



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2020년01월16일
(11) 등록번호 10-2067112
(24) 등록일자 2020년01월10일

(51) 국제특허분류(Int. Cl.)
G06N 3/063 (2006.01) G06N 3/04 (2006.01)
(52) CPC특허분류
G06N 3/063 (2013.01)
G06N 3/049 (2013.01)
(21) 출원번호 10-2017-0134748
(22) 출원일자 2017년10월17일
심사청구일자 2017년10월17일
(65) 공개번호 10-2019-0042996
(43) 공개일자 2019년04월25일
(56) 선행기술조사문헌
KR1019980081691 A*
(뒷면에 계속)

(73) 특허권자
한양대학교 산학협력단
서울특별시 성동구 왕십리로 222(행당동, 한양대학교내)
(72) 발명자
송윤흡
경기도 성남시 분당구 수내로 74, 112동 1304호
(수내동, 양지마을금호1단지아파트)
이정
서울특별시 성동구 왕십리로 222, 산학기술관 615호(사근동)
(74) 대리인
양성보

전체 청구항 수 : 총 4 항

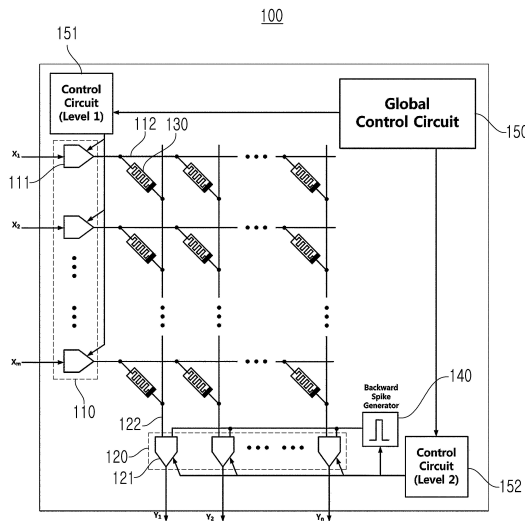
심사관 : 유진태

(54) 발명의 명칭 PCM 기반의 뉴런 네트워크 소자

(57) 요약

일 실시예에 따르면, PCM 기반의 뉴런 네트워크 소자는, 입력 레이어 및 출력 레이어 별로 배치되는 복수의 뉴런들; 상기 입력 레이어의 입력 라인 및 상기 출력 레이어의 출력 라인 사이를 연결하는 복수의 PCM(Phase Change Material)들; 및 상기 복수의 뉴런들에 의해 공유되며, 상기 출력 레이어의 뉴런들 각각에서 출력되는 출력 펄스를 기초로 스파이크를 생성하는 적어도 하나의 BSG(Backward Spike Generator)를 포함한다.

대표도 - 도1



(56) 선행기술조사문헌
KR1020100129741 A*
US20170017879 A1*
US20120109864 A1
KR1020170080431 A

*는 심사관에 의하여 인용된 문헌

이 발명을 지원한 국가연구개발사업

과제고유번호 201700000000924

부처명 미래창조과학부

연구관리전문기관 (재)한국연구재단

연구사업명 이공분야기초연구사업 / 중견연구자지원사업 / 핵심연구

연구과제명 칼코지나이드 물질을 기반으로 한 신경 모방형 3차원 세포체 소자 개발

기여율 1/1

주관기관 한양대학교 산학협력단

연구기간 2017.05.01 ~ 2018.04.30

명세서

청구범위

청구항 1

입력 레이어 및 출력 레이어 별로 배치되는 복수의 뉴런들;
 상기 입력 레이어의 입력 라인 및 상기 출력 레이어의 출력 라인 사이를 연결하는 복수의 PCM(Phase Change Material)들;
 상기 복수의 뉴런들에 의해 공유되며, 상기 출력 레이어의 뉴런들 각각에서 출력되는 출력 펄스를 기초로 스파이크를 생성하는 적어도 하나의 BSG(Backward Spike Generator); 및
 상기 복수의 뉴런들 각각에서 출력되는 펄스의 타이밍을 동기화하는 적어도 하나의 컨트롤 회로를 포함하고,
 상기 입력 레이어의 뉴런들 각각은,
 입력 펄스를 처리하는데 필요한 기능만을 구현하도록 역방향 펄스 드라이버를 제외한 PMOS 및 NMOS로 구성되며,
 상기 출력 레이어의 뉴런들 각각은,
 출력 펄스를 처리하는데 필요한 기능만을 구현하도록 순방향 펄스 드라이버를 제외한 PMOS 및 NMOS로 구성되고,
 상기 적어도 하나의 컨트롤 회로는,
 상기 입력 레이어의 뉴런들 각각에서 출력되는 펄스의 타이밍을 동기화하는 레벨 1 컨트롤 회로;
 상기 출력 레이어의 뉴런들 각각에서 출력되는 출력 펄스의 타이밍을 동기화하는 레벨 2 컨트롤 회로; 및
 상기 레벨 1 컨트롤 회로 및 상기 레벨 2 컨트롤 회로를 제어하는 글로벌 컨트롤 회로를 포함하는 PCM 기반의 뉴런 네트워크 소자.

청구항 2

제1항에 있어서,
 상기 복수의 뉴런들 각각은,
 상기 레이어 별로 서로 다른 구성요소를 포함하는, PCM 기반의 뉴런 네트워크 소자.

청구항 3

삭제

청구항 4

삭제

청구항 5

삭제

청구항 6

삭제

청구항 7

입력 레이어, 히든 레이어 및 출력 레이어 별로 배치되는 복수의 뉴런들;

상기 입력 레이어의 입력 라인 및 상기 히든 레이어의 연결 라인 사이와, 상기 히든 레이어의 연결 라인 및 상기 출력 레이어의 출력 라인 사이를 연결하는 복수의 PCM(Phase Change Material)들;

상기 복수의 뉴런들에 의해 공유되며, 상기 히든 레이어의 뉴런들 각각에서 출력되는 펄스 또는 상기 출력 레이어의 뉴런들 각각에서 출력되는 출력 펄스를 기초로 스파이크를 생성하는 적어도 하나의 BSG(Backward Spike Generator); 및

상기 복수의 뉴런들 각각에서 출력되는 펄스의 타이밍을 동기화하는 적어도 하나의 컨트롤 회로

를 포함하고,

상기 입력 레이어의 뉴런들 각각은,

입력 펄스를 처리하는데 필요한 기능만을 구현하도록 역방향 펄스 드라이버를 제외한 PMOS 및 NMOS로 구성되며,

상기 히든 레이어의 뉴런들 각각은,

상기 입력 레이어의 뉴런들로부터 전달되는 펄스를 상기 출력 레이어의 뉴런들로 전달하는 기능을 구현하도록 WTA(Winner-Takes-All) 드라이버를 제외한 PMOS 및 NMOS로 구성되고,

상기 출력 레이어의 뉴런들 각각은,

출력 펄스를 처리하는데 필요한 기능만을 구현하도록 순방향 펄스 드라이버를 제외한 PMOS 및 NMOS로 구성되며,

상기 적어도 하나의 컨트롤 회로는,

상기 입력 레이어의 뉴런들 각각에서 출력되는 펄스의 타이밍을 동기화하는 레벨 1 컨트롤 회로;

상기 히든 레이어의 뉴런들 각각에서 출력되는 펄스의 타이밍을 동기화하는 레벨 2 컨트롤 회로;

상기 출력 레이어의 뉴런들 각각에서 출력되는 출력 펄스의 타이밍을 동기화하는 레벨 3 컨트롤 회로; 및

상기 레벨 1 컨트롤 회로, 상기 레벨 2 컨트롤 회로 및 상기 레벨 3 컨트롤 회로를 제어하는 글로벌 컨트롤 회로

를 포함하는 PCM 기반의 뉴런 네트워크 소자.

청구항 8

제7항에 있어서,

상기 복수의 뉴런들 각각은,

상기 레이어 별로 서로 다른 구성요소를 포함하는, PCM 기반의 뉴런 네트워크 소자.

청구항 9

삭제

청구항 10

삭제

청구항 11

삭제

청구항 12

삭제

발명의 설명

기술 분야

[0001] 아래의 설명은 사람의 신경계를 PCM(Phase Change Material) 기반으로 모델링한 뉴런 네트워크 소자에 관한 것

으로, 보다 상세하게는 모델링된 회로 면적을 감소시키는 뉴런 네트워크 소자에 대한 것이다.

배경 기술

- [0002] 기존의 뉴런 네트워크 소자는, 컬럼 입력 신호들을 증폭하여 입력 받는 복수의 입력 구동 증폭기들 및 로우 출력 신호들을 증폭하여 출력하는 복수의 출력 구동 증폭기들을 포함하는 회로로 모델링된다. 이 때, 기존의 뉴런 네트워크 소자는 복수의 입력 구동 증폭기들 및 복수의 출력 증폭기들을 동일한 구조(예컨대, 역방향 펄스 드라이버, 순방향 펄스 드라이버 및 WTA(Winner-Takes-All) 드라이버를 포함하는 구조)로 형성하며, 복수의 입력 구동 증폭기들 및 복수의 출력 증폭기들 각각이 모두 스파이크를 생성하는 SG(Spike Generator)를 포함하도록 형성한다. 이와 같은 기존의 뉴런 네트워크 소자에 대한 기술은 등록특허공보 제10-0183406호에 개시되어 있다.
- [0003] 이에, 기존의 뉴런 네트워크 소자는 복수의 입력 구동 증폭기들 및 복수의 출력 증폭기들 각각의 회로 면적이 넓게 모델링되는 단점을 가지며, 이로 인하여 전체적인 회로 면적 역시 증가되는 단점을 갖는다. 또한, 기존의 뉴런 네트워크 소자는 복수의 입력 구동 증폭기들 및 복수의 출력 증폭기들 각각이 기능(예컨대, 펄스 입력 또는 펄스 출력)과 무관한 불필요한 구성요소를 포함하기 때문에, 에너지 소모가 큰 단점이 있다.
- [0004] 따라서, 아래의 실시예들은, 이러한 기존의 뉴런 네트워크 소자의 단점을 해결하는 기술을 제안하고자 한다.

발명의 내용

해결하려는 과제

- [0006] 일 실시예들은 사람의 신경계를 모델링된 회로 면적 및 에너지 소모를 감소시킨 PCM 기반의 뉴런 네트워크 소자를 제공한다.
- [0007] 구체적으로, 일 실시예들은 기존의 뉴런 네트워크 소자에서 복수의 입력 구동 증폭기들 및 출력 구동 증폭기들에 대응하는 복수의 뉴런들 각각이 SG를 포함하는 대신에, 적어도 하나의 BSG(Backward Spike Generator)를 공유하도록 구성되는 뉴런 네트워크 소자를 제공한다.
- [0008] 또한, 일 실시예들은 복수의 뉴런들이 레이어 별로 서로 다른 구성요소를 포함하도록 구성되는 뉴런 네트워크 소자를 제공한다.

과제의 해결 수단

- [0009] 일 실시예에 따르면, PCM 기반의 뉴런 네트워크 소자는, 입력 레이어 및 출력 레이어 별로 배치되는 복수의 뉴런들; 상기 입력 레이어의 입력 라인 및 상기 출력 레이어의 출력 라인 사이를 연결하는 복수의 PCM(Phase Change Material)들; 및 상기 복수의 뉴런들에 의해 공유되며, 상기 출력 레이어의 뉴런들 각각에서 출력되는 출력 펄스를 기초로 스파이크를 생성하는 적어도 하나의 BSG(Backward Spike Generator)를 포함한다.
- [0010] 일측에 따르면, 상기 복수의 뉴런들 각각은, 상기 레이어 별로 서로 다른 구성요소를 포함할 수 있다.
- [0011] 또 다른 일측에 따르면, 상기 입력 레이어의 뉴런들 각각은, 역방향 펄스 드라이버를 제외한 PMOS 및 NMOS를 포함하고, 상기 출력 레이어의 뉴런들 각각은, 순방향 펄스 드라이버를 제외한 PMOS 및 NMOS를 포함할 수 있다.
- [0012] 또 다른 일측에 따르면, 상기 PCM 기반의 뉴런 네트워크 소자는, 상기 복수의 뉴런들 각각에서 출력되는 펄스의 타이밍을 동기화하는 적어도 하나의 컨트롤 회로를 더 포함할 수 있다.
- [0013] 또 다른 일측에 따르면, 상기 적어도 하나의 컨트롤 회로는, 상기 레이어 별로 각각 구비될 수 있다.
- [0014] 또 다른 일측에 따르면, 상기 적어도 하나의 컨트롤 회로는, 상기 입력 레이어의 뉴런들 각각에서 출력되는 펄스의 타이밍을 동기화하는 레벨 1 컨트롤 회로; 상기 출력 레이어의 뉴런들 각각에서 출력되는 출력 펄스의 타이밍을 동기화하는 레벨 2 컨트롤 회로; 및 상기 레벨 1 컨트롤 회로 및 상기 레벨 2 컨트롤 회로를 제어하는 글로벌 컨트롤 회로를 포함할 수 있다.
- [0015] 일 실시예에 따르면, PCM 기반의 뉴런 네트워크 소자는, 입력 레이어, 히든 레이어 및 출력 레이어 별로 배치되는 복수의 뉴런들; 상기 입력 레이어의 입력 라인 및 상기 히든 레이어의 연결 라인 사이와, 상기 히든 레이어의 연결 라인 및 상기 출력 레이어의 출력 라인 사이를 연결하는 복수의 PCM(Phase Change Material)들; 및 상기 복수의 뉴런들에 의해 공유되며, 상기 히든 레이어의 뉴런들 각각에서 출력되는 펄스 또는 상기 출력 레이어

의 뉴런들 각각에서 출력되는 출력 펄스를 기초로 스파이크를 생성하는 적어도 하나의 BSG(Backward Spike Generator)를 포함한다.

- [0016] 일측에 따르면, 상기 복수의 뉴런들 각각은, 상기 레이어 별로 서로 다른 구성요소를 포함할 수 있다.
- [0017] 또 다른 일측에 따르면, 상기 입력 레이어의 뉴런들 각각은, 역방향 펄스 드라이버를 제외한 PMOS 및 NMOS를 포함하고, 상기 히든 레이어의 뉴런들 각각은, WTA(Winner-Takes-All) 드라이버를 제외한 PMOS 및 NMOS를 포함하며, 상기 출력 레이어의 뉴런들 각각은, 순방향 펄스 드라이버를 제외한 PMOS 및 NMOS를 포함할 수 있다.
- [0018] 또 다른 일측에 따르면, 상기 PCM 기반의 뉴런 네트워크 소자는, 상기 복수의 뉴런들 각각에서 출력되는 펄스의 타이밍을 동기화하는 적어도 하나의 컨트롤 회로를 더 포함할 수 있다.
- [0019] 또 다른 일측에 따르면, 상기 적어도 하나의 컨트롤 회로는, 상기 레이어 별로 각각 구비될 수 있다.
- [0020] 또 다른 일측에 따르면, 상기 적어도 하나의 컨트롤 회로는, 상기 입력 레이어의 뉴런들 각각에서 출력되는 펄스의 타이밍을 동기화하는 레벨 1 컨트롤 회로; 상기 히든 레이어의 뉴런들 각각에서 출력되는 펄스의 타이밍을 동기화하는 레벨 2 컨트롤 회로; 상기 출력 레이어의 뉴런들 각각에서 출력되는 출력 펄스의 타이밍을 동기화하는 레벨 3 컨트롤 회로; 및 상기 레벨 1 컨트롤 회로, 상기 레벨 2 컨트롤 회로 및 상기 레벨 3 컨트롤 회로를 제어하는 글로벌 컨트롤 회로를 포함할 수 있다.

발명의 효과

- [0021] 일 실시예들은 사람의 신경계를 모델링된 회로 면적 및 에너지 소모를 감소시킨 PCM 기반의 뉴런 네트워크 소자를 제공할 수 있다.
- [0022] 구체적으로, 일 실시예들은 복수의 뉴런들 각각이 BSG를 포함하는 대신에, 적어도 하나의 BSG를 공유하도록 구성되는 뉴런 네트워크 소자를 제공할 수 있다.
- [0023] 또한, 일 실시예들은 복수의 뉴런들이 레이어 별로 서로 다른 구성요소를 포함하도록 구성되는 뉴런 네트워크 소자를 제공할 수 있다.

도면의 간단한 설명

- [0024] 도 1 내지 2는 일 실시예에 따른 2-레이어 뉴런 네트워크 소자를 설명하기 위한 도면이다.
- 도 3 내지 4는 일 실시예에 따른 2-레이어 뉴런 네트워크 소자의 활용성을 설명하기 위한 도면이다.
- 도 5a 내지 5b는 일 실시예에 있어서, 2-레이어 뉴런 네트워크 소자가 펄스 타이밍을 동기화하는 것을 설명하기 위한 도면이다.
- 도 6 내지 7은 일 실시예에 따른 3-레이어 뉴런 네트워크 소자를 설명하기 위한 도면이다.
- 도 8 내지 10은 일 실시예에 따른 뉴런 네트워크 소자에서 뉴런의 셀 동작을 설명하기 위한 도면이다.
- 도 11 내지 12는 일 실시예에 따른 2-레이어 뉴런 네트워크 소자의 동작을 설명하기 위한 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0025] 이하, 본 발명에 따른 실시예들을 첨부된 도면을 참조하여 상세하게 설명한다. 그러나 본 발명이 실시예들에 의해 제한되거나 한정되는 것은 아니다. 또한, 각 도면에 제시된 동일한 참조 부호는 동일한 부재를 나타낸다.
- [0026] 또한, 본 명세서에서 사용되는 용어(Terminology)들은 본 발명의 바람직한 실시예를 적절히 표현하기 위해 사용된 용어들로서, 이는 시청자, 운용자의 의도 또는 본 발명이 속하는 분야의 관례 등에 따라 달라질 수 있다. 따라서, 본 용어들에 대한 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.
- [0027] 도 1 내지 2는 일 실시예에 따른 2-레이어 뉴런 네트워크 소자를 설명하기 위한 도면이다.
- [0028] 도 1 내지 2를 참조하면, 일 실시예에 따른 2-레이어 뉴런 네트워크 소자(100)는 입력 레이어(110) 및 출력 레이어(120) 별로 배치되는 복수의 뉴런들(111, 121), 입력 레이어(110)의 입력 라인(112) 및 출력 레이어(120)의 출력 라인(122) 사이를 연결하는 복수의 PCM들(130), 출력 레이어(120)의 뉴런들(121) 각각에서 출력되는 출력 펄스를 기초로 스파이크를 생성하는 하나의 BSG(140)를 포함한다.

- [0029] 복수의 뉴런들(111, 121) 각각은 레이어 별로 서로 다른 구성요소를 포함한다. 특히, 복수의 뉴런들(111, 121) 각각은 배치되는 레이어 별로 불필요한 기능을 구현하기 위한 구성요소를 배제하고, 필요한 기능을 구현하기 위한 구성요소만을 포함할 수 있다. 즉, 복수의 뉴런들(111, 121) 각각은 배치되는 레이어에 따라 서로 다른 구성요소를 포함할 수 있다.
- [0030] 예를 들어, 복수의 뉴런들(111, 121) 중 입력 레이어(110)의 뉴런들(111) 각각은 입력 펄스를 처리하는데 필요한 기능을 구현하기 위한 구성요소만을 포함하고, 복수의 뉴런들(111, 121) 중 출력 레이어(120)의 뉴런들(121) 각각은 출력 펄스를 처리하는데 필요한 기능을 구현하기 위한 구성요소만을 포함할 수 있다. 더 구체적인 예를 들면, 복수의 뉴런들(111, 121) 중 입력 레이어(110)의 뉴런들(111) 각각은 역방향 펄스 드라이버를 제외한 PMOS 및 NMOS를 포함하며, 복수의 뉴런들(111, 121) 중 출력 레이어(120)의 뉴런들(121) 각각은 순방향 펄스 드라이버를 제외한 PMOS 및 NMOS를 포함할 수 있다.
- [0031] 이처럼, 복수의 뉴런들(111, 121) 각각은 배치되는 레이어에 따른 기능만을 구현하도록 최소화된 구성요소를 포함함으로써, 회로 면적 및 에너지 소모를 기존의 뉴런에 비해 감소시킬 수 있다.
- [0032] 또한, 복수의 뉴런들(111, 121) 중 입력 레이어(110)의 뉴런들(111)은 입력 펄스를 처리하는데 필요한 기능을 구현하기 위한 구성요소만을 포함하는 가운데, 불규칙한 펄스를 발생시키는 대신에 규칙적인 사각형 펄스를 생성하기 위한 구성요소를 포함할 수 있다. 이에, 입력 레이어(110)의 뉴런들(111)에서 발생하는 이벤트의 에너지 계산이 단순화될 수 있다.
- [0033] 복수의 PCM들(130) 각각은 결정화 전류에 응답하여 결정화됨으로써, 다치화된 비트를 구현한다. 복수의 PCM들(130) 각각은 기존의 뉴런 네트워크 소자에서 사용되는 복수의 용량성 소자들(PCM들)과 동일하므로, 이에 대한 상세한 설명은 생략하기로 한다.
- [0034] BSG(140)는 출력 레이어(120)의 뉴런들(121) 각각에서 출력되는 출력 펄스를 기초로 스파이크를 생성함에 있어서, 복수의 뉴런들(111, 121)에 의해 공유된다. 다시 말해, BSG(140)는 출력 레이어(120)의 뉴런들(121) 각각과 연결되도록 뉴런 네트워크 소자(100)의 출력단에 배치되어, 출력 레이어(120)의 뉴런들(121)에 의해 공유될 수 있다.
- [0035] 상술한 바와 같이, 복수의 뉴런들(111, 121) 각각이 기존의 뉴런(기존의 뉴런들 각각은 배치되는 레이어와 무관한 불필요한 구성요소를 포함하며, SG를 포함함)에 비해 감소된 회로 면적을 갖게 됨으로써, 뉴런 네트워크 소자(100)의 전체 회로 면적 역시 최소화될 수 있다.
- [0036] 또한, 복수의 뉴런들(111, 121)이 하나의 BSG(140)를 공유함에 따라, BSG(140)의 변경만을 통하여, 뉴런 네트워크 소자(100)는 다양하게 활용될 수 있다. 이에 대한 상세한 설명은 도 3 내지 4를 참조하여 기재하기로 한다.
- [0037] 또한, 뉴런 네트워크 소자(100)는 복수의 뉴런들(111, 121) 각각에서 출력되는 펄스의 타이밍을 동기화하는 적어도 하나의 컨트롤 회로(150, 151, 152)를 포함할 수 있다. 여기서, 적어도 하나의 컨트롤 회로(151, 152)는 레이어 별로 각각 구비될 수 있다.
- [0038] 예를 들어, 레벨 1 컨트롤 회로(151)는 입력 레이어(110)의 뉴런들(111) 각각에서 출력되는 펄스의 타이밍을 동기화하도록 구비되고, 레벨 2 컨트롤 회로(152)는 출력 레이어(120)의 뉴런들(121) 각각에서 출력되는 출력 펄스의 타이밍을 동기화하도록 구비될 수 있다. 또한, 레벨 1 컨트롤 회로(151) 및 레벨 2 컨트롤 회로(152)를 제어하는 글로벌 컨트롤 회로(150)가 더 구비될 수 있다. 따라서, 뉴런 네트워크 소자(100)는 적어도 하나의 컨트롤 회로(150, 151, 152)를 포함함으로써, 복수의 뉴런들(111, 121)이 동일한 타이밍으로 펄스를 출력하도록 하며, 동기적으로 작동하도록 할 수 있다. 이에 대한 상세한 설명은 도 5a 내지 5b를 참조하여 기재하기로 한다.
- [0039] 도 3 내지 4는 일 실시예에 따른 2-레이어 뉴런 네트워크 소자의 활용성을 설명하기 위한 도면이다.
- [0040] 도 3을 참조하면, 도 1 내지 2를 통하여 상술된 2-레이어 뉴런 네트워크 소자(300)는, 유니폴라(Unipolar)의 스파이크를 생성하도록 구현된 적어도 하나의 BSG(310)를 포함함으로써, PCM 시냅스 소자로 활용될 수 있다.
- [0041] 또한, 도 4를 참조하면, 도 1 내지 2를 통하여 상술된 2-레이어 뉴런 네트워크 소자(400)는, 바이폴라(Bipolar)의 스파이크를 생성하도록 구현된 적어도 하나의 BSG(410)를 포함함으로써, ReRAM으로 활용될 수 있다.
- [0042] 도 5a 내지 5b는 일 실시예에 있어서, 2-레이어 뉴런 네트워크 소자가 펄스 타이밍을 동기화하는 것을 설명하기 위한 도면이다. 구체적으로, 도 5a는 기존의 뉴런 네트워크 소자를 나타낸 도면이고, 도 5b는 일 실시예에 따

른 뉴런 네트워크 소자를 나타낸 도면이다.

- [0043] 도 5a를 참조하면, 기존의 뉴런 네트워크 소자에서 입력 뉴런들 각각은 동기적으로 작동하지 않기 때문에, 각각이 출력하는 펄스의 타이밍 역시 서로 차이가 나게 된다.
- [0044] 반면에, 5b를 참조하면, 도 1 내지 2를 통하여 상술된 2-레이어 뉴런 네트워크 소자(500)는 복수의 뉴런들 각각에서 출력되는 펄스의 타이밍을 동기화하는 적어도 하나의 컨트롤 회로(510, 520, 530)를 포함함으로써, 입력 레이어(540)의 뉴런들(541) 각각에서 출력되는 펄스의 타이밍을 동기화하고, 출력 레이어의 뉴런들 각각에서 출력되는 펄스의 타이밍을 동기화할 수 있다.
- [0045] 예를 들어, 레벨 1 컨트롤 회로(510)는 입력 레이어(540)의 뉴런들(541) 각각이 동일한 타이밍을 갖는 펄스를 출력하도록 동기화하고, 레벨 2 컨트롤 회로(520)는 출력 레이어의 뉴런들 각각이 동일한 타이밍을 갖는 펄스를 출력하도록 동기화할 수 있다. 이 때, 글로벌 컨트롤 회로(530)는 레벨 1 컨트롤 회로(510) 및 레벨 2 컨트롤 회로(520)가 입력 레이어(540)의 뉴런들(541) 및 출력 레이어의 뉴런들을 동일한 타이밍으로 동기화되도록 제어할 수 있다.
- [0046] 따라서, 일 실시예에 따른 뉴런 네트워크 소자(500)에서 출력되는 펄스의 타이밍이 모두 동일하게 됨으로써, 뉴런 네트워크 소자(500)에서 발생하는 이벤트의 에너지 계산이 단순화될 수 있으며, 시냅스 가중치 업데이트 과정의 복잡도가 현저히 낮아질 수 있다.
- [0047] 이상, 도 1 내지 5b를 참조하여, 2-레이어 뉴런 네트워크 소자에 대해 설명하였으나, 일 실시예에 따른 뉴런 네트워크 소자는 3-레이어 구조로 확장될 수 있다. 이에 대한 상세한 설명은 아래에서 기재하기로 한다.
- [0048] 도 6 내지 7은 일 실시예에 따른 3-레이어 뉴런 네트워크 소자를 설명하기 위한 도면이다.
- [0049] 도 6 내지 7을 참조하면, 일 실시예에 따른 3-레이어 뉴런 네트워크 소자(600)는 입력 레이어(610), 히든 레이어(620) 및 출력 레이어(630) 별로 배치되는 복수의 뉴런들(611, 621, 631), 입력 레이어(610)의 입력 라인(612) 및 히든 레이어(620)의 연결 라인(622) 사이와, 히든 레이어(620)의 연결 라인(622) 및 출력 레이어(630)의 출력 라인(632) 사이를 연결하는 복수의 PCM들(640), 히든 레이어(620)의 뉴런들(621) 각각에서 출력되는 펄스 또는 출력 레이어(630)의 뉴런들(631) 각각에서 출력되는 출력 펄스를 기초로 스파이크를 생성하는 두 개의 BSG(650)를 포함한다.
- [0050] 복수의 뉴런들(611, 621, 631) 각각은 레이어 별로 서로 다른 구성요소를 포함한다. 특히, 복수의 뉴런들(611, 621, 631) 각각은 배치되는 레이어 별로 불필요한 기능을 구현하기 위한 구성요소를 배제하고, 필요한 기능을 구현하기 위한 구성요소만을 포함할 수 있다. 즉, 복수의 뉴런들(611, 621, 631) 각각은 배치되는 레이어에 따라 서로 다른 구성요소를 포함할 수 있다.
- [0051] 예를 들어, 복수의 뉴런들(611, 621, 631) 중 입력 레이어(610)의 뉴런들(611) 각각은 입력 펄스를 처리하는데 필요한 기능을 구현하기 위한 구성요소만을 포함하고, 복수의 뉴런들(611, 621, 631) 중 히든 레이어(620)의 뉴런들(621) 각각은 입력 레이어(610)의 뉴런들(611)로부터 전달되는 펄스를 출력 레이어(630)의 뉴런들(631)로 전달하는 기능을 구현하기 위한 구성요소만을 포함하며, 복수의 뉴런들(611, 621, 631) 중 출력 레이어(630)의 뉴런들(631) 각각은 출력 펄스를 처리하는데 필요한 기능을 구현하기 위한 구성요소만을 포함할 수 있다. 더 구체적인 예를 들면, 복수의 뉴런들(611, 621, 631) 중 입력 레이어(610)의 뉴런들(611) 각각은 역방향 펄스 드라이버를 제외한 PMOS 및 NMOS를 포함하고, 복수의 뉴런들(611, 621, 631) 중 히든 레이어(620)의 뉴런들(621) 각각은 WTA(Winner-Takes-All) 드라이버를 제외한 PMOS 및 NMOS를 포함하며, 복수의 뉴런들(611, 621, 631) 중 출력 레이어(630)의 뉴런들(631) 각각은 순방향 펄스 드라이버를 제외한 PMOS 및 NMOS를 포함할 수 있다.
- [0052] 이처럼, 복수의 뉴런들(611, 621, 631) 각각은 배치되는 레이어에 따른 기능만을 구현하도록 최소화된 구성요소를 포함함으로써, 회로 면적 및 에너지 소모를 기존의 뉴런에 비해 감소시킬 수 있다.
- [0053] 또한, 복수의 뉴런들(611, 621, 631) 중 입력 레이어(610)의 뉴런들(611)은 입력 펄스를 처리하는데 필요한 기능을 구현하기 위한 구성요소만을 포함하는 가운데, 불규칙한 펄스를 발생시키는 대신에 규칙적인 사각형 펄스를 생성하기 위한 구성요소를 포함할 수 있다. 이에, 입력 레이어(610)의 뉴런들(611)에서 발생하는 이벤트의 에너지 계산이 단순화될 수 있다.
- [0054] 복수의 PCM들(640) 각각은 결정화 전류에 응답하여 결정화됨으로써, 다치화된 비트를 구현한다. 복수의 PCM들(640) 각각은 기존의 뉴런 네트워크 소자에서 사용되는 복수의 용량성 소자들(PCM들)과 동일하므로, 이에 대한

상세한 설명은 생략하기로 한다.

- [0055] 두 개의 BSG들(650, 651) 중 히든 레이어(620)의 뉴런들(621) 각각과 연결되도록 배치되는 제1 BSG(650)는 히든 레이어(620)의 뉴런들(621) 각각에서 출력되는 펄스를 기초로 스파이크를 생성함에 있어서, 히든 레이어(620)의 뉴런들(621)에 의해 공유된다. 즉, 제1 BSG(650)는 히든 레이어(620)의 뉴런들(622) 각각과 연결되도록 히든 레이어(620)의 출력단에 배치되어, 히든 레이어(620)의 뉴런들(621)에 의해 공유될 수 있다. 만약, 히든 레이어(620)가 복수의 레이어들로 구성되는 경우, 제1 BSG(650)는 복수의 히든 레이어들 중 출력 레이어에 인접한 출력단에 하나 배치될 수 있으나, 이에 제한되거나 한정되지 않고, 복수의 히든 레이어들 각각에 배치되도록 복수 개로 구현될 수도 있다.
- [0056] 두 개의 BSG들(650, 651) 중 출력 레이어(630)의 뉴런들(631) 각각과 연결되도록 배치되는 제2 BSG(651)는 출력 레이어(630)의 뉴런들(631) 각각에서 출력되는 펄스를 기초로 스파이크를 생성함에 있어서, 출력 레이어(630)의 뉴런들(631)에 의해 공유된다. 다시 말해, 제2 BSG(651)는 출력 레이어(630)의 뉴런들(631) 각각과 연결되도록 뉴런 네트워크 소자(600)의 출력단에 배치되어, 출력 레이어(630)의 뉴런들(631)에 의해 공유될 수 있다.
- [0057] 상술한 바와 같이, 복수의 뉴런들(611, 621, 631) 각각이 기존의 뉴런(기존의 뉴런들 각각은 배치되는 레이어와 무관한 불필요한 구성요소를 포함하며, SG를 포함함)에 비해 감소된 회로 면적을 갖게 됨으로써, 뉴런 네트워크 소자(600)의 전체 회로 면적 역시 최소화될 수 있다.
- [0058] 또한, 복수의 뉴런들(611, 621, 631)이 두 개의 BSG들(650, 651)를 공유함에 따라, 두 개의 BSG들(650, 651) 중 출력 레이어(630)의 뉴런들(631) 각각에 연결된 BSG(651)의 변경만을 통하여, 뉴런 네트워크 소자(600)는 다양하게 활용될 수 있다. 이에 대한 상세한 설명은 도 3 내지 4를 참조하여 상술되었으므로, 생략하기로 한다.
- [0059] 또한, 뉴런 네트워크 소자(600)는 복수의 뉴런들(611, 621, 631) 각각에서 출력되는 펄스의 타이밍을 동기화하는 적어도 하나의 컨트롤 회로(660, 661, 662, 623)를 포함할 수 있다. 여기서, 적어도 하나의 컨트롤 회로(661, 662, 663)는 레이어 별로 각각 구비될 수 있다.
- [0060] 예를 들어, 레벨 1 컨트롤 회로(661)는 입력 레이어(610)의 뉴런들(611) 각각에서 출력되는 펄스의 타이밍을 동기화하도록 구비되고, 레벨 2 컨트롤 회로(662)는 히든 레이어(620)의 뉴런들(621) 각각에서 출력되는 출력 펄스의 타이밍을 동기화하도록 구비되며, 레벨 3 컨트롤 회로(663)는 출력 레이어(630)의 뉴런들(631) 각각에서 출력되는 출력 펄스의 타이밍을 동기화하도록 구비될 수 있다. 또한, 레벨 1 컨트롤 회로(661), 레벨 2 컨트롤 회로(662) 및 레벨 3 컨트롤 회로(663)를 제어하는 글로벌 컨트롤 회로(660)가 더 구비될 수 있다. 따라서, 뉴런 네트워크 소자(600)는 적어도 하나의 컨트롤 회로(660, 661, 662, 663)를 포함함으로써, 복수의 뉴런들(611, 621, 631)이 동일한 타이밍으로 펄스를 출력하도록 하며, 동기적으로 작동하도록 할 수 있다. 이에 대한 상세한 설명은 도 5a 내지 5b를 참조하여 상술되었으므로, 생략하기로 한다.
- [0061] 이상, 히든 레이어(620)가 하나의 레이어로 구성되는 경우를 설명하였으나, 이에 제한되거나 한정되지 않고, 히든 레이어(620)는 복수의 레이어들로 구성될 수도 있다. 이러한 경우 역시 상술된 구조로 설명 가능하다.
- [0062] 도 8 내지 10은 일 실시예에 따른 뉴런 네트워크 소자에서 뉴런의 셀 동작을 설명하기 위한 도면이다. 보다 상세하게, 도 8은 뉴런의 동작을 설명하기 위해 뉴런의 구성요소 중 활성화되는 구성요소를 나타낸 도면이고, 도 9는 뉴런의 동작을 설명하기 위한 플로우 차트이며, 도 10은 뉴런의 동작에 따른 타이밍 다이어그램을 나타낸 도면이다.
- [0063] 도 8을 참조하면, 도 1 내지 2를 통하여 상술된 뉴런 네트워크 소자에 포함되는 복수의 뉴런들 각각은 기록 페이즈(Write phase), 판독 페이즈(Reset phase) 및 리셋 페이즈(Reset phase)의 3 단계로 작동한다. 이하, 복수의 뉴런들 각각의 동작은 동일하므로, 출력 레이어의 뉴런들 중 어느 하나의 뉴런을 대상으로 해당 동작을 설명한다. 그러나 이에 제한되거나 한정되지 않고, 입력 레이어 또는 히든 레이어 각각에 포함되는 뉴런 역시 동일하게 동작할 수 있다.
- [0064] 예를 들어, 기록 페이즈에서, 크로스바에 의해 통합된 입력 전류는 PCM(복수의 PCM들 중 뉴런에 대응하는 PCM)의 컨덕턴스가 증가하도록 PCM을 결정화하기 위해 복사되어 결정화 전류를 형성하고 PCM으로 인가된다. 이에, 뉴런에 포함되는 구성요소 중 도면 상 짙게 표시된 구성요소만이 활성화되고 도면 상 흐리게 표시된 구성요소는 비활성화될 수 있다.
- [0065] 다른 예를 들면, 판독 페이즈에서, PCM의 컨덕턴스가 임계 값에 도달하는지 여부가 검출된다. 따라서, 뉴런에 포함되는 구성요소 중 도면 상 짙게 표시된 구성요소만이 활성화되고 도면 상 흐리게 표시된 구성요소는 비활성

화될 수 있다.

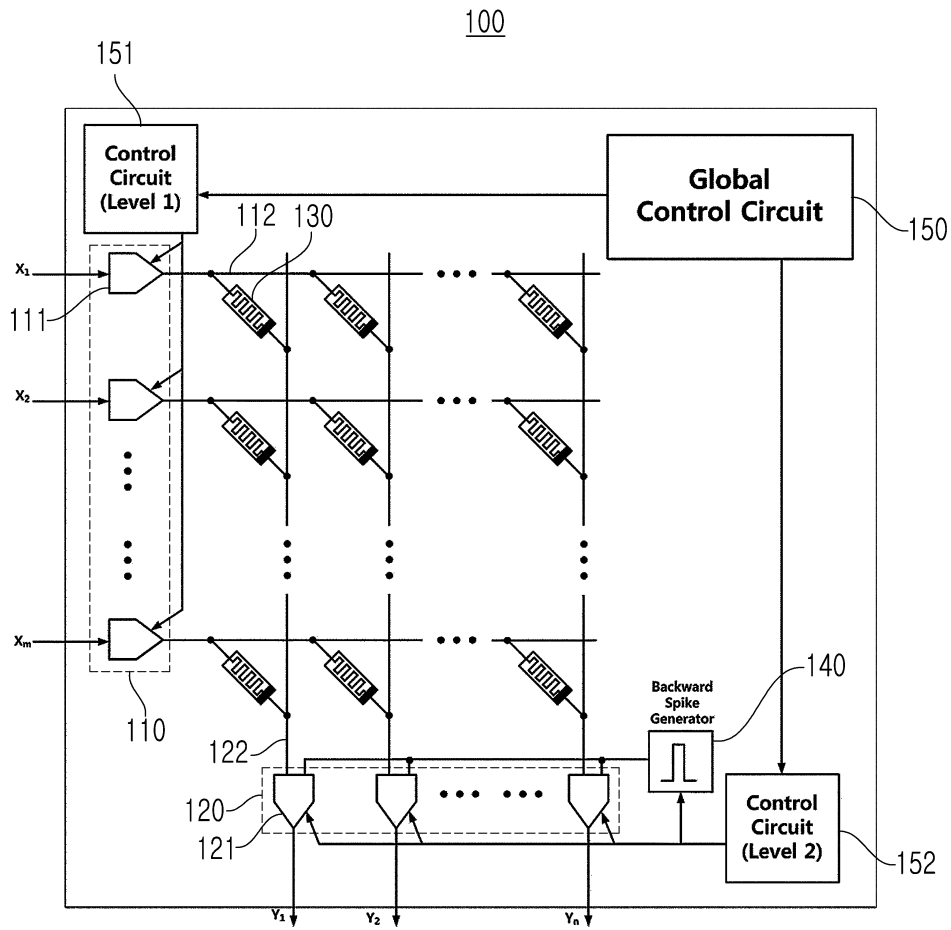
- [0066] 또 다른 예를 들면, 리셋 페이지에서, PCM의 컨덕턴스가 완전히 낮은 상태로 리셋된다. 이에, 뉴런에 포함되는 구성요소 중 도면 상 짙게 표시된 구성요소만이 활성화되고 도면 상 흐리게 표시된 구성요소는 비활성화될 수 있다.
- [0067] 이러한 뉴런의 Integrate 및 Fire(스파이크가 생성되어 뉴런에서 시냅스들로 제공되는 것) 동작은 도 9과 같은 순서로 수행된다. 뉴런은 초기 상태로 완전히 리셋된다(910).
- [0068] 이어서, 시냅스 가중 전류의 공간 합계(Spatial summation)이 적용됨에 따라 뉴런은 기록 페이지로 동작한다(920).
- [0069] 그 다음, 뉴런은 판독 페이지로 설정되어, 판독 페이지에서 판독 전류가 뉴런으로 인가되고, 시냅스들에 펄스가 적용되도록 하는 Fire또는 펄스로부터 시냅스들을 보호하는 No fire가 발생할 수 있다(930).
- [0070] 그 후, 뉴런에서 Fire가 발생되었는지 여부가 판단됨으로써(940)(예컨대, Fire 발생 여부는 외부 회로에 의해 감지되고 제어됨), Fire가 발생된 경우, 뉴런의 동작은 완료되고 뉴런은 불응기(Refractory period) 동안에 리셋된다(950). 이 때, 940 단계에서 뉴런은 전도도가 임계 값에 도달했는지 여부에 따라 Fire가 발생되었는지 판단될 수 있다.
- [0071] 만약, 뉴런의 전도도가 임계 값에 도달한 경우, 뉴런에서 Fire가 발생된 것으로 판단되어 950 단계가 수행될 수 있다. 반면에, 940 단계에서 Fire가 발생되지 않은 경우로 판단되면(뉴런의 전도도가 임계 값에 도달하지 않은 경우), 뉴런은 920 단계부터 다시 동작할 수 있다.
- [0072] 이처럼 동작하는 뉴런의 타이밍 다이어그램은 도 10과 같이 표현된다.
- [0073] 도 11 내지 12는 일 실시예에 따른 2-레이어 뉴런 네트워크 소자의 동작을 설명하기 위한 도면이다. 구체적으로, 도 11은 2-레이어 뉴런 네트워크 소자의 동작을 설명하기 위한 플로우 차트이고, 도 12는 2-레이어 뉴런 네트워크 소자의 동작에 따른 타이밍 다이어그램을 나타낸 도면이다.
- [0074] 도 11을 참조하면, 뉴런 네트워크 소자에 포함되는 복수의 뉴런들은 초기 상태로 완전히 리셋된다(1110).
- [0075] 이어서, 패턴 펄스들이 전류로 제공되거나 그렇지 않음에 따라, 복수의 뉴런들 중 입력 레이어의 뉴런들은 기록 페이지로 동작한다(1120).
- [0076] 그 다음, 복수의 뉴런들 중 입력 레이어의 뉴런들이 판독 페이지로 설정됨에 따라, 복수의 뉴런들 중 출력 레이어의 뉴런들은 기록 페이지(이 때, 패턴 펄스는 다음 패턴 펄스로 대체될 수 있음)로 동작한다(1130).
- [0077] 그 다음, 복수의 뉴런들 중 입력 레이어의 뉴런들은 휴식을 취하고, 복수의 뉴런들 중 출력 레이어의 뉴런들은 판독 페이지로 설정된다(1140).
- [0078] 그 다음, 뉴런들에서 Fire가 발생되었는지 여부가 판단됨으로써(1150)(예컨대, Fire 발생 여부는 외부 회로에 의해 감지되고 제어됨), Fire가 발생된 경우, Learning 동작 또는 Testing 동작이 수행될 수 있다(1160, 1170).
- [0079] Learning 동작의 경우, 복수의 뉴런들 중 입력 레이어의 뉴런들은 Learning 펄스를 제공한다(1160). 예를 들어, 1160 단계에서 일반적인 BSG가 동작을 시작할 수 있다.
- [0080] Testing 동작의 경우, 복수의 뉴런들 중 입력 레이어의 뉴런들은 휴식을 취하고, Fire가 발생된 출력 레이어의 뉴런들은 신호를 출력한다(1170).
- [0081] 그 후, 뉴런들의 동작은 완료되고, 불응기 동안에 리셋된다(1180).
- [0082] 만약, 1150 단계에서 Fire가 발생되지 않은 경우로 판단되면, 뉴런 네트워크 소자는 1130 단계부터 다시 동작할 수 있다.
- [0083] 이처럼 동작하는 뉴런의 타이밍 다이어그램은 도 12와 같이 표현된다. 도 11에서 WRITE 1 및 READ 1은 입력 레이어의 뉴런들에 대한 인에이블 신호이고, WRITE 2 및 READ 2는 출력 레이어의 뉴런들에 대한 인에이블 신호를 의미한다. 이에, Testing 동작에서 READ1은 시냅스 가중치 업데이트를 완료하기 위해 독립적으로 FIRE 신호로 활성화될 수 있다.
- [0084] 이상과 같이 실시예들이 비록 한정된 실시예와 도면에 의해 설명되었으나, 해당 기술분야에서 통상의 지식을 가

진 자라면 상기의 기재로부터 다양한 수정 및 변형이 가능하다. 예를 들어, 설명된 기술들이 설명된 방법과 다른 순서로 수행되거나, 및/또는 설명된 시스템, 구조, 장치, 회로 등의 구성요소들이 설명된 방법과 다른 형태로 결합 또는 조합되거나, 다른 구성요소 또는 균등물에 의하여 대치되거나 치환되더라도 적절한 결과가 달성될 수 있다.

[0085] 그러므로, 다른 구현들, 다른 실시예들 및 특허청구범위와 균등한 것들도 후술하는 특허청구범위의 범위에 속한다.

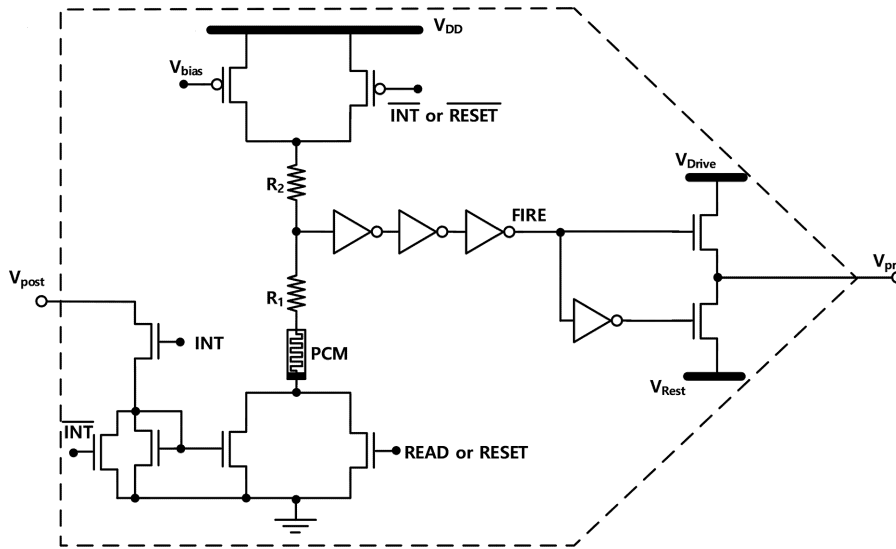
도면

도면1

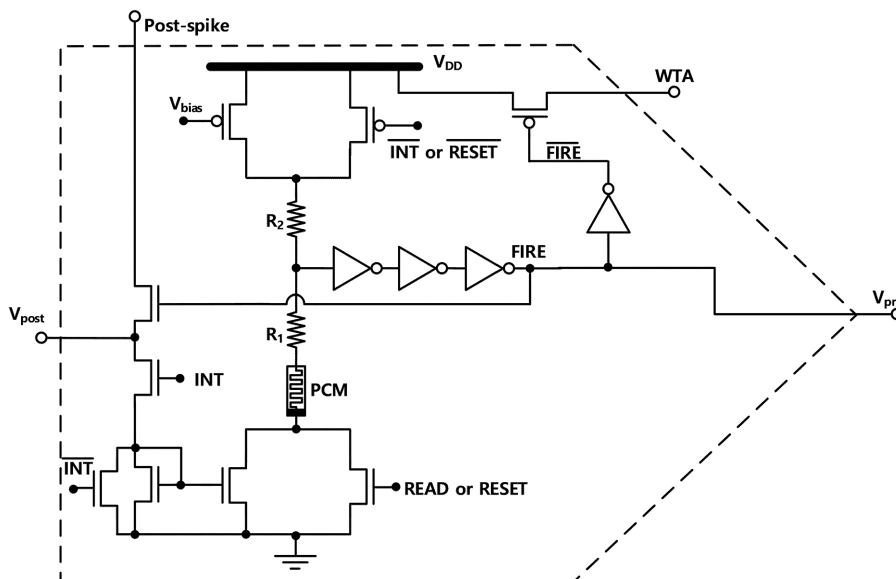


도면2

입력 레이어의 뉴런(111)

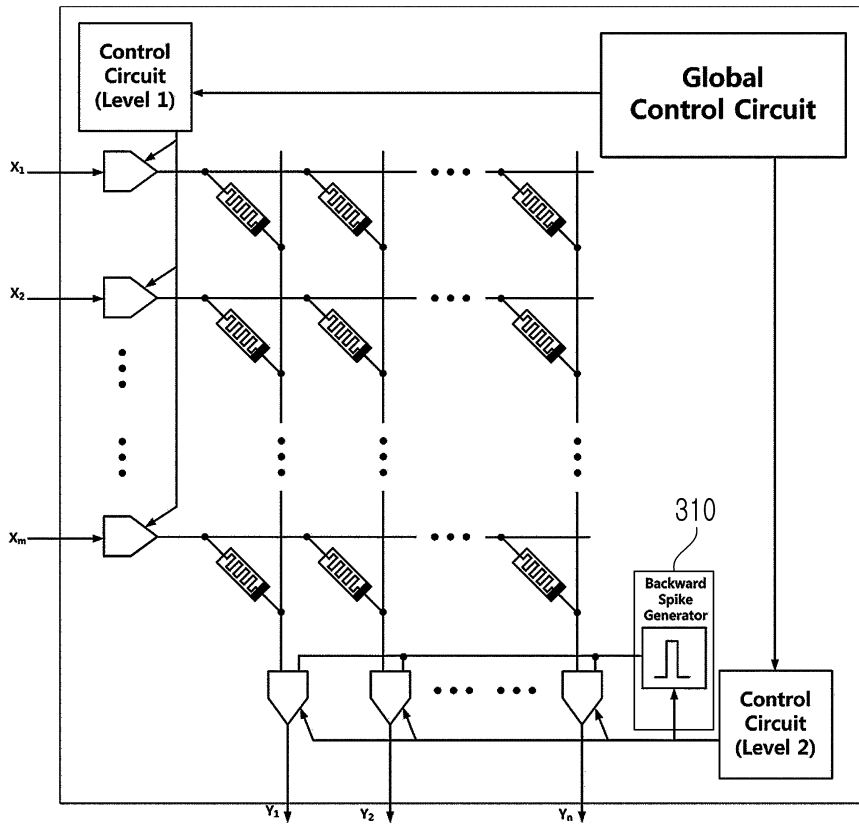


출력 레이어의 뉴런(121)



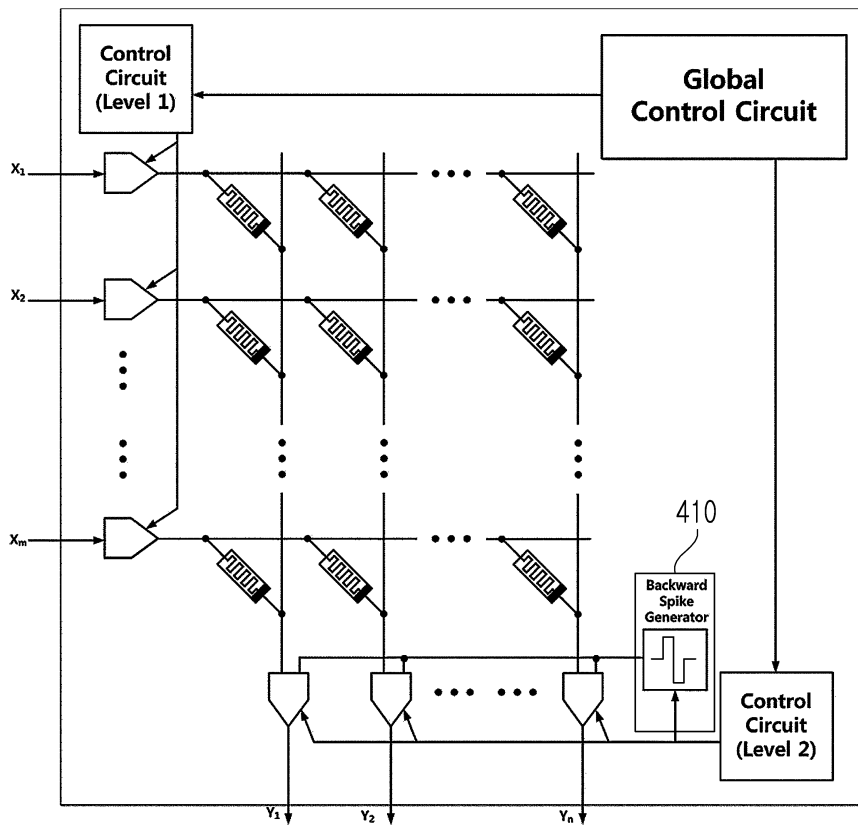
도면3

300

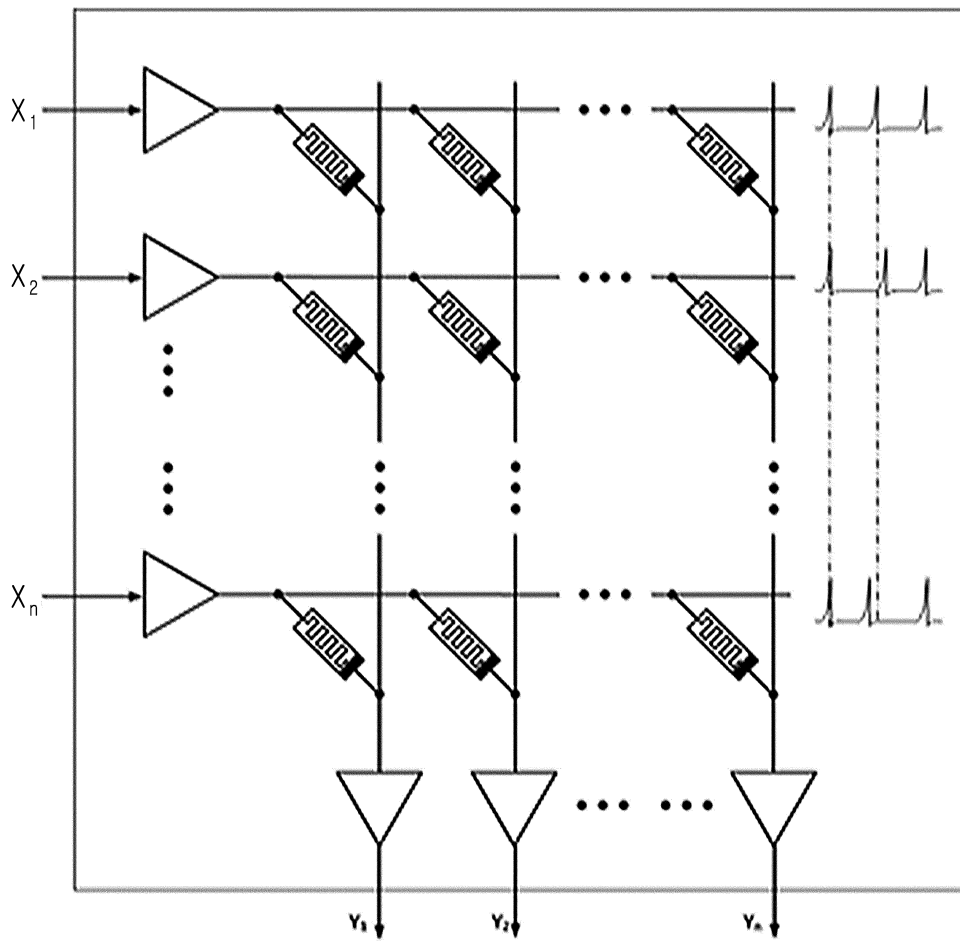


도면4

400

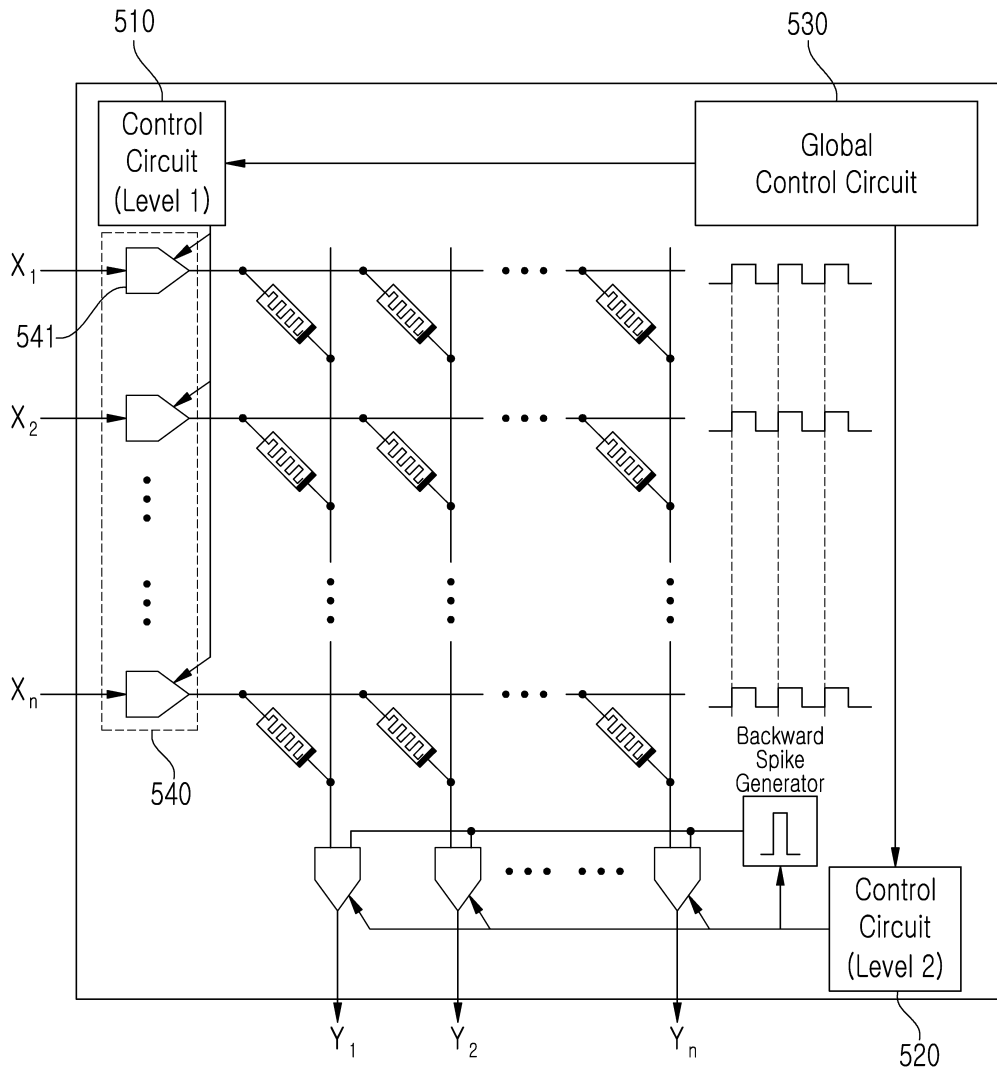


도면5a

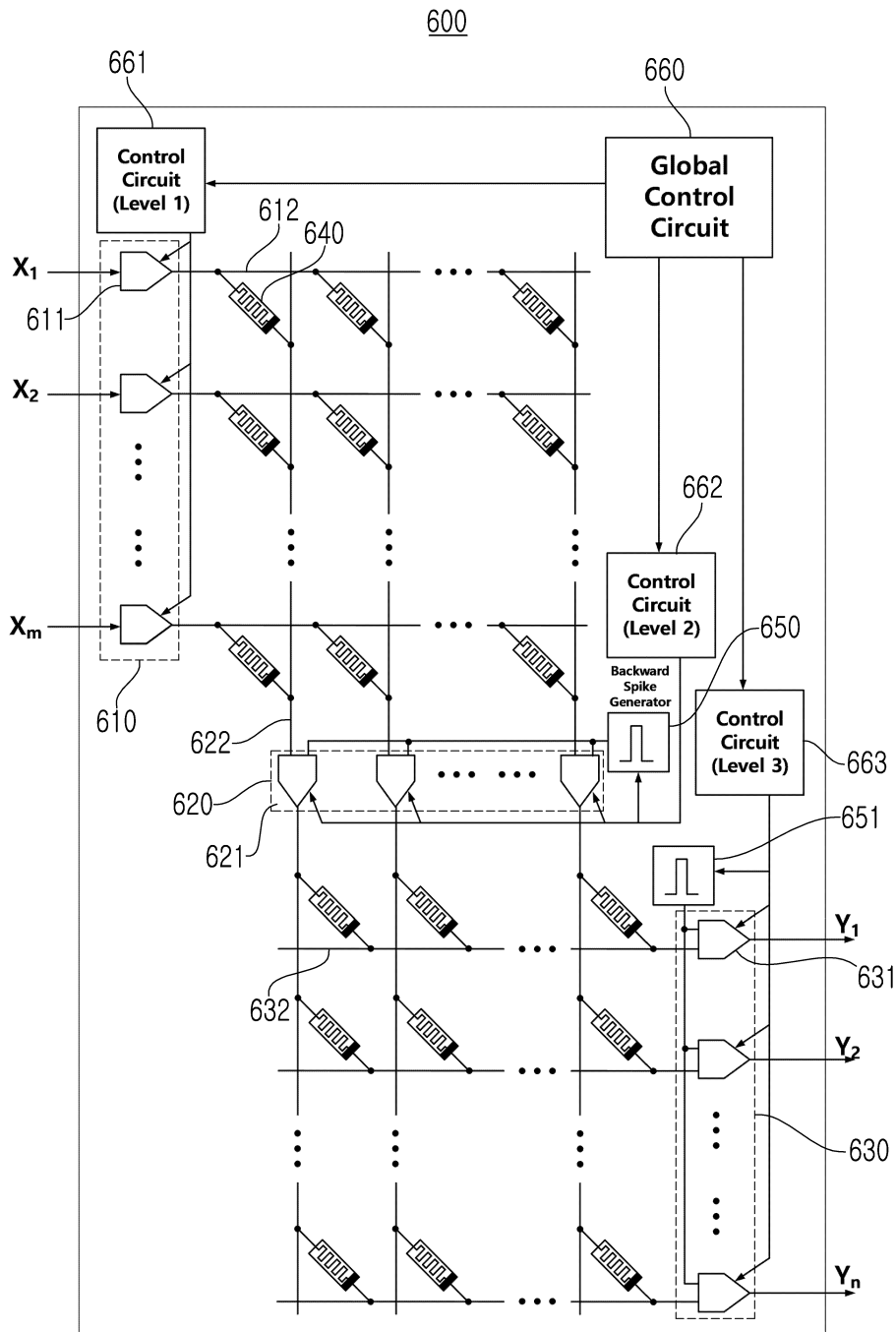


도면5b

500

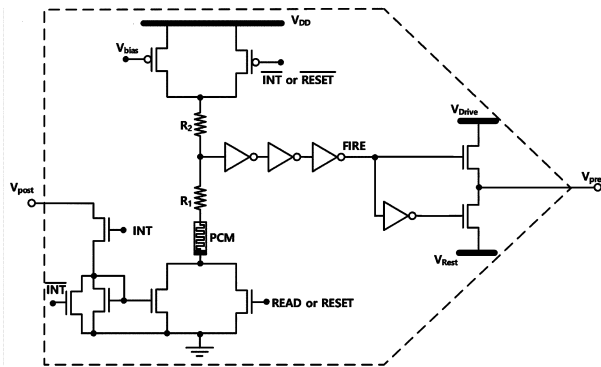


도면6

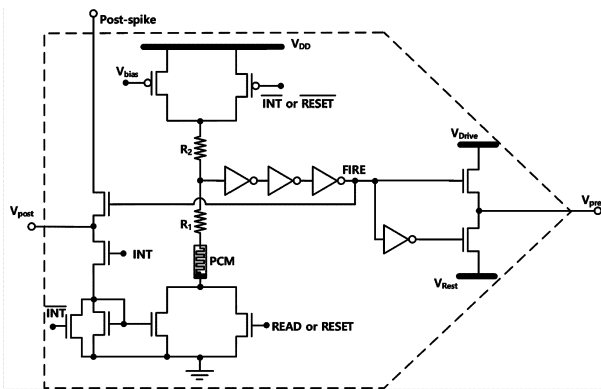


도면7

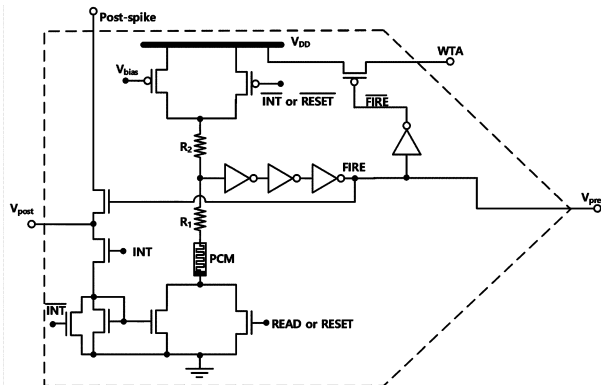
입력 레이어의 뉴런(611)



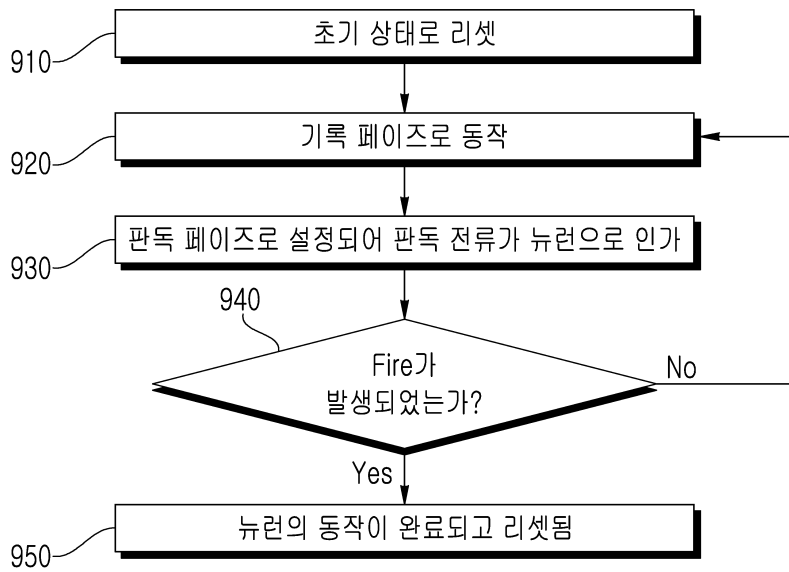
히트 레이어의 뉴런(621)



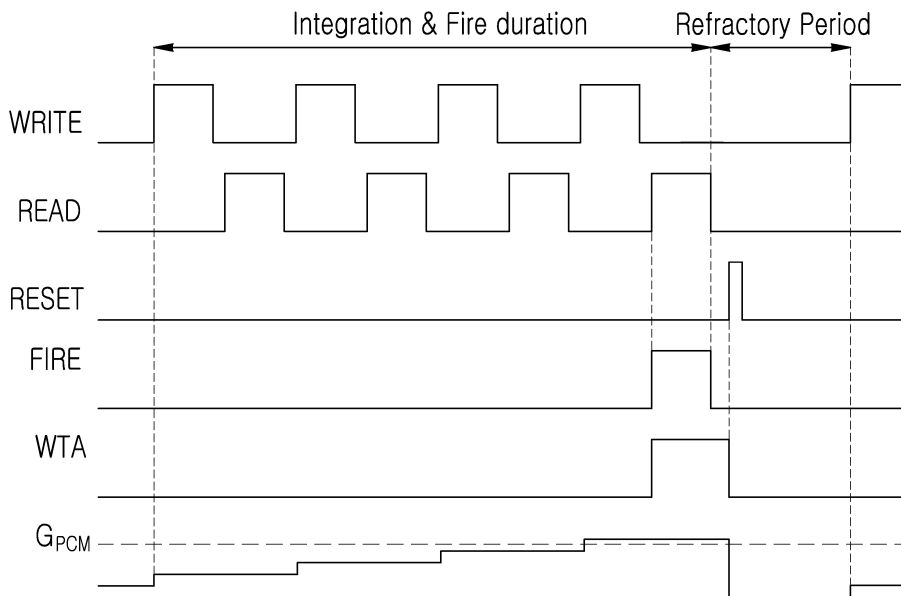
출력 레이어의 뉴런(631)



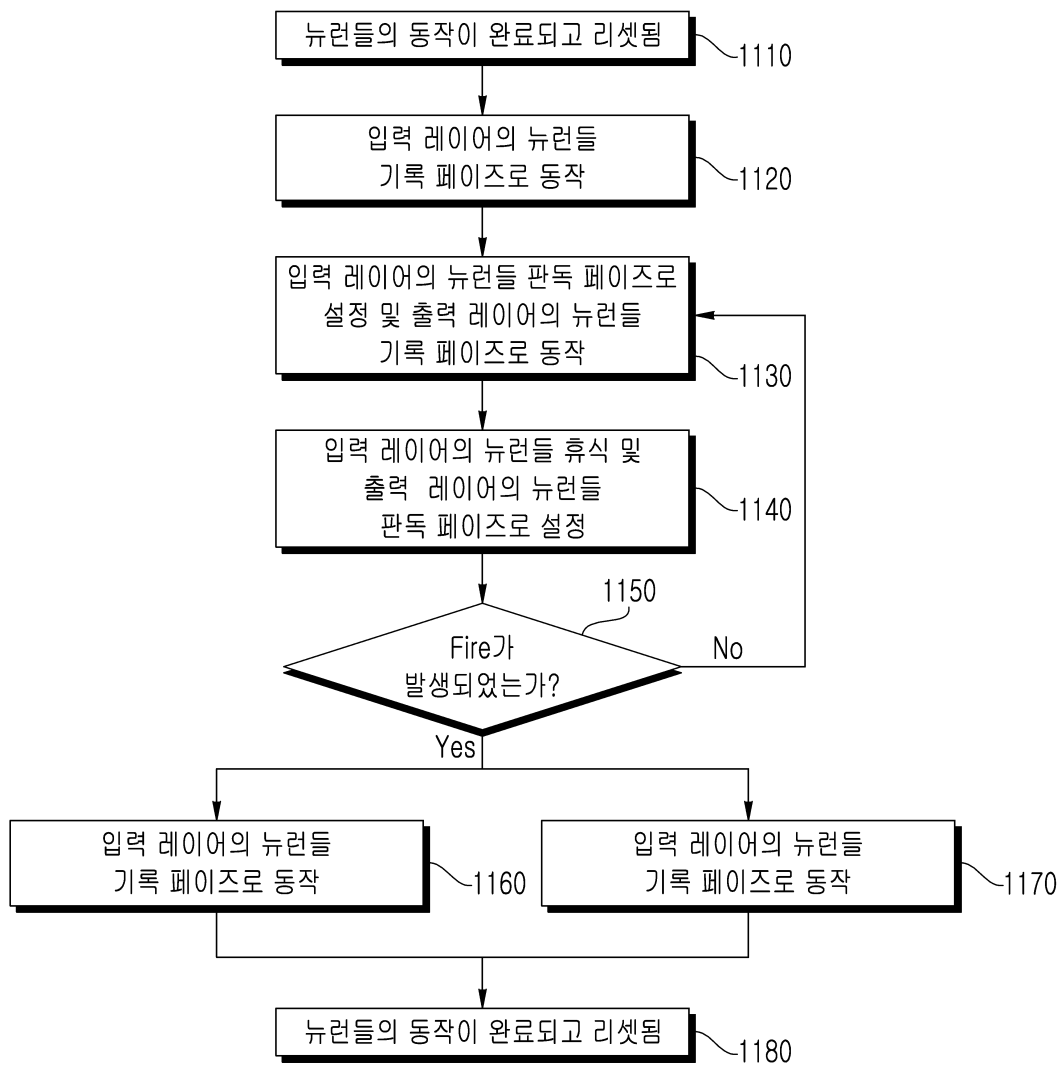
도면9



도면10



도면11



도면12

