Near-Field Optics for Heat-Assisted Magnetic Recording (Experiment, Theory, and Modeling)

William A. Challener and Amit V. Itagi

Seagate Technology, 1251 Waterfront Place, Pittsburgh, PA 15222, USA, william.a.challener@seagate.com; amit.itagi@seagate.com

I. NEAR-FIELD TRANSDUCERS FOR HEAT-ASSISTED MAGNETIC RECORDING

One application of near-field transducers (NFT) is in heat-assisted magnetic recording (HAMR). HAMR is similar to conventional magneto-optical (MO) recording in that the data are stored in magnetic bits on a disk by heating the area of the bit with a laser beam in the presence of an external field to set the magnetic orientation of the bit as it cools. The optical head in conventional MO recording is mounted on an actuator and optical feedback signals are used to maintain a constant spacing between the head and the recording medium, which is generally on the order of tens or hundreds of nanometers. Also, for conventional MO recording the applied magnetic field is very small (approximately 0.02 T), typically generated by a large fixed external magnet, and the laser energy rather than the magnetic field is modulated with the input data stream. On the other hand, for HAMR the integrated optical–magnetic head is mounted...
on a slider, which flies over the surface of the recording medium at 10 nm or less. The applied field for HAMR is highly localized, very large in magnitude (up to 1 T or more), and generated by a miniature recording pole positioned within tens of nanometers of the optical spot. For HAMR the magnetic field from the pole is modulated with the input data stream, while the laser energy on the medium can remain constant.

Conventional magnetic recording technology records magnetic bits with down-track and cross-track dimensions less than 100 nm. Areal recording densities of up to 400 Gb/in.$^2$ have been demonstrated. Unfortunately, it is difficult with conventional recording technology to achieve substantially larger densities. As the storage density increases, the area of each bit decreases, but to maintain the same level of signal-to-noise ratio, the number of magnetic grains within each bit must not decrease. Therefore, greater areal densities require smaller magnetic grains. The magnetic grain diameter is presently on the order of 10 nm. As the volume of a magnetic grain is reduced, it reaches a point where the magnetic orientation of the grain becomes thermally unstable. Essentially, the average thermal energy within the grain, which is proportional to $k_B T$, becomes comparable to the magnetic anisotropy energy, $K_u V$, where $k_B$ is Boltzmann’s constant, $T$ is the absolute temperature, $K_u$ is the magnetic anisotropy constant of the grain, and $V$ is the volume of the grain. This has been termed the “superparamagnetic limit” of magnetic storage density. Although it is possible to increase the stability of the recording medium by increasing the magnetic anisotropy of the recording material, eventually the applied magnetic recording field from the recording head is insufficient to switch the magnetic state of the medium. It requires new technologies to achieve areal densities beyond this point. In HAMR the magnetic anisotropy of the medium is momentarily reduced to enable recording by raising its temperature. The recorded bit is then quickly cooled back to its high-anisotropy state at ambient temperature to stabilize it. In this manner, extremely high magnetic anisotropy materials such as FePt can be recorded in a HAMR system, thereby potentially enabling areal storage densities in the range of 1–40 Tb/in.$^2$.\textsuperscript{1,2}

At these storage densities, the recorded domains are only tens of nanometers in length and width. Hence, the optical spot used to heat the recording medium in HAMR must be an order of magnitude smaller than the optical wavelength of low-cost and high-power
semiconductor lasers, which is in the range of 650–830 nm. A conventional lens can only focus light to a spot size defined by the diffraction of light from the clear aperture of the lens. The diffraction limit for the focused optical spot is given by

\[ d = \frac{0.5 \lambda}{n \sin \theta}, \]

where \( d \) is the full-width spot diameter at the half maximum point (FWHM), \( \lambda \) is the wavelength, \( n \) is the refractive index of the medium in which the light is focused, and \( \theta \) is the half angle of the cone of focused light. In other words, conventional optics are able to focus light to a spot size of approximately a half wavelength. For example, the new Blu-ray technology operates at a wavelength of 405 nm and a numerical aperture (equivalent to \( n \sin \theta \)) of 0.85, which corresponds to a focused optical spot size of approximately 240 nm. Although this is a very small optical spot, it is still much too large for use with HAMR. In a sense, HAMR replaces the difficulty of surpassing the superparamagnetic limit with the difficulty of focusing light below the diffraction limit.

A solid immersion lens (SIL) is a somewhat unconventional focusing optic that is able to bring light to a focus inside a transparent high-index material, resulting in a spot size that is \( n \) times smaller in diameter than that for light brought to a focus in air by a lens with the same numerical aperture, where \( n \) is the refractive index of the SIL. Such an optic may be an essential part of a HAMR disc drive. However, this spot size is still at least twice as large as that required for a 1-Tb/in.\(^2\) HAMR areal storage density. Therefore, a HAMR disc drive requires a new approach for concentrating light energy into a spot smaller than the diffraction limit. Such devices are possible by making use of the “near field,” that is, by concentrating energy that consists of both propagating and nonpropagating components. Because the nonpropagating components are evanescent – they decay exponentially with the distance from their source – the NFT can only generate a sub-diffraction-limited spot within a distance that is much smaller than a wavelength. Fortunately, even in a conventional disc drive, the recording head flies within 20 nm of the recording surface, well within the near field of a HAMR transducer.

NFTs are becoming popular in various areas of spectroscopy (e.g., surface-enhanced Raman spectroscopy with a lithographically defined surface of gold or silver nanoparticles of various shapes).
NFTs may also become useful for extremely high-density optical lithography and optical imaging. However, the requirements for the NFT in HAMR are substantially greater than those for NFTs in spectroscopy, lithography, or near-field imaging. In all cases, the NFT must concentrate optical energy into a spot much smaller than the diffraction limit, or in the time-reversed sense, scatter or transmit light from an optical region much smaller than the diffraction limit. For HAMR, however, the efficiency of the NFT is also of primary importance. A NFT which confines the light energy to a 20-nm spot but which only conducts one part in $10^5$ of the incident laser power into this spot is not useful for HAMR even though it might work for spectroscopy. To make use of low-cost, commercial semiconductor lasers for HAMR disc drives, the NFT coupling efficiency must be approximately 5%. Although this may seem like a very small efficiency, it should be remembered that the efficiency of light transmission through a near-field tapered optical fiber with a 50-nm aperture is only approximately 0.001%. Thus, the HAMR NFT must have a power coupling efficiency into the recording medium that is orders of magnitude greater than the transmission efficiency of tapered optical fibers.

This immediately raises the question whether it is correct to make a comparison between the transmission efficiency of a NFT and its power coupling efficiency into a recording medium. Transmission efficiency is a far-field property, while coupling efficiency is a near-field property. Is it possible that a very tiny aperture with a far-field transmittance of $10^{-5}$ could nevertheless in the near field couple optical power efficiently into a recording medium? Are the far-field and near-field properties of NFTs related and, if so, in what way? Of course the far-field transmittance of a NFT is only defined in the absence of a recording medium. When a recording medium is placed within the near field of a transducer, does that significantly affect the optical properties of the NFT itself? What is the best way to judge the merit of a NFT?

In the literature for NFTs a variety of approaches have been reported for judging the merit or efficiency of NFTs. One popular efficiency measure is the value of the enhancement of the electric field in the vicinity of the NFT relative to that of an incident plane wave. Another figure of merit (FOM) is the amount of power that is transmitted through a NFT aperture relative to the incident power on the NFT integrated over the surface of the NFT aperture. This FOM...
assumes that the transmitted power (a far-field quantity) is directly related to the near-field coupling efficiency of the transducer. It is not simple to apply this FOM to NFT antennas. A sensible FOM for HAMR is the power coupling efficiency, i.e., the ratio of the power dissipated within the optical hot spot of the recording medium to the total power in the incident beam. This FOM is generally somewhat more difficult to compute than the other FOMs because it requires an incident focused light beam with a well-defined power rather than a simple incident plane wave.

With an appropriate FOM, it is possible to study, optimize, and compare different NFT designs in detail theoretically. The NFTs can generally be categorized as either antennas or apertures, although there are some NFT designs that incorporate aspects of both. Several mechanisms can be identified in these different designs that enhance the coupling efficiency of light into the recording medium. For example, in most cases the NFT is chosen to support surface plasmons that resonate in the incident optical field and thereby greatly enhance the optical field amplitude in the near field of the transducer. Often these NFT designs incorporate sharp tips to further increase the field amplitude via the lightning rod effect. A small gap between two regions of the NFT can be used to enhance field amplitudes via the dual-dipole effect. Other mechanisms are designed for more efficiently funneling energy from the incident beam into the active region of the NFT.

The outline of this chapter is as follows. In Sect. II we discuss the modeling techniques employed in this study. In Sect. III the difference between the near field and the far field is considered and it is argued that any FOM based on far-field quantities is not appropriate for HAMR. Various FOMs are considered in Sect. IV as they relate to HAMR. Several mechanisms that may be employed by NFTs for enhancement of the coupling efficiency are discussed in Sect. V. In Sect. VI these mechanisms are studied for a variety of transducer designs and a FOM is used to compare them. Because both antennas and apertures may be useful for HAMR, we discuss the relationship between these different transducer approaches in Sect. VII. The relationship between the far field and the near field, especially in so far as far-field measurements may be used to characterize NFTs, is discussed further in Sect. VIII. Finally, the means for efficiently illuminating the NFT is an important topic which we address in Sect. IX in a discussion on photonic nanojets.
II. MODELING TECHNIQUES

Analytical solutions for the electromagnetic fields can be obtained for only a small set of physical objects which generally exhibit some form of symmetry. A solution to the problem of the scattering of plane waves by spherical or ellipsoidal objects was found by Mie.\(^8\) Many useful insights can be obtained from this semianalytical theory and we make use of it in this chapter to discuss the local field enhancement due to the surface plasmon resonance of metallic spheres. However, in general it is not possible to study the wide variety of NFT designs analytically. We have found that the scattered field finite difference time domain (FDTD) technique\(^9\) is well suited to our transducer studies. In this technique, the incident electric field is defined analytically throughout the computation space, but the scattered field is computed numerically in the time domain at specific points throughout the computational space on a Yee cell lattice as shown in Fig. 1. As can be seen from this figure, the individual electric and magnetic field components are specified at different points within each cell.

Figure 1. The finite difference time domain (FDTD) computation space is composed of Yee cells which define the locations of the electric and magnetic field components on the cell edges and cell faces, respectively.
The Yee cells must be chosen sufficiently small that the numerical approximation is accurate. In practice, for modeling surface plasmon phenomena at optical wavelengths for highly conducting metals, we have found that a cell size of \((2.5 \text{ nm})^3\) is generally reasonable. The cell size and the computational resources in turn limit the size of the computation space. At the boundaries of the space, appropriate boundary conditions must be implemented so that scattered fields do not get reflected back into the computation space. For the simulations in this chapter, we used either reradiating boundary conditions or perfectly matched layers. The size of the Yee cell also determines the maximum size of the time step which can be used to avoid numerical instability. The Courant time is an upper limit on the step size, but in practice it is found that somewhat smaller time steps are required for stability. Smaller Yee cells require shorter time steps. For plane wave scattering problems, it is generally necessary to run the simulation for five or more complete periods of the wave to reach nearly steady state conditions. At integral values of the time step the scattered electric field at each Yee cell is updated from the incident electric field, the scattered electric field, and the scattered magnetic field. At half-integral time steps the scattered magnetic field is updated from the scattered electric field. In the scattered field FDTD technique, as opposed to the total field FDTD technique, the update equations are significantly more complex for materials that include optical losses. We model metals as Debye materials in the FDTD calculation with a separate set of Debye parameters for each wavelength.

III. NEAR FIELD COMPARED WITH FAR FIELD

It is not unusual to find articles on NFTs that begin with the classic result of Bethe for the far-field transmission efficiency of a circular aperture. Bethe was able to solve analytically for the light transmitted through a circular aperture in an infinitesimally thin perfectly conducting sheet. He discovered that when the aperture diameter is small compared with the wavelength of the incident light wave, the transmission efficiency is given by

\[
T = \left( \frac{64 \pi^2}{27} \right) \left( \frac{d}{\lambda} \right)^4.
\]  

(2)
where $T$ is the ratio of the power per unit area transmitted through the aperture to the power per unit area incident upon the aperture and $d$ is the diameter of the aperture. Obviously, the fourth-power dependence on the ratio of the diameter to the wavelength causes the transmitted power to fall drastically with aperture diameter. This discouraging result convinced many people that it was impossible to efficiently conduct optical energy into volumes much smaller than $\lambda^3$.

However, it is not difficult to demonstrate that far-field measurements are not necessarily a measure of near-field efficiency. Perhaps the simplest example is to consider a SIL that is illuminated only for angles greater than $\sin^{-1}(1/n)$, where $n$ is the index of refraction of the SIL, as shown in Fig. 2. In this case no light is transmitted into the far field – it is all internally reflected at the bottom interface of the SIL. This optical transducer would fare very poorly with any FOM that is based on a far-field property. However, if any object such as a recording medium is placed adjacent to this surface, the light energy in a highly concentrated spot at the focus of the SIL will be coupled into the medium. The near-field coupling efficiency is not zero and in fact may be quite respectable. This simple example, therefore, demonstrates that there is not necessarily a one-to-one correspondence between far-field transmittance and near-field coupling efficiency.

IV. FIGURES OF MERIT

How does one know if a particular NFT is promising for use in a specific application? To optimize a particular NFT design or to make comparisons between different NFTs, it is important to have a
FOM. A variety of FOMs have been used in the literature to judge the performance of NFTs. Examples include far-field transmittance, peak field intensity in the neighboring medium, percent dissipated power in the medium, and temperature rise in the medium. We shall consider each of these in turn and discuss their advantages and limitations.

1. Far-Field Transmittance

For a NFT that is an aperture, the simplest theoretical and experimental procedure for evaluating NFT efficiency is to calculate or measure the far-field transmittance. The total power transmitted through the aperture must be normalized in some manner. The incident beam in a theoretical calculation is frequently a plane wave; however, the incident power in a plane wave is infinite. Because only a finite amount of power is transmitted through an aperture, the transmittance of the aperture as a ratio of transmitted power to incident power is exactly zero for a plane wave; therefore, the transmittance of an aperture for a plane wave is not a useful FOM. However, there is a finite amount of power in a plane wave in the cross-sectional area of the aperture. A popular FOM is the ratio of the transmitted power from a plane wave incident upon the NFT (or the absorbed power of the medium next to the NFT) to the power/area of the plane wave multiplied by the cross-sectional area of the aperture. For periodic arrays of NFTs, the FOM is the ratio of the transmitted power to the power/area of the plane wave multiplied by the area of a unit cell. The power/area for a plane wave is

$$ S = \frac{|E_0|^2}{2\eta} \left( \text{W/m}^2 \right), $$

where $E_0$ is the amplitude of the incident plane wave in volts per meter and $\eta$ is the impedance of the medium of propagation in ohms. For free space,

$$ \eta = \sqrt{\frac{\mu_0}{\varepsilon_0}} \approx 377 \Omega. $$

Unfortunately, as discussed in the previous section, it is easily demonstrated that the far-field transmittance is not necessarily related to the near-field coupling efficiency for a NFT meant to be
used for HAMR. Although this FOM may be appropriate for NFTs designed for some applications, it is not appropriate for designing or comparing HAMR transducers and will not be used in this study.

2. Peak Field or Field Intensity

Another common FOM in studies of NFTs is the ratio of the peak electric field amplitude in the vicinity of the NFT to the electric field amplitude of an incident plane wave. This is a particularly appropriate FOM for designing NFTs for surface-enhanced Raman scattering. The Raman signal from various organic compounds is experimentally found to be enhanced by many orders of magnitude\textsuperscript{12,13} when the organic molecules are attached to rough silver surfaces or to gold or silver nanoparticles of different shapes. Because the Raman effect is a two-photon process, the intensity of the scattered light is proportional to the fourth power of the electric field in the vicinity of the molecule. Indeed, the amplification of the Raman spectrum is so large that individual molecules can be detected.\textsuperscript{14,15} By optimization of the NFT design for the greatest field enhancement, arrays of NFTs on a substrate can be optimized for surface-enhanced Raman scattering. On the other hand, the local field enhancement of the NFT when it is suspended in free space or some other dielectric medium is not a particularly appropriate FOM for HAMR. This is because the field enhancement from a NFT in free space can be significantly different from the field enhancement in the presence of any metallic or lossy medium. This will be demonstrated in the studies of triangle and bow-tie antennas discussed in Sect. V.

The $|E|^2$ field intensity in a lossy medium is directly proportional to the dissipated power. In particular,

$$P_{\text{diss}} = \frac{1}{2} \text{Re}(\sigma) |E|^2,$$  \hspace{1cm} (5)

where $\sigma$ is the complex optical conductivity of the lossy material. The optical conductivity is directly related to the complex optical dielectric constant of the lossy material,

$$\sigma = -\frac{i2\pi\epsilon\epsilon_0}{\lambda} \left( \frac{\epsilon}{\epsilon_0} - 1 \right) (\Omega m)^{-1},$$  \hspace{1cm} (6)
where \( c \) is the speed of light and \( \varepsilon_0 \) is the permittivity of free space, \( 8.854 \times 10^{-12} \). Therefore, the peak field intensity within the medium normalized by the incident field intensity is closely related to the total power dissipation within the medium and can be a useful FOM for HAMR studies of transducers. It will be used for the studies discussed in Sect. V. There are two disadvantages of this FOM, however. First of all, the spatial distribution of the field intensity in the recording medium can differ greatly for different NFTs. The total power dissipated in the medium is proportional to the \(|E|^2\) field intensity integrated over the volume of the medium, not just the peak \(|E|^2\) at some point within the medium. A second issue with this FOM is more subtle. The peak field intensity in the incident beam is a function of the wavelength and polarization of the incident focused beam. If two NFTs couple light into a medium with the same peak \(|E|^2\) FOM, but at two different wavelengths, then the NFT which operates at the shorter wavelength will be more efficient at coupling power into the medium. At the shorter wavelength, the focusing optics will generate a smaller spot with dimensions proportional to \( \lambda^2 \). Therefore, for the same peak field amplitude of the incident beam, there is more optical power in the vicinity of the transducer at the shorter wavelength to be coupled into the medium. An alternative way of looking at this is that for a given optical power in the incident beam, the field intensity at the focus of the beam is proportional to \( 1/\lambda^2 \). Therefore, shorter-wavelength operation of a NFT is an advantage. This factor is not explicitly taken into account in a FOM based solely on \(|E|^2\).

### 3. Percent Dissipated Power in the Recording Medium

Although every FOM has certain advantages and disadvantages, one of the best optical FOMs for HAMR is based on the total optical power dissipated within a certain region of the recording medium. Once a calculation has been performed for the electric field around the NFT and within the medium, it is straightforward to apply (5) to determine the dissipated power within any region of the computational space. This FOM does account for the wavelength dependence of the focused spot. For high-density HAMR storage, the bit cell will be smaller than 50 nm. Therefore, the NFTs considered in Sect. V will be evaluated on the basis of the percentage of the power in the incident beam that is dissipated within a circular area of 50 nm in diameter in the recording medium.
4. Temperature Rise in the Recording Medium

Finally, if the thermal properties of the recording medium are known, then thermal models may be applied to convert the dissipated power within the medium into a corresponding temperature profile. The FWHM size of the thermal spot and its peak temperature for a given input optical power to the transducer are directly related to the capability of the NFT for HAMR. Because this FOM depends on the specific thermal properties of the multilayer film stack in the recording medium, and these properties are often not known with precision, this FOM is of limited usefulness.

V. MECHANISMS FOR ENHANCEMENT OF THE FIGURE OF MERIT

As mentioned in Sect. I, there are several different mechanisms that operate in a well-designed NFT for enhancing the FOM. For the purposes of this article either the peak $|E|^2$ intensity in the medium or the dissipated power in the medium will be chosen as the FOM for studying these enhancement mechanisms for HAMR. Depending on the specific NFT design, the order of importance of these mechanisms may vary, but in general the best NFTs will combine most or all of these mechanisms. The ones we will consider in this section are localized surface plasmon resonance (LSPR), the lightning rod effect, and the dual-dipole effect.

1. Localized Surface Plasmon Resonance

Small metallic particles are well known to exhibit LSPRs. Surface plasmons are collective excitations of surface charge which under suitable conditions can be excited by an external optical field. Localized surface plasmons (LSPs) are oscillations of surface charge on a finite structure with fields that decay exponentially from the surface of the structure in both directions normal to the surface. The structure may be composed of a metal surrounded by a dielectric, or it may be composed of a dielectric surrounded by a metal. Examples include metallic nanoparticles and nanobubbles embedded in metals. Nanoholes in metal films also support LSPRs even though a hole is not entirely surrounded by the metal film. The surface plasmon resonance wavelength is determined by the size, shape, and material of the structure and the surrounding medium.
At resonance, the nanoparticles absorb the incident optical energy much more efficiently and generate enhanced electric fields at their surfaces from the oscillating surface charge. The enhanced absorption from LSPR of silver and gold nanoparticles embedded in glass has been used since medieval times to make stained glass windows with yellow and red colors. In Fig. 3 the extinction coefficient for a 60-nm gold sphere is shown calculated from Mie theory and graphed along with the electric field intensity at the surface of the particle. The LSPR is observed at approximately 550 nm by the peak in both the extinction coefficient and the field intensity at the surface. It should be noted that the peak $|E|^2$ field intensity at the surface of the sphere is more than 70 times larger than the field intensity of the incident plane wave. A plot of the field intensity in the neighborhood of the sphere is shown in Fig. 4.

The resonance wavelength of LSPs is determined in part by the refractive index of the surrounding medium. This is illustrated in Fig. 5 by plotting the peak field intensity for the 60-nm gold sphere versus wavelength for several different surrounding dielectrics. As the index of the dielectric increases, the resonance shifts towards longer wavelengths.
Figure 4. Contour plot of the field intensity in the xy plane of a 60-nm gold sphere embedded in a dielectric of index 1.5 when excited by a plane wave of unit amplitude which is polarized along the x axis at a wavelength of 550 nm. The points in the plane are computed with an increment of 1 nm.

Figure 5. Effect of the surrounding dielectric index on the surface plasmon resonance wavelength of a 60-nm gold sphere. Increasing the dielectric index shifts the resonance to longer wavelengths and enhances the peak field intensity.
Although a nanosphere is an excellent structure for illustrating the surface plasmon resonance enhancement of electric fields, it is not a particularly well designed NFT for HAMR. A better NFT for HAMR is the triangle antenna. This structure, which can also be considered to be a nanoparticle, exhibits a LSPR. A plane wave incident upon the triangle antenna and polarized along its length can drive surface currents back and forth along the antenna. For appropriate antenna dimensions, the antenna becomes a resonant structure of oscillating surface currents which is a LSPR. As previously stated, it is not possible to compute the resonance fields analytically or semianalytically for most NFT structures, which have much lower symmetry than spherical nanoparticles. Therefore, in this article such calculations were carried out with the scattered field FDTD numerical approach. In the FDTD calculation a plane wave of unit amplitude is incident onto the triangle antenna propagating in the $-z$ direction. A plot of the peak $|E|^2$ at the tip of a triangle antenna versus wavelength is shown in Fig. 6. The LSPR occurs at a wavelength of 775 nm. At this wavelength the peak field intensity, as computed

![Figure 6. Peak field intensity at the tip of a gold triangle antenna embedded in free space versus wavelength for excitation by an incident plane wave of unit amplitude polarized along the length of the antenna. The antenna has an apex angle of 45°, a length of 200 nm, and a thickness of 80 nm.](image-url)
Figure 7. Local field intensity for a gold triangle antenna with a length of 200 nm, radius of curvature at the apex of 20 nm, apex angle of 45°, and thickness of 80 nm. The incident plane wave of unit amplitude is polarized along the x axis. The FDTD cell size is (2.5 nm). The field intensity at this wavelength is plotted in the xy plane through the center of the antenna in Fig. 7, showing that the peak field intensity occurs at the edge of the apex of the antenna as would be expected from the lightning rod effect (to be discussed in the next section). A plot of the field intensity along the x axis through the center of the apex in Fig. 8 demonstrates the characteristic exponential decay of the field strength on either side of the edge of the antenna.

The LSPR is also affected by the dimensions of the nanoparticles. In Fig. 9, the peak intensity is plotted for the triangle antenna versus the length of the antenna. The moral of the story is that if NFTs are designed properly, their dimensions, their optical properties and those of the surrounding materials will all be chosen so as to maximize the field enhancement in the recording medium by operating at the resonance of the LSP. Although the isolated nanoparticles considered in this section give theoretical field intensity enhancements of over 2 orders of magnitude, it should
Figure 8. Field intensity computed along the $x$ axis of the triangle antenna in Fig. 6 showing the exponential decay characteristic of the field from surface plasmons. The decay for negative $x$ into the gold antenna is of course much faster than the decay into the surrounding dielectric.

Figure 9. Field intensity at the apex of the triangle antenna as a function of the antenna length computed for plane wave excitation at a wavelength of 775 nm.
be remembered that these values are not very relevant to HAMR. When the NFTs are in the presence of lossy metallic materials like the recording medium, the field within the medium is shielded and greatly reduced. Moreover, the power absorption of the medium greatly reduces the $Q$ of the resonance, leading to much smaller field enhancements.

2. Lightning Rod Effect

The lightning rod effect refers to the well-known fact that sharp metallic objects tend to generate very large localized fields.\(^{22,23}\) Electric field lines must terminate normally to the surface of a perfect conductor. This effect tends to concentrate the field lines at any sharp points of highly conducting materials.\(^{24}\) This is a shape effect, not a resonance effect, and therefore does not have any particular wavelength dependence. It may or may not be associated with a LSPR. For example, as a spherical gold nanoparticle is pulled into an ellipsoidal shape, the LSPR splits into resonances at two different wavelengths. One of the resonances shifts towards shorter wavelengths with increasing obliquity and one shifts towards longer wavelengths. The longer-wavelength resonance corresponds to surface charge oscillating along the long axis of the ellipsoid and it is found that the fields at the tips of the ellipsoid at the resonance get stronger as the end of the ellipsoid gets narrower and sharper.\(^{25}\) This effect is shown in Fig. 10. The lightning rod effect can generate extremely large field enhancements.

The triangle antenna also provides an excellent illustration of the lightning rod effect. In this case the FDTD technique is used to compute the fields at the apex of the antenna as the radius of curvature at the apex is varied. All calculations are carried out with a cell size of (2.5 nm).\(^3\) The results are graphed in Fig. 11. The peak field at the apex for this particular antenna design and within the accuracy of the FDTD calculation is somewhat smaller than the absolute peak field as can be seen from Fig. 7. Clearly it is beneficial to design the NFT with a sharp point(s) to both enhance the field intensity and localize it within the recording medium.

A contour plot of the field intensity for the antenna with a 5-nm radius of curvature is shown in Fig. 12 for comparison with the plot for a 20-nm radius of curvature in Fig. 7.

The strong effect on field enhancement of the lightning rod effect leads directly to a remark which, although obvious, nevertheless
Figure 10. Field enhancement at the tip of a prolate spheroid as a function of its aspect ratio.

Figure 11. Peak field intensity at the apex of a triangle antenna versus radius of curvature. The antenna is 200 nm long, 80 nm thick, with a 45° apex angle. The incident plane wave has a wavelength of 775 nm.
Figure 12. Local field intensity for a gold triangle antenna with a length of 200 nm, radius of curvature at the apex of 5 nm, apex angle of 45°, and thickness of 80 nm. The incident plane wave of unit amplitude is polarized along the x axis. The FDTD cell size is (2.5 nm).³

seems to be often neglected in the literature. In particular, the peak field intensity is also necessarily a function of the cell size used in the numerical simulation. It is well known that the electric field amplitude at the edge of a semi-infinite perfectly conducting straight edge has a logarithmic divergence.⁶ If this were modeled numerically with a finer and finer mesh, the peak field amplitude would be found to continuously increase. Therefore, when comparisons are made between different NFTs using numerical calculations of peak field amplitude, care should be taken to ensure that the same numerical algorithm and same cell size are being employed in the comparison. Otherwise the results are meaningless.

3. Dual-Dipole Resonance

A third technique for field enhancement is the dual-dipole effect. In this case, two resonant particles are brought close enough together to interact with each other. In the gap region between the two particles, the field can become much more intense than that from either
particle separately. As a simple example, we first consider the case of two 60-nm gold particles with an incident plane wave polarized along the axis connecting them. The peak field intensity is plotted versus wavelength for a 10-nm gap between the spheres in Fig. 13. There is clearly a resonance wavelength at 650 nm for excitation of the LSPs on the spheres. The field intensity distribution at this wavelength is plotted in Fig. 14. The peak field intensity of approximately 1,200 is in the region between the two spheres. The peak field intensity in the gap between the spheres is plotted versus gap distance in Fig. 15. The intensity falls very rapidly with increasing gap distance.

VI. COMPARISON OF NEAR-FIELD TRANSDUCERS

In this section the results of the previous two sections are combined to compare several NFT designs that have been suggested for use in data storage. In particular, the triangle antenna and the bow-tie antenna are compared with the circular aperture, the tapered rectangular aperture, the bow-tie aperture, and the C aperture. All NFTs are illuminated by a highly focused beam using a SIL with a refractive index of 1.5 to obtain an optical spot size with dimensions...
Figure 14. Field intensity distribution at resonance for the dual gold spheres showing the large field enhancement in the gap between the spheres.

Figure 15. Field intensity in the gap between dual gold spheres as a function of gap distance for a background dielectric with index 1.5 at a wavelength of 650 nm, and a background dielectric with index 1.0 at a wavelength of 525 nm.
of $0.49\lambda \times 0.38\lambda$ as calculated using the stationary-phase approximation in the Richards–Wolf theory. A simple recording medium consisting of 10 nm of cobalt laminated to a 100-nm gold heat sink is placed 7.5 nm below the NFT. The separation distance of the NFT from the medium is determined by several considerations. At terabit per square inch storage densities, the down-track distance between magnetic transitions is only approximately 10–15 nm. With such small spacing between transitions, it is necessary for the magnetic reader to fly extremely close to the surface of the medium. Moreover, the fields generated in the medium by the NFT are primarily evanescent fields. If the medium is spaced too far from the NFT, the amplitude of these fields is too small to couple power efficiently.

As previously demonstrated, an efficient NFT should make use of a LSPR effect. This effect requires a metallic surface that is highly conductive at optical frequencies. There are relatively few metals that satisfy this criterion. Silver and aluminum can support LSPs throughout the visible region. Gold and copper can support LSPs in the near infrared region and slightly into the red region of the visible spectrum. These elements and their alloys are the only reasonable choices for NFTs in device applications. However, pure silver, copper, and aluminum all have problems with corrosion. This leaves gold as the material of choice for the NFT and, therefore, gold is used for all NFT comparisons in this section.

The FOM for making the NFT comparisons is the peak field intensity within the recording medium. The FWHM of the spot size within the top layer of the recording medium is required to be 50 nm or less for a realistic HAMR storage device at terabits per square inch densities. The minimum dimension allowed within the NFT structure is 20 nm for all NFTs. This ensures that no NFT design is given an “unfair” advantage in the comparison by making use of the lightning rod effect to a greater extent than the other designs. The cell size in the FDTD calculations is (2.5 nm). With these restrictions, it is possible to make reasonable comparisons of the NFTs. However, it should be remembered that if the desired FWHM optical spot size within the transducer is specified, then a better FOM is the dissipated power within this area.

1. Circular Aperture

The circular aperture in an opaque film is the simplest NFT. It has traditionally been given a poor rating as a NFT based in large part on
the discouraging far-field transmittance of such apertures as found both theoretically and experimentally. However, as has been previously discussed, the far-field transmittance of an aperture does not necessarily correlate to its near-field power coupling efficiency. The geometry of this NFT is shown in Fig. 16.

The peak field intensity in the medium exhibits a LSPR and a maximum value at a wavelength of approximately 650 nm regardless of hole diameter as shown in Fig. 17. Unfortunately, there are two problems with this NFT. The peak field intensity within the medium is extremely small and the dissipated power within the medium spreads over an area that is much larger than the hole, as shown in Fig. 18. On the other hand, the dissipated power within a 50-nm-diameter cylinder in the medium is 0.14%, which is much larger than the value of $10^{-5}$ that might be expected for the far-field transmittance based on the theory of Bethe. By filling the hole with a high-index dielectric, one can reduce the optical spot size and increase the field intensity in the medium.

2. Tapered Rectangular Aperture

The efficiency of the circular aperture can be improved significantly by tapering the side walls. Moreover, because the circular aperture produces an oblong dissipated power spot along the direction of the incident polarization as shown in Fig. 18, it makes sense to widen...
Figure 17. Peak field intensity in the recording medium versus wavelength for circular apertures of various diameters. The gold film is 40 nm thick.

Figure 18. Field intensity within the recording medium for a circular aperture in a gold film with a 40-nm diameter and a thickness of 50 nm. (Reprinted from Ref. [7]. Copyright 2006 with permission from the Institute of Pure and Applied Physics.)
Figure 19. Tapered gold rectangular aperture. The aperture is filled with the glass of the solid immersion lens with refractive index 1.5.

Figure 20. Field intensity in the medium versus wavelength for the tapered gold rectangular aperture.

The peak field intensity within the recording medium as a function of wavelength is shown in Fig. 20 for several different aperture...
dimensions. The 20 nm × 40 nm rectangular aperture in the 50-nm-thick gold film with a 45° slope to the side walls generates the largest field intensity in the medium at a wavelength of approximately 650 nm. The field intensity within the medium is shown in Fig. 21. The total power dissipated in the central 50 nm of the medium is 0.92%. Moreover, the optical spot within the medium is smaller than the desired 50-nm FWHM. This is a substantial improvement over the air-filled circular aperture with straight side walls.

3. Bow-Tie Aperture

An aperture in the shape of a bow tie, also called a “bow-tie slot antenna” is shown in Fig. 22. This aperture is essentially a rectangular aperture with a constriction in the center. When it is illuminated with light polarized across the gap as shown in the figure, a LSPR is excited which oscillates surface charge into the two tips in the center. The sharpened tips enhance the field strength in this region via the
lightning rod effect. Moreover, the two tips separated by a small gap provide field enhancement via the dual-dipole effect. Therefore, all three field-enhancement mechanisms are present in this NFT.

With so many dimensions to specify for this aperture, the optimization process is lengthy. Variation of the length of the aperture with wavelength indicates an optimum length of 300 nm or greater although the LSPR wavelength is approximately 800 nm and only weakly dependent on aperture length. Variation of the aperture width gives a similar result for the optimum value and the optimum thickness is approximately 80 nm. As the gap is made narrower, the field intensity increases in the gap via the dual-dipole effect. There is some variation in efficiency with apex angle, but values in the range of 60°–90° are good.

The wavelength dependence of the peak field intensity in the medium is plotted in Fig. 23 for an aperture with a length of 300 nm, a width of 290 nm, a thickness of 80 nm, a gap of 20 nm, and an apex angle of 90°. There is a narrow LSPR at 725 nm. The field intensity in the medium at this wavelength is plotted in Fig. 24. The FWHM optical spot size in the medium is somewhat larger than the desired 50 nm. The percentage of power delivered to a 50-nm cylinder in the medium is 1.7%. This NFT is not as successful at confining the optical energy as some of the other designs.
Figure 23. Peak field intensity in the medium versus wavelength for a bow-tie aperture with dimensions given in the text.

Figure 24. Field intensity from the bow-tie aperture in the medium at a wavelength of 725 nm. (Reprinted from Ref. [7]. Copyright 2006 with permission from the Institute of Pure and Applied Physics.)
Next, the C aperture or ridge waveguide is considered. The C aperture was originally proposed by Shi et al., but they originally considered apertures in perfectly conducting metal films in the absence of a recording medium and for an incident plane wave. Although these calculations indicated 3 orders of magnitude greater field intensities from the C aperture than from a square aperture, these results are not directly relevant for HAMR. Many additional studies have been made which include the effects of real metals and focused incident beams. The C aperture is shown in Fig. 25.

The ridge waveguide is a well-known geometry for transporting microwaves. Like the bow-tie aperture, the C-aperture length can be less than the cutoff dimension for a rectangular aperture. For a rectangular aperture in which the incident field is polarized parallel to the short dimension, the field amplitude tends to zero at the short edges of the aperture and is maximum in the central region. The ridge in the center of the C aperture squeezes the field and thereby further enhances the field strength between the ridge and the opposite side. This can also be considered a dual-dipole effect, where the opposite side serves as an image surface to the ridge. For a C aperture in a real metal there is also a LSPR. Finally, the ridge itself enhances the field via the lightning rod effect. Therefore, this NFT also makes use of all the field-enhancement mechanisms. Propagating surface plasmon polaritons can be excited between the bottom of the SIL and the aperture. In principle these surface plasmons may siphon energy away from the LSP within the aperture, thereby reducing the coupling efficiency. However, with clever engineering these surface plasmons can actually be made to contribute additional energy to the LSP.

![Figure 25. Dimensions of the C aperture.](image-url)
Near-Field Optics for Heat-Assisted Magnetic Recording

Figure 26. Peak field intensity in the medium versus wavelength for the C aperture. The results are plotted for an aperture filled with air (n = 1) and an aperture filled with glass (n = 1.5).

There are again many dimensions to be optimized for this NFT. A length of approximately 300 nm is found to be near optimum. The LSPR occurs at approximately 700 nm and the width is optimized at 55 nm for a ridge that is 20 nm wide and has a gap of 20 nm. The optimum thickness is approximately 100 nm. The wavelength dependence of the field intensity in the medium is shown in Fig. 26. As the index of the material inside the aperture increases, the resonance is found to shift towards longer wavelengths. A plot of field intensity within the medium in Fig. 27 shows that the light is very well confined. This NFT delivers 2.1% of the incident power into the central 50-nm region of the recording medium.

5. Triangle Antenna

Antennas have also been proposed as NFTs for HAMR. The simplest antenna design may be the triangle, as shown in Fig. 28. The lightning rod effect was demonstrated in Sect. IV.2 for a triangle antenna in free space. This antenna also exhibits a LSPR. It does not
make use of the dual-dipole effect for field enhancement. When the antenna is adjacent to a lossy metallic recording medium, however, it behaves very differently. The LSPR wavelength is a very sensitive function of antenna length. A length greater than 150 nm places the resonance at wavelengths greater than 900 nm. A 100-nm antenna
has a resonance at a wavelength of approximately 800 nm. Confinement of the optical spot is difficult to achieve, however, with large apex angles, so an apex angle of 30° is chosen. The wavelength dependence of the peak field intensity in the medium is shown in Fig. 29 for a 100-nm-long antenna that is 50 nm thick. Although the LSPR occurs at 750 nm, the field intensity within the medium is not confined, as shown in Fig. 30. By operating the antenna at shorter wavelengths, one obtains better field confinement at the expense of field intensity, as shown in Fig. 31. The dissipated power within the medium at a wavelength of 650 nm is approximately 1.1%. However, the field intensity in the medium tends to spread out away from the tip of the antenna even at this wavelength.

It is interesting to compare these results with calculations of the triangle antenna in free space. Plots of the extinction, scattering, and absorption cross sections for the 100-nm triangle antenna in free space along with the peak field intensity at the apex are shown in Fig. 32. The resonance occurs at 675 nm, significantly shifted from the resonance wavelength in the presence of the medium. The peak field intensity occurs at the apex of the antenna, and is clearly
Figure 30. Field intensity in the medium from a 100-nm-long triangle antenna with a 30° apex angle at a wavelength of 750 nm. (Reprinted from Ref. [7]. Copyright 2006 with permission from the Institute of Pure and Applied Physics.)

Figure 31. Field intensity in medium from a 100-nm-long triangle antenna with a 30° apex angle at a wavelength of 650 nm.
Figure 32. Extinction, scattering, and absorption cross sections normalized by the area of the antenna for the 100-nm triangle antenna on a glass substrate as a function of wavelength. The peak $|E|^2$ field intensity versus wavelength is also plotted.

Figure 33. “Beaked” triangle antenna.

not useful for predicting the distribution of dissipated power in the medium. This clearly exhibits the unreliability of using peak field intensity for an antenna or aperture in free space as a FOM for HAMR.

One way in which the problem of lack of confinement of the coupled power to the medium can be solved is to cant the antenna so that only the tip is close to the medium. Another approach is to add a small “beak” at the end of the antenna as shown in Fig. 33.
If a beak with a 20-nm width, length, and height is added to the triangle antenna, then the resonance wavelength is slightly shifted to 725 nm, but the field intensity within the medium at resonance is much better confined, as shown in Fig. 34. The dissipated power in a 50-nm cylinder in the medium is 2.9%.

6. Bow-Tie Antenna

The bow-tie antenna was first proposed as a NFT by Grober et al.\textsuperscript{36} In the microwave frequency range, the bow tie is a well-known antenna design. As shown in Fig. 35, the antenna is composed of two triangular metallic plates with a narrow gap between them. This NFT is the complement of the bow-tie aperture. All three NFT enhancement mechanisms are clearly present in the design. Optimizing the antenna design proceeds along lines similar to those for the triangle antenna. An antenna that is 200 nm long, 50 nm thick, with a 20-nm gap, 20-nm apex width, and 30° apex angle exhibits two LSP resonances at approximately 625 nm and 750 nm as shown in Fig. 36.
As in the case of the triangle antenna, however, the long-wavelength resonance with the highest field intensity corresponds to an unconfined spot, as shown in Fig. 37. The shorter-wavelength resonance, on the other hand, does generate a small spot in the medium, as shown in Fig. 38, and delivers 1.9% of the incident optical power to the medium at a wavelength of 625 nm. Again, to obtain a confined spot at the peak of the resonance curve, the bow-tie antenna
Figure 37. Field intensity in the medium for a bow-tie antenna at a wavelength of 750 nm.

Figure 38. Field intensity in the medium for a bow-tie antenna at a wavelength of 625 nm.
can be canted so that only the high-field region between the tips is in close proximity to the medium. A 20° cant of the two antenna halves generates a much smaller spot at the resonance wavelength of 750 nm, as shown in Fig. 39. The canted bow tie delivers 2.1% of the incident power into the central 50 nm of the medium.

VII. ANTENNA AND APERTURE RELATIONSHIP

In the earlier sections we considered different near-field structures. These structures were one of three types: apertures, antennas, and hybrid structures. The apertures have a finite dielectric opening in a metal thin film. The resonant near field of interest is located within and in the vicinity of the opening. Antennas are finite metallic structures located in an infinite dielectric region. The resonant near field of antennas is located around the metallic structure. Then there are hybrid structures such as a metal-coated, tapered optical fiber, which on one hand do not have an aperture and on the other have metal going to infinity. In such hybrid structures, the fact that the metal
goes to infinity is not important. In fact, the resonant field of interest is in a geometrically localized region around the metal – the region around the tip in the case of the tapered fiber. If the metal is terminated at a certain distance (a distance on the order of the decay length of the associated surface plasmons), the field in the geometrically localized region does not change. Thus, these hybrid structures can be converted into an aperture or an antenna structure without considerably altering the physics of the near field. Thus, we assume that all the near-field structures of our interest are either of the aperture or of the antenna type. The calculation of the cross sections goes along different lines for the two types. Hence, this classification is needed.

If we interchange the dielectric in the aperture opening and the thin film metal, we get a complementary structure, which is an antenna. Is there any relation between the resonance properties of the two structures? If we assume that the aperture metal film is infinitesimally thick, and that the metal is a perfect electrical conductor (PEC), the aperture and the complementary antenna structure are connected by a form of Babinet’s principle. Suppose that the aperture is illuminated by an incident electric field $\vec{E}_i$. The interaction with the aperture will set up a total electric field $\vec{E}_1$. Now, suppose that the complementary antenna structure is illuminated by an incident magnetic field that is vectorially equal to the incident electric field in the aperture case. Thus, the incident magnetic field in the antenna case is $\vec{H}_i$. In this case, let $\vec{H}_2$ be the total magnetic field. The particular form of Babinet’s principle states that

$$\vec{E}_1 + \vec{H}_2 = \vec{E}_i.$$  \hspace{1cm} (7)

Of course, in the case of a real metal film that is not infinitesimally thick, the principle is not expected to be perfectly satisfied.

**VIII. NEAR-FIELD AND FAR-FIELD RELATIONSHIP**

It is difficult to design experiments to characterize the near field of the structures. Any probe such as the scanning near field optical microscope, which probes the near field directly, could end up altering the structure of the near field. This could have an effect of shifting the wavelength of the desired resonance. Experiments that account for the light radiated in the far field can also be designed.
But then, is the amount of far-field radiation a good measure of the near-field enhancement? We discuss this connection in the following section.

The source of radiation, in the classical electromagnetic theory, is an accelerated charge. For time-harmonic fields, electrical current serves as the source. There is a considerable amount of literature on the radiation properties of apertures and antennas at radio and microwave frequencies. At these frequencies, the penetration of the fields into a metal is small. Thus, it is frequently quite acceptable to model these structures by assuming the metals are PECs. At optical frequencies, a significant portion of the incident energy can be dissipated in the metal. In addition, typical metals exhibit surface plasmon resonances at optical frequencies. Associated with a surface plasmon is an oscillating charge distribution on the surface of the structure, localized within the skin depth of the metal.

In the absence of sources outside a closed surface, the tangential electric and magnetic fields on the surface uniquely define the field distribution outside the surface. In particular, the tangential fields can be interpreted as electric and magnetic currents on the surface. The equivalent currents replace the physical sources. The fields generated by the physical and the hypothetical sources are the same outside the surface. Inside the surface the field due to the hypothetical sources is zero. This theorem is used to calculate the far-field radiation pattern from the near-field FDTD simulation. We apply the theorem in the special case of the region outside the surface being a homogeneous dielectric medium. In FDTD simulations the infinite domain is converted into a finite computational domain using matching boundary conditions. This is true even in the case of stratified media (e.g., thin film structures) that extend to infinity. Thus, in the rest of the discussion, we assume that the domain of interest is infinite.

Figure 40 shows a scatterer embedded in a hypothetical closed surface. The surface currents are defined on the closed surface. The equations that connect the surface currents to the tangential fields and govern the radiation from the currents can be found in the popular FDTD texts.

The Poynting vector has units of power per unit area. When the normal (outward) component of the Poynting vector is integrated over a closed surface, it represents the electromagnetic power leaving the surface. For monochromatic fields the Poynting vector oscillates harmonically about a direct-current offset. The frequency of the
oscillation is twice that of the oscillating fields. The average value of the Poynting vector over an oscillation period is the measure of the net power flow across the surface in one direction. In lossless regions, the divergence of the Poynting vector is zero. Thus, in accordance with the Gauss divergence theorem, if we choose a closed surface in a lossless region, the surface integral of the energy flux is zero. The arbitrariness in the choice of the surface in the equivalence theorem does not change the net energy flux through the surface. Let the time-harmonic electric and magnetic field (at frequency $\omega$) at a point be given by

$$\vec{E} = \vec{E}_o \exp(-i\omega t)$$

and

$$\vec{H} = \vec{H}_o \exp(-i\omega t),$$

respectively. The vector quantities $\vec{E}_o$ and $\vec{H}_o$ contain the amplitude and phase information, and are hence complex. The time-averaged Poynting vector is given by

$$< \vec{S} > = \frac{1}{2} \text{Re} \left( \vec{E}_o \times \vec{H}_o^* \right).$$

Here, $\text{Re}$ and $^*$ stand for the real part and complex conjugation, respectively. In the rest of the discussion we will only be interested in
the time-averaged Poynting vector for monochromatic radiation. The only aspect of (10) that we carry forward is the linear dependence of the Poynting vector on the electric and magnetic fields. Hence, we simplify the notation by dropping the angular brackets and the subscript $o$ on the fields. We represent the bilinear form by

$$\vec{S} = (\vec{E}, \vec{H}).$$  \hspace{1cm} (11)$$

Henceforth, Poynting vector refers to the time-averaged Poynting vector.

The fields $\vec{E}$ and $\vec{H}$ can be decomposed into the incident field (indicated with subscript $i$) and the scattered field (indicated with subscript $s$). The incident field is the field that would have been present if the scatterer were absent. This assumes that the optical source excitation is the same. We have been vague in our definition of the scatterer. To be specific, we choose a geometrical arrangement as our starting point. This is the incident geometry, and the field is the incident field. We then alter the geometry. The change is small enough so that the optical source can still be assumed to be unperturbed. In particular, the change that we make would either be placing a microscopic particle (the antenna) in the geometry, or punching a hole in a metal film (the aperture). The difference between the field in the changed geometry and the field in the incident geometry is defined as the scattered field. In the context of the equivalence theorem, the change that we make is done inside the hypothetical surface. No matter how the incident geometry is defined, we assume that the region outside the hypothetical surface is lossless and homogeneous. With this decomposition of the fields, the Poynting vector is given by

$$\vec{S} = \vec{S}_i + \vec{S}_s + \vec{S}_c,$$  \hspace{1cm} (12)$$

where

$$\vec{S}_i = (\vec{E}_i, \vec{H}_i),$$  \hspace{1cm} (13)$$

$$\vec{S}_s = (\vec{E}_s, \vec{H}_s),$$  \hspace{1cm} (14)$$

and

$$\vec{S}_c = (\vec{E}_i, \vec{H}_s) + (\vec{E}_s, \vec{H}_i).$$  \hspace{1cm} (15)$$

$\vec{S}_i$ and $\vec{S}_s$ are the Poynting vectors of the incident and the scattered field, respectively. $\vec{S}_c$ is an interference term. In a homogeneous dielectric, $\vec{S}$, $\vec{S}_i$, and $\vec{S}_s$ are divergenceless; hence, $\vec{S}_c$ is also
The total energy flux from the volume inside the hypothetical surface is the integral of the component of the Poynting vector along the outward normal, over the closed surface. Let us represent this integral by $I$ and the integral of the three terms on the right-hand side by $I_i$, $I_s$, and $I_c$. Thus,

$$I = \iiint_A \vec{S} \cdot d\vec{A},$$

and

$$I = I_i + I_s + I_c.$$ 

Here, $A$ represents integral over the closed surface. If the closed surface is distorted such that the volume swept in distorting the surface is in a homogeneous dielectric medium, then owing to the divergenceless property, the integrals $I$, $I_i$, $I_s$, and $I_c$ remain invariant.

1. Radiation from Antennas

Let the incident geometry be the infinite free space and the incident field be a plane wave. We use spherical polar coordinates such that the polar and the azimuthal angles are denoted by $\theta$ and $\phi$, respectively. The polar angle is measured with respect to the $+Z$ axis. The azimuthal angle is measured with respect to the $+X$ axis in the $XY$ plane. The incident plane wave is propagating along the $\theta = 0^\circ$ direction ($+Z$ direction), and the polarization of the incident beam is along the ($\theta = 90^\circ$, $\phi = 0^\circ$) direction ($+X$ direction). Since the Poynting vector is divergenceless in this medium, $I_i = 0$.

In the far field, the radiation field in a certain direction appears locally like a plane wave propagating in that direction. The plane wave in a particular direction can further be decomposed into two mutually orthogonal polarizations. An analysis using Green’s function indicates that only the plane wave propagating in the same direction as the incident wave and possessing the same polarization can contribute to the term $I_c$.\(^{15}\) Thus, $I_c$ is proportional to the appropriately polarized scattered field radiation in the direction of the incident beam. Energy conservation considerations indicate that $I$ is precisely the negative of the power being absorbed inside the hypothetical surface, averaged over a field oscillation. We denote this quantity by $I_a$. In fact, $I_s$ and $I_a$ divided by $|\vec{S_i}|$ are called the scattering and the absorption cross sections, respectively. Their sum, the
total or “extinction” cross section, is thus directly proportional to the field strength of the scattering in the direction of the incident wave. This is more commonly known as the “optical theorem.” A cross section has physical dimensions of area, and can be normalized by the physical cross section of the antenna to obtain a dimensionless normalized cross section. If we consider all possible slices of the antenna normal to the propagation direction, the physical cross section of the antenna is the area of the slice with the largest cross section.

2. Radiation from Apertures

Consider a plane polarized wave incident normally on a metal film of finite thickness. The incident energy is transmitted across the film, reflected back, or absorbed in the film (see Fig. 41). Considering that the plane wave is infinite in extent, each of the three energy contributions is infinite. Let the field distribution in the presence of the film be termed the “incident field.” Let us now etch an aperture of finite cross section in this film. Let the difference of the field after and before etching the aperture be termed the “scattered field.” We follow an analysis similar to the case of the antennas. The aperture is the source of the scattered field. Owing to the loss in the metal, this scattered field decays inside the metal with increasing lateral distance from the aperture. Since the metal film is infinite in its plane, the hypothetical surface used in defining the equivalent currents has to wrap around the metal at infinity. We define the surface (see Fig. 42) to be $S_1 - S_2 - S_3$ on one side, and $S_4 - S_5 - S_6$ on the other. The expression for the power flux, (17), is applicable here. However, the

Figure 41. Incident geometry.
key difference from the antenna case is that \( I_i \) is no longer zero. In fact, the principle of energy conservation implies that \( I_i \) and the net power absorbed inside the closed surface before etching the aperture, \( I_{af} \), add up to zero. Note, both are infinite quantities. Also, \( I \) and the net power absorbed inside the closed surface after etching the aperture, \( I_{aa} \), add up to zero. Hence,

\[ I_s + (I_{aa} - I_{af}) = -I_c. \]  

The scatterer (aperture) is finite in extent. Moreover, owing to the loss in the metal, fields decay in the film away from the aperture; hence, \( I_s, I_{aa} - I_{af} \), and \( I_c \) are finite quantities. The definition of the scattering cross section is analogous to the antenna case. However, in the definition of the absorption cross section, we replace \( I_a \) with \( I_{aa} - I_{af} \). In the case of the antenna, the term \( I_c \) was stated to be directly proportional to the radiation intensity in the forward direction. For the aperture, a Green's function analysis similar to the antenna case indicates that the contribution to \( I_c \) from the surface \( S1 - S2 - S3 \) is proportional to the radiation intensity in the forward direction. Similarly, the contribution from the surface \( S4 - S5 - S6 \) is proportional to the radiation intensity in the backward direction. Thus, \( I_c \) is a linear combination of the radiation intensity in the forward and backward directions.

To calculate the radiation pattern of the apertures using FDTD, we need to define the equivalent currents on the hypothetical surface. To overcome the difficulty of dealing with an infinite surface, we choose the closed surface to be \( S2 - S7 - S5 - S8 \). The assumption is
that the surfaces S7 and S8 are chosen far away from the aperture, so that the scattered field is negligible on them. Our claim is as follows: it is possible to choose surfaces S1, S3, S4, and S6 infinitesimally close to the metal film surface and surfaces S7 and S8 sufficiently far from the aperture, such that the scattered field on the surfaces S1, S3, S4, S6, S7, and S8 is infinitesimally small. This is possible owing to the dissipation in the metal. Thus, the equivalent currents are essentially present only on S2 and S5. Thus, in the FDTD code, the radiation pattern can be calculated exactly as in the antenna case – by using the closed surface S2 – S7 – S5 – S8.

3. Numerical Modeling

We apply the concepts discussed in the last few sections to the case of a C aperture in aluminum. The thickness of the aluminum film is chosen to be 100 nm. The dimensions of the C aperture are as follows: aperture length 155 nm, aperture width 70 nm, tongue width 25 nm, and gap width 25 nm. The incident field is X-polarized. The XZ plane is a mirror symmetry plane for the C aperture. The surrounding dielectric is assumed to be free space. The normalized scattering and absorption cross sections as a function of wavelength are shown in Fig. 43.

Figure 43. Cross sections of the C aperture (normalized by the area).
The near-field intensity is calculated at a point, in the gap, 5 nm beyond the transmission side of the aperture. The cross sections are normalized with respect to the physical area of the aperture. For comparison with the near-field intensity, we normalize the cross sections such that the peak cross section is unity. To distinguish this from the area normalization, we call this the “magnitude normalization.” The cross sections and the near-field intensity are shown in Fig. 44.

The three quantities have been magnitude-normalized in this plot. Geometrically, the C aperture is a ridge waveguide of finite extent. For a PEC waveguide of the same cross section, the cutoff wavelength for the lowest-order transverse electric mode is around 500 nm. As one approaches the cutoff wavelength from shorter wavelengths, the longitudinal wave vector decreases in magnitude. Hence, for the same length of the waveguide, the field has a larger number of transverse traversals in the aperture. On the other hand, if one moves away from the cutoff wavelength towards longer wavelengths, the longitudinal wave vector becomes imaginary, indicating evanescent decay. Hence, the strongest resonance is expected to be at the cutoff wavelength. Three things about the aluminum C aperture are different from the PEC waveguide: the metal can support surface plasmons, the incident field (field in the metal film before the aperture is etched) has a Fabry–Perot resonance, and leaky modes that
are not seen in an infinite waveguide can be excited in the case of a waveguide whose length is a fraction of the wavelength. One or more of these effects could cause a shift in the resonance wavelength. In fact, we observe a resonance at approximately 650 nm. The radiation pattern of the aperture at a wavelength of 650 nm is plotted in Fig. 45. The corresponding near field intensity is shown in Fig. 46.

An electric dipole is induced in the gap of the C aperture. The far-field radiation pattern of the dipole is expected to have a
doughnut shape with \( X \) as the cylindrical symmetry axis. However, the presence of the infinite metal in the \( xy \) plane is expected to quench the radiation pattern in that plane. This would cause a pinching of the radiation pattern in the \( xy \) plane. This is seen in Fig. 45.

To see the effect of the medium, we place a cobalt film 5 nm from the aperture. The magnitude-normalized near-field intensity with and without the medium is plotted in Fig. 47. A considerable shift in the resonance wavelength is seen. Thus, the medium loads the C aperture.

We then consider the antenna structure complementary to the C aperture – the C antenna. Even though we do not have a PEC antenna, we would like to test the agreement with Babinet’s principle. Instead of rotating the polarization of the incident beam by 90\(^\circ\), we rotate the antenna structure by 90\(^\circ\). The normalized cross sections of the C antenna are shown in Fig. 48. The resonance wavelength is the same as that of the C aperture; however, the scattering cross section is much more enhanced in this case. If we assume that a resonance enhancement of the electric field has an associated enhancement in the magnetic field, and vice versa, then Babinet’s principle suggests that the complementary structure should also have a resonance in the same spectral region. An enhanced magnetic field of opposite phase

![Figure 47. Effect of the medium on the near-field resonance of a C aperture.](image_url)
would be needed to nullify the enhancement in the electric field of the complementary structure.

The radiation pattern of the antenna at a wavelength of 650 nm is plotted in Fig. 49. This radiation pattern shows the cylindrical symmetry of the doughnut-shaped dipolar radiation pattern.
IX. PHOTONIC NANOJETS

Up to this point we have been considering NFTs for applications such as HAMR. The transducer itself, however, is only one part of a complete device for recording. The transducer will not be effective unless it is situated at the position of a large field amplitude from the incident laser beam. This is generally accomplished by focusing the beam onto the transducer. A simple objective lens may be quite satisfactory for this purpose. In this section we discuss techniques for highly concentrating an incident beam into a “nanojet,” i.e., a narrow beam of energy with an extended path length. In principle, nanojet optics could form one part of the complete system for near-field recording.

In the geometrical optics description of conventional lens focusing, the focus is the point where all the light rays converge. In Fig. 50, the focusing of a plane wave by a lens is shown.

The focal point is situated in the middle of a sphere. If the refractive index of the sphere is greater than unity and the sphere is truncated to a hemisphere (part on the right of the dashed line is removed), we end up with a SIL. The key feature of this geometry is that all the rays converge to a point – the focal point. If, instead, the lens is removed from the system such that all parallel incident rays fall directly on the sphere (Fig. 51), then all the rays will not converge to a single point.

![Figure 50. Focusing by a lens.](image-url)
Nonetheless, for an appropriate choice of the sphere refractive index, there exists a surface such that several rays converge at every point of the surface. In other words, the focal point degenerates into a surface. This surface is termed a “caustic.” Owing to the symmetry of the problem, the caustic has a cylindrical symmetry. The caustic has a cusp at the point where the caustic intersects the symmetry axis. Our discussion so far has been based on geometrical optics. When one goes to a complete electromagnetic description, the focal point of a lens does not have a field singularity. Nonetheless, there is a focal region of high field concentration. Similarly, for the caustic one ends up with a region of high field concentration in the neighborhood of the geometrical caustic. In addition, in the electromagnetic description, the wavelength adds a length scale to the phenomenon. Thus, for a fixed radius and refractive index of the sphere, the caustic region will depend on the wavelength of light (in free space).

In Fig. 51, the cusp of the caustic is shown to lie inside the sphere. In such a situation, the cusp can be pushed to the surface of the sphere by reducing the refractive index of the sphere. In the geometrical optics description, this will happen when the refractive index of the sphere is twice that of the surrounding medium (assumed to be free space here). In the physical optics description, the choice of the refractive index ratio depends on the ratio of the radius of the sphere to the wavelength. Typically, it is found to be smaller than 2. When the cusp region is chosen close to the sphere surface, an interesting phenomenon of “photonic nanojet” emerges. On the free space side of the surface, an intense optical-jet-like region is generated. A two-dimensional FDTD model of this phenomenon is shown in Fig. 52.
The photonic nanojets display two remarkable features. Even though the spot size is comparable to a high-numerical-aperture diffraction-limited SIL spot, the depth of focus is much larger. For a comparable spot size from a lens, the depth of focus would have been much smaller. Secondly, from Fig. 52, the decay length of the two-dimensional photonic nanojet (distance between the peak field and the $1/e$ field in the longitudinal direction) is larger than the wavelength. The spot size of the nanojet at its waist is marginally larger than the size of a spot generated by a two-dimensional lens of unit numerical aperture\(^{38}\) (see Fig. 53).

The angular spectrum (spatial frequency content) of the photonic nanojet is shown in Fig. 54. The amplitude distribution of the angular spectrum alone does not explain the long decay length of the photonic nanojet. The bathtub-shaped phase distribution plays a key role. Different spatial frequencies gain different phases on propagation. This dephasing causes spot divergence. The bathtub shape counteracts the typical dephasing factor that decreases with increasing magnitude of the spatial frequency.

When a nanoparticle is placed in the light path, light is scattered. Assuming that the incident light is a plane wave, the light that radiates back towards the source is termed the “backscattered light.” When nanoparticles are illuminated by a plane wave, the intensity of backscattered light is small compared with the intensity
Figure 53. Comparison of the two-dimensional nanojet spot size with that of lens focusing. $k$ is the wave vector in free space. (Reprinted from Ref. [38]. Copyright 2005 with permission from the Optical Society of America.)

Figure 54. Angular spectrum of a two-dimensional nanojet. $s$ and $k$ are the spatial frequency and the free-space wave vector, respectively. (Reprinted from Ref. [38]. Copyright 2005 with permission from the Optical Society of America.)

of the incident light. If instead, a lens is used to focus light onto the nanoparticle, the backscattered light intensity increases by a few orders of magnitude. However, if the nanoparticle is placed in the photonic nanojet, the backscattering increases by several orders of magnitude.\textsuperscript{39} In the two-dimensional case, the effect is still seen, but it is not as pronounced. The enhanced backscattering for the two-dimensional case is shown in Fig. 55. The effect of the particle size on the backscattering enhancement is shown in Fig. 56.
Figure 55. Differential cross section of a particle placed in a nanojet. (Reprinted from Ref. [39]. Copyright 2004 with permission from the Optical Society of America.)

Figure 56. Backscattering enhancement in the nanojet as a function of particle size. (Reprinted from Ref. [39]. Copyright 2004 with permission from the Optical Society of America.)
Chen et al. have argued that while the large intensity of the nanojet provides a lenslike enhancement of the backscattering, it is the coordination between the backscattering and the modes of the nanojet-creating sphere that generates the superenhancement of the nanojet.

X. CONCLUSION

A variety of mechanisms have been discussed for enhancing the efficiency of NFTs for use in HAMR. These include the LSPR effect, the lightning rod effect, and the dual-dipole effect. Several common FOMs for NFTs have been discussed and it has been shown that peak $|E|^2$ field intensity within the recording medium, or even better, the dissipated power within the recording medium are the best FOMs. On the other hand, far-field transmittance or even peak field amplitude in the absence of a recording medium are not useful for judging the merits of NFTs for HAMR. Several transducer designs have been analyzed theoretically and compared using a standard geometry that approximates the situation found in HAMR. The results are summarized in Table 1. Surprisingly large power coupling efficiencies can be obtained theoretically for the best transducer designs, lending credibility to the engineering challenge of building such a data storage device.

Our study of the C aperture indicates that the resonance wavelength for an aperture of finite length can be shifted from the cutoff

Table 1. Summary of near-field transducer (NFT) performance. The peak $|E|^2$ intensity is normalized by that of the incident beam.

| NFT design          | $\lambda_{res}$ | Peak $|E|^2$ | $P_{diss}$ | FWHM spot size (nm$^2$) |
|---------------------|----------------|-------------|------------|------------------------|
| Circular aperture   | 650            | 0.07        | 0.14%      | 113 $\times$ 142       |
| Rectangular aperture| 650            | 0.80        | 0.92       | 43 $\times$ 25         |
| Bow-tie aperture    | 725            | 1.38        | 1.7        | 59 $\times$ 56         |
| C aperture          | 700            | 2.42        | 2.1        | 34 $\times$ 39         |
| Triangle antenna    | 650            | 0.77        | 1.1        | 55 $\times$ 54         |
| Beaked triangle     | 725            | 2.82        | 2.9        | 43 $\times$ 41         |
| Bow-tie antenna     | 650            | 1.41        | 1.4        | 39 $\times$ 36         |
| Canted bow tie      | 750            | 2.61        | 2.1        | 31 $\times$ 36         |

*FWHM* full width at half maximum
wavelength of the corresponding PEC waveguide, especially at optical frequencies. The far-field cross sections and the near field intensity have resonances at around the same wavelength. The strong currents associated with the near-field enhancement are also responsible for the absorption and far-field radiation. It might be possible to come up with a current distribution of certain orientation and phase relationship such that the far-field radiation is small. Whether there is a geometry in which this current distribution can be excited by a plane wave is an open question. Complementary aperture/antenna systems that seem to be resonating in completely different modes can still have similar resonance properties in accordance with Babinet’s principle.

Finally, we have briefly considered one interesting optical technique for exciting the NFT via a nanojet.

ACKNOWLEDGMENTS

We would like to acknowledge many useful conversations with our Seagate colleagues Ed Gage, Eric Jin, Terry McDaniel, and Chubing Peng during the course of this work.

REFERENCES

15. S. Nie and S. R. Emory, Science 275 (1997) 1102.
Modern Aspects of Electrochemistry No. 44
Modelling and Numerical Simulations II
(Ed.) M. Schlesinger
2009, XXI, 518 p. 221 illus., Hardcover