UNIVERSITY OF CALIFORNIA
RIVERSIDE

Heat Assisted Magnetic Recording for Areal Densities Beyond 1Tbit/in^2

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

by

Rabee Muhieddine Ikkawi

December 2008

Dissertation Committee:
Dr. Sakhrat Khizroev, Chairperson
Dr. Robert Haddon
Dr. Ilya Dumer

The Dissertation of Rabee Ikkawi is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ACKNOWLEDGEMENTS

First and foremost, I thank Allah^{swt} for His blessings.

The completion of this work would not have been possible without many great people. Predominantly, I am exceptionally grateful and indebted to my advisor Dr. Sakhrat Khizroev for being my graduate studies advisor. He has been instrumentally compassionate with his technical and non-technical advice to broaden my engineering and non-engineering horizons. He stands behind where I am today. I am delighted to say that he has been and will always be my mentor.

I am pleased to have world-class, nationally recognized committee members starting with Dr. Robert Haddon, Dr. Ilya Dumer, Dr. Alexander Balandin, and Dr. Cengiz Ozkan.

I want to thank my friends and research colleagues for their continuous support and collaboration which was the building block of this thesis. I start with a great friend before anything: Nissim Amos. Our friendship since the undergraduate years has resulted in professional research solidarity. We have concurrently been through all the milestones to present this thesis since 2003. Yazan Hijazi, indeed a great friend and colleague, has been exceedingly influential. I thank Jihad Boura for his support and sponsorship. He played a major role in molding my career path.

I am also grateful for the experience I gained at Western Digital (San Jose 1) over the past two years and would like to thank Pradeep Thayamballi and Jih-Ping Peng for showing confidence in me and giving me the opportunity out of so many available PhD students. From Western Digital also, I would like to acknowledge John Al- Sammerraie, Adam Torabi, Jinshan Li, Rahmat Karimi, Jim Tung, Joel Shummaker, and Jose Mendez for their relentless support and professional advice.

DEDICATION

I dedicate this thesis to my lovely wife, Mayssaa and our little gorgeous Yasmina. Having both next to me is unparalleled with anything else in life. Their continuous support, patience, and love are the reasons behind what I am today.

My parents are the ones who believed in me before I did myself. Their sacrifice and love empowered me to work hard and satisfy their dream of seeing me awarded a doctoral degree in engineering. My brothers and sisters: Tarek, Souhair, Mohammad, and Neveen have always been loving and supportive. I will always strive to make them proud of me like I am proud of them. My in-laws, the whole family, have been very considerate and accommodating throughout. I would like to thank every individual of them.

ABSTRACT OF THE DISSERTATION

Heat Assisted Magnetic Recording for Areal Densities Beyond 1Tbit/in^2

by

Rabee Muhieddine Ikkawi

Doctor of Philosophy, Graduate program in Electrical Engineering
University of California, Riverside, December 2008
Dr. Sakhrat Khizroev, Chairperson

The magnetic recording industry is keeping up with the ultra high demand of high capacity hard drives by improving the areal recording densities of these devices. Such imposing advancement in utilization and performance is due to successive scaling in the geometrical dimensions of the device. This progression is truncated by the fundamental limit known as the superparamagnetic limit which occurs when bits of digital data are aggressively decreased that ambient heat demagnetizes them, leading to loss of the stored data.

To overcome this problem, the use of large magnetic anisotropy energy density alloys is compulsory, but the write fields that are required by these alloys are prohibitively large, rendering these media effectively unwritable. Fortunately, heat assisted magnetic recording (HAMR) enables the use of the smallest possible thermally stable grains, irrespective of the ultra-large intrinsic anisotropy. HAMR exploits the substantial drop of coercivity of ferromagnetic material to a level attainable by the magnetic writing head when the disk temperature is elevated close to its Curie level, consequently enhancing the areal density dramatically.

In this thesis, a theoretical and experimental study underlying the design of a heating element based on ultra-high-efficiency near-field optics suitable for extending areal densities beyond 1Tbit/in^2 is presented. Near-field apertures are fabricated using focused ion beam milling and characterized via near-field optical microscopy. A breakthrough in the localization of adequate amount of heat into a 30 nm spot size is reported. HAMR demonstration is presented utilizing a customized spinstand tester, and the associated issues are thoroughly addressed.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER I: Data Storage Perspectives

1. Technological Evolution of Hard Disk Drive

"How much information is created in a year?" A question raised and answered by a study [1] conducted by the school of Information Management and Systems at the university of California, Berkeley. According to the study, about 5 exabytes (1exabyte = 10^{18} bytes) of new information was created in 2002 and 92% of this information was stored in magnetic media. Currently, hard drives are the dominant magnetic storage devices because they offer the best overall combination in non-volatility, reliability, large capacity, high data transfer rate, and low production cost. Data from disk drive manufacturers showed that the estimated number of total hard drives sold in 2006 exceeded 400 million units and expected to exceed 650 million in 2010, generating worldwide market revenue of the order of tens of billions dollars per year.

Hard disk data storage technology is evolving at a remarkable pace that has few equals. The phenomenal advancement in this technology has been achieved as a result of continuous and coordinated progress in such diverse areas as physics and processing of novel nano-film structures, mechanics of positioning and flying with extremely small tolerances as well as the development of novel techniques for coding, detection and digital signal processing of high-density recorded data. The design of modern disk drives is a fusion of various scientific disciplines entailing magnetics, surface physics, material science, tribology, control theory, information and communication theories.

Since the invention of the first hard drive by IBM in 1956, the technological evolution of hard drives has been nothing short of phenomenal. The first commercial hard drive, called RAMAC (Random Access Method of Accounting and Control), had a total

capacity of 5MB, contained 50 disks each 24 inches in diameter, had a access rate of one second and data transfer rate of 0.5MB per second, and cost $50,000 [2]. By contrast, today's typical hard drives can hold capacities in excess of 500GB, use 2.5-3.5 inch disks, have an access time of 6ms, and data transfer of 150MB per second, and cost about $120. The cost per megabyte has decreased 25 million fold in half a century.

Research in magnetic recording has being pursued in the directions of increasing the areal density of storage, increasing the rate of data transfer, decreasing the data access time, and decreasing the cost per megabyte. These are the four figures of merit for hard disk drives. Since the inception of the first hard drive half a century ago, areal density has explosively increased by strictly scaling the device physical dimensions while refining both material and manufacturing processes. However, this progression has been hindered by a fundamental limitation known as the superparamagnetic limit.

2. Superparamagnetic limit and the need for a new technology

The data on a magnetic recording medium is stored by means of recording a certain spatial variations of the magnetization, where the magnetization variations represent the data. The relation between the data and the magnetization pattern is defined by the encoding scheme used. Figure 3 shows a simplified schematic of a conventional longitudinal recording system. The recording media are engineered such that the preferred direction of the magnetization, a so-called easy axis, lies in the plane of the recording layer. Using an inductive "ring"-type writer, the magnetization of the grains is aligned along the track in either positive or negative direction. The data is read back using a magnetoresistive element. A change or no change in the magnetization direction at the bit transitions corresponds to a 1 or to a 0, respectively. The lateral dimension of a

bit, i.e. the smallest feature realized in a particular drive design, defines the areal bit density that such a drive supports.

A conventional magnetic medium has granular structure such that each bit consists of several magnetic grains or magnetic clusters. The magnetic clusters/grains are usually shaped irregularly and are randomly packed, as shown in Figure 1a. Consequently, the recording bits and bit transitions are usually not perfect, which is illustrated in Figure 1b. These imperfections lead to noise in the playback signal. The noise is kept below a certain acceptable level by means of including a sufficiently large number of magnetic grains into each bit. The resulting averaging reduces the level of noise. As the areal density increases, the bit size and the size of the grains that constitutes the bit, decreases. Typical grains in today's media range from 5 to 15nm.

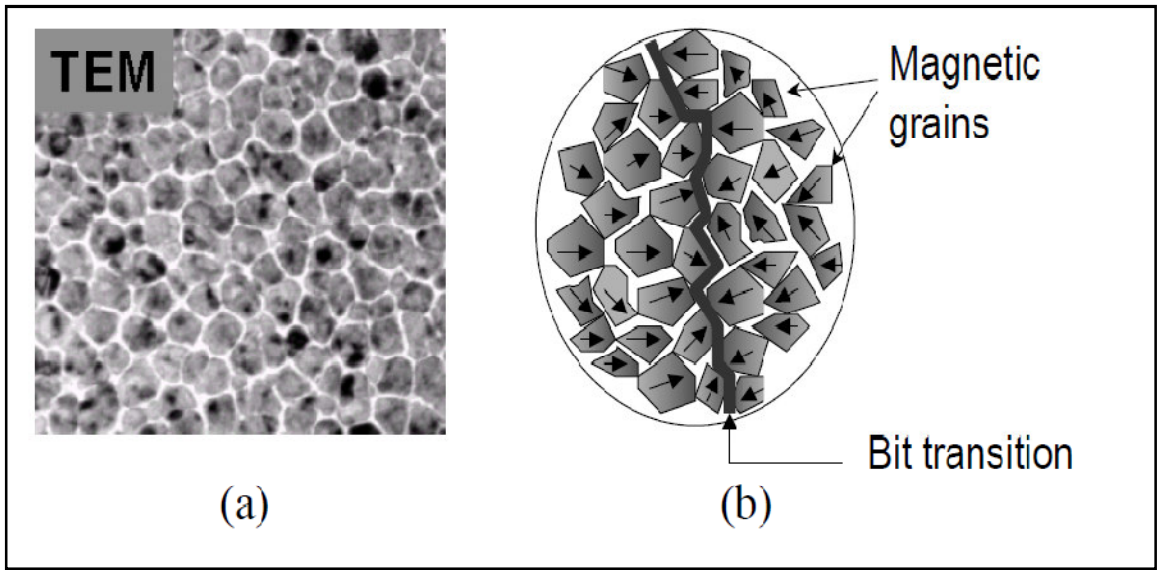

Figure 1: (a) A transmission electron micrograph of a typical granular medium; (b) a schematic of a single bit transition in a granular medium.

One of the critical factors characterizing the reliability of a data storage device is data stability. Various parameters control the stability of the data against the external factors.

With respect to the external temperature, which is manifested by thermal fluctuations in the recording media, the magnetic anisotropy energy stored in each magnetic grain is one of the major determinants (assuming that the grains are magnetically independent). The magnetic anisotropy energy approximately defines the amount of energy necessary to reverse the direction of the magnetization of a grain. For a single grain, it is equal to K_UV, where K_U is the magnetic anisotropy energy per unit volume and V is the volume of the grain. For a medium to be thermally stable, the above quantity K_UV should be substantially greater (40-60 times) than the energy of a single quantum of thermal fluctuation in order to store information for 20 years or longer, k_BT, where k_B is Boltzman's constant and T is the temperature [5] as shown in Figure 2.

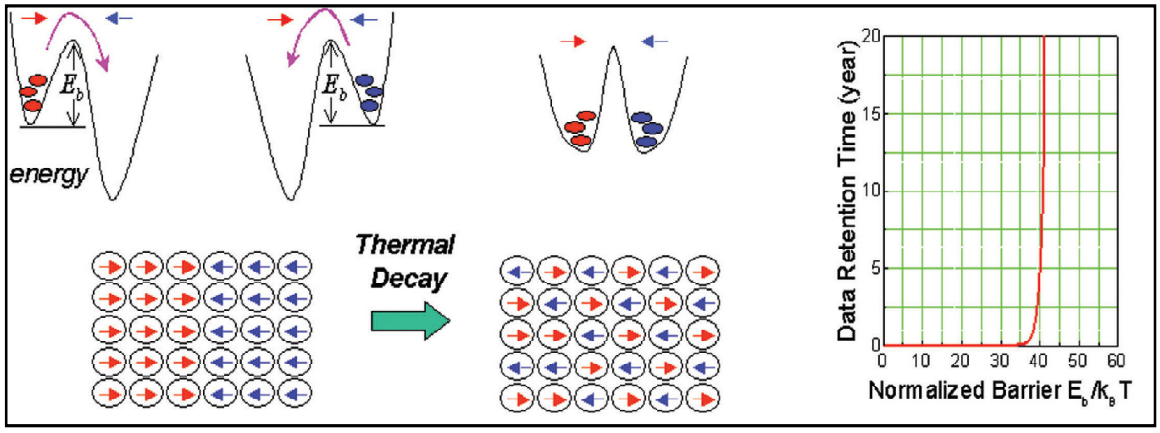

Figure 2: Spontaneous erasure of a recorded transition under thermal excitation. A transition is in a thermal nonequilibrium state (left) and the asymmetry between the two energy wells is caused by the demagnetization field generated in the transition. Thermal excitation will yield spontaneous magnetization reversal of the grains: when thermal equilibrium is reached, the transition is completely erased. In other words, data retention time in a recording medium is always finite and it is just a matter of how long before data are thermally erased [3,4].

As mentioned above, the higher areal densities require smaller grain sizes. It follows that to sustain thermal stability; K_U of a magnetic medium material should increase with the grain size decreases. Unfortunately, as K_U increases, so does the write field necessary to

efficiently write onto the medium. In conventional longitudinal recording, the upper limit of the write field that a recording head can generate is equal to $2\pi MS$ where MS is the saturation magnetization moment of the head material. The highest value of $4\pi MS$ of the materials available today is rapidly approaching what is believed to be a fundamental limit of ~25kGauss. This defines the upper limit of the K_U values that can be employed in a longitudinal medium and, consequently, the maximum areal density achievable with conventional longitudinal recording. It has been predicted that with the materials available today, the highest areal density achievable with conventional longitudinal recording is ~100Gbit/in^2 [5]. To achieve the optimum transition sharpness, the field magnitude needs to be optimized such that its gradient at the center of the transition is a maximum. This means that the maximum write head field in the medium has to be approximately twice the coercivity. The field generated by a write head is determined by the magnetic flux density inside the head poles at the gap. In current HDDs, the maximum head field is slightly less than half the magnetic flux density in the head poles. Over the years, the head pole material in the gap region has changed from permalloy ($Ni_{81}Fe_{19}$) with saturation flux density Bs = 1.0 T, to $Ni_{45}Fe_{55}$ with B_s = 1.5 T, to $Fe_{65}Co35$ alloy with B_s = 2.45 T, which is used in write heads manufactured today5-7. At the same time, the coercivity of thin film recording media has increased from 2000 Oe to 4800 Oe. However, the alloy $Fe_{65}Co_{35}$ used in today's write heads has the highest saturation magnetic flux density (Bs = 2.45 T) of any existing magnetic material. This means that the maximum write field inside recording media in longitudinal mode is limited to 1.2 T. As an estimate, data writing in a medium where coercivity exceeds 6000 Oe would not be optimal. As shown in Figure 3, the field in the recording medium is the

fringing field generated across the head gap by the magnetized write head poles on both sides. Inside the gap, the magnetic flux density is essentially the same as that inside the magnetic poles on both sides of the gap, almost twice the fringing flux density. However, if the gap material is superconducting, the magnetic flux is repelled out of the gap and the field intensity in the recording media would be more than doubled. To achieve sufficient writing efficiency, this is defined as:

$$\varepsilon = \frac{field\ intensity\ in\ gap \times gap\ length}{number\ of\ \ head\ turns \times current\ amplitude}$$

The head material needs to be reasonably soft, i.e. its permeability must be sufficiently high and its coercivity sufficiently small so that the field intensity within the head pole and yoke becomes negligible. It has been found that depositing a crystalline, body-centered cubic (bcc) $Fe_{65}Co_{35}$ film on top of a face-centered cubic (fcc) underlayer, such as Cu, yields very small grains with low coercivity.

3. Deferment of the superparamagnetic limit: Alternative Technologies

After many attempts to extend longitudinal magnetic recording as the main technology for at least another couple of years, researchers are coming to terms with the reality. It is commonly recognized that the conventional recording media become highly unstable as the density increases beyond ~ 100 Gbit/in^2 (a bit with a 160 x 40 nm^2 cross-section). At nanoscale, the energy required to switch magnetization becomes comparable to the thermal energy (kT), thus making the recorded information unstable at room temperature [5].This is especially disappointing considering that during the last half-a-decade, the industry had been enjoying a relatively steady areal density growth rate of ~ 100 % per year [6]. Among key alternative technologies are 1) perpendicular magnetic recording [7,

8], 2) heat-assisted magnetic recording [9, 10, 11], 3) patterned media [12, 13, 14, 15, 16, 17], and the less traditional, 4) protein-based memory [18, 19, 20]. Each of the alternatives has its advantages and limitations. A number of open questions to be answered are associated with the non-conventional physics underlying the recording mechanisms at nanoscale. The main objective of this thesis is to implement a harmonious experimental methodology to comparatively study the physics underlying the properties of one of the key alternative data recording devices and systems and thus understand their limits and potential. The two alternatives promoted in this thesis are Heat Assisted Magnetic Recording (HAMR) and protein based memory.

A. Alternative One: Perpendicular Magnetic Recording

Perpendicular recording has been the technology adopted to succeed the conventional longitudinal method of recording information. It is believed that heat-assisted magnetic recording (HAMR) [21] is the most promising alternatives for extending the areal data density to 1Terabit/in^2 and beyond (a square bit cell with a side of less than 26 nm).

A schematic diagram of a typical longitudinal system is shown in Figure 3. In this recording mode, they use a ring recording head for writing and a single-layer longitudinal magnetic recording media (with magnetization in the plane of the disk). The recording is produced by the recording field fringing from the gap region. The maximum value of the recording field which can be achieved is about $2\pi M_s$, where M_s is the saturation magnetization of the magnetic material of which the recording head is made [22]. The

magnetic read head (typically, a GMR sensor) is attached along the track next to the write head.

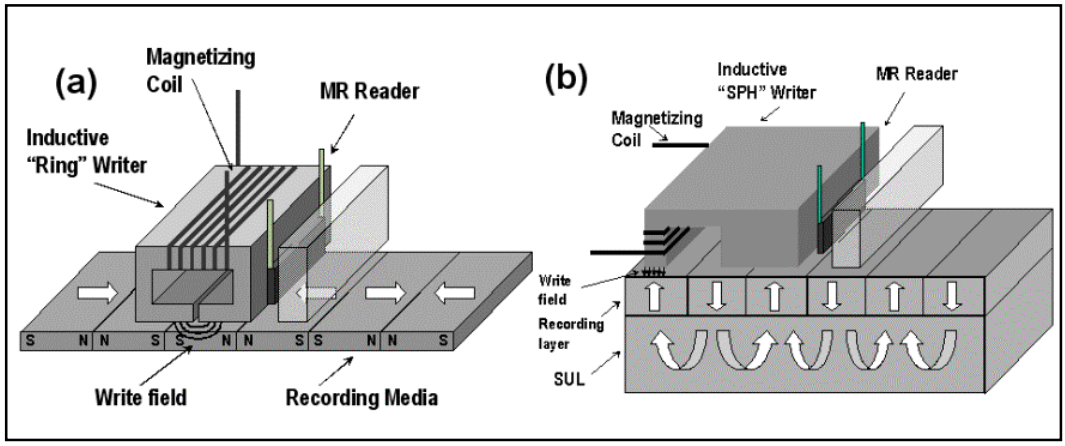

Figure 3: (a) conventional LMR, (b) PMR

A schematic diagram of a recording system of probably the most popular mode out of several perpendicular recording modes is shown in Figure 3b. In this perpendicular mode, they use a single pole head (SPH) for writing and a double-layer recording media consisting of a "hard" perpendicular magnetic layer and a "soft" magnetic underlayer (SUL). The purpose of the soft underlayer is to "close" the path for the magnetic flux generated in the coil of the recording head and to increase the magnetic recording field by approximately a factor of two (i.e. to almost $4\pi M_s$, where M_s is the magnetization of the material the pole is made of). The increased recording field means that recording media with higher anisotropy can be utilized thus deferring the superparamagnetic limit to even higher areal densities. Similar to longitudinal recording, the magnetic read head (e.g. a GMR or another magnetoresistive sensor) is placed along the main track next to the single pole write head.

B. Alternative Two: Heat-Assisted Magnetic Recording

HAMR has potential to extend areal densities to at least 50 Terabit/in^2, with perpendicular recording deferring superconducting limit barely above the 1 Terabit/in^2 mark and patterned media on its own making above 10 Terabit/in^2. As the title implies, to record information, HAMR relies on using not only magnetic but also thermal energy. Therefore, a recording medium with substantially larger anisotropy energy could be utilized for storing information. In turn, larger anisotropy energy leads to a larger thermal stability.

A schematic diagram of one of the popular modes of HAMR is shown in Figure 4. As it can be seen, a laser beam heats the region in the recording media where the final information is going to be recorded before the field is applied. Due to the local temperature increase in this region, the field necessary to switch the magnetization is going to be sufficiently small despite the ultra-large anisotropy of the recording media. In other words, both thermal and magnetic energies are combined together to switch the magnetization. In this far-field light delivery implementation, the bit track and length are determined by the applied magnetic field.

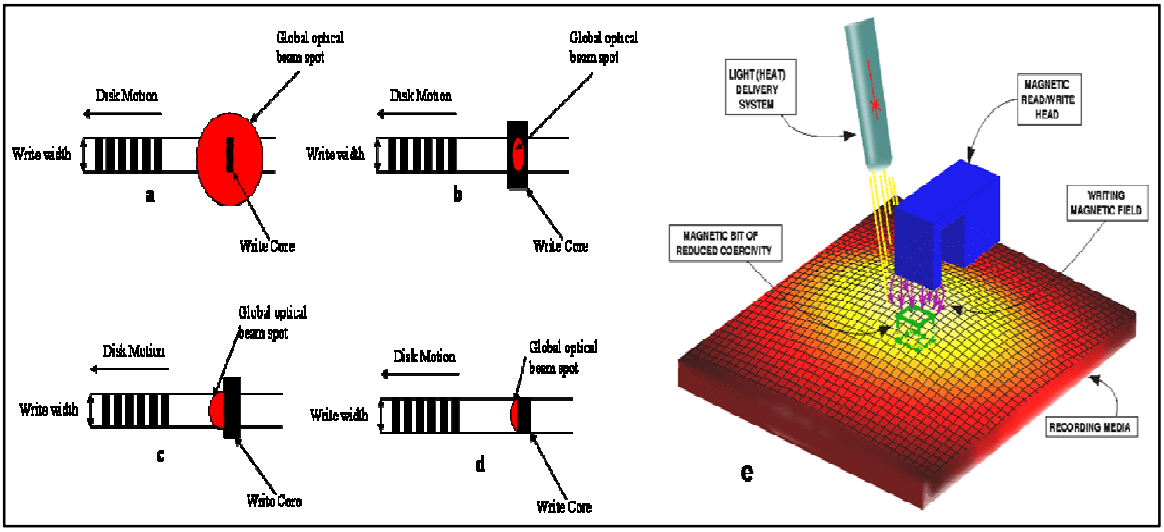

Figure 4: Four different modes of HAMR: (a) far-field light and localized magnetic field, (b) near-field light and large-scale field, (c) near-field light and semi-localized field to define the bit length, and (d) collocated near-field light and magnetic field.

A schematic diagram of one of the popular modes of HAMR is shown in Figure 4. A laser beam heats the media region where the final information is going to be recorded before the field is applied. Due to the local temperature increase (closer to Curie), the field necessary to switch the magnetization is going to be accessibly small despite the ultra-large anisotropy. In other words, both thermal and magnetic energy parts will contribute together to switch the magnetization. In this far-field light delivery implementation, the bit track and length are determined by the applied magnetic field. For simplicity, in this figure a single pole perpendicular magnetic head is shown as the source of the recording field. A variety of combinations of magnetic recording systems and optical heat sources will be thoroughly discussed in chapter II. Finally, HAMR could be integrated with patterned media.

There are at least four modes of HAMR to be studied, as shown in Figure 4a-d, respectively. Figure 4a illustrates the implementation with far-field light and a localized magnetic field, while Figure 4-b illustrates the implementation with near-field light and a

large-scale magnetic field. In the first case, the magnetic field defines the bit width and length, while in the latter case the light defines the bit width and length. The implementation with near-field light and a semi-localized magnetic field is shown in Figure 4-c. In this case, the light defines the bit trackwidth while the magnetic field defines the bit length. Finally, the implementation with collocated near-field light and magnetic field is shown in Figure 4-d.

C. Alternative Three: Patterned Media

Patterned media is believed to be the ultimate magnetic data storage solution for areal densities even beyond 10 Terabit/in^2 (a square bit cell with a side smaller than 8 nm). At such high areal densities, the entire library of Congress can be stored in a device of a size of a dime coin. To illustrate the concept of patterned media, one should be reminded that in the conventional recording mode, every bit consists of hundreds of grains for maintaining the statistically determined signal to noise ratio (SNR) or by the random $1/N^{0.5}$ noise. To keep up sufficient (for the data channel) SNR with the areal density increase, a grain size is linearly reduced as the bit size is reduced. This, in turn, results in the superparamagnetic limit of the conventional recording: the anisotropy energy of one grain (proportional to its volume) eventually becomes comparable to the thermal energy (kT) and consequently media becomes unstable. In the case of patterned media, every bit consists of only one grain. The issue with the statistically defined SNR is solved by periodically organizing (assembling) the bit structure. Due to the periodicity, the bits can be synchronized with the signal channel. In the simplest scheme, the magnetic elements could have only one axis of magnetization. The direction of magnetization is interpreted as a binary 1 or 0. Comparative diagrams of conventional (disordered) longitudinal media

and patterned perpendicular media with a soft underlayer are shown in Figure 5a-b, respectively. The other implementations of patterned media include single-layer perpendicular and longitudinal magnetic patterns.

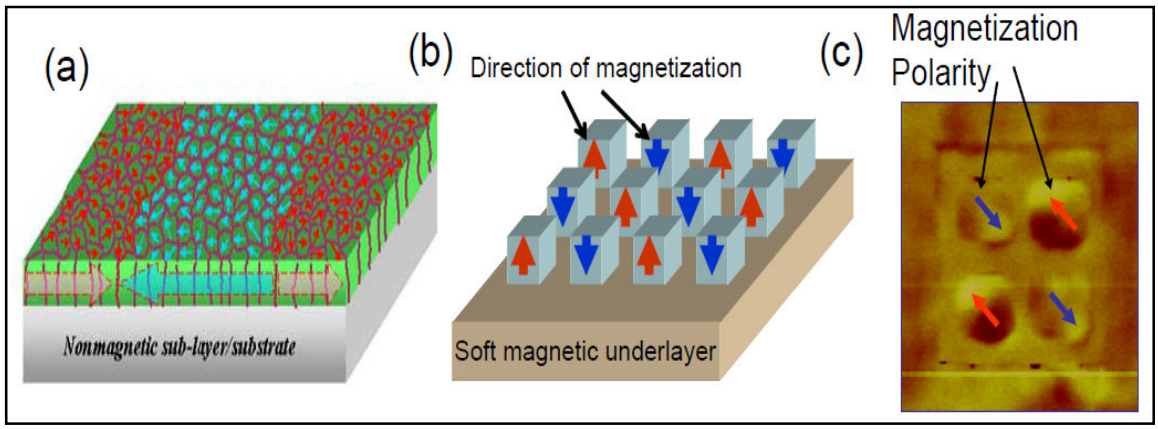

Figure 5: (a) Conventional "random" longitudinal Media and (b) Patterned perpendicular media. (c) MFM image of FIB-made patterned longitudinal media.

The problem with patterned media is that, until now, there have been no economical and mass-production-compatible manufacturing methods capable of making them. For example, electron beam is resolution limited and too slow to cover large areas [23], and laser-interferometric lithograpy [24] is laser wavelength limited. In this project, a comparative study of patterned media manufactured with different methods will be made. Among other fabrication methods, FIB, self-assembled particles [25], nanoimprinting lithography [26] are currently being investigated.

D. Alternative Four: protein based memory

Protein memory in the proposed form is the direct result of cross-fertilization of magnetic recording and microbiology and will be studied as a technology potentially capable of recording at a single-molecule level.

In the past, protein-based memory has been proposed by Robert Birge. His approach has relied on the attempt to utilize protein as a three-dimensional (3-D) means of memory.

This task has been complicated by the presence of two unknowns. First, it is an open question how to immobilize protein into a 3-D form. The recording media used in his experiments has been in a gel form. As a result, it has not been trivial not only to store information in the media but also to store the media itself. Secondly, researchers have tried to implement optical devices to record and read information from the 3-D media. However, even today, holographic or other 3-D optical methods have not succeeded in the demonstration of effective areal densities comparable to the values in routine demonstrations of magnetic recording.

There is no fundamental difficulty in using the same optical recording transducer configuration as used in HAMR. As described above, in this recording mode, a FIB-fabricated near-field optical transducer will be used to focus photons in a spot size substantially smaller than the physical wavelength to drive proteins through their thermal cycle. This alternative will be thoroughly discussed in chapter IV.

4. Objective and arrangement of the thesis

The main goal of this thesis is to contribute to the development of near-field optical transducers that can be effectively utilized for heat assisted magnetic recording and protein based memory. A study underlying the design of a heating element based on ultra-high-efficiency near-field optics suitable for extending areal densities beyond 1Tbit/in^2 is presented. A focus is given to the experimental and theoretical study of the issues associated with the localization of adequate amount of heat into a spot size smaller than 30 nm on the recording media. Theoretically, the thesis thoroughly presents the design issues of near-field transducers and the heat induced by them using the finite element method numerical analysis. This thesis also presents an experimental study of

recording characteristics of various near-field transducers fabricated via focused ion beam milling (FIB). To count the number of photons emitted in the near field, a scanning near-field optical microscopy system has been implemented. A far-field HAMR spinstand has been designed and built to analyze the recording physics in HAMR, a demonstration is also presented. Since the near-field transducers are not fundamentally limited to HAMR, they can be effectively utilized for protein based memory. Thus, a detailed description on this alternative technology is presented.

Chapter one introduces the fundamental limitations hindering the advancement of the current data storage technologies and presents alternative ones.

Chapter two expands on the design, fabrication, and characterization of near field transducers forming the core of this thesis.

Chapter three investigates the heat issues associated in the design of HAMR heads and media. Numerical analysis is presented to evaluate heat induced by laser and their effect on the morphology of the disk lubricant.

Chapter four discusses the design of the HAMR spinstand and presents the far-field demonstration.

Chapter five presents protein based memory as an alternative to extend areal density growth, and investigates its key elements.

CHAPTER II: Design, Fabrication, and Characterization of Near-field Transducers

1. Theoretical Background

A. Far and Near Field Optics

Traditional optical microscopy, based on lenses, has been the most popular microscopical method in existence for over 300 years. Reasons for that are the ease of use of the microscope, low cost and identical optical contrast mechanisms as observed with the naked eye, e.g. spectroscopically resolved absorption, reflection and/or transmission contrast at non-homogeneous samples. Despite the advantages, the optical resolution is diffraction limited to approximately half the optical wavelength, i.e. about 300 nm. The next sections will show the origin of this limit and will reveal how to overcome this limit, by looking at the optical image formation [27].

Consider an object, of limited size, with a transmittance of $f(x, y, 0)$. The spatial frequency spectrum $F(u, v)$ can be written as,

$$F(u,v) = \iint f(x,y,0) e^{-j2\pi(ux+vy)} dxdy \qquad (1)$$

Because of the limited size of the object the spectrum contains frequencies from zero to infinity. If a plane wave illuminates the object, the field $U(x, y, 0)$ just behind the object (**Figure 6**) can be written as,

$$U(x,y,0) \equiv f(x,y,0) \qquad (2)$$

This can be described as the inverse Fourier transform of the spatial frequency spectrum

$$U(x,y,0) = \iint F(u,v) e^{j2\pi(ux+vy)} dudv \qquad (3)$$

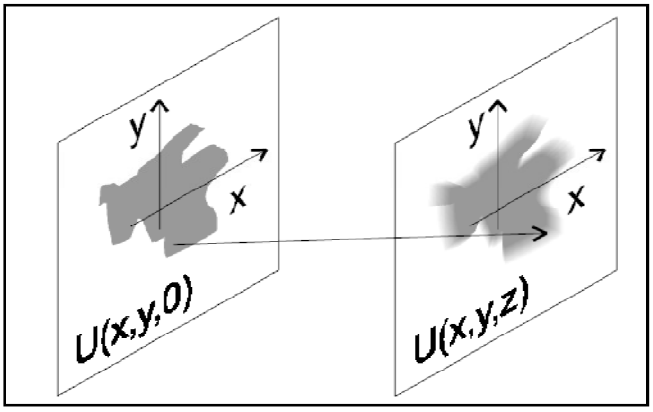

Figure 6: Optical image formation: field distribution at z due to the field at z=0

The field at the object can thus be regarded as a superposition of plane waves propagating in the direction $k = (k_x, k_y, k_z) = (\alpha, \beta, \gamma) 2\pi / \lambda$

With the direction cosines:

$$\alpha = \lambda u, \beta = \lambda v \text{ and } \gamma = \sqrt{1 - \alpha^2 - \beta^2} \qquad (4)$$

The spatial frequency spectrum contains the angular frequency spectrum of the object field $U(x, y, 0)$. Choosing the observation plane at distance z, the angular frequency spectrum can be written as,

$$F(\alpha k / 2\pi, \beta k / 2\pi, z) = F(\alpha k / 2\pi, \beta k / 2\pi) e^{jkz\sqrt{1-\alpha^2-\beta^2}} \qquad (5)$$

This equation reveals that only a limited range of angular frequencies propagates in the z-direction. Two cases can be distinguished:

- $\alpha^2+\beta^2<1$: the argument of the exponential function in the angular frequency spectrum is imaginary corresponding to a wave propagating in the z-direction towards the observation plane.
- $\alpha^2+\beta^2>1$: Angular frequency spectrum can be rewritten as

$$F(\alpha k / 2\pi, \beta k / 2\pi, z) = F(\alpha k / 2\pi, \beta k / 2\pi) e^{-kz\sqrt{\alpha^2+\beta^2-1}}$$

where the argument of the exponential function is real. As a consequence the amplitude of the wave decreases exponentially in the z-direction. Since α and β are proportional to the spatial frequencies u and v, it is seen that for low spatial frequencies and thus for low angular frequencies the waves propagate in the z-direction towards the observation plane. These components are the far-field components of the angular frequency spectrum. The high spatial frequency components are only present near the sample and decay exponentially in the z-direction. The region near the sample containing the high spatial frequency components is called the near-field zone.

In conventional microscopy, lenses with a limited numerical aperture (NA = sin (q)), are placed in the far-field. Consequently, only waves propagating with their k-vector within the NA will reach the detector,

$$\sin\left(\vec{z},\vec{k}\right) < NA$$

$$\alpha^2 + \beta^2 < NA^2$$

Using equation (4), this means that only spatial frequencies (u, v) which are smaller than NA/λ are detected, corresponding to lateral distances in $U(x, y, 0)$ larger than λ/NA. As a result, the maximum achievable resolution at the image plane is limited to λ/NA; the *diffraction limit*, Abbe [28].Similarly, Rayleigh defined the diffraction limit as the distance between two objects for which the intensity maximum of one object coincides with the first minimum in intensity of the second object. This resolution limit can be determined by looking at optical image formation [29] as, $0.61.\lambda\times/NA$, which is the Rayleigh criterion.

The conventional Optical Pick-up Head is considered a far field optical system, where the data is written and read using a conventional diffraction limited optical system and the smallest resolvable mark size is defined by Equation:

$$d_{mark} = \frac{\lambda}{4NA}$$

Where λ is the wavelength and *NA* is the numerical aperture. With DVD specifications (λ = 650 nm and *NA* = 0.6) *dmark* = 271 nm. The minimum mark size in DVD is set at 400nm to overcome the interference when bits are placed at *dmark* distance from each other due to the modulation of the far field signal. The formula suggests that in order to reduce the bit/mark size, λ has to be reduced and the NA has to be increased. However, the shortest wavelength laser diode can support is 375nm and the largest NA for an objective lens is 0.8[A]. As a result, fundamental limitations have been reached.

Equation (6) reveals that waves containing the high spatial frequency information of the object do not propagate but decay exponentially with the distance from the object. Near-field scanning optical microscopy (NSOM) is based on detection of these non-propagating evanescent waves in the near-field zone, in order to obtain the high spatial frequency information of the object. For detection of near-field light, a probe has to be brought into the near-field zone, close to the sample to either detect the near-field directly, by means of a nanometer-size detector, or to convert the evanescent waves into propagating waves and detect these in the far-field, by using a nanometer-size scatter source or a waveguide with subwavelength size aperture. As these methods detect the already present near-field this mode of operation is called the collection mode. Another method is to use the illumination mode in which high spatial frequency waves are introduced near the sample, by a sub-wavelength light source, and propagating waves,

resulting from an interaction between the nearfield and the sample, are detected in the far-field.

B. Diffraction Theory and its implication on super resolution

This chapter reference Bethe's work in 1942 to explain the diffraction theory and it's implications on super resolution. Other works are considered yet not utilized as they lacked vital elements in our design as the work done by Harrington and Auckland [30] which used the method of moments to calculate the transmission coefficients but not the spatial distribution; spatial distribution is of great importance in our study. In parallel to that, the work of Neerhoff and Mur [31] who used Green's function formulation to analyze field amplitudes within and beyond collimation area is also considered. However, they considered slits in the order of ~λ/2 whereas slits of ~λ/10 are of interest. Lastly, Kirchoff's solution is evaluated for a potential solution. Yet, it does not satisfy the boundary conditions whereas it assigns a non zero tangential component of the field in the metallic screen. Kirchoff's theory also assumes a scalar theory while electromagnetic solutions are vectorial. Bethe considers a hole in a cavity as a radiation aperture, and then he calculates the amount and spatial distribution of the emitted radiation. We shall reintroduce Bethe work as it presents the physics of the model we adopted for the design of our apertures. Bethe assumes E_0, H_0 to be the fields at the input side of the aperture fulfilling the boundary conditions at the location of the screen; thus,

$E_0 \tan = 0 \Leftrightarrow n \times E_0 = 0 \Leftrightarrow H_0 \cdot n = 0$ while E_{0n} and $H_{0\tan}$ are in general different from zero. On the output side of the aperture, field in zero approximation vanishes. In the hole, H and the normal component of E are discontinuous, creating the following boundary conditions in the hole:

$$E_{1\tan} = E_{2\tan} \qquad \text{inside the hole} \qquad (1.1)$$
$$E_{1\tan} = E_{2\tan} = 0 \text{ for } x = 0 \qquad \text{outside the hole} \qquad (1.2)$$
$$H_{2\tan} - H_{1\tan} = H_{0\tan} \qquad \text{in the hole} \qquad (1.3)$$

Boundary conditions for the normal components are automatically fulfilled if those for the tangential component are fulfilled. Thus, the boundary conditions for E_1 and H_1 are satisfied if E_2 satisfies boundary condition (1.2). So for any $x > 0$, and any y and z:

$$E_{1y}(-x, y, z) = E_{2y}(x, y, z) \qquad (1.4) \quad \Leftrightarrow \quad H_{1x}(-x, y, z) = H_{2x}(x, y, z) \qquad (1.6)$$
$$H_{1y}(-x, y, z) = -H_{2y}(x, y, z) \qquad (1.5) \quad \Leftrightarrow \quad E_{1x}(-x, y, z) = -E_{2x}(x, y, z) \qquad (1.7)$$

Inserting (1.5) into (1.3) we find that H_2 must satisfy the boundary condition

$H_{2\tan} = \frac{1}{2} H_{0\tan}$ (1.8) in the hole and from (1.7), we get $E_{2x} = \frac{1}{2} E_{0,x}$ (1.9) in the whole where $E_{0,x}$ and

H_0 are known from the input field. The problem now is to calculate E_2 and H_2 subject to boundary conditions (1.2),(1.8), and (1.9). H_{0y}, H_{0z}, and E_{0x} are considered constant over the aperture since it is assumed smaller than wavelength of the incident wave.

It was assumed by Stratton that if there are no currents or charges at the output of the aperture, the electric field at a point r in that space is given by the following equation:

$$E(r) = \frac{1}{4\pi} \int d\sigma \left[ikn \times H(r')\varphi - (n \times E(r')) \times \nabla\varphi - n \cdot E(r')\nabla\varphi \right] \qquad 1.10$$

where *n* is the unit vector in the direction of the inward normal to the surface and *r'* is the source point on the surface where the field is given. First term in the equation is omitted since there are no charges or currents ($d\sigma = 0$) . The second term would be produced by a distribution of "magnetic currents" over the hole, and instead of just assuming current

density to be proportional to $n \times E$, Bethe determined it so it satisfies both equations (1.8) and (1.9).

Now, the magnetic current density J^* (magnetoelectric) and the charge density ρ^* can be introduced into Maxwell's equations in the same way as electric charge and currents:

$$\nabla \cdot H = 4\pi\rho^* \qquad (1.10)$$

$$\nabla \times E + \frac{1}{c}\frac{\partial H}{\partial t} = -4\pi J^* \qquad (1.11)$$

The continuity equation relating (1.10) and (1.11) is

$$\nabla \cdot J^* + \frac{1}{c}\frac{\partial \rho^*}{\partial t} = 0 \qquad (1.12)$$

Assuming time dependence of all quantities to be e^{-iwt} so that the Green function $\varphi(r) = e^{ikr} / r$ is an outgoing spherical wave, so (1.11) becomes

$$\nabla \cdot J^* = i.k.\rho^* \text{ where } k = w/c.$$

Bethe used magnetic surface currents density (K) and surface charges density (η) where the first corresponds to the discontinuity of the tangential electric field at the surface (1.11) and the second corresponds to the discontinuity of the normal magnetic field at the surface. Both densities satisfy the continuity equation (1.12). Thus, can be written as

$$\nabla \cdot K = ik\eta \ .$$

The electric field and the magnetic field at the aperture can be expressed in terms of K and η with the help of scalar and vector potential.

$$E = \nabla \times F \qquad (1.14)$$

$$H = \frac{1}{c}\frac{\partial F}{\partial t} - \nabla\psi \qquad (1.15)$$

Both satisfying Maxwell's equations as well:

$$\nabla \cdot E = 0$$

$$\nabla \times H = \frac{1}{c}\frac{\partial E}{\partial t}$$

From (1.10) and (1.11) it is found that the vector potentials F and ψ are inexact analogy to the electric case. If only surface charges are present, the vector potentials are evaluated to be:

$$E(r) = \int K(r') \times \nabla \varphi . d\sigma \qquad (1.16)$$

$$H(r) = \int \left(ikK(r')\varphi - \eta(r')\nabla\varphi \right) d\sigma \qquad (1.17)$$

(1.16) already satisfies $E_{1\tan} = E_{2\tan} = 0$ for $x = 0$ outside the hole, and it also satisfies the condition that $E_{\tan}$ vanishes on the screen everywhere outside the hole. The tangential component of the electric field and magnetic fields in the hole are given by Stratton:

$$E_{\tan} = 2\pi n \times K \text{ and } H = 2\pi\eta$$

K and η are evaluated to be:

$$\eta = -\frac{1}{\pi^2 (a^2 - r'^2)^{1/2}} H_0 \cdot r' \qquad (1.18)$$

$$K = \frac{ik}{\pi^2}(a^2 - r'^2)^{1/2} \cdot H_0 \qquad (1.19)$$

Distribution of magnetic currents given in (1.19) must satisfy boundary conditions of equation (1.16) ruling that $E_{2n} = \frac{1}{2} E_{0n}$. However, the current density does not contribute appreciably to the electric field. Therefore, additional magnetic current distribution is required to fit the boundary condition of the normal component of the electric field without giving rise to additional magnetic charge density.

The additional current distribution K_E was estimated to be:

$$K_E = \frac{1}{2\pi^2 (a^2 - r'^2)^{1/2}} r' \times E_0$$

so that the total magnetic current density is modified to be:

$$K = K_H + K_E = \frac{1}{\pi^2}\left(ik(a^2 - r'^2)^{1/2} H_0 + \frac{1}{2(a^2 - r'^2)^{1/2}} r' \times E_0 \right)$$

Now the fields can be evaluated while the vector potentials are satisfying Maxwell's equations and all the boundary conditions. This results in the following:

$$E = \frac{1}{3\pi} k^2 a^3 \varphi_0 \kappa \times (2H_0 + E_0 \times \kappa)$$
$$H = \frac{1}{3} k^2 a^3 \varphi_0 \kappa \times (2H_0 \times \kappa - E_0)$$

Then the total radiation can be estimated as the poynting vector of the diffracted field:

$$S = \frac{c}{4\pi} E \times H = \frac{c}{27\pi^2} k^2 a^6 \left(4{H_0}^2 + {E_0}^2 \right).$$

Fields in and outside the whole are used to calculate power flow anywhere desired. The fields are estimated a circular aperture:

$$E_n = \tfrac{1}{2} E_{0,n}$$
$$H_{\tan} = \tfrac{1}{2} H_{0\tan}$$
$$H_n = 2\pi\eta = -\frac{2}{\pi} \frac{H_0 \cdot r}{(a^2 - r^2)^{1/2}}$$
$$E_{\tan} = \frac{1}{\pi} \frac{r}{(a^2 - r^2)^{1/2}} E_{0,z}$$

The results above are for fields within the hole showing that the normal electric and tangential magnetic are of half the value of the unperturbed fields, and that the tangential electric is directed radially outwards from the center of the hole. Outside the hole, $H_n = E_{\tan} = 0$ while E_n and $H_{\tan}$ behave as $\frac{1}{(a^2 - r^2)^{1/2}}$.

The tangential component of the magnetic field determines the dissipation of energy in the screen if the latter has a finite conductivity. Then the power dissipation at the exit is estimated to be: $P \sim H_0^2 vDa^2 \ln(a/D)$.

C. Waves from Apertures

This section sets the platform for analyzing the electromagnetic waves as they exit the aperture. It is vital to understand their behavior to estimate near-field intensity distribution, power throughput, and extrapolation of near-field patterns from far-field ones.

In free space, Maxwell's equations, which relate spatial and time variations of the electric and magnetic field intensities *E* and *H*, are

$$\nabla \times E = -\mu_0 (\partial H / \partial t)$$
$$\nabla \times H = \varepsilon_0 (\partial E / \partial t)$$
and
$$\nabla . E = \nabla . H = 0$$

where ε_0 and μ_0 are the free space permittivity. Combining and eliminating *H* yields

$$\nabla^2 E - \mu_0 \varepsilon_0 (\partial^2 E / \partial t^2) = 0$$

and similarly for *H*. In Cartesian coordinates, this equation becomes,

$$\nabla^2 E_x - \mu_0 \varepsilon_0 (\partial^2 E_x / \partial t^2) = 0$$

and similarly for E_y and E_z. If E_x varies with z only, we have

$$\frac{\partial^2 E_x}{\partial z^2} = \mu_0 \varepsilon_0 \frac{\partial^2 E_x}{\partial t^2} \qquad (1).$$

A time harmonic time variation of the fields is usually assumed (an arbitrary time variation of the field $E_x(t)$ can be expressed as the sum of a Fourier series of harmonic time variations, each harmonic having amplitude $a(w)$.

$$E_x(t) = \int_{-\infty}^{\infty} a(w) e^{jwt}\, dw \text{ where } a(w) = 1/(2\pi) \int_{-\infty}^{\infty} E_x(t) e^{-jwt}\, dt\,.$$

If the fields vary with time t as e^{jwt}, where $w = 2\pi f$ is the angular frequency, then $\partial / \partial t$ may be replaced by jw. A solution of equation (1) is then

$$E_x = A.\exp[-jk(z-ct)] + B\exp[jk(z+ct)]$$

where $k = w\sqrt{\mu_0 \varepsilon_0}$ is the free space propagation constant, $c = w/k = (\mu_0 \varepsilon_0)^{-1/2}$ is the free space wave velocity and A and B are the real constants. This complex exponential form of spatial and time variations of fields is used for convenience. It is understood that actual field quantities are the real part; i.e. for the plane wave in the z-direction the field is

$$E_x = \Re\{A.\exp[-jk(z-ct)]\}$$

Time-averaged power flux density is conveniently obtained from $\frac{1}{2}\Re\left(E \times H^*\right)$ where * indicates the complex conjugate.

Now, for finding solutions for this plane wave, consider and arbitrary direction ζ, fields behave exp (-$jk\zeta$). ζ measured from the origin of Figure 7, and with direction cosines $\sin\theta\cos\phi, \sin\theta\sin\phi,$ and $\cos\theta$.

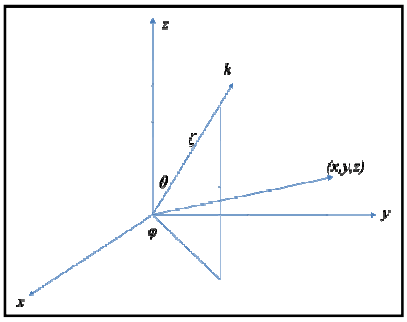

Figure 7: Coordinates for plane wave propagation

$$\zeta = x\sin\theta\cos\phi + y\sin\theta\sin\phi + \mathrm{z}\cos\theta$$

which defines a plane surface. A unit vector normal to this plane is

$$\hat{k} = \bar{k}/k = \hat{x}\sin\theta\cos\phi + \hat{y}\sin\theta\sin\phi + \hat{z}\cos\theta$$

where $\bar{k}$ is the vector propagation constant. Since $\bar{r} = \hat{x}x + \hat{y}y + \hat{z}z$, exp (-$jk\zeta$) can be written as

$$\exp(-j\bar{k}.\bar{r}) = \exp\left[-jk(x\sin\theta\cos\phi + y\sin\theta\sin\phi + z\cos\theta)\right]$$

For plane waves of this form, with spatial dependence of the fields on ζ only, the operator $\nabla = \hat{k}\,\partial/\partial\zeta$ can be replaced by $-j\bar{k}$ in Maxwell's equations to be expressed as:

$$\bar{k}\times E = w\mu_0 H$$
$$-\bar{k}\times H = w\varepsilon_0 E$$
$$\bar{k}\cdot E = \bar{k}\cdot H = 0$$

These equations yield $k = w\sqrt{\mu_0\varepsilon_0}$, the propagation constant magnitude, which when used in Maxwell's equation $\hat{k}\times E = Z_0 H$ where $Z_0 = \sqrt{\mu_0/\varepsilon_0} \cong 120\pi\Omega$ the wave impedance of free space is resulting in a time-averaged power flux in the direction of propagation $\frac{1}{2}\Re\left(E\times H^*\right)\cdot\hat{k} = \frac{1}{2}Y_0|E|^2$ where the plane wave admittance is $Y_0 = Z_0^{-1}$. These waves are considered homogenous if their equiamplitude surfaces, which are parallel planes, are also equiphase surfaces. A two-dimensional field uniform in, say, the y-direction in the α direction may be written as Figure 8:

$$\exp\left[-jk(x\sin\alpha + z\cos\theta)\right] = \exp\left[-jkr\cos(\theta-\alpha)\right]$$

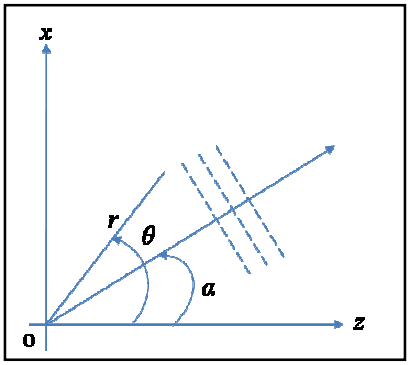

Figure 8: two-dimensional field uniform the α direction

where $x = r\sin\theta, z = r\cos\theta$. For real angles α, the wave is homogenous. When direction angles are complex, the wave becomes inhomogeneous

$$\exp[-jkr\cosh\alpha_2\cos(\theta-\alpha_1)]\exp[kr\sinh\alpha_2\sin(\theta-\alpha_1)]$$

which propagates in a direction α_1 from the z-axis and decays exponentially at right angles to this direction. Such waves are often able to propagate along a plane boundary between two media and are also found in waveguides operating at a frequency below cut-off frequency. Waves then propagate at the input and are exponentially attenuated down the guide axis.

A. Two and Three-Dimensional Waves Spectra

To analyze aperture antennas for the design of near-field optical heads, it is essential to evaluate the spectra of the wave in order to relate far and near field patterns. It was originated by Woodward and Lawson (1948) and Booker and Clemmow (1950), Collin and Rothschild (1963) and Rhodes (1964). The plane wave spectrum representation of electromagnetic fields is well described by Clemmow (1966).

Fields are assumed to be uniform in the y-direction and its electric field lies entirely in the x-z plane. Then any plane wave radiated in the direction α from the aperture in the $z = 0$ plane has an x-component of electric field given by

$$E_x(x,z) = A(\alpha)\exp[-jk(x\sin\alpha + z\cos\alpha)]$$

where $A(\alpha)$ is the amplitude of the wave. Consider a collection of such plane waves propagating in all directions α from the $z = 0$ plane with amplitude $A(\alpha)$ for each direction of propagation. The field equations are linear so each wave may be added and the total field got by summing over all directions α. It is convenient to let $k_x = k\sin\alpha, k_z = k\cos\alpha,$ and $F(k_x) = \lambda A(\alpha)$. The total E_x field in $z \geq 0$ is then

$$E_x(x,z) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(k_x)\exp[-j(k_x x + k_z z)]dk_x$$

where

$$k_z = \sqrt{k^2 - k_x^2} \text{ when } k_x^2 \leq k^2$$
$$= -j\sqrt{k_x^2 - k^2} \text{ when } k_x^2 > k^2$$

All real values of k are included in the above equation so that homogeneous and inhomogeneous plane waves are included. For the former $k_x^2 < k^2$ and the positive root is chosen for k_z in order that waves propagate outward in the z-direction, the latter, for which $k_x^2 > k^2$, are waves which propagate at complex angles. With k_z negative imaginary, they decay exponentially with increasing z. These evanescent waves transverse the aperture plane but do not carry energy away. They affect the reactance, rather than the resistance, of the antenna.

Putting $z = 0$ in the same equation gives the aperture field required to sustain the x-component of the total field

$$E_x(x,0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(k_x)\exp - j(k_x x)dk_x$$

$F(k_x)$ is usually called the angular spectrum of the field. The inverse Fourier transform of the aperture distribution may be written as

$$F(k_x) = \int_{-\infty}^{\infty} E(x,0)\exp(jk_x x)dk_x$$

The electric field vector of a plane wave is normal to the direction of propagation $\hat{k}$. Thus each plane wave here will generally have both x and z components, whose relative magnitudes may be calculated from $\hat{k}\cdot E = 0$. The total electric and magnetic fields are

$$E(x,z) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\hat{x} - \frac{k_x}{k_z}\hat{z}\right)F(K_x)\exp\left[-j(k_x x + k_z z)\right]dk_x$$

$$H(x,z) = \frac{\hat{y}Y_0 k}{2\pi}\int_{-\infty}^{\infty}\frac{F(K_x)}{k_z}\exp\left[-j(k_x x + k_z z)\right]dk_x$$

Two spectrum functions are required to specify an arbitrarily polarized field radiated from an aperture. These are the Fourier transforms of the two orthogonal components of aperture field, electric or magnetic. Fields in Figure 9 are polarized in the x-z plane; no y-component of electric field.

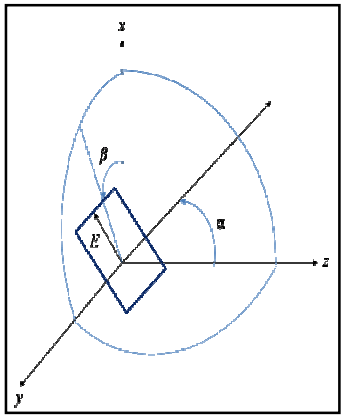

Figure 9: Plane wave radiating from an aperture in the z=0 plane.

The x-component of a plane wave in the direction α, β can be written

$$E_x(x,y,z) = \lambda^2 / F(k\sin\alpha\sin\beta, k\sin\alpha\sin\beta)\exp[-jk(x\sin\alpha\cos\beta + y\sin\alpha\sin\beta + z\cos\alpha]$$

Let $k_x = k\sin\alpha\cos\beta, k_y = k\sin\alpha\sin\beta,$ and $k_z = k\cos\alpha$

Then *x*-component of the total field for the complete spectrum of plane waves is

$$E_x(x,y,z) = \frac{1}{4\pi^2}\iint F(k_x,k_y)\exp[-j(k_x x + k_y y + k_z z)]dk_x dk_y$$

Where

$$k_z = \begin{cases} \sqrt{k^2 - (k_x^2 + k_y^2)}, & \text{when } k_x^2 + k_y^2 \le k^2 \\ -j\sqrt{(k_x^2 + k_y^2) - k^2}, & \text{when } k_x^2 + k_y^2 > k^2 \end{cases}$$

Waves for which $k_x^2 + k_y^2 \le k^2$, corresponding to real angles α, β, contribute to the radiation field of the aperture, whereas those for which $k_x^2 + k_y^2 > k^2$ correspond to complex angles α and β and form the evanescent field of the antenna aperture. At z=0, the aperture field is

$$E_x(x,y,0) = \frac{1}{4\pi^2}\iint F(k_x,k_y)\exp[-j(k_x x + k_y y)]dk_x dk_y$$

Where

$$F(k_x, k_y) = \iint E_x(x, y, 0) \exp\left[j(k_x x + k_y y)\right] d_x d_y$$

The z-component of the electric field in each plane wave satisfy the null of the convergence of the electric field with the k vector,

$$\bar{k} \cdot E = (\hat{x}k_x + \hat{y}k_y + \hat{z}k_z) \cdot (\hat{x}E_x + \hat{z}E_z) = 0$$

or

$$E_z(x, y, 0) = -\frac{k_x}{k_y} E_x(x, y, 0)$$

And the complete electric and magnetic fields can be represented as

$$E(x, y, z) = \frac{1}{4\pi^2} \iint (\hat{x}k_z - \hat{z}k_x) F(k_x, k_y) \exp\left[-j(k_x x + k_y y + k_z z)\right] \frac{dk_x dk_y}{k_z}$$

$$H(x, y, z) = \frac{Y_0}{4\pi^2 k} \iint (\hat{x}k_y k_x + \hat{y}(k^2 - k_y^2) - \hat{z}k_y k_z) F(k_x, k_y) \exp\left[-j(k_x x + k_y y + k_z z)\right] \frac{dk_x dk_y}{k_z}$$

For an aperture field linearly polarized in the y-z plane, with no electric field in the x-direction, the radiated fields can be written as

$$E(x, y, z) = \frac{1}{4\pi^2} \iint (\hat{y}k_z - \hat{z}k_y) F(k_x, k_y) \exp\left[-j(k_x x + k_y y + k_z z)\right] \frac{dk_x dk_y}{k_z}$$

It is known that in the far field the angular spectrum is the radiation pattern of the aperture. However, here we are not interested in apertures that have physical dimensions greater than λ generating propagation fields. We are to discuss evanescent waves that occur in the near field zones due to apertures smaller than the wavelength of light illuminated and thus generate spots smaller than that diffraction limited in the far field.

B. Near-field radiation patterns

In far field optics, radiation patterns are considered at distances from the aperture sufficiently large that all ray paths from the aperture to the field point are essentially parallel. In near field optics, this assumption is removed and the radiation field in what is known in optics as Fresnel zone of the aperture.

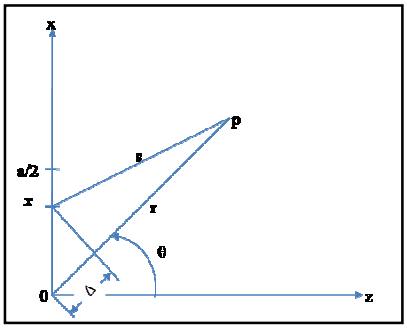

Figure 10: Path length differences in the near field of the aperture

Figure 10 shows ray paths *xp and op* in the x-z plane which is not parallel. From the geometry

$$s^2 - r^2 = (s-r)(s+r) = x^2 - 2rx\sin\theta$$

If Δ is the path difference at p of radiation from o and x in the aperture $(s-r) = -\Delta$ and, since $s + r \approx 2r$,

$$\Delta \approx \frac{-x^2}{2r} + x\sin\theta\,.$$

The effect of a finite range on the path difference is the quadratic term which vanishes as $r \to \infty$. The usual requirements for the far field patterns is that the maximum error in path length due to finite ranges is less than $\lambda/16$ for radiation from all parts of the aperture. For an aperture of width a the criterion is $a^2/8r < \lambda/16 \to r > \frac{2a^2}{\lambda}$

This is rather arbitrary division between the near field or Fresnel zone of an aperture and the far field. The effect of the finite range on the radiation pattern will depend also on the aperture distribution. The patterns of distributions highly tapered towards the aperture edges will be less affected by a particular finite range than the patterns of more uniform distributions. In consequence, this criterion may be inadequate in the latter case and more than adequate in the former. Near-field effects are also larger for aperture distributions tapered in phase towards aperture edges than for in-phase distributions.

The phase difference between contributing waves from $\Delta \approx \frac{-x^2}{2r} + x\sin\theta$ is $2\pi\Delta/\lambda = -\beta x^2 + k_1 x$ where $\beta = \pi/r\lambda$ and $k_1 = k\sin\theta$. If the aperture field distribution is $E_1(x)$, the near-field radiation pattern is

$$D_1(k_1) = \int_{-\infty}^{\infty} E_1(x)\exp(-j\beta x^2)\exp(jk_1 x)dx$$

which in optics is called a Fresnel transform.

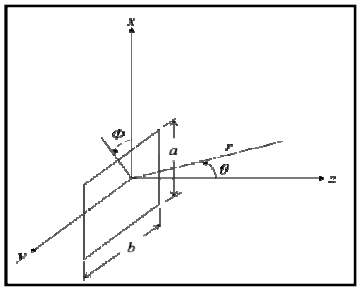

Figure 11: Coordinates for radiation from a rectangular aperture

In the *y*-*z* plane, the near-field radiation pattern is

$$E(r,\theta,\phi) = AD_1(k_1)D_2(k_2),$$

where

$$D_1(k_1) = \int_{-a/2}^{a/2} E_1(x) \exp\left[-j(\beta x^2 - k_1 x)\right] dx$$

$$D_2(k_2) = \int_{-b/2}^{b/2} E_2(y) \exp\left[-j(\beta y^2 - k_2 y)\right] dy$$

if $E_1(x) = 0, |x| > a/2$ and $E_2(y) = 0, |y| > b/2$

where

$$k_1 = k \sin\theta \cos\phi$$
$$k_2 = k \sin\theta \sin\phi$$

For simplicity, since the stated integrals are too involved than those of the far-field radiation pattern but can be reduced to Fresnel integrals, consider $E_1(x) = 1$, therefore,

$$\begin{aligned} D_1(k_1) &= \exp\left[j(k_1^2/4\beta)\right] \int_{-a/2}^{a/2} \exp\left[-j\beta(x - k_1/2\beta)^2\right] dx \\ &= \sqrt{\frac{r\lambda}{2}} \exp\left[j(r\lambda k_1^2/4\pi)\right] \int_{u_1}^{u_2} \exp\left[-j(\pi/2)u^2\right] du \\ &= \sqrt{\frac{r\lambda}{2}} \exp\left[j(r\lambda k_1^2/4\pi)\right] \{C(u_2) - C(u_1) - j[S(u_2) - S(u_1)]\} \end{aligned}$$

where

$$u_{1,2} = \pm \frac{a}{\sqrt{2r\lambda}} - \frac{k_1}{2\pi}\sqrt{2r\lambda}$$

and the Fresnel integrals defined by

$$\int_0^{v_1} \exp\left[-j(\pi/2)v^2\right] dv = C(v_1) - jS(v_1)$$

are tabulated by Jahnke and Emde, 1945; Abramowitz and Stegun, 1964.

The near-field pattern for a co sinusoidal distribution can be treated similarly,

$$E_2(y) = \cos\frac{\pi y}{b} = \frac{1}{2}\{\exp[j(\pi y/b)] + \exp[-j(\pi y/b)]\}$$

then,

$$D_2(k_2)=\frac{1}{2}\left(\exp\left[j\frac{1}{4\beta}\left(k_2+\frac{\pi}{b}\right)^2\right]\int_{-b/2}^{b/2}\exp\left\{-j\beta\left[y-\frac{1}{2\beta}\left(k_2+\frac{\pi}{b}\right)\right]^2\right\}\times dy+\exp\left[j\frac{1}{4\beta}\left(k_2-\frac{\pi}{b}\right)\right]\int_{-b/2}^{b/2}\exp\left\{-j\beta\left[y-\frac{1}{2\beta}\left(k_2-\frac{\pi}{b}\right)\right]^2\right\}dy\right)$$

$$=\frac{1}{2}\sqrt{\frac{r\lambda}{2}}\left(\exp\left[j\frac{r\lambda}{4\pi}\left(k_2+\frac{\pi}{b}\right)^2\right]\int_{v_1}^{v_2}\exp\left[-j(\pi v^2/2)\right]dv+\exp\left[j\frac{r\lambda}{4\pi}\left(k_2-\frac{\pi}{b}\right)^2\right]\int_{w_1}^{w_2}\exp\left[-j(\pi w^2/2)\right]dw\right)$$

$$=\frac{1}{2}\sqrt{\frac{r\lambda}{2}}\left(\exp\left[j\frac{r\lambda}{4\pi}\left(k_2+\frac{\pi}{b}\right)^2\right]\times\{C(v_2)-C(v_1)-j[S(v_2)-S(v_1)]\}+\exp\left[j\frac{r\lambda}{4\pi}\left(k_2-\frac{\pi}{b}\right)^2\right]\times\{C(w_2)-C(w_1)-j[S(w_2)-S(w_1)]\}\right)$$

where

$$v_{2,1}=\pm\sqrt{\frac{b}{2r\lambda}}-\frac{k_2}{2\pi}\sqrt{2r\lambda}-\frac{1}{b}\sqrt{\frac{r\lambda}{2}}$$

$$w_{2,1}=\pm\sqrt{\frac{b}{2r\lambda}}-\frac{k_2}{2\pi}\sqrt{2r\lambda}+\frac{1}{b}\sqrt{\frac{r\lambda}{2}}$$

Aperture distributions which are sinusoidal or cosinusoidal functions can be dealt with in this way, by expressing the entire integrand as a complex exponential and completing the square to reduce the expression to Fresnel integrals. Compound distributions can be decomposed into simple distributions and so treated also, but direct numerical integration will be used.

2. Alternatives to Dodge the Diffraction Limit

Recent advances in near-field optics achieve spatial resolution significantly better than the diffraction limit. Solid immersion lenses [32,33], apertures on good metallic conductors [34,35], bow-tie antennas [36,37], ridge waveguides[38, 39, 40], tapered optical fibers[41,42] silicon pyramidal probes [43] and apertureless near-field microscopy [44] are among the possible ways to achieve intense optical spots with small sizes.

The so-called “aperture-NSOM” exploits sensors based on tapered optical fibers, ending with a nanoscopic aperture (30–100 nm). The aperture diameter is defined by a lateral metallization with a thin film (typically Au or Al 100 nm thick) deposited by thermal evaporation or sputtering techniques. Multiple reflections at the metallized sides confine the light towards the apical aperture. Such reflections have a twofold effect: they reduce the throughput of the probes (to 10^{-6}–10^{-3}), and increase the total amount of light absorbed by the metal coating (20%–30%) of the coupled power, to be compared to 8% of the aluminum absorption at normal incidence. As a result, the metallic film can heat up to several hundreds of degrees [45] and undergo thermal expansion phenomena.

Possible utilization of the aforementioned transducers in a HAMR system is currently under investigation. Several of these transducers, however, are not suitable for HAMR. For solid immersion lenses with high index materials and small wavelengths, the theoretical FWHM formula yields minimum spot sizes of 55–65 nm [46] which are too large. The transmission efficiency of the tapered optical fiber or the silicon pyramidal probe is too low for a HAMR system. The performance of the other transducers requires investigation, especially in the presence of a magnetic recording medium.

The ridge waveguide has been widely used in microwave and antenna systems because of its low cutoff frequency, wide bandwidth, and low impedance. In the literature, the field transmission through ridge waveguides has been extensively studied at microwave frequencies [47/48/49]. The ridge waveguide has been recently suggested for optical data storage, because of its ability to generate intense fields in the gap between the ridge and the opposite surface of the waveguide. This was demonstrated by three-dimensional finite

element method (FEM) simulations of ridge waveguides in perfect conductors surrounded by free space.

The following section summarizes some design tips and design procedures of nanoapertures with strong field enhancement, based on the theoretical calculations and analyses of near-field distributions of unconventional ridge apertures by FEM.

A. Numerical Analysis of single ridge waveguides using FEM

A waveguide which has been widely used in many broadband microwave equipments is the ridge waveguide. The original ridge waveguide consisted of a regular rectangular waveguide with one or two ridges inserted. The main advantage of ride waveguides is the wider separation between the cut-off numbers of its dominant and first higher order modes resulting in longer cutoff wavelength than conventional rectangular waveguides having the same internal dimensions. Another advantage is the fact that it their low characteristic impedance is bracketed between that of the regular rectangular waveguide (377Ω) and those of coaxial structures (50 Ω).

This work utilizes ridge waveguides to focus visible light beyond the diffraction limit in the near field vicinity of the waveguide, so knowledge of the cutoff wavelength and characteristic impedance is crucial for design purposes. In this work a reference to Chen's work [50] is made to calculate of the dominant mode characteristics of rectangular waveguides.

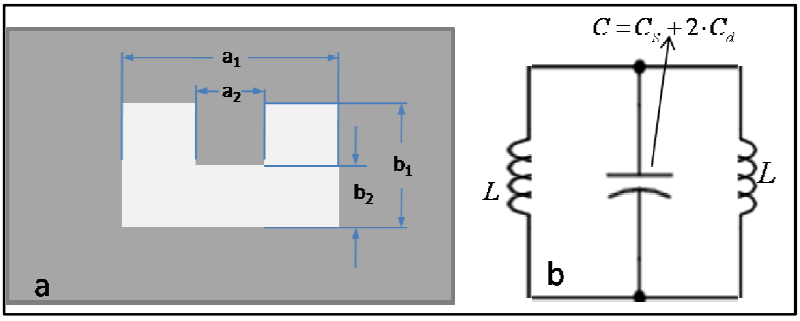

Figure 12: Cross section view of (a) single-ridge and (b) equivalent lumped circuit

Figure 12 represents single ridge (a, b) waveguides in a lumped-constant equivalent circuit. The capacitance in parallel with the two inductors consists of the electrostatic capacitance, C_s, and the discontinuity capacitance, C_d. During dominant mode operation of the waveguide, C_s mainly depends on the region between the ridge and the bottom plate, where a strong electric field exists. Electrostatic capacitance is approximated in terms of the permittivity of the medium and its characteristic inner and outer dimensions:

$$C_s = \frac{\varepsilon \cdot a_2}{b_2}$$

The ridge in the waveguide represents discontinuities to the electromagnetic waves causing local and higher order waves. The effect of local charges is accounted for by inserting discontinuity susceptance at the proper locations. A series for the discontinuity capacitance was developed by Whinnery [51] using Hahn's method of field matching [52] then compared to Schwartz-Christoffel fringing capacitance in a constricted conductor [53].

$$C_d = \frac{\varepsilon}{\pi}\left[\frac{(b_2/b_1)^2+1}{b_2/b_1}\cosh^{-1}\left(\frac{1+(b_2/b_1)^2}{1-(b_2/b_1)^2}\right) - 2\ln\frac{4(b_2/b_1)}{1-(b_2/b_1)^2}\right]$$

The total capacitance of the waveguide is then

$$C = \frac{\varepsilon \cdot a_2}{b_2} + 2C_d$$

The inductances in the equivalent circuit are determined by the sections of the waveguide on both sides of the ridge. Inductance to either section is given in terms of the permeability of the medium and the section characteristic dimensions:

$$L = \frac{\mu(a_1 - a_2)}{2}(b_1)$$

It is vital to mention that cutoff frequencies remains unchanged for double and single ridged waveguides. However, for double ridge waveguides, the capacitance is halved and the inductance is doubled.

At cutoff, the waves travel from side to side in the crosswise direction without longitudinal propagation representing anti-resonance condition in the lumped circuit. Therefore, cutoff frequency of the ridge can be derived to be:

$$f_c^{'} = \frac{1}{2\pi\sqrt{(L/2)C}} = \frac{1}{\pi\sqrt{\mu\varepsilon}\sqrt{\left(\frac{a_2}{b_2} + \frac{2C_d}{\varepsilon}\right)(a_1 - a_2)(b_1)}}$$

Knowledge of the fields in any waveguide is necessary in order to calculate its power flow, its attenuation and its impedance. The purpose of this section is to summarize some approximate closed form relationships for the fields in this type of waveguide of waveguide and present some calculations based on the finite element method (FEM) procedure.

The finite element method (FEM) is a numerical technique for finding approximate solutions of partial differential equations (PDE) as well as of integral equations. The solution approach is based either on eliminating the differential equation completely (steady state problems), or rendering the PDE into an approximating system of ordinary differential equations, which are then solved using standard techniques such as Euler's method, Runge-Kutta, etc.

In solving partial differential equations, the primary challenge is to create an equation that approximates the equation to be studied, but is numerically stable, meaning that errors in the input data and intermediate calculations do not accumulate and cause the resulting output to be meaningless. There are many ways of doing this, all with advantages and disadvantages. The Finite Element Method is a good choice for solving partial differential equations over complex domains, when the domain changes, when the desired precision varies over the entire domain. Its development can be traced back to the work by Alexander Hrennikoff (1941)[54] and Richard Courant (1942) [55]. While the approaches used by these pioneers are dramatically different, they share one essential characteristic: mesh discretization of a continuous domain into a set of discrete sub-domains, usually called elements. Hrennikoff's work discretizes the domain by using a lattice analogy while Courant's approach divides the domain into finite triangular subregions for solution of second order elliptic partial differential equations (PDEs) that arise from the problem of torsion of a cylinder.

B. Methodology and Results

The finite element method (FEM) is a well known numerical algorithm for the solution of Maxwell's equations [56/57]. Triangular elements are used to discretize the computational sub domains and boundaries, which allows modeling of arbitrarily shaped three-dimensional geometries. Edge basis functions and second order interpolations functions are used to expand the functions. Adaptive mesh refinement is employed to improve the coarse solution regions with high field intensities and large field gradients Figure 13.

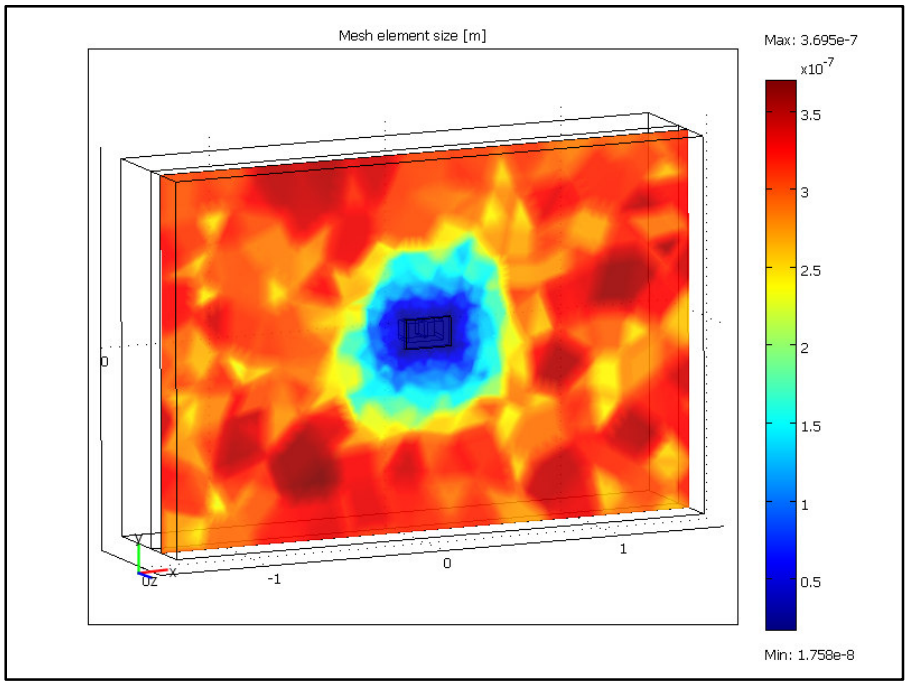

Figure 13: FEM mesh case: refined element size at subdomains and boundaries of interest

Debye equation is used to model the frequency dependant dielectric function of the materials. The first order Debye equation can be written as:

$$\varepsilon = \varepsilon_{\infty} + \frac{\varepsilon_s - \varepsilon_{\infty}}{1 + iwt_0}$$

where ε_∞ is the infinite frequency dielectric constant, ε_s is the zero frequency dielectric constant, w is the angular frequency and t_0 is the relaxation time. The optical properties of materials used in this simulation are retrieved from literature [58/59] and summarized in the following table:

Material	Wavelength (nm)	Dielectric index*(n, k)*	Infinite permittivity	Static permittivity	Conductivity [$(\Omega m)^{-1}$]	Relaxation Time (s)
Gold	516	(0.61, 2.11)	1			
Gold	550	(0.36, 2.69)	1	−149.95	1.093e6	1.223e−15
Gold	700	(0.16, 3.95)	1	−2831.23	5.173e6	4.843e−15
Cobalt	516	(1.95, 3.65)	1			
Cobalt	700	(2.30, 4.43)	1	−23.21	7.673e5	2.793e−16
Silver	516	(0.13, 3.06)	1			
Aluminum	516	(0.82, 6.26)	1		3.07e6	

The model makes use of the predefined port boundary condition. The input matched boundary condition consists of two parts: an incident planar wave and an absorbing boundary condition. The matched boundary condition is also used at the output boundaries to eliminate any reflections there. At the output boundaries there is no excitation. The walls of the waveguide are considered to be good conductors, so the perfectly electric conductive (PEC) boundary condition can be used.

For specifying the absorbing boundary condition, the propagation constant, β, of the wave has to be defined. The propagation constant is determined from eigenmode analysis of the waveguide cross section. Entering the correct propagation constant eliminates all waves with that wavenumber in the propagating direction. To make the boundary condition perfectly absorbing, the propagating direction needs to be the same as the

normal direction of the output boundary. All ports are assumed to be connected to matched loads, that is, there is no reflection directly at a port.

The propagation constant can be given in three different ways. One method is to only specify the propagation constant β for the current frequency. This identical to the way the matched boundary condition is specified. The only difference is that the port boundary condition uses a normalized power level.

The next method is to specify a reference frequency ν_0 and the propagation constant β_0 for that frequency. The propagation constant as a function of frequency is then calculated as

$$\beta = \sqrt{\beta_0^2 - 4\pi^2 \varepsilon\mu(\nu_0^2 - \nu^2)}$$

The third option is to specify the cutoff frequency, ν_{cutoff}. The propagation constant as a function of frequency is in this case calculated as

$$\beta = \sqrt{4\pi^2 \varepsilon\mu(\nu_0^2 - \nu_{cutoff}^2)}\,.$$

Figure 14 explains how the port boundary condition is utilized to excite the waveguide embedded in metal film. The surrounding boundaries, as explained above, are modeled as matched boundary conditions to eliminate any reflections of the guided wave by perfectly absorbing it. Propagation constant is calculated to be $3.14e^{15}$ as the excitation port is 3x2 μm^2. The front boundaries of the metal are modeled as PEC, while all the internal boundaries generated by the waveguide follow the continuity condition.

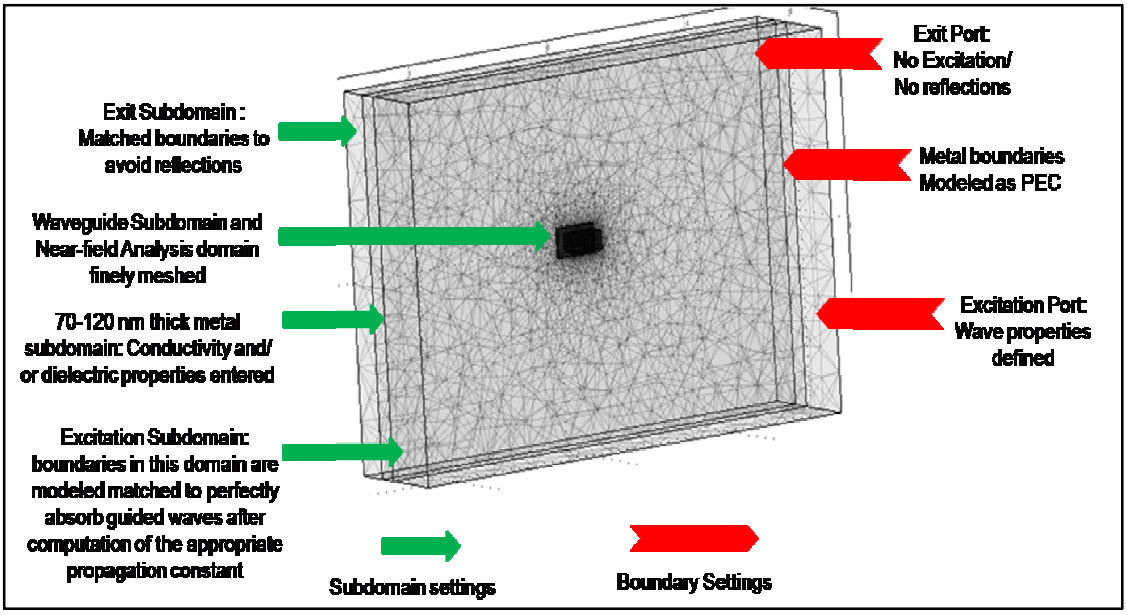

Figure 14: Meshed geometry of the modeled waveguide

The first model is to corroborate the poor power throughput of conventional shaped apertures, which is the motivation to design new aperture waveguides that can enhance power throughput significantly. Figure 10 shows that almost extremely low throughput is observed when the waveguide characteristic dimension is smaller than 200nm. Even at that size, the circular aperture contains the light merely, rather than focusing it.

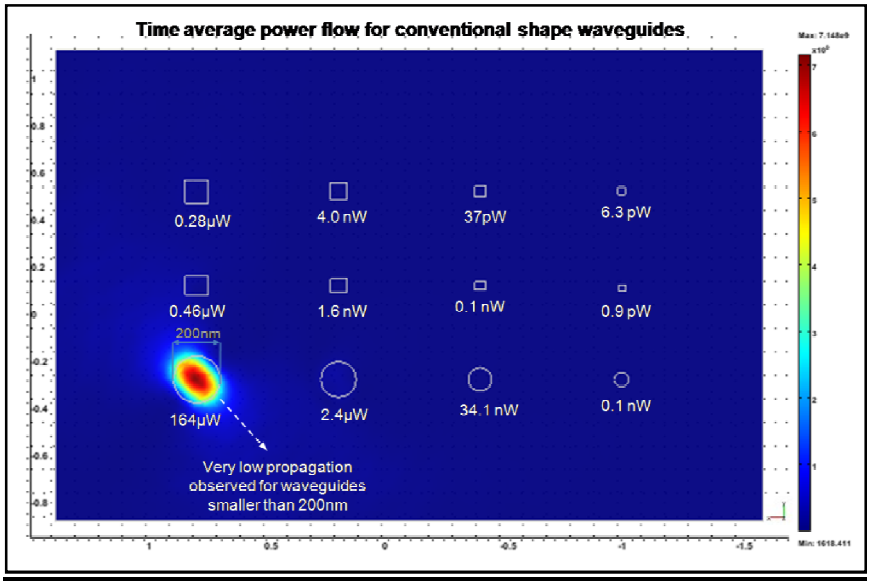

Figure 15: Power output of conventional shaped waveguides

Now, the performance of the ridge waveguide as a near-field aperture for data storage systems is investigated. The ridge waveguide geometry is illustrated in Figure 15. Gold, Silver, and Aluminum will be considered for the metal choice, thicknesses will range from 70-120 nm based on the skin depth of the material at the selected wavelength; the wavelength spectrum will envelop the visible regime. The structure is illuminated by TE or TM- polarized plane wave propagating in the z-direction.

The gap (a_2) defines the focusing capabilities of the waveguide when the other elements of the structure are contributing appropriately to the cutoff wavelength and charge allocation. In the HAMR system, the head flies anywhere between 5-15 nm above the magnetic disk, so all our analysis on the spot size, intensity distribution, and power throughput will be concentrated in the extreme near-field vicinity of the head. The power focused by the transducer will be computed by integrating the Poynting vector across a boundary (200x200 nm^2) at various distances from the exit plane of the head.

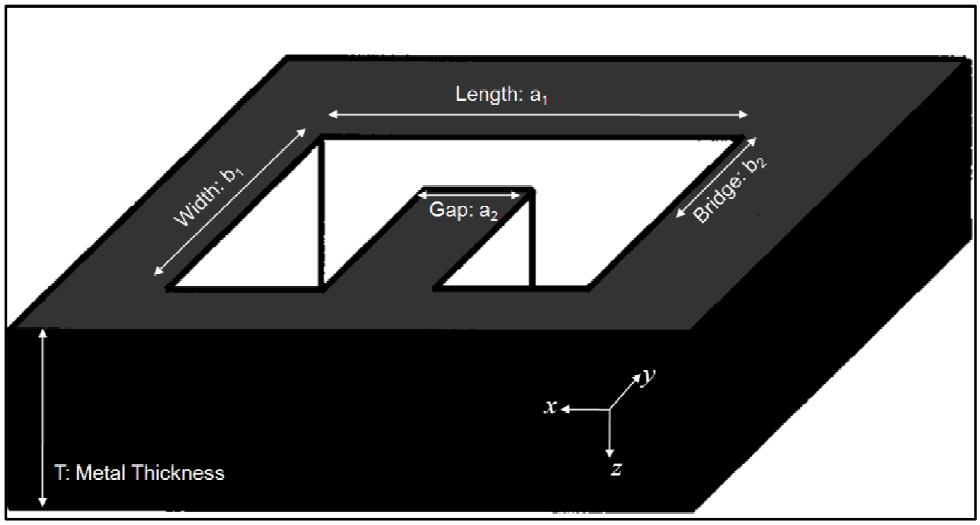

Figure 16: Single Ride Waveguide Transducer

The optical spot size and transmission efficiency are greatly enhanced by choosing optimal waveguide dimensions, metal thickness, material properties, and the wavelength used. It is vital to mention that appropriate selection of the mentioned elements is a directly dependant on exciting surface plasmon resonances, which are the biggest contributors to transmission enhancement at optical frequencies. Previous studies [60] of surface palsmons effects in bow tie antennas indicated that the best performance was obtained by using silver at short wavelengths of the visible spectrum. Gold was found to damp the surface plasmon modes at short wavelengths. The material properties of silver support surface plasmons at higher frequencies with smaller damping. This enables smaller spot sizes as well as higher transmission efficiencies. As a result, Silver is selected as the metal, and the operational wavelength is selected as 405 nm.

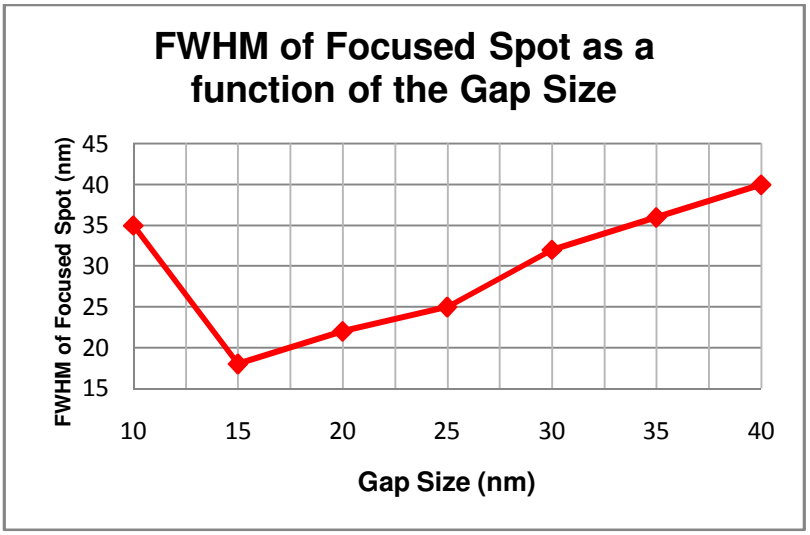

Figure 17: Spot size dependence on the gap size

First, the effect of the length and the width of the waveguide on the output power is investigated. Since the interest is mainly in spot sizes smaller than 30nm, and knowing that the gap in single ridge structures plays the role of the “effective ridge” to focus light

(Figure 16, Figure 17), the gap a_2 is selected to be 25nm. The thickness of the metal is chosen to be five times its skin depth at 458 nm, any thickness over ~45nm should attenuate the light completely (Figure 18[61]), so 70nm thickness should be adequate. It is crucial for the thickness of the metal not to exceed the threshold thickness, as more attenuation of the propagating modes along the waveguide can happen. When an aperture in a film of nonnegligible thickness is irradiated at the aperture resonance wavelength, transmission decays rapidly with increasing film thickness, because evanescent fields decay rapidly along the aperture channel and many fewer charges can be excited at the exit plane to radiate. However, when operating at wavelengths shorter than the waveguide cutoff, additional thickness induced resonances can be observed that are facilitated via the waveguide propagation mode. Total transmission in a thick metal film depends on the coupling of the surface-charge oscillation and the charge oscillation inside the aperture channel. A large departure from either the surface resonance condition or the thickness resonance condition leads to little coupling of the charge to the exit plane and in poor transmission.

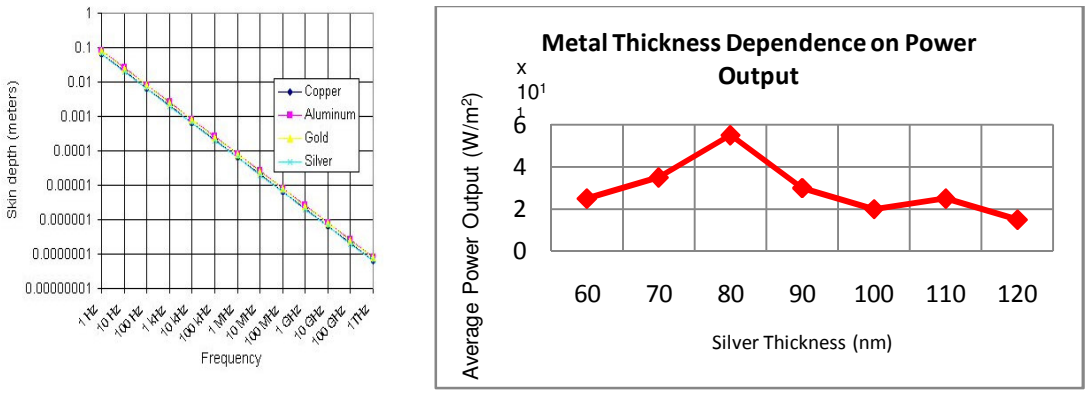

Figure 18: Skin Depth of various metals as a function of incident light frequency, and silver thickness dependence on power output

Results suggests that the width of the waveguide is selected between 75 and 105 nm to maintain a small focused spot size down to 30nm and for maximum power throughput

(Figure 19). The suggestion of Schlesinger *et al* [62] to keep the sum of the length and the width of the waveguide to equal half the illuminating wavelength is adopted.

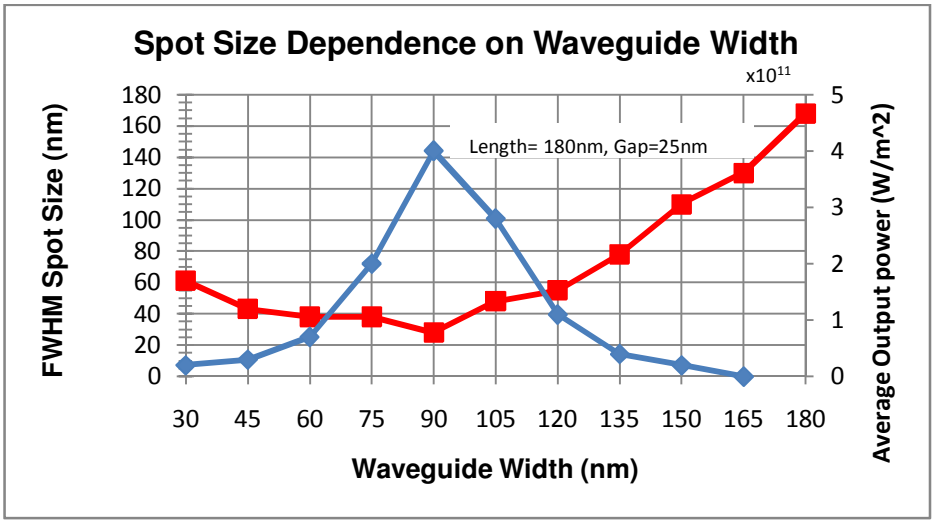

Figure 19: Power output and spot size dependence on the waveguide width

Knowing the optimal width of the waveguide, the length of the waveguide is varied to investigate the effects on power and spot size as well. It is recognized that the length of the waveguide should range between 180 and 210 nm to maintain small spots and high power throughput (Figure 20).

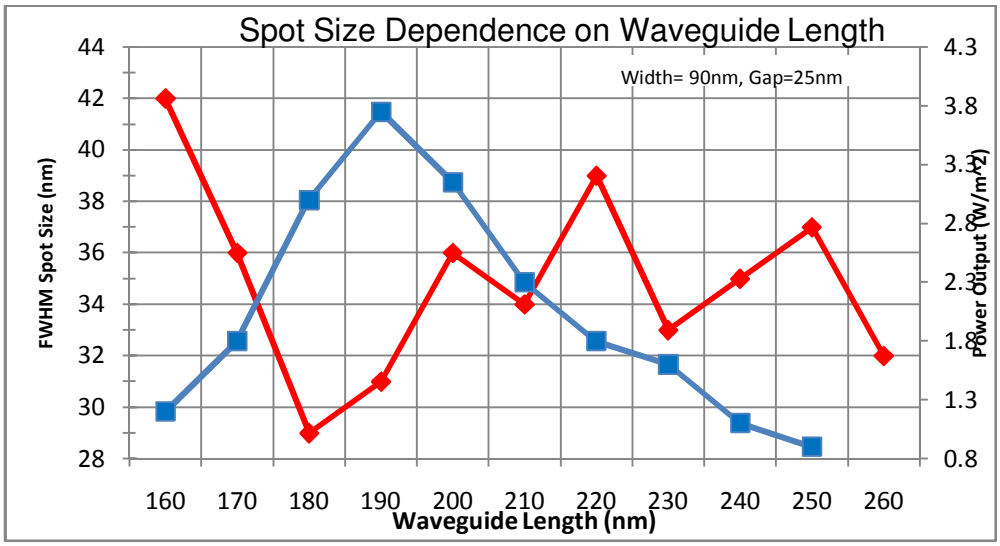

Figure 20: Power output and spot size dependence on the waveguide length

Finally the effect of the bridge is considered. The bridge determines the intensity of the fields of the electric dipole resulting from charge separation between the ridge and the opposite side of the waveguide. Selecting a large bridge will result in a weaker electric dipole due to large charge separation. Conversely, selecting a small bridge will result in a very close charge separation, and a weak electric dipole fringing field to interact with the medium. Figure 21 suggests selecting the bridge to be between 25 to 30 nm.

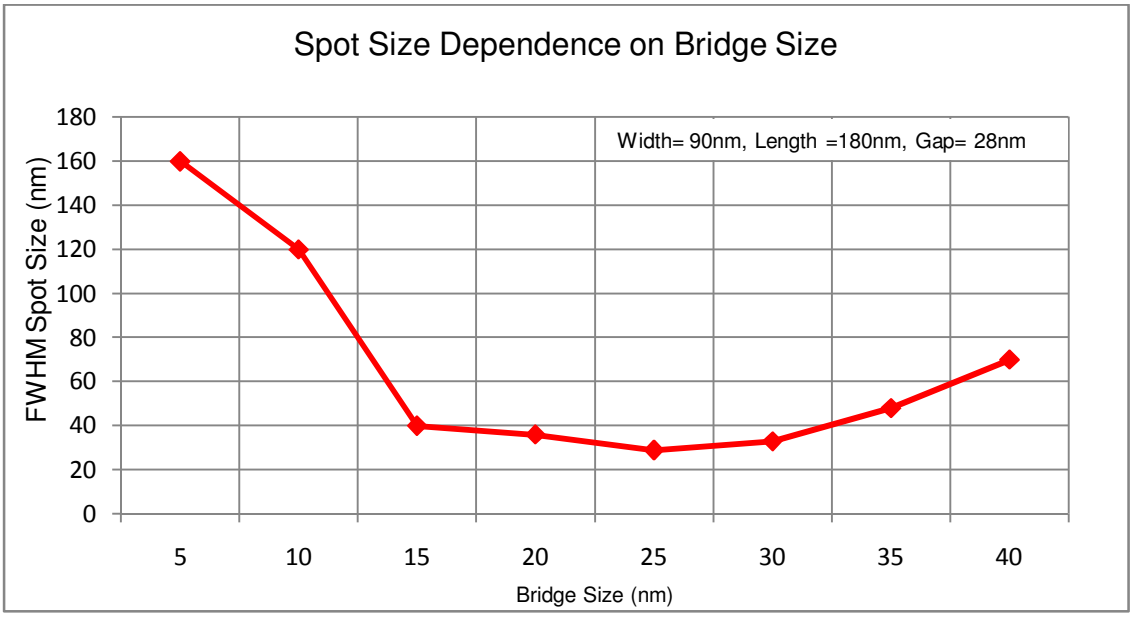

Figure 21: FWHM of spot size dependence on the bridge size

Accordingly, the optimal waveguide is designed according to the suggested dimensions. Figures show that the gap evidently functions as the effective ridge. Light is focused in the vicinity of the gap and well determined by its size. This can be related to the literature we previously discussed concerning the accumulation of charges on the edges. As incident waves arrive at the metal plane, current is induced by the incident electric field, and the presence of the aperture forces the current paths to separate at one edge and recombine at the opposite edge in the polarization direction. The continuity equation dictates that charges of opposite signs are deposited at the back edge and the ridge of the

C shape as current flows around the aperture. As the induced current oscillates back and forth with the incident frequency, the induced charges also oscillate, much like an oscillating dipole. The transmission enhancement that is absent in the conventional metal waveguide is due to the radiation from the oscillating dipole formed by the induced charges.

The power focused by the ridge is measured by integrating the Poynting vector across the boundary of interest. Figure 15 shows that for 200mW of TE_{10} mode light, 3.53mW is coupled into the waveguide, of which 2.19mW is transmitted through the aperture to the near-field of the transducer (Figure 22) .

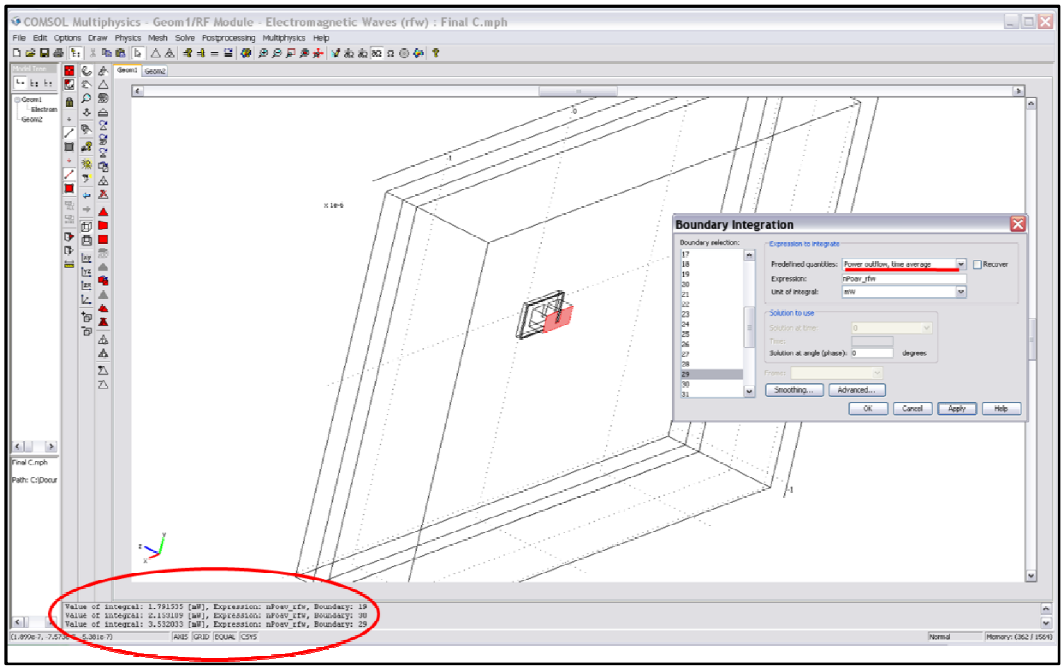

Figure 22: Snapshot of power calculation

Now, a comparison of the performance of Gold and Silver under the same simulation conditions is shown. Even though the model slightly accounts for surface plasmon resonance, Silver is still capable of focusing more light into a slightly smaller spot size

than that done by Gold (Figure 23). For silver a broad maximum in transmitted power occurs at a wavelength of 400 nm corresponding to a surface plasmon (SP) resonance. Surface plasmons, which are collective longitudinal oscillations of surface charge with their associated electromagnetic field cannot be modeled in ideal conductor which does not allow an electric field in the longitudinal direction at its surface.

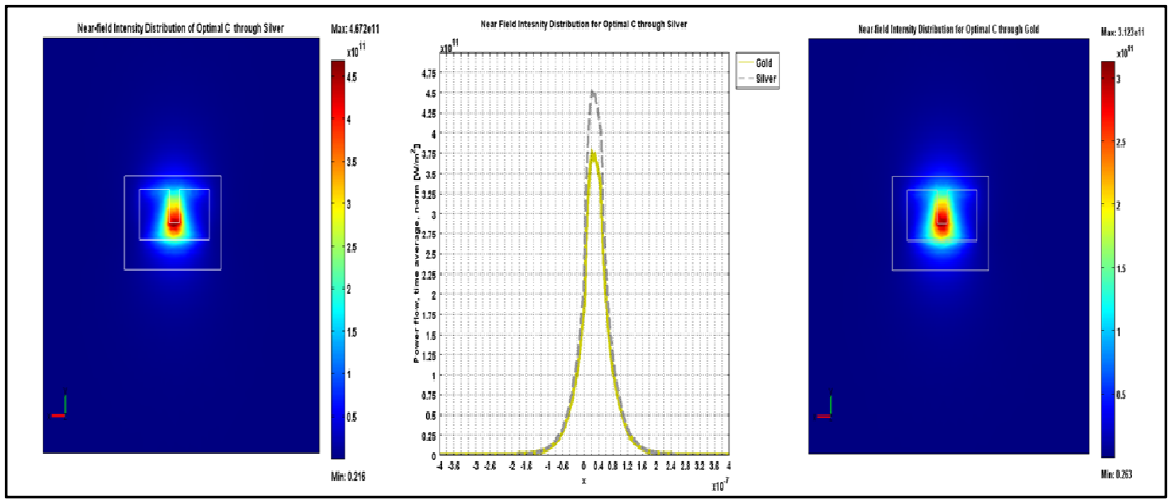

Figure 23: Near-field Intensity comparison between Silver and Gold

The Near-field intensity distribution is a direct correspondence to the spatial distribution of the focused spot. The near-field distribution of a small aperture can be well confined by the aperture size up to a distance of z $=a_1/2$. Inside this region, the field distribution evolves with the distance z, while the peak intensity drops as z increases (Figure 24).

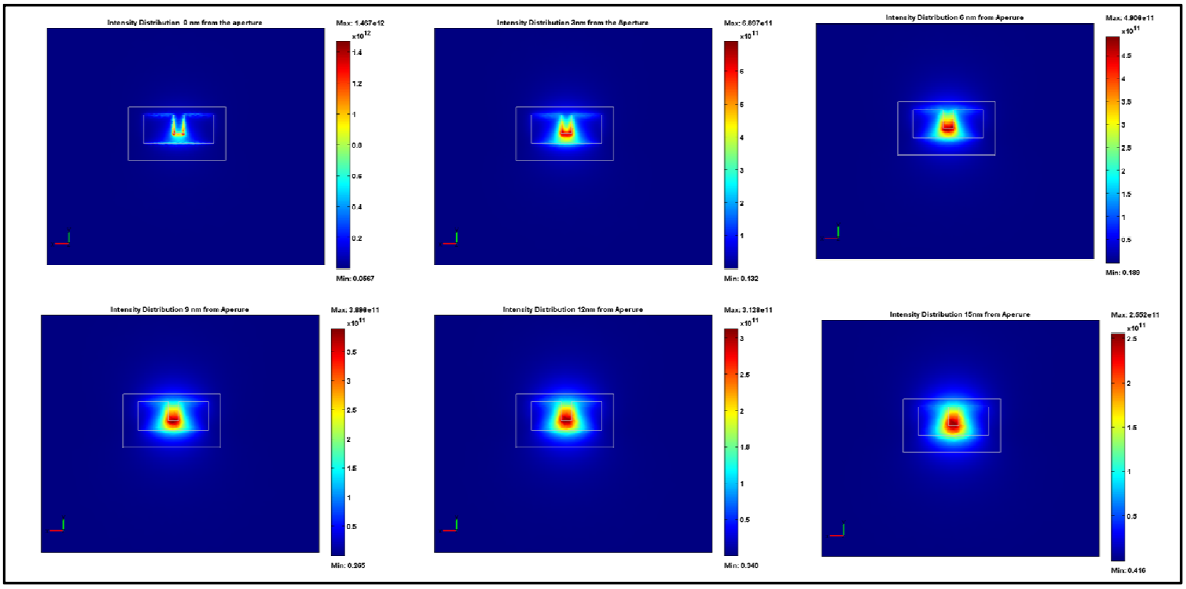

Figure 24: Contour representation of intensity distribution at different distances from the plane of the exit

Figure 25 presents a closer look on how the spot evolves along the z-direction. It is worth mentioning here that the dipole at the aperture exit results in two hot spots which defines a two lobe spot in the very close vicinity of the exit. These two lobes start merging until they form a single spot 6-8nm away from the exit of the aperture.

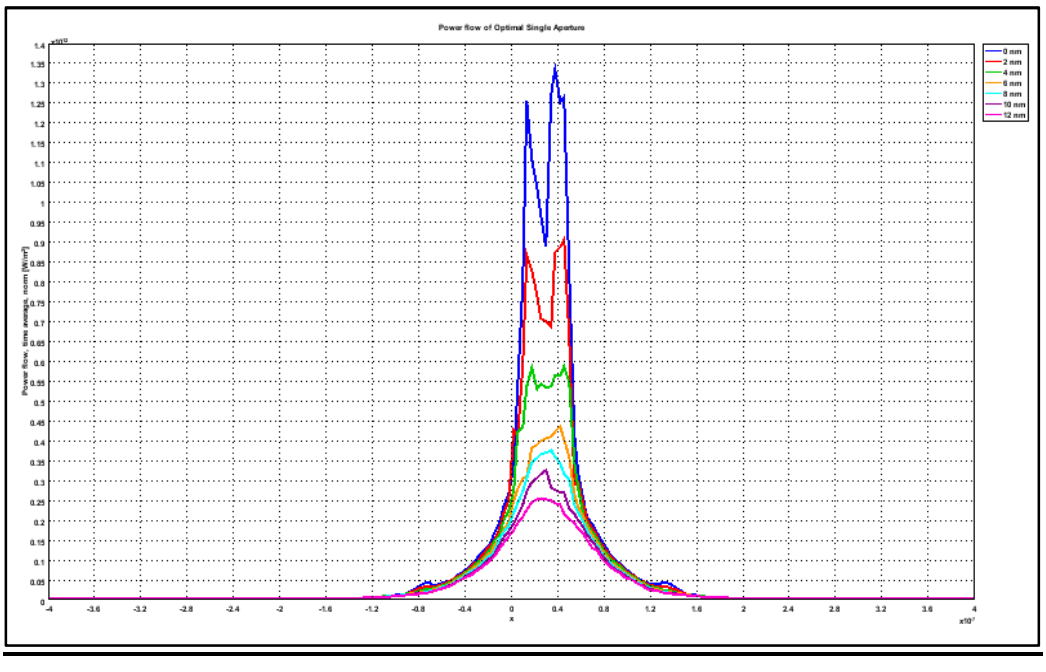

Figure 25: Spatial distribution of intensity at different distances from the plane of the exit

Figure 26 shows the FWHM of the focused spot 8nm away from the exit of the aperture in free space.

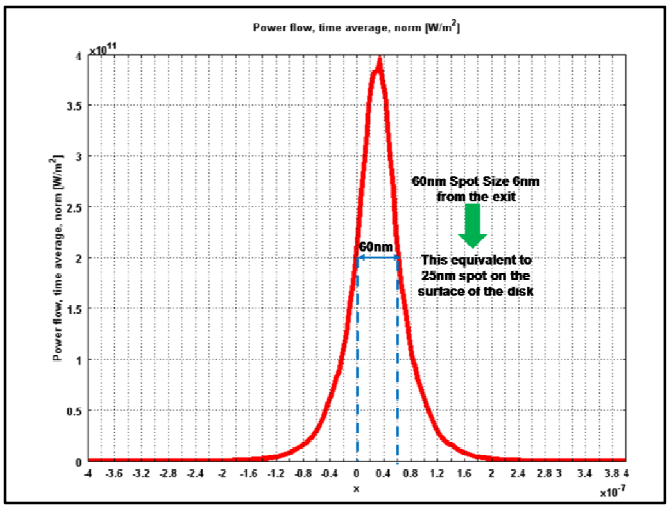

Figure 26: FWHM of focused spot

Taking into consideration the magnetic medium, the electromagnetic waves couple into the magnetic disk described in chapter 3 will result in a spot size smaller than that in free space (Figure 27). However, significant portion of the power will be lost. An intuitive explanation to this can be obtained by considering the boundary conditions between free space and the magnetic medium. Since permittivity of magnetic medium is extremely high at optical frequencies compared to air, the normal component of the light which forms the majority of the propagation fields in discontinuous at this boundary as a function of the air to medium permittivity. Even though the tangential components are continuous, but they do not account as significant to the near-field distribution.

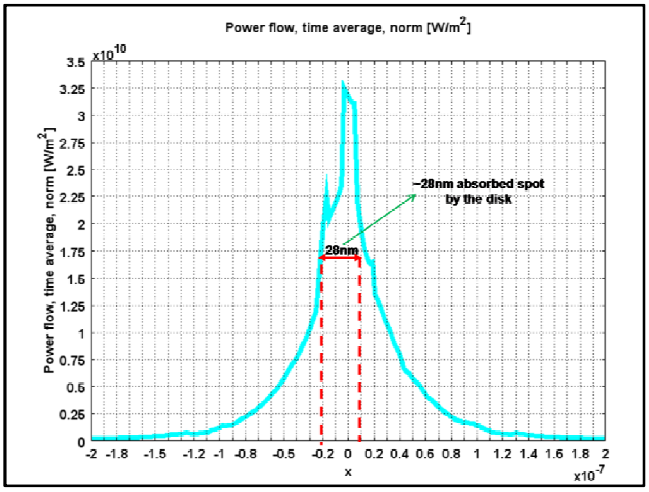

Figure 27: FWHM of absorbed power on the magnetic disk

Another important influencing factor must be considered, namely, the incident light's polarization state. Due to changes in boundary condition, the light beam generates depolarization, and components different from the incident light's polarization direction appear.

The light intensity distribution of the C aperture in the measure plane appears as some characteristics in the near-field zone. The incident beam polarized linearly in the *y* direction generates *Ex* and *Ez* components after going through the nanoaperture due to the depolarization effect (Figure 28).

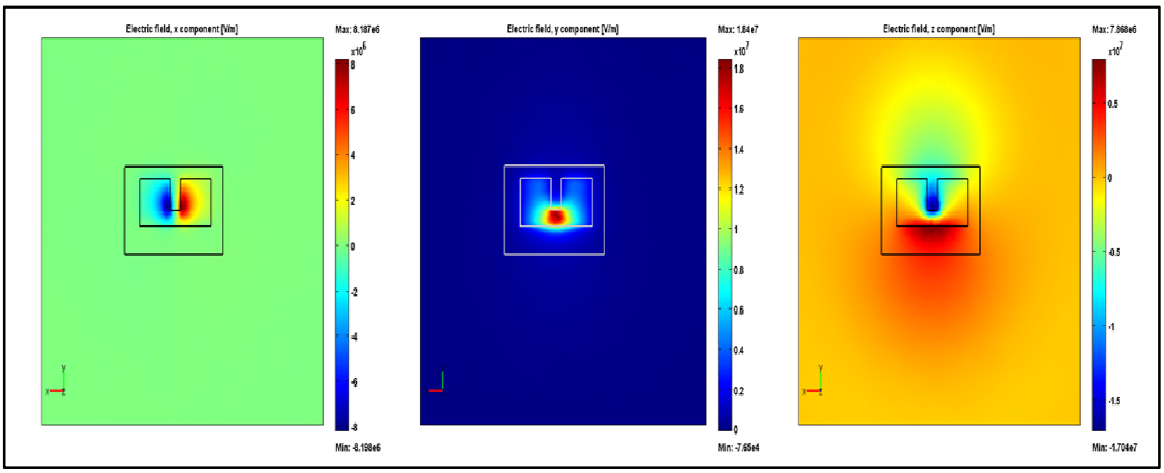

Figure 28: Electric Field Components in the Near field zone

From the calculation results, it can be concluded as follows. Ex^2 distributes mainly on the two ends of the C aperture perpendicular to the x direction and appears as two symmetrical spots. It elongates the spot in the y direction. Ez^2 presents the field enhancement property on the edges. The Ez component is along the incident direction, which indicates that the electromagnetic wave is propagated along the metallic film's surface. Ez^2 generates on the edges perpendicular to the x direction in the C aperture and forms two separate abnormal spots in the cross section. But one is relatively weak. Ez^2 has a great effect on the total field distribution. Ey^2, corresponding to the incident light's polarization direction, is comparable with Ex^2, but both are smaller than Ez^2 in quantity. In the measure plane, the total field Et^2 is enhanced by the evanescent field in the z direction.

This behavior also corresponds to the boundary conditions of the aperture: At the two edges parallel to the incident-light polarization, the electric field is forced to be zero. Therefore the fields are pushed away from these two edges; for the two edges perpendicular to the polarization, the electric field is not constrained. Instead, the electric field of the incident light drives the motion of the charges in the metal resulting charges of opposite signs accumulated along these two edges. These charges generate strong local fields, which lead to the two hot spots at a very small distance z.

3. Aperture Fabrication

A. Metal Deposition:

Thin film deposition is a process applied in the semiconductor industry to grow electronic materials. The deposition process can be broadly classified into physical vapor deposition (PVD) and chemical vapor deposition (CVD). In CVD, the film growth takes place at

high temperatures, leading to the formation of corrosive gaseous products, and it may leave impurities in the film. The PVD process can be carried out at lower deposition temperatures and without corrosive products, but deposition rates are lower and it leaves residual compressive stress in the film. Electron beam physical vapor deposition, however, yields a high deposition rate from 0.1 μm / min to 100 μm / min at relatively low substrate temperature with very high material utilization efficiency.

In an EBPVD system, the deposition chamber is evacuated to a pressure of 10^{-4} Torr. The material to be evaporated is in the form of ingots. Electron beams can be generated by thermionic emission, field emission or the anodic arc method. The generated electron beam is accelerated to a high kinetic energy and focused towards the ingot. When the accelerating voltage is between 20 kV – 25 kV and the beam current is a few amperes, 85% of the kinetic energy of the electrons is converted into thermal energy as the beam bombards the surface of the ingot. The surface temperature of the ingot increases resulting in the formation of a liquid melt. Although some of incident electron energy is lost in the excitation of X-rays and secondary emission, the liquid ingot material evaporates under vacuum. The ingot itself is enclosed in a copper crucible, which is cooled by water circulation.

PURPOSE/ REAL DEPOSIT	E-BEAM EVAPORATION CONDITION	DEKTEK (PROFILERMETER) STEP HEIGHT (CROSS SAMPLE)	AFM MEASUREMENT STEP HEIGHT (CROSS SAMPLE)	AFM MEASUREMENT ROUGHNESS (GLASS SAMPLE)
150NM/ 125NM	BP : 2.92 X 10(-6) POWER : 16.6 ~ 17.2% RATE : 3.8 ~ 4.1 A/SEC	1. 1615.0 A 2. 1579.5A 3. 1590.5A (1+2+3)/3=1595 A	174.72nm	Figure 24-1
125NM/ 100NM	BP : 5.63 X 10(-6) POWER : 16.3 ~ 16.7% RATE : 3.9 ~ 4.3 A/SEC	1. 1374.2 A 2. 1362.3 A 3. 1551.6 A (1+2+3)/3=1429.4 A	147.01 nm	Figure 24-2
100NM/ 75NM	BP: 5.34 X 10(-6) POWER: 16.4 ~ 16.9 % RATE: 3.8 ~4.1 A/SEC	1. 993.5 A 2. 980.1 A 3. 987.9 A (1+2+3)/3=987.17 A	107.05 nm	Figure 24-3

The table above shows projected thickness to be deposited and the measured values using an AFM and profiler meter. Figure 29 confirms the step height measurements using AFM and also presents the surface roughness in terms of mean average.

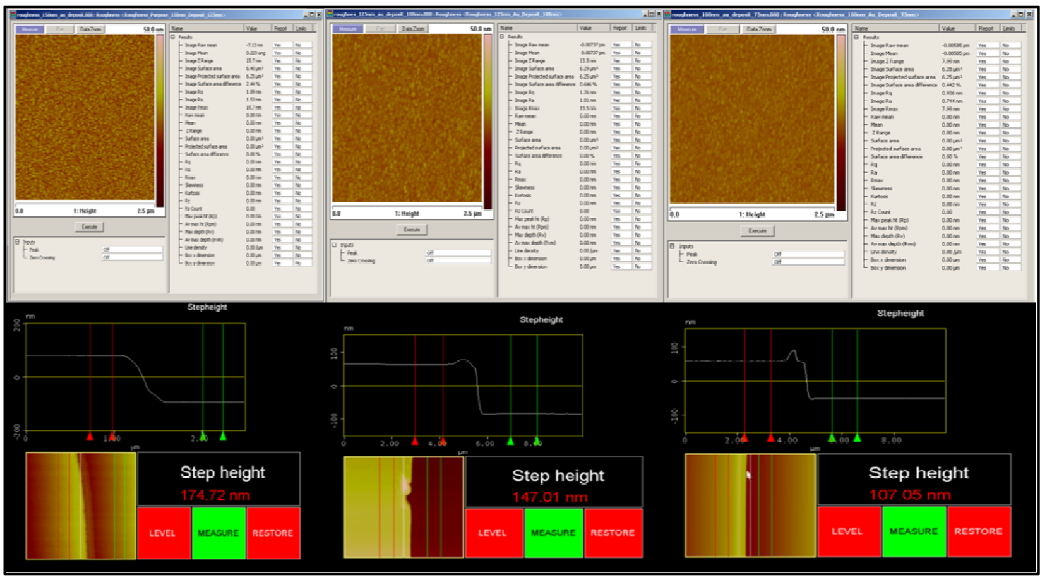

Figure 29: Gold roughness and step height measurements

B. FIB Fabrication

i. Focused ion beam milling theory

Focused ion beam (FIB), is a scientific instrument that resembles a scanning electron microscope. However, whereas the SEM uses a focused beam of electrons to image the sample in the chamber, a FIB instead uses a focused beam of gallium ions. The ability to mill, image, and deposit material using a FIB instrument critically depends on the nature of the interactions between ion beam and solid structure [63]. Figure 30 shows a schematic diagram illustrating some of the possible ion beam and target material interactions that can take place during ion bombardment of a solid. Milling takes place as a result of physical sputtering of the target. An understanding of sputtering requires investigation of the interaction between an ion beam and the target. When the high-

energy gallium ions strike the sample, they will sputter atoms from the surface as the result of a series of elastic collisions where momentum is transferred from the incident ions to the target atoms within a collision cascade region. Interactions between the incident ion and the solid occur due to the initial kinetic energy of the ion. Gallium atoms will also be implanted into the top few nanometers of the surface, and the surface will be made amorphous. If the ion is not backscattered out of the target surface, the ion will eventually come to rest, implanted within the target at some depth below the surface. In general, the number of secondary electrons generated per incident ion is ~ 1 and is 10-1000x greater than the number of secondary ions [64]. Because of the sputtering capability, the FIB is used to modify or machine materials at the nanoscale.

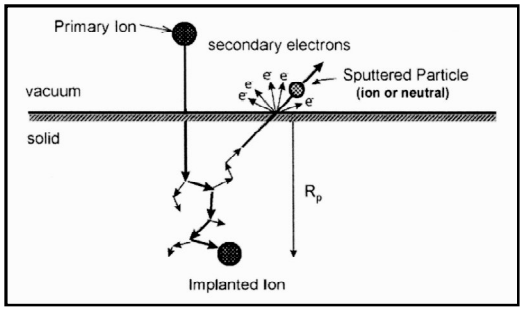

Figure 30: Schematic diagram of the sputtering process and ion-solid interactions.

The sputtering yield is defined as the number of expelled particles per incident gallium ion. Sputtering can be investigated as phenomena based on statistical surface erosion on an atomic scale. Sputtered particles generally have energy in a range between 2 and 5 eV. The emission of sputtered particles depends on the beam incident angle and generally follows a cosine distribution for normal ion bombardment. Back sputtering occurs when an incident ion is scattered either directly, or after a series of multiple collisions, out of the target. This effect is similar to backscattering in electron-solid interactions. The back sputtering yield of the incident ion increases with angle of incidence. Generally, materials

with a higher sputtering yield have a correspondingly higher back sputtering yield of incident ions.

Focused ion beam (FIB) systems utilize a gallium liquid metal ion source (LMIS) and two electrostatic lens focusing column to bombard the sample surface with an high energy focused ion beam used for precision material etching, deposition, and high-resolution imaging. FEI Company's flexible, reliable, and safe gas injection system (GIS) provides the capability of enhanced metal and dielectric removal, as well as metal and insulating material deposition that allow fast turnaround circuit modification. The Magnum column operates at a working distance of 18 mm with an extractor voltage range of 10–12 kV, with 12 kV maximum.

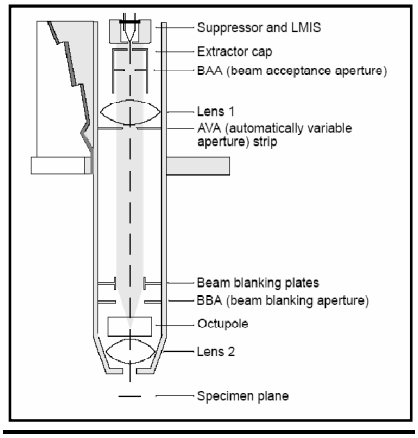

Figure 31: Magnum Column Alignment Diagram.

Gallium liquid metal ion sources are the most widely used ion source, because Gallium has a low melting temperature (T_m=30°C) and remain liquid at the room temperature conditions for a long period of time, a low vapor pressure (10^{-8} Torr), and high surface tension at its melting point. The source does not require continual heating to

achieve and maintain the ion emission.. In addition, the combination of tungsten wire substrate (emitter) and Gallium is very stable, and the reaction rate at the melting point of the Ga liquid is very low. Since Ga only reacts slightly with W at its melting point, this leads to good wetting of W substrate and good flow of the liquid along the emitter (the reason that it works so well is that Ga reduces native tungsten oxide to Ga_2O_3 which is volatile when heating the source). The sources provide long life (1000 or more hours) in a modest vacuum (10^{-5}-10^{-7} torr) and reliable operation in electrostatic focusing columns.

Magnum ion column alignment procedures consist of mechanical and electronic beam manipulations to correct asymmetries. These corrections are made to give the best beam performance over a full range of operations.

Because ions are so much heavier than electrons, the use of magnetic lenses to focus ion beams from LMIS is not as feasible as for electrons. Instead ion optics requires electrostatic lenses. As shown in Figure 31, the complete optics section of the FIB system consists of a condenser lens (lens 1), an objective lens (lens 2), steering quadrupole, deflecting octupole system, and selectable variable sized apertures. The ion beam emitting from the LMIS is focused and condensed by lens 1, and limited by an automated variable aperture (AVA). The AVA is a beam defining aperture system consisting of a molybdenum aperture strip with different size of apertures in order to vary the beam current incident on the sample. The condenser and objective lenses are very compact and their focusing properties are independent of the mass. The optics are designed in such a way that the current from the source was reduced from about 2 microamperes to about 20 nA by the biggest aperture located on the aperture strip. When varying the beam current, a corresponding aperture on the aperture strip is selected and mechanically moved the aperture hole into pre-determined position of the assembly. The beam current is thus increased with increasing the aperture

size. Table 1 shows the relationship between the aperture size, beam current, ion spot size, and Ga^+ ion flux in the FEI XP 800.

Aperture (nm)	Beam Current (pA)	Beam Diameter (nm)	Current Density (A/cm^2)	Ion Flux (Ga^+/s_cm^2).10^{19}
8	1	6	3.5	2.3
19.6	10	10	12.7	12.7
35.2	30	12.5	24.5	15.9
40.1	50	15	28.3	18.4
55.3	100	20	31.8	20.7
84.4	300	25	61.1	39.4
112.8	500	30	70.8	46
150.2	1000	35	104	67.6
245.2	3000	60	106	69
309.4	5000	100	63.7	41.4
357.3	7000	150	39.6	25.8
1000.1	20000	600	7	4.6

ii. Pattern calculation

The FEI XP 800 has a beam control system for FIB sputtering and deposition with a digital scan moving the beam pixel by pixel. By design, the beam size of an FIB is in the order of 10nm for the smallest reachable current to hundreds of nanometers in diameter. In order to mill a feature larger than the beam sizes, or make patterns on an area, the fine ion beam must be scanned over the desired area. The pattern is an assemblage of discrete steps taken by the beam as it moves within the defined raster area. The operator has control of the size of the beam stepping and the number of steps it takes between contiguous locations by inputting pixel dwell time and pixel overlap, respectively, in a material file with an extension of ".mtr" that is used by the FIB software control system. The step size s is relating to the overlap OL and beam diameter d by the following equation: $s = d\,(1\text{-}100\boldsymbol{OL})$.

In practice, the patterned feature size is always larger than the nominal beam diameter. For example, the minimum feature size is about 30 nm in Si when 1 pA beam current is used in our FEI XP 800 FIB system with a beam spot size of less than 10 nm. Overlaps between features would occur if the –100 OL were used. This has to be taken into account when defining the overlap parameters. The FEI software on FEI XP 800 FIB has the ability to calculate and store the material files, which are based on the sputtering yield for a particular material. After the definition of a pattern area for milling or deposition, the FIB software calculates the milling time for a given beam current, a given material depth and a given material sputtering rate.

This calculation is based on the statement that the accurate material sputtering rate is known. This might be true for most commonly used materials like Si. However, the material sputtering rates are not known for all materials and will need to be defined experimentally or calculated. It is possible to override this calculation, by entering a milling time directly into the software interface or by stopping the milling manually during milling process. In this work, the thin film magnetic materials were milled with different materials files with same OL and dwell times but different total deposition dose. But if the milling time was controlled manually, the sputtering rates used were irrelevant.

iii- Fabricated Apertures

A Metal film (70-120nm thick), usually aluminum, gold, or silver, is firstly deposited on the glass window of a laser diode via the electron beam vapor deposition described previously. Secondly, apertures are fabricated using the high resolution focused ion beam (FIB) FEI XP 800 system, described above, with 30 keV Ga^+, 1-3 pA beam current, and Gallium ion flux of 1.3×10^{19} $Ga^+/s\cdot cm^2$ (Figure 32).

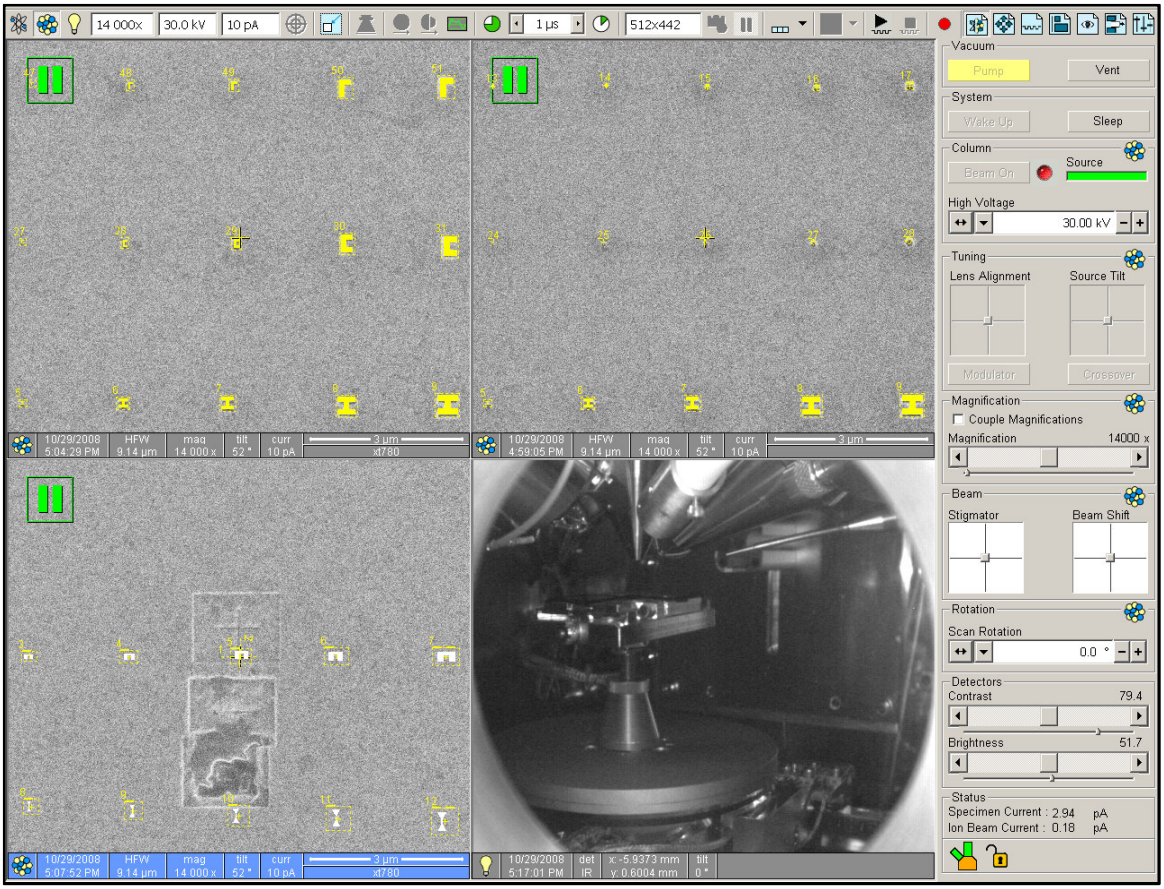

Figure 32: Snapshot of FIB Operation Panel

The theoretical minimum beam spot size for this case is 8 nm, and the real spot size is estimated to be 10 nm. The pixel-by-pixel milling method was used to raster patterns with 10 nm step size. The milling was subsequent to the numerically estimated patterns through a personalized developed software codes based on Matlab to minimize the secondary sputtering and redeposition effects. The milling rate was measured with an atomic force microscopy (AFM), and aperture depth was controlled by total milling time that ranged between 5 and 15 seconds, depending on the shape and deepness. Ion images of the apertures are shown in Figure 33.

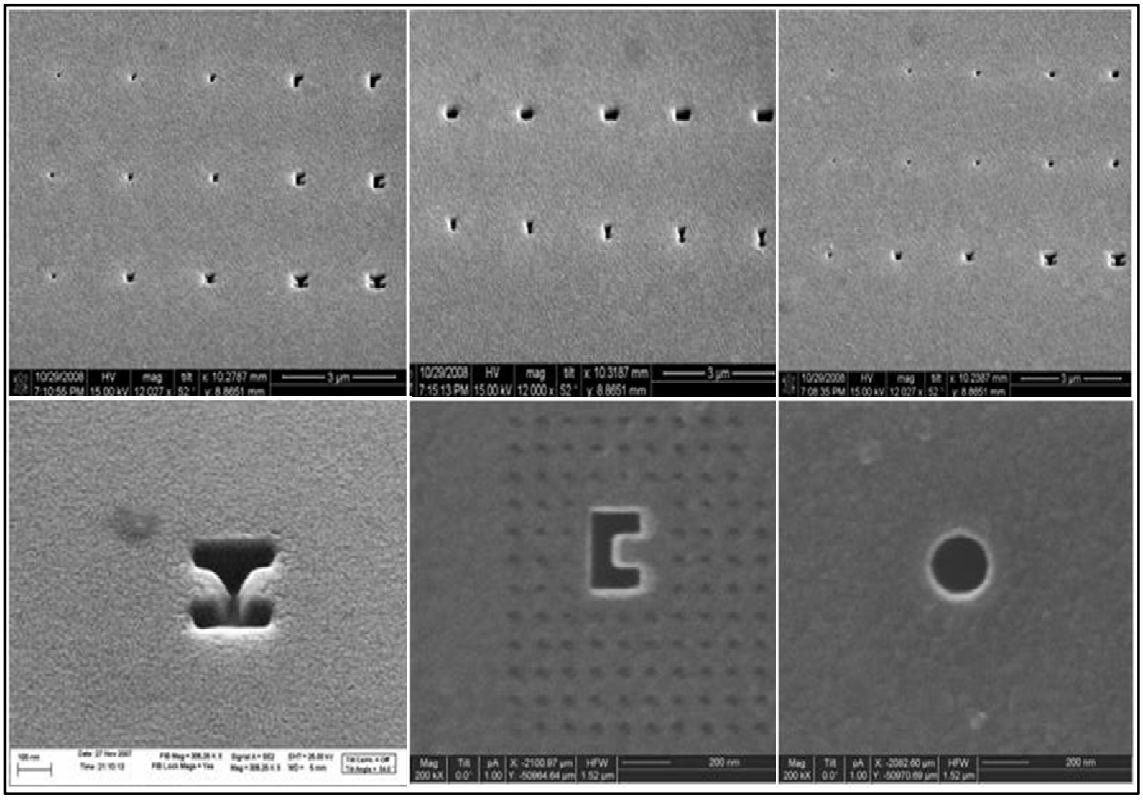

Figure 33: Ion images of the fabricated apertures

4. Characterization:

Near-field Scanning Optical Microscopy (NSOM) is an imaging technique that combines Scanning Probe Microscopy (SPM) with optical microscopy. While the purpose of an NSOM system is to obtain optical data, it is the precise position detection and control capability of an SPM that allows the probe to scan a defined region of the sample surface and closely track the topography. By maintaining a sufficiently small probe-sample distance regime (<10nm), near-field optical imaging is possible. For each scan, a topographic image is also produced, allowing for direct comparison and analysis of topographical and optical sample characteristics.

Historically, optical microscopy has been limited by one factor: the diffraction limit of light. This limitation has frustrated attempts to image features approximately the size of a

wavelength of light. As far back as the 1930's, a theoretical solution to this problem was suggested [65] however, it was not until the early 1980's that the electronic control and feedback capability existed to realize this solution [66]. In theory, to overcome the limitations imposed by diffraction, the aperture exposing the sample must be kept very small—on the order of 50 nm—and the aperture must be kept close to the sample surface—generally less than 10 nm. The development of scanning probe microscopes has provided the technology needed to maintain such close tip-sample spacing while a tip is being scanned over a sample. The laser light source in an NSOM system is launched into a straight optical fiber. The end of the fiber tapers to a diameter of 50 nm. The fiber is then coated with aluminum approximately 100 nm thick. The fiber becomes a "light funnel" directing light onto the sample. Photodetectors are placed behind the sample (transmission mode) or beside the tip (reflection mode) to collect light emitted from the sample.

The NSOM instrument is a unique combination of a scanning probe microscope and an ultra-high resolution optical microscope. Generally both modes—topography and optical—are used together. Topography mode alone may be used initially in order to identify potential features of interest on the sample surface. The NSOM, used as an optical microscope, can be operated either in tip illumination or tip collection mode. Tip illumination mode is perhaps the most common NSOM mode. The tip is used as a "light funnel" or "nano-flashlight" to illuminate the sample in a precise, controlled manner. This mode can be further subdivided into reflection, transmission, and lithography modes. Reflection mode gathers light that has been reflected from the sample. It is used for opaque samples, such as semiconductor surfaces. Reflection mode might suffer from

image artifacts created by tip shadowing on the sample surface, as the tip creates a wave guide with the surface, through which the light must pass before being collected. Transmission mode is more commonly used and is more efficient. The collector is placed behind the sample and collects a majority of the light as it passes through the sample. The drawback to this mode is that it is limited to the use of thin, transparent samples. In tip collection mode, the sample is the light source, and the tip acts as a way to collect this light. This method is best for samples such as waveguides and laser diodes.

SNOM makes use of a technique where an aluminum-coated tapered optical fiber probe is placed at a distance (from an imaged sample) which is a fraction of the wavelength of the emitted light. SNOM probes are used to collect emitted radiation in the near-field to examine the evanescent fields of waveguides (emission profiles) or as an illumination source to profile surfaces, excite local photoconductivity in the semiconductor, locally heat active devices, etc… Simultaneous shear-force measurements provide an independent measure of the surface topography to maintain a fixed proximity (5-10 nm) between tip and sample. The fiber tip is mounted on a quartz tuning-fork and dithered at its resonant frequency with a small piezo element where any tip-surface interaction quenches the resonance, providing a height measurement with nanometer resolution. The shear-force topography is exploited to correlate various optical profiles with the physical device structure. In this particular case, the imaged sample is the emitting edge of a diode laser.

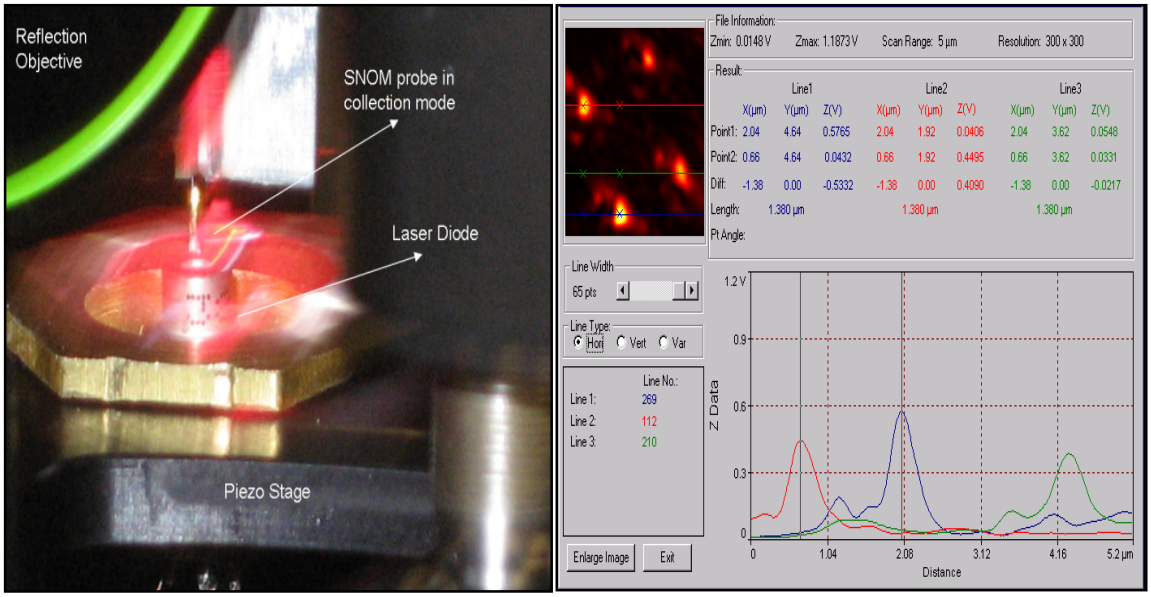

Figure 34: (left) Experimental setup showing SNOM probe in collection mode scanning the emitting surface of the laser diode (right) point analysis for (2) C apertures (lines 1and 2) and "I" shape apertures.

In our experiment, the laser diode is mounted face up on a piezo actuated flexure stage and scanned in the xy-plane underneath the probe tip. Collected light is coupled into the optical fiber and transmitted to be analyzed with a photo-multiplier tube (PMT) detector (Figure 34). Distortions at the edge of apertures due to the deep etch, possible back deposition of aluminum during fabrication, aperture size which is comparable to the tip size, and melting of the tip aluminum coat due to excessive power exposure impinged the quality of the topography image, yet had slight effect on the optical image. It was recorded for an input power of 80µW the output was measured to be 0.53µW. So for 250mW of input power, assuming linear dependence, 1.65mW will be focused onto an approximately 30.10nm radius. This estimation, if validated, counterpart our simulated results where we estimated ~ 2.19mW of power for 250mW input power (Figure 35) and 0.712µW for 80µW input (Figure 36).

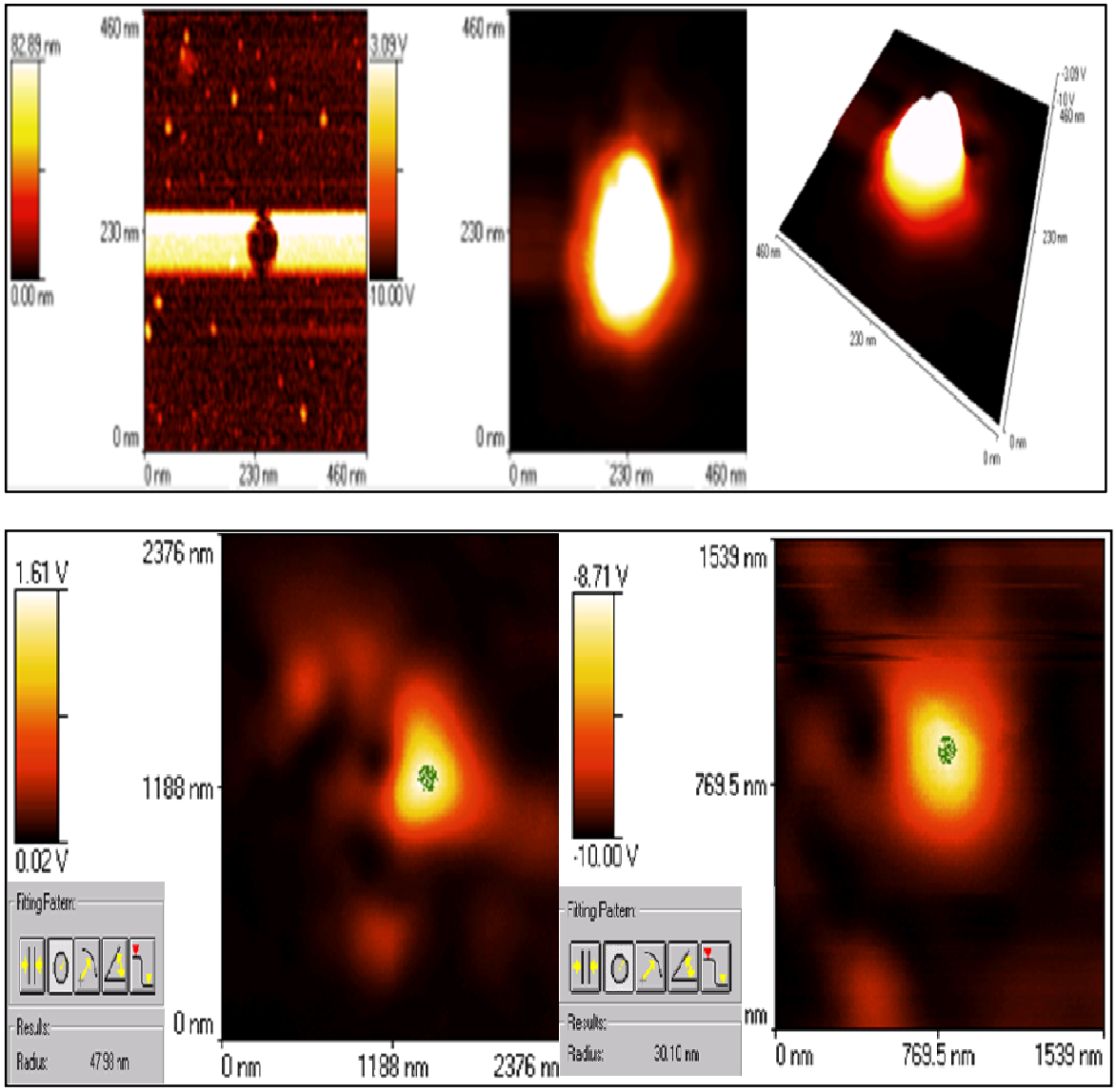

Figure 35:(first row) topographical representation of 100nm circular aperture and 2D and 3D representation of the light collected, (second row) Spot size of "C" aperture with 40nm ridge size and (right) "C" with 30nm.

These measurements were recorded for the "C" shape aperture after it was extrapolated from batch scanning (Figure 34-2) that apertures with "C" shape claimed superiority over other shaped aperture in terms of locality and throughput.

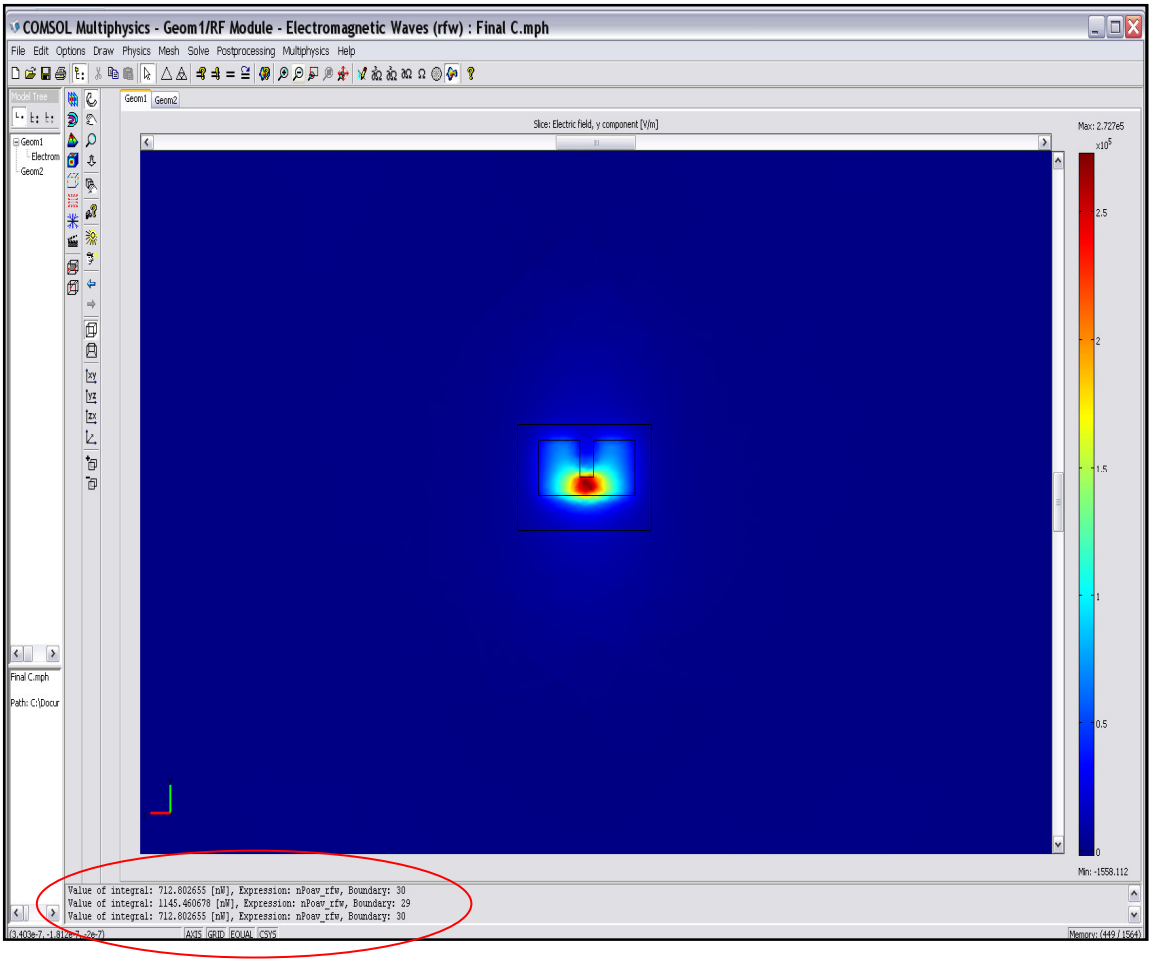

Figure 36: Intensity Distribution for optimal ridge illuminated by 80 μW

A. Analysis of Light Losses in NSOM Tips

The transmission of light through an NSOM tip is a lossy endeavor. The light in the fiber begins to penetrate the fiber cladding once the core size shrinks (due to the tapered region) to values below a few wavelengths. The light that penetrates the cladding is reflected by the metal coating and thus never reaches the aperture. Due to this process, only 1 out of 10^4 - 10^5 photons is transmitted through the tip. Increasing the power input to the fiber to compensate for this loss can only be done up to a few mW. If the power input into the fiber is greater than a few mW, heating at the taper becomes inadequately severe and damages the metal/glass interface so that the coating becomes compromised

or lost completely in the case of melting (Figure 37) [67 , 68 69, 70, 71]. Therefore, the upper limit to the continuous-wave transmitted power of an NSOM tip is of the order of 100 nW (10^{12} photons/s). In practice, our 100 nm apertures transmit the power of the order of 10 nW (10^{11} photons/s) for an input power of a few mW. This setback limited our collection scanning capabilities at high input powers and restricted the experiments to the lower input power, which NSOM probes can tolerate. To estimate the throughput at the highest input power, we assumed (extrapolated) the same linear dependence of the throughput measured at the lowest input power. The overheating problem discussed above provides an additional motivation of the study of thermal conduction in materials used for HAMR and NSOM tips.

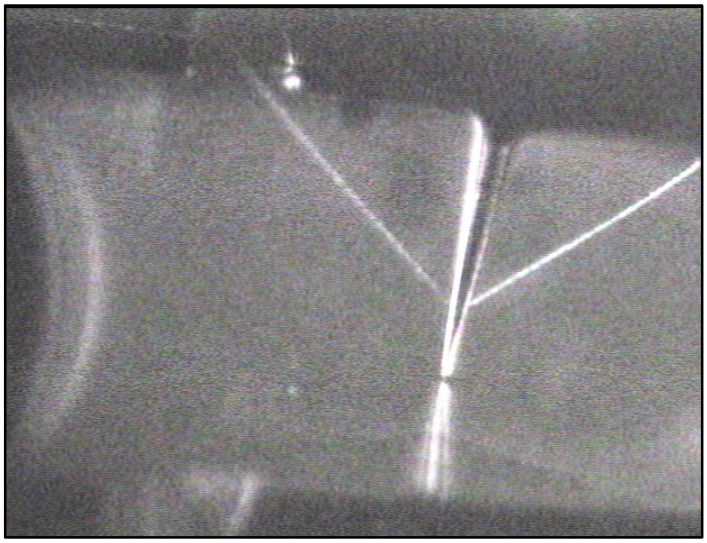

Figure 37: Snapshot of melting NSOM probe.

5. Conclusion

Near-field optics is exploited to design a transducer capable of delivering more than 1 m-W into a spot in the order of 30 nm in diameter. This is fundamental requirement for these transducers to be utilized in HAMR systems. Thorough background theoretical

explanation has been disclosed in this thesis to set a strong platform for the numerical analysis of these transducers. FEM has been used to simulate nanoscale apertures embedded in metal films. These apertures showed strong dependence on the metal used, its geometrical effects, wavelength and polarization of light. Ridge shaped aperture outperformed other shapes apertures in terms of near-field intensity distribution and transmission efficiency. Ridge apertures were then fabricated on Gold coated laser diodes using FIB, and then characterized using NSOM system. Simulation and experimental results complemented more than 75%. 2.19mW was estimated numerically for a spot size of 30nm, our measurement shows 1.65mW.

CHAPTER III: Thermal Behavior Induced by LASER Heating in HAMR Systems

In magnetic recording systems, an emphasis has been laid on achieving higher areal density, which has forced certain characteristic lengths of the devices to nanometer scale. In HAMR nanoscale energy management becomes important in the operation and reliability; nanometers sized hot spot is formed when a laser pulse is incident on the disk, which results in the lowering of the magnetic coercivity of the media. The energy transport from the hot spot to the lubricant film influences its adsorption/desorption and replenishment. Also, the subsequent transport of energy to the air above the disk could influence the air bearing stability. Therefore, a fundamental understanding of the subcontinuum heat conduction phenomena is critical [72,73,74]. It is known that the surface morphology of the lubricant film affects the performance of read and write of the head. Lubricant depletion due to disk rotation and fluctuation behavior by the head-lubricant interaction, when the head flying height need to be decreased for higher aerial densities, has been extensively studied. However, hardly any studies have dealt with the thermo-optic behavior of the loop which includes lubricant depletion induced by laser heating which theoretically can be related to evaporation followed by thermo-capillary stress and thermo-viscosity.

Absorption of the electromagnetic waves by the medium will induce heat due to electron-phonon interplay. Thus the thermal behavior of energy carriers, particularly, electrons and phonons should be considered during the numerical analysis for predicting the transient sub-continuum thermal transport in the magnetic media [75,76,77,78,79,80].

Our strategy is to evaluate lubricant depletion numerically then experimentally arrive at the most favorable head/media-lubricant configuration for a reliable HAMR system.

1. LASER light: Propagation and Absorption

For laser light to cause a tough effect on a material, it should be absorbed. The absorption process can be thought of as a secondary source of energy inside the material. It is the secondary source that will determine the dynamics of the irradiated material rather than the beam generated by the laser device; this will be our approach to model a HAMR system based on laser assisted writing. Following, we will summarize optical behavior of metal relevant to electromagnetic wave absorption:

The laser beam will be sufficiently approximated as an electric field of a linearly polarized plane wave propagating in a homogenous medium. Before absorption occurs, the electric field can be represented as follows:

$$\vec{E} = \vec{E}_0 \exp\left[i\left(\frac{2.\pi.z}{\lambda} - w.t \right) \right]$$

where z is the coordinate along the direction of propagation, w is the angular frequency, and λ is the wavelength. Angular frequency and the wavelength of light can then be related by the phase velocity c/n, c being the speed of light and n being the refractive index of the medium:

$$\lambda = \left(c \middle/ n \right).\left(2.\pi \middle/ w \right)$$

Analogous expressions can also hold for the magnetic field H, and can be related by:

$$H_0 = E_0 n_1 \varepsilon_0 c$$

where ε_0 is the dielectric constant in vacuum. Generally, the electric and the magnetic fields carry equivalent amount of energy. However the force exerted by the

electromagnetic wave on an electron is contributed to the electric field by a factor of *v/c* compared to that by the magnetic field component (*v* is the electron velocity), and hence it is usually neglected:

$$F = -e\left[\vec{E} + (n_1 / c)(\vec{v} \times \vec{H})\right]$$

Irradiance[2] is the energy flux per unit area of the wave and can be formulated as:

$$I = \left|\vec{E} \times \vec{H}\right| = n_1 . \varepsilon_0 . c . E_0^2$$

In quantum mechanics, a wave with irradiance I corresponds to a flux of $I / \hbar . w$ of photons carrying $\hbar . w$ of energy. Photon fluxes in intense laser beam are significantly vast, making classical concepts adequate to describe beam interaction with materials.

Commonly, the lateral beam distribution is a cylindrically symmetric Gaussian where irradiance is maximum near the optical axis and exponentially decaying laterally along its radius *w*; It rigorously apply to TEM_{00} laser mode, but it will shall be used as an approximation throughout:

$$I(r) = I_0 \exp(-r^2 / w^2)$$

where I_0 is the irradiance along the symmetry axis . The total power of the beam can then be formulated as:

$$P = \pi . w^2 . I_0$$

The propagation of the electric field over a distance z within the lossy medium is attenuated by a factor $\exp(w n_2 z / c)$ where n_2 is the extinction coefficient of the absorbing media equivalent to the complex refractive index. The attenuation of the electric field indicates that some of the light energy is absorbed. The absorption coefficient can then be derived as (inverse of α is the absorption length):

$$\alpha = -(1/I)\left(dI/dz\right) = 2.w.n_2/c = 4.\pi.n_2/\lambda$$

Reflectance relates the ratio of reflected to incident irradiance and is given by Fresnel expression:

$$R = \left|(n-1)/(n+1)\right|^2$$

Reflectance and the absorption coefficient will both contribute to determine the amount of power absorbed by the material. For a normal incident beam of irradiance I, the power density deposited at depth z is the expression that represents the secondary source in the material:

$$J_a(z) = I(1-R).\alpha\{1-\exp\left[-\int_0^z \alpha(z')dz'\right]\}$$

Since the refractive index of the treated material is of great importance to model the absorbed light, it is commendable to connect it with the properties of the medium. This is provided by Maxwell's equations:

For metals, the optical response is dominated by the conduction electrons. Only the electrons close to the Fermi level, contribute to the optical properties. These electrons interact with the lattice merely by collisions. The dielectric function of free-electron metal can be expressed by:

$$\varepsilon = 1 + w_p^2(-\tau_e^2 + i\tau_e/w)(1 + w^2\tau_e^2)$$

where

$$w_p = \sqrt{N_e e^2/m_e\varepsilon_0}$$

is the electron plasma frequency, N_e is the number of bound electrons, τ_e is the collision time .It is stated that for $w = w_p$: both ε_1 and n_1 vanish (in the ultraviolet regime for most metals), for $w < w_p$ reflectance (R) and absorption (α) are high, and finally low R and α for $w > w_p$. Implementing Drude expression, the optical properties of a free-electron metal for $w < w_p$ are related to its DC conductivity σ_0:

$$\sigma_0 = N_e e^2 \tau_e / m_e = w_p^2 \varepsilon_0 \tau_e$$

For most metals in the visible regions, we have $n_1 \cong w_p / 2w^2 \tau_e \cong 0$ and $n2 \cong w_p / w$ which yields reflectance and absorption of:

$$1 - R \cong 2 / w_p \tau_e = 2\varepsilon_0 w_p / \sigma_0$$

and

$$\alpha \cong 2w_p / c$$

2. Heating by Laser Light

The main advantage for laser heating is its localization, both in space and time. By matching laser beam properties to those of the optical and thermal properties of the material, precise heating can be selected to ensemble the needs of the process. Laser heating can be restricted to surfaces by matching a large absorption coefficient material and a short interaction time. Similarly, rapid cooling can be accomplished using short pulses or employing a moving beam and/or surface. The main effect of laser irradiation in our application is to raise the temperature of the material which then obeys its own thermodynamics laws, so calculating the temperature distribution in the solid has to be investigated.

The first product of absorbed laser light by the material is absolutely not heat, however it is the excitation energy of bound electrons, kinetic energy for free electrons, and possibly excess phonons.

The degradation of the localized primary excitation energy into uniform heat evolves over three steps:

1. Spatial and temporal randomization of the motion of the excited particles, proceeding with the momentum relaxation time of the particles (time is shorter than the pulse time).
2. Energy equi-partition which involves a large number of elementary collisions and intermediate states. Energy transfer mechanisms will also be involved.
3. Heat flow subsequent to laser energy conversion to heat.

The mathematical theory of heat conduction is based on the assumption that the heat flux across a plane in a solid is proportional to the local temperature gradient:

$$\phi(z_0) = -K\left(\frac{dT}{dz}\right)_{z_0}$$

K is the thermal conductivity of the material.

The energy balance of a slab of material bounded between z and z +Δz in terms can be expressed in terms of its volumetric heat capacity c_p/V:

$$\Delta t[\phi(z) - \phi(z + \Delta z)] = \Delta T(c_p / V)\Delta z$$

Δt is the change in temperature brought about by a net heat flux across the boundaries. The heat flow equation in one dimension can be derived then by making Δz arbitrarily small, which then yields:

$$\frac{\partial}{\partial z}\left(\frac{K\partial T}{\partial z}\right)=\left(\frac{c_p}{V}\right).\left(\frac{\partial T}{\partial t}\right)$$

If heat is produced in the material, the power density of the heat source is added to the left hand side of the equation.

To solve the heat flow equation, we assume the power density of heat production to equal that of the absorbed laser light:

$$J_a(z)=I(1-R)\times\alpha\left\{\left(1-\exp\left[-\int_0^z\alpha\left(z'\right)dz'\right]\right)\right.$$

1. All material parameters are constants.
2. Consider a homogenous material in the form of a slab between the planes z = 0 and z = L.
3. Initial temperature is zero- a uniform temperature can be added to the equation.
4. Slab is assumed to be thermally insulated, i.e., no heat flows across the boundaries is allowed: $\partial T/\partial z=0$ for z = 0 and z =L, at all times.
5. The laser beam is incident on the plane z = 0 and taken to be of cylindrical symmetry. The heat flow equation can now be written as :

$$(\partial T/\partial t)=\kappa.\nabla^2T+[\alpha.I_a(x,y,t).V/c_p].e^{-\alpha.z}$$

Where $\kappa=KV/c_P$ is the heat diffusivity and I_a is the unreflected part of the incident irradiance. This will be the equation solved for the model with minor adjustments that include using thermal conductivity instead of diffusivity and accounting for the reflected/unreflected part of the beam by substituting it with a heat line source.

3. Heat Transfer: Theoretical Background

Heat transfer is defined as the movement of energy due to difference in temperature e.g. an object or fluid, is at a different temperature than its surroundings or another body, transfer of thermal energy, also known as heat transfer, occurs in such a way that the body and the surroundings reach thermal equilibrium. It is characterized by the following three mechanisms, all of which our model supports:

A. Conduction: is the transfer of thermal energy from a region of higher temperature to a region of lower temperature through direct molecular communication within a medium or between mediums in direct physical contact without a flow of the material medium. The transfer of energy could be primarily by free electron diffusion as predominant in metals or phonon vibration as predominant in insulators. In other words, heat is transferred by conduction when adjacent atoms vibrate against one another, or as electrons move from atom to atom. Typical for heat conduction is that the heat flux is proportional to the temperature gradient: $\boldsymbol{q = -k\nabla T}$ where $\boldsymbol{k}$ is the thermal conductivity defined as the quantity of heat, $\boldsymbol{Q}$, transmitted in time $\boldsymbol{t}$ through a thickness $\boldsymbol{L}$, in a direction normal to a surface of area $\boldsymbol{A}$, due to a temperature difference $\boldsymbol{\Delta T}$.

B. Convection: It the term used for the heat dissipation from a solid surface to a fluid by circulation or movement of the hot particles in bulk to cooler areas in a material medium, where the heat transfer coefficient $\boldsymbol{h}$ and the temperature difference across a fictive film describe the heat flux: rate of convection between an object and the surrounding fluid. The heat transfer coefficient is not a material property. The heat transfer coefficient $\boldsymbol{h}$ depends upon the geometry, fluid, temperature, velocity, and other characteristics of the system in which convection occurs. Therefore, the heat

transfer coefficient must be derived or found experimentally for every system analyzed. Formulae and correlations are available in many references to calculate heat transfer coefficients for typical configurations and fluids.

C. Radiation: is the transfer of heat through electromagnetic radiation. Hot or cold, all objects radiate energy at a rate equal to their emissivity times the rate at which energy would radiate from them if they were a black body.

4. Solving heat transfer equation

A. Adding heat transfer by conduction:

Heat equation: $\rho.C_p.\frac{\partial T}{\partial t}+\nabla.(-K\nabla T)=Q$

Variables and quantities in this equation:

- T is the temperature,
- ρ is the density,
- C is the *heat capacity*,
- k is the *thermal conductivity*,
- Q is a *heat source* or *heat sink.*

If the thermal conductivity is anisotropic, k becomes the thermal conductivity tensor, k as in

$$k=\begin{bmatrix} k_{xx} & k_{xy} & k_{xz} \\ k_{yx} & k_{yy} & k_{yz} \\ k_{zx} & k_{zy} & k_{zz} \end{bmatrix}$$

Boundary Conditions: two types of boundary conditions can be used, the *Dirichlet* type and the *Neumann* type. Use the *Dirichlet* type to set a temperature on a boundary: $\boldsymbol{T=T_0}$, and use the *Neumann* type to set the heat flux on a boundary: $\boldsymbol{-n.q=q_0}$

where:

- q is the heat flux vector,
- n is the normal vector of the boundary,
- $\boldsymbol{q_0}$ is inward heat flux, normal to the boundary.

B. Adding Convective Cooling to the System:

This section describes the equations and the boundary conditions used for modeling our problem that involve convective heat transfer namely convective cooling or heating whereby a fluid cools a surface by natural or forced convection. We will use a heat transfer coefficient on the convection-cooled surfaces as follows: convection cooling will be modeled by specifying the heat flux on the boundaries that interface with the cooling fluid as being proportional to the temperature difference across a fictive thermal boundary layer. The flux will be described in terms of a heat transfer coefficient, $\boldsymbol{h}$, according to the equation:

$$-n.\left(-K\nabla T\right)=h(T_{\text{inf}}-T)$$

The main difficulty in using heat transfer coefficients is in calculating or specifying the appropriate value of the h coefficient. That coefficient depends on the cooling fluid, the fluid's material properties, and the surface temperature—and, for forced-convection cooling, also on the fluid's flow rate. In addition, the geometrical configuration affects the coefficient.

It is possible to divide convection cooling into four main categories depending on the type of convection conditions (natural or forced) and on the type of geometry (internal or external convection flow). In addition, these four cases can all experience either laminar or turbulent flow conditions, resulting in a total of eight types of convection,

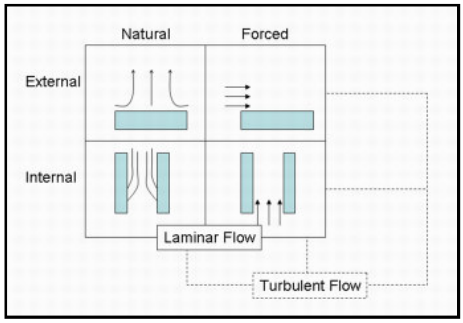

Figure 38 The eight possible categories of convection cooling.

The difference between natural and forced convection is that in the latter case an external force such as a fan creates the flow. In natural convection, buoyancy forces induced by temperature differences and the thermal expansion of the fluid drive the flow.

For each category, various relationships for the heat transfer coefficient have been proposed in the literature. For most cases, the equation describing the ***h*** coefficient varies significantly with the geometrical shape. The coefficient library uses handbook expressions based on the following set of dimensionless numbers:

Nusselt number: measures the enhancement of heat transfer from a surface due to convection compared to just conduction occurring. It is formulated as follows:

$$Nu_L(Re, Pr, Ra) = hk/L$$

- ***Reynolds number***- ratio of inertial to viscous forces in a fluid. Thus, it used to identify laminar or turbulent flow: $\boldsymbol{Re_L = \rho UL/\eta}$
- ***Prandtl number***- dimensionless number approximating the ratio of momentum diffusivity (viscosity) and thermal diffusivity: $\boldsymbol{Pr = \rho C_p/\eta}$
- ***Rayleigh number***- it indicates the presence and strength of convection within a fluid body and is defined as the product of the ***Grashof*** number, which describes the relationship between buoyancy and viscosity within a fluid, and the ***Prandtl***

number, which describes the relationship between momentum diffusivity and thermal diffusivity: $Ra = GrPr = \rho^2 g \beta C_p \Delta T L^3 / k \eta^2$.

where h is the heat transfer coefficient, L gives the characteristic length, ΔT denotes the temperature difference between surface and cooling fluid bulk, and g represents the gravitational constant. The equations for these dimensionless numbers also contain several properties of the cooling fluid: ρ is the density, U gives the bulk velocity, η denotes the viscosity, C_p equals the heat capacity, and β is the thermal expansivity. In the coefficient library, the h coefficient is based on Nusselt number correlations from handbooks and is expressed as a function of the material properties, temperature, flow rate, and geometry. The above numbers can be evaluated for the fluid (air) using the following scalar expressions:

Expression	***Description***	***Value***
rho_air	density : ρ	$1.013e^{5}*28.8e^{-3}/(8.314*T)$
eta_air	Viscosity: η	$6e^{-6}+4e^{-8}*T$
$C_{p_}$air	heat capacity: C_p	$1.1e^{3}$
k_air	Thermal Conductivity: k	10^(-3.723+0.865*log10(abs(T))

The following values will be used to evaluate Prandtl and Reynolds numbers to arrive at the average heat transfer coefficient (***h***). Despite the fact that cooling is not forced in a real circumstances and since the model relies on the relative motion of the laser relative to the disk to demonstrate the spinning effect (disk is not spinning- no cooling effect was observed for natural air flow), forced external air flowing over a heated plate case was

implemented to approximate ***h***, where the air is designated a linear velocity equivalent to that of the disk (25m/s). ***h*** is estimated to be: $h_{average} = \left(\frac{k}{L}\right)\frac{0.928\,\mathrm{Pr}^{0.33}\,\mathrm{Re}^{0.5}}{\left(1+\left(\frac{0.0207}{\mathrm{Pr}}\right)^{0.67}\right)^{0.25}}$

C. Adding Radiative Heat Transfer

Thermal radiation is electromagnetic radiation emitted from the surface of an object which is due to the object's temperature. Radiative heat transfer does not have an impact on the heat equation for the subdomain (bulk of the disk). In this case we can assume heat transfer by conduction in the subdomain, as described in the previous section.

In the laser heating problem, radiation will occur from the heated surface to the ambient. This will add a radiative heat flux term into the boundary condition equation:

$$-n.(-k\nabla T) = q_0 + h(T_{\inf} - T) + \varepsilon\sigma(T_{amb}^4 - T^4)$$

Where

- ε is the surface emissivity estimated at 0.3,
- σ is the Stefan-Boltzmann constant being 5.67*10^-8,
- T is the ambient radiation temperature.

D. Adding Highly Conductive Layers

The highly conductive layer feature is efficient for modeling heat transfer in thin layers without the need to create a fine mesh for them. The material in the thin layer should be a good thermal conductor. This is the case when modeling a thin magnetic layer (100nm) on a glass substrate (0.635mm). Because the layer is very thin and has a high thermal conductivity, we can assume that no variations in temperature and in-plane heat flux exist

along the layer's thickness. Furthermore, the heat flux in the layer's normal direction is modeled as heat sources that are smeared out along the layer thickness.

To describe heat transfer in highly conductive layers, the model uses a variant of the heat equation that describes the in-plane heat flux in the layer:

$$d_s.C_P.\rho_s.\frac{\partial T}{\partial t}+\nabla_t.(-d_s k_s \nabla_t T) = q_{\partial\Omega} - q_\Omega$$

The properties in the equation are:

- ρ_s is the layer density,
- C_p is the layer heat capacity,
- k_s is the layer thermal conductivity tensor,
- d_s is the layer thickness,
- $q_{\partial.\Omega}$ is the heat flux from the boundary into the layer, and
- q_Ω is the heat flux from the layer into the subdomain.

5. Model Definition and Results

A laser beam will be used to locally heat a magnetic thin film deposited on either aluminum or glass substrates. Typically, the disk spins at 5400-7200 rpm with respect to the stationary light source with the capability of the disk to be laterally translated to address specific tracks over the disk. In our model we will be using the relative motion effect and that by spinning the laser at the specified frequencies.

The model treats the localized transient heating caused by a laser beam that moves in circles over a magnetic material coated Al/ or glass disk. The beam's penetration depth, which can be described with an absorption coefficient, depends on the ambient

temperature. So the model will examine the penetration depth and the influences of the disk rotation on the transient temperature distribution.

Initially, the model will consider the laser beam having an infinitesimal width and thus treats it as a line heat source. However, the beam will be assigned a diameter in proceeding work.

The model makes use of the conduction mode to describe the transient heat transfer in the 3D geometry. The transient energy-balance equation for heat conduction will be modeled as:

$$\rho.C_p.\frac{\partial T}{\partial t} + \nabla.(-K\nabla T) = Q$$

where ρ is the density of the treated solid, C_p is the specific heat capacity, K is the thermal conductivity, and Q is the heat source.

Boundaries are assumed to be isolating and the model use the weak form, subdomain application to model the laser penetration. The equation describing the penetration is based on a law called the *Beer-Lambert law* which states that in any medium that is absorbing light, the decrease in intensity I per unit length z is proportional to the instantaneous value of I. In mathematical form this is: $\frac{\partial I}{\partial x} = -K_{abs} I$

where I represents the relative laser intensity, K_{abs} is the absorption coefficient and the relative laser intensity has the solution: $I(z) = I_0.e^{-K_{abs}.z}$,

$\boldsymbol{I_0}$ being the intensity of the electromagnetic radiation at the surface of the absorbing medium.The volumetric specific heat then can then be formulated in the 3D geometry as: $Q = P_{in} K_{abs} I$, where P_{in} is the total power of the incoming laser beam.

The model implements the heat source relative motion to the disk when it couples the 3D temperature variable to the 1D equation. The coupling is done via transforming extruded subdomain variable which in this case will be the laser. The time-dependant transformation expression that results in a moving heat source is summarized as follows:

$$x = R\sin(wt),\ y = R\cos(wt),\ z = x'$$

where *x, y,* and *z* corresponds to the 3D coordinates, x' represents the 1D coordinate, ***R*** is the radius of the track , and *w* is the angular velocity in rad/sec.

Solving heat transfer due to conduction alone would require the following physical properties for the 200 nm Cobalt thin film on silica glass substrate:

Property	**Silica Glass**	**Cobalt**	**Aluminum**
Thickness (nm)	10,000	50	10,000
Density (kg/m^3)	2203	8900	2700
Heat Capacity(J/Kg.K)	703	421	900
Absorption Coefficient (1/m)	Theoretically: 0	~5e8	~8e8
Thermal Conductivity (W/m.K)	1.38	100	400

R is chosen to be 5mm to specify the radius of the trajectory and the angular velocity ***w*** is varied according to the rpm used. Input power is set to be 200mW in both cases:

1. 200mW line source of $9\mu m^2$ cross section
2. 200mW line source of $9\mu m^2$ cross section illuminating near-field apertures described in chapter II.

Assessing cooling effects resulted by air-convective cooling and surface radiation; a boundary condition on the heated boundary of the disk was assigned as using boundary conditions:

$$-n.(-k\nabla T) = q_0 + h(T_{inf} - T) + \varepsilon\sigma(T_{amb}^4 - T^4)$$

Normal heat flux at the boundary of interest

Inward heat flux due to conduction set on the boundary using the Neumann type boundary condition and mathematically represented in the previous section as: $\boldsymbol{Q = P_{in} K_{abs} I}$

Heat lost from the surface to the surrounding between the moving fluid - induced by the spin of the disk- and the disk will be represented by the product of the ***h Coefficient,*** contact surface area between the solid and the fluid, and the temperature difference between the solid and the fluid. Evaluating the Reynolds number as follows:

$$\mathbf{Re}_L = \frac{\rho.U.L}{\eta} = 0.1624 << 5e^5 \Rightarrow h_{ave} = \left(\frac{k}{L}\right)\frac{0.928.\mathbf{Pr}^{0.33}\,\mathbf{Re}^{0.5}}{\left(1+\left(\frac{0.0207}{\mathbf{Pr}}\right)^{0.67}\right)^{0.25}}$$

where ***L*** is the characteristic length expressing the thickness of the modeled boundary, **Pr** is estimated to have a 0.7 value for air, and the **Re** was evaluated to be 0.1624. Consequently, the heat transfer coefficient is approximated to be $1.93e^{16}$.

Radiative heat flux with surface-to-ambient condition is also termed into the boundary condition as the ambient surroundings behave as black body with unity absorptivity, zero reflectivity, and also having constant temperature. Although radiation is expected to occur; its effects on cooling the heated surface are insignificant as the term is evaluated to reduce the heat flux by 2066.72 (~13 orders of magnitude less than convective cooling).

Figure 3 shows the temporal behavior of temperature induced by various source powers. The source in figure one is the source feeding the near-field apertures discussed in chapter II. The maximum temperature induced by the source is ~2300K when outputting 200mW into $9\mu m^2$ cross sectional area of cobalt on a glass disk. The rotational speed adopted here is the standard 5400 rpm and convective cooling is taken into consideration by inserting the appropriate boundary conditions as described earlier.

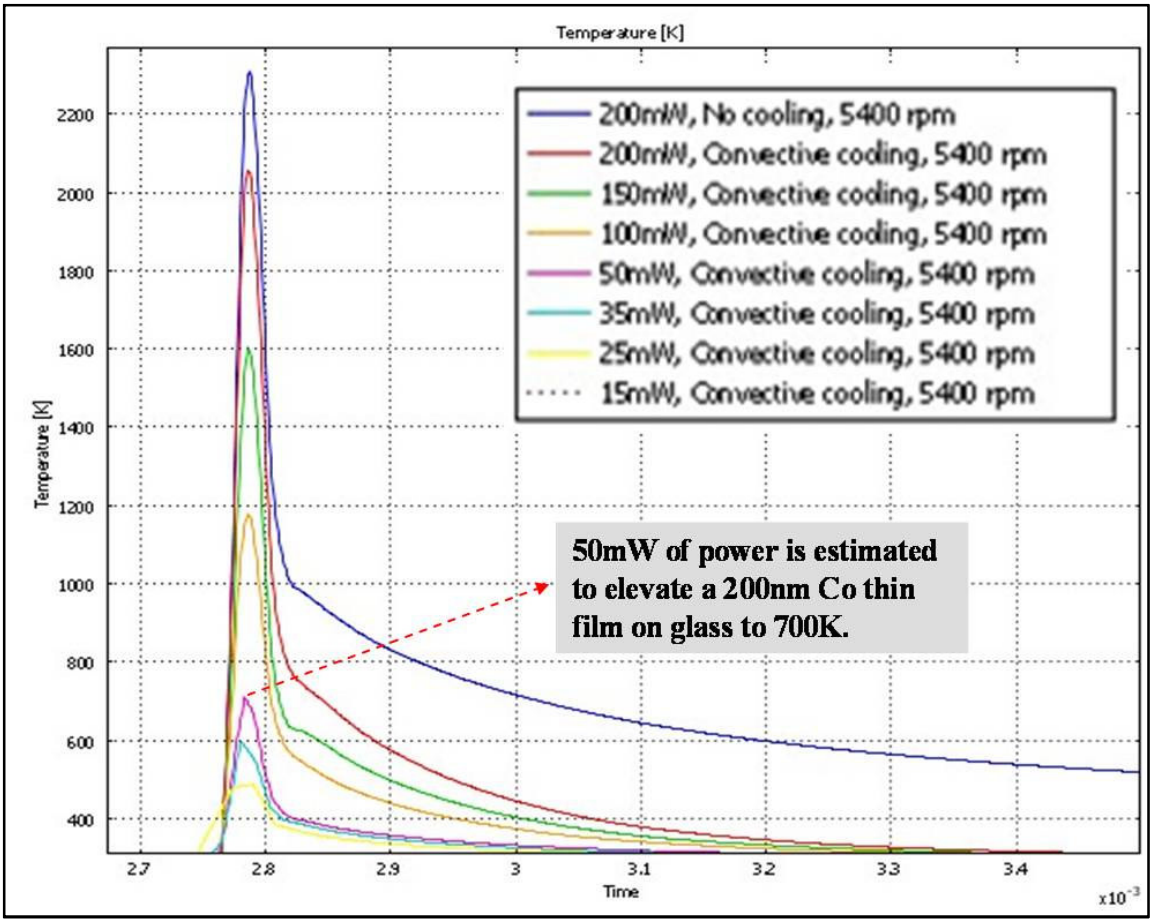

Figure 39: Thermal behavior induced by a 200mW line source given the cross-sectional dimensions (9μm^2) of the waveguide feeding the near-field apertures.

Figure 3 shows temperature induced by the near-field transducer described in chapter II when illuminated by 200mW source. The maximum temperature reached is ~900K when the disk is spinning at 3600 rpm and cooling effects are accounted for. Since other rotational speeds are used in industry, the model considered the cases where the spindle is spinning at: 3600, 5400, and 7200 rpm respectively. It is evident that the slower the angular velocity, the higher the temperature induced. It can be clearly verified that 200K difference is observed for the same power level and under the same conditions, when the rotational speed was changed from 7200 to 3600rpm respectively. Convective cooling shows a major enhancement in heat dissipation temporally as cooling times changed from

few microseconds to sub microns. It is vital to mention that this model did not consider heat sinks, which will decrease cooling time and spatial distribution significantly. This will be a subject for further investigation upon the acquisition of the physical properties of HAMR media, as the subject is still under aggressive research.

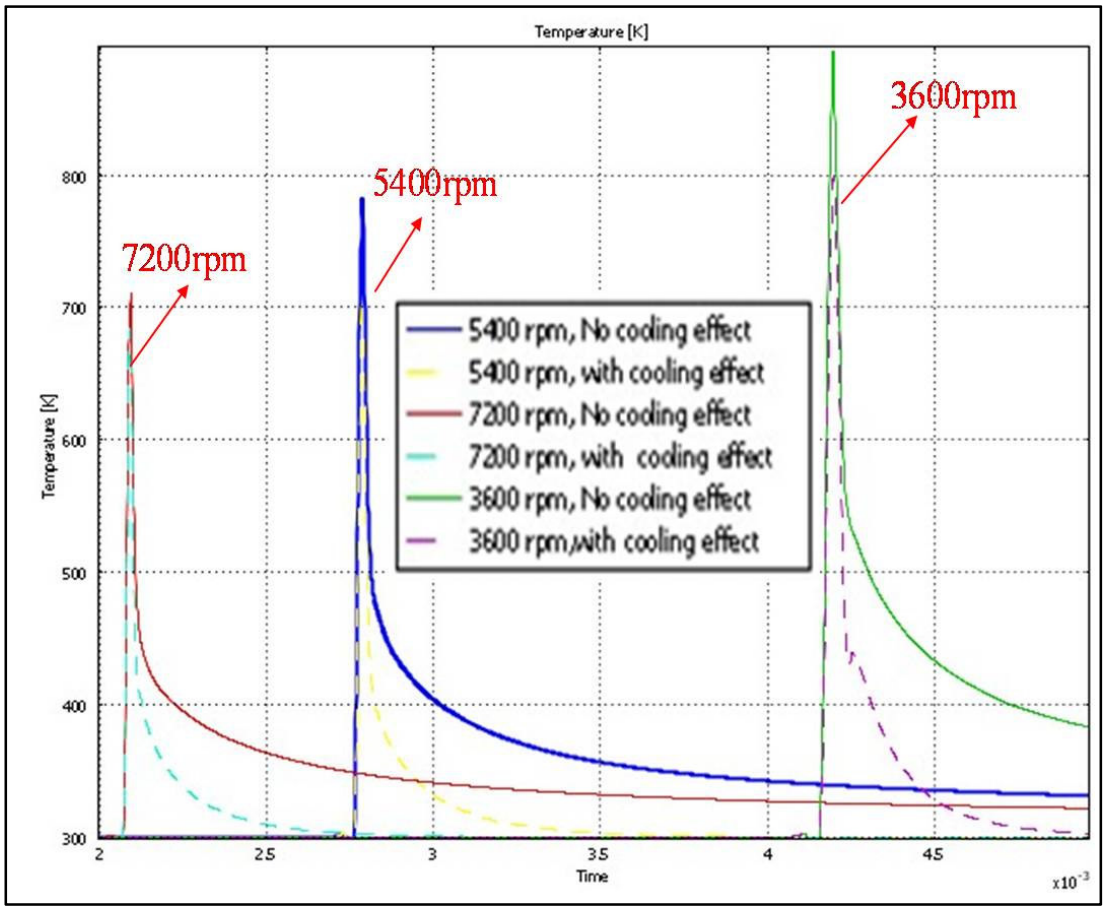

Figure 40: Heat induced by the optimal near-field transducer when illuminated with 200mW input power

CHAPTER IV: HAMR Experimental Setup and Methodology

This section thoroughly explains the setup and methods followed to demonstrate HAMR. Firstly, the components of the far-field light delivery system are discussed. Secondly, the techniques followed to align the magnetic and the optical fields on the recording disk are explained. Thirdly, the media used for the purpose are presented. Fourthly, HAMR recorded tracks are shown followed by a discussion on the results. Finally, a major issue associated with laser heating which is the lubricant depletion and its effect on the stability of the system is addressed.

1. Far-field Light Delivery System

A. Collimation and Focusing of LASER Diodes using Diffractive Limited Performance

Firstly, the specifications of the diode laser are presented:

1. High Efficiency: 0.95W/A (typ.)
2. Visible Light: 660nm (typ.)
3. Low Aspect Ratio ($\theta_{\perp} / \theta_{//}$): 1.7 (typ.)
4. Low Astigmatic Distance: 1mm (typ.)
5. Emitter x: $L_1 = 1\mu m$
6. Emitter y: $L_2 = 100\mu m$
7. Wavelength: $\lambda = 655$ nm

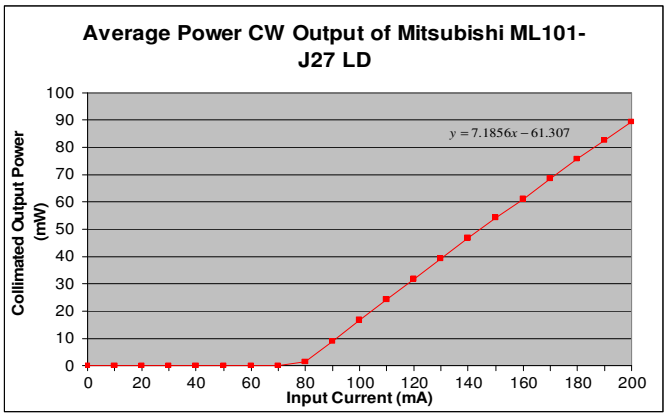

Figure 41: P-I curve of the laser diodes used for HAMR spinstand

i: Calculating Divergence from the source:

Divergence of the fast axis (HWHM): $G_1 = \frac{12.0°}{2} = 6°$,

Divergence of the slow axis (HWHM): $G_2 = \frac{20.0}{2} = 10.0°$

Divergence on $1/e^2$ (half angle to x-axis): $E_1 = a\tan\left(1.699\tan\left(G_1 \cdot \frac{\pi}{180}\right)\right) = 0.17671$

Divergence on $1/e^2$ (half angle to x-axis) $E_2 = a\tan\left(1.699\tan\left(G_2 \cdot \frac{\pi}{180}\right)\right) = 0.29107$

Therefore, numerical aperture of the optic used for collimation must be greater than $\sin(E_2) = 0.2869$.

ii. Utilizing optics to collimate the source:

I. Tube labeled **CT1** (N.A. 0.55, f = 4.5mm).
II. Tube labeled **CT2** (N.A. 0.505, f = 8mm).

The principle plane of the lens is approximately 6.97mm for CT1 and 10.47mm for CT2 from the laser mounting surface of the laser as the emitting edge of the laser is 2.47 mm from the mount, plus the 4.5 mm focal length of the optics for CT1 and 8mm for CT2.

iii. Calculating collimated beam utilizing different optics:

When **CT1** is used, the output diameter of the collimated beam is approximated to be:

$$D_1 = 2.f.\tan(E_1) + L_1 = 1.605mm$$
$$D_2 = 2.f.\tan(E_2) + L_2 = 2.9462mm$$

This is the beam diameter estimation is exactly at the exit of the collimator. However, the focusing optics sits approximately 2.2" a way. Beam at different distances from the collimator can be estimated as following (calculations will be done for our actual setup):

Beam Divergence behind optics:

$$E_3 = a.\tan\left(\frac{L_1}{2.f}\right).1000 = 0.2222mrad$$

$$E_4 = a.\tan\left(\frac{L_2}{2.f}\right).1000 = 2.7770mrad$$

Therefore at $d = 0.05588m$ from the collimator tube, beam diameter is estimated to be:

$$D_3 = d.2000\tan\left(\frac{E_3}{1000}\right) + D_1 = 1.6298mm$$

$$D_4 = d.2000\tan\left(\frac{E_4}{1000}\right) + D_2 = 3.2565mm$$

Repeating this calculation for **CT2** implies:

$$D_1 = 2.f.\tan(E_1) + L_1 = 3.2168mm \quad E_3 = a\tan\left(\frac{L_1}{2.f}\right).1000 = 0.13888mrad$$

$$D_2 = 2.f.\tan(E_2) + L_2 = 5.6493mm \quad \& \quad E_3 = a\tan\left(\frac{L_2}{2.f}\right).1000 = 1.3888mrad$$

Therefore,

$$D_3 = d.2000\tan\left(\frac{E_3}{1000}\right) + D_1 = 3.2323mm$$

$$D_4 = d.2000\tan\left(\frac{E_4}{1000}\right) + D_2 = 5.8046mm$$

In conclusion:

- CT1 yields a collimated beam of 1.6298 x 3.2565mm.

- CT2 yields a collimated beam of 3.2323x 5.8046mm.

iv: Focusing of collimated beam utilizing different optics:

$$F_1 = 2.44.\lambda.\frac{f}{D_3} =$$

$$F_2 = 2.44.\lambda.\frac{f}{D_4} =$$

Optics Used	N.A	Focal Length (mm)	Working Distance(mm)	Spot Size(µm) for CT1	Spot Size (µm) for CT2
Objective M- 40X	0.65	4.5	0.6	4.44x2.22	2.24x1.24
Objective M- 20X	0.40	9.0	1.7	8.89x4.45	4.48x2.49
Objective M- 10X	0.25	10.6mm	0.6	10.09x4.89	5.18x2.78

2. Magnetic and Optical fields Alignment

Experiments were performed on a HAMR test platform based on a modified commercial Guzik spin stand (1701B) and read/write analyzer RWA2550 with an attachment of optical laser power delivering system. Figure 42 shows images of the planned optical setup and executed on the spinstand platform. The setup provides an optical spot ranging between 2-10µm on the surface of the disk. As shown below the spot "globally" heats up the region, while the high resolution magnetic head defines the transition width.

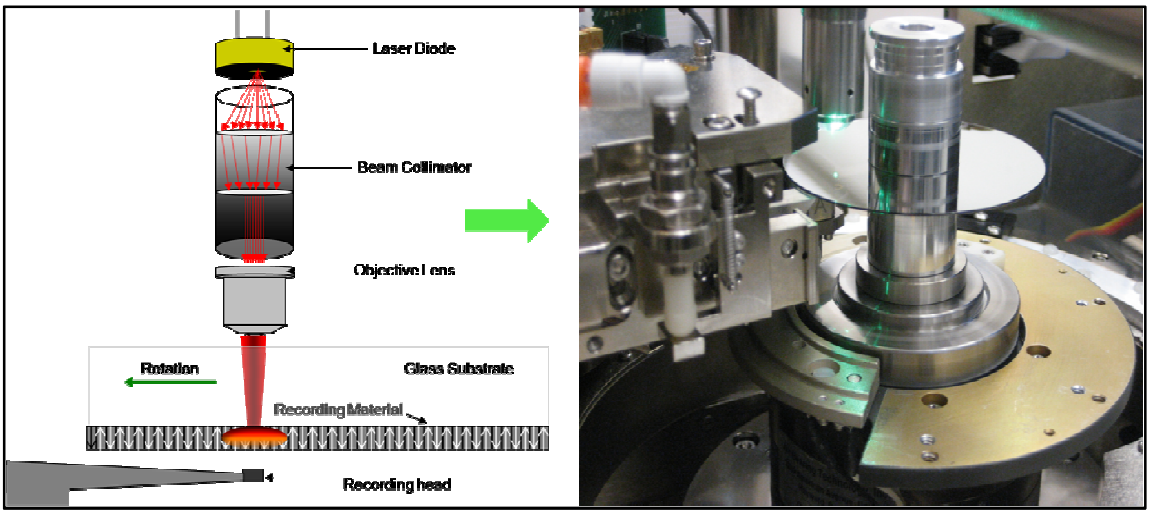

Figure 42: Far-field light delivery system plan (left), attached to the spinstand platform (right)

For aligning the optical spot with the magnetic writer, the laser spot is focused on the read/write element region through a glass disk (Figure 43). Two approaches are followed to precisely locate the position of the beam in terms of track numbers on the disk, as the

spinstand can provide unprecedented mechanical stability and positioning rather than allowing the beam to scan the disk instead. In other words, the beam holds stationary on the ABS as close as possible to the read write element. The two experiments are as follow:

1. Write a series of tracks covering the lateral position of the beam on the ABS with the laser turned OFF. The laser is then turned ON and it is expected that the laser erase recorded tracks in the regions it exposes.
2. Utilizing the thermal correspondence of the GMR reader in the magnetic head. It is expected that the resistance of the reader correlate in a linear way with temperature rise induced by the laser.

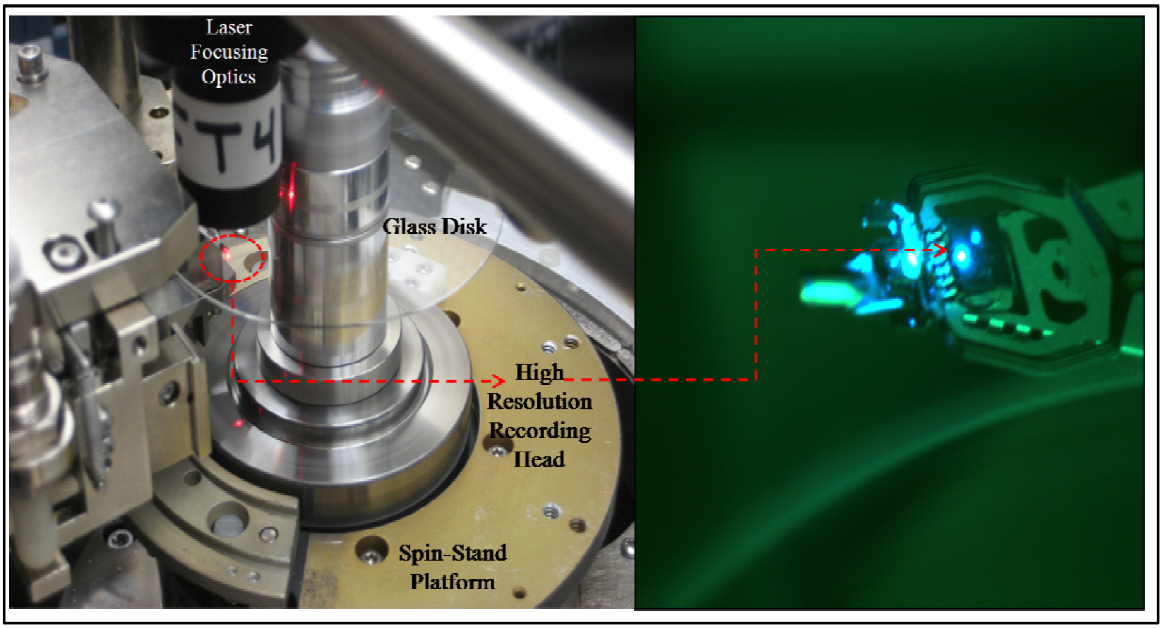

Figure 43: Laser focused through glass disk on the ABS of the magnetic head

Experiment 1: Tracks Erasure with Laser

A series of tracks are written at head current of 35mA with the laser power turned OFF. Approximately 2000 tracks are written around the central expected location of the spot.

Scanning for the spot will cover a linear distance of 200µm. Track profile for the recorded tracks is displayed (Figure 44).

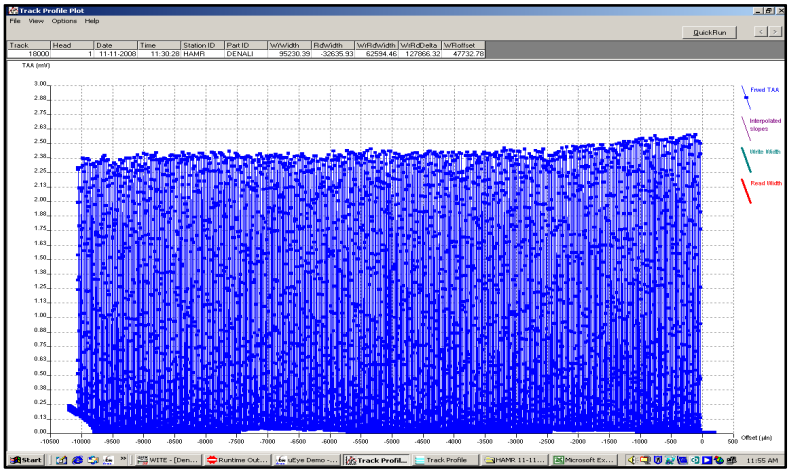

Figure 44: Track profile of recorded tracks, LASER OFF.

Secondly the laser is turned ON while the head is reading the recorded tracks. Due to the energy provided by the laser, the region exposed results in erasure of the recorded tracks. Figure 45 clearly shows erased tracks. Thus, specifying the position of the laser beam in terms of exact track numbers. It is of importance here to mention that the erasure geometry is superimposed with the Gaussian distribution of the beam. Approximately, 800µinches (FWHM of 10.16µm) is erased by the beam which has FWHM of 10µm.

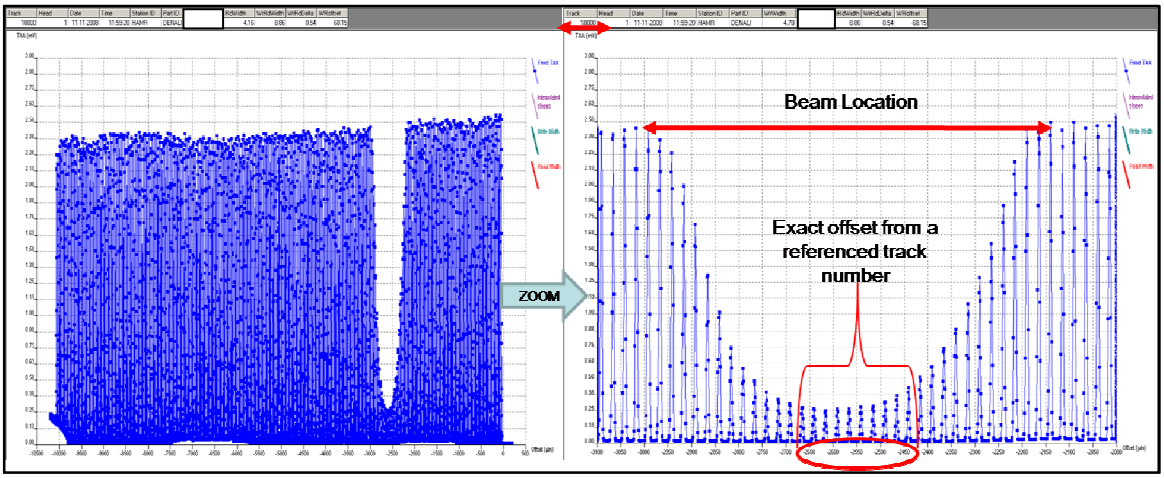

Figure 45: Track profile after LASER was turned ON showing erasure corresponding to the beam geometry.

Experiment 2: Thermal behavior of GMR Heads

For extended precision, experiment two is utilized. The head scans track by track only reading MR-resistance. Since the scan covers a smaller margin than that of experiment as the subsequent one locates the beam effectively, experiment two employs a different focusing system yielding a spot size of ~2.24μm. Resistance of MR reading heads depends on temperature in a simple manner,

$$R(T) = R_0(1 + \alpha\Delta T),$$

where α is the experimental temperature coefficient of resistance. α is estimated to be 0.0562Ω/°C for a 51Ω slider. This GMR thermal characteristic is taken to advantage when aligning the optical and the magnetic field in a HAMR system. The laser is focused onto the air bearing surface of the head and then the head is allowed to scan several tracks reading magnetoresistive impedance at each position. The maxima of the MR-reading should be positioned where the focused spot meets the reading element of the head. Then the head is allowed to move the writer-reader offset value so the writer accesses the heated spot on the disk.

The magnetic head used for the purpose has the following properties:

- Magnetic read width (MRW) 109 ± 25 nm
- Shield to shield gap=50 nm (ref)
- Read/Write Alignment = Zero Offset@ MD
- HGA MR Resistance= 46 ohms (ref)
- Max MR Bias =140 mV
- Track Pitch: 1/TPI=1/140000= 127nm

- Track-to-Track offset= 8nm

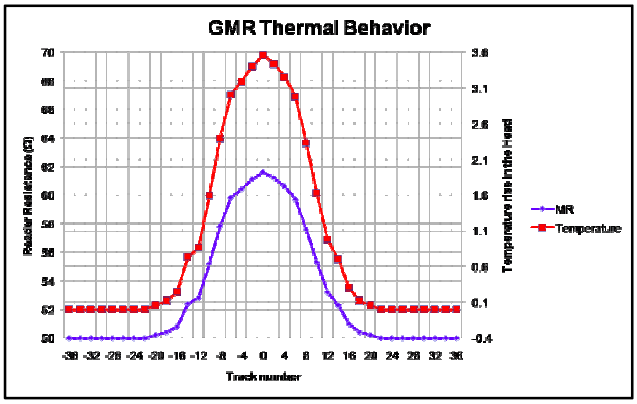

Figure 46: Thermal behavior of MR heads induced by laser heating through a spinning glass disk (5400rpm, 10mW, 2.2µm)

Figure 5, shows the maximum resistance at the central track. Central track here was reset t zero to show the position of the with respect to normalized number of tracks. Obviously, a spot of 2.2µm had double its size effect on the reader element. Temperature of the

3. Magnetic Media

Because the size of magnetic grains is approaching the superparamagnetic limit in current perpendicular media, it is necessary to produce thin film media made with magnetic alloys with larger anisotropy energies to achieve higher recording densities. Due to its high anisotropy field and good environmental stability, FePt (L_{10}) is the most promising media for achieving such ultra-high recording densities. In order to use FePt as a high-density recording media, very small (less than 5nm), uniform, highly-ordered, and isolated FePt (L_{10}) columnar grains with excellent perpendicular texture are desired.

$L1_0$-ordered FePt thin films are a leading candidate for HAMR because of their excellent material properties. FePt features high magnetocrystalline anisotropy, K_u, which is approximately 10^8 ergs/cm^3, allowing for grain sizes of only a few nanometers while maintaining good thermal stability [81,82]. However, in order to achieve the L10-phase

requires post-annealing and/or deposition at elevated substrate temperature. Several underlayers have been investigated that can reduce both the ordering temperature and the in-plane contributed variant, thus accelerating the ordering transformation temperature and inducing the (001) preferred orientation [83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94]. In this work, FePt thin films were prepared in order to investigate its thermomagnetic properties for HAMR.

High temperature VSM is studied with a temperature range of 30˚C to 400˚C. Both magnetization (M_s) and coercivity (H_c) were measured. We can approximate that the Curie temperature, that is, the temperature at which the magnetization of the film goes to zero, to be 416˚C (689 K). Figure 47 shows that we only need to heat the FePt film to 250˚C (523 K) in order to lower the coercivity enough to write with a normal perpendicular head.

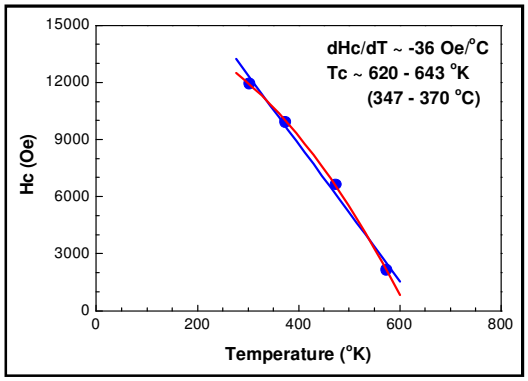

Figure 47: Coercivity dependence on temperature using a high temperature VSM.

Lubrication of the media surface is essential to ensure fly-ability of the head over it; the head is flying anywhere between 5 and 10nm only from the surface while the disk is spinning at very high frequencies up to 10,000 rpm. Without lubrication, it is almost guaranteed that the head will crash on the media

Unfortunately, and due to facilities limitation in academia, lubrication of the developed disks was not possible. This hindered the attempt to do HAMR on FePt disks until the technology is developed in academia. For future projection, we have collaborated strongly with the leader of disk drive industry, Western Digital Corporation, to lubricate our media. Western Digital support was extended to provide us with conventional perpendicular heads and media to perform our HAMR demonstration.

The media provided is CoCrPt-X deposited on glass substrate with the appropriate soft under-layers and overcoats. High temperature VSM is studied with a temperature range of 30˚C to 250˚C. Both magnetization (M_s) and coercivity (H_c) are measured.

Figure 48 shows how magnetization and coercivity both drop with temperature increase. At 250 degrees, coercivity is approximately 1200 Oe which compared to 4800 Oe at room temperature. The results look irreversible that when temperature is decreased back to room temperature, both coercivity and magnetization go to larger values than their initial ones. This is an important observation although the magnitude is not that significant. However, for media with much higher coercivity, like our candidate FePt, the reversibility becomes an issue.

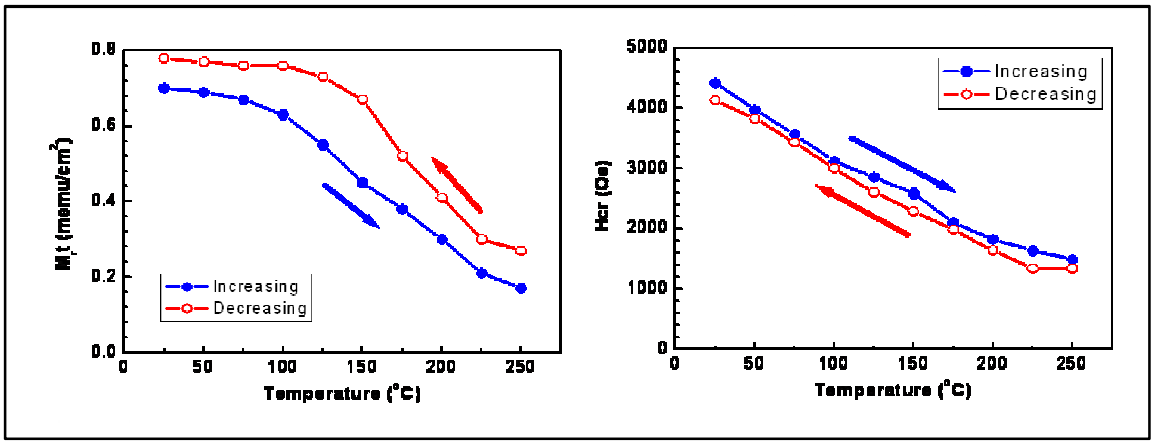

Figure 48: Temperature Dependence of Coercivity and Remnant Magnetization

Although this media is writable with conventional heads, it is with reduced head currents and reflect HAMR enhancement on the signal.

4. HAMR Demonstration

Saturation writing test is performed to investigate heat influence on writing current. The writing current ranges between 0–50mA while the laser power was in the range of 0–120mW. Saturation writing current test data is obtained and saturation curves are plotted in Figure 49.

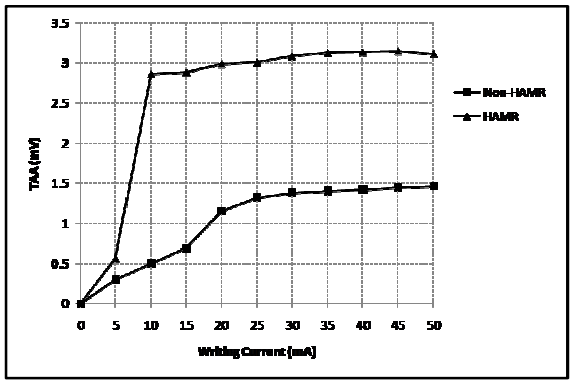

Figure 49: Saturation curves for HAMR and non-HAMR cases

A saturation writing current 25 mA was determined for the non-HAMR case. This value corresponds to coercivity at room temperature of 4500 Oe. It is observed that the saturation writing current reduces in the HAMR case to 10mA. This is a direct correspondence to the rise in temperature in the test medium. Consequently, the medium coercivity is reduced and the saturation writing current value becomes smaller.

The technique followed to locate the beam in section 1 is followed here for recording. After it is determined that recording is extremely weak at 5mA head current, this value is used for HAMR proof of concept demonstration. A set of tracks is written along the expected the location of the beam. Ideally, recording should only happens where the spot is.

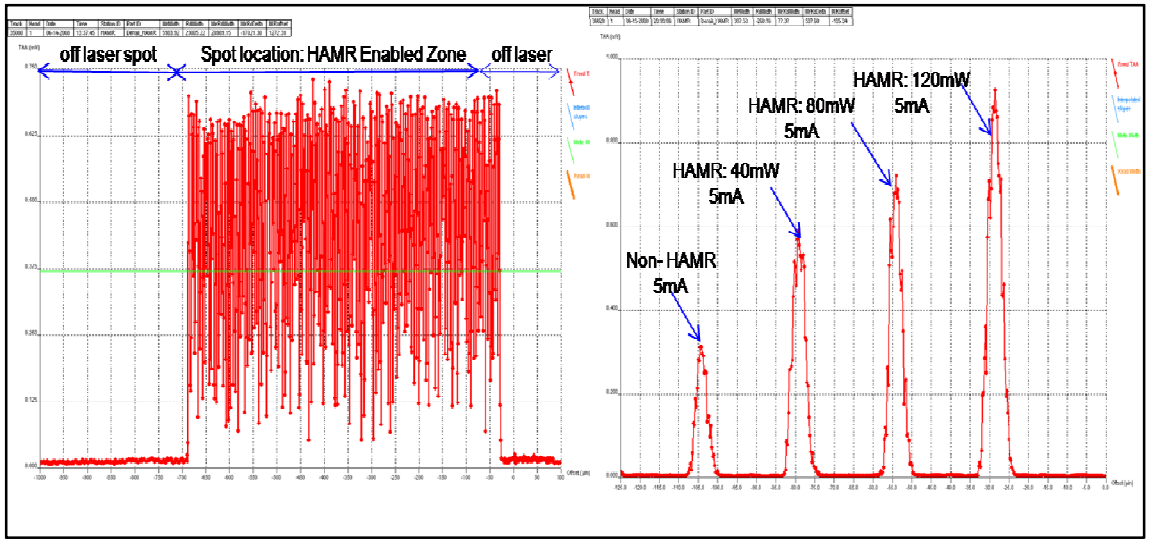

Figure 50: (left) Track profile of HAMR enabled zone (right) central track recorded with 5mA at various power levels

Figure 50 shows enabled recording with the impossible stand alone current of 5mA when 80mW expose the disk with a cross-sectional FWHM of 7.5µm. Figure 50 (left) shows amplitude variation with different powers.

To illustrate between hybrid and conventional recording cases, the head is allowed to saturate the medium while the laser is ON and have it compared to when the laser is OFF.

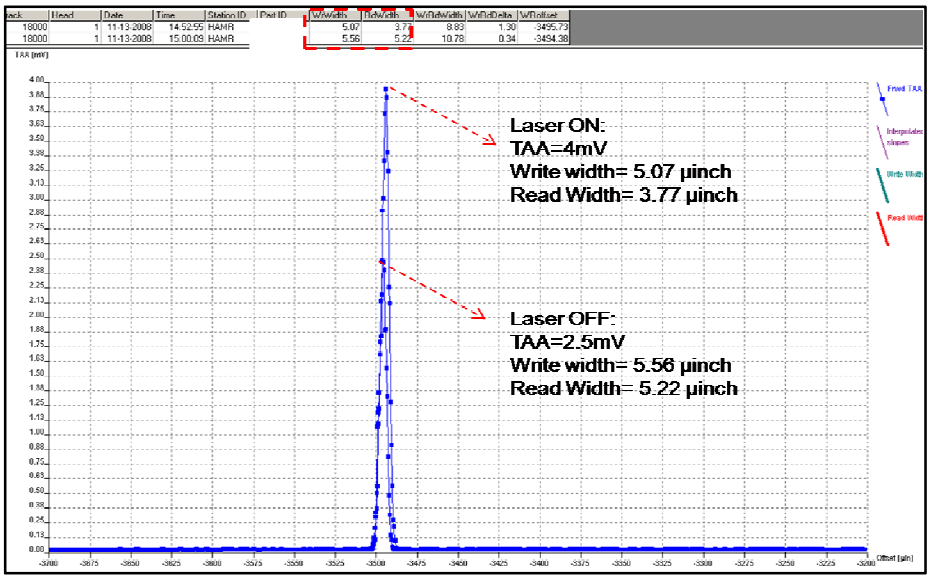

Figure 51: HAMR comparison at saturation current

Figure 51shows the enhancement of the signal level on perpendicular media when HAMR is utilized. Approximately 40% increase in signal level, write width reduction of 10% even though the head is the element that defines the track width, and finally 30% reduction in read width. It is logical to conclude that HAMR facilitates recording significantly even when the media is not HAMR designed. It is very comprehensible that heat lowers the system requirements considerably and allows the magnetic writer to dictate its specifications irrespective of media noise.

5. Lubricant Depletion

In disk drives, the integrated read/write transducer (head) is placed at the trailing edge of an air bearing slider gliding above a rotating disk where magnetic information is stored. One HAMR scheme is to integrate a waveguide to the read/write transducer, so a laser beam is focused onto the magnetic grains directly underneath the transducer during the writing cycle. The air bearing slider is supported by a thin layer air cushion (air bearing layer) to reduce friction and wear. The air bearing layer provides the primary lubrication to the head/disk interface. The estimated spacing for the air bearing layer at the trailing edge of the slider (flying height) is about 2.5 nm to achieve a 1 Tbits/in^2 storage density. In addition to the air bearing layer, a molecularly thin layer of Ztype perfluoropolyether (PFPE) lubricant (about 1.5 nm thick) is coated on the surface of the magnetic disk to prevent its scratching by the flying head when intermittent contact occurs.

The lubricant layer also serves as a corrosion protection layer. The reliability of head/disk interface strongly depends on the interface's capability to maintain the thickness of the lubricant layer. A severe lubricant depletion under the sliding head may lead to the failure of the whole system. Different slider/lubricant interaction mechanisms have been

previously identified that are able to induce lubricant depletion for a traditional (non-HAMR) disk drive system.

The lubricant surface under the sliding head is distorted by the air flow, surface tension and disjoining pressure gradient. The slider/lubricant interaction has been demonstrated to have a significant impact on the dynamics and flyability the head [95,96]. It was experimentally observed that the head/lubricant interaction not only induces head vibration, but also leads to the formation of imprinted depletion tracks and washboard patterns on the lubricant layer. Lubricant redistribution on the disk surface also occurs as a consequence of lubricant pick-up and re-deposition by the head. The dominant lubricant depletion mechanism can be attributed to the air shearing stress under the sliding head for a traditional disk drive system. The lubricant is subjected to a much harsher environment in a HAMR system than that of a traditional system, i.e. the lubricant has to sustain a very high non-uniform flash temperature induced by the laser beam heating during the writing cycles, in addition to all the other factors that have a negative impact on the lubricant performance in a traditional disk drive system. The laser beam can increase the disk surface temperature by 100–400 K above the ambient environment in nanoseconds within a very small region surrounding the beam (almost as small as the size of the laser beam). The disk temperature returns to the ambient value in nanoseconds after the laser is turned off. The increased non-uniform temperature may lead to severe lubricant depletion due to enhanced evaporation and thermocapillary stress. The mass loss rate due to evaporation is quickly increased when the liquid surface temperature is increased. For common liquids, the surface tension decreases when the surface temperature is increased. A non-uniform surface temperature results in an interfacial thermocapillary stress that tends to pull

liquids away from the hot region to cold region (thermocapillarity) [97]. At raised temperature, the lubricant also becomes more mobile due to a reduced viscosity (thermoviscosity). It is critical for the polymer engineers to understand how the non-uniform flash temperature affects the lubricant depletion in order to engineer new lubricants with desirable properties for the new HAMR system.

This thesis presents experimental data on lubricant depletion observed during the HAMR demonstration. The data was collected from an OSA Candela.

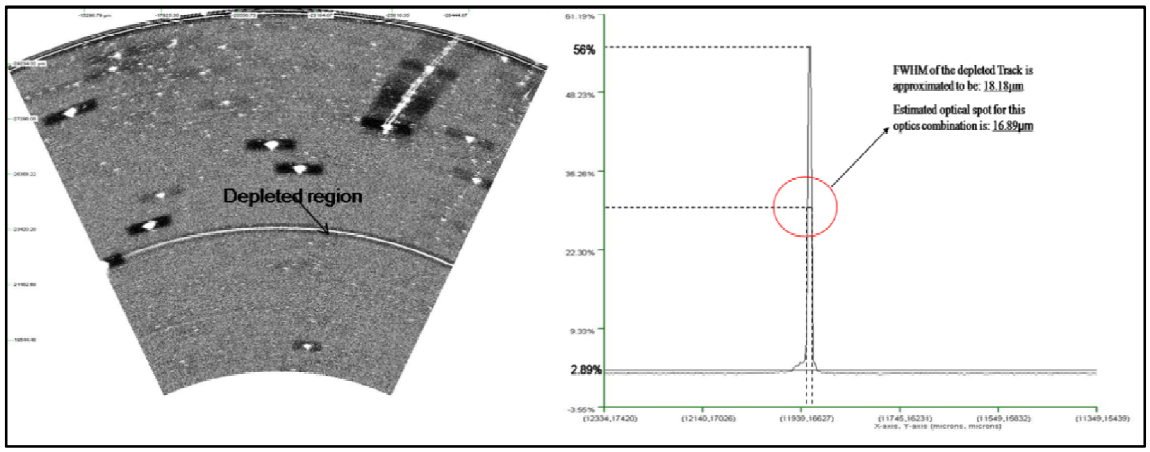

Figure 52: (left) Lubricant depleted observed with laser spot size of 18µm and 120mW CW power, (right) cross section of depleted region

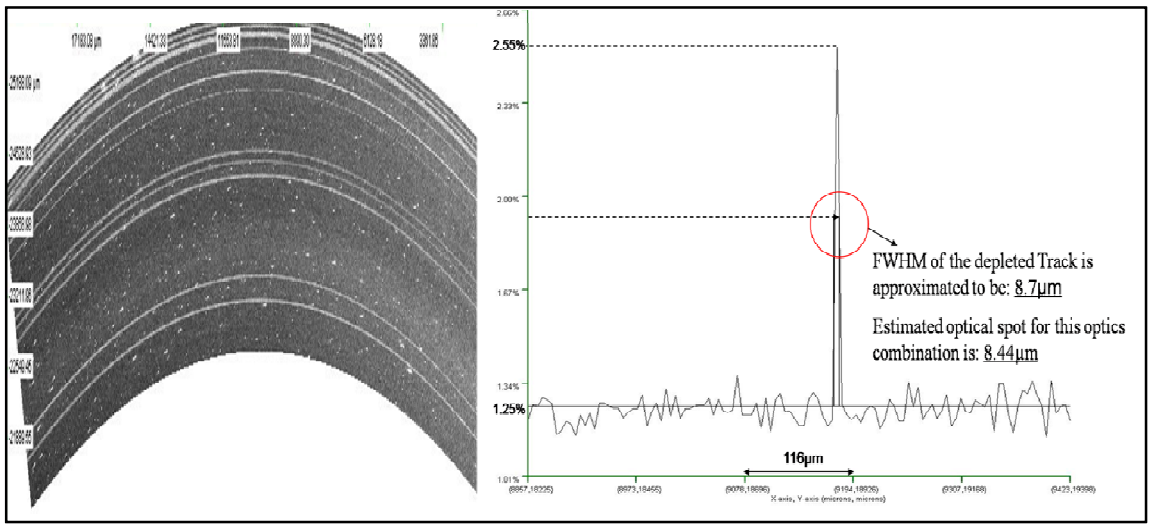

Figure 53: (left) Lubricant depleted observed with laser spot size of 8.7µm and 120mW CW power, (right) cross section of depleted region

Findings (Figure 52,Figure 53) of the experiment can summarized that (1) lube depletion occurs at very high laser beam power (100mW CW/ >500mW QCW), (2) time of exposure and spinning frequencies were found to be vital parameters: no depletion is recorded for short time exposures <60 secs, and less severe depletion is observed for higher spinning frequencies.

CHAPTER V: Protein-based Memory

1. Historical and Technological Background

The concept of optical protein-based memory has been around since the early 1970s. Yet, no commercially available protein-based memory devices exist. This review presents an analysis of the main challenges associated with the practical implementation of such devices. Furthermore, the feasibility of making the most of the truly unparalleled properties of photochromic proteins by realizing optical data storage disk drive with unmatched features and, particularly, record-high data densities and rates is discussed in detail.

It is a historical time for the information storage industry. Conventional technology is facing an inevitable end due to a fundamental limit to the laws of scaling. Reaching the superparamagnetic limit recently forced the industry to abandon longitudinal magnetic recording in favor of perpendicular recording. Perpendicular recording offers a factor of three to five improvements in the data density, which makes it an incremental solution to the growing demand for information storage. A number of technologies have been proposed to further defer the limit. The two most popular alternatives are patterned media, and heat-assisted magnetic recording (HAMR). These technologies promise to defer the superparamagnetic limit beyond 1 terabit (Tb)/in.^{2}. However, because of the large number of open questions associated with any of the technologies, it might still be premature for a reliable forecast on how far and how fast any of these alternatives could take the industry. The industry does not have a consensus on the technology to pursue beyond three to five years from today. Nonetheless, the demand for higher data density of storage and memory devices is exponentially growing,

especially with the ever-increasing use of the Internet, explosive growth of broadband communication, increasingly complex multimedia mobile devices, and rapid expansion of on-demand databases serving multinational businesses.

The concept of optical protein-based memory has been around since the early 1970s [98]. Yet, no commercially available protein-based memory devices exist. In this review, we analyze the main challenges associated with the practical implementation of such devices. Furthermore, we discuss the feasibility of making the most of the truly unparalleled properties of photochromic proteins by realizing an optical data storage disk drive with unmatched features and, particularly, record-high data densities and rates.

It appears that much of the early research on protein-based electronic devices was conducted in the former Soviet Union during the Cold War, in an effort to jump ahead of Western computing technologies, mostly for military purposes [99]. Although most results remain classified, Soviet work on bacteriorhodopsin (BR) has become widely known and has triggered extensive research in this field in Western countries and Japan [100, 101, 102, 103]

Figure 54: Photograph of salt lakes in Australia. The origin of the purple color is attributed to the abundant presence of bacteriorhodopsin. Image credit: Cheetham Salt Limited.

Bacteriorhodopsin is a purple-colored protein found in microbes that live in harsh environments such as salt marshes and salt lakes. In such extreme conditions, few organisms other than these microbes, called *Halobacteria salinarium,* can survive. Their notoriously bright purple color explains the color of the natural and artificial saline basins and salt lakes as seen from an airplane (Figure 54). When nutrients become scarce, cell membranes that are rich in BR come into play. They serve as a light-converting enzyme that keeps the organism's lifecycle going. Specifically, as a photochromic protein, BR converts light energy into chemical energy through a proton-pumping action across the *Halobacterium salinarium* cell membrane [104]. In this way, BR acts a protein powerhouse that turns on in times of famine, changing color from purple to yellow as it absorbs light [105].

For billions of years, evolution has perfected the chemical and physical functions of BR, with the protein often performing better than synthetic materials ever could. BR can function under conditions of intense light, survive in salt marshes, and withstand temperatures as high as 140º C and pressures above 300 GPa. This unprecedented thermal and chemical stability is probably the main reason that BR has become the most attractive candidate for a protein-based building block of future electronic devices [106]. Last but not the least, the BR molecule is extremely simple and efficient.

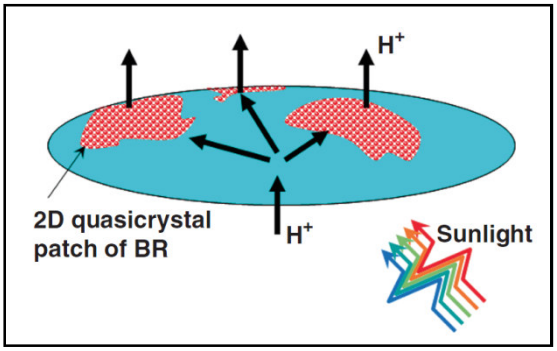

Figure 55: Schematic of a halobacterial cell covered with patches of self-assembled 2D arrays of bacteriorhodopsin molecules. The light-driven proton-pumping process is illustrated.

BR is found in the form of two-dimensional (2D) crystal patches integrated into the halobacterial cell membrane, as illustrated in Figure 55. Each patch is a self-assembled array of BR molecules. The tendency of molecules of BR to self-assemble into a crystal lattice explains the protein's unique stability [107]. Whereas the stability of the protein is determined by the collective assembly, the unique optical properties of BR are inherent to individual molecules. The characteristic diameter of each BR molecule is only 2 nm. Each molecule acts as a light-driven proton pump with an unprecedented (high) quantum efficiency of approximately 60% [108]. Such high quantum efficiency implies that less than two photons are necessary to trigger each proton-pumping process. The molecules are perfect optical cell/switches.

One ambitious goal has been to learn to implement this (almost-quantum) optical cell to leapfrog advances in electronics and, particularly, information-storage-related technologies. Furthermore, with the modern advances of genetic engineering, any additional features could be custom-tailored to suit practically any other specific technology requirement [109, 110]. Finally, the feasibility of cost-effective mass production through genetic engineering is also not out of the question.

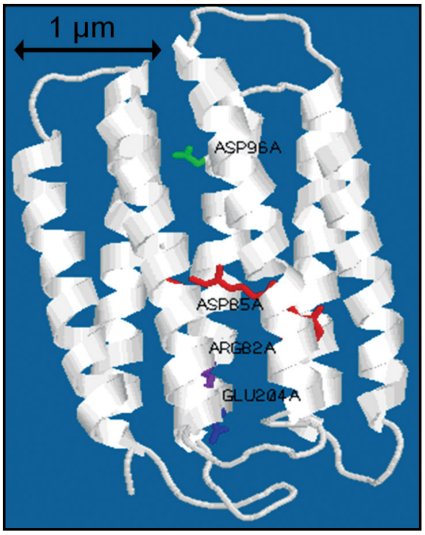

Figure 56: Three-dimensional structure of the bacteriorhodopsin molecule. The light-absorbing retinal chromophore (linked to Lys-216 via a protonated Schiff Base) is shown in red, and is in its ground-state (all-trans) configuration. Also shown are the important residues *asp-85* (yellow), *arg-82* (purple), *glu-204* (blue) and *asp-96* (green).

A three-dimensional (3D) structure of a BR molecule is shown in Figure 56. The image was created with the popular freeware application RasMol [111]. The structure of the molecule was elucidated with an accuracy of 3.5 Å in the Henderson group [112, 113]. BR consists of 262 amino acids, from which a propeptide of 13 amino acids is cleaved. Even a simple picture such as this can display the main features of bacteriorhodopsin. Its structural signature, the amino acid chain folded into seven transmembrane helices, is represented as seven gray strands. Several key groups are highlighted: The light-absorbing retinal chromophore (linked to Lys-216 through a protonated Schiff base) is shown in red and is in its ground-state (all-trans) configuration. Also shown are the important residues asp-85 (yellow), arg-82 (purple), glu-204 (blue), and asp-96 (green). The structural changes during the photocycle result in the irreversible proton-transport or proton-pumping mechanism [114]. The structural changes in BR are preceded by deprotonation of the Schiff base [115]. The longest side of the molecule is on the order of only 2 nm.

2. Photocycle of BR as the Key to Enable Information Storage beyond a 10 Tb/in^2

When bacteriorhodopsin absorbs a photon of light, it passes through a cycle of structural changes, that is, the photocycle of BR [116,117]. The key result of the changes is the isomerization from all-trans to 13-cis of the retinal chromophore bound at Lys-216 [118]. Each change in the molecular structure is referred to as an intermediate. Figure 57 shows the intermediates in the photocycle of BR at room temperature. The intermediates are identified with capital letters, and the numbers in parentheses indicate the wavelengths (in nanometers) of the absorption maxima. The thin black arrows indicate the transitions (between the respective intermediates) through thermal fluctuations, whereas the thick colored arrows reflect the photochemically induced transitions. The color of each thick arrow represents the color of the excitation frequency.

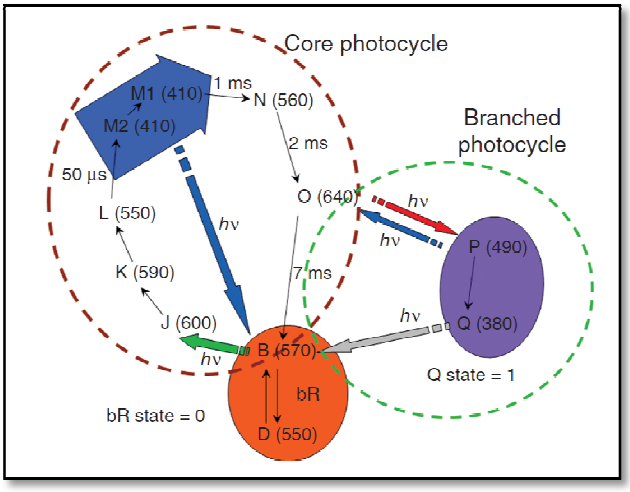

Figure 57: Schematic of the photocycle of bacteriorhodopsin at room temperature. bR is the ground state, and the intermediates (states) are identified with capital letters. The numbers in parentheses indicate the wavelengths (in nanometers) of the absorption maxima. The thin black arrows indicate the transitions (between the respective states) through thermal fluctuations, whereas the thick colored arrows reflect the photochemically induced transitions. The color of each arrow represents the color of the excitation frequency.

Naturally, these conformational light-induced changes are also temperature-dependent, implying that an intermediate can be stabilized ("frozen") at a certain temperature, usually below –10°C (specific to each intermediate). At temperatures below the freeze temperature, the transition to the next intermediate becomes suppressed. The characteristic relaxation times (lifetimes) of some of the intermediates at room temperature are shown next to the respective thin arrows. The photocycle with intermediates can be treated as a state diagram with intermediates presented as states. Each intermediate has its own absorption spectrum, and thus, intermediates (states) can be optically distinguished from each other. The characteristic absorption spectra of some of the key intermediates (in information-storage-related applications) of BR at room temperature are shown in Figure 58.

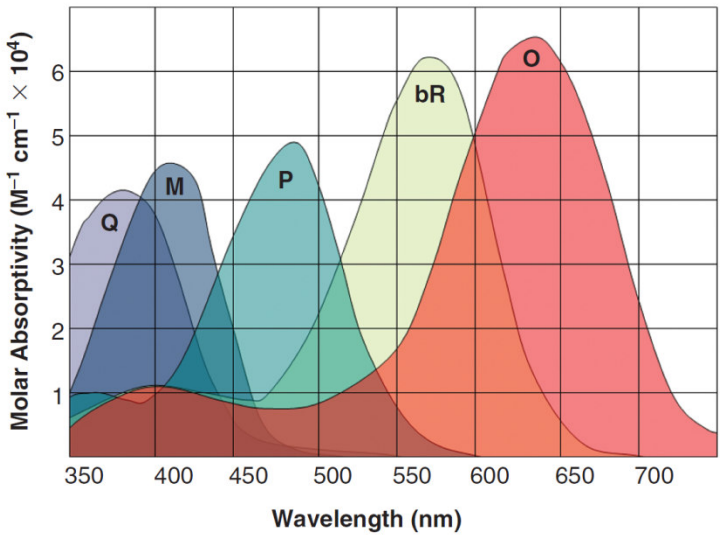

Figure 58: Schematic illustrating typical absorption spectra of the five key states, bR, M, O, P, Q, in the photocycle of BR.

The following key properties of the photocycle of BR should be highlighted with regard to information-storage-related applications:

- The photocycle of BR consists of several branches [119]. The two most significant (for information-related applications) branches are the core and branched

photocycles. During its normal function, the BR molecule remains in the core photocycle. The intermediates in the branched photocycle are not populated at physiological conditions [120]. As described in a following point, for the molecule to go into the branched photocycle, the protein should be exposed to a certain sequence of light.

- In general, the ground state of BR, or the bR state, consists of a 6:4 mixture of the D and B intermediates. Upon exposure to light, all of the protein molecules transition into the B intermediate. From there, they transition into the core photocycle, as illustrated in Figure 4.
- In the core photocycle, every transition between any two adjacent intermediates is reversible, with the exception of the transition from M2 to M1. The transition from M2 to M1 is responsible for the irreversible proton transport during the photocycle (deprotonation and reprotonation) [121]. It should be noted that the M2 and M1 intermediates have practically indistinguishable absorption spectra. Consequently, the combination of M2 and M1 is often referred to as the M state.
- The core photocycle could be used for dynamic short-term (volatile) memory applications, such as random-access memory (RAM) and devices for fast nondestructive optical processing [122]. In this case, the ground state (bR) and the complex M state are usually used as the two binary states, that is, “0” and “1,” respectively. The lifetime of the M state at room temperature varies from approximately 1 s for the wild-type BR to a few minutes for certain variants of BR, such as D96N. The wild-type BR is the original form found in nature while variants stand for genetically engineering mutants of BR. Today both types, the wild-type BR

and BR variants, are commercially available [123]. Furthermore, as mentioned in the introduction, with the modern advances of genetic engineering, it is very likely that the lifetime of the M state could be further optimized to fit specific applications. However, because the lifetime of the M state is still relatively short, the core photocycle cannot yet be used for nonvolatile memory applications.

- It is the branched photocycle that is being considered for most long-term nonvolatile memory applications. Nature itself manufactured a device with a photocycle that is almost ideally suited for long-term information-storage applications [124]. In this case, the two binary states (i.e., 0 and 1) are represented by the ground state (bR) and either the P or Q intermediate in the branched cycle. The branched photocycle originates at the O intermediate, that is, after deprotonation has taken place. When the protein is illuminated by red light, it branches off from O to P. Then, the protein thermally converts from P to Q. The Q intermediate is isolated from the core photocycle by a relatively large energy gap, which underlies the nonvolatile memory application of BR. To transition from the Q intermediate back to the bR state in the core photocycle, the protein is exposed to blue light. The process results in reisomerization of the chromophore to the all-trans configuration and the return of the protein to the ground (bR) state. As illustrated in Figure 5, the bR state and the Q intermediate have quite different absorption spectra and thus can be optically distinguished.
- To implement 3D volumetric optical memory, various paging principles have been proposed to access information in selected regions during both writing and reading steps. With paging, the intersection of two beams of light is used to select specific

regions. This often makes 3D recording devices relatively bulky and complex. In contrast, without the need to use sophisticated paging methods to read information, the branched photocycle could also be used to enable relatively cost-effective 2D protein-based information storage with areal densities above 10 Tb/in.^{2}. This review article focuses on the 2D approach.

- The characteristic time of the optically induced transitions between certain intermediates can be on the order of a few picoseconds. This implies that the fundamental limit of the data rate in a protein-based system is far beyond the fundamental limits for competing magnetic or semiconducting technologies [125,126,127].

3. Two-Dimensional Protein-Based Information Storage Systems

Three-dimensional protein-based memory technologies have been extensively studied for over a decade [128,129]. However, no commercial products based on this approach have yet been developed. For protein-based memory to be able to compete with conventional technologies, it is necessary to demonstrate the feasibility of achieving substantially higher data densities and/or data rates. A two-dimensional implementation of protein-based recording might be a more adequate and convincing way to demonstrate and take advantage of the unmatched characteristics of the photochromic protein compared to conventional alternatives. In this implementation, a disk coated with a thin protein could be used similarly to a magnetic or optical recording disk. As shown in Figure 59, in this case, a recording transducer consisting of write and read elements scans the disk with a fixed separation between the disk and the transducer. The goal would be to demonstrate substantially higher areal data densities and perhaps data rates compared

to magnetic and optical disk recording. In the following sections, we discuss the principles of writing and reading information in the case of protein disk recording.

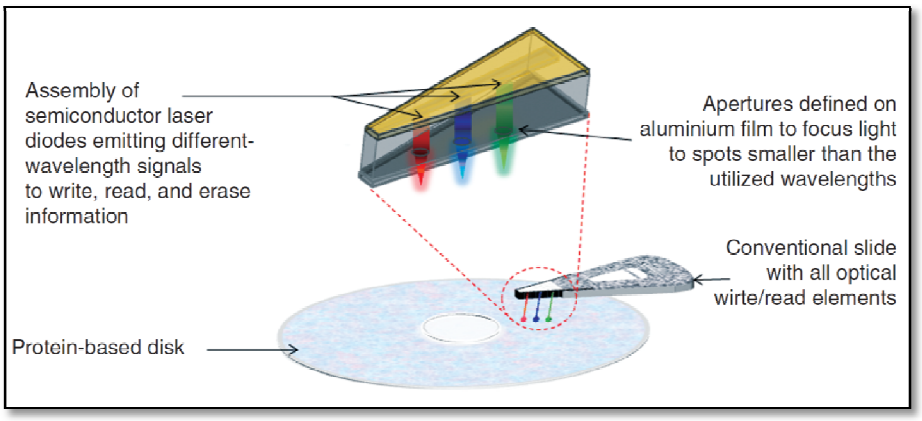

Figure 59: Schematic illustrating protein-based disk recording implementation.

A. Two-Step Write Mechanism

As mentioned earlier, to enable nonvolatile information storage, it is necessary to use the branched photocycle, with the binary states 0 and 1 defined as the bR (ground) state (in the core photocycle) and either the P or Q intermediate (in the branched photocycle), respectively. There are many different ways to implement write and read mechanisms in protein disk recording [130]. We discuss some of the most obvious ways to accomplish these tasks. To set off the branched photocycle, two diodes with two different wavelengths, namely, green and red, need to be used. First, the green light is used to convert the protein from the ground state to the O intermediate. Then, the red light is used to convert BR from O to the P intermediate in the branched photocycle (within a few picoseconds), from which it thermally transitions into the stable Q intermediate. To put this process into practice and implement a write process in protein-based disk recording, the following two-step mechanism can be used.

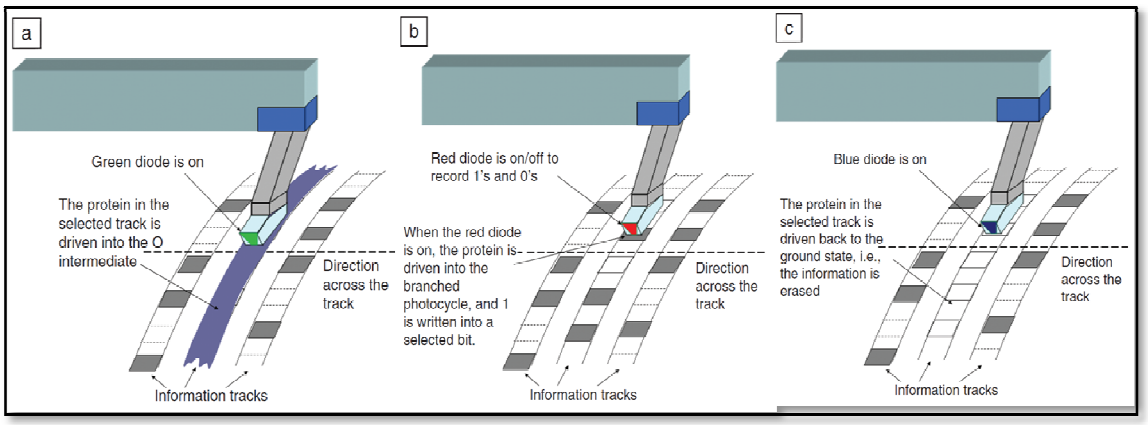

Figure 60: Schematics illustrating the two-step write process: : (a) A selected track is exposed to green light to drive the protein in this region to the O intermediate. (b) The information is recorded by turning the red diode on and off. If a certain bit is exposed to red light, the protein in this region transitions into the branched photocycle, and the binary 1 is recorded. Otherwise, with the red diode off, the binary 0 remains written in the bit. (c) When the blue diode is on, the information in a bit is erased, that is, the protein in the bit converts back to the ground state (i.e., the binary 0 state).

As the first step (absorption of the first photon), a certain track can be selected through exposure to the green light. In this case, the green diode is on for the duration of one disk rotation (period). The characteristic time for the rotation period is in the range of a few milliseconds, depending on the rotation rate. For example, it would take approximately 4 ms for the rate of 15,000 rotations per minute (rpm). As a result of this exposure, the protein along the selected track is driven to the O intermediate, as shown in Figure 60a. During the second step (absorption of the second photon), the red diode is used to record certain information in the track. The red laser is turned on and off to record 1's and 0's, respectively, in the bit regions along the track, as shown in Figure 60b. The relaxation time of the O intermediate is on the order of 7 ms, which is longer than the period of the disk rotation (3 ms) and thus perfectly suited for the described recording two-step write mechanism. To further increase the efficiency of the recording process, faster rotation of the disk might be preferred. However, to maintain an adequate temporal separation between the two photon absorption events, it might be necessary to spin the

disk in the idle mode (with all three diodes being off) for one or more rotations before the second-step photons are fired.

Finally, to erase the information in a bit, the blue diode is turned on. When exposed to the blue light, the protein in the exposed bit region converts back to the ground (bR) state, as shown in Figure 60.

Genetic engineering might become a cost-effective tool to further increase the recording efficiency. In this case, the goal of the genetic engineering would be to design a BR variant with a substantially increased yield of the O intermediate [131]. Because the O intermediate is the gateway to the branched photocycle, by increasing chances for the protein to transition into this intermediate, we also increase the efficiency of the described write process. The efficiency of the write process is defined as the fraction of photons in the source that result in a successful transition into the branched photocycle.

It should be noted that, with this relatively trivial write mechanism, the diodes do not use the wavelengths that would directly interact with either the P or Q intermediate. Therefore, the described write scheme has inherently negligible, if any, erasure; that is, it is nondestructive with respect to any previously written information. In contrast, erasure of an adjacent track by a stray field or the field generated by part of the recording head is one of the most critical issues in modern magnetic recording technologies [132].

B. Read Mechanism

The read process in the 2D case is substantially more straightforward and probably more economical than the popular read processes used in the 3D case. In the 2D case, no paging is necessary, because every bit of information can be read back with one shot (step) and the coordinates of the bit are determined by the location of the read

element, as shown in Figure 61. This implies that, to precisely track the position of the bits, data servo methods similar to those employed in conventional magnetic and optical disk drives could be implemented.

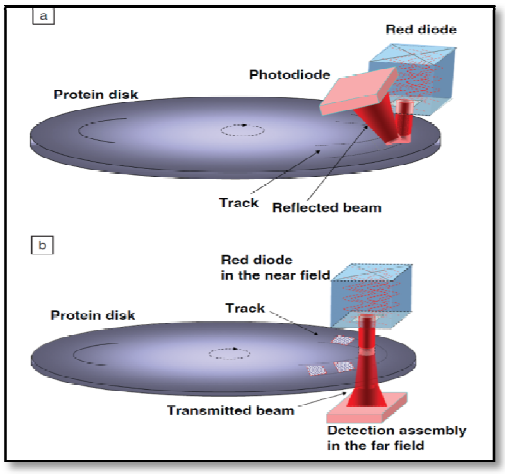

Figure 61: Schematics (not to scale) of protein disk recording systems with signal detection in the(a) reflection and (b) transmission modes.

The two most obvious ways to read information from the protein disk rely on the difference between the two binary states in the absorption maximum and the amount of fluorescence.

Reading through absorption measurements is possible because the two binary states, the bR and P/Q states, have quite different absorption maxima, as shown in Figure 58. Thus, the read signal could be chosen as either reflected or transmitted power, as shown in Figure 61a-b, respectively [133]. For example, if a red diode with a wavelength of 593 nm is used as the source, then the signal detected by the photodiode, which needs to operate at the same wavelength, will depend on whether the protein in the bit region is in the bR or P/Q state. Because the bR state absorbs substantially more signal, the detected reflected or transmitted signal will be smaller for the bR case (binary 0) than the P/Q state (binary 1) [134]. To enhance the signal, the disk substrates can be made to be opaque and reflective for the cases of reflection and transmission, respectively. Ideally,

one would expect twice as much signal difference (between the two states) in the case of reflection because the effective distance through which the light passes in the disk is twice as long as it would be in the case of transmission. However, for the same reason, the absolute value of the signal might also be weaker and, therefore, more noise-sensitive in the case of reflection. Therefore, to maximize the signal-to-noise ratio (SNR), it is critical to consider this wavelength-sensitive tradeoff, which depends on the medium thickness and the optical density of the protein film under study.

It has been theoretically predicted that fluorescence-based reading could be more efficient, that is, provide a factor-of-two higher SNR than equivalent absorption-based reading [135]. In these calculations, the shot noise was considered as the main source of noise because of the finite number of photons involved in the read process [136]. This assumption is indeed valid for measurements with conventional SNOM (scanning near-field optical microscope) systems, which must operate with inadequately low power values (on the order of a fraction of a nanowatt, when over 100 nW is required to achieve an SNR of above 15 dB). In this particular study, because of the implementation of the nanolaser technology described in the next section, the authors obtained substantially higher power values, and therefore, the relative contribution to the SNR of the shot noise could be greatly diminished. Nevertheless, the mechanism of reading through fluorescence is different and therefore should be experimentally explored with regard to this project before any definite conclusions can be drawn about the most efficient implementation of protein-based disk drives.

In the case of fluorescence measurements, the detection assembly should operate at a characteristic wavelength of the fluorescent light (and not necessarily at the

wavelength of the near-field source). The fluorescence detection is usually performed in the range from ultraviolet to red light wavelengths, that is, from ~150 nm to 650 nm [137]. When excited by a near-field source at a certain wavelength, an intermediate of BR undergoes a certain excitation and relaxation sequence, which, in turn, triggers fluorescence with a certain wavelength spectrum. It appears that different intermediates have different fluorescence spectra and therefore could be distinguished from each other [138]. Consequently, the detection assembly could be chosen to function at an independent wavelength at which the difference in the amount of fluorescence between the two binary states of BR is at its maximum.

4. Key Challenge in the Design of Write and Read Transducers

The key challenge to accomplish any of the above-described write and read mechanisms is to provide adequate (for a required SNR, i.e., 10–20 dB) power in a nanoscale beam spot [139]. Probably the most traditional way to image (read) information recorded in photochromic films would be similar to the mechanism used in scanning near-field optical microscopy (SNOM). By this, we mean not the relatively slow positioning through 2D or 3D piezoactuators used in conventional scanning probe microscopy (SPM) tools but rather the truly refined optical system used in SNOM. Indeed, the optics of state-of-the-art SNOM systems is capable of delivering record-low signal power (~0.1 nW) into a focusing spot size on the order of 30 nm. Literally, a few photons at different frequencies could be controlled in the near-field region. However, such relatively low power values are not adequate for implementing either of the recording processes, especially at room temperature, where thermal noise becomes a prevalent factor. In fact, to generate an adequate SNR (of $\gtrsim$15 dB), it is estimated

(through computer simulations) that the power of the light focused into a nanoscale spot with a diameter of ~30 nm should be on the order of hundreds of nanowatts, or at least 1,000 times higher than what could be produced by commercial SNOM systems [140, 141].

The reason for the low power-delivery capability (throughput) of commercial SNOM systems is the use of fiber probes with modified tips. The ideal dimensions of the fiber are such that the light is propagated with no detectable power loss due to the existence of the waveguide conditions. However, to focus light into a nanoscale spot size ($\lesssim$30 nm), the region close to the tip of the fiber should be modified accordingly. Now, if the light is delivered to a spot size substantially smaller than the wavelength, the waveguide in the region of the modified tip cannot comply with the waveguide conditions. As a result, over 99% of the energy (generated by the source) is lost on its way through the region of the modified tip. This explains the relatively inefficient energy-delivery capability (low throughput) of a fiber with a modified nanoprobe. Barely enough for reading, such a low power throughput is not at all adequate for writing information in protein films.

5. Solution to the Power Loss Problem in the Near-Field Optical Regime

To address the issue of inadequately low energy throughput of near-field systems, researchers worldwide have studied several near-field designs [142,143]. Among the popular designs are tapered fibers with a nonzero aperture, bow-tie antennas [144], and so called “apertureless” probes [metal-coated diode lasers with a nanoscale aperture attached to an air-bearing surface (ABS) of a recording head]. Numerical simulations have indicated that the apertureless design might be the most efficient alternative

[145,146]. However, until recently, no experiment had been conducted to demonstrate power values above a few nanowatts (for a 30-nm aperture). The main challenge was associated with the complexity of fabricating and testing such small optical transducers. [147].

Fortunately, the problem was recently resolved with the development of focused-ion-beam- (FIB-) fabricated high-throughput nanolasers at University of California–Riverside [148]. Based on the results of simulations with the commercial software COMSOL to consider thermal effects in the near field as well [149]. Ikkawi et al. conducted a groundbreaking experiment to demonstrate the feasibility of nanolasers with output power in the near field finely controllable in the range of 1 μW [150]. The researchers used FIB to fabricate near-field optical transducers (nanolasers) with apertures as narrow as 10 nm [151]. A 100-nm-thick film of aluminum was coated on the emitting edge of a 5-μW diode laser, and various aperture shapes were etched using FIB, as shown in Figure 62a. The emitting edge of the fabricated nanolaser was scanned using SNOM. The power was collected by a platinum-coated silicon probe through a photon-multiplier tube (PMT) into an avalanche photodetector PIN-FET (*p*-intrinsic-*n* field effect transistor). The light emitted by the nanolaser at every scan point was scattered by the vibrating silicon probe and thus the evanescent light in the near field was converted into propagating waves in the far field to then be detected by the PMT. The C-shaped aperture (with a 40-nm-long side) demonstrated the highest power, approximately 250 nW, focused into a spot size on the order of 30 nm (Figure 62b). The smallest power, approximately 160 nW, was demonstrated by the circular aperture. This technology is scalable to a beam diameter as narrow as 5 nm and powers ranging from 0 to above 1

μW, and it is not sensitive to the wavelength of the diode. Therefore, it could be used for both the writing and reading mechanisms described previously.

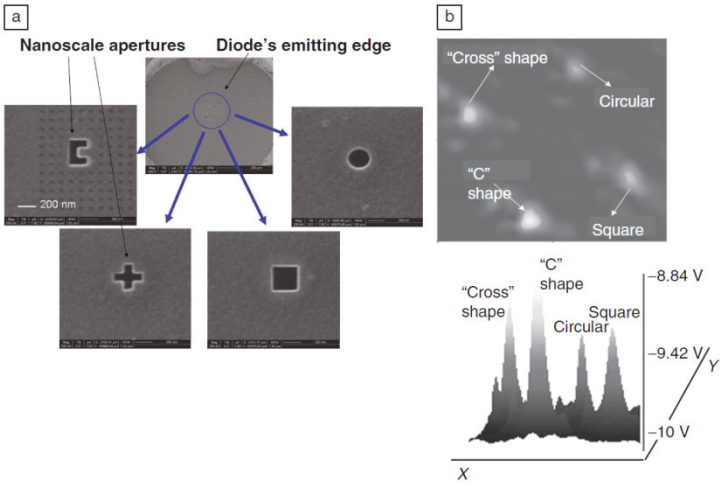

Figure 62: (a) Ion image of four different apertures shapes, namely, C-shaped, cross-shaped, circular, and square, etched by focused ion beam on the emitting edge of a diode laser. (b) SNOM scan to demonstrate a nanolaser with a 40-nm C shape to focus light into a 30-nm spot. For comparison, SNOM images from the other three aperture shapes under study are also shown. To quantitatively compare the signals from the four aperture shapes, the digitized SNOM profiles for all the four apertures are shown (bottom).

Today, a number of laboratories throughout the world are able to coat disk substrates with protein films of various properties. However, it is often perceived that one of the main remaining challenges to be overcome before this technology can be commercialized is the development of thin overcoats. These overcoats are necessary to protect the protein films in the disk drive environment and to ensure an adequate lifetime of the memory device. Fortunately, this is not a fundamental problem, and it is just a matter of time before one should be able to buy the first product. As never before, the multi-billion-dollar data storage industry is looking into finding a viable technology of the future. Protein-based disk recording is one of the most promising ones. Companies in the industry have vast experience in solving technical issues associated with the development of thin overcoats and packaging into commercial devices. Sooner or later,

this experience will be applied to the emerging technology of protein-based disk recording.

REFERENCES

1 http://www.sims.berkeley.edu/research/projects/how-much-info-2003

2 E. Daniel, C. Mee, and M. Clark, Magnetic Recording The First 100 Years, pp.270-299, IEEE Press 1999

3 Zhu, J. -G., IEEE Trans. Magn. (1992) 27, 5040

4 Zhu, J. -G., and Bertram, H. N., J. Appl. Phys. (1988) 63, 3248

5 S. H. Charap, P.-L. Lu, Y. He "Thermal Stability of Recorded Information at High Densities, "*IEEE Trans. Magn*., **33** (1), 978-83 (1997).

6 M. H. Kryder, "Future trends in magnetic storage technology," Proceedings of North American Perpendicular Magnetic Recording Conference (NAPMRC), (2003).

7 S. Iwasaki and Y. Nakamura, "An analysis for the magnetization mode for high density magnetic recording, " IEEE Trans. Magn., **13**, 1272 (1977).

8 S. Khizroev and D. Litvinov**,** *Perpendicular Magnetic Recording*, Kluwer Academic Publishers, 2004; ISBN 1-4020-2662-5

9 H. F. Hamann, Y. C. Martin, and H. K. Wickramasighe, "Thermally assisted recording beyond traditional limits*," Appl. Phys. Lett.* **84** (5), 810 (2004).

10 T. McDaniel, W. Challener, "Issues in heat-assisted perpendicular recording," *IEEE Trans. Magn.* **39** (4), 1972-9 (2003).

11 P. Herget, T. Rausch, A. C. Shiela, D. D. Stancil, T. E. Schlesinger, J.-G. Zhu, and J. A. Bain, "Mark shapes in hybrid recording," *Appl. Phys. Lett*. **80** (11), 1835-7 (2002).

12 M. Albrecht, C. T. Rettner, A. Moser, M. E. Best, and B. D. Terris, "Recording performance of high-density patterned perpendicular media," *Appl. Phys. Lett*. **81** (15), 2875-7 (2002).

13 P. B. Fischer and S. Y. Chou, "10 nm Electron Beam Lithography and Sub-50 nm Overlay Using A Modified Electron Microscope", *Appl. Phys. Lett*. **62**, 2989 (1993).

14 C. Chappert, H. Bernas, J. Ferré, V. Kottler, J.-P. Jamet, Y. Chen, E. Cambril, T. Devolder, F. Rousseaux, V. Mathet, and H. Launois, Science ***280****, 1919 (1998)*.

15 R. L. White, R. M. H. New, and R. F. W. Pease, IEEE Trans. Magn. ***33****, 990 (1997)*.

16 C. A. Ross, H. I. Smith, T. Savas, M. Schattenburg, M. Farhoud, M. Hwang, M. Walsh, M. C. Abraham, and R. J. Ram, J. Vac. Sci. Technol. *B **17**, 3168 (1999)*.

17 D. Weller and A. Moser, IEEE Trans. Magn. ***35**, 4423 (1999)*.

18 R. R. Birge, *Scientific American* **3**, 90-95 (1995).

19 N. Hampp, "Bacteriorhodopsin as a photochromic retinal protein for optical memories," *J. Chem. Rev*. **100**, 1755-76 (2000).

20 V. Renugopalakrishnan, A. Strzelczyk, Pingzou Li, A. A. Mokhnatyuk, S. H. Gursahani, M. Nagaraju, and S. L. Lakka, "Retroengineering Bacteriorhodopsins: Design of Smart Proteins by Bionanotechnology" *International J. Quantum Chem.* **95**, 627-631 (2003).

21 P. Herget, T. Rausch, A. C. Shiela, D. D. Stancil, T. E. Schlesinger, J.-G. Zhu, and J. A. Bain, "Mark shapes inybrid recording," Appl. Phys. Lett. **80** (11), 1835-7 (2002).

22 S. Khizroev, Y. Liu, K. Mountfield, M. Kryder, D. Litvinov, "Physics of perpendicular magnetic recording: writing process," JMMM **246** (1-2), 335-44 (2002).

23 P. B. Fischer and S. Y. Chou, "10 nm Electron Beam Lithography and Sub-50 nm Overlay Using A Modified Electron Microscope", Appl. Phys. Lett. **62**, 2989 (1993).

24 C. A. Ross, H. I. Smith, T. Savas, M. Schattenburg, M. Farhoud, M. Hwang, M. Walsh, M. C. Abraham, and R. J. Ram, J. Vac. Sci. Technol. B **17**, 3168 (1999).

25 S. Sun, C. B. Murray, D. Weller, L. Folks, A. Moser, "Monodisperse FePt Nanoparticles and Ferromagnetic FePt Nanocrystal Superlattices", Science **287**, 1989 (2000).

26 S. Y. Chow, P. R. Krauss, and P. Renstrom, "Imprint Lithography with 25-nanometer Resolution", Science **272**, 85 (1996).

27 D. Courjon, C. Bainier, C. Girard and J.M. Vigoureux, Ann. Physik 2, p. 149 (1993)]

28 E. Abbe. J. Roy. Micr. Soc. 2, 300 and 460 (1882).3 T.

29 Wilson, Confocal Microscopy, Academic Press, London, UK, 1990.

30 D. T. Auckland and R. F. Hamington, "Electromagnetic transmission through a filled slit in a conducting plane of finite thickness, TE Case," Report TR-77-9, Dept. of Electrical and Computer Engineering, Syracuse university, Syracuse, NY 13210, 1977)

31 (F. L. Neerhoff and G. Mur, "Diffraction of a plane electromagnetic wave by a slit in a thick screeu placed between two different media," Appl. Sci. Res., vol. 28, pp. 73 – 88, July 1973)

32 S. M. Mansfield and G. S. Kino, Appl. Phys. Lett. **57**, 2615 (1990)

33 B. D. Terris, H. J. Mamin, and D. Rugar, Appl. Phys. Lett. **68**, 141 (1996)

34 F. J. G. de Abajo, Opt. Express **10**, 1475 (2002).

35 R. Wannemacher, Opt. Commun. **195**, 107 (2001)

36 R. D. Grober, R. J. Schoelkopf, and D. E. Prober, Appl. Phys. Lett. **70**, 1354 (1997).

37 E. Oesterschulze,G. G. G., M. Muller-Wiegand, A. Vollkop, and O. Rudow, J. Microsc. **202**, 39 (2001)

38 X. Shi, R. Thornton, and L. Hesselink, Proc. SPIE **4342**, 320 (2001)

39 X. Shi and L. Hesselink, Jpn. J. Appl. Phys., Part 1 **41**, 1632 (2001)

40 T. E. Schlesinger, T. Rausch, A. Itagi, J. Zhu, J. A. Bain, and D. Stancil,Jpn. J. Appl. Phys., Part 1 **41**, 1821 (2002)

41 M. Ohtsu and H. Hori, *Near-Field Nano-Optics* (Kluwer Academic, New York, NY, 1999).

42 . Abbe. J. Roy. Micr. Soc. 2, 300 and 460 (1882).3 T.

43 C. Mihalcea, W. Scholz, S. Werner, S. Munster, E. Oesterschulze, and R. Kassing, Appl. Phys. Lett. **68**, 3531 (1996).

44 A. Hartschuh, E. J. Sánchez, X. S. Xie, and L. Novotny, Phys. Rev. Lett. **38**, 1719 (2002)

45 High-Power Laser Light Source for Near-Field Optics and its Application to High-Density Optical Data Storage," Afshin Partovi et al., Applied Physics Letters Vol. 75, pp. 1515-1517 (1999).

46 21W. A. Challener, T. W. McDaniel, D. M. Mihalcea, K. R. Mountfield, K. Pelhos, and I. K. Sendur, Jpn. J. Appl. Phys., Part 1 **42**, 2043 (2002)

47 Y. Utsumi, IEEE Trans. Microwave Theory Tech. **33**, 111 (1985)

48 W. J. Getsinger, IRE Trans. Microwave Theory Tech. **9**, 41 (1962)

49 C. A. Balanis, *Advanced Engineering Electromagnetics* (Wiley, New York, NY, 1989)

50 Calculation of the parameters of Ridge Waveguides, IRE Transactions on Microwave Theory and Techniques, Jan 1957, 5, 12-17

51 J. R. Whinnery and H. W. Jamieson, "Equivalent circuits of discontinuities in transmission lines," PROC. IRE, vol. 32, pp. 98--1 16; February, 1944

52 W. C. Hahn, "A new method for the calculation of cavity resonators, " J. Ap@. Phys., vol. 12, pp. 62–68; January, 1941

53 13 Miles Walker, "Conjugate Functions for Engineers, " Oxford University Press, Cambridge, Mass.; 1933

54 Hrennikoff, A. "Solution of Problems in Elasticity by the Framework Method", J. Appl. Mech. 8, pp. 169-175, 1941.

55 Courant, R., "Variational Methods for Solution of Equilibrium and Vibration", *Bull. Am. Math Soc*., Vol. 49, 1943, pp. 1-43.

56 J. M. Jin, *The Finite Element Method in Electomagnetics* (Wiley, New York, NY, 2000

57 G. Strang and G. J. Fix, *An Analysis of Finite Element Method* (Prentice- Hall, Englewood Cliffs, NJ, 19730

58 E. D. Palik, *Handbook of Optical Constants of Solids* (Academic, SanDiego, CA, 1998)

59 L. G. SCHULZ, The Optical Constants of Silver, Gold, Copper, and Aluminum, The Absorption Coefficient, Journal of Optical Society of America, Volume 44, Number 5, May 1954

60 K. Sendur and W. Challener, J. Microsc. **210**, 279 (2003)

61 www.microwave101.com

62 T. E. Schlesinger, T. Rausch, A. Itagi, J. Zhu, J. A. Bain, and D. Stancil, Jpn. J. Appl. Phys., Part 1 **41**, 1821 (2002).

63 Lucille A. Giannuzzi, Brenda I. Prenitzer, Brian W. Kempshall, "ION - SOLID INTERACTIONS," chapter from "Introduction to Focused Ion Beams Instrumentation Theory Techniques and Practice," Springer, (2005)

64 Orloff J, Utlaut M, and Swanson L, "High Resolution Focused Ion Beams: FIB and its Applications," Kluwer Academic/Plenum Publishers, NY, (2003).

65 E.H. Synge: A Suggested Model for Extending Microscopic Resolution into the Ultramicroscopic Region. Phil. Mag. 6, 356-362 (1928).

66 D.W. Pohl, W. Denk, and M. Lanz: Optical Stethoscopy: Image Recording with Resolution 1/20. Appl. Phys. Lett. 44, No. 7, 651-653 (1984).

67 B.I. Yakobson, A. LaRosa, H.D. Hallen, and M.A. Paesler. Thermal optical effects in NSOM probes. Ultramicroscopy 59, 334 (1995).

68 A. H. La Rosa and H. D. Hallen, Compact Method for Optical Induction of Proximal Probe Heating and Elongation. J. Appl. Opt. 41, 2015-2019 (2002)

69 . Kann, T.Milster, Heating mechanisms in a near-field optical system. Applied Optics, Vol. 36, No. 24, 5951. (1997)

70 P. G. Gucciardi, S. Patanè, A. Ambrosio and M. Allegrini. Observation of tip-to-sample heat transfer in near-field optical microscopy using metal-coated fiber probes. Applied Physics Letters 86, 203109 (2005).

71 A. H. La Rosa, B. I. Yakobson, and H. D. Hallen. Optical imaging of carrier dynamics with sub-wavelength optical probes. Appl. Phys. Lett. 67, (18), 2597-2599 (1995).].

72 Majumder J; Steen WM (1980) Heat transfer model for cw laser material processing. J Appl Phys 51(2): 941-947.

73 Yilbas BS; Sami M (1997) Heat transfer analysis of a semi-infinite solid heated by a laser beam. Heat and Mass Transfer 32: 242-253

74 Zhang HJ (1990) Non-quasi-steady analysis of heat conduction from a moving heat source. ASME, J Heat Transfer 112: 777-779.

75 M. S. Jhon, Adv. Chem. Phys. 129, 1 (2004).

76 A. A. Joshi and A. Majumdar, J. Appl. Phys. 74, 31 (1993).

77 G. Chen, J. Heat Transfer 124, 320 (2002).

78 S. Narumanchi, J. Y. Murthy, and C. H. Amon, J. Heat Transfer 125, 896 (2003).

79 Ready, J. F.: Effects due to absorption of laser radiation. J. of Applied Physics 36 (11) (1965) 462–470.

80 Masud Mansuripur and G. A. N. Connell Laser-induced local heating of moving multilayer media 666 APPLIED OPTICS / Vol. 22, No. 5 / 1 March 1983.

81 [1]O. A. Ivanov et al., Fiz. Met. Metalloved, **35**, 81 (1973).

82 D. Weller et al., IEEE Trans. Magn. **36**, 10 (2000).

83 [3]Y.N. Hsu et al., J. Appl. Phys. **89**, 7068 (2001).

84 H.Y Sun et al., Appl. Phys. Lett. **88**, 192501 (2006).

85C. Feng et al., Appl. Phys. Lett. **88**, 232109 (2006).

86 Y. F. Xu et al., Appl. Phys. Lett. **80**, 3325 (2002).

87 C. H. Lai et al., Appl. Phys. Lett. **85**, 4430 (2004).

88 Y. Zhu et al., Appl. Phys. Lett. **87**, 032504 (2005).

89 Y. C. Lai et al., J. Appl. Phys. **101**, 053913 (2007).

90 C. C. Chiang et al., Appl. Phys. Lett. **88**, 152508 (2006).

91 S. Jeong et al., J. Appl. Phys. **87**, 6950 (2000).

92 J. S. Chen et al., Appl. Phys. Lett. **81**, 1848 (2002).

93 Z. G. Zhang et al., Appl. Phys. Lett. **83**, 1785 (2003).

94 D. H. Wei et al., J. Appl. Phys. **103**, 07E138 (2008).

95 Ma X D, Tang H, Stirniman M and Gui J 2002 *IEEE Trans. Magn.* **38** 112

96 Deoras S K and Talke F K 2003 *IEEE Trans. Magn.* **39** 2471

97 Oron A, Davis S H and Bankoff S G 1997 *Rev. Mod. Phys.* **69** 931

98 D. Oesterhelt, W. Stoeckenius, *Proc. Natl. Acad. Sci. U.S.A.* **70**, 2853 (1973).

99 R.E. Armstrong, J.B. Warner, "Biology and the Battlefield," in *Defense Horizons 25* (Center for Technology and National Security Policy, National Defense University, Washington, DC, March 2003).

100 S. Hunter, F. Kiamilev, S. Esener, D. Parthenopoulos, P.M. Rentzepis, *Appl. Opt.* **29**, 2058 (1990).

101 R.R. Birge, *Sci. Am.* **82**, 348 (1994).

102 R. Thoma, N. Hampp, *Opt. Lett.* **17** (16), 1158 (1992).

103 T. Tsujioka, F. Tatezono, T. Harado, K. Kuroki, M. Irie, *Jpn. J. Appl. Phys.* **33**, 5788 (1994).

104 J.A. Stuart, D.L. Marcy, K.J. Wise, R.R. Birge, *Synth. Met.* **127**, 3 (2002).

105 R. Henderson, G.F.X. Schertler, *Philos. Trans. R. Soc. London, Ser. B* **326** (1236), 379 (1990).

106 A. Miniewicz, V. Renugopalakrishnan, in *Bionanotechnology : Proteins to Nanodevices,* V. Renugopalakrishnan, R.V. Lewis, Eds. (Springer, New York, 2006), chap. 12.

107 E.M. Landau, J.P. Rosenbusch, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 14532 (1996).

108 J.L. Spudich, *Science* **288** (5470), 1358 (2000).

109 N. Hampp, A. Popp, D. Oesterhelt, C. Brauchle, European Patent EP 0 655 162 (1992).

110 V. Renugopalakrishnan, A. Strzelczyk, P.Z. Li, A.A. Mokhnatyuk, S.H. Gursahani, M. Nagaraju, M. Prabhakaran, H. Arjomandi, S. Lakka, *Int. J. Quantum Chem.* **95**, 625 (2003).

111 www.umass.edu/microbio/rasmol.

112 N. Grigorieff, T.A. Ceska, K.H. Downing, J.M. Baldwin, R. Henderson, *J. Mol. Biol.* **259**, 393 (1996)

113 E.G. Pebay-Peyroula, G. Rummel, J.P. Rosenbusch, E.M. Landau, *Science* **277**, 1676 (1997).

114 H. Luecke, H.T. Richter, J.K. Lanyi, *Science* **280**, 1934 (1998).

115 T. Oka, K. Inoue, M. Kataoka, Y. Yagi, *Biophys. J.* **88** (1), 436 (2005).

116 G. Varo, J.K. Lanyi, *Biochemistry* **30** (20), 5016 (1991).

117 T. Rink, M. Pfeiffer, D. Osterhelt, K. Gerwert, H.-J. Steinhoff, *Biophys. J.* **78**, 1519 (2000)

118 A. Campion, J. Terner, M.A. El-Sayed, *Nature* **256**, 659 (1977).

119 S.P. Balashov, *Isr. J. Chem.* **35**, 415 (1995).

120 A. Popp, M. Wolperdinger, N. Hampp, C. Brauchle, D. Oesterhelt, *Biophys. J.* **65**, 1449 (1993).

121 G. Varo, J.K. Lanyi, *Biochemistry* **30**, 5008 (1991).

122 M. Wolperdinger, N. Hampp, *Biophys. Chem.* **56**, 189 (1995).

123 Actilor GmbH, www.mib-biotech.de.

124 R.R. Birge, U.S. Patent 5,559,732 (1994).

125 C. Denis Mee, E.D. Daniel, *Magnetic Recording* (McGraw-Hill, New York, ed. 2, 1996).

126 J.C. Mallinson, *The Foundation of Magnetic Recording 2E* (Academic Press, New York, ed. 2, 1993).

127 R.S. Muller, T.I. Kamis, M. Chan, *Device Electronics for Integrated Circuits* (Wiley, New York, ed. 3, 2002).

128 R.R. Birge, *Sci. Am.* **82**, 349 (1994).

129 N. Hampp, *J. Chem. Rev.* **100** (5), 1755 (2000).

130 F. Pina, *J. Am. Chem. Soc.* **119**, 5556 (1997).

131 D.W. McCamant, P. Kukura, R.A. Mathies, *J. Phys. Chem. B* **109** (20), 10449 (2005).

132 E. Svedberg, D. Litvinov, R. Gustafson, S. Khizroev, *J. Appl. Phys.* **93** (3), 2828 (2003).

133 K. Bagley, G. Dollinger, L. Eisenstein, A.K. Singh, L. Zimanyi, *Proc. Natl. Acad. Sci. U.S.A.* **79** (16), 4972 (1982).

134 P.J. Booth, A. Farooq, *Eur. J. Biochem.* **246** (3), 674 (1997).

135 T. Tsujioka, M. Irie, *Appl. Opt.* **37** (20), 4419 (1998).

136 R. Sarpeshkar, T. Delbruck, C.A. Mead, *IEEE Circuits Devices Mag.* 23 (November 1993).

137 N.B. Gillespie, L. Ren, L. Ramos, H. Daniel, D. Dews, K.A. Utzat, J.A. Stuart, C.H. Buck, R. Birge, *J. Phys. Chem. B* **109** (33), 16142 (2005).

138 Y. Yokoyama, M. Sonoyama, S. Mitaku, *J. Biochem.* **131**, 785 (2002).

139 D. Sarid, presented at INSIC Annual Meeting, Monterey, CA, 2002.

140 T. Tsujioka, M. Irie, *Appl. Opt.* **38**, 5066 (1999).

141 A. Partovi, D. Peale, M. Wuttig, C.A. Murray, G. Zydzik, L. Hopkins, K. Baldwin, W.S. Hobson, J. Wynn, J. Lopata, L. Dhar, R. Chichester, J.H.-J. Yeh, *Appl. Phys. Lett.* **75**, 1515 (1999).

142 F. Chen, A. Itagi, J.A. Bain, D.D. Stancil, T.E. Schlesinger, *Appl. Phys. Lett.* **83** (16), 3245 (2003).

143 M.E. Itkis, F. Borondic, A. Yu, R.C. Haddon, *Science* (312), 413 (2006).
144 R.D. Grober, R.J. Schoelkopf, III, D.E. Prober, U.S. Patent 5,696,372 (December 9, 1997

145 T. Milster, F. Akhavan, M. Bailey, J. Erwin, D. Felix, K. Hirota, S. Koester, K. Shimura, Y. Zhang, *Jpn. J. Appl. Phys.*, **1** (40), 1778 (2001).

146 X. Shi, R. Thornton, L. Hesselink, *Proc. SPIE* **4342**, 320 (2001).

147 F. Chen, A. Itagi, J.A. Bain, D.D. Stancil, T.E. Schlesinger, *Appl. Phys. Lett.* **83** (16), 3245 (2003).

148P. Patel-Predd, *MITs Technol. Rev.* (November/December 2007).

149 www.comsol.com.

150 R. Ikkawi, N. Amos, A. Krichevsky, R. Chomko, D. Litvinov, S. Khizroev, *Appl. Phys. Lett.* **91** (15), 3115 (2007).

151 S. Khizroev, D. Litvinov, *Nanotechnology* **14**, R7–15 (2004).