(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년07월05일
(11) 등록번호 10-2272518
(24) 등록일자 2021년06월28일

(51) 국제특허분류(Int. Cl.)
G06F 15/80 (2006.01) *G06F 17/16* (2006.01)
G06F 7/544 (2017.01) *G06F 7/78* (2006.01)
G06N 3/063 (2006.01)
(52) CPC특허분류
G06F 15/8046 (2013.01)
G06F 17/16 (2013.01)
(21) 출원번호 10-2020-0031518
(22) 출원일자 2020년03월13일
심사청구일자 2020년03월13일
(56) 선행기술조사문헌
CN104462021 A
US8406334 B1
KR1020190099931 A
KR1020060017703 A

(73) 특허권자
인천대학교 산학협력단
인천광역시 연수구 아카데미로 119 (송도동)
성균관대학교산학협력단
경기도 수원시 장안구 서부로 2066 (천천동, 성균관대학교내)
(72) 발명자
정재용
인천광역시 연수구 해송로 70, 110동 1303호(송도동, 송도웰카운티2단지)
인야엣울라
인천광역시 연수구 함박안로 94, 103호(연수동)
(뒷면에 계속)
(74) 대리인
특허법인충정

기술이전 희망 : **기술양도**

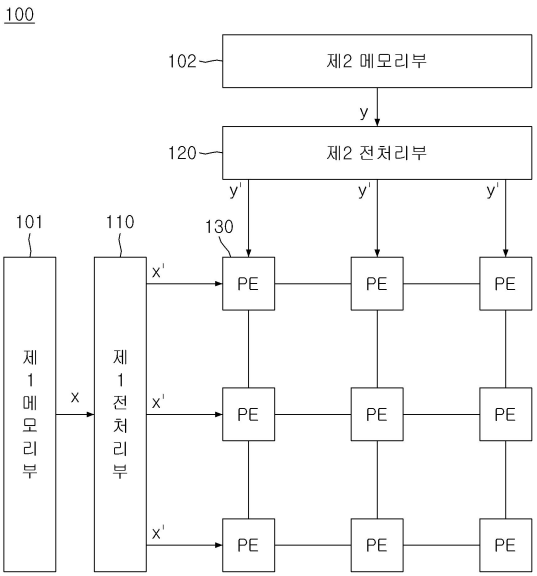
전체 청구항 수 : 총 10 항 심사관 : 윤혜숙

(54) 발명의 명칭 **시스톨릭 어레이 장치**

(57) 요 약

시스톨릭 어레이 장치는 시스톨릭 어레이를 구성하는 복수의 프로세싱 엘리먼트(Processing Element)에서 반복되는 전용 로직을 제거하여 PE의 외부의 입력단에 배치하여 전처리함으로써 PE의 연산 지연, 면적, 전력 소모, 하드웨어 자원을 줄일 수 있는 효과가 있다.

대 표 도 - 도4



(52) CPC특허분류
G06F 7/5443 (2013.01)
G06F 7/78 (2013.01)
G06N 3/063 (2013.01)
G06F 2207/3892 (2013.01)
(72) 발명자
카시프 이나얏
인천광역시 미추홀구 인하로134번길 31, B01호(용현동, 조은빌라)
양준성
경기도 성남시 분당구 분당로 244(분당동, 하나빌라)

이 발명을 지원한 국가연구개발사업

과제고유번호	SRFC-TB1803-02
부처명	삼성전자㈜
과제관리(전문)기관명	삼성전자 미래기술육성센터
연구사업명	삼성미래기술육성사업
연구과제명	Hybrid HW-SW 기반 학습면역 보안 IoT 시스템 아키텍처
기 여 율	1/1
과제수행기관명	성균관대학교
연구기간	2018.09.01 ~ 2020.08.31

명 세 서

청구범위

청구항 1

제1 메모리부에 저장된 값인 X를 X'로 변형하여 전달하는 제1 전처리부;

제2 메모리부에 저장된 값인 Y를 Y'로 변형하여 전달하는 제2 전처리부; 및

2차원 배열로 상호 연결되고, 상기 제1 전처리부가 좌측에서 병렬로 각각 연결되고, 상기 제2 전처리부가 상단에서 병렬로 각각 연결되고, 상기 제1 전처리부로부터 좌측에서 상기 X'를 각각 입력받고, 상기 제2 전처리부로부터 상단에서 상기 Y'를 각각 입력받는 복수의 PE(Processing Element)를 포함하는 것을 특징으로 하는 시스톨릭 어레이 장치.

청구항 2

제1항에 있어서,

상기 제1 전처리부와 상기 제2 전처리부는 곱셈 로직의 일부를 포함하는 것을 특징으로 하는 시스톨릭 어레이 장치.

청구항 3

제1항에 있어서,

상기 제1 전처리부는 상기 제1 메모리부에 저장된 값인 n-bit X를 전처리하여 m-bit X'로 변형한 후 상기 각각의 PE로 전달하고,

상기 제2 전처리부는 상기 제2 메모리부에 저장된 값인 n-bit Y를 전처리하여 m-bit Y'로 변형한 후 상기 각각의 PE로 전달하며,

상기 각각의 PE는 상기 제1 전처리부로부터 좌측에서 상기 m-bit X'를 각각 입력받고, 상기 제2 전처리부로부터 상단에서 상기 m-bit Y'를 각각 입력받으며, m은 n보다 큰 정수인 것을 특징으로 하는 시스톨릭 어레이 장치.

청구항 4

제1항에 있어서,

상기 제1 전처리부는 상기 제1 메모리부에 저장된 값인 n-bit X를 전처리하여 Signed Digit Number 형태의 X'로 변형한 후 상기 각각의 PE로 전달하고,

상기 제2 전처리부는 상기 제2 메모리부에 저장된 값인 n-bit Y를 전처리하여 Signed Digit Number 형태의 Y'로 변형한 후 상기 각각의 PE로 전달하며,

상기 각각의 PE는 상기 제1 전처리부로부터 좌측에서 상기 Signed Digit Number 형태의 X'를 각각 입력받고, 상기 제2 전처리부로부터 상단에서 상기 Signed Digit Number 형태의 Y'를 각각 입력받는 것을 특징으로 하는 시스톨릭 어레이 장치.

청구항 5

제1항에 있어서,

상기 제1 전처리부는 상기 제1 메모리부에 저장된 값인 X를 부스 인코딩하여 부스 인코딩된 형태의 X'로 변형한 후 상기 각각의 PE로 전달하고,

상기 제2 전처리부는 상기 제2 메모리부에 저장된 값인 Y를 Y+2Y = 3Y로 전처리하여 Y와 3Y를 상기 각각의 PE로 전달하며,

상기 각각의 PE는 상기 제1 전처리부로부터 좌측에서 상기 부스 인코딩된 형태의 X'를 각각 입력받고, 상기 제2

전처리부로부터 상단에서 상기 Y와 3Y를 각각 입력받는 것을 특징으로 하는 시스톨릭 어레이 장치.

청구항 6

2차원 배열로 상호 연결되고, 입력 행렬 X가 저장된 제1 메모리부가 좌측에서 병렬로 각각 연결되고, 입력 행렬 Y가 저장된 제2 메모리부가 상단에서 병렬로 각각 연결되고, 상기 제1 메모리부로부터 좌측에서 상기 X를 각각 입력받고, 상기 제2 메모리부로부터 상단에서 상기 Y를 각각 입력받는 복수의 PE(Processing Element); 및

상기 각각의 PE로부터 출력된 P를 후처리하여 P'로 변형하여 출력하는 후처리부를 포함하는 것을 특징으로 하는 시스톨릭 어레이 장치.

청구항 7

제6항에 있어서,

상기 후처리부는 상기 각각의 PE의 출력된 n-bit P를 후처리하여 m-bit P'로 변환한 후 출력하며, n은 m보다 큰 정수인 것을 특징으로 하는 시스톨릭 어레이 장치.

청구항 8

제6항에 있어서,

상기 후처리부는 상기 각각의 PE의 출력된 Redundant Binary Number P를 후처리하여 Binary Number P'로 변환한 후 출력하는 것을 특징으로 하는 시스톨릭 어레이 장치.

청구항 9

제6항에 있어서,

상기 후처리부는 상기 각각의 PE의 출력된 복수의 Binary Number들의 결합 P를 덧셈기로 후처리하여 하나의 Binary Number P'로 변환한 후 출력하는 것을 특징으로 하는 시스톨릭 어레이 장치.

청구항 10

제6항에 있어서,

상기 후처리부는 곱셈 누산 로직의 일부를 포함하는 것을 특징으로 하는 시스톨릭 어레이 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 시스톨릭 어레이 장치에 관한 것으로서, 더욱 상세하게는 시스톨릭 어레이를 구성하는 복수의 프로세싱 엘리먼트(Processing Element)에서 반복되는 전용 로직을 제거하여 PE의 외부의 입력단에 배치하여 전처리함으로써 PE의 연산 지연, 면적, 전력 소모를 줄일 수 있는 시스톨릭 어레이 장치에 관한 것이다.

배 경 기 술

[0002] 인공 신경망(Artificial Neural Network; ANN)은 생물학적인 신경망과 유사한 방식으로 데이터 또는 정보를 처리할 수 있다. 인공 신경망의 은닉 계층이 2개 이상인 경우를 심층 신경망(Deep Neural Network)이라 하며, 최근 활발히 연구되고 있다.

[0003] 심층 신경망은 데이터 처리를 위하여 방대한 양의 계산량이 요구하는데, 이를 위하여 행렬 곱셈에 최적화된 회로인 시스톨릭 어레이(Systolic Array)를 적용한다.

[0004] 시스톨릭 어레이는 동일한 동작을 하는 많은 PE(Processing Element)들로 이루어져 있으며, 동시에 많은 연산을 할 수 있다.

[0005] 시스톨릭 어레이를 이용한 연산 기술은 오래 전부터 사용되는 방법이며, 최근에 위의 컨볼루션 신경망과 같은 딥 신경망(Deep Neural Network)을 처리하기 위해 많이 사용되고 있다.

[0006] 도 1 및 도 2는 종래 기술에 따른 시스톨릭 어레이 구조를 나타낸 도면이다.

[0007] 종래 기술은 2차원 배열로 상호 연결된 PE(10)이고, 제1 메모리부(20)로부터 입력 행렬 A를 입력받고, 제2 메모리부(30)로부터 입력 행렬 B를 입력받고, 두 개의 3×3 입력 행렬 A와 B의 곱을 수행한 후, 동일한 크기의 행렬 C를 생성하는 3×3 시스톨릭 어레이 구조이다.

[0008] 입력 행렬 A는 왼쪽 가장자리의 PE(10)에 제공되고, 오른쪽 방향으로 전달되고, 입력 행렬 B는 상부의 PE(10)에 제공되고, 하부 방향으로 전달된다.

[0009] 각각의 PE(10)는 곱셈 연산을 수행하는 곱셈기와 누적 유닛을 포함하고, 누적 결과를 누적 레지스터에 축적된다.

[0010] 예를 들어, PE_{11}은 제1 클록 사이클(t1)에서 입력 a_{11}과, 입력 b_{11}을 곱셈 연산을 수행한다. PE_{11}은 제2 클록 사이클(t2)에서 입력 a_{12}과, 입력 b_{21}을 곱셈 연산을 수행하고, 이전 사이클인 제1 클록 사이클에서 입력된 a_{11}과 b_{11}을 각각 PE_{12}와 PE_{21}로 전달한다.

[0011] 동일한 클록 사이클에서 PE_{12}는 입력 a_{11}와 입력 b_{12}을 곱셈 연산하고, PE_{21}은 입력 a_{21}과 입력 b_{11}을 곱셈 연산한다.

[0012] PE_{11}은 제3 클록 사이클(t3)에서 입력 a_{13}과 입력 b_{31}을 곱셈 연산하고, 제4 클록 사이클(t4)에서 최종 결과 C_{11}를 출력한다. PE_{12}와 PE_{21}은 제5 클록 사이클(t5)에서 행렬 c_{12}와 c_{21}의 결과를 출력한다. 이러한 방법으로 입력 행렬의 값이 시스톨릭 어레이의 각 구성요소로 전파되고, 출력 행렬이 계산된다.

[0013] 도 3은 종래 기술에 따른 시스톨릭 어레이 구조의 상세 구성을 나타낸 도면이다.

[0014] 종래 기술에 따른 시스톨릭 어레이 구조는 2차원 배열로 상호 연결된 PE(10)로 이루어져 있다.

[0015] 각각의 PE(10)는 XDL(X Dedicated Logic)(11), YDL(Y Dedicated Logic)(12), ML(Main Logic)(13), 누산기(14) 및 PDL(P Dedicated Logic)(15)를 포함한다. 종래 기술의 시스톨릭 어레이 구조는 XDL(11), YDL(12), ML(13)이 하나의 곱셈기를 구성한다.

[0016] XDL(11)은 곱셈기 로직 중 입력 X에만 의존하는 로직이고, YDL(12)은 곱셈기 로직 중 입력 Y에만 의존하는 로직이고, ML(13)은 곱셈기 로직 중 XDL(11)과 YDL(12)이 아닌 나머지 부분으로 정의된다. 즉, 하나의 곱셈기는 XDL(11), YDL(12), ML(13)로 분해 가능하다.

[0017] XDL(11)은 제1 메모리부(20)로부터 임의의 사이클에서 X축 방향으로 전송되는 입력 행렬 A를 수신하여 연산을 수행한다.

[0018] YDL(12)은 제2 메모리부(30)로부터 임의의 사이클에서 Y축 방향으로 전송되는 입력 행렬 B를 수신하여 연산을 수행한다.

[0019] ML(13)은 XDL(11)로부터 처리된 결과와 YDL(12)로부터 처리된 결과를 이용하여 곱셈 연산의 핵심 부분을 수행한다.

[0020] 누산기(14)는 곱셈기(13)의 곱셈 결과를 누적할 수 있다.

[0021] PDL(15)은 누산기(14)에서 누적된 결과를 최종 처리하여 PE(10)의 밖으로 내보낸다.

[0022] PE(10)는 곱셈기(13), 누산기(14), PDL(15)를 이용하여 곱셈 누산 연산을 수행하고, 곱셈 누산 연산을 반복적으로 수행하여 내적 연산(Dot-Product)을 수행할 수 있다.

[0023] 이와 같이, 시스톨릭 어레이는 각각의 PE(10)를 구성하는 복수의 엘리먼트가 처리해야 하는 연산량으로 인하여 하드웨어 자원의 요구가 증가하며, 빠른 속도와 높은 효율을 달성하지 못하는 문제점이 있었다.

선행기술문헌

특허문헌

[0024] (특허문헌 0001) 한국 공개특허번호 제10-2019-0065144호

발명의 내용

해결하려는 과제

[0025] 이와 같은 문제점을 해결하기 위하여, 본 발명은 시스톨릭 어레이를 구성하는 복수의 프로세싱 엘리먼트(Processing Element)에서 반복되는 전용 로직을 제거하여 PE의 외부의 입력단에 배치하여 전처리함으로써 PE의 연산 지연, 면적, 전력 소모를 줄일 수 있는 시스톨릭 어레이 장치를 제공하는데 그 목적이 있다.

[0026] 본 발명은 시스톨릭 어레이를 구성하는 복수의 프로세싱 엘리먼트(Processing Element)에서 반복되는 전용 로직을 제거하여 PE의 외부의 출력단에 배치하여 후처리함으로써 PE의 연산 지연, 면적, 전력 소모를 줄일 수 있는 시스톨릭 어레이 장치를 제공하는데 그 목적이 있다.

과제의 해결 수단

[0027] 상기 목적을 달성하기 위한 본 발명의 특징에 따른 시스톨릭 어레이 장치는,

[0028] 제1 메모리부에 저장된 값인 X를 X'로 변형하여 전달하는 제1 전처리부;

[0029] 제2 메모리부에 저장된 값인 Y를 Y'로 변형하여 전달하는 제2 전처리부; 및

[0030] 2차원 배열로 상호 연결되고, 상기 제1 전처리부가 좌측에서 병렬로 각각 연결되고, 상기 제2 전처리부가 상단에서 병렬로 각각 연결되고, 상기 제1 전처리부로부터 좌측에서 상기 X'를 각각 입력받고, 상기 제2 전처리부로부터 상단에서 상기 Y'를 각각 입력받는 복수의 PE(Processing Element)를 포함하는 것을 특징으로 한다.

[0032] 본 발명의 특징에 따른 시스톨릭 어레이 장치는,

[0033] 2차원 배열로 상호 연결되고, 입력 행렬 X가 저장된 제1 메모리부가 좌측에서 병렬로 각각 연결되고, 입력 행렬 Y가 저장된 제2 메모리부가 상단에서 병렬로 각각 연결되고, 상기 제1 메모리부로부터 좌측에서 상기 X를 각각 입력받고, 상기 제2 메모리부로부터 상단에서 상기 Y를 각각 입력받는 복수의 PE(Processing Element); 및

[0034] 상기 각각의 PE로부터 출력된 P를 후처리하여 P'로 변형하여 출력하는 후처리부를 포함하는 것을 특징으로 한다.

발명의 효과

[0035] 전술한 구성에 의하여, 본 발명은 시스톨릭 어레이를 구성하는 복수의 프로세싱 엘리먼트(Processing Element)에서 반복되는 전용 로직을 제거하여 PE의 외부의 입력단에 배치하여 전처리함으로써 PE의 연산 지연, 면적, 전력 소모, 하드웨어 자원을 줄일 수 있는 효과가 있다.

도면의 간단한 설명

[0036] 도 1 및 도 2는 종래 기술에 따른 시스톨릭 어레이 구조를 나타낸 도면이다.

도 3은 종래 기술에 따른 시스톨릭 어레이 구조의 상세 구성을 나타낸 도면이다.

도 4는 본 발명의 실시예에 따른 전처리부를 구비한 시스톨릭 어레이 장치의 구성을 나타낸 도면이다.

도 5는 본 발명의 제1 실시예에 따른 전처리부를 구비한 시스톨릭 어레이 장치의 일례를 나타낸 도면이다.

도 6은 본 발명의 제2 실시예에 따른 전처리부를 구비한 시스톨릭 어레이 장치의 다른 일례를 나타낸 도면이다.

도 7은 본 발명의 실시예에 따른 후처리부를 구비한 시스톨릭 어레이 장치의 구성을 나타낸 도면이다.

도 8은 본 발명의 실시예에 따른 후처리부를 구체화한 모습을 나타낸 도면이다.

도 9는 본 발명의 다른 실시예에 따른 후처리부의 구성을 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0037] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0038] 이와 같이, 시스톨릭 어레이는 각각의 PE(10)를 구성하는 복수의 엘리먼트가 처리해야 하는 연산량으로 인하여

하드웨어 자원의 요구가 증가하며, 빠른 속도와 높은 효율을 달성하지 못하는 문제점이 있었다.

[0039] 본 발명의 시스톨릭 어레이 장치는 텐서 프로세싱 유니트(Tensor Processing Unit), GPU(Graphic Processing Unit) 등에 적용될 수 있다.

[0040] 도 4는 본 발명의 실시예에 따른 전처리부를 구비한 시스톨릭 어레이 장치의 구성을 나타낸 도면이다.

[0041] 본 발명의 실시예에 따른 전처리부를 구비한 시스톨릭 어레이 장치는 제1 메모리부(101)에 저장된 값인 X를 X'로 변형하여 전달하는 제1 전처리부(110)와, 제2 메모리부(102)에 저장된 값인 Y를 Y'로 변형하여 전달하는 제2 전처리부(120) 및 2차원 배열로 상호 연결되고, 제1 전처리부(110)가 좌측에서 병렬로 각각 연결되고, 제2 전처리부(120)가 상단에서 병렬로 각각 연결되고, 제1 전처리부(110)로부터 좌측에서 X'를 각각 입력받고, 제2 전처리부(120)로부터 상단에서 Y'를 각각 입력받는 복수의 PE(Processing Element)를 포함한다.

[0042] 제1 전처리부(110)와 제2 전처리부(120)는 PE(130)를 구성하는 복수의 엘리먼트(Element) 중에서 공통되는 일부 엘리먼트 또는 연산 기능을 PE(130)의 외부에 배치하여 전처리함으로써 PE(130)의 내부에서 연산량을 감소시키는 기능을 수행한다.

[0043] 여기서, 엘리먼트는 PE(130)의 내부 구성요소로서 XDL(11), YDL(12), ML(13), 누산기(14) 및 PDL(15)를 나타낸다.

[0044] 시스톨릭 어레이는 제1 메모리부(101)에 저장된 n-bit X를 전처리하여 m-bit X'로 변환한 후, PE(130)로 전달하고, 제2 메모리부(102)에 저장된 n-bit Y를 전처리하여 m-bit Y'로 변환한 후, PE(130)로 전달할 수 있다. 여기서, m은 n보다 큰 정수를 의미한다.

[0045] 다른 실시예로서, 시스톨릭 어레이는 제1 메모리부(101)에 저장된 X를 전처리하여 Signed Digit Number 형태의 X'로 변환한 후, PE(130)로 전달하고, 제2 메모리부(102)에 저장된 Y를 전처리하여 Signed Digit Number 형태의 Y'로 변환한 후, PE(130)로 전달할 수 있다.

[0046] 또 다른 실시예로서, 시스톨릭 어레이는 제1 메모리부(101)에 저장된 Binary Encoding된 X를 전처리하여 Non Binary Encoding 형태의 X'로 변환한 후, PE(130)로 전달하고, 제2 메모리부(102)에 저장된 Binary Encoding 형태의 Y를 전처리하여 Non Binary Encoding 형태의 Y'로 변환한 후, PE(130)로 전달할 수 있다.

[0047] 제1 전처리부(110)와 제2 전처리부(120)의 일례를 하기의 도 5와 도 6에 상세하게 설명한다.

[0049] 도 5는 본 발명의 제1 실시예에 따른 전처리부를 구비한 시스톨릭 어레이 장치의 일례를 나타낸 도면이다.

[0050] 본 발명의 제1 실시예에 따른 시스톨릭 어레이는 각각의 PE(130)에서 반복되는 전용 로직을 제거하여 PE(130)의 외부의 입력단을 나타내는 제1 전처리부(110)와 제2 전처리부(120)에 배치함으로써 PE(130)의 연산 지연, 면적, 전력 소모를 줄일 수 있다.

[0051] 제1 전처리부(110)는 XDL(11)이고, 제2 전처리부(120)는 YDL(12)이며, PE(130)는 ML(13), 누산기(14)를 포함한다.

[0052] 본 발명의 제1 실시예는 PE(130)의 내부에 배치된 XDL(11)과 YDL(12)를 외부에 배치하여 전용 전처리함으로써 각각의 PE(130)에서 수행되는 연산량과 하드웨어 자원을 감소시키는 효과가 있다.

[0054] 도 6은 본 발명의 제2 실시예에 따른 전처리부를 구비한 시스톨릭 어레이 장치의 일례를 나타낸 도면이다.

[0055] 본 발명의 제2 실시예에 따른 시스톨릭 어레이 장치는 각각의 PE(130)에서 반복되는 전용 전처리 로직을 제거하여 PE(130)의 외부에 입력단을 나타내는 제1 전처리부(110)와 제2 전처리부(120)에 배치하여 PE(130)의 연산 지연, 면적, 전력 소모를 줄일 수 있다.

[0056] 제1 전처리부(110)는 부스 리코더(Booth Recoder)를 적용하여 Binary Number를 Signed Digit Number로 변환한다.

[0057] 제2 전처리부(120)는 덧셈기를 이용하여 Y로부터 3Y로 연산한다. 이렇게 전처리된 결과물인 Signed Digit Number, Y, 3Y를 PE(130)에 전달함으로써 PE(130)의 연산 속도를 높일 수 있다.

[0058] 시스톨릭 어레이는 왼쪽 가장자리에서 입력되는 입력 행렬 X을 병렬로 사전 처리하는데, X를 Radix-8 부스 인코딩으로 각각 사전 처리하는 제1 전처리부(110)와, 상단 가장자리에서 입력되는 입력 행렬의 요소 Y를 Y+2Y = 3Y로 각각 사전 처리하는 제2 전처리부(120)를 포함한다.

[0059] 제1 전처리부(110)는 Radix-8 부스 인코딩 처리된 결과를 각각의 PE(130)로 전달한다. 다시 말해, 제1 전처리부(110)는 승수(Multiplier)(X)를 입력받아 이를 부스 알고리즘에 따라 인코딩하고, 인코딩된 승수를 각각의 PE(130)로 전달한다.

[0060] 제2 전처리부(120)는 피승수(Multiplicand) Y+2Y를 사전 처리하여 Y와 3Y를 각각의 PE(130)로 전송한다.

[0061] 부스 인코딩 곱셈기는 승수(Multiplier)의 비트들을 부스 알고리즘에 의해 인코딩하여 부스 인코딩 단위로 피승수(Multiplicand)의 인코드 값을 결정하고, 부스 인코딩 단위의 인코드 값을 이용하여 부분 곱들을 가산함으로써 승수와 피승수의 곱셈 연산을 수행하는 회로이다. 이때, 승수를 인코딩하는 비트수의 단위인 기수(Radix)를 높일수록 승수와 피승수의 부분 곱의 개수를 줄일 수 있다.

[0062] 각각의 PE(130)는 부분곱 생성부, 웰러스 트리(Wallace Tree) 회로 및 캐리 전파 가산기(Carry Propagate Adder)를 포함한다.

[0063] 부분곱 생성부는 제1 전처리부(110)로부터 수신한 인코딩된 승수와 제2 전처리부(120)는 피승수(Multiplicand) Y와 3Y를 이용하여 복수의 부분곱(Partial Product)을 생성한다.

[0064] 웰러스 트리 회로는 복수의 부분곱을 3 : 2의 압축기를 이용하여 제1 부분합과 제2 부분합으로 압축하여 캐리 전파 가산기로 전달한다.

[0065] 캐리 전파 가산기는 제1 부분합과 제2 부분합을 입력받고, 이를 합하여 곱셈 연산을 수행한다.

[0066] 각각의 PE(130)는 부스 인코딩과 피승수 3Y의 연산이 포함되어 있지 않고, Radix 8을 이용하기 때문에 파이프라인 컷 셋(Pipeline Cut Sets) 레지스터 비용을 상당히 감소시킬 수 있다.

[0067] 시스톨릭 어레이는 제1 메모리부(101)에 저장된 X를 부스 인코딩하여 부스 인코딩된 형태의 X'로 변환한 후, PE(130)로 전달하고, 제2 메모리부(102)에 저장된 Y를 Y+2Y = 3Y로 전처리하여 Y, 3Y를 PE(130)로 전달한다.

[0069] 도 7은 본 발명의 실시예에 따른 후처리부를 구비한 시스톨릭 어레이 장치의 구성을 나타낸 도면이고, 도 8은 본 발명의 실시예에 따른 후처리부를 구체화한 모습을 나타낸 도면이다.

[0070] 본 발명의 실시예에 따른 시스톨릭 어레이 장치는 2차원 배열로 상호 연결되고, 입력 행렬 X가 저장된 제1 메모리부(101)가 좌측에서 병렬로 각각 연결되고, 입력 행렬 Y가 저장된 제2 메모리부(102)가 상단에서 병렬로 각각 연결되고, 제1 메모리부(101)로부터 좌측에서 X를 각각 입력받고, 제2 메모리부(102)로부터 상단에서 Y를 각각 입력받는 복수의 PE(Processing Element)(130) 및 각각의 PE(130)로부터 출력된 P를 후처리하여 P'로 변형하여 출력하는 후처리부(140)를 포함한다.

[0071] 후처리부(140)는 복수의 PDL(142)로 이루어져 있다.

[0072] 각각의 PDL(142)는 각각의 PE(130)에서 출력된 Redundant Binary Number P를 처리하여 Binary Number 'P'로 변환한 후 출력할 수 있다.

[0073] 각각의 PDL(142)는 각각의 PE(130)에서 출력된 n-bit P를 후처리하여 m-bit P'로 변환한 후 출력할 수 있다. 여기서, n은 m보다 큰 정수를 의미한다.

[0074] 각각의 PDL(142)는 각각의 PE(130)에서 출력된 Non-Binary Encoding 형태의 P를 전처리하여 Binary Encoding 형태의 P'로 변환한 후 출력할 수 있다.

[0075] 각각의 PDL(142)는 각각의 PE(130)에서 출력된 복수의 Binary 수들의 결합 P를 덧셈기로 덧셈 후 하나의 Binary 수 P'로 변환한 후 출력할 수 있다.

[0076] 각각의 PDL(142)는 각각의 PE(130)에서 출력된 정규화되지 않은 부동 소수점 수 P를 정규화된 부동 소수점 수 P'로 변환하여 출력할 수 있다.

[0077] 각각의 PDL(142)는 각각의 PE(130)에서 출력된 수 P를 올림, 반올림 또는 버림한 어림수 P'로 변환하여 출력할 수 있다.

[0078] 도 9는 본 발명의 다른 실시예에 따른 후처리부의 구성을 나타낸 도면이다.

[0079] 도 9는 도 8의 PDL(142)를 덧셈기로 구성한 일례를 나타낸 것이다.

[0080] 각각의 PDL(142)는 각각의 PE(130)에서 출력된 P=(p0, p1)를 p0 + p1로 덧셈한 P'로 변환하여 출력할 수 있다.

[0081] 이상에서 본 발명의 실시예는 장치 및/또는 방법을 통해서만 구현이 되는 것은 아니며, 본 발명의 실시예의 구성에 대응하는 기능을 실현하기 위한 프로그램, 그 프로그램이 기록된 기록 매체 등을 통해 구현될 수도 있으며, 이러한 구현은 앞서 설명한 실시예의 기재로부터 본 발명이 속하는 기술분야의 전문가라면 쉽게 구현할 수 있는 것이다.

[0082] 이상에서 본 발명의 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속하는 것이다.

부호의 설명

[0083] 101: 제1 메모리부

102: 제2 메모리부

110: 제1 전처리부

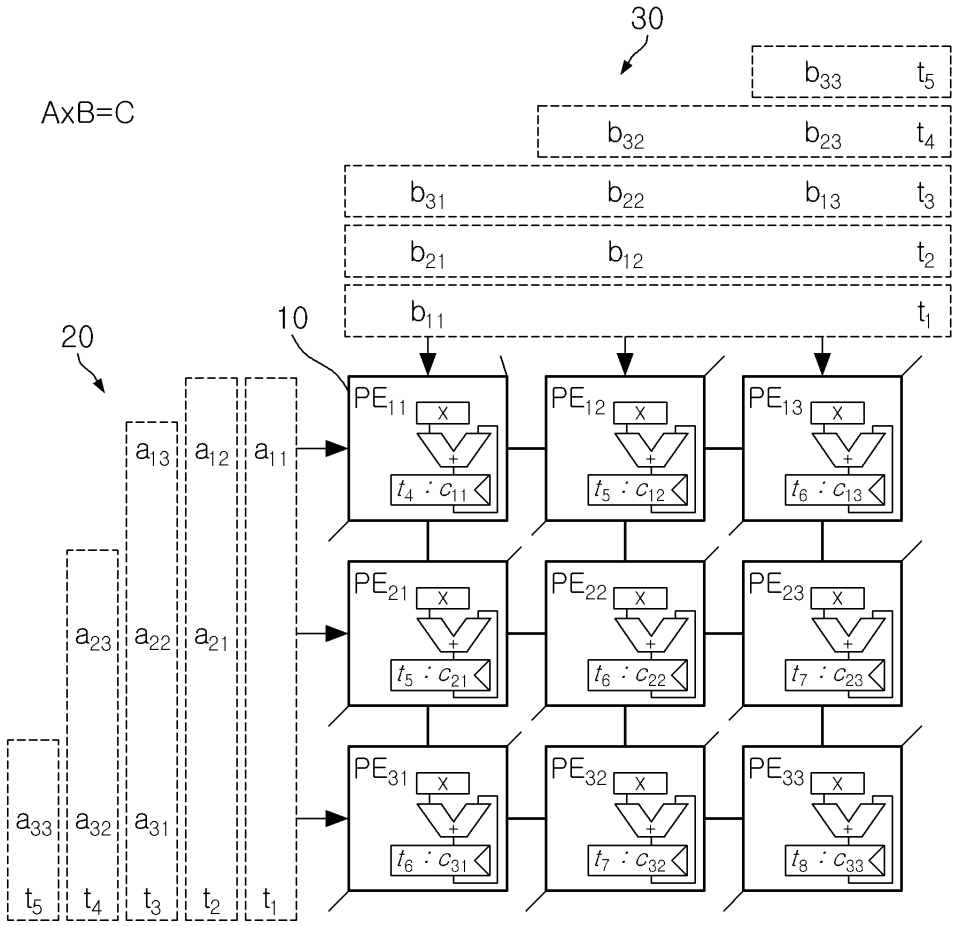
120: 제2 전처리부

130: PE

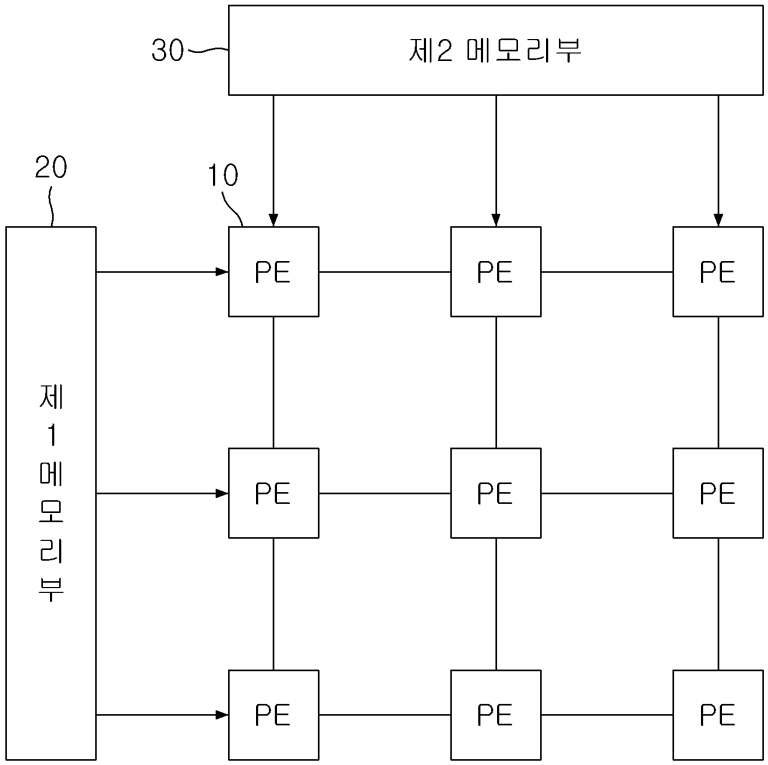
140: 후처리부

도면

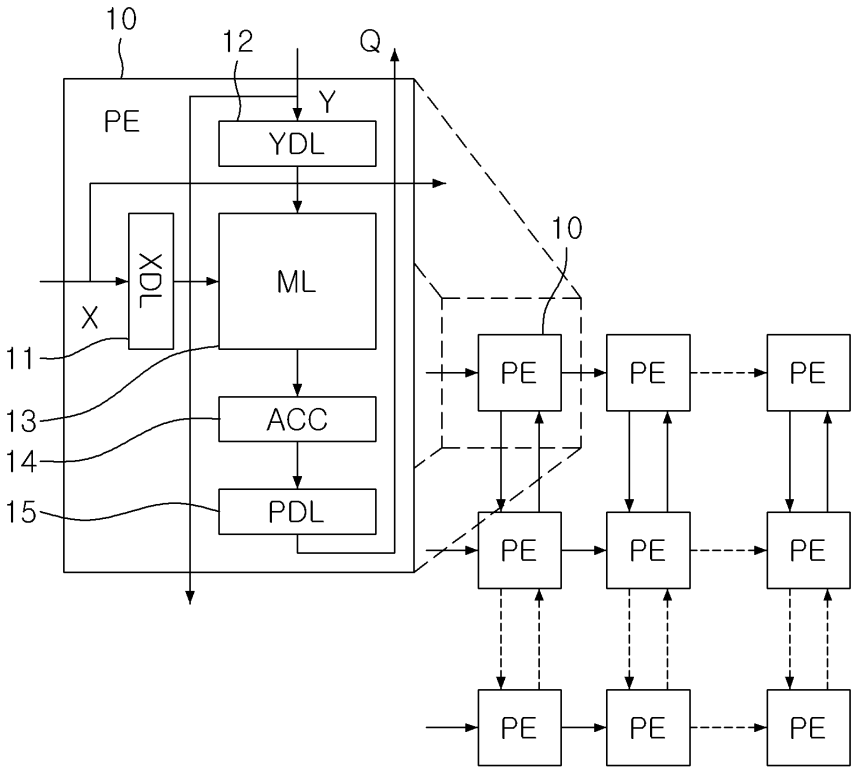
도면1



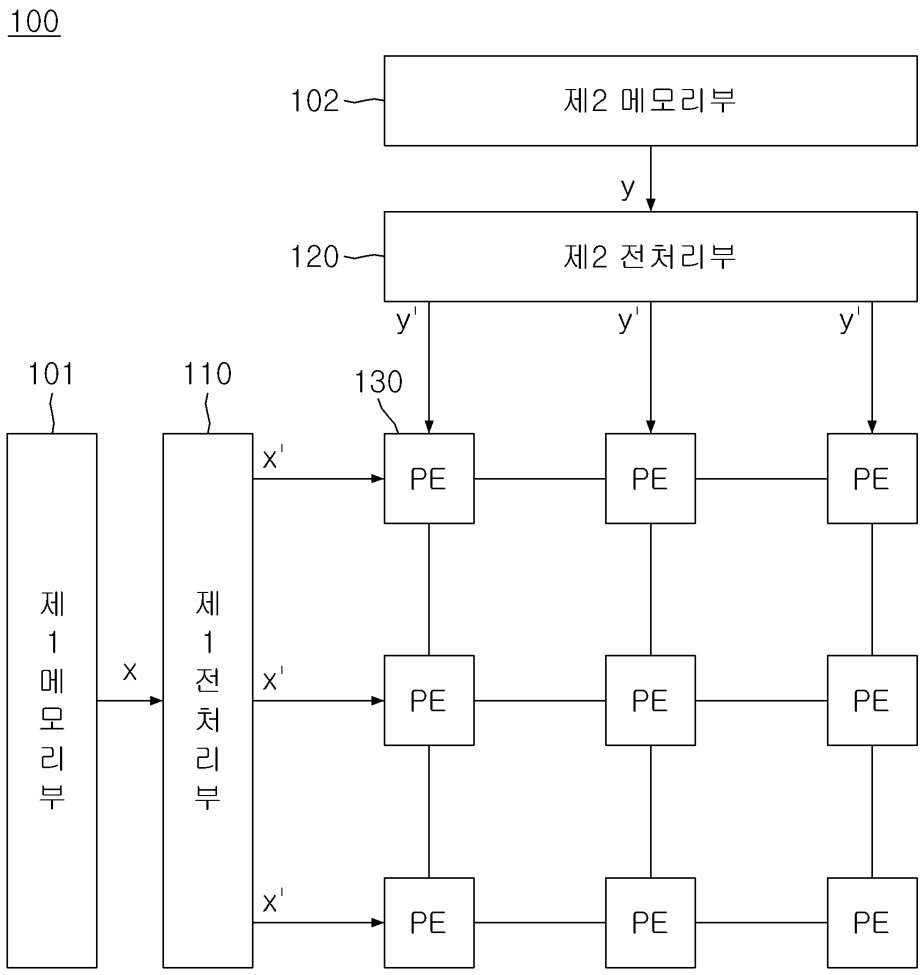
도면2


30
제2 메모리부
20
10
제1메모리부
PE
PE
PE
PE
PE
PE
PE
PE
PE

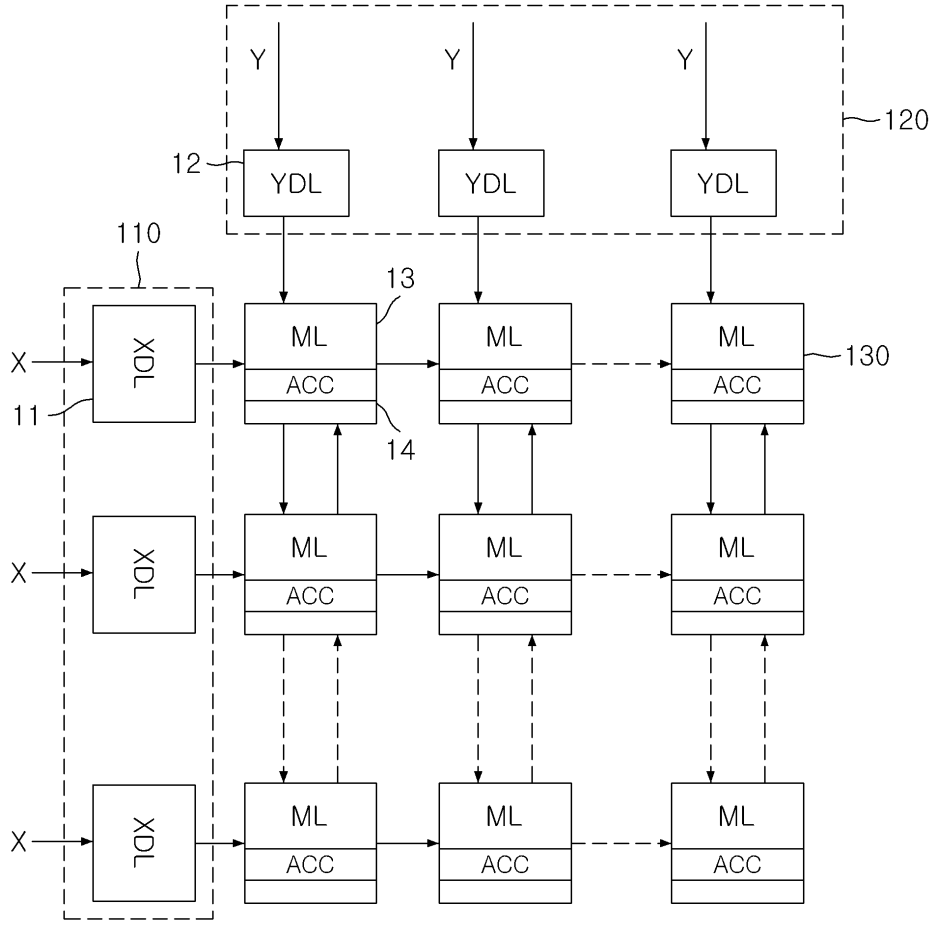
도면3


10
12
Q
PE
Y
YDL
XDL
ML
X
11
13
14
15
ACC
PDL
10
PE
PE
PE
PE
PE
PE
PE
PE
PE

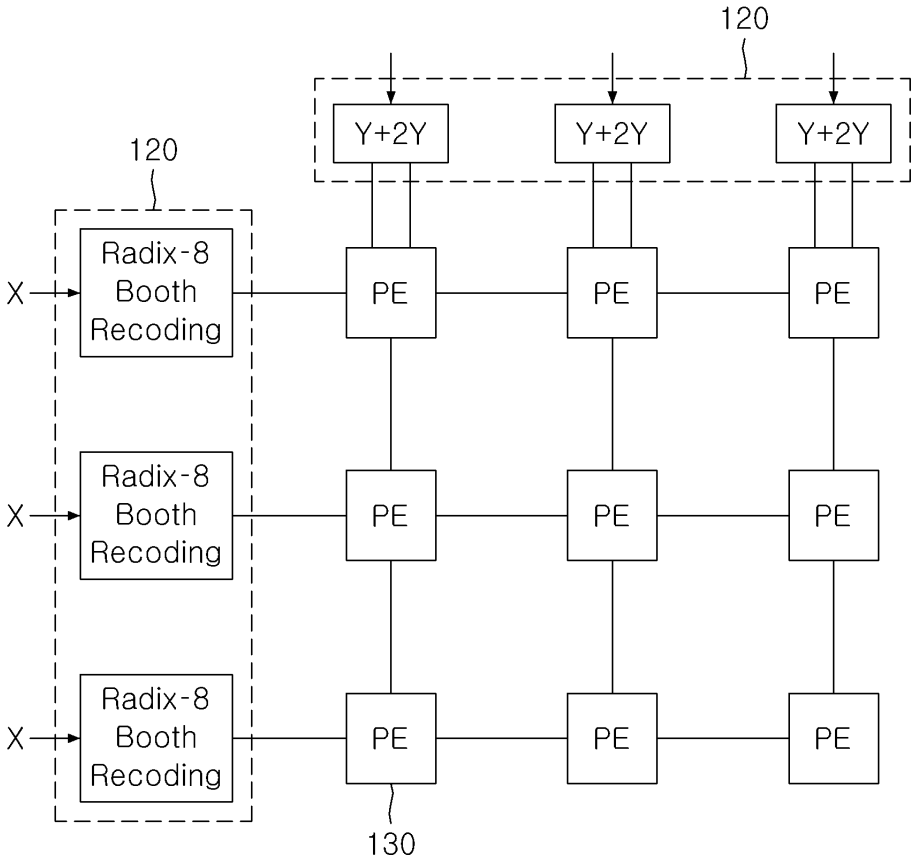
도면4


100
102
제2 메모리부
y
120
제2 전처리부
y'
y'
y'
101
110
130
제 1 메모리부
x
제 1 전처리부
x'
x'
x'
PE
PE
PE
PE
PE
PE
PE
PE
PE

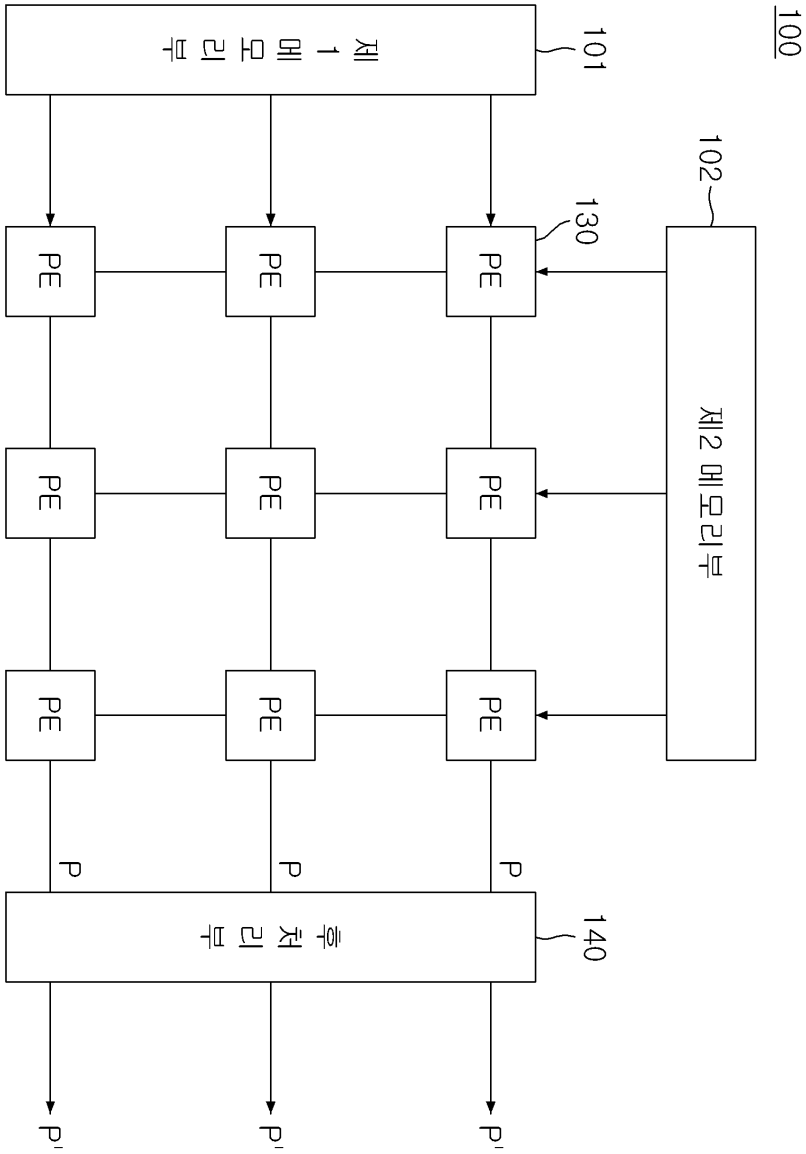
도면5



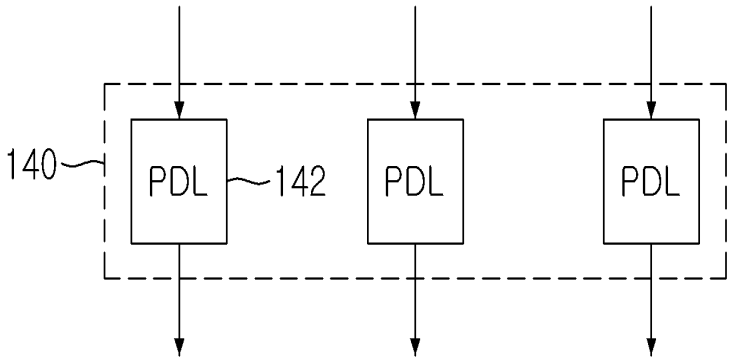
도면6


120
Y+2Y
Y+2Y
Y+2Y
120
Radix-8
Booth
Recoding
X
PE
PE
PE
Radix-8
Booth
Recoding
X
PE
PE
PE
Radix-8
Booth
Recoding
X
PE
PE
PE
130

도면7


100
제1 메모리부
101
102
130
PE
제2 메모리부
P
후처리부
140
P'

도면8


140
PDL
142
PDL
PDL

도면9

