

# ENERGY CONVERSION IN SILICON NANOSTRUCTURES AND DEVICES

A DISSERTATION

SUBMITTED TO THE DEPARTMENT OF ELECTRICAL ENGINEERING

AND THE COMMITTEE ON GRADUATE STUDIES

OF STANFORD UNIVERSITY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Jeremy Alexander Rowlette

June 2010

© 2010 by Jeremy Alexander Rowlette. All Rights Reserved.  
Re-distributed by Stanford University under license with the author.



This work is licensed under a Creative Commons Attribution-Noncommercial 3.0 United States License.  
<http://creativecommons.org/licenses/by-nc/3.0/us/>

This dissertation is online at: <http://purl.stanford.edu/yv233gq8296>

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

**Kenneth Goodson, Primary Adviser**

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

**Mark Brongersma**

I certify that I have read this dissertation and that, in my opinion, it is fully adequate in scope and quality as a dissertation for the degree of Doctor of Philosophy.

**R Pease**

Approved for the Stanford University Committee on Graduate Studies.

**Patricia J. Gumpert, Vice Provost Graduate Education**

*This signature page was generated electronically upon submission of this dissertation in electronic format. An original signed hard copy of the signature page is on file in University Archives.*





# Abstract

We examine the physics of nanoscale energy conversion in low-dimensional silicon structures and devices. The theoretical and experimental results of this work will facilitate improved designs of the silicon nanoelectronic and nanophotonic devices vital to emerging high-density, high-bandwidth information systems.

We begin by quantifying the reductions in nanotransistor performance by nonequilibrium optical phonons (OPs) generated in the drain as a result of hot electron energy relaxation. An inefficient conversion of OPs into the long-wavelength acoustic phonons (APs) responsible for heat conduction can cause an energy conversion bottleneck leading to reduced device currents and negative differential conductance instability, which poses a threat to the continued scaling of CMOS technology. We develop fully-coupled, electron-phonon, Monte Carlo Boltzmann transport simulations to assess the impact of this effect on device operation. The new approach includes electron and phonon dispersion and introduces a novel, occupation number-based stochastic rejection algorithm to efficiently couple the electron-phonon dynamics and nonequilibrium particle interaction. To address gaps in existing data, we use full-band, anharmonic perturbation theory to calculate the 2-phonon joint densities of states and lifetimes of OPs critical to hot electron energy conversion. We find that the reduction in electron mobility and the increase in leakage current due to hot OPs were greatly overestimated in previous studies and that drive current will not be impacted significantly for power densities below  $200 \text{ TW/cm}^3$ . This corresponds to 20x the existing device densities and 2x what is predicted at the end of the technology roadmap.

Next, we develop a new theory of the electron dynamics and nonradiative energy transfer in optically-pumped silicon nanocrystals (NCs) embedded in a dielectric host. This class of materials is promising in the development of inexpensive, CMOS-compatible, optical sources due to strong luminescence in the 600-1000 nm band and an ability to sensitize codoped  $\text{Er}^{3+}$  emitting near 1540 nm. We study the nonradiative recombination (NR) processes which limit the optical gain of these materials and

which prevent their adoption into commercial applications using a custom built, two-color optical pump-probe system which demonstrates sub-10 picosecond temporal resolution and sensitivity within 3 dB of the shot noise limit. From an improved data set, we establish a new theory which shows the importance of long-range Coulombic dipole-dipole interaction between excited electron states at high excitation levels, an important loss mechanism neglected in all previous work. The proposed model accurately reproduces the data on bulk and quasi-2D NC films down to 10 nm over more than four decades of pump intensity and unifies existing models by showing that their solutions are limiting forms of the new model. We experimentally observe the transition from 3D to 2D behavior as a systematic reduction in NR at high excitation levels by as much as 2x which is accompanied by a characteristic change in power-law dependence on pump intensity from  $1/3$  to  $1/4$  with decreasing film thickness in exact agreement with our theory. Further reductions in NR losses and a change in power-law to  $1/7$  are predicted in 1-D NC laden ridges. The substantial reduction in NR losses observed in 2-D films and even greater reductions predicted in 1-D ridges and 0-D islands reduces the intrinsic lasing threshold of these materials and demonstrates a path forward to realizing silicon NC lasers.



## Acknowledgments

I am humbled by the number of individuals who have blessed me with their time, their patience, and their knowledge over the course of completing my formal education. I am indebted to many but wish to especially acknowledge the few individuals who made this dissertation possible. First, I would like to thank my research advisor, Prof. Ken Goodson, for supervising my doctoral research and for guiding me through the natural ups and downs of scientific research. Prof. Goodson taught me, an electrical engineer, about the nature of heat, from phonons to heat sinks. I could not have asked for a better teacher. In addition to his technical advice, Prof. Goodson provided invaluable personal and professional career advice for which I have benefited greatly. I am also grateful for the research supervision I received from Prof. Mark Brongersma regarding silicon nanophotonic materials. In addition to what became a highly fruitful research collaboration, I cherish the many hours spent together at the white board discussing the physics of light interaction with nanomaterials. Among my thesis committee, I have been privileged to know Prof. Fabian Pease the longest. In 2001, while taking his course on charged particle beams and plasmas, I came to assist one of Prof. Pease's former students, Theresa Kramer, on the fabrication and testing of silicon nanopillar transistors. Upon her graduation, Dr. Kramer took a post-doc position within the Goodson Group and would eventually become the link between Prof. Goodson and myself some years later. I am thankful for the many technical conversations Prof. Pease and I shared over the years and am especially fond of the discussions concerning the *Kelvin Generator* and its close resemblance to the modern flip-flop. One of the most instrumental people in a graduate students' life can often be the group administrator. This was especially the case for me and so I want to give special thanks to Cecilia Gichane-Belle for her ability to manage the daily chaos of more than 20 faculty and students, while always maintaining her warm and calming disposition. I wish to acknowledge two former Goodson Group members, Dr. Eric Pop and Dr. Sanjiv Sinha who enabled me to ramp up quickly on developing nanoscale

transport simulations. Having colleagues who challenge you technically each day and who you can trust to give you critical feedback can make all the difference in doctoral program. I was fortunate to have two very close friends and collaborators, Matt Panzer and John Reifenberg, who filled this role. Matt and I spent countless hours building and rebuilding the picosecond pump-probe system from scratch and tracking down every stray photon and every nanovolt offset in the electronics. I also benefited greatly from my friendship and collaboration with Dr. Rohan Kekatpure and Dr. Aaron Hryciw, both members of the Brongersma Lab. Dr. Kekatpure and Dr. Hryciw were instrumental in supplying high-quality, thin-film silicon nanoparticle samples for the pump-probe studies. Outside of Stanford, there were many others who contributed to my personal and technical development to whom I wish to give special thanks. Dr. Gary Woods, a former colleague during my early days at Intel, and alumnae of the Ginzton Lab, is a close friend and technical confidant who I have sought advice from on numerous occasions. I wish to also thank Dr. William Lo for his technical advice from time-to-time and especially for his assistance in locating much-needed surplus equipment for completing the pump-probe system. In a former life, I had the great fortune to work with some truly fantastic people while an engineer at Intel. I would like to acknowledge Rick Livengood, Travis Eiles, Cameron Wagner, Seth Fortuna, Michael DiBattista, Lise Varner, and Paul Winer. Finally, I thank my family, Anita, John, Bobbi, Zoë, Mom and Dad, for being behind me for every step of this long journey. They have seen me at my best and at my worst and have kept me anchored in reality throughout. I am so thankful for their love and unwavering support. I am blessed to have a loving wife who understands me and who inspires me to be a better person every day. I thank my Mom, who taught me how to care about others, to be patient, to listen, and to never give up. I thank my Dad, my first Physics teacher, who taught me how to hit a jump shot, to fly cast, and to appreciate the subtle beauty of physics. Thank you.

## **Dedication**

This dissertation is dedicated to my family: Anita, John, Bobbi, Zoë, Mom and Dad

# Table of Contents

1	Introduction .....	1
1.1	Energy Conversion in Silicon Nanoelectronic Devices .....	3
1.2	Energy Conversion in Silicon Nanophotonic Materials .....	11
1.3	Overview.. .....	12
2	Fundamentals of Nanoscale Energy Conversion.....	14
2.1	Energy.....	14
2.2	Energy Conversion .....	19
2.3	Relevant Length and Time Scales .....	33
2.4	Energy Conversion in Nanostructures.....	36
2.5	Energy Conversion in Silicon Nanostructures .....	43
3	The Energy Eigen States of Silicon.....	44
3.1	Vibrational Eigen States.....	44
3.1.1	Bulk Phonon Modes .....	48
3.1.2	Vibrational Modes of Nanoscale Crystallites.....	55
3.2	Electron Eigen States.....	56
3.2.1	Bulk Electron States .....	60
3.2.2	Electrons in Nanoscale Crystallites .....	65
4	Phonon-phonon Energy Conversion in Silicon .....	67
4.1	Anharmonic Perturbation Theory .....	67
4.2	Acoustic Phonon Interaction and Lattice Conductivity.....	69

4.3	Optical to Acoustic Phonon Energy Conversion .....	70
4.4	Optical Phonon Lifetimes .....	75
5	Energy Conversion in Silicon Nanoelectronic Devices .....	77
5.1	Background .....	78
5.2	Nanoscale Heat Generation .....	80
5.3	Electron Transport Model .....	82
5.4	Phonon Transport Model .....	86
5.5	Coupled Electron-phonon Transport Model .....	89
5.6	Hot Optical Phonons .....	97
5.7	The Size Effect .....	102
5.8	Limitations of the Analytical Electron Band Model .....	103
5.9	Summary .....	103
6	Optical Pump-probe Methods.....	105
6.1	Experimental Layout .....	105
6.2	Frequency Domain Analysis .....	109
6.3	Thermo-reflectance Measurement Technique.....	115
6.3.1	General Solution to the 4-D Temperature Field .....	118
6.3.2	Temperature Field at the Transducer Layer .....	121
6.3.3	Frequency Domain Thermo-reflectance Signal.....	124
6.3.4	Calibration Standard (Al/SiO <sub>2</sub> /Si) .....	127
6.4	Transient Differential Transmission Measurements .....	128
6.5	Application Notes	
6.5.1	Beam Divergence and Beam Steering.....	131
6.5.2	Offsets and Signal Normalization .....	131
6.5.3	Photodiode Saturation .....	133
6.5.4	Optimizing SNR .....	135

7	Energy Conversion in Silicon Nanophotonic Materials .....	138
7.1	Silicon Nanophotonics .....	138
7.2	Physical Properties of Silicon Nanocrystal-Oxide Matrices .....	141
7.2.1	PECVD SRO .....	141
7.2.2	E-beam SRO .....	144
7.3	Differential Transmission Pump-probe Studies .....	147
7.4	Carrier Dynamics Modeling .....	151
7.4.1	Coupled Rate Equation Analysis .....	151
7.4.2	Long-range Dipole-Dipole Energy Transfer .....	152
7.4.3	Steady-state Analytical Model .....	155
7.4.4	Monte Carlo Analysis .....	160
7.4.5	Dipole-dipole Coupling in Low-D Systems .....	164
7.5	Experimental Evidence of Dipole-dipole Coupling in Quasi-2D.....	168
7.6	Conclusions .....	171
8	Conclusion .....	173
8.1	Summary .....	174
8.1.1	Nonequilibrium Energy Conversion in Nanotransistors .....	174
8.1.2	Energy Conversion in Silicon Nanophotonic Materials .....	177
8.2	The Future: Controlling Energy Conversion in Nanosystems .....	179
	Bibliography .....	183



## List of Tables

2.1:	Estimates of the electron coherence time, length and mean free path in silicon. ....	35
2.2:	Estimates of the phonon coherence time, length and mean free path in silicon. Note that the group velocity is taken to be the acoustic velocity averaged over the low-frequency [100] longitudinal and acoustic modes, i.e. $v_g = (v_{ac,l} + 2v_{ac,t})/3$ where $v_{ac,l}$ and $v_{ac,t}$ are 8.43 and 5.84 Km/s respectively. ....	36
3.1	Valence force potential constants for bulk silicon as determined by Tubino et al. and used in this work. ....	54
3.2	Summary of on-site and two-center overlap integrals used in computing the bulk and finite crystal bandstructures. The values are from Boykin et al. ....	63
4.1	Energy and representative wave vectors for four dominant LA + TA phonon pairs which a g-type LO phonon may create during spontaneous decay while conserving energy and crystal momentum. ....	74
5.1	Summary of electron and phonon transport model parameters used in simulations discussed in this work. ....	85
7.1	Comparison of extracted parameter values for the three model variations plotted in Fig. 7.5. ....	158

7.2	Summary of the laws of scaling for dipole-dipole coupling between excited states in systems of Poisson distributed optical centers. ....	166
-----	--	-----

## List of Figures

- 1.1 The standard CMOS inverter logic gate and its idealized equivalent RC circuit model. shown during a logical (b) 0 to 1 and (c) 1 to 0 transition. One of the notable idealizations is the assumption that the opposing transistor is completely turned off during the opposite half cycle. ....4
- 1.2 (a) The RC circuit, the most basic of circuit models helps build our intuition for far more complicated transport models. As the complexity of our physical models increase to capture details of the microscopic transport, we must return these concepts. (b) An equivalent circuit where resistor is replaced by an unknown device with nanoscale dimensions. Modeling the transport within the box can be quite challenging but the energy flowing into and out of the boundaries of the device must be conserved (on average). Dissipative processes within the box act like a time-varying resistor. ....4
- 1.3 The instantaneous power dissipated and the total cumulative heat dissipated over time for two different hypothetical resistor-time paths. In the top panel, the normalized instantaneous heat generation (red lines) and cumulative heat generated (blue lines) as a function of the normalized time are shown. The solid and dashed lines represent fixed and variable resistance conditions respectively. In the bottom panel, the I-V curves of the two resistors is shown. ....7
- 1.4 Volumetric heat generation rates as a function of transistor channel length through the end of the existing ITRS roadmap. ....9

1.5	The effects of high-density energy conversion in small volumes leads to non-equilibrium particle distribution functions.....	10
2.1	Schematic representation of the three lowest order perturbation expansion terms concerning the probability amplitude of taking the electron from state 1 with energy $E_1$ to state 3 with energy $E_3$ when there are a total of four eigenstates of the system. The first order term describes only the direct path whereas the second order includes all intermediate paths. The intermediate states are also referred to as “virtual” states. ....	24
2.2	Schematic representation of the one electron, two state example described by (2.29). ....	26
2.3	Illustration of how the uncertainty in the energy of a perturbation (caused here by the assumption that it is turned on abruptly at $t = 0$ ) leads to the equivalence of energy conservation breaking at short time scales. The energy of the coherent wave is given by $E_p$ at long time scales. When $t \sim 4\pi\hbar / E_{21}$ , there is an appreciable overlap of the spectral density function of the perturbation. However, this overlap vanishes as $t$ tends to infinity. ....	31
2.4	Schematic illustration of the Fourier representation of a Bloch wavefunction (right) describing an electron in a periodic crystal. The composite phase-space wavefunction arises from the convolution of two components: (top left) the non-periodic phase factor and (bottom left) the periodic function. ....	38
2.5	Schematic of the frequency space representation of a particle wavefunction (left) residing in a hypothetical one-dimensional periodic	

	lattice and having good quantum numbers $n$ and $k_0$ and (right) the same particle in a finite one-dimensional crystal with a linear dimension equal to $L \sim 4a$ . Each of the Fourier components of the wavefunction exhibit a characteristic broadening of order $1/L$ . ....	39
2.6	(left) The potential energy of a single unit cell of a hypothetical 1-D crystal having two bound energy states $E_0$ and $E_1$ in the isolated, i.e. one unit cell, case. (right) Energy manifold showing how $E_0$ and $E_1$ split into a total of eight levels, two of which are degenerate, when four unit cells are included and periodic boundary conditions are enforced. ....	40
2.7	(Left) Energy band diagram for the periodic crystal having 4 unit cells (red open circles). The black dotted lines represent the form of the bands as the number of unit cells is increased to 40. (Right) Projection of the two lowest energy state wave functions (ground state = black square ; 1 <sup>st</sup> excited state = blue squares) onto a plane-wave (i.e. Fourier) basis for (a, c) a periodic crystal having 4 unit cells (a ring) and (b, d) a finite crystal with exactly 4 unit cells. ....	41
2.8	(a, c) Real-space and (b, d) $k$ -space representations of the ground state wavefunction for the (a, b) finite crystal with 1 (purple), 2 (blue), 3 (green), and 4 (black) unit cells in comparison with (c, d) a periodic crystal having four unit cells. The four-unit cell periodic potential is shown in (a) and (c) for qualitative comparison. ....	42
3.1	Atomic structure of silicon including the primitive cell and the 12 neighboring primitive cells which include up to fourth nearest neighbor. ....	46

3.2	The primitive cell of the reciprocal lattice of silicon, i.e. the first Brillouin zone (FBZ) showing the irreducible wedge and Group symbols for the high-symmetry points and directions. ....	47
3.3	Schematic representation of two basic internal coordinates, including (left) bond stretching and (right) angle bending. The bond stretching internal coordinate involves two atoms whereas the bond angle displacement internal coordinate involves the motion of three atoms. The orientation of the displacement vectors for each of the atoms is indicated by a straight arrow. ....	50
3.4	Schematic representation of the types of atomic displacements which make up the three internal coordinates (top row) and the coupling between internal coordinates (bottom row) used to describe the motion of the silicon lattice. There is a total of 16 internal coordinates defined for the bulk silicon lattice. ....	53
3.5	(top) Bulk phonon dispersion relationship for the silicon lattice computed using the VFP model. (bottom) Constant energy contours of the six phonon branches in the $K_x$ , $K_y$ plane. The acoustic modes are shown on the top row and the optical modes on the bottom row. The red octagons indicate the boundary of the FBZ in the specified plane. Lighter regions indicate higher energy. ....	55
3.6	Energy distribution of the vibrational states as a function of crystal size as indicated by the number of atoms. The distribution rapidly converges to that of the bulk crystal. ....	57
3.7	The effect of boundary conditions on the energetic distribution of vibrational modes. The top two distributions represent the cases where	

	the surface atoms are take to be equivalent to the silicon mass or taken to be infinite in mass respectively. ....	58
3.8	The effect of boundary conditions on the energetic distribution of vibrational modes. ....	59
3.9	(top) Electronic bandstructure of bulk silicon along high-symmetry directions and (bottom) constant energy surfaces for (left) valence band holes and (right) conduction band electrons. ....	64
3.10	Electron state histogram (discrete density of states) for (top) $\text{Si}_{18}\text{H}_{40}$ 0.543 nm cube ( $E_g = 3.42$ eV, $\lambda = 362$ nm) (bottom) $\text{Si}_{95}\text{H}_{124}$ 1.08 nm cube ( $E_g = 2.33$ eV, $\lambda = 532$ nm) . ....	65
3.11	Energy gap of cubic silicon nanocrystals as a function of crystal size as calculated using the empirical tight binding approach and the simple effective mass approximation. ....	66
4.1	Constant energy contours of the silicon phonon band structure along the Kx-Ky plane for the longitudinal optical, longitudinal acoustic, and transverse acoustic branches (left to right). These contours were calculated using the Valence Force Potential (VFP) approach discussed in Chapter 3. ....	72
4.2	Figure 4.2: Visualization of the mathematical problem required in solving the joint density of states in 3-D and 2-D systems. ....	72
4.3	$g_2(E, E_0 - E)$ for an LO phonon of initial energy $E_0$ and wavevector directed along $\Gamma$ -X at the points $(2\pi/a)[0,0,\alpha]$ where $\alpha = 0, 0.3, 0.5, 0.7$ , and 1.0 (labeled (a)-(e) respectively) decaying into two lower energy,	

	crystal momentum conserving normal modes of energies $E$ and $E_0 - E$ . The single-phonon density of states $g(E)$ is shown in light blue for reference on top of (a). (bottom) Zoomed in view of the $g_2(E, E_0 - E)$ spectrum for the $g$ -LO ( $\alpha = 0.3$ ) phonon. The vertical axes are in arbitrary units. ....	73
4.4	Calculated lifetime for LO phonon modes along $\Gamma$ -X. The inset shows the calculated spectral full-width at half maximum (FWHM) for the R-LTO mode calculated using semi-empirical approach (black solid) and by the DFT calculations (red dotted curve) compared to experimental data (solid blue dots). Note that the axes ranges for the inset are the same as that shown in Fig. 2 of [46].....	76
5.1	Schematic representation of the broadening effect on the phonon generation spectra with increasing power density. The energy contours represent energy spacing of approximately one phonon energy such that a transition from one valley to the other corresponds to a reduction in energy by the amount equal to $f$ -type phonon. ....	81
5.2	Electron conduction band structure for silicon computed using (blue) an empirical tight binding parameter set and (red) a single six-valley analytic nonparabolic electron band used in this work. The shaded region corresponds to the region of the conduction band below 1 eV where electrons are likely to be found in sub-1V devices. As the figure shows, the biggest limitation in the single, analytic, non-parabolic band in low-voltage transport is the inability to capture the structure of the second lowest energy conduction band, which is important in the neighborhood of the X and K,U points. ....	83

5.3	Phonon dispersion for silicon computed using (blue) a valence force field model compared to (red) an isotropic model using simple quadratics as used in this work. ....	84
5.4	Comparison of the single phonon density of states spectrum as computed using a full-band calculation and the isotropic model with quadratic dispersion relations for all branches. ....	84
5.5	The algorithm for combining the electron MC simulation with the phonon BTE solution.....	90
5.6	A schematic representation of the quasi-1D two-terminal n <sup>+</sup> /n/n <sup>+</sup> device structure modeled in this work and a comparison to a more realistic device structure. ....	93
5.7	Electrostatic characteristics of the 1-D n <sup>+</sup> /n/n <sup>+</sup> device simulated in this work. The device has three regions: two 150 nm n <sup>+</sup> (10 <sup>20</sup> cm <sup>-3</sup> ) source and drain regions separated by a 20 nm n-type (10 <sup>16</sup> cm <sup>-3</sup> ) “channel”. The doping is uniform within each region and 1.25 nm/decade characterizes the doping concentration roll-off between regions. (top) current density (left axis) and peak electric field (right axis) vs. source-drain bias voltage. (bottom) electron potential energy vs. position within the device for three bias conditions. ....	94
5.8	Volumetric power generation rate vs. position within the 1-D n <sup>+</sup> /n/n <sup>+</sup> device. The power generation at each grid point was computed according to Eq. 5.3.....	95
5.9	(left) Phonon dispersion and (right) phonon generation spectra computed at 4 locations within the device for the 1V case. The red	

	(right-most) curve corresponds to the location of the peak power dissipation. The remaining curves correspond to $r = 10, 20$ , and $30$ nm displaced from the peak generation point or hotspot within the drain. ....	96
5.10	Average electron energy as a function of position for an optical phonon lifetime of $10$ ps for uncoupled (dashed line) and fully-coupled/closed-loop (solid line) simulations. ....	96
5.11	(top) Effective lattice temperature ( $T_{\text{eff}}$ ) computed using the self-consistent e-MC/p-SFBTE method. (bottom) $T_{\text{eff}}$ for the LO phonon branch (solid black - b) compared to that of the g-LO phonon (dashed black - c) for the $1$ V bias condition. The temperature computed from classic heat diffusion equation is also shown for reference (solid blue – a). ....	98
5.12	Phonon distribution at peak power generation point ( $r = 25$ nm). The inset shows a zoomed in view of optical mode occupation as a function of optical phonon lifetime ( $\tau_{\text{op}} = 0.5, 1, 10$ ps). ....	99
5.13	Excess occupation number for the g-type longitudinal optical (g-LO) phonon as a function of volumetric power density as calculated for uniform power generation (upper dashed line data set) and as calculated at the peak power generation point in the 1-D device (lower solid line data set). ....	100
6.1	Experimental layouts for the two basic configurations of the pump-probe system. In the top panel, the system is configured in “reflection” mode which can be used for measurements such as transient thermal reflectance thermometry. In the bottom panel, the system is configured	

	in “transmission mode.” In this configuration, the system can be used for transient differential absorption measurements. ....	106
6.2	Images of the picosecond pump-probe system developed by myself and Matt Panzer. The pump-probe system is used in two basic configurations: reflection and transmission mode. Reflection mode was used extensively for extracting thermal properties of nanostructured thin films using time-domain thermo-reflectance technique. The transmission mode configuration was used extensively for measuring carrier dynamics in silicon nanoparticle oxide films, the subject of Chapter 7. ....	107
6.3	Real-time sequence of the unmodulated pulse train. ....	109
6.4	An idealization of the frequency spectrum of the unmodulated laser pulse train. ....	111
6.5	Pump modulation signal approximated as a rectangular wave, $0 \leq A_L < A_H \leq 1$ ....	111
6.6	Relative amplitude of the pump spectrum vs. duty cycle fraction. ....	113
6.7	The response of the probe beam due to a fast excitation in a thin film caused by the pump beam. The initial change of the excitation is assumed to be sub-picosecond and therefore the rising edge of the response is a good measure of the convolution of the pump and probe intrinsic temporal widths. From this data, we see that the characteristic width is roughly 20 ps which corresponds to the 2x the pulse width of the source. The sample used to make this measurement was a thin	

	silicon rich silicon oxide film containing 5 nm silicon nanocrystals with a density of approximately $2 \times 10^{18} \text{ cm}^{-3}$ . ....	114
6.8	Schematic of the heating geometry of the sample looking down on the sample along the beam path. ....	117
6.9	Schematic representation of a material stack to be modeled. The z-component of the temperature field is treated using transmission matrix approach, a well-known technique routinely used in electromagnetics and quantum mechanics but which is less known within the thermal physics community.....	118
6.10	A typical amplitude spectrum (green) of $H(f)$ for the positive frequency domain from 0 to 100 MHz. The very low frequency limit is independent of frequency and in the high-frequency limit, $ H(f)  \propto 1/\sqrt{\omega}$ . The amplitude spectrum and real part of H is even with respect to the origin whereas the phase and imaginary part are odd. ....	124
6.11	Graphical representation of the convolution which represents the measured signal on the lock-in amplifier for each probe delay. ....	126
6.12	Graphical representation of the convolution represents the measured signal on the lock-in amplifier for each probe delay zoomed into spectral region of interest. ....	127
6.13	Measured thermal response for an Al/SiO <sub>2</sub> /Si material stack demonstrating the accuracy of the pump-probe thermo-reflectance technique. The negative time delay signal has been subtracted from the signal during analysis. ....	128

6.14	Pictures of the pump-probe system configured for transient differential transmission measurements. The liquid helium cryostat with integrated heater enabled measurements from about 5 K up to 400 K. ....	129
6.15	(top) A typical trace of the average detected probe signal voltage over the full time-delay scanning window (3.8 ns). The signal drops off with increasing delay because of finite beam steering of the delay stage as well as divergence of the beam. (bottom) A typical differential transmission signal obtained by exciting a silicon nanocrystal oxide film with 532 nm light and probing at 1064 nm. The red curve is the uncorrected lock-in magnitude and the blue curve shows the same signal after dividing by the normalization signal (top). Beyond about 500 ps, the need for correcting for finite beam steering and divergence is clearly evident. . ....	132
6.16	(Top) (Blue diamond) Measured RMS voltage fluctuation on probe detector in a 1 Hz bandwidth as a function of average detector signal voltage. (Magenta square) Data points showing calculated shot noise voltage for given photocurrent. The inset shows the ratio of the measured voltage noise to the shot noise limited voltage noise. For signal levels on the order of 150 mV, one can operate 3 dB of the shot noise floor. (Bottom) Plot of the relative SNR as a function of average detector voltage. Increasing the signal above about 150 mV does not improve SNR. ....	138
7.1	Physical properties of the SRO films measured in this work. $\tau_{\text{rad}}$ was obtained from transient PL experiments described in and $\tau_{\text{A},2}$ was computed using the formula, $\tau_{\text{A},2}^{-1} = C_{\text{A}}(2/V)^2$ , with $C_{\text{A}} = 4 \times 10^{-31} \text{ cm}^6/\text{s}$ and $V = (\pi/6)D^3$ . The distribution function was calculated using the methodology described. ....	142

7.2	High resolution TEM cross-section images for a PECVD deposited silicon rich oxide, which was subsequently annealed at 1100 C for one hour to produce silicon nano-crystallites with diameters in the vicinity of 5 nm and a density of $2 \times 10^{18} \text{ cm}^{-3}$ . ....	143
7.3	High resolution TEM cross-section images for a e-beam evaporated SiO, which was subsequently annealed at 1100 C for one hour to produce silicon nano-crystallites. ....	145
7.4	High resolution TEM cross-section images for a e-beam evaporated SiO/SiO <sub>2</sub> superlattice film containing 10 periods of 20 nm nc-Si-SiO <sub>2</sub> and 20 nm SiO <sub>2</sub> barrier layer. The sample was annealed at 1100 C for one hour to produce silicon nanocrystallites as can be seen as bright spots in the energy filtered HRTEMs. ....	146
7.5	Transient, differential transmission measurement results for the PECVD sample. Both the background signal and the peak amplitude of losses increase with increasing pump intensity. The initial decay time constant reduces with increasing pump intensity as a consequence of Coulombic interaction between excited electrons. ....	147
7.6	(a) The peak (red square) and negative time (blue circles) delay relative transmission loss as a function of pump intensity. (b) The difference between the peak value and the negative time delay (steady-state) value vs. pump intensity showing linearity of pump excitation cross-section. (c) The initial time constant as a function of pump intensity showing a -2/3 law. ....	148

7.7	Two experimental methods used to determine the optically-induced FCA losses in SRO films by measuring the (a) loss-induced spectral broadening of the 895 nm PL emission line of an SRO microdisk resonator and the (b) steady-state (negative time-delay) component of picosecond transient differential absorption (TDA) measurements performed on a 500 nm thick SRO film deposited on a fused-silica substrate. ....	149
7.8	The normalized free-carrier absorption coefficient ( $\alpha_{FC,max} = 24.6 \pm 0.3 \text{ cm}^{-1}$ at 1064 nm) vs. pump intensity for the two complementary pump-probe measurement techniques described in the text and in Fig 1. The solid red curve is the best fit for the proposed model. Linear, 1/3 and 1/2 power-law curves are included as visual guides. The inset is a zoomed in view of the maximum at $200 \text{ KW/cm}^2$ . ....	150
7.9	(a) Comparison of the three model variants including the independent (i.e. noninteracting) NC model with values of $e_{21} = 2$ and $e_{21} < 1$ (black and blue solid curves respectively) and the proposed model which includes near-field interaction between $N_1$ states (red solid curve). Each model variant is optimal with respect to fitting the data. The curves in the inset are plotted on a logarithmic scale zoomed into the region where the 1/3 power law region is observed. (b) Relative error for the three model variants compared to the experimental uncertainty (shaded region). ....	159
7.10	(a, b) Visualization of model sensitivity to $\Gamma$ and $e_{21}$ (c) Data Set B (solid circles), $N_I + e_{21}N_2$ (A) scaled to unity and (B) normalized by $N_T$ , (C) $N_I/N_T$ , and (D) $e_{21}N_2/N_T$ . The sequence of pictures in (c) depicts the dominant nonradiative recombination channels vs. pump intensity ( $I_p$ )	

	and the solid and dashed lines in the manifold indicate active and inactive FCA states respectively. ....	160
7.11	The physical domain of a typical Monte Carlo calculation for (a) the case a thick ( $> 100$ nm) film considered to be infinite in all directions and (b) a film with finite thickness, which we use to calculate the dynamics of ultra-thin (10 nm) quasi-2D films. (c) A top down schematic cross-sectional view of either 3D or quasi-2D domains. The middle cell is generated by randomly positioning the nanocrystals according to the specified average density. Then this cell is repeated to impose cyclic boundary conditions in directions representing infinite sample extent. The period of the cell is typically on the order of 100 nm or larger in order to ensure that any particular nanocrystal does not see a mirror image of itself or of its neighbors. ....	161
7.12	Typical transient Monte Carlo simulation output showing average nanocrystal occupancy vs. time in response to a step function pump intensity input. The red highlighted region of the curve is the portion of the curve which is used to calculate the steady-state parameters as a function of pump intensity. ....	162
7.13	Results of MC simulations (open circles) which treat the near-field coupling between NCs in the $N_1$ state by a term of the form $\gamma R_1^{-6}$ where $R_1$ is the distance between the two interacting NCs. The MC results are compared to the steady-state analytical model where the near-field interaction is described by $\Gamma \cdot N_1^2$ (solid curves labeled A, C, & D as defined in Fig. 4). (a) Results of MC simulation calibration performed to fine-tune the ratio $\gamma/\Gamma$ and also to demonstrate the physical equivalence of $\gamma \cdot R_1^{-6}$ and $\Gamma \cdot N_1^2$ . To assist the calibration process, the material parameters were intentionally skewed ( $\sigma = 3.8 \times 10^{-16} \text{ cm}^2$ ; $\tau_1 =$	

	380 $\mu\text{s}$ ; $\tau_{A,2} = 59 \text{ ns}$ ; $\Gamma = 2.3 \times 10^{-29} \text{ cm}^6 \text{ s}^{-1}$ ; $e_{21} = 0.5$ ) in order to force the near-field coupling to dominate the recombination physics, and thereby exaggerate the 1/3 power-law region over an extended range of pump intensities. From the calibration, we find $\gamma/\Gamma = 1/110$ . (b) MC simulation results using the calibrated value of $\gamma = \Gamma/110$ and the optimal parameter set described in the text ( $\sigma = 3.8 \times 10^{-16} \text{ cm}^2$ ; $\tau_1 = 38 \text{ ns}$ ; $\tau_{A,2} = 5.9 \text{ ns}$ ; $\Gamma = 2.9 \times 10^{-29} \text{ cm}^6 \text{ s}^{-1}$ ; and $e_{21} = 0.5$ ). The analytical curve is the same as in Figs 7.8, 7.9, and 7.10 and accurately reproduces the experimental data. ....	163
7.14	Steady state Monte Carlo calculations of a quasi-2D film of thickness 20 nm. The scattered blue dots in each figure represents the average occupancy at a particular location for single simulation. The red curve is an ensemble average of all simulation runs. The sequence of figures represents pump intensities of 1, 10, and 25 $\text{KW}/\text{cm}^2$ . As the pump intensity increases, so does the overall occupancy of the nanocrystals. However, near the edges of the film, the nanocrystals have higher occupancy on average due to the reduced neighbor interaction. At 10 $\text{KW}/\text{cm}^2$ , the ratio of the occupancy at the edge to the center is approximately 2:1 and decrease at higher pump intensity due to higher order Auger processes limiting the occupancy. ....	167
7.15	Excited carrier losses vs. pump intensity for three different nc-Si/SiO <sub>2</sub> samples fabricated using the e-beam evaporation technique as described in Section 7.2. The samples were prepared by Dr. Aaron Hryciw of the Brongersma Lab. ....	168
7.16	Steady-state photoluminescence spectroscopy measurements of the three samples: 200 nm bulk film (Th), 20 nm x 10 superlattice (SL1) and 10 nm x 20 superlattice film (SL2). The data show an increasing	

	blue shift with decreasing film thickness as well as a maximum PL signal by the 20 nm SL sample. The PL data was taken by Dr. Aaron Hryciw in the Brongersma Lab (Stanford University). ....	170
7.17	High resolution and energy filtered (bottom four images) TEMs of the thick and SL1 samples shown side-by-side for direct comparison. ....	171
8.1	Steady-state Monte Carlo simulation of an optically pumped (CW), nanoparticle (NP) system having a spherical geometry with radius 10 nm. The simulation model includes inter-NP dipole-dipole coupling and nonradiative energy exchange between excited electron states as was described in Chapter 7. The top panel shows a scatter plot of NP occupancy, defined by the relation $N_1/N_T + 2*N_2/N_T + 3*N_3/N_T + \dots$ , as a function of radial position. The solid light-blue line indicates the ensemble average occupancy vs. radial position. As expected, the occupancy tends to increase near the edges due to the reduction in interacting neighbors. The apparent reduction near the center of the sphere is an artifact of the simulation and is not indicative of any physical process. The binning excluded NPs within the inner most 1 nm region. The bottom panel is a scatter plot of the steady-state heat generation rate per second normalized by the bandgap energy of the NP system. The solid green line indicates the ensemble averaged heat generation rate vs. radius. In contrast to the occupancy plot, the heat generation per particle at steady-state conditions is uniform along the radius. ....	181



# Chapter 1

## Introduction

Curiosity demands that we ask questions, that we put the things together, to understand this multitude of aspects as perhaps the action of a small number of elemental things.

- R. P. Feynman

When we think of *energy conversion*, solar panels, wind turbines, hydroelectric dams, and nuclear power plants likely come to mind. As a scientist or engineer, our thoughts may quickly turn to energy conversion devices found in the lab, such as fuel cells and thermoelectric modules. Each of these is an example of a macroscopic system or device whose intended function is, by design, to convert energy from one form into another. In the examples listed, solar, gravitational, nuclear, chemical, and thermal energy are converted into a more convenient form of electrical energy. If we think a little harder, we may start to include devices not commonly listed as energy converters, such as photodetectors (optical to electrical), piezoelectric actuators (electrical to mechanical), and diode lasers (electrical to optical). If we then consider the fact that any device which performs useful work must generate entropy and thereby *dissipate* energy in the form of heat in accordance with the second law of thermodynamics [1], we must conclude that all devices which perform useful work\* are essentially energy converting devices, whether or not their intended function warrants their classification as such in common practice. Work is required to process, transport, and store information and so it follows that information handling devices necessarily produce heat and are thus fundamentally energy converting devices.

---

\* By definition, a device is something which does useful work.

Take for example the silicon transistor, the basic building block of modern information systems. Its function is simple: to *translate* a voltage level at its gate electrode into a corresponding voltage at its drain electrode. An input voltage at the gate controls the flow of current through the transistor channel leading to the charging or discharging of an output capacitance whose voltage represents binary information, i.e. a “1” or a “0”. One needs only to place their hand on the backside of their “laptop” computer to discover where all of the energy flowing through the billions of transistors making up the CPU ultimately winds up.

Modern microprocessors dissipate on the order of 100 J of heat every second or the equivalent of 1.5 incandescent light bulbs. Since a single microprocessor chip can execute roughly a billion instructions every second, the average energy required to execute a single instruction is of order  $10^{12}$  eV, which is 14 orders of magnitude larger than the theoretical minimum of  $k_B T \cdot \ln(2) \sim 10^{-2}$  eV required to execute a single binary operation [2]. Clearly, there is plenty of room at the bottom for improving efficiency!

What dictates the amount of heat generated during a single binary operation? How long does it take to convert this energy into heat? What are the elementary particles and processes which participate in the conversion of energy into heat and thereby act to increase the entropy of the system? Where does the heat go once it is generated? These are important questions to be asked of any system. However, asking such questions is vitally important when it comes to building and analyzing devices that are made so small that the number of atoms comprising the device become countable<sup>†</sup>.

This dissertation describes a theoretical and experimental investigation of the fundamental energy conversion processes which strongly influence the basic operating characteristics and stability of emerging silicon-based nanoelectronic and nanophotonic information processing devices. Let us consider the foundation of the problems to be addressed in this work by examining more closely the nature of energy dissipation in nanoscale electronic and photonic devices individually.

---

<sup>†</sup> Roughly speaking, countable means systems containing less than  $10^6$  atoms or volumes of material less than a few tens of thousands of cubic nm.

## 1.1 Energy Conversion in Silicon Nanoelectronic Devices

The silicon-based Complementary Metal Oxide Semiconductor (CMOS) technology has dominated the information systems landscape for more than three decades. The device structures may change substantially over time but the same basic operating principles should continue to apply even as novel device structures such as carbon nanotube (CNT) field effect transistors (FET) get closer to supplanting the silicon MOSFET.

The simplest of CMOS logic gates, the inverter, and its equivalent circuit as it “processes” either a logical 1 or 0 at its input is shown in Figure 1.1. Here, p and n subscripts identify the majority carrier type, hole or electron, respectively. Although this dissertation is concerned primarily with microscopic transport physics and energy conversion occurring within a single transistor, it is useful to briefly review the macroscopic behavior of the CMOS inverter because the complexity of the systems we will model have great potential to diminish our physical intuition of the system behavior.

During a logical transition, only one of the transistors in the CMOS inverter is intended to be fully turned on. Making this increasingly poor assumption, we can model the system as an network of resistance and capacitance with the transistor being modeled as a time varying resistor like that drawn in Figures 1.1b and 1.1c for the case of a  $1 \rightarrow 0$  and  $0 \rightarrow 1$  transition at the gate, respectively.

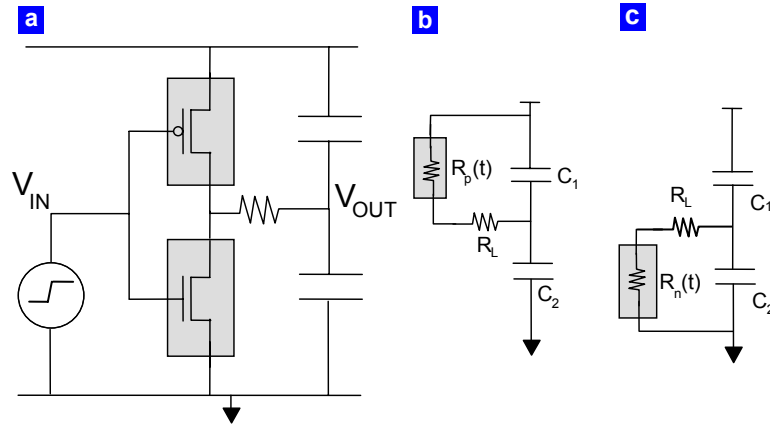


Figure 1.1: The standard CMOS inverter logic gate and its idealized equivalent RC circuit model, shown during a logical (b) 0 to 1 and (c) 1 to 0 transition. One of the notable idealizations is the assumption that the opposing transistor is completely turned off during the opposite half cycle.

In analyzing even this simplified form of the circuit, it is instructive to begin with an even simpler example, our old friend, the switched RC circuit drawn in Figure 1.2a. The capacitor is initially discharged and the switch is opened at time  $t = 0$ .

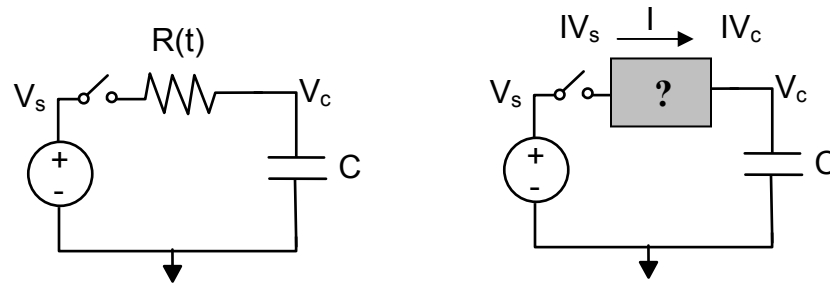


Figure 1.2: (a) The RC circuit, the most basic of circuit models helps build our intuition for far more complicated transport models. As the complexity of our physical models increase to capture details of the microscopic transport, we must return these concepts. (b) An equivalent circuit where resistor is replaced by an unknown device with nanoscale dimensions. Modeling the transport within the box can be quite challenging but the energy flowing into and out of the boundaries of the device must be conserved (on average). Dissipative processes within the box act like a time-varying resistor.

Heat generated by the circuit is entirely the result of the resistor element, the instantaneous heat generation being given by the Joule heating rate,

$$\dot{Q}(t) = I_R^2(t)R(t) = \frac{V_R^2(t)}{R(t)} \quad (1.1)$$

where  $I_R$  and  $V_R$  is the current flowing through and the voltage dropped across the resistor  $R$  respectively. If the resistance remains unchanged during the charging cycle, i.e.  $R(t) = R_0$ , we obtain the well-known result,

$$\begin{aligned} Q &= \int_0^{t' \rightarrow \infty} \dot{Q}(t') dt' = \frac{V_R^2(0)}{R_0} \int_0^{t' \rightarrow \infty} e^{-2t'/R_0 C} dt' \\ &= \frac{1}{2} C V_R^2(0) = U_E \end{aligned} \quad (1.2)$$

which states that the total heat generated  $Q$  by the charging of the capacitor is equal to the amount of potential energy stored in the electric field of the capacitor  $U_E$  upon completion of the charging<sup>‡</sup>. The value of  $V_R(0)$  is the voltage dropped across the resistor just after the switch has been closed, which in this case is equal to the value of the source voltage. Although, all of the heat is dissipated in the resistor, the magnitude of the resistor plays absolutely no role in dictating the total amount of heat dissipated; in combination with the capacitance value, it only modifies the rate at which energy is dissipated and transferred to the capacitor by the ideal voltage source.

Furthermore, if the capacitor is periodically charged and discharged every  $T = 1/f$  seconds, then the total amount of energy dissipated every second is given by

$$P_{tot} = C V_R^2(0) f \quad (1.3)$$

---

<sup>‡</sup> Mathematically speaking, this of course takes an infinite amount of time but in a practical sense, about 5 time constants.

But does this result hold if the resistor is time-varying, as for example in the case of a CMOS logic gate during its switching between binary states? The total heat generated during the transition is now given by the following expression

$$Q = V_R^2(0) \int_0^{t' \rightarrow \infty} \frac{\exp\left(-\frac{2}{C_0} \int_{t'}^{t''} \frac{dt''}{R(t'')}\right)}{R(t')} dt' \quad (1.4)$$

The increased complexity of the function masks a simple and most elegant result. Provided that  $R(t)$  remains positive<sup>§</sup> over all time, the total amount of heat generated over the course of the charging cycle is identical to the case of when the resistance remains constant, that is  $Q = (1/2)CV_R^2(0)$ .

As an example, let us compare the instantaneous power and total heat generated as a function of time for an RC circuit for two cases: first, the resistance is fixed with a value  $R_0$  and second, the resistance is time-varying having an arbitrary temporal dependence given by  $R_0 \cdot f(t)$  where  $f(t)$ . Let us normalize the time base by the time constant of the standard RC problem, i.e.  $t^* = t/\tau$  where  $\tau = R_0 C$ , and the instantaneous heat generation by  $CV_R^2(0)/\tau$ , and the total heat generated up to time  $t^*$  by  $CV_R^2(0)$ . For the case of the variable resistance, let us take as our arbitrary function to be

$$R(t^*) = R_0 [2 + \sin(10 \cdot t^*)] \quad (1.5)$$

The results of these two cases are plotted in Figure 1.3. In the top panel, the normalized instantaneous heat generation (red lines) and cumulative heat dissipated (blue lines) as a function of the normalized time are shown. The solid and dashed lines represent fixed and variable resistance conditions respectively. In the bottom panel, the I-V curves of the two resistor models are shown. For the case of the temporally varying resistor, we observe over the range of voltage plotted eight negative

---

<sup>§</sup> A negative resistance, not to be confused by a negative differential resistance, implies gain of the system which must be facilitated by an additional energy source.

differential conductance zones. Such behavior is typical of resonant tunneling diodes [3] and quantum cascade lasers [4] where intrinsically quantum mechanical effects dominate the macroscopic behavior of the device.

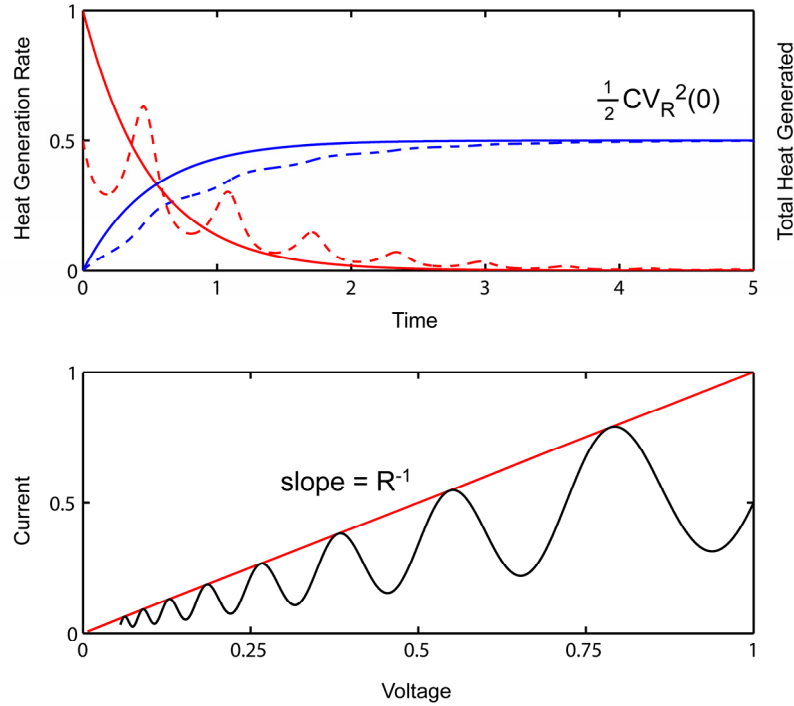


Figure 1.3: The instantaneous power dissipated and the total cumulative heat dissipated over time for two different hypothetical resistor-time paths. In the top panel, the normalized instantaneous heat generation (red lines) and cumulative heat generated (blue lines) as a function of the normalized time are shown. The solid and dashed lines represent fixed and variable resistance conditions respectively. In the bottom panel, the I-V curves of the two resistors is shown.

The point of this simple exercise is to make it clear that regardless of the complexity of the objects placed inside the gray box of Figure 1.2b, or what level of physical detail we chose to describe the objects in the box, the total amount of energy transferred to the load capacitance and the total amount of heat generated in the box is fixed by the external parameters of the system as defined at its input and output ports<sup>\*\*</sup>.

If we were interested only the total amount of energy dissipated by the device, and not in how long it took to perform the transfer of energy to the reactive load, or where

<sup>\*\*</sup> Admittedly, I have conveniently sidestepped the tricky issue of defining the edge of the box!

exactly the heat was generated between the terminals of the device, this would essentially be the end of the story. The only apparent levers available to the device engineer would be stray capacitance, operating voltage, and clock frequency. However, this is only part of the story. We care very much about how quickly and where the heat is generated within the device as it directly impacts the performance of the device and ultimately the system. How the energy is distributed within the device controls the rate of information transfer, the stability of the information contained at its terminals and the long-term reliability of the device, three critically important aspects of an information processing device.

Efficient removal of heat from the transistor and interconnect layers will be a growing challenge to the successful scaling of digital nanotechnologies for the foreseeable future. This challenge is growing due to the continued scaling down of transistor dimensions and the increasing density of devices combined with the move to thin-body single and multiple gate devices that provide improved control of the channel electric field. The most promising multi-gate device is the FinFET [5] of which numerous derivatives have been proposed in recent years [6] [7]. While these devices offer superior subthreshold slopes approaching the theoretical limit of 60 mV/decade, they exhibit higher thermal resistance than bulk devices because of geometric confinement by low thermal conductivity materials as well as enhanced phonon boundary scattering in the active layers.

The amount of energy converted to heat during the flipping of a single logical bit is reducing over time due to a gradual reduction in supply voltage and a continued reduction in device capacitance. However, these elements are not reducing as rapidly as the volume where the heat is initially dissipated, typically the drain extension region of the device. As a consequence, the volumetric power density is increasing with each technology generation. Based on simple volumetric scaling arguments and considering a modest rate of reduction in voltage as outlined by the ITRS, the volumetric power density can be shown to scale as  $L_g^{-1.7}$ , where  $L_g$  is the physical channel length of the transistor [8]. This basic trend is shown in Fig. 1.4 for sub-20 nm single gate (SG), double gate (DG) and triple gate (TG) thin body silicon

transistors. The upper two curves correspond to ITRS [9] values for TG and DG devices and the lower curve represents the scaling set forth in [10] for TG, DG, and SG devices. The volumetric power density of a 20 nm device is of the order  $10 \text{ TW/cm}^3$  and a 6 nm device at the end of the roadmap is expected to increase by one order of magnitude. The calculations assume uniform current distribution [7] and that heat generation in the S/D fin extension dominates as we will discuss in Chapter 5.

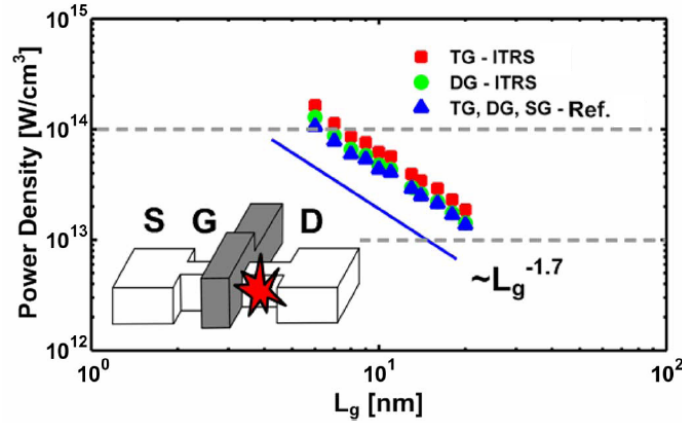


Figure 1.4: Volumetric heat generation rates as a function of transistor channel length through the end of the existing ITRS roadmap [8].

Reducing channel lengths  $L_g$  in order to increase packing density and to reduce the energy-delay product [10] has a direct impact on the departure from equilibrium of the electron and phonon systems within devices, thus increasing both the complexity in modeling the electrothermal behavior. If the rate of energy conversion becomes sufficiently strong within a volume of material such that a quasi-equilibrium can not be maintained between all particles, energy may accumulate in a particular segment of the multi-particle distribution. In such a case, a single temperature no longer suffices in characterizing the microscopic ensemble. A subgroup of particles may become “hot” relative to some background particle distribution until the system can relax to a near-equilibrium state by inter-particle scattering and conversion of energy between the particle sub-systems. Figure 1.5 shows an example of this effect in the electron system within a nanoscale transistor. On the left of the figure, the electron potential energy function is drawn for two different cases, with zero bias (top) and with an

applied gate to source  $V_{GS}$  and drain to source  $V_{DS}$  voltage equal to the supply voltage  $V_{CC}$ . On the right hand side, the electron distribution in energy is sketched.

When no bias is applied (top image), the electrons in the system are in thermal equilibrium with themselves and the lattice and therefore have a common temperature throughout the system. When a bias is applied (bottom), a portion of the electron distribution in the source (S) is injected into the channel region and are subject to the large electric field and potential dropped across the channel. At very short time scales, e.g. less than 100 fs, electrons do not typically have time to exchange energy and momentum. As a consequence, the segment of the electron distribution which is injected into the channel will have a distribution resembling that of a thermal distribution which is displaced in energy by the amount of potential energy dropped in the channel.

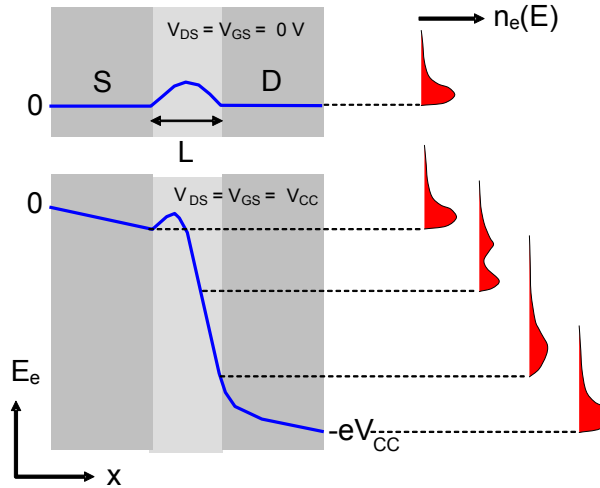


Figure 1.5: The effects of high-density energy conversion in small volumes leads to non-equilibrium particle distribution functions.

The electron system can then be seen as consisting of two sub-populations, a nonequilibrium component and a near-equilibrium component. The superposition of these two systems reveals a bimodal distribution as drawn. After a few nanometers into the channel, however, the electrons will have had sufficient time to thermalize yielding an equilibrium distribution for the entire electron population but one which is characterized by an effective temperature exceeding 1000 K. The elevated temperature

is enabled by the inability to transfer energy to the lattice vibrational modes efficiently. The “hot” electron system proceeds to cool in the drain by scattering with phonons, the quanta of the lattice vibrational modes. As electrons tend to favor some phonon modes more than others, a nonequilibrium population of phonon modes may occur within the first 10 nm of the drain junction. Eventually, all of the excess energy is converted to heat and is conducted away to the heat sink of the device by long wavelength acoustic phonons. The nature of this cascading process of energy conversion will be the subject of Chapters 4 and 5.

## **1.2 Energy Conversion in Silicon Nanophotonic Devices**

Integrating optics into high-density digital information processing devices such as microprocessors, graphics chips, or network processors has been a long held dream shared by countless engineers, scientists, and technologists. Having the ability to build information systems that combine the innate advantages of silicon electronics with those of an optical communication network, would likely spawn a new era in computing capability that would mimic the accelerated growth in performance seen in the transition from mainframe computers to the introduction of the personal computer in the 1980s. Silicon has been the material choice for the electronics industry for decades but it has played a highly diminished role in the telecommunications sector for fundamental reasons. Because silicon is an indirect bandgap semiconductor, meaning it requires a phonon, an impurity or some lattice defect to assist in the collection and generation of light, it has relatively poor optoelectronic properties compared to the III-V semiconductors such as GaAs and InP. Despite its limitations, silicon is increasingly finding its way into optoelectronic systems in the form of waveguides, modulators, detectors, and filters. The critical optoelectronic component missing, however, is an efficient electrically pumped, coherent silicon light source.

A promising path forward to developing an all-silicon CMOS compatible source is through the modification of the intrinsic optoelectronic properties of silicon by reducing the physical size of the crystal to only a few nanometers on a side. In doing

so, relatively efficient photo- and electro-luminescence have been reported on account of momentum uncertainty created by the confined electronic wavefunctions in and radiative surface states – the latter being particularly true for silicon nanoparticles embedded in silicon nitride host. Despite the great interest in their optical properties, the energy losses due to nonradiative processes such as free-carrier absorption (FCA) and fast multi-charge Auger recombination severely limit the prospects of achieving population inversion and consequently the point of reaching the condition of lasing. Progress towards understanding these intrinsic loss mechanisms which compete unfavorably with stimulated emission has been greatly impeded by the fact that the systems of silicon nanocrystals typically produced are inhomogeneous and are plagued with complicated carrier dynamics physics. To make progress in this field, there needs to be a complete understanding of how the energy of an excitation source, either an electrical or optical pump, is converted to the either desired light output, the creation of defects, or the generation of heat.

Chapter 7 discusses my work identifying some of the key physical processes which limit population inversion in dense silicon nanocrystal composite films. In support of these efforts, a custom, two-color optical pump-probe system with sub-10 picosecond temporal resolution is developed to measure the carrier dynamics of silicon NCs films through indirect probing of the excited carrier intraband absorption.

### **1.3 Overview**

Chapter 2 reviews the fundamentals of energy conversion in bulk and finite nanostructures. Chapter 3 discusses the electron and lattice vibrational energy eigenstates of bulk and finite nanometer scale silicon crystals which were modeled using semi-empirical numerical techniques. Chapters 4 and 5 present my work in advancing the state-of-the-art in modeling detailed energy conversion and transport in quasi-ballistic silicon electronic devices. Such highly scaled quasi-1D transistors are routinely operated under high-field, and thus highly nonequilibrium, conditions where both the electrons and phonons deviate substantially from thermal equilibrium

distributions. Chapter 4 presents calculations of optical phonon lifetimes which are based on the full-band dispersion calculations discussed in Chapter 3. Chapter 5 develops rigorous, fully-coupled, self-consistent, electron-phonon Monte Carlo simulations which culminate in the simulation of a realistic 20 nm silicon devices operating under typical bias conditions. From this work, we are able to assess the magnitude of the role in which hot optical phonons will play in silicon devices within the horizon of the latest semiconductor roadmap.

Chapters 6 and 7 are devoted to describing the development of a two-color laser pump-probe technique with picosecond time resolution and its use in quantifying nonradiative energy conversion in dense systems of silicon nanocrystals embedded in host amorphous dielectric films. The unique dependence of the excited carrier losses vs. probe delay and pump intensity is used to garner new insights into the physics of energy transfer between nanoparticles and a non-local electron-electron interaction model is developed for bulk and low-dimensional systems.

Chapter 8 concludes this dissertation by summarizing the major contributions of this work. A proposal for new research is proposed which builds on the results of this work.

## Chapter 2

### Fundamentals of Nanoscale Energy Conversion

I must say that I also do not like indeterminism. I have to accept it because it is certainly the best that we can do with our present knowledge.

-P. A. M. Dirac

In this chapter, I will review the fundamental physics of microscopic energy conversion and specifically as it applies to silicon nanostructures. In doing so, we will examine the key differences between finite nanoscale crystals and their bulk equivalent, while placing special emphasis on the characteristics of the silicon lattice.

#### 2.1 Energy

In the framework of non-relativistic quantum mechanics, the total energy or Hamiltonian of a single particle such as an electron may be written as a complex differential operator  $H$  which acts on a two-component (arising from the two spin state basis +/-) particle field or wavefunction  $\Psi_{\pm}(\vec{r}, t)$  according to

$$H\Psi_{\pm}(\vec{r}, t) = i\hbar \frac{\partial}{\partial t} \Psi_{\pm}(\vec{r}, t) \quad (2.1)$$

For a single electron coupled to the electromagnetic field  $\vec{A}(\vec{r}, t)$ , but neglecting spin interactions, i.e.  $\nabla \times \vec{A} \sim 0$ ,  $H$  acquires the following form:

$$H = \frac{1}{2m} \left( \frac{\hbar}{i} \nabla - e\vec{A}(\vec{r}, t) \right)^2 + V(\vec{r}, t) = i\hbar \frac{\partial}{\partial t} \quad (2.2)$$

the two spin states corresponding to the same spatial wavefunction are degenerate in energy. We gain physical insight into the meaning of energy by examining the eigenfunctions of the non-relativistic Hamiltonian operator, which satisfy the equation

$$i\hbar \frac{\partial}{\partial t} \Psi(\vec{r}, t) = E \Psi(\vec{r}, t) \quad (2.3)$$

where  $E$  is a constant. The eigenfunctions have the following form

$$\Psi(\vec{r}, t) = \Psi(\vec{r}, 0) e^{-i\theta(t)} = \varphi(\vec{r}) e^{-i\theta(t)} \quad (2.4)$$

where the temporal phase  $\theta$  is linear in time with the eigenvalue  $E$  setting the rate of angular variation according to

$$\theta(t) = (E / \hbar) t \quad (2.5)$$

An important characteristic of the energy eigenfunctions is that a measurement of the total energy of an electron residing in such an energy eigenstate is independent of time, i.e. is stationary, and will always return the value  $E$  with 100 % certainty. In mathematical terms,  $\langle \Psi | H | \Psi \rangle = \langle H \rangle = E$ , and  $\langle H^2 \rangle - \langle H \rangle^2 = 0$ .

Although  $E$  is an eigenvalue of the operator  $i\hbar \partial / \partial t$ , it is not necessarily an eigenvalue of the left-hand side of (2.3), a necessary condition if  $E$  is to represent the total energy. To generalize the case, then let  $H$  be time-dependent but for the sake of simplicity, allow for  $H(\vec{r}, t) = H_0(\vec{r}) + H'(\vec{r}, t)$  where  $H'(\vec{r}, t) = g(\vec{r})f(t)$ . Operating on the separable eigenfunction (2.4), we have

$$\begin{aligned} H(\vec{r}, t)\varphi(\vec{r})e^{-i\theta(t)} &= H_0(\vec{r})\varphi(\vec{r})e^{-i\theta(t)} + g(\vec{r})f(t)\varphi(\vec{r})e^{-i\theta(t)} \\ &= E\varphi(\vec{r})e^{-i\theta(t)} \end{aligned} \quad (2.6)$$

Then, multiplying both sides by  $\varphi^*(\vec{r})e^{+i\theta(t)}$ , integrating over all space, and using the orthonormality of the eigenfunctions, we obtain the final result

$$E(t) = \langle H_0 \rangle + f(t)\langle g \rangle = E_0 + f(t)\langle g \rangle \quad (2.7)$$

where

$$\langle g \rangle = \int d\vec{r} \varphi^*(\vec{r}) g(\vec{r}) \varphi(\vec{r}) \quad (2.8)$$

is referred to as the *matrix element*. We conclude that, if the Hamiltonian operator has a time-dependence, then  $E$  must also depend on time and the assumption  $E = \text{constant}$  fails to hold. Despite this contradiction, the form of the wavefunction is maintained albeit with the phase angle modified to include an additional time-dependent term:

$$\theta(t) = (1/\hbar)E_0t + (1/\hbar)\langle g \rangle \int_0^t f(t')dt' \quad (2.9)$$

Evidently, the time-dependent part of  $H$  changes the energy of the particle continuously and manifests itself as a continuous change in the quantum mechanical phase angle, which depends explicitly on the temporal history of  $H(t)$ . Clearly, the total energy of the electron is not conserved in such a case; the time-dependence of  $H$  is the result of an external *force* pumping energy into (+) or out of (-) the system. This force must arise from the interaction with another particle or the interactions of a group of particles not explicitly described in the specified system.

For a time-independent  $H$ , there is an infinite set of linearly independent energy eigenfunctions which span the *Hilbert space*. The complete time-dependent

wavefunction can be expressed as a linear combination of these orthonormal basis states as

$$\Psi(\vec{r}, t) = \sum_n c_n \varphi_n(\vec{r}) e^{-i\theta_n(t)} \quad (2.10)$$

The expected value of the total energy for such a state is

$$\langle H \rangle = \sum_n E_n |c_n|^2 = E = \text{constant} \quad (2.11)$$

However, there is now an intrinsic uncertainty in the total energy as quantified by the variance  $\sigma_E^2$

$$\sigma_E^2 = \langle H^2 \rangle - \langle H \rangle^2 = \frac{1}{N} \sum_{m,n} |c_n|^2 \left[ E_n^2 - N E_m E_n |c_m|^2 \right] \quad (2.12)$$

Here,  $N$  is the total number of basis states necessarily included in the expansion to describe the state. Considering the normalization condition,  $\sum |c_n|^2 = 1$ , it can easily be shown that if  $|c_n|^2 = 1$  for any  $n$ , i.e. the electron is fully described by a single energy eigenfunction, then  $\sigma_E^2 = 0$  as expected.

A measurement of an observable  $Q$  performed on a superposition state  $\Psi$  generally will have a time-dependence, irrespective of whether the operator  $Q$  includes a time dependence or not. For example, take the case of an electron residing in a superposition of two energy eigen states,

$$\Psi(\vec{r}, t) = c_1 \varphi_1(\vec{r}) e^{-iE_1 t / \hbar} + c_2 \varphi_2(\vec{r}) e^{-iE_2 t / \hbar} \quad (2.13)$$

where, by the normalization condition we have  $|c_1|^2 + |c_2|^2 = 1$ . If we perform the measurement  $\langle Q \rangle = \langle \Psi | Q | \Psi \rangle$ , and  $Q$  has no explicit time dependence, we obtain the general result

$$\langle Q \rangle = |c_1|^2 Q_{11} + |c_2|^2 Q_{22} + |c_1| |c_2| Q_{21} \cos(E_{21}t/\hbar - \delta_{21}) \quad (2.14)$$

where  $\delta_{21} = \angle c_2 - \angle c_1$ ,  $E_{21} = E_2 - E_1$ ,  $Q_{ij} = \langle \varphi_i | Q | \varphi_j \rangle$  and  $Q_{ij} = Q_{ji}^*$  due to the assumed Hermiticity of the observable  $Q^{\dagger\dagger}$ . For example, if  $Q = \delta(\vec{r} - \vec{r}')$ , the measurement  $\langle \Psi | \delta(\vec{r} - \vec{r}') | \Psi \rangle = |\Psi(\vec{r}', t)|^2$ , yields the probability of finding the particle in a volume  $d\vec{r}'$  about  $\vec{r}'$  and depends on time according to the relation

$$|\Psi(\vec{r}', t)|^2 = |c_1|^2 |\varphi_1(\vec{r}')|^2 + |c_2|^2 |\varphi_2(\vec{r}')|^2 + 2|c_1| |c_2| \varphi_1(\vec{r}') \varphi_2(\vec{r}') \cos(E_{21}t/\hbar - \delta_{21}) \quad (2.15)$$

The energy difference  $E_{21}$  dictates how rapidly the expected measurement value changes as it sets the period of oscillation of the phase angle. Finally, if  $Q = H$  and  $H$  is time-independent, then the measurement  $\langle H \rangle$  yields

$$\langle H \rangle = |c_1|^2 E_1 + |c_2|^2 E_2 = E = \text{constant} \quad (2.16)$$

as  $H$  is diagonalized ( $H_{ii} = E_i$ ) through the choice of the energy eigenfunction basis. By use of (2.12) and the requirement  $|c_1|^2 = 1 - |c_2|^2$ , the intrinsic uncertainty in the measurement of the total energy is found to increase in proportion to the energy separation between the two states,

$$\sigma_E = |c_1| (1 - |c_1|^2)^{1/2} E_{21} \quad (2.17)$$

---

<sup>$\dagger\dagger$</sup>   $Q$  must be Hermitian if it is to represent an observable quantity.

where  $E_{21} = E_2 - E_1$ . By inspection, the maximum uncertainty is  $E_{21}/2$  and occurs when  $|c_1|^2 = |c_2|^2 = 1/2$ .

Having discussed what energy is within the framework of non-relativistic quantum mechanics, and identified the special properties arising when choosing to represent the particle wavefunction in the energy eigenfunction basis, we are now in position to describe the exchange of energy between particles and systems of particles. We begin with establishing the mechanism for exchange of energy between an isolated particle and an “external” force. We then proceed to establish how energy is exchanged between interacting particles within a closed system and how the distribution of energy amongst a population of interacting particles can be adequately characterized.

## 2.2 Energy Conversion

In the previous section, we saw that if the Hamiltonian of a system is independent of time, then the expectation value of the total energy of that system will be constant. Furthermore, this result holds even if there is an intrinsic uncertainty in the total energy due to the wavefunction being described as a superposition of eigenstates (2.10). In essence, the constancy of  $\langle H \rangle$  is a quantum mechanical statement of energy conservation; for the average total energy to be a conserved quantity, the Hamiltonian describing all particles and their interactions within the volume of a closed system must be time-independent. Particles may interact and exchange energy within the system but the total expected value of energy must be conserved at all times. In this section, we build the foundations for describing the exchange of energy between particles.

Consider the case of a single electron in a closed system whose wavefunction is expressed as a superposition of the complete set of energy eigenstates of  $H_0(\vec{r})$  having eigenvalues  $E_n$ . At time  $t = 0$ , an external force is turned on adding a contribution to the total Hamiltonian operator which can be described as

$H'(\vec{r}, t) = g(\vec{r})f(t)$ . The amplitudes of the basis states are now time-dependent so that the description of the electron can be written as

$$\Psi(\vec{r}, t) = \sum_n c_n(t) \varphi_n(\vec{r}) e^{-iE_n t / \hbar} \quad (2.18)$$

The fact that the amplitudes are time-dependent means that the probability of finding the particle in a particular eigenstate at time  $t$ , given by  $|c_n(t)|^2$ , is changing with time. This is equivalent to saying that the particle makes transitions between eigenstates of different energy thereby exchanging energy with the external force in the process. The explicit time-dependence of these coefficients is found by solving the time-dependent Schrödinger equation as follows:

$$[H_0 + g(\vec{r})f(t)] \sum_n c_n(t) \varphi_n(\vec{r}) e^{-iE_n t / \hbar} = i\hbar \frac{\partial}{\partial t} \sum_n c_n(t) \varphi_n(\vec{r}) e^{-iE_n t / \hbar}$$

$$\sum_n c_n(t) g(\vec{r}) f(t) \varphi_n(\vec{r}) e^{-iE_n t / \hbar} = i\hbar \sum_n \dot{c}_n(t) \varphi_n(\vec{r}) e^{-iE_n t / \hbar}$$

Using the orthonormality of the energy eigenfunctions of the closed system, we multiply both sides by  $\varphi_m^*(\vec{r}) e^{+iE_m t / \hbar}$  and integrate over all space to obtain the following result

$$\dot{c}_m(t) = \frac{1}{i\hbar} \sum_n c_n(t) f(t) g_{mn} e^{+iE_{mn} t / \hbar} \quad (2.19)$$

where the matrix elements are given by

$$g_{mn} = \int d\vec{r} \varphi_m^*(\vec{r}) g(\vec{r}) \varphi_n(\vec{r}) \quad (2.20)$$

Integrating over time, we find the amplitude of the  $m^{\text{th}}$  basis state to be

$$c_m(t) = \frac{1}{i\hbar} \sum_n g_{mn} \left\{ \int_0^t dt' c_n(t') f(t') e^{+iE_{mn}t'/\hbar} \right\} + c_m(0) \quad (2.21)$$

The explicit time-dependence of the amplitudes can be determined using an array of available numerical techniques. It should be noted that solving the coupled equations for the instantaneous values of the amplitudes is equivalent, albeit with a potential reduction in the parameter space, to solving the full time-dependent Schrödinger equation using a finite difference method. However, the mapping of the instantaneous solution onto the energy eigenstates provides the physical intuition of electrons making discrete transitions between these basis states over time as well as may lead to important simplifying assumptions.

Once the amplitudes are determined, the average rate of energy conversion due to the external force can be calculated using the following expression<sup>\*\*</sup> (Appendix B)

$$\frac{d}{dt} \langle H \rangle = \sum_n E_n \frac{d}{dt} |c_n(t)|^2 + i\hbar \sum_n \frac{d}{dt} [c_n^*(t) \dot{c}_n(t)] \quad (2.22)$$

Another important characteristic is the instantaneous spread or uncertainty in the system energy which can be measured as the energy variance which can be shown to be given by

---

<sup>\*\*</sup> The second series term in 2.22 is not intuitive and is often neglected by assumption of strong de-coherence.

$$\sigma_E^2(t) = (1/N) \sum_{m,n} \left\{ \begin{aligned} &E_n^2 |c_n|^2 - \hbar^2 c_n^* \ddot{c}_n + i\hbar E_n c_n^* \dot{c}_n \\ &- N[E_m E_n |c_m|^2 |c_n|^2 + 2i\hbar E_m |c_m|^2 c_n^* \dot{c}_n - \hbar^2 c_m^* c_n^* \dot{c}_m \dot{c}_n] \end{aligned} \right\} \quad (2.23)$$

It is easy to show that (2.23) reduces to the time-independent expression (2.12) when the amplitudes themselves are time-independent.

We now discuss the most widely used approximation technique for determining the time-dependent amplitudes  $c_n(t)$ . This approximation, referred to as *time-dependent perturbation theory* (TDPT), has been employed with great success since the early days of Quantum Mechanics. The principle assumption enabling the use of TDPT is that the magnitude of the time-dependent component of the  $H$  is small compared to the energy separation between the interacting states. Consequently, the external force can be treated as a perturbation. Furthermore, each basis state amplitude may then be expanded as a perturbation expansion. For book-keeping purposes, the amplitudes are expanded as a power series using a perturbation series parameter  $\lambda$ , which we set to unity at the end of the analysis. We begin with the substitution

$$H'(\vec{r}, t) \rightarrow \lambda H'(\vec{r}, t) \quad (2.24a)$$

and for each amplitude, generate the series

$$c_n(t) = c_n^{(0)}(t) + \lambda c_n^{(1)}(t) + \lambda^2 c_n^{(2)}(t) + \lambda^3 c_n^{(3)}(t) + \dots + \lambda^p c_n^{(p)}(t) + \dots \quad (2.24b)$$

where  $^{(p)}$  indicates the order of the expansion. Inserting (2.24) into (2.19), we get

$$[\dot{c}_m^{(0)}(t) + \lambda \dot{c}_m^{(1)}(t) + \lambda^2 \dot{c}_m^{(2)}(t) + \dots] = \frac{1}{i\hbar} \sum_n [c_n^{(0)}(t) + \lambda c_n^{(1)}(t) + \lambda^2 c_n^{(2)}(t) + \dots] \lambda f(t) g_{mn} e^{+iE_{mn}t/\hbar}$$

Equating like-powers of  $\lambda$  and then allowing  $\lambda \rightarrow 1$ , we obtain the varying orders of the perturbation expansion:

$$\dot{c}_m^{(0)}(t) = 0 \rightarrow c_m^{(0)}(t) = \text{constant} \quad (2.25a)$$

$$\dot{c}_m^{(p)}(t) = \frac{1}{i\hbar} \sum_n c_n^{(p-1)}(t) f(t) g_{mn} e^{+iE_{mn}t/\hbar} \quad ; \quad p > 0 \quad (2.25b)$$

If we assume that the electron initially resides in one of the eigenstates labeled  $i$ , then the coefficient for a final state labeled  $f$  is given by the sequence

$$c_f(t) = c_f^{(0)}(t) + c_f^{(1)}(t) + c_f^{(2)}(t) + \dots$$

It follows from 2.25 that

$$c_f^{(0)}(t) = \begin{cases} 1 & ; f=i \\ 0 & \end{cases} \quad (2.26a)$$

$$\dot{c}_f^{(1)}(t) = \frac{1}{i\hbar} f(t) g_{fi} e^{+iE_{fi}t/\hbar} \quad (2.26b)$$

$$\dot{c}_f^{(2)}(t) = \frac{1}{i\hbar} \sum_n c_n^{(1)}(t) f(t) g_{fn} e^{+iE_{fn}t/\hbar} \quad (2.26c)$$

where  $n$  is an index running over all intermediate states. Furthermore, we set the initial conditions to be

$$c_f^{(p)}(0) = \begin{cases} 1 & ; f=i ; p=0 \\ 0 & ; \text{else} \end{cases} \quad (2.26d)$$

Figure 2.1 illustrates schematically the three lowest order terms in the expansion. In the first-order correction, only direct paths between the initial and final state of interest enter into the expression. In the second-order expansion, the summation includes intermediate paths. The total amplitude to take an electron from the initial (state 1) to the final state (state 3) includes a summation of all possible paths.

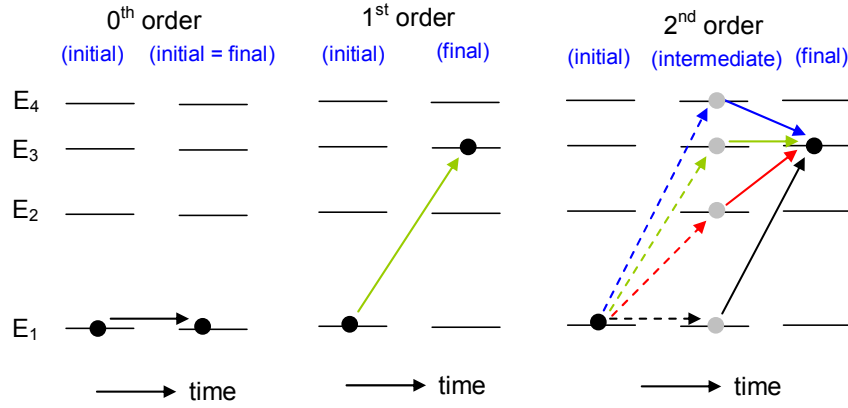


Figure 2.1: Schematic representation of the three lowest order perturbation expansion terms concerning the probability amplitude of taking the electron from state 1 with energy  $E_1$  to state 3 with energy  $E_3$  when there are a total of four eigenstates of the system. The first order term describes only the direct path whereas the second order includes all intermediate paths. The intermediate states are also referred to as “virtual” states.

As an example, consider once again the two-level system. From (2.25), we arrive at the two coupled differential equations:

$$\dot{c}_1(t) = \frac{1}{i\hbar} \left\{ c_1(t) f(t) g_{11} + c_2(t) f(t) g_{21}^* e^{-iE_{21}t/\hbar} \right\} \quad (2.27a)$$

$$\dot{c}_2(t) = \frac{1}{i\hbar} \left\{ c_1(t) f(t) g_{21} e^{+iE_{21}t/\hbar} + c_2(t) f(t) g_{22} \right\} \quad (2.27b)$$

2.27 can be solved exactly using standard finite difference techniques or approximately using TDPT.

Assuming that the electron starts out in state 1, the three lowest order TDPT approximations are, in accordance with 2.26,

**0<sup>th</sup> order**

$$c_1^{(0)}(t) = 1 \quad (2.28a)$$

$$c_2^{(0)}(t) = 0 \quad (2.28b)$$

**1<sup>st</sup> order**

$$c_1^{(1)}(t) = \frac{1}{i\hbar} g_{11} \int_0^t dt' f(t') \quad (2.28c)$$

$$c_2^{(1)}(t) = \frac{1}{i\hbar} g_{21} \int_0^t dt' f(t') e^{+iE_{21}t'/\hbar} \quad (2.28d)$$

**2<sup>nd</sup> order**

$$\begin{aligned} c_1^{(2)}(t) = & -\frac{1}{\hbar^2} g_{11} g_{11} \int_0^t \int_0^{t''} dt'' dt' f(t'') f(t') \\ & -\frac{1}{\hbar^2} g_{21} g_{12} \int_0^t \int_0^{t''} dt'' dt' f(t'') f(t') e^{+iE_{21}(t''-t')/\hbar} \end{aligned} \quad (2.28e)$$

$$c_2^{(2)}(t) = -\frac{1}{\hbar^2} g_{21} [g_{11} + g_{22}] \int_0^t \int_0^{t''} dt'' dt' f(t'') f(t') e^{+iE_{21}(t''+t')/\hbar} \quad (2.28f)$$

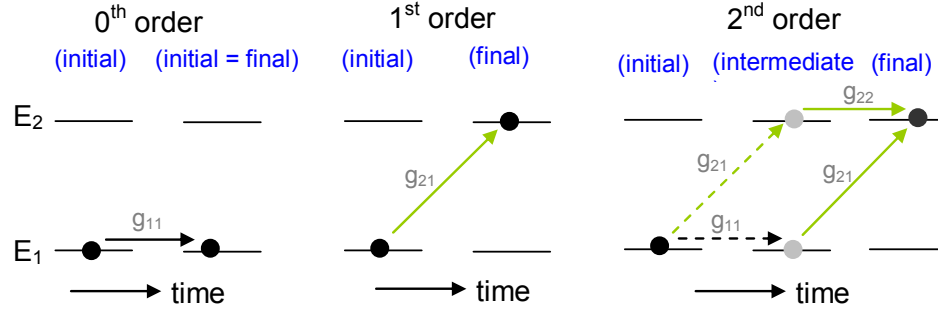


Figure 2.2: Schematic representation of the one electron, two state example described by (2.28).

In general, the time-dependent part of the perturbation  $f(t)$  can be expanded in the Fourier basis,

$$f(t) = \int d\omega F(\omega) e^{i\omega t} \quad (2.29)$$

thus allowing us to transform 2.29d into

$$c_2^{(1)}(t) = \frac{1}{i\hbar} g_{21} \int d\omega \left\{ F(\omega) \int_0^t dt' e^{+i(\omega + \omega_{21})t'} \right\} \quad (2.30)$$

If  $f(t)$  takes the simple form of a sinusoidal perturbation, e.g.

$$f(t) = 2A \cos(\omega_0 t) ; F(\omega) = A[\delta(\omega + \omega_0) + \delta(\omega - \omega_0)] \quad (2.31)$$

then 2.28d becomes

$$c_2^{(1)}(t) = \frac{2A}{i\hbar} g_{21} \left\{ e^{+i(\omega_{21}-\omega_0)t/2} \frac{\sin[(\omega_{21}-\omega_0)t/2]}{\omega_{21}-\omega_0} + e^{+i(\omega_{21}+\omega_0)t/2} \frac{\sin[(\omega_{21}+\omega_0)t/2]}{\omega_{21}+\omega_0} \right\} \quad (2.32)$$

The probability of finding the particle in state 2 after a time  $t$  has elapsed is found by taking the modulus square of 2.32 to get

$$|c_2^{(1)}(t)|^2 = \frac{4A^2}{\hbar^2} |g_{21}|^2 \left\{ \frac{\sin^2[(\omega_{21}-\omega_0)t/2]}{(\omega_{21}-\omega_0)^2} + \frac{\sin^2[(\omega_{21}+\omega_0)t/2]}{(\omega_{21}+\omega_0)^2} + \chi(t) \right\} \quad (2.33a)$$

where  $\chi(t)$  represents the cross-terms:

$$\chi(t) = \frac{1}{2} \left\{ \frac{1 - \cos[(\omega_{21}-\omega_0)t] + \cos[(2\omega_0)t] - \cos[(\omega_{21}+\omega_0)t]}{\omega_{21}^2 - \omega_0^2} \right\} \quad (2.33b)$$

The first term in the brackets of 2.33a will clearly dominate over the second term near resonance, i.e.  $\omega_0 \approx \omega_{21}$ , provided  $\omega_{21} > 0$  and  $|\omega_{21} - \omega_0| \ll |\omega_{21} + \omega_0|$  thus allowing us to safely ignore the second term. This is almost always the case since the far-from resonant condition would place the probability of a transition to be of order  $\sim |g_{21}|^2 / E_{21}^2$  or smaller. The third term in the brackets represents a rapidly oscillating term, which on resonance, has a high-frequency oscillation of  $\omega \sim 3\omega_0 + \omega_{21}$  and diminished magnitude  $\chi(t) \rightarrow \sin(2\omega_{21}t)/2\omega_{21}$  which makes it small compared to the first term. Due to the diminished magnitude and the fact that the rapid oscillations wash out in the presence of system decoherence, we eliminate all but the first term. We are now left with the expression

$$|c_2^{(1)}(t)|^2 = \frac{4A^2}{\hbar^2} |g_{21}|^2 \left\{ \frac{\sin^2[(\omega_{21} - \omega_0)t/2]}{(\omega_{21} - \omega_0)^2} \right\} \quad (2.34)$$

We see that the probability of finding the electron in state 2 oscillates between a minimum value 0 and a maximum of  $A^2 |g_{21}|^2 / (E_{21} - E_0)^2 \ll 1$ . We require that the maximum be much less than unity in order for the use of TDPT to be valid. Furthermore, the rate of change in the probability, or rather the transition rate in going from state 1 to 2,  $r_{21}$ , is also oscillatory, and is given by taking the derivative of 2.34.

$$\begin{aligned} r_{21} = \frac{d}{dt} |c_2^{(1)}(t)|^2 &= \frac{4A^2}{\hbar^2} |g_{21}|^2 \frac{d}{dt} \left\{ \frac{\sin^2[(\omega_{21} - \omega_0)t/2]}{(\omega_{21} - \omega_0)^2} \right\} \\ &= \frac{4A^2}{\hbar^2} |g_{21}|^2 \left\{ \frac{\sin[(\omega_{21} - \omega_0)t/2] \cos[(\omega_{21} - \omega_0)t/2]}{(\omega_{21} - \omega_0)} \right\} \\ &= \frac{2A^2}{\hbar^2} |g_{21}|^2 \left\{ \frac{\sin[(\omega_{21} - \omega_0)t]}{(\omega_{21} - \omega_0)} \right\} \end{aligned} \quad (2.35)$$

The oscillation frequencies for the two quantities are  $(\omega_{21} - \omega_0)/2$  and  $\omega_{21} - \omega_0$  respectively near resonance and are predicted to go to zero when  $\omega_0$  is on resonance within the framework of TDPT. However, a non-perturbative analysis shows that the oscillation frequency of  $|c_2^{(1)}(t)|^2$  does not fall to zero on resonance. Rather, it is described by the Rabi (flopping) frequency (2.36) which reaches a minimum value of  $|g_{21}|/(2\hbar)$  at resonance<sup>§§</sup>.

$$\omega_r = (1/2) \sqrt{(\omega_{21} - \omega_0)^2 + |g_{21}|^2 / \hbar^2} \quad (2.36)$$

---

<sup>§§</sup> We note that sensitive measurements of  $|g_{21}|$  can be made by performing reflection spectroscopy on quasi-two-level systems such as III-V semiconductor quantum dots coupled to high-Q cavities and using 2.36.

The oscillatory behavior of 2.34 and 2.35 is the result of a *coherent* perturbation, that is the perturbation has a stable (either constant or slowly varying) phase angle. If instead the perturbation is the sum of a number of incoherent sinusoidal sources, i.e. having no correlation in phase angle, then we can add the probabilities rather than taking the modulus squared of the amplitudes. In doing so, the additional cross-terms which arise in 2.33 rapidly average to zero. Converting  $A^2(\omega)$  into a spectral density (in units of inverse frequency), we have for the incoherent case the following expression

$$|c_2^{(1)}(t)|^2 = \frac{4}{\hbar^2} |g_{21}|^2 \left\{ \int d\omega A^2(\omega) \frac{\sin^2[(\omega_{21} - \omega)t/2]}{(\omega_{21} - \omega)^2} \right\} \quad (2.37)$$

Furthermore, since the function  $\sin^2[(\omega_{21} - \omega)t/2]/(\omega_{21} - \omega)^2$  is strongly peaked around the value  $\omega = \omega_{21}$ , having a time-dependent characteristic bandwidth of  $\Delta\omega = 4\pi/t$ , we can make remove  $A^2(\omega)$  from the integral by making the approximation  $A^2(\omega) \approx A^2(\omega_{21})$  to get

$$|c_2^{(1)}(t)|^2 = \frac{4}{\hbar^2} |g_{21}|^2 A^2(\omega_{21}) \left\{ \int d\omega \frac{\sin^2[(\omega_{21} - \omega)t/2]}{(\omega_{21} - \omega)^2} \right\} \quad (2.38)$$

Finally, if the perturbation is turned on for a time long enough such that  $\Delta\omega/\omega_{21} \ll 1$ , the integral in 2.38 can be converted to

$$\lim_{t \rightarrow \infty} \int_{-\infty}^{+\infty} d\omega \frac{\sin^2[(\omega_{21} - \omega)t/2]}{(\omega_{21} - \omega)^2} = (t/2) \int_{-\infty}^{+\infty} dx \frac{\sin^2(x)}{x^2} = \frac{\pi}{2} t \int_{-\infty}^{+\infty} dx \delta(x) \quad (2.39)$$

where  $x \equiv (\omega - \omega_{21})t/2$ . We then arrive at the expression

$$|c_2^{(1)}(t)|^2 = \frac{2\pi}{\hbar^2} |g_{21}|^2 A^2(\omega_{21})t \quad (2.40)$$

The probability is linear with time and thus the transition rate is constant and given by

$$r_{21} = \frac{2\pi}{\hbar^2} |g_{21}|^2 A^2(\omega_{21}) \quad (2.41a)$$

Alternatively, normalizing the spectral density to unity, i.e.  $A^2(\omega) = 1$ , we have the more common form<sup>\*\*\*</sup>

$$r_{21} = \frac{2\pi}{\hbar} |g_{21}|^2 \delta(E - E_{21}) \quad (2.41b)$$

Expression 2.42, and more commonly 2.42b, is referred to as Fermi's Golden Rule and is perhaps the most practically useful expression arising from all of Quantum Mechanics. We will make great use of this expression throughout the work presented in this thesis. A key feature of 2.42 is that energy is strictly conserved as indicated explicitly by the delta function 2.42b. Recall that the use of the delta function relied upon the assumption that the perturbation was turned on for a very long time satisfying the condition  $E_{21}t \gg 4\pi\hbar$ . At very short time scales where this condition is not met a transition can occur even if the perturbation is not on resonance. This is because there is an uncertainty in the energy of the perturbation due its having a finite duration.

---

<sup>\*\*\*</sup> Note that the property  $\delta(ax) = \delta(x)/a$  was used here.

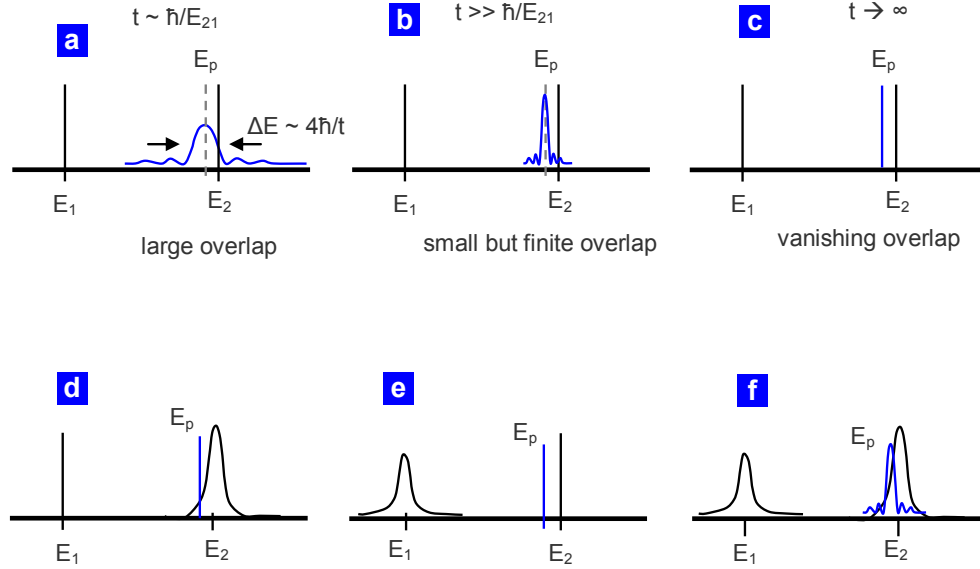


Figure 2.3: Illustration of how the uncertainty in the energy of a perturbation (caused here by the assumption that it is turned on abruptly at  $t = 0$ ) leads to the equivalence of energy conservation breaking at short time scales. The energy of the coherent wave is given by  $E_p$  at long time scales. When  $t \sim 4\pi\hbar/E_{21}$ , there is an appreciable overlap of the spectral density function of the perturbation. However, this overlap vanishes as  $t$  tends to infinity.

For a one electron, multi-level system exposed to an incoherent source such as we have been discussing, we can write a system of coupled equations describing the probability of finding the electron in any one of the levels described by an index  $i$  and  $f$  as follows

$$\frac{dp_i}{dt} = \sum_f (r_{if} p_f - r_{fi} p_i) \quad ; \quad r_{if} = r_{fi} = 1/\tau \quad (2.42)$$

with the requirement that  $\sum_i p_i = 1$ . Furthermore, the expected value of energy is

$$\langle E \rangle = \sum_i p_i E_i \quad (2.43)$$

which can be compared to 2.22. Taking the derivative, we obtain the rate of energy conversion to be

$$\frac{d\langle E \rangle}{dt} = \sum_i E_i \frac{dp_i}{dt} = \sum_{i,f} E_i (r_{if} p_f - r_{fi} p_i) \quad (2.44)$$

For the two-level system, we obtain

$$\frac{d\langle E \rangle}{dt} = \frac{E_{21}}{\tau} [p_1 - p_2] = \frac{E_{21}}{\tau} [2p_1 - 1] = \frac{E_{21}}{\tau} [1 - 2p_2] \quad (2.45)$$

Intuitively, we find that the maximum average rate of energy conversion is  $\pm E_{21}/\tau$ . For a continuous or quasi-continuous distribution of energy levels, the instantaneous probability of finding an electron within an energy  $E$  is given by the distribution function,  $f(E, t)$ . At equilibrium,  $t$

$$f = \frac{1}{e^{(E-\mu)/\beta} + 1} \quad (\text{Fermi-Dirac}) \quad (2.46a)$$

$$f = \frac{1}{e^{E/\beta} - 1} \quad (\text{Bose-Einstein}) \quad (2.46b)$$

where  $\beta = k_B T$ .

$$\frac{du_E}{dt} = \frac{d}{dt} \int E \cdot f(E, t) g(E) dE \quad (2.47)$$

where  $f(E, t)$  is the time-dependent particle distribution function.

## 2.3 Relevant Length and Time Scales

In the previous section, we saw that the transition rate due to a coherent perturbation exhibited rapid oscillations at short time-scales,  $t \sim 2h/E_{21}$ . Recalling Eq. 2.17, we generalize this notion to multi-level systems by defining a *coherence time* as

$$t_c \sim h/\sigma_E \quad (2.48)$$

The uncertainty in energy can be thought to arise from either of two sources or a combination of the two: (1) it may arise due to the creation of a superposition state as in the example in the previous section leading to the result 2.17 even in the absence of a perturbation<sup>†††</sup> or (2) it arises due a perturbation that has a finite duration and therefore a finite energy spectral bandwidth. For a thermally equilibrated state, the uncertainty in energy of an electron will be of order  $\sigma_E \sim k_B T$ . Thus we can estimate the coherence time to be  $t_c \sim h/k_B T$ .

As a natural extension to the concept of temporal coherence, we can define a coherence length  $l_c$  as the mean displacement of the centroid of a particle wavepacket occurring in a time equal to  $t_c$  and is given by the expression

$$l_c = v_g t_c \quad (2.49)$$

where  $v_g$  is the magnitude of the group velocity with the wavepacket centered in momentum space around the wavevector  $\bar{k}_0$  and mean energy  $E_0$ . The vector relation can be obtained by taking the gradient in k-space of the energy about a point in k-space.

---

<sup>†††</sup> Or, the perturbation that produced the superposition state has been turned off.

$$\vec{v}_g = \frac{1}{\hbar} \nabla_{\vec{k}} E \Big|_{\vec{k}=\vec{k}_0} \quad (2.50)$$

For an electron having a free-particle-like bandstructure  $E = \hbar^2 k^2 / 2m^*$ . The group velocity of an electron having an average thermal energy of  $1.5 \cdot k_B T$ , is

$$v_g = \frac{\hbar k}{m^*} \sim \sqrt{\frac{3k_B T}{m^*}} \quad (2.51)$$

It follows that the coherence length is of order

$$l_c \sim \sqrt{\frac{3\hbar^2}{m^* k_B T}} \quad (2.52)$$

Furthermore, we may define a mean-free-path  $\Lambda$  as the average displacement of the wavepacket before experiencing a momentum-randomizing scattering event with an average rate of  $R = 1/\tau$ .

$$\Lambda = v_g \tau \quad (2.53)$$

The mean relaxation time can be extracted from electrical or thermal conductivity measurements using the following relations

$$\sigma_e = \frac{e^2 N_D \tau_{el}}{m^*} \rightarrow \tau_{el} = \frac{\sigma_e m^*}{e^2 N_D} \quad (2.54a)$$

$$k = \frac{1}{3} C_v v_g^2 \tau_{ph} \rightarrow \tau_{ph} = \frac{3k}{C_v v_g^2} \quad (2.54b)$$

where  $m^*$  is taken to be the conductivity effective mass, which is found by taking a weighted average over the six conduction band valleys using (2.55a).

$$m_c^* = \left( \frac{1}{3m_l^*} + \frac{2}{3m_t^*} \right)^{-1} = 0.259 \cdot m_e \quad (2.55a)$$

$$m_{(DOS)}^* = 6^{2/3} (m_l^* m_t^{*2})^{1/3} = 1.182 \cdot m_e \quad (2.55b)$$

To determine the effective group velocity of an electron near the conduction band minimum of silicon, we make use of the density of states effective mass given by 2.55b in combination with (2.51). Tables 2.1 and 2.2 provide estimates for average values of  $v_g$ ,  $t_c$ ,  $l_c$ , and  $\Lambda$  for electrons and phonons in bulk silicon at 77 K and 300 K. Note that for equilibrium conditions, the average coherence time is common to all particle types within a system at equilibrium, provided that the distribution functions can be approximated using a Boltzmann distribution, i.e. when the temperatures are far from absolute zero.

Table 2.1: Estimates of the electron coherence time, length and mean free path in silicon. Note that the density of states (DOS) effective mass is used to estimate the average group velocity and the conductivity effective mass is used to determine mean-free scattering times and lengths.

$T$	$N_D$	$1/\sigma_e$	$v_g$	$t_c$	$\tau$	$l_c$	$\Lambda = v_g \tau$
[K]	[cm <sup>-3</sup> ]	[μΩ-cm]	[x10 <sup>7</sup> cm/s]	[fs]	[fs]	[nm]	[nm]
77	10 <sup>13</sup>	N/A	0.54	623	N/A	67	N/A
77	10 <sup>20</sup>	N/A	0.54	623	N/A	67	N/A
300	10 <sup>13</sup>	5x10 <sup>8</sup>	1.07	160	841	17	90
300	10 <sup>20</sup>	400	1.07	160	105	17	11

Table 2.2: Estimates of the phonon coherence time, length and mean free path in silicon. Note that the group velocity is taken to be the acoustic velocity averaged over the low-frequency [100] longitudinal and acoustic modes, i.e.  $v_g = (v_{ac,l} + 2v_{ac,t})/3$  where  $v_{ac,l}$  and  $v_{ac,t}$  are 8.43 and 5.84 Km/s respectively.

$T$	$k_{th}$	$C_v$	$v_g$	$t_c$	$\tau$	$l_c$	$\Lambda = v_g \tau$
[K]	[Wm <sup>-1</sup> K <sup>-1</sup> ]	[MJ/K/m <sup>3</sup> ]	[Km/s]	[ps]	[ps]	[nm]	[nm]
77	1,442	0.41	6.70	0.62	234	4.17	1568
300	148	1.66	6.70	0.16	5.95	1.07	39.9

The coherence length of electrons is more than an order of magnitude longer than that of phonons, a discrepancy which arises from the differences in dispersion relations for the two particle types. Consequently, it is far more challenging to create phonon superlattices than it is for electronic superlattices. For the case of electrons, quantum coherence effects can be readily observed for structures with physical dimensions can on the order of 10 nm or even larger. However, to see similar type effects with phonons, the physical size of the structure needs to be of order 1 nm, at which point, atomic level surface roughness and lattice defects must be controlled to extreme levels. In Chapter 3, we will examine the nature of the bandstructure for both electrons and phonons in nanoscale cubic sized silicon crystals.

## 2.4 Energy Conversion in Nanostructures

In this section, we discuss how energy conversion processes are modified when one or more of the physical dimensions of a solid are reduced to lengths comparable to either the coherence length  $l_c$  of the dominant energy carriers. Two significant effects are observed. First, the distribution of energy levels, or energy spectrum, of the solid are shifted with respect to the spectrum of the bulk material. Secondly, the wavefunctions are modified such that the... We begin by reviewing the nature of the particle wavefunction in an infinite crystal and then see that happens when we one or more of the crystal dimensions is reduced to a scale comparable to the coherence length of the energy carrying particles.

In an infinite crystal, we know from Bloch's theorem, that the wavefunction can be written as the product of a function,  $u_{n,\vec{k}_0}$ , which is periodic in the primitive crystal lattice vectors  $\vec{a}_i$ , and a plane wave envelope function  $e^{i\vec{k}_0 \cdot \vec{r}}$  where  $\vec{k}_0$  represents a wavevector contained within the first Brillouin Zone (FBZ), i.e.

$$\Psi_0(\vec{r}, t) = u_{\vec{k}_0}(\vec{r}) e^{i(\vec{k}_0 \cdot \vec{r} + \omega(\vec{k}_0)t)} = \phi(\vec{r}) e^{i\omega(\vec{k}_0)t} \quad (2.56)$$

2.54 has the property

$$\Psi(\vec{r} + \vec{R}, t) = u_{\vec{k}_0}(\vec{r} + \vec{R}) e^{i(\vec{k}_0 \cdot (\vec{r} + \vec{R}) + \omega t)} = \Psi(\vec{r} + \vec{R}, t) = \Psi(\vec{r}, t) e^{i\vec{k}_0 \cdot \vec{R}} \quad (2.57)$$

where

$$\vec{R} = m\vec{a}_1 + n\vec{a}_2 + l\vec{a}_3 \quad ; \quad m, n, l \in \text{integers} \quad (2.58)$$

Taking the Fourier transform of 2.56 with respect to the spatial coordinate, we have

$$\phi(\vec{k}) = \mathfrak{T} \left\{ u_{n,\vec{k}_0}(\vec{r}) e^{i\vec{k}_0 \cdot \vec{r}} \right\} = \mathfrak{T} \left\{ u_{\vec{k}_0}(\vec{r}) \right\} * \mathfrak{T} \left\{ e^{i\vec{k}_0 \cdot \vec{r}} \right\} \quad (2.59)$$

where (\*) represents the three-dimensional convolution integral. Since  $u_{\vec{k}}(\vec{r})$  is a periodic function, it will have a discrete periodic spectrum in reciprocal lattice vectors,  $\vec{b}_i$  which are related to the real-space lattice vectors according to

$$\vec{G} = p\vec{b}_1 + q\vec{b}_2 + r\vec{b}_3 \quad ; \quad p, q, r \in \text{integers} \quad (2.60)$$

with the primitive reciprocal lattice vectors being defined in the standard way as

$$\bar{b}_1 = 2\pi \frac{\bar{a}_2 \times \bar{a}_3}{V_{p.c.}} ; \bar{b}_2 = 2\pi \frac{\bar{a}_3 \times \bar{a}_1}{V_{p.c.}} ; \bar{b}_3 = 2\pi \frac{\bar{a}_1 \times \bar{a}_2}{V_{p.c.}} \quad (2.61)$$

and  $V_{p.c.} = \bar{a}_1 \cdot (\bar{a}_2 \times \bar{a}_3)$ . Using the Fourier transform property,  $\mathfrak{T}\{f \cdot g\} = \mathfrak{T}\{f\} * \mathfrak{T}\{g\}$ ,

2.57 can be written as

$$\phi(\bar{k}) = \left\{ \sum_{\bar{G}} c_{\bar{k}_0 + \bar{G}} \delta(\bar{k} + \bar{G}) \right\} * \{2\pi \delta(\bar{k} - \bar{k}_0)\} = 2\pi \sum_{\bar{G}} c_{\bar{k}_0 + \bar{G}} \delta(\bar{k} - \bar{k}_0 + \bar{G}) \quad (2.62)$$

An infinite series of delta functions with magnitudes  $2\pi |c_{\bar{k}_0 + \bar{G}}|$  are located at the points defined by  $\bar{k} = \bar{k}_0 + \bar{G}$  where  $\bar{G}$  is the complete set of reciprocal lattice vectors. The nature of Eq. 2.59 is illustrated in Figure 2.4 for a hypothetical one-dimensional crystal.

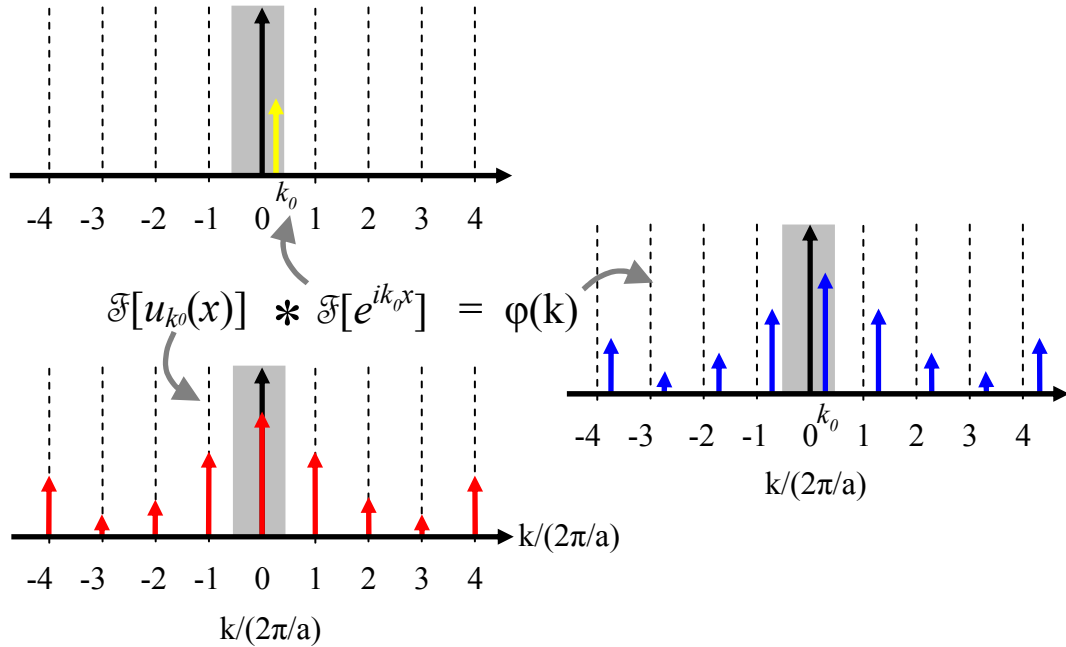


Figure 2.4: Schematic illustration of the Fourier representation of a Bloch wavefunction (right) describing an electron in a periodic crystal. The composite phase-space wavefunction arises from the convolution of two components: (top left) the non-periodic phase factor and (bottom left) the periodic function.

If non-periodic boundary conditions are imposed upon the crystal, as is the case for any real finite crystal, then the periodic portion of the wavefunction is necessarily lost and the Bloch theorem does not strictly apply. Of course for large crystals, the Bloch function is still an excellent approximation. An important question then becomes: what is the critical length scale at which the Bloch function ceases to provide a useful description of the wavefunction? To see the effect in practice, we take a one-dimensional crystal and abruptly truncate the edges to a finite length,  $L$ . This can be modeled by applying a top-hat function to the infinite periodic crystal with lattice period  $a$  having a value of one over the region  $x = \pm L/2$  and zero outside of this region. The Fourier transform of the modulation function is

$$M(k) = L \cdot \text{sinc}\left(\frac{Lk}{2}\right) \quad (2.63)$$

which has a characteristic spatial bandwidth of  $\Delta k_x = 4\pi/L$ . We find that each of the discrete Fourier components exhibited by the bulk crystal wavefunction will essentially be convolved with a broadening function closely resembling 1.61. The difference between bulk and finite crystals is shown qualitatively in Fig. 2.5. The key results is that the broadening of each peak is of order  $1/L$  and is therefore negligible until  $L \sim a$ .

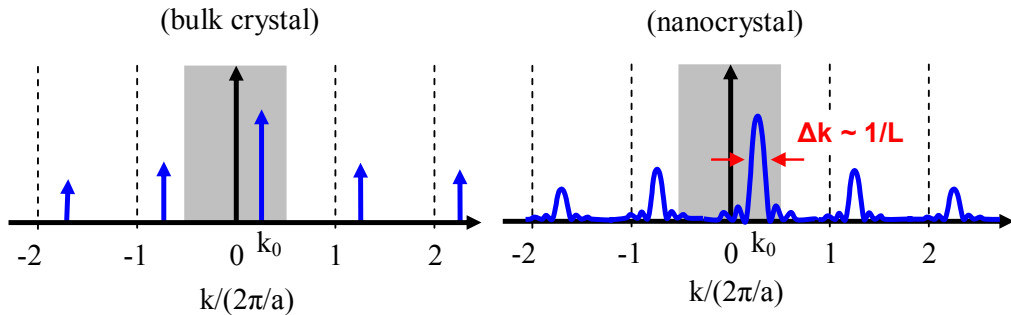


Figure 2.5: Schematic of the frequency space representation of a particle wavefunction (left) residing in a hypothetical one-dimensional periodic lattice and having good quantum numbers  $n$  and  $k_0$  and (right) the same particle in a finite one-dimensional crystal with a linear dimension equal to  $L \sim 4a$ . Each of the Fourier components of the wavefunction exhibit a characteristic broadening of order  $1/L$ .

We now solidify this result with a simple but extremely important example, which provides essential insight into the characteristics of real 3-D solids, including silicon. We take the standard Kronig-Penny model and see what happens when we truncate the boundary of the crystal to only a handful of unit cells. Fig. 1.9 shows the unit cell for the example which has dimension of  $b = 1.25 \text{ \AA}$ ,  $L = 2.5 \text{ \AA}$  and  $a = 5 \text{ \AA}$ . We compare two cases: (case 1) four unit cells with periodic boundary conditions (a ring of four cells) and (case 2) four unit cells with constrained boundary conditions where the barrier is extended sufficiently far outside of the crystal so as to not impact the wavefunction. The energy of the barrier,  $E_b$ , was chosen to be 3 eV which leads to two bound states with energies  $E_0$  and  $E_1$  in the finite single unit cell case. Increasing the number of unit cells to four causes the two energy levels to split into a total of eight energy levels as shown in the manifold on the right hand side of Fig. 2.6. In the case that periodic BCs are imposed, two of the energy levels are degenerate due to symmetry requirements and is indicated as such in the figure.

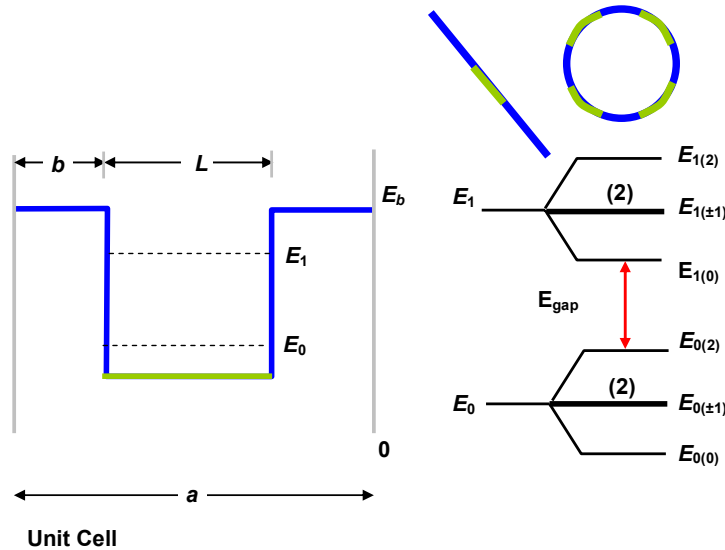


Figure 2.6: (left) The potential energy of a single unit cell of a hypothetical 1-D crystal having two bound energy states  $E_0$  and  $E_1$  in the isolated, i.e. one unit cell, case. (right) Energy manifold showing how  $E_0$  and  $E_1$  split into a total of eight levels, two of which are degenerate, when four unit cells are included and periodic boundary conditions are enforced.

The left side of Fig. 2.6 shows the energy band diagram in the extended zone scheme for the periodic case. The red circles highlight points in  $k$ -space for the four-unit cell crystal. The solid black dots are  $E$ - $k$  points for the same crystal but with the number of unit cells increased to 40 in order to show that the shape of the band diagram is unchanged; only the density of levels increases to the point of a continuum in an infinite crystal.

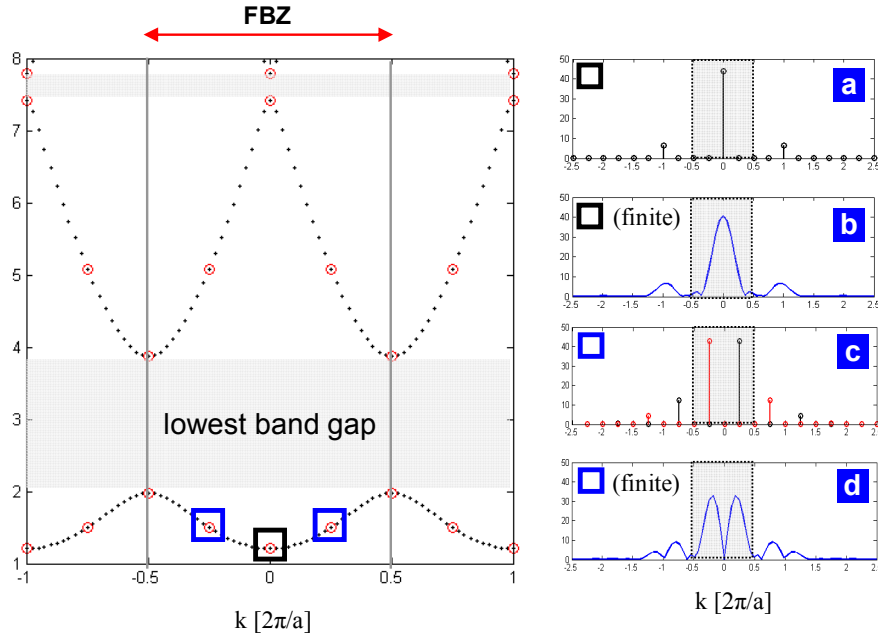


Figure 2.7: (Left) Energy band diagram for the periodic crystal having 4 unit cells (red open circles). The black dotted lines represent the form of the bands as the number of unit cells is increased to 40. (Right) Projection of the two lowest energy state wave functions (ground state = black square ; 1<sup>st</sup> excited state = blue squares) onto a plane-wave (i.e. Fourier) basis for (a, c) a periodic crystal having 4 unit cells (a ring) and (b, d) a finite crystal with exactly 4 unit cells.

On the right side of Fig. 2.7, the Fourier spectrums of the two lowest energy levels for the four-unit cell crystal having periodic and finite BCs are shown. Fig. 2.7a and 2.7b are the Fourier spectrums of the ground state (black square) of the four unit cell crystal having periodic and fixed BCs respectively. Fig. 2.7c and 2.7d are the Fourier spectrums (magnitudes) of the first excited state (blue squares) of the four unit cell crystal having periodic and fixed BCs respectively. As was predicted, the finite crystal exhibits broadening about each Fourier component of the periodic crystal spectrum

with a characteristic bandwidth of  $\sim 4\pi/L = \pi/a$  which is exactly half the width of a single Brillouin Zone. Finally, we compare the real-space representation of the ground state wavefunction with its Fourier-space representation for the two cases in Fig. 2.8.

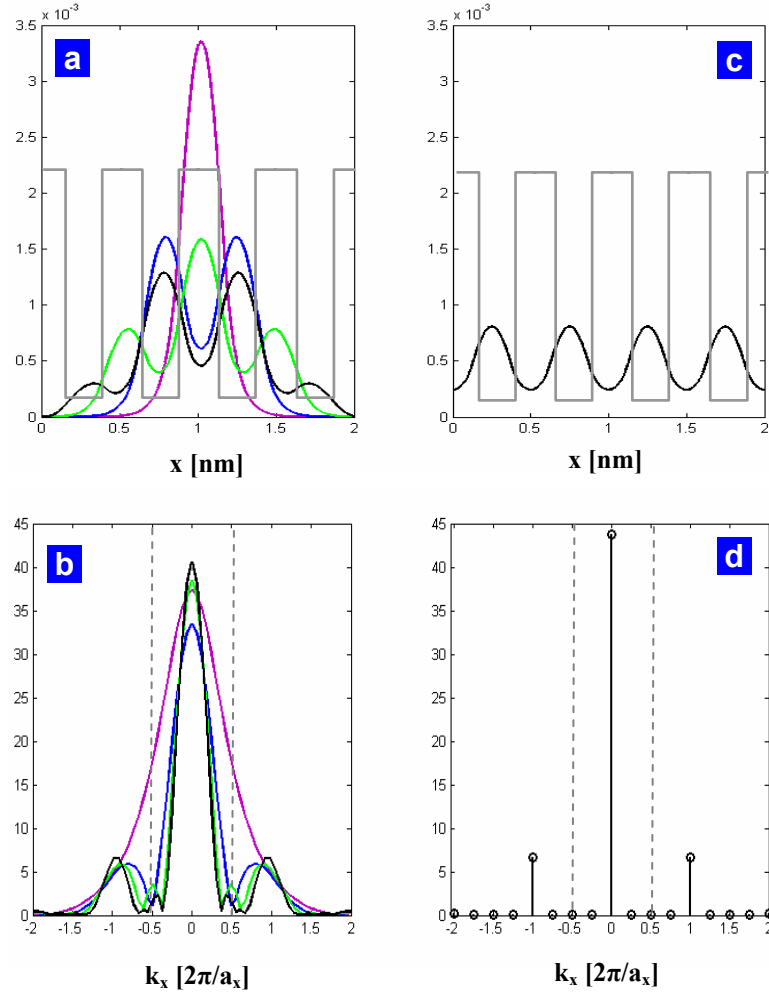


Figure 2.8: (a, c) Real-space and (b, d)  $k$ -space representations of the ground state wavefunction for the (a, b) finite crystal with 1 (purple), 2 (blue), 3 (green), and 4 (black) unit cells in comparison with (c, d) a periodic crystal having four unit cells. The four-unit cell periodic potential is shown in (a) and (c) for qualitative comparison.

We are now in position to understand a key consequence to reducing the dimensionality of the crystal to a length comparable to the lattice constant.

## 2.5 Special Properties of the Silicon Crystal

By scaling down the physical size of the silicon crystal, the basic electronic, optoelectronic, and thermal-mechanical properties of silicon are modified. Through advances in fabrication techniques, the size and shape of silicon crystal can be tailored to adjust these physical properties to meet the demands of a particular device application. For example, despite being a very poor source for photons in its bulk form, silicon nanocrystals having dimension of order 1-10 nm have been shown to exhibit efficient luminescence under optical and more recently electrical pumping. The enhanced light emission is attributed to the relaxation of the k-conservation rule as the physical dimension of the crystal becomes comparable to the lattice constant. This behavior was shown numerically in Section 2.4 with the 1-D multi-quantum well examples.

By decreasing the physical size of the crystal to some small multiple of the crystal lattice constant (5.43 Å), the physical meaning of a Brillouin zone and a k-vector associated with a wavefunction begins to be blurred. At some point, we must begin to ask ourselves if it makes sense to treat silicon as an indirect bandgap material? In many respects, we have already addressed this question in Section 2.4. Strictly speaking, the concept of a Brillouin zone scheme, is only valid for infinite periodic crystals. For a finite crystal, each of the k-vector components begins to broaden into continuous k-space distribution but which is still highly peaked around the positions defined for the infinite crystal. This broadening increases to the point that the full-width of each peak extends across the entire FBZ at roughly when the crystal is one lattice constant wide. Therefore, we see that silicon will still retain its characteristic as an indirect bandgap material until the crystal size is just a few nanometers on an edge. We will examine the modifications to the bandstructure of silicon for small nanocrystals in the following chapter.

## Chapter 3

### The Energy Eigen States of Silicon

In this chapter I will describe the computational methods I used to compute the electron and phonon energy eigen states for silicon in both bulk and finite nanoscale structures. The electron eigen states are calculated using an empirical tight binding (ETB) model and the phonon eigen states are computed using a valence force potential (VFP) model. Both are semi-empirical models which accurately reproduce the full electron and phonon bandstructure of the bulk crystal and can be easily adapted to the treatment of large (5 nm) nanocrystals like those which will be discussed in Chapter 7. The computed eigen states will be used in subsequent chapters to determine electron-phonon and phonon-phonon transition probabilities and energy conversion rates using time-dependent perturbation theory.

#### 3.1 Background

As we discussed extensively in Chapter 2, time-dependent perturbation theory (TDPT) can often be used effectively to compute transition probabilities and therefore energy transfer rates between particles provided the conditions for its validity are met. In order to employ TDPT, explicit knowledge of the energy eigen states of the unperturbed crystal are required. However, calculating the eigen states of the unperturbed crystal exactly from first principles is a monumental undertaking even for the most simplest of crystals. Even the most rigorous calculations performed today using super computers are forced to make simplifying assumptions, particularly when it concerns the exceptionally difficult task of treating many-body electron-electron (e-e) interactions. The most successful “first principles” calculations for either electron or phonon eigen states has been density functional theory (DFT), which addresses the e-e interaction through an effective potential determined through self-consistency of

the Poisson and Schrödinger equations. Despite many successes in its predictive powers, DFT calculations are computationally expensive and are therefore limited to the treatment of very simple systems, including small molecules and crystals with long-range periodicity. Furthermore, DFT calculations do not always produce superior results compared to semi-empirical techniques which are physics based but which rely on a small number of adjustable parameters optimized to reproduce the bandstructure at a large number of measurable points. As an example of a glaring limitation, DFT calculations routinely underpredict the bandgaps for bulk crystals by more than 2x [11]. Surprisingly, this is despite the ability of DFT calculations to accurately predict the dispersion within individual bands. The inability to predict the bandgap of bulk semiconductor materials greatly diminishes our confidence in using it to predict the energy eigen states of small particles.

In contrast, DFT calculations have been very successful in predicting the phonon bandstructure of a wide range of semiconductor materials including heterogeneous alloyed semiconductor materials [12]. However, DFT calculations fall short when it comes to modeling relatively large nanocrystals and the coupling between eigen states with reduced crystal symmetry, i.e. those modes which do not fall along a high-symmetry axis including the zone center.

Molecular dynamics (MD) simulations using a rigid inter-atomic potential is also a powerful technique for describing phonon propagation and energy relaxation as a result of high-order lattice anharmonicity [13]. Though, for many cases, including silicon, this approach does not adequately treat the valence electron density redistribution which acts to dampen the effective inter-atomic force constants. This leads to an increase in an effective stiffness of the lattice and errors in the phonon energies computed [13].

Given the limitations of ab initio calculations and our need to compute the electronic and vibrational eigen states of relatively large silicon nanocrystals as well as the interactions of states with reduced symmetry, semi-empirical techniques offer a suitable balance of accuracy and computational cost to meet the needs of this work. An additional benefit to using a semi-empirical approach is that we are able to make a

more meaningful comparison between eigen states computed for bulk and finite nanometer scale crystals.

In the next two sections, we will discuss the semi-empirical techniques employed to determine the energy eigen states of the atomic lattice, i.e. the vibrational states, and of the valence electrons. Before doing so, let us first review briefly the basic chemical bonding structure of silicon. Figure 3.1 shows the primitive cell of silicon along with the atoms contained within the 12 nearest neighbor primitive cells. The basic bonding structure is tetrahedral with an unstressed bond angle of  $109.5^\circ$ , the solid lines drawn between atoms indicating the primary bond axes. The crystal has a diamond lattice structure, which can be described as a face centered cubic (FCC) lattice with a basis of two nonequivalent atoms displaced by the vector  $(a/4)(1\ 1\ 1)$  where  $a$  is the lattice constant ( $5.43\ \text{\AA}$  at room temperature). The two nonequivalent silicon atoms are indicated in the figure as filled and unfilled spheres.

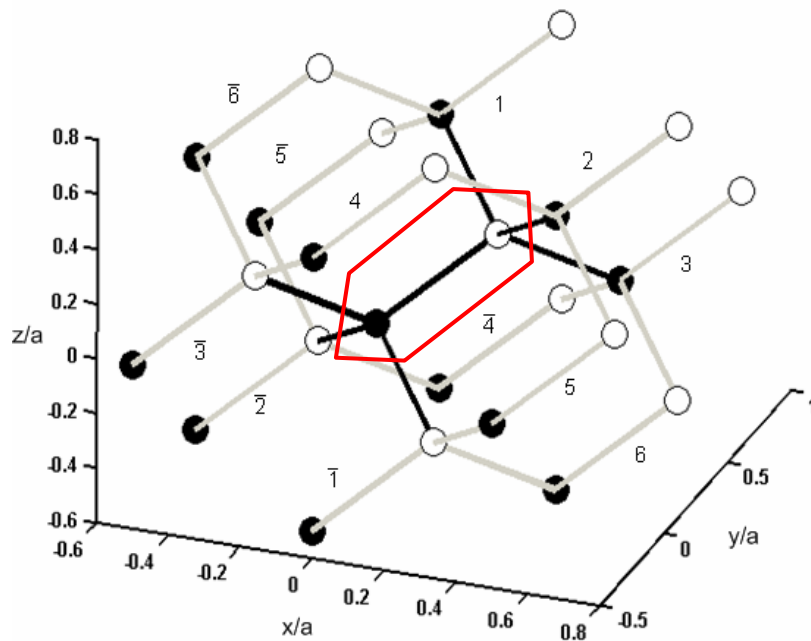


Figure 3.1: Atomic structure of silicon including the primitive cell and the 12 neighboring primitive cells which include up to fourth nearest neighbor.

Each of the primitive cells is labeled with a barred or unbarred number from 1-6, in accordance with the labeling convention used by Tubino et al. [14]. The barred and unbarred nomenclature simplifies the analysis by identifying the primitive cell pairs which share a common lattice vector  $\vec{R}$  but with opposite sign, i.e. their phase factors are related by  $\exp(\pm i\vec{G} \cdot \vec{R}_j) = \exp(\pm i\theta_j)$  ;  $j = 1, 2, \dots, 6$ .

Finally, we recall that the reciprocal space representation is that of a body-centered cubic lattice, whose primitive cell (the first Brillouin zone) is a 14 sided truncated octahedron as shown in Figure 3.2. The high symmetry points of the reciprocal space lattice and the corresponding Group symbols are indicated for the irreducible wedge, which contains a reciprocal space volume equal to  $1/48^{\text{th}}$  of the FBZ.

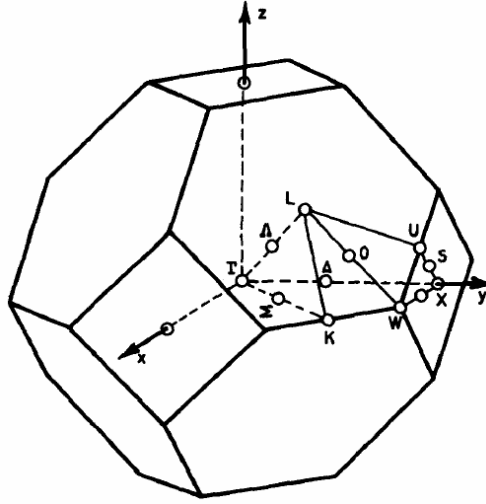


Figure 3.2: The primitive cell of the reciprocal lattice of silicon, i.e. the first Brillouin zone (FBZ) showing the irreducible wedge and Group symbols for the high-symmetry points and directions.

### 3.2 The Vibrational Eigen States

The vibrational energy eigen states of the silicon crystal are computed using a six-parameter semi-empirical valence force potential (VFP) model developed by Tubino et al. [14], which accurately reproduces the phonon bandstructure which is known from angle-resolved neutron diffraction experiments. The core elements of the VFP model was originally developed by E. B. Wilson [15, 16] as a way to predict the infrared absorption and Raman scattering spectra of molecules in the gas phase. This early approach, referred to as the “GF” method, was later adapted to the modeling of periodic crystals by Piseri and Zerbe [17] and then specifically to silicon, germanium and diamond by Tubino and co workers [14]. At its core, the approach is no different than any other lattice dynamics model which uses a set of effective force constants between the massive atomic cores in order to describe the Newtonian motion of the molecule. That is, the crystal is treated essentially as a coupled mass-spring system. However, there are several key attributes which make this technique particularly appealing over other semi-empirical lattice dynamics approaches. To appreciate these advantages, a brief review of lattice dynamics calculations will be beneficial.

First, consider the case of a finite crystal or molecule. Recall that, the motion of the  $N$  atoms about their equilibrium positions can be described in relation to the total kinetic ( $T$ ) and total potential ( $V$ ) energies as follows:

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} + \frac{\partial V}{\partial q_j} = 0 \quad ; \quad j = 1, 2, \dots, 3N \quad (3.1)$$

where  $q_j$  is the  $j^{\text{th}}$  displacement coordinate corresponding to atom  $n = \text{floor}(j/3)$ . There are  $3N$  such coordinates, one for each of the three established orthogonal coordinate axes per atom of the crystal.  $T$  and  $V$  are defined in the standard ways as

$$T = \frac{1}{2} \sum_{j=1}^{3N} m_j \dot{q}_j^2 \quad (3.2)$$

$$V = \frac{1}{2} \sum_{i,j=1}^{3N} f_{ij} q_i q_j + \text{higher order terms} \quad (3.3)$$

where the  $9N^2$  ( $3N \times 3N$ ) values of  $f_{ij}$  represent the full set of inter-atomic harmonic force constants. When inserted into 3.1, we arrive at the set of  $3N$  coupled equations

$$m_j \ddot{q}_j + \sum_{i=1}^{3N} f_{ij} q_i = 0 \quad j = 1, 2, \dots, 3N \quad (3.4)$$

The  $3N$  solutions to this equation retain the form

$$q_i = A_i \cos(\omega t + \phi) \quad (3.5)$$

where  $\omega =$  are the  $3N$  eigen frequencies of the lattice which are solutions to the secular equation

$$\sum_{i=1}^{3N} (f_{ij} - \delta_{ij} m_j \lambda^2) A_i = 0 \quad (3.6)$$

Each eigen state is characterized by each atom in the crystal having a simple harmonic oscillatory motion about its equilibrium position at a frequency and phase which is constant for all atoms. However, the relative amplitudes of the displacement of each atom is unique and characterized by the  $3N$  values of  $A_i$ . Note that there are three values of  $A$  for each atom for each orthogonal coordinate direction. In the special case that the masses are the same for all atoms in the crystal,  $\omega = \sqrt{m} \lambda$ .

One of the great benefits of the VFP approach comes in its ability to account for intermediate range forces, i.e. forces spanning several neighbor distances, with a highly reduced set of adjustable force constants  $f_{ij}$ . The reduction in effective force constants is achieved by grouping the inter-atomic coupling into a smaller set of natural molecular motions such as the linear stretching of a bond between two atoms or the bending of an angle defined by the intersection of two bonds with an apex atom as shown in Figure 3.3.



Figure 3.3: Schematic representation of two basic internal coordinates, including (left) bond stretching and (right) angle bending. The bond stretching internal coordinate involves two atoms whereas the bond angle displacement internal coordinate involves the motion of three atoms. The orientation of the displacement vectors for each of the atoms is indicated by a straight arrow.

By building up from simple molecular motion, physical intuition into the crystal vibrational spectrum is maintained as the size one examines the crystal being built from a small number of atoms into an infinite periodic crystal.

The simple molecular motions are defined as linear combinations of atomic displacements and can be treated as a new coordinate called an “internal coordinate.” To be consistent with Wilson’s notation, the internal coordinates are represented by the symbol  $S_t$  where  $t$  represents the label of the internal coordinate. The Newtonian motion of the crystal can then be recast in terms of these internal coordinates as

$$T = \frac{1}{2} \sum_{t,t'=1}^P G^{-1}_{tt'} \dot{S}_t \dot{S}_{t'} \quad (3.7)$$

and

$$V = \frac{1}{2} \sum_{t,t'=1}^P F_{tt'} S_t S_{t'} \quad (3.8)$$

where  $P$  is the total number of defined internal coordinates for the finite crystal or molecule under study. The internal coordinates, labeled with the  $t, t'$  subscripts running over the internal coordinate space, are expressed as linear combinations of the real coordinate displacements through the  $B$  matrix, i.e.

$$S_t = \sum_{i=1}^{3N} B_{ti} q_i \quad (3.9)$$

The inverse  $G$  matrix in (3.7) is defined in relation to the  $B$  matrix and the atomic masses as follows

$$G_{tt'} = \sum_{i=1}^{3N} \frac{1}{m_i} B_{ti} B_{t'i} \quad (3.10)$$

Again, it should be noted that here are only  $N$  available values of  $m_i$  but there are  $3N$  values of the subscript  $i$  in order to account for the three degrees of freedom for each atom. It follows that the secular equation for each eigen value can be written in the following form

$$|GF - I\lambda| = 0 \quad (3.11)$$

which was the motivation for calling it Wilson's "GF" method. The  $G$  matrix is essentially a matrix describing the geometric relationship between the internal and

crystal (true) coordinates of the molecule. The  $F$  matrix is the matrix describing the forces which translate displacements of the internal coordinates to the contribution to the potential energy function.

This technique is extended to periodic crystals in the following way. First, the phonon coordinate dynamical matrix  $D$  is defined as

$$D(\vec{K}) = M^{-1/2} F_q(\vec{K}) M^{-1/2} \quad (3.12)$$

where the  $M$  matrix is a diagonal matrix containing the masses related to each atomic coordinate. The  $F$  matrix is related by the  $B$  matrix through the transformation

$$F_q(\vec{K}) = B^+(\vec{K}) F_s(\vec{K}) B(\vec{K}) \quad (3.13)$$

where, again, the  $B$  matrix transforms the force constant description from crystal displacement coordinates ( $q$ ) to internal coordinates ( $S$ ). Note that the  $B$  matrices are in general not square matrices. Their size is instead given by  $(P) \times (3N)$ , where  $P$  is the number of internal coordinates. Finally, the phonon space ( $\vec{K}$ -space) representations of the  $B$  and  $F_s$  matrices are given by the expressions

$$B(\vec{K}) = \sum_l B_l \exp[-i\vec{K} \cdot \vec{R}_l] \quad (3.14)$$

$$F_s(\vec{K}) = \sum_l F_l \exp[-i\vec{K} \cdot \vec{R}_l] \quad (3.15)$$

where  $\vec{K}$  is the reciprocal space vector and  $\vec{R}_l$  is the displacement vector corresponding to the  $l^{\text{th}}$  unit cell relative to the origin.  $B_l$  and  $F_l$  are the geometric transformation and force constant matrices which account for the atoms involved in the  $l^{\text{th}}$  unit cell. Combining these expressions, we obtain the following secular equation where  $L(\vec{K})$  is the eigenvector matrix whose columns represent the mass weighted displacement coordinates of the atoms in the primitive cell.

$$[D(\vec{K}) - \omega^2 I]L(\vec{K}) = 0 \quad (3.16)$$

For the case of silicon, there are two atoms in the primitive cell and thus the matrix is  $6 \times 6$ . For each value of  $\vec{K}$ , there will be six eigenvalues  $\omega$ . These values represent the six different phonon branches which are classified according to their polarization along the symmetry axis  $\vec{K} = (2\pi/a)[1\ 0\ 0]$ .

Tubino and coworkers [14] were able to model the full dispersion of the silicon lattice using only three internal coordinates along with three off-diagonal force constants totaling six force constants in the internal coordinate representation. These six collective atomic motions are represented schematically in Figure 3.4 and the force constant values are listed in Table 3.1. The total number of internal coordinates reaches 16 after accounting for all possible interactions between the primitive cell atoms and their neighbors through these three internal coordinates.

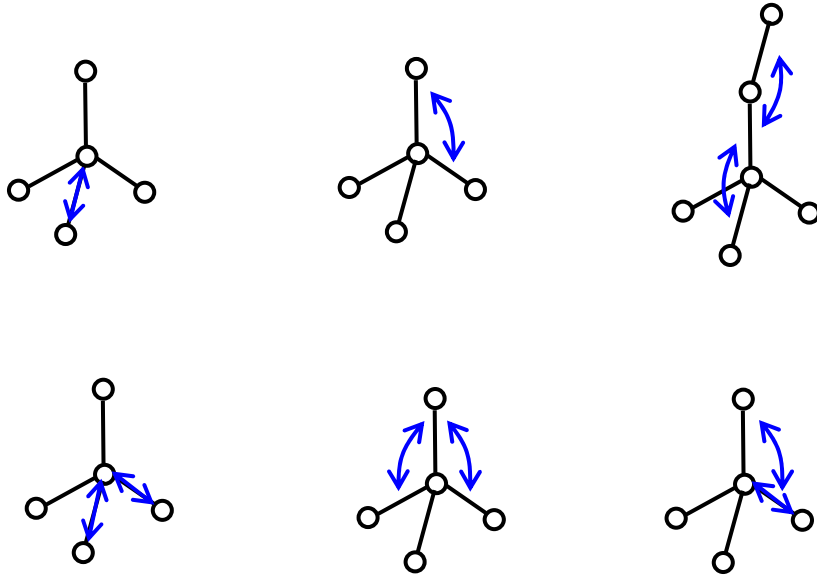


Figure 3.4: Schematic representation of the types of atomic displacements which make up the three internal coordinates (top row) and the coupling between internal coordinates (bottom row) used to describe the motion of the silicon lattice. There is a total of 16 internal coordinates defined for the bulk silicon lattice.

Table 3.1 Valence force potential constants for bulk silicon as determined by Tubino et al. [14] and used in this work.

$F_{ii'}$	value	unit
$K_R$	$1.469 \pm 0.031$	mdynes/Å
$H_\Delta$	$0.164 \pm 0.011$	mdyne-Å-rad <sup>-2</sup>
$F_R$	$0.052 \pm 0.008$	mdynes/Å
$F_{R\Lambda}$	$0.073 \pm 0.008$	mdynes/rad
$F_{\Lambda'}$	$-0.025 \pm 0.013$	mdyne-Å-rad <sup>-2</sup>
$F_{\Lambda''}$	$0.132 \pm 0.021$	mdyne-Å-rad <sup>-2</sup>

### 3.2.1 Bulk Phonon Modes

In this section, we examine the phonon bandstructure and constant energy surfaces for the bulk (infinite, periodic) silicon crystal computed using the valence force potential approach described in the previous section. Figure 3.5 shows the full phonon bandstructure and constant energy surfaces in k-space along the high symmetry directions of the crystal.

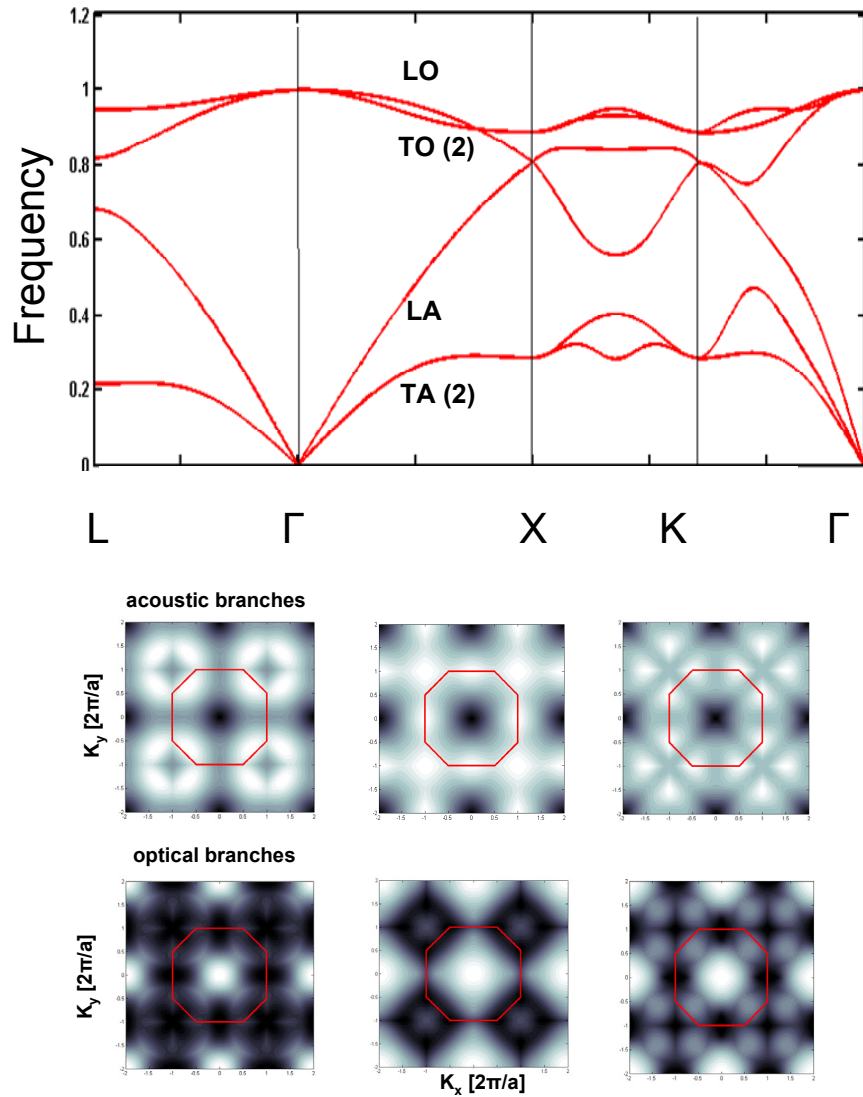


Figure 3.5: (top) Bulk phonon dispersion relationship for the silicon lattice computed using the VFP model. (bottom) Constant energy contours of the six phonon branches in the  $K_x$ ,  $K_y$  plane. The acoustic modes are shown on the top row and the optical modes on the bottom row. The red octagons indicate the boundary of the FBZ in the specified plane. Lighter regions indicate higher energy.

Along the equivalent  $[1\ 0\ 0]$  directions ( $\Gamma X$ ), the six phonon branches are labeled according to the polarization vector of the eigenfunctions, i.e. the displacement vectors of the primitive cell atom, and the relative phase angle of the displacement vectors of the two atoms in the primitive cell. The three optical branches correspond to the case where the two primitive cell atoms are out of phase ( $\phi = \pi$ ) whereas the three acoustic modes are defined by the case when the two atoms are in phase ( $\phi = 0$ ). Within these two categories, we have longitudinal and transverse modes corresponding to the cases that the direction of the polarization vector falls along the  $\Gamma X$  axis or orthogonal to it respectively. By symmetry, the transverse modes are degenerate along  $\Gamma X$  but are split along other lower-symmetry axes, as for example along the  $\Gamma K$  and  $XK$ . Furthermore, the transverse and longitudinal modes are degenerate at the zone center ( $\Gamma$ ) for the acoustic and optical modes. The acoustic frequency vanishes at the zone center whereas the optical phonon frequency of (63 meV). This mode is optically active and can be probed using Raman or IR absorption or reflection spectroscopy.

### 3.2.2 Vibrational Modes of Silicon Nanocrystals

In this section, we examine the results of computing the vibrational spectrum of finite silicon crystals using the VFP approach and the internal coordinate force constants which reproduce the bulk phonon dispersion.

Figure 3.6 shows the evolution of the distribution of eigen energies as a function of increasing the size of the nanocrystal. The number of atoms included in the crystal are indicated in each plot and range from a 28 atoms to 730 atoms with the assumption that the basic tetrahedron bonding structure is maintained. The number of atoms correspond to crystals having  $1 \times 1$ ,  $2 \times 2$ ,  $3 \times 3$ , and  $4 \times 4$  unit cells. Note that there are 8 “whole” atoms within the silicon unit cell. However, in our construction of a crystal with the same size, the atoms at the corners of the unit cell are included which gives us a larger number of atoms.

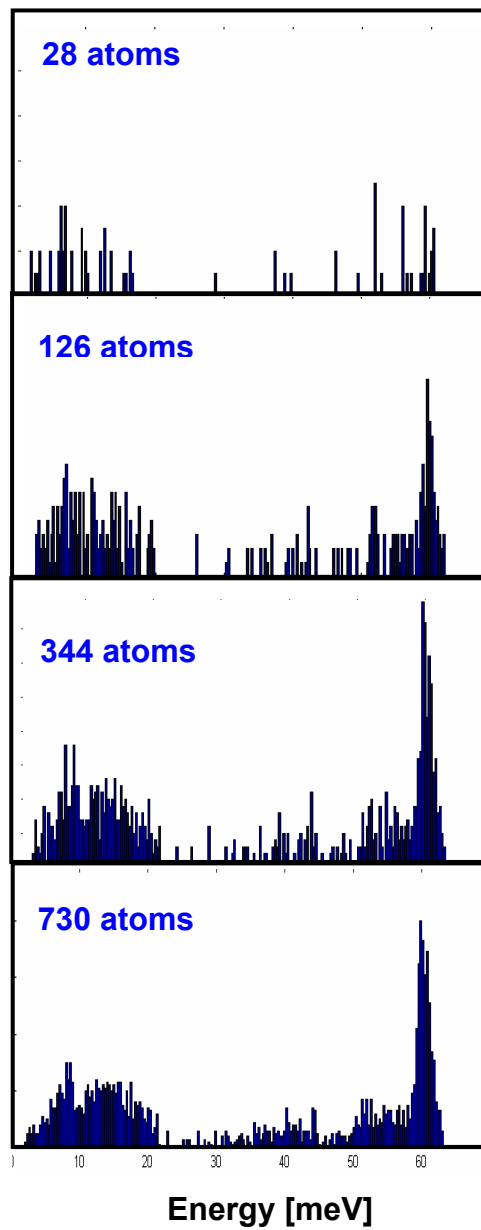


Figure 3.6: Energy distribution of the vibrational states as a function of crystal size as indicated by the number of atoms. The distribution rapidly converges to that of the bulk crystal.

In the typical experimental scenario, the silicon nanocrystal will be surrounded by a host material or fluid and the energetic coupling to the local environment would need to be accounted for. However, modeling the coupling to the environment requires knowledge of the atomic structure and inter-atomic forces which is not readily available. Therefore, it is instructive to examine the effective of the boundary atoms on the vibrational spectrum. In Figure 3.7, we examine the impact of the choice of boundary condition imposed on the lattice dynamics by simply changing the mass of the boundary atoms. Here, we examine the case that  $m = m_{\text{Si}}$  and  $m = \infty$  for a cubic nanocrystal containing 730 atoms. The vibrational spectrum of the finite nanocrystal for these two cases are compared to the density of states of the bulk crystal, which was computed using a Brillouin zone Tetrahedral integration method to be described in more detail in Chapter 4.

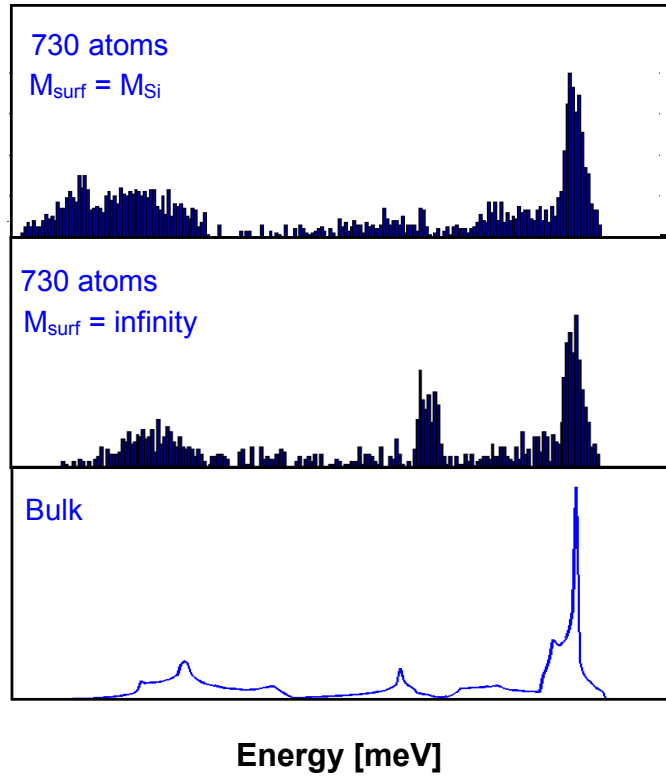


Figure 3.7: The effect of boundary conditions on the energetic distribution of vibrational modes. The top two distributions represent the cases where the surface atoms are taken to be equivalent to the silicon mass or taken to be infinite in mass respectively.

Once the energy distribution is known, one can proceed to compute thermodynamic properties such as the heat capacity of a single silicon nanocrystal. Using expression (3.17), we compute the heat capacity per atom as a function of temperature for four different sizes of nanocrystals and compare it to the heat capacity of the bulk crystal computed using (3.18) where  $\rho(E)$  is the volumetric density of states. The results are shown in Figure 3.8.

$$C_a = \frac{1}{N} \frac{\partial}{\partial T} \sum_i^{3N} \frac{E_i}{\exp(E_i / k_B T) - 1} \quad (3.17)$$

$$C_V = \frac{\partial}{\partial T} \int_0^{\infty} \frac{E \rho(E) dE}{\exp(E / k_B T) - 1} \quad (3.18)$$

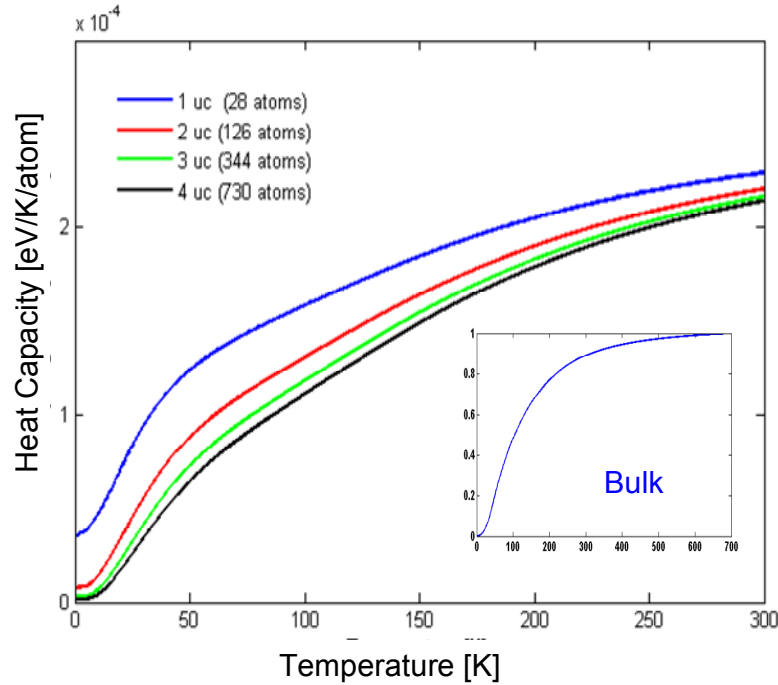


Figure 3.8: The effect of boundary conditions on the energetic distribution of vibrational modes.

### 3.3 The Electronic Eigen States

In this section, we discuss the methodology in computing the electronic bandstructure for silicon using nearest-neighbor empirical tight binding (ETB) approach, which uses a 10 orbital plus spin basis in an  $sp^3d^5s^*$  configuration developed by Boykin and coworkers [18]. The foundation of the ETB formalism, equivalent to the method of linear combination of atomic orbitals (LCAO), was developed by Slater and Koster [19] in the 1950s. and has remained essentially unchanged despite refinements in the parameter and basis state optimization made by notable researchers over the subsequent decades. An excellent reference for the method as well as its history is the text by W. Harrison [20]. For immediate reference, we briefly review this method below.

In the tight binding (TB) or LCAO method, the electron eigenfunctions are taken as linear combinations of the atomic orbital functions of the isolated atoms  $|\phi_{\alpha,\tau}\rangle$  where the index  $\alpha$  specifies a particular orbital label, e.g.  $s, p_x, p_y, p_z$ , and  $\tau$  is an index which specifies the atom within the crystal. In this context, the orbital functions associated with a given atom are taken to be mutually orthogonal. In the case of the full crystal, the  $n^{\text{th}}$  eigenfunction is described as the linear combination of the form

$$|\Psi_n\rangle = \sum_{\alpha,\tau} c_{\alpha,\tau}^n |\phi_{\alpha,\tau}\rangle \quad \alpha = 1, 2, \dots, N_{\text{orbitals}} ; \quad \tau = 1, 2, \dots, N_{\text{atoms}} \quad (3.19)$$

where the  $c_{\alpha,\tau}^n$  are complex coefficients and the indices  $\alpha$  and  $\tau$  run over the labels of the particular orbital and atom within the crystal respectively. It is then easy to see that the probability of finding an electron *about* a single atom of the crystal labeled  $\lambda$  when it is in the  $n^{\text{th}}$  eigen state, is then given by the summation

$$P_\lambda = \sum_{\alpha} |c_{\alpha,\lambda}^n|^2 \quad (3.20)$$

The  $n^{\text{th}}$  eigen energy is then found by solving the time-independent Schrödinger equation (3.21)

$$H|\Psi_n\rangle = E_n|\Psi_n\rangle \quad (3.21)$$

which upon substitution of (3.19) into (3.21), gives

$$H \sum_{\alpha,\tau} c_{\alpha,\tau}^n |\phi_{\alpha,\tau}\rangle = E_n \sum_{\alpha,\tau} c_{\alpha,\tau}^n |\phi_{\alpha,\tau}\rangle \quad (3.22)$$

or

$$E_n = \frac{\sum_{\alpha',\alpha,\tau',\tau} (c_{\alpha',\tau'}^n)^* (c_{\alpha,\tau}^n) \langle \phi_{\alpha',\tau'} | H | \phi_{\alpha,\tau} \rangle}{\sum_{\alpha',\alpha,\tau',\tau} (c_{\alpha',\tau'}^n)^* (c_{\alpha,\tau}^n) \langle \phi_{\alpha',\tau'} | \phi_{\alpha,\tau} \rangle} \quad (3.23)$$

Using the variational theorem, the above equation can be recast into the following secular equation

$$E_n = \frac{\sum_{\alpha',\alpha,\tau',\tau} (c_{\alpha',\tau'}^n)^* (c_{\alpha,\tau}^n) \langle \phi_{\alpha',\tau'} | H | \phi_{\alpha,\tau} \rangle}{\sum_{\alpha',\alpha,\tau',\tau} (c_{\alpha',\tau'}^n)^* (c_{\alpha,\tau}^n) \langle \phi_{\alpha',\tau'} | \phi_{\alpha,\tau} \rangle} \quad (3.24)$$

The challenge then becomes the determination of the values of the two classes of overlap integrals

$$H_{\alpha',\alpha,\tau',\tau} = \langle \phi_{\alpha',\tau'} | H | \phi_{\alpha,\tau} \rangle \quad (3.25a)$$

$$S_{\alpha',\alpha,\tau',\tau} = \langle \phi_{\alpha',\tau'} | \phi_{\alpha,\tau} \rangle \quad (3.25b)$$

Computing these values from first principles is extremely difficult and even the most rigorous approaches often fail to produce satisfactory results. Therefore, these values are typically determined by adjusting their values in order to minimize the error in reproducing the experimentally determined electron bandstructure. The first step in computing these values is therefore to minimize the number of adjustable parameters. Since the overlap integrals become fitting parameters at this point, the first simplifying assumption is to assume that the orbital functions are mutually orthogonal, even for orbitals associated with different atoms. This allows us to write  $S_{\alpha',\alpha,\tau',\tau} = \delta_{\alpha',\alpha,\tau',\tau}$  and limits the task to computing the overlap integrals in (3.25a). Beyond this, the best approach is then to begin with the smallest number of atomic orbitals necessary to reproduce the bandstructure with the necessary level of accuracy.

There is a choice in the number and type of orbitals to use in representing the electronic eigenfunctions. Each representation results in a different set of values of the overlap integrals. Given that the valence of silicon is 4 and the bonding is predominantly covalent, it would seem that the bandstructure should be satisfactorily reproduced by using only a basis set which includes  $sp^3$  hybridization. Indeed, pioneering work by Ren and Harrison [21] using tight binding approach to reproduce the silicon electronic bandstructure used this minimal basis set while limiting the overlap integrals to include only nearest neighbor interactions. With this choice of basis, the bandstructure can be reproduced with reasonable qualitative agreement. However, as Klimeck et al. [22] highlighted, the minimal basis configuration does not adequately describe the effective masses and even leads to some absurd results such as an infinite transverse effective mass at the X point. Niquet et al. [23] were able to resolve this limitation by including interactions up to third nearest-neighbor. However, this approach severely complicates the analysis of computing optical and electron-phonon matrix elements, the latter requiring the determination of the effects of local strain on the overlap integrals. After reviewing the various approaches to represent the

silicon lattice, the nearest neighbor  $sp^3d^5s^*$  configuration was deemed to be the most suitable for computing the energy eigen states for finite crystals. The values of the same-site and two-center overlap integrals used to compute the eigen energies of the bulk and finite nanocrystals are summarized below in Table 3.2 [18]. The notation is the same used by Slater and Koster [19].

Table 3.2: Summary of on-site and two-center overlap integrals used in computing the bulk and finite crystal bandstructures. The values are from Boykin et al. [18].

$H_{\alpha',\alpha}$	Value [eV]
$E_s$	-2.15168
$E_p$	4.22925
$E_{s^*}$	19.11650
$E_d$	13.78950
$\lambda$	0.01989
$ss\sigma$	-1.95933
$s^*s^*\sigma$	-4.24135
$ss^*\sigma$	-1.52230
$sp\sigma$	3.02562
$s^*p\sigma$	3.15565
$sd\sigma$	-2.28485
$s^*d\sigma$	-0.80993
$pp\sigma$	4.10364
$pp\pi$	-1.51801
$pd\sigma$	-1.35554
$pd\pi$	2.38479
$dd\sigma$	-1.68136
$dd\pi$	2.58880
$dd\delta$	-1.81400

### 3.3.1 Bulk Electron States

In this section, we examine the electronic bandstructure of bulk silicon using the ETB method described in the previous section. Figure 3.9 shows the full band dispersion of electrons along the high-symmetry directions as well as the constant energy surfaces of both the valence band hole states and the conduction band electron states. The conduction band has three six equivalent valleys located at the equivalent  $(2\pi/a)[0.85\ 0\ 0]$  positions. In the bottom right panel of Figure 3.9, the constant energy surface in the  $K_y$ - $K_x$  plane shows four of these minima.

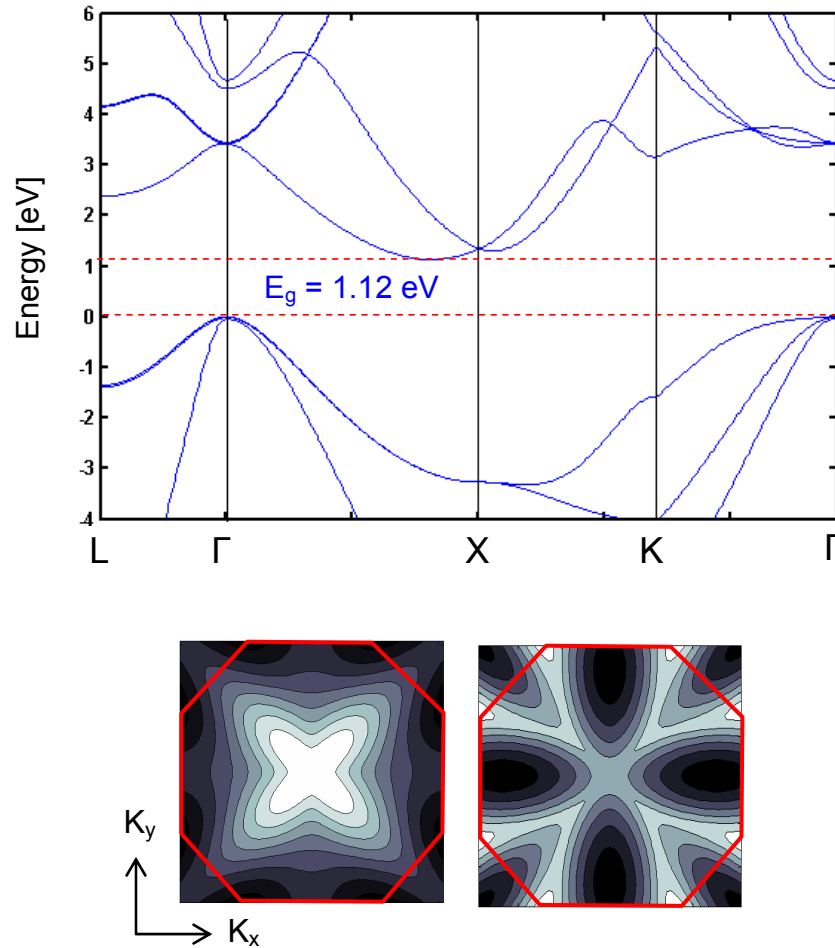


Figure 3.9: (top) Electronic bandstructure of bulk silicon along high-symmetry directions and (bottom) constant energy surfaces for (left) valence band holes and (right) conduction band electrons.

### 3.3.1 Electrons in Nanoscale Silicon

In this section, we calculate the distribution of eigen energies for small silicon nanocrystals and show that the dependence of the bandgap follows the effective mass approximation well. Figure 3.10 shows the histogram of energy states for two cubic hydrogen terminated nanocrystals of size 0.54 nm and 1.08 nm. The fundamental bandgap is located about 0 eV and is seen to increase from 2.33 eV to 3.42 eV in going from 1.08 nm to 0.54 nm.

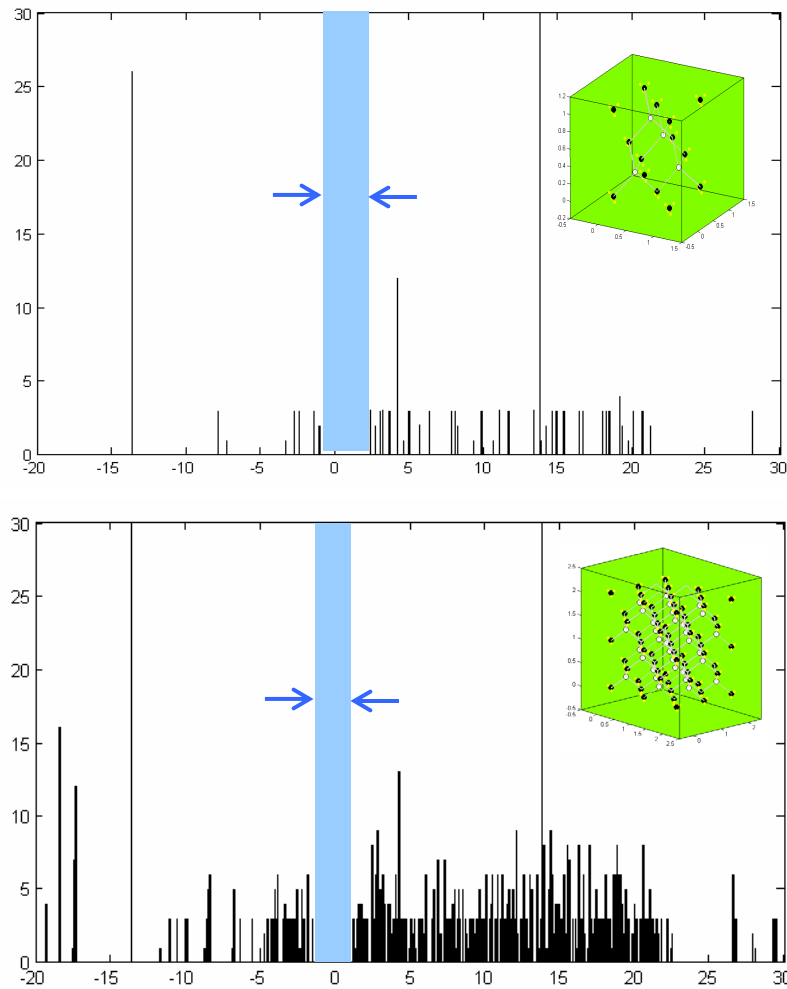


Figure 3.10: Electron state histogram (discrete density of states) for (top)  $\text{Si}_{18}\text{H}_{40}$  0.543 nm cube ( $E_g = 3.42$  eV,  $\lambda = 362$  nm) (bottom)  $\text{Si}_{95}\text{H}_{124}$  1.08 nm cube ( $E_g = 2.33$  eV,  $\lambda = 532$  nm).

Figure 3.11 shows the bandgap computed for four different finite nanocrystals as well as the bulk crystal (blue open circles). The red dotted line represents the bandgap as predicted by the effective mass approximation (3.26) using the conduction band longitudinal effective mass and the valence band heavy hole effective mass.

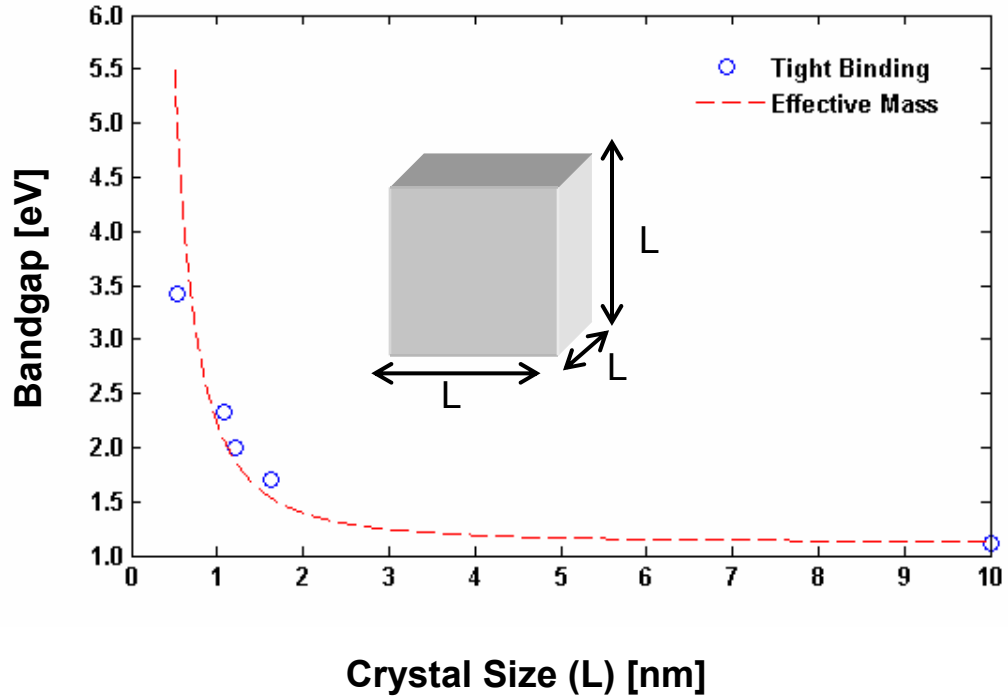


Figure 3.11: Energy gap of cubic silicon nanocrystals as a function of crystal size as calculated using the empirical tight binding approach and the simple effective mass approximation.

$$\Delta E_G = \frac{3\hbar^2}{2} \left( \frac{\pi}{L} \right)^2 \left( \frac{1}{m_l^*} + \frac{1}{m_{hh}^*} \right) \quad (3.26a)$$

$$m_l^* = 0.916m_e \quad (3.26b)$$

$$m_{hh}^* = 0.537m_e \quad (3.26c)$$

## Chapter 4

### Phonon-Phonon Energy Conversion

We should be able to put into the mathematical machine the interatomic force function and receive from the other end the value of the thermal conductivity at any temperature. If this has not been done, it is because of the great complexity of the computation.

- J. M. Ziman

In Chapter 3, we discussed the methods for computing the basis (eigen) states for electrons and phonons of the silicon crystal. In the determination of these states, the fundamental assumption is that the Hamiltonian is time-invariant. As was discussed in some detail in Chapter 2, in this picture, the eigen states are linearly independent and therefore a particle residing in one of these states can not interact with another state. The physical world would indeed be quite uneventful should such a condition ever be met.

#### 4.1 Anharmonic Perturbation Theory

Interactions between phonon modes due to third-order terms of the anharmonic crystal potential can be expressed as a second-order perturbation expansion with a perturbing Hamiltonian expressed as [24]

$$H' \equiv \frac{1}{\sqrt{N}} U^{(3)}(Q_0, Q_1, Q_2) \hat{Q}_0 \hat{Q}_1 \hat{Q}_2 \quad (4.1)$$

where

$$U^{(3)}(Q_0, Q_1, Q_2) \equiv \frac{\partial^3 U_E}{\partial Q_0 \partial Q_1 \partial Q_2} \quad (4.2)$$

is the third-order derivative of the crystal potential with respect to the three sets of normal mode coordinates  $Q_0, Q_1, Q_2$ . Here, the subscripts refer to a phonon mode which is characterized by a reduced wavevector  $\bar{q}_l$  and polarization vector  $\hat{e}_l$ . When the phonon modes are expressed in the number basis (Fock Space), the amplitudes of the normal modes can be expressed as operators

$$\hat{Q}_l = \sqrt{\frac{\hbar}{2M\omega_l}} \left( \hat{e}_l a_l e^{i\bar{q}_l \cdot \bar{r}} + \hat{e}_l^* a_l^\dagger e^{-i\bar{q}_l \cdot \bar{r}} \right) \quad (4.3)$$

where  $a_l^\dagger$  and  $a_l$  are the creation and annihilation operators for phonon mode  $l$  respectively, having frequency  $\omega_l$ . In this space, the following bosonic algebra can be employed to calculate the phonon-phonon matrix elements:

$$a_l |..., N_l, ... \rangle = \sqrt{N_l} |..., N_l - 1, ... \rangle \quad (4.4a)$$

$$a_l^\dagger |..., N_l, ... \rangle = \sqrt{N_l + 1} |..., N_l + 1, ... \rangle \quad (4.4b)$$

$$a_l^\dagger a_l |..., N_l, ... \rangle = N_l |..., N_l, ... \rangle \quad (4.4c)$$

$$[a_l, a_l^\dagger] = 1 \quad (4.4d)$$

$$\langle N_m | a_l | N_l \rangle = \sqrt{N_l} \delta_{N_m, N_l - 1} \quad (4.4e)$$

$$\langle N_m | a_l^\dagger | N_l \rangle = \sqrt{N_l + 1} \delta_{N_m, N_l + 1} \quad (4.4f)$$

After a series of tedious algebra, we arrive at the final expression for the square of the matrix element

$$\begin{aligned}
& \left| \langle \dots, N_2, \dots, N_1, \dots, N_0, \dots | H | \dots, N_0, \dots, N_1, \dots, N_2, \dots \rangle \right|^2 \\
&= \frac{\hbar^3}{2M^3 \omega_0 \omega_1 \omega_2} \left| U^{(3)}(Q_0, Q_1, Q_2) \right|^2 (N_0)(N_1 + 1)(N_2 + 1) e^{i(\bar{q}_0 - \bar{q}_1 - \bar{q}_2) \cdot \vec{r}}
\end{aligned} \tag{4.5}$$

When integrated over all modes, the phase factor yields the momentum conservation rule discussed in Chapter 2,  $\bar{q}_0 = \bar{q}_1 + \bar{q}_2 + \vec{G}$ .  $E_0 = E_1 + E_2$ .

## 4.2 Acoustic Phonon Interaction

Empirically determined phenomenological scattering rates ( $\tau \sim 1/\Gamma$ ) for acoustic phonons for each type of scattering mechanism were discussed extensively in [25]. The forms of the equations are based on the early works of Klemens [26], Callaway [27], and Holland [28] and arise from limiting forms of the BTE and the assumption of an isotropic dispersion relation. The general forms for normal (N) and Umklapp (U) phonon-phonon (p-p) as well as phonon-impurity/defect (p-i) scattering rates for acoustic phonons are summarized in Eq. 4.6

$$\Gamma_{p-p,N}^{LA} = A_{N,LA} \omega^2 T^3 \tag{4.6a}$$

$$\Gamma_{p-p,N}^{TA} = A_{N,TA} \omega T^4 \tag{4.6b}$$

$$\Gamma_{p-p,U}^{TA} = \begin{cases} A_{U,TA} / \sinh(\hbar \omega / k_B T) & ; \quad \omega_1 < \omega < \omega_2 \\ 0 & ; \quad \omega < \omega_1 \end{cases} \tag{4.6c}$$

$$\Gamma_{p-i} = (A_{\delta M} + A_{\delta R}) \omega^4 \tag{4.6d}$$

where the set of ‘A’ coefficients are taken to be independent of frequency and temperature. The best-known values for these coefficients, along with the frequency parameters  $\omega_1$  and  $\omega_2$  appearing in (4.6c) for TA Umklapp scattering are listed in Table 1. In (4.6d), the coefficients  $A_{\delta M}$  and  $A_{\delta R}$  correspond to impurity scattering

caused by mass differences and local lattice distortion respectively and are both taken to be proportional to the impurity concentration.

### 4.3 Optical to Acoustic Phonon Energy Conversion

The thermal energy is transported out away from the transistor hotspot primarily through low energy acoustic modes which have group velocities between 5000 and 9000 m/s in silicon. However, as was shown in [29], a significant amount of the thermal energy (as much as 2/3 [30]) is initially stored in the optical phonon modes which have group velocities less than about 1000 m/s. Thermal conduction, therefore, has the potential to be impeded locally as a result of the additional energy decay step required for optical modes to decay into acoustic modes. Many researchers have cited the potential of an energy bottleneck arising from a relatively long relaxation time for optical phonons compared to the electron-phonon scattering time ( $\sim 100$  fs) [31]. Such an intermediate decay process is believed to set an upper limit on the frequency performance for some important III-V quantum well optoelectronic devices [32]. The lifetimes of the optical modes are therefore very important parameters in understanding nonequilibrium heat conduction near the transistor hotspot. In particular, the *g*-type longitudinal optical (*g*-LO) phonon decay rate is thought to play an important role in nonequilibrium energy relaxation in the drain of silicon based transistors and in determining the onset of hot phonon effects [31, 33, 34]. This is because of its strong coupling with high energy electrons [35] and its relatively low modal heat capacity as determined by the inverse of its group velocity. However, despite the importance of the decay processes of the *g*-LO phonon, among others, very little work has been performed in determining either the lifetime or in illuminating the decay channels available to this mode, *cf.* [36]. One of the main reasons for this is that the *g*-LO phonon is not optically active like the zone-center Raman active LO-TO mode (R-LTO). Furthermore, because of the reduced symmetry of the *g*-LO mode, simulations of these phonons require large supercells and therefore extensive computational resources. Sinha *et al.* [13] recently performed detailed MD simulations

of g-LO phonon wavepackets. The key results from that work were that normal three-phonon processes dominated in the relaxation process and that the primary decay channels were of  $\text{LO} \rightarrow \text{LA} + \text{TA}$  type. Despite the complexity and rigor of the MD simulations, a subtle limitation in the choice of interatomic potentials was that they do not reproduce the exact harmonic eigen frequencies of the Si crystal. Hence, the decay channels described in that work are necessarily different than what can occur in Si. To provide additional insight into the physics of the decay process for an arbitrary optical phonon mode in silicon, and in particular the g-LO phonon, we calculate the density of final states  $g_2(\omega\omega_0-\omega)$  [37] for pairs of phonons which conserve both energy ( $\omega_{s_0}(\vec{q}_0) = \omega_{s_1}(\vec{q}_1) + \omega_{s_2}(\vec{q}_2)$ ) and crystal momentum ( $\vec{q}_0 = \vec{q}_1 + \vec{q}_2 + \vec{G}$ ) for an optical phonon with initial wavevector  $\vec{q}_0$  and branch index  $s_0$ . This “final states spectrum” for three-phonon processes arise naturally from third order anharmonic perturbation theory as applied to the calculation of the energy relaxation time for a single mode. Such calculations were rigorously performed using density functional theory (DFT) for the R-LTO mode of silicon by Debernardi and coworkers [12]. The expression for the transition rate under the assumption that only three phonon processes are present is given by Eq. 10 [38]

$$\Gamma(\omega_{s_0}(\vec{q}_0)) \propto \sum_{\vec{q}_1, \vec{q}_2, s_1, s_2} \left| U \begin{pmatrix} \vec{q}_0, \vec{q}_1, \vec{q}_2 \\ s_0, s_1, s_2 \end{pmatrix} \right|^2 \frac{N(\omega_{s_1}(\vec{q}_1)) + N(\omega_{s_2}(\vec{q}_2)) + 1}{\omega_{s_0}(\vec{q}_0) \omega_{s_1}(\vec{q}_1) \omega_{s_2}(\vec{q}_2)} \times \delta(\omega_{s_0}(\vec{q}_0) - \omega_{s_1}(\vec{q}_1) - \omega_{s_2}(\vec{q}_2)) \quad (4.7)$$

where  $U$  is the third-order matrix element. Here, all third order processes except for the process in which the initial phonon decays into two lower energy modes are retained, a reasonable approximation considering the phonon dispersion relation for silicon [12].

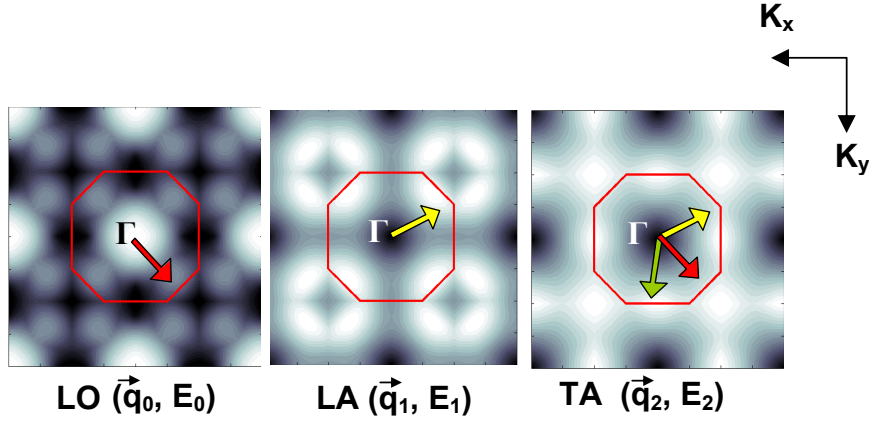


Figure 4.1 Constant energy contours of the silicon phonon band structure along the  $K_x$ - $K_y$  plane for the longitudinal optical, longitudinal acoustic, and transverse acoustic branches (left to right). These contours were calculated using the Valence Force Potential (VFP) approach discussed in Chapter 3.

Using a tetrahedral B.Z. integration method [39-42], we calculate  $g_2(\omega, \omega_0 - \omega)$  for various initial LO phonon modes along  $\Gamma$ -X (including the R-LTO mode, the  $g$ -LO mode and approximately the  $f$ -LO mode as discussed in Section II) by integrating over the entire B.Z. using the full-band phonon dispersion relation in Fig. 2 [14][43]. We restrict our calculations to normal processes, i.e.  $\vec{G} = 0$ . Fig. 3 shows the results of the calculations for LO phonons with initial wavevectors of  $\vec{q}_0 = (2\pi/a)[0, 0, \alpha]$  where  $\alpha = 0, 0.3, 0.5, 0.7$ , and  $1.0$  as labeled (a)-(e) respectively.

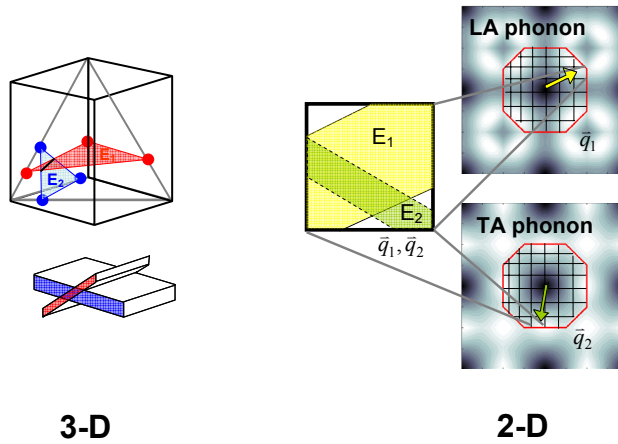


Figure 4.2: Visualization of the mathematical problem required in solving the joint density of states in 3-D and 2-D systems.

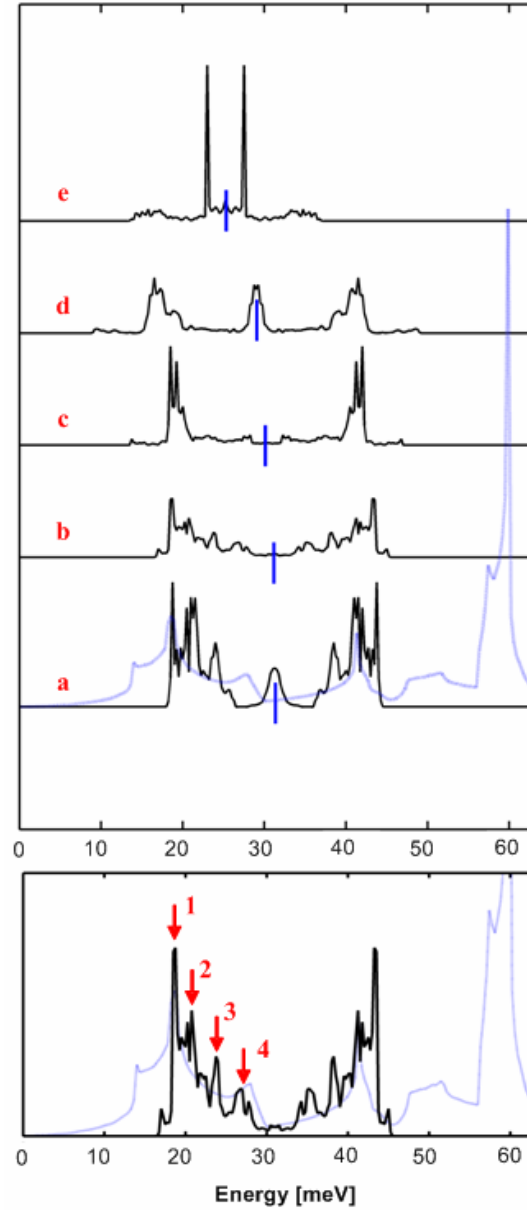


Figure 4.3:  $g_2(E, E_0 - E)$  for an LO phonon of initial energy  $E_0$  and wavevector directed along  $\Gamma$ -X at the points  $(2\pi/a)[0,0,\alpha]$  where  $\alpha = 0, 0.3, 0.5, 0.7$ , and  $1.0$  (labeled (a)-(e) respectively) decaying into two lower energy, crystal momentum conserving normal modes of energies  $E$  and  $E_0 - E$ . The single-phonon density of states  $g(E)$  is shown in light blue for reference on top of (a). (bottom) Zoomed in view of the  $g_2(E, E_0 - E)$  spectrum for the g-LO ( $\alpha = 0.3$ ) phonon. The vertical axes are in arbitrary units.

The single phonon density of states function  $g(\omega)$  for silicon is superimposed on top of the  $g_2(\omega, \omega_0 - \omega)$  plot for the R-LTO mode for easy comparison. We now point out several key features in these plots. First,  $g_2(\omega, \omega_0 - \omega)$  is always even with respect to  $\omega = \omega_0 / 2$ . The monotonic red shift in  $\omega_0 / 2$  seen in going from (a) to (e) is a consequence of the LO dispersion along  $\Gamma$ -X. Secondly, we see that  $g_2(\omega, \omega_0 - \omega)$  is strongly peaked at regions where both  $g(\omega)$  and  $g(\omega_0 - \omega)$  have comparable strength. This is why combinations of low frequency and high frequency modes contribute negligibly to  $g_2(\omega, \omega_0 - \omega)$  in silicon despite the very large  $g(\omega)$  at high frequencies. Finally, we point out the relatively weak central peak around  $\omega_0 / 2$  for the R-LTO mode. This is the so-called ‘‘Klemens channel’’ [44]. As evident in Fig. 3, this channel has a relatively small  $g_2(\omega, \omega_0 - \omega)$  and hence should not dominate in the decay for the R-LTO mode unlike Klemens’ initial postulation. These results are consistent with the work of Debernardi *et al.* [12][45].

Because of the stated importance of the g-LO phonon, we examine its  $g_2(\omega, \omega_0 - \omega)$  spectrum more closely in the bottom panel of Fig. 3. We identify four dominant final state phonon pairs, labeled 1-4, and their representative wavevectors and energies are summarized in Table 2. For each of these pairs, we observe only combinations of the form  $\text{LO} \rightarrow \text{LA} + \text{TA}$  which is consistent with the detailed MD calculations of [13].

Table 4.1. Energy and representative wave vectors for four dominant LA + TA phonon pairs which a g-type LO phonon may create during spontaneous decay while conserving energy and crystal momentum.

	Initial	Pair I		Pair II		Pair III		Pair IV	
	LO	LA	TA	LA	TA	LA	TA	LA	TA
$q_x/(2\pi/a)$	0.0	0.95	-0.95	0.75	-0.75	0.63	-0.63	0.53	-0.53
$q_y/(2\pi/a)$	0.0	0.23	-0.23	0.48	-0.48	0.18	-0.18	0.05	-0.05
$q_z/(2\pi/a)$	0.3	-0.13	-0.43	0.25	0.05	0.53	-0.23	0.68	-0.38
Energy [meV]	62	44	18	41	21	38	24	35	27

#### 4.4 Optical Phonon Lifetimes

Finally, we take our calculations one step further and estimate the intrinsic phonon lifetime for each of the modes shown in Fig. 6 as a function of temperature. We do so by computing Eq. 12 and assuming the third order matrix elements to be a constant equal to  $U_0$  which we fit to the Raman line width data of [46]. After obtaining  $U_0$  for the R-LTO mode, we apply this same factor to the remaining modes of interest and compute the lifetime as a function of temperature. These results are shown in Fig. 4. The inset of Fig. 4 shows the R-LTO linewidth as calculated using our semi-empirical method and compared to the more rigorous calculations of [12] and the experimental data of [46]. Our results are within 10 % of the experimental data over a range of nearly 0-700 K using the single fit parameter,  $U_0$ . The success of this calculation supports the conclusion that the 3<sup>rd</sup> order matrix element magnitudes are only weakly dependent on wavevector. From these results, we see that the g-LO mode is estimated to be around 8.5 ps near 0 K and reduces to about 5 ps at room temperature. Over any operating temperature typical of integrated circuits, the g-LO lifetime is seen to be about twice that of the R-LTO mode but is still of the same order of magnitude. At high power densities, this lifetime is expected to decrease on account of higher phonon occupation. The temperature dependence of the lifetimes arises from the occupation factors in Eq. 10 which are given by the B-E distribution at equilibrium. For nonequilibrium conditions, it is straightforward to compute these lifetimes provided we know the appropriate nonequilibrium occupation factors to apply [47].

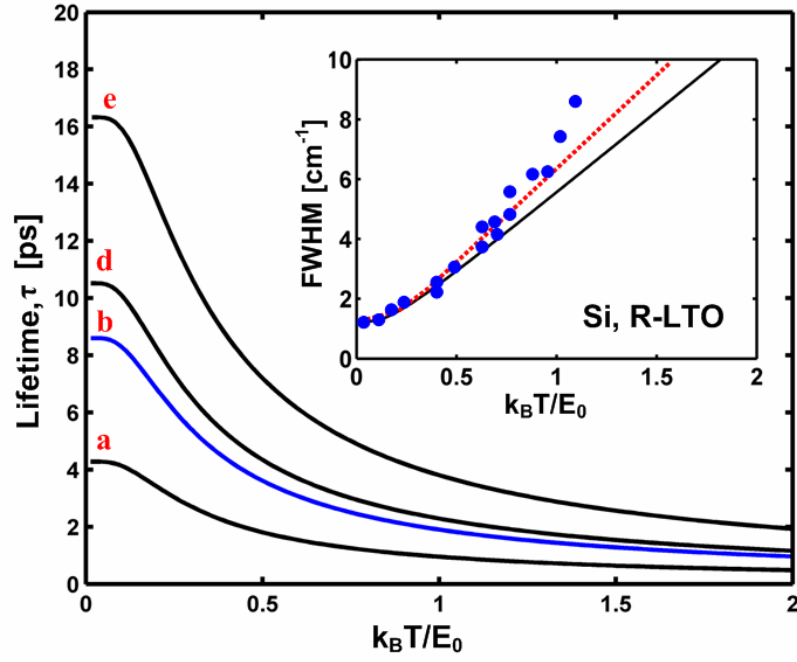


Figure 4.4: Calculated lifetime for LO phonon modes along  $\Gamma$ -X. The inset shows the calculated spectral full-width at half maximum (FWHM) for the R-LTO mode calculated using semi-empirical approach (black solid) and by the DFT calculations (red dotted curve) of [37] compared to experimental data (solid blue dots) [46]. Note that the axes ranges for the inset are the same as that shown in Fig. 2 of [46].

## Chapter 5

### Energy Conversion in Silicon Nanoelectronic Devices

Since a given system can never of its own accord go over into another equally probable state but into a more probable one, it is likewise impossible to construct a system of bodies that after traversing various states returns periodically to its original state, that is a perpetual motion machine.

– L. E. Boltzmann

This chapter describes my work in modeling detailed coupled nonequilibrium electron-phonon transport in nanoscale silicon devices. The Boltzmann transport equation (BTE) is solved for both electron and phonon populations at each grid point in the simulated device in an iterative fashion until convergence is reached. These simulations rely heavily on the Monte Carlo technique, which has been an invaluable tool for nearly four decades in studying semiconductor devices under high-field conditions. This work builds upon the research of two former graduate students in the Goodson Lab. The base of the Monte Carlo simulation code using quadratic phonon dispersion relations was developed by Eric Pop and the idea of the split-flux BTE model was developed by Sanjiv Sinha. Critical modifications to the Monte Carlo simulations were implemented in order to allow for the feedback of the updated phonon populations during the phonon transport sub-routines. From this coupled transport approach, we are able to make important conclusions about the likelihood of hot phonon effects being relevant in silicon devices.

## 5.1 Background

There has been a strong interest in studying electro-thermal processes in silicon electronic devices since the early 1970s, only a short time after large-scale integrated (LSI) silicon logic devices hit the market in high-volume. This was a time when only n-MOS transistors were being employed in integrated circuits, a consequence of the many integration challenges inherent to complementary MOS technology yet to be overcome. The cost to moving to the far-less power hungry CMOS technology imposed an even greater economic barrier, whereby, only until power densities reached the point where cooling chips with a fan became impractical, did it become justifiable. The power density problem appeared to be solved by the move to CMOS technology in the late 1970s, and ushered in the era of ultra-large-scale-integration (ULSI) and the reality of personal computing. However, as devices were aggressively reduced in size into the 1980s in accordance with Moore's Law, a new power density issue began to emerge. Rather than being concerned with overall power consumption and the challenges of maintaining junction temperatures below 90 °C, researchers began to be concerned about the rapidly growing volumetric power densities within individual transistors. Despite the adoption of so-called constant electric field scaling, electric fields in fact continued to increase to the point where the electron population within a few 10's of nanometers of the peak electric field would be sufficiently energized by the field to render it no longer in equilibrium with the lattice, i.e. the phonon system. The electron population near the drain of a transistor could reach nonequilibrium temperatures exceeding 2000 K. This presented a significant reliability problem for CMOS devices and warranted deeper investigation.

The idea of assigning a temperature, a quantity strictly defined at thermodynamic equilibrium, to an electron system which is not in equilibrium with its surroundings would seem meaningless. However, because the electron population thermalization times are typically less than 100 fs, it can typically be assumed that the electrons population itself in equilibrium with itself, i.e. it has redistributed its energy amongst

all of the electrons within a mean-free-path such that it approaches an equilibrium distribution at elevated temperature.

Due to the enormous complexity of transport physics under highly non-equilibrium conditions and to the inability to experimentally probe such microscopic system properties directly, Monte Carlo simulations became an invaluable tool for analyzing the dynamics of energy gain and loss of the electron system. An historical overview of the Monte Carlo technique applied to CMOS, and most often n-MOS devices, is provided by Fischetti and coworkers [48] and an early review of the technique is provided by Jacoboni and Reggiani [49], both pioneers of this field. In the 1990s, a new experimental technique relying on near-infrared electroluminescence of the *hot* electrons in switching transistors was developed [50, 51], enabling researchers to indirectly measure the effective temperature of the electron distribution. Temperatures approaching 2000 K were reported [52].

The continued scaling of transistor dimensions and spatial density continues to cause major thermal management challenges on the chip. Effective removal of heat from the transistor and interconnect layers will be a growing challenge to the successful scaling of digital nanotechnologies for the foreseeable future. These thermal management challenges are necessarily being addressed at all levels of design, from the transistor, to the circuit and microarchitecture, to the package and enclosure. While the challenges are growing more significant at all of these levels, the electrothermal phenomena occurring within transistors are particularly challenging because of the multi-carrier transport physics involved. Reducing channel lengths  $L_g$  in order to increase packing density and to reduce the energy-delay product [10] has a direct impact on the departure from equilibrium of the electron and phonon systems within devices, thus increasing both the complexity and the importance of nanoscale electrothermal phenomena. Transistor-level thermal management is made more important by the move to thin-body single and multiple gate devices that provide improved control of the channel electric field. The most promising multi-gate device is the FinFET [5] of which numerous derivatives have been proposed in recent years [6] [7]. While these devices offer superior subthreshold slopes approaching the theoretical

limit of 60 mV/decade, they exhibit higher thermal resistance than bulk devices because of geometric confinement by low thermal conductivity materials as well as enhanced phonon boundary scattering in the active layers.

This chapter discusses the fundamental heat generation and transport mechanisms in silicon devices and shows why nanoscale and nonequilibrium thermal phenomena are being factored into device technology decisions. The paper is organized into four main sections. Sections II and III deal with the detailed electron transport and the generation and transport of heat at length and time scales less than 100 nm and 10 ps respectively. In Section IV, we close the transport loop by describing efficient simulation techniques for coupling the heat and charge transport, an essential requirement for understanding the thermal impact on electrical characteristics in future devices. In Section V, we use the results of Sections II-IV to understand the implications of nonequilibrium coupled charge-heat transport at nanometer length scales and their impact on leakage power, electrical drive current, and reliability. Additionally, we address the topics of anomalous temperature rise near nanometer-scale heat sources as well as the issue of hot optical phonons in silicon, both topics being heavily debated over the past two decades.

## 5.2 Nanoscale Heat Generation

Within the transistor, thermal energy is predominantly stored and transported by the vibrational modes of the lattice, or phonons, of the semiconducting material. Heat generation is the result of electrons transferring their excess energy gained from the electric field to the phonon population by means of scattering. To model the details of the heat generation process, we use an electron Monte Carlo simulator (*e*-MC) developed by Pop *et al.* in [53] and later modified in [54].

The local volumetric heat generation rate  $Q''(\vec{r})$  (W/cm<sup>3</sup>) is equal to the energy-weighted difference of emitted (*ems*) and absorbed (*abs*) phonons times the ratio of the electron density  $n_e(\vec{r})$  to the number of simulated electrons  $N_{sim}$  (typically 10,000) divided by the simulation time interval step  $\Delta t$  [29] [55]:

$$Q'''(\vec{r}) = \sum_{q,s} Q'''_{q,s}(\vec{r}) = \frac{n_e(\vec{r})}{N_{\text{sim}} \Delta t} \sum_{q,s} \left( \hbar \omega_{q,s}^{\text{ems}} - \hbar \omega_{q,s}^{\text{abs}} \right)_{|\vec{r}} \quad (5.1)$$

where we define the modal volumetric heat generation rate as

$$Q'''_{q,s}(\vec{r}) \equiv \frac{n_e(\vec{r})}{N_{\text{sim}} \Delta t} \left( \hbar \omega_{q,s}^{\text{ems}} - \hbar \omega_{q,s}^{\text{abs}} \right)_{|\vec{r}} \quad (5.2)$$

The frequency spectrum of net emitted phonons for both bulk and strained silicon was computed as a function of electric field and for various doping conditions by Pop *et al.* [29]. At low fields, the heat generation is restricted to intravalley acoustic modes as the electrons do not have sufficient energy to transfer between valleys. At intermediate fields, sharp peaks centered around the signature *g*-type and *f*-type phonons are observed. At higher fields, the emission spectrum broadens about each of the intervalley peaks as a result of the finite phonon dispersion which leads to a gradual relaxation of the *k*-conservation rule associated with the *f* and *g*-type phonons. Fig. 5.1 illustrates this concept.

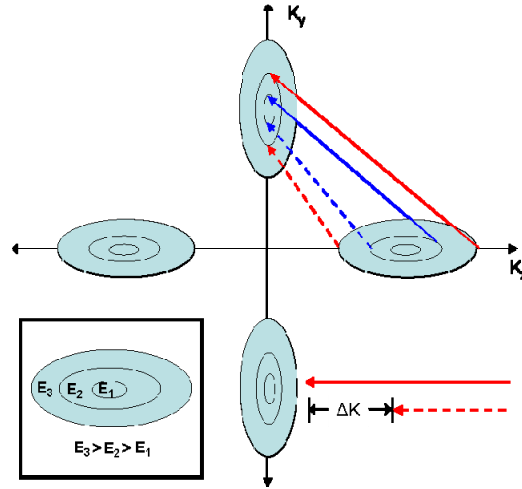


Figure 5.1: Schematic representation of the broadening effect on the phonon generation spectra with increasing power density. The energy contours represent energy spacing of approximately one phonon energy such that a transition from one valley to the other corresponds to a reduction in energy by the amount equal to *f*-type phonon.

### 5.3 Electron Transport

While the models are summarized here, the reader is referred to the original works as well as a large body of work describing the general Monte Carlo technique [48, 49]. The *e*-MC code employs a six valley, analytic, nonparabolic, single conduction band model described by (5.3) [49].

$$E_k (1 + \alpha E_k) = \frac{\hbar^2}{2 m_e} \sum_i \frac{(k_i - k_{vi})^2}{m_{vi}^*} ; i = x, y, z ; v = 1, 2, 3, \dots, 6 \quad (5.3)$$

The effective mass  $m_{vi}^*$  is along the direction index  $i$  for valley  $v$  and the nonparabolicity factor  $\alpha$  is taken to be  $0.5 \text{ eV}^{-1}$  for silicon. The valleys are centered at the six equivalent  $(2\pi/a)[0.85, 0, 0]$  points within the Brillouin Zone (B.Z.). The analytic electron band structure is shown in Fig. 5.2 (red) along with an accurate full band model (blue) [18]. The density of states (DOS) for the analytic band closely matches that for the full-band description below about 1.5 eV [53].

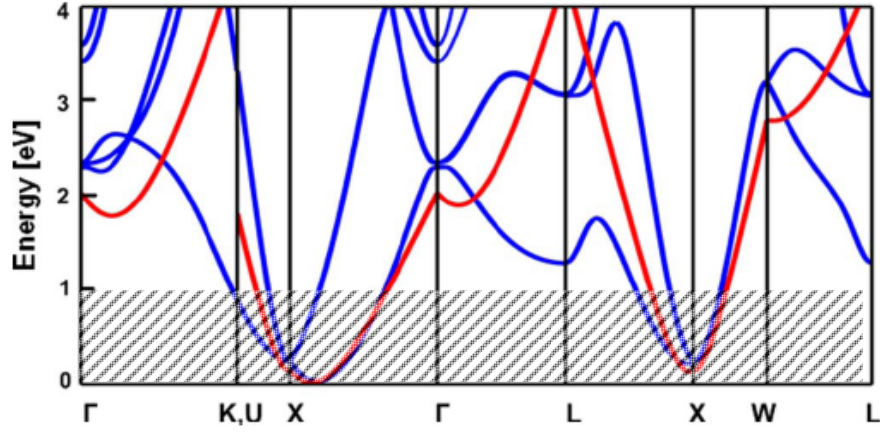


Figure 5.2: Electron conduction band structure for silicon computed using (blue) an empirical tight binding parameter set [18] and (red) a single six-valley analytic nonparabolic electron band [53] used in this work. The shaded region corresponds to the region of the conduction band below 1 eV where electrons are likely to be found in sub-1V devices. As the figure shows, the biggest limitation in the single, analytic, non-parabolic band in low-voltage transport is the inability to capture the structure of the second lowest energy conduction band, which is important in the neighborhood of the X and K,U points.

Since gate voltages are not expected to rise above 1 V for future nanotransistor technologies, the analytic nonparabolic band provides reasonable accuracy while achieving a significant reduction in computational cost compared to more accurate full-band codes [56-58]. By reducing the complexity of the electron band model, the *e*-MC program is able to efficiently incorporate a more detailed dispersion relation model for both acoustic and optical phonons for computing scattering probabilities and energy and crystal momentum conserving final states. The biggest limitation in the single, analytic, non-parabolic band in low-voltage transport is the inability to capture the structure of the second lowest energy conduction band, which is important in the neighborhood of the X and K,U points.

The dispersion for each polarization branch, *s*, is modeled by a quadratic, relation given by  $\omega_{q,s} = \omega_0^s + v^s q + c^s q^2$  where *q* is the wavevector magnitude along an arbitrary direction in the crystal [53]. The bottom panel of Fig. 5.3 shows the isotropic dispersion relation (red) for silicon for directions of high symmetry along with a full-band phonon dispersion calculated using a six-parameter, empirical valence force field

(VFF) model as described in Chapter 3 [14]. The single phonon density of states of these two models is compared in Fig. 5.3. The quadratic parameters used in the *e*-MC simulations here were optimized to fit the dispersion relation along  $\langle 100 \rangle$  direction.

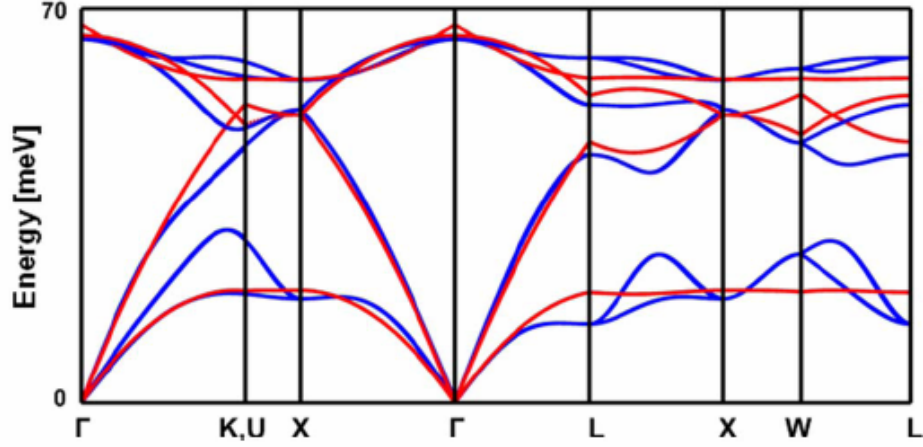


Figure 5.3: Phonon dispersion for silicon computed using (blue) a valence force field model [14] compared to (red) an isotropic model [53] using simple quadratics as used in this work.

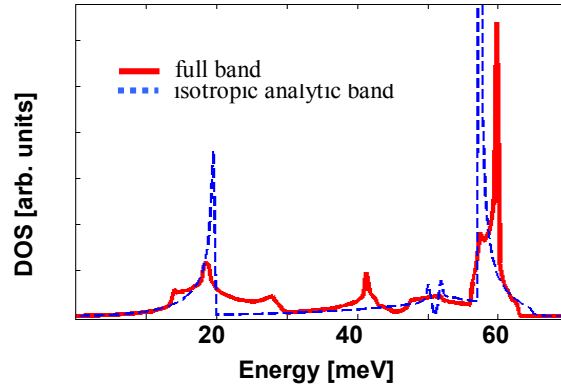


Figure 5.4: Comparison of the single phonon density of states spectrum as computed using a full-band calculation [14] and the isotropic model [53] with quadratic dispersion relations for all branches.

In addition to elastic ionized impurity scattering, both intravalley and intervalley phonon scattering are treated inelastically [59]. Intervalley scattering is modeled by three *g*-type (equivalent valley) and three *f*-type (non-equivalent valley) transitions, both of which are Umklapp (U) processes [60]. The *g*-type phonons are directed along  $\langle 100 \rangle$  and are located at  $(2\pi/a)[0.3, 0, 0]$  and at equivalent points within

the B.Z. whereas the  $f$ -type phonon is directed  $11^\circ$  off the  $\langle 100 \rangle$  equivalent directions at  $(2\pi/a)[1, 0.15, 0.15]$  and at equivalent points within the B.Z. To enable the isotropic phonon dispersion assumption, the  $f$ -type phonons are considered to be directed along  $\langle 100 \rangle$  and positioned at the edge of B.Z. at the **X** point [53].

Electron-phonon scattering is treated in the standard way using Fermi's Golden Rule. The forms of the scattering rates for both intravalley acoustic and intervalley acoustic and optical scattering are, respectively [53]

$$\Gamma_i = \frac{D_{a,s}^2 m_d}{4\pi\rho\hbar^2 k_s} \int_{\omega_{q,s}} \frac{1}{\left(N_{q,s} + \frac{1}{2} \mp \frac{1}{2}\right)} I_q^2 q^3 dq \quad (5.4a)$$

$$\Gamma_{if} = \frac{\pi\Delta_{if}^2 Z_f}{2\rho\omega_{q,s}} \left(N_{q,s} + \frac{1}{2} \mp \frac{1}{2}\right) g_{df}(E_k \pm \hbar\omega_{q,s}) \quad (5.4b)$$

The values for the effective intravalley and intervalley deformation potentials,  $D_a$  and  $\Delta_{if}$  used in this work are reported in Table 1. The upper and lower signs correspond to absorption and emission processes respectively throughout this paper.  $N_{q,s}$  is the average phonon occupation number given by the Bose-Einstein (B-E) distribution,  $N_{q,s} = [\exp(\hbar\omega_{q,s}/k_B T) - 1]^{-1}$  at equilibrium. Under nonequilibrium conditions  $N_{q,s}$  must be determined by solving the Boltzmann Transport Equation (BTE), which we will discuss in subsequent sections.

Table 5.1: Summary of electron and phonon transport model parameters used in simulations discussed in this work.

Electron Transport Parameters							
$\Delta_f^{TA}$	$\Delta_f^{LA/LO}$	$\Delta_f^{TO}$	$\Delta_g^{TA}$	$\Delta_g^{LA}$	$\Delta_g^{LO}$	$D_{LA}$	$D_{TA}$
[eV/cm]	[eV/cm]	[eV/cm]	[eV/cm]	[eV/cm]	[eV/cm]	[eV]	[eV]
$0.5 \times 10^8$	$3.5 \times 10^8$	$1.5 \times 10^8$	$0.3 \times 10^8$	$1.5 \times 10^8$	$6.0 \times 10^8$	6.39	3.01
Phonon Transport Parameters							
$A_{N,LA}$	$A_{N,TA}$	$A_{U,TA}$	$A_{\delta M} + A_{\delta R}$	$\omega_1$	$\omega_2$	$\tau_{op}$	
[s·K <sup>-3</sup> ]	[K <sup>-4</sup> ]	[s <sup>-1</sup> ]	[s <sup>3</sup> ]	$n_0 = 1 \times 10^{20} \text{ cm}^{-3}$	[THz]	[THz]	[ps]
$2.0 \times 10^{-24}$	$9.3 \times 10^{-13}$	$5.5 \times 10^{-18}$	$1.3 \times 10^{-44} (n/n_0)$	23.56	27.49	0.5-10	

It needs to be stated that the choice of phonon deformation potentials used in this work (Table 1) are not unique and that a wide range of values have been reported over the past three decades [e.g., 6,17]. Full band models typically yield lower ( $\sim 2\text{-}3\times$ ) effective phonon deformation potentials than the simpler, non-parabolic band models due to the differences in the electron-phonon joint density of states (JDOS). Even for cases where the bulk observable electron transport properties, such as the mobility, are captured correctly, the large variations in deformation potentials strongly influence the internal specifics of the simulations including the ratio of phonon energy absorption attributed to the various phonon modes. The present calculations employ the multi-valley, nonparabolic band (NPB) model because of its computational simplicity, and does not attempt to resolve the specific limitations of this approximate method. We discuss the possible impact of the differences that may arise due to the use of the NPB model in greater detail in Section 5.8.

#### 5.4 Phonon Transport

The electron system cools in the drain end of the transistor through net phonon emission. The phonon population then proceeds to evolve in a manner which tends to bring the phonon population back towards the equilibrium B-E distribution. This evolution of the phonon distribution can be described by a phonon BTE. Scattering mechanisms include phonon-phonon ( $p\text{-}p$ ), phonon-electron ( $p\text{-}e$ ), phonon-impurity/vacancy ( $p\text{-}i$ ), and phonon-boundary ( $p\text{-}b$ ) types. In this work we use a split-flux form of the phonon BTE ( $p\text{-SFBTE}$ ) introduced by Sinha *et al.* [34] to describe the phonon transport. The  $p\text{-SFBTE}$  was derived under the relaxation time approximation and ensures macroscopic energy conservation. It captures ballistic phonon conduction near the transistor hotspot but also yields a convenient interface to continuum calculations (*i.e.* diffusive conduction) far from the hotspot. In effect, the phonon distribution is split into two populations, the first being a near-equilibrium component,  $N_{q,s}(T_F)$  which has the B-E distribution corresponding to a temperature  $T_F$  which obeys Fourier's heat conduction law. The second population  $n_{q,s}$  is a

nonequilibrium departure component, which dominates the transport near the hotspot and is determined by solving the phonon BTE in the relaxation time approximation given by (5.5).

$$\mathbf{v}_{q,s} \cdot \nabla_{\mathbf{r}} n_{q,s} = -\frac{n_{q,s}}{\tau_{q,s}} + \dot{n}_{q,s} \quad (5.5)$$

Here,  $\mathbf{v}_{q,s}$ ,  $\tau_{q,s}$ , and  $\dot{n}_{q,s}(\vec{r})$  are the modal group velocity, lifetime, and source term (in units of #/s) respectively for a phonon of mode  $q$  and branch  $s$ .  $\dot{n}_{q,s}(\vec{r})$  is determined from the  $e$ -MC simulation output and is directly related to the modal volumetric heat generation rate  $\dot{Q}_{q,s}''(\vec{r})$  given by (5.1) through division by the number of phonon states per unit volume  $g(\omega_{q,s})\Delta\omega$

$$\dot{n}_{q,s}(\vec{r}) = \frac{\dot{Q}_{q,s}''(\vec{r})}{\hbar \omega_{q,s} g(\omega_{q,s}) \Delta\omega} \quad (5.6)$$

Once  $n_{q,s}$  is determined, macroscopic energy conservation is used to determine  $T_F$  according to (5.7).

$$\frac{1}{(2\pi)^3} \int \frac{n_{q,s}(\vec{r})}{\tau_{q,s}} \hbar \omega_{q,s} d\vec{q} + \vec{k} \cdot \nabla_{\mathbf{r}}^2 T_F(\vec{r}) = 0 \quad (5.7)$$

Here,  $\vec{k}$  is the thermal conductivity tensor, which for thin films can be appropriately modified to account for increased boundary scattering. It is important to recognize that while phonon-boundary scattering can reduce the effective thermal conductivity of thin films, this additional scattering mechanism may not necessarily force the phonon system farther from equilibrium. On the contrary, the additional scattering may serve to allow the local phonon system to more closely approximate the B-E distribution at a

given temperature. In contrast, the introduction of specific phonon modes due to electron scattering is inherently disruptive to the phonon distribution function and can cause severe departure from equilibrium. Boundary scattering becomes important if the smallest dimension of a material domain is comparable to the mean-free-path ( $\lambda$ ) of a phonon mode, given by the product of the modal group velocity and lifetime ( $\lambda_{q,s} = v_{q,s} \cdot \tau_{q,s}$ ) [61] [62]. For LA modes, which conduct most of the heat,  $\lambda$  is of the order of 100 nm near room temperature. However, for optical phonons,  $\lambda < 10$  nm and so only in ultrathin films or inversion layers will boundary scattering play an important role in optical phonon transport.

Once  $N_{q,s}(T_F(\vec{r}))$  and  $n_{q,s}(\vec{r})$  have been determined from (5) and (7), an effective temperature,  $T_{eff}$ , can be defined by equating the total energy density at a particular location to that for an equivalent B-E distributed population and integrating over the appropriate polarization branches and wavevector space according to 5.8.

$$\frac{1}{(2\pi)^3} \sum_s \int N_{q,s}(T_{eff}) \hbar \omega_{q,s} d\vec{q} = \frac{1}{(2\pi)^3} \sum_s \int (N_{q,s}(T_F) + n_{q,s}) \hbar \omega_{q,s} d\vec{q} \quad (5.8)$$

If we restrict the integration to a particular polarization branch then we obtain an effective branch temperature. If we further restrict the integration to include only a single wavevector and branch, then we obtain an effective temperature for a single mode at  $\omega_{q,s}$  [63]. The use of an effective temperature is merely a convenient means for communicating the degree to which a particular segment of the phonon population deviates from equilibrium. Ultimately, what matters is the modal occupation number since this determines the strength of the scattering processes through (5.4).

The transient form of the  $p$ -SFBTE can also be found in [34]. In that work, Sinha *et al.* examined the important issue of phonon population build-up between successive clock cycles using a typical phonon generation spectrum calculated by Monte Carlo simulations. They concluded that the optical phonon lifetimes were sufficiently short as to prevent phonon accumulation from cycle to cycle for typical operating frequencies. Although the transient problem is important, we will continue to focus

our attention on the steady-state solutions within this paper. We now turn our attention to the determination of the modal lifetimes  $\tau$  which are essential for capturing the transport physics of the microscopic system.

### 5.5 Coupled Electron-Phonon Transport

To model the effects of self-heating in a transistor, the electron and phonon systems must be fully coupled together. It is not sufficient for electron-phonon scattering to be simply included in an electron transport model since the phonons that are generated during the simulation are not “sensed” by the simulated electrons. There needs to be a way to feed the updated occupation numbers back into the calculation of the  $e$ - $p$  scattering rates. Furthermore, the phonons generated during the simulation must be allowed to propagate and decay as they would in a real device as discussed in Section III. The complexity and magnitude of such a task has prevented truly rigorous solutions of the coupled transport physics at such length and time scales. Various approximations in either the electron or phonon models are typically necessary to make the problem tractable. In references [64-66], either moments of the phonon BTE or the use of a ballistic-diffusive form of the BTE using major simplifications in the electron and phonon dispersions were performed. Lake and Datta [67] used a nonequilibrium Green’s-function formalism to study detailed energy transfer between electrons and phonons in a mesoscopic diode. As discussed briefly in [54, 68], we have chosen to fully couple the electron and phonon populations by combining the  $e$ -MC technique described in Section II with the  $p$ -SFBTE described in Section III and by solving the two transport problems in sequential iteration. Provided the simplifications made in modeling both the electron and phonon systems as well as in the solution techniques for solving the phonon BTE, we are able to examine the coupled transport physics for realistic devices while retaining valuable physical insights into the spectral decomposition of the heat within the device at all segments of the calculation.

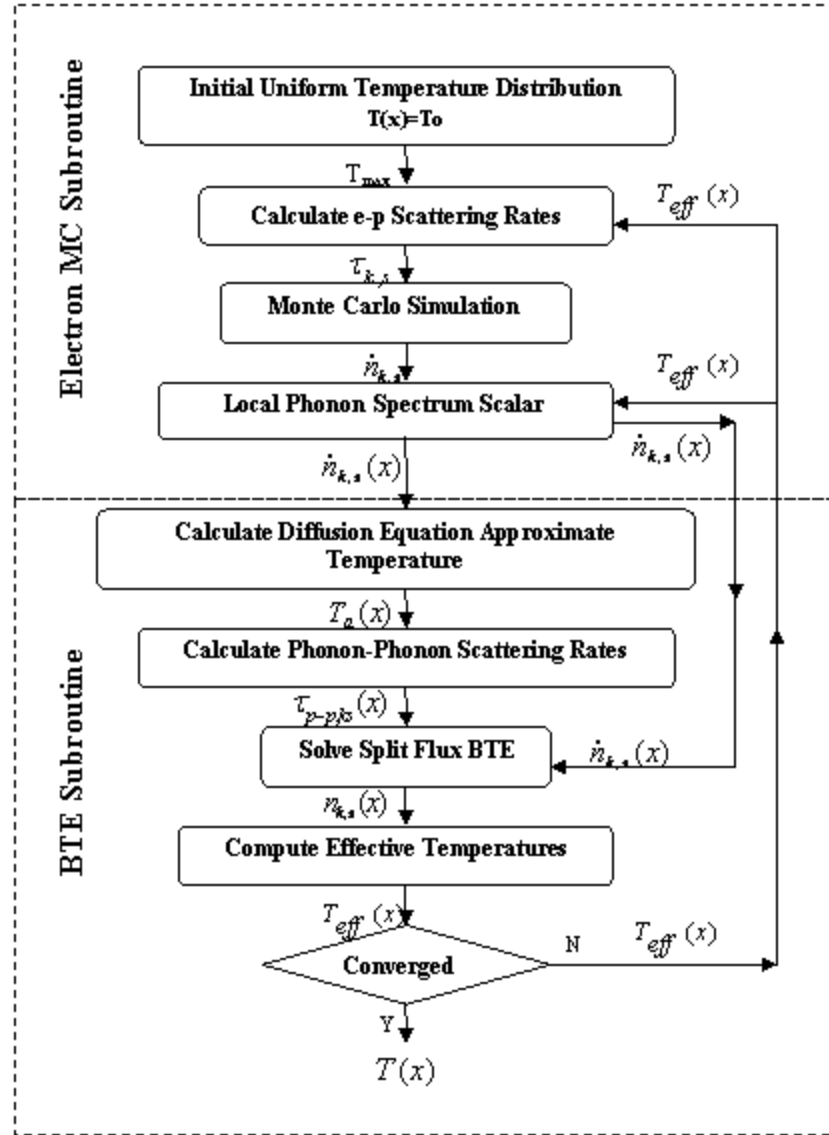


Figure 5.5: The algorithm for combining the electron MC simulation with the phonon BTE solution.

For each iteration, two independent simulations are performed: one for the electron system and the other for the phonon system. Outputs from each simulation are fed back into the other and the simulation proceeds until satisfactory convergence is achieved. We find that this method typically achieves convergence within five iterations. The coupled simulation begins with an isothermal (300 K) *e*-MC simulation whose initial conditions are given by a drift-diffusion device simulator such as

MEDICI. The *e*-MC computes electron transport self-consistently with the electric field by solving the Poisson equation at all steps across the device grid. Net phonon generation rates as a function of position and phonon frequency are gathered from the *e*-MC and fed into the *p*-SFBTE portion of the code. In the latter, the phonons are allowed to propagate in the absence of the electron system and only *p-p* and *p-i* scattering are included. The scattering rates for acoustic modes are determined at the beginning of the *p*-SFBTE simulation using (4.6). The temperature dependence of the *p-p* scattering rates are treated by using the temperature field calculated from the classical heat diffusion equation and taking the volumetric power generation rate from the *e*-MC simulation as the source term. Because of a lack of experimental data, we have typically assigned a single lifetime for all optical modes to be in the range of 0.5-10 ps which is consistent with Raman data and theoretical calculations. At the end of the *p*-SFBTE calculation, an updated distribution of phonons as a function of position is computed. This distribution of phonons is then used to compute the electron-phonon scattering rates for the subsequent *e*-MC simulation by updating the phonon occupation numbers. It can be impractical to use separate phonon occupation numbers for all modes and branches and for all grid points in computing the scattering rates. Therefore, we typically compute an effective temperatures for the dominant optical *f* and *g*-type phonons as well for the LA and TA branches. In all, we compute and pass six position dependent effective temperature vectors back to the *e*-MC. These temperature vectors are then used to adjust the scattering rates in a manner we will discuss shortly. Before doing so, we note that aside from the added complexity and the reduction in computational speed, there is nothing fundamentally preventing the use of additional temperatures to account for the occupation of individual phonon modes or ranges of phonon modes during scattering rate calculations.

The maximum scattering rate for each electron-phonon scattering type is then calculated using the corresponding maximum effective temperature and the simulation begins. To include the dependence of the local phonon occupation, we then employ a temperature-based (or rather occupation number-based) rejection algorithm, a technique commonly used in the Monte Carlo technique [49]. When an electron at a

grid location  $r_i$  is chosen to scatter with a particular phonon of type  $j$ , the local effective temperature  $T_{eff,j}(r_i)$  is compared to the maximum effective temperature  $T_{eff,j,MAX}$  through (5.9)

$$\eta_j(r_i) = \frac{N_{q,s}(T_{eff,j}(r_i)) + \frac{1}{2} \mp \frac{1}{2}}{N_{q,s}(T_{eff,j,MAX}) + \frac{1}{2} \mp \frac{1}{2}} \quad (5.9)$$

where  $0 \leq \eta_j(r_i) \leq 1$ . A random number  $X$  with uniform probability density over the unit interval  $[0, 1]$  is then generated and if  $\eta_j(r_i) < X$  then the scattering event is allowed to take place. Otherwise, it is rejected and the electron continues on its initial trajectory unperturbed, *i.e.* is treated as a self-scattering event. With the phonon generation source term output from the *e*-MC and the electron-phonon scattering rates being adjusted to account for phonon occupation (via the effective temperature and the rejection algorithm), the electrons and phonons form a closed-loop system. We now discuss the application of this algorithm to the simulation of a simple 1-D silicon device.

Figures 5.6 and 5.7 show the 1-D n+/n/n+ silicon device along with its electrical characteristics which we have used extensively in developing the *e*-MC and *p*-SFBTE and coupled simulation algorithms [33, 54, 68]. Although the device is infinite in extent in the transverse plane and lacks a gate terminal, such a device structure resembles the core of transistor structures such as the FinFET. The band diagram along the channel is similar to that along the channel of typical CMOS devices, especially DG devices where vertical (transverse) symmetry exists. Additionally, this device structure allows us to extract physical insight into the energy relaxation processes which can be applied to more sophisticated 3-D device structures.

The device consists of three regions. Two 150 nm “source/drain” regions are doped to  $10^{20} \text{ cm}^{-3}$  and are separated by a 20 nm lightly doped ( $10^{16} \text{ cm}^{-3}$ ) “channel” region [69]. For thermal boundary conditions the departure from equilibrium phonon

population,  $n_{q,s}$  is taken to be zero for all modes at the left and right contacts, *i.e.* the contacts are treated as perfect thermal reservoirs. Furthermore, the temperature  $T_F$ , which determines the distribution of the near-equilibrium phonon population, is set to 300 K at both contacts and its spatial derivative is continuous.

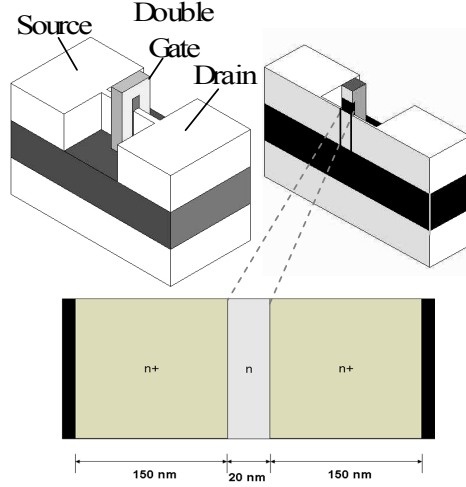


Figure 5.6: A schematic representation of the quasi-1D two-terminal n+/n/n+ device structure modeled in this work and a comparison to a more realistic device structure.

Fig. 6 shows the steady-state power density generated within the device at three different bias conditions (0.6, 0.8, and 1.0 V) and were computed using (5.1). As discussed in [33], nearly all of the power generation within this device structure occurs within the drain, a common characteristic for transistors operating in the quasi-ballistic transport regime. Furthermore, the generation profile is exponentially decaying with a characteristic length of about 20 nm which is essentially independent of applied drain voltage. This lack of dependence of drain voltage was attributed to the fact that the average electron velocity as well as the electron-phonon scattering rates both scale approximately as  $\sim(E - E_g)^{1/2}$  and thus the energy relaxation length remains essentially constant. The peak average kinetic energy gained by the electrons across the channel is found to be proportional to the drain-to-source voltage  $(E - E_g) \sim 0.4(eV_{ds})$  [33]. The impact on energy relaxation length by a fully-silicided drain located within 20 nm of the channel/drain boundary has not been investigated.

However, because of the amorphous nature of the silicide, the scattering length is expected to be of the order of the disorder length,  $\sim 1$  nm. We believe that in the case that the fin extension is less than 20 nm, the energy relaxation length will likely be reduced and will be comparable to the fin extension length itself.

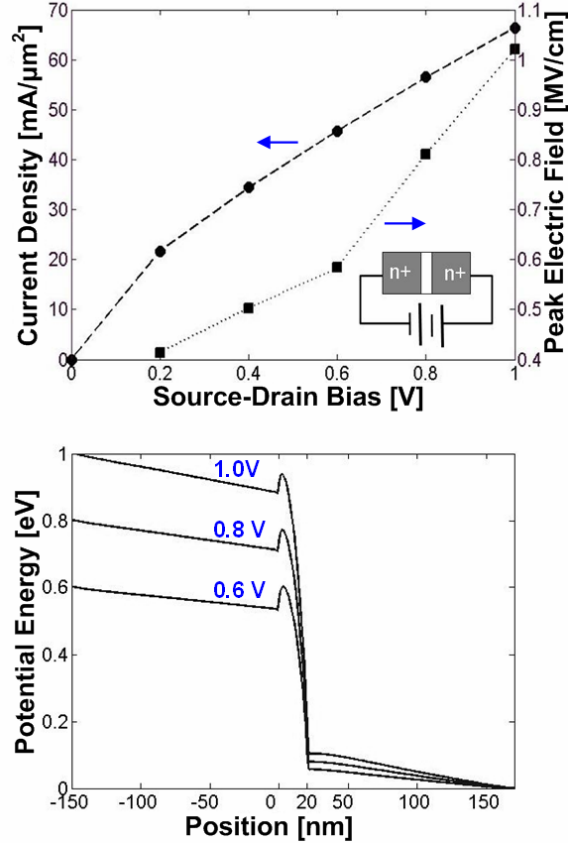


Figure 5.7: Electrostatic characteristics of the 1-D n+/n/n+ device simulated in this work. The device has three regions: two 150 nm n+ ( $10^{20}$  cm $^{-3}$ ) source and drain regions separated by a 20 nm n-type ( $10^{16}$  cm $^{-3}$ ) “channel”. The doping is uniform within each region and 1.25 nm/decade characterizes the doping concentration roll-off between regions. (top) current density (left axis) and peak electric field (right axis) vs. source-drain bias voltage. (bottom) electron potential energy vs. position within the device for three bias conditions.

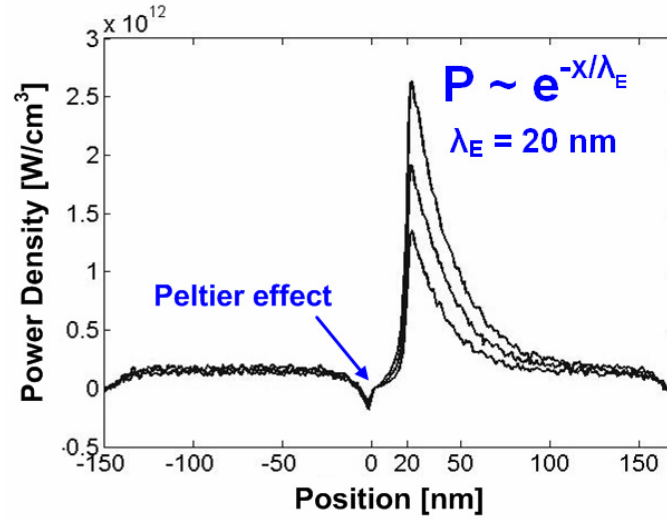


Figure 5.8: Volumetric power generation rate vs. position within the 1-D  $n^+/n/n^+$  device. The power generation at each grid point was computed according to (5.1).

Fig. 5.9 shows the phonon generation spectra computed at four positions within the drain region beginning with the peak power generation point. As the electron system continues to cool deeper into the drain, the phonon emission spectrum becomes more concentrated about the  $f$  and  $g$ -type phonons. The impact on average electron energy is shown in Fig. 10. The average energy increases in both the source and drain regions but there is little effect in the channel region, a result of the near-ballistic transport across the channel. Furthermore, despite an appreciable temperature rise within the device, it is found that the device current is reduced by only about 1 % at 1 V. The reduction in drive current is attributed to an increase in the drain scattering, which leads to a slight increase in the channel barrier height and ultimately a reduction in the electron source injection rate. A slightly larger drain current reduction would occur if it wasn't for the partial compensation of the higher injection velocity of source electrons due to the temperature rise at the source. We see that overall, the effective of temperature on a device in the on-state is reduced.

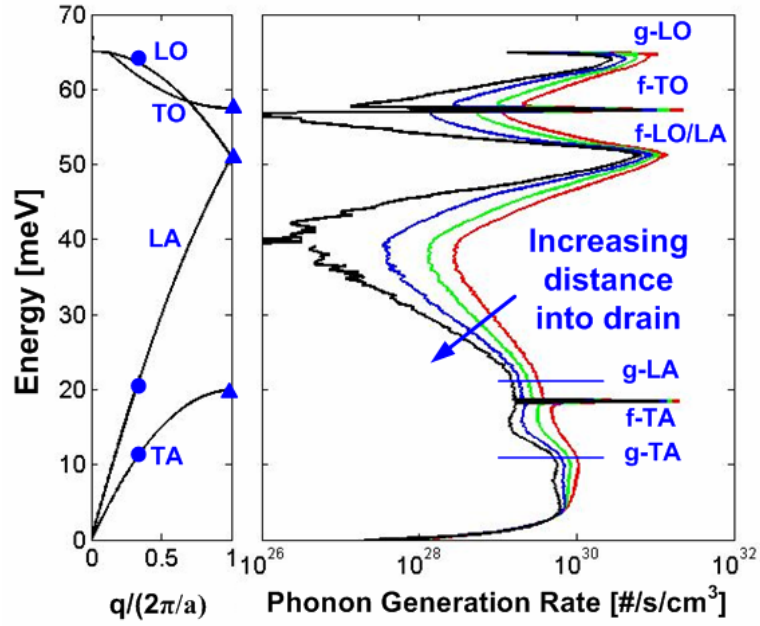


Figure 5.9: (left) Phonon dispersion and (right) phonon generation spectra computed at 4 locations within the device for the 1V case. The red (right-most) curve corresponds to the location of the peak power dissipation. The remaining curves correspond to  $r = 10, 20$ , and  $30$  nm displaced from the peak generation point or hotspot within the drain.

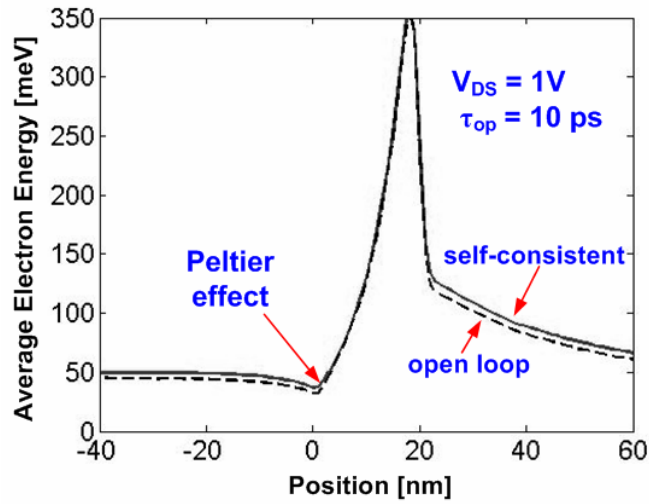


Figure 5.10: Average electron energy as a function of position for an optical phonon lifetime of  $10$  ps for uncoupled (dashed line) and fully-coupled/closed-loop (solid line) simulations.

## 5.6 Hot Optical Phonons

Highly nonequilibrium optical phonon populations have the potential to significantly impact electrical transport and impede thermal conduction near the hotspot. The effects of hot optical phonons have been widely observed in III-V semiconductor devices through observation of extended energy relaxation times in ultrafast optical measurements [70] as well as through excess noise at microwave frequencies [71]. Recently, negative differential conductance in suspended 2-terminal, metallic, single-wall carbon nanotube (CNT) devices were attributed to hot optical phonons [72]. There it was postulated that the optical phonon lifetime was significantly enhanced by lack of a substrate. A large population of nonequilibrium optical phonons in the drain of a transistor may cause the energy relaxation length to extend beyond the 20 nm characteristic length calculated in this work. Furthermore, the increased scattering rate would lead to a marked reduction in mobility in the drain and perhaps even lead to negative conductance effects. In order for these types of effects to take place, the optical phonon occupation numbers need to be comparable to or larger than unity such that the absorption processes are comparable to the emission processes.

The upper panel of Fig. 5.11 shows the profiles of the effective temperature for the entire phonon population. Near the hotspot, the phonon temperature is dominated by the effects of ballistic transport and an anomalous temperature rise is observed.

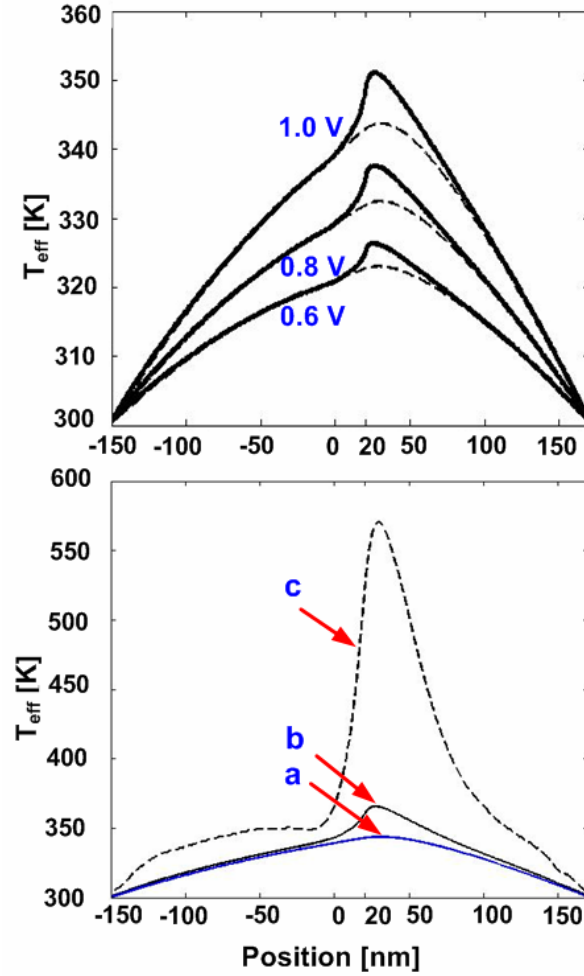


Figure 5.11: (top) Effective lattice temperature ( $T_{eff}$ ) computed using the self-consistent e-MC/p-SFBTE method. (bottom)  $T_{eff}$  for the LO phonon branch (solid black - b) compared to that of the g-LO phonon (dashed black - c) for the 1 V bias condition. The temperature computed from classic heat diffusion equation is also shown for reference (solid blue – a).

Beyond about 50 nm into the drain, the temperature profiles are predicted well by the classical heat diffusion equation (dashed lines). In the bottom panel of Fig. 5.11, we compare the effective temperatures for the LO branch and the isolated g-LO mode obtained by appropriately restricting the integration in Eq. (8). From this figure we can see how poorly an effective LO branch temperature would describe the large departure from equilibrium exhibited by certain individual modes within the LO branch.

The phonon energy distribution at the peak power generation point in the device is shown in Fig. 5.12. The acoustic phonons ( $< 51$  meV) are well-behaved in that they closely adhere to a B-E distribution of elevated temperature  $T_{eff,(LA,TA)} = 343$  K.

However, the optical phonons ( $> 51$  meV) exhibit significant departures from the B-E equilibrium distribution. Temperatures near the zone center approach 1000 K and the g-LO phonon peaks at about 570 K for an optical phonon lifetime ( $\tau_{op}$ ) of 10 ps. The inset of Fig. 5.12 shows the dependence of the phonon distribution on the optical phonon lifetime parameter as it is varied from 0.5 to 10 ps. The deviation of the optical phonon modes from the equilibrium distribution diminishes in direct proportion to the phonon lifetime and a B-E distribution describes the entire phonon population well for  $\tau_{op} < 1$  ps for the power densities simulated.

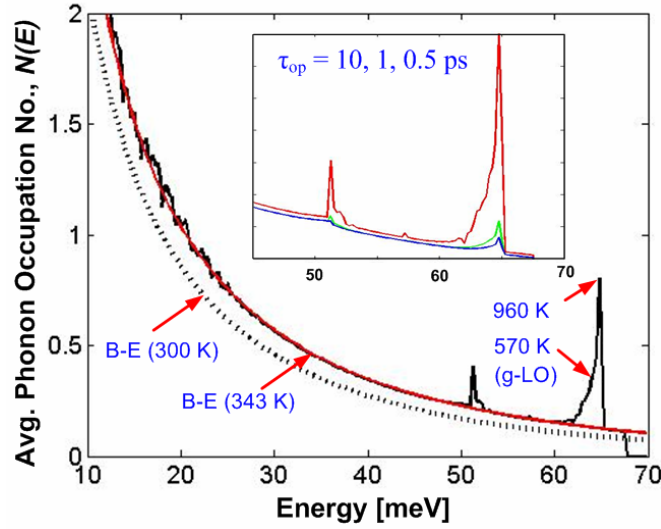


Figure 5.12: Phonon distribution at peak power generation point ( $r = 25$  nm). The inset shows a zoomed in view of optical mode occupation as a function of optical phonon lifetime ( $\tau_{op} = 0.5, 1, 10$  ps).

As discussed in Section III, the g-LO phonon has been flagged as a good indicator for hot phonons in silicon. A simple expression for the excess occupation number, derived from a single mode rate equation is given by (5.10).

$$n_{g-LO} \approx \frac{Q_{g-LO}''' \tau_{g-LO}}{\hbar \omega_{g-LO} g(\omega_{g-LO}) \Delta \omega} \quad (5.10)$$

where  $Q_{g-LO}'''$  is the volumetric power being transferred to the g-LO phonon by the electrons in the spectral width  $\Delta\omega$  and  $\tau_{g-LO}$  is the modal relaxation time, estimated to be about 5 ps at room temperature. This expression is valid only under homogeneous power generation conditions and tends to be an overestimate of the population by a factor of about 2. Fig. 5.13 shows the excess occupation number, for the g-LO mode as a function of the volumetric power density for bulk silicon as well as at the hotspot of the 1-D device of Section IV.

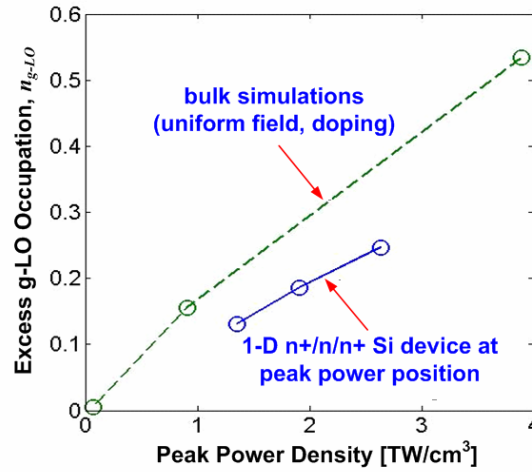


Figure 5.13: Excess occupation number for the g-type longitudinal optical (g-LO) phonon as a function of volumetric power density as calculated for uniform power generation (upper dashed line data set) and as calculated at the peak power generation point in the 1-D device (lower solid line data set).

We find that the excess occupation number increases approximately as  $n_{g-LO} \sim 1 \times 10^{-10} Q_{tot}'''^{3/4}$  where  $Q_{tot}'''$  is the total volumetric power density measured in  $W/cm^3$ . Reducing this number by a factor of 1/2 to account for nonuniform heat generation, we find that the nonequilibrium occupation number will reach unity for a power density of about  $20 \text{ TW/cm}^3$ . This assumes that the phonon lifetime does not decrease with increasing power density. If we assume that the phonon lifetime is reduced to subpicosecond levels at high power densities, as some recent experimental work has indicated may be likely [73], then the critical power density would exceed  $200 \text{ TW/cm}^3$ . According to our estimates in Figure 1.4, this level is just beyond the

projected power density for well-behaved, end-of-the-roadmap FinFET devices. Since the transistor volumetric power density is sensitive to the fin (body) thickness, statistical process variations may cause a significant fraction of the transistors across a chip to cross into the regime where nonequilibrium heating plays a significant role in the electrical transport. This may lead to an enhancement in the variation of performance across die and wafer. In addition to a critical volumetric power density, we can estimate a critical electric field by relating the peak power density for ballistic transport to the peak power density given by the classical joule heating term as  $\max\{Q_{\text{ballistic}}'''\} \sim \max\{\vec{J} \cdot \vec{E}\} / \gamma$  where  $\gamma$  can be taken to be  $\sim 2-3$  [33]. By assuming the current density to be given by  $|\vec{J}| \sim \beta / L_g \text{ } \mu\text{A}/\text{nm}^2$  where the magnitude of  $\beta$  is equivalent to the number of gates electrodes (also  $\sim 2-3$ ) times  $1 \text{ } \mu\text{A}/\text{nm}$ , we arrive at an expression for the critical electric field,  $|\vec{E}|_{\text{crit}} \sim (\gamma/\beta) \max\{Q_{\text{ballistic}}'''\} \cdot L_g$ . For a 10 nm device, a critical power density of  $200 \text{ TW}/\text{cm}^3$  would correspond to a peak electric field of about  $20 \text{ MV}/\text{cm}$ , about 40x the field strength associated with dielectric breakdown in silicon.

Interfacial or boundary scattering was not included in this simulation but would act in a way to reduce the optical phonon temperature as well as impede acoustic phonon conduction especially in thin-body devices such as FinFETs. The effective thermal conductivity for a 22 nm thin film of undoped single-crystalline silicon was measured to be only  $20 \text{ W}/\text{m}\cdot\text{K}$  near room temperature, about an order of magnitude lower than the bulk undoped value of  $148 \text{ W}/\text{m}\cdot\text{K}$  which was used in this work. A 10 nm film is expected to be reduced to just  $13 \text{ W}/\text{m}\cdot\text{K}$  [74] as the conductivity scales approximately as  $\delta = d_s / \Lambda_b$  [61] where  $d_s$  is the film thickness and  $\Lambda_b$  is the bulk phonon mean-free-path ( $\sim 100 \text{ nm}$ ). The consequences of thermal boundary scattering for ultra-thin body silicon and germanium devices were recently discussed by Pop *et al.* [74]. Reductions in the effective thermal conductivity for the silicon layer in quasi-1D devices such as FinFETS are expected to be even more severe than 2-D films.

## 5.7 The Size Effect

In this section, we discuss some of the consequences of nanoscale and nonequilibrium transport which were discussed in detail within the previous three sections on leading edge and future thin-body silicon FET devices. First, we address the origins of what various researchers have referred to as the “size-effect” of the heat generation source in bulk materials [75]. For realistic materials, there may be an increase in a local temperature as we have defined in Eq. 8 relative to the temperature field which would be calculated using diffusive conduction equations, *i.e.* Fourier’s Law. This was quite evident in Fig. 8 for the 1-D device. This rise in effective temperature is the consequence of a skewed phonon distribution near the heat generation source (Fig. 10) which tends to favor higher energy modes (with lower group velocities) at high power densities. The additional thermal resistance which leads to this increased temperature rise near the source is directly related to the energy relaxation processes between optical and acoustic phonons as discussed in Section III. This additional thermal resistance disappears with a vanishing optical phonon lifetime and the size of the source does not dictate the magnitude of the anomalous temperature rise, unlike what has been hypothesized [34, 76, 77].

The anomalous increase in the effective temperature within the first 20-50 nm of the channel/drain boundary acts to reduce the device current and may have an appreciable impact on the reliability of the transistor. Despite the ballistic nature of nanometer scale devices, the drain and source are still electrically coupled; to maintain current continuity, the potential barrier seen by the source electrons will increase in the presence of increased scattering in the drain which acts to reduce the source injection rate. In contrast, an anomalous temperature rise at the source-channel boundary acts to increase the electron injection rate. For the device studied in Section IV, we observed a reduction in the drive current by about 1 %, a rather weak effect. Although direct evidence of an anomalous increase in the device temperature is lacking, there are some preliminary indications that transistor reliability may be impacted by the effects of nonequilibrium phonon populations [78].

### 5.8 Limitations of the Analytic Electron Band Model

Finally, we address the consequences of our choice in using a simplified analytic nonparabolic band (NPB) model. As we mentioned briefly in Section II, the intervalley deformation potentials used in this work are about twice the values computed using full-band models [57, 79]. This is despite the fact that the NPB model reproduces the electron DOS reasonably well up to about 1.5 eV. Fischetti and Laux [57] attributed the smaller deformation potentials of the full-band model to the ability of the electron population to reach a lower energy configuration through equivalent valley repopulation by means of an enhanced contribution from Bloch oscillation, which the NPB model fails to describe adequately. Using the NPB model it was shown that the ratio of energy being dissipated by optical and acoustic phonon modes is  $\sim 2:1$  [29]. However, full-band MC codes have shown the reverse ratio of  $\sim 1:2$ . To account for the deficiencies of the NPB model, we can somewhat crudely add an additional factor of  $\frac{1}{2}$  to Eq. 12 and increase the estimates for the onset of hot optical phonons to power densities approaching  $500 \text{ TW/cm}^3$ . Full-band calculations would undoubtedly improve the accuracy of these estimates but they should not change one of the more important conclusions of this work, that hot optical phonons are unlikely to play a significant role in the electrical behavior of well-behaved silicon devices described by the existing technology roadmap.

### 5.9 Summary

Transistor designs over the next decade will feature confined geometries with increasing surface-to-volume ratios and rising volumetric power densities. Thermal conductance within the transistor will be dramatically reduced due to increased surface scattering and by the confinement of intrinsically low-conductivity materials. Furthermore, thermal boundary resistance at interfaces between dissimilar materials, including Si and SiGe alloys, will lead to higher junction temperatures and therefore

higher leakage currents than what may be expected. Reliability will be impacted by the increase in junction temperatures caused by lower thermal conductance and may also be influenced by nonequilibrium optical phonons in the drain. Mobility reduction due to hot optical phonons does not seem to be a major threat to the near-term evolution of CMOS technology. However, this assessment only stands for well-behaved devices operating under ideal conditions through the end of the current roadmap. In some analog or high-power applications, hot phonons could play a major role. Based on our work, we estimate that hot phonon effects will not play a significant role for power densities below about  $100 \text{ TW/cm}^3$  for silicon based devices if we consider optical phonon lifetimes to reduce below 1 ps under typical operating conditions. For silicon-based quantum well devices, the models used in this work will need to be modified appropriately and the issue of hot phonons will need to be reevaluated. Our conclusions are largely dependent on the optical phonon lifetime parameter. Therefore, experimental validation of the optical phonon lifetimes for realistic device conditions is essential.

## Chapter 6

### Optical Delay-line Pump-Probe Measurements

To acquire lore, one needs a big sandbox and long uninterrupted stretches of time to spend there, absorbed in the play.

To improve a measurement, you have to know what limits it.

- P. C. D. Hobbs

In this chapter, I will discuss my development of a custom two-color optical delay line pump probe system which was used to measure short timescale energy transfer physics in silicon nanoparticle oxide films and other nanostructured materials. I developed the system in collaboration with Matt Panzer, a fellow graduate student in the Goodson Lab at Stanford University. The system was used extensively in two basic experimental configurations which enabled differential reflection and differential transmission style measurements. The focus of my work was on the development of differential transmission measurements for the purpose of characterizing carrier dynamics and energy transfer in optically pumped silicon nanoparticle-oxide films. However, the co-development of transient thermo-reflectance measurements proved to be an invaluable technique for calibrating the accuracy of the pump-probe system on account of the linearity of the technique and the ability to rely on well-known thermal properties of common materials produced in an academic fabrication facility, e.g. thin film aluminum and thermally grown silicon oxide. The subject of Chapter 7, will be the results of the pump-probe studies performed on silicon nanoparticle oxide films.

#### 6.1 System Layout

Figures 6.1 and 6.2 shows a simplified schematic and images of the existing picosecond pump-probe optical metrology platform developed to probe energy

transfer physics in thin films and nanostructured materials. The heart of the system is a 9 ps pulsewidth, 82 MHz repetition rate, mode-locked Nd:YVO<sub>4</sub> laser from Time Bandwidth Products (Switzerland) emitting at 1064 nm. The maximum pulse energy is 7 nJ, corresponding to an average output power of about 1 W. After exiting the laser cavity and an isolator (Faraday rotator), the output stream of pulses are split into a pump path (green) and a probe path (red) using a polarizing beam splitter (PBS). A half-wave plate before the PBS allows for control of the balance of power between the two paths.

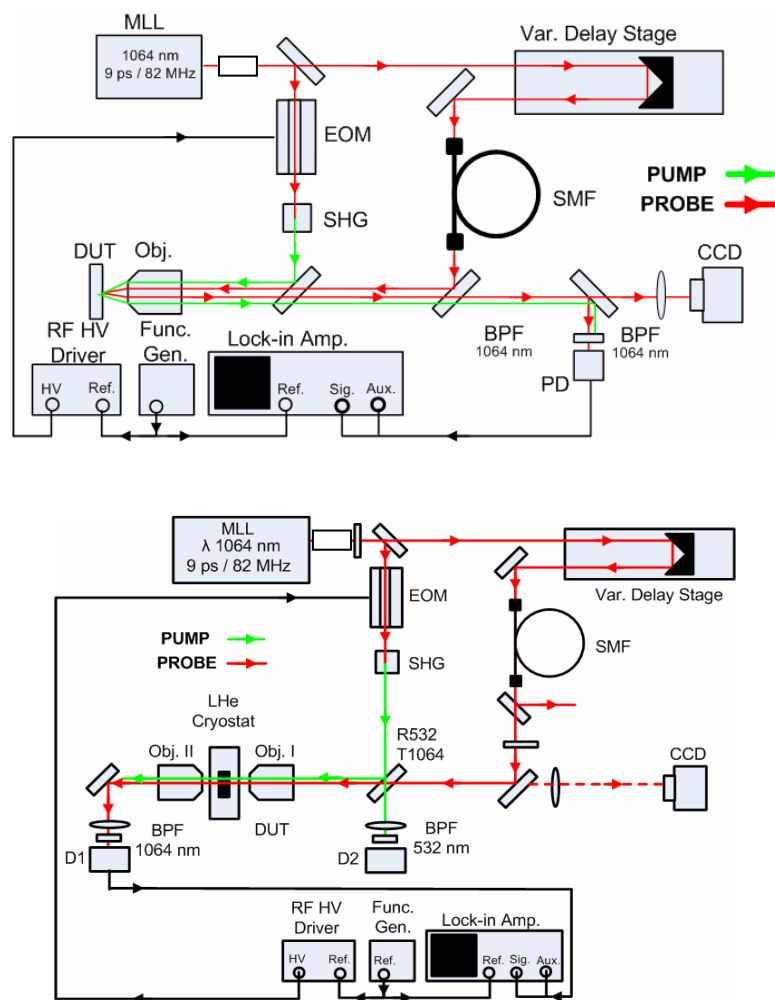


Figure 6.1: Experimental layouts for the two basic configurations of the pump-probe system. In the top panel, the system is configured in “reflection” mode which can be used for measurements such as transient thermal reflectance thermometry. In the bottom panel, the system is configured in “transmission mode.” In this configuration, the system can be used for transient differential absorption measurements.

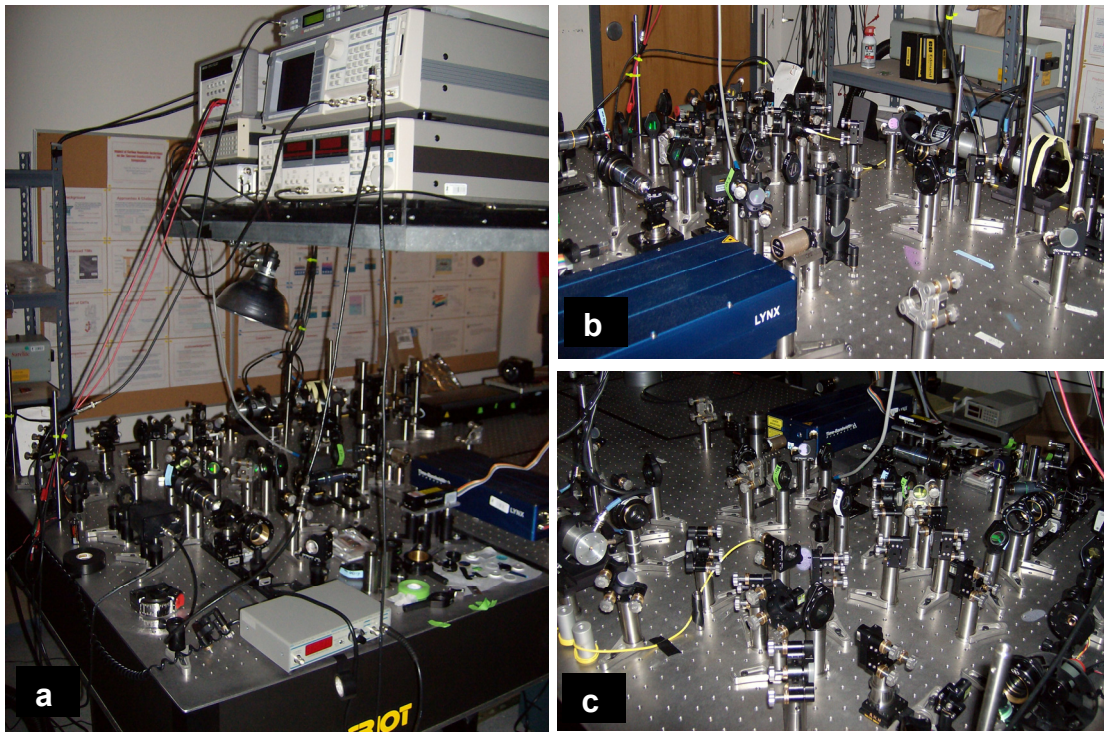


Figure. 6.2: Images of the picosecond pump-probe system developed by myself and Matt Panzer. The pump-probe system is used in two basic configurations: reflection and transmission mode. Reflection mode was used extensively for extracting thermal properties of nanostructured thin films using time-domain thermo-reflectance technique. The transmission mode configuration was used extensively for measuring carrier dynamics in silicon nanoparticle oxide films, the subject of Chapter 7.

After converting to 532 nm, the pump beam is imaged onto the sample to a spot size of the order of  $10\text{ }\mu\text{m}$  in diameter using an infinity-corrected, 0.41 N.A. long working distance objective lens (20x,  $f = 10\text{ mm}$ , Mitatoyo). The reflected pump beam is then imaged back along the primary optical axis and imaged onto a CCD camera where the beam properties of either the pump and probe can be precisely analyzed. The combination of the objective lens and the 500 mm tube lens form a basic 50x microscope which allows the operator to accurately determine the spatial properties of both the pump and probe beams on the sample.

The probe beam, which remains at 1064 nm, takes a path whose length can be precisely controlled using a variable optical delay stage (Newport) and a broadband retroreflector. The delay stage provides a total delay of about 4 ns (maximum roundtrip time). However, in practice it is very difficult to achieve accuracy of the

measurement over such a large time delay. This is due to several issues. First, the beam can be “steered” by several milliradian (mrad) by moving the stage position from one end to the other. This is the best that can be achieved despite the best efforts of alignment. When considering a 10 mm focal length objective lens, such as the 20x Mitatoyo, 1 mrad translates into a 10  $\mu\text{m}$  deviation on the sample, which is comparable to the focused spot size. This means the probe beam would walk completely off of the pump beam on the sample surface during the course of the measurement. Second, because the beam is divergent, the beam width before entering the objective lens will change substantially from one end of the stage to the other ( $\sim 3\times$ ). This will cause both the spot size and the focal position to change continuously throughout the measurement. Neither of these effects are easily corrected in the data analysis. Therefore, we adopted the following approach to eliminate these effects.

After traversing the delay stage, we pass the probe beam through a fiber “mode scrambler” (FMS) which consists of a 2 m long single mode fiber which is wrapped around two mandrels, clockwise around the first and counter-clockwise around the second. The beam profile at the output of the FMS is insensitive to the beam shape entering the fiber and therefore eliminates the effects of beam steering as well as beam divergence. However, there remains an output intensity dependence on stage position. This intensity dependence can be easily removed in the final signal by dividing the lock-in signal with the average power illuminating the detector. We do so by splitting the output of the photodetector into two signals, one signal going into the input channel of the lock-in and the other going to one of the auxiliary channels. The auxiliary channel simply digitizes the average value of the detected signal and the input channel detects the first harmonic of the modulated probe signal. The division is performed digitally. Care must be taken to assure that no appreciable DC offsets exist in either the primary signal or in the normalization signal. Photodetector saturation, particularly intensity saturation, can be particularly troublesome in performing this normalization scheme which we discuss at the end of this section. After emerging from the FMS, the probe beam is imaged onto the samples through the same 20x Mitatoyo objective. The reflected signal is reflected back along the principal optical

axis and is directed onto the photodetector. A 10 nm bandpass filter centered at 1064 nm is located in front of the PD in order to prevent the 532 nm pump beam from leaking into the detector. Before every measurement, pump leakage can be determined by simply blocking the probe beam, measuring the noise floor and then blocking the pump beam and measuring the noise floor. If there is no pump leakage, then the noise floor should be the same for both measurements.

## 6.2 Frequency Domain Analysis

The output of the MLL (before modulation occurs) is a pulse train with period  $T = 12.2$  ns (82 MHz repetition rate) with each pulse having a finite width of about 9 ps. The probe beam is a scaled (amplitude) and delayed version of the pump beam as illustrated in Figure 6.3.

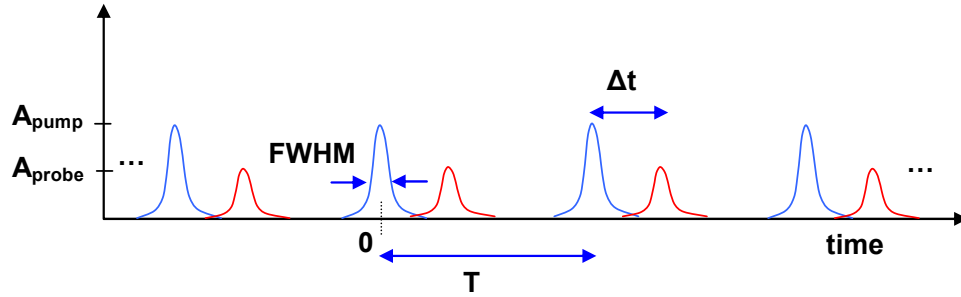


Figure 6.3: Real-time sequence of the unmodulated pulse train.

The laser pulse is split into two components,  $p_{\text{pump}}(t)$  and  $p_{\text{probe}}(t)$  which are given by

$$I_{\text{pump}}(t) = A_{\text{pump}} \exp(-a_{\text{pump}} t^2) * \sum_{m=-\infty}^{+\infty} \delta(t - mT) \quad (6.1)$$

$$I_{\text{probe}}(t) = A_{\text{probe}} \exp(-a_{\text{probe}} t^2) * \sum_{m=-\infty}^{+\infty} \delta(t - mT - \Delta t) \quad (6.2)$$

where the asterisk (\*) indicates the convolution integral,

$$f(t) * g(t) = g(t) * f(t) = \int_{-\infty}^{+\infty} f(t') g(t - t') dt' \quad (6.3).$$

$\Delta t$  is the time delay of the probe pulse relative to the pump pulse as determined by the optical delay line and the parameter  $a$  defines the temporal pulse width. The pulse width is commonly defined by the full-width at half maximum (FWHM) and is related to  $a$  by the expression

$$FWHM = 2\sqrt{\frac{\ln(2)}{a}} \quad (6.4)$$

or

$$a = \frac{4 \ln(2)}{FWHM^2} = \frac{2.773}{FWHM^2} \quad (6.5)$$

The Fourier Transform of (..) is the product of a Gaussian envelope function with characteristic width of  $\sqrt{a/4 \ln(2)}$  and a comb function with frequency spacing  $2\pi/T$

$$I_{pump}(\omega) = \sqrt{\frac{\pi}{a_{pump}}} \exp(-\omega^2/4a_{pump}) \left(\frac{2\pi}{T}\right) \sum_{m=-\infty}^{+\infty} \delta\left(\omega - m \frac{2\pi}{T}\right) \quad (6.6)$$

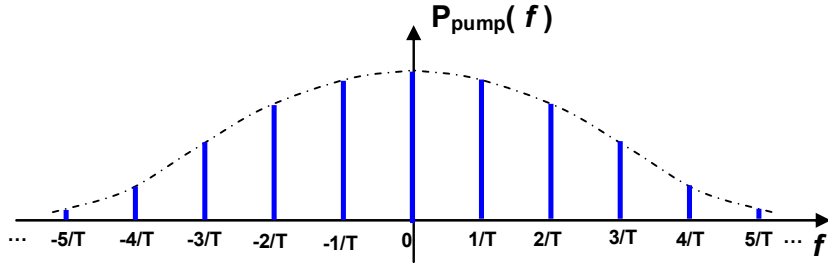


Figure 6.4: An idealization of the frequency spectrum of the unmodulated laser pulse train.

The pump beam is typically amplitude modulated by a rectangular transmission function defined by  $m(t)$ , given by

$$m(t) = \begin{cases} A_H ; & nT_{\text{mod}} \leq t \leq (n+\eta)T_{\text{mod}} \\ A_L ; & (n+\eta)T_{\text{mod}} \leq t \leq (n+1)T_{\text{mod}} \end{cases} \quad n = \text{set of all integer} \quad (6.7)$$

where  $\eta$  is the fraction of the cycle which is high (duty cycle =  $\eta \cdot 100$ ) and  $0 \leq A_L < A_H \leq 1$ . The resultant intensity function of the heating pump pulse is then given by the product

$$h(t) = I_{\text{pump}}(t) \cdot m(t) \quad (6.8)$$

which is as illustrated in Figure 6.5.

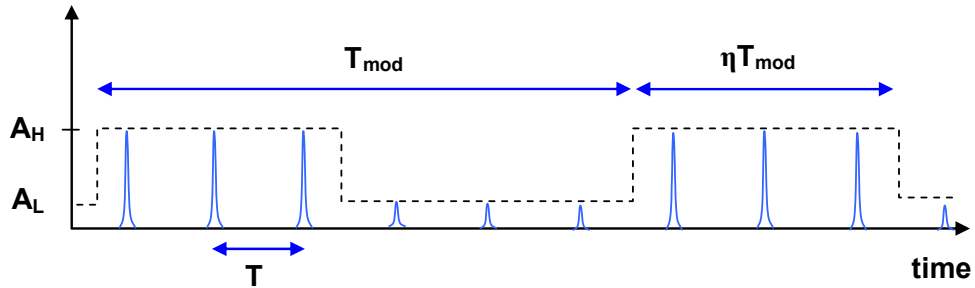


Figure. 6.5: Pump modulation signal approximated as a rectangular wave,  $0 \leq A_L < A_H \leq 1$

The Fourier transform (F.T.) of (6.8) is found by the convolution

$$\hat{I}_{pump}(\omega) = I_{pump}(\omega) * M(\omega) \quad (6.9)$$

where  $I_{pump}(\omega)$  was given by (6.6) and  $M(\omega)$  can be shown to be

$$M(\omega) = \sum_{n=-\infty}^{+\infty} \left[ (A_H - A_L) \eta T_{mod} \frac{\sin(\omega \eta T_{mod} / 2)}{\omega \eta T_{mod} / 2} \right] \left( \frac{2\pi}{T_{mod}} \right) \delta \left( \omega - n \frac{2\pi}{T_{mod}} \right) + A_L \delta(\omega) \quad (6.10)$$

Proceeding with the convolution, we obtain the final result:

$$\begin{aligned} \hat{I}_{pump}(\omega) = \sum_{m,n=-\infty}^{+\infty} & \left\{ \sqrt{\frac{\pi}{a_{pump}}} \exp\left(- (m2\pi/T)^2 / 4a_{pump}\right) \left( \frac{2\pi}{T} \right) \right. \\ & \times (A_H - A_L) \eta T_{mod} \frac{\sin((\omega - m2\pi/T) \eta T_{mod} / 2)}{(\omega - m2\pi/T) \eta T_{mod} / 2} \\ & \times \left( \frac{2\pi}{T_{mod}} \right) \delta \left( \omega - m \frac{2\pi}{T} - n \frac{2\pi}{T_{mod}} \right) \\ & \left. + A_L \sqrt{\frac{\pi}{a_{pump}}} \exp\left(- \omega^2 / 4a_{pump}\right) \left( \frac{2\pi}{T} \right) \sum_{m=-\infty}^{+\infty} \delta \left( \omega - m \frac{2\pi}{T} \right) \right\} \end{aligned}$$

It can easily be shown that the maximum signal will occur when  $\eta = 50\%$ .

With the exception of a scalar constant, the amplitude spectrum of the probe beam will have the same spectrum as the unmodulated pump beam (Fig. 6.4). In addition to a real scaling factor, each spectral component will have a phase angle which depends on the relative delay to the pump pulse train,  $\exp(if_n \Delta t)$  where  $n$  is the  $n$ th harmonic of the laser repetition frequency and  $\Delta t$  is the relative delay between the pump and probe signals. If the pump and probe beams are focused onto a sample that will respond instantaneously to the pump signal, then the temporal properties of the beam can be measured indirectly by examining the interaction on the sample. By the width property

of the convolution integral, the width of the response function will be the sum of the widths of the pump and probe beams.

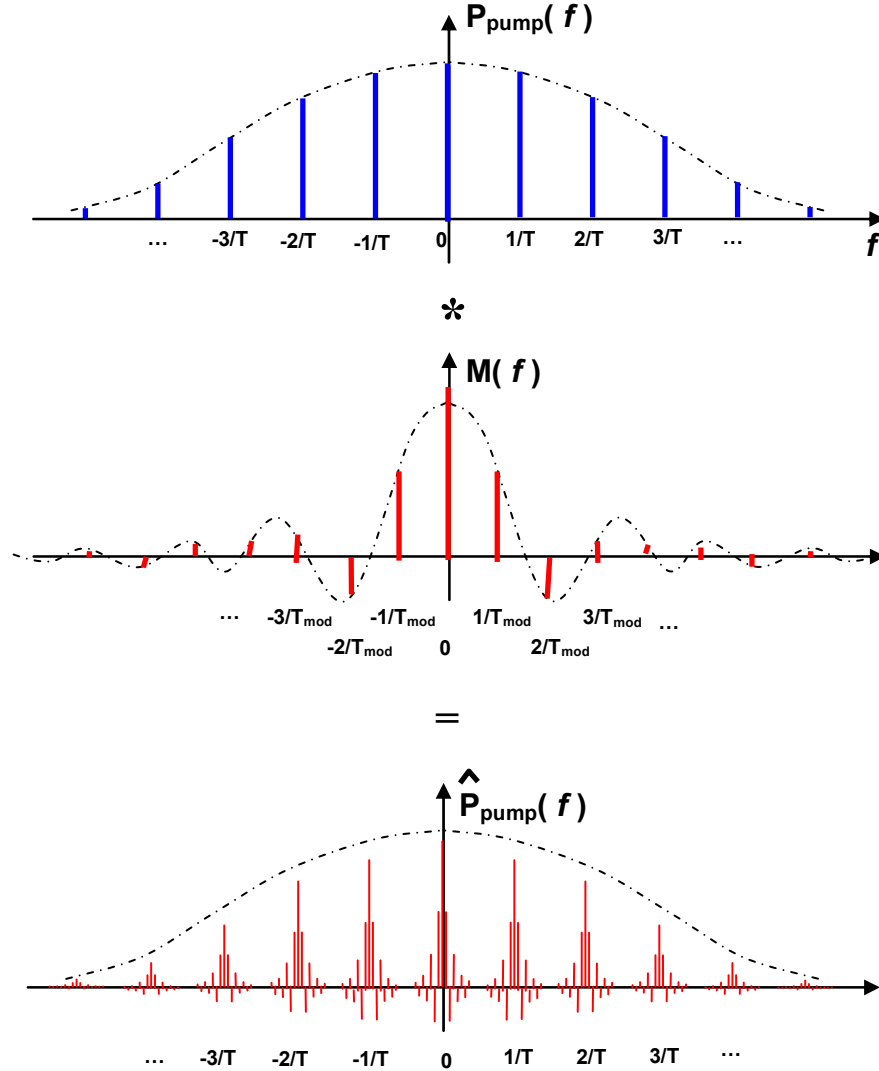


Fig. 6.6: Visualization of the spectrum of the square-wave amplitude modulated pump pulse train when the pulse repetition period is much smaller than the modulation period, i.e.  $T \ll T_{\text{mod}}$ .

Figure 6.7 shows the results of such a measurement performed on a 500 nm thick film of silicon nanocrystal doped silicon oxide grown by Rohan Kekatpure of the Brongersma Lab (Stanford University). In the absence of the pump beam (532 nm), there is negligible absorption in the sample by the probe beam. However, the pump beam is able to promote ground state electrons to a series of upper excited states which are then capable of absorbing the probe light. This upward transition is estimated to be

much less than 1 ps and can thus be taken to be instantaneous. The relaxation time however takes much longer, on the order of 100 ps for the initial decay and a much longer (10s of ns) beyond that due to multi-level recombination processes which will be the subject of the Chapter 7. Because the initial response is effectively instantaneous, we can use this to measure the intrinsic temporal width of the pump and probe beams. The response time is approximately 20 ps and corresponds to 2x the expected temporal pulse width of the source in exact agreement with theory.

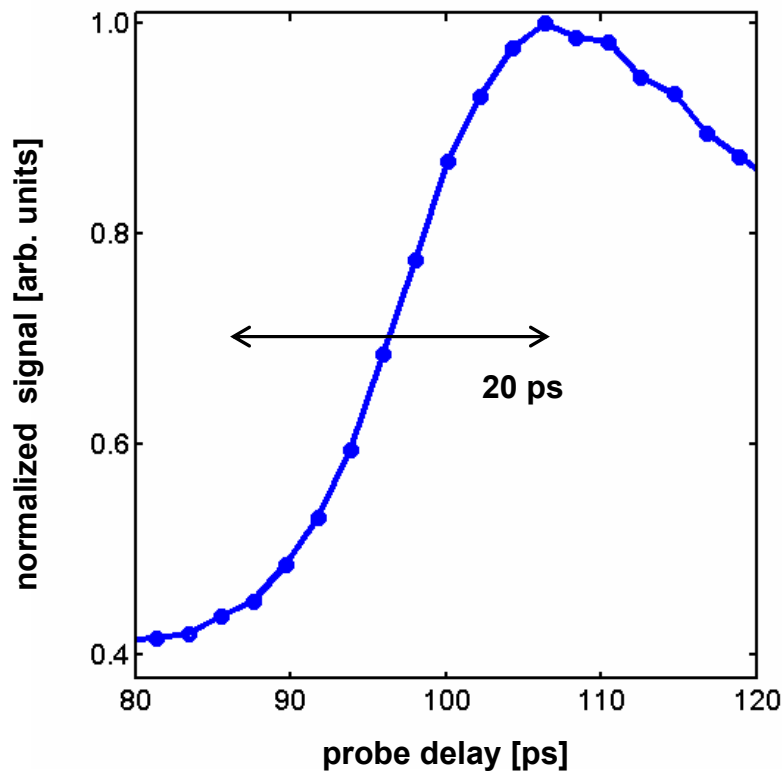


Fig. 6.7: The response of the probe beam due to a fast excitation in a thin film caused by the pump beam. The initial change of the excitation is assumed to be sub-picosecond and therefore the rising edge of the response is a good measure of the convolution of the pump and probe intrinsic temporal widths. From this data, we see that the characteristic width is roughly 20 ps which corresponds to the 2x the pulse width of the source. The sample used to make this measurement was a thin silicon rich silicon oxide film containing 5 nm silicon nanocrystals with a density of approximately  $2 \times 10^{18} \text{ cm}^{-3}$ .

### 6.3 Time Domain Thermo-Reflectance Thermometry

The pump-probe system described can be used to extract thermal properties of ultra-thin films and nanostructured materials using a time domain thermal reflectance (TDTR) technique [80]. In addition to being a powerful tool for probing thermal conduction physics of thin films and nanostructured materials, the TDTR technique is an excellent approach to calibrating the pump-probe system due to the linearity of the transient system response at low excitation levels to materials easily deposited on a substrate, e.g. aluminum on thermally grown silicon oxide. Prior to performing any critical set of measurements, it became common lab practice to measure what became our calibration standard: 30 nm Al on 102 nm thermal SiO<sub>2</sub> on a Si substrate. At the time scales of measurement, the thermal oxide is thermally infinite and thus thermal interface effects in the SiO<sub>2</sub>/Si boundary can be neglected. Because this technique was used heavily for researching the thermal properties of novel samples as well as a calibration technique, I will spend time here discussing the details of this measurement.

The measurement concept is simple. A thin (~30 nm) metal transducer layer is deposited on top of a material or stack of materials whose thermal properties are of interest. A small fraction of a pump pulse directed at the sample is absorbed within a skin depth of the transducer layer. The resulting heat generation induces a temperature field which evolves over time as the heat diffuses through the underlying layers and does so in a manner which depends uniquely on the thermal properties of the underlying material layers. A delayed attenuated probe beam is directed at the same spot sampling the spatially distributed surface temperature rise through the thermo-reflectivity effect. For small changes in temperature, (~ 1 K), the reflectivity of the metal surface will change in proportion to the change in surface temperature and is characterized small-signal thermo-reflectance coefficient, denoted by  $dR/dT$ . Typical magnitudes of  $dR/dT$  are in the range of  $10^{-4} - 10^{-3} \text{ K}^{-1}$ . The temperature diffuses vertically ( $z$ -direction, out-of-plane) as well as laterally ( $r$ , in-plane). The vertical diffusion will dominate if the pump beam spot size is substantially larger than the

effective diffusion length of the material stack for the time scale of the measurement.

If the heating pump pulse is axially symmetric, the temperature profile at any given time  $t$  will also be axially symmetric, depending only on the radial position  $r$ . Analyzing the region of the surface of the metal transducer layer described by an annulus of radius  $r$  and infinitesimal thickness  $dr$ , we find that the total reflected probe power for this region is given simply by the product of the probe power striking this region, the temperature rise at this radius, and the thermo-reflectance coefficient, the total incident probe power of course being the product of the intensity at radius  $r$  and the differential area of the region.

$$\Delta P_{probe}(r,t) = \left( \frac{dR}{dT} \right) 2\pi \Delta T(r,t) I_{probe}(r,t) r dr \quad (6.11)$$

where  $I_{probe}(r,t)$  is the radial distribution of probe intensity.

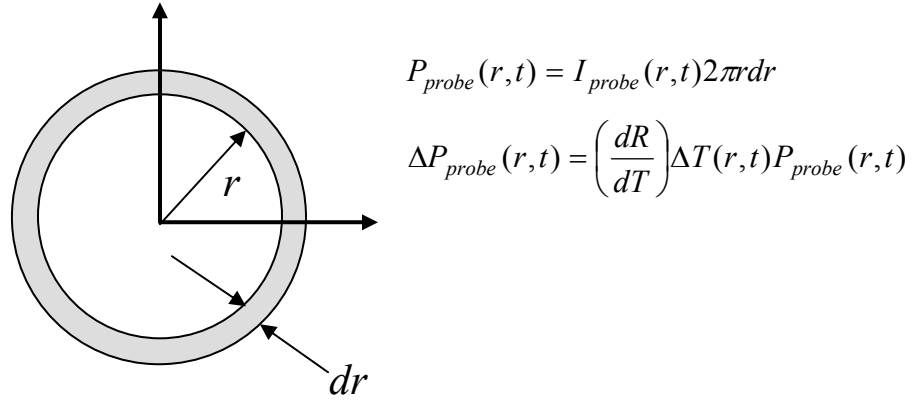


Figure 6.8: Schematic of the heating geometry of the sample looking down on the sample along the beam path.

The total change in reflected probe power is then found by integrating over all space:

$$\Delta P_{probe}(t) = 2\pi \int_0^{\infty} \Delta P_{probe}(r,t) r dr \quad (6.12)$$

By modulating the pump beam as described in previous sections, high-SNR lock-in techniques can be employed. Thus, we are ultimately interested in extracting  $\Delta P_{probe}(\omega_{mod}, \Delta t)$ , the frequency domain modulated probe signal collected at the fundamental modulation frequency as a function of probe delay time. Taking the Fourier transform (FT) of (6.12), we have

$$\Delta P_{probe}(\omega, \Delta t) = \int_{-\infty}^{+\infty} \left\{ 2\pi \int_0^{\infty} \left( \frac{dR}{dT} \right) \Delta T(r, \omega') I_{probe}(r, \omega - \omega', \Delta t) r dr \right\} d\omega' \quad (6.13)$$

where we used the well-known FT relation

$$f(r, t)g(r, t) \Rightarrow F(r, \omega) * G(r, \omega) \quad (6.14)$$

It remains then to determine the functions  $\Delta T(r, \omega)$  and  $I_{probe}(r, \omega, \Delta t)$  and is the subject of the following sub-sections. After measuring the temperature response indirectly through  $\Delta P_{probe}(\omega, \Delta t)$ , we rigorously solve the heat diffusion equation in space (radial) and time coordinates (frequency space) and vary the unknown material parameters such as thermal conductivity and interface boundary conductances in order to fit the data. In the following sections, we describe the solution to the heat diffusion equation which we use to obtain the surface temperature distribution as a function of radial position and time delay as measured on the lock-in amplifier. Before doing so, we discuss a few key characteristics of this technique which complicate the analysis.

Because the temperature decay due to a single pump pulse occurs over a long timescale ( $\mu s$ ) compared to the repetition period of the laser ( $\sim 10$  ns), the instantaneous temperature rise is significantly impacted by the temperature response of previous pump pulse. Therefore, the signal at any time must be represented as a convolution of the temperature responses induced by all previous pump pulses. Furthermore, since the pump beam is modulated, the heating response will have

Fourier components at integral multiples of the modulation frequency and therefore the probe signal will also have Fourier components at these frequencies.

### 6.3.1 General Solution of the Temperature Field

We begin by assuming that for each layer within the material stack, the thermal properties are homogeneous with the exception that the in-plane ( $r$ ) and out-of-plane ( $z$ ) thermal conductivities may be different.

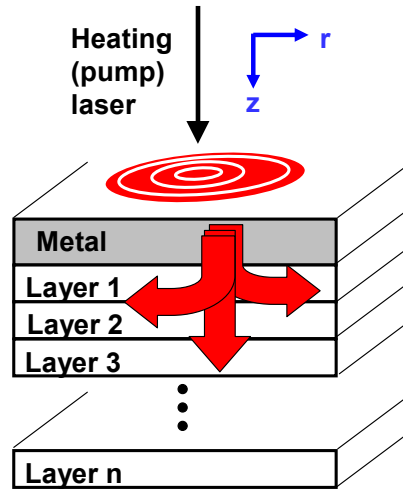


Figure 6.9: Schematic representation of a material stack to be modeled. The  $z$ -component of the temperature field is treated using transmission matrix approach, a well-known technique routinely used in electromagnetics and quantum mechanics but which is less known within the thermal physics community.

For each of these layers, we take the form of the heat diffusion equation to be axially symmetric, described by the general form

$$k_r \left( \frac{\partial^2 T(r, z, t)}{\partial r^2} + \frac{1}{r} \frac{\partial T(r, z, t)}{\partial r} \right) + k_z \frac{\partial^2 T(r, z, t)}{\partial z^2} = c_v \frac{\partial T(r, z, t)}{\partial t} \quad (6.15a)$$

$$k_{r,z} \left( \frac{\partial^2 T(r,z,t)}{\partial r^2} + \frac{1}{r} \frac{\partial T(r,z,t)}{\partial r} \right) + \frac{\partial^2 T(r,z,t)}{\partial z^2} = \frac{1}{\alpha} \frac{\partial T(r,z,t)}{\partial t} \quad (6.15b)$$

where  $k_{r,z} \equiv k_r / k_z$  the ratio of the in-plane to out-of plane thermal conductivity and  $\alpha = k_z / c_v$  is the out-of-plane thermal diffusivity. The solution of the temperature profile in each layer is separable in the two spatial coordinates  $r$  and  $z$ . Furthermore, we transform the equation into the frequency domain to obtain the solution

$$T(r,z,t) = \int_{-\infty}^{+\infty} T(r,z,\omega) e^{i\omega t} d\omega \quad (6.16)$$

where  $T(r,z,\omega) = R(r,\omega)Z(z,\omega)$ . We do so because ultimately, the temperature response is measured in the frequency domain by the lock-in amplifier. Through direct substitution, (6.15) is transformed into

$$\frac{1}{R(r,\omega)} \left( \frac{\partial^2 R(r,\omega)}{\partial r^2} + \frac{1}{r} \frac{\partial R(r,\omega)}{\partial r} \right) = \frac{\frac{i\omega}{\alpha} - \frac{1}{Z(z,\omega)} \frac{\partial^2 Z(z,\omega)}{\partial z^2}}{k_{r,z}} = -k^2 \quad (6.17)$$

where  $-k^2$  is the separation constant for a particular value of  $\omega$ . The radial (in-plane) equation is the well-known Bessel's equation of zero order

$$\frac{\partial^2 R(r,\omega)}{\partial r^2} + \frac{1}{r} \frac{\partial R(r,\omega)}{\partial r} + k^2 R(r,\omega) = 0 \quad (6.18)$$

which has the solution  $R(r) = J_0(-kr) = J_0(kr)$ . The  $z$ -equation (out-of-plane) is given by

$$\frac{\partial^2 Z(z, \omega)}{\partial z^2} = \gamma^2 Z(z, \omega) \quad (6.19)$$

where

$$\gamma^2 \equiv \left( \frac{i\omega}{\alpha} + k_{r,z} k^2 \right) \quad (6.20)$$

and has a solution of the form

$$Z(z, \omega) = A_z e^{\pm \gamma z} \quad (6.21)$$

We then have for layer  $n$  a frequency-domain solution of the form

$$T_n(r, z, \omega) e^{i\omega t} = A_{n\pm}(\omega) J_0(kr) e^{\pm \gamma z + i\omega t} \quad (6.22)$$

The task then becomes determining the radial temperature distribution at the top metallic surface. We discuss how this is done in the following section.

### 6.3.2 Temperature Field at Surface

The frequency spectrum of the temperature response  $\Delta T(r, \omega)$  due to an axially symmetric heating source  $Q'''(r, \omega)$  can be found by convolving the source function with the system thermal response to a time-harmonic radial point source function  $g(r, \omega)$ . In rectangular coordinates, this is written as

$$\Delta T(x, y, \omega) = \iint Q'''(x', y', \omega) g(x - x', y - y', \omega) dx' dy' \quad (6.23)$$

Using once again the FT relation (6.14), we obtain

$$\Delta T(k_x, k_y, \omega) = Q'''(k_x, k_y, \omega)g(k_x, k_y, \omega) \quad (6.24)$$

or, in radial coordinates

$$\Delta T(k, \omega) = Q'''(k, \omega)g(k, \omega) \quad (6.25)$$

where the real-space variables are transformed into frequency-space variables as  $x, y \rightarrow k_{x,y}$  and  $r \rightarrow k$ . The heat generation term is proportional to the modulated pump power  $\hat{P}_{pump}(\omega)$  at the surface, and can be written as

$$Q'''(r, \omega) = \frac{2A}{\pi w_0^2} \exp\left(-2 \frac{r^2}{w_0^2}\right) Q'''(\omega) \quad (6.26)$$

Here,  $2w_0$  is the beam waist at the sample and  $A$  is the average pump power absorbed on the surface of the metal. The actual value is not important since the final signal is normalized. Performing a Hankel transform ( $r \rightarrow k$ ) on

$$g(k, \omega) = 2\pi \int_0^{+\infty} g(r, \omega) J_0(kr) r dr \quad (6.27)$$

$$Q'''(k) = 2\pi \int_0^{+\infty} Q'''(r) J_0(kr) r dr \quad (6.28)$$

$$= 2\pi \int_0^{+\infty} \frac{2A}{\pi w_0^2} \exp\left(-2 \frac{r^2}{w_0^2}\right) J_0(kr) r dr \quad (6.29)$$

$$= A \exp\left(-\frac{\pi^2 k^2 w_0^2}{2}\right) \quad (6.30)$$

where  $J_0$  is the 0<sup>th</sup> order Bessel equation. We then have

$$\begin{aligned} \Delta T(k, \omega) &= g(k, \omega) \cdot Q'''(k, \omega) \\ &= g(k, \omega) \cdot A \exp\left(-\frac{\pi^2 k^2 w_0^2}{2}\right) Q'''(\omega) \end{aligned} \quad (6.31)$$

$$\Delta T(r, \omega) = 2\pi \int_0^{+\infty} g(k, \omega) \cdot Q'''(k, \omega) J_0(kr) k dk \quad (6.32)$$

$g(k, \omega)$  can be found using a transfer matrix approach in combination with separation of variables technique discussed in the previous sub-section. We refer the reader to Feldman [81] or References [82, 83] for more detailed derivation of the transfer matrices and only summarize the results here.

$$g(k, \omega) = \frac{1}{\gamma_1} \left( \frac{B_1^+ + B_1^-}{B_1^- - B_1^+} \right) \quad (6.33)$$

where

$$\begin{pmatrix} B_1^+ \\ B_1^- \end{pmatrix} = T_{1,n} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and, for each layer  $n$ , the material-dependent parameters

$$\gamma_n = k_{z,n} u_n \quad (6.34)$$

$$u_n = \left(4\pi k^2 + q_n^2\right)^{1/2} \quad (6.35)$$

$$q_n^2 = \frac{i\omega}{\alpha_n} \quad (6.36)$$

Here,  $T_{1,n}$  is the transfer matrix along the z-direction for the layered medium relating the temperature in the final layer (typically treated as an infinite medium) and top surface, i.e.  $n = 1$ . In going from layer  $n + 1$  to layer  $n$  (begin from layer furthest from metal transducer), we have the following relations between the “forward” (+) and “backward” (-) propagating complex thermal fields:

$$\begin{pmatrix} B_n^+ \\ B_n^- \end{pmatrix} = T_{n,n+1} \begin{pmatrix} B_{n+1}^+ \\ B_{n+1}^- \end{pmatrix} \quad (6.37)$$

where

$$T_{n,n+1} = T_i(n \rightarrow n+1) \cdot T_n \quad (6.38)$$

and  $T_i(n \rightarrow n+1)$  and  $T_n$  describes the transfer across interface between  $n$  and  $n + 1$  and across the region  $n$  respectively.

$$T_i(n \rightarrow n+1) = \frac{1}{2\gamma_n} \begin{pmatrix} \gamma_n + \gamma_{n+1} + \gamma_n \gamma_{n+1} R_{n,n+1} & \gamma_n - \gamma_{n+1} - \gamma_n \gamma_{n+1} R_{n,n+1} \\ \gamma_n - \gamma_{n+1} + \gamma_n \gamma_{n+1} R_{n,n+1} & \gamma_n + \gamma_{n+1} - \gamma_n \gamma_{n+1} R_{n,n+1} \end{pmatrix} \quad (6.39)$$

$$T_n = \begin{pmatrix} \exp(-u_n \cdot L_n) & 0 \\ 0 & \exp(u_n \cdot L_n) \end{pmatrix} \quad (6.40)$$

In the above transfer matrix expressions,  $R_{n,n+1}$  is the boundary resistance (in units of  $\text{K}\cdot\text{m}^2/\text{W}$ ) between layers  $n$  and  $n + 1$  and  $L_n$  is the physical thickness for layer  $n$ .

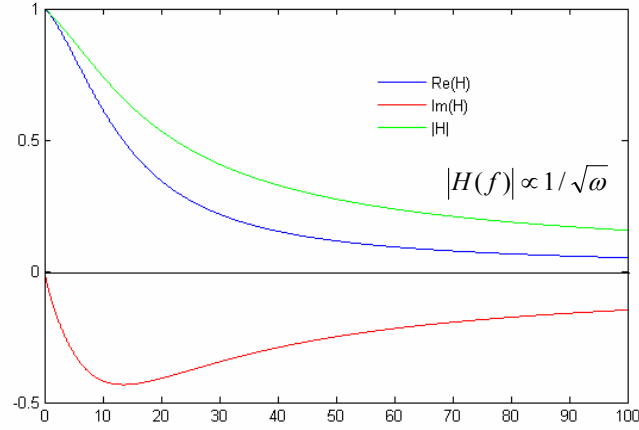


Figure. 6.10: A typical amplitude spectrum (green) of  $H(f)$  for the positive frequency domain from 0 to 100 MHz. The very low frequency limit is independent of frequency and in the high-frequency limit,  $|H(f)| \propto 1/\sqrt{\omega}$ . The amplitude spectrum and real part of  $H$  is even with respect to the origin whereas the phase and imaginary part are odd.

### 6.3.3 Frequency Domain Thermo-reflectance Signal

The full solution for  $\Delta P_{probe}(\omega, \Delta t)$  can be shown to be given by (6.41)

$$\Delta P_{probe}(\omega, \Delta t) = A_{pump} A_{probe} \left( \frac{dR}{dT} \right) H(\omega) Q'''(\omega) * P_{probe}(\omega) \quad (6.41)$$

where

$$H(\omega)Q'''(\omega) * P_{probe}(\omega) =$$

$$\begin{aligned} & \frac{4\pi^3(A_H - A_L)}{T\sqrt{a_{pump}a_{probe}}} \left( \eta \frac{\sin(\pi\eta)}{\pi\eta} \right) \\ & \times \sum_{p=-\infty}^{+\infty} \left\{ H\left( \pm \frac{2\pi}{T_{mod}} - p \frac{2\pi}{T} \right) \right. \\ & \times \exp\left( -\left( p \frac{2\pi}{T} \right)^2 \frac{1}{4} \left( \frac{1}{a_{pump}} + \frac{1}{a_{probe}} \right) \right) \\ & \times \exp\left( i \left( p \frac{2\pi}{T} \right) \Delta t \right) \left. \right\} \end{aligned} \quad (6.42)$$

and

$$H(\omega) \equiv 2\pi \int_0^{\infty} g(k, \omega') \cdot \exp\left( -\frac{\pi^2 k^2 (w_0^2 + w_1^2)}{2} \right) k dk \quad (6.43)$$

The  $\pm$  indicates an additional summation over the positive and negative components of the fundamental modulation frequency. If the pulse width becomes infinitesimal, i.e. the laser is modeled as a comb function,

$$H(\omega)Q'''(\omega) * P_{probe}(\omega) =$$

$$\begin{aligned} & \frac{4\pi^3(A_H - A_L)}{T\sqrt{a_{pump}a_{probe}}} \left( \eta \frac{\sin(\pi\eta)}{\pi\eta} \right) \\ & \times \sum_{p=-\infty}^{+\infty} \left\{ H\left( \pm \frac{2\pi}{T_{mod}} - p \frac{2\pi}{T} \right) \times \exp\left( i \left( p \frac{2\pi}{T} \right) \Delta t \right) \right\} \end{aligned} \quad (6.44)$$

and we recover the expression obtained by Cahill [82]. The convolution integral just evaluated analytically, is represented in Figure 6.11

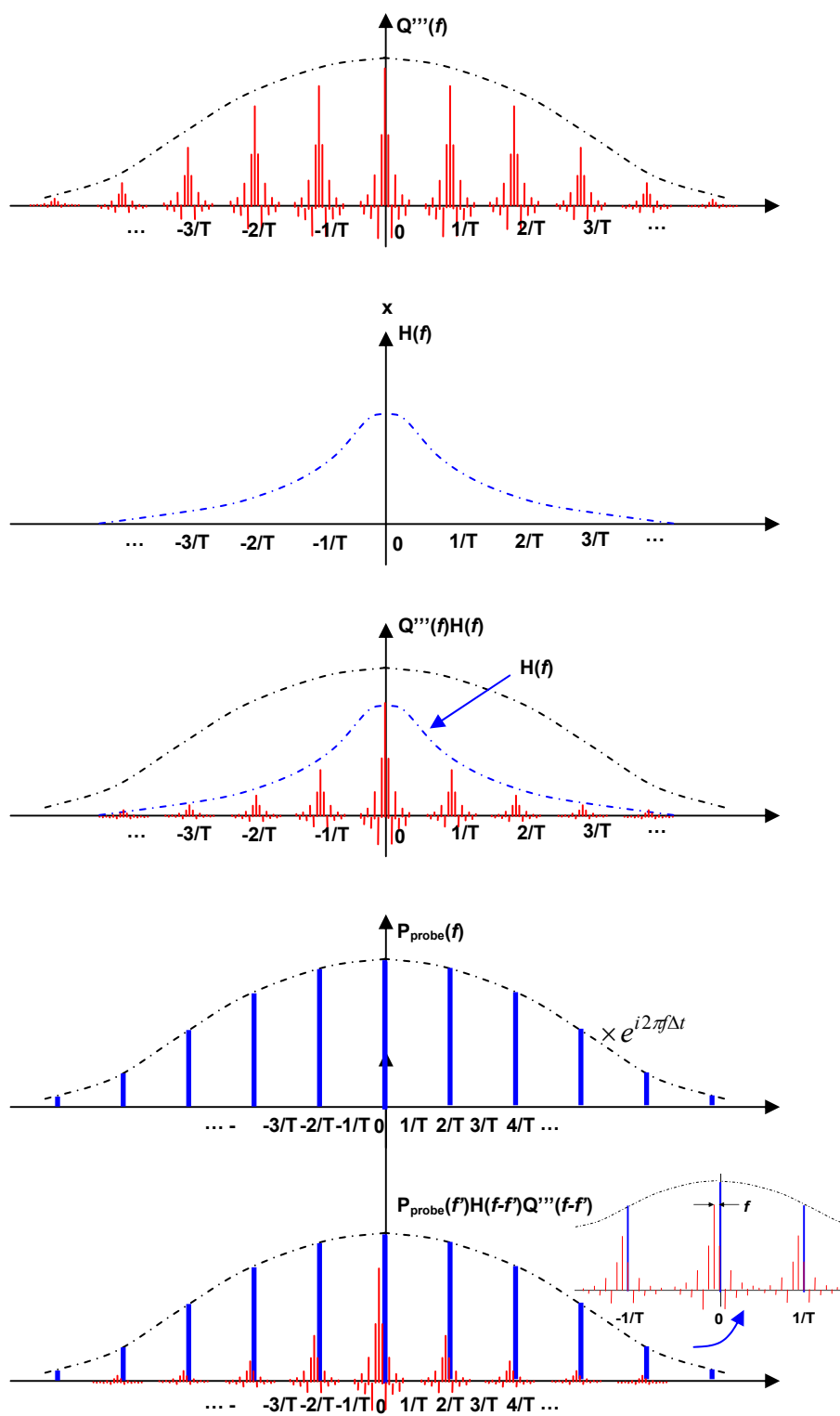


Figure 6.11: Graphical representation of the convolution which represents the measured signal on the lock-in amplifier for each probe delay.

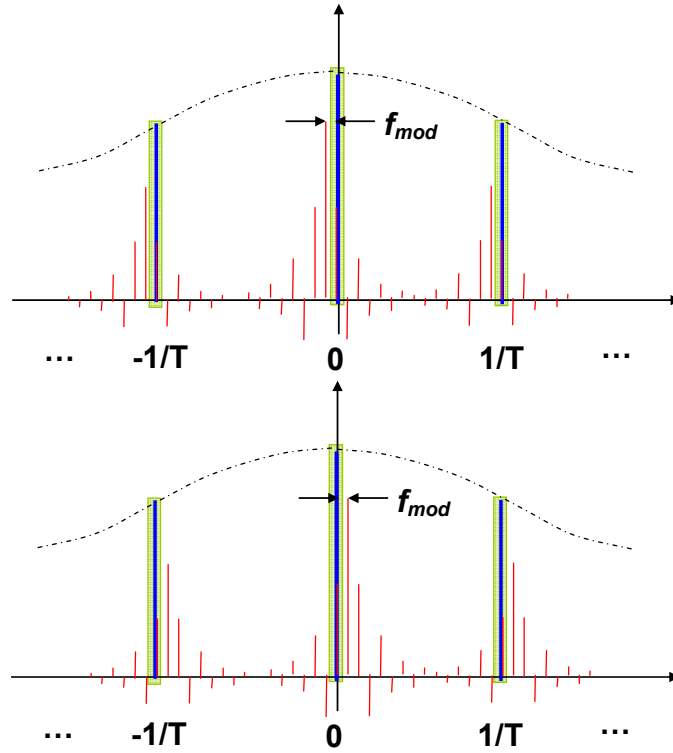


Figure 6.12: Graphical representation of the convolution represents the measured signal on the lock-in amplifier for each probe delay zoomed into spectral region of interest.

#### 6.3.4 Calibration Standard (Al/SiO<sub>2</sub>/Si)

For calibration, we use a sample which has a 102 nm thermal silicon oxide grown on a 740  $\mu\text{m}$  silicon substrate and a 30 nm aluminum layer deposited on the silicon oxide. A typical trace using the system described in Section II is shown in Figure 10 along with an analytic fit using a code which implements the theory described in Section III. The analytic fit was performed using a MATLAB code written to implement the theory described in the previous section. We extract a thermal conductivity of 1.4 W/m/K for the thermal oxide which is the industry established value.

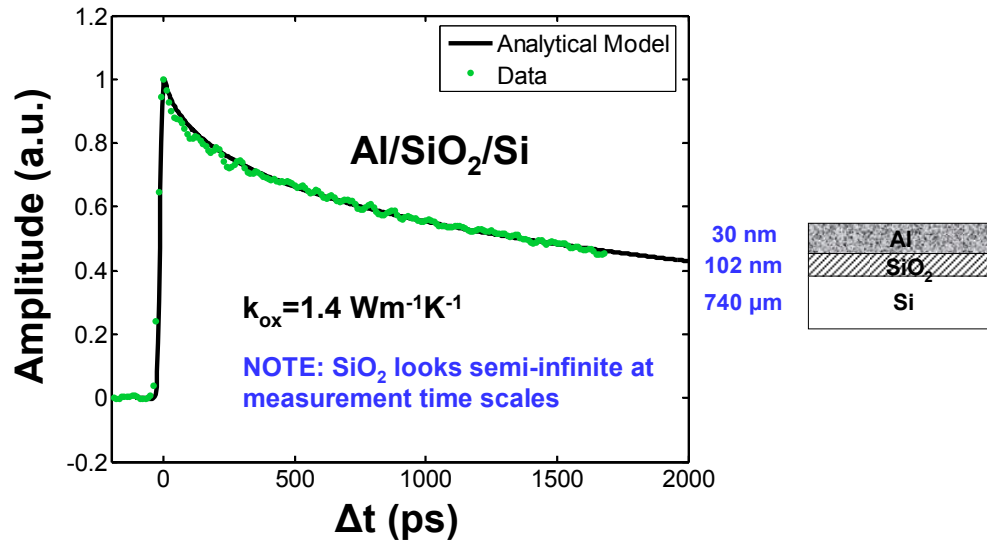


Figure 6. 13: Measured thermal response for an Al/SiO<sub>2</sub>/Si material stack demonstrating the accuracy of the pump-probe thermo-reflectance technique. The negative time delay signal has been subtracted from the signal during analysis.

## 6.4 Transient Differential Transmission Measurements

In this section, we will discuss the differential transmission technique developed to probe short-time scale carrier dynamics in dense silicon nanocrystal composite materials, the subject of Chapter 7. Figure 6.14 show the differential transmission configuration in combination with a liquid helium cryostat enabling measurements from about 10 K up to 400 K.

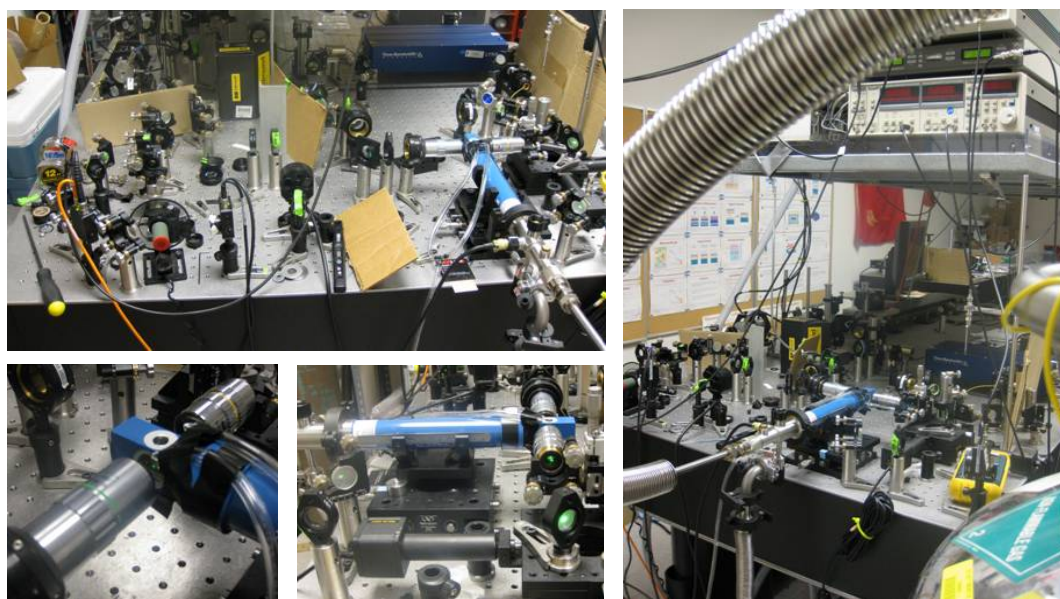


Figure 6.14: Pictures of the pump-probe system configured for transient differential transmission measurements. The liquid helium cryostat with integrated heater enabled measurements from about 5 K up to 400 K.

The pump (532 nm) and probe (1064 nm) beams are aligned to be collinear using a the combination of a mirror placed at the sample focal plane and an imaging camera as shown in Figure 6.1. The pump and probe beams are both focused down onto the sample and collected using long working distance Mitatoyo NIR objectives. The input objective was a 10 mm WD 20x 0.42 NA infinity corrected lens and the collecting objective was a 20 mm WD 10x 0.25 NA infinity corrected lens. The collecting optic is able to efficiently collect the transmitted beam because the full aperture of the focusing objective is not used. The beam diameters of the pump and probe beams were typically 5 and 8  $\mu\text{m}$  respectively at the sample surface, substantially larger than the diffraction limit for the lens due to the underfilling of the entrance pupil, and corresponded to an effective NA of less than 0.1. A narrow bandpass filter is placed between the output of the collection objective and the single element photodetector (ThorLabs – PDA10CS InGaAs detector). The output of the probe beam is fed directly into the input channel of a low-frequency lock-in amplifier (SRS SR850) whose

reference signal is the function generator which sets the pump square wave modulation frequency.

The signal measured on the lock-in is the differential transmission signal of the probe beam passing through the sample due to the influence of the pump excitation of the sample. We are assured that the change in transmission is due to the pump because the lock-in extracts only those signal components with frequency at the fundamental pump modulation frequency. It is easy to verify that the signal is due to sample excitation by blocking the pump to assure that the signal drops to the noise floor. Because of the large separation in laser frequency, the off-the-shelf 1064 nm bandpass filter is more than adequate to suppress pump leakage to below the noise floor of the detector.

In studying excited carrier dynamics, the pump generates excited electrons which occupy conduction band states. Over time, these electrons relax back to the ground state in a unique manner which can be used to study the physical recombination processes. Neglecting two-photon absorption processes, the probe photon energy is insufficient to excite electrons from the valence band to the conduction band. However, the probe photon energy is sufficient to excite an electron already in the conduction band due to the initial excitement by the pump beam which passes through the same interaction volume of the sample. Therefore, the probe beam will be attenuated slightly in the presence of the pump-excited carriers and will produce a modulation signal at the detector related to the pump modulation spectrum. If the system is perfectly linear or only weakly nonlinear, then the probe modulation spectrum will be identical, with the exception of a scaling factor. The probe beam is then delayed in time with respect to the pump beam using the delay line and at each equivalent position of the delay line, the rms signal at the fundamental pump modulation frequency is collected to produce the transient full transient signal over several nanoseconds with sub 10 ps temporal resolution. In the following section, I will discuss some of the issues that arise when trying to build such an experimental system.

## 6.5 Application Notes

Building a pump-probe system which relies on an optical delay technique appears straightforward and in retrospect is not terribly difficult to construct. However, there are a few subtle aspects of the system which need to be communicated to the next generation of students, which unfortunately does not appear in standard text books on optics and photonics. In this section, I will spell out some of the more important lessons learned during the development of the system. In addition to these application notes, I refer the reader to the excellent text by P. C. D. Hobbs [84]. A copy of Hobbs' text should be on the shelf of any practicing or aspiring experimentalist.

### 6.5.1 Beam Divergence and Beam Steering

In the pump-probe system described here, the probe pulse is delayed in time with respect to the pump pulse using a 0.6 m optical delay line (Newport) giving a maximum time delay of 4 ns. One of the challenges of making delay-line measurements over long time scales (nanoseconds) is that the roundtrip distance is sufficient to cause appreciable beam expansion due to standard beam divergence. In addition to beam divergence, the delay line can introduce a small but appreciable position dependent angular deflection. Both of these effects introduce an amplitude signal which can not easily be separated from the time-dependent sample response that is of interest. In practice, these effects can be subtracted from the true signal by first converting these two effects into a pure amplitude signal. We achieve this by running the probe beam through a single mode fiber (SMF) stage after traversing the delay line. Therefore, the beam divergence and steering only affects the transmission signal but the beam quality on the output of the fiber is independent of the optical delay line position. During data analysis, this calibration curve can then be used to normalize the signal and therefore extract the true signal. Figure 6.15 shows a typical probe signal amplitude on the detector as a function of probe delay time using the SMF spatial filter.

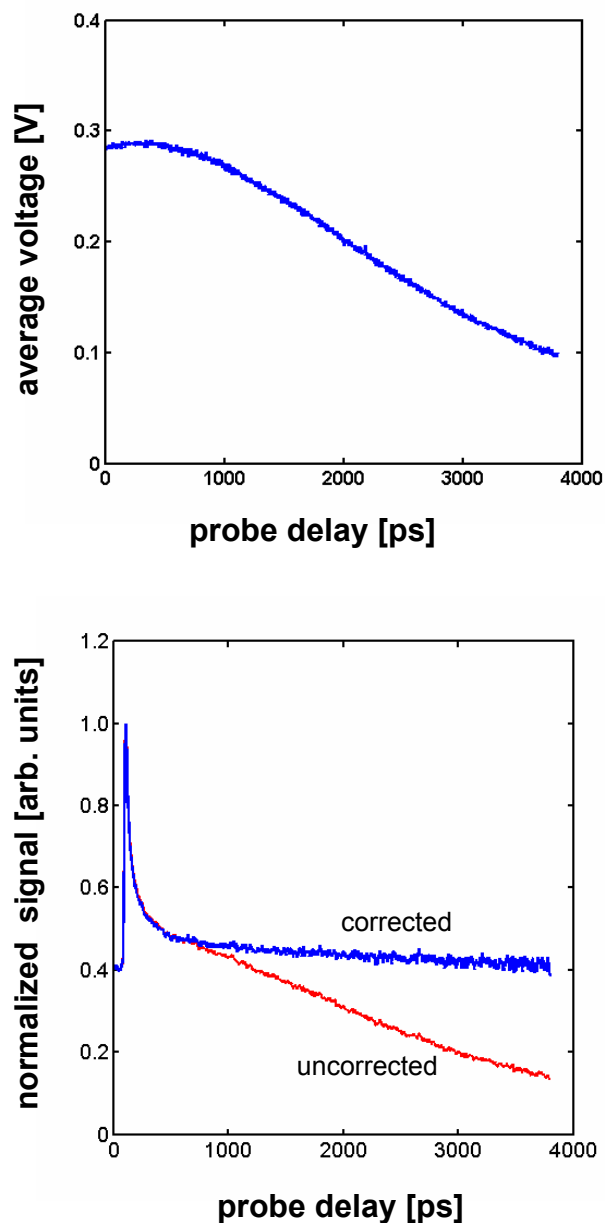


Figure 6.15: (top) A typical trace of the average detected probe signal voltage over the full time-delay scanning window (3.8 ns). The signal drops off with increasing delay because of finite beam steering of the delay stage as well as divergence of the beam. (bottom) A typical differential transmission signal obtained by exciting a silicon nanocrystal oxide film with 532 nm light and probing at 1064 nm. The red curve is the uncorrected lock-in magnitude and the blue curve shows the same signal after dividing by the normalization signal (top). Beyond about 500 ps, the need for correcting for finite beam steering and divergence is clearly evident.

### 6.5.2 Offsets and Signal Normalization

In order to mitigate the effects of beam expansion and beam steering as the delay line changes position (the round trip path length changing up to six feet), we pass the probe beam through a single-mode optical fiber mode scrambler stage. This essentially converts any beam expansion and beam steering into changes in amplitude. It would seem straightforward to simply normalize the raw signal by the nominal probe power in order to correct for the delay-line position dependent intensity. However, it is difficult to mitigate the effects of small offsets which show up at the modulation frequency. Such offsets can arise from internal device components as in the case of insufficient grounding, or through RF pickup through the cables.

$$\Delta P'_{probe}(\omega_{mod}, \Delta t) = \frac{\Delta P_{probe}(\omega_{mod}, \Delta t) + \varepsilon_N}{P_{probe}(0, \Delta t) + \varepsilon_D} \quad (6.45)$$

where  $\varepsilon$  represents a small offset for the numerator (N) and denominator (D). If  $P_{probe}(0, \Delta t) \gg \varepsilon_D$  and  $\Delta P_{probe}(\omega_{mod}, \Delta t) \gg \varepsilon_N$  are large compared to for all  $\Delta t$ , then this problem does not present a significant problem. There are two main sources of offsets that need to be eliminated or at least accounted for. Because the measurement is performed over a narrow frequency band by method of lock-in amplifier, an offset in this case represents a small leakage signal at the fundamental modulation frequency.

The largest source of an interfering signal at this frequency is of course the experimental rig itself. The square wave amplitude modulation of the pump is achieved by using an electro-optic modulator (EOM) in combination with a polarizing beam splitter. The EOM requires over 100 V modulation to achieve the  $\pi/2$  phase shift and therefore the broadcasted signal can be quite large if proper electrical isolation is not implemented. In practice, we found that this could easily add a microvolt of baseline. This signal, if stable, can be subtracted from the measured

signal during data processing but it is best to try to minimize it as much as possible.

A more subtle source of offset can come from stray (random scatter) pump light which has not been converted to 532 nm and is therefore at the probe wavelength. It is necessary to include a bandpass filter at the pump wavelength at the output of the second harmonic generator to prevent 1064 nm leakage into the detection path. However, sources of scatter upstream need to be watched carefully. If blocking the pump beam does not cause the signal on the detector to drop to the dark noise floor, try moving a beam block (e.g. a piece of cardboard) around the table to see if you can block an stray signal coming into the detector. If you still can't get rid of the offset, it is likely to be electronic coupling.

### **6.5.3 Photodiode Saturation**

One of the largest causes of error in any photonic measurement arises from the operation of the photodetector in a nonlinear regime. It would seem that such a scenario could only happen to the most junior of experimentalist but detector nonlinearity can fool even the most seasoned lab veterans. Never assume that the detector is operating linearly. Prove to yourself that it is by measuring it each time before making a set of measurements.

There are two basic types of photodiode saturation to consider: (1) power saturation and (2) intensity saturation. Although the effects of either forms may take may present themselves as subtle artifacts in a measurement requiring a trained eye, the first form of saturation is easiest to detect. Increasing the power striking the photodetector should increase linearly. This should be verified each time.

Intensity saturation of the photodetector occurs when the incident intensity on the photodetector exceeds the level at which the detector behaves linearly. The real problem with this scenario is that the DC voltage readout of the detector at which the intensity saturation occurs can be much lower than the level which would indicate power saturation of the detector, a value typically provided in datasheets. The only way to determine if one is in this regime is by measuring output voltage as a function

spot size on the detector. This can be done by putting the detector on a linear positioning stage. Diagnosing this problem can be rather tedious but needs to be done.

#### 6.5.4 SNR Optimization

A common mistake of a junior experimentalist is to optimize signal level rather than signal to noise ratio (SNR). At the end of the day, the signal level is meaningless. It only matters what information you can extract from your measurement which is set by the SNR. So, it is understandable that one would first reach for the gain knob on a detector in order to improve SNR if one encounters a noisy signal. The fact is that the gain knob will amplify both the signal and the noise in at best equal amounts but in all real cases, additional noise gets added to the system thus degrading the SNR. Therefore, it is necessary to compute the SNR as a function of signal level or gain or other experimental system parameters in order to identify the optimal SNR.

The fundamental limit of measuring a signal in the pump-probe system is shot noise of probe laser beam, which is given by the expression (6.46)

$$v_{rms} = i_{rms} R = R \sqrt{2e \mathfrak{R} P_{probe} \Delta f} \quad (6.46)$$

where  $R$  is the transimpedance gain (V/A) of the detector amplifier Probe is the probe power and  $\mathfrak{R}$  is the responsivity of the detector measured in A/W. The ThorLabs switchable gain InGaAs detector has a responsivity of about 0.75 A/W at 1064 nm and a transimpedance gain of  $1.5 \times 10^3$  V/A (0 dB setting, high-Z load). Therefore, an average 200 mV output signal corresponds to an average power of about 180  $\mu$ W or a photodiode current of 134  $\mu$ W. At this power level the rms voltage fluctuation due to shot noise at the input to the lock-in amplifier would then be approximately 10 nV in a 1 Hz measurement bandwidth. This sets the ultimate detection limit of the system. In practice, however, it is difficult to reach the shot noise level for a number of reasons. First, the laser itself has excess amplitude noise which can greatly exceed the shot noise limit. Furthermore, vibrational motion of mirrors and etalon effects in the system

can lead to additional amplitude modulation but is typically limited to audio frequencies. Within the detector itself, there is a wealth of additional sources of noise which may bring the total noise floor above the shot noise limit. Figure 6.16 shows the rms noise signal for the ThorLabs PDA10CS detector used in this work as a function of average probe laser power measured as average detector output voltage.

The blue diamond data is the actual noise signal measured and the purple square data set represent the predicted shot noise rms voltage. The inset in the top plot shows the ratio of the two voltages and indicates how close we are to the shot noise limit. We see that for detector output voltages less than 200 mV, we achieve a noise floor within 3 dB of the shot noise limit. This is achieved without any laser noise canceling techniques which indicates that the laser amplitude noise is shot noise limited at 100 KHz. In bottom plot, the relative SNR for a measurement of a 20 nm thick silicon nanoparticle-oxide film is shown as a function of probe power (again measured as the average output voltage of the detector). Here, the pump power incident on the sample was fixed while the probe power was changed. We can see from this measurement that the SNR is optimal for a value of 200 mV detector voltage. All of the measurements described in the following chapter were taken at this probe power.

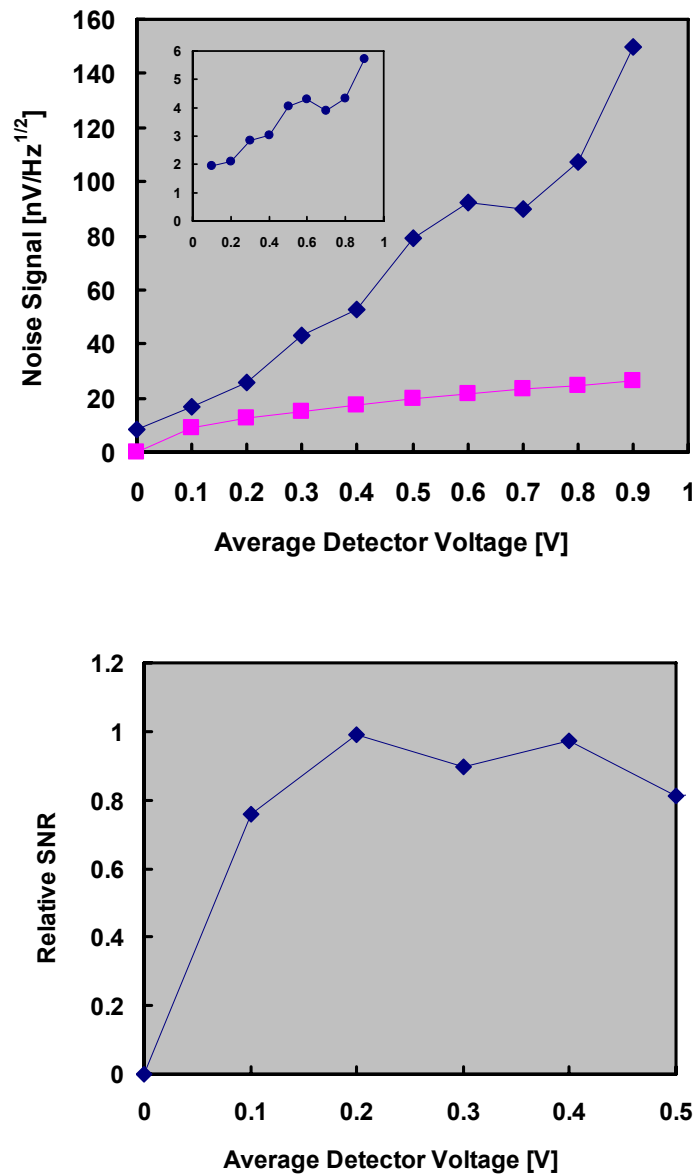


Figure 6.16: (Top) (Blue diamond) Measured RMS voltage fluctuation on probe detector in a 1 Hz bandwidth as a function of average detector signal voltage. (Magenta square) Data points showing calculated shot noise voltage for given photocurrent. The inset shows the ratio of the measured voltage noise to the shot noise limited voltage noise. For signal levels on the order of 150 mV, one can operate 3 dB of the shot noise floor. (Bottom) Plot of the relative SNR as a function of average detector voltage. Increasing the signal above about 150 mV does not improve SNR.

## Chapter 7

### Energy Conversion in Silicon Nanophotonic Materials

In this chapter, I will present my work investigating the dynamics of energy conversion in dense systems of luminescent silicon nanocrystals, a class of materials which is promising in the development of silicon-compatible optical sources. The carrier dynamics of optically pumped silicon nanocrystals embedded in amorphous silicon oxide films were measured using picosecond transient and steady-state two-color optical pump probe techniques described in part in Chapter 6. Rohan Kekatpure and Dr. Aaron Hryciw of the Brongersma Group in the Materials Science Department at Stanford University provided thin-film silicon nanocrystal-oxide samples in support of this work.

#### 7.1 Silicon Nanophotonics

The transistor has been reduced in size to the point that its intrinsic signal delay contribution is often negligible compared to that of the electronic interconnect bridging two logic gates. This is particularly true when signals must span long distances (100s of micrometers to millimeters) such as the case of clock distribution networks. The case for replacing copper wire with either free-space or waveguide optical interconnects between chips and perhaps even on-chip has been argued by many but perhaps most effectively by Miller [85]. Optical signaling provides a number of fundamental advantages which include significantly higher signal bandwidth (100-1000 THz), which allows for higher data rates as well as a reduction in cross-talk through frequency filtering, and potentially lower power consumption. The latter

advantage arises from the fact that the low-loss waveguides are available and the entire signal line does not need to be charged to the signal voltage during transmission.

In theory, a strong case can be made for replacing copper wire with optical interconnects. However, to realize a simple optical communication link, low cost CMOS compatible optics and optoelectronic components are needed, including a source, a modulator, an interconnect, and a receiver. Tremendous strides in developing viable silicon-based solutions for the latter three link elements have been made in recent years [86-88]. However, a suitable light source remains elusive. A low-cost, CMOS compatible coherent light source which can be detected by silicon or germanium detectors is desperately needed to complete the short-haul communication link.

For fundamental reasons, this goal is a monumental challenge. It is well known that silicon in its bulk crystalline state is a poor source of photons due to its indirect bandstructure. That is, when making a transition from an excited conduction band state near the band edge, the electron must interact with a phonon of large wavevector ( $\sim 1 \text{ \AA}^{-1}$ ) in order to conserve crystal momentum. Hence, the radiative transition is a second order process and substantially reduced. A silicon Raman laser was demonstrated by Rong and coworkers in 2003 [89] and was a technological advancement in silicon photonics. However, such a laser is not suitable for optical interconnection since it fundamentally requires an optical pumping source. In essence, it is a frequency conversion device.

The fundamental efficiency of light emission in bulk silicon can be greatly increased by reducing the size of the silicon lattice to a length scale which is comparable to the unit cell of the silicon lattice,  $a = 5.43 \text{ \AA}$ . As discussed in Chapter 2, the reduction in the size of the crystal leads to broadening of the wavefunction peaks in momentum or  $k$ -space on the order of  $\sim 1/\Delta L$ . Si nanocrystals (NCs) formed inside of a host dielectric film exhibit strong and stable room-temperature visible and near-infrared luminescence [90-97] as well as the ability to efficiently sensitize co-doped  $\text{Er}^{3+}$  ions which fluoresce in the important telecomm band centered around 1540 nm [98, 99]. Such properties offer exciting possibilities in developing practical Si-based

coherent light sources for on-chip and chip-to-chip interconnects. Thus, these materials are of great technological interest. In addition, the fundamental physics of efficient light emission in low-dimensional semiconductors having indirect band gaps as well as the dynamics of closely-packed interacting quantum dots or NCs are rapidly advancing fields of study which continue to provide intriguing questions of fundamental importance.

Despite great interest in their optical properties, it has now been established that free-carrier absorption (FCA) and fast nonradiative carrier recombination (NRCR), such as Auger processes, severely limit the net optical gain in Si NC-dielectric films to impractical levels [90, 93, 100-102]. Progress towards fully understanding these intrinsic loss mechanisms which compete unfavorably with stimulated emission has been impeded by the substantial size inhomogeneity that is characteristic of most common forms of these films, e.g. porous Si [94, 95], Si rich Si oxides (SRO) [90-93] and nitrides (SRN) [97], making it difficult to disentangle the effects of competing optoelectronic phenomena [103, 104]. Our limited understanding of NRCR and FCA losses in PS and SRO films is evident by the establishment of multiple seemingly contradicting sets of carrier rate equations over the past two decades, cf. Refs. [95, 102, 105-107]. In order to reconcile these differences, we have precisely measured the optically induced FCA losses of dense SRO films over an extended pump intensity range and have compared the results to existing carrier dynamics models. We find that none of the existing rate equation models is capable of accurately reproducing the pump intensity dependence over the full range of data. However, by introducing a single near-field, Coulombic, dipole-dipole interaction term of similar form as described in Ref. [108] between excited NCs to the independent (i.e. non-interacting) NC coupled rate equation model by Kovalev *et al.* [105], we are able to accurately reproduce the data over the full range of pump intensities. Furthermore, we find that the existing rate equation models can be interpreted as limiting forms of this model.

The optically induced free-carrier absorption coefficient ( $\alpha_{FC}$ ) of densely-packed ( $2 \times 10^{18} \text{ cm}^{-3}$ ) Si nanocrystals (NCs,) embedded in a glass matrix was accurately measured in the near-infrared using two complementary pump-probe techniques.

When compared to established carrier dynamical models, the dependence of  $\alpha_{FC}$  on optical pump intensity reveals enhanced nonradiative recombination at high pump intensity ( $> 10 \text{ KW/cm}^2$ ) which can be explained by the interaction of carriers located in spatially separated NCs. The rate of recombination due to the interaction between a pair of excited NCs is determined to increase inversely with the sixth power of their separation distance ( $\tau^{-1} \sim \gamma R^{-6}$  ;  $\gamma \approx 2.9 \times 10^{-31} \text{ cm}^6 \text{ s}^{-1}$ ) and is indicative of near-field dipole-dipole energy transfer. These results explain the source of an apparent inconsistency in power-law behavior found in previous carrier dynamics studies on similar materials which did not account for this interaction.

## **7.2 Physical Properties of Silicon Nanocrystal – Oxide Matrices**

In these studies, we examine the optical properties of silicon nanocrystals formed from silicon rich oxides (SRO) fabricated using two different methods, a plasma enhanced chemical vapor deposition (PECVD) method and an e-beam evaporation method. Both methods employ the same basic nanocrystal formation physics. A non-stoichiometric silicon oxide is first deposited onto a substrate (e.g. fused silica or silicon) with an excess amount of silicon. After deposition, a high-temperature anneal step is performed which induces the nucleation and growth of small  $\sim 5 \text{ nm}$  nanocrystallites with a distribution of size and density of order of  $10^{18} \text{ cm}^{-3}$ . The next two sections will describe the fabrication details and the film properties using the two techniques.

### **7.2.1 PECVD Deposited SRO**

Thin films of silicon rich silicon oxide were prepared by PECVD and subsequent annealing by Rohan Kekatpure in the Stanford Nanofabrication Facility (NSF). The conditions for processing are described in Ref. [93] but are stated here for reference. The deposition was carried out using a 2% silane  $\text{SiH}_4$  which was diluted in  $\text{N}_2$  and a precursor nitrous oxide ( $\text{N}_2\text{O}$ ) gas on a fused silica substrate at a temperature and pressure of  $350^\circ\text{C}$  and  $650 \text{ mTorr}$  respectively. The  $\text{SiH}_4:\text{N}_2\text{O}$  flow ratio was set to

control the silicon concentration which indirectly controlled the size and density of silicon nanocrystals. After deposition, the films were annealed at 1100 °C for 1 hour to form the nanocrystallites.

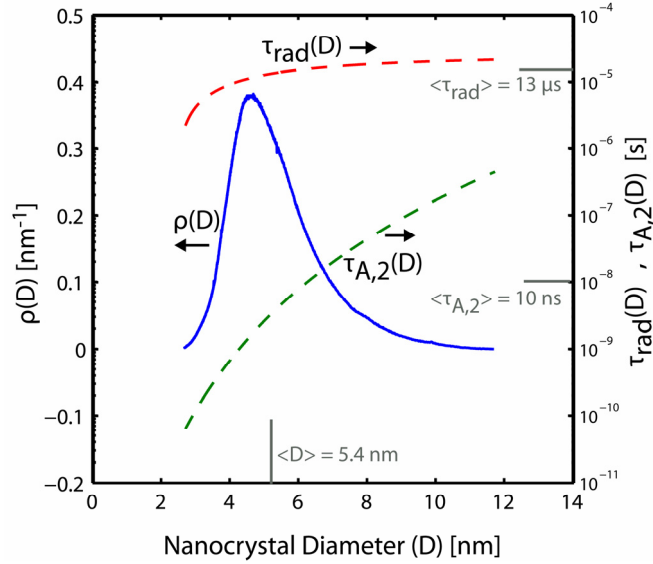


Figure 7.1: Physical properties of the SRO films measured in this work.  $\tau_{\text{rad}}$  was obtained from transient PL experiments described in [4] and  $\tau_{A,2}$  was computed using the formula,  $\tau_{A,2}^{-1} = C_A(2/V)^2$ , with  $C_A = 4 \times 10^{-31} \text{ cm}^6/\text{s}$  and  $V = (\pi/6)D^3$ . The distribution function was calculated using the methodology described in [93, 102].

Figure 7.2 shows high resolution HRTEM images of 500 nm thick SRO films grown using this technique.

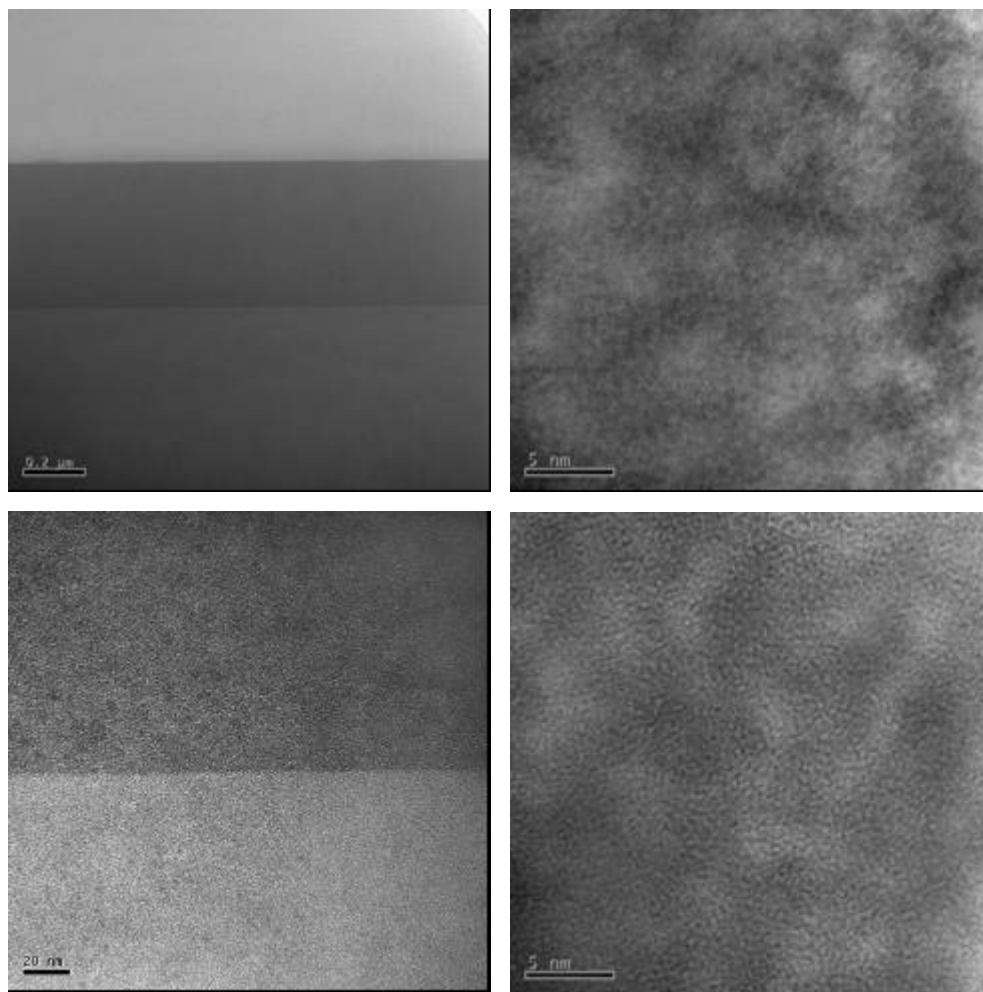


Figure 7.2: High resolution TEM cross-section images for a PECVD deposited silicon rich oxide, which was subsequently annealed at 1100 C for one hour to produce silicon nano-crystallites with diameters in the vicinity of 5 nm and a density of  $2 \times 10^{18} \text{ cm}^{-3}$ .

### 7.2.2 E-beam Evaporated SRO

In subsequent sections, we will discuss the motivation for examining SRO films down to 10 nm within multi-layer superlattice structures. In order to control the thickness of SRO layers to nm resolution, an e-beam evaporation method was preferred over the PECVD process. This is because the first 20-30 nm of the PECVD film are believed to have a differing film composition than the remainder of the film. Dr. Aaron Hryciw performed the deposition of e-beam evaporated silicon monoxide (SiO) in various thicknesses and followed the deposition with a 1 hour anneal at 1000 °C. High resolution TEMs of a 230 nm thick sample and a superlattice consisting of 10 alternating layers of 23 nm SRO and 30 nm SiO<sub>2</sub> are shown in Figures 7.3 and 7.4 respectively.

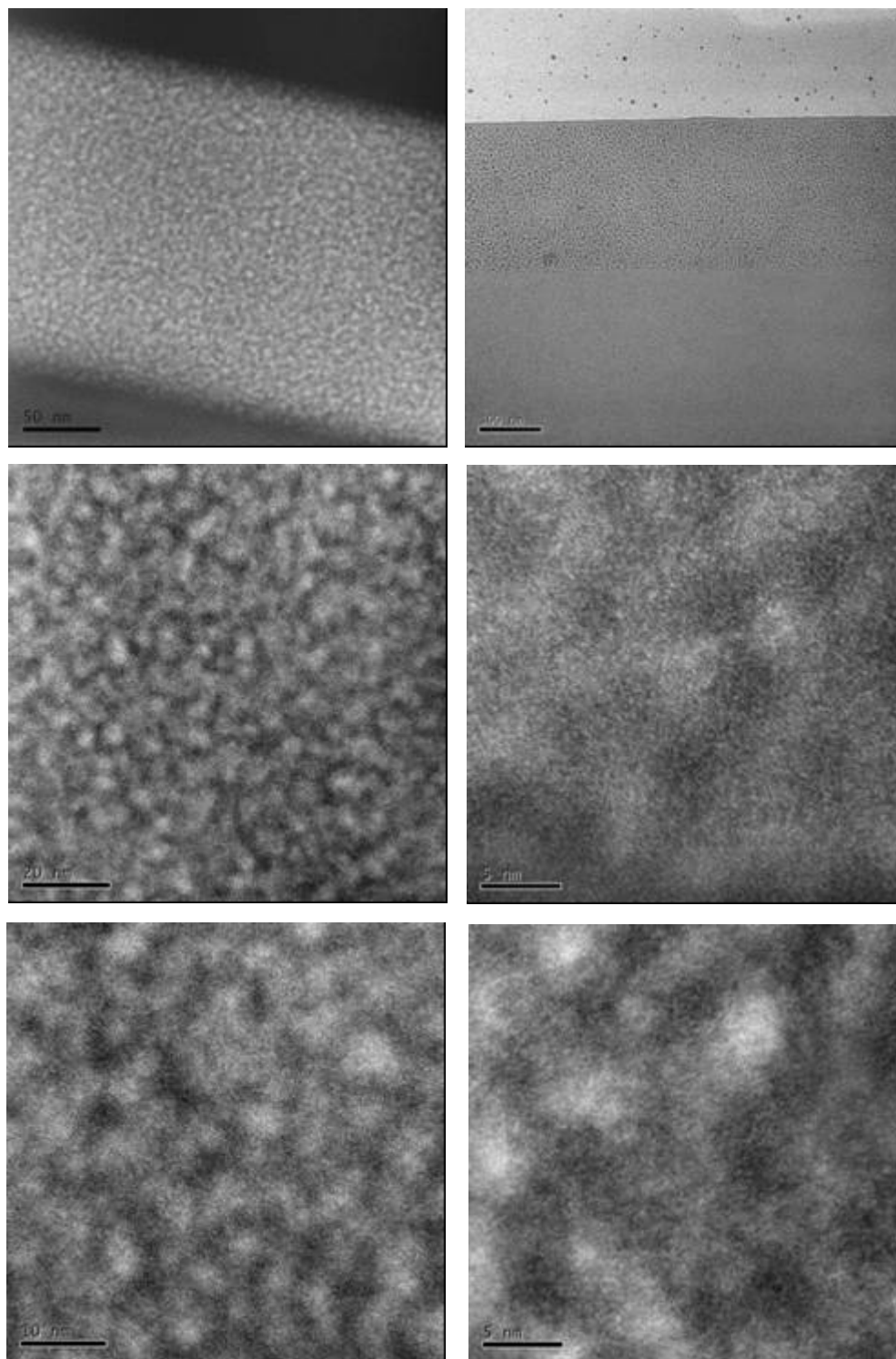


Figure 7.3: High resolution TEM cross-section images for a e-beam evaporated SiO<sub>2</sub>, which was subsequently annealed at 1100 C for one hour to produce silicon nano-crystallites.

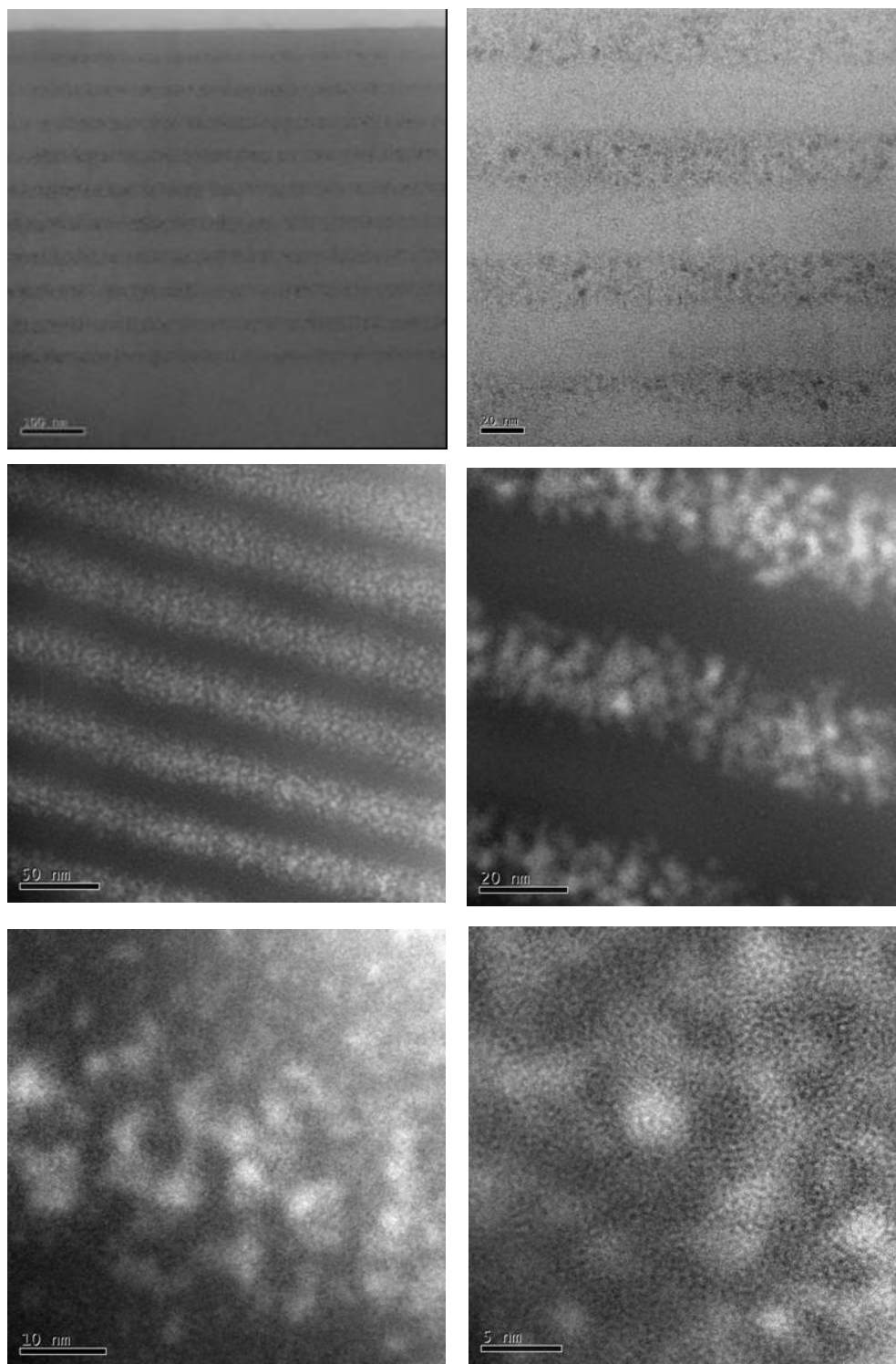


Figure 7.4: High resolution TEM cross-section images for a e-beam evaporated SiO/SiO<sub>2</sub> superlattice film containing 10 periods of 20 nm nc-Si-SiO<sub>2</sub> and 20 nm SiO<sub>2</sub> barrier layer. The sample was annealed at 1100 C for one hour to produce silicon nanocrystallites as can be seen as bright spots in the energy filtered HRTEMs.

### 7.3 Transient and Steady-State Differential Transmission Pump-probe Measurements

In this section, I will present the results of transient and steady-state differential transmission measurements made on systems of silicon nanocrystals embedded in silicon oxide matrix. The samples are those discussed in the previous section. The pump-probe technique I employed is that described in Chapter 6.

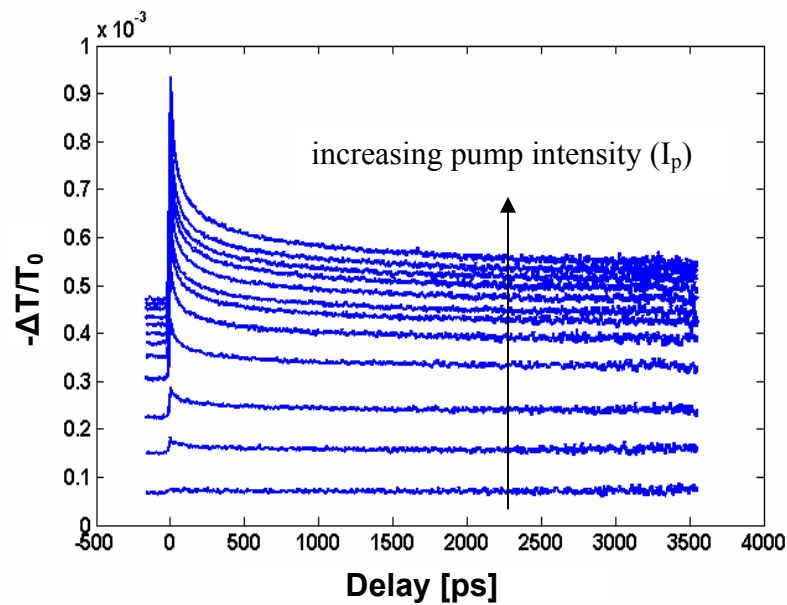


Figure 7.5: Transient, differential transmission measurement results for the PECVD sample. Both the background signal and the peak amplitude of losses increase with increasing pump intensity. The initial decay time constant reduces with increasing pump intensity as a consequence of Coulombic interaction between excited electrons.

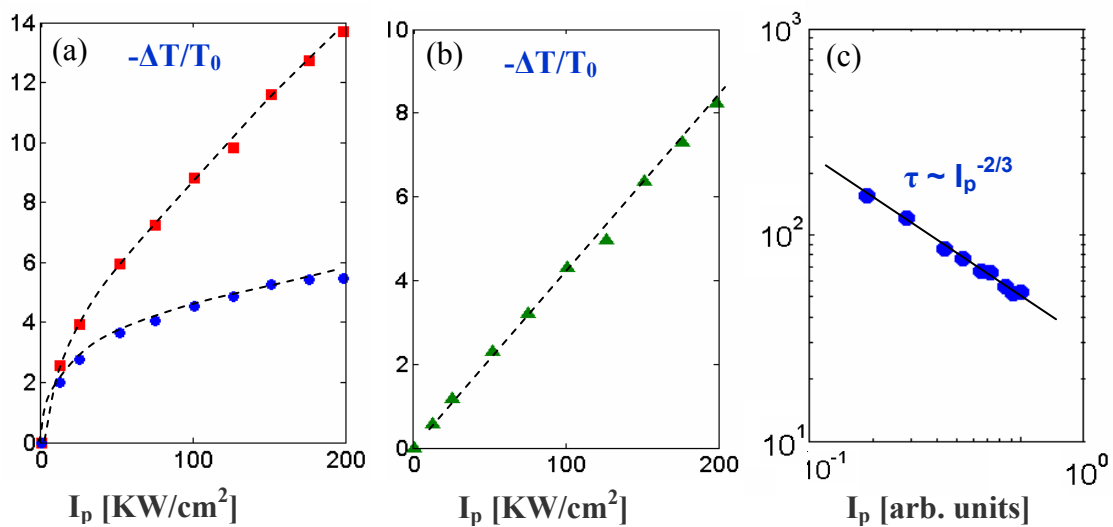


Figure 7.6: (a) The peak (red square) and negative time (blue circles) delay relative transmission loss as a function of pump intensity. (b) The difference between the peak value and the negative time delay (steady-state) value vs. pump intensity showing linearity of pump excitation cross-section. (c) The initial time constant as a function of pump intensity showing a  $-2/3$  law.

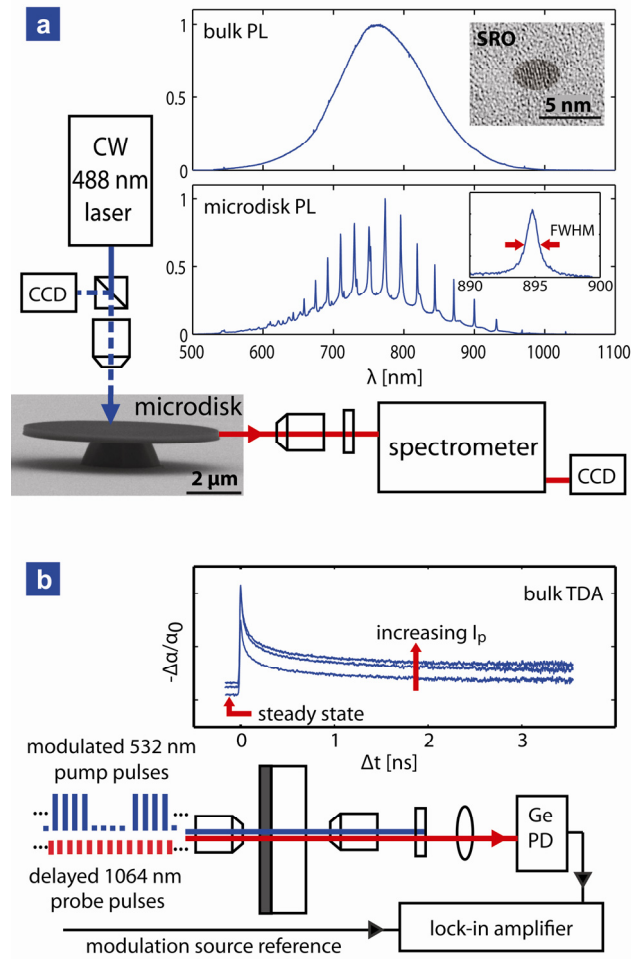


Figure 7.7: Two experimental methods used to determine the optically-induced FCA losses in SRO films by measuring the (a) loss-induced spectral broadening of the 895 nm PL emission line of an SRO microdisk resonator and the (b) steady-state (negative time-delay) component of picosecond transient differential absorption (TDA) measurements performed on a 500 nm thick SRO film deposited on a fused-silica substrate.

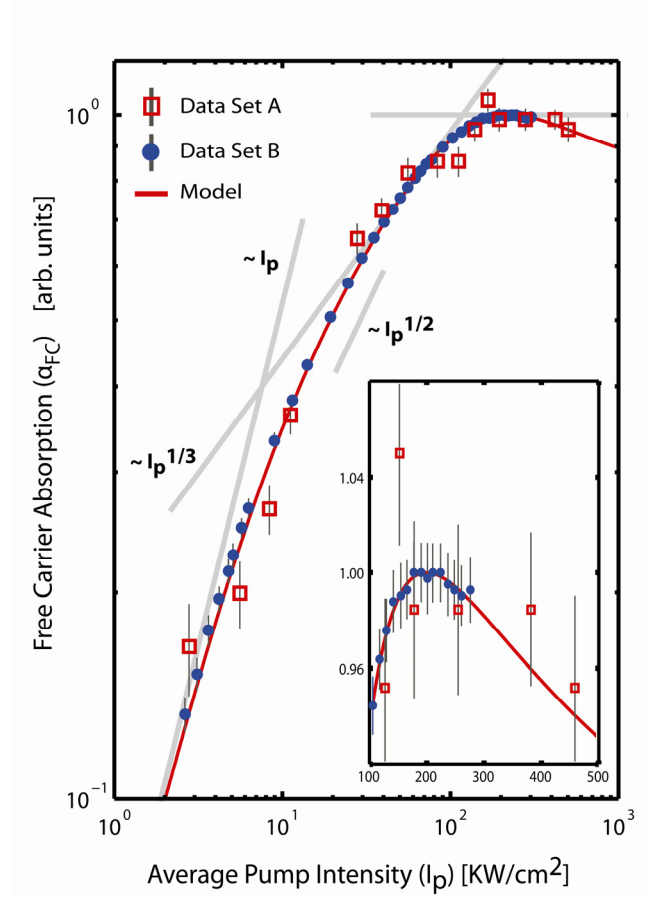


Figure 7.8: The normalized free-carrier absorption coefficient ( $\alpha_{FC,max} = 24.6 \pm 0.3 \text{ cm}^{-1}$  at 1064 nm) vs. pump intensity for the two complementary pump-probe measurement techniques described in the text and in Fig 1. The solid red curve is the best fit for the proposed model. Linear, 1/3 and 1/2 power-law curves are included as visual guides. The inset is a zoomed in view of the maximum at 200 KW/cm<sup>2</sup>.

## 7.4 Carrier Dynamics Modeling

### 7.4.1 Coupled Rate Equation Analysis

Charge densities exhibiting 1/2 and 1/3 power-law dependences on pump intensity are characteristics of bulk semiconductors owing to two (e.g. bimolecular Auger [95]) and three-particle Auger processes [109] respectively. Therefore, given the observed power-law behavior of the data, bulk-like linear-plus-quadratic [95, 106, 107] or linear-plus-cubic [102] loss rate models could explain the data below 10 KW/cm<sup>2</sup> and 100 KW/cm<sup>2</sup> respectively. However, these models do not predict saturation at high intensities, which is a key signature of isolated or independent NCs [105, 110]. The independent NC model by Kovalev et al. [105] provides the closest fit to the data but, in its current form, can not reproduce the data to within experimental uncertainty. Motivated by the observation of the 1/3 power-law behavior, we modify the isolated NC model [16] to include a single cubic interaction term between NCs resulting in the following set of coupled rate equations:

$$\dot{N}_0 = -\phi\sigma_{01}N_0 + N_1/\tau_1 + \Gamma N^{*2}N_1 \quad (7.1a)$$

$$\dot{N}_1 = \phi\sigma_{01}N_0 - \phi\sigma_{12}N_1 - N_1/\tau_1 + N_2/\tau_{A,2} - \Gamma N^{*2}N_1 \quad (7.1b)$$

$$\dot{N}_2 = \phi\sigma_{12}N_1 - N_2/\tau_{A,2} \quad (7.1c)$$

$$N_T = N_0 + N_1 + N_2 \quad (7.1d)$$

where  $N_T(D)$  and  $N_j(D, t)$  are the densities of NCs with diameter  $D$  and those having  $j$  EHPs respectively,  $\phi$  is the instantaneous pump photon flux,  $\sigma_{ij}$  is the absorption cross-section for promoting a NC from the  $i^{\text{th}}$  to the  $j^{\text{th}}$  e-h state,  $\tau_1$  is the single EHP recombination lifetime including both radiative (rad) and nonradiative (nr) channels, ( $\tau_1^{-1} = \tau_{rad}^{-1} + \tau_{nr}^{-1}$ ), and  $\tau_{A,2}$  is the two EHP Auger recombination lifetime. Higher excitation levels ( $O(N_3)$ ) have been omitted because their steady-state occupations are deemed to be negligible for all measured intensities. For the same considerations, we

neglect interactions between NCs of higher excitation levels than those which are in the  $N_l$  state and write the cubic interaction term as the product of  $N_l$  and  $(N^*)^2$ . Here, we have defined  $N^*$  as the instantaneous average concentration of NCs containing one EHP, i.e.  $N^*(t) \equiv \int \rho(D) N_l(t, D) dD$  where  $\rho(D)$  is the unit normalized distribution of NC

diameters which we have plotted in Fig. 3 [93, 102]. We interpret the physical meaning of this interaction as an increase in the nonradiative decay of a selected NC in the  $N_l$  state due to the presence of neighboring NCs also in the  $N_l$  state, whose strength increases in proportion to  $\langle\langle N_l \rangle\rangle^2$  or inversely to the sixth power of their average separation distance, i.e.  $\tau^{-1} \sim \langle\langle R_l \rangle\rangle^{-6}$  where  $\langle\langle R_l \rangle\rangle \sim \langle\langle N_l \rangle\rangle^{-1/3}$  [111]. Such a dependence is indicative of near-field Coulombic dipole-dipole coupling [108]. However, unlike the case explicitly treated in Ref. [108], where an EHP in one NC interacts with the ground state of a neighboring NC, here the interaction occurs between two EHPs located in separated NCs in close proximity to each other. Such a resonant nonradiative energy transfer mechanism was first investigated by Fröhlich and Mahr [112] for a system of excited F-centers in KI crystals.

#### 7.4.2 Long-range Coulombic Dipole-Dipole Coupling Between Excited States

Recently, evidence of Auger recombination involving charge carriers located in separate nanocrystals (NCs) was reported for a dense system of silicon NCs embedded in a thick (500 nm) silicon oxide glass matrix [113]. The recombination rate of an electron-hole pair (EHP) located in a randomly selected NC due to the presence of an EHP located in a neighboring NC located a distance  $R$  away was determined to have the form  $\tau^{-1} = \gamma R^{-6}$ , where  $\gamma$  controls the pair-interaction strength. This form of an interaction arises from the near-field expansion of the Coulomb potential as described by the Dexter-Forster (D-F) theory [114]. However, in contrast to the standard definition of Dexter-Forster nonradiative energy transfer whereby an EHP in one NC or molecule interacts with the ground state of a neighboring site, here the interaction is

between two EHPs located in separate sites. In both cases, two (labeled  $i$  and  $f$ ) two-electron(hole) wavefunctions  $\psi_{i,f}(\vec{r}_1, \vec{r}_2)$  are coupled through the mutual Coulomb potential of having the form

$$H' = \frac{-e^2}{\kappa |\vec{r}_1 - \vec{r}_2|} \quad (7.2)$$

where  $\kappa$  is the dielectric permittivity function. The distinction here is in the description of the initial and final states  $\psi_{i,f}(\vec{r}_1, \vec{r}_2)$ ; the physical process though is essentially the same. We shall therefore refer to the latter case as Förster-Auger (F-A) recombination.

$$\tau^{-1} = \frac{2\pi}{\hbar} |\mu_{if}|^2 \delta(E_i - E_f) \quad (7.3)$$

where, the matrix element is given by

$$\mu_{if} = -\frac{e^2}{\kappa} \iint d\vec{r}_1 d\vec{r}_2 \frac{\psi_f(\vec{r}_1, \vec{r}_2) \psi_i(\vec{r}_1, \vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|} \quad (7.4)$$

which to lowest non-zero order, the dipole-dipole interaction becomes

$$\mu_{if} \approx \mu_{if}^{(d-d)} = -\frac{e^2}{\kappa R^3} \iint d\vec{r}_1 d\vec{r}_2 \psi_f(\vec{r}_1, \vec{r}_2) \psi_i(\vec{r}_1, \vec{r}_2) \quad (7.5)$$

Taking the modulus squared of (7.5), we get the expected result which predicts that the transition rate scales as

$$\tau^{-1} \sim \kappa^{-2} R^{-6}. \quad (7.6)$$

One of the consequences of D-F theory is that energy is typically transferred from smaller NCs to larger NCs as a result of the bandgap scaling through the quantum confinement effect. However, in the DFA theory, this trend is not necessarily expected. For NCs or quantum dots (QDs) larger than about 1 nm in diameter, the quantum state properties, e.g. wavefunction and density of states, closely resemble that of the bulk material for energies greater than about 1 eV above and below the band edges. Therefore, given that NC bandgap energies typically exceed 1 eV for small NCs, it should not be surprising to find that the DFA energy transfer is less dependent on the size of the participating NCs and energy transfer is essentially bidirectional and far more efficient than D-F energy transfer.

In the framework of the DFA model, the total recombination rate of an EHP in a randomly selected NC due to such an interaction is found by summing over all of the interactions by neighboring NCs containing an EHP. Assuming a uniform density of NCs containing a single EHP ( $N_1$ ), a constant value of  $\gamma$ , and that the interactions between NCs can be treated independently, the total recombination rate in three spatial dimensions (3-D) is

$$\tau^{-1} = 2 \int_{R_{\min}}^{\infty} \frac{\gamma N_1 4\pi R^2 dR}{R^6} = \left( \frac{8\pi\gamma}{3\alpha^3} \right) N_1^2 = \Gamma N_1^2 \quad (7.7)$$

where  $R_{\min} = \alpha_{\nu} N_1^{-1/3}$ , the proportionality constant  $\alpha_{\nu}$  is of order unity. The factor of 2 in the front of the integral accounts for the fact that an excited nanoparticle can act as either a donor or an acceptor with equal probability [112]. From this analysis, we find that the rate of recombination in 3-D is proportional to  $N_1^2$ . Furthermore, we find that the ratio between the effective, or macroscopic, dipole-dipole interaction strength to the microscopic interaction strength, is given by

$$\Gamma / \gamma = \frac{8\pi}{3\alpha_{\nu}^3} \approx 110 \quad (7.8)$$

where  $\alpha_V$  is determined to be approximately 0.42 in three dimensional systems which are Poission distributed. It is interesting to note that this value falls between the value which would give the average nearest neighbor distance ( $\alpha_V = 0.55$ ) and the  $R^{-6}$  weighted nearest neighbor average ( $\alpha_V = 0.26$ ).

Neglecting all other recombination channels, the rate equation for NCs in the one EHP state can then be written as

$$\dot{N}_1 = G - N_1 / \tau = G - \Gamma N_1^3 \quad (7.9)$$

where  $G$  is the average generation rate of an EHP. At steady-state,  $N_1 \sim G^{1/3}$ . The 1/3 power-law dependence on the generation rate for a system of physically unconnected NCs is indicative of DFA coupling but can be confused with continuum transport dynamics which also predicts a 1/3 law behavior owing to three-particle Auger processes [109]. It should be stressed that this ambiguity exists only at intermediate generation rates where multiple EHP probability is low. At higher generation rates, multiple EHP generation within a NC leads to rapid decay to the single EHP state and exhibits saturation. Such a behavior is not predicted by the standard bulk Auger recombination theory.

### 7.4.3 Analytical Model

By treating each parameter as a constant [115] and setting  $\sigma_{ij} = \sigma$  for all  $i$  and  $j$ , we obtain an analytical steady-state expression for  $\langle\langle N_1 \rangle\rangle$  and  $\langle\langle N_2 \rangle\rangle$ :

$$\Gamma \langle\langle N_1 \rangle\rangle^3 + [\varphi\sigma + \varphi^2\sigma^2\tau_{A,2} + 1/\tau_1] \langle\langle N_1 \rangle\rangle - \varphi\sigma \langle\langle N_T \rangle\rangle = 0 \quad (7.10a)$$

$$\langle\langle N_2 \rangle\rangle = \varphi\sigma\tau_{A,2} \langle\langle N_1 \rangle\rangle \quad (7.10b)$$

where  $\langle\langle N_1 \rangle\rangle$  is found by taking the positive real root of (2a) and  $\langle\langle N_T \rangle\rangle$  is the density of all NCs. The effective FCA coefficient for the entire NC system can be written as

$$\alpha_{\text{FC}} = \sigma_{\text{FC}} \left[ \langle\langle N_1 \rangle\rangle + e_{21} \langle\langle N_2 \rangle\rangle + e_{31} \langle\langle N_3 \rangle\rangle \dots \right] \quad (7.11)$$

where  $e_{jl}$  is a dimensionless weighting coefficient which indicates the contribution of the  $j^{\text{th}}$  EHP state to  $\sigma_{\text{FC}}$  relative to that of the one EHP state. Typically, the assumption  $e_{jl} = j$  is made, a direct extension of the linear FCA theory for nondegenerate bulk material [116]. However, recent calculations for CdSe QDs have cast doubt on the general validity of this assumption for low-dimensional systems [117]. As we will show, we find that  $e_{21}$  must drop below unity in the steady-state case in order to reproduce the maximum observed in the data.

To the best of our knowledge, the maximum in  $\alpha_{\text{FC}}$  has never been reported before. It is best explained by allowing  $e_{21} < 1$ , i.e.  $N_2$  states have a lower contribution to  $\sigma_{\text{FC}}$  relative to the  $N_1$  states. For, if  $e_{21} \geq 1$ , then the  $\alpha_{\text{FC}}$  can not decrease because of the reaction  $N_1V + \text{photon} \rightarrow (N_1V - 1) + (N_2V + 1)$ ,  $V$  being a unit volume, which leads to  $\Delta\alpha_{\text{FC}} \sim |\Delta N_1|(e_{21} - 1) < 0$ . Hence,  $e_{21} < 1$  is necessary to produce a maximum, generally. However, this view conflicts with observed picosecond TDA data which have shown that FCA losses *increase* shortly after pump excitation under conditions when the fraction of newly created  $N_2$  states should be substantial [118]. These conflicting observations suggest that  $e_{21}$  is time dependent, e.g.  $e_{21}(t_0) = 2 \rightarrow e_{21}(t_\infty) = 0.5$ . This could occur if the  $N_2$  states are composed of multiple 2 e-h pair states, one group contributing to FCA ( $N_{2,a}$ ) and the remaining group being inactive ( $N_{2,b}$ ), perhaps mediated by a surface trap state.

The solid red curves in Figs. 2, 4, 5 and 6b are the best fits to the data using Eqs. 2 and 3 and the following optimal parameters:  $\sigma = (3.84 \pm 0.19) \times 10^{-16} \text{ cm}^2$  ;  $\tau_1 = 37.72 \pm 1.89 \text{ ns}$  ;  $\tau_{A,2} = 5.94 \pm 2.61 \text{ ns}$  ;  $\Gamma = (2.88 \pm 1.76) \times 10^{-29} \text{ cm}^6 \text{ s}^{-1}$  ; and  $e_{21} = 0.5 \pm 0.06$ . Fig. 4 plots the values of  $\langle\langle N_1 \rangle\rangle$  and  $e_{21} \langle\langle N_2 \rangle\rangle$  normalized by the NC density,  $\langle\langle N_T \rangle\rangle$ , and shows their individual contributions according to the model. The model is

capable of fitting the data exceptionally well with the optimal parameter set yielding an average relative error magnitude of 0.4 % and a peak error of 2.5 % occurring at the lowest measured intensity. The sensitivity of the model to  $\Gamma$  and  $e_{21}$  is conveyed in Fig. 4a and 4b respectively.

It is useful to compare the extracted parameter values to diameter-weighted average values using the distribution function shown in Fig. 3 obtained using the methods described in Ref. [102]. For the proposed model, we find  $\sigma$  to be  $\approx 2x$  higher than the expected value of  $\langle\sigma\rangle = 1.71 \times 10^{-16} \text{ cm}^2$  obtained from transient PL measurements but is well within the accepted range of values for annealed SRO materials [105]. Furthermore, assuming the relation  $\tau_{A,2}^{-1} = C_A(2/V)^2$  holds [119], where  $V = (\pi/6)D^3$ , and taking  $C_A = 4 \times 10^{-31} \text{ cm}^6/\text{s}$  [118], we find  $\langle\tau_{A,2}\rangle = 10.26 \text{ ns}$ , which is also within about a factor of two of the value we extract. Despite the good agreement for  $\sigma$  and  $\tau_{A,2}$ ,  $\tau_1$  is found to be  $\approx 350x$  lower than the average radiative lifetime,  $\langle\tau_{\text{rad}}\rangle = 13.2 \text{ }\mu\text{s}$  [93]. It is important to note that this value of  $\tau_1$  which we extract is not particularly sensitive to the model used because  $\tau_1$  primarily impacts data fitting in the low intensity regime where existing models are in agreement. We attribute the reduced value of  $\tau_1$  to the large fraction of nonradiating (nrad) NCs which are sampled by the pump-probe method but which are not sampled directly in PL measurements. The PL quantum efficiency (PLQE), a good measure of the fraction of radiating (rad) NCs, was recently reported by Dal Negro *et al.* [97] for PL optimized SRO films exhibiting nearly identical optical properties as the films measured in this work to be about 5 %. A low PLQE supports the conclusion that the pump-probe methods used in this work primarily sample the nrad-NC population. The fact that the value of  $\sigma$  is of similar magnitude for both rad-NCs and nrad-NCs suggests that localized defect states near the band edge are responsible for the reduced value of  $\tau_1$ .

Before discussing these results in greater detail, we first demonstrate the effectiveness of the proposed analytical model over that of the independent NC model [105] in reproducing the experimental data. In Fig. 5, we compare the best fits to the data using the independent NC model [105] with values of  $e_{21}=2$  (Model 1a) and  $e_{21} < 1$  (Model 1b) to that of the proposed model which includes both a reduced  $e_{21}$  and the

near-field coupling term. Each model variant was optimized to minimize the root-mean-square error and no further improvements were possible through continued variation of the parameter set. As Fig. 5b shows, both Models Ia and Ib with optimized parameters oscillate about the experimental data trend and have peak error magnitudes which substantially deviate outside of the bounds of experimental uncertainty. For reference, the optical parameters we extract from all three model variants are tabulated in Table 1.

Table 7.1: Comparison of extracted parameter values for the three model variations plotted in Fig. 7.5.

Model	$e_{21}$	$\sigma$ ( $\times 10^{-16} \text{ cm}^2$ )	$\tau_{A2}$ (ns)	$\tau_1$ (ns)	$\Gamma$ ( $\times 10^{-29} \text{ cm}^6/\text{s}$ )
Ia	2.0	3.7	49	85	NA
Ib	0.5	1.4	1.5	122	NA
this work	0.5	3.8	5.9	37	2.9

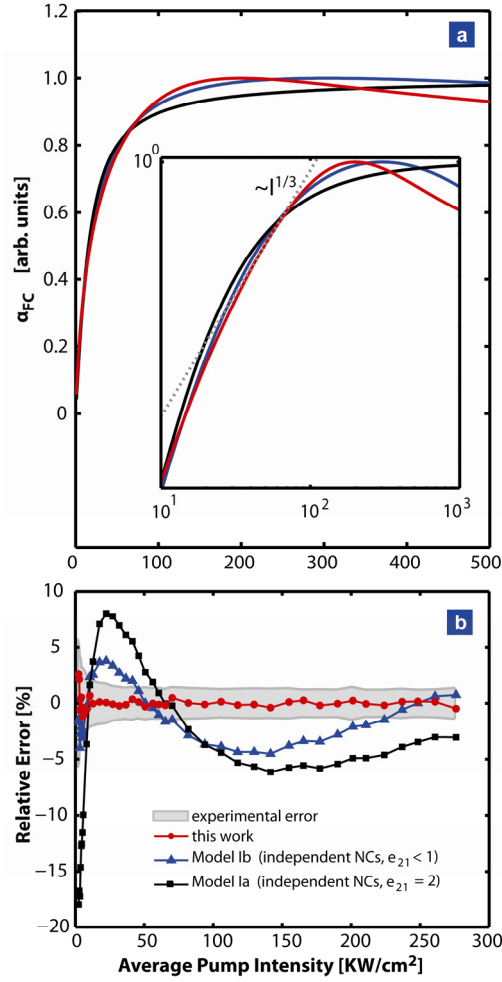


Figure 7.9: (a) Comparison of the three model variants including the independent (i.e. noninteracting) NC model [105] with values of  $e_{21} = 2$  and  $e_{21} < 1$  (black and blue solid curves respectively) and the proposed model which includes near-field interaction between  $N_1$  states (red solid curve). Each model variant is optimal with respect to fitting the data. The curves in the inset are plotted on a logarithmic scale zoomed into the region where the  $1/3$  power law region is observed. (b) Relative error for the three model variants compared to the experimental uncertainty (shaded region).

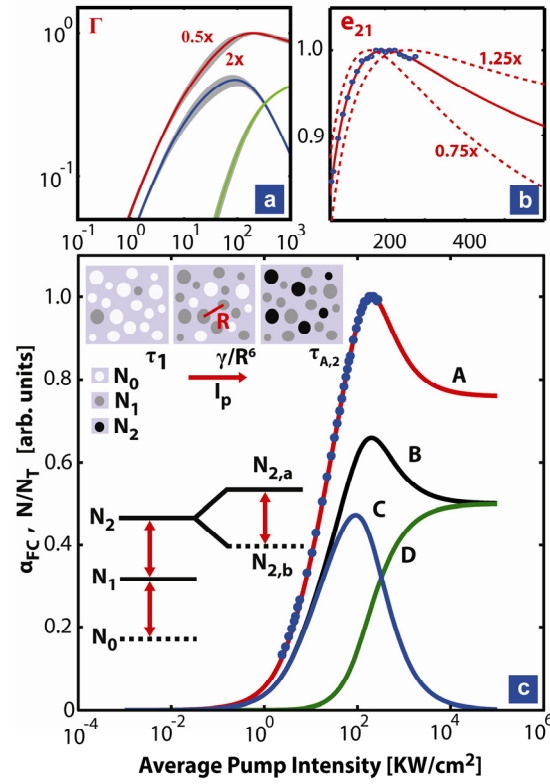


Figure 7.10: (a, b) Visualization of model sensitivity to  $\Gamma$  and  $e_{21}$  (c) Data Set B (solid circles),  $N_l + e_{21}N_2$  (A) scaled to unity and (B) normalized by  $N_T$ , (C)  $N_l/N_T$ , and (D)  $e_{21}N_2/N_T$ . The sequence of pictures in (c) depicts the dominant nonradiative recombination channels vs. pump intensity ( $I_p$ ) and the solid and dashed lines in the manifold indicate active and inactive FCA states respectively.

#### 7.4.4 Monte Carlo Analysis

To gain further insight into the near-field coupling physics, we rigorously solved Eqs. 1 and 3 using Monte Carlo (MC) methods under steady-state pump conditions for a NC population having a uniform probability density of  $2 \times 10^{18} \text{ cm}^{-3}$ . The near-field coupling between two NCs in the  $N_l$  state located a distance  $R_1$  apart was modeled using a term of the form  $\tau^{-1} \sim \gamma R_1^{-6}$  where  $\gamma$  was the only adjustable material parameter. The physical domain of a typical Monte Carlo calculation is shown below in Figure X for two different cases. The first case is for a bulk film, assumed to be infinite in extent in all directions. The second is a finite thickness film, which we use to calculate the dynamics of ultra-thin (10 nm) quasi-2D films.

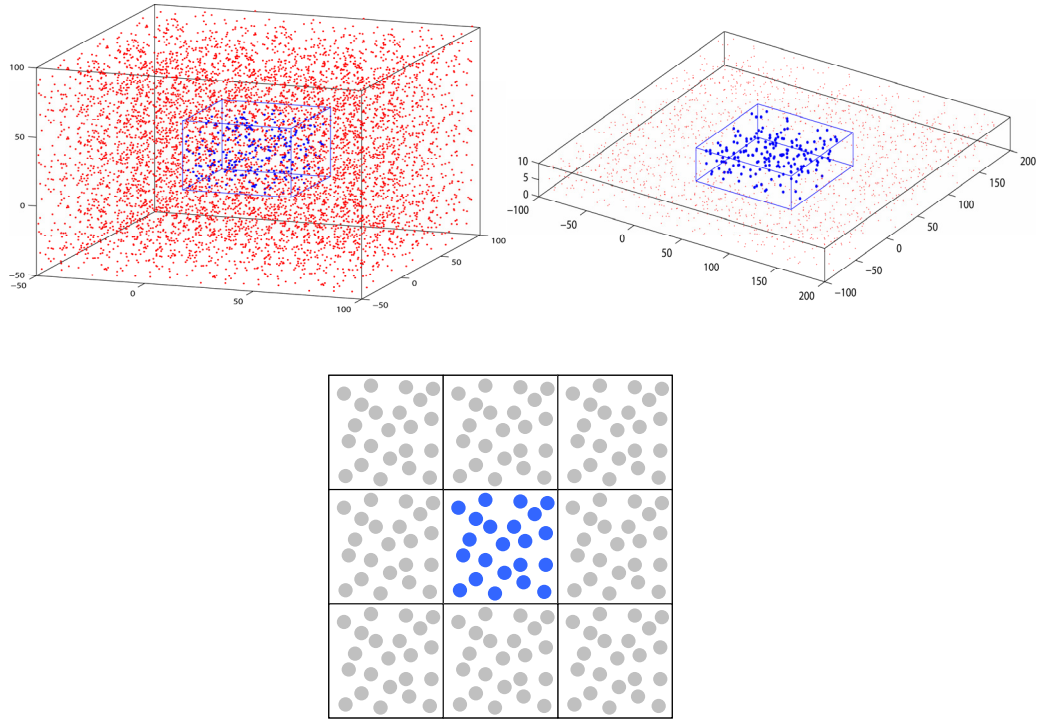


Figure 7.11: The physical domain of a typical Monte Carlo calculation for (a) the case a thick ( $> 100$  nm) film considered to be infinite in all directions and (b) a film with finite thickness, which we use to calculate the dynamics of ultra-thin (10 nm) quasi-2D films. (c) A top down schematic cross-sectional view of either 3D or quasi-2D domains. The middle cell is generated by randomly positioning the nanocrystals according to the specified average density. Then this cell is repeated to impose cyclic boundary conditions in directions representing infinite sample extent. The period of the cell is typically on the order of 100 nm or larger in order to ensure that any particular nanocrystal does not see a mirror image of itself or of its neighbors.

We calibrated the value of  $\gamma$  by matching the results of MC simulations to the results of the steady-state analytical model (Eqs. 2 and 3) for a given value of  $\Gamma$ . To assist in the calibration of  $\gamma$  and to demonstrate the physical equivalence of the two forms of the near-field coupling terms (i.e.  $\gamma \cdot R_1^{-6} \leftrightarrow \Gamma \cdot N_1^2$ ) under steady-state conditions, we selected material parameters which forced the near-field term to dominate the recombination physics over an extended range of pump intensities. A

comparison of the MC calibration simulations to the analytical model with the skewed parameter set are shown in Fig. 6a where the 1/3 power-law region arising from the near-field interaction term is clearly visible over four orders of magnitude. A value of  $\gamma = \Gamma/110$  was found to accurately reproduce the analytical results. The scaling factor (1/110) is entirely accounted for by integrating the near-field contribution of all neighboring NCs in the  $N_I$  state acting on a randomly selected NC also in the  $N_I$  state. After calibrating the value of  $\gamma$ , we performed MC simulations using the optimal parameter set values which provided the best fit to the data using the analytical model. The results of these simulations are shown in Fig. 6b along with the analytical model which was already shown to accurately fit the experimental data. Through the MC simulations, we are able to extract the effective pair interaction strength between two NCs in the  $N_I$  state to be  $\gamma = \Gamma/110 = (2.55 \pm 1.56) \times 10^{-31} \text{ cm}^6/\text{s}$ . This value is nearly identical to the bulk Auger coefficient for silicon [109, 118] and supports the conclusion that the microscopic near-field mechanism is Auger recombination involving two EHPs located in separate NCs.

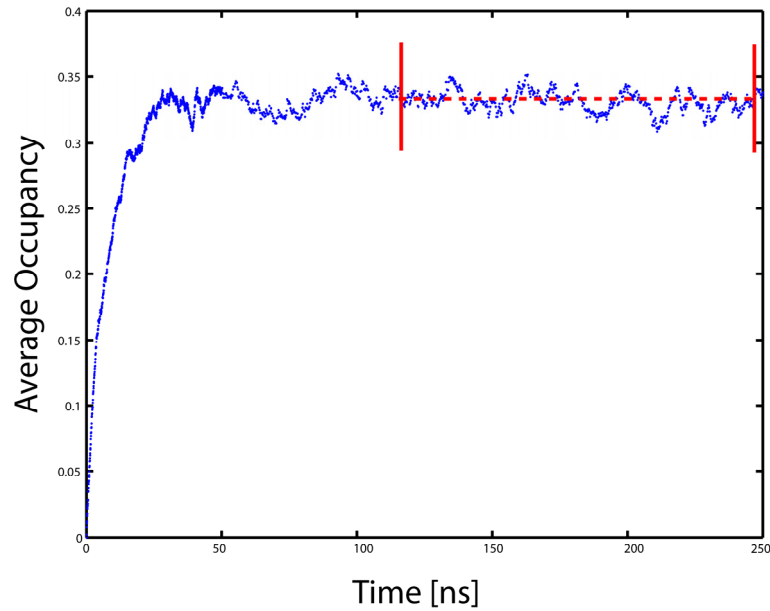


Figure 7.12: Typical transient Monte Carlo simulation output showing average nanocrystal occupancy vs. time in response to a step function pump intensity input. The red highlighted region of the curve is the portion of the curve which is used to calculate the steady-state parameters as a function of pump intensity.

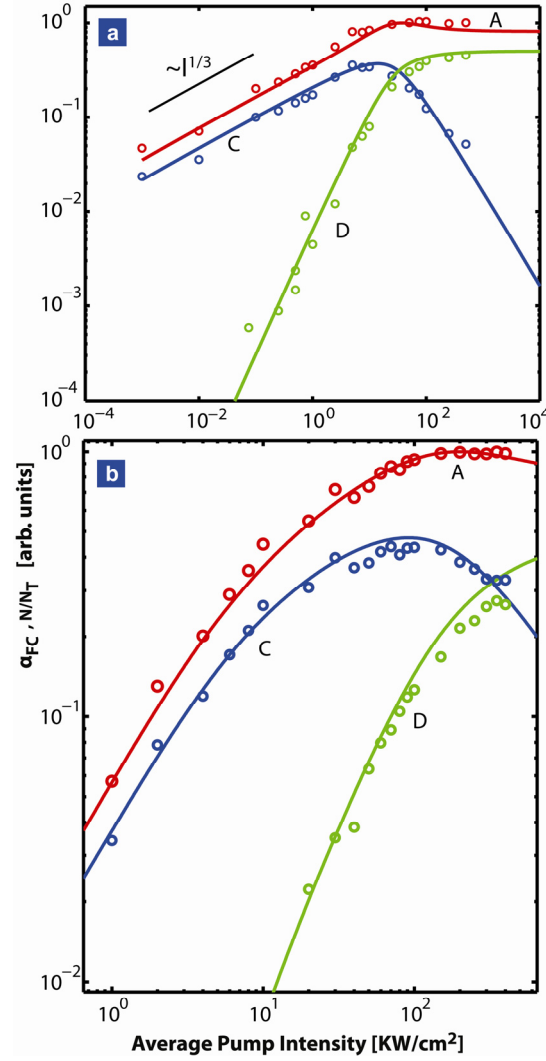


Figure 7.13: Results of MC simulations (open circles) which treat the near-field coupling between NCs in the  $N_1$  state by a term of the form  $\gamma R_1^{-6}$  where  $R_1$  is the distance between the two interacting NCs. The MC results are compared to the steady-state analytical model (Eqs. 2 and 3) where the near-field interaction is described by  $\Gamma \cdot N_1^2$  (solid curves labeled A, C, & D as defined in Fig. 4). (a) Results of MC simulation calibration performed to fine-tune the ratio  $\gamma/\Gamma$  and also to demonstrate the physical equivalence of  $\gamma \cdot R_1^{-6}$  and  $\Gamma \cdot N_1^2$ . To assist the calibration process, the material parameters were intentionally skewed ( $\sigma = 3.8 \times 10^{-16} \text{ cm}^2$ ;  $\tau_1 = 380 \text{ } \mu\text{s}$ ;  $\tau_{A,2} = 59 \text{ ns}$ ;  $\Gamma = 2.3 \times 10^{-29} \text{ cm}^6 \text{ s}^{-1}$ ;  $e_{21} = 0.5$ ) in order to force the near-field coupling to dominate the recombination physics, and thereby exaggerate the  $1/3$  power-law region over an extended range of pump intensities. From the calibration, we find  $\gamma/\Gamma = 1/110$ . (b) MC simulation results using the calibrated value of  $\gamma = \Gamma/110$  and the optimal parameter set described in the text ( $\sigma = 3.8 \times 10^{-16} \text{ cm}^2$ ;  $\tau_1 = 38 \text{ ns}$ ;  $\tau_{A,2} = 5.9 \text{ ns}$ ;  $\Gamma = 2.9 \times 10^{-29} \text{ cm}^6 \text{ s}^{-1}$ ; and  $e_{21} = 0.5$ ). The analytical curve is the same as in Figs 7.8, 7.9, and 7.10 and accurately reproduces the experimental data.

### 7.4.5 Scaling of Long-range Energy Transfer in Low-Dimensional Systems

Despite the potential for ambiguity of the 1/3 law behavior in 3-D, we find that in lower dimensions, some ambiguity is removed by the uniqueness of the power-law behavior. In analogous fashion to the above treatment in 3-D, we find for the 2-D and 1-D cases, the recombination rate for a selected EHP are

$$\tau_S^{-1} = 2 \int_{R_{\min}}^{\infty} \frac{\gamma N_1^{(S)} 2\pi R dR}{R^6} = \frac{\pi \gamma N_1^{(S)3}}{\alpha_S^4} \quad (7.12)$$

$$\tau_L^{-1} = 4 \int_{R_{\min}}^{\infty} \frac{\gamma N_1^{(L)} dR}{R^6} = \frac{4 \gamma N_1^{(L)6}}{5 \alpha_L^5} \quad (7.13)$$

where  $R_{\min} = \alpha_S N_1^{-1/2}$  and  $R_{\min} = \alpha_L N_1^{-1}$  for the 2-D and 1-D cases respectively. Obviously, the definition of  $N_1$  has necessarily changed to account for the dimensionality of the system, e.g. in 2-D,  $N_1^{(S)}$  represents the number of singly excited NCs per unit area. At steady-state, the limiting forms for 2-D and 1-D are then easily shown to be  $N_1 \sim G^{1/4}$  and  $N_1 \sim G^{1/7}$  respectively following the development of (7.9).

It is now instructive to examine the rate equation analysis in non-dimensional form for comparison. We have

$$\hat{N}_1^{1+6/p} \Gamma_{V,S,L} \tau_1 + [\phi \sigma \tau_1 + \phi^2 \sigma^2 \tau_{A,2} \tau_1 + 1] \hat{N}_1 - \phi \sigma \tau_1 = 0 \quad (7.14)$$

where  $p$  is the dimensionality of the system, and  $\hat{N}_1$  is the fractional occupancy of NCs in the single EHP excited state as defined for each dimension as

$$\frac{N_1^{(V)}}{N_T^{(V)}}, \frac{N_1^{(S)}}{N_T^{(S)}}, \text{ or } \frac{N_1^{(L)}}{N_T^{(L)}} \quad (7.15)$$

In an experiment, the fractional occupancy is measured indirectly as a measure of the total excited carrier losses. In rewriting the rate equations in this non-dimensional form, we are able to compare the extracted dipole-dipole coupling strength parameters for the various dimensions. We obtain the following relations:

$$\frac{\Gamma_S}{\Gamma_V} = \frac{3\alpha_V^3}{8\pi\gamma N_T^{(V)2}} \times \frac{\pi\gamma N_T^{(S)3}}{\alpha_S^4} = \frac{3}{8} \left( \frac{\alpha_V^3}{\alpha_S^4} \right) \approx \frac{9}{8} \quad (7.16)$$

$$\frac{\Gamma_L}{\Gamma_V} = \frac{3\alpha_V^3}{8\pi\gamma N_T^{(V)2}} \times \frac{4\gamma N_T^{(S)6}}{5\alpha_L^5} = \frac{3}{10\pi} \left( \frac{\alpha_V^3}{\alpha_L^5} \right) \approx \frac{9}{10} \quad (7.17)$$

where we have used the fact that  $N_T^{(S)} = \left(N_T^{(V)}\right)^{2/3}$  and  $N_T^{(L)} = \left(N_T^{(V)}\right)^{1/3}$ . Table 7.2 summarizes the key results of this section regarding the scaling of the long-range Coulombic dipole-dipole coupling between excited states of neighboring NCs.

Table 7.2: Summary of the laws of scaling for dipole-dipole coupling between excited states in systems of Poisson distributed optical centers.

$p$	$N_1^{(p)} \sim G^q$	$\frac{\Gamma_{(p)}}{\gamma}$	$\frac{\Gamma_{(p)}}{\Gamma_V}$	$\alpha_p$
3 (V)	$q=1/3$	$\frac{8\pi}{3\alpha_V^3} \approx 110$	1	$\alpha_V = 0.46$
2 (S)	$q=1/4$	$\frac{\pi}{\alpha_S^4}$	$\frac{3}{8} \left( \frac{\alpha_V^3}{\alpha_S^4} \right) \approx \frac{9}{8}$	$\alpha_S = 0.42$
1 (L)	$q=1/7$	$\frac{4}{5\alpha_L^5}$	$\frac{3}{10\pi} \left( \frac{\alpha_V^3}{\alpha_L^5} \right) \approx \frac{9}{10}$	$\alpha_L = 0.39$

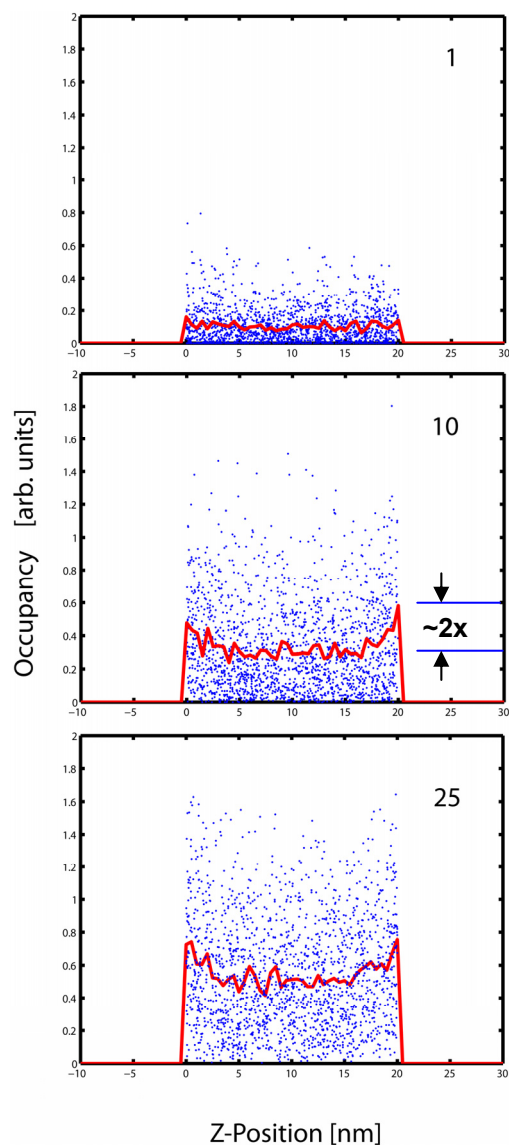


Figure 7.14: Steady state Monte Carlo calculations of a quasi-2D film of thickness 20 nm. The scattered blue dots in each figure represents the average occupancy at a particular location for single simulation. The red curve is an ensemble average of all simulation runs. The sequence of figures represents pump intensities of 1, 10, and 25  $\text{KW}/\text{cm}^2$ . As the pump intensity increases, so does the overall occupancy of the nanocrystals. However, near the edges of the film, the nanocrystals have higher occupancy on average due to the reduced neighbor interaction. At 10  $\text{KW}/\text{cm}^2$ , the ratio of the occupancy at the edge to the center is approximately 2:1 and decrease at higher pump intensity due to higher order Auger processes limiting the occupancy.

## 7.5 Experimental Evidence of Suppressed Long-range Energy Transfer in Quasi-2D Systems

In this section, we discuss additional experimental evidence of the d-d interaction between excited states in silicon nanocrystal – silicon oxide matrices by showing that the predicted behavior of the loss rate vs. pump intensity scales as theory predicts as the film thickness is reduced to the order of the characteristic length of the d-d interaction ( $\sim 10$  nm). Figure 7.15 shows steady-state differential transmission data vs. pump intensity data taken on three different samples of silicon nanocrystals formed inside of a silicon dioxide host.

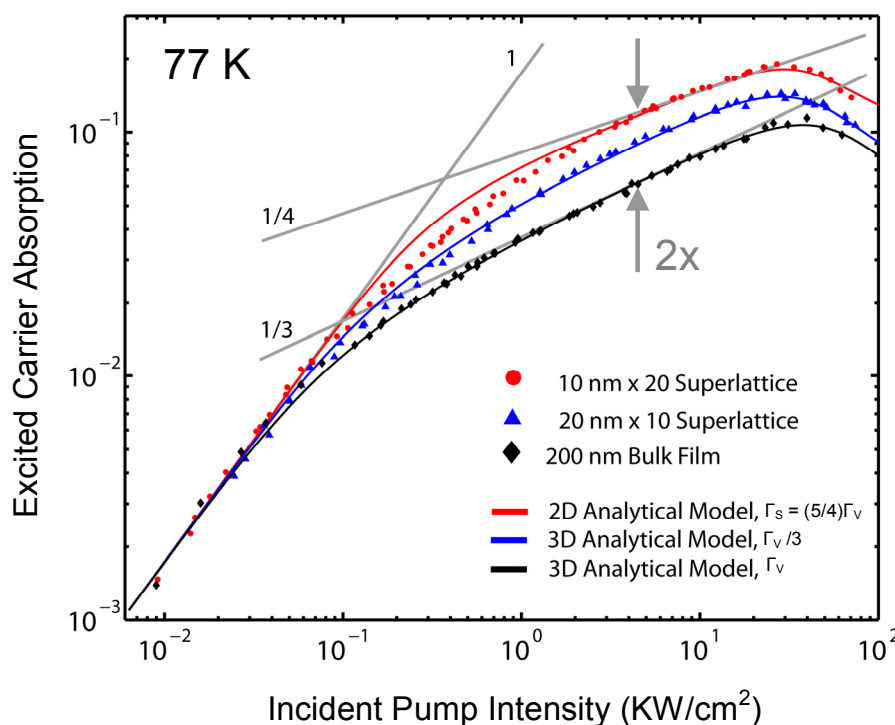


Figure 7.15: Excited carrier losses vs. pump intensity for three different nc-Si/SiO<sub>2</sub> samples fabricated using the e-beam evaporation technique as described in Section 7.2. The samples were prepared by Dr. Aaron Hryciw of the Brongersma Lab.

The samples were prepared by e-beam evaporation of SiO onto fused silica substrates followed by a high-temperature anneal at 1100 °C for 1 hour as described in Section 7.2. Sample ‘Th’ (black diamonds) is a 230 nm thick (200 nm targeted) nc-Si/SiO<sub>2</sub> film. Sample ‘SL1’ (blue triangles) is a superlattice structure consisting of 10 periods

of alternating 20 (23 actual) nm nc-Si/SiO<sub>2</sub> and 20 (23 nm actual) nm of SiO<sub>2</sub> buffer layer. The final sample, referred to as ‘SL2’, is also a superlattice structure but consisting of 20 periods of 10 nm (actual unknown at this time) nc-Si/SiO<sub>2</sub> and 20 nm (actual unknown at this time) of SiO<sub>2</sub> buffer layer. The idea behind the sample design was to reduce the individual layer thicknesses while maintaining a consistent interaction volume of active material for all three samples.

The key results are as follows. At low pump intensity, all three samples exhibit essentially identical pump-induced loss rates vs. pump intensity, i.e. the linear slopes at low pump intensity are identical to within experimental uncertainty. However, at high pump intensities,  $> 0.1 \text{ KW/cm}^2$ , the curves deviate significantly with the thinner films exhibiting higher loss rates than the bulk film in proper order, i.e. ‘SL2’  $>$  ‘SL1’  $>$  ‘Th’. A general conclusion which can be derived from this result is that the effects of finite film thickness only has a significant impact on the carrier dynamics when the nanocrystal occupancy is appreciable. This is conclusive evidence that there is indeed electron interaction between excited nanocrystals, independent of the exact mechanism enabling the interaction. Now, we discuss the evidence supporting the dipole-dipole interaction mechanism.

The solid curves represent the best-fit curves using the both the 3-D (‘Th’ and ‘SL1’ samples) and 2-D (‘SL2’ sample) analytical models. The 3-D analytical model fits both the Th and ‘SL1’ samples extremely well and the 2-D model fits the ‘SL2’ reasonably well with the exception of the intermediate region between the linear and 1/4 law regions. Most importantly, the limiting form at high pump intensity is an exact 1/4 law as predicted by the 2-D theory. The discrepancy in the intermediate transition region is the subject of an ongoing investigation using Monte Carlo analysis. It is hypothesized that the discrepancy is due to the finite thickness effects of the film which are not captured in the 2-D model. Despite this discrepancy, we find that the analytical model shows excellent agreement in terms of the parameter selection between the ‘Th’ and ‘SL2’ curves. By first fitting the extremes of the ‘SL2’ data with the 2-D model, i.e. by neglecting the intermediate regime and then using these same parameters in the 3-D model, the ‘Th’ curve is perfectly reproduced with the exception

that the  $\Gamma_V$  needs to be reduced by a factor of 1.25 or 5/4. Recalling Eq. 7.16, it was shown that in going from 2D to 3D, the ratio  $\Gamma_S / \Gamma_V$  was approximately 9/8, almost identical to the value we extract from experiment. An additional evidence supporting the dipole-dipole interaction is that the maximum ratio of occupancies between the quasi-2D film and the bulk film approaches 2, which has been a consistent trend in the Monte Carlo results as demonstrated in Figure 7.14.

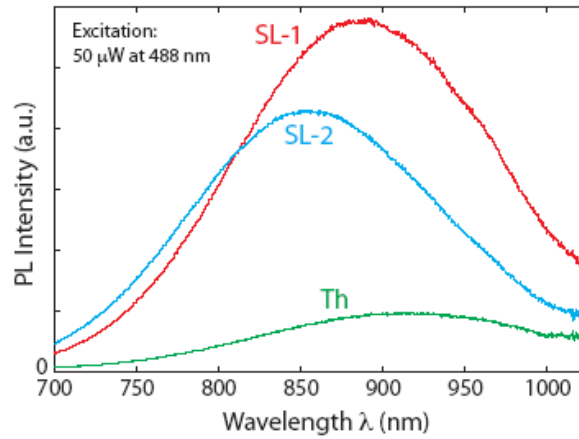


Figure 7.16: Steady-state photoluminescence spectroscopy measurements of the three samples: 200 nm bulk film (Th), 20 nm x 10 superlattice (SL1) and 10 nm x 20 superlattice film (SL2). The data show an increasing blue shift with decreasing film thickness as well as a maximum PL signal by the 20 nm SL sample. The PL data was taken by Dr. Aaron Hryciw in the Brongersma Lab (Stanford University).

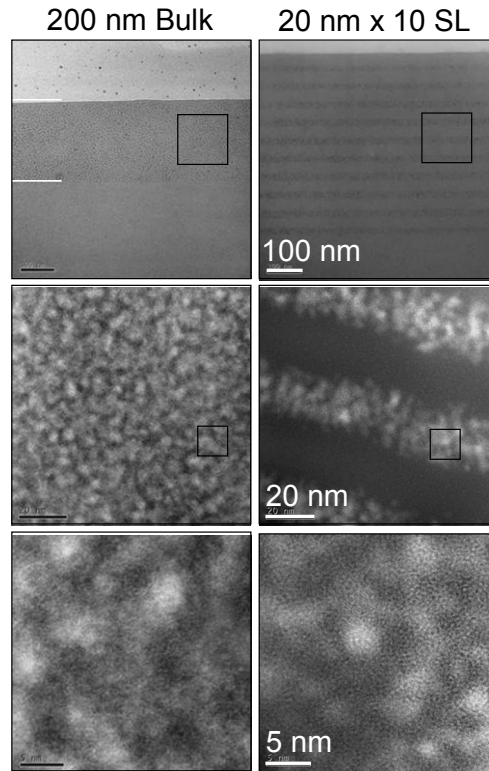


Figure 7.17: High resolution and energy filtered (bottom four images) TEMs of the thick and SL1 samples shown side-by-side for direct comparison.

## 7.6 Conclusions

Experimental evidence of inter-nanocrystal coupling between optically excited electrons in dense silicon nanocrystal systems embedded in amorphous silicon oxide films was presented. A theoretical model to explain the observation of enhanced nonradiative recombination rates at high pump intensities was proposed. The theory asserts that the dominant interaction between excited carriers is a dipole-dipole interaction resulting from long-range Coulombic forces. The strength of this coupling increases inversely to the sixth power of separation which I have shown to scale as the square of the density of excited nanocrystals in 3-D or as the cube of the density in 2-D, or as the sixth power of the density in 1-D. I presented experimental evidence

supporting this theory showing that as the film thickness reduces, the transition theory is able to predict the trend in going from 3-D to quasi-2D. Although, additional experiments are encouraged to continue test this theory, the evidence collected thus far is overwhelmingly in favor of the proposed theory.

The scaling of the Forster-Auger process to low-dimensions has not been treated to date to my knowledge. Therefore, this subject provides an interesting field of new research. One of the interesting outcomes of this process is the notion of non-local heat generation and transfer. Heat can be generated approximately 10 nm away from the source electron. It is interesting to think about the possibility of controlling heat generation on the nanoscale in order for selective growth of nanostructures.

## Chapter 8

### Conclusion

And it turns out that all of the information that man has carefully accumulated in all the books in the world can be written in this form in a cube of material one two-hundredth of an inch wide-which is the barest piece of dust that can be made out by the human eye. So there is *plenty* of room at the bottom!

– R. P. Feynman

Along with the creation, transport, or storage of information, energy must be converted. That is, entropy must be created and heat dissipated in accordance with fundamental thermodynamic principles. Considering the current trajectory of information technology, it is evident that that every atom within a few nanometers of a substrate surface will count [120] and every  $k_B T$  of energy will need to be accounted for! Therefore, as we seek to reduce the volume of material used to store and process data down to atomic dimensions, we must simultaneously advance our capability to model, experimentally measure, and ultimately manipulate the flow and conversion of energy on similar length scales.

This dissertation contributed to the fundamental understanding of energy conversion and transport physics in the silicon nanostructures likely to be the building blocks of future electronic and photonic information systems. In the following section, I will summarize the key contributions of this work as well as the conclusion which resulted from this work. In the final section, I will describe future research directions which build upon the insights developed in this dissertation.

## 8.1 Concluding Remarks

### 8.1.1 Nonequilibrium Energy Conversion in Nanotransistors

In the first half of this dissertation, we examined the physics of energy conversion between electrons and phonons in nanoscale silicon transistors at conditions far from thermal equilibrium. A distinguishing factor of this work is that we examine the case when both the electron and phonon subsystems are significantly perturbed from equilibrium such that both particle types exhibit quasi-ballistic transport behavior. This work was motivated by a long-held concern, initially brought to attention by Artaki and Price [31] in the late 1980s, that hot phonon effects may eventually thwart the continued scaling of silicon transistor technology. If the rate of energy conversion between the hot electrons and the relatively stationary optical phonons (OPs) is sufficiently fast, and subsequent decay of OPs into the long-wavelength, high-velocity, acoustic phonons (APs) responsible for heat conduction is sufficiently slow, an energy conversion bottleneck can exist leading to a growing population of hot OPs in the drain junction. If the OP occupancy  $n_{op}$  begins to exceed unity, the rate of OP absorption by the electrons ( $\tau^{-1} \sim n_{op}$ ) becomes comparable to OP emission ( $\tau^{-1} \sim n_{op} + 1$ ) and the electrons may begin to efficiently absorb OP energy. Consequently, the electron momentum relaxation time, which directly impacts the mobility, will be reduced while simultaneously extending the energy relaxation time in the drain junction. A reduction in device current for a given bias voltage is expected but in extreme cases, where OP occupancy greatly exceeds unity, negative differential conductance instability like that observed recently in suspended metallic carbon nanotube devices [72], may occur. Accurate determination of the conditions required to reach such a condition in nanometer silicon devices is therefore essential but inherently difficult to quantify.

To properly simulate the negative feedback effects of hot phonons, the electron and phonon subsystems must be fully-coupled and the simulation must be self-consistent if the perturbation is substantial. It is not sufficient to examine hot phonon

effects by simply including electron-phonon scattering (at any level of detail) in an electron transport model. Accurately simulating this scenario is inherently challenging because it requires individually tracking a wide range of electron and phonon eigen states and their interactions in space and time. In order to make the problem tractable, some level of approximations are needed. In previous numerical studies gross simplifying assumptions were made regarding the electron-phonon scattering mechanisms and the phonon dispersion. In particular, previous coupled electron-phonon or electro-thermal studies did not adequately model the nonequilibrium phonon distributions existing near ( $< 20$  nm) the peak energy density region of the device or include the mitigating effects of finite optical phonon dispersion and reductions in OP lifetimes under high power density conditions.

A key contribution of this work was to show that the effects of hot OPs on silicon nanotransistor performance were substantially overestimated in previous studies as a result of the limitations listed above. To examine the hot phonon effect, this work developed fully-coupled and self-consistent, device-level, Boltzmann transport (BT) simulations which efficiently coupled the electron and phonon subsystems while treating electron and phonon dispersion with a sufficient and comparable level of detail. This fully-coupled approach is the most rigorous coupled electron-phonon transport simulation implemented for semiconductor devices to date and greatly expands upon previous work by Pop [53] and Sinha [121] who developed electron Monte Carlo and split-flux Boltzmann models respectively to treat independent electron and phonon transport. Efficient coupling between the electrons and phonons was achieved through the introduction of a new stochastic nonequilibrium phonon occupation number rejection algorithm implemented within a well-established electron Monte Carlo framework. The electron-phonon scattering rates are dynamically adjusted within each device computational cell volume and the rejection algorithm becomes more efficient as the level of deviation from equilibrium increases.

In addition to the development of coupled simulations, the lifetimes and two-phonon final states spectrum for OPs determined to have high occupancy during transistor switching were computed using a full-band, anharmonic perturbation

approach. The computed nominal (equilibrium) OP lifetimes were then hardwired into the coupled electron-phonon transport simulations to provide an increased level of accuracy. The silicon phonon bandstructure was modeled using a six-parameter valence force-field model which accurately reproduces the 3D anisotropic phonon dispersion.

Through these self-consistent coupled device simulations, we determined that the reduction in drive current predicted by previous studies was overestimated by more than an order of magnitude and that energy densities greatly exceeding  $100 \text{ TW/cm}^3$  will be required in order to reach an unstable operating condition. To determine the likelihood of reaching this critical power density, the drain junction volumetric power density trend was computed for various transistor device designs using the current ITRS and was found to have a channel length dependence scaling as  $L^{-1.7}$ . We find that the critical power densities will not be reached within the existing roadmap timeline and consequently hot phonon effects should not contribute significantly to device performance. In addition to this important conclusion, we find that, contrary to earlier estimates [122], the effect of hot phonons on leakage current is negligible due to the rapid quenching ( $\sim 5 \text{ ps}$ ) of the nonequilibrium condition. Furthermore, it is found that self-heating at the source-channel interface during switching increases current injection in the quasi-ballistic devices and partially offsets small mobility reductions in the drain due to excess phonon scattering.

Although the coupled model was developed for silicon devices, the approach is sufficiently general to be easily applied to other semiconductor systems. Finally, it should be noted that this work completes more than a decade of theoretical work performed at Stanford by multiple generations of graduate students who have addressed this complex topic with increasing degree of accuracy.

### 8.1.2 Energy Conversion in Silicon Nanophotonic Materials

The second half of this dissertation examined the physics of carrier dynamics and the associated energy conversion in dense systems of optically pumped silicon nanocrystals (NCs) embedded in amorphous dielectric films. These materials exhibit stable, room-temperature visible and near-infrared luminescence as well as the ability to sensitize codoped  $\text{Er}^{3+}$  ions making them promising materials in the development of inexpensive, CMOS-compatible short-range coherent optical sources. Due to inherent experimental ambiguity, previous pump-probe and luminescent studies resulted in the proposal of multiple, seemingly contradicting sets of carrier rate equations to describe the NRCR physics in these materials. Specifically, bulk-like, i.e. continuous charge distribution models, as well as discrete non-interacting nanocrystal models have been employed to describe previous results.

A key contribution of this work was the development of a custom, two-color optical pump-probe system operating at 532 nm (a 2<sup>nd</sup> harmonic of the fundamental 1064 nm laser line) and 1064 nm and having sub-10 picosecond temporal resolution and a detection sensitivity within 3dB of the shot noise limit set by the probe beam. This system was used to measure the carrier dynamics of thin silicon NC films down to isolated films of 20 nm thick and superlattice stacks with active layers down to 10 nm through indirect probing of the excited carrier intraband absorption. The unique dependence of the excited carrier losses vs. probe delay and pump intensity revealed enhanced NRCR at high NC occupancies, which could not be reproduced using existing models. Specifically, saturation at higher excitation levels were in contradiction to proposed bulk continuum models.

A new coupled carrier dynamics theory was developed to model the optically-pumped NCs embedded in amorphous silicon oxide thin films. The new theory explained an enhanced rate of nonradiative recombination observed at high pump intensities as the result of long-range Coulombic dipole-dipole coupling between excited electron states in neighboring NCs. This coupling efficiency of this process is shown to increase inversely with the sixth power of the average separation distance

between excited NCs in similar form as Förster resonant energy transfer (FRET) theory would predict and has a cubic dependence on average excitation level in 3D systems. Although arising from the same origin and having similar properties, the resonant energy transfer physics identified and researched in this work differs from the FRET process in that the coupling is between two excited states in different NCs whereas in FRET, coupling is between an excited state in the donor and the ground state of the acceptor. In this sense, the process is also equivalent to Auger recombination at a distance. In 3D, the scaling of the d-d coupling is shown to scale as the cube of the average excitation level of the NC system and can therefore be confused with the standard bulk Auger scaling. An analytical theory of the scaling of this effect is developed and the stated ambiguity in 3D is shown to disappear in lower spatial dimensions, e.g. ultra-thin films (2D) or ridges (1D) containing NCs, with the d-d coupling scaling as the fourth and seventh powers of the average excitation levels for 2D and 1D respectively. The validity of the analytical model was substantiated using rigorous Monte Carlo simulation techniques to solve the coupled dynamics for low-dimensional systems of particles. Furthermore, the 2D model was experimentally validated by measuring excited carrier absorption losses in silicon NC films with critical dimensions down to 10 nm. The 10 nm and 20 nm films were grown in superlattice structures with 30 nm silicon oxide spacers in order to enhance the signal strength and maintain an approximately constant active device material thickness in all samples. The transition from 3D to 2D behavior is captured experimentally as a systematic reduction in carrier losses and a characteristic change in power-law behavior from  $1/3$  to  $1/4$ . An important consequence of the reduced NC interaction is that the rate of nonradiative recombination is reduced by as much as factor of 2 by going to ultra-thin films. This result has the potential to impact future nanoparticle optical device designs.

The development of the coupled excited d-d interaction model is not limited to silicon systems and should be relevant to other semiconductor quantum dot or nanoparticle composite materials.

## 8.2 Future Work: Energy Conversion in Interconnected Nanosystems

The final portions of this dissertation focused on the discovery of strong dipole-dipole coupling between excited states in systems of densely packed silicon NCs as well as the characteristic signature of this coupling in systems having lower spatial dimensions. There has been a considerable amount of work concerning dipole-dipole interaction between an excited state of a donor nanoparticle (NP) or quantum dot (QD) and the ground state of a nearby acceptor NP/QD. However, surprisingly, no prior work can be found regarding the dipole-dipole coupling between excited states in NP/QD pairs despite the discovery of this effect six decades ago in impurity centers [112] and again much later in the rare-earth doped glass fibers, c.f. Refs. [123, 124], heavily utilized in long-haul telecommunications systems. Although the work here focused on systems of silicon NCs, the physical processes described are not unique to this material system and should play an equally important role in a wide variety of densely packed ( $> 10^{16} \text{ cm}^{-3}$ ) quantum dot systems.

As mentioned, there has been a considerable amount of work examining the effect of d-d coupling between excited states in rare-earth doped glass fibers. Referred to as “cooperative up-conversion”, the interaction of two closely spaced erbium ions ( $\text{Er}^{3+}$ ) both initially residing in the first excited state ( $I^{13/2}$ ) may result in the nonradiative de-excitation of one ion ( $I^{13/2} \rightarrow I^{15/2}$ ), the donor ion, and transfer of energy to the acceptor ion which ends in the radiative emission of a photon at twice the initial energy. The acceptor ion then makes a transition to the upper  $I^{11/2}$  state via an intermediate vibrational coupled transition to the  $I^{9/2}$  state before finally relaxing to the ground state ( $I^{15/2}$ ) where it subsequently emits a photon of energy 1.27 eV (980 nm). Having been initially excited by photons of energy 0.838 eV (1480 nm), this d-d coupling between neighboring ions results in an up-conversion in photon energy, hence the attributed name. It is interesting to point out that, in three dimensions, the rate of de-excitation of an any given  $\text{Er}^{3+}$  ion is commonly taken to be proportional to the square of the density of excited atoms rather than the cube [124]. For a Poisson distributed population, we saw in Chapter 7 that such a process should scale as the cube of the density. However, a square-law density dependence may result in the case

that the ions are not Poisson distributed but are rather distributed in a manner which favors the formation of ionic pairs. In such a case, the average distance between excited neighbors has little effect on the deexcitation rate and is effectively removed from the integral over all states. Studies which control the density of particles effectively could be used to examine this subtle effect.

There exists a large body of work addressing the physics of FRET and its scaling to lower dimensional systems of molecules, defect centers, and more recently, quantum dot (QD) systems. A large portion of this work arose from studies performed on biological systems where spectroscopic techniques based on this process have been used to great advantage in probing biochemical reactions and transport of energy across nanometer thick lipid bilayers. A resurgence in research activity in this area in recent years has been spurred by the commercialization of semiconductor QDs in solution, e.g. II-VI core/shell heterostructures, which have increased chemical stability and are relatively robust to extended optical excitation. The size of the QDs, and therefore their bandgap energy, can be controlled with great precision. If one can control the positioning of QDs to nanometer precision, then one can manipulate the flow of charge through the system by employing the FRET effect. For example, an excited electron-hole pair in a particular QD will have a strong tendency to transfer to larger, smaller band gap, QDs. Could an analogous approach be used to control energy flow using the Auger-Förster RET process, perhaps to deposit heat in a desired location?

Monte Carlo studies have been conducted to quantify the rate of energy conversion in low-dimensional systems of nanoparticles (NPs). The results of a study performed on one exemplary nanosystem is shown in Figure 8.1. The system is a dielectric sphere of radius 10 nm which contains Poisson distributed NPs (or equivalently QDs) with a density of  $2 \times 10^{18} \text{ cm}^{-3}$ . The physical parameters of the system are identical to those used in modeling silicon NCs in silicon oxide as described in Chapter 7.

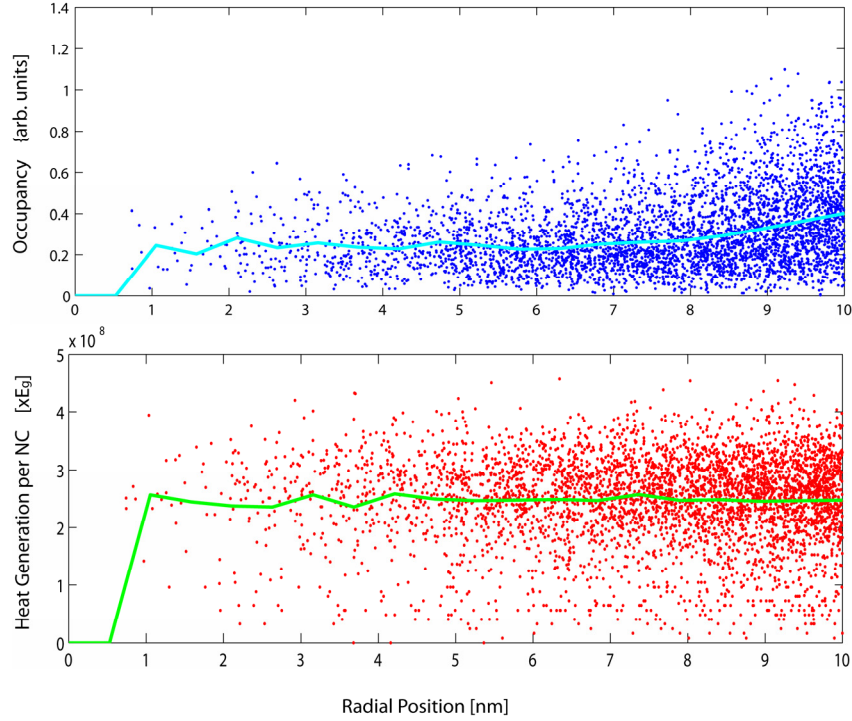


Figure 8.1: Steady-state Monte Carlo simulation of an optically pumped (CW), nanoparticle (NP) system having a spherical geometry with radius 10 nm. The simulation model includes inter-NP dipole-dipole coupling and nonradiative energy exchange between excited electron states as was described in Chapter 7. The top panel shows a scatter plot of NP occupancy, defined by the relation  $N_1/N_T + 2*N_2/N_T + 3*N_3/N_T + \dots$ , as a function of radial position. The solid light-blue line indicates the ensemble average occupancy vs. radial position. As expected, the occupancy tends to increase near the edges due to the reduction in interacting neighbors. The apparent reduction near the center of the sphere is an artifact of the simulation and is not indicative of any physical process. The binning excluded NPs within the inner most 1 nm region. The bottom panel is a scatter plot of the steady-state heat generation rate per second normalized by the bandgap energy of the NP system. The solid green line indicates the ensemble averaged heat generation rate vs. radius. In contrast to the occupancy plot, the heat generation per particle at steady-state conditions is uniform along the radius.

As expected, the occupancy tends to increase near the edges due to the reduction in interacting neighbors. The apparent reduction near the center of the sphere is an artifact of the simulation. The statistical binning excluded NPs within the inner most 1 nm region. The bottom panel is a scatter plot of the steady-state heat generation per second normalized by the bandgap energy of the NP system. The solid green line indicates the ensemble average heat generation rate vs. radius. In contrast to the occupancy distribution, the heat generation per particle at steady-state conditions is uniform throughout the sphere. This arises from the necessity to have energy transfer balance at steady-state. With the assumption that the energy excitation rate from the

optical beam is uniform, the occupancy will adjust itself in order to just balance the rate of dipole-dipole energy transfer.

The model assumes that heat is generated when either an electron-hole pair recombines through the standard nonradiative process or as the result of the long-range interaction with neighboring excited particles via the Forster-Auger mechanism. In each case, the amount of heat generated is taken to be equal to the bandgap of the particle. However, in the former case, the heat is assumed to be generated within the particle where the electron recombined and in the latter case, the heat is assumed to be dissipated in the acceptor particle, as the result of a rapid (assumed immediate for simulation purposes) intraband relaxation. Since there is a distribution of acceptors in which a donor particle can exchange energy with, selecting which acceptor particle receives the energy presents an additional challenge. The choice is made during the simulation by ranking the probabilities according to the strength of the coupling, which scales as  $1/R^6$ . The selection is then made in the standard Monte Carlo approach. The dipole-dipole energy transfer mechanism could potentially contribute to the macroscopic heat transfer rate of low-thermal conductivity materials. One could conceivably control the rate of heat transfer across a membrane through external charge injection or optical pumping.

## References

- [1] H. K. Charles Kittel, *Thermal Physics*, 2nd ed. New York: W. H. Freeman and Company, 1980.
- [2] Q. C. J. D. Meindl, J. A. Davis, "Limits of Silicon Nanoelectronics for Terascale Integration," *Science*, vol. 293, pp. 2044-2049, 2001.
- [3] R. Lake and S. Datta, "Nonequilibrium Green's-function method applied to double-barrier resonant-tunneling diodes," *Physical Review B*, vol. 45, p. 6670, 1992.
- [4] F. C. J. Faist, D. L. Sivco, C. Sirtori, A. L. Hutchinson, A. Y. Cho, "Quantum Cascade Laser," *Science*, vol. 264, pp. 553-556, 1994 1994.
- [5] H. Xuejue, *et al.*, "Sub 50-nm FinFET: PMOS," in *IEEE International Device Meeting Technical Digest* pp. 67-70, 1999.
- [6] R. Chau, *et al.*, "A 50 nm depleted-substrate CMOS transistor (DST)," pp. 29.1.1-29.1.4, 2001.
- [7] B. Doyle, *et al.*, "Tri-Gate fully-depleted CMOS transistors: fabrication, design and layout," in *Symposium on VLSI Technology Digest of Technical Papers*, pp. 133-134, 2003.
- [8] K. E. G. J. A. Rowlette, "Fully Coupled Nonequilibrium Electron-Phonon Transport in Nanometer-Scale Silicon FETs," *IEEE Transactions On Electron Devices*, vol. 55, pp. 220-232, 2008.
- [9] "International Technology Roadmap for Semiconductors (ITRS) 2005 Update " in <http://www.itrs.net/Links/2005ITRS/Home2005.htm>. vol. 2005, 2005.
- [10] R. Chau, *et al.*, "Silicon nano-transistors and breaking the 10 nm physical gate length barrier," in *IEEE Device Research Conference*, pp. 123-126, 2003.
- [11] A. W. J. Muscat, N. M. Harrison, "On the prediction of band gaps from hybrid functional theory," *Chemical Physics Letters*, vol. 342, pp. 397-401, 2001.
- [12] A. Debernardi, S. Baroni, and E. Molinari, "Anharmonic phonon lifetimes in semiconductors from density-functional perturbation theory," *Physical Review Letters*, vol. 75, pp. 1819-22, 1995.
- [13] S. Sinha, P. K. Schelling, S. R. Phillpot, and K. E. Goodson, "Scattering of g - process longitudinal optical phonons at hotspots in silicon," *Journal of Applied Physics*, vol. 97, pp. 023702-9, 2005.
- [14] R. Tubino, L. Piseri, and G. Zerbi, "Lattice Dynamics and Spectroscopic Properties by a Valence Force Potential of Diamondlike Crystals: C, Si, Ge, and Sn," *The Journal of Chemical Physics*, vol. 56, pp. 1022-1039, 1972.
- [15] J. E. B. Wilson, *Journal of Chemical Physics*, vol. 9, 1941.
- [16] J. E. Bright Wilson, J. C. Decius, P. C. Cross, *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*. New York: Dover Publications, 1955.
- [17] L. Piseri and G. Zerbi, "A generalization of GF method to crystal vibrations," *Journal of Molecular Spectroscopy*, vol. 26, pp. 254-261, 1968.

- [18] T. B. Boykin, G. Klimeck, and F. Oyafuso, "Valence band effective-mass expressions in the  $sp(3)d(5)s(*)$  empirical tight-binding model applied to a Si and Ge parametrization," *Physical Review B*, vol. 69, p. 115201, 2004.
- [19] J. C. Slater and G. F. Koster, "Simplified LCAO Method for the Periodic Potential Problem," *Physical Review*, vol. 94, p. 1498, 1954.
- [20] W. A. Harrison, *Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond*, 2nd ed. New York: Dover Publications, Inc. , 1989.
- [21] S.-Y. Ren and W. A. Harrison, "Semiconductor properties based upon universal tight-binding parameters," *Physical Review B*, vol. 23, p. 762, 1981.
- [22] G. Klimeck, R. C. Bowen, T. B. Boykin, and T. A. Cwik, " $sp^3s^*$  tight-binding parameters for transport simulations in compound semiconductors," *Superlattices and Microstructures*, vol. 27, pp. 519-24, 2000.
- [23] Y. M. Niquet, C. Delerue, G. Allan, and M. Lannoo, "Method for tight-binding parametrization: Application to silicon nanostructures," *Physical Review B*, vol. 62, p. 5109, 2000.
- [24] R. A. Cowley, "Anharmonic crystals," *Reports on Progress in Physics*, vol. 31, pp. 123-166, January, 1968 1968.
- [25] M. Asheghi, K. Kurabayashi, R. Kasnavi, and K. E. Goodson, "Thermal conduction in doped single-crystal silicon films," *Journal of Applied Physics*, vol. 91, pp. 5079-88, 2002.
- [26] P. G. Klemens, "The thermal conductivity of dielectric solids at low temperatures (theoretical)," *Proceedings of the Royal Society of London, Series A (Mathematical and Physical Sciences)*, vol. 208, pp. 108-133, 1951.
- [27] J. Callaway, "Model for lattice thermal conductivity at low temperatures," *Physical Review*, vol. 113, pp. 1046-1051, 1959.
- [28] M. G. Holland, "Analysis of lattice thermal conductivity," *Physical Review*, vol. 132, pp. 2461-2471, 1963.
- [29] E. Pop, R. W. Dutton, and K. E. Goodson, "Monte Carlo simulation of Joule heating in bulk and strained silicon," *Applied Physics Letters*, vol. 86, pp. 1-3, 2005.
- [30] In this work, we find  $\sim 2/3$  of the energy is dissipated by optical phonon modes with the remaining energy going into acoustic modes. Full-band MC codes which compute the individual matrix elements using pseudo-potential theory have shown that the situation is reversed with the acoustic modes receiving the  $\sim 2/3$  majority of the thermal energy.
- [31] M. Artaki and P. J. Price, "Hot phonon effects in silicon field-effect transistors," *Journal of Applied Physics*, vol. 65, pp. 1317-20, 1989.
- [32] J. Shah, A. Pinczuk, A. C. Gossard, and W. Wiegmann, "Hot carrier energy loss rates in GaAs quantum wells: large differences between electrons and holes," *Physica B & C*, vol. 134B+C, pp. 174-8, 1985.
- [33] E. Pop, J. A. Rowlette, R. W. Dutton, and K. E. Goodson, "Joule heating under quasi-ballistic transport conditions in bulk and strained silicon devices," in *International Conference on Simulation of Semiconductor Processes & Devices, 1-3 Sept. 2005, Tokyo, Japan*, pp. p.307-10, 2005.

- [34] S. Sinha, E. Pop, R. W. Dutton, and K. E. Goodson, "Non-equilibrium phonon distributions in sub-100 nm silicon transistors," *Journal of Heat Transfer-Transactions of the ASME*, vol. 128, pp. 638-647, 2006.
- [35] The exact strength of the coupling as determined by the deformation potential is controversial. Full-band Monte Carlo codes have yielded values of about a factor of 2 lower than the results used in this work but are still of the same order of magnitude.
- [36] S. Narasimhan and D. Vanderbilt, "Anharmonic self-energies of phonons in silicon," *Physical Review B (Condensed Matter)*, vol. 43, pp. 4541-4, 1991.
- [37] The subscript reminds us that the density of states is for phonon pairs with energies  $E$  and  $E_0 - E$ .
- [38] Summing over the third order matrix elements over all normal modes yields an additional delta function of the form  $q_0 = q_1 + q_2 + G$  where  $G$  is a reciprocal lattice vector. The delta functions appearing in (10) explicitly and implicitly constrain the summation over the B.Z. to include only pairs of phonons which satisfy the energy and momentum conservation relations.
- [39] G. Gilat and L. J. Raubenheimer, "Accurate Numerical Method for Calculating Frequency-Distribution Functions in Solids," *Physical Review*, vol. 144, p. 390, 1966.
- [40] G. Lehmann and M. Taut, "On the numerical calculation of the density of states and related properties," *Physica Status Solidi B*, vol. 54, pp. 469-77, 1972.
- [41] J. Rath and A. J. Freeman, "Generalized magnetic susceptibilities in metals: Application of the analytic tetrahedron linear energy method to Sc," *Physical Review B*, vol. 11, p. 2109, 1975.
- [42] P. B. Allen, "A tetrahedron method for doubly constrained Brillouin zone integrals. Application to silicon optic phonon decay," *Physica Status Solidi B*, vol. 120, pp. 529-38, 1983.
- [43] We note that only for the R-LTO mode would we be able to take advantage of the full symmetry of the crystal to reduce the size of the calculation; for a phonon of arbitrary initial wavevector, integrations must be performed over the entire B.Z. However, in our analysis, we restrict our calculations to LO phonons with initial wavevector falling along the  $\Gamma$ -X direction which allows us to reduce the volume of the integration by a factor of 1/8 of the B.Z. compared to 1/48 for the irreducible wedge.
- [44] P. G. Klemens, "Anharmonic Decay of Optical Phonons," *Physical Review*, vol. 148, p. 845, 1966.
- [45] We made similar calculations for diamond and found that the Klemens channel does correspond to a dominant , also consistent with the results of Ref. [28]
- [46] J. Menendez and M. Cardona, "Temperature dependence of the first-order Raman scattering by phonons in Si, Ge, and alpha -Sn: anharmonic effects," *Physical Review B (Condensed Matter)*, vol. 29, pp. 2051-9, 1984.
- [47] The simplicity of the semi-empirical methodology enables efficient computation of modal lifetimes by integrating the occupation number-

weighted density of states over frequency. This would enable one to perform self-consistent phonon transport studies efficiently.

- [48] M. V. Fischetti, S. E. Laux, P. M. Solomon, and A. Kumar, "Thirty years of Monte Carlo simulations of electronic transport in semiconductors: their relevance to science and to mainstream VLSI technology," in *2004 10th International Workshop on Computational Electronics, 24-27 Oct. 2004, West Lafayette, IN, USA*, pp. p.47-8, 2004.
- [49] C. Jacoboni and L. Reggiani, "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials," *Reviews of Modern Physics*, vol. 55, pp. 645-705, 1983.
- [50] N. Tsutsu, *et al.*, "New detection method of hot-carrier degradation using photon spectrum analysis of weak luminescence on CMOS VLSI," in *ICMTS 1990. Proceedings of the 1990 International Conference on Microelectronic Test Structures, 5-7 March 1990, San Diego, CA, USA*, pp. p.143-8, 1990.
- [51] J. A. Kash and J. C. Tsang, "Dynamic internal testing of CMOS circuits using hot luminescence," *IEEE Electron Device Letters*, vol. 18, pp. 330-2, 1997.
- [52] J. Bude, N. Sano, and A. Yoshii, "Hot-carrier luminescence in Si," *Physical Review B (Condensed Matter)*, vol. 45, pp. 5848-56, 1992.
- [53] E. Pop, R. W. Dutton, and K. E. Goodson, "Analytic band Monte Carlo model for electron transport in Si including acoustic and optical phonon dispersion," *Journal of Applied Physics*, vol. 96, pp. 4998-5005, 2004.
- [54] J. Rowlette, *et al.*, "Thermal simulation techniques for nanoscale transistors," in *IEEE/ACM International Conference on Computer-Aided Design, 2005. ICCAD-2005.*, pp. 225-228, 2005.
- [55] E. Pop, S. Sinha, and K. E. Goodson, "Heat generation and transport in nanometer-scale transistors," *Proceedings of the IEEE*, vol. 94, pp. 1587-1601, 2006.
- [56] P. D. Yoder and K. Hess, "First-principles Monte Carlo simulation of transport in Si," *Semiconductor Science and Technology*, vol. 9, pp. 852-854, 1994.
- [57] M. V. Fischetti and S. E. Laux, "Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects," *Physical Review B (Condensed Matter)*, vol. 38, pp. 9721-45, 1988.
- [58] T. Kunikiyo, *et al.*, "A Monte Carlo simulation of anisotropic electron transport in silicon including full band structure and anisotropic impact-ionization model," *Journal of Applied Physics*, vol. 75, pp. 297-312, 1994.
- [59] Unlike in early Monte Carlo codes, all phonon scattering processes, including intravalley acoustic scattering, are treated inelastically. This is important because an appreciable amount of the heat is generated in small-wavevector acoustic modes, even at large power densities.
- [60] D. Long, "Scattering of conduction electrons by lattice vibration in silicon," *Physical Review*, vol. 120, pp. 2024-2032, 1960.
- [61] C. T. M. Flik, "Size effect on the thermal conductivity of high-Tc thin-film superconductors," *Journal of Heat Transfer*, vol. 112, pp. 872-881, 1990.

- [62] M. Asheghi, Y. K. Leung, S. S. Wong, and K. E. Goodson, "Phonon-boundary scattering in thin silicon layers," *Applied Physics Letters*, vol. 71, pp. 1798-800, 1997.
- [63] In this limit we are essentially determining the appropriate temperature that when included in the argument of the B-E distribution function for a given frequency, will produce the appropriate occupation number in the nonequilibrium condition.
- [64] J. Lai and A. Majumdar, "Concurrent thermal and electrical modeling of sub-micrometer silicon devices," *Journal of Applied Physics*, vol. 79, pp. 7353-61, 1996.
- [65] A. Majumdar, K. Fushinobu, and K. Hijikata, "Effect of gate voltage on hot-electron and hot-phonon interaction and transport in a submicrometer transistor," *Journal of Applied Physics*, vol. 77, pp. 6686-94, 1995.
- [66] P. G. Sverdrup, Y. S. Ju, and K. E. Goodson, "Sub-continuum simulations of heat conduction in silicon-on-insulator transistors," *Transactions of the ASME. Journal of Heat Transfer*, vol. 123, pp. 130-7, 2001.
- [67] R. Lake and S. Datta, "Energy balance and heat exchange in mesoscopic systems," *Physical Review B (Condensed Matter)*, vol. 46, pp. 4757-63, 1992.
- [68] J. Rowlette, *et al.*, "Thermal phenomena in deeply scaled MOSFETs," in *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, pp. 984-987, 2005.
- [69] The choice of 150 nm source/drain regions was made in order to ensure that the contacts do not significantly affect the transport near the device hotspot.
- [70] A. R. Vasconcellos, R. Luzzi, C. G. Rodrigues, and V. N. Freire, "Hot-phonon bottleneck in the photoinjected plasma in GaN," *Applied Physics Letters*, vol. 82, pp. 2455-7, 2003.
- [71] J. Liberis, *et al.*, "Hot phonons in Si-doped GaN," *Applied Physics Letters*, vol. 89, 2006.
- [72] E. Pop, *et al.*, "Negative differential conductance and hot phonons in suspended nanotube molecular wires," *Physical Review Letters*, vol. 95, pp. 155505-4, 2005.
- [73] J. J. Letcher, K. Kang, D. G. Cahill, and D. D. Dlott, "Effects of high carrier densities on phonon and carrier lifetimes in Si by time-resolved anti-Stokes Raman scattering," *Applied Physics Letters*, vol. 90, pp. 252104-3, 2007.
- [74] E. Pop, R. Dutton, and K. Goodson, "Thermal analysis of ultra-thin body device scaling [SOI and FinFet devices]," in *IEEE International Electron Devices Meeting 2003, 8-10 Dec. 2003, Washington, DC, USA*, pp. p.36.6.1-4, 2003.
- [75] We wish not to confuse this topic with the "size effect" associated with boundary scattering in low-D semiconductors. To be clear, we are talking about an infinite medium in which the heat source is much smaller than the mean-free-path.
- [76] G. Chen, "Nonlocal and nonequilibrium heat conduction in the vicinity of nanoparticles," *Transactions of the ASME. Journal of Heat Transfer*, vol. 118, pp. 539-45, 1996.

- [77] G. Chen, "Ballistic-Diffusive Heat-Conduction Equations," *Physical Review Letters*, vol. 86, p. 2297, 2001.
- [78] Y. Wang, K.P. Cheung, A. S. Oates, and P. mason, "Ballistic Phonon Enhanced NBTI," in *IEEE International Reliability Physics Symposium*, Phoenix, AZ, U.S.A., 2007.
- [79] A. Abramo, *et al.*, "A comparison of numerical solutions of the Boltzmann transport equation for high-energy electron transport silicon," *IEEE Transactions on Electron Devices*, vol. 41, pp. 1646-54, 1994.
- [80] G. L. E. C. A. Paddock, "Transient Thermorefectance From Thin Metal Films," *Journal of Applied Physics*, vol. 60, pp. 285-290, 1986.
- [81] A. Feldman, *High Temperatures - High Pressures*, 1999.
- [82] D. G. Cahill, "Analysis of heat flow in layered structures for time-domain thermorefectance," *Review of Scientific Instruments*, vol. 75, pp. 5119-5122, 2004.
- [83] A. M. T. Tong, "Reexamining the 3-omega technique for thin film thermal characterization," *Review of Scientific Instruments*, vol. 77, 2006.
- [84] P. C. D. Hobbs, *Building electro-optical systems : making it all work* New York: Wiley, 2000.
- [85] D. A. B. Miller, "Device requirements for optical interconnects to silicon chips," *Proceedings of the IEEE*, vol. 97, p. 1166, 2009.
- [86] P. M. Fauchet, J. H. Shen, and Eds., "Issue on silicon photonics," *IEEE Journal of Selected Topics on Quantum Electronics*, vol. 12, p. 2, 2006.
- [87] G. T. Reed and A. P. Knights, *Silicon Photonics-An introduction*. . Chichester, U.K.: Wiley, 2004.
- [88] L. Pavesi, D. J. Lockwood, and Eds., *Silicon Photonics*. Berlin, Germany: Springer-Verlag, 2004.
- [89] A. L. Haisheng Rong, Richard Jones, Oded Cohen, Dani Hak, Remus Nicolaescu, Alexander Fang, Mario Paniccia, "An all-silicon Raman Laser," *Nature*, vol. 433, pp. 292-294, 2005.
- [90] L. Pavesi, *et al.*, "Optical gain in silicon nanocrystals," *Nature*, vol. 408, pp. 440-444, 2000.
- [91] M. L. Brongersma, *et al.*, "Size-dependent electron-hole exchange interaction in Si nanocrystals," *Applied Physics Letters*, vol. 76, pp. 351-3, 2000.
- [92] R. J. Walters, *et al.*, "Photoluminescence quantum efficiency of dense silicon nanocrystal ensembles in SiO<sub>2</sub>," *Physical Review B (Condensed Matter and Materials Physics)*, vol. 73, pp. 132302-4, 2006.
- [93] R. D. Kekatpure and M. L. Brongersma, "Fundamental photophysics and optical loss processes in Si-nanocrystal-doped microdisk resonators," *Physical Review A (Atomic, Molecular, and Optical Physics)*, vol. 78, pp. 023829-13, 2008.
- [94] M. V. Wolkin, *et al.*, "Electronic States and Luminescence in Porous Silicon Quantum Dots: The Role of Oxygen," *Physical Review Letters*, vol. 82, p. 197, 1999.
- [95] P. Malý, *et al.*, "Picosecond and millisecond dynamics of photoexcited carriers in porous silicon," *Physical Review B*, vol. 54, p. 7929, 1996.

- [96] K. S. Cho, *et al.*, "High efficiency visible electroluminescence from silicon nanocrystals embedded in silicon nitride using a transparent doping layer," *Applied Physics Letters*, vol. 86, pp. 071909-3, 2005.
- [97] L. D. Negro, *et al.*, "Light emission from silicon-rich nitride nanostructures," *Applied Physics Letters*, vol. 88, pp. 183103-3, 2006.
- [98] P. G. Kik, M. L. Brongersma, and A. Polman, "Strong exciton-erbium coupling in Si nanocrystal-doped SiO<sub>2</sub>," *Applied Physics Letters*, vol. 76, pp. 2325-7, 2000.
- [99] L. Kimerling, *et al.*, "Si-rich dielectrics for active photonic devices," in *Device Applications of Silicon Nanocrystals and Nanostructures*: Springer US, 2009.
- [100] P. G. Kik and A. Polman, "Gain limiting processes in Er-doped Si nanocrystal waveguides in SiO<sub>2</sub>," *Journal of Applied Physics*, vol. 91, pp. 534-536, 2002.
- [101] L. Dal Negro, *et al.*, "Stimulated emission in plasma-enhanced chemical vapour deposited silicon nanocrystals," *Physica E: Low-dimensional Systems and Nanostructures*, vol. 16, pp. 297-308, 2003.
- [102] R. D. Kekatpure and M. Brongersma, "Quantification of free-carrier absorption in silicon nanocrystals with an optical microcavity," *Nano Letters*, vol. 8, pp. 3787-3793, 2008.
- [103] V. I. Klimov, C. J. Schwarz, D. W. McBranch, and C. W. White, "Initial carrier relaxation dynamics in ion-implanted Si nanocrystals: Femtosecond transient absorption study," *Applied Physics Letters*, vol. 73, pp. 2603-2605, 1998.
- [104] M. Sykora, *et al.*, "Size-Dependent Intrinsic Radiative Decay Rates of Silicon Nanocrystals at Large Confinement Energies," *Physical Review Letters*, vol. 100, pp. 067401-4, 2008.
- [105] D. Kovalev, *et al.*, "Optical absorption cross sections of Si nanocrystals," *Physical Review B*, vol. 61, p. 4485, 2000.
- [106] D. Navarro-Urrios, *et al.*, "Quantification of the carrier absorption losses in Si-nanocrystal rich rib waveguides at 1.54  $\mu\text{m}$ ," *Applied Physics Letters*, vol. 92, pp. 051101-3, 2008.
- [107] M. Yamaguchi, *et al.*, "Carrier dynamics in porous silicon studied with a near-field heterodyne transient grating method," *Chemical Physics Letters*, vol. 427, pp. 192-196, 2006.
- [108] G. Allan and C. Delerue, "Energy transfer between semiconductor nanocrystals: validity of Forster's theory," *Physical Review B (Condensed Matter and Materials Physics)*, vol. 75, pp. 195311-1, 2007.
- [109] D. B. Laks, G. F. Neumark, A. Hangleiter, and S. T. Pantelides, "Theory of Interband Auger Recombination in n-type Silicon," *Physical Review Letters*, vol. 61, p. 1229, 1988.
- [110] V. I. Klimov, "Spectral and Dynamical Properties of Multiexcitons in Semiconductor Nanocrystals," *Annual Review of Physical Chemistry*, vol. 58, pp. 635-673, 2007.
- [111] The notation...

- [112] D. Fröhlich and H. Mahr, "Resonant Energy Transfer between Excited F Centers in KI," *Physical Review*, vol. 148, p. 868, 1966.
- [113] J. A. Rowlette, *et al.*, "Nonradiative recombination in strongly interacting silicon nanocrystals embedded in amorphous silicon-oxide films," *Physical Review B (Condensed Matter and Materials Physics)*, vol. 80, 2009.
- [114] D. L. Dexter, "A Theory of Sensitized Luminescence in Solids," *The Journal of Chemical Physics*, vol. 21, pp. 836-850, 1953.
- [115] Explicitly...
- [116] W. Spitzer and H. Y. Fan, "Infrared absorption in N-type silicon," *Physical Review*, vol. 108, pp. 268-271, 1957.
- [117] A. Franceschetti and Y. Zhang, "Multiexciton Absorption and Multiple Exciton Generation in CdSe Quantum Dots," *Physical Review Letters*, vol. 100, pp. 136805-4, 2008.
- [118] F. Trojanek, K. Neudert, M. Bittner, and P. Maly, "Picosecond photoluminescence and transient absorption in silicon nanocrystals," *Physical Review B (Condensed Matter and Materials Physics)*, vol. 72, pp. 075365-6, 2005.
- [119] V. I. Klimov, *et al.*, "Quantization of Multiparticle Auger Rates in Semiconductor Quantum Dots," *Science*, vol. 287, pp. 1011-1013, February 11, 2000 2000.
- [120] The double meaning of count here is intentional. To first order, every atom will be important in the sense of statistical fluctuations in data storage. This is already very evident in traditional classically scaled transistor devices where dopant concentration fluctuations are severely impacted by the absence of a handful of dopant atoms within the channel region.. To second order, in the distant future, each atom or small cluster of atoms could potentially act as a local independent computational stage, i.e. quantum computing in the true sense.
- [121] S. Sinha, E. Pop, and K. E. Goodson, "A split-flux model for phonon transport near hotspots," *American Society of Mechanical Engineers, Electronic and Photonic Packaging, EPP*, vol. 4, pp. 75-85, 2004.
- [122] S. Sinha, "Non-equilibrium phonons in silicon field-effect transistors " in *Department of Mechanical Engineering*. vol. Ph. D. Stanford, CA: Stanford University, p. 107, 2005.
- [123] S. P. C.-R. B. J. Ainslie, S. T. Davey, J. R. Armitage, C. G. Atkins, J. F. Massicott, R. Wyatt, "Erbium doped fibers for efficient optical amplifiers," *Opt. Amplif. Commun. IEE Proc.*, vol. 137, pp. 205-208, 1990.
- [124] J. N. P. Blixt, T. Carlnas, B. Jaskorzynska, "Concentration-dependent upconversion in Er<sup>3+</sup>-doped fiber amplifiers: experiments and modeling," *IEEE Transactions on Photonics Technology Letters*, vol. 3, pp. 996-998, 1991.