Charge and Energy Transport in 0D/2D Systems
Probed Using Linear and Nonlinear Spectroscopy

by

Aaron J. Goodman

Submitted to the Department of Chemistry
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Physical Chemistry

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

© Massachusetts Institute of Technology 2018. All rights reserved.
Abstract

Low-dimensional nanostructured semiconductors are promising technologies for next generation optoelectronics. Colloidal quantum dots (QDs) have been applied in many light emitting applications such as ambient lighting downconverters, LEDs, and displays. Proof-of-concept transistors, phototransistors and LEDs have been demonstrated using two dimensional atomically thin transition metal dichalcogenides (TMDs). To realize efficient next generation optoelectronics using these materials it is crucial to understand the dynamics and transport of energy and charge in these materials and hybrid structures built from them. The first Chapter of this thesis briefly motivates the technological importance of QDs and TMDs. The remainder explores exciton dynamics and transport in these low-dimensional semiconductor systems.

In Chapter 2, low temperature spectroscopy and time-resolved photoluminescence spectroscopy are used to probe the complex energy landscape seen by excitons in acid-treated MoS$_2$. We show that deeply trapped “dark” exciton states are responsible for the surprisingly long lifetime of band-edge photoluminescence in acid-treated single-layer MoS$_2$. Temperature-dependent transient photoluminescence spectroscopy reveals an exponential tail of long-lived states extending hundreds of meV into the band gap. These subband states, which are characterized by a 4 $\mu$s radiative lifetime, quickly capture and store photogenerated excitons before subsequent thermalization up to the band edge where fast radiative recombination occurs. By intentionally saturating these trap states, we are able to measure the “true” 150 ps radiative lifetime of the band-edge exciton at 77 K, which extrapolates to ~600 ps at room temperature. These experiments reveal the dominant role of dark exciton states in acid-treated MoS$_2$, and suggest that excitons spend > 95% of their lifetime at room temperature in trap states below the band edge. We hypothesize that these states are associated with native structural defects, which are not introduced by the superacid treatment; rather, the superacid treatment dramatically reduces nonradiative recombination through these states, extending the exciton lifetime and increasing the likelihood of eventual radiative recombination.

In the second half of Chapter 2, we study exciton diffusive transport in MoS$_2$
using time-resolved diffusion imaging. We also probe exciton-exciton dynamics and elucidate the role that exciton traps play in both exciton transport and dynamics. Atomically thin semiconductors such as monolayer MoS$_2$ and WS$_2$ exhibit nonlinear exciton-exciton annihilation at notably low excitation densities (below $\sim$10 excitons/µm$^2$ in MoS$_2$). We show that the density threshold at which annihilation occurs can be tuned by two orders of magnitude by varying the refractive index of the underlying supporting substrate. Using spatially-resolved transient photoluminescence spectroscopy in conjunction with numerical simulations, we find that this behavior arises from screening interactions between trapped and mobile exciton pairs. We measure the effective exciton diffusion coefficient in superacid-treated MoS$_2$ to be $D = 0.06 \pm 0.01$ cm$^2$/s, corresponding to a diffusion length of $L_D = 350$ nm for an exciton lifetime of $\tau = 20$ ns. Exciton-exciton annihilation limits the overall efficiency of 2D semiconductor devices operating at high exciton densities. The ability to tune these interactions via the dielectric environment is an important step toward more efficient optoelectronic technologies featuring atomically thin materials.

In Chapter 3 we investigate coupled QD/TMD hybrid structures. In the first half of Chapter 3, we investigate dipole-dipole coupling in the regime that the QD and TMD exhibit weak electronic coupling by engineering the interface to be insulating. We report highly efficient Förster resonant energy transfer from cadmium selenide (CdSe) quantum dots to monolayer and few-layer molybdenum disulfide (MoS$_2$). The quenching of the donor quantum dot photoluminescence increases as the MoS$_2$ flake thickness decreases with the highest efficiency (> 95%) observed for monolayer MoS$_2$. This counterintuitive result (that deviates from the predictions of Förster theory) arises from reduced dielectric screening in thin layer semiconductors having unusually large permittivity and a strong in-plane transition dipole moment, as found in MoS$_2$.

In the second half of Chapter 3, we investigate QD/TMD hybrid structures in the strongly coupled regime. We demonstrate tunable electronic coupling between CdSe QDs and monolayer WS$_2$ using variable length alkanethiol ligands on the QD surface. Using femtosecond time-resolved second harmonic generation (SHG) microscopy, we show that electron transfer from photoexcited CdSe QDs to single-layer WS$_2$ occurs on ultrafast (10 fs - 1 ps) timescales. Moreover, charge transfer excites coherent acoustic phonons in the donor QDs, which modulate the SHG response of the underlying WS$_2$ layer on picosecond timescales. These results reveal surprisingly strong electronic coupling at the QD/TMD interface and demonstrate the usefulness of time-resolved SHG for exploring ultrafast electronic-vibrational dynamics in TMD heterostructures.

In Chapter 4, we improve the sensitivity of traditional SHG spectroscopy by adding an additional field. We demonstrate the dramatic enhancement of weak second-order nonlinear optical signals via stimulated sum and difference frequency generation. We present a conceptual framework to quantitatively describe the interaction and show that the process is highly sensitive to the relative optical phase of the stimulating field. To emphasize the utility of the technique, we demonstrate stimulated enhancement of second harmonic generation (SHG) from bovine collagen-I fibrils. Using a stimulating pulse fluence of only 3 nJ·cm$^2$, we obtain an SHG enhancement $> 10^4$ relative to the spontaneous signal. We discuss the conceptual differences between optical heterodyning of SHG signals and our presented “stimulation” framework. In
the second half of Chapter 4, we use the additional field to resolve the SHG signal phase revealing the sign and magnitude of the sample nonlinear susceptibility. We perform phase-resolved SHG imaging of polycrystalline MoS$_2$ flakes and determine the absolute grain orientations using the phase information.

Thesis Supervisor: William A. Tisale
Title: Associate Professor without Tenure
Acknowledgments

I would like to thank sources of financial funding that made my Ph.D. possible. Thanks to the National Science Foundation for paying my stipend through the Graduate Research Fellowship Program, to the Department of Energy for funding much of my research through Will’s Early Career Award and through the Center for Excitonics, an Energy Frontier Research Center at MIT and Harvard. Thanks to the Samsung Global Research Opportunities program for funding work investigating charge transfer at QD/TMD interfaces.

Countless thanks to Will for being an outstanding advisor. In a very practical sense, Will always enables his students to do good science: he provides funds, creativity, and rigorous scientific thinking all of which are important in creating a successful research environment. I know that I pursued a few hare-brained ideas; Will enabled that exploration even when it didn’t directly address an existing research goal. I especially enjoyed and valued the scientific conversations in which we solved problems collaboratively. Perhaps most importantly, Will always puts the needs and goals of his students first valuing their success rather than treating them as a means to an end (this is surprisingly uncommon at MIT). He does this conscientiously and with great kindness. I also thank him for actively creating an inclusive work environment where all group members feel welcome. I couldn’t have asked for a better group of coworkers and that starts at the top.

My thesis committee has helped me immensely. Professor Nelson has provided valuable insight. I always enjoyed our annual meeting; his passion for science is unmatched and his deep understanding of ultrafast optics made our research discussions a true pleasure. Professor Bawendi contributed considerable expertise in the physics of nanostructured materials. Talking with him and his group members really strengthened my knowledge of exciton transport and dynamics empowering me to complete this work.

Special thanks to Ferry Prins for being a mentor and friend in the Tisdale group. Ferry brought two dimensional TMDs into the group and spearheaded the QD/TMD
energy transfer story. Additionally, Ferry was an enthusiastic, capable mentor and force for positive culture in the group during the formative early years. Though there is stiff competition, Ferry may have been the most capable drinking partner to grace the Muddy Charles with me.

In a tie for close second are Dr. Mark Weidman and Leo Spiegel. Mark is a phenomenal scientist whose productivity is a thing of Tisdale Lab legend. Mark helped train me on the diffusion imaging apparatus that he and Matt built together and was a valuable collaborator and friend.

Yunan Gao (Steve, the first of his name) was another outstanding postdoctoral researcher who developed the phase-resolved degenerate parametric amplification technology to completion and made important contributions to my pump-probe SHG optical setup. Yunan thought about science in a creative and simultaneously precise way and we shared many stimulating discussions. I appreciate his kindness and friendship.

Dr. Jolene Mork was a fantastic mentor who was always willing to help me. I appreciate her willingness to not take things too seriously and her companionship that one time we were stranded in Texas. With her considerable wet lab expertise she set many good precedents that made the lab more safe and productive, and was a willing aid when I had to do things with chemicals.

Thanks to Dr. Rachel Gilmore for teaching me ultrafast optics and for building the ultrafast bay from scratch. I appreciate her willingness to take things just seriously enough. The depth and completeness with which she answered her scientific questions always impressed me. I am glad to have her as a friend.

Thanks Matt Ashner for introducing me to climbing. He probably increased my life expectancy by showing me a physical activity that captures my imagination and reduces my stress. He is an ambassador to the sport.

Thank you to all of the other Tisdale Lab members for being talented, wonderful coworkers and collaborators. Nabeel thank you for tirelessly taking point making sure the lab continued to function. Sam, thank you for being an even better social chair than I could ever hope to be. Liza, thank you for your piercing insight and for your
friendship. Thank you for members of the Gossip Club for gossiping with me.

For the last two years, Michelle and I have lived in East Campus on 41W. Thank you to Rob Miller, Sandy Alexandre, and Carrie Wicks for being a supportive house team. Each cares deeply about the students in East Campus and provided bottomless support for the Graduate Resident Tutors (GRTs) there. Likewise, the other GRTs are beautiful people whom I’m glad to know. Living on 41W has taught me to practice my empathy and also made me question some of my assumptions. It is special to live in a community where everyone is accepted as their authentic self.

I was lucky to work in Professor Bart Bartlett’s lab at the University of Michigan. My experience there drove me to start my graduate studies by showing me how fun basic research could be. Bart was a formative undergraduate advisor who never compromised and produced good science as a result. I appreciate the opportunity to work in his lab.

Thank you to my dear friends Adam, David, Max, Russ, Sean, and Ted. You offer consistent support when things are difficult. Thanks for being there.

While at MIT, I met and married my best friend Michelle. Thank you Michelle for always supporting me, convincing me of my worth, and loving me unconditionally. We make a good team, and I look forward to what comes next.

My family has always supported me in my endeavors. Thanks to my parents for providing me with a safe home and a college education. I broke a lot of things in our house when I did “science” “experiments”, and my parents were generally supportive of that creativity. Thank you to Sarah and David for being wonderful siblings who always want the best for me. I love you guys.

**Buddy**

tail wagging at work
he is happy that you’re here
he will steal your food.
This doctoral thesis has been examined by a Committee of the
Department of Chemistry as follows:

Signature redacted
Professor Keith A. Nelson............
Chairman, Thesis Committee
Haslam and Dewey Professor of Chemistry

Signature redacted
Professor William A. Tisdale............
Thesis Supervisor
ARCO Career Development Associate Professor of Chemical Engineering

Signature redacted
Professor Moungi G. Bawendi............
Member, Thesis Committee
Lester Wolfe Professor of Chemistry
## Contents

1 Motivation for Low Dimensional Materials 35  
1.1 2D TMDs for Lighting and Computing 35  
1.2 Quantum Dots for Displays and Lighting 38  

2 Exciton Transport and Dynamics in Transition Metal Dichalcogenide Monolayers 41  
2.1 Excitons in Transition Metal Dichalcogenides 41  
2.1.1 Exciton Background Information 41  
2.1.2 Transition Metal Dichalcogenides Background 43  
2.2 Exciton Trapping Is Responsible for the Long Apparent Lifetime in Acid-Treated MoS$_2$ 45  
2.2.1 Background for Chemical Treatment of MoS$_2$ Samples 45  
2.2.2 Background for Radiative Rates and Apparent Decay Rates in TMDs 47  
2.2.3 Trapped Exciton States in MoS$_2$ 49  
2.2.4 Equilibration between Band-Edge and Trapped Exciton States 51  
2.2.5 Observing the True Radiative Rate 53  
2.2.6 Numerical Modelling Details and Results 56  
2.2.7 Trap States in CVD-Grown MoS$_2$ 58  
2.2.8 Conclusions 59  
2.2.9 Experimental Methods 61  
2.3 Suppressing Diffusion-Mediated Exciton-Exciton Annihilation in 2D Semiconductors Using the Dielectric Environment 62
3 2D/0D Hybrid Structures: CdSe Quantum Dots on Transition Metal Dichalcogenides

3.1 Hybrid Quantum Dot / 2D Devices ........................................... 89
3.2 Exciton States in Transition Metal Dichalcogenides and Quantum Dots 91
  3.2.1 Transition Metal Dichalcogenide Electronic Structure ........ 91
  3.2.2 CdSe Quantum Dot Electronic Structure .................. 95
3.3 Reduced Dielectric Screening and Enhanced Energy Transfer in Single-
   and Few-Layer MoS$_2$ .................................................. 97
  3.3.1 Förster Resonance Energy Transfer .............................. 97
  3.3.2 Measuring the Rate of Energy Transfer Using Time-Correlated
        Single Photon Counting ........................................... 99
  3.3.3 Energy Transfer in CdSe/CdZnS Quantum Dot/MoS$_2$ Hybrid
        Structures ......................................................... 101
  3.3.4 Conclusions .......................................................... 108
  3.3.5 Energy Transfer from to MoS$_2$ from QDs with Thin Shells and
        Short Ligands ...................................................... 110
3.4 Theory of Charge Transfer in 0D/2D Donor/Acceptor Systems 111
3.5 Charge Transfer in CdSe Quantum Dot/WS₂ Monolayer Heterostructures 113

4 Stimulated Second Harmonic Generation 123
4.1 Second Harmonic Generation 123
4.2 Optical Stimulation Background 126
4.3 Enhancement of Second-Order Nonlinear-Optical Signals by Optical Stimulation 128
  4.3.1 Second Order Nonlinear Optical Spectroscopies: Successes and Opportunities 128
  4.3.2 Amplifying the Second Harmonic Generation Signal: Results 130
4.4 Philosophical and Practical Differences between Optical Stimulation and Heterodyne Detection 137
  4.4.1 Comparing Optical Stimulation to Heterodyne Detection 137
  4.4.2 Contrasting Optical Stimulation to Heterodyne Detection 139
4.5 Phase-Sensitive Nonlinear Optical Imaging of Two-Dimensional Materials 141
  4.5.1 Phase-Sensitive Imaging of 2D Materials 143
  4.5.2 Spatial Resolution of DPA 146
  4.5.3 Conclusion 147
  4.5.4 Methods 148
List of Figures

1-1 Members of the layered TMD family typically possess one of three crystallographic symmetries (2H, 1T, or 1T'). The thermodynamically favored structural phase depends on material composition with some compositions possessing metastable phases. .......................... 36

1-2 (a) CIE chromaticity diagram. QDs exhibit narrow emission line widths and can be tuned to emit across the visible spectrum. Their color purity allows mixtures of dots to accurately reproduce a large fraction of the humanly perceivable color space. (b) Plot showing luminous efficacy (lm W⁻¹) and color rendering index (CRI, a metric indicating a light source’s ability to accurately render colors). Due to their color purity and spectral tunability QD-based ambient lighting devices can be engineered with good luminous efficiency and CRI. Figure reproduced from ref. [1], Nature Publishing Group ................................. 39

2-1 Exciton varieties. (a) Bulk Wannier-Mott exciton (b) Bulk Frenkel exciton (c) An exciton confined in one dimension (d) An exciton confined in three dimensions .................................................. 42

2-2 Acid-treated MoS₂ optical properties at room temperature. (a) The absorbance spectrum (blue) shows two sharp, excitonic resonances arising from the A and B excitons that originate from the K point in the Brillouin zone. (b) Time-correlated single photon counting histogram of the A exciton photoluminescence at room temperature exhibits single-exponential decay dynamics with a time constant, \( \tau_{obs} = 20 \) ns. ........ 47
Temperature-dependent photoluminescence spectra (a) Photoluminescence spectra (thin lines) at low temperature exhibit emission at energies below the band-edge exciton. The lineshape is fit well by an exponential density of states occupied by a Fermi-Dirac Distribution (thick lines). (b) At low temperature, trapped exciton emission dominates the photoluminescence spectra, while at high temperatures most luminescence comes from the band-edge exciton. (c) Power-dependent photoluminescence spectra at 77 K reveal that the band-edge exciton emission becomes more prominent relative to the trapped exciton emission at higher laser powers. This observation indicates that a significant portion of the traps are filled at moderate excitation intensities.

A three-state model explains the observed photoluminescence decay time and the absence of trap state emission at room temperature. (a) Apart from the ground state and the band-edge exciton, there are also trapped exciton states that have \(~1000\times\) slower recombination rate than the band-edge exciton. At room temperature, thermalization up to the band edge and subsequent radiative recombination is most probable, whereas at 77 K direct recombination from trap to ground state is observed. (b) Time-resolved trapped exciton emission reveals these two pathways. At 77 K, trapped excitons decay with their slow radiative rate (\(~4 \mu s\)). At higher temperatures, some thermalize up to the band-edge leaving behind only deeply trapped excitons, which decay slowly. The shaded region of the inset indicates the portion of the spectrum that is collected by the time-resolved detector.
Spectrally-resolved transient photoluminescence spectroscopy reveals the true radiative rate. (a) At room temperature, nearly all of the emission comes from the band-edge exciton, with a noticeable contribution from the B exciton. The exciton emission has long-lived components mediated by the long-lived, dark trapped exciton states. (b, c) At 77 K, the majority of emission comes from trapped excitons (~1.65−1.85 eV). The pulsed laser operates at 76 MHz (13 ns between pulses), so the long-lived trapped exciton sites are always nearly completely occupied in this experiment. When a new pulse arrives, band edge excitons cannot trap and instead decay radiatively with a rate reflecting the true band-edge radiative rate, which is 150 ps at 77 K.

Power-dependent PL at 77 K. Power-dependent PL spectra (light gray) were fitted to band edge and trapped exciton contributions (colored lines) to extract density-dependent steady state populations.

Transient photoluminescence modelling. (a) The trap state energy spectrum used in the numerical model is illustrated here. The PL spectrum is plotted in red with the x-axis energy centered about the band edge exciton energy, $E_X$. The density of trap states is exponentially decaying from the band edge with a slope $\alpha = 5 \text{ eV}^{-1}$ found from fitting the PL spectra. This exponential decay is plotted in blue. The transient ODE model was solved five times at each temperature with trap state energies represented by the blue rectangles. The weight of each solution parameterized by a specific trap state energy was determined by the density of states. The five solutions for each temperature were added together to generate the model’s output. (b) The experimental trapped exciton emission at multiple temperatures is plotted along with the model output, which was generated with no free parameters.
Temperature dependent PL spectra from acid-treated CVD-grown MoS$_2$. The CVD-grown sample also exhibits trapped exciton emission at low temperature. The trapped exciton emission line shape differs reflecting a different energetic distribution of trap states.

Proposed superacid treatment mechanism. The TFSI acid chemically oxidizes the MoS$_2$, removing dopant electrons. The conjugate base anion then passivates the structural defect.

Three carrier Auger processes (left) A relaxing electron scatters off another electron promoting increasing its energy (right) A relaxing electron scatters off of a hole, pushing it deeper into the valence band. In each case, energy and momentum are conserved.

Effect of substrate on the optical properties of superacid-treated MoS$_2$. (a) Transmission optical micrograph of exfoliated MoS$_2$ with monolayer region indicated. The inset depicts the chemical structure of the superacid used in the treatment. The dark thick line is the border of the polymer capping layer on top of the MoS$_2$ flake. (b) Schematic of the substrate/MoS$_2$/polymer stack. (c) Photoluminescence spectra of treated MoS$_2$ on quartz (blue), sapphire (green), and STO (red). (d) Steady-state QY measured as a function of exciton generation rate. Data were recorded for MoS$_2$ supported on quartz (blue trace), sapphire (green trace), and strontium titanate (red trace). (e) Time-resolved photoluminescence traces. The traces were globally fit to extract $k_X$ and $k_{XX}$ as described in the text. (f) Exciton-exciton annihilation rate constants, $k_{XX}$, inferred from steady-state QY measurements (open circles) and time-resolved photoluminescence measurements (filled squares). $k_{XX}$ values for MoS$_2$ on quartz (blue), sapphire (green), and strontium titanate (red) are plotted against the supporting substrates' refractive indices (top). Analogous data for WS$_2$ samples are shown in the bottom panel.
Exciton diffusion imaging. (a) Schematic diagram of the optical apparatus used to resolve exciton diffusion in space and time. (b) top: simulation of the photoluminescence intensity along a line cut of the radially symmetric exciton population as a function time. With an incident fluence of 0.5 nJ/cm$^2$, almost all excitons decay radiatively and the spot broadens due to diffusion. bottom: same as top panel, but with an incident fluence of 50 nJ/cm$^2$, many excitons decay due to annihilation. This results in artificial broadening of the density profile; excitons decay most rapidly in the center of the spot where exciton density is highest. (c) Experimentally observed broadening of the exciton population with time at 0.5 nJ/cm$^2$, 5 nJ/cm$^2$, and 50 nJ/cm$^2$ incident fluences in the top, middle, and bottom panels, respectively. White lines indicate the evolution of the standard deviation with time. (d) Steady-state measurement of exciton diffusion. The intensity profile of luminescence coming from a film of quantum dots is indicated in green. Excitons do not diffuse in this sample, so the intensity profile indicates the performance of our optical system. The photoluminescence intensity profile collected from MoS$_2$ with a generation rate of $4.8 \times 10^{15}$ cm$^{-2}$s$^{-1}$ is broadened due to diffusion. The diffusive contribution to the width of the intensity distribution is indicated in red in the inset. (e) This cartoon illustrates the role that trapped excitons play in influencing the measured diffusivity. Though excitons diffuse quickly at the band edge with the band edge diffusivity, they fall into immobile traps. Consequently the measured diffusivity ($D_{\text{measured}}$) is much smaller than the true band edge diffusivity.
2-13 Trapped exciton states and their equilibration with the band edge. (a) 77 K photoluminescence spectra of MoS$_2$ supported on quartz (top), sapphire (middle), and strontium titanate (bottom). At low temperature, trapped excitons emit at energies below the free exciton. The photoluminescence spectra (solid black lines) are modelled as a Fermi-Dirac distribution over an exponential density of states. (b) State diagram for trapped and band-edge (free) excitons.

2-14 Monte Carlo simulation of trapped excitons contributing to annihilation. (a) Diffusing excitons meet long-lived trapped excitons and annihilate. The equilibrium between trapped and free excitons is responsible for the large annihilation rate constants. (b) A Monte Carlo model invoking trapped and diffusing band edge excitons quantitatively reproduces steady-state QY data. The model output is plotted here (thick transparent lines) against the experimental data (open circles) for MoS$_2$ on quartz (blue), sapphire (green), and strontium titanate (red). (c) The annihilation rate constant inferred from the model (crossed diamonds) is plotted along with the experimentally derived values (open circles, QY; filled squares, time-resolved photoluminescence). The annihilation rate is plotted against the supporting substrates’ refractive indices.

2-15 Screened Coulomb interactions in TMDs. The analytical form is included as an inset. $H_0$ and $Y_0$ are Struve and Bessel functions, respectively, $\epsilon_{\text{sub}}$ and $\epsilon_{\text{pmma}}$ are the permittivity of the substrate and PMMA respectively, and $r_0$ is a characteristic screening length scale proportional to the TMD polarizability.
Diffusion imaging experiments performed on MoS$_2$ supported on sapphire (a), and STO (b). Top panels show the spatially-integrated PL decay. Middle panels show exciton diffusion imaging as discussed in 2.3.4. Bottom panels show the PL spatial distribution variance extracted from the middle panels evolving in time showing diffusive broadening.

Absolute quantum yield values for WS$_2$ supported on quartz (green circles), sapphire (yellow triangles), and STO (red squares) extracted from calibrated PL spectra as a function of carrier generation rate. The data are consistent with and qualitatively similar to the data for MoS$_2$ on the three substrates. At low generation rates, the samples emit with near unity quantum yield. As the generation rate increases, the exciton-exciton annihilation interaction lowers the quantum yield dramatically.

Optical imaging apparatus characterization. Left: PL image collected from a thin film of well insulated CdSe quantum dots in which exciton diffusion does not occur. The dots were excited with a focused laser ($\lambda = 520$ nm, NA = 1.4). Middle: PL image collected from MoS$_2$ with the same imaging system using a sufficiently low fluence to avoid exciton-exciton annihilation. Right: Radial intensity profiles of the images in the left and center panels showing that the MoS$_2$ PL intensity profile is broadened due to exciton diffusion.

Photoluminescence spectra collected upon cooling (left) to and heating (right) from liquid nitrogen temperature. The color gradation indicates temperature (room temperature is red, 77 K is blue). The cooling and heating cycle took approximately one hour.

Integrated photoluminescence intensity upon cooling (blue) to and heating (red) from liquid nitrogen temperature. The cooling and heating cycle took approximately one hour.
2-21 Schematic illustration of the two sub populations that contribute to low-temperature PL spectra. .............................................. 86

3-1 *QD/TMD hybrid devices* (a) A proposed hybrid LED device in which electrodes inject charges into the TMD. The charges meet and form excitons. Excitons transfer to the adjacent QD, which handles light emission. (b) A hybrid QD/TMD photodector takes advantage of the QDs tunable broadband absorption. The QD undergoes photoinduced charge transfer injecting free carriers into the TMD. The free carriers are swept out of the device by an applied field leading to photoinduced gain. ................................................................. 91

3-2 (a) left: A molybdenum disulfide unit cell with lattice vectors $a_1$, $a_2$. right: A 2D hexagonal lattice Brillouin zone with labelled critical points as well as reciprocal lattice vectors $b_1$, $b_2$. (b) Band structure of monolayer MoS$_2$ adapted from reference [2]. (c) Monolayer MoS$_2$ differential reflectance spectrum showing absorption features corresponding to the A and B excitons as well as interband transitions in the band nesting region, C. ................................................................. 92

3-3 (a) Electric fields lines emanating from charges in atomically thin TMDs can sample the surrounding dielectric environment, which is typically less polarizable than the TMD. (b) The Coulomb interaction between two point charges in a TMD is sensitive to the surrounding dielectric. Here the Coulomb potential energy due to two point charges separated by a distance $r_{ij}$ is plotted. Bulk TMDs (such as MoS$_2$) are highly polarizable and the Coulomb potential is strongly screened (green). Quartz is much less polarizable and the Coulomb potential is less screened (blue). Charges in a MoS$_2$ monolayer encased quartz experience a Coulomb potential similar to charges separated in quartz except at very small distances, where the Coulomb potential approaches that of the bulk TMD (red). ................................................................. 94
3-4 (a) Colloidal quantum dots are nanometer-scale semiconductor crystallite cores covered with organic small molecule capping ligands that impart colloidal stability. (b) When the nanocrystal size is reduced below a critical size, the carriers begin to feel size-dependent quantum confinement effects. The continuous bands of states present in the bulk transition into atom-like discrete states near the band edge whose energies are sensitive to the quantum dot size.

3-5 Qualitative description of Förster resonant energy transfer (FRET) (a) FRET can be described as a donor dipole emitting a near field which is then received by a lossy acceptor dipole. The acceptor dipole must have some absorptive power by which to receive the radiated field. This interaction is screened by the intervening dielectric. (b) This analogy carries over to excitonic systems. Donor and acceptor dipoles correspond to the acceptor and donor transition dipole moments. Similarly to the macroscopic antenna picture, the dipoles are screened by a dielectric and must be aligned to facilitate energy transfer.

3-6 An excited quantum dot exhibits an effective decay rate, $k_{\text{total}}$, that is a sum of all the first order decay rates available to the exciton: the radiative rate $k_{\text{rad}}$, first order nonradiative rates $k_{\text{NR}}$, and the energy transfer rate, $k_{\text{ET}}$. We can infer the energy transfer rate by measuring the total decay rate in the presence and absence of an energy acceptor.

3-7 Time-correlated single photon counting experimental apparatus. A pulsed diode laser excites the sample. Emitted photons are directed to an avalanche photodiode (APD). Photon detection events at the APD result in a voltage pulse being sent to the counting board, which is synchronized to the laser excitation. Pulse arrival - photon emission delay times are recorded and histogrammed by the counting board.
3-8 (a) Schematic of the hybrid QD/MoS$_2$ energy transfer system with indicated length scales. (b) Monolayer MoS$_2$ absorption spectrum [3], overlaid with the CdSe/CdZnS emission spectrum. Inset: transmission electron micrograph of a monolayer of the CdSe/CdZnS QDs used in this study. (c) Scanning electron micrograph showing the submonolayer of CdSe/CdZnS QDs on top of a MoS$_2$ flake (dark section) and the surrounding SiO$_2$ substrate (lighter section).

3-9 (a) Optical micrograph of a mechanically exfoliated MoS$_2$ flake with indicated layer thicknesses of the different regions. (b) QD fluorescence lifetime map of the same flake with a saturated color scale to emphasize the lifetime variations within the flake area. (c) Energy transfer efficiency map of the same flake.

3-10 (a) Fluorescence lifetime histograms of CdSe QDs on different thicknesses of MoS$_2$. The black curve is obtained from QDs on top of SiO$_2$. The dashed gray curve represents the instrument’s response function. Inset: spectral overlap between CdSe QD emission and MoS$_2$ absorption spectra. (b) Same as (a) but for InAs quantum dots with an exciton resonance energy below the MoS$_2$ exciton resonance.

3-11 Energy transfer rate as a function of the number of MoS$_2$ layers. The dashed line represents the energy transfer rate to thick, bulklike MoS$_2$. Blue circles indicate experimental data with associated uncertainty that is smaller than the size of the symbols. The solid line is a fit to eqn. 3.10. Inset: schematic representation of the definition of the thickness (t) and distance (d) parameters.
3-12 (a) Representative fluorescence lifetime histograms of the thin-shell CdSe QDs on different thicknesses of MoS₂. The black curve represents the instrument's response function (IRF). Inset: spectral overlap between thin-shell CdSe QD emission and MoS₂ absorption spectra. (b) Enhancement factor as a function of $t/d$, where $t$ is the underlying MoS₂ thickness, and $d$ is the distance from the quantum dot center to the MoS₂ surface. Red circles represent the thin-shell quantum dot data, which is in good agreement with the thick-shell quantum dot data presented in Fig. 3-11 in section 3.3.3 (blue circles) as well as with the fit of eqn. (3.10) (black line) to the thick-shell data.

3-13 Characterization of the quantum dot (QD)/WS₂ heterostructure. (a) Schematic of the QD/WS₂ interface. QDs are coated with alkanethiol ligands of tunable length. (b) Estimated band alignment at the hybrid interface, based on published ionization potentials, exciton binding energies, and measured optical gaps. (c) CdSe QD (dark, red) and WS₂ (light, orange) absorption spectra. The pump pulse spectrum (black, dashed) selectively excites the quantum dots in the heterostructure.

3-14 Time-resolved second harmonic generation (SHG). (a) Schematic of the time-resolved SHG microscopy instrument. (b) Transient SHG response of different samples. Data are shown for quantum dots capped with ethanethiol (blue, circles), butanethiol (green, squares), octanethiol (red, diamonds), and dodecanethiol (violet, triangles). Symbols represent the experimental data, while transparent lines represent fits to a simple kinetic model. (c) Inverse electron transfer rate constants determined by fitting the data to our kinetic model. Values measured at different locations on the spatially heterogeneous sample are indicated by open circles, while the average values are denoted by filled circles. The gray line at the bottom of the figure represents the limitation of our instrument resolution.
Recovery of the transient SHG response. (a) Following photoexcitation, an electron in the QD excited state can return to the QD ground state or transfer to the acceptor WS$_2$. From there, the charge can recombine directly to the QD or diffuse within the WS$_2$. (b) Representative SHG signal recovery dynamics are shown for WS$_2$ covered with ethanethiol- (blue, circles), butanethiol- (green, squares), octanethiol- (red, diamonds), and dodecanethiol- (violet, triangles) capped QDs. Across multiple samples, long time transient SHG traces show similar relaxation dynamics.

Coherent phonon oscillations. (a) Transient SHG response for ethanethiol- and butanethiol-capped QD samples on WS$_2$, exhibiting coherent oscillation in the signal intensity. (b) Residuals of a bi-exponential fit to the ethanethiol data in (a) are plotted in black, and fitted to a damped sine wave, plotted in red, with a frequency of $\sim 14$ cm$^{-1}$. (c) The fit extracted an oscillation frequency $\sim 14$ cm$^{-1}$. (c) Raman spectrum of the same QDs, plotted in blue. The radially symmetric QD breathing mode scatters inelastically at a frequency matching the picosecond oscillations in the transient SHG data.

(a) In the field picture second harmonic generation occurs via a nonlinear polarization response. The nonlinear polarization emits a field with components at the second harmonic frequency. (b) In the photon picture, second harmonic generation occurs when two photons with energy $h\omega$ combine to form one photon of energy $2h\omega$. The total photon momentum is conserved.

Nonlinear polarization response. When a sinusoidal field $\vec{E}(t)$ (black, dashed) is applied to a medium, it generates a polarization response. The polarization response of a linear (red), nonlinear centrosymmetric (black, solid), and nonlinear noncentrosymmetric (blue) are shown.
A two level system. State \( j \) has population \( N_j \) and energy \( E_j \). States 1 and 2 are separated by energy \( h\nu_0 \).

(Color online) The stimulated SHG experiment (a) A schematic of the optical system used to perform stimulated SHG. (PMT: photomultiplier tube) (b) A visualization of the signal modulation scheme used to differentiate SHG signal from stimulating photons. (c) Stimulated SHG in \( \beta \)-barium borate.

Analysis of stimulated SHG using the coupled-wave formalism. Open circles are data from BBO, while black lines are analytical fits predicted by eqns. (4.14–4.16) (a) Solution of eqns. (4.14–4.16) with initial condition \( \theta = -\frac{\pi}{2} \). (b) Dependence of the stimulated SHG signal on the composition of the incident fields. (c) Dependence of the signal enhancement \( \gamma \) on the total incident intensity. (d) Dependence of the stimulated SHG signal on the fundamental intensity. (e) Dependence of the stimulated SHG signal on the stimulating intensity.

Effect of phase on stimulated SHG (a) Relative optical phase relations that lead to stimulated SHG (left) and DFG (right). (b) Effect of stimulated SHG and DFG on the signal phase \( \Omega \) measured by the lock-in amplifier. (c) Dependence of the lock-in detected signal on the relative optical phase \( \theta \). Grey regions indicate an SHG signal in phase with the reference \( (\Omega = 0^\circ) \), whereas the white regions represent a stimulated DFG signal measured to be out of phase with the reference \( (\Omega = 180^\circ) \).

Illustration of stimulated SHG in collagen I. At low incident powers, stimulated SHG produces more than \( 10^4 \) times as many signal photons as spontaneous SHG. An optical micrograph of the collagen I sample is presented as an inset.
4-8 When a local oscillator pulse follows behind the signal pulse with a delay, $\tau$ it interferes with the signal to produce the illustrated spectrum. The spectrum oscillates between constructive and destructive interference as different frequency components of the pulses have differing relative phases. The red line shows the local oscillator and signal spectra summed in the absence of interference effects for comparison.

4-9 Demonstration of the difference between heterodyne SHG, as classically implemented (a), and stimulated SHG, as described in [4] (b). In stimulated SHG or DFG, power is transferred between the fundamental and second harmonic fields.

4-10 Phase-modulated degenerated parametric amplification (DPA) microscopy. (a) Schematic illustration of the optical apparatus and associated components. In this embodiment, $\lambda = 830$ nm and 415 nm pulses generated by a Ti:sapphire laser are cross-polarized at the sample and the fundamental field intensity is measured in transmission using a standard photodiode. (b) Illustration of the working principle of the technique: the electro-optic modulator (EOM) modulates the relative optical phase, $\phi$, between the two laser pulses at 1 MHz according to a saw tooth waveform at $2\pi$ modulation depth, generating a sinusoidal variation of the intensity in each field. (c) DPA signal amplitude map of a triangular monolayer MoS$_2$ flake. (d) Bright-field optical image of the same flake.

4-11 Definition of the coordinate system used in this section relative to the MoS$_2$ crystal lattice. The angle $\theta$ is the positive angle measured from the $x$ axis to the fundamental field polarization vector. The fundamental and second harmonic fields are cross-polarized.
4-12 DPA imaging of MoS$_2$ crystals. (a,b) Rotational dependence of the DPA amplitude and phase for the triangular MoS$_2$ single-crystal flake shown in the inset. Red and blue colors indicate a lock-in phase of 0° or 180°, respectively. (c) Bright-field image of the multi-grain MoS$_2$ flake analyzed in panels d-i. (d,e) DPA amplitude and phase maps. (f,g) Line cuts taken along the dash-dotted lines through the images in panels d and e. (h) Universal curve used to determine the crystal orientation of each grain labelled in panel d. (i) Illustrated crystal orientation of the seven grains labelled in panel d.  

4-13 Characterization of grain boundaries. DPA amplitude (a-c) and phase (d-f) mapping near a mirror twin boundary in a monolayer MoS$_2$ flake.  

31
List of Tables

2.1 Parameters used in the Monte Carlo modelling of exciton diffusion, trapping, and annihilation. The band edge diffusivity \( D_{\text{band edge}} \), trap density, and detrapping attempt frequency \( f_{\text{detrap}} \) were the same when modelling each substrate. The trap state exponential tail parameter \( \alpha \) and annihilation radius \( R \) each depended on the supporting substrate in the system being modelled. 78
Chapter 1

Motivation for Low Dimensional Materials

This thesis focuses on exciton transport and dynamics in zero dimensional (0D) quantum dots (QDs), two dimensional (2D) transition metal dichalcogenides (TMDs) as well as coupling at interfaces between those two materials classes. The unique properties of 0D and 2D nanostructured semiconductor materials make them strong candidates for next generation optoelectronic technologies in lighting, computing, and energy conversion. In order to develop next generation technologies, it’s crucial to first understand exciton and charge dynamics in these nanostructured materials.

1.1 2D TMDs for Lighting and Computing

The family of TMDs with chemical formula MX$_2$ (M a 2+ transition metal cation, X a chalcogen anion) contains many members and exhibits structural and electronic diversity (illustrated in Fig. 1.1). Depending on the chemical composition, 2D TMDs exhibit band gaps suitable for metal, semimetal, or semiconducting applications [5]. Insulating hexagonal boron nitride, metallic graphene, and other 2D materials can easily be incorporated into van der Waals heterostructures with tunable characteristics [6].

Monolayer TMDs exhibit many characteristics that arise from their atomically
Figure 1-1: Members of the layered TMD family typically possess one of three crystallographic symmetries (2H, 1T, or 1T'). The thermodynamically favored structural phase depends on material composition with some compositions possessing metastable phases.

thin nature. They are mechanically flexible and can accommodate large mechanical strains [7]. This property enables flexible, transparent transistor technologies [8], which would be difficult to achieve with traditional bulk semiconductors. TMDs are sensitive to their external environment; this enables facile tuning through electrostatic doping [9], applied external fields [10], the dielectric environment [11], and mechanical strain [12, 13].

Despite their atomically thin nature, TMDs possess moderate to high charge carrier mobilities [14-17]. For example, monolayer MoS$_2$ exhibits a room temperature charge carrier mobility of $\sim 200\ \text{cm}^2\text{V}^{-1}\text{s}^{-1}$ when encapsulated in a high index dielectric. This property is an important strength when using TMDs in optoelectronic devices.
strongly bound excitons - Due to reduced dielectric screening, excited electron-hole pairs in TMDs form strongly bound (300-400 meV) excitons. The large exciton binding energy corresponds to a highly localized exciton (Bohr radius \( a_B \approx 1 \text{ nm} \)). The coalescence of many band-edge states at the K point to a single 1s excitonic state results in strong light-matter interactions. This manifests in fast exciton radiative recombination \( (k_{\text{rad}} \sim 10 - 100 \text{ ps}) \) [18, 19] and large absorption cross section at the exciton resonance \( (A_0 \approx 0.1 - 0.3 \text{ at the exciton resonance}) \) [3, 20]. This strong light-matter interaction is important for photodetectors and energy conversion applications [21] and to phenomena that require strong coupling to the excitonic state such as excitonic energy transfer [22] and the formation of exciton-polaritons [23-26].

Excitonic light emitting devices fabricated from monolayer TMDs have been demonstrated [27-31]. Withers et al. reported a tunnel-junction LED with external quantum efficiency approaching 10% at room temperature. The design incorporated hexagonal boron nitride tunnel barriers and graphene electrodes. The entire device could be constructed on a transparent, flexible substrate [31]. TMD LED device efficiency will improve as higher quantum yield materials are obtained.

One potential mechanism by which 2D LEDs can outperform traditional organic LED architectures is via their in-plane oriented transition dipole moment. In traditional organic LED architectures, external quantum efficiency is fundamentally limited by light extraction bottlenecks resulting from waveguiding and coupling to lossy surface plasmons [32]. These fundamental outcoupling loss pathways are mitigated with the emitter transition dipole orientation is parallel to the device surface [33]. Orienting the emitters parallel to the device plane increases the efficiency of such a device architecture by 10% relative to the case in which excitons are isotropically oriented all other things being equal. Though this can be done with anisotropic solution-phase materials [34], using such materials requires controlled placement of the emitters in the device. In contrast, the transition dipole moment in TMD layers is oriented in the plane of the TMD [35], which is optimal for light outcoupling. In addition to potential efficiency gains due to improved light outcoupling, 2D TMD LEDs have been designed to take advantage of the spin-valley coupling in TMDs. Circularly
polarized emission of controlled handedness has been observed from valley-polarized excitons in LEDs [31].

There is some effort to use atomically thin TMDs to improve upon traditional silicon transistors in key metrics [36, 37]. MoS$_2$ possesses a relatively low in-plane permittivity and heavier effective carrier mass than silicon enabling record-breaking small channel lengths in a carbon nanotube-gated FET device [36] with workable source-drain leakage currents suitable for certain low power applications.

1.2 Quantum Dots for Displays and Lighting

Colloidal QDs are nanometer-scale semiconductor crystals with surface-bound organic ligands that impart colloidal stability. Because they can be prepared as colloids, QD suspensions can be deposited with processing techniques not normally accessible to bulk semiconductors such as inkjet and roll-to-roll printing. In contrast to bulk semiconductor crystalline materials, which are typically grown at high temperature and require compatible substrates with matching lattice constants, colloidal quantum dots can be deposited from solution and do not require substrates amenable to high temperature processing.

The defining feature of colloidal QDs is their easily tunable optical gap. QDs are grown in a size regime in which their carriers are subject to quantum confinement (particle sizes smaller than the parent bulk material’s Bohr radius). In this regime, the bulk band-like states near the band gap give way to discrete atom-like electron and hole states that can be described by solutions to the particle in a sphere model. These states are similar to the electronic states of a hydrogen atom and possess analogous principal quantum numbers and angular momenta. Synthetically tuning the QD size in the quantum confined regime effectively changes the size of the potential well invoked by the particle in a sphere model, shifting the absolute energy and relative energies of the resulting atom-like states. In this way, the QD optical gap is tied to the QD size and can be easily synthetically tuned without changing the QD chemical composition.
The result is a valuable technology useful for optoelectronic devices that require tunable absorption and/or emission. QDs exhibit broad band absorption with tunable onset and narrowband emission with tunable emission energy. Since their discovery, QDs with high photoluminescence quantum yield have been obtained. These properties make QDs an attractive technology for ambient lighting and display applications [1].

![Diagram](image-url)

**Figure 1-2:** (a) CIE chromaticity diagram. QDs exhibit narrow emission line widths and can be tuned to emit across the visible spectrum. Their color purity allows mixtures of dots to accurately reproduce a large fraction of the humanly perceivable color space. (b) Plot showing luminous efficacy (lm W⁻¹) and color rendering index (CRI, a metric indicating a light source's ability to accurately render colors). Due to their color purity and spectral tunability QD-based ambient lighting devices can be engineered with good luminous efficiency and CRI. Figure reproduced from ref. [1], Nature Publishing Group.

Narrow emission linewidths allow mixtures of QDs to accurately reproduce a large portion of the humanly perceivable color space. Any visible light field can be expressed by its intensity spectrum $I(\lambda)$. When this spectrum is observed by the human eye it is projected into a three dimensional vector space, because the human eye contains three types of color sensors with different spectral responses. This three dimensional projection uniquely represents the color observed by the human eye (a perceived color can originate from multiple unique spectra $I(\lambda)$). We break this three dimensional color into its chromaticity (a two dimensional subspace of human perceivable color) and its luminance. A color's chromaticity represents its hue, and that two dimensional
space is mapped in the CIE chromaticity diagram shown in Fig. 1.2a. The border of the chromaticity diagram represents peaked spectra with infinitely narrow spectral width. Colors within the diagram can be formed as linear combinations of other colors. In order to accurately produce a wide color gamut covering a large fraction of this space, it is important to have emitters with narrow line widths. The narrow emission linewidth accessible with QDs allows for the accurate reproduction of a large fraction of human-perceivable colors. For this reason, QDs have been commercialized in display technologies where accurate color reproduction is important.

Additionally the suitability of ambient lighting for color reproduction is quantified by a metric called the color rendering index (CRI). Ambient lighting with higher CRI is generally more pleasant, and accurately reproduces object colors in a scene. QD based light solutions are capable of beating many traditional lighting technologies in CRI while maintaining workable luminous efficacies. The state of the ambient lighting field and QD lighting’s place in that field is shown in Fig. 1.2b. QDs are a highly attractive technology for display and lighting applications due to their size-tunable narrowband emission and their solution processability.
Chapter 2

Exciton Transport and Dynamics in Transition Metal Dichalcogenide Monolayers

2.1 Excitons in Transition Metal Dichalcogenides

2.1.1 Exciton Background Information

Excitons are electronic excited states of materials in which the excited electron remains bound to the hole from which it originated through a coulombic interaction. There are different types of excitons that are classified based on the strength of the electron-hole interaction. Typically the strength of the interaction is mediated by the polarizability of the medium. Whether excitations create bound excitons or dissociated free charges depends on the thermal energy available to the exciton, $k_BT$, and the exciton binding energy, $E_X$, which is the difference in energy between the exciton state and uncorrelated charge state. The fraction of bound excitons in a population of excitons can be approximated using Boltzmann statistics:

$$f_{\text{bound}} = 1 - \exp \left( \frac{-E_X}{k_BT} \right). \tag{2.1}$$
In materials with appreciable permittivities such as bulk inorganic semiconductors (e.g. silicon, germanium, III-V semiconductors, metal oxides), the electron-hole Coulomb interaction is well screened and the binding energy is relatively small ($E_X < 100 \text{ meV}$). Such excitons, called Wannier-Mott excitons have large spatial extent ($\sim 5 - 10 \text{ nm}$) and behave analogously to a hydrogen atom composed of charges with reduced effective mass and with a strongly screened Coulomb interaction. Consequently, the exciton binding energy is much smaller than the hydrogen atom ionization energy. Some semiconductors exhibit a progression of excitonic states analogous to the electronic energy levels of a hydrogen atom, which can be observed using absorption spectroscopy at low temperature [38]. A Wannier-Mott exciton is illustrated in Fig. 2.1.1a.

In materials with small permittivities (such as polymer solids, organic crystals, and inorganic crystalline insulators) photoexcited electrons remain strongly correlated with the hole from which the electron originates. The Coulomb attraction is poorly screened and the exciton binding energy can consequently be quite large ($E_X \approx 0.1 - 1 \text{ eV}$). In this regime, the exciton is highly localized; the bound exciton radius is less than a nanometer and often less than the crystal lattice constant. This type of exciton is called a Frenkel exciton, which is illustrated in Fig. 2.1.1b.
In nanostructured semiconductors, it is possible to induce prominent electron hole interactions by spatially confining photoexcited charges. Monolayer transition metal dichalcogenides for instance have exciton binding energies on the order of 200-400 meV [39-41]. This is much larger than one might expect based on the bulk materials' large permittivities [42]. The surprisingly large binding energy is caused by inefficient screening in atomically thin materials. Though the material itself is highly polarizable, the surrounding medium usually is not. Consequently, charges can still interact strongly by bypassing the polarizable medium in favor of the surrounding dielectric. This idea is illustrated in Fig. 2.1.1c.

In spherical nanocrystals, electron-hole interactions are forced by confining the charges in all three dimensions. Though it can be difficult to decouple the size-dependent valence and conduction band edge from the size-dependent exciton binding energy in quantum dots, Meulenberg et al. measured the band alignments using X-ray absorption and photoelectron spectroscopy [43]. They found that the exciton binding energy in CdSe quantum dots ranged from 0.2 – 1 eV, which is much larger than the binding energy in bulk CdSe (15 meV) due to confinement of the electron and hole to the same region. This motif is illustrated in Fig. 2.1.1d.

The strength of a material’s light-matter interaction with respect to a given transition \( g \rightarrow e \) is governed by the magnitude of the transition dipole moment, which is mediated by the excited and ground state wavefunction overlap: \( \mu_{eg} = \langle \psi_e | q r | \psi_g \rangle \). Strongly bound excitons typically possess large transition dipole moments and appear in semiconductor absorption spectra as narrow features red shifted from the electronic band gap by the exciton binding energy. Similarly, excitons can emit, but this process occurs in competition with thermal exciton dissociation.

### 2.1.2 Transition Metal Dichalcogenides Background

*Electronic properties of TMDs:* In 2010, Mak et al. isolated atomically thin MoS\(_2\) [20], building upon the Nobel Prize winning isolation of graphene in 2004 [44] and expanding the field of two dimensional (2D) crystalline materials to include transition metal dichalcogenides (TMDs). Similar to graphene, TMDs exhibit high carrier
moieties. In marked contrast to graphene, however, most TMDs possess a direct bandgap with energies suitable for semimetal, insulating, and – most promisingly – semiconducting behavior. Semiconductor devices have been built, including transistors [14, 27, 28, 45] that incorporate the atomically thin nature of the semiconducting channel and exceed state of the art silicon technologies in key transistor metrics such as channel length [36].

Excitons in TMDs: In addition to exhibiting favorable charge carrier mobilities, TMDs possess a strongly bound excitonic excited state that couples very strongly to the ground state [39-41]. Members of the diverse TMD family possess excitonic resonances that can emit from the near infrared to visible regions of the spectrum. The large oscillator strength of the excitonic transition in TMDs results in strong light-matter interactions, which have been leveraged in microcavities and waveguides to produce exciton-polaritons, strongly coupled light-matter excited states [23-26]. Despite their atomically thin nature, TMDs can be engineered to emit with near unity quantum yield [46, 47], an important property for light-emitting applications including LEDs [27-30] and lasers [48]. Unique to 2D materials, the exciton transition dipole moment is oriented in the TMD plane [35], which is advantageous for efficient outcoupling of light in both vertical and lateral LED architectures. Additionally, different TMDs can be mixed and matched to design multilayer TMD heterostructures engineered for a specific applications [49-52].

Dielectric confinement in TMDs: TMDs possess unique physical properties that arise from their crystal structure and atomically thin nature. The surrounding dielectric environment strongly influences monolayer TMDs’ electronic and excitonic properties. Charges in monolayer TMDs are poorly screened despite generally large permittivities. This reduced dielectric screening manifests in several interesting phenomena: nonhydrogenic excitonic Rydberg series [39, 40], tuning exciton and trion energies using the external environment [11], strong tunable interlayer coupling in TMD heterostructures [24, 50-52], and prominent many-body effects [53] including strongly bound trions [3, 41, 54] and biexcitons [55].

Valleytronics in TMDs: Due to the lack of inversion symmetry in most monolayer
TMDs, excitonic excited states and free charge carriers possess a valley degree of freedom [56], which is largely analogous to electron spins. Excitons possess a valley index of +K or –K, and populations of a given index can be selectively excited using circularly polarized light [57–59]. Similar to electron spins, the energy of these valley-indexed exciton states can be tuned with an external magnetic field in a physical phenomenon termed the Valley Zeeman Effect [60]. In addition to the unique physics that arise from this binary valley degree of freedom, technologists envision encoding and manipulating information with it in an emerging field, valleytronics [56]. Toward this end, scientists have demonstrated the physical manipulation of charges using the Valley Hall Effect [61, 62], and the electrically driven generation of circularly polarized light using valley-polarized excitons [63].

2.2 Exciton Trapping Is Responsible for the Long Apparent Lifetime in Acid-Treated MoS$_2$

Much of Section 2.2 has appeared in print as Goodman, Willard, and Tisdale Phys. Rev. B (2017), [64].

2.2.1 Background for Chemical Treatment of MoS$_2$ Samples

As-exfoliated and CVD-grown MoS$_2$ samples exhibit poor quantum yields ($\sim$ 0.1-1%), which severely limits their use in optoelectronic devices that require efficient emission or long distance exciton transport. Monolayer MoS$_2$ is susceptible to sulfur vacancies [47, 65]. Scanning tunneling microscopy reveals single sulfur vacancies ($V_S$) as well as more complex structural defects such as disulfur vacancies ($V_{S2}$), molybdenum sulfur complex vacancies ($V_{MoS3}$, $V_{MoS6}$) and antisite defects ($Mo_{S2}$, $S_{2Mo}$) [65]. With the exception of the antisite defects, the structural defects remove negative charge from the TMD, which must be compensated by free electrons to maintain charge neutrality. As a result, as-prepared MoS$_2$ monolayers are usually $n$-doped.

In 2015, Amani et al. reported a chemical superacid treatment [46] that increases
the photoluminescence quantum yield from < 1% as-exfoliated to > 95%. Simultaneously, they observed an increase in the photoluminescence lifetime from roughly 250 ps to 10 ns. Though the treatment resulted in no observable structural changes, transistors prepared from the acid-treated MoS$_2$ exhibited intrinsic transfer characteristics in contrast to the n-type as-prepared material. Further studies extended the superacid treatment to WS$_2$ [47], but found the treatment did not improve photoluminescence quantum yield in selenium-based TMDs. Tantalizingly, the superacid treatment was found to be moderately effective in improving the quantum yield in large area CVD-grown MoS$_2$ [66].

There are two other notable papers that investigate the chemical treatment of MoS$_2$. Kim et al. used a capping fluoropolymer layer to make the superacid treatment previously reported by Amani et al. more robust [67]. Kim et al. coated MoS$_2$ flakes with a layer of CYTOP and then treated the flakes with the superacid, bis-trifluoromethylsulfonimide (TFSI). In the absence of the capping polymer, treated MoS$_2$ samples did not retain their improved quantum yield upon rinsing or exposure to vacuum, while the capping polymer protected the treated MoS$_2$ against those conditions. Kim et al. also saw that increasing the capping polymer thickness increased the reaction time necessary for effective superacid treatment. Through detailed analysis, they showed that the treatment required the TFSI molecule to diffuse through the capping polymer and reach the MoS$_2$ surface.

Atallah et al. reported a different route to improving the quantum yield of CVD-grown MoS$_2$ [68]. Atallah et al. found that the quantum yield could be improved by covering MoS$_2$ flakes with an ionic liquid. The ionic liquid only acted to improve the quantum yield when the individual MoS$_2$ flakes were connected to ground via an ohmic contact however. While the ionic liquid is unable to oxidize the n-type MoS$_2$ on its own, it is able to passivate nonradiative recombination centers when free charge-compensating electrons are able to leave the MoS$_2$ flake.
2.2.2 Background for Radiative Rates and Apparent Decay Rates in TMDs

Absorption and photoluminescence spectroscopy reveal the high material quality achievable with the bis(trifluoromethane)sulfonimide (TFSI) superacid treatment described by Amani et al. The absorbance and emission spectra of acid-treated MoS$_2$ are plotted in Fig. 2-2. The absorbance spectrum in Fig. 2-2a exhibits two narrow absorbance features corresponding to the A and B excitons originating from the spin-split conduction band at the K point of the Brillouin zone. The peaked absorbance
reflects the excitonic nature of the transition. The red line shows the narrow photoluminescence spectrum exhibiting clean luminescence from the A exciton. Fig. 2-2b shows the photoluminescence decay dynamics. The data were collected at sufficiently low fluence to avoid exciton-exciton interactions as evidenced by the monoexponential decay. The observed lifetime is roughly 20 ns, consistent with prior work.

Extensive theoretical work has been done to predict radiative lifetimes in 2D quantum wells [69], and later TMDs [19]. Such approaches begin with Fermi’s Golden Rule, which describes the transition rate between two states, $\Gamma_{i\rightarrow f}$:

$$\Gamma_{i\rightarrow f} = \frac{2\pi}{\hbar} |\langle \psi_f | q_r | \psi_i \rangle|^2 \rho,$$

where $\langle \psi_f |$ and $| \psi_i \rangle$ are the crystal ground state and excited exciton state respectively, $q_r$ is the dipole operator, and $\rho$ is the density of states available to the emitted photon. The term, $\langle \psi_f | q_r | \psi_i \rangle$ is the transition dipole moment and can be rewritten in terms of the momentum operator:

$$\langle \psi_f | q_r | \psi_i \rangle = \frac{i\hbar}{E_X m} \langle \psi_f | p | \psi_i \rangle$$

where $E_X$ is the exciton energy, $m$ is the exciton mass, and $p$ is the momentum operator. The transition dipole moment of an exciton is roughly proportional to the exciton momentum expected value (as long as there is overlap between the exciton and hole state). Strongly bound excitons with small exciton Bohr radii such as those in TMDs have large transition dipole moments and correspondingly large radiative rates. At zero Kelvin, an exciton in a TMD is expected to have a radiative lifetime [19]

$$\tau_{rad}^0 = \frac{2}{\Gamma_0} = \frac{\hbar \epsilon}{2k_0} \left( \frac{E_X}{\epsilon \hbar v} \right)^2 (a_B)^2.$$

where $\epsilon$ is the dielectric constant of MoS$_2$, $k_0$ is the magnitude of the wavevector of exciton emission, $v \approx \sqrt{E_g/(2m_\epsilon)}$ the Kane velocity, and $a_B$ the exciton Bohr radius. However, at finite temperature, only a small subset of a thermalized exciton
population can emit while satisfying momentum conservation. This effect modifies the radiative rate according to

\[ \tau_{\text{rad}}^{\text{eff}} = \frac{3 k_B T}{2 E} \tau_{\text{rad}}^0, \]  

(2.5)

where \( E \) is a bound on the kinetic energy an exciton can possess and still emit radiatively. Using values for the MoS\(_2\) dielectric constant [42] and exciton Bohr radius [53] from the literature, equation (2.5) implies a room temperature exciton lifetime of 470 ps. This is in reasonable agreement with the results of first principles calculations, which predict a 270 ps radiative lifetime at room temperature [18], but is more than 40 times smaller measured lifetime seen experimentally (Fig. 2-2b).

### 2.2.3 Trapped Exciton States in MoS\(_2\)

We believe long-lived trapped excitons explain the difference between the expected and measured decay times. Crystallographic defects are known to permeate exfoliated MoS\(_2\) [65, 70], and persist even in well passivated acid-treated MoS\(_2\) [46, 47]. Furthermore, sub-band emission has been observed in acid-treated MoS\(_2\) and attributed to defect-bound excitons [71]; here we investigate the interaction of these states with the band-edge states. Though not evident in room temperature absorbance or photoluminescence spectra, long-lived emission from sub-band traps comprises a large fraction of the photoluminescence spectrum at cryogenic temperatures. Fig. 2-3a presents photoluminescence spectra acquired at many temperatures between 77 K and 300 K. The spectra are normalized to the maximum emission intensity below the band-edge exciton to highlight and compare trapped exciton emission. The percentage of the total emission intensity from the trapped excitons is presented in Fig. 2-3b. At high temperature, nearly all of the emission comes from the band-edge exciton, whereas at 77 K, nearly all emission comes from trapped excitons. The intensity and spectral distribution of trap state emission varies from sample to sample, but its presence is observed in both exfoliated MoS\(_2\) and CVD-grown samples (see Supporting Information).
Figure 2-3: Temperature-dependent photoluminescence spectra (a) Photoluminescence spectra (thin lines) at low temperature exhibit emission at energies below the band-edge exciton. The lineshape is fit well by an exponential density of states occupied by a Fermi-Dirac Distribution (thick lines). (b) At low temperature, trapped exciton emission dominates the photoluminescence spectra, while at high temperatures most luminescence comes from the band-edge exciton. (c) Power-dependent photoluminescence spectra at 77 K reveal that the band-edge exciton emission becomes more prominent relative to the trapped exciton emission at higher laser powers. This observation indicates that a significant portion of the traps are filled at moderate excitation intensities.

The trapped exciton emission lineshape is well represented by a Fermi-Dirac distribution over an exponential tail of states extending into the band gap (thick lines), as shown by the fits in Fig. 2-3a. The density of traps states, \( \rho \propto \exp \left[ -\alpha (E - E_{\text{band edge}}) \right] \), is parametrized by \( \alpha \sim 5 \text{ eV}^{-1} \). Importantly, this distribution of trap state energies reproduces steady-state photoluminescence lineshapes as well as the spectrum of detrapping rates observed in transient trapped exciton photoluminescence decay histograms. We motivate the use of a Fermi-Dirac distribution by noting that the trapped exciton emission saturates at moderate laser power. Fig. 2-3c shows the 77
K emission excited under laser power spanning 4 orders of magnitude. At higher steady-state exciton population densities, fewer trap sites are available for exciton trapping, making band edge emission more likely.

We note that in all cases the effective temperature of the trapped exciton distribution is higher than the lattice temperature, indicating that the exciton population and lattice do not reach thermal equilibrium over the course of the exciton lifetime. This is rationalized by a model in which trap states—each having a well-defined energy—are spatially separated from one another. In this case, complete thermalization of the trapped exciton population is only achieved if the exciton is able to sample many different trap sites during its lifetime through multiple trapping/de-trapping and transport events.

### 2.2.4 Equilibration between Band-Edge and Trapped Exciton States

Fig. 2-4a illustrates the paths that a trapped exciton takes in this energy landscape. Following photoexcitation, band edge excitons are quickly captured into trap states. From there, trapped excitons can either decay radiatively to the ground state directly (the blue path in Fig. 2-4a) or use some of the available thermal energy to reach the band-edge before quickly decaying with the fast band-edge radiative rate (the red path in Fig. 2-4a). We probe these processes by monitoring the emission dynamics from the trap state (using a long-pass filter to isolate the trap state emission) at a set of temperatures between 77 K and 240 K using a 250 kHz diode laser. The results of these measurements are presented in Fig. 2-4b. At 77 K, there is not sufficient thermal energy to promote trapped excitons to the band edge; consequently, excitons decay radiatively from their trapped state with the trapped exciton lifetime ($\sim 4 \mu$s). In contrast, at 240 K there is sufficient thermal energy present to promote a portion of the trapped exciton population to the band-edge with rate $k_{\text{detr}} = k_{\text{trap}} \exp(-\Delta E/k_B T)$, where $\Delta E$ represents the depth of the trap state. This detrapping appears as fast decay components in the trap-state emission dynam-
Figure 2-4: A three-state model explains the observed photoluminescence decay time and the absence of trap state emission at room temperature. (a) Apart from the ground state and the band-edge exciton, there are also trapped exciton states that have \(~1000\times\) slower recombination rate than the band-edge exciton. At room temperature, thermalization up to the band edge and subsequent radiative recombination is most probable, whereas at 77 K direct recombination from trap to ground state is observed. (b) Time-resolved trapped exciton emission reveals these two pathways. At 77 K, trapped excitons decay with their slow radiative rate (\(~4\ \mu s\)). At higher temperatures, some thermalize up to the band-edge leaving behind only deeply trapped excitons, which decay slowly. The shaded region of the inset indicates the portion of the spectrum that is collected by the time-resolved detector.
ate temperatures exhibit a spectrum of detrapping rates reflecting the distribution of trapped exciton energies.

Using the three-state model illustrated in Fig. 2-4a, we can construct the rate equations governing band-edge and trapped exciton dynamics. Those equations are:

\[
\frac{d[X]}{dt} = R_{\text{gen}} - k_{\text{rad}}^{\text{band-edge}}[X] - k_{\text{trap}}[X] \left( 1 - \frac{[X_T]}{N_0} \right) + k_{\text{detrap}}[X_T]
\]

\[
\frac{d[X_T]}{dt} = -k_{\text{rad}}^{\text{trapped}}[X_T] - k_{\text{detrap}}[X_T] + k_{\text{trap}}[X] \left( 1 - \frac{[X_T]}{N_0} \right)
\]

where \([X]\) and \([X_T]\) are the band-edge and trapped exciton densities respectively, \(R_{\text{gen}}, k_{\text{rad}}^{\text{band-edge}}, k_{\text{trap}}, k_{\text{detrap}}\) are the rates of generation, band-edge radiative decay, trapped exciton decay, trapping and detrapping respectively, and \(N_0\) is the average area density of traps. The model represented by eqns. (2.6-2.7) captures many salient optical properties of MoS\(_2\). The absence of non-radiative decay pathways reflects the near-unity QY reported by Amani et al. [46]. The term \((1 - [X_T]/N_0)\) represents the fraction of trap sites that are currently unoccupied and reproduces saturation behavior as the generation rate is increased. The trapping rate and trap density can be found quantitatively by examining the data under steady-state continuous excitation (Fig. 2-3c) in the context of eqns. (2.6-2.7) at equilibrium when their left-hand-sides are zero. Assuming thermally-activated detrapping and using the \((k_{\text{rad}}^{\text{trapped}})^{-1} = 4 \mu s \text{ measured at } 77 \text{ K (see Fig. 2-4b)}, we find \(k_{\text{trap}} = 3 \times 10^{11} \text{ s}^{-1}\) and \(N_0 = 4 \times 10^6 \mu \text{m}^{-2}\). With the same parameters we can fit the out-of-equilibrium, transient data presented in Fig. 2-4b using a numerical ODE solver.

### 2.2.5 Observing the True Radiative Rate

Spectrally-resolved transient photoluminescence spectroscopy reveals the final relevant rate constant, \(k_{\text{rad}}^{\text{band-edge}}\). Fig. 2-5a shows the time-resolved photoluminescence spectrum recorded at room temperature. Nearly all of the emission is from the band-edge exciton with some initial contribution from the B exciton, consistent with steady-
state photoluminescence spectra.

Fig. 2-5b shows the time-resolved photoluminescence spectrum collected at a sample temperature of 77 K. In addition to the blue-shifted A exciton emission, there is prominent, long-lived emission from the trapped exciton states. These data were collected with a 76 MHz pulsed laser operating at 570 nm (2.17 eV) with 10 μW of power at the sample (50 μJ/cm²). During the time between pulses, the trap state emission is nearly static with the exception of a small, fast component due to hot trapped exciton luminescence. The high repetition rate laser generates $\sim 2 \times 10^{13}$ excitations$\cdot\mu$m$^{-2}\cdot$s$^{-1}$ whereas long-lived trapped excitons (despite their high area density) can only relax at a rate of $10^{12}$excitons$\cdot\mu$m$^{-2}\cdot$s$^{-1}$. Consequently, the trap sites are always nearly completely saturated and at equilibrium with the band-edge exciton state in this experiment. When the next laser pulse arrives, no empty sites are available for trapping and the only available relaxation pathway is direct radiative recombination to the ground state. Consequently, the fast decay of the A exciton emission reflects the temperature-adjusted intrinsic radiative lifetime. Fig 2-5c shows the decay dynamics of the A exciton emission integrated between 1.9 and 2 eV along with the experiment's impulse response function (IRF). The red line shows a single exponential convolved with the IRF and fitted to the experimental data. The extracted radiative lifetime is 150 ps. Extrapolating to room temperature using eqn. (2.5) this implies a room temperature lifetime $\tau_{\text{rad}}(300 \text{ K}) = 580$ ps in reasonable agreement with the estimate made using eqns. (2.4-2.5). The observed 20 ns exciton lifetime at room temperature (Fig. 2-2b) is $\sim$20 times longer than the lifetime of a free band-edge exciton, indicating that excitons in acid-treated MoS$_2$ spend $>95\%$ of their lifetime occupying trap states below the band edge.
Figure 2-5: Spectrally-resolved transient photoluminescence spectroscopy reveals the true radiative rate. (a) At room temperature, nearly all of the emission comes from the band-edge exciton, with a noticeable contribution from the B exciton. The exciton emission has long-lived components mediated by the long-lived, dark trapped exciton states. (b, c) At 77 K, the majority of emission comes from trapped excitons (1.65–1.85 eV). The pulsed laser operates at 76 MHz (13 ns between pulses), so the long-lived trapped exciton sites are always nearly completely occupied in this experiment. When a new pulse arrives, band edge excitons cannot trap and instead decay radiatively with a rate reflecting the true band-edge radiative rate, which is 150 ps at 77 K.
2.2.6 Numerical Modelling Details and Results

Figure 2-6: Power-dependent PL at 77 K. Power-dependent PL spectra (light gray) were fitted to band edge and trapped exciton contributions (colored lines) to extract density-dependent steady state populations.

Steady State Density Dependence – Power-dependent spectra were collected at 77 K to extract the trapping rate, \( k_{\text{trap}} \), and the trap state density, \( N_0 \). The power-dependent spectra were analyzed to extract density-dependent steady-state populations of the band edge \([X]\) and trapped \([X_T]\) populations. The spectra and their fits are plotted in Fig. 2-6. After integrating the counts from each state and considering the two states’ markedly different radiative rates, the fraction of the total population at the band edge, \([X]/([X] + [X_T])\) was found and used for modelling.

The expected power-dependent equilibrium populations were predicted by considering eqns. (2.6-2.7). In the CW PL spectroscopy experiment, the populations reach a steady-state, and their rates of change (the left hand sides of each equation) are zero. This establishes a set of two coupled equations in \([X]^\text{eq}\) and \([X_T]^\text{eq}\) with \( R_{\text{gen}} \) determined by the laser power. The equations were solved numerically while varying \( k_{\text{trap}} \) and \( N_0 \) to reproduce the experimentally found fluence-dependent band-edge population fraction, \([X]/([X] + [X_T])\). Fitted values \( k_{\text{trap}} = 3 \times 10^{11} \text{ s}^{-1} \) and \( N_0 = 4 \times 10^6 \mu\text{m}^{-2} \) best reproduced the data.
Transient ODE Modelling – The temperature-dependent trapped exciton emission (Fig. 2-4b in 2.2.4 reproduced in Fig. 2-7b here) was numerically modelled using eqns. (2.6-2.7). The time-dependent model was seeded with initial conditions $[X](t = 0) = 0$ and a trap state population determined by the experimental pulse energy. The pulse generated 100,000 total trapped excitations that were distributed over five energetic bins below the band gap. The excitations were partitioned into the bins according to the experimentally determined trap density of states. This partitioning is illustrated in Fig. 2-7a. Each set of initial conditions was numerically integrated and the solutions for each trap state energy were added together to generate a solution at each temperature to compare to the experimental data. With no free parameters, the model produces solutions that closely match the multi-exponential behavior shown in the data (Fig. 2-7b). The model captures behavior at low temperatures: trapped excitons do not have sufficient thermal energy to escape to the band edge. Conversely, at higher temperatures, excitons trapped at different energies can be thermally promoted to the band edge with different average rates determined by the depth of their traps.
Figure 2-7: Transient photoluminescence modelling. (a) The trap state energy spectrum used in the numerical model is illustrated here. The PL spectrum is plotted in red with the x-axis energy centered about the band edge exciton energy, $E_x$. The density of trap states is exponentially decaying from the band edge with a slope $\alpha = 5 \text{ eV}^{-1}$ found from fitting the PL spectra. This exponential decay is plotted in blue. The transient ODE model was solved five times at each temperature with trap state energies represented by the blue rectangles. The weight of each solution parameterized by a specific trap state energy was determined by the density of states. The five solutions for each temperature were added together to generate the model’s output. (b) The experimental trapped exciton emission at multiple temperatures is plotted along with the model output, which was generated with no free parameters.

### 2.2.7 Trap States in CVD-Grown MoS$_2$

Figure 2-8: Temperature dependent PL spectra from acid-treated CVD-grown MoS$_2$. The CVD-grown sample also exhibits trapped exciton emission at low temperature. The trapped exciton emission line shape differs reflecting a different energetic distribution of trap states.
Multiple exfoliated samples as well as CVD-grown samples exhibit the trapped exciton emission at low temperature. Notably, the trapped exciton emission varies in prominence from sample to sample, and the exact lineshape also varies. This suggests that the average area-density and energetic distribution of traps may vary sample to sample, supporting our hypothesis that the trap states are determined by the quality of the initial starting material rather than caused by the superacid treatment itself. Importantly, the dark traps influence exciton behavior even when not apparent in room temperature absorption and emission spectra.

Here, spectra from a CVD-grown sample are plotted in Fig 2-8. The trapped exciton emission was fitted to an exponentially decaying density of trap states filled with a Fermi-Dirac distribution as in the 2.2.3, but the fitted parameter describing the energetic distribution of trap states was found to be $\alpha \sim 10 \text{ eV}^{-1}$ as opposed to the $\alpha \sim 5 \text{ eV}^{-1}$ found in the exfoliated sample.

2.2.8 Conclusions

![Chemical structure](image)

Figure 2-9: Proposed superacid treatment mechanism. The TFSI acid chemically oxidizes the MoS$_2$, removing dopant electrons. The conjugate base anion then passivates the structural defect.

These results, in combination with recent work in the field [47, 67, 68], lead us to speculate with regards to the mechanism of photoluminescence enhancement. As
discussed in 2.2.1, monolayer MoS$_2$ is susceptible to sulfur vacancies [47, 65]. We hypothesize that the action of the superacid is to simultaneously passivate structural defects with the conjugate base anion (TFSI$^-$) and remove dopant electrons via chemical reduction of the acidic proton (2H$^+ + 2e^- \rightarrow H_2$). Sulfur vacancies in MoS$_2$ are known to be catalytically active toward hydrogen evolution [72]. Such a hypothesis is consistent with Kim et al.’s recent demonstration of enhanced acid-treated MoS$_2$ stability in the presence of a capping fluoropolymer (protecting the residual surface-bound TFSI anion) [67], and the superacid’s tendency to make n-doped materials more intrinsic [46]. These processes are illustrated in Fig. 2-9.

In agreement with this speculation, we hypothesize that the trap states observed here are not introduced by the superacid treatment. Rather, we believe they are associated with native structural defects (most likely sulfur vacancies) that survive the acid treatment [47]. In as-exfoliated samples, these trap states act as recombination centers increasing non-radiative recombination and reducing the QY. The action of the acid treatment is to dramatically reduce the rate of non-radiative recombination from trap-to-ground state, which together with the intrinsically slow radiative rate of this transition, leads to longer exciton lifetimes. Consequently, the superacid treatment does nothing but improve the photophysical properties of MoS$_2$. To achieve trap-free exciton dynamics limited only by intrinsic radiative recombination, more pristine starting materials are needed.

This work demonstrates that trapped excitons significantly affect room temperature exciton dynamics even though they don’t appear in absorbance or photoluminescence spectra at room temperature. The trapped excitons possess small transition dipole moments and degeneracy relative to the band-edge exciton states, which explains their absence in room temperature measurements. That these traps extend so far into the bandgap, however, ensures that these states play a role in the lives of excitons at room and elevated temperatures. It is important to understand the true energy landscape seen by excitons in acid-treated MoS$_2$ as the presence of deep traps could affect exciton transport and dynamics, processes critical to most excitonic optoelectronic devices. With better understanding of the identity, energy, and distribution
of these traps, it may be possible to alter the trap state energy distribution using the external dielectric environment or adjust the trap occupancy via electrostatic doping.

2.2.9 Experimental Methods

Sample Preparation – Flakes were prepared by mechanically exfoliating an MoS$_2$ single crystal (Graphene Supermarket). For most measurements, samples were exfoliated to Si/SiO$_2$ to aid in monolayer identification. Sample thicknesses were confirmed with Raman spectroscopy [73]. For absorbance measurements, flakes were exfoliated to a No. 1 glass coverslip. Samples were superacid treated following the procedure reported by Amani et al. [46]. bis(trifluoromethane)sulfonimide (TFSI) was dissolved into anhydrous 1,2-dichloroethane in a nitrogen glovebox forming a 0.2 mg/mL solution. Samples were immersed in a freshly prepared solution for 10 minutes. After removing the vial containing the sample from the glovebox, the sample was removed from solution and blown dry under nitrogen. The sample was then placed in a 100°C oven and heated in air for 5 minutes.

Steady State Spectroscopy – Spatially-resolved spectroscopy was performed in an inverted microscope. For steady-state spectra, a CW diode laser (Coherent, Sapphire SF, 532 nm) was focused at the sample (Nikon, CFI S Plan Fluor ELWD, 40×, 0.6 NA). Fluorescence was passed by a long-pass dichroic reflector and imaged at the entrance slit of a 0.5 m focal length spectrograph and dispersed onto a cooled charge-coupled device (Princeton Instruments, Pixis). Emission spectra were transformed from wavelength (nm) to photon energy (eV) using the Jacobian transformation [74]. Absorbance spectra were taken by illuminating a transparent substrate from above with a broad-band incoherent source (tungsten halogen lamp) and recording spectra transmitted through the substrate-only as well as the substrate/sample stack.

Time-Correlated Single Photon Counting – For the data shown in Fig. 2-2b and Fig. 2-5, a 76 MHz pulsed excitation source was used (Coherent OPO, PP automatic, 570 nm, 76 MHz, < 1 ps pulse duration). For long lifetime dynamics shown in Fig. 2-4b, a 250 kHz pulsed excitation source was used (Picoquant, LDH-D-C405M, 405 nm, 0.4 ns pulse duration). Fluorescence was detected with a silicon avalanche photodiode
(Micro Photon Devices, PDM50, 50 ps resolution at the detection wavelength), which sent a voltage pulse to a counting board (Picoquant, PicoHarp 300). Time-resolved spectra were acquired by first passing the fluorescence through a spectrograph and measuring decay histograms at each monochromated spectral slice.

Numerical Modelling – Steady-state spectra were modelled by taking the differential eqns. (2.6 & 2.7) at equilibrium resulting in a system of nonlinear equations that were solved numerically using fsolve in Matlab.

Time-dependent dynamics were modelled by numerically integrating eqns. (2.6-2.7) using the ODE solver in Matlab.

2.3 Suppressing Diffusion-Mediated Exciton-Exciton Annihilation in 2D Semiconductors Using the Dielectric Environment

2.3.1 Exciton Exciton Annihilation

Exciton exciton annihilation is a bimolecular process in which one exciton decays dissipating its energy nonradiatively by imparting it to the other exciton. The decay path is more prominent at high exciton densities, \(N\), and the rate at which the process occurs, \(R_{XX}\), is generally proportional the square of the exciton density due to the process' bimolecular nature, \(R_{XX} \propto k_{XX}N^2\). Here, \(k_{XX}\) is a rate constant that reflects the exciton diffusivity and annihilation capture radius in a given system. Exciton exciton annihilation removes one exciton from the system nonradiatively, and can consequently limit photoluminescence quantum yields at high exciton densities. Exciton exciton annihilation typically occurs via excitonic Auger recombination.

Excitonic Auger Recombination – Auger recombination is a known significant process in bulk semiconductors [75]. Auger recombination is a nonradiative three-carrier process in which an excited electron relaxes into an unoccupied state (hole) scattering off a third charge, which is then promoted in energy. Two possibilities for this process
Figure 2-10: *Three carrier Auger processes* (left) A relaxing electron scatters off another electron promoting increasing its energy (right) A relaxing electron scatters off of a hole, pushing it deeper into the valence band. In each case, energy and momentum are conserved.

are illustrated in Fig. 2.3.1. The relaxing electron can scatter off of another electron, which is then promoted to a higher energy state; alternatively, the relaxing electron can scatter off of a hole, pushing it into a lower energy state. In bulk inorganic semiconductors where excitons are likely dissociated at room temperature, Auger recombination is typically a three carrier process whose rate is governed by the number of states that the third charge can scatter into (conserving energy and momentum ± a phonon) and a matrix element $M_{ee}$ ($M_{eh}$), that describes the magnitude of the electron-electron (electron-hole) interaction [76].

In systems with strongly bound excitons, a similar process can occur when two excitons are separated by a small distance. For example, Sun *et al.* observed exciton exciton annihilation in monolayer molybdenum disulfide, which they attributed to an excitonic Auger process; they commented that annihilation rate was anomalously rapid relative to Auger recombination processes in III-V quantum well systems [77]. Kavoulakis and Baym exhaustively considered the Auger scattering processes available to interacting excitons in cuprous oxide [78]. Weng *et al.* extended that
theoretical treatment to Auger recombination of excitons in one-dimensional systems [79]. Compared to three-carrier Auger processes, there are more scattering pathways for a pair of excitons, because there is an additional charge present. The scattering processes are dictated by analogous matrix elements, which reflect the charge-charge Coulomb interaction and require wavefunction overlap [79]. Low-dimensional inorganic semiconductors foster strong exciton exciton interactions by limiting the dimensions along which excitons can move and by promoting strong Coulomb interactions due to reduced dielectric screening.

2.3.2 Exciton Diffusion

Wannier-Mott excitons diffuse via a similar mechanism to the constituent free carriers, while remaining strongly correlated. In this formalism, the exciton is localized and scatters with lattice phonons, which impart momentum to the exciton stochastically. This thermally activated transport assumes that exciton transport occurs via hopping (similar to Brownian motion) and not band-like transport. At nonzero temperature, on average excitons have $k_B T/2$ thermal energy per degree of freedom. This thermal energy represents the exciton’s interactions with phonons that drive diffusive motion.

Excitons have some average thermal velocity $v_{th}$ at which they move and scatter with frequency $1/\tau_c$. Consequently, on average excitons travel a mean free path $l$ between scattering events:

$$l = v_{th} \tau_c$$

(2.8)

The product of the mean free path and the thermal velocity is the diffusivity, $D = lv_{th}$, which describes the rate at which excitons diffuse. While the diffusivity can be difficult to measure directly, it can approximately related to the constituent charge carrier mobilities according to the Einstein equation:

$$D = lv_{th} = v_{th}^2 \tau_c = \frac{kT \mu}{q}$$

(2.9)
per degree of freedom.

A uniform distribution of randomly diffusing particles exhibits no net transport. Any particle moving in one direction is on average "cancelled out" by a particle moving in the other direction. The distribution of particles can however change in time when there is a nonzero density gradient. In particular, the time dependent particle distribution $N(r, t)$ evolves at a rate governed by the particle’s diffusivity according to Fick’s second law:

$$\frac{\partial N}{\partial t} = -D \nabla^2 N(r, t)$$  \hfill (2.10)

A particularly important solution to Fick’s second law is the time evolution of a Gaussian particle density distribution. When a distribution is initialized at $t = t_0$ to the Gaussian distribution:

$$N(r, t_0) = N_0 \exp\left(-\frac{r^2}{2\sigma_0^2}\right),$$  \hfill (2.11)

it evolves diffusively such that at time $t = t_0 + \Delta t$ it will have the broadened Gaussian distribution:

$$N(r, t_0 + \Delta t) = \frac{N_0}{\sqrt{4\pi D\Delta t}} \exp\left(-\frac{r^2}{2(\sigma_0^2 + 2D\Delta t)}\right).$$  \hfill (2.12)

Put another way, the variance of the distribution, $\sigma^2(t)$ evolves according to

$$\sigma^2(t) = \sigma_0^2 + 2D\Delta t.$$  \hfill (2.13)

In a typical experiment performed in a microscope, the microscope objective focuses the excitation laser to a Gaussian profile at the sample surface, where it produces a Gaussian distribution of excitons (when excited in the linear regime). By monitoring the time-dependent broadening of this exciton profile (the variance), it is possible to extract the diffusivity [80].

The dimensionality of the sample plays a couple of roles in the exciton diffusion
process. Excitons that are restricted to diffuse in two dimensions for example have fewer degrees of freedom and consequently a lower thermal velocity at a given temperature, which affects the absolute value of the diffusivity for systems with a given scattering frequency and mean free path.

2.3.3 Supressing Exciton-Exciton Annihilation

This section discusses the interplay between exciton diffusion and exciton exciton annihilation in MoS$_2$ and WS$_2$ monolayers. The rate of exciton exciton annihilation is tuned by altering the supporting substrate dielectric constant.

MoS$_2$ monolayers were mechanically exfoliated from bulk single crystals onto SiO$_2$/Si substrates. The samples were transferred to quartz, sapphire, or strontium titanate (STO) substrates chosen for their varied dielectric constant. MoS$_2$ samples were then encapsulated with a PMMA capping layer and treated according to the superacid treatment detailed by Amani et al. [46, 47]. The PMMA encapsulation mimics a similar strategy described in the literature by Kim et al. that uses fluoropolymer encapsulation to stabilize the superacid treatment against solvent washing and vacuum exposure [67]. An optical transmission micrograph of an exemplar flake is shown in Fig. 2-11a and the corresponding substrate/sample/polymer stack is illustrated in Fig. 2-11b.

After the superacid treatment, the PL of the MoS$_2$ monolayers supported on quartz, sapphire, and STO were markedly enhanced (normalized PL spectra are shown in Fig. 2-11c, where the peak energies and profiles are identical. Calibrated PL intensity was measured as a function of the generation rate, allowing the extraction of steady-state QYs, which are plotted in Fig. 2-11d. The QY series for MoS$_2$ monolayers supported by all three substrates are qualitatively similar; the QY is near unity at low generation rate and decreases as the generation rate is increased. The QY’s dependence on the generation rate can be described by the ratio of the radiative
decay rate to the sum of the rates of all decay paths available to the exciton:

\[ QY = \frac{k_X}{k_X N + k_{NR} N + k_{XX} N^2} \]  

(2.14)

where \( k_X \) is the radiative decay rate, and \( k_{NR} \) and \( k_{XX} \) are the first order nonradiative decay rate and the exciton exciton annihilation rate constants respectively. This recombination model successfully captures the efficiency droops at high generation rate where the QY drops precipitously because exciton exciton annihilation \( (k_{XX} N^2) \) begins to outpace radiative decay \( (k_X N) \). In treated MoS\(_2\), the QY is near unity at low generation rate, implying that \( k_{NR} \) is negligible. This allows us to extract \( k_{XX} = 0.8, 0.02, 0.005 \text{ cm}^{-2}\text{s}^{-1} \) for the samples on quartz, sapphire, and STO, respectively (the dash lines in Fig 2-11d are the fits using eqn. (2.14). The threshold generation rate at which the QY drops below 50% can be extended by two orders of magnitude; for samples on quartz, sapphire, and STO, those generation rates are \( 1.5 \times 10^{16}, 2.1 \times 10^{17}, \) and \( 1.1 \times 10^{18} \text{ cm}^{-2}\text{s}^{-1} \).

The exciton exciton annihilation rate constant can also be extracted from transient measurements. Time correlated single photon counting (TCSPC) measurements were performed at varied incident laser fluences (corresponding to different initial exciton concentrations, \( N(0) \)) to reveal recombination dynamics. By stitching together the decay curves with varied \( N(0) \), a single decay curve with over four decades of dynamic range is obtained. The decay curves for the treated MoS\(_2\) supported by all three substrates are shown in Fig. 2-11e. The decay curves are multiexponential containing fast components due to annihilation at high exciton density and slower radiative decay at low exciton density. The rates \( k_X \) and \( k_{XX} \) can be extracted by fitting the decay curves to a simple kinetic model, in which the excited exciton density, \( N(t) \), decays according to the equation,

\[ \frac{dN(t)}{dt} = -k_X N(t) - k_{XX} N^2(t). \]  

(2.15)

The values of \( k_{XX} \) obtained by this fitting for TCSPC are in good agreement with the values extracted from the steady-state QY measurements, as plotted in the top
Figure 2-11: Effect of substrate on the optical properties of superacid-treated MoS$_2$. (a) Transmission optical micrograph of exfoliated MoS$_2$ with monolayer region indicated. The inset depicts the chemical structure of the superacid used in the treatment. The dark thick line is the border of the polymer capping layer on top of the MoS$_2$ flake. (b) Schematic of the substrate/MoS$_2$/polymer stack. (c) Photoluminescence spectra of treated MoS$_2$ on quartz (blue), sapphire (green), and STO (red). (d) Steady-state QY measured as a function of exciton generation rate. Data were recorded for MoS$_2$ supported on quartz (blue trace), sapphire (green trace), and strontium titanate (red trace). (e) Time-resolved photoluminescence traces. The traces were globally fit to extract $k_X$ and $k_{XX}$ as described in the text. (f) Exciton-exciton annihilation rate constants, $k_{XX}$, inferred from steady-state QY measurements (open circles) and time-resolved photoluminescence measurements (filled squares). $k_{XX}$ values for MoS$_2$ on quartz (blue), sapphire (green), and strontium titanate (red) are plotted against the supporting substrates’ refractive indices (top). Analogous data for WS$_2$ samples are shown in the bottom panel.

panel of Fig. 2-11f. Notably, $k_{XX}$ can be similarly tuned in WS$_2$ using the supporting substrate; the values of $k_{XX}$ found in WS$_2$ are shown in the bottom panel of Fig. 2-11f. $k_X$ remains constant for MoS$_2$ on all three substrates. This differs from III-V thin film devices in which the radiative recombination rate is highly dependent on the optical mode density and refractive index of the medium [81]. The constant $k_X$
in MoS$_2$ indicates that the refractive index of the substrate affects the nonradiative recombination rate only.

### 2.3.4 Measuring Exciton Diffusion in TMDs

The exciton Bohr radius of the lowest-energy band-edge exciton in MoS$_2$ has been calculated to be roughly 5-10 Å [53]. In quartz-supported MoS$_2$, the photoluminescence QY dropped to 90% at a generation rate of 0.5 excitons $\mu$m$^{-2}$ per exciton lifetime (20 ns). One potential explanation for exciton exciton interaction at such a small generation rate is highly diffusive excitons that diffuse long distances before meeting and annihilating.

To probe exciton transport, we followed excitons’ motion in space and time using transient photoluminescence microscopy [80, 82]. The optical setup is depicted in Fig. 2-12a. A pulsed laser is focused to a diffraction-limited excitation spot at the sample using an oil-immersion objective, and the epi-fluorescence is collected by the same objective. A 360× magnified image of the fluorescing exciton population is scanned by an APD detector, which is synchronized to the pulsed laser to collect PL decay histograms. A PL decay trace was collected at each detector position in the image plane effectively recording a video of exciton transport and decay pixel-by-pixel.

The top panel of Fig. 2-12b depicts the simulated time evolution of an exciton population initialized with a Gaussian spatial profile, indicated by the dashed black trace, designed to mimic the exciton population instantaneously excited by a 0.5 nJ/cm$^2$ laser pulse focused to a diffraction limited spot ($\lambda = 405$ nm). At this fluence, excitons only decay radiatively. The exciton population decays exponentially as time progresses (coded in the trace colors). Simultaneously, excitons diffuse out of the initial excitation spot, broadening the distribution. We model this decay process in terms of a continuum model in which the exciton density, $N(r,t)$, evolves as a function of space and time according to

$$\frac{dN}{dt} = D \nabla^2 N - (k_X + k_{NR}) N - k_{XX} N^2,$$  \hspace{1cm} (2.16)
where $D$ is the exciton diffusivity. The bottom panel of Fig. 2-12b depicts an analogous simulation performed with a higher excitation fluence. In this regime, the exciton exciton annihilation term in eqn. (2.16) becomes prominent. Excitons still undergo radiative decay and diffusion, but additionally annihilate with a rate that depends non-linearly on the local exciton density. The non-radiative decay channel increases the overall decay rate of the exciton population and also changes the shape of the distribution. The peak of the excited distribution decays more quickly than the tails, leading to artificial flattening and broadening.

In Fig. 2-12c, we present the time-resolved spatial broadening of the exciton population measured in quartz-supported monolayer MoS$_2$ at three different incident laser fluences. The top panel depicts data in which each laser pulse excited on average 1.2 excitons per square micron. The white traces track the standard deviation of the distribution with time. For purely diffusive broadening, the change in the distribution variance grows linearly in time:

$$\sigma^2(t) - \sigma^2(0) = 2Dt$$

(2.17)

This behavior is seen at low fluences and fitting the variance as a function of time allows us to extract a diffusivity $D_{\text{qs}} = 0.06 \pm 0.01 \text{ cm}^2/\text{s}$, corresponding to a diffusion length $L_D = \sqrt{D\tau} = 350 \text{ nm}$. In contrast, with higher excitation fluences, the spot appears to broaden more quickly. However, this is due to faster rates of exciton exciton annihilation in the center of the distribution rather than faster exciton transport. In these cases, the variance of the intensity distribution $I(x, t)$ grows sublinearly in time. Such details are captured by the data and reproduced by the simulation parameterized by our measured values for $k_{\text{xx}}$ and $D$. We performed the same measurement for samples supported on sapphire and STO and extracted diffusivities $D_{\text{sapphire}} = 0.04 \pm 0.01 \text{ cm}^2/\text{s}$ and $D_{\text{STO}} = 0.06 \pm 0.02 \text{ cm}^2/\text{s}$. The choice of substrate did not appear to significantly affect the exciton diffusivity.

To corroborate the time-resolved measurement of the diffusion coefficient on quartz, we performed a separate measurement of the exciton diffusion length using steady-
state PL imaging. A CW laser was focused to a diffraction limited spot at the sample and the emission was imaged on a CCD camera. To calibrate the optical system, we first performed the experiment with a thick film of well-separated CdSe quantum dots, in which no exciton diffusion could occur [82]. This demonstrated that the convolved excitation and collection point spread functions (PSFs) of our microscope are nearly diffraction limited (339 nm measured vs 304 nm expected, see Fig. 2-12d). We then performed the same measurement with MoS₂ at sufficiently low fluence to avoid exciton exciton annihilation. Under the same conditions, the spot is broadened due to exciton diffusion, as shown in Fig. 2-12d. During the excitons’ lifetime, the variance increased by 0.026 µm² implying a diffusivity of 0.03 ± 0.01 cm²/s, which is consistent with the time-resolved measurements.
Figure 2-12: Exciton diffusion imaging. (a) Schematic diagram of the optical apparatus used to resolve exciton diffusion in space and time. (b) Top: simulation of the photoluminescence intensity along a line cut of the radially symmetric exciton population as a function of time. With an incident fluence of 0.5 nJ/cm², almost all excitons decay radiatively and the spot broadens due to diffusion. Bottom: same as top panel, but with an incident fluence of 50 nJ/cm², many excitons decay due to annihilation. This results in artificial broadening of the density profile; excitons decay most rapidly in the center of the spot where exciton density is highest. (c) Experimentally observed broadening of the exciton population with time at 0.5 nJ/cm², 5 nJ/cm², and 50 nJ/cm² incident fluences in the top, middle, and bottom panels, respectively. White lines indicate the evolution of the standard deviation with time. (d) Steady-state measurement of exciton diffusion. The intensity profile of luminescence coming from a film of quantum dots is indicated in green. Excitons do not diffuse in this sample, so the intensity profile indicates the performance of our optical system. The photoluminescence intensity profile collected from MoS₂ with a generation rate of $4.8 \times 10^{15}$ cm$^{-2}$s$^{-1}$ is broadened due to diffusion. The diffusive contribution to the width of the intensity distribution is indicated in red in the inset. (e) This cartoon illustrates the role that trapped excitons play in influencing the measured diffusivity. Though excitons diffuse quickly at the band edge with the band edge diffusivity, they fall into immobile traps. Consequently the measured diffusivity ($D_{\text{measured}}$) is much smaller than the true band edge diffusivity.
2.3.5 Exciton Trapping Influences Exciton Transport and Annihilation

The exciton diffusivity, which was measured by two independent methods (transient imaging and CW imaging), is surprisingly small. For Wannier excitons, one would expect the exciton diffusivity to be comparable to the diffusivity of the individual charge carriers. One can estimate the exciton diffusivity using the charge carrier mobility in MoS2, which has been measured extensively in the literature [14-17]. Single-layer MoS2 exhibits a room temperature charge carrier mobility of \( \sim 200 \text{ cm}^2\text{V}^{-1}\text{s}^{-1} \) when encapsulated in a high index dielectric. This mobility predicts a room temperature exciton diffusivity of 5 cm\(^2\)s\(^{-1}\) using the Einstein relation,

\[
D \approx \frac{\mu k_B T}{q_e},
\]

where \( \mu \) is the charge carrier mobility and \( q_e \) is the charge of an electron. The predicted exciton diffusivity is roughly two orders of magnitude larger than the observed diffusivity in MoS2.

This notable discrepancy points to the important role that trapped excitons play in exciton transport and dynamics in acid-treated MoS2. We found in our previous work that excitons equilibrate between free and bound states with exciton trapping energies ranging from 0.1-0.4 eV below the band edge [64]. Defect-bound excitons are well-studied in bulk semiconductors [83, 84]. Shallow defect-bound excitons exhibit smaller oscillator strengths due to localization effects, resulting in longer radiative lifetimes and slower recombination dynamics [84, 85]. In nanostructured semiconductors such as CdSe quantum dots, holes are strongly localized to unpassivated surface states leading to low energy emission deep below the delocalized exciton emission [86–88]. By analogy, excitons localized to neutral and/or structural defects in monolayer TMDs exhibit broad emission deep below the free exciton [64]. The trapped exciton states possess roughly 4 orders of magnitude smaller radiative rate and are hidden in room temperature absorption and emission spectra. This is consistent with highly localized charge-separated defect-bound excitons, an assignment supported [89] by
the dramatic solvatochromism exhibited by the bound exciton emission induced by the supporting dielectric (Fig. 2-13a).

At room temperature, excitons in acid-treated MoS$_2$ spend roughly 95% of their lifetime immobilized in trapped defect states below the band edge [64]. Equilibration between dark trapped exciton states and bright band edge exciton states lengthens the effective exciton lifetime from $\sim 500$ ps to $\sim 20$ ns at room temperature. The experimentally observed exciton diffusivity, which is a time-weighted average of the free and bound states, is small despite the fact that excitons may diffuse very quickly while at the band edge. This principle is illustrated in Fig. 2-12e. By dividing the measured diffusivity ($D_{\text{measured}}$) by the ratio of the band edge exciton radiative rate and the apparent radiative rate ($\tau_{\text{rad}}$ and $\tau_{\text{apparent}}$ respectively) we can infer the band edge diffusivity from our measurement:

$$D_{\text{band edge}} = D_{\text{measured}} \times \left( \frac{\tau_{\text{apparent}}}{\tau_{\text{rad}}} \right).$$

This results in a band edge diffusivity $D_{\text{band edge}} = 2$ cm$^2$ s$^{-1}$ ($\tau_{\text{apparent}} = 20$ ns), which is close to the expected diffusivity based on the charge carrier mobility in monolayer MoS$_2$.

The presence of trapped exciton states was confirmed using low temperature photoluminescence spectroscopy. Fig. 2-13a shows the photoluminescence spectra taken at 77 K for acid-treated MoS$_2$ supported on quartz, sapphire, and STO. Note that the PL intensity and spectrum are reversible after the temperature recovered from 77 K to room temperature. The spectra (plotted on a logarithmic scale) exhibit a dominant peak centered at roughly 1.95 eV that is characteristic of band edge exciton emission. The three spectra also exhibit low energy emission below the free exciton, which is not evident in room temperature photoluminescence spectra. These low energy features correspond to emission from weakly-radiative defect-bound exciton states [64]. The spectra in Fig. 2-13a are successfully fit to a linear combination of a Gaussian density of states corresponding to band-edge emission as well as an exponential tail density of trapped exciton states; in all three fits, the ratio of the degeneracy of the band
edge and trapped exciton states is held constant. The energetic distribution of the trap states and their steady-state occupation are inferred from the fit. In agreement with prior results [64], the majority of excitons are trapped in the steady state despite clean room temperature PL spectra. The relevant exciton states and the rate constants coupling those states are diagrammed in Fig. 2-13b.

Figure 2-13: Trapped exciton states and their equilibration with the band edge. (a) 77 K photoluminescence spectra of MoS$_2$ supported on quartz (top), sapphire (middle), and strontium titanate (bottom). At low temperature, trapped excitons emit at energies below the free exciton. The photoluminescence spectra (solid black lines) are modelled as a Fermi-Dirac distribution over an exponential density of states. (b) State diagram for trapped and band-edge (free) excitons.

We hypothesize that the substrate dielectric constant moderates the exciton annihilation rate by screening the interaction between trapped and freely diffusing excitons. We emphasize that our samples were exfoliated and chemically treated prior to transfer to the substrates used in these experiments. Though the starting MoS$_2$ samples should be of similar quality, it is possible that unfavorable interactions with
the substrate may create new exciton trapping sites. Another possibility is that the substrate alters the exciton band edge diffusivity. To rule out this possibility, we experimentally measured the exciton diffusivity and effective exciton lifetime, which were invariant from sample to sample, within experimental uncertainty.

To test our physical model of the substrate dependence, we performed Monte Carlo simulations of exciton transport and annihilation. The results of these simulations, shown in Fig. 2-14, quantitatively reproduced all of our experimental observations. Key parameters involving the interaction between trapped and freely diffusing excitons are illustrated in Fig. 2-14a. In the model, long-lived, immobile trapped excitons act as nonradiative recombination centers for diffusing excitons. When diffusing excitons come within a critical radius $R$ of a trapped excitons, they annihilate in a bimolecular process. While many trapped excitons have the opportunity to detrap and either decay radiatively or find another trapped exciton to annihilate, a non-negligible portion of the trapped exciton population (which spans roughly 400 meV) does not possess sufficient thermal energy to detrap. This subpopulation of deeply trapped excitons persists for a long time when compared to the apparent exciton lifetime (10s of ns) and the band edge radiative lifetime (less than 1 ns).

A simple kinetic model that includes the effects of trapped excitons quantitatively reproduces all of our spatiotemporal measurements of exciton transport and dynamics in MoS$_2$. This model describes excitons as mobile particles in a two-dimensional landscape that includes explicit spatially localized traps. Excitons were initialized at the band edge, with a spatial distribution coinciding with that of focused laser pulse, and then allowed to diffuse in two dimensions. The dynamics of an exciton persisted until the exciton underwent either radiative decay, simulated stochastically with a rate $k_X$, or exciton exciton annihilation, by diffusing within a separation $R$ of another exciton. Exciton trapping and detrapping was simulated using a Monte Carlo algorithm. Namely, when an exciton encountered a trap, it trapped with unity probability and became immobilized. A trapped exciton detrapped stochastically
Figure 2-14: Monte Carlo simulation of trapped excitons contributing to annihilation. (a) Diffusing excitons meet long-lived trapped excitons and annihilate. The equilibrium between trapped and free excitons is responsible for the large annihilation rate constants. (b) A Monte Carlo model invoking trapped and diffusing band edge excitons quantitatively reproduces steady-state QY data. The model output is plotted here (thick transparent lines) against the experimental data (open circles) for MoS₂ on quartz (blue), sapphire (green), and strontium titanate (red). (c) The annihilation rate constant inferred from the model (crossed diamonds) is plotted along with the experimentally derived values (open circles, QY; filled squares, time-resolved photoluminescence). The annihilation rate is plotted against the supporting substrates' refractive indices.

with a rate determined by the energetic depth of the trap it occupied, as given by,

\[
k_{\text{detrap}} = f_{\text{detrap}} \exp \left[ -\frac{\Delta E_{\text{trap}}}{k_B T} \right], \tag{2.20}
\]

where \( f_{\text{detrap}} \) specifies the base detrapping attempt frequency, \( \Delta E_{\text{trap}} \) denotes the depth of the trap, and \( k_B T \) is the Boltzmann constant times temperature. Formulated in this way, the trapping and detrapping rates obey detailed balance. Traps were distributed randomly in space and each trap was assigned a random trapping energy, \( \Delta E_{\text{trap}} \), distributed on the interval (0.15, 0.4) eV and weighted by exponential distribution \( \rho(\Delta E_{\text{trap}}) \propto \exp[-\alpha \Delta E_{\text{trap}}] \).

### 2.3.6 Results and Conclusions

This model requires very few parameters for input, most of which are available from experiment or the literature. The parameters and their values used in the model are enumerated in Table 2.1. The band edge diffusivity was modelled as \( D_{\text{band edge}} = 5.0 \)
Quartz & Sapphire & Strontium Titanate \\
\hline
$D_{\text{band edge}}$ [cm$^2$s$^{-1}$] & 5.0 & 1000 & $5 \times 10^{13}$ \\
Trap Density [\mu m$^{-1}$] & & & \\
$f_{\text{detrap}}$ [s$^{-1}$] & 6.0 & 7.5 & 11 \\
$\alpha$ [eV$^{-1}$] & 4 & 0.6 & 0.4 \\
$R$ [nm] & & & \\
\hline
Table 2.1: Parameters used in the Monte Carlo modelling of exciton diffusion, trapping, and annihilation. The band edge diffusivity ($D_{\text{band edge}}$), trap density, and detrapping attempt frequency ($f_{\text{detrap}}$) were the same when modelling each substrate. The trap state exponential tail parameter ($\alpha$) and annihilation radius ($R$) each depended on the supporting substrate in the system being modelled.

cm$^2$s$^{-1}$ in agreement with the measured charge carrier mobility in the literature [14, 16] as well as our diffusivity measurements when the large fraction of time that excitons spend immobile bound to defects is considered. The trapping density was chosen based on the spatial density of charged point defects observed in scanning tunneling microscopy images of acid-treated MoS$_2$ [47], which reveal a defect density of roughly 1000 per square micron. In order to reproduce the experimentally measured apparent exciton lifetime (tens of nanoseconds), it was necessary to choose a detrapping attempt frequency $f_{\text{detrap}} = 5 \times 10^{13}$ s$^{-1}$.

There are two critical parameters that depend on the supporting substrate refractive index: $\alpha$, which parametrizes the distribution of trap energies, and $R$, the radius at which two excitons annihilate. The parameter $\alpha$ describes the energetic distribution of the trapping sites and consequently the rate of thermally activated detrapping from those sites. The extent to which Coulomb interactions are screened in atomically thin materials is heavily influenced by the supporting substrate [22, 39, 64], and the external dielectric environment has been used to tune the energy of the band edge exciton in monolayer MoS$_2$ [11]. The energetic distribution of the trapping sites is similarly sensitive to the external dielectric environment. Trapped excitons are solvated by a polarizable dielectric environment and exhibit small oscillator strengths; these characteristics are consistent with a charge-separated trapped exciton state.

The second substrate-dependent parameter is the separation, $R$, at which two
excitons annihilate. This parameter is fitted by comparing simulation results to the steady-state QY measurements. The fitted values of $R$ depend strongly on the supporting substrate as well. As the refractive index of the substrate is increased, the separation, $R$, at which excitons interact gets smaller. This is consistent with exciton exciton annihilation mechanisms that require electron wavefunction overlap, or interaction via Coulomb or dipole interactions, both of which would be modulated by the surrounding dielectric environment, with less screening (i.e. lower index) favoring more distant interactions (i.e. larger $R$).

The output of this model is plotted along with the experimental steady-state QY data in Fig. 2-14b; The model accurately reproduces the steady-state experiments. Importantly, the model also reproduces the annihilation dynamics and transport behavior observed in the transient experiments. The agreement with experiment and the single fitting parameter ($R$) reflect that this model is an accurate representation of the physical system.

The dependence of $R$ on the substrate refractive index is striking; the extracted radius ranges from 4 nm (quartz), 0.6 nm (sapphire), to 0.4 nm (STO) while the substrate index ranges from 1.45 (quartz), 1.76 (sapphire), to 2.5 (STO). This trend is consistent with dielectric screening of the Coulomb interaction in TMDs and its dependence on the surrounding substrate. In Fig. 2-15, we plot the screened Coulomb potential as a function of radial separation in MoS$_2$ supported on quartz (blue), sapphire (green), and STO (red). For a constant value of the interaction potential (y-axis), changing the substrate dielectric constant from sapphire to STO results in an order-of-magnitude reduction in the corresponding radius, consistent with the results of our Monte Carlo simulations (Table 2.1).

Though TMD lasers [48, 90] and LEDs [28–30] have been demonstrated, practical use requires operation at high exciton densities. For instance, a MoTe$_2$ laser [90] exhibited a threshold pump generation rate, $R_{\text{pump}} \approx 4 \times 10^{18}$ cm$^{-2}$s$^{-1}$, in the regime where exciton exciton annihilation is dominant. Achieving high brightness LEDs or sufficiently high exciton densities for lasing or polariton condensation in the presence of competitive second order nonradiative decay channels necessitates excessive
Figure 2-15: Screened Coulomb interactions in TMDs. The analytical form is included as an inset. $H_0$ and $Y_0$ are Struve and Bessel functions, respectively, $\epsilon_{\text{sub}}$ and $\epsilon_{\text{pmma}}$ are the permittivity of the substrate and PMMA respectively, and $r_0$ is a characteristic screening length scale proportional to the TMD polarizability.

pump rates. Exciton exciton annihilation places a fundamental limit on the operating efficiency of such devices. Understanding the mechanism behind this efficiency loss and raising the maximum achievable operating efficiency by tuning the dielectric environment are critical advances for the future of TMD optoelectronic devices.

Strong exciton exciton interactions and the tuning of those interactions through the surrounding dielectric are both manifestations of reduced dielectric screening in 2D materials. Coulomb interactions are poorly screened in monolayer TMDs [11, 22, 39, 40] resulting in large exciton, trion, and biexciton binding energies. Though these many-body interactions can be exploited to observe physics unique to 2D materials, here they facilitate exciton exciton annihilation, limiting radiative efficiency. Dielectric screening is weak in these materials because coulomb interactions can circumvent the highly polarizable TMD by leaking through the surrounding environment. We take advantage of that fact here by tuning the dielectric constant in the external environment to suppress exciton exciton interactions. Tuning the strength of many-body interactions through the dielectric environment is a powerful design paradigm unique to low-dimensional materials.
2.3.7 Exciton Diffusivity Measurements: MoS$_2$ on Sapphire and STO

Exciton diffusivity measurements were also performed on acid-treated MoS$_2$ supported on sapphire and STO. In each case the experiment was performed as described in sections 2.3.4 and 2.3.12. In each case, the experiment was performed at sufficiently low fluence to avoid exciton exciton annihilation as indicated by the monoexponential PL decay curves shown in the top panels of Fig 2-16. The PL spatial distribution was fitted to a Gaussian at each time point. The variances from the Gaussian fits are plotted as a function of time in the bottom panel. The variance grows linearly in time at a rate that reflects the diffusivity. In both cases, the measured diffusivity was similar than the diffusivity measured for MoS$_2$ on quartz: $0.04 \pm 0.01$ and $0.06 \pm 0.02$ cm$^2$/s for MoS$_2$ on sapphire and STO respectively.

2.3.8 Generation Rate Dependent Quantum Yield in WS$_2$

The advantageous tuning of the exciton exciton annihilation rate constant with the supporting substrate refractive index is also observed in WS$_2$. Generation rate de-
Figure 2-17: Absolute quantum yield values for WS$_2$ supported on quartz (green circles), sapphire (yellow triangles), and STO (red squares) extracted from calibrated PL spectra as a function of carrier generation rate. The data are consistent with and qualitatively similar to the data for MoS$_2$ on the three substrates. At low generation rates, the samples emit with near unity quantum yield. As the generation rate increases, the exciton-exciton annihilation interaction lowers the quantum yield dramatically.

2.3.9 Optical System Characterization

For CW imaging, the emission spot is a convolution of the excitation point spread function (PSF) (i.e. the Gaussian laser spot size) and the collection PSF. A diffraction limited intensity distribution of excitons is excited, and then as each emits, it is localized within the accuracy of the imaging optics. It is possible to assess these two PSFs independently. To assess the collection PSF, a sparse film of isolated CdSe quantum dots was drop cast. This was then illuminated by an LED and the PL was imaged. Each point-like emitter appeared as a Gaussian spot with a width representing the imaging optics collection PSF.

For the excitation PSF, a complete film of dots was cast and excited with a laser. For these experiments, we used CdSe QDs coated with a thick (2-3 nm) ZnCdS shell and long-chain oleate ligands that were previously shown to prevent any measurable
exciton diffusion [82]. The imaged PL spot is shown in the left panel of Fig. 2-18. The width of this distribution represents the convolved excitation and collection PSFs. With this characterization in hand, the imaged PL emission spot recorded on MoS₂ represents the convolution of the excitation and collection PSFs as well as additional broadening due to diffusion. This broadening can be used to extract a diffusivity, $D = 0.03 \pm 0.01 \text{ cm}^2\text{s}^{-1}$ as described in section 2.3.4.

### 2.3.10 PL Spectra Under Extended Vacuum and Cooling

To assess the effect of vacuum and low temperature on the polymer-capped samples, the samples were cooled to 77 K and then heated to room temperature under vacuum. Upon cooling and concurrent exposure to vacuum, the total PL intensity decreased. Simultaneously, the exciton emission blue shifted and trap state emission became more prominent as less thermal energy was available to promote trapped excitons to the band edge [64]. The cooling and heating spectra are presented in Fig. 2-19. Upon heating, the spectrum red shifted and the trap state emission became less prominent again. Notably, the PL intensity recovered only somewhat until the cryostat was returned to atmospheric pressure, at which point the PL intensity completely recovered.
Figure 2-19: Photoluminescence spectra collected upon cooling (left) to and heating (right) from liquid nitrogen temperature. The color gradation indicates temperature (room temperature is red, 77 K is blue). The cooling and heating cycle took approximately one hour.

The total emission intensity collected as the sample was cooled, heated, and returned to atmospheric pressure is plotted in Fig. 2-20.
2.3.11 Fitting Low Temperature PL Spectra

Low-temperature PL spectra exhibit contributions from band-edge excitons and trapped exciton states. These contributions are presented in Fig. 2-21. Equilibration between these states is thermodynamically driven; their relative occupation is determined by the states’ degeneracies and the energy difference between the states. Band-edge and trapped excitons can also cool within their respective distributions. The two densities of states are occupied by a Fermi-Dirac distribution. Cooling within the trap site density of states is diffusion-mediated and kinetically limited. The “effective temperature” of the trapped exciton population is much higher than the lattice temperature as the population does not have sufficient time to reach equilibrium. The PL spectra are well fit using the trapped exciton and band-edge exciton states’ degeneracies and radiative rates.
\[ F_{f.d.} = \frac{1}{\exp \left( \frac{E - \mu_{\text{B.E.}}}{k_B T_{\text{lattice}}} \right) + 1} \]

\[ \frac{N_{\text{B.E.}}}{N_{\text{trap}}} = \frac{g_{\text{B.E.}}}{g_{\text{trap}}} \exp \left( \frac{-\Delta E}{k_B T_{\text{lattice}}} \right) \]

\[ F_{f.d.} = \frac{1}{\exp \left( \frac{E - \mu_{\text{trap}}}{k_B T_{\text{trapped excitons}}} \right) + 1} \]

Figure 2-21: Schematic illustration of the two sub populations that contribute to low-temperature PL spectra.

2.3.12 Experimental Methods

Sample Preparation – Mineral MoS\(_2\) (SPI) was exfoliated on SiO\(_2\)/Si substrates and then transferred to other substrates including quartz, sapphire or STO substrates by a dry transfer technique via a polymethyl methacrylate (PMMA) membrane as a transfer media. The transferred MoS\(_2\) was treated by the following procedure: 20 mg of bis(trifluoromethane)sulfonimide was dissolved in 5 ml of 1,2-dichloroethane and then diluted with 45 ml 1,2-dichlorobenzene to make a superacid solution. The transferred MoS\(_2\) with PMMA was then immersed in the superacid solution for 30 seconds at room temperature. The sample was blow dried with nitrogen. Note that the enhancement depends on the initial quality of the sample and only a portion of sample can reach $\geq 95\%$ QY.

Transient PL – Samples were excited using a 405 nm pulsed laser diode (Picoquant, LDH-D-C-405M, 40 MHz repetition rate, 0.4 ns pulse duration) with fluences as indicated in 2.3.3. The laser was focused to a nearly diffraction-limited spot (Nikon, CFI S Plan Fluor ELWD, 40×, 0.6 NA). Fluorescence was collected with the same
objective, and passed through a dichroic mirror and 600 nm - 700 nm bandpass filter before being focused onto a Si avalanche photodiode (Micro Photon Devices, PDM50, 50 ps resolution at the detection wavelength). The detector was connected to a counting board for TCSPC (Picoquant, PicoHarp 300).

**Quantum Yield Measurement** – The calibrated PL QY measurement method has been previously described in detail [46]. Briefly, the 514.5 nm line of an Ar ion laser (Lexel 95) was focused to the sample using a 60× ultra-long working distance objective (NA = 0.7). PL was collected by the same objective, filtered and dispersed by a spectrograph. The emission was detected by a Si CCD camera (Andor, iDus BEX2DD). The excitation power and optical system spectral sensitivity were externally calibrated. The instrument function was cross-calibrated using rhodamine 6G (QY close to 100%) and a spectralon as reference samples. The measured PL spectra were integrated and converted into external quantum efficiencies and corresponding QYS.

**Transient PL Microscopy** – $\lambda = 570$ nm pulses from a synchronously pumped optical parametric oscillator (Coherent, PP automatic, 76 MHz, < 1 ps) were spatially filtered by a single-mode optical fiber and used to excite the sample. The laser was focused to a diffraction-limited spot (Nikon, CFI Plan Apo Lambda, 60× Oil, 1.4 NA). Fluorescence was collected by the same objective and filtered by a dichroic mirror and 600 nm - 700 nm bandpass filter. The APD detector was placed in the 360× magnified image plane outside the microscope. The detector position in the image plane was controlled by two orthogonal motorized actuators (Thorlabs, ZFS25B). The evolution of the photoluminescence spatial profile with time was acquired by scanning the detector across the magnified emission profile and collecting a photoluminescence decay histogram at each position.

**PL Spectroscopy, Low Temperature** – A 532 nm continuous wave laser (Coherent, Sapphire SF 532-20 CW) was focused at the sample (Nikon, CFI S Plan Fluor ELWD, 40×, 0.6 NA). Fluorescence was collected by the same objective and filtered by a dichroic mirror before being dispersed by a spectrograph (Princeton Instruments, Acton SP2500) and imaged on a cooled CCD camera (Princeton Instruments, Pixis
Low-temperature data were collected under vacuum in a microscope-mounted cryostat (Janis, ST-500-P).

**Numerical Simulation** — Exciton dynamics were simulated with a fixed time step Monte Carlo algorithm. Excitons were initialized to the band edge according to a spatial profile matching the excitation laser intensity profile. At each time step free excitons hopped a fixed distance in a random direction. Excitons trapped with unit probability if the center position of the exciton was within 0.4 nm of the center of an empty trap. Upon moving within $R$ nm of an occupied trap, the exciton annihilated and was removed from the simulation. Annihilation between pairs of band-edge excitons were rare due to the low population of detrapped excitons and were thus neglected. Trapped excitons detrapped probabilistically as described in section 2.3.5.
Chapter 3

2D/0D Hybrid Structures: CdSe Quantum Dots on Transition Metal Dichalcogenides

3.1 Hybrid Quantum Dot / 2D Devices

QDs and TMDs possess complementary strengths. QDs are capable of size-tunable broadband absorbance and efficient narrowband emission. On the other hand, TMDs exhibit impressive charge carrier mobilities on the order of $100 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ at room temperature. Broadly, QDs excel at optical processes such as light emission and absorption, while TMDs excel at charge conduction.

While QD devices have been realized, their performance is often limited by poor charge conduction in the QD active layer. QD solids engineered for charge carrier conduction have mobilities on the order of $0.01 \text{ cm}^2\text{V}^{-1}\text{s}^{-1}$ [91]. As a consequence, QD LEDs exhibit external quantum efficiency droop at high driving voltages [92]. In order to achieve high brightnesses, LEDs must be driven with large currents and consequently large voltages. This can lead to nonradiative multi-carrier interactions and direct field-driven exciton dissociation limiting QD LED operating efficiencies [92].
TMDs possess superior charge carrier mobilities, but despite the strongly bound emissive exciton states, TMDs generally do not photoluminesce with high quantum efficiency. As exfoliated, most materials exhibit 0.1-1% photoluminescence quantum efficiency. Through superior crystal growth methods or chemical treatments, it is possible to engineer TMDs with quantum yields approaching unity. Even so, strong exciton exciton interactions resulting from poor dielectric screening of Coulomb interactions can limit device operating efficiencies at moderate charge carrier densities. Furthermore, it is difficult to tune TMD emission energies to the degree that is possible with colloidal QDs.

With these complementary strengths and weaknesses in mind, we envision hybrid 0D/2D devices that take advantage of QDs' favorable optical properties and TMDs' favorable charge conduction. Two such devices are illustrated in Fig. 3-1. A hybrid LED device is illustrated in Fig. 3-1a. In this figure, charges are injected into a TMD where they form excitons. Excitons transfer to the coupled QD, which handles light emission taking advantage of each constituent materials' relative strengths. A hybrid photodetector is illustrated in Fig. 3-1b. Here, a QD acts as a tunable absorber and injects free charges into an underlying TMD. An applied field passes current through the TMD, which is sensitive to the photoinduced conductivity. High detectivity hybrid QD/TMD photodetectors have been demonstrated that take advantage of the QD spectral tunability [93].

To further develop hybrid QD/TMD devices, it is important to understand the coupling between zero dimensional QDs and two dimensional TMDs. This chapter focuses on coupling between QDs and TMDs in the limit that they are weakly coupled as well in the limit where there is strong tunable electronic coupling.
3.2 Exciton States in Transition Metal Dichalcogenides and Quantum Dots

3.2.1 Transition Metal Dichalcogenide Electronic Structure

Monolayer TMDs have a hexagonal structure illustrated in Fig. 3-2a. The reciprocal lattice and first Brillouin zone are also shown. Line cuts of the band structure calculated for monolayer MoS$_2$ along the $\Gamma$-K and K-M directions are shown in Fig. 3-2b [2]. Monolayer TMDs are typically direct gap semiconductors with the band gap located at the K point. When additional layers are present, the conduction band minimum along the $\Gamma$-K direction lowers in energy and the valence band maximum at $\Gamma$ rises in energy forming an indirect gap smaller in energy than the direct gap at the K point [20].

In most monolayer TMDs, the lowest energy transition is the direct gap at the K point, and mixing of delocalized states at the band extrema at the K point gives rise to strongly bound excitons. The valence band maximum at the K point is split due to spin-orbit coupling resulting in the A and B exciton transitions indicated in Fig. 3-2b. The conduction and valence band are nearly parallel in the region along the $\Gamma$-K direction labelled C in Fig. 3-2b, which results in many interband transitions.
in the joint density of states [2]. Fig. 3-2c shows the differential reflectance spectrum for a monolayer MoS$_2$ flake. The A and B exciton transitions are indicated as well as the broad, intense absorbance that arises from the parallel bands along the $\Gamma$-K direction.

Excitons in monolayer TMDs are modelled using an adapted 2D hydrogen model [53]. The electron and hole effective masses are calculated based on the curvature of the conduction and valence bands at the K point. The energy of the system is
modelled with the Hamiltonian

$$H = - \sum_i \frac{\nabla_i^2}{2m_i} + \sum_{i<j} q_i q_j V(rij). \quad (3.1)$$

The first term represents the kinetic energy of each localized charge and the second term represents the Coulombic attraction between the electron and hole separated by distance $rij$. In traditional 2D hydrogen models, the Coulomb potential $V(rij)$ has the standard $r^{-1}$ dependence characteristic of homogeneous media

$$V_{\text{bulk}}(rij) = \frac{1}{4\pi \epsilon r_{ij}}, \quad (3.2)$$

where $\epsilon$ is the permittivity of the medium. In sharp contrast, charges in two dimensional materials interact via a unique Coulomb potential

$$V(rij) = \frac{\pi}{(\epsilon_1 + \epsilon_2) r_{ij}} \left[ H_0(rij/r_0) - Y_0(rij/r_0) \right], \quad (3.3)$$

where $\epsilon_1$ and $\epsilon_2$ are the permittivities of the surrounding dielectric media, $H_0$ is the Struve function, and $Y_0$ is the Bessel function of the second kind, and $r_0$ is a characteristic screening length scale in the two dimensional material related to its 2D polarizability, $\alpha_{2D}$, according to

$$r_0 = 2\pi \alpha_{2D}. \quad (3.4)$$

This Coulomb potential unique to atomically thin two dimensional materials differs from the homogeneous Coulomb potential, $V_{\text{bulk}}$, in two salient ways. First, the Coulomb potential’s dependence on the external dielectric environment (represented by the permittivities of the surrounding media in eqn. 3.3) results from charges being able to interact through the surrounding dielectric. Two charges can be connected through a path that samples the surrounding media dielectric environment where the electric field is typically less strongly screened. The electric potential energy resulting from two adjacent charges can consequently be much larger than it would be if the
entirety of their interaction were screened by the TMD permittivity. Additionally, this sensitivity to the dielectric environment allows engineering of many body interactions through tuning the surrounding media [11]. These ideas are illustrated in Fig. 3-3a, and will be further discussed in Section 2.3.

The second consequence of the anomalous Coulomb screening is that many body interactions are poorly screened over moderate length scales when compared to the screening lengths in bulk TMDs. This is illustrated in Fig. 3-3b. The potential energy resulting from two charges in a monolayer TMD is much larger than the energy resulting from two charges with the same geometry in the bulk. This anomalous screening has been most conclusively shown in the non-Hydrogenic exciton Rydberg series in monolayer TMDS [39, 40], and the surprisingly strongly many body interactions in TMDs including strongly bound excitons, trions [3, 41, 54], and biexcitons [55].

Figure 3-3: (a) Electric fields lines emanating from charges in atomically thin TMDs can sample the surrounding dielectric environment, which is typically less polarizable than the TMD. (b) The Coulomb interaction between two point charges in a TMD is sensitive to the surrounding dielectric. Here the Coulomb potential energy due to two point charges separated by a distance \( r_{ij} \) is plotted. Bulk TMDs (such as MoS\(_2\)) are highly polarizable and the Coulomb potential is strongly screened (green). Quartz is much less polarizable and the Coulomb potential is less screened (blue). Charges in a MoS\(_2\) monolayer encased quartz experience a Coulomb potential similar to charges separated in quartz except at very small distances, where the Coulomb potential approaches that of the bulk TMD (red).
3.2.2 CdSe Quantum Dot Electronic Structure

Quantum dots (QDs) are nanometer-scale semiconductor crystals that in the simplest case are spherical Fig. 3-4a. When the physical size of the nanocrystal is reduced below a characteristic length scale, $a_B$ (the Bohr radius), the energies of the carriers in a QD are shifted relative to their energy in the bulk material. The Bohr radius is given by

$$a_B = \epsilon \frac{m}{m^* a_0},$$

where $\epsilon$ is the permittivity of the material, $m$ is the rest mass of the particle of interest, $m^*$ is the effective mass of the particle in the material of interest (reflecting the local curvature of the band the particle occupies), and $a_0$ is the Bohr radius of the hydrogen atom. When the nanocrystal is shrunk below the constituent material’s Bohr radius, the charges inside experience quantum confinement and the electronic energy levels allowed inside the nanocrystal are determined by the nanocrystal size. Near the band edge, the continuous electronic bands of electronic states allowed for the bulk semiconductor give way to discrete atom-like electronic states. This idea is illustrated in Fig. 3-4b. Quantum confinement is very important for technological applications, because it allows facile tuning of the QD absorption and emission energy by changing the QD size (without altering the parent semiconductor composition). Such tuning is much more difficult for bulk semiconductors. Though the absorption onset is size-tunable in QDs, they exhibit broadband absorbance above the lowest excitation energy with transitions between discrete atom-like states giving way to continuous interband transitions at higher energies. QDs can be engineered to exhibit efficient narrow-band photoluminescence, which has been leveraged in technologies that require color purity such as wide-color gamut displays.

When the nanocrystal is substantially larger than the parent crystal unit cell, the electronic states in a QD have the form

$$\psi(r) = u_k(r)f(r)$$
Figure 3-4: (a) Colloidal quantum dots are nanometer-scale semiconductor crystalline cores covered with organic small molecule capping ligands that impart colloidal stability. (b) When the nanocrystal size is reduced below a critical size, the carriers begin to feel size-dependent quantum confinement effects. The continuous bands of states present in the bulk transition into atom-like discrete states near the band edge whose energies are sensitive to the quantum dot size.

where \( u_k(r) \) are the Bloch functions of the parent crystal lattices with momenta \( k \), and \( f(r) \) are the envelope functions that modulate the underlying Bloch wavefunctions. The envelope functions \( f(r) \) are the part of the QD electronic structure that is sensitive to the QD size and shape. For spherical QDs the envelope functions are well-represented by the solutions to a particle in a finite-barrier potential well sphere (the particles have effective masses characteristic of the bulk semiconductor). The single electron (hole) eigenstates that arise from that treatment are similar to the states of a hydrogen atom and have analogous principal quantum numbers and angular momenta to the atomic hydrogen orbitals. Though charges are mostly confined to the semiconductor core, the potential barrier at the QD surface is finite and the resulting electronic wavefunctions partially escape the nanocrystal boundaries decaying exponentially away from the QD surface. This wavefunction leakage plays an important role in charge transfer processes involving QDs.

It’s worth considering whether or not the charges in a QD are Coulombically bound like constituent charges of an exciton in an extended semiconductor. The charges are
constrained by the QD geometry to occupy the same space, but the kinetic energy resulting from the spatial confinement (the spatial derivative term of the particle in a sphere Hamiltonian) of the charges is substantially larger than the Coulomb potential energy gained by cohabitating the same nanocrystal [94]. Regardless, there is an attractive interaction between the electron and hole similarly to bound excitons in extended nanostructured semiconductors (such as nanotubes and two dimensional materials). Furthermore, in spherical QDs, the exciton transition dipole moment is roughly isotropically oriented in contrast to one- and two-dimensional excitonic materials.

3.3 Reduced Dielectric Screening and Enhanced Energy Transfer in Single- and Few-Layer MoS\textsubscript{2}

Much of Section 3.3 has appeared in print as Prins, Goodman, and Tisdale Nano Lett. (2014), [22].

3.3.1 Förster Resonance Energy Transfer

Förster resonance energy transfer (FRET) is one mechanism by which energy transfer from a donor to acceptor can occur. Förster described the theoretical framework for FRET in 1948 [95]. FRET is a near field interaction in which a donor dipole transfers energy to a nearby acceptor dipole. One way of viewing this interaction is that a charge-neutral exciton excited state transfers from the donor to acceptor. Another view is that the donor dipole radiates a near field, which is then dissipated by the lossy acceptor dipole (through some imaginary component of the acceptor dipole’s susceptibility). This second formalism in which the donor and acceptor dipoles act as radiating and receiving antennae is useful for describing FRET and is illustrated in Fig. 3-5.

This qualitative picture is described quantitatively by an energy transfer rate,
FRET can be described as a donor dipole emitting a near field which is then received by a lossy acceptor dipole. The acceptor dipole must have some absorptive power by which to receive the radiated field. This interaction is screened by the intervening dielectric. (b) This analogy carries over to excitonic systems. Donor and acceptor dipoles correspond to the acceptor and donor transition dipole moments. Similarly to the macroscopic antenna picture, the dipoles are screened by a dielectric and must be aligned to facilitate energy transfer.

\[ k_{ET} = k_{rad} \left( \frac{R_0}{d} \right)^6 \]  
\[ R_0^6 = \frac{9\eta\kappa^2}{126\pi^5 n^4} \int \lambda^4 F_D(\lambda) \sigma_A(\lambda) d\lambda, \]  

where the Förster radius, \( R_0 \), is the donor-separation distance at which energy transfer is 50\% efficient, \( k_{rad} \) and \( \eta \) are the donor radiative rate and quantum yield respectively, \( d \) is the donor-acceptor separation, \( \kappa^2 \) is an orientation factor that requires the donor and acceptor dipoles to be aligned to maximize transfer efficiency, \( n \) is the intervening medium’s refractive index, and the integral calculates the spectral overlap between the donor emission spectrum \( F_D(\lambda) \) and the acceptor absorption spectrum \( \sigma_A(\lambda) \).
\( F_D(\lambda) \) is normalized such that \( \int F_D(\lambda) \, d\lambda = 1 \) and \( \sigma_A(\lambda) \) is the donor absorption cross section.

### 3.3.2 Measuring the Rate of Energy Transfer Using Time-Correlated Single Photon Counting

![Diagram of energy transfer and radiative emission](image)

The total observed decay rate will be

\[
 k_{\text{total}} = k_{\text{rad}} + k_{\text{NR}} + k_{\text{ET}}
\]

Figure 3-6: An excited quantum dot exhibits an effective decay rate, \( k_{\text{total}} \) that is a sum of all the first order decay rates available to the exciton: the radiative rate \( k_{\text{rad}} \), first order nonradiative rates \( k_{\text{NR}} \), and the energy transfer rate, \( k_{\text{ET}} \). We can infer the energy transfer rate by measuring the total decay rate in the presence and absence of an energy acceptor.

We measure the rate of energy transfer from the quantum dot donor to the MoS\(_2\) acceptor by monitoring the quantum dots’ apparent radiative lifetimes, \( k_{\text{total}} \). When an exciton forms on an excited quantum dot with high quantum yield, it radiates with a characteristic lifetime, \( \tau_{\text{rad}} \). In addition to the radiative decay pathway, in most real systems, the exciton can relax nonradiatively eventually dissipating its energy as heat. Nonradiative exciton traps and nonradiative many body interactions are examples of processes that can decrease quantum dots’ photoluminescence quantum yield. In the low excitation regime, nonradiative decay pathways are typically first-order decay processes that can be combined into a single rate constant, \( k_{\text{NR}} \), and the quantum
dot radiative rate is the sum of all the decay rates available to the quantum dot:

\[ k_{\text{total}} = k_{\text{rad}} + k_{\text{NR}} + k_{\text{ET}}. \]  

(3.9)

We can infer the energy transfer rate, \( k_{\text{ET}} \) by looking at the quantum dot radiative rate in the presence and absence of an energy acceptor. These concepts are illustrated in Fig. 3-6.

We measure photoluminescence decay rates in our system using a technique called time-correlated single photon counting (TCSPC). The TCSPC experiment is performed with the optical apparatus illustrated in Fig. 3-7. The sample is placed in a microscope and excited with a pulsed laser. The pulsed laser is synchronized with an electronic counting board. When the sample fluoresces, the fluorescence is collected and directed to an avalanche photodiode (APD), which detects single photons and outputs voltage pulses that are directed to the laser-synchronized counting board. The pulsed experiment is repeated millions of times while the counting board records all of the measured photon arrival-detection time delays. When these delays are histogrammed, they form a detailed record of the sample’s time-dependent emission dynamics.
Figure 3-7: Time-correlated single photon counting experimental apparatus. A pulsed diode laser excites the sample. Emitted photons are directed to an avalanche photodiode (APD). Photon detection events at the APD result in a voltage pulse being sent to the counting board, which is synchronized to the laser excitation. Pulse arrival - photon emission delay times are recorded and histogrammed by the counting board.

3.3.3 Energy Transfer in CdSe/CdZnS Quantum Dot/MoS₂ Hybrid Structures

Here, we demonstrate nonradiative energy transfer from colloidal quantum dots to exfoliated monolayer and few-layer molybdenum disulfide (Fig. 3-8a). We use high quantum yield (> 80%) CdSe/CdZnS core/shell quantum dots having an emission spectrum that overlaps with the strong MoS₂ direct excitonic absorption features (see Fig. 3-8b). Monolayer and few-layer MoS₂ flakes were deposited on Si/SiO₂ substrates by mechanical exfoliation from bulk single-crystalline MoS₂ (Graphene Supermarket). Core-shell CdSe/CdZnS quantum dots (QD Vision Inc.) were deposited onto the samples using spin-coating from a dilute solution (0.5 mg/mL in toluene, 1500 rpm) to form a submonolayer that is ≤ 1 QD thick in all locations (Fig. 3-8c). The CdSe QD cores are 4.8 nm in diameter and surrounded by a 2.3 nm thick Cd₀.₅Zn₀.₅S shell that is capped with a layer of organic ligands (octadecyl phosphonic acid) 1.6 nm in length. The quantum dots are well passivated, resulting in solution phase quantum yields of > 80% and a native 1/e lifetime of 20.4 ± 0.1 ns.

Scanning photoluminescence lifetime measurements were performed using an in-
Figure 3-8: (a) Schematic of the hybrid QD/MoS₂ energy transfer system with indicated length scales. (b) Monolayer MoS₂ absorption spectrum [3], overlaid with the CdSe/CdZnS emission spectrum. Inset: transmission electron micrograph of a monolayer of the CdSe/CdZnS QDs used in this study. (c) Scanning electron micrograph showing the submonolayer of CdSe/CdZnS QDs on top of a MoS₂ flake (dark section) and the surrounding SiO₂ substrate (lighter section).

verted microscope (Nikon, Ti Eclipse), equipped with an x-y piezo stage (Nanonics, MV2000). The samples were excited using a 405 nm pulsed laser diode (LDHD-C-405M, Picoquant, 10 MHz repetition rate, 0.5 ns pulse duration) with an average pulse fluence of 60 nJ/cm² across a near-diffraction limited spot ~ 1 μm in diameter (Nikon, 20× objective, N.A. = 0.4). The photoluminescence was collected with the same objective, passed through a dichroic mirror and long-pass filter, and focused onto a Si avalanche photodiode (Micro Photon Devices, PDM50, 32 ps resolution). The detector was connected to a counting board for time-correlated single photon counting (Picoquant, PicoHarp 300). During laser scanning fluorescence lifetime imaging, the sample was scanned by the piezo stage, recording a fluorescence lifetime trace for each pixel. Raman imaging was performed using a scanning Raman microscope (Horiba, LabRAM) with a 532 nm light source (Kaiser Optical Systems, Inc., Invictus).

Fig. 3-9a shows a bright-field optical micrograph of a MoS₂ flake prior to QD deposition. The different colors result from absorptive as well as thin-film interference effects and indicate several different thicknesses present in this sample [96]. To determine the layer thickness of each region, we performed micro-Raman measurements of this same flake. The energy difference between the A₁₆ and E₂₈ Raman modes is sensitive to the layer thickness and can be used to quantify the thickness of few-layer MoS₂ [73]. We perform this characterization for a variety of flakes and identify
thickness regions ranging from bulk-like thicknesses (> 8 layers) down to monolayer thickness. The sample shown in Fig. 3-9a contains facets ranging from three to more than eight layers thick.

A submonolayer of quantum dots was deposited across the different samples, and we used laser-scanning microscopy to construct maps of QD exciton lifetime (Fig. 3-9b). Compared to the much larger volume-integrated absorption cross-section and efficient emission of the QDs, MoS$_2$ was only weakly fluorescent and contributed negligibly to total photoluminescence. For each pixel in the laser scanning microscopy image, we recorded a fluorescence lifetime decay curve and extracted the 1/e lifetime. For all samples, we found that the fluorescence lifetime of QDs away from the flake was close to their native lifetime of 20.4 ± 0.1 ns. In contrast, the fluorescence lifetime of the QDs on top of the MoS$_2$ was significantly shortened, indicating strong quenching of the QD photoluminescence by energy transfer to MoS$_2$. Interestingly, we observed that the lifetime varied as a function of flake thickness with strongest QD quenching occurring for the thinnest regions of MoS$_2$.

![Image](image_url)

Figure 3-9: (a) Optical micrograph of a mechanically exfoliated MoS$_2$ flake with indicated layer thicknesses of the different regions. (b) QD fluorescence lifetime map of the same flake with a saturated color scale to emphasize the lifetime variations within the flake area. (c) Energy transfer efficiency map of the same flake.

In Fig. 3-9b, we show the lifetime map of the multifaceted sample of Fig. 3-9a. The color scale is saturated above 5 ns to highlight the thickness dependence of the quenching. The bulk-like section of this sample has a spatially averaged lifetime of
2.82 ± 0.04 ns while the trilayer section has a spatially averaged lifetime of 1.70 ± 0.01 ns. We can express the lifetime variations in terms of an energy transfer efficiency, given by \( \eta_{ET} = 1 - (\tau_{DA}/\tau_D) \), in which \( \tau_D \) is the native 1/e lifetime of the donor QDs on SiO₂ (in other words, in the absence of MoS₂) and \( \tau_{DA} \) the 1/e lifetime of the QDs on top of the MoS₂ acceptor. The resulting efficiency map is shown in Fig. 3-9c.

For all flakes that we investigated, we found that the energy transfer efficiency is consistently above 80%, reaching as high as 95.1 ± 0.1% for monolayer sections.

The observed quenching of QD fluorescence is consistent with nonradiative energy transfer from the QDs to MoS₂ through dipolar interactions. It should be noted, though, that the quenched intensity and shortened lifetime of the CdSe/CdZnS QDs could in principle also originate from individual charge transfer to MoS₂. To confirm nonradiative energy transfer as the responsible mechanism, we performed control experiments with InAs QDs having a bandgap energy smaller than that of MoS₂. Fig. 3-10 shows representative fluorescence decay histograms of both the CdSe and InAs QDs. In contrast to the CdSe QDs and consistent with the energy transfer mechanism, the photoluminescence lifetime of the InAs QDs remains unchanged when placed on top of a MoS₂ surface.

We now turn to the layer thickness dependence of energy transfer. The resonant MoS₂ absorption features shown in Fig. 3-8b are due to spin-orbit split direct excitonic transitions at the K-point of the MoS₂ Brillouin zone [97]. The oscillator strength and spectral shape of these transitions are largely unaffected by the MoS₂ flake thickness [5, 20, 98], and therefore the large (> 300%) increase in the energy transfer rate with decreasing thickness cannot be explained by changes in the spectral overlap. Consequently, the observation of more efficient energy transfer for thinner MoS₂ acceptor layers is inconsistent with the most basic approximations of energy transfer, as described in Förster theory [95]. Within this model, the energy transfer rate is calculated assuming dipolar coupling between donor and acceptor point dipoles. Förster theory can be expanded to describe energy transfer between a point emitter and an acceptor surface by treating the acceptor surface as an array of noninteracting point dipoles [99]. A larger volume of acceptor material automatically leads
to more efficient energy transfer, because each additional dipole constructively contributes to the energy transfer rate. This theoretical framework has been successfully applied to describe energy transfer in a variety of nanomaterial systems, including energy transfer in colloidal quantum dot assemblies [100, 101], between emitters and graphene surfaces [102, 103], as well as from molecular dyes to plasmonic modes in thin metallic films [104]. Moreover, as expected within this framework but contrasting with our own observations, recent experiments using graphene as the acceptor for QD donors showed a monotonically decreasing energy transfer rate for decreasing...
thickness of few-layer graphene flakes [102].

Interestingly, it has been suggested that the assumption of noninteracting point dipoles in Förster theory is an oversimplification in the limit of thin semiconductor acceptor films with high polarizability [103, 105]. The thin film geometry can lead to enhanced energy transfer as a result of reduced dielectric screening of in-plane components of the donor dipole field [106]. This geometric effect is accentuated in acceptor media with in-plane oriented transition dipole moments and may in extreme cases lead to the counterintuitive result where thinner acceptor layers yield more efficient energy transfer [105]. Thin MoS$_2$ flakes are characteristic examples of materials that should exhibit this effect. MoS$_2$ has a large permittivity [3, 96], and its dielectric function is reported to be highly anisotropic with the resonant transitions polarized completely in the plane of the material [35].

![Figure 3-11: Energy transfer rate as a function of the number of MoS$_2$ layers. The dashed line represents the energy transfer rate to thick, bulklike MoS$_2$. Blue circles indicate experimental data with associated uncertainty that is smaller than the size of the symbols. The solid line is a fit to eqn. 3.10. Inset: schematic representation of the definition of the thickness (t) and distance (d) parameters.](image)

To quantify the enhancement of energy transfer in our QD/MoS$_2$ system, we compare the energy transfer rate corresponding to the different thicknesses with the energy transfer rate in the bulk limit. The energy transfer rate is given by $k_{\text{ET}} = \tau_{\text{DA}}^{-1} - \tau_{\text{D}}^{-1}$ and is calculated by spatially averaging data collected from each thickness region across multiple samples. From a total of five flakes, we obtained data for ten
different thickness regions, ranging in thickness from 1 to 8 layers thick. In addition, for visibly thick bulk-like flakes that we estimate to be $> 20$ layers, we obtain an average $k_{ET} = k_{ET}^{3D} = 2.4 \pm 0.3 \times 10^8 \text{ s}^{-1}$. In Fig. 3-11, we plot $k_{ET}$ as a function of the number of layers, with $k_{ET}^{3D}$ depicted as a dotted line. The enhancement factor $k_{ET}/k_{ET}^{3D}$ (right axis in Fig. 3-11) shows that energy transfer from the quantum dots to a MoS$_2$ monolayer is more than three times faster than energy transfer to the bulk.

Using the continuum model derived by Gordon and Gartstein [103], we can express the enhancement factor in the case of anisotropic permittivity as a function of the ratio between the flake thickness $t$ and the distance $d$ between the donor and the surface of the acceptor (see inset to Fig. 3-11)

\[
\frac{k_{ET}}{k_{ET}^{3D}} = \text{Im} \left[ \int_0^\infty d\rho \rho^2 e^{-2\rho} \frac{(\beta^2 - 1) \left(1 - e^{-2\rho d(t/d)}\right)}{(\beta + 1)^2 - (\beta - 1)^2 e^{-2\rho d(t/d)}} \right] \\
\times \left( \text{Im} \left[ \int_0^\infty d\rho \rho^2 e^{-2\rho} \frac{(\beta^2 - 1)}{(\beta + 1)^2} \right] \right)^{-1},
\]

where $\beta = (\epsilon_\parallel/\epsilon_\perp)^{1/2}$ is the effective permittivity of the acceptor, $\delta = (\epsilon_\parallel/\epsilon_\perp)^{1/2}$ is a parameter representing the dielectric anisotropy, $\lambda$ is the donor emission wavelength, and $\rho$ is a variable of integration. Taking the distance from the center of the QD to the MoS$_2$ surface to be $d = 6.3$ nm and a flake thickness of $t = 0.65$ nm per S-Mo-S layer [73], we fit the data in Fig. 3-11 to eqn. 3.10, allowing the real and imaginary parts of $\epsilon_\parallel$ and $\epsilon_\perp$ to be varied as fit parameters. A least-squares regression yielded the strongly anisotropic dielectric functions $\epsilon_\parallel = 18.2 + 11.3i$ and $\epsilon_\perp = 7.1$, where $\epsilon_\perp$ has a negligible imaginary component. The large difference in the parallel and perpendicular dielectric functions is consistent with recent reports on the complete in-plane polarization of the direct excitonic transition dipoles in MoS$_2$ [35]. Additionally, these values compare favorably to an isotropically averaged dielectric function reported recently for MoS$_2$ [3]. It should be noted that the values obtained here represent the “effective” permittivity of MoS$_2$ within our hybrid architecture, since eqn. 3.10 assumes vacuum as the immersing medium whereas our MoS$_2$
flakes are supported between SiO$_2$ ($n \approx 1.5$) and a layer of organic-capped quantum
dots ($n \approx 1.8$). Consequently, placing MoS$_2$ on substrates with lower permittivity or
suspending over air could further emphasize the enhancement.

To confirm the universality of the reduced dielectric screening effect in MoS$_2$, we
tested the predictions of eqn. (3.10) using CdSe QDs with thinner (< 1 nm thick)
CdZnS shells and shorter capping ligands (see section 3.3.5). As expected for the
reduced QD-MoS$_2$ distance in this thin-shell CdSe system ($d = 3.9$ nm, contrasted
to $d = 6.3$ nm for the thick-shell CdSe/CdZnS QDs shown in Fig. 3-8) we observe
an overall faster energy transfer rate. Moreover, consistent with the predictions of
eqn. (3.10) we find that the energy transfer rate increases with decreasing layer thick-
ness for this set of QDs as well. Finally, the scaling of the enhancement factor $k_{ET}/k_{3D}^{ET}$
with the $t/d$ ratio for the thin-shell CdSe system shows excellent agreement with the
thick-shell CdSe system (see section 3.3.5), suggesting that the enhancement factor
is largely independent of the QD donor and is mainly determined by the effective
dielectric environment of MoS$_2$ and the underlying substrate.

### 3.3.4 Conclusions

The scaling of the energy transfer rate with donor-acceptor separation distance, $d$,
can reveal interesting information about the underlying physics of the system. In the
case of classical Förster energy transfer the scaling exponent $m$, where $k_{ET} \propto d^{-m}$
[99], depends on the dimensionality of the system. For energy transfer between a
zero-dimensional (0D) emitter and a semi-infinite 3D acceptor the scaling exponent
is $m = 3$, whereas $m = 4$ for energy transfer between a 0D donor and a 2D acceptor
[103]. Using the experimentally measured energy transfer rates for the two CdSe QD
donors studied here, we found that the scaling exponent for energy transfer to bulk
MoS$_2$ was close to $m = 3$, as expected from Förster theory. However, for single-
layer MoS$_2$ the scaling exponent was less than 3 ($m \approx 2.5$), which is in contrast to
the $m = 4$ behavior for a 2D acceptor predicted by Förster theory (and observed
experimentally for energy transfer from molecular donors to graphene [103]). The
slower scaling in thin layers is predicted by eqn. (3.10) and clearly shown by the
shape of the curve in Fig. 3-11: for a given layer thickness, \( t \), the enhancement effect is amplified at distances \( d \) farther away from the acceptor surface (in other words, \( t/d \) is smaller, moving higher on the \( k_{ET}/k_{ET}^{3D} \) curve; see cautionary note below).

We emphasize that Förster theory fails to capture salient features of our experimental data for thin MoS\(_2\) acceptors, such as distance scaling and thickness dependence. Specifically, the assumptions of additive and noninteracting acceptor dipoles and immersion within an infinite homogeneous dielectric medium are violated for few-layer MoS\(_2\) and, consequently, the concept of a Förster radius is invalid. However, one may define an effective Förster radius, \( R_0 \), as the distance at which the probability of energy transfer to MoS\(_2\) is equal to the probability of radiative or nonradiative decay on the QD. Because of the energy transfer rate enhancements in single- and few-layer samples, both \( R_0 \) and the distance-scaling exponent are functions of the flake thickness. Using our experimentally determined distance scaling exponents and the experimentally measured energy transfer rates for the thick shell CdSe/CdZnS QD donors, the calculated Förster radii are \( R_0 \approx 11 \) nm for bulk-like MoS\(_2\) acceptors and \( R_0 \approx 19 \) nm for single-layer MoS\(_2\) acceptors, where the distance is measured from the center of the QD to the surface of MoS\(_2\). The unusually long-range interaction distance for single-layer MoS\(_2\) arises from the reduced dielectric screening phenomena described in eqn. (3.10) and shown in Fig. 3-11, the large absorption cross section of MoS\(_2\) [21, and the high quantum efficiency of the QDs used in this study. We caution, however, that these effective Förster radii may not be applicable at actual length scales comparable to \( R_0 \). The universal curve for MoS\(_2\) shown in Fig. 3-11 predicts that at distances greater than \( \sim 9 \) nm away from the MoS\(_2\) surface (\( t/d < 0.075 \)) the enhancement factor begins to decrease with increasing separation distance, and that at even larger separation distances bi- or trilayer MoS\(_2\) acceptors may exhibit the fastest energy transfer.

In conclusion, we have demonstrated efficient (> 95%) energy transfer between zero-dimensional semiconductor quantum dots and a two-dimensional semiconductor TMD. We have shown that, as a result of reduced dielectric screening in thin layers, the energy transfer rate is enhanced relative to the bulk. These results can lead
to improved optoelectronic devices in which the performance is enhanced through surface modification, electrostatic gating, and functionalization of TMDs.

### 3.3.5 Energy Transfer from to MoS$_2$ from QDs with Thin Shells and Short Ligands

As a control experiment, we measured fluorescence quenching of a second set of CdSe QDs. While the CdSe QDs in the main text have a thick CdZnS shell and long ligands (octadecylphosphonic acid), this second set has a thin CdZnS shell and short aromatic ligands (benzylphosphonic acid). As a result, the QD-MoS$_2$ distance in this second system is shortened to 3.9 nm, compared to 6.3 nm in the CdSe system discussed in section 3.3.3. Consistent with expectations, the reduced QD-MoS$_2$ distance results in faster energy transfer rates (see Fig. 3-12a). Qualitatively, we observe the same thickness scaling, with a monotonically decreasing lifetime with decreasing MoS$_2$ layer thickness, indicating an increase in the energy transfer rate to thinner MoS$_2$ flakes.
Figure 3-12: (a) Representative fluorescence lifetime histograms of the thin-shell CdSe QDs on different thicknesses of MoS$_2$. The black curve represents the instrument's response function (IRF). Inset: spectral overlap between thin-shell CdSe QD emission and MoS$_2$ absorption spectra. (b) Enhancement factor as a function of $t/d$, where $t$ is the underlying MoS$_2$ thickness, and $d$ is the distance from the quantum dot center to the MoS$_2$ surface. Red circles represent the thin-shell quantum dot data, which is in good agreement with the thick-shell quantum dot data presented in Fig. 3-11 in section 3.3.3 (blue circles) as well as with the fit of eqn. (3.10) (black line) to the thick-shell data.

3.4 Theory of Charge Transfer in 0D/2D Donor/Acceptor Systems

Much of this discussion follows the argument presented by Tisdale and Zhu [107].

Electronic coupling between a zero dimensional donor and an extended two dimensional electron acceptor is governed the energetic and spatial overlap of the donor and acceptor states. When a discrete donor and discrete acceptor interact the magnitude of their coupling is determined by the electronic exchange coupling energy $t = \int \psi_{\text{donor}}^{*} \hat{H} \psi_{\text{acceptor}} \, d\tau$, which requires wavefunction overlap between the donor and acceptor states. When a discrete donor interacts with an extended solid, the interaction occurs between the discrete donor wavefunction and across all $k$ states of the
acceptor. The strength of the electronic coupling in this case is represented by

$$\Delta = \sum_k \int_{\text{space}} \psi^*_{\text{donor}} \hat{H} \psi_{\text{acceptor}}^k \, d\tau. \quad (3.11)$$

When a discrete molecular state couples with an extended solid, its density of states broadens into a Lorentzian line shape with width $2\Delta$,

$$\rho_{\text{adsorbate}}(\epsilon) = \frac{1}{\pi} \frac{\Delta(\epsilon)}{(\epsilon - \epsilon_a - \Lambda(\epsilon))^2 + \Delta(\epsilon)^2}, \quad (3.12)$$

where $\epsilon_a$ is the original unperturbed discrete donor energy, $\Delta(\epsilon)$ is the projection of the bulk density of states onto the discrete adsorbate, and $\Lambda(\epsilon)$ is its Hilbert transform [108].

There are three regimes of electronic coupling leading to qualitatively different charge transfer behavior [107]. For a given system, the relative magnitudes of the electronic coupling $\Delta$ and the system's reorganization energy upon charge transfer $\lambda$ determine the charge transfer behavior.

When electronic coupling is relatively weak ($\Delta \ll 2\lambda/\pi$), the system is in the nonadiabatic regime. This regime occurs when the donor and acceptor are separated in space by nonconductive spacer ligands or if the energy of the donor state lies in a gap in the acceptor’s density of states. In this regime, the activation energy for charge transfer is mostly determined by the free energy change of charge transfer $\Delta G^0$ and the reorganization energy $\lambda$. The rate of electron transfer in this regime is governed by Marcus theory.

In the intermediate regime when $k_B T < \Delta < 2\lambda/\pi$, charge transfer is still a nonadiabatic thermally assisted process, but the activation energy for transfer is lowered by the electronic coupling $\Delta$.

In the strong coupling regime, $\Delta > 2\lambda/\pi$, charge transfer occurs adiabatically. This regime is achieved when the donor and acceptor are in close proximity or bridged by electronically conductive spacer ligands and requires that the donor energy fall in a region of the acceptor’s density of states with nonzero density. In this regime, the
donor and acceptor states hybridize to form a delocalized state and charge transfer corresponds to dephasing of the electronic coherence between the donor and acceptor.

3.5 Charge Transfer in CdSe Quantum Dot/WS$_2$ Monolayer Heterostructures

Hybrid structures containing both quantum dots (QDs) and 2D transition metal dichalcogenides (TMDs) leverage both the constituent 0D and 2D materials' favorable properties. 2D TMDs contribute high charge carrier mobilities [15, 17, 45] to mixed-dimensional heterostructures, while 0D QDs can provide strong, spectrally tunable broad-band absorption and efficient narrowband emission. Through careful selection of the 2D material and QD, the hybrid structure can be tuned to a specific application. Electronically insulated QDs have been used to sensitize 2D materials via resonant energy transfer, revealing anomalous energy transfer phenomena at 0D/2D interfaces [22, 103, 109] and enabling tunable interfacial coupling [110]. Furthermore, photodetectors with large gain and facile spectral tunability have been demonstrated by strongly coupling QDs to 2D materials by exchanging the QD native ligands to short, conductive spacer ligands [93, 111].

Boulesbaa et al. reported observing ultrafast (< 45 fs) charge transfer at the interface of monolayer WS$_2$ and CdSe QDs [112]. This interpretation is surprising because the CdSe cores were passivated by a 2 nm thick insulating ZnS shell and 18-carbon octadecylamine ligands (~ 2 nm dot-to-dot spacing), and strong electronic coupling is mediated by wave function overlap between donor and acceptor.

QD/2D hybrid interfaces are difficult to probe with traditional linear spectroscopies, since both the QD and TMD possess large absorption cross sections with overlapping spectral features. In addition, 2D materials are extremely sensitive to the surrounding dielectric environment [11, 39], and it can be difficult to distinguish the effects of dielectric screening from direct electronic coupling. To investigate tunable electronic coupling at a QD/WS$_2$ interface, we use time-resolved SHG.
Figure 3-13: **Characterization of the quantum dot (QD)/WS$_2$ heterostructure.** (a) Schematic of the QD/WS$_2$ interface. QDs are coated with alkanethiol ligands of tunable length. (b) Estimated band alignment at the hybrid interface, based on published ionization potentials, exciton binding energies, and measured optical gaps. (c) CdSe QD (dark, red) and WS$_2$ (light, orange) absorption spectra. The pump pulse spectrum (black, dashed) selectively excites the quantum dots in the heterostructure.
WS$_2$ monolayers were mechanically exfoliated from a bulk single crystal (2D Semiconductors) on a Si/SiO$_2$ substrate. CdSe QDs were synthesized by a previously reported method [113] and suspended in toluene. The QD excitonic peak absorption occurred at $\lambda = 646$ nm (Fig. 3-13c) corresponding to nanocrystals roughly 7 nm in diameter [114]. QDs were deposited onto the WS$_2$/SiO$_2$/Si substrate by spin coating and the native octadecylphosphonic acid surface ligands were subsequently exchanged for variable length alkanethiols in a layer-by-layer process [115]. In this method, $\sim 50 \mu$L of a 10 mg/mL QD suspension in toluene was spin-coated onto the WS$_2$/SiO$_2$/Si substrate at 1500 rpm for 30s. The surface was then covered with a 0.1 M solution of the desired alkanethiol (ethanethiol, 1-butanethiol, 1-octanethiol, or 1-dodecanethiol) in acetonitrile and allowed to soak for 30 s. The exchange solution was then spun off. Lastly, the sample was covered with pure acetonitrile and spun at 1500 rpm for 30 s to wash away free ligand. This process was repeated five times to form the nanocrystal film. The resulting interface is illustrated in Fig. 3-13a.

The estimated band edge alignment at the QD/WS$_2$ interface is shown in Fig. 3-13b. The valence band maxima for the CdSe cores and WS$_2$ relative to vacuum (the ionization potential, $\phi$) were taken from literature. Jasieniak et al. reported size-dependent CdSe QD valence band maxima energies, which the measured using photoelectron spectroscopy [116]. Keyshar et al. reported the work functions of monolayer TMDs on SiO$_2$ measured using photoelectron microscopy [117], reporting values for WS$_2$ in good agreement with electronic structure calculations [118]. Meanwhile, the energy of the excitonic transitions in WS$_2$ and our CdSe QDs are easily obtained from absorption spectroscopy. The WS$_2$ band gap was inferred by adding the reported exciton binding energy [39] to the exciton transition energy yielding a gap, $E_g$, in good agreement with scanning tunneling spectroscopy experiments [119]. The QD layer and WS$_2$ monolayer form a type-II heterojunction, which favors electron transfer from photoexcited QDs to WS$_2$. The QD (solution and film, see Supporting Information) and WS$_2$ absorbance are shown in Fig. 3-13c; the QD absorbance extends to significantly lower energy than the WS$_2$ monolayer, enabling selective photoexcitation of the QD in the heterostructure.
We use time-resolved SHG microscopy to take advantage of the symmetry properties of the separate QD and 2D components to isolate the pump-induced response of the TMD [120]. Monolayer 2H TMD materials, such as the WS$_2$ used in this study, are highly nonlinear and lack inversion symmetry. This results in a large second order nonlinear susceptibility that depends on the number of TMD layers [121, 122]. Meanwhile, a film of randomly oriented CdSe quantum dots has no appreciable SHG response. SHG is sensitive to charge transfer at polarized interfaces, rendering it uniquely well-suited to study this material system [123].
Our time-resolved SHG microscopy instrument is illustrated in Fig. 3-14a. A 76 MHz Ti:sapphire oscillator (Coherent, Mira-HP) generates 100 fs, $\lambda = 830$ nm pulses, which synchronously pump an optical parametric oscillator (OPO, Coherent, Mira-OPO). The OPO generates 100 fs pulses of tunable wavelength ($\lambda = 550$ to 700 nm), which we use to photoexcite the sample. A 0.15 mJ/cm$^2$ pulse at $\lambda = 650$ nm excites the QDs, and after a variable waiting period, SHG response of the sample is probed by a $\lambda = 830$ nm, 2 mJ/cm$^2$ pulse. Some fraction of the probe pulse is converted to the second harmonic, which is then transmitted through a short pass filter and detected by a photomultiplier tube (PMT). The pump-induced change in the SHG response is found by mechanically chopping the pump and using a gated counter to count SHG photons in the presence and absence of the pump pulse. Prior to entering the microscope, the pump and probe lines are spatially filtered and collimated to form diffraction-limited Gaussian spots at the sample (spot size $\sim 800$ nm in diameter). The sample was kept under vacuum in a sealed vacuum chamber, situated on top of the microscope stage, to prevent photoinduced damage in air.

The rate of electron transfer, which reflects the degree of electronic coupling in the hybrid system, is captured in early-time dynamics of the transient SHG response. We present the early time transient SHG signal, plotted as the change in the SHG intensity as a fraction of the intensity in the absence of the pump, in Fig. 3-14b. Experimental data are plotted as open symbols while fits to a simple kinetic model described later are plotted as solid lines. When bare WS$_2$ is excited with the $\lambda = 650$ nm pump, no transient SHG response is observed (see Supporting Information). In contrast, excitation of the QD/WS$_2$ heterostructure results in a sub-picosecond change in the SHG intensity. QDs capped with ethanethiol show the fastest rise time in the SHG signal, reflecting the fastest electron transfer rate. Longer alkanethiol capping ligands result in monotonically decreasing electron transfer rates as reflected in the elongated rise times.

The ethanethiol sample additionally shows a fast recovery of the SHG signal, likely reflecting direct electron-hole recombination across the interface after the initial electron transfer event, facilitated by particularly strong electronic coupling with this
short ligand. The fast recovery feature is less prominent in butanethiol coated dots and disappears entirely for octanethiol and dodecanethiol coated QDs. Fig. 3-14c shows the extracted electron transfer rate constants $k_{ET} = \tau_{ET}^{-1}$ for QDs capped with each ligand. The open circles represent values extracted from experiments performed at different locations on the sample. Filled circles represent the average time constant value (error bars represent the standard error). There is considerable variation in the electron transfer time constant at different sample locations, reflecting heterogeneity in the QD/WS$_2$ interface – presumably due to inconsistency in the degree of ligand coverage and orientation of the faceted QDs on the WS$_2$ surface. However, on average, we observe a monotonic increase in the electron transfer time constant as the capping ligand length is increased.

The observed electron transfer rate constants range from 1.4 to $> 100$ ps$^{-1}$, which is surprisingly fast for electron transfer from a CdSe QD. For comparison, optimal electron hopping rates between CdSe cores separated by atomically thin inorganic linkers are calculated to be 1 ps$^{-1}$ even when the interdot geometry is optimized for transfer [124]. Effective hopping rate constants in the most conductive CdSe QD solids are at most $\sim$ 4 ps$^{-1}$ [125]. One possible explanation for the transient SHG signal is ultrafast hole trapping at the QD surface [126]. Though this process is known to occur in CdSe QDs, the transient SHG signal dependence on capping ligand length makes such an interpretation less likely.

Fig. 3-15a illustrates several possible pathways available to a photoexcited charge in this experiment. The excited electron can relax back to the QD ground state before it has a chance to transfer ($k_{\text{decay}}$). Alternatively, it can transfer to the neighboring WS$_2$ conduction band with the electron transfer rate constant $k_{ET}$. Once transferred, the electron can recombine directly to the QD ground state with rate constant $k_{\text{recombine}}$ or diffuse in the WS$_2$ plane, leading to an effective relaxation rate constant $k_{\text{diff}}$. SHG data were fit to this kinetic model using the coupled differential
The time-dependent area densities of excited charges on QDs, \([QD^*](t)\), and WS\(_2\), \([WS_2^*](t)\), were initialized with initial condition \([QD^*](0) = N_0\) and \([WS_2^*](0) = 0\), reflecting QD-selective excitation. Traces were fit to this kinetic model to extract key rates such as \(k_{ET}\) (Fig. 3-14b). For direct comparison to experimental data, the modeled kinetics were convolved with the pump-probe sum frequency generation cross correlation.

Fig. 3-15b shows the SHG signal recovery dynamics for samples coated with the four capping ligands. The decaying portions of the traces are fit to biexponentials, which are plotted as thick, transparent lines. The long time dynamics reflect a complicated set of processes that occur following electron transfer, including lateral diffusion within the WS\(_2\) plane and subsequent recombination with a hole. The long-time relaxation rate constant, \(k_{\text{diff}}\), varied somewhat by sample location but did not show a clear dependence on QD ligand.

QDs capped with ethanethiol and butanethiol exhibited the fastest electron transfer dynamics. When transient SHG signals from these samples were collected for longer integration times (to improve signal-to-noise ratio), coherent oscillations in the transient SHG data became apparent (Fig. 3-16a); Notably, oscillations were not observable in octanethiol and dodecanethiol samples, supporting the interpretation that ultrafast charge transfer – rather than QD photoexcitation alone – leads to the intensity oscillations. Subtracting fitted biexponential decays from the data yields residuals that oscillate about the fit. Residuals from the sample with butanethiol-capped dots are plotted in Fig. 3-16b (black line) along with a damped sine wave (red line). The fitted frequency of this oscillatory component is \(\Omega \sim 14 \text{ cm}^{-1}\). In the low frequency region of their vibrational spectrum, CdSe quantum dots possess collective acoustic modes analogous to those of an elastic sphere [127, 128]. The low-
Figure 3-15: *Recovery of the transient SHG response*. (a) Following photoexcitation, an electron in the QD excited state can return to the QD ground state or transfer to the acceptor WS$_2$. From there, the charge can recombine directly to the QD or diffuse within the WS$_2$. (b) Representative SHG signal recovery dynamics are shown for WS$_2$ covered with ethanethiol- (blue, circles), butanethiol- (green, squares), octanethiol- (red, diamonds), and dodecanethiol- (violet, triangles) capped QDs. Across multiple samples, long time transient SHG traces show similar relaxation dynamics.

The highest energy collective vibration corresponds to a radially symmetric breathing mode. This mode is Raman active, and appears in the low-frequency Raman spectrum of the dots used in this system (collected with a previously reported experimental apparatus [114]). The Raman spectrum scattered by the 7 nm diameter dots is plotted in blue in Fig. 3-16c. The 14 cm$^{-1}$ Raman breathing mode matches the frequency of the picosecond oscillations in the transient SHG traces. The fitted damping rate, $\Gamma \sim 0.2$ ps$^{-1}$, corresponds to a $\sim 6$ cm$^{-1}$ linewidth, which is similar to the acoustic phonon linewidth in the Raman spectrum ($\sim 4$ cm$^{-1}$) and consistent with other time-domain measurements performed on CdSe QDs in solution [129].

We interpret the oscillations in SHG intensity as resulting from a coherent QD vibration, which modulates the nonlinear polarizability of the QD/WS$_2$ heterostructure by changing the local dielectric environment surrounding the WS$_2$ monolayer. Alternatively, the SHG intensity oscillation could result from the QD vibration modulating the charge density distribution at the interface. In either case, ultrafast electron transfer excites coherent acoustic vibration of the QD layer, which dephases over time.
Figure 3-16: Coherent phonon oscillations. (a) Transient SHG response for ethanethiol- and butanethiol-capped QD samples on WS$_2$, exhibiting coherent oscillation in the signal intensity. (b) Residuals of a bi-exponential fit to the ethanethiol data in (a) are plotted in black, and fitted to a damped sine wave, plotted in red, with a frequency of $\sim 14$ cm$^{-1}$. (c) The fit extracted an oscillation frequency $\sim 14$ cm$^{-1}$. (c) Raman spectrum of the same QDs, plotted in blue. The radially symmetric QD breathing mode scatters inelastically at a frequency matching the picosecond oscillations in the transient SHG data.

In conclusion, we engineered a QD/TMD system to facilitate fast charge transfer by placing CdSe QD cores directly on a WS$_2$ monolayer separated by short alkanethiol ligands. We adjusted the energetics of the system to exclude all processes except charge transfer and probed the charge transfer dynamics using time-resolved SHG. We were able to tune the rate of charge transfer by changing the QD capping ligand, adjusting the degree of electronic coupling at the QD/WS$_2$ interface. In the case of ethanethiol and butanethiol, the ultrafast transfer process coherently excited a QD acoustic phonon, modulated the SHG response in the time domain. Charge transfer is a fundamental process underlying 0D/2D hybrid optoelectronic devices and this work probes that process spectroscopically at a model interface. The work also demonstrates the advantages of using SHG to probe dynamics at TMD interfaces, which can be difficult to study with linear spectroscopy.
Chapter 4

Stimulated Second Harmonic Generation

4.1 Second Harmonic Generation

This discussion closely follows the presentation of nonlinear optics by Boyd [130].

Consider an electric field \( \tilde{E}(t) \) (here the tilde represents that the field is time-varying at optical frequencies). When the field \( \tilde{E}(t) \) is incident on a medium its electric field interacts with the charged particles in the medium (electrons and atomic nuclei) exerting a Coulomb force on the particles. The charged particles respond and are displaced from their equilibrium positions forming an electric dipole. The dipole moment per unit volume, \( \tilde{P}(t) \), is called the polarization. In the limit where the incident field is weak and the charge displacement reacts linearly to the field, the generated polarization has magnitude

\[
\tilde{P}(t) = \epsilon_0 \chi^{(1)} \tilde{E}(t),
\]

(4.1)

where \( \epsilon_0 \) is the vacuum permittivity and \( \chi^{(1)} \) is called the linear susceptibility. When the incident field is strong however, the polarization may not respond linearly to the field. In the time domain, this means a sinusoidal field may produce a polarization whose magnitude oscillates non-sinusoidally. In this strong field nonlinear regime, the
generated polarization can be expressed as a power series in the field strength:

\[ P(t) = \varepsilon_0 \left[ \chi^{(1)} \vec{E}(t) + \chi^{(2)} \vec{E}^2(t) + \chi^{(3)} \vec{E}^3(t) + \cdots \right] \]

\[ \equiv \tilde{P}^{(1)}(t) + \tilde{P}^{(2)}(t) + \tilde{P}^{(3)}(t) + \cdots . \]  

(4.2)  
(4.3)

The terms dictating the higher order polarization responses, \( \chi^{(2)} \) and \( \chi^{(3)} \) are the second- and third-order nonlinear susceptibilities.

---

Figure 4-1: (a) In the field picture second harmonic generation occurs via a nonlinear polarization response. The nonlinear polarization emits a field with components at the second harmonic frequency. (b) In the photon picture, second harmonic generation occurs when two photons with energy \( \hbar \omega \) combine to form one photon of energy \( 2\hbar \omega \). The total photon momentum is conserved.

The simplest example of a nonlinear optical process is second harmonic generation (SHG). In SHG, an intense field at frequency \( \omega \) interacts with a nonlinear medium and generates a second order nonlinear polarization response \( \tilde{P}^{(2)}(t) \) at twice the frequency, \( 2\omega \). This polarization response then radiates a second harmonic field converting a portion of the incident field to its second harmonic. This process is illustrated schematically in Fig. 4-1a. When light is considered as discrete occupied photon modes, SHG occurs when two photons of energy \( \hbar \omega \) are destroyed and a single photon of energy \( 2\hbar \omega \) is simultaneously created. This process does not require \( \hbar \omega \) or \( 2\hbar \omega \) to match transition energies in the nonlinear medium as the photons can interact.
with the medium \textit{via} virtual levels.

One salient feature of SHG (and all even-order nonlinear optical processes) is that nonlinear media with a center of inversion do not facilitate SHG. Many crystal classes have inversion symmetry and importantly, materials that are isotropic on optical length scales (such as liquids, some polymers, and randomly oriented nanocrystals) also possess pseudo inversion symmetry with regard to this symmetry consideration. The result that centrosymmetric media have vanishing even order nonlinear susceptibilities is only valid when the medium’s polarization responds instantaneously to the incident field, which isn’t strictly true. Regardless centrosymmetric nonlinear media facilitate SHG with very small, near-zero efficiency. This can be seen with a simple argument: the second order nonlinear polarization is again given by

\begin{equation}
\tilde{P}^{(2)}(t) = \epsilon_0 \chi^{(2)} \tilde{E}^2(t),
\end{equation}

where the applied field is \( \tilde{E}(t) = E_0 \cos(\omega t) \). If the sign of the applied field is changed, then the sign of the induced polarization \( \tilde{P}^{(2)}(t) \) must also change, because of the material’s assumed inversion symmetry. Consequently upon changing the sign of the applied field, eqn. 4.4 becomes

\begin{equation}
-\tilde{P}^{(2)}(t) = \epsilon_0 \chi^{(2)} \left[ -\tilde{E}(t) \right]^2 = \epsilon_0 \chi^{(2)} \tilde{E}^2(t).
\end{equation}

By considering eqns. (4.4-4.5), the assumption of inversion symmetry has implied that \( \tilde{P}^{(2)}(t) = -\tilde{P}^{(2)}(t) \). Consequently, \( \tilde{P}^{(2)}(t) \) and also the second order susceptibility \( \chi^{(2)} \) must be zero. This argument is illustrated in Fig. 4-2. The sinusoidal field (black, dashed) generates a sinusoidal response (red) in a linear medium. In a nonlinear medium, the sinusoidal field in general generates a complex, nonsinusoidal response. In a centrosymmetric medium, the response is symmetric about zero as shown in Fig. 4-2 (black, solid). This response contains exclusively odd harmonics of the incident field. In contrast, a noncentrosymmetric medium’s response to a sinusoidal field can be asymmetric about zero as shown in Fig. 4-2 (blue). This polarization response contains both odd and even harmonics of the incident field illustrating that
Figure 4-2: Nonlinear polarization response. When a sinusoidal field $E(t)$ (black, dashed) is applied to a medium, it generates a polarization response. The polarization response of a linear (red), nonlinear centrosymmetric (black, solid), and nonlinear noncentrosymmetric (blue) are shown.

4.2 Optical Stimulation Background

Before it was derived with quantum mechanics, stimulated emission was postulated by Albert Einstein using simple arguments involving the Planck Spectrum and a two level system depicted in Fig. 4-3.

Einstein's argument began from the Planck spectrum, the spectrum at which black-bodies were observed to emit. At thermal equilibrium, the absorption and emission balance such that the spectral density of radiation is characterized by the temperature, resulting in the Planck spectrum:

$$\rho(\nu, T) = \frac{8\pi h \nu^3}{c^3} \frac{1}{\exp(-\nu/k_bT) - 1}. \quad (4.6)$$

The argument for the existence of stimulated emission begins by considering a two level system (illustrated in Fig. 4-3) coupled to the radiation field. If one looks at the
spectral density at the frequency coupling the two states in the two level system, one can write the spectral density as a function of the ratio of the populations of the two states assuming a Boltzmann distribution,

$$\rho(\nu_0) = \frac{8\pi h \nu^3}{c^3} \frac{1}{N_1/N_2 - 1}. \quad (4.7)$$

This equation can be rearranged and both sides multiplied by the spontaneous emission rate, $A_{21}$, to yield

$$\left(\frac{A_{21}}{8\pi h \nu_0^3}\right) n_1\rho(\nu_0) = A_{21} N_2 + \left(\frac{A_{21}}{8\pi h \nu_0^3}\right) n_2\rho(\nu_0). \quad (4.8)$$

This equation represents the rates of absorption (left-hand-side), spontaneous emission (first term, right-hand-side), and stimulated emission (second term, right-hand-side) for the two level system depicted in Fig. 4-3. This argument resulting from Einstein’s analysis of the Planck spectrum implied that a two level system interacting with a radiation field couldn’t be completely described by (simulated) absorption and spontaneous emission. The analysis yielded stimulated emission with a rate proportional to the excited state population, $N_2$, and the spectral density at the frequency coupling the two states, $\rho(\nu_0)$. Eqn. (4.8) also shows that the constants related to
stimulated emission \((B_{21})\) and absorption \((B_{12})\) are equal:

\[ B_{21} = B_{12} = \frac{A_{21} c^3}{8\pi h \nu_0^3}. \]  

(4.9)

With these labels, eqn. (4.8) can be simplified to

\[ B_{12} N_1 \rho(\nu_0) = A_{21} N_2 + B_{21} N_2 \rho(\nu_0), \]  

(4.10)

where \(A_{21}\) and \(B_{21} \& B_{12}\) are commonly referred to as Einstein A and B coefficients respectively. Remarkably, Einstein had the foresight to postulate spontaneous emission, absorption, and stimulated emission by analyzing the Planck spectrum more than a decade before they were derived using quantum mechanics [131].

Although this result was derived through simple arguments centered around the Planck spectrum, a well-known result of quantum theory of radiation is that the rate of stimulated emission into a single mode of the field is equal to the spontaneous emission rate into that mode times the average number of photons occupying that mode. This formalism provides one qualitative explanation for all stimulated optical processes, included nonlinear optical processes such as stimulated Raman scattering and parametric amplification.

### 4.3 Enhancement of Second-Order Nonlinear-Optical Signals by Optical Stimulation

Much of Section 4.3 has appeared in print as Goodman and Tisdale Phys. Rev. Lett. (2015), [4].

#### 4.3.1 Second Order Nonlinear Optical Spectroscopies: Successes and Opportunities

Second-order nonlinear optical interactions are useful for probing the electronic and vibrational properties of surfaces and interfaces, measuring time-resolved interfacial
dynamics, and studying the structure of biological tissue. Because second-order nonlinear optical processes are dipole-forbidden in centrosymmetric media, second-order nonlinear signals are inherently surface and interface-selective. For instance, vibrational sum frequency generation (SFG) can inform our understanding of chemical bonding at solid surfaces and aqueous interfaces [132, 133] and time-resolved second harmonic generation (SHG) can be used to study the ultrafast dynamics of charge transfer at donor-acceptor interfaces [123, 134, 135]. In biological tissues, the inherent nonlinearity of SHG enables label-free 3D imaging of protein scaffolds [136].

Second-order nonlinear optical experiments are often limited by low nonlinear conversion efficiencies. The efficiency of these nonlinear optical processes is determined by the nonlinearity of the sample, the volume of material probed by the laser beam, and the incident pulses’ energies and durations. Even with the arrival of ultrafast pulsed lasers, weakly nonlinear media do not support efficient non-resonant SHG and SFG. Increasing the incident laser fluence can increase the conversion efficiency, but is often accompanied by sample photodamage. For many experiments, signal photon count rates are <100-1000 Hz; in such scenarios, the experiment’s signal-to-noise ratio (SNR) is bounded by $\sqrt{n}$ due to the unfavorable statistics of counting small numbers of photons, n.

Optical stimulation is an approach that has been successfully used to enhance other weak, inelastic scattering phenomena such as Raman and resonant inelastic X-ray scattering [137, 138]. In order to increase the efficiency of these inelastic scattering processes, light of the scattered signal frequency is coincident on the material with the pump, seeding the nonlinear conversion of the pump to the signal frequency. In the case of Raman scattering, optical stimulation has enabled many new technologies such as spectrally tailored microscopy and label-free video-rate imaging [139–141]. While optical parametric amplification is an example of a stimulated second-order process used in many labs to shift the frequency of ultrafast laser pulses, the use of optical stimulation to enhance weak signals in SHG spectroscopy or imaging has not been demonstrated.
4.3.2 Amplifying the Second Harmonic Generation Signal: Results

Here, we show the stimulated enhancement of SHG and difference frequency generation (DFG) in a configuration that is suitable for a wide variety of samples. We quantitatively describe the observed power and phase dependences using a coupled-wave formalism and achieve signal amplification of >10^4 in the biologically relevant sample collagen I. Our analyses indicate that the degree of signal amplification scales inversely with the sample's nonlinear susceptibility \( \chi^{(2)} \) and the distance over which the stimulating and fundamental fields interact, signifying that optical stimulation is most advantageous in systems with the weakest spontaneous signals.

In order to realize stimulated SHG, it is necessary to overlap pulses at the fundamental and second harmonic frequency in space, time, and direction at the sample. To achieve this, we used the modified Mach-Zehnder interferometer diagrammed in Fig. 4-4a. The output of a 76 MHz repetition rate Ti:sapphire oscillator producing 100 fs pulses in the near-infrared (\( \lambda = 830 \text{ nm} \)) was split into two equal intensity beams. The stimulating light (\( \lambda = 415 \text{ nm} \)) was generated in one arm of the interferometer using a phase-matched nonlinear optical crystal, \( \beta \)-barium borate (BBO). A mechanical delay stage and a piezo-mounted mirror were used to control the overall time delay and the relative optical phase, respectively, between the fundamental and stimulating laser pulses. The two beams were recombined using a dichroic mirror and focused collinearly onto the sample. The signal generated at the second harmonic frequency was sent either to a balanced photodiode (for detection of stimulated signals) or to a photomultiplier tube and gated photon counter (for detection of spontaneous signals and absolute-intensity calibration of the photodiode).

In contrast to spontaneous SHG, stimulated SHG signals are no longer background-free due to the presence of the incident stimulating field. Following analogous strategies used in stimulated Raman scattering [139], we modulated the fundamental beam at 3 kHz, as illustrated in Fig. 4-4b, to separate incident second harmonic intensity from that which was generated within the sample. The stimulated SHG signal was
Figure 4-4: (Color online) The stimulated SHG experiment (a) A schematic of the optical system used to perform stimulated SHG. (PMT: photomultiplier tube) (b) A visualization of the signal modulation scheme used to differentiate SHG signal from stimulating photons. (c) Stimulated SHG in $\beta$-barium boarate.

then demodulated by a phase-sensitive lock-in amplifier.

Initial studies were performed using the ideal nonlinear medium BBO as a model system. When the fundamental and stimulating pulses are not overlapped in time, there is a small $2\omega$ signal at the chopping frequency due to spontaneous SHG. When the time delay, $\Delta t$, approaches zero, the SHG signal is dramatically enhanced due to stimulation, as shown in Fig. 4-4c.

To describe the interaction between the fundamental and stimulating fields we adapt some of the arguments made in the seminal 1962 paper by Armstrong et al. [142], with the assumption that there is no momentum mismatch between the fields. When fields at the fundamental and second harmonic frequencies interact in a non-
Figure 4-5: Analysis of stimulated SHG using the coupled-wave formalism. Open circles are data from BBO, while black lines are analytical fits predicted by eqns. (4.14–4.16) (a) Solution of eqns. (4.14–4.16) with initial condition $\theta = -\frac{\pi}{2}$. (b) Dependence of the stimulated SHG signal on the composition of the incident fields. (c) Dependence of the signal enhancement $\gamma$ on the total incident intensity. (d) Dependence of the stimulated SHG signal on the fundamental intensity. (e) Dependence of the stimulated SHG signal on the stimulating intensity.

linear medium, their amplitudes are coupled. The fields exchange energy over a characteristic length, $l$, where

$$l^{-1} = 2\omega^2 \left( \frac{2\pi d_{\text{eff}}}{c^2} \right) k_{\omega}^{-1} \sqrt{I_{\text{total}}}.$$

(4.11)

Here, $d_{\text{eff}}$ is the effective susceptibility, which takes into account the orientation of the fields with respect to the sample, $c$ is the speed of light, $k_{\omega}$ is the fundamental field wave vector, and $I_{\text{total}}$ is the sum of the fundamental and second harmonic fields' intensities. If we normalize the field amplitudes to the total intensity, defining normalized field amplitudes $u_\omega$ and $u_{2\omega}$ such that

$$u_\omega^2 + u_{2\omega}^2 = 1$$

(4.12)
and defining the relative phase between the fundamental and stimulating waves to be

$$\theta = 2\phi_\omega - \phi_{2\omega},$$  \hspace{1cm} (4.13)

then exchange of energy between the normalized fields is described by the coupled set of differential equations

$$\frac{du_\omega}{d\zeta} = u_\omega u_{2\omega} \sin(\theta)$$  \hspace{1cm} (4.14)

$$\frac{du_{2\omega}}{d\zeta} = -u_\omega^2 \sin(\theta)$$  \hspace{1cm} (4.15)

$$\frac{d\theta}{d\zeta} = \frac{\cos(\theta)}{\sin(\theta)} \frac{d}{d\zeta} \left( \ln(u_\omega^2 u_{2\omega}) \right)$$  \hspace{1cm} (4.16)

where $\zeta = z/l$ is the [dimensionless] normalized propagation distance.

If the relative phase between the two waves is initially $\theta = \pm \frac{\pi}{2}$, then the relative phase does not change with propagation. The well-known $\theta = -\frac{\pi}{2}$ solution to eqns. (4.14–4.16), plotted in Fig. 4-5a, can be used to conceptually understand stimulated SHG. The derivative of the second harmonic intensity $dI_{2\omega}/d\zeta$ represents the rate of growth of the second harmonic field intensity; it is included in the figure as a dashed line. Spontaneous SHG occurs in the limit of $I_{2\omega} \to 0$ (the left edge of the plot), where conversion is slow and initiated by vacuum fluctuations. As $I_{2\omega}$ grows, its presence further accelerates conversion from $\omega$ to $2\omega$. Conceptually, the effect of stimulation can be understood as moving from a regime where $dI_{2\omega}/d\zeta$ is small to a regime where it is much larger. We note that introduction of an additional field in eqns. (4.14–4.16) accounting for the stimulating beam is unnecessary, since the stimulating field is identical to the spontaneously generated field, and the rate of $2\omega$ generation isn’t explicitly dependent on the field amplitude at prior zeta.

This effect can be seen clearly in Fig. 4-5b, where the intensity of the second harmonic generated within the sample is plotted against the fraction of the incident intensity in the stimulating field, while the total intensity was held constant. Data points on this plot can be mapped to $dI_{2\omega}/d\zeta$ at different values of $\zeta$ in Fig. 4-5a. For fixed total incident power, optimal nonlinear conversion occurs when one third of
the incident intensity is in the stimulating field, reflecting the maximum in $dI_{2\omega}/d\zeta$ visible in Fig. 4-5a.

We define the ratio of the stimulated signal intensity to the spontaneous signal intensity (under conditions of constant total intensity and 2:1 $I_\omega:I_{2\omega}$ stimulating ratio) as the "signal enhancement", $\gamma$. By examining the power dependences of the spontaneous and optimal stimulated signals, we predict the behaviour of $\gamma$:

\[
\frac{I_{\text{stim}}}{I_{\text{spont}} \propto I_{\text{total}}^{3/2} \times d_{\text{eff}} \times z}
\]

\[
\frac{I_{\text{spont}}}{I_{\text{total}}^{2} \times d_{\text{eff}} \times z^{2}}
\]

\[
\gamma = \frac{I_{\text{stim}}}{I_{\text{spont}}} \propto I_{\text{total}}^{-1/2} \times d_{\text{eff}}^{-1} \times z^{-1}
\]

Eqn. (4.17) implies that the signal enhancement $\gamma$ grows without bound as the extent of the nonlinear interaction decreases (i.e. the normalized propagation distance $\Delta \zeta$ decreases). Consequently, the advantage of stimulated SHG is greatest in precisely the situations where it is most needed: in weakly nonlinear media, small interaction volumes, or in media that do not admit the use of large incident powers. To demonstrate this point, the $I_{\text{total}}^{-1/2}$ dependence of $\gamma$ was tested in BBO. The results are in agreement with the prediction of eqn. (4.17) and are shown in Fig. 4-5c. The dependences of stimulated SHG on $I_\omega$ with fixed $I_{2\omega}$ and on $I_{2\omega}$ with fixed $I_\omega$ are shown in Fig. 4-5d and e respectively, also in agreement with the predictions of eqn. (4.17).

The relative optical phase $\theta$ defined in eqn. (4.13) determines the nature of the nonlinear conversion. The $\theta$-dependence of eqns. (4.14-4.16) suggests the possibility of changing the direction of energy flow between the fundamental and second harmonic fields. The two special cases $\theta = \pm \pi/2$ are illustrated in Fig. 4-6a. When $\theta = -\pi/2$, SHG occurs, whereas a phase of $\theta = \pi/2$ induces DFG. Though the stimulated SHG and DFG signals have similar magnitudes, they have the opposite lock-in signal phase $\Omega$ relative to the modulation of the fundamental beam, as illustrated in Fig. 4-6b. A stimulated SHG signal is in-phase with the reference modulation, while a stimulated DFG signal is $180^\circ$ out-of-phase. The relative optical phase of the two fields and, accordingly, the direction of energy flow between the fundamental and second har-
Figure 4-6: Effect of phase on stimulated SHG (a) Relative optical phase relations that lead to stimulated SHG (left) and DFG (right). (b) Effect of stimulated SHG and DFG on the signal phase $\Omega$ measured by the lock-in amplifier. (c) Dependence of the lock-in detected signal on the relative optical phase $\theta$. Grey regions indicate an SHG signal in phase with the reference ($\Omega = 0^\circ$), whereas the white regions represent a stimulated DFG signal measured to be out of phase with the reference ($\Omega = 180^\circ$).

monic can be finely controlled using a piezo-mounted mirror, as shown in Fig. 4-6c. We note that unwanted phase fluctuations due to air currents and optical table vibrations are the dominant noise source in our experiments and present an additional difficulty compared to spontaneous SHG. However, there are many possible strategies for alleviating phase instability [143–145] and an improved apparatus is under development.

To demonstrate the technique's utility, stimulated SHG signals were collected from bovine collagen I. Collagen I is a naturally abundant nonlinear material that
is frequently the target of bioimaging studies. SHG imaging of collagen fibrils can distinguish diseased and wild-type tissue morphologies [146] and has recently been used to determine single fibril diameters smaller than the Abbe limit [147].

Spontaneous and stimulated SHG signals from collagen I were recorded as a function of incident fundamental power. The results are plotted in Fig. 4-7. The spontaneous signal grows quadratically with incident power, while the stimulated signal increases linearly. Even at very small incident stimulating fluences (2.7 nJ/cm²), the signal amplification exceeds four orders of magnitude. In all other measurable ways, stimulated SHG in collagen I behaved identically to stimulated SHG in BBO.

It is important to note, however, that in thick, semi-ordered collagen samples, dispersion and additional momentum contributions from spatial frequencies present in the sample morphology would prevent phase matching and complicate analysis [148]. Our prepared collagen sample is 1-2 fibrils thick, which is much shorter than the length over which the fundamental and second harmonic fields’ relative optical phase would change due to momentum mismatch. Such quasi-perfect phase matching is also realized in surface and interfacial spectroscopy [123, 132-135], where we anticipate stimulated SHG will significantly advance our ability to explore interfacial physics at femtosecond timescales and sub-micron length scales.
Finally, it is worth noting similarities and differences between stimulated SHG and heterodyne SHG [149–155] or SFG [156, 157]. Both approaches utilize an additional optical field to amplify an otherwise weak nonlinear optical signal, leading to linear (instead of quadratic) dependence on laser power and interaction volume. Whereas heterodyne SHG typically involves mixing of a time-delayed reference signal with the SHG signal generated from the sample in a spectrometer, stimulated SHG involves mixing of temporally coincident fields within the sample itself. The salient difference is that in stimulated SHG power is actually transferred between two freely propagating fields. This has the notable benefit of allowing one to directly detect the intensity of the SHG field using a single channel detector, such as a photodiode, rendering the method suitable for high-speed imaging. Another potential benefit of the optical stimulation approach is the ability to detect changes in $I_\omega$ instead of $I_{2\omega}$, which could enable access to structures and interfaces buried within media that are absorptive or highly scattering at the second harmonic frequency.

4.4 Philosophical and Practical Differences between Optical Stimulation and Heterodyne Detection

Much of section 4.4 has appeared in print as Goodman and Tisdale *Phys. Rev. Lett.* (2016), [158] in response to a comment by Grechko *et al.*, [159].

4.4.1 Comparing Optical Stimulation to Heterodyne Detection

This subsection is a summary of a Comment that Grechko *et al.* submitted in response to the work described in section 4.3 [159].

The crux of the comment submitted by Grechko *et al.* is that the dependence of the stimulated SHG signal on the incident fundamental and second harmonic fields ($E_\omega$ and $E_{2\omega}$ respectively) can be explained using heterodyne arguments. In this scenario, the fundamental field generates some nonlinear polarization $P^{(2)}(E_{2\omega}; E_\omega, E_\omega)$ which
radiates a second harmonic field $E_{2\omega}$. As is observed in traditional spontaneous SHG experiments, the generated second harmonic field $E_{2\omega}^\text{SHG}$ is proportional to the incident fundamental field intensity $I_\omega$. This field is then amplified by the presence of the heterodyne (local oscillator) field $E_{2\omega}^\text{LO}$ through classical interference effects:

$$I_{2\omega}^\text{signal} = \left|E_{2\omega}^\text{signal}\right|^2 = \left|E_{2\omega}^\text{SHG} + E_{2\omega}^\text{LO}\right|^2 = I_{2\omega}^\text{SHG} + I_{2\omega}^\text{LO} + 2E_{2\omega}^\text{SHG} E_{2\omega}^\text{LO}.$$  \hspace{1cm} (4.18)

The amplified signal term in this traditional heterodyne amplification formalism is the term $2E_{2\omega}^\text{SHG} E_{2\omega}^\text{LO}$, which has the same proportionality as stimulated SHG, $|E_\omega|^2 E_{2\omega}^\text{LO}$.

Stimulated SHG as presented in section 4.3 and heterodyned SHG behave the same when the incident fundamental and second harmonic fields are scaled reflecting a fundamental similarity. Heterodyne amplification and detection are usually performed in two ways that differ fundamentally from the signal enhancement provided by stimulated SHG however: In pulsed experiments, the local oscillator is often trailing in time behind the signal pulse. Secondly, in many implementations, the local oscillator is interfered with the signal field after the sample using a beam splitter.

When the local oscillator (LO) is delayed behind the signal, the pulses are directed into a spectrograph and dispersed and interfered. The collected spectrum is a linear combination of the signal spectrum, the LO spectrum, and the product of the signal and LO fields multiplied by a cosine term reflecting the time delay (see Fig. 4-8). Though there is signal amplification in some regions of the spectrum, the spectral power is subtracted from other regions due to the different frequency components of the LO and signal pulse having different relative phases. Thus, no net power is transferred into the signal or removed from the fundamental pulse which is frequency doubled to the second harmonic.

Another common implementation combines the LO pulse with the signal pulse after the sample with a beam splitter (after all other field interactions occur). In this implementation the fields interact spontaneously at the sample and then the signal pulse is amplified by constructively interfering with the LO pulse. The amplification does not occur due to enhanced interaction at the sample, rather constructive
Figure 4-8: When a local oscillator pulse follows behind the signal pulse with a delay, $\tau$ it interferes with the signal to produce the illustrated spectrum. The spectrum oscillates between constructive and destructive interference as different frequency components of the pulses have differing relative phases. The red line shows the local oscillator and signal spectra summed in the absence of interference effects for comparison.

interference in the detected beam path after the beam splitter balances destructive interference in the other beam path. While traditional heterodyne detection implementations are similar to optical stimulation in some ways as Grechko et al. asserted in their comment, there are salient differences discussed in the following section.

4.4.2 Contrasting Optical Stimulation to Heterodyne Detection

With these descriptions of heterodyne detection in mind, it is important to emphasize that describing the signal enhancements described in section 4.3 as optical stimulation is valid. In general, a signal may be classified as stimulated when it is detected at a field mode that is populated by one of the incident fields [160]. However, a quantum description of the fields is not always necessary to adequately describe stimulated nonlinear optical processes. For instance, stimulated Raman scattering can be fully explained by a classical interference model [160]. In section 4.3, stimulated SHG and stimulated difference frequency generation (DFG) were described using a fully classi-
cal nonlinear wave mixing formalism [4, 142]. That classical description is equivalent to degenerate parametric amplification.

The argument by Grechko et al. that the observations in section 4.3 can also be explained by heterodyne detection is true. The seemingly disparate physical pictures of interference at the detector (heterodyne detection) and wave mixing in a nonlinear medium (stimulated SHG) can be reconciled within a quantum electrodynamical treatment of the process. When a quantum description of field and matter is invoked, the local oscillator (LO) is interpreted as a stimulating field that also interacts with the sample [161]. In this case, nonlinear wave mixing and heterodyne detection are two equivalent classical descriptions of the same interaction.

There are, however, practical differences between the approach described in section 4.3 [4] and heterodyne detection, as it is traditionally performed [149, 151, 157, 162] (see Fig. 4-9). When the LO bypasses the sample, or when it trails behind the fundamental pulse in time, the residual fundamental field is unperturbed. In contrast, when the fundamental and harmonic fields are temporally coincident in the sample, power is transferred between the fundamental and harmonic fields.

The stimulated SHG technique has several features of potential practical importance: (1) The SHG signal is mirrored in the fundamental field. This could be useful when a medium is scattering or absorptive at the harmonic wavelength but considerably more transparent at the fundamental (typically near-infrared) wavelength such as biological tissue [163]. (2) In highly scattering media (again, biological tissue is an example), spatial coherence of the backscattered spontaneous SHG signal may be lost [148], making the signal difficult to amplify by traditional heterodyne approaches. Stimulated SHG can succeed under these circumstances. (3) As with parametric amplification, stimulated SHG can be performed in a noncollinear geometry (provided the sample is thin enough). This presents the possibility of obtaining stimulated SHG signals background-free, which would deliver signal-to-noise improvements beyond what is achievable by traditional approaches to heterodyne detection.
Figure 4-9: Demonstration of the difference between heterodyne SHG, as classically implemented (a), and stimulated SHG, as described in [4] (b). In stimulated SHG or DFG, power is transferred between the fundamental and second harmonic fields.

4.5 Phase-Sensitive Nonlinear Optical Imaging of Two-Dimensional Materials

The stimulated SHG signal is sensitive to the relative optical phase of the incident fundamental field $E_\omega$ and stimulating field $E_{2\omega}$. The relative optical phase (eqn. 4.13) $\theta = 2\phi_\omega - \phi_{2\omega}$ affects the rate of power transfer between the two fields as described in eqn. (4.14). The rate of power transfer between the fields is proportional to $\sin (\theta)$ (in other words, $\theta$ affects the magnitude and direction of the rate of power transfer). In this section, we describe an alternative implementation of stimulated SHG; The relative optical phase of the incident fields $\theta$ is modulated between 0 and $2\pi$ resulting in MHz modulation of the stimulated signal between SHG and degenerate difference frequency generation. We call this technique Degenerate Parametric Amplification (DPA).
Figure 4-10: Phase-modulated degenerated parametric amplification (DPA) microscopy. (a) Schematic illustration of the optical apparatus and associated components. In this embodiment, $\lambda = 830$ nm and 415 nm pulses generated by a Ti:sapphire laser are cross-polarized at the sample and the fundamental field intensity is measured in transmission using a standard photodiode. (b) Illustration of the working principle of the technique: the electro-optic modulator (EOM) modulates the relative optical phase, $\phi$, between the two laser pulses at 1 MHz according to a saw tooth waveform at $2\pi$ modulation depth, generating a sinusoidal variation of the intensity in each field. (c) DPA signal amplitude map of a triangular monolayer MoS$_2$ flake. (d) Bright-field optical image of the same flake.
The phase-modulated degenerate parametric amplification microscopy technique, as illustrated schematically in Fig. 4-10a, is performed by rapidly oscillating the relative optical phase between the incident $\omega$ and $2\omega$ fields so that power transfer changes direction at MHz frequency. The transmitted laser power of the fundamental ($\lambda = 830$ nm) or second harmonic ($\lambda = 415$ nm) beam is detected using a standard photodiode, and the magnitude and phase of the DPA signal is retrieved using a lock-in amplifier. We found that high-frequency phase modulation overcomes difficulties associated with low-frequency phase noise from vibrations, air currents, and other instabilities on the optical table. Furthermore, careful choice of modulation depth (exactly $2\pi$) and waveform (saw tooth) is essential for making the measurement robust against long-term thermal drift.

The excellent signal contrast achievable with DPA microscopy is demonstrated on a single-layer MoS$_2$ flake, shown in Fig. 4-10c-d. For these measurements we used cross-polarized fundamental and second harmonic laser pulses ($\sim 110$ fs) generated by a Ti:sapphire laser and detected the residual fundamental field intensity in transmission mode using a photodiode.

4.5.1 Phase-Sensitive Imaging of 2D Materials

One of the most critical challenges in characterizing two-dimensional materials is unambiguous determination of the absolute crystal orientation [164]. This information is essential for the fabrication of anisotropic 2D heterostructures and devices [165], and for the understanding of 2D materials growth processes [166]. In addition, grain boundaries themselves provide opportunities for the discovery of novel physical phenomena [167, 168]. Photoluminescence and Raman are not strongly sensitive to domain rotations, and polarized SHG cannot distinguish between domains with mirror symmetry [169–171]. Advanced transmission electron microscopy (TEM) techniques such as high resolution TEM combined with electron diffraction or scanning-TEM (STEM) can be used to precisely determine crystal orientation and identify grain boundaries, but these techniques are not compatible with optical and electronic substrates [65, 172, 173], thus preventing in-situ examination and requiring sample
preparations that are delicate and tedious. Here, we show that DPA microscopy is a facile way to retrieve this structural information.

Monolayer MoS$_2$ has $D_{3h}$ symmetry with non-vanishing second-order nonlinear susceptibility tensor elements $\chi^{(2)}_{yy} = -\chi^{(2)}_{xx} = -\chi^{(2)}_{xy} = \chi^{(2)}_{yx}$ [121, 169], where the coordinate system is defined with respect to the MoS$_2$ atomic lattice as shown in Fig. 4-11. For the case of cross-polarized fundamental and second harmonic fields, the change in fundamental field intensity is given by

$$\Delta I_0 \propto \chi^{(2)} \cos(3\theta) I_0 \sqrt{I_{2\omega}} \sin(2\phi_\omega - \phi_\omega),$$

where $\theta$ denotes the orientation of the MoS$_2$ crystal with respect to the fundamental field polarization, as illustrated in Fig. 4-11. Note that both the sign of $\chi^{(2)}$ and the orientation of the MoS$_2$ crystal determine whether $\Delta I_0$ is positive or negative. A 180° rotation of the crystal ($\theta = 0°$ and 180°) will generate an equal amplitude signal, but opposite lock-in phase, as demonstrated in Fig. 4-12a-b.

In lower-symmetry materials, DPA can be used to determine each component of the full nonlinear susceptibility tensor. Each combination of wave vector (with respect to a defined crystal axis), fundamental field polarization, and second harmonic

\[\chi^{(2)}_{yx} = \chi^{(2)}_{xy}, \chi^{(2)}_{yy} = \chi^{(2)}_{xx}, \chi^{(2)}_{yx} = \chi^{(2)}_{xy}.\]
Figure 4-12: DPA imaging of MoS$_2$ crystals. (a,b) Rotational dependence of the DPA amplitude and phase for the triangular MoS$_2$ single-crystal flake shown in the inset. Red and blue colors indicate a lock-in phase of 0° or 180°, respectively. (c) Bright-field image of the multi-grain MoS$_2$ flake analyzed in panels d-i. (d,e) DPA amplitude and phase maps. (f,g) Line cuts taken along the dash-dotted lines through the images in panels d and e. (h) Universal curve used to determine the crystal orientation of each grain labelled in panel d. (i) Illustrated crystal orientation of the seven grains labelled in panel d.
polarization can be chosen to isolate a specific $\chi^{(2)}$ matrix element [130]. By systematically varying the crystal orientation and input field polarizations, the complete second-order nonlinear susceptibility tensor can be mapped. This approach can be extended to surface and interfacial phenomena in centrosymmetric media, too.

In Fig. 4-12c, we show a CVD-grown monolayer MoS$_2$ flake composed of multiple crystal grains with varied orientation. The corresponding DPA amplitude and phase maps are shown in Fig. 4-12d-e. The combined amplitude and phase information was used to determine the absolute orientation of each crystal grain, as demonstrated in Fig. 4-12h-i. After identifying the crystal orientation of all grains, the tilt angle between adjacent grains could be determined. The grains I, II, III and IV are commonly observed neighboring mirror twins with a relative tilt angle of 60°, and their grain boundaries are comprised of 4|8-member or 4|4-member rings [65, 172]. The grains V, VI and VII have a relative tilt angle of 51°; IV and V have a relative tilt angle of 27°; and I and VII have a relative tilt angle of 33°. These grain boundaries have been predicted by atomistic simulations or observed by other techniques, but the atomic structures have not been identified [167, 170].

4.5.2 Spatial Resolution of DPA

Lastly, we demonstrate the capability of resolving grain boundaries with high signal contrast and spatial resolution using DPA microscopy. In Fig. 4-13, we show DPA amplitude and phase maps in the vicinity of a mirror twin boundary in a monolayer MoS$_2$ flake. The DPA amplitude goes to zero at the grain boundary due to destructive interference between the second harmonic fields originating from the left and right sides of the boundary. Line scans of the amplitude (Fig. 4-13c) and phase (Fig. 4-13f) reveal diffraction-limited resolution in the amplitude map and sub-diffraction limited resolution in the phase map. The excellent spatial resolution of the phase map is due to the binary nature of this signal; it must be either 180° or 0°, and the spatial resolution is ultimately only limited by the signal intensity.
Figure 4-13: Characterization of grain boundaries. DPA amplitude (a-c) and phase (d-f) mapping near a mirror twin boundary in a monolayer MoS$_2$ flake.

4.5.3 Conclusion

Phase-modulated degenerate parametric amplification offers many advantages over conventional polarized SHG or heterodyne SHG. First, the signal is contained in both the fundamental and harmonic fields, so that either beam can be detected as experimental circumstances necessitate. This is particularly useful if the medium – or substrate – is highly scattering or absorptive at the second harmonic wavelength. Furthermore, the signal intensities are large enough to be detected with a standard silicon photodiode under normal room lighting, which is amenable to high-speed signal acquisition. Beyond 2D materials characterization, we expect DPA to find use in bioimaging, interfacial spectroscopy, and ultrafast microscopy.
4.5.4 Methods

*Phase-modulated degenerate parametric amplification (DPA) microscopy* – A Ti:sapphire oscillator (Coherent Mira HP) generated 830 nm laser pulses of ~110 fs pulse duration at 76 MHz repetition rate. The $\lambda = 830$ nm laser beam passed through a small aperture (to produce a near-Gaussian beam), a half-wave plate, a Glan-Taylor polarizer (GT5, Thorlabs), and was then focused into a 0.1 mm thick $\beta$-barium borate crystal (BBO, Type 1, Eksma Optics) by a 50 mm focal length lens ($\sim 8$ nJ pulse energy) to generate $\lambda = 415$ nm laser pulses ($\sim 40$ pJ pulse energy). The output beams were re-collimated by a 50 mm focal length lens, and then separated by a dichroic beam splitter (042-0845, > 99.5% reflectivity at 800 nm, Eksma Optics). Residual 830 nm light was removed from the 415 nm beam using a colored glass filter (FGS550, Thorlabs), then the 415 nm beam was sent directly to the EOM (see details below) and then into an optical delay line consisting of mirrors mounted on a linear translation stage (462-X-M, Newport Corporation) driven by a piezoelectric inertia actuator (ZBT225B, Thorlabs). The second harmonic and fundamental paths were re-combined by a beam combiner (042-4805 Eksma Optics, > 99.5% reflectivity at 390-410 nm), and were directed into an inverted optical microscope (Nikon Ti-U). The two beams were focused onto the sample by a microscope objective lens (Nikon, CFI S Plan Fluor ELWD, 40×, 0.6 numerical aperture), and then directed to an amplified Si photo-detector (PDA 36A, Thorlabs) by a 25 mm focus length condenser lens. When the 830 nm (415 nm) laser beam was detected, a long (short) pass filter FGL550S, Thorlabs (FGB 39, Thorlabs) was place in front of the detector. The signal from the photodetector was fed directly into a lock-in amplifier (HF2LI, Zurich Instruments), and the output was recorded and analyzed by a computer. For spatial mapping, the sample was scanned relative to the stationary focal point using a piezo stage (P-545.xR8S PI nano XY Piezo System, Physikinstrumente). For the sample rotation dependent experiments, a motorized precision rotation stage (PRM1Z8, Thorlabs) was used.

A function generator (4063, BK Precision) generated the saw-tooth wave for the
EOM driver (Model 275, Conoptics) and supplied the reference signal for the lock-in amplifier. The EOM was a potassium deuterium phosphate (KD*P) phase modulator (M350-160 phase, Conoptics). To minimize phase noise, the separated beam paths (dual beam paths from separator to combiner, including the EOM) were placed inside a homemade box to minimize the effects of external disturbances. Inside the box, the EOM was placed on a water-cooled breadboard (Thorlabs) fed by a re-circulating chiller (Coherent).
Bibliography


