Near field scanning luminescence and photothermal microscopy

Jon Rogers
Near Field Scanning Luminescence
and Photothermal Microscopy

by

Jon K. Rogers

B.S., California Polytechnic State University (1992)

A thesis submitted in partial fulfillment of the
requirements for the degree of Master of Science
in the Center for Imaging Science
at the

ROCHESTER INSTITUTE OF TECHNOLOGY

November 1994

Signature of Author

Jon K. Rogers
21 November 1994

Accepted by

Dr. Dana G. Marsh
Coordinator, M.S. Degree Program
Certificate of Approval

M.S. Degree Thesis

This M.S. Degree Thesis of Jon K. Rogers has been examined and approved by the thesis committee as satisfactory for the thesis requirement for the Master of Science degree.

Thesis Advisor: Dr. Mehdi Vaez-Iravani

Thesis Committee Member: Dr. Pantazis Mouroulis

Thesis Committee Member: Dr. Zoran Ninkov

DECEMBER 1, 1994

Date
Thesis Release Permission Form

Rochester Institute of Technology
Center for Imaging Science

Title of Thesis: Near Field Scanning Luminescence and Photothermal Microscopy

I, Jon K. Rogers, hereby grant the Rochester Institute of Technology permission to reproduce and to distribute copies of this thesis document in whole or in part. Any reproduction will not be for commercial use or profit.

Jon K. Rogers
Near Field Scanning Luminescence
and Photothermal Microscopy

by

Jon K. Rogers

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in the Center for Imaging Science at the Rochester Institute of Technology

ABSTRACT

A near field optical scanning probe microscope with force regulation is presented. The microscope force regulation uses a differential interferometer for monitoring the tip frequency while the tip provides a sub-wavelength aperture providing a system which simultaneously records topographical information (by the force microscope) and an optical image. The flexibility of the system is evident in the specific applications pursued in this research. Simultaneous force and luminescence images derived from the investigation of porous silicon are presented. We observed variation in luminescence over sub-micron distances. Furthermore, variation in the spectral distribution of small particles of porous silicon was also observed. Both these results support the quantum wire theory proposed to explain the luminescence properties of the porous silicon and its formation.

The same system was slightly reconfigured to provide imaging of thermal properties of microcircuitry. The imaging was performed by detecting changes in reflectivity as a function of temperature. The first near field photothermal probe microscope is demonstrated along with a simultaneous scanning force microscope. The system shows high resolution of thermal signal, but needs some noise reduction techniques to improve image quality.
ACKNOWLEDGEMENTS

I wish to thank my advisor, Dr. Mehdi Vaez-Iravani, for his constant support. I would also like to thank Mr. Ricardo Toledo-Crow, for without his aid, many elements of the project would have gone awry. I want to acknowledge the help and cooperation of the RIT Microelectronic Engineering Department, and the department head, Dr. Fuller. In particular, I would like to thank the Equipment Engineers, Mr. Tom Grimsley, Mr. Paul Waldrop and Mr. Scott Blondell, for the use of their department’s equipment. I thank Dr. Santosh Kurinac for her aid in my understanding of porous silicon. Thanks to Mr. Fred Seiferth, Mr. Jens Von Behren, Mr. Sid Drakash, and Mr. Joe Summa who all provided the various samples used to complete this research. I would like to thank the Mechanical Engineering Department’s Tool Makers, Mr. Jim Graenier, Mr. Tom Locke, and Mr. Dave Hathaway, for their help in machining parts for the system. I would especially like to thank the inventor of super glue and the Loctite Corporation for their contribution to our project. Finally, I thank my family, and friends who have been behind me the whole way.
DEDICATION

I dedicate this work to my father, Mr. John Rogers. Without your support, I would not have made it this far. And to my wife, Kathleen, for her undying support through the long nights of research.
## Contents

1 Introduction ............................................. 1
   1.1 Objectives ........................................... 1
   1.2 Near Field Optics ..................................... 3
   1.3 Background ........................................... 13
   1.4 Scanning Force Microscopy and Near Field Optical Microscopy: The System 22
   1.5 Benefits and Limitations of the Near Field System .................. 26

2 Correlative Imaging of Porous Silicon Topography and Photoluminescence 29
   2.1 Introduction ............................................ 29
   2.2 Photon Emission and Bandgap Theory .......................... 30
   2.3 Theories on Porous Silicon and Its Luminescence Properties .......... 32
   2.4 Anodization of Silicon in Hydrofluoric Acid .................. 34
   2.5 Imaging System Configuration ............................. 36
      2.5.1 Image Interpretation ................................. 39
   2.6 Images, Spectra and Discussion ................................ 44
      2.6.1 Samples ............................................. 44
      2.6.2 Sample Specific Interpretation Considerations ............. 44
      2.6.3 N-Type Images ....................................... 46
      2.6.4 P-Type Images ....................................... 50
      2.6.5 Spectra ............................................. 64
2.7 Porous Silicon Particles ........................................ 69
  2.7.1 Images ....................................................... 71
  2.7.2 Spectra ...................................................... 75
2.8 Conclusions and Further Research .............................. 80

3 Photothermal Imaging with SFM and NSOM .................. 81
  3.1 Introduction .................................................. 81
  3.2 System Configuration and Imaging Considerations .......... 86
  3.3 Images, Results and Discussion ............................. 93
  3.4 Conclusions ................................................... 101

4 Conclusions ....................................................... 102

A Fabrication of Fiber Optic Probes for Simultaneous SFM and NSOM 104
  A.1 Tip Fabrication ............................................... 104
         A.1.1 Tip Performance Concerns .............................. 106
  A.2 Coating Thickness on SFM/NSOM Tips ........................ 108

B Photomultiplier Detectors and Associated Noise ............... 111
  B.1 Photomultiplier Tubes (PMT) ................................ 111
  B.2 Noise .......................................................... 112
  B.3 Noise in photomultiplier tubes. ............................. 114
  B.4 Minimum Detectable Power for Dark Current Limit and Shot Noise Limit 116
List of Figures

1-1 Diagrams of Diffraction angle calculations (a) and Numerical Aperture calculations (b). ........................................ 4
1-2 Schematic of oblique illumination. Allows for maximum angle between the two collected orders for maximum resolution. ....................... 5
1-3 Schematic of a propagating planar wave. Dashed lines represent wavefronts, perpendicular to the direction of propagation. .......................... 8
1-4 Square modulus of evanescent components for different distances. z in terms of wavelengths. .............................................. 10
1-5 Square modulus of propagating components for different distances. z in terms of wavelengths. .............................................. 11
1-6 Square modulus of both propagating and evanescent components for different distances. z in terms of wavelengths. ...................... 11
1-7 Plots of edge reproduction for different tip sample spacing. z in terms of wavelengths. ................................................. 13
1-8 Schematic of setup for the optical stethoscope by Pohl et al. (a) before tip formation, (b) overall configuration, (c) aperture formation. .............. 15
1-9 Experimental setup for force gradient measurement (after Martin et al.). .. 18
1-10 Differential interferometric method of feedback for "tipless" Scanning Force Microscopy. ..................................................... 20
1-11 Schematic for resonance frequency calculation. .......................... 23
1-12 Setup used for reflection mode NSOM with "tipless" interferometric regulation of tip and sample spacing. D1 and D2 are photodetectors.

2-1 Simplified Energy Diagram of a Semiconductor.

2-2 Electrolytic cell for porous silicon anodization.

2-3 Schematic of SFM/NSOM setup for porous silicon analysis.

2-4 Force (a, top and bottom), luminescence (b), and direct reflected (c) images of a demarcation between polished and porous silicon on an n-type wafer.

2-5 Schematic of an idealized tip/sample orientation to illustrate the problem of false reproductions in optical signal collection.

2-6 Force (left) and reflected optical (right) scans of polysilicon lines on silicon. White light illumination.

2-7 Force (left) and reflected optical (right) scans of polysilicon lines on silicon. Coherent illumination.

2-8 Cross sectional schematic diagram of a macropore and the effect of non-absorbed source light on luminescence. The collection microscope will see both primary and secondary luminescence.

2-9 Force (a, top and bottom), luminescence (b) and direct reflected (c) images of n-type P-Si.

2-10 Force (a) and luminescence images of n-type P-Si.

2-11 Force (a), luminescence (b) and direct reflected (c) images of n-type P-Si.

2-12 Force (a) and luminescence (b) of thin film region of p-type P-Si.

2-13 Micrograph of p-type P-Si surface.

2-14 Force (a), luminescence (b), and direct reflected (c) images of p-type P-Si. Each pair shows a 24x12 micron area.

2-15 Force (a), luminescence (b) and direct reflected (c) images of p-type P-Si. White boxes on left side scans indicate region of right side scans.

2-16 Force (a, top and bottom), luminescence (b) and direct reflected (c) images of p-type P-Si.
2-17 Force (a, top and bottom), luminescence (b) and direct reflected (c) images of p-type P-Si. ......................................................... 58
2-18 Force (a, top and bottom), luminescence (b) and direct reflected (c) images of p-type P-Si. ......................................................... 60
2-19 Force (a) and luminescence (b) images of p-type P-Si showing ultra-high resolution. ................................................................. 61
2-20 Force (a) and luminescence (b) images of p-type P-Si showing "halo" like reproduction of a macropore. ........................................... 62
2-21 Force (a) and luminescence (b) and reflected optical (c) images of p-type P-Si showing macro columnar structures. ......................... 63
2-22 Force (a), and luminescence (b) images of p-type P-Si, showing flat region between ridges. ......................................................... 65
2-23 Schematic of SFM/NSOM setup for spectral analysis of porous silicon. ................................................................. 66
2-24 Spectrum from ridge region of p-type P-Si. Peak at 675nm. ............... 67
2-25 Spectrum from thin film region of p-type P-Si. Peak at 672nm. ............... 68
2-26 Transmission configuration for P-Si dust investigation. For spectral analysis D1 was replaced with monochromator and cooled CCD array. 70
2-27 Force (a) and luminescence (b) images of P-Si dust particle. ............... 72
2-28 Force (a) and luminescence (b) images of P-Si dust particles. ............... 73
2-29 Force (a) and luminescence (b) images of P-Si dust particles. ............... 74
2-30 Force (a) and luminescence (b) images of P-Si dust particles. Large range scan. 76
2-31 Force (a) and luminescence (b) images of P-Si dust particles showing extreme localization of luminescence. ................................. 77
2-32 Spectra from 3 different dust particles. For shape comparison only. ....... 78
2-33 Macro spectra from (a) etched, and (b) scraped regions. ........................ 79
3-1 Various detection schemes for thermal microscopy. ............................ 83
3-2 Various detection schemes for thermal microscopy. ............................ 84
3-3 Schematic of layer structure used in reflectivity change as a function of heat. 87
3-4 SFM/NSOM setup for photothermal investigation. ........................................... 94
3-5 Far field scanning microscope images showing direct reflected (a) and pho-
    tothermal (b) signals. Current modulation frequency at 1kHz. .................... 96
3-6 Far field scanning microscope images showing direct reflected (a) and pho-
    tothermal (b) images using current frequencies of 158Hz (bottom) and 1kHz
    (top). ............................................................................................................. 97
3-7 Near field force (a) and photothermal (b) images of circuit. Current frequency
    at 20KHz. ...................................................................................................... 99
3-8 Near field direct reflected (a) and photothermal (b) images of the circuit.
    Current modulation frequency at 50kHz. ................................................... 100

A-1 Isometric perspective diagram of tip rotator, front (a) and side (b) views,
    and schematic of evaporation procedure setup (c). ................................. 105
A-2 Schematics of different tip geometries and corresponding performance. .... 107
A-3 Schematic of spherical radiating source, for coating calculations. .......... 108
Chapter 1

Introduction

1.1 Objectives

The primary goals of this research are to describe a system which can simultaneously image both optical characteristics and topography on a microscopic level. The general system can be configured for many possible applications, ranging from the study of tunneling current variation to phase contrast imaging. At this point in research, many applications of the system have been proposed and implemented, while other configurations remain ideas, still to be tested. Many of the systems also have the ability to be multi-faceted. That is, a system, originally configured for reflection imaging, may be easily modified to perform transmission imaging. Particular samples warrant the desire for different investigations, and the flexibility of the systems can help to achieve a thorough investigation of the particular sample in question. The research here introduces two new applications of a reflection imaging near field microscope system. We will present the general setup, and describe the workings of each element. The system is flexible, so for the different samples studied, a description of any minor setup modifications and their impact on the resultant imaging will be presented. The research hopes to show that the system designed and utilized here can provide new insights and can be easily reconfigured to provide sample-specific information. The system has been used in an investigation of porous silicon and photothermal imaging. We establish
the importance of the sub-wavelength resolution of the system for each application, and compare that to information currently achieved using typical far-field lens systems.

Our first application was the investigation of porous silicon, a form of silicon that emits light under certain conditions. Up to now, research has concentrated primarily on the macro-analysis of this material. One current topic in semiconductor technology is finding a cost-efficient method of incorporating light sources into silicon chips. One idea is to create localized areas of porous silicon, a room-temperature photoluminescent material, within the devices. For this reason, there is a need for a thorough microscopic analysis of the material. Furthermore, semiconductor light sources can typically be used as detectors as well. Because silicon is the most used semiconductor today, learning about its light emitting or collecting qualities is an important area of study. This research hopes to show that the described system provides a superior way of analyzing porous silicon, and may provide information leading to the confirmation of certain theories pertaining to the luminescent processes of the material.

This thesis will further describe a similar system configuration and its use in photothermal imaging. This will show the system versatility, as well as adding a new method of photothermal investigation of thermally active samples. Photothermal imaging is important in a variety of fields, including both surface and sub-surface imaging. It is particularly relevant in today’s computer world, for as the critical dimensions of microchip circuitry are reduced, thermal effects of current in those circuits become an ever increasing problem. Heating of the wires from current crowding and electron migration can weaken the wires within circuitry, leading to failure. The system described here will generate highly localized thermal images of electrical circuitry. It may also lead to a better understanding of the problem of heating and weakness in microcircuitry, and to methods of evaluating possible solutions.
1.2 Near Field Optics

The demand for more memory per unit space, more circuitry per unit space, or better resolution imaging, has created the search for imaging methods that resolve beyond the conventional far field microscope. To calculate the resolution limits of a far field microscope, we use a simple diffraction angle analysis. Light impinging on the sample will be diffracted at angles related to the wavelength of the incoming light $\lambda$, and the smallest sample detail, $d$ (see figure 1-1). The diffraction angle of the $m^{th}$ order is given by:

$$\theta_m = \sin^{-1}\left(\frac{m\lambda}{d}\right)$$  \hspace{1cm} (1.1)

In a far field system, to resolve detail, the collection lens must collect at least two orders. These orders will produce interference whose resultant period is equal to the sample dimensions. The collection angle of the microscope lens is finite and associated with the numerical aperture (NA) which is defined by:

$$NA = n \sin(\alpha)$$  \hspace{1cm} (1.2)

where $n$ is the index of refraction of air (the medium surrounding the lens) and $\alpha$ is the maximum acceptance half-angle (see figure 1-1). Comparing the acceptance angle and the diffraction angle equations, we can find the maximum resolution possible for a given lens. Using oblique illumination, the zeroth order light will enter the lens at one edge. The first order, then, will be diffracted and collected by the lens on the other edge (see figure 1-2). Then we see that

$$d_{\text{min}} = \frac{\lambda}{\sin(\theta)} = \frac{\lambda}{\sin(2\alpha)}$$  \hspace{1cm} (1.3)
(a) Diffraction angle
\[ \theta_m = \sin^{-1}(m\lambda/d) \]

(b) Numerical Aperture
\[ NA = \sin(\alpha) \]

Figure 1-1: Diagrams of Diffraction angle calculations (a) and Numerical Aperture calculations (b).
Figure 1-2: Schematic of oblique illumination. Allows for maximum angle between the two collected orders for maximum resolution.

The \( \sin(2\alpha) \) can be approximated by \( 2\sin\alpha \) for small angles.\(^1\) So putting that approximation in, we see that

\[
d_{\text{min}} = \frac{\lambda}{2 \sin(\alpha)} = \frac{\lambda}{2NA}
\]

(1.4)

For example, a typical 10x objective with a numerical aperture of 0.25, and illumination centered around 500nm would show detail down to about 1 micron. To achieve better resolution, there are two methods. One is using shorter wavelength illumination. As the wavelength decreases, we see in equation 1.4, that the minimum resolvable distance, d, will also decrease. This has been achieved using x-rays and electron beam imaging. X-ray radiation varies in wavelength from about 1nm to about 5pm (picometers). This range is potentially a few hundred times resolution improvement over visible light imaging (400nm-700nm). Electron beams wavelengths are shorter than x-rays, and can theoretically give still better resolution. In fact, the primary cause of resolution limitations in electron microscopy is lens aberrations. These methods achieve good results but are not useful for all samples

\(^1\)Because \( \sin(2\alpha) = 2 \sin\alpha \cos\alpha \), for \( \alpha < 0.1 \), \( \cos\alpha > 0.99 \). We can use the approximation \( \sin(2\alpha) \approx 2 \sin\alpha \) for these rough calculations.
and do not provide the flexibility of simultaneously imaging multiple parameters. Another common method is to use oil-immersion lenses to improve the numerical aperture by raising the index of refraction within the lens field. But again, the lens aberrations and wavelength of illumination still define the primary resolution limitations of the system.

Yet another option is near field imaging. Recall that the above calculations assumed a far field lens system, and far field detection. The near field, which can be defined as within a fraction of a wavelength of the illuminated sample, the light does not exhibit normal diffraction effects. In fact, it is shown below, that the light remains "collimated" for a distance equal to the aperture diameter (depending on the strictness of the word collimated). This leads to the idea of near field microscopy.

High frequency components of the light modulated by the sample characteristics contain the fine detail, beyond classical diffraction limited system resolution. However, these components decay rapidly during propagation and are called evanescent waves. Because of the exponential decay, these waves can only be observed if the detection is performed very near the modulation region, the "near field."[1] If these waves are detected, the resolution is no longer defined by the diffraction angle, wavelength, or lens quality, but primarily by the detection area, and its separation from the sample.

The following derivation shows the spatial distribution of light (both evanescent and propagating waves) emerging from a small (sub-wavelength) aperture2, and analyzes the expected resolution of an edge, as a function of separation between sample and detection. Note that the following derivation includes many assumptions, the primary one being that the system can adequately be analyzed using scalar mathematics.

We start by assuming a planar, monochromatic wave is travelling in the +z direction. The electric field can be written as

\[ u(x, z < 0, t) = \exp[-i(kz - \omega t)] \]

\[ (1.5) \]

2We will see that this derivation can be used as an approximation for the action of the tips used in near field microscopy.
where \( k \) is the wavenumber, given by \( \frac{2\pi}{\lambda} \), \( \lambda \) is the wavelength of the incoming light, and \( \omega \) is the angular frequency given by \( 2\pi\nu \) where \( \nu \) is the temporal frequency. We can disregard the time dependent part of the equation because its sinusoidal variations will average out over time.

For this analysis, we approximate the aperture as a linear object whose transmission profile can be written as \( \tau(x) \). The field just beyond this object, setting the object position to \( z=0 \), is given by

\[
u(x, 0^+) = \tau(x) \exp[-i(kz)]\big|_{z=0^+} = \tau(x) \quad (1.6)
\]

We can decompose this into its frequency components and their relative amplitudes using the Fourier Transform.

\[
U(\xi, 0^+) = \int_{-\infty}^{\infty} u(x, 0^+) \exp[-i(2\pi x \xi)] dx = \int_{-\infty}^{\infty} \tau(x) \exp[-i(2\pi x \xi)] dx \quad (1.7)
\]

Notice that the \( U(\xi, 0^+) \), and \( u(x, 0^+) \) are transform pairs. Also notice that the final expression is the transform of the object’s transmission function. So we can express \( \tau(x) \) and \( \Upsilon(\xi) \) as transform pairs recognizing that

\[
\Upsilon(\xi) = \int_{-\infty}^{\infty} \tau(x) \exp[-i(2\pi x \xi)] dx \quad (1.8)
\]

This also leads to a relation between the electric field and the transform of the transmission function. By inverse transforming the above equality, and combining with equation 1.6, we get

\[
u(x, 0^+) = \int_{-\infty}^{\infty} \Upsilon(\xi) \exp[i2\pi x \xi] dx \quad (1.9)
\]

The Fourier Transform can be thought of as a method of procuring spatial frequency data from the image data. But in optics, the frequency domain expression can also be analyzed as a decomposition of the propagating wavefront into a superposition of planar waves each
Figure 1-3: Schematic of a propagating planar wave. Dashed lines represent wavefronts, perpendicular to the direction of propagation.

with a different amplitude and angle of propagation (determined by $\xi$). Each plane wave is expressed as a propagating plane wave in the form

$$\exp[2\pi i (x\xi + z\eta)]$$

in which

$$\xi^2 + \eta^2 = \frac{1}{\lambda^2}$$

If $\lambda$ is the wavelength in the direction of propagation, then $\frac{1}{\xi}$ represents the wavelength in the $x$ direction, and $\frac{1}{\eta}$ the wavelength in the $z$ direction (see figure 1-3). So at a point $z$ beyond the origin, the field is given by

$$u(x, z > 0) = \int_{-\infty}^{\infty} \gamma(\xi) \exp[2\pi i (x\xi + z\eta)] \, dx$$

Solving for $\eta$ using the geometric relationships and substituting into equation 1.12, we arrive
at the equation for the amplitude distribution at a point $z$.

\[
\begin{align*}
u(x, z_>0) &= \int_{-\infty}^{\infty} \mathcal{T}(\xi) \exp \left[ i2\pi \left( x\xi + z \left( \frac{\sqrt{1-(\lambda \xi)^2}}{\lambda} \right) \right) \right] dx \\
&= \int_{-\infty}^{\infty} \mathcal{T}(\xi) \exp \left[ i2\pi x\xi \right] \exp \left[ ikz \left( \frac{\sqrt{1-(\lambda \xi)^2}}{\lambda} \right) \right] dx
\end{align*}
\]

(1.13)

This can be separated into two pieces: (1) homogeneous wave components for which $\lambda \xi < 1$. These waves propagate in both $x$ and $z$ directions. (2) inhomogeneous wave components for which $\lambda \xi > 1$. These propagate in $x$ but decay exponentially in $z$. The inhomogeneous wave components are the evanescent waves. This breaks equation 1.13 down to the following:

\[
\begin{align*}
u(x, z_>0) &= \int_{\lambda \xi \leq 1} \mathcal{T}(\xi) \exp \left[ i2\pi x\xi \right] \exp \left[ ikz \left( \frac{\sqrt{1-(\lambda \xi)^2}}{\lambda} \right) \right] dx \\
&\quad + \int_{\lambda \xi > 1} \mathcal{T}(\xi) \exp \left[ i2\pi x\xi \right] \exp \left[ ikz \left( \frac{\sqrt{1-(\lambda \xi)^2}}{\lambda} \right) \right] dx
\end{align*}
\]

(1.14)

This equation gives the spatial distribution of light as it propagates through an aperture with a transmission function $\tau(x)$.

Numerical solutions were performed assuming a one dimensional case in which the tip aperture was $\tau(x)=\text{rect}(x)^3$, and the corresponding transform, $\mathcal{T}(\xi)=\text{sinc}(\xi)^4$. The figures included show the progress of evanescent and propagating waves and the combination of the two over varying distances from the aperture. The aperture was taken to be one-fifth of the wavelength and results were plotted using Matlab, matrix manipulation software. The plots in figure 1-4 show some sharp details (high frequency information) retained in the evanescent waves. Notice, also, that the evanescent waves do not propagate far. After about $\frac{\lambda}{3}$, the average intensity is down to nearly zero. Alternatively, the propagating terms (in figure 1-5) start out smooth and slowly dissipate. To see the entire contents of the output light, we look at the figure 1-6 which is the combination of the evanescent waves and the propagating terms (square modulus). This indicates that we have collimation for about

\[
^3\text{rect}(x) = \begin{cases} 
0, & |x| < \frac{1}{2} \\
\frac{1}{2}, & |x| = \frac{1}{2} \\
1, & |x| > \frac{1}{2}
\end{cases}
\]

\[
^4\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}.
\]
Figure 1-4: Square modulus of evanescent components for different distances. $z$ in terms of wavelengths.

$\frac{\lambda}{2}$. This can be used as a general rule, that the distance of collimation is roughly equal to the aperture diameter. This collimation length roughly defines the near field region. Although this analysis gives a reasonable description of the situation, for completeness, we continued from this point, to achieve results for the reproduction of an edge by the theoretical aperture, for different separations.

This edge represents a boundary between a perfectly transparent region and a completely opaque region. As the aperture scans across an edge, the detected signal will be the convolution\(^5\) of the two entities, aperture and sample. This can be easily calculated

\(^5\)The convolution of $f(x)$ and $h(x)$ can be defined by the integral

$$f(x) \ast h(x) = \int_{-\infty}^{\infty} f(\alpha)h(x - \alpha)d\alpha$$
Figure 1-5: Square modulus of propagating components for different distances, $z$ in terms of wavelengths.

Figure 1-6: Square modulus of both propagating and evanescent components for different distances, $z$ in terms of wavelengths.
using the previous results. We can simply convolve the amplitude with a step function.

By taking the amplitude at each of the different z values calculated above, we will see how the reproduced edge is affected by distance from the tip.

Mathematically, this can be written as

\[ g(x, z) = u(x, z) \ast \text{step}(x, z = 0) \]  \hspace{1cm} (1.15)

We can move to the frequency domain, by simply replacing the convolution (\( \ast \)) with a product. This implies

\[ G(\xi, z) = U(\xi, z)F\{\text{step}(x)\} \]  \hspace{1cm} (1.16)

where \( G(\xi, z) \), and \( U(\xi, z) \) are corresponding transform pairs of \( g(x, z) \) and \( u(x, z) \). To get the intensity, we simply calculate the square modulus of the amplitude by the following

\[ i(x, z) = u(x, z) \cdot u^*(x, z) \]  \hspace{1cm} (1.17)

where \( u^*(x, z) \) represents the complex conjugate of \( u(x, z) \).

This was calculated and the results are plotted in figure 1-7. Notice how quickly the resolution goes down, with increasing aperture/detector spacing. For this reason, we want our aperture to be as close as possible to the sample.

It should be noted that the concept of resolution in the near field is a subject of intense theoretical and experimental research. The above analysis is simply a scalar analysis. A more accurate and rigorous analysis would require vector space mathematics. These more rigorous approaches have been performed theoretically starting in 1944 with H.A. Bethe's "Theory of the Diffraction by Small Holes".[2]

In addition to the theoretical and experimental investigations that are currently being performed regarding resolution, there are situations in which the sample itself will influence

\[ \text{step}(x) = \begin{cases} 
0, & x < 0 \\
0.5, & x = 0 \\
1, & x > 0 
\end{cases} \]
the resolution. The above analysis gives rough guidelines for the expected resolution of a "typical" tip and system, but realistically, these concerns of sample influence must be individually accounted for when analyzing each set of data. It should also be noted that resolution is not the only parameter of concern. There are far field systems whose resolution capability rivals some Near Field Scanning Optical Microscope (NSOM) systems. However, these systems will typically have a large background illumination which can make imaging of certain parameters impossible. Using a near field system allows for a highly localized source light and low background illumination. For imaging concerns in fluorescence experiments, and others, this will make the difference between an impossible experiment and useful images.

1.3 Background

Near field optical microscopy started with the "optical stethoscopy" of D.W. Pohl, W. Denk, and M. Lanz, in 1983[3]. The idea was derived from previous research by Ash and Nichols, of
electromagnetic radiation in imaging which found sub-wavelength resolution systems using microwaves[4]. Before the work by Pohl et al, no one had established a method to make a sub-wavelength aperture for visible light. Such apertures are now routinely made by pulling a length of fiber to a sharp tip (radii ranging from 10nm to 100nm), and coating the sides with aluminum (See Appendix A).

Pohl’s method used a quartz sample which was cut down to approximately 10mm length and 2mm width. Anisotropic etching with HF created a pointed tip on one end of the crystal, whose radius was about 30 nm. They coated the entire piece with a layer of chromium, a layer of aluminum and a final layer of gold to prevent oxidation. Afterpolishing the back end of the quartz and setting this up on a bimorph (a piezo-bending micro positioner), they could launch laser light into the back. Because the tip had been coated over in the process, they needed a way to open up a tiny hole at the tip. They were able to accomplish this by the following method. Using a sample consisting of e-beam micro-lithographed test patterns (lines ranging from 2000nm to 60nm), they brought the tip into contact with the sample using the bimorph. They monitored contact by looking for a closed circuit between the tip’s aluminum coating and the sample’s foil lines. Once the contact signal was observed, the bimorph was pushed forward slightly more to exert pressure on the tip. Because of the small area at the tip, only a small force was required to bring the pressure above the plastic deformation threshold. The metal film became thinner and thinner until it was transparent. Monitoring this through a microscope, they stopped the bimorph’s forward motion as soon as they could see light through the tip. They proceeded to scan the tip in a raster fashion. For each point the tip was pushed into contact with the sample, and the light was collected by a microscope on the back side of the sample (see figure 1-8). The tip was then lifted and moved to the next point of detection, and placed back on the sample. This signal was received by a photo-multiplier tube. Once a reading was taken, the tip was moved off the sample, scanned in the x direction a given distance, then placed back on the sample for the next point’s reading. The results were conclusive in that they had reached sub-wavelength resolutions. In fact, the resolution of this system was 10-20nm, nearing that of conventional
scanning electron microscopes (SEM) without the necessity of a conducting sample, or the vacuum viewing conditions. Also, Pohl proposed using this setup on biological media as a way to probe into the bulk of a soft tissue or fluid sample[3].

The next experiment of importance in the field of NSOM was Massey et al studying sub-wavelength resolution in the far-infrared[5]. The purpose of this study was to show that the evanescent waves can be used for imaging. They used the evanescent waves created by 118μm far-infrared illumination of a 10μm hole in aluminum foil. With two identical pieces of foil, each with a 10μm pinhole, one served as the source aperture and the other served as an object. They scanned these two identical holes (a convolution) and recorded the evanescent wave power emitted through the "object" hole. Several scans were tested, each
with a different separation between the two pieces of foil. The best results had resolutions of about $20\mu m$, or $\frac{1}{5}$. The results of this experimental work were in close agreement with Massey’s theoretical paper on the evanescent waves[6]. These evanescent waves fall off exponentially. To detect them, the detector must be within a wavelength of the sample light (within the "near field" regime). This was accomplished above with a pair of foil pinholes. But the distance regulation was to be a concern for further developments of the optical stethoscope (soon to be renamed Near Field Scanning Optical Microscopy or NSOM).

In 1986, Pohl et al published another paper which expanded on the earlier work and included more complete results of the experimentation. They called the system a near field optical-scanning (NFOS) microscope. It was argued that cost considerations and ease of use might allow NFOS imaging to serve as a replacement for SEM imaging in certain cases. However, the biggest benefit of NFOS, it was claimed, was that NFOS could examine objects with resolution near the resolution of SEM, operate in air or fluids, and did not require a conductive sample. This would allow for in situ investigation of biological samples, without coating the sample with gold or requiring a vacuum environment. The aperture used here was the same metal coated quartz tip as mentioned before. However, a new method of controlling the tip-sample spacing was explored which was based on the scanning tunneling microscopy (STM), introduced earlier by Binnig and Rohrer[7]. Pohl successfully incorporated an STM mechanism into his system[8]. The basic principle behind STM is the following. A voltage difference is setup between the conducting tip and the back side of the sample, which, when the sample is in close proximity to the tip, allows current tunneling. This tunneling current goes up as the tip and sample approach each other. This signal can be sent through a series of feedback electronics which control the z-axis positioner of the sample. This will maintain a constant tip-sample spacing. This meant that the system, used by Pohl and others, could raster scan and keep the tip a given distance away from the sample using the tunneling current as a feedback signal. Because the quartz tip was already coated with a metal, they could apply a voltage to that and observe the tunneling current.
One of the benefits of this method of tip position regulation is that non-contact imaging could be performed. Previous methods actually dragged a tip across the sample. Simply the force of the sample would push the tip up. An interferometric detection of the tip height could be recorded, thereby establishing a topographical map. In STM and attractive mode force microscopy (explained below), the tip and the sample never actually contact which makes this set up ideal for looking at biological media and other soft media that would be damaged by a contact method of imaging.

The next important development was the first demonstration of non-contact atomic force microscopy (AFM, also called SFM\textsuperscript{7}) by Y. Martin, C.C. Williams, and H.K. Wickramasinghe[9]. The tip, made of tungsten, was vibrated at its resonance and brought in towards the sample. Atomic forces from the sample would damp the oscillation of the tip. This oscillation was monitored by a laser heterodyne interferometer (see figure 1-9). Knowing the structure of the cantilever, the vibration amplitude was measured as a function of the tip-sample distance. From this, the gradient of the force at the tip was deduced, as well as the force itself. This force gradient could also be used as a feedback signal to keep the tip-sample spacing constant, like the STM mentioned above. In this feedback mode, the correction signal sent to the z-piezo was recorded, resulting in a profile of the sample in question.

Another mode of operation used by Martin \textit{et al}[9], was as follows. They noted that the long range force is an attractive one. As the tip is brought in toward the sample, at some point it will be pulled toward the sample by this attractive Van der Waals force. Once it reaches a potential minimum, then the force required to remove the tip can be measured which will lead to a derivation of the peak strength of the attractive force, and a characterization of the sample material. They were able to achieve resolutions of 5nm

\textsuperscript{7}True non-contact atomic force microscopy (AFM) can only be performed in a vacuum environment. When a sample is in the presence of air, normal atmosphere, there is a micro contamination layer that lays on top of it. This micro layer causes drag forces that damp amplitude and frequency. The atomic forces do affect the tip vibration, but in air, the contamination drag forces are more significant. For this reason, we will call our system Scanning Force Microscopy, because we are using all forces, atomic and otherwise, to retard the tip's oscillation and detect our proximity to the sample.
Figure 1-9: Experimental setup for force gradient measurement (after Martin et al.).
for the topographical maps. Several of the above methods were later combined to give different benefits, including combined STM and optical stethoscopy. Betzig, Issacson and Lewis set up a method by which the light was not piped through the tip, but collected by it. A light source was imaged through a microscope onto the sample, then the tip served as a collection aperture[10]. There was an earlier demonstration of reflection near field microscopy. Light was scattered through a small aperture. This light was reflected off the sample back through the aperture, and was collected by a optical microscope[11]. This group used the STM method of tip-sample separation regulation.

Meanwhile Reddick, Warmack and Ferrell, and others[12],[13] pursued the evanescent wave approach. These works were considering the possibility of using evanescent waves for imaging, as opposed to using the tunneling current or the optical signal. The general approach was to deposit the sample on a glass substrate. Light launched into the substrate would, without the presence of the sample, experience total internal reflection (TIR). The TIR would leave the evanescent waves to penetrate through the sample[6]. As the tip was scanned across the sample, the strength of the evanescent waves (exponentially decaying) was used as the feedback signal to control the z-piezo. It was an analogous system to STM, but it took advantage of the evanescent waves instead of the tunneling current. This was called the Photon Scanning Tunneling Microscope (PSTM) or Evanescent Field Optical Microscope (EFOM).

The next breakthrough in the field was by P.C. Yang, Y. Chen and M. Vaez-Iravani[14], who devised an orthogonal cantilever/sample configuration for SFM. Previous methods employed a cantilevered tip. The attractive force of the sample was in the same direction as the AC vibration of the tip. In this setup, tip crashing was a common problem. For situations where the gradient of the force exerted on the tip is greater than the cantilever stiffness, there is an unstable position and a "jump-to-contact" is likely to occur. In an effort

---

6Note, that initially, TIR may extinguish all but the evanescent waves. But once the sample is placed into the system, the conditions for TIR will change. For this reason, it would be more accurate to say that imaging was achieved using "primarily" exponentially decaying evanescent waves which were collected with some background illumination, due to changed TIR conditions.
to solve some of these problems inherent with the previous SFM methodology, Iravani's group offered the straight cantilever configuration (or "tipless"). In this case, the tip is vibrated in the x-y direction. Instead of getting a jump-to-contact (which could result in a broken tip), the drag and atomic forces simply damp the vibration of the tip. This tip frequency and amplitude was monitored by a differential interferometric system (see figure 1-10).

Feedback electronics were set up such that when the amplitude of the vibration was reduced due to the force damping, the tip would pull back. Once the vibration returned to its maximum value, it would push forward slightly. This set up an equilibrium position for the tip, so that, as the tip was raster scanned across the sample, the tip-sample spacing remained constant. As before, the correction signal was monitored to give a topographical map. Another benefit of the system described here is that, provided the tip is made of
light guiding material, one can perform simultaneous near field optical and atomic force microscopy. Indeed, in 1992, R. Toledo-Crow et al used this method to publish the first paper describing a near field optical microscope with orthogonal tip force regulation[15]. This provided both topographical and optical images simultaneously. Because the images were taken from the same portion of the sample, correlative analysis could be used to provide more information than either image alone.

E. Betzig devised yet another tip-sample spacing regulation method. This method relies on the "detection of shear forces between the end of a near-field probe and the sample of interest."[16] In Betzig’s system, developed in 1992, they independently but simultaneously had chosen an orthogonal tip configuration similar to the system developed by Yang, Chen and Iravani[14]. Laser light was piped down the fiber and through the tip. There was a collection microscope behind the sample. The tip was vibrated near its resonance. Small shear forces between the sample and tip during engagement caused the resonance frequency of the system to shift. This shift was measured by analyzing the light collected by the microscope. The microscope imaged the light emitted by the probe onto a position sensitive detector. The resulting DC signal was recorded to give the NSOM image (transmission as a function of position), and the AC portion of the signal could be demodulated to give either phase or amplitude information, which would be used as the feedback signal.

However, this method limited the system’s capability to imaging samples which were transparent and whose transmissivity was roughly uniform. Furthermore, any slight variation in opacity, that may not have been seen previously, would influence the recorded topographical image. One final problem was that any region of sufficient opacity could render the feedback system useless and might cause a crash between the tip and the sample[15]. It should be noted that in further work, Betzig et al reconfigured the system to include an external method of tip oscillation detection, in which the tip’s shadow is cast onto a split detector which results in an AC signal due to the tip’s vibration. This detection method overcomes the problem of sample dependence; however, the system and signal must be recalibrated for each new tip.
The work mentioned above is only a highlight of the work that has been completed in the field. As mentioned before, several regulation methods have been proposed and implemented[19]-[25]. But the force regulation has now been universally adopted by the NSOM community as the method of choice for maintaining a constant tip/sample spacing separation.

1.4 Scanning Force Microscopy and Near Field Optical Microscopy: The System

Simultaneous imaging is achieved by making the probe for one modality act as the probe for another modality. The following discussion explains the basic workings of the system used for this research. At the heart of the system is the tip. The tips are pulled from fiber optic lengths. The method is a simple heat and pull procedure which softens the fiber then stretches the now malleable fiber out until a tip forms and air cooling solidifies it again. (see Appendix A.1). This conical tip is then coated with aluminum (see Appendix A.2) creating a final tip diameter (with coating) of about 200nm for the force resolution with a 20nm aperture for the optical point source. Light is piped down the fiber and exits through the 20nm aperture. This produces a virtual point source at the end of this tip. The tip is placed on a bimorph which is driven with a high frequency sine wave signal. Because the resonance curve is roughly Lorentzian, we operate at just below the resonance where the curve is linear which gives the best sensitivity. The resonance is a function of the bulk modulus of the material (the fiber), and the spring constant[26]. Assuming a cantilever with mass \( m \) for a lever of length \( l \) (see figure 1-11), the spring constant can be calculated from

\[ k = \frac{3mgl}{9L^3} \]

This is not the ultimate resolution. Depending on the exact tip size and shape, and the coating characteristics, resolutions can be in the tens of nanometers range. But to have good optical resolution, the tip must be sufficiently coated to prevent light from exiting the fiber anywhere but at the tip. Imperfections during coating can lead to pinholes along the side of the tip. The thick coating that we use tends to eliminate these pinholes, but does reduce our force resolution to the above indicated, around 100-200nm.
Figure 1-11: Schematic for resonance frequency calculation.

\[ k = \frac{3EI}{l^3} \]  

(1.18)

and so resonance frequency of the tip is at

\[ f = \frac{1}{2\pi} \sqrt{\frac{k}{0.23m}} \]  

(1.19)

where \( E \) = Young's Modulus, \( I = \int x^2 dA \), \( A \) is the cross sectional area and \( x \) is the direction of motion. Because \( E \) and \( I \) remain relatively constant between tips, the primary concern in the calculation is the length of fiber that is free to vibrate, \( l \).

To monitor this vibration in frequency and amplitude, we use a differential interferometric detection method. A HeNe laser is passed through a Wollaston prism, and is focused on the side of the tip (see figure 1-12). The polarization of the incoming laser is adjusted so it impinges on the Wollaston at 45° to its axes. The corresponding polarizations split into two separate parallel beams which propagate through the lens and are focused on the side of the tip. The separation of these two beams along the tip is an important factor in the sensitivity. Ideally, we would like to have one spot far from the tip, where the amplitude of vibration is small, and one very near the tip. This would produce the best sensitivity. In the setup used for this research, the spots are \( \sim 100\mu m \) apart. Each beam is reflected off the tip and propagates back through the lens and the Wollaston prism. The beam splitter sends the now co-propagating beams to a photodetector. Because the beams traverse different
path lengths, they have a different phase. This phase shift introduces a polarization rotation. When the beams are recombined, a crossed polarized analyzer is placed just before the photodetector, allowing only the interfering parts to pass. This interference modulation frequency corresponds to that of the tip, and is detected by a photodetector. The signal from the photodetector is sent through a lock-in amplifier, where its phase and amplitude are measured. When the sample and tip come into close proximity (engagement), the force gradient of the sample will change the effective stiffness of the tip, causing the resonance of the system to shift up. This shift results in a decrease in the signal at the detector. This signal is compared with a reference voltage by a differential amplifier, and the output serves as the feedback. As the signal drops below the reference, the feedback pulls the sample away from the tip. The signal begins to rise and the feedback stabilizes. This insures a constant tip-sample spacing. The tip can then be raster scanned across the sample. The feedback signal is recorded during the scan and provides the topographical information about the sample.

With the tip-sample spacing regulated by the interactive forces, the optical signal is generated by collecting the light which interacts with the sample (see figure 1-12). The system can be configured for reflection or transmission, and can be configured to detect any contrast mechanism already used with far field microscopy. The tip can act as either a source or a collector. As a collector, the sample would be illuminated by a "large" spot (a few microns in diameter), and the collected light would be that which was reflected off, or transmitted through the sample and launched into the tip. The resolution of the optical signal would be defined by the aperture of collection, the tip aperture.

For this research, we used the tip as the source, "illumination mode". In this method, as diagramed above, the source light is launched into the fiber and exits at the tip aperture. The collection microscope, in reflection or transmission, serves to collect any light that was acted on by the sample. The resolution, again, is defined by the spot size, the tip aperture. The light collected by the far field scope is detected and monitored by a photomultiplier tube (PMT) or some other fast, high sensitivity detector.
Figure 1-12: Setup used for reflection mode NSOM with "tipless" interferometric regulation of tip and sample spacing. D1 and D2 are photodetectors.

Both the force and the optical signal can be simultaneously recorded. In fact, the system is not limited to two contrast modalities. With the proper equipment and configurations, the system could conceivably collect four or five different modalities, simultaneously. The simultaneous imaging can lead to a better understanding of a given sample's chemical and physical characteristics.
1.5 Benefits and Limitations of the Near Field System

With this relatively new system of imaging, there are still concerns of applicability. Is the system beneficial to all areas of research? What does this system provide, that previous methods cannot? To answer these kinds of questions, we must look at the benefits of the system, as well as some inherent limitations.

Unlike Scanning Electron Microscopy (SEM), whose resolution is typically superior to NSOM, but requires a vacuum environment and a conducting sample, NSOM can be performed in air, with virtually any sample. Specifically, the degree of sample opacity or transmissivity is not an obstacle for successful imaging, resulting in great flexibility. Furthermore, because we are using light, instead of electron beams, to perform the imaging, the sample need not be conducting. We can use all contrast methods already in practice with far field microscopy, such as, reflectivity, transmissivity, polarization, phase contrast imaging, etc.

The system is a non-contact imaging system. This allows for non-destructive imaging of biological samples, and other pliable, fragile materials. But, one of the biggest benefits of the system is its inherent capability for correlative imaging. Since only one tip is required to perform both topographical mapping and optical imaging, we can collect at least two images from the identical part of the sample, each with different characteristics. For image analysis, sample identification, and unambiguous imaging, this technique can be very helpful.

Despite these benefits, there are inherent limitations to the system as well. One common problem with the system is complex image interpretation. Because the tip is functioning as the probe for both force and optical images, there is the chance that one is influencing the other. That is, if a certain structure detected by the force probe changes the tip/sample spacing in such a way as to affect the optical resolution, the optical image will show some detail, that does not exist. With regular topographies, these discrepancies can be easily identified and accounted for. However, with more complex topographies, identifying the errors can be much more difficult (as we will see with porous silicon).

Another limitation is the available photon count. Because the tip is small, the aperture
is only letting a small amount of light out, typically a few nanowatts. This must then be acted on by the sample, leaving less than a nanowatt for detection. This varies with sample and the parameter being detected, but it means that images must be taken with little background illumination, and, often, with long integration times. This can lead to position shifts due to instabilities in the piezo scanners.

Our system is also limited to samples with topography depth of no greater than a few microns. The bimorph that holds the sample performs all the z-direction motion during a scan. This bimorph has a maximum travel of a few microns, which imposes this depth limitation. Furthermore, on samples with deep topographies, there is much more chance of causing damage to the tip. The feedback electronics have a limited reaction time. If the scan is sufficiently rapid on a sample with large topography, the feedback may be too slow to keep the tip/sample spacing constant and could lead to a crash between the tip and sample. For these reasons, we limit our studies to samples with topographical depths of less than a few microns.

For research purposes, the entire system is constructed on a threaded optical table, with individual optical components, micro-positioners, along with some specially made. Though the system is flexible, this kind of construction can lead to less stability and lower resolution than some systems. At this point, our system requires about 1.5 square meters of table space. This includes force regulation, as well as optical source and detection, and spectroscopic analysis equipment. Within that space, we can easily switch between reflection and transmission imaging.

This basic setup was used throughout the following research. Modifications were constantly being tried to increase resolution, stability, and sensitivity. Specific alterations to the system for sample specific concerns will be described as necessary.

As described, the system can be used for many different applications. Initially, the system was configured and some test images were taken using line width measurement samples, which consisted of chrome on quartz. These insured the system was in working order.
This led to the investigation of porous silicon. The interest in porous silicon stemmed from its optical characteristics (photoluminescent at room temperature), and the possibility of incorporating it into current silicon processing technology. Much of the research done previously concentrated on a macro analysis of the material, and there are several opposing theories behind the luminescent properties. We felt that a near field approach might lead to a better understanding of these processes, with the idea of perhaps incorporating porous silicon into a near field system as a light source.
Chapter 2

Correlative Imaging of Porous Silicon Topography and Photoluminescence

2.1 Introduction

Silicon is the most used semiconductor today. Current research on silicon includes various methods of incorporating light sources into silicon. The benefit of making a light source from silicon, of one form or another, is that the source could then be easily adapted and built right into existing technology. Successful creation of in situ light sources in silicon could open up a large area of continuing research into displays and opto-electronic devices.

When crystalline silicon is etched with acid or other means, it can become "porous." If the remaining silicon has nano-meter sized structures, after etching, it will photoluminesce. There are several theories behind this behavior. One goal of this research is to provide a means of examining porous silicon from a new perspective in hopes to clarify the chemical and physical means by which this photoluminescence takes place.
To understand the interest in porous silicon (P-Si)\(^1\), we must first look at silicon and its energy configuration before and after etching, and how the structure influences the photo-chemical properties.

### 2.2 Photon Emission and Bandgap Theory

In a crystalline structure, electrons can occupy only certain energy levels, which form various energy bands. The two highest energy bands are labelled the conduction band and the valence band. For metals, these levels overlap. For semiconductors and insulators, these levels do not overlap. The energy region between the two levels is called the "forbidden" region, because no allowed energy states exist here. The band gap energy for a given crystal is defined as the difference in energy between these two levels (See figure 2-1). Materials can be categorized by their band gap. Those whose band gap is between 0.5eV and 3eV are considered semiconductors. To promote electrons into the conduction band requires applied energy, in the form of either electronic absorption, or optical photon absorption. If

---

\(^1\)Porous silicon can be made of both n-doped and p-doped silicon. For notation purposes, any references to P-Si will imply porous silicon in general. N-type specifies negatively doped porous silicon, and p-type specifies positively doped.
this absorbed energy is sufficient, larger than the band gap, the electron will be promoted to the higher energy level. In terms of photon absorption, the energy of a photon can be calculated by:

\[ E = \frac{hc}{\lambda} \]  

(2.1)

where \( h \) is Planck's constant, \( c \) is the speed of light, and \( \lambda \) is the wavelength of the absorbed photon. When the electron leaves the valence band, it leaves an excess hole behind. Because this is not an equilibrium state, the electron will look for a method of returning to the valence band. The electron can return to the valence band by way of a band gap transition, either direct or indirect. In a direct transition, the electron returns from the conduction band straight down to the valence band. Energy is released that is equal to the band gap energy. Silicon is not a direct semiconductor and is more likely to decay by way of one or more intermediate energy states, indirectly. The presence of midgap levels typically results from impurities, or lattice defects, whose energy levels are located in the "forbidden" region. The excited electron will drop to this intermediate state, which is located between the valence and conduction bands, decaying by way of phonon emission, which consists of vibrational energy released into the crystalline lattice. The electron may continue to decay over several forbidden region energy levels, releasing phonons at each decay down to the valence band. In this case, no light is emitted. Because this phonon transition is much more probable than a photon decay, indirect semiconductors, like silicon, are typically not chosen for light emitters. However, it has been found that with the preferential chemical etching, a network of pores can be created in silicon which, when excited with sufficient optical or electrical energy, will luminesce[28]. This luminescence has been attributed to many different phenomena. In the following section, the most prominent theories of porous silicon's photoluminescent qualities are discussed.
2.3 Theories on Porous Silicon and Its Luminescence Properties

It has just be explained that silicon is an indirect semiconductor, and thus more likely, if excited, to decay by way of phonons. This would indicate that it is not a good source for light emission. But porous silicon has been shown to act quite unlike bulk silicon by many researchers[29]-[31],[33]. The structure of porous silicon is not well understood, though it is very dependent on the conditions during the etching process. This includes wafer dopant type and dopant concentration, acid concentration, etching time, and current density. Pore sizes range from the nanometer range to the micron range, and porosity (the percentage of the sample that is air) can range from 10% to 90%[30]. The processes of photoluminescence are also not well understood. Three of the primary proposed mechanisms are described here.

In 1990, Canham reported findings that if silicon was properly etched with hydrofluoric acid (HF), it became porous[29]. This porous silicon would photoluminesce when excited (illuminated) by ultraviolet light. He found broad band room temperature luminescence whose peak varied between 600nm and 800nm, depending on the preparation. This was a large blue shift from bulk crystalline silicon’s 1.1μm bang gap. With more etch time, he found porosity increasing, and column size decreasing. This decrease also lead to blue shifts in the peak wavelength of luminescence. The primary support for the quantum wire theory from Canham’s paper was that he showed that longer etch time lead to blue shifts. This blue shift would indicate smaller columns, and therefore higher band gap.

Two years later, Ohno first proposed that the etching leads to the creation of nanometer sized columns. These "quantum wires" were so called because they were small enough that quantum confinement properties stretch energy bands apart. The band gap energy in these wires was increased sufficiently so as to warrant a photon emission in the visible range.[34]
Further evidence for the quantum wire theory came from D. Naylor et al who looked at decay times of the luminescence by exciting the porous silicon with ultra-short 248nm pulses and clocking the luminescence. Their results showed agreement with current quantum theories[33].

Also V. Lehmann et al showed that the quantum wire theory not only accounted for many of the optical characteristics of porous silicon, but also explained the formation. This, he reported, was one more piece of evidence that the quantum wire theory was valid[35].

One of the opposing theories was proposed by Prokes et al who argued that hydrides present at the surface of the etched silicon (SiHx) made the structure similar to hydrogenated amorphous silicon (a-Si:H) which exhibits room temperature photoluminescence. They were able to form porous layers which, unlike Canham's and others' samples, did not exhibit blue shifting with increased etch time. He began with low porosity samples (~22%) and continued to etch and found no blue shifting. The luminescence was examined in a high vacuum environment. In this way, they were able to show that photoluminescence decreases with hydrogen desorption.[37]

Previous studies by Wolford et al indicated similar results. They found that higher concentrations of hydrogen in a-Si:H lead to blue shifting in the low temperature spectra (5K), and that the highest hydrogen content sample examined in the study, showed room temperature photoluminescence. He indicated that 30% hydrogen lead to a band gap of 1.7eV. This was proposed to be due to polysilane chains (SiH2)n. The presence of polysilane chains in a material increased band gap energies.[36]

Still others believe the luminescence comes from the presence of siloxene related compounds in the porous silicon. Fuchs et al saw the similarities between siloxene and porous silicon, and found that the luminescent spectral peak (for the particular P-Si sample they examined), and the peak from two different derivatives of siloxene (Si6O3H6 [Kautsky recipe] and Si6O3+nH6-m [Wöhler recipe] ) matched[31]. The Kautsky prepared sample showed a much larger spread in the spectrum, but that was disregarded due to "rather uncontrolled substitution of hydrogen by OH groups during the anneal." The Wöhler compound

33
spectrum matched well to the photoluminescent spectrum. They concluded that the origin of the photoluminescence must be the same in both the siloxene and porous silicon samples. They proposed a silicon ring structure which, with different ligands, could account for shifting photoluminescence peaks and changing radiative recombinations among different samples.[31]

Because understanding this mechanism could be an important part in the development of porous silicon as a useful light source in optoelectronics, the following research hopes to introduce a new method of investigating and analyzing porous silicon in an attempt to clarify the actual processes in porous silicon luminescence.

2.4 Anodization of Silicon in Hydrofluoric Acid

Porous silicon can be made by many different means. One common method of production is etching silicon wafers in a solution of hydrofluoric acid. Etching is done by anodization which produces a network of randomly spaced pores. The etching occurs preferentially to the pores due to surface energy effects and surface defects. As structures become small, the band gap energy goes up. This makes it more difficult for electrons to leave the silicon, and combine with the ionized acid. This means that area will be less likely to etch than the surrounding region. This leads to the creation of the pores. During the etch, to speed the process, a current can be passed through the wafer\(^2\). Further etching intensity can be increased by adding illumination upon the sample during the etch time. The pores, space where there is no silicon remaining, typically range in size from 50nm to 2nm. This leaves behind silicon columns. If these columns are less than 5nm, the bandgap is sufficient to allow visible photon decays[29], i.e. photoluminescence.

The typical setup for etching is diagramed in figure 2-2. The silicon wafer is sealed against a bath of HF, which can be buffered with ethanol, methanol or water. To speed the

\(^2\)This defines the difference between anodization and chemical etching. Both are wet processes involving an acid etch. But anodization relies on a current passing through the sample and acid solution during etching to increase etch rate and vigor.
etching process, a voltage potential is setup between the cathode, an ohmic contact on the back of the wafer, and the anode, a platinum plate in the HF solution. This generates a current through the solution and the silicon wafer. The current flux can be measured and current density calculated. Typical current densities during etch time should not exceed 50 mA/cm². For current densities above this value, etching is no longer preferential to the pores. The etching is so vigorous that electropolishing occurs, stripping the wafer clean of pores and nano-crystals. To aid in even etching spatially across the wafer, an ohmic contact can be made on the backside of the wafer before etching. This provides more even current densities, and more uniform etching. Both n-type and p-type wafers required at least 5 minutes of etch to begin to photoluminesce. Typically the photoluminescent intensity would increase with etch time. Typical etch durations were between 5 and 30 minutes.

Once the samples were made, we had to configure the particular system with which we would be able to collect both images and spectroscopy.[38]

---

3The contact on the back of the wafer is ohmic to provide equal potential over the entire surface of the wafer. For current density, we assume this ohmic contact and divide current by area. This is only a rough estimate, since variations will occur if the back contact is not perfectly ohmic.
2.5 Imaging System Configuration

For the porous silicon images that follow, we used the basic setup described earlier. The particulars of the setup for examining porous silicon are pointed out here (see figure 2-3). In general, the luminescence from porous silicon is a broad band of reddish light (typically centered between 650nm and 700nm). For our imaging, we needed to establish a method of isolating the red luminescence from the collected light, which included, in addition to the desired luminescence, the illuminating source, argon ion (488nm and 514nm), and the force detection light (633nm). The source laser was modulated by a Bragg cell before being launched into the fiber and piped to the tip. The frequency of modulation varied throughout the images, between 1 kHz and 10 kHz. This light was collected after reflection
off the sample by a microscope and focused onto a photomultiplier tube (PMT). The PMT signal was sent through additional amplification stages, and to a lockin amplifier whose reference frequency was the same as the Bragg modulation frequency. This allowed us to isolate the modulated light from any DC (primarily the force detection light). Because the blue reflected light was also being modulated, we introduced an orange filter (passes 550nm and up) just before the PMT. This isolated the photoluminescence from the blue signal. To further insure that the system was only collecting photoluminescence, we placed a 633nm notch filter before the PMT. This kept out any force detection light that may have been modulated from the tip vibration, thermal effects of the tip, etc.

Tests were periodically run during the imaging process to confirm the fact that the luminescence signal collected was indeed from the silicon’s luminescence and not from luminescence or fluorescence of any of the optical components in the system. One basic difficulty with this experiment is that many things fluoresce naturally, e.g. paper, plastics, glue. That natural fluorescence tends to span the same rough wavelength range where we expect to find the silicon’s photoluminescence. If any single element in the system, filters, optics, etc. was fluorescing, that emission would be modulated as the blue and could be collected and interpreted as photoluminescence. An image with strong fluorescence from a component in the system would mimic the direct reflected image. In order to insure the system’s discrimination, we performed some test scans. Figure 2-4 shows the force, luminescence and blue images of an edge on sample B (specific etch rates are included in the following section). The left portion of each image is the polished silicon wafer. The lower right region is the porous area. Notice, the luminescence is only present on the porous region (as expected). This served as a good test for the system. Also note, in the blue image, there is plenty of light with which to image. If the blue and luminescent images had matched (in this particular test) we would not be able to conclude anything. However, the difference between the two images indicates that there was no leakage of blue, nor significant fluorescence or luminescence from other elements of the system. The absence of luminescence within the polished region indicates that the system is properly separating the blue and red compo-
Figure 2-4: Force (a, top and bottom), luminescence (b), and direct reflected (c) images of a demarcation between polished and porous silicon on an n-type wafer.
ments of the collected light. Also in the blue image, there is a dark line between the regions. This may be accounted for by the shadowing effects described in the following section.

2.5.1 Image Interpretation

After all the benefits have been considered, one must remember that the images are not free of complications. Because of certain optical characteristics, such as interference, and some system setup requirements, there are a few "errors" that need to be accounted for in certain imaging modes. Because the primary imaging mode for this thesis is reflection, this discussion will concentrate on the problems associated with that imaging mode.

One of these problems is a basic shadowing of the illumination light, by the sample. Imagine the following scenario in a reflection mode system. Figure 2-5 shows a "well-defined" sharp edged sample of opaque lines on a flat substrate. During a raster scan, the tip will approach the wall of the sample. As the tip nears the wall, the reflected light will be trapped by the wall. This results in a false decrease in the reflected signal. On the opposite side of the wall, there is a false increase in the signal due to multiple reflections. These false detections depend highly on the actual surface topography and surface dimensions. Also the tip-sample orientation and sample-collection microscope orientation can play an important role in these discrepancies. Because the tip and collection microscope are not co-linear, there must be some finite angle between them, which is dependent on the physical dimensions of the lens, the actual configuration of the system, the focal length and working distance of the lens, etc. This angle, along with the surface topography, will define the amount of shadowing and enhancement seen. A smaller angle will lead to less prominent effects. But creating a smaller angle, means either a longer working distance lens or a smaller lens (physical dimensions). The best configuration can only be found from empirical data, which inherently include equipment limitations, space considerations, etc. Furthermore, if these effects are known, they can be effectively accounted for when analyzing an image.

As an actual example, we performed some scans on a sample of polysilicon lines on silicon[27]. These samples are commonly used in micro-lithography for resolution testing
Figure 2-5: Schematic of an idealized tip/sample orientation to illustrate the problem of false reproductions in optical signal collection.
and line width measurements. Because these samples are well defined, it was easy to identify shadowing and false enhancement of the optical signal. For figure 2-6, the collection microscope was pointing down, from the top right. The left images are the force, and the right images are the reflected optical. Notice on the top pair of images that the lower walls are shadowed while the upper walls show enhanced signal. The bottom pair of images shows the same sample from the bottom of the structure, and one can see the shadowing and enhancement reversing their positions. There is reflection enhancement in the grooves and shadowing along the bottom wall. Notice these images are denoted "white light". This brings us to the next problem with image interpretation, which is interference.

With the previous images in figure 2-6, white light was used specifically to reduce interference effects. A multimode fiber was used to produce a tip, and a xenon lamp served as a white light source, reducing temporal coherence. Without coherence, interference effects were averaged out. However, in the next images, figure 2-7, we see that with monochromatic light, in this case a HeNe laser, interference plays a role. The presumption is that not all of the light interacts with the surface directly below the tip. This light can be reflected and collected (see figure 2-5). But this light has now followed a different path from the directly reflected light. This path length difference will lead to interference. Notice the dark bands in the enhanced signal in figure 2-7 which are interference fringes. This sample was used to demonstrate the shadowing effects because of its regular structure. The sharp edges and straight lines make these errors easily identifiable.

The interpretation concerns mentioned are, by no means, the only concerns when interpreting images. However, these have proven the most prominent "false" image characteristics.
Figure 2-6: Force (left) and reflected optical (right) scans of polysilicon lines on silicon. White light illumination.
Figure 2-7  Force (left) and reflected optical (right) scans of polysilicon lines on silicon  Coherent illumination.
2.6 Images, Spectra and Discussion

2.6.1 Samples

The samples used in this study were made by Fred Seiferth in the Microelectronic Engineering department at Rochester Institute of Technology. The samples chosen had the following etch parameters.

- **Sample A**: bulk n-type wafer, etched in HF:H\textsubscript{2}O=1:1 for 40min with illumination, average current density = 6.3 mA/cm\textsuperscript{2}.

- **Sample B**: bulk n-type wafer, etched in HF:H\textsubscript{2}O=1:1 for 5min with illumination, average current density = 10 mA/cm\textsuperscript{2}.

- **Sample C**: bulk n-type wafer, etched in HF:H\textsubscript{2}O=1:1 for 40min with illumination, average current density = 10 mA/cm\textsuperscript{2}.

- **Sample D**: bulk p-type wafer, etched in HF:C\textsubscript{2}H\textsubscript{5}OH:H\textsubscript{2}O=1:1 for 30min without illumination, average current density = 10 mA/cm\textsuperscript{2} (C\textsubscript{2}H\textsubscript{5}OH=Ethanol).

- **Sample E**: bulk p-type wafer, etched in HF:C\textsubscript{2}H\textsubscript{5}OH:H\textsubscript{2}O=1:1 for 15min with illumination, average current density = 40 mA/cm\textsuperscript{2}, then for 15min average current density = 56 mA/cm\textsuperscript{2}.

It should be mentioned that the investigation was focused on photoluminescence, and not electroluminescence. For this reason, the word luminescence will from now on imply photoluminescence.

2.6.2 Sample Specific Interpretation Considerations

Image interpretation, as explained above, can include several considerations, including those which are sample specific. In the case of porous silicon, there can be some interpretation
problems when the tip probes into a pore. Consider a situation in which a tip is probing into a pore, so the light coming out of the tip will illuminate the silicon just below it (see figure 2-8). Not all of that light will be absorbed by the silicon. This extra scattered light could excite luminescence along the neighboring walls of the pore. All the light within the collection spot size of the microscope will contribute to a signal. For this reason, when a tip probes into a pore, we might expect to see a falsely enhanced signal. At the same time, if this pore is deep enough, it may act as a trapping site for any light, blue or otherwise. This would be manifested as a lower signal than the actual. Both these considerations must be taken into account when interpreting the following images.
2.6.3 N-Type Images

N-type (negatively doped) samples tended to be very rough. The forty minute etch, sample A, had macro pores with dimensions between 10 to 20 microns wide and deep. This kind of topography is difficult for our system to handle. The engagement signal only allows a certain range of movement, as does the actual z motion bimorph. For this reason, most of the force images were taken from sample B. It had sufficient photoluminescent qualities while having more manageable topography.

We see in the images both positive and negative contrast when comparing the luminescent image with either direct reflected, or force microscope. Figure 2-9 shows a large range scan (12 square microns) from sample B. The overall contrast is positive between the force and luminescence. There is considerable speckle present in the direct reflected blue image. This can be attributed to the interference effects described in above. The smaller scan (6 square microns, see figure 2-10) This shows both regions of positive and negative contrast. Again, here the interpretation can be difficult. But it appears that there is more luminescence from certain pores than others. Recall the sample specific interpretation problems. In interpreting this particular image, shadowing seems evident in the luminescence in the bottom central region of the scan (identified by the two arrows). Notice the large vertically oriented pore in the force image. The shadowing effect described above would account for the bright and dark regions shown in the corresponding luminescent image.

Another small scan image (figure 2-11) shows the complexity of the topography. There is good detail in the force image. Notice that the reflected blue image is basically uniform, disregarding speckle. However, the luminescent image shows significant contrast. Also the higher luminescing regions correspond to both pores and peaks in the topography. That is, the contrast shown is both negative and positive. This is a substantial characteristic for several reasons. First, it indicates that the system is not showing much coupling between force and optical. If that were the case, most of the valleys would show one value for luminescence, while the peaks would show different luminescence. Furthermore, it indicates that there is indeed topographical dependence on luminescence. That is, not only does the
Figure 2.9: Force (a, top and bottom), luminescence (b) and direct reflected (c) images of n-type P-Si (Sample B)
Figure 2-10: Force (a) and luminescence (b) images of n-type P-Si (Sample B).
Figure 2-11 Force (a), luminescence (b), and direct reflected (c) images from n-type P-Si (Sample B).
physical topographic shape affect the resulting image, which we have established, but different regions of the sample contain different characteristics and exhibit different luminescent efficiencies. This is very important in terms of the previously mentioned research ideas. If this type of material is to be integrated into chips, for quality control and reliability, the process should make each tiny region of porous silicon act the same as every other. If the etching process has the inherent randomness shown by these images, the chance of maintaining quality control will be limited.

2.6.4 P-Type Images

The p-type (positively doped) topography varied from sample to sample, and was distinctly different from the general n-type topography. In general, the p-type was less rough than n-type which tended to etch in large valley like pores. The p-type samples investigated showed ridge structures with relatively flat regions in between. Because our system has a limited z-movement, the p-type samples were easier to image. Their relatively shallow topography was within the limits of our imaging capabilities.

One of the characteristics noticed during the p-type etching, was that, under the proper conditions, a thin film developed. This film appears shiny and "brushed" to the eye. The film is quite reflective and hides a more detailed topography beneath. The formation of this thin film layer seems dependent on both acid concentration and current density[38]. The presumption is that the top of the wafer etches into a "mesoporous" state. This mesoporous region allows acid to seep underneath it where the nanopore etching takes place. The etching that occurs beneath this surface region creates H\(_2\) as a by-product, which collects forming bubbles beneath the "mesoporous" region. If these bubbles grow large enough, the surface film will burst and separate from the sample. This has lead to some current investigations on the thin film properties, which are being pursued by our group.

Images of sample D (Figure 2-12) show very even luminescence over the 12 micron scan. This sample appeared very smooth, compared to the others. It had a shiny, reflective surface and showed very little topographical detail. The images show here are from sample D whose
Figure 2-12: Force (a) and luminescence (b) of thin film region of μ-type P-Si (Sample E).
film is apparently undisturbed. Recall that sample D had no illumination during etching. The illumination increases the etching strength by influencing the electronic properties of the silicon during the etch. Without illumination, the etch proceeds more slowly and with less vigor. This would indicate that less hydrogen was formed beneath this layer and was insufficient to take the thin film off. The fact that we get even luminescence may indicate one of two things. Either the layer beneath is luminescing and the luminescent light is being sufficiently scattered by the thin film so as to average it, or, and perhaps more likely, the film itself is in fact luminescing. From previous images, we saw a correlation between the luminescence and the topography. Rough topography did not produce even luminescence. These images might indicate that the film itself is evenly luminescing, and therefore may not be entirely mesoporous. Again, these results do not conclude either of the above hypotheses, but warrant further investigation, which is currently underway.

Sample E was a very interesting sample in our investigation. Figure 2-13 is a micrograph of the sample. Recall this is a p-type wafer in which the current density, and etching was sufficient to remove the majority of this thin film covering. The topography shown, in the micrograph, and the topography investigated for the following images show the "exposed" regions of the sample. Here one problem in the imaging was the dynamic range. Looking at the micrograph, notice the dark lines which make up the "designs", which are the areas where the photoluminescence is concentrated. Within these lines, small "ridge-like" structures existed which are not visible in the photograph. These ridges are approximately half the width of the dark lines. They showed photoluminescence to background (neighboring flat regions) of 50:1 or more. For this reason, small variations of luminescence within one ridge were difficult to see over large range scans. The gray, and relatively flat, regions show very low luminescence, perhaps 1:4 ratio of photoluminescence to background. We began with large scans, 24 microns, to achieve a general context and size, as well as to locate a good region for scanning. An example of these large scans is shown in image 2-14. Each individual scan is 12 microns square. The two neighboring images are displayed together to

---

4Indeed, far field micrographs showed a strong correlation between film and good luminescent qualities.
Figure 2-13: Micrograph of p-type P-Si surface (Sample E).
Figure 2-14: Force (a), luminescence (b) and direct reflected (c) images of p-type P-Si. Each pair shows a 24x12 micron area (Sample E).
individual scan is 12 microns square. The two neighboring images are displayed together to show a 24x12 micron area. Small topographical detail can be seen rightmost of the two the force images (top scans in figure 2-14). Inherent in our current data collection and imaging is a scaling from the collected data to an image format. This scaling is a linear interpolation whose slope is defined by the minimum and maximum values in the recorded data. For this reason, an overall flat image will be enhanced to show the small detail, whereas an image with large detail, such as the left force image, will lose some of the smaller detail⁵. These scans were taken in pairs, in an effort to alleviate this dynamic range problem. Notice that the ridge, the large area in the bottom left corner shows very high luminescence. Also, notice the reflected blue image has much less contrast. This is further indication that we are actually getting luminescence from the sample and not remnants of luminescence from optics, filters, fibers or any other piece in the system. The narrow dark region just above and to the left of the ridge in the blue image can be accounted for by the shadowing effect explained above.

Once the context was established, smaller scans were taken to investigate the finer detail of the sample. The following images, figures 2-15, 2-16, and 2-17, all show different ridges with the corresponding blue and luminescent images. The blue image is direct reflected light. In each case, the ridge is well defined topographically and shows strong photoluminescence. We see positive contrast between the force and luminescent images. This seems to be the general case when looking at the larger scans. Because of dynamic range, and resolution (defined by the pixel size, not the tip), again, we cannot see the finer details on the larger scans. For this reason, smaller images of the ridge were taken. Images in figure 2-18 are taken on the top of one of these ridges. First, we can see much finer detail in the topography. Furthermore, we see negative contrast between the topography and the luminescence. Notice the slight spatial shift between scans⁶. The blue scan is shifted slightly

⁵Our collection methods allow 16 bit (2¹⁶ gray values) inputs. But for display purposes, we have 8 bits (2⁵⁵ grays). For this reason, we scale the 16 bits linearly to 8 bits and much of the small detail can be quickly swamped by large surface detail.

⁶Our current system can only capture two images at once. This is not an inherent limit of the system, simply a limit from our computing power. Because we can only capture two parameters, we typically take
Figure 2-15: Force (a), luminescence (b) and direct reflected (c) images of p-type P-Si. White boxes on left side scans indicate regions of right side scans (Sample E).
Figure 2-16: Force (a, top and bottom), luminescence (b) and direct reflected (c) images of p-type P-Si (Sample E).
Figure 2-17: Force (a, top and bottom), luminescence (b) and direct reflected (c) images of p-type P-Si (Sample E).
in the vertical direction. This is due primarily to temperature variation effects of the piezo translators which control the x and y motion of the scan. Taking that into account, the dark regions in both optical scans match spatially. The luminescence, however, is showing a relatively bright center which is not evident in the blue image. This negative contrast can be seen on the smaller scale as these scans show.

The smaller images have occasionally shown some questionable data. The images just discussed look reasonably well defined but not as well-resolved as the following image. Figure 2-19 shows very fine resolution in the luminescent image. The resolution apparent in the luminescent scan seems to be attributable to a coupling between the force and optical images as discussed above.

Figure 2-20 is interesting in that it is large enough to see the positive contrast, and small enough to notice some negative contrast. This image also shows a ridge region. Notice the dark "nib" in the bottom center of the image. This is a macropore in the sample which shows a "halo" around the edge in the luminescent image. This is probably due to the enhancement as explained above. If the non-absorbed blue scattered and excited nearby areas to luminesce, it would follow from the above argument that near the walls of the pore, a brighter signal should be collected, since in the center, there are no walls near enough to enhance the luminescence in that way. This is evident in image 2-20.

There have been SEM images showing columnar structures in porous silicon,[30] not fine enough to see the individual "nano-columns" but enough to see the general column-like orientation of the silicon. This can also be seen in our image along the side of a ridge. Figure 2-21 shows the column like structures. Examining the corresponding optical image, this is another set of images that display both positive and negative contrast. The negative contrast may not be obvious. In the force images, there are three large columns pointing to the upper right corner. The upper two seem to be positively seen in the luminescent image, but the third one, which seems larger than the upper two, shows somewhat less

force as one of each pair of images taken. By taking force images, for each optical scan, we can verify and account for any slight positional shift that may have occurred.
Figure 2-18: Force (a. top and bottom), luminescence (b) and direct reflected (c) images of p-type P-Si (Sample E).
Figure 2-19: Force (a), and luminescence (b) images of p-type P-Si, showing ultra-high resolution (Sample E).
Figure 2-20: Force (a), and luminescence (b) images of p-type P-Si, showing "halo" like reproduction of a macropore (Sample E).
luminescence.

As mentioned before, the p-type topography tends to be flatter than the n-type. As an example, figure 2-22 shows a topography and photoluminescent image which indicates the lack of topography between the ridges. The luminescence is present, but is much lower in the flat regions. The image shows a fall off in luminescence over the 12 micron square area, whereas the force show no significant change. Also, the small "nib" in the force also corresponds directly with a bright point of luminescence.

From all these images, it appears that almost any combination of topography and luminescence is possible. We have shown pores which glow, and those which do not. This is indicating that the sample is not uniform topographically, which could have been established with a farfield microscope, but also non-uniform structurally. If the luminescence is a direct result of the micro-structure, as is expected, then that structure is varying over the etched area of the sample, and quickly varying spatially, as our images portray.

2.6.5 Spectra

From the above scans, we noticed that there is strong luminescence from both the "brushed" or thin film region and the "exposed" regions of the sample. We set out to establish any possible distinction between these regions, by obtaining and comparing their respective luminescent spectral distributions. For, although a spectrum, in itself, may not be a viable method to compare samples made by different etching processes, for samples that have undergone the same processing, the spectrum can be a reasonable parameter to analyze similarities and differences between two regions.

The spectra were obtained using a setup similar to the configuration used for the imaging except that the blue detection leg was replaced by a monochromator and a cooled CCD (see figure 2-23). The CCD was receiving light in the 10^-15 Watts/pixel regime. We were collecting about a nanowatt of reflected blue light. The luminescent efficiency was less than 1/2000, leaving sub-picowatt power, which was further spread by the monochromator over a range of about 200nm. For this reason, the spectra were obtained over integrations times
Figure 2-21: Force (a), luminescence (b), and reflected optical (c) images of p-type P-Si, showing macro columnar structure (Sample E).
Figure 2-22: Force (a), and luminescence (b) images of p-type P-Si, showing flat region between ridges (Sample E).
Figure 2-23: Schematic of SFM/NSOM setup for spectral analysis of porous silicon.
of 100 seconds, and several spectra, each 50nm in width, were individually taken and spliced together to achieve the spectra included here in figures 2-24 and 2-25.

These spectra were taken on both regions, one on top of a ridge, and the other on the thin film region. Notice the peaks, which are at 672nm and 675nm respectively, match quite well. The notch centered at 633nm is due to a holographic notch filter. This filter prevented force detection light (HeNe) from being collected by the system and saturating the CCD. Excluding the notch, the plots are reasonably smooth and show similarities in the overall shapes. The full-width at half-maximum (FWHM) may initially appear larger for the ridge region. But notice that the dark current level is not zero. That is, the background has not been sufficiently subtracted out, and the zero levels for the two plots do not match. This makes the measurement for FWHM difficult to make accurately. Also, due to reflectivity changes across the different regions of the sample, the noise values may be more influential on one spectrum than the other.

In general, it appears that the two regions have the same spectra and perhaps have the same concentrations of nano-pores. These spectra also warrant further research in the thin film region.

### 2.7 Porous Silicon Particles

Once the images from the wafer samples had been collected and analyzed, we wanted to isolate smaller particles for imaging in order to reduce the image interpretation problems. To do this, we started with sample C which had deep topographies. The porous region was carefully scraped creating dust particles which were placed into a drop of poly-methyl-meth-acrylate (PMMA) and spun onto a microscope cover slip. The PMMA acted as a non luminescing adhesive to keep the particles in place during the analysis. With this sample, we were able isolate and image single particles of a few microns diameter. This analysis removed the possibility of secondary luminescence and shadowing problems which made the previous images difficult to interpret.

Because the particles were small, and the penetration depth of silicon for red light
Figure 2-24: Spectrum from ridge region of p-type P-Si. Peak at 675nm.
Figure 2-25: Spectrum from thin film region of p-type p-Si. Peak at 672nm.
Figure 2-26: Transmission configuration for P-Si dust investigation. For spectral analysis D1 was replaced with monochromator and cooled CCD array.

(luminescence) is at a few microns\textsuperscript{7}[39], these particular experiments were done using a transmission mode configuration (see figure 2-26). In the transmission mode, the shadowing effect and secondary luminescence considerations are much less of a concern. In fact, the primary imaging consideration is optical penetration. As the sample becomes thicker, more light will be absorbed by the sample, and less will be collected. For direct transmission imaging, we would expect direct correlation with force imagery.

\textsuperscript{7}The penetration depth for silicon for luminescence (assuming $\lambda=670\text{nm}$) is $5\mu$m, and for the blue illumination source ($\lambda=488\text{nm}$) is $1\mu$m. This helps us isolate the blue since typical particle size in the images is about 2 microns. This allows red through but retards the blue to less than the 1/e point.
2.7.1 Images

The first set of images show a particle approximately 2 microns across. The luminescence image shows a strong signal on the bottom half of the particle and less luminescence on the top half (see figure 2-27). Initially, this image looks as though it is a direct transmission image. As mentioned, the stronger luminescence is evident from the most thin area of the sample. But the next image shows that the image here is truly the luminescence, without significant influence from the topography.

The next image, figure 2-28, shows two particles, again each one about 2 microns across. The two particles look very similar from the force image, as far as topographical characteristics are concerned. However notice the luminescence image shows two very different particles. The upper left particle, which appears slightly thicker in the force image, shows a highly localized region of strong luminescence and low luminescence around the outer edge. Whereas the lower right particle shows the opposite. The center of that particle shows low luminescence while the outer edge shows slightly stronger luminescence.

Our next image, figure 2-29 show further evidence that the luminescence varies dramatically from region to region. The force shows several particles near one another. The luminescent image shows the drastic variation in efficiency from one particle to the next. The image brightness has been enhanced to show that there is luminescence in the lower and right most particles. This accounts for the apparent saturation of the central left particle. Notice also, that the thickest particle shows the most luminescence. In terms of the quantum wire theory, this would appear to be reasonable. A larger particle will contain more columns and therefore luminesce more than a thin particle. Furthermore, it is obvious that the effect of the penetration depth is insignificant. That is, the absorption of any luminescent power by the particle is negligible compared to the amount of generated luminescence.

Finally, to insure that the proximity of the particles from the previous image was not influencing the luminescence, we took a large range scan (10 micron square) which included two particles (see figure 2-30). Again, there is the dramatic variation of luminescence from
Figure 2-27: Force (a) and luminescence (b) images of P-Si dust particle (made from Sample C).
Figure 2-28: Force (a) and luminescence (b) images of P-Si dust particle (made from Sample C).
Figure 2-29: Force (a) and luminescence (b) images of P-Si dust particle (made from Sample C).
one to the other, while topographically, there are only slight variations evident.

Finally, we saw high resolution imaging, and highly localized luminescence in figure 2-31. The size of the luminescent spot appears to on the order of a few hundred nanometers. This indicates both good optical resolution and a dependence on the topography down to the tenths of microns regime. This is further support for the quantum wire theory.

2.7.2 Spectra

If the quantum wire theory holds, smaller particles should show more narrow spectra. As the particle gets smaller, it will contain less pores and the "statistical spread" that we see in the spectra here should be reduced in width. Taken to an extreme, one nano-crystal should emit nearly monochromatic light. Furthermore, as there is evidence of regional variations of luminescent strength, the question about spectral variation arises. To complete our investigation of the dust, we setup the system for spectroscopy. To take spectra, we simply found strongly luminescent particles, and placed the tip, stationary, on that region. Because of signal considerations, the spectra were taken over 100 seconds. In this time period, there was probably some x-y shifting due to drift in the piezo translators. However, it did not prove to be significant enough to affect the results.

Three spectra were obtained from three different dust particles. Two of the three spectra match closely, while the third is blue shifted by nearly 40nm (figure 2-32)\(^8\).which is a significant shift. We also saw some change in the overall shape of the spectrum. We proceeded to take macro-spectra of the same sample, both on the etched region and the "scraped" region from which the dust was generated. We noticed that the macro spectra seemed to have a double peak (figure 2-33). It appears that the first two dust spectral peaks are more closely related to the left most peak of the macro spectra, while the shifted dust spectral peak is closer to the right macro spectral peak. This indeed shows that, as well as being different in intensity distribution, the spectral distribution can change from

---

\(^8\)Note that this figure is simply for overall shape comparison. The relative intensities from one spectrum to the next are arbitrary.
Figure 2-30: Force (a) and luminescence (b) images of P-Si dust particle. Large range scan (made from Sample C).
Figure 2-31: Force (a) and luminescence (b) images of P-Si dust particle showing extreme localization (made from Sample C).
Figure 2-32: Spectra from 3 different dust particles. For shape comparison only.
Figure 2-33: Macro spectra from (a) etched, and (b) scraped regions.
region to the next within one sample. This may also lead to notion that the macro spectra is actually a superposition of the spectra taken at many different regions, in which the peak values are a result of the particular etching procedure and etching parameters.

These spectra further support the quantum wire theory.

2.8 Conclusions and Further Research

Based on the experimental results, data, and model for collection optics, we have shown that the topography and luminescence are inter-related. From one region to the next, on a micro scale, the luminescence can vary significantly, as seen particularly with the dust images. If this material is going to be incorporated into chips or other opto-electronic devices, there must be some more research on production of uniform performance for each sample. If there exists any variation over small spatial dimensions, there will be some significant variation from sample to sample. The process of etching must be such that a flat, uniform luminescent region is obtained.

It appears, at this stage in our analysis, that the quantum confinement theory is the closest of the three presented to appropriately modeling the processes behind the luminescence of porous silicon. We expect further analysis into the thin-film regions of the p-type wafers, and small particles to provide further evidence in support of that theory.

This research has incited many questions into the p-type thin film area. Further research will include this thin film investigation. Furthermore, we have presented the beginning of our silicon dust analysis, but we would like to extend it to still smaller particles and continue the spectral analysis. If these small particles could be isolated and attached to the end of a tip, we would have a NSOM/SFM tip with an integrated light source. This could be achieved with current micro-machined silicon tips which could be etched at the tip. The etching would make porous silicon at the very end of the tip. This could be illuminated with blue light, creating a glowing red tip of luminescence, or perhaps set up with appropriate voltages to excite electroluminescence. These are all areas of research that will be pursued in the near future of this group.
Chapter 3

Photothermal Imaging with SFM and NSOM

3.1 Introduction

Photothermal imaging and microscopy techniques have been developing rapidly over the last two decades. They have proved important in the non-destructive imaging fields and in material composition analysis fields. Thermal imaging consists of analyzing samples by studying their thermal characteristics directly, or by utilizing optical methods to see secondary characteristics resulting from thermal activity in the sample.

Photothermal imaging is typically performed by heating the sample in question with an energy beam, often a laser, and monitoring the thermal waves, or results of the thermal waves, as they propagate through the sample. To create a thermal wave, there must be a periodic source of some kind. The energy source can be modulated, creating an oscillating source which will induce thermal waves. The periodicity of these waves can be easily distinguished from the DC heating. Thermal waves, act like evanescent waves, in that they decay quickly, within a wavelength. However, within the decay length, they behave like a conventional propagating wave.

There are many different approaches to thermal imaging which are each based around
the fact that the heat source is heating both the sample, and the surrounding air (through radiation). To probe the thermal variations, the system should be designed to analyze the air, and its changes due to heating, or the sample and its variations with heat. This leads to many detection schemes, six of which, originally designed for far field microscopy, are described here.

For each of the systems (see figures 3-1 and 3-2), the heat source is a modulated laser, whose modulation frequency will determine the frequency of the thermal waves. One of the earliest systems involved encapsulating the sample in air. When heated, the sample heats the air layer in its vicinity which leads to acoustic waves that will propagate within the enclosed air region. These acoustic waves can then be monitored by a microphone\cite{42}. Because the sample heating will induce structural changes in the sample, the surface changes can be monitored by a position sensitive detector. This includes an interferometric detection method, or simply a direct motion detection\cite{43}\cite{44}. Acoustic wave detection on the sample can also be used by placing the sample in contact with a transducer\cite{45}. Temperature variation can be monitored by infrared detection methods\cite{46}. And finally, the detection can be performed by using a probe beam which is detected after deflection by index of refraction changes in the heating air above the sample, also called the "mirage" effect.\cite{47}

Imaging of these modes can be performed on samples exhibiting any kind of thermal variations, heat capacity, thermal conductivity, shape deformation, etc. One of the problems inherent with photothermal microscopy is that the resolution is related to the frequency of the source modulation.\cite{42} Besides resolution problems due to high frequency thermal waves, existing systems using far field microscopy are limited by their optical resolution as well. So for high resolution of imaging, high frequency thermal waves are required, but also high resolution detection is necessary to see the results.

With these concepts in mind, we emphasize that the field of thermal imaging is not a new one, but the applications to high resolution imaging is fairly recent. To establish the state of technology, we present some of the more recent experiments in thermal microscopy.

Far field systems dominate today's thermal microscopy experiments. As well as the
Acoustic Detection

Surface Deformation Detection

Figure 3-1: Various detection schemes for thermal microscopy.
aforementioned experiments by Rosencwaig, Ameri et al which were performed within the last 10 years, there is still the interest in farfield systems today. Mansanares et al published a study in April 1994 comparing theoretical and experimentally procured data on the influence of planar thermal barriers on photothermal reflectivity.\[48\] A derivation of a heat diffusion model for isotropic layered systems was presented by Reichling et al in January 1994\[41\]. Note that this topic had already been published by Vaez-Iravani et al in 1988\[40\], but this shows that there is still interest in the field. Furthermore, U. Seidel et al used a far field system, in May 1994, to calculate photothermal contrast from deviations in thermal conductivity to find optimal modulation frequencies for analyzing samples\[52\].

These currently dominate the field due to ease of use and ease of setup, but the imaging quality, in terms of resolution and contrast, could be improved. In 1986, Rosencwaig and Opsal designed a method of performing high-resolution thermal imaging using an SEM as their heat source\[49\]. The electron beam was intensity modulated to generate a modulating thermal wave within the sample. But for imaging, they concentrated on the accompanying thermoacoustic generated waves. These waves have a much slower decay but much longer wavelengths. For this reason, they detected the acoustic waves which carried with them the
information from the thermal waves. This coupling between the two can be an important factor in thermal imaging. This provides the capacity for use in penetration imaging.

One of the early near field probe systems was setup in 1988. Williams and Wickramasinghe introduced the Scanning Thermal Profiler[22]. Starting with a small probe, not unlike a NSOM tip, they constructed a thermocouple sensor at the tip using the junction of dissimilar conductors. The conductors covered the tip, and were separated by an insulator everywhere except the very end of the probe. The thermocouple region then produced a voltage that was dependent on the temperature of the probe. This voltage was recorded as the thermal signal. The probes were constructed with dimensions of 100nm and claimed 0.001 degree sensitivity with 100Hz bandwidth. This construct was able to thermally image samples with very high resolution. But they also proposed the use of this probe as a surface profiler. As a heated tip is brought into near proximity of a solid sample, because solid heat transfer properties are much better than air, the tip will be cooled as heat is transferred from the tip to the sample. As the tip is scanned across the material, the heat of the tip can be monitored and will serve as a feedback signal for z-position control of the sample.

So even though there are methods of doing near field thermal microscopy, many have not yet been exploited. Using thermal imaging, this research hopes to present a high resolution method of examining current crowding in micro circuitry, such as computer chips. This kind of imaging could prove to be important in the future of circuit design. Because current flow leads to heating of the conducting material, with NSOM photothermal imaging, we can investigate the spatial current distribution over a wire which may be on a sub-micron scale. If there is significant heating in certain areas, this means heavy current crowding, which can lead to electron migration\(^1\), which in turn can result in weaker wires, and eventually a breakdown. If we are able to isolate these heavily populated regions on the circuit, before failure, changes can be made to the circuit design to reduce crowding, improving energy

---

\(^1\)Electron migration is the physical movement of aluminum atoms due to multiple collisions with the conducting electrons. Sufficient current will lead to movements significant enough to alter the current densities in the wire or track. This will lead to higher localized current, thus more rapid movement. Eventually this can weaken the track a point such that the track can no longer support the current flowing, which will result in failure.
loss, and lifetimes of the circuits.

3.2 System Configuration and Imaging Considerations

Photothermal effects are caused by variations in structural properties due to heat, and can be observed with visible light imaging. In this investigation, we were interested in analyzing the current distribution in micron sized aluminum tracks on quartz. As current passes through the metal, it will heat up. This heat will, in turn, change the index of refraction. For metals, the index of refraction is represented by a complex number where the imaginary part is the absorptivity. As absorption of the light changes, so will the reflectivity. From this reflectivity change, we can estimate the kinds of temperature changes that are occurring.

Once we had decided to pursue this investigation, we performed some preliminary tests using a far field system to see that the basic idea was sound. Though this yielded some reasonable images, we wanted to see where our limitations would be if we went to the near field system. We know that the light emitted from the tip is typically about 1 nanowatt. We need to see the amount that will get converted to an oscillating signal due to reflectivity changes and compare that to the noise limitations of the detectors. For this reason, we calculate the temperature variation and establish a relationship with the reflectivity. With that relationship, we can arrive at the expected reflectivity change, and compare that, finally, with theoretical noise limitations of the system (see Appendix B).

Starting from the basic heat flow equation, the following derivation of the variation of temperature in the metal is a simplification of the actual three dimensional case to a one dimensional case. It serves as an approximation of the kind of heating we expect to see, and will indicate the amount of reflectivity change expected. For this cursory analysis, we will assume one-dimension, one interface of air on quartz with an infinitesimally thin heating source, the aluminum, on the quartz (see figure 3-3). This assumption is reasonable since we know that the track is, in fact, thin, as well as the fact that aluminum is a good heat conductor. This means the temperature at the top and bottom faces of the track will be virtually identical. We have also assumed that current crowding is not taking place.
Figure 3-3: Schematic of layer structure used in reflectivity change as a function of heat.

That is, the current density over the surface of the aluminum is equal which means equal heating from all portions of the track. This calculation is intended to establish, roughly, the expected signal to noise. We have also assumed semi-infinite materials with only one boundary, the aluminum track.

To simplify things further, we will assume, instead of a current heating source, that we have a laser source whose absorption is achieved within a very small distance, within the boundary region. This assumption may, at first, seem strange. However, we can reasonably assume uniform heating of the aluminum track by the current, because the track is thin and aluminum is a good conductor of heat. The heating from a laser source would be nearly identical to heating from a current source. Later in the derivation, we will see that the absorptivity term contained in the laser source equation, actually is not significant in the final analysis.

With these assumptions in mind, we start with the one-dimensional case. The heat flow equation is in the form

\[ \frac{\partial^2 \Psi}{\partial z^2} - \frac{1}{\alpha'} \frac{\partial \Psi}{\partial t} = \text{Source} \]  

where \( \Psi \) is the temperature in terms of \( z \) and \( t \), and \( \alpha' \) is the thermal diffusivity, and is
related to the thermal wave number by the following equality.

\[ k^2 = \frac{i\omega}{\alpha'} \]  

(3.2)

and

\[ \alpha' = \frac{\kappa}{\rho c} \]  

(3.3)

in which \( \kappa \) is the thermal conductivity of the sample, \( \rho \) is the density, and \( c \) is the specific heat[50].

In this case, the heat source is assumed to be a laser source, absorbed within the aluminum layer, i.e. no penetration into the quartz. If we also assume that the aluminum is infinitesimally small, its structural properties play no role in the equation. For that reason, all the parameters, such as specific heat and thermal conductivity, will be those related to the quartz. As the derivation progresses, it will be clear that these assumptions are reasonable. We can write the source term as

\[ Source = \frac{I_o \alpha_q}{\kappa_q} \exp(-\alpha_q z) \exp(i\omega t) \]  

(3.4)

in which \( I_o \) is the source intensity, \( \alpha_q \) is the absorptivity of quartz, \( \kappa_q \) is the thermal conductivity of quartz, and \( \omega \) is the modulation frequency of the source. When we put equations 3.1 and 3.4 together, we get

\[ \frac{\partial^2 \Psi}{\partial z^2} - \frac{1}{\alpha'} \frac{\partial \Psi}{\partial t} = \frac{-I_o \alpha_q}{\kappa_q} \exp(-\alpha_q z) \exp(i\omega t) \]  

(3.5)

We assume that \( \Psi(z, t) \) is separable, and can be written as the product of the corresponding \( z \) and \( t \) parts. We also assume that because the heating source is modulated at \( \omega \), the temperature modulation will also vary at that frequency. So \( \Psi(t) \) must be also a sinusoidal function, and \( \Psi(z, t) \) can be written as

\[ \Psi(z, t) = \Psi(z) \Psi(t) = \Psi(z) \exp(i\omega t) \]  

(3.6)
Putting this into equation 3.5, we have

\[
\frac{\partial^2}{\partial z^2} [\Psi(z) \exp(i\omega t)] - \frac{1}{\alpha'} \frac{\partial}{\partial t} [\Psi(z) \exp(i\omega t)] = \frac{-I_0 \alpha_q}{\kappa_q} \exp(-\alpha_q z) \exp(i\omega t) \quad (3.7)
\]

\[
\Rightarrow \exp(i\omega t) \frac{\partial^2 \Psi(z)}{\partial z^2} - \frac{i\omega}{\alpha'} \Psi(z) \exp(i\omega t) = \frac{-I_0 \alpha_q}{\kappa_q} \exp(-\alpha_q z) \exp(i\omega t) \quad (3.8)
\]

\[
\Rightarrow \frac{\partial^2 \Psi(z)}{\partial z^2} - \frac{i\omega}{\alpha'} \Psi(z) = \frac{-I_0 \alpha_q}{\kappa_q} \exp(-\alpha_q z) \quad (3.9)
\]

So the equation is now only spatially dependent. This can be easily solved to yield

\[
\Psi(z) = \Psi_1 \exp(-ikz) + \frac{-I_0 \alpha_q}{\kappa_q} \exp(-\alpha_q z) \frac{1}{\left(\alpha_q^2 - \frac{i\omega}{\alpha'}\right)} \quad (3.10)
\]

This equation represents the thermal wave in the quartz. Assuming the source is associated with the quartz layer only, the equation for air will be the same without the source term, and with an opposite sign for \(z\), the propagation direction.

\[
\Psi(z)_{air} = \Psi_a \exp(+ik_\alpha z) \quad (3.11)
\]

Now that we have these equations, we must determine the boundary conditions for the situation, and solve accordingly. Referring back to figure 3-3, notice that the boundary is specified to be at \(z=0\). The two thermal boundary conditions associated with this interface come from the continuity of temperature and continuity of the temperature flux. This can be written as

\[
[\Psi+]_{z=0} = [\Psi-]_{z=0} \quad (3.12)
\]

and

\[
\left[-\kappa_+ \frac{\partial \Psi_+}{\partial z}\right]_{z=0} = \left[-\kappa_- \frac{\partial \Psi_-}{\partial z}\right]_{z=0} \quad (3.13)
\]

Substituting sample specific 3.10 equations into the appropriate sides of the boundary con-
dition equations 3.12 and 3.13, our initial equations are

\[ [\Psi_0]_{\text{quartz}} - \frac{I_0 \alpha_q}{\kappa_q \left( \alpha_q^2 - \frac{\imath \omega}{\alpha_q^2} \right)} = [\Psi_0]_{\text{air}} \]  

(3.14)

and

\[ -\kappa_q (-\imath k_q) [\Psi_0]_{\text{quartz}} + \kappa_q \frac{I_0 \alpha_q^2}{\kappa_q \left( \alpha_q^2 - \frac{\imath \omega}{\alpha_q^2} \right)} = -\kappa_a (\imath k_a) [\Psi_0]_{\text{air}} \]  

(3.15)

Notice that at this stage, we see that if the absorptivity is assumed to be large, which we have done by requiring all the source heating to take place in a very small thickness, then \( \alpha \), typically around \( 10^7 \), will dominate the denominator in both equations. For equation 3.14, this will mean that the \( \frac{\Psi}{\alpha_q^2} \) becomes negligible, and the \( \alpha_q \) in the numerator will cancel, leaving only the first power of \( \alpha_q \) in the denominator. One can see that if the absorptivity is relatively large compared with the \( \frac{\Psi}{\alpha_q^2} \) ratio, the absorbing material properties (aluminum) are negligible.

After making these adjustments, and solving the two equations, we end up with the temperature fluctuation in terms of the source intensity.

\[ \Psi(I_o)_{x=0} = \left[ 1.05 \times 10^{-5} \frac{K m^2}{W} \right] I_o \]  

(3.16)

We now must replace the source term \( I_o \) with the appropriate form of the current related power.

Power from current can be calculated from \( P = \eta I^2 R \). Here, \( \eta \) represents an efficiency term, since not all of the current will contribute to heating. If we assume that the source is sinusoidal, an assumption that was expected from the beginning of the derivation, then the power due to the current would be in the form

\[ P_c = i_o^2 R = i_o^2 R \sin^2(i \omega' t) \]

\[ \Rightarrow P_c = \eta i_o^2 R \left( \frac{1}{2} (1 + \cos(2i \omega' t)) \right) \]  

(3.17)

where \( \omega' = \frac{\omega}{2} \), \( R \) is the resistance, and \( i_o \) represents the average (DC) current. Because
we are immediately concerned with only the AC value, we will disregard the DC term. Furthermore, we can equate \( P_c \) with \( I_o \) with appropriate units. Since we have solved this using the laser source assumption, the equation requires \( I_o \) to be in units of intensity, power per unit area. Our source term is simply energy. To accommodate, we can set

\[
I_o = \frac{P_c}{A} \tag{3.18}
\]

where \( A \) is the area of the track that is being heated, and \( P_c \) is our current induced power. If we assume that the majority of heating is taking place in 100 square microns (using approximate dimensions of the tracks we investigated), then \( I_o \) will be given by

\[
I_o = \frac{\eta i_{c0} R \left( \frac{1}{2} (1 + \cos(2iw't)) \right)}{A} \tag{3.19}
\]

Substituting in the following estimated values, \( A = 1 \times 10^{-10} \text{m}^2 \), \( R = 3k \Omega \), \( i_{c0} = 5mA \), and \( \eta = 10\% \), we get the following equation for \( I_o \)

\[
I_o = \left( 3.75 \times 10^6 \frac{W}{m^2} \right) (1 + \cos(2iw't)) \tag{3.20}
\]

Combining equations 3.20 and 3.16, we get an equation for temperature fluctuation as a function of time.

\[
\Psi(t)_{t=0} = \left( 3.93 \times 10^1 \right) (1 + \cos(2i\omega t))K \tag{3.21}
\]

To find the maximum heat fluctuation, we set the cosine term to unity, and we get maximum temperature fluctuations of 40 degrees Kelvin.

To convert to reflectivity, we need to know the aluminum reflectivity change as a function of temperature. Following an analysis by Rosencwaig[51], we can take the sample reflectivity to be \( R \) where

\[
R = R_o + \frac{dR}{dT} \Delta T = R_o + \Delta R \tag{3.22}
\]
which leads to
\[
\frac{\Delta R}{R_o} = \frac{1}{R_o} \frac{dR}{dT} \Delta T
\]

(3.23)

\(R_o\) represents sample reflectivity at temperature \(T_o\), and \(\Delta T\) is the change in temperature from \(T_o\). We know that for aluminum, the temperature coefficient of reflectivity, or \(\frac{1}{R_o} \frac{dR}{dT}\), is \(2.5 \times 10^{-5}\). So using the following numbers, \(\Delta T = 40K\), \(T_o = 300K\), \(R = 0.92\), in the equation 3.23, we get \(\Delta R = 0.001\). This means we get about 0.1% change. Assuming we have 1nW of light, of which about 50% is collected and 0.1% of that is modulated, then the modulated power that arrives at the detector is on the order of tenths of picowatts or \(10^{-13}W\).

Now that we have the estimated reflectivity change, we must look at the minimum detectable power and see how the two compare. A noise analysis is presented in Appendix B. We found that the minimum detectable power\[54\] for our system is given by
\[
P_{\text{min}} = \frac{h v_s}{\eta} \Delta v
\]

(3.24)

For these particular experiments, our detector bandwidth, \(\Delta v\), was around 10Hz, the efficiency of the detector was \(\sim 20\%\), and the source light was an argon-ion laser whose temporal frequency, \(v_s\), is \(\sim 6 \times 10^{14} Hz\). Substituting these numbers in, we find the minimum detectable noise is about \(2 \times 10^{-17}Watts\). So the theoretical SNR should be on the order of 10000. From this analysis, the detector should be sensitive enough to see the variation\[2\].

There was another consideration that we became aware of during initial tests of the system. As indicated previously, the nature of the particular sample under examination has to be considered in interpreting any resulting image. For this sample, the current only flows in the aluminized region, and the heating effects are confined to areas very close to the current carrying regions. So any signals we detect should be concentrated around the edges of the track, where current crowding is expected. But if we consider only the inherent

\[2\] Bandwidth can be roughly estimated by \(\frac{1}{t_i}\), where \(t_i\) is the integration time. Typical integration times were 100ms, which leads to 10Hz bandwidth.
optical characteristics of the sample, we notice a large discrepancy in the optical reflectivity that exists between aluminum and quartz regions. As we raster scan across a track, the optical signal collected will be essentially a step function. A step function can be represented in Fourier space[53] as

\[ \mathcal{F}\{\text{step}(x)\} = \frac{1}{2} \delta(\xi) + \frac{1}{i2\pi \xi} \] (3.25)

The frequency contents of the signal fall off as \( \frac{1}{\xi} \). The collected signal will then include that temporal frequency distribution. Depending on the lockin frequency, scan speed, etc., this distribution could overlap with the detectable bandwidth leading to false signal reproduction. There are at least two methods to reduce or avoid this problem: 1) to scan slowly, which reduces the sharpness of the step function as seen by the detector, which, in turn, reduces the higher frequency content; or 2) modulate the current at a high frequency, such that the frequency remnants from the step function are far from the lockin bandwidth. For our experiments, we adopted a combination of both options to empirically determine a working compromise.

Our basic system configuration is the same as the porous silicon setup, reflection mode imaging (see figure 3-4). The only significant difference is that we now have a current source on the sample. The frequency of that current is sent to the lockin reference, and we are recording any collected light at twice that frequency. Also, only one detector is used since the light collected for both direct optical and thermal images is blue and no optical filtration is needed. The signal is filtered after collection into DC and AC components which contain direct reflected and thermal information, respectively.

### 3.3 Images, Results and Discussion

Our results were as follows. Concentrating our scans on junction regions from a small wire to a large pad, we wanted to analyze the thermal waves, which in this case, would indicate the spatial concentration of the current. As there is more current, there will be increased
Figure 3-4: SFM/NSOM setup for photothermal investigation.
heating. Theories behind current spreading at a junction indicate that the spatial spread is inversely related to current modulation[51]. That is, as the current modulation frequency goes up, the current spread at the junction becomes smaller. The resolution of our system is sufficient to see this kind of change, but the images obtained indicate one limitation, namely the low signal level.

The sample we looked at included several circuits of varying linewidths, from 0.25 microns up to 1 micron. We began with large range far field scans to see that the heating was, in fact, taking place at the junctions where it was expected. Figure 3-5 shows a 12x12 micron area at a junction. The circuit in this particular image had lines of 0.25 microns. They were not resolved in the lithography processing and bled together to form one thick line, about 3 microns, with a junction wire on the order of 0.5 microns. We see heat concentrated on that junction and further down the wires. With the system working we went on to take a small scan using far field microscopy.

Figure 3-6 shows direct reflected and photothermal images of a junction with current in the center wire, modulated at 158Hz and 1kHz. The heating is occurring, as expected, primarily at the immediate junction of the thin wire to the pad. The rest of the image is noisy but there is no heating signal. Notice the noise is primarily located on the regions where the pad exists. This is related to the light level. As explained above, because the pad is very reflective, the collected optical signal over the pad region is relatively large. This puts it into the shot noise region. Where there is little reflected light, over the quartz region, the signal collected by the PMT is low and most likely dark current limited. The bottom pair of images shows the same scan region, but the current frequency is only 158Hz. These scans seem to be less well resolved at the edges. This is probably due to the step function presence. Recall that for lower frequencies, the step function distribution is stronger. Its influence at 158Hz will be much greater than at 1kHz.

Our NSOM/SFM scans proved to be much more difficult, due to much reduced light levels. SNR improves as the square root of power. If we assume our light power was 10μW/μm² for the SOM scans, and now we have ~5nW/μm², the decrease in power is
Figure 3-5: Far field scanning microscope images showing direct reflected (a) and photothermal (b) signals. Current frequency at 1kHz.
Figure 3-6: Far field scanning microscope images showing direct reflected (a) and photothermal (b) signals, using current frequencies of 1kHz (top) and 158Hz (bottom).
roughly 1000, so the SNR will go down by a factor of 30.

Scans in figure 3-7 show heating, as expected in the center wire, where current is running. There is also evidence of noise along the other wires. This is due to the shot noise, and the step function presence. The current frequency for these scans was 20kHz. We see in the force image, a piece of dust on the center track. It seems to show considerable heating. In these images, the step function presence should be reduced by two factors. One is the high modulation frequency of the current. However the main reason is that the sample was scanned in such a way as to reduce the number of step functions encountered. Notice that if the raster scans are in line with the tracks, not perpendicular to them, there will be less reflectivity jumps and therefore, less noise from the step functions. Figure 3-8 shows the direct reflected optical image on the left, and photothermal on the right. It was performed with a very bright tip. We estimated a microwatt of light at the tip output. This indicates that the aperture was large, due to loss of coating. It was still sub-wavelength, but because the resolution is defined by that aperture, what these scans gain in SNR is lost in resolution. The light collected was saturating the PMT so a low noise photodetector was used. The thermal image is very well localized, in fact, the only evidence of signal is from the middle, or heated track.

A close inspection of the thermal image reveals high resolution imaging. Notice that maximum heating is taking place right at the junction of the center track and the large pad. This is an interesting result. The effects of electron migration were mentioned before, and we actually witnessed it while performing these experiments. In many cases, the circuit failed before we had a chance to get reasonable images. When examined under a microscope, after failure, the break always occurred at that junction. Because this image shows the highest heating at the junction, this indicates that there is maximum current crowding at that region. This, in turn, implies that the electron migration will be most significant here, and therefore, that would be the most likely area for failure.
Figure 3-7: Near field force (a) and photothermal (b) images of circuit. Current frequency of 20kHz.
Figure 3-8: Near field direct reflected (a) and photothermal (a) images of the circuit: Current modulation frequency was 50kHz.
3.4 Conclusions

We have demonstrated that near field photothermal microscopy can work, and provides high resolution thermal imaging.

One of the aspects that we would like to pursue in the future is to analyze the thermal wave spread as a function of current modulation frequency. As mentioned, it has been theoretically shown that the thermal wavelength changes with frequency[51]. A future experiment would include investigating the current as it leaves the junction into the large pad, and how the spreading there is affected by the current frequency. From all the images taken during this branch of the research, none showed sufficient heating in that region, indicating that, perhaps, a more sensitive system is required. We have been able to predict the thermal effects that may be encountered using theoretical modelings of the systems examined. Our particular setup may have been introducing noise above shot noise which was not accounted for in our rough noise calculations. Furthermore, the percentage of current associated with heating was a rough value. Nevertheless, the results obtained thus far are very encouraging.
Chapter 4

Conclusions

A force-regulated near field scanning optical microscope has been presented. The benefits of high resolution and localized illumination have been explained and presented with the help of images obtained in specific analyses of porous silicon and photothermal imaging. The system is not without faults and its applicability to various imaging configurations must be evaluated for each specific case. Yet, it has been shown to be a worthy companion to SEM.

In the area of porous silicon, this research has demonstrated the following:

- A near field optical system is a reasonable method to obtain new insights into the microstructural characteristics of porous silicon topography and its corresponding photoluminescence.

- There is a correspondence between luminescence and topography, with variations in luminescence over very small regions of the samples. It is apparent from these studies that there is a dependence of luminescence on topography, though not a direct correspondence.

- The spectra of different regions also show variation. The small silicon particles showed spectral peak shifts and different overall shapes.

- Both of the above results are in direct accordance with implications of the quantum wire theory.
As mentioned, we would like to continue a similar analysis on ever diminishing particles. The premise being if the spectral distribution is, in fact, quantum mechanically defined, we should see a narrowing of the spectra with smaller particles. Taken to the extreme case, a particle which consisted of one single quantum wire should, if the quantum wire theory holds, emit nearly monochromatic light.

Along those same lines of thought, if we were able to isolate a small piece of silicon, and affix it to the tip of a probe, we would have a NSOM probe with an in situ source. Development of a tip of this nature would be beneficial to the production model probe microscopes which are currently being marketed. By including the in situ source, much of the space, mechanics and launching mechanisms required for a laser source would be eliminated. Furthermore, the tip would not have to be a fiber optic. Silicon tips which are currently being produced using micro machining techniques, could be adapted to include the source.

The system was further shown to be easily reconfigured for photothermal imaging. The photothermal investigation demonstrated the following:

- The first photothermal imaging with a near field optical probe microscope was performed successfully.
- Images yielded high resolution and showed localized heating of microcircuitry.
- There were noise limitations apparent in some images. A more sensitive system might be necessary for more detailed results. However, this could be performed with a near field microscope, by simply altering the system in such a way as to reduce the noise levels.

As mentioned, further research plans would include tuning the system to reduce noise levels, and increase the sensitivity. Assuming that were accomplished, the next phase would include performing imaging of a circuit with different current frequencies and attempting to image the current crowding and the current spreading as a function of the modulation frequency.
Appendix A

Fabrication of Fiber Optic Probes for Simultaneous SFM and NSOM

A.1 Tip Fabrication

All the tips used for the research included here were manufactured in house. We used standard single and multimode fibers, which we pulled into tips using the basic method of heating the fiber under tension. This procedure was controlled using a commercially available micropipette puller. The puller included parameters which could be used to alter the final tip shape, which is an important factor in the ultimate performance, both optical and force, of the system.

Once a tip was made and determined to be reasonable in terms of shape, cleanliness, symmetry, etc., it required coating. This consisted of a thin chromium layer (a few nanometers), which enhances the adhesion of aluminum, followed by a layer of aluminum, estimated at ~150nm (calculation follows in Appendix A.2). This aluminum helps to guide any light that would have otherwise been unguided in the last few microns of the tip. It also confines the exit light to a single aperture plane.

Following figure A-1, we see the general setup for coating. We use a large evaporation chamber, inside which are two sources, chromium and aluminum. The fibers must be
Figure A-1: Isometric perspective diagram of tip rotator, front (a) and side (b) views, and schematic of evaporation procedure setup (c).
rotated during the process to ensure even coating of the surface. For this purpose, we have constructed a gear system as diagramed. This consists of four gears each with hollow shafts. These shafts, which are large enough for a fiber diameter, hold each tipped fiber individually, and are driven by a small DC motor. This gear system, once loaded with fibers, can be mounted in the vacuum chamber at any height and angle. These two factors, height and angle, play an important role in the resulting coat. If there is not a sufficient angle between the fiber and the source, the tip of the fiber will get coated closing the aperture and rendering the fiber useless for NSOM. To prevent this, the tips are generally anchored about 10° from the horizon. Assuming that the evaporating source can be approximated as a point source, we see that spherical waves of the coating will propagate from that source, coating the tip. With the tip at a slight angle from horizontal, the very end of the tip is shadowed during the coating procedure and remains uncoated. This creates the aperture. We estimate this aperture to be around 10-20nm on a good tip.

With the tip pulled and coated, it is ready for use. The average success rate is around 12%. One tip in eight survives the entire process from pulling to the final mounting in the system.

A.1.1 Tip Performance Concerns

These tips, as explained, come in many different lengths and shapes. Long narrow tips provide good force resolution and good depth penetration, but typically have low light or lossy optical qualities because the length of fiber that is non-guiding is longer. Figure A-2 shows three possible tip geometries and the corresponding optical and topographical implications. When the tip is pulled, there is some point along the cone at which the fiber loses its light guiding properties. This is due to the fact that as the cone gets smaller, at some point, its diameter will be less than the core diameter of the fiber. At this point, the fiber is no longer a wave guide, but simply a conical piece of glass. It is at this point that the aluminum becomes the primary guiding mechanism. Long narrow tips have a longer region for which the aluminum is guiding since the core diameter is reached sooner. They
are therefore less efficient (generally) providing less output light. Blunt tips tend to have longer optical confinement, and retain more of the power output at the tip, but poor depth penetration and topographical resolution. A tip somewhere in between these two yields a suitable compromise of optical quality and force resolution.

Once we have completed these procedures of pulling and coating, the tip must be investigated for its optical qualities. We cleave the back end of the fiber and launch light into the tip. If the tip had any dust on it during the coating phase, the aluminum will not coat sufficiently, which can be seen during a close examination. A good tip will show light emitted only at the tip, and that spot will be diffraction limited. Another concern, as previously mentioned, is that the tip might be blocked, no aperture. In this case, there are methods of "opening" up the aperture. We have found that occasionally, light power launched into the fiber, if sufficient, will heat the tip enough to make the aluminum begin to flake off. In some cases, it will actually "blow" a hole at the tip. If this method fails, we setup the SFM feedback system and bring the tip into engagement on a sample. With a very "rough" engagement, i.e. a large feedback oscillation amplitude, the tip will actually contact the sample just enough to loosen some of the aluminum at the tip, so that it flakes off.
Figure A-3: Schematic of spherical radiating source, for coating calculations.

Using this basic methodology, we were able to create tips within about one day’s time for use. Much of the "technology" simple came from trial and error. But we have found a reasonably consistent method, that with enough patience, will provide us with tips.

A.2 Coating Thickness on SFM/NSOM Tips

We are interested to see how much coating is being deposited onto the aluminum because that will affect the force resolution. To calculate this value, we proceed as follows. Assuming that the coating surface is a sphere, and following figure A-3, we start with the fact that the volume of the Al pellet, given by $V = \frac{m}{\rho}$, is entirely deposited onto the tip. Thus that volume must be equal to the to the difference between the volumes of the two spheres, one with radius $R$, and the other of radius $R - \Phi$.

\[
V = \frac{m}{\rho} = V_R - V_{R-\Phi} = \frac{4}{3} \pi R^3 - \frac{4}{3} \pi (R - \Phi)^3 \tag{A.1}
\]

\[
= \frac{4}{3} \pi [R^3 - (R^2 - 2R\Phi + \Phi^2)(R - \Phi)] \tag{A.2}
\]

If we then assume that $\Phi$ is small compared to $R$, then the $\Phi^2$ terms will drop out and
we will be left with the following:

$$\Phi = \frac{m}{4\pi \rho R^2} \tag{A.3}$$

where $m =$ mass of Al, $\rho =$ density of Al, and $R =$ distance from Al source to surface being coated.

Now, if we take the mass evaporation to be a function linear in time, then $m = m(t) = \alpha t$, where $\alpha$ is some constant. Using this assumption we see that:

$$d\Phi = \frac{a}{4\pi \rho R^2} dt \tag{A.4}$$

Our cylinder is rotating at some angular velocity $\omega_0 = \frac{\theta}{t}$ and so $dt = \frac{d\theta}{\omega_0}$. Putting this all together along with a projection factor of $\sin(\theta)$, we get the following:

$$d\Phi = \frac{a}{4\pi \rho R^2} dt \cdot [\sin(\theta)] = \frac{a}{4\pi \rho R^2} \sin(\theta) \frac{d\theta}{\omega_0} \tag{A.5}$$

To see the effect of a full rotation, we simply integrate over $2\pi$.

$$\Phi_{2\pi} = \int_{0}^{\pi} \frac{a}{4\pi \rho R^2} \sin(\theta) d\theta = \frac{a}{2\pi \rho R^2} \omega_0 \tag{A.6}$$

The integration limits are 0 to $\pi$ to account for the fact that half of the cylinder is shadowed the whole time. If we assume that in a time $\sigma$ there are $n$ turns, where $n = \frac{\omega_0 \sigma}{2\pi}$ then the total thickness evaporated in a given time $\sigma$ is given by $\Phi_{\sigma} = \frac{a\sigma}{4\pi^2 \rho R^2} (NOTE: \Phi_{static} = \Phi_{rotating \pi})$

From the static case, we can calculate the coating from the pellet of Al (0.54g with a density of 27 g/cm$^3$) and a target placed 20cm from the source. We find that the coating thickness is

$$\Phi_{static} = \frac{(0.54g)}{4\pi (27 \frac{g}{cm^3})(20cm)^2} \approx 400nm \tag{A.7}$$
\[ \Phi_{\text{rotating}} = \frac{\Phi_{\text{static}}}{\pi} \approx 125 \text{nm} \]

So if our tip size estimation is correct at 50nm, then our overall tip diameter after coating should be around 300nm, \((50nm + 125nm + 125nm)\). This diameter limits the force resolution, but not the optical resolution which is defined by the aperture size.
Appendix B

Photomultiplier Detectors and Associated Noise

Because in the near field optics, the source light level and collected light level are typically small, there is a need for highly sensitive, relatively fast, and low-noise detection systems. For this reason, the photomultiplier tube was our detector of choice for most applications.

B.1 Photomultiplier Tubes (PMT)

Photomultiplier tubes work by collecting light at a photocathode. This photocathode gives off electrons when photons strike it. The electrons are drawn from this plate, toward the first dynode due to a voltage potential. Each successive dynode has an increased potential so the electrons continue to follow a path from the cathode to the anode. The typical gains of \( G=10^5 \) comes from the dynodes where

\[
G = \delta^N
\]

(B.1)

\( N \) is the number of dynodes and \( \delta \) is the number of secondary electrons generated for each incident electron. When the dynode is hit with an electron, typically at about 100eV,
secondary emission of electrons causes an increase in the current. This process is repeated at each terminal until the anode is reached, by which time, the number of electrons has increased by five or six orders of magnitude. This high gain system is very useful in near field imaging.

B.2 Noise

Noise is, in general, the deviation of the collected signal from the actual signal. In this section and the following two, we calculate the expected noise performance of a PMT following derivation included in Yariv's book on optoelectronics[54]. We begin with the a situation in which the optical signal field to be detected is \( v_s \), where

\[
v_s(t) = V_s \cos(\omega t)
\]  

(B.2)

This assumes a clean sinusoidal signal. We can include a noise term, with the assumption that instantaneous noise can be broken into two pieces, an in-phase component and an orthogonal component. This noise term can then be written as

\[
v_N(t) = V_{NC}(t) \cos(\omega t) + V_{NS}(t) \sin(\omega t)
\]  

(B.3)

where \( V_{NC} \) and \( V_{NS} \) are random, slowly varying, uncorrelated values with a mean value equal to zero. The total field collected at the detector is then represented by the sum of the noise and signal fields.

\[
v_{total}(t) = v_s(t) + v_N(t) = V_s \cos(\omega t) + V_{NC}(t) \cos(\omega t) + V_{NS}(t) \sin(\omega t)
\]  

(B.4)

Using Euler's relations, the equation can be rewritten as

\[
v_{total}(t) = Re\{V(t)e^{i\omega t}\}
\]  

(B.5)
where $V(t)$ is the phasor representation of the total field. From ensemble averaging\(^1\), it can be shown that $V_{NC}$ is zero while the mean square value is given by

$$
V^2_{NC} = V^2_{NS} = \sigma^2
$$

Here $\sigma$ represents the mean value. Power can be obtained then from time averaging for time $T$.

$$
P(t) = [v_{total}(t) \cdot v^*_{total}(t)] = V_S^2 + 2V_S V_{NC} + V^2_{NC} + V^2_{NS}
$$

Then time averaging this, gives

$$
P(t) = V_S^2 + V^2_{NC} + V^2_{NS} = V_S^2 + 2\sigma^2 \tag{B.7}
$$

This gives us the average power over a long integration period. Because each measurement is subject to the noise, the uncertainty associated with this power measurement will provide some useful noise information. To get the uncertainty, we take the root mean square of the power fluctuations.

$$
\Delta P = [(P(T) - \overline{P})^2]^{\frac{1}{2}} \tag{B.8}
$$

Substituting in for $P(t)$ and $\overline{P}$, and using the relationship above showing that $V^2_{NC} = V^2_{NS} = \sigma^2$, it can be shown that

$$
\Delta P = 2\sigma \sqrt(V_S^2 + 2\sigma^2) \tag{B.9}
$$

From equation B.7, we can let $P_S = V_S^2$ assuming $P_S$ is signal measured with no noise. Now that we have an idea of the deviation of the real signal due to noise, we must relate that

---

\(^1\)The ensemble average is a statistical tool which can be found by taking the average of simultaneous measurements from a large array of "identical" systems. For randomly varying functions in time, the ensemble average is equal to the time average of one of the particular systems.
to the actual power recorded. If we assume that our measured minimum power can be no lower than the fluctuation, then $P_s$ must be equal to $\Delta P$, so that

$$P_{\text{min}} = 2\sigma \sqrt{P_{\text{min}} + 2\sigma^2}$$  \hfill (B.10)

Then solving for $P_{\text{min}}$, we find

$$P_{\text{min}} = 2\sigma^2 (1 + \sqrt{2})$$  \hfill (B.11)

The $2\sigma^2$ term can be renamed $P_N$ which is the noise power. The assumption that $P_s$ must be equal to $\Delta P$ for minimum detection may seem a bit arbitrary but logical also. If the signal is lower than the noise, distinguishing it from the maximum deviation of power may be difficult. Typical calculations for noise limits, however, assume that the signal and noise power are equal (Signal to Noise (SNR) is unity). This makes for ease of calculations. For this cursory analysis, that approximation will be sufficient.

### B.3 Noise in photomultiplier tubes.

There are three main noise mechanisms present in photomultiplier detection: shot-noise, thermal or Johnson noise, and dark current noise. Shot noise is primarily a result of cathode shot-noise, though there is the presence of dynode shot-noise as well. Because noise from the dynodes is much less amplified than the signal from the cathode, the dynode noise is small enough it can be ignored in this analysis.

For the following analysis, we can assume there is a source light, modulated by some chopping frequency, which impinges on the sample and is collected by the photocathode. This modulated (video) signal contains optical intensity information in the amplitude. Assuming the input signal is

$$e_s(t) = E_s (1 + m \cos \omega_m t) \cos \omega_s t$$  \hfill (B.12)
where \( \omega_m \) represents the modulation frequency, and \( \omega_s \) represents the optical frequency. This leads to a photocathode current which will be proportional to the square of the incoming optical signal.

\[
i_c(t) \propto [E_s(1 + m \cos \omega_m t)]^2 \quad (B.13)
\]

This represents the current time averaged over, at least, a few wavelengths of light (a few femtoseconds).

We can get the time averaged photocathode current, and the proportionality constant using the quantum efficiency of the photocathode, \( \eta \), the frequency of the light, \( \nu \), where \( \nu = \frac{c}{\lambda} \), along with the average power, \( P \).

\[
\bar{i_c} = \frac{P \eta}{h \nu} \quad (B.14)
\]

Knowing all this, we find that the signal output current will be

\[
i_s = \frac{G P \eta}{h \nu} (2m) \cos \omega_m t \quad (B.15)
\]

where \( G \) is the gain of the multiplier tube, typically around \( 10^6 \). Assuming the detector is limited by a bandwidth \( \Delta \nu \) centered around the modulation frequency \( \omega_m \), then the shot-noise due to that system will be

\[
\bar{(i_{shot}^2)} = 2G^2 e(\bar{i_c} + i_d) \Delta \nu \quad (B.16)
\]

The dark current is \( i_d \). Thermal noise, or Johnson noise, is thermally generated noise and can be written from thermodynamic equations. Its average power is given by

\[
\bar{(i_{Nj}^2)} = \frac{4kT \Delta \nu}{R} \quad (B.17)
\]

\( R \) here is the resistance connected across the anode.

Putting all this together, we want to look at the minimum detectable power incident on
the detector. We can compare that number to the experiment in question and analyze our approach.

SNR, signal to noise ratio, can be calculated using the above equations B.15, B.16, and B.17. Noise adds linearly so we have

$$\text{SNR} = \frac{S}{N} = \frac{i_s^2}{(i_{N,\text{shot}}^2 + i_{N,j}^2)} = \frac{2\left[\frac{GP\eta}{hv}\right]^2}{2G^2e(i_c + i_d)\Delta v + \frac{4kT\Delta v}{R}}$$

(B.18)

## B.4 Minimum Detectable Power for Dark Current Limit and Shot Noise Limit

Dark current is the noise signal associated thermal excitation only; that is, dark current is the signal from the detector with no incident photons. This can be an important factor in NSOM due to the low illumination power. Typical tips emit between 1nW and 10nW. This requires the use, in most cases, of a high gain, high sensitivity detector like a photon counter or a photo-multiplier tube. Often the desired imaging modality requires collection of very low signals, as in the photoluminescence experiments performed. For this reason, certain systems/investigations can requires systems which are dark current limited. For a signal to noise (SNR) ratio to be one, minimum detectable signal, the incoming signal must be at least equal to the dark current.

To find the minimum detectable power, simply solve this for P. For the dark current limited case, we assume \(i_d \gg i_c\) and SNR=1. We also assume that because \(G\) is large, the Johnson noise term in the denominator becomes negligible and can be ignored. Substituting \(P_{\text{min}}\) for \(P\), and solving, we get

$$P_{\text{min}} = \frac{hv_s}{\eta} \sqrt{\frac{i_d\Delta v}{e}}$$

(B.19)

For typical values, \(v_s = 5 \times 10^{14}\ Hz\), \(\eta = 10\%\), \(\Delta v = 1Hz\), \(e = 1.602 \times 10^{-19}C\), and \(h = 6.626 \times 10^{-34}Js\), we can solve for \(P_{\text{min}}\). These numbers give \(P_{\text{min}} \approx 4\times10^{-8} \cdot \sqrt{i_d}\). Assuming a typical dark current value of \(1 \times 10^{-20}A\), then dark current noise would be on
the order of $4 \times 10^{-18} J$.

For minimum detectable power in a shot noise limited case, we make the same assumptions, that $G$ is large enough to disregard Johnson noise, and that $i_c \gg i_d$. By making the assumptions here that there is no contribution from dark current or thermal noise, we can actually get an approximate minimum power from equations B.18 and B.14. We get the shot noise limit as with dark current.

$$P_{\text{min}} = \frac{h v_s}{\eta} \sqrt{\frac{i_c \Delta v}{e}} \quad \text{(B.20)}$$

Then substitute in the value of the average cathode current to get

$$P_{\text{min}} = \frac{h v_s}{\eta} \sqrt{\frac{P_{\text{min}} \eta e \Delta v}{e}} \quad \text{(B.21)}$$

Then we rearrange, and see that the shot noise limited, minimum detectable power can be written as

$$P_{\text{min}} = \frac{h v_s}{\eta} \Delta v \quad \text{(B.22)}$$

This equation is interesting because the dependence on the actual current has dropped out. The only variable here is the bandwidth. If we use the "typical values" stated above, we see that $P_{\text{min}}$ is on the order of $20\text{eV}$, or $10^{-18} J$. This is sometimes called the quantum limit. The benefit of being in the shot noise limited region of operation is that the only limitation is the Poisson noise from quantized photons. There is no dependence on any detector parameters. This is the ultimate limit in detectability.
Bibliography


