Thermal/Optical Effects in NSOM Probes

B.I. Yakobson, A. LaRosa, H.D. Hallen, and M.A. Paesler

Physics Department, North Carolina State University, Raleigh, NC 27695-8202, USA

Abstract

The coupling of thermal and optical effects in near-field scanning optical microscope (NSOM) fitted with a coated tapered fiber probe is considered. Heating of the probe occurs due to the absorption of light in the metal coating in the tapered region of the probe at relatively large (tens of microns) distances from the probe apex. Conduction of heat that occurs with a time constant of tens of milliseconds results in an elevated probe temperature over a rather large region of the taper. This in turn, affects the optical throughput of the NSOM. Experimental measurement of thermal/optical coupling and a discussion of coupling mechanisms is included.

Key words: near-field scanning optical microscopy (NSOM), thermal effects

1. Introduction

Fundamental advances in near-field microscopy occur on one of two fronts: enhanced resolution and increased signal. Technological improvements in instrument design have focused on issues such as independent topological feedback and control; or on the development of a variety of contrast mechanisms. Resolution and signal studies, however, have led to the most basic advances in NSOM instrument design. While considerable attention has been paid to NSOM resolution, very few studies consider the mechanisms limiting NSOM signal.

In the proceedings of the previous conference in this series, we suggested that the study of throughput limitation in an NSOM probe should focus on an investigation of the behavior of the probe at considerable - i.e. far-field - distances from the probe tip [1]. That argument involved a simple thermodynamic consideration of energy losses in the probe. In brief, an injected power W of several milliwatts typically results in a transmitted power of a few nanowatts. It is not conceivable that these milliwatts could all be lost in the near-field region. There are simply no thermodynamically reasonable mechanisms that involve dissipation of this much power in the sub-wavelength structure that comprises the NSOM probe apex. Indeed, assuming the dissipating volume to be as large as $a \approx 1\mu m$ in diameter, one can estimate its temperature for radiation channel from $a^2 \cdot \sigma T^4 = W$, which gives a very large temperature. For the alternative path of conduction back up the fiber channel, $T \cdot a \cdot \kappa \sim W$, and the temperature can be estimated to be on the order of a thousand degrees.

In the contribution below, we extend these arguments by using a repeated zone scheme (also introduced at NFO-2 [2]) to determine the region of an NSOM probe where the milliwatts of power are lost. We further analyze the thermal

diffusion in an NSOM probe to determine the role of various heat transport and loss mechanisms. We also model the coupling between the thermal response of the probe to the effects of light passing through it and the ability of that probe to deliver useful light through an aperture at its apex.

Experimental studies of the thermal response of an NSOM probe confirm the conclusions reached in the modeling studies. The data show that the thermal response of a typical NSOM probe under normal operation has a time constant on the order of tens of milliseconds. Such a response is entirely consistent with the thermal diffusion calculations. Finally, we suggest that our modeling is consistent with probe temperature measurements [3] and independent modeling work [4] as presented elsewhere in these proceedings.

2. The Role of Sideput-Density

In the illumination mode NSOM, light is injected into the core of the cylindrical fiber from which the probe is drawn. In the region where the probe begins to taper, the core has insufficient diameter to support propagation. Light is ejected from the core. Propagation proceeds in the cladding, and light is contained by the reflective metallic coating. The optical signal that can be provided in this mode is limited by the ability of the probe to tolerate and dissipate the heat generated by light absorbed by the metal with its non-unity reflection coefficient. One simple measure of this lost light - which we call sideput - is the power dissipated per unit area of the metal coating. If one assumes that the reflectivity does not change appreciably along the taper, then this measure scales with the density of reflections, which can be modeled in a simple ray tracing scheme. Insofar as the actual light profile may be represented in terms of rays in this far-field region, then one need only determine the density of ray-metal reflections to obtain the desired measure. We call this quantity the sideput density.

A long standing method for tracing rays in waveguides of non-constant but linear dimensions involves a repeated zone scheme [5] that was recently applied [2] to NSOM tapers. In this scheme, one follows not the actual trajectory of a ray in an NSOM probe, but a series of images of that trajectory in an expanded representation of the probe. For a conical probe, the scheme is straightforward. For more complicated geometries, it is possible to extend the scheme by considering any profile to be a series of truncated cones [2]. Indeed, any monotonically tapering geometry can be studied in this manner. In the following analysis, we consider only linear tapers - i.e. cones - but suggest that extension to other shapes is straightforward.

We characterize a bundle of rays by the angle β by which they differ from being paraxial and the distance r from the axis and ignore any skew component. The half angle of the conical taper is ϕ. We show in figure 1 the situation for $\beta = 0°$ and $\phi = 6°$ with the bundle entirely filling the probe. In the figure this bundle of rectangular cross section distribution is shown in the taper (which corresponds to the power density distribution inversely proportional to the distance from the axis, $1/r$, consistent with the translational invariance in the

long fiber). The interactions of this bundle with the walls in the first zone are shown. The total number of confrontations, that is the sideput density, is plotted as a function of distance along the taper in the bottom of the figure. (We ignore the dependence of the reflectivity on the incident angle.) The somewhat chopped profile is not an artifact of the calculation. It is a true measure of the rather quantized nature of bundle-wall interactions. For this particular set of (β, ϕ) values, a strongly peaked sideput density results. The abscissa in figure 1 is a normalized measure of the probe taper. Typically, near-conical tapers are on the order of 1 mm long. For such a probe, the scale thus represents fractions of a millimeter. We assume a structure of this size in the following discussion.

Shown in figure 2 are several plots of sideput density for a variety of (β, ϕ) values. While they differ somewhat, one thing is clear: The large majority of the sideput in an NSOM probe is lost at distances of several tens of microns from the probe apex. That is, the sideput density peaks where ray tracing is valid. This fact has been experimentally confirmed with a conventional microscopic investigation of damaged tips [6].

3. Heat Dissipation

The heat generated by the sideput is dissipated through one of several conceivable channels.

i) Lateral heat transfer through the SiO_2 results in thermal equilibration across the cross section of the taper with a time constant $\tau_{\perp} = 0.4\text{x}10^{-6} \left[\frac{r}{\mu m}\right]^2$ seconds. For the tapers considered, one has $r \lesssim 60\ \mu m$ so that this time is relatively short. That is, thermal equilibrium across the fiber is quite rapidly achieved.

ii) The longitudinal thermal conductance $\bar{\kappa}$ of the probe results in a conductive dissipation. This conductance is a composite containing contributions from both the metal coating and the glass. The high conductance through the Al coating dominates conductance near the apex ($\kappa_{Al} = 250\ \kappa_{glass}$), but the glass becomes increasingly important with increasing radius ($\bar{\kappa} = 2.5\ \kappa_{glass}$ at a distance of 250 μm from the apex). This behavior of the thermal conductance of a probe as a function of distance from its apex is shown schematically in figure 3. In the figure $\bar{\kappa}(z)$, the effective conductance of the probe as a function of distance from the apex, is plotted as the solid curve. The probe has a radius r at a distance z from the apex, and the aluminum coating has a constant thickness t along its entire length. The arbitrary length scale might be considered to be on the order of microns thus the 1000 represents a probe taper length in this figure of 1 mm, a typical value.

To calculate the temporal response through this conductive channel, an approximation to $\bar{\kappa}$ is made. This approximation, called $\kappa_{app}(z)$, is shown as the dashed-dotted line in the figure. Because $\bar{\kappa}(z)$ increases so dramatically near the apex (i.e. at low z), we assume in our calculation that the conductance is essential infinite from the apex out to some critical value z_c (in the figure z_c is arbitrary chosen to be 160 arbitrary units). For larger distances a constant value is assumed. In the figure, this constant value is chosen to be equal to roughly three times its value at $z = \infty$, that is $\kappa_{app}(z > z_c = 160) \approx 3\kappa(z = \infty)$. We also assume that conduction alone is responsible for dissipation and that this conduction can be approximated as being due to a sphere of some radius z_c, cooling into all space. In this approximation, a characteristic time constant of $\tau_{||} = \frac{\Delta z^2}{4\kappa} = 0.02$ seconds is obtained.

iii) Lateral convection to the ambient is relatively slow with a characteristic time that scales with the surface area. For a tip in laboratory air, a time constant associated with these convective losses given by $\tau_c \approx 0.2 \left[\frac{r}{\mu m}\right]$ seconds can be calculated.

iv) Radiation also scales with the radiating area, so the functional dependence of the time constant associated with dissipation through the lateral radiation channel is the same as that for lateral convection. These losses scale with the temperature difference between the probe and the environment to the fourth power. In a room temperature laboratory and for the highest conceivable temperature at which the probe will avoid thermal damage (quite conservatively taken to be 10^3 K), the time constant associated with such lateral radiative losses is $\tau_r \approx 0.6 \left[\frac{r}{\mu m}\right]$.

Because $\tau_{||} >> \tau_{\perp}$, τ_c, and τ_r, the hot tip may be thought of as a cone almost insulated laterally but uniformly heated in some domain of length Δz. This justifies the assumption used to calculate the conductive time constant, i.e. that the conduction from a length Δz heated near the apex of an insulated cone is equivalent to the cooling of a sphere of radius Δz. It is of particular interest to keep the relative magnitude of these various time constants in mind in modeling tip thermodynamics. In particular, for a tip radius r = 50 nm, one obtains a time constant for convection cooling of τ_c = 10 ms. This value, however, is of no significance in thermal modeling since the heating is fed by a much faster process, namely the conductive heating from a nearby region (at $z \approx z_c$) which is coupled to the tip by a region of very high conductance, as is shown in figure 3. Thus a tip of this radius will retain the temperature of the more thermally massive probe shank.

4. Measurement of Thermal Relaxation of Probe

In an experimental set-up quite similar to one described in a presentation at the previous conference in this series [7] the time constant associated with conductive cooling, i.e. $\tau_{||}$, has been determined [6] for an NSOM with no sample in place. The response of a typical probe is shown in figure 4. At the top of the figure, the infrared response to a step function (on this time scale) is shown. The data are fit to a single exponential of $\tau = 0.02$ seconds. Displayed at the bottom of the figure is the relative magnitude of the synchronously detected infrared signal as a function of frequency. Probe response is independent of frequency ν below $\nu \approx 10$ Hz. To one significant figure, the results of these two experiments agree. That is, $\nu \approx \frac{1}{2\pi\tau}$. Essentially all probes measured had temporal responses of this magnitude, thus suggesting that conductive cooling is universally dominant in NSOM operation.

5. Thermal/Optical Coupling

The measured temporal response of an NSOM probe's throughput to pulsed radiation involves a time constant that is essentially identical to the modeled temporal response associated with probe heating. What remains is a determination of the physical link between probe heating and throughput that renders the thermal response optically detectable.

Heating can affect the passage of light through the tip in several ways:

[a] Enlarged Aperture Effect. The light in an NSOM probe passes through a subwavelength metal aperture. Since this aperture will grow in diameter with increased temperature, probe throughput should also scale with temperature as a result of this mechanism. The dependence of transmission through a subwavelength aperture on aperture radius is a matter of some discussion. Bethe [8] showed that for an aperture of radius r in an infinitely thin infinitely conducting metal sheet, transmission goes as $[\frac{r}{\lambda}]^6$. Experimental studies [9, 10] suggest that the dependence is somewhat weaker at $[\frac{r}{\lambda}]^4$. Using the former dependence and the known thermal expansion coefficient of Al, the relative *increase* in the magnitude of the infrared transmission due to this enlarged aperture effect can be determined [6].

[b] Elongated Probe Effect. Throughput can decrease due to thermal elongation of the region of the probe beyond cut-off. While quantification of this effect is somewhat more complicated since the exponential fall-off of intensity with distance is a function of the taper diameter, we can estimate the exponential to be on the order of $2\pi/\lambda$. The relative *decrease* in

the magnitude of the infrared transmission due to this elongated probe effect due to this elongated probe effect is also easily calculable [6].

[c] Diminished Reflectivity Effect. Both the infrared and the visible light used in the experiment described above fall into the relaxation region where Al is strongly reflecting. The relative *decrease* in the magnitude of the infrared transmission due to this diminished reflectivity effect is also easily evaluated [6].

The rather curious result of the three calculations outlined above is that these effects are essentially equal in absolute value for the probe shapes, probe materials, and illumination conditions typically involved in NSOM imaging. To within a factor of 3 or 4, calculations show [6] that for a given temperature rise, the (positive) enlarged aperture effect, the (negative) elongated probe effect, and the (negative) diminished reflectivity effect possess nearly equal influence on probe throughput. All effects have a normalized magnitude near $\pm 10^{-4}/\mathrm{K}$.

A more careful examination of the experimental data allows one to study the relative contributions of the mechanisms responsible for the heat-sensitive nature of probe throughput. Particular interesting is the fact that the throughput can either rise or fall with increasing temperature! That is, the modulated ir signal can go up or down as the visible light is turned on or off. In figure 5, the ir throughput as a function of time is plotted for two different probes. Initially no visible light passes through the system. With the introduction of pulsed visible light at a time of 5.5 seconds, the ir signal is modified in a square wave pattern manifest in the figure. The visible light is modulated in a square wave pattern with a period of one second. In the trace at the top of the figure the ir intensity *increases* with the introduction of the visible light and thus with increased temperature. In the bottom trace the ir signal *decreases* with temperature.

These data suggest that mechanisms resulting in both an enhancement and in a diminution of probe throughput must sometimes be dominant in NSOM probes. Since only the enlarged aperture effect (described in paragraph [a] above) results in an increased throughput at higher temperatures, this must be dominant in at least some experimental situations. The opposite behavior - decreased throughput at higher temperatures - suggests that either the elongated probe effect (paragraph [b]) or the diminished reflectivity effect (paragraph [c]) can be dominant.

6. Measurement of Intensity Dependence

We have measured the intensity dependence of the normalized infrared modulation $\frac{\Delta I}{I}$ as shown in figure 6. In this experiment, the cw infrared intensity is held constant and as the intensity of the visible light is increased. The synchronously detected modulated normalized infrared signal is measured and plotted in the figure. For all tips studied in these investigations, the data appear qualitatively quite similar to the curves shown as figures 6. That is, the

magnitude of the modulation of the infrared signal tracks linearly with intensity at low light levels. For increased visible radiation, this magnitude falls off to a weaker dependence (lower slope) at higher incident intensities.

The data of figure 6 clarify interpretation of both the nature of the heat dissipation (§3) and thermal/optical coupling (§5). In particular, in the linear intensity regime the scaling of the modulation with intensity suggests that neither the heat dissipation mechanism nor the thermal/optical coupling balance changes with intensity.

When the curve of figure 6 bends to lower slopes, however, one of two things has happened. In one scenario, another channel for thermal dissipation has begun to play a role and the temperature of the probe no longer scales with light intensity in the same way. For example, convection or radiation might begin to provide a significant cooling channel at higher intensities. In an alternative interpretation, the balance between thermal/optical coupling changes, and a different mechanism (a, b, or c in §V) begins to play a role in the coupling.

VII. Conclusions. Calculations involving both thermodynamic and optical arguments suggest that the principal cause of NSOM probe breakdown lies in diffusion of the metal coating away from the probe apex. This optically-induced probe heating results from the large sideput density of light absorbed by the metal coating in a region several tens of microns from the apex of the probe. The thermal response of the tip is temporally dominated by a mechanism (of time constant $\approx 10^{-2}$ s) involving the conduction of the heat away from this region where the sideput density is at a maximum. The optical performance of the probe is affected by the optically-induced heating in several ways. In some cases the dominant effect involves an increased aperture which allows for greater throughput. In other cases, the optical performance is dominated either by taper elongation or a decrease in metal reflectivity or a combination of these two effects.

Finally, we underscore the fact that a consensus is beginning to emerge among researchers studying tip heating. Both the modeling work in Mikhail Libenson's group as well as the experimental work in Alfred Meixner's laboratory agree substantively with the principal conclusions drawn in this contribution. These two works are presented in separate papers in this volume. The temperature measurements from the latter group are particularly useful. We are currently incorporating these measurements into our modeling of probe optics and thermal kinetics.

Acknowledgments

The authors acknowledge helpful conversations with Catherine Jahncke. Support was provided by the US Army Research Office through grants DAAH04-94-G-0156 and DAAH04-93-G-0194.

References

1. Yakobson, B.I. and M.A. Paesler, Ultramicroscopy 57 (1995) 241.

2. Yakobson, B.I. and M.A. Paesler, "Tip optics for illumination NSOM: Extended zone approach." Ultramicroscopy 57 (1995) 204.

3. A. Meixner, this volume

4. M. Libenson, this volume.

5. W.B. Allen, *Fiber Optics* (Plenum, NY) 1973.

6. A. LaRosa, C. Jahncke, B.I. Yakobson, and H.D. Hallen, Applied Physics Letters, in press.

7. A. LaRosa, C.L. Jahncke, and H.D. Hallen, Ultramicroscopy 57 (1995) 303.

8. Bethe, H.A., "Theory of Diffraction by Small Holes." Physical Review 66 (1944) 163.

9. Massey, G.A., "Microscopy and pattern generation with scanned evanescent waves." Applied Optics 23 (1984) 658.

10. Massey, G., J. Davis, S. Katnik and E. Omon, "Subwavelength resolution far-infrared microscopy." Applied Optics 24 (1985) 1498.

Figure captions

Figure 1. The sideput density for a conical probe. A bundle of rays of half-angle $\phi = 6°$ that diverge from being co-axial by an angle $\beta = 0°$ and that fills the conical taper is shown across the middle of the figure. The sideput density plot represents a measure of the interactions of the bundle of rays with the reflecting walls of the cone. The interactions can be tracked by following the image of the trajectory in the repeated zones.

Figure 2. Sideput density versus distance from the apex for a number of conical probes for a variety of (β, ϕ) values (given in degrees). While the plots differ somewhat, common to all is an indication of a large sideput density well removed from the taper apex.

Figure 3. Thermal conductance of a probe as a function of distance from probe apex (solid curve). The conductance falls as the fraction of (relatively high conductivity) Al decreases while that of (relatively low conductivity) glass increases. To model the behavior of thermal diffusion, the dash-dotted profile was chosen. Near the apex, conductance is assumed to be infinite, while at larger distances ($z > z_c = 160$ arbitrary units) a value approximately three times the value at $z = \infty$ is chosen.

Figure 4. Temporal response of a typical probe to a rapid 0.6 mW pulse of visible radiation as monitored by a 0.35 mW input of cw ir light. The infrared signal as a function of time is plotted at the top of the figure. The data fit a single exponential decay of $\tau = 0.02$ seconds quite well. The infrared signal as a function of frequency ν is shown at the bottom of the figure. Within the precision of the measurement $\nu_{knee} = \frac{1}{2\pi\tau}$.

Figure 5. Infrared throughput as a function of time for two probes. In the trace at the top (bottom) of the figure the ir intensity *increases* (*decreases*) with the introduction of the pulsed visible light and thus with increased temperature. The visible pulses with a one second period begin at a time $t = 5.5$s. Data imply that mechanisms resulting in both an enhancement and in a diminution of probe throughput must sometimes be dominant in NSOM probes.

Figure 6. Intensity dependence of the modulation of the infrared signal as a function of the input intensity. The intensity of the cw infrared light is held constant. At low intensities, linear scaling of the modulation with intensity suggests that neither the heat dissipation

mechanism nor the thermal/optical coupling balance changes with intensity in this region.

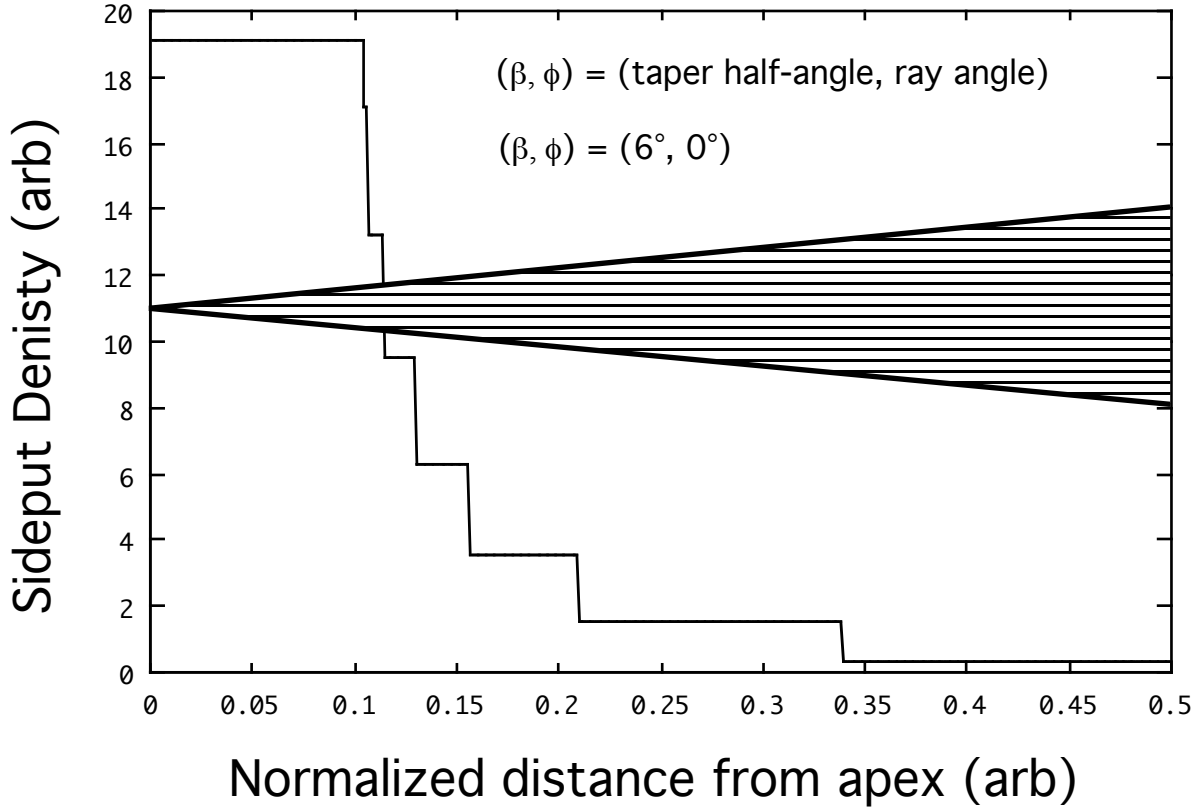
(β, ϕ) = (taper half-angle, ray angle)
(β, ϕ) = (6°, 0°)
Sideput Denisty (arb)
Normalized distance from apex (arb)
0
2
4
6
8
10
12
14
16
18
20
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5

Figure 1.

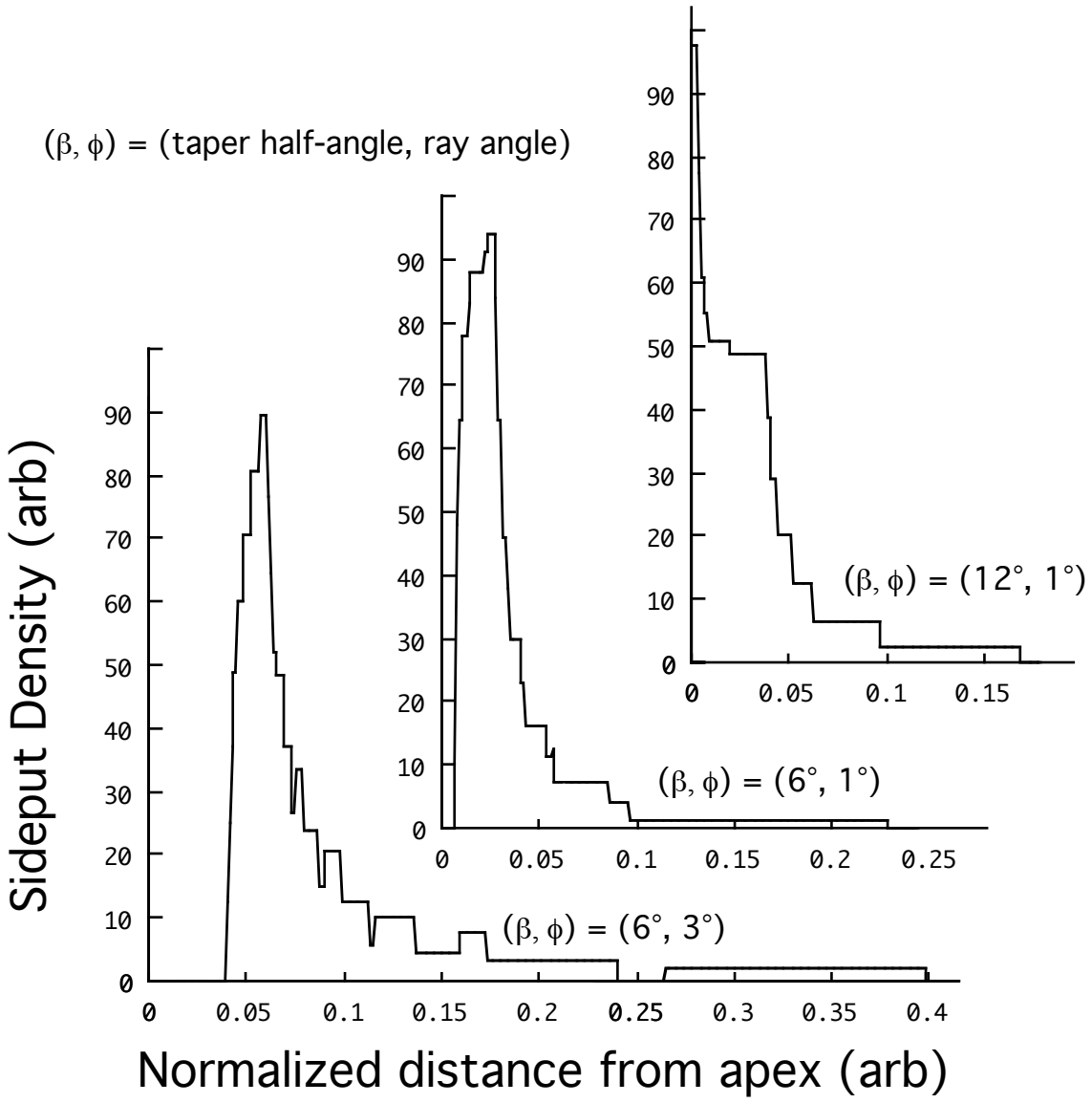
(β, ϕ) = (taper half-angle, ray angle)
Sideput Density (arb)
Normalized distance from apex (arb)
(β, ϕ) = (6°, 3°)
(β, ϕ) = (6°, 1°)
(β, ϕ) = (12°, 1°)

Figure 2.

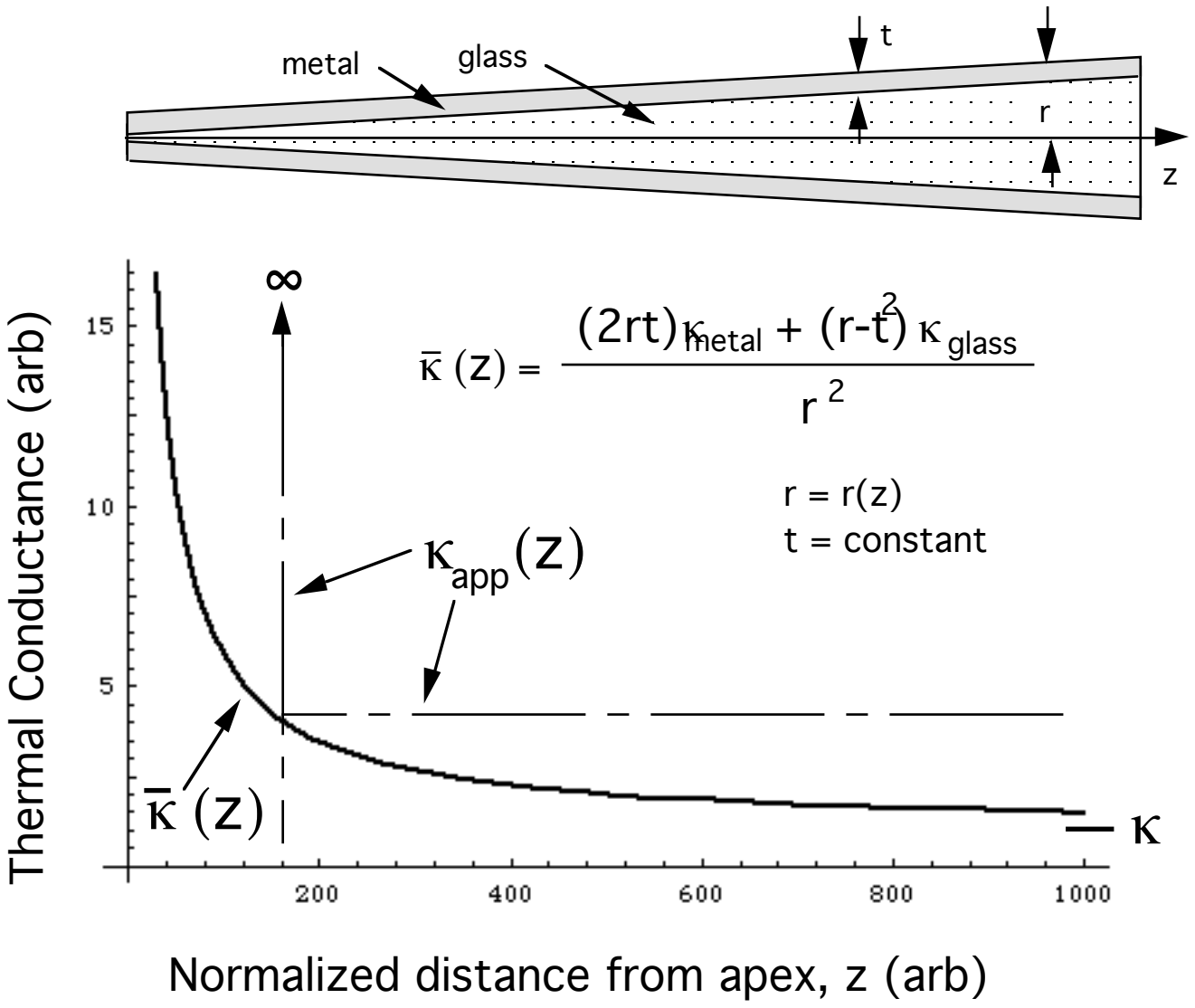
metal
glass
t
r
z
Thermal Conductance (arb)
15
10
5
200
400
600
800
1000
∞
$\bar{\kappa}(z) = \frac{(2rt)\kappa_{metal} + (r-t)^2 \kappa_{glass}}{r^2}$
r = r(z)
t = constant
$\kappa_{app}(z)$
$\bar{\kappa}(z)$
κ
Normalized distance from apex, z (arb)

Figure 3.

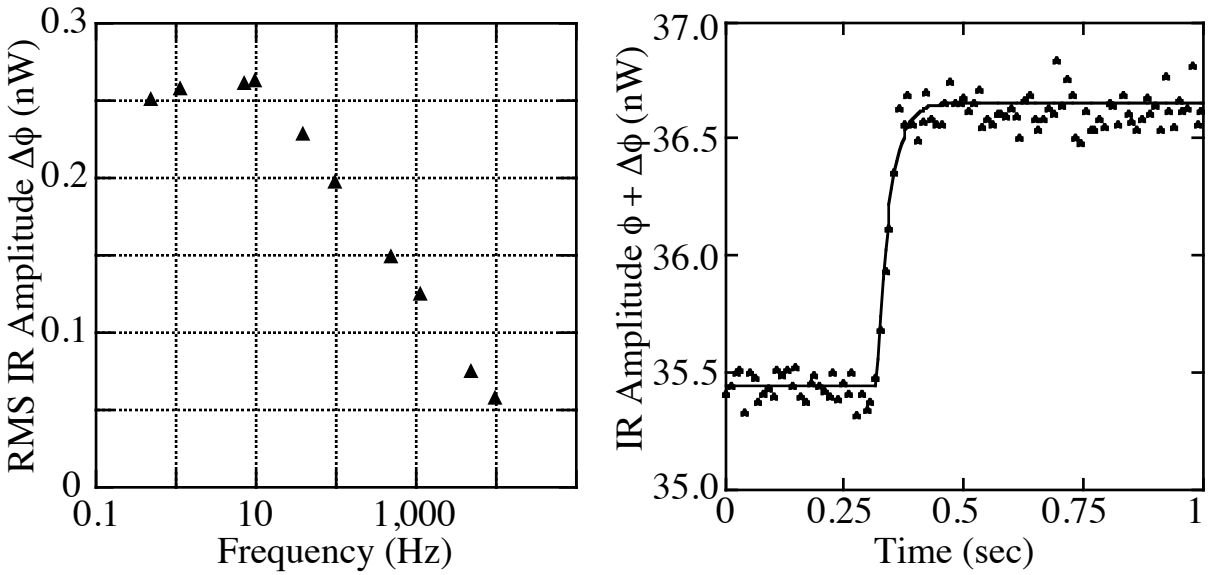
RMS IR Amplitude Δφ (nW)
0.3
0.2
0.1
0
0.1
10
1,000
Frequency (Hz)
IR Amplitude φ + Δφ (nW)
37.0
36.5
36.0
35.5
35.0
0
0.25
0.5
0.75
1
Time (sec)

Figure 4.

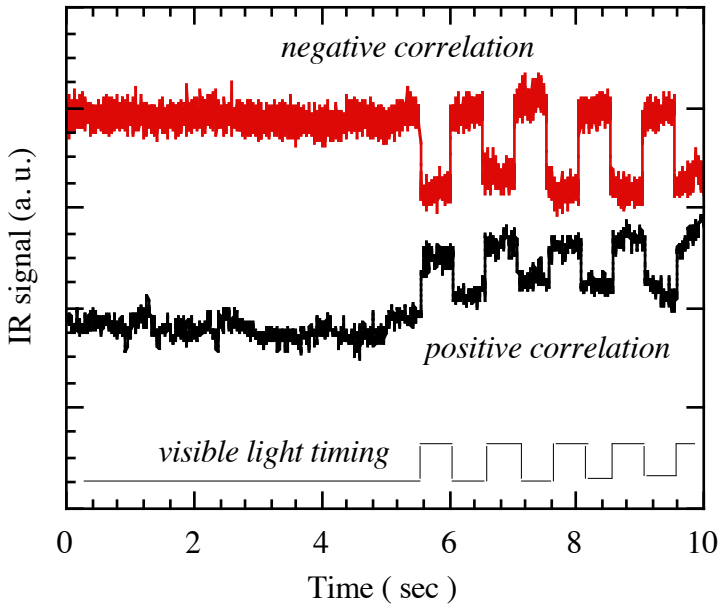
negative correlation
positive correlation
visible light timing
IR signal (a. u.)
0
2
4
6
8
10
Time (sec)

Figure 5.

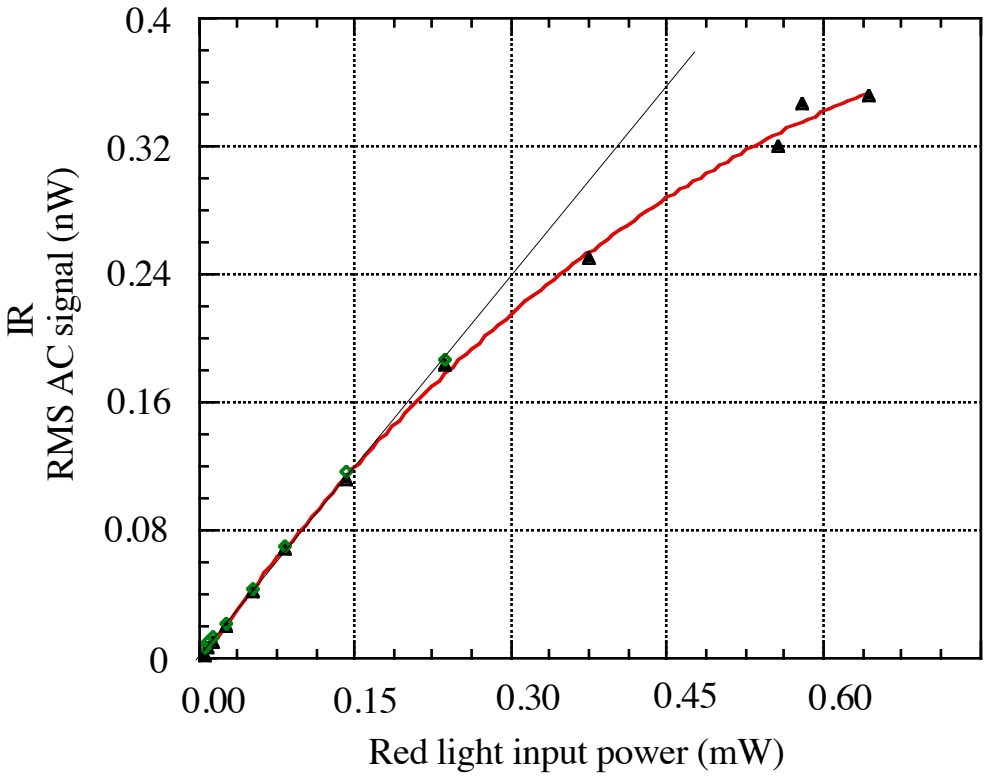
0.4
0.32
0.24
0.16
0.08
0
IR
RMS AC signal (nW)
0.00
0.15
0.30
0.45
0.60
Red light input power (mW)

Figure 6.