



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0002159
(43) 공개일자 2024년01월04일

- (51) 국제특허분류(Int. Cl.)
G06F 15/78 (2006.01) G06F 7/544 (2017.01)
G06N 3/063 (2023.01) H03M 1/66 (2006.01)
- (52) CPC특허분류
G06F 15/7821 (2013.01)
G06F 7/5443 (2013.01)
- (21) 출원번호 10-2023-0037496
(22) 출원일자 2023년03월22일
심사청구일자 2023년03월22일
- (30) 우선권주장
1020220078594 2022년06월28일 대한민국(KR)

- (71) 출원인
고려대학교 산학협력단
서울특별시 성북구 안암로 145, 고려대학교 (안암동5가)
- (72) 발명자
박중선
서울특별시 서초구 신반포로 32
김준형
서울특별시 영등포구 영신로 237(당산동 4가, 당산롯데캐슬프레스티지)
이경호
서울특별시 성북구 고려대로11길 20(안암동2가)
- (74) 대리인
특허법인주연케이알피

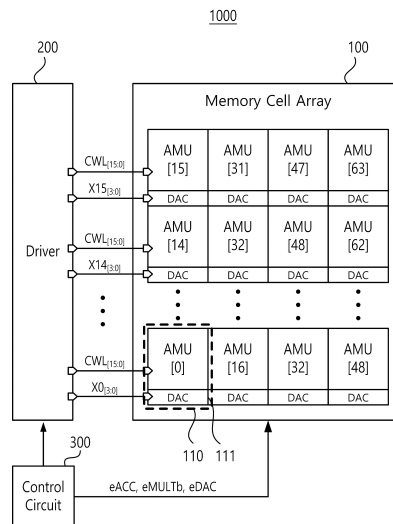
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 메모리 구조에 기반한 디지털-아날로그 변환기를 포함하는 컴퓨팅 인 메모리 장치

(57) 요약

컴퓨팅 인 메모리 장치가 개시된다. 본 컴퓨팅 인 메모리 장치는, 기저장된 가중치와, 다중 비트 입력 데이터에 대응하는 제 1 아날로그 전압에 대한 MAC(Multiply-Accumulate) 연산을 수행하는 아날로그 곱셈 유닛을 포함하는 메모리 셀 어레이, 및 다중 비트 입력 데이터를 아날로그 곱셈 유닛에 인가하는 드라이버를 포함하고, 아날로그 곱셈 유닛은, 다중 비트 입력 데이터를 제 1 아날로그 전압으로 변환하기 위한 디지털-아날로그 변환기를 포함한다.

대표도 - 도2



(52) CPC특허분류

G06N 3/063 (2013.01)

H03M 1/662 (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호 1711159051
과제번호 2020R1A2C3014820
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 개인기초연구(과기정통부)
연구과제명 온-디바이스 개인화를 위한 에너지 효율적인 딥러닝 가속기 설계
기 여 율 1/10
과제수행기관명 고려대학교 산학협력단
연구기간 2022.03.01 ~ 2023.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호 1425162856
과제번호 S3235750
부처명 중소벤처기업부
과제관리(전문)기관명 중소기업기술정보진흥원
연구사업명 틱스(창업성장기술개발사업 TIPS 과제)
연구과제명 후방 자동 긴급제동을 위한 광각카메라 기반 AI 인식 솔루션 개발
기 여 율 2/10
과제수행기관명 마인드포지(총괄주관연구개발기관), 고려대학교산학협력단(위탁기관)
연구기간 2022.03.01 ~ 2023.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호 1711173065
과제번호 2022M3I7A2079267
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 PIM인공지능반도체핵심기술개발(소자)
연구과제명 스핀케도토크 자성메모리 기반 PIM 소자
기 여 율 3/10
과제수행기관명 한국과학기술원
연구기간 2022.04.20 ~ 2023.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호 1711170645
과제번호 2022-0-00266-001
부처명 과학기술정보통신부
과제관리(전문)기관명 한국연구재단
연구사업명 PIM인공지능반도체핵심기술개발(설계)
연구과제명 초저전력 Low-Bit Precision 혼성모드 SRAM PIM 개발
기 여 율 4/10
과제수행기관명 한국과학기술원
연구기간 2022.04.01 ~ 2023.12.31

명세서

청구범위

청구항 1

기저장된 가중치와, 다중 비트 입력 데이터에 대응하는 제 1 아날로그 전압에 대한 MAC(Multiply-Accumulate) 연산을 수행하는 아날로그 곱셈 유닛을 포함하는 메모리 셀 어레이; 및

상기 다중 비트 입력 데이터를 상기 아날로그 곱셈 유닛에 인가하는 드라이버;를 포함하고,

상기 아날로그 곱셈 유닛은,

상기 다중 비트 입력 데이터를 상기 제 1 아날로그 전압으로 변환하기 위한 디지털-아날로그 변환기;를 포함하는, 컴퓨팅 인 메모리 장치.

청구항 2

제 1 항에 있어서,

상기 아날로그 곱셈 유닛은,

상기 다중 비트 입력 데이터의 비트 수에 대응하는 복수의 로컬 어레이들;을 포함하고,

상기 복수의 로컬 어레이들 각각은,

가중치를 저장하는 메모리 셀, 및 상기 디지털-아날로그 변환기를 구성하는 주변 회로를 포함하는, 컴퓨팅 인 메모리 장치.

청구항 3

제 2 항에 있어서,

상기 복수의 로컬 어레이들 중 상기 다중 비트 입력 데이터의 제 1 자릿수에 대응하는 로컬 어레이의 개수와 제 2 자릿수에 대응하는 로컬 어레이의 개수는 서로 다른, 컴퓨팅 인 메모리 장치.

청구항 4

제 3 항에 있어서,

상기 아날로그 곱셈 유닛은,

상기 다중 비트 입력 데이터의 각 자릿수에 대응하는 로컬 어레이의 주변 회로를 통해, 상기 다중 비트 입력 데이터의 각 자릿수에 대응하는 비트 값을 각각 인가받는, 컴퓨팅 인 메모리 장치.

청구항 5

제 4 항에 있어서,

상기 디지털-아날로그 변환기는,

서로 다른 자릿수에 속하는 주변 회로들이 전기적으로 분리된 상태에서 상기 다중 비트 입력 데이터가 인가되면, 상기 서로 다른 자릿수에 속하는 주변 회로들을 전기적으로 연결하여 상기 다중 비트 입력 데이터를 상기 제 1 아날로그 전압으로 변환하는, 컴퓨팅 인 메모리 장치.

청구항 6

제 5 항에 있어서,

상기 아날로그 곱셈 유닛은,

상기 드라이버로부터 상기 메모리 셀에 기저장된 가중치를 독출하기 위한 제어 신호가 인가되면, 상기 메모리

셀에 저장된 가중치와 상기 제 1 아날로그 전압의 곱셈 연산을 수행하는, 컴퓨팅 인 메모리 장치.

청구항 7

제 6 항에 있어서,

상기 아날로그 곱셈 유닛은,

상기 복수의 로컬 어레이 각각에서 수행된 상기 곱셈 연산 결과들을 누적 연산하고, 상기 누적 연산 결과를 제 2 아날로그 전압으로 출력하는, 컴퓨팅 인 메모리 장치.

청구항 8

제 7 항에 있어서,

상기 제 2 아날로그 전압을 디지털 값으로 변환하기 위한 아날로그-디지털 변환기;를 포함하는 컴퓨팅 인 메모리 장치.

청구항 9

제 6 항에 있어서,

상기 드라이버는,

연산 워드 라인을 통해 상기 아날로그 곱셈 유닛에 상기 제어 신호를 인가하고, 상기 연산 워드 라인과는 별도의 입력 라인을 통해 상기 디지털-아날로그 변환기에 상기 다중 비트 입력 데이터를 인가하는, 컴퓨팅 인 메모리 장치.

청구항 10

제 2 항에 있어서,

상기 메모리 셀은,

SRAM(Static Random Access Memory) 셀인, 컴퓨팅 인 메모리 장치.

발명의 설명

기술 분야

[0001] 본 발명은 메모리 구조에 기반한 디지털-아날로그 변환기를 포함하는 컴퓨팅 인 메모리 장치에 관한 것이다.

배경 기술

[0002] 기존의 컴퓨터 구조는 읽기/쓰기 동작을 지원하는 메모리와 데이터 연산을 지원하는 연산기가 분리되어 있다. 이때, 메모리와 연산기 사이에서 데이터가 이동할 때 소모되는 에너지에 비해, 연산기가 동작을 수행할 때 소모되는 에너지가 매우 크다. 이러한 특성 때문에, 데이터 이동이 적고 복잡한 연산이 필요한 기존의 어플리케이션에서는 기존의 컴퓨터 구조가 에너지 측면에서 매우 효율적이다.

[0003] 하지만, 인공 신경망의 합성곱 층에서 사용되는 곱셈-누적(Multiply-Accumulate, MAC) 연산은 막대한 양의 데이터의 이동을 필요로 하므로, 기존의 컴퓨터 구조는 에너지 측면에서 매우 비효율적이다.

[0004] 이를 해결하기 위해, 기존 읽기/쓰기 동작만이 가능한 메모리 안에 연산 동작을 추가하여 메모리와 연산기 간의 데이터 이동을 줄여주는 컴퓨팅-인-메모리 (CIM)라는 메모리 기술이 개발되고 있다.

[0005] 인공 신경망의 전체 연산 중 대부분을 차지하는 MAC 연산을 메모리에서 수행하기 위해서는 디지털 입력을 아날로그 값으로 변환하는 과정이 필요하다. 이를 위해 기존 CIM 기술에서는 일반적인 디지털-아날로그 변환기(Digital-Analog Converter, DAC)를 활용하였다.

발명의 내용

해결하려는 과제

[0006] 본 발명의 목적은, 메모리 내부에서 MAC 연산을 수행함으로써 DNN(Deep Neural Networks)과 같은 인공지능 연산에서 메모리와 연산기 간의 데이터 이동 횟수를 줄이고, 메모리 내부의 DAC를 이용하여 다중 비트 입력 데이터를 아날로그 전압으로 변환함으로써 MAC 연산의 정확도를 높일 수 있는 컴퓨팅 인 메모리 장치를 제공함에 있다.

과제의 해결 수단

[0007] 본 발명의 일 실시 예에 따른 컴퓨팅 인 메모리 장치는, 기저장된 가중치와, 다중 비트 입력 데이터에 대응하는 제 1 아날로그 전압에 대한 MAC(Multiply-Accumulate) 연산을 수행하는 아날로그 곱셈 유닛을 포함하는 메모리 셀 어레이, 및 상기 다중 비트 입력 데이터를 상기 아날로그 곱셈 유닛에 인가하는 드라이버를 포함하고, 상기 아날로그 곱셈 유닛은, 상기 다중 비트 입력 데이터를 상기 제 1 아날로그 전압으로 변환하기 위한 디지털-아날로그 변환기를 포함할 수 있다.

[0008] 또한, 상기 아날로그 곱셈 유닛은, 상기 다중 비트 입력 데이터의 비트 수에 대응하는 복수의 로컬 어레이들을 포함하고, 상기 복수의 로컬 어레이들 각각은, 가중치를 저장하는 메모리 셀, 및 상기 디지털-아날로그 변환기를 구성하는 주변 회로를 포함할 수 있다.

[0009] 또한, 상기 복수의 로컬 어레이들 중 상기 다중 비트 입력 데이터의 제 1 자릿수에 대응하는 로컬 어레이의 개수와 제 2 자릿수에 대응하는 로컬 어레이의 개수는 서로 다를 수 있다.

[0010] 또한, 상기 아날로그 곱셈 유닛은, 상기 다중 비트 입력 데이터의 각 자릿수에 대응하는 로컬 어레이의 주변 회로를 통해, 상기 다중 비트 입력 데이터의 각 자릿수에 대응하는 비트 값을 각각 인가받을 수 있다.

[0011] 또한, 상기 디지털-아날로그 변환기는, 서로 다른 자릿수에 속하는 주변 회로들이 전기적으로 분리된 상태에서 상기 다중 비트 입력 데이터가 인가되면, 상기 서로 다른 자릿수에 속하는 주변 회로들을 전기적으로 연결하여 상기 다중 비트 입력 데이터를 상기 제 1 아날로그 전압으로 변환할 수 있다.

[0012] 또한, 상기 아날로그 곱셈 유닛은, 상기 드라이버로부터 상기 메모리 셀에 기저장된 가중치를 독출하기 위한 제어 신호가 인가되면, 상기 메모리 셀에 기저장된 가중치와 상기 제 1 아날로그 전압의 곱셈 연산을 수행할 수 있다.

[0013] 또한, 상기 아날로그 곱셈 유닛은, 상기 복수의 로컬 어레이 각각에서 수행된 상기 곱셈 연산 결과들을 누적 연산하고, 상기 누적 연산 결과를 제 2 아날로그 전압으로 출력할 수 있다.

[0014] 또한, 상기 제 2 아날로그 전압을 디지털 값으로 변환하기 위한 아날로그-디지털 변환기를 포함할 수 있다.

[0015] 또한, 상기 드라이버는, 연산 워드 라인을 통해 상기 아날로그 곱셈 유닛에 상기 제어 신호를 인가하고, 상기 연산 워드 라인과는 별도의 입력 라인을 통해 상기 디지털-아날로그 변환기에 상기 다중 비트 입력 데이터를 인가할 수 있다.

[0016] 또한, 상기 메모리 셀은, SRAM(Static Random Access Memory) 셀일 수 있다.

발명의 효과

[0017] 상술한 본 발명의 다양한 실시 예들에 따르면, 메모리 내부에서 MAC 연산을 수행함으로써 DNN(Deep Neural Networks)과 같은 인공지능 연산에서 메모리와 연산기 간의 데이터 이동 횟수를 줄일 수 있다. 이에 따라, 인공지능 연산의 에너지 효율을 높일 수 있다. 또한, 다중 비트 입력 데이터를 아날로그 전압으로 변환할 때 메모리 내부의 DAC를 이용함으로써 공정 수율에 이점을 갖고, MAC 연산의 정확도를 높일 수 있다. 또한, 4비트 이상의 입력 데이터 사용이 가능해지므로, 1 비트나 2 비트 입력 데이터에 대한 연산보다 더 높은 인공지능 연산 정확도를 얻을 수 있다.

도면의 간단한 설명

[0018] 도 1a는 종래 컴퓨팅 인 메모리 기술을 설명하기 위한 도면이다.

도 1b는 본 발명의 일 실시 예에 따른 컴퓨팅 인 메모리 장치의 블록도이다.

도 2는 본 발명의 일 실시 예에 따른 컴퓨팅 인 메모리 장치의 구현 예를 도시한 블록도이다.

도 3은 본 발명의 일 실시 예에 따른 아날로그 곱셈 유닛의 구성을 설명하기 위한 도면이다.

도 4a는 본 발명의 일 실시 예에 따른 로컬 어레이의 구성을 도시한 도면이다.

도 4b는 본 발명의 일 실시 예에 따른 주변 회로의 회로도이다.

도 4c는 본 발명의 일 실시 예에 따른 주변 회로의 회로도이다.

도 5는 본 발명의 일 실시 예에 따른 아날로그 곱셈 유닛의 디지털-아날로그 변환 동작을 설명하기 위한 도면이다.

도 6a는 본 발명의 일 실시 예에 따른 아날로그 곱셈 유닛의 디지털-아날로그 변환 동작을 설명하기 위한 도면이다.

도 6b는 본 발명의 일 실시 예에 따른 아날로그 곱셈 유닛의 디지털-아날로그 변환 동작을 설명하기 위한 도면이다.

도 7a는 본 발명의 일 실시 예에 따른 아날로그 곱셈 유닛의 곱셈 연산 동작을 설명하기 위한 도면이다.

도 7b는 본 발명의 일 실시 예에 따른 아날로그 곱셈 유닛의 곱셈 연산 동작을 설명하기 위한 도면이다.

도 8a는 본 발명의 일 실시 예에 따른 아날로그 곱셈 유닛의 누적 연산 동작을 설명하기 위한 도면이다.

도 8b는 본 발명의 일 실시 예에 따른 아날로그 곱셈 유닛의 누적 연산 동작을 설명하기 위한 도면이다.

도 9는 본 발명의 일 실시 예에 따른 컴퓨팅 인 메모리 장치의 각종 신호의 파형을 도시한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0019] 이하 도면을 참조하여 설명되는 본 발명의 다양한 실시 예들은 특정한 실시 형태에 대해 범위를 한정하려는 것이 아니며, 다양한 변경(modifications), 균등물(equivalents), 및/또는 대체물(alternatives)을 포함하는 것으로 이해되어야 한다. 도면의 설명과 관련하여, 유사한 구성요소에 대해서는 유사한 참조 부호가 사용될 수 있다.
- [0020] 본 발명을 설명함에 있어, 관련된 공지 기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우 그 상세한 설명을 생략한다. 또한, 동일한 구성의 중복 설명은 되도록 생략하기로 한다.
- [0021] 이하에서 사용된 용어는 실시 예를 설명하기 위해 사용된 것으로, 본 발명을 제한 및/또는 한정하려는 의도가 아니다. 단수의 표현은 문맥상 명백하게 다르게 뜻하지 않는 한, 복수의 표현을 포함한다.
- [0022] 이하에서, '포함하다' 또는 '가지다' 등의 용어는 명세서상에 기재된 특징, 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것이 존재함을 지정하려는 것이지, 하나 또는 그 이상의 다른 특징들이나 숫자, 단계, 동작, 구성요소, 부품 또는 이들을 조합한 것들의 존재 또는 부가 가능성을 미리 배제하지 않는 것으로 이해되어야 한다.
- [0023] 이하에서 사용된 "제1," "제2," "첫째," 또는 "둘째," 등의 표현들은 다양한 구성요소들을, 순서 및/또는 중요도에 상관없이 수식할 수 있고, 한 구성요소를 다른 구성요소와 구분하기 위해 사용될 뿐 해당 구성요소들을 한정하지 않는다.
- [0024] 어떤 구성요소(예: 제1 구성요소)가 다른 구성요소(예: 제2 구성요소)에 "(기능적으로 또는 통신적으로) 연결되어(operatively or communicatively) coupled with/to)" 있다거나 "접속되어(connected to)" 있다고 언급된 때에는, 상기 어떤 구성요소가 상기 다른 구성요소에 직접적으로 연결되거나, 다른 구성요소(예: 제3 구성요소)를 통하여 연결될 수 있다고 이해되어야 할 것이다. 반면에, 어떤 구성요소(예: 제1 다른 구성요소(예: 제2 구성요소)에 "직접 연결되어" 있다거나 "직접 접속되어" 있다고 언급된 때에는, 상기 어떤 구성요소와 상기 다른 구성요소 사이에 다른 구성요소(예: 제 3 구성요소)가 존재하지 않는 것으로 이해될 수 있다.
- [0025] 본 발명의 실시 예들에서 사용되는 용어들은 다르게 정의되지 않는 한, 해당 기술 분야에서 통상의 지식을 가진 자에게 통상적으로 알려진 의미로 해석될 수 있다.
- [0026] 이하에서 첨부된 도면을 참조하여 본 발명의 다양한 실시 예를 상세히 설명한다.
- [0027] 도 1a는 종래의 컴퓨팅 인 메모리 기술을 설명하기 위한 도면이다.
- [0028] 디지털 회로인 메모리를 이용하여 MAC 연산과 같은 아날로그 연산을 하기 위해서는, 디지털 입력을 아날로그 값

으로 변환하는 디지털-아날로그 컨버터(DAC)와, MAC 연산 결과를 디지털로 출력하기 위한 아날로그-디지털 컨버터(ADC)가 필요하다.

- [0029] 도 1a를 참조하면, 디지털 값인 입력값($i.Act$)이 DAC로 입력되면, DAC는 입력값($i.Act$)을 아날로그 전압으로 변환한다. 아날로그 값으로 변환된 입력값($i.Act$)은 워드 라인(WL)을 통해 SRAM(Static Random Access Memory) 어레이의 각 셀로 입력된다.
- [0030] 이에 따라, 각 셀에 저장된 인공지능 모델의 웨이트값(W)과 아날로그 전압으로 변환된 입력값($i.Act$)에 대한 곱셈 연산이 수행되며, 각 셀에서 수행된 곱셈 연산 결과는 비트 라인(BL) 별로 누적되어 아날로그 전압인 V_{MAC} 으로 출력되게 된다. 한편, MAC 연산의 결과값인 V_{MAC} 은 ADC로 입력되어 다시 디지털 값으로 변환되며, 인공지능 연산에 사용되게 된다.
- [0031] 이때, 종래 컴퓨팅 인 메모리 기술에서는, 디지털 입력을 아날로그 값으로 변환하기 위해 도 1a에 도시된 바와 같은 일반적인 DAC가 이용되었다.
- [0032] 일반적으로 입력의 비트 수가 높아질수록 인공지능 연산의 정확도는 높아지게 된다. 하지만, 그에 따라 면적, 복잡도, 에너지 등과 같은 측면에서 DAC의 오버헤드가 증가하게 된다. 따라서, 도 1a에 도시된 바와 같이 일반적인 DAC를 활용하여 컴퓨팅 인 메모리 하드웨어를 구성하는 경우, 인공지능 연산의 정확도를 높이는 데 한계가 있다.
- [0033] 도 1b는 본 발명의 일 실시 예에 따른 컴퓨팅 인 메모리 장치의 블록도이다.
- [0034] 도 1b에 따르면, 컴퓨팅 인 메모리 장치(1000)는 메모리 셀 어레이(100) 및 드라이버(200)를 포함할 수 있다.
- [0035] 드라이버(200)는 메모리 셀 어레이(100)를 구동한다. 구체적으로, 드라이버(200)는 메모리 셀 어레이(100)로 다중 비트 입력 데이터(X_0 내지 X_{M-1})를 인가할 수 있다. 이때, 다중 비트 입력 데이터는 디지털 입력이다.
- [0036] 또한, 드라이버(200)는 메모리 셀 어레이(100)로 메모리 셀 어레이(100)의 각 셀에 기저장된 가중치를 독출하기 위한 제어 신호(CWL)를 인가할 수 있다. 이에 따라, 메모리 셀 어레이(100)의 각 셀에서는 입력 데이터에 대응하는 아날로그 전압과 독출된 가중치의 곱셈 연산이 이루어질 수 있다.
- [0037] 메모리 셀 어레이(100)는 복수의 아날로그 곱셈 유닛(Analog Multiply Unit(AMU), 110)을 포함할 수 있다. 이때, 각 아날로그 곱셈 유닛(110)은 기저장된 가중치와, 다중 비트 입력 데이터에 대응하는 아날로그 전압에 대한 MAC(Multiply-Accumulate) 연산을 수행할 수 있다.
- [0038] 이를 위해, 아날로그 곱셈 유닛(110)은 인공지능 연산을 위해 필요한 가중치를 저장하는 메모리 셀들, 및 디지털 입력을 아날로그 전압으로 변환하기 위한 디지털-아날로그 변환기(Digital Analog Converter(DAC), 111)를 포함할 수 있다. 이때, DAC(111)는 메모리 셀들의 구조를 활용하여 메모리 셀 어레이(100) 내부에 구현될 수 있다. 이에 관한 구체적인 내용은 후술된다.
- [0039] 구체적으로, 드라이버(200)로부터 다중 비트 입력 데이터가 인가되면, DAC(111)는 인가된 다중 비트 입력 데이터를 아날로그 전압으로 변환할 수 있다. 이때, 아날로그 전압의 크기는 인가된 다중 비트 입력 데이터에 대응할 수 있다.
- [0040] 이후 드라이버(200)로부터 가중치를 독출하기 위한 제어 신호(CWL)가 인가되면, AMU(110)는 메모리 셀에 기저장된 가중치를 독출하고, 독출된 가중치와 DAC(111)에서 생성된 아날로그 전압에 대한 곱셈 연산을 수행할 수 있다.
- [0041] 이와 같은 곱셈 연산은 AMU(110)에 포함된 메모리 셀들에서 동시에 수행될 수 있으며, AMU(110)는 각 셀에서 수행된 곱셈 연산 결과들을 누적 연산하여 MAC 연산 결과를 출력할 수 있다.
- [0042] 이와 같이, 본 발명의 일 실시 예에 따르면, 메모리 셀 어레이(100) 내부에 메모리 구조를 활용한 DAC(111)를 구현하고, 이를 이용하여 다중 비트의 디지털 입력을 아날로그 값으로 변환할 수 있다. 이에 따라, 컴퓨팅 인 메모리 장치(1000)의 공정 수율을 높일 수 있다. 또한, DAC(111)의 입, 출력 간에 높은 선형성이 확보되어 MAC 연산의 정확도를 높일 수 있다. 또한, 4비트 이상의 입력 데이터 사용이 가능해지므로, 1비트나 2비트 입력 데이터에 대한 연산보다 더 높은 인공지능 연산 정확도를 얻을 수 있다.
- [0043] 도 2는 본 발명의 일 실시 예에 따른 컴퓨팅 인 메모리 장치의 구현 예를 도시한 블록도이다.

- [0044] 도 2에 따르면, 컴퓨팅 인 메모리 장치(1000)는 드라이버(200), 메모리 셀 어레이(100) 및 제어 회로(300)를 포함할 수 있다.
- [0045] 메모리 셀 어레이(100)는 복수의 AMU(110)들을 포함할 수 있다. 도 2에서는 AMU[0]부터 AMU[63]까지 64개의 AMU(110)가 메모리 셀 어레이(100)에 포함된 예를 도시하였으나, 메모리 셀 어레이(100)에 포함될 수 있는 AMU(110)의 개수가 이에 한정되는 것은 아니다. 일 실시 예에 따르면, 메모리 셀 어레이(100)는 SRAM(Static Random Access Memory) 메모리 셀 어레이일 수 있으나, 이에 한정되는 것은 아니다.
- [0046] AMU(110)는 MAC 연산을 위해 필요한 가중치를 저장하는 메모리 셀들 및 디지털 입력 데이터를 아날로그 전압으로 변환하기 위한 DAC(111)를 포함할 수 있다. 이에 따라, AMU(110)는 메모리 셀에 기저장된 가중치와, 다중 비트 입력 데이터에 대응하는 아날로그 전압에 대한 MAC 연산을 수행할 수 있다.
- [0047] 드라이버(200)는 메모리 셀 어레이(100)를 구동할 수 있다. 특히, 드라이버(200)는 다중 비트 입력 데이터(X0 내지 X15), 및 가중치를 독출하기 위한 제어 신호(CWL)를, 메모리 셀 어레이(100)의 AMU(110)들로 각각 인가하여 메모리 셀 어레이(100)를 구동할 수 있다.
- [0048] 이때, 드라이버(200)는 입력 라인을 통해 AMU(110)에 포함된 DAC(111)로 다중 비트 입력 데이터를 인가할 수 있다. 또한, 드라이버(200)는 입력 라인과는 별도의 연산 워드 라인(Computation Word Line, CWL)을 통해 AMU(110)에 포함된 메모리 셀로 제어 신호(CWL)를 인가할 수 있다.
- [0049] 도 2를 참조하면, 드라이버(200)는 입력 라인을 통해 4비트 입력 데이터(X0_[3:0] 내지 X15_[3:0])를 각 AMU(110)의 DAC(111)로 인가할 수 있다. 또한, 드라이버(200)는 연산 워드 라인을 통해 16비트의 제어 신호 CWL_[15:0]을 각 AMU(110)의 메모리 셀들에 인가할 수 있다.
- [0050] 이에 따라, 4비트 입력 데이터는 DAC(111)를 통해 대응되는 아날로그 전압으로 변환될 수 있다. 또한, 4비트 입력 데이터에 대응하는 아날로그 전압과 기저장된 가중치의 곱셈 연산이 16개의 메모리 셀들 각각에서 동시에 수행될 수 있다. 각 메모리 셀에서 수행된 곱셈 연산 결과는 누적 연산된 후 MAC 연산 결과로 출력될 수 있다.
- [0051] 제어 회로(300)는 컴퓨팅 인 메모리 장치(1000)의 전반적인 동작을 제어할 수 있다. 특히, 제어 회로(300)는 컴퓨팅 인 메모리 장치(1000)가 전술한 MAC 연산 동작을 수행하도록 드라이버(200) 및 메모리 셀 어레이(100)의 동작을 제어할 수 있다.
- [0052] 예를 들어, 제어 회로(300)는 드라이버(200)의 동작을 제어하여, 다중 비트 입력 데이터(X0 내지 X15)가 인가되는 타이밍과, 가중치를 독출하기 위한 제어 신호(CWL)가 인가되는 타이밍을 조절할 수 있으나, 이에 한정되는 것은 아니다.
- [0053] 또한, 제어 회로(300)는, 전술한 AMU(110)의 MAC 연산 동작을 위해, PCHb, eACC, eMULTb, eDAC와 같은 각종 제어 신호를 메모리 셀 어레이(100)로 인가할 수 있다. 제어 회로(300)가 인가하는 각종 제어 신호에 관한 구체적인 내용은 후술된다.
- [0054] 이하에서는, 도 3 내지 도 4c를 참조하여, 본 발명의 일 실시 예에 따른 AMU(110)의 구성을 설명한다.
- [0055] 도 3은 본 발명의 일 실시 예에 따른 아날로그 곱셈 유닛의 구성을 설명하기 위한 도면이다.
- [0056] 일 실시 예에 따르면, AMU(110)는 다중 비트 입력 데이터의 비트 수에 대응하는 복수의 로컬 어레이들을 포함할 수 있다. 이때, 복수의 로컬 어레이들 각각은, 가중치를 저장하는 메모리 셀, 및 디지털-아날로그 변환기(111)를 구성하는 주변 회로를 포함할 수 있다.
- [0057] 도 3은 4비트 입력 데이터 X_[3:0]에 대응하는 16개의 로컬 어레이들(LA_[0] 내지 LA_[15])이 행 방향으로 이어진 AMU(110)의 일 예를 도시하고 있다. 이때, LA_[0] 내지 LA_[7]는 4비트 입력 데이터의 첫번째 자릿수에 대응하는 로컬 어레이들일 수 있다. LA_[8] 내지 LA_[11]는 4비트 입력 데이터의 두번째 자릿수에 대응하는 로컬 어레이들일 수 있다. LA_[12] 및 LA_[13]은 4비트 입력 데이터의 세번째 자릿수에 대응하는 로컬 어레이들일 수 있다. LA_[14]는 4비트 입력 데이터의 네번째 자릿수에 대응하는 로컬 어레이일 수 있다. LA_[15]는 디폴트 로컬 어레이일 수 있다.
- [0058] 도 3을 참조하면, 16개의 로컬 어레이들(LA_[0] 내지 LA_[15]) 각각은 제어 신호 CWL_[15:0]가 인가되는 메모리 셀과 4비트 입력 데이터 X_[3:0]가 입력되는 주변 회로를 포함하는 것을 볼 수 있다. 이때, 각 로컬 어레이의 주변 회로

는 A 타입 또는 B 타입일 수 있다. A 타입 주변 회로는 입력 비트 라인(input Bit Line, iBL)을 통해 인접한 주변 회로들과 전기적으로 연결되어 전하를 공유할 수 있다. B 타입 주변 회로는 패스 게이트(PG_{DAC})를 포함하며, PG_{DAC}의 온/오프에 따라 인접한 주변 회로들을 전기적으로 연결하거나 전기적으로 분리할 수 있다.

[0059] 보다 구체적으로, 도 3을 참조하면, LA_[7], LA_[11], LA_[13], LA_[14]의 주변 회로는 B 타입이고, 나머지 LA_[0] 내지 LA_[6], LA_[8] 내지 LA_[10], LA_[12], LA_[15]의 주변 회로는 A 타입이다. 따라서, B 타입 주변 회로의 PG_{DAC}가 오프된 상태에서 iBL_[3], iBL_[2], iBL_[1], iBL_[0]의 캐패시턴스 비율은 8 : 4 : 2 : 1이 된다.

[0060] 한편, 로컬 어레이들(LA_[0] 내지 LA_[15])의 주변 회로들이 전술한 DAC(111)를 구성할 수 있다. 따라서, DAC(111)는 iBL_[3], iBL_[2], iBL_[1], iBL_[0]의 캐패시턴스 비율을 이용하여 4비트 입력 데이터(X_[3:0])를 아날로그 전압으로 변환할 수 있다.

[0061] 한편, 각 로컬 어레이의 메모리 셀에서는 아날로그 전압으로 변환된 입력 데이터와 가중치의 곱셈 연산이 이루어질 수 있다.

[0062] 도 4a는 본 발명의 일 실시 예에 따른 로컬 어레이의 구성을 도시한 도면이다. 도 4a에 도시된 BL은 비트 라인을, BLB는 상보 비트 라인을, CBL은 연산 비트 라인을, ABL은 누적 비트 라인을, CWL은 연산 워드 라인을 나타낸다.

[0063] 본 발명의 일 실시 예에 따르면, 로컬 어레이의 메모리 셀은 P-8T SRAM 셀 구조를 가질 수 있다. 도 4a를 참조하면, P-8T SRAM 셀 구조는, 읽기/쓰기 동작만 가능한 기존 6T SRAM 구조에 2개의 PMOS 트랜지스터(P0, P1)가 도시된 바와 같이 추가적으로 연결된 구조일 수 있다. 이와 같은 구조를 통해, CBL에서 곱셈 연산이 수행될 수 있다.

[0064] 한편, 로컬 어레이는, DAC(111)를 구성하는 주변 회로를 포함할 수 있다. 이때, 주변 회로는 전술한 바와 같이, A 타입과 B 타입 두가지가 있다.

[0065] 도 4b는 A 타입 주변 회로의 일 예를 도시하고 있다. 도 4b를 참조하면, A 타입 주변 회로는 입력 비트 라인(input Bit Line, iBL)을 통해 인접한 주변 회로와 전기적으로 연결되어 서로 전하를 공유할 수 있다.

[0066] 도 4c는 B 타입 주변 회로의 일 예를 도시하고 있다. 도 4c를 참조하면, B 타입 주변 회로는, 패스 게이트(PG_{DAC})를 더 포함하는 것을 제외하고, 도 4b의 A 타입 주변 회로와 동일한 구성을 가질 수 있다. PG_{DAC}는 제어 신호 eDAC에 따라 온/오프될 수 있다. PG_{DAC}가 온된 상태에서 B 타입 주변 회로는 A 타입 주변 회로와 동일한 구성이 되며, iBL을 통해 인접한 주변 회로와 서로 전하를 공유할 수 있다. 한편, PG_{DAC}가 오프된 상태에서 B 타입 주변 회로는, 인접한 주변 회로들 중 다른 자릿수에 속하는 주변 회로와 전기적으로 분리되며, 다른 자릿수에 속하는 주변 회로와의 전하 공유가 차단될 수 있다.

[0067] 한편, eACC, eMULTb, eDAC와 같은 제어 신호에 따른 주변 회로의 구체적인 동작은 후술 된다.

[0068] 이하에서는, 도 5 내지 도 6b를 참조하여, 본 발명의 일 실시 예에 따른 AMU(110)의 디지털-아날로그 변환 동작을 설명한다.

[0069] 도 5는 본 발명의 일 실시 예에 따른 아날로그 곱셈 유닛의 디지털-아날로그 변환 동작을 설명하기 위한 도면이다. 구체적으로, 도 5는 도 3의 AMU(110)에 4비트 입력 데이터(X_[3:0])가 인가된 상태를 도시하고 있다.

[0070] 도 5를 참조하면, CBL[0] 내지 CBL[7]은 iBL_[3]에 연결되고, CBL[8] 내지 CBL[11]은 iBL_[2]에 연결되고, CBL[12] 및 CBL[13]은 iBL_[1]에 연결되고, CBL[14]는 iBL_[0]에 연결된다. 한편, N15의 게이트에는 항상 그라운드 전압(VSS)이 인가되고 있으므로, CBL[15]는 항상 VDD로 유지된다.

[0071] 일 실시 예에 따르면, 4비트 입력 데이터(X_[3:0])가 인가되기 전에, CBL[0] 내지 CBL[15]는 먼저 구동 전압(VDD)으로 초기화될 수 있다. 이때, LA_[7], LA_[11], LA_[13], LA_[14]의 주변 회로에 포함된 PG_{DAC}는 오프된 상태를 유지할 수 있다.

[0072] 이후, 4비트 입력 데이터(X_[3:0])가 드라이버(200)로부터 AMU(110)로 인가될 수 있다. 이때, AMU(110)는 4비트 입력 데이터의 각 자릿수에 대응하는 로컬 어레이의 주변 회로를 통해, 4비트 입력 데이터의 각 자릿수에 대응

하는 비트 값을 각각 인가 받을 수 있다. 도 5를 참조하면, 4비트 입력 데이터($X_{[3:0]}$) 중 $X_{[3]}$ 은 $LA_{[0]}$ 내지 $LA_{[7]}$ 의 주변 회로들을 통해, $X_{[2]}$ 은 $LA_{[8]}$ 내지 $LA_{[11]}$ 의 주변 회로들을 통해, $X_{[1]}$ 은 $LA_{[12]}$ 및 $LA_{[13]}$ 의 주변 회로들을 통해, $X_{[0]}$ 은 $LA_{[14]}$ 의 주변 회로를 통해 각각 인가되는 것을 볼 수 있다.

[0073] 구체적으로, $X_{[3]}$ 가 1이면, N_0 내지 N_7 는 온되며, $iBL_{[3]}$ 은 그라운드 전압(VSS)으로 방전된다. $X_{[3]}$ 가 0이면, N_0 내지 N_7 는 오프된 상태를 유지하며, $iBL_{[3]}$ 은 프리 차지된 구동 전압(VDD)을 유지한다.

[0074] 한편, $X_{[2]}$ 가 1이면, N_8 내지 N_{11} 는 온되며, $iBL_{[2]}$ 은 그라운드 전압(VSS)으로 방전된다. $X_{[2]}$ 가 0이면, N_8 내지 N_{11} 는 오프된 상태를 유지하며, $iBL_{[2]}$ 은 구동 전압(VDD)을 유지한다.

[0075] 한편, $X_{[1]}$ 가 1이면, N_{12} 및 N_{13} 는 온되며, $iBL_{[1]}$ 은 그라운드 전압(VSS)으로 방전된다. $X_{[1]}$ 가 0이면, N_{12} 내지 N_{13} 는 오프된 상태를 유지하며, $iBL_{[1]}$ 은 구동 전압(VDD)을 유지한다.

[0076] 한편, $X_{[0]}$ 가 1이면, N_{14} 는 온되며, $iBL_{[0]}$ 은 그라운드 전압(VSS)으로 방전된다. $X_{[0]}$ 가 0이면, N_{14} 는 오프된 상태를 유지하며, $iBL_{[0]}$ 은 구동 전압(VDD)을 유지한다.

[0077] 4비트 입력 데이터($X_{[3:0]}$)가 인가되는 동안에도 $LA_{[7]}$, $LA_{[11]}$, $LA_{[13]}$, $LA_{[14]}$ 의 주변 회로에 포함된 PG_{DAC} 는 오프된 상태를 유지할 수 있다.

[0078] 따라서, 예를 들어, $X_{[3:0]}$ 이 1000₍₂₎인 경우, 도 5에 도시된 바와 같이, $iBL_{[3]}$ 는 VSS가 되고, $iBL_{[2]}$ 는 VDD가 되고, $iBL_{[1]}$ 은 VDD가 되고, $iBL_{[0]}$ 는 VDD가 될 수 있다.

[0079] 일 실시 예에 따르면, 다중 비트 입력 데이터의 각 자릿수에 대응하는 로컬 어레이의 개수는 자릿수 마다 서로 다를 수 있다. 도 5를 참조하면, 4비트 입력 데이터($X_{[3:0]}$)의 각 자릿수($X_{[3]}$, $X_{[2]}$, $X_{[1]}$, $X_{[0]}$)에 대응하는 로컬 레이어의 개수는, 8개, 4개, 2개, 1개로 자릿수에 따라 서로 다른 것을 확인할 수 있다.

[0080] 도 6a는 도 5와 같이 4비트 입력 데이터가 인가된 상태에서, $LA_{[7]}$, $LA_{[11]}$, $LA_{[13]}$, $LA_{[14]}$ 의 주변 회로에 포함된 PG_{DAC} 가 제어 신호 eDAC를 통해 온된 경우, DAC(111)의 동작을 도시하고 있다.

[0081] 도 6a를 참조하면, 4비트 입력 데이터가 인가된 상태에서 $LA_{[7]}$, $LA_{[11]}$, $LA_{[13]}$, $LA_{[14]}$ 의 주변 회로에 포함된 PG_{DAC} 가 온되면, $iBL_{[3]}$, $iBL_{[2]}$, $iBL_{[1]}$, $iBL_{[0]}$ 를 통해 행 방향으로 전하가 공유되어 4비트 입력 데이터에 대응하는 하나의 아날로그 전압이 CBL[0] 내지 CBL[15]에 인가될 수 있다.

[0082] 일 실시 예에 따르면, 4비트 입력 데이터와 아날로그 전압은 아래 수학적 식 1과 같은 관계를 가질 수 있다.

수학적 식 1

$$V_{DAC} = \left(\sum_{N=0}^3 (2^N \cdot \overline{X_{[N]}}) + 1 \right) \cdot \frac{VDD}{16}$$

[0083]

[0084] 여기서, V_{DAC} 는 CBL[0] 내지 CBL[15]에 인가되는 아날로그 전압 V_{CBL} 을 나타내고, N은 4비트 입력 데이터의 각 비트의 자릿수를 2^N 으로 표현할 때 N값을 나타내고, $\overline{X_{[N]}}$ 는 4비트 입력 데이터의 2^N 자릿수에 해당하는 비트 값의 반대값을 나타내고, VDD는 AMU(110)의 구동 전압을 나타낸다. 여기서, 1의 반대값은 0이고, 0의 반대값은 1이다.

[0085] 한편, 도 6b는 상기 수학적 식 1에 따른 계산 결과를 도시한 표이다. 도 6b에서 4-bit i.Act $X_{[3:0]}$ 는 4비트 입력 데이터를, 16 level Analog V_{CBL} 는 각 입력 데이터에 대응하는 16개의 아날로그 전압을 나타낸다.

[0086] 예를 들어, 4비트 입력 데이터가 0001인 경우, V_{DAC} 는 $\{((2^0 * 0) + (2^1 * 1) + (2^2 * 1) + (2^3 * 1)) + 1\} * (VDD/16)$ 이 되어,

(15/16)*VDD가 되며, 이는 도 6b의 결과와 일치함을 확인할 수 있다.

- [0087] 도 5 내지 도 6b에서 설명한 바와 같이, 로컬 어레이들(LA_[0] 내지 LA_[15])의 주변 회로들로 구성된 DAC(111)는, 서로 다른 자릿수에 속하는 주변 회로들이 전기적으로 분리된 상태에서 다중 비트 입력 데이터가 인가되면, 서로 다른 자릿수에 속하는 주변 회로들을 전기적으로 연결하여 다중 비트 입력 데이터를 아날로그 전압(V_{DAC})으로 변환할 수 있다.
- [0088] 도 7a는 본 발명의 일 실시 예에 따른 아날로그 곱셈 유닛의 곱셈 연산 동작을 설명하기 위한 도면이다. 일 실시 예에 따르면, AMU(110)는 드라이버(200)로부터 메모리 셀에 기저장된 가중치를 독출하기 위한 제어 신호(CWL)가 인가되면, 메모리 셀에 기저장된 가중치와, 아날로그 전압의 곱셈 연산을 수행할 수 있다.
- [0089] 도 7a를 참조하면, 제어 신호 eMULTb에 따라 주변 회로의 iBL과 메모리 셀의 연결이 차단되고, CWL의 전압이 VSS가 되면서, 아날로그 레벨로 변환된 4비트 입력 데이터와 메모리 셀 내부에 저장되어 있던 가중치의 곱셈 연산이 이루어지는 것을 확인할 수 있다.
- [0090] 이때, 메모리 셀에 기저장된 가중치 W[i]는 0 또는 1의 값을 가질 수 있다. 따라서, W[i]가 0인 경우, P0 및 P1은 모두 온되며, CBL의 전압 V_{CBL}은 VDD로 충전된다. 반면, W[i]가 1인 경우에는 P0가 오프되므로, V_{CBL}은 V_{DAC}를 유지하게 된다.
- [0091] 도 7b는 가중치에 따른 4비트 입력 데이터의 곱셈 연산 결과를 표로 도시하고 있다.
- [0092] 도 8a 및 도 8b는 본 발명의 일 실시 예에 따른 아날로그 곱셈 유닛의 누적 연산 동작을 설명하기 위한 도면이다. 일 실시 예에 따르면, AMU(110)는 AMU(110)에 포함된 복수의 로컬 어레이 각각에서 수행된 곱셈 연산 결과들을 누적 연산하고, 누적 연산 결과를 아날로그 전압으로 출력할 수 있다.
- [0093] 도 8a를 참조하면, AMU(110)의 각 로컬 어레이(LA)의 CBL에 아날로그 전압으로 형성되어 있는 곱셈 결과가 열 방향으로 전하 공유를 통해 누적 연산되고, 최종적으로 ABL(Accumulate Bit Line)에 MAC 연산 결과가 아날로그 값으로 표현되는 것을 확인할 수 있다. 이때, ABL<0> 내지 ABL<63>은 AMU[0] 내지 AMU[63]에 각각 대응된다.
- [0094] 구체적으로, 도 8b를 참조하면, AMU(110)는 각 로컬 어레이(LA₀ 내지 LA₁₅)의 곱셈 결과가 CBL에 아날로그 전압(V_{CBL,0} 내지 V_{CBL,15})으로 형성된 상태에서 제어 신호 eACC가 인가되면, 곱셈 결과를 누적하여 ABL에 아날로그 전압으로 출력할 수 있다.
- [0095] 이때, 일 실시 예에 따르면, ABL을 통해 출력되는 아날로그 전압(V_{ABL})은 아래 수학적 식 2와 같이 계산될 수 있다.

수학적 식 2

$$V_{ABL} = \frac{\sum_{i=0}^{15} (C_{CBL} \cdot V_{CBL,i}) + C_{ABL} \cdot VDD}{16 C_{CBL} + C_{ABL}}$$

- [0096]
- [0097] 여기서, V_{ABL}은 ABL을 통해 출력되는 아날로그 전압을 나타내고, C_{CBL}은 CBL의 캐패시턴스를 나타내고, V_{CBL,i}는 i번째 로컬 어레이의 곱셈 연산 결과를 나타내고, C_{ABL}은 ABL의 캐패시턴스를 나타낸다.
- [0098] 한편, ABL로 출력되는 MAC 연산 결과는 아날로그-디지털 변환기(ADC)를 이용하여 디지털 값으로 변환된 후 사용될 수 있다. 이를 위해, 컴퓨팅 인 메모리 장치(1000)는 아날로그 전압을 디지털 값으로 변환하기 위한 아날로그-디지털 변환기(ADC)를 더 포함할 수 있다.
- [0099] 도 9는 본 발명의 일 실시 예에 따른 컴퓨팅 인 메모리 장치의 각종 신호의 파형을 도시한 도면이다.
- [0100] 도 9를 참조하면, 프리차지(Precharge) 단계에서 모든 CBL, iBL과 ABL는 VDD로 초기화될 수 있다. 이때, 도 8b 및 도 9를 참조하면, 제어 신호 PChb가 인가되는 PMOS에는 로우 신호가 인가되고, 제어 신호 eACC가 활성화될 수 있다.
- [0101] 이후, 4비트 입력 데이터가 AMU(110)의 DAC(111)로 인가되면, 입력 데이터의 각 비트값에 따라 iBL[3], iBL[2], iBL[1], iBL[0]에는 VDD 또는 VSS 전압이 각각 형성된다. 이후, DAC(111)는 제어 신호 eDAC에 따라 행

방향으로 전하를 공유하여 4비트 입력 데이터에 대응하는 아날로그 전압을 CBL에 형성할 수 있다.

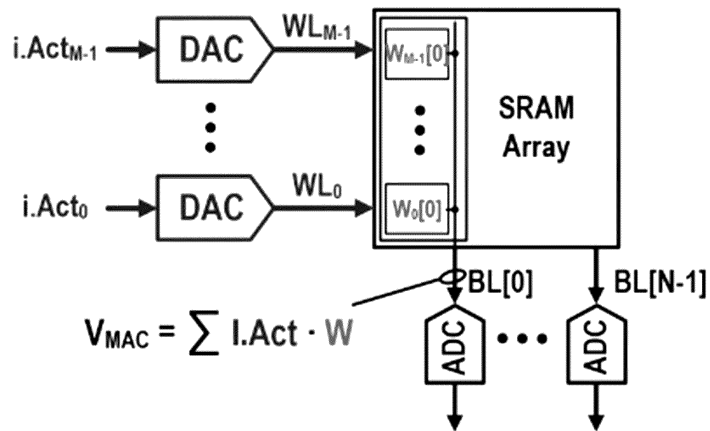
- [0102] 이후, 제어 신호 eMULTb가 비활성화 되면, AMU(110)는 비활성화된 제어 신호 eMULTb에 따라 주변 회로의 iBL과 메모리 셀의 연결을 차단할 수 있다. 또한, 제어 신호 CWL을 통해 CWL에 VSS가 인가되면, AMU(110)는 각 메모리 셀에서 곱셈 연산을 진행할 수 있다.
- [0103] 한편, 도면에는 도시되지 않았지만 이후 제어 신호 eACC가 활성화되면, AMU(110)는 활성화된 제어 신호 eACC에 따라 각 로컬 어레이에서 형성된 곱셈 결과를 열 방향으로 전하 공유할 수 있다. 이에 따라, ABL에 최종적으로 MAC 연산 결과가 아날로그 전압으로 표현될 수 있다. 이후 ABL에 인가된 아날로그 전압은 아날로그-디지털 변환기(ADC)를 통해 디지털 값으로 변환되어 다양하게 사용될 수 있다.
- [0104] 이상에서는, 다중 비트 입력 데이터가 4비트 입력 데이터인 경우를 예로 들어 설명하였으나, 실시 예가 이에 한정되는 것은 아니며, 더 많은 비트 수의 입력 데이터에 대한 MAC 연산에도 본 발명의 다양한 실시 예들이 변형되어 이용될 수 있다.
- [0105] 한편, 본 발명의 다양한 실시 예들은, 임베디드 메모리를 사용하는 모든 제품들에 적용될 수 있다. 또한, 캐시 메모리를 사용하는 AI(artificial intelligence) 기기나 AP(application processor), CPU(central processing unit), NPU(neural processing unit) 등과 같은 각종 연산 프로세서에 적용될 수도 있다.
- [0106] 이상과 같은 본 발명의 다양한 실시 예에 따르면, 상술한 본 발명의 다양한 실시 예들에 따르면, 메모리 내부에서 MAC 연산을 수행함으로써 DNN(Deep Neural Networks)과 같은 인공지능 연산에서 메모리와 연산기 간의 데이터 이동 횟수를 줄일 수 있다. 이에 따라, 인공지능 연산의 에너지 효율을 높일 수 있다.
- [0107] 또한, 다중 비트 입력 데이터를 아날로그 전압으로 변환할 때 메모리 내부의 DAC를 이용함으로써 공정 수율에 이점을 갖고, MAC 연산의 선형성을 높여 정확한 연산이 수행될 수 있다.
- [0108] 또한, 4비트 이상의 입력 데이터 사용이 가능해지므로, 1비트나 2비트 입력 데이터에 대한 연산보다 더 높은 인공지능 연산 정확도를 얻을 수 있다.
- [0109] 이상의 설명은 본 개시의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 개시가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 개시의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 또한, 본 개시에 따른 실시 예들은 본 개시의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시 예에 의하여 본 개시의 기술 사상의 범위가 한정되는 것은 아니다. 따라서, 본 개시의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 개시의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

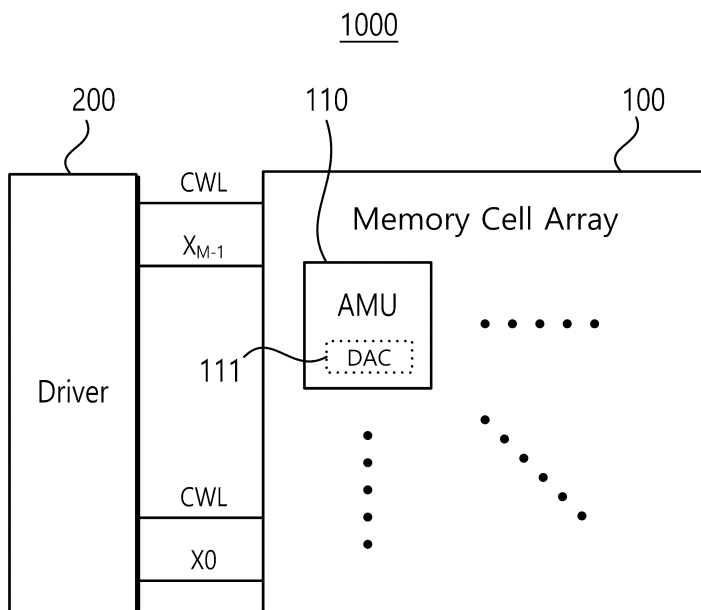
- [0110] 1000: 컴퓨팅 인 메모리 장치
- 100: 메모리 셀 어레이
- 200: 드라이버
- 300: 제어 회로
- 110: 아날로그 곱셈 유닛
- 111: 디지털-아날로그 변환기

도면

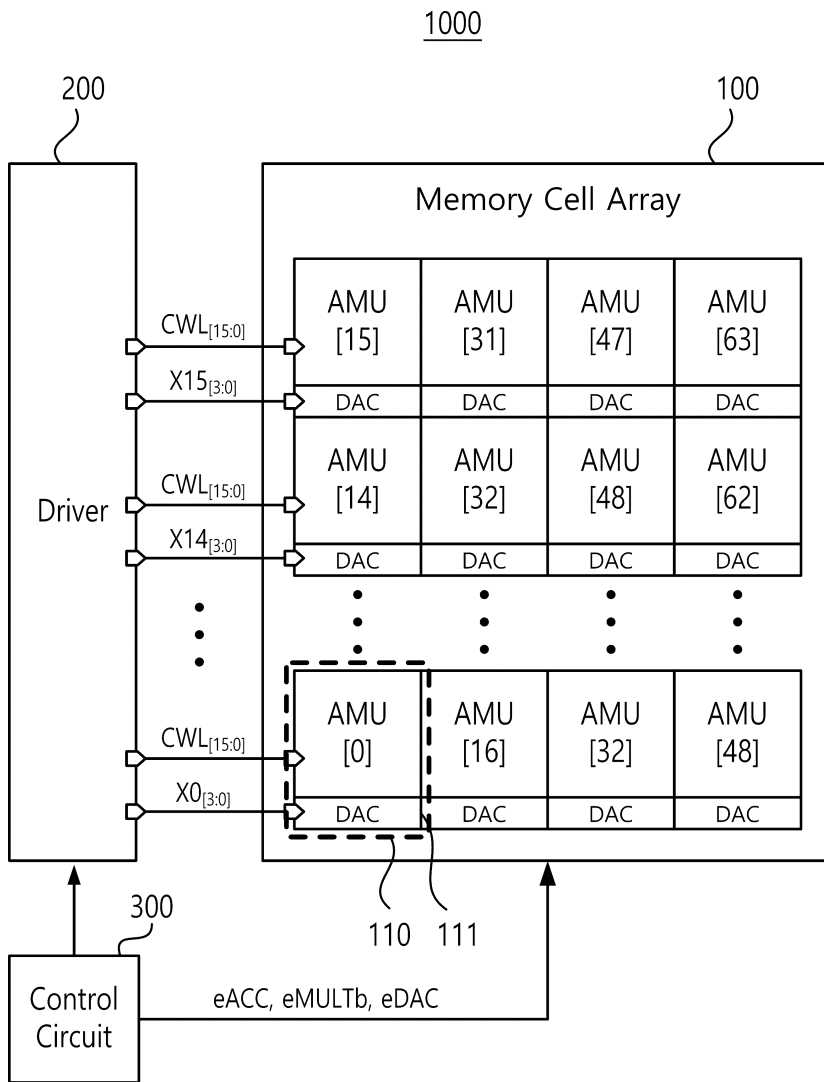
도면1a



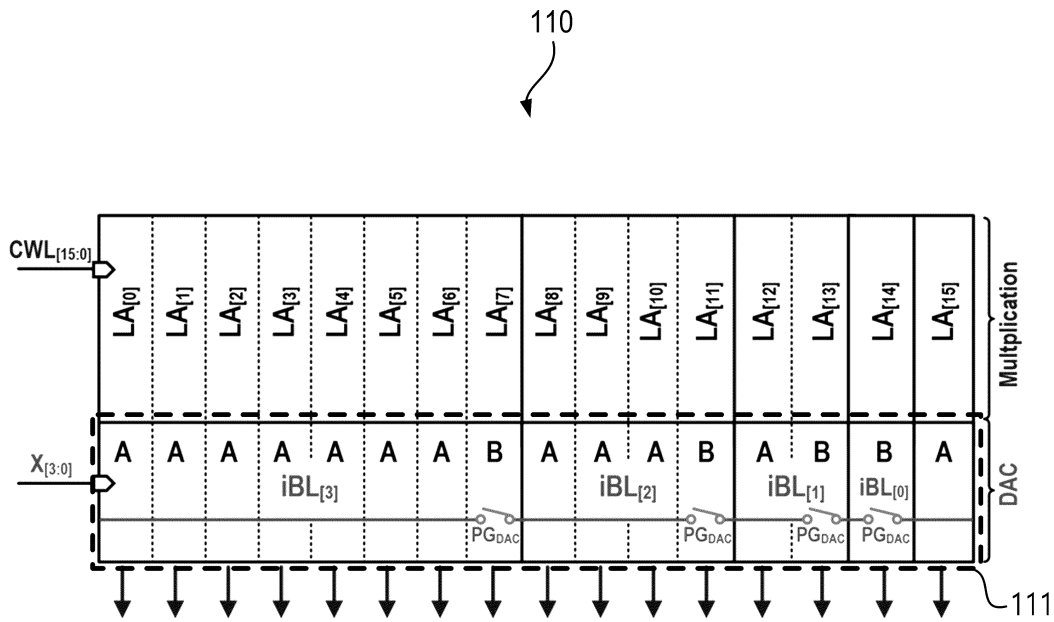
도면1b



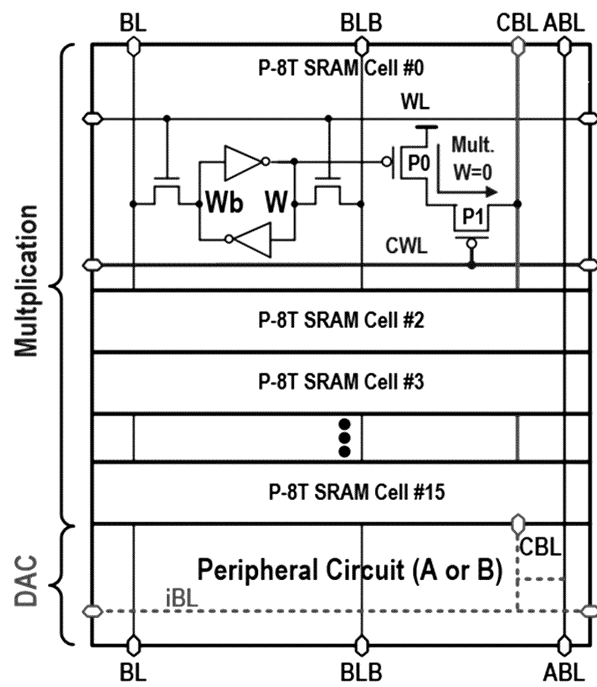
도면2



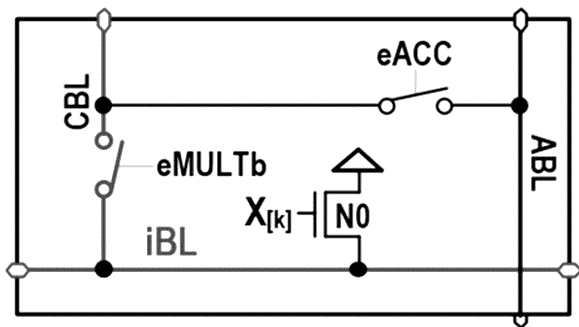
도면3



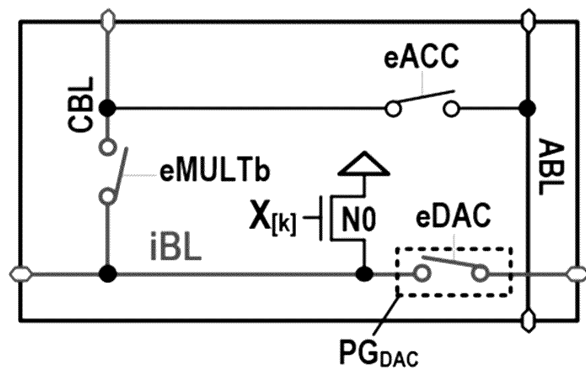
도면4a



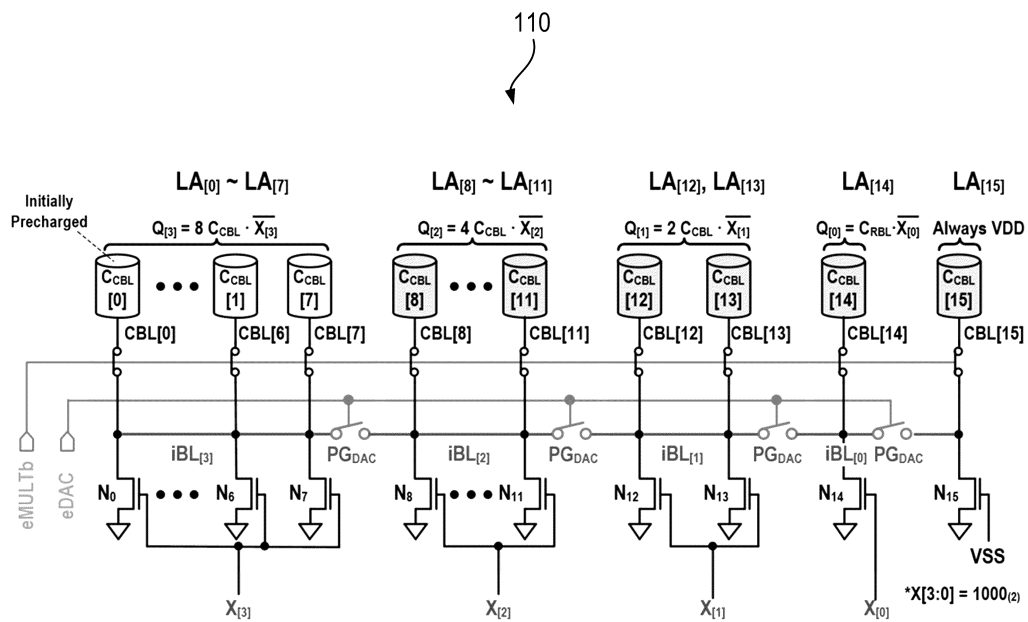
도면4b



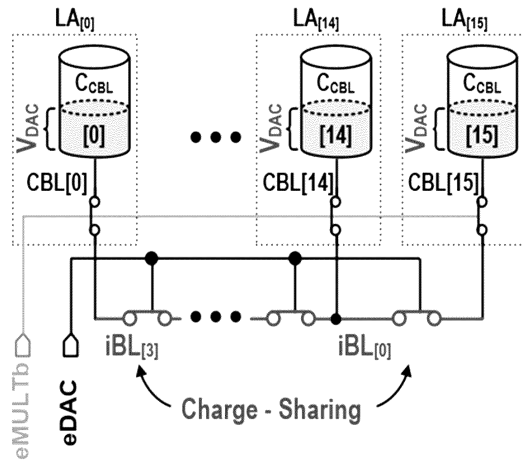
도면4c



도면5



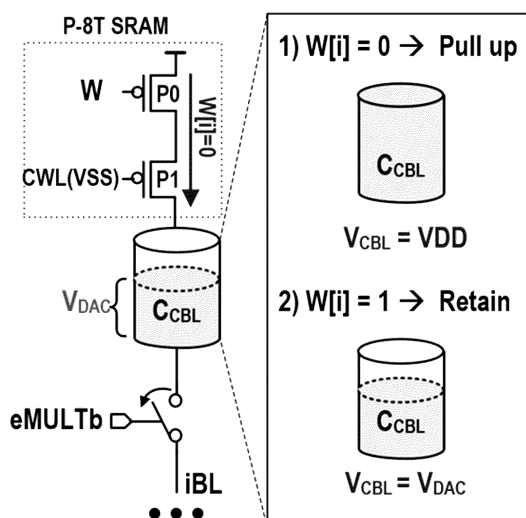
도면6a



도면6b

4-bit i.Act $X_{[3:0]}$	16 level Analog V_{CBL}
0000	$VDD : 0_{(10)}$
0001	$15/16 \cdot VDD : 1_{(10)}$
0010	$14/16 \cdot VDD : 2_{(10)}$
$\vdots$	$\vdots$
1110	$2/16 \cdot VDD : 14_{(10)}$
1111	$1/16 \cdot VDD : 15_{(10)}$

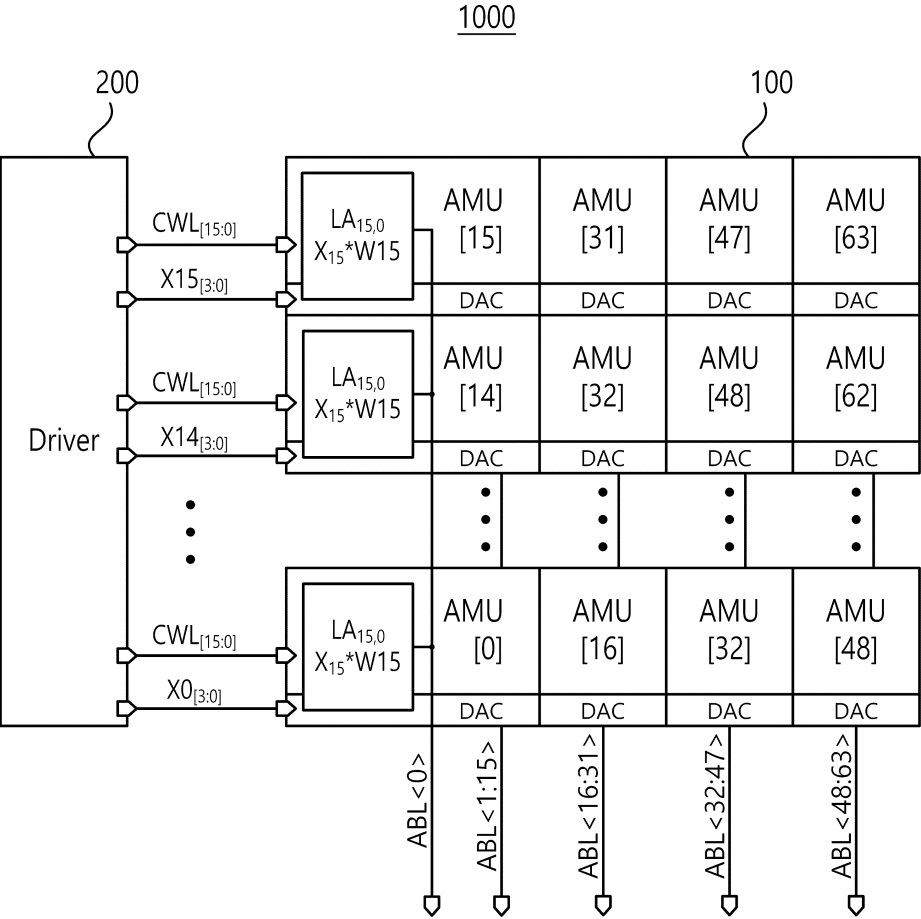
도면7a



도면7b

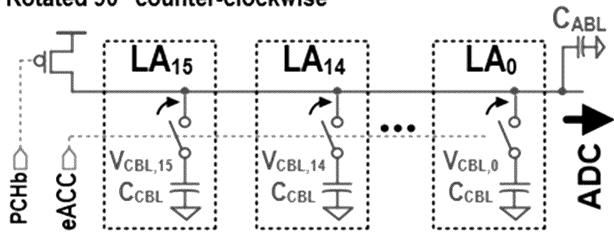
4-bit i.Act $X_{[3:0]}$	1-bit Weight (W)	16 level Mult. Result V_{CBL}
don't care	0 $V_{DAC} \rightarrow VDD$	$VDD : 0_{(10)}$
0000	1 $V_{DAC} \rightarrow V_{DAC}$	$VDD : 0_{(10)}$
0001		$15/16 \cdot VDD : 1_{(10)}$
0010		$14/16 \cdot VDD : 2_{(10)}$
$\vdots$		$\vdots$
1110		$2/16 \cdot VDD : 14_{(10)}$
1111		$1/16 \cdot VDD : 15_{(10)}$

도면8a



도면8b

Rotated 90° counter-clockwise



도면9

