All about Physical Data Recovery

: How to recover overwritten data from magnetic disk

proneer

proneer@gmail.com

http://forensic-proof.com

JK Kim

Outline

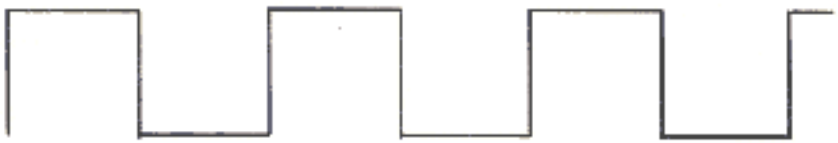
1. 데이터 기록 방식 (Data Recoding Mode)
2. 데이터 인코딩 기법 (Data Encoding Techniques)
3. 데이터 완전삭제 기법 (Data Wiping Techniques)
4. 물리적 데이터 복구 가능성 (Probability of Physical Data Recovery)

Data Recoding Mode

기본 개념 (Basics)

- **데이터(Data)** (pron.: /ˈdeɪtə/ day-tə or /ˈdætə/)
  - the **quantities**, **characters**, or **symbols** on which operations are performed by a computer, being stored and transmitted **in the form of electrical signals** and recorded on **magnetic**, **optical**, or **mechanical recording media** – *Wikipedia*
  - In computer science, data is anything **in a form suitable for use with a computer** – *Wikipedia*

Analog Signal



Digital Signal

저장매체 기록 방식 (Storage Recoding Mode)

- **자기 기록(Magnetic Recording)**
  - Hard Disk, Magnetic Tape, Magnetic Stripe Card, …
- **광학 기록(Optical Recording)**
  - Compact Disk(CD), Digital Versatile Disk(DVD), …
- **전자 기록(Electronic Recording)**
  - RAM, ROM, Flash Memory, …

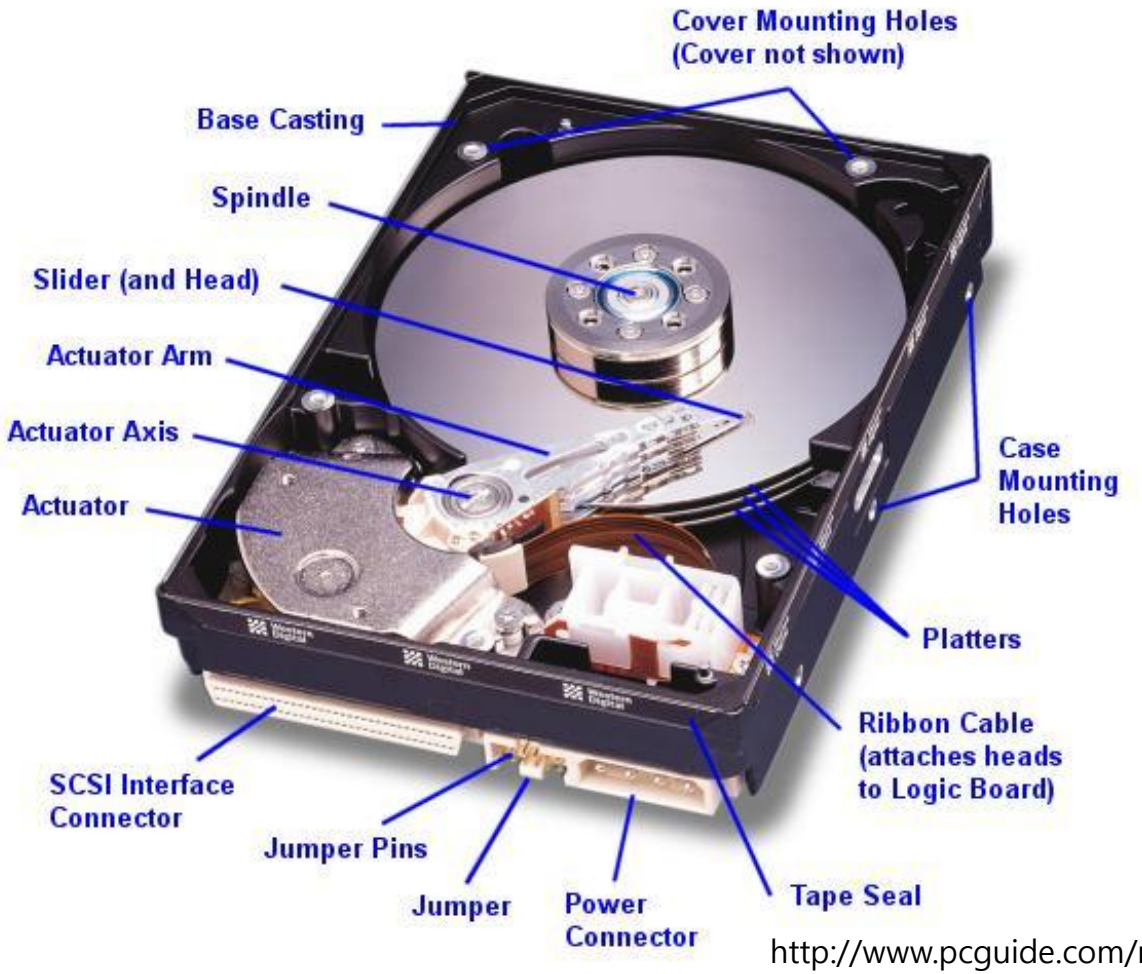
하드디스크 드라이브 (HDD)

- 디스크 내부

http://www.pcguide.com/ref/hdd/index.htm

하드디스크 드라이브 (HDD)

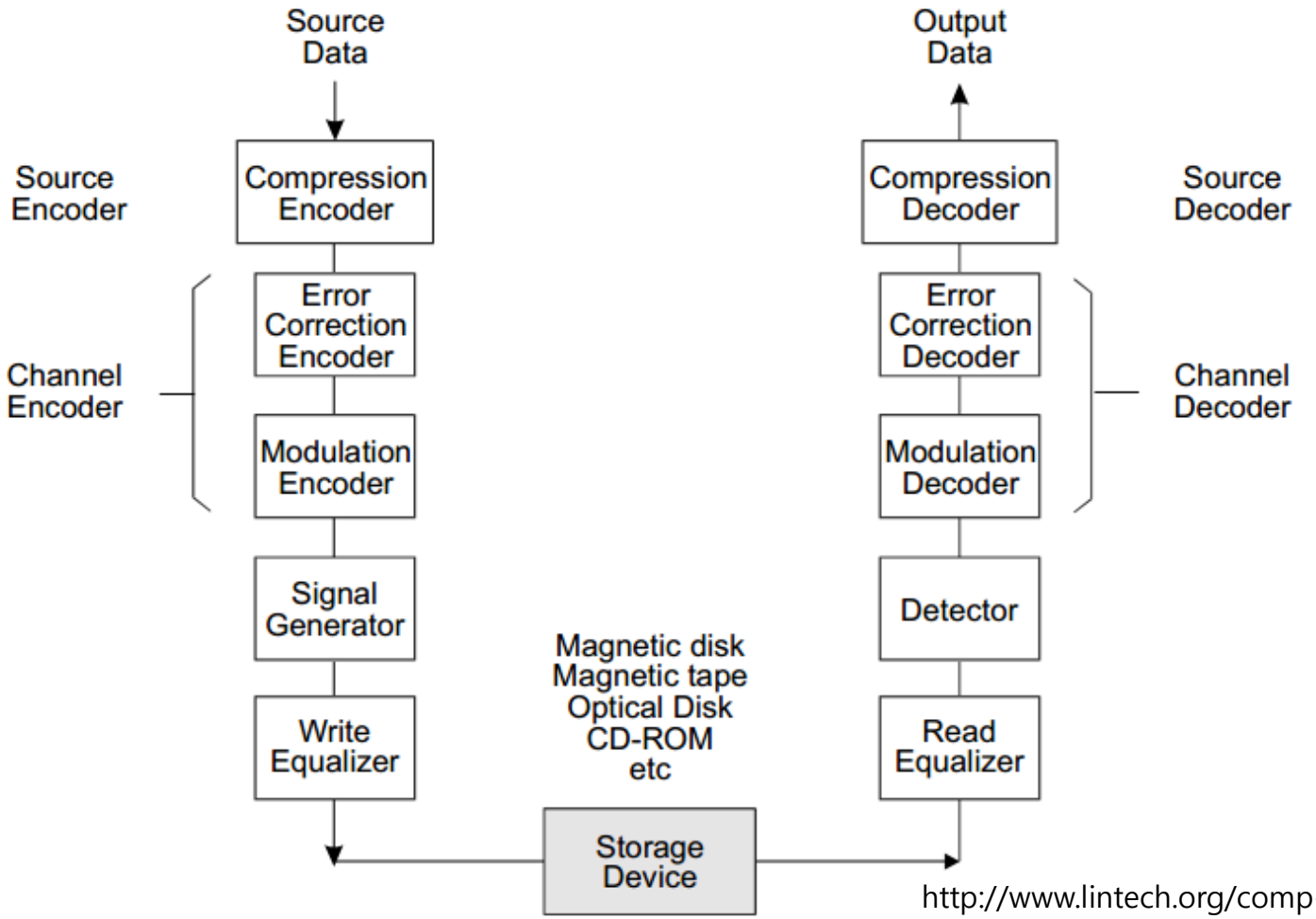
- 주요 구성 요소



http://www.pcguide.com/ref/hdd/op/index.htm

하드디스크 드라이브 (HDD)

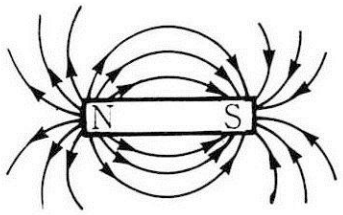
- 디지털 데이터 읽기/쓰기 채널



http://www.lintech.org/comp-per/07MAGREC.pdf

자성체 (Magnetic Substance)

- **자기장 내에서 자화되는 물질 ➔ 지구상의 모든 물질**



- **자성체 분류**
  - **강자성체(ferromagnetic substance)** : 외부의 강한 자기장이 있을 때, 그 자기장 방향으로 강하게 자화된 뒤 자기장이 사라져도 자화가 남아있는 물질 (철, 코발트, 니켈 등)
  - **반자성체(diamagnetic substance)** : 외부 자기장에 의해 자기장과 반대 방향으로 자화되는 물질 (금속과 산소를 제외한 기체, 물 등)
  - **상자성체(paramagnetic substance)** : 자기장안에서는 자기장 방향으로 약화게 자화하고, 자기장이 제거되면 자화하지 않는 물질 (알루미늄, 주석, 백금, 이리듐 등)

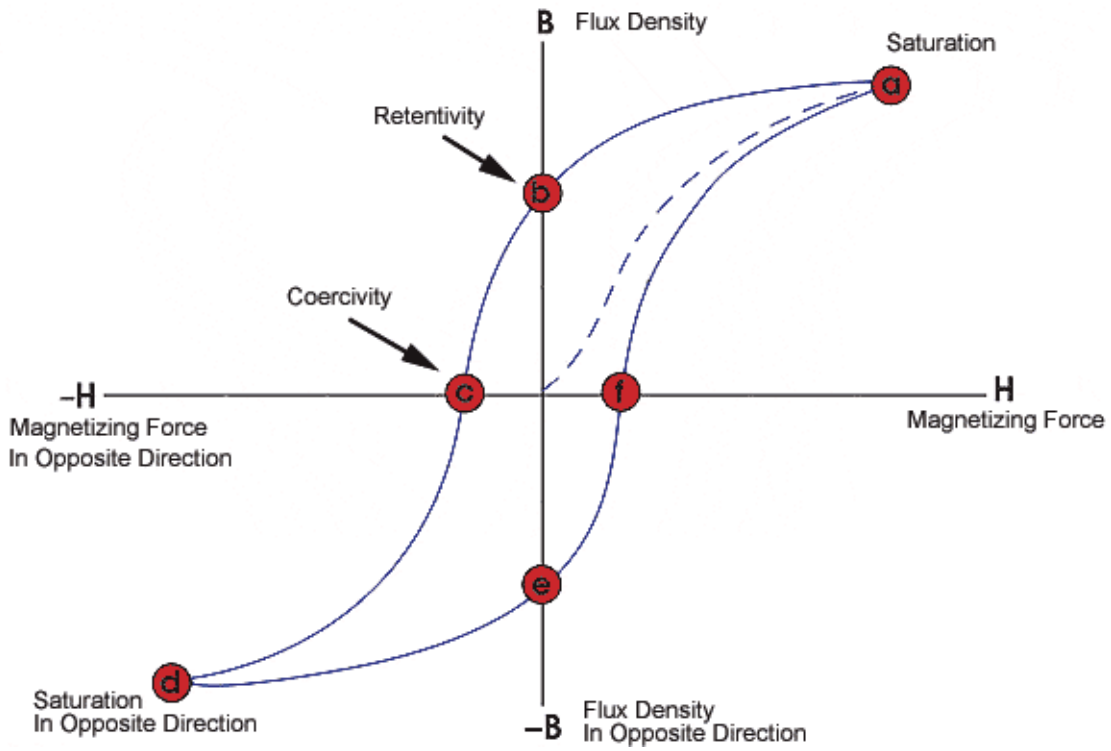
자성체 (Magnetic Substance)

- **자성체의 특징 (Magnetic Material Characteristics)**
  - **포화 상태(Saturation)** : 자화되지 않은 상태에서 자기장 세기(H)를 증가시켰을 때 더 이상 증가하지 않는 상태
  - **잔류 자기(Retentivity, Remanence)** : 자기 포화상태에서 자화를 감소하여, 자기장를 제거했을 때 자성체에 남아있는 자력, 강자성체는 잔류 자기가 남기 쉬움 (영구 자석)
  - **보자력(Coercivity) :** 잔류 자기를 없애기 위해(0으로) 역방향으로 가해야 하는 자기장의 세기
  - **자속 밀도(magnetic flux density)** : 자기장의 크기를 나타내는 것으로 단위 면적을 수직으로 지나는 자기력선의 수
  - **자기력(magnetic force)** : 두 극 사이에 작용하는 힘으로 서로 밀거나 당기는 힘

http://www.lintech.org/comp-per/07MAGREC.pdf

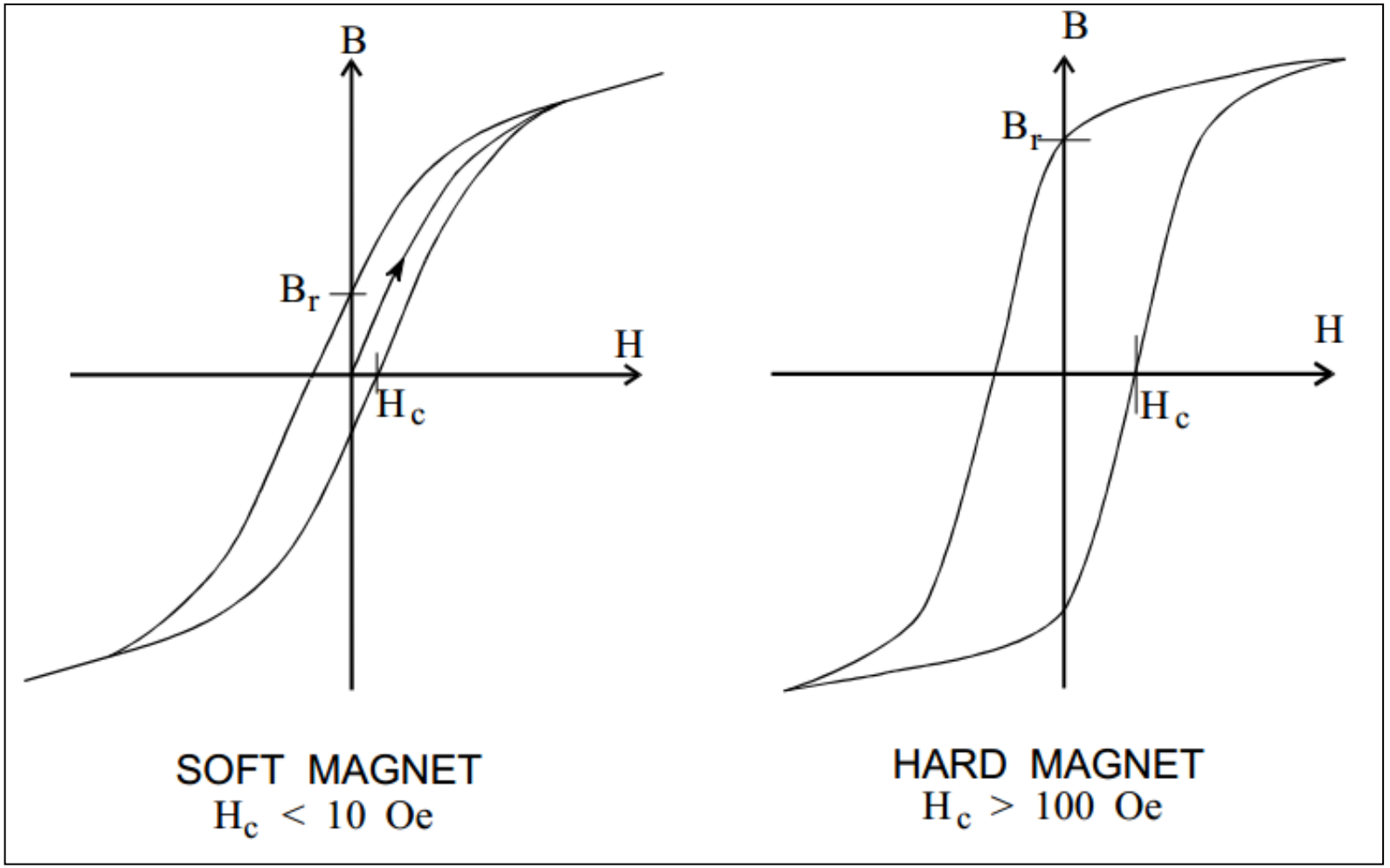
자성체 (Magnetic Substance)

- **자기 이력 곡선 (Hysteresis Loop), B-H 곡선**
  - 디스크의 경우, 전압 제한, 타이밍, 쓰기 밀도 등으로 인해 포화 상태에 이르지 못함
  - 온도, 덮어쓰기 같은 요인으로 인해 각 쓰기 작업 마다 B-H 곡선이 다름



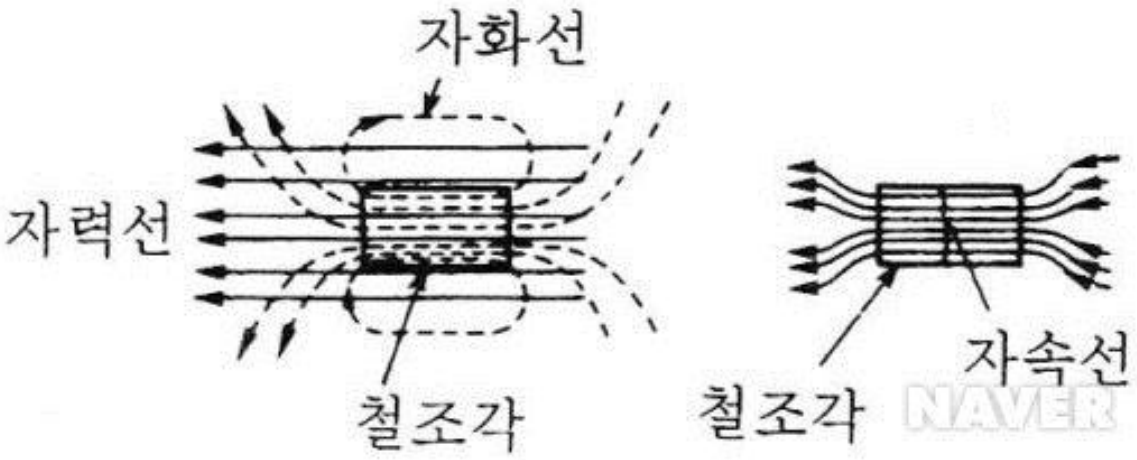
자성체 (Magnetic Substance)

- 자기 이력 곡선 (Hysteresis Loop), B-H 곡선



자성체 (Magnetic Substance)

- **자화(magnetization) vs. 자속(magnetic flux)**
  - **자화**
    - ✓ 자기장 내에 자성체를 놓았을 때, 해당 자성체도 자기를 띄게 되는 현상
  - **자속**
    - ✓ 자화된 자성체를 통과하는 자속선을 하나로 합한 총 자속선 수



데이터 기록 (Data Recording)

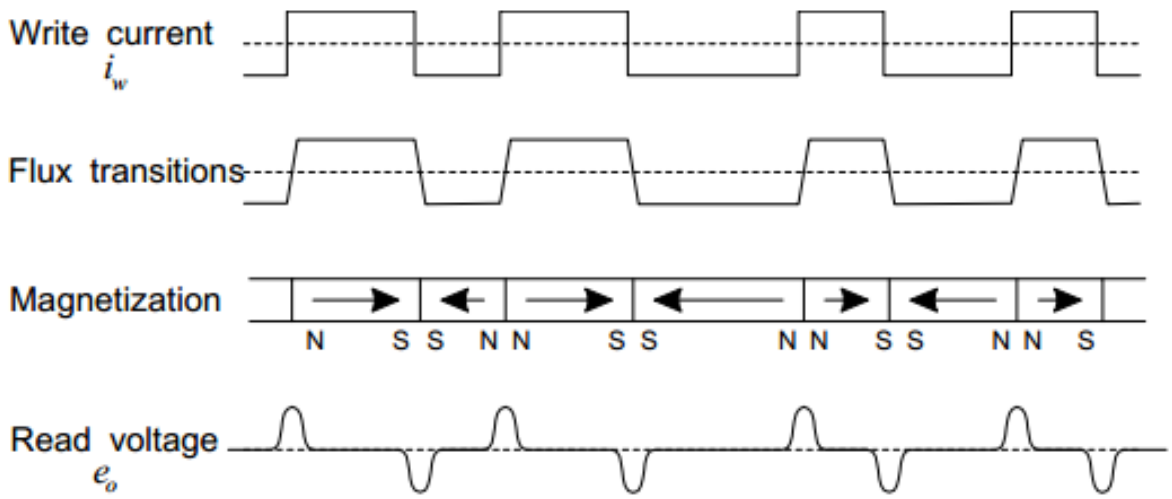
- **0과 1 표현 방법 ➔** 자기는 두 극성이 존재 (North, South)
  - **자극(magnetic pole) 측정을 이용한 방식**
    - ✓ N 극(또는 S 극) : 0
    - ✓ S 극(또는 N 극) : 1
  - **자화 반전(flux reversal)을 이용한 방식**
    - ✓ S → N(또는 N → S) : 0
    - ✓ N → S(또는 S → N) : 1
  - **절대적인 극성이 아니라 자화 반전을 측정하는 방식으로 발전 ➔ 왜???????**

데이터 기록 (Data Recording)

- **자화 반전을 사용하는 이유**
  - **필드(Field) vs. 반전(Reversal)**
    - ✓ 각 필드 값(극성)보다 반전을 측정하기가 더 쉬움
    - ✓ 반전을 자기저항 소자로 감지 ➔ 전압으로 변환 ➔ 헤드를 이용해 탐지
    - ✓ 디스크 밀도가 증가함에 따라 임계 감도를 탐지할 수 있는 자기장 세기는 계속 감소
  - **동기화(Synchronization)**
    - ✓ 각 비트의 시작과 끝이 어디인가?
    - ✓ 연속된 0을 1,000개 정도 기록할 경우 ➔ SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS.....
  - **필드 구분(Field Separation)**
    - ✓ 동일한 데이터(연속된 0혹은 1)가 기록될 경우, 하나의 큰 자기장 생성
    - ✓ 각 필드의 구분이 필요 ➔ 추가적인 필드 구분자가 필요함

데이터 기록 (Data Recording)

- **데이터 기록에 따른 데이터, 자속 전이, 자화 흐름, 읽기 전압**

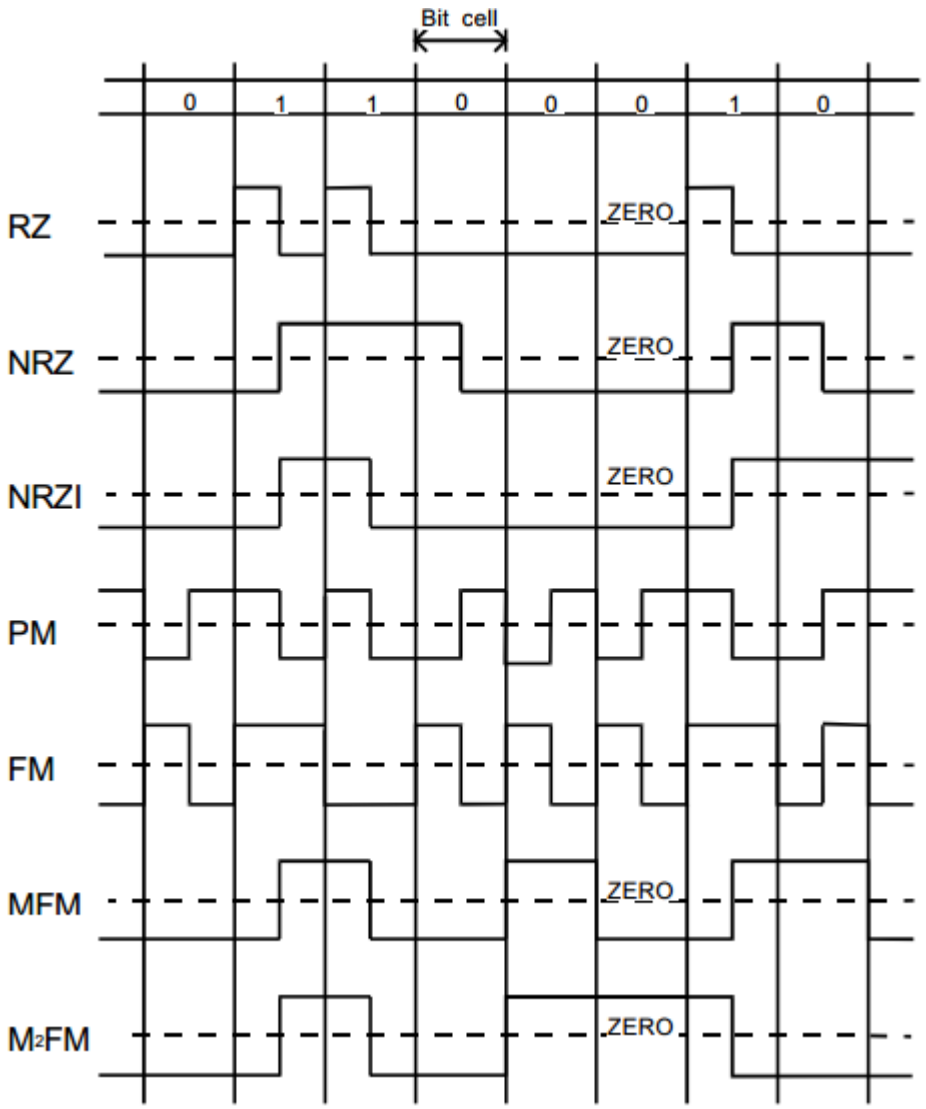


- 자성체에 외부 자계를 가하면 특정 방향으로 자화
- 자화 반전의 역할은 데이터를 구분하여 클록 동기화(clock synchronization)
- 클록 동기화에 사용되는 자화 반전을 어떻게 줄이냐? ➔ 인코딩 기법의 발달

Data Encoding Techniques

인코딩 기법 (Encoding Techniques)

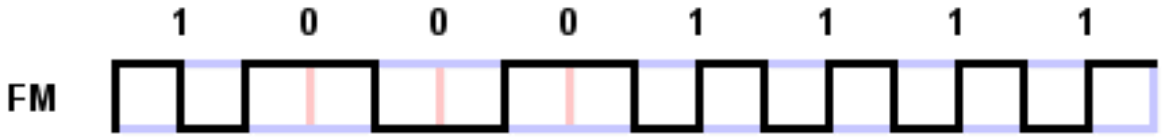
- 다양한 기법
  - **RZ** (Return to Zero)
  - **NRZ** (Non-Return to Zero)
  - **NRZI** (Non-Return to Zero, Invert)
  - **PM** (Phase Modulation)
  - **FM** (Frequency Modulation)
  - **MFM** (Modified FM)
  - **MMFM** (Modified MFM)



http://www.lintech.org/comp-per/07MAGREC.pdf

FM (Frequency Modulation)

- **주파수 변조** (0에 비해 1이 두 배의 반전 수를 가짐)
  - **0** : 자화 반전 다음에 자화 반전 X
  - **1** : 연속된 2개의 자화 반전
  - 1970년 말~1980년 초에 플로피 디스크에 사용 ➔ 후에 MFM으로 변경
  - 클록을 위해 추가된 자화 반전으로 인해 낭비가 매우 심함

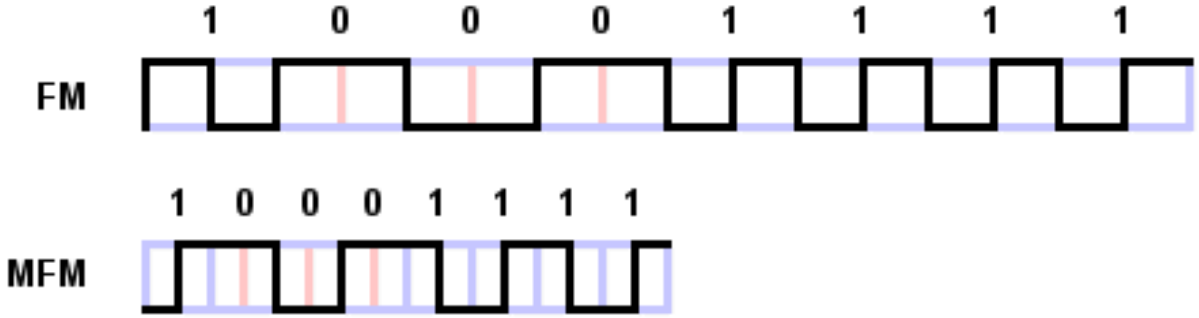


Bit Pattern	Encoding Pattern	Flux Reversals Per Bit	Bit Pattern Commonality In Random Bit Stream
0	RN	1	50%
1	RR	2	50%
Weighted Average		1.5	100%

* **N** : No flus reversal, **R** : flux Reversal

MFM (Modified Frequency Modulation)

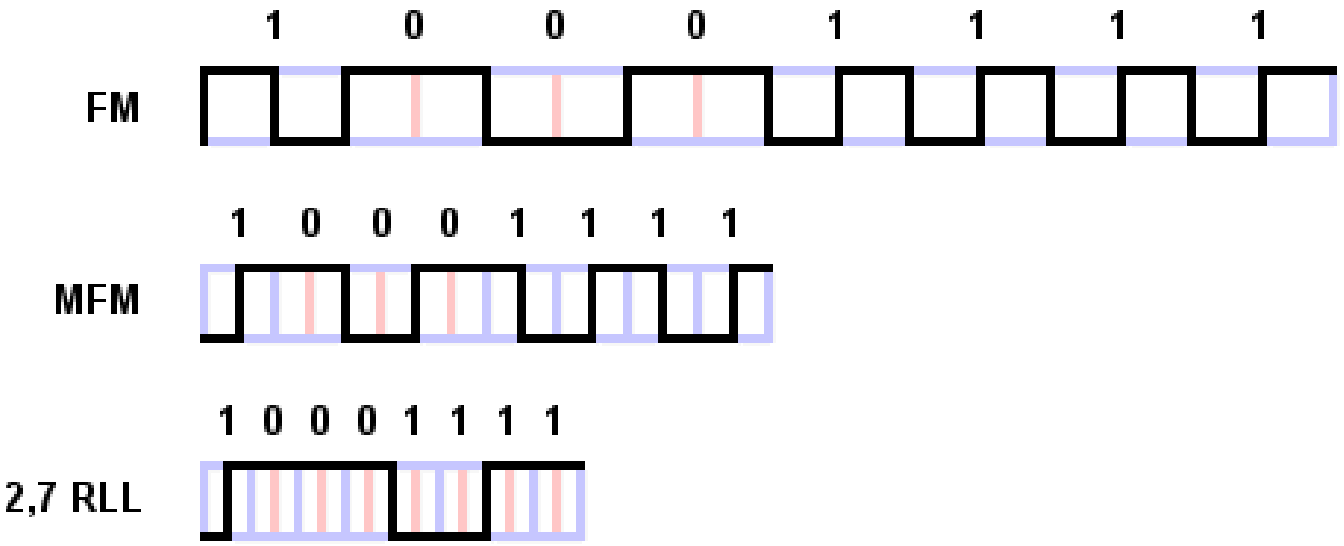
- **변형 주파수 변조**
  - 클록에 사용되는 자화 반전 수를 줄여 FM을 향상 ➔ 연속된 0일 때만 자화 반전 사용
  - 알고리즘, 인코딩/디코딩 회로의 복잡 ➔ 어차피 컨트롤러가 하는 일이라
  - 플로피 디스크와 초기 하드디스크에 사용 ➔ 현재까지 플로피 디스크의 표준



Bit Pattern	Encoding Pattern	Flux Reversals Per Bit	Bit Pattern Commonality In Random Bit Stream
0 (preceded by 0)	RN	1	25%
0 (preceded by 1)	NN	0	25%
1	NR	1	50%
Weighted Average		0.75	100%

RLL (Run Length Limited) (cont'd)

- **런 길이 제한**
  - 하나의 비트를 인코딩 하는 것이 아닌 몇 비트를 묶어 패턴으로 기록 ➔ 클록/자화 반전 혼합
  - **런 길이(Run Length)** : 자화 반전 사이의 최소 간격
  - **런 제한(Run Limited)** : 자화 반전 사이의 최대 간격
  - 다양한 변형 존재 ➔ RLL (1,7), RLL (2,7)



RLL (Run Length Limited)

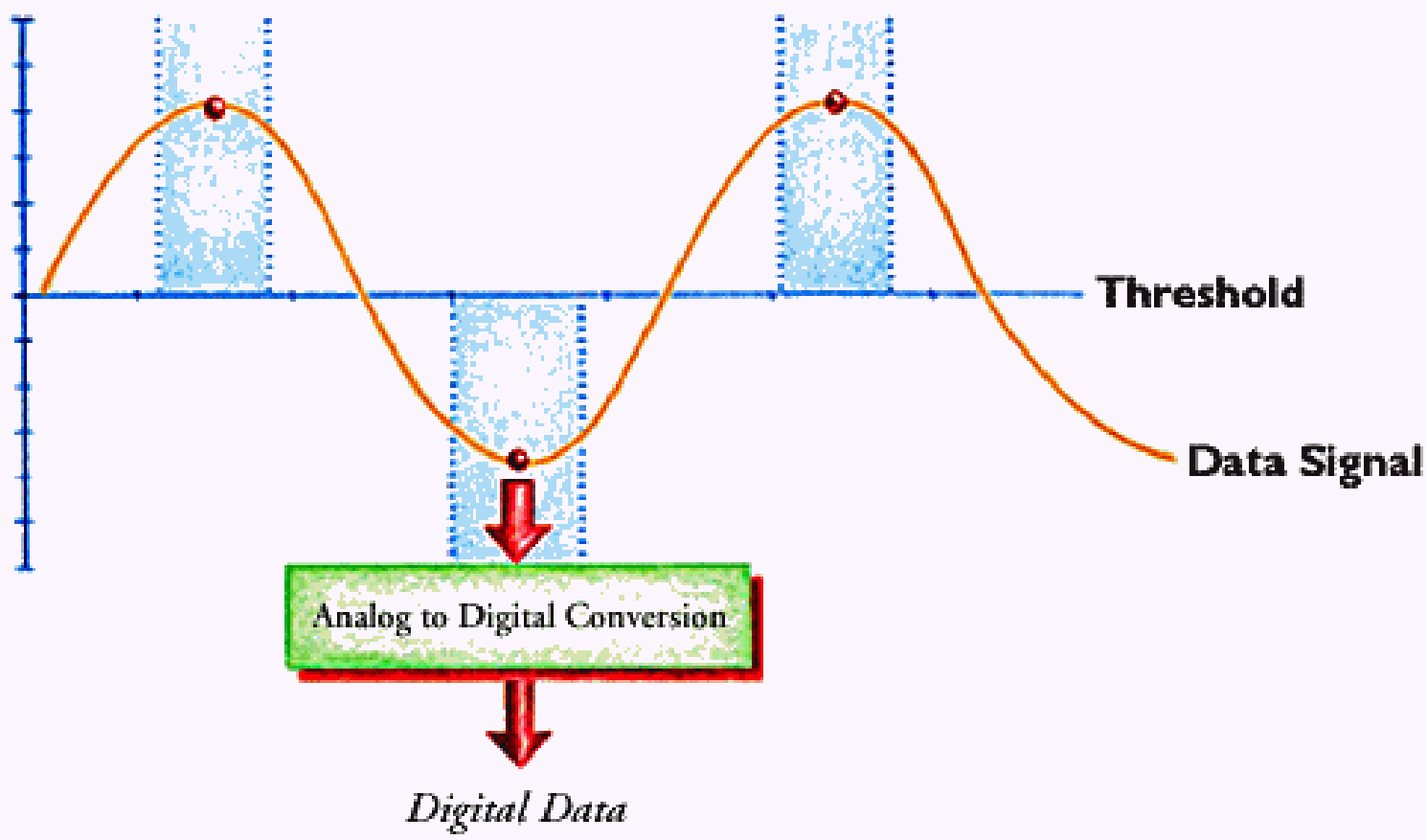
- **RLL (2, 7)**

Bit Pattern	Encoding Pattern	Flux Reversals Per Bit	Bit Pattern Commonality In Random Bit Stream
11	RNNN	1/2	25%
10	NRNN	1/2	25%
011	NNRNNN	1/3	12.5%
010	RNNRNN	2/3	12.5%
000	NNNRNN	1/3	12.5%
0010	NNRNNRNN	2/4	6.25%
0011	NNNNRNNN	1/3	6.25%
Weighted Average		0.4635	100%

- RLL 등장으로 하드디스크 시장은 MFM에서 RLLL로 교체
- 플로피 디스크는 여전히 MFM 사용

PRML (Partial Response, Maximum Likelihood) (cont'd)

- **최대치 검출 (Peak Detection)**
  - RLL 인코딩은 최대치 검출 방식

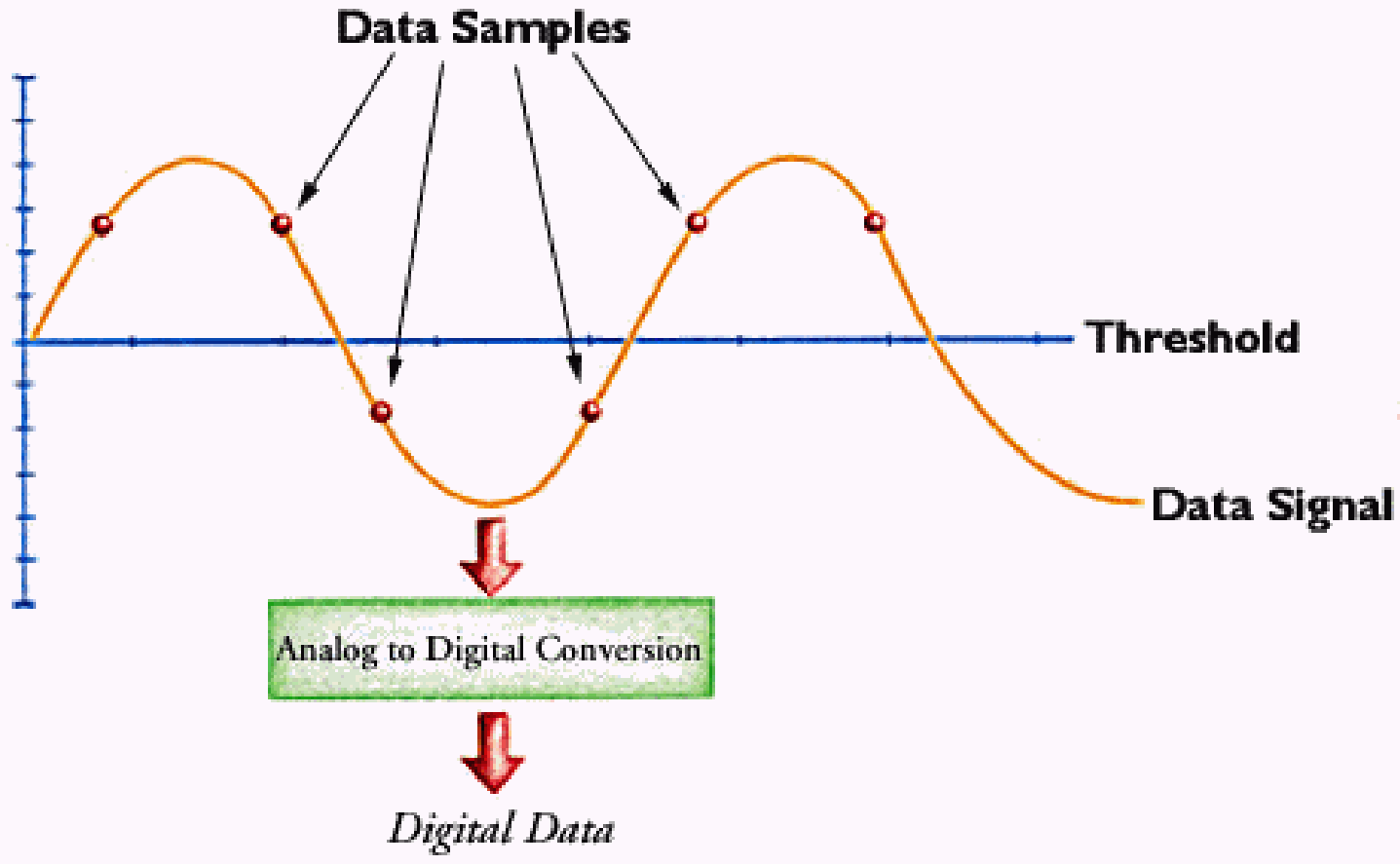


PRML (Partial Response, Maximum Likelihood) (cont'd)

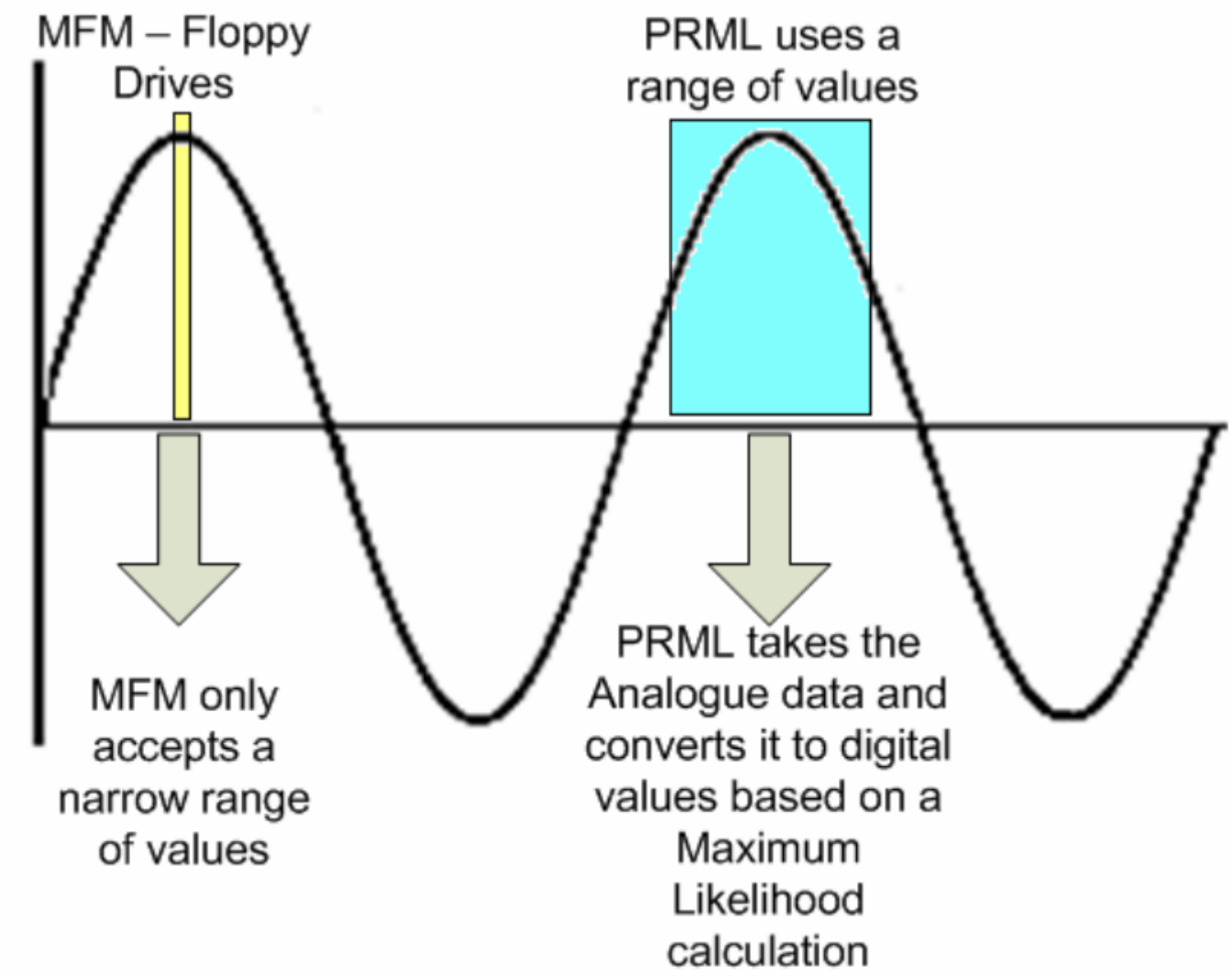
- **최대치 검출의 한계**
  - 배경 잡음보다 최대치가 충분히 클 경우 잘 동작
  - 기록 밀도의 증가로 자화 반전의 최대치는 서로 더 밀접하게 위치 ➔ 간섭 발생
  - 간섭을 줄이고자 자기장 세기를 줄임 ➔ 최대치 검출의 어려움
  - 어려움을 해결하고자 등장한 것이 PRML (부분 응답, 최대 유사)
  - PRML은 RLL에 비해 30~40%의 기록밀도를 증가시킴

PRML (Partial Response, Maximum Likelihood)

- **부분 응답, 최대 유사**
  - 데이터 샘플링 (부분 응답) ➔ 가장 큰 시퀀스 탐지 (최대 유사)



MFM vs. PRML


MFM – Floppy Drives
PRML uses a range of values
MFM only accepts a narrow range of values
PRML takes the Analogue data and converts it to digital values based on a Maximum Likelihood calculation

EPRML (Extended PRML)

- **확장 PRML**
  - 더 효과적인 알고리즘과 처리 회로를 사용 ➔ 더 정확하게 해석
  - 오류 발생률 증가 없이 PRML에 비해 약 20%~70%까지 기록 밀도 증가
  - 현대의 하드디스크는 모두 EPRML 인코딩 방식을 사용

Data Wiping Techniques

완전삭제 방식 (Wiping/Sanitization/Secure Erase/Destruction Methods)

- **덮어쓰기 (Overwriting)**
  - 삭제하고자 하는 데이터의 위치에 [0, 1, 임의 패턴, 랜덤 데이터]를 덮어씀
- **암호화 (Encryption)**
  - 안전한 방식을 이용해 디스크 혹은 파일 암호화
- **디가우징(Degaussing)**
  - 강력한 자기장을 이용해 자기디스크의 표면의 자력 흐름을 파괴
- **물리적 파괴, 천공, 파쇄(Physical Destruction)**
  - 강력한 기계를 사용해 물리적으로 구멍을 내거나 파쇄

디가우징 vs. 물리파괴

덮어쓰기 기반의 와이핑 표준 (Standard Overwritten Techniques)

- **http://en.wikipedia.org/wiki/Data_erasure**

표준	년도	반복	패턴	비고
U.S. Navy	1993	3	문자, 보수, 랜덤	검증 필수
U.S. Air Force	1996	4	0, 1, 문자	검증 필수
Peter Gutmann	1996	1-35	매우 다양	원래 현재는 사용되지 않는 MFM, RLL을 위해
Bruce Schneier	1996	7	0, 1, 5번의 유사 랜덤	
U.S. DoD	2001	3	문자, 보수, 다른 패턴	
German Federal	2004	2-3	불규칙 패턴, 보수	
CSEC	2006	3	0(1), 보수	분류되지 않은 매체를 위해
NIST	2006	1	?	
U.S. NISP	2006	?	?	더 이상 지정하지 않음
NSA/CSS	2007	0	?	디가우즈 또는 파괴
Australian	2008	1	?	디가우즈 또는 일급 비밀 매체 파괴
New Zealand	2008	1	?	기밀 데이터를 위해

덮어쓰기 방식 (Overwritten Methods)

- **디스크 영역 와이핑**
  - 물리 섹터 시작~마지막 까지 덮어쓰기
  - 비할당 영역, 슬랙 공간 와이핑
  - HPA, DCO 고려
- **파일 단위 와이핑**
  - 파일 삭제 시 파일 데이터를 안전하게 와이핑
  - 파일 메타데이터와 관련 아티팩트(프리패치, 레지스트리 등)도 고려
- **표준 와이핑 기법을 지원하는 자동화 도구 사용**

와이핑 도구 (Wiping Tools)

- **BCWipe** (file, folder, free space, windows artifacts, file slack)
- **Hardwipe** (file, drive, space)
- **Eraser** (file, folder, free space, slack space)
- **CCleaner** (windows artifacts)
- **File Shredder** (file, folder)
- **SDelete** (file, folder, free space)
- **Darik's Boot And Nuke** (drive)
- **dd** (*nix) (file, drive)
- … …

Probability of Physical Data Recovery

피터 구트만 논문 (Peter Gutmann's Paper)

- **Secure Deletion of Data from Magnetic and Solid-State Memory (1996)**
  - MFM(Magnetic Force Microscopy)로 복구 가능성 언급
  - 0에다 1을 덮어쓰면 0.95, 1에다 1을 덮어쓰면 1.05 형태의 값이 기록 ➔ 이전 데이터 유추 가능
  - 기록할 때 쓰기 헤더 위치의 오차 발생 ➔ 트랙 가장자리의 잔여 데이터 남김
  - 최대 35번의 덮어쓰기가 필요 ➔ MFM, (1,7) RLL, (2,7) RLL 모두 대상
  - PRML 방식을 사용하는 디스크는 랜덤 데이터를 몇 번만 덮어써도 충분

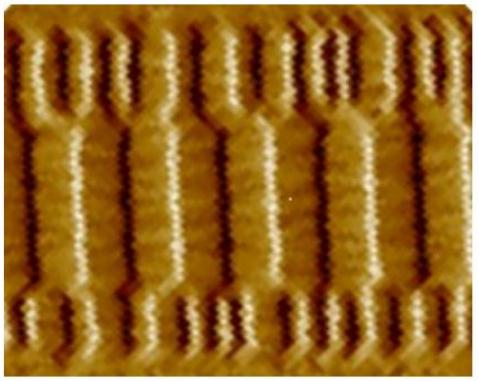
피터 구트만 덮어쓰기 데이터 (Peter Gutmann's Overwrite Data)

Overwrite Data				
Pass No.	Data Written	Encoding Scheme Targeted		
1	Random			
2	Random			
3	Random			
4	Random			
5	01010101 01010101 01010101 0x55	(1,7) RLL		MFM
6	10101010 10101010 10101010 0xAA	(1,7) RLL		MFM
7	10010010 01001001 00100100 0x92 0x49 0x24		(2,7) RLL	MFM
8	01001001 00100100 10010010 0x49 0x24 0x92		(2,7) RLL	MFM
9	00100100 10010010 01001001 0x24 0x92 0x49		(2,7) RLL	MFM
10	00000000 00000000 00000000 0x00	(1,7) RLL	(2,7) RLL	
11	00010001 00010001 00010001 0x11	(1,7) RLL		
12	00100010 00100010 00100010 0x22	(1,7) RLL		
13	00110011 00110011 00110011 0x33	(1,7) RLL	(2,7) RLL	
14	01000100 01000100 01000100 0x44	(1,7) RLL		
15	01010101 01010101 01010101 0x55	(1,7) RLL		MFM
16	01100110 01100110 01100110 0x66	(1,7) RLL	(2,7) RLL	
17	01110111 01110111 01110111 0x77	(1,7) RLL		
18	10001000 10001000 10001000 0x88	(1,7) RLL		
19	10011001 10011001 10011001 0x99	(1,7) RLL	(2,7) RLL	
20	10101010 10101010 10101010 0xAA	(1,7) RLL		MFM
21	10111011 10111011 10111011 0xBB	(1,7) RLL		
22	11001100 11001100 11001100 0xCC	(1,7) RLL	(2,7) RLL	
23	11011101 11011101 11011101 0xDD	(1,7) RLL		
24	11101110 11101110 11101110 0xEE	(1,7) RLL		
25	11111111 11111111 11111111 0xFF	(1,7) RLL	(2,7) RLL	
26	10010010 01001001 00100100 0x92 0x49 0x24		(2,7) RLL	MFM
27	01001001 00100100 10010010 0x49 0x24 0x92		(2,7) RLL	MFM
28	00100100 10010010 01001001 0x24 0x92 0x49		(2,7) RLL	MFM
29	01101101 10110110 11011011 0x6D 0xB6 0xDB		(2,7) RLL	
30	10110110 11011011 01101101 0xB6 0xDB 0x6D		(2,7) RLL	
31	11011011 01101101 10110110 0xDB 0x6D 0xB6		(2,7) RLL	
32	Random			
33	Random			
34	Random			
35	Random			

또 다른 논문 (Another Paper)

- **Overwriting Hard Drive Data: The Great Wiping Controversy (2008)**
  - 복구가 불가능하게 덮어쓰려면 몇 번 시도를 해야 하는지에 대한 논란이 많음
  - 피터 구트만 논문 이후 데이터를 한번만 덮어쓰면 복구할 수 있다는 논란이 커짐
  - 다수의 와이핑은 너무 많은 시간이 요구됨
  - 이런 논란을 검증하고자 테스트 해봄 ➔ 포렌식적으로 의미가 있는지?
  - **잘못된 오해**
    - ✓ 1을 기록할 때, 0을 덮어쓰면 "0.95", 1을 덮어쓰면 "1.05"에 근접한다는 추정이 맞는가?
    - ✓ 각 쓰기 작업 시 정확한 "1"의 값을 쓰는 것이 가능한가?

MFM (Magnetic Force Microscopy)



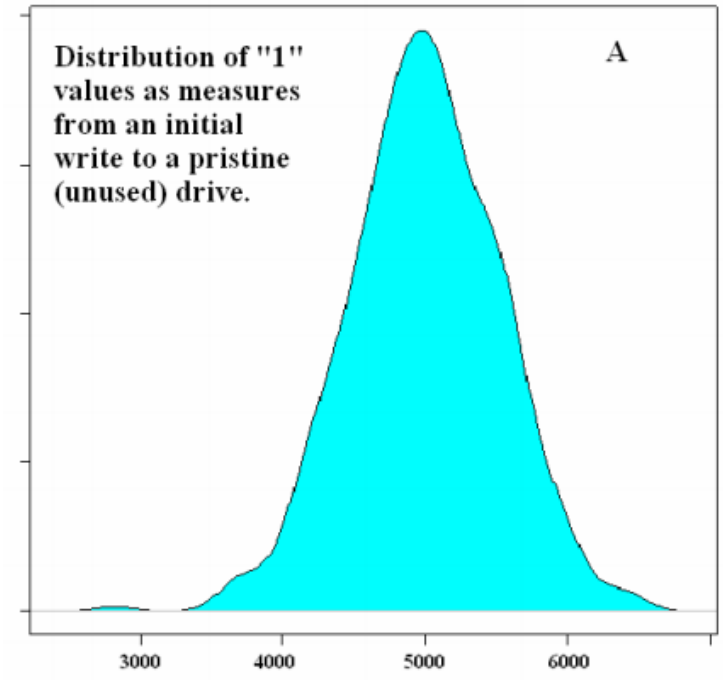
- **MFM 기능**
  - 자기력에 의한 샘플 표면의 공간적 변이를 형상화
  - **1 바이트**를 복구하는데 **4분** 소요
  - 형상화로 알 수 있는 정보
    - ✓ 트랙의 너비와 왜곡
    - ✓ 전이 이상
    - ✓ 디스크 상의 기록 영역과 덮어쓴 영역의 차이

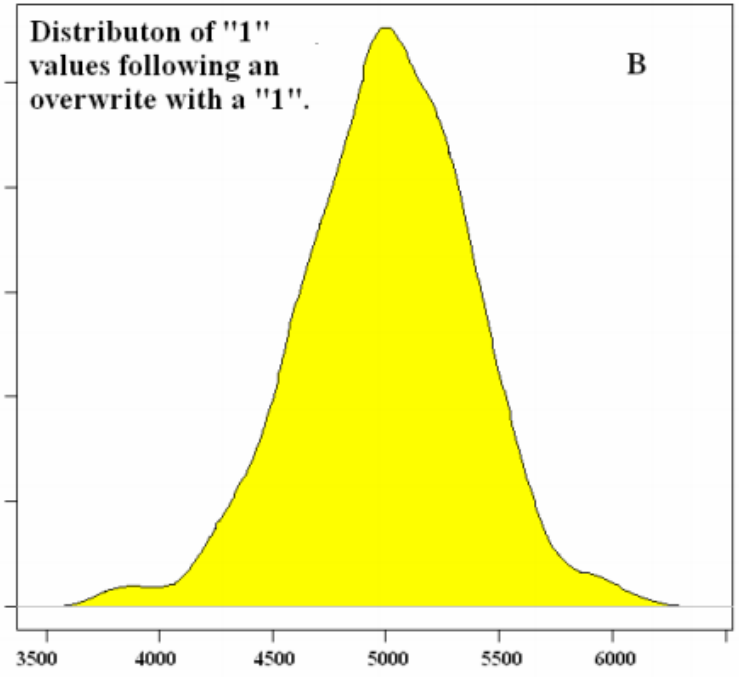
MFM (Magnetic Force Microscopy)

- **형상화 패턴의 변이**
  - **기록 패턴에 영향을 미치는 요인**
    - ✓ 헤드의 움직임 ➔ 내부적으로 진동, 공기 흐름 발생 ➔ 항상 정확한 곳에 위치?
    - ✓ 온도 ➔ 디스크 회전에 따른 내부 열로 자기력 변동 ➔ 플래터 온도가 올라가면 자기력이 감소됨
    - ✓ 랜덤 에러 ➔ 수많은 컴포넌트의 동작 중 발생하는 오류
    - ✓ 기록된 이전 데이터
  - **보안책**
    - ✓ 열 보정(Thermal Recalibration) 알고리즘을 사용하여 디스크 수축/팽창에 따른 속도 차이 개선
    - ✓ 알고리즘 개선, 내부 부품 개선, 인코딩 스키마 개선
  - 다양한 노력에도 이력 현상과 많은 외부 요인에 의해 동일한 패턴을 기록하기는 어려움

밀도 분포 (Distribution of Density)

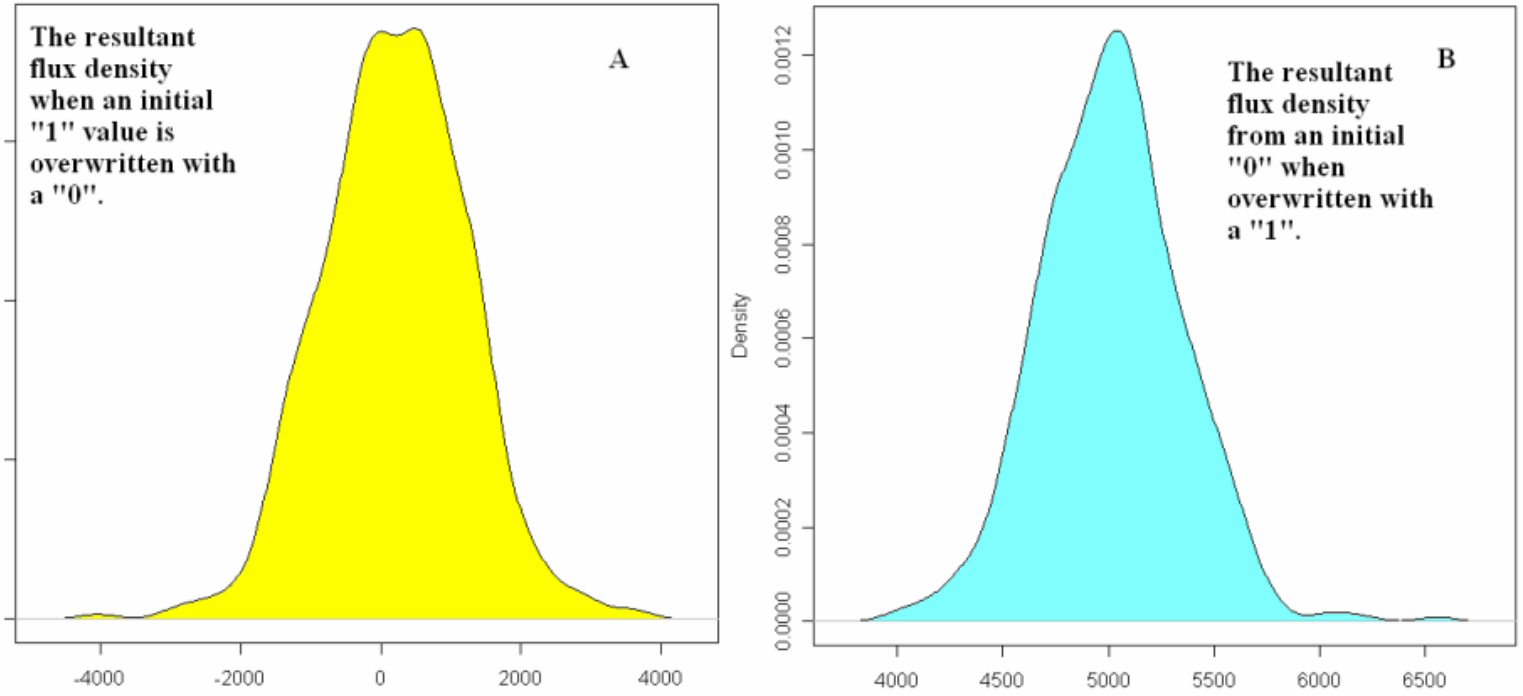
- **밀도 분포의 차이**
  - 디스크에 "1"을 기록하는 매 기록 작업마다 정확히 "1" 값이 기록될 수 있는가?
  - 온도 변동, 습도, 진동, 시간에 따른 부식으로 인해 와이핑 단계마다 밀도 분포는 다름
  - 미세한 차이지만 값을 예측하기는 어려움





밀도 분포 (Distribution of Density)

- **밀도 분포의 차이**
  - "1.06"이라는 값을 얻었을 때, 이전 값의 영향인가? 온도의 영향인가? ➔ 판단 어려움
  - 시간이 지남에 따라 드라이브 자기장은 부식 (약해짐?)



밀도 분포 (Distribution of Density)

- **복구 가능성**
  - **깨끗한 디스크 또는 와이핑된 디스크에서 데이터를 복구할 수 있는 가능성은 낮음**
    - ✓ 0.01% 보다 더 작을 수 있음
  - **사용하고 있는 디스크(포맷된 디스크 포함)에서 복구할 수 있는 가능성도 낮음**
    - ✓ 가능성은 좀 더 높지만 의미있는 단어를 얻어낼 가능성은 적음
  - **전자 현미경 기술이 더 발전한다면?**
    - ✓ 전자 현미경의 에러는 하드디스크 에러보다 더 적음
    - ✓ 전자 현미경의 기술적 한계가 아닌 하드디스크의 기술적 한계

에러 관리 로직 (Read Error Severities and Error Management Logic)

- **디스크 컨트롤러의 에러 최소화 기법**
  - **ECC Error Detection :** 섹터 Servo 영역에 저장된 ECC 활용
  - **ECC Error Correction :** ECC 에러 탐지 시 오류 정정 방식으로 복구
  - **Automatic Retry :** 갑작스런 움직임이나 온도 변화로 정확한 위치를 못 찾을 때, 정정 후 재시도
  - **Advanced Error Correction :** 고급 에러 정정 알고리즘 사용 ➔ 속도 느림
  - **Failure :** 섹터를 읽을 없는 경우, 에로 복구가 불가능

- 제조사마다 공통적인 에러 최소화 기법 사용 ➔ 에러의 영향은 줄어듬
- 현재 인코딩 스키마(PRML, EPRML)에서는 아날로그 값의 허용이 넓음 ➔ 굳이 보안책을
- 결과적으로, 이전 값을 알아내는 것은 확률 게임

실험 데이터와 방법 (Data and Method) (cont'd)

- **카테고리 A**
  - 사용되지 않은 깨끗한 디스크
  - 포맷된 디스크 (NTFS의 기본 섹터 크기를 이용해 한번 포맷)
  - 시뮬레이션 드라이브 (랜덤한 데이터를 32번 덮어쓰기, /dev/random) ➔ 0으로 덮어씀

- **카테고리 B**
  - 최초 기록과 연속된 덮어쓰기에 5가지 패턴 사용
    - ✓ 모두 0
    - ✓ 모두 1
    - ✓ 01010101 패턴
    - ✓ 00111011 패턴
    - ✓ 00001111 패턴

실험 데이터와 방법 (Data and Method)

- **17개의 선정된 모델 사용**
  - 오래된 Quantum 1GB에서 2006년 출시된 모델까지 (SCSI, IDE 등)
  - 56개의 하드디스크 테스트

- **실험**
  1. 1KB 파일을 이용해 데이터 기록
  2. 드라이브 왜곡과 비트는 모두 읽음
  3. 76,800 데이터 포인트 분석을 위해 각 절차를 5번 반복함
     - ✓ 사전 분포를 이용해 베이즈 정리(Bayes' Theorem)를 사용
     - ✓ 실제 포렌식 업무에서는 사전 데이터를 알 수 없음

복구 가능성 (Probability of Recovery)

- **오래된 드라이브 모델에 대한 확률 분포 테이블**
  - 초기 1을 기록한 후, 0으로 덮어씀 (이상적인 상황)

Probability of recovery	Pristine drive	Used Drive (ideal)
1 bit	0.92	0.56
2 bit	0.8464	0.3136
4 bit	0.71639296	0.098345
8 bits[5]	0.51321887	0.009672
16 bits	0.26339361	9.35E-05
32 bits	**0.06937619**	**8.75E-09**
64 bits	0.00481306	7.66E-17
128 bits	2.3166E-05	5.86E-33
256 bits	5.3664E-10	3.44E-65
512 bits	2.8798E-19	1.2E-129
1024 bits	8.2934E-38	1.4E-258

복구 가능성 (Probability of Recovery)

- **새로운 드라이브 모델에 대한 확률 분포 테이블**
  - 초기 1을 기록한 후, 0으로 덮어씀 (이상적인 상황)
  - 추가 덮어쓰기를 1회, 3회한 후 결과 비교

Probability of recovery	Pristine drive (plus 1 wipe)	Pristine drive (plus 3 wipe)
1 bit	0.87	0.64
2 bit	0.7569	0.4096
4 bit	0.57289761	0.16777216
8 bits	0.328211672	0.028147498
16 bits	0.107722901	0.000792282
32 bits	**0.011604223**	**6.2771E-07**
64 bits	0.000134658	3.9402E-13
128 bits	1.81328E-08	1.55252E-25
256 bits	3.28798E-16	2.41031E-50
512 bits	1.08108E-31	5.8096E-100
1024 bits	1.16873E-62	3.3752E-199

복구 가능성 (Probability of Recovery)

- **새로운 드라이브 모델에 대한 확률 분포 테이블**
  - 고밀도의 EPRML 사용 드라이브의 복구 확률은 **어림 짐작 확률과 유사**
  - **2006년 모델 테스트**
    - ✓ 모두 0으로 와이핑된 디스크에 1을 덮어썼을 때 ➔ 최대 49.18%(+/- 0.11) 복구 확률
    - ✓ 모두 0으로 와이핑된 디스크에 다른 패턴 ➔ 최대 36.08%(+/- 0.24) 복구 확률
    - ✓ 일반적으로 사용하던 디스크의 복구 확률은?

복구 가능성 (Probability of Recovery)

▪ 복구 데이터 분포

- 8 비트를 정확히 읽었을 때, "1"로 표시

```
  [1]   0 0 1 0 1 0 0 1 0 0 1 1 1 0 1 0 1 1 1 1 1 0 0 1 0 0 1 0 1 0 1 1 1 0 1 0 0 0 0 0 0 0 1 0 1 1 1
 [48]  0 1 0 1 0 0 0 1 0 1 1 1 1 1 0 0 0 0 0 0 1 0 0 0 0 1 1 0 1 1 1 0 0 0 1 0 1 0 1 1 1 1 0 1 0 0 0
 [95]  0 1 1 0 0 1 0 0 1 1 0 1 0 0 1 0 1 0 0 0 1 0 1 1 1 0 0 1 0 0 1 1 0 0 0 0 1 0 1 0 1 0 1 0 0 1 1
[142]  1 1 1 1 1 1 1 1 0 0 1 1 1 1 0 0 0 0 0 1 0 0 1 0 0 0 1 0 0 1 0 1 1 0 1 1 1 0 0 1 1 0 0 0 1 0 0
[189]  1 0 1 1 0 1 0 1 0 0 1 1 1 0 1 0 1 1 0 1 0 1 1 1 1 0 1 1 0 0 1 0 0 1 0 1 0 0 1 1 0 0 0 0 1 0 1
[236]  0 0 1 0 1 1 1 0 1 1 0 0 1 1 1 1 1 1 1 0 0 1 1 1 0 0 1 0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 1 0 1 0 1
[283]  0 1 0 1 0 1 1 0 0 0 1 1 0 0 0 1 0 0 1 0 0 1 1 1 0 0 1 1 1 1 1 1 0 0 1 1 0 0 1 1 1 1 1 0 0 0 0
[330]  0 0 0 0 0 0 1 1 0 1 0 1 1 1 0 1 0 1 1 0 0 0 1 1 0 0 0 1 1 0 1 0 1 0 1 0 0 0 0 1 1 0 0 1 0 1 0
[377]  1 1 1 0 1 0 1 0 0 1 0 0 0 0 0 0 0 1 0 0 0 1 0 0 1 1 1 1 1 0 1 1 0 1 1 0 1 0 1 1 1 0 1 0 0 0 1
[424]  1 1 0 0 1 1 1 0 0 0 0 1 1 1 0 1 0 0 0 0 1 0 0 1 0 0 0 1 1 0 0 0 1 0 0 1 1 1 1 0 1 1 1 0 0 0 1
[471]  1 1 1 1 0 1 1 1 1 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 1 0 1 1 0 0 1 0 1 1 1 1 1 1 1 0 1 0 1 0 0 0 0
[518]  1 0 0 1 0 1 1 1 1 1 0 0 1 0 1 1 0 1 0 1 1 1 1 1 0 0 1 0 0 0 1 0 0 0 0 1 1 0 0 0 0 1 1 0 0 0 0
[565]  1 1 1 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 1 1 1 1 0 0 0 0 0 1 0 1 1 1 1 1 1 1 0 1 1 1 1 0 0 1 0 0 0
[612]  1 1 0 1 1 1 1 0 1 1 1 1 0 0 0 1 1 0 0 0 1 1 0 0 1 0 1 1 0 0 1 0 0 0 0 1 0 1 0 1 0 0 0 0 1 0 0
[659]  1 1 0 1 0 1 0 1 1 0 0 0 0 0 0 1 1 0 1 0 0 1 1 1 0 0 0 1 1 1 0 1 1 0 0 0 0 0 1 1 1 1 1 0 0 1 1
[706]  1 0 0 1 1 1 0 1 0 1 0 1 1 0 1 0 0 0 0 0 1 1 0 1 1 1 0 1 0 1 0 1 0 0 1 1 0 0 0 0 0 0 0 1
[753]  1 0 0 1 1 1 0 0 1 0 1 1 0 0 0 0 1 1 1 1 1 0 0 0 1 0 0 1 1 1 0 0 0 0 1 0 1 0 0 0 1 0 1 1 0 0 0
[800]  0 0 1 0 0 0 1 0 0 0 0 0 0 1 0 1 1 1 1 0 1 0 1 0 0 0 0 1 0 1 1 1 1 1 0 0 1 0 0 1 0 0 0 0 0 0 0
[847]  0 0 0 0 1 0 1 0 0 0 0 0 0 1 0 1 1 0 1 1 1 0 1 1 0 0 0 1 1 1 1 1 1 0 0 1 0 0 0 0 1 1 1 0 0 1 0
[894]  1 1 1 0 1 1 0 0 1 0 0 0 0 1 1 0 1 1 1 1 0 1 1 0 0 1 0 0 1 0 0 1 1 1 0 1 1 0 0 1 0 0 1 0 0 1 1
[941]  1 0 0 0 1 0 1 0 1 0 1 1 1 1 1 1 0 1 1 0 0 0 1 0 0 1 1 1 1 0 0 1 0 1 1 1 1 1 1 0 0 1 1 1 0 1 0
[988]  0 0 1 0 1 1 1 0 1 0 0 1 1 1 0 0 1 1 0 0 1 0 0 0 0 1 1 1 1 1 0 1 0 0 1 1 0
```

복구 가능성 (Probability of Recovery)

- **예제 테스트**
  - **테스트 데이터**

> *Secure deletion of data - Peter Gutmann - 1996*
> *Abstract*
> *With the use of increasingly sophisticated encryption systems, an attacker wishing to gain access to sensitive data is forced to look elsewhere for information. One avenue of attack is the recovery of supposedly erased data from magnetic media or random-access memory.*

  - **복구 결과 (최선의 방법)**

> *‰*
> *‘cKræ}d8Œeti²n•of0daÊI0Ptr0G§tWÇíï_¼Á1u960eb8tÈñutW00000Dç•Ã#Ì0*
> *Hf$00¦000%£z0\0ã0000á0áä«it|tþÛ0u³e•Ffºi™%|eàsinqTyøîopÚ”Ë:i†aze0*
> *®Mcryption0sîÙtems?DKtA""cÐÏ0+¢sinŒ0toK–ai2z ÷c(ns~0tü0;e*
> *½iti)e""daÆa>s0foôce¸ÑtÒÍl2o–*
> *ìeII¶˜$eöe›Ÿr""inf¬rm‰ion.0OnRïavem>egoN0- ¨tRÁ"1i*
> *läßh ±0"eÛoie=y0Czsu• ¨s/`lÜ{era’Jd0dataF ¨ro>•magne³;&£õãÈáã%or*r‰ndoª-Qcc«ÇŸ0mà*
> *@ryl000000000000000000*

복구 가능성 (Probability of Recovery)

- **결과적으로…**
  - 개별 비트의 복구 가능성은 있지만 **의미 있는 데이터는 복구 불가능**
  - 디스크 제조사나 디스크 상태에 따라 차이 발생 ➔ **표준화하기는 어려움**
  - GB/TB의 정보를 자동으로 검색하는 **도구를 개발하는 것은 불가능**
  - 오래된 드라이브에서 조차도 **가능성이 낮음**
  - 법정에서 **증거로 활용하기에는 부적합**
  - 덮어쓴 드라이브에서의 복구 가능성에 대한 **논란은 종식시킬 필요가 있음**

- **참고 자료 (Reference)**

1. Peter Gutmann, **Secure Deletion of Data from Magnetic and Solid-State Memory**
2. Craig Wright, Dave Kleiman, Shyaam Sundhar R.S, **Overwriting Hard Drive Data: The Great Wiping Controversy**
3. PC Guide's **Hard Disk Data Encoding and Decoding** (http://www.pcguide.com/ref/hdd/geom/data.htm)
4. Ian McLoughlin, **Magnetic Recording Fundamentals** (http://www.lintech.org/comp-per/07MAGREC.pdf)