Precision Optical Characterization on Nanometer Length and Femtosecond Time Scales

by

Daniel Bender

B.S., Physics, University of New Mexico, 2002

DISSERTATION

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
Optical Science and Engineering

The University of New Mexico

Albuquerque, New Mexico

December, 2008
©2008, Daniel Bender
Dedication

To my mother and father.
Acknowledgments

The completion of this dissertation comes with the help of many people. The success of this work is a result of the counsel I received from my adviser Professor Mansoor Sheik-Bahae. His support and guidance are gratefully acknowledged. My deepest appreciation goes to Professor Michael Hasselbeck for his enduring patience and technical expertise. I thank Professor Richard Epstein for his ideas and encouragement. The contributions from these great minds made this work possible.

Support provided through the Cross-disciplinary Optics Research and Education (CORE) program, as well as, the Multi-University Research Initiative (MURI) was instrumental and is gratefully appreciated.

Among faculty members I extend my personal thanks to Prof. Wolfgang Rudolph, Prof. Kevin Malloy, Prof. Jean-Claude Diels, Prof. Stephen Boyd and Dr. Andreas Stintz. I greatly appreciate the teaching, equipment and support each of you has granted over the years.

I also would like to thank cooperative researchers Jeff Nicholson, Balakishore Yellampalle, Sarah Kurtz and Mark Hoffbauer.

My personal thanks also goes to my fellow students: Wendy Patterson, Denis Seletskiy, Babak Imangholi, Chengao Wang, Chia-Yeh Li, Seth Melgaard, Nima Vadiiee, Shahed Azariyan, Ed Soto, as well as, post-docs Mahendra Shakya and Luke Emmert, who have assisted in discussions and discovery from beginning to end.

Many thanks are also due to the Physics and Astronomy Staff: Daniel Sandoval, Sandra Ortiz, Jaye Jenson, (the) Carl Allen, John DeMoss, John Behrendt, Gary Harrison and Steve Portillo.
Precision Optical Characterization on Nanometer Length and Femtosecond Time Scales

by

Daniel Bender

ABSTRACT OF DISSERTATION

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
Optical Science and Engineering

The University of New Mexico

Albuquerque, New Mexico

December, 2008
Precision Optical Characterization on Nanometer Length and Femtosecond Time Scales

by

Daniel Bender

B.S., Physics, University of New Mexico, 2002
Ph.D, Optical Science and Engineering, University of New Mexico, 2008

Abstract

This dissertation details progress made in ultrafast optics and time resolved near-field microscopy. First, I will discuss experiments in ultrafast optics. Careful characterization of ultrashort laser pulses is critical for many applications. Because pulse durations are less than a fraction of a picosecond, direct electronic measurement is not possible and optical techniques must be used. I will describe a simple, real-time scheme for visualizing subtle pulse distortion that occurs on the femtosecond time scale. The procedure for generating the modified spectrum autointerferometric correlation (MOSAIC) from the second harmonic generation frequency resolved optical gating (SHG FROG) dataset is shown. The MOSAIC trace provides complimentary visual pulse assessment to SHG FROG. Spectral phase retrieval from the pulse spectrum and MOSAIC is examined. Next, I discuss the concept of laser cooling of solids, i.e. all optical refrigeration in glasses, crystals and semiconductors. Near-field
scanning optical microscopy (NSOM) is used in tandem with atomic force microscopy (AFM) to characterize semiconductor heterostructures for potential use in laser cooling. A novel combination of normal force feedback AFM and collection mode NSOM provide unique coincident imagery. Subsurface defects, which adversely affect laser cooling efficiency, are identified and lower nonradiative recombination lifetime by a factor of 10 compared to defect free areas. Time correlated single photon counting (TCSPC) resolves near-field fluorescence lifetimes and simultaneously gives correlation with surface topography. Fluorescence lifetime images (FLIMs) give direct correspondence to and supply supplemental data for AFM.
Contents

List of Figures xiv

Glossary xxiv

I Ultrafast Optics 1

1 Introduction to Ultrafast Optics 2

1.1 Overview ......................................................... 2

1.2 Pulse Propagation in Materials ................................. 5

1.3 Summary .......................................................... 8

2 Pulse Characterization 9

2.1 Overview .......................................................... 9

2.2 Techniques ........................................................ 9

2.3 The MOSAIC algorithm ......................................... 13

2.4 Envelope MOSAIC ............................................... 18
Contents

4 Ti:sapphire Laser System 63

4.1 Overview .......................... 63

4.2 Ti:sapphire Laser ..................... 63

4.3 Autocorrelator ......................... 66

5 Summary and Future Work: Ultrafast Optics 70

II Time-Resolved Microscopy and Its Application to Laser Cooling in Semiconductors 72

6 Laser Cooling of Solids 73

6.1 Overview .......................... 73

6.2 Laser Cooling ........................ 73

6.3 Laser Cooling in Glasses and Crystals ...................... 74

6.4 Laser Cooling in Semiconductors ...................... 81

6.5 The Nature of Nonradiative Recombination ................. 86

6.5.1 Shockley-Read-Hall model .................. 86

6.5.2 Surface States ...................... 87

6.6 Carrier Diffusion ........................ 90

6.7 Carrier Dynamics ........................ 92

6.8 Summary .......................... 97
7 Near-Field and Atomic Force Microscopy

7.1 Overview .................................................. 98
7.2 Microscopy .................................................. 99
7.3 Atomic Force Microscopy ................................ 101
7.4 Near-field Microscopy ..................................... 104
7.5 AFM Characterization of GaInP/GaAs Double Heterostructures ... 108
7.6 Time Correlated Single Photon Counting ............... 112
7.7 Integration of Photon Counting with Near-Field Microscopy ... 117
7.8 Fluorescence Lifetime Imaging: Understanding the effects of surface defects in GaInP/GaAs DHS .................. 119
7.9 Fluorescence Lifetime Imaging: Unpassivated GaAs Edge ...... 122
7.10 Fluorescence Lifetime Imaging: Subsurface Defect at GaAs Edge ... 130
7.11 Fluorescence Lifetime Imaging: Bulk GaAs ..................... 131
7.12 Fluorescence Microscopy (Far-Field) ....................... 137
    7.12.1 Diffraction Limited Fluorescence Imaging: Subsurface Defects 138
    7.12.2 Far-field Fluorescence Imaging: GaInP Passivation Layer .... 141
    7.12.3 Far-field Fluorescence Imaging: Processing Induced Defects . 141
7.13 Summary .................................................... 145

8 AFM-NSOM System ............................................. 147

8.1 Overview .................................................. 147
Contents

8.2 Instrumentation ................................................................. 147

9 Summary and Future Work: Time-Resolved Microscopy ............... 154

Appendices .............................................................................. 156

A Solutions to the carrier recombination equation ....................... 157

B Fitting the Decay Tail of Semiconductor Luminescence ........... 161

C Comprehensive Fitting of the Full Luminescence Decay ........... 165

References .............................................................................. 168
List of Figures

1.1 A pulse shown unchirped (left) and chirped (right). Chirp is shown by an evolving frequency across the pulse. ........................................... 6

1.2 Calculated group velocity dispersion for fused silica, SF10 and BK7. 7

2.1 Experimental setups for pulse characterization using PICASO, SHG FROG and SPIDER. ................................................................. 13

2.2 A Michelson interferometer used to produce a second order interferogram. The pulse is interfered with itself as a function of delay, \( \tau \). ........................................................................................................... 15

2.3 (a) an IAC calculated from and unchirped pulse. (b) MOSAIC trace for the IAC of (a), note the flat baseline. A chirped pulse used in (c) IAC and (d) MOSAIC calculation. The MOSAIC trace shows a flat baseline for unchirped pulses and shoulders for chirped pulses. The IAC traces look nearly the same. ...................................................... 17
List of Figures

2.4 MOSAIC signals generated from the same pulse and detected using (a) a BBO crystal and a linear detector (pulse duration: 60 fs FWHM) and (b) a two photon absorbing LED (76 fs FWHM). Insets shows fringe resolved MOSAIC. Measured IAC signals from which the MOSAIC traces in (a) and (b) were derived are shown in (c) and (d), respectively. The structure of the IAC waveforms appears almost identical, while the MOSAIC traces reveal chirp induced by the detection method. 19

2.5 MOSAIC traces obtained from (a) an 85 fs Ti:sapphire laser and (b) intentionally chirping the laser pulse with 2 mm of ZnSe. MOSAIC rendered from a SHG spectral measurement (circles) and a second-order IAC (solid line). Insets show same data on a semi-log scale. 21

2.6 Linear and nonlinear autocorrelations 22

2.7 a) Simulation of H-MOSAIC for a pulse having a symmetric spectrum with (a) no dispersion, (b) GVD, (c) TOD and (d) an asymmetric spectrum with no dispersion. The dashed line indicates zero delay. 24

2.8 Experimentally obtained IAC and MOSAIC traces 27

2.9 a) Single IAC trace just above the noise level. b) Averaged MOSAIC waveform produced from 1200 noisy IAC traces (red triangles); the chirp of the same pulse obtained with negligible noise is reproduced (solid line) 28

2.10 Experimental data (red triangles) and fit (green lines) that reveal high-order components of chirp on a 120 fs Ti:sapphire laser pulse. 29
List of Figures

2.11  a) Single IAC trace just above the noise level of a frequency doubled Ti:sapphire laser pulse ($\lambda = 415$ nm). b) Averaged MOSAIC waveform produced from acquiring 1000 noisy IAC traces.  

2.12  Chirp response in MOSAIC originating from LEDs that differ in manufacturer and color. The lowest chirp measurement was obtained through SHG in a BBO crystal.  

2.13  A cross-sectional view of the metal-semiconductor-metal detector structure.  

2.14  (Left) Chirp response in MOSAIC. (Right) signal-to-noise of the second order IAC as an MSM detector is brought through focus.  

2.15  Calculated SHG FROG traces from an unchirped (a) and chirped (b) pulse. Corresponding MOSAIC traces produced from the SHG FROG traces, unchirped (c) and chirped (d). It is important to note the FROG traces appear identical; complimentary MOSAIC traces reveal pulse chirp without the need for iterative reconstruction.  

3.1  The R-MOSAIC algorithm. Higher order phase terms are accessed sequentially to accommodate spectral complexity.  

3.2  Simulation of MOSAIC based retrieval with GVD.  

3.3  Simulation of MOSAIC based retrieval with TOD.  

3.4  Simulation of MOSAIC based retrieval with GVD and TOD.  

3.5  Simulation of MOSAIC based retrieval with noise.
List of Figures

3.6  Error maps for (a) an IAC, (b) a fringe resolved MOSAIC, (c) E-MOSAIC and (d) E-MOSAIC with DFP on a pulse having GVD, TOD and a symmetric spectrum. ........................................ 52

3.7  (a) MOSIAC traces computed for a pulse with a symmetric spectrum and -GVD/+TOD and +GVD/+TOD. (b) DFP for cases in (a) as well as no dispersion. ........................................ 54

3.8  Normalized DFP signals showing sensitivity to relative sign on GVD and TOD dispersion coefficients. ........................................ 55

3.9  100 noisy normalized DFP traces for 2% additive white noise to pulse spectrum (red lines). Averaged DFP (blue dots) and target DFP (black line) are also shown. ........................................ 56

3.10 a) An experimental H-MOSAIC (solid line) and reconstructed H-MOSAIC (circles) from the phase (red dots) and measured spectrum of (b). ........................................ 57

3.11  (a) Measured second-order IAC (b) Experimental MOSAIC (pink lines) and reconstructed MOSAIC (dots) from the measured spectrum and retrieved phase of (c). Time domain pulse (d). ........ 58

3.12  A retrieved pulse from E-MOSAIC and DFP ........................................ 59

3.13  Normalized algorithm performance for spectral phase retrieval on our 60 fs Ti:sapphire laser pulses. All algorithms are evaluated with the MATLAB software platform. ......................... 61

4.1  Ti:sapphire cavity producing 60 fs pulses ........................................... 65
List of Figures

4.2 Photo of the pulsed Ti:sapphire laser cavity. The pump laser path is shown in green, Ti:sapphire cavity is shown in red. Fused silica prisms are used for intracavity dispersion compensation. . . . . . . . 66

4.3 Michelson interferometer with component manufacturers listed. . . . 68

6.1 The laser cooling cycle. Electrons are excited from the top of the ground state to the bottom of the excited state by a pump photon of frequency $\nu$ (red) where they acquire additional energy from the surrounding lattice (yellow). Fluorescence is emitted from the system at energy $h\nu_f$ which carries away thermal energy of the lattice (blue). 75

6.2 Schematic of an optical refrigeration system. Pump light is efficiently generated by a semiconductor diode laser and carried to the mirrored cooler element by an optical fiber. The laser enters the cooler through a pinhole in one mirror and is trapped by the mirrors until it is absorbed. Isotropic fluorescence escapes the cooler element and is absorbed by the vacuum casing. A sensor or other load is connected in the shadow region of the second mirror. . . . . . . . . . . . . . . 76

6.3 (a) Cooling cycle in laser refrigeration of a semiconductor in which absorption of laser photons with energy $h\nu$ creates a cold distribution of electron-hole carriers (only electron distribution is shown for clarity). The carriers then heat up by absorbing phonons followed by an up-converted luminescence at $h\nu_f$. (b) Typical anti-Stokes luminescence observed in GaAs/GaInP double heterostructure . . . 83

6.4 A GaAs/GaInP double heterostructure is bonded to a nearly indexed matched dome (ZnS or ZnSe) lens for luminescence extraction. . . . 84
List of Figures

6.5 (a) Nonradiative recombination through multiphonon emission. (b) Radiative recombination giving off a photon of energy $h\nu_f$. (c) Auger recombination. ................................................. 85

6.6 (a) A semi-infinite periodic chain of atoms representing a semiconductor with a surface at $z = 0$. (b) Solution to the time independent Schrödinger equation showing positional dependence of surface states at a semiconductor/vacuum interface. ........................................... 88

6.7 Defect mediated recombination .......................................................... 89

6.8 Photoluminescence from Eq. 6.24. (a) $N_d << N_o$ late time single exponential behavior is evident (blue line). When SNR < 1 in the late time, single exponential behavior becomes buried beneath the noise floor (red dashed line). (b) $N_d \geq N_o$ single exponential behavior spans the measurement time. Experimental examples for the conditions in (a) and (b) are shown in (c) and (d) respectively. .... 96

7.1 Near-field and far-field spot size comparison ..................................... 100

7.2 Normal force feedback mechanism for AFM ...................................... 102

7.3 The four modes of operation in NSOM. (a) Transmission: local excitation global collection from below sample, (b) Reflection: local excitation global collection from above sample, (c) Collection: global excitation and local collection and (d) Illumination/Collection mode: local excitation and collection are performed through the tip. .... 106

7.4 Dual AFM/NSOM setup. This scheme is used to create spatially overlapping, simultaneously acquired AFM and NSOM images. .... 108
List of Figures

7.5 A cross-sectional view of an MOCVD sample grown by NREL. The sample has a 30 nm passivation layer on top. 109

7.6 Schematic view showing CCD incorporation to AFM imaging. 110

7.7 (a) Optical image taken with a 10x magnification microscope objective. Surface defects are seen as black irregularly shaped objects. Probe tip is also visible. (b) AFM image of the area outlined by the white box in (a). The surface defect has a height of $\approx 3.5 \ \mu m$ from the planar background. 111

7.8 The concept of TCSPC 115

7.9 Reverse start-stop timing method of TCSPC. The clock starts on a photon event and stops on the next laser pulse. 116

7.10 (a) Fluorescence lifetime image of a surface defect. The lifetime is reduced by a factor of 10 relative to the defect free regions. (b) An AFM image of the defect in (a). The AFM image was obtained in a second scan. 120

7.11 Dots arrayed on a sample surface. The mesa structure is shown from the top and side views. The circumference of the GaAs layer is unpassivated. 122

7.12 (a) An AFM, (b) integrated intensity image, (c) 100x optical image and (d) FLIM of a GaAs double heterostructure. Green box in (c) depicts the approximate scan area. Dashed lines in (b) and (d) indicate the position of the sample edge as obtained from the simultaneously acquired AFM image (a). 124
List of Figures

7.13  (a) Average nonradiative lifetime as a function of position approaching an unpassivated sample edge and (b) on a log scale. Carrier recombination < 8 \( \mu \)m from the sample edge experiences very fast (<10 ns) nonradiative decay. ................................. 128

7.14  (a) Measured PL intensity approaching an unpassivated surface in GaAs (positioned at 22 \( \mu \)m). (b) Calculated transient carrier diffusion in GaAs following impulse excitation. (c) Calculated PL intensity approaching an unpassivated GaAs edge. Inset shows region highlighted in red. Intensity reduction is due to diffusion and surface recombination. ................................. 129

7.15  Dots arrayed on a sample surface. Area of second FLIM is denoted by the black box. ................................. 130

7.16  (a) An AFM, (b) integrated intensity image, (c) 100x optical image and (d) FLIM of a GaAs double heterostructure. Green box in (c) depicts the approximate scan area. Dashed lines in (b) and (d) indicate the position of the sample edge as obtained from the simultaneously acquired AFM image (a). ................................. 132

7.17  (a) AFM image on bulk GaAs passivated with 30 nm of GaInP. No surface features were found. (b) Integrated intensity image from the luminescence collect during the AFM of (a). (c) Optical image showing sample surface. Surrounding surface defects are highlighted with purple circles. Area of scan is denoted by the green box. White lines connecting surface defects overlap with features of the superimposed intensity (or lifetime) image. (d) Lifetime image from the luminescence collected during the AFM measurement of (a). ................................. 134
7.18 (a) An AFM, (b) intensity image, (c) optical image and (d) FLIM of a bulk GaAs double heterostructure. No surface structure was found in the AFM image. FLIM and intensity images show a trench with poor optical performance that spatially overlaps with the white line connecting the surface defects (highlighted with purple circles) of (c). Scan area of the AFM/FLIM is shown with a green box. Data in (b) and (d) was collected in NSOM illumination mode. 

7.19 Apparatus for fluorescence microscopy. The CCD image can be displayed at video refresh rates.

7.20 (a) Optical image with surface defects (highlighted in red circles). (b) Fluorescence image of a candidate GaAs laser cooling sample. The dark lines in the fluorescence image reveal degraded optical performance despite a predominantly featureless surface.

7.21 (a) Optical image of sample MH701. (b) Fluorescence image of the same area. The dark spot in the fluorescence image reveals degraded optical performance despite the uniform surface.

7.22 (a) GaAs and (b) GaInP fluorescence images of the same area. Green circle marks the position of a defect in GaAs, but not in GaInP. Surface scratches appear in GaInP fluorescence, but don’t affect the GaAs fluorescence image.

7.23 GaAs fluorescence images of the same dot taken (a) before and (b) after liftoff and bonding. Three stages of process induced defects are noted. (c) An optical image of the same area as (b) shows surface damage.
List of Figures

8.1 Instruments and communication pathways used in time-resolved simultaneous AFM and lifetime imaging. . . . . . . . . . . . . . . . . . . 149
8.2 Single photon counting AFM-NSOM apparatus. . . . . . . . . . . . . 150
8.3 Instruments response function for the TCSPC system. Laser diode driver triggered (a) internally and (b) externally with a DG535. . . . 153

B.1 (a) Measured fluorescence decay (black line) and best fit (yellow line). The gray window is the region over which the fitting is performed. (c) $\tilde{\chi}^2$ (blue line) and fitted lifetime (green line) over the extending gray shaded region in (a), optimum window width is $\approx$1300 points. (b) The $\tilde{\chi}^2$ at optimum window width as a function of signal-to-noise. (d) Lifetime of the fits calculated from (b). . . . . . . . . . . . . . . . 163

C.1 Semiconductor luminescence decay recorder in the near-field (black line). A fit under the nonradiative approximation (blue line) and radiative/nonradiative approximation (red dashed line). . . . . . . . 166
Glossary

FROG  Frequency resolved optical gating

IAC  2nd order Interferometric autocorrelation

ICE  Interferometric correlation envelope

LED  Light emitting diode

MIIPS  Multiphoton intrapulse interference phase scan

MOSAIC  Modified spectrum auto-interferometric correlation

PICASO  Phase and intensity from correlation and spectrum only

SHG  Second harmonic generation

SPIDER  Spectral phase interferometry for direct electric field reconstruction

GVD  Group velocity dispersion

TOD  Third order dispersion

AFM  Atomic Force Microscopy

FLIM  Fluorescence Lifetime Image

PL  Photoluminescence
**Glossary**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>GaAs</td>
<td>Gallium Arsenide</td>
</tr>
<tr>
<td>GaInP</td>
<td>Gallium Indium Phosphide</td>
</tr>
<tr>
<td>EQE</td>
<td>External Quantum Efficiency</td>
</tr>
<tr>
<td>DHS</td>
<td>Double Heterostructure</td>
</tr>
<tr>
<td>RE</td>
<td>Rare Earth</td>
</tr>
</tbody>
</table>
Part I

Ultrafast Optics
1.1 Overview

To begin, what is implied in the term "ultrafast"? Typically this refers to events on the femtosecond timescale, $1\text{ fs} = 10^{-15}\text{ s}$. To get a feel for the 15 orders of magnitude covered in the prefix "femto" it’s insightful to do a back of the envelope calculation to reveal that if the age of the universe is one second, one femtosecond would roughly correspond to the amount of time it takes you to read this introduction. Exploration and management of phenomena on this timescale comes through the application of ultrashort laser light pulses, packets of electromagnetic waves oscillating at optical frequencies. The production, control and measurement of such pulses is crucial to accessing phenomena such as atomic orbital period and molecular dynamics. The high optical intensity associated with ultrashort pulses allows for study of nonlinear processes. Consider a laser producing 100 fs pulses at a repetition rate of 100 MHz and an average power of 1 Watt. Each pulse has a peak power of 100 kW. By focusing the pulse to a spot size of 100 $\mu\text{m}^2$ a peak intensity of 100 Gigawatt/cm$^2$ can be achieved!
Chapter 1. Introduction to Ultrafast Optics

The measurement of pulses on the femtosecond timescale is challenging. Without an accurate measure of the pulse it’s impossible to know just what kind of pulse one is dealing with. What kinds of nonlinearities and peak intensities can it achieve? How will it distort as it propagates through a material? A detailed understanding of the pulse is required for full use. Detection of a pulse by direct electronic means is limited to a few picoseconds ($10^{-12}$ s) [1]. Streak cameras merge the generation of photoelectrons and their temporal resolution into one device, but even these only have response times on the order of a picosecond [2]. What remains is an all optical method of measurement or more practically a two step process where an ultrafast optical gate ensures sufficient bandwidth followed by an electronic sampling; converting photons to electrons at a reduced bandwidth.

Unlike electronic systems, optical systems have ample bandwidth. From Fourier analysis is known that more bandwidth supports shorter pulses. One laser material that provides a large bandwidth is a titanium doped sapphire (Ti:sapphire) crystal. Ti:sapphire has a bandwidth that spans over 400 nm from $\approx 600$ nm to $> 1$ micron [3]. This bandwidth can support laser pulses as short as 5 fs, corresponding to less than 2 optical cycles of the electric field. A cavity surrounding the Ti:sapphire crystal provides optical feedback associated with gain exceeding loss and lasing. For the generation of short pulses the energy contained in the cavity must be confined to as small an interval as possible. This can be achieved with an intracavity amplitude modulator [4], but if an acousto-optic or electro-optic modulator is used, bandwidth associated with the control electronics again limits the minimum pulse duration.

Issues associated with electronic bandwidth can be avoided by considering an all optical modulator. One such example of an optical modulator is a semiconductor saturable absorber [2]. This device is placed intracavity and absorbs photons. At sufficient intensities these materials become partially transparent. When this happens a power fluctuation within the cavity will experience lower loss and become
amplified, concentrating the cavity energy into a short period of time. Relaxation of the semiconductor material is not instantaneous however and as a consequence bandwidth limitations again restrict the minimum pulse duration.

One of the most successful techniques employed for the generation of near single cycle pulses is based on Kerr lens modelocking (KLM). A Kerr nonlinearity causes high intensities to be delayed with respect to low intensities. When the center and most intense spatial portion of the beam is phase retarded this corresponds to focusing, or the creation of a Kerr lens. An aperture around the center region of the beam can act as an affective absorber, providing lower loss to focused high intensity light. This longitudinal effect produces a red-shift on the leading part of the pulse and a blue-shift on the trailing part of the pulse and has been named self-phase modulation (SPM) [5]. An important property of SPM is that it also spectrally broadens the pulse, leading to additional bandwidth. Because no carrier dynamics are involved (in contrast to the semiconductor saturable absorber) response times of less than 1 fs can be achieved [2]. An example of KLM in tandem with a semiconductor saturable absorber mirror has been demonstrated by Sutter et al. [6]. In this geometry a Kerr nonlinearity is produced within the Ti:sapphire gain medium. The primary advantages of this configuration are ease of cavity alignment, increased average power and long term power stability compared with KLM alone.

A femtosecond pulse can be regarded as a collection of electromagnetic radiation with an electric field dependence in space and time. The evolution of these fields and their interactions with matter are described by Maxwell’s equations with a material response given by a macroscopic polarization. An electromagnetic light pulse is characterized by measurable quantities related to the electric field. A complex representation of the electric field is used to define the pulse. By neglecting the spatial dependence of the electric field, $\tilde{E}(r, t) = \tilde{E}(t)$, a full description can be given in either the frequency or time domain. Starting with a complex $\tilde{E}(t)$, one can define
Chapter 1. Introduction to Ultrafast Optics

a field in the frequency domain using the complex Fourier transform [7]

\[ \tilde{E}(\Omega) = F[\tilde{E}(t)] = \int \tilde{E}(t)e^{-i\Omega t}dt = |\tilde{E}(\Omega)|e^{i\Phi(\Omega)}. \] (1.1)

The term, \(|\tilde{E}(\Omega)|\), is the spectral amplitude, \(\Phi(\Omega)\) is the spectral phase and the variable \(\Omega\) denotes angular frequency. When \(\Phi(\Omega) = 0\), the pulse is designated as transformed limited. Similarly, the complex inverse Fourier transform can be applied to complex electric field in the frequency domain to arrive at the time domain pulse

\[ \tilde{E}(t) = F^{-1}[\tilde{E}(\Omega)] = \frac{1}{2\pi} \int \tilde{E}(\Omega)e^{i\Omega t}dt = |\tilde{E}(t)|e^{i[\omega_o t + \phi(t)]} \] (1.2)

where \(\omega_o\) denotes the center (or carrier) frequency of the pulse and \(\phi(t)\) is referred to as the pulse chirp. A pulse having \(\phi(t) = 0\) is classified as unchirped. As an example, the temporal dependence of the real part of the electric field for a chirped and unchirped pulse is shown in Fig. 1.1. The electric field amplitude, \(|\tilde{E}(t)|\), is chosen to be a Gaussian. For this figure the temporal pulse chirp is shown visually in two ways: 1) the frequency or color of the pulse is shown constant for the unchirped case Fig. 1.1(left) and evolving with time for the chirped case Fig. 1.1(right), 2) the period of oscillation is constant across the pulse, unchirped, and changing with time, chirped. The depicted chirped pulse is an example of linear upchirp, that is, its frequency increases with increasing time.

1.2 Pulse Propagation in Materials

For an understanding of a pulse propagating in matter more consideration is needed beyond the temporal and spectral characteristics. An electric field now taken with both spatial and temporal dependence is considered for the reduced wave equation

\[ \left[ \frac{\delta^2}{\delta z^2} + \frac{\Omega^2}{c^2}\epsilon(\Omega) \right] \tilde{E}(z,\Omega) = 0 \] (1.3)
where the assumption that the electric field is linearly polarized and propagating in the positive z-direction as a plane wave has been made. The term $\epsilon(\Omega)$ is the dielectric constant and $c$ is the speed of light in vacuum. The general solution to this is

$$\tilde{E}(z, \Omega) = \tilde{E}(0, \Omega) e^{-ik(\Omega)z} \quad (1.4)$$

with the propagation constant, $k(\Omega)$, is given by $k(\Omega) = \Omega n(\Omega)/c$ with $n$ being the index of refraction. It is useful to expand the propagation constant in a Taylor series about the carrier frequency, $\omega_o$

$$k(\Omega) = k_o + \frac{d k}{d \Omega} \bigg|_{\omega_o} (\Omega - \omega_o) + \frac{1}{2} \frac{d^2 k}{d \Omega^2} \bigg|_{\omega_o} (\Omega - \omega_o)^2 + \frac{1}{6} \frac{d^3 k}{d \Omega^3} \bigg|_{\omega_o} (\Omega - \omega_o)^3 + ... \quad (1.5)$$

and rewrite the general solution as

$$\tilde{E}(z, \Omega) = \tilde{E}(0, \Omega) e^{-ik_o z - i\delta k z} \quad (1.6)$$

where $k_o = 2\pi n(\lambda_o)/\lambda_o$ or $k_o = w_o n(\omega_o)/c$. The term $\frac{d^2 k}{d \Omega^2} \bigg|_{\omega_o}$ is referred to as group velocity dispersion (GVD) and $\frac{d^3 k}{d \Omega^3} \bigg|_{\omega_o}$ is third order dispersion (TOD). As a consequence of materials having a finite response time, a pulse propagating through
the material will acquire the phase term, $i\delta k z$. In the frequency domain, the non-
instantaneous response gives rise to a dispersion. The GVD for a few common glasses
is shown for reference in Fig. 1.2. Fused silica and BK7 are glasses frequently used as
prism materials in pulse compressor geometries due to their low dispersion. A more
dispersive glass, SF10, is shown for comparison. These curves are calculated from
the readily available Sellmeier equations found for example in Ref. [8] for fused silica
and Ref. [9] for BK7 and SF10. The increasing dispersion with increasing frequency
of these materials is known as positive GVD.

Simple physical insight can be used to produce an estimate of pulse broadening
from GVD. Suppose a pulse has a transform limited pulse duration of $\tau_{tl}$ and a
spectral bandwidth, $\Delta \omega_p$. The spectral components of the pulse each travel with
their own group velocity, $v_g(\Omega)$. The difference of the group velocities across the
Chapter 1. Introduction to Ultrafast Optics

pulse spectrum is

\[ \Delta v_g = \left. \frac{dv_g}{d\Omega} \right|_{\omega_p} \Delta \omega_p. \]  

\hspace{1cm} (1.7)

After a pulse travels a distance, \( L \), it can broaden by as much as

\[ \Delta \tau_{tl} = \Delta \left( \frac{L}{v_g} \right) \]  

\hspace{1cm} (1.8)

which after differentiation gives

\[ \Delta \tau_{tl} = L \left. \frac{d^2 k}{d\Omega^2} \right|_{\omega_p} \Delta \omega_p. \]  

\hspace{1cm} (1.9)

While a pulse my have a high bandwidth it may not be as short as dictated by
the Fourier transform. By assigning some typical number to the quantities in Eq.
1.9, one can get a feel for the temporal broadening of high bandwidth pulses in
the laboratory. Consider a pulse having a 100 nm bandwidth (transform limit \( \approx 
10 \text{ fs} \)) centered at 800 nm. The GVD for fused silica at 800 nm is 350 fs²/cm. If
this pulse passes through a 1 mm thick glass slide it will broaden by roughly 10 fs,
approximately doubling its transform limited duration.

1.3 Summary

The large amounts of broadening and phase accumulation encountered in propagating
high bandwidth ultrashort pulses motivates the need for careful characterization
beyond a simple measure of the spectrum.
Chapter 2

Pulse Characterization

2.1 Overview

This chapter introduces techniques used to measure ultrashort laser pulses and briefly describes a few of the more popular techniques. Modified spectrum autointerferometric correlation (MOSAIC) is a simple and sensitive technique for interpreting subtle pulse distortions. Its introduction, detailed description and experimental demonstration are presented here. Applications of MOSAIC are shown. The last section describes how to produce a MOSAIC trace from the second harmonic generation frequency resolved optical gating (SHG FROG) dataset.

2.2 Techniques

Detailed characterization of amplitude and phase of ultrashort laser pulses is vital to the controlled use of femtosecond laser systems [7, 10]. The diagnosis of such short pulses by means of direct electronic detection is limited by instrumental bandwidth.
Chapter 2. Pulse Characterization

To measure a short event in time requires an even shorter event in time, but what measures the shortest event? For femtosecond laser systems the pulse is used to measure itself. This process is called autocorrelation and is expressed as

\[ g(\tau) = \int f(t)f(t + \tau)dt \]  

(2.1)

where \( f(t) \) is the intensity of the ultrashort pulse and the limits of integration are for all \( t \). The autocorrelation trace is always symmetric and hence its Fourier transform is real. This property means the intensity autocorrelation can only be used to provide an estimate of the pulse profile. Furthermore, since the intensity autocorrelation contains no information about the phase present on the pulse, \( g(\tau) \) is not unique, meaning many different pulses can give the same intensity autocorrelation. Despite its shortcomings, the intensity autocorrelation still has value due to its ease of use and confirmation of whether a laser is producing femtosecond mode-locked pulses.

Optical sampling techniques based on nonlinear autocorrelation or cross correlation remain the most viable method for characterizing such short pulses. The first order or linear correlation contains spectral amplitude information, but provides no information on the phase of the ultrashort pulse. Nonlinear schemes such as second harmonic generation (SHG), two-photon fluorescence, two-photon conductivity and Kerr gating provide intensity autocorrelations that are routinely used to estimate the laser pulse width. However, ambiguities have been reported [11]. No chirp (or phase) information is gained unless an interferometric setup such as second-order interferometric auto-correlation (IAC) is used [10]. Furthermore, while IAC contains certain amplitude and phase information, it will not yield a full characterization of the amplitude and phase of the pulse electric field. A number of elegant techniques have been introduced that reconstruct the full electric field [12, 13, 14, 15, 16]. The variety of these techniques include measurements of frequency resolved intensity autocorrelation, frequency resolved interferograms, spectrograms, and sonograms. The
uniqueness of the retrieved spectral phase (or its ambiguity) typically varies with the degree of complexity in the implementation of these measurements. The most successful schemes such as frequency resolved optical gating (FROG) [12] and spectral phase interferometry for direct electric field reconstruction (SPIDER) [13] can completely retrieve the amplitude and phase with only a single laser shot, but require involved setups and multi-pixel detectors. The SPIDER technique is experimentally complicated. It is based off of spectral shearing interferometry and arrives at the phase in a non-iterative analytic manner [13]. A downside of the SPIDER technique is that the delay between the two replica pulses to be interfered with the broadened pulse must be maintained to very high precision. Mechanical stabilities on the order of nanometers and beam pointing stabilities of milliradians are required for 10% accuracy on pulse width [17]. The SHG FROG setup is very similar to a balanced intensity autocorrelation setup, with the exception that the nonlinear detector is replaced with an SHG crystal and a spectrometer. The recorded FROG dataset is two-dimensional, a map in time-frequency, and requires a computationally intense iterative algorithm to extract the electric field amplitude and phase. More details on FROG can be found at the end of this chapter. Simpler methods analogous to FROG are available that are computationally straightforward and thus applicable for real-time phase retrieval [18], but require a spectrally resolved two dimensional dataset and a well characterized reference pulse. Multiphoton intrapulse interference phase scan (MIIPS) is an adaptive pulse characterization technique that can retrieve the electric field and compensate for spectral phase to achieve transform limited pulses [19]. This technique requires a configurable amplitude/phase mask as well as a two-dimensional CCD. The preceeding techniques are all successful for reconstruction of the amplitude and phase of ultrashort pulses, but do so with either experimental complexity leading to higher cost or computational burden leading to longer times.
Chapter 2. Pulse Characterization

Simpler methods have been developed requiring only a single element detector and autocorrelator. One example of this is the phase and intensity from correlation and spectrum only (PICASO) [20] method. Figure 2.1 shows the experimental setup for SHG FROG, SPIDER and PICASO. Initially, the PICASO method retrieved the phase of the ultrashort pulse using only the intensity autocorrelation and pulse spectrum. Later it was found that retrieval based on this dataset could contain ambiguities [11]. According to Naganuma et. al. a unique retrieval is possible using a second order IAC and pulse spectrum [21]. A revised version of PICASO, however, was developed with the only nontrivial ambiguity being direction of time. The retrieval method makes use of a three envelope dataset obtained from linear and nonlinear, balanced or unbalanced autocorrelation traces [22]. The experimental setup for PICASO is shown in Fig. 2.1 for the three most common versions of PICASO: balanced correlation (IAC), unbalanced correlation and dual correlations. For unbalanced correlations a material of a known transfer function is placed in one of the arms of the Michelson interferometer. The transfer function can unbalanced the amplitude or phase of the pulse to be characterized. Similarly, in the dual correlation case, a material of a known transfer function is placed after the output of the interferometer, but before the detector. The detection nonlinearity is either second or third order. Third order correlations give less ambiguity in phase retrieval, but give less signal to noise than second order correlations. The pulse is retrieved by an iterative algorithm. An iterative, three envelope phase retrieval technique based on a population split genetic algorithm has recently shown promise for improved computational efficiency and accuracy [23]. Other pulse characterization techniques depict temporal asymmetry using unbalanced interferometric correlation envelope (ICE) functions [24]. The modified spectrum autointerferometric correlation (MOSAIC) is similar to PICASO, but has the advantages of time-domain signal averaging and straightforward visual chirp interpretation. The remaining focus of this chapter will be on the MOSAIC technique and its development. The SHG FROG and MOSAIC
Chapter 2. Pulse Characterization

Figure 2.1: Experimental setups for pulse characterization using PICASO, SHG FROG and SPIDER.

techniques are very similar and later in this chapter it will be shown how a MOSAIC can be generated from an SHG FROG trace, providing complimentary non-iterative pulse assessment.

2.3 The MOSAIC algorithm

In many applications, however, the full field retrieval may not be necessary, and only a semi-quantitative yet sensitive measure of the phase distortion is of interest. In the instance of adjusting intracavity prism separation, precise knowledge of the
electric field is not needed. A fast simple monitor of pulse chirp is all that is needed to optimally set the cavity prism configuration. Indeed, a graphical indication of pulse chirp is desired over a quantitative monitoring such as FROG, for ease of interpretation and real-time feedback. The MOSAIC algorithm detailed here achieves this by a very simple approach: an IAC trace is converted to a fringe free trace that provides a visual and unambiguous indication of the phase distortion (chirp) with very high sensitivity [25, 26]. The algorithm runs efficiently on a PC to allow adjusting and monitoring of mode-locked lasers in real-time.

The production of the second order IAC is done experimentally with an interferometer. An example of a widely used Michelson interferometer setup is shown in Fig. 2.2. A pulse (shown chirped) is incident on a 50/50 beam splitter. The first arm of the interferometer contains a pulse reflected off the beam splitter and retro-reflected off a corner cube, while the second arm of the interferometer contains a transmitted pulse that is retro-reflected off of a corner cube and delayed with respect to the first arm. The retro-reflected pulses recombine at the beam splitter and are directed to a nonlinear detection scheme. One of the most commonly used nonlinear detection schemes is SHG followed by a linear detector. With this setup a spectral filter is typically used to removed any residual fundamental laser light. An alternative method is two-photon absorption-induced photocurrent in a light-emitting diode [27]. The detected signal is recorded as a function of delay, \( \tau \), between the two arms.

The simple principle of generating a MOSAIC trace can be described in the frequency domain as follows: The laser pulse is assumed to have an electric field given by

\[
\tilde{E}(t) = \sqrt{f(t)}e^{j[\omega_0 t + \phi(t)]}
\]

where \( \phi(t) \) denotes the temporal chirp. The well known second order IAC trace produced by a second order autocorrelator is given by [7]:
Chapter 2. Pulse Characterization

Figure 2.2: A Michelson interferometer used to produce a second order interferogram. The pulse is interfered with itself as a function of delay, \( \tau \).

\[
S_{IAC}(\tau) = 1 + 2 \int f(t)f(t+\tau)dt + \int f(t)f(t+\tau)\cos(2\omega\tau + 2\Delta\phi)dt + 2 \int f^{1/2}(t)f^{3/2}(t+\tau)\cos(\omega\tau + \Delta\phi)dt + 2 \int f^{3/2}(t)f^{1/2}(t+\tau)\cos(\omega\tau + \Delta\phi)dt
\]

(2.2)

where \( \Delta\phi(t,\tau) = \phi(t + \tau) - \phi(t) \) and \( \int f(t)dt = 1 \). The limits of integration are from \(-\infty\) to \(\infty\). The spectrum (Fourier transform) of the above IAC contains three components at 0, \( \omega \) and 2\( \omega \), where \( \omega \) is the fringe frequency. The spectrum is then modified by retaining the dc term, removing the \( \omega \) term and amplifying the 2\( \omega \) component by 2 [28]. The inverse Fourier transform produces a fringe-resolved MOSAIC trace in the time domain [25].
Chapter 2. Pulse Characterization

\[ S_{MOSAIC}(\tau) = 1 + 2 \int f(t)f(t+\tau)dt \]
\[ + 2 \int f(t)f(t+\tau)\cos(2\omega\tau + 2\Delta\phi)dt \]  
(2.3)

this spectral modification eliminates the linear component at \( \omega \) and was originally proposed as a filter for Kerr-lens autocorrelation measurements [28]. Further analysis can be performed by expanding the cosine term in Eq. 2.3 to derive an expression for the fringe visibility [25]

\[ S_{MOSAIC}(\tau) = g(\tau) + |\tilde{g}_{2\omega}(\tau)|\cos[2\omega\tau + \Phi(t)] \]
\[ = g(\tau) + [g_s(\tau)^2 + g_c(\tau)^2]^{1/2}\cos[2\omega\tau + \Phi(t)] \]  
(2.4)

where

\[ g_s(\tau) = \int f(t)f(t+\tau)\sin[2\Delta\phi(t)]dt \]  
(2.5)

\[ g_c(\tau) = \int f(t)f(t+\tau)\cos[2\Delta\phi(t)]dt \]  
(2.6)

are the sine and cosine intensity autocorrelations and \( g(\tau) = \int f(t)f(t+\tau)dt \) is the intensity autocorrelation. The term

\[ \tilde{g}_{2\omega}(\tau) = \int f(t)f(t+\tau)e^{-2i\Delta\phi(t,\tau)}dt \]  
(2.7)

is the amplified \( 2\omega \) component of the IAC, which is also the envelope of the second harmonic field autocorrelation. The quantity, \( \Phi(t) = -\tan^{-1}(Im(\tilde{g}_{2\omega}(\tau)))/Re(\tilde{g}_{2\omega}(\tau)) \)
Chapter 2. Pulse Characterization

is the fringe phase. The fringe phase is not important to realize the sensitivity of the MOSAIC trace to chirp, however it will play a role in subsequent sections that deal with spectral phase retrieval.

A comparison of ordinary IAC and MOSAIC traces calculated for unchirped and linearly chirped pulses is illustrated in Fig. 2.3. An unchirped pulse is used to produce the IAC, Eq. 2.2, Fig. 2.3a, and fringe resolved MOSAIC, Eq. 2.3, Fig. 2.3b. The case of a linearly chirped pulse is shown in Fig. 2.3(c, d). It is important to note that while the IAC traces appear very similar, the MOSAIC traces display a sharp difference.

Figure 2.3: (a) an IAC calculated from and unchirped pulse. (b) MOSAIC trace for the IAC of (a), note the flat baseline. A chirped pulse used in (c) IAC and (d) MOSAIC calculation. The MOSAIC trace shows a flat baseline for unchirped pulses and shoulders for chirped pulses. The IAC traces look nearly the same.
Chapter 2. Pulse Characterization

2.4 Envelope MOSAIC

To appreciate the sensitivity of the MOSAIC trace it is not necessary to display the interferometric fringes. The upper and lower envelopes of the MOSAIC trace are given by [29]

\[ S_{MOSAIC}(\tau) = g(\tau) \pm |\tilde{g}_{2\omega}(\tau)| \]  \hspace{1cm} (2.8)

Note that the lower envelope,

\[ S_{min}(\tau) = g(\tau) - |\tilde{g}_{2\omega}(\tau)| \]  \hspace{1cm} (2.9)

exhibits a flat feature (i.e. equals zero for all \( \tau \)) when no chirp is present \((\Delta \phi = 0)\) and thus provides a sensitive and background-free signal indicative of pulse chirp. This can be understood by noting that when the pulse is unchirped the second harmonic field envelope is identical to the fundamental field intensity envelope, \(E^2(t) = I(t)\), indicating that the envelopes \(g(\tau)\) and \(|\tilde{g}_{2\omega}(\tau)|\) are equal and hence their difference, \(S_{min}(\tau)\), is zero. It is this property that makes the MOSAIC trace a sensitive visual measure of pulse chirp. The second harmonic field autocorrelation is related to the second harmonic spectrum by

\[ \tilde{g}_{2\omega}(\tau) = F^{-1}(|\tilde{E}^2(\omega)|^2) \]  \hspace{1cm} (2.10)

where the second-harmonic power spectrum is represented by \(|\tilde{E}^2(\omega)|^2\) and \(F^{-1}\) denotes the inverse Fourier transform operation.

Examples of two different chirp conditions in an experimental envelope MOSAIC trace is shown in Fig. 2.4(a, b) where the presence of shoulders on \(S_{min}\) indicates
Chapter 2. Pulse Characterization

Figure 2.4: MOSAIC signals generated from the same pulse and detected using (a) a BBO crystal and a linear detector (pulse duration: 60 fs FWHM) and (b) a two photon absorbing LED (76 fs FWHM). Insets shows fringe resolved MOSAIC. Measured IAC signals from which the MOSAIC traces in (a) and (b) were derived are shown in (c) and (d), respectively. The structure of the IAC waveforms appears almost identical, while the MOSAIC traces reveal chirp induced by the detection method.

chirp; the corresponding IAC traces appear indistinguishable, Fig. 2.4(c, d). The upper MOSAIC envelope is a measure of pulse duration but does not possess information about chirp information beyond $S_{\text{min}}$. The insets show the fringe resolved MOSAIC trace. We find it more useful to replace the upper envelope by the intensity autocorrelation $g(\tau)$ so that $S_{\text{max}}(\tau) = g(\tau)$ and $S_{\text{min}}(\tau) = g(\tau) - |\tilde{g}_{2\omega}(\tau)|$ represent amplitude and phase profiles respectively [26].

The fringe-free MOSAIC envelope $S_{\text{min}}$ is obtained from the IAC by the homodyne operation: $g_s(\tau) = 2 \langle S_{\text{IAC}}(\tau) \sin(2\omega \tau) \rangle_\Omega'$ and $g_c(\tau) = 2 \langle S_{\text{IAC}}(\tau) \cos(2\omega \tau) \rangle_\Omega'$.
additionally, \( g(\tau) = \langle S_{IAC}(\tau) \rangle'_{\Omega}/2 \) is the intensity autocorrelation, where the brackets, \( \langle \cdots \rangle \), indicate low pass filtering (time averaging) and the upper cutoff, \( \Omega' \), is chosen based on the bandwidth of the pulse [8]. In practice, we find that \( \frac{\omega mN}{N} \leq \Omega' < \omega/2 \) where \( N \) is the number of fringes in the FWHM of the IAC and \( 1 \leq m \leq 6 \) [8]. The fringe-free traces facilitate extensive time averaging to improve signal-to-noise ratio as will be shown in a later section.

2.5 Single Shot MOSAIC

Fringe-resolved autocorrelation (e.g. IAC) traces may not be practical for single shot measurements, which are often needed in low-repetition rate amplifier systems. A fringe-free method of MOSAIC based on an intensity autocorrelation and second harmonic spectrum can, however, be generated in a single shot. This procedure is called Envelope-MOSAIC (or E-MOSAIC) which provides the same information as first-generation MOSAIC generated from a fringe resolved IAC. An experimental demonstration of E-MOSAIC is as follows: A second-order IAC is produced from mode-locked 85 fs Ti:sapphire laser pulses centered at a wavelength of 825 nm using SHG. The SHG spectrum is collected by blocking the delay arm and routing the frequency doubled signal to an Ocean Optics HR4000 spectrometer with an 8 \( \mu m \) core diameter optical fiber. The inverse Fourier transform of the second harmonic power spectrum is computed over a spectral bandwidth consistent with the total time-delay of the intensity autocorrelation. The normalized time domain trace is then subtracted from the normalized intensity autocorrelation to produce, Fig. 2.5(a, open circles). This result is compared with \( S_{min} \) obtained from the second-order IAC using the same experimental setup, Fig. 2.5(a, solid lines). In Fig. 2.5b, the laser pulse is deliberately chirped before the autocorrelator by passing it through a 2 mm thick ZnSe window. The data in the insets compare the signals on a semi-log scale.
Figure 2.5: MOSAIC traces obtained from (a) an 85 fs Ti:sapphire laser and (b) intentionally chirping the laser pulse with 2 mm of ZnSe. MOSAIC rendered from a SHG spectral measurement (circles) and a second-order IAC (solid line). Insets show same data on a semi-log scale.

and validates the fidelity of E-MOSAIC.

### 2.6 Hybrid MOSAIC

Temporal chirp is easily identified with a MOSAIC trace by the appearance of shoulders on the minimum envelope. Fourier analysis shows that spectral phase distortion (dispersion) does not necessarily imply temporal phase distortion (chirp). A pulse having a symmetric spectrum and any odd order dispersion, for example, will exhibit a flat MOSAIC minimum (i.e. no temporal chirp) even though it is not transform limited. An asymmetric spectrum and flat spectral phase, in contrast, will exhibit a nonzero minimum envelope and bandwidth limited intensity autocorrelation. To deal with these situations, we introduce an enhanced implementation of MOSAIC that is uniquely sensitive to both spectral dispersion and temporal chirp. The added capability requires the pulse spectrum in addition to a second-order IAC. In rapid-
Chapter 2. Pulse Characterization

Figure 2.6: Linear and nonlinear autocorrelations from a Michelson interferometer.

scan IAC schemes, the spectrum can be obtained by splitting off a small amount of the autocorrelator output and directing it to a linear detector [8, 12 and 14], see Fig. 2.6.

The pulse spectrum, $|\tilde{E}(\Omega)|$, is found from the Fourier transform of the resulting linear interferogram (or a spectrometer in a noninterferometric or single shot arrangement). A transform limited pulse is created by assigning a flat phase to the measured pulse spectrum which is then used to compute the transform limited intensity autocorrelation, $g^{TL}(\tau)$. And the complex transform limited amplified $2\omega$ component of the IAC by employing the convolution integral:

$$\tilde{g}^{TL}_{2\omega}(\tau) = F^{-1}\left\{ \int |\tilde{E}(\omega')||\tilde{E}(\omega - \omega')|d\omega' \right\}$$

(2.11)

The phase of $\tilde{g}^{TL}_{2\omega}(\tau)$ or $\tilde{g}_{2\omega}(\tau)$ is not important for visual appreciation of pulse distortion at this stage, but may play a role for phase retrieval as will be discussed in the
Chapter 2. Pulse Characterization

phase retrieval section. The difference between the deduced transform limited autocorrelation and measured intensity autocorrelation, \( g(\tau) \), results in a background free trace displaying a symmetric double-hump for broadened pulses and flat baseline for transform limited pulses. We call this implementation hybrid MOSAIC (H-MOSAIC) [29].

The autocorrelation is symmetric about zero delay, so we use the average of both halves for the difference computation, \( g(\tau) - g^{TL}(\tau) \) and the measured lower envelope of MOSAIC. We define the lower envelope of H-MOSAIC as

\[
S_{HYB}(\tau) = \begin{cases} 
  g(\tau) - g^{TL}(\tau) & : \tau < 0 \\
  g(\tau) - |\tilde{g}_{2\omega}(\tau)| & : \tau \geq 0.
\end{cases}
\] (2.12)

The lower envelope of H-MOSAIC is sensitive to spectral dispersion for \( \tau < 0 \) and temporal chirp for \( \tau > 0 \). It is important to note the non-orthogonality of the H-MOSAIC peaks: dispersion can give rise to a temporal chirp and vice versa. In Fig. 2.7 we present H-MOSAIC simulations for the special cases of a pulse having a symmetric spectrum with (a) no dispersion, (b) group velocity dispersion (GVD) only and (c) third order dispersion (TOD) only. For these select cases there is an added ambiguity in the intensity autocorrelation, \( g(\tau) \), caused by the sign-independence of even and odd orders of spectral phase [22]. This point will be revisited in chapter 3.

The pulse distinguished by the trace in (a) can be regarded as the ”perfect pulse”, which means it is symmetric in time and frequency, and without chirp or dispersion. Any small deviation from this symmetry will appear as shoulders in the H-MOSAIC traces. Note that in case (c) when only TOD is present we have \( g(\tau) - |\tilde{g}_{2\omega}(\tau)| = 0 \).

Fig. 2.7d shows an H-MOSAIC produced from an asymmetric spectrum and flat spectral phase. This case is transform-limited, which leads to \( S_{HYB}(\tau) = 0 \) for \( \tau < 0 \). Due to the asymmetry in the pulse spectrum, temporal phase is produced upon Fourier transformation of the pulse to the time domain. The presence of this
Figure 2.7: a) Simulation of H-MOSAIC for a pulse having a symmetric spectrum with (a) no dispersion, (b) GVD, (c) TOD and (d) an asymmetric spectrum with no dispersion. The dashed line indicates zero delay.

time domain phase is seen in the shoulder of the lower envelope of the H-MOSAIC trace in Fig. 2.7d for \( \tau > 0 \). It is important to note there is no new information in H-MOSAIC not already present in a MOSAIC or IAC trace and the spectrum. The power of H-MOSAIC is the ability to clearly extract this information and display it in an easily interpreted graphical representation.

2.7 Averaging MOSAIC Traces

The generation of the MOSAIC trace from the IAC trace is done by spectral filtering. The location of the filter in the Fourier domain is based on the fringe frequency of
Chapter 2. Pulse Characterization

the IAC. We obtain the fringe frequency $\omega$ in a time window at the peak of the IAC. When the signal-to-noise ratio approaches unity, accurate measurement of $\omega$ is no longer possible. The fundamental fringe frequency $\omega$ can, however, be determined with high accuracy even if the second-order autocorrelation signal is weak and noisy. This determination is accomplished by directing a small amount of the interferometer output to a linear detector to generate a linear interferogram from which $\omega$ can be obtained. This independent measurement of $\omega$ allows the processing of very noisy IAC signals in the MOSAIC algorithm by averaging Eq. 2.9:

$$\{S_{\text{min}}(\tau)\}_{\text{ave}} = \bar{g}(\tau) - \eta|\bar{g}_2\omega(\tau)|$$  \hspace{1cm} (2.13)

where $\eta$ is a distortion correction factor that is generated by the algorithm. The IAC may be distorted as a result of detector bandwidth limitations, autocorrelator misalignment and/or nonquadratic detector response leading to $S_{\text{min}} \neq 0$ at $\tau = 0$. As the ideal condition of $S_{\text{min}} = 0$ can only be approached in practice, it becomes necessary to force $S_{\text{min}}$ to zero at zero delay. This correction takes the form

$$\eta = \frac{g(0) - S_{\text{min}}(0)}{|\bar{g}_2\omega(0)|}$$  \hspace{1cm} (2.14)

and is needed to render a correct MOSAIC, separating pulse chirp information from distortion associated with autocorrelator misalignment and noise in the detection electronics.

The utility of the corrective factor $\eta$ is displayed in an experimental example of autocorrelator misalignment distortion, see Fig. 2.8. We use a mode-locked Ti:sapphire laser (see chapter 4) producing 120 fs pulses at a center wavelength of 820 nm. An aligned IAC signal is shown in Fig. 2.8(a and c) while Fig.2.8e depicts a signal from a deliberately misaligned autocorrelator. The misalignment was produced by a vertically tilting one arm of the interferometer. The spatial fringes across the detector
Chapter 2. Pulse Characterization

mimic a restrictive electronic bandwidth. Corresponding MOSAIC waveforms are displayed in Figs. 2.8(b, d and f). The IAC traces before and after the misalignment appear identical. The MOSAIC signal in Fig. 2.8f, however, reveals misalignment by a visible departure of $S_{\text{min}}$ from the baseline at zero delay. It is important to eliminate such deviations for correct chirp characterization. Real-time calculation of $\eta$ is performed using Eq. 2.14; the corrected MOSAIC shown in Fig. 2.8d (triangles) reproduces the undistorted shoulder height to within the measurement noise. MOSAIC can produce a correct shoulder height for non-quadratic detectors with $I^{2+\epsilon}$, $-0.2 \leq \epsilon \leq 0.2$ for chirped pulses producing a shoulder height of 13% or less with an error of 2%. Equivalently, MOSAIC can determine the order of detection to within 7% for $-0.3 \leq \epsilon \leq 1.0$ using unchirped pulses [26].

The MOSAIC generation process with a very noisy IAC is demonstrated in Fig. 2.9. The IAC waveform is barely visible above the noise in Fig. 2.9a. After a 1200 shot average and implementation of the algorithm in Eq. 2.13, a clear MOSAIC waveform can be resolved in Fig. 2.9b (red triangles). This MOSAIC trace reproduces the chirp extracted from a noise-free IAC of the same pulse (solid lines in Fig. 2.9b). The signal to noise in the IAC is adjusted by making use of a polarizing beam splitter and half-waveplate prior to the autocorrelator. Adjusting the waveplate scales the pulse power while maintaining dispersive elements in the beam path. The MOSAIC signal averaging approach is very useful for characterizing ultrashort pulses deep in the ultraviolet or far-infrared where the production of second-order IAC signals is often plagued by distortion and noise. Averaging also allows for efficient, in-situ pulse monitoring since only a relatively small amount of power needs to be diverted to the autocorrelator. It is important to emphasize that it is difficult to produce an averaged IAC with chirp sensitivity because of fringe jitter.

The shoulders of an averaged MOSAIC trace have structure and temporal separation that are sensitive measures of the order of the chirp. In Fig. 2.10, we fit
章 2. 脉冲特性化

图 2.8: 实验获得的 IAC 轨迹：a) 无啁啾脉冲，c) 喇叭脉冲和 e) 喇叭脉冲与失真。失真是由移位分光计产生的。相应于 MOSAIC 最大和最小包络线的 MOSAIC 包络线显示在 b)，d) 和 f)。校正的最小包络线的 MOSAIC 在 d) 中以三角形表示。

图 2.8: 实验获得的 IAC 轨迹：a) 无啁啾脉冲，c) 喇叭脉冲和 e) 喇叭脉冲与失真。失真是由移位分光计产生的。相应于 MOSAIC 最大和最小包络线的 MOSAIC 包络线显示在 b)，d) 和 f)。校正的最小包络线的 MOSAIC 在 d) 中以三角形表示。

脉冲包络在图 2.9b 假设一个激光脉冲的电场 \( \tilde{E}(t) = \sqrt{f(t)} e^{i[\omega_0 t + \phi(t)]} \) 与啁啾 \( \phi(t) = a(t/t_p)^2 + b(t/t_p)^3 + c(t/t_p)^4 \) 和脉冲持续时间 \( t_p \)。脉冲包络，\( f(t) \) 被假设为一个 sech 形状。系数a, b, 和 c 调整以适合最小 \( S_{min} \) 中的折边以及肩峰高度。图 2.10 中的拟合计算使用 \( a = 0.18 \pm 0.01, |b| = 0.2 \pm 0.02 \) 和 \( c = -0.082 \pm 0.002 \)。肩峰结构表明存在较高阶的啁啾（即 \( b \neq 0 \) 和/或 \( c \neq 0 \)）。

脉冲包络在图 2.9b 假设一个激光脉冲的电场 \( \tilde{E}(t) = \sqrt{f(t)} e^{i[\omega_0 t + \phi(t)]} \) 与啁啾 \( \phi(t) = a(t/t_p)^2 + b(t/t_p)^3 + c(t/t_p)^4 \) 和脉冲持续时间 \( t_p \)。脉冲包络，\( f(t) \) 被假设为一个 sech 形状。系数a, b, 和 c 调整以适合最小 \( S_{min} \) 中的折边以及肩峰高度。图 2.10 中的拟合计算使用 \( a = 0.18 \pm 0.01, |b| = 0.2 \pm 0.02 \) 和 \( c = -0.082 \pm 0.002 \)。肩峰结构表明存在较高阶的啁啾（即 \( b \neq 0 \) 和/或 \( c \neq 0 \)）。
Figure 2.9: a) Single IAC trace just above the noise level. b) Averaged MOSAIC waveform produced from 1200 noisy IAC traces (red triangles); the chirp of the same pulse obtained with negligible noise is reproduced (solid line).

In the case of fitting the MOSAIC trace of Fig. 2.10 *a priori* knowledge was required of the pulse (i.e. that it had a $sech$ envelope in the time domain). Later it will be shown that when the pulse spectrum is known in addition to the MOSAIC trace a unique full field reconstruction is possible.

2.8 Applications

2.8.1 Ultraviolet Autocorrelation

In this section, MOSAIC pulse characterization is performance in high noise conditions where stand-alone IAC is not possible. The characterization is demonstrated by applying the MOSAIC averaging technique to the low signal-to-noise case of autocorrelated frequency doubled Ti:sapphire at $\lambda = 415$ nm. The doubled pulses are autocorrelated with an ultraviolet Michelson interferometer and a suitably cut BBO
Chapter 2. Pulse Characterization

Figure 2.10: Experimental data (red triangles) and fit (green lines) that reveal high-order components of chirp on a 120 fs Ti:sapphire laser pulse.

crystal to produce 208 nm light that is detected with a photo-multiplier tube (Fig. 2.11a. A small amount of the autocorrelator output is directed to a linear detector from which the fringe frequency $\omega$ is determined. Both pulse duration and chirp are difficult to evaluate with the IAC due to low signal-to-noise. The MOSAIC algorithm overcomes this obstacle with signal averaging. In Fig. 2.11b, we show an averaged MOSAIC waveform generated from 1000 low signal IAC traces. Significant pulse chirp is seen in the shoulders of the averaged MOSAIC trace. It is important to note that this chirp information is inaccessible without MOSAIC.
Chapter 2. Pulse Characterization

Figure 2.11: a) Single IAC trace just above the noise level of a frequency doubled Ti:sapphire laser pulse ($\lambda = 415$ nm). b) Averaged MOSAIC waveform produced from acquiring 1000 noisy IAC traces.

### 2.8.2 LED and BBO Comparison

The advantages of chirp sensitivity in MOSAIC make possible sensitive detector characterization. Reid *et. al.* have shown that an unbiased LED exhibits a nonlinear power dependence that can be used to generate the second order IAC [27]. MOSAIC reveals spectral phase distortion introduced by common two-photon photovoltaic detectors. By changing the nonlinear detection scheme on the output of the Michelson interferometer it is possible to directly compare detected signals from both LEDs (via two photon photoconductivity) and SHG followed by a linear detector [30].

Using the setup shown in Fig. 2.2 we produce an IAC and associated MOSAIC trace from four different LED manufacturers. All IAC traces are generated under the same conditions with the LEDs being used in the unbiased (photoconductive) mode. The Ti:sapphire laser (described in chapter 4) produces nominally 60 fs pulses at a center wavelength of 815 nm. The average power incident on the autocorrelator is greater than 500 mW. Each MOSAIC trace is averaged 100 times to ensure a noise free comparison. Chirp response in MOSAIC from the different LED manufactures
Chapter 2. Pulse Characterization

is displayed in Fig. 2.12. All LEDs have a band gap energy greater than the photon energy of the laser pulse. The advertised color of the LED emission is specified in the figure. A comparison of the pulse detected with SHG using a BBO crystal is also shown. The entry labeled "BBO start" refers to chirp response using SHG in BBO taken prior to all LED measurements, while the label "BBO finish" refers to an autocorrelation measurement taken after all LED measurements. The fact that the start and finish BBO measurements are the same represents a control in the experiment and indicates that the pulsed laser was stable during data collection. Of the LED manufacturers tested (Xicon, Kingbright, Fairchild, and Chicago Miniature) all showed significantly more chirp in MOSAIC than SHG from BBO. This increased chirp response is due to pulse propagation in the LED structure prior to two-photon absorption.

The study of the LEDs also quantified the affect of the epoxy dome lens of the LED on pulse chirping. For all measurements shown in Fig. 2.12 the epoxy dome was removed except for cases denoted with a "*". In these cases a with and without dome measurement was made. It can be seen from the data in the figure that for the three instances when the dome was removed a negligible difference in the amount of chirp was recorded. For the Kingbright LED the chirp response increased slightly upon dome removal. This increase is most likely due to a different positioning of the LED with respect to the focused laser after removal of the dome. It was noted that removing of the dome increased the signal to noise on the IAC. It is therefore desirable to remove the dome to achieve the highest fidelity data and ease in the alignment of the focused laser onto the LED active area. It is important to emphasize that the effect of the dome is negligible for our 60 fs pulses; shorter pulses may experience a measurable amount of pulse broadening due to their increased bandwidth. A substantial variation in pulse chirp was found among LEDs from a given manufacturer. For example, two different red LEDs from Chicago Miniature revealed a factor of two difference in pulse chirp. These two LEDs are the same
Figure 2.12: Chirp response in MOSAIC originating from LEDs that differ in manufacturer and color. The lowest chirp measurement was obtained through SHG in a BBO crystal.

color, have the same manufacturer part number and came out of the same bag. At most, a factor of 10 increase in MOSAIC chirp response is measured in LEDs. Pulse broadening as much as 15% is found using LEDs compared to BBO.

2.8.3 Metal-Semiconductor-Metal Detector

The planar metal-semiconductor-metal (MSM) structure can be used as a two-photon detector in second order autocorrelations; we examine a ZnSe MSM device for induced pulse chirp [31]. The structure is a single crystal ZnSe substrate with interleaved titanium electrodes and a gold cap layer. Titanium provides high adhesion to the ZnSe [32]. A $3.5\times$ microscope objective focuses the autocorrelator output on the
Chapter 2. Pulse Characterization

region between the electrodes (bias: 30 V), see Fig. 2.13 [30].

A 500 average MOSAIC is obtained for different lens positions; shoulder height (chirp) is plotted in Fig. 2.14 Left. Error bars are due to slight asymmetry in MOSAIC traces. SNR of the second order IAC is shown in Fig. 2.14 Right. Lowest chirp occurs with the focusing geometry that produces the highest SNR, but this alignment optimization must be made with caution. Changing the lens position shifts the region of two-photon absorption from the surface to deeper in the bulk ZnSe. Significant chirp is introduced as the interaction length in the material increases, while the SNR decreases by < 10%.

2.9 MOSAIC and Frequency Resolved Optical Gating (FROG)

One of the most common and commercially available techniques for pulse characterization is based on FROG. Performing pulse characterization using FROG consists of two parts: 1) an experimental apparatus, in which a pulse is split and mixed in a nonlinear optical medium. The mixed signal output is recorded temporally and
Figure 2.14: (Left) Chirp response in MOSAIC. (Right) signal-to-noise of the second order IAC as an MSM detector is brought through focus.

spectrally as a function of delay between the two pulses, 2) and iterative phase retrieval algorithm that retrieves the intensity and phase [33]. The FROG measurement and retrieval algorithm have been shown to uniquely and quantitatively reconstruct the complex electric field of ultrashort pulses [34]. A FROG measurement is performed in the time-frequency domain rather than just the time domain as in the case of autocorrelation. Measurements of both time and frequency are performed simultaneously. A mathematically rigorous expression of the FROG spectrogram (time-frequency dataset) is

\[
S_{FROG}(\omega, \tau) = |\int E(t)G(t - \tau)e^{i\omega\tau} dt|^2 \tag{2.15}
\]

where \(G(t)\) is the gate function. Ideally, the gate function would be known and a straight forward deconvolution would be possible. However, for ultrashort pulses no such known gate function is accessible. As in the case of autocorrelation, the pulse (or some variation of it) is used to measure itself. The variation of the pulse gate function used determines the type of FROG trace measured. Two examples of the
Chapter 2. Pulse Characterization

gate function are a polarization gate (PG) and the pulse itself

\[
G(t) = \begin{cases} 
E(t - \tau) & : \text{SHG FROG} \\
|E(t - \tau)|^2 & : \text{PG FROG}.
\end{cases}
\] (2.16)

When the pulse itself is used as the gate function SHG FROG is the result. In the case of SHG FROG the nonlinear optical medium relies on a \(\chi^{(2)}\) effect and the resulting data set is sometimes referred to as a "spectrally resolved autocorrelation" [35]. The mathematical constraint imposed on any FROG dataset comes from the known form of the nonlinear optical gate function and is used in the reconstruction algorithm. The extracted intensity and phase of the pulse are obtained using an iterative algorithm based on generalized projections [36]. Depending on the complexity of the pulse the time to reconstruct the FROG trace can vary.

Ideally, the SHG FROG trace would give direct insight on the phase of the pulse from a visual inspection. In such a case no time consuming iterative retrieval would be needed. Some thought has been given to visually interpreting the SHG FROG trace, see for example Ref. [33]. In this work it was noted that a horseshoe shape about zero delay was characteristic of excessive third order dispersion. However, the observation of the horseshoe shape is problematic in that it is not a background free estimation. That is, there is a significant amount of additional structure in the FROG trace that does not resemble a horseshoe shape. Additionally, a spectrally normalized SHG FROG trace has been developed [37]. All spectral slices within the FROG trace are normalized to a peak value of unity. In this rendering, a pulse with an arbitrary spectrum and flat spectral phase would give a streak of uniform thickness about zero delay. This trace is again subjective. The confirmation of a uniform thickness for near transform limited pulses is visually difficult. Subtle departures from the transform limit are likely to go unnoticed as the trace is not background free. Unfortunately, SHG FROG traces are largely regarded as unintuitive due to their symmetry along the delay axis [33, 34, 36, 38, 37].
Chapter 2. Pulse Characterization

The SHG FROG trace can be made more intuitive by augmenting it with MOSAIC. The MOSAIC trace can be obtained directly from the SHG FROG trace. No additional experimental hardware beyond that necessary to measure the FROG trace is needed. The rendering of MOSAIC from SHG FROG facilitates direct, immediate insight on pulse chirp. No iterative reconstruction is required and hence the production of the MOSAIC trace is real-time in all cases; independent of pulse complexity. The SHG FROG trace and the generalized projection algorithm do a good job at quantitative pulse reconstruction, while MOSAIC performs well as a qualitative background free tool for visualizing pulse chirp. The merging of these two techniques provides complimentary pulse assessment.

The straightforward generation of MOSAIC from an SHG FROG trace is understood by considering the functional form of the FROG spectrogram

\[ S_{\text{FROG}}^{\text{SHG}}(\omega, \tau) = \left| \int E(t)E(t-\tau)e^{i\omega \tau} \, dt \right|^2 \]  \hspace{1cm} (2.17)

where the gate function is the pulse itself. Note the slice at \( \tau = 0 \) is simply

\[ S_{\text{FROG}}^{\text{SHG}}(\omega, \tau = 0) = \left| \int E(t)E(t)e^{i\omega \tau} \, dt \right|^2 \]  \hspace{1cm} (2.18)

which is also the second harmonic power spectrum, \( |\tilde{E}^2(\omega)|^2 \). From Eq. 2.10 it is known that the inverse Fourier transform of \( |\tilde{E}^2(\omega)|^2 \) yields \( g_{2\omega}(\tau) \):

\[ g_{2\omega}(\tau) = F^{-1}[S_{\text{FROG}}^{\text{SHG}}(\omega, \tau = 0)]. \]  \hspace{1cm} (2.19)

Since the SHG FROG trace is a spectrally resolved intensity autocorrelation, the 1-dimensional intensity autocorrelation, \( g(\tau) \), can easily be obtained by spectrally integrating the SHG FROG dataset. The lower envelope of the MOSAIC trace can then be computed

\[ S_{\text{min}}(\tau) = g(\tau) - |F^{-1}[S_{\text{FROG}}^{\text{SHG}}(\omega, \tau = 0)]| \]  \hspace{1cm} (2.20)
Chapter 2. Pulse Characterization

while the upper trace is (as before) \( g(\tau) \).

As was shown in Section 2.4, the MOSAIC trace is visually far more sensitive to pulse chirp than an IAC. In an analogous way, MOSAIC is more sensitive to displaying chirp than an SHG FROG trace. This enhanced sensitivity is due to the background free nature of the MOSAIC trace; any departure from an unchirped pulse results in a deflection from zero in the lower MOSAIC envelope. The visual power of MOSAIC from SHG FROG is demonstrated in Fig. 2.15.

The transform limited pulse used for computation of the SHG FROG trace shown in Fig. 2.15a has a pulse width of \( \approx 10 \) fs and a symmetric spectrum. The case of Fig. 2.15b is generated from the same spectral amplitude but having \( 10 \) fs\(^2\) of spectral dispersion. The MOSAIC trace of Fig. 2.15c is determined from the SHG FROG trace of Fig. 2.15a and immediately confirms that the pulse is unchirped without the need for an iterative retrieval. The sensitivity of the MOSAIC trace is apparent from Fig. 2.15d. While the SHG FROG trace of Fig. 2.15b shows little difference from its unchirped counterpart, the MOSAIC trace illustrates a dramatic difference. The chirped MOSAIC trace is computed from the SHG FROG dataset and easily displays pulse chirp embedded within the FROG trace.

In practice, the quality of the retrieved amplitude and phase from FROG is judged based on a 2-dimensional RMS difference between the measured and reconstructed FROG traces. The functional form of the rms difference to be minimized in a FROG based reconstruction is given by [33]

\[
G_{\text{rms}} = \left\{ \frac{1}{N^2} \sum_{\omega, \tau=1}^{N} \left[ S_{\text{Meas}}^{\text{FROG}}(\omega, \tau) - S_{\text{Recon}}^{\text{FROG}}(\omega, \tau) \right]^2 \right\}^{1/2} \tag{2.21}
\]

where \( N^2 \) is the number of points in the FROG dataset. The measured and reconstructed FROG traces are normalized to a peak value of unity before \( G_{\text{rms}} \) is
Figure 2.15: Calculated SHG FROG traces from an unchirped (a) and chirped (b) pulse. Corresponding MOSAIC traces produced from the SHG FROG traces, unchirped (c) and chirped (d). It is important to note the FROG traces appear identical; complimentary MOSAIC traces reveal pulse chirp without the need for iterative reconstruction.

computed. For the example of the two distinct pulses shown in Fig. 2.15, the RMS difference between the two SHG FROG traces is $G_{rms} = 0.0043$. Alternatively, the MOSAIC traces from the two pulses of Fig. 2.15 have an RMS difference of 0.0246. The sensitivity to the difference in phase between the two pulse of Fig. 2.15 is nearly 6x greater in MOSAIC than SHG FROG. The computation of the
Chapter 2. Pulse Characterization

1-dimensional RMS error for MOSAIC is based on

\[
\delta = \frac{1}{2} \left\{ \frac{1}{N} \sum_{\tau=1}^{N} \left[ g^{\text{Meas}}(\tau) - g^{\text{Recon}}(\tau) \right]^2 \right\}^{1/2}
+ \left\{ \frac{1}{N} \sum_{\tau=1}^{N} \left[ S^{\text{Meas}}_{\text{min}}(\tau) - S^{\text{Recon}}_{\text{min}}(\tau) \right]^2 \right\}^{1/2}.
\]

A critically important point is published values for pulses retrieved using SHG FROG have exceeded \( G_{\text{rms}} = 0.01 \) [39]. Such large errors in reconstructed FROG traces suggest structure is being missed in the electric field to be retrieved. The production of a MOSAIC trace on both the measured and reconstructed FROG trace is an additional way to judge the quality of a retrieval effort. It is important to note this can easily be implemented on pulses already retrieved because no additional information or measurements are needed beyond the measured and reconstructed FROG traces.

2.10 Summary

In this chapter the MOSAIC algorithm was presented. Three methods for generating a MOSAIC trace were developed from measurements of the IAC, SHG spectrum and intensity autocorrelation, and SHG FROG dataset. Averaged MOSAIC traces were exhibited high sensitivity to chirp with a signal to noise level of \( \approx 1 \). The high signal to noise found on averaged MOSAIC envelopes was applied to chirp characterization of frequency doubled Ti:sapphire laser pulses. Pulse chirp response in MOSAIC was found to increase substantially \( (2\% \rightarrow 17\%) \) as an MSM detector was brought through a focus. Lowest chirp was shown to coincide with highest signal to noise. Pulse distortion arising from the detection method was characterized using MOSAIC.
Chapter 2. Pulse Characterization

A factor of 10 increase in MOSAIC shoulder height and a nearly 15% increase in pulse width was measured using LEDs compared to SHG with BBO. The method of rendering a hybrid MOSAIC trace was introduced and provided distinction between pulse chirp and spectral dispersion.
Chapter 3

Spectral Phase Retrieval from MOSAIC

3.1 Overview

This dissertation now turns to the problem of retrieving the phase of ultrashort laser pulses. While the strength of MOSAIC lies in its visual exhibition of pulse chirp, more information is available. The MOSAIC trace in combination with the pulse spectrum allow a unique retrieval of the full electric field. The only ambiguities are absolute phase, pi phase slips, arrival time and most notably the direction of time. Retrieval of synthetic and experimental pulses is shown. A sequential retrieval algorithm is presented and discussed. Different optimization routines are investigated for computational efficiency and accuracy.
3.2 Retrieval MOSAIC: Spectral Phase Reconstruction

Naganuma et. al. have shown that a combination of pulse spectrum and IAC is sufficient to uniquely reconstruct the complex electric field with only a time direction ambiguity [21]. The SNR required to uniquely reconstruct the phase, however, may not be experimentally practical [11]. By combining MOSAIC data with the first-order interferogram and performing additional analysis, the spectral phase of the electric field can be recovered [40]. A number of synthesized pulses of varying complexity were shown to be reconstructed successfully using an iterative method. Averaging of the MOSAIC dataset allows for resolution of the ambiguities discussed in ref. [11], as shown by ref. [41]. The MOSAIC technique has recently been used to reconstruct ultrashort pulses in the mid-IR [42], because MOSAIC is an algorithm, it can be performed in any spectral region or experimental condition where an IAC or intensity autocorrelation and second harmonic spectrum can be measured.

The usefulness of MOSAIC has been run with a fringe resolved trace, as well as, in an envelope rendering with homodyne detection and signal averaging [26]. High fidelity traces are extracted using fringe-free averaging technique in a high noise environment where the signal to noise ratio approaches unity. Different pulses can therefore be distinguished even when producing essentially identical IAC traces [11]. Additionally, the MOSAIC envelope is not distorted by intensity imbalance in the autocorrelator and any residual linear absorption that may be present in two photon absorbing detectors, unlike an IAC.

We demonstrate MOSAIC for the retrieval of the spectral phase and hence a full field reconstruction of an ultrashort pulse. The speed and simplicity of this 1-dimensional algorithm comes at the expense of time-direction ambiguity, but such precise knowledge is not necessary for many applications. Unique quantitative anal-
Chapter 3. *Spectral Phase Retrieval from MOSAIC*

Analysis gives information on the electric field, amplitude and phase in the presence of extreme noise [26, 41]. Algorithm structure and efficiency is discussed. A comparison to IAC based phase retrieval will also be presented.

As was shown in the previous chapter, the H-MOSAIC trace is a convenient visual representation of the data set used in 1-dimensional retrieval schemes. Experimental reconstruction using MOSAIC and the pulse spectrum has been demonstrated using an iterative line minimization technique [40]. In this reconstruction method, all points in the spectral phase are optimized individually at the expense of processing time. Processing time can be reduced about 7x by analyzing phase with a fourth-order Taylor-series expansion and adjusting the coefficients using the Retrieval (R)-MOSAIC algorithm [29]. The R-MOSAIC algorithm is implemented by sequential optimization of the Taylor-series coefficients and works as follows: Group velocity dispersion is optimized while third order dispersion and fourth order dispersion (FOD) are ignored. If convergence is not realized with GVD alone, the retrieved GVD coefficient seeds a second stage of optimization that includes TOD. A new GVD coefficient is generated along with a TOD coefficient. If satisfactory convergence is still not obtained, a third stage is performed to give a new GVD, TOD and FOD coefficients. This sequence accounts for the predominant contribution of the lowest order spectral phase coefficients encountered in realistic pulses. Our simulations have shown that simultaneous optimization of GVD, TOD and FOD is less likely to converge on the optimal phase coefficients. Simultaneous optimization can weight higher order terms too heavily causing the algorithm to fall into a local well far from the correct solution in the search space. This was noted as a poorly reconstructed MOSAIC with a higher root mean square (RMS) error compared to the sequential routine. The structure of the iterative sequential algorithm is shown in Fig. 3.1. The efficiency of different search algorithms will be discussed in section 3.5.

The sequential R-MOSAIC algorithm is run until a minimum error is found. An
iterative search algorithm minimizes the RMS error, $\Delta$, between the measured and reconstructed MosaIC traces. The function to be minimized is given by:

\[
\Delta = \left\{ \frac{1}{2N} \left[ \sum_{\tau=1}^{N} \left[ g_{\text{Meas}}(\tau) - g_{\text{Recon}}(\tau) \right]^2 \right] + \sum_{\tau=1}^{N} \left[ S_{\text{Meas}}^{\text{min}}(\tau) - S_{\text{Recon}}^{\text{min}}(\tau) \right]^2 \right\}^{1/2}
\]  

(3.1)

where $N$ is the number of points used in the reconstruction, $g_{\text{Meas}}(\tau), S_{\text{Meas}}^{\text{min}}(\tau)$ and $g_{\text{Recon}}(\tau), S_{\text{Recon}}^{\text{min}}(\tau)$ represent measured and computed quantities, respectively. This error signal contains equal weighting for both the upper and lower envelopes of MosaIC. Minimization of the RMS defines convergence of the algorithm.
Chapter 3. Spectral Phase Retrieval from MOSAIC

3.2.1 Sample Case: GVD

To demonstrate the capability of R-MOSAIC, a few synthetic pulses are modeled. The model consists of an ultrashort pulse with a spectrum described by the hyperbolic secant function and phase expressed as a 4th order polynomial:

\[
\tilde{E}(\Omega) = \frac{2}{e^{a_1\Omega/\Delta\omega} + e^{-a_2\Omega/\Delta\omega} e^{i[\phi_{GVD}(\Omega-\omega_0)^2 + \phi_{TOD}(\Omega-\omega_0)^3 + \phi_{FOD}(\Omega-\omega_0)^4]}}
\]  

(3.2)

where \(a_1\) and \(a_2\) are asymmetry parameters initially set equal to one. The spectrum is centered at \(\omega_0 = 0\) and has a transform limited duration of \(\approx 10\) fs. In the first pulse modeled, a spectral phase is assigned having \(\phi_{GVD} = 25\) fs\(^2\) and \(\phi_{TOD} = \phi_{FOD} = 0\). This pulse represents a frequently encountered situation in the laboratory; pulse distortion is due to GVD originating from propagation in a material. In this case we neglect any TOD that may be acquired from propagation in a material (subsequent test cases will include this contribution). The simulated ultrashort pulse defined in the frequency domain is Fourier transformed to the time domain where it is used to compute the upper and lower envelopes of a MOSAIC trace, Eq. 2.8. This computed trace is the target MOSAIC to be reconstructed. Results of the retrieval effort are illustrated in Fig. 3.2.

Fig. 3.2a shows the target (solid lines) and reconstructed (dots) MOSAIC for a pulse possessing only GVD. Successful reconstruction to the target MOSAIC is performed using a MATLAB optimization routine based on a pattern search algorithm [43]. In this test case the reconstruction took less than half a second to recover the target GVD coefficient (performed on an IBM ThinkPad 1.8 GHz Intel Centrino Duo processor). The initial starting point for the spectral phase is zero (i.e., a flat spectral phase or transform limited). The retrieved and target pulses are plotted in the frequency domain across the given pulse spectrum, Fig. 3.2b, and time domain Fig. 3.2c. Solid lines represent target values and dots are the retrieval results. It
Chapter 3. Spectral Phase Retrieval from MOSAIC

Simulated Pulse Case 1:
\[ \phi_{GVD} = 25 \text{ fs}^2 \]

Figure 3.2: Simulation of MOSAIC based retrieval with a pulse have a symmetric spectrum and \( \phi_{GVD} = 25 \text{ fs}^2 \). a) MOSAIC trace b) pulse in the frequency and c) time domain. Solid lines indicate target and dots reconstruction.

is important to realize that the difference between the target and retrieved spectral phase is a trivial ambiguity and arises from the symmetric nature of autocorrelation [21]. This ambiguity is often referred to as "direction of time" [22, 21, 40]. Mathematically, this ambiguity is expressed as \( \tilde{E}(\Omega) = \tilde{E}^*(\Omega) \) or \( E(t) = E^*(-t) \). Hence, MOSAIC (and autocorrelation in general) is insensitive to the absolute sign of the spectral phase. It is important to note that retrievals based on SHG FROG [33] or grating-eliminated no-nonsense observation of ultrafast incident laser light e-fields (GRENOUILLE) [44] also have the direction of time ambiguity. The MOSAIC trace is, however, sensitive to the relative sign between GVD and TOD coefficients, this will be discussed in Section 3.3. For the test case in Fig. 3.2 the target \( \phi_{GVD} \) was
Chapter 3. Spectral Phase Retrieval from MOSAIC

25 fs$^2$ and the retrieved was -25 fs$^2$. In experiment, this ambiguity can be removed by unbalancing the autocorrelator (cross-correlation) [22], or making an additional measurement with a known dispersive element prior to the pulse entering the autocorrelator [2].

3.2.2 Sample Case: TOD

The second case considered for simulation is a pulse described by Eq. 3.2 where $\phi_{GVD} = \phi_{FOD} = 0$ and $\phi_{TOD} = 50$ fs$^3$. Such a pulse could be encountered in the laboratory by compensation of the GVD (using a prism/grating compressor [45] or chirped mirrors [46]) leaving only a residual TOD component to the spectral phase. The target and successfully reconstructed MOSAIC trace can be seen in Fig. 3.3a. In this case the MOSAIC indicates an unchirped pulse (flat baseline), which is confirmed in noting that the time domain phase is flat. This pulse is broadened beyond its transform limit, but this information is only conveyed in the upper envelope of the MOSAIC trace, $g(\tau)$. Spectral phase recovery and retrieved time domain amplitude and phase are shown in Fig. 3.3 (b) and (c) respectively. Once again the recovered spectral phase demonstrates the direction of time ambiguity, target: $\phi_{TOD} = 50$ fs$^3$, retrieved: $\phi_{TOD} = -50$ fs$^3$. The algorithm took 1.2 seconds and 110 iterations to converge on the target.

3.2.3 Sample Case: GVD and TOD

The third case modeled in simulation is that of a pulse containing both GVD and TOD coefficients and having a mathematical description given by Eq. 3.2. This scenario is representative of a pulse propagating through a dispersive media that imparts a non-negligible amount of TOD. Pulses containing second and third order dispersion require two stages of optimization in the R-MOSAIC algorithm. An ex-
Simulated Pulse Case 2:
\[ \phi_{TOD} = 50 \text{ fs}^3 \]

Figure 3.3: Simulation of MOSAIC based retrieval with a pulse have a symmetric spectrum and \( \phi_{TOD} = 50 \text{ fs}^3 \). a) MOSAIC trace b) pulse in the frequency and c) time domain. Solid lines indicate target and dots reconstruction.

ample pulse is created with \( \phi_{GVD} = 25 \text{ fs}^2 \) and \( \phi_{TOD} = 50 \text{ fs}^3 \). After 88 iterations of the pattern search algorithm and 1.1 s convergence on the target was obtained. Results are depicted in Fig. 3.4. The MOSAIC trace reconstructed in Fig. 3.4a shows a significant amount of pulse chirp as a result of the spectral phase recovered in Fig. 3.4b. The correctly retrieved time domain pulse shows not only amplitude asymmetry, but considerable temporal phase, Fig. 3.4c. In this case, the spectral dispersion coefficients were retrieved with the correct sign. However, \( \tilde{E}^*(\Omega) \) is also a solution.
Chapter 3. Spectral Phase Retrieval from MOSAIC

Simulated Pulse Case 3:

\[ \phi_{GVD} = 25 \text{ fs}^2 \text{ and } \phi_{TOD} = 50 \text{ fs}^3 \]

Figure 3.4: Simulation of MOSAIC based retrieval with a pulse have a symmetric spectrum, \( \phi_{GVD} = 25 \text{ fs}^2 \) and \( \phi_{TOD} = 50 \text{ fs}^3 \). a) MOSAIC trace b) pulse in the frequency and c) time domain. Solid lines indicate target and dots reconstruction.

3.2.4 Sample Case: Effect of Noise

The effect of noise on the R-MOSAIC algorithm is considered. Noise on the measured quantities used in the reconstruction is created by adding random noise to the pulse spectrum, \( \tilde{E}(\Omega) \). In this case a random number generator is used producing a Gaussian distribution whose FWHM is 3% the peak spectral amplitude. The target phase is assigned to the noisey spectrum, \( g(\tau) \) and \( S_{\text{min}}(\tau) \) are then computed. The averaging algorithm detailed in section 2.7 is used to average 100 MOSAIC traces from the 100 spectra containing random additive noise. Results presented in Fig. 3.5 are for a pulse have a symmetric \textit{sech} spectrum and \( \phi_{GVD} = 25 \text{ fs}^2 \) and \( \phi_{TOD} = 50 \text{ fs}^3 \).
Simulated Pulse Case 4:
\[ \phi_{GVD} = 25 \text{ fs}^2 \text{ and } \phi_{TOD} = 50 \text{ fs}^3 \]
3\% noise

The RMS error for the reconstructed MOSAIC trace was 0.0018. The retrieved dispersive coefficients in the simulation were \( \phi_{GVD} = 24.96 \text{ fs}^2 \) and \( \phi_{TOD} = 49.74 \text{ fs}^3 \) and took 1.2 seconds to recover. These retrieved coefficients show better accuracy when compared to the result from the unaveraged MOSAIC trace of \( \phi_{GVD} = 23.96 \text{ fs}^2 \), \( \phi_{TOD} = -43.01 \text{ fs}^3 \) and \( \phi_{FOD} = -4.78 \text{ fs}^4 \). It is important to note the good agreement between the target and reconstructed MOSAIC traces in Fig. 3.5c spans nearly two decades on the log scale and is made possible by signal averaging.
Chapter 3. Spectral Phase Retrieval from MOSAIC

3.3 Error Mapping

Spectral phase retrieval using the three envelope dataset outlined by Naganuma et al. can be improved with the application of the MOSAIC algorithm. Here we show that preprocessing of the IAC to a MOSAIC trace leads to better localization of the error minimal in the parameter space explored by iterative retrieval algorithms. To demonstrate this a simulation is performed with an ultrashort pulse having a symmetric spectrum given by $|\tilde{E}(\Omega)| = \exp[-\Omega^2/(\Delta\omega)^2]$. The spectrum is centered at 800 nm and has a FWHM $\approx 57$ nm. A spectral phase is assigned to the pulse having $\phi_{GVD} = 50$ fs$^2$ and $\phi_{TOD} = 175$ fs$^3$. From this pulse a target IAC, Eq. 2.2, and a fringe resolved MOSAIC, Eq. 2.3, are computed. Trial reconstruction pulses are produced by taking the target spectral amplitude and assigning a trial spectral phase for all values of $\phi_{GVD}$ between -100 and 100 fs$^2$ and $\phi_{TOD}$ between -600 and 600 fs$^3$. To visualize ambiguities in the retrieval we use these trial pulses to produce an error map in a manner similar to ref. [22]. An error map as a function of $\phi_{GVD}$ and $\phi_{TOD}$ is produced by computing the RMS error between the target IAC and the trial IAC traces, Fig. 3.6a. Regions of dark blue indicate low error. The corresponding error map for the target fringe resolved MOSAIC and trial fringe resolved MOSAIC trace is shown in Fig. 3.6b, [30]. It is important to note that the region of low error surrounding the solution shrinks in the case of the fringe resolved MOSAIC. This highly localized solution suggests faster convergence and better accuracy for iterative phase retrieval schemes. In the case of an intensity imbalanced autocorrelator or residual linear absorption on the detector the IAC error map would be affected. Because the $\omega$ term is removed and it is background free, MOSAIC error maps are insensitive to these distortions.

We further compute the error map for the case of the E-MOSAIC, Eq. 2.8. Results are presented in Fig. 3.6c, [30]. Here it can be seen that regions of low error become broader and less localized relative to the fringe resolved case. In addition,
there is also fourfold degeneracy for the region of lowest error. This degeneracy results in ambiguity on the sign of the $\phi_{GVD}$ and $\phi_{TOD}$ coefficients. The ambiguity and poor localization is due to neglecting the fringe phase, $\Phi(\tau)$, from Eq. 2.3. For visual interpretation of the E-MOSAIC trace, $\Phi(\tau)$ is not needed, however, for phase retrieval its inclusion is important particularly when the pulse spectrum is symmetric.

Experimentally, $\Phi(\tau)$ can be signal averaged in the same way as a MOSAIC
envelope. As a function of delay the fringe resolved MOSAIC trace given by Eq. 2.3 cannot be directly averaged due to fringe jitter. The question becomes how to extract the $\Phi(\tau)$ term for use in spectral phase reconstruction. The convolution theorem states that the product of two functions in the time domain is equivalent to the convolution of those two functions in the frequency domain. This means SHG expressed as product in the time domain can be regarded in the frequency domain as

$$\tilde{E}(2\omega) = \int \tilde{E}(\omega)\tilde{E}(\omega' - \omega)d\omega$$

(3.3)

where $\tilde{E}(2\Omega)$ is the second harmonic spectrum and $\omega'$ is a dummy variable. Since the fundamental pulse spectrum is measured (through linear autocorrelation) we can compute the transform limited second harmonic spectrum

$$\tilde{E}^{TL}(2\omega) = \int |\tilde{E}(\omega)||\tilde{E}(\omega' - \omega)|d\omega$$

(3.4)

and subsequently the transform limited $g_{2\omega}^{TL}(\tau)$:

$$g_{2\omega}^{TL}(\tau) = F^{-1}[|\tilde{E}(2\omega)|^2] = |g_{2\omega}(\tau)|e^{i[2\omega\tau + \Phi^{TL}(\tau)]}$$

(3.5)

where $F^{-1}$ indicates the inverse Fourier transform. From the IAC we obtain the term

$$g_{2\omega}(\tau) = |g_{2\omega}(\tau)|e^{i[2\omega\tau + \Phi(\tau)]}.$$  

(3.6)

It can be difficult to separate the $2\omega\tau$ term from the fringe phase due to uncertainties in $\omega$ determination. However, uncertainties present on $g_{2\omega}^{TL}(\tau)$ are also present on $g_{2\omega}(\tau)$ since they are both produced from the same spectrum. By taking a division of these two quantities a background free measure of the fringe phase is available

$$\frac{g_{2\omega}(\tau)}{g_{2\omega}^{TL}(\tau)} = \frac{|g_{2\omega}(\tau)|}{|g_{2\omega}^{TL}(\tau)|}e^{i[\Phi(\tau) - \Phi^{TL}(\tau)]}.$$  

(3.7)
The differential fringe phase (DFP), $\Phi(\tau) - \Phi_{TL}(\tau)$, is no longer coupled with $\omega$ if a single interferometer is used. It has the additional benefit of being insensitive to spectral asymmetries; dependence is removed in the same manner as $2\omega\tau$. Thus DFP is background free and only sensitive to spectral phase. The ambiguity on the sign on the $\phi_{GVD}$ and $\phi_{TOD}$ coefficients in E-MOSAIC is depicted in Fig. 3.7a. The example pulses consist of spectral phase with both GVD and TOD components of varying sign applied across a symmetric spectrum. In all four cases of sign on the spectral components the MOSAIC trace is the same (two are shown). The DFP is different depending on whether GVD and TOD have the same or opposite sign, see Fig. 3.7b. The DFP can be used to remove sign ambiguities associated with phase retrieval on a symmetric spectrum. It cannot resolve -GVD/-TOD vs. +GVD/+TOD or -GVD/+TOD vs. +GVD/-TOD, hence DFP is not sensitive to the absolute sign.

The value of the DFP can become large far from zero delay. To accentuate the relevant features of the DFP we weight it to the amplitude of $g(\tau)$. The normalized
Chapter 3. Spectral Phase Retrieval from MOSAIC

Figure 3.8: Normalized DFP signals showing sensitivity to relative sign on GVD and TOD dispersion coefficients.

DFP is then [30]

$$\delta_{DFP}(\tau) = g(\tau)[\Phi(\tau) - \Phi^{TL}(\tau)].$$  \hspace{1cm} (3.8)

If both GVD and TOD have the same sign, the normalized DFP will show a peak followed by a valley, Fig. 3.8(blue line), while an opposite sign between the coefficients is seen as a valley followed by a peak, Fig. 3.8(red line).

The inclusion of the normalized DFP to E-MOSAIC is shown in the error map of Fig. 3.6d. The error function in this case includes equal weighting for the upper and lower envelopes of MOSAIC as well as the normalized DFP. Spurious solutions are eliminated and highly localized solutions are restored, indicating DFP is an important contribution to E-MOSAIC based phase retrieval for pulses having a symmetric spectrum.

To mimic the experimental determination of the normalized DFP we put 2% additive white noise on a symmetric spectrum possessing only +GVD and generate
Chapter 3. Spectral Phase Retrieval from MOSAIC

Figure 3.9: 100 noisy normalized DFP traces for 2% additive white noise to pulse spectrum (red lines). Averaged DFP (blue dots) and target DFP (black line) are also shown.

the DFP. The results of 100 normalized DFP computations are shown in Fig. 3.9(red lines). Since the production of a MOSAIC envelope in the lab is accompanied by a normalized DFP, we can signal average the DFP just as we do a MOSAIC envelope, see Fig. 3.9(blue dots). The averaged DFP reproduces the target DFP, Fig. 3.9(black line), after 100 shots.

3.4 Experimental

We use the sequential optimization technique to reconstruct the electric field from the measured pulse spectrum and H-MOSAIC of Fig. 3.10a. An iterative simplex algorithm minimizes the RMS error. The measured spectrum and retrieved spectral phase from R-MOSAIC are shown in Fig. 3.10b, phase is the dashed line. Execution of the sequential reconstruction algorithm using 128 points is accomplished on an IBM laptop computer in less than 1 second with a 1.8 GHz Centrino Duo processor.
Figure 3.10: a) An experimental H-MOSAIC (solid line) and reconstructed H-MOSAIC (circles) from the phase (red dots) and measured spectrum of (b).

for $\Delta = 0.0072$. This minimum achievable RMS error is set by experimental noise and higher order phase terms not accounted for in the Taylor-series phase expansion.

Further reduction of RMS error can be obtained by individual optimization of each point in the spectral phase using a line search method similar to Ref. [40]. Our line-search algorithm is seeded with the sequential reconstruction result and run until a new minimum achievable RMS is found. The retrieved phase and reconstructed H-MOSAIC trace are depicted in Fig. 3.10(b, dots) and Fig. 3.10(a, circles), respectively. Processing time increases to 30 seconds giving a reduction of RMS error from 0.0072 to 0.0038. We find this accuracy improvement to be of little practical benefit.

For pulses not having a Taylor-series expandable phase it becomes necessary to use an individual point line search method to adequately reconstruct MOSAIC traces. Such algorithms can be seeded with the output of the resulting spectral phase from R-MOSAIC to allow for more rapid convergence. An experimental example of a phase recovered from individual point line search is shown across a measured spectrum in
Figure 3.11: (a) Measured second-order IAC (b) Experimental MOSAIC (pink lines) and reconstructed MOSAIC (dots) from the measured spectrum and retrieved phase of (c). Time domain pulse (d).

Fig. 3.11c. Simultaneous recording of both linear and second-order interferometric autocorrelation traces is done on a two-channel digital oscilloscope controlled with National Instruments LabVIEW software. The associated IAC and 100x averaged MOSAIC trace can be seen in Fig. 3.11(a, b) with the retrieved time domain pulse displayed in Fig. 3.11d. While a low error was achieved in the pulse retrieval, we show in the next section that ambiguities in the phase retrieval error map require additional information to be used to correctly identify the spectral phase.

To demonstrate the experimental relevance of the DFP we consider the retrieved
pulse from Fig. 3.11. The highly symmetric spectrum allows for the ambiguity displayed in Fig. 3.6c. Identically reconstructed MOSAIC traces having the same RMS error were produced with different retrieved pulses by changing the sign of the starting point in the iterative line search algorithm. The retrieved pulses from the different starting points are shown in Fig. 3.12(a, b) [30]. The red line is a polynomial fit across the FWHM of the pulse spectrum to the retrieved spectral phase. The fit in Fig. 3.12a is \( \Phi_1(\Omega) = -466f s^2 \Omega^2 - 11700 f s^3 \Omega^3 \) while the fit in Fig. 3.12b is \( \Phi_2(\Omega) = -1000f s^2 \Omega^2 + 11900 f s^3 \Omega^3 \). The sign difference between the GVD and TOD coefficients in the two fits is a nontrivial ambiguity that can be resolved with the DFP. The DFP for each reconstructed pulse is shown in the insets. The correlation of the peaks and valleys of the reconstructed DFP with the measured DFP indicates correct sign for the pulse in Fig. 3.12a. Similarly, the anti-correlation of the peaks and valleys of the DFP shown in Fig. 3.12b indicate that it is not the correct pulse; despite the fact that its reconstructed E-MOSAIC is the same.

Note that the advantage of the E-MOSAIC trace is that it can be averaged for
Chapter 3. Spectral Phase Retrieval from MOSAIC

pulse reconstruction at very low pulse energies. Additionally, fewer points can be used in the reconstruction (provided all necessary temporal features are adequately sampled) since the fringe structure does not have to be reproduced. However, if sufficient pulse energy is available, the fringe-resolved MOSAIC trace can be used for full pulse reconstruction, and issues of ambiguities in the E-MOSAIC error map or inclusion of the DFP can be avoided. In either fringe MOSAIC or E-MOSAIC with DFP a spectral reconstruction error map shows higher localization of the solution when compared to IAC.

3.5 Algorithm Efficiency

Several optimization algorithms have been investigated for effectiveness in phase retrieval using the R-MOSAIC procedure outlined above. Performance is defined by the minimum achievable RMS error and processing time. We tested line search, simplex, Levenberg-Marquardt, genetic and pattern search algorithms. Each of these algorithms was unmodified from the standard code available in the MATLAB software platform. Analysis of mode-locked Ti:sapphire laser pulses are presented in Fig. 3.13 [30].

The best phase reconstruction is obtained with either a simplex or line search routine. The genetic and pattern search algorithms also produce satisfactory fits, but do so at the expense of computation time.

3.6 Summary

The current chapter provided a background for phase retrieval based on a three envelope dataset (MOSAIC plus pulse spectrum). A real-time method for full electric
field reconstruction was developed using a sequential algorithm. For pulses having a Taylor expandable phase, the sequential algorithm reduced processing by 7x relative to existing techniques. Examples of the retrieved synthetic pulses having varying amounts of GVD and TOD were presented. A simulated example with measurement noise was also presented. Retrieval error maps displaying ambiguities in IAC, fringe MOSAIC, envelope MOSAIC and envelope MOSAIC with DFP were shown. MOSAIC based retrieval error maps exhibited steeper gradients and higher localization of solutions compared to IAC. Giving better accuracy and faster convergence for optimization algorithms exploring the parameter space. The DFP was developed to remove ambiguities associated with the envelope MOSAIC reconstruction of pulses having symmetric spectra. An experimental example was presented of a retrieved pulse having GVD and TOD. The pulse was retrieved in < 1 second. A more complicated pulse possessing a solution degeneracy found in the error map

Figure 3.13: Normalized algorithm performance for spectral phase retrieval on our 60 fs Ti:sapphire laser pulses. All algorithms are evaluated with the MATLAB software platform.
Chapter 3. Spectral Phase Retrieval from MOSAIC

investigation was measured and successfully retrieved using E-MOSAIC and DFP. Different optimization algorithms were tested for their accuracy and computational efficiency.

The MOSAIC algorithm is very useful in circumstances where real-time chirp feedback is desired. The field of microscopy is an example. Pulse broadening due to propagation through dispersive elements such as microscope objectives and optical fibers can reduce pulse peak power and ultimately result in lowered signal to noise. Pre-chirping with a pulse compressor followed by characterization using MOSAIC (either visual or field reconstruction) can diminish pulse distortion. Chapter 7 describes some of the microscopy techniques that can benefit from MOSAIC when ultrashort laser pulses are used.
Chapter 4

Ti:sapphire Laser System

4.1 Overview

This chapter outlines the Ti:sapphire laser used to produce $\approx 60$ fs pulses. The layout of the oscillator cavity is shown with critical dimensions labeled. The second section details the autocorrelator used to make measurements reported elsewhere in this dissertation.

4.2 Ti:sapphire Laser

For femtosecond laser pulse characterization a mode-locked (ML) Ti:sapphire laser is constructed. A broadband oscillator capable of supporting sufficient bandwidths for 50-60 fs laser pulses is used in the cavity configuration of Fig. 4.1. A population inversion in the Ti$^{3+}$ dopant atoms is created with a diode pumped solid state (DPSS) intra-cavity frequency doubled continuous wave (CW) laser. For our system a neodymium doped vanadate (Nd:YVO) laser from Coherent Inc. (Verdi V5)
produces a maximum 5 Watts of output power at a wavelength of 532 nm. The polarization of the pump laser is s-polarized. Prior to focusing the pump laser into the Ti:sapphire crystal a periscope is used to rotate to p-polarization and adjust the pump beam height. A plano-convex lens with a 120 mm focal length is used for focusing into the Ti:sapphire. The lens provides mode matching between the pump and Ti:sapphire laser beams. Mirrors M4 and M5 collect and collimate fluorescence from the 7.75 mm Ti:sapphire crystal. Ti:sapphire laser beam path is shown in red. The mirrors M4 and M4 have a 10 cm radius of curvature. All cavity mirrors, M1-M5, have a low dispersion highly reflective dielectric coating. Mirror M1 is designed for 45 degree angle of incidence (AOI) while M2 and M3 are 0 degree AOI. The prisms P1 and P2 provide intra-cavity dispersion compensation. The majority of the GVD in the cavity comes from the Ti:sapphire crystal with a small amount being added from the mirror coatings and propagation through the prism glass. Fused silica is chosen for the prism material due to its relatively low third order dispersion. The prisms are cut at Brewster’s angle and are set at minimum deviation during the alignment process. The tip to tip prism separation distance is approximately 54 cm. A slit across mirror M2 is used for wavelength selectivity. Laser light incident on mirror M2 is spectrally dispersed across the mirror. A spatial translation of the slit across M2 selects a spectral mode and hence allows wavelength tunability. The ultrafast output coupler has a reflectivity of approximately 88% across the bandwidth of the pulse. The laser output is p-polarized due to the Brewster angle cut on the crystal. The Ti:sapphire crystal is mounted in an aluminum holder. Thermal grease provides high thermal conductivity between the aluminum holder and the crystal. The assembly is water cooled to 17 °C.

A photograph of the assembled Ti:sapphire laser on a breadboard is shown in Fig. 4.2. Out of frame is the Verdi pump laser. The cavity optics are enclosed in box to minimize air currents that adversely affect modelocking stability. It is important to note that a tightly constructed box reduces the frequency of required cleanings needed
Chapter 4. Ti:sapphire Laser System

Figure 4.1: Ti:sapphire cavity producing 60 fs pulses. Critical distances are shown.

to maintain optimum laser performance. A fiber optic routed to an Ocean Optics spectrometer is used to monitor the modelocked spectrum by picking up residual laser light leaked through mirror M1. Immediately after the output coupler a glass slide reflects a small amount of power into a fast photo diode (Thorlabs DET200) for verification and observation of the pulse train. The repetition rate of the laser is measured to be 86 MHz. The average output power is just over half a Watt and gives a pulse energy of 6 nJ.
Figure 4.2: Photo of the pulsed Ti:sapphire laser cavity. The pump laser path is shown in green, Ti:sapphire cavity is shown in red. Fused silica prisms are used for intracavity dispersion compensation.

4.3 Autocorrelator

Pulses emitted from the Ti:sapphire laser are characterized by autocorrelation using the Michelson interferometer sketched in Fig. 4.3. Upon entering the interferometer the pulse is split using a 50% reflecting 50% transmitting at 800 nm beamsplitter. The pulse in arm 1 retro-reflects off of a stationary gold coated corner cube (Newport Corp.), but displaced a few millimeters with respect to the incoming beam. The pulse in arm 2 is retro-reflected off of a gold coated corner cube mounted on an acoustical subwoofer (Computer Acoustics). The subwoofer is driven sinusodially at
a frequency of approximately 2 Hz with a Lodestar function generator model FG-2100A. The spatial extent of the pulses coming out of the laser is on the order of tens of microns, much less that the range of displacement of the delay stage. The magnitude of the displacement of the corner cube delay is a few millimeters, ensuring sufficient delay for complete overlap during the autocorrelation process. The response of the subwoofer was investigated using the Ti:sapphire laser operating in the CW mode. CW laser operation is produced by creating a dispersion imbalance in the oscillator cavity. The imbalance is accomplished by inserting an excess amount of glass from prism P1 into the beam path of Fig. 4.1. Interference fringes at the output of the interferometer as a function of subwoofer position (delay) are recorded using a linear detector (Thorlabs DET110) and an oscilloscope (Tektronix TDS 520). It is observed that the period of oscillation of the fringes is constant over a broad range (ps) of delay. Fringe period distortion is seen only near the edge of the delay range where the speaker is slowing down. The response of the subwoofer is found to be linear in a range much larger than the range over which fs autocorrelation takes place. The retro-reflected beam off the delayed corner cube is displaced a few millimeters with respect to the incoming beam, see Fig. 4.3. This displacement is to prevent unwanted feedback to the oscillator which can disrupt modelocking. Triggering of the oscilloscope is done through a synchronization signal provided by the function generator. The trigger signal is delayed with a digital delay stage (BNC model 7010), and allows for precise overlap of the produced and recorded interferogram.

The two outputs of the autocorrelator are used for simultaneous linear and nonlinear detection. Linear detection is performed by directing the retro-reflected autocorrelator output to a linear detector connected to a digital oscilloscope through a variable terminator (Thorlabs VT1). The variable terminator ensures there is sufficient electronic bandwidth to fully resolve the interferometric fringes produced during a delay scan. Nonlinear detection is produced by making use of the orthogonal autocorrelator output. The output is focused using a microscope objective (Bausch
Chapter 4. Ti:sapphire Laser System

Figure 4.3: Michelson interferometer with component manufacturers listed.

and Lomb 3.5x). In the case of SHG a nonlinear crystal is used (BBO, KDP or BiBO), cut at an angle suitable for phase matching at 800 nm. A filter (Schott FG3) is used to remove any residual fundamental laser light. Linear detection is then done on the SHG light only. For the case of two-photon photoconductive detectors (e.g. LED’s), fundamental laser light is focused directly onto the detector active area. The communication of the detector to the oscilloscope is again done through a variable terminator. The amount of termination is determined by gradually reducing the termination resistance until the detected IAC (or MOSAIC) shows no visible distortion. With the above mentioned equipment and settings, a termination of 1 kΩ is typical. The oscilloscope record length is set at 15k points. This corresponds to between 30 and 40 points per fringe in the IAC. A lower sampling can be used, however, care must be taken to ensure adequate sampling of all relevant features.
Chapter 4. Ti:sapphire Laser System

The calibration of the time (delay) axis from the autocorrelation comes from an independent spectral measurement of the pulse wavelength. The measurement originates from the fiber optic, seen in Fig. 4.2. Collected laser light is routed to an Ocean Optics HR2000 spectrometer. The peak wavelength of the spectral bandwidth is chosen for axis calibration. The laser spot is scanned across the surface of the fiber to ensure collection from the central most portion of the beam. However, spatial chirp is found to be negligible, only a few nanometers across the beam waist in both lateral directions. One period of oscillation in the IAC corresponds to a translation of one wavelength. If the wavelength is for example $\lambda = 800$ nm and there are $N = 30$ points per fringe, then the $\Delta t$ between two consecutive points in the IAC is $\Delta t = \lambda / (cN) = 0.09$ fs/point where the speed of light, $c$, is $3 \times 10^8$ m/s.
Chapter 5

Summary and Future Work:
Ultrafast Optics

Ultrashort laser pulse characterization using MOSAIC has been investigated for both visual chirp interpretation and full electric field reconstruction. The high visual sensitivity of MOSAIC makes it unique among pulse characterization schemes. The presence of chirp is easily identified without the need for a time consuming iterative retrieval algorithm. However, when more quantitative information about the pulse is desired, MOSAIC has been shown to perform real-time phase retrieval using a sequential optimization algorithm.

A method for obtaining high signal to noise MOSIAC traces was presented. This method uses an averaging technique and was experimentally demonstrated to yield a high fidelity MOSAIC trace in the presence of extreme noise (SNR ≈ 1). The high signal to noise found on averaged envelope MOSAIC traces served as a sensitive measure for electric field reconstruction. A combination of pulse spectrum and averaged MOSAIC envelopes reconstructed a pulse in real-time using a single autocorrelator. Gradients in reconstruction error maps are found to be steeper using MOSAIC when
Chapter 5. Summary and Future Work: Ultrafast Optics

compared to IAC. A corrective measure has been developed to accommodate signal
distortion arising from electronic bandwidth limitations, interferometer misalignment
and nonquadratic detector response.

The applicability of MOSAIC was extended to the single shot regime. By making
use of the intensity autocorrelation and second harmonic spectrum, a MOSAIC trace
can be computed. This important result was confirmed experimentally; bringing the
utility of visual pulse characterization to low repetition rate laser systems. This
development was done in tandem with a new sensitive graphical representation of
ultrashort phase quality; Hybrid-MOSIAC, which delineates the difference between
temporal and spectral phase distortions.

The production of a MOSAIC trace from an SHG FROG trace was discussed. A
simulated pulse exhibited greater visual sensitivity to pulse chirp in MOSAIC than
SHG FROG.

Based on this the following suggestions for future work are logical:

1) Make the visual sensitivity to pulse chirp comparison experimentally. It is
expected that by making two measurements of a laser pulse (one with chirp and one
without chirp) using SHG FROG followed by a MOSAIC computation, a revealed
increase in visual sensitivity will be found in MOSAIC.

2) Evaluate the accuracy of a pulse retrieved using SHG FROG by computing a
MOSAIC for both the measured and reconstructed FROG traces.

3) Exploration of non-interferometric techniques for pulse characterization. In
the case of very long (≈ps) pulses producing sufficient delay can be impractical, a
phase retrieval phase on a measure of nonlinear spectra would be tractable.
Part II

Time-Resolved Microscopy and Its Application to Laser Cooling in Semiconductors
Chapter 6

Laser Cooling of Solids

6.1 Overview

The focus of this dissertation now turns to the concept of laser cooling. In this chapter the reader will find a review of the laser cooling field in both glasses and semiconductors. Carrier recombination processes in semiconductors are introduced and framed in the context of a radiative external quantum efficiency for laser cooling. The photoluminescence lifetime is used as a measure of carrier recombination dynamics. The unwanted consequence of enhanced nonradiative recombination in the cooling cycle is discussed and motivates the work of subsequent Chapters.

6.2 Laser Cooling

The idea behind laser cooling of solids is the lowering the temperature of a sample by irradiating it with laser light. The essential concept of radiative cooling is the thermal energy of a material is carried away by luminescence photons. For this to
happen, the excitation wavelength of light must be longer than the mean fluorescence wavelength, $\lambda_f$. This process is known as anti-Stokes luminescence. The possibility of cooling solids by fluorescence up-conversion was initially proposed by Pringsheim in 1929 [47]. However, the first experimental observation of laser cooling was not until 1995 by Epstein et. al. [48]. It took many decades between prediction and experimental realization because of a lack of a suitable monochromatic light source (the laser) and development of sufficiently pure materials.

### 6.3 Laser Cooling in Glasses and Crystals

For net laser cooling to be realized, thermal energy must be removed radiatively. The cooling cycle is described by the following three step process, see Fig. 6.1: 1) an excitation photon of energy $h\nu$ is absorbed creating an excited state with lower energy than the mean fluorescence, $h\nu_f$ 2) the excited state electrons acquire additional energy by thermalizing with the surrounding lattice through phonon absorption and 3) electrons relax back to the ground state through the emission of photons that have a higher average energy than the pump photons. If the excitations relax primarily by emitting light and this light is removed efficiently, net cooling of the solid can occur.

At first, Pringsheim’s theory was rejected by some scientists due to its apparent violation of the second law of thermodynamics; which states for a given system not in equilibrium, entropy must increase with time. In 1946 Landau showed the concept was valid by accounting for the entropy of the fluorescent light [49]. The lost entropy of the sample is compensated by the increase in entropy of the fluorescent light; which radiates isotropically, loses coherence and increases in spectral distribution.

Laser cooling of solids can be used to make an all solid-state cryocooler, a possible representation is shown in Fig. 6.2 [50]. Such coolers could serve in in space-borne
Figure 6.1: The laser cooling cycle. Electrons are excited from the top of the ground state to the bottom of the excited state by a pump photon of frequency $\nu$ (red) where they acquire additional energy from the surrounding lattice (yellow). Fluorescence is emitted from the system at energy $h\nu_f$ which carries away thermal energy of the lattice (blue).

applications. Mechanical cryocoolers can achieve low temperatures but do so while producing unwanted vibrations. Thermoelectric coolers are all solid state, but a study by Ball Aerospace Corporation showed that in low-power space-borne applications ytterbium-based optical refrigerators could outperform conventional thermo-electric and mechanical coolers in the temperature range between 80-170K [51].

The advantages of an all-optical solid state cooler are considerable. Such a cooler would contain no moving parts and is therefore applicable to vibration sensitive systems (such as cameras mounted on satellites). Additionally, there are no parts to wear out or degrade. The system would be long-lived and reliable. Issues associated with cryogenic fluids and compressors are also avoided.

Solid state, thermo-electric (TE) coolers, based on the Peltier effect are the pri-
Chapter 6. Laser Cooling of Solids

Figure 6.2: Schematic of an optical refrigeration system. Pump light is efficiently generated by a semiconductor diode laser and carried to the mirrored cooler element by an optical fiber. The laser enters the cooler through a pinhole in one mirror and is trapped by the mirrors until it is absorbed. Isotropic fluorescence escapes the cooler element and is absorbed by the vacuum casing. A sensor or other load is connected in the shadow region of the second mirror.

mary competitors to an all-optical solid state cooler. Even though TE coolers have no moving parts or flowing fluids their absolute temperature is limited to $\approx 150 \text{ k}$ in a multi-stage configuration, which is comparable to temperatures reached with the best performance in laser cooled glasses yet reported [52]. A limitation of the TE cooler design is the trade-off between absolute temperature and cooling power. Unlike TE coolers, potential laser coolers using a glass can achieve temperatures in the range of TE coolers in a single stage [51].

The initial observation of optical refrigeration (laser cooling) by Epstein et. al. was done with a high purity glass sample doped with ytterbium (Yb). The laser cooling cycle was carried out in the Yb atomic system and removed heat from the surrounding glass host material to accomplish net cooling. Additionally, net cooling has also been observed in different host glasses such as fluorochloride ($\text{Yb}^{3+}:\text{CNBZn}$).
as well as fluoride ($\text{Yb}^{3+}$:BIG) by Fernandez et al. [53]. Although these systems have been successful in achieving net cooling, the largest amount of temperature reduction was exhibited with ($\text{Yb}^{3+}$:ZBLAN); showing a decrease of nearly $\Delta T \approx 100K$ from the surroundings [52]. Additionally, laser cooling has been observed in crystal hosts: $\text{Yb}^{3+}$:KGd(WO$_4$)$_2$ [54], $\text{Yb}^{3+}$:YAG [55], $\text{Yb}^{3+}$:Y$_2$SO$_5$ [55], and $\text{Yb}^{3+}$:KPb$_2$CL$_5$ [56]. The crystals offer several advantages over glasses: high thermal conductivity, more mechanically robust, potential for increased absorption cross-section and possible lower thermal emissivity.

In 2000, a demonstration by Hoyt et al. showed net cooling of a thulium-doped ZBLAN glass sample ($\text{Tm}^{3+}$:ZBLAN) at twice the laser wavelength used in the Yb cooling experiments [57]. A peak cooling per absorbed power of -30 °C/W was demonstrated at a net temperature drop of 1.2 °C. This important result showed the energy gap scaling predictions and led to a record amount of cooling power (or heat lift) at $\approx 73$ mW. It was also the first demonstration of laser cooling in the presence of excited-state absorption. It is important to note that this led to a temperature decrease of 26 K from ambient [58]. Recently, cooling by 1.5 K in a Tm doped crystal, ($\text{Tm}^{3+}$:BaY$_2$F$_8$), was demonstrated by Patterson et al. [59]. Even more recent is cooling of Yb:YLF using cavity enhanced resonant absorption. This method allowed cooling of 70 K from ambient [60].

The efficiency of the cycle shown in Fig. 6.1 is understood by the quantum limit; the ratio of the extracted energy to the input energy,

$$\eta_c = \frac{h\nu f - h\nu}{h\nu} = \frac{\lambda - \lambda_f}{\lambda_f}$$

(6.1)

where $\nu = c/\lambda$. Equation 6.1 represents the best possible efficiency. Rearranging Eq. 6.1 gives

$$\eta_c = \frac{\nu f}{\nu} - 1$$

(6.2)
Chapter 6. Laser Cooling of Solids

This limit is based upon that all luminescence photons leaves the system (i.e., are absorbed by a heat sink). However, if the luminescence photons are recycled, such as with a photovoltaic supplying power to the laser, the efficiency can approach the Carnot limit \[50\].

From Eq. 6.1 it would seem that the efficiency could be increased by increasing the wavelength of the pump laser. There is however a practical limit to this: absorption will dramatically decrease when pumping in the long wavelength tail of the transition. The absorbed power is frequency dependent and generally falls off rapidly below the bandgap. As the pump photon energy approaches the mean luminescence energy, the efficiency goes to zero and laser cooling is unattainable. Additionally, the presence of parasitic absorption can lead to undesired heating that can overwhelm net cooling. These considerations imply pumping should be at a photon energy about \(k_B T\) below the mean luminescence energy leading to

\[
\eta_c \approx \frac{k_B T}{E_{\text{gap}}}.
\]  

(6.3)

The mean fluorescent energy, \(h \nu_f\), can be approximated by the energy gap of the material, \(E_{\text{gap}}\).

Thus far it has been assumed that the sole means of excited state relaxation is through photon emission. Unfortunately, there are also nonradiative ways of system relaxation that do not result in the production of a luminescence photon and therefore inhibit cooling. A nonradiative decay process is of particular importance because it results in the production of heat. While the heat lift during one iteration of the cooling cycle is on the order of \(k_B T\), heating due to nonradiative recombination is much larger and on the order of \(E_{\text{gap}}\). The efficiency at which radiative relaxation occurs can be regarded as an internal quantum efficiency of the form

\[
\eta_{\text{int}} = \frac{W_r}{W_r + W_{nr}}
\]  

(6.4)
Chapter 6. Laser Cooling of Solids

where $W_r$ and $W_{nr}$ are the radiative and nonradiative recombination rates respectively.

While a photon may be be produced through radiative recombination, it does not remove heat from the host material unless it escapes. Only a fraction of the radiative recombination results in cooling, $\eta_c$. Hence, an extraction efficiency must be taken into account, resulting in an external quantum efficiency [61].

$$\eta_{ext} = \frac{\eta_c W_r}{\eta_c W_r + W_{nr}}. \quad (6.5)$$

Heating in the sample can occur due to luminescence reabsorption and subsequent nonradiative decay. Care must be taken to maximize the extraction efficiency thereby giving the highest external quantum efficiency possible.

By taking into consideration the external quantum efficiency, the cooling efficiency is then modified to

$$\eta_c = \eta_{ext} \frac{\nu f}{\nu} - 1. \quad (6.6)$$

Not all of the power absorbed in the sample goes into the desired state. The existence of sample impurities for example can give rise to a background absorption, $\alpha(\nu)$. The presence of the background absorption is realized as a departure from unity in the resonant absorption efficiency, expressed as

$$\eta_{abs} = \frac{\alpha(\nu)}{\alpha(\nu) + \alpha_b}. \quad (6.7)$$

where $\alpha(\nu)$ is the frequency dependent resonant absorption and $\alpha_b$ is the parasitic background absorption. Accounting for the presence of parasitic background absorption, which is assumed to be frequency-independent at and around the excitation energy further modifies the cooling efficiency to

$$\eta_c = \eta_{abs} \eta_{ext} \frac{\nu f}{\nu} - 1. \quad (6.8)$$
From Eq. 6.3 the absolute limit of the cooling efficiency is based on the temperature. This limit is due to absorption narrowing from lack of phonons. With rare earth doped glasses and crystals, the excitation for the cooling transition occurs from the top of a ground state manifold to the bottom of the desired excited state manifold. The population of electrons in each manifold is governed by the Boltzmann distribution under a local equilibrium assumption [62]. The quantities $\eta_{\text{abs}}$ and $\eta_{\text{ext}}$ are both functions of temperature. As the temperature of the sample drops, the availability of electrons needed to participate in the cooling cycle decreases. This requires an increase in the pump frequency $\nu$ in order to access electrons of lower energy in the ground state manifold. As $\nu$ approaches $\nu_f$ the efficiency goes to zero. The Boltzmann statistics result in a limit of $\approx 100$ K for cooling Yb [51]. Other condensed matter systems that do not rely on Boltzmann statistics are an attractive alternative as they may offer the prospect of reaching temperatures below 100 K. In RE-doped systems the cooling transition occurs in localized donor ions within the host material, however, in a material such as a semiconductor free carriers are governed by Fermi-Dirac statistics instead of Boltzmann statistics. The transition occurs between extended valence and conduction bands. The maximum cooling power density (rate of heat removal) is proportional to the photo-excited electron (-hole) density and inversely proportional to the radiative recombination time. In semiconductors the optimal density is limited due to many-body processes (derived in the next section) and does not exceed that of moderately doped RE systems. However, one can gain 5 to 6 orders of magnitude in cooling power density because the radiative recombination rates in semiconductors ($\approx$ ns) are much faster than in RE ions ($\approx$ ms) [50].
6.4 Laser Cooling in Semiconductors

The potential of cooling to temperatures lower than 100 K makes semiconductors an attractive alternative to rare earth doped glasses and crystals. Semiconductors offer three advantages 1) the potential for greater cooling power due to a higher excitation density per unit volume, 2) a highly developed growth and characterization community allowing incorporation with optoelectronic elements and 3) the possibility of achieving temperatures as low as 10 K [61].

The primary difference between semiconductors and RE-doped materials is in the cooling cycle. In RE-doped systems the cooling transition takes place in donor ions embedded with the host material, while in semiconductors the cycle takes place between extended valence and conduction bands. Donor ions of RE systems are spatially localized within the host material and are therefore distinguishable. This localization leads to Boltzmann statistics for charge carriers. In contrast, the cooling cycle in semiconductors takes place in the host itself and not in localized donors. Charge carriers are indistinguishable and obey Fermi-Dirac statistics. Charge carriers in Fermi-Dirac distributions may allow semiconductors to get much colder than RE-doped systems. The highest energy levels of the ground state manifold in the RE-doped system become depleted as the temperature is lowered, due to Boltzmann statistics. The efficiency of the cooling cycle is reduced when the Boltzmann constant times the lattice temperature becomes comparable to the width of the ground state [50].

The best choice of a semiconductor for use in laser cooling experiments is from those having a direct band gap. Semiconductors possessing a direct band gap have the minimum of the conduction band and the maximum of the valence band overlapping in k-space (momentum space). This alignment facilitates the direct radiative transition from conduction to valence band, without the need for phonon interaction.
By contrast, indirect band gap semiconductors do not have overlapping conduction band minima and valence band maxima. Radiative transitions in such materials are phonon assisted to conserve momentum. Because of this, radiative transitions have a low quantum efficiency and such semiconductors are therefore undesirable for laser cooling. The primary means of recombination in these materials is through phonon generation.

The laser cooling cycle in semiconductors is very similar to that of glasses and crystals. Electron-hole pairs are created by absorption of pump laser light (at energy $h\nu$) from the top of the valence band to the bottom of the conduction band, see Fig. 6.3a. The initial electron charge carrier distribution following direct excitation is shown in red in the conduction band. The excited state charge carrier distribution thermalizes with the surrounding lattice (blue curve in the conduction band) by absorbing phonons. Radiative recombination emits photons with an average frequency $\nu_f$ higher than that of the pump. This emission at shorter wavelengths from the pump excitation is known as anti-stokes luminescence. Figure 6.3b is an example of anti-Stokes luminescence. A GaAs semiconductor is excited with a pump wavelength of 890 nm and produces broadband luminescence with a mean wavelength of $\lambda_f \approx 860$ nm [63].

The removal of luminescence is a much more formidable problem in semiconductors compared to glasses or crystals. The problem is due to the high index of refraction in semiconductors, $n_0 \approx 3$ to 4. Glasses have a much lower index of refraction, $n_0 \approx 1.5$. Total internal reflection inhibits the escape of light in the semiconductor as stipulated by the condition for the critical angle, $\theta_c = \sin^{-1}(n_1/n_0)$, where $n_1$ is the index of refraction for the surrounding medium, typically air (or vacuum), $n_1 \approx 1$. The critical angle is the angle above which light cannot leave the sample. The higher index of refraction in semiconductors limits the trajectories of light rays that can exit the sample due to total internal reflection. To enhance the
Chapter 6. Laser Cooling of Solids

Figure 6.3: (a) Cooling cycle in laser refrigeration of a semiconductor in which absorption of laser photons with energy $h\nu$ creates a cold distribution of electron-hole carriers (only electron distribution is shown for clarity). The carriers then heat up by absorbing phonons followed by an up-converted luminescence at $h\nu_f$. (b) Typical anti-Stokes luminescence observed in GaAs/GaInP double heterostructure

Luminescence extraction efficiency, geometric coupling schemes such as nearly index matched dome lenses have been used, see Fig. 6.4 [61]. The GaAs/GaInP double heterostructure shown is lifted off a parent GaAs substrate and van der-Waals bonded to a ZnS dome. The GaInP lattice matched layers provide surface passivation as well as carrier confinement [50].

Light rays above the critical angle are totally internally reflected. Which in turn gives rise to luminescence trapping, reexcitation of electron-hole pairs and subsequent reemission [64]. This reexcitation is called photon recycling. Only a fraction of the photons, $\eta_e$, participate in the cooling cycle. As a consequence, the external quantum efficiency for semiconductors takes the form [63]

$$\eta_{ext} = \frac{\eta_e B N^2}{AN + \eta_e B N^2 + CN^3}$$  \hspace{1cm} (6.9)
Chapter 6. Laser Cooling of Solids

Figure 6.4: A GaAs/GaInP double heterostructure is bonded to a nearly indexed matched dome (ZnS or ZnSe) lens for luminescence extraction.

where $A$, $B$ and $C$ are decay coefficients. The $BN^2$ term is the radiative recombination rate. Nonradiative carrier recombination is represented with the $AN$ coefficient (which is dominated by surface recombination in high purity semiconductors) and $CN^3$ [65]. If the carrier density (per unit volume), $N$, is sufficiently high, the three-body Auger recombination term can contribute and is represented empirically in Eq. 6.9 by $CN^3$.

From Eq. 6.9 it is evident that as nonradiative recombination from the $A$ coefficient increases, the external quantum efficiency decreases. The optimum carrier density for maximizing Eq. 6.9 can be found by taking derivative

$$\frac{d}{dN} \eta_{ext} = \frac{\eta_e B (A - CN^2)}{AN + N(\eta_e B + CN)^2}$$

(6.10)

and setting it equal to zero. Which gives an optimum carrier density of $N = \sqrt[3]{\frac{A}{C}}$. 84
Figure 6.5: (a) Nonradiative recombination through multiphonon emission. (b) Radiative recombination giving off a photon of energy $h\nu_f$. (c) Auger recombination.

The $A$, $B$ and $C$ mechanisms of recombination are depicted in Fig. 6.5. Nonradiative recombination through multiphonon emission is shown in (a). Radiative conduction to valence band recombination is shown in (b) with the production of a photon at energy $h\nu_f$. The three body Auger recombination is shown in part (c). The radiative and Auger coefficients are fixed material parameters. Once a semiconductor material is chosen, $B$ and $C$ are constants. However, this is not the case with the nonradiative $A$ coefficient. Variations in growth conditions and sample purity can lead to not only varying values of $A$, but also a positional dependence. Spatial nonuniformity in $A$ will create a spatially nonuniform external quantum efficiency. It is the goal of Chapter 7 to measure the spatial nonuniformity in $A$.

For the laser cooling process, it is desirable to have the dominant recombination mechanism come from $BN^2$ and have as small as possible contribution from the nonradiative processes $AN$ and $CN^3$. Auger recombination can be avoided by limiting the laser excitation to restrict carrier densities to regions where Auger recombination is negligible. Nonradiative recombination, $AN$, however cannot be so
easily controlled. As this dissertation will show, nonradiative recombination is found to have a strong spatial dependence beneath a surface passivation layer. The presence of subsurface defects enhancing nonradiative recombination is confirmed. This means that $A \rightarrow A(r)$ and hence $\eta_{ext} \rightarrow \eta_{ext}(r)$. Through the use of unique spatially and temporally resolved microscopy techniques, characterization of the subsurface demonstrates significant nonradiative structure leading to a strong spatial dependence in external quantum efficiency.

6.5 The Nature of Nonradiative Recombination

The physics of nonradiative recombination is different depending on where the recombination takes place. There are two separate origins for carrier recombination: the bulk and the surface.

6.5.1 Shockley-Read-Hall model

In the bulk, semiconductor nonradiative recombination is described by the Shockley-Read-Hall model [66] and is based on point defects having a single deep-level within the forbidden gap. The energy of this level is assumed to be far from the band edges [67]. The localized mid gap state can also supply an electron. An electron in the conduction band is captured by the neutral state. The energy difference between the conduction band and the mid gap state is given off in the form of a multi-phonon emission. If the trap state is charged, the mid gap states supplies an electron to the conduction band via multi-phonon absorption. The same process can also occur for holes. The energy level of the mid gap state can be above or below the Fermi level.

Point defects giving rise to localized trap states can be created by deep-level impurities such as C, Fe, Ni, Co, W or Au. Additionally, defects can be induced
after the growth of the semiconductor, these handling induced defects are vacancies, interstitials, anti-sites, grain boundaries and dislocations [68, 69, 70, 61].

The surface defects can serve as efficient nonradiative recombination centers that produce heat in GaInP/GaAs semiconductor double heterostructures (DHS). Recombination is a two step process: 1) electrons in the conduction band are captured in a trap and 2) electrons recombine with holes in the valence band. This recombination dissipates energy equal to the gap energy. For GaAs the band gap energy is 1.42 eV at 300 K [68]. The energy is released by multi-phonon emission [71, 70, 69].

6.5.2 Surface States

The atomic structure within a bulk semiconductor is periodic. However, the finite physical size of the semiconductor sample requires it to have edges and therefore surfaces. At these edges, atoms are bound on one side (the bulk of the semiconductor), but unbound on the other side (air/vacuum interface). The semiconductor structure ceases to be periodic and as a result dangling bonds allow discrete or continuous energy states within the forbidden band gap.

Understanding the origin of surface states can be gained from considering a one dimensional semi-infinite chain of atoms, see Fig. 6.6a. Which gives rise to a periodic potential of the form

\[ V(z) = 2V_o \cos\left(\frac{2\pi z}{a}\right) \]

for \( z < 0 \) [72]. The spacing of the atoms is \( a \). The electronic state as a function of spatial coordinate \( z \) is given by the solution to the time independent Schrödinger equation

\[
\left[-\frac{\hbar^2}{2m} \frac{d^2}{dz^2} + V(z)\right] \Psi(z) = E \Psi(z)
\]
Figure 6.6: (a) A semi-infinite periodic chain of atoms representing a semiconductor with a surface at \( z = 0 \). (b) Solution to the time independent Schrödinger equation showing positional dependence of surface states at a semiconductor/vacuum interface.

where \( m \) is the mass of the electron, \( \hbar \) is Planck’s constant divided by \( 2\pi \) and \( V(z) \) is the periodic potential, Eq. 6.11. The solution to Eq. 6.12 is a decaying exponential for \( z > 0 \) and a decaying oscillating exponential for \( z < 0 \). The numerical solution is plotted in Fig. 6.6b. A nonzero wavefunction, \( \Psi(z) \), indicates the presence of surface states extending into the bulk. The energy level of the surface states lies within the forbidden bandgap of the bulk states and can serve as a nonradiative recombination pathway. The effect of these states falls off exponentially and typically doesn’t exist for more than a few angstroms into the semiconductor [72]. The small spatial extent of the surface states can be passivated by only a few atomic layers.

Figure 6.7 shows the different ways carrier recombination can occur in the bulk or at the surface. Shown in the figure is an active layer of GaAs passivated above and below by GaInP. Charge carriers (green dots) are created uniformly throughout the active layer from globally uniform optical illumination. In case (a), carrier diffusion and surface recombination are shown. Carriers within a diffusion length of the unpassivated side edge of the GaAs can experience surface recombination. A localized luminescence collection probe, shown above the sample, records dimin-
ished luminescence due to the presence of the surface states. Even though surface states extend only a few angstroms into the GaAs bulk, their presence is felt from many micrometers away due to carrier diffusion, see section 6.6 and 7.9. Case (b) is that of a defect visible from the top GaInP surface. The defect is not located near the edge of the sample. Such a defect perturbs the periodic potential of Eq. 6.11 and gives rise to interface states at the bulk/defect junction. Like the case of the vacuum/semiconductor interface of Fig. 6.6 surface states will exist and are accessible to carriers through diffusion. Figure 6.7b shows carriers on both the left and right sides of the defect and the collection probe on only the left side of the defect. Carriers created within a diffusion length but to the right of the defect will not diffuse through the defect to the region of luminescence collection. Because of this, diminished luminescence is again recorded at position (b).

Figure 6.7: A GaAs/GaInP double heterostructure illuminated globally with pulsed excitation generating a uniform density of e-h pairs. (a) Carrier diffusion and surface recombination at an unpassivated GaAs edge (sample edge). (b) A defect propagating through all layers of the GaAs/GaInP double heterostructure serves as a nonradiative recombination site. (c) A point defect located only in the GaAs. The presence of this defect is not seen in the surface topology of the GaInP. (d) A defect at the GaAs/GaInP interface layer. Such a defect is not visible from the surface of the passivating GaInP layer. (e) The ideal case where all generated charge carriers are free to participate in radiative recombination.
Chapter 6.  Laser Cooling of Solids

Figure 6.7c depicts a point defect located in the GaAs layer only. While the point defect maybe small and the spatial extent over which the defect perturbs the bulk periodic potential short, its effects are accessed over a broader region due again to diffusion. Example (d) of Figure 6.7 shows a defect existing in the interface between the GaAs and the GaInP passivation layer. If the passivation layer is too thin or not spatially uniform, the top surface of the GaAs will not be adequately passivated and surface states will again arise. Defects depicted in examples (c) and (d) would not be visible from the top surface of the DHS. To identify such defects, characterization beyond surface morphology is needed. Chapter 7 details a unique simultaneous surface/subsurface characterization suited for this task. Figure 6.7e depicts the ideal situation excitation far from the sample edge, having adequate passivation and no defects nearby. Luminescence is the strongest here. The area is uniform and the $A$ coefficient is constant in this region. This case would have the highest and most uniform $\eta_{ext}$. Identifying such homogeneous regions is critical for success in semiconductor based optical refrigeration.

6.6 Carrier Diffusion

In the experimental condition that a sample is globally illuminated, charge carriers would be evenly distributed throughout the bulk. However, in the case of a localized measurement, such as shown in Fig. 6.7, charge carriers created in the bulk can diffuse to an unpassivated edge or defect and undergo nonradiative recombination. In the absence of an external electric field influencing carrier position the density rate equation becomes [73]

$$\frac{dN(r,t)}{dt} = G - [A(r) + \eta_e B N_d]N - \eta_e B N^2 - C N^3 + D_a \nabla^2 N$$  \hspace{1cm} (6.13)
Chapter 6. Laser Cooling of Solids

where $N$ is now a function of both space and time. $D_a$ is the ambipolar diffusion coefficient. The spatial diffusion accounted for here is only considered in the horizontal plane as the vertical dimension of our semiconductor samples is thin (750 $\mu$m) and therefore assumed to be uniformly populated with carriers.

The diffusion length, i.e., the average distance an excited carrier travels before recombining, can be estimated for GaAs samples. If the diffusion length becomes comparable to either the laser spot size or the sample area, carrier diffusion can be important. Because electrons are less massive than holes and therefore have a higher mobility, the ambipolar diffusion coefficient is dominated by the electron diffusion such that, $D_a \approx D_e$. The diffusion length, $L_e$ is related to the diffusion coefficient and the recombination lifetime, $\tau$, by $L_e = \sqrt{D_e \tau}$. The electron diffusion coefficient is arrived at from the Einstein relation $D_e = \mu_e k_b T/e$ where $e$ is the charge of the electron and $\mu_e$ is the electron mobility [73]. The recombination lifetime depends on the carrier density, but for high quality samples (small value for $A$), radiative recombination is dominant. Published values for the radiative coefficient range from $2 \times 10^{-16} \leq B \leq 7 \times 10^{-16}$ $m^3/s$ [64]. Taking a value of $B = 4 \times 10^{-16}$ $m^3/s = 4 \times 10^{-10}$ $cm^3/s$ and assuming a carrier density of on the order of $10^{17}$ $cm^{-3}$ the radiative lifetime is $\tau = 1/(BN) = 25$ $ns$ at room temperature. The electron mobility can also be estimated by assuming a background dopant (impurity) density of $10^{15}$ in which case the mobility is reported as $\approx 8500$ $cm^2V^{-1}s^{-1}$ [73]. From the Einstein relation, this gives a diffusion coefficient of $D_e \approx 260$ $cm^2s^{-1}$. Finally, the average diffusion length can be estimated as $L_e \approx 25\mu m$. A similar analysis can be performed for holes. The mobility for holes is $\mu_h = 400$ $cm^2V^{-1}s^{-1}$ and the diffusion coefficient is then $D_h = 10$ $cm^2/s$. The diffusion length for holes using a radiative lifetime of $\tau = 25$ $ns$ is $L_h = \sqrt{D_h \tau} = 5$ $\mu m$. This means that electron-hole pairs created within a diffusion length of the unpassivated sample edge can still experience surface recombination, despite the fact that surface states do not extend more than a few angstroms into the bulk. While the nonradiative recombination coefficient might
be small in high quality samples, electron-hole pair generation near an unpassivated edge can lead to short nonradiative lifetimes due to carrier diffusion and subsequent surface recombination.

6.7 Carrier Dynamics

Chapter 7 of this dissertation will spatially characterizes defects using time-resolved photoluminescence as a measure of the carrier dynamics. The time dependence of carrier recombination is

\[
\frac{dN}{dt} = G - A(r)N - \eta_e B N_d N - \eta_e B N^2 - C N^3
\]

where \(BN_d\) is radiative recombination to an impurity level (donor or acceptor), \(N_d\) is the density of the ionized impurities, \(G\) is the generation term: \(G = \frac{\alpha(\nu,N)I}{h\nu}\) with laser irradiance \(I\) and \(B\) is the radiative coefficient for GaAs. Radiative recombination is accomplished through dopant mediated \((BN_dN)\) and bimolecular \((BN^2)\) pathways. The nonradiative lifetime is \(\tau_{nr} = 1/A\). The generation is assumed to be an impulse excitation such as from a laser pulse much shorter than the lifetimes of the recombination mechanisms. Following this impulse excitation \(N_o\) photocarriers are generated, and their decay is expressed as

\[
\left. \frac{dN}{dt} \right|_{\text{decay}} = -(A(r) + \eta_e B N_d)N - \eta_e B N^2 - C N^3.
\]

The nonradiative lifetime is mostly dominated by surface recombination, but can also be significant in the bulk near surface defects. For sufficiently low excitation densities or late times in the decay, both the radiative and Auger recombination
terms can be neglected and Eq. 6.15 can be solved analytically. The carrier density solution is

\[ N(t, r) = N_0 e^{-A(r)'t} \]  

(6.16)

where the substitution \( A(r)' = A(r) + \eta_e BN_d \) has been made because of the indistinguishability of the terms that are linear in \( N \) from Eq. 6.15.

In our experimental condition, we can adjust the excitation power to assure \( CN^3 \) (Auger) is negligible. At moderately low enough initial excitation, the radiative recombination term is included and only the Auger term can be ignored, the carrier recombination equation can again be solved analytically:

\[ N(t, r) = N_0 e^{-A(r)'t} \left(1 + \frac{\eta_e BN_o}{A(r)'(1 - e^{-A(r)'t})} \right). \]  

(6.17)

A derivation of this solution can be found in appendix A.

If, however, the carrier density is low enough that Auger can be ignored, but high enough that nonradiative can also be ignored, the following radiative approximation can be made

\[ \frac{dN}{dt} = -\eta_e BN^2. \]  

(6.18)

The solution, see appendix A, is

\[ N(t) = \frac{N_0}{1 + \eta_e BN_0 t} \]  

(6.19)

when the dominant recombination pathway is through photon emission. This solution can also be obtained by taking the limit of small \( A't \) or early time in Eq. 6.17.

The photoluminescence (PL) incident on a detector is expressed as:

\[ PL = \eta_{sys}[BN N_d + BN^2] \]  

(6.20)
where $\eta_{sys}$ is a constant of the experimental system and depends not only on the extraction efficiency, but also on the collection optics used to route the light to a detector and quantum efficiency of the detector used to record the PL. For convenience, Eq. 6.20 is rewritten in the form

$$PL \propto B^2 NN_d + (BN)^2.$$  \hfill (6.21)

The expression is independent of $A'$ when a purely radiative solution is used. This means signals measured with an electron-hole density in the radiative regime contain no information beyond an upper limit on $A'$.

The experimentally detected PL signal can then be expressed in a functional form by substituting any of the analytic solutions to the carrier recombination equation (Eq’s. 6.16, 6.17 or 6.19) into Eq. 6.21. Furthermore the PL expression can be normalized to its peak values at $t = 0$. The resulting expressions for the normalized PL under the aforementioned approximations are

$$\tilde{PL}(t,r)_{A'} \propto \frac{e^{-A't}(e^{-A't} + \frac{N_d}{N_o})}{1 + \frac{N_d}{N_o}} \quad (6.22)$$

$$\tilde{PL}(t,r)_B \propto \frac{1}{(\eta_e BN_o t + 1)^2(1 + \frac{N_d}{N_o})} + \frac{1}{(\eta_e BN_o t + 1)^2(1 + \frac{N_o}{N_d})} \quad (6.23)$$

$$\tilde{PL}(t,r)_{A'B} \propto \frac{e^{-2A't} + \frac{N_d}{N_o}[1 + \frac{\eta_e BN_o}{A'}(1 - e^{-A't})]e^{-A't}}{(1 + \frac{N_d}{N_o})[1 + \frac{\eta_e BN_o}{A'}(1 - e^{-A't})]^2} \quad (6.24)$$

where Eq. 6.22 is from the nonradiative approximation, Eq. 6.23 is from the radiative and Eq. 6.24 contain both radiative and nonradiative recombination terms. All solutions assume Auger can be ignored. For clarity, $A'(r)$ has been written as $A'$. At late times when the carrier density is low, Eq. 6.21 can further be simplified by
Chapter 6. Laser Cooling of Solids

assuming radiative recombination is not bimolecular, but strictly dopant mediated. Therefore, $PL \propto BNN_d$ and the normalized luminescence is simply

$$\bar{PL}_{A'd} \propto e^{-A'(r)t} + \Delta.$$  \hspace{1cm} (6.25)

In experimentation, Eq. 6.25 does not decay to zero, but rather to a constant noise floor, $\Delta$. The offset is a consequence of low signal in the late time PL intensity. If signal-to-noise permits, the single exponential decay tail of the PL can be resolved and the nonradiative lifetime can be extracted directly from it.

A few sample cases of PL will now be discussed. Equation 6.24 describes the measured signal when both radiative and nonradiative recombination mechanisms are present. Figure 6.8a (blue line) shows a semiconductor luminescence decay with $\eta_eBN_o = 1 \times 10^{-10}$ cm$^3$/s, $N_o = 5 \times 10^{17}$ cm$^{-3}$, $N_d = 1 \times 10^{13}$ cm$^{-3}$, $A = 1/500$ ns$^{-1}$ and $\Delta = 0$. In this case $A' \approx A$. Note that the plot is semilog. At early times the decay is curved, indicating that it is not single exponential. As the curve progresses to later times, the late-time single exponential approximation takes over and the decay appears as a straight line for the last $\approx 600$ ns. The slope of this line is $A'$ and is the quantity measured in experiment. The dashed red line is plotted under the same conditions with the exception that $\Delta = 3 \times 10^{-4}$. Here it can be seen that the decay is curved throughout the time window plotted and becomes flat at the end of the decay. The late time single exponential behavior is buried beneath the offset $\Delta$. Figure 6.8b is plotted under the condition $\eta_eBN_o = 1 \times 10^{-10}$ cm$^3$/s, $N_o = 1 \times 10^{15}$ cm$^{-3}$, $N_d = 1 \times 10^{16}$ cm$^{-3}$ and $A = 1/500$ ns$^{-1}$. As the ionized impurity density increases with respect to the initial optically created density, $N_o$, single exponential behavior is exhibited at earlier times. When $N_d \geq N_o$ the entire duration of the decay is single exponential. Measured luminescence decays exhibiting the behavior of Fig. 6.8(a, b) are shown in (c, d) respectively. The measured decay of Fig. 6.8c will be revisited in appendix C.
Chapter 6. Laser Cooling of Solids

Figure 6.8: Photoluminescence from Eq. 6.24. (a) $N_d \ll N_o$ late time single exponential behavior is evident (blue line). When SNR < 1 in the late time, single exponential behavior becomes buried beneath the noise floor (red dashed line). (b) $N_d \geq N_o$ single exponential behavior spans the measurement time. Experimental examples for the conditions in (a) and (b) are shown in (c) and (d) respectively.

For defect-free regions within the bulk, the nonradiative recombination coefficient is much smaller than $BN_d$. This means $A' \approx BN_d$. However, near interfaces the $A$ coefficient can become large such that $A \gg BN_d$ and $A' \approx A$. Because $A'$ is the measured slope it is not possible to separate $A$ from $BN_d$ unless a temperature dependent study is performed [65]. For a given semiconductor sample, the impurity density does not change by more than a few percent across the wafer [74, 75]; however, as the next chapter will show, $A$ can change by an order of magnitude.
6.8 Summary

The concept of laser cooling was discussed in this chapter. Progress of the optical refrigeration field for glasses and crystals was reviewed. Expressions for cooling efficiency were presented for semiconductors. Carrier dynamics were modeled using analytical solutions to the carrier recombination equation. The recombination equation was solved under several approximations corresponding to different initial carrier concentrations. Semiconductor photoluminescence measurements served as an indication of carrier recombination.

The next chapter will detail the microscopy techniques used to spatially and temporally resolve the defect dependent luminescence signature of GaInP/GaAs double heterostructures used in laser cooling experiments.
Chapter 7

Near-Field and Atomic Force Microscopy

7.1 Overview

The nonradiative recombination coefficient of GaInP/GaAs DHS used in laser cooling is spatially and temporally resolved at room temperature in this chapter. Using a unique combination of atomic force and near-field microscopy subsurface defects are identified and quantitatively analyzed in both spatial extent and nonradiative lifetime through fluorescence lifetime images (FLIMs). Time correlated single photon counting (TCSPC) is used to record photoluminescence decays in the near-field. The chapter begins with a brief explanation of atomic force and near-field microscopy methods. The chapter ends by describing a simple yet sensitive method for real-time optical screening of semiconductor samples.
Chapter 7. Near-Field and Atomic Force Microscopy

7.2 Microscopy

The capability to observe and measure samples in new and revealing ways is an essential tool in science. The field of microscopy has provided great insight to many areas of study from physics to biology. Microscopy consists of the three well known branches. They are optical, electrical and scanning probe. Optical and electron microscopy are based on the reflection, diffraction and refraction of radiation to create an image. The first of these fields to be studied was far-field optical microscopy. Far field refers to the separation of the object and its image being much greater than the wavelength used in producing the image. In 1873 the first satisfactory theory of spatial resolution using coherent radiation was formed by Ernest Abbe [76], see also [77]. In this work, Abbe treated objects to be imaged as a diffraction grating, every element of the object must be taken into account in determining the image plane. This was in addition to accounting for every diffracting element on the aperture of the objective. The double diffraction theory leads to the minimum resolvable value of the grating spacing, $a_{\text{min}}$ [78]:

$$a_{\text{min}} = \frac{0.5\lambda}{n \sin \theta}$$

(7.1)

where $n$ is the index of refraction at the object, $\lambda$ is the wavelength and $2\theta$ is the angle over which the light is collected. According to this equation, the minimum resolvable spatial frequency using visible light is about 200 nm.

In contrast, scanning probe microscopy is based on a the interaction of a probe with a sample surface and is not limited by the far-field diffraction limit. Raster scanning of the sample (or probe) builds up an image point by point (pixel by pixel). Probe based microscopy has advantages and disadvantages when compared to far-field optical microscopy. There are three main advantages to scanning probe microscopy:
Chapter 7. Near-Field and Atomic Force Microscopy

1) Spatial resolution of near-field scanning optical microscopy (NSOM) is limited by the size of the probe used. Commercial nanofabrication allows probe tip size to be less than 200 nm in diameter [79]. This means the minimum resolvable feature size is not constrained by Eq. 7.1 and sub-diffraction image features can be seen, see Fig. 7.1. A diffraction limited Gaussian beam is shown illuminating three dots in the focal plane. These three dots are separated by a wavelength and cannot be individually resolved. However, an aperture having a diameter less than the wavelength of light and placed in the near-field of the dots can illuminate a single dot.

![Image of spot size comparison for near-field and far-field illumination schemes.](image)

2) The image created is not necessarily based on optical intensities. The type of image is specified by the probe used. For example, thermoresistive probes allow resistance measurements as a function of position while thermocouple probes perform temperature mapping.

3) Multiple images can be obtained in a single scan. Often probes are dual use, making two simultaneous measurements for every point in a scan. This yields two images, such as surface morphology and surface temperature.
Chapter 7. Near-Field and Atomic Force Microscopy

The primary disadvantage of probe based microscopy is image acquisition speed. Because images are created in a pixel by pixel fashion, scanning times of minutes to hours are encountered. As will be shown later, scanning times needed for time resolving optical signatures in the near-field can exceed 7 hours for a 50 by 50 pixel image.

7.3 Atomic Force Microscopy

One of the most common scanning probes techniques relies on an atomic force interaction between the probe tip and the sample. This technique is known as atomic force microscopy or AFM. An AFM image is a measure of surface topography. A sample is placed beneath a needle-like AFM probe and is scanned to create an image. For each point in the scan, the AFM probe "feels" the surface; gauging its height with respect to other positions in the scan.

The principle of AFM is based upon on an electromagnetic interaction of the atoms in the probe tip and the sample surface. As the probe tip approaches the sample, electronic structure in one component induces a structure in the other and an attractive force is generated between the two [80]. The electronic interaction between the tip and sample is governed by the van der Waals force which is proportional to $1/R^6$ where $R$ is the separation distance. Due to the strong dependence on separation, tip to sample distances involved with AFM scanning are in the nm range. A measure of the height of the sample is realized through a deflection of the tip due to the attractive van der Waal force. The AFM probe is itself a spring oscillating at resonance. The natural resonant frequency is given by $\omega = \sqrt{k/m}$ where $k$ is the spring constant and $m$ is the mass. When the probe tip is close enough to the sample the attractive van der Waal force bends the probe. The magnitude of the deflection of the probe is measured and is proportional to the tip displacement, see Fig. 7.2.
In practice, the tip displacement (or bend) is maintained throughout the scan and the sample is moved up or down (depending on topography) based on feedback from the measured probe deflection.

One implementation of the feedback mechanism is to reflect a laser beam off the AFM probe and onto a position sensitive detector (PSD), see Fig. 7.2. As the tip interacts with the sample scanned beneath it the detector records extension or flexion of the probe and adjusts the sample height accordingly. This type of feedback is referred to as normal force feedback; the sample moves up and down, tracing out surface structure. Normal force feedback AFM can be performed in two modes of operation contact and non-contact. 1) With the contact mode of operation the probe tip touches the surface, tracing out surface morphology as the scan progresses (sometimes this is referred to as profilometry). Contact mode has the disadvantage of wearing the tip and shortening its lifetime. 2) Non-contact mode. The probe tip does not touch the sample but resides just above it. Probes have long operating life in non-contact mode. With non-contact mode AFM, the probe tip oscillates at its natural resonance ($\approx$ 1-100 kHz). In turn, this modulates the detected signal at the PSD. Lock-in detection is used to isolate the signal at the natural resonance frequency,
Chapter 7. Near-Field and Atomic Force Microscopy

separating it from spurious vibrations. The average tip to sample separation depends on a chosen amplitude, but at its nearest point the tip is only 1-2 nm from the sample surface [81]. Non-contact mode is sometimes termed non-contact tuning referring to the signal selection on resonance. While non-contact AFM is always performed with an oscillating tip, contact mode may or may not be. A second type of feedback in use is known as shear force. With the shear force implementation, a piezo scanner with attached AFM probe form an electromechanical system. The probe is dithered about and average position and the power dissipation on resonance is sensitive to the change in damping force as the tip approaches and interacts with the sample [82]. The primary advantage of this feedback system is that it is non-optical and hence eliminates unwanted excitation or detection associated with the feedback laser used in normal force configurations. Non-optical feedback is of particular interest when performing a simultaneous AFM and optical measurement. However, it has been shown that many near field optical images obtained with shear force techniques contain serious artifacts that are connected with the coupling of shear force feedback to the sample characteristics at the surface. Thus, as one changes from one sample to another, it is difficult to discern whether the signal change is due to a change in the frictional force of the sample or to an actual topographic change [79, 83]. Normal force feedback is insensitive to these changes. Additionally, the contact mode of operation is inaccessible with shear force feedback.

The first demonstration of AFM was done by Binnig et. al. in 1986 [84]. In their work, the sample was mounted to an $xyz$ piezoelectric scanner. The tip was also mounted on a piezoelectric driver. Feedback signaling the tip to sample separation was monitored by passing a small current from the tip through the sample. The current tunnels across the tip to sample gap providing a sensitive measure of separation. Piezo drives adjust the separation distance to maintain the tunneling current at a constant level. Four modes of operation were investigated: 1) the sample is modulated in the z-direction at its resonant frequency. This in turn modulates the
Chapter 7. Near-Field and Atomic Force Microscopy

detected current used in the AFM feedback. 2) and 3) the probe tip is modulated in the z-direction. A change in tip to sample distance changes both the amplitude and phase of the modulated monitor current. The amplitude was used for mode (2) and the phase was used for mode (3) as a signal to drive feedback circuits. 4) Both piezo feedback circuits are used in such a way that sample and tip were driven in opposite directions. The fourth method proved to be the most reproducible and showed nano-Newton force sensitivity on a ceramic (Al$_2$O$_3$) sample. This same technique was later used to image the crystalline structure of graphite, molybdenum disulfide and boron nitride on the atomic scale [85].

7.4 Near-field Microscopy

The basic principle of near-field microscopy is based on light illuminating a sample through a sub-wavelength aperture. The aperture is placed in the near-field of the sample, a distance much less than the wavelength of light. The resolution achieved is much better than what can be obtained with conventional far-field microscope. The limiting factor in the minimum achievable resolution is the diameter of the aperture (usually 50-100 nm) [79].

The procedure of creating an image using near-field microscopy is comprised of three major components:

1) A point light source must be brought to within a few nanometers of the sample surface.

2) The point light source is scanned across the sample surface or the sample is scanned beneath the light source. The light source to sample separation is maintained throughout the scan.

3) The optical signal from the region illuminated by the point light source must
be collected and detected.

There are several different ways to obtain a point light source. One of the most common ways is to use a pulled or etched fiber optic. By heating a fiber and drawing it out a taper can be produced. The tapered end is coated with a metal except for the very tip, where a small aperture remains uncoated and allows optical transmission. These apertures can be manufactured as small as 50 nm [79]. Light is focused into the opposite end of the fiber and is emitted from the tapered metal coated tip. A standard AFM probe can also be used. In this implementation, the AFM probe is created with a small (sub-wavelength) hole through its pyramidal tip. Laser light is focused through the hole. Additionally, a nanopipette can be filled with a fluorescence material and excited by either a voltage or an optical pump.

The placement of the point light source in the near-field of the sample is accomplished by means of a feedback mechanism. This done in an analogous way to AFM; consisting of two primary methods. The normal force feedback scheme of Fig. 7.2 can be used with cantilevered tapered optical fibers and AFM cantilevered probes made with a hole in the tip. Shear-force (or tuning fork) feedback can be used for straight tapered fibers. The amplitude of the fiber tip oscillation is dependent upon the tip to sample distance and can be monitored in a feedback circuit. The near-field work reported in this dissertation is done exclusively with the normal force feedback mechanism.

Four distinct modes of operation are available in NSOM. They are shown in Fig. 7.3. Figure 7.3a illustrates transmission mode operation. The sample is excited locally in the near-field through the aperture (red line). Sample luminescence or transmitted excitation (blue line) is collected globally from below (typically with a microscope). Similarly in Fig. 7.3b, reflection mode operation is shown with local excitation and global collection of luminescence from above. The third mode of operation is collection mode; sample luminescence or excitation light is collected
Chapter 7. Near-Field and Atomic Force Microscopy

Figure 7.3: The four modes of operation in NSOM. (a) Transmission: local excitation global collection from below sample, (b) Reflection: local excitation global collection from above sample, (c) Collection: global excitation and local collection and (d) Illumination/Collection mode: local excitation and collection are performed through the tip.

locally in the near-field through the aperture, while excitation is global, Fig. 7.3c. Lastly, illumination/collection mode is shown in 7.3d. The sample is illuminated through the same tip that collects the optical signal. This mode is sometime referred to as local-local operation because illumination and collection are both done through the near-field aperture.

Detection of the collected light can be performed in a variety of ways. Detectors that can be used are avalanche photo diodes (APD), photo multiplier tubes (PMT) or spectrometers. The production of an image using NSOM is done by collecting through one of the four modes of operation followed by detecting at each position of the scanner. An example might be illumination through the near-field aperture, global collection from above using a conventional far-field microscope and detection of the reflected illumination off the sample surface using a PMT. This would yield an intensity image proportional to the reflectivity of the sample surface. Other modes of operation can be used with different detectors to create different images. It is this flexibility that makes NSOM a powerful tool for optical characterization of semiconductor samples used in laser cooling.
Chapter 7. Near-Field and Atomic Force Microscopy

A historical perspective reveals that the idea of extending the spatial resolution of microscopy is not new and the concept for the use of near-field apertures was originally proposed by Synge in 1928 [86]. He suggested using either a quartz cone coated with metal everywhere except the tip or a pinhole in a metal plate to produce a sub-wavelength point light source. In 1956 O’Keefe proposed the idea of near-field microscopy but was aware of the primary difficulty in achieving the short controlled distance between the sample and pinhole [87]. The first demonstration of scanning near-field microscopy was done in 1972 by Ash and Nichols [88]. This work showed spatial resolution on the order of $\lambda/15$ using 3 cm wavelength radiation. It wasn’t until 1984 that NSOM was shown to be a practical scientific tool by Pohl et al. [89], who showed 25 nm sized details using 488 nm radiation.

The experimental implementation of NSOM and AFM is very similar. A feedback mechanism brings a probe close to a sample surface and builds up an image scanning point by point. These two forms of microscopy can be combined in a single apparatus to perform a simultaneous topographic and optical characterization. The setup for this dual measurement scheme is shown in Fig. 7.4. A tapered, cantilever fiber optic serves as both a deflecting AFM probe while at the same time collecting optical information from the region of the sample directly below the aperture. The setup in Fig. 7.4 shows a global illumination and local collection, but all four modes of NSOM operation shown in Fig. 7.3 can be used. A great benefit of global illumination and local collection setup is NSOM and AFM images can be spatially correlated. A direct comparison can be made between any $AFM(x, y)$ pixel and any $NSOM(x, y)$ pixel. The same can be said for the local-local configuration. Transmission mode also facilitates direct comparison, but is only useful when a sample is transparent. It is important to note that local illumination with global collection does not give a direct comparison of AFM and NSOM images, rather there is a constant spatial offset such that, $AFM(x, y) = NSOM(\alpha_x x, \alpha_y y)$. The spatial offset is because the NSOM image from global collection originates from light collected around the probe tip and
Chapter 7. Near-Field and Atomic Force Microscopy

Figure 7.4: Dual AFM/NSOM setup. This scheme is used to create spatially overlapping, simultaneously acquired AFM and NSOM images.

not directly beneath it where the AFM measurement is made. Regions beneath the aperture are obscured by the fiber. The spatial offset is of particular importance when a sample has both a physical and optical edge. However, this issue can be avoided provided an NSOM/AFM configuration that is insensitive to the spatial offset is chosen.

7.5 AFM Characterization of GaInP/GaAs Double Heterostructures

Using the microscopy techniques outlined in the previous sections, surface characterization of a GaInP/GaAs double heterostructure is performed. Characterization begins with size quantification of any defects present on the sample. High quality semiconductor samples are provided by the National Renewable Energy Laboratory (NREL). The samples are grown using Metal-Organic Chemical Vapor Deposition (MOCVD). The cross-sectional structure of the NREL sample MF438 is illustrated
in Fig. 7.5. Sample MF438 is a uniquely grown sample because of its 30 nm passivation cap layer. All other samples from NREL have a symmetric passivation around the GaAs active layer, 750 nm above and below. The AlAs serves as a release layer. A selective etch dissolves only the AlAs layer, releasing the GaAs active layer sandwiched between two GaInP passivation layers. It is important to note that the sides of the GaAs active layer remain unpassivated.

While the hope is for the sample surface to be smooth, a random distribution of surface defects is found. The presence of defects is noted not only on MF438, but on the majority of samples. By using a conventional far-field microscope mounted above the AFM, pictures of the surface as well as surface morphology can be recorded. Schematically, this configuration is shown in Fig. 7.6. In practice, a charge coupled device (CCD) serves as a camera displaying images at video refresh rates. The CCD is mounted into an ocular port of the microscope and is interfaced to a computer for image display and archiving.
An example view of the MF438 sample surface is shown in Fig. 7.7a [90]. This optical image was taken with 10x magnification microscope objective. Several large irregularly shaped black objects are seen in the center and top center portion of the image. These are surface defects similar to the one depicted in Fig. 6.7b. The cantilevered probe tip is also visible as the out of focus streak across the middle left half of the image. Because the microscope is focused on the sample surface, the AFM probe body (above the sample) is blurred. The white box in the scan indicates the approximate area over which a 33 by 33 µm AFM scan is being performed, see Fig. 7.7b. The surface topology is recorded with a Nanonics AFM/NSOM-100 (see Chapter 8). Figure 7.7b shows a defect height of about 3.5 µm above the planar background. The scan consists of 256 by 256 individual AFM deflection measurements made in the non-contact mode. The pixel dwell time is 5-10 ms and the scan duration was < 20 minutes. The normal force feedback mechanism is used to maintain a constant tip to sample separation with a proportional-integral-derivative (PID) control system. The settings of the Nanonics AFM/NSOM Topaz feedback controller were $P = 0.1 \times Maximum$, $I = Maximum$ and $D = 0$. A Nanonics AFM/NSOM probe (similar to that pictured in Fig. 7.4) with a 500 nm tip size
Figure 7.7: (a) Optical image taken with a 10x magnification microscope objective. Surface defects are seen as black irregularly shaped objects. Probe tip is also visible. (b) AFM image of the area outlined by the white box in (a). The surface defect has a height of $\approx 3.5 \, \mu m$ from the planar background.

was used. Features smaller than 500 nm would not be observed in this AFM scan. Prior to the AFM measurement, the sample received a through methanol based swab cleaning. Attempts were also made to clean the sample in an ultrasonic methanol bath. Surface defects could not be removed with either of these two procedures.

The frequency and height of the defects found in this sample have consequences for potential laser cooling experiments. As was mentioned in Section 6.4, efficient luminescence extraction is critical for laser cooling success; luminescence must leave
the sample to carry away heat. Luminescence removal is accomplished with a nearly index matched dome lens as seen in Fig. 6.4 [64]. Domes manufactured from ZnS or ZnSe are transparent in the infrared, provide near index matching and present a minimal heat load from blackbody radiation. This relaxes the critical angle restraint and allows more luminescence to leave the sample, thereby increasing $\eta_e$. Analogous schemes are used for light extraction in LEDs. Bonding, however, can be problematic when surface defects are present. Defects between the sample and dome prevent uniform bonding. Luminescence removal in the regions of defects is reduced due to a lack of contact. Photon recycling effects are enhanced in the unbonded areas of the sample. Photon recycling can likely lead to increased heating in these areas and can potentially overwhelm net cooling of the semiconductor.

The composition of the surface defects is found to be made of the same material constituting the semiconductor [91]. Energy dispersive X-ray spectroscopy (EDXS) confirms the elemental make-up of the defects. The method of EDXS works by inelastically scattering electrons off of electrons bound to atoms in a solid. Ejection of bound electrons can occur if the incident electron energy is high enough. When electronic vacancies are filled, X-ray photons are emitted. Elemental analysis is performed by resolving the unique X-ray emission with an energy dispersive spectrometer.

Later it will be shown that the nonradiative recombination coefficient, $A$, is enhanced on and around surface defects. The next section will discuss how the semiconductor luminescence lifetime is measured.

### 7.6 Time Correlated Single Photon Counting

One of the drawbacks of measuring an optical signal using NSOM is low photon throughput and the commensurate extended pixel dwell times needed to acquire
adequate signal. From Fig. 7.4 it is evident that light collected through the near-field aperture must propagate through a tapered fiber and ultimately to a detector. To achieve sub-diffraction limited imaging, the size of the aperture must be less than half the wavelength of light being used. One method of producing such a small diameter aperture is to use a fiber optic, by heating and drawing a fiber, the cross-section is tapered down to a fine point which can be used as a near-field aperture. In the region of the fiber optic between the near-field aperture and where the tapered fiber diameter becomes equal to or greater than half the wavelength of light propagating through it, the fiber optic behaves as a lossy waveguide. The power flow in this region is described by

\[ P(z) = P_0 e^{-2\beta_\lambda z} \]  

(7.2)

where \( z \) is the axial propagation direction and \( P_0 \) is the power incident on the aperture. The attenuation constant, \( \beta_\lambda \), can be found by either solving the boundary-value problem with boundary conditions appropriate for finite conductivity, or by calculating the ohmic losses [92]. The power loss experienced in propagating through a tapered fiber optic near-field probe restricts signal throughput.

The small signal available at the output end of the fiber optic requires a very sensitive detection method. The demands on the detection system become even more extreme when the optical signal is to be time resolved; signal-to-noise is effectively reduced proportionally by the number of time bins in which photons are to be resolved. For time resolving semiconductor luminescence from our GaInP/GaAs DHS samples in the near-field the method of time correlated single photon counting (TCSPC) is required.

The principle of TCSPC is based on counting photons from the leading edge of an electrical trigger signal. The concept works as follows: a photon is incident on a detector, the detector produces a short electrical pulse (\( \approx 30 \text{ ns} \)) indicating
detection of the photon. The electrical pulse is routed to TCSPC hardware. For a full description of the TCSPC hardware and laser system used in this dissertation see Chapter 8. Only the leading edge of the electrical pulse is used to trigger the time bin into which the photon is counted. The temporal width of the time bins, $\Delta t$, is tens of picoseconds and much less than the width of the triggering electrical pulse. Because the timing electronics are triggered from the leading edge of the electrical pulse, time resolution is not limited by the instrument response FWHM, but rather the timing jitter associated with the reproducibility of the leading edge. Timing jitter here refers to the uncertainty in the time between when the photon is detected to when the photon is counted. While the FWHM of detection electronics may be as short as a few nanoseconds, the timing jitter to reproduce the nanosecond electronic response is typically 10-100 picoseconds. As photons arrive, they are sorted into their corresponding time bins. A record for the number of counts in each time bin is maintained and at the end of the signal acquisition a histogram of all photon records from many laser shots is displayed reproducing the original waveform, see Fig. 7.8.

Instrumentation used in TCSPC is only capable of counting no more than one photon per excitation laser pulse. This property is due to the nature of the counting electronics; measuring the time between when a photon is counted and when the sample was excited. Typically, the arrival time of luminescence photons is done by starting a clock when the laser is pulsed and stopping the clock when the photon is detected, see for example Ref. [65]. In contrast, the implementation of TCSPC used in this work starts the clock when a photon is detected and stops the clock when the next laser pulse occurs. The arrival time of the photon relative to the laser pulse that originated it is determined from the repetition rate of the laser and is depicted in Fig. 7.9. This method of timing is called reverse start-stop. Due to the nature of the reverse start-stop operation it is critically important to have a low jitter laser pulse train. If the time between two consecutive laser shots is unstable, timing errors will result in the arrival times of detected photons. Additionally, TCSPC electronics
only have one clock and can therefore only count at most one photon per laser shot. It is possible to record multiple photons per laser shot, however, such systems require multiple detectors.

Thus far it has been assumed that TCSPC can respond to all photon events provided photon count rate does not exceed one photon per excitation laser pulse. Statistically, however, there is a greater probability of counting the single photon per laser pulse at early times in the decay process where there are more photons to

Figure 7.8: The concept of TCSPC. No more than one photon per excitation/emission cycle is counted.
be counted. If additional photons are present on the detector after the first photon, they will go undetected and the tail of the decay will be unresolved. The measured waveform will become distorted and fail to reproduce the intended original waveform. This obstacle is overcome by reducing the photon count rate at the detector to no more than one photon per 20-100 laser pulses [93]. This empirical 1-5% rule is the margin determined by TCSPC hardware manufacturers needed to ensure single photon statistics and undistorted waveform measurement.

An additional consideration for TCSPC is the maximum count rate the electronics can handle. After the electronics record a photon event, a recovery time is needed before the electronics are available to count again. The recovery time or dead time of the TCSPC hardware is determined by whichever component of the system takes the longest to recover from a counting event. A Picoquant TimeHarp 200 TCSPC card has an electronic dead time of 350 ns [94]. If, for example, the laser repetition rate is 1 GHz, the time between laser shots is 1 ns. Invoking the 1-5% rule, a photon can
be counted every 20-100 ns. However, this period of time is less than the dead time of TimeHarp 200 and so photon events will be missed and single photon statistics violated; the intended waveform will be distorted due to detector saturation. Issues associated with detector dead time are only applicable at high repetition rate (i.e. GHz) lasers. If GHz laser repetition rates must be used, photon count rates must be lowered such that the time between photon counting events is longer than then electronic dead time.

### 7.7 Integration of Photon Counting with Near-Field Microscopy

The concept of integrating TCSPC with near-field microscopy is not new. The aim of this section is to identify what has been done in the field of time resolved near-field microscopy and furthermore how the work of this dissertation fulfills a previously undemonstrated experimental void. A brief review of relevant literature will establish a context into which the microscopy work of subsequent sections can be appreciated.

In 1997 the first study of a single GaAs quantum wire was reported by Richter et. al. [95]. Near-field optical spectroscopy was used to resolve the spectrum of a single GaAs quantum wire as a function of lateral position from the wire. This was experimentally implemented in both the illumination and collection modes. A single line perpendicular to the GaAs quantum wire was scanned. At each point in the scan the spectrum was recorded and a two-dimensional image was created (1-dim. in space and 1-dim. in wavelength). Time resolving of the semiconductor luminescence was also performed, but only at a single point (on the wire). This early work did not result in a lifetime image or an AFM image, but was an important step in demonstrating the capabilities of NSOM. The same group has also detailed their
Chapter 7. Near-Field and Atomic Force Microscopy

low temperature near-field microscope in Ref. [96]. An important feature to note is the use of the shear-force feedback mechanism which is subject to artifacts in AFM and doesn’t allow contact mode topographic imaging [79].

Individual InGaAs/GaAs quantum dots have been resolved using NSOM and TCSPC [97]. Local illumination and local collection done through the near-field probe have identified quantum dots and allowed single dot lifetime measurements. Lifetime measurements are made only on the quantum dots themselves. In this work, the local-local measurement scheme has been used to create a time-integrated intensity image, but no simultaneously obtained AFM measurement was made. The collection efficiency of photoluminescence was improved through the use of a double-tapered near-field probe and allowed examination under weak excitation [98]. Time-integrated optical intensity images were produced with simultaneous AFM. However, the AFM was produced using shear-force feedback.

Simultaneous topographic and near-field optical imaging has been done to produce AFM and fluorescence lifetime images (FLIM) images [99]. In this work, a modified AFM probe is used to create a high local electric field enhancement under laser illumination, which provides a strong electric field interaction between the AFM tip and the fluorescencing molecule under the tip. A fluorescencing nanosphere was imaged to simultaneously produce AFM, time-integrated optical intensity and FLIM contrast images. Illumination and collection were done through a far-field microscope objective, while the near-field probe is placed beneath the nanosphere. The authors of Ref. [99] found a change in lifetime of 66% due to the presence of the AFM probe indicating that a FLIM produced with an NSOM is a lifetime contrast image (further explanation can be found section 7.9). This work is in close comparison to that of Kwak et. al. who have also produced lifetime contrast images in the near-field [100]. However, the experimental implementation is different in that Kwak and co-workers used a local illumination and global collection mode of NSOM
Chapter 7. Near-Field and Atomic Force Microscopy

operation. The shear-force feedback mechanism was used and a simultaneous AFM image was not produced.

The simultaneous AFM/FLIM work of this dissertation and detailed in subsequent sections is performed in a mode of operation not previously accessed; global illumination and local collection using normal force feedback. Local collection when combined with the normal force feedback mechanism provides two advantages: 1) and AFM image that is insensitive to artifacts can be produced and 2) AFM deflection measurements and photon counting are done in the same area and at the same time. This mode of operation is in contrast to previous simultaneous AFM/FLIM image methods which are tip enhanced and only possible with either thin films or single molecules; the sample can be measured with AFM from above and luminescence collected from below. Single molecules placed on an opaque substrate cannot be simultaneously imaged with AFM and FLIM using the tip enhanced configuration.

7.8 Fluorescence Lifetime Imaging: Understanding the effects of surface defects in GaInP/GaAs DHS

To better understand the effect of surface defects on carrier recombination a FLIM is created. Using a modified NSOM interfaced to TCSPC hardware (see Chapter 8) we have the unique capability to spatially characterize defects and assess surface quality in addition to temporally characterizing recombination mechanisms. A FLIM using a 500 nm aperture cantilevered NSOM probe reveals a reduction in lifetime by an order of magnitude in a 10 by 10 pixel image over a 73 by 73 µm area encompassing a surface defect, results are presented in Fig. 7.10a. Fitting of the luminescence decays from individual pixels is accomplished by means of the iterative algorithm
Figure 7.10: (a) Fluorescence lifetime image of a surface defect. The lifetime is reduced by a factor of 10 relative to the defect free regions. (b) An AFM image of the defect in (a). The AFM image was obtained in a second scan.

Outline in Appendix B. Each of the pixels in the FLIM has an acquisition time of 100 seconds. An 256 by 256 pixel AFM scan of the same surface defect shown in the FLIM is depicted in Fig. 7.10b. The spatial extent over which the reduced nonradiative lifetimes can be observed is found to be greater than the spatial extent of the surface defect and is due to carrier diffusion as explained in Section 6.6.

Figure 7.10 is the first FLIM produced after TCSPC integration. The FLIM luminescence signal is collected in a primary scan while the AFM data (non-contact mode) is obtained in a second scan. The dual scan nature of the measurement is done to avoid what is found to be a significant source of noise in our NSOM work; the AFM feedback laser. The reported issues here seek to educate future users of NSOM to the experimental obstacles encountered with normal force feedback. The wavelength of the CW AFM feedback laser for this work is 667 nm (measured with...
Chapter 7. Near-Field and Atomic Force Microscopy

an Ocean Optics HR2000). The laser wavelength is above the band gap of our GaAs samples. As a result, residual laser light that misses reflection off the NSOM probe strikes the sample near the probe tip. Excitation in the sample occurs from the CW feedback laser and produces substantial background luminescence at the detector. These background counts are equal to or greater than the luminescence produced from the pulsed excitation source making it impossible to time resolve the optical signature due to insufficient signal to noise. As a temporary fix, the feedback laser is turned off during scanning and data collection of the FLIM in Fig. 7.10a. This has the unwanted consequence of disabling the PID feedback for AFM imagery because no oscillating feedback signal is present on the PSD. The lack of signal is interpreted in the control electronics as an impact of the sample with the probe tip (prohibited in non-contact tuning). Control electronics attempt to correct this by moving the piezo scanner to bottom of its z-axis translation increasing the tip to sample distance in an attempt to pull the tip off the sample and restore the oscillatory feedback signal. However, the feedback laser is off during the data collection of the FLIM, therefore, the z-axis position remains pegged at the minimum of its translation range (3.5 \mu m). Hence, the tip to sample separation for the FLIM of Fig. 7.10a is not in the near-field, but on the order of 3.5 \mu m.

All subsequent FLIMs produced with our AFM-NSOM system have the PID feedback enabled and measurements are performed in the near-field. Spurious sample excitation issues associated with the AFM feedback laser are eliminated by changing the feedback laser wavelength to 980 nm, i.e. above the band gap of GaAs.

The dual scan of Fig. 7.10 means an alternative method could be used to arrive at the same information. A far-field lifetime measurement in tandem with a separate AFM scan would yield the image. However, the next section will show complimentary simultaneously acquired AFM/FLIM images. Simultaneous images allow conclusions to be made beyond those of a far-field lifetime and separate AFM image.
For success in laser cooling of semiconductors, it is critical to avoid surface defects. The enhanced nonradiative recombination on surface defects shown in Fig. 7.10 will lower the external quantum efficiency. It is equally important to avoid illuminating within $\approx 50 \, \mu\text{m}$ of a surface defect. Carriers can diffuse to the defect and experience rapid nonradiative recombination. Diffusion and recombination to an unpassivated GaAs surface is discussed in the following section.

### 7.9 Fluorescence Lifetime Imaging: Unpassivated GaAs Edge

Further characterization of GaInP/GaAs heterostructures is performed by analyzing the edge of a etched wafer. The lift-off and bonding process done for semiconductor external quantum efficiency experiments is preceded by sample etching [63]. An array of mesas is etched onto the semiconductor surface, see Fig. 7.11.

![Figure 7.11: Dots arrayed on a sample surface. The mesa structure is shown from the top and side views. The circumference of the GaAs layer is unpassivated.](image)

The AlAs layer in the mesa structure is used as a release layer, when etched away
Chapter 7. Near-Field and Atomic Force Microscopy

the GaInP/GaAs double heterostructure is lifted off the substrate. Lifted-off dots are then van der Waal bonded to a ZnS or ZnSe dome lens to facilitate light extraction. However, the measurements reported here take place prior to lift-off and bonding to avoid the possibility of processed induced lifetime degradation (see Section 7.12). A point of particular importance in Fig. 7.11 is the unpassivated GaAs circumference of each dot. Producing a FLIM in the region outlined by the black box reveals the effects of an unpassivated edge.

To gain insight into the role of edge effects on semiconductor fluorescence lifetime, we simultaneously produce a high resolution (50 x 50 pixel) FLIM near the edge of a GaAs double heterostructure, Fig. 7.12d and an AFM surface characterization Fig. 7.12a. The color map of the FLIM displays \( A' = 1/\tau \). Black indicates low nonradiative recombination and increasing red color indicates an increasing nonradiative recombination coefficient. The integrated PL (or brightness) is show in Fig. 7.12b. The brightest positions of the integrated intensity image are noted as those farthest from the edge. Uniform optical intensities and lifetimes are observed in directions parallel to the sample edge. A gradual roll-off in the optical intensity and an increase in the nonradiative recombination coefficient are seen as the near-field probe approaches the unpassivated edge. However, no luminescence is recorded above the background noise for a distance of \( \approx 8 \) microns from the edge. In this region recombination is dominated by a carrier diffusion and subsequent surface recombination, detection of radiative recombination is below our experimental noise. An optical image of the probed area is displayed with 100x magnification in Fig. 7.12c. The green box indicates the approximate region over which the scan is performed. Lifetime and intensity are collected simultaneous with AFM imagery, which assures a direct, one-to-one correspondence with surface morphology. The AFM image reveals the mesa height to be 3 \( \mu \)m and shows the edge to be at a position of \( \approx 22 \) \( \mu \)m.

The FLIM is produced using collection mode NSOM (global illumination with
Figure 7.12: (a) An AFM, (b) integrated intensity image, (c) 100x optical image and (d) FLIM of a GaAs double heterostructure. Green box in (c) depicts the approximate scan area. Dashed lines in (b) and (d) indicate the position of the sample edge as obtained from the simultaneously acquired AFM image (a).

The near-field probe is serving two functions. The tip is oscillating at 60-80 kHz for non-contact tuning AFM image creation while simultaneously collecting luminescence from the region directly beneath aperture. This method of data collection guarantees that not only the AFM and FLIM images are created at the same time but ensures a spatial overlap as well. The FLIM presented in the previous section was also produced in this mode but with the disadvantage of the PID system being disabled. By swapping the AFM feedback laser to a 980 nm wavelength, several advantages have been added: 1) the luminescence is now recorded in the near-field
Chapter 7. Near-Field and Atomic Force Microscopy

and as a consequence the spot size is now limited by the aperture (500 nm). While this is not below the far-field diffraction limit, it does allow simultaneity with AFM. 2) signal to noise improves dramatically by a factor of 10 (compared to the data of Fig. 7.10). This allows a 10 (instead of 100) second pixel dwell time. More pixels can be measured within a scan time to give a more detailed image. 3) The FLIM and AFM images are produced from one scan. By simultaneously measuring luminescence and surface topography a direct correlation can be made between features found in both images. Dual image collection is also advantageous because scanner reproducibility is uncertain with multiple scans.

The luminescence data of Fig. 7.12 is fit using the routine detailed in Appendix C. However, there are a few important modifications. Bian et. al. have shown that recording a decay in the near-field can significantly change the emission characteristics that would be observed in the far-field [101]. A reduction in lifetime by nearly a factor of two was noted in the measured lifetime as a function of tip to sample lateral separation. This has been attributed to nonradiative energy transfer from the excited molecule to the aluminum coating of the NSOM tip [101, 102, 103]. In the words of the authors [101], ”...a perturbation to the radiative rate can hardly be avoided.” This effect makes it nearly impossible to extract a correct lifetime from a decay measured in the near-field. The probe to sample distance is, however, constant throughout the scan and in turn makes the lifetime distortion constant over the scan as well. Thus a FLIM produced using NSOM is a lifetime contrast image and does not measure the absolute lifetime unless care has been taken to account for the near-field probe distortion [100, 99].

Because global excitation is identical for every pixel, changes in the peak of the measured luminescence decay are proportional to $BN_o$ in the measurement spot, and are due to recombination at defect sites located between where the sample is excited and where the luminescence is measured. This situation was depicted in
Fig. 6.7. Because of the inability to illuminate the area directly beneath the near-field probe, there is a small spatial offset between where the sample is excited and where the luminescence is collected. This means that the density $N_o$ created at the point of excitation is different than the $N_o$ at the point of measurement. The doping concentration, $B N_d$, is assumed to be constant over the small scan ranges in use here and only changes in nonradiative recombination are therefore measurable. This assumption is reasonable given that doping densities typically do not vary by more than a few percent across a wafer ($\approx$ inches) [74, 75]. For fitting, the variables $A'$ and $\eta_e BN_o$ are adjusted on the first pixel only. The doping density is fixed such that $N_d/N_o = 0.1$ (based on an initial carrier concentration of $1e16 \text{ cm}^{-3}$ and a background doping concentration of $1e15 \text{ cm}^{-3}$, see appendix C). The background doping density is assumed to be constant for the length scales involved in producing a FLIM ($< 100 \mu \text{m}$). In all subsequent pixels $\eta_e BN_o$ and $N_d/N_o$ are not varied in the fitting algorithm, but are scaled with the measured peak value. For example, if in the first pixel the peak count value of the unnormalized decay was 1000 counts, the data was fit with $A' = 0.01 \text{ ns}^{-1}$, $\eta_e BN_o = 1e-3 \text{ ns}^{-1}$ and $N_d/N_o = 0.1$. For the second pixel, the peak count value of the unnormalized decay was 500 counts, $\eta_e BN_o = 0.5 \times 1e-3 \text{ ns}^{-1}$, $N_d/N_o = 0.1/0.5$ and $A'$ is adjusted by the fitting routine to fit the data. Therefore, $\eta_e BN_o$ and $N_d/N_o$ are scaled while $A'$ is varied. The noise floor parameter, $\Delta$, in Eq. 6.25 is varied for every pixel. While there is some latitude in choosing the ratio $N_d/N_o$, it has little consequence. The effect of starting with a different value of $N_d/N_o$ only changes the scaling of the $A'$ coefficient, but the contrast remains the same. Areas having high nonradiative recombination will still have high nonradiative recombination relative to areas having minimal nonradiative recombination. The changing of the $A'$ scale is inconsequential as a FLIM produced in the near-field is a lifetime contrast image to start with [100].

Regions designated with black in the FLIM of Fig. 7.12 represent only an upper limit on $A'$. Nonradiative recombination is too weak to resolve in the window of
Chapter 7. Near-Field and Atomic Force Microscopy

time over which the decay has appreciable signal to noise. Increased signal to noise in the tail of these decays would be needed to resolve the nonradiative lifetime. Such signal to noise is, at present, inaccessible with our NSOM system. The window of time over which fitting is performed is arrived at as follows: a computation over the last 25 data points in the dataset is regarded as the noise floor. Next, the algorithm discerns the fitting window by starting from the peak and progressing to later times until a value equal to the noise floor is reached. All data points contained within the SNR > 1 window are used for fitting. This procedure not only ensures the data over which the fitting is performed contains meaningful signal, but also allows freedom to accommodate pixels with low signal by adapting the number of points over which to fit.

By taking the sum of all the lifetime measurements in the columns of Fig. 7.12d one can produce an average lifetime as a function of position across the sample edge. This spatially averaged lifetime is shown in Fig. 7.13a and again in Fig. 7.13b on a log scale. The error bars represent measured uncertainty (standard deviation of the mean) for the 50 data points in each column of the FLIM. The sample edge from the AFM scan is denoted as a red dotted line at 22 µm on the horizontal position axis. It is important to note that the fluorescence lifetime is not discernible for distances ≈ 8 µm from the sample edge. This effect is due to the presence of nonradiative surface recombination at the unpassivated sample edge. A gradual increase in the lifetime can be observed with a progression away from the sample edge. The measurement of the nonradiative lifetime in regions far from the sample edge becomes noisy due to the diminishing effect of the unpassivated surface at 22 µm.

The total time integrated PL as a function of position across the sample is shown in Fig. 7.14a on a semilog scale. A dramatic reduction in the recorded intensity is noted as the measurement position approaches the unpassivated sample surface. The reduction is due to carrier diffusion from the bulk to the sample edge. Rapid
Chapter 7. Near-Field and Atomic Force Microscopy

Figure 7.13: (a) Average nonradiative lifetime as a function of position approaching an unpassivated sample edge and (b) on a log scale. Carrier recombination < 8 µm from the sample edge experiences very fast (<10 ns) nonradiative decay.

nonradiative recombination takes place at the surface. The process is described by the transport equation in one spatial dimension following pulsed light excitation:

\[
\frac{dN}{dt} = -A'N + D_e \frac{\partial^2 N}{\partial x^2}
\]

(7.3)

and the solution is [73]

\[
N(x, t) = \frac{N_a}{\sqrt{4\pi D_e t}} e^{-\frac{x^2}{4D_e t}} e^{-A't}.
\]

(7.4)

The number of electrons (or holes) per unit area is \( N_a \) and \( D_e \) is the electron diffusion coefficient. The solution is plotted in Fig. 7.14b as a function of time and space. The origin of the plot represents the location in space and time of carriers created by an impulse excitation. As time progresses, the excess carriers diffuse out along the spatial dimension. The calculation is performed for GaAs: \( A' = 400 \text{ ns}^{-1} \) and \( D_e = 260 \text{ cm}^2\text{s}^{-1} \). Figure 7.14c shows the time integrated PL (\( \propto BNN_d \)) as a function of position approaching a sample edge at zero. The inset depicts the PL roll-off over
Chapter 7. Near-Field and Atomic Force Microscopy

the same length scale as Fig. 7.14a. The diminished signal is due to carrier diffusion to the sample edge and subsequent rapid surface recombination.

Figure 7.14c shows the PL roll-off for a distance of 200 µm from the sample edge. For laser cooling, it is important to avoid regions ≈ 50 µm from the sample edge. Excitation within this distance will result in external quantum efficiency degradation due to surface recombination.

Figure 7.14: (a) Measured PL intensity approaching an unpassivated surface in GaAs (positioned at 22 µm). (b) Calculated transient carrier diffusion in GaAs following impulse excitation. (c) Calculated PL intensity approaching an unpassivated GaAs edge. Inset shows region highlighted in red. Intensity reduction is due to diffusion and surface recombination.
Chapter 7. Near-Field and Atomic Force Microscopy

7.10 Fluorescence Lifetime Imaging: Subsurface Defect at GaAs Edge

Additional characterization of etched GaAs dots is done by scanning a second area of the same dot. The second scan is also performed at the dot edge in an attempt to reproduce the effects seen in Fig. 7.12. The area over which the measurements are taken is shown in Fig. 7.15 with a black box. The conditions to create the second AFM/FLIM scan are the same as those of the previous section.

Figure 7.15: Dots arrayed on a sample surface. Area of second FLIM is denoted by the black box.

Results of the scan are presented in Fig. 7.16. The AFM image of Fig. 7.16a shows the edge of the sample spanning a diagonal line through the scan area. The height of the mesa top is again measured to be 3 μm. The green box in the optical image shows the region over which the scan is preformed, Fig. 7.16c. This region is near the AFM/FLIM scan of the previous section but does not overlap with it. A feature worthy of attention is the large mark seen in the middle of the both the intensity and lifetime images, see Fig. 7.16b and d respectively. This large region represents an area of severely degraded optical performance; it is both dim and
dominated by nonradiative recombination. Nonradiative lifetimes in this area are 10x shorter than regions having no defects and being far from the unpassivated sample edge. A critical point is that the AFM scan of the same area shows no corresponding spatial surface features. Because all three images are produced at the same time and in the same scan, unambiguous spatial comparison is possible. The surface of the sample is featureless with the exception of the sample edge. Unlike Fig. 7.10, there is structure in the optical images not seen in the AFM. The luminescence signature originates from the GaAs, but since it does not mirror the AFM characterized GaInP surface passivation layer, the optical defect is subsurface.

The $\approx 8 \, \mu m$ offset between the physical and optical images is reproduced. Carriers generated within this $8 \, \mu m$ window diffuse to the unpassivated surface edge and rapidly recombine nonradiatively. These unpassivated edges would produce heating in a laser cooling experiment, allowing optical excitation energy to be converted into heat through the nonradiative phonon generation process. This issue can be avoided by selecting regions far from the unpassivated sample edge ($\approx 50 \, \mu m$) for use in laser cooling experiments. The next section will detail AFM/FLIM analysis done in the bulk semiconductor thereby avoiding the undesirable effects of surface recombination.

### 7.11 Fluorescence Lifetime Imaging: Bulk GaAs

The previous sections showed a 10x reduction in the nonradiative lifetime near unpassivated edges. A subsurface defect was also presented displaying enhanced nonradiative recombination. The aim of the current section is to study the optical properties of bulk GaAs passivated above and below by GaInP. By moving away from the edges, bulk behavior is observed and the optical performance of a candidate GaAs laser cooling sample is evaluated. Experiments conducted to measure the external
quantum efficiency do so in the bulk semiconductor [63]. It is therefore important to quantify not only the nonradiative lifetime but also optical uniformity in the bulk. A perfect crystalline lattice would be repeated in all directions [72]. In such a case, the optical performance of the semiconductor would be uniform. The AFM/FLIM measurement is well suited to assess optical performance of bulk semiconductor. AFM can confirm or deny the presence of surface defects while simultaneous FLIM allows commentary on optical processes.
Chapter 7. Near-Field and Atomic Force Microscopy

A 50 by 50 pixel AFM/FLIM is produced in an area of bulk semiconductor. The area chosen for the scan is free from visible defects (under 100x magnification) ensuring surface uniformity and is far from any sample edge or surface scratch (>100 µm). The majority of the NREL MF438 sample surface is free from surface defects and visible scratches. For positions far from a sample edge surface recombination can be neglected and bulk carrier dynamics observed. Using the same GaInP/GaAs double heterostructure we produce a FLIM on the sample bulk. A representative area of the GaAs double heterostructure surface is shown in Fig. 7.17c. The region outlined by the green box (70 µm per side) in the optical image denotes the approximate area over which the scan is performed. Several surface defects are visible in the optical image; however, none reside in the area scanned. The lack of defects is confirmed in the AFM image seen in Fig. 7.17a. Only a small protrusion is seen in the extreme upper right hand corner of the scan. The remainder of the scanned area is featureless, displaying a surface roughness of about 100 nm. The commensurate intensity and lifetime (displayed as 1/Lifetime) images are shown in Fig. 7.17b and d respectively. The data for all three images was collected simultaneously; therefore, there is a direct one-to-one correspondence between pixels in each image. The intensity and lifetime images show an unexpectedly rich structure. Note the black trenches in the intensity image of Fig. 7.17b. These regions indicate almost no semiconductor luminescence. From the FLIM it is clear that these same regions exhibit an enhancement in the nonradiative recombination coefficient. It is also important to note that the structure observed in the intensity and lifetime is not spatially correlated with anything found in the AFM image. The optical nonuniformity of Fig. 7.17b and d is indeed surprising. Structure in these images is again attributed to subsurface defects and features in the GaAs. The orientation of the subsurface trench network can, however, be spatially correlated to lines connecting surface defects found in the optical image of Fig. 7.17c. By superimposing the intensity (or lifetime) image onto the optical image in the scan area and connecting the dots representing the surrounding
surface defects, we find the resulting lines qualitatively overlap with the structure found in the intensity (or lifetime) image, see Fig. 7.17c white lines. It is expected that increased strain in these lines begins and ends at surface defects resulting in increased nonradiative recombination.

Figure 7.17: (a) AFM image on bulk GaAs passivated with 30 nm of GaInP. No surface features were found. (b) Integrated intensity image from the luminescence collect during the AFM of (a). (c) Optical image showing sample surface. Surrounding surface defects are highlighted with purple circles. Area of scan is denoted by the green box. White lines connecting surface defects overlap with features of the superimposed intensity (or lifetime) image. (d) Lifetime image from the luminescence collected during the AFM measurement of (a).
Further confirmation of the intricate optical signature is obtained by conducting a second FLIM in the bulk on the same GaAs double heterostructure. The second FLIM is produced in a different area of the same sample chosen to lie between two prominent surface defects, see Fig. 7.18c. The green box denotes the approximate scan area and surface defects are highlighted by purple circles. The AFM of Fig. 7.18a again shows a smooth surface, devoid of any surface structure or feature. The lifetime image, Fig. 7.18d and its complimentary intensity image, Fig. 7.18b, display significant structure. A feature to note in the two optical images is the trench cutting a diagonal from the upper right corner to the mid-upper left edge of the scan area. Short nonradiative lifetimes and low luminescence counts are complimentary in the intensity and lifetime images. The diagonal trench is not spatially correlated with any morphological feature contained in the scan area, however, a line connecting surface defects just outside the scan area traces out the trench. This important result confirms two things: 1) surface defects can serve as beginning or ending points for subsurface boundaries and 2) the subsurface network of defects exists in at least two separate areas, those of 7.17 and Figs. 7.18. The surface defects create strain between the defects and beneath the surface. A question that now presents itself is, how widespread is the network of subsurface defects? Is it localized to areas only having surrounding surface defects or is it more universal? Section 7.12 will answer these questions.

A final point prior to proceeding will be made in regards to the mode of NSOM operation used to create Fig. 7.18. Thus far AFM/FLIM images have been produced using the NSOM collection mode; the sample is excited globally from above and luminescence is collected locally through the near-field fiber probe. With this configuration, carriers are generated in an area surrounding the probe tip. Luminescence that is detected originated from carriers that experienced a diffusion from areas surrounding the tip to areas directly beneath the tip. The data of Fig. 7.18 was collected using illumination mode NSOM; that is, excitation light illuminates
Figure 7.18: (a) An AFM, (b) intensity image, (c) optical image and (d) FLIM of a bulk GaAs double heterostructure. No surface structure was found in the AFM image. FLIM and intensity images show a trench with poor optical performance that spatially overlaps with the white line connecting the surface defects (highlighted with purple circles) of (c). Scan area of the AFM/FLIM is shown with a green box. Data in (b) and (d) was collected in NSOM illumination mode.

the sample through the tip and subsequent luminescence is gathered with a far-field microscope. This experimental configuration is achieved by exchanging the positions of the excitation laser and APD detector of Fig. 7.4. In this configuration, carriers are generated directly beneath the near-field probe but must diffuse outward before
Chapter 7. Near-Field and Atomic Force Microscopy

experiencing radiative recombination and detection. This configuration is the reverse of collection mode NSOM and provides confidence that the structures measured in Figs. 7.17 and 7.18 is not an experimental artifact. Additionally, the area of Fig. 7.18 was scanned a second time and the optical signature was reproduced with the exception of a spatial offset originating from the scanners inability to reset to the original position.

7.12 Fluorescence Microscopy (Far-Field)

While simultaneous AFM and FLIM imaging is successful and essential for mapping out integrated luminescence, surface morphology and the nonradiative recombination coefficient, they do so at the expense of time. A typical AFM/FLIM scan time takes greater than 7 hours and generates several gigabytes of data. Such scans contain 2500 individual lifetime measurements and can take anywhere from 12 to 24 hours to process and fit. This section details the development of a method capable of identifying regions of poor optical performance in candidate GaAs samples at video refresh rates. The characterization allows rapid, real-time screening of semiconductor luminescence.

The experimental apparatus for characterizing luminescence is shown in Fig. 7.19. An unfocused 30 mW CW laser (LambdaPro: $\lambda=532$ nm) is used to uniformly excite the sample at grazing incidence. A far-field 10x microscope objective is used to collect luminescence and image it onto a CCD. The luminescence passes through a bandpass filter (Omega Optical No. 800-900XRD), 800-900 nm, to eliminate any stray laser light or luminescence from the GaInP passivation layer ($\lambda \approx 670$ nm). The CCD is interfaced to a video capture card and the luminescence image is displayed on a computer screen with video refresh rates. Semiconductor samples are mounted on an $xyz$ translation stage allowing interrogation of the entire sample surface.
Chapter 7. Near-Field and Atomic Force Microscopy

Figure 7.19: Apparatus for fluorescence microscopy. The CCD image can be displayed at video refresh rates.

7.12.1 Diffraction Limited Fluorescence Imaging: Subsurface Defects

An example of a fluorescence image from sample NREL MF438 is shown in Fig. 7.20b. Brighter areas indicate more luminescence. The optical image from the same area is displayed in Fig. 7.20a. The generation of the optical image is done by removing the bandpass filter and replacing the laser with a white light source. Sample position is maintained during the switch from the optical to the fluorescence mode of operation and facilitates direct comparison of the area shown in the two images. The optical image has no discernible structure with the exception of a two surface defects in the left center portion of the image (highlighted with red circles). Unlike the optical image, the fluorescence image reveals substantial structure. The subsurface network discovered in the near-field lifetime images of Figs. 7.17 and 7.18 is reproduced in the fluorescence image. No corresponding AFM image is generated with Fig. 7.20
but the work of the previous sections strongly suggests the luminescence structure is uncorrelated with surface topography. Two types of defects are revealed in the fluorescence image: 1) point defects which randomly populate the image and are not necessarily correlated with surface defects and 2) line defects that cross each other and span hundreds of microns. It is important to note that not all line defects begin or end with surface defects. Only two surface defects can be seen in the optical image, however, many more line defects are discernible in the fluorescence image. In surveying the sample surface, many of the line defects traverse the entire sample (approximately 3 x 6 mm) terminating at the edge. Samples exhibiting a preponderance of defects, such as those of Fig. 7.20b, are rejected for laser cooling attempts.

Spatial luminescence interrogation through fluorescence imaging is performed on a second sample, NREL grown MH701. This sample is identical in structure to MF438 with two notable exceptions: 1) MH701 has no AlAs release layer and 2) the top level passivation layer is 750 nm thick as opposed to the 30 nm thickness of MF438, see Fig. 7.5. A picture of the sample surface under white light illumination is depicted in Fig. 7.21a. It can be seen from the figure that no surface blemishes or defects are present. The area shown is an arbitrary but representative sampling of the sample provided. Figure 7.21b is a fluorescence picture of the same area. A prominent dark spot is noted in the middle of the image. This subsurface point defect produces almost no luminescence and is similar to case (c) of Fig. 6.7. Defects of this nature are rare on the MH701 sample and an effort was made to locate even one defect for depiction in this dissertation. No subsurface trench defects were found anywhere on the sample.
Chapter 7. Near-Field and Atomic Force Microscopy

Figure 7.20: (a) Optical image with surface defects (highlighted in red circles). (b) Fluorescence image of a candidate GaAs laser cooling sample. The dark lines in the fluorescence image reveal degraded optical performance despite a predominantly featureless surface.

Figure 7.21: (a) Optical image of sample MH701. (b) Fluorescence image of the same area. The dark spot in the fluorescence image reveals degraded optical performance despite the uniform surface.
7.12.2 Far-field Fluorescence Imaging: GaInP Passivation Layer

Additional optical evaluation is done by changing the bandpass filter in front of the CCD to 600-700 nm (Omega Optical No. 600-700XRD). An exchange of bandpass filters allows selective luminescence imaging; GaInP passivation layer luminescence is detected while GaAs luminescence is filtered out. The selectivity allows the optical defects to be bracketed to either the GaAs or GaInP layers or both. An example of luminescence images from each layer originating from the same area of NREL sample MH700 is presented in Fig. 7.22. GaAs fluorescence is shown in (a) and GaInP in (b). The green circle in (a) marks the position of a subsurface point defect. The green circle in (b) is in the same position as (a), but but shows no evidence of degraded optical performance. This defect is therefore located in the GaAs layer only. In a complimentary fashion, the dark streaks in the upper left corner of the GaInP fluorescence picture do not appear in the corresponding GaAs image. The dark streaks in the GaInP image are from surface scratches, but are not deep enough to affect GaAs passivation. The majority of point defects found in the GaAs luminescence images of MH700 and MH701 samples also existed in the GaInP. The point defect shown in Fig. 7.22 is an interesting exception that serves to demonstrate the selective capabilities of the fluorescence imaging technique. Unfortunately, the GaInP passivation layer of MF438 is too thin to produce a detectable signal and the creation of GaInP fluorescence images is not possible with this sample.

7.12.3 Far-field Fluorescence Imaging: Processing Induced Defects

The application of fluorescence imaging is further extended by assessing defects induced during the liftoff and bonding process. A detailed description of the procedure
Chapter 7. Near-Field and Atomic Force Microscopy

Figure 7.22: (a) GaAs and (b) GaInP fluorescence images of the same area. Green circle marks the position of a defect in GaAs, but not in GaInP. Surface scratches appear in GaInP fluorescence, but don’t affect the GaAs fluorescence image.

used in the van der Waal bonding process can be found in Refs. [104, 105]. The idea is that dots created in the etching process, such as those depicted in Fig. 7.15, are removed from the substrate by means of a selective solvent etch that dissolves the AlAs layer. Upon dissolving of the AlAs layer, the dots release from the substrate and are briefly suspended in the solvent prior drying and placement on a dome lens. Semiconductor dots analyzed in this section are bonded to a glass slide for convenience as shown in Fig. 7.19 [106].

Luminescence images of 24 individual dots are taken before liftoff and bonding. Dots are created from the MH701 double heterostructure. Of the 24 dots, 13 had surface defects to start. The surface of each dot is thoroughly inspected. If a dot has a defect, the luminescence image taken encompasses the defect. When multiple defects are present a sufficient number of images are taken to record the position of each defect. If no defect is present, an arbitrary region is selected for luminescence recording. The dots are carefully tracked throughout the liftoff and bonding pro-
cess to ensure correct identification needed for before and after comparison. Results are presented in Fig. 7.23. The luminescence picture before the liftoff and bonding process is shown in Fig. 7.23a and the same dot post-processing is Fig. 7.23b. An optical image of the sample surface is shown in Fig. 7.23c. The curvature of the dot is visible in all three images. The before luminescence image shows no subsurface defects and this particular dot has no surface defects either. However, after processing three distinct stages of defect progression are noted. The first stage, marked by the first black circle, shows a bubbling up of the sample surface, see Fig. 7.23c. The area outlined by first black circle is verified as bubbling up by changing the z-axis (perpendicular to the image plane) of the microscope objective and bringing the bubble top into focus. The same area in the luminescence image displays luminescence nonuniformity (noted by concentric light/dark rings), but no degradation (i.e. the area is still bright). The second stage of defect progression is marked by the second black circle in Figs. 7.23b and c. The optical image reveals this area to be bubbled up in a similar manner to the first stage of defect progression. However, upon inspection of the luminescence, a sharp drop in the integrated intensity can be detected with the dimmest portion of the luminescence coming from the highest point (middle) of the bubble. The degraded luminescence suggests damage has been done to the GaAs layer and nonradiative recombination has become enhanced as a result. The third and final stage of defect progression is identified by the third black circle. This most advanced stage of process induced defect creation results in visible surface damage. From Fig. 7.23c it is evident that the surface has been damaged by a "punch-through" defect. A characteristic "X" shape in the middle of the third circle is a tear with the pedals opening upward. The corresponding area in the luminescence image shows extensive optical degradation spanning nearly 50 µm. Punch-through defects are observed in 4 of the 24 processed samples.

The punch-through defects of Fig. 7.23c are likely caused by trapping of solvent or water between the GaAs/GaInP double heterostructure and the glass slide. When
light pressure is applied to the semiconductor to bond to the glass slide, pressure inside the bubble increases causing it to rupture. It is also worth noting that the bubbles can be manipulated. That is, pushing on the sample surface causes the position of the bubbles to change. An attempt is made to gradually apply pressure to the dot from one side to the other in hopes of pushing the bubbles to the sample edge and relieving pressure. This attempt is done by rolling a swab across the semiconductor surface. Ultimately, the technique proved ineffective as the swab is not smooth enough to uniformly apply pressure. It is likely that the bubbles can lead to a degraded external quantum efficiency due to less than perfect bonding. A bubble above the number 2 black circle in Fig. 7.23b appears brighter than the surrounding area. The brighter region is due to less luminescence escaping the semiconductor.
through the glass slide bonded beneath it and more emitting from the top. Total internal reflection is enhanced over a bubble due to a greater mismatch in the index of refraction. The nonuniform bond will also lead to a nonuniform external quantum efficiency.

7.13 Summary

This chapter began with a review of microscopy techniques. Of specific emphasis were atomic force and near-field microscopy. Surface defect morphology on GaInP/GaAs heterostructures was mapped out using AFM. A characterized defect was found to be 3.5 $\mu$m in height and $\approx$ 10 by 20 $\mu$m in spatial footprint. The integration of TCSPC to near-field microscopy allowed time resolving of photoluminescence from laser cooling semiconductor samples. Fluorescence lifetime images were presented illustrating:

1) a factor of 10 enhancement in nonradiative recombination on surface defects compared to defect-free measurements.

2) a dramatic reduction in nonradiative lifetime and PL approaching an unpas-sivated GaAs edge. This reduction is attributed to diffusion and surface recombina-tion. Excitation within 50 $\mu$m of the sample edge experiences enhanced nonradiative recombination due to the surface.

3) subsurface defects were identified using both illumination and collection mode NSOM with simultaneous AFM. An order of magnitude reduction in nonradiative lifetime was noted on these subsurface defects.

All three of these effects result in a lower or nonuniform external quantum efficiency, which can lead to excess heating and overcome potential net cooling.
Chapter 7. Near-Field and Atomic Force Microscopy

The end of the chapter detailed the development of a fluorescence microscopy characterization technique capable of real-time screening of semiconductor luminescence. The method was used to uncover process induced defects and bonding nonuniformity.
Chapter 8

AFM-NSOM System

8.1 Overview

The experimental setup used to acquire the fluorescence lifetime and atomic force images found elsewhere in this dissertation is detailed here. Where possible, manufacturers of the equipment used is noted. An equipment map is presented that shows what instruments are used and the communication pathways between them. The minimum timing resolution of the system is 622 ps and is dominated by the detector. Measured system instrument response functions are presented.

8.2 Instrumentation

A novel system capable of mapping the lifetime simultaneously with topography as a function of position on a semiconductor sample has been constructed, see Fig. 8.1. A FLIM is created by interfacing a commercial normal force feedback Atomic Force Microscope (Nanonics NSOM-100) to time-correlated single photon counting
Chapter 8. AFM-NSOM System

(TCSPC) hardware. The AFM is placed between upright and inverted confocal microscopes (Olympus) as seen in Fig. 8.2. The presence of these microscopes allows for excitation or collection of fluorescence from either above or below the sample. For the measurements reported in this dissertation, excitation is either globally from above using the upright microscope and luminescence is collected through the near-field probe or locally through the tip and luminescence is collected with the upright microscope. The excitation source is a diode laser (Picoquant PDL-405) producing 70 ps pulses at a wavelength of 405 nm and is driven by a Picoquant PDL 800-B laser diode driver. The average laser power at 80 MHz is 2.45 mW. The excitation strikes the sample in the vicinity of the near-field probe. For collection mode NSOM fluorescence and scattered laser light are collected in the near-field with a fiber coupled probe possessing an aperture of 500 nm. Collected light is then passed through the fiber optic and emitted inside a light tight box. Residual laser light is removed using a bandpass filter (Omega Optical XRD800-900); the luminescence is focused onto a thermo-electrically cooled avalanche photo diode (APD) (Perkin-Elmer SPCM-ACR-15) having a dark count rate of 50 ct/s. Photon counts from the APD are registered with a TCSPC card (Picoquant TimeHarp 200). Time correlation of the photons is achieved by means of a synchronization signal from the excitation laser driver. Spatial correlation of photon counts is made possible by accommodating a pixel marker signal found in the Data Translation (DT) controller of the AFM scanner. As a scan is being performed, the TCSPC card is recording time-tagged photons and spatial pixel markers. Recorded data is processed using custom MATLAB software upon completion of the scan.

The interface module, PID controller, DT controller, HP computer and AFM head comprise the commercial Nanonics NSOM 100 system. Every time the AFM scanner indexes, a pixel marker triggering signal from the DT controller is routed to a Stanford Research DG535 digital delay generator. This externally triggered DG535 serves three functions: 1) a 5 ms TTL pulse is sent to a timing interface box. For
Chapter 8. AFM-NSOM System

Figure 8.1: Instruments and communication pathways used in time-resolved simultaneous AFM and lifetime imaging.

this 5 ms window the timing interface box holds pin 1 of the 15 pin D-connector on the TimeHarp 200 at a TTL low state, disabling photon counts. The remainder of the time pin 1 is held at a TTL state enabling photon counting. Count gating is done to blank out periods where the scanner is transitioning from one pixel (or line) to the next. 2) a TTL pulse > 200 ns is sent to pin 5 of the 15 pin TimeHarp 200 D-connector. The TTL pulse serves as a pixel marker in the dataset. 3) a TTL pulse ≈ 1 second in width is sent to a National Instruments data acquisition pad (model number: DAQPad-6020E). At the start of the experiment the DAQ pad is continuously queried until the TTL high state is read and subsequently signals the start of data collection. The DAQ pad is interfaced to an IBM computer through a USB 2.0 port.
System instruments are controlled with the LabVIEW software platform. Custom LabVIEW programming allows for sequential instrument initialization and system timing. Embedded within the LabVIEW program is a MATLAB kernel that performs fitting to the measured fluorescence decays. Once the fitting has been performed the resulting lifetimes of the fits are rendered in LabVIEW or MATLAB as a function of position. As a rule of thumb, decay rates $1/10^{th}$ of the measured signal can be recovered through deconvolution of the measured decay with the instrument response function [94]. The minimum timing resolution of the system, $\Delta T_{\text{min}}$, is given by the square root of the sum of the squared timing jitters, $\delta$, for each component of the
Chapter 8. AFM-NSOM System

TCSPC system. That is:

$$\Delta T_{\text{min}} = \sqrt{\delta_{\text{Laser}}^2 + \delta_{\text{Card}}^2 + \delta_{\text{Detector}}^2 + \ldots}$$  \hspace{1cm} (8.1)

From the product manual the timing jitter for the laser and TCSPC card is 30 ps and 40 ps respectively. The timing jitter for the Perkin-Elmer SPCM detector series has been studied at length in Ref. [107]. The study revealed a timing jitter ranging from 190 to 620 ps using 12 different detectors. By using the worst case scenario number of 620 ps the minimum timing resolution of the TCSPC NSOM system is 622 ps. The minimum timing resolution of the system is dominated by the timing jitter of the detector. This means timing resolution can be improved by using lower jitter detector such as a Picoquant PDM series which gives timing resolution down to 50 ps [108]. The instrument response function of the PDM detector series has the additional advantage of the being insensitive to count rate. The dependence of the instrument response function on photon count rate for both Perkin-Elmer SPCM and Picoquant PDM detectors can be found in Ref. [109]. Instrument response and its dependence on count rate has been ignored during the fitting of luminescence decays for semiconductors in this work because the lifetime is much longer than the response function. It is important to note that the minimum timing resolution of 622 ps can only be achieved for measurement time of 164 ns. The minimum is because the TCSPC card has 4096 time bins and each bin has a $\delta t$ of 40 ps. The number of bins is fixed for every measurement. Extended time base measurements are made possible by increasing the $\delta t$ of the time bins. The production of FLIMs reported here is done with a $\delta t$ of 298 ps per bin. This gives a measurement time of 4096*298 ps = 1.22 $\mu$s. The maximum $\delta t$ of the TimeHarp 200 is 1.1909 ns which gives a maximum measurement window of 4.878 $\mu$s.

The laser diode driver for the PDL-405 has an internal oscillator capable of 80, 40, 20, 10 and 5 MHz repetition rates. The lifetimes of interest in semiconductors
Chapter 8. AFM-NSOM System

span hundreds of nanoseconds to microseconds. To accommodate such long decays, the repetition rate of the laser must be decreased. Lower repetition rates are accomplished by means of externally triggering the laser diode driver with an internally triggered DG535, see Fig. 8.1. Interfacing is done through a GPIB port to an IBM computer or the instrument front panel. The DG535 can provide TTL triggering pulses to the PDL 800-B from DC to 1 MHz. The 1 MHz setting is chosen to allow full carrier relaxation to occur in the semiconductor before the next excitation laser pulse arrives. The most significant consequence of using the DG535 is there is now an additional term under the square root of Eq. 8.1. The jitter of the DG535 exceeds that of the SPCM detector. Figure 8.3 shows the instrument response function (IRF) for the TCSPC system trigger off the internal oscillator of the PDL 800-B (a) and off the DG535 (b). The IRF for each measurement is produced by directing a small amount of laser light to the detector, no sample is present. The FWHM of the IRF for the DG535 trigger is nearly 3.5 times longer than the internally triggered PDL 800-B. Despite the reduced timing resolution, long lived decays encountered in with semiconductors can be recovered without the need for instrument response deconvolution. Decays on the order of the FWHM of the DG535 IRF (<10 ns) are not recoverable without deconvolution. However, if such short decays are to be measured, the repetition rate of the laser can be increased and the better timing resolution of the internally triggered PDL 800-B appreciated.

Spatial resolution of the system is limited by the aperture size of the near-field probe. A good compromise between spatial resolution and fluorescence throughput can be achieved with a 500 nm aperture for our mean luminescence wavelength of 870 nm. While this is not below the diffraction limit of \(\lambda/2\) it is sufficient to resolve the features detailed in this dissertation and still allow simultaneous AFM image creation.
Chapter 8. AFM-NSOM System

Figure 8.3: Instruments response function for the TCSPC system. Laser diode driver triggered (a) internally and (b) externally with a DG535.
Chapter 9

Summary and Future Work: Time-Resolved Microscopy

The experimental work of previous chapters revealed subsurface defects. While AFM analysis indicated no surface structure, temporally and spatially resolving photoluminescence showed a resulting enhancement in nonradiative recombination. The enhancement was attributed to the presence of defects in the GaAs layer. In the context of laser cooling, this can lead to sample heating and potentially overwhelm any cooling. This unique characterization was done using normal force feedback and collection mode NSOM; an experimental avenue not previously traveled. The primary advantage of this approach is direct spatial correlation between optical and morphological images. The mode of operation is insensitive to semiconductor carrier diffusion causing spatial blurring of the measurement spot size.

A new method for rapid screening of semiconductor samples was implemented using fluorescence microscopy. By assessing luminescence radiating from the GaAs layer over a wide field of view, regions of poor optical performance can be identified and avoided for cooling trials. Images features resulting from this characterization
Chapter 9. Summary and Future Work: Time-Resolved Microscopy

technique support those produced in the near-field.

The primary drawback to near-field microscopy is the small photon throughput. Thus, future work should therefore center on increasing throughput. Two possible methods for doing so include:

1) Double angle tapered near-field probes have been shown to increase transmission by a factor of 10 compared to single angle tapers [98]. If such a probe could be integrated into the NSOM used in this dissertation work, image collection time would reduced from hours to minutes.

2) Tuning fork feedback for tip to sample stabilization should be considered provided artifacts are accounted for. With the current experimental configuration, the normal force feedback laser is always present in detection. While this can be abated with optical filters, it is never zero. Tuning fork feedback would eliminate spurious optical sources.
Appendices
Appendix A

Solutions to the carrier recombination equation

The time dependent carrier density, $N$, can be obtained analytically under several approximations of the carrier recombination Eq. 6.15:

$$\left. \frac{dN}{dt} \right|_{\text{decay}} = -(A + \eta_e B N_d)N - \eta_e B N^2 - C N^3.$$  \hspace{1cm} (A.1)

Here the carrier density is assumed to be a function of time only. The simplest approximation allows only nonradiative recombination

$$\left. \frac{dN}{dt} \right|_{\text{decay}} = -(A + \eta_e B N_d)N = -A'N.$$  \hspace{1cm} (A.2)

This ordinary differential equation belongs to a special class of problems known as variable separable. That is, dependent and independent variables can be separated to opposite sides of the equation

$$\frac{dN}{-A'N} = dt.$$  \hspace{1cm} (A.3)
Appendix A. Solutions to the carrier recombination equation

allowing direct integration

\[
\frac{1}{-A'} \int \frac{dN}{N} = \int dt \quad \text{(A.4)}
\]

evaluating the indefinite integral gives

\[
\frac{1}{-A'} \ln(N) = t + c \quad \text{(A.5)}
\]

where \(c\) is a constant of integration. Rigorously speaking, only the absolute value of the argument in the natural log is allowed. However, as \(N\) is known to be only a positive quantity the \(|N|\) is dropped and replaced with \(N\). Solving for \(N\) and making use of the initial condition that at \(t = 0\), \(N = N_o\) the final solution is obtained:

\[
N(t) = N_o e^{-A't}. \quad \text{(A.6)}
\]

A second approximation uses both nonradiative and radiative components. In this condition only the Auger recombination term, \(CN^3\), is ignored. The recombination equation is

\[
\frac{dN}{dt} \bigg|_{\text{decay}} = -(A + \eta_e B N_d)N - \eta_e B N^2 = -A'N - \eta_e B N^2. \quad \text{(A.7)}
\]

Noting again that the differential equation is variable separable,

\[
- \int \frac{dN}{A'N + \eta_e B N^2} = \int dt \quad \text{(A.8)}
\]

which integrates to give

\[
\frac{1}{A'} \ln\left( \frac{\eta_e B N + A'}{N} \right) = t + c_0 \quad \text{(A.9)}
\]
Appendix A. Solutions to the carrier recombination equation

rearranging yields

\[ \frac{\eta_e BN + A'}{N} = c_1 e^{At} \]  \hspace{1cm} (A.10)

where \( c_1 \) is again a constant of integration. Solving for \( N \)

\[ N(t) = \frac{-A'}{\eta_e B - c_1 e^{At}}. \]  \hspace{1cm} (A.11)

Evaluation of the constant is based on the initial condition, \( N(t = 0) = N_o \):

\[ N_o = \frac{-A'}{\eta_e B - c_1} \]  \hspace{1cm} (A.12)

which is solved to give \( c_1 = \eta_e B + A'/N_o \). Plugging the value of this constant back into Eq. A.10 and performing some algebraic manipulations gives the final solution:

\[ N(t) = \frac{N_o e^{-At}}{1 + \frac{\eta_e BN_o}{A'}(1 - e^{-At})}. \]  \hspace{1cm} (A.13)

As a check, note that in the limit \( t \to 0, N(t) \to N_o \). As an additional check note that if the radiative term is ignored, \( B = 0 \), the nonradiative solution, Eq. A.6, is recovered.

A final condition of the recombination dynamics is considered. If the initial carrier density is low enough to ignore Auger, but sufficiently high to ignore nonradiative recombination, then recombination is dominated by the \( BN^2 \) term:

\[ \left. \frac{dN}{dt} \right|_{\text{decay}} = -\eta_e BN^2. \]  \hspace{1cm} (A.14)

In this purely radiative regime, the differential equation can be solved through the straightforward integration

\[ \int \frac{dN}{N^2} = -\eta_e B \int dt \]  \hspace{1cm} (A.15)
which gives the following solution

\[- \frac{1}{N} = -\eta_e B t + c_o \]

(A.16)

where \( c_o \) is a constant of integration. Use of the initial condition, \( N(t = 0) = N_o \), gives \( c_o = N_o \) and the final solution under the radiative approximation is

\[ N(t) = \frac{N_o}{1 + \eta_e B N_o t}. \]

(A.17)
Appendix B

Fitting the Decay Tail of Semiconductor Luminescence

In this appendix the subject of fitting in the decay tail of semiconductor luminescence is discussed. Chapter 5 derived an expression for the late-time behavior of carrier recombination. Under the approximation that the decay is purely linear, $A' = A + BN_d$, the PL can be described by a single exponential, Eq. 6.25. What is needed, however, is an analytical means of determining the tail. Where is the tail? How long is the tail? And how can it be found in an objective manner?

To process a measured decay, a custom algorithm is developed. A measured decay displayed on a semi-log scale is depicted in Fig. B.1a. The goal of the algorithm is to extract a single exponential nonradiative lifetime representative of very low excess carrier density. At early times in the decay signal, the data displays a non-linear slope that indicates complicated (i.e. multi-exponential) recombination processes. The radiative, nonradiative recombination processes from Eq. 6.15 are all contributing at early times. At sufficiently long times, nonlinear recombination is minimized, but the data becomes noisy. The fitting procedure is tailored to find the range of data
Appendix B. Fitting the Decay Tail of Semiconductor Luminescence

where an optimum signal can be recovered. This occurs between the regions of non-exponential behavior and low signal-to-noise. The quality of a single exponential fit to a measured decay is quantified with the function $\chi^2$:

$$\chi^2 = \sum_{k=1}^{N} \frac{(O_k - E_k)^2}{E_k}$$  \hspace{1cm} (B.1)

where $E_k$ is the observed value at the $k^{th}$ point, $E_k$ is the expected value at that point, and $N$ is the number of points in the dataset. Next, $\chi^2$ is normalized to obtain

$$\tilde{\chi}^2 = \chi^2 / N.$$  \hspace{1cm} (B.2)

A value of $\tilde{\chi}^2 = 0$ represents a perfect fit and $\tilde{\chi}^2 \leq 1$ is regarded as a good fit [110].

The fitting algorithm works as follows: The last 20 data points of the measured decay are averaged; this region is outlined in Fig. B.1a by the open orange circle. This number is the average noise value for the data set. The algorithm searches the dataset, starting at the end for an average value of 20 data points whose value is ten times greater than the average noise. Once the location of a point within the dataset having a signal-to-noise of ten has been established, the data window is extended (see Fig. B.1a) indicated by the gray box. The window starts at low signal and progresses to high signal. As the window is extended to larger signal, all the included data points are fit with a single exponential function. At each fit, $\tilde{\chi}^2$ is calculated. In this way, $\tilde{\chi}^2$ is obtained as a function of the data window width (Fig. B.1c blue line). The corresponding lifetime is also shown (Fig. B.1c green line). Note that $\tilde{\chi}^2$ has a distinct minimum for a window width of $\approx 1300$ data points. For more data points, $\tilde{\chi}^2$ increases because the decay signal becomes increasingly non-exponential, i.e. the fit to a single exponential function is poorer. For smaller data windows, $\tilde{\chi}^2$ is degraded due to increasing noise.

The next step in the algorithm is to increment the position of the noise measurement window from ten to eleven and repeat the above procedure. This generates
Appendix B. Fitting the Decay Tail of Semiconductor Luminescence

Figure B.1: (a) Measured fluorescence decay (black line) and best fit (yellow line). The gray window is the region over which the fitting is performed. (c) \( \tilde{\chi}^2 \) (blue line) and fitted lifetime (green line) over the extending gray shaded region in (a), optimum window width is \( \approx 1300 \) points. (b) The \( \tilde{\chi}^2 \) at optimum window width as a function of signal-to-noise. (d) Lifetime of the fits calculated from (b).

\( \tilde{\chi}^2 \) as a function of the signal-to-noise (Fig. B.1b). The nature of this graph is the similar to the graph in Fig. B.1c; it displays a saddle point that identifies the region between non-exponential behavior and poor signal-to-noise. This defines the best fit in terms of the optimum data range with respect to the noise floor. The optimum fit is shown in Fig. B.1a with a solid yellow line. It is important to note that this same decay was fit by eye prior to fitting with the algorithm detailed here. The results of this subjective fit were 1.502 \( \mu s \) lifetime with a \( \tilde{\chi}^2=7.33e-5 \). In contrast, the algorithm produced a 1.69 \( \mu s \) lifetime and a \( \tilde{\chi}^2=9.626e-6 \), a nearly 10x reduction.
Appendix B. Fitting the Decay Tail of Semiconductor Luminescence

in error. The extracted lifetime corresponding to minimum the $\chi^2$ is shown in Fig. B.1d. Note the sharp drop in lifetime when signal-to-noise is higher than at the optimum. This means that higher signal-to-noise can come at the expense of non-single exponential decay.
Appendix C

Comprehensive Fitting of the Full Luminescence Decay

Fitting in the tail of the luminescence decay is only possible provided signal to noise is sufficient to resolve the tail. For situations where the decay tail is not accessible, fitting data in the early time must be used to quantify recombination. Of particular relevance to this dissertation is measurement of luminescence decays done through near-field probes, see Chapter 7. In this application low photon count rates prevent resolving the decay tail and fitting is more involved due to contributions from both radiative, nonradiative and possibly Auger mechanisms. The strength of Auger recombination can be controlled, however, through judicious of the excitation laser; restricting the carrier density to values low enough that Auger can be ignored, leaving only radiative and nonradiative avenues to participate in recombination.

A decay recorded in the near-field is shown in Fig. C.1 (black line) on a semi-log scale. Such a decay would represent the data recorded in one pixel of a fluorescence lifetime image. The acquisition and fitting of many such decays would comprise an image. Use of the fitting algorithm described in the previous section is not advised
for decays recorded in the near-field as only a small portion of the dataset can be fit. The data does not show evidence of linear behavior in the \(\approx 550\) ns time scale. This conclusion is further confirmed by the attempted iterative curve fit of a linear function (blue line). The fitting routine used here seeks to minimize \(\tilde{\chi}^2\) over the entire dataset, unlike the algorithm described in the previous section which fits only a portion of the data. Only the variables \(A'\) and \(\Delta\) are adjusted to minimize \(\tilde{\chi}^2\). The fit to normalized data is of the functional form expressed in Eq. 6.25. Note that only a relatively small region of the fit overlaps with the data, but overall the quality of the fit is poor. The poor fit suggests additional recombination mechanisms are at work and their presence must be accounted for.

Alternatively, a fit under the radiative/nonradiative approximation of Eq. 6.24, shows a very good fit (red dashed line). Inclusion of both \(A'\) and \(B\) coefficients is
Appendix C. Comprehensive Fitting of the Full Luminescence Decay

necessary to satisfactorily fit the data. Fitting is again based on reducing $\tilde{\chi}^2$. In principle, fitting is accomplished with the variation of three variables: $A'$, $\eta_e BN_o$ and $N_d$. However, the fit is not unique. The fit can be made unique by applying an approximation and making an additional measurement. The additional measurement is that of $N_o$ and is done by recording the intensity of the excitation laser used, knowing the wavelength of the laser light and knowing the absorption at the laser wavelength for the semiconductor used. From this information, $N_o$ can be computed, $N_o = \frac{\alpha(\nu)I}{h\nu}$, where $\alpha(\nu)$ is the absorption and $I$ is the laser intensity. From measurement the laser spot size on the sample is $\approx 4 \, \mu m$, the energy per pulse is 15 pJ and the wavelength is 405 nm. The absorption of GaAs at 405 nm is $10^5 \, cm^{-1}$ [111], which gives an initial carrier density of $N_o = 6.4e16 \, cm^{-3}$. The sample is known to be undoped so a background doping concentration of $N_d = 2.5e15 \, cm^{-3}$ is assumed [65]. The radiative recombination coefficient, $B$, is assumed to be constant across the sample. Thus only $A'$ is adjusted to fit the data shown in Fig. C.1. The nonradiative lifetime is found to be $\tau = 1/A' = 147$ ns with a $\tilde{\chi}^2 = 0.0358$. 

167
References


References


References


References


References


References


[79] www.nanonics.co.il


References


References


[106] Liftoff and bonding done by C. Wang, Univ. of New Mexico, 2008.


References

