Title
Probing Optical 'Dark' Effects in Artificial and Natural Nano-structures

Permalink
https://escholarship.org/uc/item/1pk207g9

Author
Ye, Ziliang

Publication Date
2013

Peer reviewed|Thesis/dissertation
Probing Optical ‘Dark’ Effects in Artificial and Natural Nano-structures

By
Ziliang Ye

A dissertation submitted in partial satisfaction of the requirements for the degree of Doctor of Philosophy in Applied Science & Technology in the Graduate Division of the University of California, Berkeley

Committee in charge:
Professor Xiang Zhang, Chair
Professor Feng Wang
Professor Jie Yao

Fall 2013
Abstract

Probing Optical ‘Dark’ Effects in Artificial and Natural Nano-structures

by

Ziliang Ye

Doctor of Philosophy in Applied Science and Technology

University of California, Berkeley

Chair: Professor Xiang Zhang

In the nano-scale world, many interesting optical effects cannot be detected in the far field zone by linear probes. Nevertheless, they are promising to bring numerous novel optical functionalities while some already play dominant roles in determining the overall optical responses. This dissertation presents a consistent endeavor to study these fascinating optical dark effects, using a wide range of tools including, but not limited to, the nearfield scanning optical microscopy, the nonlinear spectroscopy/microscopy and the photocurrent scanning microscopy. The studied systems are classified into two categories: artificial plasmonic structures and natural two dimensional transition metal dichalcogenide.

Among the plasmonic artificial structures, we focus on plasmonic antennas. By fine-tuning the nearfield coupling between them, we demonstrate a range of classical analogs to many interesting atomic effects, such as electromagnetic induced transparency and anti-Hermitian coupling induced superradiance. Both of them are dark to the far field, but can serve as effective color sorters below the diffraction limit within the near field zone. Employing the resonance nature of plasmonic antenna, we demonstrate for the first time the photonic spin Hall effect at the metasurface. A large spin orbit coupling induced deflecting is observed even at the normal incidence condition. Also, with nearfield microscopy, we map the mode profile of the hybrid waveguide. The optical mode is found to be highly confined within the gap between the metal and dielectric layer, which has a good potential to become the future platform for integrated active optical components.

In the study of 2D transition metal dichalcogenide, we discover a series of dark excitonic resonance with a set of selection rules complimentary to the linear transition, provided by two-photon absorption process. These resonances are originated from a very large binding energy in a unique 2D exciton model. A TMDC exciton based light emitting device is further demonstrated as a potential application. With the nonlinear probes, both the crystal orientation and an edge state resonance are identified, which has an otherwise zero or negligible response to the linear probe. Finally, we extend the study to explore the valley Hall effect in TMDC. Different photocurrent mechanisms are investigated and a laser polarization dependent transverse photocurrent is observed.
Dedicated to my parents
Jun Huang
and
Min Ye
Contents

1. Introduction 1
   1.1. Introduction and motivation 1
   1.2. Outline of thesis 2

Section I: Nearfield effects in plasmonic nano structures 4

2. Apertureless nearfield scanning optical microscope 4
   2.1. Diffraction limit 4
   2.2. Principle and configuration of a-NSOM 6
   2.3. High harmonic demodulation and pseudoheterodyne detection 7

3. Nearfield coupling in the plasmonic antenna array 11
   3.1. Introduction to the plasmonic antenna 11
   3.2. Coupling between antennas: coupled resonator model 13
   3.3. Dark transition in Plasmonic Interference Transparency (PIT) 15
       3.3.1 Introduction to Electromagnetic Induced Transparency (EIT) 15
       3.3.2 Analog with coupled plasmonic antenna 17
       3.3.3 Experimental results: from far field to near field 18
   3.4. Anti-Hermitian coupling in plasmonic antenna array 22
       3.4.1. Introduction to Anti-Hermitian coupling in open quantum systems 22
       3.4.2. Analog in plasmonic antenna array 23
       3.4.3. Numerical investigation of the critical coupling point 25
       3.4.4. Experimental results 28
   3.5. Conclusion and outlook 30

4. Photonic spin Hall effect at the metasurface 32
   4.1. Introduction 32
       4.1.1. Optical spin Hall effect at the interface 32
       4.1.2. Plasmonic antenna based metasurface 35
   4.2. Experimental results and discussion 36
       4.2.1. Polarization rotation induced by SHE 37
       4.2.2. Trajectory shift induced by SHE 40
   4.3. Numerical simulation 41
   4.4. Conclusion 42

5. Imaging the optical confinement effect in the hybrid waveguide 43
   5.1. Introduction to the hybrid waveguide theory 43
   5.2. Experimental demonstration and discussion 46
       5.2.1. Nearfield characterization 46
   5.3. Conclusion 51
Section II: Dark excitons and their optical properties in TMDC monolayer

6. Transition Metal Dichalcogenide (TMDC): structures, properties, characterizations and fabrications
   6.1. Introduction to few layer MoS$_2$ 53
      6.1.1. Atomic structure 53
      6.1.2. Electronic structure and optical properties 54
   6.2. Sample preparation and characterizations 56
      6.2.1. Sample preparation and monolayer identification 56
      6.2.2. Monolayer characterizations 59
   6.3. Transfer methods 60
   6.4. Fabrication processes 61
      6.4.1. Electron beam lithography 61
      6.4.2. Metal contacts 62
      6.4.3. Etching 63
   6.5. Conclusion and outlook 64

7. Dark excitons and nonlinear responses in TMDC 65
   7.1. Introduction to the nonlinear response of monolayer TMDC 65
   7.2. TMDC Dark exciton probed by TPA spectroscopy 66
      7.2.1. TMDC exciton and TPA selection rule 66
      7.2.2. One-photon and two-photon absorption spectrum 68
      7.2.3. Binding energy and exciton radius 73
   7.3. Crystal orientation and edge state imaging by SHG 75
      7.3.1. Imaging the layer thickness and crystal orientation 75
      7.3.2. Edge state nonlinearity 78

8. Electroluminescence from a silicon-MoS$_2$ PN junction 81
   8.1. Introduction 81
   8.2. PN junction: design, fabrication and electrical performance 81
   8.3. Electroluminescence from direct bandgap exciton and defect exciton 83
   8.4. Conclusion and outlook 88

9. Towards the valley Hall effect in TMDC 90
   9.1. Introduction to the valleytronics and valley Hall effect 90
   9.2. Device design and electrical performance 93
   9.3. Origin of the longitudinal photo current 95
   9.4. Polarization dependence of the transverse photocurrent 102
   9.5. Conclusion and outlook 105

Reference 107
List of figures

Fig. 2.1: Constant frequency contour in an isotropic medium, the k circle is the boundary between the propagating zone and evanescent zone.

Fig. 2.2: Schematic of the a-NSOM configuration. A silver-gallium alloy needle tip is used in the plasmonic attena array experiment, to be discussed in details in Chapter 3.

Fig. 2.3: A Z dimension scan of the demodulated near field signal.

Fig. 2.4: Simulated high harmonic signals together with the side band signals produced by the phase modulation of the reference beam.

Fig. 2.5: Schematic of a-NSOM with the interferometer section.

Fig. 2.6: (a) Phase of the propagating of the surface plasma. (b) Amplitude of the field. Scale bar: 2µm.

Fig. 3.1: Plasmonic antenna bridge the gap between the free space propagation and quantum emitter (a), as well as receiver (b). Adapted from Ref. 27.

Fig. 3.2: (a) Illustration of the first resonance condition of the classical antenna coupled with a transmission line. (b) Effective wavelength for the plasmonic antenna resonance in the optical range. Dotted black line, dashed blue line and solid red lines represents plasmonic antennas of different diameters. Adapted from Ref. 1.

Fig. 3.3: The energy-level diagram and simulated near-field intensity spectra for 30-nm-high, 50-nm-wide, and 110-nm-long (arm length) symmetric two-wire gold antennas with gaps of 6 nm (black dashed line) and 16 nm (gray dash–dotted line), as well as for a single-wire antenna with the same dimensions (black solid line) Adapted from Ref. 1.

Fig. 3.4: Two spring-ball systems are coupled by another spring. The coupling results in two new collective modes, one eigen mode of higher eigen energy is called the anti-bonding mode and the other one of lower eigen energy is called the bonding mode. Adapted from Ref. 1.

Fig. 3.5: EIT Energy levels and allowed transition path between the levels. The control beam and the probe beam are denoted by $\omega_p$ and $\omega_c$. Adapted from Ref. 36.

Fig. 3.6: Absorption of EIT. When a coupling laser beam is applied, the original absorption peak becomes a dip and a large group index is introduced into the medium.
Fig. 3.7: Illustration of the coupling design in the PIT structure. (a) PIT can be seen as the deconstructive interference between two different paths. (b) A hybridization picture between a bright and a dark antenna.

Fig. 3.8: (a) The SEM images of the PIT sample fabricated by e-beam lithography with the dimensions indicated. Scale bar is 300nm (b) The transmission spectrum with the polarization of incident light, indicated in (a), measured by Fourier Transform Infrared Spectrometer (solid) and the numerical simulation results (dashed) by CST Microwave Studio. (c) The near-field spectrum of SPP oscillation amplitude of the electric dipole (red dashed) and quadruple (green solid), respectively. Three typical conditions with different contrasts between dipole and quadrupole are marked by dash lines of different colors. The near field experiments are operated at these three frequencies.

Fig. 3.9: Local electric field intensity mapped by the a-NSOM. (a) The near field result measured at 1310 nm. The NSOM signal are rendered in green and superimposed onto AFM topography results in grey. Only four corners of the gold rod pair light up, which stands for the dominant population of SPP in electric quadrupole over the dipole. The mere excitation of the quadrupole accounts for the transparency in the far field. Scale bar is 300nm. (b) The near field result measured at 1530 nm. The excitation locations of the localized SPP are transferred to the ends of dipole antenna.

Fig. 3.10: Near-field result at the dress state frequency. (a) The near-field mapping result at 1310 nm, the dressed state wavelength of another sample. The optical signal, represented by the blue hotspots at the ends of both single rod and rod pair, indicate that SPP in the two resonators have comparable populations at this condition. Scale bar is 300 nm. (b) Numerical simulation of the structure at the same frequency. Blue color stands for positive $\pi/2$ and red represents $-\pi/2$ phase (c) The schematic of the phase of the electric field based on the experimental and the numerical results. The signs stand for the instantaneous charge accumulated at the ends of the antenna.

Fig. 3.11: A plasmonic system consisting of two optical antennas made of gold on a quartz substrate. The lengths of the antennas are 430 nm and 470 nm. The width and thickness of both antennas are 40 nm and 25 nm, respectively. A beam is incident at the normal direction onto the antennas, with polarization along the longitudinal direction of the antennas.

Fig. 3.12: (a) The retrieved real part (black) and imaginary part (red) of optical coupling constant between the two antennas, and the dissipations of the long (blue) and short antennas (green) as a function of their edge-to-edge separation $s$. (b) The simulated (solid) and fitted (dashed) spectra of the electric field probed at the center of the gap of two coupled antennas with lengths of 430 nm (black) and 470 nm (red) at separations of 30 nm, where the coupling between the two antennas is almost purely anti-Hermitian. The responses of the two antennas show highly
asymmetric profiles, a characteristic of Fano lineshape. The vertical dashed lines mark the eigen-frequencies in each figure.

Fig. 3.13: The schematic of a plasmonic antenna array consisting of five optical antennas with gradually varying lengths, ranging from 430 nm to 590 nm at a step of 40 nm. The thickness and width of the metal strips are 25 nm and 40 nm, respectively. The edge-to-edge separation between the nearest neighbors is 45 nm.

Fig. 3.14: (a) The calculated electric field magnitude at the gap center of each antenna versus the optical wavelength. At the resonance of each antenna, all the other antennas are strongly suppressed, leaving a highly selective excitation of a single antenna. (b) The spectral response for five uncoupled antennas with the same geometries as (a).

Fig. 3.15: Experimental verification of the selective excitation of individual antennas in the plasmonic antenna array. (a-e) The simulated near-field distributions of the antenna array at five different wavelengths: 1200 nm (250 THz), 1300 nm (230.8 THz), 1400 nm (214.3 THz), 1500 nm (200 THz) and 1600 nm (187.5 THz), showing selective excitation of individual antennas. (f-j) The corresponding experimental observations show very good agreement with the simulations. The optical image (red), measured by NSOM, is overlapped onto the topographical image (grey) measured by the atomic force microscopy.

Fig. 3.16: The measured intensity of light integrated along the y direction, versus x direction, at each measured wavelength labeled in fig. 3.15.

Fig. 3.17: The near field measurement at 1400 nm wavelength on the control sample, which consists of an array of antennas with the same geometry specification but at a large nearest neighbor separation of 300 nm. Without the anti-Hermitian coupling, all the antennas within the focus spot, which is around 2 mm, are excited simultaneously due to the spectral overlapping of the resonances. Scale bar is 300 nm.

Fig. 4.1: Illustration of the optical spin Hall effect at the interface between air and SiO2 (index: 1.515). As the beam is obliquely incident onto the interface, the orthogonally circular polarized beam are shifted laterally towards different directions.

Fig. 4.2: Plane wave decomposition of the Gaussian beam. The polarization of wave component with non-zero in-plane wave vector will rotate after transmitting through the interface.

Fig. 4.3: Calculation based on the geometry in fig. 4.1. H and V are the incident polarization directions; + and – stand for the left and right circular polarized components in the output beam. Adapted from Ref. 63.

Fig. 4.4: Scanning electron microscope (SEM) image of the metasurface structure working in the near infrared range. Adapted from Ref. 65.
Fig. 4.5: The illustration of the new wave front composed of the scattering wave from eight different V shape antennas. The tilting angle is $\lambda_0 / \Gamma$.

Fig. 4.6: (a) Schematic of PSHE. When light is propagating along a curved trajectory, the transversality of electromagnetic waves requires a rotation of polarization vector in the transverse planes along the trajectory. (b) The strong spin-orbit interaction within the optically thin metamaterial leads to the accumulation of circular components of the beam in the transverse directions ($y'$-directions) of the beam even when the incident angle is surface normal. Adapted from Ref. 64.

Fig. 4.7: Light from a broadband source was focused onto the sample with a lens ($f = 50$ mm). The polarization can be adjusted in both $x$- and $y$-orientations with a half-wave plate. The regularly and the anomalously refracted far-field transmission through the metasurface was collected using a lens ($f = 50$ mm) and imaged on a CCD camera. The polarization state of the transmission is resolved by using an achromatic quarter-wave plate, a half-wave plate, and a polarizer with high extinction ratio.

Fig. 4.8: (a) Observation of PSHE: The helicity of the anomalously refracted beam. The incidence is from the silicon side onto the metasurface and the polarized in $x$-direction along the phase gradient. The incidence angle is zero degree. The pseudo color red and blue represent right- and left-circularly polarizations. (b) The helicity of the refracted beam from a $y$-direction polarized input beam.

Fig. 4.9: (a) The wavelength dependences of the refraction angles when excitation is normally incident onto the metasurface. The measurement was conducted for three samples with different phase gradients, 3.6, 4.0, and 4.4 rad/mm, respectively. The measured refraction angles agree well with the theoretical predictions (solid curves). The inset schematically shows the light trajectory for the anomalous refraction for the surface normal incidence. (b) The transverse motion between the weight centers of the right- and left-circular polarized refractions, showing anomalous PSHE effect over a broad range of wavelengths.

Fig. 4.10: Simulated PSHE of circularly polarized light for a metasurface with a phase gradient of 4.4 rad/\textmu{}m.

Fig. 5.1: A typical hybrid waveguide geometry. The plasmonic mode at the metal surface is hybridized with the photonic mode in the dielectric rod, providing a new mode of small mode volume and long propagation length. Over half of the optical energy is concentrated in the nanometer sized dielectric gap between the waveguide and the metal layer. Adapted from Ref. 100.

Fig. 5.2: (a) Propagation length as a function of cylinder diameter, d and gap, h. The upper and lower horizontal broken lines represent the propagation lengths for the uncoupled metal/oxide and metal/semiconductor interfaces respectively (i.e. without cylinder). (b) Effective mode area, $A_m / A_0$ as a function of cylinder diameter, d and gap width, h. The cusp features occur when the position of maximum $W(r)$ shifts from the center of the cylinder to its edge nearest the metal. Lower panels show electromagnetic energy density distributions for $c$, $[d, h] = [400, 100]$ nm; d, [d,
Fig. 5.3: (a) Normalized energy density along the y-axis, showing the confinement in the low index dielectric region (no shading). The broken line in the left inset is x=0. The metal and semiconductor cylinder regions are shaded gray and green respectively. (b) Normalized energy density along the x-axis between metal and host dielectric. The broken line in the right inset is y = -d/2. The left inset shows the FWHM of the energy density distribution in the x-direction and the normalized mode area, $A_m / A_0$, as a function of h. Furthermore, the near linear increase in gap energy density (reciprocal of $A_m / A_0$) with decreasing h is an indicative of capacitive energy storage of this geometry (inset b).

Fig. 5.4: Schematic of HPP mode size mapping via apureless NSOM. The HPP strips consist of a semiconductor (high-k) separated from a metallic surface by a nanometer scale low-k gap. They are excited by illuminating the slit at the input end. Adapted from Ref. 100.

Fig. 5.5: 3D image overlap of the deep sub-wavelength HPP mode signal (red spot) offering optical confinement significantly below the diffraction limit of light. Scale bar = 125 nm. MgF$_2$ gap $h = 10$ nm, illumination wavelength $l = 633$ nm. Inset: Height profile of tapered strip for free-space to HPP strip coupling, scale bar = 1 mm. Focused ion beam (FIB) etching was used to define the illumination-port (IN) and the access point for the NSOM probe of the confined optical mode (OUT).

Fig. 5.6: Line cross section of the nearfield intensity along the horizontal broken line (solid line). Red dashed line is the cross section in the numerical simulation of the same structure. Black dashed line stands for the simulated cross section at 10 nm away from the surface.

Fig. 5.7: Summary of the optical confinement along the horizontal and vertical direction in the hybrid waveguides of different widths.

Fig. 5.8: Increasing the wavelength of the illumination beam to 808 nm and 1427 nm results in mode height FWHM of 63 nm and 74 nm respectively, featuring broadband, deep sub-wavelength operation of HPP-based devices. Solid lines are Gaussian fits.

Fig. 5.9: (a) Far field images of illuminated HPP strips showing incident white light reflected from the strip input and the corresponding emission from the distal strip output facet. The strips lengths, $L$, are indicated on top of each image and the dashed line is a guide the eye to the output signal. (b) The HPP mode offers propagation lengths of more than 10 times its free space wavelength at $\lambda = 633$ nm. Without the low-k gap-layer (plasmonic control) the propagation length is only about 1 $\mu$m (black dashed line). Inset: the HPP mode’s propagation length increases with its operating wavelength and exceeds 20 times its vacuum wavelength at near infrared wavelength ($\lambda = 800$ nm).

Fig. 5.10: (a) Illustration of the excitation and collection setup used to characterize the transmission of HPP waveguide. (b) Group index dispersion of the HPP
waveguide retrieved from the Febry-Perot resonance in the transmission spectrum of a series of waveguides with different lengths. The red dashed line and the back broken line correspond to two extreme cases, where $h \to \infty$ and $h \to 0$, respectively.

Fig. 6.1: Atomic structure of monolayer and bulk MoS$_2$. The side view of the structure are featured in (a) & (c), while the top view is shown in (b) & (d).

Fig. 6.2: The gradual transition of the electronic band structure in MoS$_2$. The band gap becomes a direct one at K point as the material is reduced from few layers to the monolayer form. Adapted from Ref. 16.

Fig. 6.3: PL spectrum of MoS$_2$ monolayer excited by a 488 nm laser.

Fig. 6.4: Monolayer MoS$_2$ contrast on silicon oxide of different thickness on heavily doped silicon, at different wavelengths. Adapted from Ref. 119.

Fig. 6.5: White light image of monolayer MoS$_2$ on 270 nm SiO$_2$/Si substrate. A bilayer is also visible in the image, which has about twice the contrast than the monolayer. The insect shows the centimeter scale sample chip with an array of alignment markers.

Fig. 6.6: A collage of PL (red) & scattering (green) map of a 1cmx1cm sample. Each picture is about 200x200 µm large and has 256x256 pixels. Therefore a submicron resolution can be achieved.

Fig. 6.7: (a) literature result of the dependence of $E_{2g}$ & $A_{1g}$ splitting and the layer thickness. (b) Experimental results measured in our lab on three different layers on a SiO2/Si substrate. (c) AFM image and cross section of the three-layer sample.

Fig. 6.8: SEM image of the monolayer and bilayer of the MoS$_2$.

Fig. 6.9: Optical image of a MoS$_2$ Hall device after EBL lithography patterning and development.

Fig. 6.10: AFM image of a Hall device after the metal deposition and lift-off process.

Fig. 6.11: (a) Optical image of a Hall bar device after the etching. White lines highlight the In/Au contacts. Green lines are the external boundary of the etching windows. (b) Only the PL from MoS$_2$ channel are left after the etching.

Fig. 7.1: First principle calculation of the linear absorption spectrum of the monolayer MoS$_2$ with (green), and without (red) considering the exciton effect. Black dashed line marks the emission energy. (a) Zero temperature result. (b) Finite temperature result with a large broadening. Adapted from Ref. 117.

Fig. 7.2: Illustration of the two-photon absorption and luminescence process after the relaxation. Blue arrows represent the excitation process while the red arrow means the emission process. The green lines are the excited exciton states and the red line is the ground exciton state.
Fig. 7.3: Linear transmission spectrum of monolayer WS$_2$ on quartz substrate.

Fig. 7.4: (a) An emission spectrum of WS$_2$. The excitation laser wavelength is 1000 nm. Pulse duration is about 200 fs. The power level is about 10 kW/cm$^2$. The spectrum is taken at 10K. (b) Transmissivity of the microscopy system. The strong dispersion is mainly caused by the high NA objective.

Fig. 7.5: Power dependence of SHG (red) and TPL (blue) signals. The SHG signal has a good quadratic power dependence while the TPL signal has a clear saturation behavior at the high power level.

Fig. 7.6: 2D emission-excitation plot with the normalized excitation intensity. The SHG sloped line is saturated in order to achieve an optimal contrast of the TPL signals.

Fig. 7.7: Overlap plot of one-photon transmission (green) and two-photon absorption (blue) spectrum. Arrows are pointed at the emission energy levels and short bars highlight the TPL peak positions.

Fig. 7.8: Two-photon absorption spectrum with different top capping layer. The experiments are operated at the room temperature.

Fig. 7.9: DFT calculation with GW approximation. Dashed line is the result including BSE and the solid line is without BSE approximation. The difference between them is caused by the exciton effect. The red arrows are pointed at 1s and 2s bright excitons and the green arrows are pointed at 2p and 3p ark excitons. The band edge position is highlighted by a yellow arrow.

Fig. 7.10: Calculated 1s exciton wavefunction both in real space and k space. The real space wavefunction has an in-plane extension of about 1 nm in radius. The wavefunction is highly localized at K points.

Fig. 7.11: White light image (a), confocal luminescence mapping (b) and SHG signal mapping (c) of a sample with mono-, bi-, and trilayer areas.

Fig. 7.12: (a) Polarization dependence of the SHG intensity. The incident and detection polarization are the same, and labeled in the fig. (b). SHG is dependent on the angle between the polarization axis and the crystal orientation. Without phase sensitivity, the intensity pattern has a six fold symmetry.

Fig. 7.13: White light image (a), total SHG signal (b) and y-axis polarized SHG signal (c) of a continuous CVD grown MoS$_2$ monolayer. (d) Illustrations of the crystal orientation distribution of A and B areas to demonstrate the polycrystalline nature of the CVD film.

Fig. 7.14: (a) SHG image pumped at 1280 nm. (b) SHG image of the same sample pumped at 1300 nm. The brim of crystal shows a strong nonlinear optical edge states. The scale bare is 20 µm. (c) Cross-sections (white dashed lines shown in (a)
and (b)) compare the SHG of the same crystal under the pump wavelength of 1280 nm and 1300 nm, respectively.

Fig. 7.15: Nonlinear edge state resonance of MoS$_2$ monolayer membrane. (a – f), SHG images of the same monolayer MoS$_2$ sample for the fundamental wavelengths of 1280 nm, 1290 nm, 1300 nm, 1310 nm, 1320 nm, and 1330 nm, respectively. The scale bar is 20 µm. (g) Resonance of the edge nonlinear response at ~ 0.947 eV (1310 nm). The ratio is determined by the integrated SHG with the areas circled in (a).

Fig. 8.1: Band alignment of between p-type silicon and n-type MoS$_2$. (a) The relative band positions when they are separate. (b) When the two materials get contacted and reach an equilibrium state as a PN junction. (c) A forward bias is applied, inducing a current and consequent emission in MoS$_2$. The blue dots represent the holes in silicon and the green dots represent the electron in MoS$_2$.

Fig. 8.2: Processes of site-controlled MoS$_2$ transfer method. The dashed circle highlights the monolayer MoS$_2$ flake across the Si/SiO$_2$ interface. Scale bar: 20µm.

Fig. 8.3: IV curve of the PN junction shows an ideal diode behavior.

Fig. 8.4: The experimental setup measures both electroluminescence and photoluminescence at cryogenic temperature with a laser scanning capability.

Fig. 8.5: (a) Schematic of the PN junction and the emission hot spot. Pink color represents a 300-nm-thick silicon layer while the purple layer is a SiO$_2$ layer and In/Au contacts are drawn in yellow. (b) Experimental observed emission pattern. White dashed lines outline the monolayer boundary and the grey dashed line is the silicon step edge.

Fig. 8.6: (a) Scattering mapping and (b) photocurrent mapping of the device. The yellow and white dash lines define the In/Au electrode and monolayer MoS$_2$ boundaries respectively. Scalebar: 5 µm.

Fig. 8.7: Electroluminescence and photoluminescence spectra from the same device. The electroluminescence spectrum is measured at the current level of 42 µA. Besides the strong direct exciton emission (labeled AX), a weaker satellite peak at a lower energy is observed (labeled DX).

Fig. 8.8: (a) Electroluminescence spectra of a MoS$_2$ diode recorded at different currents from 17 to 86 µA. The data can be fitted well with two Lorentz contributions, which are attributed to A exciton (labeled AX) and bound exciton (labeled DX) emission. (b) Red dots: the A exciton emission (AX) shows an approximately linear increase with current. Green squares: bound exciton emission (DX). The electroluminescence saturates as the current exceeds ~65 µA. The dashed lines are fitting results. (c) Electroluminescence spectrum at a current of 86 µA. A new exciton (labeled NX) feature emerges with a higher energy of 2.26 eV. Red curves are the Lorentz fitting result.
Fig. 8.9: Low temperature electroluminescence spectra and electrical power dependence of the MoS$_2$ diode shown in Fig. 8, recorded at different current levels from 13 to 82 µA. The experimental temperature is 10K.

Fig. 9.1: Band structure of MoS$_2$. Electron at different k points have different valley indexes, indicating different effective spins and bit information.

Fig. 9.2: (a) Polarization resolved photoluminescence from MoS$_2$ at 10K. (b) The helicity is retrieved based on fig. 9.1 (a). (c) The temperature dependence of the helicity ranging from 10K to room temperature.

Fig 9.3: Schematic of the valley Hall device. Different circular polarized photon could excite the excitons in different valleys. Berry curvature together with the longitudinal electric field will generate the opposite anomalous velocities along the transverse direction, thus building up a transverse voltage. A back gate is used to tune the carrier density inside the channel.

Fig. 9.4: SEM images of a fabricated sample. (a) Low magnification image showing the electrode lines and the contacting pads. (b) High magnification image showing the drain/source contacts and the Hall contacts. Scale bar is 2µm.

Fig. 9.5: (a) IV relationships between different pairs of contacts showing a good Ohmic contact. (b) The transfer curve of the device. Red line: $V_g$ is swept from negative to positive. Black line: $V_g$ is swept from positive to negative. Sweeping speed is 10 V/s.

Fig. 9.6: IV curves measured in two-probe (a) and four-probe (b) configurations from a parallel-probe device at different back-gate levels.

Fig. 9.7: Schematic of the scanning photo current microscope. An EO modulator is used as a dynamic half wave plate to modulate the polarization of the laser beam. Depending on experiment design, a polarizer or a quarter waveplate is placed before the objective.

Fig. 9.8: Photocurrent mapping of an edge of the silicon detection. Cross section along the dashed line indicates the focus spot is 1.7 µm in diameter.

Fig. 9.9: A long time log of the slow photocurrent response in MoS$_2$ detector. The bottom figure shows the treatment to the sample. Red bars stand for the laser treatment in the channel, with the height representing the laser power. Yellow and blue bars mean the heating and air-venting action. The top plot shows the dynamic of the current level in correspondence to different treatments.

Fig. 9.10: A water ion absorption/desorption model can explain the slow photo response. After the exciton is generated by incident photon, the ion becomes neutralized and desorbed from the transistor channel.
Fig. 9.11: An overlap image of the photocurrent mapping and the electrode positions. Scale bar: 2μm.

Fig. 9.12: (a) Scattering mapping. (b) Photocurrent mapping at zero source-drain bias. (c) Schematic of the photothermal electric potential distribution across the device.

Fig. 9.13: Band alignment between MoS₂ and Titanium contact with an external voltage bias applied. The built-in electric field at the high potential end results in a higher photocurrent enhancement than the low potential one.

Fig. 9.14: (a) Photovoltage signal mapping of an Indium contact device. (b) Scattering image is collected at the same time.

Fig. 9.15: (a) Amplitude of the photovoltage as a function of $I_{ds}$ (x axis), and back gate voltage $V_g$ (y axis). (b) The phase of the photovoltage relative to the EO modulation phase. A $\pi$ phase change, equivalent to the voltage sign change, occurs when the applied current is reversed.

Fig. 9.16: (a) Electrical bridge model for explaining the measured transverse photovoltage pattern. Both the relative phase (b) and the amplitude (c) of the photovoltage show a four-fold symmetry.

Fig. 9.17: (a) Time sequence schematic plot of the intensity modulation by an EO modulation together with a polarizer. Below is shown the corresponding periodic high voltage applied to the modulator. (b) An ideal Hall measurement modulation sequence should be a pure periodic helicity modulation. (c) Nevertheless, EO modulator will cause an unbalance pulse train due to the different transmissivity for orthogonally polarized light.

Fig. 9.18: The drift of the difference signal of EO modulation transmissivity for orthogonal polarized light. The stable period is about 4 hours long after 6 hours’ oscillation.

Fig. 9.19: (a) Transverse photovoltage as a function of the retardance from the quarter waveplate with a positive $V_{ds}$. (b) The curve is flipped when the applied voltage is reversed. (c) Longitudinal photovoltage showing a strong $2\omega$ dependence.
List of table

Tab. 9.1: Summary of the 2P, 4P and contact resistance at different back-gate levels.
Acknowledgement

My past five years of Ph.D. study at Berkeley has been a fruitful and joyful journey. I am indebted mostly to my advisor, Professor Xiang Zhang, who brought me to this wonderland and endowed me all the passion and resources for research from the very beginning. Without his optimistic full support throughout the journey, I couldn’t have achieved even a corner of the works presented in this dissertation. His non-stop enthusiasm, crave for novelty, and constant strictness on details have all been molded into my soul.

I am more than thankful to all my four past academic mentors in XLab: Professor Shuang Zhang, Professor Thomas Zentgraf, Professor Xiaobo Yin, and Dr. Yuan Wang. From setting up the first lens to writing up the first draft, their patient guidance was the biggest treasure to a novice in the research world. As the leading and mentoring role was passed to me, I also really enjoyed working with the new ’Moly’ subgroup and collaborating with competent postdocs as Majid Gharghi, Yu Ye, Yimin Li, Kedar Hippalgaonkar, and talented students as Hanyu Zhu, Mervin Zhao, Jun Xiao, Ying Wang, Quanwei Li, Yang Xia.

XLab is a big family full of love. Everyone in the lab is a family member to me. They include past and current postdocs: Jie Zhu, Guy Bartal, Rupert Oulton, Bin Guo, Avi Niv, Hu Cang, Atsushi Ishikawa, Jia Zhu, Changgui Lu, Jenson Li, Chih-Wei Chang, Tae-Jin Yim, Li Zeng, Changshi Lao, Ajithkumar Warrier, Xiaodong Yang, Sheng Shen, Pavel Kolchin, Stefano Palomba, Maiken Mikkelsen, Daniel Lanzillotti Kimura, Boubacar Kante, Renmin Ma, Peng Zhang, Tongcang Li, Haim Suchowski, Alessandro Salandrino, Xuefeng Zhu, Pankaj Jha, Chihhui Wu, Liang Feng, Xingjie Ni, Aimin Wu, Baolei Li, Dominc Zerulla; my peer students: Junsuk Rho, Kevin O’Brien, Sadao Ota, Ania Labno, Christopher Gradden, Ong Pholchai, Zeev Abrams, Yongshik Park, Volker Sorger, Ming Liu, Lee Fok, Jie Yao, Jason Valentine, Liang Pan, Sheng Wang, Yongmin Liu, Yi Xiong, Zi Jing Wong, Shaomin Xiong, Sui Yang, Michael Mrejen, David Barth, Jeongmin Kim, Taiki Hatakeyama, and our dear administrator Nga Thien Bui. In addition, I am particularly thankful to our group alumni, Professor Zhaowei Liu in UC San Diego and Professor Nicholas Fang in MIT, for their supports throughout my study.

I am grateful to Professor Feng Wang for serving in my qualify/thesis committee and his kindness and readiness to address my question at any time in the past five years. Many thanks also go to his group members, Long Ju, Sufeii Shi, Kaihui Liu, Jonghwan Kim, Xiaoping Hong, Jason Horng. They have introduced me the basics of two dimensional materials and valleytronics.

Professor David Attwood in Electrical Engineering department, the chair of AS&T program, has spent numerous time with me in arranging my prelim, qualify exam, career plan, and the department colloquium. His kindness and patience are deeply appreciated in my heart. I would like to thank Professor Jie Yao, Professor Steve
Leone, Professor Ivan Kaminow, Professor Hartmut Häffner and Professor Junqiao Wu for kindly serving in my prelim, qualify, and thesis committee. Outside the program, I want to thank Professor Ziqiang Qiu, Professor Steve G. Louie, Professor James Analytis and Professor Peter Yu, in Physics department, for all the inspiring discussions. The staff of the machine shop in Mechanical engineering department and the device characterization lab in Electrical engineering department have been very helpful to me over the past years, and I appreciate very much their support as well.

My final thanks go to my undergraduate advisor, Professor Xiaodong Cui in Hong Kong University. His encouragement and support have benefited me till now.
Chapter 1.

Introduction

1.1 Introduction and motivation

If not detectable within the normal vision range, should the observed phenomena be accepted as a reality? At the advent of modern science, this used to be an important philosophical question. Today we have all accepted that the scientific proof can be verified by indirect measurements. Nevertheless, the light-matter interaction, an extension of our vision, has remained one of the most important probes for exploring the novel phenomena. A recent negative example is the ‘dark’ matter: about 85% of the matter in the universe does not interact with the radiation via the electromagnetic force and are named dark matter. They are therefore extremely hard to detect and turn out to be an important topic in modern cosmology.

Now let’s zoom into the other extreme. On the scale of a few nanometers, we can see a similar urgent need to investigate the light-matter interaction. As predicted by Moore’s law, the number of transistors in a processor has been doubling every year. The minimum feature size in the state-of-art industry standard has been reduced to 22 nm. Such an integration trend is strongly motivated by the pursuit of higher computation speed while maintaining the total energy consumption level. As a result, optics for nanoscale objects, or nano-optics, has become very critical to the emergent nanotechnology, ranging from fabrication to characterization, from computation to communication. One example is the rise of plasmonics, where an optical surface wave at metal-dielectric interface has been proposed to communicate the information across the transistor, in the place of the old-fashioned transmission line. It is a nano version of optical fibers used in the telecommunication industry and much larger bandwidth can be provided theoretically.

The fundamental challenge in nano-optics is that most of the novel effects cannot be detected in the far field. They are hereby named optical ‘dark’ effects. One reason for the darkness is that most of the feature lengths in nano-optics are much smaller than the photon wavelength in free space. Such a large mismatch blocks the crosstalk between them. Consequently, in nano-optics, the optical energy is highly confined in the nearfield, a type of rim zone of the matter, and cannot radiate to free space. Another reason is that in order to achieve versatile functions, many nano-structures are designed to acquire very dramatic optical phase contrasts at the subwavelength level. As a result, the conventional dipole approximation is no longer valid. For example, in one of our studied coupled antenna systems, called plasmon interference transparency, the dipole moment is completely cancelled out and the remaining higher order poles cannot receive or emit farfield photon efficiently.

In order to access these dark information, our method is to use a scanning optical microscopy to approach the nearfield zone of the matter. By monitoring the dynamics of the nearfield distribution at different excitation conditions, we studied
the coupling effects in a range of artificial nano-structures, from plasmonic antennas\textsuperscript{11} to hybrid waveguides.\textsuperscript{12} Besides, these new understandings are applied to achieve even more exotic functionalities, such as the giant photonic spin Hall effect at the metasurface.\textsuperscript{13}

Interestingly, these dark effects are not just limited to the artificial nano-structures. In the natural nano-crystals, there also exist some optical transitions beyond the dipole approximation, appearing as dark transitions. For example, the exciton can have excited dark levels, depending on the its overall envelop wavefunction.\textsuperscript{14} These dark levels can provide lots of information about the material.

Hard as they are for the detection in the farfield by linear probes, nonlinear optical probes are very useful tools for such studies since they share different selection rules and different contrasts with the linear probes.\textsuperscript{15} Using the nonlinear probe, we investigate these dark excitons in 2D transition metal dichalcogenide (TMDC), a new type of low dimensional semiconductor whose electronic structure becomes optically interesting only after the quantum confinement.\textsuperscript{16,17} A very tightly bond exciton is found in it. In addition, we find the nonlinear properties are strongly affected by the oddness of the layer number, the crystal orientation and the emergent states at the edges. All of them are not sensitive to the linear probes. With the new knowledge, we have built a new TMDC based light emitting device\textsuperscript{18} and gone to great lengths to demonstrate the first valley sensitive detector with TMDC monolayer.

1.2 Outline of thesis

This dissertation is organized into two sections. The first section (Chapter 2-5) is devoted to the nearfield investigation on the artificial nano-structures. In Chapter 2, we will start with introducing the background and working mechanism of the apertureless nearfield scanning optical microscope, the pillar of the nearfield imaging. The apertureless scanning optical microscope we developed can achieve a spatial resolution of 15 nm over a very broad excitation range. In Chapter 3, we will demonstrate two well-designed plasmonic antenna structures, and show how the nearfield imaging technique can help us understand the coupling effects in them. Following that, we will show an application of this understanding in Chapter 4. The coupled antennas are employed to generate the photonic spin Hall effect on metasurface in an unprecedented regime. Chapter 5 is focused on nearfield-imaging the functionality of a hybridized plasmonic-dielectric waveguide. The mode size inside this waveguide can be confined to 60 nm in one direction.

The second section (Chapter 6-9) is on our effort to study and engineer 2D TMDC as an emergent optical material. Chapter 6 will introduce the background of this material, including the preparation, characterization and fabrication techniques. Chapter 7 is on how we can use the nonlinear optical probes to study a series of optically dark effects, such as the dark exciton, the binding energy, the crystal orientation and the resonance states at the edge. A TMDC based light emitting
device is discussed in Chapter 8 as a practical application. Finally, in Chapter 9, we will discuss our effort towards achieving the first electrical detector for valley polarization, a new degree of freedom promising for future quantum computing applications.
Chapter 2.

Apertureless nearfield scanning optical microscope

Nearfield scanning optical microscopy (NSOM) is a type of scanning probe microscope detecting the local optical property near the surface. \(^{19-21}\) NSOM can provide an optical resolution lower than the diffraction limit, because it accesses the evanescent wave in the near field. This is particularly useful for the research of plasmonics, where the typical 'hotspot' feature size is much smaller than the wavelength. \(^{22}\) During my PhD, I constructed an apertureless-NSOM (a-NSOM) with an optical resolution of 20nm in the visible range, and it has a broad working bandwidth, ranging from visible to the near infrared, which is only limited by the availability of the light source and photo detector. \(^{23}\) Furthermore, we introduce a technique called pseudoheterodyne detection into the system, which can effectively suppress the background signal and provide us the phase information of near field.

2.1 Diffraction limit

Even a perfect conventional lens based optical imaging system has a resolution limit about 200nm, which is about half the wavelength of the light. \(^{24}\) This can be understood qualitatively by Heisenberg’s uncertainty principle, which says the product of the uncertainty in the spatial position and the uncertainty of the momentum in the same direction has to be larger than the Plank constant. It can be written in this way:

\[
\Delta(hk_x) \cdot \Delta x \geq \hbar / 2
\]  
(2.1)

Since we are working with photons, the largest momentum spread after the focusing is \(k = 2\pi/\lambda\), which leads to

\[
\Delta x \geq \lambda / 4\pi
\]  
(2.2)

which result means the smallest feature we can resolve depends on the wavelength of the photon. In the real system, the numerical aperture of the lens needs also be taken into consideration. A more quantitative analysis is based on the angular spectrum representation.

The angular spectrum representation is a mathematical tool to describe the optical field in a homogeneous media. Any optical field in this representation is composed of a series of propagating plane waves and evanescent waves. The bandwidth of the angular momentum finally limits the resolution in the far field. For example, given the electric field at certain spatial position \(E(r)\), we can Fourier transform it into momentum space. \(^{25}\)
\[ E(k_x, k_y; z) = \frac{1}{4\pi^2} \iint E(x, y, z)e^{-i(k_x x + k_y y)} \, dx \, dy \]  

(2.3)

where \(x, y\) are the Cartesian transverse coordinates and \(k_x, k_y\) are the spatial frequency. \(Z\) is defined as the axis transverse to the electric field. Reversely, the inverse Fourier transform reads as

\[ E(x, y; z) = \iint E(k_x, k_y, z)e^{i(k_x x + k_y y)} \, dk_x \, dk_y \]  

(2.4)

In a homogeneous medium and source free zone, the vector Helmholtz equation is reduced into the following form

\[ (\nabla + k^2)E(r) = 0 \]  

(2.5)

If we insert eq2.3 into 2.5 and define \(k_z = \sqrt{k^2 - k_x^2 - k_y^2}\), we will get the solution to the Helmholtz equation as

\[ E(k_x, k_y; z) = E(k_x, k_y; 0)e^{i k_z z} \]  

which means the each k component in the optical field will propagate without inferring each other. Then at position \(z\), the total field can be expressed as

\[ E(x, y, z) = \iint E(k_x, k_y, 0)e^{i(k_x x + k_y y + i k_z z)} \, dk_x \, dk_y \]  

(2.6)

As we can see, when the transverse spatial frequency exceeds the maximum frequency supported by the medium \(k\), the Fourier component will decay exponentially. We define these components as the evanescent waves. The difference can be seen more clearly with the illustration in Fig. 2.1.

---

**Fig. 2.1:** Constant frequency contour in an isotropic medium, the \(k\) circle is the boundary between the propagating zone and evanescent zone.

As a result, the optical wave cannot carry the information below the diffraction limit to the far field. In order to retrieve these lost information, we need to go to the near field, whose spatial decaying constant is determined by \(\delta = 1 / k_z\), and this is the basic principle of the a-NSOM developed in our lab.
2.2 Principle and configuration of a-NSOM

Our a-NSOM system is developed based on a commercial AFM scanner (Veeco, Bioscope). As the AFM tip scan over the sample, the sample is illuminated from the bottom and the local evanescent field is perturbed by the tip and scattered into the propagating field. These lights are guided into detector and recorded as the signals in the computer, in one to one correspondence to the local field strength at the scanning point. A schematic is show as Fig. 2.2

![Schematic of the a-NSOM configuration. A silver-gallium alloy needle tip is used in the plasmonic attena array experiment, to be discussed in details in Chapter 3.](image)

The AFM scanner is consisted of a 3 piezoelectric scanner, capable of positioning the probe with sub-nanometer precision in x, y and z directions. An additional piezo stack on the AFM probe holder drives the cantilevers of AFM probes in the tapping mode, which oscillates in the z-direction at a resonant frequency, about 300 kHz. Operated in the tapping mode, the vibration amplitude of the cantilever is measured by laser reflection angle from the scanner head. When the force from the sample overdamps the oscillating tip, the tip is elevated from the surface. During the scan, a fast PID circuit keeps a constant tip-sample interaction, equivalently constant tip-sample separation, which is about 10-50nm in the usual experiment. Depending the on local field strength, both silicon tip and the silver gallium alloy nano needle tip have been employed in our experiments.

The a-NSOM is intrinsically a broadband system, and we can couple many different light sources into the system. As shown in the schematic, either high-power laser or supercontinuum white light source (Fianium light, SC 450) is focused onto the point of interest on the sample with a high NA objective (Nikon CFI Plan Fluor ELWD DM 40x C, N.A. 0.60, W.D. 3.7-2.7mm) to excite either localized hot spots or guided surface waves which cannot be detected in the far field. When the probe approaches to the nearfield region of the sample from the top, the light scattered from the sharp tip is collected and collimated to form a signal beam by an aspheric collection lens.
(Newfocus 5726H, N.A. 0.16 NA, F.L. 15.4mm) which has been aligned to focus exactly at the apex of a-NSOM probe from the side view. A polarizer is used to select Ez component of nearfield scattered from the tip. Different APDs were employed to optimize the detection efficiency at the operational frequency.

2.3 High harmonic demodulation and pseudoheterodyne detection

An intrinsic challenge facing the a-NSOM technique is the large background signal. Because the tip is a non-resonant scatter, not only the apex of the tip, but also the body of the macroscopic tip together with the cantilever will scatter a lot of light, which contains no nearfield information. The ratio between signal and background can be estimated as the following. If the excitation light is focused into a diffraction limited spot, with the diameter around 500 nm, and we want to achieve a resolution about 20 nm, which means we only want to pick out the signal from a 20 nm-diameter spot. The signal-background ratio is about $(25)^2=625$. The final signal-background ratio at least has to be two to one, which means a minimal suppression ratio over one thousand is needed.

This is achieved through a nearfield modulation and farfield demodulation process. If we assume the polarizability of the tip is constant, then the scattered electric field is $E = (E_i + E_r) \cdot \alpha$, proportional to the total field at that position. We choose to work in the tapping mode, it means we oscillate the tip of AFM vertically at several hundred kilohertz, $\Omega$, then the observed signal is fed back to lock-in amplifier and locked at higher harmonics ($n\Omega$). On one hand, the propagating wave is smooth and relative uniform throughout the oscillation space, so the tapping cannot introduce any modulation in the scattered plane wave signal. On the other hand, the evanescent wave decay exponentially along z direction, which implies the scattered evanescent wave will oscillate dramatically in time. This contrast helps us to pick out the signal from the large plane wave background. And such a contrast gain will go exponentially as the harmonic order increases. In our system, two function generators (FG1 & FG2, Stanford Research Systems, DS345) are synchronized with the same 10 MHz time base. FG1 controls the vibration frequency $\Omega$ and driving amplitude of the a-NSOM probe, while FG2 generates the frequency $n\Omega$ for the reference of the lock-in amplifier (Stanford Research Systems, SR844). After the demodulation process, a near-field image of the sample can be reconstructed together with the topography mapping from the AFM system. A typical near field Z scan is show in the Fig. 2.3. When the tip approached the surface, the demodulated signal increases exponentially, indicating a pure nearfield contribution.

![Fig. 2.3: A Z dimension scan of the demodulated near field signal.](image-url)
However, high harmonic demodulation is still not enough to suppress all the background, when the signal is very weak. An interferometric technique called pseudoheterodyne detection is installed.

The mechanism can be understood theoretically first. If we analyzed the background signal more carefully, we can see most of them are caused by the scattering of propagating components from the rear part of the tip and AFM cantilever.

\[ E_s = E_N + E_B = E_N^0 e^{i\phi} e^{-(a_i \cos \Omega t + Z_0) / \delta} + E_B^0 e^{i\phi_B + \delta / \Omega t} \]

Since normal detectors use quadratic detection, which is only sensitive to intensity of light, background will inevitably appear in the high harmonics. Of course, the contrast between nearfield and background signal will be improved if we increasingly raise the order of harmonics, but that will also sacrifice our absolute signal strength exponentially. Alternatively we can add another reference beam to interfere with collected scattered light, and modulate its relative phase in a sinusoidal way at a frequency, \( \Delta \), by a mirror mounted on a piezostage. Then the total electric field can be written in such a way:

\[ E_i = E_R + E_s = E_R + E_N + E_B \]
\[ = E_R^0 e^{i\phi + \delta / \Omega t + \phi^*} + E_N^0 e^{i\phi} e^{-(a_i \cos \Omega t + Z_0) / \delta} + E_B^0 e^{i\phi_B + \delta / \Omega t} \]

Where \( \gamma \) is the modulation depth of the mirror normalized to the wavelength, \( 4\pi A_m / \lambda \). Accordingly, the intensity will be

\[ I = (E_R + E_N + E_B) \cdot (E_R + E_N + E_B)^* \]
\[ = E_R^0 e^{2i\phi} + E_N^0 e^{2i\phi} + E_B^0 e^{-2i(a_i \cos \Omega t + Z_0)} + E_B^0 E_N^0 e^{-(a_i \cos \Omega t + Z_0)} \cos(\Omega t + \phi_R + \gamma \cos \Delta t) \]
\[ + E_B^0 E_N^0 e^{-(a_i \cos \Omega t + Z_0)} \cos(\Omega t) + E_R^0 E_N^0 \cos(-\Omega t + \phi_R + \gamma \cos \Delta t) + E_R^0 E_N^0 e^{-(a_i \cos \Omega t + Z_0)} \cos(\phi_R + \gamma \cos \Delta t) \]

It’s obvious that the Fourier transform of \( e^{-(a_i \cos \Omega t + Z_0)} \) will generate high harmonics of \( \Omega \) and the Fourier transform of \( \cos(-\phi_R - \gamma \cos(\Delta t)) \) will produce high harmonics of frequency \( \Delta \). The first two terms are time independent and they can be directly filtered out with a high pass filter. The third term denotes the pure nearfield signal; however it cannot be measured directly, because of the following term. The fourth term represents the interference between the background and nearfield even without any reference beam, and it shares the same exponential decay character as the 3rd term. The 5th term is the interference between background and reference beam, which only has the Fourier components at the frequency \( \Omega + m \Delta \) (m = 0, \pm 1, \pm 2...). The last term represents the interference between the nearfield and reference beam, which contributes at the frequency \( \Omega + m \Delta \) (m = 0, \pm 1, \pm 2...) and more importantly at the higher harmonic frequencies and their side bands n \( \Omega + m \)
Δ (n > 1, m = 0, ± 1, ± 2...), shown as the Figure 2.4. Therefore, by measuring the
demodulated signal at the high harmonic’s side band, we can get a background free
nearfield signal.

\[ \Delta(n > 1, m = 0, \pm 1, \pm 2...) \]

Figure 2.4: Simulated high harmonic signals together with the side band signals produced
by the phase modulation of the reference beam.

Besides, such a detection method can provide additional phase information of
nearfield according to the equation below:

\[ E_R = E_R^0 \exp(i\gamma \cos(\Delta t) + i\phi_R) = \sum_m E_m \exp(i m \Delta t) \]

where \( E_m = E_R^0 J_m(\gamma) \exp(i\phi_R + \frac{m \pi}{2}) \)

so \( I_{nm} = (E_R^0 J_m(\gamma) E_N^0)^2 \cos(\phi_R + \frac{m \pi}{2}) = c \cdot I_N \cos(\phi_R + \frac{m \pi}{2}) \)

If correct \( \gamma \) is chosen to ensure \( J_1(\gamma) = J_2(\gamma) \), we can retrieve the absolute
amplitude and phase of the pure nearfield by simultaneously measuring the first
and second side band of the same harmonic signal. \( \gamma \) value can be reached by
continuously adjusting the piezo mirror oscillation amplitude until the first and
second side band signal become the same in the calibration sample and the
subsequent phase and amplitude can be synthesized according to the following
equations.

\[ \text{Amplitude}(x, y) = \sqrt{I(n\Omega + \Delta)^2 + I(n\Omega + 2\Delta)^2} ; \text{Phase}(x, y) = \arctan\left( \frac{I(n\Omega + \Delta)}{I(n\Omega + 2\Delta)} \right) \]  \hspace{1cm} (2.9)

Figure 2.5: Schematic of a-NSOM with
the interferometer section.
The experimental configuration of the setup implementing the pseudoheterodyne detection is illustrated in the figure 2.5. In addition to the figure 2.2, a reference beam is first generated at a pellicle beam splitter (BS1 in Fig. 5.1, Thorlab BP108, 8:92 (R:T) split ratio), and travel a delay line to minimize the difference between the signal and reference beams. A reference mirror (PM) is continuously vibrated at a frequency of $\Delta$ with a piezo stack actuator. The back scattered light from the tip is joined with the reference beam at the second pellicle beam splitter (BS2) and the combined beam is sent to the APD detector. Finally, the demodulation of the detector output is necessary to reconstruct high-resolution optical images of the sample. Three function generators (FG1, FG2 and FG3 in Fig 5.1, Stanford Research Systems, DS345) are synchronized with the same 10 MHz time base. FG1 controls the s-NSOM probe vibration frequency $\Omega$ and driving amplitude, FG3 handles the dithering frequency $\Delta$ of the reference mirror (PM), and FG2 determines the lock-in frequency $n \Omega + m \Delta$.

Fig. 2.6: (a) Phase of the propagating of the surface plasma. (b) Amplitude of the field. Scale bar: 2$\mu$m.

The phase information is important for imaging the propagating surface plasma. An example is show in the figure 2.6. A 633 nm surface plasma wave at the interface between silver and air is excited by a curved slit at the middle of the image, and propagate towards the top right and bottom left direction. Fig. 2.6 (b) shows the amplitude of the surface plasma where nothing but an exponential decay from the slit can be observed. However in Fig. 2.6 (a), the phase accumulation during the wave propagation is clearly seen. The difference wave vector measured is a beating result between the surface wave and the background scattering.
Chapter 3

Nearfield coupling in the plasmonic antenna array

Plasmonic antenna is a resonating optical device that efficiently converts the free propagating optical radiation to localized plasmonic hotspots. In this chapter, I will first introduce the plasmonic antenna theory and discuss the traditional coupling between the antennas. Then I will focus on some novel coupling phenomena between these mesoscopic structures, resulted from the fine tuning of the radiative continuum coupling channels. I will also discuss their experimental demonstrations with the home-built a-NSOM introduced in Chapter 2. Finally I will conclude the chapter with an outlook for the possible applications.

3.1 Introduction to the plasmonic antenna

![Fig. 3.1: Plasmonic antenna bridge the gap between the free space propagation and quantum emitter (a) as well as receiver (b). Adapted from Ref. 27.](image)

Plasmonic antenna, or optical antenna, is a nano version of the classical antenna traditionally exploited in the microwave regime. As shown in the fig. 3.1, it serves as a bridge between the diffraction limit focus spot, which is around hundreds of nanometers due to the diffraction limit (Chapter 2.1), and the quantum emitter/ receiver, which usually has a size of nanometer. The gap in between has been well known as a major barrier for any efficient emission and absorption process. This can be illustrated by considering a time-harmonic point dipole, oscillating at the frequency $\omega$. If we assume the dipole P has a small size $\Delta l$, then the effective oscillating current has the value of $I = P \cdot \omega / \Delta l$. As a result the total radiation power of this dipole can be rewritten in the following form:

$$ P_{rad} = \frac{\pi}{3} I^2 Z_w \left(\frac{\Delta l}{\lambda}\right)^2 $$

(2.1)

$Z_w$ is the free space impedance $\sqrt{\mu_0 / \varepsilon_0} = 377 \Omega$. Accordingly, the radiation resistance can be defined as $P_{rad} / I^2$. Apparently, the larger mismatch between the wavelength and the dipole length, the lower radiation efficiency it will have.
The resonance condition of the classical microwave antenna can be understood in the figure 3.2 (a). Then the wavelength of the emission wave is equal to twice of the antenna length, we reach the first standing wave resonance condition. The current in the whole antenna radiates as a large harmonic dipole. In the language of the antenna convention, the reactance disappearance and only the radiative resistance is left at this moment. This condition can be met every time the antenna length is equal to the odd order of the radiation wavelength, i.e. $1/2 \lambda, 3/2 \lambda, 5/2 \lambda$ and so on. 31

![Diagram](image)

**Fig. 3.2:** (a) Illustration of the first resonance condition of the classical antenna coupled with a transmission line. (b) Effective wavelength for the plasmonic antenna resonance in the optical range. Dotted black line, dashed blue line and solid red lines represents plasmonic antennas of different diameters. Adapted from Ref. 1.

In reality, the scaling rule above is not perfectly accurate when it is applied in the optical regime, because the penetration effect into skin depth can no longer be neglected. When the antenna’s length and diameter is comparable to the skin depth, the finite plasma frequency will result in a delay between the current response and driving electric field. But an equivalent picture can be developed that the electrons in the antenna are following the driving field of a new equivalent wavelength.

Phenomenologically, we can write down the dependence of the effective wavelength and the free space driving field’s wavelength. 32

$$\lambda_{\text{eff}} = n_1 + n_2 \frac{\lambda}{\lambda_p} \quad (2.2)$$

The parameters in the eqn. 2.2 have been numerically estimated and plotted in the fig. 3.2 (b). As can be seen, the effective wavelength, with which the gold antenna resonates at 600 nm excitation condition, is only 300 nm, which means the antenna length only needs to be as short as 150 nm.

Another important difference between plasmonic antenna and classical antenna made of perfect conductors is the kinetic inductance. 33 This is an additional effective inductance, which comes from the inertia of the electron, rather than the magnetic energy stored in the circuit.
In the optical regime, the current is mainly the polarization current, \(-i\omega P\) and 
\[ P = \varepsilon_0(\varepsilon - 1)E. \]
As a result, the current is written in the following form.

\[ J = \omega\varepsilon_0(\varepsilon'' - i(\varepsilon' - 1))E \]  \hspace{1cm} (2.3)

The imaginary part of the current stands for a current with \(\pi/2\) phase delay compared with the driving field, which is consistent with the transitional inductance definition. If we use the Drude model in the metal, the imaginary part of the current can be neglected at the frequency much lower than the collision frequency, and therefore it is usually not important in the integrated circuit. However at the working frequency of plasmonic antenna, the imaginary part is significant and cannot be neglected. It finally sets the up limit of the frequency range what plasmonic antenna can work with.

### 3.2 Coupling between antennas: coupled resonator model

When two plasmonic antennas are positioned near each other’s hotspot, they are coupled via the nearfield electromagnetic wave. One example is shown in figure 3.3. After the coupling, they are usually called the coupled-dipole antenna, or the gap antenna. The coupling strength depends on the gap width, depth as well the respective resonating frequency of both antennas.

![Fig. 3.3: The energy-level diagram and simulated near-field intensity spectra for 30-nm-high, 50-nm-wide, and 110-nm-long (arm length) symmetric two-wire gold antennas with gaps of 6 nm (black dashed line) and 16 nm (gray dash-dotted line), as well as for a single-wire antenna with the same dimensions (black solid line) Adapted from Ref. 1.](image)

The coupling effect manifests itself as the generation and splitting of new eigenmodes within the structure. As shown in the fig. 3.3 left diagram, the new modes are defined as the bonding and anti-bonding modes, in accordance with the
hybridization theory. They are labeled in such way because in the bonding mode, the current oscillation is in phase in both antennas, while it is out of phase in the anti-bonding mode. Consequently, in the bonding mode, the instantaneous charges across the gap are of opposite signs, which result in attraction force and therefore a lower level in the electrostatic energy. The case is reversed in the anti-bonding case. The instantaneous charge of the same sign gives a repulsive force and results in a increase in the eigen frequency. Evidently, the coupling induced blue and red shift is sensitive to the gap size. For a pair of gold antennas, with 30-nm height, 50-nm width, and 110-nm length, the new modes’ frequency splitting ranges from 0 to 100THz with a gap limit as small as 6nm. (fig. 3.3, right diagram)

The result of the hybridization is that the new bonding mode becomes a collective dipole, with a much larger oscillation strength, and the anti bonding mode acquire s a much smaller oscillation strength, because the two dipoles are radiating in the opposite phases and cancel each other. The elimination of the radiation channel also leads to a smaller linewidth, or a higher quality factor. The reciprocity theorem shows if it’s not radiating efficiently, it is also difficult to excite efficiently. Therefore, we also call the anti-bonding mode a dark mode.

Fig. 3.4: Two spring-ball systems are coupled by another spring. The coupling results in two new collective modes, one eigen mode of higher eigen energy is called the anti-bonding mode and the other one of lower eigen energy is called the bonding mode. Adapted from Ref. 1.

To develop a more intuitive picture on the bonding-anti-bonding modes, we analyze a classical coupled resonator model in the following, which for example describes a system of two balls coupled by a spring. (Figure 3.4)

For a harmonic oscillator driven by a periodic external force, the equation of motion is a differential equation:

\[ m \ddot{a} + \gamma \dot{a} + ka = F \]  

(2.4)

m is the mass of the oscillator, k is the spring constant, F is the driving force and \( \gamma \) is the damping coefficient. If F is a proportional to an oscillating electric field \( ge^{-i\omega t} \), we can accordingly guess oscillation strength ‘a’ also has a time harmonic form: \( a = a_0 e^{-i\omega t} \). Then the solution to eq. 2.4 is
\[ a_0 = ((\omega_0 - \omega)(\omega_0 + \omega) + i\Gamma \omega)^{-1} gE \]  
\[ \omega_0 = \sqrt{k/m} \text{ is the intrinsic resonating frequency.} \]
\[ \Gamma = \gamma / m \text{ is the line width of the resonance.} \]
Considering the near resonance regime, where \( \omega_0 \sim \omega \), and define \( \delta = \omega_0 - \omega \), eq. 2.5 can be reduced to

\[ a_0 = (\delta + i\Gamma)^{-1} gE \]  
(2.6)

Here \( g \) is just a constant. The absorption in the resonator will be proportional to \( \text{Im}(a) = \Gamma / (\delta^2 + \Gamma^2) \). We can extend this solution to coupled-dipole model:

\[
\begin{pmatrix}
    a \\
    b
\end{pmatrix} = \begin{pmatrix}
    \delta_a + i\Gamma_a & \kappa \\
    \kappa & \delta_b + i\Gamma_b
\end{pmatrix}^{-1} \begin{pmatrix}
    g_aE \\
    g_bE
\end{pmatrix}
\]  
(2.7)

\( a \) and \( b \) are the oscillation strengths of two coupled resonators. \( \delta_a, \delta_b \) are the frequency differences between each resonator’s intrinsic resonance frequency and the driving field frequency; \( \Gamma_a, \Gamma_b \) are the damping coefficients and \( g_a, g_b \) are the coupling coefficients between the driving field and the resonators. \( \kappa \) is the effective coupling constant between the resonators. After the coupling, the motion of individual resonator becomes a superposition of two collective modes, which are the bonding and the anti-bonding modes. The frequency splitting between the new modes can be acquired by solving the eigenvalues of the coupling matrix, which is

\[ \Delta = 2\sqrt{(-i\Gamma_a - i\Gamma_b - \delta_a - \delta_b)^2 + 4(\Gamma_a \Gamma_b - i\Gamma_b \delta_a - i\Gamma_a \delta_b - \delta_a \delta_b + \kappa^2)} \]  
(2.8)

Clearly, the larger the coupling constant \( \kappa \) is, the larger the frequency splitting between bonding and anti-bonding modes will be.

A simple model as it is, the coupled resonator model can generate a lot of interesting phenomena by finely tuning the linewidth, the resonating frequency, the coupling coefficients and most importantly the coupling constant in the model. In the following section, I will give out two specific examples. Chapter 3.3 is devoted to realizing a plasmonic version of EIT effect by coupling a bright antenna with a dark antenna. In chapter 3.4, we expand the coupling constant space from the real to the complex one, through which we can effectively reduce the linewidth of each individual antenna.

### 3.3 Dark transition in Plasmonic Interference Transparency (PIT)

#### 3.3.1 Introduction to Electromagnetic Induced Transparency (EIT)

Electromagnetic induced transparency is a technique for eliminating the effect of a medium on a propagating beam of electro-magnetic radiation. \(^{36}\) It was initially
proposed and demonstrated in atomic system with three energy levels. One example is shown below.

*Fig. 3.5: EIT Energy levels and allowed transition path between the levels. The control beam and the probe beam are denoted by $\omega_p$ and $\omega_c$. Adapted from Ref. 36.*

There are three energy levels in the figure 3.5, labeled as state 1, 2 and 3. State 1 is the ground state of the system. State 2 is highest excited state which has a dipole allowed transition to state 1 as well as state 3. State 3 is an intermediate state that has the same symmetry with the ground state, and therefore has no dipole transition to the state 1. Due to the distinguishing spontaneous emission rate, any population trapped in the state 3 has a much longer life time and consequently its energy level is much better defined with less radiative broadening, equivalently higher quality factor. This is one of the requirements of the EIT technique.

Another important requirement for EIT is the strong coupling between state 2 and 3. In the experiment, a very strong laser beam, labeled as control beam in fig. 3.5 and tuned to the energy difference between state 2 and 3, is incident together with the probe beam, labeled as $\omega_p$. The intensity of the control beam is much stronger than the probe beam, so that a new transition loop, $|1\rangle \rightarrow |2\rangle \rightarrow |3\rangle \rightarrow |2\rangle$, is possible and has a similar probability compared with the direct transition from 1 to 2. However, the loop transition has a $\pi$ phase shift compared with the direct transition due to the resonant transition between 2 and 3. Because the final transition probability is the sum of all the possible quantum paths, the loop transition and the direct transition cancels each other out, which means at the transparency point, no atom will be excited from the ground state 1 to the excited 2. The absorption of the probe beam in the media is completely eliminated.

*Fig. 3.6: Absorption of EIT. When a coupling laser beam is applied, the original absorption peak becomes a dip and a large group index is introduced into the medium.*
When the coupling beam is applied, the absorption spectrum of the medium will be largely modified, as shown in the figure 3.6. At the original peak position, the absorption drops to zero due to the quantum transition path interference and the original single peak is split into double peaks. At the dip position, all the excited atoms are populated in a superposition state between 1 and 3. No atom will be populated in state 2.

Because this technique can introduce a large spectral dispersion while reducing the absorption, it becomes a very effective approach to enhance the nonlinearity and slow down the light speed within the medium. Recently, analogs of the EIT system using coupled micro-size optical resonators have been proposed, which bring the original quantum phenomena into the realm of classical optics. In order to achieve the ultimate goal of a nano-sized active optical circuit with switching and modulation capabilities, we proposed and demonstrated here the plasmonic version of EIT effect, which we call PIT effect.

### 3.3.2 Analog with coupled plasmonic antenna

A plasmonic analog structure of EIT effect consists of a single plasmonic rod in close vicinity to a double rod structure as schematically illustrated in fig. 3.7. The double rod structure acts like a quadrupole antenna that possesses an anti-symmetric plasmon mode and does not interact directly with the incident planer electromagnetic waves, appearing as a ‘dark element’ consequently. A single plasmonic rod can strongly couple to the external field and therefore appears as a ‘bright element’, as a resonating dipole antenna. When they are placed in the close proximity, the quadrupole antenna can be indirectly excited through the near-field coupling from the resonating bright element. As the result, the destructive interference can occur between the two alternative excitation pathways of the dipole antenna: the direct excitation from the incident plane wave and the indirect excitation through the near-field coupling between the quadrupole and the dipole antennas (see fig. 3.7 (a)). The strength of such a mutual coupling is determined by the separation between the two resonators and the induced transparency takes place when the coupling strength is in the strong coupling regime where the strength of the indirect near-field excitations is comparable to that of the direct excitations but with a $\pi$ phase difference, $\pi/2$ phase shift from each resonator. Alternatively, the formation of the PIT can be interpreted as the result of eigen-modes hybridization of the two coupled plasmonic elements; the bonding and anti-bonding modes, which are the linear superposition of the radiative and non-radiative modes (see fig. 3.7 (b)). Excitation at the frequency centering between the two resonant frequencies receives opposite contributions from each resonance that cancel each other, thereby rendering the whole system transparent.
The above design is based on the analogy to mimic the transition pathways between ground states, excited states (bright states 2) as well as metastable states (dark states 3) in EIT atoms. The electron population at each state has its classical counterpart in classical analogues, measured by the local electric field amplitude. Therefore, the near-field visualization of bright and dark elements in PIT offers a new opportunity in understanding the underlying coherent processes that leads to PIT. The understanding of population distribution within each of states is as important as the transitions across them since a transition cannot be established without adequate population occupying given states.

In atomic processes, the population of excitations is distributed in the energy space, which requires complicated pumping schemes and ensemble averages in identically prepared atoms. In stark contrast, the major advantage of our plasmonic analogue is that the different energy levels correspond to the electromagnetic resonance of the spatially separated antenna elements (bright and dark). The equivalent energy states are therefore correlated with the spatial location of the electromagnetic hotspot. As a result, by mapping the local electric field intensity of SPPs with an Apertureless NSOM, we can visualize the equivalent population distribution in a single meta-atom level.

### 3.3.3 Experimental results: from far field to near field

In experiment, an array of PIT structures were fabricated on an Indium Titanium Oxide (ITO) coated cover glass with electron beam lithography, followed by metallization with gold and a subsequent lift-off process. Fig. 3.8 (a) shows the scanning electron microscope (SEM) image of the resulting PIT structures in an
array. The geometric parameters are: L1=275 nm, W1=54 nm, S1=85 nm, L2=340 nm, W2=86 nm, S2=43 nm. The periodicity is 690 nm and the metal thickness is 40 nm. The spectral far-field response was characterized using a Fourier Transform Infrared spectrometer (NA=0.4). The far-field transmission spectrum exhibits the typical induced transparency feature (fig. 3.8 (b)), namely, a narrow transparency band at 1.33 µm wavelength between the two transmission dips at the wavelengths of 1.2 µm and 1.5 µm. The Transmission peak doesn’t reach 100 per cent as the metal loss cause a finite lifetime in the antenna. On the other hand, the spectrum is not completely symmetric regarding the transparent frequency due to the slight misalignment between the dipole and the quadrupole resonance frequency. This spectrum is confirmed by the finite-difference-time-domain (FDTD) simulation, shown as the red dashed line. Using the model of two coupled resonators as discussed above, we retrieved the mutual coupling strength between the quadrupole and the dipole resonators from the experimental far-field transmission spectrum. The damping rate of the dipole is $g_d = 89$ THz and for quadrupole is $g_q = 47$ THz; the effective coupling rate $k = 84$ THz, which is stronger than the geometric mean of the damping rates of the two resonators. This result experimentally confirmed that the strong coupling occurs between the two plasmonic elements.

![Diagram](image)

**Fig. 3.8**: (a) The SEM images of the PIT sample fabricated by e-beam lithography with the dimensions indicated. Scale bar is 300nm (b) The transmission spectrum with the polarization of incident light, indicated in (a), measured by Fourier Transform Infrared Spectrometer (solid) and the numerical simulation results (dashed) by CST Microwave Studio. (c) The near-field spectrum of SPP oscillation amplitude of the electric dipole (red dashed) and quadruple (green solid), respectively. Three typical conditions with different contrasts between dipole and quadrupole are marked by dash lines of different colors. The near field experiments are operated at these three frequencies.

With the parameters retrieved from the far field transmission experiment, we obtained the near-field oscillation amplitude of the resonators spectrum, as shown in fig. 3.8(c). The resonant peak of the dipole SPP (the red curve), due to the coupling, splits into a ‘M’ shape and shows an anti-resonance valley at the transparency frequency of 1310 nm, while that of the quadrupole SPP (the green curve) exhibits a strong enhancement plateau comprising of two largely overlapped
resonant peaks. This shares the similar trend as the electronic population distributions in non-radiative state and the radiative state in atomic systems. In the following, we investigate the localized SPP intensity at three different frequencies relative to the transparency point, marked by the dash lines of different colors in fig. 3.8(c) using our a-NSOM system.

![Image]

**Fig. 3.9:** Local electric field intensity mapped by the a-NSOM. (a) The near field result measured at 1310 nm. The NSOM signal are rendered in green and superimposed onto AFM topography results in grey. Only four corners of the gold rod pair light up, which stands for the dominant population of SPP in electric quadrupole over the dipole. The mere excitation of the quadrupole accounts for the transparency in the far field. Scale bar is 300nm. (b) The near field result measured at 1530 nm. The excitation locations of the localized SPP are transferred to the ends of dipole antenna.

Fig. 3.9 (a) displays the local electric field intensity (in green) at the transparency wavelength of 1310 nm, where the retrieved oscillation amplitude at the 'dark' state exceeds the oscillation amplitude at the 'bright' state. The intensity distribution is superimposed with the topographic image (in grey), simultaneously taken with atomic force microscopy and NSOM in the same scan. The near-field image shows that, in each single PIT structure, the non-radiative quadrupole antenna is excited, forming four localized SPP hot spots at all ends of the metallic double bar and the radiative dipole is significantly suppressed due to the deconstructive interference between the direct and indirect excitation paths described in fig. 3.7(a). As a result, the absorption and the scattering of the dipole antenna are greatly suppressed, and the overall system appears transparent in the far field. This observation that the PIT system is purely excited in the non-radiative state at the transparency frequency, confirms the importance of exciting the dark element in achieving the induced transparency.

At the wavelength of 1530 nm, far from the transparency window, as shown in fig. 3.9(b), the near-field measurement reveals an excitation of the dipole with localized SPP forming at the ends of the dipole antenna. This agrees well with the near-field spectrum obtained in fig. 3.8(c). At this wavelength, only the bright element in the PIT structure is excited because the dark element has a relatively higher Q factor and is far off the excitation frequency. Therefore, the extinction at this frequency can be mainly attributed to the electromagnetic loss of the bright element.
Fig. 3.10: Near-field result at the dress state frequency. (a) The near-field mapping result at 1310 nm, the dressed state wavelength of another sample. The optical signal, represented by the blue hotspots at the ends of both single rod and rod pair, indicate that SPP in the two resonators have comparable populations at this condition. Scale bar is 300 nm. (b) Numerical simulation of the structure at the same frequency. Blue color stands for positive $\pi/2$ and red represents $-\pi/2$ phase (c) The schematic of the phase of the electric field based on the experimental and the numerical results. The signs stand for the instantaneous charge accumulated at the ends of the antenna.

Interestingly, we also observed the field distribution of a state hybridized of both dipole and quadrupole state. At the frequency marked by the blue dash line in fig. 3.8 (c), the transmission is at one of the dips corresponding to a hybridized mode wherein the non-radiative resonator shows comparable oscillation strength as that of the radiative resonator. At this condition, the near-field image in fig. 3.10(a) indeed displays the excitations of both bright and dark elements. For this sample, the geometry has been modified to match the dressed state frequency with the available laser frequency. The geometry of the sample is $L_1=306$ nm, $W_1=61$ nm, $S_1=105$ nm, $L_2=347$ nm, $W_2=72$ nm, $S_2=50$ nm. The periodicity is 690 nm and the thickness of the metal film is 30 nm. The retrieved parameters are respectively $g_A=34$ THz and $g_B=124$ THz; the effective coupling rate $k_A=113$ THz. The new dress state wavelength is at 1530 nm.

As can be seen in fig. 3.10(a), the electric field occurring at the quadrupole ends closer to the dipole is oscillating stronger than the ends away from the dipole. In Fig. 4(b), another FDTD simulation confirms this effect. The physics behind it is shown schematically in fig. 3.10(c), where the instantaneous charge accumulation associated with the oscillating electric field is labeled by the positive and negative signs at the ends of the bright and dark resonators. As shown in the fig. 3.10(c), the closer-to-dipole end of the quadrupole is oscillating in phase with the neighboring dipole end. Therefore, the Coulomb repulsion between the charges of the same sign in the dipole and quadrupole SPP renders the closer ends weaker than the distant ends in intensity, which can explain the observations in the experiment and simulation. Such a spatial visualization of the near-field interactions effectively reveals the interference result between dark and bright elements in the PIT atom, reassembling the coherent processes among energy states in atomic EIT.
3.4 Anti-Hermitian coupling in plasmonic antenna array

3.4.1 Introduction to Anti-Hermitian coupling in open quantum systems

The indirect coupling among quasi-bound states through common continuum decay channels have been widely studied in various open quantum systems, in particular for the investigation of the resonance phenomena in nuclei, atoms, molecules, and quantum dots.\textsuperscript{46-48} Interesting features of these systems have been observed, such as the restructing of the eigen-states with contrasting lifetimes in the system: some long-lived states are deprived of the coupling strength to the decay channels, while others with enhanced coupling to the decay channels having very short lifetime, i.e. the so called “super-radiant” states that was first proposed by Dicke for describing the coherent spontaneous radiation of a gas confined to a sub-wavelength scale.\textsuperscript{49} Spectroscopically, this phenomenon is manifested as sharp resonances super-imposed on the broad super-radiant states, as observed, for example, in the decay of compound nuclear states in certain nuclear reactions.\textsuperscript{50-52}

A general treatment of the open systems, the standard projection formulism, divides the Hilbert space of an open system into two subspaces, \{\textit{Q}\} subspace that consists of interior quasi-bound (discrete) states \{|\textit{q}\}\rangle and \{\textit{P}\} subspace that only consists of the continuum decay channels\{|\textit{c}\}\rangle, with \textit{Q} and \textit{P} being the projection operators of the two subspaces.\textsuperscript{53} The effective Hamiltonian acting on the bound states by taking into account the indirect coupling mediated by the continuum decay channels can be written as\textsuperscript{54},

\[
H_{\text{eff}} = H_{\text{QQ}} + H_{\text{QP}}(E - H_{\text{PP}})^{-1}H_{\text{PQ}}
\]  

(2.9)

where \(H_{\text{QQ}}\) is the Hamiltonian projected in the subspace of quasi-bound states, \(H_{\text{QP}}\) and \(H_{\text{PQ}}\) are the coupling matrices between the \textit{Q} and \textit{P} subspaces, and \((E - H_{\text{PP}})^{-1}\) is the propagator in the subspace of open channels. The second term on the RHS of Eqn (1) represents the indirect interaction among the bound states mediated by the open (continuum) channels. The matrix elements of the effective Hamiltonian can be further expressed as\textsuperscript{46},

\[
\langle q_1 | H_{\text{eff}} | q_2 \rangle = \langle q_1 | H_{\text{QQ}} | q_2 \rangle + \sum_{c=1}^{K} PV \int dE \frac{V_{q_1}^c V_{q_2}^c}{E - E'} - \frac{i}{2} \sum_{c=1}^{K} V_{q_1}^c V_{q_2}^c
\]

(2.10)

or in the matrix form,

\[
H_{\text{eff}} = H_{\text{QQ}} + PV \int dE \frac{VV^*}{E - E'} - \frac{i}{2} \frac{VV^*}{2}
\]

(2.11)

where P.V denotes the Cauchy principal value of the integral, \(V_q^c = \sqrt{2\pi} \langle q | H_{\text{QP}} | c \rangle\), \(K\) is the number of decay channels. Thus, within the quasi-bound states subspace, the effective Hamiltonian of the system contains an imaginary, or anti-Hermitian
coupling matrix \((-iVV^*/2\) that arises from the indirect coupling among the bound states mediated by the open channels.\(^{55,56}\) For an open quantum system consisting of \(M\) bound states coupled to \(K\) continuum decay channels, the coupling matrix \(V\) between the discrete and continuum subspaces is of dimension \(M \times K\), and the indirect anti-Hermitian coupling matrix \(-iVV^*/2\) has a dimension of \(M \times M\), but with a rank of \(K\) in the case of \(M > K\). For an open system with a single common decay channel \((K = 1)\), \(V = [V_1, V_2, \ldots, V_M]^T\) is simply a column vector. Note that the Cauchy principal value of the indirect coupling is the \(2^{nd}\) term on RHs of eq. 2.11, which is real and generally non-zero, leading to a shift of the energies of eigen-states, which corresponds to the so-called collective Lamb shift.\(^{57-59}\)

### 3.4.2 Analog in plasmonic antenna array

The governing Hamiltonian for a quantum open system with a single continuum channel can be mapped onto a plasmonic system consisting of an array of plasmonic dipole antennas with the same orientation positioned in proximity to each other (separations < wavelength). Being an open system, the coupling can generally be divided into two parts, the direct near field coupling, and the indirect coupling mediated by radiative channels, i.e., continuum states. For antennas with small separations and the same orientation, the antennas primarily couple to the same dipolar radiation mode, i.e., a single decay channel with \(K = 1\). In addition, the separations between the antennas can be carefully designed so that the effect of the direct near field coupling and the Cauchy principal value of the indirect coupling counteract each other, and the system is dominated by the anti-Hermitian part of the indirect coupling.

Continuum mediated anti-Hermitian coupling in an array of densely positioned plasmonic antennas can be utilized for spatially manipulating light in the deep subwavelength scale. We consider an array of plasmonic antennas with different resonance frequencies \(\omega_j \neq \omega_i\), and investigate the coupling phenomena of antennas under the excitation of a plane wave. According to the discussion in Ch. 3.2, the general form of the coupled equations can be written as,

\[
\begin{bmatrix}
-\omega + \omega_1 - i\gamma_1 & -\kappa_{1p} & -\kappa_{1M} \\
\cdot & -\omega + \omega_p - i\gamma_p & -\kappa_{pM} \\
\cdot & \cdot & -\omega + \omega_M - i\gamma_M \\
\end{bmatrix}
\begin{bmatrix}
A_1 \\
A_p \\
A_M \\
\end{bmatrix} =
\begin{bmatrix}
g_1 \\
g_p \\
g_M \\
\end{bmatrix} E_0 \quad (2.12)
\]

where \(\gamma_\phi\) is the dissipation term, \(\kappa_{pq}\) (\(\kappa_{pq} = \kappa_{qp}\)) is the coupling coefficient between the \(p\)th and \(q\)th antennas, \(g_p\) denotes the coupling strength between the incident plane wave and the antenna, with \(g_p \propto V_p G\), and \(E_0\) is the electric field of the incident wave. For dipole antennas operating at infrared frequencies, the overall loss is
usually dominated by the radiation loss (coupling to the continuum channel). For simplicity of analysis, we assume that \( \gamma_p \) only contains the radiation loss. We further assume an ideal case that the Hermitian part of the overall coupling, which includes the direct near field coupling, and the Cauchy principal value of the indirect coupling, vanishes, and \( \kappa_{pq} \) only consists of the indirect coupling mediated by the single decay channel. With the above assumptions, \( i\gamma_p \) and \( \kappa_{pq} \) are just the diagonal and off-diagonal elements of the anti-Hermitian coupling tensor \( iVV' / 2 \), respectively, i.e.

\[
i\gamma_p = iV_p^2 / 2, \quad \kappa_{pq} = iV_p V_q / 2
\]

(2.13)

At the resonance frequency of the pth antenna (\( \omega = \omega_p \)), the pth column of the coupling matrix in eq. 2.12 is reduced to,

\[
\begin{bmatrix}
-\kappa_{1p} \\
\vdots \\
-i\gamma_p \\
\vdots \\
-\kappa_{Mp}
\end{bmatrix} = -i \frac{V_p}{2} \begin{bmatrix}
V_1 \\
\vdots \\
V_p \\
\vdots \\
V_M
\end{bmatrix}
\]

(2.14)

which has a linear relation with the excitation vector \([g_1, g_2, \ldots]^T\) on the right hand side of eq. 2.12. Thus, eq. 2.12 can be solved as,

\[
A_p = \frac{2ig_p}{V_p^2}, \quad A_q(p \neq q) = 0
\]

(2.15)

Eq. 2.15 suggests that only a single resonator can be selectively excited at its resonance frequency, whereas all the others remain completely unexcited. In a strongly coupled system the eigen-states of the multi-antenna system are hybridized excitations of more than one resonator. This indicates that, under plane wave excitation of the antenna array, multiple eigen-states are simultaneously excited, which interfere constructively at a single resonator, and destructively at all the others, leading to a highly localized state. Since the spatial separation between the neighboring antennas can be much less than the wavelength of the incident electromagnetic wave, the plasmonic system governed by anti-Hermitian coupling offers an interesting platform for manipulation of light in the deep subwavelength scale.

The assumption of purely anti-Hermitian coupling constants among all the antennas is apparently too ideal in a realistic plasmonic system. Nonetheless, this condition can be relaxed, and only the coupling constants among those antennas with close resonance frequencies are critical for achieving a good contrast in selective excitation of individual antennas. To the first order approximation, a small deviation \( \Delta\kappa_{pq} \) of the coupling coefficient \( \kappa_{pq} \) from the ideal coupling strength of \(-iV_p V_q / 2\)
results in a non-zero $A_q \sim \Delta \kappa / (\omega_q - \omega_p) \cdot A_p$ at the resonance frequency of the $p$th antenna. Therefore, the deviation of coupling constant from the idea case for two resonators with very different resonance frequencies has very little effect on the performance of the system.

3.4.3 Numerical investigation of the critical coupling point

A simple implementation of a plasmonic system exhibiting anti-Hermitian indirect coupling mediated by a single open channel is shown in fig. 3.11. The system consists of two optical dipole antennas of slightly different lengths with deep sub-wavelength spacing. Each optical antenna is capable of focusing light into the gap region of nanometer scales.

![Image of a plasmonic system](image)

**Fig. 3.11:** A plasmonic system consisting of two optical antennas made of gold on a quartz substrate. The lengths of the antennas are 430 nm and 470 nm. The width and thickness of both antennas are 40 nm and 25 nm, respectively. A beam is incident at the normal direction onto the antennas, with polarization along the longitudinal direction of the antennas.

Numerical simulation (CST Microwave Studio) was carried out to calculate the optical response of each antenna at their gap centers under the illumination of a plane wave at normal incidence, and numerical fitting was subsequently employed to retrieve the coupling constant between the two antennas. In the simulation, Drude model was used for the gold dielectric parameters, with $\omega_p = 1.37 \times 10^{16} \text{ rad} / s$, and $\gamma = 4.1 \times 10^{13} \text{ rad} / s$. A typical near field intensity plot is illustrated in fig. 3.12 (b) and the retrieved parameters are shown in fig. 3.12 (a) for a range of separations $s$ between the two antennas.
Fig. 3.12: (a) The retrieved real part (black) and imaginary part (red) of optical coupling constant between the two antennas, and the dissipations of the long (blue) and short antennas (green) as a function of their edge-to-edge separation $s$. (b) The simulated (solid) and fitted (dashed) spectra of the electric field probed at the center of the gap of two coupled antennas with lengths of 430 nm (black) and 470 nm (red) at separations of 30 nm, where the coupling between the two antennas is almost purely anti-Hermitian. The responses of the two antennas show highly asymmetric profiles, a characteristic of Fano lineshape. The vertical dashed lines mark the eigen-frequencies in each figure.

The coupling constant has an imaginary part (red curve) that is only slightly less than the dissipation of both antennas (blue and green) over a broad range of $s$ up to about 250 nm. The approximate equity $\text{Im}(\kappa_{12})^2 \approx \gamma_1 \gamma_2$ indicates the presence of an anti-Hermitian term of rank 1 in the governing Hamiltonian of the plasmonic system, i.e. both antennas primarily couple to a single radiative decay channel. The slight difference between the anti-Hermitian coupling strength $\text{Im}(\kappa_{12})$ and the dissipations at very small antenna separation is mainly due to the intrinsic ohmic loss of the antennas. As $s$ increases above 250 nm, the effect of other radiative channels, such as electric quadrupole and magnetic dipole radiation modes, start to become significant, and the assumption of a single continuum channel does not hold anymore. Consequently, the imaginary part of the coupling constant deviates significantly from the dissipation rates of the antennas, and the foregoing analysis no longer applies in this regime.

At very small separations, the real (Hermitian) part of the coupling coefficient exhibits a large negative real part, which is dominated by the direct near-field coupling between the two antennas. With increasing distance, the near field coupling decreases rapidly, and the real part of the coupling crosses zero at a separation around 30 nm, where the direct near field coupling is cancelled by the Cauchy principal value of the indirect coupling, leaving a purely anti-Hermitian coupling between the two antennas. Around this separation, highly asymmetric Fano profiles for the response of each antenna are observed, as indicated by fig. 3.12(b). There exists a strong suppression of the excitation of each antenna as manifested by the spectral dip at 225 and 241 THz, respectively (marked by the black and red dashed lines). Although the spectral dips at these two frequencies do
not go to zero as given by eq. 2.15 for the ideal case, there appears a large contrast between the amplitudes of the two antennas. Thus, light localization can be switched from one antenna to the other by slightly tuning the frequency of incident light. As fig. 3.12(a) shows, the imaginary part of coupling constant is greater than the real parts for a broad antenna separation range from 20 nm to 350 nm. Thus, the coupled optical antennas serve as a model system for studying the interesting physics in a physical open system dominated by the anti-Hermitian coupling.

Next we extend the plasmonic system to include five optical antennas with deep sub-wavelength spacing (Fig. 3.13). The length of the antennas ranges from 450 nm on one side to 610 nm on the other side with a step of 40 nm. The edge-to-edge spacing between the nearest neighboring antennas is 45 nm, and the separations between the second, third and forth nearest neighbors are 130 nm, 215 nm, 300 nm, respectively.

![Fig. 3.13: The schematic of a plasmonic antenna array consisting of five optical antennas with gradually varying lengths, ranging from 430 nm to 590 nm at a step of 40 nm. The thickness and width of the metal strips are 25 nm and 40 nm, respectively. The edge-to-edge separation between the nearest neighbors is 45 nm.](image)

As indicated by fig. 3.14(a), this configuration gives rise to imaginary (anti-Hermitian) part dominated coupling coefficients among the antennas in the array. The anti-Hermitian part of the coupling constants roughly satisfy $\text{Im}(\kappa_{pq})^2 \approx \gamma_p \gamma_q$ up to the third nearest neighbor ($s = 215$ nm), but not for the two antennas at the two ends (forth nearest neighbor). However, as the two antennas at the two ends have dramatically different lengths and therefore very different resonance frequencies, this deviation has only a small effect on the overall performance. The major deviation from the ideal case described by eq. 3.14 is that the there exists real (Hermitian) part of the coupling constants among the antennas. Nonetheless, it will be shown later through both numerical simulations and experimental observations that the selective excitation of individual antennas can still be achieved despite the presence of Hermitian coupling among the antennas.
Fig. 3.14: (a) The calculated electric field magnitude at the gap center of each antenna versus the optical wavelength. At the resonance of each antenna, all the other antennas are strongly suppressed, leaving a highly selective excitation of a single antenna. (b) The spectral response for five uncoupled antennas with the same geometries as (a).

We performed the numerical simulation on the antenna array using the commercial software, CST Microwave Studio™. In the simulation, a plane wave is incident at normal incidence onto the array of antennas. As shown in fig. 3.14(a), each antenna exhibits a large resonance peak at approximately its uncoupled resonance frequency (fig. 3.14(b)), where all the other antennas are strongly suppressed, leading to selective excitation of an individual antenna. This observation is consistent with the theoretical analysis given by eq. 15, despite the presence of Hermitian coupling constants and ohmic loss that result in a deviation of $\kappa_{pq}$ from $i\gamma_p\gamma_q$ for interaction between certain antennas. In contrast to the selective excitation and sharp spectral features for the coupled antenna array, an array of uncoupled antenna exhibits very poor excitation selectivity due to the relative broad line-width of the resonance peak, as shown in fig. 3.14(b).

To visualize the selective excitation of individual antennas, we numerically obtained the out-of-plane electric field intensity distribution of the antenna array at the resonance frequencies of each antenna, as shown in Fig. 3.15(a-e). Note that the out-of-plane component of the electric field represents the charge distribution of the plasmonic system. The field distribution confirms that the antennas are individually excited at their resonance wavelengths, and the hot spots shifted from the shortest antenna at 1200 nm wavelength to the longest antenna at a wavelength of 1600 nm.

3.4.4 Experimental results

Experimentally, we fabricated the plasmonic antenna array and directly observed selective excitation of individual antennas at carefully chosen wavelengths, as shown in Fig. 3.15. The plasmonic antenna array was fabricated on a quartz wafer using electron beam lithography, followed by metal evaporation and a lift-off process. To characterize the optical response of the plasmonic antenna array, a-NSOM system was employed to measure the near-field distribution at the top surface of the plasmonic multiplexers. A supercontinuum light source together with
an acoustic-optical modulator is used as a tunable light source in the near-infrared range with a linewidth of 5 nm. Fig. 3.15(f-j) shows the near field optical distribution for the array of optical antennas at five different wavelengths from 1200 nm to 1600 nm at a step of 100 nm. At each optical excitation wavelengths, only a single antenna was strongly excited, whereas all the others were suppressed, which is in good agreement with the simulation results.

![Figure 3.15](image)

*Fig. 3.15: Experimental verification of the selective excitation of individual antennas in the plasmonic antenna array. (a-e) The simulated near-field distributions of the antenna array at five different wavelengths: 1200 nm (250 THz), 1300 nm (230.8 THz), 1400 nm (214.3 THz), 1500 nm (200 THz) and 1600 nm (187.5 THz), showing selective excitation of individual antennas. (f-j) The corresponding experimental observations show very good agreement with the simulations. The optical image (red), measured by NSOM, is overlapped onto the topographical image (grey) measured by the atomic force microscopy.*

The NSOM measured localization of light in the antenna array is further quantified by a plot of intensity of light integrated along the y direction, versus x, at each measured wavelength, as shown in fig. 3.17. On average, light is strongly confined within a scale of 70 nm along the x direction at each excited wavelength. Both simulation and measurement show that light can be confined and manipulated in the deep subwavelength scale in an array of optical dipole antenna, as the center-to-center distance between the neighboring antennas is only 85 nm that is about 1/15 of the optical wavelengths.

![Figure 3.16](image)

*Fig. 3.16: The measured intensity of light integrated along the y direction, versus x direction, at each measured wavelength labeled in fig. 3.15.*
To highlight the role that the coupling among the antennas plays, a control sample, which consists of an array of antennas with the same geometrical parameters, but spatially separated far (300 nm) from each other, was measured, at a wavelength of 1300 nm. At this separation, there exists a large difference between the imaginary part of the coupling constant and the dissipation rate of each antenna, and therefore a single open channel approximation does not apply. As shown in fig. 3.17, there is no selective excitation of individual antennas as several antennas were simultaneously excited. The comparison between the control sample and the well-designed antenna array implies that the coupling among antennas mediated by a single decay channel plays a significant role in achieving selective excitation of individual antennas.

Fig. 3.17: The near field measurement at 1400 nm wavelength on the control sample, which consists of an array of antennas with the same geometry specification but at a large nearest neighbor separation of 300 nm. Without the anti-Hermitian coupling, all the antennas within the focus spot, which is around 2 mm, are excited simultaneously due to the spectral overlapping of the resonances. Scale bar is 300nm.

3.5 Conclusion and outlook

In this chapter, we have discussed the fundamentals of plasmonic antenna theory, the coupled resonator model and show that the coupling between antennas can offer many interesting optical phenomena at the nano scale. By coupling the dark and the bright element, we achieve a nanoscopic analog of EIT effect, which can serve as a compact delay line element for the future integrated photonic circuit. In the other work, the anti-Hermitian coupling between antennas effectively reduces the line width of each resonator so their spectral responses on longer overlap with each other. Both of them might play a very important role in the future photonic circuit.

One particularly interesting application is using these devices as Nanoscale spectrum splitter for the solar cell. When a photon of energy much larger than the band gap is absorbed in the solar cell, it will relax to the band gap and generate a voltage no more than the band gap energy, which is known as the thermalization loss. To save these energy, tandem cell based on multiple junctions has been heavily employed in many extreme situations such as extraterrestrial exploration. However, the current serial architecture in the tandem cell is suffering from many problems.
For example, the current matching in the serial circuit limits the whole cell’s current performance to be the poorest junction’s performance. Optically, each junction must share the same concentration factor and no optimization according to the solar spectrum is allowed. Also, it is very challenging to fabricate multiple different epitaxial single crystal semiconductor layers.

In the future, we can use both our EIT and anti-Hermitian device to sort the light into different in-plane hotspots, thus changing the serial architecture into a parallel one. In addition, the plasmonic resonating elements can help enhance the absorption in the cells originally with low absorption rate. After the spectral splitter, even more narrow band photon managing elements can be added to further improve the efficiency. For example, non-reciprocal propagation has recently been demonstrated in nano optics with broken PT symmetry, but it’s an intrinsically narrow-band effect. If integrated with the spectral splitter in the solar cell, this can further boost the cell’s open circuit voltage by prohibiting the spontaneous re-emission back into the source.
Chapter 4

Photonic spin Hall effect at the metasurface

In this chapter, we will discuss an enhanced photonic spin Hall effect (SHE) at metasurface, a very recent example of utilizing array of plasmonic antennas to achieve an very weak effect. The rapidly varying phase discontinuities along a metasurface, breaking the axial symmetry of the system, enable the direct observation of large transverse motion of circularly polarized light, even at normal incidence. Within the chapter, I will first review the fundamental physics behind the optical spin Hall effect and explain the basic concept of plasmonic antenna based metasurface. Then I will discuss our experimental demonstration of photonic SHE at the metasurface, which provide a new degree of freedom in transferring the angular momentum between photon spin and orbit degree of freedom.

4.1 Introduction

4.1.1 Optical spin Hall effect at the interface

The spin Hall effect (SHE) of photon is an photonic analogy to the spin Hall effect of charged particle: during the transportation of the particle, its trajectory will be splitted in the direction transverse to the transport one according to the spin polarization.\(^6\) It is induced by the finite berry curvature effect from the band structure or impurity scattering.\(^6\) In analogy to SHE in electron system, the spin-1 photon take the place of spin-1/2 charge and the refractive index gradient take the role of the electric potential gradient.

![Fig. 4.1: Illustration of the optical spin Hall effect at the interface between air and SiO2 (index: 1.515). As the beam is obliquely incident onto the interface, the orthogonally circular polarized beam are shifted laterally towards different directions.](image)

Fig. 4.1: Illustration of the optical spin Hall effect at the interface between air and SiO2 (index: 1.515). As the beam is obliquely incident onto the interface, the orthogonally circular polarized beam are shifted laterally towards different directions.

One example of the optical SHE at the interface between air and SiO2 is shown in the figure 4.1. When a light beam with finite diameter is incident with an oblique angle, \(\theta_i\), onto the interface, it will be refracted to a new angle \(\theta_T\), according to the classical Fresnel law. In addition to that, the transmitted beam will experience a lateral shift, transverse to the plane of wave vector and surface norm, with the shift sign dependent on the incident photon’s helicity. This can be understood as the conservation of photon’s total angular momentum along the z direction. When the
\( \sigma \), beam is bent more towards the z direction by the interface, the spin angular momentum is increased. In our time reversal symmetric system, the total angular moment, \( J = j_{\text{spin}} + j_{\text{orbital}} \), is conserved.\(^{62} \) Consequently, the orbital angular momentum has to decrease, corresponding to a lateral beam shift towards the left. For the photon of \( \sigma \) helicity, the spin angular momentum is decreased as the beam bends, and therefore the orbital part is increased, resulting a right-direction lateral shift. This is defined to be the optical spin Hall effect. Only in a finite-sized excitation beam can it be observed, since a plane wave will have an infinite beam width which obscures the observation of the lateral shift.

Fig. 4.2: Plane wave decomposition of the Gaussian beam. The polarization of wave component with non-zero in-plane wave vector will rotate after transmitting through the interface.

The optical SHE can also be understood in the context of classical optics language. (fig. 4.2) The finite width incident beam can be spatially Fourier transformed into a series of plane wave of different wave vectors. They form a Gaussian distribution with the center at the incident wave vector, \( k_0 \). If the incident beam is prepared as the vertical polarization, as the figure shows, they are not purely P polarized light for all the k components, and they have to be decomposed into two orthogonal components, S and P. As the figure shows, the wave component with the largest in-plane wave vector will posses the largest S polarization component. On the other hand, the transmissivity of S wave and P wave will be different, depending on the incident angle. Then, the transmitted light will have an electric field of a different direction compared with the incident one. This can also be considered as the result of the transversality. A new horizontal component will emerge for wave component. For the wave of positive y wave vector, the new polarization is rotated clockwise by an angle proportional to \( k_y \), and the wave of negative \( k_y \) will acquire an anti-clockwise polarization rotation of the same value. In reference of the circular polarized beam, this can be seen as the superposition of an increase in the phase of \( \sigma \) light and a decrease in the phase of \( \sigma \) light:
\[ |V\rangle|k_y\rangle \rightarrow \frac{1}{\sqrt{2}}(\exp(-ik_0\delta)|+\rangle + \exp(ik_0\delta)|-\rangle)|k_y\rangle = \sum \exp(-ik_0\sigma_0\delta)|s\rangle \quad (3.1) \]

The total wave packet can be expanded in both \( y \) and \( k_y \) space:

\[ |\Psi_s\rangle = \int dy\Psi(y)|y\rangle = \int dk_y\Phi(k_y)|k_y\rangle|s\rangle \quad (3.2) \]

After the fraction, each \( k_y \) component experience a different phase delay

\[ |\Psi_s\rangle = \int dk_y\Phi(k_y)\exp(-ik_0\sigma_0\delta)|k_y\rangle|s\rangle = \int dy\Psi(y - s\delta)|y\rangle|s\rangle \quad (3.3) \]

This indicates that the circular polarized photon will experience a real space side jump after the refraction, with the sign depending the helicity. The magnitude of the jump can be determined from the Fresnel law analysis and is plotted in the figure 4.3.

![Diagram showing calculations based on geometry in fig. 4.1. H and V are the incident polarization directions; + and – stand for the left and right circular polarized components in the output beam. Adapted from Ref. 63.](Image 156x357 to 441x644)

Fig. 4.3: Calculation based on the geometry in fig. 4.1. H and V are the incident polarization directions; + and – stand for the left and right circular polarized components in the output beam. Adapted from Ref. 63.

The optical SHE is a weak effect because of the negligible photon momentum together with a weak spin-orbit interaction. However, it was detected via a sensitive measurement called weak measurement. 63 It is based on the pre and post selection process, which can arbitrarily amplify the signal strength that is otherwise very small, and cannot be distinguished. The newly defined weak value is defined as \( A_w = \langle \psi_2|A|\psi_1\rangle / \langle \psi_2|\psi_1\rangle \), where \( \psi_1 \) and \( \psi_2 \) are the pre and post selected wave function. Evidently, if they are close to orthogonal to each other, the weak value can be much larger than the measurable \( \langle \psi_1|A|\psi_1\rangle / \langle \psi_1|\psi_1\rangle \). Specifically in the SHE experiment, the post selection can be a polarization state perpendicular to the incident one. Although most of the signal light is suppressed, the shifted value is well within the detection limit.
4.1.2 Plasmonic antenna based metasurface

Metasurface is defined to be a surface tiled with artificial resonating structures. The strong light-matter interaction on the surface can efficiently reroute the transmitted/reflected light. Since the recent introduction of metasurface concept, many interesting optical functions have been demonstrated. One example based on the resonating plasmonic antenna is shown in fig. 4.4.

![Image of metasurface structure]

Fig. 4.4: Scanning electron microscope (SEM) image of the metasurface structure working in the near infrared range. Adapted from Fig. 65.

The goal of the metasurface design is to introduce a phase gradient in the scattered wave, so that the new wave front can be arbitrarily modified. The specific design here is composed of an array of plasmonic antennas. Each unit cell contains eight V shape antenna. The eigen modes of each V shape antenna can be viewed as a hybridization result between two single antennas, as discussed in the chapter 3.2, with one labeled as symmetric mode and the other as anti-symmetric mode. The symmetric mode is polarized along the mirror axis of the antenna and the anti-symmetric one is along the orthogonal direction. They are resonating at different frequencies due to the coupling term in the coupled oscillator model (Ch. 3.2).

When the incident electric field is polarized along the horizontal or vertical directions, which is not polarization direction of the eigen-modes, a new electric field component along the orthogonal directions will be generated. The amplitude and phase of the new cross-polarized components are dependent the oscillating current phase and amplitude within the antenna, and therefore can be engineered by tuning the length of antenna as well as the angle between two arms. Once four antenna configuration are identified with an overall $\pi$ phase coverage, the second half of the cell can just repeat the first half with a 90-degree real-space rotation. Such a rotation can introduce another $\pi$ phase change in the scattered wave, so that the whole unit cell can achieve a $2\pi$ phase coverage with equal scattering amplitude.
**Fig. 4.5:** The illustration of the new wave front composed of the scattering wave from eight different V shape antennas. The tilting angle is $\lambda_0 / \Gamma$. Adapted from Ref. 64.

The resulting wavefront engineering effect can be seen in the figure 4.5. Although the electromagnetic wave only passes a very thin layer of gold film, the resonating structure endows the scattered wave a very dramatic phase change relative to the incident phase. With the continuous in-plane phase gradient, the wave front is effective tilted by an angle, $\lambda / \Gamma$. Furthermore, because the phase difference exists in the scattered wave, both refracted and reflected waves will experience such a tilting. Compared with the traditional phase accumulation during the wave propagation, the phase change is very abrupt at the interface. Therefore, they can allow the optical functions to happen at the sub-diffraction-limited level, regardless of the refractive index of the substrate. In addition, the new wave front can be engineered into different regimes, for example, large beam bending at the normal incidence, negative refraction or reflection, and complete conversion between the free space propagating wave and the surface wave at the metasurface.

### 4.2 Experimental results and discussion

The experimental observation of spin Hall effect of light is fundamentally challenging since the amount of momentum that a photon carries is exceedingly small. The exploration of such a weak process relies on the accumulation of the effect through many multiple reflections\(^7\) or ultra-sensitive quantum weak measurements with pre- and post-selections of spin states.\(^{63,71}\) Moreover, the present theory of photonic spin Hall effect (PSHE) assumes the conservation of total angular momentum over the entire beam\(^7\), which may not explicitly hold especially when tailored wavelength-scale photonic structures are introduced. Here we demonstrate experimentally the strong interactions between the spin and the orbital momentum of light in a thin metasurface, a two-dimensional electromagnetic nano-structure with designed in-plane phase retardation over the wavelength scale.\(^\text{13}\) In such an optically thin material, the resonance-induced anomalous “skew-scattering” of light destroys the axial symmetry of the system and we observed PSHE even at the normal incidence. In stark contrast, for conventional interfaces between two homogeneous media, the spin-orbit coupling does not exist at the normal incidence.
4.2.1 Polarization rotation induced by SHE

The PSHE or the spin-orbit interaction arises from the non-collinear momentum and velocity (the change of trajectory) of light. When light is propagating along a curved trajectory, as schematically shown in fig. 4.6 (a), the time-varying momentum along the light path has to introduce a geometric polarization rotation to maintain the polarization transverse to its new propagation direction\textsuperscript{73}, \( \hat{e} = -k(\hat{e} \cdot \hat{k}) / k^2 \). Here \( \hat{e} \) and \( k \) are the polarization vector and the wave vector, respectively. Similar to the role of Rashba field in electron SHE, the rotation of the polarization depends on the helicity of light and may be considered as circular birefringence with a pure geometric origin\textsuperscript{74}. Moreover, as the back-action from geometric polarization rotations, the spin-orbit interaction also changes the propagation of light, as we will show in later sections, resulting in a polarization-dependent transverse displacement for light.

![Fig. 4.6: (a) Schematic of PSHE. When light is propagating along a curved trajectory, the transversality of electromagnetic waves requires a rotation of polarization vector in the transverse planes along the trajectory. (b) The strong spin-orbit interaction within the optically thin metamaterial leads to the accumulation of circular components of the beam in the transverse directions (y'-directions) of the beam even when the incident angle is surface normal.](image)

We consider first an ordinary interface between two homogeneous media. When a Gaussian beam impinging onto the interface at normal incident angle, the axial symmetry eliminates the spin-orbit coupling and the total angular momentum of the entire beam is conserved. By designing a metasurface with a rapid gradient of phase discontinuity, \( \nabla \Phi \) along the interface in the \( x \)-direction (fig. 4.6 (b)), however, we introduce a strong spin-orbit coupling when the light is refracted off the interface. The rapid, wavelength scale phase retardation can be incorporated in the optical path by suitable design of the interface\textsuperscript{75}. Such a position-dependent phase discontinuity removes the axial symmetry of the interface and therefore allows us observing PSHE even at the normal incident angle.
Figure 4.6 (b) shows schematically PSHE for the light beam that is refracted off a metasurface with rapid in-plane phase retardation. The momentum conservation at the metasurface now has to take into account that the position dependent phase retardation, and the induced effective circular birefringence are determined by the gradient of the in-plane phase change or the curvature of the ray trajectory, \( (\mathbf{k} \times \mathbf{k}) / k^2 \). As a result, for a linearly polarized incidence, light of opposite helicities will be accumulated at the opposite edges of anomalously refracted beam in the transverse direction (fig. 3.6 (b)). The faster the in-plane phase changes, the stronger the effect is. Since both the local phase response and its gradient are tailored through metamaterial design, the optical spin-orbit interaction from the metasurface is strong and widely tunable.

\[ \text{Fig. 4.7: Light from a broadband source was focused onto the sample with a lens (f = 50 mm). The polarization can be adjusted in both x- and y- orientations with a half-wave plate. The regularly and the anomalously refracted far-field transmission through the metasurface was collected using a lens (f = 50 mm) and imaged on a CCD camera. The polarization state of the transmission is resolved by using an achromatic quarter-wave plate, a half-wave plate, and a polarizer with high extinction ratio.} \]

To experimentally observe the strong PSHE at the metasurface, we use the polarization-resolved detection setup as shown schematically in fig. 4.7, which allows precise measurement of Stokes parameters of the refracted beam, providing the spin information of the light in the far-field. A super-continuum light source is utilized for broadband measurement and the beam is focused onto the sample with a spot size of \( \sim 50 \mu m \). In the experiment anomalous refracting metasurface that consist of V-shaped gold antennas are utilized. By changing the length and orientation of the arms of the V-shaped structures, the subwavelength antennas at resonance introduce tunable phase retardations between the incident and the forward propagating fields. Eight antennas with optimized geometry parameters are chosen for linear phase retardation along the x-direction. Samples with different phase gradients were measured. The anomalously refracted far-field transmission through the metasurface is collected using a 50 mm lens and imaged on an InGaAs CCD camera. The polarization of the incidence is linear and can be adjusted in either x- or y- orientations with a half-wave plate. The polarization state of the anomalously and the regularly refracted beams are resolved at far-field using an achromatic quarter-wave plate, a half- wave plate, and a polarizer with an extinction ratio greater than \( 10^5 \) for all the wavelengths of interest.
Since the polarization state of light can be unambiguously determined by the Stokes vector $S$ on a unit Poincaré sphere, the evolution of the polarization due to spin-orbit interaction is well described by the precession of Stokes vector on a Poincaré sphere. The helicity or the handedness of light is uniquely given by the circular Stokes $S_z$ parameter (circular polarization): $S_z = (I_{\sigma^+} - I_{\sigma^-})/(I_{\sigma^+} + I_{\sigma^-})$, where $I_{\sigma^+}$ and $I_{\sigma^-}$ are the intensities of the anomalously refraction with circular polarization basis, which were imaged successively using the CCD camera. The coordinates of the image represent the in-plane wavevectors of the refracted beams. The right circular $\sigma^+$ and left circular $\sigma^-$ polarizations are discriminated by setting appropriately the angle of the wave-plates and polarization analyzer. With the polarization resolved detection, the resulted circular Stokes parameter of the anomalously refracted beam was calculated from measurements for each pixel and shown in fig. 4.8 (a) for $x$-polarized incidence. The in-plane wave-vector dependence of the circular Stokes parameters shows a maximal value of $\sim 0.1$ located at $\pm \pi/2$ and the sign of $S_z$ is reversed between $+\pi/2$ and $-\pi/2$, showing a transverse motion of the polarization. While the incident angle is kept at zero throughout the experiment, the phase gradient along the interface removes the axial symmetry of the optical system and enables the direct observation of PSHE for the skewing-refracting anomalous beams with different circular polarized light. Figure 4.8 (b) shows the PSHE effect for $y$-polarized incidence, the helicity of the refracted beam is clearly inverted due to the 90-degree phase rotation of the incidence. As the transverse spin current is solely determined by the longitudinal components of electromagnetic fields, such a unique PSHE can only be observed in the anomalously refracted beam. As a control experiment, the spin-orbit coupling vanishes for the regularly transmitted beam, which exits in the direction of surface normal of the metasurface.

![Fig. 4.8: (a) Observation of PSHE: The helicity of the anomalously refracted beam. The incidence is from the silicon side onto the metasurface and the polarized in $x$-direction along the phase gradient. The incidence angle is zero degree. The pseudo color red and blue represent right- and left-circularly polarizations. (b) The helicity of the refracted beam from a $y$-direction polarized input beam.](image-url)
4.2.2 Trajectory shift induced by SHE
The photonic spin-orbit interaction in a curved light path not only manifests a helicity-dependent circular birefringence but also influences the trajectory of light. This effect resembles the Imbert-Fedorov shifts in the case of total internal reflections\cite{76} and the recently observed optical Magnus effect\cite{77,78}. When considering spin-orbit coupling, the Fermat principle based on ray optics is not sufficient in accurately depicting the light trajectory and a polarization-dependent transverse motion of light has to be incorporated to compensate the geometric phases. For the incident with a pure spin state (circularly polarized), a transverse motion of beam’s center of mass is therefore expected.

![Fig. 4.9: (a) The wavelength dependences of the refraction angles when excitation is normally incident onto the metasurface. The measurement was conducted for three samples with different phase gradients, 3.6, 4.0, and 4.4 rad/mm, respectively. The measured refraction angles agree well with the theoretical predictions (solid curves). The inset schematically shows the light trajectory for the anomalous refraction for the surface normal incidence. (b) The transverse motion between the weight centers of the right- and left-circular polarized refractions, showing anomalous PSHE effect over a broad range of wavelengths.](image)

Such a transverse motion is in real space and can be directly measured. Replacing the polarization optics shown in fig. 4.7 with a variable liquid-crystal phase retarder, the spin-dependent motion of photons (the relative displacement between the anomalously refracted $I_{\sigma^+}$ and $I_{\sigma^-}$) was then measured by an InGaAs quadrant detector with polarized incident beam periodically modulated in either left- $\sigma_+$ and right-handed $\sigma_-$ polarization states. Throughout the experiment the incidence angle was kept at normal incidence to the sample. The tailored in-plane phase gradient induces anomalous transmission at different refracting angles determined by the gradient. Figure 4.9 (a) shows the refraction angle at different incident wavelength for multiple samples with different phase gradients. The anomalous refraction angle approaches to 90 degrees when the incident wavelength approaching to eight times of the period. In fig. 4.9 (b), the transverse motions of the beams are shown at the normal incident for wavelengths over hundreds nanometer bandwidth. The strong spin-orbit interactions induced by the resonant V-shaped antenna enable the observation of the transverse displacement that diverges as the anomalous refraction angle approaches 90 degrees. Current PSHE theory assumes the
conserved momentum solely over the Gaussian wave-packet. However, the rapid phase gradient along the metasurface supplies additional momentum, making such a theoretical treatment incomplete.

4.3 Numerical simulation

By considering the energy transport at the interface, the polarization-dependent transverse motion of light can be analyzed by integrating the Poynting vectors (including the evanescent fields) over the half-space of the exiting medium. Any transverse motion of light is captured by the transverse components of Poynting vectors. Integrating the Poynting vectors (including the evanescent fields) over the full half-space above the metasurface allows us to evaluate the overall transverse displacement for a transmitted beam with finite size,

$$\Delta = \frac{1}{S_z} \int_0^\infty S_y(z)dz$$  \hspace{1cm} (3.4)

Here $S_y$ denotes the Poynting vector component that is perpendicular to the plane of incidence (i.e., along transverse $y$-direction in fig. 4.6 (b)). $S_z$ is the $z$ (surface normal) component of the Poynting vector of the transmitted beam at the interface ($z = 0$). The procedure is general and applicable to both transmitted and reflected fields. A similar approach has been employed to analyze the Göös-Hänchen effect$^{79}$, a longitudinal beam displacement occurs at Total Internal Reflection, by integrating the longitudinal components of Poynting vector $\delta_{GH} = \frac{1}{S_z} \int_0^\infty S_x(z)dz$

Fig. 4.10: Simulated PSHE of circularly polarized light for a metasurface with a phase gradient of 4.4 rad/µm.

To evaluate the effect numerically we use the rigorous Fourier modal method for solving the Poynting vectors of the transmitted fields$^{80}$. Figure 4.10 shows the numerically integrated PSHE shifts for a metasurface with a phase gradient of 4.4 rad/µm. The
incident polarization is circular and the result agrees well with the experiments. The finite sized Gaussian beam is simulated by the superposition of a set of plane waves (11×11 plane waves) spanning in both polar and azimuthal directions,

\[ E(r) = \int_{-\infty}^{\infty} e(k)E_0 \exp\left(-\frac{k_\parallel^2}{\sigma^2_k}\right)e^{-ik_\parallel r} dr^3 \]  

(3.5)

Here \( e(k) \) is the electrical field vector of each individual plane wave, and \( E_0 \) is the amplitude of the incident Gaussian wavepacket. \( k_\parallel \) is the transverse components of the wave vector \( k \). \( \sigma_k \) measures the angular spectral width of the beam and a 4-degree spectral width is chosen for the calculations shown in fig. 4.10. For each individual plane wave, sufficiently large numbers of Fourier orders (21×21) were retained in the simulation to ensure the convergence.

4.4 Conclusion

The strong photonic SHE at a metasurface with a designed phase discontinuity over the wave-length scale enables the observation of transverse motions of circularly polarized light. The anomalous skew scattering of light simultaneously breaks the rotational symmetry in the polarization space and the axial symmetry along its trajectory, giving rise to a broadband PSHE, even at the normal incidence. The generation and manipulation of strong spin-orbit interaction of light with tailored nanomaterials provide a new degree of freedom in information transfer between spin and orbital angular momentum of photons.
Chapter 5

Imaging the optical confinement effect in the hybrid waveguide

In the emerging field of nano-photonics, the manipulation of light on scales much smaller than the wavelength is involved. However, very few feasible approaches exist so far. Surface plasmon polaritons (SPP), the optical excitation at the interfaces between metal and dielectric, is one of the most promising candidates for deep sub-diffraction-limited optical confinement. Nevertheless, experiments have only been successful in demonstrating moderate optical confinement, which is comparable to that of conventional dielectric waveguides due to prohibitive fabrication demands and optical losses. In this chapter, we demonstrate a new approach combining the conventional waveguide with plasmonics to address the trade-off between strong field confinement and propagation distance. By fine-tuning the coupling between the photonic waveguide mode and the surface plasma mode, we show that “capacitor-like” energy storage allows deep sub-wavelength confinement in the dielectric gap, thus dramatically reducing the metallic loss.

5.1 Introduction to the hybrid waveguide theory

Photonics has become the key driver in global data communications. The ever growing demand for higher data bandwidth and lower power consumption of photonic devices have set a roadmap for reducing the physical photonic component size down to the nanoscale beyond the diffraction limit of light with increasing functionality. While various compact technologies have been developed for reducing the physical size, the optical mode confinement offers enhanced light matter interactions towards low driving power and fast modulation speed. Optical confinement of conventional photonic components is restricted by the diffraction limit of the light resulting in weak light-matter interaction that often demands long device sizes to increase the interaction. For example, linear and ring resonator often require lengths of 10-1000 microns, which leads to undesirable large footprints, thermal instabilities and high bending induced radiation losses upon downscaling ring resonators. Surface plasmon polaritons (SPPs), collective oscillations of electrons at metal-dielectric interfaces, were proposed as a potential solution for nanoscale photonics since their wavelength can be scaled down below diffraction limit. However, the direct experimental demonstration in realizing low-loss propagation of deep sub-wavelength optical modes has not been realized due to a rapid increase in the metal loss upon scaling down the optical mode, which pushes the electromagnetic field into the metal.
A hybrid plasmon polariton concept has been proposed to overcome this challenge\textsuperscript{100}. This approach uses a high-k, semiconductor dielectric strip separated from a metal surface by a nanoscale low-k gap. Since the hybridized plasmon energy is concentrated in the low loss, low-k gap, this novel method offers ultra-small mode confinement ($I^2/400$) over a broad range of frequencies and, simultaneously allows for reduced optical loss compared to metal-semiconductor interface design\textsuperscript{101}.

![Fig. 5.1: A typical hybrid waveguide geometry. The plasmonic mode at the metal surface is hybridized with the photonic mode in the dielectric rod, providing a new mode of small mode volume and long propagation length. Over half of the optical energy is concentrated in the nanometer sized dielectric gap between the waveguide and the metal layer. Adapted from Ref. 100.](image)

The hybrid waveguide geometry, shown in fig. 5.1, consists of a high permittivity semiconductor cylinder ("cylinder waveguide") embedded in a low-permittivity dielectric near a metal-dielectric interface ("SPP waveguide"). In the following study, we vary the cylinder diameter, \(d\), and dielectric gap width, \(h\), to control the propagation distance, \(L_m\), mode area, \(A_m\) and electromagnetic field distribution of a single hybrid mode.

Figures 5.2(a) and 5.2(b) present numerical calculations of propagation distance and normalized mode area, respectively, as a function of \(d\) and \(h\). For large cylinder diameter and gap width (\(d > 200\) nm, \(h > 20\) nm) the hybrid waveguide supports a low loss cylinder-like mode with electromagnetic energy confined to the high-permittivity dielectric core (Fig. 5.2(c)). Conversely, a small diameter cylinder (\(d < 200\) nm) results in an SPP-like mode, localized mainly to the metal-dielectric interface and suffering the typical loss of an SPP. At moderate cylinder diameters (\(d \sim 200\)nm) the coupling between the underlying modes forms a hybrid mode that features both cylinder and SPP characteristics; namely, its electromagnetic energy is distributed over both the cylinder and adjacent metal-dielectric interface (fig 5.2(d)). However, when the gap width is reduced towards the nanometer scale, the hybrid mode no longer displays characteristics of either the cylinder or the SPP mode; instead, it is strongly confined (figs. 5.2(e), (f)). Despite this strong confinement, the hybrid mode’s propagation length exceeds that of SPPs of the equivalent metal-semiconductor interface (lower broken line in Fig 5.2(a)). Evidently, the gap width provides significant control over the character of the hybrid mode and provides the
means to store electromagnetic energy leading to nano-scale optical fields with low mode loss.

Fig. 5.2: (a) Propagation length as a function of cylinder diameter, \( d \) and gap, \( h \). The upper and lower horizontal broken lines represent the propagation lengths for the uncoupled metal/oxide and metal/semiconductor interfaces respectively (i.e. without cylinder). (b) Effective mode area, \( A_m / A_0 \) as a function of cylinder diameter, \( d \) and gap width, \( h \). The cusp features occur when the position of maximum \( W(r) \) shifts from the center of the cylinder to its edge nearest the metal. Lower panels show electromagnetic energy density distributions for \( c, [d, h] = [400, 100] \) nm; \( d, [d, h] = [200, 100] \) nm; \( e, [d, h] = [200, 2] \) nm; and \( f, [d,h] = [400, 2] \) nm. Adapted from Ref. 100.

Fig. 5.3: (a) Normalized energy density along the \( y \)-axis, showing the confinement in the low index dielectric region (no shading). The broken line in the left inset is \( x=0 \). The metal and semiconductor cylinder regions are shaded gray and green respectively. (b) Normalized energy density along the \( x \)-axis between metal and host dielectric. The broken line in the right inset is \( y = -d/2 \). The left inset shows the FWHM of the energy
density distribution in the x-direction and the normalized mode area, $A_m / A_0$, as a function of $h$. Furthermore, the near linear increase in gap energy density (reciprocal of $A_m / A_0$) with decreasing $h$ is an indicative of capacitive energy storage of this geometry (inset b). Adapted from Ref. 100.

On first inspection, the confinement appears to be in the metal; however, the energy density distribution in fig. 5.2(d) indicates that a significant portion of energy resides in the low permittivity gap between the cylinder and the metal-dielectric interface. Figure 5.3(a) resolves the electromagnetic energy density along the y axis for x = 0 showing that the peak energy density is in the low index gap between cylinder and metal. Strong sub-wavelength confinement is also evident along the x direction (Fig. 5.3(b)). Further calculations (not shown here) indicate that at least 30% of the mode’s energy resides in the gap for $h = 2$ nm (Fig. 5.2(e)) with more than 50% for $h = 100$ nm (Fig. 5.2(d)).

The strong energy density in the gap region occurs for two reasons: firstly, it is directly related to the continuity of the displacement field at the material boundaries, which leads to a strong normal electric field component in the gap; and secondly, in both the uncoupled SPP and cylinder geometries, the electric field components normal to material interfaces are dominant, amplifying the first effect. Physically, this corresponds to energy storage by polarization and plasma oscillations along the semiconductor/oxide and the metal/oxide interfaces respectively; i.e. the gap region has an effective optical capacitance (Fig 5.3). Note that a higher contrast in dielectric permittivity contrast with optimized geometrical parameters can enhance the discontinuity of the uncoupled cylinder’s radial electric field component and the hybrid mode confinement.

5.2 Experimental demonstration and discussion

5.2.1 Nearfield characterization

In order to directly visualize the highly confined mode within the dielectric gap, we applied our a-NSOM to map the optical intensity profile within the waveguide. The experimental scheme is illustrated in fig. 5.4.

Fig. 5.4: Schematic of HPP mode size mapping via apertureless NSOM. The HPP strips consist of a semiconductor (high-k) separated from a metallic surface by a nanometer scale low-k gap. They are excited by illuminating the slit at the input end.
Experimentally, we replace the cylindrical waveguide with a square waveguide for the convenience of fabrication. Except that the field enhance peak is different at the apex point, the majority of the physical effect discussed in the above section is approximately the same here. The lithographically defined hybrid plasmon mode strips are prepared by evaporating a high index semiconductor, ZnS, \( n = 2.2 \) onto an Ag film separated by a thin, \( h = 10 \) nm, MgF\(_2\) film, both on a quartz substrate followed by lift off in Acetone. The rough-mean-square (RMS) of the Ag film is about 1.0 - 1.5 nm, measured via atomic-force-microscopy (AFM). The fabricated HPP-based strips have lateral dimensions of about 200 nm in height, \( H \), with varying width, \( W \) (150 - 800 nm). The strip input is designed to be tapered in order to act as an optical funnel to increase the signal strength and to reduce the impedance mismatch between free space and the HPP mode (Inset fig. 5.5). As the final step, focused-ion-beam (FIB) milling creates the input and the 45 degree angled output port to enable the scanning NSOM tip to access the HPP mode’s field profile.

![Figure 1](image1.png)

Fig. 5.5: 3D image overlap of the deep sub-wavelength HPP mode signal (red spot) offering optical confinement significantly below the diffraction limit of light. Scale bar = 125 nm. MgF\(_2\) gap \( h = 10 \) nm, illumination wavelength \( l = 633 \) nm. Inset: Height profile of tapered strip for free-space to HPP strip coupling, scale bar = 1 mm. Focused ion beam (FIB) etching was used to define the illumination-port (IN) and the access point for the NSOM probe of the confined optical mode (OUT).

Figure 5.5 shows a topological AFM scan of a 1 mm\(^2\) region around the output of such a strip superimposed with the NSOM image acquired simultaneously at an illumination wavelength of 633 nm. It can be seen that the optical confinement is indeed as small as about 50 nm (Full Width at Half Maximum, FWHM) and is situated at the low-k dielectric gap region directly confirming the HPP mode.

The HPP strips are excited by illuminating a metal slit at the input (IN) end of the strip (inset of fig. 5.5) from the substrate side with polarization perpendicular to the slit. Since our nearfield imaging technique is mostly sensitive to electric fields along the axis of the NSOM detection tip, the z-component of the HPP mode is imaged, which is expected to be strongest in the gap region. To gain access to the internal
fields, we sliced the strip open, allowing the NSOM tip to probe the HPP mode’s cross-section directly. When the polarization of illumination is parallel to the metal slit, HPP modes are not excited in the strip and only a background signal was measured by NSOM.

![Graph](image)

*Fig. 5.6: Line cross section of the nearfield intensity along the horizontal broken line (solid line). Red dashed line is the cross section in the numerical simulation of the same structure. Black dashed line stands for the simulated cross section at 10 nm away from the surface.*

We have compared line-scans of the NSOM field image (broken white lines in fig 5.5) with numerical simulations. Taking into account the HPP mode’s field diffraction along the z-direction, we were able to correlate experiment and theory well at tip-sample separations of about 10 nm (fig 5.6). A similar correlation is found for the mode-width data along the y-direction. The optical mode confinement in the y-direction is controllable by the lateral HPP strip width, \( W \), with the smallest measured mode area, \( A_{\text{HPP}} = 53 \text{ nm} \times 63 \text{ nm} \), proving a non-resonant, deep sub-wavelength (< \( 1/12, l = 633 \text{ nm} \)) nano-scale mode.

![Graph](image)

*Fig. 5.7: Summary of the optical confinement along the horizontal and vertical direction in the hybrid waveguides of different widths.*

Fig 5.7 summarizes this property by comparing the FWHM of the HPP mode for varying HPP strip widths. Whereas the mode height in the z-direction remains essentially constant, the mode width follows the strip width. While deep sub-wavelength mode sizes are expected in the z-direction due to the involvement of
plasmons, lateral confinement can also be strongly sub-wavelength. This arises from the accumulation of polarization surface charges concentrated around the center (y-direction) of the HPP strip. Furthermore, we find the hybrid plasmon mode’s optical confinement to be relatively insensitive to the illumination wavelength. We demonstrate this broadband feature by illuminating the HPP strips with visible and near-infrared light (l = 633, 808 and 1427 nm) and find that the vertical mode confinement remains essentially constant at about 60 ~ 70 nm (fig 5.8 (a)). Since the chosen HPP strip height, H, is optimized for an operation wavelength of 633 nm, the subtle mode height increase from 60 nm to 70 nm is expected for changing the illumination wavelength from visible to near-infrared. These remarkably small mode sizes are among the smallest for propagating electromagnetic waves demonstrated to date.

![Graph showing signal E^2 vs. wavelength for VIS, NIR, and IR](image)

*Fig. 5.8: Increasing the wavelength of the illumination beam to 808 nm and 1427 nm results in mode height FWHM of 63 nm and 74 nm respectively, featuring broadband, deep sub-wavelength operation of HPP-based devices. Solid lines are Gaussian fits.***

### 5.2.1 Farfield characterization

In addition to providing strong optical confinement, another advantage of the HPP mode over other non-resonant plasmonic schemes is its reduced optical loss by allocating most of optical energy in the dielectric gap instead inside metal. We have measured the transmission through HPP strips of varying length under white light illumination (fig 5.9). The sub-wavelength confined HPP mode propagates more than 10 times its vacuum wavelength at visible wavelengths near 633 nm, which is more than six times further than plasmonic control strips consisting of a metal-semiconductor interface without the low index gap region (fig 5.9(b)). The propagation length increases further with increasing wavelength, and exceeds 20 times its vacuum wavelength at near infrared (Inset fig 5.9(b)), thus allowing for sufficiently long interaction lengths, which, combined with strong optical confinement, can create strong light-matter-interaction effects for active photonic components.
Fabry–Perot interference spectroscopy and using NSOM imaging. We consider the dispersion of the HPP mode, enabling compact and large data bandwidth densities (Supplementary Fig. S3). For the latter, the optical confinement, another advantage of the HPP mode, over other non-metals, can create strong light-matter interaction effects for active control (black dashed line) confirming the hybrid nature of the HPP mode.

In conclusion, we directly demonstrated waveguiding of ultra-wideband light at moderate data bandwidth densities. For the latter, the optical confinement, another advantage of the HPP mode, over other non-metals, can create strong light-matter interaction effects for active control (black dashed line) confirming the hybrid nature of the HPP mode. The abrupt change in the effective refractive index at the end of the HPP strip acts as a partial reflector for the HPP modes. The resulting optical cavity displays Fabry–Perot interference fringes, corresponding to longitudinal cavity modes. The group index can be determined from the spectral mode spacing, \( \Delta \lambda = \left( \frac{\lambda^2}{2n_gL} \right) \), where \( L \) is the strip length, \( \lambda \) is the wavelength, \( n_g = n - \frac{dn}{d\lambda} \cdot \lambda \), the group index and \( n \) is the effective mode index. As a result, we can measure the group index of HPP

**Fig. 5.9:** (a) Far field images of illuminated HPP strips showing incident white light reflected from the strip input and the corresponding emission from the distal strip output facet. The strips lengths, \( L \), are indicated on top of each image and the dashed line is a guide the eye to the output signal. (b) The HPP mode offers propagation lengths of more than 10 times its free space wavelength at \( \lambda = 633 \) nm. Without the low-k gap-layer (plasmonic control) the propagation length is only about 1 \( \mu \)m (black dashed line). Inset: the HPP mode's propagation length increases with its operating wavelength and exceeds 20 times its vacuum wavelength at near infrared wavelength (\( \lambda = 800 \) nm).
waveguide by measuring the transmission spectrum of the waveguide. The results are shown in fig. 5.10(b).

![Illustration of the excitation and collection setup used to characterize the transmission of HPP waveguide.](image)

*Fig. 5.10: (a) Illustration of the excitation and collection setup used to characterize the transmission of HPP waveguide. (b) Group index dispersion of the HPP waveguide retrieved from the Fabry-Perot resonance in the transmission spectrum of a series of waveguides with different lengths. The red dashed line and the back broken line correspond to two extreme cases, where $h \to \infty$ and $h \to 0$, respectively.*

The increasing group index with photon energy confirms the dispersive, plasmonic mode character of the HPP mode, which manifests itself in the deep sub-wavelength confinement (fig 5.10). The dispersion of the HPP relation lies between the two extreme cases of gap width, $h$: namely, a semiconductor strip ($h \to \infty$) and a semiconductor strip in direct contact with the silver film ($h \to 0$). This demonstrates controlled hybridization of the modes of a semiconductor strip and a metal-dielectric interface.

### 5.3 Conclusion

Plasmonic waveguides can be deployed in two distinct application areas; (1) routing information passively and (2) actively altering optical signals. For the latter, the optical confinement can strengthen light-matter-interactions, thereby reducing the required devices size to utilize specific effect (e.g. a 2nd or 3rd order non-linearity). The trade-off lies in enhancing the effect, i.e. mode confinement, and incurring loss over the designed device length providing the effect. With this in mind, deep-subwavelength metal optics hold promise to produce a significant nonlinear enhancement, while maintaining sufficient propagation distances. In addition, hybrid plasmonic waveguides allow for an increased packing density of photonic components as the HPP mode strongly suppresses the waveguide crosstalk compared to dielectric waveguides.
In conclusion, we directly demonstrated waveguiding of ultra-small propagating waves at visible and near infrared frequencies via NSOM imaging. We confirm the appropriate optical field hybridization between a surface plasmon and a dielectric mode by Fabry-Perot interference spectroscopy and find a dispersive group index as a result. This HPP concept has reduced metal losses offering reasonably long light-matter interaction lengths, potentially enabling compact and efficient nano-scale photonic components, since it elegantly interfaces plasmonics with semiconductors. This novel mode design holds a great potential for truly nanoscale photonic applications such as intra-chip optical communication\textsuperscript{102}, signal modulation\textsuperscript{103}, nanoscale lasers\textsuperscript{104} and bio-sensing\textsuperscript{105}. 

Chapter 6

Transition Metal Dichalcogenide (TMDC): structures, properties, characterizations and fabrications

The success of graphene research has inspired the researchers to seek opportunities in other 2D crystals. Among many, few layer transition metal dichalcogenide (TMDC) has recently become another research focus, thanks to its interesting electronic and optical properties, which can potentially generate future applications beyond the graphene. For example, the direct bandgap at the visible frequency in TMDC has been a highly desired yet still missing property in graphene. In this chapter, we will first give an introduction to the atomic structure, electronic structure, the optical property of the materials. Following that, we will discuss the foundations of TMDC study: how to prepare the materials; the transfer method as well as the fabrication methods. The device performance and experimental results will be presented in the following chapters. In this chapter, I will focus on Molybdendum disulfide (MoS₂) as an example, though other TMDC might be utilized in the following chapters.

6.1 Introduction to few layer MoS₂

6.1.1 Atomic structure

TMDC is a class of materials with the formula MX₂, where M is a transition metal element from group IV, group V or group VI, and X is a chalcogen (S, Se or Te). These elements form layered structures of the form X-M-X, with the transition metal in a hexagonal plane sandwiched by two chalcogen hexagonal atomic planes (fig. 6.1 (a) - (d), yellow spheres are the sulfide atom and purple spheres are the molybdenum atom). Most natural MoS₂ bulk crystals form so-called 2H stacking, which is of hexagonal symmetry and has two effective layers in the repeating unit (fig. (c) & (d)). The transition metal atom at the middle layer has a trigonal prismatic coordination and the inversion symmetry is kept in the bulk crystal.

On the other hand, as can be seen in the fig. 6.1, the in-plane lattice constant is much smaller than the inter layer spacing (a,b = 3.14 Å, c = 6.5 Å). The in-plane bonding in MoS is much stronger than the interlayer bonding, which is a type of weak Van der Waals bonding. Therefore, it’s possible to break the repeating unit cell to reduce the crystal into a single effective sandwich layer by mechanical exfoliation, as how graphene is discovered. The resulting structure is shown in fig. (a) & (b). The crystal group is reduced from D₆h to D₃h and the inversion symmetry is broken. This is very important to some physical properties such as the second harmonic generation and the valley Hall effect discussed in the following.
6.1.2 Electronic structure and optical properties

The bulk MoS$_2$ crystal is an indirect bandgap semiconductor, with a 1.2 eV band gap across the K point of the valence band and the middle point of the conduction band between $K$ and $\Gamma$. However, as the material is gradually reduced to its monolayer form, the band gap progressively shifts to a direct band gap across the K point.\textsuperscript{16,17} The change in the band structure with layer number is due to quantum confinement and the resulting change in hybridization between $p_z$ orbitals on S atoms and $d$ orbitals on Mo atoms. For MoS$_2$, density functional theory (DFT) calculations show that the conduction-band states at the K-point are mainly due to localized $d$ orbitals on the Mo atoms, located in the middle of the S–Mo–S layer sandwiches and relatively unaffected by interlayer coupling. However, the states near the $\Gamma$-point are due to combinations of the antibonding $p_z$-orbitals on the S atoms and the $d$ orbitals on Mo atoms, and have a strong interlayer coupling effect. Therefore, as the layer numbers change, the direct excitonic states near the K-point are relatively unchanged, but the transition at the $\Gamma$-point shift significantly from an indirect one to a larger, direct one.

The DFT calculation with local density approximation (LDA) in fig. 6.2 shows the direct band gap is about 1.9 eV in MoS$_2$.\textsuperscript{16} This is very important to the application of MoS$_2$ monolayer. Firstly, this enables many optoelectronic devices based on the monolayer. The shift to the direct bandgap means the radiative combination is much more efficient than before, and therefore it is more promising to build light emission
device (LED) or even stimulated emission with monolayer MoS$_2$. In addition, the oscillation strength across the direct band gap is also enhanced. As only monolayer thick as it is, the

![Diagram]

Fig. 6.2: The gradual transition of the electronic band structure in MoS$_2$. The band gap becomes a direct one at $K$ point as the material is reduced from few layers to the monolayer form. Adapted from Ref. 16.

Absorption peak in WS$_2$ is measured to be over ten percent, which means the monolayer based photo detector/solar cell will be much more efficient than some bulky devices. Secondly, the large band gap can also help the efficiency of electronic transistor. Owing to the finite value of band gap size, the current within the MoS$_2$ channel is very sensitive the doping. An on/off ratio larger than $10^6$ has been demonstrate in the monolayer MoS$_2$ transistor.$^{107}$

Fig. 6.3 shows the photoluminescence (PL) spectrum of a single layer of MoS$_2$. The monolayer is excited by a 488 nm laser and the emission is collected by the same objective in a back reflection configuration. The spectrum manifests a peak at 680 nm, which is indeed the direct band gap emission. This peak is orders of magnitude lower in the bilayer or any few layer sample thicker than monolayer. It is because the band gap is an indirect one and therefore the quantum efficiency is much lower than the monolayer. This is a very fascinating example showing that the less material we have, the more photoluminescence we can get.

Furthermore, a shoulder at the 630 nm is also observed which is from the direct bandgap transition from the other valence band at $K$ point, evidently in fig. 6.2. The valence is split due to the spin orbit coupling in the Mo, which is previously prohibited by the inversion symmetry in the bulk form. In the monolayer form, the inversion symmetry is broken and the spin orbit coupling becomes unnegligible. In MoS$_2$, this energy is about 140 meV, which is approximately same as the splitting we observe in fig. 6.3. The spin orbit coupling induced splitting is larger in heavier compounds, such as WS$_2$, which can be as large as 400 meV.
Another small feature in the figure 6.3 is the large hump around 800 nm. This is below the smallest energy band gap and therefore cannot come from the direct band gap transition. We carried out a temperature dependent PL experiment and found this peak reaches its maximum at the low temperature, and therefore we expect it to be related to the defect luminescence, in agreement with the literature. 114

![Fig. 6.3: PL spectrum of MoS$_2$ monolayer excited by a 488 nm laser.](image)

Further DFT studies suggest that the strong emission and absorption is not only caused by the direct band gap. 115–118 The exciton effect is also very strong in MoS$_2$ monolayer, i.e., the binding energy of these 2D excitons can be as large as 1 eV. We will discuss the exciton effect in details in Chapter 7.

6.2. Sample preparation and characterizations

6.2.1 Sample preparation and monolayer identification

The sample is mechanically exfoliated in a similar way with graphene. We start with a natural MoS$_2$ crystal (spi), and use the tape to pick up a large flat flake. Then we folded the tape for multiple times so the flakes can be thin down to semi-transparent level. Finally we press the tape against the substrate and gently scrub the tape. After the tape is slowly removed, some small MoS$_2$ flakes will be left on the substrate and a fraction of them are few layers, which we can use multiple ways to identify. Experimentally, we tried different tapes to optimize the yield and the Ultratape with least adhesion performs the best among all. One big difference between MoS$_2$ and graphene is that the MoS$_2$ crystal is more fragile, and therefore the MoS$_2$ monolayer is usually smaller than the exfoliated graphene.

In order to visualize the exfoliated MoS$_2$, we exfoliated the sample on the substrate of silicon oxide on degenerate silicon. Similar to graphene preparation, the multiple reflection from the top and bottom surface of the oxide layer will interfere with each
other and the shift of such an interference due to the presence of TMDC gives an optical contrast. An estimation of the monolayer’s optical contrast based the bulk properties of TMDC has been published. As can be seen, the contrast reaches its maximum at 550 nm, when the silicon oxide is 90 nm or 270 nm thick. And such a contrast is enough to identify under normal wide field microscope.

![Fig. 6.4: Monolayer MoS₂ contrast on silicon oxide of different thickness on heavily doped silicon, at different wavelengths. Adapted from Ref. 119.](image)

![Fig. 6.5: White light image of monolayer MoS₂ on 270 nm SiO₂/Si substrate. A bilayer is also visible in the image, which has about twice the contrast than the monolayer. The insect shows the centimeter scale sample chip with an array of alignment markers.](image)

A typical monolayer image is shown in fig. 6.5. The substrate has a purple color under white illumination and the few layer will show additional purple contrast than the substrate. In the few layer limit (n<=4), the contrast is approximately proportional to the layer thickness. As can be seen, we also use the optical lithography technique to fabricate an array of gold alignment marker with alphabetic labels across the whole centimeter sized chip. This helps to record the few layer position and enables the following fabrication process.

In some situations, we couldn’t use a well-designed SiO₂/Si substrate. For example, in order to measure the transmission spectrum of the monolayer, it has to be situated on a transparent substrate; or the monolayer is prepared on the thick
Polymethyl methacrylate (PMMA) layer for the transfer purpose. Therefore we develop another laser-scanning PL approach to search for the monolayer. It is based on the fact that monolayer emits much more PL than bilayer or even bulk materials. We use a laser scanning confocal microscope (Nikon 3000) to scan the sample. The excitation laser is at 488 nm and the photomultiplier tube (PMT) detects the signal from 640 nm to 800 nm. The PL signal is recorded in correspondence to the scanning mirror position, and a PL map is obtained after a single scanning. If there’s a monolayer in the scanning field of view, the PL map can recognize its location. In addition, we use another PMT to simultaneously detect the scattering of the excitation laser, so that we can include the prepared gold alignment marker into the image. After comparing the two channel maps, the relative location of the monolayer with regards to its closest marker can be obtained, which are used for future processes.

![Image of PL and scattering map](Fig. 6.6: A collage of PL (red) & scattering (green) map of a 1cm x 1cm sample. Each picture is about 200x200 µm large and has 256x256 pixels. Therefore a submicron resolution can be achieved.)

A confocal scan over a centimeter-by-centimeter chip is shown in fig. 6.6. The scattering channel is coded in green and the PL channel is coded in red. Because a single scan has a very limited field of view (200 µm x 200 µm in this case to achieve a good enough resolution), an automated scanning synchronizing program is developed. After the automated scan, we can make a collage of PL map and an analysis program is used to pick out the high PL intensity pixel. Then we can return to the monolayer position and double check the monolayer quality. The whole process takes about 30 minutes. Different band pass filters can be added into PL
collection path, so that different 2D semiconductors with different direct band gap energy can be identified.

### 6.2.2 Monolayer characterizations.

The identification method discussed above is confirmed by the Raman measurement, atomic force microscope (AFM) measurement and PL measurement. As we will show, once the white light image and confocal PL mapping are calibrated by Raman and AFM, we can fully rely on these identification methods.

The Raman feature of MoS$_2$ monolayer has been discussed in the literature. As the fig. 6.7 (a) show, the major distinguishable peaks are $E_{2g}$ and $A_{1g}$ peaks. The former peak stands for the in-plane symmetric oscillation mode between the transition metal and chalcogen atoms; and the latter represents the out of plane optical phonon. Evidently, the out of plane mode is more sensitive to the layer spacing. The smaller layer number there is, the smaller restoration force will be and therefore the oscillation frequency will be reduced in the monolayer. Given the in-plane mode is not very sensitive to the layer number, the difference between $E_{2g}$ and $A_{1g}$ becomes a very good gauge for the layer thickness. As the material is reduced to the monolayer, the frequency difference shifts from 24 cm$^{-1}$ to about 19 cm$^{-1}$.

![Fig. 6.7: (a) literature result of the dependence of $E_{2g}$ & $A_{1g}$ splitting and the layer thickness. (b) Experimental results measured in our lab on three different layers on a SiO2/Si substrate. (c) AFM image and cross section of the three-layer sample.](image)
Our experimental result is show in fig. 6.7 (b). The peak around 390 \text{ cm}^{-1} is the \textit{E}_{2g} peak and the one at 410 \text{ cm}^{-1} is the \textit{A}_{1g} peak. On a sample with mono, bi and trilayer areas, we get the Raman peak splitting dependence. The splitting between the two is about 23 \text{ cm}^{-1} for monolayer, 21 \text{ cm}^{-1} for bilayer and 19 \text{ cm}^{-1} for the monolayer case, which agree well with the literature values. The peak around 525 \text{ cm}^{-1} is the Raman signal from the silicon substrate. As the flake gets thinner, MoS$_2$ peak becomes smaller while the silicon peak becomes stronger, which agrees with the simple transmission-reflection model.

This result is double confirmed by AFM measurement in fig. 6.7 (c). Here, we use a DI (dimension 3000) AFM to resolve the individual layer thickness. Silicon tips working in the tapping mode are employed in the measurement. The step across each layer is about 0.7 nm. The step between the substrate and the first layer is much rougher but still smaller than 1 nm, which indicates a monolayer of MoS$_2$.

### 6.3 Transfer methods

In our lab, we use two different types of transfer methods. One is for the CVD grown MoS$_2$ film and the other is for the exfoliated MoS$_2$ flake.\textsuperscript{121}

For the CVD grown MoS$_2$, we follow the graphene transfer method that we spin coat a layer of PMMA (500nm) on top of the film and soak it in KOH, the oxide underneath the film is slightly etched and the film together with PMMA comes off naturally and float in the KOH solution.

Regarding the exfoliated flake, it is harder to transfer it to a precise location due to the limit of small flake size, so we adopt an approach without any liquid solution. In the dry transfer process, we spin coat a very thick PMMA layer (about 6 \textmu m) on a flat substrate. This extraordinary thickness can be achieved by spinning A10 PMMA at 1500 RMP for two times. After a soft baking at 80 degree, we directly exfoliated MoS$_2$ on top of it. Experimentally, we found that the exfoliation yield is higher than on the SiO$_2$/Si substrate, due to the possible special adhesion between MoS$_2$ and PMMA. Although the MoS$_2$ monolayer no longer has a high visible contrast on such a non-optimized substrate, we can still use the confocal PL method to find the monolayer flake.

After that, we use an up-right microscope equipped with a set of micro-manipulating tool (Ted Pella inc.) to implement the transfer. At first, we use a micro cutter to crop a small PMMA window (50x50 \textmu m) containing a monolayer flake. The surface adhesion between PMMA and SiO$_2$ is weak enough to rip the small PMMA window off the substrate. Then we skew the window onto a micro needle mounted on a micrometer stage. The rotating axis on the stage helps to flip the PMMA by 180 degree, so that the surface with the flake is facing downward. After that, we move the PMMA frame to the destination substrate we want to transfer to. Under the microscope, we can align precisely the flake with target feature. After laying the
frame onto the substrate, we can slowly pull out the needle. Some small adjustment can be done by nudging the PMMA frame after the laying.

A post annealing process (150 °C for 30 minutes) is necessary to fix the flake to the final substrate. After the annealing, the color of the flake is changed due to the disappearance of air gap between the flake and the substrate. Then we can use the acetone to wash off the PMMA frame and flake would be adhered to the substrate solidly.

Similar method has been tried to suspend the flake on a pre-patterned hole arrays in our lab. The challenge is the adhesion surface area is much reduced, and therefore the flake will drift off easily in the PMMA washing process. In order to secure the monolayer in the critical drying process, some metallic clamping structures are tried out.

6.4 Fabrication processes.

In order to build electrical/optoelectrical devices on MoS\textsubscript{2} monolayer, different fabrication processes have been developed. In this section, I will mainly discuss the general lithography and etching technique. The project specific processing will be left to the following chapters.

6.4.1 Electron beam lithography

In our lab, we use the electron beam lithography (EBL) to define the electrode and the etching area, due to the small size of exfoliated MoS\textsubscript{2} based device. Usually the whole fabrication takes three electron beam exposures: one for identifying the monolayer location, one for defining the electrodes and one for defining the etching window.

Firstly, the directly exfoliated sample is soaked in the Isopropanol Alcohol (IPA) at 80 °C for one hour, in order to remove the residual from the exfoliation tape. Then the monolayer area in the sample is imaged by a dedicated EBL machine in the nanolab in UC Berkeley (Crestec CABL-9510CC). This scan is to have a precise measurement of the relative location information between the monolayer and the alignment marker as well as the fine resolution information of the monolayer to be fabricated. The monolayer shows up a high contrast in the reflection type SEM image due to the high electronic conductivity in the material (fig. 6.8). Nevertheless, the contrast between the monolayer, bilayer and bulk is much lower than under the optical microscope.

![Fig. 6.8: SEM image of the monolayer and bilayer of the MoS\textsubscript{2}](image)
Then we spin coat a thin layer of PMMA (A4, molecule weight 495) on the whole chip and follow it by a soft baking at 80 °C. The estimated PMMA thickness is about 300 nm. At this stage, the monolayer completely looses the contrast, in both the optical microscope and the electron microscope. But we can still design the electrode according to our previously obtained SEM image and the white light image. A great attention is paid in order to avoid the bulk crystal and the alignment marker, which can potentially short the circuit. The alignment under EBL is implemented through the prepatterned gold alignment markers. With a good dose test on the dummy sample, a series of electrodes are patterned around the monolayer. Then the sample is developed in MIBK/Acetone (1:3) solution for 30s. After the development, the electrode can be seen clearly under the optical microscope.

![Optical image of a MoS2 Hall device after EBL lithography patterning and development.](image)

**Fig. 6.9:** Optical image of a MoS2 Hall device after EBL lithography patterning and development.

### 6.4.2 Metal contacts

Different metals have been discussed in the literature to form a good electrical contact with MoS2. The early transistor paper shows that the gold contact with annealing or gold/Titanium contact can form very Ohmic contacts with MoS2 channel. The experiments in our lab shows that even though I-V curve is linear between MoS2 and Au/Ti contacts, the contact resistance is still too large, about the same level as the channel resistance. And such a contact resistance is also dependent on the doping level within the channel, which is understood as a result of doping dependent transfer length.

In order to improve the contact resistance, we try a range of metal in place of Ti. At this stage, we find the thermal evaporated Indium contacts have the lowest contact resistance. Indium is chosen here because MoS2 has a very low work function according to the DFT calculation and Indium is known to have the lowest work function among all the stable metals at the ambient condition.

The contacting metal stack is a 15 nm Indium layer followed by 50 nm gold. Both of them are evaporated in a thermal evaporator with a vacuum level of 10⁻⁶ Torr. After
the evaporation, the undesired films are lifted off in Acetone. The final contact resistance is usually about one forth of the channel resistance. An AFM scan image is shown in the figure 6.10.

![AFM image of a Hall device after the metal deposition and lift-off process.](image)

*Fig. 6.10: AFM image of a Hall device after the metal deposition and lift-off process.*

To further get rid of the contact resistance, we also tried to dope the contact area by using organic molecules. The specific molecule we used is phenylethylamine (PEA), which is known to provide n doping to carbon nanotube and graphene.\(^{125, 126}\) Experimentally, we soaked the sample with contact window open into a PEA solution for 10 hours. After a brief rinsing in DI water, the sample is sent to the evaporator for metal evaporation. After the doping, the contact resistance is found to drop by another 2-3 times.

### 6.4.3 Etching

A regular device shape is desired in some experiments and we therefore develop an etching process to shape the irregular exfoliated monolayer flake and CVD grown continuous monolayer. We first tried the CF\(_4\) gas in the plasma-assisted etching. Although the few layer can be etched off efficiently, some fluorite will be left in the etching window as a residual, which has PL in the similar wavelength range as MoS\(_2\) PL. Plus, sometimes the residual might contaminate the channel area as well. Therefore, we change to using a pure Argon gas in the plasma etching. The etching rate is about 30nm/min. After about 30 seconds, the unwanted monolayer can be removed neatly. A device after the etching process is shown below. The device is etched into a Hall bar device and the green line indicates the external boundaries of the etching window. As the confocal PL image (fig. 6.11 (b)) shows, the PL is eliminated except the channel area.

![Fig. 6.11: (a) Optical image of a Hall bar device after the etching. White lines highlight the In/Au contacts. Green lines are the external boundary of the etching windows. (b) Only the PL from MoS\(_2\) channel are left after the etching.](image)
6.5 Conclusion and outlook

In this chapter, we have presented the basic structural electronic and optical properties of TMDC, with an example of MoS$_2$. Other TMDC shows a similar trend, except different bandgap size and the spin orbit coupling strength. Different preparation, identification, characterization techniques are discussed here. The transfer and fabrication processes developed in our lab have also been introduced which serves as a good foundation for the MoS$_2$ based device projects in the following chapters.

Among the topics covered above, one important direction but left untouched is the CVD growth of TMDC. The CVD grown TMDC could potentially extend the current research into large-scale fabrication and bring us a lot more interesting applications. The possibility has been confirmed in large area MoS$_2$ growth at the early stage. 122,127,128 Another important direction is based on the transfer technique. Different heterostructures between TMDC, graphene and other 2D materials can give us many more opportunities, and very pristine sample could be prepare by directly transfer TMDC in between two boron nitride layers. This type of device could potentially elevate the device performance to a new level.
Chapter 7

Dark excitons and nonlinear responses in TMDC

7.1 Introduction to the nonlinear response of monolayer TMDC

Bulk TMDC is a hexagonal lattice crystal, as discussed in the chapter 6, and belongs to $D_{6h} (6/mmm)$ group with inversion symmetry. On the other hand, when it's reduced to the monolayer, the mirror sub-lattice is removed and the point group is changed to $D_{3h}$, with broken inversion symmetry. Such a change has a fundamental impact on the nonlinear optical properties in TMDC: in $D_{6h}$ type crystal, no element in the second nonlinear tensor is nonzero; while in $D_{3h}$ type crystal, all the in-plane $\chi^{(2)}$ elements have finite values:\(^{129,130}\):

$$\chi = \chi^{(2)}_{xx} = \chi^{(2)}_{yy} = -\chi^{(2)}_{xy} = -\chi^{(2)}_{yx}$$ \hspace{1cm} (7.1)

This property is significant since we can utilize it to resolve the crystal orientation, as well as to enhance the contrast between the monolayer and bilayers. Experimentally, we utilize this parametric second harmonic generation (SHG) effect to observe the interference and the annihilation of the nonlinear waves at the boundaries between adjacent crystals. More interestingly, we discovered an enhancement in the second harmonic generation signal from the edge state at certain crystalline edge.

On the other hand, the two-photon absorption effect, a non-parametric third order nonlinear process proportional to the imaginary part of $\chi^{(3)}$, becomes another important process for our monolayer experiment. In comparison with second order nonlinearity, third order nonlinearity, $\chi^{(3)}$, is not significantly different between $D_{3h}$ and $D_{6h}$ groups. Both of them have the same relationship for the in-plane $\chi^{(3)}$ elements:

$$\chi^{(3)}_{xxx} = \chi^{(3)}_{yyy} = \chi^{(3)}_{xxy} + \chi^{(3)}_{xyx} + \chi^{(3)}_{yxx} + \chi^{(3)}_{yx}$$ \hspace{1cm} (7.2)

However, two-photon absorption spectroscopy is a powerful spectroscopic tool for determining the positions of the energy levels that are not connected to the ground by one-photon transition.\(^{14,131}\) In TMDC monolayer the exciton effect has been predicted to play a dominating role due to the reduced dimension effect\(^{115-118}\). In this chapter, I will discuss our work of probing the dark exciton, which is the excited levels of exciton, via the two-photon absorption (TPA) spectroscopy. Our work, together with the first-principle calculation, demonstrate that a very large binding
energy exists in monolayer TMDC, with the direct band gap size much larger than the previous understanding.

### 7.2 TMDC Dark exciton probed by TPA spectroscopy

#### 7.2.1 TMDC exciton and TPA selection rule

Exciton is a bound state formed by electron and hole due to the strong Coulomb interaction between them. Since the discovery of strong photoluminescence enhancement in single layer TMDC, exciton has been expected a very important factor in determining the optical properties for the following two reasons. First of all, the absorption spectrum of monolayer TMDC, unlike that of the traditional bulk direct band gap semiconductor, exhibits very sharp peaks. So far, no step-like feature has been observed that can be associated with the direct band gap edge absorption, an important featuring result of 2D joint density of state. On the other hand, previous researches on nano-materials, such as carbon nanotube, semiconducting quantum dot and conducting polymer, have shown the exciton effect is very prominent in these reduced dimension materials since the dielectric screening is greatly deducted.

A first-principle calculation result on linear absorption spectrum of monolayer MoS$_2$ is depicted in fig. 7.1. In fig. 7.1 (a), the red dashed curve is the calculation result without considering the exciton effect. A clear band edge absorption step shows up at about 2.8 eV. The green solid curve is a calculation result based on GW plus Bethe Salpeter equation (GW-BSE) approach, with exciton taken into consideration. Many peaks show up in the spectrum, indicating the existence of multiple excitons lines. The difference between lowest peak, ‘A’ peak at 1.8 eV, and the step edge is therefore the binding energy, which in this calculation is about 1 eV. ‘A’ and ‘B’ have the double exciton peak character are due to the spin orbit coupling induced valence band splitting. In addition to them, there’s another large absorption peak at 2.5 eV, labeled as ‘C’ peak, which is mainly from the direct band gap exciton at Γ point. All the other peaks such as ‘A’ and ‘B’ peaks are the excited excitons peaks features.

![Fig. 7.1: First principle calculation of the linear absorption spectrum of the monolayer MoS$_2$ with (green), and without (red) considering the exciton effect.](image)

Black dashed line marks the emission energy. (a) Zero temperature result. (b) Finite temperature result with a large broadening. Adapted from Ref. 117.
The main reason that neither the clear band edge absorption nor the distinct exciton feature can be seen is due to the large overlap between these features and the broadening at finite temperature. (fig. 7.1 (b)) Nothing but a large slow hump can be observed around the 2.5 eV in the absorption spectrum, as predicted in fig. 7.1 (b) and confirmed by the experiments. Experimentally, only ‘A’ and ‘B’ exciton peaks have been clearly resolved so far.

In order to probe these excited exciton features and reveal the real band edge position, we use the two-photon spectroscopy technique, which is a very powerful tool that has been exploited to detect the exciton effect, from the bulk semiconductor to nanomaterials such as two dimensional quantum well and one dimensional carbon nanotube. The physical mechanism is described as the following: with the centripetal potential, the electron hole pair forms a series of Rydberg states with definite parity, like ‘A’ and ‘B’ peaks in fig. 7.1 (a). One-photon transition can only reach the final states with even parity, while the two-photon transition will reach the final states of odd parity. As a result, the selection rule enables the two-photon transition to access the states complimentary to the one-photon transition. The difference measured between the excited energy levels and ground energy level of exciton is a good gauge of the exciton binding strength. Because the two-photon measurement can largely suppress the one-photon transition background from other reciprocal points, we can more easily observe these discrete two-photon features.

The situation is more complicated in TMDC monolayer because the inversion symmetry is broken in the odd layers, and in other words, parity is not a good quantum number. But just as GaAs quantum well can be approximated into $D_{4h}$ group with inversion symmetry at the center of the reciprocal space, if we are only concerned about the transition within a single valley and ignore the spin degree freedom, we can assume the point group to be $D_{6h}$ group. The parity definition can be approximately restored, and so is the selection rule contrast. To note, the selection rule here is different from the valley selection rule involving circular polarized light, but the combination of them is worth further exploration.

In our experiments, instead of directly measuring the TPA spectrum, we monitor the TPA induced luminescence (TPL). (fig. 7.2) Physically, we scan the excitation laser around the half of the exciton absorption energy. Given the strong electric field during the 100s femto-second laser pulse, a portion of the valence electron will simultaneously absorb two photon, the energy of which we define here as the excitation energy, and excited to a new level, forming an exciton pair with the hole left behind. The Exciton can have more energy than the ground exciton level, given the specific excitation energy level. In analogy with the hydrogen atom model, we call these states as 2p, 3p, 4p... They are named in the ‘p’ series because their envelope functions have to be of odd in symmetry so that the total exciton wavefunction is even. To note, even though we use the conventional definition in hydrogen model, the energy series can be very different from the hydrogenic series.
After these excited excitons relax back to the exciton ground level (1s state), they would annihilate themselves by emitting a photon, of which the energy is defined to be the emission energy. If we assume the relaxation and emission process have the same quantum efficiency, independent of the excitation energy, a larger emission intensity means a larger two-photon absorption cross section. Therefore, by monitoring TPL strength at different excitation energy, we can obtain TPA spectrum.

![Diagram of the two-photon absorption and luminescence process after the relaxation. Blue arrows represent the excitation process while the red arrow means the emission process. The green lines are the excited exciton states and the red line is the ground exciton state.]

As an intrinsically weak nonlinear process, the measurement requires a high quantum efficiency in the emission. We choose to work with the exfoliated WS$_2$ flakes from a CVD-synthesized crystal, which has less defects and impurity doping compared with the natural MoS$_2$ crystal. The monolayer WS$_2$ flakes is directly exfoliated onto the transparent quartz substrate, in order to avoid any interference effect. As introduced in the chapter 6.2, the monolayer is identified with a confocal photoluminescence microscope and characterized respectively with tools including AFM and Raman spectroscopy.

### 7.2.2 One-photon and two-photon absorption spectrum

Firstly, we start from understanding the ground exciton levels by measuring the linear transmission spectrum of the monolayer WS$_2$ (fig. 7.3). The absorption peaks at 610 and 510 nm are recognized as the excitonic transitions across the direct band gap at K points, usually defined as A and B excitons. The splitting between them, which is about 400meV, is caused by the strong spin orbital coupling induced valence band splitting, which is in good agreement with the photoluminescence measurement. As can be seen, the excitonic transition greatly enhances the light matter interaction within the material, resulting in an absorption of more than 10 per cent in three layers of atoms, at the A exciton resonance peak. Apart from them, there's a new peak at 470nm, which we call C exciton, and has been attributed to the excitonic direct bandgap transition at Γ point.
In TPL experiment, we scan the two-photon excitation energy from just below the A exciton energy to more than half eV higher than that, so that any excited exciton will be fully captured by the experiments, as shown in fig 7.1. In order to scan all the interesting TPL feature with a span of 300 nm (0.6eV) we utilize the tunable output beam of a Ti:sapphire laser pumped optical parametric oscillator. (Inspire HF 100) The spectrum resolution is about 10 meV. The laser is focused onto a sample through a 100x objective lens with a numerical aperture of 0.9 and emission signal is collected in the back scattering configuration and analyzed by a spectrometer with cooled CCD. The transmissivity of the optical system has been carefully calibrated in the experiment in order to get the absolute power level at the objective focus plane. (fig. 7.4 (b))

**Fig. 7.3: Linear transmission spectrum of monolayer WS\(_2\) on quartz substrate.**

**Fig. 7.4: (a) An emission spectrum of WS\(_2\). The excitation laser wavelength is 1000 nm. Pulse duration is about 200 fs. The power level is about 10 KW / cm\(^2\). The spectrum is taken at 10K. (b) Transmissivity of the microscopy system. The strong dispersion is mainly caused by the high NA objective.**
A typical emission spectrum excited by an ultrafast laser excitation with a wavelength of 500 nm is shown in fig. 7.4 (a). The strong emission peak at 500 nm is the second harmonic generation (SHG) signal, which will be discussed in detail in the following section. More interestingly, we observe the emissions at 610 and 620 nm, which are even higher in energy than the excitation energy. These florescent peaks cannot be induced by anything other than the two-photo absorption process. The two peaks have been known as the exciton and trion emission peaks, with a splitting of 30 meV, which can be clearly resolved at 10K. Both of them are good indicators of the two-photon absorption strength, since the partition between them are not relevant to the excitation energy level, to be shown in fig. 7.6. The two-photon absorption induced florescence is further confirmed by checking the power dependence of the excitation and emission power levels. (fig. 7.5)

![Power dependence of SHG (red) and TPL (blue) signals. The SHG signal has a good quadratic power dependence while the TPL signal has a clear saturation behavior at the high power level.](image)

Both SHG and TPL signals are plotted in the same log-log plot in fig. 7.5, demonstrating their respective power dependence. Evidently, the SHG signal can be fitted well with a straight line, showing a perfect quadratic power dependence. The TPL signal is also very quadratic at a low power level. This demonstrates that indeed the florescence is induced by the two-photon absorption. Nevertheless, when the power exceeds certain threshold, the TPL slope starts to get slower, and approaches the linear power slope at the high power limit. We explain this saturation effect by introducing a power dependent lifetime into the picture. This lifetime reduction can be caused by the exciton-exciton annihilation process, which is determining the exciton lifetime when the exciton density is higher than $1/(\text{exciton radius})^2$. This effect is found to be very strong in electroluminescence, which is going to be discussed in Chapter 8.
After understanding the saturation effects TPL, we start to scan the excitation wavelength. The emission spectra at different excitation energies are compiled into a 2D image, depicted in fig. 7.6 (a). At every excitation energy, the power is limited carefully to make sure we are operating in the non-saturation regime, and the emission spectrum is normalized by the excitation power at the focus. The SHG signal manifests itself as a straight line with a unit slope in the 2D plot. Since we focus on the TPL process here, only a small portion of SHG spectrum is shown in the figure and the colorbar is adjusted to best visualize the TPL contrast. TPL features show up as two horizontal lines at 605 and 618 nm. Clearly, the TPL signal has a non-constant level at different excitation energies. The emission lines show two clear peaks when the two-photon wavelength matches 540 nm and 495 nm. As mentioned above, the emission intensities have similar trends between the excitation and trion peaks. Then, we plot the emission peak values at 618 nm versus the excitation energy. (fig. 7.7)

---

**Fig. 7.6:** 2D emission-excitation plot with the normalized excitation intensity. The SHG sloped line is saturated in order to achieve an optimal contrast of the TPL signals.

---

**Fig. 7.7:** Overlap plot of one-photon transmission (green) and two-photon absorption (blue) spectrum. Arrows are pointed at the emission energy levels and short bars highlight the TPL peak positions.
In fig. 7.7, the trion and exciton emission energy are represented by two black arrows at 2 and 2.04 eV. This agrees with the A exciton peak in linear absorption spectrum, plotted as the green line in the same figure. Above the emission energy, there is a significant gap where no TPL is detected. Above the gap, at about 2.28 eV, a large TPL peak is observed with a linewidth about 80 meV. Then TPL signal drops to certain level until another peak occurs at 2.48 eV. This peak shares a similar peak width but has a lower intensity than the low energy one. According to our excited exciton picture, the 2.28 eV peak correspond to the 2p state of A exciton and the 2.48 eV peak is linked to the 3p state of A exciton. The energy difference between them and the emission energy, which is 0.24 eV and 0.44 eV, correspond to the 1s-2p and 1s-3p separation in hydrogen atoms, and they show the exciton binding energy is no smaller than this, and should have the same order of magnitude with this separation. Apparently, these are the two-photon active only levels in the energy space and therefore no corresponding features can be observed in one-photon spectrum. The large binding energy we detect also qualitatively agrees with the estimation made from the trion binding energy.

The excited of B exciton and C exciton should exist in the even higher energy levels, but we are not able to detect them experimentally since the relaxation and emission process become very inefficient when the excitation energy level is too high.

The experiments are operated at the room temperature.

The monolayer material is usually sensitive to the environment, and the binding effect might be screened by the dielectric background. In order to investigate study it, we capped the monolayer WS$_2$ with different materials, including water, immersion oil and aluminum oxide. (fig. 7.8) The average dielectric constant at the emission energy can be estimated by $\varepsilon_{\text{av}} = (\varepsilon_{\text{top}} + \varepsilon_{\text{sub}}) / 2$, which varies from 1.7 to 2.5. After the capping, the A exciton emission is not shifted within the experimental error. This is due to the well-known screening effect on both the self-energy and the binding energy. As the environment screening reduce the binding energy, the self-
energy is also reduced by the same amount. These two factors cancel each other, leading to an invariance of emission energy. The results shown in fig. 7.8 are measured at the room temperature. Therefore, the exciton emission is red shift to 2 eV. On the other hand, we find the binding energy, which is approximately proportional to the A exciton peak and 2p peak separation, doesn’t change either within the experimental error. 1s-2p separation is about 0.24 eV and 1s-3p separation is about 0.45 eV, in good agreement with the low temperature results. This also means the measured 1s-2p splitting is already the intrinsic property of the monolayer even though it is lying on the quartz substrate, which can be explained by the fact that the binding energy is so large and the radius of the exciton is very small so that the highly confined electron hole pair couldn’t see the screening effect of the dielectric background.

### 7.2.3 Binding energy and exciton radius

In order to retrieve the total binding energy of the A exciton, we resort to the GW-BSE based density function calculation. The result of the A exciton and its excited levels are depicted in fig. 7.9.

![Fig. 7.9: DFT calculation with GW approximation. Dashed line is the result including BSE and the solid line is without BSE approximation. The difference between them is caused by the exciton effect. The red arrows are pointed at 1s and 2s bright excitons and the green arrows are pointed at 2p and 3p dark excitons. The band edge position is highlighted by a yellow arrow.](image)

The calculated linear absorption spectra with (dashed line) and without (solid line) exciton effect are plotted in fig. 7.9. Evidently, the excitons can modify the absorption spectrum to a great extent. The star symbol labels the linear oscillation strength of each calculated exciton eigen state. At every absorption peak, there is always an exciton state. These excitons are so called 'bright' excitons with even envelope function and can be excited by one-photon transition. Besides them, there
are many excitons levels outside the absorption peaks, and have very low linear oscillation strengths. They are ‘dark’ excitons with odd parity in the envelope function and cannot be excited by one-photon transition. These dark excitons contribute to the resonance in the two-photon absorption spectrum.

By checking the group character of these excitons’ envelope functions, we can assign them to the s and p symmetry groups, in consistence with the previous definitions. The 1s state is calculated to be 2.05 eV, 2p is 2.29 eV, 3p is 2.49 eV. These values agree quantitatively very well with the experimental results, up to an error of 10 meV. And similar to MoS$_2$ calculation, we also found a 2s state between 2p and 3p states, which has not been experimentally observed due to the possible thermal broadening as well as the B exciton contribution. Unlike the 3D hydrogen model, 2p and 3P states only have 4 eigen states, rather than 6. Because the wavefunction is limited to 2D, it can only rotate in the plane and no angular momentum can be projected into the plane. Therefore, the magnetic number, m, can only be +1 and −1. If we consider the additional valley degeneracy, K and K', each l series can only have 4 eigenstates. (green arrows in fig. 7.9)

More interestingly, we find the total binding energy, which is the separation between 1s state, marked by the red arrow, and the band edge energy, marked by yellow arrow, is 0.65 eV. This is a very significant value considering the exciton binding energy in normal semiconductor is on the meV level. The enhancement could come from the fact that the phase space that allow the electron and hole to be separated by r has been reduced from $r^2$ to $r$. As a result, the electron hole pair tends to stay closer in the two-dimensional system like here.

![Fig. 7.10: Calculated 1s exciton wavefunction both in real space and k space. The real space wavefunction has an in-plane extension of about 1 nm in radius. The wavefunction is highly localized at K points.](image)

It is also interesting to check the wave function of the exciton. (fig. 7.10) For the ground exciton, it is located in the minimal energy locations in the k space, which is the K point at the hexagonal corners. As can be seen, the wavefunction spread is
only about three unit cells, or 1 nm meter in space, which is 20 times the Bohr radius. This corresponds to a 20 times reduction of the 1s binding energy in the hydrogen, that is \( \frac{13.6}{20} = 0.68 \text{ eV} \), in good agreement with the energy level calculation. Along the z dimension, the wavefunction is highly localized at the Mo layer, which might explain the insensitivity of the binding energy to the dielectric environment.

7.3 Crystal orientation and edge state imaging by SHG

Because SHG is strongly affected by the lowered symmetry of a material, the structure and symmetry properties of MoS\(_2\) atomic membranes can be probed by second-harmonic emission, which allows us to demonstrate the nonlinear optical crystallography on these unique 2D materials\(^{111-113,137}\). Here we show the crystal orientation and the edge state at Mo edge can be directly imaged with SHG signal.

7.3.1 Imaging the layer thickness and crystal orientation

As the symmetry is different in odd layer TMDC and even layer TMDC, SHG signal is oscillating at full amplitude as the layer number increases. Experimentally, we exfoliate the flakes on the SiO\(_2\)/Si substrate, and from time to time we can find a sample with a slow gradient in the layer thickness, depicted in fig. 7.11.

![Fig. 7.11: White light image (a), confocal luminescence mapping (b) and SHG signal mapping (c) of a sample with mono-, bi-, and trilayer areas.](image)

The layer thickness can be approximated by the white field image contrast in fig. 7.11 (a). The tip with the bright purple color is the monolayer. On the right, it is connected with a bilayer and a trilayer which increasing absorption. This is also confirmed by Raman and AFM methods. A photoluminescence confocal mapping excited by a 488 nm laser is shown in fig. 7.11 (b). Consistent with other sample, only monolayer has very bright photoluminescence intensity. Both bilayer and trilayer are as dark as the bulk park in the image. The unique contrast of even-odd layer thickness is shown in fig. 7.11 (c). As we scan a 1300 nm laser across the multilayer sample, we only collect the SHG emission around 650 nm, which is similar to the results in fig. 7.4 (a), as our total signal at the scanning point. The
mapping clearly indicates that mono- and trilayer will give out very strong SHG, and the bilayer gives very little SHG even though it has twice material than the monolayer. The reason can be understood as the nonlinear dipoles in the neighboring layers are oscillating out of phase and therefore cancel each other. This situation happens in every even layer samples. Below the few layer regime, there is a bulk with many different thickness areas and some of the areas give out the similar level signals as the mono and trilayers. Very likely, they are the odd layer areas, but we couldn’t tell their absolute layer thickness by AFM. To note, the resolution of this nonlinear imaging technique doubles the linear imaging resolution at the pumping frequency, and can reach the sub-micron level in this case.

From the geometry point of view, SHG will have the strongest signal when its polarization is along the inversion symmetry broken direction. To be more specific, if the SHG signal is detected along the incident polarization direction, the intensity will be strongest when the polarization is along the mirror axis. This can be shown by calculating the nonlinear polarization induced by an arbitrarily linear polarized incident light. If we assume the laser polarization and detection direction has an angle $\theta$ to one of the mirror axis, the SHG intensity $\theta$ dependence is calculated as following:

$$
E_x = E \cdot \cos \theta; \quad E_y = E \cdot \sin \theta
$$

$$
P_x = \chi^{(2)}_{xx} \cdot E_x^2 + \chi^{(2)}_{yy} \cdot E_y^2 = (\cos^2 \theta - \sin^2 \theta) \chi E^2 = \cos(2\theta) \chi E^2
$$

$$
P_y = \chi^{(2)}_{yy} \cdot E_x E_y + \chi^{(2)}_{yx} \cdot E_x E_y = -2\sin \theta \cos \theta \chi E^2 = -\sin(2\theta) \chi E^2
$$

$$
P(\theta) = P_x \cdot \cos \theta + P_y \sin \theta = \cos(3\theta) \chi E^2
$$

$$
I(\theta) = \cos^2(3\theta) \chi^2 E^4 = \cos^2(3\theta) \chi^2 I^2
$$

(7.2)

Therefore SHG has a strong anisotropic dependence between the polarization and the crystal orientation, which is plotted in fig. 7.12 (a). The angle definition is illustrated in fig. 7.12 (b). The x axis is the armchair direction and y axis is the zigzag direction. Such a contrast only shows up when SHG is detected along the excitation direction, because the orthogonal nonlinear polarization is proportional to $\sin(3\theta)$. The sum of them becomes isotropic again.

---

*Fig. 7.12: (a) Polarization dependence of the SHG intensity. The incident and detection polarization are the same, and labeled in the fig. (b). SHG is dependent on the angle between the polarization axis and the crystal orientation. Without phase sensitivity, the intensity pattern has a six fold symmetry.*
Fig. 7.13: White light image (a), total SHG signal (b) and y-axis polarized SHG signal (c) of a continuous CVD grown MoS$_2$ monolayer. (d) Illustrations of the crystal orientation distribution of A and B areas to demonstrate the polycrystalline nature of the CVD film.

Because we measure the SHG signal without phase sensitivity, the emission pattern should have a six-fold symmetry, without the resolution of sulfide and molybdenum atom. It can still help to differentiate the crystal orientation, which is particularly useful for imaging the grains in continuous MoS$_2$ film grown by chemical vapor deposition (CVD) method. Figure 7.13 shows the linear and nonlinear optical images of a continuous CVD grown monolayer MoS$_2$ membrane. Oxide-on-silicon substrates with 285 nm thick oxide films are utilized for optimized optical contrast. The monolayer samples show high quality photoluminescence and they are optically uniform over large areas (fig 7.13 (a)). In stark contrast, the SHG image (fig 7.13 (b)) reveals the polycrystalline nature of the uniform monolayer, and the averaged grain size ranges from 20 to 40 µm, and the pump wavelength is 1300 nm. Here the incident polarization of the pump beam is along the y-axis, and the total second-harmonic radiation is collected. The uniform SHG intensity within each grain indicates that the individual grains are single crystals. At the grain boundaries, however, the SHG is substantially suppressed due to the destructive interference and annihilation of the nonlinear waves generated from the neighboring grains with different orientations. Though the grain boundaries are only a few atom sites in width, the crystal boundaries are clearly seen in fig 7.13 (b).

Moreover, the nonlinear generation not only reveals the symmetry properties of the crystal, but also allows the rapid mapping of crystal orientations due to its two-dimension crystalline nature. Here, by adding a polarizer along the incident direction, we map the crystal orientation of the polycrystalline atomic membrane respective to the y-axis. This approach allows us to capture a complete map of the crystal grain structures, color-coded by its crystal orientations, as shown in fig 7.13 (c). Large area CVD-grown films clearly has complex grain geometries and orientations.

As shown above, due to the three-fold rotation symmetry of monolayer MoS$_2$, the SHG imaging without phase information does not discriminate between opposite crystal orientations. Consequently, it only determines grain rotations modulo 60 degrees and the measurable difference between grain orientations is from 0 to 30 degrees as illustrated in Fig. 1c. For example, crystals I and II have no contrast in the color-coded orientation maps (fig 7.13 (c)), but the clearly visible grain boundary
between (fig 7.13 (b)) indicates the two crystal grains are in the opposite orientations. On the other hand, the grain boundary between crystals II and III has much less contrast (fig 7.13 (b)), but the two crystals are distinctive when looking at the crystal orientations in fig c (~ 12° difference in orientations).

Twin grain structures can be identified by six-fold symmetric patterns, as shown in fig 7.13 (b) and fig 7.13 (c), where crystals in group A have a cyclic twin structure while each crystal in group B instead has a ~ 30° difference in orientations with respect to its neighbors. Such a characterization process conventionally requires diffraction-filtered dark-field transmission electron microscopy as well as a sophisticated and time-consuming analysis procedure. The rapidly nonlinear optical crystallography, however, allows us to directly visualize the grain structure in a non-invasive manner, providing key information for optimizing synthesis strategies for large area CVD growth. Though the results shown here were obtained for a fundamental wavelength of 1300 nm, the symmetry and the polar patterns of SHG are ubiquitous for all pump wavelengths.

7.3.2 Edge state nonlinearity

Nonlinear optical susceptibilities are highly surface-specific and sensitive to extremely small changes in materials’ electronic properties. SHG has therefore been one of the most outstanding tools for surface science. Similar to the surface of a three-dimensional crystal, breaking the translational symmetry at the edge of a 2D crystal potentially creates one-dimensional atomic edges with localized electronic properties that may substantially differ from the 2D bulk.

Fig. 7.14: (a) SHG image pumped at 1280 nm. (b) SHG image of the same sample pumped at 1300 nm. The brim of crystal shows a strong nonlinear optical edge states. The scale bare is 20 µm. (c) Cross-sections (white dashed lines shown in (a) and (b)) compare the SHG of the same crystal under the pump wavelength of 1280 nm and 1300 nm, respectively.

Fig. 7.14 shows SHG images of discrete triangular-shaped CVD-grown monolayer islands at pump wavelengths of 1280 nm and 1300 nm, respectively. While actually rough at the atomic scale, the edges of these monolayer islands are the more energetically favorable termination of molybdenum zigzag edges, and appear straight and sharp under the optical microscope. The SHG images with two different fundamental wavelengths, however, show striking difference at these atomic edges.
When pumped at 1300 nm, a brim of enhanced SHG is observed at the edges of the crystals (fig. 7.14 (b)), which is substantially stronger than that of the central region. At the corners, the SHG is even stronger. As a comparison no such enhancement is observed at the edges at the pump wavelength of 1280 nm. The edge enhanced SHG is further seen in the cross-sections in fig. 7.14 (c). Structure and field discontinuities at the edges of these 2D crystals can potentially lead to enhanced in-plane nonlinear polarizability, like the surface enhanced SHG often observed on metal and dielectric surfaces\textsuperscript{138,139}. Such an enhancement, however, shall be equally effective for different pump wavelengths as long as there are no additional electronic states located at the edge. The stronger second-harmonic edge responses pumped at 1300 nm than that at 1280 nm is therefore the results of the electronic structure variations between the sharp crystalline edges and the bulk monolayer crystals. Moreover, assuming that the region where the electronic states are affected by the existence of the atomic edges is about the same as the thickness, \textit{i.e.}, less than one nanometer, it requires more than three order magnitude of nonlinear enhancement to exceed the nonlinear generation in the central region given an illumination pump laser spot size greater than 1 µm.

![Fig. 7.15: Nonlinear edge state resonance of MoS\textsubscript{2} monolayer membrane. (a–f), SHG images of the same monolayer MoS\textsubscript{2} sample for the fundamental wavelengths of 1280 nm, 1290 nm, 1300 nm, 1310 nm, 1320 nm, and 1330 nm, respectively. The scale bar is 20 µm. (g) Resonance of the edge nonlinear response at \textasciitilde 0.947 eV (1310 nm). The ratio is determined by the integrated SHG with the areas circled in (a).]

To further elucidate these phenomena, we imaged spectroscopically the SHG at the edges and corners of the samples. We observed a resonance wavelength is about 1310 nm (\textasciitilde 0.95 eV), a slightly higher photon energy than the half transition energy of the A exciton complex (fig. 7.15). Quantum-mechanical expression for the second order nonlinear susceptibility of a material involves matrix elements with one occupied and two unoccupied intermediate states. When there is an additional localized energy level nearly coincident with virtual transition levels, the coupling between the radiation and the nonlinear optical susceptibility becomes large. The nonlinear optical transition matrix elements at the edges may involve additional
occupied edge states when compared to that of the crystalline atomic membrane, resulting in strongly enhanced second-order nonlinear susceptibilities.

Density functional theory has predicted the existence and provides atomic-scale insight into these edge states\textsuperscript{140,141}; such one-dimensional edges have been identified on MoS\textsubscript{2} clusters on single crystal gold by scanning tunneling microscopy in ultra-high vacuum.\textsuperscript{142} The existence of special electronic edge states for finite low-dimensional crystals has also been reported in graphene nanoribbons\textsuperscript{143-145} and later at the end of one-dimensional atomic chains (end states)\textsuperscript{146}. We report here, for the first time, the strong optical nonlinear responses of such edges in 2D MoS\textsubscript{2} membranes, which have for a long time been suspected to be the active sites for electrocatalytic hydrogen evolution\textsuperscript{147,148}. As a control experiment, there is no extraordinary photoluminescence responses observed on these edges.

7.4 Conclusion and outlook

In conclusion, we have shown the nonlinear spectroscopy is a very ideal tool to study the material property of low-dimensional TMDC. We discovered an extraordinarily large exciton binding energy in TMDC monolayer, via the two-photon absorption spectrum, which is confirmed with the DFT calculation and exciton-trion separation. This large binding energy is found to be not sensitive to the dielectric environment. On the other hand, we utilize the second harmonic generation to directly image the oddness, crystal orientation of TMDC few layer and discovered an electronic edge state in the CVD grown sample.

These nonlinear phenomena should be universal among TMDC systems and therefore, in the future, it would be interesting to explore the binding energy and SHG coefficient in other TMDC materials. More importantly, because of the large exciton effect, the band gap size as well as the conduction/valence band locations are very different from previous understandings. All the electronic devices and optoelectronic devices have to be redesigned based on the new understanding of the electronic structures. Also, the observed photo current effect has to be re-explained in the strong exciton picture.
Chapter 8

Electroluminescence from a silicon-MoS₂ PN junction

8.1 Introduction

As a 2D direct bandgap semiconductor, TMDC is a very promising candidate as an electrical emitter. Nevertheless, all the previous effort has been focused on the photoluminescence and absorption study. Here, we demonstrate the electroluminescence from monolayer MoS₂ fabricated on top of a heavily p-type doped silicon substrate.\(^{18}\) We investigate the microscopic origin of electroluminescence, the temperature dependence of it and the internal quantum efficiency is estimated to be around \(10^{-7}\). By comparing the photoluminescence and electroluminescence of a MoS₂ diode, multiple excitonic peaks are identified. Auger recombination of the exciton-exciton annihilation of bound exciton emission is observed under a high electron-hole pair injection rate at room temperature. Hereby, we demonstrate the electroluminescence as an ideal complimentary tool to study monolayer TMDC.

8.2 PN junction: design, fabrication and electrical performance

The exfoliated MoS₂ from the natural crystal has a very high level of electron doping, about \(10^{12}\ cm^{-2}\).\(^{107}\) It comes from either the impurities levels or the defects within the crystal. As a result, the easiest way to generate the electroluminescence in MoS₂ is to combine it with a heavily hole doped material, thus forming a PN junction between them. So far no P type MoS₂ has been reported except the device electrically doped by liquid ion at a high gating level. We choose the P-dope silicon and let it form a hetero-junction with MoS₂. The band diagram is depicted in fig. 8.1.

![Fig. 8.1: Band alignment of between p-type silicon and n-type MoS₂. (a) The relative band positions when they are separate. (b) When the two materials get contacted and reach an equilibrium state as a PN junction. (c) A forward bias is applied, inducing a current and consequent emission in MoS₂. The blue dots represent the holes in silicon and the green dots represent the electron in MoS₂.](image-url)
As we know, silicon has a bandgap of 1.1 eV and, according to Chapter 7, MoS$_2$ has an intrinsic band gap of 2.8 eV$^{117}$, while the affinity energy has a relatively same value for both materials (fig. 8.1 (a)). When they get into contact, a natural PN junction will form at the interface between them, and since MoS$_2$ has a small electron density, it is easy to deplete the contacting area in MoS$_2$. Due to the difference in the band gap size, both the conduction band and the valence band have an offset between the materials. (fig. 8.1 (b)) When a higher electrical potential is applied on the MoS$_2$ side, the bands will bend accordingly, as depicted in fig. 8.1 (c). The bands will first become flat, and above a certain threshold, the band will be bended upward, opposite to the original trend, and a charge current will be generated along the bias direction. Following the convention, we define this bias as the forward bias. On the other hand, if a negative bias is applied, the original band bending will become even stronger, and no current, except a leaky one, will be generated.

The forward bias current is consisted of two components. One is from the conduction electron flowing from MoS$_2$ to silicon and the other is the hole current, which flows from the silicon to MoS$_2$. These additional holes will combine with the majority carrier, electrons, in MoS$_2$ radiatively and emits light from the monolayer. Most of the emission has the bandgap energy in MoS$_2$ monolayer as hot electrons and holes relax fast to the band edge. On the other hand, not much of the hot electron will generate emission in silicon since its band gap is an indirect type.

Experimentally, we first prepare an array pattern of highly doped p-type silicon on silicon oxide. This is done by doping the SOI wafer (Soitec Inc.) assisted by plasma. RIE etching was is to remove the undesired silicon area until the SiO$_2$ layer is reached. The Si/SiO$_2$ step is where we expect to form PN junction with MoS$_2$. Therefore, we developed a site-control transfer method to precisely position the monolayer MoS$_2$ across the silicon/silicon oxide step. Details are discussed in Chapter 6.2. Basically, after the monolayer is identified, we crop it out by cutting a PMMA around. Under the optical microscope, we carefully place the monolayer with PMMA onto the Si/SiO$_2$ step. Some intermediate steps are shown in fig. 8.2. And the final device schematic is depicted in fig. 8.5(a).

![Fig. 8.2: Processes of site-controlled MoS$_2$ transfer method. The dashed circle highlights the monolayer MoS$_2$ flake across the Si/SiO$_2$ interface. Scale bar: 20μm.](image)
After the transfer, In/Au contacts are fabricated (described in Chapter 6.3) to contact both MoS$_2$ and silicon. IV curve in fig. 8.3 shows an ideal diode response: a very large conductance at the forward bias up to 3V, and a very small current upon the reverse bias till -3V.

![IV curve of the PN junction shows an ideal diode behavior.](image)

**8.3 Electroluminescence from direct bandgap exciton and defect exciton**

The Electroluminescence is measured in a cryostat-based scanning optical microscope, illustrated in fig. 8.4. Depending on the flipping mirror, the emission can be guided to either a single-photon sensitive camera (Andor DL-604M-#VP) or a fiber coupled spectrometer, so that both the emission pattern and the spectrum can be obtained. We use a 50x objective with a NA of 0.55 to optimize the collection efficiency, given the limited working distance, and a 633 nm HeNe laser is included to excite photoluminescence as well as to help the alignment. A 3D piezostage is attached to the objective so that the excitation laser spot can be scanned over the sample.

![The experimental setup measures both electroluminescence and photoluminescence at cryogenic temperature with a laser scanning capability.](image)
Electroluminescent emission captured by a single-photon sensitive camera from a device at forward bias voltage of 5 V and 42 µA current at the room temperature is depicted in fig. 8.5 (b). After superimposing a white light scattering image of the device, we find that the electroluminescence occurs via injected electron-hole recombination processes and is localized in the edge of the heterojunction. By applying an in-plane bias voltage, the largest voltage drop occurs across the heterojunction edge due to the semiconducting characteristics of MoS$_2$. The schematic of the device as well the emission hot spot are illustrated in fig. 8.5 (a).

![Schematic and emission hot spot](image)

**Fig. 8.5:** (a) Schematic of the PN junction and the emission hot spot. Pink color represents a 300-nm-thick silicon layer while the purple layer is a SiO$_2$ layer and In/Au contacts are drawn in yellow. (b) Experimental observed emission pattern. White dashed lines outline the monolayer boundary and the grey dashed line is the silicon step edge.

This is further confirmed by the electrostatic potential mapping by a scanning photocurrent microscopy. (fig. 8.6) Since our setup combines laser scanning function with electrical measurement capability, we can map out the photocurrent distribution. The photo current is originated from the desociation of excitons by the built-in electric field in the depletion regime of the PN junction. Simultaneously, we can collect the back scattering signal of the laser scanning (fig. 8.6(a)), which reflects the geometry of the device. In the scattering image, we highlight the In/Au ohmic contact and monolayer MoS$_2$ with yellow and white dash lines separately. As shown in the photocurrent map in fig. 8.6 (b), we only find a pronounced photoresponse around the edge of the heterojunction. Such a strong photo response indicates a strong built-in electric field across the MoS$_2$/silicon junction. The photocurrent hotspot location also agrees well with the electroluminescence hotspot location.

![Scattering and photocurrent mapping](image)

**Fig. 8.6:** (a) Scattering mapping and (b) photocurrent mapping of the device. The yellow and white dash lines define the In/Au electrode and monolayer MoS$_2$ boundaries respectively. Scalebar: 5 µm.
To measure the photoluminescence and electroluminescence spectra, we coupled the emitted light to a fiber coupled spectrometer (Andor Shamrock 303) as shown in Fig. 8.4. Figure 8.7 shows the room-temperature electroluminescence spectrum of monolayer MoS$_2$ at a current of 42 µA. The two principal luminescence features at 667 nm (labeled AX) and 720 nm (labeled DX) are associated with the A exciton and the defect related bound exciton of monolayer MoS$_2$, respectively. Compared with the photoluminescence spectrum plotted in the same figure, the high electrical bias causes a spectral broadening. The B exciton, with its energy close to that of the A exciton, is indistinguishable in the electroluminescence due to the low efficiency of the B exciton emission process and the broadening of the electroluminescence.

![Graph showing PL and EL spectra](image)

*Fig. 8.7: Electroluminescence and photoluminescence spectra from the same device. The electroluminescence spectrum is measured at the current level of 42 µA. Besides the strong direct exciton emission (labeled AX), a weaker satellite peak at a lower energy is observed (labeled DX).*

Figure 8.8 depicts the room-temperature electroluminescence spectra under varied current. The electroluminescence exhibits a current threshold of about 15 µA in this device; the two main features of the AX and DX excitation emissions are clearly observed at currents exceeding this threshold. They can be well fitted with two Lorentzian contributions. In Fig. 3b, we present the current dependence of the AX and DX emission intensities as extracted from Fig. 3a. The A exciton emission, AX, shows a linear dependence with increasing current. However, the bound exciton, DX, rises linearly at low currents but saturates as the current exceeds about 65 µA. Saturation of the DX exciton emission cannot be caused by a phonon-assisted nonradiative process, as the two peaks display different current dependencies. We propose this is an effect involving multiple exciton-exciton interactions, which is similar to Auger recombination, a process well documented in tightly confined carbon nanotube systems. Auger recombination may lead to rapid exciton-exciton annihilation when extra excitons or multiple excitations are present. At a low electron-hole pair injection rate of $I/2q$ ($q$ as the electron charge), there may only be one electron-hole pair in excited monolayer MoS$_2$. Thus, we observe linear dependence with increasing current at low injection rate.
Fig. 8.8: (a) Electroluminescence spectra of a MoS$_2$ diode recorded at different currents from 17 to 86 µA. The data can be fitted well with two Lorentz contributions, which are attributed to A exciton (labeled AX) and bound exciton (labeled DX) emission. (b) Red dots: the A exciton emission (AX) shows an approximately linear increase with current. Green squares: bound exciton emission (DX). The electroluminescence saturates as the current exceeds ~65 µA. The dashed lines are fitting results. (c) Electroluminescence spectrum at a current of 86 µA. A new exciton (labeled NX) feature emerges with a higher energy of 2.26 eV. Red curves are the Lorentz fitting result.

When the electron-hole pair injection rate exceeds the inverse carrier lifetime $\tau_{\Gamma}^{-1}$, more than one electron-hole pair is present in monolayer MoS$_2$. The Auger process opens up a nonradiative recombination channel for electron-hole pair recombination. If the Auger process is efficient, it will quickly deplete the population of electron-hole pairs. The annihilation of the electron-hole pairs comes to a stop when only a single electron-hole pair remains in the monolayer MoS$_2$. Thus, we observe the saturation of the exciton emission at sufficiently high injection current. The sudden saturation further suggest that the DX-DX annihilation lifetime is much smaller than the total carrier lifetime, $\tau_{A}^{DX} \ll \tau$. On the other hand, due to the
absence of any noticeable AX saturation, we expect the AX-AX annihilation lifetime $\tau_{A}^{AX}$ to be much longer than $\tau_i$.

It is well known the excitonic transition is highly temperature dependent$^{114}$. With decreasing the device temperature to 10 K, the electroluminescence resonance exhibited a blue shift from 667 nm to 662 nm for the AX peak and from 720 nm to 701 nm for DX peak, while the full-width at half-maximum linewidth decreased down to 32 nm. The electroluminescence spectrum is nearly consistent with the photoluminescence spectrum at low temperature. Deviations are caused by the slight red shift and spectral broadening arising from the high electrical current induced local temperature.

**Fig. 8.9:** Low temperature electroluminescence spectra and electrical power dependence of the MoS$_2$ diode shown in Fig. 8, recorded at different current levels from 13 to 82 $\mu$A. The experimental temperature is 10K.

Again, the two main features of the AX and DX excitation emissions can be clearly read out under all the currents above threshold (Fig. 4a). In Fig. 4b, we present the current dependence of the AX and DX emission intensities as extracted from Fig. 4a. Surprisingly, the DX emission exhibits different current dependence behavior compared to that of room temperature. Both AX and DX emissions show a linear
increase with current. The absence of the saturation of the DX emission under a high electron-hole injection rate at low temperature is due to the slowing down of the Auger rate\textsuperscript{156}. Past work has already demonstrated that the lifetime of the electron-hole pair recombination at low temperature (4.5 K) is prolonged by more than one order compared to that of room temperature\textsuperscript{154}. As the prolonged DX-DX annihilation lifetime $\tau_{\text{A} DX}$ is comparable or longer than $\tau_L$, there will be no observable saturation of the DX emission even under high injection rate.

Under a high electron-hole injection rate, a new high-energy peak emerges in the electroluminescence spectra both for the room temperature (fig. 8.8 (c)) and low temperature (Fig. 8.9 (c)) case. We fit the electroluminescence spectrum at room temperature under current of 86 $\mu$A with Lorentzian contributions and find a new emerged feature (labeled NX) peaks at 550 nm with energy of 2.26 eV. Further experimental and theoretical work is needed to identify the microscopic nature of the exciton recombination.

The carrier-to-photon conversion efficiency is defined as the number of emitted photons per injected electron-hole pair. After calibrating the collection efficiency of our setup, we estimate an electroluminescence efficiency of $\sim 10^{-7}$ photons per injected electron-hole pair, which is much lower than the photoluminescence efficiency of up to up to $10^{-3}$ in monolayer MoS$_2$. To account for this difference of four orders in magnitude, we propose three possible reasons. Due to the big valence band offset between monolayer MoS$_2$ and silicon, holes injected from silicon to monolayer MoS$_2$ only carry a small fraction of the current weight. However, only the holes injected to monolayer MoS$_2$ have the possibilities to recombine with the majority carriers (electrons) in monolayer MoS$_2$ to emit light. Additionally, only a fraction of the electrically induced electron-hole pairs carries the correct spin to populate the radiative singlet exciton states\textsuperscript{16}. Finally, the short non-radiative lifetime of the electron-hole pairs may lead to an additional reduction in efficiency. Pump-probe experiment has demonstrated that the non-radiative lifetime is one order shorter than that of radiative process in monolayer MoS$_2$.

### 8.4 Conclusion and outlook

In conclusion, we demonstrate the electroluminescence of monolayer MoS$_2$ fabricated on a heavily $p$-type doped silicon substrate. The high signal-to-noise ratio allows for the identification of emission from different optical transitions. Auger recombination of the exciton-exciton annihilation of bound exciton emission is observed under a high electron-hole pair injection rate at room temperature for the first time. Under high electron-hole pair injection rate, a higher energy exciton peak of 2.26 eV is observed for the first time in monolayer MoS$_2$ system.

Further direction includes replacing the current solid-state substrate with a flexible substrate. Because the active regime is now atomically thin, the MoS$_2$ based emitting device should be very flexible to conform to any shaped surface. In addition, theories have predicted that the strong in-plane bond in MoS$_2$ can sustain very high strain before breaking down. This should work hand in hand with the emerging
flexible electronics field, such as the graphene based circuit. Another interesting direction is to couple the MoS$_2$ with a spin polarized contact. As the next chapter will show, the spin and valley degree of freedom has a one-to-one correspondence in monolayer MoS$_2$. By injecting spin-polarized current in to the junction, we might be able to get the circular polarized radiation from the light emission device.
Chapter 9

Towards the valley Hall effect in TMDC

Momentum valleys in the electronic band structure have been proposed to be a new degree of freedom for information storage and processing. In this new architect, the information is encoded in the index of the valley where electrons are populated, or in short, valley polarization\textsuperscript{157-160}. So far, valley polarization has only been injected and detected in optical way. In this chapter, I will discuss our effort to electrically detect the valley polarization via the so-called valley Hall effect. The effect manifests itself as a photocurrent, transverse to the applied voltage and purely dependent on the helicity of the incident photon.

9.1 Introduction to the valleytronics and valley Hall effect

A core question of condensed matter physics is how to understand the electron behaviors in crystal lattice. In the early age, electrons were simply treated as charged particles. However, people latter on realized that the behaviors of electrons are more complicated, which is associated with different degrees of freedom. The coupling between lattice, charge, spin and orbital degrees of freedom provides a rich platform to study and manipulate the behaviors of electrons in condensed matter, which might lead to novel functionalities. For instance, the coupling of charge and spin degrees of freedom is the foundation of spintronic devices, which has been proposed to be the next-generation electronic devices\textsuperscript{161}.

Recently, valley degree of freedom has become another new dimension for the computing application\textsuperscript{160}. In this case, the information is encoded in the position of the electron in the momentum space. As conduction electrons usually sit in the near-band edge valleys in the energy-momentum space, we define the valley index to be our information bit. For example, in the hexagonal lattice, if the electron is populated in the K valley, it stands for the ‘1’ state and if the electron is in the K' valley, it means the ‘0’ state. (fig. 9.1) Theoretically, one can use this degree of freedom to build a completely new type of transistor. All the signal processing and data storage functions can be implemented on this new platform.

![Fig. 9.1: Band structure of MoS$_2$. Electron at different k points have different valley indexes, indicating different effective spins and bit information.](image-url)
Similar to spintronics, valley based electronics, or valleytronics, is a subject involving quantum effects. Because the two states are separated far away from each other in the momentum space, only phonons with very large momentum can interfere with the encoded information, through a rare process of inter-valley scattering. Therefore, the coherence of the valley polarization can be kept for long in the crystal. Compared with spin polarization, which is susceptible to many decoherence mechanism such as the nuclear spin of the ion background, valley polarization definitely has certain advantages in realization the quantum computing in solids.

Nevertheless, the challenging point is that most valleys are located at highly symmetric K points, and therefore degenerate with each other in terms of the energy level. So it is not easy to initialize and read out these information hidden in the momentum space. At the current stage, the orbital angular momentum associated with each valley has been heavily exploited to distinguish valleys in an optical way. A popular material system to realize it is TMDC monolayer.

In the electronic structure, TMDC monolayer shares many features with bilayer graphene with broken inversion symmetry. Among them, the most important implication to valleytronics is that electrons and holes in a single valley will acquire different Berry phase, and thus different orbital angular momentum. The difference in the orbital angular momentum leads to an exact selection rule for the polarization of excitation photon during the interband transition process. The valley selection rule is opposite for opposite valleys. Consequently, one can utilize the polarization of the photon to selectively excite the electrons in the target valley. In addition, since monolayer TMDC is a direct band gap material, the excited electron hole pair would recombine according to the same selection rule as in the excitation process, thus releasing, via the helicity, the index of the valley they are residing in. Hereby, by detecting the polarization of the emitted photon, we can measure the valley polarization within the material.

![Fig. 9.2: (a) Polarization resolved photoluminescence from MoS$_2$ at 10K. (b) The helicity is retrieved based on fig. 9.1 (a). (c) The temperature dependence of the helicity ranging from 10K to room temperature.](image)

One example from MoS\textsubscript{2} is depicted in fig. 9.2. Circularly polarized light, $\sigma_-$ or $\sigma_+$ light first excite MoS\textsubscript{2} monolayer, the helicity resolved photoluminescence spectra are collected in fig. 9.2 (a). Evidently, there’s a large contrast between the $\sigma_+$ and $\sigma_-$ emissions. The polarization of the emission, calculated by $\rho = (I_{\sigma_-} - I_{\sigma_+}) / (I_{\sigma_-} + I_{\sigma_+})$, is plotted in fig. 9.2 (b) for both helicity excitations, which approaches 50 percent as the emission energy is getting close to the excitation one at 10K. The emission will lost the helicity as the temperature approaches the room temperature (fig. 9.2 (c)). The temperature mechanism has been attributed the phonon and the non-perfect selection rule as the excitation deviates from the K point.\textsuperscript{167}

In order to achieve the holy grail of valleytronics, it’s necessary to prepare and detect the valley polarization in an electrical way. One of the possible detection schemes is via the valley Hall effect\textsuperscript{168}.

Valley Hall effect is an electrical transport effect similar to the normal Hall effect, but instead of different charge, electrons of different valley index will drift to opposite directions under a longitudinal electric field. It is caused by the non-trivial Berry curvature distributed in the momentum space\textsuperscript{169}. When the electron or hole moves in a crystal along the electric field direction, the berry curvature will behave as a pseudo magnetic field in the momentum space and generate an additional velocity perpendicular to the electric field in the real space, known as the anomalous velocity. This is a high order correction to the group velocity, so the total group velocity can be formulated as below:

$$v_g(k) = \frac{\partial E}{\hbar \partial k} - \frac{eE}{\hbar} \times \Omega(k) \quad (9.1)$$

$v_g$ is the total group velocity in the crystal, $\frac{\partial E}{\hbar \partial k}$ is the group velocity induced by band dispersion, $E$ is the electric field and $\Omega(k)$ is the Berry curvature at the k point in that specific band.

Since both $E$ and $v_g$ are polar vector, $\Omega$ has to be an axial vector, which means in an inversion symmetric system $\Omega(k) = \Omega(-k)$. On the other hand, if there is a time-reversal symmetry in the system, $\Omega(k) = -\Omega(-k)$ has to be met. This result implies that in a system with both time reversal and inversion symmetries, Berry curvature has to be zero everywhere in the bands. Here, we choose the monolayer molybdenum disulfide as our material system to demonstrate the valley Hall effect, since its inversion symmetry is broken in the monolayer limit.

When a photon of certain helicity generates an electron-hole pair in one valley, in addition to a longitudinal photocurrent, the electron will contribute a transverse current, $j_{\sigma} = n \cdot E \cdot e \cdot \mu_{\sigma}$, with the equivalent transverse mobility $\mu_{\sigma}$ equal to $2 \text{ cm}^2 / (V \cdot s)$ according to the first principle calculation\textsuperscript{164}, while the hole will drift towards the other side of the sample with a comparable velocity. Here, $n$ is the
photo injected electron density; E is the electric field along the longitudinal direction and e is the electron charge. Consequently, in the equilibrium state, there will be a net transverse voltage across the sample to compensate the drift current. Thanks to the protection of time reversal symmetry, the berry curvature is of the same value but opposite sign in the other valley. Therefore, when the helicity of the photon is changed, the current flow direction will be reversed, and so is the net voltage. The scheme is depicted in the figure 9.3. When incident light is linearly polarized, both valleys will be populated, and no net transverse photocurrent will be generated. In this way, we can detect the valley polarization in the material by measuring the transverse voltage.

![Schematic of the valley Hall experiment](image)

*Fig 9.3: Schematic of the valley Hall experiment. Different circular polarized photon could excite the excitons in different valleys. Red spheres represent holes and blue spheres represent holes. Different green circles indicate they are situated in different valleys, consequently acquiring different orbital angular momentum. A back gate is used to tune the carrier density inside the channel.*

### 9.2 Device design and electrical performance

A number monolayer of samples based on the above design have been fabricated using the method described in Chapter 6.4. SEM image of a wire-bonded sample is shown in fig. 9.4. The MoS$_2$ channel is etched into a rectangular shape with a length to width ratio about 2:1. The typical channel width is 3-5 $\mu m$, which is more than twice the laser spot diameter we will use. The source and drain contacts are wide enough to cover the whole channel width. Depending on the monolayer size, one or two pairs of Hall contacts are fabricated in the device, which are typically 1.5$\mu m$ in width. All the contacts are made of 10 In/ 50nm Au, with an overlap by more than 1$\mu m$ with MoS$_2$ underneath it. The electrical doping is achieved through the back
gating. For the later reference, we label the drain and source contacts as 1 & 4, and the Hall contacts and 2, 3, 5 and 6.

![SEM images of a fabricated sample. (a) Low magnification image showing the electrode lines and the contacting pads. (b) High magnification image showing the drain/source contacts and the Hall contacts. Scale bar is 2µm.](image)

The electrical performance is measured in both ambient and vacuum condition. With In/Au contacts, the IV curve is a straight line, indicating a good Ohmic contact between the channel and contacts. (fig. 9.5 (a)) At zero back gating, $R_{14}$ is usually on the order of $M\Omega$. Every pair of contacts is characterized electrically to ensure a good ohmic contact over the experimental range. The transfer curve is also measured. (9.5 (b)) Our device is effectively a good transistor, with a large on/off ratio $>10^5$ demonstrated by the back gate tuning. The slope in the on state reveals that the free electron mobility in our device is about $19 \text{ cm}^2/(\text{V} \cdot \text{s})$. This value can be further improved by capping the top oxide gate according to the literature. The electron density level at this point is estimated to be $\sim 10^{12} \text{ cm}^{-2}$. Two curves with different voltage sweep directions are depicted in fig. 9.5 (b). The hysteresis between them is attributed to the water molecule absorption/desorption effect, as discussed in the following section.

![IV relationships between different pairs of contacts showing a good Ohmic contact. (b) The transfer curve of the device. Red line: $V_g$ is swept from negative to positive. Black line: $V_g$ is swept from positive to negative. Sweeping speed is 10 V/s.](image)
We study the contact resistance between MoS$_2$ and In/Au contacts with a 4-parallel-probe device. As depicted in fig. 9.6, both the two-probe (2p) and four-probe (4p) show a linear relationship. The difference between them is the contact resistance, summarized in the table 9.1. Evidently, not only the 2p and 4p resistance, but also the contact resistances are functions of the back-gate doping. This is expected to be originated from the electron density dependent transfer length.

![IV curves measured in two-probe (a) and four-probe (b) configurations from a parallel-probe device at different back-gate levels.](image)

<table>
<thead>
<tr>
<th>BG</th>
<th>0V</th>
<th>10V</th>
<th>20V</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{2p}$ ($\Omega$)</td>
<td>0.34M</td>
<td>0.14M</td>
<td>0.085M</td>
</tr>
<tr>
<td>$R_{4p}$ ($\Omega$)</td>
<td>0.28M</td>
<td>0.12M</td>
<td>0.074M</td>
</tr>
<tr>
<td>$R_{contact}$ ($\Omega$)</td>
<td>0.06M</td>
<td>0.02M</td>
<td>0.011M</td>
</tr>
</tbody>
</table>

*Tab. 9.1: Summary of the 2P, 4P and contact resistance at different back-gate levels.*

**9.3 Origin of the longitudinal photo current**

Experimentally, we observe four types of photo current when the device is illuminated by the laser: a slow one with a time scale of minutes; a fast one at the channel-contact edge independent of $V_{ds}$, one dependent on $V_{ds}$ but still at the edge, and one at the center of the channel. Here, we attribute them to four different origins: absorption and desorption of water molecules, photothermalelectric effect, Schottky effect and the photo injected carrier driven by the applied electric field.

The photocurrent response is measured in a liquid Helium cryostat mounted in an optical scanning microscope, as depicted in fig. 9.7. A 633 nm Helium-Neon laser used in the experiment to excite the interband transition. The scanning function is achieved via modulating the position of a piezo stage, on which an objective of 0.5 NA is mounted. As the piezo stage is moving in y and z direction, the focused beam spot is moving relative to the sample. At every scanning position, the back scattering of the laser is collected by a beam splitter with a photodetector, and finally a scattering image can be acquired together with the photo current mapping. In our experiment, depending on which pair of electrodes the electrical signal is measured,
either constant current is applied along the source and drain direction and voltage is measured, or constant voltage is applied and the photo induced current is measured. The cryostat is usually pumped down to $10^{-6}$ Torr and the temperature down to 10K has been tested.

![Schematic of the scanning photo current microscope](image)

*Fig. 9.7: Schematic of the scanning photo current microscope. An EO modulator is used as a dynamic half wave plate to modulate the polarization of the laser beam. Depending on experiment design, a polarizer or a quarter waveplate is placed before the objective.*

To calibrate the resolution of our system, we mount a silicon photo detector inside the cryostat. As the laser is scanning over the detector’s active area, a photocurrent can be measured and compiled into a 2D image in corresponding to the piezo stage position. Similar to the razor blade method, a sharp edge of the photoactive mapping is used to calibrate the diameter of the beam. (fig. 9.8). The cross section of the edge shows the beam spot is about 1.7 µm. This is achieved by expanding the laser beam in the free space to about 6mm in diameter, to utilize the full NA of the objective.

![Photocurrent mapping of an edge of the silicon detection](image)

*Fig. 9.8: Photocurrent mapping of an edge of the silicon detection. Cross section along the dashed line indicates the focus spot is 1.7 µm in diameter.*
At first, a large but very slow photocurrent is observed in the experiment. (fig. 9.9) In the constant voltage mode, as the laser is incident on the channel area, $I_{ds}$ will have a large increase. Nevertheless, as the light is turned off, the current doesn’t return to its initial value immediately. After half an hour long continuous dropping, the current will reach an equilibrium, still higher than the original value. We found that the higher optical power we use, the higher peak current will be reached. The long relaxation time is always on the order of hour. Similar effect is seen when the sample is heated in the vacuum environment. We test a moderate heating, up to 150 °C can increase the equilibrium current value. Similar effect is found when the sample is left in the vacuum condition for hours. Finally, when air is vented into the cryostat, $I_{ds}$ decreases dramatically, and even though the laser treat can still boost the current peak, the relaxation becomes shorter, reduced to approximately minute’s level.

![Fig. 9.9: A long time log of the slow photocurrent response in MoS$_2$ detector. The bottom figure shows the treatment to the sample. Red bars stand for the laser treatment in the channel, with the height representing the laser power. Yellow and blue bars mean the heating and air-venting action. The top plot shows the dynamic of the current level in correspondence to different treatments.](image)

We understand this slow photoresponse as the result of water molecule absorption/desorption$^{170-172}$. As we know, the MoS$_2$ exfoliated from natural crystals are n doped heavily and hydrophilic, in agreement with the bulk property. After the channel is fabricated into device and left in the ambient environment for some time, water molecules will adhere to the channel surface. These molecules are usually negatively charged, forming H$_2$O$^-$, and become immobile ions. These ions would expel the conducting electron in the channel, thus turning off the MoS$_2$ transistor.

On the other hand, when the laser is incident on the channel, electron hole pair, or the exciton, will be generated and the free holes would be easily captured by the water ion and render them into neutral ions. These neutral molecules would be easily desorbed from the channel and increase the conductance in the channel as a result. (fig. 9.10) The water molecules can also be removed by the heating or staying in the low humidity environment for long. On the other hand, the water molecule in the air will increase the possibility of absorption, thus reducing the channel...
conductance. All these trends can explain our experimental observation with the right time scale.

Fig. 9.10: A water ion absorption/desorption model can explain the slow photo response. After the exciton is generated by incident photon, the ion becomes neutralized and desorbed from the transistor channel.

Because the photo-injected carrier has an intrinsically short life inside the channel material, we expect to see a fast photo response from the device. In order to study the dynamics on the millisecond time scale, we add an EO modulator into the beam pass. The retardance is set to half wave, so the modulator functions a dynamic half wave plate, which periodically alter the polarization from vertical to horizontal. After it and before the objective, a polarizer is added, so that the polarization modulation is converted into periodic intensity modulation, usually at the frequency of 2 kHz. The fast optical response should be consequently modulated at 2 kHz, which is much slower than the RC constant of the device. Such a small modulation on top of a large DC background is finally picked out by a lock-in amplifier (Standford 830) and recorded as a photocurrent map. A photocurrent mapping from an early stage sample is depicted in fig. 9.11.

Fig. 9.11: An overlap image of the photocurrent mapping and the electrode positions. Scale bar: 2μm.

It is evident from the fig. 9.11, not all the photo responses are in the channel area. A large portion of photocurrent hotspot is located at the boundary of the contacts, which means they are not intrinsic photo injected induced conductivity. These extrinsic photo responses cannot contribute to the valley Hall effect, measured at the middle of the channel. Such an extrinsic component can be divided into two
parts, one is not dependent on the source-drain voltage, and the other is dependent on the polarity of the voltage applied.

Fig. 9.12: (a) Scattering mapping. (b) Photocurrent mapping at zero source-drain bias. (c) Schematic of the photothermal electric potential distribution across the device.

The first contribution can be seen more clearly when $V_{dd}$ is not applied. (fig. 9.12) At the zero bias condition, there are still multiple photocurrent hot spots at the contacts edge, not only of the source-drain contacts but also of the Hall contacts. More interestingly, the hot spots show different signs at the opposite sides of the contacts. The sides closer to the high potential contact show negative signals, whereas the sides closer to the low one show positive signals. We attribute these remote hot spots at zero bias to the photothermal electric effect$^{173}$. When the laser is focused at the contact edges, the local temperature is raised, which induces an electrical potential across the boundary and the boundaries at the opposite sides have opposite polarities. The voltage magnitude is dependent on the monolayer Seebeck coefficient, which is estimated to be extraordinarily large at the intrinsic doping level. The local heating effect has different power dependence compared with the photo carrier related effect, and at high power level ($I > 5kW/cm^2$), the photoresponse is found dominated by the thermoelectric signal.

Another origin of the edge photo response is the Schottky contact between the 2D semiconductor and the metal contacts$^{174}$. We confirm it by observing the bias-dependent photocurrent-hot-spots ratio at the source and drain sides. When MoS$_2$ and metal contacts don’t share the same work function, a Schottky barrier will occur at the interface. The electric field built at the interface would accelerate the photo carrier and provides a very large current. Such a response has a very similar trend as the photothermalelastic effect, as discussed above, except that it will be sensitive to the bias applied, which determines the local band bending at the contact.
In our Ti/Au contact sample, the band alignment is depicted in fig. 9.13. Schottky barriers will exist at both source and drain sides, and the built-in electric field can enhance the photo response. As shown in the figure, the electric field at the junction is also dependent on the voltage applied. At the high potential end, the barrier becomes sharper and electric field becomes stronger at the low potential end. As a result, the negative voltage contact always has a stronger photo response than the high voltage contact. And this is experimentally observed in our early devices. (fig. 9.11) When the polarity of the voltage is changed, the hotspot intensity ratio is also switched. In order to eliminate the Schottky barrier as much as possible, we replace Ti with In, which has a lower work function and can align better with MoS$_2$ in the conduction band.

By replacing the Ti/Au contact with In/Au contact and limiting the power to a low level, the photo response at the contact edge can be avoided to a great extent. One Indium contact sample mapping is depicted in fig. 9.14. Clearly, the majority of the hot spot is located in between the Hall contacts and away from any of the contacts. The dependence of the photovoltage (in this case, constant current source is applied) on the applied current strength and the backgate voltage level is depicted in fig. 9.15.
Fig. 9.15: (a) Amplitude of the photovoltage as a function of $I_{ds}$ (x axis), and back gate voltage $V_g$ (y axis). (b) The phase of the photovoltage relative to the EO modulation phase. A $\pi$ phase change, equivalent to the voltage sign change, occurs when the applied current is reversed.

The photovoltage is very small when $I_{ds}$ is small, and the phase channel from the lock-in amplifier, which represents the relative phase relation between the periodic electrical signal and the periodic modulation, becomes dominated by the random noise. When the device becomes more resistive, as the back gate voltage is increased in the negative direction, the photovoltage amplitude increases. This can be understood as a result of Ohmic relationship $\Delta V_{ph} = R \cdot \Delta I_{ph}$, $\Delta I_{ph}$ is a constant in this case. In order to get a measurable photovoltage signal, we tune the channel into the high resistive regime.

Fig. 9.16: (a) Electrical bridge model for explaining the measured transverse photovoltage pattern. Both the relative phase (b) and the amplitude (c) of the photovoltage show a four-fold symmetry.

Before we proceed to the helicity effect, it is interesting to investigate how the laser beam affects the transverse photovoltage. This is done by measuring the voltage across the Hall contacts, while a constant current is applied along the longitudinal direction and the incident laser's intensity is modulated periodically. Ideally, no
Photovoltage can be generated in the direction transverse to the applied current source, except a small projection of the longitudinal voltage due to the misalignment of the Hall contacts. Experimentally, we can see an interesting spatially four-fold photovoltage pattern. (fig. 9.16 (c)) Every neighboring hotspot pair acquires opposite signs, as the figure 9.16 (b) shows. This is due to the finite channel size seen by the focused laser spot and can be explained by a balance electrical bridge model. As depicted in fig. 9.16 (a), the whole channel can be viewed as a resistance network. When no light is applied, the Hall contacts are of equal potential. Whereas, as light is incident at any of the corner, the mirror symmetry of the channel is broken. In other words, the electrical bridge is no longer balanced. A current will flow from left to right, when the top left or bottom right corners are illuminated. An opposite current flow will happen when the top right and the bottom left corners are illuminated. The current flows means the Hall contacts are not in equal potential any more and a voltage can be measured across them. Such an observation double confirms the photo carrier origin of the measured photo voltage along the longitudinal direction.

9.4 Polarization dependence of the transverse photocurrent

After we can get the photo response from the middle of the channel, we park the laser spot at the center and investigate the polarization dependence of it.

In order to measure the valley Hall effect, which is essentially a helicity dependent transverse photovoltage, we change the modulation type from intensity modulation to polarization modulation, so the lock-in signal directly comes from the helicity information. This is achieved by replacing the polarizer with a quarter wave plate, so the laser polarization is periodically altered from $\sigma_+ \to \sigma_-$. An ideal modulation series is depicted in fig. 9.17 (b).

---

**Fig. 9.17:** (a) Time sequence schematic plot of the intensity modulation by an EO modulation together with a polarizer. Below is shown the corresponding periodic high voltage applied to the modulator. (b) An ideal Hall measurement modulation sequence should be a pure periodic helicity modulation. (c) Nevertheless, EO modulator will cause an unbalance pulse train due to the different transmissivity for orthogonally polarized light.
In the real experimental setup, the EO modulator has a transmission difference, which introduces an unwanted intensity modulation in addition to the polarization modulation. (fig. 9.17 (c)) This is compensated by fine-tuning the angle of the final beam splitter before the objective. Since $t_s$ and $t_p$ has different dependences on the incident angle, the intensity modulation can be removed, with the help of a photo detector directly mounted in front of the objective. Nevertheless, the EO modulator’s transmissivity difference is found to be a non-constant value (fig. 9.18), because it is caused by the continuous temperature drift within the EO crystal. As shown in fig. 9.18, the transmissivity has a very slow oscillation period that lasts about half an hour. After the thermal equilibrium is reached, the modulator will be stable for another 4 hours until the new oscillation starts again. Accordingly, we operate our experiment in this stable window. In the future, a feedback system is planned to be included in the system to increase the stability of the pulse.

![Fig. 9.18: The drift of the difference signal of EO modulation transmissivity for orthogonal polarized light. The stable period is about 4 hours long after 6 hours’ oscillation.](image)

A primary transverse photovoltage result is depicted in fig. 9.19. The measured signal is plotted as black dots, against the retardance, or the angle between the fast axis of the wave plate and the laser polarization axis. When zero retardance is applied, the polarization is not changed, which means the incident polarization sequence is $|\uparrow\rangle\rightarrow|\rightarrow\rangle\rightarrow|\uparrow\rangle\rightarrow$...When the wave plate is rotated by 45 degree, a quarter wave retardance ($\pi/4$) is introduced and the polarization sequence becomes $|\sigma_-\rangle|\sigma_+\rangle|\sigma_-\rangle|\sigma_+\rangle$...A $\pi/2$ retardance is equivalent to zero retardance. A $3\cdot\pi/4$ retardance changes the retardance sequence to $|\sigma_+\rangle|\sigma_-\rangle|\sigma_-\rangle|\sigma_+\rangle$... Interestingly, at the $\pi/4$ retardance condition, we observe a large negative transverse photovoltage; at the $3\pi/4$ condition, the large photovoltage changes its sign to a positive one. This agrees with our valley Hall effect picture that the $\sigma_+$ light would generate photo carrier in one valley and the finite Berry curvature will give
rise to a positive transverse photovoltage. On the other hand, the $\sigma_-$ light would inject the carrier into the other valley and generate a negative net voltage. Such an oscillating signal is filtered out as a dc voltage by a lockin amplifier. At $3 \cdot \pi / 4$ retardance, the sign is changed because the transverse voltage changes the sequence and is out of phase with the high voltage we applied in the EO modulator. Hence, the filtered dc voltage changes its sign. The $\pi$ retardance is equivalent to zero retardance and the signal pattern will keep repeating itself as we rotate the quarter wave plate. (fig. 9.19 (a))

![Fig. 9.19: (a) Transverse photovoltage as a function of the retardance from the quarter waveplate with a positive $V_{ds}$. (b) The curve is flipped when the applied voltage is reversed. (c) Longitudinal photovoltage showing a strong $2\omega$ dependence.](image)

In the figure 9.19 (b), we change the direction of the longitudinal electric field and find the sign of the signal is change. This agrees with our expectation that the transverse photovoltage signal is proportional to the longitudinal voltage.

Interestingly, we observe, at $\pi / 2$ and $\pi$ retardance, a small negative signal peak in fig. 9.19 (a), and this peak also changes it sign when the longitudinal field is versed in fig. 9.19 (b). Evidently, this peak has a repetition frequency twice and the valley Hall signal and reaches its maximum whenever the incident laser is linearly polarized. As a result, the total signal can be fit by a combination of a fundamental oscillation ($\omega$), and the $2^{\text{nd}}$ harmonic oscillation ($2\omega$), as the blue curve shows. The fitted amplitude between $\omega$ and $2\omega$ components is about 2.
We found the $2\omega$ signal is originated in the intensity modulation as the wave plate is rotated. As depicted in fig. 9.19 (c), the photo voltage also has a non-zero oscillation in the longitudinal direction, which should not happen if the intensity of the laser beam is a constant as in fig. 9.17 (b). The valley Hall effect shouldn’t either have a contribution in the direction parallel to the applied electric field. Therefore, the longitudinal signal can only come from a variable laser intensity modulation at different wave plate angle, which can be understood as a linear dichroism in the system. Our $2\omega$ transverse photovoltage component is a projection of such a longitudinal photovoltage. Compared with the intensity modulation experiment carried out in the same device, the undesired intensity difference in our polarized modulation experiment is about one percent of the total laser intensity. If we assume 5 percent of the longitudinal voltage will be projected onto the transverse direction, which is the ratio we measure in the DC experiment, the valley induced transverse signal is about 0.1 percent of the intensity induced longitudinal photovoltage, indicating $\mu_v \sim 0.001\mu_x \sim 0.01 \, cm^2 / (V \cdot s)$, which is much smaller than the theoretical prediction value.

Such a one-per-cent modulation may come from the linear dichroism of different optical components in the system. One possibility is the different transmissivity of the quarter wave plate, with their fast axis orthogonal to each other. The zero order quarter waveplate is essential made of two quartz crystals. Each quartz surface will give rise to about 0.4% transmission different between ordinary light and extraordinary light at 633 nm. The orthogonal configuration will allow a repetition of the modulation at every 90 degrees.

### 9.5 Conclusion and outlook

In this chapter, we have introduced the basic concepts of valleytronics and discuss the possibility of its implementation in TMDC system. We have shown a series of efforts to demonstrate the valley Hall effect, which is an electrical approach to access the valley polarization inside TMDC. MoS$_2$ based transistor is fabricated and four different photocurrent origins have been identified in it. Finally, a fast transverse photovoltage depending on the helicity of the incident photon has been measured, suggesting the existence of the valley Hall effect.

In the future, in order to get a purer valley Hall signal, a feedback system can be introduced to correct the long-term EO modulator drift. The rotational quarter waveplate can be replaced by a Soleil Babinet compensator, so as to reduce unwanted intensity modulation without sacrificing the polarization modulation.

Currently there’s an orders of magnitude of the discrepancy between the theoretically predicted valley Hall coefficient and the measured results in our experiments. The measured signal is two orders smaller than the expectation. We suspect this might come from neglecting the exciton effect inside TMDC, which could reduce the photocurrent by a large amount. Therefore, in addition to further
investigations of the origin of intrinsic photovoltage, we can cap the channel with other 2D material, such as boron nitride films, which can effectively reduce the exciton binding energy according to the theoretical prediction\textsuperscript{115}. Also, the carrier mobility is expected to be enhanced inside such a sandwich structure.

To form a complete valleytronic circuit, on the other hand, the electrical injection of valley polarized electron is also very critical and has not been demonstrated so far. It is difficult since the electrons injected from the metal contact are in the thermal equilibrium in the momentum space, and therefore populated over all the valleys. A point contact in TMDC with certain edge configurations has been proposed to be a good valley filter\textsuperscript{162}. Another possibility is to utilize the large spin orbital coupling within TMDC, which results in a one-to-one correspondence in spin and valley degree of freedom. By injecting a spin polarized hole current, we can effectively polarized the charge in side the channel.
Reference:

46. Okołowicz, J., Płoszajczak, M. & Rotter, I. Dynamics of quantum systems
67. Aieta, F. *et al.* Aberration-free ultrathin flat lenses and axicons at telecom


107. Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V. & Kis, A. Single-layer


146. Crain, J. N. & Pierce, D. T. End states in one-dimensional atom chains. *Science*


