Three-dimensional topological insulator based nanospaser

Hari P. Paudel, Vadym Apalkov, and Mark I. Stockman

Three-dimensional topological insulator based nanospaser

Hari P. Paudel, Vadym Apalkov,* and Mark I. Stockman†

Center for Nano-Optics (CeNO) and Department of Physics and Astronomy,
Georgia State University, Atlanta, Georgia 30303, USA

After the discovery of spaser (surface plasmon amplification by stimulated of radiation), first proposed by Bergman and Stockman in 2003, it become possible to deliver optical energy beyond the diffraction limit and generate an intense source of optical field. Spaser is a nanoplasmonic counterpart of laser. One of the major advantages of spaser is its size: spaser is truly a nanoscopic device whose size can be made smaller than skin depth of the material to a size as small as the nonlocality radius (~ 1 nm). Recently, an electrically pumped graphene based nanospaser has been prosposed that operates in the mid-infrared region and utilizes nanopatch of graphene as a source of plasmons and a quantum-well cascade as its gain medium. Here we propose an optically pumped nanospaser based on 3-dimensional topological insulator (3D TI) materials such as Bi_2Se_3 that operates at an energy close to the bulk bandgap energy ~ 0.3 eV and uses the surface as a source for plasmons and its bulk as a gain medium. The population inversion is obtained in the bulk and radiative energy of exciton recombination is transferred to surface plasmons of the same material to stimulate spasing action. This is truly a nanoscale spaser as it utilizes same material for dual purposes. We show theoretically a possibility of achieving spasing with 3D TI. As the spaser operates in the mid-infrared spectral region, it can be a useful device for number of applications such as nanoscopy, nanolithography, nanospectroscopy, and semi-classical information processing.

I. INTRODUCTION

Nano-optics is a fast growing field of research in optical condensed matter science. New concepts and ideas are coming along both theoretically and experimentally in an effort to concentrate optical energy into subwavelength (few tens of nanometers) dimensions. Source of concentrated optical energy into a nanoscale regime has number of applications such as detection and spectroscopy of biological molecules based on surface-enhanced Raman scattering (SERS)[1–3], SP-assisted thermal cancer treatment[4], scanning near-field optical microscopy[5,6], thermally assisted magnetic recording[7], plasmonic enhanced generation of extreme ultraviolet (EUV) pulses[8], extreme ultraviolet to soft x-ray (XUV) pulses[9] and semi-classical information processing[10,11]. After the discovery of spaser first proposed by Bergman and Stockman in 2003[12], it become possible to concentrate the optical energy beyond the diffraction limit and obtain the intensity required to induce nonlinear phenomenon. A spaser is a nanoscale quantum generator of nanolocalized coherent and intense optical fields[11–14]. Spaser is a nanoplasmonic counterpart of the laser, however, it has number of advantages over the laser. One of the major advantages is the size: the spaser is truly a nanoscopic device whose size can be made smaller than skin depth of the material $l_s = \frac{\omega}{c} \left[Re \left(\frac{-\varepsilon_m^2}{\varepsilon_m + \varepsilon_d} \right) \right]$ ($l_s \sim 25$ nm for single valence plasmonic metals such as silver and gold), where ε_m is the metal dielectric function and ε_d is the medium dielectric constant, to a size as small as the nonlocality radius $l_n \sim v_f/\omega$, where v_f is the Fermi velocity, which is just around a nanometer.

Several experimental studies and investigations have been performed on the spaser to integrate it in practical devices where the gain medium consists of dye molecules[15–18], quantum dots[19,20], quantum wells[21] and other semiconductor nanostructures[22–24]. On the theoretical side, after the original prediction[12], number of ideas and concepts have been proposed with several different gain medium engineered wisely to understand spasing modes and reduce losses. References on the theoretical studies include, but not limited, to Refs. 11,13,25–27.

In Ref. 28 authors have proposed a nanospaser as a coherent quantum generator of surface plasmons in nanostructured patch of graphene sheet placed on top of a stack of quantum cascade wells and electrodes to electrically pump the carriers in the active element which is the quantum well in this case. Interesting electronic and optical properties of graphene arise due to the fact that it has valence and conduction bands touching each other at two distinct high symmetry points $\mathbf{K}$ and $\mathbf{K}'$ in the Brillouin zone[29]. At the vicinity of those points, energy dispersion of electrons can be described by the low energy Dirac equations, $H_G(\mathbf{k}) = \mathbf{h}(\mathbf{k}) \cdot \sigma$, where $\mathbf{h}(\mathbf{k}) = \{h_x(\mathbf{k}), h_y(\mathbf{k}), 0\}$ and $\sigma = \{\sigma_x, \sigma_y, \sigma_z\}$ are Pauli matrices. For the small $\mathbf{q} = \mathbf{k} - \mathbf{K}$ ($\mathbf{k} - \mathbf{K}'$), we have $\mathbf{h}(\mathbf{k}) = \hbar v_f \mathbf{q}$, showing that the electronic states are helical in nature "the one way traffic". This helical nature of electronic states greatly enhances the charge mobilities and results in reduced plasmonic losses, and therefore high quality factor in a doped graphene. This is one of the great advantages of the graphene nanospaser over the spaser with metallic nanoparticles placed in adjacent to a dielectric gain medium where the loss is hindering to

achieve spasing practically. In addition to that, tight confinement of spatial distribution of graphene plasmons to regions 40 times smaller than the free-space light wavelength results unprecedented level of optical energy concentration in a nanoscale[30–32].

In this article, we propose a nanospaser based on a 3D topological insulator (TI) material (Bi_2Se_3) nanostructure of few nanometer thick geometry (nanopatch) that holds distinct symmetric and antisymmetric surface plasmon modes at mid infrared frequencies. 3D TI materials such as Bi_2Se_3 are narrow bandgap semiconducting materials where the topologically protected gapless surface/interface states appear within the bulk insulating gap[33,34]. The surface electronic states of the 3D TIs resemble as that of a graphene sheet but with an additional remarkable property of spin-momentum locking in 2D plane in perpendicular direction to each other due to the presence of strong spin-orbit coupling[33,35,52] and such property has been observed in the number of experiments[34,35,37]. The Fermi velocity of the Dirac electron in 3D TI is $v_f \sim 6 \times 10^5$ m/s[34], which is less almost by a factor of 2 than the Fermi velocity in graphene. The spatial distribution of the plasmon fields is, therefore, expected to be confined in smaller regions than in graphene. Using the carrier mobility of $\mu = 10^4$cm^2/Vs at the carrier density of $n = 2 \times 10^{11}$cm^{-2} for the surface states in a thin film of Bi_2Se_3[55], for the plasmon energy $\hbar\omega_o < \hbar\omega_q < 2E_F$, where $\hbar\omega_o$ is the phonon energy and the maximum phonon energy in Bi_2Se_3 is shown to be in the 3 meV and E_F is the Fermi energy relative to surface Dirac point, we calculate the plasmon relaxation rate as $\tau = (\mu\hbar/ev_f)\sqrt{4\pi n} = 0.8$ ps, which yields the plasmon quality factor of $Q = \omega_q\tau/2 = 122$ at the plasmon energy of $\hbar\omega_q = 0.2$ eV.

Bi_2Se_3 has the rhombohedral crystal structure, and consists of five atomic layers arranged along the z-direction, known as quintuple layers (QLs). Two such QLs are coupled by a weak interaction, predominantly of the van der Waals type. At the vicinity of the Γ point in the Brillouin zone of Bi_2Se_3, the low energy of Dirac particles in the xy-plane can be described by the effective Hamiltonian[38,52], $H_{TI} = \hbar v_f (\sigma \times \mathbf{k} \cdot \hat{\mathbf{z}}) - E_F$, where $\hat{\mathbf{z}}$ is the unit vector in z-direction. The chiral nature of the electronic states with the spin locked in perpendicular direction to the momentum results in unique type of collective excitations which are, indeed, accompanied by the transverse spin fluctuations. At this point, it is pertinent to remind the fact that current can be written in terms of in-plane spin component as[38] $\mathbf{J} = v_f\sigma \times \hat{\mathbf{z}}$. The fluctuation of density in time can be calculated using the continuity equation $\partial n_q/\partial t = -\nabla \cdot \mathbf{J}$, which suggests that the charge density n_q oscillates in time together with the in-plane transverse spin density $s^T = \hat{\mathbf{z}} \cdot (\hat{\mathbf{q}} \times \mathbf{s})$. In a slab of 3D TI, current and spin density oscillations on the top surface are either in phase or out of phase with the respective components from the bottom surface, depending on the symmetric or antisymmetric plasmon modes. The strong spin-orbit interaction leads

spin to follow momentum oscillation as an electron is displaced in the momentum space. The ratio of the transverse spin amplitude and the charge density amplitude at long wavelength limit can be much larger than unity: $s^T/n_q = 1/\hbar v_f [E_F\alpha_D v_f/2q]^{1/2}$, where $\alpha_D = e^2/\hbar v_f$ is the fine structure constant for Dirac system and is 3.7 for Bi_2Se_3.

In our model, top and bottom layers of 3D TI nanopatch of thickness d (> 5) nm are spatially separated to avoid interaction between Dirac fermions from the opposite surfaces that would, otherwise, results a conventional semiconductor. However, there exists a long range Coulomb interaction that couples the opposite surfaces. At the limit $qd \to 0$, such geometry naturally supports two types of plasmon modes: symmetric (photon like) modes with $\omega_q \sim q$ and antisymmetric (plasmon like) modes with $\omega_q \sim q^{1/2}$. At a smaller thickness of the nanopatch, symmetric modes are highly Landau damped and antisymmetric modes are the one observed in the experiments. In this article, we present a proof-of-concept of a 3D TI nanospaser where self-sustained excitations of surface plasmons are fed by the radiative transitions (in resonance with the surface plasmons) occurring in the bulk part of the same materials. In Fig. 1, we show schematic of spasing processes in Bi_2Se_3. In principle, the spasing processes can be achievable in both the undoped and p-type doped crystal. Fermi level can be tuned either by doping the bulk with p-type dopants such as Sn, Mg and Ca or by introducing dopants to the surface directly which can be achieved by dosing with O_2 gas that supplies p-type carriers leaving the surface electronic states intact. The Ref. 42 shows different doping schemes in 3D TI to tune the Fermi energy. As shown in Fig. 1a, in order to acquire the necessary population inversion in the bulk of undoped Bi_2Se_3, a photon of energy $\hbar\omega \geq E_F$, pumps electrons deep from the valence band to the conduction band, creating holes in the valence band. The excited carriers from the bulk may diffuse to the surfaces after pumping and can alter the plasmon resonance frequency. However, the carrier diffusion can be controlled by pumping the carriers close to the resonance. In less than a picosecond (ps), generated holes are redistributed and populated on the top of the valence band[42], as shown in Fig. 1 **c**. The same is expected for the excited electrons in the conduction band but they are populated above the Fermi level. The excited holes form excitons with the already occupied electrons at the bottom of conduction band. The excitons undergo a radiative recombination, emitting surface plasmons. It is to be noted that due to the high dielectric constant of ~ 100 in Bi_2Se_3, there is a strong screening of the Coulomb interaction between the carriers. This results a relatively larger excitonic Bohr radius, $a_{ex} = (\epsilon m_o/\mu) a_B$ where ϵ, m_o, μ and a_B are respectively the dielectric constant, free electron mass, reduced mass and Bohr radius, of nearly 200 nm for $\mu = 0.24$ m_o. This gives rise a shallow exciton with energy for the lowest excitonic level of 0.33 meV, which is negligible as compared to the bandgap energy of
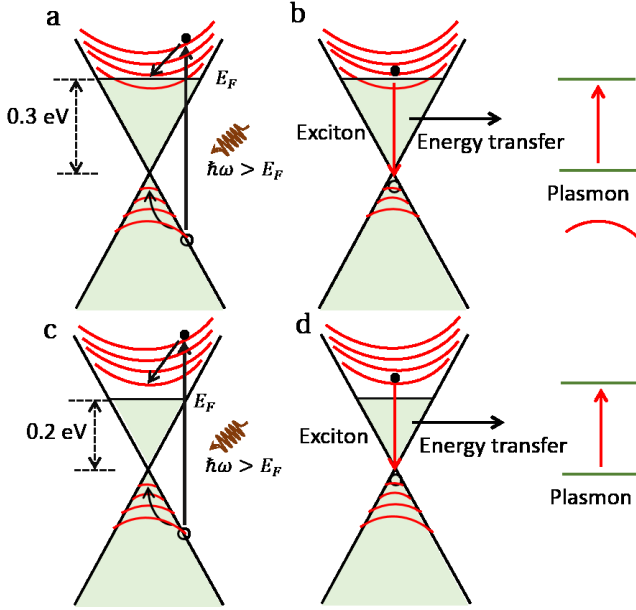
Figure 1: Schematic of spasing processes with undoped (**a**, **b**) and p-doped (**c**, **d**) Bi_2Se_3. **a**, Electron-hole pairs in the bulk can be generated with photon energy of $\hbar\omega \geq E_F$. Parabolic curves represent the bulk valence and conduction bands and the straight lines passing the gap represent the Dirac surface states. Light blue area indicate the filled Fermi sea of electrons. **b**, Within 200 fs electrons acquire quasi-equilibrium Fermi distribution in the conduction band and the same is true for holes in the valence band. Due to intraband relaxation, the excited holes are populated at the top of the valence band while the excited electrons are populated above the Fermi level, forming excitons. Those excitons radiatively recombine and the released energy is transferred to the surface plasmons. **c** and **d**, Fermi energy is reduced to 0.2 eV by p-type doping. The spasing can be achieved in a same way as in the undoped case.

0.3 eV. The plasmonic oscillations on the surface stimulate this emission that supply the feedback for the spaser action. Similar spasing mechanism can be obtained with the doped Bi_2Se_3 as shown in Fig. 1 **c** and **d**, where the Fermi energy is set at 0.2 eV above the Dirac point.

II. MODEL BASED ON RPA THEORY

The validity of the RPA theory in case of the Bi_2Se_3 can be justified as follow: the ratio of the characteristic Coulomb interaction energy to the kinetic energy is $r_s = e^2/\varepsilon_{ave}\hbar v_f$, where $\varepsilon_{ave} = (\varepsilon_{TI} + \varepsilon_m)/2$, ε_m is the medium dielectric constant, ($\varepsilon_{TI} \approx 100$,). In Bi_2Se_3, $\varepsilon_{ave} \sim 50$ for $\varepsilon_{TI} \approx 100$ and $\varepsilon_m = 1$, and $v_f = 6 \times 10^7$ cm/s implies $r_s \sim 0.07$, satisfying the condition $r_s \ll 1$ of RPA applicability. In the present case, we consider a double layer model (Fig. 2 **a**) where the bulk properties of 3D TI is incorporated by its dielectric constant that

separates the top and bottom surfaces with metal-like conductivities. This is valid at low temperature where the phonon's effects are submissive[43]. At temperature above the Debye temperature, $\Theta_D = 180$ K in Bi_2Se_3, due to increase in the rate of electron-phonon scattering events, excited carriers are transferred from the bulk conduction band to the surface states which alters the surface conductivities in experiments.

Consider a Hamiltonian that captures the properties of collective oscillations on top and bottom surfaces of 3D TI material as

$$H^{l/l'} = \hbar v_f \sum_{\alpha,\beta,\mathbf{k},l,l'} a^+_{k,l,\alpha}(\boldsymbol{\tau} \times \mathbf{k}) \cdot \hat{z} a_{k,l,\beta}$$
$$+ \frac{1}{2S} \sum_{\mathbf{q},l,l'} v_{l,l'}(\mathbf{q}) n_{\mathbf{q},l} n_{-\mathbf{q},l'}, \qquad (1)$$

where l, l' =1 (top), 2 (bottom) surfaces, α, β are spin indices and a^+ (a) is the creation (annihilation) operator for an electron. S is the area of 2D surface. $n_{q,l} = \sum_{\mathbf{k},\alpha} a^+_{\mathbf{k}-\mathbf{q},l,\alpha} a_{\mathbf{k},l,\alpha}$ is the density operator and $\boldsymbol{\tau} = \{\tau_x, \tau_y, \}$ are the Pauli matrices. First, we proceed by calculating the linear response function under RPA theory that allows us to find the excitations of collective modes at low temperature. For two electronically decoupled 3D TI surfaces the transverse spin and density response function, $\chi(\mathbf{q}, \omega)$, is a 2×2 block diagonal matrix whose off diagonal elements are zero. The block diagonal elements in density and spin (n, s^T) basis are[38],

$$\chi^{ll}(\mathbf{q}, \omega) = \begin{pmatrix} 1 & (-1)^{l+1} x \\ (-1)^l x & -x^2 \end{pmatrix} \chi^{ll}_{nn}(\mathbf{q}, \omega), \qquad (2)$$

where $x = \frac{\omega}{v_f q}$, and the off diagonal elements determine spin and charge correlation. $\chi^{ll}_{nn}(q, \omega)$ is the density-density response function. The $\pm$ signs in $\chi^{ll}(\mathbf{q}, \omega)$ indicate that the velocity of the Dirac fermion on top and bottom surfaces has opposite sign. Under long-range Coulomb interaction within the RPA, the response function can be written as,

$$\chi^{RPA} = \frac{\chi(\mathbf{q}, \omega)}{1 - v(\mathbf{q}) \chi(\mathbf{q}, \omega)}, \qquad (3)$$

where $\chi(\mathbf{q}, \omega) = \begin{pmatrix} \chi^{11} & 0 \\ 0 & \chi^{22} \end{pmatrix}$. It is to be noted that $v(\mathbf{q})$ is a 2×2 matrix whose diagonal and off diagonal elements are two dimensional Fourier transform of Coulomb potentials and can be obtained by solving the Poisson's equation[41]:

$$v^{11/22}(q) = \frac{4\pi e^2 \varepsilon_2 \left(\cosh(qd) + (\varepsilon_{3/1}/\varepsilon_2)\sinh(qd)\right)}{q\left(\varepsilon_2(\varepsilon_1 + \varepsilon_3)\cosh(qd) + (\varepsilon_1\varepsilon_3 + \varepsilon_2)\sinh(qd)\right)} \qquad (4)$$

and

$$v^{12}(q) = \frac{4\pi e^2 \varepsilon_2}{q\left(\varepsilon_2(\varepsilon_1+\varepsilon_3)\cosh(qd) + (\varepsilon_1\varepsilon_3+\varepsilon_2)\sinh(qd)\right)}, \tag{5}$$

where ε_1 and ε_3 are the dielectric constants of the surrounding medium on the top and bottom surfaces, and $\varepsilon_2 = \varepsilon_{TI}$ is the dielectric constant of the bulk Bi_2Se_3. The response function χ provides an important information about the collective states that are excited at small transferred momentum. Poles of the response function χ^{RPA} give the excitation energies of collective modes. That means, collective mode frequencies of the system are obtained by solving $\text{Det}\left[1 - v(\mathbf{q})\chi(\mathbf{q},\omega)\right] = 0$. It is to be noted that the coupling exists solely due to long range Coulomb interaction, only the charge part of the opposite surfaces are coupled. In the region $v_f q < \omega < 2E_f/\hbar - v_f q$, collective modes of oscillations are undamped, and in other regions such modes are not observed as the energy of the modes is transferred to the particle-hole excitations. In the region $v_f q < \omega$, $\chi^{ll}_{nn}(\mathbf{q},\omega)$ is real and its long wavelength limit $q \to 0$ is given by

$$\chi^{ll}_{nn}(\mathbf{q},\omega) = \left(\frac{q^2 E^l_F}{4\pi\hbar^2\omega^2}\right)\left[1 + \frac{\hbar\omega}{4\mu^l}\ln\left|\frac{2E^l_F - \hbar\omega}{2E^l_F + \hbar\omega}\right|\right]. \tag{6}$$

Using Eq. (6), and assuming that $qd \to 0$, we obtain an antisymmetric plasmon mode as

$$\omega^2_{ant} = \frac{v^2_f k_f \alpha_D}{\varepsilon_1 + \varepsilon_3}\left[\left(\frac{1+\xi}{2}\right)^{1/2} + \left(\frac{1-\xi}{2}\right)^{1/2}\right]q \tag{7}$$

and a symmetric plasmon mode as

$$\omega^2_{sym} = \frac{v^2_f k_f \alpha_D}{\sqrt{2}\varepsilon_2}\left[\frac{(1-\xi)^{1/2}(1+\xi)^{1/2}}{(1-\xi)^{1/2} + (1+\xi)^{1/2}}\right]dq^2, \tag{8}$$

where $\xi = n_1 - n_2/n_1 + n_2$ is the density polarization and $k_f = \sqrt{4\pi(n_1+n_2)}$ is the Fermi wave vector. The mode ω_{ant} corresponds to in phase and the mode ω_{sym} corresponds to out of phase oscillations of longitudinal electric field. Note that ω_{ant} is thickness d and ε_2 independent while ω_{sym} changes with d and ε_2. In the limit $dq \to 0$, ω_{sym} disappears and only the graphene like plasmon mode exists. The ω_{ant} mode merges into the particle-hole continuum above $\hbar\omega^c_{ant} = 2E_F\alpha_D/\left[\sqrt{2}(\varepsilon_1+\varepsilon_3)\right]$, which is 0.33 eV for Bi_2Se_3 with $\varepsilon_1 = 1$, $\varepsilon = 3.8$ and $E_F = 0.3$ eV. If no coupling between top and bottom surfaces is considered, we obtain two decoupled graphene like modes. In this case, $\hbar\omega^c_{ant} = \alpha_D E_F/\left[2\varepsilon_{avg}\right] = 11$ meV for Bi_2Se_3 with $\varepsilon_{avg} \approx 50$ and $E_f = 0.3$ eV[38]. The general solution of the $\chi(\mathbf{q},\omega)$ has a singularity at $\omega = v_f q$[51]. The frequency given by Eq. (8) leads to a divergence in the general solution of $\chi(\mathbf{q},\omega)$ as the group velocity $v_{sym} = \left|\frac{d\omega_{sym}}{dq}\right|$ approaches v_f. The most general solution at $q \to 0$ for this mode is obtained in Ref. 44. Here
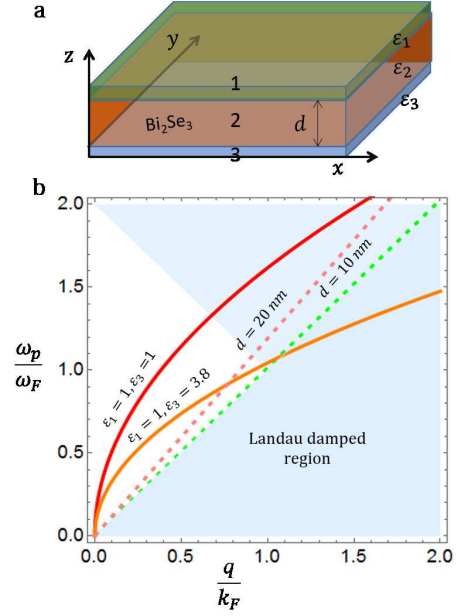

Figure 2: Geometry (a) and plasmon dispersion (b) in a nanopatch of Bi_2Se_3. Plasmon dispersion is shown for the thickness of $d = 10$ and 20 nm, and for top and bottom dielectric constant of $\varepsilon_1 = 1$ and $\varepsilon_2 = 1, 3.8$. The antisymmetric mode (red and orange solid lines) has a plasmon like dispersion and depends on the parameters E_F, ε_1, and ε_2, while the symmetric mode (pink and green dotted lines) has a photon like dispersion, and depends on the parameters E_F, and ε_2. As $dq \to 0$ symmetric mode approaches the frequency line $\omega = v_f q$.

we write that solution for $\xi = 0$ as

$$\omega_{sym} = \left(\sqrt{2}\varepsilon_2 + \alpha_D k_f d\right)v_f q/\left(\sqrt{2\varepsilon_2\left(\varepsilon_2 + \sqrt{2}\alpha_D k_f d\right)}\right). \tag{9}$$

This solution and the solution given by Eq. (7) are shown in Fig. 2 b for 10 and 20 nm thick nanopatch of Bi_2Se_3. The symmetric mode traverses along the line $\omega = v_f q$, and is pinned to the electron-hole continuum.

The extend of the topological surface states inside the bulk is measured by the localization length, $l_o = \hbar v_f/E_g$, where E_g is the bulk bandgap, and it is equal to 1.5 nm ($\sim$2 quintuple layer (QL)) in Bi_2Se_3. The quantum confinement of the carriers created at the surfaces due to downward band bending generates two distinct transport channels with two different conductivities: one due to topological surface states that extend to 2 QL and other due to two-dimensional electron gas (2DEG) trapped beneath the surface that extends up to 8 QL[45]. Considering the 2DEG may change the plasmon behaviors, nevertheless, it should be possible to separate out the effect of 2DEG, if any, from the topological surface states and write the response functions as $\chi = \chi^{TI} + \chi^{2DEG}$ for two independent channels. Assuming no particle transfer occurs between these channels, the plasmon modes

at the interface of topological surface states and 2DEG should pin to the particle-hole continuum and the modes obtained above should not be affected.

III.   SPASING WITH 3D TOPOLOGICAL INSULATOR

After understanding the behavior of the plasmon modes in 3D TI, we now analyze the nanospasing, which can be achieved for the undamped plasmon mode. To substantiate the functioning of the 3D TI spaser, we need to calculate the electric field distribution inside and outside of the geometry that supports the excitation of plasmon modes at the interfaces and interacts with the dipole transition in the bulk. The solution of the Maxwell's equations for TM modes in the regions 1, 2 and 3, shown in Fig. 2 **a**, can be written as

$$E_x^1(z) = A_1 e^{\alpha_1(z+d/2)} e^{iqx}, \qquad (10)$$

$$E_x^2(z) = (A_2 \sin(\alpha_2 z) + B_2 \cos(\alpha_2 z)) e^{iqx}, \qquad (11)$$

$$E_x^3(z) = B_3 e^{-\alpha_1(z-d/2)} e^{iqx}, \qquad (12)$$

where A_1, A_2, B_2 and B_3 are the field amplitudes, α_i, $i = 1$, 2, and 3, are the wave vectors in the regions 1, 2 and 3, respectively. The spatial decay of the electromagnetic field away from the interface is determined by the imaginary wave vectors α_i, which are given by $\alpha_i^2 = q^2 - \varepsilon_i(\omega/c)^2$. The z-component of the electric field and y-component of the magnetic field can be obtained as $E_z^i = -\frac{iq}{\alpha_i^2}\frac{dE_x^i}{dx}$ and $H_y^i = \frac{\varepsilon^i E_z^i}{qc}$. Using Maxwell's equation with the boundary conditions on the components $E_{x/z}$ and H_y at the interfaces, one can find the field amplitudes. From to the discontinuity in the tangential component of the magnetic field due to the surface current we find the determinantial equation as

$$\left[\begin{array}{cc} \left(\frac{\varepsilon_1}{\alpha_1} + \frac{4\pi i\sigma_1}{\omega}\right)\left(1 + e^{-\alpha_2 d}\right) + \frac{\varepsilon_2}{\alpha_2}\left(1 - e^{-\alpha_2 d}\right) & \left(\frac{\varepsilon_3}{\alpha_3} + \frac{4\pi i\sigma_2}{\omega}\right)\left(1 + e^{-\alpha_2 d}\right) + \frac{\varepsilon_2}{\alpha_2}\left(1 + e^{-\alpha_2 d}\right) \\ \left(\frac{\varepsilon_1}{\alpha_1} + \frac{4\pi i\sigma_1}{\omega}\right)\left(1 - e^{-\alpha_2 d}\right) + \frac{\varepsilon_2}{\alpha_2}\left(1 + e^{-\alpha_2 d}\right) & \left(\frac{\varepsilon_1}{\alpha_1} + \frac{4\pi i\sigma_2}{\omega}\right)\left(1 + e^{-\alpha_2 d}\right) + \frac{\varepsilon_2}{\alpha_2}\left(1 - e^{-\alpha_2 d}\right) \end{array} \right] \left[\begin{array}{c} A_1 \\ B_3 \end{array} \right] = 0, \qquad (13)$$

where σ_1 and σ_2 are frequency dependent conductivites for the top and bottom surfaces. Solutions of Eq. (13) to exist determinant of 2×2 matrix should be zero. For symmetric dielectric environment (equal dielectric constant on top and bottom, $\varepsilon_1 = \varepsilon_3$), we have $\sigma_1 = \sigma_2$, therefore, we obtain an equation with two different modes (symmetric and antisymmetric) as,

$$0 = \left\{ \left(\frac{\varepsilon_1}{\alpha_1} + \frac{4\pi i\sigma}{\omega}\right)\cosh\left(\frac{\alpha_2}{2}d\right) + \frac{\varepsilon_2}{\alpha_2}\sinh\left(\frac{\alpha_2}{2}d\right) \right\}$$
$$\times \left\{ \left(\frac{\varepsilon_1}{\alpha_1} + \frac{4\pi i\sigma}{\omega}\right)\sinh\left(\frac{\alpha_2}{2}d\right) + \frac{\varepsilon_2}{\alpha_2}\cosh\left(\frac{\alpha_2}{2}d\right) \right\} \qquad (14)$$

For $\alpha_2 d \to \infty$ no electromagnetic coupling exist between top and bottom surfaces and the solution reduces to two independent graphene like modes for each individual surface.

For simplicity, we define new parameters C_1, C_2 and C_3 (see Appendix) and write the coefficients A_2, B_2 and B_3 as $A_2 = C_1 A_1$, $B_2 = C_2 A_1$ and $B_3 = C_3 A_1$. The coefficient A_1 can be calculated in terms of number of plasmons using the Brillouin expression[28]. The magnetic energy in the Brillouin expression is of the order of v_f/c ($\ll 1$) smaller than the electric energy and therefore, can be neglected. For the plasmon frequency under consideration, we have $q \ll \sqrt{\varepsilon_i}\omega/c$, therefore we can write $\alpha_i \approx q$. Then, A_1 in terms of the number of plasmons, N, is obtained as

$$|A_1|^2 = \frac{8\pi N\hbar\omega_p q}{S\left[\varepsilon_1 + 2\left(\varepsilon + \omega\frac{d\varepsilon}{d\omega}\right) + \varepsilon_2\left\{C_1^2\left(\sinh qd - qd\right) + C_2^2\left(\sinh qd + qd\right)\right\} + C_3^2\left\{\varepsilon_3 + 2\left(\varepsilon + \omega\frac{d\varepsilon}{d\omega}\right)\right\}\right]}. \qquad (15)$$

The solution of Eq. (14) gives conductivity $\sigma(\omega)$ for each modes that satisfies the proper boundary conditions at the interfaces. Now with the help of RPA solutions for the dispersion relations and two-dimensional dielectric function $\varepsilon(\omega) = \varepsilon_b + 2\pi i\frac{\sigma(\omega)}{\omega}q$, where ε_b is the background dielectric constant, we can easily calculate the quantity $\varepsilon + \omega\frac{d\varepsilon}{d\omega}$ for the Dirac fermions in Eq. (15). In the limit $qd \to 0$, only antisymmetric mode appears and

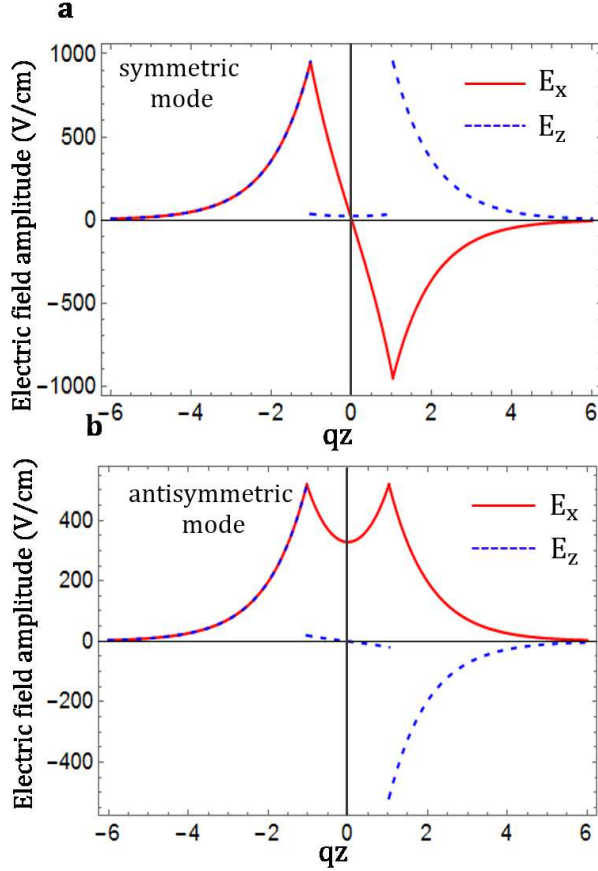Figure 3: Electric field distributions are shown for the symmetric (**a**), and the antisymmetric (**b**) plasmon modes in a 10 nm thick nanopatch of Bi_2Se_3 with the number of plasmons, $N = 2$, and symmetric dielectric environment ($\varepsilon_1 = \varepsilon_3 = 3.8$). The red solid line indicates the component E_x and the blue dotted line indicates the component E_z. Note that E_z is suppressed inside the material due to large dielectric constant ($\varepsilon_2 = \varepsilon_{TI} \approx 100$).

Eq. 15 reduces to

$$|A_1| = \sqrt{4\pi\hbar\omega_q q/[S(\varepsilon_1 + \varepsilon_2)]}. \quad (16)$$

In the limit of $qd \gg 1$, plasmon modes are degenerate with the behavior $\omega_q \sim q^{1/2}$, and in the opposite limit, $qd \ll 1$, the mode with behavior $\omega_q \sim q$ is pinned to the particle-hole continuum, and the second mode $\omega_q \sim q^{1/2}$ is similar to the one in graphene. Therefore, spasing action must be easier to obtain for the anitsymmetric mode. The electric field distributions inside and outside of the nanopatch with the number of plasmons $N = 2$, which should be sufficient to get spasing, are shown in Fig. 3. Note that the behavior of the amplitude with the number of plasmons: $A_1 \sim N^{1/2}$. There should not be any confusion in the defination of the field, since we have adapted the defination with respect to the symmetry of E_z.

For the self-sustained excitations of plasmon modes on the surface, we need a coupling between the plasmonic field and the bulk dipole transition as shown in Fig. 1. Here we take the antisymmetric mode field because this mode is undamped for $dq \to 0$. The interaction Hamiltonian describing the coupling between the plasmonic field and the electron system in the bulk is given by following expression

$$H_I = \sum_{\mathbf{q},\mathbf{k}} \Omega_{\mathbf{k},\mathbf{q}} \left(b^+_{c,\mathbf{k}-\mathbf{q}} b_{v,\mathbf{k}} c^+_{\mathbf{q}} + b^+_{v,\mathbf{k}+\mathbf{q}} b_{c,\mathbf{k}} c_{\mathbf{q}} \right), \quad (17)$$

where $b_c(b_v)$ electron annihilation operator in conduction (valence) band, c_q is the plasmon annihilation operator. Here $\Omega^*_{\mathbf{k},\mathbf{q}} = \Omega_{\mathbf{k},\mathbf{q}}$ is the Rabi frequency with plasmon's wave vector $\mathbf{q}$. The Rabi frequency for the given mode is

$$\Omega_{\mathbf{k},\mathbf{q}} = e \int \Psi^*_{c,\mathbf{k}'} \mathbf{E}_{\mathbf{q}} \cdot \mathbf{r} \Psi_{v,\mathbf{k}} d^3\mathbf{r}, \quad (18)$$

where Ψ_c and Ψ_v are the electron wavefunctions in conduction and valence band. We can write $\Psi_{c,\mathbf{k}'} = u_{c,\mathbf{k}'}(\mathbf{r}) e^{i\mathbf{k}' \cdot \mathbf{r}}$ and $\Psi_{v,\mathbf{k}} = u_{v,\mathbf{k}}(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}}$, where $u_{c,\mathbf{k}'}(\mathbf{r}) = \frac{1}{V^{1/2}}\sum_G C_c(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}}$ and $u_{v,\mathbf{k}}(\mathbf{r}) = \frac{1}{V^{1/2}}\sum_G C_v(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}}$ are periodic parts of the Bloch functions. Here $C_c(C_v)$ is the expansion coefficient for conduction (valence) band and $\mathbf{G}$ is the reciprocal lattice vector,. It is to be noted that the time dependent of the Bloch function leads to a delta function which ensures the conservation of the energy. Writing $\mathbf{r} = \mathbf{R}_j + \mathbf{r}'$, where $\mathbf{r}'$ lies within one unit cell and $\mathbf{R}_j$ is the lattice vector. Because of the periodicity of $u_\mathbf{k}(\mathbf{r})$, Rabi frequency reduces to

$$\Omega_{\mathbf{k},\mathbf{q}} = e \int u^*_{c,\mathbf{k}'}(\mathbf{r}) \mathbf{E}_{\mathbf{q}} \cdot \mathbf{r} \Psi_{v,\mathbf{k}} d^3\mathbf{r}. \quad (19)$$

Since the filed is homogeneous over the sample, we can separate out the integral over field distribution from the rest. Then Eq. (19) reduces to

$$\Omega_{\mathbf{k},\mathbf{q}} = \langle d_\mathbf{k} \rangle S \int \mathbf{E}_\mathbf{q}(z) dz, \quad (20)$$

where S is the area of the sample and $\langle d_\mathbf{k} \rangle$ is the dipole matrix element at point $\mathbf{k}$,

$$\langle d_\mathbf{k} \rangle = e \int_{\text{unit cell}} u^*_{c,\mathbf{k}'}(\mathbf{r}) \mathbf{r} u_{v,\mathbf{k}}(\mathbf{r}) d^3\mathbf{r}. \quad (21)$$

To calculate the dipole matrix element, we need to know the periodic part of the functions, $u_\mathbf{k}(\mathbf{r})$, which can be calculated using the density functional theory (DFT)[10]. Here we use the PAW (Projector Augmented Wave) approximation within the framework of DFT as implemented in VASP (Vienna Ab-initio Simulation Package). In the bulk structure of Bi_2Se_3, energy dispersion along the high symmetry lines $\Gamma \to Z \to F \to L$ are extracted and the bandgap is calculated to be 0.35 eV along $\Gamma \to Z$

direction. In the experiments, it has been observed that Bi_2Se_3 is a direct bandgap material with the bandgap of 0.3 eV at the Γ point[50]. For particular $\mathbf{k}$, periodic functions $u_{c,v}(\mathbf{r})$ are expanded over several $\mathbf{G}$ vectors. When the quasi-equilibrium condition is achieved, carriers are settled at their respective band extrema and the most significant contribution to the dipole transitions comes from the radiative transitions occuring between the band extrema. At the minimum bandgap position, the dipole matrix element, defined by Eq. (21), is calculated to be $\langle d_{\mathbf{k}} \rangle = 21$ Debye (6.2976×10^{-18} e.s.u). Using this value, the Rabi frequency, given by Eq. (20), can be readily solved for the given mode.

The dynamics of the spaser within the density matrix approach can be described by solving the equation for the density matrix $\rho_{\mathbf{k}}(t)$ in the gain medium,

$$i\hbar \dot{\rho}_{\mathbf{k}}(t) = [\rho_{\mathbf{k}}(t), H]. \qquad (22)$$

For the stationary regime, the condition for spasing can be written as[12,28],

$$\frac{(\Gamma_{cv}+\gamma)^2}{(\omega_{c,v}-\omega_q)^2+(\Gamma_{cv}+\gamma)^2}\sum_{\mathbf{k}}\left|\frac{\Omega_{c,v}(\mathbf{k},\mathbf{q})}{\hbar}\right|^2 \geq \gamma\Gamma_{cv}, \qquad (23)$$

where Γ_{cv} is the polarization relaxation rate and $\gamma = 1/\tau$ is the plasmon relaxation rate.

At the resonance, i.e. $\omega_q = \omega_{c,v}$, we obtain the condition for spasing as,

$$\left|\frac{\Omega_{k,q}}{\hbar}\right|^2 \geq \gamma\Gamma_{cv}. \qquad (24)$$

Then, the spasing condition as a limitation on τ can be written as

$$\tau \geq \tau_{min} = \frac{\Gamma_{cv}}{\left|\frac{\Omega_{k,q}}{\hbar}\right|^2}. \qquad (25)$$

Quantity τ_{min} is the minimum plasmon relaxation times necessary to drive the system in the spasing mode. The τ_{min} as a function of plasmon frequency ω_p is shown in Fig. 4 for different Fermi energies and for nanopatch thickness $d = 10$ and 20 nm. Here we take $\Gamma_{cv} = 6$ ps for Bi_2Se_3[49]. For example, for $d = 10$ nm, the plasmon frequency of 4.5×10^{14} Hz (~ 0.3) eV with τ_{min} of 0.5 ps, for example, can be achieved at Fermi energy of 0.3 eV. This plasmon frequency corresponds to the bandgap energy. It is to be noted that for $d = 20$ nm, the required plasmon frequency for spasing is red-shifted relative to same τ_{min} for $d = 10$ nm. For $d = 10$ nm, there is a broad range of plasmon frequencies in the mid-infrared range of 3×10^{14} to 6×10^{14} Hz (0.2 to 0.4 eV) where the plasmon relaxation time is short enough to occur spasing. There are just a few experimental evidences of plasmon relaxation times in 3D TI available to date. In Ref. 47 the plasmon relaxation time of few picoseconds was observed in micro-ribbon arrays. If we take a small momentum transfer limit (large mirco-ribbon thickness), a
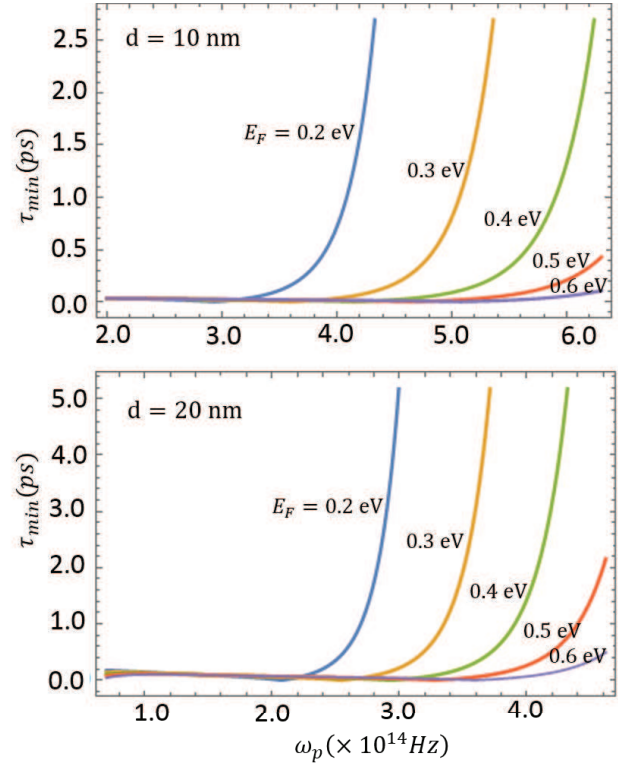

Figure 4: Minimum relaxation times of plasmons in Bi_2Se_3 sufficient for spasing are shown for $d = 10$ and 20 nm as a function of plasmon frequency at different Fermi energies. Spasing occurs for plasmon relaxation times of $\tau > \tau_{min}$. Results are shown for symmetric dielectric environment.

plasmon relaxation time of less than a picosecond is obtained. Thus, this observation along with the plasmon energy dispersion as shown in Fig. 2 tell us that it is possible to achieve the spasing in 3D TI with the plasmon frequency equal to the bulk bandgap energy.

The useful quantity that measures how strongly the plasmons are confined before they decay is plasmon quality factor Q, which is given by $Q = \omega_p\tau/2$. The spasing condition given by Eq. (25) sets a minimum limit on Q, i.e., $Q_{min} \leq Q$, required for the spasing. The minimum quality factor, Q_{min}, as function of frequency has a minimum possible value for each Fermi energy, which is determined by condition $\lambda_p \approx 2d$. For $\lambda_p < 2d$, a strong confinement of plasmons with high Q value is possible for the given parameters such as Fermi energy. In Fig. 5 we show Q_{min} as a function of ω_p for $d = 10$ nm and 20 nm, and with different Fermi energies in the symmetric dielectric environment. At the plasmon frequency of 4.5×10^{14} Hz (0.3 eV), and the Fermi energy of 0.3 eV, which is required for spasing, for $d = 10$ nm we find that $Q_{min} \approx 20$. To date, no experimental results for 3D TI are available to determine the value of Q, however, insight can be obtained from the result for Dirac plasmons in graphene where the quality factor of $Q \approx 50$ is obtained using the experimental values for the carrier's mobility and density[28]. Here, at $\lambda_p \approx 2d$, we
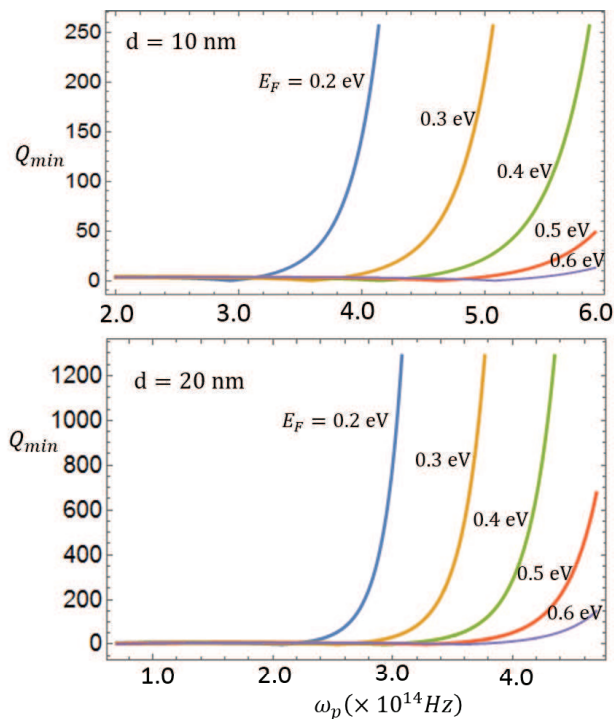
Figure 5: The minimum plasmon quality factor Q_{min} is shown as a function of plasmon frequency ω_p for $d = 10$ and 20 nm at different Fermi energies. Results are shown for symmetric dielectric environment.

have $Q_{min} = [\omega_p \tau_{min}/2]_{\lambda_p = 2d}$, which gives a deep minimum of $Q_{min} \approx 3$, a relatively low value of Q and this can be compared with the result for graphene, for which $Q_{min} \approx 1.5$[28]. At larger thickness, Q increases and the plasmon frequency is red-shifted at a given Fermi energy. If we compare the case of $d = 10$ nm and $d = 20$ nm of our result (Fig. 5), we see that the plasmon frequency, which is in the resonance with the bulk dipole transition frequency, is easily achievable at $d = 10$ nm than at $d = 20$ nm.

It is possible that electron-surface phonons scattering may provide additional channel for the decay of surface plasmons in Bi_2Se_3. The time and angle resolved photoemission spectroscopy (TrARPES) measurement[40] for the intrinsic cooling behavior of Dirac fermions in Bi_2Se_3 shows that at temperatures above the Debye temperature $\Theta_D = 180$ K, electrons are scattered to the surface from the bulk by phonons due to an elevated electron-phonon scattering rate. In another experiment[48], a hallmark of the low energy surface acoustic phonon in a slab of Bi_2Se_3 has been presented with a maximum phonon frequency of nearly 4 THz (≈ 3 meV). These observation starkly avoid any interaction of the phonon and plasmon modes at low temperatures in the energy range that we are interested at for the spasing.

IV. CONCLUSION

We theoretically study the nanospasing based on 3D TI material, bulk part of which works as a gain medium and surface supports the collective oscillation of the Dirac fermions in the mid-infrared regime. The 3D TI materials are small bandgap semiconducting materials with surface states occurring inside the bulk bandgap and such surface states are robust against back-scattering by non-magnetic impurities. In addition to that, spin-momentum locking property of surface states provide number of possibilities including a channel for spin accumulation and spin-plasmon excitation. A nanopatch of thickness d with electrostatically interacting top and bottom surfaces can support antisymmetric and symmetric modes of plasmons. Under RPA theory, with a double layer model, symmetric modes are shown to be pinned to the Landau damped region at a thickness of $d = 10$ nm. We show that it is possible to obtain spasing, generation of a coherent source of surface plasmons, with the plasmon energy around 0.3 eV at which the resonance with the dipole transition in bulk is possible. In a 3D TI nanopatch, we obtain the minimum plasmon relaxation rate for spasing to be less than a picosecond, the same order of magnitude for plasmon decay rate obtained in the experiment[47], with the plasmon energy of 0.3 eV. A minimum quality factor of $Q_{min} = 3$ is obtained at a plasmon wavelength of $\lambda_p = 2d$, which should be achievable in the experiments even in the strong damping limit. Our results provide a nanoscale model of spaser ever proposed with a single material that supports plasmons and at the same time also works as a gain medium.

V. DISCUSSION

A nanospaser using a sheet of graphene was proposed previously in the Ref. 28 as a coherent quantum generator of surface plasmons that utilizes the radiative emission of exciton recombinations in cascaded quantum wells. The present manuscript has two fold differences from the proposal in the Ref. 28: First, present nanospaser utilizes the bulk of the material (Bi_2Se_3) which also supports the surface plasmon modes and the carriers in the bulk are optically pumped, whereas the nanospaser proposed in the Ref. 28 utilizes quantum wells as a gain medium and carriers in there are electrically controlled. The present model uses single material for both the generation of surface plasmons and the source of feedback rather than using seperate element as a gain medium. This provides effectively couple the surface plasmon modes with the exciton recombinations. Therefore, this unique model of 3D TI based nanospaser helps to realize truly a nanospaser. Theoretically and also experimentally, graphene and 3D TIs consist of several fundamental differences. Second: the proposal in Ref. 28 provides a qualitative description for the plasmon modes and their interaction with the quantum wells. The response function has a simple

description obtained from the Drude model. Here our proposal consists of an optically controlled nanospaser based on the 3D TI with a full analytically description of plasmons modes obtained under RPA on the line of Ref. 51, and presents details about the interaction of the plasmonic field with the electronic transitions in the bulk, strength of which, as measured by dipole matrix elements, are calculated using DFT. The working wavelength regime of the nanospaser here depends on the bulk bandgap (0.3 eV) of the material which also supports surface plasmons. It should be understood that the same set of fundamental equations can always be used to describe the light-matter interactions wherever necessary and the condition for the spasing is always described by the same universal equation.

The scheme of the nanospaser we propose here is truly a nanoscopic in nature which functions in the mid infrared regime. Nevertheless, questions may arise on the mode and the performance of the spasing with the 3D TIs due to possibility of affects coming from such as carrier transfer processes, impurity bands and excitons recombination kinetics. Here we discuss these affects in more details.

The major carrier transfer processes from the bulk to the surface may occur due to diffusion and phonon assisted scattering events. It is possible that electron-surface phonons scattering may provide additional channel for the decay of surface plasmons in Bi_2Se_3. The time and angle resolved photoemission spectroscopy (TrARPES) measurement[40] for the intrinsic cooling behavior of Dirac fermions in Bi_2Se_3 shows that at temperatures above the Debye temperature $\Theta_D = 180$ K, electrons are scattered to the surface from the bulk by phonons due to an elevated electron-phonon scattering rate. In another experiment[48], a hallmark of the low energy surface acoustic phonon in a slab of Bi_2Se_3 has been presented with a maximum phonon frequency of nearly 4 THz (≈ 3 meV). These observation starkly avoid any interaction of the phonon and plasmon modes at low temperatures in the energy range that we are interested at for the spasing.

Complications to achieve spasing in Bismuth based 3D TIs which are alloys may arise due to impurity bands lying within the bandgap that may affect the electron-hole recombination kinetics. The alloys such as Bi_xSb_{1-x} contain random substitution disorder, their dispersion branches are extremely difficult to describe, and only be possible to define their electronic structures within the mean field approximation. They tend to have impurity bands inside the nominal bulk energy gap, which sometimes overlap with the surface states[52]. However, Bismuth based compound 3D TIs such as Bi_2Se_3 and Bi_2Te_3 have fairly simple bandstructures and can be grown to be extremely clean in the experiments[53]. Nevertheless, there are impurity bands which are situated below the conduction band edge and are conductive. However, their density is small as compared to the conduction band carrier density. The mobility in the impurity bands, for example, is 10 cm^2/Vs which is two order of magnitude smaller than the Hall mobility in the conduction band (~ 1000 cm^2/Vs)[54,55]. Therefore, the role of impurities on the optical excitation can be safely disregarded.

Finally, the surfaces states have few nm penetration length of around 2-3 QL. Electrons and holes can hop to surface and fill single-particle surface states. As shown by Bansal et al[45] and have also mentioned above, beneath those states, there are 2D electron-gas states extending up to 8 QL due to band bending effect and then followed by the bulks states. Theoretically, it is always possible to separate out the 2D electron states and the bulk contribution from the surface states. The problem arising due to possibile filling of single-particle surface states as a consequence of diffusion of the carriers from the bulk to the surface can be circumvent in the experiment by pumping carriers in the bulk with a laser that transfers small kinetic energy to the carriers in the bands so that there is a negligible probability of diffusing to the surface leaving the plasmon dispersion for Dirac particles unaffected.

VI. CONCLUSION

We theoretically present the nanospasing using the 3D TI material, bulk part of which works as a gain medium and surface supports the collective oscillation of Dirac fermions in the mid-infrared regime. The 3D TI materials are small bandgap semiconducting materials with surface states occurring inside the bulk bandgap and such surface states are robust against back-scattering by non-magnetic impurities. In addition to that, spin-momentum locking property of surface states provide number of possibilities including a channel for spin accumulation and spin-plasmon excitation. A nanopatch of thickness d with electrostatically interacting top and bottom surfaces can support antisymmetric and symmetric modes of plasmons. Under RPA theory, with a double layer model, symmetric modes are shown to be pinned to the Landau damped region at a thickness of $d = 10$ nm. We show that it is possible to obtain spasing, generation of a coherent source of surface plasmons, with the plasmon energy around 0.3 eV at which the resonance with the dipole transition in bulk is possible. In a 3D TI nanopatch, we obtain the minimum plasmon relaxation rate for spasing to be less than a picosecond, the same order of magnitude for plasmon decay rate obtained in the experiment[47], with the plasmon energy of 0.3 eV. A minimum quality factor of $Q_{min} = 3$ is obtained at a plasmon wavelength of $\lambda_p = 2d$, which should be achievable in the experiments even in the strong damping limit. Our results provide a nanoscale model of spaser ever proposed with a single material that supports plasmons and at the same time also works as a gain medium.

Acknowledgments

Appendix

We consider TM waves with $E_y = 0$, $H_x = 0$ and $H_z = 0$. Therefore, the solution of the Maxwell's equations for the electric fields in regions 1, 2 and 3, shown in Fig. 2, is $E_x^1(z) = A_1 e^{\alpha_1(z+d/2)} e^{iqx}$, $E_x^2(z) = (A_2 \sin(\alpha_2 z) + B_2 \cos(\alpha_2 z)) e^{iqx}$, and $E_x^3(z) = B_3 e^{-\alpha_1(z-d/2)} e^{iqx}$, respectively. The z-component of the electric field and y-component of the magnetic field are obtained using the relations $E_z^i = -\frac{iq}{\alpha_i^2} \frac{dE_x^i}{dx}$ and $H_y^i = \frac{\varepsilon^i E_z^i}{qc}$. The continuity of the tangential component of the electric field at the interfaces $z = \pm d/2$ (Fig. 2 **a**) requires that $E_x^i(z = \pm a/2) = E^{i+1}(z = \pm a/2)$. Similarly, the discontinuity of the normal component of the electric field at $z = \pm a/2$ requires that $\varepsilon_i E_z^i(z = \pm a/2) = \varepsilon_{i+1} E_z^{i+1}(z = \pm a/2)$. Using these relations, we solve for the field amplitudes A_2 and B_2

$$A_2 = \frac{B_3 - A_1}{2 \sinh\left(\frac{\alpha_2 d}{2}\right)} \qquad (26)$$

and

$$B_2 = \frac{A_1 + B_3}{2 \cosh\left(\frac{\alpha_2 d}{2}\right)}. \qquad (27)$$

Similarly, another set of A_2 and B_2 can be obtained using the discontinuity of E_z.

The frequency dependent conductivities σ_1 and σ_2 on the top and bottom surfaces create the discontinuity of the tangential components of the magnetic field. The Maxwell's equation takes the form

$$\frac{\varepsilon_i}{\alpha_i^2} \frac{\partial^2 E_x^i}{\partial z^2} - \varepsilon_i E_x^i = \frac{4\pi i}{\omega} \sum_{i=1,2} \sigma_i E_x^i \delta(z - z_i), \qquad (28)$$

where z_1 and z_2 are the positions of the top and bottom surfaces, respectively.

Integrating Eq. (28) around $z_i = \pm d/2$ and writing the resulting equations in terms of the field amplitudes A_1 and B_3, we obtain the following equations

$$\left(\frac{\varepsilon_1}{\alpha_1} + \frac{4\pi i \sigma_1}{\omega} + \frac{\varepsilon_2}{\alpha_2} \tanh\left(\frac{\alpha_2 d}{2}\right)\right) A_1$$
$$+ \left(\frac{\varepsilon_3}{\alpha_3} + \frac{4\pi i \sigma_2}{\omega} + \frac{\varepsilon_2}{\alpha_2} \tanh\left(\frac{\alpha_2 d}{2}\right)\right) B_3 = 0, \qquad (29)$$

$$- \left(\frac{\varepsilon_1}{\alpha_1} + \frac{4\pi i \sigma_1}{\omega} + \frac{\varepsilon_2}{\alpha_2} \coth\left(\frac{\alpha_2 d}{2}\right)\right) A_1$$
$$+ \left(\frac{\varepsilon_3}{\alpha_3} + \frac{4\pi i \sigma_2}{\omega} + \frac{\varepsilon_2}{\alpha_2} \cosh\left(\frac{\alpha_2 d}{2}\right)\right) B_3 = 0. \qquad (30)$$

Above equations can be easily cast in the determinantial form as shown in Eq. (13). We obtain two solutions for B_3 in terms of A_1 corresponding to two different modes (symmetric and antisymmetric)

$$B_3^{sym} = -\frac{\frac{\varepsilon_1}{\alpha_1} + \frac{4\pi i \sigma_1}{\omega} + \frac{\varepsilon_2}{\alpha_2} \tanh\left(\frac{\alpha_2 d}{2}\right)}{\frac{\varepsilon_3}{\alpha_3} + \frac{4\pi i \sigma_2}{\omega} + \frac{\varepsilon_2}{\alpha_2} \tanh\left(\frac{\alpha_2 d}{2}\right)} A_1 \qquad (31)$$

and

$$B_3^{ant} = \frac{\frac{\varepsilon_1}{\alpha_1} + \frac{4\pi i \sigma_1}{\omega} + \frac{\varepsilon_2}{\alpha_2} \coth\left(\frac{\alpha_2 d}{2}\right)}{\frac{\varepsilon_3}{\alpha_3} + \frac{4\pi i \sigma_2}{\omega} + \frac{\varepsilon_2}{\alpha_2} \coth\left(\frac{\alpha_2 d}{2}\right)} A_2. \qquad (32)$$

Using Eqs. (26), (27), (31), and (32), for each mode, we can express B_2, A_2, and B_3 in terms of A_1 as $A_2 = C_1 A_1$, $B_2 = C_2 A_1$, and $B_3 = C_3 A_1$, where C_1, C_2, and C_3 are some constants.

* Electronic address: vapalkov@gsu.edu
† Electronic address: mstockman@gsu.edu

[1] M. Moskovits, Rev. Mod. Phys., **57**, 783 (1985).

[2] M. I. Stockman, V. M. Shalaev, M. Moskovits, R. Botet, and T. F. George, Phys. Rev. B **46**, 2821 (1992).

[3] K. Kneipp, M. Moskovits, and H. Kneipp, eds., Electromagnetic Theory of SERS, vol. 103 (Springer, Heidelberg, 2006).

[4] L. R. Hirsch, R. J. Stafford, J. A. Bankson, S. R. Sershen, B. Rivera, R. E. Price, J. D. Hazle, N. J. Halas, and J. L. West, Proc. Natl. Acad. Sci. USA **100**, 13549 (2003).

[5] L. Novotny and B. Hecht, Principles of Nano-Optics (Cambridge University Press, Cambridge, New York, 2006).

[6] A. Israel, M. Mrejen, Y. Lovsky, M. Polhan, S. Maier, and A. Lewis, Laser Focus World **43**, 99 (2007).

[7] W. A. Challener, C. Peng, A. V. Itagi, D. Karns, W. Peng, Y. Peng, X. Yang, X. Zhu, N. J. Gokemeijer, Y. T. Hsia, G. Ju, R. E. Rottmayer, M. A. Seigler, and E. C. Gage, Nat. Phot. **3**, 220 (2009).

[8] S. Kim, J. H. Jin, Y. J. Kim, I. Y. Park, Y. Kim, and S. W. Kim, Nature **453**, 757 (2008).

[9] . I.-Y. Park, S. Kim, J. Choi, D.-H. Lee, Y.-J. Kim, M. F. Kling, M. I. Stockman, and S.-W. Kim, Nat. Phot. **5**, 677 (2011).

[10] H. P. Paudel and M. N. Leuenberger, Nano Lett. **12**, 2690 (2012).

[11] M. I. Stockman, J. Opt. **12**, 024004 (2010).

[12] D. Bergman and M. I. Stockman, Phys. Rev. Lett. **90**, 027402 (2003).

[13] M. I. Stockman, Nat. Photonics **2**, 327 (2008).

[14] [5] M. I. Stockman, Opt. Express **19**, 22029 (2011).

[15] M. A. Noginov, G. Zhu, A. M. Belgrave, R. Bakker, V.M. Shalaev, E. E. Narimanov, S. Stout, E. Herz, T. Suteewong, and U. Wiesner, Nature **460**, 1110 (2009).

[16] J.Y. Suh, C. H. Kim, W. Zhou, M. D. Huntington, D. T. Co, M. R. Wasielewski, and T.W. Odom, Nano Lett. **12**, 5769 (2012).

[17] J. K. Kitur, G. Zhu, A. B. Yu, and M. A. Noginov, J. Opt. **14**, 114015 (2012).

[18] X. Meng, J. Liu, A. V. Kildishev, and V. Shalaev, Laser Photonics Rev. **8**, 6, 896 (2014).

[19] N. I. Zheludev, S. L. Prosvirnin, N. Papasimakis, and V. A. Fedotov, Nat. Photonics **2**, 351 (2008).

[20] E. Plum, V. A. Fedotov, P. Kuo, D. P. Tsai, and N. I. Zheludev, Opt. Express **17**, 8548 (2009).

[21] R. A. Flynn, C. S. Kim, I. Vurgaftman, M. Kim, J. R. Meyer, A. J. Makinen, K. Bussmann, L. Cheng, F. S. Choa, and J. P. Long, Opt. Express **19**, 8954 (2011).

[22] R. F. Oulton, V. J. Sorger, T. Zentgraf, R.-M. Ma, C. Gladden, L. Dai, G. Bartal, and X. Zhang, Nature **461**, 629 (2009).

[23] R.-M. Ma, R. F. Oulton, V. J. Sorger, G. Bartal, and X. Zhang, Nat. Mater. **10**, 110 (2010).

[24] M. J. H. Marell, B. Smalbrugge, E. J. Geluk, P. J. van Veldhoven, B. Barcones, B. Koopmans, R. Notzel, M. K. Smit, and M. T. Hill, Opt. Express **19**, 15109 (2011).

[25] M. I. Stockman, Phys. Rev. Lett. **106**, 156802 (2011).

[26] E. S. Andrianov, A. A. Pukhov, A. V. Dorofeenko, A. P. Vinogradov, and A. A. Lisyansky, Opt. Lett. **36**, 4302 (2011).

[27] S. D. Campbell and R. W. Ziolkowski, Opt. Commun. **285**, 3341 (2012).

[28] V. Apalkov and M. I. Stockman, Light: Science and Applications **3**, 191 (2014).

[29] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).

[30] F. H. Koppens, D. E. Chang, F. J. de Abajo, Nano Lett. **11**, 3370 (2011).

[31] J. Chen, M. Badioli, P. Alonso-Gonzalez, S. Thongrattanasiri, F. Huth, J. Osmond, M. Spasenovic, A. Centeno, A. Pesquera, P. Godignon, A. Z. Elorza, N. Camara, F. J. Garcia de Abajo, R. Hillenbrand, and F. H. Koppens, Nature **487**, 77 (2012).

[32] 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. Castro Neto, C. N. Lau, F. Keilmann, and D. N. Basov, Nature (London) **487**, 82 (2012).

[33] H. P. Paudel and M. N. Leuenberger, J. Phys.: Condens. Matter **26**, 082201 (2014).

[34] D. Hsieh, Y. Xia, D. Qian, L. Wray, J. H. Dil, F. Meier, J. Osterwalder, L. Patthey, J. G. Checkelsky, N. P. Ong, A. V. Fedorov, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature **460**, 1101 (2009).

[35] M. Hajlaoui, E. Papalazarou, J. Mauchain, G. Lantz, N. Moisan, D. Boschetto, Z. Jiang, I. Miotkowski, Y. P. Chen, A. Taleb-Ibrahimi, L. Perfetti, and M. Marsi, Nano Lett. **12**, 3532 (2012).

[36] H. Zhang , C. X. Liu, X. L. Qi, X. Dai, Z Fang, and S.-C. Zhang, Nature Phys. **5**, 438 (2009).

[37] Y. Xia, D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature Phys. **5**, 398, (2009).

[38] S. Raghu, S. B. Chung, X-L. Qi, and S-C. Zhang, Phys. Rev. Lett. **104**, 116401 (2010).

[39] D. Kim, Q. Li, P. Syers, N. P. Butch, J. Paglione, S. D. Sarma, and M. S. Fuhrer, Phys. Rev. Lett. **109**, 166801 (2012).

[40] Y. H. Wang, D. Hsieh, E. J. Sie, H. Steinberg, D. R. Gardner, Y. S. Lee, P. Jarillo-Herrero, and N. Gedik, Phys. Rev. Lett. **109**, 127401 (2012).

[41] R. E. V. Profumo, M. Polini, R. Asgari, R. Fazio, and A. H. MacDonald, Phys. Rev. B **82**, 085443 (2010).

[42] Bo Zhou, Z. K. Liu, G. Analytis, K. Igarashi, S. K. Mo, D. H. Lu, R. G. Moore, I. R. Fisher, T. Sasagawa, Z. X. Shen, Z. Hussain, and Y. L. Chen, Semicond. Sci. Tech. **27**, 124002 (2012).

[43] D. Hsieh, F. Mahmood, J. W. McIver, D. R. Gardner, Y. S. Lee, and N. Gedik, Phys. Rev. Lett. **107**, 077401 (2011).

[44] R. E. V. Profumo, M. Polini, R. Asgari, R. Fazio, and A. H. MacDonald, Phys. Rev. B **85**, 085443 (2012).

[45] N. Bansal, Y. S. Kim, M. Brahlek, E. Edrey, and S. Oh, Phys. Rev. Lett. **109**, 116804 (2012).

[46] G. Kresse and J. Furthmuller, Phys. Rev. B **54**, 11169 (1996).

[47] P. Di Pietro, M. Ortolani, O. Limaj, A. Di Gaspare, V. Giliberti, F. Giorgianni, M. Brahlek, N. Bansal, N. Koirala, S. Oh, P. Calvani, and S. Lupi, Nature Nanotechnology **8**, 556 (2013).

[48] X. Zhu, L. Santos, R. Sankar, S. Chikara, C. Howard, F. C. Chou, C. Chamon, and M. El-Batanouny, Phys. Rev. Lett. **107**, 186102 (2011).

[49] J. A. Sobota, S. Yang, J. G. Analytis, Y. L. Chen, I. R. Fisher, P. S. Kirchmann, and Z.-X. Shen, Phys. Rev. Lett. **108**, 117403 (2012).

[50] I. A. Nechaev, R. C. Hatch, M. Bianchi, D. Guan, C. Friedrich, I. Aguilera, J. L. Mi, B. B. Iversen, S. Blugel, Ph. Hofmann, and E. V. Chulkov, Phys, Rev. B **87**, 121111(R) (2013).

[51] T. Stauber and G. Gomez-Santos, Phys. Rev. B **88**, 205427 (2013).

[52] H. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Zh. Fang, and S.-C. Zhang, Nature Physics **5**, 438 (2009).

[53] M. Neupane, A. Richardella, J. Sanchez-Barriga, S.-Y. Xu, N. Alidoust, I. Belopolski, Ch. Liu, G. Bian, D. Zhang, D. Marchenko, A. Varykhalov, O. Rader, M. Leandersson, T. Balasubramanian, T.-R. Chang, H.-T. Jeng, S. Basak, H. Lin, A. Bansil, N. Samarth, and M. Zahid Hasan, Nature Comm. **5**, 3841 (2014).

[54] J. G. Checkelsky, Y. S. Hor, R. J. Cava, and N. P. Ong, Phys. Rev. Lett. **106**, 196801 (2011).

[55] D. Kim, S. Cho, N. P. Butch, P. Syers, K. Kirshenbaum, S. Adam, J. Paglione, and M. S. Fuhrer, Nature Physics **8**, 459 (2012).