Lowering HAMR Near-Field Transducer Temperature via Inverse Electromagnetic Design

Samarth Bhargava and Eli Yablonovitch, *Fellow, IEEE*

Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley, CA 94709 USA

Heat-assisted magnetic recording (HAMR) allows for data writing in hard disks beyond 1 Tb/in^2 areal density, by temporarily heating the area of a single bit to its Curie temperature. The metallic optical antenna or near-field transducer (NFT), used to apply the nanoscale heating to the media, may self-heat by several hundreds of degrees. With the NFT reaching such extreme temperatures, demonstrations of HAMR technology experience write-head lifetimes that are orders of magnitude less than that required for a commercial product. Hence, heating of the NFT is of upmost importance. In this paper, we first derive fundamental limits on the temperature ratio NFT/Media to drive NFT design choices for low-temperature operation. Next, we employ inverse electromagnetic design software, which solves for unexpected geometries of the NFT and waveguide. We present computationally generated designs for the waveguide feeding the NFT that offer a 50% reduction in NFT self-heating (∼220 °C) compared with typical industry designs.

Index Terms—**Adjoint method, computational electromagnetics, gradient methods, hard disks, heat-assisted magnetic recording (HAMR), inverse problem, nanophotonics, near-field transducer (NFT), optical antenna, optimization, plasmonics, thermal management.**

I. Introduction

TO ENSURE thermal stability of data over a 10-year lifetime in hard disks of beyond 1 Tb/in^2 areal density, the magnetocrystalline anisotropy of the magnetic granular media must be increased, while scaling the magnetic grains to smaller dimensions [1]. Recording information to such a medium is a monumental challenge. The current state-of-the-art employs writing electromagnets that are already limited by magnetic field saturation of permeable metals, placing an upper bound on the strength of magnetic field that can be applied during the recording process. Heat-assisted magnetic recording (HAMR) promises writing to highly anisotropic media, by temporarily heating the area of a single datum to its Curie temperature, while simultaneously applying a magnetic field from a conventional electromagnet [2], [3]. In practice, a metallic optical antenna or near-field transducer (NFT) focuses light onto the highly absorbing magnetic recording layer in the media and locally heats a 30 nm $\times$ 30 nm spot on the medium to near 700 K [4], [5]. Since the metal comprising the NFT, typically gold, is itself absorbing at optical frequencies, the NFT also heats by several hundreds of degrees [6]. This NFT self-heating is a significant cause of failure in HAMR systems and limits the lifetime of today's prototype HAMR write heads to be orders of magnitude less than the desired 10-year lifespan [6]. Hence, an important figure of merit (FOM) for reliability in an HAMR write head is the temperature ratio between the media hotspot and NFT.

The difficulty in designing a low-temperature HAMR write head is twofold: 1) the fundamental limits on the NFT temperature are not well-known and 2) designing the light delivery system that produces nanoscale heating requires understanding the complex electromagnetic interactions of the illuminating waveguide, metallic NFT, magnetic write pole, and multi-layered hard disk medium. On the first note, we derive in this paper, a simple analytic model for the ratio of temperature rise in the NFT to the temperature rise in the media hotspot. This provides important limits and constraints on the structural design of the NFT that must be satisfied for low-temperature operation. On the second note, because of the wave nature of light, the optimal shapes of electromagnetic structures are often unexpected [7]. Traditional design approaches, based on intuition or highly constrained optimization, such as parameter sweeps, are inadequate. In this paper, we propose inverse electromagnetic design software, which provides fast optimization of 3-D electromagnetic structures with thousands of degrees of freedom. With such a large parameter space, an optimization can search for unexpected shapes of the NFT, or the feeding waveguide that offers superior performance for HAMR. The drawback of such an optimization is computational expense. For applications like HAMR, a single 3-D Maxwell simulation of nanoscale metallic structures and a multi-layered medium is computationally demanding even on modern high-performance computing clusters. Since heuristic algorithms like particle swarm and genetic algorithms rely on random trials, they are too computationally burdensome for practical engineering design for optics in the nanoscale. In contrast, the inverse electromagnetic design software performs gradient-based optimization using the adjoint method, which results in fast deterministic optimization that is computationally inexpensive [8]–[10].

In this paper, our strategy toward achieving a low-temperature HAMR write head is to make major NFT design choices based on a simple analytic expression for the NFT/media temperature ratio. Then, we use our inverse electromagnetic design software to find unexpected shapes of

Manuscript received July 18, 2014; revised August 22, 2014; accepted August 26, 2014. Date of current version May 18, 2015. Corresponding author: S. Bhargava (e-mail: samarth@eecs.berkeley.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMAG.2014.2355215

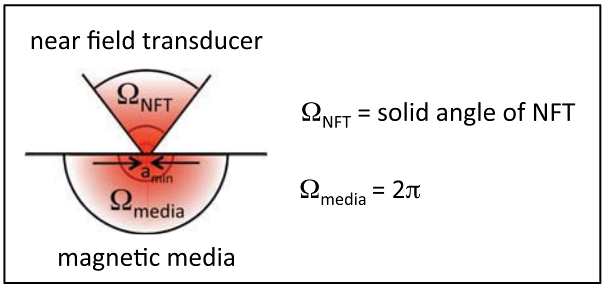

Fig. 1. Model of spherical heat conduction in a hemispherical media and conical NFT.

the waveguide feeding the proposed NFT design to provide the desired optical performance.

II. Media/NFT Temperature Ratio

To derive the temperature ratio between the hard disk media and NFT, one may start with a simple model of spherical heat conduction from heat sources due to optical absorption in the medium and the tip of the NFT. As shown in Fig. 1, we approximate the metallic NFT as a cone and the multi-layered media stack as a homogenous hemisphere. It is also assumed that all the significant optical absorption is in the NFT tip and in the media hotspot.

Relative to ambient temperature infinitely far away, the temperature rise in the NFT tip and media hotspot is described by

$$\Delta T_{\mathrm{NFT}} = \frac{P_{\mathrm{NFT}}}{\Omega_{\mathrm{NFT}} \times K_{\mathrm{NFT}} \times a_{\min}} \tag{1}$$

$$\Delta T_{\mathrm{media}} = \frac{P_{\mathrm{media}}}{\Omega_{\mathrm{media}} \times K_{\mathrm{media}} \times a_{\min}} \tag{2}$$

where P is the heat generated in the NFT tip or media, Ω is the solid angle, K is the thermal conductivity, and $a_{\min}$ is the minimum diameter of the NFT tip. It is assumed that the diameter of the hotspot in the media is also $a_{\min}$. Next, we relate the heat generated to the optical absorption in these respective regions by

$$P_{\mathrm{NFT}} = \frac{1}{2}\omega\varepsilon_0\ \varepsilon''_{\mathrm{NFT}}\,|\mathbf{E}_{\mathrm{NFT}}|^2 \tag{3}$$

$$P_{\mathrm{media}} = \frac{1}{2}\omega\varepsilon_0\ \varepsilon''_{\mathrm{media}}\,|\mathbf{E}_{\mathrm{media}}|^2 \tag{4}$$

where ω is the frequency of the excitation laser light, ε_0 is the free-space permittivity, ϵ'' is the imaginary part of the permittivity, and $|E|^2$ is the light intensity in the NFT tip or media. We can estimate the ratio of light intensity in the NFT tip to media according to electromagnetic boundary conditions at the NFT-air-media interface. Assuming the electric field is perpendicular to the interface, the light intensities in the NFT and media are proportional as shown in

$$|\epsilon_{\mathrm{NFT}}\mathbf{E}_{\mathrm{NFT}}|^2 = |\epsilon_{\mathrm{media}}\mathbf{E}_{\mathrm{media}}|^2. \tag{5}$$

By combining (1)–(5), we derive the following dimensionless ratio for media/NFT temperature:

$$\frac{\Delta T_{\mathrm{media}}}{\Delta T_{\mathrm{NFT}}} = \frac{|\epsilon_{\mathrm{NFT}}|^2}{|\epsilon_{\mathrm{media}}|^2} \times \frac{\varepsilon''_{\mathrm{media}}}{\varepsilon''_{\mathrm{NFT}}} \times \frac{K_{\mathrm{NFT}}}{K_{\mathrm{media}}} \times \frac{\Omega_{\mathrm{NFT}}}{\Omega_{\mathrm{media}}}. \tag{6}$$

For a low-temperature NFT and write head, this ratio must be as high as possible. Clearly, there are significant factors that are not accounted for in this expression, such as the anisotropic thermal conductivity of HAMR granular media and its under-layers, or the exact structural design of the NFT. Nevertheless, this expression correctly emphasizes some key design requirements for low-temperature NFT operation.

1) The media must have minimum heatsinking.
2) The NFT metallurgy must be optimized for $K_{\mathrm{NFT}} \times (|\epsilon_{\mathrm{NFT}}|^2/\varepsilon''_{\mathrm{NFT}})$.
3) The NFT structural design must have the largest solid angle of heat conduction at the NFT tip.

III. Inverse Electromagnetic Design

Inverse electromagnetic design is based on two concepts: 1) freeform optimization where the shapes are not constrained by a small number of parameters, but rather thousands of parameters and 2) gradient-based optimization via the adjoint method to efficiently optimize the freeform shape. In the context of HAMR, important electromagnetic FOM, include the optical absorption in the media hotspot, the ratio of absorption in the media versus the NFT, and the ratio of absorption in the hotspot versus secondary unwanted hotspots in the media. The gradient is the derivative of the chosen FOM with respect to all of the geometric parameters, which may be the shape boundaries of the NFT and waveguide structures in the HAMR write head. The gradient allows us to use a fast deterministic optimization algorithm like steepest decent. In contrast, heuristic methods like genetic algorithms and particle-swarm optimizations rely on random trials, whose computational burden is too cumbersome for applications in which even a single 3-D simulation of Maxwell's equations are only feasibly on high-performance computing resources. The simplest but most inefficient method to calculating the gradient is finite-difference, which would require at least $N+1$ simulations, where N is the number of parameters. The adjoint method allows us to calculate the gradient with only two simulations regardless of the number of parameters and is crucial for the computational efficiency of the inverse design software. The adjoint method will be described mathematically here in the context of electromagnetics [8], [9].

First, let us denote the FOM as the integral of an arbitrary function of electric field at locations $\boldsymbol{x}$ within a particular volume V_{FOM}

$$\mathrm{FOM} = \int_{V_{\mathrm{FOM}}} f\,(\mathbf{E}\,(\boldsymbol{x}))d^3x. \tag{7}$$

The electric field in this region is a function of electromagnetic sources and of geometric structures. For a geometry optimization, we must model the electromagnetic effects of a small perturbation to the geometry. In this paper, we consider two possible structural perturbations. First, a sparse perturbation is the inclusion of an isolated small sphere of permittivity ϵ_2 displacing a volume within a sea of permittivity ϵ_1, as shown in Fig. 2(a). Second, a boundary perturbation at the interface between two objects of permittivity ϵ_1 and ϵ_2 is the inclusion of a bump of ϵ_2 replacing a volume of ϵ_1, as shown in Fig. 2(b). For either perturbation type, if the

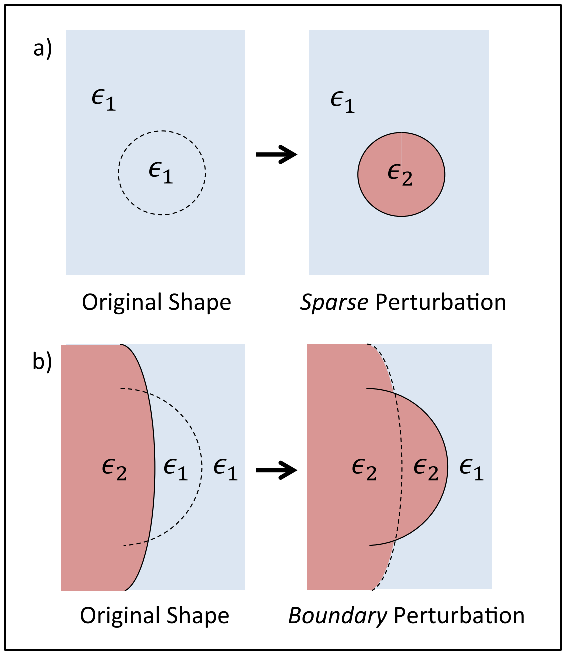

Fig. 2. (a) Sparse perturbation is the inclusion of an isolated small sphere displacing a material of different permittivity. (b) Boundary perturbation is the inclusion of a locally flat bump at the interface between materials of different permittivity.

perturbation is electrically small, the electric field in this perturbed volume of ϵ_2 is the same as the original electric fields in the displaced volume of ϵ_1, only differing by a different set of boundary conditions. For the sparse perturbation, applying boundary conditions around the perturbed sphere leads to (8), relating the electric field in the sphere to the original electric field in the sea of ϵ_1 by the Clausius–Mossotti factor [11]. Similarly, for the boundary perturbation, we arrive at (9), which are the familiar boundary conditions of a flat interface, where $\|$ and $\perp$ denote the parallel and perpendicular vector components of the electric field. In the following equations, $\boldsymbol{x}'$ denotes the location of the perturbed volume:

$$\mathbf{E}_{\mathbf{perturbed}}\left(\boldsymbol{x}'\right) \approx \frac{3}{\left(\frac{\epsilon_2}{\epsilon_1}\right)+2}\mathbf{E}_{\mathbf{orig}}\left(\boldsymbol{x}'\right) \tag{8}$$

$$\mathbf{E}_{\mathbf{perturbed}}\left(\boldsymbol{x}'\right) \approx \mathbf{E}_{\mathbf{orig}\|}\left(\boldsymbol{x}'\right) + \frac{\epsilon_1}{\epsilon_2}\mathbf{E}_{\mathbf{orig}\perp}\left(\boldsymbol{x}'\right). \tag{9}$$

The electromagnetic effects of these perturbations are effectively modeled by a change in dipole moment density, $\mathbf{P}_{\mathbf{pert}}$, in the perturbed volume as described in the following equations for the sparse and boundary perturbations [12]–[14]:

$$\frac{d\mathbf{P}_{\mathbf{pert}}}{dV_{\text{sparse}}}\left(\mathbf{x}'\right) \approx \left(\epsilon_2-\epsilon_1\right)\frac{3}{\left(\frac{\epsilon_2}{\epsilon_1}\right)+2}\mathbf{E}_{\mathbf{orig}}\left(\boldsymbol{x}'\right) \tag{10}$$

$$\frac{d\mathbf{P}_{\mathbf{pert}}}{dV_{\text{bnd}}}\left(\boldsymbol{x}'\right) \approx \left(\epsilon_2-\epsilon_1\right)\left(\mathbf{E}_{\mathbf{orig}\|}\left(\boldsymbol{x}'\right) + \frac{\epsilon_1}{\epsilon_2}\mathbf{E}_{\mathbf{orig}\perp}\left(\boldsymbol{x}'\right)\right). \tag{11}$$

This change in dipole moment causes a change in electric field elsewhere in space. Of interest, the electric field in the volume where the FOM is evaluated is perturbed according to (12), where $\mathbf{G}^{\mathbf{EP}}\left(\boldsymbol{x},\boldsymbol{x}'\right)$ is electromagnetic Green's function relating a unit current source at the perturbation location $\boldsymbol{x}'$ to the electric field induced at location $\boldsymbol{x}$ within V_{FOM}. In application to the complex optical systems used for HAMR, this Green's function can only be evaluated by a full 3-D Maxwell simulation with a current source at $\boldsymbol{x}'$ and observing the numerically calculated electric fields at $\boldsymbol{x}$

$$\mathbf{E}_{\mathbf{perturbed}}\left(\boldsymbol{x}\right) = \mathbf{E}_{\mathbf{orig}}\left(\boldsymbol{x}\right) + \mathbf{P}_{\mathbf{pert}}\left(\boldsymbol{x}'\right)\mathbf{G}^{\mathbf{EP}}\left(\boldsymbol{x},\boldsymbol{x}'\right). \tag{12}$$

By differentiating (7) and using the chain rule, we arrive at an expression for the gradient, which is the derivative of the FOM with respect to a volumetric change in permittivity, shown in (13). The $2Re\{\ \}$ is a result of carefully taking the total derivative with respect to the complex valued functions $\mathbf{E}$ and $\mathbf{P}$, which is not shown in detail here for brevity. By replacing $\mathbf{E}\left(\boldsymbol{x}\right)$ in (13) with the perturbed electric field given by (12) and carrying out the differentiation with respect to $\mathbf{P}_{\mathbf{pert}}$, we arrive at

$$\frac{\partial\text{FOM}}{\partial V_{\text{pert}}\left(\boldsymbol{x}'\right)} = 2Re\left\{\int_{V_{\text{FOM}}}\frac{df}{d\mathbf{E}}\left(\boldsymbol{x}\right)\cdot\frac{\partial\mathbf{E}\left(\boldsymbol{x}\right)}{\partial\mathbf{P}_{\mathbf{pert}}\left(\boldsymbol{x}'\right)}\cdot\frac{d\mathbf{P}_{\mathbf{pert}}}{dV_{\text{pert}}}\left(\boldsymbol{x}'\right)d^3x\right\} \tag{13}$$

$$\frac{\partial\text{FOM}}{\partial V_{\text{pert}}\left(\boldsymbol{x}'\right)} = 2Re\left\{\int_{V_{\text{FOM}}}\frac{df}{d\mathbf{E}}(\boldsymbol{x})\cdot\left[\frac{d\mathbf{P}_{\mathbf{pert}}}{dV_{\text{pert}}}\left(\boldsymbol{x}'\right)\cdot\mathbf{G}^{\mathbf{EP}}\left(\boldsymbol{x},\boldsymbol{x}'\right)\right]d^3x\right\}. \tag{14}$$

Note that when using (14), if we desire the unique value of the gradient at N possible boundary perturbations at locations $\boldsymbol{x}'$, the term $\left[d\mathbf{P}_{\mathbf{pert}}/dV_{\text{pert}}\left(\boldsymbol{x}'\right)\cdot\mathbf{G}^{\mathbf{EP}}\left(\boldsymbol{x},\boldsymbol{x}'\right)\right]$ must be evaluated by N individual Maxwell simulations of a current source equal to the dipole moment of each possible perturbation at each $\boldsymbol{x}'$, respectively. This is computationally expensive. Instead, we leverage reciprocity in electromagnetics, which Lorentz proved in 1896 for any arbitrary arrangement of conducting or dielectric, isotropic or anisotropic bodies [15], which generically describes even the most complex optical structures used for HAMR. He proved that among arbitrary structures, two current distributions $\mathbf{J_1}$ and $\mathbf{J_2}$ that individually induce the electric field distributions $\mathbf{E_1}$ and $\mathbf{E_2}$, respectively, are related by the following simple expression:

$$\int\mathbf{J_1}\cdot\mathbf{E_2}dV = \int\mathbf{J_2}\cdot\mathbf{E_1}dV. \tag{15}$$

It is useful for us to consider the case of a unit current at $\boldsymbol{x}'$ and the electric field it induces at $\boldsymbol{x}$, in which case the well know reciprocity of electromagnetic Green's functions is obtained

$$\mathbf{G}^{\mathbf{EP}}\left(\boldsymbol{x},\boldsymbol{x}'\right) = \mathbf{G}^{\mathbf{EP}}\left(\boldsymbol{x}',\boldsymbol{x}\right). \tag{16}$$

After substituting (16) into (14), we obtain a fundamentally different expression for calculating the gradient that is computationally inexpensive, shown in (17). According to (10) and (11), the first term depends on the electric fields from a single forward simulation of the original source and the unperturbed geometry. The latter term comprises the electric fields from a single adjoint simulation, where the source is the superposition of current sources of amplitude $df/d\mathbf{E}\left(\boldsymbol{x}\right)$ in the volume of the FOM region. Hence, only two Maxwell simulations are required to obtain the gradient at all potential perturbation positions $\boldsymbol{x}'$

$$\frac{\partial \mathrm{FOM}}{\partial V_{\mathrm{pert}}(\boldsymbol{x}')} = 2Re\left\{\frac{d\mathbf{P}_{\mathbf{pert}}}{dV_{\mathrm{pert}}}(\boldsymbol{x}')\cdot\int_{V_{\mathrm{FOM}}}\frac{df(\boldsymbol{x})}{d\mathbf{E}(\boldsymbol{x})}\mathbf{G}^{\mathbf{EP}}(\boldsymbol{x}',\boldsymbol{x})\,d^3x\right\}. \tag{17}$$

By substituting (10) and (11) in (17), we arrive at the gradient calculations for the sparse and boundary perturbations that are used in this paper

$$\frac{\partial \mathrm{FOM}}{\partial V_{\mathrm{bnd}}(\mathbf{x}')} \approx 2Re\left\{\left[(\epsilon_2-\epsilon_1)\left(\mathbf{E}_{\mathbf{orig}\parallel}(\boldsymbol{x}')+\frac{\epsilon_1}{\epsilon_2}\mathbf{E}_{\mathbf{orig}\perp}(\boldsymbol{x}')\right)\right]\cdot\left[\int_{V_{\mathrm{FOM}}}\frac{df}{d\mathbf{E}}(\boldsymbol{x})\cdot\mathbf{G}^{\mathbf{EP}}(\boldsymbol{x}',\boldsymbol{x})\,d^3x\right]\right\} \tag{18}$$

$$\frac{\partial \mathrm{FOM}}{\partial V_{\mathrm{sparse}}(\mathbf{x}')} \approx 2Re\left\{\left[\frac{3(\epsilon_2-\epsilon_1)}{\left(\frac{\epsilon_2}{\epsilon_1}\right)+2}\mathbf{E}_{\mathbf{orig}}(\boldsymbol{x}')\right]\cdot\left[\int_{V_{\mathrm{FOM}}}\frac{df}{d\mathbf{E}}(\boldsymbol{x})\cdot\mathbf{G}^{\mathbf{EP}}(\boldsymbol{x}',\boldsymbol{x})\,d^3x\right]\right\}. \tag{19}$$

Using this gradient calculation, one can easily implement an iterative optimization using steepest descent, where every iterative geometry update is in the direction of the gradient. In this paper, we also used finite-difference between consecutive iterations to approximate the second derivative $\partial^2\mathrm{FOM}/\partial V_{\mathrm{bnd}}(\boldsymbol{x}')^2$, specifically the diagonal of the Hessian, to implement a quasi-Newton update method. This demonstrated superior convergence as compared with steepest descent. In addition, the freeform nature of the boundary optimization was implemented by representing the boundaries on a binary bitmap, where 1s and 0s represent the material inside and outside the various boundaries. Every pixel along the boundary was treated as a separate degree of freedom, and the boundary could expand outward or contract inward on a per-pixel basis. Hence, the optimized boundaries were allowed to completely diverge from the initial shape fed to the optimization algorithm. This freeform geometric representation combined with gradient-based optimization allows for creative objective-first design of 3-D electromagnetic structures, which we call inverse electromagnetic design.

IV. Maxwell Simulation Methods

The ability to perform accurate 3-D electromagnetic simulations is imperative to computational optimization. Simulation results in this paper use a commercial finite difference time domain (FDTD) Maxwell solver, Lumerical FDTD, in which a pulse of light is injected into the waveguide of the HAMR system and propagated in the time domain toward the NFT, write pole, and media until pulse energy has decayed beyond our desired precision. A detailed mesh convergence test was performed to ensure minimal computational error due to discretization. The most crucial and computationally demanding mesh requirements were 1 nm cubic Yee cells in the metallic NFT and 0.5 nm Yee cell thicknesses in the media to resolve the various nanometer-thin layers of the media stack. We used an in-house high-performance computing cluster consisting of 336 cores and 668 GB RAM over 26 nodes. By parallelizing the solver through a message-passing interface, Open MPI [16], and 40 Gb/s Infiniband interconnects, our in-house cluster can simulate FDTD models of half a million Yee cell nodes. Typically, we ran simulations on 64 to 128 cores at a time, with which we could run iterative optimizations of HAMR structures in roughly one day's time.

TABLE I

Structural and Optical Properties in Proposed HAMR System

Device	Dimensions	n	k
Au NFT	125 nm radius, 60 nm thick	0.16	5.08
Au NFT Peg	50 nm wide at ABS, 30 nm thick	0.16	5.08
Ta_2O_5 Waveguide	100 nm thick	2.1	-
SiO_2 Cladding	-	1.4	-
CoFe Writepole	120 wide at ABS	3	4
Head Overcoat	2.5 nm thick	1.6	-
Air Gap	2.5 nm thick	1.0	-
Media Overcoat	2.5 nm thick	1.2	-
FePt Recording Layer	10 nm thick	2.9	1.5
MgO Interlayer	15 nm thick	1.7	-
Au Media Heatsink	80 nm thick	0.26	5.28
Media Substrate	infinite	1.5	-

TABLE II

Thermal Properties in Proposed HAMR System

Material	Specific Heat (J/m^3K)	Thermal Conductivity (W/mK)
Au	$3\cdot10^6$	100
Ta_2O_5	$2\cdot10^6$	2
SiO_2	$2\cdot10^6$	1
CoFe	$3.5\cdot10^6$	20
FePt - Lateral	$3\cdot10^6$	5
FePt - Vertical	$3\cdot10^6$	50
MgO	$2\cdot10^6$	3

Fig. 4 shows 3-D views of the HAMR structure that was modeled, and Table I contains structural and optical properties at the operation laser wavelength of 830 nm for the numerous write head and media components. These properties were chosen to closely mimic designs from industry publications and patent literature. Table II shows thermal properties that were assumed for a thermal finite-element model performed in COMSOL multi-physics to predict the media and NFT temperatures.

For low-temperature operation, a crucial NFT design choice is to have the largest solid angle of heat conduction from the tip of the NFT. The left of Fig. 3 shows a simplified HAMR system consisting of a magnetic write pole, media stack, gold lollipop NFT, and incident light in a slab waveguide similar to that of seagate's parabolic solid-immersion mirror (PSIM) [6]. Through 3-D FDTD modeling, we observe that this system produces a sharply confined hotspot in the media's recording layer, whose dimensions are defined by the cross-sectional dimensions of the NFT tip at the air-bearing surface (ABS).

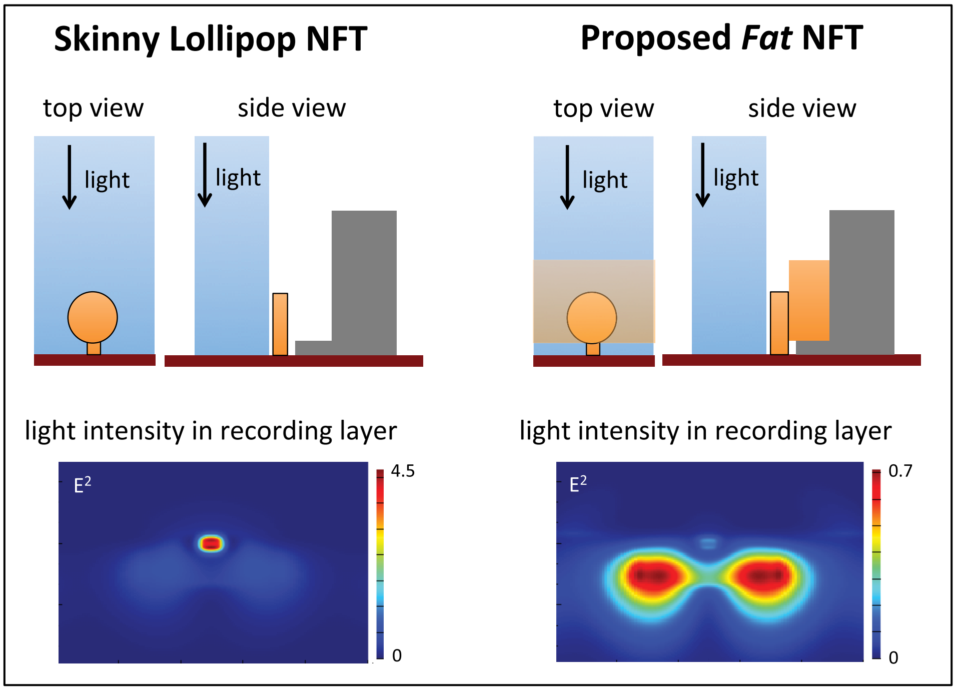

Fig. 3. HAMR optical system composed of a Ta_2O_5 waveguide (blue), gold NFT (yellow), CoFe write pole (gray), and magnetic media (red). Left: typical skinny lollipop NFT produces a confined hotspot in the storage layer. Right: proposed fat NFT has different electromagnetic behavior and is a poor mode match to a PSIM-like waveguide mode.

However, the lollipop NFT is skinny and has a very little solid angle of heat conduction. Such NFTs experience a rise in temperature by hundreds of degrees, which is a significant cause of failure in HAMR systems. Therefore, we need a fat NFT that has a large solid angle of heat conduction.

We propose a fat NFT, shown on the right of Fig. 3, consisting of a thin-film gold pattern embossed on a bulk chunk of gold. The only part of the thin-film pattern that does not touch the bulk gold is the NFT tip, because the magnetic write pole tip was fixed at a 30 nm offset from the NFT tip at the ABS. This new fat NFT is not the familiar lollipop NFT. We illuminated the fat NFT with the same PSIM-like waveguide mode, and the observed light intensity in the media is shown in the bottom right of Fig. 3. The PSIM-like waveguide mode is a poor mode-match to the fat NFT and the system delivers a poorly confined hotspot. The side lobes in the light intensity pattern are unacceptable, because high temperatures in the media outside of the hotspot would unintentionally erase information. A typical data storage specification is to allow for 100 000 writes to a particular track without erasing data on nearby tracks. To meet this specification, the peak light intensity in the hotspot versus the peak intensity elsewhere in the media should be at least five, and achieving this light intensity ratio with the proposed fat NFT was the goal in this paper.

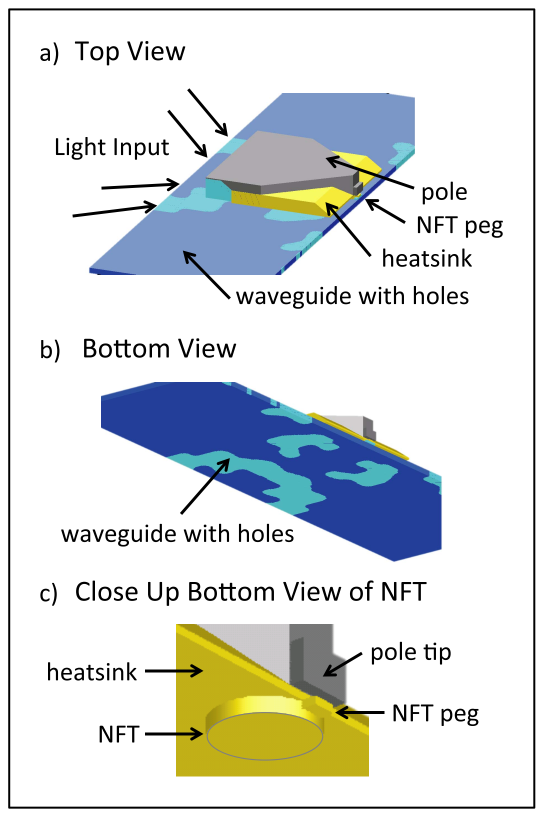

Fig. 4. 3-D views of the proposed HAMR light delivery structure. Fat NFT is a thin-film disk embossed on a bulk chunk of gold, and the slab waveguide contains a pattern of low-index material.

V. Inverse Design Results

The strategy to improve the mode match between the incident waveguide light and the proposed fat NFT is to insert an array of holes of low-index material etched into the high-index slab waveguide. A 3-D perspective view of the proposed HAMR structures is shown in Fig. 4. The incident light enters a slab Ta_2O_5 waveguide and evanescently couples to the fat NFT. The waveguide is patterned with holes of SiO_2 to reshape the incident mode to better couple to the fat NFT. The NFT

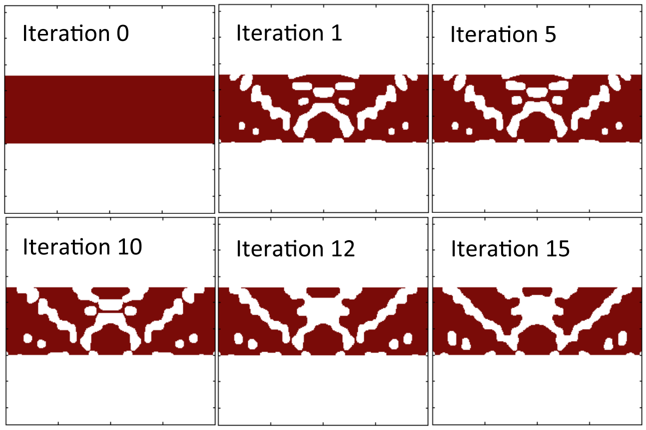

Fig. 5. Top view and iterative evolution of a Ta_2O_5 slab waveguide core (red) patterned with SiO_2 holes (white). This computer-generated pattern offers more absorption in the hotspot and reduced unintentional erasure of adjacent tracks.

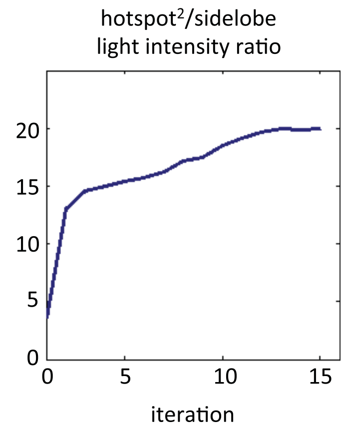

Fig. 6. Convergence plot of the optimized FOM versus iteration. FOM was the square of the peak light intensity in the media hotspot divided by the peak light intensity in the unwanted sidelobes.

consists of a gold disk embossed directly underneath a bulk chunk of gold with a gold peg protruding toward to ABS. A CoFe magnetic write pole sits on top of the NFT, and the write pole tip is 30 nm above the top surface of the NFT peg. Not shown in these 3-D views is the magnetic media stack, described in Table I, which is adjacent to the right side of the waveguide, NFT peg, and write pole tip.

We used the inverse electromagnetic design software to computationally generate the optimal waveguide pattern. The FOM for the optimization was the light intensity ratio between the hotspot and unwanted side lobes in the media. The geometry that was optimized was a 2-D binary bitmap of 75 000 pixels, where each pixel represented a 3-D voxel of either SiO_2 or Ta_2O_5 of dimensions 8 nm × 8 nm × 100 nm occupying a total volume of 4 μm × 1.2 μm × 0.1 μm, which was the region of the waveguide core under the NFT and adjacent to the ABS. Using mathematical morphology, additional constraints on the binary bitmap were employed to enforce that the minimum diameter of a SiO_2 hole was greater than 128 nm and the radius of curvature of any boundary was at least 64 nm. Fig. 5 shows the iterative optimization of the holey waveguide pattern over 15 iterations, representing a total of only 30 simulations to optimize 75 000 degrees of freedom. In the first iteration, the software was configured to use the sparse gradient and added many new SiO_2 holes into the Ta_2O_5 waveguide core. In the latter iterations, the software calculated the boundary gradient and was constrained to make boundary changes only. Fig. 6 shows a plot of the FOM versus iteration, showing smooth stable convergence toward a locally optimal design. Note that, between iterations 10 and 15, the geometry kept changing with little change to the FOM, suggesting that the optimal solution is robust to small variations in the boundaries of the waveguide pattern. In addition, although not specifically studied here, this also suggests that the optimal solution is robust to small variations in material permittivity.

For a fair comparison, we modeled a typical heatsink structure for a lollipop NFT, shown in Fig. 7, that consists of a 120 nm diameter gold cylinder connecting the center of the NFT to a bulk chunk of gold of the same dimensions

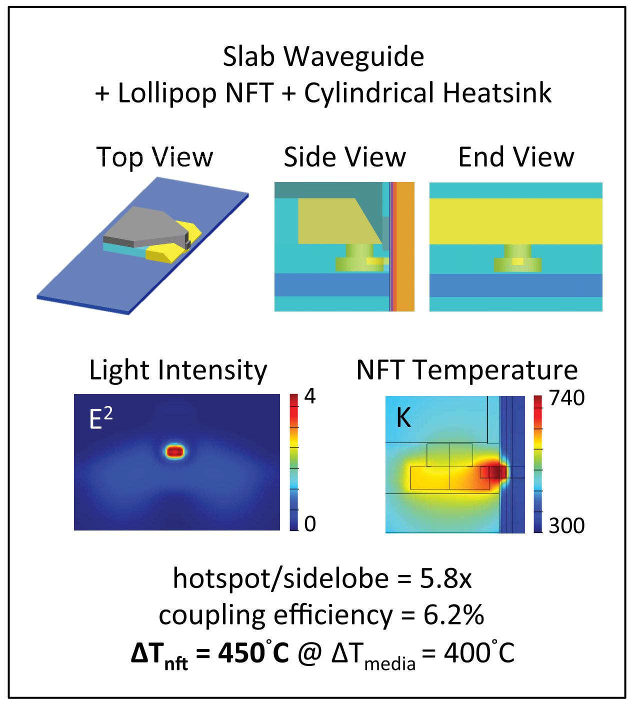

Fig. 7. Top: structural model of a slab waveguide, lollipop NFT, narrow cylindrical heatsink, and write pole. Mid: simulated light intensity in the media and side-view temperature profile of the NFT, heatsink, and media. This design suffers from severe self-heating.

used in the fat NFT design. The narrowness of the cylindrical heatsink limits the heat conduction out of the NFT peg. The same geometries of the waveguide, cladding, write pole, and media used in the fat NFT model were used in the lollipop NFT model. We imported the optical absorption profile as a volumetric heat source in a thermal FEM model, in which we observed a peak temperature rise in the NFT peg of 450 °C above ambient when injecting enough light into waveguide to achieve a desired 400 °C temperature rise in the media hotspot.

The optimized waveguide pattern coupled to the proposed fat NFT is shown in Fig. 8. Using identical simulation models, we observed that the new proposed structure produces nearly identical optical properties of the media hotspot. In particular,

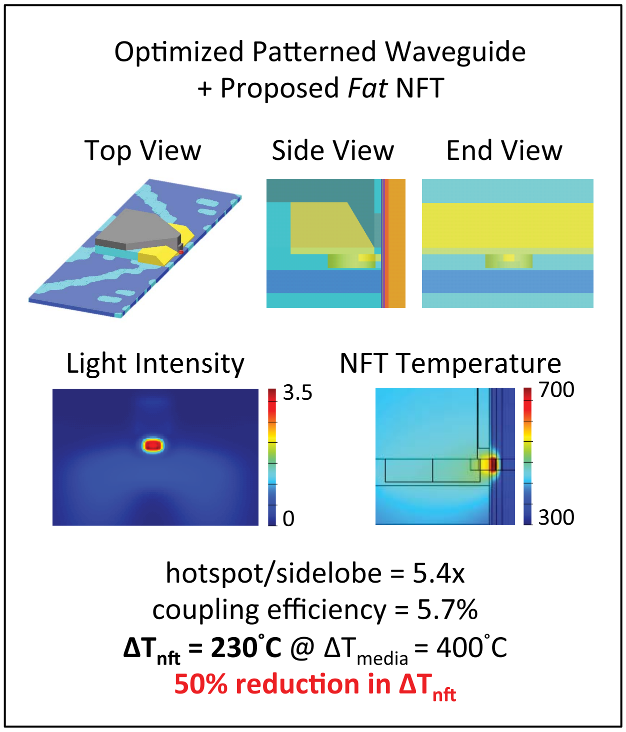

Fig. 8. Top: structural model of the proposed HAMR structure consisting of a patterned waveguide, fat NFT, and write pole. Mid: simulated light intensity in the media and side-view temperature profile of the NFT, heatsink, and media. This design achieves desirable optical properties and significantly reduced self-heating.

the hotspot shape is well defined by the NFT peg dimensions, the power absorbed in the hotspot normalized to power injected into waveguide is ∼6%, and the light intensity ratio between the hotspot and undesired sidelobes is greater than 5. More importantly, we observed that the proposed fat NFT has a peak temperature rise in the NFT peg of only 230 °C above ambient when injecting enough light into the waveguide to achieve the same 400 °C temperature rise in the media. This represents a ∼50% lower temperature rise (220 °C) compared with a lollipop NFT with a typical cylindrical heatsink. This is expected, because the typical cylindrical heatsink offers little solid angle of heat conduction. A ∼50% reduction in NFT temperature could result in exponential improvements to HAMR write head lifetimes.

VI. Conclusion

The simple thermal analysis in this paper mandates that we have a fat NFT with a large solid angle of heat conduction. The fat NFT proposed in this paper has different electromagnetic properties than the familiar skinny lollipop NFT. Thus, a different incident waveguide mode is required to properly excite the proposed NFT. With the power of the inverse electromagnetic design, we computationally generated unexpected waveguide patterns that mode-matched to the fat NFT. The combined system of a patterned waveguide and fat NFT produced the desired optical properties for HAMR nanoscale light delivery as well as a greatly reduced operation temperature. Reliability of structural and electronic devices often varies with the exponential of temperature. Hence, the new structures proposed in this paper may offer orders of magnitude improvements to reliability by reducing the NFT self-heating by ∼50% (220 °C) compared with typical industry designs. Of course, the exact waveguide mode that illuminates the NFT and exact material properties in the write head and media of a commercial product must be accounted for in the optimization itself. Accordingly, the inverse electromagnetic design software will be made available online at http://optoelectronics.eecs.berkeley.edu/PhotonicInverseDesign for the purpose that it may be directly applied to industry's custom HAMR designs. An important future paper will include simultaneous optimization of the NFT and waveguide geometries, co-optimization of numerous FOM, and direct optimization of thermal physics in addition to electromagnetics. We expect that computationally generated electromagnetic structures will have an important role in commercial HAMR technology.

Acknowledgment

This work was supported in part by the Advanced Storage Technology Consortium and in part by the National Science Foundation Graduate Research Fellowship Program.

References

[1] T. W. McDaniel, "Ultimate limits to thermally assisted magnetic recording," *J. Phys., Condens. Matter*, vol. 17, no. 7, p. R315, Feb. 2005.
[2] M. H. Kryder *et al.*, "Heat assisted magnetic recording," *Proc. IEEE*, vol. 96, no. 11, pp. 1810–1835, Nov. 2008.
[3] R. E. Rottmayer *et al.*, "Heat-assisted magnetic recording," *IEEE Trans. Magn.*, vol. 42, no. 10, pp. 2417–2421, Oct. 2006.
[4] B. C. Stipe *et al.*, "Magnetic recording at 1.5 Pb m^{-2} using an integrated plasmonic antenna," *Nature Photon.*, vol. 4, no. 7, pp. 484–488, Jul. 2010.
[5] M. A. Seigler *et al.*, "Integrated heat assisted magnetic recording head: Design and recording demonstration," *IEEE Trans. Magn.*, vol. 44, no. 1, pp. 119–124, Jan. 2008.
[6] W. A. Challener *et al.*, "Heat-assisted magnetic recording by a near-field transducer with efficient optical energy transfer," *Nature Photon.*, vol. 3, no. 4, pp. 220–224, Apr. 2009.
[7] S. Bhargava and E. Yablonovitch, "Inverse design of optical antennas for sub-wavelength energy delivery," in *Proc. CLEO*, Jun. 2014, p. FW3K.6.
[8] O. D. Miller, "Photonic design: From fundamental solar cell physics to computational inverse design," Ph.D. dissertation, Dept. Elect. Eng., Univ. California, Berkeley, CA, USA, 2012.
[9] S. Bhargava, "Inverse electromagnetic design of optical antennas for heat-assisted magnetic recording," Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, CA, USA, Tech. Rep. UCB/EECS-2012-248, 2012.
[10] C. M. Lalau-Keraly, S. Bhargava, O. D. Miller, and E. Yablonovitch, "Adjoint shape optimization applied to electromagnetic design," *Opt. Exp.*, vol. 21, no. 18, pp. 21693–21701, Sep. 2013.
[11] J. D. Jackson, *Classical Electrodynamics*, 3rd ed. New York, NY, USA: Wiley, 1998.
[12] S. G. Johnson, M. Ibanescu, M. A. Skorobogatiy, O. Weisberg, J. D. Joannopoulos, and Y. Fink, "Perturbation theory for Maxwell's equations with shifting material boundaries," *Phys. Rev. E*, vol. 65, no. 6, p. 066611, Jun. 2002.
[13] S. G. Johnson, M. L. Povinelli, M. Soljačić, A. Karalis, S. Jacobs, and J. D. Joannopoulos, "Roughness losses and volume-current methods in photonic-crystal waveguides," *Appl. Phys. B*, vol. 81, nos. 2–3, pp. 283–293, 2005.
[14] M. S. Dadash, N. K. Nikolova, and J. W. Bandler, "Analytical adjoint sensitivity formula for the scattering parameters of metallic structures," *IEEE Trans. Microw. Theory Techn.*, vol. 60, no. 9, pp. 2713–2722, Sep. 2012.
[15] H. A. Lorentz, "The theorem of Poynting concerning the energy in the electromagnetic field and two general propositions concerning the propagation of light," in *Proc. Collected Papers*, vol. 3. 1896, pp. 1–15.
[16] E. Gabriel *et al.*, "Open MPI: Goals, concept, and design of a next generation MPI implementation," in *Proc. 11th Eur. PVM/MPI Users' Group Meeting*, 2004, pp. 97–104.