(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2014년08월18일
(11) 등록번호 10-1430553
(24) 등록일자 2014년08월08일

(51) 국제특허분류(Int. Cl.)
G06F 11/08 (2006.01)
(21) 출원번호 10-2013-0066045
(22) 출원일자 2013년06월10일
심사청구일자 2013년06월10일
(56) 선행기술조사문헌
Cheng Huang, Jin Li and Minghua Chen, "On Optimizing XOR-Based Codes for Fault-Tolerant Storage", Applications Information Theory Workshop(ITW), September, 2007.IEEE), pp.218-223
Kwanghee Park et al. "Reliability and Performance Enhancement Technique for SSD array storage system using RAID mechanism", ISCIT 2009
송재석 외, "신뢰성 향상과 성능개선을 위해 다양한 Erasure 코드를 적용한 SSD 기반 RAID-6 시스템 구조" 전자공학회지 47권 c1편 제6호
전체 청구항 수 : 총 3 항

(73) 특허권자
인하대학교 산학협력단
인천광역시 남구 인하로 100, 인하대학교 (용현동)
(72) 발명자
김덕환
서울 양천구 목동로 230, 103동 1004호 (목동, 목동아이파크)
메디히
인천 남구 인하로 115-16, 201호 (용현동)
(74) 대리인
특허법인 정안

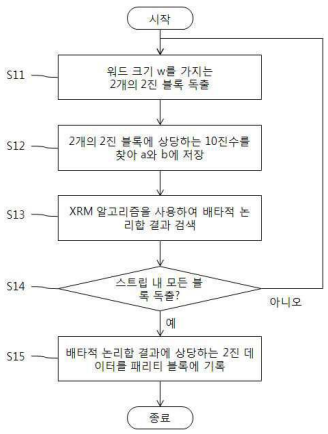
심사관 : 김정진

(54) 발명의 명칭 **에스에스디 기반 레이드 시스템의 인코딩 방법**

(57) 요 약

본 발명에 따른 에스에스디 기반 레이드 시스템의 인코딩 방법에 관한 것으로, 본 발명은 저장장치로부터 소정 워드 크기를 가지는 2개의 2진 블록을 독출하는 단계와, 상기 독출된 2개의 2진 블록에 각각 상당하는 10진수를 피연산자로 배타적 논리합 게이트로 전송하는 단계와, 상기 전송된 피연산자에 대한 배타적 논리합(이하, XOR) 결과를 XOR 참조 매트릭스(XRM)에서 검색하는 단계와, 상기 XOR 결과를 검색한 후, 상기 저장장치 내 모든 블록 독출여부를 확인하는 단계와, 상기 저장장치 내 모든 블록을 독출하면, XOR 결과에 상당하는 2진 데이터를 패리티 블록에 기록하는 단계를 포함하는 것을 특징으로 한다.

대 표 도 - 도5



이 발명을 지원한 국가연구개발사업

과제고유번호	H0502-13-1016
부처명	미래창조과학부
연구사업명	IT/SW창의연구과정
연구과제명	Flash Array를 이용한 Net-Centric 기반의 Cloud Storage 어플라이언스 개발
기 여 율	1/1
주관기관	인하대학교 산학협력단
연구기간	2013.05.01 ~ 2014.04.30

특허청구의 범위

청구항 1

SSD((Solid State Disk) 저장장치에서 SSD 제어기가 데이터 블록에 대하여 소거 코드를 적용하는 인코딩 방법에 있어서,

상기 SSD 제어기는 상기 저장장치로부터 소정 워드 크기를 가지는 2개의 2진 블록을 독출하는 단계와,

상기 SSD 제어기는 상기 독출된 2개의 2진 블록에 각각 상당하는 10진수를 피연산자로 배타적 논리합 게이트로 전송하는 단계와,

상기 SSD 제어기는 상기 전송된 피연산자에 대한 배타적 논리합(이하, XOR) 결과를 XOR 참조 매트릭스(XRM)에서 검색하는 단계와,

상기 SSD 제어기는 상기 XOR 결과를 검색한 후, 상기 저장장치 내 모든 블록 독출여부를 확인하는 단계와,

상기 SSD 제어기는 상기 저장장치 내 모든 블록을 독출하면, XOR 결과에 상당하는 2진 데이터를 패리티 블록에 기록하는 단계를 포함하고,

상기 XOR 결과 검색 단계는,

규칙이 매칭되지 않으며 워드 크기가 2가 아니면 대칭 규칙에 의해 분할된 4개의 영역 중 상기 XOR 결과의 위치를 검색하는 단계와,

상기 검색된 XOR 결과의 위치가 제2영역 또는 제3영역이면 상기 검색된 XOR 결과의 위치를 제2영역 또는 제3영역에서 제1영역으로 이동시키는 단계를 포함하는 것을 특징으로 하는 인코딩 방법.

청구항 2

제1항에 있어서,

상기 SSD 제어기는 상기 저장장치 내 모든 블록을 독출하지 않은 경우, 다음 2개의 2진 블록을 독출하는 단계를 더 포함하는 것을 특징으로 하는 인코딩 방법.

청구항 3

삭제

청구항 4

제1항에 있어서,

상기 SSD 제어기는 상기 검색된 XOR 결과의 위치가 제4영역이면 상기 검색된 XOR 결과의 위치를 제1영역으로 이동시키는 단계를 더 포함하는 것을 특징으로 하는 인코딩 방법.

명 세 서

기 술 분 야

[0001] 본 발명은 SSD 기반 RAID-6 시스템의 인코딩 성능 및 에너지 소비를 개선하기 위한 에스에스디 기반 레이드 시스템의 인코딩 방법에 관한 것이다.

배 경 기 술

[0002] 저장 시스템 제조업체들은 에스에스디(Solid State Disk: SSD)용으로 소모되는 작은 공간과 함께 고성능의 측면

의 돌파구를 만들 것을 연구자들에게 요구하고 있다. 소거 코드는 고장에 저항하도록 저장 시스템에 채용된다.

[0003] 레이드(Redundant Array of Independent Disk: 이하, RAID)-6 시스템에서의 소거 코딩은 RAID-1 시스템에서의 데이터 복제보다 적은 수의 디스크를 필요로 한다. 소거 코딩에서 인코딩 과정은 데이터 저장이 벡터로 기술되고 코딩 저장은 소거 코드로 기술되는 갈루아체(Galois Field) 연산에서의 매트릭스 벡터 연산을 이용하여 이루어진다. 코딩 매트릭스 내에는 비트 매트릭스를 만드는 1과 0만 있으므로, 매트릭스-벡터 연산은 XOR 연산을 이용하여 정의된다. 따라서, 소거 코드의 성능은 인코딩이 실행되는 동안 연산되는 XOR 연산의 수에 좌우된다. 몇 가지 방법들이 인코딩 성능의 향상을 위해 제안되었다. 이러한 문제를 개선하기 위한 최근의 방법 두가지는 부호 특정 혼성 복원(Code Specific Hybrid Reconstruction)과 공통 영역식(common subexpression)을 이용하여 알고리즘을 매칭하는 것이다.

발명의 내용

해결하려는 과제

[0004] 본 발명은 상기한 종래기술에 근거하여 안출된 것으로, SSD 기반 RAID-6 시스템의 성능, 공간 효율, 및 에너지 소모를 개선하여 SSD 기반 RAID-6 시스템용 플랫폼의 성능을 향상시키는 에스에스디 기반 레이드 시스템의 인코딩 방법을 제공하는데 목적이 있다.

과제의 해결 수단

[0005] 상기한 과제를 해결하기 위하여, 본 발명에 따른 에스에스디 기반 레이드 시스템의 인코딩 방법은 저장장치로부터 소정 워드 크기를 가지는 2개의 2진 블록을 독출하는 단계와, 상기 독출된 2개의 2진 블록에 각각 상당하는 10진수를 피연산자로 배타적 논리합 게이트로 전송하는 단계와, 상기 전송된 피연산자에 대한 배타적 논리합(이하, XOR) 결과를 XOR 참조 매트릭스(XRM)에서 검색하는 단계와, 상기 XOR 결과를 검색한 후, 상기 저장장치 내 모든 블록 독출여부를 확인하는 단계와, 상기 저장장치 내 모든 블록을 독출하면, XOR 결과에 상당하는 2진 데이터를 패리티 블록에 기록하는 단계를 포함하는 것을 특징으로 한다.

[0006] 또한, 본 발명은 상기 저장장치 내 모든 블록을 독출하지 않은 경우, 다음 2개의 2진 블록을 독출하는 단계를 더 포함하는 것을 특징으로 한다.

[0007] 또한, 상기 XOR 결과 검색 단계는, 규칙이 매칭되지 않으며 워드 크기가 2가 아니면 대칭 규칙에 의해 분할된 4개의 영역 중 상기 XOR 결과의 위치를 검색하는 단계와, 상기 검색된 XOR 결과의 위치가 제2영역 또는 제3영역이면 상기 검색된 XOR 결과의 위치를 제2영역 또는 제3영역에서 제1영역으로 이동시키는 단계를 포함하는 것을 특징으로 한다.

[0008] 또한, 본 발명은 상기 검색된 XOR 결과의 위치가 제4영역이면 상기 검색된 XOR 결과의 위치를 제1영역으로 이동시키는 단계를 더 포함하는 것을 특징으로 한다.

발명의 효과

[0009] 본 발명은 XRM을 통해 액세스 시간을 빠르게 개선하며 메모리 사용량을 줄일 수 있다.

[0010] 또한, 본 발명은 기존의 XOR 연산 기반 소거 코드와 비교할 때 소거 코드를 인코딩하거나 디코딩하는 중앙처리장치의 오버헤드를 제어할 수 있다.

도면의 간단한 설명

[0011] 도 1은 본 발명에 따른 SSD 기반 RAID-6 시스템의 전체 구조.

도 2는 본 발명과 관련된 3개의 데이터 장치와 2개의 코딩 장치를 가지는 소거 코드를 도시한 일 예.

도 3은 본 발명과 관련된 XOR 참조 매트릭스를 도시한 도면.

도 4는 본 발명과 관련된 w=16일 때 XOR 연산에 대한 패턴의 클러스팅을 도시한 예.

도 5는 본 발명에 따른 인코딩 과정을 도시한 흐름도.

도 6은 본 발명에 따른 XRM 알고리즘을 도시한 흐름도.

발명을 실시하기 위한 구체적인 내용

[0012] 이하, 첨부된 도면들을 참조하여 본 발명의 실시예를 상세하게 설명한다.

[0013] 본 발명은 대용량 저장장치로 사용되는 기존의 하드디스크 기반의 레이드(Redundant Array of Inpendent Disk: 이하, RAID) 시스템에서 하드디스크를 에스에스디(Solid State Disk: 이하, SSD)로 대체한 RAID 시스템을 제안한다. 그리고, 본 발명에서 사용되는 RAID 시스템은 2개의 패리티(parity)를 사용하는 RAID-6 시스템이고, 이 RAID-6 시스템에 소거 코드(erasure code)를 적용한다.

[0014] 도 1은 본 발명에 따른 SSD 기반 RAID-6 시스템의 전체 구조를 도시한다.

[0015] 도 1에 도시된 바와 같이, SSD 기반 RAID 시스템은 RAID-6 계층과 SSD 저장 플랫폼으로 구성된다.

[0016] RAID-6 계층의 경우, RAID-6 제어기는 소거 코드를 채택하는 인코딩 및 디코딩 스케줄러, 소거 코드를 사용하는 내부 고장 검출기를 가지는 임의 및 순차 읽기/쓰기 모듈, 추적 정보를 갱신하는 로그 생성기를 포함한다.

[0017] 멀티 SSD 저장 플랫폼은 호스트 인터페이스 계층(host interface layer: HIL), 플래시 변환 계층(flash transslation layer: FTL), 낸드(NAND) 플래시 칩으로 구성된다.

[0018] 플래시 변환 계층은 플래시 메모리와 파일 시스템 사이에 위치하며 플래시 메모리를 디스크처럼 블록 디바이스로 사용하게 하는 것으로, 파일 시스템에서 요구하는 블록을 플래시 메모리의 블록으로 맵핑(mapping)하는 역할을 한다.

[0019] 도 2는 본 발명과 관련된 3개의 데이터 장치와 2개의 코딩 장치를 가지는 소거 코드를 도시한 일 예이다.

[0020] 도 2를 참조하면, 3개의 데이터 장치와 2개의 코딩 장치를 가지는 RAID-6 소거 코드의 예를 보여준다. 데이터 장치들은 $\{D_{0,0}, D_{0,1}, \ldots, D_{1,0}, D_{i,j}\}$으로 설명된다. 더욱이, 코딩 장치들은$\{P_{0,0}, P_{0,1}, \ldots, P_{j,0}\}$와 $\{Q_{0,0}, Q_{0,1}, \ldots, Q_{j,0}\}$로도 나타낼 수 있고, 여기서, i는 데이터 장치의 수를 나타내고 j는 소거 코드 워드의 수에 기반한 청크 크기(chunk size)를 나타낸다

[0021] 배타적 논리합(XOR) 연산 동안, 각 2진 워드는 그에 대응되는 10진수로 나타낼 수 있다. 하지만, 운영시스템 성능에 따른 메모리 한계로 인해 5 워드를 초과하는 숫자에 대해 XOR 연산을 실행하면 오버플로우(overflow)가 나타날 수 있다.

[0022] 본 발명은 RAID-6 시스템에서 소거 코드의 계산을 위한 XOR 연산의 실행시간을 줄이기 위해 XOR 참조 매트릭스(XOR reference matrix: 이하, XRM)을 사용한다. 본 발명은 XOR 연산을 실시간으로 실행하는 대신에 XRM을 통해 결과를 검색할 수 있다. 하지만, 워드 크기가 증가할수록, 복잡도는 높아진다. 도 3에 기반한 XOR 연산의 특징들은 표 1에 표시하였다. XRM의 특성을 분석하여 인코딩 과정을 채택하는데 적용하였다.

표 1

[0023]

규칙	a	b	c	조건
1	a	0	n	$0 < a < m$
2	0	b	n	$0 < b < m$
3	a	b	0	$a = b$
4	a	b	m	$a + b = m$
5	m	b	m-b	$0 < b < m$
6	a	m	m-a	$0 < b < m$

XOR (a, b) = c; m= {3, 7, 15, 31, 63, 127, 255 ... }

[0024] 표 1에서, a와 b는 XOR 연산의 피연산자를 나타내고 c는 XOR 연산의 결과값이다. 대칭 규칙이라 불리는 이러한 특징들을 XRM을 최적화하는데 이용할 것이다. 이외에도, 규칙 7 및 8, 9를 사용하여 메모리 내에 XRM을 저장하는데 필요한 공간을 축소할 수 있다. 도 4에서, 다른 색상비(color rate)를 이용하여 XOR 연산 결과를 클러스터링(clustering)하여 패턴을 분석할 수 있다. 각 포인트는 상당하는 10진수를 나타내는 0에서 255까지의 파란색으로 기술되었다. 이 수치는 규칙 7, 8, 9 규칙들이 워드 크기의 다른 값을 가지는 XRM에 적용될 수 있음을 보여준다. $c_{a,b}$는 a와 b의 XOR 연산의 결과이다.

[0025] **규칙 7**: a나 b가 2^{w-1} 보다 작다면 결과 위치 $c_{a,b}$는 4개 영역 중에서 제2영역(2) 또는 제3영역(3) 안에서 찾을 수 있다. 그러나, a와 b가 2^{w-1} 보다 크다면 결과 위치는 제4영역(4) 안에 있다.

[0026] **규칙 8**: 도 4에 도시된 바와 같이, 대칭 특성은 XRM을 4개의 영역으로 분할하고, 제1영역(1)과 제4영역(4) 또는 제2영역(2)과 제3영역(3)은 같은 결과를 가진다. 이러한 이점을 활용하여 XRM의 크기를 실제 크기보다 50% 작게 줄일 수 있다. $a=a-2^{w-1}$ 및 $b=b-2^{w-1}$ 를 적용하므로 결과 위치를 제4영역에서 제1영역으로 이동시킨다.

[0027] **규칙 9**: 도 3에 도시된 바와 같이, 제1영역과 제3영역 또는 제1영역과 제2영역은 유사한 패턴을 가진다. a 또는 b로부터 오프셋(offset) 2^{w-1} 를 빼므로, 이것들의 XOR 연산은 동일한 결과값을 얻을 수 있다. 따라서, $a=a-2^{w-1}$ 및 $b=b-2^{w-1}$ 를 적용함으로써 XRM의 검색 범위를 좁힐 수 있다.

[0028] a) $C_{a,b}$가 XRM의 제2영역 안에 위치된다면,

[0029] $C_{a,b}=C_{a-2^{w-1},b}+2^{w-1}$;

[0030] 여기서, $2^{w-1}=C_{a,b}-C_{a-2^{w-1},b}$ 이고,

[0031] b) $C_{a,b}$가 XRM의 제3영역 안에 위치된다면,

[0032] $C_{a,b}=C_{a,b-2^{w-1}}+2^{w-1}$;

[0033] 여기서, $2^{w-1}=C_{a,b}-C_{a,b-2^{w-1}}$ 이다.

[0034] 도 5는 본 발명에 따른 인코딩 과정을 도시한 흐름도이다.

[0035] 본 발명의 소거 코드 방법은 소거 코드의 종류와 무관한 XRM을 사용함으로써 인코딩 프로세스를 위한 소거 코드를 채용한다.

[0036] 도 5에 따르면, SSD 제어기(SSD 저장 플랫폼)는 w^2 인 블록 크기를 가지는 2개의 2진 블록 D_1과 D_2을 RAID 제어기의 스트립(strip)으로부터 독출한다(S11). 여기서, 스트립은 데이터를 저장하는 저장장치이다.

[0037] SSD 제어기는 2진 블록들에 상당하는 10진수를 찾아 XOR의 피연산자 a와 b로 저장한다(S12). 즉, RAID 제어기는 스트립으로부터 독출한 2개의 2진 블록에 각각 상당하는 10진수를 SSD 제어기의 XOR 게이트로 전송한다.

[0038] SSD 제어기는 XRM 알고리즘을 이용하여 피연산자 a와 b의 XOR 연산 결과를 검색한다(S13). 즉, SSD 제어기는 피연산자 a와 b의 XOR 연산 결과 위치를 XOR 참조 매트릭스에서 검색한다.

[0039] SSD 제어기는 스트립 내 존재하는 모든 블록을 읽었는지를 확인한다(S14). 예를 들어, 스트립트 내에 2진 블록 D_1 내지 D_4이 존재하는 경우, 마지막 2진 블록인 D_4까지 독출했는지를 확인한다.

[0040] SSD 제어기는 상기 피연산자 a와 b의 XOR 연산 결과를 낸드 플래시 칩의 패리티 블록에 기록한다(S15).

[0041] 상기 단계(S14)에서 상기 저장장치 내 모든 블록을 독출하지 않은 경우, 다음 2개의 2진 블록을 독출하여 상기 제12단계(S12) 내지 제13단계(S13)를 다시 수행한다. 즉, 저장장치 내 모든 블록을 독출할 때까지 제11단계(S11) 내지 제14단계(S14)를 반복적으로 수행한다.

[0042] 도 6은 본 발명에 따른 XRM 알고리즘을 도시한 흐름도이다.

[0043] 본 발명에서 XRM 알고리즘은 실제로 XOR 연산을 수행하는 대신 표 1의 대칭 규칙을 사용하는 XRM에서 피연산자 a와 b의 XOR 연산 결과를 검색한다.

[0044] 도 6에 도시된 바와 같이, SSD 제어기는 두 개의 10진수 a와 b(피연산자)로 입력되는 인코딩 데이터로부터 크기가 w^2 인 2비트(bit) 매트릭스를 읽는다(S131). 여기서, w는 인코딩 프로시저의 워드 크기이다.

[0045] XRM 알고리즘은 루프 조건이 만족하는 동안, 제132단계(S132)부터 제140단계(S140)를 반복하여 주어진 a와 b에 대하여 표 1의 대칭 규칙을 하나씩 점검한다. w의 값이 감소함에 따라 점검을 반복할 때마다 XRM의 검색 범위가 좁아진다.

[0046] SSD 제어기는 규칙이 매칭되지 않고 워드 크기가 2가 아니면, 규칙 7을 적용하여 XRM의 4개 영역들 중에서 피연산자의 XOR 결과 위치를 검색한다(S132, S133).

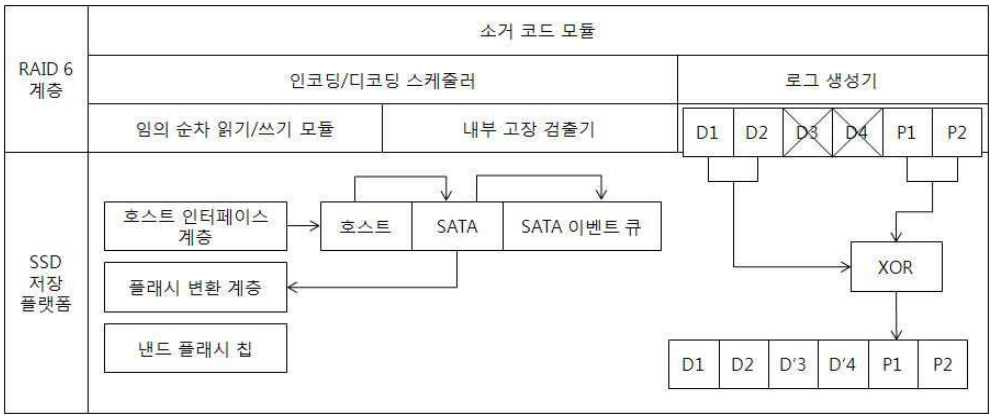
[0047] 상기 검색된 XOR 결과 위치가 제2영역 또는 제3영역이면, 규칙 9를 적용하여 상기 검색된 XOR 결과 위치를 제2영역 또는 제3영역에서 제1영역으로 이동시킨다(S134, S135). 그리고, 범위는 오프셋 2^{w-1}를 반영하여 갱신한다. 즉, a 또는 b의 값을 감소시킴으로써 상기 검색된 XOR 결과 위치를 이동시킨다.

[0048] 상기 검색된 XOR 결과 위치가 제4영역이면 규칙 8을 적용하여 제4영역에서 제1영역으로 이동시킨다(S137, S138).

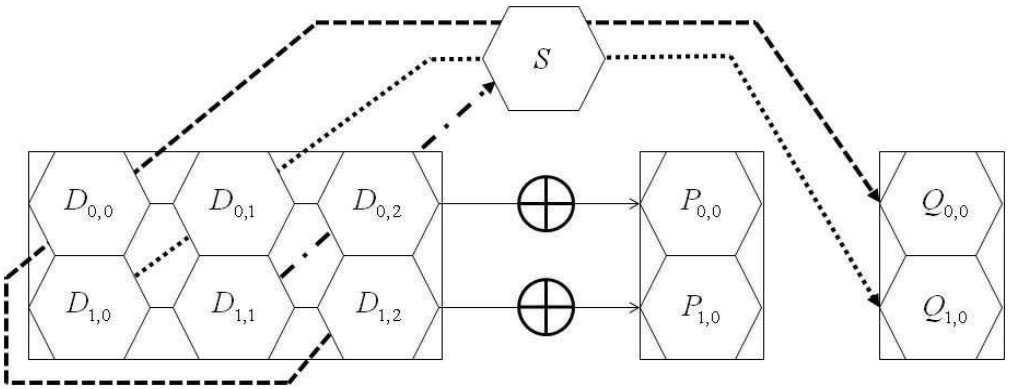
[0049] 그리고, SSD 제어기는 워드 크기 w의 값을 감소시키고, a와 b의 XOR 연산 결과와 범위의 합을 결과로 갱신하여 도 5의 단계(S14)로 리턴한다.

도면

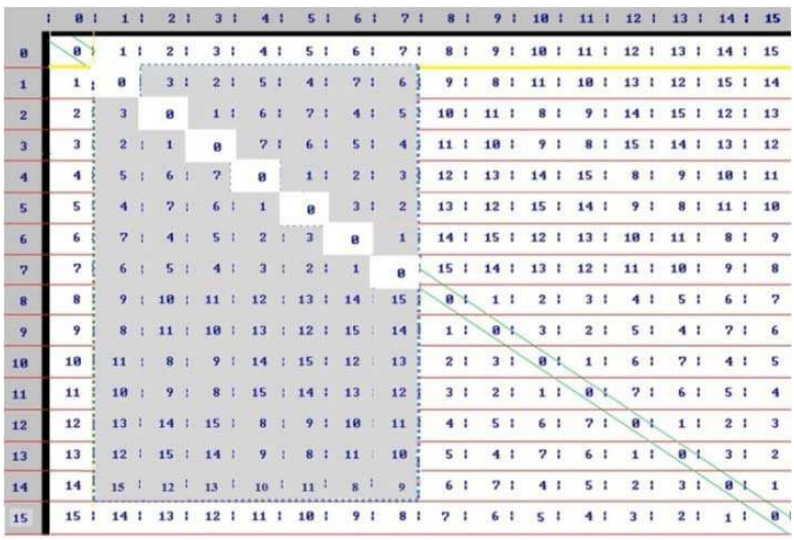
도면1



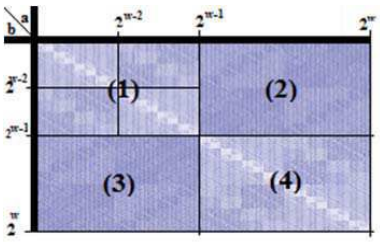
도면2



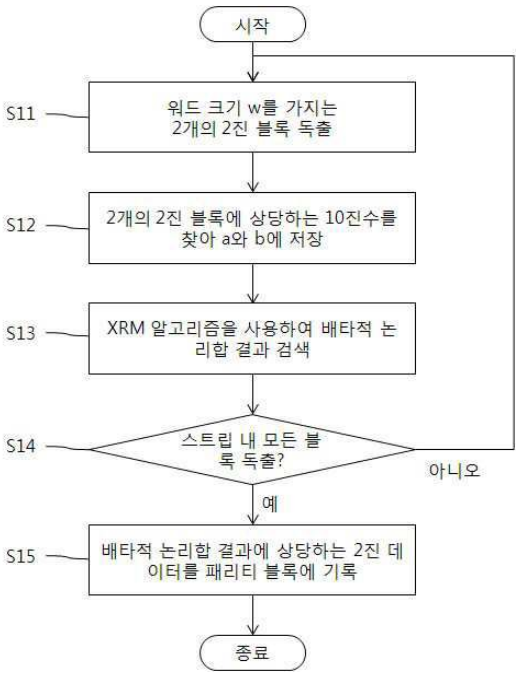
도면3



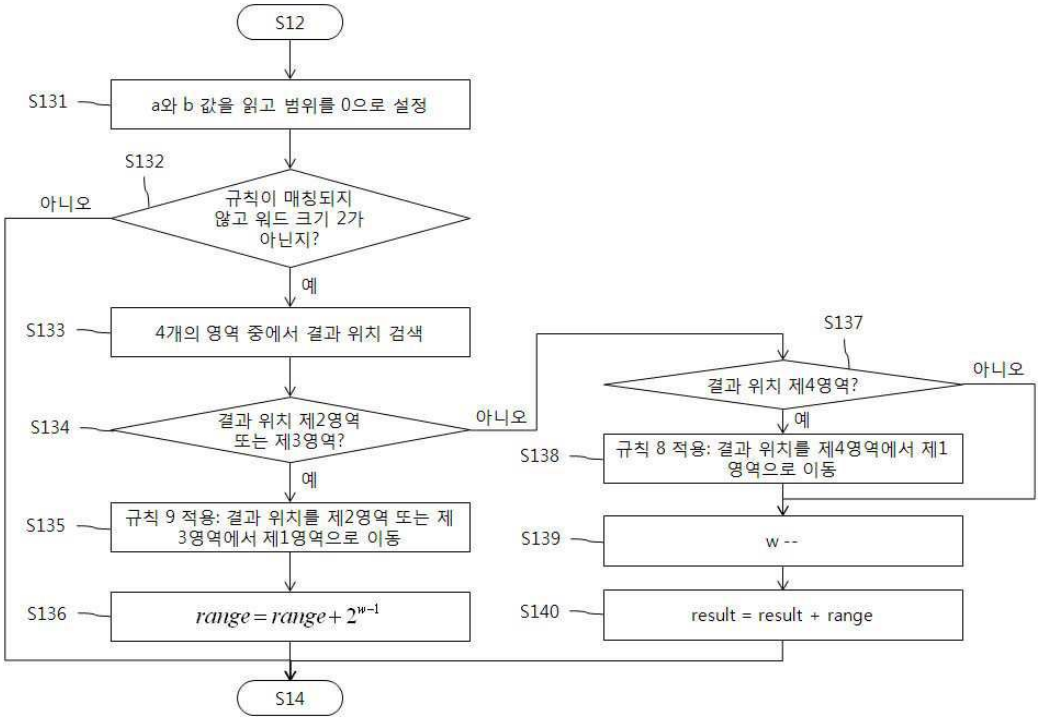
도면4


a
b
2^{w-2}
2^{w-1}
2^{w}
(1)
(2)
(3)
(4)

도면5


시작
S11
워드 크기 w를 가지는 2개의 2진 블록 독출
S12
2개의 2진 블록에 상당하는 10진수를 찾아 a와 b에 저장
S13
XRM 알고리즘을 사용하여 배타적 논리합 결과 검색
S14
스트립 내 모든 블록 독출?
아니오
예
S15
배타적 논리합 결과에 상당하는 2진 데이터를 패리티 블록에 기록
종료

도면6


S12
S131
a와 b 값을 읽고 범위를 0으로 설정
S132
규칙이 매칭되지 않고 워드 크기 2가 아닌지?
아니오
예
S133
4개의 영역 중에서 결과 위치 검색
S134
결과 위치 제2영역 또는 제3영역?
아니오
예
S135
규칙 9 적용: 결과 위치를 제2영역 또는 제3영역에서 제1영역으로 이동
S136
$range = range + 2^{w-1}$
S137
결과 위치 제4영역?
아니오
예
S138
규칙 8 적용: 결과 위치를 제4영역에서 제1영역으로 이동
S139
w --
S140
result = result + range
S14