Title

Tunneling of two-dimensional surface polaritons through nanogaps in atomically thin crystals

Journal

Physical Review B, 99(16)

ISSN

2469-9950

Authors

Kang, Ji-Hun
Wang, Sheng
Wang, Feng

Publication Date

2019-04-01

DOI

10.1103/physrevb.99.165408

Peer reviewed

Tunneling of two-dimensional surface polaritons through nanogaps in atomically thin crystals

Ji-Hun Kang[1*], Sheng Wang[2, 3], and Feng Wang[2, 3, 4]

[1] Department of Physics and Astronomy, Seoul National University, Seoul 08826, Republic of Korea

[2] Department of Physics, University of California at Berkeley, Berkeley, California 94720, USA

[3] Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

[4] Kavli Energy NanoSciences Institute at the University of California, Berkeley and the Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

Corresponding author: *jihun.kang@snu.ac.kr

We theoretically investigate the tunneling of two-dimensional surface polaritons (2DSPs) through nanometer-wide gaps in atomically thin crystals. For quantitatively accurate results, we developed a rigorous model based on the diffraction of 2DSPs for strongly confined surface polaritons (i.e., the polariton wavelength much shorter than the free-pace photon wavelength). We find distinctive features of the tunneling of 2DSPs. First, radiation loss during the tunneling is shown to be negligible. Second, the reflection coefficient R and tunneling coefficient T are shown to exhibit an anomalous logarithm singularity in their dependency on the gap width. Even for a gap size over two orders of magnitude smaller than the surface polariton wavelength, an appreciable reflection

coefficient was observed in our calculation. Finally, we show that when the gap size increases, the phase of R saturates very rapidly to a non-trivial value of $\pi/4$. Based on these results, we further examine resonant tunneling of 2DSP through two identical gaps separated by a distance L, and establish a resonance condition defined by $L \approx \lambda_{sp}(4n\text{-}1)/8$ with a positive integer n.

I. Introduction

Two-dimensional (2D) surface polaritons (SPs), coupled excitation of photons and conduction electrons [1–4], optical phonons [5,6], or excitons residing in atomically thin 2D crystals [7–10], exhibit deep sub-wavelength confinement and hold great promises for nanoscale integration of polariton optics [2,11–15]. Understanding of propagation behaviors of 2DSPs is at the core of further manipulation of low-dimensional polariton optics. Recently there have been extensive studies of the reflection of 2DSPs at the crystal's edge, which have enabled direct visualization of 2DSPs in near-field infrared nanoscopy [16–19]. The edge reflection of 2DSP has almost a unity reflection amplitude, but the reflection phase is non-zero due to the near-field evanescence wave close to the 2D crystal edge [16–18]. The existence of the evanescent wave naturally leads to the question of near-field coupling, i.e., 2DSP tunneling across a finite-sized gap in a 2D crystal. Such near-field tunneling of 2DSP was investigated by recent works [20]. Due to the reduced dimensionality of atomically thin 2D crystals, the tunneling of 2DSP can exhibit different gap-size-dependent tunneling efficiency and phase shifts compared with conventional evanescence wave coupling.

In this letter we investigate theoretically the behavior of strongly confined 2DSP's tunneling through nanometer-sized gaps in atomically thin 2D crystals. Through our model that guarantees quantitative accuracy, we find several distinctive properties of the tunneling of strongly confined 2DSPs (i.e., $\lambda_{sp} << \lambda_0$ where λ_{sp} and λ_0 are wavelengths of polaritons and free-space photon, respectively). First, for ideal lossless 2D crystals, radiation loss during the tunneling is shown to be negligible, implying that the tunneling always satisfies the conservation $|T|^2 + |R|^2 \approx 1$ where T and R are tunneling and reflection coefficients, respectively. Second, we find that the tunneling

efficiency shows a strong gap size dependence when the gap size approaches zero: the reflection and tunneling coefficient of 2DSP shows a logarithm singularity close to a zero gap, in contrast to a smooth exponential decay of conventional evanescence wave coupling across a three-dimensional gap. This logarithm singularity is found to originate from the multiple interaction between the planewave modes in the gap and the unbounded modes in the 2D crystal region. Finally, we reveal that the phase of the reflection coefficient also changes rapidly with the increase of the gap size, and saturates at a non-trivial value of $\pi/4$ for large gap size. This unusual phase information allows further examination of resonant tunneling of 2DSPs across two identical nanogaps in 2D crystals, and we show that the resonant tunneling condition is defined by $L \approx \lambda_{sp}(4n\text{-}1)/8$ where L is the distance between the two gaps and n is a positive integer.

II. Theoretical model

We model the system as an infinitely thin 2D sheet with a gap of width g, as shown in Fig. 1. We assume that the 2D sheet is embedded in a surrounding medium with uniform permittivity of ε_s. 2DSPs propagate along the crystal in $+x$ direction. 2DSPs can be described by localized electromagnetic (EM) waves in the proximity of 2D crystal surfaces [11,16,18]. Specifically, with the configuration shown in Fig. 1, Maxwell equations characterize the z-component of magnetic field in terms of a surface-bounded mode $|p\rangle$ of which the real-space representation $\langle y|p\rangle$ forms an anti-symmetric distribution, $\langle y|p\rangle = \exp(iq_y|y|)y/|y|$ with $q_y \equiv \sqrt{\varepsilon_s k_0^2 - q_x^2}$, and k_0 and q_x are momenta of the photon and the 2DSP, respectively [21]. Here, we assume that there is no material loss in 2D crystals. Then, 2DSPs propagating in x direction can be described by taking

into account the phase change as $|p\rangle\exp(iq_x x)$. When 2DSPs arrive at the edge of the crystal (x=0), there is a mismatch between eigen-functions of the gap ($0\le x\le g$) and crystal ($x\le 0$) regions. This results in the coupling to the reflected 2DSP as well as diffracting waves. Bases for the diffracting waves can be represented by unbounded modes $|u_{ky}\rangle$ in 2D crystal region and plane-wave modes $|f_{ky}\rangle$ in the gap that will be coupled to $|p\rangle$ and $|u_{ky}\rangle$ at $x=g$ [18]. In Figs. 2(a) and (b), those 2DSP and diffracting waves are separately shown for an exemplary tunneling case. The coupling between $|p\rangle$ and $|u_{ky}\rangle$ eventually determines the tunneling efficiency, and as shown in Fig. 2(c) and (d), the efficiency changes dramatically with the gap size.

All those macroscopic procedures can be explicitly described by the following field expansion of z-component magnetic field,

$$\begin{aligned}
|H_z^{x\le 0}\rangle &= \left(e^{iq_x x} - Re^{-iq_x x}\right)|p\rangle + \int_{-\infty}^{\infty} dk_y \rho_{ky} |u_{ky}\rangle e^{-ik_x x}, \\
|H_z^{0\le x\le g}\rangle &= \int_{-\infty}^{\infty} dk_y \alpha_{ky} |f_{ky}\rangle e^{ik_x x} + \int_{-\infty}^{\infty} dk_y \beta_{ky} |f_{ky}\rangle e^{-ik_x x}, \\
|H_z^{x\ge g}\rangle &= T|p\rangle e^{iq_x (x-g)} + \int_{-\infty}^{\infty} dk_y \tau_{ky} |u_{ky}\rangle e^{ik_x (x-g)}.
\end{aligned} \tag{1}$$

Here, R and T respectively are the reflection and tunneling coefficients of 2DSP that will be fixed later, and $k_x \equiv \sqrt{\varepsilon_s k_0^2 - k_y^2}$. The real-space representations for $|u_{ky}\rangle$ and $|f_{ky}\rangle$ respectively are given by

$$\langle y|u_{ky}\rangle = \frac{y}{|y|}\left(e^{ik_y|y|} + \frac{k_y - q_y}{k_y + q_y} e^{-ik_y|y|}\right), \quad \langle y|f_{ky}\rangle = e^{ik_y y}. \tag{2}$$

Note that $\left|u_{ky}\right\rangle$ describes unbounded diffracting waves in the 2D crystal region. The completeness of our modal expansion in the crystal region is justified by the orthogonal relationship between $\left|p\right\rangle$ and $\left|u_{ky}\right\rangle$ as

$$\begin{aligned}
&\left\langle p\middle|u_{ky}\right\rangle = \int_{-\infty}^{\infty} dy \left\langle p\middle|y\right\rangle\left\langle y\middle|u_{ky}\right\rangle = 0, \\
&\left\langle u_{ky1}\middle|u_{ky}\right\rangle = 4\pi\delta\left(k_y - k_{y1}\right) + \frac{k_{y1}+q_y}{k_{y1}-q_y} 4\pi\delta\left(k_y + k_{y1}\right).
\end{aligned} \tag{3}$$

After the diffraction at the nanogap, the coupling of incident 2DSP to the reflected 2DSP and diffracting waves can be determined by the following basis dependencies,

$$\begin{aligned}
&\left\langle f_{ky}\middle|u_{ky1}\right\rangle = \frac{1}{i}\frac{4k_{y1}}{k_{y1}+q_y} K\left(k_y, k_{y1}\right) + 2\pi \frac{q_y}{k_{y1}+q_y}\left[\delta\left(k_y - k_{y1}\right) - \delta\left(k_y + k_{y1}\right)\right], \\
&\left\langle f_{ky}\middle|p\right\rangle = -\frac{2ik_y}{k_y^2 - q_y^2}.
\end{aligned} \tag{4}$$

Here, the kernel K is defined as

$$K\left(k_y, k_{y1}\right) \equiv \frac{k_y}{k_y^2 - k_{y1}^2}. \tag{5}$$

Corresponding electric fields can be easily obtained from Eq. (1) by using the Maxwell's equations. We can then readily apply the boundary conditions. Boundary conditions require the continuity of tangential components of EM waves at two interfaces, x=0 and x=g. The continuity of H_z and E_y at x=0 gives continuity conditions,

$$\begin{aligned}
&\left(1-R\right)\left|p\right\rangle + \int_{-\infty}^{\infty} dk_y \rho_{ky}\left|u_{ky}\right\rangle = \int_{-\infty}^{\infty} dk_y \left(\alpha_{ky} + \beta_{ky}\right)\left|f_{ky}\right\rangle, \\
&q_x\left(1+R\right)\left|p\right\rangle - \int_{-\infty}^{\infty} dk_y \rho_{ky}\left|u_{ky}\right\rangle k_x = \int_{-\infty}^{\infty} dk_y \left(\alpha_{ky} - \beta_{ky}\right)\left|f_{ky}\right\rangle k_x.
\end{aligned} \tag{6}$$

Also, at $x=g$, we have

$$\begin{aligned} T\left|p\right\rangle+\int_{-\infty}^{\infty}dk_y\tau_{ky}\left|u_{ky}\right\rangle&=\int_{-\infty}^{\infty}dk_y\left(\alpha_{ky}e^{ik_xg}+\beta_{ky}e^{-ik_xg}\right)\left|f_{ky}\right\rangle,\\ q_xT\left|p\right\rangle+\int_{-\infty}^{\infty}dk_y\tau_{ky}\left|u_{ky}\right\rangle k_x&=\int_{-\infty}^{\infty}dk_y\left(\alpha_{ky}e^{ik_xg}-\beta_{ky}e^{-ik_xg}\right)\left|f_{ky}\right\rangle k_x. \end{aligned} \tag{7}$$

Then, we can project the first lines of Eqs. (6) and (7) onto $\left|p\right\rangle$ and $\left|u_{ky}\right\rangle$, and project the remaining second lines onto the free-space plane-wave basis $\left|f_{ky}\right\rangle$ [16,18]. Projections onto $\left|p\right\rangle$ and $\left|u_{ky}\right\rangle$ yield four coupled integral equations,

$$\begin{aligned} 1-R&=\frac{1}{\left\langle p\right|\left.p\right\rangle}\int_{-\infty}^{\infty}dk_y\left(\alpha_{ky}+\beta_{ky}\right)\left\langle p\right|\left.f_{ky}\right\rangle,\quad T=\frac{1}{\left\langle p\right|\left.p\right\rangle}\int_{-\infty}^{\infty}dk_y\left(\alpha_{ky}e^{ik_xg}+\beta_{ky}e^{-ik_xg}\right)\left\langle p\right|\left.f_{ky}\right\rangle,\\ \rho_{ky}&=-\frac{1}{2}\frac{q_y}{k_y-q_y}\left(\alpha_{ky}+\beta_{ky}\right)-\frac{1}{2\pi i}\frac{k_y}{k_y-q_y}\int_{-\infty}^{\infty}dk_{y1}\left(\alpha_{ky1}+\beta_{ky1}\right)K\left(k_{y1},k_y\right),\\ \tau_{ky}&=-\frac{1}{2}\frac{q_y}{k_y-q_y}\left(\alpha_{ky}e^{ik_xg}+\beta_{k_y}e^{-ik_xg}\right)-\frac{1}{2\pi i}\frac{k_y}{k_y-q_y}\int_{-\infty}^{\infty}dk_{y1}\left(\alpha_{ky1}e^{ik_{x1}g}+\beta_{ky1}e^{-ik_{x1}g}\right)K\left(k_{y1},k_y\right), \end{aligned} \tag{8}$$

where T, R, ρ_{ky}, τ_{ky}, α_{ky} and β_{ky} can be determined by another set of two coupled integral equations obtained from the projection onto $\left|f_{ky}\right\rangle$,

$$\begin{aligned} \alpha_{ky}-\beta_{ky}+\rho_{ky}\frac{2q_y}{k_y+q_y}&=\left(1+R\right)\frac{1}{2\pi}\frac{q_x}{k_x}\left\langle f_{ky}\right|\left.p\right\rangle+\frac{2}{\pi i}\int_{-\infty}^{\infty}dk_{y1}\rho_{ky1}\frac{k_{x1}}{k_x}\frac{k_y}{k_{y1}+q_y}K\left(k_{y1},k_y\right),\\ \alpha_{ky}e^{ik_xg}-\beta_{ky}e^{-ik_xg}-\tau_{ky}\frac{2q_y}{k_y+q_y}&=T\frac{1}{2\pi}\frac{q_x}{k_x}\left\langle f_{ky}\right|\left.p\right\rangle-\frac{2}{\pi i}\int_{-\infty}^{\infty}dk_{y1}\tau_{ky1}\frac{k_{x1}}{k_x}\frac{k_y}{k_{y1}+q_y}K\left(k_{y1},k_y\right). \end{aligned} \tag{9}$$

Here, $k_{x1}\equiv\sqrt{\varepsilon_sk_0^2-k_{y1}^2}$. Note that integral equations in Eqs. (8) and (9) are in the Fredholm form. The complexity of Eqs. (8) and (9) is due to the presence of the kernel $K(k_{y1},k_y)$, which originated from the non-vanishing dependency between two vectors $\left|u_{ky1}\right\rangle$ and $\left|f_{ky}\right\rangle$ with $k_y\neq k_{y1}$. To solve the

coupled integral equations, we adopt super-lattice approximation (SLA) that allows semi-analytic treatment via the quantization of the integrals.

III. Super-lattice approximation (SLA)

SLA is a method to approximate the system as a periodic one with sufficiently large periodicity, allowing quantitatively accurate calculation via quantization of the bases. To make a periodic system, consider that the 2D crystal is embedded in two perfect electric conductor (PEC) plates, located at $y = \pm d$ and parallel to x-axis. We assume that d is much larger than the 2DSP wavelength. The system is now effectively periodic in y-direction through the PEC boundary condition,

$$\frac{\partial}{\partial y}\langle y|H_z\rangle\Big|_{y=\pm d} = 0. \tag{10}$$

By Eq. (10), the momenta of diffracting waves can be quantized. After the same field expansions and projection procedures, we can rewrite Eq. (8) as

$$\begin{aligned}
&1-R = \frac{1}{\langle p|p\rangle}\sum_{m=1}^{\infty}\langle p|f_{ky,m}\rangle(\alpha_m+\beta_m), \quad T = \frac{1}{\langle p|p\rangle}\sum_{m=1}^{\infty}\langle p|f_{ky,m}\rangle\left(\alpha_m e^{ik_{x,m}g}+\beta_m e^{-ik_{x,m}g}\right),\\
&\rho_n = \frac{1+e^{-2i\kappa_{y,n}d}}{2d+\dfrac{\sin\left(2\kappa_{y,n}d\right)}{\kappa_{y,n}}}\sum_{m=1}^{\infty}\left(\alpha_m+\beta_m\right)\frac{k_{y,m}}{k_{y,m}^2-\kappa_{y,n}^2},\\
&\tau_n = \frac{1+e^{-2i\kappa_{y,n}d}}{2d+\dfrac{\sin\left(2\kappa_{y,n}d\right)}{\kappa_{y,n}}}\sum_{m=1}^{\infty}\left(\alpha_m e^{ik_{x,m}g}+\beta_m e^{-ik_{x,m}g}\right)\frac{k_{y,m}}{k_{y,m}^2-\kappa_{y,n}^2},
\end{aligned} \tag{11}$$

and Eq. (9) as

$$\alpha_m - \beta_m = \frac{1}{d}\frac{q_x}{k_{x,m}}(1+R)\langle f_{ky,m}|p\rangle - \frac{1}{d}\sum_{n=1}^{\infty}\rho_n \frac{\kappa_{x,n}}{k_{x,m}}\frac{2k_{y,m}}{k_{y,m}^2-\kappa_{y,n}^2}\left(1+e^{2i\kappa_{y,n}d}\right),$$
$$\alpha_m e^{ik_{x,m}g} - \beta_m e^{-ik_{x,m}g} = T\frac{1}{d}\frac{q_x}{k_{x,m}}\langle f_{ky,m}|p\rangle + \frac{1}{d}\sum_{n=1}^{\infty}\tau_n \frac{\kappa_{x,n}}{k_{x,m}}\frac{2k_{y,m}}{k_{y,m}^2-\kappa_{y,n}^2}\left(1+e^{2i\kappa_{y,n}d}\right). \quad (12)$$

Here, $k_{y,m} \equiv (2m\text{-}1)\pi/2d$ is the quantized momentum of diffracting waves in the nanogap with mode number m, and $k_{x,m} \equiv \sqrt{\varepsilon_s k_0^2 - k_{y,m}^2}$. The momentum of diffracting waves in 2D crystal region is denoted by $\kappa_{y,m}$ that can be obtained by the following eigenvalue equation,

$$2q_y e^{iq_y d + i\kappa_{y,m}d} - q_y e^{2i\kappa_{y,m}d} - \kappa_{y,m} e^{2i\kappa_{y,m}d} = q_y - \kappa_{y,m}. \quad (13)$$

Then, $\kappa_{x,m} \equiv \sqrt{\varepsilon_s k_0^2 - \kappa_{y,m}^2}$ can also be fixed. One can find that Eq. (13) satisfies the orthogonal relationship between bases of 2DSP and diffracting waves. We note that, to simplify the problem, the momentum of 2DSP is not quantized. This assumption is a reasonable approximation when d is much larger than the 2DSP wavelength.

Now, we can truncate the total number of the quantized modes, and rewrite the second and third lines of Eq. (11) and Eq. (12) in the following coupled matrix equations,

$$\boldsymbol{\rho} = \mathbf{M_a}(\boldsymbol{\alpha}+\boldsymbol{\beta}), \quad \boldsymbol{\tau} = \mathbf{M_a}(\mathbf{G_+}\boldsymbol{\alpha} + \mathbf{G_-}\boldsymbol{\beta}),$$
$$\boldsymbol{\alpha} - \boldsymbol{\beta} = (1+R)\mathbf{c} - \mathbf{M_b}\boldsymbol{\rho}, \quad \mathbf{G_+}\boldsymbol{\alpha} - \mathbf{G_-}\boldsymbol{\beta} = T\mathbf{c} + \mathbf{M_b}\boldsymbol{\tau}. \quad (14)$$

Here, $\boldsymbol{\rho}$, $\boldsymbol{\tau}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\mathbf{c}$ are column vectors defined as

$$\boldsymbol{\rho}^T \equiv [\rho_1, \rho_2, \cdots, \rho_N], \ \boldsymbol{\tau}^T \equiv [\tau_1, \tau_2, \cdots, \tau_N], \ \boldsymbol{\alpha}^T \equiv [\alpha_1, \alpha_2, \cdots, \alpha_N], \ \boldsymbol{\beta}^T \equiv [\beta_1, \beta_2, \cdots, \beta_N],$$
$$\mathbf{c}^T \equiv [c_1, c_2, \cdots, c_N], \ c_m \equiv \frac{1}{d}\frac{q_x}{k_{x,m}}\langle f_{ky,m}|p\rangle, \quad (15)$$

and $\mathbf{M_a}$, $\mathbf{M_b}$, $\mathbf{G_+}$, and $\mathbf{G_-}$ are $N \times N$ matrices with matrix elements

$$M_a(m,n) \equiv \frac{1+e^{-2i\kappa_{y,m}d}}{2d+\dfrac{\sin\left(2\kappa_{y,m}d\right)}{\kappa_{y,m}}}\frac{k_{y,n}}{k_{y,n}^2-\kappa_{y,m}^2},\ M_b(m,n) \equiv \frac{1}{d}\frac{\kappa_{x,n}}{k_{x,m}}\frac{2k_{y,m}}{k_{y,m}^2-\kappa_{y,n}^2}\left(1+e^{2i\kappa_{y,n}d}\right),$$
$$G_{\pm}(m,m) \equiv e^{\pm ik_{x,m}g}. \tag{16}$$

N is the number of considered quantized modes. Note that $\mathbf{G}_{\pm}$ are diagonal matrices. After some manipulations, Eq. (14) can be simplified as

$$\boldsymbol{\alpha} = -T\mathbf{M_{f1}c} + (1+R)\mathbf{M_{f2}c}, \quad \boldsymbol{\beta} = (1+R)\mathbf{G}_{+}\mathbf{M_{f1}c} - T\mathbf{G}_{+}\mathbf{M_{f2}c}, \tag{17}$$

where $\mathbf{M_{f1,2}}$ are defined as

$$\mathbf{M_{f1}} \equiv \left(\mathbf{M}_{\pm}\mathbf{G}_{+}\mathbf{M}_{\mp}^{-1}\mathbf{M}_{\pm}\mathbf{G}_{+} - \mathbf{M}_{\mp}\right)^{-1},\ \mathbf{M_{f2}} \equiv \left(\mathbf{M}_{\pm}\mathbf{G}_{+} - \mathbf{M}_{\mp}\mathbf{G}_{-}\mathbf{M}_{\pm}^{-1}\mathbf{M}_{\mp}\right)^{-1},$$
$$\mathbf{M}_{\pm} \equiv (\mathbf{I}+\mathbf{M_b}\mathbf{M_a})\mathbf{G}_{-}, \quad \mathbf{M}_{\mp} \equiv (\mathbf{I}-\mathbf{M_b}\mathbf{M_a})\mathbf{G}_{+}. \tag{18}$$

Here, **I** is the identity matrix. By inserting Eq. (17) into the first line of Eq. (11), we can finally obtain reflection and tunneling coefficients *R* and *T* as

$$T = \frac{2B_2}{B_1^2 - B_2^2}, \quad R = -1 + \frac{2B_1}{B_1^2 - B_2^2}, \tag{19}$$

with B_1 and B_2 given by

$$B_1 \equiv 1 + \mathbf{a}^T\left(\mathbf{M_{f2}} + \mathbf{G}_{+}\mathbf{M_{f1}}\right)\mathbf{c}, \quad B_2 \equiv \mathbf{a}^T\left(\mathbf{M_{f1}} + \mathbf{G}_{+}\mathbf{M_{f2}}\right)\mathbf{c},$$
$$\mathbf{a}^T \equiv [a_1, a_2, \cdots a_N], \ \ a(m) \equiv \frac{\langle p | f_{ky,m}\rangle}{\langle p | p\rangle} = \frac{\int_{-d}^{d} dy \dfrac{y}{|y|} e^{iq_y|y|}\sin\left(k_{y,m}y\right)}{\int_{-d}^{d} dy e^{2iq_y|y|}}. \tag{20}$$

Shown in Fig. 3 are gap-size-dependent tunneling and reflection coefficients. We find that the tunneling efficiency decays anomalously rapidly with deep sub-λ_{sp} gap size, e.g., $g<0.01\lambda_{sp}$, while it exhibits much slower decay when the gap size gets larger. For comparison, we also calculated

the coefficients by using the finite-difference time-domain (FDTD) method. Results from SLA, based on Eq. (8) and (9), are in excellent agreement with FDTD results. We also note that the radiative loss involved by the tunneling is negligibly small for strongly-confined 2DSP ($q_x>>k_0$) with ideal lossless crystals, as $|T|^2+|R|^2$ in Fig. 3 is very close to 1. This means that the diffracted waves in both free-space and crystal regions are mostly in the form of evanescent waves that do not transfer the electromagnetic energy to the far-field.

IV. Born Approximation

SLA as we have discussed so far allows very accurate results, but it is not easy to find a closed form of 2DSP tunneling behavior in this approximation. To understand the surprisingly strong gap size dependency of the tunneling behavior with $g\rightarrow 0$ limit, we apply the Born approximation (BA) to the integral equations to obtain closed forms of T and R. A straight forward calculation with the first BA gives rise to the closed-forms of tunneling and reflection coefficients [22]:

$$T_{BA} = \frac{2I_2}{I_1^2 - I_2^2}, \quad R_{BA} = -1 + \frac{2I_1}{I_1^2 - I_2^2}. \tag{21}$$

Here, for an extremely narrow gap ($g<<\lambda_{sp}$), two factors I_1 and I_2 can be written as

$$I_1 \approx 1 - \frac{2}{3\pi}\frac{q_x}{q_y}\ln\left(-\frac{iq_y g}{2}\right), \quad I_2 \approx -\frac{2}{3\pi}\frac{q_x}{q_y}\ln\left(-\frac{iq_y g}{2}\right). \tag{22}$$

Also, for a strongly-confined 2DSP, q_y can be approximated as $q_y \equiv \sqrt{\varepsilon_s k_0^2 - q_x^2} \approx iq_x\left(1 - O\left(\varepsilon_s k_0^2 / q_x^2\right)\right)$, and this gives rise to

$$T_{BA} \approx 1 - \frac{1}{1 + \frac{4i}{3\pi} \ln\left(\frac{\pi g}{\lambda_{sp}}\right)}, \quad R_{BA} \approx \frac{1}{1 + \frac{4i}{3\pi} \ln\left(\frac{\pi g}{\lambda_{sp}}\right)}. \tag{23}$$

Equation (23) clearly shows that the tunneling and reflection coefficients are governed by anomalous logarithmic dependency on the gap size, which is consistent with an earlier study [20]. This leads to very strong reflection of 2DSP even with an extremely narrow gap although no reflection is expected for exact zero gap. The origin of the logarithmic singularity is the multiple interaction between $\left| f_{ky} \right\rangle$ and $\left| u_{ky} \right\rangle$ at x=0 and g [22]. However, it should be noted that in such deep sub-λ_{sp} gap size case, quantum effects are expected to start playing important roles, especially when the gap size is in sub-nanometer scale. Several models including so-called quantum corrected model [23,24] and nonlocal hydrodynamic model [25,26] can be adopted to take into account the quantum effects.

We also note that, in Eq. (21) and (22), the effect of the surrounding medium only appears at $q_y = \sqrt{\varepsilon_s k_0^2 - q_x^2}$ for a given q_x. This suggests that, as long as $|q_x| >> |\varepsilon_s k_0|$ is satisfied, the presence of a substrate does not have a significant impact on the tunneling and reflection behavior, and that we can use the same equations for T and R by considering only the polariton momentum change induced by the substrate.

V. Phase of *R*, and the resonant tunneling

Another important physical quantity that should be taken into account is the phase information of T and R. Near-field infrared nanoscopy measurements enables real-space visualization of 2DSPs

where the phase of R directly determines the positions of nodes and anti-nodes of the 2DSPs interference patterns. Understanding the phase information is critical for quantitative interpretation of near-field studies. Figure 4(a) shows phases of T and R, calculated by SLA. In the zero-gap limit, we can see that the phase of R and T goes to $\pi/2$ and 0, respectively. However, as the gap size gets larger, the phases of R saturates very rapidly to $\pi/4$, the known value of phase shift of edge-reflected (infinitely large gap) 2DSP in the limit $q_x/k_0 \rightarrow \infty$ [18]. The phase of T is shown to saturate to $-\pi/4$ with increasing gap size. Here, we note that the difference between those two phases is shown to be $\pi/2$ independent on the gap size. This fact is quite interesting because the $\pi/2$ phase difference can be found in other completely different systems such as the tunneling of a free-space plane-wave through a lossless metallic slab.

So far, we have discussed properties of tunneling of 2DSPs in terms of gap-dependent tunneling efficiency and phase shift. In analogy with resonant quantum tunneling through multiple potential barriers, 2DSP resonant tunneling can also happen with additional nanogaps in the 2D crystal. The simplest structure is a 2D crystal with two identical gaps separated by a 2D crystal island of width L, as shown in Fig. 4(b). To be rigorous, we can use the same field expansions as in Eq. (1) for all five separated regions. However, as shown in Fig. 2(a), the diffracting waves exhibit a rapid decay with increasing distance from the nanogap. This nature suggests that the interaction between two nanogaps is mainly mediated by the 2DSP $|p\rangle$ and the role of unbounded diffracting modes $|u_{ky}\rangle$ in the nanogap interaction is relatively small, especially when two nanogaps are separated by a sufficiently wide 2D crystal island. Then, the system can be reduced to a simple cavity in which only a multiple reflection of 2DSP on the island needs to be taken into account. This simple cavity model immediately gives rise to the tunneling and reflection coefficients T_{double} and R_{double} as

$$T_{double} = \frac{T^2 e^{iq_x L}}{1 - R^2 e^{2iq_x L}}, \quad R_{double} = R + \frac{T^2 R e^{2iq_x L}}{1 - R^2 e^{2iq_x L}}, \tag{24}$$

where T and R respectively are tunneling and reflection coefficients in Eq. (19). Figure 4(b) shows tunneling of 2DSP with varying island width L. We can clearly see that resonant tunneling arises periodically with the island width. Also, for three exemplary gap sizes, $0.05\lambda_{sp}$, $0.1\lambda_{sp}$ and $0.2\lambda_{sp}$, the resonance condition is shown not to be highly dependent on them. This stems from the rapid saturation of phase of R with increasing gap size, as shown in Fig. 4(a). Therefore, when the gap size is not extremely small, the phase of R is close to $\pi/4$, and we can approximately define the resonant tunneling condition as

$$L_{res} \approx \frac{\lambda_{sp}}{8}(4n-1), \quad n = 1, 2, 3... \tag{25}$$

The first two solutions for Eq. (25) are denoted as two vertical dashed lines in Fig. 4(b). We also point out that, owing to the nearly non-radiative nature of the tunneling, resonant tunneling with almost 100% efficiency is possible with an ideal lossless 2D crystal. Overall, analytic results based on the simple cavity-model show good agreement with FDTD results, justifying our assumption that the interaction between two gaps is dominated by $|p\rangle$, and that role of $|u_{ky}\rangle$ in the resonant tunneling is small.

VI. Conclusion

In conclusion, we have discussed the tunneling of two-dimensional surface polaritons (2DSPs) through nanogaps in 2D crystals, applicable to all types of 2DSPs that support strong

electromagnetic field in the proximity of the crystal's surface. Through rigorous analytic calculations based on the Maxwell's theory, three main properties of the 2DSP tunneling are revealed including anomalous logarithmic singularity in the gap-dependent tunneling efficiency, nearly non-radiative nature, and phase information. Our theory provides not only a quantitative understanding of the 2DSP tunneling through nanogaps in 2D crystals which is essential for further development of relevant theories in low-dimensional polaritonics and applications, but also full details of the electromagnetic field configuration of the system, particularly important for the understanding of the near-field infrared nano-imaging studies.

Acknowledgments

This work was supported by Office of Basic Energy Science, Department of Energy under contract No. DE-AC02-05CH11231 (SP2 program). J. H. Kang was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIP) (NRF-2018R1C1B6009007).

Appendix – First Born approximation

The first Born approximation (BA) is equivalent to letting $\langle u_{k1} | f_{ky} \rangle = 0$ with $k \neq k_1$ in Eq. (4), (8) and (9), by ignoring the coupling between two different diffracting waves in the nanogap and the 2D crystal. The first BA applied to the third and fourth lines of Eq. (8) yields

$$\rho_{ky} \approx -2\frac{q_y}{k_y - q_y}\left(\alpha_{ky} + \beta_{ky}\right), \quad \tau_{ky} \approx -2\frac{q_y}{k_y - q_y}\left(\alpha_{ky}e^{ik_x g} + \beta_{k_y}e^{-ik_x g}\right), \tag{A1}$$

and from Eq. (9), we have

$$\begin{aligned} \alpha_{ky} - \beta_{ky} &\approx (1+R)\frac{1}{2\pi}\frac{q_x}{k_x}\langle f_{ky} | p\rangle - \rho_{ky}\frac{2q_y}{k_y + q_y}, \\ \alpha_{ky}e^{ik_x g} - \beta_{ky}e^{-ik_x g} &\approx T\frac{1}{2\pi}\frac{q_x}{k_x}\langle f_{ky} | p\rangle + \tau_{ky}\frac{2q_y}{k_y + q_y}. \end{aligned} \tag{A2}$$

From Eq. (A1) and (A2), we can suppress ρ_{ky} and τ_{ky}, and obtain

$$\begin{aligned} \alpha_{ky} + \beta_{ky} &= \langle f_{ky} | p\rangle\left[F_{ky}(1+R) - G_{ky}T\right], \\ \alpha_{ky}e^{ik_x g} + \beta_{ky}e^{-ik_x g} &= \langle f_{ky} | p\rangle\left[G_{ky}(1+R) - F_{ky}T\right], \end{aligned} \tag{A3}$$

with F_{ky} and G_{ky} given by

$$\begin{aligned} F_{ky} &\equiv \frac{1}{2\pi}\frac{q_x}{k_x}\frac{\left(\left(k_y^2 - 2q_y^2\right) + e^{2ik_x g}k_y^2\right)\left(k_y^2 - q_y^2\right)}{\left(k_y^2 - 2q_y^2\right)^2 - e^{2ik_x g}k_y^4}, \\ G_{ky} &\equiv \frac{1}{2\pi}\frac{q_x}{k_x}\frac{2e^{ik_x g}\left(k_y^2 - q_y^2\right)^2}{\left(k_y^2 - 2q_y^2\right)^2 - e^{2ik_x g}k_y^4}. \end{aligned} \tag{A4}$$

By substituting Eq. (A3) into the first line of Eq. (8), we arrive at Eq. (21) with two factors I_1 and I_2 defined by

$$I_1 \equiv 1 + \frac{1}{\langle p | p\rangle}\int_{-\infty}^{\infty} dk_y \left|\langle f_{ky} | p\rangle\right|^2 F_{ky}, \quad I_2 \equiv \frac{1}{\langle p | p\rangle}\int_{-\infty}^{\infty} dk_y \left|\langle f_{ky} | p\rangle\right|^2 G_{ky}, \tag{A5}$$

that can be reduced to Eq. (22) with the logarithmic dependency.

To find the origin of the logarithm dependency, we look at Eq. (A1) and (A2) again. We note that, under the first Born approximation, the factor $-q_y/2(k_y-q_y)$ in Eq. (A1) is a result of projection of $\left|f_{ky}\right\rangle$ onto $\left|u_{ky}\right\rangle$, representing the dependency between two vectors. Likewise, the factor $2q_y/(k_y+q_y)$ is a result of the reversed projection. By defining $C_{f\to u} \equiv -q_y/2(k_y-q_y)$ and $C_{u\to f} \equiv 2q_y/(k_y+q_y)$, we can rewrite Eq. (A2) as

$$\begin{aligned}
&\alpha_{ky}\left(1+C_{u\to f}C_{f\to u}\right)-\beta_{ky}\left(1-C_{u\to f}C_{f\to u}\right)=\left(1+R\right)\frac{1}{2\pi}\frac{q_x}{k_x}\left\langle f_{ky}\middle|p\right\rangle,\\
&\alpha_{ky}e^{ik_xg}\left(1-C_{u\to f}C_{f\to u}\right)-\beta_{ky}e^{-ik_xg}\left(1+C_{u\to f}C_{f\to u}\right)=T\frac{1}{2\pi}\frac{q_x}{k_x}\left\langle f_{ky}\middle|p\right\rangle.
\end{aligned} \tag{A6}$$

The physical meaning of the factor $C_{u\to f}\,C_{f\to u}$ is the coupling strength between $\left|f_{ky}\right\rangle$ and $\left|u_{ky}\right\rangle$, determining the electromagnetic coupling of two different vectors at given interfaces (x=0 and g). Here, we newly introduce the coupling strength W such that

$$W \equiv C_{u\to f}C_{f\to u} = -\frac{q_y^2}{k_y^2-q_y^2}. \tag{A7}$$

Then, α_{ky} and β_{ky} can be simplified as

$$\begin{aligned}
\alpha_{ky} &= \frac{1}{2\pi}\frac{q_x}{k_x}\left\langle f_{ky}\middle|p\right\rangle\frac{-\left(1+R\right)\left(1+W\right)e^{-ik_xg}+T\left(1-W\right)}{e^{ik_xg}\left(1-W\right)^2-e^{-ik_xg}\left(1+W\right)^2},\\
\beta_{ky} &= \frac{1}{2\pi}\frac{q_x}{k_x}\left\langle f_{ky}\middle|p\right\rangle\frac{-\left(1+R\right)\left(1-W\right)e^{ik_xg}+T\left(1+W\right)}{e^{ik_xg}\left(1-W\right)^2-e^{-ik_xg}\left(1+W\right)^2}.
\end{aligned} \tag{A8}$$

Clearly, the terms involving W and g in Eq. (A8) are in the multiple reflection form, implicating the multiple interaction between $\left|f_{ky}\right\rangle$ and $\left|u_{ky}\right\rangle$ at x=0 and g. These terms are the origin of the logarithmic dependency.

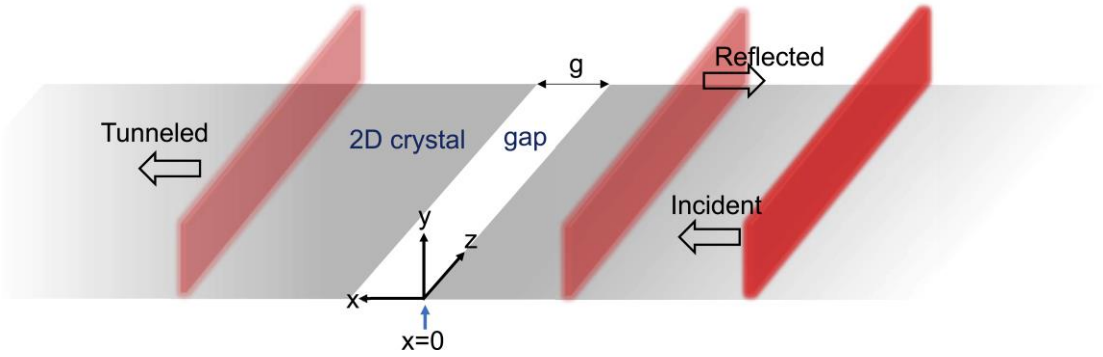

Figure 1. Schematic of 2DSP tunneling through an abrupt free-space gap in a 2D crystal.

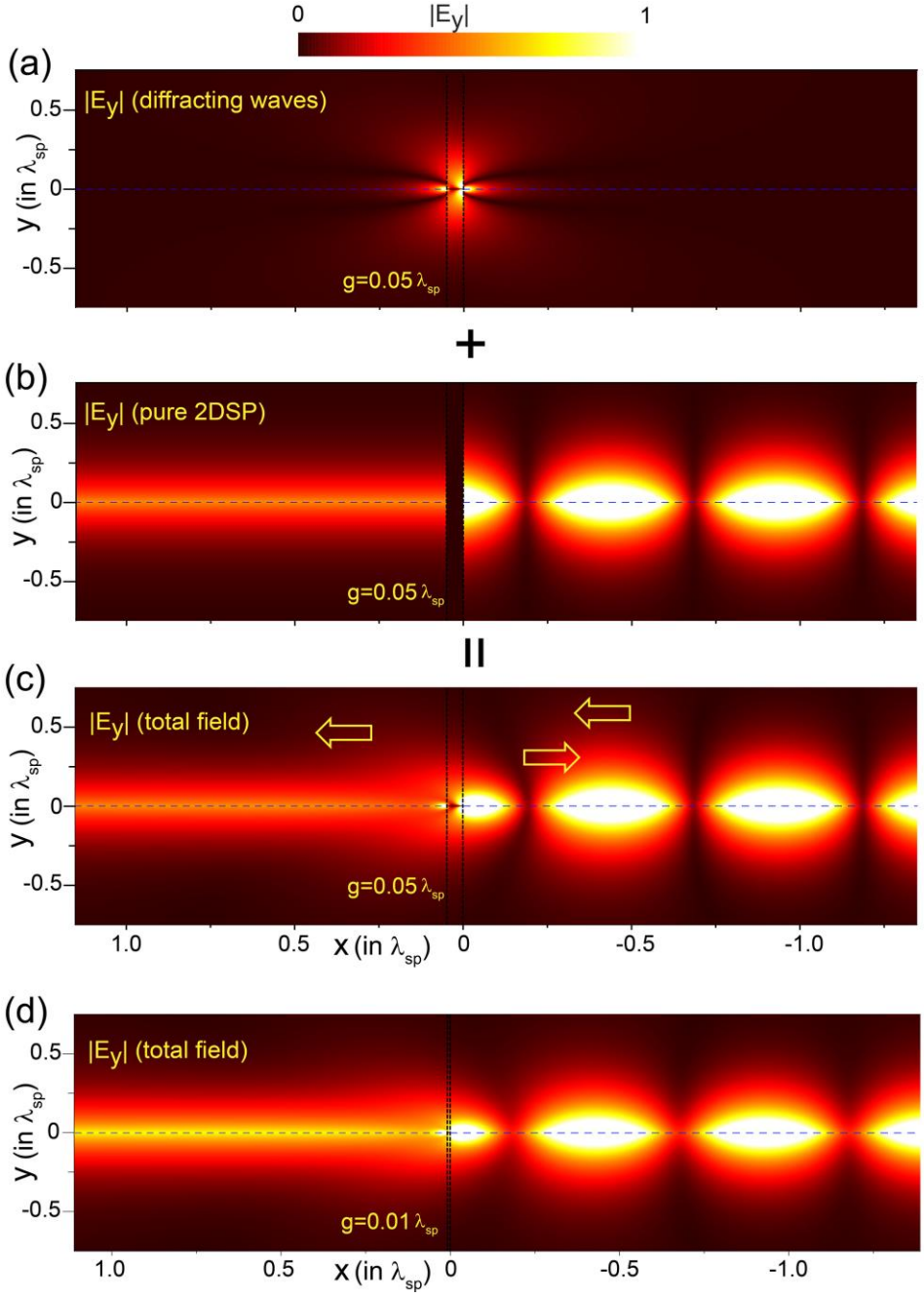

Figure 2. Analytically calculated field maps of $|E_y|$ for separated (a) diffracting waves and (b) 2DSP, obtained from super-lattice approximation (SLA) based on Eq. (2) and (3). (c, d) Maps of total field for two different gap sizes. The vertical dashed lines denote the two boundaries between the crystal and the gap. We set $\varepsilon_s = 1$.

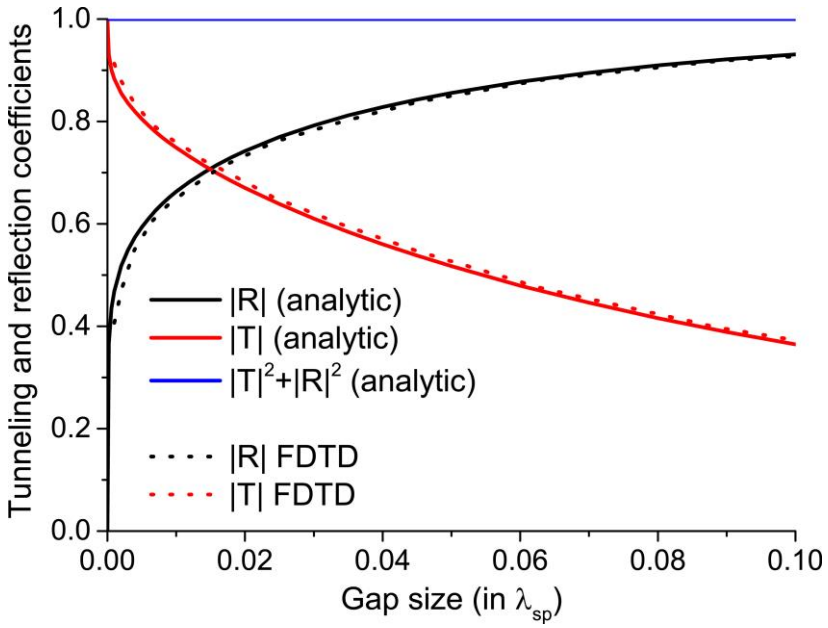

Figure 3. Analytically (SLA) and numerically (FDTD) calculated tunneling and reflection amplitudes against varying gap size. For both analytic and numerical calculations, we set q_x/k_0=50 which is close to the momentum ratio of graphene plasmons near 6 μm photon wavelength. In the FDTD numerical calculation, the thickness of 2D crystal is set to be $\lambda_{sp}/400$ that corresponds to a crystal with 0.3 nm thickness and 120 nm 2DSP wavelength. For both cases, we set ε_s = 1.

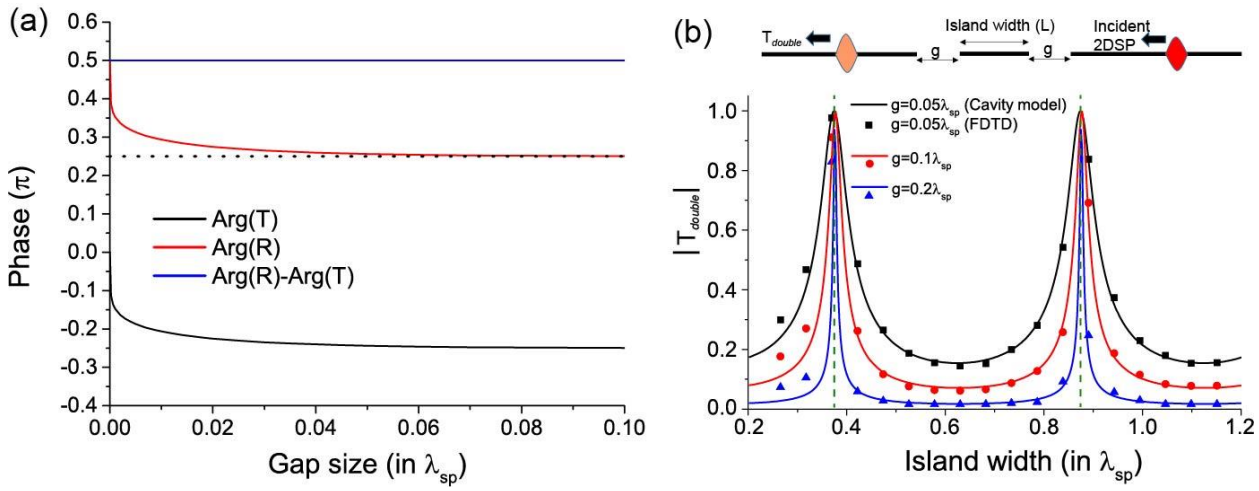

Figure 4. (a) Phases of T and R against the gap size. The horizontal dotted line denotes $\pi/4$, the limiting value of phase of R with infinitely large gap size [18]. (b) Tunneling of 2DSP through two identical gaps. The gaps are separated by an island 2D crystal of width L. Two green vertical dashed lines denote the first two resonance conditions predicted by Eq. (25). For both (a) and (b), q_x is set as $50k_0$. We considered free-standing crystals ($\varepsilon_s = 1$)

References

[1] W. Zhou, J. Lee, J. Nanda, S. T. Pantelides, S. J. Pennycook, and J.-C. Idrobo, Nat. Nanotechnol. **7**, 161 (2012).

[2] Z. Fei, A. S. Rodin, G. O. Andreev, W. Bao, A. S. McLeod, M. Wagner, L. M. Zhang, Z. Zhao, M. Thiemens, G. Dominguez, M. M. Fogler, A. H. C. Neto, C. N. Lau, F. Keilmann, and D. N. Basov, Nature **487**, 82 (2012).

[3] A. Woessner, M. B. Lundeberg, Y. Gao, A. Principi, P. Alonso-González, M. Carrega, K. Watanabe, T. Taniguchi, G. Vignale, M. Polini, J. Hone, R. Hillenbrand, and F. H. L. Koppens, Nat. Mater. **14**, 421 (2014).

[4] H. Con, T. M. Plasmonics, V. W. Brar, M. S. Jang, M. Sherrott, J. J. Lopez, and H. A. Atwater, Nano Lett. **13**, 2541 (2013).

[5] Z. Shi, H. A. Bechtel, S. Berweger, Y. Sun, B. Zeng, C. Jin, H. Chang, M. C. Martin, M. B. Raschke, and F. Wang, ACS Photonics **2**, 790 (2015).

[6] S. Dai, Z. Fei, Q. Ma, A. S. Rodin, M. Wagner, A. S. McLeod, M. K. Liu, W. Gannett, W. Regan, K. Watanabe, T. Taniguchi, M. Thiemens, G. Dominguez, A. H. C. Neto, A. Zettl, F. Keilmann, P. Jarillo-Herrero, M. M. Fogler, and D. N. Basov, Science (80-.). **343**, 1125 (2014).

[7] C. Chakraborty, L. Kinnischtzke, K. M. Goodfellow, R. Beams, and a. N. Vamivakas, Nat. Nanotechnol. **10**, 507 (2015).

[8] Z. Ye, T. Cao, K. O'Brien, H. Zhu, X. Yin, Y. Wang, S. G. Louie, and X. Zhang, Nature **513**, 214 (2014).

[9] D. Li, R. Cheng, H. Zhou, C. Wang, A. Yin, Y. Chen, N. O. Weiss, Y. Huang, and X. Duan, Nat. Commun. **6**, 7509 (2015).

[10] J. Lagois and B. Fischer, Solid State Commun. **18**, 1519 (1976).

[11] D. N. Basov, M. M. Fogler, and F. J. Garcıa De Abajo, Science (80-.). **354**, 195 (2016).

[12] J. D. Caldwell, A. Kretinin, Y. Chen, V. Giannini, M. M. Fogler, Y. Francescato, C. T. Ellis, J. G. Tischler, C. R. Woods, A. J. Giles, M. Hong, K. Watanabe, T. Taniguchi, S. a. Maier, and K. S. Novoselov, Nat. Commun. **5**, 5221 (2014).

[13] W. Gao, G. Shi, Z. Jin, J. Shu, Q. Zhang, R. Vajtai, P. M. Ajayan, J. Kono, and Q. Xu, Nano Lett. **13**, 3698 (2013).

[14] L. Ju, B. Geng, J. Horng, C. Girit, M. Martin, Z. Hao, H. a Bechtel, X. Liang, A. Zettl, Y. R. Shen, and F. Wang, Nat. Nanotechnol. **6**, 630 (2011).

[15] H. Yan, X. Li, B. Chandra, G. Tulevski, Y. Wu, M. Freitag, W. Zhu, P. Avouris, and F. Xia, Nat. Nanotechnol. **7**, 330 (2012).

[16] A. Y. Nikitin, T. Low, and L. Martin-Moreno, Phys. Rev. B **90**, 041407(R) (2014).

[17] L. Du, D. Tang, and X. Yuan, Opt. Express **22**, 749 (2014).

[18] J.-H. Kang, S. Wang, Z. Shi, W. Zhao, E. Yablonovitch, and F. Wang, Nano Lett. **17**, 1768 (2017).

[19] J. A. Gerber, S. Berweger, B. T. O'Callahan, and M. B. Raschke, Phys. Rev. Lett. **113**, 055502 (2014).

[20] B.-Y. Jiang, E. J. Mele, and M. M. Fogler, Optics Express **26**, 17209 (2018).

[21] N. Alù, Andrea, Engheta, J. Opt. Soc. Am. B **23**, 571 (2006).

[22] see the Appendix for detail.

[23] R. Esteban, A. G. Borisov, P. Nordlander, and J. Aizpurua, Nat. Commun. **3**, 825 (2012).

[24] L. Mao, Z. Li, B. Wu, and H. Xu, Appl. Phys. Lett. **94**, 243102 (2009).

[25] F. J. García de Abajo, J. Phys. Chem. C **112**, 17983 (2008).

[26] J. M. McMahon, S. K. Gray, and G. C. Schatz, Nano Lett. **10**, 3473 (2010).