3 June 2004

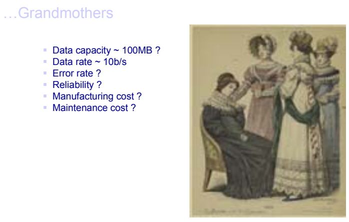
Disk Drive Science

Steve Legg

This paper, which accompanies the lecture slides, is a narrative examination of the history of disk drive science, comprising four short chapters: These chapters explore some of the evolutionary and commercial pressures as well as the scientific and engineering considerations that have led to the current state of disk drive technology.

Introduction to data storage

THE FIRST 200,000 YEARS - RELY ON GRANDMA



People have wanted to record information, be it art, commercial records or personal experiences - since people have been around. Before the discovery of media to do this with (such as cave walls, clay tablets or paper) people simply had to rely on the data storage capacity of their own heads. The elders of the society acted as the data storage repositories. Since in many societies, women who had survived past child bearing age tended to last longer than their more belligerent mates, it is fair to consider Grandmothers as the data storage medium of choice for most of the history of the human race.

To set some benchmarks for more modern data storage devices, some estimates of the parameters of a person as a data storage device are:

DATA STORAGE USING “MEATWARE”

Storage capacity: Brute-force capacity estimates based on the number of neurons in a brain, or the number of synapses between them [1] are flawed in that they are based on the assumption that brain storage is analogous to the kind of RAM that we are used to in computers. There is evidence [2][3] that the brain is far less like RAM; storing memories in redundant ways that involve many neurons per “memory”, while at the same time involving each neuron in many “memories”. A good estimate of the storage capacity of the human brain was made by Landauer, by treating the acquisition of memories as a communications process [4]. The upper limit of capacity that was reached by this estimate was of the order of 100MB - the kind of raw capacity that today we carry in a digital camera, or an MP3 player. In terms of current state of the art disk capacities, this would enable the total memories of the student population of Southampton University to fit on half a dozen hard disk drives.

Data rate: Storing data is one thing - getting it back from the storage medium is another. Again from the approach that he took to his estimate of capacity, Landauer concluded that people can acquire memory at the rate of a couple of bits per second. Replaying existing memory is a little faster - speaking at 50 words per minute with perhaps a word encoded as 10 - 15 bits yields a data rate approaching 10 bits per second. Neither of these rates look high compared with disk data channel rates of 40 megabytes per second - typical for a medium performance disk drive today.

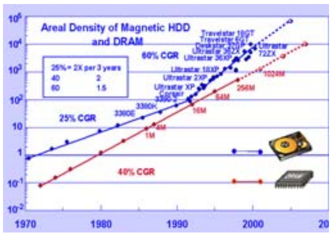
Error rate, reliability, manufacturing cost etc. It is not necessary to belabour the comparison, simply to point out that when it comes to memorising the kind of data that current commerce, science and engineering need, disk storage is probably a better way to do it.

Data storage with disk drives

HOW DID WE GET STARTED WITH DISK DRIVES?

It is almost fifty years since the introduction of the magnetic disk drive (the “Random Access Method for Accounting and Control” or “IBM 305 RAMAC”[5], designed and manufactured by IBM in 1953 was the first computer offered for sale with magnetic disk storage). As an interesting aside, it was the somewhat arduous journey from IBM head office in Armonk, NY to San Jose, CA that allowed a group of engineers in San Jose to work on what was essentially a bootleg project. If the flight across the continent had not taken 12 hours and been expensive, they would have got more management attention sooner. The effect of that quite likely would have closed down their project in the early stages. The plan was to paint some aluminium disks (50 of them, each 2 feet in diameter) with the iron oxide primer that was (and indeed still is) used to paint the Golden Gate bridge in San Francisco. The bridge, and it’s characteristic paint being very familiar to the engineers in San Jose, being only an hour’s drive North along US101. A read / write head similar to those used in a tape recorder would then be positioned over the spinning disks, and moved radially by a mechanical actuator. This entire arrangement weighing around a ton could store the huge capacity of 5 million characters (bytes hadn’t been invented then, these were 7 bit characters in ‘EBCDIC’ code), or 4.4MB in today’s terms.

WHY ARE WE STILL USING DISK DRIVES?



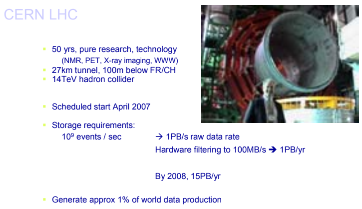
Looking a little further along the development it is interesting to plot the “areal density” (that is the number of bits per unit area) for both disk drives and for silicon based storage (DRAM)[6]. Plotting the areal density of these two technologies on a log scale versus time, it is no surprise to see that DRAM density is a straight line - popularised as “Moore’s Law” after the observation made 40 years ago by the Intel co-founder that the ability to pack features onto the surface of silicon doubles every 18 months [7]. It is worth noting that as memory density was following Moore’s Law, hard disk areal density was growing at a slower (though still exponential) rate with a compound annual 23% increase during the 1970’s and 1980’s. If this trend had continued, then the two lines would have crossed in the mid 1990’s. The raw materials for both of these technologies are relatively cheap (sand for silicon chips and aluminium and rust for disk storage), so the cost of the products is dominated by the processing of these materials. As soon as the density of one exceed that of the other, then the more dense one would dominate the market. Of course disk storage still had one significant advantage - that it is non-volatile (it retains its memory when the power is switched off), however the relentless density increases of DRAM are also matched by the non-volatile silicon technologies (EEPROM and FLASH memory), these would surely have overtaken disk storage also.

If these lines had crossed, disk drive technology would have gone the way of the thermionic valve, relegated to niche applications.

The thing that caused the growth rate for areal density of disk drives to pick up again in the mid 1980's was the introduction (by IBM) of disk read heads based on the Magnetoresistive effect and the Giant Magnetoresistive effect - so called "MR" and "GMR" heads. This allowed the area density growth for disk drives to resume at 60% CGR. In the time between then and now, this has increased the density of disk storage to be two orders of magnitude higher than silicon storage. The Magnetoresistive effect is explained by work on electron scattering in solids carried out by Sir Nevill Mott at Cambridge in the 1940's and 1950's [8] and for which he (with two others) was awarded the Nobel prize for physics in 1977.

Data storage demand

SO WHO NEEDS ALL OF THIS STORAGE?

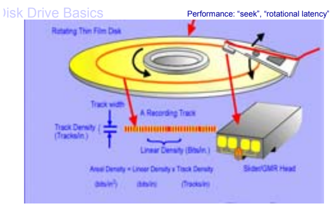


Estimates of the total amount of data stored on disk drives vary, but a good proportion of the total unique knowledge that the human race has is stored (or storable) on magnetic disks [9]. Total disk capacity shipped by all manufacturers currently averages around 190PB per quarter [10] or a little under 1EB per annum. The rate of growth of capacity shipped has recovered from the slow-down of 2001 / 2002 and is now back to the 30% - 35% compound annual growth that has been seen for the past decade or more. It is worth noting that while the *capacity* of disk storage shipped has increased continually, the *revenue* from this disk capacity has remained flat, or even declined slightly - such is the fierce competitive pressure on price per MB.

Individual large users of disk storage are typically in the engineering and scientific areas, such as oil exploration companies (each with a few PB of disk storage installed) and research institutes working on genomics and protein folding. At the extreme of the engineering and scientific scale, CERN's Large Hadron Collider experiment [11] which is being prepared to start operation in 2007, will yield raw data at the rate of 1PB/s. These raw data will be filtered in hardware to look for "interesting" events, taming this torrent of data into a more manageable deluge for further analysis. The resulting stream of data that will need to be stored on disk being closer to 100MB/s or a "couple of petabytes per year".

Disk technologies - aerodynamics and magnetics

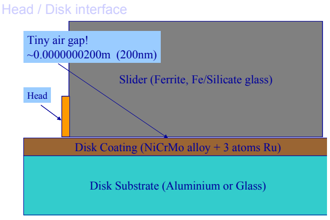
UP CLOSE AND PERMANENT

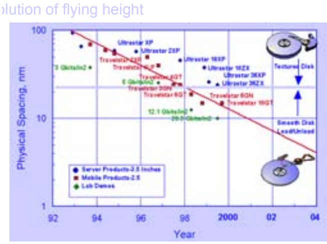


On a disk drive platter, data is recorded in concentric tracks, accessed by moving an arm (a.k.a. "actuator") to position a read / write head over the track. As the disk rotates, recorded domains in the track pass under the head where they are detected as bits and decoded. Areal density in disk drives is driven by having narrow tracks, and short bit lengths. These two parameters are in their turn driven by the ability of the read / write head to fly very close to the surface of the recording medium. Of course the ability of the head to write and read back very small bits would be of little use unless they can be made non-volatile, so the medium must have moderately high coercivity, and high remanence. The early magnetic materials used as disk coatings were based on iron oxides with H_c around 50A/m. Today, metallic films are used, with H_c values 20 times or more greater than that of iron oxides. While this allows smaller track widths and bit lengths, it places new challenges on designing heads that can reliably write bits onto such media.

Disk drive design provides engineering challenges in a number of different disciplines, from aerodynamics to vibration and many in between. For the purposes of this narrative, only two aspects will be considered further - how to make the read / write head fly close to the disk, and how to read back data that has been written.

FLYING AT LOW ALTITUDE





The write head of a disk drive is essentially a toroid of magnetically soft material with a gap in it and a number of turns of wire around it. The fringing field from the gap is allowed to 'leak' into the surface of the recording medium on the disk, causing it to become magnetised. To achieve bit lengths that are close to the size of the gap in the write head, it is necessary to have the write head to recording medium spacing much less than the gap width. The basic physics of this was well known by the time the first disk drives were designed, having been calculated originally by the Swedish physicist Olle Karlqvist in 1950[12]. It is notable that his calculations are based on magnetic drum recording since at that time no-one had considered using disks. The change of geometry from drums to disks could of course have been anticipated since the early Edison wax cylinder had been displaced by disks for sound recording for the same reason - volumetric efficiency.

Very small spacing then between the write head and the disk was the key to high capacity. It was clear early on however that allowing the head to touch the disk surface was not a good thing. Since the materials from which the heads were made were hard (typically ferrites or other glasses) and the coating on the disk itself was thin, allowing the two to come into contact resulted in instant, permanent loss of data. The head then was required to fly above the disk surface, at a controlled height. Since it is hard to manufacture a disk to be perfectly flat and smooth, read / write heads that were fixed at a certain distance from the disk were limited in how low they could "fly". The breakthrough was to hold the read / write head against the disk surface and control its separation by the use of an "air bearing". Air bearings that were externally pressurised were used first in the IBM RAMAC, and allowed the head to fly just 20µm above the disk surface. The requirement to increase density drove the need to fly lower, which challenged the design of externally pressurised bearings. The concept of a "self acting" air bearing, where the shape of the underside of the head itself created the pressure for the bearing was introduced in the IBM 1301 disk drive in 1962. This allowed the head to fly as low as 6.4µm [13]. In the intervening 30 years, head design and the use of smoother substrates and coatings for disks have permitted flying heights to be reduced to 20nm in common disk drives. This height is considerably less than the mean free path of a molecule in air - making it hard to imagine what is meant by "flying" at this altitude.

HEAD CRASH - WHEN IT ALL GOES HORRIBLY WRONG

Of course a very thin air bearing means that the disk enclosure, the metal case that holds the disk platters, motor and head assemblies, must be kept free of particles that could get stuck in the air bearing gap. If a particle does get into the gap, it tends to abrade other particles from the disk surface, which may cause further contamination and leading to a catastrophic failure - a so called "head crash". A drive which has failed in this way is readily identified since it makes a hideous squealing noise in the last few hours of its life. When opened the disk enclosure is full of black dust - the remains of the ferrite heads, and bits of the disk coating. Needless to say the data that was on the drive is not recoverable after this has happened.

To avoid this, disk drive components are carefully cleaned prior to assembly. Assembly itself is done in clean rooms with air filtered to a level around 100X cleaner than an operating theatre[1]. A small filter is installed in each drive to catch any residual particles, and trap them so that they cannot contaminate the air bearing(s). Opening a disk drive outside of such conditions leads to its inevitable death from a head crash - though most drives will function for a short while after limited contamination.

Making aluminium disk platters that were coated with a magnetic layer smooth enough to allow lower and lower flying heights required more and more elaborate treatments. Initially pressed disks were flat enough. Later, as flying heights became lower, disks were milled, then ground and finally diamond turning was required to get an extremely smooth surface prior to coating. The flying height limit for aluminium disks was reached at around 200nm, and disk drives in the last 10 years have used glass substrates for their platters since these can be made very smooth by annealing.

DON'T LOSE YOUR HEAD

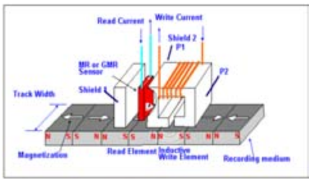
One of the consequences of *very* smooth disks and heads however is stiction - the tendency for two surfaces that are smooth at an atomic level to weld together when they come into contact. This difficulty arises when the head stops flying - for example when a disk drive is powered off. The heads "land" on the surface, and weld to it. The disk spindle motor has to provide additional torque on start-up to overcome this stiction. The next time the disk is started, one of two things happens. Either the head drags small particles of the disk coating away when the disk starts to move (this is a bad thing since these particles on its underside now make it rough and it ploughs furrows in the smooth disk surface, filling the disk enclosure with debris and contaminating the other heads' air bearings). Alternatively the head detaches from its suspension as the motor supplies enough torque to tear it away (this is also a bad thing since it renders the drive useless). Managing stiction is a major challenge at low flying heights and it has been tackled in a number of ways. One is to deliberately roughen a "landing zone" an annular area in the centre of the disk, onto which the head is allowed to land. The actuator can be made to always move the head to this area as the disk spins down on power off. Indeed most designs use the residual rotational energy of the disk with the spindle motor acting as a generator to provide the electrical power to move the actuator to the landing zone. This caters for the situation where electrical power to the drive is lost suddenly before the head has been moved to the landing zone.

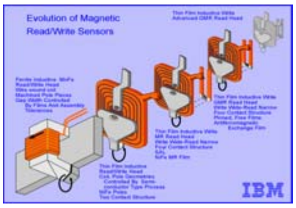
The use of a landing zone however consumes some of the area that could otherwise be used for data storage. It also takes a further preparation step on the disk platter after it has been coated. Many current drives use an alternative technique which is to arrange for the actuator to move the head off of the disk altogether prior to the disk slowing down [14]. The increase in mechanical complexity of such an arrangement is more than offset by the saving in disk surface area, reduction in process steps, and the lower flying height achievable.

1. Typically disk assembly is in class 10 cubicles in a class 100 cleanroom. Class 100 refers to the air having fewer than 100 particles of > 0.5µm per cubic foot(sic).

WRITE THIS DOWN

agnetic Recording - schematic



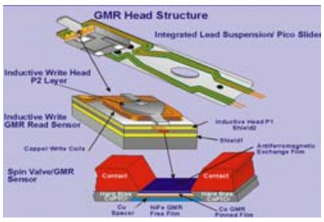


Having arranged to fly a read / write head over the disk surface, the next challenge is to record data bits permanently (but erasably) on the magnetic medium. Early read / write heads were a toroid of soft magnetic material (typically ferrite) with a small air gap, and a coil of wire around the toroid. The fringing field from the gap magnetises the recording medium on the disk in an orientation depending on the direction of current flow in the head coil. The same coil was used for writing (by passing a current through to induce a magnetic flux in the toroid) and for reading (by sensing the voltage induced as the coil passed over a recorded bit). These two functions tended to place conflicting demands on the design of the read / write head. To make it possible to write quickly, the write coil needed few turns since its inductance was the limiting factor in changing the magnetic flux quickly. Conversely to make a sensitive read head needed many turns, since the sensing of a bit was made by measuring the induced voltage on the head.

Several generations of read / write heads came and went - each a compromise between write bandwidth and read sensitivity. Initially the heads were small fabricated ferrite cores, with hand wound coils. Later the coils were printed using photolithographic techniques, and the magnetic pole pieces were fabricated using similar techniques to those used for fabricating features on integrated circuits - so called “thin film” heads.

The breakthrough in 1991 (that put the kink in the areal density curve) was the separation of the write element in the head from the read element, and to base the read element on an entirely different technology - the magnetoresistive effect (“MR heads”)[15]. These sensed the bits on the disk medium by a change in resistance of a thin metal (CoFe) film under the influence of the tiny magnetic field from the bits. Apart from the ability to optimise the write head and the read element independently, the other advantage of separating these was to allow them to be made different widths. Having a wide write head and a narrower read head, allowed the tracks to be squeezed closer together since it reduced the chance of reading part of one track and part of the adjacent one.

ENTER THE GIANT

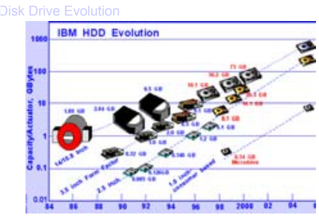


The “giant magnetoresistive” (GMR) effect was discovered independently in the late 1980s by Peter Gruenberg in Germany, and Albert Fert in the University of Paris-Sud. They both saw large resistance changes in materials comprised of alternating very thin layers of various metals when these films were subject to an external magnetic field. Their experiments only worked however at cryogenic temperatures and very high magnetic fields. Their techniques for producing these films were difficult to reproduce, and not suitable for manufacturing. IBM researchers at the Almaden research lab. in San Jose CA tried a quicker and less precise method for producing these films and discovered a method that would work with common manufacturing processes, and at room temperature[16]. The resultant structure, called a “spin valve”, comprises a thin layer of a metal such as copper sandwiched between two layers of an alloy (such as NiFe) which exhibits the GMR effect. When the magnetizations of both of the GMR layers are parallel, the spin valve exhibits a lower resistance since electrons of one spin orientation (either “up” or “down”) are not scattered. When the GMR layers are magnetised in antiparallel (each layer in an opposite direction) all electrons are scattered (the spin-up ones by one layer and the spin-down ones by the other). This results in a higher resistance to current flow (Mott’s two-current theory).

To create a practical spin valve, the layers are fabricated so that the magnetisation orientation of one of them is “pinned” - typically by exchange coupling to an adjacent antiferromagnetic layer (such as FeMn)[17]. The other GMR layer is free to orient itself to the magnetic field that it is in (the so-called “free” layer). As this structure passes over a magnetic bit recorded on the surface of the disk, the free layer magnetises in the direction of the field from the bit. If this is parallel to the magnetisation direction of the pinned layer, a lower resistance is seen, if it is antiparallel, a higher resistance. Thus the spin valve is able to discriminate between oppositely magnetised domains that were recorded earlier on the disk surface.

Such a read sensor is narrower, far more sensitive, and has far higher bandwidth than an inductive read head. This permits narrower tracks and smaller bit lengths, contributing to areal density growth.

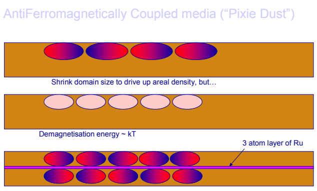
DRIVE EVOLUTION



The net effect of these technology advances has been to drive more an more capacity per disk drive. One consequence has been the evolution of drive form factors. Initially to get sufficient capacity per “spindle”, large disks and several of them in a stack needed to be contained in one disk enclosure. As density grew there were more data accessed by only one actuator, and the mechanical performance of that actuator (the number of “Input / output Operations Per Second” or “IOPS”) became the limiting factor. A disk drive has to deliver both capacity (MB, or GB) and ability to access that capacity (IOPS). Most commercial applications have a certain characteristic “access density” measurable as a number of IOPS per GB. If a disk drive cannot deliver this access density it is necessary to either make the actuator faster or to make the capacity smaller (by making a physically smaller disk drive with the same areal density).

Making actuators faster is limited by how much power you can put into the actuator (which is a specialised motor) before I^2R losses cause it to heat up too much. Power dissipation in a disk drive goes up as a little over the fourth power of the disk diameter (actually $P \alpha D^{4.6} \times R^{2.8}$ where D is the diameter and R the rotational rate[6]). Making disk drives smaller helps in a number of ways: Smaller disks mean faster actuators since the actuator has to move less far. Smaller disks may be spun more quickly since the frictional losses go as the square of the disk radius. Smaller disks have a larger surface area to volume ratio and may be cooled more easily (the same reason why elephants need cooling fins [ears] to lose heat and mice need fur to retain it). All of these factors have driven the evolution from large disks (RAMAC’s 24” diameter disks) through to the 3.5” diameter disks that are most common today. For lower power applications still smaller disks are used, 2.5” disks for laptop computers and 0.9” disks in the Hitachi Microdrive - an entire 4GB disk drive the size of a book of matches.

A PINCH OF PIXIE DUST



What are the limits for magnetic recording on disks? As track widths and bit lengths become smaller, thermodynamics creeps in like an uninvited guest to spoil the party. There is a certain amount of energy required to “flip” a magnetic domain on a disk surface from one orientation to the other (from a “0” to a “1”). When this amount of energy becomes comparable to the thermal energy that a domain has at room temperature (kT), then domains will spontaneously flip orientation at random. This so-called “superparamagnetic limit” seemed like a ceiling on the areal density that could be achieved.

Another breakthrough in materials science from IBM allowed this ceiling to be raised yet again. This time, by putting a very thin layer (three atoms thick) of a diamagnetic material such as Ruthenium under the surface of the disk coating[18], the magnetic energy of a bit could be increased - avoiding spontaneous demagnetisation. The fact that such a dramatic effect could be caused by just a three atom thick layer led to it being described as magic - like sprinkling “pixie dust” onto the disk to make it work.

BEYOND MAGNETIC RECORDING

There will be a limit however to the areal densities that can be achieved by magnetic recording on disk drives, and that limit is expected to be of the order of 60 - 70 Gb/ sq. inch., or around 1TB in a 3.5” form factor disk drive. Other technologies to store data are being researched and some of these will be brought to the market as data storage devices within the coming decade. Such a current research device is IBM’s “Millipede” [19] which is in fact a miniaturised punched card technology - harking back to the company’s roots when it started as the “Computing-Tabulating-Recording company” using Hollerith’s punched card technology to store and collate data. Millipede promises to push areal density to several hundred Gb/sq. inch, almost an order of magnitude higher than magnetic disks will reach.

RAID - SAFETY IN NUMBERS

In conclusion, through a long evolution and successive refinement we have disk drive technology that allows vast amounts of data to be stored and retrieved quickly at low cost. The fundamental technology - dangling a magnetic coil over a spinning disk coated with rust seems unlikely - indeed if you were to invent it today it might be hard to get financial backing to develop it. Of course every technology has a failure rate, and one thing to consider is “just how safe *is* your data on a disk drive”. The need to ensure that error rates are lower and lower, and the need to design systems that will not lose data even if one or more disk drives fail has led to the development of schemes to group together the capabilities of more than one disk drive to create “super-reliable” disk arrays[20]. It is these RAID arrays, not individual disk drives that are the basis for the kind of data storage that is demanded by current commercial applications.

References

[1] Principles of Neural Science, E R Kandel & J H Schwartz, 2^{nd} ed. Elsevier, 1985

[2] From Neuron to Brain, S W Kuffler, J G Nichols, A R Martin, 2^{nd} ed. Sinauer, 1984

[3] How the Mind Works, Steven Pinker, Penguin, 1998

[4] “How much do people remember? Some estimates of the quantity of learned information in long-term memory” T K Landauer, Cognitive Science 10, 477-493, 1986.

[5] http://www-1.ibm.com/ibm/history/documents/pdf/faq.pdf and http://www.research.ibm.com/research/gmr/history.htm

[6] Technological Impact of Magnetic Hard Disk Drives on Storage Systems, Ed Grochowski & R D Halem, IBM Systems Journal 42, No.2, 2003. (Several other figures also after Grochowski from IBM Storage Technology website 2001).

[7] Cramming more components onto integrated circuits. Gordon E Moore, Electronics, Volume 38, Number 8, April 19th 1965

[8] Mott N F, The electrical conductivity of transition metals, Proc. Roy. Soc., 153, 699 (1936) and Adv. Phys. 13, 325 (1964). [For Mott's "two current" model see also http://www.physicstoday.com/pt/vol-54/iss-5/p31.html]

[9] Wonders: The sum of human knowledge? Philip Morrison and Phylis Morrison, Scientific American, 279 (July 1998):115. Also here http://web.media.mit.edu/~saul/windup/wub/windup/0798wonders.htm

[10] IDC report, Computer Reseller News December 5, 2003

[11] Private communication. For more details on the LHC project, see the LHC link on the CERN home page http://public.web.cern.ch/public/

[12] Calculation of the magnetic field in the ferromagnetic layer of a magnetic drum, Olle Karlqvist, Transactions of the Royal Institute of Technology Stockholm, 86, 1954

[13] For a history of disk drive air bearing design, see http://www.hitachigst.com/hdd/research/storage/hdi/abshistory.html

[14] For a description of head load / unload technology see http://www.pcguide.com/ref/hdd/perf/qual/featuresHead-c.html

[15] Magnetic recording handbook: technology and applications, edited by C Denis Mee, Eric D Daniel, McGraw Hill, 1989, pp284 -

[16] http://www.research.ibm.com/research/gmr.html

[17] An introduction to magnetic recording heads, Mark H Kryder, published at http://www.seagate.com/docs/pdf/whitepaper/intromagneticheads.pdf

[18] http://www.research.ibm.com/resources/news/20010518_pixie_dust.shtml

[19] http://www.zurich.ibm.com/st/storage/millipede.html

[20] "A Case for Redundant Arrays of Inexpensive Disks (RAID)", David Patterson, Garth Gibson, Randy Katz, ACM, 1988