# Physics of Digital Data Storage

Tenzin Rigden[1]

*Carleton College, Department of Physics, Northfield, MN 55057*

(Dated: 13 February 2015)

Word count:

# I. INTRODUCTION

On a fundamental level, a computer stores everything in a series of 1s or 0s called bits. Movies, pictures, and applications are all stored as a series of billions of these bits. In this binary system, 1 represents a true/on state while a 0 represents a false/off state.

# II. OPTICAL STORAGE

## A. Background

Optical storage devices are devicecs that use optical methods to record and read information. For optical disks, data is recorded and read using a laser beam. To record data, a laser beam is shone on the reflective surface creating a pit with a depth of the wavelength of the laser divided by 2. Thus lands are the areas that were not shone on by the laser. To read, a very low power laser is shone on the disk and the reflected signal is converted to an electrical signal using a scanning photo detector. A "1" is interperted as when there is a change in the surface, i.e. changing from pit to land or land to pit, and a "0" is interperted as when there is no change in the surface. The way this is manifested is that when at a changing point for the surface, the laser will be shining at both the land and the pit. This means that the light reflected off of the pit will have a phase change of $\lambda/2$ behind the light reflected off of the land causing destructive interference and no reflected light, "1". When there is no change in the surface, the light simply reflects back normally and is interpreted as a "0." (CITE MEMORY MASS STORAGE)

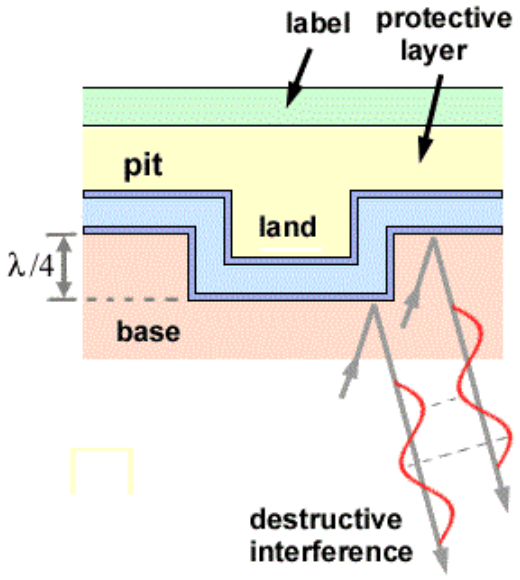


FIG. 1.

## B. Optical Disks

Optical Disks come in a few different styles: Compact Disk ROM (CD-ROM), CD-Recordable (CD-R), CD-Rewriteable (CD-RW), and Digital Versatile Disk (DVD). These disc technologies are written sequentially in a continous spiral track eminating from the middle out. This causes it to be slower than other technologies that use cocentric tracks. A CD-ROM is a disc that's only meant to be readable and typically used for things such as installation of programs. Since it is only written once in the factory, a glass master disk is created using a high intensity laser. Liquid polycarbonate is inserted into the master disk to recreate the disk which is then covered with a reflective layer then a protective one. A CD-R is slightly different because it needs to be able to written to once outside of the factory. This is one by having a painted layer between the reflective layer and transparent layer. The painted layer is initially transparent and becomes dark to simulate a pit when impacted by a high-intensity laser. Lastly, in a CD-RW, the reflective layer has two states, a slightly reflective amorphous state and a highly reflective crystalline state. A high intensity laser shone on this layer will cause it to turn into its amorphous state and simulates a pit, while a medium intensity laser will cause it to change back into its crystalline state and simulates a land. And again, a low intensity laser is used to read the pits and lands without

affecting the state.

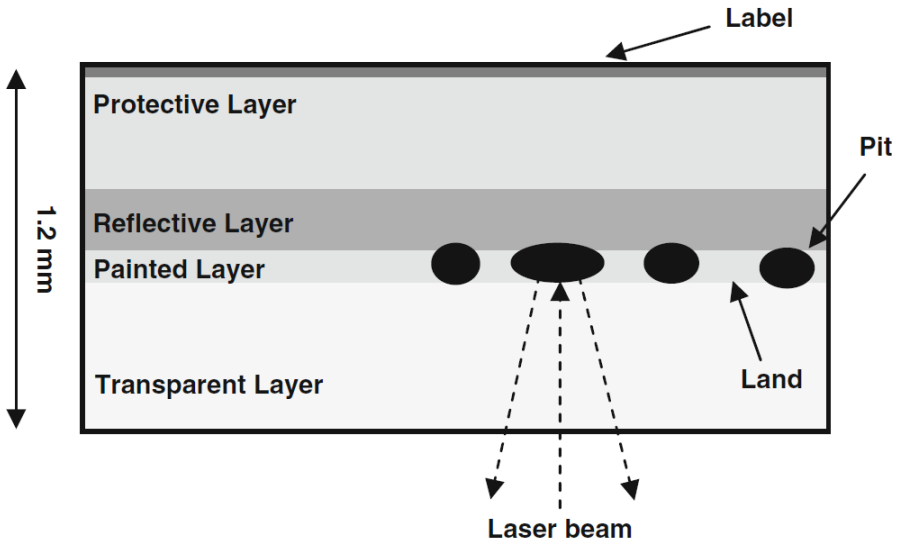


FIG. 2.

While CDs were useful and easy to carry around, their size was limited to only 700MB. Demand for higher storage capacities in the 1990s led to the invention of the DVD which offered data storage of at least 4.7GB. The first main difference between CDs and DVDs is that DVDs use a shorter wavelength light, 650nm, than CDs use, 780nm. What this allows is that pits and lands in the DVD can be much smaller and can be placed closer together than a CD could, increasing storage density. Another innovation that allowed storage of up to 8.5GB was the ability to have two different recording layer: a semi reflective layer in the middle and a fully reflective layer at the other end. The different layers can be read by changing the focus of the laser. Additionally, instead of having a protective layer on one side of the disk, you can put another disc on it to effectively double the capacity again. However, this requires the manual flipping of the disc when accessing the other half of the data. By having double sided and a double layer, DVDs can achieve a storage of 17GB. (CITE MEMORY MASS STORAGE INTRO SECTIONS)

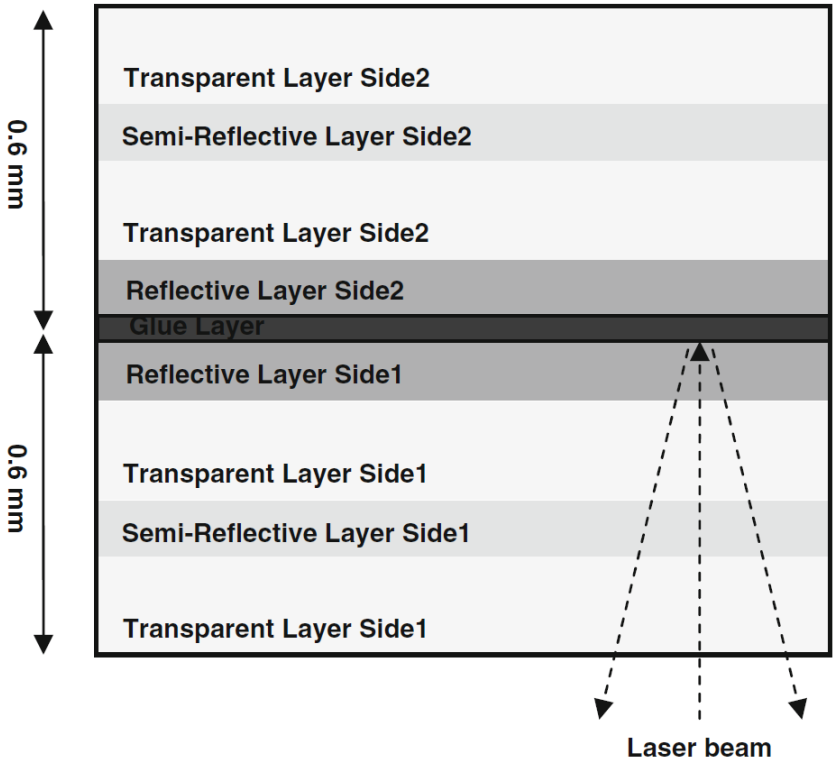


FIG. 3.

A more recent innovation in storage of optical disks was the invention of the Blu-ray disc. Just like how using a smaller wavelength light in DVD compared to the ones in CD allowed the reduction in size of the pits and track size, Blu-ray discs also use a much smaller wavelength light to increase capacity. Blu-ray uses 405nm light which allows both the pit sizes and the track size, space between tracks, to be reduced to .15$\mu$m and .32$\mu$m respectively compared to the track size of .74$\mu$m for DVDs.

### C. Holographic Data Storage

However, these all still pale in comparison to the potential storage capabilities of holographic data storage (HDS). In the most basic sense, a holography is the method of capturing and recreating optical information through interference patterns. To record a hologram, a coherent light source is split into a reference and object beam using a beam splitter. The object beam is reflected off the object being recorded on to the photographic plate. The reference beam, also being shone on to the photographic plate, experience interference with

the object beam and that interference pattern is recorded on the plate as a hologram. Destructive interference appears as dark spots and constructive interference appears as bright spots. This can all be seen in Figure 4.

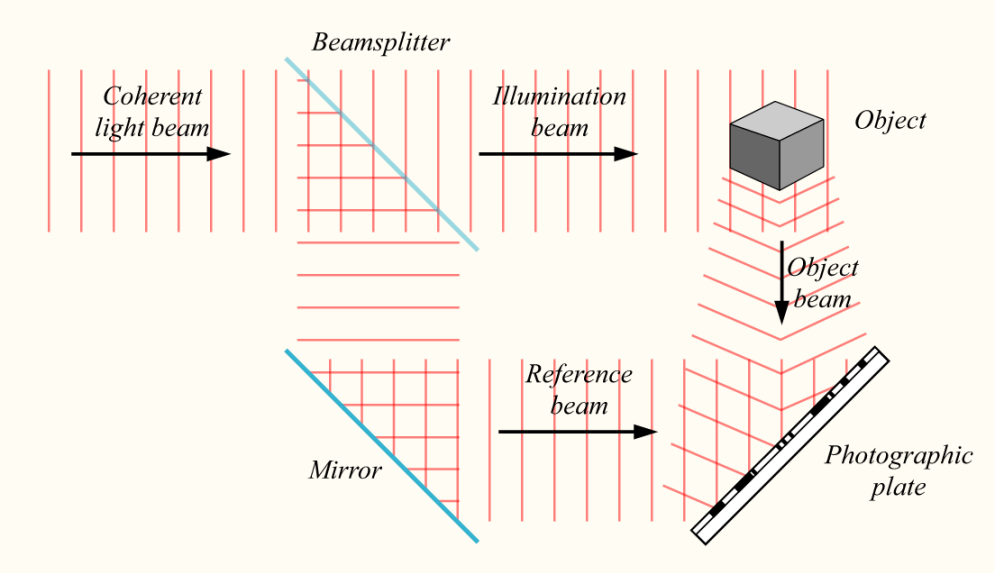


FIG. 4.

To recreate the image from the hologram, the reference beam is shone back at the hologram at the same angle and by looking at where the object should be through the hologram, a virtual image can be seen of the object as seen in Figure 5.

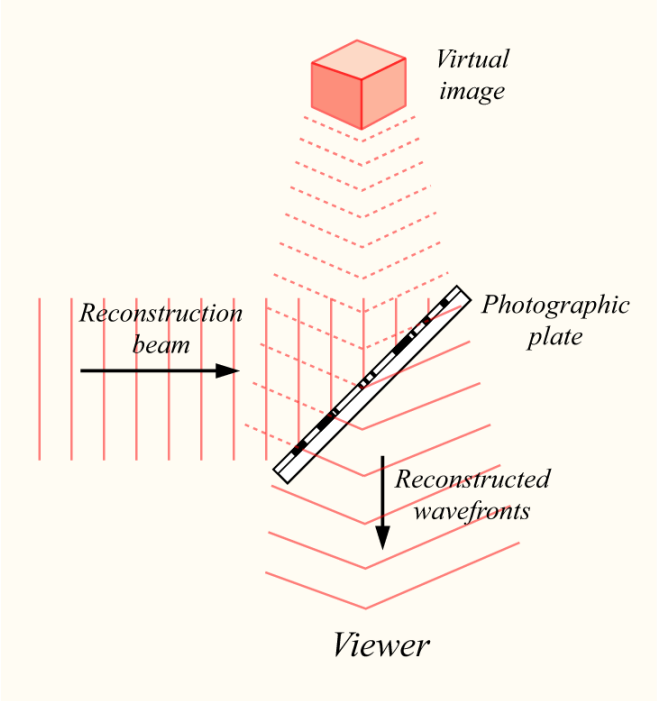


FIG. 5.

Holograms can be created in both thick and thin regimes. The thin regime is unsuitable for high density storage because thin holograms like those written on photographic film confine their diffractive interaction to a single plane and can't be used for multiplexing. (CITE HOLOGRAPHIC DATA STORAGE BOOK) The criterea used quantify the thickness or thinness of a hologram is known as the $Q$ parameter. The $Q$ parameter is given by

$$Q = \frac{2\pi\lambda L}{n_0 \Lambda^2}, \tag{1}$$

where $\lambda$ is the wavelength of the light, $L$ is the thcikness of the recording layer, $n_0$ is the index of refraction of the medium, and $\Lambda$ is the grating period. A hologram is considered in the thin regine if its $Q$ parameter is $Q < 1$ and is considered in the thick regime if its $Q$ parameter is $Q > 1$. The thickness of the medium can range from $500\mu$m to a few millimeters. (CITE HOLOGRAPHIC DATA STORAGE book)

To store digital data in holograms, a spatial light modulator (SLM) is needed. The SLM converts the 1s and 0s of the digital data into light and dark pixel respectively in a pattern. The number of bits that can be stored per SLM image can be over one million, but this varys with the SLM. So to record digital data, instead of reflecting the object beam off of the object, it is shone through the SLM then onto the storage medium as seen in figure 6.

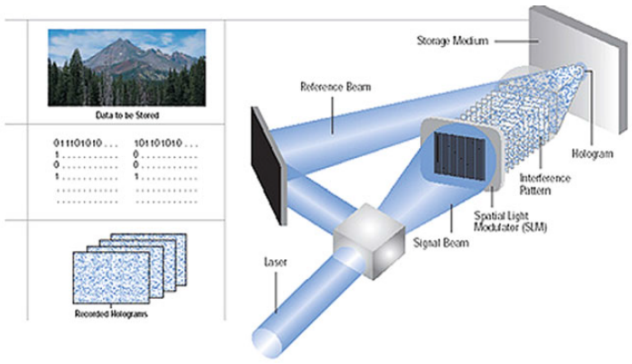


FIG. 6.

Reading the hologram is the same as it would be for a thin hologram. The reference beam is shone at the hologram, but now there is a detector array to convert the SLM image of dark and bright spots into 1s and 0s for the computer. Since a page of bits can be read/written all at once at the speed of light instead of bit by bit, HDS has the ability to have much greater read write speeds over other competing technologies such as magnetic data storage.

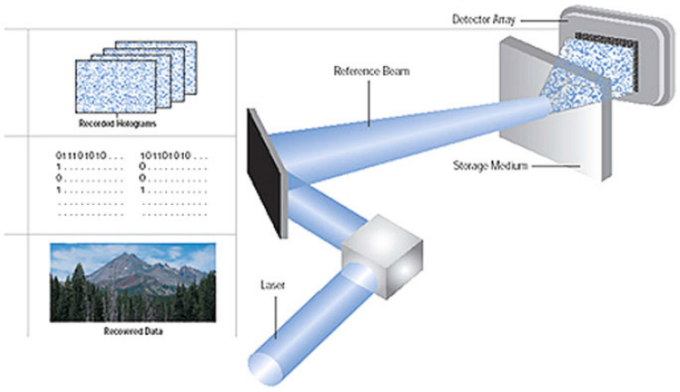


FIG. 7.

The reason why HDS can be so dense is because of the various multiplexing methods available. One type is angular multiplexing. Since the angle of the reference beam must match exactly the angle it had when recording the hologram, that means that a small change in the angle when recreating the image will cause the image not to appear. We can use this sensitivity by having one page of data from the SLM stored at one angle and by changing the angle of of entry. Another method to add multiplexing is to vary the wavelength of the the source and reference beams. This method is limited due to the small tuning range of lasers.

Yet another method is to vary the point of entry of the beams into the medium. While there are more methods of multiplexing, the true benefit of multiplexing occurs when combining multiple multiplexing techniques at once. For example, hybrid wavelength and angular multiplexing systems have been tested before. (CITE MEMORY MASS STORAGE)

## III. MAGNETIC STORAGE

### A. Background

Another popular technology used to store data is magnetic storage. Available in Magnetic storage is a property can be attributed to the ferromagnetic property of certain metals such as iron and cobalt. Ferromagnetic materials are divided into small magnetic domains each with their own magnetic field in a direction. An unmagnetized ferromagnet will have these domains in random directions such that the net effect is 0 as seen in Figure 8a. However, if an external magnetic field is applied, these domains will align with the external field and will stay aligned even after the external field is removed as in Figure 8b.

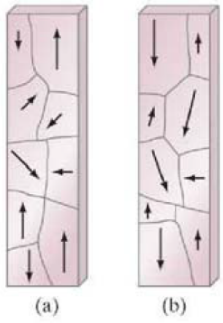


FIG. 8.

This is due to the hysteresis loop experienced by the ferromagnetic material. If we start with an unmagnetized ferromagnetic material with no external magnetic field, point a in Figure 9, and increase the external magnetic field, $B_0$, we get to point b where there is a both a nonzero external amgnetic field and total field from the external and material. If the external magnetic field is then decreased to zero, instead of returning to point a, it instead reaches point c. Here, even though there is no external magnetic field, there is still an internal field from the ferromagnetic material. This phenomenom is what allows magnetic

storage to store bits. Aligning a certain domain's magnetic field in one direction can signify a 1 while the other direction's signifies a 0.

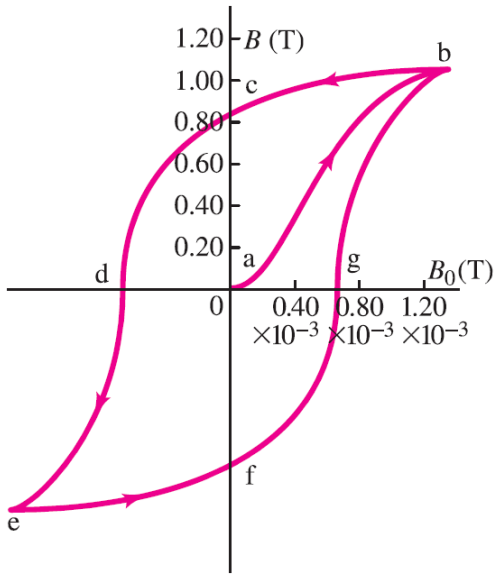


FIG. 9.

Reading and writing to the material is done by inducing a magnetic field using a head and coil of wire. The mechanics behind this are governed by Maxwell's equations. The two relevant equations are

$$\nabla x E = -\frac{\partial B}{\partial t}, \tag{2}$$

which is known as Faraday's law of induction and

$$\nabla x B = \mu_0 J + \mu_0 \epsilon_0 \frac{\partial E}{\partial t}, \tag{3}$$

which is known as Ampere's law (cite purcell). Faraday's law tells us that a changing electric field will yield and changing magnetic field, and Ampere's law tells us the opposite.

## B. Magnetic Disk Storage

A magnetic disk drive is the most commonly used data storage technology used in consumer products today. First introduced in 1956 by IBM, modern drives consist of mainly of a read/write head and a hard platter with a ferromagnetic layer. Originally, the head was used for both reading and writing but compromises had to be made. The read head wants a

thicker gap to be better able to penetrate the medium while the write head wants a smaller gap to get better accuracy. As domains got smaller, it became more and more difficult for the read/write heads to be able to properly read the intended domain from the increasing noise of the other domains. Nowadays, reading is done instead by a giant magnetoresistive (GMR) head. Magnetoresistance (MR) is the property that when a coiled wire will not only generate an electric field/current, it will also experience a change in resistance. This change in resistance can be used as a more accurate way to read the bits by applying a constant potential across a MR sensor and recording how the potential changes due to the change in the resistance of the wire because of Ohm's law. GMR is similar to MR but actually is a quantum effect. It is appropriately named because of its ability to change the resistance is giant in comparison to MR. (CITE HIGH DENSITY DATA STORAGE)

TALK MORE ABOUT GMR The mechanism behind this goes goes into the spin state of the moving electrons. These electrons can only have one of two possible states, up or down. If the spin states align with the direction of the magnetic field, they move forward in the wire simulating a lower resistance. However, if the spin state is against the field, then the electrons will travel in the opposite direction of the current causing more collisions and thereby increasing the resistance of the wire. (CITE HIGH DENSITY DATA STORAGE)

The innovation of GMR read head and a dedicated write head allowed the domain bits to shrink even more. However, a new issue began to arise. The ability to retain stored data in the material is dependent on the product of $K_u$ and $V$ where $K_u$ is the magnetic anisotropy constant of the material. This product gives the energy barrier that must be overcome that allows the material to keep its magnetic field even without the external field. However, as the domains decrease in size to increase density, the volume decreases and so does the energy barrier. If the volumte of the domain decreases to the point where this energy barrier is on the same order of magnitude as thermal energy, $K_bT$, the domain may randomly switch its field direction and "flip bits." This created a limit to the storage density that longitudinal recording could attain.(CITE PERPENDICULAR RECORDING ARTICLE)

To surpass this limit, the orientation of the recorded magnetic field changed from longitudinal recording to perpendicular recording. In addition to the change in orientation, the writing head was changed to a single-pole head and a highly permeable soft underlayer(SUL) was added below the recording material. The permeability of the SUL allowed it to act as a magnetic mirror to the actual head. This greatly increased the induced magnetic field used

to write data. The stronger magnetic field means that materials with higher anisotropy values could be used that were not feasible for the weaker magnetic field in longitudinal recording. Because the anisotropy constant can be increased, the volume of the domain can be decreased further without worrying about thermal fluctuations.

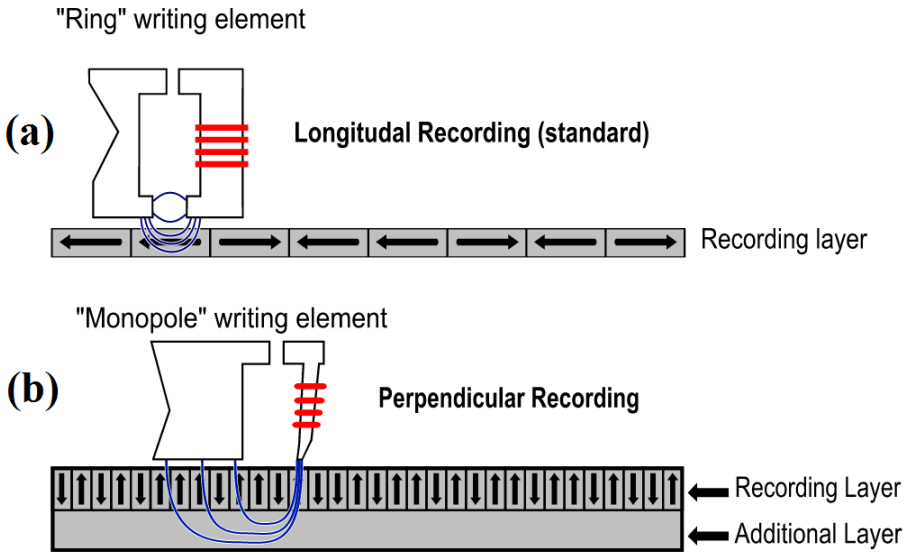


FIG. 10.

However, this too will have a limit at a smaller domain size where thermal fluctuations will be an issue again. One method still currently in development is called heat-assisted magnetic recording (HAMR). This method is based on the fact that all ferromagnetic materials have a curie temperature where the internal magnetic field it had is reduced to zero, and in the presence of an external magnetic field will align itself to that. In HAMR, when something needs to get written, a laser is used to heat up the desired domain to the curie temperature allowing our magnetic field to allign the domain, and then allowing the material to cool back down. This means that the recording material can be changed to one with a much higher anisotropy value allowing the domain sizes to decrease even more, further increasing the storage density. (CITE HAMR ARTICLE)

## IV. FLASH MEMORY

One of the most, if not the most, important device in modern computing as a whole is the transistor. A transistor is able take in a circuit input and act as a switch turning another circuits on or off without any moving parts. This ability to switch off and on a circuit will prove crucial in saving bits of memory.

### A. Semiconductors

To explain how transistors work, first we need to explain what a semiconductor is since a transistor is a semiconductor device. A semiconductor is any material that are more conductive than insulators but less conductive than conductors. A bandgap diagram is useful in describing semiconductors. Band structure represents the specific energy levels that electrons can occupy due to quantum mechanics. However, only the valence band is relevant when discussing conductivity. The valence band is the highest energy levels that electrons can occupy at absolute zero temperature. As seen in Figure 11, the conductor is any material whose valence band is not completely full because it overlaps with the next higher energy band. This means that the valence electrons are free to move around and can carry a current. An insulator is any material whose conduction band is separated by a large band gap from its valence band. This means that the valence electrons can not easily move to its conduction band and thus are not free too move. Lastly, a semiconductor is any material whose conduction band is only separated by a small gap meaning that valence electrons can easily be excited into the conduction band and carry a current.

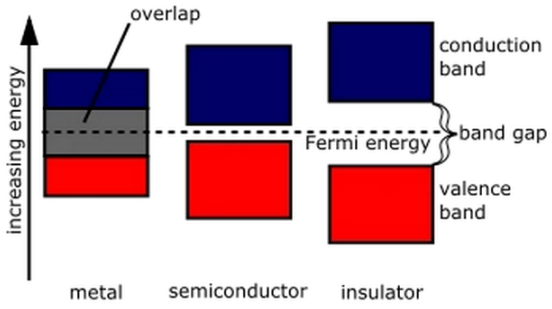


FIG. 11.

However, most commercial applications of semiconductors use impurity semiconductors.

Impurity semiconductors are materials that became conductive by adding an impurity, dopant, to the material. For example, say we have silicon with 4 valence electrons where each valence electron is shared with another silicon atom to create tetrahedron. Then if we replace one of those silicon atoms with a phosphorus atom with 5 valence electrons, 4 of the 5 valence electrons will be shared with the 4 nearby silicon atoms, but the last one is loosly bound to the phosphorus. At room temperature, thermal energy is enough to release it from this loosly bound state and it becomes a conduction electron. This is called n-type doping. If instead of adding an element with 5 valence electrons and instead added one with only 3, we would then have one less electron than its pure state which is represented as a hole, which act as positive charge carriers. At room temperature, this holes moves down to the the valence band where it is free to move.

A simple example of a semiconductor device is the pn junction diode. This diode has the property of only allowing current to flow in one direction. This device is created when a p type semiconductor is combined with an n type semiconductor. As the two are brought together, electrons in the n type fill the hole states in the p type. This continues until the energy levels of the p and n type semiconductors until it is no longer favorable for the electrons to keep moving. After reaching equilibrium, the n type ends up being positively charged and the p type negatively charged. Additionally, at the junction of the two semiconductors, there is a depletion region where there are almost no conduction electrons and or holes and there exists a strong local electric field due to charged bilayer created by the two semiconductors.

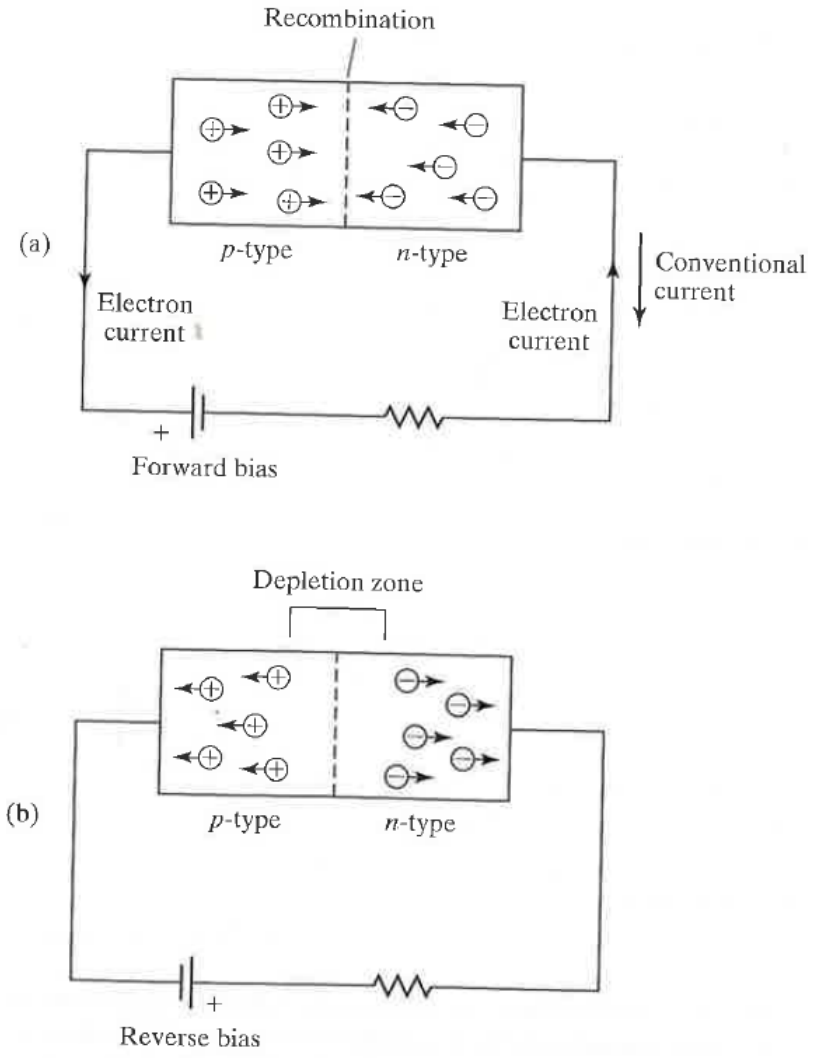


FIG. 12.

The reason that current can only flow one direction can be seen in Figure 12. When there is a clockwise current as in Figure 12a, some holes from the p type and electrons from the n type combine at the interface and and produce heat. So the diode loses electrons and holes at the interface, however the current pulls electrons from the p type resupplying the p type with holes. The converse happens in the n type so a current can flow indefinitely. This is known as forward bias. When the current flows counter-clockwise as in Figure 12b, the holes in the p type and the electrons in the n type flow away from the interface. This continues until the electric field due to the fixed impurity charges left behind is strong enough to balance the external field, causing the current to stop; this is known as reverse bias.

## B. Transistors

So now that we know what p type and n type doped semiconductors and pn junction diodes are, we can now talk about transistors. While there are many different types of semi conductors, the one that flash memory uses is the metal oxide semiconductor field effect transistor (MOSFET). A MOSFET, as seen in Figure 13a consists of a p type silicon with two n type regions on each side, source and drain, and a metal gate separated from the semiconductor by an insulating oxide layer (typically $SiO_2$) in the middle. When no voltage is applied to the gate as seen in a, the MOSFET acts as two pn diodes, and so any voltage difference between the source and drain will reverse bias one of the two pn diodes so no current will flow. However, if a positive voltage is applied to the gate, the positive charge attracts electrons and repels holes as in Figure 13b. If the voltage applied to the gate is above a certain threshold of around 1V, a thin layer of electrons will be formed under the oxide. Also known as the inversion layer, this creates a conducting channel between the source and drain, and the p type silicon acts as an n type allowing current to flow. This change created by the field from the gate is why it is called a field effect transistor. One advantage MOSFETS have over other transistors such as the Bipolar Junction Disorder is that the MOSFET requires nearly no current, so the input power is much less and there will be less issues with self-heating. (CITE MODERN PHYSICS TAYLOR)

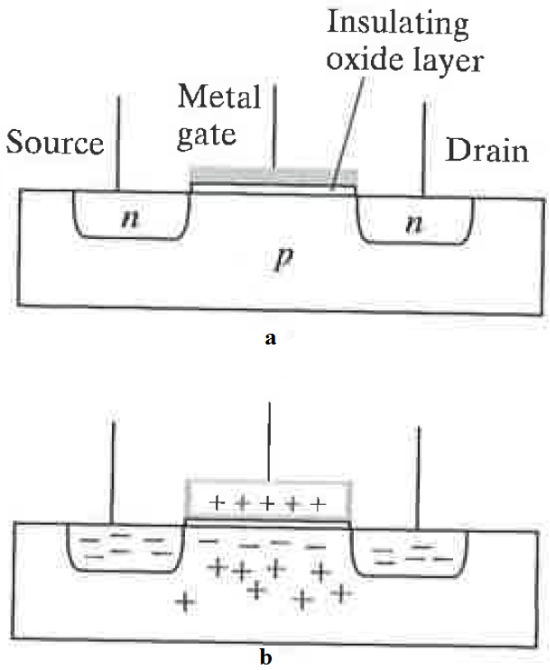


FIG. 13.

What is actually used to store bits is called the floating gate transistor. It is the same as the MOSFET except instead of only having one gate above the insulating oxide, you have another layer of insulating oxide on the other side of the gate followed by another gate called the control gate. So now the middle gate, also known as the floating gate, is electrically isolated due to the oxide layer on each side. This means that when positive charge is stored on this gate, current will pass through and send a 1 and when it's off, that data will be saved since the charges can't leave normally. Since the floating gate is electrically isolated, the only have to add and remove charge is through Fowler-Nordheim tunneling.

## V. PROBE BASED STORAGE

While still very much in development, probe based storage could offer an alternative to the more common storage systems such has hard drives and flash memory and compete with holographic storage. Based on the success of the scanning tunneling microscope (STM) and the atomic force microscope (AFM), there was much hope in using these new technologies to store massive amounts of data in much tinier scales than their traditional counterparts. While there is research being done on quite a few different approaches for probe based stor-

age, such as magnetic storage medium, phase change medium, and ferroelectric approach, I will talk briefly about thermomechanical method of storing and retrieving data in nanometer-scale indentations in thin polymer films. (CITE PROBE BASED) While the write and read speed of a single probe is relatively slow, it can be greatly improved by instead using a large array of probes in parallel that all can read, write, and erase its spot. Instead of traditional rotating disks in hard drives, the storage medium is positioned by microelectromechanical system (MEMS) based x and y actuators, which typically have actuation distances of around 100 $\mu$m. In thermomechanical probe-based storage, 1s and 0s correspond to indentations or lack of indentations respectively. Also, these indentations are created by softening up the medium through local heating by the tip. As for the arrays themselves, a large array consisting of up to 4096 (64x64) cantilevers with tips and sensor has been successfully created using silicon micromachining techniques. (CITE PROBE BASED) However, having this many tips brings up other issues such as physically controlling that many tips accurately and also the large power consumption needed for that many tips. Nevertheless, probe based storage, with its ability to achieve ultrahigh densities of over 1 Tb/in$^2$, has potential to become the next popular storage method.

## VI. APPLICATIONS

### A. Colloids

### B. Optical Thermal Ratchets

### C. Molecular Motors

## VII. BROADER IMPACTS

## REFERENCES