



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0080716
(43) 공개일자 2025년06월05일

(51) 국제특허분류(Int. Cl.)
G06F 15/78 (2006.01) G06F 7/544 (2017.01)
G06N 3/065 (2023.01) G11C 13/00 (2006.01)
(52) CPC특허분류
G06F 15/7821 (2013.01)
G06F 7/5443 (2013.01)
(21) 출원번호 10-2024-0032407
(22) 출원일자 2024년03월07일
심사청구일자 2024년03월07일
(30) 우선권주장
1020230167382 2023년11월28일 대한민국(KR)

(71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
(72) 발명자
박중선
서울특별시 서초구 신반포로 32
장윤호
서울특별시 성북구 고려대로7다길 38 (안암동1가)
(74) 대리인
특허법인주연케이알피

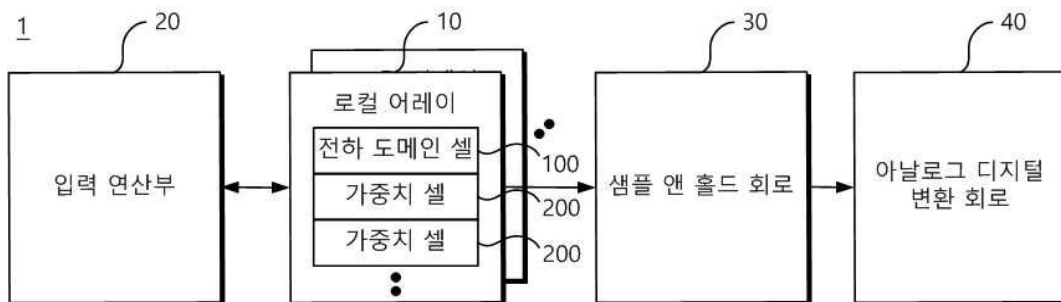
전체 청구항 수 : 총 11 항

(54) 발명의 명칭 저항성 메모리 기반의 프로세싱 인 메모리 장치 및 그 동작 방법

(57) 요약

본 발명은 저항성 메모리 기반의 프로세싱 인 메모리 장치에 관한 것으로, 본 발명에 따른 프로세싱 인 메모리 장치는, 적어도 하나 이상의 로컬 어레이 및 상기 로컬 어레이의 연산에 대한 예측 값에 기초하여 입력 값을 재설정하고, 상기 재설정된 입력 값을 상기 로컬 어레이에 인가하는 입력 연산부를 포함할 수 있고, 상기 로컬 어레이는 복수의 가중치를 저장하는 적어도 하나 이상의 가중치 셀 및 상기 재설정된 입력 값에 기초하여 저장된 상기 가중치를 전압 값으로 변환하는 전하 도메인 셀을 포함할 수 있다.

대표도 - 도1



(52) CPC특허분류

G06N 3/065 (2023.01)

G11C 13/0021 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711195797
과제번호	00229028
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	PIM인공지능반도체핵심기술개발(설계)
연구과제명	e-MRAM에 기반한 고신뢰성 저전력 인증 하드웨어 개발
과제수행기관명	고려대학교 산학협력단
연구기간	2023.04.01 ~ 2026.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711199104
과제번호	2022M3H4A1A04096339
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	나노소재기술개발
연구과제명	비휘발성 임베디드 메모리용 스핀스와핑 기반 스핀토크 소재 개발
과제수행기관명	고려대학교 산학협력단
연구기간	2022.10.01 ~ 2027.06.30

명세서

청구범위

청구항 1

저항성 메모리 기반의 프로세싱 인 메모리 장치에 있어서,

적어도 하나 이상의 로컬 어레이; 및

상기 로컬 어레이의 연산에 대한 예측 값에 기초하여 입력 값을 재설정하고, 상기 재설정된 입력 값을 상기 로컬 어레이에 인가하는, 입력 연산부를 포함하고,

상기 로컬 어레이는,

복수의 가중치를 저장하는 적어도 하나 이상의 가중치 셀; 및

상기 재설정된 입력 값에 기초하여 저장된 상기 가중치를 전압 값으로 변환하는 전하 도메인 셀을 포함하는, 프로세싱 인 메모리 장치.

청구항 2

제1항에 있어서,

상기 가중치 셀은 저항성 메모리를 포함하고,

상기 전하 도메인 셀은 래치(latch) 구조의 SRAM(Static Random Access Memory)를 포함하는, 프로세싱 인 메모리 장치.

청구항 3

제1항에 있어서,

상기 로컬 어레이는,

상기 전하 도메인 셀의 소스 라인 전압 값을 전하 공유(charge sharing)함으로써, 곱셈 누적 연산을 수행하는, 프로세싱 인 메모리 장치.

청구항 4

제3항에 있어서,

상기 프로세싱 인 메모리 장치는,

상기 곱셈 누적 연산의 결과 값을 저장하는 적어도 하나 이상의 샘플 앤 홀드 회로; 및

적어도 하나 이상의 상기 샘플 앤 홀드 회로에 저장된 상기 결과 값들을 전하 공유하고, 상기 결과 값들을 디지털 값으로 변환하는 아날로그 디지털 변환 회로를 더 포함하는, 프로세싱 인 메모리 장치.

청구항 5

제3항에 있어서,

상기 예측 값은 상기 가중치 및 상기 입력 값에 곱셈 누적 연산을 수행한 결과이고,

상기 입력 연산부는,

상기 예측 값이 0인 경우에, 상기 입력 값을 0으로 재설정하는, 프로세싱 인 메모리 장치.

청구항 6

제2항에 있어서,

상기 적어도 하나 이상의 가중치 셀은,

상기 복수의 가중치 중에서 필터의 가중치 패턴에 기초하여 가중치를 저장하고,

상기 복수의 가중치 중에서 상기 가중치 패턴이 반복되는 경우에는 상기 가중치를 제거하는, 프로세싱 인 메모리 장치.

청구항 7

제6항에 있어서,

상기 입력 연산부는,

상기 가중치 패턴에 따른 상기 가중치와 곱셈 연산하는 적어도 하나 이상의 입력 값을 합산하고,

상기 합산된 입력 값을 상기 로컬 어레이에 인가하는, 프로세싱 인 메모리 장치.

청구항 8

저항성 메모리 기반의 프로세싱 인 메모리 장치의 동작 방법에 있어서,

복수의 가중치를 저장하는 단계;

로컬 어레이의 연산에 대한 예측 값에 기초하여 입력 값을 재설정하는 단계;

상기 재설정된 입력 값을 상기 로컬 어레이에 인가하는 단계;

상기 재설정된 입력 값에 기초하여 저장된 상기 가중치를 전압 값으로 변환하는 단계;

상기 전압 값을 전하 공유(charge sharing)함으로써, 곱셈 누적 연산을 수행하는 단계;

상기 곱셈 누적 연산의 결과 값을 저장하는 단계; 및

저장된 상기 결과 값들을 전하 공유하고, 상기 결과 값들을 디지털 값으로 변환하는 단계를 포함하는, 프로세싱 인 메모리 장치의 동작 방법.

청구항 9

제8항에 있어서,

상기 예측 값은 상기 가중치 및 상기 입력 값에 곱셈 누적 연산을 수행한 결과이고,

상기 입력 값을 재설정하는 단계는,

상기 예측 값이 0인 경우에, 상기 입력 값을 0으로 재설정하는, 프로세싱 인 메모리 장치의 동작 방법.

청구항 10

제8항에 있어서,

상기 복수의 가중치를 저장하는 단계는,

상기 복수의 가중치 중에서 필터의 가중치 패턴에 기초하여 가중치를 저장하는 단계; 및

상기 복수의 가중치 중에서 상기 가중치 패턴이 반복되는 경우에는 상기 가중치를 제거하는 단계를 포함하는, 프로세싱 인 메모리 장치의 동작 방법.

청구항 11

제10항에 있어서,

상기 재설정된 입력 값을 상기 로컬 어레이에 인가하는 단계는,

상기 가중치 패턴에 따른 상기 가중치와 곱셈 연산하는 적어도 하나 이상의 입력 값을 합산하는 단계; 및

상기 합산된 입력 값을 상기 로컬 어레이에 인가하는 단계를 더 포함하는, 프로세싱 인 메모리 장치의 동작 방법.

발명의 설명

기술 분야

[0001] 본 발명은 전하 도메인 연산이 가능한 저항성 메모리 기반의 프로세싱 인 메모리 장치 및 그 동작 방법에 관한 것이다.

배경 기술

[0002] 최근 높은 정확도를 달성하는 인공 신경망의 발전으로, 요구되는 곱셈-누적 연산의 수가 증가했고 이는 프로세서와 메모리 간의 막대한 양의 데이터 이동을 야기한다. 이로 인해 연산 뿐만 아니라 데이터 이동에 대한 에너지 소모 또한 해결해야 하는 중요한 문제로 대두되고 있다.

[0003] 하지만 기존의 컴퓨터 구조 상 많은 연산과 막대한 데이터 이동을 처리하기에 매우 에너지 비효율적이다. 따라서 프로세싱-인-메모리(PIM, Processin-In-Memory)이라는 메모리 기술이 개발되었다. 프로세싱-인-메모리는 기존 읽기/쓰기 동작을 지원하며 연산 기능을 메모리 내부에서 추가로 지원하여 데이터 이동에 따른 에너지 소모를 해결한다. 그 중에서도 ReRAM, MRAM, PCRAM 등 비휘발성 저항성 메모리를 이용한 프로세싱-인-메모리는 전원이 없이도 인공신경망의 가중치를 작은 면적으로 모두 저장할 수 있다. 그러므로 저항성 메모리 기반의 프로세싱-인-메모리는 가중치 데이터 이동을 제거하고, 메모리 내부에서 전류 합산 기법으로 효율적인 곱셈-누적 연산이 가능하다. 그러나 저항성 메모리 기반의 프로세싱-인-메모리가 실제로 칩으로 제작될 때, 발생하는 공정 변이로 인하여 신뢰성 높은 연산을 수행하기가 어렵다.

[0004] 저항성 메모리 기반의 프로세싱-인-메모리가 신뢰성 높은 연산을 수행하기 위하여 연산 시에 활성화하는 로우라인(row-line)의 개수를 줄여야 한다. 이에 따라, 활성화되는 워드라인(word-line)의 개수가 줄어들면서 한 사이클에 수행되는 곱셈-누적 연산의 크기가 줄어든다. 줄어든 연산만큼 저항성 메모리에 접근하여 워드라인을 활성화하는 횟수는 늘어나기 때문에, 곱셈-누적 연산 속도 및 에너지 효율성은 떨어진다는 문제가 있다.

선행기술문헌

특허문헌

[0005] (특허문헌 0001) 한국 등록 특허 제10-2298607호

발명의 내용

해결하려는 과제

[0006] 본 발명의 목적은 전하 도메인 연산이 가능한 저항성 메모리 기반의 프로세싱 인 메모리 장치 및 그 동작 방법을 제공함에 있다.

과제의 해결 수단

[0007] 본 발명에 따르면, 저항성 메모리 기반의 프로세싱 인 메모리 장치는 적어도 하나 이상의 로컬 어레이 및 상기 로컬 어레이의 연산에 대한 예측 값에 기초하여 입력 값을 재설정하고, 상기 재설정된 입력 값을 상기 로컬 어레이에 인가하는, 입력 연산부를 포함할 수 있고, 상기 로컬 어레이는 복수의 가중치를 저장하는 적어도 하나 이상의 가중치 셀 및 상기 재설정된 입력 값에 기초하여 저장된 상기 가중치를 전압 값으로 변환하는 전하 도메인 셀을 포함할 수 있다.

[0008] 그리고 상기 가중치 셀은 저항성 메모리를 포함할 수 있고, 상기 전하 도메인 셀은 래치(latch) 구조의 SRAM(Static Random Access Memory)를 포함할 수 있다.

[0009] 그리고 상기 로컬 어레이는, 상기 전하 도메인 셀의 소스 라인 전압 값을 전하 공유(charge sharing)함으로써, 곱셈 누적 연산을 수행할 수 있다.

[0010] 그리고 상기 프로세싱 인 메모리 장치는 상기 곱셈 누적 연산의 결과 값을 저장하는 적어도 하나 이상의 샘플

앤 홀드 회로 및 적어도 하나 이상의 상기 샘플 앤 홀드 회로에 저장된 상기 결과 값들을 전하 공유하고, 상기 결과 값들을 디지털 값으로 변환하는 아날로그 디지털 변환 회로를 더 포함할 수 있다.

[0011] 그리고 상기 예측 값은 상기 가중치 및 상기 입력 값에 곱셈 누적 연산을 수행한 결과일 수 있고, 상기 입력 연산부는 상기 예측 값이 0인 경우에, 상기 입력 값을 0으로 재설정할 수 있다.

[0012] 그리고 상기 적어도 하나 이상의 가중치 셀은, 상기 복수의 가중치 중에서 필터의 가중치 패턴에 기초하여 가중치를 저장할 수 있고, 상기 복수의 가중치 중에서 상기 가중치 패턴이 반복되는 경우에는 상기 가중치를 제거할 수 있다.

[0013] 그리고 상기 입력 연산부는, 상기 가중치 패턴에 따른 상기 가중치와 곱셈 연산하는 적어도 하나 이상의 입력 값을 합산할 수 있고, 상기 합산된 입력 값을 상기 로컬 어레이에 인가할 수 있다.

[0014] 또한, 본 발명에 따르면, 저항성 메모리 기반의 프로세싱 인 메모리 장치의 동작 방법은 복수의 가중치를 저장하는 단계, 로컬 어레이의 연산에 대한 예측 값에 기초하여 입력 값을 재설정하는 단계, 상기 재설정된 입력 값을 상기 로컬 어레이에 인가하는 단계, 상기 재설정된 입력 값에 기초하여 저장된 상기 가중치를 전압 값으로 변환하는 단계, 상기 전압 값을 전하 공유(charge sharing)함으로써, 곱셈 누적 연산을 수행하는 단계, 상기 곱셈 누적 연산의 결과 값을 저장하는 단계 및 저장된 상기 결과 값들을 전하 공유하고, 상기 결과 값들을 디지털 값으로 변환하는 단계를 포함할 수 있다.

[0015] 그리고 상기 예측 값은 상기 가중치 및 상기 입력 값에 곱셈 누적 연산을 수행한 결과일 수 있고, 상기 입력 값을 재설정하는 단계는, 상기 예측 값이 0인 경우에, 상기 입력 값을 0으로 재설정할 수 있다.

[0016] 그리고 상기 복수의 가중치를 저장하는 단계는, 상기 복수의 가중치 중에서 필터의 가중치 패턴에 기초하여 가중치를 저장하는 단계 및 상기 복수의 가중치 중에서 상기 가중치 패턴이 반복되는 경우에는 상기 가중치를 제거하는 단계를 포함할 수 있다.

[0017] 그리고 상기 재설정된 입력 값을 상기 로컬 어레이에 인가하는 단계는, 상기 가중치 패턴에 따른 상기 가중치와 곱셈 연산하는 적어도 하나 이상의 입력 값을 합산하는 단계 및 상기 합산된 입력 값을 상기 로컬 어레이에 인가하는 단계를 더 포함할 수 있다.

발명의 효과

[0018] 본 발명의 실시예에 따른 저항성 메모리 기반의 프로세싱 인 메모리 장치 및 그 동작 방법은 전하 도메인에서 연산을 수행하므로 신뢰성을 향상시킬 수 있다. 즉, 본 발명에 따른 저항성 메모리 기반의 프로세싱 인 메모리 장치 및 그 동작 방법은 한번에 많은 워드라인을 활성화할 수 있어, 연산 에너지 효율 및 연산 처리 속도를 향상시킬 수 있다.

[0019] 또한, 본 발명의 실시예에 따른 저항성 메모리 기반의 프로세싱 인 메모리 장치 및 그 동작 방법은 가중치 압축 기법을 사용하여 종래의 먼적 오버헤드 문제를 완화시킬 수 있다.

도면의 간단한 설명

[0020] 도 1은 본 발명의 실시예에 따른 저항성 메모리 기반의 프로세싱 인 메모리 장치의 블록구성도이다.

도 2a 및 도 2b는 본 발명의 실시예에 따른 로컬 어레이의 예시이다.

도 3은 본 발명의 실시예에 따른 로컬 어레이의 동작에 대한 예시이다.

도 4a 및 도 4b는 본 발명의 실시예에 따른 가중치 셀의 동작에 대한 예시이다.

도 5는 본 발명의 실시예에 따른 입력 연산부의 동작에 대한 예시이다.

도 6은 본 발명의 실시예에 따른 저항성 메모리 기반의 프로세싱 인 메모리 장치의 동작 방법에 대한 순서도이다.

발명을 실시하기 위한 구체적인 내용

[0021] 이하에서는 본 발명의 기술사상을 본 발명의 기술분야에서 통상의 지식을 가진 자가 용이하게 실시할 수 있을 정도로 상세히 설명하기 위하여, 본 발명의 실시 예들이 첨부된 도면을 참조하여 설명될 것이다.

- [0022] 도 1은 본 발명의 실시예에 따른 저항성 메모리 기반의 프로세싱 인 메모리 장치(1)의 블록구성도이고, 도 2a 및 도 2b는 본 발명의 실시예에 따른 로컬 어레이(10)의 예시이다.
- [0023] 도 1을 참조하면, 프로세싱 인 메모리 장치(1)는 로컬 어레이(10), 입력 연산부(20), 샘플 앤 홀드 회로(30) 및 아날로그 디지털 회로(40)를 포함할 수 있다.
- [0024] 프로세싱 인 메모리 장치(1)는 여러 개의 로컬 어레이(local array, 10)로 구성될 수 있고, 하나의 로컬 어레이(10)는 여러 개의 가중치 셀(weight cell, 100)과 하나의 전하 도메인 셀(charge domain cell, 200)로 구성되어 있다.
- [0025] 도 2a 및 도 2b를 참조하면, 전하 도메인 셀(100)은 재설정된 입력 값에 기초하여 저장된 가중치를 전압 값으로 변환할 수 있다. 전하 도메인 셀(100)은 래치(latch) 구조의 SRAM(Static Random Access Memory)를 포함할 수 있다. 전하 도메인 셀(100)은 가중치가 변환된 전압 값은 전하 기반으로 연산이 수행될 수 있다.
- [0026] 적어도 하나 이상의 가중치 셀(200)은 복수의 가중치를 저장할 수 있다. 저항성 메모리를 포함할 수 있는데, 용량이 큰 고집적 저항성 메모리를 포함할 수 있다. 다만, 이는 예시일 뿐 이에 한정하지 않는다.
- [0027] 다시 도 2b를 참조하면, 가중치 셀(200)이 가중치 '1'을 저장할 때 가중치 셀(200)의 왼쪽 소자는 작은 저항 상태가 될 수 있고, 오른쪽 소자는 높은 저항 상태가 될 수 있다. 전하 도메인 셀(100)의 LBL(Local Bit Line)과 LSL(Local Source Line)를 전압 VDD로 프리 차지(pre-charge)시킴으로써, 두 개의 소자에 전류를 흘릴 수 있다. 두 개 소자에 인가되는 전류가 확연히 차이가 날 때, 전하 도메인 셀(100)은 SEb(Sensing Enable signal Bar)에 전압 VSS를 인가하여 소자의 저항 상태를 감지할 수 있고, 가중치를 전압 값으로 변환할 수 있다. 전하 도메인 셀(100)의 오른쪽 소자가 높은 저항일 때는 LSL의 전압은 VDD가 될 수 있다. 이후, 전하 도메인 셀(100)의 CL(Computing Line)에 1bit 입력 값에 해당되는 전압을 인가할 수 있고, 1bit X 1bit 연산을 수행할 수 있다. 그리고 전하 도메인 셀(100)은 소스 라인(SL, Source Line)에 곱셈 결과값을 저장할 수 있다. GSL(Global Source Line) 전압은 곱셈 결과값들의 누적값일 수 있다. 이후, 로컬 어레이(10)는 전하 도메인 셀(100)의 소스 라인(SL) 전압 값을 전하 공유(charge sharing)함으로써, 곱셈 누적 연산을 수행할 수 있다 이후 GSL(Global Source Line) 전압은 곱셈 결과값들의 누적값일 수 있다. 여기에서, 전하 공유는 전하량이 서로 다른 두 개의 축전기가 트랜지스터로 연결되어 있는 경우, 트랜지스터에 전류가 흐르면 전하량이 많은 축전기의 전하가 적은 축전기로 나뉘어서 평형을 이루는 일을 의미한다.
- [0028] 가중치 셀(200)은 복수의 가중치 중에서 필터의 가중치 패턴에 기초하여 가중치를 저장할 수 있다. 즉, 가중치 셀(200)은 가중치 패턴이 반복되는 경우에, 가중치를 제거할 수 있다. 여기에서, 필터는 열방향 가중치 압축 기법에 의해 가중치 패턴을 설정하는 필터로, 가중치 셀(200)에 저장될 수 있다. 이에 대한 구체적인 설명은 도 4a의 설명에서 후술하기로 한다.
- [0029] 입력 연산부(20)는 로컬 어레이(10)의 연산에 대한 예측 값에 기초하여 입력 값을 재설정할 수 있고, 재설정된 입력 값을 로컬 어레이(10)에 인가할 수 있다. 여기에서, 예측 값은 가중치 및 입력 값에 곱셈 누적 연산을 수행한 결과일 수 있다.
- [0030] 입력 연산부(20)는 예측 값이 0인 경우에, 상기 입력 값을 0으로 재설정할 수 있다.
- [0031] 입력 연산부(20)는 가중치 패턴에 따른 가중치와 곱셈 연산하는 적어도 하나 이상의 입력 값을 합산할 수 있다. 입력 연산부(20)는 합산된 입력 값을 상기 로컬 어레이에 인가할 수 있다.
- [0032] 샘플 앤 홀드 회로(30)는 곱셈 누적 연산의 결과 값을 저장할 수 있다.
- [0033] 아날로그 디지털 변환 회로(40)는 샘플 앤 홀드 회로(30)에 저장된 결과 값들을 전하 공유할 수 있고, 결과 값들을 디지털 값으로 변환할 수 있다.
- [0034] 상술한 바와 같이, 로컬 어레이(10)는 단일 면적은 크지만 연산 신뢰성이 높은 SRAM의 특성과 단일이 면적이 작은 저항성 메모리의 특성이 결합될 수 있다.
- [0035] 로컬 어레이(10)는 면적 증가를 최소화할 수 있고, 곱셈-누적 연산은 SRAM에서 많이 사용한 전하 도메인에서 수행되기 때문에 연산 신뢰성을 높일 수 있다. 고신뢰 연산은 PIM 구조에서의 한 사이클에 수행되는 연산의 양을 늘릴 수 있으므로, 결과적으로는 연산 속도 및 에너지 효율을 향상시킬 수 있다.

- [0037] 도 3은 본 발명의 실시예에 따른 로컬 어레이(10)의 동작에 대한 예시이다.
- [0038] 도 3을 참조하면, 활성화 가능한 열 라인(row-line)의 개수가 하나의 행에 존재하는 로컬 어레이(10) 개수보다 작을 때는 (a)에 도시된 구조를 사용할 수 있다. 한편, 활성화 가능한 열 라인(row-line)의 개수가 로컬 어레이(10) 개수보다 많을 때는 (b)에 도시된 구조를 사용할 수 있다. 여기에서, MSB(Most Significant Bit)는 데이터의 최상위 비트를, LSB(Least Significant Bit)는 데이터의 최하위 비트를 의미할 수 있다.
- [0039] 예를 들어, (a) 구조에서는 I1,1~I1,4 와 A1,1~A1,4, B1,1~B1,4의 곱셈 누적 연산을 위하여 먼저 I1,1 and I1,3의 LSB부터 먼저 로컬 어레이(10)에 입력될 수 있다. 이후, 곱셈 누적 연산을 수행한 결과 값인 아날로그 전압 값이 샘플 앤 홀드 회로(30)에 저장될 수 있다. 이후, 아날로그 디지털 변환 회로(40)는 샘플 앤 홀드 회로(30)에 저장된 결과 값들을 디지털 값으로 변환할 수 있다.
- [0040] 한편, (b) 구조에서는 I1,1 and I1,3의 LSB가 로컬 어레이(10)에 입력될 수 있다. 이후, 곱셈-누적 연산의 결과 값을 제1 샘플 앤 홀드 회로(30)에 저장할 수 있다. 이때, 아날로그 디지털 변환 회로(40)는 바로 동작하지 않을 수 있고, I1,2 and I1,4가 바로 입력되어 곱셈-누적 연산의 결과 값이 제2 샘플 앤 홀드 회로(30)에 저장될 수 있다. 제1 및 제2 샘플 앤 홀드 회로(30)에 저장된 전압 즉 결과 값들을 전하 공유한 이후에, 아날로그 디지털 변환 회로(40)를 통해 한꺼번에 결과 값들이 디지털 값으로 변환될 수 있다.
- [0041] 상술한 바와 같이, 본 발명의 실시예에 따른 저항성 메모리 기반의 프로세싱 인 메모리 장치(1)는 전하 도메인에서 연산을 수행하므로 신뢰성을 향상시킬 수 있다. 즉, 본 발명에 따른 저항성 메모리 기반의 프로세싱 인 메모리 장치(1)는 한번에 많은 워드라인을 활성화할 수 있어, 연산 에너지 효율 및 연산 처리 속도를 향상시킬 수 있다.
- [0042] 또한, 본 발명의 실시예에 따른 저항성 메모리 기반의 프로세싱 인 메모리 장치(1)는 가중치 압축 기법을 사용하여 종래의 면적 오버헤드 문제를 완화시킬 수 있다.
- [0044] 도 4a 및 도 4b는 본 발명의 실시예에 따른 가중치 셀(100)의 동작에 대한 예시이다.
- [0045] 도 4a를 참조하면, 본 발명에 따른 프로세싱 인 메모리 장치(1)는 전하 도메인 셀(100)로 인하여 면적이 증가하는 문제가 있다. 이를 해결하기 위하여 가중치 셀(200)에 열방향 가중치 압축 기법이 적용될 수 있다. 열방향 가중치 압축 기법은 열 방향으로 메모리에 저장되어 있는 하나의 필터를 오직 2^N (가중치가 N-bit인 경우) 패턴으로 압축한 기법이다. 가중치를 압축을 했음에도 오류없이 연산을 수행하기 위해서, 로컬 어레이(10)에 입력 값 주입 전에, 가중치 셀(200)은 동일한 패턴의 가중치와 곱해져야 하는 입력 값끼리 먼저 합산할 수 있다. 이때, 입력 값과 가중치의 패턴 관계에 대한 정보는 일반적인 고정적 저항성 메모리에 저장될 수 있다. 결과적으로는 면적이 큰 가중치 셀(200)에 불과 몇 개의 가중치 패턴만 저장할 수 있고, 많은 입력 값 인덱싱(indexing) 정보는 고정적 저항성 메모리에 저장될 수 있으므로, 면적 오버헤드의 크기를 상당히 감소시킬 수 있다.
- [0046] 열방향 가중치 압축 기법에서는 하나의 필터 내부에 있는 가중치들 중에서 하나의 패턴씩만 가중치 셀(100)에 저장시킬 수 있고, 나머지는 제거할 수 있다. 예를 들어, 2-bit 가중치 8개를 가지고 있는 필터의 경우에 00, 01, 10, 11 4가지의 가중치 패턴만 가중치 셀(100)에 저장할 수 있고 나머지는 제거할 수 있다. 참고로, 동일한 필터 내부에 있는 가중치들은 열방향으로 메모리에 매핑되기 때문에 열방향 가중치 압축 기법이라고 칭할 수 있다.
- [0047] 하지만, 이러한 압축 기법은 연산에서 오류가 발생할 수가 있으므로, 입력 값이 전하 도메인 셀(100)에 입력되기 전에, 입력 연산부(20)는 먼저 같은 가중치와 곱해질 입력 값끼리 합산할 수 있다. 도 4b를 참조하면, 제1 필터의 01과 곱해져야 하는 제1 입력 값은 10, 01, 01 이 있으며, 전술한 3개의 입력 값 중 하나의 입력 값으로 먼저 누적될 수 있고, 로컬 어레이(10)에 입력될 수 있다. 이에 따라, 열방향 가중치 압축 기법뿐만 아니라, 가중치 공유 기법을 사용하여 더욱 가중치의 데이터양을 줄일 수가 있다. 열방향 가중치 압축 기법을 사용하면 로컬 어레이(10)에 저장되는 모든 필터의 가중치 패턴들이 동일하기 때문에 하나의 로컬 어레이(10) 내부에서는 하나의 필터의 가중치 패턴들만 남겨놓고 나머지는 공유하여 가중치를 더욱 압축시킬 수 있다.
- [0049] 도 5는 본 발명의 실시예에 따른 입력 연산부(20)의 동작에 대한 예시이다.
- [0050] 가중치 셀(200)은 열방향 가중치 압축 기법으로 인해 고정될 수 있고, 제한된 가중치 패턴이 로컬 어레이(10)에

저장될 수 있다는 점을 고려하여, 입력 연산부(20)는 제로 변환(Zero-conversion) 기법과 제로 스킵(Zero-skipping) 기법이 적용될 수 있다.

- [0051] 제로 변환 기법에서는 로컬 어레이(10)에서 가중치와 곱셈 누적 연산을 하여 결과물이 0이 되는 입력 값들을 예측할 수 있다. 예측 결과를 로컬 어레이(10)에 주입하기 전에 '0' 값으로 만들 수 있다. 도 5를 참조하면, IC-2와 IB-2는 모두 '0' 이 아닌 '1' 이지만, 각각 로컬 어레이(10)에 에 인가되어 1X01+1X11 연산이 이루어질 수 있다. 이에 따라, 연산 결과는 '0' 이 나올 수 있다. 결과값 '0' 이 나오는 경우에 굳이 연산할 필요가 없으므로, 입력 연산부(20)는 입력 IC-2 와 IB-2를 모두 '0' 으로 변환시킬 수 있다. 이와 같은 방법으로 입력 연산부(20)는 제로 변환 기법을 이용하여 '0' 의 비율이 높일 수 있다. 또한, 입력 연산부(20)는 이를 기반으로 '0' 의 비율이 높은 입력 값들을 제로 스킵 기법에 적용할 수 있다. 이에 따라 로컬 어레이(10)의 연산 속도와 에너지 효율을 향상시킬 수 있다.
- [0052] 도 6은 본 발명의 실시예에 따른 저항성 메모리 기반의 프로세싱 인 메모리 장치(1)의 동작 방법에 대한 순서도이다.
- [0053] 도 6을 참조하면, S1000 단계에서, 복수의 가중치 중에서 필터에 대응하는 가중치 패턴에 따른 가중치가 저장될 수 있다. 예를 들어, 도 1의 가중치 셀(200)은 복수의 가중치 중에서 필터에 대응하는 가중치 패턴에 따른 가중치를 저장할 수 있다.
- [0054] S1100 단계에서, 복수의 가중치 중에서 반복되는 가중치 패턴을 가지는 가중치가 제거될 수 있다. 예를 들어, 가중치 셀(100)은 복수의 가중치 중에서 반복되는 가중치 패턴을 가지는 가중치를 제거할 수 있다.
- [0055] S1200 단계에서, 가중치 패턴에 따른 가중치와 곱셈 연산하는 적어도 하나 이상의 입력 값이 합산될 수 있다. 예를 들어, 입력 연산부(20)는 가중치 패턴에 따른 가중치와 곱셈 연산하는 적어도 하나 이상의 입력 값을 합산할 수 있다.
- [0056] S1300 단계에서, 로컬 어레이(10)의 연산에 대한 예측 값에 기초하여 입력 값이 재설정될 수 있다. 예를 들어, 도 1의 입력 연산부(20)는 로컬 어레이(10)의 연산에 대한 예측 값에 기초하여 입력 값을 재설정할 수 있다. 입력 연산부(20)는 입력 값을 제로 변환(Zero-conversion) 기법과 제로 스킵(Zero-skipping) 기법에 적용할 수 있다. 입력 연산부(20)는 예측 값이 0인 경우에, 입력 값을 0으로 재설정할 수 있다. 여기에서, 예측 값은 가중치 및 입력 값에 곱셈 누적 연산을 수행한 결과일 수 있다.
- [0057] S1400 단계에서, 재설정된 입력 값에 기초하여 저장된 가중치를 전압 값으로 변환될 수 있다. 도 1의 전하 도메인 셀(100)은 재설정된 입력 값에 기초하여 저장된 가중치를 전압 값으로 변환할 수 있다.
- [0058] S1500 단계에서, 입력 값이 로컬 어레이(10)에 인가될 수 있다. 예를 들어, 입력 연산부(20)는 재설정되거나 또는 합산된 입력 값을 로컬 어레이(10)에 인가할 수 있다.
- [0059] S1600 단계에서, 전압 값을 전하 공유함으로써, 곱셈 누적 연산이 수행될 수 있다. 예를 들어, 로컬 어레이(10)은 전압 값을 전하 공유함으로써, 곱셈 누적 연산을 수행할 수 있다.
- [0060] S1700 단계에서, 곱셈 누적 연산의 결과 값이 저장될 수 있다. 예를 들어, 샘플 앤 홀드 회로(30)는 곱셈 누적 연산의 결과 값을 저장할 수 있다.
- [0061] S1800 단계에서, 저장된 결과 값들이 전하 공유될 수 있고, 결과 값들이 디지털 값으로 변환될 수 있다. 예를 들어, 아날로그 디지털 변환 회로(40)는 저장된 결과 값들을 전하 공유하고, 결과 값들을 디지털 값으로 변환할 수 있다.
- [0062] 상술한 바와 같이, 본 발명의 실시예에 따른 저항성 메모리 기반의 프로세싱 인 메모리 장치의 동작 방법은 전하 도메인에서 연산을 수행하므로 신뢰성을 향상시킬 수 있다. 즉, 본 발명에 따른 저항성 메모리 기반의 프로세싱 인 메모리 장치의 동작 방법은 한번에 많은 워드라인을 활성화할 수 있어, 연산 에너지 효율 및 연산 처리 속도를 향상시킬 수 있다.
- [0063] 또한, 본 발명의 실시예에 따른 저항성 메모리 기반의 프로세싱 인 메모리 장치의 동작 방법은 가중치 압축 기법을 사용하여 종래의 면적 오버헤드 문제를 완화시킬 수 있다.
- [0065] 상술한 내용은 본 발명을 실시하기 위한 구체적인 실시 예들이다. 본 발명은 상술한 실시 예들 이외에도, 단순

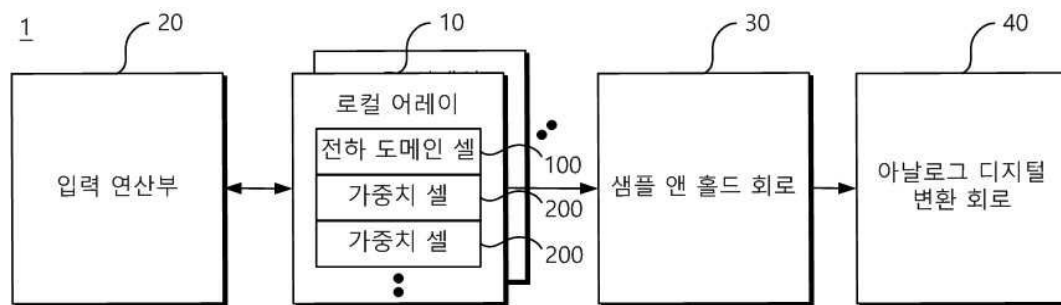
하게 설계 변경되거나 용이하게 변경할 수 있는 실시 예들도 포함될 것이다. 또한, 본 발명은 실시 예들을 이용하여 용이하게 변형하여 실시할 수 있는 기술들도 포함될 것이다. 따라서, 본 발명의 범위는 상술한 실시 예들에 국한되어 정해져서는 안되며, 후술하는 특허청구범위뿐만 아니라 이 발명의 특허청구범위와 균등한 것들에 의해 정해져야 할 것이다.

부호의 설명

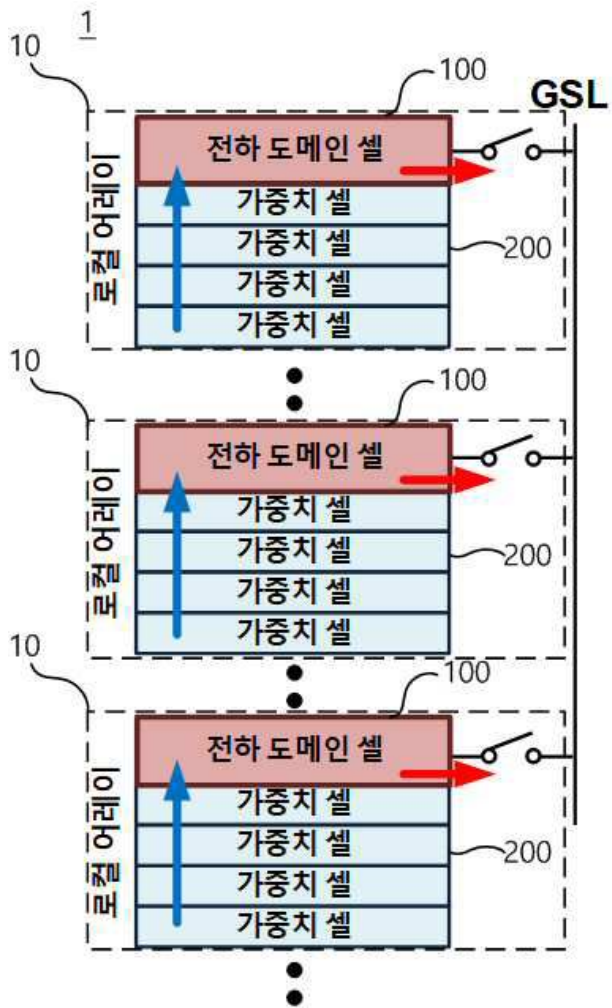
- [0066]
- 1 : PIM 장치
 - 10 : 로컬 어레이
 - 100 : 가중치 셀
 - 200 : 전하 도메인 셀
 - 20 : 입력 연산부
 - 30 : 샘플 앤 홀드 회로
 - 40 : 아날로그 디지털 변환 회로

도면

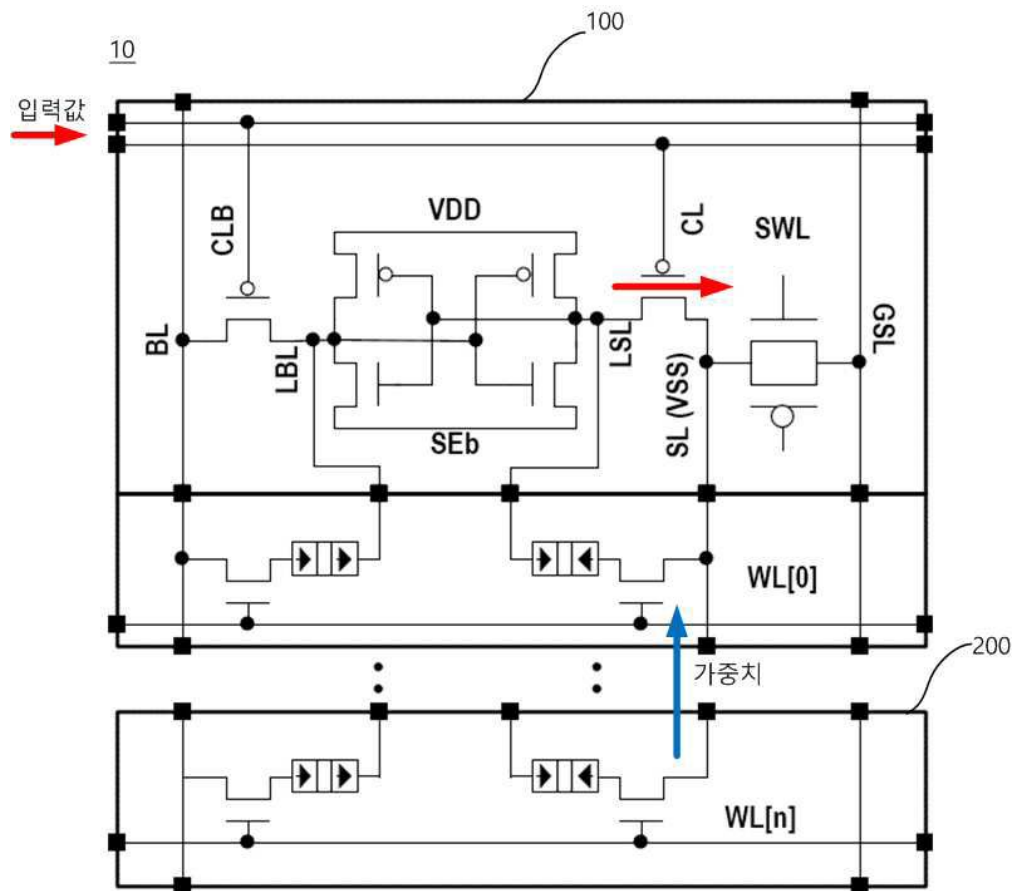
도면1



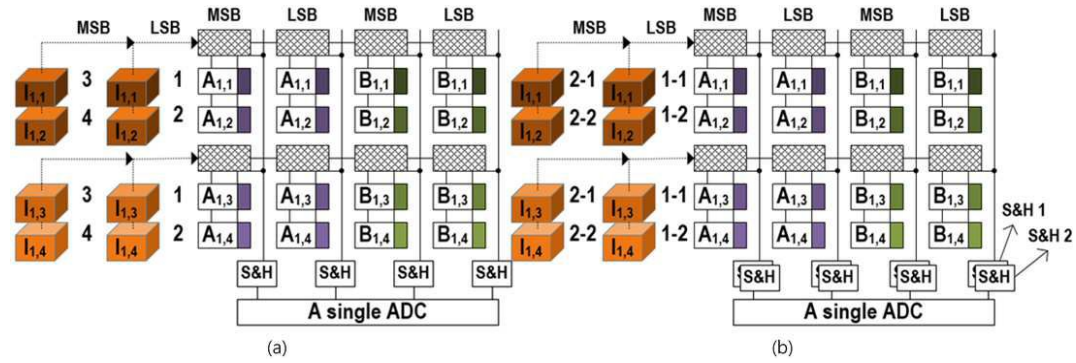
도면2a



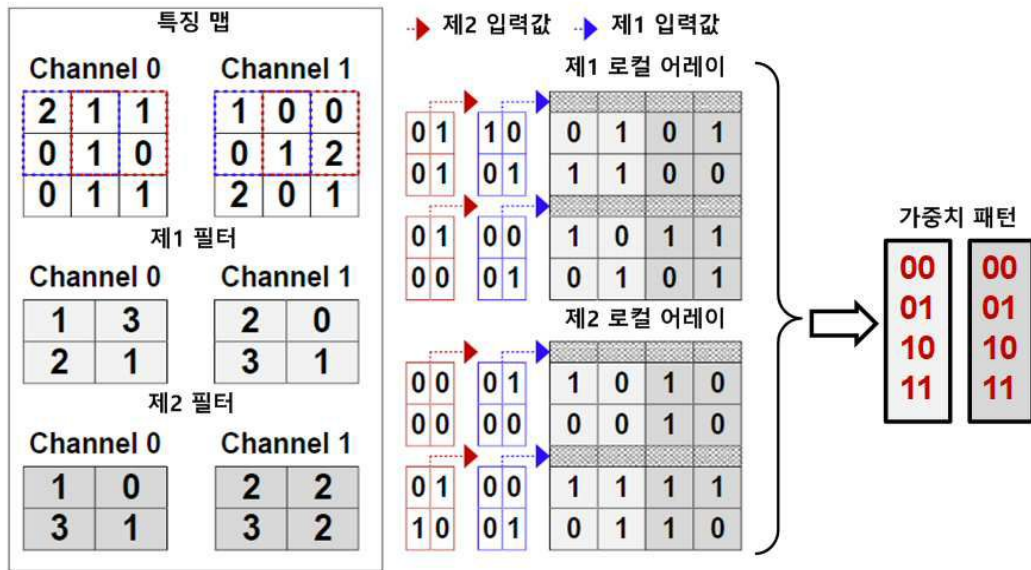
도면2b



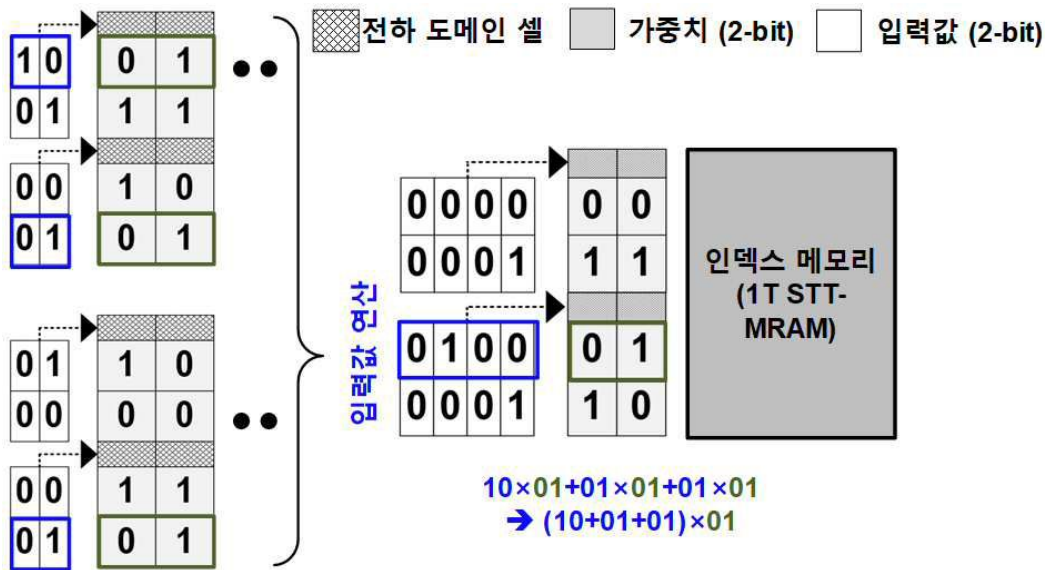
도면3



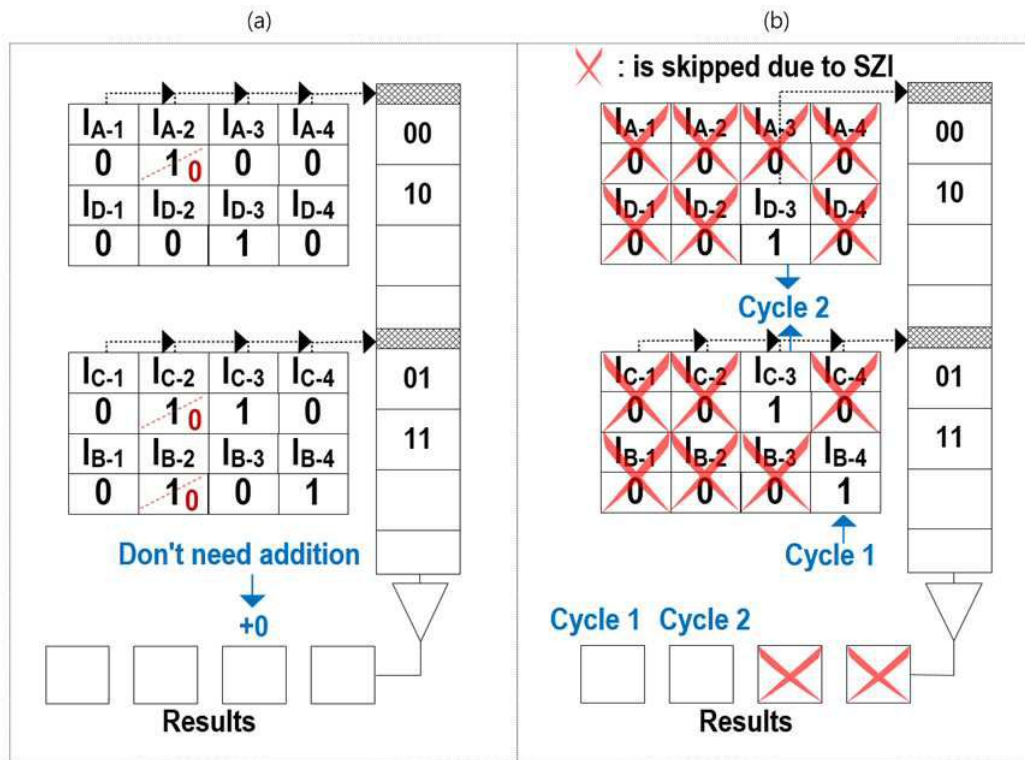
도면4a



도면4b



도면5



도면6

