



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년10월27일
(11) 등록번호 10-2318819
(24) 등록일자 2021년10월22일

(51) 국제특허분류(Int. Cl.)
G11C 11/419 (2015.01) G06N 3/063 (2006.01)
G11C 11/56 (2021.01) G11C 7/12 (2006.01)
(52) CPC특허분류
G11C 11/419 (2018.05)
G06N 3/063 (2013.01)
(21) 출원번호 10-2020-0070430
(22) 출원일자 2020년06월10일
심사청구일자 2020년06월10일
(56) 선행기술조사문헌
Xin Si et al., 'A Twin-8T SRAM
Computation-in-Memory Unit-Macro for Multibit
CNN-Based AI Edge Processors', IEEE JSSC,
Vol.55. No.1, Jan. 2020.

(73) 특허권자
연세대학교 산학협력단
서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)
(72) 발명자
정성욱
서울특별시 서대문구 연세로 50, 제3공학관 C513 (신촌동, 연세대학교)
안홍근
서울특별시 서대문구 연세로 50, 제3공학관 C420 (신촌동, 연세대학교)
이영규
서울특별시 서대문구 연세로 50, 제3공학관 C420 (신촌동, 연세대학교)
(74) 대리인
특허법인우인

전체 청구항 수 : 총 11 항

심사관 : 손윤식

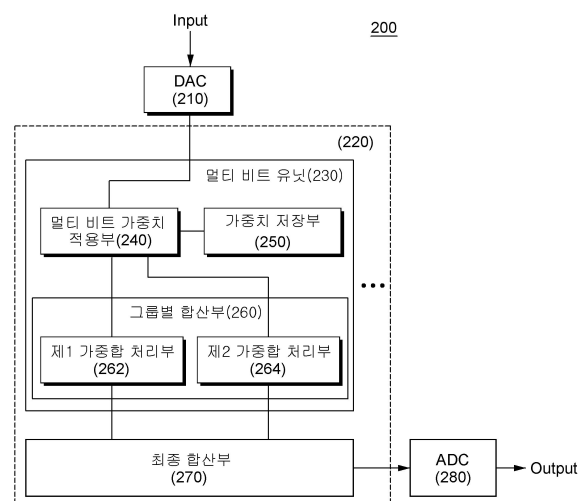
(54) 발명의 명칭 멀티 비트 가중치의 연산을 위한 인 메모리 장치

(57) 요약

멀티 비트 가중치의 연산을 위한 인 메모리 장치를 개시한다.

본 발명의 실시예에 따른 멀티 비트 메모리 셀 어레이는, 입력 신호에 기초한 입력 데이터를 저장하고, 저장된 입력 데이터에 멀티 비트 가중치를 적용하여 그룹별로 합산한 그룹별 합산값을 출력하는 적어도 하나의 멀티 비트 유닛; 및 상기 적어도 하나의 멀티 비트 유닛 각각과 연결되며, 그룹별로 비율을 조정하여 상기 그룹별 합산값을 입력받고, 입력된 상기 그룹별 합산값을 합산 처리하여 최종 출력값을 출력하는 최종 합산부를 포함할 수 있다.

대표도 - 도2



(52) CPC특허분류

G11C 11/56 (2013.01)

G11C 7/12 (2013.01)

명세서

청구범위

청구항 1

입력 신호에 기초한 입력 데이터를 복수의 멀티 비트 적용 노드 각각에 저장하고, 상기 복수의 멀티 비트 적용 노드 각각의 입력 데이터에 멀티 비트 가중치를 적용하며, 상기 복수의 멀티 비트 적용 노드를 그룹화하고 각각의 그룹별로 합산한 그룹별 합산값을 출력하는 적어도 하나의 멀티 비트 유닛; 및

상기 적어도 하나의 멀티 비트 유닛 각각과 스위치로 연결되며, 상기 스위치의 스위칭 동작 제어에 따라 그룹별로 비율을 조정하여 상기 그룹별 합산값을 입력받고, 입력된 상기 그룹별 합산값을 합산 처리하여 최종 출력값을 출력하는 최종 합산부

를 포함하는 것을 특징으로 하는 멀티 비트 메모리 셀 어레이.

청구항 2

제1항에 있어서,

상기 적어도 하나의 멀티 비트 유닛 각각은,

제1 비트라인(BL) 및 제2 비트라인(BLB) 사이에서 전원 전압을 공급받고, 상기 전원 전압에 기초하여 멀티 비트 가중치를 저장하는 가중치 저장부;

병렬로 연결된 복수의 멀티 비트 적용 노드를 포함하고, 상기 복수의 멀티 비트 적용 노드 각각에 연결된 커패시터에 저장된 상기 입력 데이터에 상기 멀티 비트 가중치 중 하나인 비트 가중치를 적용하여 가중치 적용값을 출력하는 멀티 비트 가중치 적용부; 및

상기 복수의 멀티 비트 적용 노드를 그룹화하여 상기 가중치 적용값을 그룹별로 합산하여 그룹별 합산값을 출력하는 그룹별 합산부

를 포함하는 것을 특징으로 하는 멀티 비트 메모리 셀 어레이.

청구항 3

제2항에 있어서,

상기 멀티 비트 가중치 적용부는,

복수의 멀티 비트 적용 노드 각각에 인가되는 상기 입력 신호에 대응하는 입력 전압을 통해 상기 커패시터에 상기 입력 데이터를 저장하고, 소정의 스위칭 동작을 통해 상기 비트 가중치가 입력되면, 멀티 비트 적용 노드 각각에서 상기 입력 데이터에 상기 비트 가중치를 적용하여 상기 가중치 적용값을 출력하는 것을 특징으로 하는 멀티 비트 메모리 셀 어레이.

청구항 4

제3항에 있어서,

복수의 멀티 비트 적용 노드 각각에 연결된 상기 커패시터는,

서로 상이한 정전용량을 가지거나, 일부 커패시터만 동일한 정전 용량을 가지며, 상기 멀티 비트 가중치 적용부는 복수의 멀티 비트 적용 노드 각각에서 상기 정전용량에 대응되는 입력 데이터에 상기 비트 가중치를 적용하여 상기 가중치 적용값을 출력하는 것을 특징으로 하는 멀티 비트 메모리 셀 어레이.

청구항 5

제4항에 있어서,

상기 멀티 비트 가중치 적용부는,

상기 복수의 멀티 비트 적용 노드 각각에서 상기 입력 데이터에 상기 비트 가중치를 곱셈 처리하여 산출된 상기 가중치 적용값을 출력하는 것을 특징으로 하는 멀티 비트 메모리 셀 어레이.

청구항 6

제4항에 있어서,

상기 그룹별 합산부는,

상기 복수의 멀티 비트 적용 노드 중 제1 그룹에 포함된 멀티 비트 적용 노드 각각에서 출력되는 복수의 상기 가중치 적용값을 합산 처리하여 산출된 제1 그룹 합산값을 출력하는 제1 가중합 처리부; 및

상기 복수의 멀티 비트 적용 노드 중 제2 그룹에 포함된 멀티 비트 적용 노드 각각에서 출력되는 복수의 가중치 적용값을 합산 처리하여 산출된 제2 그룹 합산값을 출력하는 제2 가중합 처리부를 포함하며,

상기 제1 그룹에 포함된 멀티 비트 적용 노드는 상기 제2 그룹에 포함된 멀티 비트 적용 노드와 서로 다른 멀티 비트 적용 노드인 것을 특징으로 하는 멀티 비트 메모리 셀 어레이.

청구항 7

제6항에 있어서,

상기 그룹별 합산부는,

상기 멀티 비트 가중치 중 상위 자릿수의 상위 비트 가중치에 대응하는 멀티 비트 적용 노드를 상기 제1 그룹에 포함시켜 상기 제1 그룹 합산값을 출력하고,

상기 멀티 비트 가중치 중 상기 상위 비트 가중치를 제외한 하위 자릿수의 하위 비트 가중치에 대응하는 멀티 비트 적용 노드를 상기 제2 그룹에 포함시켜 상기 제2 그룹 합산값을 출력하는 것을 특징으로 하는 멀티 비트 메모리 셀 어레이.

청구항 8

제7항에 있어서,

상기 최종 합산부는,

상기 적어도 하나의 멀티 비트 유닛 각각에서 출력되는 상기 제1 그룹 합산값을 입력받아 제1 최종 합산값을 산출하는 제1 최종 합산라인; 및

상기 적어도 하나의 멀티 비트 유닛 각각에서 출력되는 상기 제2 그룹 합산값을 입력받아 제2 최종 합산값을 산출하는 제2 최종 합산라인을 포함하며,

상기 최종 합산부는, 상기 적어도 하나의 멀티 비트 유닛 각각과 상기 제1 최종 합산라인 사이에 연결된 제1 최종 스위치와 상기 적어도 하나의 멀티 비트 유닛 각각과 상기 제2 최종 합산라인 사이에 연결된 제2 최종 스위치의 스위칭 동작에 따라 상기 제1 그룹 합산값 및 상기 제2 그룹 합산값을 상기 비율로 조정하고,

상기 제1 최종 합산라인 및 상기 제2 최종 합산라인 사이에 연결된 제3 최종 스위치의 스위칭 동작을 통해 상기 제1 최종 합산값 및 상기 제2 최종 합산값을 합산 처리하여 상기 최종 출력값을 출력하는 것을 특징으로 하는 멀티 비트 메모리 셀 어레이.

청구항 9

복수의 비트라인 및 복수의 워드라인;

디지털 입력신호를 아날로그 형태의 입력 전압으로 변환하는 DAC;

상기 입력 전압을 기반으로 멀티 비트 가중치 및 입력 데이터를 저장하고, 상기 멀티 비트 가중치 및 상기 입력 데이터를 이용하여 곱셈 처리(Multiplication) 및 합산 처리(Summation) 동작을 수행하여 최종 출력값을 출력하는 적어도 하나의 멀티 비트 메모리 셀을 포함하는 메모리 셀 어레이; 및

상기 최종 출력값을 입력받고, 상기 최종 출력값을 디지털 출력신호로 변환하는 ADC

를 포함하는 것을 특징으로 하는 인 메모리 장치.

청구항 10

제9항에 있어서,

상기 메모리 셀 어레이는,

입력 신호에 기초한 입력 데이터를 복수의 멀티 비트 적용 노드 각각에 저장하고, 상기 복수의 멀티 비트 적용 노드 각각의 입력 데이터에 멀티 비트 가중치를 적용하며, 상기 복수의 멀티 비트 적용 노드를 그룹화하고 각각의 그룹별로 합산한 그룹별 합산값을 출력하는 적어도 하나의 멀티 비트 유닛; 및

상기 적어도 하나의 멀티 비트 유닛 각각과 스위치로 연결되며, 상기 스위치의 스위칭 동작 제어에 따라 그룹별로 비율을 조정하여 상기 그룹별 합산값을 입력받고, 입력된 상기 그룹별 합산값을 합산 처리하여 최종 출력값을 출력하는 최종 합산부

를 포함하는 것을 특징으로 하는 인 메모리 장치.

청구항 11

제10항에 있어서,

상기 적어도 하나의 멀티 비트 유닛 각각은,

제1 비트라인(BL) 및 제2 비트라인(BLB) 사이에서 전원 전압을 공급받고, 상기 전원 전압에 기초하여 멀티 비트 가중치를 저장하는 가중치 저장부; 병렬로 연결된 복수의 멀티 비트 적용 노드를 포함하고, 상기 복수의 멀티 비트 적용 노드 각각에 연결된 커패시터에 저장된 상기 입력 데이터에 상기 멀티 비트 가중치 중 하나인 비트 가중치를 적용하여 가중치 적용값을 출력하는 멀티 비트 가중치 적용부; 및 상기 복수의 멀티 비트 적용 노드를 그룹화하여 상기 가중치 적용값을 그룹별로 합산하여 그룹별 합산값을 출력하는 그룹별 합산부를 포함하되,

상기 멀티 비트 가중치 적용부는, 복수의 멀티 비트 적용 노드 각각에 인가되는 상기 입력 신호에 대응하는 입력 전압을 통해 상기 커패시터에 상기 입력 데이터를 저장하고, 소정의 스위칭 동작을 통해 상기 비트 가중치가 입력되면, 멀티 비트 적용 노드 각각에서 상기 입력 데이터에 상기 비트 가중치를 적용하여 상기 가중치 적용값을 출력하며,

복수의 멀티 비트 적용 노드 각각에 연결된 상기 커패시터는, 서로 상이한 정전용량을 가지거나, 일부 커패시터만 동일한 정전 용량을 가지며, 상기 멀티 비트 가중치 적용부는 복수의 멀티 비트 적용 노드 각각에서 상기 정전용량에 대응되는 입력 데이터에 상기 비트 가중치를 적용하여 상기 가중치 적용값을 출력하는 것을 특징으로 하는 인 메모리 장치.

발명의 설명

기술 분야

[0001] 본 발명은 멀티 비트 가중치를 연산하기 위한 인 메모리 장치에 관한 것이다. 본 발명의 연구는 삼성미래기술육성사업의 지원을 받아 수행된 ICT 창의과제의 실시간 초해상도 복원 작업에 적용 가능한 저전력 신경망 단일 칩 기술 개발 기술과 관련된다(No. 2018-11-1438).

배경 기술

[0002] 이 부분에 기술된 내용은 단순히 본 발명의 실시예에 대한 배경 정보를 제공할 뿐 종래기술을 구성하는 것은 아니다.

[0003] 딥러닝(Deep Learning)과 같은 인공지능 처리 방식에서 많은 양의 데이터 처리가 필요해짐에 따라 컴퓨팅(computing)에서 발생하는 전원(power) 및 지연(delay)에 대한 오버헤드(overhead)보다 메모리 데이터 액세스(memory data access) 및 전송(transfer)에서 발생하는 전원(power) 및 지연(delay)에 대한 오버헤드(overhead)가 더 커지고 있다. 따라서, 이러한 오버헤드를 줄이기 위하여 도 1a에 도시된 것과 같이, 메모리(Memory)에서 연산하는 기술인 CIM(Computing In Memory) 기술이 필요하다.

- [0004] 도 1a 및 도 1b는 종래의 인 메모리 장치를 설명하기 위한 도면이다.
- [0005] 도 1a의 (a)는 초해상도 복원 작업(Super-resolution)을 위한 신경망(neural network)을 나타낸다. 이미지 프로세싱(Image processing) 분야는 인공지능 신경망의 발달과 더불어 급속한 성능 발전이 진행되고 있다. 최근, 하드웨어로 알고리즘을 구현하기 위하여 가중치(weight) 압축과 같은 기법을 통한 이진 네트워크(binary network)가 개발되었다.
- [0006] 하지만, 초해상도 복원 작업과 같이 픽셀 값(pixel value) 자체를 중요시 하는 작업(task)에서는 알고리즘의 비트 폭(bit-width)이 감소함에 따라 성능이 급감하게 된다. 알고리즘의 최대 신호 대 잡음비(PSNR: Peak Signal-to-noise ratio)를 유지하기 위해서는 최소 8 bit 이상의 가중치(weight)를 사용해야 한다. 이에, 인 메모리 장치에서의 멀티 비트 가중치(Multi-bit Weight) 연산이 반드시 필요하게 된다.
- [0007] 도 1a의 (b)는 비트 라인 기반의 가중치 합산을 수행하는 종래의 인 메모리 장치를 나타내고, 도 1a의 (c)는 워드 라인 기반의 가중치 합산을 수행하는 종래의 인 메모리 장치를 나타낸다. 도 1a의 (b) 및 (c)를 참고하면, 종래의 인 메모리 장치는 분류 과업 타겟(Classification task target)을 이용하여 이진 가중치 기반의 신경망(Binary neural network)을 구현한다. 다만, 종래의 인 메모리 장치는 멀티 비트의 입력 신호 및 출력신호에 이진 가중치를 합산하여 동작함에 따라 성능 저하가 발생하고, 성능 저하로 인한 정상적인 초해상도 복원 작업의 적용(SR application) 동작이 불가능하다.
- [0008] 도 1b는 전하 공유(charge sharing) 방식의 종래의 인 메모리 장치를 나타낸다.
- [0009] 도 1b의 (a)는 단일 비트 가중치를 적용하는 종래의 인 메모리 장치를 나타낸다. 도 1b의 (a)의 인 메모리 장치는 프리차지 단계, 가중치 곱셈 처리 단계, 합산 단계로 동작한다.
- [0010] 프리차지 단계는 디지털 형태의 입력 신호(6 bit I_x)를 아날로그 형태의 전압(V_x)로 변환하여 프리차지 동작을 수행한다. 가중치 곱셈 처리 단계는 셀 데이터에 대응하는 이진 데이터(가중치)를 전압(V_x)에 적용하고, 가중치가 적용된 전압(V_x)을 방전한다. 합산 단계는 전압(V_x)이 방전되는 노드를 전하 공유를 위해 연결하여 합산된 합산 전압(V_{sum})을 출력하고, 출력된 합산 전압을 디지털 형태의 출력 신호(6 bit)로 변환하여 출력한다.
- [0011] 도 1b의 (b)와 같이, 전하 공유(charge sharing) 방식의 종래의 인 메모리 장치는 멀티 비트 가중치를 적용하기 위해서는 추가적인 커패시터(capacitor)를 사용하여 가중치 비트 폭(Weight bit-width)을 확장할 수 있다. 구체적으로, 종래의 인 메모리 장치는 N 개의 멀티 비트 가중치를 적용하기 위하여 Array column을 N 배 확장하기 위하여 추가적인 커패시터를 각각의 노드에 추가 연결하여 멀티 비트 가중치를 적용할 수 있는 장치를 구현할 수 있다.
- [0012] 다만, 멀티 비트 가중치를 적용할 수 있는 종래의 인 메모리 장치는, 2의 승수로 증가하는 커패시터가 요구됨에 따라 가중치 비트 폭이 1 비트 증가하는 경우, 1 비트당 약 2 배의 면적이 증가하게 된다. 예를 들어, 8 비트의 멀티 비트 가중치를 적용하는 장치의 경우, 1C, 2C, ..., 128C 등의 정전용량의 커패시터가 필요하며, 이로 인해 면적 오버헤드(Area overhead)가 발생하게 된다.

발명의 내용

해결하려는 과제

- [0013] 본 발명은 계층적(Hierarchy)으로 멀티 비트 가중치에 대한 합산을 처리하는 멀티 비트 가중치의 연산을 위한 인 메모리 장치를 제공하는 데 주된 목적이 있다.

과제의 해결 수단

- [0014] 본 발명의 일 측면에 의하면, 상기 목적을 달성하기 위한 멀티 비트 메모리 셀 어레이는, 입력 신호에 기초한 입력 데이터를 저장하고, 저장된 입력 데이터에 멀티 비트 가중치를 적용하여 그룹별로 합산한 그룹별 합산값을 출력하는 적어도 하나의 멀티 비트 유닛; 및 상기 적어도 하나의 멀티 비트 유닛 각각과 연결되며, 그룹별로 비율을 조정하여 상기 그룹별 합산값을 입력받고, 입력된 상기 그룹별 합산값을 합산 처리하여 최종 출력값을 출력하는 최종 합산부를 포함할 수 있다.
- [0015] 또한, 본 발명의 다른 측면에 의하면, 상기 목적을 달성하기 위한 인 메모리 장치는, 복수의 비트라인 및 복수의 워드라인; 디지털 입력신호를 아날로그 형태의 입력 전압으로 변환하는 DAC; 상기 입력 전압을 기반으로 멀티

터 비트 가중치 및 입력 데이터를 저장하고, 상기 멀티 비트 가중치 및 상기 입력 데이터를 이용하여 곱셈 처리(Multiplication) 및 그룹별로 합산 처리(Summation) 동작을 수행하여 최종 출력값을 출력하는 적어도 하나의 멀티 비트 메모리 셀을 포함하는 메모리 셀 어레이; 및 상기 최종 출력값을 입력받고, 상기 최종 출력값을 디지털 출력신호로 변환하는 ADC를 포함할 수 있다.

발명의 효과

- [0016] 이상에서 설명한 바와 같이, 본 발명은 멀티 비트 가중치를 적용하기 위하여 추가로 요구되는 커패시터를 현저히 감소시킬 수 있는 효과가 있다.
- [0017] 또한, 본 발명은 멀티 비트 가중치를 적용하기 위하여 요구되는 면적을 줄일 수 있는 효과가 있다.
- [0018] 또한, 본 발명은 연산하고자 하는 멀티 비트 가중치의 비트 폭이 증가할수록 더 많은 커패시터와 면적을 줄일 수 있는 효과가 있다.
- [0019] 또한, 본 발명은 합산 단계를 증가시킴에 따라 커패시터의 감소 효율을 증가할 수 있는 효과가 있다.

도면의 간단한 설명

- [0020] 도 1a 및 도 1b는 종래의 인 메모리 장치를 설명하기 위한 도면이다.
- 도 2는 본 발명의 실시예에 따른 인 메모리 장치를 개략적으로 나타낸 블록 구성도이다.
- 도 3은 본 발명의 실시예에 따른 인 메모리 장치의 동작 단계를 설명하기 위한 순서도이다.
- 도 4는 본 발명의 실시예에 따른 인 메모리 장치의 회로 구성을 나타낸 도면이다.
- 도 5는 본 발명의 실시예에 따른 인 메모리 장치의 1차 합산 동작을 설명하기 위한 도면이다.
- 도 6은 본 발명의 실시예에 따른 인 메모리 장치의 2차 합산 동작을 설명하기 위한 도면이다.
- 도 7은 본 발명의 실시예에 따른 인 메모리 장치의 실험결과를 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0021] 이하, 본 발명의 바람직한 실시예를 첨부된 도면들을 참조하여 상세히 설명한다. 본 발명을 설명함에 있어, 관련된 공지 구성 또는 기능에 대한 구체적인 설명이 본 발명의 요지를 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명은 생략한다. 또한, 이하에서 본 발명의 바람직한 실시예를 설명할 것이나, 본 발명의 기술적 사상은 이에 한정하거나 제한되지 않고 당업자에 의해 변형되어 다양하게 실시될 수 있음은 물론이다. 이하에서는 도면들을 참조하여 본 발명에서 제안하는 멀티 비트 가중치의 연산을 위한 인 메모리 장치에 대해 자세하게 설명하기로 한다.
- [0022] 초해상도 복원 작업(Super-resolution)과 같은 작업(task)에서는 같이 픽셀 값(pixel value)이 중요하기 때문에, 알고리즘 성능 유지를 위하여 인 메모리 장치의 하드웨어(hardware)는 멀티 비트 가중치(multi-bit weight)를 적용하는 기능을 반드시 지원해야만 한다.
- [0023] 종래의 인 메모리 장치의 구조는 전하 공유(charge sharing) 방식을 이용하여 가중합 처리(weighted summation)를 수행하고 있으나, 가중치의 비트 폭(bit-width)은 1 bit(Binary)만을 지원하고 있다.
- [0024] 종래의 인 메모리 장치에서 멀티 비트 가중치를 적용하기 위해서는 멀티 비트 가중치의 연산을 위하여 1 bit 증가 시 약 2 배의 추가적인 커패시터 면적이 요구된다. 예를 들어, 7 bit의 멀티 비트 가중치의 경우 인 메모리 장치는 $1, 2, \dots, 2^6$ 의 비율을 갖는 커패시터가 필요하고, 8 bit의 멀티 비트 가중치의 경우 인 메모리 장치는 $1, 2, \dots, 2^7$ 의 비율을 갖는 커패시터가 필요하다. 이에, 비트 수의 증가에 따라 요구되는 커패시터가 기하급수적(exponentially)으로 증가하게 된다.
- [0025] 즉, 종래의 인 메모리 장치에서는 커패시터로 인한 칩 면적(chip area)이 증가하게 되고, 이는 실리콘(Silicon)에서 멀티 비트 가중치를 적용하기 위한 하드웨어를 구현하는데 큰 문제가 된다.
- [0026] 본 발명에서는 계층적(hierarchy)으로 합산 처리를 수행하는 컴퓨팅 인 메모리(CIM: Computation-in-Memory)의 구조를 제안하며, 추가적인 커패시터의 사용을 최소화하기 위한 연산 방법을 제안한다.

- [0027] 도 2는 본 발명의 실시예에 따른 인 메모리 장치를 개략적으로 나타낸 블록 구성도이다.
- [0028] 본 실시예에 따른 인 메모리 장치(200)는 복수의 비트라인, 복수의 워드라인, DAC(Digital Analog Converter, 210), 메모리 셀 어레이(220) 및 ADC(Analog Digital Converter, 280)를 포함한다. 인 메모리 장치(200)에 포함된 DAC(210) 및 ADC(280)는 종래의 구성과 동일함으로 그 자세한 설명은 생략하도록 한다.
- [0029] 본 실시예에 따른 메모리 셀 어레이(220)는 적어도 하나의 멀티 비트 유닛(230) 및 최종 합산부(270)를 포함한다.
- [0030] 도 2의 인 메모리 장치(200)는 일 실시예에 따른 것으로서, 도 2에 도시된 모든 블록이 필수 구성요소는 아니며, 다른 실시예에서 인 메모리 장치(200)에 포함된 일부 블록이 추가, 변경 또는 삭제될 수 있다.
- [0031] 본 실시예에 따른 인 메모리 장치(200)는 SRAM(Static RAM) 기반의 컴퓨팅 장치인 것으로 기재하고 있으며, 메모리 셀 어레이(220)은 SRAM 메모리 셀을 포함할 수 있으나 반드시 이에 한정되는 것은 아니며, 다양한 종류의 휘발성 메모리(volatile memory) 기반의 구성으로 구현될 수도 있다.
- [0032] 메모리 셀 어레이(220)는 DAC(210)로부터 입력 전압을 입력받아 입력 데이터를 저장하고, 저장된 입력 데이터에 멀티 비트 가중치를 적용하여 합산 처리한 최종 출력값을 ADC(280)로 출력한다. 이하, 메모리 셀 어레이(220)에 포함된 구성 및 그 동작에 대해 설명하도록 한다.
- [0033] 적어도 하나의 멀티 비트 유닛(230)은 입력 신호에 기초한 입력 데이터를 저장하고, 저장된 입력 데이터에 멀티 비트 가중치를 적용하여 그룹별로 합산한 그룹별 합산값을 출력한다. 본 실시예에 따른 멀티 비트 유닛(230)은 멀티 비트 가중치 적용부(240), 가중치 저장부(250) 및 그룹별 합산부(260)를 포함한다.
- [0034] 멀티 비트 가중치 적용부(240)는 병렬로 연결된 복수의 멀티 비트 적용 노드를 포함한다.
- [0035] 멀티 비트 가중치 적용부(240)는 복수의 멀티 비트 적용 노드 각각에 연결된 커패시터에 저장된 입력 데이터에 멀티 비트 가중치 중 하나인 비트 가중치를 적용하여 가중치 적용값을 출력한다. 예를 들어, 멀티 비트 가중치가 '10001011'인 경우, 멀티 비트 가중치 적용부(240)는 각각의 멀티 비트 적용 노드에 '1', '0', '0', '0', '1', '0', '1', '1' 비트 가중치 각각을 적용한다. 여기서, 멀티 비트 가중치는 복수의 멀티 비트 적용 노드 각각에 구비된 스위치의 스위칭 제어를 통해 가중치 저장부(250)로부터 획득된 가중치를 의미하며, 멀티 비트 가중치는 이진 비트(0 또는 1)인 것이 바람직하다.
- [0036] 멀티 비트 가중치 적용부(240)는 복수의 멀티 비트 적용 노드 각각에 인가되는 입력 신호에 대응하는 입력 전압을 통해 커패시터에 입력 데이터를 저장한다. 이후, 멀티 비트 가중치 적용부(240)는 소정의 스위칭 동작을 통해 비트 가중치가 입력되면, 멀티 비트 적용 노드 각각에서 입력 데이터에 비트 가중치를 적용하여 가중치 적용값을 출력한다.
- [0037] 복수의 멀티 비트 적용 노드 각각에 연결된 커패시터는 서로 상이한 정전용량을 가지거나, 일부 커패시터만 동일한 정전 용량을 가진다. 이에, 멀티 비트 가중치 적용부(240)는 복수의 멀티 비트 적용 노드 각각에서 정전용량에 대응되는 입력 데이터에 비트 가중치를 적용하여 가중치 적용값을 출력한다. 구체적으로, 멀티 비트 가중치 적용부(240)는 복수의 멀티 비트 적용 노드 각각에서 입력 데이터에 비트 가중치를 곱셈 처리하여 산출된 가중치 적용값을 출력할 수 있다.
- [0038] 예를 들어, 8 비트의 멀티 비트 가중치 중 상위 비트 가중치(W_{1-1} 내지 W_{1-4})를 적용하기 위한 4 개의 멀티 비트 적용 노드 각각에는 8C, 4C, 2C, 1C와 같이 서로 다른 정전용량의 커패시터가 연결될 수 있다. 또한, 8 비트의 멀티 비트 가중치 중 하위 비트 가중치(W_{1-5} 내지 W_{1-8})를 적용하기 위한 4 개의 멀티 비트 적용 노드 각각에는 8C, 4C, 2C, 1C와 같이 서로 다른 정전용량의 커패시터가 연결될 수 있다.
- [0039] 한편, 본 실시예에 따른 멀티 비트 가중치 적용부(240)는 8 비트의 멀티 비트 가중치를 2 비트 단위로 구분하여 합산할 경우, 8 비트의 멀티 비트 가중치 중 제1 비트 가중치(W_{1-1} 내지 W_{1-2}), 제2 비트 가중치(W_{1-3} 내지 W_{1-4}), 제3 비트 가중치(W_{1-5} 내지 W_{1-6}), 제4 비트 가중치(W_{1-7} 내지 W_{1-8}) 각각을 적용하기 위한 2 개의 멀티 비트 적용 노드 각각에는 2C, 1C와 같이 서로 다른 정전용량의 커패시터가 연결될 수 있다.
- [0040] 멀티 비트 가중치 적용부(240)의 입력 데이터 획득 및 멀티 비트 가중치 적용에 대한 제어는 별도의 컨트롤러(미도시)에 의해 제어될 수 있으나 반드시 이에 한정되는 것은 아니다.
- [0041] 가중치 저장부(250)는 제1 비트라인(BL) 및 제2 비트라인(BLB) 사이에서 전원 전압을 공급받고, 전원 전압에 기

초하여 멀티 비트 가중치를 저장한다. 여기서, 가중치 저장부(250)는 일반적인 메모리 셀에 포함된 데이터 저장 회로로 구현될 수 있다.

[0042] 예를 들어, 데이터 저장 회로는 제1 비트라인(BL) 및 소정의 워드라인에 연결되는 제1 액세스 트랜지스터와 제2 비트라인(BLB) 및 소정의 워드라인에 연결되는 제2 액세스 트랜지스터를 포함할 수 있다. 또한, 데이터 저장 회로는 제1 액세스 트랜지스터 및 제2 액세스 트랜지스터에 연결되고, 복수의 트랜지스터를 포함하며 제1 비트라인 또는 제2 비트라인이 충전 또는 방전되도록 동작하는 교차 결합 인버터를 포함할 수 있다.

[0043] 가중치 저장부(250)는 멀티 비트 가중치를 저장하며, 스위칭 제어에 의해 멀티 비트 가중치 적용부(240)로 멀티 비트 가중치를 공급한다. 예를 들어, 가중치 저장부(250)는 '10001011'인 멀티 비트 가중치를 저장하는 경우, 가중치 저장부(250)는 멀티 비트 가중치 적용부(240)로 '10001011'인 멀티 비트 가중치를 공급한다. 여기서, 가중치 저장부(250)는 멀티 비트 가중치 적용부(240)에 포함된 복수의 멀티 비트 적용 노드 각각으로 '1', '0', '0', '0', '1', '0', '1', '1' 비트 가중치 각각을 공급할 수 있다.

[0044] 가중치 저장부(250)의 저장 및 출력 제어는 별도의 컨트롤러(미도시)에 의해 제어될 수 있으나 반드시 이에 한정되는 것은 아니다.

[0045] 그룹별 합산부(260)는 멀티 비트 가중치 적용부(240)와 연동하여 그룹별 합산값을 출력하는 동작을 수행한다. 구체적으로, 그룹별 합산부(260)는 멀티 비트 가중치 적용부(240)에 포함된 복수의 멀티 비트 적용 노드를 그룹화하고, 각각의 그룹별로 복수의 멀티 비트 적용 노드 각각에서 출력되는 가중치 적용값을 합산하여 그룹별 합산값을 출력한다.

[0046] 그룹별 합산부(260)는 제1 가중합 처리부(262) 및 제2 가중합 처리부(264)를 포함한다.

[0047] 제1 가중합 처리부(262)는 복수의 멀티 비트 적용 노드 중 제1 그룹에 포함된 멀티 비트 적용 노드 각각에서 출력되는 복수의 가중치 적용값을 합산 처리하여 제1 그룹 합산값을 산출하고, 산출된 제1 그룹 합산값을 출력한다.

[0048] 제1 가중합 처리부(262)는 멀티 비트 가중치 중 상위 자릿수의 상위 비트 가중치에 대응하는 멀티 비트 적용 노드를 제1 그룹에 포함시켜 제1 그룹 합산값을 출력한다. 예를 들어, 제1 가중합 처리부(262)는 8 비트의 멀티 비트 가중치 중 상위 4 비트인 상위 비트 가중치에 대응하는 멀티 비트 적용 노드를 제1 그룹에 포함시키고, 제1 그룹에 포함된 멀티 비트 적용 노드 각각에서 출력되는 4 개의 가중치 적용값을 합산 처리하여 제1 그룹 합산값을 출력한다.

[0049] 제2 가중합 처리부(264)는 복수의 멀티 비트 적용 노드 중 제2 그룹에 포함된 멀티 비트 적용 노드 각각에서 출력되는 복수의 가중치 적용값을 합산 처리하여 제2 그룹 합산값을 산출하고, 산출된 제2 그룹 합산값을 출력한다.

[0050] 제2 가중합 처리부(264)는 멀티 비트 가중치 중 상위 비트 가중치를 제외한 하위 자릿수의 하위 비트 가중치에 대응하는 멀티 비트 적용 노드를 제2 그룹에 포함시켜 제2 그룹 합산값을 출력한다. 예를 들어, 제2 가중합 처리부(264)는 8 비트의 멀티 비트 가중치 중 하위 4 비트인 하위 비트 가중치에 대응하는 멀티 비트 적용 노드를 제2 그룹에 포함시키고, 제2 그룹에 포함된 멀티 비트 적용 노드 각각에서 출력되는 4 개의 가중치 적용값을 합산 처리하여 제2 그룹 합산값을 출력한다.

[0051] 그룹별 합산부(260)에서 제1 그룹에 포함된 멀티 비트 적용 노드는 제2 그룹에 포함된 멀티 비트 적용 노드와 서로 다른 멀티 비트 적용 노드인 것이 바람직하다.

[0052] 그룹별 합산부(260)는 복수의 멀티 비트 적용 노드가 포함된 그룹의 개수에 따라 가중합 처리부(예: 제3 가중합 처리부, 제4 가중합 처리부 등)를 추가하거나, 삭제할 수 있다.

[0053] 최종 합산부(270)는 적어도 하나의 멀티 비트 유닛(230) 각각과 연결되어 최종 출력값을 출력한다.

[0054] 최종 합산부(270)는 적어도 하나의 멀티 비트 유닛(230) 각각과 연결되며, 그룹별로 비율을 조정하여 그룹별 합산값을 입력받고, 입력된 그룹별 합산값 각각에 대한 최종 합산값을 산출하며, 각각의 최종 합산값을 합산 처리하여 최종 출력값을 출력한다.

[0055] 최종 합산부(270)는 적어도 하나의 멀티 비트 유닛(230) 각각에서 출력되는 제1 그룹 합산값을 입력받아 제1 최종 합산값을 산출하는 제1 최종 합산라인 및 적어도 하나의 멀티 비트 유닛 각각에서 출력되는 제2 그룹 합산값을 입력받아 제2 최종 합산값을 산출하는 제2 최종 합산라인을 포함한다.

- [0056] 최종 합산부(270)는 적어도 하나의 멀티 비트 유닛(230) 각각과 제1 최종 합산라인 사이에 연결된 제1 최종 스위치와 적어도 하나의 멀티 비트 유닛(230) 각각과 제2 최종 합산라인 사이에 연결된 제2 최종 스위치의 스위칭 동작 제어에 따라 제1 그룹 합산값 및 제2 그룹 합산값을 특정 비율로 조정한다.
- [0057] 예를 들어, 최종 합산부(270)는 멀티 비트 가중치 적용부(240)에서 8 비트의 멀티 비트 가중치 중 4 비트 단위로 가중치를 구분하여 그룹별 합산을 수행한 경우, 제1 그룹 합산값 및 제2 그룹 합산값을 16 : 1 비율로 조정하여 제1 최종 합산값 및 제2 최종 합산값을 산출할 수 있다. 한편, 최종 합산부(270)는 멀티 비트 가중치 적용부(240)에서 8 비트의 멀티 비트 가중치 중 2 비트 단위로 가중치를 구분하여 그룹별 합산을 수행한 경우, 제1 그룹 합산값 및 제2 그룹 합산값을 4 : 1 비율로 조정하여 제1 최종 합산값 및 제2 최종 합산값을 산출할 수 있다.
- [0058] 최종 합산부(270)는 제1 최종 합산라인 및 제2 최종 합산라인 사이에 연결된 제3 최종 스위치의 스위칭 동작 제어를 통해 비율이 조정된 제1 그룹 합산값들을 합산한 제1 최종 합산값과 제2 그룹 합산값들을 합산한 제2 최종 합산값을 합산 처리하여 최종 출력값을 산출하고, 산출된 최종 출력값을 ADC(280)로 출력한다.
- [0059] 도 3은 본 발명의 실시예에 따른 인 메모리 장치의 동작 단계를 설명하기 위한 순서도이다.
- [0060] 인 메모리 장치(200)는 입력 신호에 근거하여 DAC로부터 입력 전압을 입력받아 입력 데이터를 복수의 멀티 비트 적용 노드 각각에 연결된 커패시터에 프리차지한다(프리차지 단계: S310).
- [0061] 인 메모리 장치(200)는 소정의 스위칭 동작을 통해 멀티 비트 가중치의 각 자리에 대응하는 비트 가중치가 입력되면, 멀티 비트 적용 노드 각각에서 입력 데이터에 비트 가중치를 적용하여 가중치 적용값을 출력한다(가중치 계산 단계: S320).
- [0062] 인 메모리 장치(200)는 복수의 멀티 비트 적용 노드를 그룹화하고, 각각의 그룹별로 복수의 멀티 비트 적용 노드 각각에서 출력되는 가중치 적용값을 합산하여 그룹별 합산값을 출력한다(제1 합산 처리 단계: S330). 구체적으로, 인 메모리 장치(200)는 복수의 멀티 비트 적용 노드 중 제1 그룹에 포함된 멀티 비트 적용 노드 각각에서 출력되는 복수의 가중치 적용값을 합산 처리하여 산출된 제1 그룹 합산값을 출력하고, 복수의 멀티 비트 적용 노드 중 제2 그룹에 포함된 멀티 비트 적용 노드 각각에서 출력되는 복수의 가중치 적용값을 합산 처리하여 산출된 제2 그룹 합산값을 출력한다.
- [0063] 인 메모리 장치(200)는 적어도 하나의 멀티 비트 유닛 각각과 연결되며, 그룹별로 비율을 조정하여 그룹별 합산값을 입력받고, 입력된 그룹별 합산값 각각에 대한 최종 합산값을 산출하며, 각각의 최종 합산값을 합산 처리하여 최종 출력값을 출력한다(제2 합산 처리 단계: S330).
- [0064] 도 3에서는 각 단계를 순차적으로 실행하는 것으로 기재하고 있으나, 반드시 이에 한정되는 것은 아니다. 다시 말해, 도 3에 기재된 단계를 변경하여 실행하거나 하나 이상의 단계를 병렬적으로 실행하는 것으로 적용 가능할 것이므로, 도 3은 시계열적인 순서로 한정되는 것은 아니다.
- [0065] 도 3에 기재된 본 실시예에 따른 인 메모리 장치의 동작 방법은 애플리케이션(또는 프로그램)으로 구현되고 단말장치(또는 컴퓨터)로 읽을 수 있는 기록매체에 기록될 수 있다. 본 실시예에 따른 인 메모리 장치의 동작 방법을 구현하기 위한 애플리케이션(또는 프로그램)이 기록되고 단말장치(또는 컴퓨터)가 읽을 수 있는 기록매체는 컴퓨팅 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치 또는 매체를 포함한다.
- [0066] 이하, 도 4 내지 도 7에서는 8 비트의 비트 폭을 가지는 멀티 비트 가중치를 적용하기 위한 인 메모리 장치(200)로 가정하여 설명하도록 한다. 인 메모리 장치(200)에 구비된 커패시터의 개수 또는 합산 처리 단계의 수에 따라 멀티 비트 가중치의 비트 폭은 변경될 수 있다.
- [0067] 도 4는 본 발명의 실시예에 따른 인 메모리 장치의 회로 구성을 나타낸 도면이다.
- [0068] 도 4를 참조하면, 8 비트의 멀티 비트 가중치를 적용하기 위한 인 메모리 장치(200)는 멀티 비트 가중치(W_{1-1} 내지 W_{1-8})를 적용하기 위한 8 개의 멀티 비트 적용 노드 각각에 커패시터가 연결된다. 여기서, 멀티 비트 적용 노드 각각에 커패시터는 서로 상이한 정전용량을 가지거나, 일부 커패시터만 동일한 정전 용량을 가진다.
- [0069] 8 비트의 멀티 비트 가중치 중 상위 비트 가중치(W_{1-1} 내지 W_{1-4})를 적용하기 위한 4 개의 멀티 비트 적용 노드 각각에는 8C, 4C, 2C, 1C와 같이 서로 다른 정전용량의 커패시터가 연결된다.
- [0070] 또한, 8 비트의 멀티 비트 가중치 중 하위 비트 가중치(W_{1-5} 내지 W_{1-8})를 적용하기 위한 4 개의 멀티 비트 적용

노드 각각에는 8C, 4C, 2C, 1C와 같이 서로 다른 정전용량의 커패시터가 연결된다.

- [0071] 본 실시예에 따른 인 메모리 장치(200)는 상위 비트 가중치(W_{1-1} 내지 W_{1-4})를 적용하기 위한 멀티 비트 적용 노드에 128C, 64C, 32C, 16C와 같은 커패시터가 아닌 8C, 4C, 2C, 1C와 같은 커패시터를 연결하여 멀티 비트 가중치를 적용한다. 이러한 커패시터의 연결을 통해 인 메모리 장치(200)의 칩 면적을 최소화하고, 추가적인 커패시터의 사용을 최소화할 수 있다.
- [0072] 또한, 인 메모리 장치(200)는 최종 합산 과정에서 기 연결된 커패시터(8C, 4C, 2C, 1C)를 재사용하여 최종 최종 출력값(V_{SUM})을 ADC로 출력할 수 있다.
- [0073] 도 5는 본 발명의 실시예에 따른 인 메모리 장치의 1차 합산 동작을 설명하기 위한 도면이고, 도 6은 본 발명의 실시예에 따른 인 메모리 장치의 2차 합산 동작을 설명하기 위한 도면이다.
- [0074] 인 메모리 장치(200)는 입력 신호(디지털 코드_6비트)를 입력받고, DAC에서 입력 신호에 따른 입력 전압(V_{in})으로 변환한다.
- [0075] 인 메모리 장치(200)는 입력 전압(V_{in})을 복수의 멀티 비트 적용 노드(멀티 비트의 각 노드) 각각에 인가하여 서로 다른 정전용량을 가지는 커패시터 각각을 충전한다. 즉, 멀티 비트 적용 노드에 연결된 커패시터는 서로 다른 커패시턴스로 충전된다. 여기서, 입력 전압은 입력 데이터로 정의될 수 있다.
- [0076] 인 메모리 장치(200)에서 복수의 멀티 비트 적용 노드는 서로 병렬로 연결되어 있으며, 서로 절연(isolated)되어 있다.
- [0077] 인 메모리 장치(200)는 워드라인을 턴 온 상태로 제어하여 전류원을 켜지게 하며, 공급되는 전류에 따라 커패시터의 동작이 결정된다. 여기서, 전류원은 기 저장된 멀티 비트 가중치 각각(0 또는 1)에 대응하는 전류를 공급한다. 또한, 공급되는 전류는 멀티 비트 가중치의 각 자릿수에 해당하는 비트 가중치일 수 있으며, 0 또는 1의 값을 가질 수 있다.
- [0078] 복수의 멀티 비트 적용 노드 중 전류원에서 공급된 전류가 1일 경우, 해당 멀티 비트 적용 노드의 커패시터는 방전된다. 한편, 복수의 멀티 비트 적용 노드 중 전류원에서 공급된 전류가 0일 경우, 해당 멀티 비트 적용 노드의 커패시터는 현재 상태(충전된 상태)를 유지하는 동작을 수행한다.
- [0079] 이후, 인 메모리 장치(200)는 워드라인을 오프 상태로 제어하여 전류원을 꺼지게 한다.
- [0080] 복수의 멀티 비트 적용 노드 각각의 커패시터들은 제1 전하 공유 노드들과 연결된다. 여기서, 제1 전하 공유 노드들은 그룹별 합산부(260)과 대응된다.
- [0081] 인 메모리 장치(200)는 상위 비트 가중치 그룹에 대한 전하 공유를 통해 제1 가중합 처리하여 상위 4 비트 가중합에 대한 제1 그룹 합산값(V_{MSB})을 출력하고, 하위 비트 가중치 그룹에 대한 전하 공유를 통해 제2 가중합 처리하여 하위 4 비트 가중합에 대한 제2 그룹 합산값(V_{LSB})을 출력한다.
- [0082] 인 메모리 장치(200)는 복수의 멀티 비트 유닛(8 비트 유닛) 각각의 제1 그룹 합산값(V_{MSB})과 연결된 노드 및 제2 그룹 합산값(V_{LSB})과 연결된 노드 각각과 연결되는 스위치를 소정의 비율(예: 16 : 1)로 조정하여 연결되도록 한다.
- [0083] 인 메모리 장치(200)는 소정의 비율로 조정된 제1 그룹 합산값(V_{MSB})을 합산한 제1 최종 합산값과 소정의 비율로 조정된 제2 그룹 합산값(V_{LSB})을 합산한 제2 최종 합산값을 추가 스위칭 제어를 통해 최종 합산 처리를 수행하여 출력 전압(V_{SUM})인 최종 출력값을 ADC로 출력한다.
- [0084] 도 7은 본 발명의 실시예에 따른 인 메모리 장치의 실험결과를 나타낸 도면이다.
- [0085] 종래의 인 메모리 장치는 8 비트의 멀티 비트 가중치를 적용하는 경우 멀티 비트 유닛 각각마다 255C(1, 2, 4, 8, 16, 32, 64, 128C)에 대응하는 커패시터가 필요하다.
- [0086] 한편, 본 발명에 따른 인 메모리 장치(200)는 상위 4 비트와 하위 4 비트를 서로 분리되어 전하 공유되도록 하여, 멀티 비트 유닛 각각마다 30C에 대응하는 커패시터가 필요하다. 이에 따라, 본 발명에 따른 인 메모리 장치(200)는 종래의 인 메모리 장치에 비해 추가되는 커패시터가 현저히 감소(약 88 %)하게 되며 전체 칩 면적(chip

area)도 감소하게 된다.

[0087]

본 발명에 따른 인 메모리 장치(200)는 멀티 가중치를 합산하는 단계를 3 단계로 변경하는 경우 즉, 8 비트의 멀티 비트 가중치를 4 개의 2 비트 그룹으로 분리하여 계층적으로 전하 공유를 수행하는 경우 추가되는 커패시터를 더 감소(약 95 %)하게 될 수 있다.

[0088]

이상의 설명은 본 발명의 실시예의 기술 사상을 예시적으로 설명한 것에 불과한 것으로서, 본 발명의 실시예가 속하는 기술 분야에서 통상의 지식을 가진 자라면 본 발명의 실시예의 본질적인 특성에서 벗어나지 않는 범위에서 다양한 수정 및 변형이 가능할 것이다. 따라서, 본 발명의 실시예들은 본 발명의 실시예의 기술 사상을 한정하기 위한 것이 아니라 설명하기 위한 것이고, 이러한 실시예에 의하여 본 발명의 실시예의 기술 사상의 범위가 한정되는 것은 아니다. 본 발명의 실시예의 보호 범위는 아래의 청구범위에 의하여 해석되어야 하며, 그와 동등한 범위 내에 있는 모든 기술 사상은 본 발명의 실시예의 권리범위에 포함되는 것으로 해석되어야 할 것이다.

부호의 설명

[0089]

200: 인 메모리 장치

210: DAC

220: 메모리 셀 어레이

230: 멀티 비트 유닛

240: 멀티 비트 가중치 적용부

250: 가중치 저장부

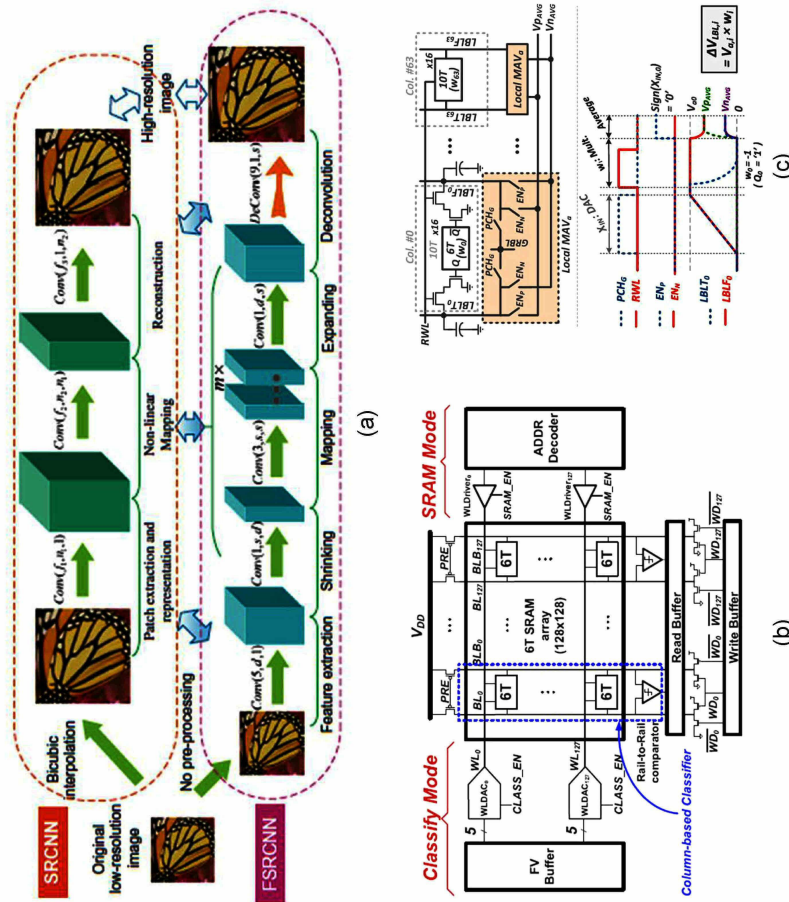
260: 그룹별 합산부

270: 최종 합산부

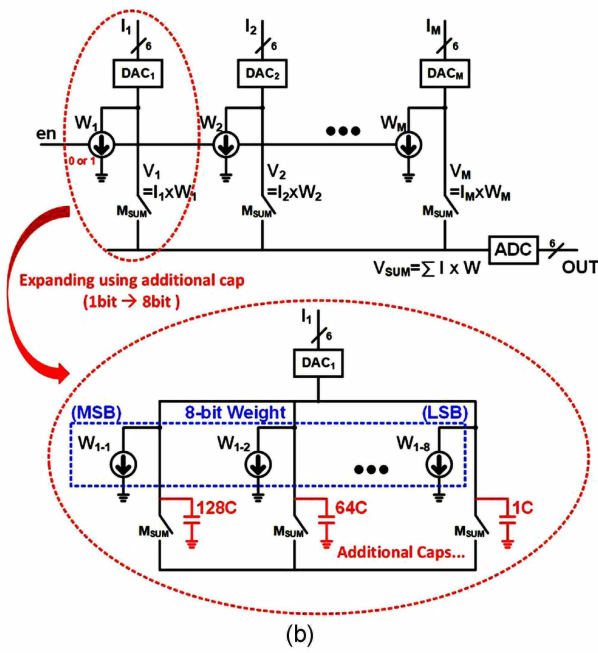
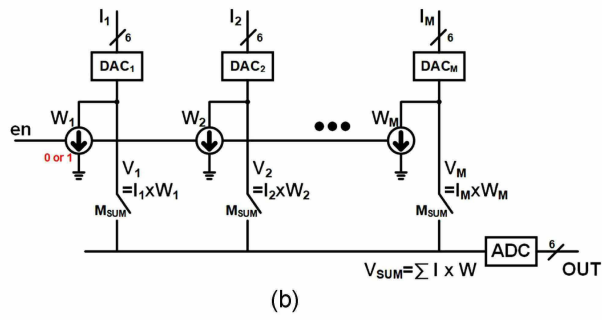
280: ADC

도면

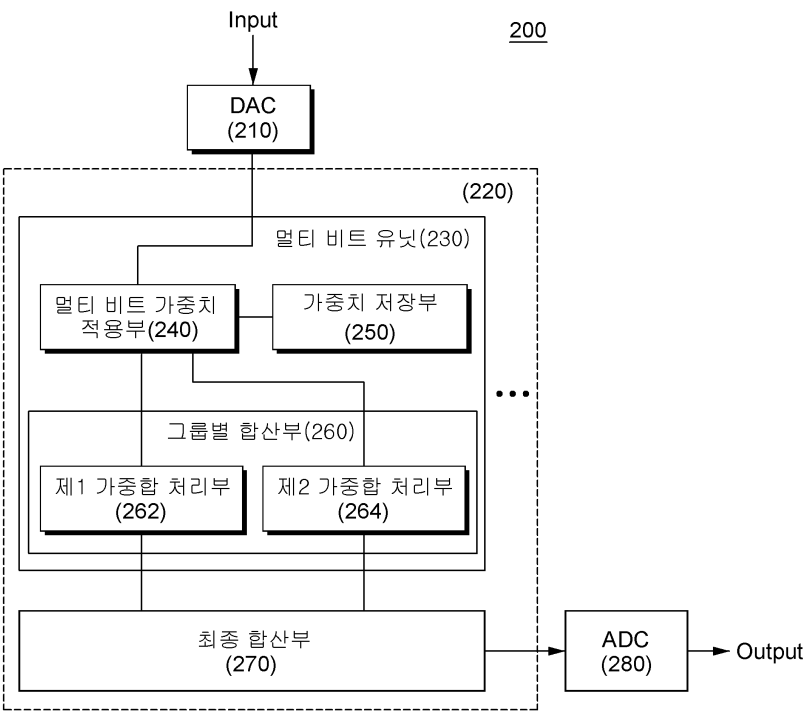
도면 1a



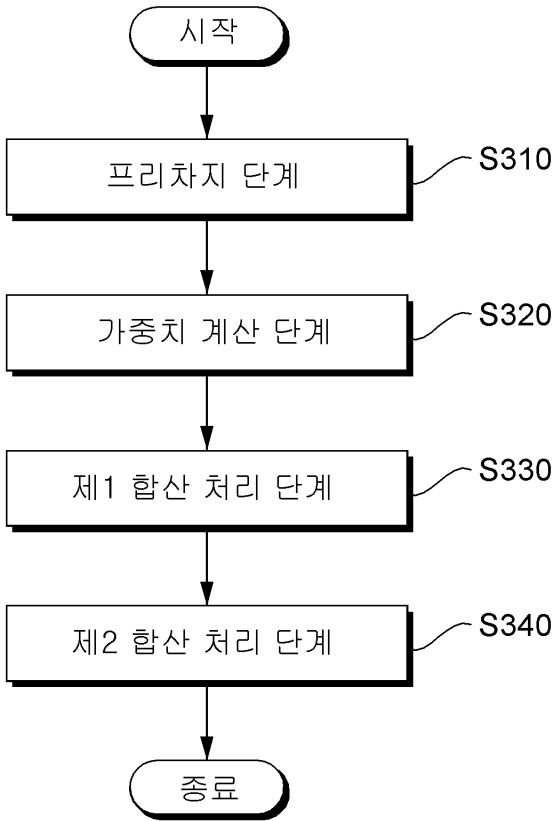
도면1b



도면2

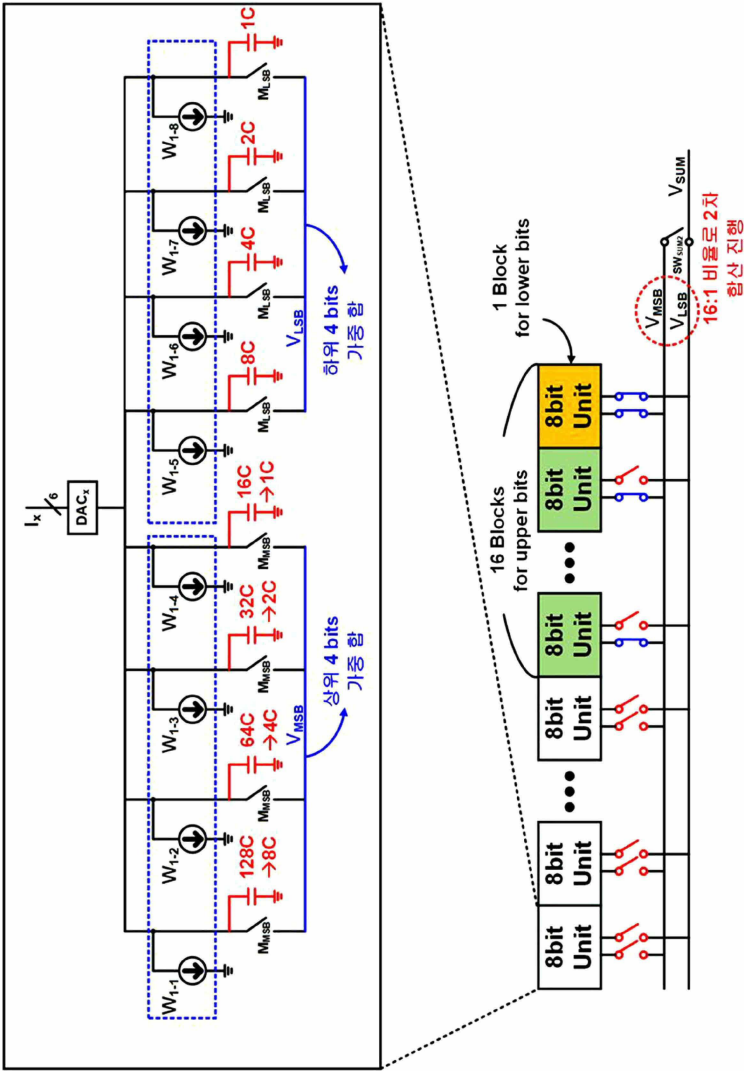


도면3

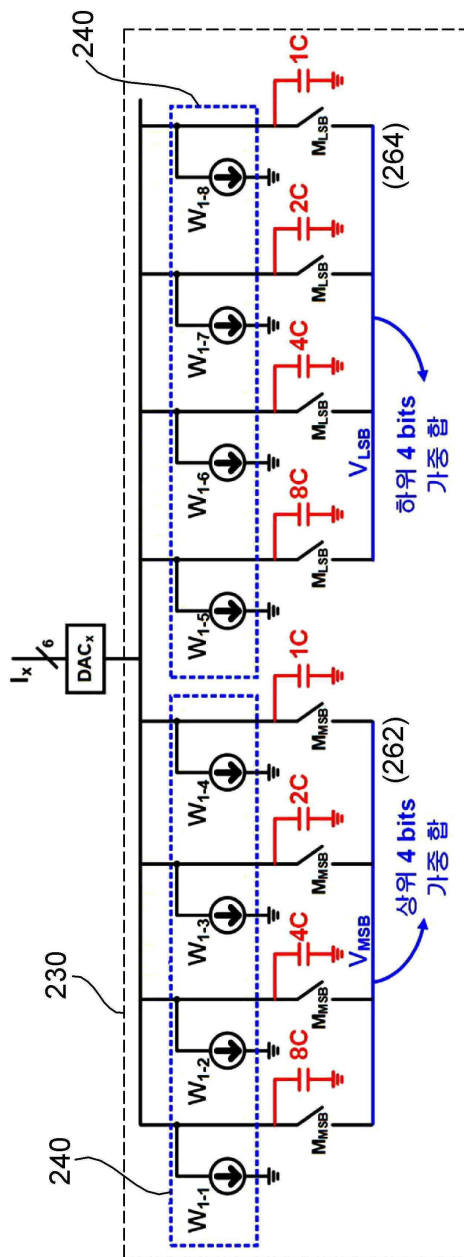


도면4

200

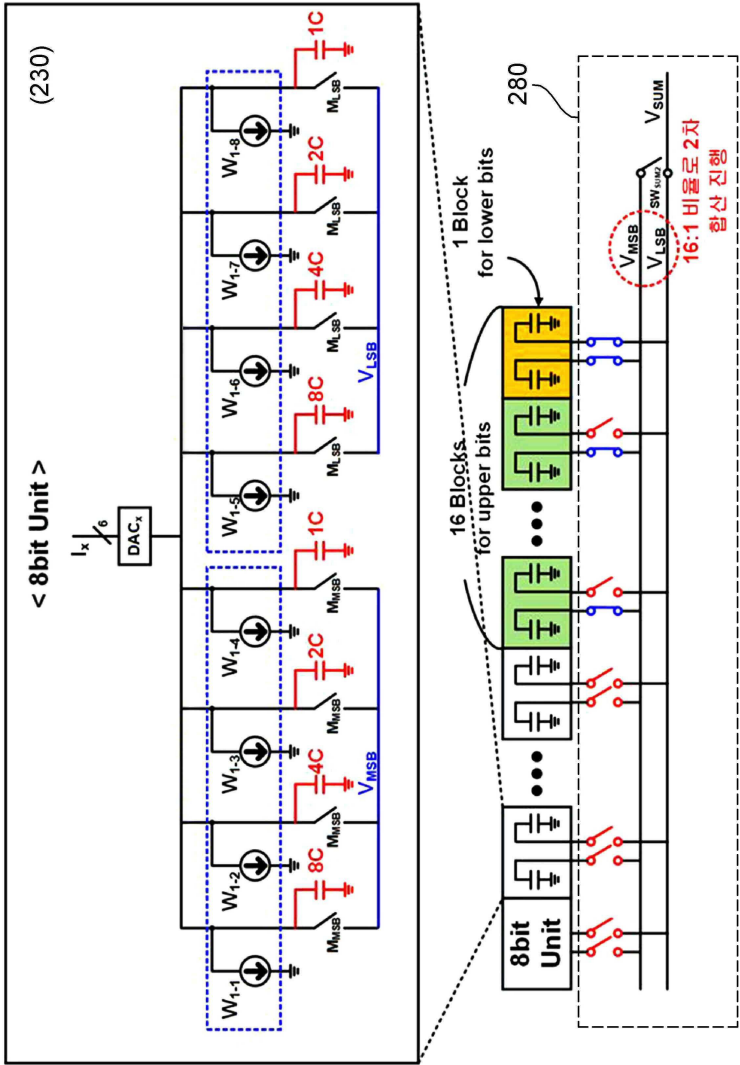


도면5



도면6

200



도면7

