(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0124830
(43) 공개일자 2024년08월19일

(51) 국제특허분류(Int. Cl.)
H10N 70/20 (2023.01) *G06F 17/15* (2006.01)
G06N 3/065 (2023.01) *H10N 70/00* (2024.01)
(52) CPC특허분류
H10N 70/253 (2024.05)
G06F 17/153 (2013.01)
(21) 출원번호 10-2024-0020513
(22) 출원일자 2024년02월13일
심사청구일자 2024년02월13일
(30) 우선권주장
1020230017639 2023년02월09일 대한민국(KR)

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
왕건욱
서울특별시 강북구 삼양로19길 25, 112동 1103호
장진곤
서울특별시 성북구 안암로 145, R&D 센터 509호
(뒷면에 계속)
(74) 대리인
특허법인시공

전체 청구항 수 : 총 20 항

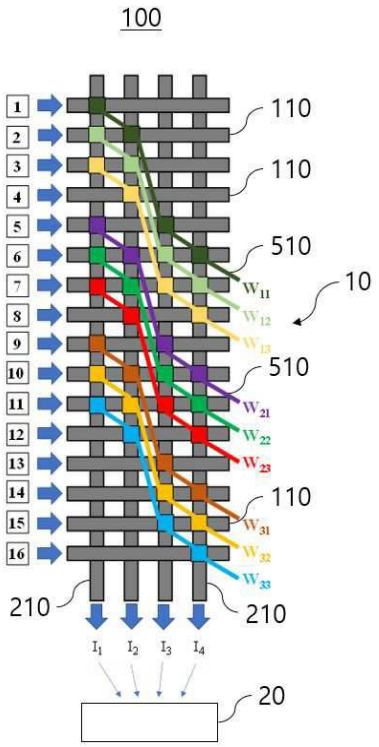
(54) 발명의 명칭 **3단자 멤트랜지스터 어레이의 다중 셀 동시적 웨이트 학습에 의한 에너지 효율적인 원-스텝 컨볼루션 연산 장치 및 방법**

(57) 요 약

3단자 멤트랜지스터 어레이 소자를 기반으로 컨볼루션 및 풀링 처리를 원-스텝으로 수행하고 동시적 다중 셀 업데이트 기술을 실현하여, 컨볼루션 연산 및 풀링 과정에 소모되는 시간과 에너지를 줄여 전체 시스템의 효율성을 높이는 에너지 효율적인 원-스텝 컨볼루션 연산 장치 및 방법이 개시된다. 본 발명의 실시예에 따른 컨볼루션 연

(뒷면에 계속)

대 표 도 - 도2a



산 장치는 입력 데이터가 입력되는 입력 라인들, 상기 입력 라인들과 교차하는 출력 라인들, 및 상기 입력 라인들과 상기 출력 라인들의 교차 영역에 형성되는 다수의 셀을 포함하는 멤트랜지스터 시냅스 어레이; 및 상기 멤트랜지스터 시냅스 어레이에 의해 출력되는 출력 값들을 풀링 처리하는 풀링 처리부를 포함한다. 상기 멤트랜지스터 시냅스 어레이는, 입력 라인들에 의해 입력 데이터를 입력받아 멀티-컬럼의 출력 라인들을 통해 컬럼 단위로 컨볼루션 특징 맵의 출력 값들을 출력하도록 구성된다. 상기 풀링 처리부는, 상기 멤트랜지스터 시냅스 어레이로부터 상기 컬럼 단위로 출력되는 출력 값들을 풀링 처리하여 컨볼루션 특징을 출력하도록 구성된다.

(52) CPC특허분류
G06N 3/065 (2023.01)
H10N 70/011 (2023.02)

(72) 발명자
함성길
인천광역시 부평구 후정로 12, 103동 404호

김용준
경기도 의정부시 호암로 256, 111동 204호

이 발명을 지원한 국가연구개발사업
과제고유번호 1711188653
과제번호 2022R1A2B5B02001455
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 개인기초연구(과기정통부)
연구과제명 초저전력 학습 및 3차원 집적화 가능한 확률제어 시냅스 및 뉴런 소재/소자 어레이 개발
과제수행기관명 고려대학교
연구기간 2023.03.01 ~ 2024.02.29
이 발명을 지원한 국가연구개발사업
과제고유번호 1711135355
과제번호 2021M3H4A1A01079420
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 나노·소재기술개발(R&D)
연구과제명 신개념 대각선-게이트 멤트랜지스터 어레이를 통한 극저전력 원-스텝 컨볼루션 신경망 원천 기술 개발
과제수행기관명 고려대학
연구기간 2021.05.10 ~ 2021.12.31
이 발명을 지원한 국가연구개발사업
과제고유번호 1711197992
과제번호 00220077
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 집단연구지원
연구과제명 칩온어파이버 기반 인터랙티브 전자섬유기술
과제수행기관명 전북대학교산학협력단
연구기간 2023.06.01 ~ 2024.02.29

명 세 서

청구범위

청구항 1

입력 데이터가 입력되는 입력 라인들, 상기 입력 라인들과 교차하는 출력 라인들, 및 상기 입력 라인들과 상기 출력 라인들의 교차 영역에 형성되는 다수의 셀을 포함하는 멤트랜지스터 시냅스 어레이; 및

상기 멤트랜지스터 시냅스 어레이에 의해 출력되는 출력 값들을 풀링 처리하는 풀링 처리부를 포함하고,

상기 멤트랜지스터 시냅스 어레이는, 입력 라인들에 의해 입력 데이터를 입력받아 멀티-컬럼의 출력 라인들을 통해 컬럼 단위로 컨볼루션 특징 맵의 출력 값들을 출력하도록 구성되고,

상기 풀링 처리부는, 상기 멤트랜지스터 시냅스 어레이로부터 상기 컬럼 단위로 출력되는 상기 멀티-컬럼의 출력 값들을 풀링 처리하여 컨볼루션 특징을 출력하도록 구성되는, 컨볼루션 네트워크 연산 장치.

청구항 2

청구항 1에 있어서,

상기 입력 데이터는 컨볼루션 처리 대상인 입력 이미지의 셀 값들에 대응되고,

상기 출력 라인들의 멀티-컬럼은 상기 입력 이미지 내에 포함되는 다수의 상이한 서브 영역들과 대응되는, 컨볼루션 네트워크 연산 장치.

청구항 3

청구항 2에 있어서,

상기 서브 영역들은 각각 컨볼루션 연산에 사용되는 커널 필터의 크기를 가지는, 컨볼루션 네트워크 연산 장치.

청구항 4

청구항 2에 있어서,

상기 멀티-컬럼의 각 컬럼은,

상기 입력 라인들 중 각 컬럼에 대응되는 서브 영역에 해당하는 제1 입력 라인들로부터 입력되는 제1 입력 데이터에 따라 상기 컨볼루션 특징 맵의 출력 값을 출력하고;

상기 입력 라인들 중 각 컬럼의 서브 영역에 대응되지 않는 제2 입력 라인들로부터 입력되는 제2 입력 데이터는 상기 컨볼루션 특징 맵의 출력에 반영되지 않는, 컨볼루션 네트워크 연산 장치.

청구항 5

청구항 2에 있어서,

상기 멀티-컬럼의 각 컬럼에 대응되는 상기 제1 입력 라인들은 컬럼에 따라 상이한, 컨볼루션 네트워크 연산 장치.

청구항 6

청구항 2에 있어서,

상기 멀티-컬럼의 제1 컬럼은 상기 입력 라인들 중 제1 입력 라인 그룹의 입력 데이터를 기초로 제1 컨볼루션 특징 맵을 출력하고,

상기 멀티-컬럼의 제2 컬럼은 상기 입력 라인들 중 상기 제1 입력 라인 그룹과 상이한 제2 입력 라인 그룹의 입력 데이터를 기초로 제2 컨볼루션 특징 맵을 출력하고,

상기 제2 입력 라인 그룹은 상기 제1 입력 라인 그룹으로부터 하나 이상의 입력 라인 단위로 쉬프트되도록 설정

되는, 컨볼루션 네트워크 연산 장치.

청구항 7

청구항 6에 있어서,

상기 제1 입력 라인 그룹 및 상기 제2 입력 라인 그룹에 각각 포함되는 입력 라인의 개수는 상기 커널 필터의 셀 개수와 동일한, 컨볼루션 네트워크 연산 장치.

청구항 8

청구항 2에 있어서,

상기 멤트랜지스터 시냅스 어레이는,

상기 다수의 셀 중 일부인 다중 셀에 공통의 학습 신호를 인가하는 하나 이상의 게이트 전극 라인을 포함하는, 컨볼루션 네트워크 연산 장치.

청구항 9

청구항 8에 있어서,

상기 멤트랜지스터 시냅스 어레이는,

상기 게이트 전극 라인에 의해 인가되는 게이트 전압에 의해 동시에 상기 다중 셀에 대한 웨이트 업데이트를 수행하는, 컨볼루션 네트워크 연산 장치.

청구항 10

청구항 9에 있어서,

상기 다중 셀은 컨볼루션 연산을 위해 상기 커널 필터의 동일한 셀에 의해 공통적으로 연산 처리되는 셀들인, 컨볼루션 네트워크 연산 장치.

청구항 11

청구항 8에 있어서,

상기 멤트랜지스터 시냅스 어레이는,

상기 게이트 전극 라인을 따라 배열된 다수의 시냅스 소자의 가중치를 동시에 학습하기 위한 가중치 제어 신호를 상기 게이트 전극 라인에 인가하도록 구성되는 게이트 전극을 더 포함하는, 컨볼루션 네트워크 연산 장치.

청구항 12

청구항 8에 있어서,

상기 입력 라인들 및 상기 출력 라인들 중 적어도 일부는 그래핀 물질로 이루어지고, 다수의 셀은 각각 시냅스 소자를 포함하고,

상기 시냅스 소자는 반도체 물질을 포함하는 반도체층; 및 상기 반도체층 상에 마련되어 페르미 준위를 조절하는 강유전체 물질로 이루어지는 조절층을 포함하는, 컨볼루션 네트워크 연산 장치.

청구항 13

청구항 12에 있어서,

상기 게이트 전극 라인을 통해 인가되는 학습 신호에 따른 상기 강유전체 물질의 잔류 분극 조절을 통해 상기 그래핀 물질과 상기 반도체 물질 사이의 쇼트키 배리어가 조절되어 아날로그 컨덕턴스 변화를 나타내는, 컨볼루션 네트워크 연산 장치.

청구항 14

청구항 8에 있어서,

상기 다수의 셀은 각각 시냅스 소자를 포함하고,

상기 시냅스 소자는 반도체 물질을 포함하는 반도체층; 및 상기 반도체층 상에 마련되는 강유전체 물질로 이루어지는 조절층을 포함하고,

상기 시냅스 소자는 상기 게이트 전극 라인을 통해 인가되는 학습 신호에 따른 상기 강유전체 물질의 잔류 분극 조절을 통해 상기 반도체층의 채널 경계의 반송자 밀도가 조절되어 시냅스 가소성이 구현되는, 컨볼루션 네트워크 연산 장치.

청구항 15

청구항 12 내지 청구항 14 중 어느 한 항에 있어서,

상기 반도체층은 IGZO(indium gallium zinc oxide)층, ZnO 층, 및 펜타센 층 중의 적어도 하나를 포함하는, 컨볼루션 네트워크 연산 장치.

청구항 16

청구항 12 내지 청구항 14 중 어느 한 항에 있어서,

상기 조절층은 PVDF-TrFE(poly-vinylidene fluoride-trifluoro-ethylene), HZO, PVDF, PZT 및 페로브스카이트 구조를 가지는 산화물 중의 적어도 하나의 강유전체 물질을 포함하는, 컨볼루션 네트워크 연산 장치.

청구항 17

(a) 입력 라인들 및 출력 라인들의 교차 영역에 시냅스 소자로 이루어지는 다수의 셀을 포함하는 멤트랜지스터 시냅스 어레이가, 상기 입력 라인들에 입력되는 입력 데이터에 따라 멀티-컬럼의 출력 라인들을 통해 컬럼 단위로 컨볼루션 특징 맵의 출력 값을 출력하는 단계; 및

(b) 풀링 처리부가, 상기 멤트랜지스터 시냅스 어레이로부터 상기 컬럼 단위로 출력되는 상기 멀티-컬럼의 출력 값들을 풀링 처리하여 컨볼루션 특징을 출력하는 단계를 포함하는, 컨볼루션 네트워크 연산 방법.

청구항 18

청구항 17에 있어서,

하나 이상의 게이트 전극 라인에 의해, 상기 다수의 셀 중 일부인 다중 셀에 공통의 학습 신호를 인가하는 단계를 더 포함하는, 컨볼루션 네트워크 연산 방법.

청구항 19

청구항 18에 있어서,

상기 멤트랜지스터 시냅스 어레이가, 상기 게이트 전극 라인에 의해 인가되는 게이트 전압에 의해 동시에 상기 다중 셀에 대한 웨이트 업데이트를 수행하는, 컨볼루션 네트워크 연산 방법.

청구항 20

청구항 19에 있어서,

상기 다중 셀은 컨볼루션 연산을 위해 상기 커널 필터의 동일한 셀에 의해 공통적으로 연산 처리되는 셀들인, 컨볼루션 네트워크 연산 방법.

발명의 설명

기 술 분 야

[0001] 본 발명은 효율적인 객체 인식/학습을 수행하는 컨볼루션 뉴럴 네트워크 시스템을 위한 3단자 멤트랜지스터 어레이 기반의 컨볼루션 연산 장치 및 방법에 관한 것으로, 보다 상세하게는 원-스텝 컨볼루션 연산 및 풀링 처리와, 다중 셀의 동시적 웨이트 학습에 의해 에너지 효율적으로 데이터 처리를 수행하는 원-스텝 컨볼루션 연산 장치 및 방법에 관한 것이다.

배 경 기 술

[0002] 객체 인식/판단 기술은 얼굴인식, 자율주행, 파형분석 등 빅데이터 응용 분야의 많은 부분을 차지하고 있는 핵심적인 요소이다. 현재 가장 발전된 객체 인식 시스템은 컨볼루션 뉴럴 네트워크를 통해 신뢰성 높은 입력 데이터 처리 기술을 사용하고 있다. 컨볼루션 뉴럴 네트워크 시스템은 여러 개의 필터를 통한 입력 신호 컨볼루션 변환을 통해 컨볼루션 특징 맵을 생성하며, 이를 기반으로 수렴적 연산 과정을 통해 정확하고 높은 학습 성능을 나타내고 있다. 이러한 컨볼루션 네트워크의 연산 소자로, 비휘발성 아날로그 스위칭이 가능한 멤리스터 및 멤트랜지스터 어레이가 각광받고 있다. 이는 신호 전달 매체 없이 자체 연산 및 데이터 저장이 가능한 장점으로 인해 기존의 CMOS 컴퓨팅에 비해 높은 에너지 효율성을 갖는다. 하지만 현재의 컨볼루션 변환 기술은 이미지나 필터의 크기가 커지거나 레이어가 많아질수록 연산해야 하는 범위가 늘어나게 되어 에너지 소모가 극심해지게 된다.

[0003] 도 1a는 종래의 컨볼루션 변환 과정을 나타낸 모식도이다. 도 1b는 종래의 순차적 웨이트 업데이트 과정을 나타낸 모식도이다. 종래의 컨볼루션 변환 과정은 예를 들어 4x4 입력 이미지에 3x3 커널 필터 연산이 적용될 경우, 총 4회의 슬라이딩 과정이 반복되며, 이러한 슬라이딩 과정에 따라 고정된 하나의 커널 시냅스 웨이트 컬럼에서 입력 이미지의 전압 벡터를 변경하여 인가하는 과정을 반복적으로 수행해야 한다.

[0004] 이와 같이 커널 필터의 슬라이딩을 순차적으로 반복하면서 컨볼루션 연산을 통해 슬라이딩 단계 별로 하나씩 특징 맵을 추출하면 4개의 특징 맵 성분이 모이게 되고, 다음으로 특징 맵 성분들을 최대 또는 평균으로 풀링 연산하여 작은 단위의 특징 맵으로 정방향 수렴하는 과정을 수행한다. 이로 인해 각각의 최대값 또는 평균값을 추출하는 풀링 과정에서의 정보 저장과 재연산 과정이 수반되기 때문에 효율적인 변환이 어려워진다. 이미지 단위 또는 필터가 커질수록 이러한 순차적인 컨볼루션 변환 과정은 매우 큰 연산량과 연산 에너지를 수반하게 된다.

[0005] 또한, 기존의 2단자 크로스바 어레이를 멀티-컬럼 컨볼루션 연산 소자로 사용할 경우, 도 1b와 같이 순차적인 2단자 어레이의 다중 셀 웨이트 업데이트가 수행되어야 한다. 이 경우, 같은 커널 시냅스 웨이트를 가져야 하는 셀들에 대해 각각 독립적인 쓰기 전압을 인가해야 하며, 2단자 워드-비트 라인을 이용하기 때문에 누설 전류로 인한 인접 셀 웨이트 영향을 야기하여 정확한 다중 셀 웨이트 구현이 어려워진다.

발명의 내용

해결하려는 과제

[0006] 본 발명은 3단자 멤트랜지스터 어레이 소자를 기반으로 컨볼루션 및 풀링 처리를 원-스텝으로 수행하고 동시적 다중 셀 업데이트 기술을 실현하여, 컨볼루션 연산 및 풀링 과정에 소모되는 시간과 에너지를 줄여 전체 시스템의 효율성을 높이는 에너지 효율적인 원-스텝 컨볼루션 연산 장치 및 방법을 제공하기 위한 것이다.

[0007] 또한, 본 발명은 일함수 변조 가능한 그래핀이 적용된 배리스터 타입 또는 강유전체의 분극 분역을 조절하여 채널 컨덕턴스 조절이 가능한 FeFET 타입으로 구현되어 소자 컨덕턴스 값을 정밀하고 정확하게 조절 가능한 수직 3단자 구조의 인공 시냅스 소자를 제공하기 위한 것이다.

[0008] 한편, 본 발명에서 이루고자 하는 기술적 과제들은 이상에서 언급한 기술적 과제들로 제한되지 않으며, 언급하지 않은 또 다른 기술적 과제들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0009] 본 발명의 실시예에 따른 컨볼루션 네트워크 연산 장치는 입력 데이터가 입력되는 입력 라인들, 상기 입력 라인들과 교차하는 출력 라인들, 및 상기 입력 라인들과 상기 출력 라인들의 교차 영역에 형성되는 다수의 셀을 포함하는 멤트랜지스터 시냅스 어레이; 및 상기 멤트랜지스터 시냅스 어레이에 의해 출력되는 출력 값들을 풀링 처리하는 풀링 처리부를 포함한다.

[0010] 상기 멤트랜지스터 시냅스 어레이는, 입력 라인들에 의해 입력 데이터를 입력받아 멀티-컬럼의 출력 라인들을 통해 컬럼 단위로 컨볼루션 특징 맵의 출력 값들을 출력하도록 구성된다.

[0011] 상기 풀링 처리부는, 상기 멤트랜지스터 시냅스 어레이로부터 상기 컬럼 단위로 출력되는 상기 멀티-컬럼의 출력 값들을 풀링 처리하여 컨볼루션 특징을 출력하도록 구성된다.

[0012] 상기 입력 데이터는 컨볼루션 처리 대상인 입력 이미지의 셀 값들에 대응될 수 있다. 상기 출력 라인들의 멀티-컬럼은 상기 입력 이미지 내에 포함되는 다수의 상이한 서브 영역들과 대응될 수 있다. 상기 서브 영역들은 각각 컨볼루션 연산에 사용되는 커널 필터의 크기를 가질 수 있다.

[0013] 상기 멀티-컬럼의 각 컬럼은, 상기 입력 라인들 중 각 컬럼에 대응되는 서브 영역에 해당하는 제1 입력 라인들로부터 입력되는 제1 입력 데이터에 따라 상기 컨볼루션 특징 맵의 출력 값을 출력할 수 있다. 상기 입력 라인들 중 각 컬럼의 서브 영역에 대응되지 않는 제2 입력 라인들로부터 입력되는 제2 입력 데이터는 상기 컨볼루션 특징 맵의 출력에 반영되지 않을 수 있다. 상기 멀티-컬럼의 각 컬럼에 대응되는 상기 제1 입력 라인들은 컬럼에 따라 상이할 수 있다.

[0014] 상기 멀티-컬럼의 제1 컬럼은 상기 입력 라인들 중 제1 입력 라인 그룹의 입력 데이터를 기초로 제1 컨볼루션 특징 맵을 출력하고, 상기 멀티-컬럼의 제2 컬럼은 상기 입력 라인들 중 상기 제1 입력 라인 그룹과 상이한 제2 입력 라인 그룹의 입력 데이터를 기초로 제2 컨볼루션 특징 맵을 출력할 수 있다. 상기 제2 입력 라인 그룹은 상기 제1 입력 라인 그룹으로부터 하나 이상의 입력 라인 단위로 쉬프트되도록 설정될 수 있다. 상기 제1 입력 라인 그룹 및 상기 제2 입력 라인 그룹에 각각 포함되는 입력 라인의 개수는 상기 커널 필터의 셀 개수와 동일할 수 있다.

[0015] 상기 멤트랜지스터 시냅스 어레이는, 상기 다수의 셀 중 일부인 다중 셀에 공통의 학습 신호를 인가하는 하나 이상의 게이트 전극 라인을 포함할 수 있다. 상기 멤트랜지스터 시냅스 어레이는, 상기 게이트 전극 라인에 의해 인가되는 게이트 전압에 의해 동시에 상기 다중 셀에 대한 웨이트 업데이트를 수행할 수 있다. 상기 다중 셀은 컨볼루션 연산을 위해 상기 커널 필터의 동일한 셀에 의해 공통적으로 연산 처리되는 셀들일 수 있다.

[0016] 상기 멤트랜지스터 시냅스 어레이는, 상기 게이트 전극 라인을 따라 배열된 다수의 시냅스 소자의 가중치를 동시에 학습하기 위한 가중치 제어 신호를 상기 게이트 전극 라인에 인가하도록 구성되는 게이트 전극을 더 포함할 수 있다.

[0017] 일 실시예에서, 상기 입력 라인들 및 상기 출력 라인들 중 적어도 일부는 그래핀 물질로 이루어질 수 있다. 다수의 셀은 각각 시냅스 소자를 포함할 수 있다. 상기 시냅스 소자는 반도체 물질을 포함하는 반도체층; 및 상기 반도체층 상에 마련되어 페르미 준위를 조절하는 강유전체 물질로 이루어지는 조절층을 포함할 수 있다.

[0018] 상기 시냅스 소자는, 상기 게이트 전극 라인을 통해 인가되는 학습 신호에 따른 상기 강유전체 물질의 잔류 분극 조절을 통해 상기 그래핀 물질과 상기 반도체 물질 사이의 쇼트키 배리어가 조절되어 아날로그 컨덕턴스 변화를 나타낼 수 있다.

[0019] 다른 실시예에서, 상기 시냅스 소자는 반도체 물질을 포함하는 반도체층; 및 상기 반도체층 상에 마련되는 강유전체 물질로 이루어지는 조절층을 포함하고, 상기 게이트 전극 라인을 통해 인가되는 학습 신호에 따른 상기 강유전체 물질의 잔류 분극 조절을 통해 상기 반도체층의 채널 경계의 반송자 밀도가 조절되어 시냅스 가소성이 구현될 수 있다.

[0020] 상기 반도체층은 IGZO(indium gallium zinc oxide)층, ZnO 층, 및 펜타센 층 중의 적어도 하나를 포함할 수 있다.

[0021] 상기 조절층은 PVDF-TrFE(poly-vinylidene fluoride-trifluoro-ethylene), HZO, PVDF, PZT 및 페로브스카이트 구조를 가지는 산화물 중의 적어도 하나의 강유전체 물질을 포함할 수 있다.

[0022] 본 발명의 실시예에 따른 컨볼루션 네트워크 연산 방법은 (a) 입력 라인들 및 출력 라인들의 교차 영역에 시냅스 소자로 이루어지는 다수의 셀을 포함하는 멤트랜지스터 시냅스 어레이가, 상기 입력 라인들에 입력되는 입력 데이터에 따라 멀티-컬럼의 출력 라인들을 통해 컬럼 단위로 컨볼루션 특징 맵의 출력 값을 출력하는 단계; 및 (b) 풀링 처리부가, 상기 멤트랜지스터 시냅스 어레이로부터 상기 컬럼 단위로 출력되는 상기 멀티-컬럼의 출력 값들을 풀링 처리하여 컨볼루션 특징을 출력하는 단계를 포함한다.

[0023] 본 발명의 실시예에 따른 컨볼루션 네트워크 연산 방법은 하나 이상의 게이트 전극 라인에 의해, 상기 다수의 셀 중 일부인 다중 셀에 공통의 학습 신호를 인가하는 단계를 더 포함할 수 있다.

[0024] 본 발명의 실시예에 따른 컨볼루션 네트워크 연산 방법은 상기 멤트랜지스터 시냅스 어레이가, 상기 게이트 전극 라인에 의해 인가되는 게이트 전압에 의해 동시에 상기 다중 셀에 대한 웨이트 업데이트를 수행하는 단계를 더 포함할 수 있다.

발명의 효과

[0025] 본 발명의 실시예에 의하면, 3단자 멤트랜지스터 어레이 소자를 기반으로 컨볼루션 및 풀링 처리를 원-스텝으로 수행하고 동시적 다중 셀 업데이트 기술을 실현하여, 컨볼루션 연산 및 풀링 과정에 소모되는 시간과 에너지를 줄여 전체 시스템의 효율성을 높일 수 있다.

[0026] 또한, 본 발명의 실시예에 의하면, 게이트 전극 라인에 의해 3단자 멤트랜지스터 시냅스 어레이의 다중 셀을 동시에 업데이트할 수 있으며, 쓰기 과정에서의 인접 셀 영향을 줄일 수 있다. 이에 따라 3단자 멤트랜지스터 시냅스 어레이에 의해 다중 셀 동시 학습을 정확하고 효율적으로 구현할 수 있다.

[0027] 또한, 본 발명의 실시예에 의하면, 컨볼루션 처리에 필요한 이미지 슬라이딩 횟수를 감소시키고, 데이터 입출력 프로세스 횟수를 감소시킴으로써 컨볼루션 연산 및 풀링 과정에 소모되는 시간과 에너지를 줄여 전체 시스템의 효율성을 높일 수 있다.

[0028] 또한, 본 발명의 실시예에 의하면, 일함수 변조 가능한 그래핀이 적용된 배리스터 타입 또는 강유전체의 분극 분역을 조절하여 채널 컨덕턴스 조절이 가능한 FeFET 타입으로 구현되어 소자 컨덕턴스 값을 정밀하고 정확하게 조절 가능한 수직 3단자 구조의 인공 시냅스 소자가 제공된다.

[0029] 한편, 본 발명에서 얻을 수 있는 효과는 이상에서 언급한 효과들로 제한되지 않으며, 언급하지 않은 또 다른 효과들은 아래의 기재로부터 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0030] 도 1a는 종래의 대표적인 4개의 인접한 컨볼루션 이미지 픽셀 변환 과정을 나타낸 모식도이다.

도 1b는 종래의 순차적 웨이트 업데이트 과정을 나타낸 모식도이다.

도 2a는 본 발명의 실시예에 따른 원-스텝 컨볼루션 연산 장치를 나타낸 개념도이다.

도 2b는 본 발명의 실시예에 따른 원-스텝 컨볼루션 연산 방법을 나타낸 개념도이다.

도 2c는 본 발명의 실시예에 따른 원-스텝 컨볼루션 연산 장치를 구성하는 3단자 멤트랜지스터 어레이의 동시적 게이트 전극 라인 업데이트 방식을 나타낸 개념도이다.

도 3은 본 발명의 일 실시예에 따른 3단자 멤트랜지스터 시냅스 어레이에 의한 원-스텝 컨볼루션 및 풀링 과정을 나타낸 개념도이다.

도 4는 본 발명의 실시예에 따른 3단자 멤트랜지스터 시냅스 어레이를 이용한 원-스텝 연산 과정과, 종래의 2단자 어레이를 통한 투-스텝 컨볼루션 및 풀링 과정을 비교하여 나타낸 개념도이다.

도 5는 본 발명의 일 실시예에 따른 3단자 멤트랜지스터 어레이 구조를 나타낸 도면이다.

도 6은 본 발명의 일 실시예에 따른 3단자 멤트랜지스터 어레이 구조에 포함된 멤트랜지스터 소자를 확대하여 나타낸 도면이다.

도 7은 본 발명의 일 실시예에 따른 3단자 멤트랜지스터 소자 시스템 제조 방법에 의해 3단자 멤트랜지스터 시스템이 제조되는 과정을 순차적으로 나타낸 도면이다.

도 8은 4x4 입력 데이터와 3x3 커널(컨볼루션 필터)의 조건 하에서, 기존의 CNN 어레이의 구동 방식을 개략적으로 나타낸 도면이다.

도 9는 도 8의 과정을 보다 상세히 나타낸 도면으로, 컨볼루션 필터의 슬라이딩에 따른 입력 데이터에 대한 컨볼루션 연산 과정을 순차적으로 나타낸 도면이다.

도 10은 도 8과 같은 조건 하에서, 본 발명의 실시예에 따른 3단자 멤트랜지스터 시스템에 의한 컨볼루션 연산 과정을 개략적으로 나타낸 도면이다.

도 11은 도 10의 과정을 보다 상세히 나타낸 도면으로, 게이트 전극 라인을 갖는 3전극 멤트랜지스터 시스템에 의하여 특징 맵(feature map) 데이터를 이용하여 별도의 정보를 저장하는 과정 없이, 풀링(pooling) 과정을 통해 한번에 컨볼루션 연산 결과가 출력되는 모습을 나타낸 도면이다.

도 12는 도 10 및 도 11에 도시된 방식으로 소스 전극을 통해 입력 데이터를 입력받고, 드레인 전극을 통해 데이터가 출력되는 모습을 나타낸 예시도이다.

도 13은 기존의 시냅스 어레이 중 어느 하나의 멤트랜지스터 소자에 대하여 가중치 학습을 위한 펄스 시퀀스를 입력하는 모습을 나타낸 도면이다.

도 14는 본 발명의 일 실시예에 따른 3단자 멤트랜지스터 시스템에 있어서 한 번의 펄스 시퀀스 입력으로 복수의 3단자 멤트랜지스터 소자에 가중치를 학습시키는 모습을 나타낸 도면이다.

도 15는 강유전체층에 분극이 일어난 모습을 나타낸 도면이다.

도 16은 3단자 멤트랜지스터 소자의 게이트 전압에 따른 드레인-소스 전류 변화 특성을 나타낸 도면이다.

도 17은 3단자 멤트랜지스터 소자의 펄스 게이트 전압 인가에 따른 PSC(postsynaptic currents) 변화 특성을 나타낸 도면이다.

도 18a는 본 발명의 실시예에 따른 배리스터 타입 수직 3단자 인공 시냅스 소자를 나타낸 도면이다.

도 18b는 도 18a의 배리스터 타입 수직 3단자 인공 시냅스 소자의 레이아웃을 나타낸 도면이다.

도 18c 및 도 18d는 도 18a의 배리스터 타입 수직 3단자 인공 시냅스 소자의 단면도 및 층별 성분 원소 분석 결과를 나타낸 도면이다.

도 18e는 도 18a의 배리스터 타입 수직 3단자 인공 시냅스 소자의 라만 스펙트럼을 나타낸 도면이다.

도 18f는 도 18a의 배리스터 타입 수직 3단자 인공 시냅스 소자의 전류 밀도 및 유전체 특성을 나타낸 도면이다.

도 19a 내지 도 19i, 도 20a 내지 도 20g, 도 21a 내지 도 21c는 본 발명의 실시예에 따른 수직 3단자 인공 시냅스 소자의 다양한 전기적 특성을 나타낸 도면이다.

도 22a는 본 발명의 또 다른 실시예에 따른 수직 3단자 시냅스 소자를 나타낸 도면이다.

도 22b 내지 도 22e는 도 22a의 수직 3단자 시냅스 소자의 제조 과정을 나타낸 도면이다.

도 23a는 본 발명의 또 다른 실시예에 따른 수직 3단자 시냅스 소자를 나타낸 도면이다.

도 23b 내지 도 23f는 도 23a의 수직 3단자 시냅스 소자의 제조 과정을 나타낸 도면이다.

도 24a 내지 도 24c, 도 25a 내지 도 25f, 도 26a 내지 도 26e, 도 27a 내지 도 27g는 본 발명의 실시예에 따른 수직 3단자 시냅스 소자의 전기적 특성을 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0031] 본 발명의 다른 이점 및 특징, 그리고 그것들을 달성하는 방법은 첨부되는 도면과 함께 상세하게 후술되는 실시예를 참조하면 명확해질 것이다. 그러나 본 발명은 이하에서 개시되는 실시예에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있으며, 단지 본 실시예는 본 발명의 개시가 완전하도록 하고, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 발명의 범주를 완전하게 알려주기 위해 제공되는 것이며, 본 발명은 청구항의 범주에 의해 정의될 뿐이다.

[0032] 만일 정의되지 않더라도, 여기서 사용되는 모든 용어들(기술 혹은 과학 용어들을 포함)은 이 발명이 속한 종래 기술에서 보편적 기술에 의해 일반적으로 수용되는 것과 동일한 의미를 가진다. 일반적인 사전들에 의해 정의된 용어들은 관련된 기술 그리고/혹은 본 출원의 본문에 의미하는 것과 동일한 의미를 갖는 것으로 해석될 수 있고, 그리고 여기서 명확하게 정의된 표현이 아니더라도 개념화되거나 혹은 과도하게 형식적으로 해석되지 않을 것이다. 본 명세서에서 사용된 용어는 실시예들을 설명하기 위한 것이며 본 발명을 제한하고자 하는 것은 아니다. 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함한다.

[0033] 명세서에서 사용되는 '포함한다' 및/또는 이 동사의 다양한 활용형들 예를 들어, '포함', '포함하는', '포함하고', '포함하며' 등은 언급된 조성, 성분, 구성요소, 단계, 동작 및/또는 소자는 하나 이상의 다른 조성, 성분, 구성요소, 단계, 동작 및/또는 소자의 존재 또는 추가를 배제하지 않는다. 본 명세서에서 '및/또는' 이라는 용어는 나열된 구성들 각각 또는 이들의 다양한 조합을 가리킨다.

[0034] 한편, 본 명세서 전체에서 사용되는 '~부'는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미할 수 있다. 예를 들어 소프트웨어, FPGA 또는 ASIC과 같은 하드웨어 구성요소를 의미할 수 있다. 그렇지만 '~부'가 소프트웨어 또는 하드웨어에 한정되는 의미는 아니다. '~부'는 어드레싱할 수 있는 저장 매체에 있도록 구성될 수도 있고 하나 또는 그 이상의 프로세서들을 재생시키도록 구성될 수도 있다. 따라서, 일 예로서 '~부'는 소프트웨어 구성요소들, 객체지향 소프트웨어 구성요소들, 클래스 구성요소들 및 태스크 구성요소들과 같은 구성요소들과, 프로세스들, 함수들, 속성들, 프로시저들, 서브루틴들, 프로그램 코드의 세그먼트들, 드라이버들, 펌웨어, 마이크로 코드, 회로, 데이터, 데이터베이스, 데이터 구조들, 테이블들, 어레이들 및 변수들을 포함한다. 구성요소들과 '~부' 안에서 제공되는 기능은 더 작은 수의 구성요소들 및 '~부'들로 결합되거나 추가적인 구성요소들과 '~부'들로 더 분리될 수 있다.

[0035] 본 명세서에서 어떠한 구성요소 A가 다른 구성요소 B '상에' 위치하거나 형성된다는 것은 특별히 언급되지 않은 한, 그 구성요소 A가 또 다른 구성요소 C가 개재되지 않은 상태로 직접 구성요소 B 상에 위치 또는 형성될 수 있는 것은 물론, 구성요소 A와 구성요소 B 사이에 구성요소 C가 개재된 상태로 구성요소 A가 구성요소 B 상에 위치 또는 형성되는 것을 포괄하는 것으로 이해되어야 한다. 이하, 본 명세서의 첨부된 도면을 참조하여 본 발명의 실시예를 상세하게 설명한다.

[0036] 도 2a는 본 발명의 실시예에 따른 원-스텝 컨볼루션 연산 장치를 나타낸 개념도이다. 구체적으로, 도 2a는 3단자 대각선 세이트 형태의 어레이 소자 구조에 따른 원-스텝 컨볼루션 연산 장치를 나타낸 개념도이다. 도 2b는 본 발명의 실시예에 따른 원-스텝 컨볼루션 연산 방법을 나타낸 개념도이다. 도 2c는 본 발명의 실시예에 따른 원-스텝 컨볼루션 연산 장치를 구성하는 3단자 멤트랜지스터 어레이의 동시적 게이트 전극 라인 업데이트 방식을 나타낸 개념도이다. 도 2a, 도 2b 및 도 2c를 참조하여 설명하면, 본 발명의 실시예에 따른 원-스텝 컨볼루션 연산 장치(100)는 3단자 멤트랜지스터 시냅스 어레이로 이루어지는 멤트랜지스터 시스템(10)와, 폴링 처리부(20)를 포함할 수 있다.

[0037] 멤트랜지스터 시스템(10)의 3단자 멤트랜지스터 시냅스 어레이는 컨볼루션 연산을 위해 다수의 행과 열을 이루도록 배열되는 다수의 셀을 포함한다. 3단자 멤트랜지스터 시냅스 어레이는 다수의 소스 전극 라인(110), 다수의 소스 전극 라인(120)과 교차하도록 배열되는 다수의 드레인 전극 라인(210), 그리고 다수의 소스 전극 라인(110)과 다수의 드레인 전극 라인(210) 간의 교차 영역에 해당하는 셀마다 소스 전극 라인(110)과 드레인 전극 라인(210) 사이에 마련되는 시냅스 소자를 포함할 수 있다.

[0038] 멤트랜지스터 시스템(10)은 3단자 멤트랜지스터 시냅스 어레이에 의해 입력 라인들(예를 들어, 소스 전극 라인들)에 입력되는 입력 데이터로부터 멀티-컬럼의 출력 라인들(예를 들어, 드레인 전극 라인들)을 통해 컬럼 단위로 컨볼루션 특징 맵의 출력 값들을 출력할 수 있다. 멤트랜지스터 시스템(10)은 도 2c와 같이 시냅스 웨이트 정보가 필요한 다중 셀을 동시에 업데이트할 수 있으며, 입력 데이터에 대해 각각의 부분 커널 컨볼루션 연산 결과인 출력 값들을 동시에 추출할 수 있다.

[0039] 멤트랜지스터 시스템(10)은 컨볼루션 연산을 위한 다수의 셀을 포함하는 멤트랜지스터 시냅스 어레이와, 멤트랜지스터 시냅스 어레이의 다수의 셀 중 컨볼루션 필터의 슬라이딩과 관련하여 설정된 다중 셀의 웨이트를 동시적으로 학습하도록, 다중 셀에 공통의 학습 신호를 인가하는 게이트 전극 라인(510)을 포함할 수 있다.

[0040] 멤트랜지스터 시냅스 어레이는 입력 데이터가 입력되는 입력 라인들, 입력 라인들과 교차하는 출력 라인들, 및 입력 라인들과 출력 라인들의 교차 영역에 형성되는 다수의 셀을 포함할 수 있다. 게이트 전극 라인(510)은 멤트랜지스터 시냅스 어레이의 다수의 셀 중 입력 라인들과 출력 라인들을 기준으로 대각 방향으로 배열되는 다중 셀의 웨이트를 동시적으로 학습하도록, 다중 셀에 공통의 학습 신호를 인가할 수 있다. 입력 라인들에는 컨볼루션 연산을 위한 커널 필터의 슬라이딩 영역에 해당하는 입력 데이터가 입력될 수 있다.

[0041] 컨볼루션 연산을 위한 3단자 멤트랜지스터 시냅스 어레이는 다수의 셀 중 일부인 다중 셀(컨볼루션 필터의 각 셀에 의해 공통적으로 연산 처리되는 셀들)에 공통의 학습 신호를 인가하여 다중 셀의 웨이트를 동시적으로 학습할 수 있다. 입력 라인들에 입력되는 입력 데이터는 컨볼루션 처리 대상인 입력 이미지의 셀 값들에 대응될 수 있다. 도 2a의 예에서, 4x4 입력 이미지는 16개의 셀들로 이루어지므로, 입력 이미지는 16개의 셀 값들을 가진다. 이에 따라 총 16개의 입력 라인들에 1번 내지 16번 입력 데이터가 입력된다.

[0042] 출력 라인들의 멀티-컬럼은 입력 이미지 내에 포함되는 다수의 상이한 서브 영역들과 대응될 수 있다. 도 2a의 예에서, 입력 이미지의 크기는 4x4이고, 커널 필터의 크기는 3x3이므로, 스트라이드가 '1'로 설정될 경우 컨볼루션 연산은 입력 이미지 내의 4개의 서브 영역에 대해 수행된다. 이에 따라 멀티-컬럼의 컬럼 개수는 '4'로 설

정되며, 출력 라인들의 개수도 4개가 된다.

[0043] 입력 이미지 내의 서브 영역들은 다른 하나 이상의 서브 영역과 부분적으로 중첩될 수 있다. 즉, 입력 이미지 내의 제1 서브 영역은 입력 이미지 내의 제2 서브 영역과 일부 셀들이 공통되지만, 서브 영역 전체가 일치하지는 않는다. 서브 영역들은 각각 컨볼루션 연산에 사용되는 커널 필터의 크기를 가질 수 있다. 도시된 예에서, 커널 필터의 크기가 3x3이므로, 서브 영역의 크기도 3x3이 된다.

[0044] 입력 이미지의 4개의 서브 영역 중 첫번째 서브 영역은 입력 이미지 중의 좌측 상부 영역이고, 두번째 서브 영역은 입력 이미지 중의 우측 상부 영역이고, 세번째 서브 영역은 입력 이미지 중의 좌측 하부 영역이고, 네번째 서브 영역은 입력 이미지 중의 우측 하부 영역이다. 도시된 예에서, 입력 이미지, 커널 필터는 모두 행과 열의 개수가 동일한 정사각 크기로 이루어져 있으나, 행과 열의 개수가 상이하게 설정될 수도 있다.

[0045] 멀티-컬럼의 각 컬럼은 입력 라인들 중 각 컬럼에 대응되는 서브 영역에 해당하는 제1 입력 라인들로부터 입력되는 제1 입력 데이터에 따라 컨볼루션 특징 맵의 출력 값을 출력하고, 입력 라인들 중 각 컬럼의 서브 영역에 대응되지 않는 제2 입력 라인들로부터 입력되는 제2 입력 데이터는 컨볼루션 특징 맵의 출력에 반영되지 않을 수 있다.

[0046] 멀티-컬럼의 각 컬럼에 대응되는 제1 입력 라인들은 컬럼에 따라 상이할 수 있다. 멀티-컬럼의 제1 컬럼은 입력 라인들 중 제1 입력 라인 그룹의 입력 데이터를 기초로 제1 컨볼루션 특징 맵을 출력하고, 멀티-컬럼의 제2 컬럼은 입력 라인들 중 제1 입력 라인 그룹과 상이한 제2 입력 라인 그룹의 입력 데이터를 기초로 제2 컨볼루션 특징 맵을 출력할 수 있다. 제1 입력 라인 그룹 및 제2 입력 라인 그룹에 각각 포함되는 입력 라인의 개수는 커널 필터의 셀 개수(도시된 예에서는 9개)와 동일할 수 있다.

[0047] 도시된 예에서, 멀티-컬럼 중 첫번째 출력 라인에 해당하는 첫번째 컬럼(제1 컬럼)은 첫번째 서브 영역(제1 서브 영역)에 해당하는 1, 2, 3, 5, 6, 7, 9, 10, 11번 입력 라인들(제1 입력 라인 그룹)로부터 입력 데이터를 입력받는다. 4, 8, 12, 13, 14, 15 16번 입력 라인들에 입력되는 입력 데이터는 첫번째 컬럼(제1 컬럼)의 컨볼루션 특징 맵의 출력에 활용되지 않는다.

[0048] 멀티-컬럼 중 두번째 출력 라인에 해당하는 두번째 컬럼(제2 컬럼)은 두번째 서브 영역(제2 서브 영역)에 해당하는 2, 3, 4, 6, 7, 8, 10, 11, 12번 입력 라인들(제2 입력 라인 그룹)로부터 입력 데이터를 입력받는다. 제2 입력 라인 그룹은 제1 입력 라인 그룹으로부터 하나 이상의 입력 라인 단위(도시된 예에서는 하나의 입력 라인 단위)로 쉬프트되도록 설정될 수 있다. 1, 5, 9, 13, 14, 15, 16번 입력 라인들에 입력되는 입력 데이터는 두번째 컬럼(제2 컬럼)의 컨볼루션 특징 맵의 출력에 활용되지 않는다.

[0049] 멀티-컬럼 중 세번째 출력 라인에 해당하는 세번째 컬럼(제3 컬럼)은 세번째 서브 영역(제3 서브 영역)에 해당하는 5, 6, 7, 9, 10, 11, 13, 14, 15번 입력 라인들(제3 입력 라인 그룹)로부터 입력 데이터를 입력받는다. 제3 입력 라인 그룹은 제2 입력 라인 그룹으로부터 하나 이상의 입력 라인 단위(도시된 예에서는 세개의 입력 라인 단위)로 쉬프트되도록 설정될 수 있다. 2, 3, 4, 8, 12, 16번 입력 라인들에 입력되는 입력 데이터는 세번째 컬럼(제3 컬럼)의 컨볼루션 특징 맵의 출력에 활용되지 않는다.

[0050] 멀티-컬럼 중 네번째 출력 라인에 해당하는 네번째 컬럼(제4 컬럼)은 네번째 서브 영역(제4 서브 영역)에 해당하는 6, 7, 8, 10, 11, 12, 14, 15, 16번 입력 라인들(제4 입력 라인 그룹)로부터 입력 데이터를 입력받는다. 제4 입력 라인 그룹은 제3 입력 라인 그룹으로부터 하나 이상의 입력 라인 단위(도시된 예에서는 하나의 입력 라인 단위)로 쉬프트되도록 설정될 수 있다. 1, 2, 3, 4, 5, 9, 13번 입력 라인들에 입력되는 입력 데이터는 네번째 컬럼(제4 컬럼)의 컨볼루션 특징 맵의 출력에 활용되지 않는다.

[0051] 멤트랜지스터 시냅스 어레이는 게이트 전극 라인(510)에 의해 인가되는 게이트 전압에 의해 동시에 다중 셀에 대한 웨이트 업데이트를 수행할 수 있다. 다중 셀은 컨볼루션 연산을 위해 커널 필터의 동일한 셀에 의해 공통적으로 연산 처리되는 셀들일 수 있다. 1, 2, 5, 6번 입력 데이터는 멀티-컬럼에서 각각 커널 필터의 가장 좌측 상단의 셀과 공통으로 연산되므로, 제1 컬럼의 1번 입력 라인에 해당하는 셀, 제2 컬럼의 2번 입력 라인에 해당하는 셀, 제3 컬럼의 5번 입력 라인에 해당하는 셀, 제4 컬럼의 6번 입력 라인에 해당하는 셀이 게이트 전극 라인(510)으로 연결되어 게이트 전극 라인(510)에 인가되는 게이트 전압에 의해 동시적으로 웨이트 업데이트가 수행된다.

[0052] 예를 들어, 1, 2, 5, 6번 입력 데이터는 멀티-컬럼에서 각각 커널 필터의 가장 좌측 상단의 셀과 공통으로 연산되므로, 제1 컬럼의 1번 입력 라인에 해당하는 셀, 제2 컬럼의 2번 입력 라인에 해당하는 셀, 제3 컬럼의 5번 입력 라인에 해당하는 셀, 제4 컬럼의 6번 입력 라인에 해당하는 셀이 게이트 전극 라인(510)으로 연결되어 게

이트 전극 라인(510)에 인가되는 게이트 전압에 의해 동시적으로 웨이트 업데이트가 수행된다.

[0053] 도 3은 본 발명의 일 실시예에 따른 3단자 멤트랜지스터 시냅스 어레이에 의한 원-스텝 컨볼루션 및 풀링 과정을 나타낸 개념도이다. 3단자 멤트랜지스터 시냅스 어레이를 구성하는 각각의 단위 시냅스 셀들은 3단자 게이트-소스-드레인 전극과, 소스-드레인 전극 라인 간의 교차 영역에 마련되는 시냅스 반도체층(예를 들어, IGZO, ZnO, 펜타센 등) 및 분극 조절층(예를 들어, PVDF-TrFE, HZO, PVDF, PZT(납 지르코네이트 티타네이트), 페로브스카이트 구조를 가지는 산화물 등의 강유전체 물질 등)으로 구성될 수 있다.

[0054] 도 3의 실시예와 같이, 첫 번째부터 네 번째 컬럼의 출력 값(출력 전류값)이 각각 0, 4, 3, 3 인 경우, 3단자 멤트랜지스터 시냅스 어레이로부터 동시에 출력 값들이 추출되는 즉시, 풀링 단위의 연산(예를 들어, 최대 풀링, 평균 풀링 등)이 수행된다. 이에 따라 중간 특징 맵의 저장이 필요없게 되어 빠르고 에너지 효율적인 컨볼루션 변환이 가능해진다. 즉, 이미지 연산 영역 내에서 풀링 과정이 컨볼루션 변환 단계에 포함되며, 이는 전체 이미지 슬라이딩 양을 크게 줄일 수 있다.

[0055] 도시된 실시예의 경우, 4x4 이미지 영역에서 입력 전압 벡터는 고정한 채로 멀티 컬럼 형태의 커널 필터 연산이 수행된다. 이러한 원-스텝 컨볼루션 변환 연산을 효율적으로 가능하게 하려면 선제적으로 멀티 컬럼의 같은 웨이트 상태를 안정적으로 구현하는 것이 중요하다. 도 2a의 3단자 멤트랜지스터 시냅스 어레이에서 같은 색깔로 표시된 9종류의 커널 시냅스 웨이트는 모든 학습 과정에서 같은 웨이트 값을 보존해야 하며, 이러한 과정은 읽기 과정의 소스-드레인 전극 라인(워드-비트 라인)과 독립적인 게이트 전극 라인의 도입으로 구현될 수 있다.

[0056] 같은 시냅스 웨이트 업데이트가 필요한 다중 셀은 도 3과 같이 소스-드레인 전극 라인(워드-비트 라인)과 독립적인 게이트 전극 라인(예를 들어, 게이트 전극 라인)으로 동시에 접근될 수 있다. 이에 따라 워드-비트 라인과 독립적인 게이트 전극 라인을 통해 다중 셀의 웨이트가 동시적으로 업데이트될 수 있으며, 동일 입력 이미지에 해당하는 전압 값에 대해 이미지 슬라이딩과 같은 효과를 내는 특징 맵을 동시적으로 추출할 수 있다. 특히 특징 맵 데이터를 이용하여 별도의 정보를 저장하는 과정 없이, 풀링 과정을 통해 한번에 컨볼루션 연산 결과가 출력될 수 있다. 풀링 처리부(20)는 멤트랜지스터 시냅스 어레이의 멀티-컬럼의 출력 라인들을 통해 컬럼 단위로 출력되는 컨볼루션 특징 맵의 출력 값들을 풀링 처리하여 컨볼루션 특징을 추출할 수 있다.

[0057] 꺾은 선을 따라 같은 웨이트를 쓰기 전압으로 확보하는데 있어서 3전극 게이트 전극 라인은 쓰기 과정에서의 인접 셀 영향을 줄일 수 있으며, 에너지 효율적인 다중 셀 동시적 업데이트를 가능하게 한다. 이런 식으로 동시 추출된 4개의 출력 전류값 (I_1, I_2, I_3, I_4)은 곧바로 풀링 과정으로 변환되며 중간 특징 맵을 거치지 않은 원-스텝 컨볼루션 변환을 구현할 수 있다.

[0058] 3전극 멤트랜지스터 시냅스 어레이를 구성하는 각 시냅스 소자는 게이트 전극 라인에 게이트 전압이 인가되면, 게이트 전압에 따른 전기장이 분극 조절층(예를 들어, PVDF-TrFE)의 상태가 결정되고, 분극 효과가 드레인 전극(예를 들어, 그래핀)을 투과하여 반도체층(예를 들어, IGZO)의 아날로그 전도성 상태를 결정하는 원리에 의해 동작한다.

[0059] 본 발명의 실시예에 의하면, 기존의 2단자 크로스바 어레이가 아닌 3단자 멤트랜지스터 시냅스 어레이를 기반으로 하기 때문에, 게이트 전극 라인을 통한 동시적 다중 셀 업데이트를 이용하여 에너지 효율적인 컨볼루션 변환 연산을 구현할 수 있다. 이러한 어레이 구조는 읽기/쓰기 과정에서 다른 전극을 통해 접근함으로써 읽기/쓰기 오류를 줄이는 데에도 기여할 수 있다.

[0060] 본 발명의 실시예에 의하면, 커널 필터에 의해 웨이트를 업데이트할 때 도 2c와 같이 게이트 전극 라인을 공유하는 다중 셀들에 동시적으로 웨이트 업데이트를 적용함으로써, 반복적 학습 펄스 인가의 생략 및 간소화가 가능하다. 이로 인해 기존의 2단자 시냅스 어레이의 단점을 보완할 수 있도록 3단자 멤트랜지스터 시냅스 어레이의 동시 학습이 가능해진다.

[0061] 도 4는 본 발명의 실시예에 따른 3단자 멤트랜지스터 시냅스 어레이를 이용한 원-스텝 연산 과정과, 종래의 2단자 어레이를 통한 투-스텝 컨볼루션 및 풀링 과정을 비교하여 나타낸 개념도이다. 도 4의 (A)는 기존의 2단자 멤리스터 어레이를 통한 컨볼루션 변환 과정을 나타낸 것으로, 전체 이미지 공간에 대해 커널 슬라이딩으로 특징 맵을 추출하고, 그것을 다시 풀링 맵으로 변환한 후에 출력 뉴런으로 연결된다.

[0062] 도 4의 (B)는 본 발명의 실시예에 따라 수행되는 3단자 멤트랜지스터 시냅스 어레이를 통한 컨볼루션 변환 과정을 나타낸 것이다. 본 발명의 실시예에 의하면, 확장된 이미지 공간에 대해 원-스텝 컨볼루션 및 풀링 연산이 가능하므로 특징 맵 변환 없이 곧바로 풀링 맵으로 이미지 변환되어 빠르고 에너지 효율적인 컨볼루션 변환이

가능하다.

[0063] 도 5는 본 발명의 일 실시예에 따른 3단자 멤트랜지스터 어레이 소자 구조를 나타낸 도면이다. 도 6은 본 발명의 일 실시예에 따른 3단자 멤트랜지스터 어레이 소자 구조에 포함된 멤트랜지스터 소자를 확대하여 나타낸 도면이다. 도 5 및 도 6을 참조하면, 3단자 멤트랜지스터 시스템(10)은 복수개의 소스 전극 라인(110), 복수개의 드레인 전극 라인(210), 복수개의 시냅스 소자(300, 400), 및 게이트 전극 라인(510)을 포함할 수 있다. 복수개의 소스 전극 라인(110), 복수개의 드레인 전극 라인(210), 복수개의 시냅스 소자(300, 400), 및 게이트 전극 라인(510)은 기판(20) 상에 형성될 수 있다.

[0064] 각 소스 전극 라인(110)은 소정의 제1 방향으로 형성될 수 있다. 복수개의 소스 전극 라인(110)은 제1 방향과 상이한 제2 방향을 따라 서로 나란하게 배열될 수 있다. 제2 방향은 제1 방향에 대해 수직한 방향일 수 있다. 소스 전극 라인(110)은 소스 전극(120)의 일측으로부터 제1 방향으로 연장되도록 구성될 수 있다. 소스 전극 라인(110) 및 소스 전극(120)은 미리 설정된 전압이 인가될 수 있다.

[0065] 일 예로, 소스 전극(120) 및 소스 전극 라인(110)에는 기준 전압(예를 들어, 접지 전압)이 인가되거나, 입력 데이터에 해당하는 입력값이 인가될 수 있다. 각 드레인 전극 라인(210)은 드레인 전극(220)으로부터 제2 방향으로 연장되게 형성될 수 있다. 복수개의 드레인 전극 라인(210)은 복수개의 소스 전극 라인(110)과 교차하도록 형성될 수 있다. 복수개의 드레인 전극 라인(210)은 제1 방향을 따라 서로 나란하게 배열될 수 있다.

[0066] 드레인 전극 라인(210)은 시냅스 소자(300, 400)의 반도체층(300) 상에 형성될 수 있다. 드레인 전극 라인(210)은 소스 전극 라인(110)과 교차되는 일부 영역(교차 영역)이 소스 전극 라인(110)의 일부 영역 상에 적층되도록 구성될 수 있다. 예를 들어, 제2 방향이 제1 방향과 수직할 경우, 드레인 전극 라인(210)의 일부 영역은 소스 전극 라인(110)의 일부 영역과 서로 수직을 이루며, 소스 전극 라인(110) 상에 적층될 수 있다. 따라서, 도 5에 도시된 바와 같이 격자 형상(크로스바 어레이 구조)을 이룰 수 있다.

[0067] 드레인 전극 라인(210) 및 드레인 전극(220)은 미리 설정된 전압이 인가될 수 있다. 일 예로, 드레인 전극(220) 및 드레인 전극 라인(210)에는 기준 전압(예를 들어, 접지 전압)이 인가되거나, 입력 데이터에 해당하는 입력값이 인가될 수 있다. 소스 전극 라인(110)과 드레인 전극 라인(210)은 금속 등의 전도성 물질로 형성될 수 있다. 실시예에서, 드레인 전극 라인(210)은 일함수 변조가 가능한 그래핀층을 포함할 수 있다.

[0068] 시냅스 소자(300, 400)는 복수개의 소스 전극 라인(110)과 복수개의 드레인 전극 라인(210) 사이의 교차 영역들에 각각 제공될 수 있다. 도 6은 본 발명의 실시예에 따른 시냅스 소자의 일 예를 나타낸 것으로, 배리스터(barristor) 타입의 수직 시냅스 소자를 나타낸 것이다. 시냅스 소자(300, 400)는 소스 전극 라인(110)과 드레인 전극 라인(210) 사이의 교차 영역에 제공되는 반도체층(300)과, 드레인 전극 라인(210)의 교차 영역의 페르미 준위(Fermi level)를 조절하기 위한 조절층(400)을 포함할 수 있다.

[0069] 반도체층(300)은 소스 전극 라인(110)과 드레인 전극 라인(210) 사이에 제공될 수 있다. 반도체층(300)은 소스 전극 라인(110)과 드레인 전극 라인(210)이 서로 겹쳐지는 영역에 있어서, 소스 전극 라인(110)과, 소스 전극 라인(110) 상에 적층된 드레인 전극 라인(210)의 사이에 제공될 수 있다. 실시예에서, 반도체층(300)은 그래핀과 쇼트키 배리어를 형성할 수 있는 IGZO(Indium Gallium Zinc Oxide)층으로 이루어지거나, ZnO(zinc oxide), 펜타센 등의 반도체 물질로 이루어질 수 있지만, 반도체층(300)을 구성하는 물질이 이에 한정되지는 않는다.

[0070] 조절층(400)은 반도체층(300)의 상면을 덮도록 구성될 수 있다. 보다 상세히 말하면, 조절층(400)은 반도체층(300)과, 반도체층(300) 상에 제공되는 드레인 전극 라인(210)을 함께 덮도록 구성될 수 있다. 조절층(400)은 게이트 전극 라인(510)을 통해 인가되는 게이트 전압에 의해 드레인 전극 라인(210)의 그래핀층의 페르미 준위를 조절하여 반도체층(300)과 그래핀층 간의 쇼트키 장벽(Shottky barrier)을 조절할 수 있는 물질로 형성될 수 있다. 이러한 물질의 예에는 PVDF-TrFE(poly-vinylidene fluoride-trifluoro-ethylene)층과 같은 강유전체층이나, 고체 전해질층이 있다.

[0071] 배리스터 타입의 3전극 수직 시냅스 소자는 강유전체의 잔류 분극 조절을 통해 그래핀층의 그래핀과 반도체층(300)의 반도체 물질 사이의 쇼트키 배리어를 조절하여 아날로그 컨덕턴스 변화 특성을 나타낸다. 조절층(400)은 그래핀층과 반도체층(300) 간의 쇼트키 배리어의 변조가 비휘발성 특성을 가지도록 하여 게이트에 인가된 전계를 유지할 수 있도록 하며 이로 인해 저전력 시냅스 소자로 동작할 수 있도록 한다.

[0072] 조절층(400)은 형성된 쇼트키 배리어의 모듈레이션이 비휘발성을 갖기 위해 게이트에 인가된 전계를 유지한다. 조절층(400)은 그래핀의 페르미 레벨을 조절할 수 있는 유/무기(PVDF-TrFE, HZO, PZO, PVDF 등) 강유전체로 이루어지는 게이트 절연층(gate dielectric layer)으로 이루어질 수 있다. 특히 HZO는 수 nm의 얇은 박막 특성으

로 저전력, 고속 동작이 가능하며, CMOS 호환성이 높은 소재/구조로, 고집적화 및 회로 연결이 용이한 이점이 있다.

[0073] 게이트 전극 라인(510)은 소스 전극 라인(110)과 드레인 전극 라인(210)의 교차 영역에서 소스 전극 라인(110) 및 드레인 전극 라인(210)과 교차하도록 형성될 수 있다. 게이트 전극 라인(510)은 소스 전극 라인(110) 및 드레인 전극 라인(210)과의 교차 영역에서 조절층(400) 상에 적층될 수 있다. 게이트 전극 라인(510)은 조절층(400)의 상면과 접촉하며, 제3 방향으로 형성될 수 있다.

[0074] 제3 방향은 제1 방향과 제2 방향의 사이 방향일 수 있다. 게이트 전극 라인(510)의 길이 방향은 소스 전극 라인(110)의 길이 방향과 드레인 전극 라인(210)의 길이 방향의 사이 방향으로 형성될 수 있다. 따라서, 도 1에 도시된 바와 같이, 3전극 멤트랜지스터 시스템(10)의 형상을 전체적으로 살펴보면, 게이트 전극 라인(510)이 대각선 방향으로 형성됨을 알 수 있다.

[0075] 3단자 멤트랜지스터 시스템(10)은 소스 전극(120), 드레인 전극(220) 및 게이트 전극(520)을 더 포함할 수 있다. 게이트 전극(520)은 게이트 전극 라인(510)을 따라 대각 방향으로 배열된 다수의 시냅스 소자(300, 400)의 웨이트(weight), 즉 가중치를 동시에 학습하기 위한 가중치 제어 신호를 게이트 전극 라인(510)에 인가할 수 있다. 소스 전극(120)은 소스 전극 라인(110)의 끝단에 제공될 수 있다. 드레인 전극(220)은 드레인 전극 라인(210)의 끝단에 제공될 수 있다. 게이트 전극(520)은 게이트 전극 라인(510)의 끝단에 제공될 수 있다.

[0076] 본 발명의 실시예에 따른 3단자 멤트랜지스터 시스템(10)은 제어부(도시 생략)를 더 포함할 수 있다. 제어부는 각 드레인 전극 라인(210) 또는 각 소스 전극 라인(110)에 인가될 입력 신호 및 각 게이트 전극 라인(510)에 인가될 가중치 제어 신호를 결정하도록 구성될 수 있다. 드레인 전극 라인(210)과 소스 전극 라인(110) 중 어느 하나는 입력으로, 다른 하나는 출력으로 활용될 수 있다.

[0077] 실시예에서, 제어부는 입력 데이터를 컨볼루션 처리하기 위한 컨볼루션 필터의 필터값들을 기초로 가중치 제어 신호를 결정하여 게이트 전극(520) 및 게이트 전극 라인(510)에 인가하고, 입력 데이터의 입력값들을 기초로 입력 신호를 생성하여 드레인 전극(220) 또는 소스 전극(120)에 인가할 수 있다. 3단자 멤트랜지스터 시스템(10)은 복수의 소스 전극 라인(110) 또는 복수의 드레인 전극 라인(210)으로 출력되는 출력 신호들을 기초로 컨볼루션 연산 결과를 산출할 수 있다. 3단자 멤트랜지스터 시스템(10)은 입력 데이터의 크기 및 컨볼루션 필터의 크기를 기초로, 복수의 소스 전극 라인(120)의 출력 신호들 중, 컨볼루션 필터의 슬라이딩에 따른 컨볼루션 연산값들과 관련된 특징 신호들을 추출하고, 추출된 특징 신호들을 풀링(pooling)하여 컨볼루션 특징 맵을 생성할 수 있다.

[0078] 도 7은 본 발명의 일 실시예에 따른 3단자 멤트랜지스터 소자 시스템 제조 방법에 의해 3단자 멤트랜지스터 시스템(10)이 제조되는 과정을 순차적으로 나타낸 도면이다. 도 7을 참조하면, 3단자 멤트랜지스터 소자 시스템 제조 방법(S10)은 S100 내지 S500 단계를 포함한다. S100 단계에서, 기판 상에 복수개의 소스 전극(120), 제1 방향을 길이 방향으로 갖는 복수개의 소스 전극 라인(110) 및 복수개의 드레인 전극(220)이 형성될 수 있다.

[0079] 일 예로, S100 단계는 SiO_2/Si 기판을 에탄올에 15분, 아세톤에 15분 순으로 초음파 세척한 후, 흐르는 순수(DI water)로 세척한 다음 질소 에어건 블로잉을 수행하는 단계(S110), AZ5214 Photoresist를 스핀 코팅한 후 1분간 120 ° C에서 베이킹(baking)하는 단계(S120), 포토마스크(Photo mask)를 활용하여 소스 전극(120), 소스 전극 라인(110) 및 드레인 전극(220)에 해당하는 패턴 부분을 노광한 후, MIF300 현상액에 15초 동안 현상하는 단계(S130) 및 열증착기(Thermally evaporator)를 이용하여 알루미늄(Al)을 3nm 두께로 증착한 후 금(Au)을 22 nm 두께로 증착한 후 아세톤을 이용하여 리프트 오프(lift off) 공정을 진행하여 PR(감광액)을 제거하는 단계(S140)를 포함할 수 있다.

[0080] S200단계에서, 소스 전극 라인(110) 상의 일부 영역(드레인 전극 라인 및 게이트 전극 라인과의 교차 영역)에 반도체층(300)이 형성될 수 있다. 일 예로, S200 단계는 반도체층(300)을 위한 포토리소그래피(photolithography)를 진행하는 단계(S210) 및 80 um^2 사이즈의 반도체층(300) 형성을 위해 RF-스퍼터링 시스템(sputtering system)을 이용하여 IGZO를 증착 조건 50 W, Ar 19sccm, O_2 2 sccm, 20 mTorr 하에서 두께 15 nm로 증착한 후 아세톤으로 리프트 오프를 진행하는 단계(S220)를 포함할 수 있다.

[0081] S300 단계에서, 반도체층(300) 상의 일부 영역과 드레인 전극(220) 상의 일부 영역을 연결하는 복수개의 드레인 전극 라인(210)이 제2 방향으로 형성될 수 있다. 일 예로, S300 단계는 소스 전극 라인(110)과 드레인 전극 라인(210) 간의 교차 영역에 해당하는 반도체층(300) 상에 그래핀(Graphene)층을 대면적 전사하고, 아세톤에 20분

담근 후 DI water로 세척함으로써 그래핀 보호층인 PMMA를 제거하는 단계(S310), 포토리소그래피 를 통해 그래핀층 위의 넓이 40 um^2의 드레인 전극 라인(210) 부분을 제외한 감광액을 제거하여 현상하는 단계(S320) 및 RIE(reactive-ion etching) 공정을 통해 드레인 전극 라인(210)을 제외한 그래핀층을 RIE 조건 40 W, Ar 10 sccm, O_2 10 sccm, 0.2 Torr 하에서 에칭 후 아세톤을 이용하여 감광액을 리프트 오프 처리하는 단계(S330)를 포함할 수 있다.

[0082] S400 단계에서, 반도체층(300)과, 반도체층(300) 상에 제공되는 드레인 전극 라인(210)의 그래핀층을 함께 덮도록 구성되는 조절층(400)이 형성될 수 있다. 일 예로, S400 단계는 유기 강유전체 고분자인 폴리비닐리덴-트리플루오로에틸렌(PVDF-TrFE)(70:30)를 DMF 용매에 8 wt%로 녹여준 용액을 스핀-코팅(spin-coating)한 후, 135 ° C에서 2시간 동안 어닐링(annealing)하여 결정화 시키는 단계를 포함할 수 있다.

[0083] S500 단계에서, 조절층(400) 상면의 일부 영역에 접촉되도록 게이트 전극 라인(510)이 제1 방향과 제2 방향의 사이 방향인 제3 방향(대각 방향)으로 형성될 수 있다. 일 예로, S500 단계는 반도체층(300)을 포함하는 각 시냅스 소자의 채널 면적을 덮을 수 있는 게이트 전극 라인(510)을 포토리소그래피를 통해 현상 후 열증착에 의해 Al을 25 nm 두께로 증착하는 단계(S510) 및 유기소재인 PVDF-TrFE위 게이트 패턴의 리프트 오프를 위해 감광액만 선택적으로 제거할 수 있는 PGMEA(Propylene glycol monomethyl ether acetate) 용매에 10분 동안 노출하여 리프트 오프를 진행하는 단계(S520)를 포함할 수 있다.

[0084] 도 8은 4x4 입력 데이터와 3x3 커널(컨볼루션 필터)의 조건 하에서, 기존의 CNN 어레이의 구동 방식을 개략적으로 나타낸 도면이고, 도 9는 도 8의 과정을 보다 상세히 나타낸 도면으로, 컨볼루션 필터의 슬라이딩에 따른 입력 데이터에 대한 컨볼루션 연산 과정을 순차적으로 나타낸 도면이다. 도 8 및 도 9를 참조하면, 기존의 CNN 어레이 구동 방식의 경우, 커널을 이용한 각각의 이미지 슬라이딩을 4회 실시하여, 이를 통해 4개의 인접한 특징 맵(feature map)을 추출한 후, 특징 맵에 최대 풀링(max pooling)을 적용하여 출력한다.

[0085] 도 10은 도 8과 같은 조건 하에서, 본 발명의 실시예에 따른 3단자 멤트랜지스터 시스템(10)에 의한 컨볼루션 연산 과정을 개략적으로 나타낸 도면이고, 도 11은 도 10의 과정을 보다 상세히 나타낸 도면으로, 대각선 어레이(게이트 전극 라인이 대각선 방향으로 형성되는 본 발명의 구조)를 갖는 3단자 멤트랜지스터 시스템(10)에 의하여 특징 맵(feature map) 데이터를 이용하여 별도의 정보를 저장하는 과정 없이, 풀링(pooling) 과정을 통해 한번에 컨볼루션 연산 결과가 출력되는 모습을 나타낸 도면이다. 도 12는 도 10 및 도 11에 도시된 방식으로 소스 전극(120)을 통해 인풋 데이터를 입력 받고, 드레인 전극(220)을 통해 데이터가 출력되는 모습을 나타낸 예시도이다.

[0086] 도 10 내지 도 11을 참조하면, 본 발명의 일 실시예에 따른 3단자 멤트랜지스터 시스템(10)은 한번의 입력 데이터의 입력을 통해 특징 맵을 도출할 수 있으며, 도출된 특징 맵을 별도로 저장하는 과정 없이, 곧바로 풀링 과정을 통해 컨볼루션 연산 결과 데이터를 출력할 수 있다. 예를 들어, 도 10 및 도 11에 도시된 바와 같이 입력 데이터가 1 부터 16까지의 총 16개의 화소값들에 해당하는 입력값들을 포함하고 있는 경우, 각각의 입력값이 각각의 소스 전극(120)에 한번에 컬럼 형태로 입력될 수 있다. 이를 위해, 소스 전극의 개수는 입력 데이터를 구성하는 입력값들의 개수(예를 들어, 컨볼루션 처리 단위인 영상 내의 화소들의 개수)와 같은 16개로 구성되거나, 그보다 많아야 한다.

[0087] 일 예로, 입력 데이터가 NXN 크기를 가지는 경우, 소스 전극 라인(110) 및 드레인 전극 라인(120)의 개수는 각각 N^2 개로 구현될 수 있다. 특징 맵 도출과 풀링 과정을 거쳐 생성된 출력 데이터는 드레인 전극(220)을 통해 출력되는데, 도시된 예의 경우, 첫번째 드레인 전극(220)을 통해 A 아웃풋이 출력되고, 두번째 드레인 전극(220)을 통해 B 아웃풋이 출력되고, 5번째 드레인 전극(220)을 통해 C 아웃풋이 출력되고, 6번째 드레인 전극(220)을 통해 D 아웃풋이 출력될 수 있다.

[0088] 이때 아웃풋 A 내지 D는 3단자 멤트랜지스터 시스템(10)에 의해 다수의 드레인 전극(220)을 통해 동시에 출력될 수 있다. 즉, 게이트 전극 라인을 공유하는 16x4 어레이 소자 구조에서 옴의 법칙(Ohm's law)과 키르히호프(Kirchhoff's) 법칙을 통한 벡터 연산을 통해 A, B, C, D 연산 값이 한번에 추출된다. 이에 따라 본 발명의 실시예에 의하면 빠르고 효율적인 컨볼루션 연산이 가능하다.

[0089] 도 13은 기존의 시냅스 어레이 중 어느 하나의 멤트랜지스터 소자에 대하여 가중치 학습을 위한 펄스 시퀀스를 입력하는 모습을 나타낸 도면이고, 도 14는 본 발명의 일 실시예에 따른 3단자 멤트랜지스터 시스템(10)에 있어서 한 번의 펄스 시퀀스 입력으로 복수의 3단자 멤트랜지스터 소자(11)에 가중치를 학습시키는 모습을 나타낸

도면이다.

[0090] 도 13을 참조하면, 붉은색으로 도시된 하나의 멤트랜지스터 소자에 대하여 가중치를 학습시키기 위해서는 소스 전극 라인과 드레인 전극 라인 각각에 펄스 시퀀스를 입력해야 한다. 이와 달리, 도 14를 참조하면, 본 발명의 일 실시예에 따른 3단자 멤트랜지스터 시스템(10)는, 게이트 전극 라인(510)을 통한 한 번의 펄스 시퀀스 인가를 통해 복수의 3단자 멤트랜지스터 소자(11)(대각 방향으로 배열된 복수의 시냅스 소자의 그룹)에 대하여 동시에 학습에 사용되는 가중치 제어 신호를 전달할 수 있다. 3단자 멤트랜지스터 시스템(10)의 경우, 각각의 색깔별로 도시된 복수의 3단자 멤트랜지스터 소자(11)에 대하여 가중치를 학습시키고자 할 경우, 3단자 멤트랜지스터 소자(11)를 지나는 하나의 게이트 전극 라인(510)에 펄스 시퀀스를 입력하는 것 만으로 목적한 바를 달성할 수 있다.

[0091] 다시 도 6을 참조하면, 3단자 멤트랜지스터 소자(11)는 3단자 구조를 갖는데, 소스 전극 라인(110)과 드레인 전극 라인(210)이 채널층인 반도체층(예를 들어, IGZO층)을 사이에 두고 수직으로 교차할 수 있으며, 그 위로 강유전체층(예를 들어, PVDF-TrFE층)이 위치하도록 하여 게이트 전극 라인(510)을 대각 방향으로 제작할 수 있다.

[0092] 상술한 바와 같은 본 발명의 실시예에 따른 3단자 멤트랜지스터 시스템(10)은 컨볼루션 네트워크 연산을 수행하는 컨볼루션 네트워크 연산 장치 및 방법으로 응용될 수 있다. 컨볼루션 네트워크 연산 장치는 입력 데이터의 입력값들을 기초로 입력 신호들을 생성하여 복수의 드레인 전극 라인(또는 복수의 소스 전극 라인)에 인가하고, 컨볼루션 필터의 필터값들을 기초로 가중치 제어 신호들을 결정하여 복수의 게이트 전극 라인에 인가할 수 있다.

[0093] 컨볼루션 네트워크 연산 장치는 복수의 소스 전극 라인(또는 복수의 드레인 전극 라인)으로 출력되는 출력 신호들을 기초로 컨볼루션 연산 결과를 산출할 수 있다. 이때, 컨볼루션 네트워크 연산 장치는 입력 데이터의 크기 및 컨볼루션 필터의 크기를 기초로 출력 신호들 중 컨볼루션 필터의 슬라이딩에 따른 컨볼루션 연산값들과 관련된 특징 신호들을 추출하고, 특징 신호들을 풀링하여 컨볼루션 특징 맵을 생성할 수 있다.

[0094] 도 11에 도시된 실시예의 경우, 첫번째 내지 여섯번째 소스 전극 라인 중, 첫번째, 두번째, 다섯번째 및 여섯번째 소스 전극 라인의 출력 신호들이 특징 신호들로 추출되고, 해당 특징 신호들을 풀링하여 컨볼루션 특징 맵을 생성하게 된다. 입력 데이터의 행과 열 순으로 순차로 순번을 부여할 경우, 출력 신호들 중에서 추출되는 특징 신호들의 순번은 예를 들어, Nk+1, Nk+2, ..., Nk+N-M+1 (k는 0, 1, ..., N-M) (N: 입력 데이터의 크기, M: 컨볼루션 필터의 크기) 일 수 있다. 이 경우, N^2 개의 출력 신호들 중에서, $(N-M+1)^2$ 개의 특징 신호들이 추출될 수 있다. 도 7의 경우, N = 4, M = 3 이므로, 1, 2번째 순번(k = 0 인 경우)과, 5, 6번째 순번(k = 1 인 경우)에 해당하는 4개(2^2개)의 특징 신호들(A, B, C, D)이 소스 전극들을 통해 출력된다.

[0095] 도 15는 강유전체층에 분극이 일어난 모습을 나타낸 도면이고, 도 16은 3단자 멤트랜지스터 소자(11)의 게이트 전압에 따른 드레인-소스 전류 변화 특성을 나타낸 도면이고, 도 17은 3단자 멤트랜지스터 소자(11)의 펄스 게이트 전압 인가에 따른 PSC(postsynaptic currents) 변화 특성을 나타낸 도면이다. 도 16의 가로축은 3단자 멤트랜지스터 소자(11)의 게이트 전극에 인가되는 게이트 전압, 세로축은 드레인-소스 간에 흐르는 전류이다. 도 16의 실험에서 드레인-소스 전압은 1V로 일정하게 유지되었다.

[0096] 도 17의 가로축은 3단자 멤트랜지스터 소자(11)의 게이트 전극에 인가되는 펄스 전압, 세로축은 펄스 전압에 따른 후시냅스 전류(postsynaptic currents)이다. 도 17의 실험은 3단자 멤트랜지스터 소자(11)의 게이트 전극에 시냅스 소자의 가중치를 증가시키는 강화 펄스(potentiating pulse)에 해당하는 30V 펄스 전압을 500 ms 주기로 30초 동안 반복적으로 인가한 후, 다시 시냅스 소자의 가중치를 감소시키는 억제 펄스(depressing pulse)에 해당하는 -30V 펄스 전압을 500 ms 주기로 30초 동안 반복적으로 인가하는 방식으로 수행되었으며, 드레인-소스 전압은 1.5 V로 일정하게 유지하였다.

[0097] 도 6, 도 15 내지 도 17을 참조하면, 게이트 전극(520) 및 게이트 전극 라인(510)에 인가된 전기장은 강유전체층(Ferroelectric layer)에 분극을 일으키고, 이 분극의 방향에 따라 그래핀의 페르미 레벨을 조절할 수 있게 된다. 이는 곧 그래핀층과 반도체층 사이의 쇼트키 장벽(Schottky barrier)의 높이 조절과 연결되며, 그에 따라 3단자 멤트랜지스터 소자(11)의 컨덕턴스(conductance) 조절이 가능하게 된다.

[0098] 드레인 전극 라인(210)의 그래핀층과 조절층(400)의 강유전체를 사용하여 쇼트키 장벽을 조절하는 배리스터(barristor) 구조 디바이스로 3단자 멤트랜지스터 소자(11)를 수직 구조로 제조하는 경우, 수평형 인공 시냅스 소자에 비해 풋-프린트(foot-print)를 줄일 수 있다. 특히, 본 발명의 실시예에 의하면, 대각화 게이트 어레이

를 구성하는 웨이트 업데이트 영역과 행렬곱연산 구동 영역을 구분하여, 동시적 학습 용이성을 확보할 수 있다.

[0099] 상술한 바와 같은 본 발명의 실시예에 따른 3단자 멤트랜지스터 시스템에 의하면, 입력 데이터와 컨볼루션 필터 간의 데이터 슬라이딩 동작이 필요하지 않아 회로를 간소화할 수 있으며, 컨볼루션과 풀링을 동시적으로 연산 처리를 수행할 수 있어 연산량을 종래 기술 대비 1/4 미만으로 줄이고, 연산 속도를 비약적으로 높일 수 있다. 또한, 시냅스 소자의 가중치 동시 학습 효율을 높일 수 있으며, 풀링 처리를 위해 특징 맵을 저장할 필요가 없고 별도의 풀링층을 필요로 하지 않아 데이터 입/출력 및 연산량을 추가로 감소할 수 있다.

[0100] 상술한 바와 같이 본 발명은 다중 셀 업데이트를 통해 동시 학습 가능한 3단자 멤트랜지스터 어레이를 제안하며, 이를 통해 기존의 컨볼루션 뉴럴 네트워크 시스템에 비해 에너지 소모를 줄이고, 연산 속도를 향상시킬 수 있다. 특히 이러한 3단자 어레이 집적 및 동시 학습 기술은 객체 및 영상 처리의 효율성이 중요한 빅데이터 시스템의 컨볼루션 연산 구동 소자로 응용될 수 있으며, 그 외에도 실시간 얼굴 및 사물 인식, 자율주행을 비롯하여 다양한 컨볼루션 연산 및 학습 플랫폼으로 활용될 수 있다.

[0101] 도 18a는 본 발명의 실시예에 따른 배리스터 타입 수직 3단자 인공 시냅스 소자를 나타낸 도면이다. 도 18a의 실시예에서, 수직 3단자 인공 시냅스 소자는 무기 강유전체에 해당하는 HZO 층과, HZO 층 상에 증착되는 IGZO 반도체층, HZO 층 상에 증착되는 Al_2O_3 층, Al_2O_3 층 상에 전사되는 그래핀 층(드레인 전극 라인), HZO 층의 하부에 형성되는 TiN 게이트 전극 라인, IGZO 반도체층 상에 증착되는 Au 소스 전극 라인을 포함한다.

[0102] 도 18b는 도 18a의 배리스터 타입 수직 3단자 인공 시냅스 소자의 레이아웃을 나타낸 도면이다. 도 18c 및 도 18d는 도 18a의 배리스터 타입 수직 3단자 인공 시냅스 소자의 단면도 및 층별 성분 원소 분석 결과를 나타낸 도면이다. 도 18e는 도 18a의 배리스터 타입 수직 3단자 인공 시냅스 소자의 라만 스펙트럼을 나타낸 도면이다. 도 18f는 도 18a의 배리스터 타입 수직 3단자 인공 시냅스 소자의 전류 밀도 및 유전체 특성을 나타낸 도면이다. 도 19a 내지 도 19i, 도 20a 내지 도 20g, 도 21a 내지 도 21c는 본 발명의 실시예에 따른 수직 3단자 인공 시냅스 소자의 다양한 이미지 객체 인식 특성을 나타낸 도면이다. 도 18 내지 도 21은 무기 강유전체로 HZO를 사용한 실시예에 해당한다. 구체적으로, 도 21은 대각선 게이트 시냅스 소자를 통한 이미지를 인식하는 인공 신경망 구조 및 인식, 학습 성능을 나타내는 도면이다.

[0103] 도 22a는 본 발명의 또 다른 실시예에 따른 수직 3단자 시냅스 소자를 나타낸 도면이다. 도 22b 내지 도 22e는 도 22a의 수직 3단자 시냅스 소자의 제조 과정을 나타낸 도면이다. 도 22a의 수직 3단자 시냅스 소자는 TiN 게이트 전극 라인, TiN 게이트 전극 라인 상의 HZO 유전체층, HZO 유전체층 상에 전사되는 그래핀 층, 그래핀 층 상에 증착되는 ZnO 반도체층, ZnO 반도체층 상에 증착되는 Au/Al 전극 라인을 포함한다.

[0104] 도 23a는 본 발명의 또 다른 실시예에 따른 수직 3단자 시냅스 소자를 나타낸 도면이다. 도 23b 내지 도 23f는 도 23a의 수직 3단자 시냅스 소자의 제조 과정을 나타낸 도면이다. 도 23a의 수직 3단자 시냅스 소자는 Au/Cr 전극 라인, Au/Cr 전극 라인 상의 IGZO 반도체층, IGZO 반도체층 상에 전사되는 그래핀 층, 그래핀 층 상의 PVDF-TrFE 층, PVDF-TrFE 층 상의 Al 전극 라인을 포함한다. 도 24a 내지 도 24c, 도 25a 내지 도 25f, 도 26a 내지 도 26e, 도 27a 내지 도 27g는 본 발명의 실시예에 따른 수직 3단자 시냅스 소자의 인공지능 인식/학습 특성을 나타낸 도면이다. 도 22 내지 도 27은 유기 강유전체로 PVDF-TrFE를 사용한 실시예에 해당한다. 구체적으로, 도 27d 내지 도 27g는 대각선 게이트 시냅스 어레이 소자를 통한 인공지능 이미지 학습 성능을 나타내는 도면이다.

[0105] 도 18 내지 도 27에 나타낸 다양한 전기적 특성들을 통해 본 발명의 실시예에 따른 수직 3단자 인공 시냅스 소자는 강유전체 기반으로 소자 컨덕턴스 값을 정밀하게 조절 가능하다는 것을 알 수 있다. 3단자 시냅스 소자는 일함수 변조 가능한 그래핀(graphene)이 적용된 배리스터 타입의 인공 시냅스 소자 및/또는 강유전체의 분극 분역 조절에 의해 채널 컨덕턴스 조절이 가능한 FeFET 타입의 인공 시냅스 소자를 포함할 수 있다.

[0106] 기존의 3단자 시냅스 소자는 2단자 시냅스 소자에 비해 낮은 집적도($10F^2$)와 노드 간 모든 인터랙션(interaction)에서 절연막 증착이 추가되어 공정 난이도가 증가한다. 본 발명의 실시예에 따른 3단자 인공 시냅스 소자는 강유전체가 게이트 유전체로 적용됨과 동시에 비트 라인과 워드 라인 사이에 수직으로 채널이 형성되는 비휘발성 수직 3단자 구조의 시냅스 소자로 제공된다.

[0107] 본 발명의 실시예에 따른 배리스터 타입의 수직 3단자 인공 시냅스 소자는 강유전체(HZO, PZT, PVDF-TrFE, PVDF 등) 물질을 기반으로 제작될 수 있다. 수직 3단자 인공 시냅스 소자는 일함수 변조가 가능한 그래핀과 쇼트키 배리어 형성을 위해 반도체 물질(IGZO, ZnO, 펜타센 등)을 수직으로 적층하여 제조될 수 있다. 적용된 강유전체

의 잔류 분극 조절로 그래핀과 반도체 사이의 쇼트키 배리어를 조절하는 아날로그 컨덕턴스 변화 특성을 가지며, 형성된 쇼트키 배리어의 변조가 비휘발성으로 인해 게이트에 인가된 전계를 유지할 수 있는 저전력 시냅스 소자로 동작할 수 있다.

[0108] 본 발명의 실시예에 따른 수직 3단자 인공 시냅스 소자는 기존의 2단자 소자와 달리, VMM 연산 영역과 웨이트 업데이트 영역이 분리되어 독립적인 동작이 가능하다. 또한, 본 발명의 실시예에 따른 수직 3단자 인공 시냅스 소자는 비트 라인과 워드 라인 사이 수직으로 채널을 형성할 수 있도록 반도체 채널 패터닝을 진행하므로, 기존 3단자 소자에 비해 고집적화에 유리한 구조($4F^2$)를 가지며, 공정 용이성이 높다.

[0109] FeFET 타입의 수직 3단자 시냅스 소자의 경우, 미세 패터닝을 통해 수직형 채널을 형성하고, 강유전체를 유전층으로 사용해 제작될 수 있다. FeFET 타입의 수직 3단자 시냅스 소자는 잔류 분극을 통하여 채널 경계의 반송자 밀도를 조절함으로써 수직 방향의 채널 컨덕턴스를 조절 및 저장할 수 있다. 즉, 강유전체를 유전층으로 사용해 강유전체의 잔류 분극을 통하여 채널 경계의 반송자 밀도를 조절함으로써 수직 방향의 채널 컨덕턴스를 조절 및 저장할 수 있다.

[0110] 이상에서 실시예를 통해 본 발명을 설명하였으나, 위 실시예는 단지 본 발명의 사상을 설명하기 위한 것으로 이에 한정되지 않는다. 통상의 기술자는 전술한 실시예에 다양한 변형이 가해질 수 있음을 이해할 것이다. 본 발명의 범위는 첨부된 특허청구범위의 해석을 통해서만 정해진다.

부호의 설명

[0111] 10 3단자 멤트랜지스터 시스템

20 기판

110 소스 전극 라인

120 소스 전극

210 드레인 전극 라인

220 드레인 전극

300 반도체층

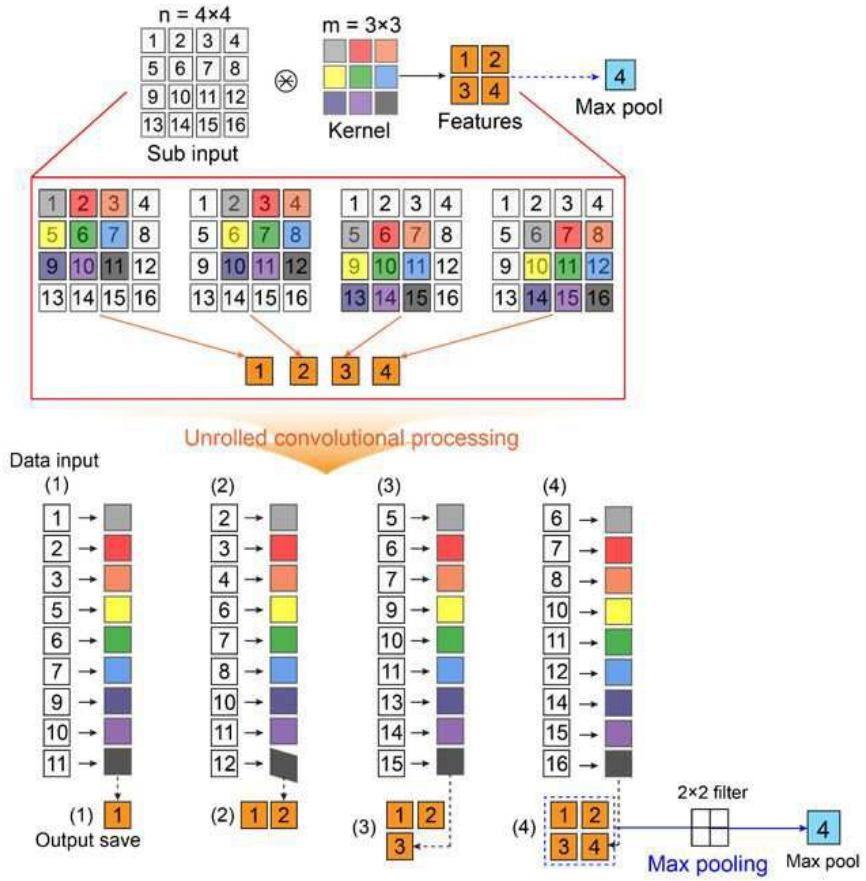
400 조절층

510 게이트 전극 라인

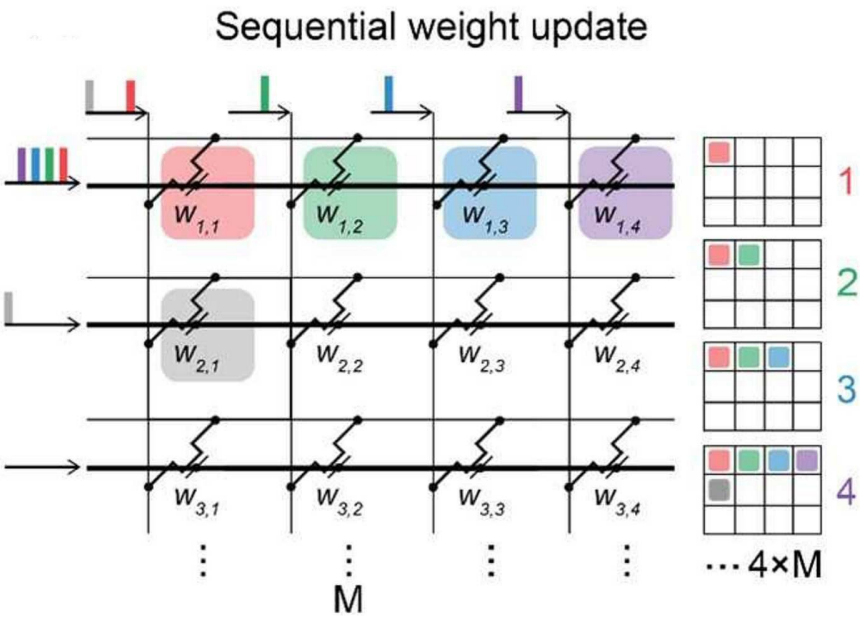
520 게이트 전극

도면

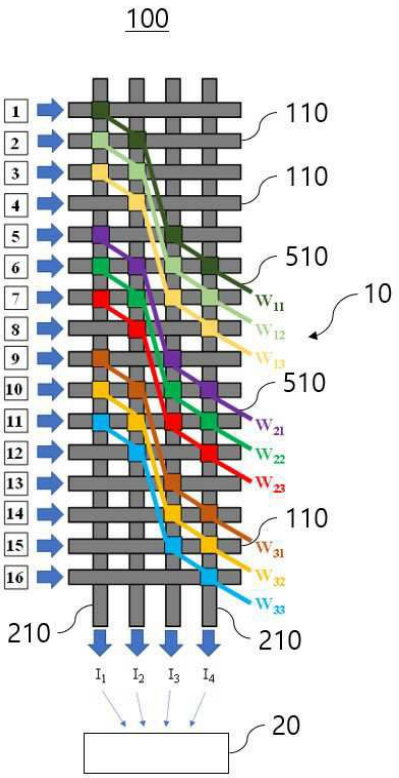
도면1a


n = 4×4
m = 3×3
Sub input
Kernel
Features
Max pool
Unrolled convolutional processing
Data input
(1)
(2)
(3)
(4)
Output save
2×2 filter
Max pooling
Max pool

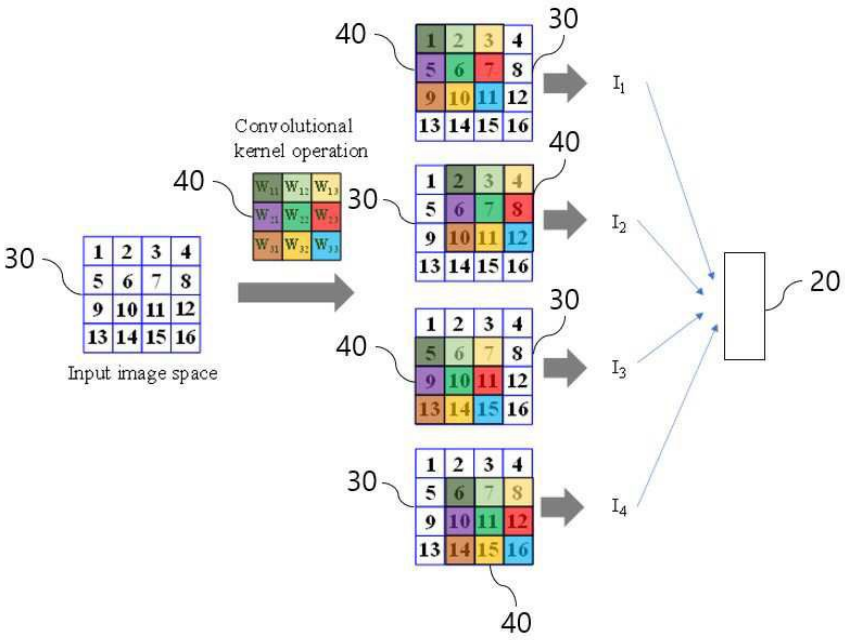
도면1b


Sequential weight update
M
4×M

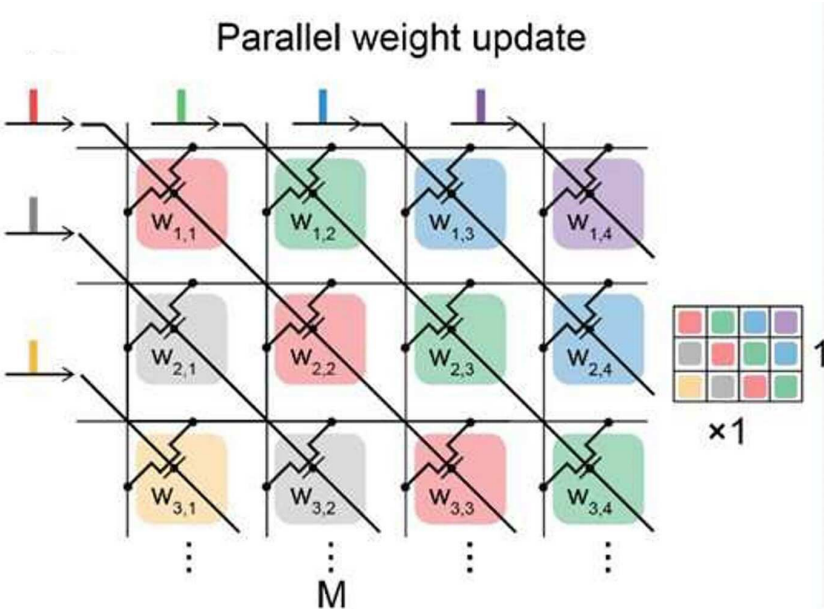
도면2a


100
110
110
510
10
510
110
210
210
20

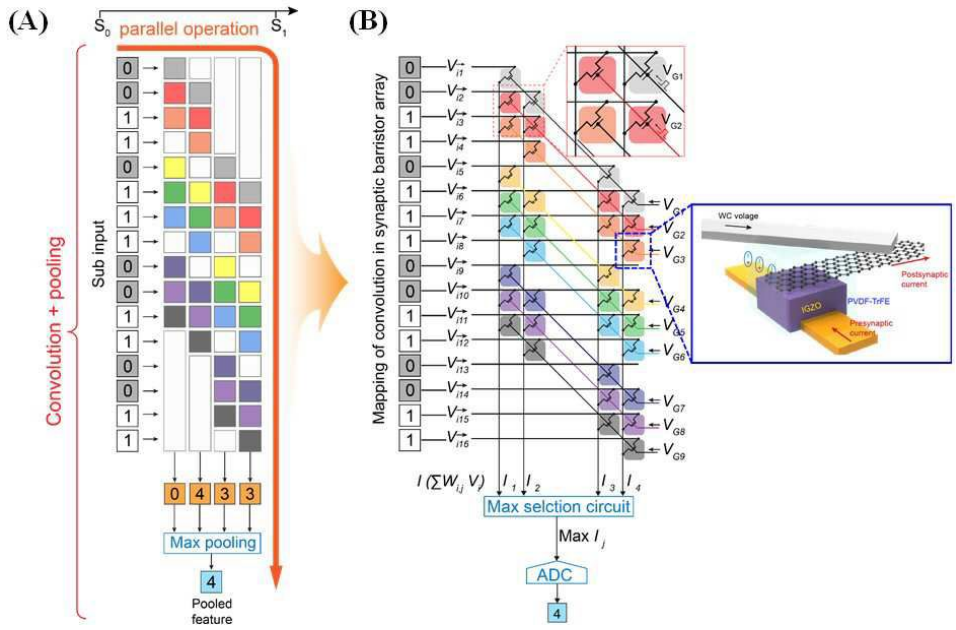
도면2b


Convolutional
kernel operation
Input image space
30
40
20

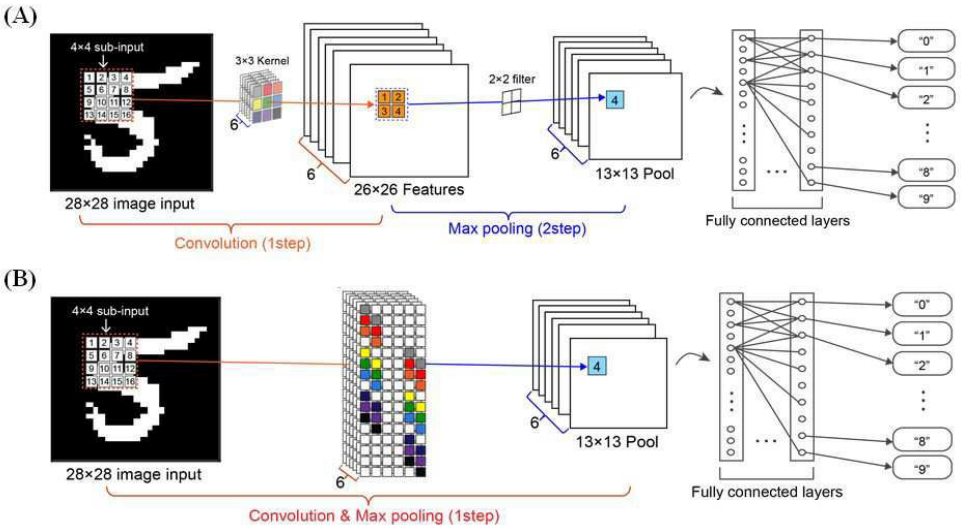
도면2c


Parallel weight update
W1,1
W1,2
W1,3
W1,4
W2,1
W2,2
W2,3
W2,4
W3,1
W3,2
W3,3
W3,4
M
×1

도면3

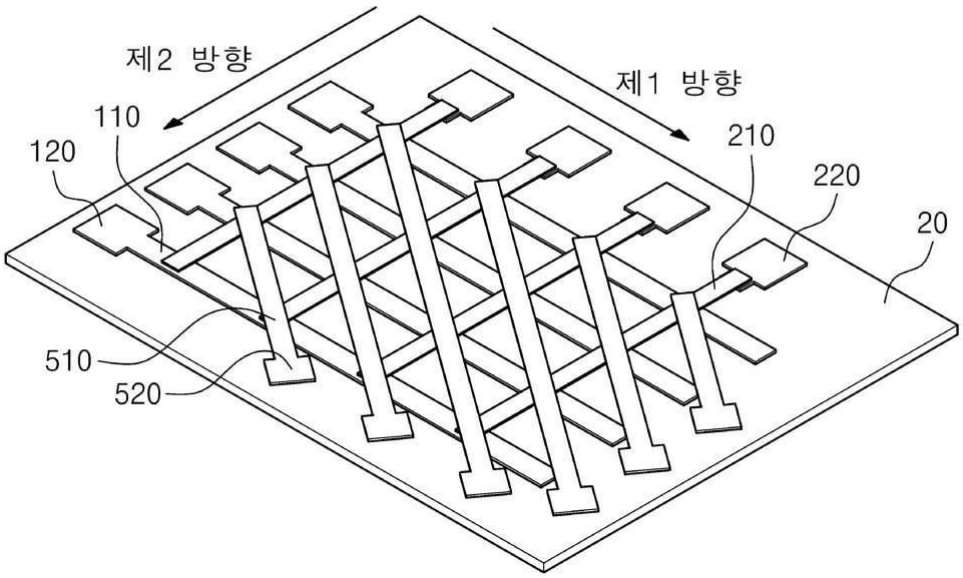

(A)
S0 parallel operation S1
Convolution + pooling
Sub input
0 4 3 3
Max pooling
4
Pooled feature
(B)
Mapping of convolution in synaptic barristor array
I (ΣWij Vj)
Max selction circuit
Max Ij
ADC
4
WC volage
IGZO
PVDF-TrFE
Postsynaptic current
Presynaptic current

도면4

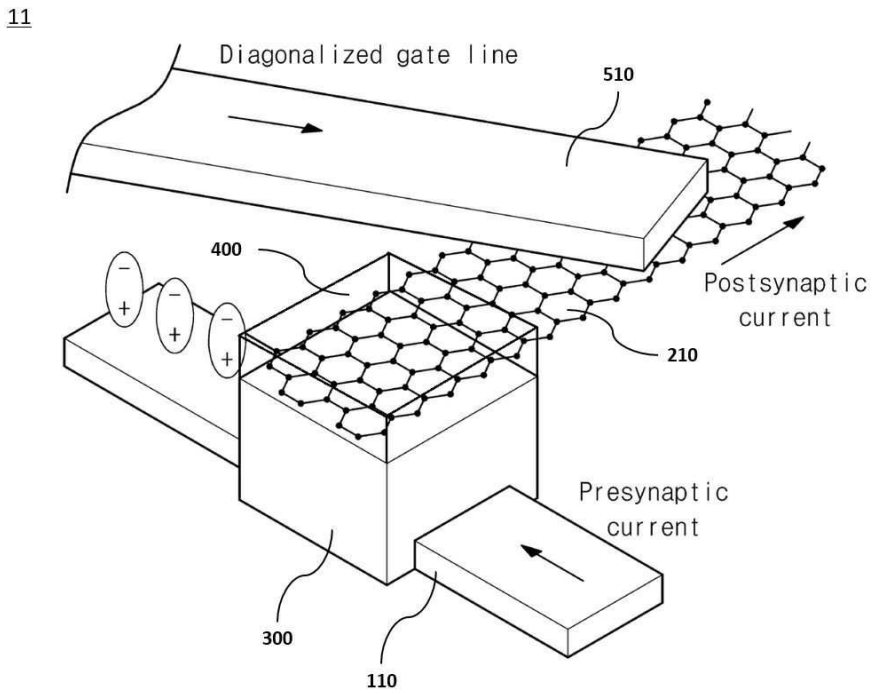

(A)
4×4 sub-input
3×3 Kernel
2×2 filter
28×28 image input
26×26 Features
13×13 Pool
Fully connected layers
Convolution (1step)
Max pooling (2step)
(B)
4×4 sub-input
28×28 image input
13×13 Pool
Fully connected layers
Convolution & Max pooling (1step)

도면5

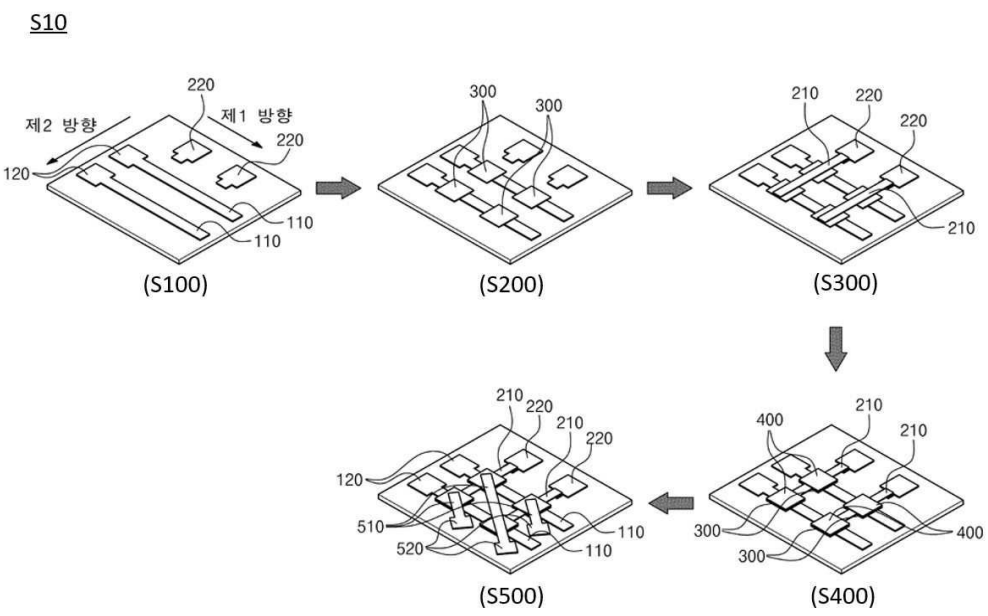
10


제2 방향
제1 방향
110
120
210
220
20
510
520

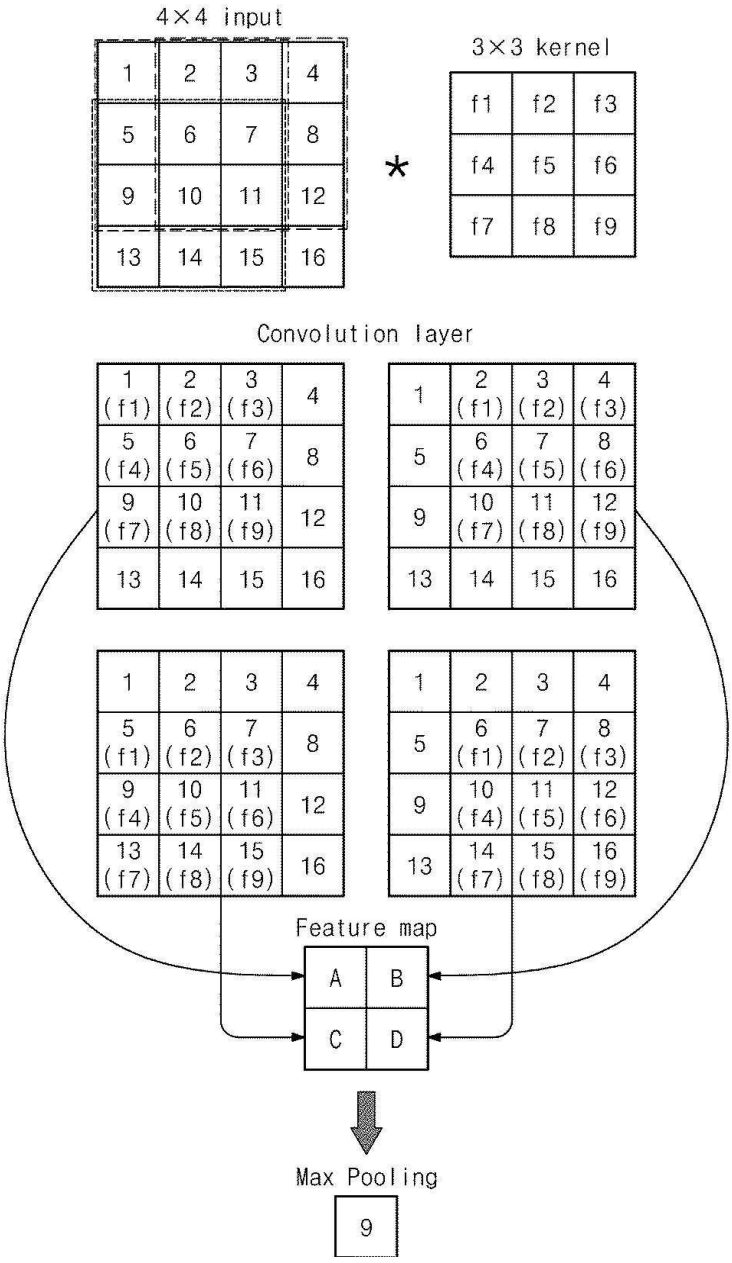
도면6


11
Diagonalized gate line
510
400
Postsynaptic
current
210
Presynaptic
current
300
110

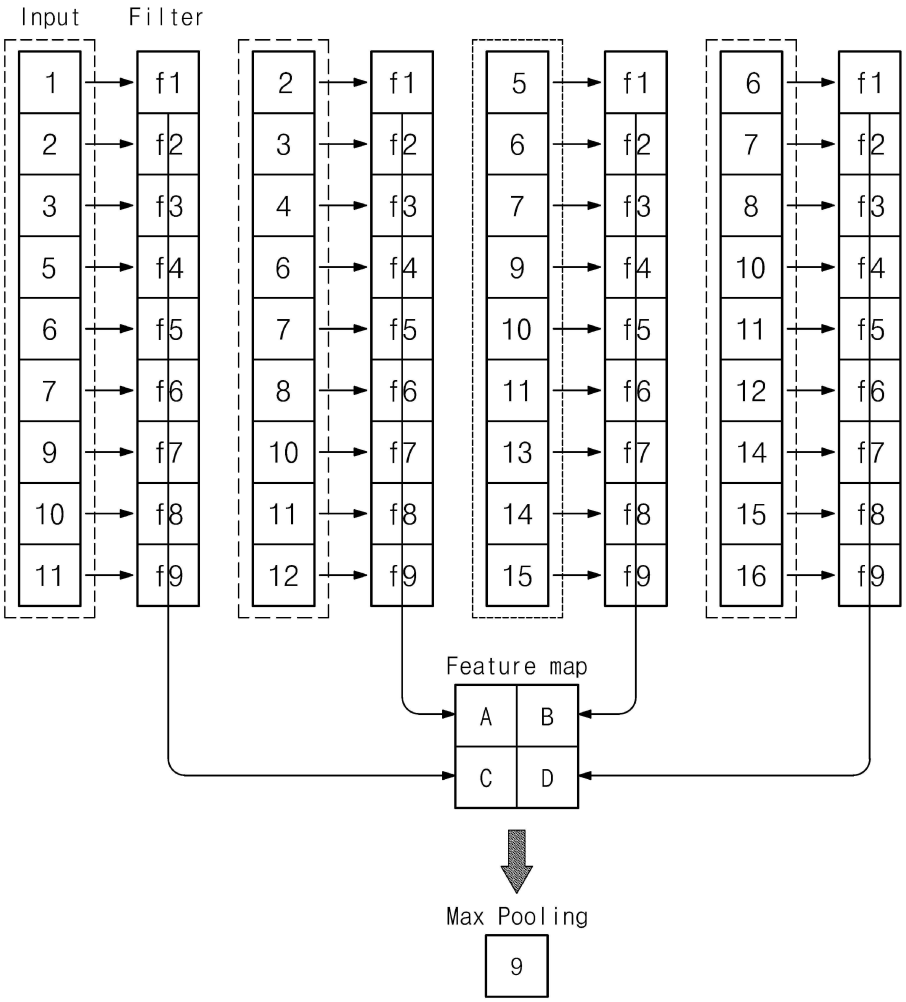
도면7


S10
220
제1 방향
제2 방향
220
120
110
110
(S100)
300
300
(S200)
210
220
220
210
(S300)
210
220
210
220
120
510
520
110
110
(S500)
400
210
210
300
300
400
(S400)

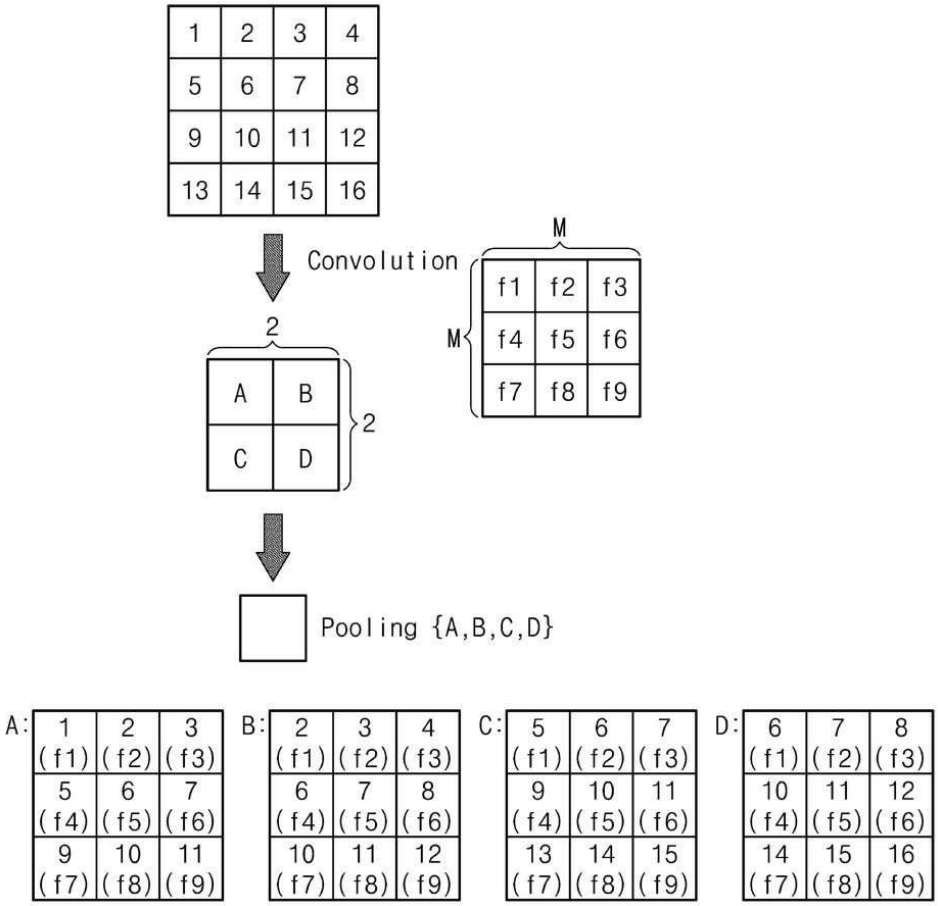
도면8


4×4 input
1 2 3 4
5 6 7 8
9 10 11 12
13 14 15 16
*
3×3 kernel
f1 f2 f3
f4 f5 f6
f7 f8 f9
Convolution layer
Feature map
A B
C D
Max Pooling
9

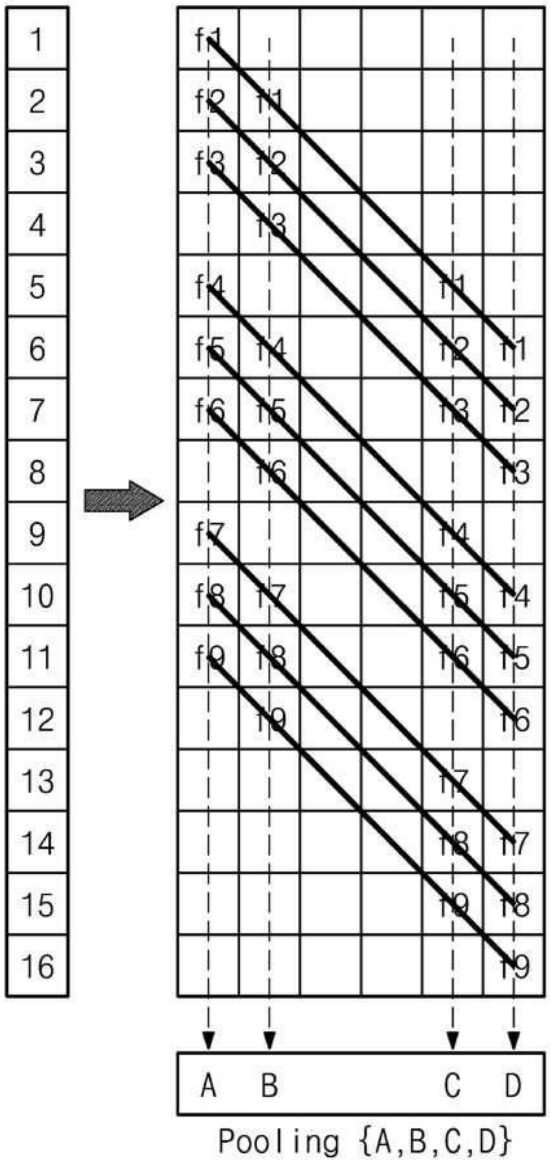
도면9


Input
Filter
1
2
3
5
6
7
9
10
11
f1
f2
f3
f4
f5
f6
f7
f8
f9
2
3
4
6
7
8
10
11
12
5
6
7
9
10
11
13
14
15
6
7
8
10
11
12
14
15
16
Feature map
A
B
C
D
Max Pooling
9

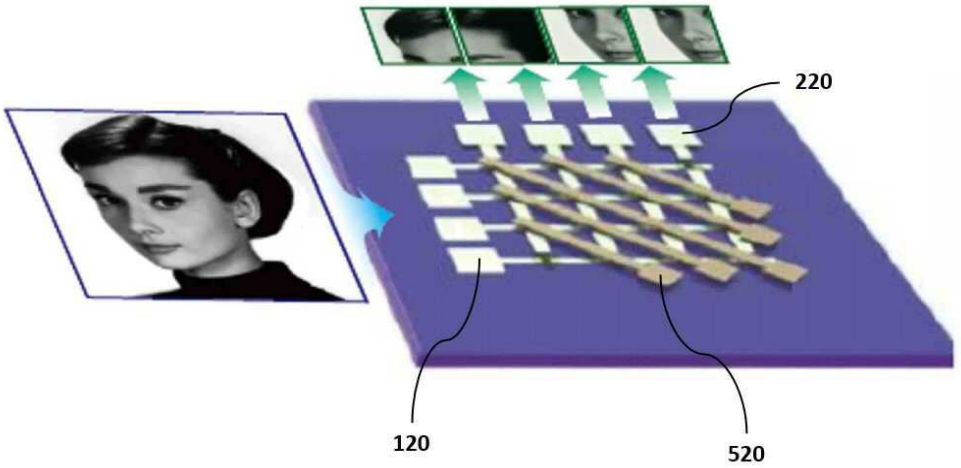
도면10


1 2 3 4
5 6 7 8
9 10 11 12
13 14 15 16
Convolution
M
f1 f2 f3
f4 f5 f6
f7 f8 f9
2
A B
C D
2
Pooling {A,B,C,D}
A:
1 (f1) 2 (f2) 3 (f3)
5 (f4) 6 (f5) 7 (f6)
9 (f7) 10 (f8) 11 (f9)
B:
2 (f1) 3 (f2) 4 (f3)
6 (f4) 7 (f5) 8 (f6)
10 (f7) 11 (f8) 12 (f9)
C:
5 (f1) 6 (f2) 7 (f3)
9 (f4) 10 (f5) 11 (f6)
13 (f7) 14 (f8) 15 (f9)
D:
6 (f1) 7 (f2) 8 (f3)
10 (f4) 11 (f5) 12 (f6)
14 (f7) 15 (f8) 16 (f9)

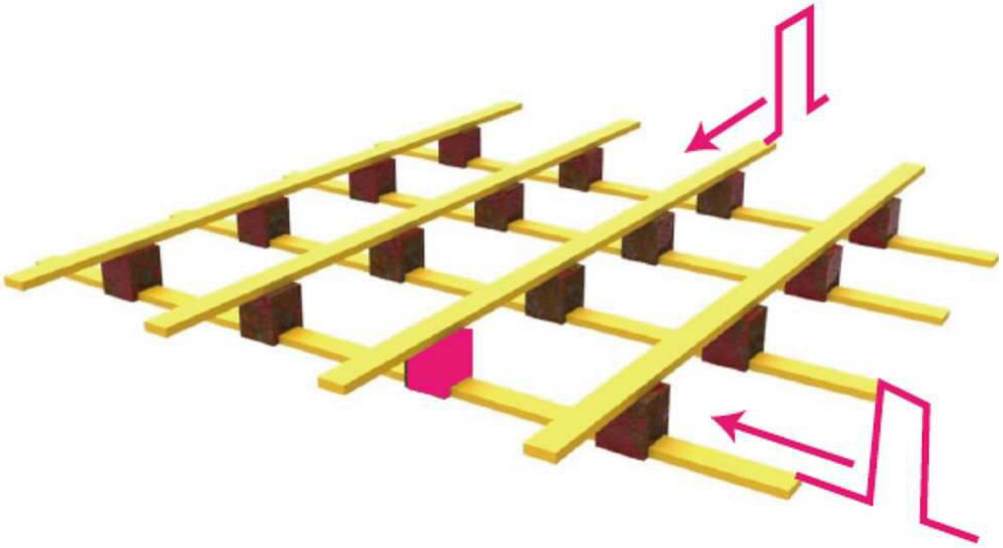
도면11



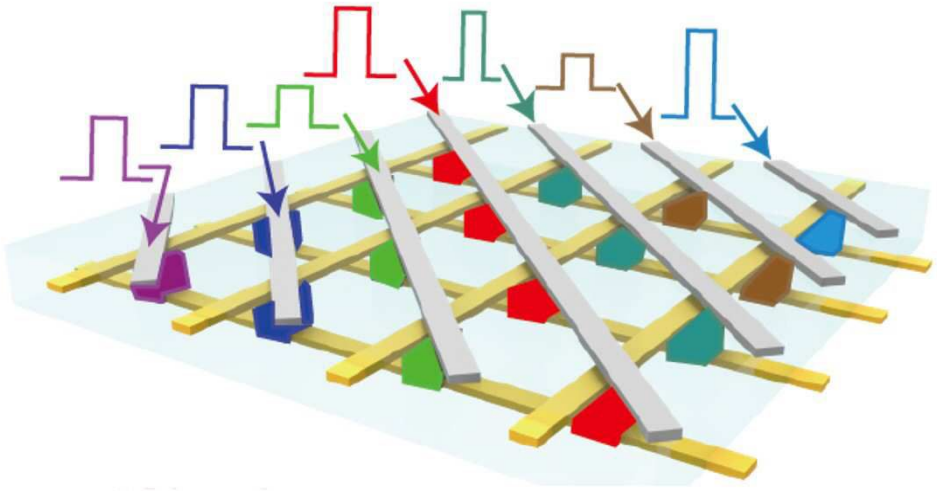
도면12



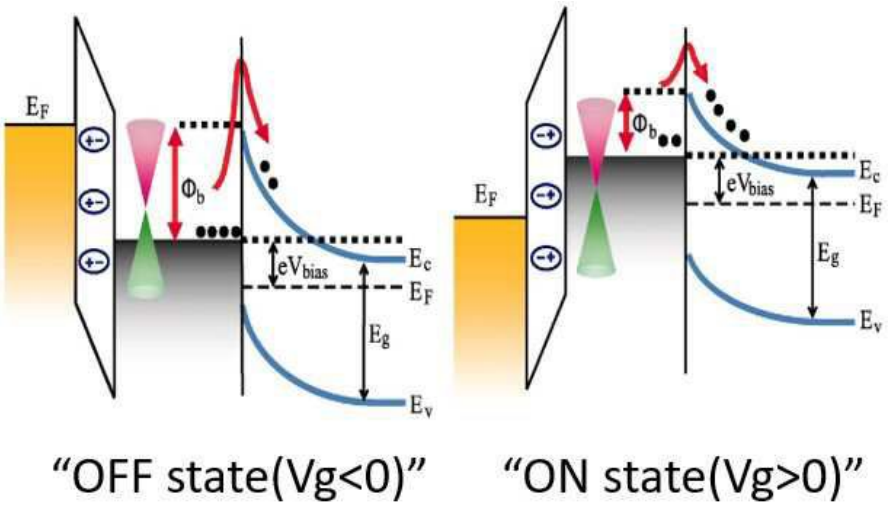
도면13



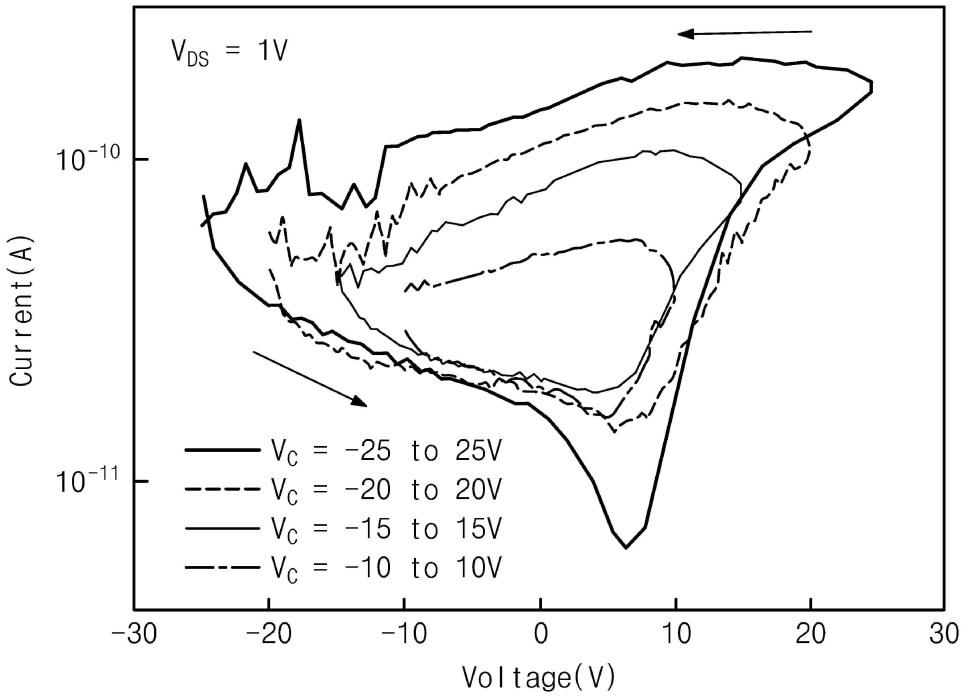
도면14



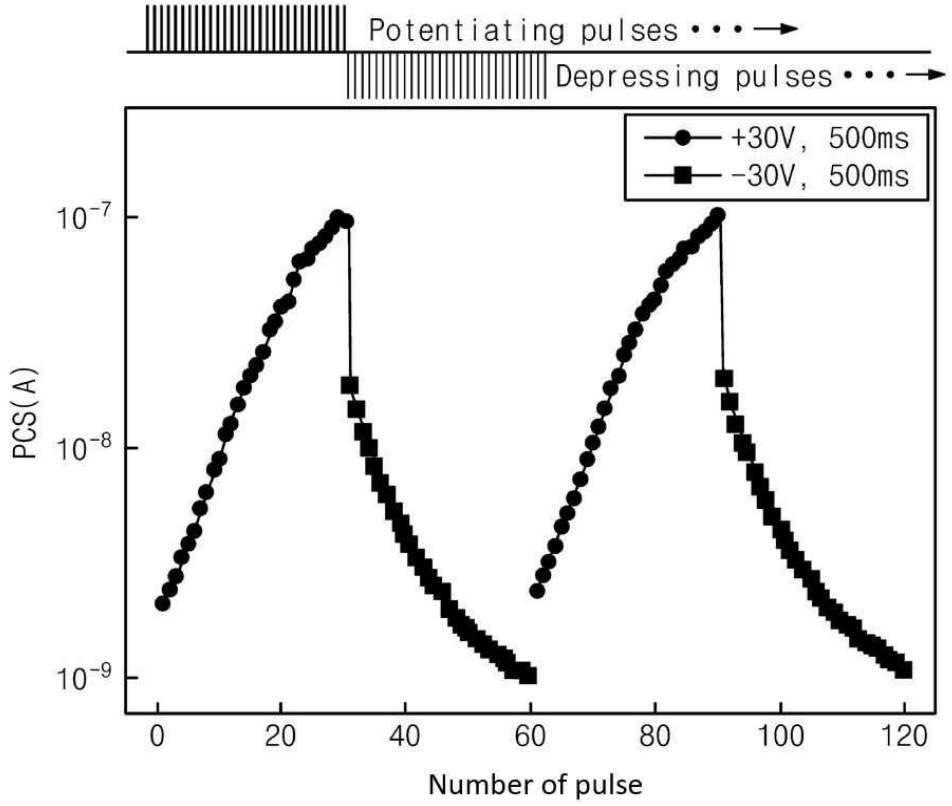
도면15



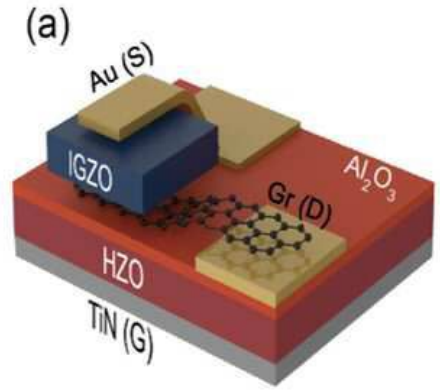
도면16


V_{DS} = 1V
10^{-10}
10^{-11}
Current(A)
V_C = -25 to 25V
V_C = -20 to 20V
V_C = -15 to 15V
V_C = -10 to 10V
-30
-20
-10
0
10
20
30
Voltage(V)

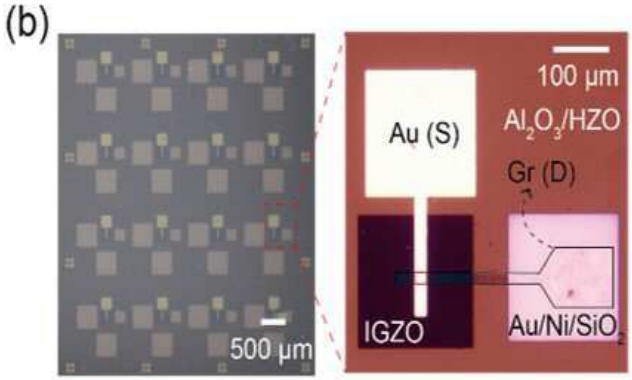
도면17


Potentiating pulses
Depressing pulses
+30V, 500ms
-30V, 500ms
10^{-7}
10^{-8}
10^{-9}
PCS(A)
0
20
40
60
80
100
120
Number of pulse

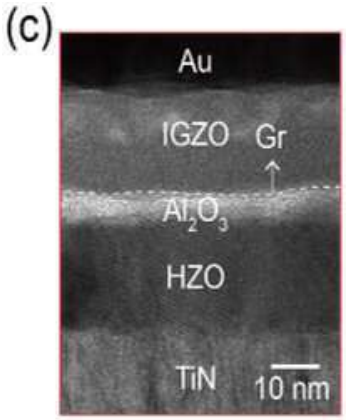
도면18a



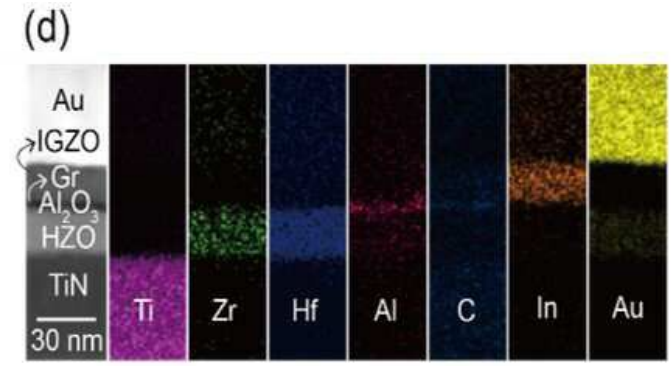
도면18b



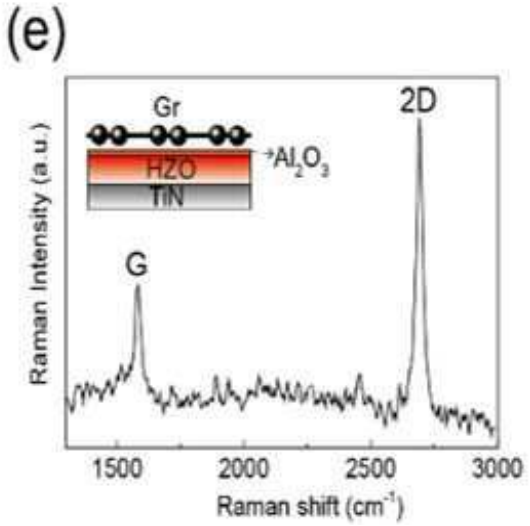
도면18c



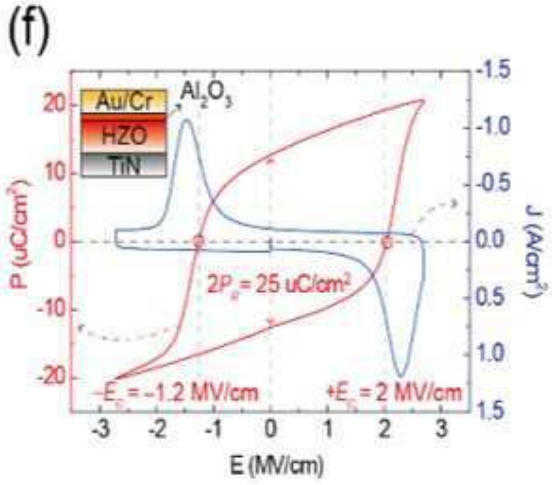
도면18d



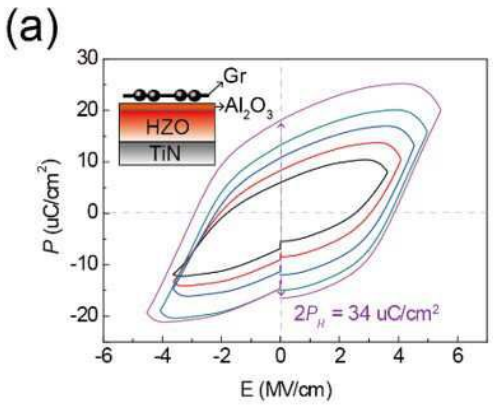
도면18e



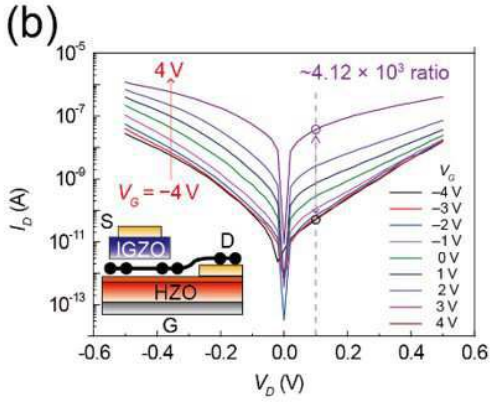
도면18f



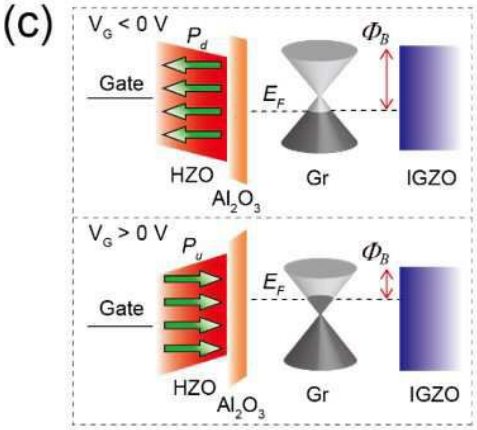
도면19a



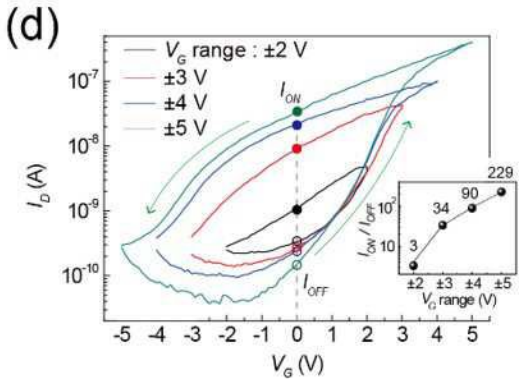
도면19b



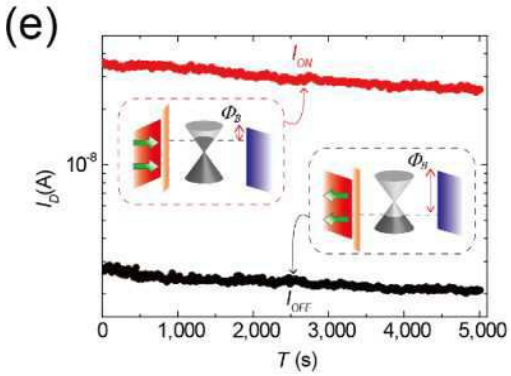
도면19c



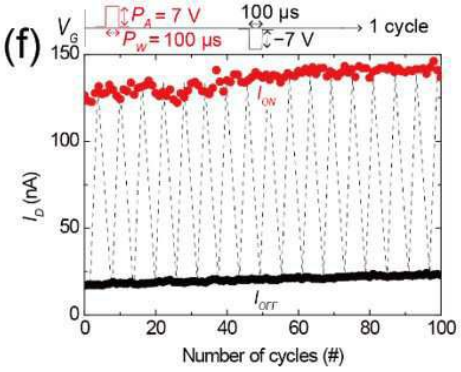
도면19d



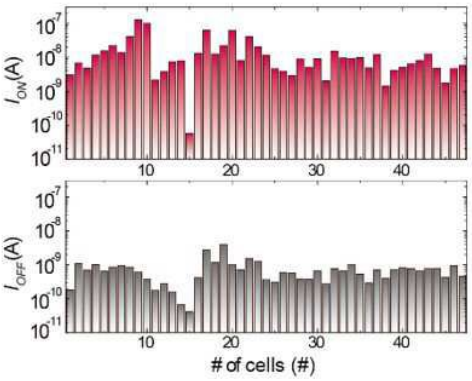
도면19e



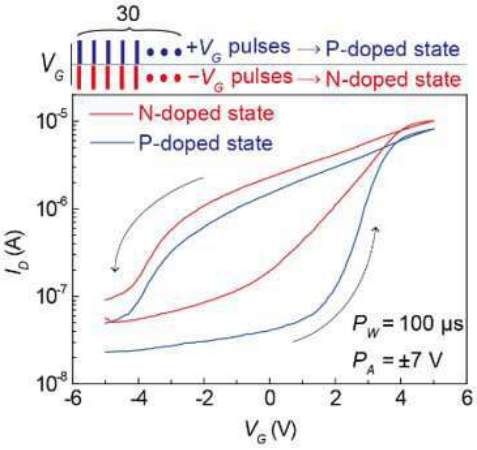
도면19f



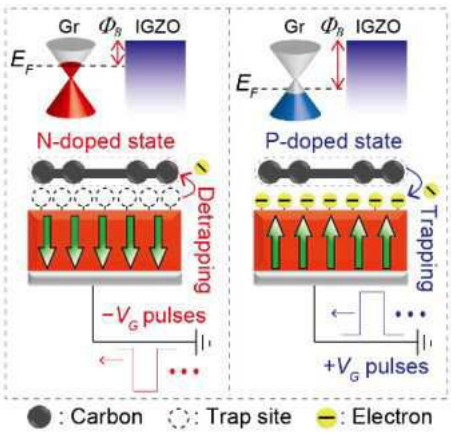
도면19g



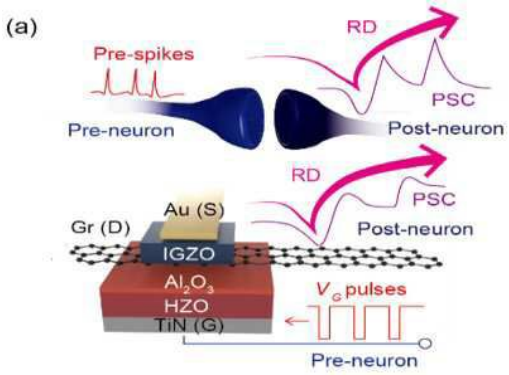
도면19h



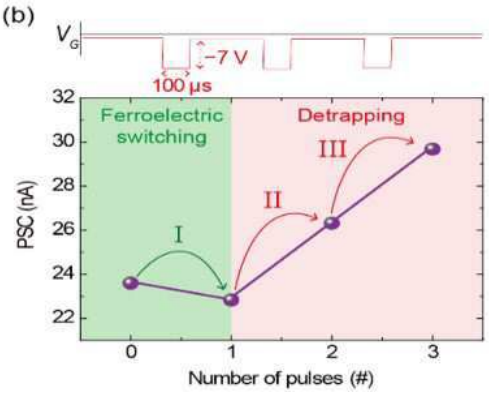
도면19i



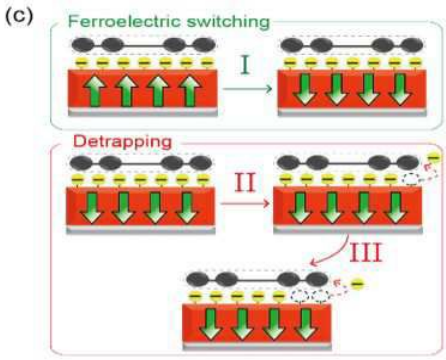
도면20a



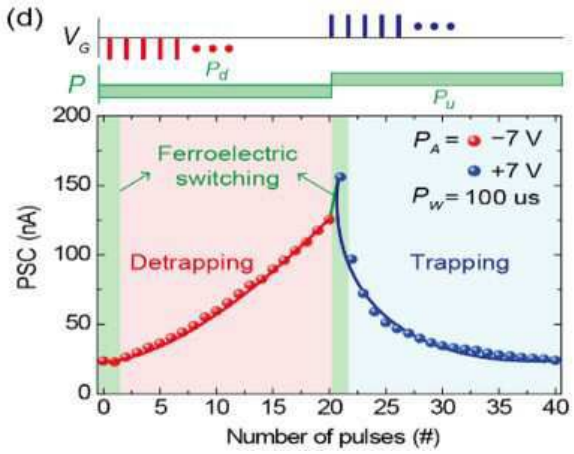
도면20b



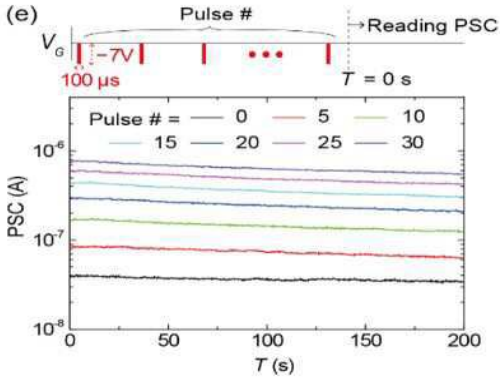
도면20c



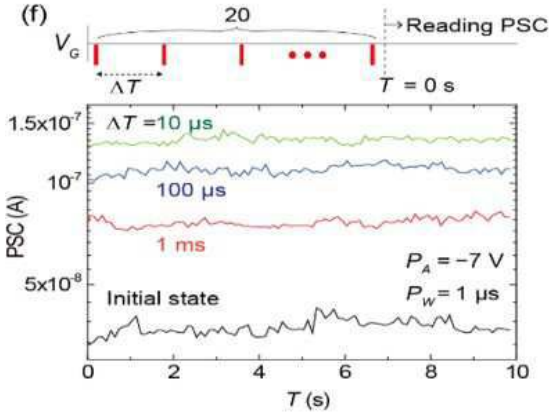
도면20d



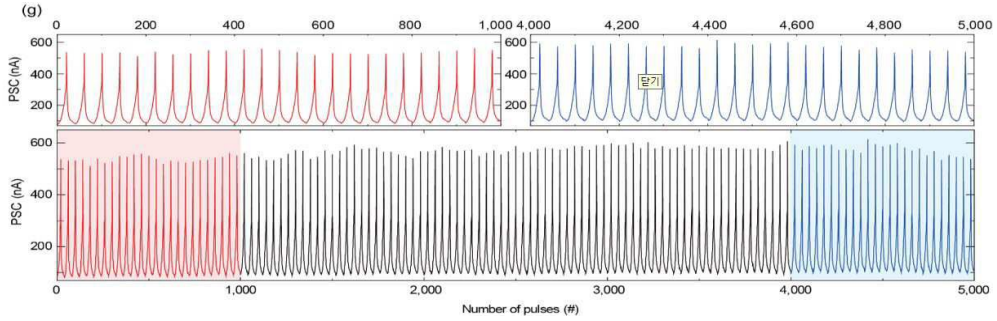
도면20e



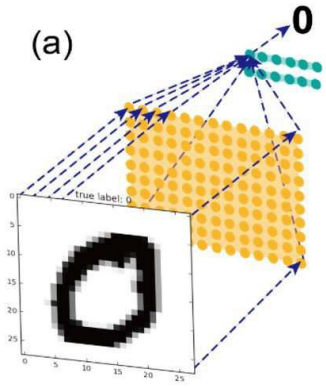
도면20f



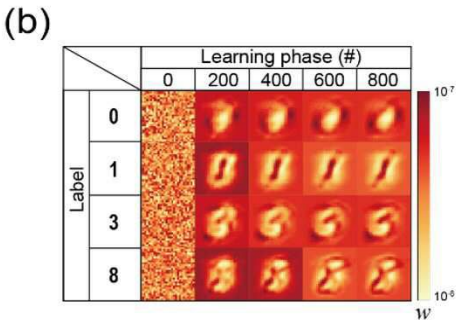
도면20g



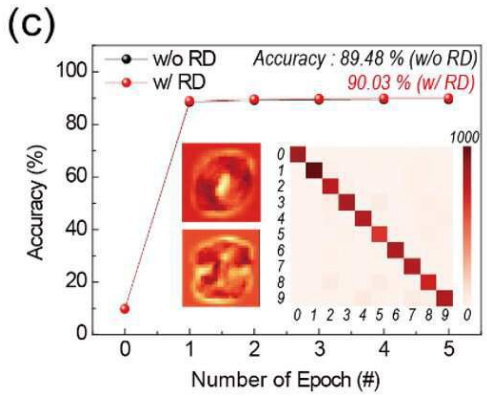
도면21a



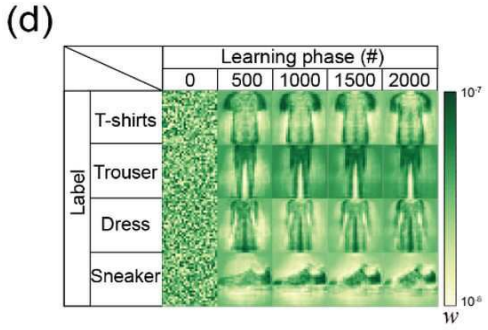
도면21b



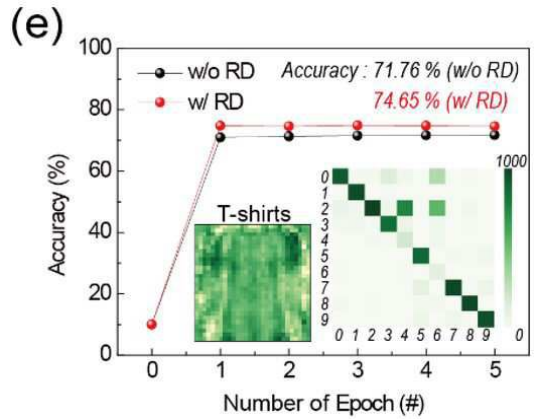
도면21c



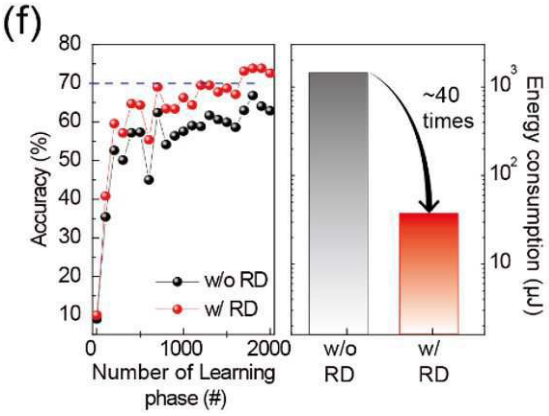
도면21d



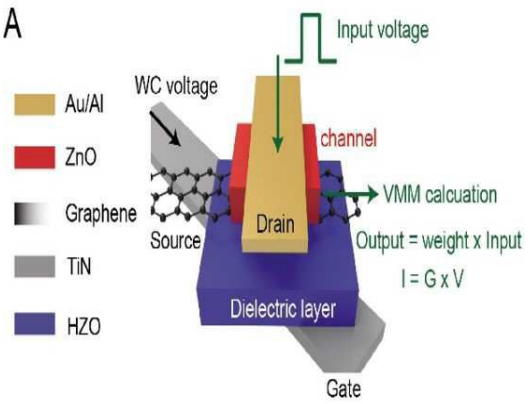
도면21e



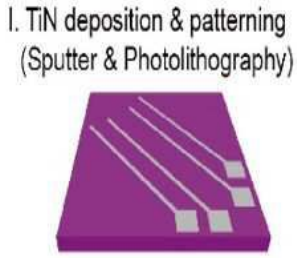
도면21f



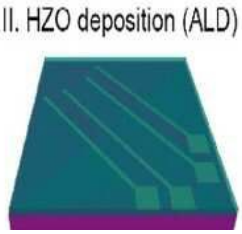
도면22a



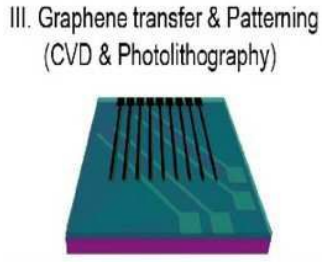
도면22b



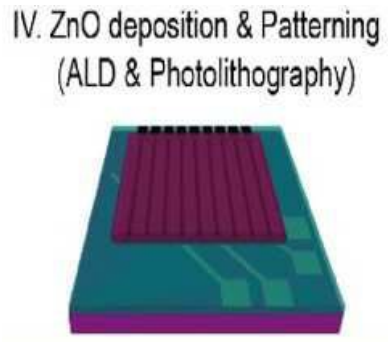
도면22c



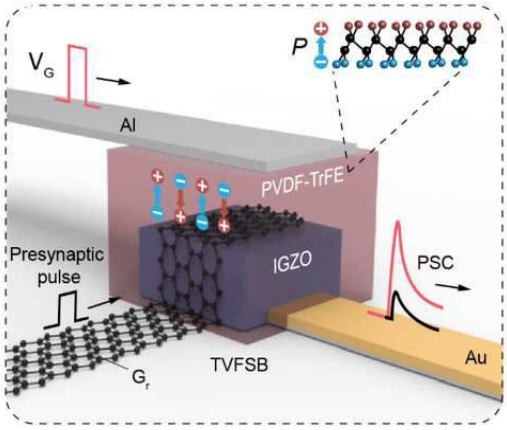
도면22d



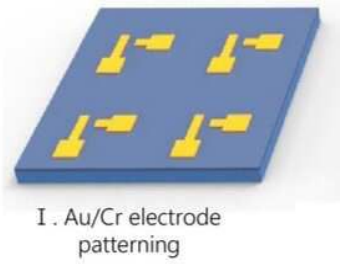
도면22e



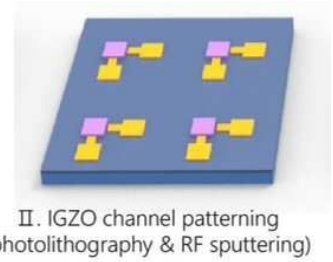
도면23a



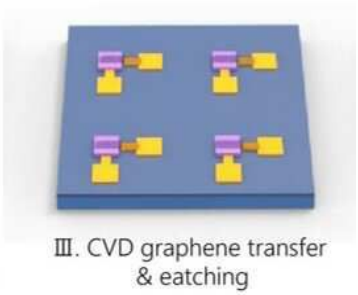
도면23b



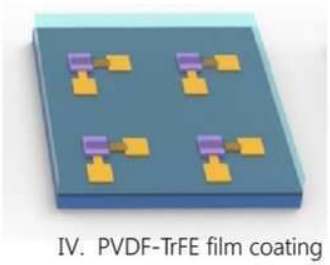
도면23c



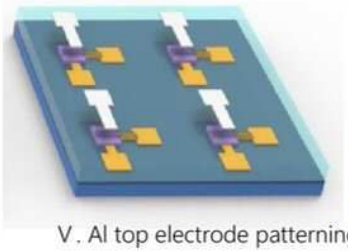
도면23d



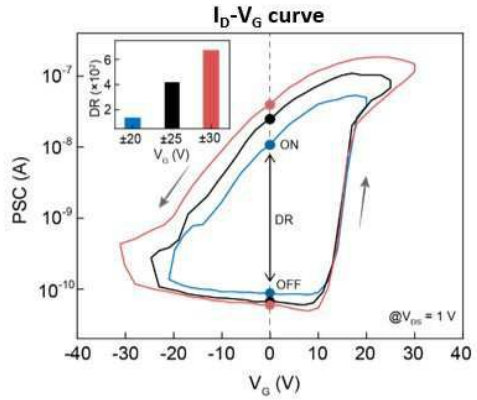
도면23e



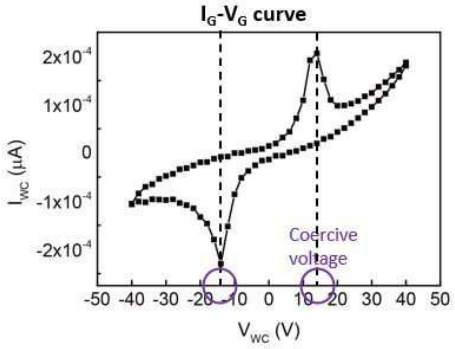
도면23f



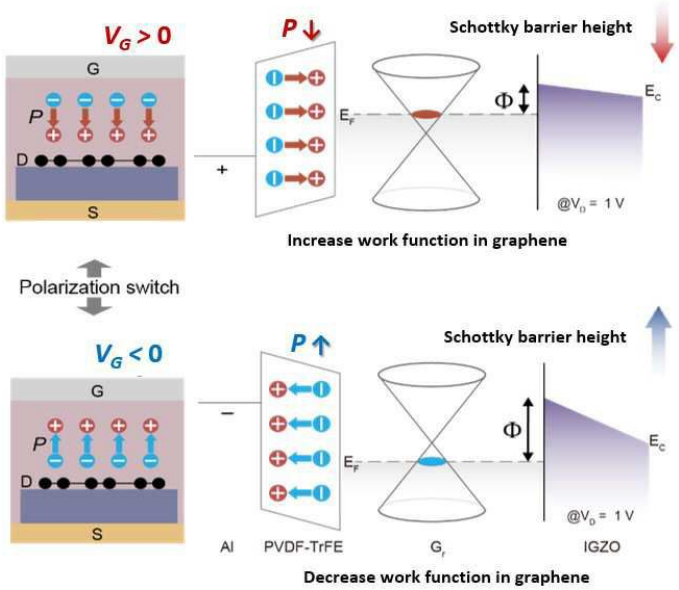
도면24a



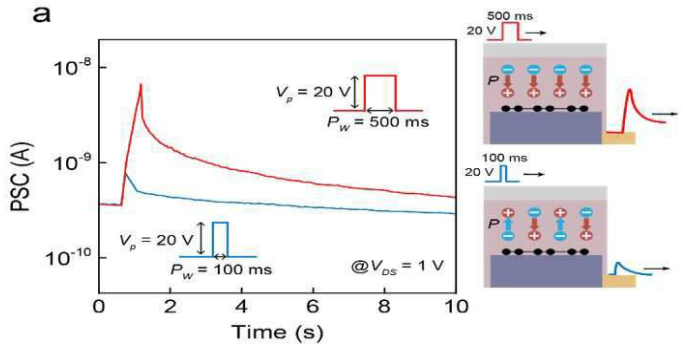
도면24b



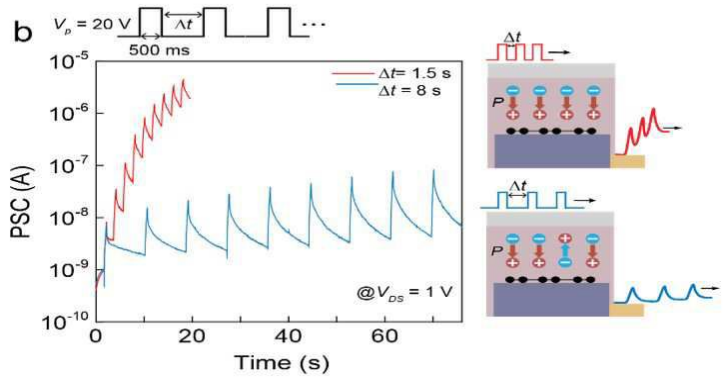
도면24c



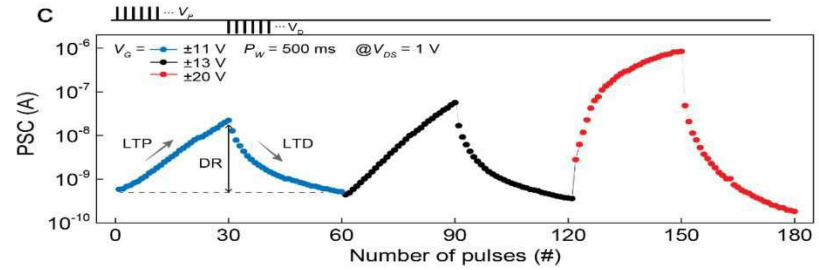
도면25a



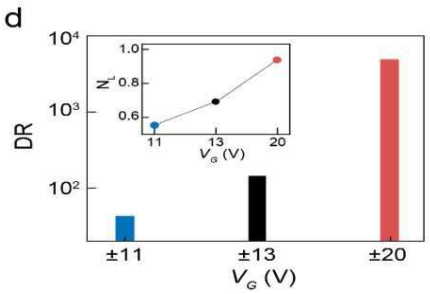
도면25b



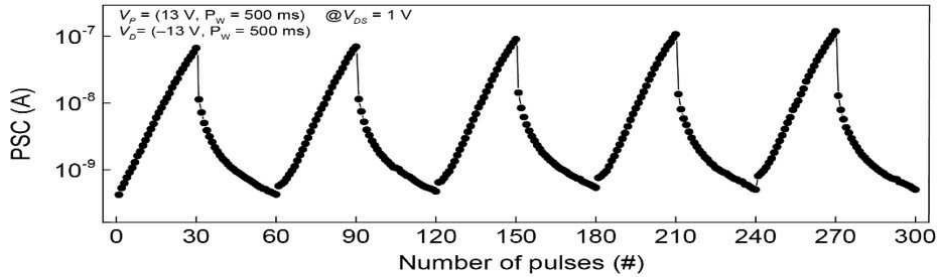
도면25c



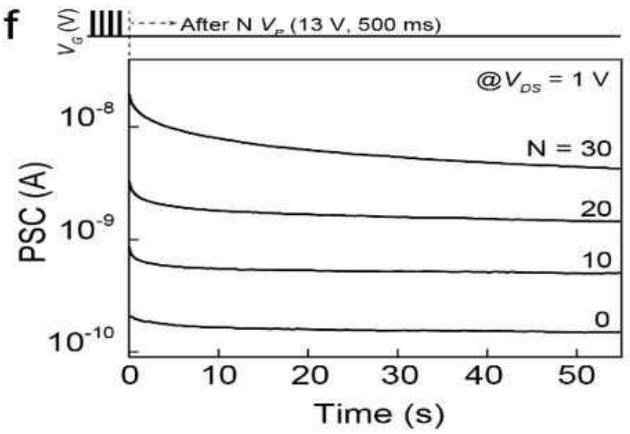
도면25d



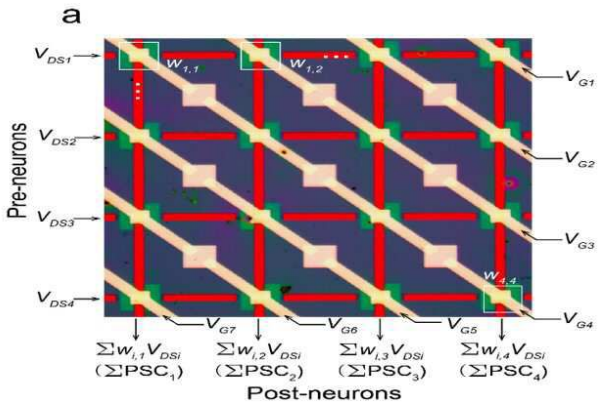
도면25e



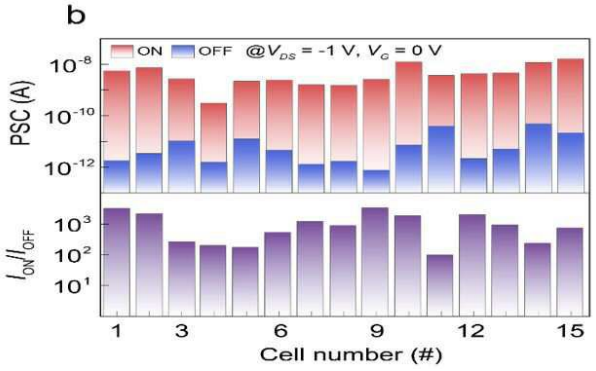
도면25f



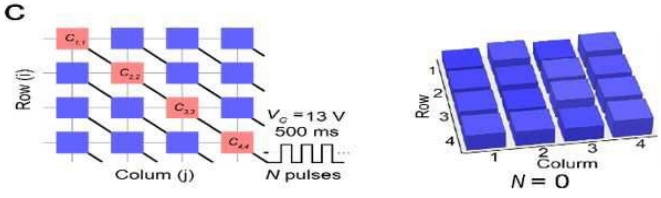
도면26a



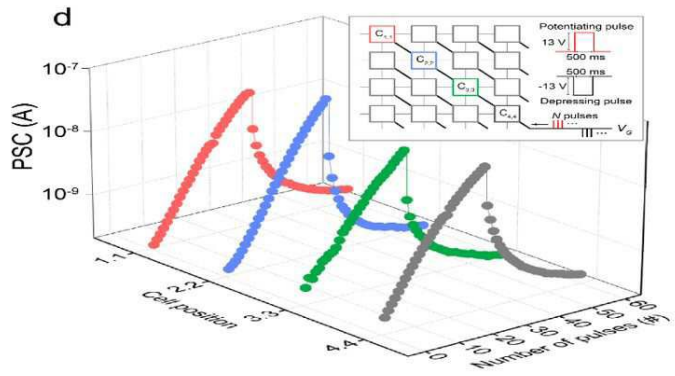
도면26b



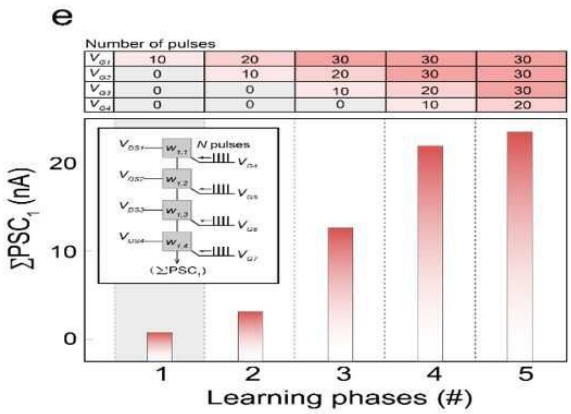
도면26c



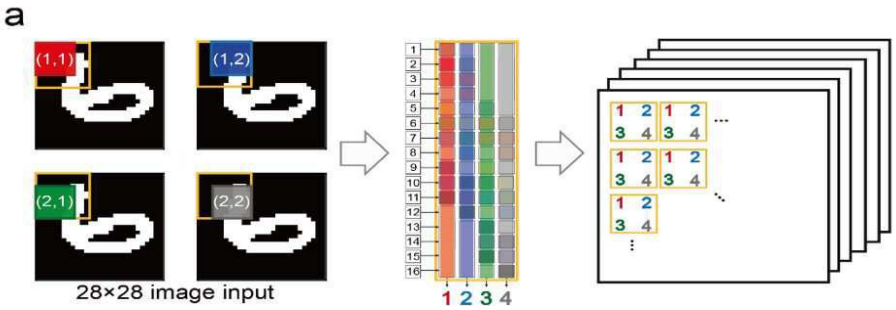
도면26d



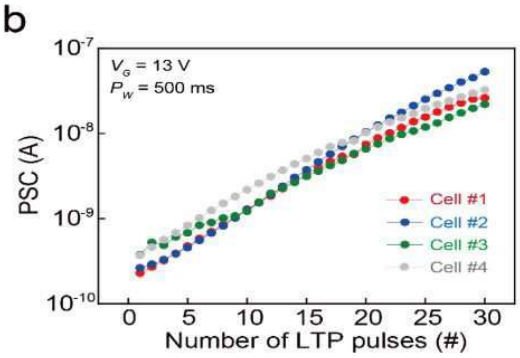
도면26e



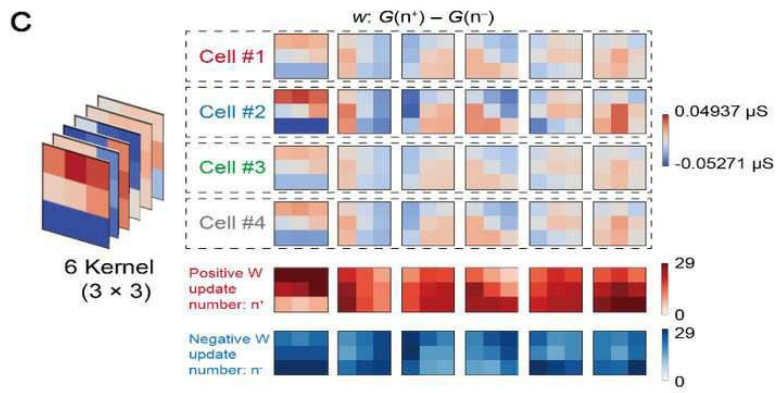
도면27a



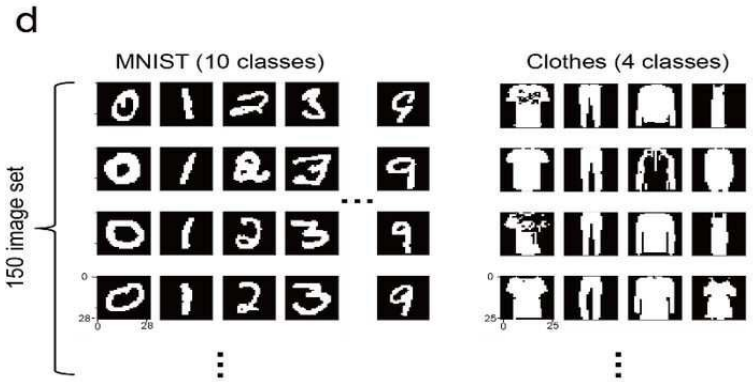
도면27b



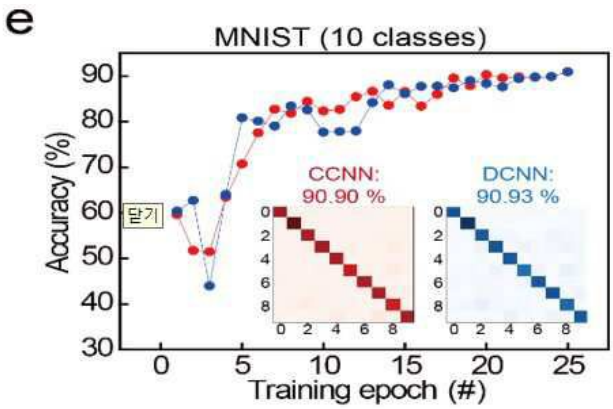
도면27c



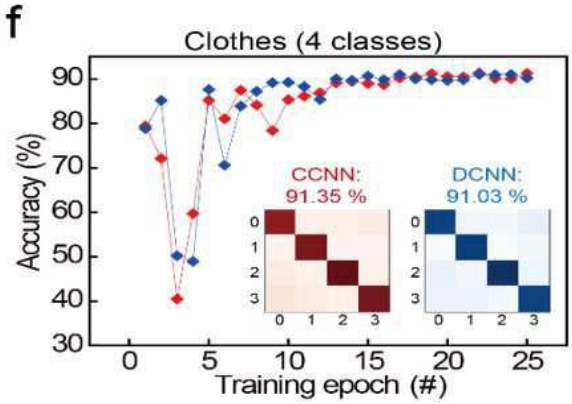
도면27d



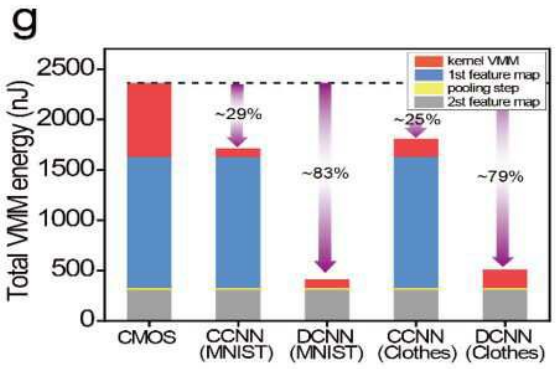
도면27e



도면27f


f
Clothes (4 classes)
Accuracy (%)
Training epoch (#)
CCNN:
91.35 %
DCNN:
91.03 %

도면27g


g
Total VMM energy (nJ)
kernel VMM
1st feature map
pooling step
2st feature map
~29%
~83%
~25%
~79%
CMOS
CCNN (MNIST)
DCNN (MNIST)
CCNN (Clothes)
DCNN (Clothes)