



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0141090
(43) 공개일자 2021년11월23일

- (51) 국제특허분류(Int. Cl.)
G06N 3/063 (2006.01) G06N 3/08 (2006.01)
- (52) CPC특허분류
G06N 3/0635 (2013.01)
G06N 3/08 (2013.01)
- (21) 출원번호 10-2020-0058240
- (22) 출원일자 2020년05월15일
심사청구일자 2020년05월15일

- (71) 출원인
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
- (72) 발명자
조만호
서울특별시 강남구 선릉로 8, 211동 1101호 (개포동, 래미안블레스티지)
- 김다솔
서울특별시 서대문구 연희로10길 69, 405호 (연희동)
- (74) 대리인
특허법인다나

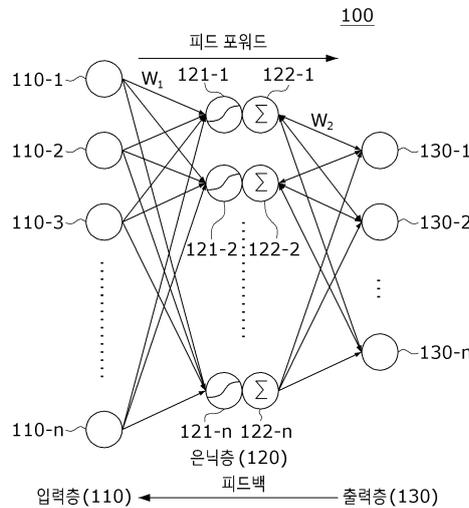
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 상변화 소자를 이용한 인공신경망 시스템

(57) 요약

실시예에 따르면, 입력 데이터를 수신하는 입력층; 상기 입력 데이터에 대응하는 출력 데이터를 출력하는 출력층; 그리고 상기 입력층과 상기 출력층 사이에 위치하는 복수의 은닉층들을 포함하며, 상기 은닉층들은 활성화 함수 연산을 수행하여 제1결과값을 출력하는 제1노드와 상기 제1노드에서 출력된 제1결과값을 합산하는 제2노드를 포함하며, 상기 제1노드의 활성화 함수는 상변화 소자의 저항-전압 특성에 따라 결정되는 상변화 소자를 이용한 인공신경망 시스템을 제공한다.

대표도 - 도1



명세서

청구범위

청구항 1

입력 데이터를 수신하는 입력층;

상기 입력 데이터에 대응하는 출력 데이터를 출력하는 출력층; 그리고

상기 입력층과 상기 출력층 사이에 위치하는 복수의 은닉층들을 포함하며,

상기 은닉층들은 활성화 함수 연산을 수행하여 제1결과값을 출력하는 제1노드와 상기 제1노드에서 출력된 제1결과값을 합산하는 제2노드를 포함하며,

상기 제1노드의 활성화 함수는 상변화 소자의 저항-전압 특성에 따라 결정되는 상변화 소자를 이용한 인공신경망 시스템.

청구항 2

제1항에 있어서,

상기 상변화 소자는 상부 메탈라인 및 하부 메탈라인,

상기 상부 메탈라인 및 상기 하부 메탈라인 사이에 배치되는 상변화 물질 및 셀렉터를 포함하는 상변화 소자를 이용한 인공신경망 시스템.

청구항 3

제2항에 있어서,

상기 상변화 소자의 전압-저항 특성은 시그모이드 함수 형태에 따라 변화하는 상변화 소자를 이용한 인공신경망 시스템.

청구항 4

제3항에 있어서,

상기 제2노드는 상기 상변화 소자에 흐르는 전류치를 합산하는 상변화 소자를 이용한 인공신경망 시스템.

청구항 5

제1항에 있어서,

상기 출력층은 상기 제2노드에서 합산한 값에 가중치를 부여하여 상기 출력 데이터를 출력하는 상변화 소자를 이용한 인공신경망 시스템.

청구항 6

제2항에 있어서,

상기 상변화 물질은 Ge-Sb-Te계 물질로 이루어지는 상변화 소자를 이용한 인공신경망 시스템.

청구항 7

제2항에 있어서,

상기 상변화 물질은 Ag-In-Sb-Te계 물질로 이루어지는 상변화 소자를 이용한 인공신경망 시스템.

청구항 8

제1항에 있어서,

상기 상변화 물질은 Ti-Sb-Te 계 물질로 이루어지는 상변화 소자를 이용한 인공신경망 시스템.

청구항 9

제1항에 있어서,

상기 상변화 소자의 전압-저항 특성은 ReLu함수 형태에 따라 변화하는 상변화 소자를 이용한 인공신경망 시스템.

청구항 10

제1항에 있어서,

상기 상변화 소자의 전압-저항 특성은 복수개의 상변화 물질이 조합된 함수 형태에 따라 변화하는 상변화 소자를 이용한 인공신경망 시스템.

발명의 설명

기술 분야

[0001] 본 발명의 일실시예는 상변화 소자를 이용한 인공신경망 시스템에 관한 것으로, 구체적으로는 3D 크로스 포인트 (3D Xpoint) 메모리 소자에 적용될 수 있는 상변화 소자를 이용한 인공신경망 시스템에 관한 것이다.

배경 기술

[0003] 3D 크로스 포인트 기술은 동작 속도가 빠르고 비휘발성인 상변화 물질에 기초한다. 해당 기술의 단기 목표는 DRAM과 NAND Flash 사이의 속도 병목을 해소하는 새로운 형태의 메모리/스토리지 모듈로 활용하는 것이고, 장기 목표는 DRAM과 NAND Flash 모두를 대체하는 스토리지 클래스 메모리로 활용되는 것이다.

[0004] 상변화 물질은 그 특성에 따라서, 3D 크로스 포인트 메모리/스토리지 모듈의 신뢰 범위를 조정할 수 있다. 이러한 3D 크로스 포인트 기술은 뉴로모픽(neuromorphic) 칩으로의 활용 가능성으로도 주목을 많이 받는데, 이는 딥러닝(deep learning)에 최적화된 병렬 계산에 적합한 구조이기 때문이다.

[0005] 3D 크로스 포인트 기술은 인가한 펄스 횟수에 따라 저항이 선형적으로 올라가는 정도를 나타내는 선형성(linearity)과 셋(set)-리셋(reset)펄스의 교번 인가에 따라 저항 레벨이 대칭적으로 변화하는 대칭성(symmetry)이 동시에 우수한 상변화 물질 개발이 절대적으로 필요한 실정이다. 이를 통하여 옴의 법칙(ohm's law)과 키르히호프의 전류 법칙(Kirchhoff's current law)를 인공신경망 시스템에 직접 적용해 1-PCM/cell로 사용할 수 있게 된다. 하지만 현재까지는 인공신경망 시스템에서 효과적인 성능을 발휘할 수 있을 정도의 선형성과 대칭성을 만족하는 상변화 물질 개발은 매우 느린 실정이다.

발명의 내용

해결하려는 과제

[0007] 본 발명이 이루고자 하는 기술적 과제는 3D 크로스 포인트 메모리 소자에 최적화 된 상변화 소자를 이용한 인공 신경망 시스템을 제공하는데 있다.

과제의 해결 수단

[0009] 실시예에 따르면, 입력 데이터를 수신하는 입력층; 상기 입력 데이터에 대응하는 출력 데이터를 출력하는 출력층; 그리고 상기 입력층과 상기 출력층 사이에 위치하는 복수의 은닉층들을 포함하며, 상기 은닉층들은 활성화 함수 연산을 수행하여 제1결과값을 출력하는 제1노드와 상기 제1노드에서 출력된 제1결과값을 합산하는 제2노드를 포함하며, 상기 제1노드의 활성화 함수는 상변화 소자의 저항-전압 특성에 따라 결정되는 상변화 소자를 이용한 인공신경망 시스템을 제공한다.

[0010] 상기 상변화 소자는 상부 메탈라인 및 하부 메탈라인, 상기 상부 메탈라인 및 상기 하부 메탈라인 사이에 배치되는 상변화 물질 및 셀렉터를 포함할 수 있다.

[0011] 상기 상변화 소자의 전압-저항 특성은 시그모이드 함수 형태에 따라 변화할 수 있다.

[0012] 상기 제2노드는 상기 상변화 소자에 흐르는 전류치를 합산할 수 있다.

[0013] 상기 출력층은 상기 제2노드에서 합산한 값에 가중치를 부여하여 상기 출력 데이터를 출력할 수 있다.

[0014] 상기 상변화 물질은 Ge-Sb-Te계 물질로 이루어질 수 있다.

[0015] 상기 상변화 물질은 Ag-In-Sb-Te계 물질로 이루어질 수 있다.

[0016] 상기 상변화 물질은 Ti-Sb-Te 계 물질로 이루어질 수 있다.

[0017] 상기 상변화 소자의 전압-저항 특성은 ReLu함수 형태에 따라 변화할 수 있다.

[0018] 상기 상변화 소자의 전압-저항 특성은 복수개의 상변화 물질이 조합된 함수 형태에 따라 변화할 수 있다.

발명의 효과

[0020] 본 발명인 상변화 소자를 이용한 인공신경망 시스템은 3D 크로스 포인트 메모리 소자에 최적화 된 인공신경망 시스템을 제공할 수 있다.

[0021] 또한, 연산 코스트를 크게 절감시킬 수 있다.

[0022] 또한, 제작 단가 인하 및 수율 향상에 크게 기여할 수 있다.

도면의 간단한 설명

[0024] 도1은 실시예에 따른 인공신경망 시스템의 개념도이다.

도2는 실시예에 따른 상변화 소자의 개념도이다.

도3 및 도4는 실시예에 따른 상변화 소자의 전압-저항 그래프이다.

도5는 Ge-Sb-Te계 상변화 물질의 전압-저항 특성을 도시한 그래프이다.

도6은 Ag-In-Sb-Te계 상변화 물질의 전압-저항 특성을 도시한 그래프이다.

발명을 실시하기 위한 구체적인 내용

- [0025] 이하, 첨부된 도면을 참조하여 본 발명의 바람직한 실시예를 상세히 설명한다.
- [0026] 다만, 본 발명의 기술 사상은 설명되는 일부 실시 예에 한정되는 것이 아니라 서로 다른 다양한 형태로 구현될 수 있고, 본 발명의 기술 사상 범위 내에서라면, 실시 예들간 그 구성 요소들 중 하나 이상을 선택적으로 결합, 치환하여 사용할 수 있다.
- [0027] 또한, 본 발명의 실시예에서 사용되는 용어(기술 및 과학적 용어를 포함)는, 명백하게 특별히 정의되어 기술되지 않는 한, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 일반적으로 이해될 수 있는 의미로 해석될 수 있으며, 사전에 정의된 용어와 같이 일반적으로 사용되는 용어들은 관련 기술의 문맥상의 의미를 고려하여 그 의미를 해석할 수 있을 것이다.
- [0028] 또한, 본 발명의 실시예에서 사용된 용어는 실시예들을 설명하기 위한 것이며 본 발명을 제한하고자 하는 것은 아니다.
- [0029] 본 명세서에서, 단수형은 문구에서 특별히 언급하지 않는 한 복수형도 포함할 수 있고, "A 및(와) B, C 중 적어도 하나(또는 한 개 이상)"로 기재되는 경우 A, B, C로 조합할 수 있는 모든 조합 중 하나 이상을 포함할 수 있다.
- [0030] 또한, 본 발명의 실시 예의 구성 요소를 설명하는 데 있어서, 제1, 제2, A, B, (a), (b) 등의 용어를 사용할 수 있다.
- [0031] 이러한 용어는 그 구성 요소를 다른 구성 요소와 구별하기 위한 것일 뿐, 그 용어에 의해 해당 구성 요소의 본질이나 차례 또는 순서 등으로 한정되지 않는다.
- [0032] 그리고, 어떤 구성 요소가 다른 구성요소에 '연결', '결합' 또는 '접속'된다고 기재된 경우, 그 구성 요소는 그 다른 구성 요소에 직접적으로 연결, 결합 또는 접속되는 경우뿐만 아니라, 그 구성 요소와 그 다른 구성 요소 사이에 있는 또 다른 구성 요소로 인해 '연결', '결합' 또는 '접속' 되는 경우도 포함할 수 있다.
- [0033] 또한, 각 구성 요소의 "상(위) 또는 하(아래)"에 형성 또는 배치되는 것으로 기재되는 경우, 상(위) 또는 하(아래)는 두 개의 구성 요소들이 서로 직접 접촉되는 경우뿐만 아니라 하나 이상의 또 다른 구성 요소가 두 개의 구성 요소들 사이에 형성 또는 배치되는 경우도 포함한다. 또한, "상(위) 또는 하(아래)"으로 표현되는 경우 하나의 구성 요소를 기준으로 위쪽 방향뿐만 아니라 아래쪽 방향의 의미도 포함할 수 있다.
- [0034] 이하, 첨부된 도면을 참조하여 실시예를 상세히 설명하되, 도면 부호에 관계없이 동일하거나 대응하는 구성 요소는 동일한 참조 번호를 부여하고 이에 대한 중복되는 설명은 생략하기로 한다.
- [0036] 도1은 실시예에 따른 인공신경망 시스템의 개념도이다. 도1을 참조하면, 실시예에 따른 상변화 소자를 이용한 인공신경망 시스템은 입력층(110), 은닉층(120) 및 출력층(130)을 포함할 수 있다.
- [0037] 상변화 소자를 이용한 인공신경망 시스템(100)은 생물학적인 뉴런의 기능을 단순화시킨 인공 뉴런들이 이용되고, 인공 뉴런들은 연결 가중치(W_1, W_2, \dots, W_n)를 가지는 연결선을 통해 상호 연결될 수 있다. 연결 가중치(W_1, W_2, \dots, W_n)는 연결선이 갖는 특정한 값으로서 연결 강도라고도 표현할 수 있다. 인공 뉴런은 노드로 지칭될 수 있다.
- [0038] 입력층(110), 은닉층(120) 및 출력층(130)은 복수의 노드(110-1, \dots , 110-n)들을 포함할 수 있다. 예를 들면, 입력층(110)은 복수의 입력 노드(110-1, \dots , 110-n)들을 포함할 수 있다. 각 은닉층은 복수의 은닉 노드(121-1, 122-1, \dots , 121-n, 122-n)들을 포함할 수 있다. 출력층(130)은 복수의 출력 노드(130-1, \dots , 130-n)들을 포함할 수 있다. 각 층(입력층, 은닉층, 출력층)에 포함된 노드의 수는 서로 다를 수 있다.
- [0039] 입력층(110)은 입력 데이터를 수신할 수 있다. 입력층(110)을 구성하는 복수의 입력 노드(110-1, \dots , 110-n)들은 각각 입력값들을 수신할 수 있다. 즉, 하나의 노드에는 하나의 입력값이 입력될 수 있다. 실시예에 따른, 상변화 소자를 이용한 인공신경망 시스템(100)은 입력에 대응하는 출력을 제공할 수 있다. 입력층(110)은 학습을 수행하기 위한 입력을 수신하여 복수의 은닉층(120)들로 전달할 수 있다. 예를 들면, 입력은 음성 데이터, 영상 데이터, 생체 데이터 및 필적 데이터 등을 포함할 수 있다.
- [0040] 복수의 은닉층(120)들은 제1 은닉층 내지 제n은닉층을 포함할 수 있다. 복수의 은닉층들은 입력층을 통해 전달된 학습 데이터를 예측하기 쉬운 값으로 변화시킬 수 있다. 각 은닉층(120)은 이전 은닉층에서 수신된 입력 값

을 특정한 기준에 기초하여다음 은닉층으로 전달하고, 전달 시 은닉층마다 서로 다른 연결 가중치를 적용하여 전달할 수 있다. 실시예에서는 설명의 편의를 위하여 하나의 은닉층을 도시하였으나, 이와는 달리 복수개의 은닉층이 본 발명에 적용될 수 있다.

- [0041] 은닉층(120)의 노드(121-1, 122-1, . . . , 121-n, 122-n)는 노드 입력들을 수신할 수 있다. 노드 입력들은 입력값들을 포함할 수 있다. 입력값들은 이전 은닉층의 노드들로부터 수신될 수 있다. 입력값들은 각각 연결 가중치와 곱해질 수 있다. 가중치는 이전 은닉층의 노드들과의 연결선과 일대일로 대응할 수 있다. 입력값들 각각과 연결 가중치 각각의 곱은 전달 함수에 의해 합해질 수 있다. 활성화 함수는 전달 함수의 출력값의 범위를 제한하여 노드 출력을 출력할 수 있다.
- [0042] 활성화 함수는 전달 함수의 출력값의 범위를 제한하고 계산의 편의성을 제공하며, 인공 신경망이 단순한 계산이 아닌 복잡한 판단을 할 수 있게끔 하는 역할을 한다. 따라서, 어떠한 활성화 함수를 사용하는지에 따라 상변화 소자를 이용한 인공신경망 시스템(100)의 학습 효율성이 달라질 수 있다.
- [0043] 은닉층(120)은 제1노드(121-1, . . . , 121-n) 및 제2노드(122-1, . . . , 122-n)를 포함할 수 있다. 은닉층(120)들은 활성화 함수 연산을 수행하여 제1결과값을 출력하는 제1노드(121-1, . . . , 121-n)와 제1노드(121-1, . . . , 121-n)에서 출력된 제1결과값을 합산하는 제2노드(122-1, . . . , 122-n)를 포함할 수 있다.
- [0044] 실시예에서, 제1노드(121-1, . . . , 121-n)의 활성화 함수는 상변화 소자의 저항-전압 특성에 따라 결정될 수 있다.
- [0045] 도2는 실시예에 따른 상변화 소자의 개념도이다. 도2를 참조하면, 상변화 소자(1210)로 구성되는 제1노드(121-1, . . . , 121-n)는 워드 메탈라인(1213) 및 컬럼 메탈라인(1211, 1212), 워드 메탈라인(1213) 및 컬럼 메탈라인(1211, 1212) 사이에 배치되는 상변화 물질 및 셀렉터(1215, 1217)를 포함할 수 있다.
- [0046] 상변화 소자(1210)는 워드(word) 라인(1213)과 컬럼(column) 라인(1211, 1212)의 교차점(cross point)에 상변화 물질로 이루어지는 메모리 셀(1214, 1216)을 포함하며, 이를 3D형태로 하여 2층으로 쌓은 구조를 가질 수 있다.
- [0047] 상변화 소자(1210)는 최상단의 메탈라인 4개가 2층의 컬럼 라인(1211)이 되고, 맨 아래 동일한 형태의 메탈라인 4개는 1층의 컬럼 라인(1212)이 된다. 또한, 가운데 컬럼 라인(1211, 1212)과 직교하는 방향의 메탈라인 4개는 워드 라인(1213)으로 1, 2층의 메모리 셀(1214, 1216)이 이를 공유하는 형태를 가질 수 있다.
- [0048] 메모리 셀(1214, 1216)은 워드 라인(1213)과 컬럼 라인(1211, 1212)이 교차하는 부분에 배치되어 있는 상변화 물질로 구성될 수 있으며, 각 메모리 셀(1214, 1216)마다 붙어있는 셀렉터(1215, 1217)에 가해지는 전압에 따라 읽기 동작 또는 쓰기 동작인지 여부가 결정될 수 있다.
- [0049] 다시 도1을 참조하면, 실시예에서 상변화 소자의 전압-저항 특성은 시그모이드 함수 형태에 따라 변화할 수 있다. 인공신경망 시스템(100)의 활성화 함수는 비선형적인 특성을 가질 수록 더 깊은 학습을 수행할 수 있다. 시그모이드 함수는 바이너리 분류(binary classification)에 적절한 함수로서, 일정 값을 기준으로 0 또는 1의 값으로 구분함으로써 입력 데이터를 분류할 수 있다. 시그모이드 함수는 인공신경망 시스템(100)에서 특정 임계치를 넘을 때만 활성화되기 때문에 활성화 함수로 사용될 수 있다.
- [0050] 도3 및 도4는 실시예에 따른 상변화 소자의 전압-저항 그래프이다. 도3을 참조하면, 실시예에 따른 상변화 소자에 셋(set) 펄스 폭(pulse width(ns))을 일정하게 유지하면서 전압을 점점 강하게 인가하면, 일정 전압에서 저항의 값이 급격하게 감소하는 구간을 발견할 수 있다. 즉, 상변화 소자는 일정 범위의 전압에서는 일정한 저항값을 유지하다가 임계 전압치를 초과하는 전압이 인가되면 저항값이 급격하게 감소하여 유지되는 S자 곡선 형태를 취하고 있다. 이러한 S자 곡선 형태는 일정값을 기준으로 0 또는 1의 값을 출력하는 시그모이드 함수와 유사한 곡선 형태를 가짐을 확인할 수 있다. 실시예에 따른 상변화 소자는 1 내지 2[V]사이의 특정 임계 전압값을 기준으로 하여 저항값이 1000배 이하로 급격하게 감소하는 경향을 보이고 있다.
- [0051] 또한, 도4를 참조하면, 실시예에 따른 상변화 소자에 셋(set) 펄스 폭(pulse width(ns))을 일정하게 유지하면서 펄스 인가 횟수를 누적하여 증가시키면, 일정 누적 인가 횟수에서 저항의 값이 급격하게 감소하는 구간을 발견할 수 있다. 즉, 상변화 소자는 일정 횟수의 펄스가 인가되더라도 일정한 저항값을 유지하다가 임계 펄스 인가 횟수를 초과하는 경우 저항값이 급격하게 감소하여 유지되는 S자 곡선 형태를 취하고 있다. 이러한 S자 곡선 형태는 일정값을 기준으로 0 또는 1의 값을 출력하는 시그모이드 함수와 유사한 곡선 형태를 가짐을 확인할 수 있다. 실시예에 따른 상변화 소자는 약 5 내지 10사이의 누적 펄스 인가 횟수를 기준으로 하여 저항값이 100배 이

하로 급격하게 감소하는 경향을 보이고 있다.

- [0052] 다시 도1을 참조하면, 제2노드(122-1, . . . , 122-n)는 상변화 소자에 흐르는 전류치를 합산할 수 있다. 제1노드(121-1, . . . , 121-n)의 메모리 셀에 흐르는 전류는 상변화 소자의 전압-저항 특성에 의존하여 옴의 법칙에 따라 함께 변화하게 된다. 즉, 제1노드(121-1, . . . , 121-n)의 메모리 셀에 흐르는 전류는 상변화 소자의 전압값에 비례하며, 저항값에는 반비례하는 값을 가질 수 있다. 제2노드(122-1, . . . , 122-n)는 키르히호프의 전류 법칙에 따라 연결되어 있는 제1노드(121-1, . . . , 121-n)의 메모리 셀에 흐르는 전류의 값을 합산하여 출력층(123)에 전달할 수 있다.
- [0053] 상변화 물질은 Ge-Sb-Te계 물질로 이루어질 수 있다. 도5는 Ge-Sb-Te계 상변화 물질의 전압-저항 특성을 도시한 그래프이다. 도5를 참조하면, 셋 펄스와 리셋 펄스의 전압을 각각 3[V], 5[V]로 유지한 상태에서 펄스 폭을 조절함으로써 상변화 물질에 상변화를 일으킬 수 있다. 셋 펄스의 경우 약 400ns의 펄스 폭을 기준으로 상변화를 일으켜 높은 저항 상태에서 낮은 저항 상태로 이동하였음을 확인할 수 있다. 리셋 펄스의 경우 약170ns의 펄스 폭을 기준으로 상변화를 일으켜 낮은 저항 상태에서 높은 저항 상태로 이동하였음을 확인할 수 있다. Ge-Sb-Te계 상변화 물질은 전술한 임계 펄스 폭을 기준으로 상변화를 일으켜 저항 상태가 급격하게 변화하였음을 확인할 수 있다. 또한, 급격한 저항의 변화에 따른 전압-저항 특성 그래프는 시그모이드 함수의 형태를 취함을 확인할 수 있다.
- [0054] 또는, 상변화 물질은 Ag-In-Sb-Te계 물질로 이루어질 수 있다. Ag-In-Sb-Te계 상변화 물질은 낮은 리셋 전류로 동작 가능하다는 점과 다른 상변화 물질의 단점으로 제기된 낮은 결정화 온도와 높은 녹는점을 보완하는 물질로 결정화 속도가 빠르다는 장점이 있다. 도6은 Ag-In-Sb-Te계 상변화 물질의 전압-저항 특성을 도시한 그래프이다. 도6을 참조하면, Ag-In-Sb-Te계 상변화 물질의 저항은 50ns의 셋 펄스폭을 가지는 전압을 인가하였을 때 전압의 크기를 증가시켜도 상변화가 일어나지 않았다. 그러나, 100ns이상의 펄스폭을 가지는 전압을 인가하였을 때, 임계 전압을 초과하는 구간에서 상변화가 일어나 높은 저항 상태에서 낮은 저항 상태로 이동하였음을 확인할 수 있다. 100n, 400nss의 셋 펄스 폭을 인가하였을 때, Ag-In-Sb-Te계 상변화 물질은 약 1.1[V]에서 상변화가 일어나 높은 저항 상태에서 낮은 저항 상태로 이동하였음을 확인할 수 있다. 단, 100ns의 셋 펄스 폭을 인가한 경우보다, 400ns의 셋 펄스 폭을 인가한 경우에 빠른 시간안에 상 변화가 완료되었음을 확인할 수 있다.
- [0055] 또는, 상변화 물질은 Ti-Sb-Te 계 물질로 이루어질 수 있다.
- [0056] 따라서, 셋 펄스 폭을 길게 인가하여 스텝 함수에 가까운 시그모이드 함수를 구현할 수 있으며, 상대적으로 셋 펄스 폭을 짧게 인가하여 완만한 S자 곡선의 시그모이드 함수를 구현할 수 있다. 이와 같이 Ag-In-Sb-Te계 상변화 물질에 인가되는 셋 펄스 폭 또는 리셋 펄스 폭을 조절함으로써 원하는 형태의 시그모이드 함수를 구현할 수 있다.
- [0057] 또한, 상변화 물질은 전기적 물성에 따라 다양한 시그모이드 함수를 구현할 수 있다. 따라서, 열 전도도(hermal conductivity), 저항, 온-오프 비율(on-off ratio) 등의 전기적 물성을 고려하여 상변화 물질을 선택함으로써 다양한 형태를 가지는 시그모이드 함수를 구현할 수 있으며, 이를 인공신경망 시스템의 활성화 함수로 적용할 수 있다.
- [0058] 실시예에 따른 상변화 물질은 인가한 펄스의 전압 크기 또는 펄스 횟수에 따라) 저항이 시그모이드 함수의 형태로 변화하게 된다. 따라서, 입력층에서 전달된 벡터값을 활성화 함수에 입력하여 연산하는 과정을 상변화 소자의 특성을 이용하여 대체할 수 있다. 아 후, 여러 메모리 셀의 저항을 한번에 읽게되면, 이 때 흐르는 전류는 각 셀의 저항으로 결정된 전류의 합일 수 있다. 이는 인공신경망 시스템의 피드 포워드(feed forward) 과정에서 합산 과정과 동일하다. 이와 같은 방법으로 활성화 함수 연산 및 합산 연산을 추가 과정 없이 상변화 소자에 전압을 인가하여 상을 바꾸고 셀들에 흐르는 전류를 읽는 방법으로 대체함으로써 연산 코스트를 크게 감소시킬 수 있다.
- [0059] 또한, 실시예에 따른 상변화 소자의 전압-저항 특성은 ReLu(Rectified linear unit)함수 형태에 따라 변화할 수 있다. RELU함수는 0보다 작은 값이 입력되는 경우 0을 반환하고, 0보다 큰 값이 입력되는 경우 그 값을 그대로 반환할 수 있다.
- [0060] 또한, 실시예에 따른 상변화 소자의 전압-저항 특성은 복수개의 상변화 물질이 조합된 함수 형태에 따라 변화할 수 있다. 예를 들면, 상변화 소자의 전압-저항 특성은 시그모이드 함수와 RELU함수가 조합된 함수 형태로 변화할 수 있다. 또한, 실시 Tanh, Leaky ReLU 등 다양한 비선형 함수가 조합된 형태로 변화할 수 있다.

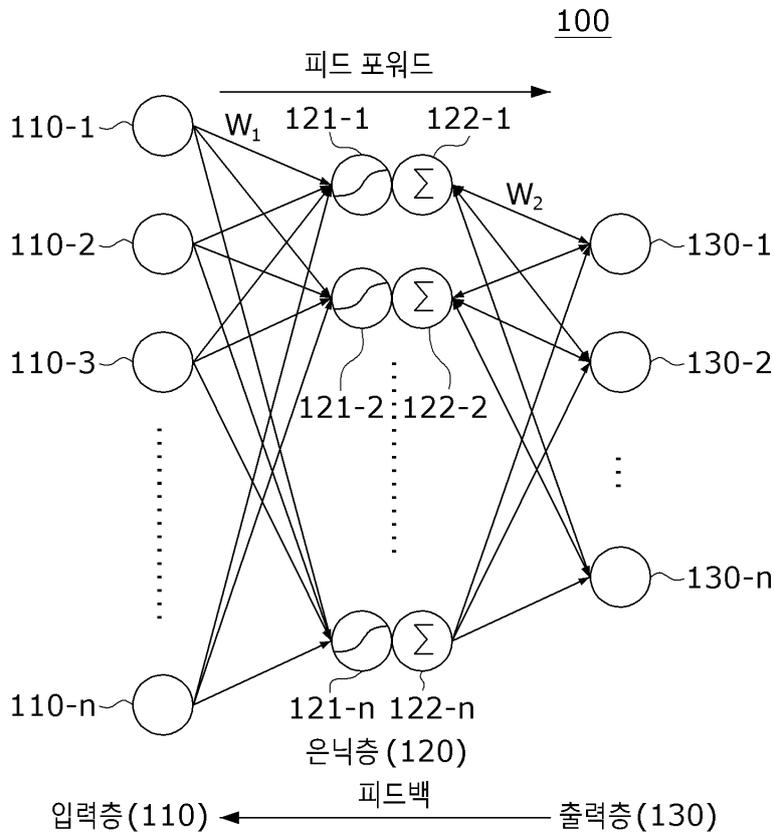
- [0061] 출력층(130)은 제2노드에서 합산한 값에 가중치를 부여하여 출력 데이터를 출력할 수 있다. 출력층(130)은 입력 데이터에 대응하는 출력 데이터를 출력할 수 있다. 출력층(130)은 은닉층(120)으로부터 수신한 신호에 기초하여 출력을 생성할 수 있다. 출력층(130)을 구성하는 복수의 출력 노드들은 각각 출력값들을 출력할 수 있다. 즉, 하나의 출력 노드에서 하나의 출력값이 출력될 수 있다. 하나의 노드와 다음 노드 사이에서 연결 가중치가 적용될 수 있다. 연결 가중치는 노드와 노드 사이에서 개별적으로 결정될 수 있다. 연결 가중치는 학습과 오류역전과 과정을 통해 업데이트될 수 있다.
- [0062] 상변화 소자를 이용한 인공신경망 시스템(100)은 감독 학습(supervised learning)을 통해 신경망을 학습시킬 수 있다. 감독 학습이란 학습 데이터와 그에 대응하는 출력 데이터를 함께 신경망에 입력하고, 학습 데이터에 대응하는 출력 데이터가 출력되도록 연결선들의 연결 가중치를 업데이트하는 방법이다. 예를 들면, 상변화 소자를 이용한 인공신경망 시스템(100)은 델타 규칙(delta rule)과 오류역전과 학습(backpropagation learning) 등을 통해 인공 뉴런들 사이의 연결 가중치를 업데이트할 수 있다.
- [0063] 오류역전과 학습은, 주어진 학습 데이터에 대해 전방 계산(forward computation)으로 오류를 추정된 후, 출력층에서 시작하여 은닉층과 입력층 방향으로 역으로 전진하여 추정한 오류를 전파하고, 오류를 줄이는 방향으로 연결 가중치를 업데이트하는 방법이다. 예를 들면, 오류역전과 학습은 입력을 각 층을 전달하여 출력을 계산하는 단계 및 출력의 오차를 줄이기 위해 신경망 내의 연결 가중치를 수정하는 단계로 구분될 수 있다. 상변화 소자를 이용한 인공신경망 시스템(100)은 현재 설정된 연결 가중치들이 얼마나 최적에 가까운지를 측정하기 위한 목적 함수(objective function)를 정의하고, 목적 함수의 결과에 기초하여 연결 가중치들을 계속 변경하고, 학습을 반복적으로 수행할 수 있다. 예를 들면, 목적 함수는 신경망이 학습 데이터에 기초하여 실제 출력한 출력 값과 출력되기로 원하는 기대값 간의 오류를 계산하기 위한 오류 함수일 수 있다. 상변화 소자를 이용한 인공신경망 시스템(100)은 오류 함수의 값을 줄이는 방향으로 연결 가중치들을 업데이트할 수 있다.
- [0064] 예를 들면, 연결 가중치를 업데이트하는 방법으로 경사 하강법(gradient descent)이 있다. 경사 하강법은 각 노드의 연결 가중치의 변화량에 따른 오류의 변화량(기울기)을 줄이는 방향으로 연결 가중치를 업데이트하는 것을 말한다. 인공신경망은 복수의 층으로 이루어져 있기 때문에 각 노드의 연결 가중치는 신호에 곱해지는 여러 연결 가중치 중에 하나를 의미할 수 있다.
- [0066] 상기에서는 본 발명의 바람직한 실시예를 참조하여 설명하였지만, 해당 기술 분야의 숙련된 당업자는 하기의 특허 청구의 범위에 기재된 본 발명의 사상 및 영역으로부터 벗어나지 않는 범위 내에서 본 발명을 다양하게 수정 및 변경시킬 수 있음을 이해할 수 있을 것이다.

부호의 설명

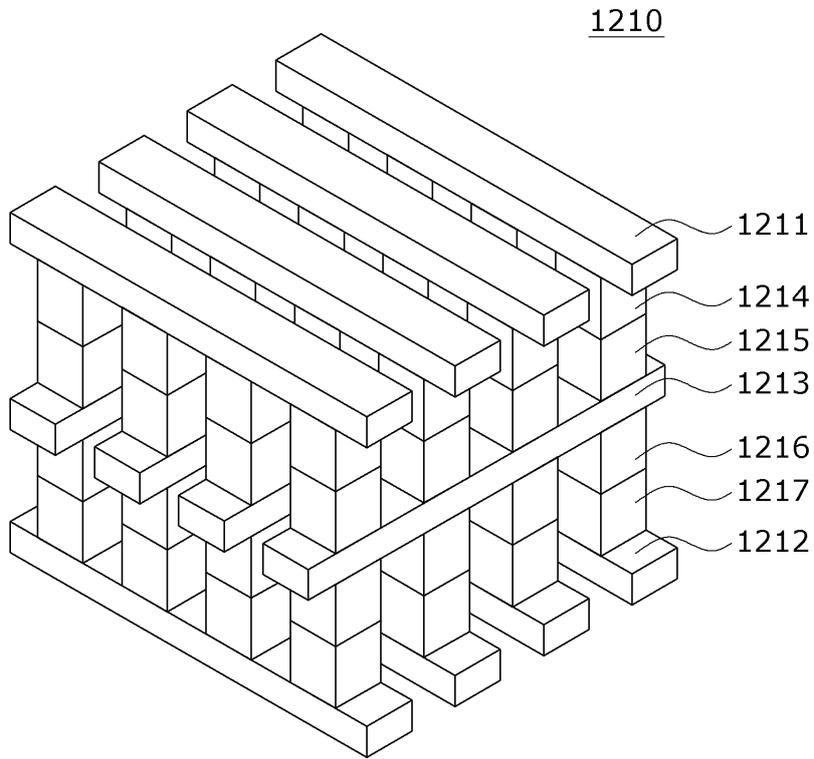
- [0068] 100: 인공신경망 시스템
- 110: 입력층
- 120: 은닉층
- 130: 출력층
- 1210: 상변화 소자
- 121-1, . . . , 121-n: 제1노드
- 122-1, . . . , 122-n: 제2노드
- 1211, 1212: 컬럼 메탈라인
- 1213: 워드 메탈라인
- 1214, 1216 메모리 셀
- 1215, 1217: 셀렉터

도면

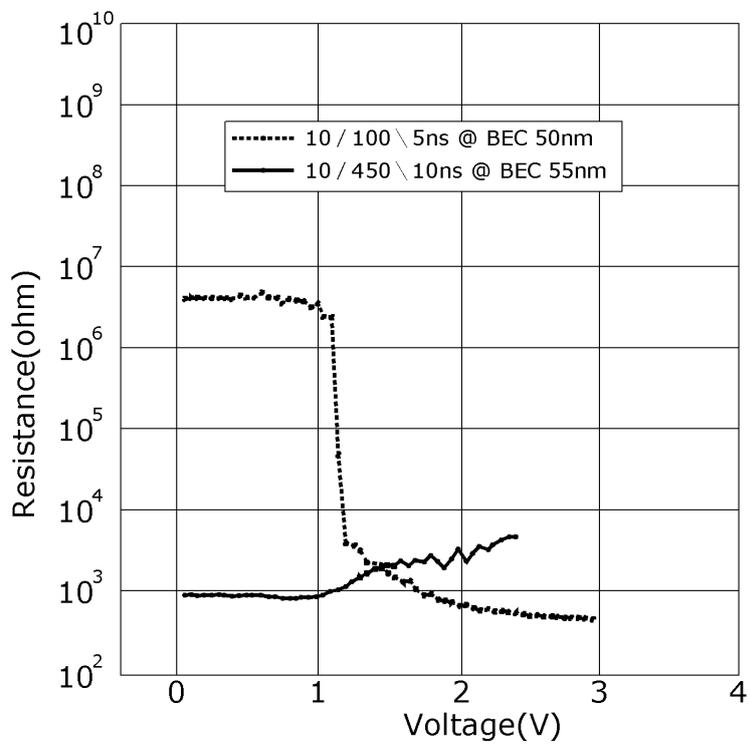
도면1



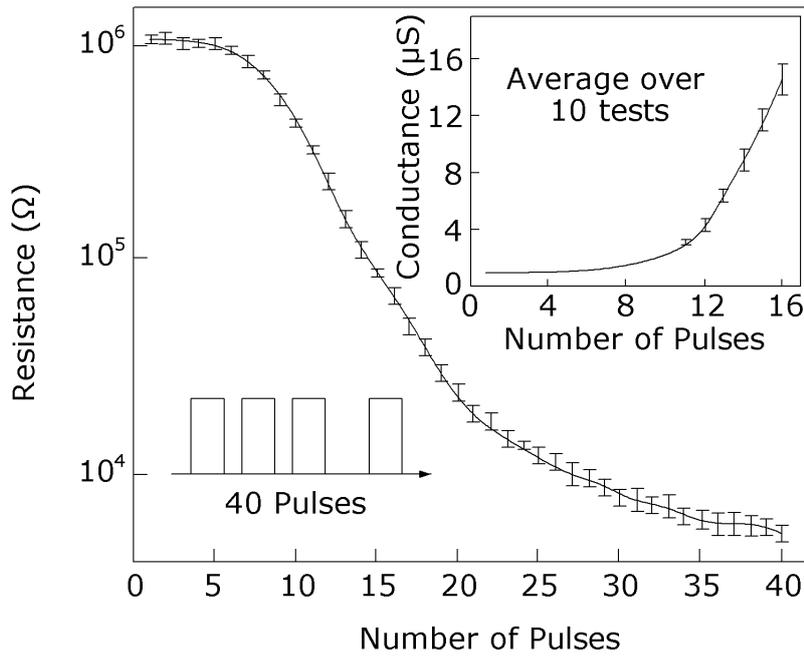
도면2



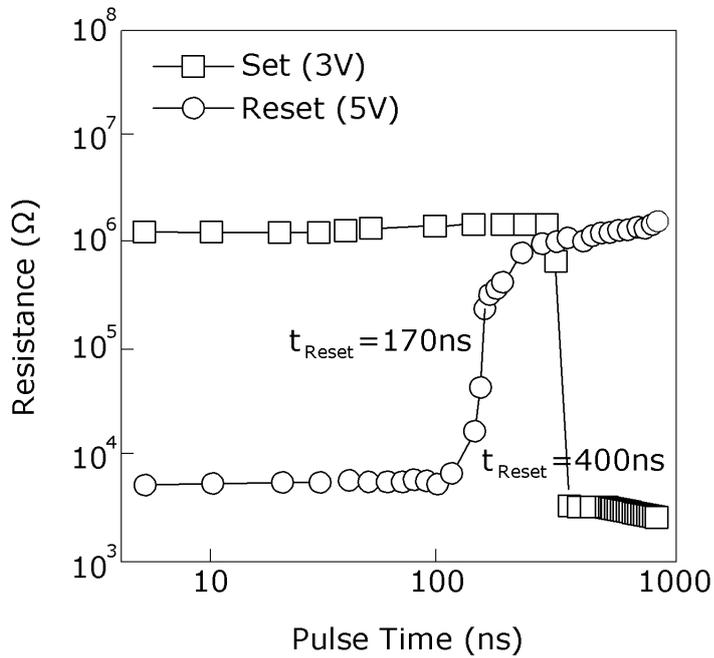
도면3



도면4



도면5



도면6

