

Disk drive

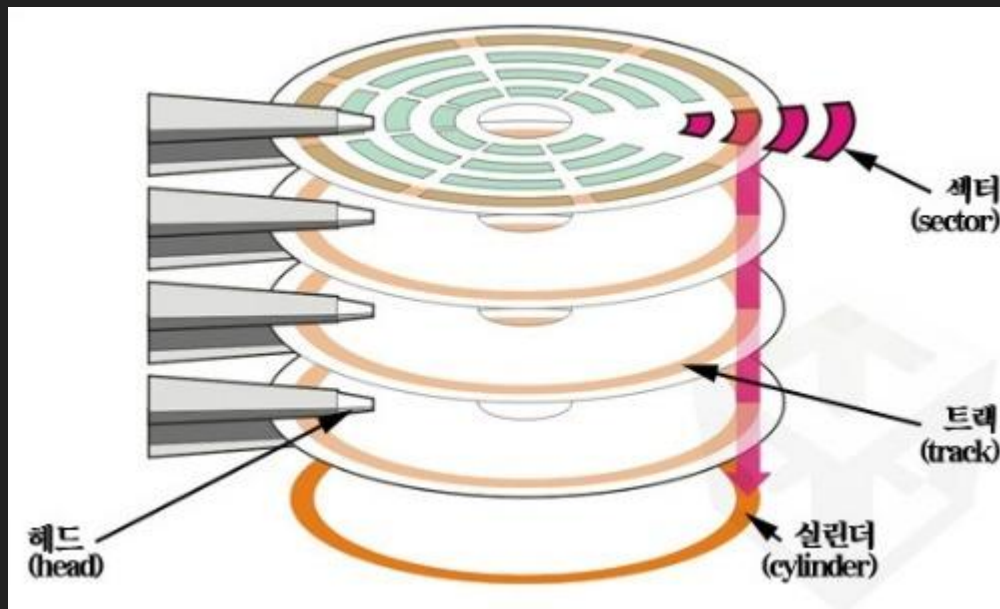
Harddisk



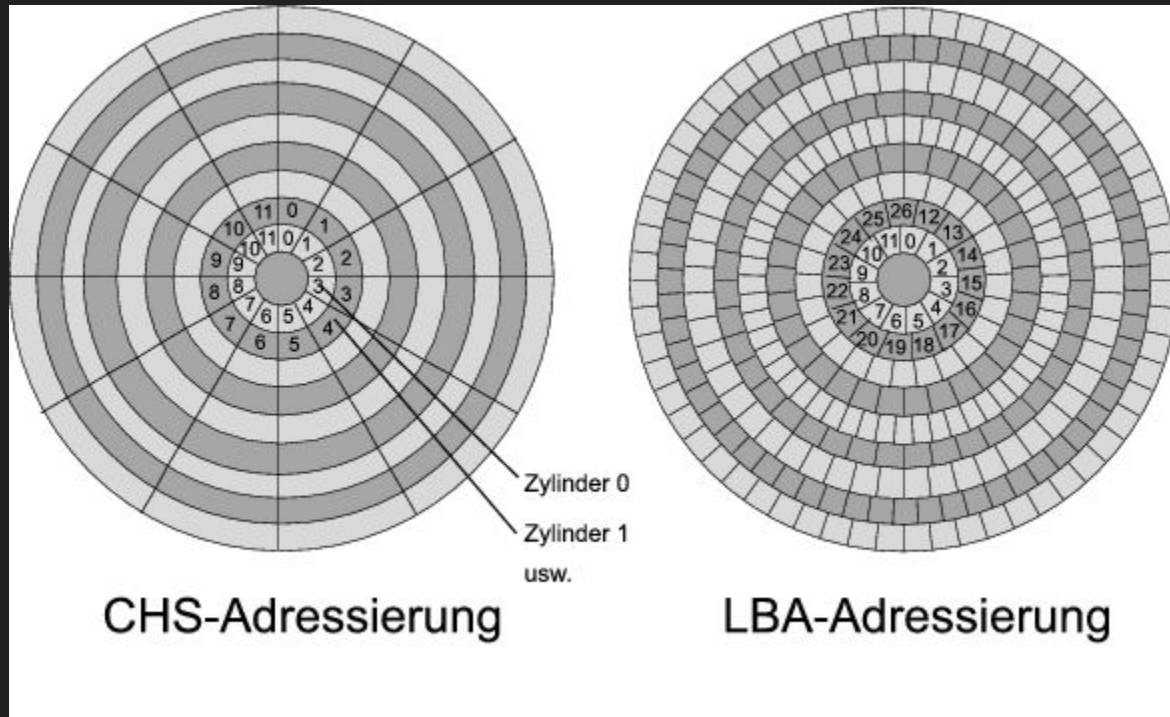
디스크의 구조



디스크의 기하학



CHS vs LBA



Sector 확인

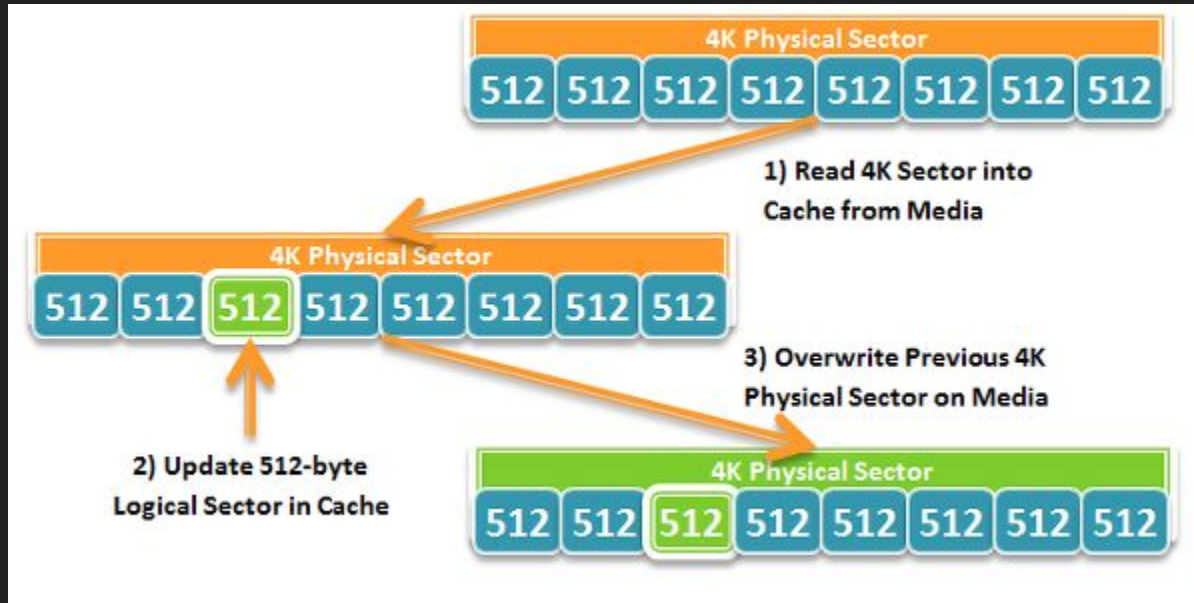
```
# fdisk -lu /dev/sda
Disk /dev/sda: 32.2 GB, 32212254720 bytes
255 heads, 63 sectors/track, 3916 cylinders, total 62914560 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk identifier: 0x00060d5a
```

Device	Boot	Start	End	Blocks	Id	System
/dev/sda1		2048	20482047	10240000	83	Linux
/dev/sda2		20482048	30722047	5120000	83	Linux
/dev/sda3		30722048	34818047	2048000	82	Linux swap / Solaris
/dev/sda4		34818048	62914559	14048256	5	Extended
/dev/sda5		34820096	36868095	1024000	83	Linux
/dev/sda6	*	36870144	62914559	13022208	83	Linux

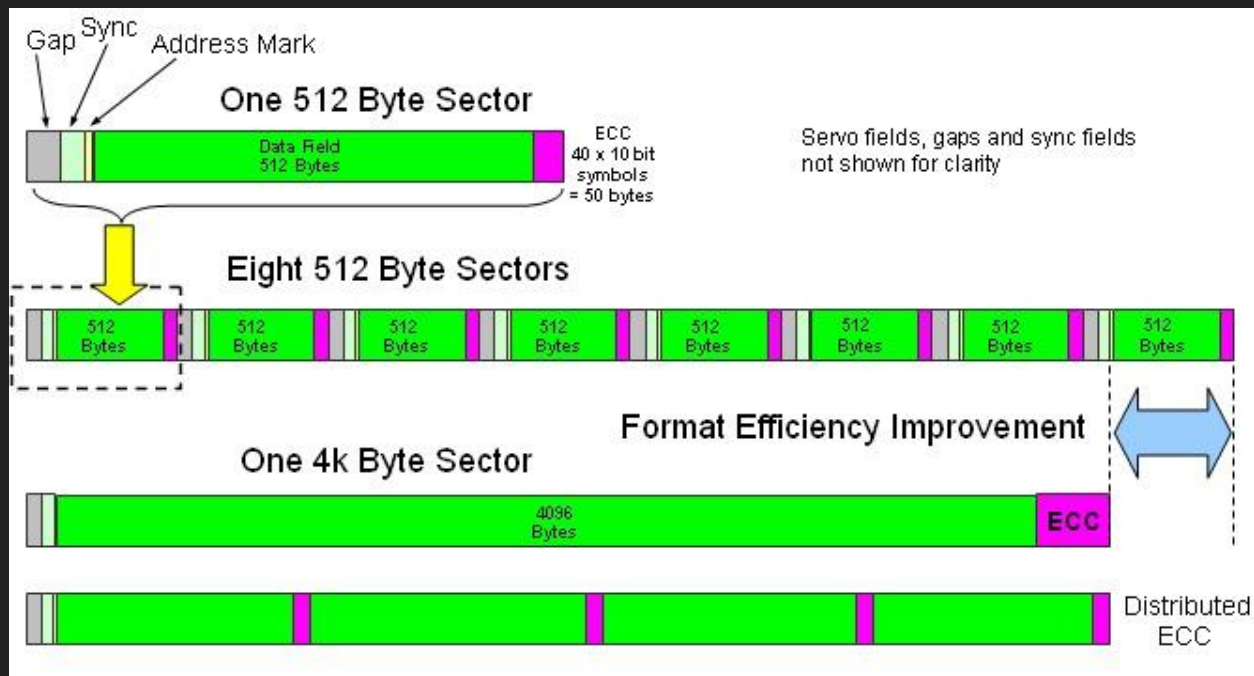
Advanced Format



4K Read-Modify-Write



AF로 얻을 수 있는 효율성



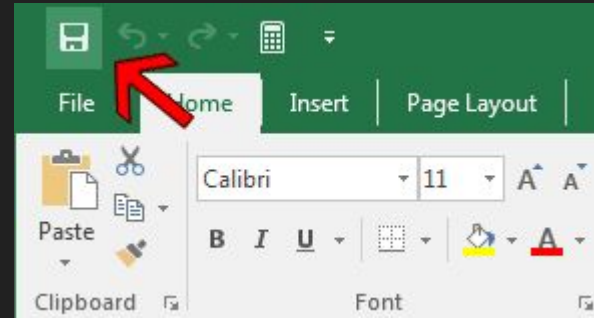
논리 섹터 크기

- Windows
 - fsutil fsinfo ntfsinfo D:
 - 섹터당 바이트: 512
- Linux
 - parted /dev/sdb print
 - 섹터 크기(논리/물리적): 512B/4096B

디스크 크기 증가가 가능했던 이유

- 플래터 기록방식
- 데이터 인코딩
 - RLL
- ZBR(Zone Bit Recording)
- 양면 기록 방식
- 공정의 발달
 - 집적도 향상

Floppy Disk



Solid State Drive



인터페이스별로



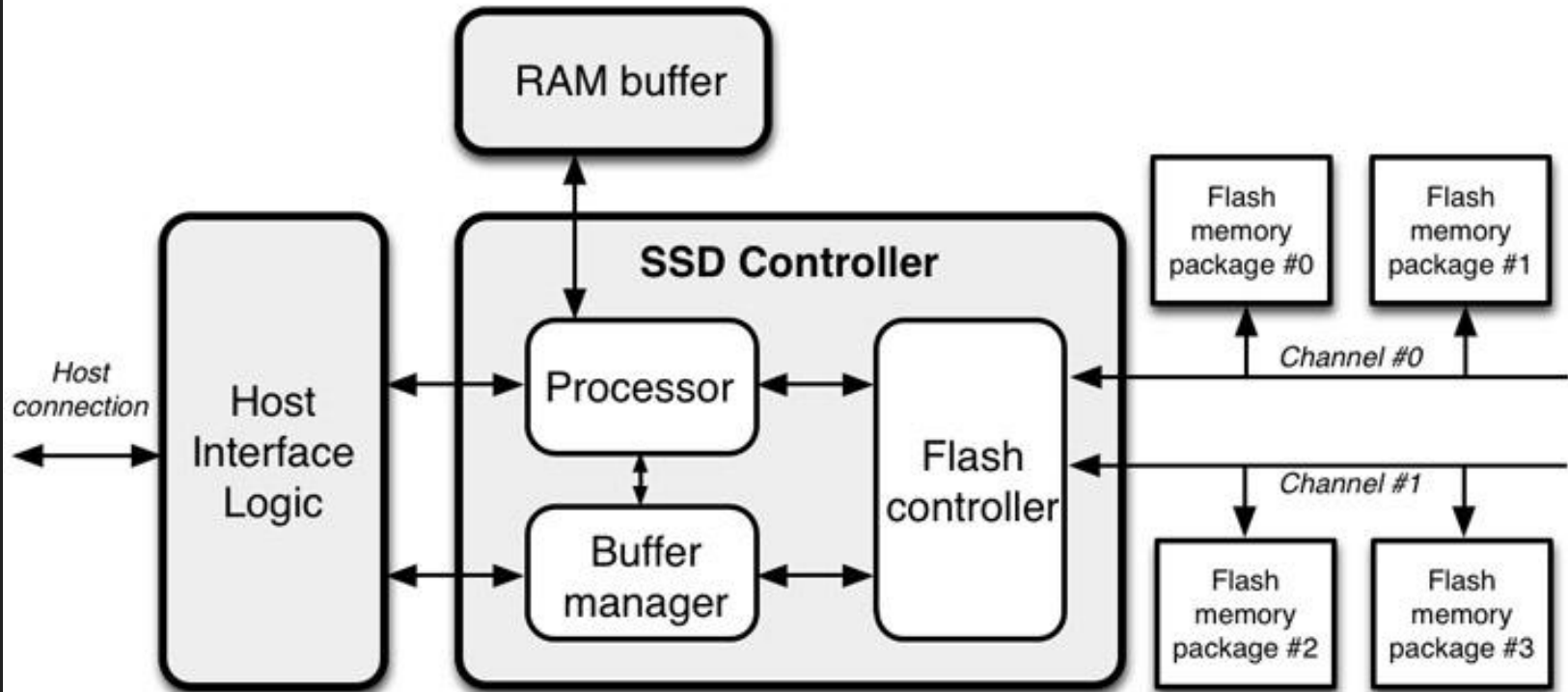
SSD

- NAND 플래시 메모리 셀
 - 수명제한 (Wearing-off)
- Cell Type
 - SLC, MLC, TLC
- Program/Erase

Latency

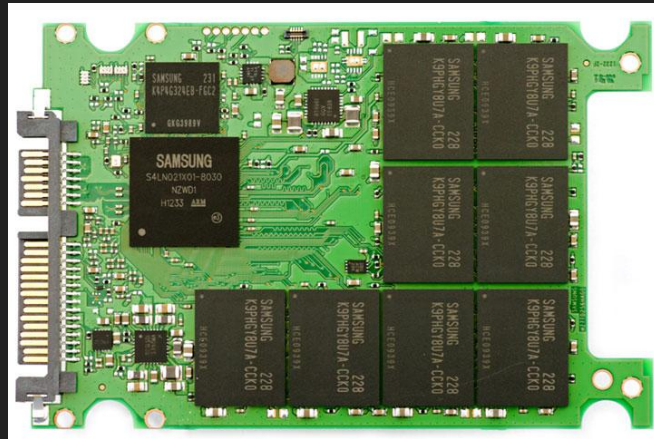
	SLC	MLC	TLC	HDD	RAM	L1 cache	L2 cache
P/E cycles	100k	10k	5k	*	*	*	*
Bits per cell	1	2	3	*	*	*	*
Seek latency (μ s)	*	*	*	9000	*	*	*
Read latency (μ s)	25	50	100	2000-7000	0.04-0.1	0.001	0.004
Write latency (μ s)	250	900	1500	2000-7000	0.04-0.1	0.001	0.004
Erase latency (μ s)	1500	3000	5000	*	*	*	*

Architecture of a solid-state drive



삼성 SSD 840 Pro (512GB) - 2013.08

- 1 SATA 3.0 인터페이스
- 1 SSD 컨트롤러 (삼성 MDX S4LN021X01-8030)
- 1 RAM 모듈 (256MB DDR2 삼성 K4P4G324EB-FGC2)
- 8 MLC NAND - 플래시 모듈, 각 모듈은 64GB (삼성 K9PHGY8U7A-CCK0)



Micron P420m Enterprise PCIe

- 8 lane PCI Express 2.0 인터페이스
- 1 SSD 컨트롤러
- 1 RAM 모듈 (DRAM DDR3)
- 64 MLC NAND - 플래시 모듈, 각 모듈은 32GB
- 2TB (provisioning제외 1.4TB 사용)



IO Access

- Read/Write
 - Page 단위 (2KB, 4KB, 8KB, 16KB)
- Erase
 - Block 단위 (256KB~4MB)
 - 128~256개의 Page
- Cannot be overwritten
 - Read-Modify-Write
 - Page state - free, stale
- Wear leveling
 - FTL

Flash Translation Layer

HDD와 동일한 인터페이스

1. Logical block mapping

- a. LBA \leftrightarrow PBA
- b. Block & Page 레벨 매핑의 trade-off
- c. Log structured filesystem과 비슷한 형태

2. Garbage-collection

- a. stale \rightarrow free

그 외

1. TRIM

- a. ext4 ~
- b. NTFS Windows 7(PCIe 제외) ~
- c. GC의 erase작업이 효율적으로 처리 될 수 있도록
- d. 컨트롤러, OS, 파일시스템에서 모두 지원해야 기능할 수 있음

2. Over-provisioning

- a. 성능 및 wear leveling에 도움

3. NCQ(Native Command Queuing)

- a. HDD에도 있던 기술이지만 SSD에서도 좋음

SAS vs SATA

요구 사항		SAS	SATA
작동 가용성		/ 하루 24시간 -/ 주 7일	8시간, -/ 주 5일
작업 부하		100%	10-20%
비용 민감도		약간 민감 비용	낮은 비용에 민감
성능	지연 시간을 찾	@ 15k RPM 5.7 밀리초	13msec 동안 @ 7200 RPM(작)
	명령어 대기열 및 재정렬	전체	제한
	회전 진동 방지	최대 21 rads/sec/초	최대 5 12 rads/sec/초
	일반적인초 당 I/o/ 드라이브	319	77
	이중 모드 작동	전체	반이중
신뢰성	잘못된 섹터 복구	일반적인 시간 초과 7-15초만	아웃 최대 30초
	범위 오차 감지	프로세서 전용 서보 및 데이터 경로	단일 프로세서 Servo/ 데이터 경로를 결합 또는 없음
	진동 센서	RV 보상 피드백 메커니즘	RV 보상
	가변 섹터 크기	활용 528- 바이트 섹터 및 I/o 컨트롤러	가변 섹터 크기(잠긴 512바이트)이 사용되지 않을 수 있습니다.
	MTBF	1.2m 시간 45°C	25°C에서 700k 시간
	내부 데이터 무결성 검사	종료 종료	메모리 버퍼에 제한적, 없음
	최대 작동 온도	최대 60°C	최대 40°C
	보증	5년	~ 3년

References

- [Kakao Tech - 개발자를 위한 SSD\(Coding for SSD\)](#)
- [Forensic-proof](#)
- [Naver D2 - SSD는 소프트웨어 아키텍처를 어떻게 바꾸고 있는가?](#)
- [Naver D2 - SSD를 쓰면 DBMS가 빨라질까?](#)
- [Intel - SAS vs SATA Spec 비교](#)