Flash-based SSDs

Solid State Storage Device - SSD.

flash → NAND-based flash.

flash page, flash block.

⇒ built only w/ transistors (like DRAM).

1. storing a single bit

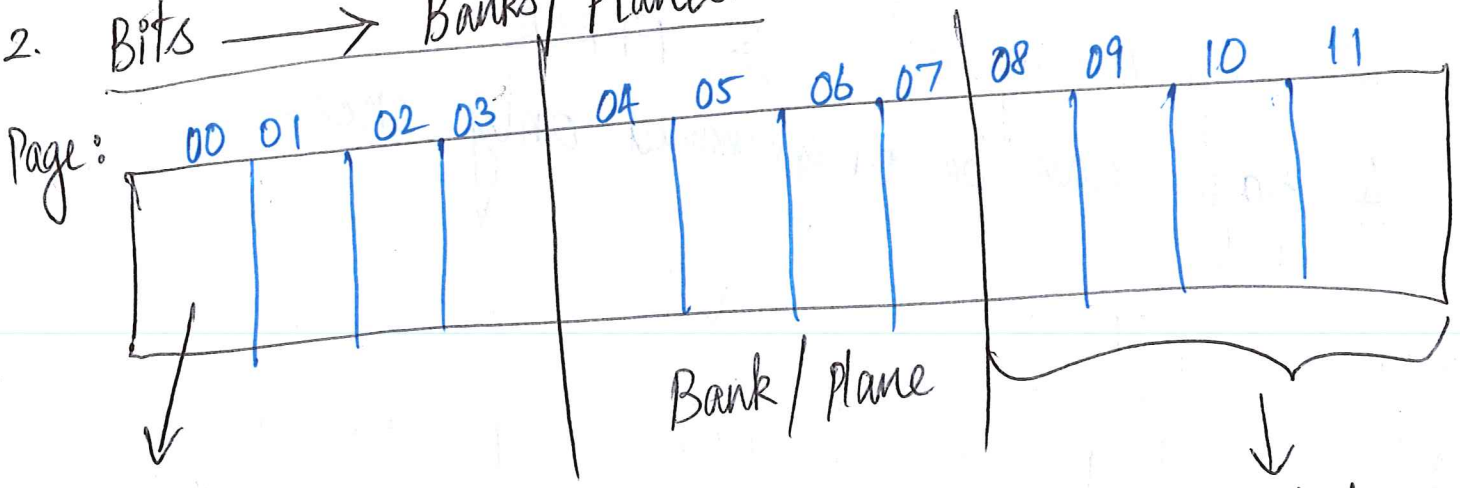
01 : SLC - Single Level Cell

00, 01, 10, 11 : MLC - Multi " "

(00: low, 01: somewhat low, 10: somewhat high, 11: high)

: TLC - Triple-level cell.

2. Bits → Banks / Planes.



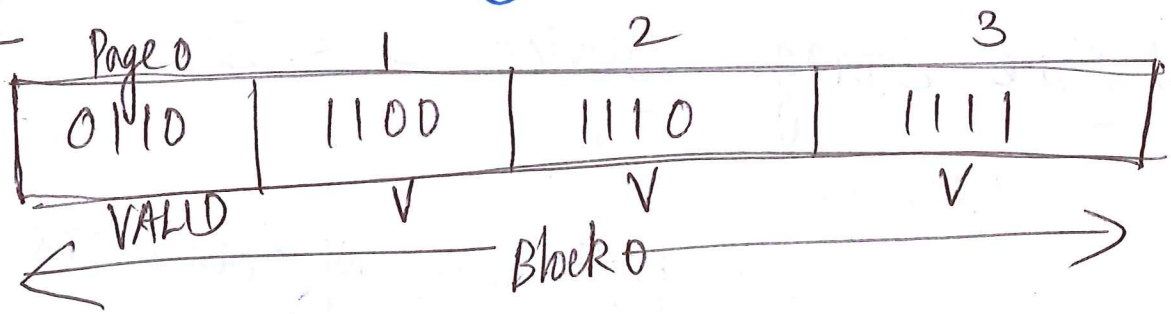
1 page (e.g., 4KB)

1 block. (e.g., 128 KB or 256KB)

3. Basic Flash Operations

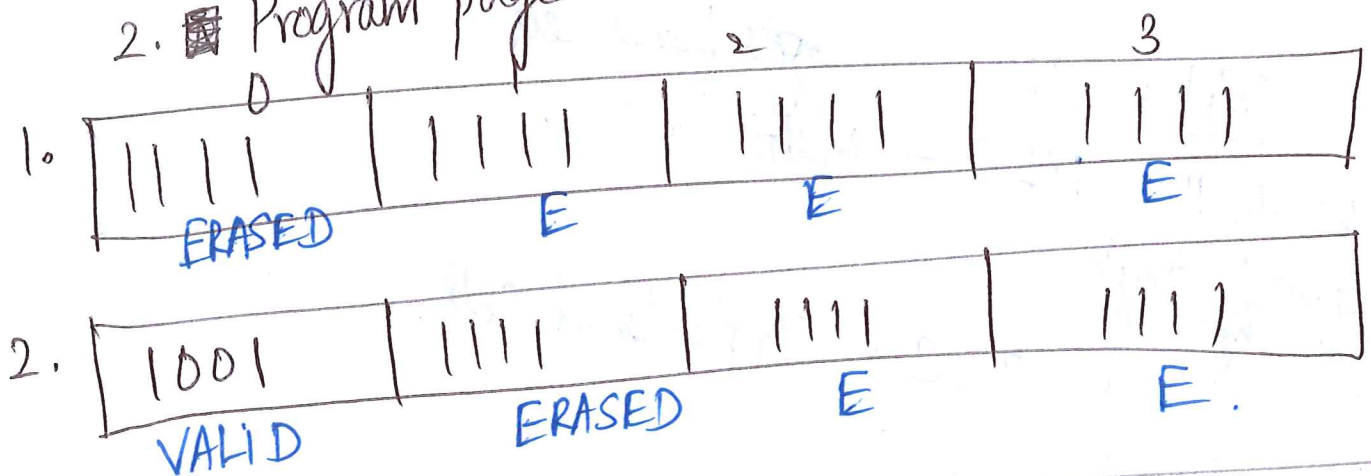
1. Read (a page) - 10s of μs
2. Erase (a block) - few ms → COSTLY!
3. Program (a page) → 100s of μs.

Example



Write to page 0:

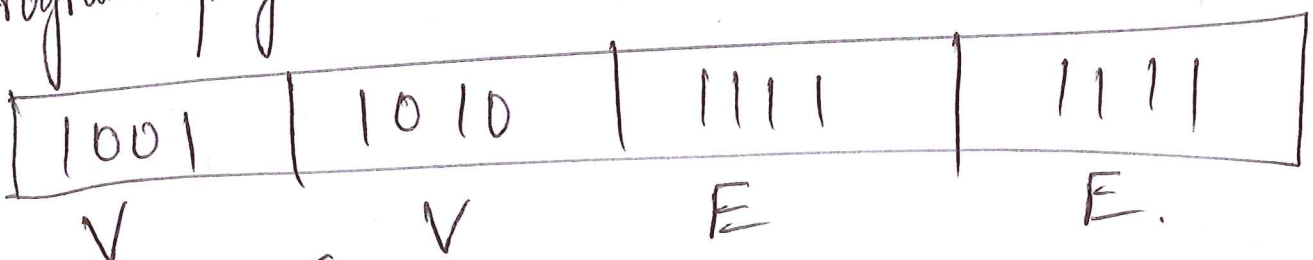
1. Erase Block 0
2. Program page 0.



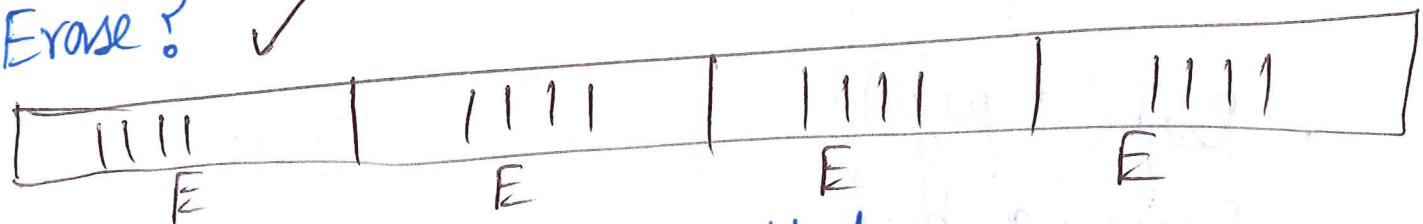
Program page 0 again? ⟹ ERROR

A page can be programmed only once.

Program page 1 now? ⇒ ✓



Erase? ✓



Now: ALL pages programmable!

4. Flash – Performance & (3) Reliability.

SHOW TABLE IN BOOK.

→ Head crash in HDD. – NOT in flash (∵ just silicon)

→ wear out.

SLC – 100,000 P/E cycles

program ↓ ↘ erase.

→ disturbance

read disturbs

program disturbs.

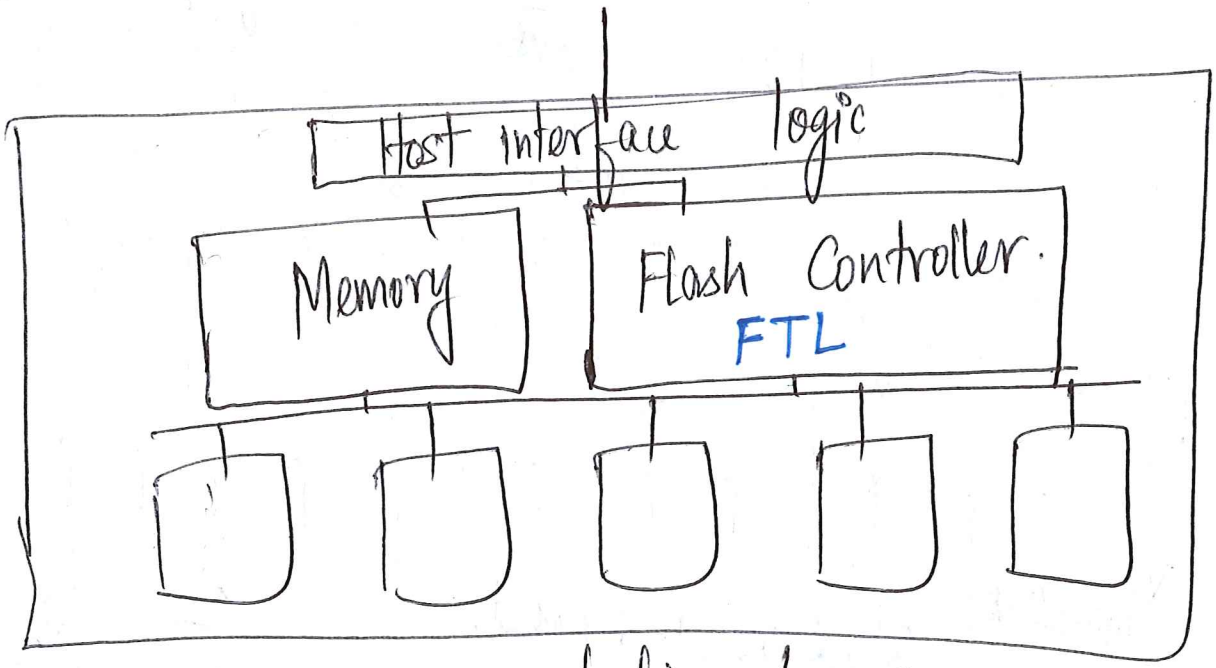
page: 0 1 2 3

0010	0011	1100	0110

→ ???

Read (2) page

5. Raw Flash ⟶ Flash based SSDs



FTL → Flash Translation Layer.

Read/Write Logical block 0 → FTL → Read, Erase, Program → Physical → blocks, pages

6. A Bad FTL organization ④

"direct mapped"

e.g. Logical ~~addr~~ page N ⟶ physical ~~addr~~ page N.

1. Read logical page N ⟶ Read phy. page N

2. WRITE. ⟶ Write Amplification ⟶
   1. Read entire block with phy. page N.
   2. ERASE the block
   3. PROGRAM the old + new pages

Problems?

→ Performance! – slower than HDDs.

→ Reliability? – Wear out of blocks

(metadata) freq. written.

7. A Log-Structured FTL

Write (100) ⟶ logical block

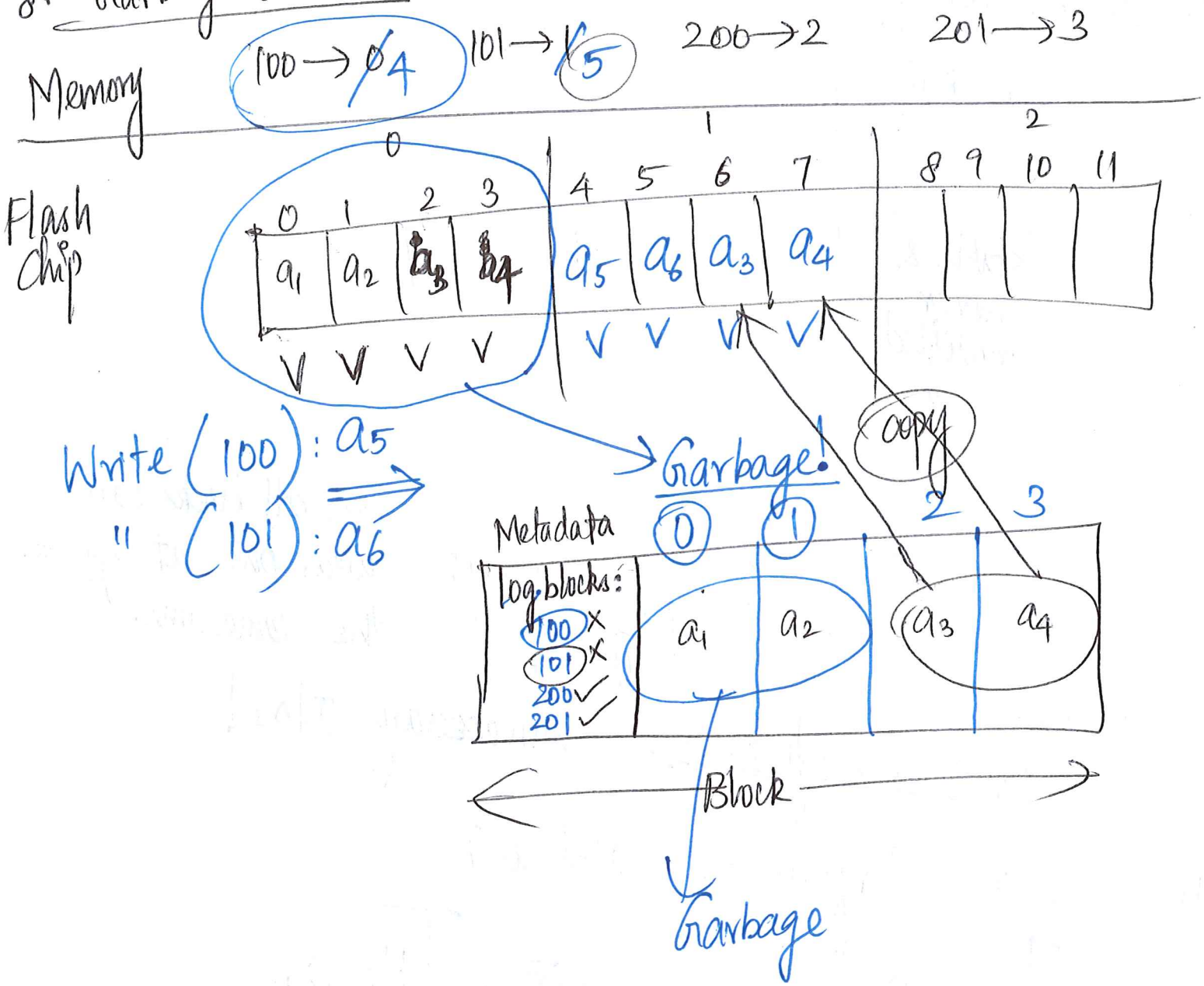
Block 0:
1. Erased. (EEEE)
2. Write page 0;
3. update the mapping table in memory.

log. block 100 → phy. page 0

CRASH! → OOB and Area (in each page) 100 → 0, 101 → 1, 200 → 2, 201 → 3

(X) Wear-leveling!

8. Garbage Collection



9. Mapping Table Size

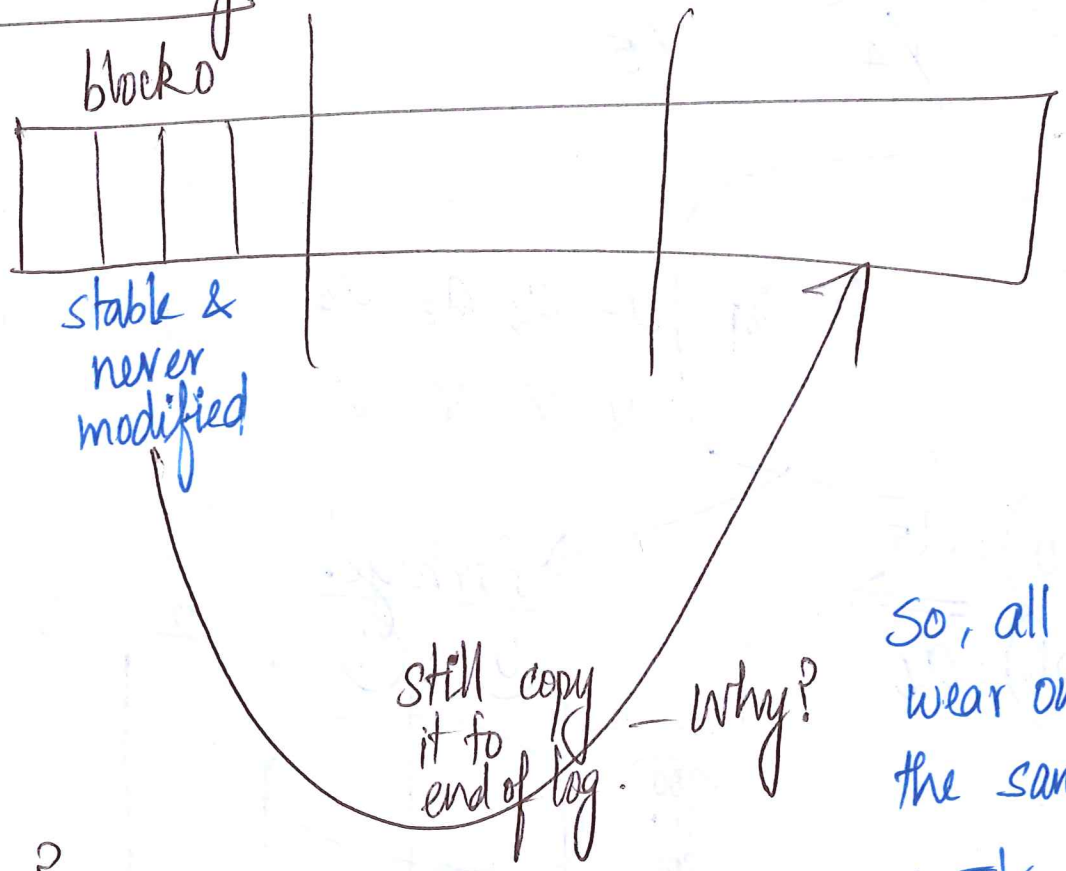
- page-level mapping.
- block-level mapping
- hybrid mapping.

READ OSTEP FOR DETAILS!

$$\frac{1TB\ (\text{Flash mem})}{4KB} = 2^{28}$$

if 1 mapping = 4 bytes
$\Rightarrow$ (1GB) for mapping table

10. Wear Leveling



So, all blocks can wear out at approx. the same time.

Con?

Write amplification — unnecessary I/Os!

11. SSD Performance and Cost

$T_{SSD\ Random\ I/O} \ll T_{HDD\ Random\ I/O}$

$T_{RandomReads_{SSD}} > T_{RandomWrites_{SSD}}$

Why?

	Cap.	Cost
SSD	250GB	$150
HDD	1TB	$50.

$Cost_{SSD} > Cost_{HDD}$

60 cents/GB 5 cents/GB