 수 있다.

발명의 효과

- [0027] 이와 같은 본 발명의 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자의 구조 및 이를 이용한 뉴럴 네트워크에 의하면 수직 적층형 크로스-포인트 커패시터 셀을 이용한 선형적인 다중 시냅틱 가중치를 가짐으로써, 학습 효율을 증가시킬 수 있다.
- [0028] 또한, 커패시터를 가중치로 사용함으로써 저항 가중치의 저항 값을 변화시켜 이에 비례하는 전류를 출력 신호로 사용하는 전도도 기반의(conductance-based) 가중치 소자에 비하여 전력 소모를 크게 감소시킬 수 있다.

도면의 간단한 설명

- [0029] 도 1은 본 발명의 일 실시예에 따른 뉴럴 네트워크를 개념적으로 도시한 블록도이다.
- 도 2는 본 발명의 일 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자의 구조를 개략적으로 보여주는 사시도이다.
- 도 3은 본 발명의 일 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자의 구조를 개략적으로 보여주는 단면도이다.
- 도 4는 본 발명의 일 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자의 동작을 개략적으로 나타내는 회로도이다.
- 도 5는 본 발명의 다른 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자의 구조를 개략적으로 보여주는 사시도이다
- 도 6은 본 발명의 다른 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자의 동작을 개략적으로 나타내는 회로도이다.

발명을 실시하기 위한 구체적인 내용

- [0030] 위 발명의 배경이 되는 기술 란에 기재된 내용은 오직 본 발명의 기술적 사상에 대한 배경 기술의 이해를 돕기 위한 것이며, 따라서 그것은 본 발명의 기술 분야의 당업자에게 알려진 선행 기술에 해당하는 내용으로 이해될 수 없다.
- [0031] 아래의 서술에서, 설명의 목적으로, 다양한 실시예들의 이해를 돕기 위해 많은 구체적인 세부 내용들이 제시된다. 그러나, 다양한 실시예들이 이러한 구체적인 세부 내용들 없이 또는 하나 이상의 동등한 방식으로 실시될 수 있다는 것은 명백하다. 다른 예시들에서, 잘 알려진 구조들과 장치들은 다양한 실시예들을 불필요하게 이해하기 어렵게 하는 것을 피하기 위해 블록도로 표시된다.
- [0032]도면에서, 레이어들, 필름들, 패널들, 영역들 등의 크기 또는 상대적인 크기는 명확한 설명을 위해 과장될 수 있다. 또한, 동일한 참조 번호는 동일한 구성 요소를 나타낸다.
- [0033]명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "간접적으로 연결"되어 있는 경우도 포함한다. 그러나, 만약 어떤 부분이 다른 부분과 "직접적으로 연결되어 있다"고 서술되어 있으면, 이는 해당 부분과 다른 부분 사이에 다른 소자가 없음을 의미할 것이다. "X, Y, 및 Z 중 적어도 어느 하나", 그리고 "X, Y, 및 Z로 구성된 그룹으로부터 선택된 적어도 어느 하나"는 X 하나, Y 하나, Z 하나, 또는 X, Y, 및 Z 중 둘 또는 그 이상의 어떤 조합 (예를 들면, XYZ, XYY, YZ, ZZ)으로 이해될 것이다. 여기에서, "및/또는"은 해당 구성들 중 하나 또는 그 이상의 모든 조합을 포함한다.
- [0034]여기에서, 첫번째, 두번째 등과 같은 용어가 다양한 소자들, 요소들, 지역들, 레이어들, 및/또는 섹션들을 설명하기 위해 사용될 수 있지만, 이러한 소자들, 요소들, 지역들, 레이어들, 및/또는 섹션들은 이러한 용어들에 한정되지 않는다. 이러한 용어들은 하나의 소자, 요소, 지역, 레이어, 및/또는 섹션을 다른 소자, 요소, 지역, 레이어, 및 또는 섹션과 구별하기 위해 사용된다. 따라서, 일 실시예에서의 첫번째 소자, 요소, 지역, 레이어, 및/또는 섹션은 다른 실시예에서 두번째 소자, 요소, 지역, 레이어, 및/또는 섹션이라 칭할 수 있다.
- [0035]"아래", "위" 등과 같은 공간적으로 상대적인 용어가 설명의 목적으로 사용될 수 있으며, 그렇게 함으로써 도면에서 도시된 대로 하나의 소자 또는 특징과 다른 소자(들) 또는 특징(들)과의 관계를 설명한다. 이는 도면 상에서 하나의 구성 요소의 다른 구성 요소에 대한 관계를 나타내는 데에 사용될 뿐, 절대적인 위치를 의미하는 것은 아니다. 예를 들어, 도면에 도시된 장치가 뒤집히면, 다른 소자들 또는 특징들의 "아래"에 위치하는 것으로 묘사된 소자들은 다른 소자들 또는 특징들의 "위"의 방향에 위치한다. 따라서, 일 실시예에서 "아래"라는 용어는 위와 아래의 양방향을 포함할 수 있다. 뿐만 아니라, 장치는 그 외의 다른 방향일 수 있다 (예를 들어, 90도 회전된 혹은 다른 방향에서), 그리고, 여기에서 사용되는 그런 공간적으로 상대적인 용어들은 그에 따라 해석된다.
- [0036]여기에서 사용되는 용어는 특정한 실시예들을 설명하는 목적이고 제한하기 위한 목적이 아니다. 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다. 다른 정의가 없는 한, 여기에 사용되는 용어들은 본 발명이 속하는 분야에서 통상적인 지식을 가진 자에게 일반적으로 이해되는 것과 같은 의미를 갖는다.
- [0037]도 1은 본 발명의 일 실시예에 따른 뉴럴 네트워크를 개념적으로 도시한 블록도이다.
- [0038]도 1을 참조하면, 본 발명의 일 실시예에 따른 뉴럴 네트워크(neural network)는 입력 뉴런(10), 출력 뉴런(20), 및 가중치 소자(30)를 포함한다. 시냅스(30) 소자는 입력 뉴런(10)으로부터 수평으로 연장하는 로우 라인(R)(row lines) 및 출력 뉴런(20)으로부터 수직으로 연장하는 컬럼 라인(C)(column lines)의 교차점에 배치될 수 있다. 설명의 편의를 위해 도 1에는 예시적으로 각각 네 개의 입력 뉴런(10) 및 출력 뉴런(20)이 도시되었으나, 본 발명은 이에 한정되지 않는다.
- [0039]입력 뉴런(10)은 학습 모드(learning mode), 리셋 모드(reset mode), 보정 또는 읽기 모드(reading mode)에서 로우 라인(R)을 통하여 가중치 소자(30)로 전기적 펄스들(pulses)을 전송할 수 있다.
- [0040]출력 뉴런(20)은 학습 모드 또는 리셋 모드 또는 보정 시 컬럼 라인(C)을 통하여 가중치 소자(30)로 전기적 펄스를 전송할 수 있고, 및 읽기 모드에서 컬럼 라인(C)을 통하여 가중치 소자(30)로부터 전기적 펄스를 수신할

수 있다.

- [0041] 가중치 소자(30)는 멀티레벨(multi-level) 값을 가질 수 있다. 실시예로서, 가중치 소자(30)는 플로팅게이트 트랜지스터들의 write/erase 여부에 따라 그 값을 달리한다. 가중치 소자(30)에 연결된 플로팅게이트 트랜지스터가 erase 상태일 때에는 트랜지스터 전류의 흐름이 원활하므로 연결된 커패시터들의 충방전 동작이 가능하며, write 상태일 때에는 플로팅게이트 트랜지스터 전류의 흐름을 막으므로 연결된 커패시터들의 충방전 동작이 불가능하다. 이때 erase 상태인 플로팅게이트 트랜지스터들의 수에 비례하여 가중치 소자(30)의 커패시터 값이 결정된다. 즉, 가중치 소자(30)에 연결된 플로팅게이트 트랜지스터의 수 만큼 멀티레벨을 가질 수 있다. 전압 펄스를 입력 신호로 인가하면 erase 상태인 플로팅게이트 트랜지스터에 연결된 커패시터만 충전 및 방전하며 이 전하들을 모아 전압으로 변환하여 출력 신호로 사용할 수 있다. 본 발명의 일 실시예에 따르면, 출력 신호를 시냅틱 가중치로 활용할 수 있다.
- [0042] 도 2는 본 발명의 일 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자의 구조를 개략적으로 보여주는 사시도이다. 도 3은 본 발명의 일 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자의 구조를 개략적으로 보여주는 단면도이다.
- [0043] 도 2 및 도 3을 참조하면, 본 발명의 일 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자(30)는 제1 방향으로 연장된 수평 도전 라인(111)들과 수평 도전 라인(111)들 사이에 위치한 수평 절연 라인(113)들을 포함하는 단위 수평 적층 구조체(110), 제2 방향과 직교를 이루며 교대로 위치하는 수직 도전층(121) 및 수직 유전체층(123)을 포함하는 단위 수직 적층 구조체(120)를 포함한다. 단위 수평 적층 구조체(110)의 일 측면을 따라 수직 유전체층(123)과 수직 도전층(121)이 교대로 형성된 구조를 단위 수직 적층 구조체(120)로 정의한다. 실시예로서, 가중치 소자(30)의 제조는 낸드(NAND) 공정을 적용할 수 있다.
- [0044] 수평 도전 라인(111)은 폴리 실리콘(Poly silicon)일 수 있다. 실시예로서, 수평 도전 라인(111)은 도전체로 기능하기 위하여 높은 도펀트 농도를 가질 수 있다. 실시예로서, 수평 도전 라인(111)은 워드 라인일 수 있다.
- [0045] 수평 절연 라인(113)은 복수 개의 수평 도전 라인(111)들 사이를 절연한다. 실시예로서, 수평 절연 라인(113)은 SiO₂, Si₃N₄, 금속 산화물, 금속 질화물 및 고분자 물질막 중 적어도 어느 하나를 포함할 수 있으나, 본 발명은 이에 한정되지 않는다.
- [0046] 단위 수직 적층 구조체(120)는 상기 복수 개의 단위 수평 적층 구조체(110) 사이에 위치할 수 있다.
- [0047] 수직 도전층(121)은 제2 방향과 직교를 이루며 형성된다. 실시예로서, 수직 도전층(121)은 폴리 실리콘일 수 있으나 이에 한정되는 것은 아니다. 수직 도전층(121)과 수평 도전 라인(111)으로 커패시터를 구성하기 위하여 수직 도전층(121)과 수평 도전 라인(111) 사이에 수직 유전체층(123)이 형성된다. 실시예로서, 수직 도전층(121)은 플레이트 일 수 있고, 수직 유전체층(123)은 커패시터의 축전 용량을 향상시키면서 효과적으로 전극 사이를 절연하기 위한 유전물질이 이용될 수 있다. 실시예로서, 수직 유전체층(123)은 SiO₂, HfO₂, ZrO₂, Si₃N₄ 및 Al₂O₃ 중 적어도 하나를 포함할 수 있다.
- [0048] 본 발명의 일 실시예에 따르면, 수평 도전 라인(111)들과 수직 도전층(121)으로 형성된 커패시터에 축적된 전하를 시냅틱 가중치(synaptic weight)로 활용한다. 실시예로서, 수평 도전 라인(111)과 전기적으로 접속되는 라인을 통해 커패시터에 저장된 전하량을 출력할 수 있다.
- [0049] 본 발명의 다른 실시예에 따르면, 크로스-포인트 커패시터 기반의 가중치 소자(30)는 제1 단위 수평 적층 구조체(110)와 제2 단위 수평 적층 구조체(110), 및 단위 수직 적층 구조체(120)를 포함하는 가중치 그룹과 수직 절연층(133)을 포함할 수 있다.
- [0050] 제1 단위 수평 적층 구조체(110)와 제2 단위 수평 적층 구조체(110)는 제1 방향으로 연장된 수평 도전 라인(111)들, 및 수평 도전 라인(111)들 사이에 제3 방향으로 수평 도전 라인(111)들과 교대로 위치하는 수평 절연 라인(113)들을 포함할 수 있다.
- [0051] 단위 수직 적층 구조체(120)는 제2 방향과 직교를 이루며 교대로 위치하는 수직 도전층(121) 및 수직 유전체층(123)을 포함할 수 있다.
- [0052] 수직 절연층(133)은 복수 개의 가중치 그룹 사이에 위치할 수 있다.
- [0053] 본 발명의 다른 실시예에 따르면, 수평 도전 라인(111)들과 수직 도전층(121)으로 형성된 커패시터에 축적된 전하를 시냅틱 가중치(synaptic weight)로 활용한다. 실시예로서, 수직 도전층(121)의 왼쪽에 위치하는 수평 도전

라인(111) 그룹과 오른쪽 수평 도선 라인(111) 그룹을 각각 positive 가중치, negative 가중치로 정의하여 활용할 수 있다.

- [0054] 도 2를 참조하면, 본 발명의 일 실시예에 따른 가중치 소자(30)를 시냅틱 가중치로 사용하기 위해 적층 크로스-포인트 커패시터 셀에 전압을 인가한다. 온-칩 인공지능 학습(On-Chip learning)에 의한 입력 신호는 일정 시간 동안에 입력되는 전압 펄스의 수를 이용하거나 전압 펄스의 전압 크기를 이용하여 다중 값을 크로스-포인트 커패시터 기반의 가중치 소자에 입력할 수 있다. 실시예로서, 선택 트랜지스터(S11, S14, S1j)로 각각의 수평 도선 라인(111)들, 즉 워드 라인을 선택한 후 공통 전극인 수직 도선층(121)에 전압을 인가한다.
- [0055] 수직 도선층(121)에 인가하는 전압 펄스는 입력(input) 상황에 맞추어 펄스 수 또는 펄스 전압의 크기를 조절할 수 있다. 가중치 값에 대응하는 워드 라인 수는 입력 펄스 수 또는 입력 전압과는 독립적으로 선택할 수 있다.
- [0056] 수직 도선층(121)과 수평 도선 라인(111)들이 교차하는 크로스 포인트는 커패시터가 형성되고 입력 전압에 의해 전하가 축적되게 된다. 축적된 전하는 합산되어 적분기(integrator)로 보내지고 적분기에서는 전압으로 변환된다. 전압으로 변환된 출력 신호는 문턱 전압 크기 이상이 될 때 다음 레이어를 구성하는 입력 노드(input node)로 전달되게 된다.
- [0057] 도 4는 본 발명의 일 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자의 동작을 개략적으로 나타내는 회로도이다.
- [0058] 도 4를 참조하면, 다중 가중치를 구성하기 위해 크로스-포인트 커패시터 셀에서 워드 라인의 수를 선택하고 커패시터의 공통 전극에 Vpp 전압을 인가함으로써 선택 트랜지스터 쪽으로 전하들이 나오도록 한다. 커패시터에 저장되는 전하량(Q)은 커패시터의 축전 용량(C)과 인가되는 전압(V)에 비례한다. 이러한 원리를 이용하여 크로스-포인트 노드의 도선 경로 수에 따라 결정되는 여러 개의 커패시터를 사용함으로써 입력 신호에 따른 출력 신호가 선형적인(linear) 값을 가질 수 있다.
- [0059] 시냅틱 가중치는 커패시턴스(capacitance)로 정의되고 출력(output)은 워드 라인에 연결된 n개의 커패시터에서 방전되는 전하(Q=nCVpp)로 결정된다. 실시예로서, 가중치 레벨은 워드 라인을 선택한 수로 결정될 수 있다.
- [0060] 본 발명의 일 실시예에 따르면, 뉴럴 네트워크로 학습 혹은 훈련을 수행할 때 입력 노드로 전압 펄스(voltage pulse)가 인가되고 각각의 가중치의 크기에 따라 커패시터 셀들의 워드 라인을 선택한다. 실시예로서, 가중치가 1이면 워드 라인 한 개를 선택하고 가중치가 5이면 다섯 개의 워드 라인을 선택한다. 각 가중치를 무작위로 (randomly), 동시에(simultaneously), 그리고 독립적(independently)으로 직접(directly) 선택함으로써 입력에 대한 출력이 매트릭스로 발생하도록 한다.
- [0061] 단위 가중치는 일정한 수의 워드 라인이 있는 커패시터 셀 블록으로 구성할 수 있다. 실시예로서, 선택할 수 있는 워드 라인의 최대 개수가 최대 다이내믹 레인지(dynamic range)가 될 수 있다. 각 가중치에서 방전되는 전하의 합은 적분기로 축적하여 전압으로 변환할 수 있다.
- [0062] 본 발명의 일 실시예에 따르면, 학습이나 training 후 최종으로 결정된 워드 라인 선택 개수의 정보는 외부 저장장치에 저장할 수 있다. 뉴럴 네트워크 회로 안에 저장기능을 장착하고자 할 때에는 선택 트랜지스터로서 플로팅 게이트 트랜지스터를 사용할 수 있다. 실시예로서, 역전파(back-propagation) 등의 알고리즘을 사용하면서 트랜스포즈 가중치 매트릭스(transpose weight matrix)를 이용하여 학습 시 가중치의 보정량을 구하고, 얻어진 가중치의 보정량에 비례하여 워드 라인을 선택할 수 있다.
- [0063] 도 5는 본 발명의 다른 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자의 구조를 개략적으로 보여주는 사시도이다
- [0064] 도 5를 참조하면, 본 발명의 다른 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자(30)는 제1 방향으로 연장된 수평 도선 라인(111)들과 수평 도선 라인(111)들 사이에 위치한 수평 절연 라인(113)들을 포함하는 단위 수평 적층 구조체(110), 제2 방향과 직교를 이루며 교대로 위치하는 수직 도선 라인(122), 수직 절연 라인(124) 및 수직 유전체층(123)을 포함하는 단위 수직 적층 구조체(120)를 포함한다. 실시예로서, 가중치 소자(30)의 제조는 낸드(NAND) 공정을 적용할 수 있다.
- [0065] 수평 절연 라인(113)은 복수 개의 수평 도선 라인(111)들 사이를 절연한다. 실시예로서, 수평 절연 라인(113)은 SiO2, Si3N4, 금속 산화물, 금속 질화물 및 고분자 물질막 중 적어도 어느 하나를 포함할 수 있으나, 본 발명에는 한정되지 않는다.

- [0066] 단위 수직 적층 구조체(120)는 상기 복수 개의 단위 수평 적층 구조체(110) 사이에 위치할 수 있다.
- [0067] 수직 도전 라인(122)은 제2 방향과 직교를 이루며 형성된다. 실시예로서, 수직 도전 라인(122)은 폴리 실리콘일 수 있으나 이에 한정되는 것은 아니다. 수직 도전 라인(122)과 수평 도전 라인(111)으로 커패시터를 구성하기 위하여 수직 도전 라인(121)과 수평 도전 라인(111) 사이에 수직 유전체층(123)이 형성되고, 수직 도전 라인(122)들 사이를 절연하기 위하여 수직 절연 라인(124)이 형성된다. 단위 수평 적층 구조체(110)의 일 측면을 따라 수직 유전체층(123)과 수직 도전 라인(122) 및 수직 절연 라인(124)이 교대로 형성된 구조를 단위 수직 적층 구조체(120)로 정의한다. 실시예로서, 수직 도전 라인(122)은 비트 라인일 수 있고, 수직 유전체층(123)은 커패시터의 축전 용량을 향상시키면서 효과적으로 전극 사이를 절연하기 위한 유전물질이 이용될 수 있다. 실시예로서, 수직 유전체층(123)은 SiO₂, HfO₂, ZrO₂, Si₃N₄ 및 Al₂O₃ 중 적어도 하나를 포함할 수 있다.
- [0068] 본 발명의 다른 실시예에 따르면, 수평 도전 라인(111)들과 수직 도전 라인(122)으로 형성된 커패시터에 축적된 전하를 시냅틱 가중치(synaptic weight)로 활용한다. 실시예로서, 수평 도전 라인(111)과 전기적으로 접속되는 라인을 통해 커패시터에 저장된 전하량을 출력할 수 있다.
- [0069] 본 발명의 다른 실시예에 따르면, 가중치 소자(30)는 도 2 및 도 3의 수직 도전층(121) 대신 수직 도전 라인(122)을 포함한다. 따라서, 수평 도전 라인(111)들과 수직 도전 라인(122)으로 형성된 커패시터 각각에 축적된 전하를 시냅틱 가중치(synaptic weight)로 활용할 수 있어, 같은 크기의 수직 도전층을 포함하는 가중치 소자와 비교할 때 다수의 멀티레벨 시냅틱 가중치를 표현할 수 있다.
- [0070] 수직 도전 라인(122)과 수평 도전 라인(111)들이 교차하는 크로스 포인트는 커패시터가 형성되고 전하가 축적되게 된다. 축적된 전하는 방전 후에 적분기(integrator)로 보내지고 적분기에서는 합산되는 동시에 전압으로 변환된다. 전압으로 변환된 출력 신호는 문턱 전압 크기 이상이 될 때 다음 레이어를 구성하는 입력 노드(input node)로 전달되게 된다.
- [0071] 본 발명의 일 실시예에 따르면, 학습이나 training 후 최종으로 결정된 워드 라인 선택 개수의 정보는 외부 저장장치에 저장될 수 있다. 뉴럴 네트워크 회로 안에 저장기능을 장착하고자 할 때에는 선택 트랜지스터로서 플로팅 게이트 트랜지스터를 사용할 수 있다. 실시예로서, 역전파(back-propagation) 등의 알고리즘을 사용하면서 트랜스포즈 가중치 매트릭스(transpose weight matrix)를 이용하여 학습 시 가중치의 보정량을 구하고, 얻어진 가중치의 보정량에 비례하여 워드 라인을 선택할 수 있다.
- [0072] 도 6은 본 발명의 다른 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자의 동작을 개략적으로 나타내는 회로도이다.
- [0073] 도 6을 참조하면, 본 발명의 다른 실시예에 따른 가중치 소자는 낸드 플래시를 활용할 수 있다. 패스 트랜지스터(Pass transistor)와 디코더는 낸드 플래시를 저장장치로 사용할 때 필요하므로 시냅틱 소자로 사용할 때에는 작동하지 않는다. 실시예로서, 낸드 플래시를 시냅틱 소자로 사용할 때는 모든 낸드 셀을 charging 혹은 discharging 상태로 하고 게이트 옥사이드(gate oxide)의 커패시턴스를 가중치로 이용할 수 있다. 플로팅 게이트 트랜지스터는 discharging 되어 있는 것이 바람직할 수 있다.
- [0074] 본 발명의 다른 실시예에 따른 가중치 소자를 동작하기 위해 입력 전압 펄스는 소스 라인으로 인가되고, 출력 전하는 워드 라인으로 방전된다. 실시예로서, 워드 라인에 선택 트랜지스터 또는 플로팅 게이트 트랜지스터를 부착하고 저장 기능을 수행하게 할 수 있다. 다른 실시예로서, 플로팅 게이트 트랜지스터를 패스 트랜지스터 및 디코더 반대편에 배치하여 낸드 플래시를 저장장치/가중치 소자 겸용으로 사용할 수 있다.
- [0075] 전술한 바와 같은 본 발명의 실시예들에 따르면, 본 발명의 일 실시예에 따른 크로스-포인트 커패시터 기반의 가중치 소자의 구조 및 이를 이용한 뉴럴 네트워크는 수직 적층형 크로스-포인트 커패시터 셀을 이용한 선형적인 다중 시냅틱 가중치를 가짐으로써, 학습 효율을 증가시킬 수 있다. 또한, 커패시터를 가중치로 사용함으로써 저항 가중치의 저항 값을 변화시켜 이에 비례하는 전류를 출력 신호로 사용하는 전도도 기반의(conductance-based) 가중치 소자에 비하여 전력 소모를 크게 감소시킬 수 있다. 또한, CMOS 기술을 사용하되 기존의 CMOS 소자의 작동 방식과 설계를 변형함으로써 시스템 크기를 줄이고 제품 공정 적용이 용이하게 할 수 있다.
- [0076] 이상과 같이 본 발명에서는 구체적인 구성 요소 등과 같은 특정 사항들과 한정된 실시예 및 도면에 의해 설명되었으나 이는 본 발명의 보다 전반적인 이해를 돕기 위해서 제공된 것일 뿐, 본 발명은 상기의 실시예에 한정되는 것은 아니며, 본 발명이 속하는 분야에서 통상적인 지식을 가진 자라면 이러한 기재로부터 다양한 수정 및 변형이 가능하다.

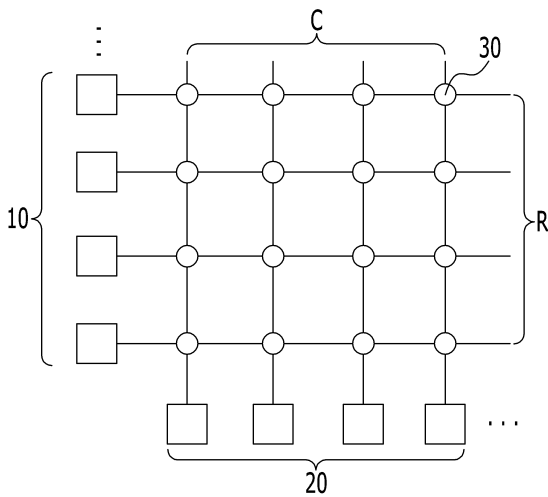
[0077] 따라서, 본 발명의 사상은 설명된 실시예에 국한되어 정해져서는 아니되며, 후술하는 특허청구범위뿐 아니라 이 특허청구범위와 균등하거나 등가적 변형이 있는 모든 것들은 본 발명 사상의 범주에 속한다고 할 것이다.

부호의 설명

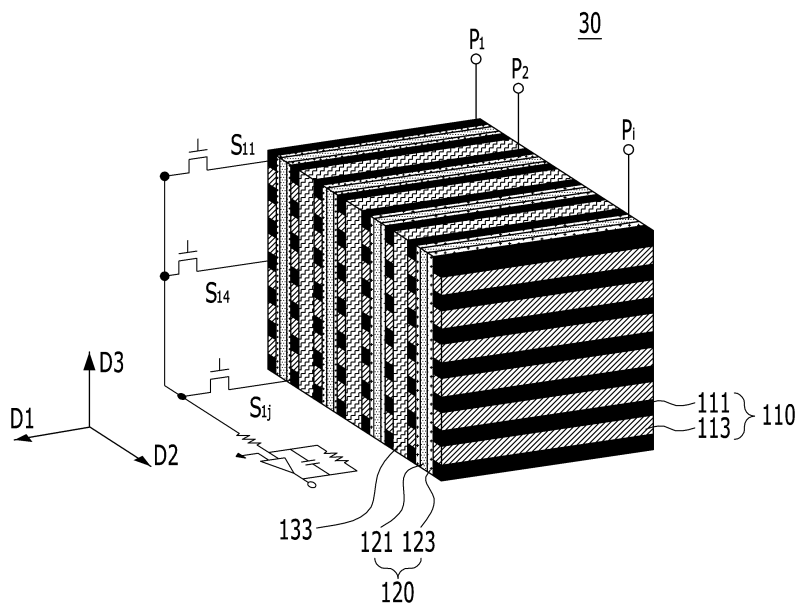
- [0078] 10: 입력 뉴런 20: 출력 뉴런
 30: 가중치 소자 110: 단위 수평 적층 구조체
 111: 수평 도전 라인 113: 수평 절연 라인
 120: 단위 수직 적층 구조체 121: 수직 도전층
 122: 수직 도전 라인 123: 수직 유전체층
 124: 수직 절연 라인 133: 수직 절연층

도면

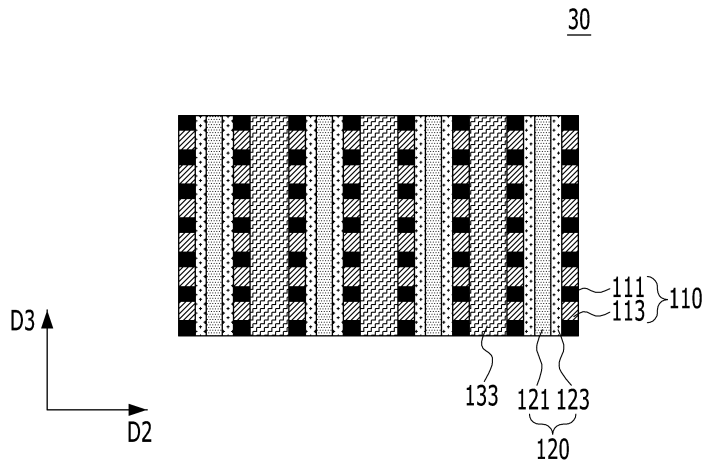
도면1



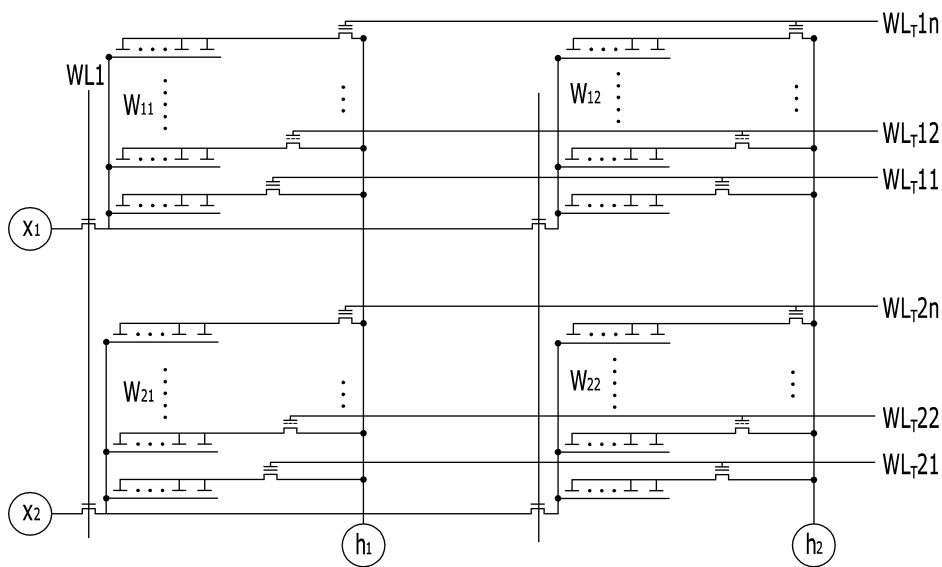
도면2



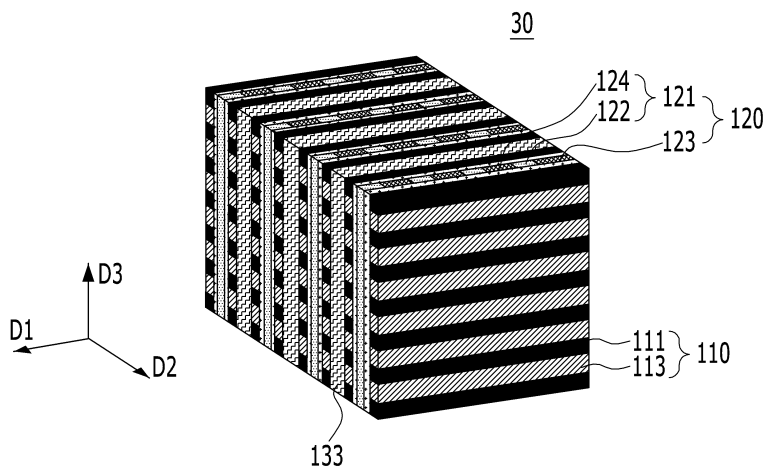
도면3



도면4



도면5



도면6

