



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0075196
(43) 공개일자 2025년05월28일

(51) 국제특허분류(Int. Cl.)

G06F 11/26 (2006.01) G06F 11/22 (2017.01)
G06F 11/263 (2006.01) G06F 7/58 (2006.01)
G06N 3/045 (2023.01)

(52) CPC특허분류

G06F 11/261 (2013.01)
G06F 11/2263 (2013.01)

(21) 출원번호 10-2023-0162405

(22) 출원일자 2023년11월21일

심사청구일자 2023년11월21일

(71) 출원인

경희대학교 산학협력단

경기도 용인시 기흥구 덕영대로 1732 (서천동, 경희대학교 국제캠퍼스내)

(72) 발명자

홍충선

경기도 용인시 수지구 상현로 30-10, 4813동 101호

보 광 히우

경기도 수원시 영통구 청명로59번길 59-10, 202호
(뒷면에 계속)

(74) 대리인

두호특허법인

전체 청구항 수 : 총 19 항

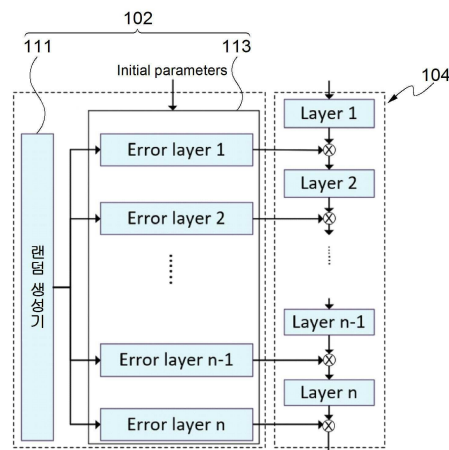
(54) 발명의 명칭 RRAM 오류 에뮬레이터 및 이를 이용한 RRAM 오류 에뮬레이션 방법

(57) 요약

RRAM 오류 에뮬레이터 및 이를 이용한 RRAM 오류 에뮬레이션 방법이 개시된다. 개시되는 일 실시예에 따른 RRAM 오류 에뮬레이터는, RRAM(Resistive Random Access Memory) 장치에서 발생하는 오류의 영향을 모방하기 위한 RRAM 오류 에뮬레이터로서, RRAM의 크로스바 어레이에 구현된 인공 신경망을 포함하는 인공 신경망 모듈 및 오류 노이즈를 생성하고, 생성한 오류 노이즈를 인공 신경망에 주입하는 오류 생성 모듈을 포함한다.

대표도 - 도2

100



(52) CPC특허분류

G06F 11/263 (2013.01)

G06F 7/588 (2013.01)

G06N 3/045 (2023.01)

홍상훈

경기도 용인시 수지구 성북1로163번길 11, 205동 101호

(72) 발명자

김녹원

경기도 용인시 기흥구 서천서로 27, 107동 1302호

김민석

서울특별시 노원구 동일로237길 40, 303호

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193491
과제번호	2019-0-01287-005
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	SW컴퓨팅산업원천기술개발
연구과제명	(SW 스타랩) 분산 엣지를 위한 진화형 딥러닝 모델생성 플랫폼
기 여 율	1/4
과제수행기관명	경희대학교 산학협력단
연구기간	2023.01.01 ~ 2023.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711194179
과제번호	00207816
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	집단연구지원
연구과제명	Meta Federated Learning 기반 Satellite-Air-Ground 통합 네트워킹 시스템 핵심 구

조 개발

기 여 율	1/4
과제수행기관명	경희대학교산학협력단
연구기간	2023.03.01 ~ 2024.02.29

이 발명을 지원한 국가연구개발사업

과제고유번호	1711193622
과제번호	2021-0-02068-003
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	정보통신방송혁신인재양성
연구과제명	인공지능 혁신 허브 연구 개발
기 여 율	1/4
과제수행기관명	고려대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

이 발명을 지원한 국가연구개발사업

과제고유번호	1711179316
과제번호	00155911
부처명	과학기술정보통신부
과제관리(전문)기관명	정보통신기획평가원
연구사업명	인공지능융합혁신인재양성
연구과제명	인공지능융합혁신인재양성(경희대학교)
기 여 율	1/4
과제수행기관명	경희대학교산학협력단
연구기간	2023.01.01 ~ 2023.12.31

명세서

청구범위

청구항 1

RRAM(Resistive Random Access Memory) 장치에서 발생하는 오류의 영향을 모방하기 위한 RRAM 오류 에뮬레이터로서,

상기 RRAM의 크로스바 어레이에 구현된 인공 신경망을 포함하는 인공 신경망 모듈; 및

오류 노이즈를 생성하고, 생성한 상기 오류 노이즈를 상기 인공 신경망에 주입하는 오류 생성 모듈을 포함하는, RRAM 오류 에뮬레이터.

청구항 2

청구항 1에 있어서,

상기 오류 생성 모듈은,

기 설정된 크기의 숫자인 랜덤 값을 생성하는 랜덤 생성기; 및

상기 랜덤 값에 기초하여 오류 노이즈를 생성하고, 상기 생성한 오류 노이즈를 상기 인공 신경망에 주입하는 오류 모델을 포함하며,

상기 오류 노이즈는, 상기 RRAM 장치에서 발생하는 오류를 모방하기 위한 것인, RRAM 오류 에뮬레이터.

청구항 3

청구항 2에 있어서,

상기 인공 신경망은, 복수의 레이어를 포함하고,

상기 오류 모델은, 상기 인공 신경망의 각 레이어에 대응하는 오류 노이즈를 각각 생성하고, 생성한 상기 오류 노이즈를 상기 인공 신경망의 각 레이어에 주입하도록 마련되는, RRAM 오류 에뮬레이터.

청구항 4

청구항 3에 있어서,

상기 인공 신경망은,

상기 RRAM의 각 메모리 셀의 상태를 낮은 저항 상태 또는 높은 저항 상태로 나타내기 위한 바이너리 신경망인, RRAM 오류 에뮬레이터.

청구항 5

청구항 3에 있어서,

상기 인공 신경망은, 소정 레이어의 출력이 다음 레이어로 입력되도록 복수 개의 레이어가 순차적으로 연결되며,

상기 오류 모델은,

상기 오류 노이즈를 상기 인공 신경망의 해당 레이어의 출력에 곱하여져서 다음 레이어로 입력되도록 하는,

RRAM 오류 에뮬레이터.

청구항 6

청구항 5에 있어서,

상기 오류 모델은,

상기 랜덤 값에 기초하여 상기 RRAM의 각 비트 라인의 실제 전류를 산출하고, 상기 비트 라인의 실제 전류와 상기 비트 라인의 이상적인 전류 간의 차이에 기초하여 상기 오류 노이즈를 생성하는, RRAM 오류 에뮬레이터.

청구항 7

청구항 6에 있어서,

상기 오류 모델은,

상기 비트 라인의 실제 전류를 하기 수학식에 의해 산출하는, RRAM 오류 에뮬레이터.

(수학식)

$$I_j = A_j N_g(0,1) + B_j$$

$$A_j = \sigma V_j a_j m$$

$$B_j = a_j V_j m \left(R_{d_id}^j + \frac{R_w(m-1)}{2} \right) + m b_j V_j$$

I_j : j번째 비트 라인의 실제 전류

$N_g(0,1)$: 랜덤 생성기에서 생성한 랜덤 값

σ : RRAM 장치에서 노이즈의 영향을 나타내는 스케일 팩터

V_j : j번째 비트 라인의 전압

m: 워드 라인의 개수

$R_{d_id}^j$: j번째 비트 라인의 이상적인 등가 저항

R_w : 비트 라인의 와이어 저항

a_j, b_j : j번째 비트 라인의 실제 등가 저항을 나타내기 위해 기 설정되는 근사 계수(approximation coefficient)

청구항 8

청구항 6에 있어서,

상기 오류 모델은,

상기 비트 라인의 실제 전류와 상기 비트 라인의 이상적인 전류 간의 차이에 기초하여 해당 비트 라인의 읽기 정확도를 산출하고, 상기 산출한 읽기 정확도를 오류 노이즈로 설정하는, RRAM 오류 에뮬레이터.

청구항 9

청구항 8에 있어서,

상기 오류 모델은,

상기 비트 라인의 읽기 정확도를 하기 수학식에 의해 산출하는, RRAM 오류 에뮬레이터.

(수학식)

$$read.accuracy_j = \frac{|I_{ideal}^j - I_j|}{I_{ideal}^j}$$

I_j : j번째 비트 라인의 실제 전류

I_{ideal}^j : j번째 비트 라인의 이상적인 전류

청구항 10

하나 이상의 프로세서들, 및

상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되고, RRAM(Resistive Random Access Memory) 장치에서 발생하는 오류의 영향을 모방하기 위한 RRAM 오류 에뮬레이션 방법으로서,

오류 노이즈를 생성하는 단계; 및

상기 생성한 오류 노이즈를 상기 RRAM의 크로스바 어레이에 구현된 인공 신경망에 주입하는 단계를 포함하는, RRAM 오류 에뮬레이션 방법.

청구항 11

청구항 10에 있어서,

상기 오류 노이즈를 생성하는 단계는,

기 설정된 크기의 숫자인 랜덤 값을 생성하는 단계; 및

상기 랜덤 값에 기초하여 오류 노이즈를 생성하는 단계를 포함하고,

상기 오류 노이즈는, 상기 RRAM 장치에서 발생하는 오류를 모방하기 위한 것인, RRAM 오류 에뮬레이션 방법.

청구항 12

청구항 11에 있어서,

상기 인공 신경망은, 복수의 레이어를 포함하고,

상기 오류 노이즈를 생성하는 단계는, 상기 인공 신경망의 각 레이어에 대응하는 오류 노이즈를 각각 생성하고,

상기 인공 신경망에 주입하는 단계는, 생성한 상기 오류 노이즈를 상기 인공 신경망의 각 레이어에 주입하는, RRAM 오류 에뮬레이션 방법.

청구항 13

청구항 12에 있어서,

상기 인공 신경망은,

상기 RRAM의 각 메모리 셀의 상태를 낮은 저항 상태 또는 높은 저항 상태로 나타내기 위한 바이너리 신경망인, RRAM 오류 에뮬레이션 방법.

청구항 14

청구항 12에 있어서,

상기 인공 신경망은, 소정 레이어의 출력이 다음 레이어로 입력되도록 복수 개의 레이어가 순차적으로 연결되며,

상기 인공 신경망에 주입하는 단계는, 상기 오류 노이즈를 상기 인공 신경망의 해당 레이어의 출력에 곱하여져서 다음 레이어로 입력되도록 하는, RRAM 오류 에뮬레이션 방법.

청구항 15

청구항 14에 있어서,

상기 오류 노이즈를 생성하는 단계는,

상기 랜덤 값에 기초하여 상기 RRAM의 각 비트 라인의 실제 전류를 산출하는 단계; 및

상기 비트 라인의 실제 전류와 상기 비트 라인의 이상적인 전류 간의 차이에 기초하여 상기 오류 노이즈를 생성하는 단계를 포함하는, RRAM 오류 에뮬레이션 방법.

청구항 16

청구항 15에 있어서,

상기 비트 라인의 실제 전류는 하기 수학식에 의해 산출하는, RRAM 오류 에뮬레이션 방법.

(수학식)

$$I_j = A_j N_g(0,1) + B_j$$

$$A_j = \sigma V_j a_j m$$

$$B_j = a_j V_j m \left(R_{d_id}^j + \frac{R_w(m-1)}{2} \right) + m b_j V_j$$

I_j : j번째 비트 라인의 실제 전류

$N_g(0,1)$: 랜덤 생성기에서 생성한 랜덤 값

σ : RRAM 장치에서 노이즈의 영향을 나타내는 스케일 팩터

V_j : j번째 비트 라인의 전압

m: 워드 라인의 개수

$R_{d_id}^j$: j번째 비트 라인의 이상적인 등가 저항

R_w : 비트 라인의 와이어 저항

a_j, b_j : j번째 비트 라인의 실제 등가 저항을 나타내기 위해 기 설정되는 근사 계수(approximation coefficient)

청구항 17

청구항 15에 있어서,

상기 오류 노이즈를 생성하는 단계는,

상기 비트 라인의 실제 전류와 상기 비트 라인의 이상적인 전류 간의 차이에 기초하여 해당 비트 라인의 읽기 정확도를 산출하는 단계; 및

상기 산출한 읽기 정확도를 오류 노이즈로 설정하는 단계를 포함하는, RRAM 오류 에뮬레이션 방법.

청구항 18

청구항 17에 있어서,

상기 비트 라인의 읽기 정확도는 하기 수학적식에 의해 산출하는, RRAM 오류 에뮬레이션 방법.

(수학적식)

$$read.accuracy_j = \frac{|I_{ideal}^j - I_j|}{I_{ideal}^j}$$

I_j : j번째 비트 라인의 실제 전류

I_{ideal}^j : j번째 비트 라인의 이상적인 전류

청구항 19

비일시적 컴퓨터 판독 가능한 저장 매체(non-transitory computer readable storage medium)에 저장된 컴퓨터 프로그램으로서,

상기 컴퓨터 프로그램은, RRAM(Resistive Random Access Memory) 장치에서 발생하는 오류의 영향을 모방하기 위한 RRAM 오류 에뮬레이션을 수행하도록 하는 하나 이상의 명령어들을 포함하고, 상기 명령어들은 하나 이상의 프로세서들을 갖는 컴퓨팅 장치에 의해 실행될 때, 상기 컴퓨팅 장치로 하여금,

오류 노이즈를 생성하는 단계; 및

상기 생성한 오류 노이즈를 상기 RRAM의 크로스바 어레이에 구현된 인공 신경망에 주입하는 단계를 수행하도록 하는, 비일시적 컴퓨터 판독 가능한 저장 매체에 저장된 컴퓨터 프로그램.

발명의 설명

기술 분야

[0001] 본 발명의 실시예는 RRAM 오류 에뮬레이션 기술과 관련된다.

배경 기술

- [0003] RRAM(Resistive Random Access Memory) 하드웨어 플랫폼 기반 크로스바 어레이(crossbar arrays)는 벡터-행렬 곱셈을 수행하는 기능을 제공하면서 성능이나 스토리지 밀도 및 확장성 등에 있어 이전 세대 메모리보다 우수하다. 이에 RRAM 기반 하드웨어 솔루션은 뉴로모픽 컴퓨팅이나 인공 신경망(Artificial Neural Network: ANN) 등에 사용되고 있다. 그러나, 인공 신경망 계산 시 무작위 노이즈(random noise) 및 고감도(high sensitivity)와 같은 작동 변형은 RRAM 장치에 구현된 아키텍처에서 여전히 어려운 문제이다.
- [0004] 도 1은 일반적인 RRAM 크로스바 어레이를 개략적으로 나타낸 도면이다. 도 1에서 비트 라인(bit line)과 워드 라인(word line)의 각 교차점에 RRAM 메모리 셀이 위치하게 된다. 그리고, 인공 신경망의 가중치는 셀 컨덕턴스로 구성될 수 있다. 행렬 곱셈을 수행할 때 RRAM은 입력 전압에 따라 저항 상태를 변경하여 전자 시냅스(electronic synapse)로 작동한다. 모든 워드 라인은 읽기 전압의 진폭으로 표시되는 벡터에 의해 바이어스되며, 여기에 가중치 맵을 나타내는 컨덕턴스 행렬이 추가로 곱해진다.
- [0005] 이상적으로는 모든 워드 라인에 특정 전압이 주어지고, 이 전압은 모든 메모리 셀에 걸쳐 균일하다. 그러나, 실제 RRAM 크로스바 어레이에서는 기생 와이어 저항으로 인해, 전원에서 멀리 떨어진 노드(즉, 메모리 셀)의 전압이 떨어지게 된다. 따라서, 전압 분포가 일관되지 않게 되고, 이는 모든 비트 라인에 대한 출력 전류의 분포 감소로 이어지게 된다.
- [0006] 이 경우, RRAM 기반 크로스바 어레이에서 구현된 인공 신경망의 정확도가 크게 저하되는 바, 회로 수준의 내부 확률(internal stochasticity) 및 기생(예를 들어, IR 드롭, 읽기 노이즈, 쓰기 노이즈 등)으로 인한 오류를 모방할 수 있는 방안이 요구된다.

선행기술문헌

특허문헌

- [0008] (특허문헌 0001) 한국등록특허공보 제10-2141385호(2020.08.05)

발명의 내용

해결하려는 과제

- [0009] 본 발명의 실시예는 RRAM 장치에서 발생하는 오류를 모방할 수 있도록 하는 RRAM 오류 에뮬레이터 및 이를 이용한 RRAM 오류 에뮬레이션 방법을 제공하기 위한 것이다.

과제의 해결 수단

- [0011] 개시되는 일 실시예에 따른 RRAM 오류 에뮬레이터는, RRAM(Resistive Random Access Memory) 장치에서 발생하는 오류의 영향을 모방하기 위한 RRAM 오류 에뮬레이터로서, 상기 RRAM의 크로스바 어레이에 구현된 인공 신경망을 포함하는 인공 신경망 모듈; 및 오류 노이즈를 생성하고, 생성한 상기 오류 노이즈를 상기 인공 신경망에 주입하는 오류 생성 모듈을 포함한다.
- [0012] 상기 오류 생성 모듈은, 기 설정된 크기의 숫자인 랜덤 값을 생성하는 랜덤 생성기; 및 상기 랜덤 값에 기초하여 오류 노이즈를 생성하고, 상기 생성한 오류 노이즈를 상기 인공 신경망에 주입하는 오류 모델을 포함하며, 상기 오류 노이즈는, 상기 RRAM 장치에서 발생하는 오류를 모방하기 위한 것일 수 있다.
- [0013] 상기 인공 신경망은, 복수의 레이어를 포함하고, 상기 오류 모델은, 상기 인공 신경망의 각 레이어에 대응하는 오류 노이즈를 각각 생성하고, 생성한 상기 오류 노이즈를 상기 인공 신경망의 각 레이어에 주입하도록 마련될 수 있다.
- [0014] 상기 인공 신경망은, 상기 RRAM의 각 메모리 셀의 상태를 낮은 저항 상태 또는 높은 저항 상태로 나타내기 위한 바이너리 신경망일 수 있다.

[0015] 상기 인공 신경망은, 소정 레이어의 출력이 다음 레이어로 입력되도록 복수 개의 레이어가 순차적으로 연결되며, 상기 오류 모델은, 상기 오류 노이즈를 상기 인공 신경망의 해당 레이어의 출력에 곱하여져서 다음 레이어로 입력되도록 할 수 있다.

[0016] 상기 오류 모델은, 상기 랜덤 값에 기초하여 상기 RRAM의 각 비트 라인의 실제 전류를 산출하고, 상기 비트 라인의 실제 전류와 상기 비트 라인의 이상적인 전류 간의 차이에 기초하여 상기 오류 노이즈를 생성할 수 있다.

[0017] 상기 오류 모델은, 상기 비트 라인의 실제 전류를 하기 수학식에 의해 산출할 수 있다.

[0018] (수학식)

$$I_j = A_j N_g(0,1) + B_j$$

[0019]

$$A_j = \sigma V_j a_j m$$

[0020]

$$B_j = a_j V_j m \left(R_{d_id}^j + \frac{R_w(m-1)}{2} \right) + m b_j V_j$$

[0021]

[0022] I_j : j번째 비트 라인의 실제 전류

[0023] $N_g(0,1)$: 랜덤 생성기에서 생성한 랜덤 값

[0024] σ : RRAM 장치에서 노이즈의 영향을 나타내는 스케일 팩터

[0025] V_j : j번째 비트 라인의 전압

[0026] m: 워드 라인의 개수

[0027] $R_{d_id}^j$: j번째 비트 라인의 이상적인 등가 저항

[0028] R_w : 비트 라인의 와이어 저항

[0029] a_j, b_j : j번째 비트 라인의 실제 등가 저항을 나타내기 위해 기 설정되는 근사 계수(approximation coefficient)

[0030] 상기 오류 모델은, 상기 비트 라인의 실제 전류와 상기 비트 라인의 이상적인 전류 간의 차이에 기초하여 해당 비트 라인의 읽기 정확도를 산출하고, 상기 산출한 읽기 정확도를 오류 노이즈로 설정할 수 있다.

[0031] 상기 오류 모델은, 상기 비트 라인의 읽기 정확도를 하기 수학식에 의해 산출할 수 있다.

[0032] (수학식)

$$read.accuracy_j = \frac{|I_{ideal}^j - I_j|}{I_{ideal}^j}$$

[0033]

[0034] I_j : j번째 비트 라인의 실제 전류

[0035] I_{ideal}^j : j번째 비트 라인의 이상적인 전류

[0036] 개시되는 일 실시예에 따른 RRAM 오류 에뮬레이션 방법은, 하나 이상의 프로세서들, 및 상기 하나 이상의 프로세서들에 의해 실행되는 하나 이상의 프로그램들을 저장하는 메모리를 구비한 컴퓨팅 장치에서 수행되고, RRAM(Resistive Random Access Memory) 장치에서 발생하는 오류의 영향을 모방하기 위한 RRAM 오류 에뮬레이션 방법으로서, 오류 노이즈를 생성하는 단계; 및 상기 생성한 오류 노이즈를 상기 RRAM의 크로스바 어레이에 구현된 인공 신경망에 주입하는 단계를 포함한다.

발명의 효과

[0038] 개시되는 실시예에 의하면, RRAM의 내부 확률적 원인과 기생 문제로 인해 발생하는 오류를 모방하도록 함으로써, 오류로 인한 전압 강하 또는 전류 누출이 인공 신경망의 성능에 미치는 영향을 정확하게 추정할 수 있어, RRAM 기반의 인공 신경망의 성능 및 정확도를 높일 수 있게 된다.

도면의 간단한 설명

[0040] 도 1은 일반적인 RRAM 크로스바 어레이를 개략적으로 나타낸 도면

도 2는 본 발명의 일 실시예에 따른 인공 신경망 기반의 RRAM 오류 에뮬레이터(error emulator)를 개략적으로 나타낸 도면

도 3은 본 발명의 일 실시예에 따른 오류 생성 모듈에서 인공 신경망의 각 층에 주입할 오류 노이즈를 생성하는 과정을 개략적으로 나타낸 도면

도 4는 본 발명의 일 실시예에 따른 오류 모델에서 읽기 정확도를 산출하는 과정을 나타낸 도면

도 5는 본 발명의 일 실시예에 따른 RRAM 오류 에뮬레이션 방법을 나타낸 흐름도

도 6은 예시적인 실시예들에서 사용되기에 적합한 컴퓨팅 장치를 포함하는 컴퓨팅 환경을 예시하여 설명하기 위한 블록도

발명을 실시하기 위한 구체적인 내용

[0041] 이하, 도면을 참조하여 본 발명의 구체적인 실시형태를 설명하기로 한다. 이하의 상세한 설명은 본 명세서에서 기술된 방법, 장치 및/또는 시스템에 대한 포괄적인 이해를 돕기 위해 제공된다. 그러나 이는 예시에 불과하며 본 발명은 이에 제한되지 않는다.

[0042] 본 발명의 실시예들을 설명함에 있어서, 본 발명과 관련된 공지기술에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략하기로 한다. 그리고, 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서 이는 사용자, 운용자의 의도 또는 관례 등에 따라 달라질 수 있다. 그러므로 그 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다. 상세한 설명에서 사용되는 용어는 단지 본 발명의 실시예들을 기술하기 위한 것이며, 결코 제한적이어서는 안 된다. 명확하게 달리 사용되지 않는 한, 단수 형태의 표현은 복수 형태의 의미를 포함한다. 본 설명에서, "포함" 또는 "구비"와 같은 표현은 어떤 특성들, 숫자들, 단계들, 동작들, 요소들, 이들의 일부 또는 조합을 가리키기 위한 것이며, 기술된 것 이외에 하나 또는 그 이상의 다른 특성, 숫자, 단계, 동작, 요소, 이들의 일부 또는 조합의 존재 또는 가능성을 배제하도록 해석되어서는 안 된다.

[0043] 또한, 제1, 제2 등의 용어는 다양한 구성 요소들을 설명하는데 사용될 수 있지만, 상기 구성 요소들은 상기 용어들에 의해 한정되어서는 안 된다. 상기 용어들은 하나의 구성 요소를 다른 구성 요소로부터 구별하는 목적으로 사용될 수 있다. 예를 들어, 본 발명의 권리 범위를 벗어나지 않으면서 제1 구성 요소는 제2 구성 요소로 명명될 수 있고, 유사하게 제2 구성 요소도 제1 구성 요소로 명명될 수 있다.

[0045] 도 2는 본 발명의 일 실시예에 따른 인공 신경망 기반의 RRAM 오류 에뮬레이터(error emulator)를 개략적으로 나타낸 도면이고, 도 3은 본 발명의 일 실시예에 따른 오류 생성 모듈에서 인공 신경망의 각 층에 주입할 오류 노이즈를 생성하는 과정을 개략적으로 나타낸 도면이다.

[0046] 도 2 및 도 3을 참조하면, RRAM 오류 에뮬레이터(100)는 오류 생성 모듈(102) 및 인공 신경망 모듈(104)을 포함할 수 있다. RRAM 오류 에뮬레이터(100)는 RRAM(Resistive Random Access Memory) 장치에서 내부의 확률적 원인 및 기생 문제로 인해 발생하는 오류의 영향을 모방하기 위한 것일 수 있다. 여기서, RRAM 장치는 RRAM 크로스바 어레이를 포함할 수 있다.

[0047] 오류 생성 모듈(102)은 인공 신경망 모듈(104)로 주입되는 오류 노이즈를 생성할 수 있다. 오류 생성 모듈(102)은 랜덤 생성기(111) 및 오류 모델(113)을 포함할 수 있다.

- [0048] 랜덤 생성기(111)는 기 설정된 크기의 숫자를 무작위로 생성할 수 있다. 일 실시예에서, 랜덤 생성기(111)는 16 비트의 숫자를 무작위로 생성할 수 있다. 랜덤 생성기(111)는 생성 과정의 분포가 가우스 정규 분포가 되도록 16 비트의 숫자를 무작위로 생성할 수 있다.
- [0049] 랜덤 생성기(111)는 난수 생성기(111a) 및 분포 반전기(111b)를 포함할 수 있다. 난수 생성기(111a)는 기 설정된 난수 생성 알고리즘에 따라 균일 분포의 16비트 난수를 생성할 수 있다. 분포 반전기(111b)는 생성된 난수에 ICDF(Inversion of Cumulative Distribution Function)를 적용하여 가우스 정규 분포의 반전을 랜덤 값($N_g(0,1)$)으로 출력할 수 있다.
- [0050] 오류 모델(113)은 랜덤 생성기(111)에서 생성된 무작위 값(랜덤 값)에 기초하여 오류 노이즈를 생성할 수 있다. 여기서, 오류 노이즈는 RRAM 크로스바 어레이에서 발생하는 내부의 확률적 원인 및 기생 문제(예를 들어, 기생 와이어 저항 또는 기생 커패시터 등)로 인해 발생하는 오류를 모방하기 위한 것일 수 있다. 오류 모델(113)은 기 설정되는 클록 주기마다 오류 노이즈를 생성할 수 있다.
- [0051] 한편, 인공 신경망 모듈(104)은 RRAM(Resistive Random Access Memory)의 크로스바 어레이에 구현된 인공 신경망을 포함할 수 있다. 일 실시예에서, 인공 신경망 모듈(104)은 RRAM 기반의 인공 신경망 가속기일 수 있다. 즉, 인공 신경망 모듈(104)은 RRAM에 기반하여 애플리케이션의 연산을 가속화 하도록 하기 위한 것일 수 있다.
- [0052] 일 실시예에서, 인공 신경망은 RRAM의 각 메모리 셀의 상태를 R_{LRS} (낮은 저항 상태, Low Resistance State) 또는 R_{HRS} (높은 저항 상태, High Resistance State)로 나타내기 위해 바이너리 신경망(Binary Neural Network)을 사용할 수 있다. 여기서, 낮은 저항 상태 또는 높은 저항 상태는 기 설정된 기준 저항을 준으로 구분될 수 있다. 즉, RRAM의 메모리 셀의 저항이 기 설정된 기준 저항 보다 낮으면 낮은 저항 상태인 것으로 하고, 메모리 셀의 저항이 기 설정된 기준 저항 보다 높으면 높은 저항 상태인 것으로 할 수 있다.
- [0053] 인공 신경망 모듈(104)은 다층의 인공 신경망을 포함할 수 있다. 인공 신경망의 각 레이어는 지속적으로 데이터를 처리하고 매 클록 주기마다 출력을 생성할 수 있다. 인공 신경망의 복수 개의 레이어는 순차적으로 연결될 수 있다. 즉, 인공 신경망의 각 레이어에서 출력되는 데이터는 다음 레이어로 입력되도록 연결될 수 있다. 이때, 인공 신경망의 레이어는 입력 데이터에 컨볼루션을 수행하는 컨볼루션 레이어일 수 있다.
- [0054] 여기서, 오류 모델(113)은 인공 신경망 모듈(104)에 포함된 인공 신경망의 각 계층(layer)에 대응하는 오류 노이즈를 각각 생성하고, 생성한 오류 노이즈를 인공 신경망 모듈(104)의 인공 신경망의 각 계층에 주입할 수 있다. 오류 모델(113)은 인공 신경망의 각 계층에 주입할 오류 노이즈를 동시에 생성할 수 있다. 오류 모델(113)은 인공 신경망의 각 계층에 대응하여 오류 노이즈를 생성하는 오류 계층(error layer)들을 포함할 수 있다.
- [0055] 일 실시예에서, 각 오류 노이즈는 인공 신경망의 각 컨볼루션 레이어의 출력에 곱해져서 다음 레이어로 주입될 수 있다. 각 컨볼루션 레이어의 출력은 주입된 오류 노이즈와 곱하여져서 컨볼루션 작업에서 행렬 곱셈 중 발생하는 오류의 영향을 에뮬레이트(emulate) 하게 된다.
- [0056] 오류 모델(113)은 랜덤 생성기(111)에서 생성된 무작위 값(랜덤 값)에 기초하여 j번째 비트 라인의 실제 전류(I_j)를 하기 수학적 식 1에 의해 산출할 수 있다.
- [0057] (수학적 식 1)
- $$I_j = A_j N_g(0,1) + B_j$$
- [0058]
- [0059] 여기서, $A_j = \sigma V_j a_j m$ 이고, $B_j = a_j V_j m \left(R_{d_id}^j + \frac{R_w(m-1)}{2} \right) + m b_j V_j$ 이다.
- [0060] $N_g(0,1)$ 는 랜덤 생성기(111)에서 생성한 랜덤 값을 나타낸다. σ 는 RRAM 장치에서 노이즈의 영향을 나타내는 스케일 팩터이다. V_j 는 j번째 비트 라인의 전압이다. m은 워드 라인의 개수를 나타낸다. $R_{d_id}^j$ 는 j번째 비트 라인의 이상적인 등가 저항을 나타낸다. R_w 는 각 비트 라인의 와이어 저항을 나타낸다. a_j 및 b_j 는 j번째 비트 라인의 실제 등가 저항을 나타내기 위해 기 설정되는 근사 계수(approximation coefficient)이다.

[0061] 오류 모델(113)은 각 비트 라인에서 실제 전류와 이상적인 전류 간의 차이를 산출할 수 있다. 오류 모델(113)은 각 비트 라인에서 실제 전류와 이상적인 전류 간의 차이를 해당 비트 라인의 오류 노이즈로 설정할 수 있다.

[0062] 오류 모델(113)은 각 비트 라인의 실제 전류와 이상적인 전류 간의 차이에 기초하여 해당 비트 라인의 읽기 정확도(read accuracy)를 산출할 수 있다. 오류 모델(113)은 비트 라인에 대한 읽기 정확도를 해당 비트 라인의 오류 노이즈로 설정할 수도 있다. 오류 모델(113)은 하기 수학적 식 2를 통해 j번째 비트 라인에 대한 읽기 정확도를 산출할 수 있다.

[0063] (수학적 식 2)

$$read_accuracy_j = \frac{|I_{ideal}^j - I_j|}{I_{ideal}^j}$$

[0064] I_j : j번째 비트 라인의 실제 전류

[0065] I_{ideal}^j : j번째 비트 라인의 이상적인 전류

[0066] 도 4는 본 발명의 일 실시예에 따른 오류 모델에서 읽기 정확도(오류 노이즈)를 산출하는 과정을 나타낸 도면이다. 도 4를 참조하면, 오류 모델(113)은 초기 전압과 전류를 기준으로 쓰기 노이즈(ϵ_{WN}) 및 읽기 노이즈(ϵ_{RN})에 영향을 받는 새로운 R_{LRS} (낮은 저항 상태, Low Resistance State), R_{HRS} (높은 저항 상태, High Resistance State)를 각각 계산할 수 있다. 여기서, R_{LRS} , R_{HRS} 는 RRAM 크로스바 어레이의 노드 상태 분포로서, 각각 균일한 분포를 가진다.

[0067] 오류 모델(113)은 균일한 분포를 갖는 R_{LRS} , R_{HRS} 를 기반으로 RRAM 크로스바 어레이의 모든 노드(즉, 메모리 셀)의 등가 저항($R_{eq}^{i,j}$)과 전압(V_{ij})을 순차적으로 계산할 수 있다. 다음으로, 오류 모델(113)은 모든 노드에서 노이즈 효과와 전압이 포함된 컨덕턴스 맵(conductance map)(G_{ij})을 계산할 수 있다. 컨볼루션 출력을 나타내는 각

[0068] 비트 라인에 대해 예시된 출력 전류는
$$I_j = \sum_{i=1}^M G_{ji} V_{ij}$$
와 같이 컨덕턴스 맵과 해당 전압으로부터 유도될 수 있다.

[0069] 한편, 각 비트 라인의 이상적인 출력 전류($I_{o_ideal}^j$)는 균일한 분포의 R_{LRS} 및 R_{HRS} 에 기초하여 RRAM 교차점의 저항을 통해 직접 계산될 수 있다. 마지막으로, 오류 모델(113)은 각 비트 라인의 이상적인 전류와 실제 전류에 기초하여 해당 비트 라인의 읽기 정확도를 산출할 수 있다.

[0070] 개시되는 실시예에 의하면, RRAM의 내부 확률적 원인과 기생 문제로 인해 발생하는 오류를 모방하도록 함으로써, 오류로 인한 전압 강하 또는 전류 누출이 인공 신경망의 성능에 미치는 영향을 정확하게 추정할 수 있어, RRAM 기반의 인공 신경망의 성능 및 정확도를 높일 수 있게 된다.

[0071] 본 명세서에서 모듈이라 함은, 본 발명의 기술적 사상을 수행하기 위한 하드웨어 및 상기 하드웨어를 구동하기 위한 소프트웨어의 기능적, 구조적 결합을 의미할 수 있다. 예컨대, 상기 "모듈"은 소정의 코드와 상기 소정의 코드가 수행되기 위한 하드웨어 리소스의 논리적인 단위를 의미할 수 있으며, 반드시 물리적으로 연결된 코드를 의미하거나, 한 종류의 하드웨어를 의미하는 것은 아니다.

[0072] 도 5는 본 발명의 일 실시예에 따른 RRAM 오류 에뮬레이션 방법을 나타낸 흐름도이다. 도시된 흐름도에서는 상기 방법을 복수 개의 단계로 나누어 기재하였으나, 적어도 일부의 단계들은 순서를 바꾸어 수행되거나, 다른 단계와 결합되어 함께 수행되거나, 생략되거나, 세부 단계들로 나뉘어 수행되거나, 또는 도시되지 않은 하나 이상의 단계가 추가되어 수행될 수 있다.

[0073] 도 5를 참조하면, RRAM 오류 에뮬레이터(100)는 기 설정된 크기의 숫자인 랜덤 값을 생성할 수 있다(S 101). 다음으로, RRAM 오류 에뮬레이터(100)는 생성한 랜덤 값에 기초하여 오류 노이즈를 생성할 수 있다(S 103). 다음

으로, RRAM 오류 에뮬레이터(100)는 생성한 오류 노이즈를 인공 신경망의 각 레이어에 주입할 수 있다(S 105).

- [0076] 도 6은 예시적인 실시예들에서 사용되기에 적합한 컴퓨팅 장치를 포함하는 컴퓨팅 환경(10)을 예시하여 설명하기 위한 블록도이다. 도시된 실시예에서, 각 컴포넌트들은 이하에 기술된 것 이외에 상이한 기능 및 능력을 가질 수 있고, 이하에 기술된 것 이외에도 추가적인 컴포넌트를 포함할 수 있다.
- [0077] 도시된 컴퓨팅 환경(10)은 컴퓨팅 장치(12)를 포함한다. 일 실시예에서, 컴퓨팅 장치(12)는 RRAM 오류 에뮬레이터(100)일 수 있다.
- [0078] 컴퓨팅 장치(12)는 적어도 하나의 프로세서(14), 컴퓨터 판독 가능 저장 매체(16) 및 통신 버스(18)를 포함한다. 프로세서(14)는 컴퓨팅 장치(12)로 하여금 앞서 언급된 예시적인 실시예에 따라 동작하도록 할 수 있다. 예컨대, 프로세서(14)는 컴퓨터 판독 가능 저장 매체(16)에 저장된 하나 이상의 프로그램들을 실행할 수 있다. 상기 하나 이상의 프로그램들은 하나 이상의 컴퓨터 실행 가능 명령어를 포함할 수 있으며, 상기 컴퓨터 실행 가능 명령어는 프로세서(14)에 의해 실행되는 경우 컴퓨팅 장치(12)로 하여금 예시적인 실시예에 따른 동작들을 수행하도록 구성될 수 있다.
- [0079] 컴퓨터 판독 가능 저장 매체(16)는 컴퓨터 실행 가능 명령어 내지 프로그램 코드, 프로그램 데이터 및/또는 다른 적합한 형태의 정보를 저장하도록 구성된다. 컴퓨터 판독 가능 저장 매체(16)에 저장된 프로그램(20)은 프로세서(14)에 의해 실행 가능한 명령어의 집합을 포함한다. 일 실시예에서, 컴퓨터 판독 가능 저장 매체(16)는 메모리(랜덤 액세스 메모리와 같은 휘발성 메모리, 비휘발성 메모리, 또는 이들의 적절한 조합), 하나 이상의 자기 디스크 저장 디바이스들, 광학 디스크 저장 디바이스들, 플래시 메모리 디바이스들, 그 밖에 컴퓨팅 장치(12)에 의해 액세스되고 원하는 정보를 저장할 수 있는 다른 형태의 저장 매체, 또는 이들의 적합한 조합일 수 있다.
- [0080] 통신 버스(18)는 프로세서(14), 컴퓨터 판독 가능 저장 매체(16)를 포함하여 컴퓨팅 장치(12)의 다른 다양한 컴포넌트들을 상호 연결한다.
- [0081] 컴퓨팅 장치(12)는 또한 하나 이상의 입출력 장치(24)를 위한 인터페이스를 제공하는 하나 이상의 입출력 인터페이스(22) 및 하나 이상의 네트워크 통신 인터페이스(26)를 포함할 수 있다. 입출력 인터페이스(22) 및 네트워크 통신 인터페이스(26)는 통신 버스(18)에 연결된다. 입출력 장치(24)는 입출력 인터페이스(22)를 통해 컴퓨팅 장치(12)의 다른 컴포넌트들에 연결될 수 있다. 예시적인 입출력 장치(24)는 포인팅 장치(마우스 또는 트랙패드 등), 키보드, 터치 입력 장치(터치패드 또는 터치스크린 등), 음성 또는 소리 입력 장치, 다양한 종류의 센서 장치 및/또는 촬영 장치와 같은 입력 장치, 및/또는 디스플레이 장치, 프린터, 스피커 및/또는 네트워크 카드와 같은 출력 장치를 포함할 수 있다. 예시적인 입출력 장치(24)는 컴퓨팅 장치(12)를 구성하는 일 컴포넌트로서 컴퓨팅 장치(12)의 내부에 포함될 수도 있고, 컴퓨팅 장치(12)와는 구별되는 별개의 장치로 컴퓨팅 장치(12)와 연결될 수도 있다.
- [0083] 이상에서 본 발명의 대표적인 실시예들을 상세하게 설명하였으나, 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자는 상술한 실시예에 대하여 본 발명의 범주에서 벗어나지 않는 한도 내에서 다양한 변형이 가능함을 이해할 것이다. 그러므로 본 발명의 권리범위는 설명된 실시예에 국한되어 정해져서는 안 되며, 후술하는 특허 청구범위뿐만 아니라 이 특허청구범위와 균등한 것들에 의해 정해져야 한다.

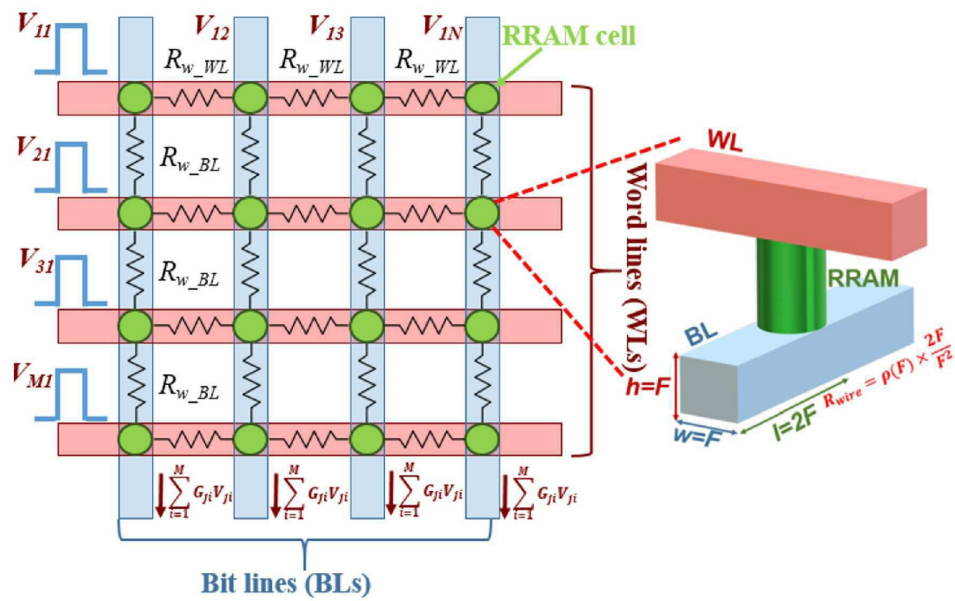
부호의 설명

- [0085] 10 : 컴퓨팅 환경
12 : 컴퓨팅 장치
14 : 프로세서
16 : 컴퓨터 판독 가능 저장 매체
18 : 통신 버스

- 20 : 프로그램
- 22 : 입출력 인터페이스
- 24 : 입출력 장치
- 26 : 네트워크 통신 인터페이스
- 100 : RRAM 오류 예블레이터
- 102 : 오류 생성 모듈
- 104 : 인공 신경망 모듈
- 111 : 랜덤 생성기
- 111a : 난수 생성기
- 111b : 분포 반전기
- 113 : 오류 모델

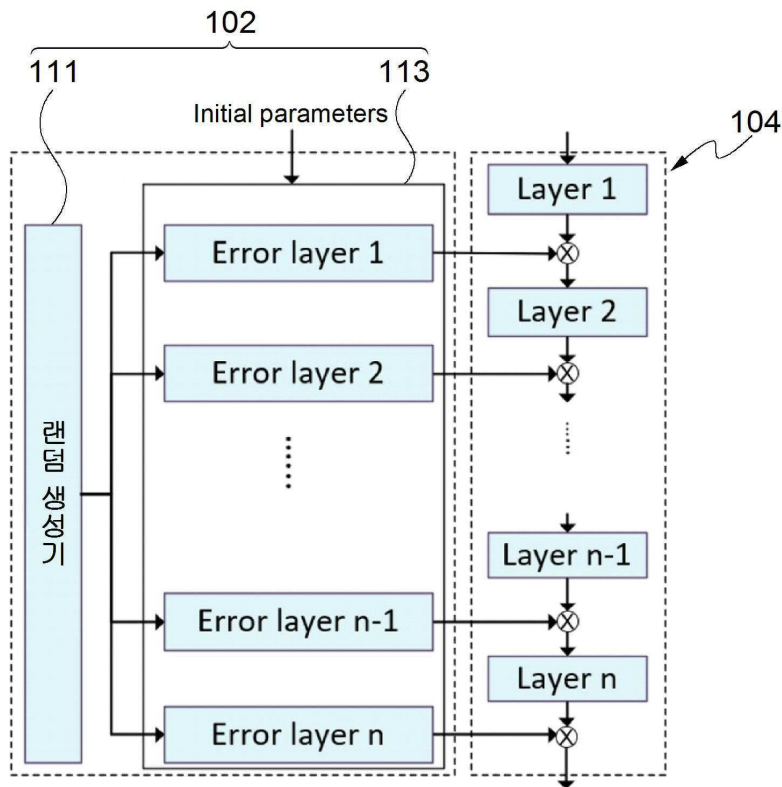
도면

도면1

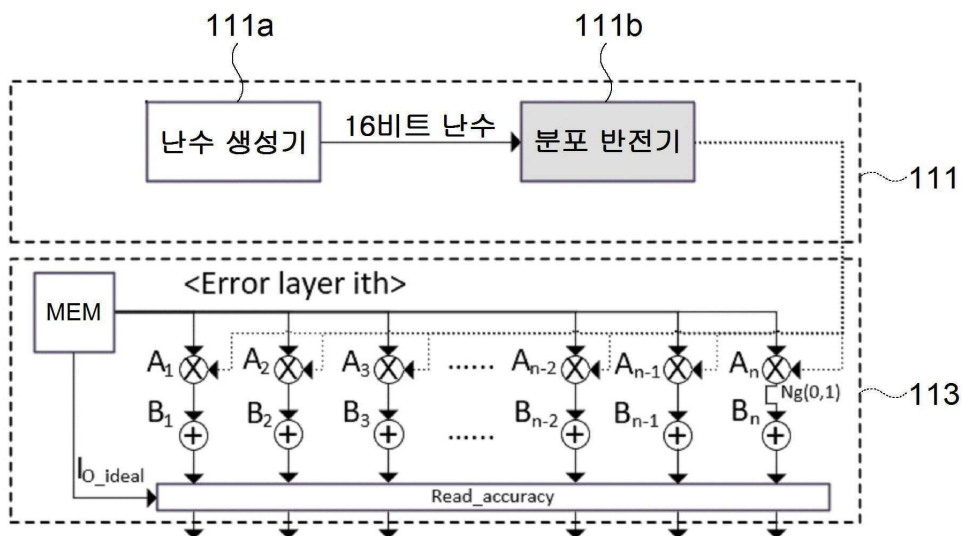


도면2

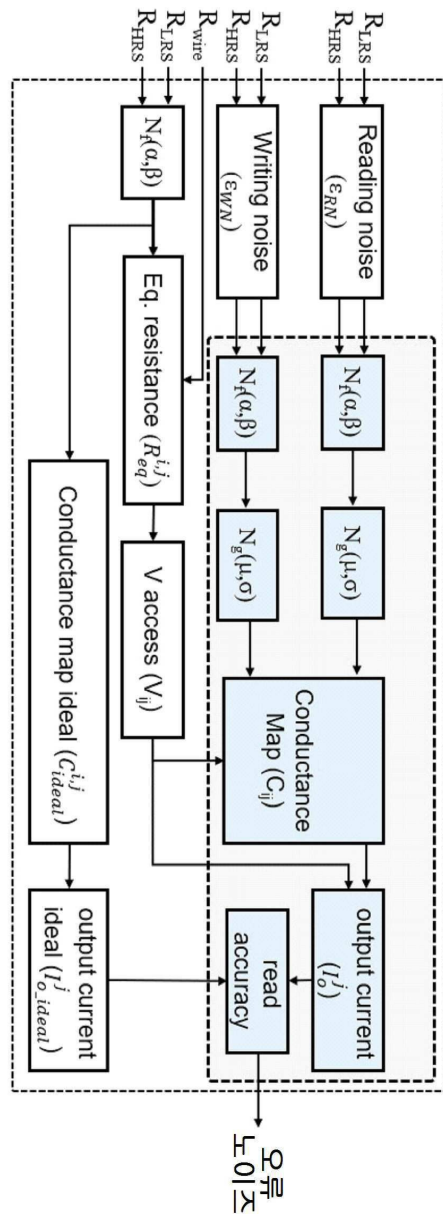
100



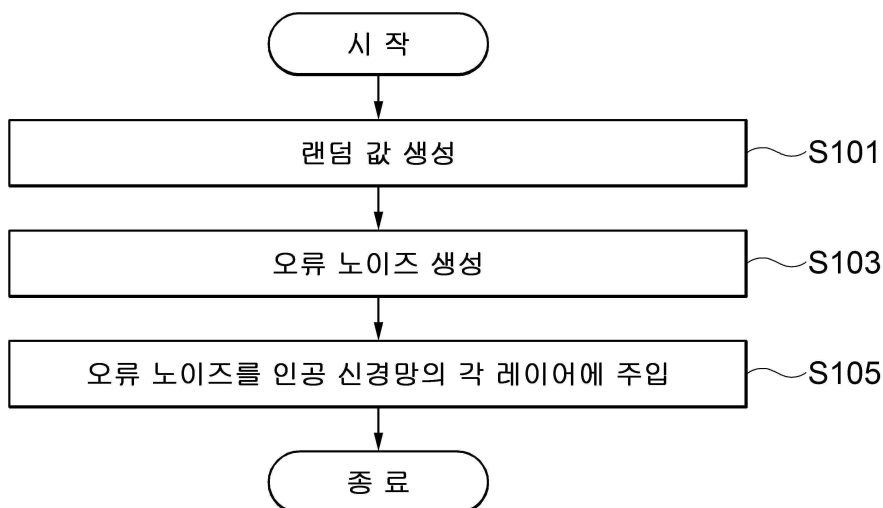
도면3



도면4



도면5



도면6

10

