Continued scaling of existing information storage paradigms appears to be near its end. A slowdown in the increase in areal storage density in magnetic hard disk drives, a system at the forefront of storage density, has been apparent in the last eighteen months and is likely to be seen in optical and solid state systems as well. However, this slowdown is a consequence of the engineering paradigm itself rather than any manifestation of fundamental physical limits on storage system performance. This paper suggests that storage densities orders of magnitude greater than those demonstrated today are possible.

Key words: fundamental limits, nanoparticles, areal density, energy dissipation

1. Introduction

A recently published National Science Foundation report observed that, “an information based (or knowledge based) economy requires a ‘cyberinfrastructure,’ the foundation of which are the electronic, magnetic, and electrooptic components of computation, storage, and communication”[1]. Over the last several decades, information storage technologies have demonstrated tremendous improvements in capacity and data transfer rates. Magnetic hard disk drives have improved in areal density at an annual rate of 60% (and in some recent years 100% or more). Solid state electronic storage systems (non-volatile memory) have improved in accordance with Moore’s Law: a 60% annual increase in capacity, while optical storage systems have advanced with the development of shorter wavelength higher numerical aperture systems. However, the continued improvement of storage systems cannot be taken for granted. Each of these types of storage systems is facing limits that may impede their continued improvement. The “superparamagnetic” limit has been identified as a serious obstacle to the continued increase in areal density in magnetic hard disk drives[2]. Optical systems have been pushed to the very limits of diffraction limited far field systems and electrical systems will only continue to improve so long as electrical devices continue to be scaled; an increasingly challenging and expensive proposition as lithographic systems move to 90 nm and 45 nm minimum dimensions.

Despite these pressures, it must be understood that the limits faced by information storage technologies are in no way insurmountable nor are they fundamental. In terms of energy usage, spatial density, or speed, storage systems have in no way approached physical limits on their performance.

2. Fundamental Limits

It seems appropriate to conjecture what might be the truly “fundamental” limits on information storage. These limits rather than being based on a particular engineering paradigm should be based on limits imposed by the laws of physics as we understand them today.

2.1 Areal density (space)

The ultimate limit on storage density should be based on the minimum dimension at which space may be defined. This minimum dimension is given by the “Planck length” defined as:

$$\left( \frac{hG}{c^3} \right)^{1/2} = 1.6 \times 10^{-35} \text{ m}$$

Where $\hbar = \hbar / 2\pi$ and $\hbar$ is Planck’s constant, $G$ is the gravitational constant, and $c$ is the speed of light. This is the minimum length that “make sense” according to current theories of physics. Thus information could not be stored at an areal density that would require “bits” smaller than this. Thus the ultimate storage density using bits of Planck length dimensions is about $10^{66}$ bits/inch$^2$.

2.2 Data Rate (time)

As with space there is a minimum increment of time that “makes sense” according to current theories of physics. This is the “Planck time” and corresponds to the time it takes light to travel across a Planck distance and is thus given by:

$$\left( \frac{hG}{c^3} \right)^{1/2} = 5.4 \times 10^{-44} \text{ sec}$$

Once could therefore conjecture an ultimate data rate of $10^{43}$ bits/sec.

2.3 Energy dissipation

The minimum energy that must be dissipated to switch a bit is not related to the energy barrier between the two states that define a “1” or a “0”. This energy barrier must be about 1 eV or more to ensure stability at room
temperature against thermal fluctuations over a period of ten years or more. Rather the amount of energy that must be dissipated in defining the state of a two state system is given from thermodynamics to be:

$$k_B T \log 2 \approx 10^{-2} \text{eV}$$

Where $k_B$ is Boltzmann’s constant and $T$ is the temperature of the system (here assumed to be room temperature).

Of course it is not possible to reach all these limits simultaneously. For example, to dissipate the minimum energy in switching a two state system would require an infinite amount of time (far from our ultimate data rate). Nonetheless these true limits tell us something about how close we are to ultimate system performance.

3. Present Day Systems

The performance of current information storage systems may be measured with a variety of metrics. In terms of areal density magnetic hard disk drives are the most advanced with industrial laboratory demonstrations recently having been announced at 170 Gbits/inch$^2$ ($1.7 \times 10^{11}$ bits/inch$^2$). While clearly not approaching a “fundamental” limit these demonstrations are becoming increasingly difficult as the superparamagnetic limit is approached. In terms of data rate modern hard disk drives operate at data rates approaching 1 Gbit/sec (10$^{-9}$ sec) again far from a “fundamental” limit. Power dissipation is also not at a limit set by fundamental physical laws. A thin film head used in magnetic hard disk drives employs a current to generate the magnetic write field of about 50 mA with the head acting as a resistive impedance of about 6 $\Omega$. Thus in $10^{-9}$ sec the energy dissipated simply by resistive heating is:

$$E = I^2R \times t = 10^{8} \text{eV}$$

Which is about ten orders of magnitude from any thermodynamic limit.

This magnitude of energy dissipation is not required in a fundamental sense in order to achieve the data rates employed in hard disk drives. If one permits the dissipation in energy to equal the energy barrier between the two states of the system (1 eV) and that this equals the uncertainty in energy (thus allowing the system to switch between the two states) one finds that the uncertainty in time is:

$$\Delta t \geq \frac{\hbar}{\Delta E} \approx \frac{10^{-34}}{10^{-16}} = 10^{-15} \text{sec}$$

and thus a data rate requiring a clock whose uncertainty is much better than $10^{-9}$ sec can easily be achieved even if the uncertainty in energy is 1 eV.

Further evidence that present day systems do not operate in any sense near the limits of information storage can be surmised from nature. A virus such as the human cytomegalovirus has a nucleopcapsid that is approximately icosahedral and about 100 nm in diameter. The genome of this virus contains 200,000 base pairs. Thus with each base pair corresponding to 2 bits this corresponds to 400,000 bits per 100 nm or an areal density of about $3 \times 10^{14}$ bits/inch$^2$. We therefore have an example of an information storage “system” that is stable and rewritable and which employs an effective areal density about 10$^3$ times greater than today’s most advanced systems.

This is not to say that it is necessarily the case that viruses (and DNA) are the best model for how information storage systems should be designed. Rather in much the same way that birds have always been a proof of the viability of heavier than air flight so too are viruses and DNA a proof of the viability of information storage at densities far in excess of today’s systems. Indeed while it is important to learn from nature it is equally important not to learn too much from nature. In designing manmade heavier than air flying systems engineers separated the function of lift and thrust. Birds essentially produce both lift and thrust through their wings. The highest performance aircraft use primarily their wings to produce lift while having separate engines to produce thrust. The separation of these two functions has allowed engineers to optimize each independently and produce heavier than air flying systems that achieve performance in terms of altitude and speed far in excess of known naturally occurring flying systems. So too with DNA and viruses. These systems have evolved over time for more than simply information storage and it therefore seems likely that storage densities far in excess of “viral” densities are possible (consistent with the fundamental limits identified).

4. Beyond Today’s Limits

The International Storage Industry Consortium (INSIC) has set the current goal for an industrial demonstration of storage density at 1 Tbit/inch$^2$ by 2008. This may be accomplished through the continued improvement in longitudinal recording systems (though this appears unlikely), through the exploitation of perpendicular recording, or by employing heat assisted magnetic recording. At the same time one can already conceive of recording systems employing established materials and components or soon to be developed materials and devices that have the potential to take information storage system performance well beyond even this aggressive density goal.

In the context of magnetic recording Weller et al$^6$ have tabulated a number of materials that may be used as the medium and have noted that FePt in the L10 phase is stable even with a particle size of 3 nm. A square array of 3 nm particles corresponds to an areal density of 71 Tbits/inch$^2$ (sufficient for 50 Tbit/inch$^2$ of user bits) if each particle is used as a single bit cell. Work by Majetich$^7$ and others has shown that at least in principle...
it is possible to produce fairly uniform arrays of nanoparticles albeit not meeting all the needed characteristics for a storage medium at this time. Further work in terms of this type of medium will be required. New issues such as the nanoparticle to nanoparticle composition variation will have to be understood since this composition variation will cause variations in coercive fields and will therefore be a source of noise in any storage system\(^9\). One suggestion for write and read mechanisms for a storage system using such nanoparticles is writing by direct spin injection and reading using a tunneling magneto-resistance\(^9\). Spin polarized current can supply the torque to magnetically flip individual grains directly, as proposed in recent work\(^{10}\) and indeed demonstrated\(^{11}\), instead of using a magnetic field. This approach is only possible when bits become small enough that the required high local current density ($>10^7$ amp/cm\(^2\)) does not result in excessive heating. Figure 1 shows how such an approach might be used with a nonmagnetic tip in contact with a magnetic medium in a perpendicular recording configuration. The medium is stacked on top of a thicker reference layer, which effectively provides a source for the excess angular momentum that is deposited in the storage layer. This nanoscale device is very appealing from a systems point of view for two reasons. First, a non-metallic tip can be used, with the reference layer providing the magnetic reference. Thus controlling magnetic domains in the tip is not necessary, and simpler fabrication techniques can be used. Secondly, it enables a two-terminal device, with the sign of the current setting the sign of the bit, thereby avoiding the need to apply both current and, for example, a magnetic field.

Reading may be accomplished with a second similar tip though this latter probe would be a multilayer structure that includes a tunneling magneto-resistance sensor fabricated at the very end of such a probe tip. Work within the DSSC at CMU by J. Zhu and coworkers has indicated that the fabrication of such a structure may be possible and that indeed certain aspects of the fabrication process lend themselves to the creation of the thin layers at the probe tip "automatically".

Higher densities and smaller bit dimensions leads to weaker readback signals, more intersymbol interference (ISI) and intertrack interference. As data rates are increased, signal to noise ratio (SNR) degrades. In mechanically rotating storage systems, shorter access times are achieved with higher rotation speed, making timing recovery system performance critical. Thus, future data storage channels must cope with not only much lower SNR’s, much more ISI, but they must be able to carry out the necessary processing in much shorter time intervals. Thus the development of advanced equalization, detection, timing recovery and coding strategies to cope with these challenges is of great importance.

As the length and width of bits decrease into the nanoscale, mechanical addressing schemes will demand ever-greater servo systems performance. The servo systems will need to compensate for formerly negligible dynamic effects that dominate at the nanoscale, such as complex hysteresis behaviors. Compensation may take the form of new algorithms or even new hardware configurations. Greater integration of the data channel and the servo system will also be necessary and it may be necessary to record PES and servo state estimator information as a preamble to a data block.

As with any increase in storage density once can choose to design a system that either has greater capacity or a smaller form factor. At 50 Tbits/inch\(^2\) one can contemplate a storage system that has a capacity of 1 Tbyte and is only 1 cm\(^2\) in size. This is an attractive dimension as it would obviously lend itself to any of a number of embedded or mobile applications. This system could be rotating or "vibrating" if probe based. If one assumes that a latency of 1 msec is acceptable then in a probe based "vibrating system" one would have to design for a vibrational frequency of 1 KHz or a rotational 60,000 rpm neither of which appear to be unreasonable given some continued progress in the mechanical systems associated with these storage systems. The data rate in a rotating system could be determined by the speed at the outer edge of a 1 cm\(^2\) "disk". This turns out to be at the above rotation rate $35$ m/sec and corresponds to a data rate of 10 Gbits/sec. Neither of which appear to be beyond consideration assuming continued progress in channels and coding.

It appears that information storage systems of orders of magnitude better performance than today’s systems and yet not even approaching the information storage densities associated with biological systems such as viruses are reasonably attainable in the next ten years.

- **Fig. 1:** Schematic view of the probe-based recording scheme utilizing spin transfer effect. When electrons travel from the polarization layer to the tip, the spin transfer torque will switch the magnetization of the storage layer in the area underneath the tip to the magnetization direction of the polarization layer. If the current direction is reversed, an opposite magnetic bit is created.
5. The Need for Information Storage

One is left to ask then “Are increases in storage system performance needed by society?” The answer appears to be a firm yes. A recent report\textsuperscript{[12]} estimates that in 2002, $5 \times 10^{18}$ bytes of new information was produced in print, film, magnetic and optical storage media with 92% stored in hard disks. At the same time information flow through electronic channels (telephone, radio, TV, and the Internet) contained $18 \times 10^{18}$ bytes of new information or about 3.5x more than was recorded. Worldwide information production has increased by 30% each year between 1999 and 2002, according to this same report. Therefore without making subjective judgments as to the value of these different sources of information one must conclude that improved storage system performance is required. In addition, the development of systems such as that described above enable applications that would seem frivolous today but may very well be taken for granted in the future. Should every school child have their own copy of the entire print collection of the U.S. Library of Congress? Certainly yes if the cost is sufficiently low. Should every law enforcement official have a device containing a photo and fingerprint of every individual on the planet? Should my cell phone contain the entire telephone directory of any cell phone user on the planet? The capacity described above makes this possible but what are the positive and negative social implications of such capabilities placed into essentially anyone’s hands? These are issues that society may have to grapple with sooner than we may imagine.

References

5) P. Frank private communication.
9) J. Zhu private communication.

Received June 8, 2004; Accepted July 2, 2004