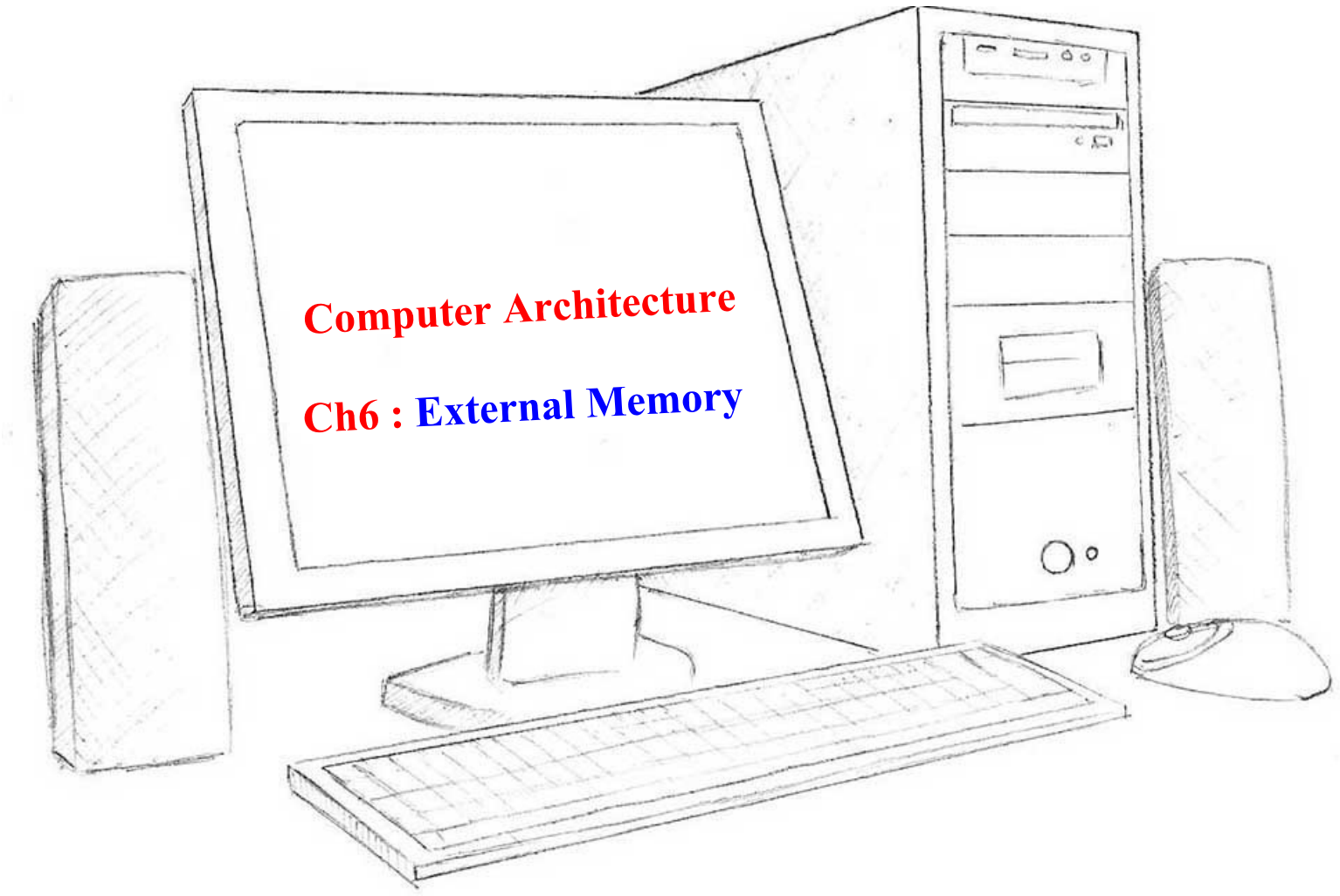


ดร. พิสิทธิ วิสุทธิเมธีกร : pisit.kmutnb@gmail.com

Part II : The Computer System

Ch 6: External Memory

6.1 Magnetic Disk

6.2 Raid

6.3 Solid State Drives

Learning Objectives

After studying this chapter, you should be able to:

- Understand the key properties of magnetic disks.
- Understand the performance issues involved in **magnetic disk** access.
- Explain the concept of **RAID** and describe the various levels.
- Compare and contrast hard disk drives and solid disk drives.
- Describe in general terms the operation of **flash memory**.
- Understand the differences among the different optical disk storage media.
- Present an overview of **magnetic tape** storage technology.

Disk เป็นแผ่นจานกลมที่สร้างจากวัสดุที่ไม่ได้มีคุณสมบัติแม่เหล็ก (Nonmagnetic Material) ที่เรียกว่า Substrate โดยส่วนนี้มักทำมาจากวัสดุที่เป็นอลูมิเนียมหรืออลูมิเนียมอัลลอย และแผ่น Disk นี้จะถูกเคลือบด้วยสารที่มีคุณสมบัติเป็นแม่เหล็กแต่ในปัจจุบัน Substrate ก็ถูกเปลี่ยนมาใช้เป็นแก้วแทนซึ่งมีข้อดีหลายประการคือ

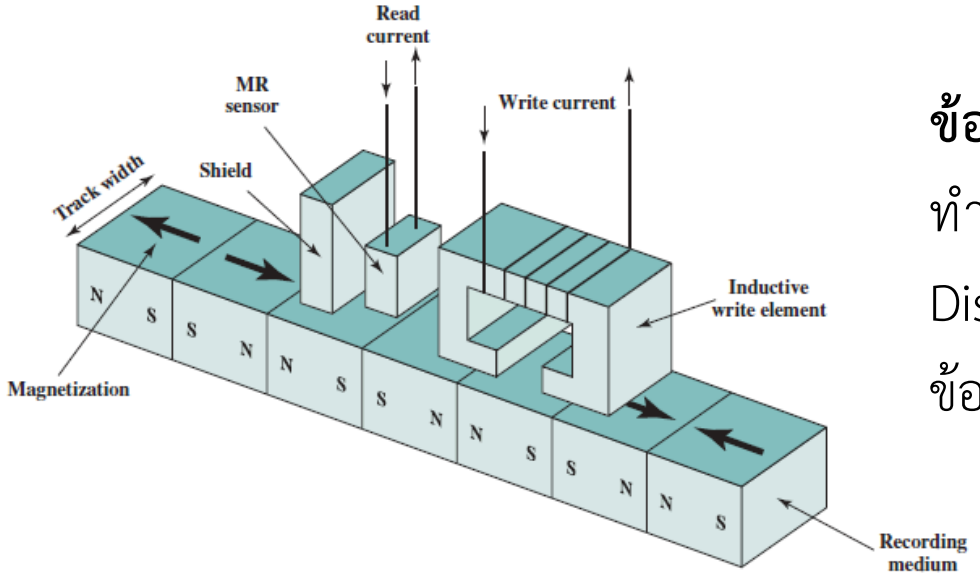
- แก้วมีผิวเรียบทำให้ magnetic film ที่เคลือบมีความสม่ำเสมอและเหมือนกันตลอดแผ่นเพิ่มความน่าเชื่อของแผ่น Disk ในการบันทึกข้อมูล
- ลดความเสียหายของผิว Disk ส่งทำให้ลดความผิดพลาดในการเขียน-อ่านข้อมูล
- สามารถทำให้ Fly height มีค่าต่ำลงได้
- มีความแข็งมากกว่าทำให้ลดการแกว่งของ Disk
- ทนต่อความสั่นสะเทือนและความเสียหายที่จะเกิดขึ้นได้ดีกว่า

Read and Write Mechanisms

ข้อมูลจะถูกบันทึกและอ่านจาก disk โดยใช้ conducting coil ซึ่งตั้งชื่อว่า head โดยแบ่งเป็นสองส่วนอิสระจากกันคือ หัวอ่านข้อมูลและหัวบันทึกข้อมูล (Read head and write head) โดยในการอ่านหรือบันทึกข้อมูลนั้นหัวอ่านข้อมูลและหัวบันทึกข้อมูลจะนิ่งอยู่กับที่โดยมี Disk เป็นส่วนที่เคลื่อนไหวภายใต้หัวอ่านและหัวบันทึกข้อมูลนั่นเอง

การบันทึกข้อมูลจะเกิดจากการป้อนสัญญาณไฟฟ้าที่เป็นพัลส์ตามรูปแบบข้อมูลที่จะบันทึกให้กับขดลวดของหัวบันทึกข้อมูลเพื่อสร้างสนามแม่เหล็ก และจะเกิดรูปแบบแม่เหล็กตามแต่รูปแบบของข้อมูลที่ผิวของแผ่น Disk ที่เคลือบด้วยสารแม่เหล็กนั่นเอง สำหรับหัวบันทึกข้อมูลนั้นจะสร้างด้วยสารที่ทำให้เป็นแม่เหล็กไฟฟ้าได้ง่าย(จ่ายกระแสไฟฟ้า) หัวบันทึกข้อมูลมีลักษณะเป็นโดนัทรูปสี่เหลี่ยมที่มี Gap คล้ายอักษรตัว C ดังรูปที่ 6.1 ด้านที่ไม่ใกล้กับแผ่น Disk ของหัวบันทึกข้อมูลจะมีขดลวดพันรอบแกนโดนัทนี้และทิศทางกระแสที่ผ่านขดลวดจะเป็นสิ่งที่กำหนดขั้วแม่เหล็กที่จะเกิดขึ้นในการบันทึกข้อมูลลงบนแผ่น Disk

การอ่านข้อมูลแต่ก่อนใช้หัวอ่านบันทึกข้อมูลที่เป็นขดลวดเมื่อเคลื่อนที่ผ่านผิวที่เป็นสนามแม่เหล็กของ Disk ก็จะเกิดกระแสที่มีทิศทางที่ขึ้นกับขั้วแม่เหล็กที่บันทึกลง Disk ซึ่งเป็นแนวคิดที่ตรงข้ามกับตอนบันทึกข้อมูลในอดีตวิธีการนี้ถูกใช้กับเครื่องอ่านเขียนข้อมูลที่เป็นแผ่น Floppy Disk หรือ Hard Disk รุ่นเก่าๆ โดยหัวอ่านและบันทึกข้อมูลจะใช้ขดลวดชุดเดียวกันทำให้ประหยัด แต่ Hard Disk ในปัจจุบันหัวอ่านจะใช้ Magnetoresistive (MR) Sensor โดยหลักการคือให้ค่าความต้านทานทางไฟฟ้าที่เปลี่ยนค่าตามสนามแม่เหล็กที่ใกล้เซนเซอร์ ซึ่งก็คือสามารถนำไปเปลี่ยนเป็นสัญญาณไฟฟ้าได้

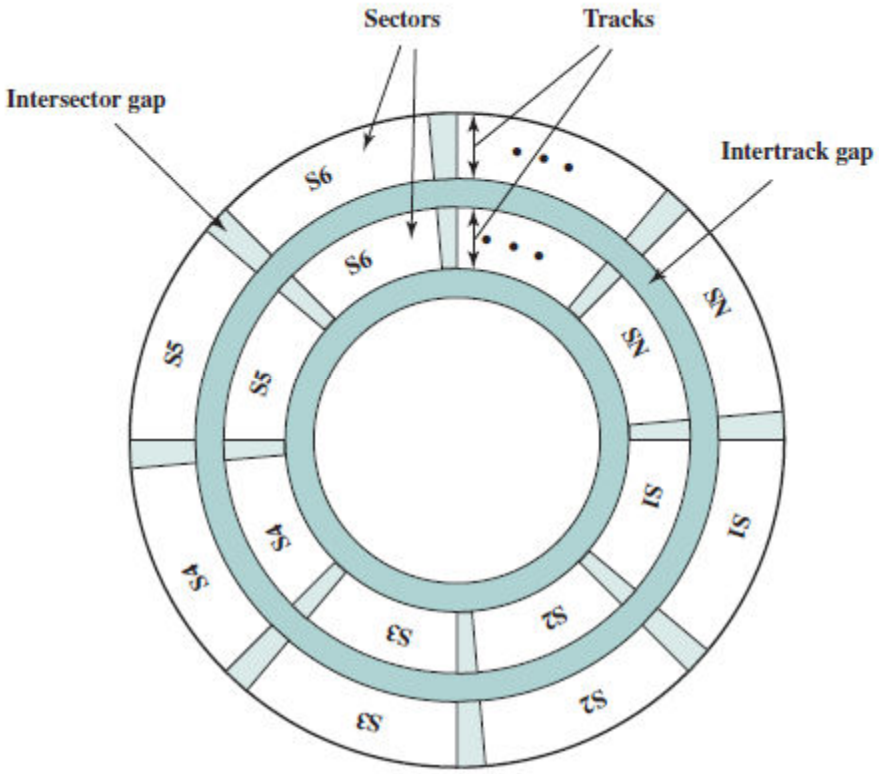


ข้อดีของการใช้ MR คือสามารถทำงานที่ความถี่สูงได้ทำให้ความจุ Disk สูงและความเร็วในการอ่านข้อมูลสูงขึ้น

รูปที่ 6.1 Inductive Write / Magneto Resistive Read Head

Data Organization and Formatting

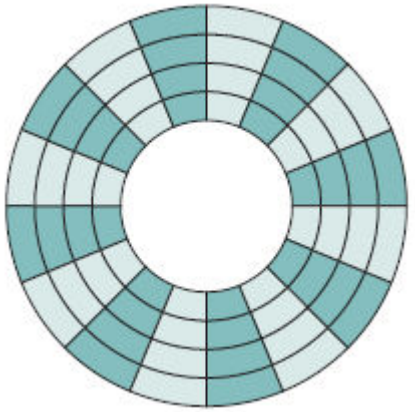
เนื่องจากขนาดของหัวอ่าน/บันทึกข้อมูลมีขนาดเล็กเมื่อเทียบกับ Disk ที่ใช้เก็บข้อมูลดังนั้นในการเก็บข้อมูลบนแผ่นจานกลมของ Disk นี้จะเก็บข้อมูลเรียงกันเป็นวงกลมเรียกว่า **Track** และแต่ละ Track จะมีความกว้างที่เท่ากันโดยในแผ่น Disk จะมี Track ที่เป็นวงกลมหลายขนาดซ้อนกันอยู่เป็นพันๆวง สำหรับ Track ที่อยู่ติดกันจะมีช่องว่างระหว่าง Track เรียกว่า Gap กั้นอยู่เพื่อป้องกันหรือทำให้เกิดความผิดพลาดน้อยที่สุดที่เกิดจากความไม่เที่ยงตรงในการเคลื่อนที่ของหัวอ่าน/บันทึกข้อมูล ตลอดจนป้องกันการรบกวนกันของสนามแม่เหล็กใน Track ที่อยู่ติดกัน นอกจากนั้นยังมีการแบ่ง Track ออกเป็นส่วนๆเรียก **Sector** โดยใน 1 Track จะมีเป็นร้อยๆ Sector และ Sector จะมี Gap ใช้แยกระหว่าง Sector เรียกว่า **Intersector Gap** หรือ Intratrack Gap สำหรับขนาดของ Sector ในแต่ละ Track ก็มีทั้งแบบเท่ากันและไม่เท่ากัน แต่โดยส่วนใหญ่ก็มักจะเป็นแบบที่มีขนาด Sector เท่ากันขนาดเท่ากับ 512 ไบต์ ซึ่งลักษณะของ Disk ที่กล่าวมาสามารถพิจารณาได้จากรูปที่ 6.2



รูปที่ 6.2 Disk Data Layout

พิจารณา Track ที่เป็นวงกลมบน Disk พบว่าแต่ละ Track มีความยาวไม่เท่ากัน Track ที่อยู่ใกล้จุดศูนย์กลางก็จะมีขนาดเล็กกว่าเส้นรอบวงก็สั้นกว่าดังนั้นการหมุนให้ครบรอบของแต่ละ Track นั้นจึงใช้เวลาไม่เท่ากัน ถ้าหากการเก็บข้อมูลบนแต่ละ Track มีจำนวนบิตข้อมูลที่เท่ากันผลก็คือจำนวนบิตข้อมูลที่เขียนหรืออ่านได้ต่อเวลาหรือ Speed ก็จะมีค่าไม่เท่ากันในแต่ละ Track นั่นเอง

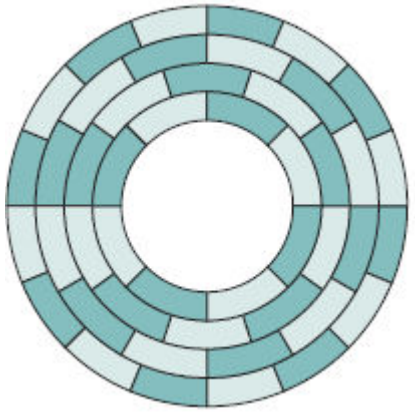
การทำให้ความเร็วในการอ่านข้อมูลคงที่นั้น วิธีการหนึ่งก็คือการเพิ่มระยะห่างระหว่างบิตข้อมูลใน Track เพื่อชดเชยให้ความเร็วในการอ่านข้อมูลของแต่ละ Track มีค่าเท่ากัน โดยที่ Disk ยังหมุนด้วยความเร็วคงที่ วิธีการนี้เรียกว่า Constant Angular Velocity (CAV) หรือจริงๆ ก็คือการทำให้ความหนานแน่นข้อมูลในแต่ละ Track ไม่เท่ากันนั่นเอง



รูปที่ 6.3 CAV Disk Layout

สำหรับข้อดีของวิธีการ CAV นี้ก็คือ การเข้าถึง Block ข้อมูลทำได้เร็วเพียงมีการระบุตำแหน่งก็คือ Sector และ Track ที่ต้องการเข้าถึง ส่วนข้อเสียของวิธีการนี้ก็คือ Track นอกสุดนั้นเส้นรอบวงยาวที่สุดแทนที่จะเก็บข้อมูลได้มากกับต้องจำกัดจำนวนบิตข้อมูลให้เท่ากับ Track วงในสุด นั่นคือความหนาแน่นข้อมูลต่อความยาวเส้นรอบวงของ Track นอกสุดจะมีค่าต่ำที่สุด ทำให้ Capacity ของ Disk นั้นขึ้นกับความสามารถในการเก็บข้อมูลของ Track ด้านในสุด

Hard disk สมัยใหม่จะใช้วิธีการที่เรียกว่า **Multiple Zone Recording** คือ Zone ที่อยู่ด้านในของ Disk จะมีจำนวนบิตข้อมูลต่ำกว่า Zone ด้านนอก

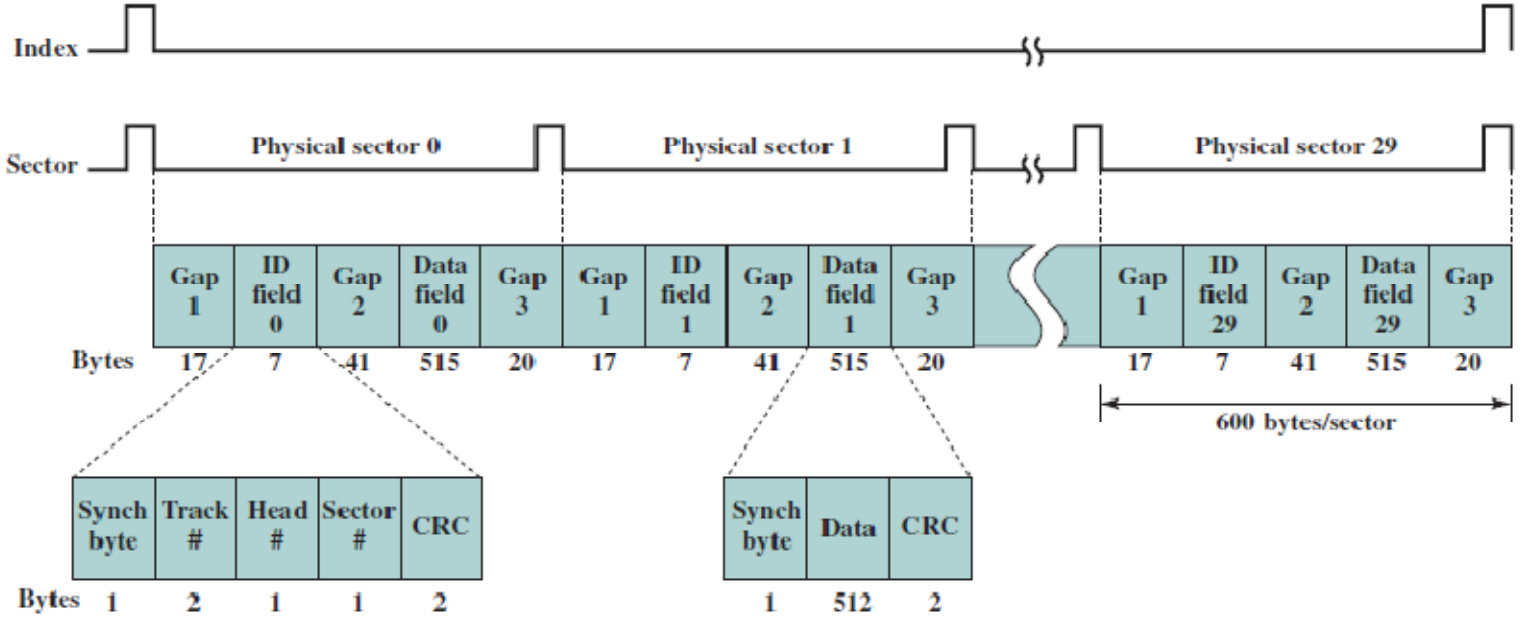


รูปที่ 6.3 Multiple Zone Recording Disk Layout

เพื่อให้ความหนาแน่นข้อมูลเท่ากันเพราะ Zone ด้านนอกมีพื้นที่ (Sector) มากกว่า ซึ่งโดยรวมแล้วทำให้ Capacity เพิ่มขึ้นแต่ก็ต้องแลกด้วยความซับซ้อนที่แบ่งออกเป็น Zone (โดยทั่วไปก็คือ 16 Zone มองเป็นชั้นๆ) แต่ละ Zone มีจำนวนบิต/Track ที่คงที่ โดย Zone ที่ห่างจากศูนย์กลาง Disk จะมีจำนวน Sector มากกว่านั่นคือบิตข้อมูลมากกว่า Zone ที่ใกล้กับจุดศูนย์กลางของ Disk จากรูป แต่ละ Zone ก็คือ 1 Track ในแต่ละ Track จะมีความยาวข้อมูล 1 bit ที่ไม่เท่ากันทำให้ Timing ในการอ่าน/เขียนเปลี่ยนไป จะมีการกำหนดจุดเริ่มต้นของ Track และจุดเริ่มต้นและสิ้นสุดของแต่ละ Sector โดยมีการบันทึก control data บน disk ซึ่งเกิดขึ้นในตอน Format

ในรูป 6.4 เป็นตัวอย่างรูปแบบ Disk รูปแบบหนึ่งโดยแต่ละ Track มี 30 Sector แต่ละ Sector มีข้อมูล 600 bytes (Data : 512 bytes + control data)

มี ID field บอกตำแหน่งของ Sector ซึ่งใน ID Field ประกอบด้วย SYNCH byte เป็นส่วนที่บอกว่าเป็นจุดเริ่มต้นของ ID , หมายเลขของ Track, Sector และ Head (Disk ชนิดนี้มีหลายหัวอ่าน/บันทึก) และ ทั้ง ID และ Data Field ยังมีส่วนของ CRC ซึ่งเป็นข้อมูลในส่วนของ Error Detecting Code



รูปที่ 6.4 Winchester Disk Format (Seagate ST506)

Physical Characteristics

Head Motion	**Platters**
Fixed head (one per track)	Single platter
Movable head (one per surface)	Multiple platter
Disk Portability	**Head Mechanism**
Nonremovable disk	Contact (floppy)
Removable disk	Fixed gap
	Aerodynamic gap (Winchester)
Sides	
Single sided	
Double sided	

Head Motion การเคลื่อนที่ของหัวอ่านใน Hard disk มีสองลักษณะคือ

แบบ Fixed Head คือหัวอ่าน/บันทึกอยู่กับที่แบบนี้จะมี 1 หัวอ่าน/บันทึกต่อ 1 Track ซึ่ง Disk ชนิดนี้ไม่ค่อยมีในปัจจุบันแล้ว

แบบ Movable Head แบบนี้มีหัวอ่าน/บันทึกเพียง 1 หัวเท่านั้นและเป็นแบบที่ติดตั้งบนแขนที่เคลื่อนที่ได้เพื่อเข้าไปยัง Track ที่ต้องการได้

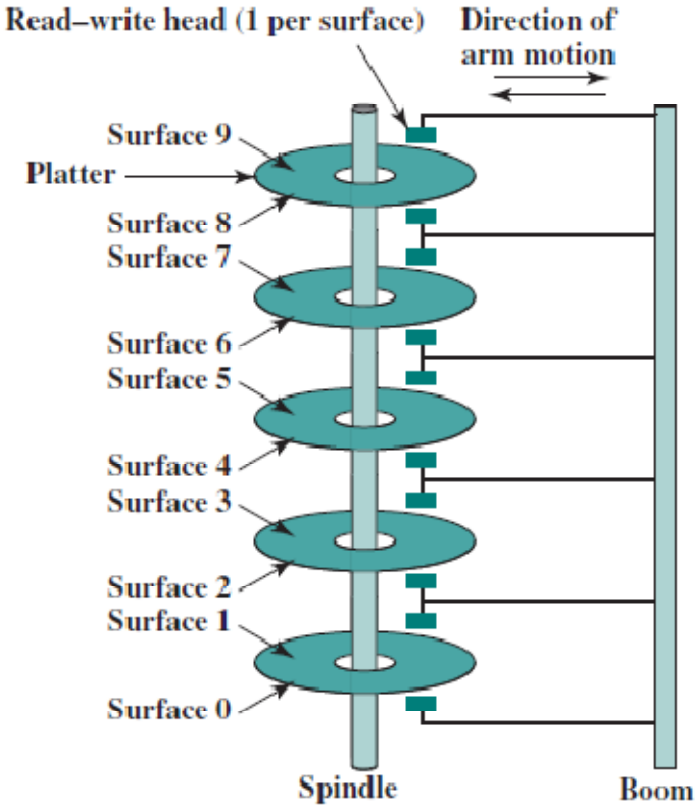
Physical Characteristics (ต่อ)

Disk Portability หมายถึงการนำ Disk ไปใช้ในที่ต่างๆซึ่งก็คือเคลื่อนย้ายจากเครื่องคอมพิวเตอร์หนึ่งไปยังเครื่องอื่นได้หรือไม่ ดังนั้นตามคุณลักษณะข้อนี้จะทำให้ disk อาจแบ่งออกเป็น 2 ลักษณะคือ

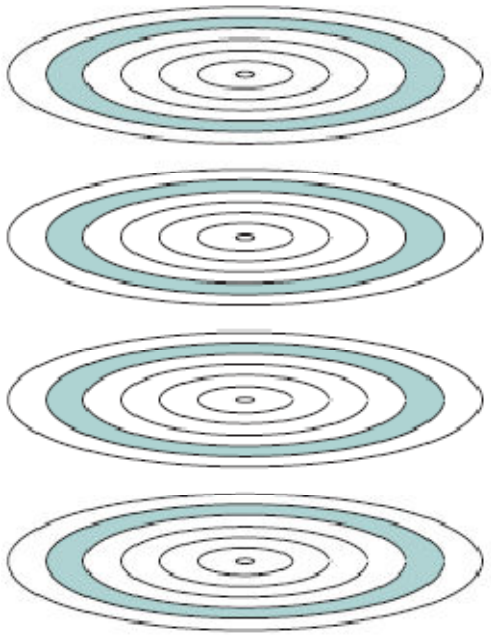
Nonremovable Disk แบบนี้ก็คือ Hard disk ที่ติดตั้งภายในเครื่องคอมพิวเตอร์นั้นเอง

Removable Disk แบบนี้เป็น Disk ที่เคลื่อนย้ายได้จากเครื่องคออมพิวเตอร์เครื่องหนึ่งไปยังเครื่องหนึ่ง disk ชนิดนี้ได้แก่ Floppy disk และ ZIP Cartridge disk

Platter ในที่นี้คือแผ่นจาน หรือ disk เก็บข้อมูลนั่นเองซึ่งจะมีการเคลือบสารที่มีคุณสมบัติทางแม่เหล็กเอาไว้ ซึ่งคุณสมบัติข้อนี้ก็คือมีทั้งแบบ single disk และ Multiple disk โดย Multiple disk จะมีการวาง disk เป็นชั้นๆห่างกัน ดังรูปที่ 6.5 และแต่ละ disk ก็มีหัวอ่าน/เขียนแยกอิสระจากกัน แต่การเคลื่อนที่ของทุกหัวนั้นจะเคลื่อนไปพร้อมกันและมีระยะห่างจากจุดศูนย์กลางของแกนกลาง disk เท่ากันซึ่งชุด track ของทุก disk ในตำแหน่งเดียวกันนั้นจะถูกเรียกว่า Cylinder ดังรูปที่ 6.6



รูปที่ 6.5 Component of Disk Drive



รูปที่ 6.6 Tracks and Cylinders

ส่วนที่เป็นสีทึบของแต่ละ disk รวมกันคือ cylinder

Physical Characteristics (ต่อ)

Sides ปกติแล้ว Disk ก็จะมีการเคลือบสารแม่เหล็กทั้งแผ่นคือทั้งสองด้านนั่นคือ Double sided แต่ก็มี Disk บางชนิดราคาถูกกว่าจะเป็นแบบ Single Sided

Head Mechanism แบ่งเป็นสามลักษณะคือ

-**แบบที่หนึ่ง** หัวอ่าน/บันทึก จะอยู่เหนือแผ่น Disk โดยมีระยะห่างที่คงที่ (air gap)

-**แบบที่สอง** หัวอ่าน/บันทึกจะสัมผัสกับแผ่น disk ขณะการอ่านหรือบันทึกข้อมูล

-**แบบที่สาม** แบบนี้ Disk จะถูกบรรจุในกล่องปิด แนวคิดของ Disk แบบนี้ก็คือหากนำ Disk แบบที่หนึ่งมาและปรับระยะห่างหัวอ่าน/บันทึก และ Disk ให้มีระยะลดลงก็จะสามารถเพิ่มความจุหรือความหนาแน่นในการเก็บข้อมูลของ disk ได้เพราะเมื่อหัวอ่าน/บันทึกอยู่ระยะใกล้กว่าจะสามารถส่งสัญญาณหรือตรวจจับแม่เหล็กจาก Disk ได้ด้วย Track ที่ละเอียดกว่าการที่หัวอ่าน/บันทึกอยู่ห่าง Disk มากกว่า ซึ่ง Disk แบบที่สามนี้เรียกว่า Winchester Disk หัวอ่านจะเป็น Foil วางบน Disk และเมื่อ Disk หมุนจะเกิดความดันยก Foil ลอยขึ้นเหนือ Disk ซึ่งเป็นระยะที่ใกล้กว่าหัวอ่าน/บันทึกในแบบที่หนึ่ง

Disk Performance Parameter ตารางที่ 6.1 แสดงคุณสมบัติของ Disk ที่เป็นชนิดที่มีสมรรถนะสูง ซึ่งปกติแล้วการติดต่อ Disk นั้นจะขึ้นกับลักษณะของระบบคอมพิวเตอร์ ระบบปฏิบัติการ และคุณสมบัติของ I/O Channel และ Disk Controller Hardware

ตารางที่ 6.1 พารามิเตอร์ของ Disk ตัวอย่าง

Characteristics	Constellation ES.2	Seagate Barracuda XT	Cheetah NS	Momentus
Application	Enterprise	Desktop	Network attached storage, application servers	Laptop
Capacity	3 TB	3 TB	400 GB	640 GB
Average seek time	8.5 ms read 9.5 ms write	N/A	3.9 ms read 4.2 ms write	13 ms
Spindle speed	7200 rpm	7200 rpm	10, 075 rpm	5400 rpm
Average latency	4.16 ms	4.16 ms	2.98	5.6 ms
Maximum sustained transfer rate	155 MB/s	149 MB/s	97 MB/s	300 MB/s
Bytes per sector	512	512	512	4096
Tracks per cylinder (number of platter surfaces)	8	10	8	4
Cache	64 MB	64 MB	16 MB	8 MB

ขณะที่ Disk Drive กำลังทำงาน Disk จะถูกหมุนด้วยความเร็วที่คงที่ เมื่อมีความต้องการที่จะอ่านหรือบันทึกข้อมูล จะมีการเคลื่อนหัวอ่าน/บันทึกไปยัง Track ที่ต้องการและเป็น Sector ที่ต้องการของ Track นั้น ซึ่งค่าพารามิเตอร์ของ Disk จะมีค่าเวลาต่างๆดังนี้

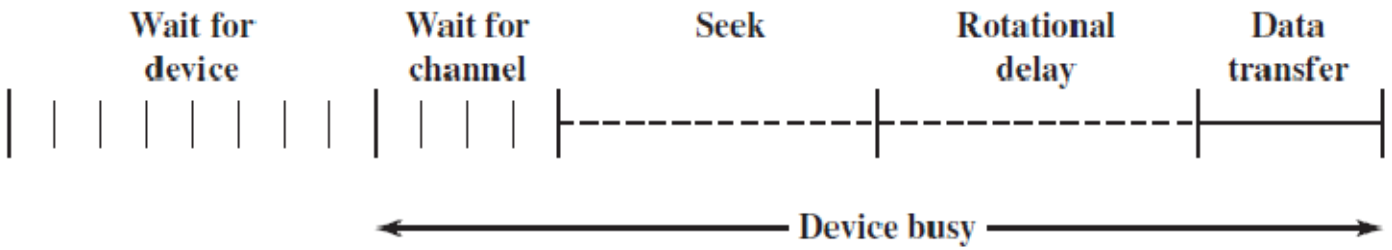
Seek Time : คือเวลาที่ใช้ในการเคลื่อนที่หัวอ่าน/บันทึกไปยัง Track ที่ต้องการ

Rotational Delay : เป็นเวลาที่เกิดขึ้นหลังจากที่หัวอ่าน/บันทึกเคลื่อนที่ไปยัง Track ที่ต้องการแล้วและรอสักระยะเพื่อให้ Sector ที่ต้องการหมุนมายังตำแหน่งของหัวอ่าน/บันทึก และค่าเวลานี้บางครั้งก็อาจเรียกว่า **Rotational latency**

Access Time คือผลรวมเวลาของ Seek Time และ Rotational Delay

Transfer Time คือเวลาหัวอ่าน/บันทึกส่งผ่านข้อมูลกับ Disk ซึ่งนับตั้งแต่ที่หัวอ่าน/บันทึกเคลื่อนที่ไปยังตำแหน่ง Sector ที่ต้องการแล้วเริ่มทำการส่งผ่านข้อมูลจนเสร็จสิ้น

นอกจากนั้นการอ่าน/เขียนข้อมูลกับ Disk ยังมีเวลาที่ต้องเสียไปกับ Disk I/O เพราะเมื่อมีการร้องขอว่าจะทำการอ่านหรือบันทึกข้อมูล ก็จะต้องมีการตรวจสอบว่า Disk Drive นั้นว่างจากการทำงานให้ process อื่นหรือไม่ถ้าว่างต้องตรวจสอบว่า I/O Channel ในการเชื่อมต่อกับ Disk ที่ต้องการว่างด้วยหรือเปล่าดังแสดงเป็น Timing รูปที่ 6.7



รูปที่ 6.7 Timing of Disk I/O Transfer

มี Hard disk บางระบบสำหรับเครื่อง Server จะมีการใช้เทคนิคที่เรียกว่า Rotational Positional Sensing **(RPS)** โดยมีการทำงานคือ เมื่อมีการกระทำคำสั่ง Seek ช่องสัญญาณ I/O จะถูกปล่อยให้นำไปใช้เพื่อตอบสนองการทำงานกับ I/O อื่นๆ จนกระทั่งกระบวนการ SEEK เสร็จสิ้น ชุดควบคุม Hard disk จะตรวจสอบว่าหัวอ่าน/บันทึกไปอยู่ในตำแหน่งเริ่มต้นของ Sector หรือยัง ถ้าพร้อมแล้วก็จะทำการติดต่อกับ Host อีกเพื่อทำการ Transfer ข้อมูล แต่ถ้า Control Unit หรือ I/O Channel ในการติดต่อไม่ว่าง ชุดควบคุม Hard Disk ก็ต้องรอจน แผ่นจานหมุนไป 1 รอบกลับมาจุดเริ่มต้น Sector ใหม่ก่อน จึงจะเริ่มกลับมาร้องขอการเชื่อมต่อกับ Control Unit อีกซึ่งเวลาที่ต้องคอยในการทำ Disk I/O Transfer นี้ก็จะถูกเพิ่มใน Timing ของรูปที่ 6.7

RAID

เพื่อเป็นการเพิ่มสมรรถนะของคอมพิวเตอร์จึงได้มีผู้คิดออกแบบ Disk Storage ที่ประกอบไปด้วย Disk หลายๆตัวที่อิสระจากกันมาทำงานร่วมกันและการที่มี Disk หลายตัวระบบจะต้องรองรับการร้องขอ I/O หลายๆตัวพร้อมกันได้ถ้าเกิดการอ่านข้อมูลจาก Disk คนละตัวในเวลาเดียวกัน

รูปที่ 6.8 Storage servers with 24 hard disk drives and built-in hardware RAID controllers supporting various RAID levels

สำหรับการเก็บข้อมูลบน Disk หลายตัวที่ทำงานร่วมกันนี้จะมีการจัดเป็นมาตรฐานที่เรียกว่า RAID (Redundant Array of Independent Disk) ซึ่งมีอยู่ 7 ระดับ คือ 0-6 โดยทั้ง 7 ระดับมีคุณสมบัติร่วมกันคือ

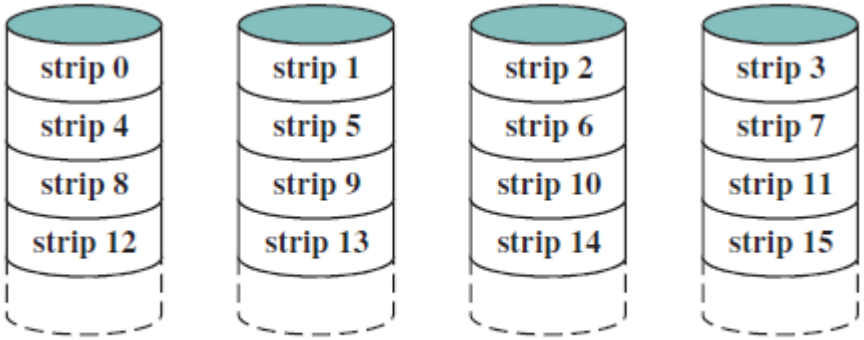
1. RAID เป็น Disk หลายตัวที่เมื่อทำงานแล้วในมุมของ OS จะเห็นเป็น Hard disk เพียงตัวเดียว (หลายๆ Physical Disk แต่มองเห็นเป็น 1 Logical Disk)
2. ข้อมูลที่จัดเก็บจะแบ่งออกเป็นส่วนๆ (Strip) แล้วกระจายกันเก็บบน Physical Disk หลายๆตัว
3. มีการเพิ่มจำนวน Disk เข้าไปเพื่อ Redundant ใช้จัดเก็บ Parity ของข้อมูลซึ่งจะมีประโยชน์ในการกู้ข้อมูลกรณีที่เกิดความเสียหายกับ Disk บางตัวในระบบ

โดยคุณสมบัติข้อ 2 และ 3 ของ Disk ที่มี Raid ระดับต่างกันก็จะมีความแตกต่างกัน โดยเฉพาะ RAID 0 และ RAID1 จะไม่รองรับคุณสมบัติในข้อที่ 3

ตารางที่ 6.2 RAID Level

Category	Level	Description	Disks Required	Data Availability	Large I/O Data Transfer Capacity	Small I/O Request Rate
Striping	0	Nonredundant	N	Lower than single disk	Very high	Very high for both read and write
Mirroring	1	Mirrored	$2N$	Higher than RAID 2, 3, 4, or 5; lower than RAID 6	Higher than single disk for read; similar to single disk for write	Up to twice that of a single disk for read; similar to single disk for write
Parallel access	2	Redundant via Hamming code	$N + m$	Much higher than single disk; comparable to RAID 3, 4, or 5	Highest of all listed alternatives	Approximately twice that of a single disk
	3	Bit-interleaved parity	$N + 1$	Much higher than single disk; comparable to RAID 2, 4, or 5	Highest of all listed alternatives	Approximately twice that of a single disk
Independent access	4	Block-interleaved parity	$N + 1$	Much higher than single disk; comparable to RAID 2, 3, or 5	Similar to RAID 0 for read; significantly lower than single disk for write	Similar to RAID 0 for read; significantly lower than single disk for write
	5	Block-interleaved distributed parity	$N + 1$	Much higher than single disk; comparable to RAID 2, 3, or 4	Similar to RAID 0 for read; lower than single disk for write	Similar to RAID 0 for read; generally lower than single disk for write
	6	Block-interleaved dual distributed parity	$N + 2$	Highest of all listed alternatives	Similar to RAID 0 for read; lower than RAID 5 for write	Similar to RAID 0 for read; significantly lower than RAID 5 for write

Note: N = number of data disks; m proportional to log N

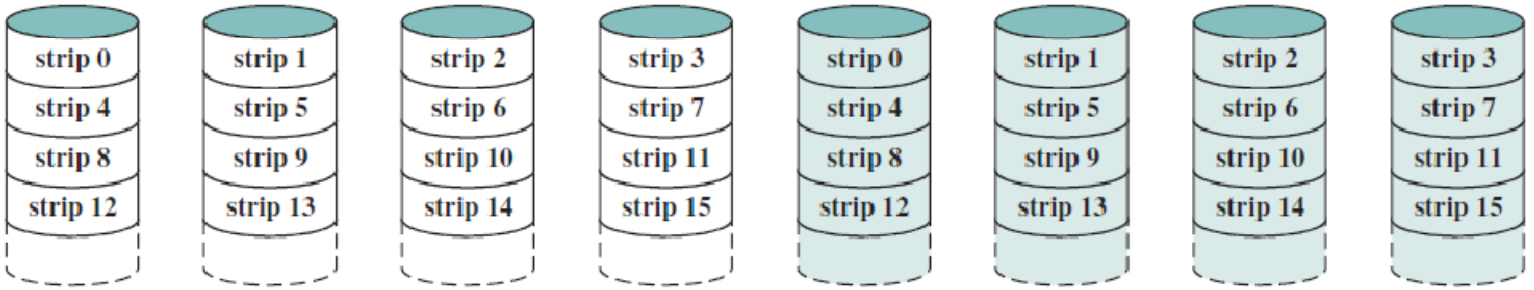


RAID 0 (Nonredundant)

RAID Level 0

รูปที่ 6.9

- No redundancy
- ข้อมูลถูกแบ่งเป็นส่วนๆถูกเก็บที่ Disk ต่างกันและเรียงลำดับกัน
- การแบ่งข้อมูลเป็น Strip และกระจายบนทุก disk (Round Robin)
- Increase speed
  - เมื่อมีการร้องขอข้อมูลหลายตัวที่อยู่บนต่าง Disk กัน
  - Disks seek สามารถเข้าถึง disk หลายตัวพร้อมๆกันได้
  - เมื่อชุดข้อมูลถูกแบ่งเป็นส่วนๆใน disk หลายตัว

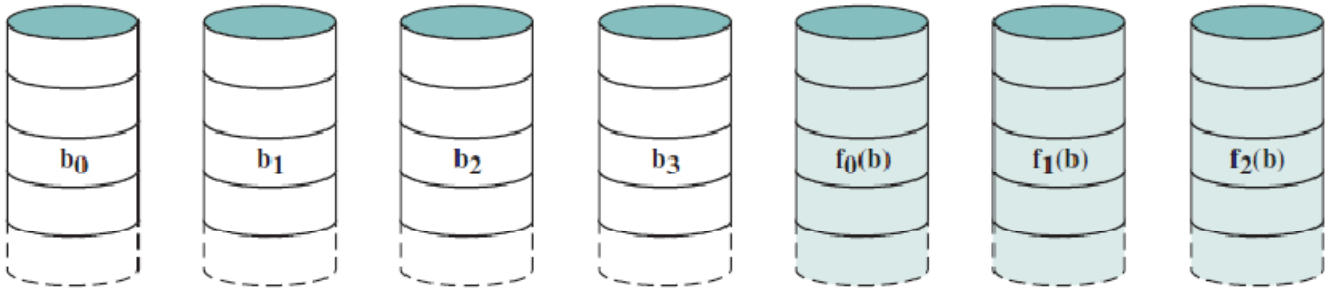


RAID 1 (Mirrored)

รูปที่ 6.10

RAID Level 1

- Disk มีข้อมูลเหมือนกัน 2
- ข้อมูลแบ่งเป็นส่วนย่อยๆกระจายไปในทุก Disk
- การอ่าน-เขียนข้อมูลต้องทำกับ Disk 2 ชุด
- การกู้ข้อมูลทำได้ง่าย เพียงสลับข้อมูลที่เสียหายออกแล้วเอาข้อมูลสำรองมาแทน แล้วทำการ update ข้อมูลบน Disk ที่เป็น Mirror ใหม่
- ราคาแพงเพราะเป็นการใช้ Hard disk 2 ชุด

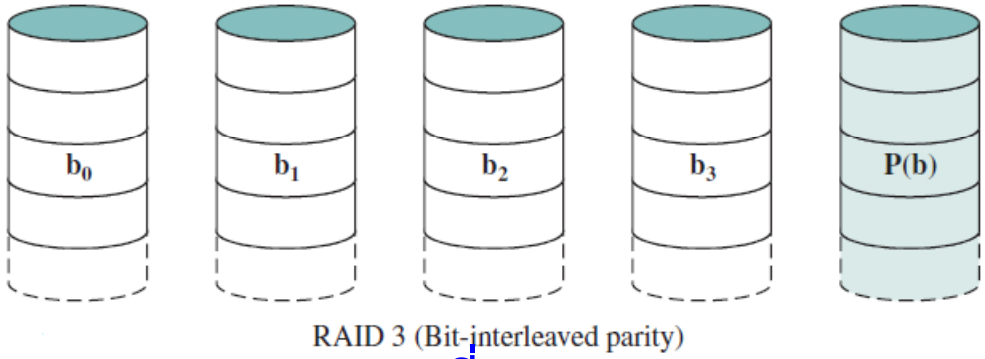


RAID 2 (Redundancy through Hamming code)

รูปที่ 6.11

RAID Level 2

- Disks จะทำงานแบบ synchronize กันคืออ่านในตำแหน่งเดียวกันทุก disk
- ข้อมูลจะแบ่งเป็น stripe เล็กๆในระดับ byte/word
- ใช้การตรวจจับแก้ไขข้อมูลผิดพลาดด้วย Hamming code สร้าง code ตรวจจับอีก 3 บิตเก็บบน Disk อีก 3 ตัวซึ่งทำให้สิ้นเปลือง จนวิธีนี้ไม่ได้ถูกใช้งานจริง



RAID 3 (Bit-interleaved parity)

RAID Level 3

รูปที่ 6.12

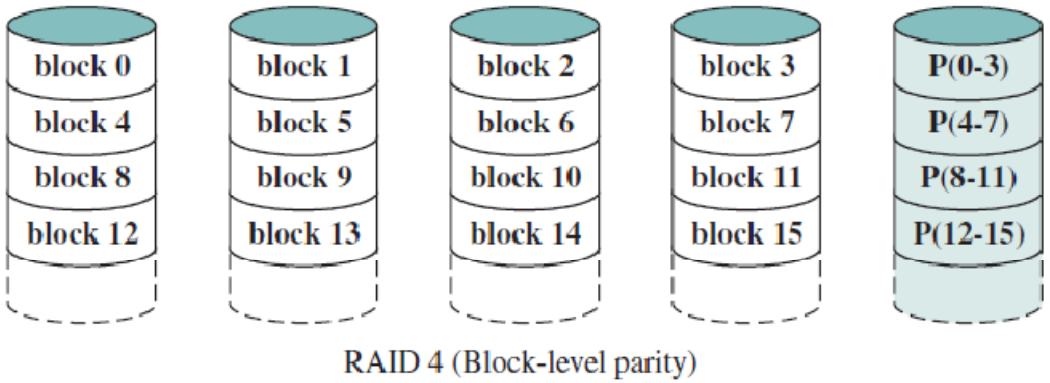
- คล้ายกับ RAID 2 แต่ต้องการ Redundant disk เพียง 1 ชุด
- ใช้ parity bit แบบง่ายแทน Error Correction Code
- หากเกิดข้อมูลเสียหายอาจกู้ข้อมูลได้โดยใช้บิตที่เหลือและ parity bit
- เนื่องจากการแบ่ง strip ในระดับบิตทำให้ระบบมี transfer rates สูง

Parity bit → $X4(i) = X3(i) \oplus X2(i) \oplus X1(i) \oplus X0(i)$

where $\oplus$ is exclusive-OR function.

Suppose that drive X1 has failed. If we add $X4(i) \oplus X1(i)$ to both sides of the preceding equation, we get

$$X1(i) = X4(i) \oplus X3(i) \oplus X2(i) \oplus X0(i)$$



RAID 4 (Block-level parity)

รูปที่ 6.13

RAID Level 4 (Block Level Parity)

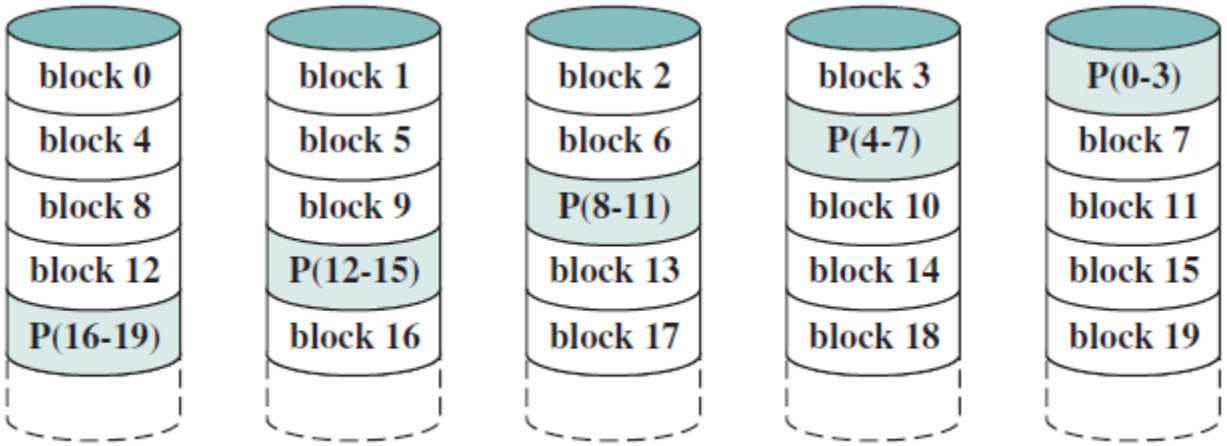
- Disk แต่ละตัวทำงานอิสระกัน (ไม่เหมือน RAID 2-3 ที่เป็น parallel access)
- เหมาะกับระบบที่มี I/O request rate สูง
- แบ่ง Strip เป็น Block
- มีการคำนวณ parity .ในระดับบิตของ Block ข้อมูลที่อยู่ต่าง disk กัน
- Parity bit จะถูกเก็บบน Parity Disk

RAID Level 4 (ต่อ)

$$X4(i) = X3(i) \oplus X2(i) \oplus X1(i) \oplus X0(i)$$

$$\begin{aligned} X4'(i) &= X3(i) \oplus X2(i) \oplus X1'(i) \oplus X0(i) \\ &= X3(i) \oplus X2(i) \oplus X1'(i) \oplus X0(i) \oplus X1(i) \oplus X1(i) \\ &= X3(i) \oplus X2(i) \oplus X1(i) \oplus X0(i) \oplus X1(i) \oplus X1'(i) \\ &= X4(i) \oplus X1(i) \oplus X1'(i) \end{aligned}$$

ด้วยวิธีการข้างต้นนี้ จะพบว่าหากมีการ update ข้อมูล X1 ใหม่ก็จะทำการ update ค่าของ parity bit ใหม่โดยใช้ข้อมูลใหม่เฉพาะบิตที่ update, ข้อมูลบิตเดิม และ parity bit เดิม ซึ่งน้อยกว่าการเอาบิตข้อมูลทุกบิตมาคำนวณ parity bit ใหม่

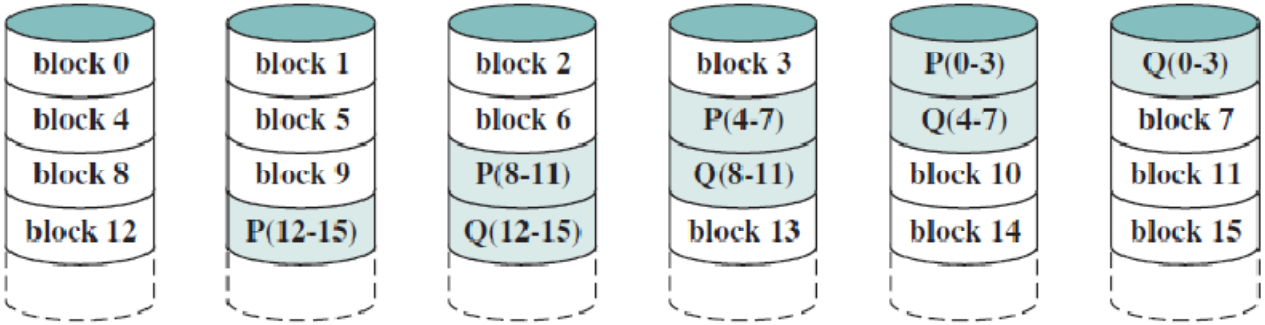


RAID 5 (Block-level distributed parity)

รูปที่ 6.14

RAID Level 5

- คล้ายกับ RAID 4 ต่างกันตรงวิธีการกระจาย Parity strip ไปยังทุก disk
- ใช้ Round robin ในการวางตำแหน่งของ parity stripe
- หลีกเลี่ยงปัญหาความน่าเชื่อถือระบบที่ฝากไว้กับ parity disk (RAID 4)
- โดยทั่วไปใช้กับเครื่อง Server ในระบบเครือข่าย

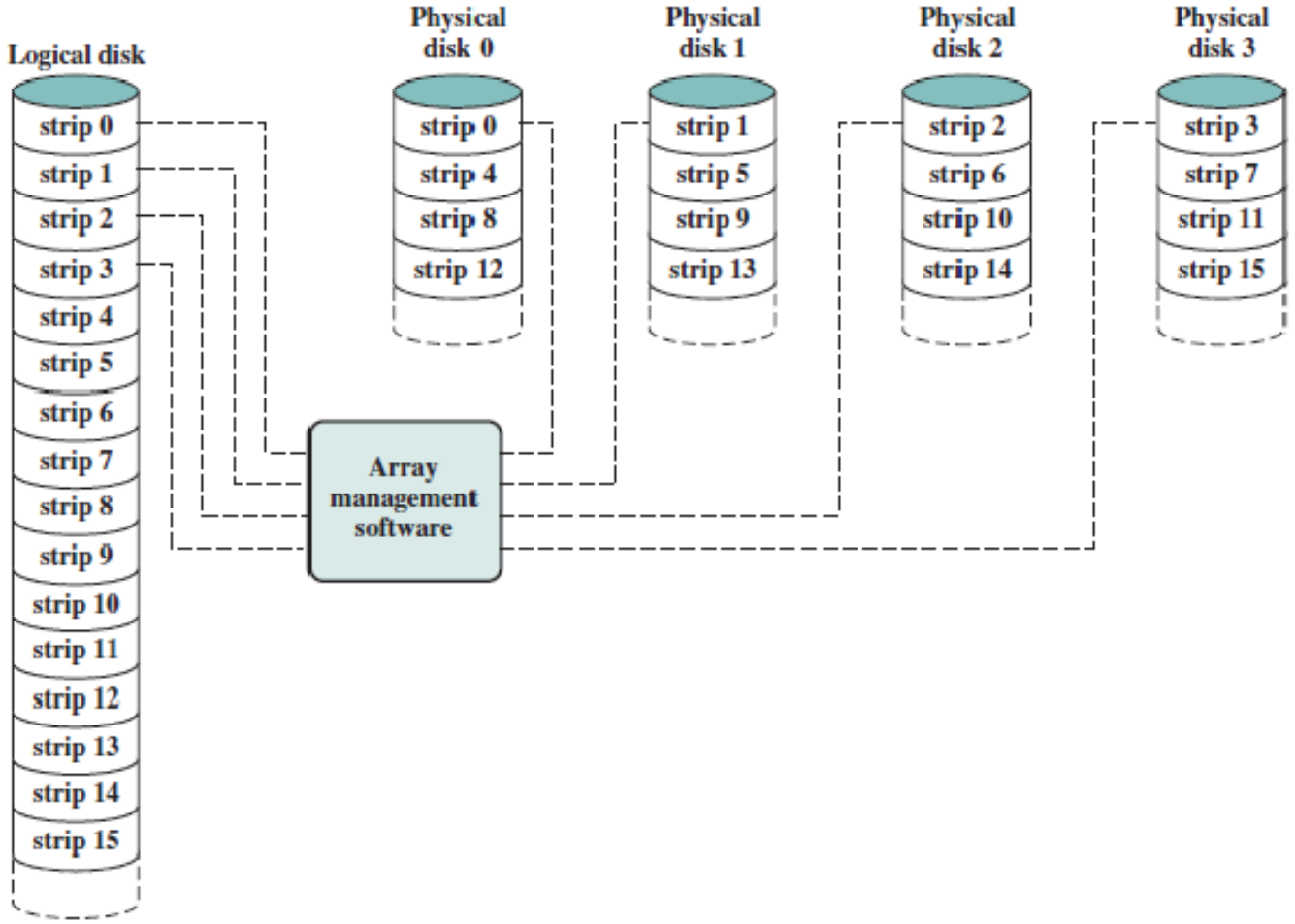


RAID 6 (Dual redundancy)

รูปที่ 6.15

RAID Level 6

- มีการคำนวณ Parity 2 ชุดคือ P และ Q (ดังรูป)
- เก็บ parity ชุดเดียวกันที่เกิดจากต่างวิธีการไว้ต่าง disk กัน เช่นในรูป P(0-3) และ Q(0-3) เก็บที่ Disk 5 และ Disk 6 ตามลำดับ เพื่อแก้ปัญหากรณี Disk ข้อมูลเกิด Fail พร้อมกัน 2 ตัวก็ยังสามารถกู้ข้อมูลขึ้นมาใหม่ได้
- กรณี Disk ข้อมูล N ตัวถ้าใช้ระบบนี้จะใช้จำนวน Disk เท่ากับ N+2



รูปที่ 6.16 Data Mapping for a RAID Level 0 Array

RAID Comparison

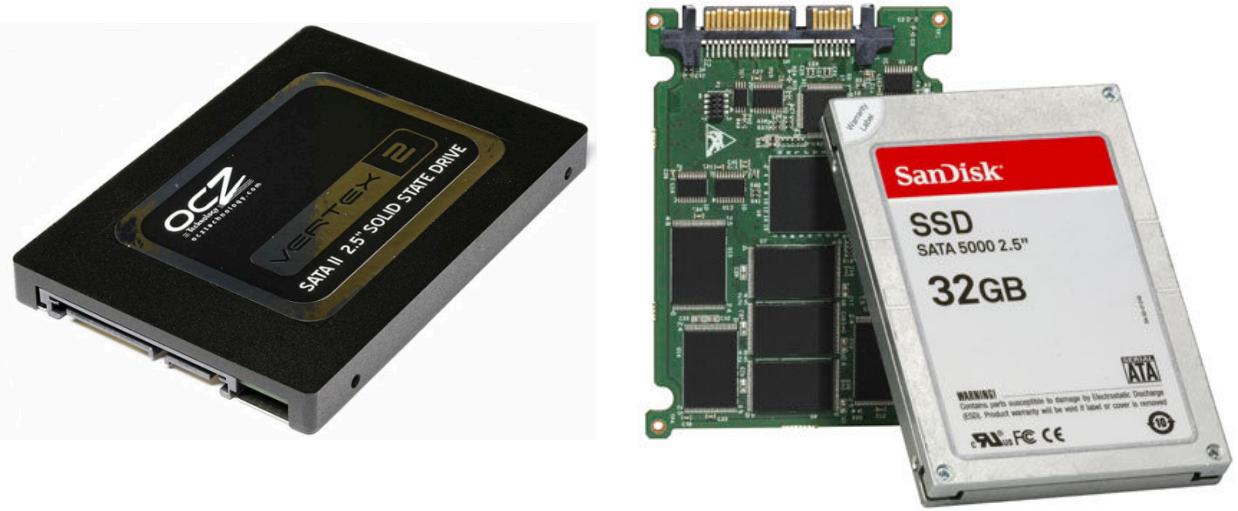
Level	Advantages	Disadvantages	Applications
0	I/O performance is greatly improved by spreading the I/O load across many channels and drives No parity calculation overhead is involved Very simple design Easy to implement	The failure of just one drive will result in all data in an array being lost	Video production and editing Image Editing Pre-press applications Any application requiring high bandwidth
1	100% redundancy of data means no rebuild is necessary in case of a disk failure, just a copy to the replacement disk Under certain circumstances, RAID 1 can sustain multiple simultaneous drive failures Simplest RAID storage subsystem design	Highest disk overhead of all RAID types (100%)—inefficient	Accounting Payroll Financial Any application requiring very high availability

RAID Comparison

Level	Advantages	Disadvantages	Applications
2	Extremely high data transfer rates possible The higher the data transfer rate required, the better the ratio of data disks to ECC disks Relatively simple controller design compared to RAID levels 3, 4, & 5	Very high ratio of ECC disks to data disks with smaller word sizes—inefficient Entry level cost very high—requires very high transfer rate requirement to justify	No commercial implementations exist/ not commercially viable
3	Very high read data transfer rate Very high write data transfer rate Disk failure has an insignificant impact on throughput Low ratio of ECC (parity) disks to data disks means high efficiency	Transaction rate equal to that of a single disk drive at best (if spindles are synchronized) Controller design is fairly complex	Video production and live streaming Image editing Video editing Prepress applications Any application requiring high throughput
4	Very high Read data transaction rate Low ratio of ECC (parity) disks to data disks means high efficiency	Quite complex controller design Worst write transaction rate and Write aggregate transfer rate Difficult and inefficient data rebuild in the event of disk failure	No commercial implementations exist/ not commercially viable

RAID Comparison

5	Highest Read data transaction rate Low ratio of ECC (parity) disks to data disks means high efficiency Good aggregate transfer rate	Most complex controller design Difficult to rebuild in the event of a disk failure (as compared to RAID level 1)	File and application servers Database servers Web, e-mail, and news servers Intranet servers Most versatile RAID level
6	Provides for an extremely high data fault tolerance and can sustain multiple simultaneous drive failures	More complex controller design Controller overhead to compute parity addresses is extremely high	Perfect solution for mission critical applications



รูปที่ 6.17 ตัวอย่าง Solid State Drives

Solid State Drives (**SSDs**) ถูกพัฒนามาเพื่อใช้งานร่วมหรือใช้แทน Hard Disk Drives (HDDs) โดยถูกใช้เป็นทั้ง internal และ external secondary memory โดยคำว่า solid state หมายถึงการใช้วงจรอิเล็กทรอนิกส์ที่สร้างมาจากสารกึ่งตัวนำ ดังนั้น solid state drive ก็จะหมายถึงหน่วยความจำที่สร้างมาจากวงจรอิเล็กทรอนิกส์ที่เป็นสารกึ่งตัวนำซึ่งจะถูกนำมาใช้ในการเก็บข้อมูลเช่นเดียวกับ hard disk drive ซึ่งในปัจจุบัน SSDs ก็ได้รับความนิยมมากขึ้นเพราะมีราคาถูกลงกว่าสมัยก่อน

Solid State Drives (SSDs) จะสร้างขึ้นจากหน่วยความจำแบบสารกึ่งตัวนำที่เรียกว่า **Flash Memory** ซึ่งชื่อนี้เป็นที่รู้จักกันมาหลายปีโดยถูกใช้ในเครื่องใช้อิเล็กทรอนิกส์หลายชนิดได้แก่ smart phones, GPS devices, MP3 players, digital cameras และ USB devices ซึ่งช่วงที่ผ่านมาราคาของ Flash Memory ราคาค่อนข้างต่ำลงมามากจนกระทั่งถึงจุดที่สามารถนำมาใช้แทน Hard Disk ได้

SSD Compared to HDD

As the cost of flash-based SSDs has dropped and the performance and bit density increased, SSDs have become increasingly competitive with HDDs. Table 6.5 shows typical measures of comparison at the time of this writing.

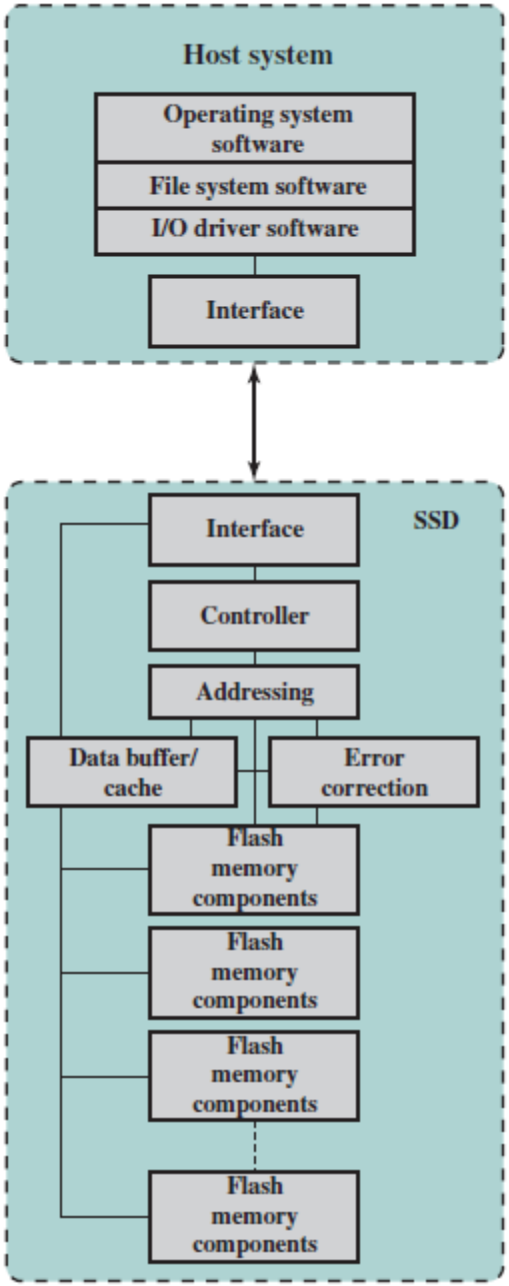
SSDs have the following advantages over HDDs:

- **High-performance input/output operations per second (IOPS):** Significantly increases performance I/O subsystems.
- **Durability:** Less susceptible to physical shock and vibration.
- **Longer lifespan:** SSDs are not susceptible to mechanical wear.
- **Lower power consumption:** SSDs use as little as 2.1 watts of power per drive, considerably less than comparable-size HDDs.
- **Quieter and cooler running capabilities:** Less floor space required, lower energy costs, and a greener enterprise.
- **Lower access times and latency rates:** Over 10 times faster than the spinning disks in an HDD.

Table 6.5 Comparison of Solid State Drives and Disk Drives

	NAND Flash Drives	Disk Drives
I/O per second (sustained)	Read: 45,000 Write: 15,000	300
Throughput (MB/s)	Read: 200+ Write: 100+	up to 80
Random access time (ms)	0.1	4–10
Storage capacity	up to 256 GB	up to 4 TB

รูปที่ 6.18 Solid State Drive Architecture



Controller: Provides SSD device level interfacing and firmware execution.

-**Addressing:** Logic that performs the selection function across the flash memory components.

-**Data buffer/cache:** High speed RAM memory components used for speed matching and to increased data throughput.

- **Error correction:** Logic for error detection and correction.

- **Flash memory components:** Individual NAND flash chips.