Overview of Flash and SSDs

Jihong Kim
Dept. of CSE, SNU

Based on:

M. Cornwell. Anatomy of a solid-state drive, ACM Queue 10(10), 2012.

Solid-State Drives (SSDs)

- A Solid-State Drive (SSD) is a data storage device that emulates a hard disk drive (HDD)

- Block devices
  - Small fixed contiguous segments of bytes as the addressable unit
  - Logical addressing to access data and abstract the physical media

- NAND Flash SSD's are essentially arrays of flash memory devices which include a controller that electrically and mechanically emulate, and are software compatible with magnetic HDD's

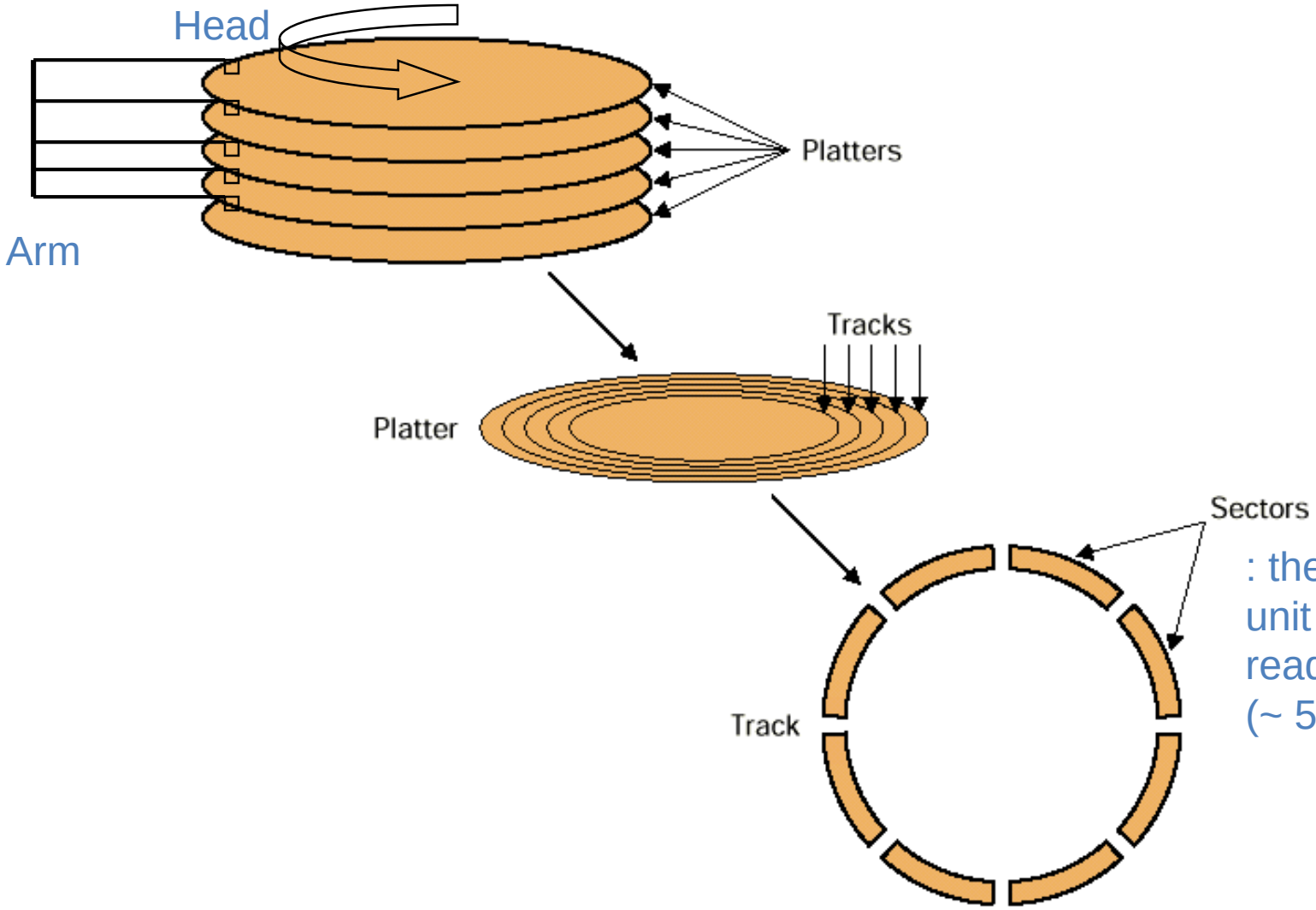
Traditional hard disk drive

Solid state hard drive

SSD over HDD

- No mechanical latency
- Erase-before-overwrite
  - Out-place update
- Asymmetric speed
  - Program: ~ 300 μs (100s of μs)
  - Read: ~ 25 μs (10s of μs)
  - Erase: ~ 2ms (a few ms)
- Asymmetric unit
  - Program & Read: page
  - Erase: block
- Limited lifetime

Magnetic Disks



: the smallest unit that can be read/written (~ 512 bytes)

Key Components of SSDs

- Three Components
  - Storage media
  - Controller
  - Host interface

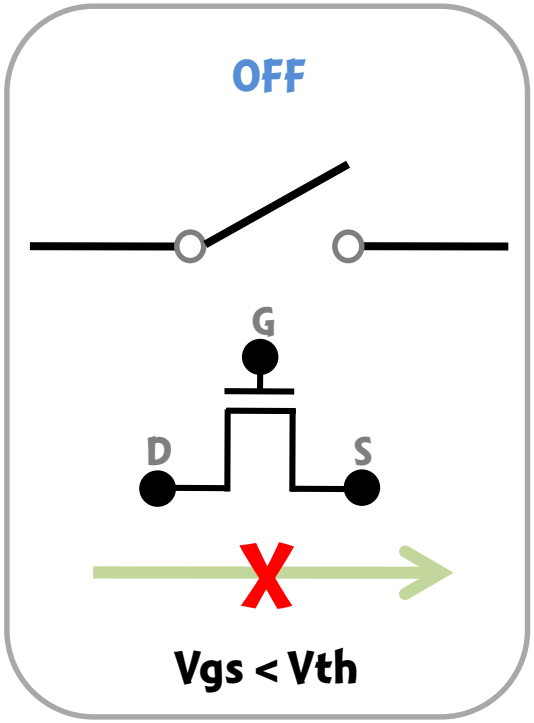
Storage Media:
NAND Flash Memory

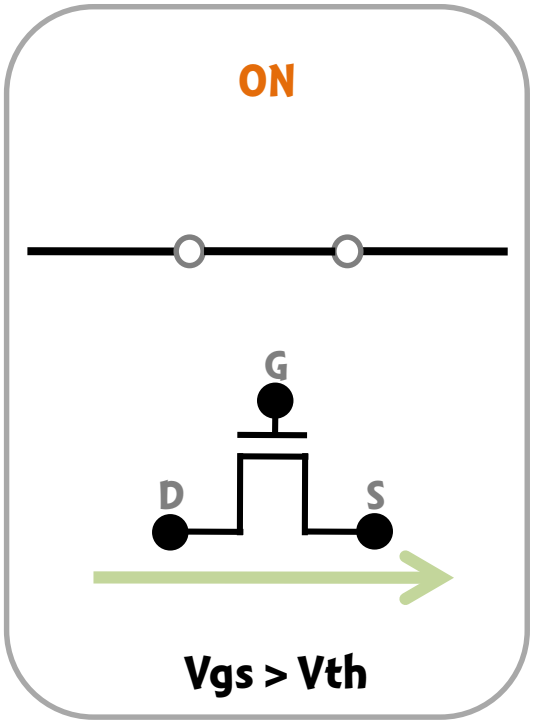
NAND Flash Memory

- **Flash memory** is a non-volatile computer storage chip that can be electrically erased and reprogrammed

Transistor as an Electrical Switch

- Control gate voltage to turn on/off the switch.
  - Turn on → Current flow
  - Turn off → No current flow





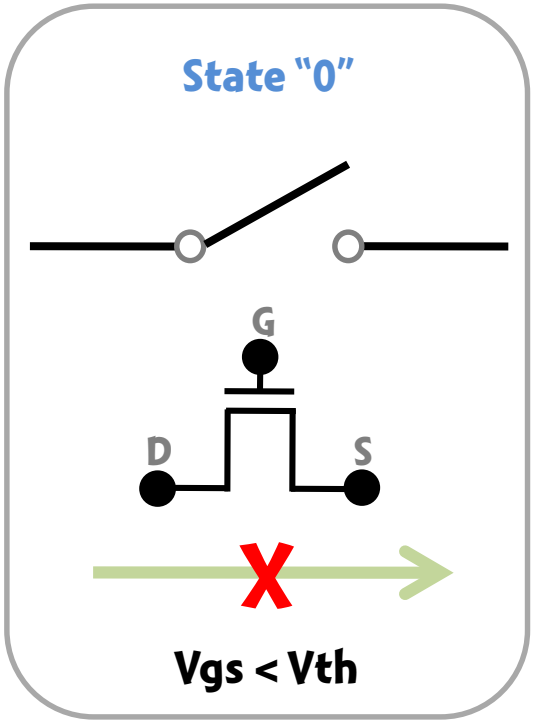
Q: Switch는 메모리는 아니나… 만약… 만약….

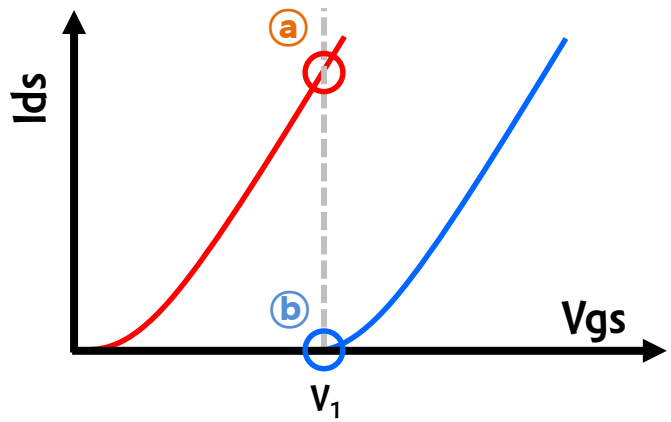
- 두 개의 상태를 구분은 하나 switch만으로는 새로운 데이터를 쓸 수가 없다.

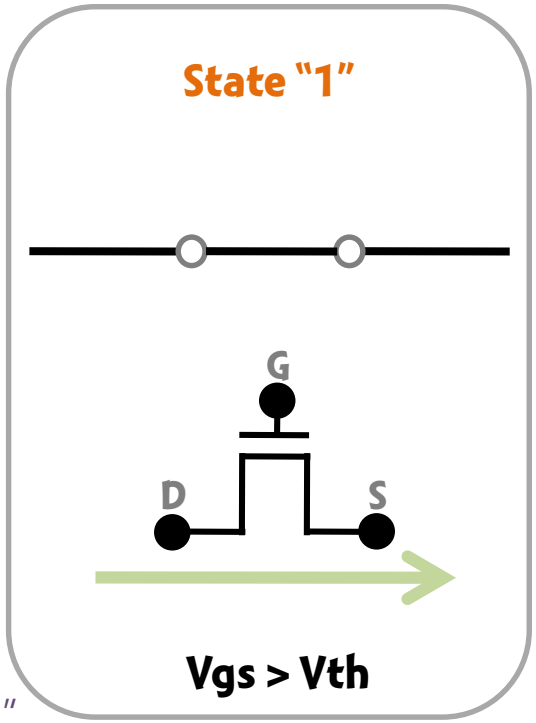
- 해결책은?
  - 고정된 Vth 값에 대해 Vgs를 조정하니…. On/Off 구분이 가능…
  - 만약 Vth값을 변화시킬 수 있으면… 같은 Vgs에 대해서… On/Off 구분이 가능…
  - 만약 다른 상태가 다른 Vth값을 가지도록 할 수 있다면 ….

Transistor as an Electrical Switch

- Control gate voltage to turn on/off the switch.
  - Turn on → Current flow → State 1
  - Turn off → No current flow → State 0







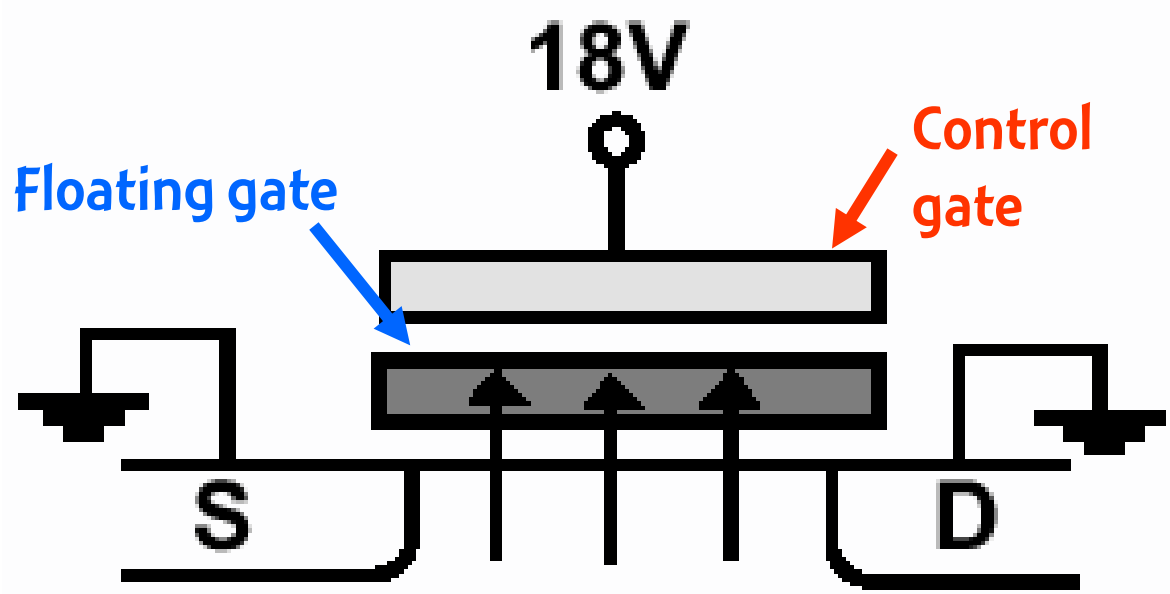
※ When V1 is applied to gate,

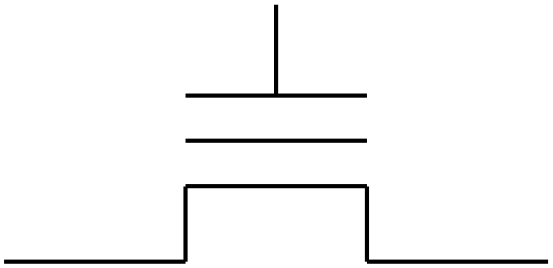
ⓐ: Vth is low → High current → State "1"

ⓑ: Vth is high → No current → State "0"

Floating Gate Transistor

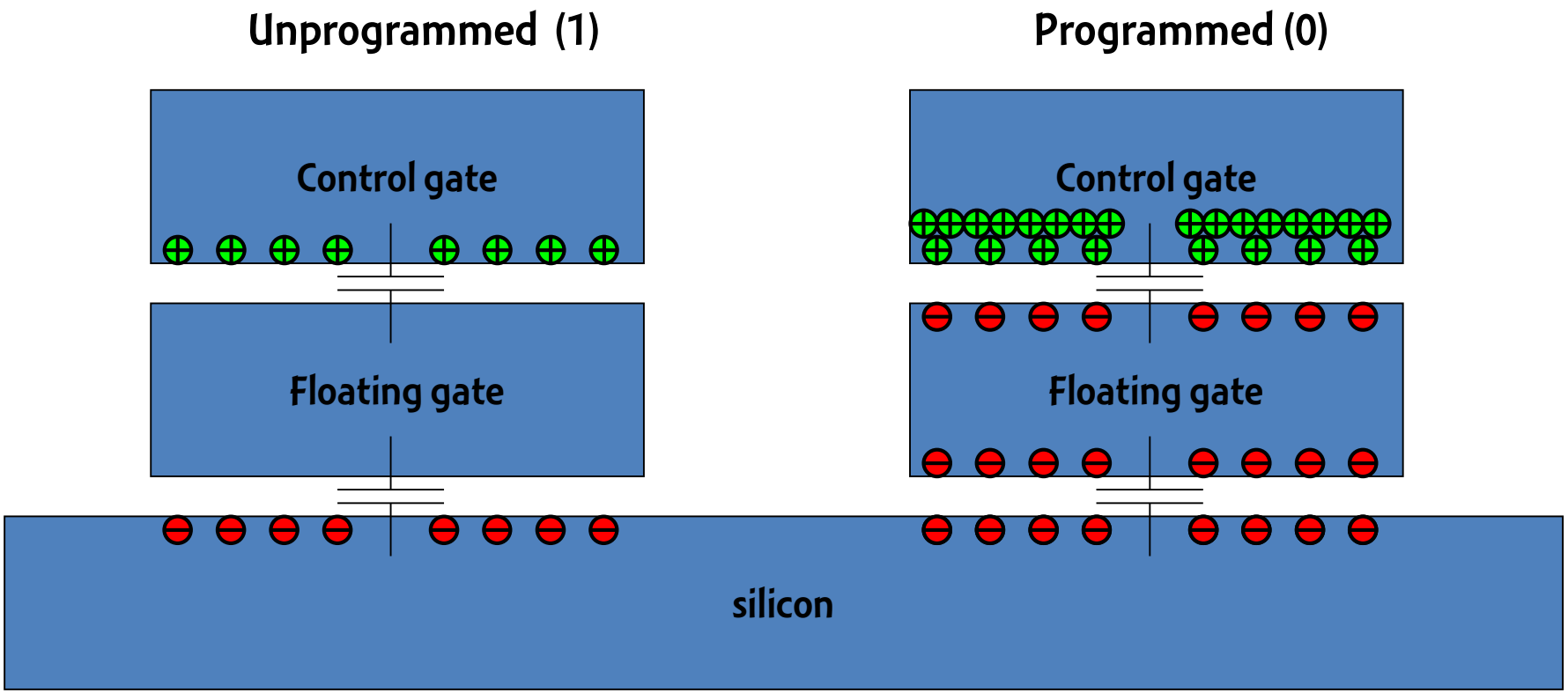
- V_T is shifted by injecting electrons into the floating gate;
- It is shifted back by removing these electrons again.





- CMOS compatible technology!

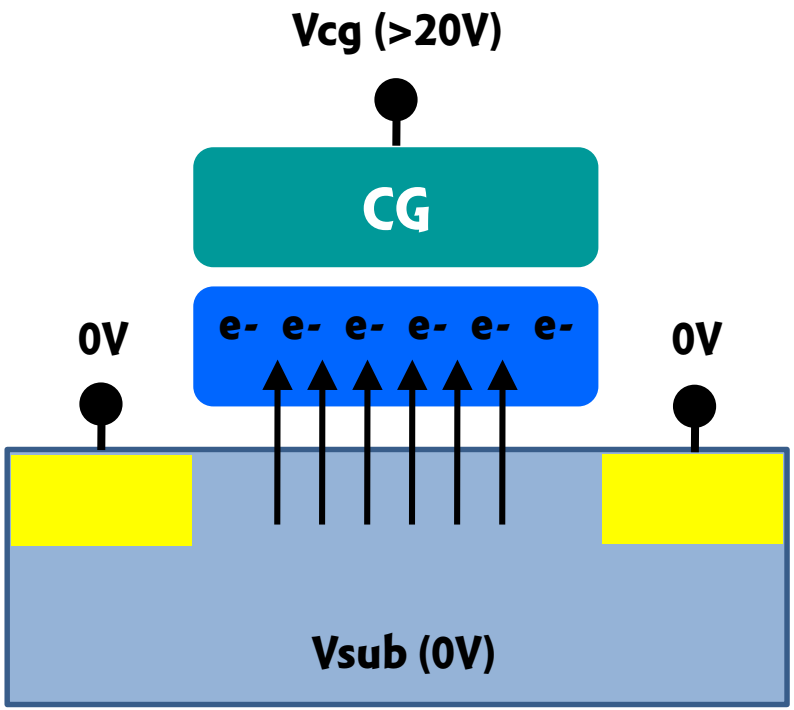
Channel Charge in Floating Gate Transistors



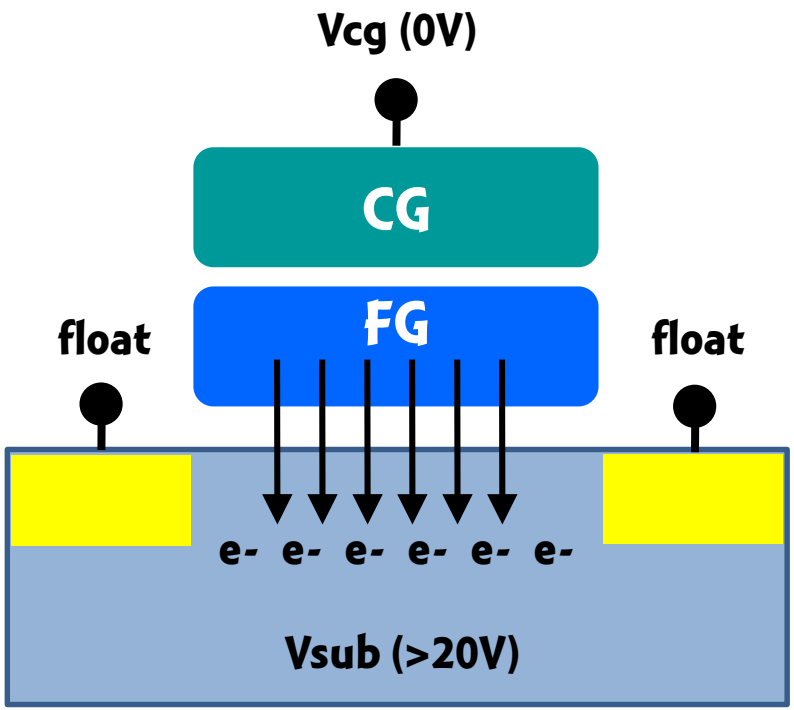
To obtain the same channel charge, the programmed gate needs a higher control-gate voltage than the unprogrammed gate

If we can control the Vth?

- Increase Vth : Program
- Decrease Vth : Erase
- Determine Vth : Read



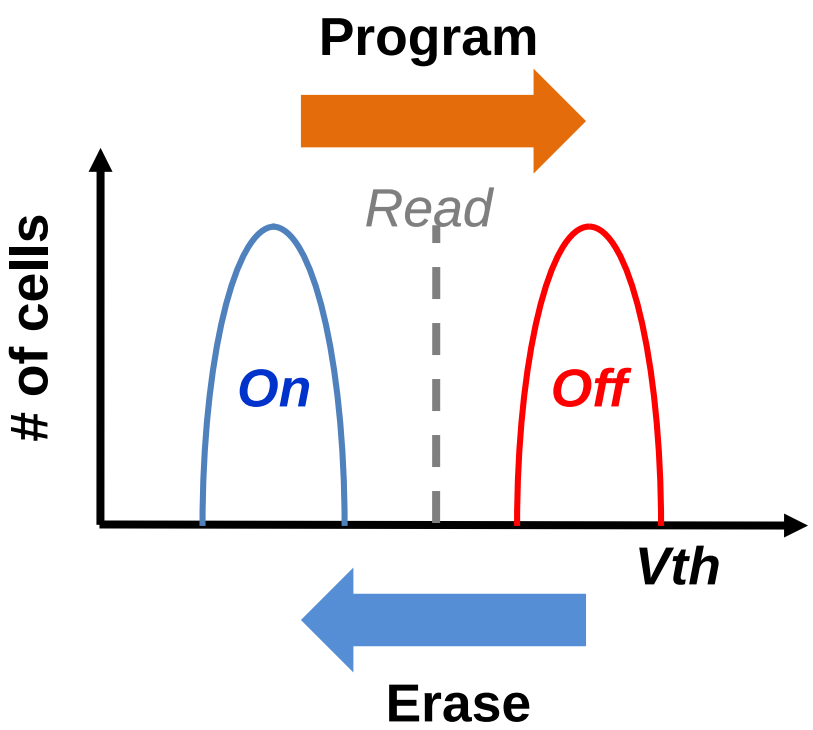
Program : FN tunneling

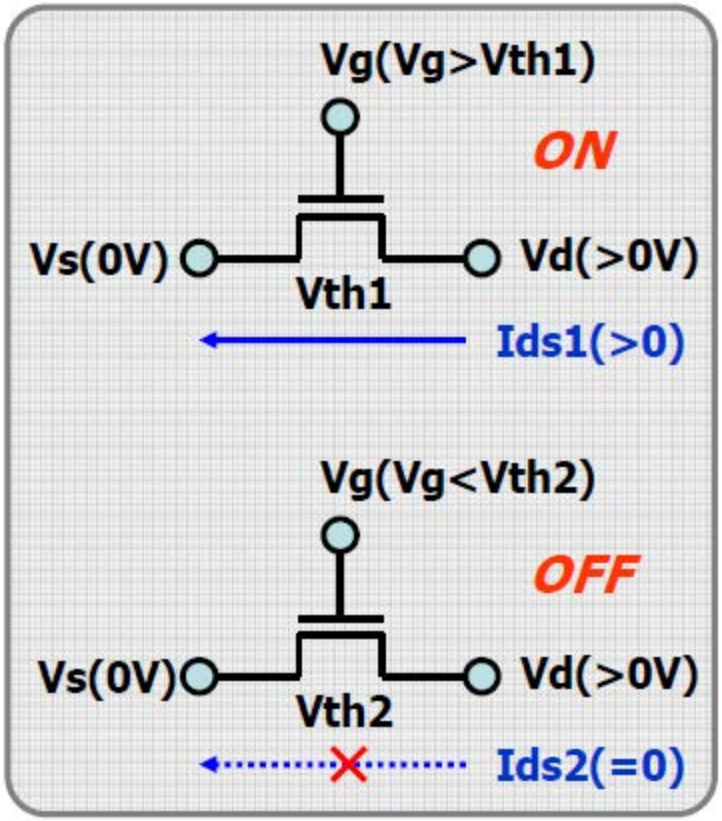


Erase : FN tunneling

NAND Operation – Overview

- **Write & Read binary data to a NAND flash cell**
  - Data "0" → Program → Shift cell Vth to high → Off state → No current flow
  - Data "1" → Erase → Shift cell Vth to low → On State → Current flow

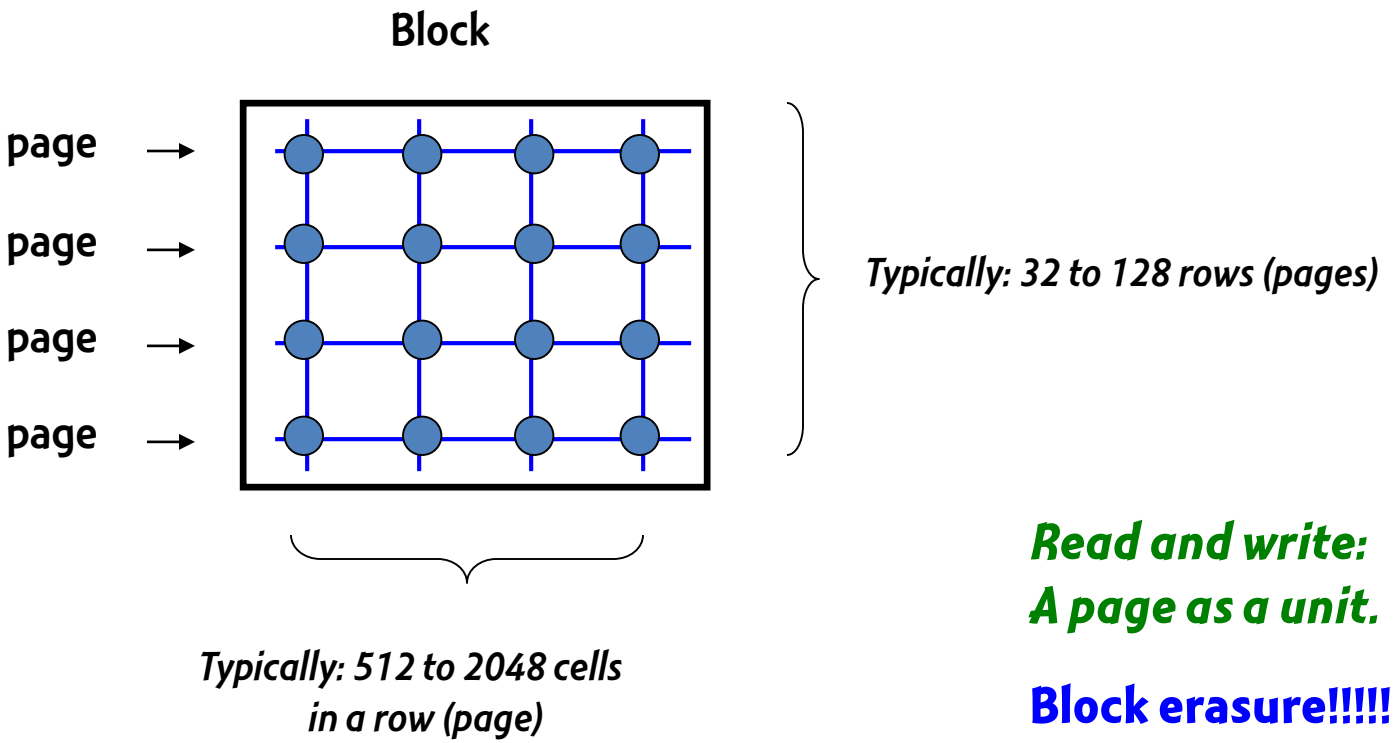




✓ ***Read : Check the current flow***

Layout of Block & Page

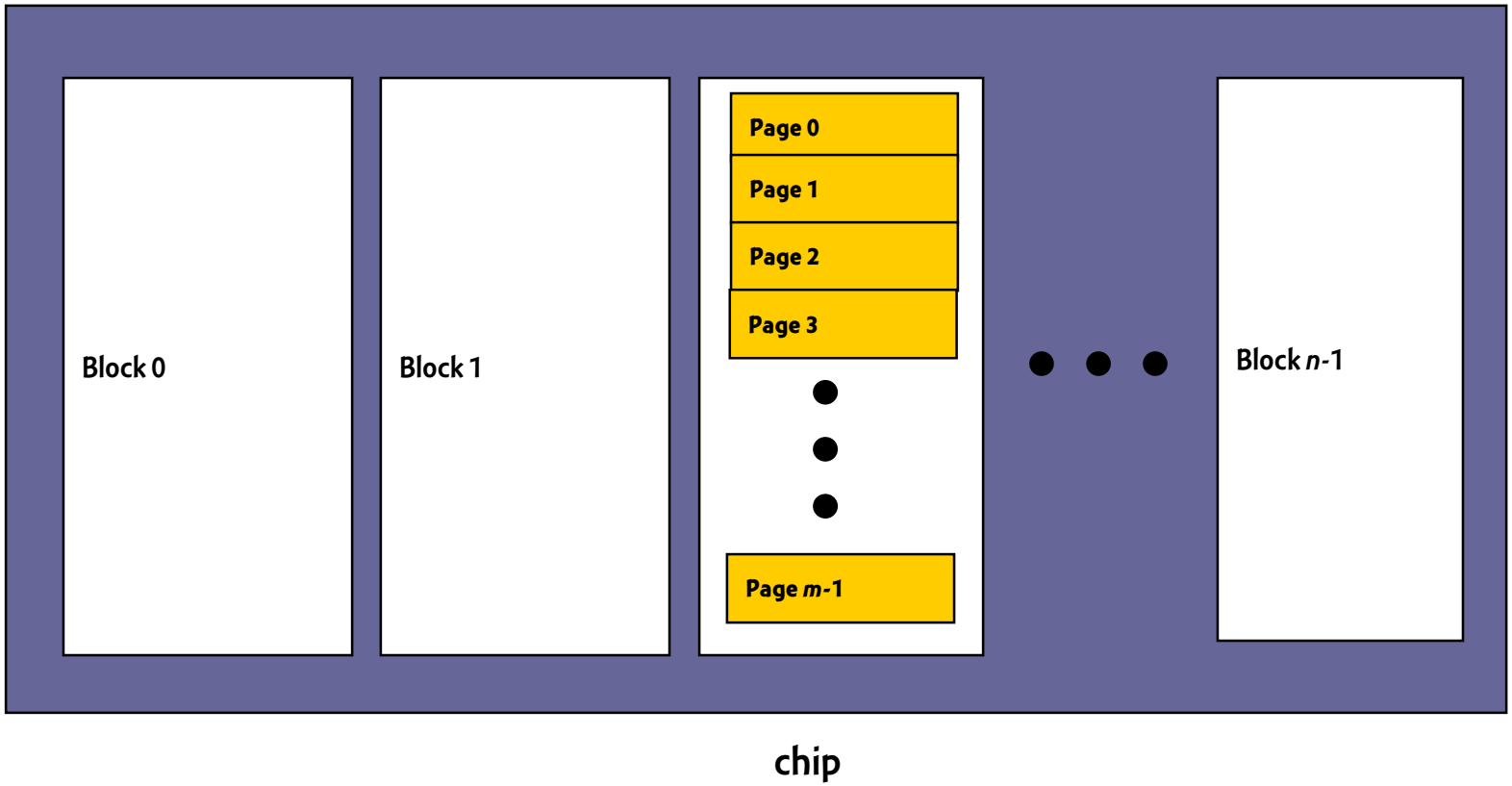
Cells form blocks. Every block is an array.
Every row is a page.



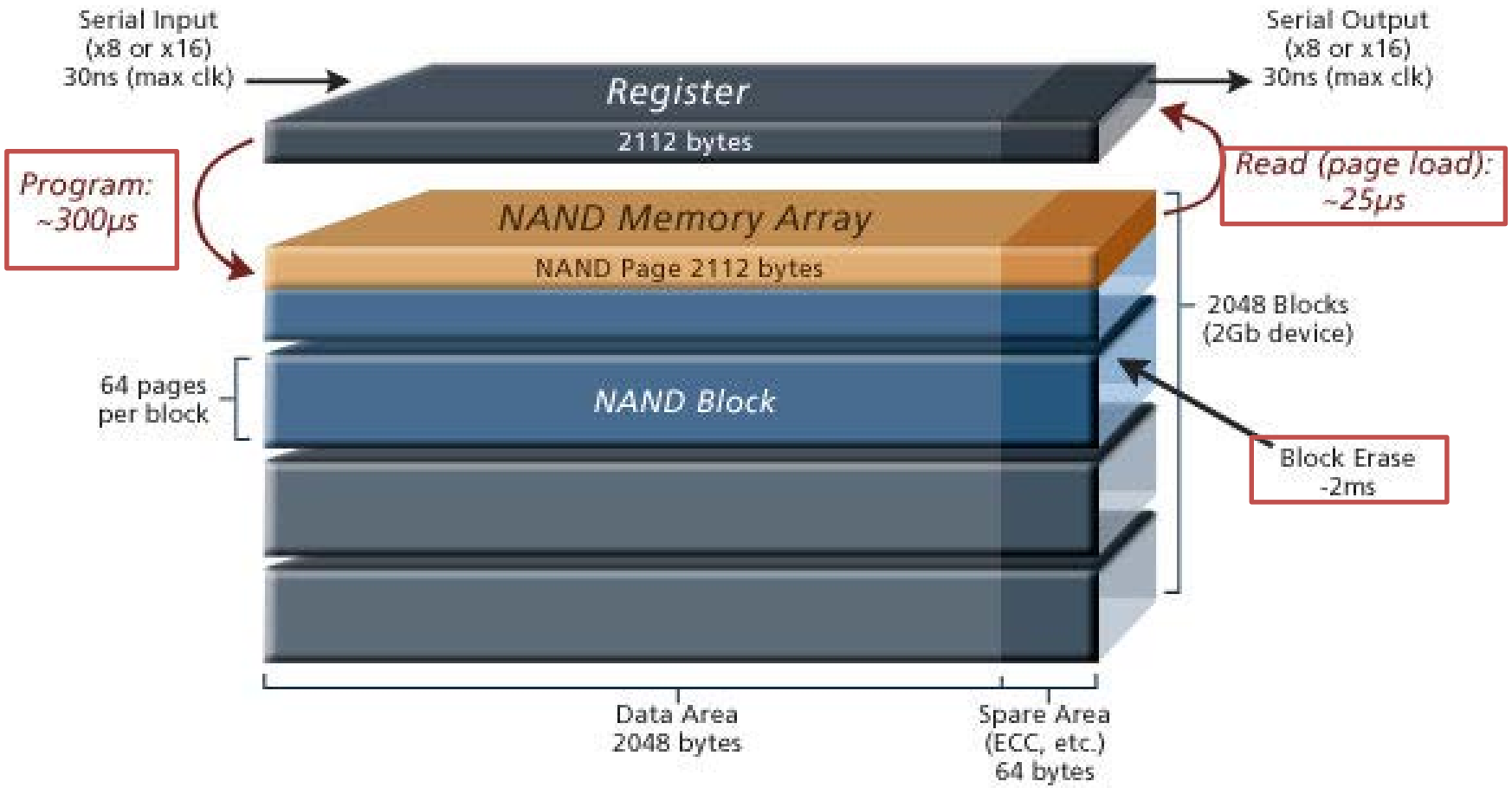
Read and write:
A page as a unit.

Block erasure!!!!!!

Organization of NAND Flash Memory

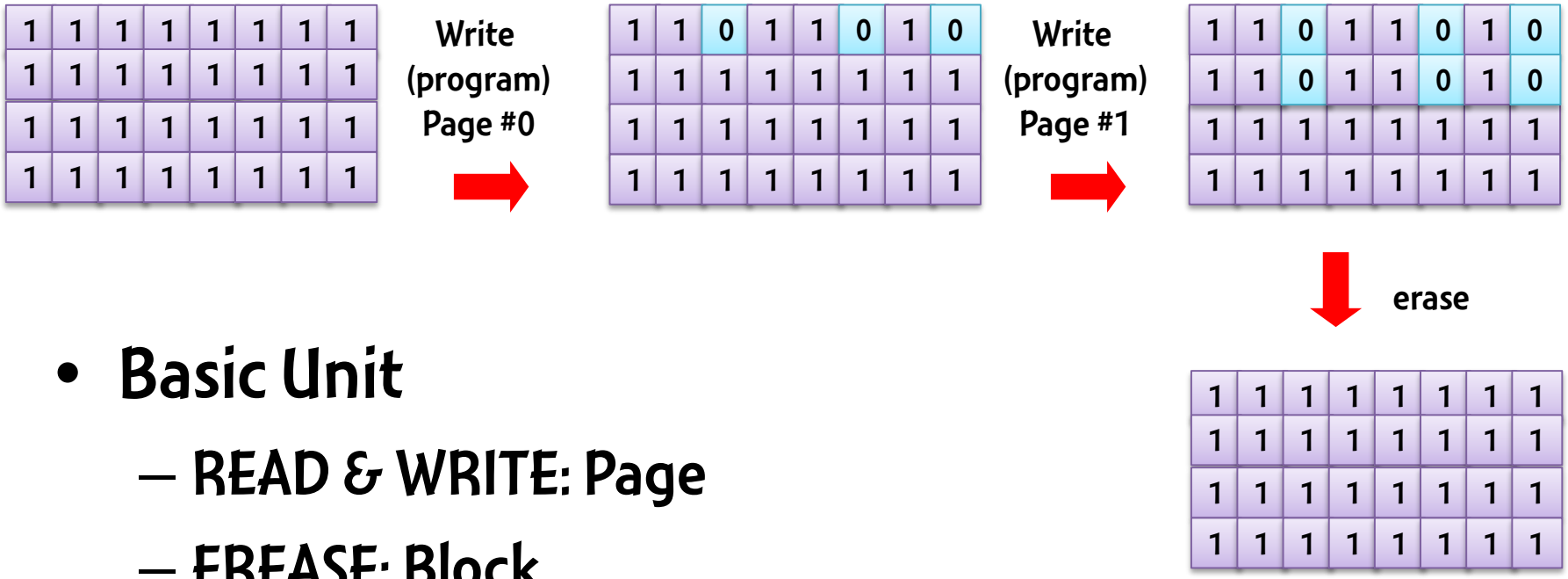


NAND Flash Operations



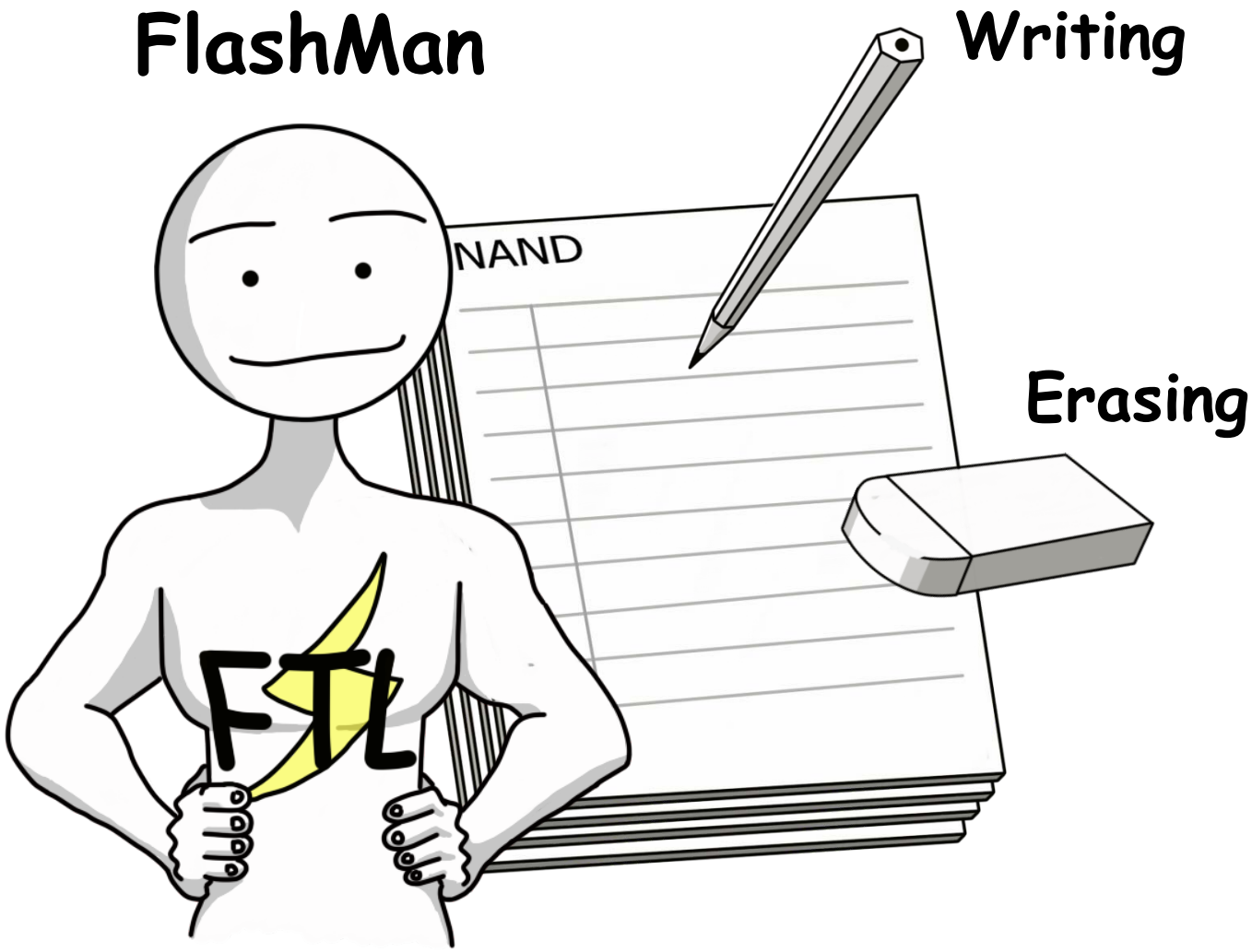
Source: Micron Technology, Inc.

Write & Erase



- Basic Unit
  - READ & WRITE: Page
  - EREASE: Block
- Program changes bit from '1' to '0'

NAND Flash Memory is like Sheets of Paper



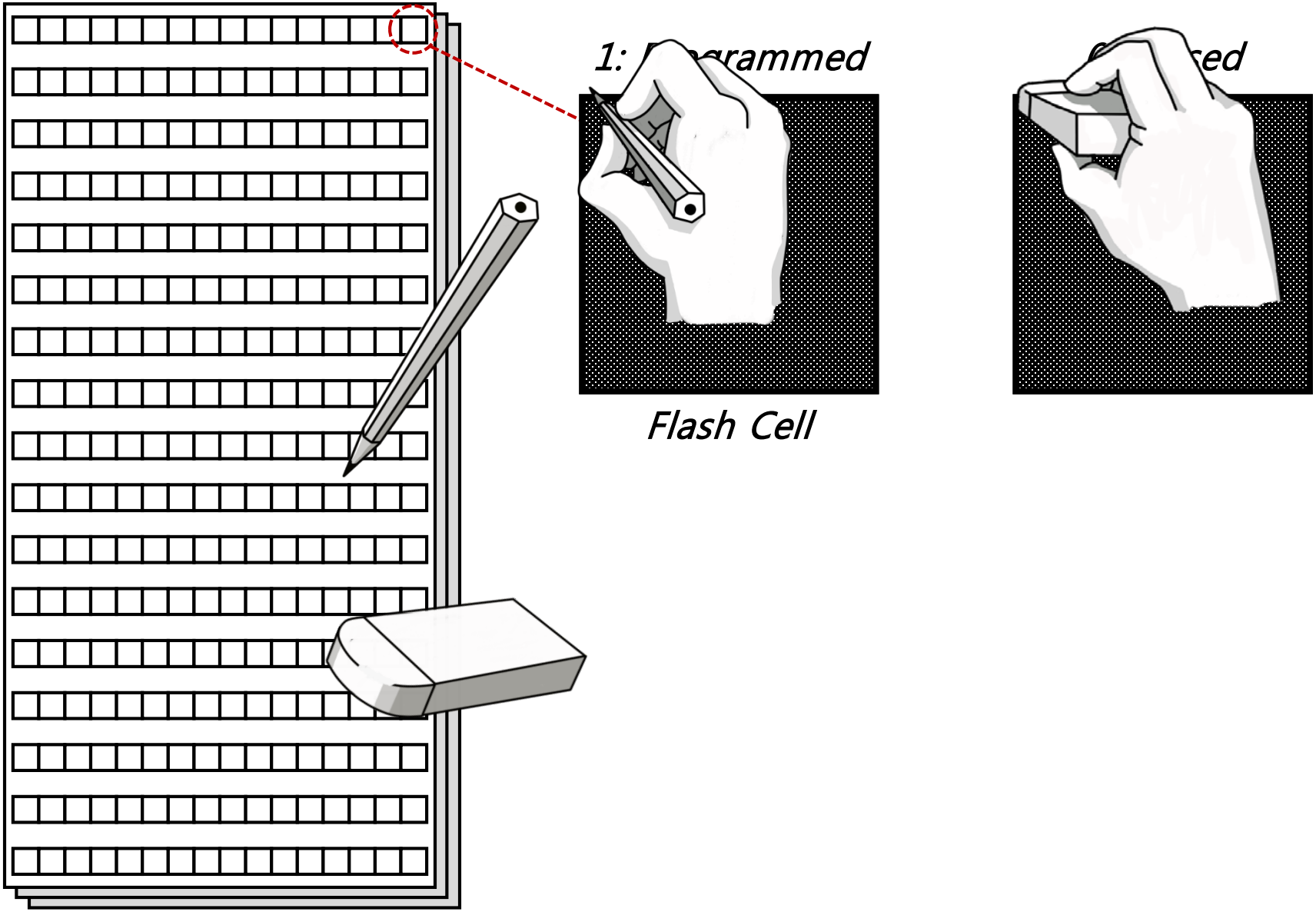
Illustrated by Jisung Park, Seoul National University

Writing Letters and Erasing Paper

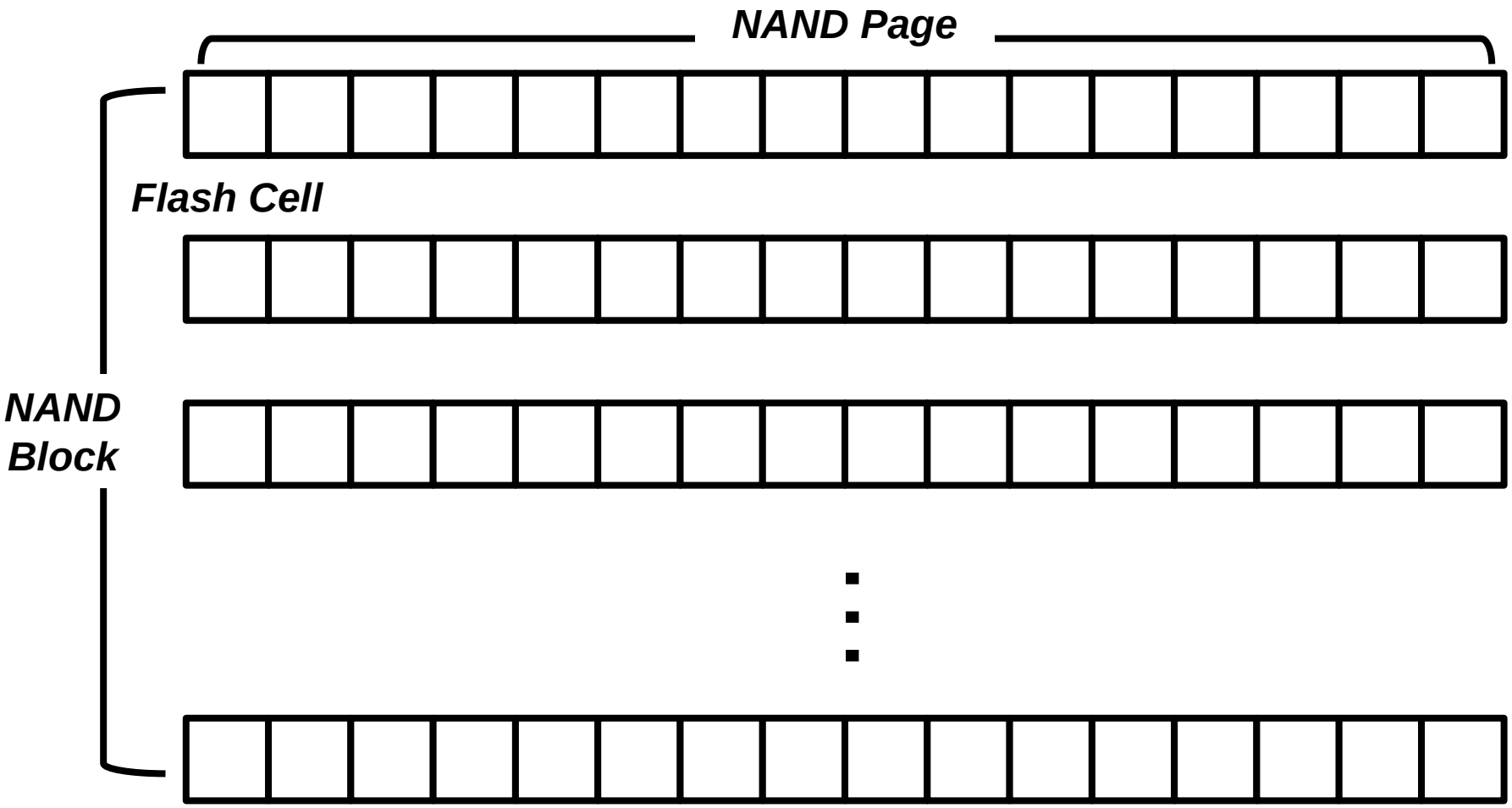
Out-Place Update
Erase before Program

(vs. In-Place Update)

NAND Flash Memory: Grid Paper

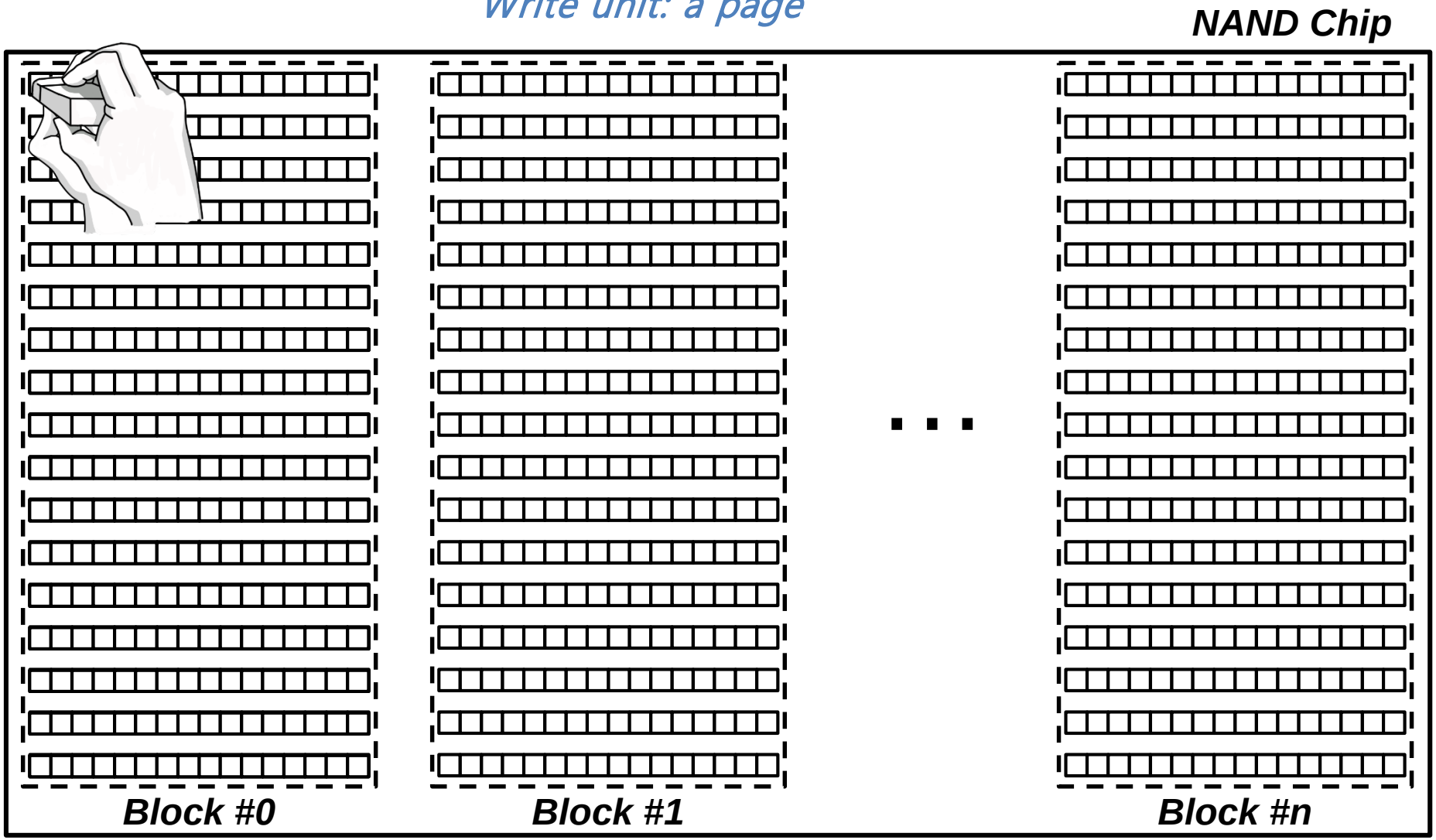


NAND Flash Architecture



NAND Flash Architecture (Cont'd)

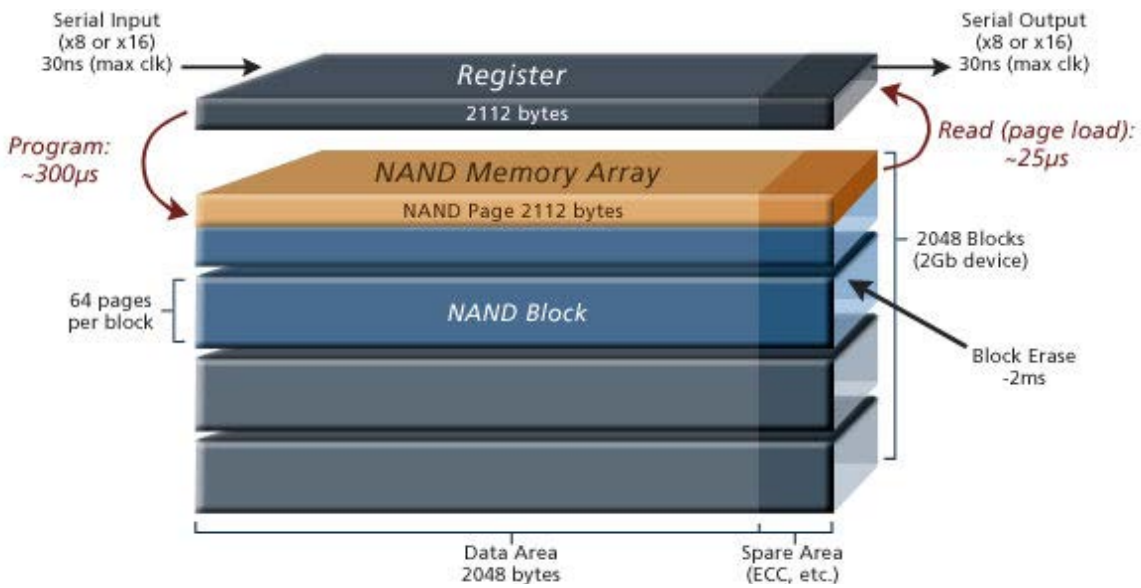
Write unit: a page



Erase unit: a block

Unique Properties of NAND Flash Memory

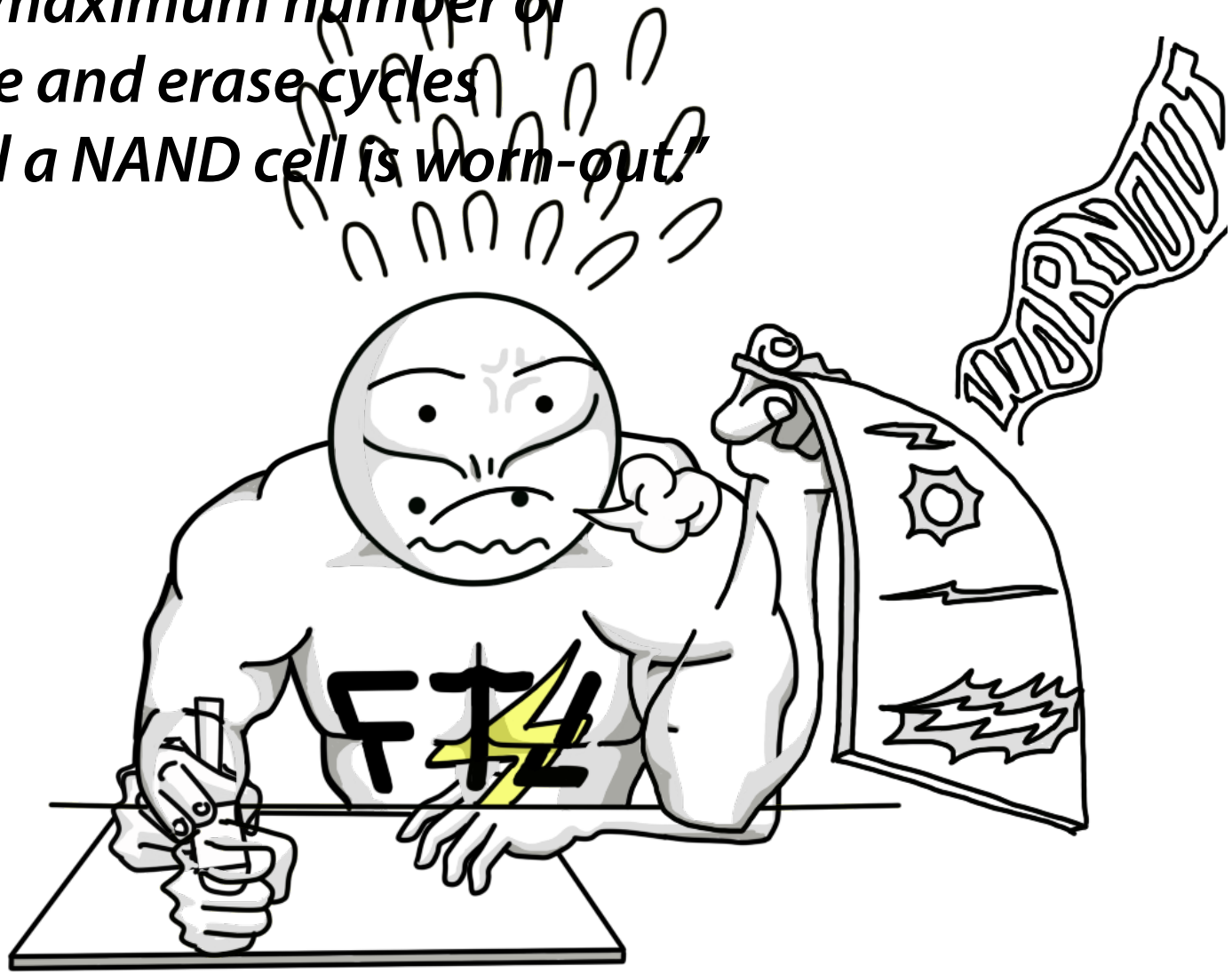
- No support for in-place update
  - Erase-before-program operation
- Asymmetric operation unit/performance
- Limited Endurance
  - A block becomes unreliable after a certain number of program/erase (P/E) cycles



NAND flash memory organization

Limited NAND Endurance

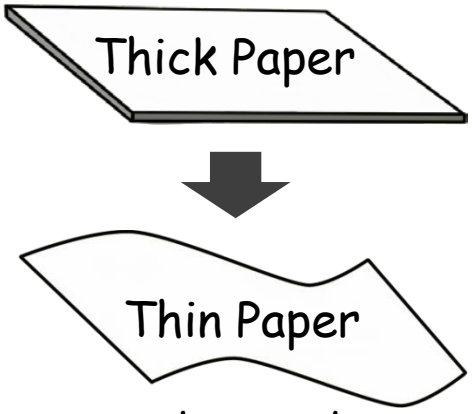
"The maximum number of write and erase cycles until a NAND cell is worn-out."



3000

Why is the NAND Endurance Decreased ?

Old technology



Advanced Semiconductor technology

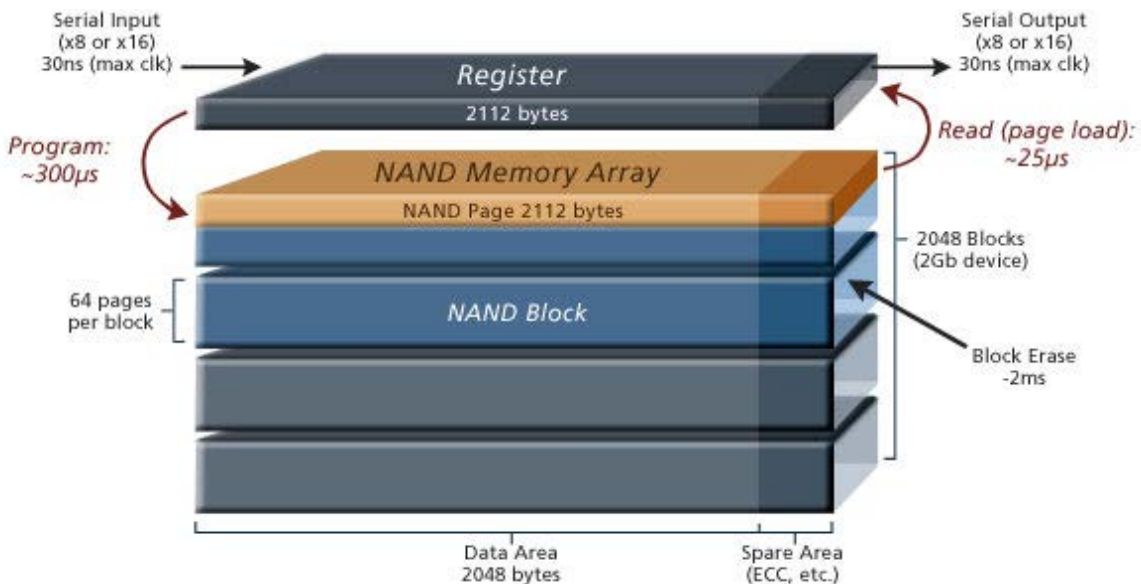
"Because thin paper wears down more easily than thick paper."

2000

Thin paper

Unique Properties of NAND Flash Memory

- Asymmetric operation unit/performance
- No support for in-place update
  - Erase-before-program operation
- Limited Endurance
  - A block becomes unreliable after a certain number of program/erase (P/E) cycles

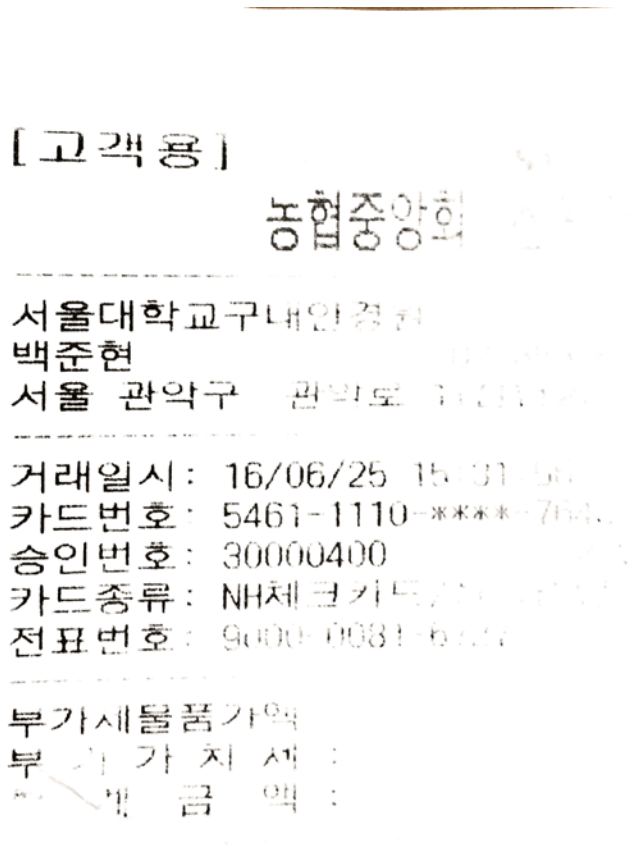


NAND flash memory organization

Like Papers ...

- Data stored in NAND flash FADES!!
  - Not ideal for storage systems
- The data dissipation speed depends on
  - The number of program/erase cycles
  - Exposure to high temperatures
  - The number of reads
  - Process technology
- Solution:
  - move data to a new location by a flash controller

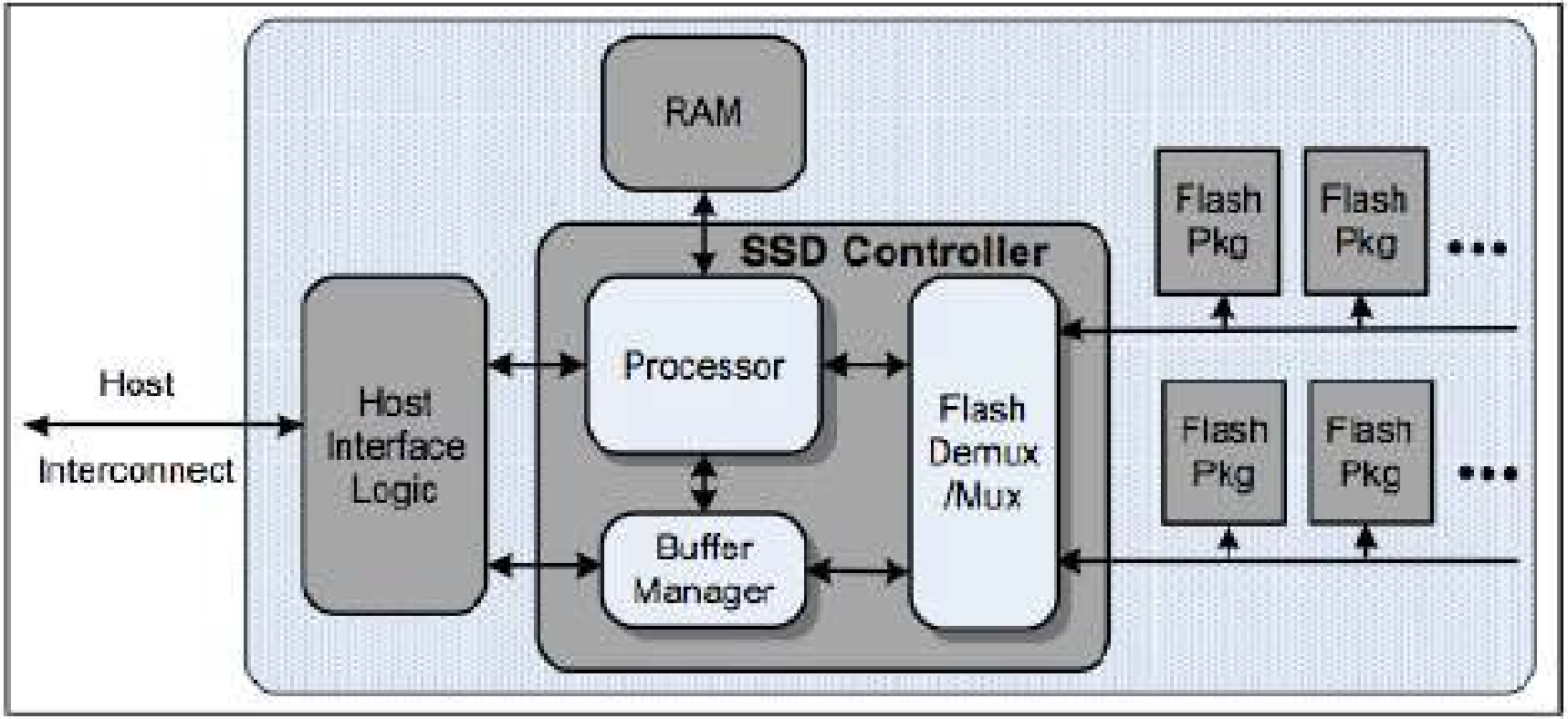
Data Retention



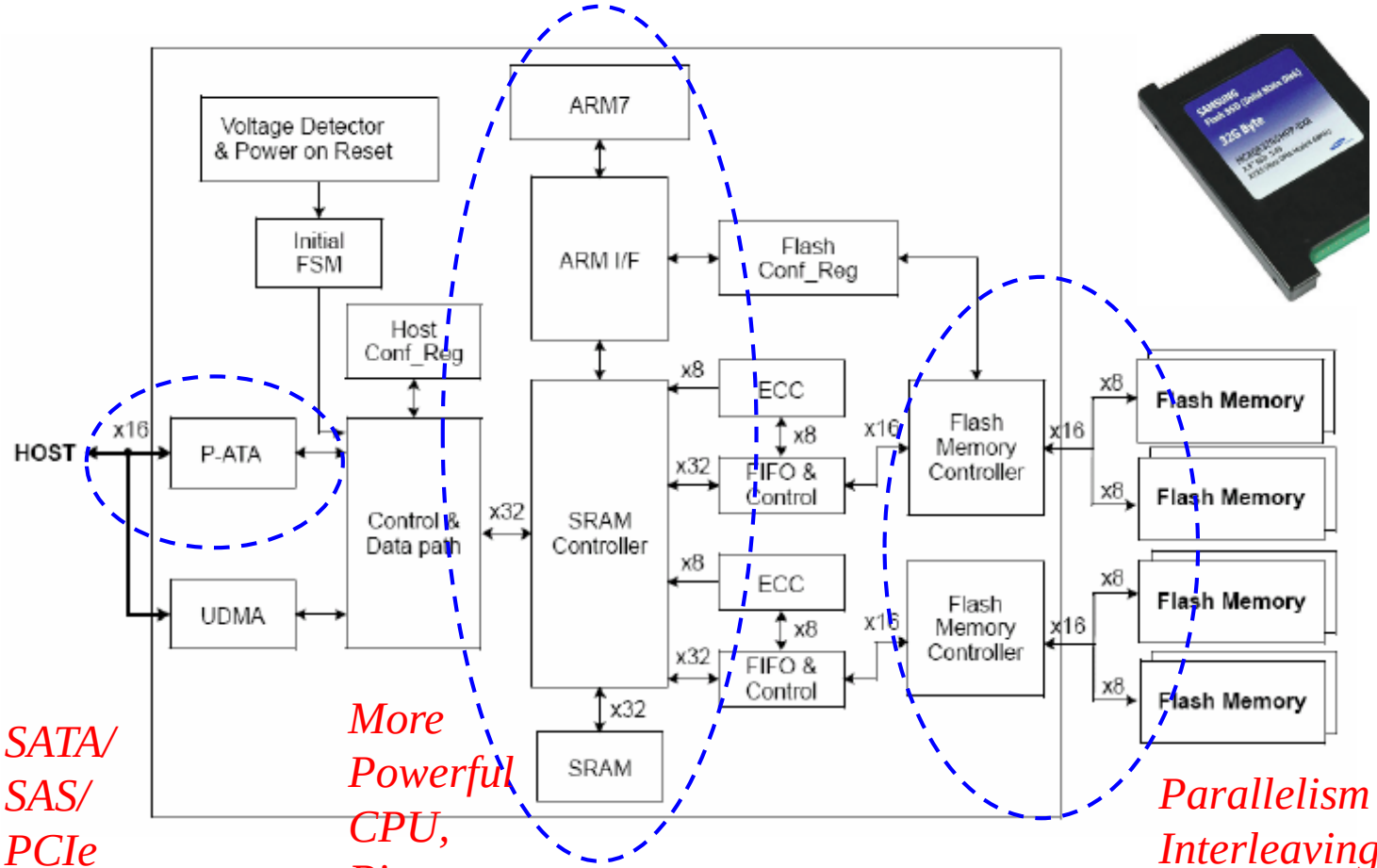
1 Year at 30 ˚C?

Flash Controller
for Performance and Reliability

Overview of Controller



Flash Controller Diagram



SATA/ SAS/ PCIe ….

More Powerful CPU, Bigger SDRAM

Parallelism / Interleaving for Large Bandwidth

Flash Controller

- Make imperfect NAND flash robust and reliable
- Based on a single ASIC die
  - SRAM for firmware
  - DRAM for caching/buffering
  - Backup power system (e.g., batteries, capacitors) for sudden power off
- Parallelism support for high performance
- Various reliability functions for flash memory

Error Correction

- For all reads and writes from/to flash, error-correcting code is used.

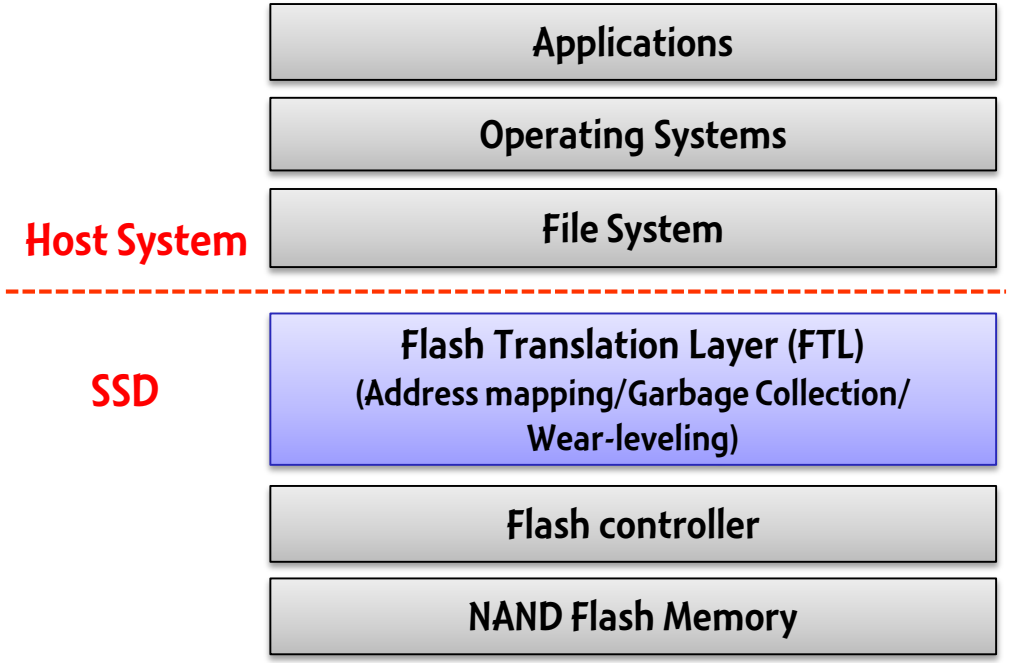
- BCH
- LDPC
- XOR/Scramblers

SSD Software

- Overcome the physical restrictions by employing system software called a flash translation layer (FTL)
  - Asymmetric operation unit/performance
  - No support for in-place update
  - Limited Endurance

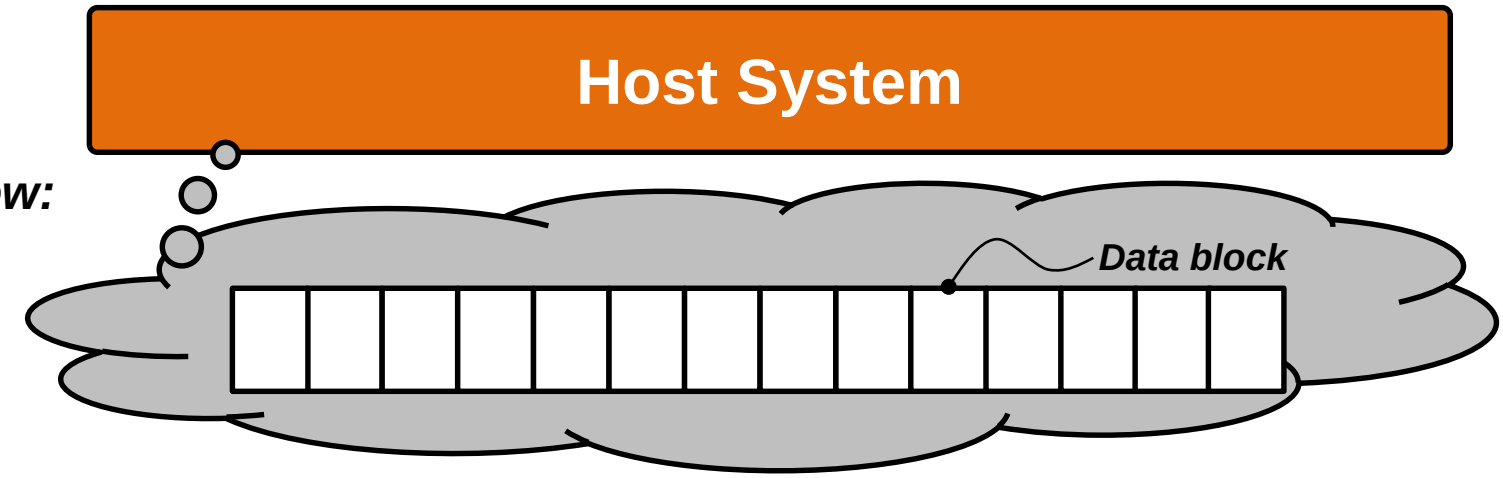
Address mapping/ Garbage collection

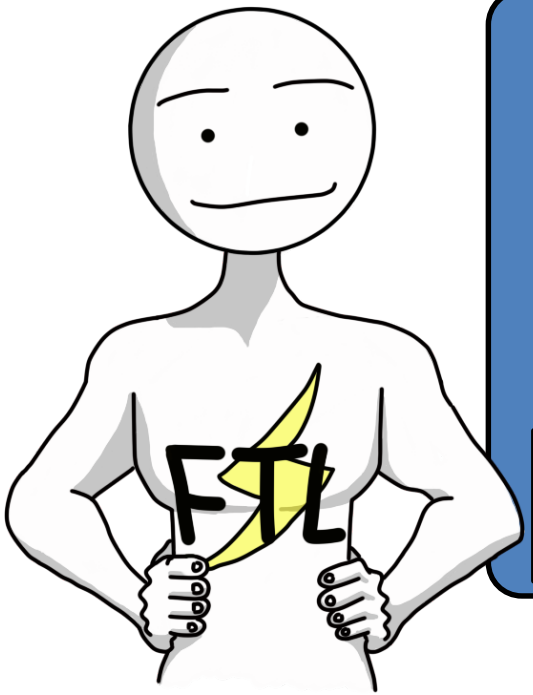
Wear-leveling

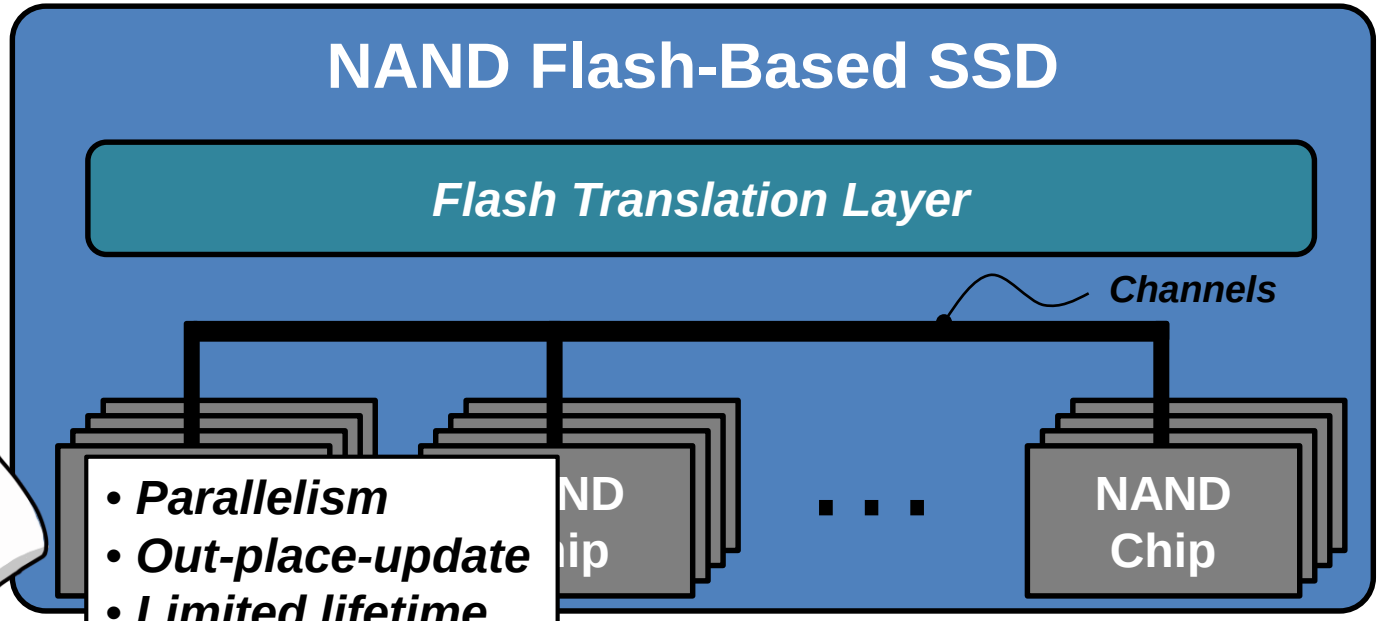


Flash Translation Layer (FTL)

Host system view:
A flat storage



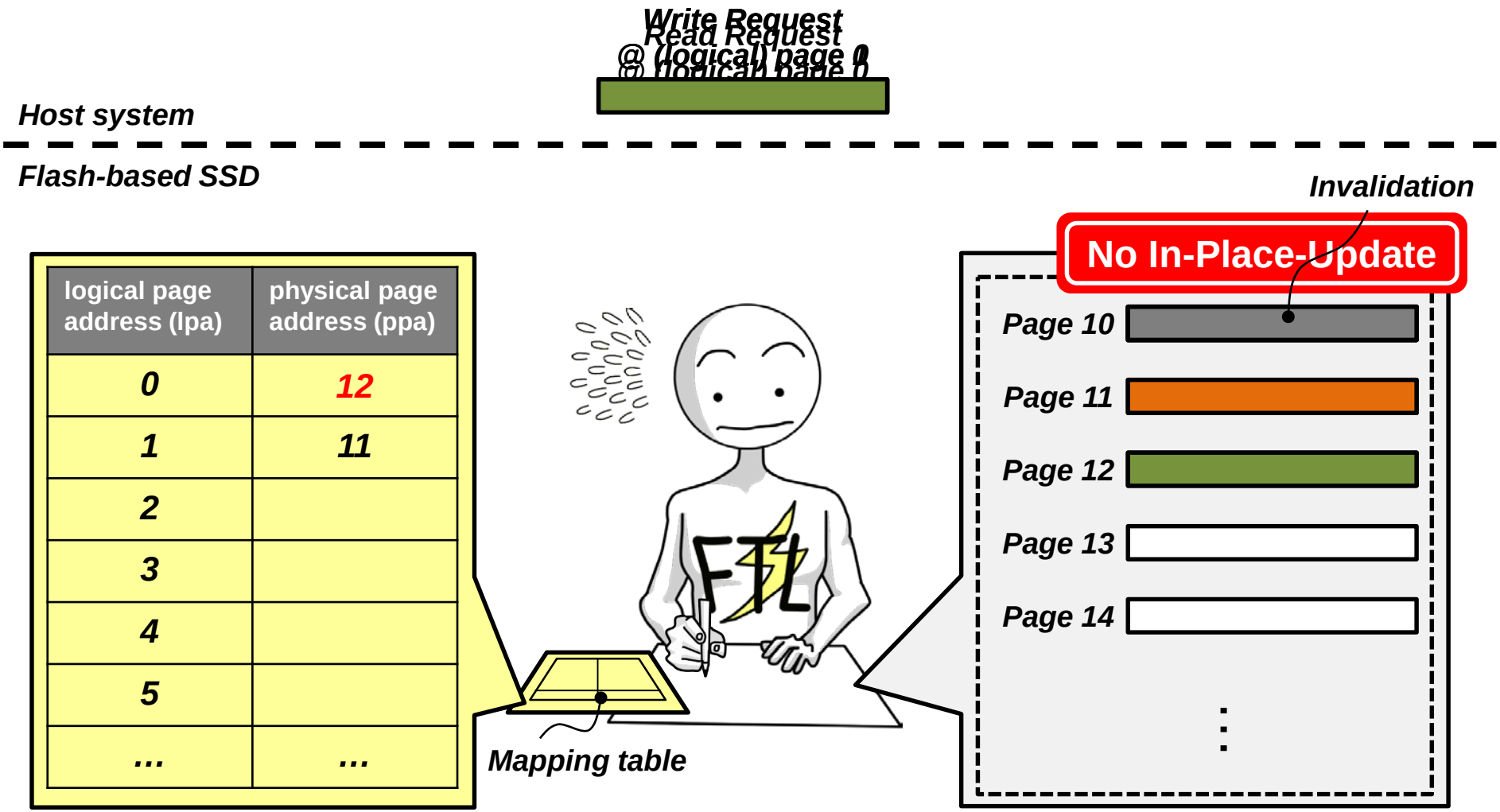




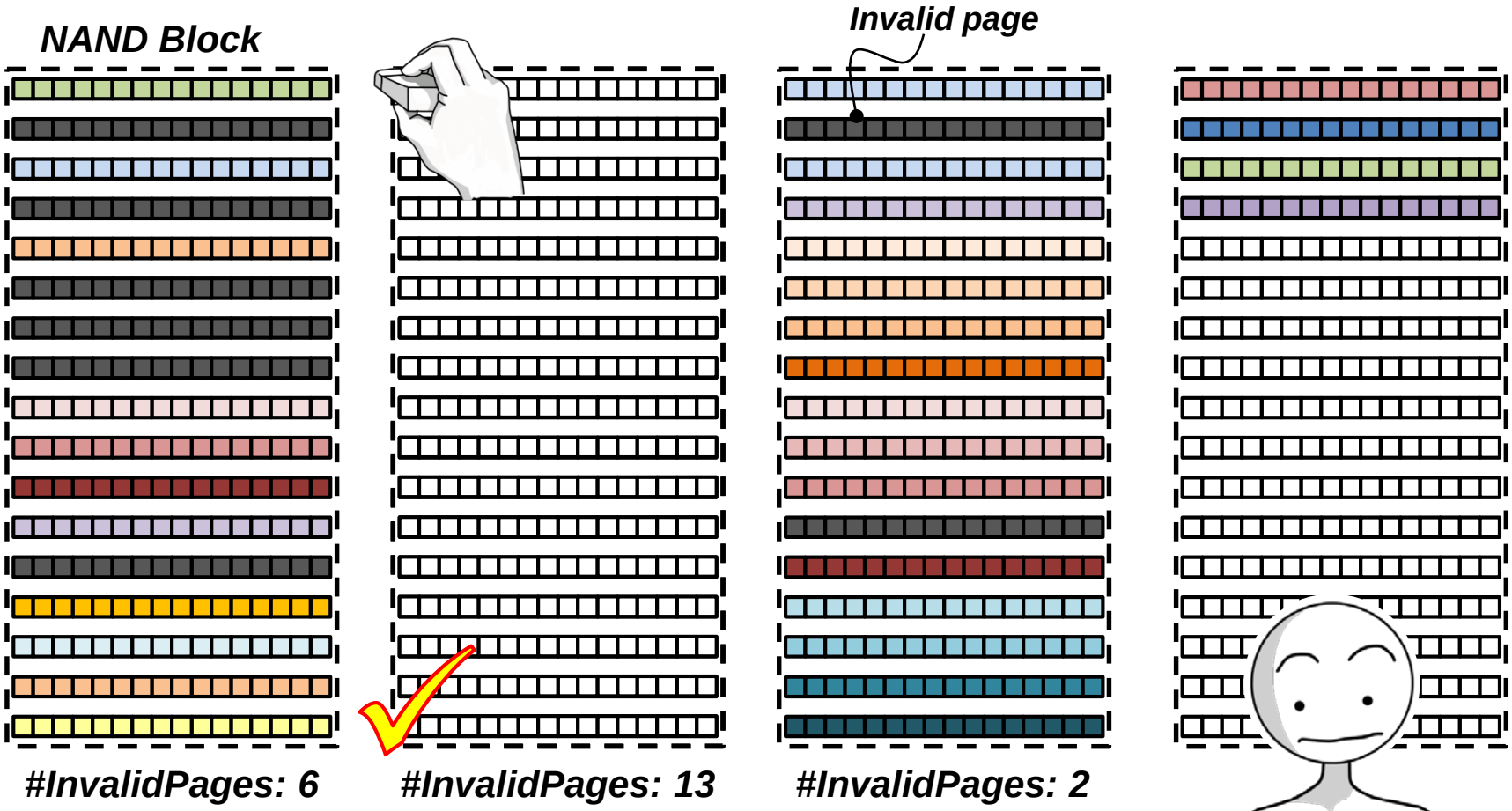
- *Parallelism*
- *Out-place-update*
- *Limited lifetime*

...

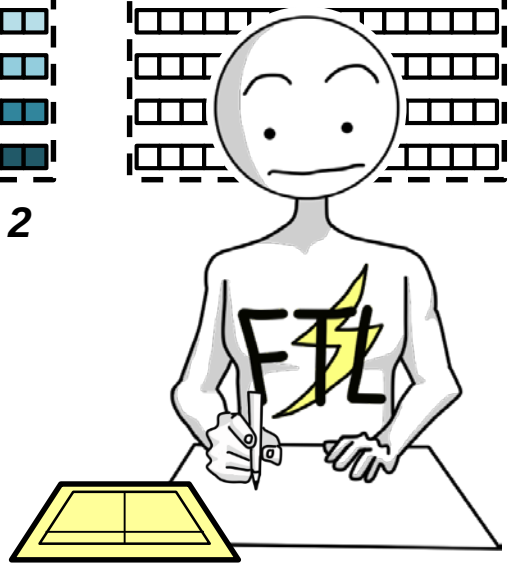
Address Translation



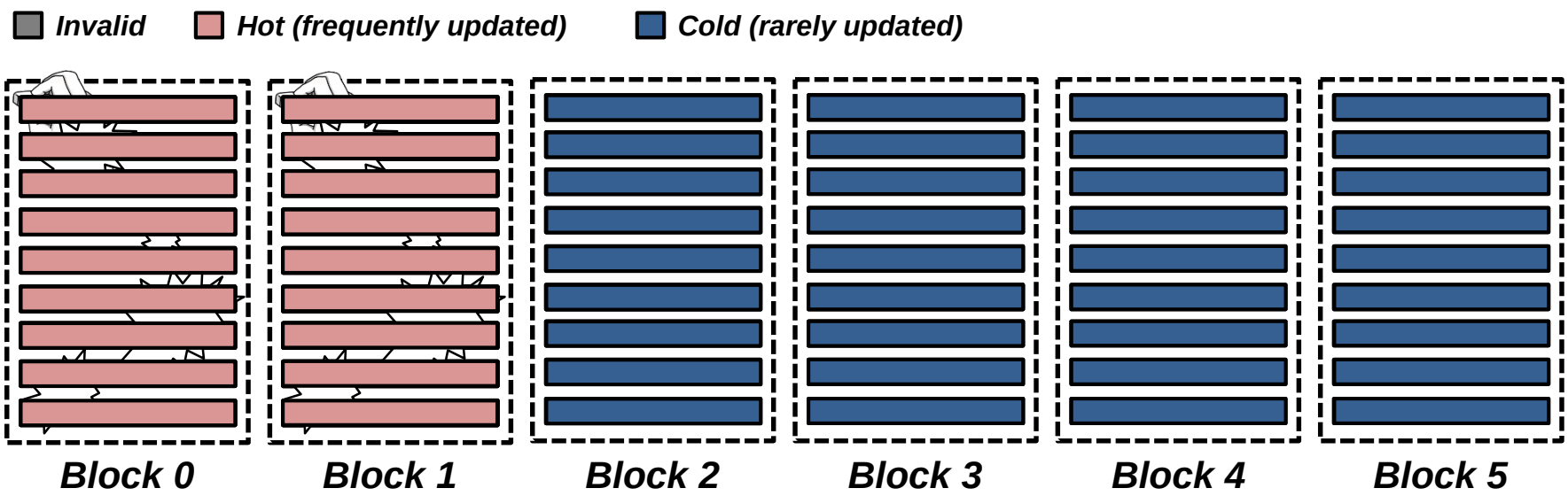
Garbage Collection



1) *Choose a victim block*
2) *Copy valid pages and update mappings*
3) *Erase the victim block*



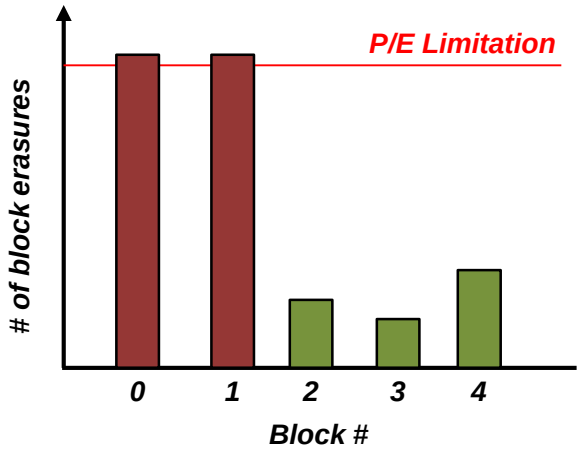
Wear Leveling

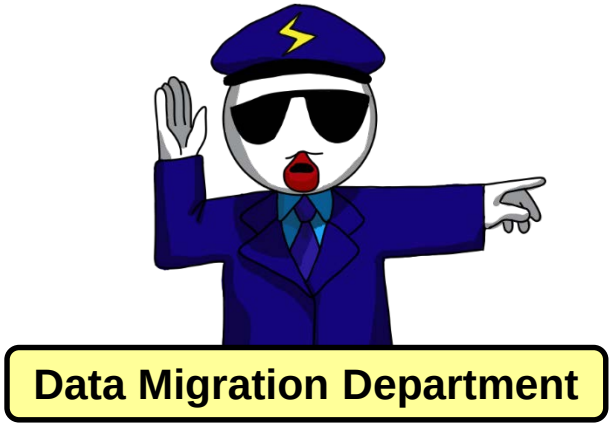


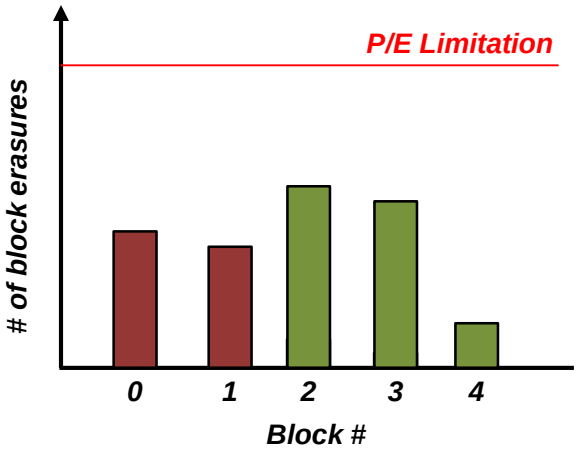
Skewed block usage!

Short device lifetime!

Hot-cold swapping





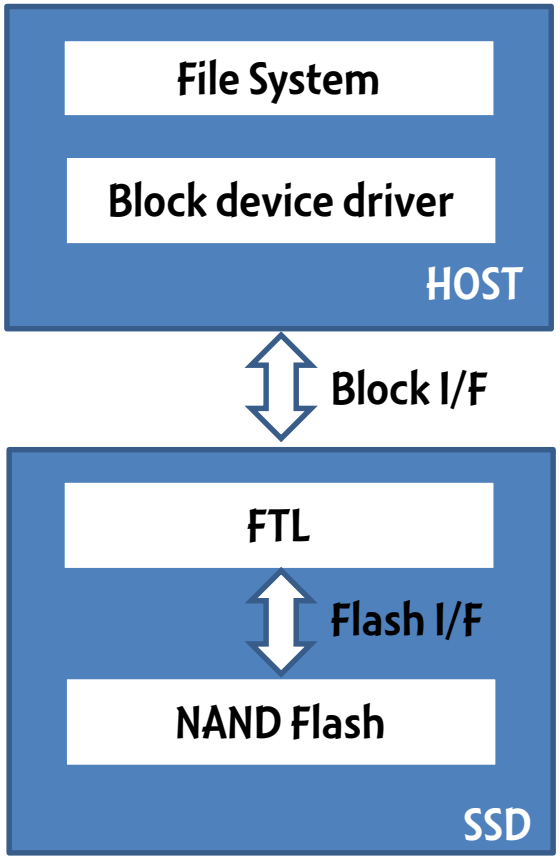


Controller Firmware

- Flash Translation Layer (FTL)
  - Mapping blocks
  - Garbage Collection
  - Wear Leveling

Flash Translation Layer

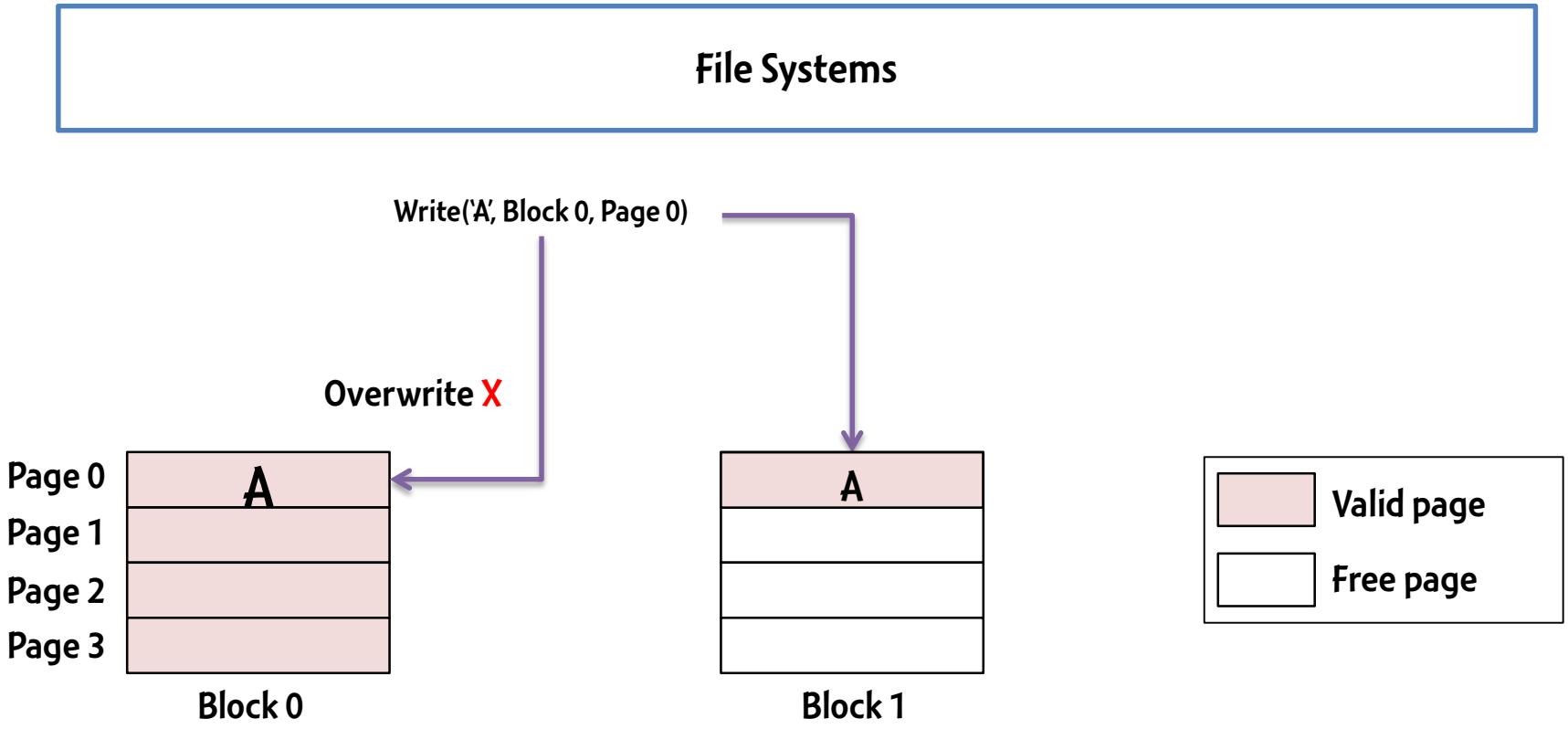
- A software layer to make NAND flash emulate traditional block devices (or disks)



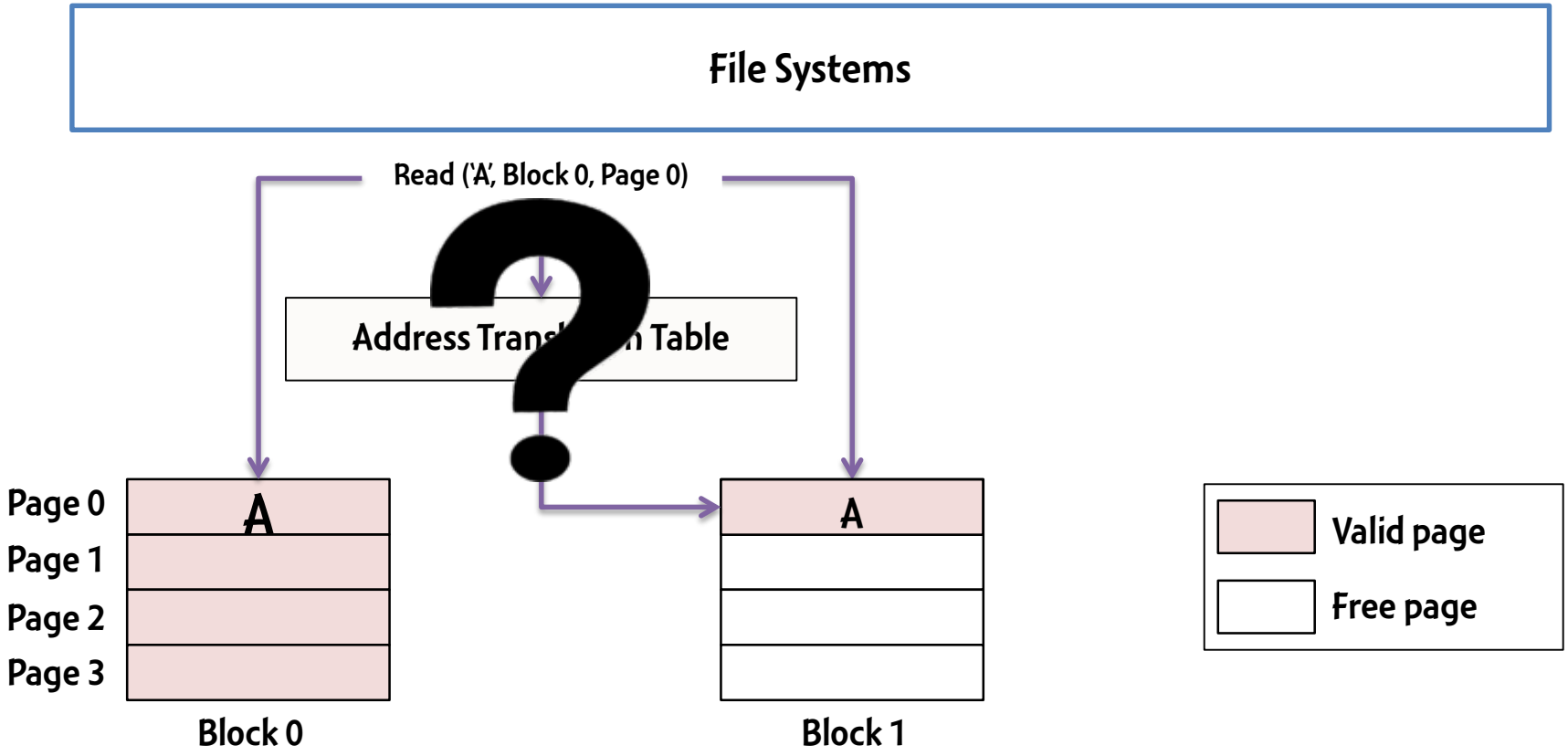
Flash Management Tasks

- Essential
  - Address Translation
    - Avoid in-place update
    - Logical Block Address (LBA) -> Physical Block Address (PBA)
  - Garbage Collection
    - Reclaim invalid blocks -> Get new free blocks
- Optimization
  - Wear Leveling
    - Erase all blocks evenly -> Extend life time

Out-place Update

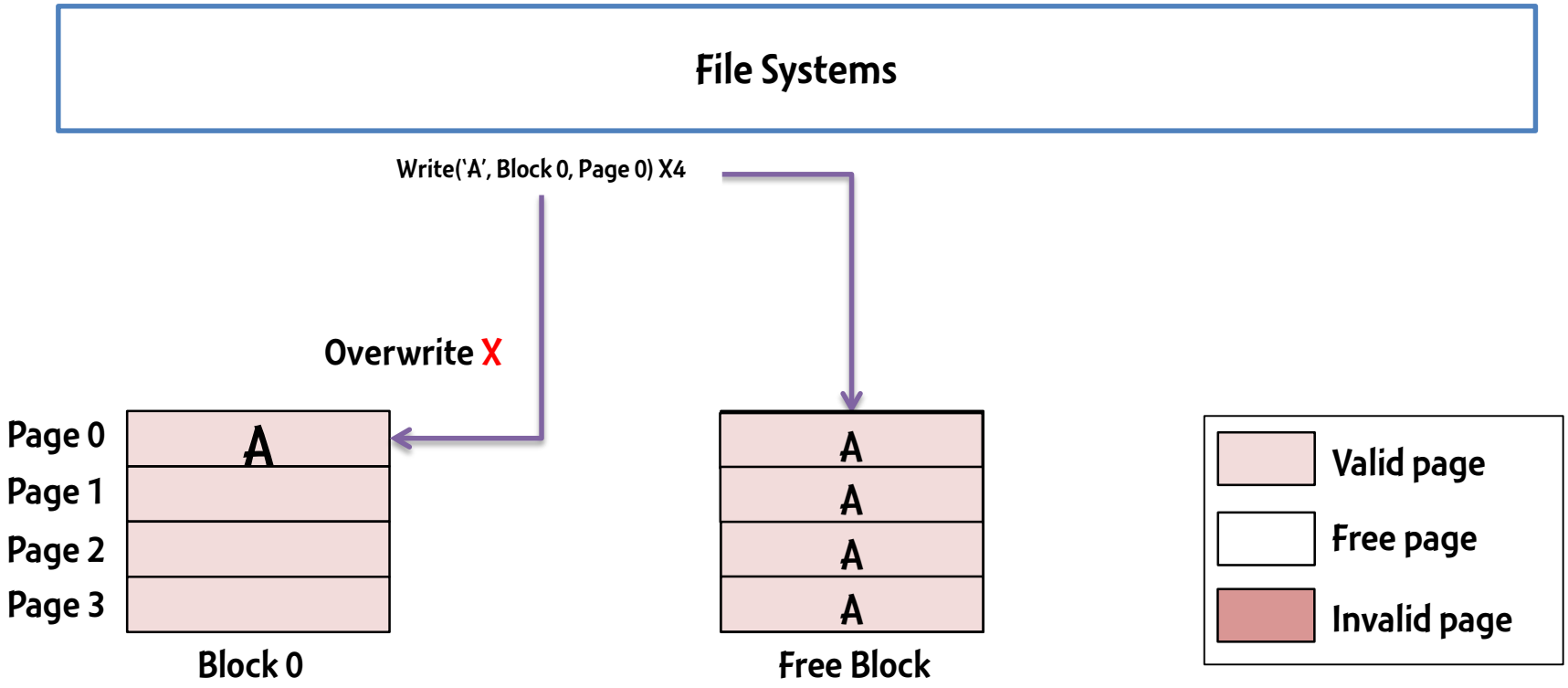


Address Translation



Garbage Collection

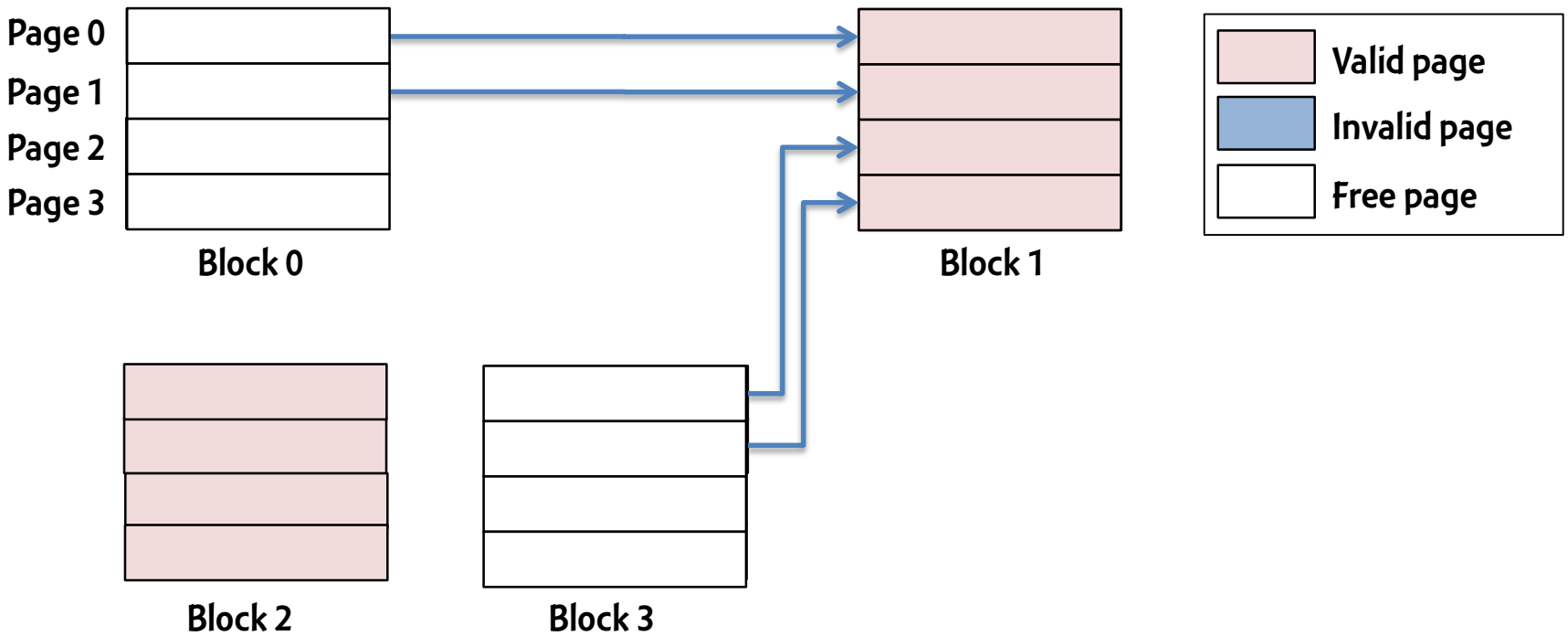
- NAND flash memory does not allow *in-place update*



- NAND flash memory will be full of invalid data…

Garbage Collection

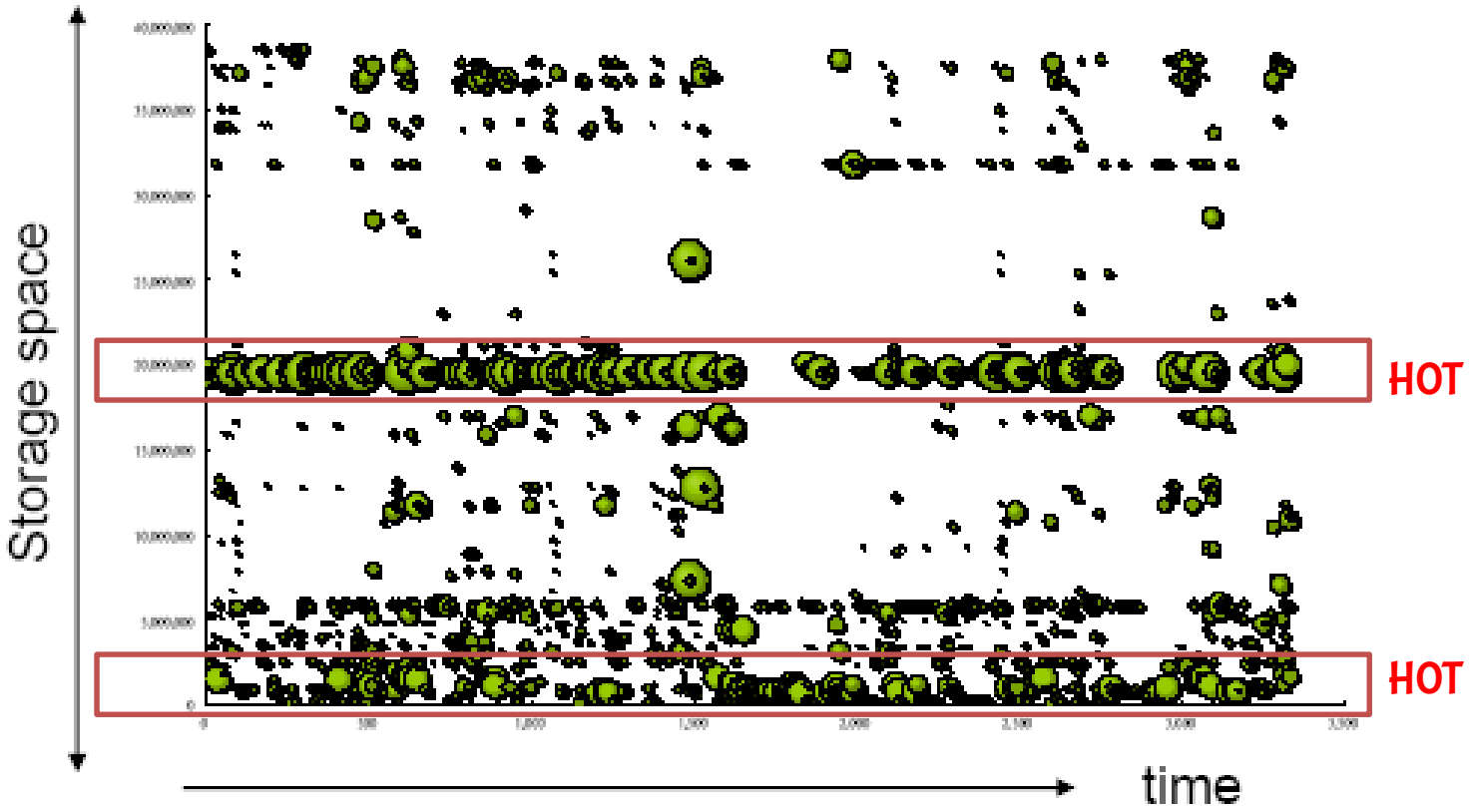
- *Gather valid data -> Erase invalid data*



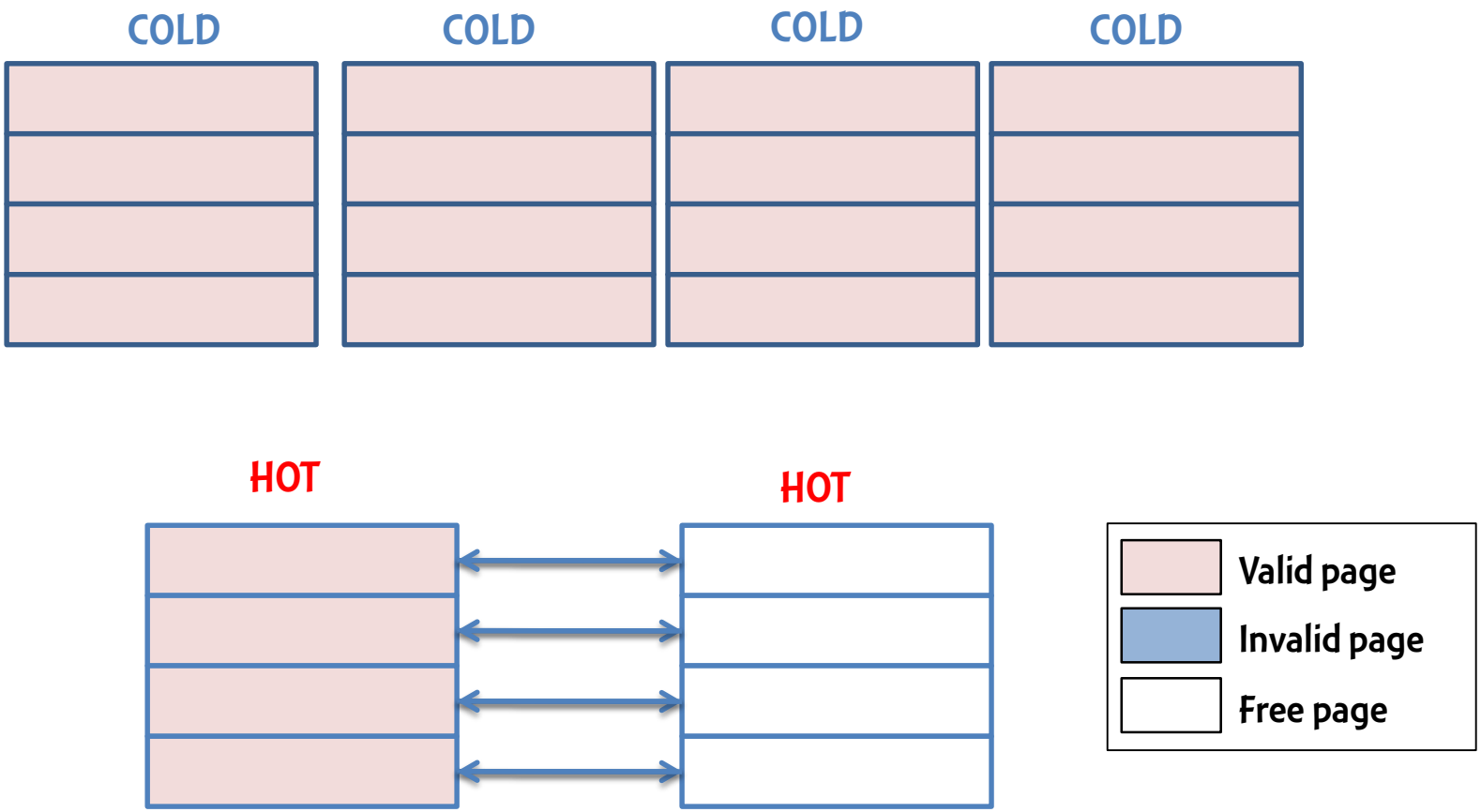
Flash Management Tasks

- Essential
  - Address Translation
    - Avoid in-place update
    - Logical Block Address (LBA) -> Physical Block Address (PBA)
  - Garbage Collection
    - Reclaim invalid blocks -> Get new free blocks
- Optimization
  - Wear Leveling
    - Erase all blocks evenly -> Extend life time

Spatial Locality of Write Request



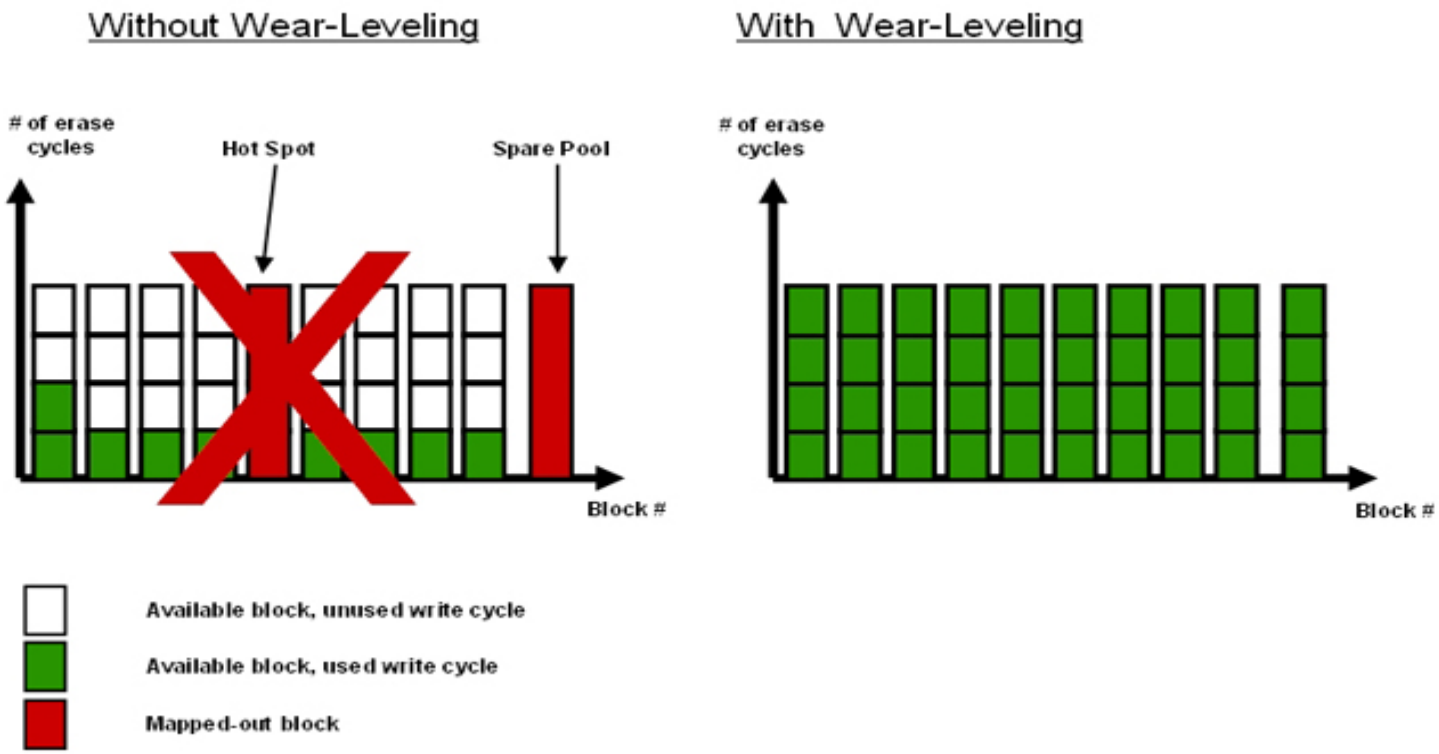
Wearing of Flash Memory Blocks



- Blocks with hot data are likely to be erased more

Wear Leveling

- Erase evenly all blocks



New Interface for SSDs

- TRIM/UNMAP
  - Let the SSD clear the LBA entries in the FTL, giving more fee space to use
- Scatter Gather
  - Reduce command overhead of one command at a time
  - Gathers multiple noncontiguous requests into a single command, reducing overhead

Host Interface

Existing Interface: SATA, SAS

New interface: PCIe