External memory

Lecture 5

HDD, SSD, interfaces

External memory

External memory is used for storing long-term information.

External memory devices:

- Hard disc drive HDD;
- Floppy disc drive FDD;
- Optical discs (CD, DVD);
- Solid-state drives SSD
- Magnetic tapes.

External memory

Access mode:

- Direct access (HDD, CD,DVD)
- Consecutive access (tapes)

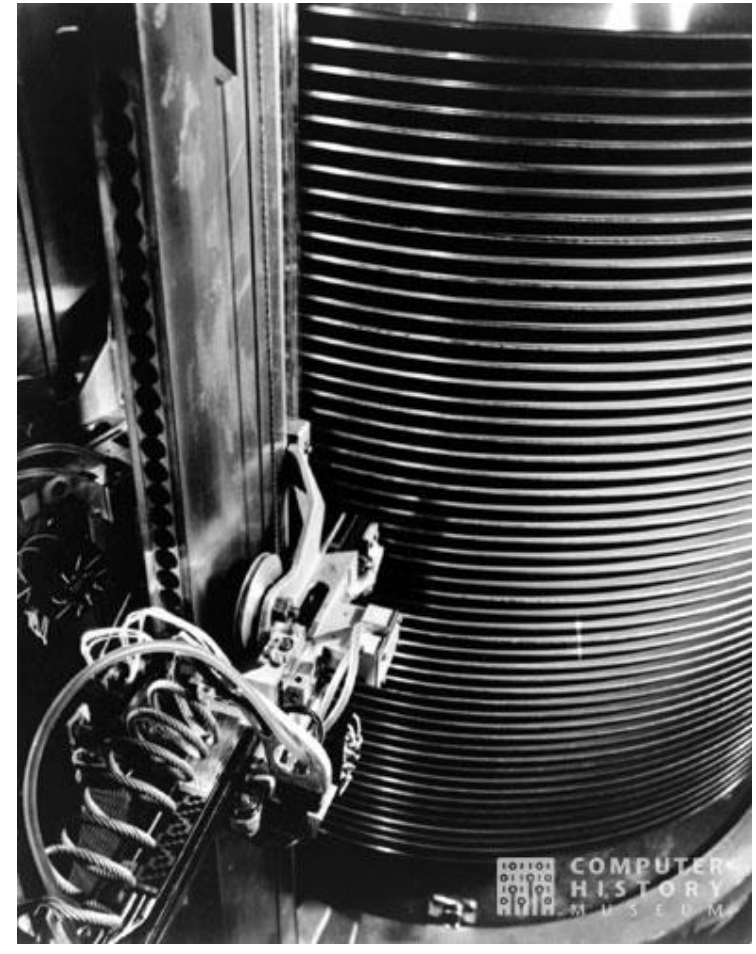
Characteristics:

- Capacity
- Latency
- Transfer rate
- Price

History

First magnetic disc 1956 m:

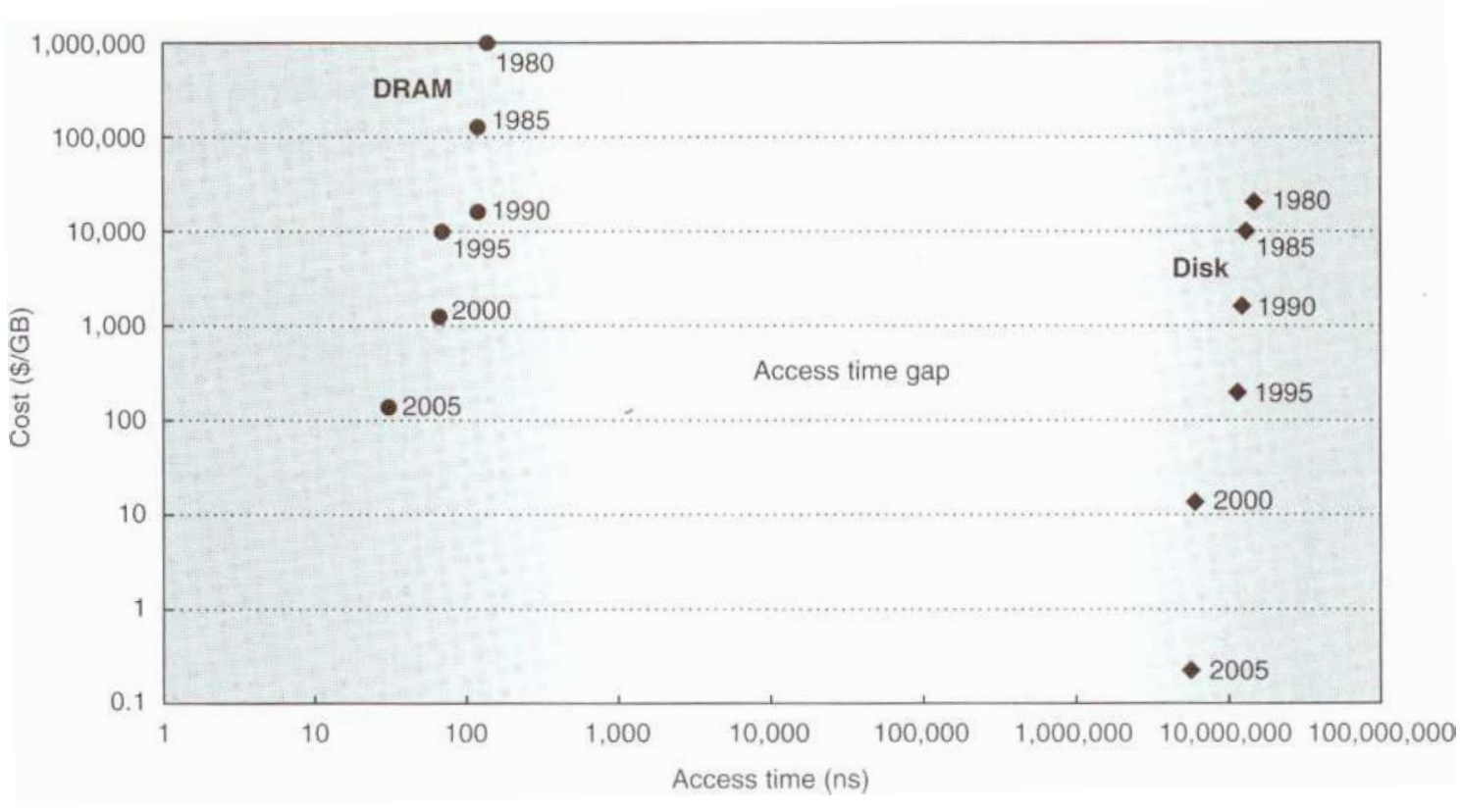
- 24” diameter,
- Total capacity 5 MB,
- 50 discs,
- 1200 rpm,
- Access time 1 s.



First HDD 1 GB, 1980 m.

- Size like freezer
- Weight 250 kg
- Price $40,000.

HDD vs RAM



Comparison of DRAM and HDD based on price and access time

Magnetic disk

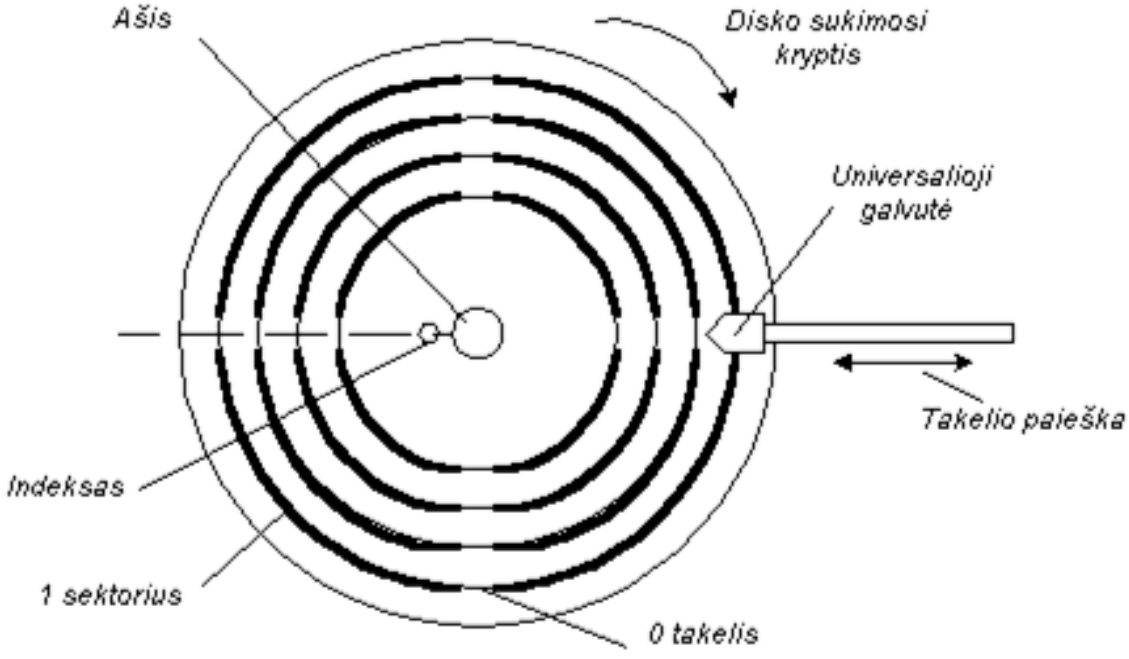
A disk is a circular platter constructed of nonmagnetic material, called the substrate, coated with a magnetizable material.

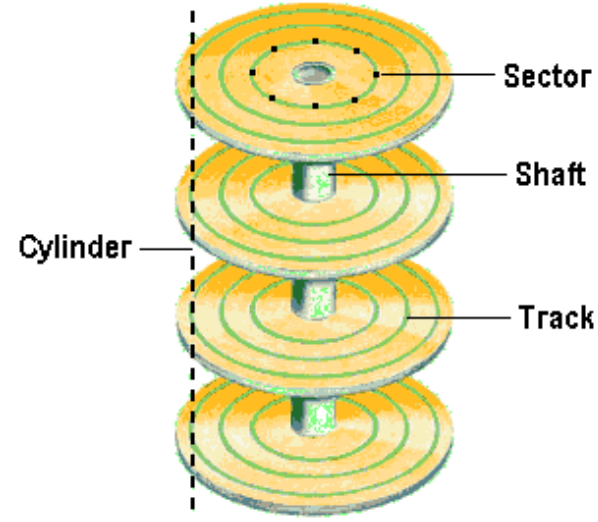
Traditionally, the substrate has been an **aluminum**

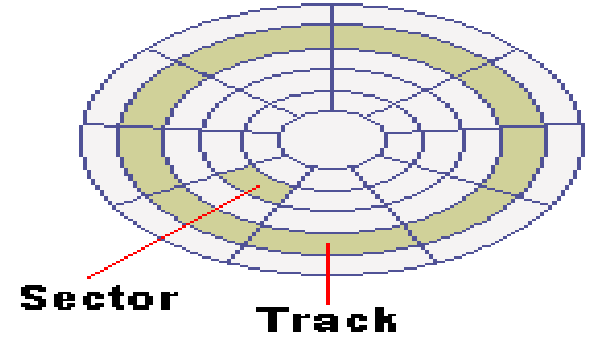
or **aluminum alloy** material.

More recently, **glass substrates** have been introduced. The glass substrate has a number of benefits like improvement in the uniformity of the magnetic film surface to increase disk reliability and etc.

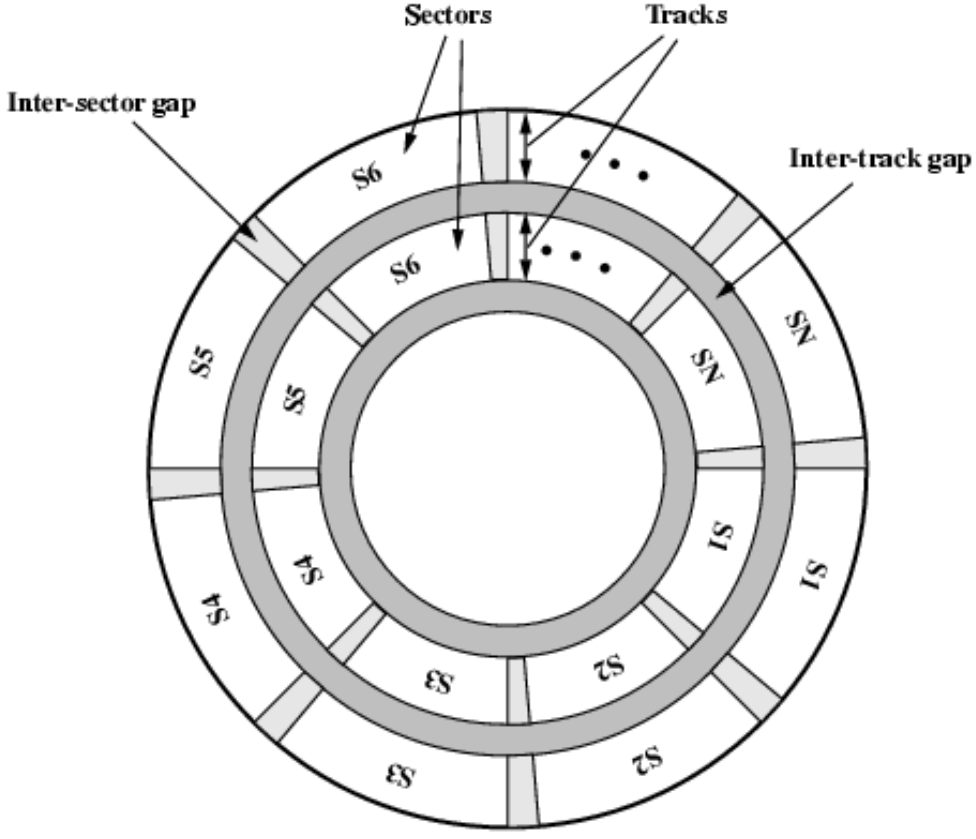
Magnetic disk

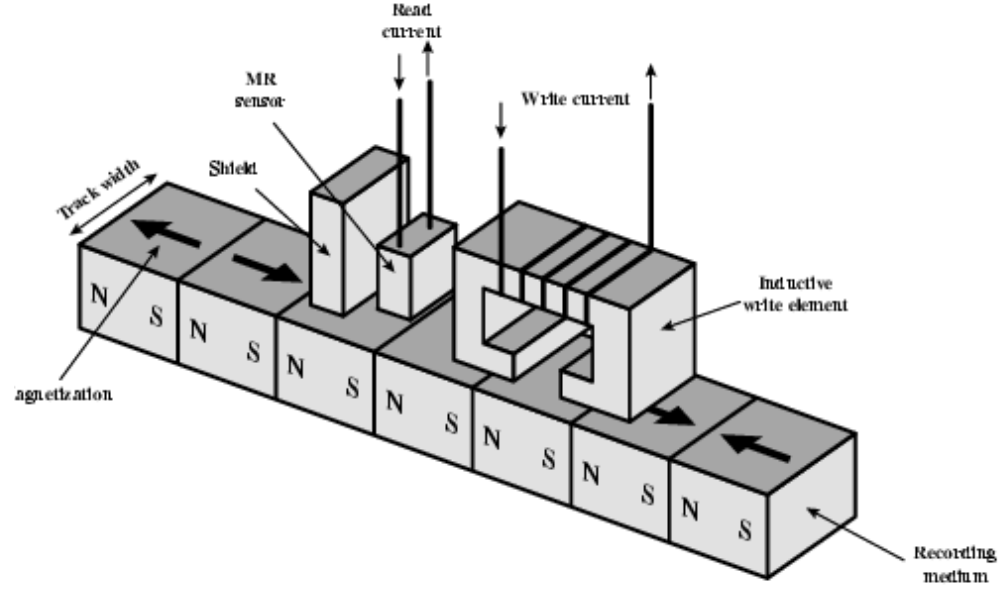






Magnetic Read and Write Mechanism





Definitions

Tracks are a concentric set of rings**.** Tracks are numbered from outside to inside**.**

Adjacent tracks are separated by **gaps**. This prevents, or at least minimizes, errors due to misalignment of the head or simply interference of magnetic fields.

Data are transferred to and from the disk in **sectors.** Size of sector in most cases is **512** bytes**.**

Clusters (allocation units) is a set of sectors**.** File allocates several clusters**.**

Magnetic disks

Control mechanism of the HHD head

HDD

HDD

Capacity = number of cylinders × number of sectors/per cylinder × size of the sector × number of surfaces

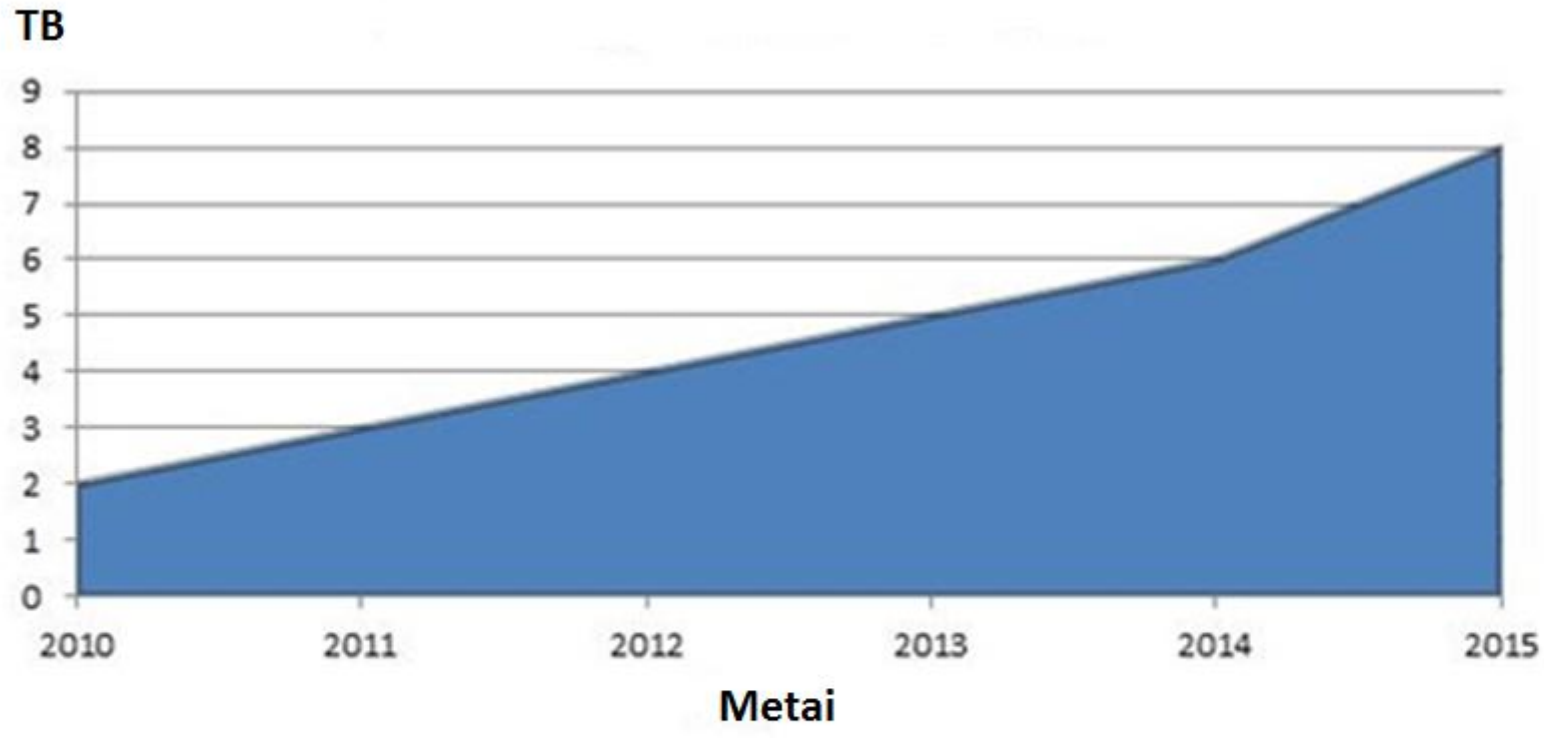
Max capacity of modern HDD is 4 TB (SATA).

Access time depends on following characteristics:

- Cylinder searching time that is positioning of the head
- Latency because of rotation
- Data transfering time

Rotational velocity of HDD - 5400, 7200, 10 000, 15 000 rpm

HDD capacity graph



Addressing system

HDD addressing systems

- CHS (***C**ylinder, **H**ead, **S**ector*)

HDD blocks are addressed using cylinder, head, and sector addresses at which data appeared on the hard disk. CHS did not map well to devices other than hard disks (such as tapes and networked storage), and was generally not used for them. CHS was used in early MFM and RLL drives, and both it and its successor, extended cylinder-head-sector (ECHS), were used in the first ATA drives.

- LBA (***L**ogical **B**lock **A**ddressing*)

LBA is a particularly simple linear addressing scheme; blocks are located by an integer index. LBA disk drives use **zone bit recording**, where the number of sectors per track depends on the track number.

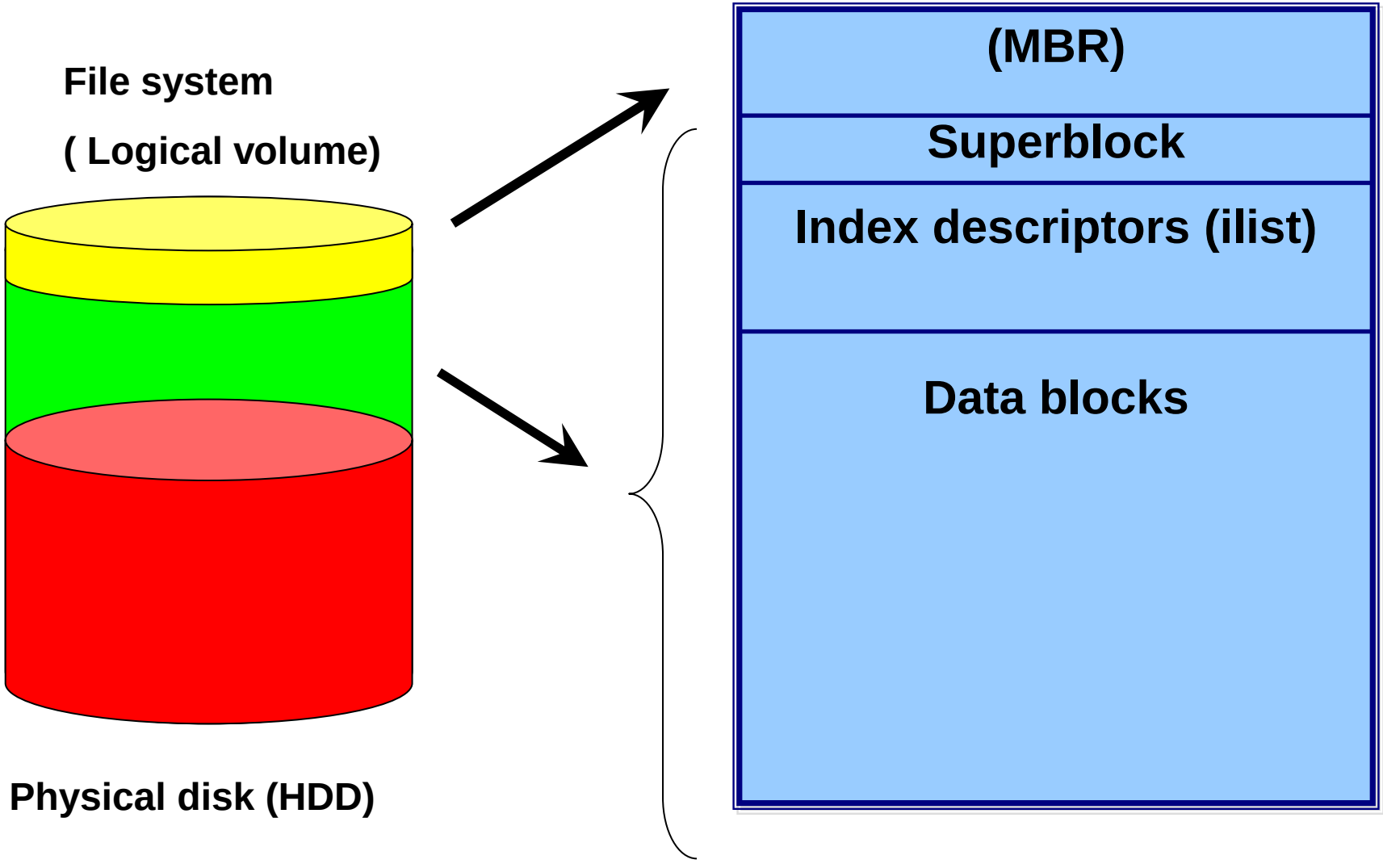
Capacity limits

Capacity limits

- ATA used 28 bits for addresses. This was supported by operating system and BIOS.

- Max number of addresses = 2^{28}
- Max number of sectors = 2^{28} = 268 435 456
- Max HDD capacity = 2^{28} sectors x 512 B = 137,4 GB

HDD logical structure



MBR

Structure of a Master Boot Record

Address			Description		Size in bytes
Hex	Oct	Dec			
0000	0000	0	Code Area		max. **440**
01B8	0670	440	Optional Disk signature		4
01BC	0674	444	Usually Nulls; 0x0000		2
01BE	0676	446	**Table of primary partitions** (Four 16-byte entries, IBM Partition Table scheme)		**64**
01FE	0776	510	55h	MBR signature; 0xAA55[1]	**2**
01FF	0777	511	AAh		
MBR, total size: 446 + 64 + 2 =					**512**

Layout of one 16-byte partition record

Offset	Description
0x00	(1 byte) Status[3] (0x80 = bootable, 0x00 = non-bootable, other = malformed[4])
0x01	(3 bytes) Cylinder-head-sector address of the first sector in the partition[5]
0x04	(1 byte) Partition type[5]
0x05	(3 bytes) Cylinder-head-sector address of the last sector in the partition[6]
0x08	(4 bytes) Logical block address of the first sector in the partition
0x0C	(4 bytes) Length of the partition, in sectors

MBR

Master Boot Record, MBR is a special type of boot sector at the very beginning of partitioned computer mass storage devices like HDD or removable drives intended for use with IBM PC-compatible systems and beyond. The concept of MBRs was publicly introduced in 1983 with PC DOS 2.0.

The MBR holds the information on how the logical partitions, containing file systems, are organized on that medium. Besides that, the MBR also contains executable code to function as a loader for the installed operating system—usually by passing control over to the loader's second stage, or in conjunction with each partition's volume boot record (VBR). This MBR code is usually referred to as a boot loader.
The organization of the partition table in the MBR limits the maximum addressable storage space of a disk to 2 TB. Therefore, the MBR-based partitioning scheme is in the process of being superseded by the GUID Partition Table (GPT) scheme in new computers. A GPT can coexist with an MBR in order to provide some limited form of backward compatibility for older systems.

HDD capacity limits

(MBR) has 32 bits for sectors addressing => 2^{32} or 2.2TB.

- Max number of addresses = 2^{32}
- Max number of sectors = 2^{32} = 4 294 967 296
- Max HDD capacity = 2^{32} sectors x 512 B = 2,2 TB

Scalable partitioning scheme GUID Partition Table (GPT) has 64-bits for addresses.

- Max HDD capacity = 2^{64} sektorių po x 512 B = 9.4 ZB

(1 ZB = 1,000,000,000 TB).

Windows Vista 64-bit and later Windows releases support **booting from GPT**, but must use UEFI (Unified Extensible Firmware Interface) firmware.

HDD characteristics

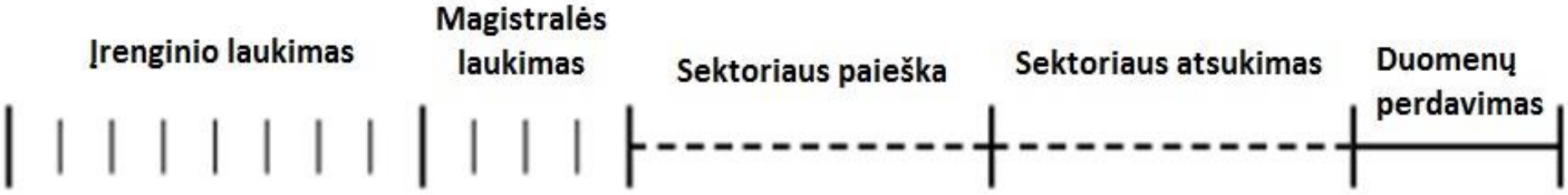
Diameter	Form	Applications
5.12"	5.25"	Old HDD (till ~1995)
3.74"	3.5"	Standard HDD for PC
3.0"	3.5"	High performance 10,000 RPM
2.5"	2.5", 3.5"	(2.5" form factor); 15,000 RPM (3.5" form factor)

HDD energy consumption

Energy consumption = Diameter^2 x RPM $^{2.8}$ x number of plates

	Capacity (GB)	Price	Platters	RPM	Diameter (inches)	Average seek (ms)	Power (watts)	I/O/sec	Disk BW (MB/sec)	Buffer BW (MB/sec)	Buffer size (MB)	MTTF (hrs)
SATA	500	$375	4 or 5	7,200	3.7	8–9	12	117	31–65	300	16	0.6M
SAS	37	$150	1	15,000	2.6	3–4	25	285	85–142	300	8	1.2M

HDD I/O



Įrenginio laukimo laikas – tai laikas, kuris sugaištamas, laukiant, kol procesorius įvykdo pertraukimo komandą.
Magistralės laukimo laikas naudojamas priėjimui prie sisteminės magistralės.
Sektoriaus paieškos laikas naudojamas nuskaitymo galvutei pozicionuoti į reikiamą cilindrą.
Sektoriaus atsukimo laikas naudojamas takelio atsukimui iki reikiamo sektoriaus.
Duomenų perdavimo laikas naudojamas fiziniam duomenų nuskaitymui iš sektoriaus.

HDD buffer

Disk buffer (often ambiguously called disk cache or cache buffer) is the embedded memory in a hard disk drive (HDD) acting as a buffer between the rest of the computer and the physical hard disk platter that is used for storage.

Modern hard disks come with 8 to 128 MB of buffer, and solid-state drives come with up to 1 GB of cache memory.

Since the late 1980s, nearly all disks sold have embedded microcontrollers and either an ATA, Serial ATA, SCSI, or Fibre Channel interface. The drive circuitry usually has a small amount of memory, used to store the bits going to and coming from the disk platter.

HDD buffer

The disk buffer is physically distinct from and is used differently from the page cache typically kept by the operating system in the computer's main memory.

The disk buffer is controlled by the microcontroller in the hard disk drive, and the page cache is controlled by the computer to which that disk is attached.

The disk buffer is usually quite small, from 8 to 64 MiB, and the page cache is generally all unused physical memory. While data in the page cache is reused multiple times, the data in the disk buffer is rarely reused.

HDD buffer

HDD buffer operates in similar way as cache:

- Write mechanisms: *write-through* or *write-back* ;
- Read mechanisms: *Read-Ahead*;
- Data replacments algorithms *LRU* or *FIFO*;
- Data block size - 512 B;

HDD buffer

In most cases FIFO replacement algorithms is used.

Waiting time = problem execution in server x I/O load / (1 - I/O load)

Example:

CPU generate HDD load 40 I/O per second.

Execution time of I/O operation is 20 ms. What is I/O load, what is average waiting time in queue, what is latency?

Solution

I/O load = 40 x 0.02 = 0.8

Waiting time = 20 x 0.8 / (1-0.8) = 80 ms

Latency = waiting time + execution time = 80 + 20 = 100 ms

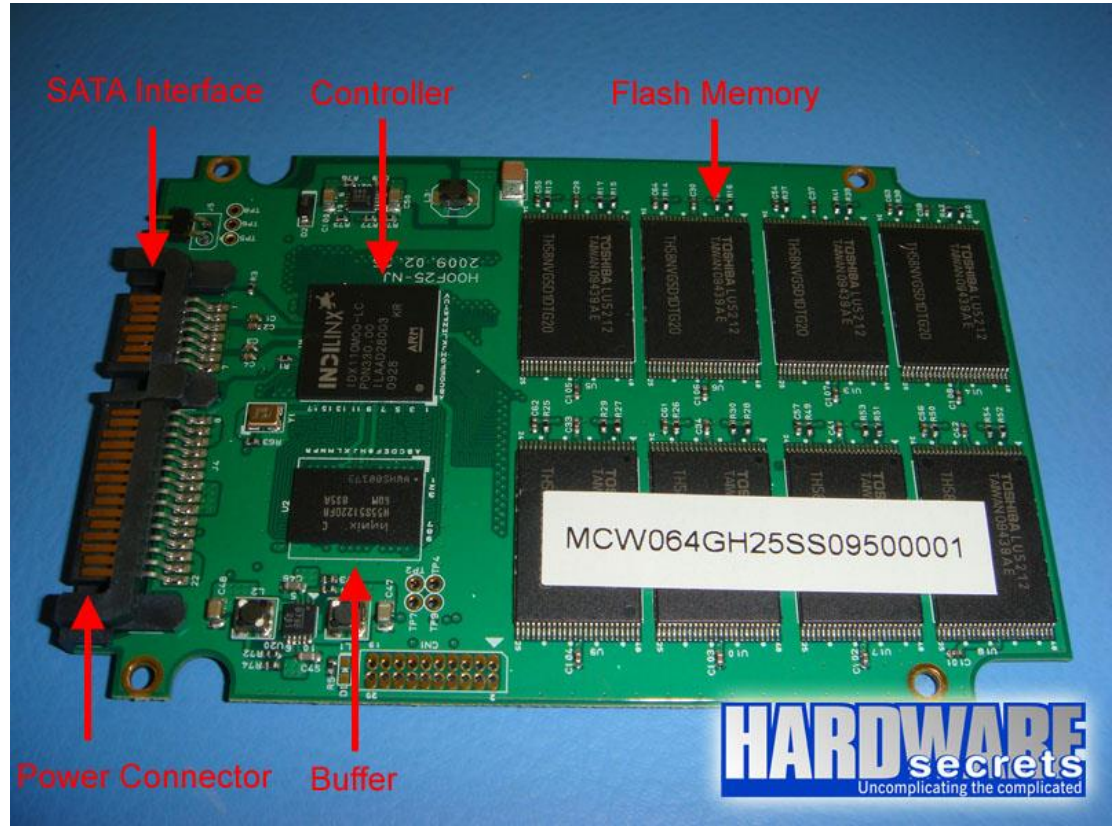
Solid state drives

A solid-state drive (SSD) is a data storage device that uses integrated circuit assemblies as memory to store data persistently. SSDs have no moving (mechanical) components.
SSD technology uses electronic interfaces compatible with traditional block input/output (I/O) hard disk drives, thus permitting simple replacement in common applications.
Additionally, new I/O interfaces, like SATA Express, have been designed to address specific requirements of the SSD technology.

Compared with electromechanical disks, SSDs are typically more resistant to physical shock, run silently, have lower access time, and less latency. However, while the price of SSDs has continued to decline over time, consumer-grade SSDs are still roughly six to seven times more expensive per unit of storage than consumer-grade HDDs.

Most SSDs use NAND-based flash memory, which retains data without power.

SSD from inside



SSD example

OCZ Colossus

- Performance 240 MB/s – read, 220 MB/s – write
- Form factor: 2,5”
- Interface: SATA 1,5 Gb/s ir 3,0 Gb/s
- Access time – 100 µs
- Capacity: 128, 256, 512, 1024 GB
- Weight: 230 gramų
- Energy Energijos sąnaudos: aktyviam režime – 0,15 W; Laikas tarp gedimų (MTBF): 1,2 mln val.

SSD storage

SSD storage RamSan 6200 (Texas Memory Systems)

- Capacity 100 TB talpą
- Performance 60 GB/s (5 mln. IOPS).
- Power – 6 kW.
- Price – 4,4 mln USD.

SSD advantages

SSD advantages:

- Better performance, short booting time
- No mechanical parts;
- Short read/write latency ~65 μs (read), ~85 μs (write);
- High transfer rate;
- Low energy consumption;
- No noise;
- The same block access time. It not depends on position of the block in memory;
- Small size and weight

SSD disadvantages

SSD disadvantages:

- Limited number of rewrite cycles (50nm MLC)
  For MLC = 10 000 times,
  For SLC = more than 100 000.
- High price
- SSD price proportional to capacity while HDD price depends on capacity but nonlineary.

MLC, SLC

MLC flash is a flash memory technology using multiple levels per cell to allow more bits to be stored using the same number of transistors.
In single-level cell (SLC) flash technology, each cell can exist in one of two states, storing one bit of information per cell.

Most MLC flash memory has four possible states per cell, so it can store two bits of information per cell. This reduces the amount of margin separating the states and results in the possibility of more errors.
Multi-level cells which are designed for low error rates are sometimes called enterprise MLC (eMLC).
MLC flash memory is its lower cost per unit of storage due to the higher data density.

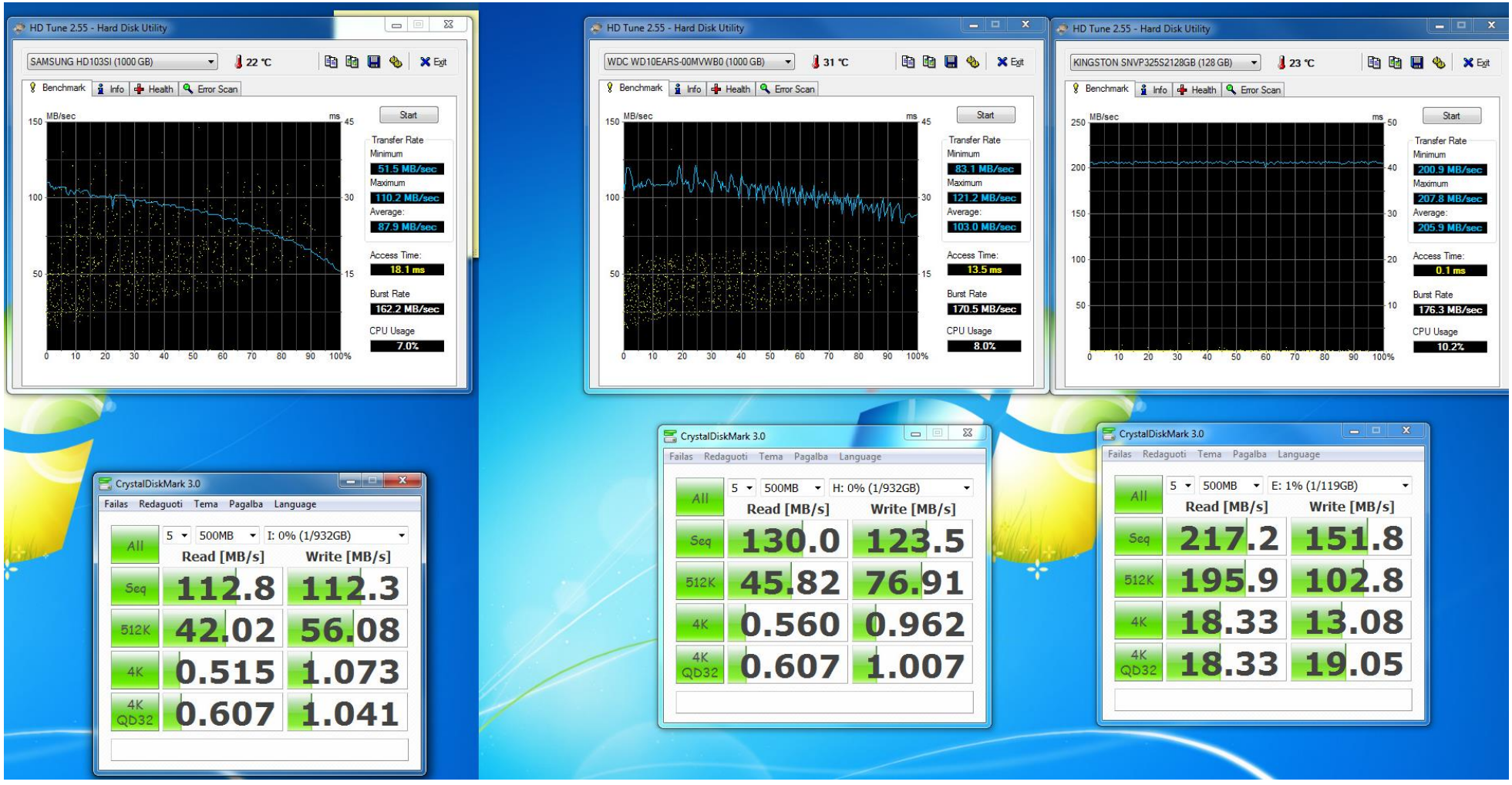
MLC ir SLC

SLC flash memory stores data in individual memory cells, which are made of floating-gate transistors.
Traditionally, each cell had two possible states, so one bit of data was stored in each cell in so-called *single-level cells*, or SLC flash memory.
SLC memory has the advantage of faster write speeds, lower power consumption and higher cell endurance. However, because SLC memory stores less data per cell than MLC memory, it costs more per megabyte of storage to manufacture. Due to faster transfer speeds and longer life, SLC flash technology is used in high-performance memory cards.

For error corrections following algorithms are used ECC (Hamming code), Bose-Chaudhuri-Hocquenghem (BCH code).

SSD vs HDD

Samsung 1TB HDD, 32 MB cache
Western Digital 1TB HDD 64 MB cache (IntelliPower)
Western Digital 128 GB SSD



HDD interface

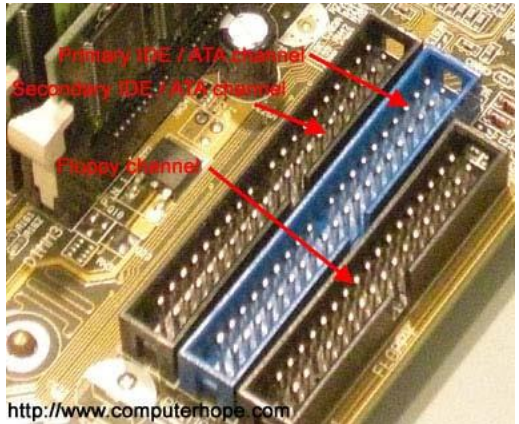
Interface types:

- IDE (Integrated Drive Electronics)
- EIDE (Exhanced Integrated Drive Electronics)
- SATA (Serial ATA)
- SCSI (Small Computer System Interface)
- SAS (Serial Attached SCSI), NL-SAS (near line SAS)
- FC – AL (Fiber Channel Arbitrated Loop)

IDE interface

IDE (*Integrated Drive Electronics*) is more commonly known as ATA or Parallel ATA (PATA) and is a standard interface for IBM compatible hard drives. IDE disc drives have controllers located on each drive, meaning the drive can connect directly to the motherboard or controller.

Max 2 HDD can be connected using IDE cable and capacity must be less than 528MB (1986 m.)



EIDE

Short for **Enhanced IDE**, a newer version of the IDE mass storage device interface standard developed by Western Digital Corporation.

It supports data rates of between 4 and 16.6 MB/s. In addition, it can support mass storage devices of up to 8.4 GB.

EIDE is sometimes referred to as Fast ATA or Fast IDE, which is essentially the same standard, developed and promoted by Seagate Technologies. It is also sometimes called ATA-2.

There are four EIDE modes defined. The most common is Mode 4, which supports transfer rates of 16.6 MBps. There is also a new mode, called ATA-3 or Ultra ATA, that supports transfer rates of 33 MBps.

ATA

Specification	*Year*	*Modes*	*Connector*	*Max transfer rate*
ATA	**1986**	**PIO 1**	**2**	**4 MB/s**
ATA-2	**1994**	**PIO 4** **DMA 2**	**2**	**16 MB/s**
ATA-3	**1996**	**PIO 4** **DMA 2**	**2**	**16 MB/s**
ATA/ ATAPI-4	**1997**	**PIO 4, DMA 2, UDMA 2**	**2 in each cable**	**33 MB/s**
ATA/ ATAPI-5	**1999**	**PIO 4, DMA 2, UDMA 5**	**2 in each cable**	**66 MB/s**

IDE

ATA / ATAPI-4 or **Ultra-DMA**, **ATA-33** ir **DMA-33**

ATA / ATAPI-5 or **ATA/66**

ATA/100 – max transfer rate – 100 MB/s

ATA/133 – max transfer rate – 133 MB/s

PIO Mode	Data Transfer Rate (MBps)	Standard
0	3.3	ATA
1	5.2	ATA
2	8.3	ATA
3	11.1	ATA-2
4	16.6	ATA-2

Serial ATA

Serial ATA (SATA) is a computer bus interface that connects host bus adapters to mass storage devices such as hard disk drives and optical drives.

Serial ATA succeeded the older Parallel ATA (PATA) standard offering several advantages over the older interface:

- reduced cable size and cost (seven conductors instead of 40 or 80)
- native hot swapping,
- faster data transfer rate
- higher signaling rates
- more efficient transfer through an (optional) I/O queuing protocol.

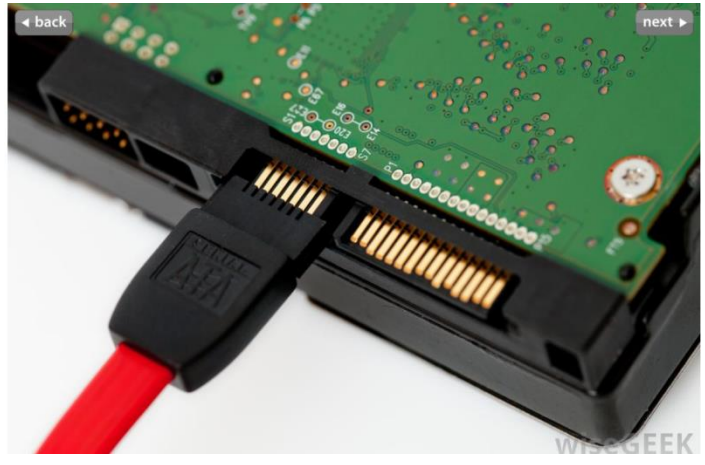
.

Serial ATA

SATA host adapters and devices communicate via a high-speed serial cable over two pairs of conductors. To ensure backward compatibility with legacy ATA software and applications, SATA uses the same basic ATA and ATAPI command-set as legacy ATA devices.

SATA has replaced parallel ATA. Serial ATA industry compatibility specifications originate from the Serial ATA International Organization (SATA-IO).

SATA1, SATA2, SATA3



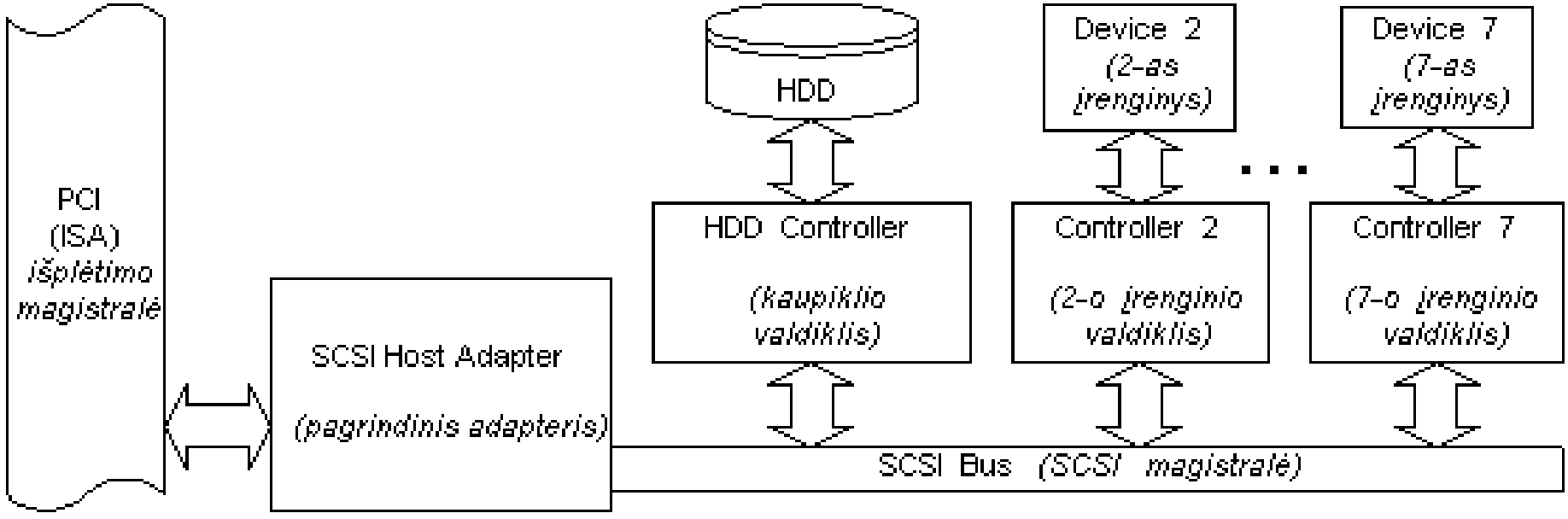
SCSI

Small Computer System Interface (SCSI) is a set of standards for physically connecting and transferring data between computers and peripheral devices.

The SCSI standards define commands, protocols and electrical and optical interfaces. SCSI is most commonly used for hard disks and tape drives, but it can connect a wide range of other devices, including scanners and CD drives, although not all controllers can handle all devices.

Up to 8 or 16 devices can be attached to a single bus. There can be any number of hosts and peripheral devices but there should be at least one host. SCSI uses handshake signals between devices.

SCSI



SCSI-1, SCSI-2 have the option of parity error checking.
Starting with SCSI-U160 (part of SCSI-3) all commands and data are error checked by a CRC32 checksum. The SCSI protocol defines communication from host to host, host to a peripheral device, peripheral device to a peripheral device.
However most peripheral devices are exclusively SCSI targets, incapable of acting as SCSI initiators—unable to initiate SCSI transactions themselves.

SCSI

SCSI-1 (or SCSI), standartizuotas ANSI 1986 m. Pralaidumas asinchroniame režime 1.5 MB/sec, sinchroniniame – 5 MB/sec. 8-bit pločio kanalas, naudojamas 50 adatų jungtis.

SCSI-2 turi tokius pagerinimus, lyginant su SCSI-1: didesnis pralaidumas, platesnė magistralė, didesnis patikimumas, geresnis pariteto skaičiavimas. SCSI-2 duomenų perdavimo greitis nuo 5 MB/sec iki 10 MB/sec. Naudojamas pavadinimas Fast SCSI-2. SCSI-2 kanalo plotis nuo 8 bits iki 16 bitų. Toks pakeitimas pažymimas kaip Wide SCSI. Sujungiant Fast SCSI-2 su Wide SCSI pasiekaimas pralaidumas iki 20 MB/sec.

SCSI-3 pagrindiniai privalumas: didesnis perdavimo greitis, palaikoma iki 32 įrenginių vienoje grandinėje ir taip pat palaiko serial jungtis.
Serial jungimai leidžia SCSI-3 panaudoti tokiose technologijose:
Serial Storage Architecture (SSA), Fibre Channel ir IEEE P1394 (FireWire).
Nuoseklus (serial) perdavimo rėžimas leidžia gauti didesnė perdavimo greitį, prijungti daugiau įrenginių, supaprastina jungtį, leidžia naudoti ilgesnius kabelius.

SCSI specifications

Specification	***Freq (MHz)***	***Transfer MB/s, Max***	***Width***	***Devices***
SCSI-1	5	5	8	8
Wide SCSI	5	10	16	8
Fast SCSI	10	10	8	8
Fast Wide SCSI	10	20	16	16
Ultra SCSI	20	20	8	8
Ultra SCSI	20	20	8	4

SCSI specifications

Specification	***Freq (MHz)***	***Transfer MB/s, Max***	***Width***	***Devices***
Wide Ultra SCSI	20	40	16	16
Wide Ultra SCSI	20	40	16	8
Wide Ultra SCSI	20	40	16	4
Ultra2 SCSI	40	40	8	8
Wide Ultra2 SCSI	40	80	16	16
Ultra3 SCSI (ULTRA 160)	80	160	16	16
Ultra320 SCSI	160	320	16	16

Serial SCSI

Serial SCSI:

- SSA (Serial Storage Architecture)
- SAS (Serial Attached SCSI)
- FC-AL (Fiber Channel)

SAS

Serial Attached SCSI (SAS) is a point-to-point serial protocol that moves data to and from computer storage devices such as hard drives and tape drives.
SAS replaces the older SCSI bus technology.

SAS, like its predecessor, uses the standard SCSI command set.

SAS offers backward compatibility with SATA, versions 2 and later. This allows for SATA drives to be connected to SAS backplanes. The reverse, connecting SAS drives to SATA backplanes, is not possible.
SAS support up to 16384 devices in one group and transfer rate up to 12 Gbit/s.

SAS connectors

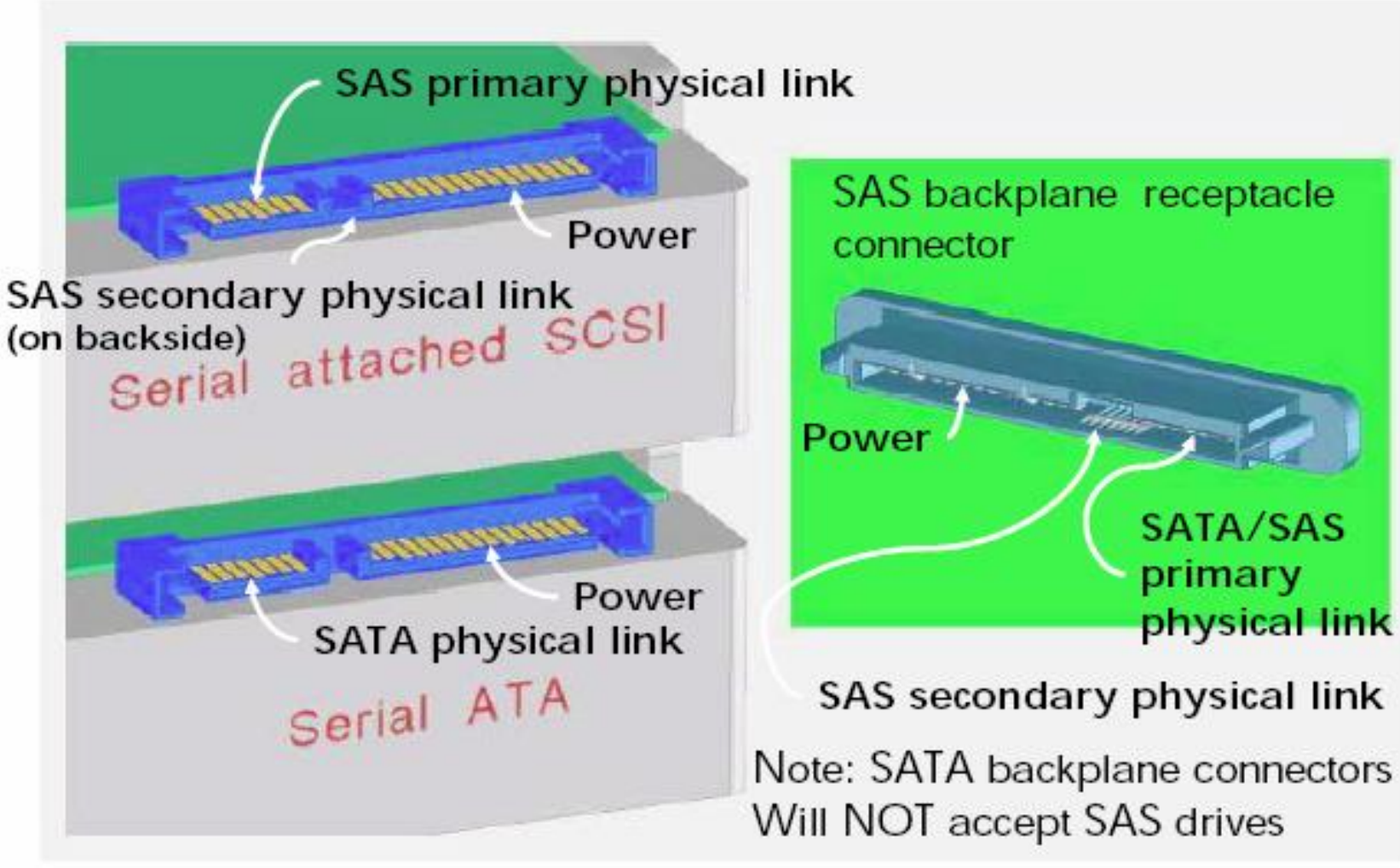
3 types of SAS connectors:

- SFF 8482 – small form factor, compatible with SATA

- SFF 8484 – middle form factor

- SFF 8470 – high density connector, compatible with Infiniband.

SAS


SAS primary physical link
Power
SAS secondary physical link
(on backside)
Serial attached SCSI
Power
SATA physical link
Serial ATA
SAS backplane receptacle connector
Power
SATA/SAS primary physical link
SAS secondary physical link
Note: SATA backplane connectors
Will NOT accept SAS drives

SAS

Serial Attached SCSI support the following protocols:

- Serial SCSI Protocol (SSP) — support *SAS disks*
- Serial ATA Tunneling Protocol (STP) — support *SATA disks*
- Serial Management Protocol (SMP) — *SAS control*