Resolving single fluorophores within dense ensembles: contrast limits of tip-enhanced fluorescence microscopy

Benjamin D. Mangum, Chun Mu, and Jordan M. Gerton

Department of Physics, University of Utah, 115 S. 1400 E., Salt Lake City, Utah. 84112
jgerton@physics.utah.edu

Abstract: We investigate the limits of one-photon fluorescence as a contrast mechanism in nanoscale-resolution tip-enhanced optical microscopy. Specifically, we examine the magnitude of tip-induced signal enhancement needed to resolve individual fluorophores within densely-packed ensembles. Modulation of fluorescence signals induced by an oscillating tip followed by demodulation with a lock-in amplifier increases image contrast by nearly two orders of magnitude. A theoretical model of this simple modulation/demodulation scheme predicts an optimal value for the tip-oscillation amplitude that agrees with experimental measurements. Further, as an important step toward the eventual application of tip-enhanced fluorescence microscopy to the nanoscale structural analysis of biomolecular systems, we show that requisite signal enhancement factors are within the capabilities of commercially available silicon tips.

© 2008 Optical Society of America

OCIS codes: (180.2520) Scanning microscopy; (180.4243) Near-field microscopy; (180.5810) Fluorescence microscopy.

References and links
1. Introduction

Tip-enhanced fluorescence microscopy (TEFM) is a type of apertureless near-field scanning optical microscopy (ANSOM) that utilizes fluorescence to generate an image. By aligning the sharp tip of an atomic force microscope (AFM) probe into the focus of a laser beam with axial polarization, enhanced fields are generated at the apex of the tip [1], as shown in Fig. 1. This field enhancement is tightly confined to the vicinity of the tip apex and has been shown to decay rapidly as \( r^{-6} \) with distance \( r \) from the tip apex [2]. These enhanced local fields can be used to beat Abbe’s diffraction limit, and various scattering processes (e.g. one- and two-photon fluorescence, Raman scattering, infrared spectroscopy, and Rayleigh scattering) have been used to image a range of samples with nanoscale resolution [2-12]. Much of the work with ANSOM to date has been on samples composed of isolated particles/molecules (e.g. fluorophores, quantum dots, nanotubes) due to the fact that ANSOM suffers from a relatively large background signal that arises from direct (non-enhanced) scattering from the laser beam. Thus, high density samples are challenging for ANSOM analysis since the background signal increases with the number of particles in the laser spot, while the tip-enhanced signal does not. This has so far prohibited the application of ANSOM to biological samples composed of a high density, heterogeneous ensemble of fluorescently-tagged biomolecules, including proteins, lipids, and nucleic acids.

Recently, a number groups have investigated various means of increasing the degree of field enhancement, including optimizing the shape of the tip to leverage plasmon and antenna resonances. These efforts have already been fruitful for increasing the enhancement, and will impact both ANSOM and sensor applications [13]. To complement these studies, it is also important to understand how much enhancement is required to image high-density samples with sufficient contrast to resolve individual molecules within the ensemble. It has been pointed out that for dense samples, the minimum (intensity) enhancement needed to achieve sufficient image contrast ultimately depends on the \( n^{th} \) root of the ratio of the area of the illuminated spot to the area under the tip, where \( n \) is the order of the scattering process being employed [13]. Naturally for linear scattering processes such as one-photon fluorescence, larger enhancement factors are needed compared to higher-order processes, such as two-photon fluorescence or Raman spectroscopy. In this paper we specifically investigate the limits of TEFM with regard to its potential for imaging high-density samples. In particular, we use a theoretical model based on experimental measurements to show that sufficient contrast can be obtained even for the relatively simple case of commercially available silicon tips and one-photon fluorescence.
Fig. 1. (Color Online) Experimental setup for TEFM. Labeled elements are as follows: He-Ne Laser — helium-neon laser (\(\lambda = 543\) nm); Mask — laser-beam mask; RPC — radial polarization converter; DM — dichroic mirror; OBJ — microscope objective; Probe — AFM probe; PZT — piezoelectric transducer; SF — spectral filters; APD — avalanche photodiode; LA — lock-in amplifier; DDS — digital synthesizer; PC — personal computer. The arrows indicate the polarization of the laser beam. Axial polarization at the sample plane can be achieved either by simply focusing a radially polarized laser beam, or by placing a laser beam mask before the microscope objective such that only super-critical rays are allowed to propagate. This focused total internal reflection fluorescence (TIRF) set-up is sometimes used because of its broadband capabilities and its large focal spot (\(\sim 1.5 \mu m \times 0.5 \mu m\)) lends itself to easy tip alignment, while radial polarization is preferred for smaller focal spots, \(\sim (250 \text{ nm})^2\).

2. Contrast in TEFM

In TEFM, the laser stimulates two distinct fluorescence signals: the far-field signal, \(S_{ff}\), resulting from direct illumination of fluorophores within the laser focus, and the near-field signal, \(S_{nf}\), resulting from field enhancement at the tip apex. The resolution of \(S_{ff}\) is at best diffraction limited, while \(S_{nf}\) has resolution given primarily by the sharpness of the tip [4]. Figure 2 shows a cartoon image composed of the superposition of \(S_{ff}\) and \(S_{nf}\) as well as a simulated profile through its center. While not shown, we also assume some noise in the far-field signal. Within this context, contrast \((C)\) and signal-to-noise ratio \((SNR)\) are defined as:

\[
C = \frac{S_{peak} - S_{ff}}{S_{ff}} = \frac{S_{nf}}{S_{ff}} \tag{1}
\]

\[
SNR = \frac{S_{peak} - S_{ff}}{\text{Noise in } S_{ff}} = \frac{S_{nf}}{\sigma_{ff}} \tag{2}
\]

where \(\sigma_{ff}\) is the standard deviation (noise) in the far-field background. The near-field signal originates from a small area on the sample surface \((a_{tip})\) given by the near-field interaction zone, which is determined mostly by the tip sharpness, while the far-field background originates from a much larger area \((A)\) given by the size of the laser focus. The total fluorescence signal for a given pixel of the raster-scanned image, \(S_{peak}\), is simply the sum of all photons collected during the pixel acquisition time \((\tau)\). The far-field signal \(S_{ff}\) is proportional to the number of
fluorophores in the focal area of the excitation beam, \( N_{FA} \), and also to a dimensionless parameter \( k \) that characterizes the total efficiency of the system: \( S_{ff} = k N_{FA} \).

The probability of an illuminated fluorophore emitting a photon follows a Poisson distribution, such that the expected average number of counts in the time interval \( \tau \) is simply \( S_{ff} \). The standard deviation is given by \( \sigma = \sqrt{S_{ff}} = \sqrt{k N_{FA}} \). In the limit of a single fluorophore in the near-field zone, \( S_{nf} = f k \), where \( f \) characterizes the fluorescence signal enhancement induced by the tip, and is a function of several parameters related primarily to its geometry and material properties. In this limit, the peak signal is given by \( S_{peak} = (f + N_{FA})k \). The overall system efficiency \( k \) is given by

\[
k = I_0 \times \sigma_0 \times \tau \times Q \times CE \times \frac{\lambda}{hc}
\]

where \( I_0 = P_0 / A \) is the intensity of the laser beam with power \( P_0 \) in a focal spot of area \( A \); \( \sigma_0 \) is the absorption cross-section of the fluorophore; \( \tau \) is the pixel acquisition time; \( Q \) is the quantum yield of the fluorophore; \( CE \) is the collection efficiency of the detection system; and \( hc/\lambda \) is the energy of a photon with wavelength \( \lambda \). A green He-Ne laser (\( \lambda = 543 \text{ nm} \)) was used for these experiments due to its low cost and the availability of fluorescent dyes and quantum dots with strong absorption at this wavelength. Although we have not done careful studies of tip-enhancement as a function of excitation wavelength, we do not expect a strong dependence since the dielectric function of silicon is fairly flat over visible wavelengths.

The lower limit for detection of a near-field signal arises from the requirement that the signal-to-noise ratio (SNR) be larger than unity,

\[
SNR = \frac{S_{nf}}{\sigma_{ff}} = f\sqrt{\frac{k}{N_{FA}}} > 1.
\]

Below this limit, the near-field signal is indistinguishable from stochastic fluctuations of the far-field background. On the other hand, to produce an image that can be interpreted visually dictates a more stringent requirement, namely that the contrast \( (C) \) be larger than unity,

\[
C = \frac{S_{nf}}{S_{ff}} = \frac{f}{N_{FA}} > 1.
\]

In this model, it is straightforward to evaluate the minimum enhancement required for sensitivity to a single fluorophore within a dense ensemble. A practical limit on density arises from
the requirement that the average spacing between fluorophores be no smaller than the microscope resolution, which is given by the near-field interaction zone. In this limit, \( N_{FA} = A/a_{tip} \), where \( A \) is the area of the laser focus. Using the focused-TIRF scheme described above, \( A = 0.75 \, \mu m^2 \) and \( a_{tip} = 100 \, nm^2 \), which suggests a signal enhancement of \( f > 7500 \) is needed to achieve contrast greater than unity. Employing a radially polarized laser beam yields a smaller focus spot, \( A = (250 \, nm)^2 \) \cite{14}, thus reducing the required enhancement to \( f \sim 600 \). Silicon tips are only capable of producing an enhancement factor of \( f \sim 20 \) \cite{4}, well below these requirements. Simple, non-optimized metal tips have been predicted to yield enhancement factors of \( f \sim 3000 \) \cite{1} and optimized metal tips that leverage antenna resonances may yield even larger enhancement factors. Although metal tips can produce much larger field enhancements than silicon, they also strongly quench fluorescence, leading to an overall reduction in the fluorescence signal and an associated decrease in the contrast. In several previous reports, silicon tips were found to yield the largest net contrast since no quenching was observed \cite{2, 4, 5}.

At first glance, the required signal enhancements predicted above cast a shadow on the potential application of TEFM to biological systems. As discussed below, however, the contrast can be improved dramatically by oscillating the AFM probe, which induces an associated modulation in the fluorescence signal, and by the subsequent application of a phase sensitive demodulation algorithm, such as lock-in amplification. Modulation/demodulation schemes are used widely in many areas of small signal processing and have also been used before in near-field microscopy \cite{2,4,5,15-18}. The analysis below demonstrates the limits of this approach for TEFM.

3. Improving contrast via phase sensitive demodulation

To calculate contrast and signal-to-noise ratio for the case of an oscillating tip, Eqs. (4) and (5) must be modified to account for the fact that the tip only intermittently contacts the sample at a particular phase of its oscillation cycle. To discuss the dependence of the near-field signal on the instantaneous height of the oscillating probe, it is useful to consider the arrival of each photon in a phase-space picture. In this scenario, each photon is assigned an angle \( \theta_i \) corresponding to the instantaneous phase of the sinusoidal tip-oscillation function at the time of detection (Fig. 3). The photon phases can be mapped to the corresponding tip-sample separation if desired.

Since the sample remains under direct laser illumination whether the tip is oscillating or not, the far-field signal for an oscillating tip is unchanged,

\[
S_{ff}^{osc} = S_{ff} = kN_{FA}.
\]

(6)

Multiple scattering of far-field photons between the tip and sample can lead to variations in the background intensity as a function of the tip height. However, these variations have been measured to be very small (\(< 5\%) for the tip-oscillation amplitudes employed here, and are thus neglected. Therefore, we assume that the far-field signal for an oscillating tip is unchanged compared to an absent tip or one which is in constant contact with the surface.

In phase-space, the maximum near-field signal occurs at a preferred phase \( \theta_p \) corresponding to tip-sample contact, and the photons are approximately Gaussian distributed around \( \theta_p \). To find the total number of near-field photons for a given pixel, \( S_n^{osc} \), the ratio \( \gamma \) defined as the number of photons collected in one oscillation cycle relative to the number that would have been collected had the tip been at the surface the entire time is calculated:

\[
\gamma = \frac{1}{2\pi} \int_{-\pi}^{\pi} \exp\left(-\frac{(\theta_i - \theta_p)^2}{2\theta_\sigma^2}\right) d\theta_i \approx \frac{\theta_\sigma}{\sqrt{2\pi}}
\]

(7)

where \( \theta_\sigma \) is the standard deviation of the photon-phase distribution, which can be obtained experimentally and is a function of oscillation amplitude. The approximation in Eq. (7) holds
in the limit that the integration limits are extended to $\pm \infty$, or equivalently when $\theta_\sigma < \pi/3$. The near-field signal for an oscillating tip is then given by

$$S_{nf}^{osc} = S_{nf} \gamma = kf \gamma = kf \frac{\theta_\sigma}{\sqrt{2\pi}}$$

(8)

Using the definitions for the oscillating signals in Eqs. (6) and (8), both the contrast and SNR for images produced by an oscillating tip (tapping mode TEFM) can now be calculated

$$C_{osc}^{sum} = \frac{f \gamma}{N_{FA}}$$

(9)

$$SNR_{osc}^{sum} = f \gamma \sqrt{\frac{k}{N_{FA}}}$$

(10)

where the subscript “sum” indicates a direct sum of the photon signals. Not surprisingly, without demodulation the contrast and SNR have been reduced by a factor of $\gamma$ compared to the non-oscillating scenario since the total number of near-field photons has decreased.

Lock-in amplification is a particularly powerful phase-sensitive demodulation technique that decomposes a modulated signal into real and imaginary components that are proportional to the cosine and sine projections in phase space, respectively. In TEFM, each detected fluorescence photon can be viewed as a unit vector pointing in the direction $\theta_i$ equal to the instantaneous phase of the tip oscillation at the time of detection (Fig. 4). In this picture, a lock-in amplifier simply performs a vector addition of the detected photons transmitted through its internal
bandpass filter. If the resultant lock-in vector $\mathbf{L}$ is divided into near-field (NF) and far-field (FF) components, both of which are vector sums, then the lock-in signal is simply the magnitude $|\mathbf{L}| = |\mathbf{NF} + \mathbf{FF}|$.

The far-field component of the lock-in vector $\mathbf{FF}$ results from an unbiased two-dimensional random walk with unit steps, and follows the probability distribution originally derived by Lord Rayleigh

$$P(r) = \frac{2r}{N_{\text{steps}}} e^{-r^2/N_{\text{steps}}}$$

where $r$ is the final end-to-end distance of the walk, and $N_{\text{steps}}$ is the number of steps in the walk [19]. This distribution has a mean $\mu_r$ and standard deviation $\sigma_r$ given by

$$\mu_r = \sqrt{\frac{\pi N_{\text{steps}}}{4}}$$

$$\sigma_r = \frac{1}{2} \sqrt{N_{\text{steps}}(4 - \pi)}.$$

In our case, $N_{\text{steps}}$ is given by the number of detected far-field photons that are transmitted by the lock-in bandpass filter, $N_{\text{steps}} = \beta \times S_{\text{ff}}$, where $\beta < 1$. This gives

$$|\mathbf{FF}| = \sqrt{\frac{\pi}{4}} \beta N_{\text{FA}}$$

$$\sigma_{|\mathbf{FF}|} = \frac{1}{2} \sqrt{\beta k N_{\text{FA}}(4 - \pi)}.$$

for the average length of the far-field component $|\mathbf{FF}|$ and its uncertainty $\sigma_{|\mathbf{FF}|}$, respectively. The near-field component $\mathbf{NF}$ comes from a biased random walk about $\theta_p$. The average value of its magnitude $|\mathbf{NF}|$ can be estimated by projecting the unit vectors corresponding to each near-field photon onto the $\theta_p$ axis and then summing the result:

$$|\mathbf{NF}| = \sum_i \cos(\theta_i - \theta_p) = S_{nf}^{\text{tot}} \times \langle \cos(\theta_i - \theta_p) \rangle.$$
where the sum runs over all the near-field photons, \( i = 1 \rightarrow S_{NF}^{osc} \). For simplification we define \( \alpha = \langle \cos(\theta_i - \theta_p) \rangle \). Since the phase of each photon \( \theta_i \) is Gaussian distributed, the normalized expectation value is

\[
\alpha = \frac{\int_0^\pi \cos(\theta_i - \theta_p) e^{-(\theta_i - \theta_p)^2/2\sigma^2} d\theta_i}{\int_0^\pi e^{-(\theta_i - \theta_p)^2/2\sigma^2} d\theta_i} \approx e^{-\theta_p^2/2}. \tag{17}
\]

Combining this result with the definition of \( \gamma \) from Eq. (7), the average magnitude of the near-field component \( |NF| \) is then approximated by

\[
|NF| = kf \gamma \alpha \approx kf \frac{\theta_p}{\sqrt{2\pi}} e^{-\theta_p^2/2}. \tag{18}
\]

When using a lock-in amplifier to demodulate the signal, an image is constructed one pixel at a time, where the value of each pixel is the magnitude of the lock-in vector, \( |L| = |NF + FF| \). The near-field component \( NF \) points along \( \theta_p \), but the far-field component \( FF \) points in a random direction. Performing the vector addition of \( NF + FF \) and averaging over all directions for \( FF \), the peak lock-in signal is given by

\[
|L|_{peak} = \sqrt{|NF|^2 + |FF|^2} = \sqrt{(kf \gamma \alpha)^2 + \frac{\pi}{4} k N_{FA} \beta}. \tag{19}
\]

The contrast \( CLI \) and signal-to-noise ratio \( SNR_{LI} \) in the lock-in signal can now be found.

\[
CLI = \frac{|L|_{peak} - |FF|}{|FF|} = \left[ \frac{4k f (\gamma \alpha)^2}{\pi N_{FA} \beta} + 1 \right]^{1/2} - 1 \tag{20}
\]

\[
SNR_{LI} = \frac{|L|_{peak} - |FF|}{\sigma_{|FF|}} \approx 2CLI \tag{21}
\]

Equation (20) can be used to calculate the minimum signal enhancement factor required to achieve contrast greater than unity:

\[
CLI > 1 \Rightarrow f > \frac{1}{\alpha \gamma} \sqrt{\frac{3\pi N_{FA} \beta}{4k}}. \tag{22}
\]

As before, we consider the case where there is only one fluorophore in the near-field zone (\(~10,000\) fluorophores/\(\mu m^2\)) and the far-field illumination area is \(~0.5 \mu m \times 1.5 \mu m\) corresponding to focused-TIRF illumination. Using typical experimental values for \( k = 10 \) and \( \beta = 0.15 \) as well as optimized values for \( \gamma = 0.4 \) and \( \alpha = 0.6 \) (see below) gives a required signal-enhancement factor of \( f > 65 \) to achieve a contrast greater than unity. Using radial polarization reduces the required enhancement to \( f > 18 \) which is very realistic for silicon tips and in fact has already been demonstrated in the case of isolated spherical quantum dots [4].

Figure 5 demonstrates how the lock-in demodulation scheme can be used to improve the contrast and SNR for samples with a high density of rod-shaped quantum dots (4 nm x 9 nm). These images were obtained using a silicon tip oscillating with an optimized amplitude of \(~30\) nm peak-to-peak (see below) and focused-TIRF illumination (\(\lambda = 543\) nm). Approach curve measurements where the tip is lowered onto isolated quantum dots and the fluorescence rate is measured as a function of tip-sample separation (data not shown) indicate an enhancement
factor of only $f \sim 4$ for these data. The small enhancement in this case results from the fact that the elongated shape of the quantum dots leads to a somewhat small spatial overlap with the region of enhanced field at the tip apex. Furthermore, the absorption dipole for these nanorods should lie predominantly along the sample surface, while the enhanced field is strongest under the tip where it is vertically polarized. This leads to relatively weak near-field excitation of the nanorods.

Our model assumes that the fluorophores, whether quantum dots or fluorescent molecules, do not blink or photobleach. In reality, both quantum dots and molecular fluorophores blink and photobleach, which alters the contrast observed in experimental images. In particular, the background signal $S_{ff}$ will be reduced for a blinking or photobleaching sample compared to an ideal one. Interestingly, this has the effect of increasing the contrast in experimental images in the limit of large fluorophore densities where the fluctuations in the far-field signal caused by blinking and bleaching are small compared to the total far-field signal $S_{ff}$. However, the probability of the tip encountering a particular fluorophore that is “on” (i.e. not in a dark or photobleached state) is reduced by the same factor as the far-field signal $S_{ff}$. The consequence of this is difficult to predict without knowledge of the blinking and photobleaching rates corresponding to the particular fluorophores of interest. This issue is highlighted by Fig. 5, which shows the topographic image of a collection of quantum dots in addition to the undemodulated and demodulated near-field fluorescence images. The total quantum dot density as observed by the AFM topography is $\sim 50 \mu m^{-2}$, however, many of the quantum dots do not fluoresce. The bright quantum dot density for this image is $\sim 14 \mu m^{-2}$ and there is clearly sufficient contrast to increase the density further; Eq. 22 predicts that a density as high as 26 bright dots/μm$^2$ can be achieved for $f \sim 4$.

Fig. 5. TEFM images of a high-density quantum dot sample. Panel (a) shows the AFM topography ($\sim 50$ total dots/μm$^2$). Panel (b) shows the scalar photon sum ($\sim 14$ bright dots/μm$^2$). Panel (c) shows the same image after lock-in demodulation. The scale bar is 200 nm.

4. Optimizing tip oscillation amplitude

The lock-in contrast and signal-to-noise ratio given in Eqs. (20) and (21) are strongly influenced by the amplitude of oscillation of the AFM tip, which determines the width of the Gaussian photon-phase distribution, $\theta_{\sigma}$, and thus the values of $\gamma$ and $\alpha$. Thus, to optimize the lock-in contrast, the product $\gamma \times \alpha$ must be maximized with respect to $\theta_{\sigma}$:

$$\frac{d}{d\theta_{\sigma}} \left( \theta_{\sigma} e^{-\theta_{\sigma}^2/2} \right) = 0$$

(23)

where the approximations in Eq. (18) have been used. Solving Eq. (23) for $\theta_{\sigma}$ gives an optimal value of $\theta_{\sigma}^{opt} = 1$ radian. The optimal oscillation amplitude, $A_{opt}$, can now be found using the equation of motion for the tip oscillation, $z = A(1 - \cos(\theta))$. To relate $\theta_{\sigma}^{opt}$ to an optimal amplitude $A_{opt}$, we define $z_{\sigma}$ as the value of tip-sample separation $z$ in an approach curve such that the integrated area under the approach curve from $0 \rightarrow z_{\sigma}$ contains 68% of the near-field
photons. The value of \( z_\sigma \) depends on the sharpness of the tip and the size and shape of the fluorescent object: sharp tips and small objects yield the smallest values of \( z_\sigma \). Substituting \( z = z_\sigma \) and \( \theta = \theta_{\sigma} \) into the equation of motion for the tip we obtain:

\[
A_{opt} = \frac{z_\sigma}{1 - \cos(1)} \approx 2.1z_\sigma. \tag{24}
\]

When the approximations made in Eq. (18) are used, a value of \( A_{opt} = 2.18z_\sigma \) is obtained compared to a value of \( A_{opt} = 2.11z_\sigma \) when complete numerical integrations are performed.

Experimental values for the contrast and signal-to-noise ratio as a function of the peak-to-peak oscillation amplitude of the tip are shown in Fig. 6, along with the theoretical predictions developed above. Isolated \((N_{FA} = 1)\) CdSe/ZnS nanorods \((4 \text{ nm} \times 9.4 \text{ nm})\) were imaged with different amplitudes using many different tips from the same fabrication wafer. Each data point was computed from the measured values of \( S_{peak} \), \( S_{ff} \), and \( \sigma_{ff} \), as used in Eqs. (1) and (2) for \( \sim 15 \) different quantum dots [6]. The values of \( f = 3.7 \pm 1.3 \), \( k = 11 \pm 5 \), \( \beta = 0.15 \pm 0.15 \), and \( z_\sigma = 7.5 \pm 2 \text{ nm} \) were all obtained from a statistical analysis of image and approach curve data. Subsequently, \( \theta_\sigma \) was computed from Eq. (24) using the measured value \( z_\sigma = 7.5 \pm 2 \text{ nm} \) to obtain \( \gamma \) and \( \alpha \) for each oscillation amplitude from Eqs. (7) and (17). Thus, the theoretical curves shown in Fig. 6 contain no free parameters whatsoever. The predicted peak-to-peak amplitude of \( 32 \pm 9 \text{ nm} \) agrees with the experimental value of \( 32 \pm 4 \text{ nm} \). This good agreement between the predictions of this theoretical model and experimental measurements lends confidence to the calculated values of the signal enhancement factors \( f \) requisite for imaging high fluorophore densities found above.

![Fig. 6. TEFM image contrast, panel (a), and signal-to-noise ratio, panel (b), for isolated quantum dots as a function of the tip oscillation amplitude. Data were obtained using BudgetSensors Multi-75 silicon tips. Data points correspond to the average value of \( \sim 15 \) measurements for the lock-in demodulation signal (closed symbols) and the scalar sum (open symbols). Dashed and dotted lines are the corresponding theoretical predictions.](image)

5. Summary of contrast limitations

We have defined the acceptable level of near-field contrast as \( C > 1 \) and have calculated the amount of signal enhancement needed in TEFM to achieve such contrast for single fluorophores within high density samples. The particular density used in our calculations was 10,000 fluorophores/\( \mu \text{m}^2 \), which corresponds to only one fluorophore of the ensemble within the measured near-field zone \((10 \times 10 \text{ nm}^2)\). Our model uses no free parameters, but rather extracts the values for the relevant parameters from experimental measurements. The model has
been validated in part by its agreement with experimental results; mathematical optimization of the tip oscillation amplitude matches experimental measurements, as seen in Fig. 6. Using this model, we have considered the two cases of a near-field probe that is not oscillating vertically above the sample surface (contact-mode AFM imaging) and one that is (tapping mode imaging) for two experimentally-relevant illumination conditions, focused-TIRF and radial polarization. For contact-mode imaging, the requisite signal enhancement factors were calculated to be $f \sim 7500$ for focused-TIRF illumination and $f \sim 600$ for radial polarization. Both of these values are well beyond the maximum measured enhancement of $f \sim 20$ for Si tips. Tapping-mode imaging coupled with lock-in demodulation significantly increases image contrast, thus reducing the requisite signal enhancement factors to $f \sim 65$ for focused-TIRF and $f \sim 18$ for radial polarization. This last case is within the capabilities of commercially available Si AFM tips. Thus we expect that the maximum density achievable with Si tips is not limited by the enhancement factor, but rather by the requirement that each fluorophore be spatially resolved from its neighbors, in this case, at least 10 nm apart.

6. Conclusions

Determining the structure of extended biomolecular networks, and relating that structure to the physical mechanisms underlying various biological functions, are very difficult and pressing problems in molecular-scale science. Current nanoscale structural analysis tools including x-ray crystallography, electron microscopy, and atomic force microscopy, have a number of limitations that prevent their application to extended networks composed of heterogeneous mixtures of various biomolecules. Fluorescence microscopy, on the other hand, is a very powerful technique for analyzing heterogeneous molecular systems, and when combined with the spatial resolution afforded by apertureless near-field microscopy, holds great promise as a future molecular-scale structural analysis tool.

Although the potential of apertureless fluorescence microscopy in structural biology has been recognized previously, a recurring criticism has been that first-order scattering processes cannot achieve the contrast needed to resolve individual molecules within a dense ensemble. In this work, we have explicitly addressed this issue and have shown both theoretically and experimentally, that it is in fact possible to achieve the needed contrast using carefully designed modulation/demodulation schemes. The key issue discussed was the need to optimize various experimental parameters, such as the oscillation amplitude and material properties of the apertureless tip. Coupled with recent and future advances in scanning probe microscopy, such as imaging in water and fast frame imaging speeds, it may ultimately be possible to combine optical resolution approaching that of electron microscopy with the ability to image bio-molecules in physiological conditions.

Acknowledgments

The authors thank Changan Xie, Jonathan Cox, David Goldenberg, and Eyal Shafran for helpful discussions. This work was supported in part by a Cottrell Scholar Award from the Research Corporation.