Size Matters

Pata

Sata

RAID

RAID 0

RAID 1

RAID 5

SSD

RAM

Is the Machine Good Enough?

Does it have enough space to store my data and analyses?

Is the storage reliable?

Does it have enough memory?

Long term storage is on Hard Drives (HDs)

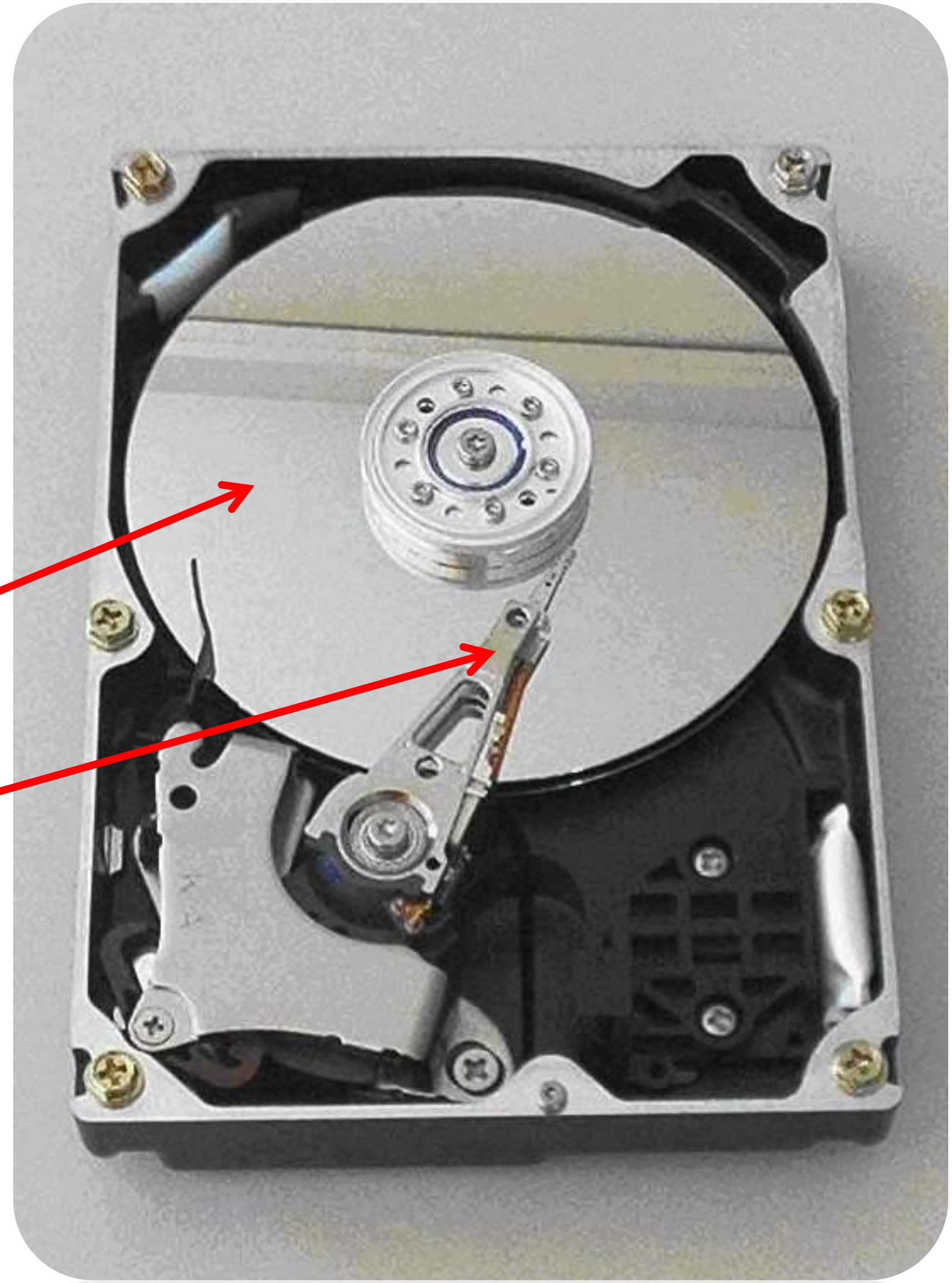
Hard Drives range from ~40 GB to ~2 TB


HITACHI
Deskstar™
SEP-2004
ATA/IDE
DO NOT COVER
THIS HOLE

Hard drive (Inside)

Platter

Read head

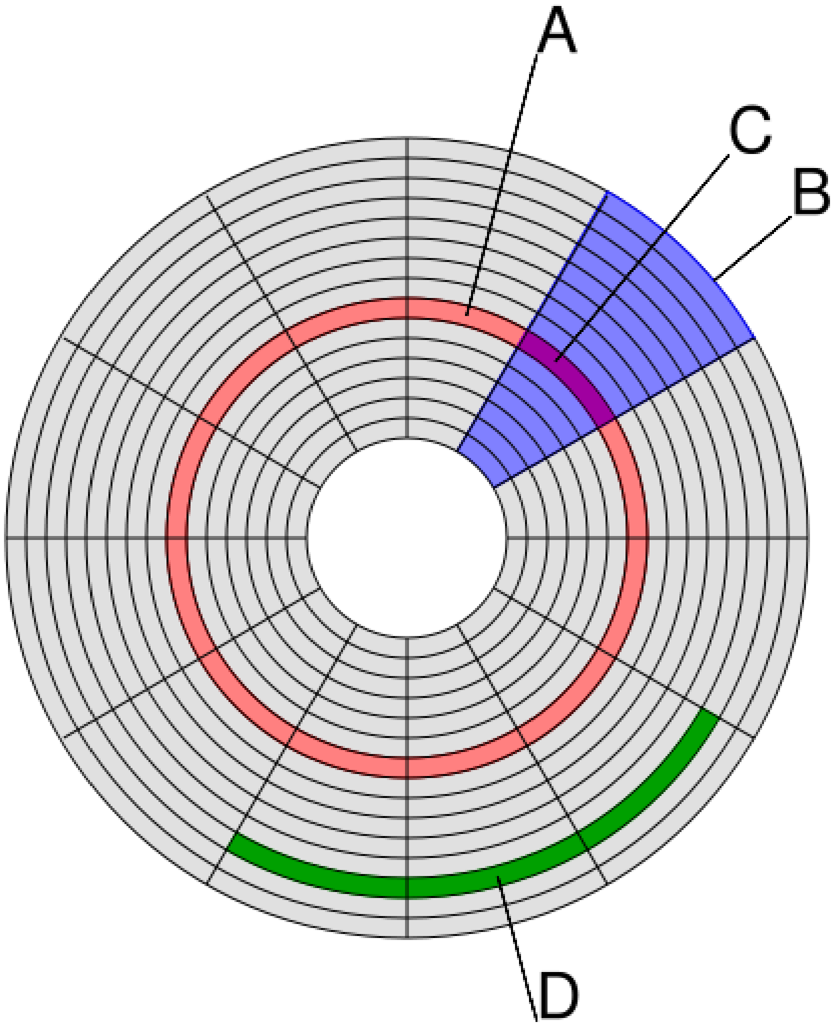


Remind you of anything?

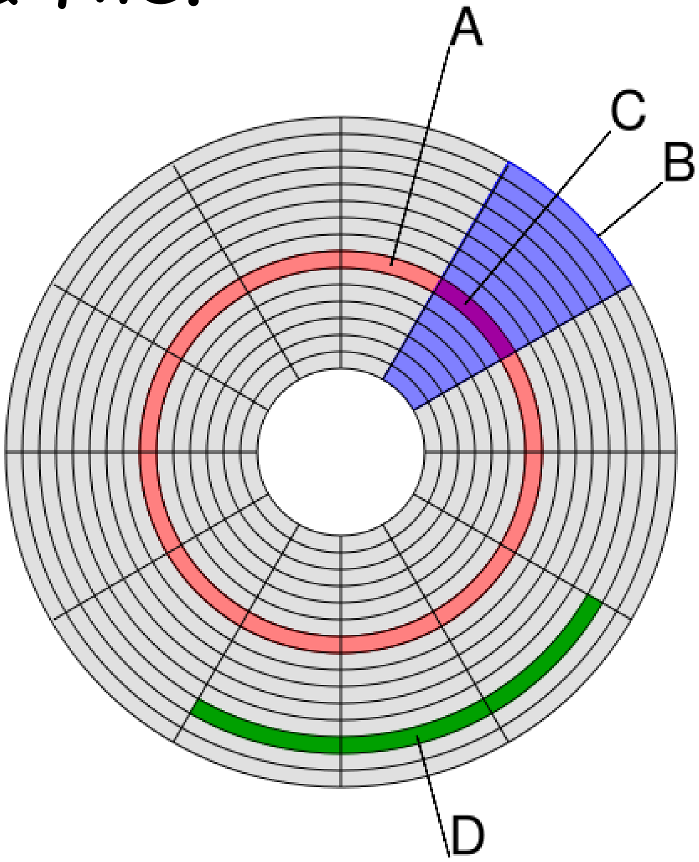
Data is organized on the platter(s):

Disk structure:

(A) track
(B) geometrical sector
(C) track sector
(D) cluster



A cluster (one or more track sectors) is the smallest logical amount of disk space that the Operating System can allocate to hold a file.



(C) track sector
(D) cluster

Because you can never use less than one cluster to store a file…

Storing small files on a file system with large clusters wastes disk space.

But large clusters reduce fragmentation and improve access speed.

i.e., It's easier to organize a few large chunks than lots of tiny ones.

In the best case, the files and the clusters are well matched in size.

(i.e., files will use one or more clusters):

So efficiency and speed are balanced against space saving considerations.

Part of your hard drive is occupied by the operating system and applications (~4-20 GB).

The rest is available for data.

Hard drives are more reliable when they are ~ $\frac{1}{2}$ full or less.

But they are cheap and easy to add.

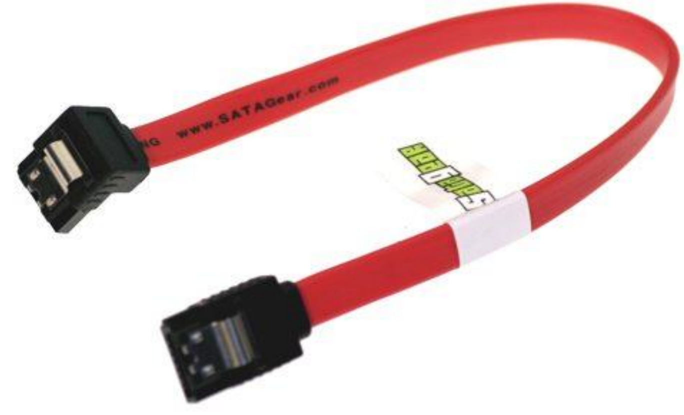
There are 2 main Hard Drive interface technologies: PATA and SATA

ATA=Advanced Technology Attachment

PATA (Parallel ATA) is the older slower technology: 33-133 MB/sec transfer speed.

ATA=Advanced Technology Attachment

SATA (Serial ATA) is the newer faster technology:150-300 MB/sec transfer speed.



ATA=Advanced Technology Attachment

The major problem with hard drives is RELIABILITY.

Google Hard Drive Study:

Drives 2 years and older fail at ~8% per year.

It is ALWAYS worth your time to back up.

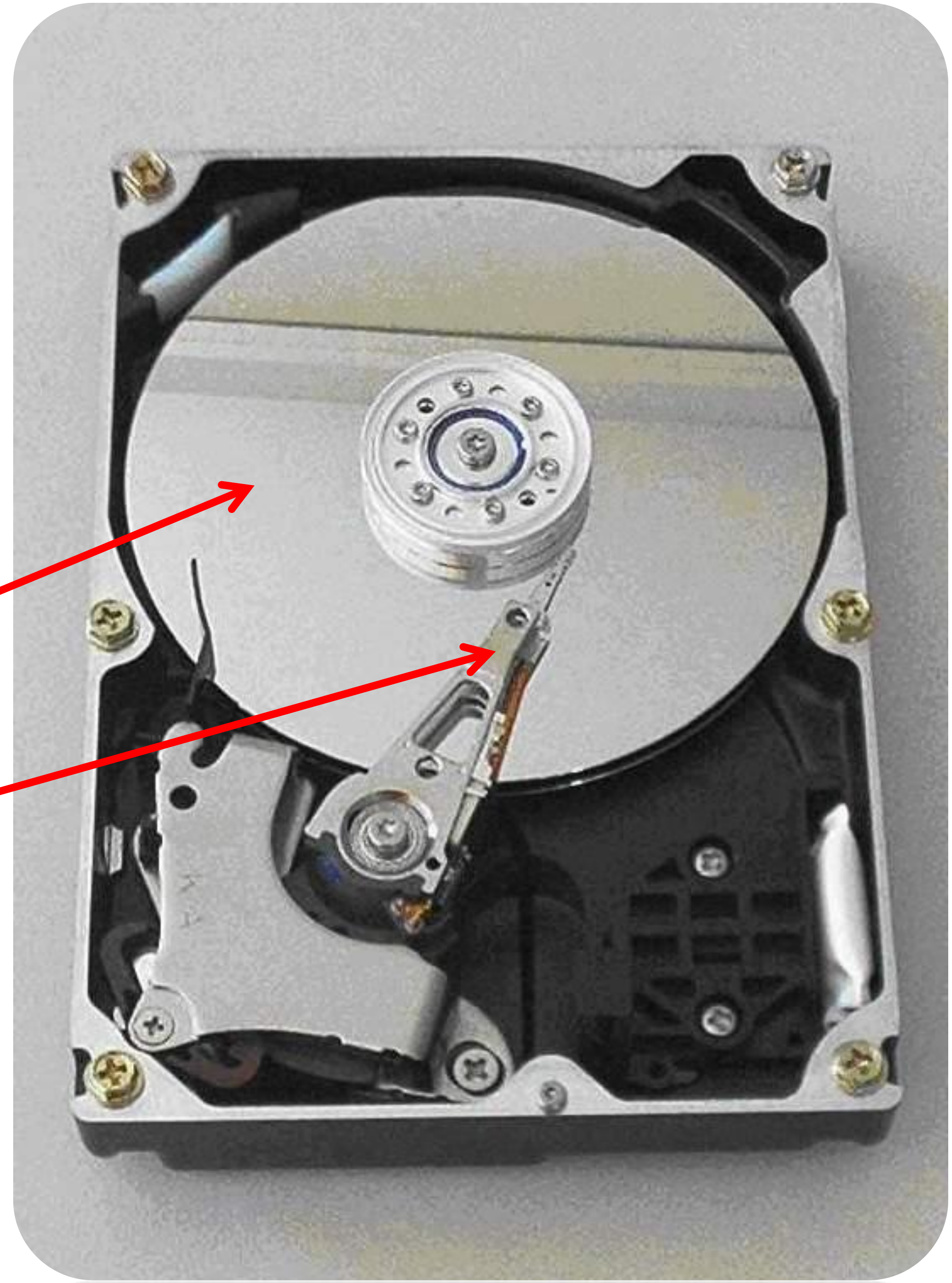
Drives that have experienced corruption are more likely to fail.

But drives can fail for no apparent reason…

Moving parts (spinning platters and a tiny read head), are easy to break.

Platter

Read head



They may grind before they seize up (giving you a chance to copy your data off).

Or, they may fail quietly and completely.

Sometimes, you can bang on them until they unseize (and you can get your stuff off).

Data recovery companies exist, but are VERY expensive.

Other than backing up and praying, what can you do?

RAID: Redundant Array of Inexpensive Disks

Raid uses several hard drives to optimize speed, safety or both.

SATAVAULT
RAID-1 V.36.B05
ESC
ENT
ALARM
MUTE

The 3 main types of RAID

RAID 0: Fast, but not safe, any disk failure causes all data to be lost.

RAID 1: mirrored drives, safe (either drive can fail), but slow. Takes 2x space to store anything.

RAID 5: Uses 3+ drives, any 1 drive can fail. Clever storage saves space & is speedier than RAID 1.

RAID logic can be in:

- hardware (better, more expensive) or
- software (cheaper, not as good).

Solid State Drives

SSD drives are made from flash memory

- no moving parts to seize up…

This means they are more reliable than platter based drives.

They use SATA* interface cables.

* Apparently some now use PCIe interfaces

Outside

Inside

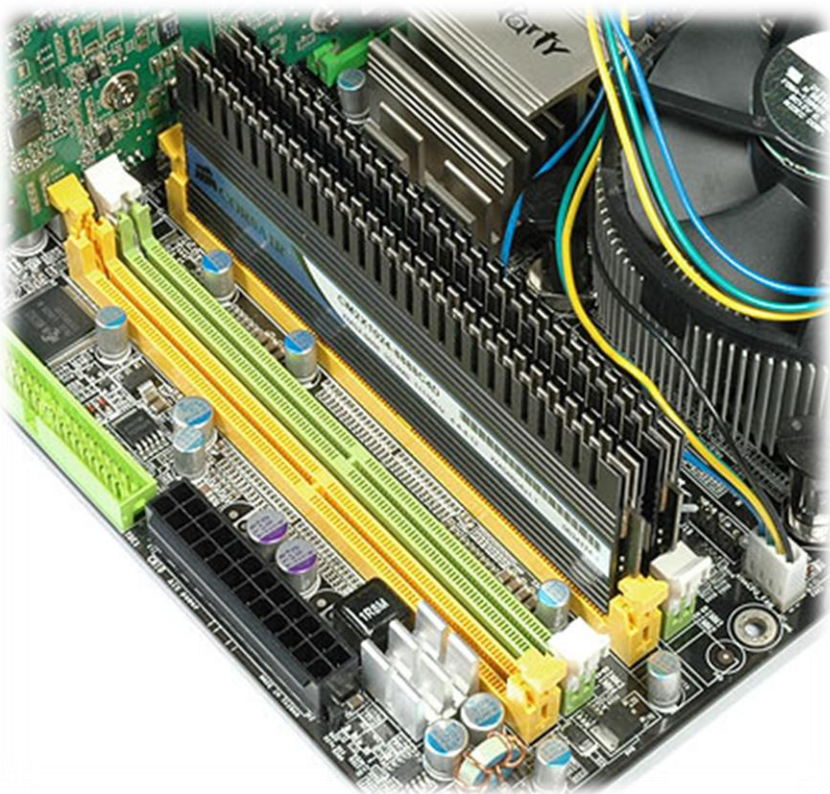
SSDs are rapidly becoming both faster and lower power use than regular HDs.

They are currently expensive, but dropping in price… and will soon be THE choice for laptops.

Summary

- Hard drives are for long term storage.
- They tend to be inexpensive, but unreliable.
- SATAs are faster than PATAs.
- Platter technology in inherently unreliable.
- RAID and SSDs help address the reliability issues.

Random Access Memory (RAM)



The more RAM you have, the more data you can manipulate at one time…
RAM is crucial to image processing.

~512 MB is very small.

3-4 GB is decent (for a 32 bit system).

You'd rather not have to share your RAM with your video card.

High end machines may have 16 to 32 GB.

What happens when there isn't enough RAM?

You can try to compensate by using some space on the hard drive for temporary storage.

This is called "swap space" a "paging file" or "virtual ram"

Heavy use of virtual memory can result in a degenerate condition where increasing resources are used to do decreasing amounts of work.

This happens because of excessive swapping of "pages" in and out of RAM.

The condition is called thrashing.

This is a BAD state of affairs, REALLY slows down data flow, and is a sign that you need more RAM.

A 32 bit platform (cpu and OS) can address 4 GB max.

64 bit systems allow 16 exabytes (16 billion GB) of RAM.

(Current 64 bit systems may have ~16 GB RAM…so there's plenty of room to grow)

Summary

RAM is like working memory

It is very important to image processing of large data sets (like dti or fmri)

- This is because we must hold lots of large images in working memory at one time in order to process them efficiently.

32 bit platforms limit the total amount of RAM the system can use.

64 bit is much better.

Data	Size
Raw 8-channel fMRI data	~1000 MB
fMRI analysis (SPM: 1 sub)	~1000 MB
Reconstructed fmri data	~200 MB
DTI (raw) (25 dirs, 3 B0s) 128x128	~50 MB
SPGR structural image	~19 MB
FLAIR 256x256 57 slices	~7.1 MB

- A CD holds ~ 700 MB
- A DVD holds ~ 4500 MB

Storage Equivalents

Raw Data from the scanner is ~200 MB per subject.

But, the 8 channel headcoil produces raw fmri data of ~1 GB per subject.

After analysis in SPM, data for one subject swells from ~200 MB to ~1 GB (maybe more)

DTI data is similar.

A CD holds ~ 700 MB.

A DVD holds about 4.5 GB (4500 MB).

Storage Amounts

Bit	Byte	Kilobyte	Megabyte	Gigabyte	Terabyte	Petabyte	Exabyte
2 states	8 Bits	8,000 Bits	8,000,000 Bits	8 billion Bits	8 trillion Bits	8 quadrillion Bits	8 quintillion Bits
	1 Byte	1,000 Bytes	1,000,000 Bytes	1 billion Bytes	1 trillion Bytes	1 quadrillion Bytes	1 quintillion Bytes
		1 KB	1000 KB	1,000,000 KB	1 billion KB	1 trillion KB	1 quadrillion KB
			1 MB	1000 MB	1,000,000 MB	1 billion MB	1 trillion MB
				1 GB	1000 GB	1,000,000 GB	1 billion GB
					1 TB	1000 TB	1,000,000 TB
						1 Petabyte	1000 PB
							1 Exabyte

Pata

Sata

RAID

RAID 0

RAID 1

RAID 5

SSD

RAM