(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2022년12월16일
(11) 등록번호 10-2478655
(24) 등록일자 2022년12월13일

(51) 국제특허분류(Int. Cl.)
G06N 3/063 (2006.01) *G06F 7/523* (2006.01)
G11C 8/08 (2006.01)
(52) CPC특허분류
G06N 3/063 (2013.01)
G06F 7/523 (2013.01)
(21) 출원번호 10-2020-0185732
(22) 출원일자 2020년12월29일
심사청구일자 2020년12월29일
(65) 공개번호 10-2022-0094485
(43) 공개일자 2022년07월06일
(56) 선행기술조사문헌
KR1020200008521 A*
US20200311523 A1*
*는 심사관에 의하여 인용된 문헌

(73) 특허권자
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
정성욱
서울특별시 서대문구 연세로 50, 제3공학관 C513 (신촌동)
이영규
서울특별시 서대문구 연세로 50, 공학원 246C (신촌동)
김기룡
서울특별시 서대문구 연세로 50, 공학원 246C (신촌동)
(74) 대리인
특허법인(유한)아이시스

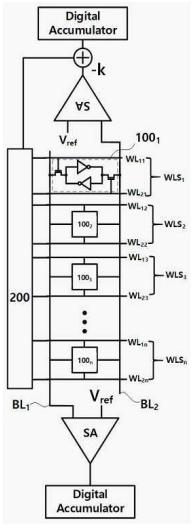
전체 청구항 수 : 총 11 항 심사관 : 노지명

(54) 발명의 명칭 **뉴럴 네트워크 연산 장치**

(57) 요 약

본 실시예에 의한 뉴럴 네트워크 연산 장치는: 가중치(weight) 비트와 반전 가중치 비트를 저장하는 메모리 소자와, 메모리 소자에 제1 입력을 제공하는 제1 워드 라인(word line)과 제2 입력을 제공하는 제2 워드 라인을 포함하는 워드 라인 세트와, 제1 입력과 가중치 비트와의 곱셈 연산 결과가 형성되는 제1 비트라인과, 제2 입력과 반전 가중치 비트가 곱셈 연산된 결과를 출력하는 제2 비트 라인을 포함한다.

대 표 도 - 도1



(52) CPC특허분류
G11C 8/08 (2013.01)

이 발명을 지원한 국가연구개발사업
과제고유번호 1711121368
과제번호 2020M3F3A2A01081918
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 원천기술개발사업
연구과제명 차세대 다치레벨 로직-메모리 융합소자를 이용한 고신뢰성 저전력 저면적 컴퓨팅-인-메모리 회로 및 아키텍처 개발
기 여 율 1/1
과제수행기관명 연세대학교
연구기간 2020.07.01 ~ 2020.12.31

명 세 서

청구범위

청구항 1

뉴럴 네트워크 연산 장치로, 상기 연산 장치는:

가중치(weight) 비트와 반전 가중치 비트를 저장하는 메모리 소자와,

상기 메모리 소자에 제1 입력을 제공하는 제1 워드 라인(word line)과 제2 입력을 제공하는 제2 워드 라인을 포함하는 워드 라인 세트와,

상기 제1 입력과 상기 가중치 비트와의 곱셈 연산 결과에 상응하는 전압이 형성되는 제1 비트라인과, 상기 제2 입력과 상기 반전 가중치 비트가 곱셈 연산된 결과에 상응하는 전압이 형성되는 제2 비트 라인을 포함하고

상기 연산 장치는,

상기 제1 비트 라인 및 상기 제2 비트 라인과 연결되며, 각각 가중치 비트 및 반전 가중치 비트를 저장하는 복수의 메모리 소자들과,

상기 복수의 메모리 소자에 각각 제1 입력과 제2 입력을 제공하는 복수의 워드 라인 세트들을 더 포함하며,

상기 제1 비트 라인에는 상기 메모리 소자 및 복수의 메모리 소자에 각각 제공된 제1 입력과 상기 메모리 소자 및 복수의 메모리 소자가 각각 저장하는 상기 가중치 비트와의 연산 결과들에 상응하는 전압이 누적되어 형성되고, 상기 제2 비트 라인에는 상기 메모리 소자 및 복수의 메모리 소자에 각각 제공된 제2 입력과 상기 메모리 소자 및 복수의 메모리 소자가 각각 저장하는 상기 반전 가중치 비트와의 연산 결과들에 상응하는 전압이 누적되어 형성되는 연산 장치.

청구항 2

제1항에 있어서,

상기 메모리 소자는

제1 노드에 입력 노드가 연결되고, 제2 노드에 출력 노드가 연결된 제1 인버터;

상기 제1 노드에 출력 노드가 연결되고, 상기 제2 노드에 입력 노드가 연결된 제2 인버터;

상기 제1 비트 라인과 연결된 제1 전극과 상기 제1 워드 라인과 연결된 제어 전극 및 상기 제1 노드에 연결된 제2 전극을 가지는 제1 스위치 및

상기 제2 비트 라인과 연결된 제1 전극과 상기 제2 워드 라인과 연결된 제어 전극 및 상기 제2 노드에 연결된 제2 전극을 가지는 제2 스위치를 포함하는 연산 장치.

청구항 3

삭제

청구항 4

제1항에 있어서,

상기 연산 장치는,

상기 제1 비트 라인에 누적되어 형성된 연산 결과에 상응하는 전압을 제공받고, 기준 전압과 비교하는 감지 증폭기를 더 포함하는 연산 장치.

청구항 5

제4항에 있어서,

상기 제1 비트 라인에 누적되어 형성된 연산 결과에 상응하는 전압을 디지털로 변환한 값은 상기 가중치와 상기 제1 입력의 곱의 누적값에 상응하는 연산 장치.

청구항 6

제4항에 있어서,

상기 제1 비트 라인에 누적되어 형성된 연산 결과에 상응하는 전압을 디지털로 변환한 값은 수학식 $\sum_{i=1}^{n} X_i W_i$ 로 표시되는 연산 장치.(W: 가중치, X: 제1 입력)

청구항 7

제4항에 있어서,

상기 제1 입력이 제공됨에 따라 상기 제1 비트 라인에 누적되어 형성된 연산 결과에 상응하는 전압을 디지털로 변환한 값을 순차적으로 누적하는 누산기(accumulator)를 더 포함하는 연산 장치.

청구항 8

제2항에 있어서,

상기 연산 장치는,

상기 제2 입력에 포함된 논리 하이 상태인 비트들의 개수를 계수(count)하는 카운터 및

상기 카운터의 계수 결과에서 상기 제2 비트 라인에 누적되어 형성된 연산 결과에 상응하는 전압을 디지털로 변환한 값을 감산하는 감산기를 더 포함하는 연산 장치.

청구항 9

제1항에 있어서,

상기 제2 비트라인에 연산되어 누적된 결과에 상응하는 전압을 디지털로 변환한 값은 반전된 가중치와 입력의 곱이 누적된 값에 상응하는 연산 장치.

청구항 10

제1항에 있어서,

상기 제2 비트라인에 연산되어 누적된 결과에 상응하는 전압을 디지털로 변환한 값은 수학식 $\sum_{i=1}^{n}\{1-W_i\}Y_i=\sum_{i=1}^{n}Y_i-\sum_{i=1}^{n}W_iY_i$ 로 표시되는 연산 장치.

(W: 가중치, Y: 제2 입력)

청구항 11

제1항에 있어서,

상기 제2 입력이 제공됨에 따라 상기 제2 비트 라인에 누적되어 형성된 연산 결과에 상응하는 전압을 디지털로 변환한 값을 순차적으로 누적하는 누산기(accumulator)를 더 포함하는 연산 장치.

청구항 12

제1항에 있어서,

상기 제1 입력은 순서를 가지는 디지털 비트에 상응하며,

상기 제2 입력은 상기 제1 입력에 대한 스트라이드(stride)된 입력으로,

상기 스트라이드된 입력은 상기 순서를 가지는 제1 디지털 비트의 순서가 시프트(shift)된 것인 연산 장치.

청구항 13

삭제

청구항 14

삭제

청구항 15

삭제

청구항 16

삭제

청구항 17

삭제

청구항 18

삭제

청구항 19

삭제

발명의 설명

기 술 분 야

[0001] 본 기술은 뉴럴 네트워크 연산 장치와 관련된다.

배 경 기 술

[0002] 딥러닝에서 많은 양의 데이터 처리가 필요해짐에 따라 데이터 이동 과정에서 필요한 전력의 부담이 증가하고, 데이터의 병목 현상(bottleneck)이 발생하여 지연(delay)의 부담(overhead)이 더욱 증가하는 추세이다. 이러한 전력과 지연을 줄이기 위해 메모리 내에서 연산을 수행하는 CIM(computing in memory) 기술이 등장하였다.

발명의 내용

해결하려는 과제

[0003] 기존의 CIM 구조는 멀티 비트(multi-bit) 센싱 마진의 문제를 해결하기 위하여 입력 비트를 직렬로 제공하여 가중치와 곱셈 연산하고 누적하여 MAC(multiply accumulate) 연산을 수행하였다. 그러나 이러한 비트 시리얼(bit serial) 방식은 처리량(throughput)이 낮다. 이를 해소하기 위하여 입력의 비트 수를 증가시키면 센싱 마진의 문제가 발생하여 검출 성능의 특성이 열화된다.

[0004] 본 기술은 상기한 종래 기술의 문제점을 개선하기 위한 것이다. 본 기술로 해결하기 위한 과제 중 하나는 종래 기술의 낮은 처리량 성능을 향상시킬 수 있는 기술을 제공하기 위한 것이다. 또한, 본 기술로 해결하기 위한 과제 중 하나는 종래 기술의 센싱 마진의 난점을 해결하기 위한 기술을 제공하는 것이다.

과제의 해결 수단

[0005] 본 실시예에 의한 뉴럴 네트워크 연산 장치는: 가중치(weight) 비트와 반전 가중치 비트를 저장하는 메모리 소자와, 메모리 소자에 제1 입력을 제공하는 제1 워드 라인(word line)과 제2 입력을 제공하는 제2 워드 라인을 포함하는 워드 라인 세트와, 제1 입력과 가중치 비트와의 곱셈 연산 결과가 형성되는 제1 비트라인과, 제2 입력

과 반전 가중치 비트가 곱셈 연산된 결과를 출력하는 제2 비트 라인을 포함한다.

[0006] 본 실시예의 일 태양에 의하면, 메모리 소자는 제1 노드에 입력 노드가 연결되고, 제2 노드에 출력 노드가 연결된 제1 인버터; 제1 노드에 출력 노드가 연결되고, 제2 노드에 입력 노드가 연결된 제2 인버터; 제1 비트 라인과 연결된 제1 전극과 제1 워드 라인과 연결된 제어 전극을 가지는 제1 스위치 및 제2 비트 라인과 연결된 제1 전극과 제2 워드 라인과 연결된 제어 전극을 가지는 제2 스위치를 포함한다.

[0007] 본 실시예의 일 태양에 의하면, 연산 장치는, 제1 비트 라인 및 제2 비트 라인과 연결되며, 각각 가중치 비트 및 반전 가중치 비트를 저장하는 복수의 메모리 소자들과, 복수의 메모리 소자에 각각 제1 입력과 제2 입력을 제공하는 복수의 워드 라인 세트들을 더 포함하며, 복수의 메모리 소자에 각각 제공된 제1 입력과 가중치 비트와의 연산 결과들은 제1 비트 라인에 누적되어 형성되고, 복수의 메모리 소자에 각각 제공된 제2 입력과 반전 가중치 비트와의 연산 결과들은 제2 비트 라인에 누적되어 형성된다.

[0008] 본 실시예의 일 태양에 의하면, 연산 장치는, 제1 비트 라인에 누적되어 형성된 연산 결과를 제공받고, 기준 전압과 비교하는 감지 증폭기를 더 포함한다.

[0009] 본 실시예의 일 태양에 의하면, 제1 비트 라인에 누적되어 형성된 연산 결과는 가중치와 제1 입력의 곱의 누적값에 상응한다.

[0010] 본 실시예의 일 태양에 의하면, 제1 비트 라인에 누적되어 형성된 연산 결과는 수학식 $\sum_{i=1}^{n} X_i W_i$ 로 표시된다.(W: 가중치, X: 제1 입력)

[0011] 본 실시예의 일 태양에 의하면, 제1 입력이 순차적으로 제공됨에 따라 제1 비트 라인에 누적되어 형성된 연산 결과를 순차적으로 누적하는 누산기(accumulator)를 더 포함한다.

[0012] 본 실시예의 일 태양에 의하면, 연산 장치는, 제2 입력에 포함된 논리 하이 상태인 비트들의 개수를 계수(count)하는 카운터 및 카운터의 계수 결과에서 제2 비트 라인에 누적되어 형성된 연산 결과를 감산하는 감산기를 더 포함한다.

[0013] 본 실시예의 일 태양에 의하면, 제2 비트라인에 연산되어 누적된 결과는 반전된 가중치와 입력의 곱이 누적된 값에 상응한다.

[0014] 본 실시예의 일 태양에 의하면, 제2 비트라인에 연산되어 누적된 결과는 수학식 $\sum_{i=1}^{n}\{1-W_i\}X_i = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} W_i X_i$ 로 표시된다. (W: 가중치, Y: 제2 입력)

[0015] 본 실시예의 일 태양에 의하면, 제2 입력이 순차적으로 제공됨에 따라 제2 비트 라인에 누적되어 형성된 연산 결과를 순차적으로 누적하는 누산기(accumulator)를 더 포함한다.

[0016] 본 실시예의 일 태양에 의하면, 제2 입력은 제1 입력에 대한 스트라이드(stride)된 입력이다.

[0017] 본 실시예에 의한 뉴럴 네트워크 출력값 검출 방법은: 카운터에 입력 비트들을 제공하는 단계와, 카운터가 입력에 포함된 턴 온 비트(turn-on bit)의 개수를 계수하는 단계와, 계수 결과 턴 온 비트의 개수가 MAC 연산 경우의 수의 k 분할 미만일 때, 센싱 마진(sensing margin)이 증가하도록 비트 라인 전압 형성 시간을 증가시킨다.(k: 2 이상 자연수)

[0018] 본 실시예의 일 태양에 의하면, 턴 온 비트의 개수가 입력 비트 수의 2 분할 미만일 때, 센싱 마진이 증가하도록 비트 라인 전압 형성시간을 적어도 2배 이상 증가시킨다,

[0019] 본 실시예의 일 태양에 의하면, 비트 라인 전압 형성 시간을 증가시키는 단계는, 억세스 트랜지스터의 도통 시간을 증가시켜 수행한다.

[0020] 본 실시예에 의한 뉴럴 네트워크 출력값 검출 방법은: 카운터에 입력 비트들을 제공하는 단계와, 카운터가 입력에 포함된 턴 온 비트(turn-on bit)의 개수를 계수하는 단계와, 계수 결과 턴 온 비트의 개수가 MAC 연산 경우의 수의 k 분할 이상일 때, 카운터가 입력 비트가 반전된 반전 입력 비트를 뉴럴 네트워크에 제공하는 단계와,

센싱 마진(sensing margin)이 증가하도록 비트 라인 전압 형성 시간을 증가시킨다.(k: 2 이상 자연수)

[0021] 본 실시예의 일 태양에 의하면, 검출 방법은 카운터에 입력 비트들을 제공하는 단계 또는 카운터가 입력에 포함된 턴 온 비트(turn-on bit)의 개수를 계수하는 단계 이전에 수행되는 뉴럴 네트워크의 가중치 값의 합을 연산하는 단계를 더 포함한다.

[0022] 본 실시예의 일 태양에 의하면, 검출 방법은 뉴럴 네트워크 출력값을 연산하는 단계를 더 포함하며, 뉴럴 네트워크 출력값을 연산하는 단계는, 반전 입력 비트가 뉴럴 네트워크에 제공되어 형성된 출력 값을 구하는 단계와, 가중치 값의 합에서 출력 값의 차이를 연산하는 단계를 수행하여 이루어진다.

[0023] 본 실시예의 일 태양에 의하면, 비트 라인 전압 형성 시간을 증가시키는 단계는, 억세스 트랜지스터의 도통 시간을 증가시켜 수행한다.

발명의 효과

[0024] 본 실시예에 의하면 가중치와 제1 입력에 대한 MAC 연산과 가중치와 제2 입력에 대한 MAC 연산을 함께 수행하여 종래 기술에 비하여 높은 처리량을 얻을 수 있다는 장점이 제공된다. 또한, 종래 기술에 비하여 향상된 센싱 마진을 얻을 수 있다는 장점이 제공된다.

도면의 간단한 설명

[0025] 도 1은 본 실시예에 의한 뉴럴 네트워크 연산 장치(1)의 개요를 도시한 도면이다.

도 2는 어느 한 메모리 소자(100)의 개요를 도시한 트랜지스터 레벨 회로도이다.

도 3(a)는 제1 인버터(I1)가 논리 로우를 출력하고, 제1 워드 라인(WL1)을 통하여 논리 하이 상태의 제1 입력(X)이 제공된 상태를 예시한 도면이다. 도 3(b)는 제1 인버터(I1)가 논리 로우를 출력하고, 제1 워드 라인(WL1)을 통하여 논리 로우 상태의 제1 입력(X)이 제공된 상태를 예시한 도면이다. 도 3(c)는 제1 인버터(I1)가 논리 하이를 출력하고, 제1 워드 라인(WL1)을 통하여 논리 하이 상태의 제1 입력(X)이 제공된 상태를 예시한 도면이다. 도 3(d)는 제1 인버터(I1)가 논리 하이를 출력하고, 제1 워드 라인(WL1)을 통하여 논리 로우 상태의 제1 입력(X)이 제공된 상태를 예시한 도면이다.

도 4(a)는 제2 인버터(I2)가 논리 하이를 출력하고, 제2 워드 라인(WL2)을 통하여 논리 하이 상태의 제2 입력(Y)이 제공된 상태를 예시한 도면이다. 도 4(b)는 제2 인버터(I2)가 논리 하이를 출력하고, 제2 워드 라인(WL2)을 통하여 논리 로우 상태의 제2 입력(Y)이 제공된 상태를 예시한 도면이다. 도 4(c)는 제2 인버터(I2)가 논리 로우를 출력하고, 제2 워드 라인(WL2)을 통하여 논리 하이 상태의 제2 입력(Y)이 제공된 상태를 예시한 도면이다. 도 4(d)는 제2 인버터(I2)가 논리 로우를 출력하고, 제2 워드 라인(WL2)을 통하여 논리 로우 상태의 제2 입력(Y)이 제공된 상태를 예시한 도면이다.

도 5(a) 및 도 5(b)는 본 실시예에 의한 뉴럴 네트워크 연산 장치의 연산 과정을 예시하기 위한 도면이다.

도 6은 본 실시예에 의한 연산 장치(1)의 동작을 설명하기 위한 개요도이다.

도 7(a) 및 도 7(b)는 본 실시예에 의한 검출 방법의 개요를 도시한 순서도이다.

도 8(a)는 카운터의 계수 결과가 4 미만인 경우에, 비트 라인에서 발생할 수 있는 전압의 분포를 나타낸 도면이다.도 8(b)는 센싱 마진이 증가한 상태를 도시한 도면이다.

발명을 실시하기 위한 구체적인 내용

[0026] 이하에서는 첨부된 도면들을 참조하여 본 실시예를 설명한다. 도 1은 본 실시예에 의한 뉴럴 네트워크 연산 장치(1)의 개요를 도시한 도면이다. 도 1을 참조하면, 본 실시예에 의한 뉴럴 네트워크 연산 장치(1)는 가중치(weight) 비트와 반전 가중치 비트를 저장하는 메모리 소자(100_1, 100_2, 100_3, ..., 100_n)와, 메모리 소자(100_1, 100_2, 100_3, ..., 100_n)에 제1 입력을 제공하는 제1 워드 라인(word line, WL_{11}, WL_{12}, WL_{13}, ..., WL_{1n})과 제2 입력을 제공하는 제2 워드 라인(WL_{21}, WL_{22}, WL_{23}, ..., WL_{2n})을 포함하는 워드 라인 세트(WLS_1, WLS_2, WLS_3, ..., WLS_n)와, 제1 입력과 가중치 비트와의 곱셈 연산 결과가 형성되는 제1 비트라인(BL_1)과, 제2 입력과 반전 가중치 비트가 곱셈 연산된 결과를 출력하는 제2 비트 라인(BL_2)을 포함한다.

[0027] 도 2는 어느 한 메모리 소자(100)의 개요를 도시한 트랜지스터 레벨 회로도이다. 도 1 및 도 2를 참조하면, 메모리 소자(100)는 제1 인버터(I1)와 제2 인버터(I2)와, 제1 비트 라인(BL1)과 연결된 제1 전극과 제1 워드 라인(WL1)과 연결된 제어 전극을 가지는 제1 스위치(TR1) 및 제2 비트 라인(BL2)과 연결된 제1 전극과 제2 워드 라인(WL2)과 연결된 제어 전극을 가지는 제2 스위치(TR2)를 포함하며, 제1 인버터(I1)의 입력은 제2 인버터(I2)의 출력에 연결되고, 제1 인버터(I1)의 출력은 제2 인버터(I2)의 입력에 연결된다.

[0028] 도 1 및 도 2로 예시된 실시예는 메모리 소자(100)로 두 개의 인버터를 이용하여 가중치 비트와 반전 가중치 비트를 저장하는 SRAM(static random access memory)을 예시한다. 제1 스위치(TR1)는 제1 워드 라인(WL1)을 통해 제공된 입력 X에 의하여 도통 및 차단이 제어된다. 제1 스위치(TR1)가 도통 및/또는 차단됨에 따라 제1 비트라인(BL1)의 전기적 상태가 변화할 수 있다. 제2 스위치(TR2)는 제2 워드 라인(WL2)을 통해 제공된 입력 Y에 의하여 도통 및 차단이 제어된다. 마찬가지로, 제2 스위치(TR2)가 도통 및/또는 차단됨에 따라 제2 비트라인(BL2)의 전기적 상태가 변화할 수 있다. 도시된 예에서 제1 스위치(TR1)와 제2 스위치(TR2)는 모두 NMOS 트랜지스터인 것을 예시하였다. 그러나, 도시되지 않은 실시예에서, 제1 스위치(TR1) 및 제2 스위치(TR2) 중 어느 하나 이상은 PMOS 트랜지스터로 구현될 수 있다. 도 2로 예시된 실시예와 같이 산업에서 널리 사용되는 여섯 개의 트랜지스터를 이용하는 SRAM을 이용함으로써 업계에서 표준화된 제조 공정을 사용할 수 있다.

[0029] 도 3(a)는 제1 인버터(I1)가 논리 로우를 출력하고, 제1 워드 라인(WL1)을 통하여 논리 하이 상태의 제1 입력(X)이 제공된 상태를 예시한 도면이다. 이하에서는, 제1 워드 라인(WL1)이 논리 하이일 때 입력(X)를 “1”이라고 하고, 제1 워드 라인(WL1)이 논리 로우일 때 입력(X)를 “0”이라고 한다. 또한, 제1 인버터(I1)의 출력이 논리 로우일 때(제2 인버터(I2)의 출력이 논리 하이일 때) 메모리(100)에 저장된 가중치 데이터를 “1”이라고 하고, 제1 인버터(I1)의 출력이 논리 하이일 때(제2 인버터(I2)의 출력이 논리 일 때) 메모리(100)에 저장된 가중치 데이터를 “0”이라고 한다. 도 2 및 도 3(a)를 참조하면, 제1 워드 라인(WL1)을 통하여 논리 하이 상태의 제1 입력(X)이 제공됨에 따라 제1 스위치(TR1)가 도통된다. 따라서, 프리차지(pre-charge)된 제1 비트 라인(BL1)은 도통된 제1 스위치(TR1)과 제1 인버터(I1)을 통하여 방전(discharge)된다. 제1 비트 라인(BL1)에서는 제공된 입력(X)과 메모리(100)에 저장된 가중치값의 곱인 “1”에 상응하는 전기적 상태가 형성된다.

[0030] 도 3(b)는 제1 인버터(I1)가 논리 로우를 출력하고, 제1 워드 라인(WL1)을 통하여 논리 로우 상태의 제1 입력(X)이 제공된 상태를 예시한 도면이다. 도 2 및 도 3(b)를 참조하면, 제1 워드 라인(WL1)을 통하여 논리 로우 상태의 제1 입력(X)이 제공됨에 따라 제1 스위치(TR1)는 차단된다. 따라서, 프리차지(pre-charge)된 제1 비트 라인(BL1)은 프리 차지된 상태를 유지하며, 제1 비트 라인(BL1)에서는 제공된 입력(X)과 메모리(100)에 저장된 가중치값의 곱인 “0”에 상응하는 전기적 상태가 형성된다.

[0031] 도 3(c)는 제1 인버터(I1)가 논리 하이를 출력하고, 제1 워드 라인(WL1)을 통하여 논리 하이 상태의 제1 입력(X)이 제공된 상태를 예시한 도면이다. 도 2 및 도 3(c)를 참조하면, 제1 워드 라인(WL1)을 통하여 논리 하이 상태의 제1 입력(X)이 제공되어 제1 스위치(TR1)가 도통됨에도 불구하고. 제1 인버터(I1)를 통한 방전 경로가 형성되지 않는다. 따라서, 프리차지(pre-charge)된 제1 비트 라인(BL1)은 프리 차지된 상태를 유지하며, 제1 비트 라인(BL1)에서는 제공된 입력(X)과 메모리(100)에 저장된 가중치값의 곱인 “1”에 상응하는 전기적 상태가 형성된다.

[0032] 도 3(d)는 제1 인버터(I1)가 논리 하이를 출력하고, 제1 워드 라인(WL1)을 통하여 논리 로우 상태의 제1 입력(X)이 제공된 상태를 예시한 도면이다. 도 2 및 도 3(d)를 참조하면, 제1 워드 라인(WL1)을 통하여 논리 로우 상태의 제1 입력(X)이 제공되므로 제1 스위치(TR1)는 차단된다. 따라서, 프리차지(pre-charge)된 제1 비트 라인(BL1)은 프리 차지된 상태를 유지하며 제1 비트 라인(BL1)에서는 제공된 입력(X)과 메모리(100)에 저장된 가중치값의 곱인 “1”에 상응하는 전기적 상태가 형성된다.

[0033] 즉, 입력(X)이 “1”이고, 메모리(100)에 저장된 가중치가 “1”일 때 프리차지된 제1 비트 라인(BL1)에서 방전(discharge)이 일어난다.

[0034] 도 4(a)는 제2 인버터(I2)가 논리 하이를 출력하고, 제2 워드 라인(WL2)을 통하여 논리 하이 상태의 제2 입력(Y)이 제공된 상태를 예시한 도면이다. 이하에서는, 제2 워드 라인(WL2)이 논리 하이일 때 입력(Y)를 “1”이라고 하고, 제2 워드 라인(WL1)이 논리 로우일 때 입력(Y)를 “0”이라고 한다. 또한, 제2 인버터(I2)의 출력이 논리 하이일 때(제1 인버터(I1)의 출력이 논리 로우일 때) 메모리(100)에 저장된 가중치 데이터를 “1”이라고 하고, 제2 인버터(I1)의 출력이 논리 로우일 때(제1 인버터(I1)의 출력이 논리 하이일 때) 메모리(100)에 저장된 가중치 데이터를 “0”이라고 한다. 도 2 및 도 4(a)를 참조하면, 제2 워드 라인(WL2)을 통하여 논리 하이 상태의 제2 입력(Y)이 제공됨에 따라 제2 스위치(TR2)가 도통된다. 그러나, 제2 인버터(I2)를 통하여 방전 경로

가 형성되지 않아 프리 차지된 제2 비트 라인은 방전(discharge)되지 않는다. 따라서, 제2 비트 라인(BL2)에서는 제공된 입력(Y)과 반전된 가중치 값의 곱인 “0”에 상응하는 전기적 상태가 형성된다.

[0035] 도 4(b)는 제2 인버터(I2)가 논리 하이를 출력하고, 제2 워드 라인(WL2)을 통하여 논리 로우 상태의 제2 입력(Y)이 제공된 상태를 예시한 도면이다. 도 2 및 도 4(b)를 참조하면, 제2 워드 라인(WL2)을 통하여 논리 로우 상태의 제2 입력(Y)이 제공됨에 따라 제2 스위치(TR2)는 차단된다. 따라서, 프리차지(pre-charge)된 제2 비트 라인(BL2)은 프리 차지된 상태를 유지하며, 제2 비트 라인(BL2)에서는 제공된 입력(Y)과 반전된 가중치 값의 곱인 “0”에 상응하는 전기적 상태가 형성된다.

[0036] 도 4(c)는 제2 인버터(I2)가 논리 로우를 출력하고, 제2 워드 라인(WL2)을 통하여 논리 하이 상태의 제2 입력(Y)이 제공된 상태를 예시한 도면이다. 도 2 및 도 4(c)를 참조하면, 제2 워드 라인(WL2)을 통하여 논리 하이 상태의 제2 입력(Y)이 제공되어 제2 스위치(TR2)가 도통되고, 제2 인버터(I2)를 통한 방전 경로가 형성된다. 따라서, 프리차지(pre-charge)된 제2 비트 라인(BL2)은 제2 인버터(I2)를 통해 방전된 상태가 형성되며, 제2 비트 라인(BL2)에서는 제공된 입력(Y)과 반전된 가중치값의 곱인 “1”에 상응하는 전기적 상태가 형성된다.

[0037] 도 4(d)는 제2 인버터(I2)가 논리 로우를 출력하고, 제2 워드 라인(WL2)을 통하여 논리 로우 상태의 제2 입력(Y)이 제공된 상태를 예시한 도면이다. 도 2 및 도 4(d)를 참조하면, 제2 워드 라인(WL2)을 통하여 논리 로우 상태의 제2 입력(Y)이 제공되므로 제2 스위치(TR)는 차단된다. 따라서, 프리차지(pre-charge)된 제2 비트 라인(BL2)은 프리 차지된 상태를 유지하며 제2 비트 라인(BL2)에서는 제공된 입력(Y)과 메모리(100)에 저장된 가중치값의 곱인 “1”에 상응하는 전기적 상태가 형성된다. 즉, 입력(X)이 “1”이고, 반전된 가중치가 “1”(제2 인버터가 논리 로우를 출력할 때) 프리차지된 제2 비트 라인(BL2)이 방전(discharge)된다.

[0038] 도 5(a)는 본 실시예에 의한 뉴럴 네트워크 연산 장치의 연산 과정을 예시하기 위한 도면이다. 도 5(a)로 예시된 연산 장치에서 메모리 소자들(100_1, 100_2, 100_3, ..., 100_n) 중 메모리(100_1, 100_2)만 가중치로 “1” 저장된 상태를 예시한다. 제1 내지 제n 워드라인 세트(WLS_1, WLS_2, WLS_3, ..., WLS_n)에 포함된 제1 워드 라인들(WL_{11}, WL_{12}, WL_{13}, ..., WL_{1n})을 통하여 입력(X1, X2, X3, ..., Xn)이 제공된다.

[0039] 제공된 입력(X1, X2, X3, ..., Xn)들에서 X1 및 X2 만 “1”이고, 나머지 입력들은 모두 “0”인 경우에, 제공된 입력과 가중치와의 곱셈 연산이 수행되고, 연산된 결과는 제1 비트라인(BL1)에 누적된다. 위에서 설명된 바와 같이 입력이 “1”이고, 가중치가 “1”일 때 프리 차지된 비트 라인(BL1)에서 방전이 일어난다.

[0040] 이와 같이 방전되어 형성된 비트 라인(BL1)의 전압은 각 입력(X)과 가중치와의 곱셈 연산 결과가 누적(MAC, multiply and accumulate)된 것에 상응한다. 비트 라인(BL1)은 감지 증폭기(SA, sense amplifier)에 제공되어 복수의 기준 전압(Vref)과 비교되어 입력(X)과 가중치와의 곱셈 연산 결과가 검출된다. 곱셈 연산 결과는 디지털화 되어 출력된다.

[0041] 디지털로 변환된 연산 결과는 아래의 수학식과 같이 표시될 수 있으며, 이로부터 입력 X와 가중치의 곱의 합을 용이하게 연산할 수 있다.

[0042] [수학식 1]

[0043] $$D(\Delta V_{bl}) = \sum WX$$

[0044] ($D(\Delta V_{bl})$: 비트라인이 방전되어 형성된 전압을 디지털로 변환한 값, W: 가중치, X: 입력)

[0045] 디지털화된 연산 결과는 디지털 누적기(accumulator)에 의하여 시프트 및 누적 연산될 수 있다. 일 예로, 입력이 LSB와 MSB를 포함하는 2 비트인 경우에, 디지털 누적기는 최초 입력 비트인 LSB에 대한 연산 결과를 시프트하고, 최후 입력 비트인 MSB에 대한 연산 결과를 도합하여 출력할 수 있다.

[0046] 도 5(b)로 예시된 연산 장치에서 메모리 소자들(100_1, 100_2, 100_3, ..., 100_n) 중 메모리(100_1, 100_2)에만 가중치로 “1” 저장된 상태를 예시한다. 제1 내지 제n 워드라인 세트(WLS_1, WLS_2, WLS_3, ..., WLS_n)에 포함된 제2 워드 라인들(WL_{21}, WL_{22}, WL_{23}, ..., $WL2_n$)을 통하여 입력(Y1, Y2, Y3, ..., Yn)이 제공된다.

[0047] 제공된 입력(Y1, Y2, Y3, ..., Yn)들에서 Y3 및 Yn 이 “1”이고, 나머지 입력들은 모두 “0”인 경우에, 제공된 입력과 반전 가중치와의 곱셈 연산이 수행되고, 연산된 결과는 제2 비트라인(BL2)에 누적된다. 위에서 설명된 바와 같이 입력이 “1”이고, 반전 가중치가 “1”일 때 프리 차지된 비트 라인(BL2)이 방전된다. 방전이 일

어난 비트 라인(BL2)의 전압은 각 입력(Y)과 반전 가중치와의 곱셈 연산 결과가 누적(MAC, multiply and accumulate)된 것에 상응한다.

[0048] 입력과 반전 가중치와의 곱셈 연산이 누적된 결과는 아래의 수학식 2와 같이 표시될 수 있다.

[0049] [수학식 2]

$$D(\Delta V_{bl}) = \sum(1-W)Y = \sum Y - \sum WY \quad \cdots\cdots ①$$

[0050]

$$\sum WY = \sum Y - D(\Delta V_{bl}) \quad \cdots\cdots ②$$

[0051] ($D(\Delta V_{bl})$: 비트라인이 방전되어 형성된 전압을 디지털로 변환한 값, W: 가중치, Y: 입력)

[0052] 수학식 2에 기재된 바와 같이, 제2 워드라인(WL2)으로 제공된 입력(Y)들은 반전된 가중치와 곱셈 연산되어 누적된다. 따라서, 가중치와 곱셈 연산되어 누적된 값(ΣWY)을 얻기 위하여는 수학식 2의 ②식으로 연산된 것과 같이 입력의 합(ΣY)에서 제2 비트라인(BL)에서 형성된 전압(V_{bl})을 디지털로 변환한 값($D\Delta V_{bl}$, K)의 차이를 연산하여야 한다.

[0053] 카운터(200)는 제2 워드라인(WL2)으로 제공되는 입력(Y)들을 제공받고, 입력에 포함된 논리 하이 비트를 계수하여 출력한다. 제2 비트 라인(BL2)에 형성된 전압값은 감지 증폭기(SA2)에 제공되고, 복수의 기준 전압들(Vref)과 비교된다. 감지 증폭기(SA2)는 반전된 가중치와 입력과의 MAC 연산 결과를 검출하고, 이를 디지털로 변환하여 출력(K)한다.

[0054] 감산기는 카운터가 계수한 입력에서의 논리 하이 비트들의 개수와 감지 증폭기(SA2)가 출력한 연산 결과의 차이를 구하여 출력한다. 상술한 바와 같이 감산기의 연산 결과는 제2 워드 라인(WL2)로 제공된 입력(Y)과 가중치와의 MAC 연산 결과에 상응한다.

[0055] 도 6은 본 실시예에 의한 연산 장치(1)의 동작을 설명하기 위한 개요도이다. 도 6을 참조하면, 제1 워드 라인들(WL_{11}, WL_{12}, WL_{13}, ..., WL_{1n})을 통하여 제공된 입력(Xi)들과 제2 워드 라인들(WL_{21}, WL_{22}, WL_{23}, ..., $WL2_n$)을 통하여 입력(Yi)이 제공된다. 도 6으로 예시된 것과 같이 제2 워드라인들(WL_{21}, WL_{22}, WL_{23}, ..., $WL2_n$)을 통하여 제공된 입력들은 제1 워드 라인들(WL_{11}, WL_{12}, WL_{13}, ..., WL_{1n})을 통하여 제공된 입력들(Xi)에 대하여 한 비트씩 스트라이드된 데이터일 수 있다.

[0056] 종래 기술에서는 입력을 한 비트씩 입력하여 가중치와의 MAC 연산을 수행하였다. 그러나, 본 실시예에 의하면, 위의 예에서 도시된 바와 같이, 제1 입력과 제2 입력의 두 비트씩 입력하여 MAC 연산을 수행할 수 있으므로, 종래 기술에 비하여 두 배 높은 처리량을 얻을 수 있다는 장점이 제공된다.

[0058] 도 7(a) 및 도 7(b)는 본 실시예에 의한 검출 방법의 개요를 도시한 순서도이다. 도 7(a)를 참조하면, 본 실시예에 의한 뉴럴 네트워크 출력값 검출 방법은: 카운터(200)에 입력 비트들을 제공하는 단계(S100)와, 카운터가 입력에 포함된 턴 온 비트(turn-on bit)의 개수를 계수하는 단계(S200)와, 계수 결과 턴 온 비트의 개수가 입력 비트 수의 k 분할 미만일 때, 센싱 마진(sensing marging)이 증가하도록 비트 라인 전압 형성 시간을 증가시켜 검출한다(S300).(k: 2 이상 자연수)

[0059] 도 7(b)를 참조하면, 본 실시예에 의한 뉴럴 네트워크 출력값 검출 방법은: 카운터(200)에 입력 비트들을 제공하는 단계(S110)와, 카운터(200)가 입력에 포함된 턴 온 비트(turn-on bit)의 개수를 계수하는 단계(S210)와, 계수 결과 턴 온 비트의 개수가 입력 비트 수의 k 분할 이상일 때, 카운터가 입력 비트가 반전된 반전 입력 비트를 뉴럴 네트워크에 제공하는 단계(S310)와, 센싱 마진(sensing marging)이 증가하도록 비트 라인 전압 형성 시간을 증가시켜 검출하는 단계(S410)를 포함한다.(k: 2 이상 자연수)

[0060] 이하에서는, 용이한 이해와 설명을 위하여 제1 비트 라인(BL1)과 제2 비트 라인(BL2)에 7 개의 메모리 소자들이 연결된 경우를 예시한다. 이로부터 제1 워드 라인으로 입력되는 제1 입력 비트수는 X1, X2, ..., X7의 7이고, 제2 워드 라인으로 입력되는 제2 입력의 입력 비트수는 Y1, Y2, ..., Y7의 7이다. 따라서, 입력과 가중치의 곱을 누적하는 MAC 연산 결과는 0 내지 7 까지의 총 8개의 상태가 있을 수 있다.

[0061] 일 실시예로, 검출 방법은 메모리에 저장된 가중치의 합을 저장하는 단계를 더 포함할 수 있다. 가중치의 합을 연산하면 한 프레임의 입력에 대해서 계속 사용할 수 있으므로 연산 시간 지연의 부담이 크지 않다. 카운터(200, 도 5(b) 참조)를 통하여 입력이 제공된다(S100, S110). 본 실시예는 예시된 것과 같이 연산 장치의 제1 워드 라인들(WL_{11}, WL_{12}, WL_{13}, ..., WL_{17})을 통하여 제공된 입력들(Xi)과 제2 워드라인들(WL_{21}, WL_{22}, WL_{23}, ..., $WL2_n$)을 통하여 제공된 입력들(Yi)은 모두 카운터(200, 도 5(b) 참조)를 통하여 제공된다.

[0062] 카운터(200, 도 5(b) 참조)는 제1 워드 라인들(WL_{11}, WL_{12}, WL_{13}, ..., WL_{17})을 통하여 제공된 입력들(Xi)에서 제1 스위치(TR1, 도 3(a) 내지 도 3(d) 참조)을 도통시키는 턴 온 비트(Turn on bit)의 개수를 계수하고, 제2 워드라인들(WL_{21}, WL_{22}, WL_{23}, ..., $WL2_n$)을 통하여 제공된 입력들(Yi)에서 제2 스위치(TR2, 도 4(a) 내지 도 4(d) 참조)을 도통시키는 턴 온 비트(Turn on bit)의 개수를 계수한다(S200, S210).

[0063] 도 3(a) 내지 도 3(d)로 예시된 실시예와 도 4(a) 내지 도 4(d)로 예시된 실시예와 같이 제1 스위치(TR1)와 제2 스위치(TR2)가 NMOS 트랜지스터인 경우에 턴 온 비트는 입력에 포함된 논리 하이 상태의 비트일 수 있으며, 카운터는 입력에 포함된 논리 하이 상태의 비트를 계수한다. 도시되지 않은 실시예에서 제1 스위치와 제2 스위치가 PMOS 트랜지스터인 경우에 턴 온 비트는 입력에 포함된 논리 로우 상태의 비트일 수 있으며, 카운터는 입력에 포함된 논리 로우 상태의 비트를 계수한다.

[0064] 카운터의 계수 결과, 턴 온 비트의 개수가 MAC 연산에서 있을 수 있는 경우의 수 절반 미만인 경우를 설명한다. 상술한 바와 같이, 가중치를 저장하는 메모리 소자가 7 개인 경우에 발생할 수 있는 경우의 수는 총 8이다. 즉, 가중치와 입력의 곱이 모두 0인 경우에서 가중치와 입력이 모두 1로, 가중치와 입력을 곱하여 누적한 경우 누적된 값은 7인 경우까지 총 8가지의 경우가 있을 수 있다.

[0065] 카운터의 계수 결과가 4 미만인 경우에, 총 7 bit의 입력을 제공받고, 카운터가 출력하는 계수 결과는 000_2, 001_2, 010_2, 011_2의 세 경우이고, 이 때 MSB는 항상 0이다. 따라서, 카운터의 계수 결과의 MSB로부터 계수 결과의 과반 여부를 파악할 수 있다.

[0066] 도 8(a)는 카운터의 계수 결과가 4 미만인 경우에, 비트 라인에서 발생할 수 있는 전압의 분포를 나타낸 도면이다. 도시된 바와 같이 입력에 대하여 0 ~ 3 까지의 전압 분포가 있을 수 있다. 0 ~ 3 사이에 분포된 전압을 검출할 때, 비트 라인에 전압을 형성하는 시간을 증가시켜 인접한 전압 분포의 거리를 증가시킨다.

[0067] [수학식 3]

$$i = C\frac{dv}{dt} \quad \cdots\cdots ①$$

$$dv = \frac{I}{C}dt \quad \cdots\cdots ②$$

[0068]

[0069] 비트라인을 등가적인 커패시터로 고려하면, 위의 수학식 3의 ②식으로 예시된 것과 같이 비트 라인에 형성된 전압을 표시할 수 있다. 즉, 비트라인에 전압이 형성되는 시간을 증가시킴으로써 비트라인에 형성되는 전압을 선형적으로 증가시킬 수 있다(S300). 일 예로, 제1 스위치(TR1, 도 3(a) 내지 도 3(b) 참조)를 도통시키는 시간을 두 배로 증가시키면 도 8(a)로 예시된 것과 같이 인접한 전압 차이가 ΔV에서 도 8(b)로 예시된 것과 2ΔV로 증가시킬 수 있으며, 이로부터 센싱 마진을 확보할 수 있다는 장점이 제공된다.

[0070] 카운터의 계수 결과, 턴 온 비트의 개수가 MAC 연산에서 있을 수 있는 경우의 수 절반 이상인 경우를 설명한다. 상술한 바와 같이, 가중치를 저장하는 메모리 소자가 7 개인 경우에 발생할 수 있는 경우의 수는 총 8이고, 카운터의 계수 결과가 4 이상인 경우에 카운터가 출력하는 계수 결과는 100_2, 101_2, 110_2, 111_2의 세 경우이고, 이 때 MSB는 항상 1이다. 따라서, 카운터의 계수 결과의 MSB로부터 계수 결과의 과반 여부를 파악할 수 있다.

[0071] 턴 온 비트의 개수가 MAC 연산 경우의 수의 절반 이상일 때, 카운터(200)는 제공된 입력을 반전하여 워드 라인들에 출력한다(S310). 즉, 턴 온비트의 개수가 절반 이상일 때 비트 라인에 형성될 수 있는 전압의 분포는 4, 5, 6, 7이 있을 수 있으나, 반전된 입력이 제공됨에 따라 비트 라인에 형성될 수 있는 전압의 분포는 0, 1, 2, 3과 같이 형성되어 턴 온 비트의 개수가 MAC 연산에서 있을 수 있는 경우의 수 절반 미만인 경우와 유사한다.

[0072] 따라서, 비트라인에 전압이 형성되는 시간을 증가시킴으로써 비트라인에 형성되는 전압을 선형적으로 증가시켜서 검출을 수행할 수 있다(S410). 상술한 구성으로부터 센싱 마진을 확보할 수 있다는 장점이 제공된다.

[0073] 비트 라인에 형성되는 전압을 디지털로 변환한 결과는 아래의 수학식 4와 같이 표시될 수 있다.

[0074] [수학식 4]

$$D(\Delta V_{bl}) = \sum W\overline{X} = \sum W(1-X) \quad \cdots\cdots ①$$

[0075]

$$\sum WX = \sum W - D(\Delta V_{bl}) \quad \cdots\cdots ②$$

[0076] 즉, 반전된 수학식 4의 ②식과 같이 입력과 가중치에 대한 MAC 연산 결과(ΣWX)는 가중치 합(ΣW)에서 비트 라인에 형성된 전압을 검출하여 디지털화한 값($D(\Delta V_{bl})$)의 차이를 연산하여 얻을 수 있다.

[0077] 이와 같이 뉴럴 네트워크 연산 장치의 연산 결과를 검출할 때, 센싱 마진을 확보할 수 있으며, 그로부터 연산된 결과를 보다 정확하게 검출할 수 있다는 장점이 제공되고, 보다 많은 비트를 가지는 연산 결과를 용이하고 정확하게 검출할 수 있다는 장점이 제공된다.

[0079] 본 발명에 대한 이해를 돕기 위하여 도면에 도시된 실시예를 참고로 설명되었으나, 이는 실시를 위한 실시예로, 예시적인 것에 불과하며, 당해 분야에서 통상적 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시예가 가능하다는 점을 이해할 것이다. 따라서, 본 발명의 진정한 기술적 보호범위는 첨부된 특허청구범위에 의해 정해져야 할 것이다.

부호의 설명

[0080] 1: 뉴럴 네트워크 연산 장치

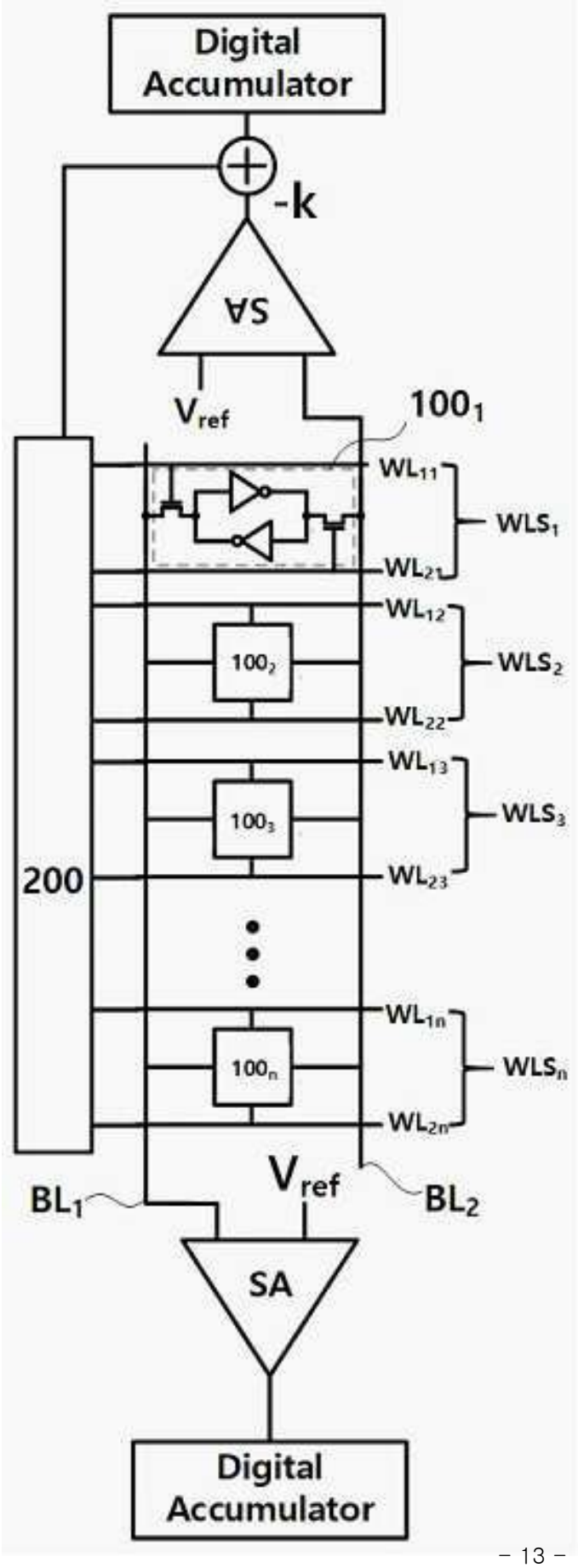
100, 100_1, 100_2, 100_3, ..., 100_n: 메모리

200: 카운터

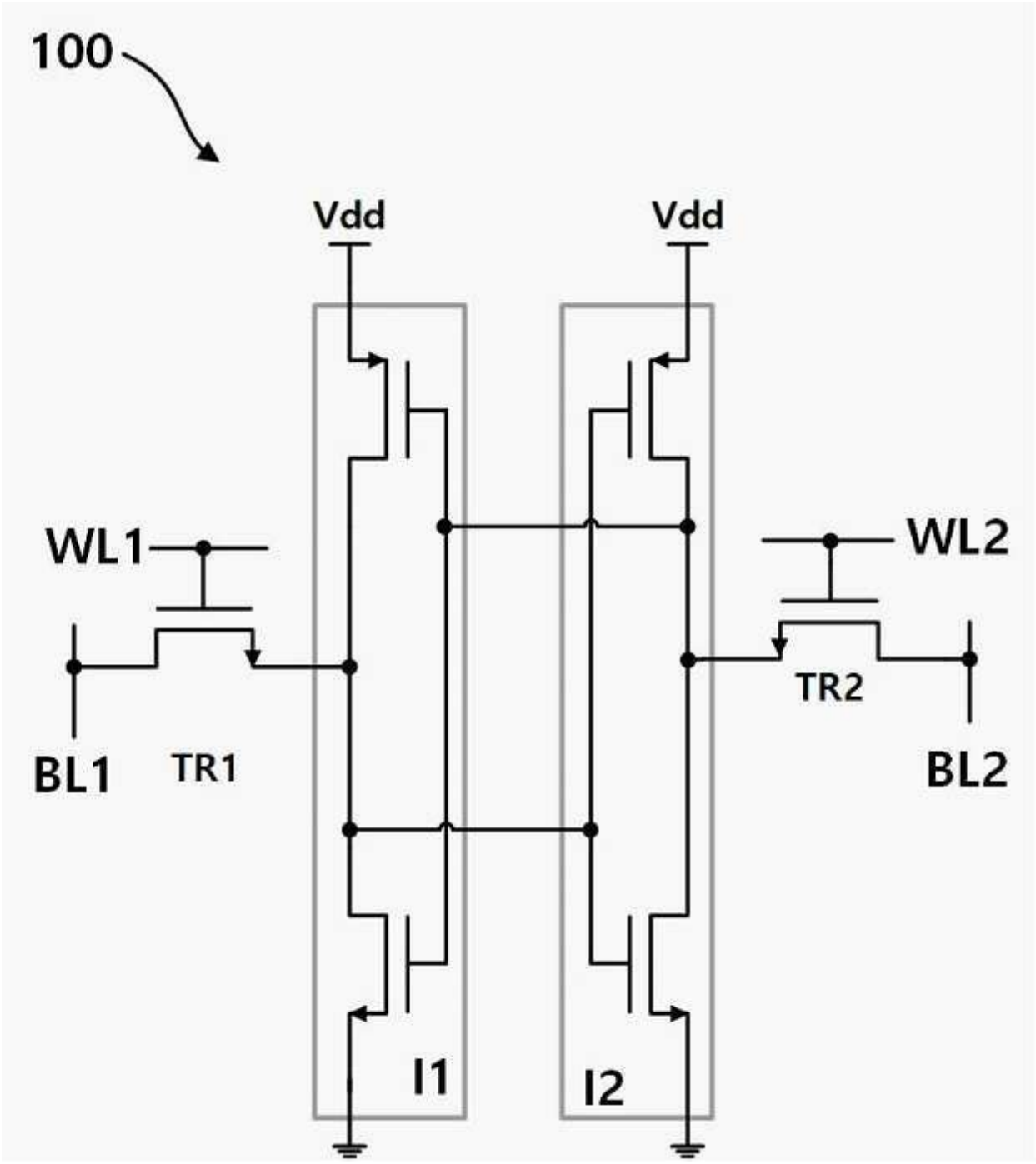
S100~S300, S110~S410: 본 실시예에 의한 검출 방법의 예시적 각 단계

도면

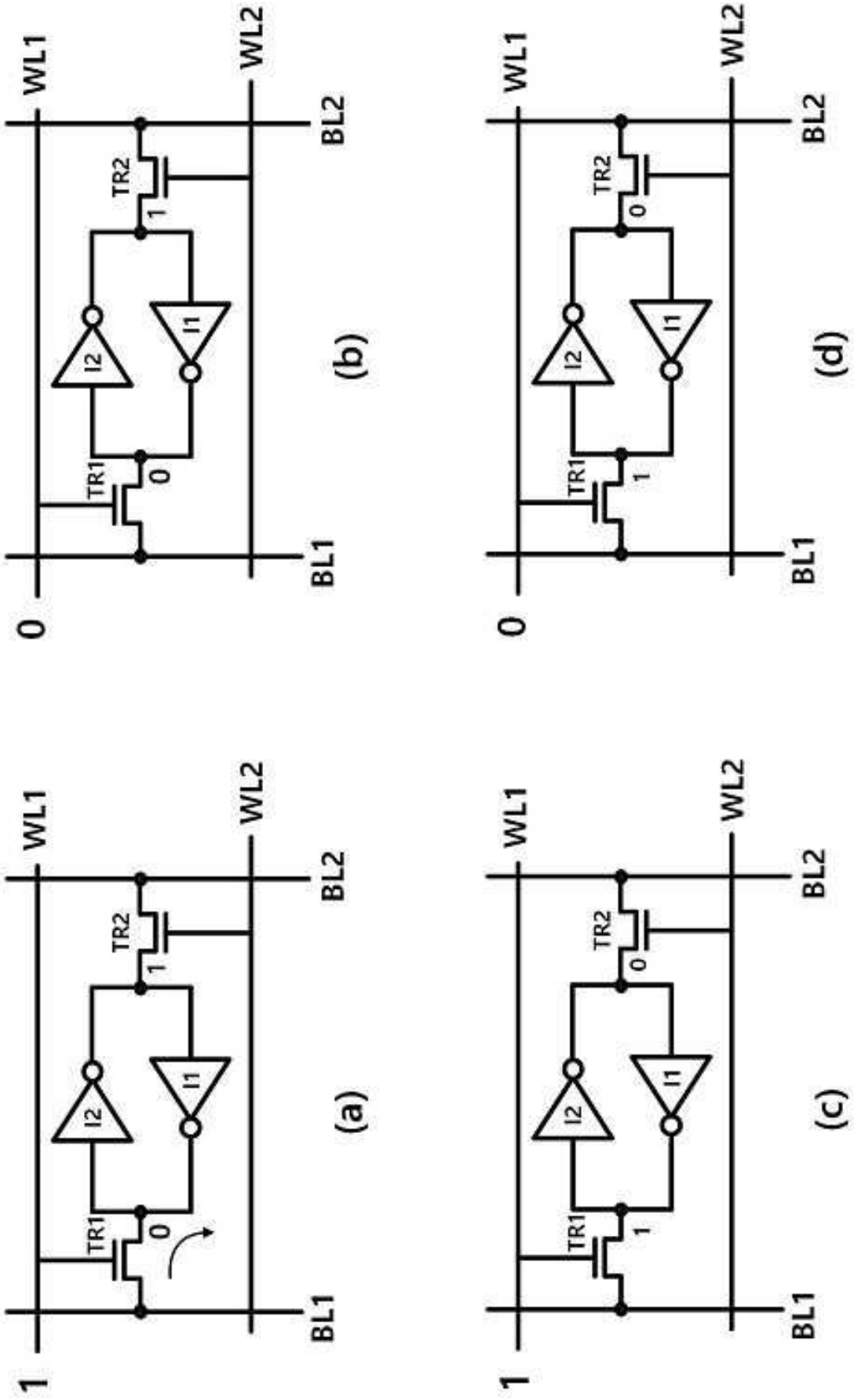
도면1


Digital Accumulator
-k
∀S
V_{ref}
100_1
WL_{11}
WLS_1
WL_{21}
WL_{12}
100_2
WLS_2
WL_{22}
WL_{13}
100_3
WLS_3
WL_{23}
200
WL_{1n}
100_n
WLS_n
WL_{2n}
V_{ref}
BL_1
BL_2
SA
Digital Accumulator

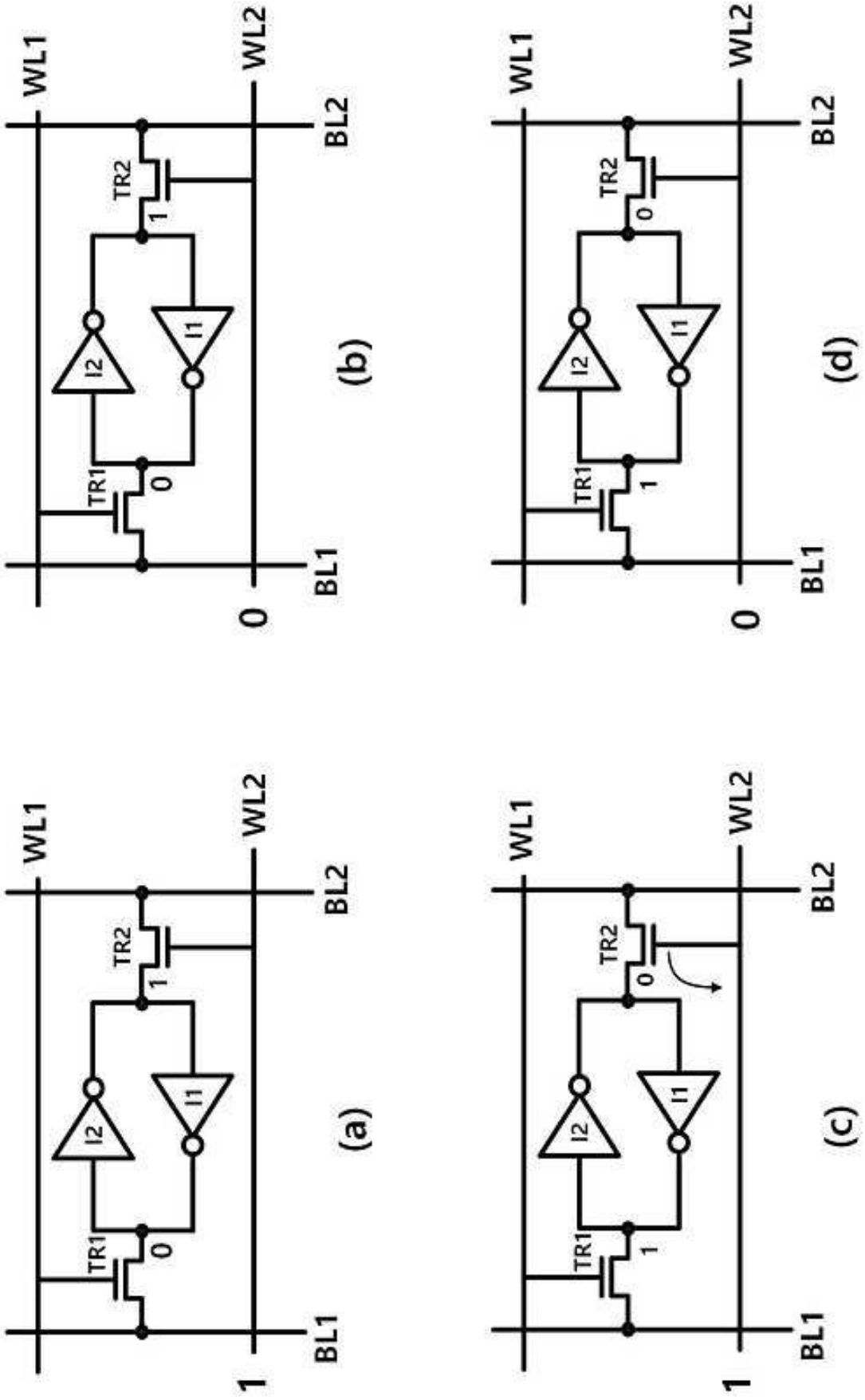
도면2



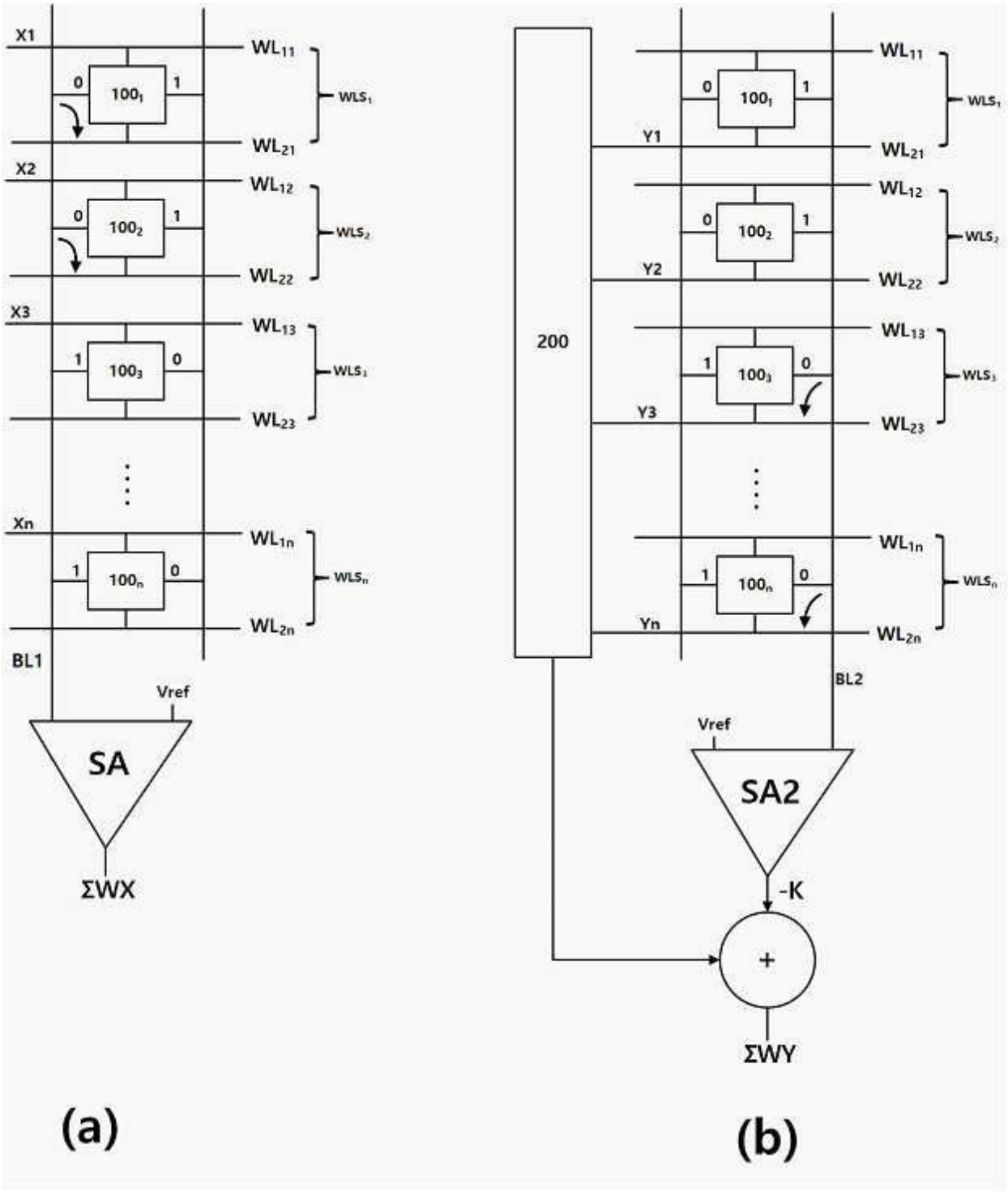
도면3



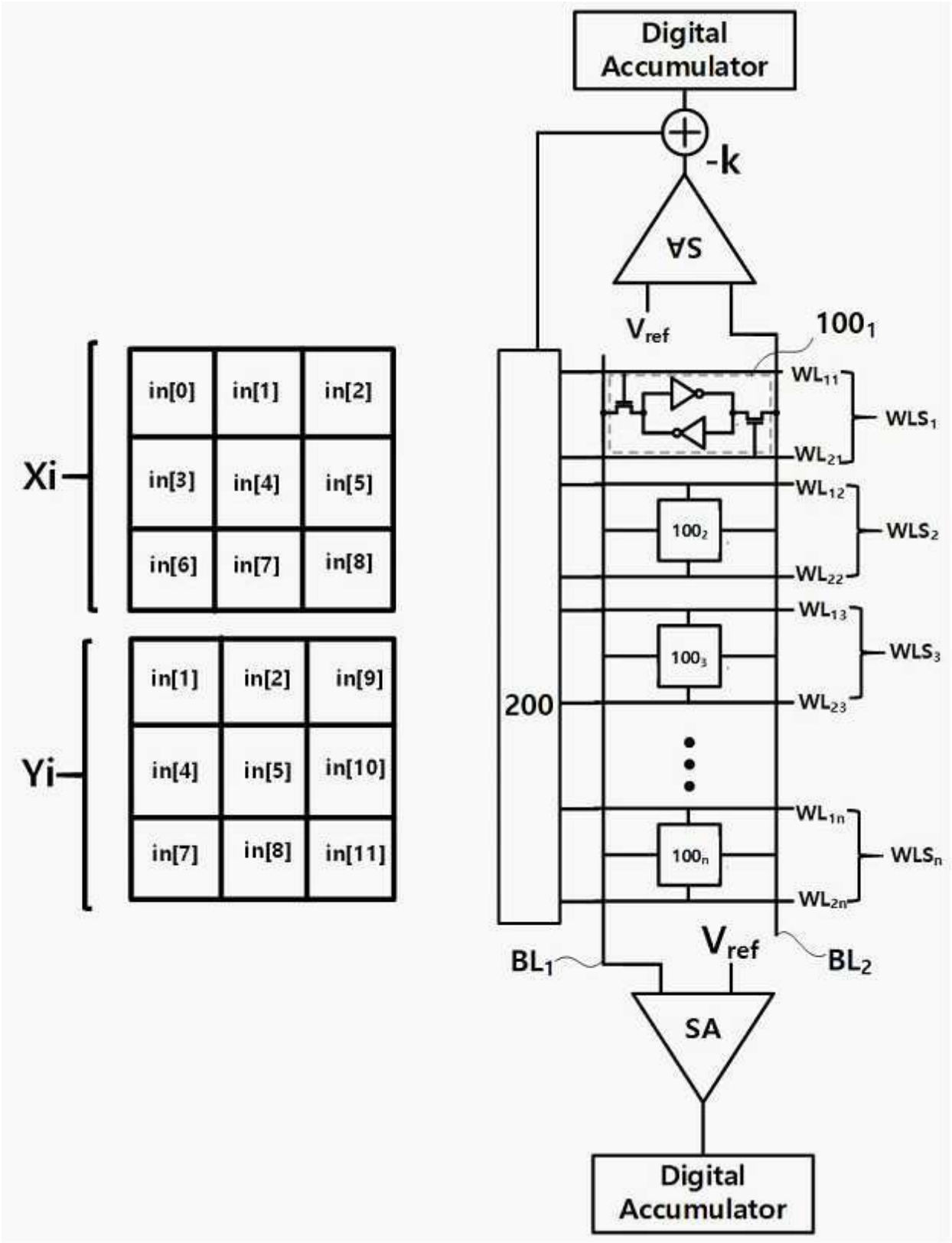
도면4



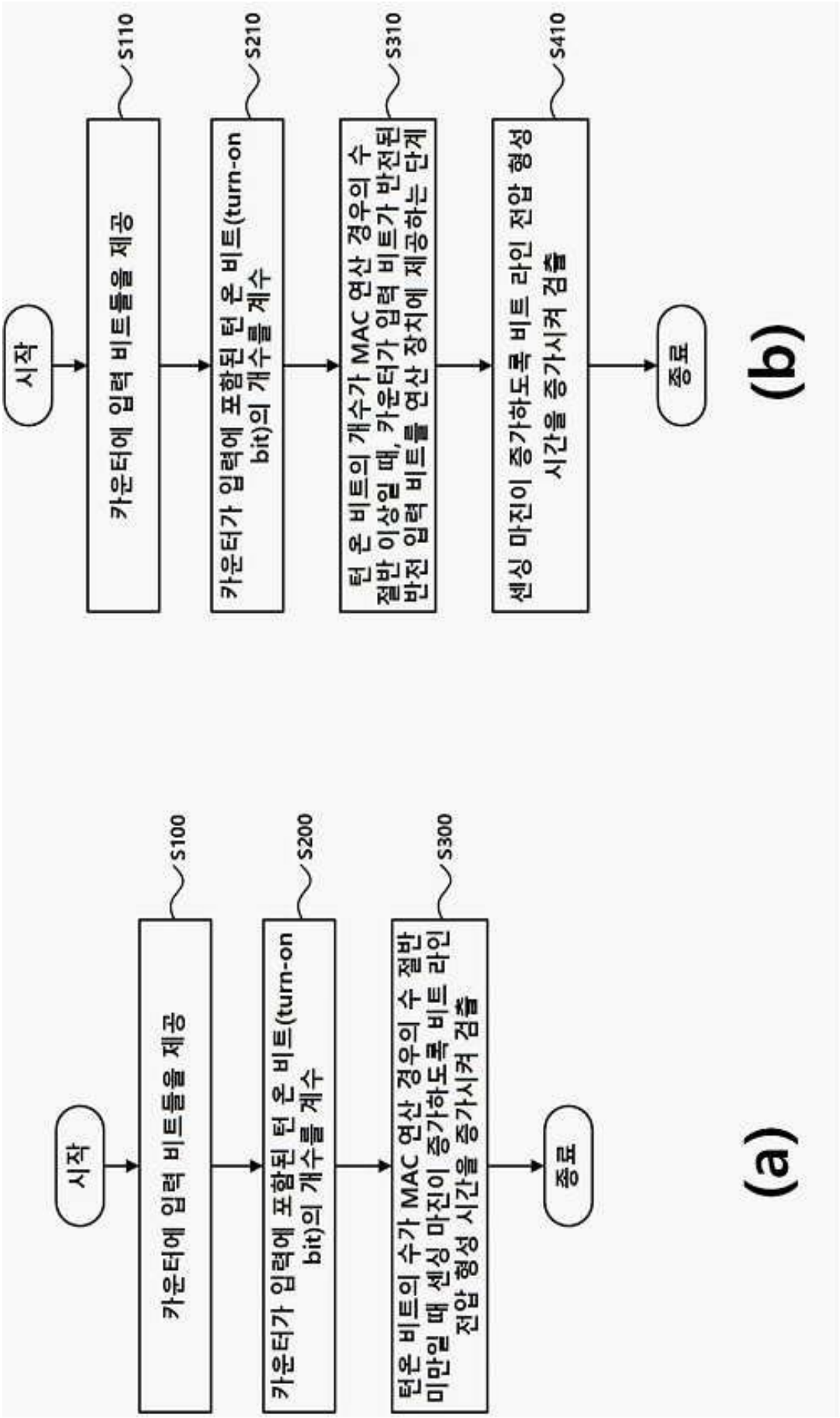
도면5


X1
X2
X3
Xn
WL11
WL21
WL12
WL22
WL13
WL23
WL1n
WL2n
WLS1
WLS2
WLS3
WLSn
1001
1002
1003
100n
0
1
BL1
Vref
SA
ΣWX
(a)
200
Y1
Y2
Y3
Yn
BL2
SA2
-K
+
ΣWY
(b)

도면6


Digital Accumulator
-k
∀S
V_{ref}
100_1
WL_{11}
WL_{21}
WLS_1
WL_{12}
100_2
WL_{22}
WLS_2
WL_{13}
100_3
WL_{23}
WLS_3
200
WL_{1n}
100_n
WL_{2n}
WLS_n
V_{ref}
BL_1
BL_2
SA
Digital Accumulator
Xi
in[0] in[1] in[2]
in[3] in[4] in[5]
in[6] in[7] in[8]
Yi
in[1] in[2] in[9]
in[4] in[5] in[10]
in[7] in[8] in[11]

도면7


(a)
시작
카운터에 입력 비트들을 제공
S100
카운터가 입력에 포함된 턴 온 비트(turn-on bit)의 개수를 계수
S200
턴온 비트의 수가 MAC 연산 경우의 수 절반 미만일 때 센싱 마진이 증가하도록 비트 라인 전압 형성 시간을 증가시켜 검출
S300
종료
(b)
시작
카운터에 입력 비트들을 제공
S110
카운터가 입력에 포함된 턴 온 비트(turn-on bit)의 개수를 계수
S210
턴 온 비트의 개수가 MAC 연산 경우의 수 절반 이상일 때, 카운터가 입력 비트가 반전된 반전 입력 비트를 연산 장치에 제공하는 단계
S310
센싱 마진이 증가하도록 비트 라인 전압 형성 시간을 증가시켜 검출
S410
종료

도면8