Polariton Photonics Using Structured Metals and 2D Materials

Ziqiang Cai, Yihao Xu, Chuangtang Wang, and Yongmin Liu*

1. Introduction

Polaritons are quasiparticles originating from strong interactions between photons and elementary excitations that could enable high tunability, tight electromagnetic field confinement, and large density of photonic states, making it possible to achieve novel and otherwise inaccessible functionalities. For these reasons, polaritons have attracted great interest in the fields of physics, materials science, and optics for both fundamental studies as well as potential applications (e.g., modulators, photodetectors, photoluminescence, etc.). In recent years, the explosive growth of research in graphene and other 2D van der Waals materials is witnessed because they provide a new platform that substantially complements conventional metals, dielectrics, and semiconductors to investigate different polariton modes. This review highlights the works published in recent years on the topic of polariton photonics based on structured metals, graphene, and transition-metal dichalcogenides (TMDs). The exotic optical properties of the polaritons in metallic structures and 2D van der Waals materials offer bright prospects for the development of high-performance photonic and optoelectronic devices.

2. Fundamentals of SPPs

Over the past decades, nano-optics based on SPPs—termed plasmonics—has become an extremely dynamic and fruitful research area, because it offers new opportunities for technology advancement in a wide range of applications, which includes spectroscopy, sensing, imaging and focusing, lithography and on-chip circuit design. These applications are made possible by SPPs’ properties of light confinement beyond the diffraction limit and also high field enhancement in the near-field region. SPPs behave as optical surface waves confined in the proximity of a metal surface, arising from the resonant interaction between electrons and EM waves. Since there are already many papers that give excellent reviews on SPPs, here we just give a brief introduction to the basic properties of SPPs. As shown in Figure 1a, SPPs are transverse magnetic (TM) polarized waves with the magnetic field perpendicular to the plane of incidence. SPPs propagate at the interface between a semi-infinite dielectric medium and a noble metal, and the electric field intensity exponentially decays into both the dielectric and metal (Figure 1b). From Maxwell’s equations and boundary conditions, we can readily prove that the in-plane wavevector $k_{||}$ of SPPs is given by...
where \( \varepsilon_d \) and \( \varepsilon_m \) are the relative permittivity of the dielectric and metal respectively. The frequency-dependent relative permittivity of metal \( \varepsilon_m \) can be expressed by the Drude model

\[
\varepsilon_m = 1 - \frac{\omega_p^2}{\omega^2 - i\Gamma}\quad (2)
\]

where \( \omega_p \) is the plasma frequency of the metal and \( \Gamma \) is the damping rate. Substituting Equation (2) into Equation (1), we can obtain the expression between \( k_b \) and frequency \( \omega \), which is also called the dispersion relation. In Figure 1c, we give an example of the dispersion relation of SPPs supported by an air–silver interface, considering the Drude model of silver.\(^{[19]}\) We can see that the in-plane wavevector \( k_b \) of SPPs is larger than the wavevector of light in free space over a broad frequency range. As a result, the out-of-plane wavevector \( k_z \) is an imaginary number, causing the wave to attenuate into the surrounding medium. The dispersion curve, however, is bent back to the light line near the asymptote around the near UV wavelengths.\(^{[22]}\)

The loss in SPPs is usually unavoidable and is mainly caused by free-electron scattering in metals, and absorption induced by interband transitions.\(^{[20]}\) Losses from free-electron scattering are strongly dependent on the roughness and defects of metals, which cannot be eliminated but can be improved to some extent with advanced fabrication techniques. The absorption losses can be greatly reduced by selecting a proper working wavelength. The impact of loss is that it limits the propagation distance of SPPs. To extend the propagation distance of SPPs, researchers have introduced two dielectric-metal interfaces using a thin metal film, enabling a mode splitting and generating a long-range SPP mode with even symmetry and a short-range SPP mode with odd symmetry, in terms of the transverse electric field distributions along the surface normal direction.\(^{[21]}\) On the other hand, some researchers have tried to utilize the “undesired” intrinsic loss of SPPs to achieve novel functions such as mode selection\(^{[22]}\) and unidirectional excitation,\(^{[23]}\) based on the Parity-Time symmetry notion in quantum mechanics.

Because of the wavevector mismatch, we cannot directly generate SPPs using free-space light. There are three common methods to excite SPPs, as shown in Figure 2. In the Kretschmann configuration (Figure 2a), free-space light is incident with angle \( \theta_{spp} \) on a dielectric prism with refractive index \( n \) to generate a horizontal wavevector that matches \( k_{spp} \), that is, \( nk_0 \sin \theta_{spp} = k_{spp} \). Consequently, SPPs on the top air–metal interface can be excited. In Figure 2b,c, scattering from gratings and surface structures can also compensate the momentum mismatch. In the case of a grating, the additional momentum could be simply expressed as \( \Delta k = N \cdot 2\pi/\Lambda \), where \( \Lambda \) is the period of the grating and \( N \) is any integer. In the case of a surface feature, the near-field component of the scattering field can excite SPPs as well, while the efficiency is generally low. Such an excitation scheme is employed in scattering-type scanning near-field optical microscopy (SNOM), in which a sharp tip can be used to excite, and detect the near field of SPPs simultaneously.\(^{[24,25]}\)

3. Manipulating SPPs with Nanostructures

3.1. Unidirectional Excitation of SPPs

Because the energy of SPPs is confined at the metal-dielectric interface and its dispersion demonstrates its ability to break the diffraction limit, SPPs are considered as one of the most promising candidates for building on-chip optical circuits. Efficient and controllable excitation and subsequent routing of SPPs are very important in this regard. In this section, we...
will focus on unidirectional excitation of SPPs, which not only shows the desired direction control of SPPs, but also provides high coupling efficiency since the energy is harnessed in only one direction.

The most direct way to achieve directional excitation is by using the interference of two SPP sources based on asymmetric structures.[26–28] For example, G. Lerosey et al. studied a system of dual dielectric cavities in a metal slab (upper diagram in Figure 3a) that can generate SPPs from light impinging on the cavities in parallel.[30] By selecting the widths of both cavities as well as the distance between the two cavities, they achieved constructive interference for the wave propagating toward the left (bottom diagram in Figure 3a). A simpler configuration was used by J. Chen et al. with only one asymmetric single-nanoslit that works like a Fabry–Pérot nanocavity, as shown by the scanning electron microscope (SEM) image in Figure 3b.[30] In this case the phase would accumulate when the SPPs go around the cavity once, which is represented as \( \Phi = 2k_{spp}L_{FP} + \varphi \), where \( \varphi \) is the phase shift by the edge of the Fabry–Pérot nanocavities. If the phase shift is an integer of \( 2\pi \), then constructive interference occurs on the right. In this experiment, the authors achieved an extinction ratio (defined as the power ratio of left- and right-propagating SPPs) of about 30:1 and an SPP generation efficiency of 1.8 times the efficiency of a symmetrical slit, at the wavelength of 830 nm. In 2012, Y. Liu et al. theoretically conceived and experimentally demonstrated the first directional SPP source with two magnetic nanoantennas with detuned resonant frequencies.[31] The arrangement of magnetic antennas and experimental results are presented in Figure 3c. Since the resonance phase and amplitude can be easily tuned by using proper geometric parameters of each antenna, the antennas can show directional excitation either toward left or right, depending on the distance between antennas. Based on this configuration, the left- and right-propagating SPPs can show an intensity contrast of 10.

In recent years, a whole new field of study was created based on the design of subwavelength nanoscale structures (“meta-atoms”) that can be arranged to construct metamaterials and metasurfaces. They can show extraordinary properties in response to EM waves, including optical waves, that can be very different from any known natural materials.[32] Various novel functions and applications have been reported, including refractive indices that are negative[31,34] zero[35,36] or positive and extremely large[37,38] and more interestingly, the ability to perform subdiffraction-limited imaging[38–41] and also create an invisibility cloak.[42–46] By manipulating the meta-atoms and their spatial distribution (i.e., distance, rotation angle, and density) on the metasurface, one can control the amplitude, phase, polarization and trajectory of EM waves in real space.[47–49] Soon after these concepts were proposed and demonstrated, researchers started to create metasurfaces to control the interactions between EM waves and SPPs.[54–56] As a result, we can go beyond the methods of SPP excitation and guidance mentioned above[56,57] to enable new capabilities and applications. In the following, we will review the recent progress of how metasurfaces help to achieve extraordinary functionality and performance for plasmonics.

The first example is a specific gradient index metasurface to convert free-space waves to surface waves in the microwave region, which was demonstrated by S. Sun et al. in 2012.[61] The idea was to design a spatial surface current distribution to match that of the surface wave. The unit cell is a subwavelength metallic H-shape structure on top of a metal plate separated by a dielectric spacer, as shown in Figure 4a (top left panel). The induced surface current phase is given by

\[
\Phi = \cos^{-1} \left( \frac{-e + \mu \tan^{-1} \left( \frac{\sqrt{e\mu}k_d}{} \right)}{e + \mu \tan^{-1} \left( \frac{\sqrt{e\mu}k_d}{} \right)} \right)
\]

which can be tuned by changing
the size of H-shape structures and hence $\varepsilon$ and $\mu$ in the equation. Consequently, it is straightforward to build the phase gradient of the surface wave, enabling the directional coupling from incident waves to surface waves. The bottom left panel in Figure 4a shows the measured and simulated electric field of the excited surface wave, and the right panel provides the far-field pattern, showing that the energy is efficiently coupled to the evanescent surface wave when the incident angle is larger than the critical angle $\theta_c$. The coupling efficiency is up to 98% in simulations. The authors emphasized that the surface wave here is different from SPPs (which are eigenmodes of the metal-dielectric interface), but can be converted to SPPs. The surface wave can only propagate under the illumination of certain incident waves. Nevertheless, it has been proved that, due to the similar surface confinement and dispersion, this surface wave can be guided out to SPP modes in other systems. Apart from the phase modulation, metasurfaces can also enable the control of SPPs by polarization, as demonstrated in the work of L. Huang et al. They used a metasurface consisting of aperture arrays as shown in the left panel of Figure 4b. By tailoring the orientation of each aperture element, there would be an additional phase for the incident right-circularly polarized (RCP) and left-circularly polarized (LCP) light with a different sign. And the phase difference is only dependent on the rotation angle, which is known as the geometric phase or Pancharatnam–Berry (PB) phase. When the different unit cells are...
arranged in an array with a certain period, it would provide a different momentum comparing to the simple grating array we mentioned in the introduction given by

$$\Delta k = \frac{2\pi}{\Lambda} \left( N + 2\sigma \frac{\Delta \phi}{2\pi} \right)$$

(3)

Here $\Lambda$ is the period and $\Delta \phi$ is the rotation angle difference between neighboring apertures. $\sigma$ indicates the helicity, which is $+1$ for RCP and $-1$ for LCP. Therefore, by carefully designing the phase gradient of the apertures, the additional momentum can only match the RCP to the right propagating SPPs or LCP to the left propagating SPPs. In other words, the ratio of excited SPPs to opposite directions can be simply manipulated by the ellipticity of the incident wave. This is the controllable directional excitation of SPPs by the helicity and wavelength of incidence. The authors have presented convincing experimental results in Figure 4b. Therefore, the electric field of the incident light with any polarization can be decomposed to excite the dipoles in orthogonal directions as two SPP sources. The overall SPP intensity along the left and right directions can then be calculated as $I_{\text{R/L}} \propto C(E_1^2 + E_2^2 + 2E_1E_2\sin \delta)$, where $C$ is a constant that depends only on the orientation of the apertures, $E_1$ and $E_2$ are the electric field components perpendicular to the two apertures and $\delta$ is a relative phase difference of $E_{1,2}$ due to the polarization. One can see that the polarization would not affect the total intensity $I_{\text{R}} + I_{\text{L}}$, while it can control the relative intensity of the SPPs field along the left and right directions. Particularly, in the case of pure LCP or RCP light, a unidirectional excitation of SPPs can be achieved with a 2D array of the “two-aperture” unit cell (bottom panel Figure 5a). The authors reported a maximum coupling efficiency and extinction ratio of $\approx 3.2\%$ and 26, respectively, at the wavelength around 620 nm. In the same year, Francisco Rodríguez-Fortuño et al. reported another method to control the SPP excitation with polarization. They demonstrated that a circularly polarized dipole with $p_x$ and $p_z$ components can excite the directional SPPs along the $x$-direction arising from the near-field interference, as shown in the top left

**Figure 5.** a) Schematic of subwavelength apertures as polarization-selective SPP plane-wave sources (top panel) and SEM image of the fabricated sample along with the measured near-field image of excited SPPs (bottom panel). b) Illustration of the circularly polarized dipole over a metal-dielectric interface (top left panel), leakage radiation microscopy results under different polarization states (top right panel) and simulated directional excitation results (bottom panel). c) SEM image for L-shaped optical slot nanoantennas (top panel) and measured SPPs excitation with $\sigma = \mp 1$ incidence (bottom panel). d) Schematic of the metal-dielectric interface with Parity-Time modulation (top panel) and simulated SPPs excitation with different modulation amplitude $V_0$ (bottom panel). a) Reproduced with permission. Copyright 2013, American Association for the Advancement of Science. b) Reproduced with permission. Copyright 2013, American Association for the Advancement of Science. c) Reproduced with permission. Copyright 2014, Wiley-VCH. d) Reproduced with permission. Copyright 2017, American Physical Society.
panel of Figure 5b. The bottom panel of Figure 5b presents the simulated directional SPPs with \((p_x, p_y) = (1, -0.705i)\), showing almost no back-propagating SPPs. By simply changing the relative phase between \(p_x\) and \(p_y\), which is highly dependent on the incident polarization, we can manipulate the contrast between left- and right-propagating SPPs. Similarly, J. Yang et al. also used the orthogonal apertures structure to realize directional excitation of SPPs.\(^{[65]}\) They built an L-shape aperture array on a thin gold film deposited on a silica substrate (Figure 5c, top panel), which would show a single-surface first-order plasmon resonance (SFPR) and a second-order plasmon resonance (SPR) mode at similar wavelength under normal incident light. The SFPR mode is simply the SPPs on the air–gold interface while SPR mode is a quadruple-like hybrid mode between air–gold and silica–gold plasmonic resonances. The interference of the SFPR dipole mode and SPR quadruple mode results in directional excitation of SPPs in either \(x\) or \(y\)-direction, as shown in the bottom panel of Figure 5c, which is readily controllable by the polarization of the incident light. The experimentally measured maximum intensity contrast is about 29.5. Interestingly, the inverse function can be realized using the same nanostructure array: If the incidence is a SPP wave, it can be coupled out by the aperture array and converted into free-space light with a certain helicity. It is noted that, although the incident RCP wave can excite SPPs propagating along the \(y\)-direction, SPPs incident from the \(y\)-direction will be coupled out as LCP, because the inverse process is like a time-reversed into free-space light with a certain helicity. It is noted that, a SPP wave, it can be coupled out by the aperture array and converted into free-space light with a certain helicity. It is noted that, although the incident RCP wave can excite SPPs propagating along the \(y\)-direction, SPPs incident from the \(y\)-direction will be coupled out as LCP, because the inverse process is like a time-reversal operation and hence the helicity would change.

All the works described above use straightforward interference notions of cavities or multiple nanostructure arrays, while there are alternative approaches, such as making use of valleytronics\(^{[66–68]}\) that will be discussed in Section 4.3, and Parity-Time (PT) symmetry, to realize directional excitation of SPPs. In 2003, Bender et al. showed that a Hamiltonian with PT symmetry can also possess real eigenvalues for energy states under certain conditions in a non-Hermitian system. Thanks to the equivalent mathematical form between the Schrödinger equation and paraxial equation of diffraction for EM waves, people can investigate PT symmetry on a photonic platform by manipulating material gain and loss, that is, complex refractive indices.\(^{[69]}\) When the gain/loss level is below a certain threshold, known as the exceptional point (EP), the eigenfrequencies could be real.\(^{[70, 71]}\) The characteristics of the optical system can be very different when crossing the EP. Interestingly, strictly balanced gain and loss are not necessary for the existence of EP. Due to gauge transformation, one can easily find that a passive system without gain materials can have the same energy eigenstates and also show the phase transition across the exceptional point as long as the loss difference is maintained.\(^{[22, 72]}\) Therefore, the “undesired” loss in the optical system could be utilized to realize extreme phenomena near the EP by tailoring complex dielectric constants such as unidirectional invisibility\(^{[73]}\) and single-mode lasing.\(^{[74, 75]}\)

In 2017, W. Wang et al. demonstrated the possibility to excite directional SPPs based on PT symmetry.\(^{[83]}\) The system consists of a metal-dielectric interface, in which the permittivity of the dielectric layer has a spatial PT symmetric distribution (Figure 5d, top panel), given by \(\varepsilon(x) = \varepsilon_d + A\cos \beta x + i\beta_0 \sin \beta x\), where \(\varepsilon_d\) is the background permittivity, \(\beta\) is the wavevector of the SPPs mode and \(\beta_0\) represents the modulation amplitude. In the simulation, the symmetric excitation of SPPs when \(\beta_0 = 0\) gradually evolves to unidirectional excitation toward the left as \(\beta_0\) increases to 1 (Figure 5d, bottom panel), which is the EP in this case. The authors analyzed this phenomenon by the mode coupling method and proved that the left/right contrast is proportional to the \(\pm 1\) order Fourier coefficient of the permittivity. Therefore, the periodically modulated permittivity can only provide \(\beta\) wavevector (or momentum) at \(\beta_0 = 1\), which only matches the left propagating mode of SPPs. More interestingly, the scattering loss of SPPs can be completely suppressed, because of the lack of necessary momentum to couple the directional SPPs back to free space. This work manifests a new approach to controlling the propagation and loss of SPPs without structuring metals.

### 3.2. Focusing and Shaping SPPs

Once SPPs are excited along the interface, the next step is to manipulate the propagation and field distribution of SPPs. For example, in order to make use of the subwavelength features of SPPs to break the diffraction limit, researchers have introduced different methods to focus SPPs into extremely small volumes, known as nanofocusing of SPPs, by using plasmonic wedges,\(^{[76]}\) conical structures\(^{[77]}\) and tapered waveguides.\(^{[79]}\) At the end of these structures, the electric field would increase anomalously due to the small mode volume, leading to a large local density of photonic states. More interestingly, the wavelength of SPPs can theoretically decrease to zero as it approaches the end point, rendering superfocusing effect promising for various applications where high field enhancement is required.\(^{[76–78]}\) In 2009, H. Choi et al. demonstrated experimentally nanoscale focusing in a V-groove structure as shown in Figure 6a.\(^{[79]}\) They used a metal-dielectric-metal structure since it does not experience mode cut-off at the nanoscale for SPPs. The SPPs would decrease its wavelength to 1/40 of free-space wavelength with a \(\approx 50\%\) power efficiency at the end of the tapered structure. A similar nanofocusing effect has been achieved on a SNOM tip by asymmetric grating structures.\(^{[80]}\) A breakthrough occurred in 2015 when V. A. Zenin et al. numerically and experimentally demonstrated significant local-field enhancement \(\approx 12,000\) by a 2D tapered-triangular nanofocusing (TSNF) made of gold on silica substrate.\(^{[81]}\) In this work, they designed gold stripe with a grating array and a tapered end with a narrow lateral dimension (Figure 6b) to channel and confine the SPP mode. The grating array is used to couple the free-space wave into SPPs. Near the tip of the TSNF, an impedance-matched nanowire antenna is coupled across a gap, in analogue of the Fabry–Pérot interferometer. It is noted that with this special design, the back reflection of SPPs can be greatly reduced due to the resonance of the combined structure.

The implementation of nanostructures and metasurfaces can also help to focus SPPs.\(^{[82, 83]}\) The early work from T. Tanemura et al. used several rectangular nanoslits to couple incoming light to SPPs as schematically illustrated in Figure 6c. Multiple wavelength focusing can then be achieved by precisely adjusting the position of each slit through iterative optimization algorithm that optimizes the phase at each focal point. The previously mentioned aperture array structure
proposed by Lin et al. can also focus SPPs in addition to the excitation direction control if the linear array is bent to form a ring.\cite{63} As a result, the polarization of incident light can manipulate the SPP excitation to be inward (focusing at the center) or outward (scattering). However, in this case, the excited phase on aperture at different angles would be different according to the orientation, thus the focus is not perfect. Later work by F. Huang et al., addressed these issues by rotating an additional angle for apertures at different positions, which would add a geometric phase to the excited SPPs.\cite{85} The authors also showed that an Archimedes-curved array (Figure 6d), instead of circular distribution of the same apertures, can focus SPPs to a centrally symmetric field with either linear, elliptical or circular polarized light.

Recently, using metasurfaces to produce hologram images has become a hot area of research.\cite{86} Novel applications have been achieved like 3D,\cite{87} highly efficient\cite{57} and multi-wavelength achromatic\cite{88} holography. However, the in-plane hologram is challenging\cite{89} because, unlike the out-of-plane case, the incident wave would be scattered several times inside the designed nanostructures. In 2015, D. Wintz et al. proposed their holographic metalens to focus SPPs with different wavelengths at different focal points.\cite{90} In their work, they first assumed two imaginary emitters with different emission
wavelengths located at the desired focal points, and the gratings were fabricated on each equiphase position of two emitters as shown in Figure 6e. If the plane waves of either wavelength are incident on the grating arrays, the excited SPPs can be focused at the corresponding focal points. They also introduced additional degrees of freedom by arranging two grating arrays together along the vertical and horizontal directions, achieving four focal points for four different wavelengths. The focal points can be turned on and off by manipulating the polarization of the incident beam with respect to orientation of the two grating arrays.

There are other works that shape SPPs into more complicated spatial distributions, such as vortices, asymmetric focusing, Bessel-like beam and realization of plasmonic Luneburg and Eaton lenses. In 2015, A. High et al. published the experimental work on hyperbolic metasurfaces that show exotic phenomena in manipulating SPPs, which were first proposed and numerically studied by Y. Liu and X. Zhang. A. High and his co-workers demonstrated negative refraction of SPPs, diffractionless SPP propagation and a plasmonic spin-Hall effect in the visible region, which are shown in the left, middle and right panel of Figure 6f, respectively. All the effects are attributed to the highly anisotropic nature of the hyperbolic metasurface.

4. Polaritons in 2D Materials

4.1. Optical Characteristics of Graphene SPPs

2D materials have attracted intensive attention in the past few decades thanks to their novel electronic, optical, mechanical and thermal properties. Graphene, which is one representative 2D material, has been explored for a wide range of photonic applications, largely because of its ability to alter its optical and electronic properties under external stimuli. It provides an exciting, atomically thin platform to study polaritonics, potentially leading to applications that require modifications of optical responses at short timescales.

Graphene is a 2D material consisting of carbon atoms that are arranged in a honeycomb lattice. The electronic band diagram of graphene is shown in Figure 7a. At low energy, the band structure is composed of two cones with the upper (conduction band) and lower (valence band) cones connected at a point. The cone-like band structure is called the Dirac cone, and the point at which they connect is called the Dirac point. The electrons occupy the states below the Fermi energy $E_F$, which can be tuned via different doping methods. The Dirac point is where the Fermi level lies when there is zero doping, which is also called the charge neutral point (CNP). If we draw a

![Figure 7](https://www.advancedsciencenews.com/...)

**Figure 7.** a) Left: the schematic of the electronic band diagram of graphene. Right: band diagram close to the Dirac point. The vertical red arrows show that only photons with energy larger than $2|E_F|$ can excite electron–hole pairs. b) Real part of graphene conductivity calculated using Equation (4), with parameters $\mu = 1500$ cm$^2$ V$^{-1}$ s$^{-1}$, $T = 300$K. Dashed lines show the intraband contribution. c) Real part of graphene conductivity for different mobility, with parameters $E_F = 0.3$ eV, $T = 300$ K. d) Graphene SPP dispersion (red curve) calculated using Equation (8), with parameters $\mu = 10^5$ cm$^2$ V$^{-1}$ s$^{-1}$, $E_F = 1$ eV, $\varepsilon_1 = 1$, $\varepsilon_2 = 4$. The red region corresponds to interband interaction, while blue region corresponds to intraband interaction.
cross-section of this Dirac cone, we can get the linear dispersion relation \[ E = \hbar v_F q \] of graphene electronic band, where \( v_F \) is the Fermi velocity and \( q \) is the in-plane wavevector. The interaction between light and graphene can be classified into two types: intraband interaction and interband interaction, as shown in Figure 7a. For intraband interaction, the incident photon energy is low, and electrons remain within the same band. In this case, graphene behaves like a 2D metal. For interband interaction, the incident photon energy is high, and electrons jump from the valence band to the conduction band. In this case, graphene behaves like a semiconductor. Whether intraband or interband interaction occurs depends on the photon energy \( \hbar \omega \) and the relative amplitude of Fermi energy \( |E_F| \), which is the energy difference between the Fermi level and Dirac point.

The key parameter to quantitatively describe the optical characteristics of graphene SPPs is the conductivity of graphene. Under random-phase approximation (RPA) and local limit \( (k_F \rightarrow 0) \), the frequency-dependent surface conductivity of graphene is\[98,99]\n
\[
\sigma(\omega) = \frac{ie^2v_Fk_B}{\hbar \omega(\omega + i\tau^{-1})} \left[ \coth \left( \frac{E_F}{2k_BT} \right) \right] + \frac{e^2}{4\hbar} \left[ 1 + \frac{1}{\pi} \arctan \left( \frac{\hbar \omega - 2E_F}{2k_BT} \right) \right] - \frac{i}{2\pi} \ln \left[ \frac{\hbar \omega + 2E_F}{\hbar \omega - 2E_F} \right] + 4(k_BT)^2 \quad (4)
\]

In this equation, \( k_B \) is the Boltzmann constant, \( T \) is the temperature, \( \tau^{-1} \) is the damping rate, and \( \tau^{-1} = \frac{e^2}{\mu E_F} \), where \( \mu \) is the carrier mobility of graphene, and \( E_F \) is the amplitude of Fermi energy, which equals to \( \hbar v_F \sqrt{n} \) with \( n \) being the carrier concentration.

Equation (4) can be divided into two parts. The first term of the equation, which has a Drude model form, describes the optical response due to intraband interaction; while the remaining three terms describe the optical response due to interband interaction. The dashed lines in Figure 7b,c represent the results calculated from the first term of Equation (4), and the solid lines are the calculation from the entire equation. It is apparent that at low frequencies, intraband interaction dominates, thus the optical response of graphene can be precisely described by only considering the intraband term. In contrast, interband interaction dominates at high frequencies, and the graphene conductivity shows a constant value, namely \( e^2/4\hbar \), corresponding to a constant absorption of about 2.3%.\[100,101\] The threshold between interband and intraband interaction, marked by vertical red dashed lines in Figure 7b,c, is \( |E_F| \), which can be derived from the last two terms of Equation (4). This \( |E_F| \) threshold can also be easily understood by checking the right diagram of Figure 7a. Since the electrons can only jump to the un-occupied states due to the Pauli-exclusion principle, the electrons have to absorb energy of at least \( |E_F| \) for the interband interaction.

The Fermi energy \( E_F \) and carrier mobility \( \mu \) are the key parameters that determine the optical characteristics of graphene, as shown in Figure 7b,c. These two parameters can be measured in multiple ways.\[102–104\] In general, large \( E_F \) and \( \mu \) values are preferred, especially for infrared and terahertz applications. The value of \( \mu \) indicates the quality of graphene, and it is a crucial factor governing the lifetime of graphene SPPs. Depending on the fabrication process, the transfer method, and the dielectric environment of graphene, the carrier mobility \( \mu \) varies from the order of \( 10^3 \) cm\(^2\) V\(^{-1}\) s\(^{-1}\) for suspended\[105\] or h-BN encapsulated exfoliated graphene,\[106,107\] to the order of \( 10^8 \)–\( 10^9 \) cm\(^2\) V\(^{-1}\) s\(^{-1}\) for centimeter-scale CVD graphene\[108–110\] or \( 10^6 \) cm\(^2\) V\(^{-1}\) s\(^{-1}\) for hundreds of micrometer-scale CVD graphene.\[111\] The Fermi level \( E_F \), on the other hand, determines the carrier concentration of graphene. There are two commonly used methods to tune \( E_F \); namely electric tuning\[112–114\] and chemical tuning.\[121–123\] We should note that \( E_F \) can also be influenced in other circumstances, such as ultraviolet light illumination,\[124\] metal contacts\[125\] and high-intensity nonequilibrium electron–hole pairs excitation.\[126\]

The dispersion relation of graphene SPPs, derived by solving Maxwell’s equations with boundary conditions\[127\] or finding the pole of Fresnel reflection coefficient,\[101\] can be expressed as

\[
\frac{\varepsilon_1}{\sqrt{k_0^2 - E_F^2}} + \frac{\varepsilon_2}{\sqrt{k_0^2 - E_i^2}} = -\frac{\sigma(\omega)i}{\omega\varepsilon_0}
\]

Here \( \varepsilon_0 \) is the vacuum permittivity, and \( \varepsilon_1, \varepsilon_2 \) are the relative permittivities of the surrounding media above and below graphene, respectively. For \( k_i \gg k_0 \), Equation (5) can be simplified to\[101,127\]

\[
k_i = i\varepsilon_0 (\varepsilon_1 + \varepsilon_2) \omega / \sigma(\omega)
\]

Also, as mentioned above, when intraband interaction dominates, we only need to consider the first term of Equation (4). At room temperature (i.e., \( T = 300 \) K), \( k_B T = 0.026 \) eV \( \ll E_F \) for moderate doping, thus Equation (4) can be simplified to

\[
\sigma(\omega) = \frac{e^2}{\pi \hbar} \frac{E_F}{\omega + i\tau^{-1}}
\]

Combining Equations (6) and (7), we can get a more explicit dispersion relation for graphene SPPs

\[
k_i = \frac{\pi \hbar^2}{e^2E_F} e_0 (\varepsilon_1 + \varepsilon_2) \omega / (\omega + i\tau^{-1})
\]

The calculated dispersion of graphene SPPs is depicted in Figure 7d. The wavevector \( k_i \) of graphene SPPs is much larger than the free-space wavevector, which would be very close to the \( \gamma \)-axis if we plot it in the same figure. Consequently, the plasmon wavelength \( \lambda_p \) is much smaller than the free-space wavelength, indicating extreme light confinement.

The ultrahigh confinement, long plasmon lifetime and gate tunability of graphene SPPs have been comprehensively studied through SNOM measurements on high mobility graphene.\[24,25,106,107\] Especially, by combining SNOM with a pump-probe setup, the nonequilibrium SPPs in graphene, in response to a femtosecond laser pulse, are revealed under a low gate voltage that otherwise cannot excite graphene SPPs in equilibrium situations.\[106\] Moreover, by taking the SNOM measurements under the liquid-nitrogen temperature, propagation length of more than 5 \( \mu \)m is observed in h-BN encapsulated high mobility graphene. It is also shown that the intrinsic limits of
graphene plasmonics are the losses from acoustic-phonon scattering and intervalley electron–phonon scattering, corresponding to an intrinsic propagation length of more than $13\,\mu\text{m}$, while the dielectric environment contributes to the extrinsic limits.\textsuperscript{107}

### 4.2. Applications of Graphene SPPs

Tremendous efforts to explore its applications have been taken ever since the first observation of stable monolayer graphene under ambient conditions.\textsuperscript{128} Researchers have been exploiting the potential of graphene SPPs in various applications, such as graphene optical modulators and graphene sensors, which will be discussed in this section.

For optical modulators, graphene naturally seems to be a suitable material for this application because of its tunable optical properties. An ideal optical modulator needs to possess both low insertion loss and high modulation depth. However, due to the large wavevector mismatch between free-space light and graphene SPPs, it is nontrivial to achieve strong interactions between light and graphene for efficient optical modulation. To overcome this issue, two major approaches have been implemented: 1) combining graphene with plasmonic structures made of noble metals; 2) patterning graphene into periodic arrays. A metallic structure can increase the light-graphene interaction through its large local EM field enhancement, while patterned periodic graphene arrays can provide additional wavevector to compensate the wavevector mismatch between light and graphene SPPs.

The first experimentally demonstrated graphene optical modulator was reported in 2011\textsuperscript{129} in which graphene was combined with a silicon waveguide to substantially increase graphene-light interactions. Following this pioneering work, numerous works were presented using either one of the two aforementioned solutions. For the first approach, different metallic metasurfaces have been applied on top of graphene\textsuperscript{98,130–133} in which deep subwavelength ($<100\,\text{nm}$) air gaps are created to increase the
local EM field intensity, and the working wavelength is determined by the resonant wavelength of plasmonic metasurfaces. By introducing a metal reflector at the bottom, which creates a so-called metal–insulator–metal (MIM) structure, the local EM field intensity can be further increased, together with the decrease of gate insulation layer thickness, leading to a higher modulation efficiency. For the second approach, graphene is patterned into nanoribbon[134] or nanodisk[135–139] arrays. The resonant wavelength of the patterned graphene is proportional to the square root of the width of nanoribbon (or diameter for nanodisk). By introducing a metal reflector at the bottom and making the thickness of middle insulation layer a quarter of the working wavelength, a Salisbury screen structure can be made, which can increase the EM field amplitude through constructive interference.[138,139] Almost all of these devices operate at mid-infrared and terahertz regions, corresponding to intraband interaction (except for ref.[130] that utilizes interband interaction), and the resonance is shifted to shorter wavelength when increasing the amplitude of Fermi energy. It should be noted that metasurfaces using patterned graphene structures have been proposed and demonstrated, aiming to manipulate and steer free-space light in the THz and infrared region.[140–143]

Although high modulation depth (~90%) has been achieved,[98,133,136] the graphene modulators mentioned above cannot achieve low insertion loss and high modulation depth simultaneously. Another problem is that a large bias voltage, up to several tens or even one hundred volts, is needed to efficiently tune the Fermi energy of graphene. Ion gel can decrease the voltage to less than 5 V,[131,136,137,198] but the inevitable absorption of ion gel can deteriorate the performance. Chemical doping can easily increase the Fermi energy to a high value, but it cannot be used in applications that require a fast modulation speed.[135] These problems associated with graphene-based modulators have been greatly alleviated by recent progress on this topic. For example, in 2016, S. Kim et al. reported a transmission-type modulator based on graphene ribbons combined with periodic metal stripes.[144] The structure is illustrated in the top two panels of Figure 8a. Compared with the earlier modulator works,[130,134,136] the modulation efficiency is greatly enhanced. At plasmonic resonance, the periodic metal stripes show extraordinary optical transmission (EOT), in which the incident wave couples into SPPs on the top metal surface, then the SPPs tunnel through the metal slit to the bottom metal surface and couple back into propagating waves. By tuning the Fermi level, the resonant wavelength of graphene ribbons inside the metal slit can match the EOT wavelength of metal stripes, and the EM wave in the metal slit would be confined around graphene ribbons instead of tunneling through. As can be seen in the bottom left and middle diagrams in Figure 8a, at the EOT wavelength of 7.46 μm, the device has large transmission and low EM field amplitude around graphene when \( E_f = 0 \) eV. In contrast, when \( E_f = -0.465 \) eV, the transmission substantially decreases, and EM field is significantly confined at graphene ribbons. The numerical simulation shows that the transmission modulation depth can reach 95.7% if graphene carrier mobility \( \mu = 15,000 \text{ cm}^2\text{V}^{-1}\text{s}^{-1} \). However, due to low quality of wet transferred CVD graphene, the mobility can only reach a value around 450 cm² V⁻¹ s⁻¹, leading to an experimentally measured modulation depth of 28.6%, as shown in the bottom right panel of Figure 8a.

S. Kim et al. further provided a comprehensive study on reflection-type modulators based on graphene combined with MIM perfect absorber structure,[145] achieving both high modulation depth and low insertion loss at the same time. As shown in top three panels in Figure 8b, they fabricated devices with three different structures, namely a Salisbury screen structure (type A), an MIM perfect absorber with graphene ribbons placed in the center of metal slits (type B), and another MIM structure with graphene ribbon off to one side in the metal slit (type C). There is a thin layer of SiO\(_2\) in between graphene and the SiN substrate. This SiO\(_2\) thin layer can reduce the permittivity of surrounding media of graphene, thus increase the light-graphene coupling strength. The three proposed devices, with structure parameters shown in these three figures, can produce local field enhancement of 4, 147, and 226, respectively. The middle three figures show the field distribution at perfect absorption conditions, which corresponds to different graphene carrier mobilities of 2271, 613, and 315 cm² V⁻¹ s⁻¹ for types A, B, and C, respectively. The relationship between carrier mobility and absorption is also shown in the bottom left panel of Figure 8b. This study reveals that high carrier mobility of graphene is not always the necessary requirement for perfect absorption, but there exists a critical value of \( \mu \) at which perfect absorption can be achieved. As we can see, the carrier mobility required for type B and C is much lower than type A structure, proving that even low quality graphene can achieve perfect absorption in these two structures. The measured performance of these three structures is shown in bottom middle and right panels of Figure 8b. The type C structure exhibits an extraordinary performance, with measured reflection varied from 3.1% to 75.2% at 7.2 μm, corresponding to a modulation depth of 95.9% and an insertion loss of only 24.8% in reflection.

Also in 2018, B. Zeng et al. reported a modulator design that can achieve high electrical tunability using low voltage (<10 V) without the aid of ion gel.[146] The device layout is shown in the top right diagram of Figure 8c. Typically, the electrically tunable graphene metasurface absorber requires a thick dielectric spacing layer, which serves as a dielectric layer for optical purpose and an electrical insulation layer for back gating. This thick dielectric layer requires a large drive voltage to effectively tune the Fermi level of graphene, which limits the modulation speed. In order to significantly reduce the drive voltage and preserve the required optical dielectric layer thickness, the authors replaced a large portion of the typically nonconductive dielectric material with slightly conductive amorphous Si (a-Si) and used a thin layer of Al\(_2\)O\(_3\) (6 nm) as the gate insulation layer. Since a-Si is transparent in mid-infrared regime, it can serve as a dielectric layer for optical purpose, and simultaneously as the back-gate electrode for electric tuning. The structure configuration is shown in top left panel of Figure 8c. The fabricated gold metasurface is shown in the bottom left panel of Figure 8c. The tiny gap (40 nm) between the I-shaped nanoantennas can provide large local field enhancement for the underneath graphene layer. The device shows peak modulation depth at around 7.3 and 8.5 μm, with the highest modulation depth of ~90% reached at 8.5 μm, corresponding to a reflection change of 33.1% at +7 V to 6.4% at ~3 V, as shown in top right panel of Figure 8c. The modulation frequency has also been measured. By excluding the influence of external resistors, this device
shows a high 3 dB cut-off modulation frequency of 7.2 GHz. The authors further demonstrated the potential of the device on mid-infrared imaging via fabricating a 6 × 6 array of such modulators, in which each of modulator served as a pixel and could be modulated individually. By raster scanning the pixels, the signal of each pixel can be recorded, and the mid-infrared picture of the object can be retrieved (bottom right diagram of Figure 8c).

Figure 8d represents another very interesting work by A. Safaei et al., which demonstrated high modulation efficiency without the use of metal-based plasmonic nanostructures.[147] In this work, the authors experimentally realized the theoretical work made by S. Thongrattanasiri et al. in 2012, who predicted 100% light absorption by graphene nanodisk arrays that are placed on a dielectric substrate with a metallic ground plate.[148] As shown in the top left panel of Figure 8d, the device utilizes a thick SU-8 layer and bottom gold layer to create a quarter wavelength cavity, which functions as a Salisbury screen structure and can increase surface plasma intensity of graphene as we have mentioned. The transparent conducting ITO layer electrically connects the graphene nanodisks, while the upper Al₂O₃ layer and Si film serve as gate dielectric and electrode, respectively. The patterned graphene nanodisk array is shown in the top right panel of Figure 8d. The Fermi level can be tuned from 0.55 eV to 1 eV, leading to a resonance shift of ≈1.11 μm, with almost 90% absorption at 1 eV Fermi level (bottom left panel of Figure 8d). Notably, the device shows polarization- and angle-insensitive light absorption over a wide incident angle up to 50° (bottom right panel of Figure 8d).

During recent years, there has been continued progress on graphene optical modulators. For example, the work reported in 2015 by B. Thackray et al. achieved a modulation depth of 20% at 1.5 μm using exfoliated monolayer graphene and BN flake on top of periodic gold nanostripes, corresponding to the largest experimentally reported modulation depth for graphene-based free-space optical modulators at the near-infrared wavelength.[149] Another example is the graphene reflection phase modulator reported in 2017 by M. Sherrott et al. in which a smooth reflection phase transition of 206° could be achieved at 8.7 μm using graphene combined with MIM structure.[150] However, this large phase tuning is accompanied with large absorption, resulting in large loss for beam steering or focal lens devices.

In addition to optical modulators, another important application is graphene-based sensors, such as graphene photodetectors and graphene molecule sensors. Graphene photodetectors are mostly fabricated in a metal–graphene–metal (MGM) architecture, where the metal electrodes are used to apply bias voltage and detect the photoinduced current. The photoinduced current can come from photovoltaic effect, in which photoinduced electron–hole (e–h) pairs are generated, or photothermoelectric effect, in which the absorbed light induces temperature change.[151,152] The different work functions between the metal contact and graphene form junctions at two edges of the graphene channel, which generate internal electric fields and spatially separate the e–h pairs. Since the internal electric fields are in opposite directions at the two edges, the bias voltage is required to get a nonzero net photocurrent.

Similar to graphene modulators, noble metal plasmonic structures can also be used to enhance the photoresponsivity of graphene photodetectors. By applying metallic cluster arrays between two graphene monolayers, Z. Fang et al. observed an 800% enhancement of photocurrent.[153] Such a huge enhancement is due to the increased excitation of e–h pairs in graphene and photo generated hot electrons transferred from the plasmonic cluster to graphene. In a typical MGM structure, incident light is only absorbed on the exposed graphene, and the portion of light incident on the electrodes is simply reflected, lowering down the efficiency. More critically, due to the semimetal property of graphene, the bias voltage applied can generate a dark current that can be even larger than the photocurrent. To solve these problems, T. Echttermeyer et al. proposed a novel metal contact design.[154] By making the electrodes into metallic grating, the light incident on the electrodes can be transformed into SPPs that propagate to the graphene area, leading to 1000% increase in photovoltaic length. They also make asymmetric metal contact design to break the symmetry of the photodetector, so that the light is mostly absorbed in one side of the graphene channel. As a result, nonzero net photocurrent can be generated without applying bias voltage, solving the dark current problem. In 2018, S. Cakmakyayan et al. demonstrated a graphene photodetector with high responsivity over a wide wavelength regime (0.6 A W⁻¹ at 0.8 μm and 11.5 A W⁻¹ at 20 μm), setting a record of the widest photodetection bandwidth with high responsivity up to date.[155] The structure is shown in figures on the left of Figure 9a, in which the metal electrodes are designed so that the incident light can be concentrated in the graphene channels within a broad spectrum from 0.8 to 20 μm, shown in the top right panel of Figure 9a. Although the optical absorption is much lower at the short wavelength regime, which corresponds to interband transition, the high responsivity is still maintained at this regime, shown in the bottom two panels in Figure 9a. The reason is that the higher energy level of electrons excited to conduction band, together with fast photocarrier transport time, makes the carrier multiplication possible and boosts the photoconductive gain. The drawback of this device is the large dark current, which is intrinsic to the MGM architecture shown in Figure 9a.

A graphene molecule sensor is based on the sensitivity of graphene plasmonics to the minor change of the permittivity of surrounding media. When molecules are absorbed around the graphene nanopatterns, the change of permittivity induced by these molecules can be detected from the change in optical responses. Compared with the metallic plasmonic molecule sensor, a graphene sensor features tighter field confinement and thus higher sensitivity, as well as dynamic tunability that enables sensing over a broad spectrum in one single device. As reported by D. Rodrigo et al.[156] by carefully tuning the optical response of graphene plasmonic nanoribbons toward the vibrational bands of protein molecules, 27% signal modulation can be detected, around threefold increase compared with modulation from the state-of-the-art metallic localized surface plasmon resonances (LSPR) sensor. The graphene molecule sensor can also be used in sensing multiple molecule species and monitoring the dynamical chemical reaction process, demonstrated by H. Hu et al.[157] As can be seen in the top left plot of Figure 9b, the graphene nanoribbons are used to identify the specific gas molecules. The featured rotational-vibrational modes (P, R) of SO₂ molecules can be captured in the extinction
spectra of graphene nanoribbons, marked as solid green areas in the top right plot of Figure 9b. Other gas molecules, including O₂, N₂O, and NO, can also be detected at different frequencies of the extinction spectrum with high sensitivity (800 zeptomole molecules per μm²). Such a high sensitivity is attributed to large physisorption of graphene nanoribbons. The dynamic plot at bottom left of Figure 9b records the real-time gas signal change when SO₂ is pumped in and then pumped out. A fast response time (<1 min) is demonstrated. This fast response time can be utilized to monitor real-time chemical reactions, such as the chemical reaction between O₂ and NO, as shown in the bottom right plot of Figure 9b.

4.3. Exciton–Polaritons in Transition Metal Dichalcogenide and Heterostructures

Since graphene does not have a bandgap and hence has limitations in the applications of electric transistors and photodetectors, researchers’ attention has recently shifted to TMDs.[158–161] TMDs are one type of semiconductors with MX₂ stoichiometry, in which M denotes a transition metal atom (e.g., Mo and W) and X denotes a chalcogen atom (e.g., S, Se, and Te). One advantage of TMD is the transition from an indirect bandgap in the bulk to direct bandgap as the number of layers gradually decreases to one. Also, the bandgap spans the energy range from 1.1 eV to 1.9 eV,[159] which corresponds to the favored wavelength of polaritons. Therefore, strong light-matter interaction can exist in such system known as exciton–polaritons.[162] One representative example of TMDs is Molybdenum disulfide (MoS₂). It has been used in applications such as emitters,[163] transistors,[164] photodetectors[165] as well as valleytronic devices[166–170] owing to its atom arrangement that breaks the inversion symmetry, as well as its novel excitonic properties such as small effective exciton Bohr radius (~1 nm) and associated large exciton binding energy (0.96 eV).[171] It should be noted that although different TMDs features qualitatively similar characteristics, there are still some differences that cannot be ignored. For example, the spin-orbit coupling can be positive or negative depending on the TMD material: in molybdenum-based TMDs (e.g., MoS₂ and MoSe₂), electrons in the lowest conduction band have the same spin as those in the highest valence band, while an opposite spin ordering is found in tungsten-based TMDs (e.g., WS₂ and WSe₂). This will result in different optically forbidden dark excitons and different spintronic properties.[161]

By combining plasmonic nanoantenna arrays with TMD monolayers, light absorption can be enhanced, thus increasing the photoluminescence (PL) of TMD monolayers. In 2014, S. Najmaei et al. reported a 65% enhancement of PL in the MoS₂-Au nanoantenna hybrid structure, in comparison with a bare MoS₂ monolayer.[172] In their experiments, by tuning the incident laser wavelength on (or off) the plasmonic resonance with the Au nanoantenna, the PL intensity is significantly increased (or almost unchanged) compared with a MoS₂ monolayer without Au nanoantenna, which unambiguously demonstrated the strong influence of the nanoantenna surface plasmons to PL intensity. The mechanism can be interpreted in terms of local EM field enhancement from the nanoantenna resonance. It induces increased light absorption, and consequently enhances PL intensity after the absorbed photons are converted into electron–hole pairs in the MoS₂ layer. Also, redshift and broadening of PL emission have been observed, arising from the temperature increase induced by the enhanced absorption of MoS₂. In 2015, S. Butun et al. reported an even higher enhancement, up to 12 times, of PL intensity using MoS₂-Ag nanodisks.[173] As shown in left top plot of Figure 10a, the structure is made up of Ag nanodisk arrays deposited on top of MoS₂ monolayer. Using different nanodisk diameters, the PL spectrum is changed, and the maximum PL emission can be 12 times higher than the pristine MoS₂ layer, shown in left bottom plot of Figure 10a. The mechanisms of this enhancement are the excitation field enhancement at pump wavelength (similar to previous work), as well as more efficient scattering at PL emission wavelengths. As illustrated in the top right plot of Figure 10a, by simulating and measuring the reflection spectrum of Ag nanodisk arrays without MoS₂, the reflection at excitation wavelength (543.5 nm) is reduced, while around emission wavelength (680 nm) the reflection is increased, proving that the plasmonic resonance couples to both excitation and emission fields. By averaging both excitation and emission field intensity underneath Ag nanodisk in finite-difference time-domain (FDTD) simulation, the calculated PL enhancement matches well with the experiments (bottom right plot of Figure 10a), further proving that excitation and emission enhancement both play an important role.

The preceding results show that the collective resonance of metal nanoantenna arrays can boost the light–matter interaction and increase the PL intensity of a TMD monolayer. The individual metal nanoparticles, on the other hand, can selectively enhance or decrease the localized PL intensity at certain exciton modes or locations. H. Lee et al. reported a structure of MoS₂ monolayer combined with Ag nanowire (NW).[174] There is a thin spacer layer (10 nm SiO₂) between Ag NW and MoS₂ to prevent direct contact as shown in the top left schematic of Figure 10b. When the laser light is incident on an Ag NW that is partially overlapped with a MoS₂ monolayer (“on NW”), the localized PL intensity can be 20 times higher than the PL intensity at location without NW (“off NW”), as shown in the bottom 3 plots of Figure 10b. Moreover, when increasing the incident energy, the PL spectra of “on NW” and “off NW” are distinctly different. For the “on NW” case, the peak wavelength, as well as the full width at half maximum, stay unchanged when incident power increased, indicating selective A⁰ exciton enhancement. In contrast, for “off NW” case, redshift of peak wavelength, accompanied with broadened spectrum, is observed with increased power, indicating emergence and evolvement of A’ and B exciton, as shown in the right plot of Figure 10b. This selective amplification of A⁰ exciton is attributed to the Fabry–Pérot cavity formed in the nonoverlapped part of the partially overlapped NW. The excitons generated at the laser incident point couple to SPPs propagating along the NW, which reflect at the NW end-point and re-excite A⁰ excitons at the laser incident point. Since the absorption of SPPs is much larger in the region where NW overlaps with MoS₂, this Fabry–Pérot cavity mode only exists at the nonoverlapped part.

W. Gao et al. reported a hybrid structure of MoS₂-Ag nanoparticles. For different morphologies of Ag nanoparticles, the characteristics of PL are different.[175] As shown in the left panel of Figure 10c, the extinction spectra of Ag octahedral (Ag OCT),
Ag nanocube (Ag NC), large Ag nanosphere (Ag SP1) and small Ag nanosphere (Ag SP2) are compared with PL spectrum of MoS2 monolayer as well as the excitation wavelength. The Ag NC shows dipolar surface resonances overlapped with both excitation wavelength and MoS2 monolayer bandgap, giving rise to the largest PL enhancement among these four morphologies. Similar to Ag NC, Ag SP1 shows weaker PL enhancement, while Ag SP2 exhibits little overlap for its resonance with either excitation wavelength or MoS2 bandgap, thus demonstrating negligible influence on the PL of MoS2. For Ag OCT, however, the PL is weakened, instead of enhanced, even though there is overlap between surface plasmon resonance and excitation wavelength of MoS2 bandgap. The reason is that Ag OCT exhibits high-order resonance modes, while other Ag nanoparticles present dipolar resonance modes. The nonradiative nature and poor directionality of Ag OCT high-order resonance modes reduce the collected PL intensity. The light enhancement (weakening) for Ag NC (Ag OCT) is demonstrated in the top two panels of Figure 10c. Utilizing these effects, the PL intensity is either weakened or enhanced at different locations by introducing both Ag NC and Ag OCT, as shown in the right bottom two diagrams of Figure 10c.

The combination of metallic structures and TMD monolayers can also introduce the coupling between optical resonant modes (including SPPs) from metallic structures and excitons from TMD monolayers. When this coupling is strong enough, the resonant modes of the hybrid system differ from the individual optical resonant mode or exciton mode. The energy separation of the new resonant modes at the crossing point (of the un-coupled case) is called Rabi splitting,[17] which is also the spectral signature for the strong coupling. X. Liu et al. proposed strong light-matter coupling and the formation of microcavity polaritons in monolayer MoS2 embedded inside a dielectric microcavity at room temperature.[176] They built a microcavity using SiO2/Si3N4 distributed Bragg reflector and MoS2 monolayer was sandwiched inside two SiO2 layers. By changing the incident angle which would shift the in-plane momentum of optical wave and extract the resonance peaks, the dispersion diagram including the upper and lower polariton branch was obtained, showing a Rabi splitting of $\hbar \Omega_{\text{Rabi}} = 46 \pm 3 \text{meV}$. The authors have also performed angle-resolved photoluminescence measurements and shown similar results. S. Wang et al. studied the strong coupling between exciton mode of WS2 monolayer and optical resonance
modes of Fabry–Pérot cavity resonance and plasmonic array resonance at the room temperature. The Fabry–Pérot cavity was made of two Ag thin films acting as mirrors, and WS2 was placed in the middle. The plasmonic arrays were fabricated by drilling a hole-array in Au film and transferring WS2 monolayer on top of it. Rabi splitting of 101 meV was observed for the Fabry–Pérot cavity case, while 60 meV of Rabi splitting was observed in WS2-plasmonic array structure. W. Liu et al. reported the strong exciton–plasmon coupling between MoS2 and Ag nanodisk plasmonic lattices. This system exhibits coupling between three types of resonances: MoS2 excitons (both A and B excitons), LSPR of individual nanodisks and lattice resonances of the array. A coupling strength of 58 meV is observed at 77 K, which comes from coupling between LSPR and A exciton. By changing the diameter of Ag nanodisks, the LSPR can be either in resonance or out of resonance with MoS2 excitons, which effectively tunes the plasmon–exciton coupling strength. X. Han et al. investigated the strong coupling between WS2 monolayer exciton and individual Ag nanocubes instead of the aforementioned plasmonic lattices. The structure was fabricated by successively depositing silver and Al2O3 thin films on top of a silicon substrate, and then transferring the WS2 monolayer and finally drop-casting Ag nanocubes. The schematic of the structure, as well as bright-field and dark-field images, are shown in the left panels of Figure 10d. The Ag nanocube and bottom silver film with Al2O3 spacing layer in the middle form a nanoscale cavity can enhance the LSPR. The thickness of the Al2O3 layer is optimized to match the resonant wavelength of LSPR to the WS2 exciton wavelength. By changing the size of Ag nanocubes, the LSPR wavelength can be finely tuned, and corresponding scattering spectra are recorded. The measured Rabi splitting is shown in the right panel of Figure 10d, where the green and black dashed lines correspond to the uncoupled WS2 exciton and the detuning of the uncoupled LSPR respectively. The blue and red dots are extracted from the peaks in the scattering spectra of WS2-Ag nanocube hybrid structures, while the fitting lines are from coupled harmonic oscillator model. The green arrows mark a large Rabi splitting value of around 145 meV measured at the room temperature, indicating a strong plasmon-exciton coupling.

Apart from the aforementioned PL intensity tuning and Rabi splitting, the metallic nanostructures can also give rise to other interesting phenomena. For example, Y. Kang et al. reported hot electron doping from Au nanoparticles to MoS2 monolayer. Hot electron doping induces a phase transition from the 2H phase to the 1T phase in a MoS2 monolayer, which is confirmed by Raman spectra. M. Wang et al. studied resonance energy transfer (RET) from single plasmonic Au nanotriangles to the MoS2 monolayer. Due to the enhanced RET from plasmonic resonance of Au nanotriangles to MoS2 monolayer, charged excitons (i.e., trions) can be observed in the scattering spectra of a single nanoparticle at room temperature. The trion population can be tuned under different dielectric constants of the surrounding medium, which is manifested via the shift of absorption peak in the spectra. J. Cuadra et al. did similar work using silver nanoprism on top of WS2 monolayer in which strong interactions between LSPR in silver nanoparticles and excitons and trions in monolayer WS2 occur. In particular, a stable trion state emerges at low temperature (6K), which forms intermixed plasmon–exciton–trion polaritons. Z. Fei et al. performed comprehensive work on the MoSe2 waveguide mode. They used a laser beam to excite the MoSe2 waveguide mode through an atomic force microscope tip. The near-field interference pattern would have a dependence on the orientation between incident beam and the sample edge directions. Therefore, by performing perpendicular and parallel measurements, the authors proved that the pattern is from the interference between scattering from the tip and the sample edge. Then they can obtain the wavelength from the fringe period and finally extract the dispersion relation. A Rabi splitting of ~100 meV indicates the existence of exciton–polaritons. In 2019, L. Sun et al. demonstrated separation of both valley excitons in real space and photon emission in momentum space with a subwavelength asymmetric structure that breaks the chiral symmetry in MoS2. The spin angular momentum performs as an additional degree of freedom of light that influences light-matter interactions at the nanoscale, which can result in a robust one-to-one relation between the helicity of incident light and the propagation direction of excited photonic modes, known as spin-momentum locking. By utilizing this principle, helicity dependent unidirectional coupling of surface plasmon is demonstrated at room temperature in a silver nanowire and also gold layer with plasmonic resonators on top of WS2 layer.

5. Conclusion and Outlook

We have reviewed many works on the fundamental study and potential application of polaritons in metallic structures and 2D van der Waals materials, including surface plasmon polaritons in graphene and exciton–polaritons in TMDs. Another major category of polaritons is phonon polaritons, which are collective oscillations induced by the coupling of photons with optical phonons in polar dielectric materials. A recent study reveals that phonon polaritons in 2D systems are simply the longitudinal optical phonon, due to the absence of splitting of longitudinal and transverse optical phonons at Γ point in 2D polar materials. Hexagonal boron nitride (hBN), as one representative polar material, possesses natural hyperbolic characteristics and provides an ideal platform to study phonon polaritons within its two Reststrahlen bands. The phonon polaritons show strong light confinement and low losses that can be even better than graphene plasmon polaritons. The experiments conducted on 2D hBN demonstrate a confinement factor of 25 that is comparable to graphene, while the corresponding loss factor is around 0.055, much smaller than graphene. Also, the wavelength of the hyperbolic phonon polariton in hBN is shown to be proportional to the sample thickness. The properties of phonon polaritons in 1D hBN nanotubes and 3D hBN nanocones have also been explored. Using near-field imaging and time-domain interferometry, researchers have experimentally demonstrated negative phase velocity and ultraslow group velocity (0.2% of speed of light in vacuum) of hBN hyperbolic phonon polaritons. Due to the aforementioned unique characteristics, the hyperbolic phonon polaritons in hBN have been
applied in near-field imaging, presenting enlarged imaging and super-resolution focusing.\[193\]

The research in recent years revealed that highly confined phonon polaritons exist not only in hBN, but also in other van der Waals materials. The study on TMDs (including MoS$_2$, MoSe$_2$, WS$_3$ and WSe$_2$) at a SiC interface discovered the ultra-confined surface phonon polaritons, with confinement factor of around 190 in bi-layer MoS$_2$.\[194\] People have also demonstrated in-plane anisotropic phonon–polaritons in $\alpha$-MoO$_3$, which is a semiconducting van der Waals material.\[195,196\] The confinement factor can reach up to 120, and the polariton amplitude lifetime can be as long as 8 ps, which are far better than graphene plasmon polaritons. Right now, the study of phonon polariton is still a rapidly growing research area and is anticipated to provide new functionalities for sub-diffraction imaging, sub-surface sensing, and so on.

There are other types of polaritons, including Cooper-pair polaritons,\[197–200\] in-plane anisotropic surface plasmon polaritons\[201–203\] and hybrid polaritons like hyperbolic plasmon–phonon polaritons,\[204–206\] which are worth further exploration. These polaritons also possess exotic optical properties and great potential in various applications. In addition, by combining different 2D materials together, the heterostructures can show better performances than individual 2D materials in some cases, including higher mobility in graphene with h-BN\[106,107\] and superior photoresponsivity ($5 \times 10^8$ A W$^{-1}$ at room temperature) in graphene-MoS$_2$ photodetectors.\[207\] Some interesting phenomena also exist in the heterostructures, such as the exciton–polaritons observed in a MoSe$_2$/h-BN heterostructure hybridized with microcavity photons.\[208\] These aforementioned polaritons and heterostructures are less exploited compared with graphene and TMDs. Even for graphene and TMDs, there is plenty of room to further push their performances. For example, most of the graphene-based free-space optical modulators reported to date operate at mid-infrared and terahertz regimes, while a large amount of metasurface devices operate at visible or near-infrared regimes. If we can push the working wavelength of graphene modulators to the shorter wavelength regime and combine them with the rapidly growing metasurface technologies, we will be able to develop more advanced functionalities, such as tunable holograms or lenses with fast response time and reconfigurability. Although there are lots of challenges along the path, these challenges also manifest opportunities for us to fully exploit the genuinely unique properties of polaritons, in order to advance the fundamental understanding and technology transfer in this frontier research area.

**Acknowledgements**

Z.C. and Y.X. contributed equally to this work. Y.L. acknowledges the financial support from the Office of Naval Research (N00014-16-1-2409) and the National Science Foundation (DMR-1654192 and CBET-1931777). The authors thank Eric Nagy for reading and polishing the manuscript during the revision process.

**Conflict of Interest**

The authors declare no conflict of interest.

**Keywords**

2D materials, graphene, metals, photonics, polaritons

Received: June 30, 2019

Revised: September 2, 2019

Published online:


