Flash Translation Layer (FTL)

RD / WR 요청이 발생했을 때 해당 동작을 SSD 내부의 NAND Flash Memory 구조에 맞게 바꾸 과정이 FTL 을 통해서 이루어진다 .

Garbage Collection
Wear Leveling

FTL 은 Logical Block Mapping 을 통해서 가능한데

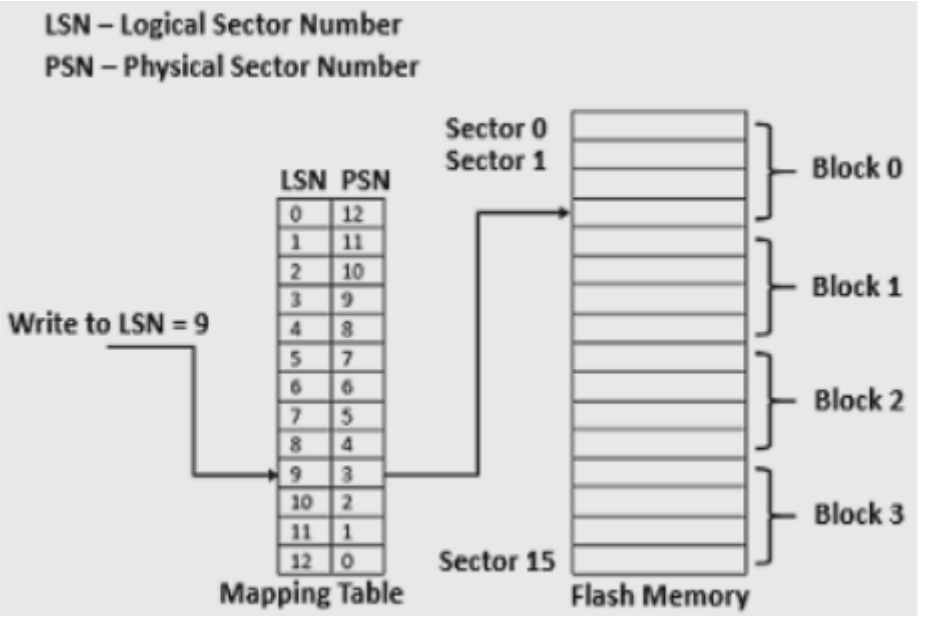
- **Page(Sector) Mapping**
- **Block Mapping**
- **Hybrid Mapping**

Flash Translation Layer (FTL)

1. Page(Sector) Mapping

1 block = 4 sector
16 logical sectors
Row-size of mapping table = 16

Pros : short access time ?
Cons : mapping table size might be too large



"LSN 9 번에 데이터를 write 해줘" 라는 요청이 들어오면 mappping table 결과에 따라서 PSN 3 번에 해당 데이터를 기록한다 .
만약 mapping table 에 해당 PSN 이 없으면 비어있는 physical pages 를 찾아서 기록하고 모든 physical pages 가 사용중이면 valid 데이터를 빈 공간으로 복사하고 mapping table 을 수정한 뒤 erase 한다 .

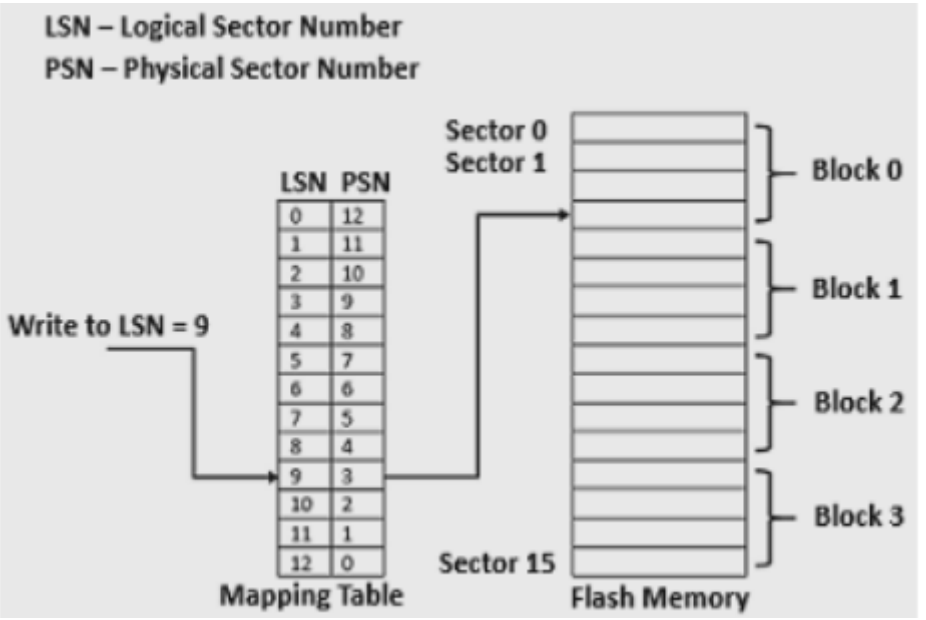
mapping table 을 플래시 메모리에 저장하거나 write 요청이 들어올 때마다 LSN 을 따로 저장하는 방식으로 언제든지 rebuild (from failure) 가능하다 .

Flash Translation Layer (FTL)

1. Page(Sector) Mapping

Table 1. Measures of Sector mapping scheme

Garbage collection cost	Block Erase is done when a block is completely utilized.
RAM requirement	Proportional to flash size
Search time	Not required
Usefulness	Useful in case strict time requirement



"LSN 9 번에 데이터를 write 해줘" 라는 요청이 들어오면 mappping table 결과에 따라서 PSN 3 번에 해당 데이터를 기록한다 .

만약 mapping table 에 해당 PSN 이 없으면 비어있는 physical pages 를 찾아서 기록하고 모든 physical pages 가 사용중이면 valid 데이터를 빈 공간으로 복사하고 mapping table 을 수정한 뒤 erase 한다 .

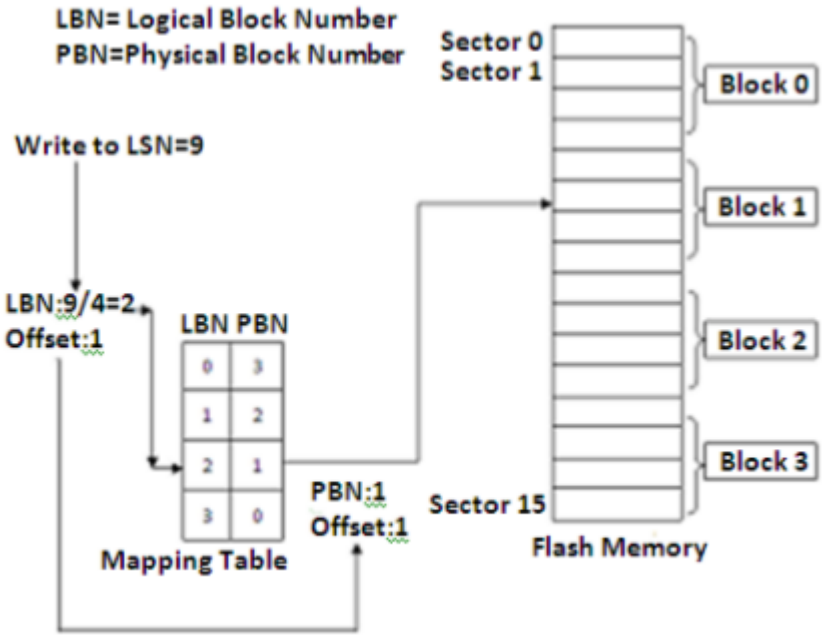
mapping table 을 플래시 메모리에 저장하거나 write 요청이 들어올 때마다 LSN 을 따로 저장하는 방식으로 언제든지 rebuild (from failure) 가능하다 .

Flash Translation Layer (FTL)

2. Block Mapping

if m pages in 1 block,
mapping table 크기는 1/m 이 됨
row-size of mapping table = num_of_blocks

Pros : size of mapping table very small
Cons : requires many read / write operation



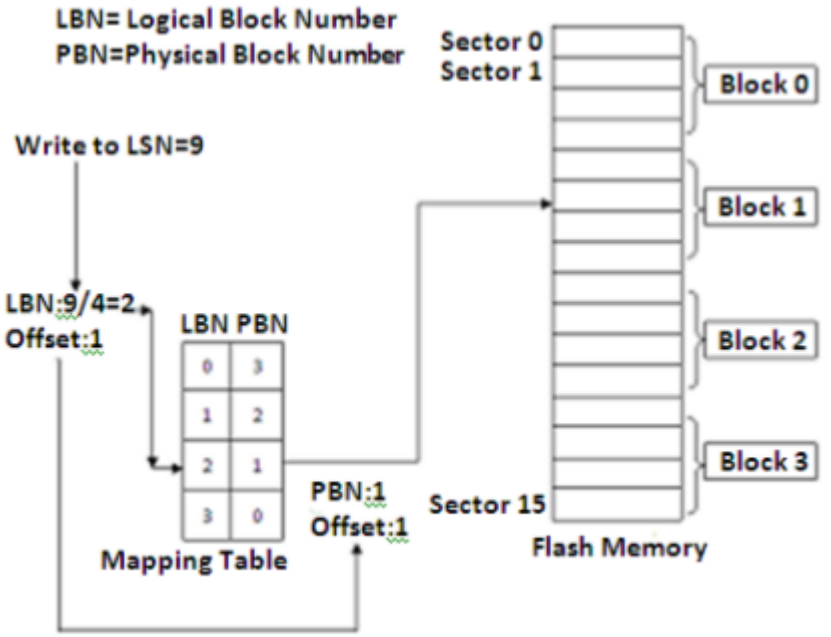
Block Mapping 에서 LBN 은 반드시 offset 정보를 나타내는 PBN 으로 매핑된다 .
“LSN 9 번에 데이터를 write 해줘” 라는 요청이 들어오면 LBN = LSN / num_of_blocks 으로 블록을 찾아가고 PBN 이 디폴트이면 그대로 기록하고 , 이미 기록되어 있으면 해당 데이터를 비어있는 블록으로 copy – erase – copy back 을 실행한다 . 이 과정에서 read / write 연산이 많이 발생한다 .

Flash Translation Layer (FTL)

2. Block Mapping

Table 2. Measures of block mapping scheme

Garbage collection cost	Block Erase is done hen a block is completely utilized.
RAM requirement	Proportional to flash size
Search time	Not required
Usefulness	Useful in case of strict time requirement



Block Mapping 에서 LBN 은 반드시 offset 정보를 나타내는 PBN 으로 매핑된다 .

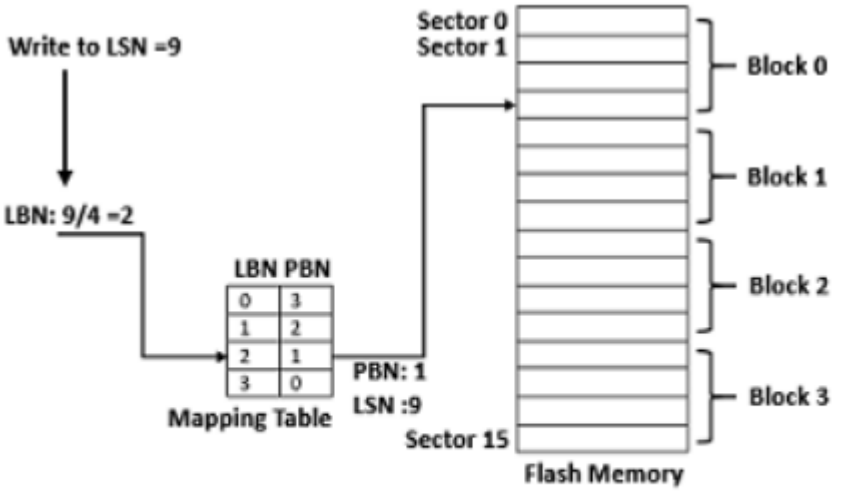
“LSN 9 번에 데이터를 write 해줘” 라는 요청이 들어오면 LBN = LSN / num_of_blocks 으로 블록을 찾아가고 PBN 이 디폴트이면 그대로 기록하고 , 이미 기록되어 있으면 해당 데이터를 비어있는 블록으로 copy – erase – copy back 을 실행한다 . 이 과정에서 read / write 연산이 많이 발생한다 .

Flash Translation Layer (FTL)

3. Hybrid Mapping

Page Mapping 과 Block Mapping 의 단점을 극복하기 위해서

Log Block : save page mapping info
Data Block : save block mapping info



데이터 수정 및 삽입 시에 Log Block 을 먼저 기록 후 Data Block 에도 기록하기

log 와 data 는 RAM inside SSD 에 저장한다 .

“LSN 9 번에 데이터를 write 해줘” 라는 요청이 들어오면 LBN = LSN / num_of_blocks 를 통해 PBN 을 찾는다 . 해당 PBN 이 비어있으면 그대로 기록하고 비어있지 않으면 다른 비어있는 블록을 찾아서 기록한 뒤에 mapping table 을 업데이트 한다 .

Log Block 과 Data Block 을 서로 consistent 하게 만드는 과정을 merge 라고 한다 .

when Log-block is full it is flushed to Data-block
into a new clean block by writing to a new clean block.

Flash Translation Layer (FTL)

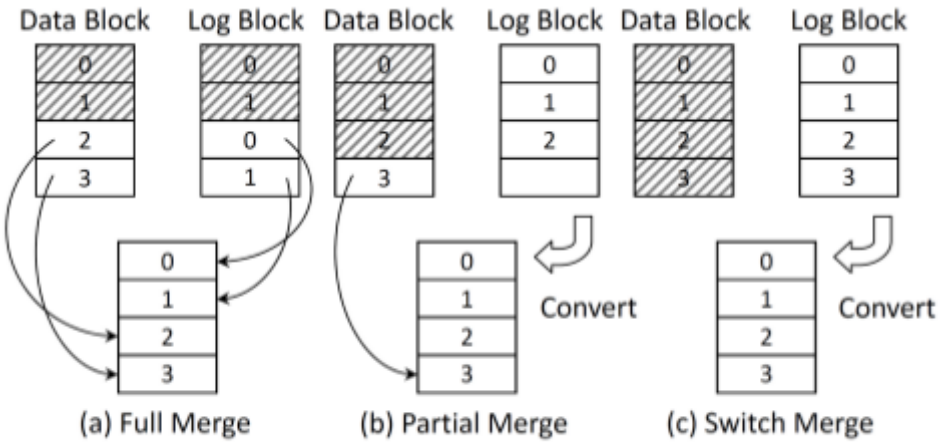


Figure 1: Three Types of Merge Operations

우리가 흔히 알고 있는 copy valid data – erase original block – copy back 방식은 Full Merge 방식이다 .

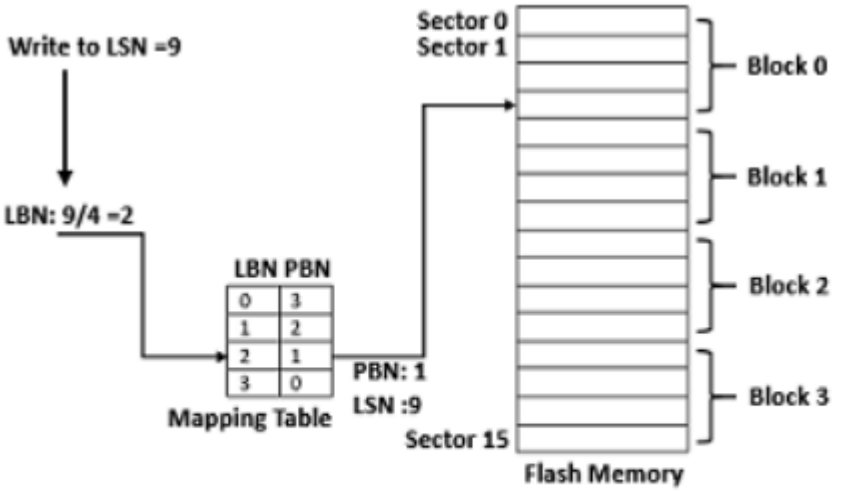
Partial Merge 와 Switch Merge 는 특수한 케이스에만 사용한다 .

Flash Translation Layer (FTL)

3. Hybrid Mapping

Page Mapping 과 Block Mapping 의 단점을 극복하기 위해서

Log Block : save page mapping info
Data Block : save block mapping info



Merge 방식은 세 가지로 분류할 수 있다 .

1) Full Merge : 어떤 log block 이 선택되었는데 first page 부터 last page 까지 NOT sequentially 데이터가 기록되어 있을 경우 블록 전체를 new clean block 으로 복사
   m read, m write, 2 erase operation

2) Switch Merge : 어떤 log block 이 선택되었는데 sequentially 데이터가 기록되어 있을 경우 해당 log block 이 data block 으로 사용됨
   1 erase operation

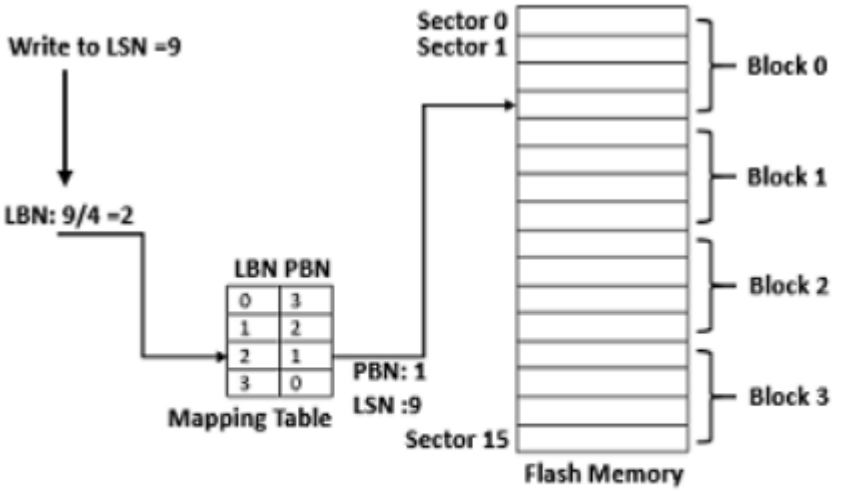
3) Partial Merge : 어떤 log block 이 선택되었는데 first page 부터 in the middle page 까지 sequentially 데이터가 기록되어 있을 경우 data block 으로 비어있는 페이지를 채움
   n read, n write, 1 erase (0 < n < m)

Flash Translation Layer (FTL)

3. Hybrid Mapping

Page Mapping 과 Block Mapping 의 단점을 극복하기 위해서

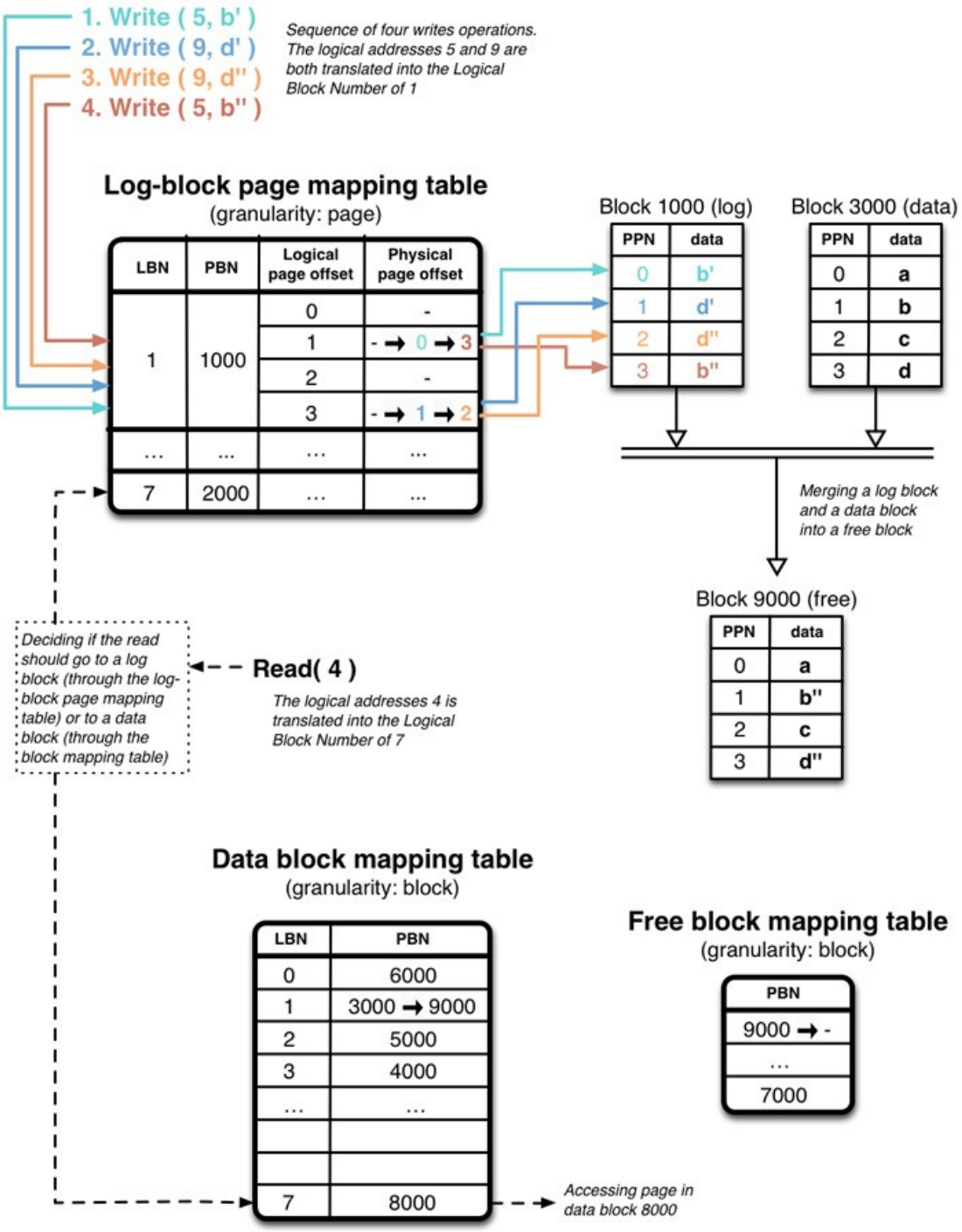
Log Block : save page mapping info
Data Block : save block mapping info



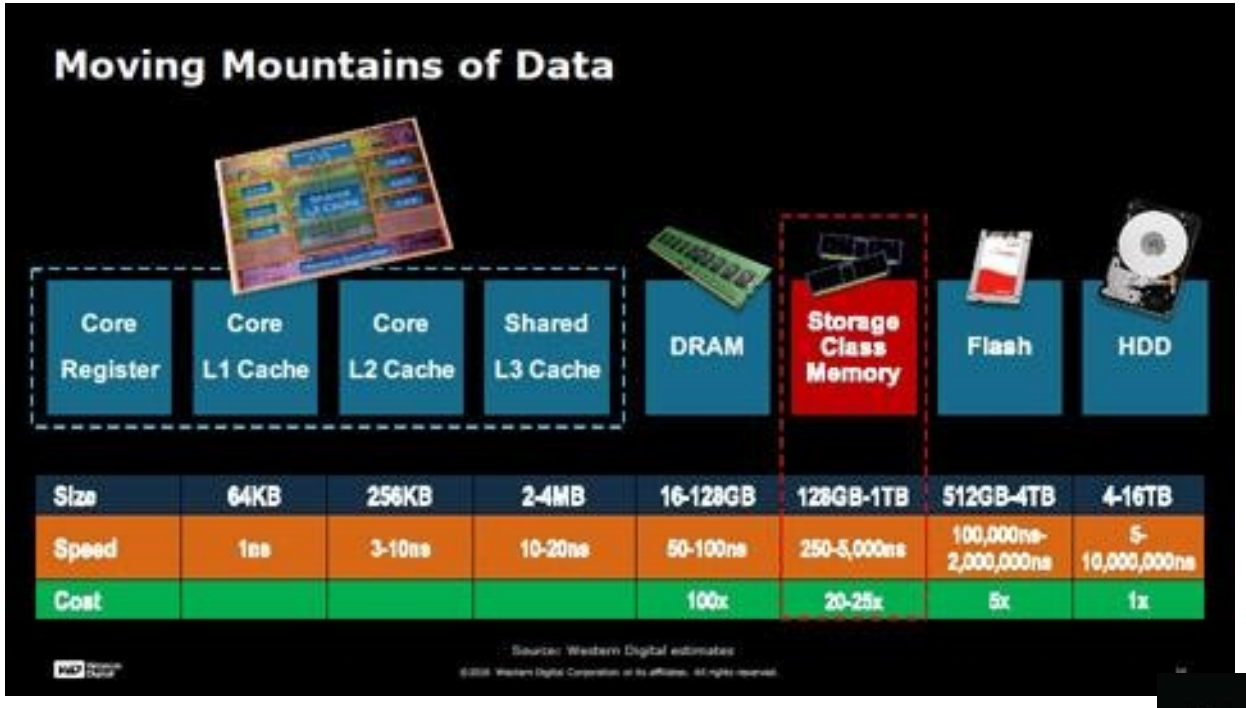
Hybrid Mapping 에는 아주 다양한 방법이 있다 .

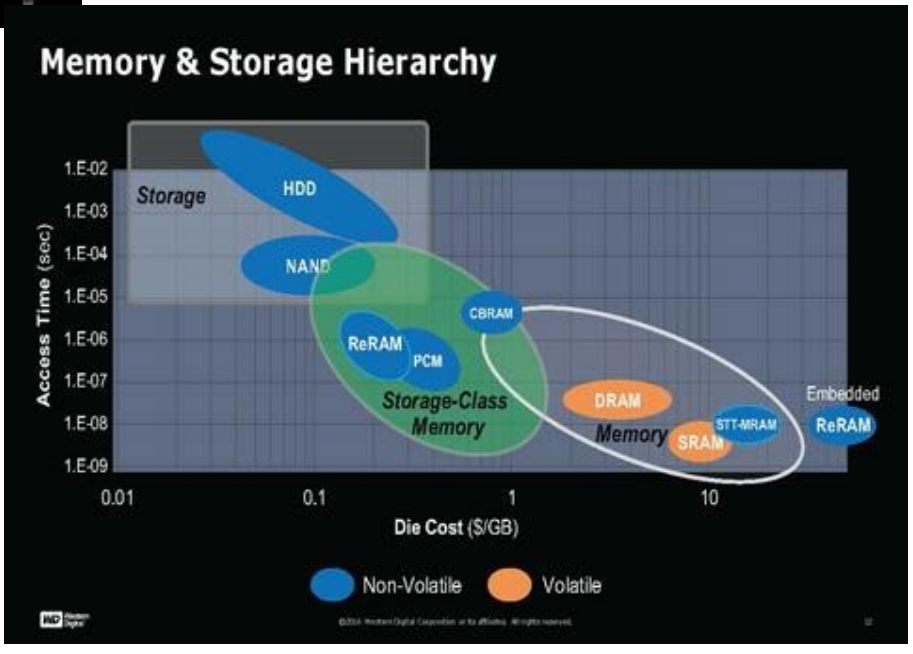
BAST
FAST
LAST
SuperBlock Reconfigurable FTL
DFTL

Hybrid log-block FTL



SCM : Storage Class Memory



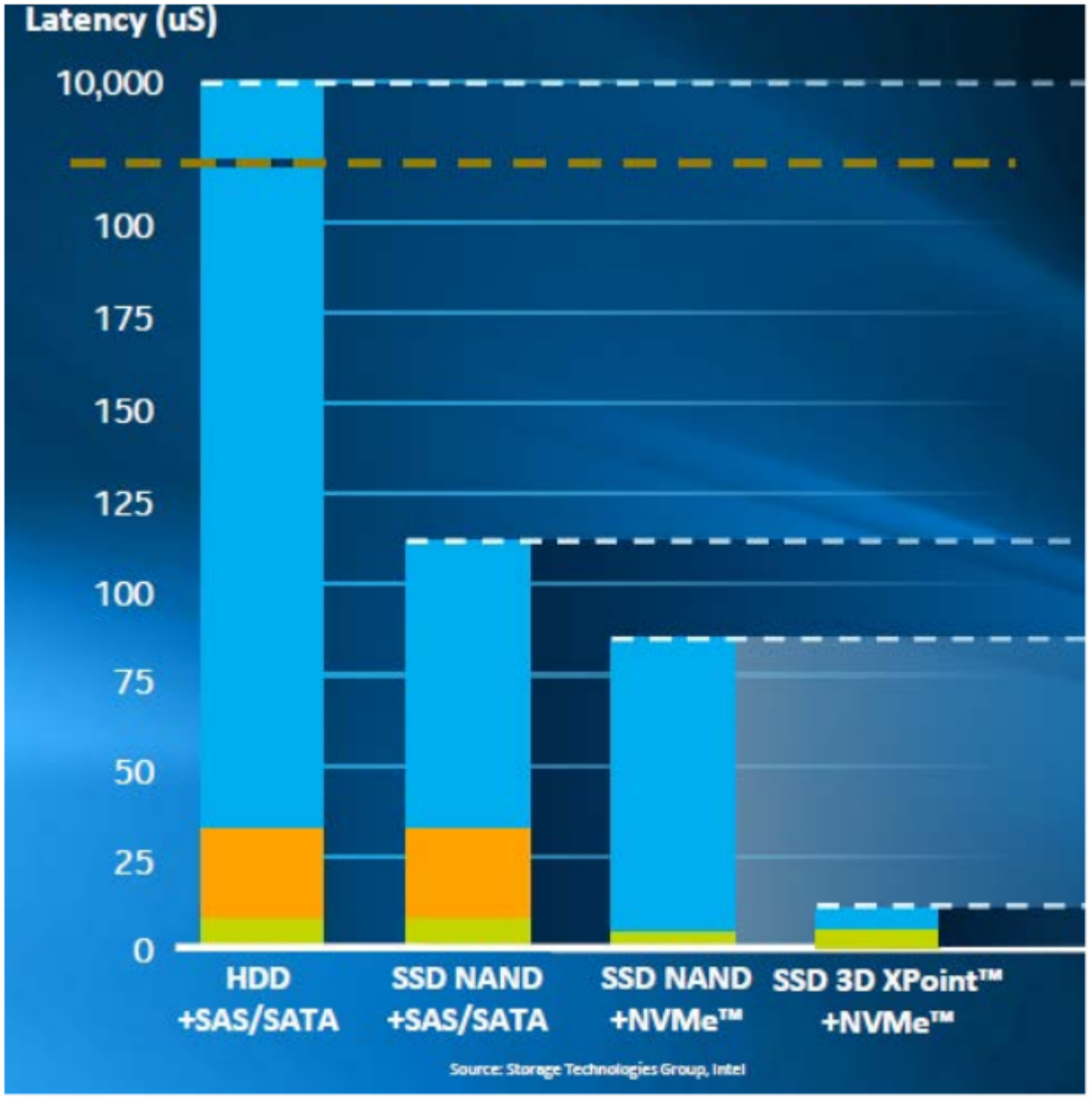


SCM : Storage Class Memory

만약에 L1 Cache I/O 처리 속도를 1 초 (sec) 로 가정한다면 ??

	Nanoseconds (ns)	Microseconds (µs)	Milliseconds (ms)	If L1 Access is 1 second
L1 Cache Reference	0.5			1 sec
L2 Cache Reference	7			14 secs
DRAM Access	200			6 mins, 40 secs
Intel Octane 3D XPoint	7,000	7		3 hours, 53 mins, 20 secs
Micron 9100 NVMe PCIe SSD Write	30,000	30		16 hours, 40 mins
Mangstor NX NVMeF Array Write	30,000	30		16 hours, 40 mins
DSSD D5 NVMeF Array	100,000	100		2 days, 7 hours, 33 mins, 20 secs
Mangstor NX NVMeF Array Read	110,000	110		2 days, 13 hours, 6 mins, 40 secs
NVMe PCIe SSD Read	110,000	110		2 days, 13 hours, 6 mins, 40 secs
Micron 9100 NVMe PCIe SSD Read	120,000	120		2 days, 18 hours,m40 mins
Disk Seek	10,000,000	10,000	10	7 months, 10 days, 11 hours, 33 mins, 20 secs
DAS Disk Access	100,000,000	100,000	100	6 years, 4 months, 19 hours, 33 mins, 20 secs
SAN Array Access	200,000,000	200,000	200	9 years, 6 months, 2 days, 17 hours, 20 mins

SCM : Storage Class Memory



SCM : Storage Class Memory

	Latency	Specified Max Bandwidth	Expected BW - 1K Byte Record	Expected BW - 8K Byte Record
HBM2	15ns	256 GB/sec	53 GB/sec	174 GB/sec
DDR4 DIMMs[1]	25ns	19.2 GB/sec	13 GB/sec	18 GB/sec
SSD – PCIe3[2]	20us	2.8 GB/sec	49 MB/sec	357 MB/sec
SSD – SATA3[3]	55us	500 MB/sec	18 MB/sec	115 MB/sec
15K SAS HDD[4]	5.5ms	246 MB/sec	0.2 MB/sec	0.8 MB/sec

Notes and sources: Rambus Analysis.

1. DDR4 @ 2400 Mbps

2. Intel SSD DC P3700 Series

3. Intel SSD DC S3710 Series

4. Cheetah 15K.5 SAS Hard Drive

	SLC	MLC	TLC	HDD	RAM
P/E cycles	100k	10k	5k	*	*
Bits per cell	1	2	3	*	*
Seek latency (µs)	*	*	*	9000	*
Read latency (µs)	25	50	100	2000-7000	0.04-0.1
Write latency (µs)	250	900	1500	2000-7000	0.04-0.1
Erase latency (µs)	1500	3000	5000	*	*
Notes	* metric is not applicable for that type of memory				
Sources	P/E cycles [20] SLC/MLC latencies [1] TLC latencies [23] Hard disk drive latencies [18, 19, 25] RAM latencies [30, 52] L1 and L2 cache latencies [52]				

SCM : Storage Class Memory

TABLE I
MEMORY TECHNOLOGY SUMMARY [1]

	Read time (ns)	**Write time (ns)**	**Read BW (MB/s)**	**Write BW (MB/s)**
DRAM	10	10	1,000	900
PCRAM	20-200	80-10^4	200-800	100-800
SLC Flash	10^4-10^5	10^4-10^7	0.1	10^{-3}-10^{-1}
ReRAM	5-10^5	5-10^8	1-1000	0.1-1000
Hard drive	10^6	10^6	50-120	50-120

TABLE II
READ AND WRITE PERFORMANCE (MB/S) ON RAW BLOCK DEVICE

	Write Random Bandwidth	**Write Sequential Bandwidth**	**Read Random Bandwidth**	**Read Sequential Bandwidth**
Optane	2174.08	2172.62	2286.15	2568.53
HDD	6.08	200.25	2.7	204.30