



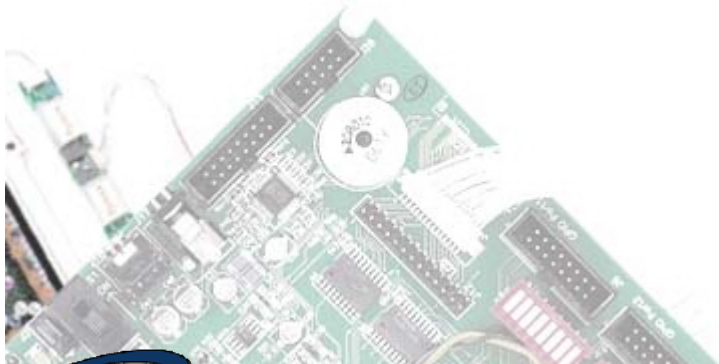
Implementation of High Performance Flash Memory Based Storage

```
...
}
/* block erase */
for (current_block = 0; current_block < NO_OF_BLOCK; current_block++) {
    FH_Erase(current_block);
}
...
}
...
```

2007. 3. 8

Jin Hyuk Yoon

Seoul National University

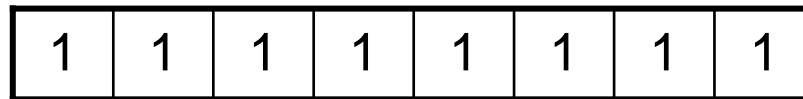


Agenda

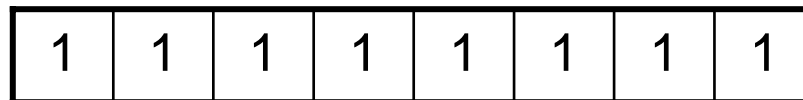
- Flash memory basics
- Performance of flash memory based storage
- Hydra: prototype SSD implementation
- Conclusions

Flash memory

- What is flash memory?
 - Partially erasable/programmable EEPROM
 - Erase: set cell (bit) to 1
 - Set all bits to 1



- Program: change cell (bit) from 1 to 0
 - Selectively change 1 bits to 0 to write value

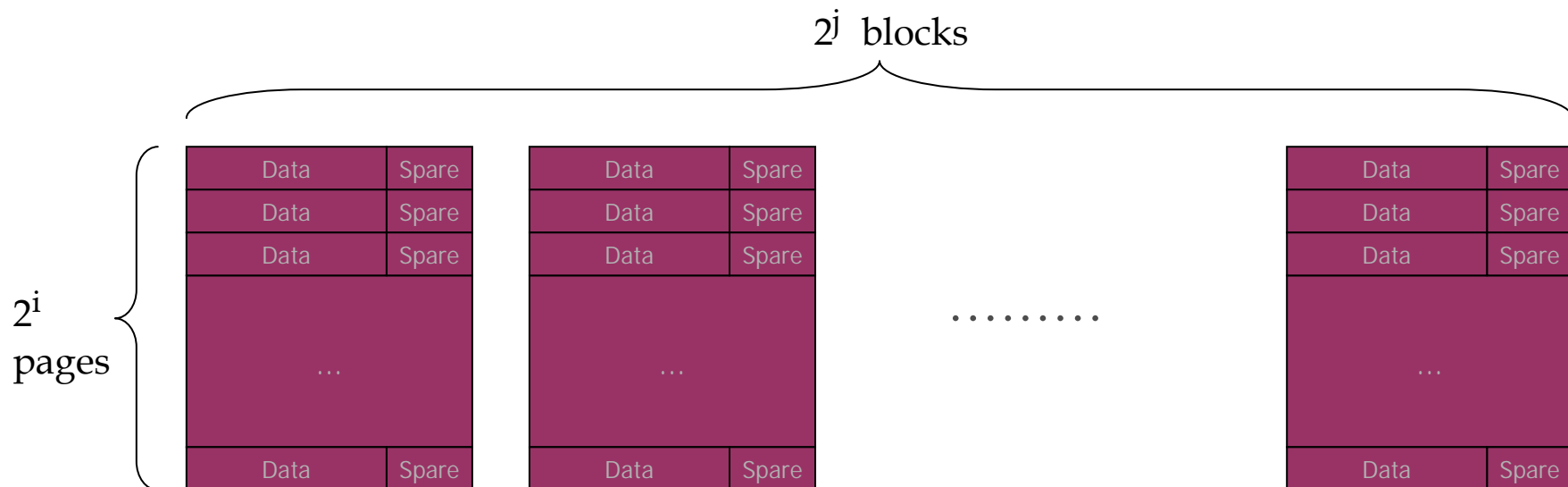


: 0xE5

Types of flash memory

- NOR type
 - Random byte access (same interface as SRAM)
 - Byte programmable
 - High cost/small capacity
 - Mainly used as code storage (ex. PC BIOS)
- NAND type
 - Random page access (special interface)
 - Page (512 B ~ 4 KB) programmable (fast write speed)
 - Low cost/large capacity
 - Mainly used as data storage (ex. USB drive)

NAND flash memory basics



- Read physical page
 - (block #, page #)
 - ~ 25 us
- Write physical page
 - (block #, page #)
 - ~ 200 us
- Erase block
 - (block #)
 - ~ 2 ms
- Data transfer
 - 33 ns/B

Types of NAND flash usage

- Raw chip itself
 - Mounted on system directly
 - Ex. PMP, mobile phone, ...
- As media of independent storage device
 - Attached to host system via well-defined host interfaces
 - Ex. CF/SD/MMC card, USB drive, SSD (Solid State Disk), ...

Examples of NAND flash usage

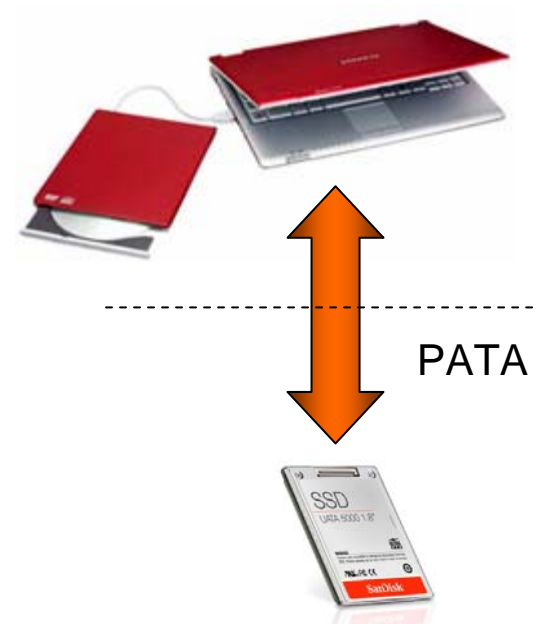
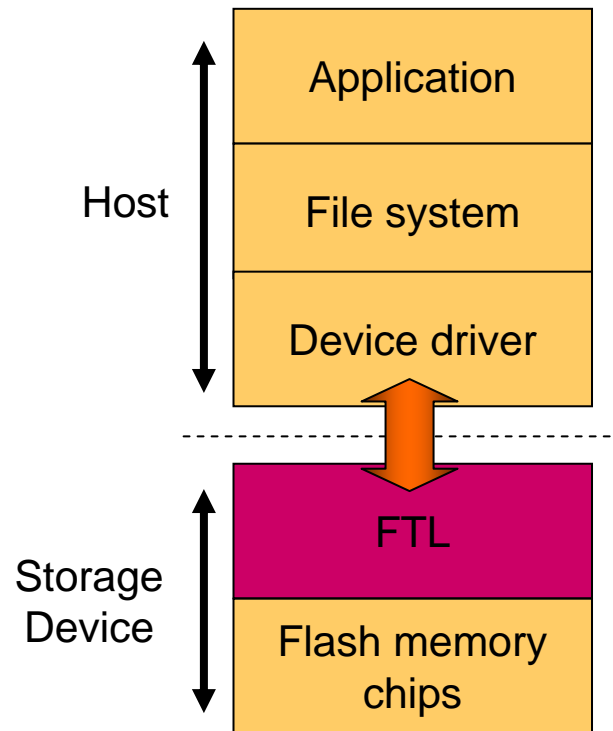
Raw chip



Independent device



Flash storage device architecture



Host-device interaction is based on widely used disk interface protocols
(e.g. ATA, SCSI, ...)

FTL (Flash Translation Layer)

- Definition
 - Software layer that makes flash memory appear to the system like a disk drive



Logical interface for a disk drive



- Operations

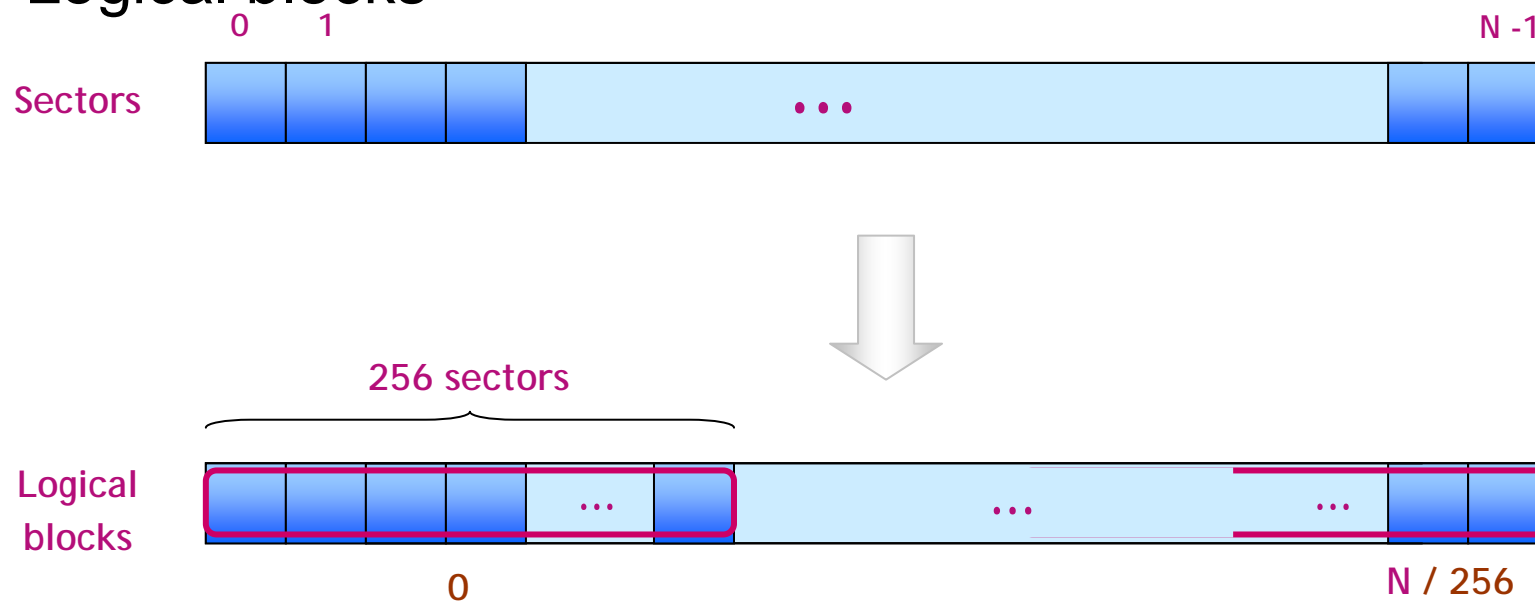
1. Identify drive(): returns N
2. Read sectors(start sector #, # of sectors)
3. Write sectors(start sector #, # of sectors)

Undesirable properties of flash memory

- No in-place update
 - Must erase before overwrite
- Large erase unit & slow erase speed
 - 2ms for 128KB block erase
- Asymmetrical read/write performance
 - 25us for 2KB page read vs. 200us for 2KB page write
- One possible solution
 - Block-level mapping & write buffering

Block level mapping

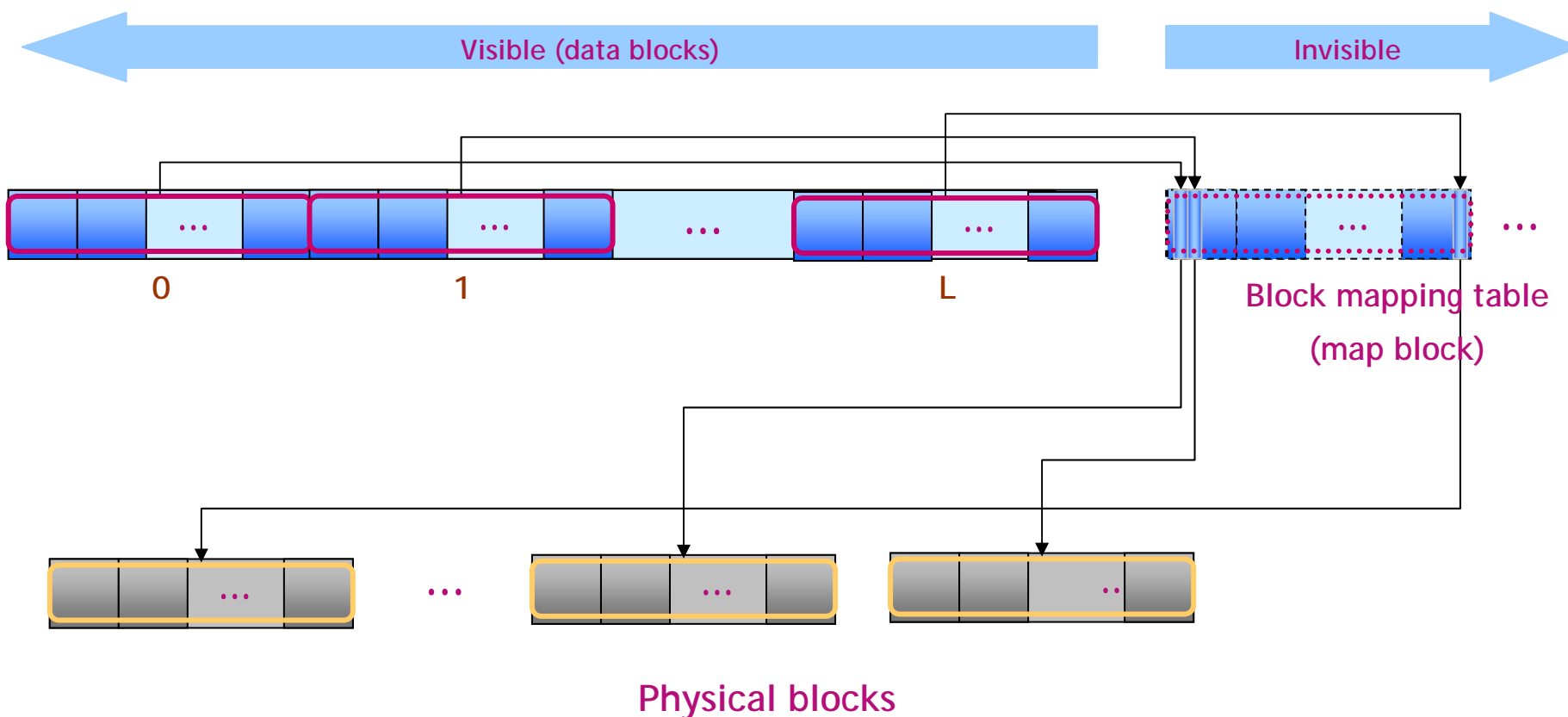
- Logical blocks



Block level mapping

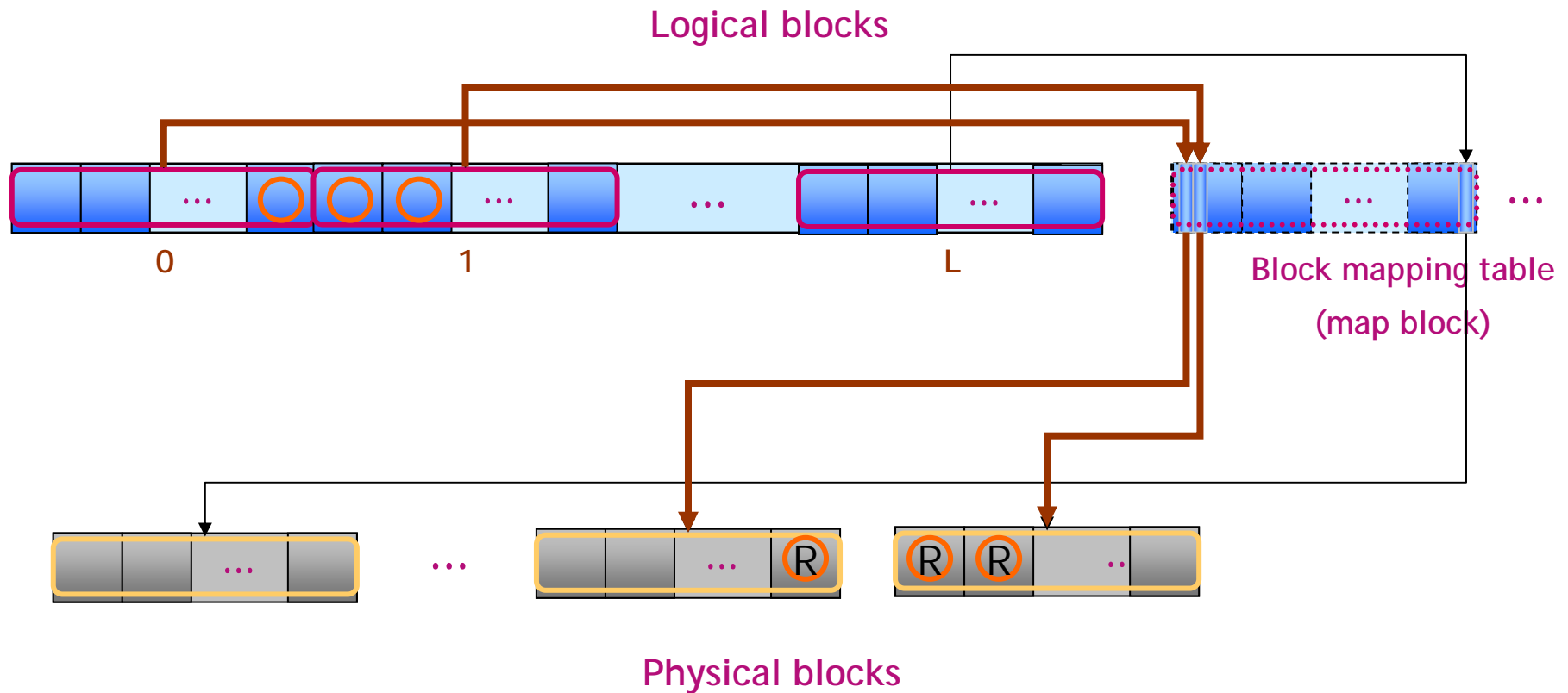
- Logical to physical block mapping

Logical blocks



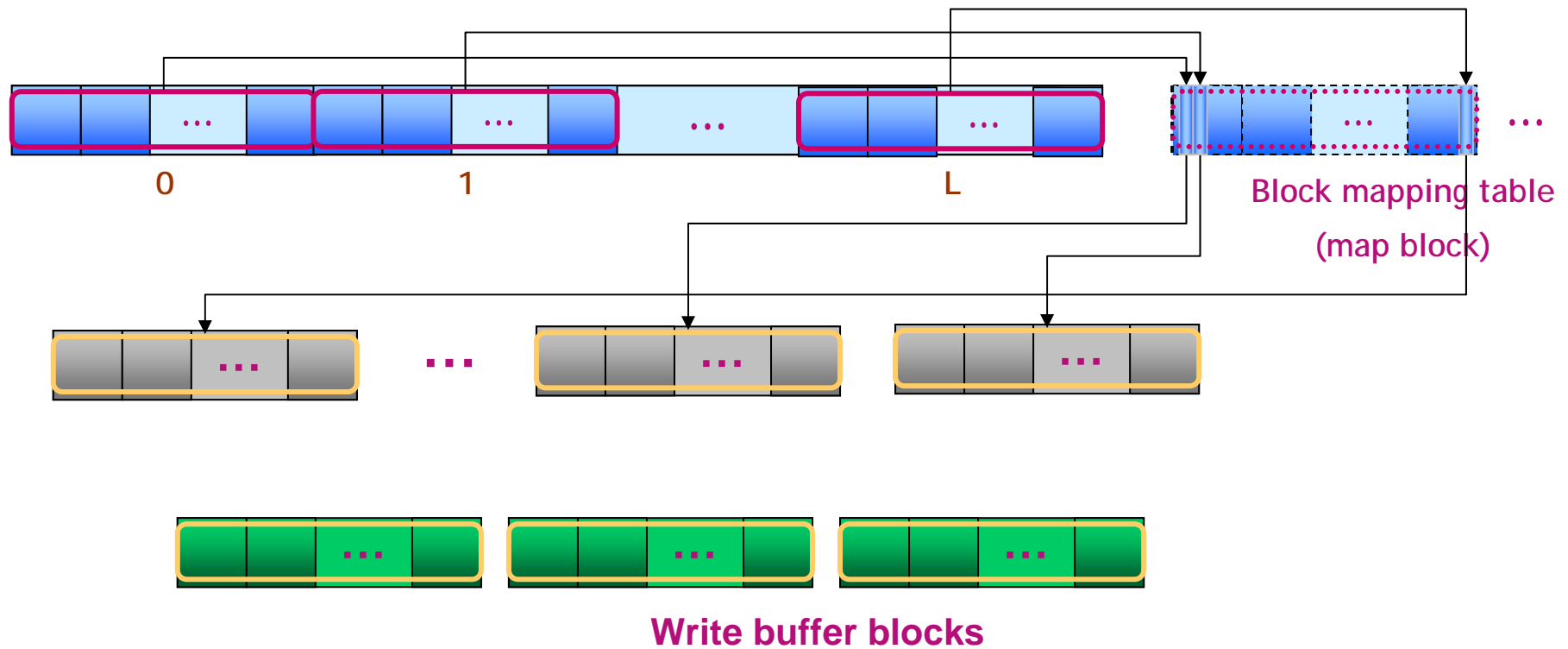
Read procedure

- Ex. read 3 sectors from 255



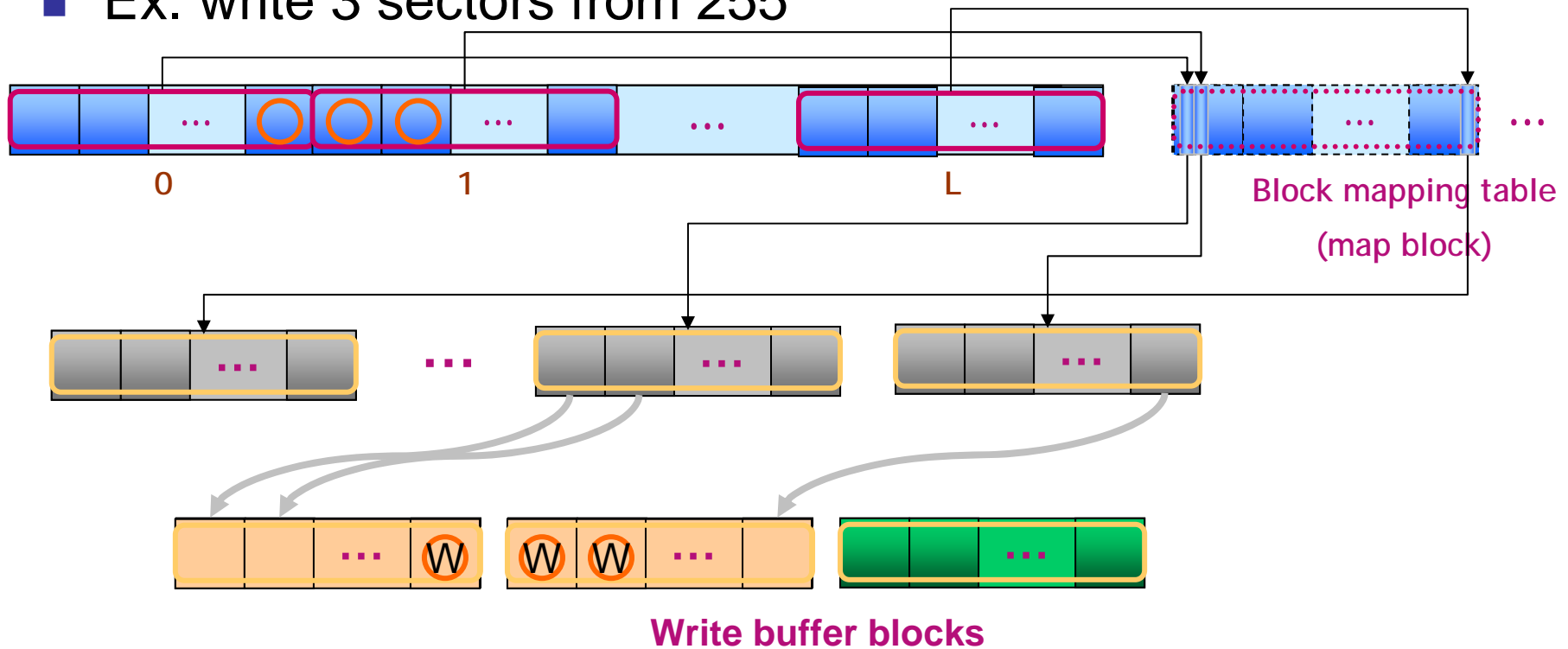
Write buffering

- Write buffer blocks: reserved for processing writes



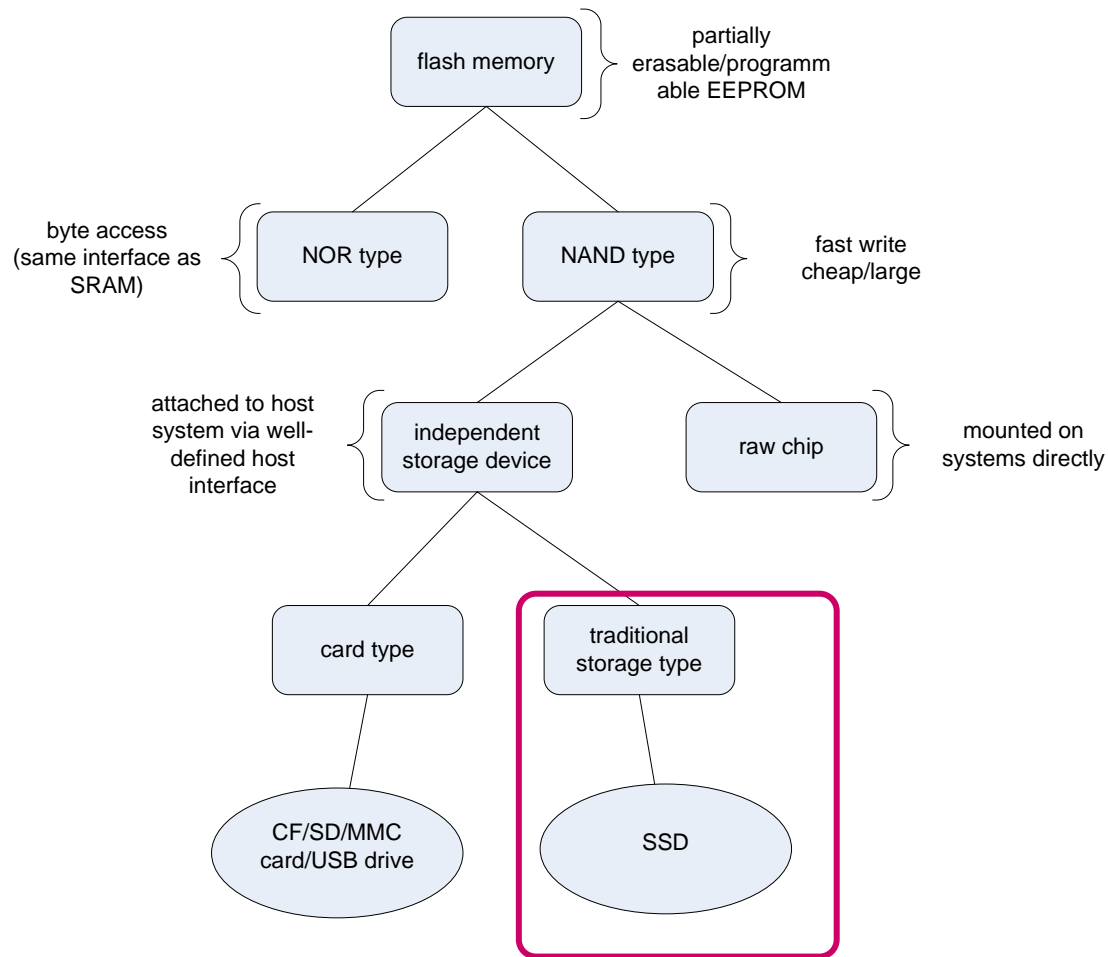
Write procedure (Data block update)

- Ex. write 3 sectors from 255



Still, update of mapping information is needed

Overall picture

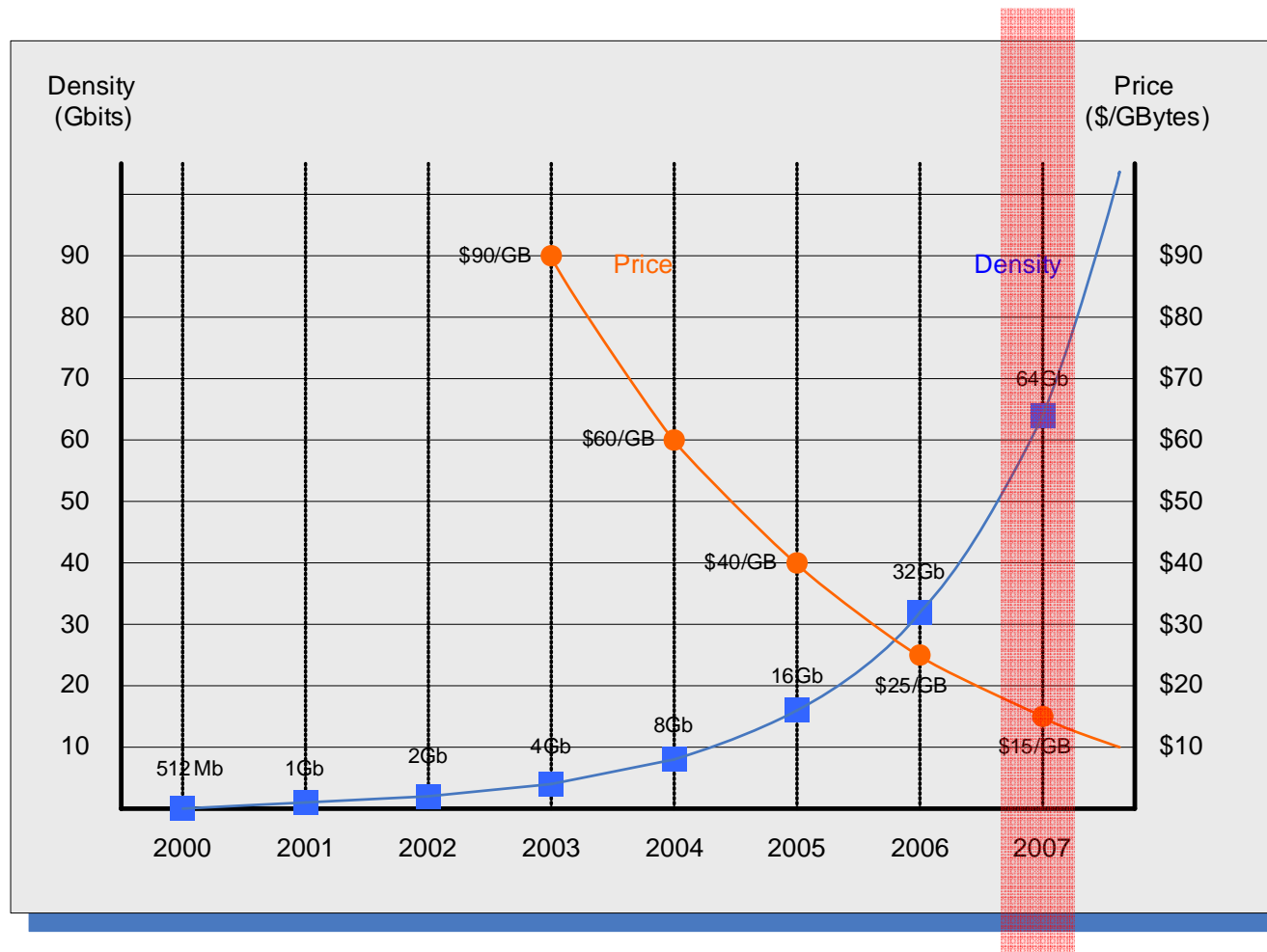


High performance in flash storage

- Generic advantages of flash memory
 - Solid state: low power, small size, high shock resistance, ...
 - Prevalent in portable storage & mobile embedded systems
- “High performance” in flash storage?
 - Technology & market trends in manufacturing flash memory chips
 - Demands on new applications
 - Replacing traditional storage (disk)
 - Large portable storage
 - ...



A critical point in portable storage



16GB SSD at less than 200\$ in 2007

Agenda

- Flash memory basics
- Performance of flash memory based storage
- Hydra: prototype SSD implementation
- Conclusions

Performance of storage

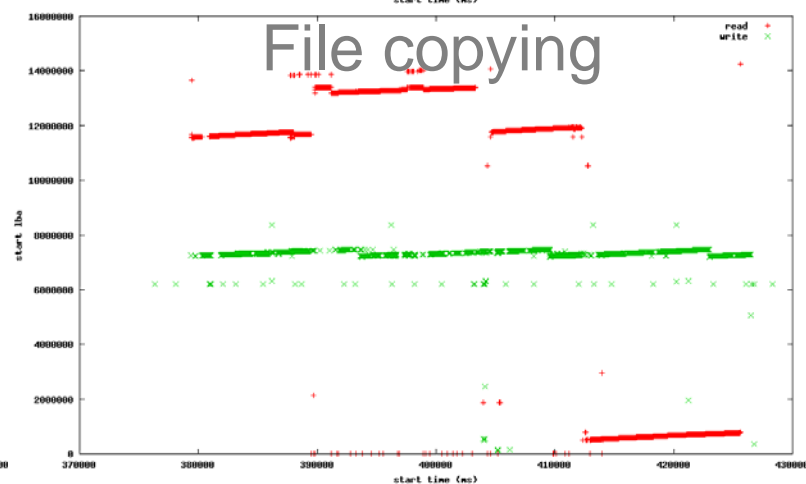
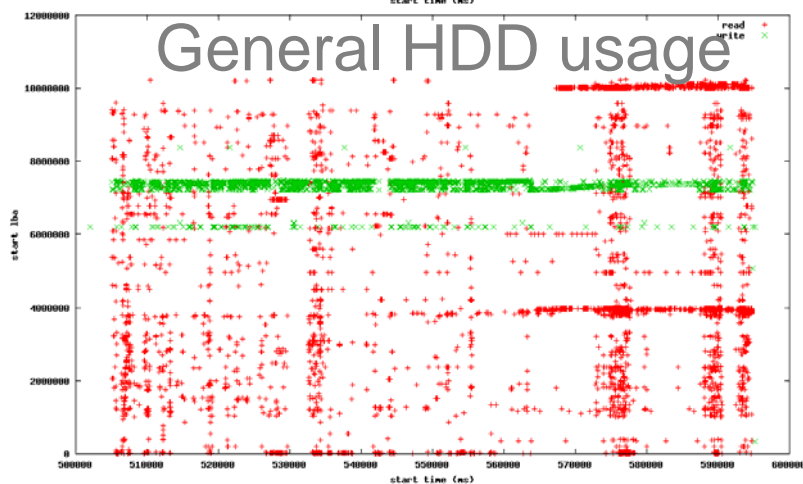
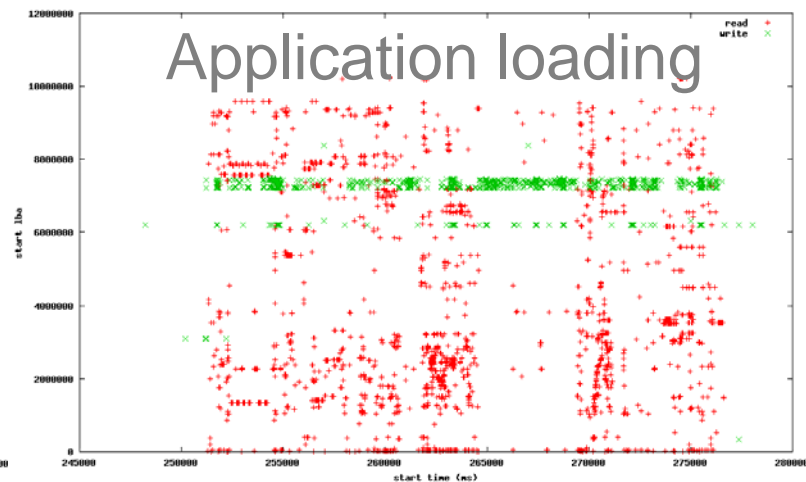
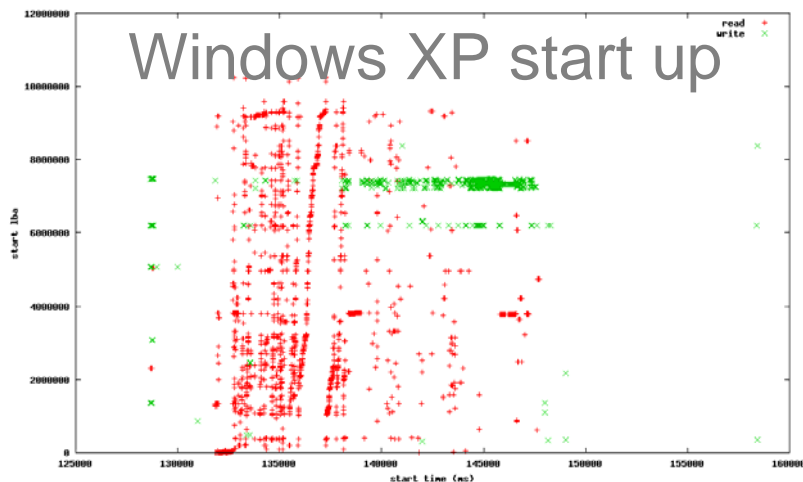
- Low level metrics
 - Throughput
 - Media transfer rate
 - Maximum data transfer speed from media
 - Latency
 - Time to start data transfer
 - Firmware processing overhead + media latency
- High level performance
 - Effective throughput on the given **workload**
 - Ex. PCMARK'04 HDD benchmark

PCMARK'04 HDD benchmark

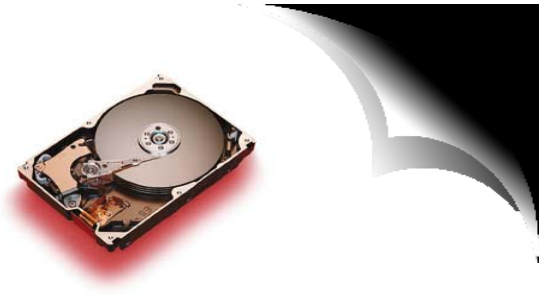


- PCMARK'04 HDD Benchmark
 - Based on real usage
 - Consists of
 - Windows XP Startup
 - ◆ Contains disk activities occurring at operating system start-up
 - Application Loading
 - ◆ Contains disk activities from loading (opening and closing) MS Word, Acrobat Reader, Windows Media Player, etc
 - General Hard Disk Drive Usage
 - ◆ Contains disk activities while executing MS Word, Winzip, Winamp, Internet Explorer, Picture Viewers, etc
 - File Copying
 - ◆ Contains disk activities from copying 400MB of files

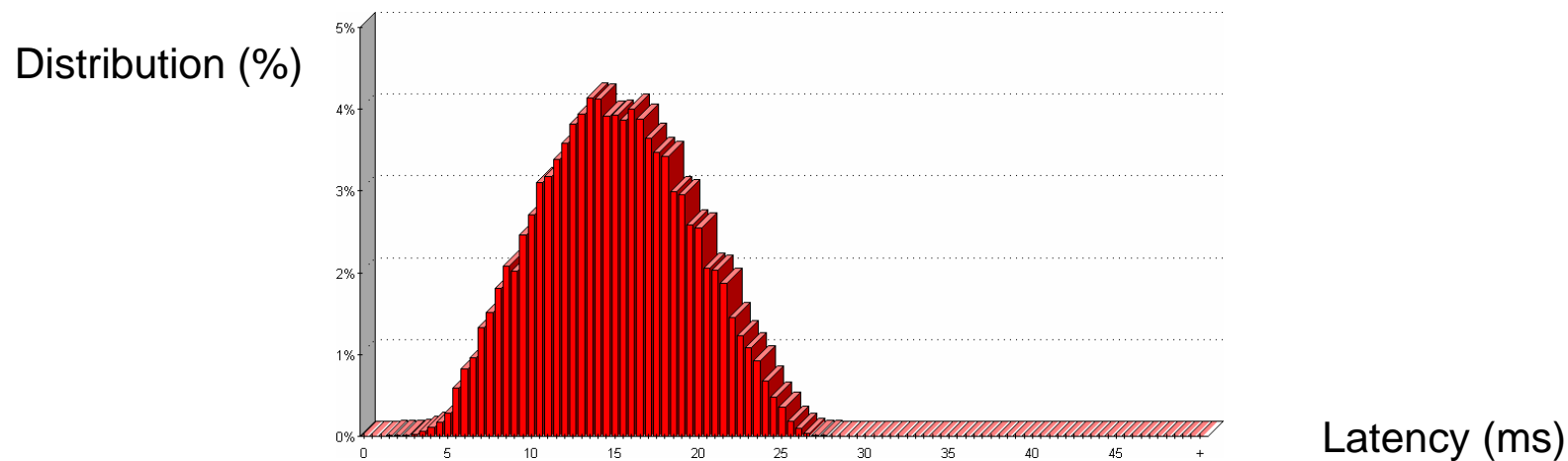
Access pattern (PCMARK'04)



Example: disk performance



- Low level metrics
 - Throughput
 - Data transfer speed
 - 60 MB/s
 - Latency
 - Media latency
 - By spec, average access time for typical disk: 12.1ms
 - By measurements, (from www.storagereview.com)



Example: disk performance



- High level performance
 - PCMARK'04 HDD benchmark

	Disk-PATA
Overall score	4646
XP Startup	8.1
Application loading	6.9
File Copying	34.4
General HDD Usage	5.6

Raw flash chip performance



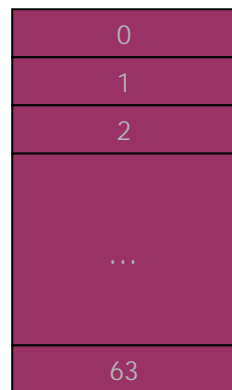
- Throughput
 - Read: $2 \text{ KB} / (25 \text{ us} + 67 \text{ us}) = 22.3 \text{ MB/s}$
 - Write: $2 \text{ KB} / (200 \text{ us} + 67 \text{ us}) = 7.7 \text{ MB/s}$
- Latency
 - Read: near 0 media latency
 - Write: need to prepare free space for new data by erase

Write latency in flash storage



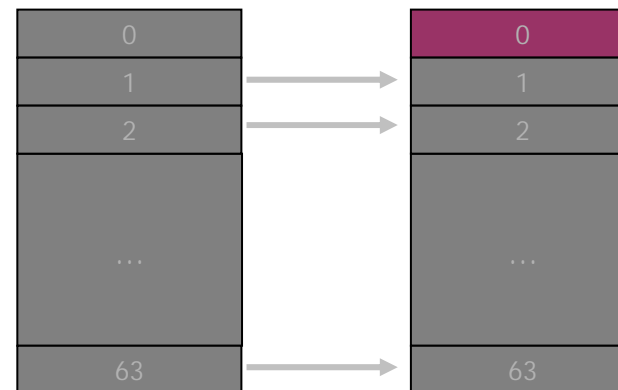
- Need to prepare free space for new data by erase
 - Incur copy overhead of valid old data
 - Problematic especially in processing **small random writes**

Whole block write



Erase write buffer block: 2 ms

Partial block write (only first page)



Erase write buffer block & copy 63 pages: 16.2 ms

Agenda

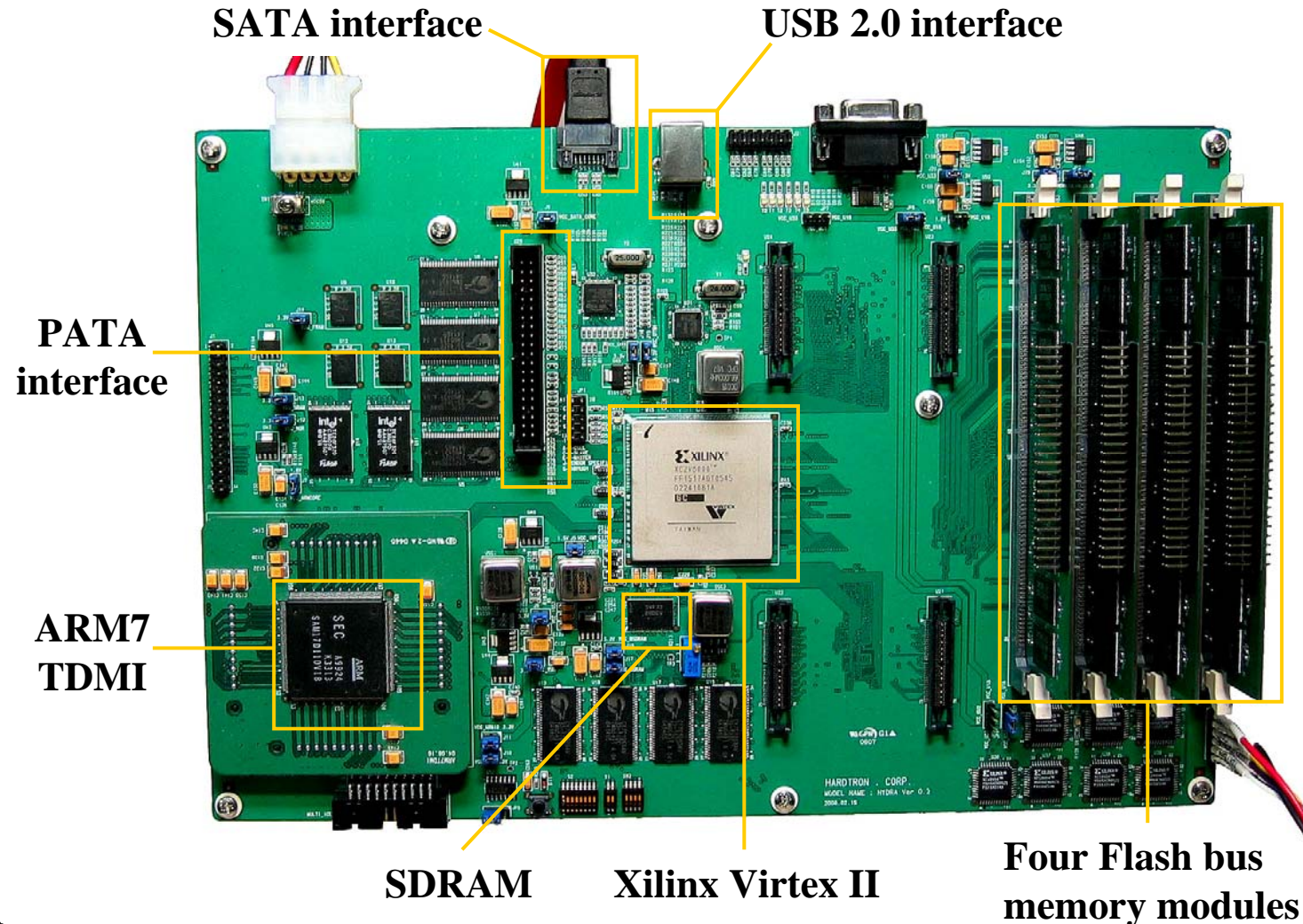
- Flash memory basics
- Performance of flash memory based storage
- **Hydra: prototype SSD implementation**
- Conclusions

Hydra

- Multiple flash memory chips abstracted as multiple independently working heads
 - Optimized for sequential performance
 - Aggressive HW automation
 - PATA UDMA133, SATA1
- Use of external SDRAM
 - For buffering written data
 - To hide flash write latency & collect small random writes

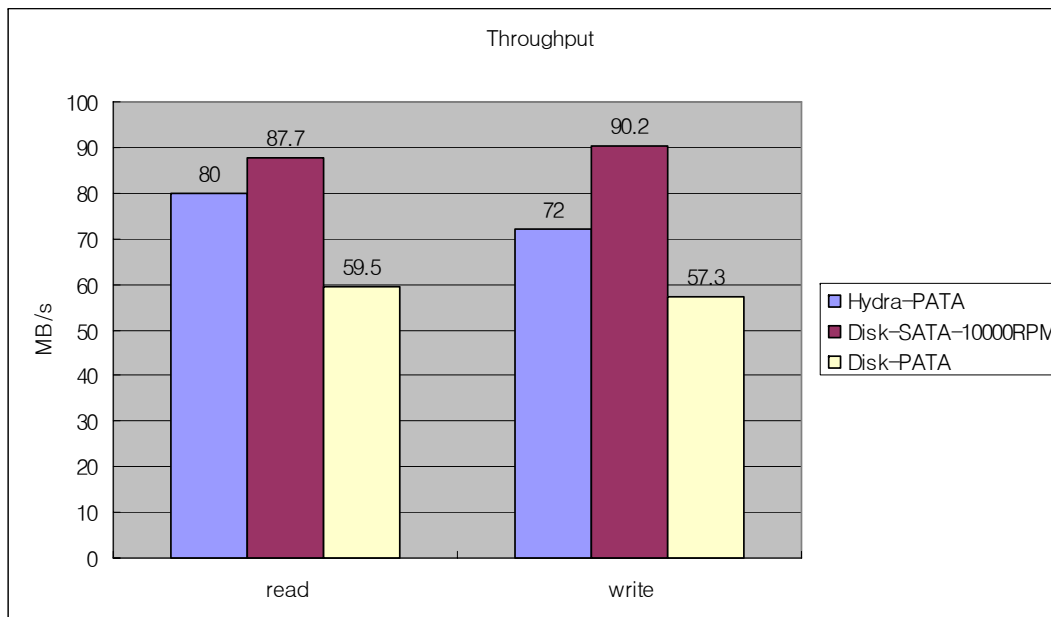


Evaluation board for Hydra (Home-made)



Performance result

- Low level metrics
 - Throughput (32MB sequential read/write)



Performance result

- PCMark04 HDD benchmark

	Hydra-PATA	Disk-SATA-10000RPM	Disk-PATA
Overall score	12566	7282	4646
XP Startup	25.0	12.4	8.1
Application loading	19.8	10.1	6.9
File Copying	51.2	58.8	34.4
General HDD Usage	17.2	8.8	5.6

Agenda

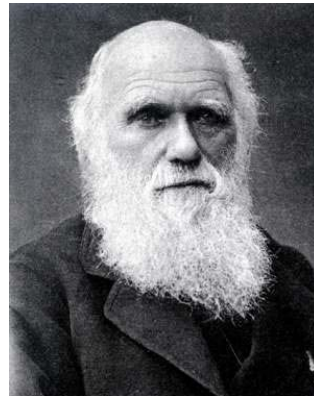
- Flash memory basics
- Performance of flash memory based storage
- Hydra: prototype SSD implementation
- **Conclusions**

Summary

- Characteristics of flash memory based storage
 - Random read: much better than disk
 - Random write: comparable with disk due to copy overhead
 - Need optimizations → use of external SDRAM
 - Sequential read/write: comparable or better than disk through bus/chip level interleaving
- Decent “rookie” of the storage world

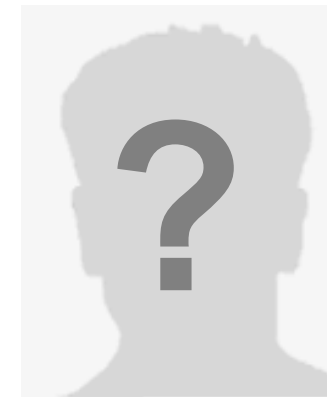


- In the animal world
 - Survival of the fittest



- In the memory world
 - Survival of the fastest or cheapest

	Volatile	Non-volatile
Fastest	SRAM	FRAM?
Cheapest	DRAM	NAND Flash HDD





Thank You

```
INFOR_HEADER_T    #hp;  
    fh(fh->PROCESS) {  
        (FT_INFORMATION);  
    }  
  
/* skip the block erase */  
for (current_block = 0; current_block < NO_OF_BLOCK; current_block++) {  
    FH_Erase(current_block);  
}  
... ..
```

Jin Hyuk Yoon
jhyoon@ssrnet.snu.ac.kr

