Active Valley-topological Plasmonic Crystal in Metagate-tuned Graphene

Minwoo Jung,1,* Zhiyuan Fan,2 and Gennady Shvets2,†

1 Department of Physics, Cornell University, Ithaca, New York, 14853, USA
2 School of Applied and Engineering Physics, Cornell University, Ithaca, New York 14853, USA

(Dated: April 3, 2018)

A valley plasmonic crystal for graphene surface plasmons (GSPs) is proposed. We demonstrate that a designer metagate, placed within a few nanometers from graphene, can be used to impose a triangular periodic Fermi energy landscape on the latter. For specific metagate geometries and bias voltages, the combined metagate-graphene structure is shown to produce sufficiently strong Bragg scattering of GSPs to produce complete propagation bandgaps, and to impart the GSPs with nontrivial valley-linked topological properties. Valley-selective kink states supported by a domain wall between differently patterned metagates are shown to propagate without reflections along sharply curved interfaces owing to suppressed inter-valley scattering. Our approach paves the way for non-magnetic dynamically reconfigurable topological nanophotonic devices.

Graphene is a promising platform for nanoscale terahertz photonic devices because it supports deeply subwavelength graphene surface plasmons (GSPs) [1]. The propagation of GSPs can be dynamically controlled by injecting free carriers (electrons or holes) using field-effect electrostatic gating [2–4]. Moreover, the electron scattering times in high-quality graphene have been extended to nearly a picosecond, thus enabling extremely long propagation distances of the GSPs [5–8]. Recently reported demonstration of electrical control of the GSP’s phase [9] suggests that graphene-based nanophotonic devices can be now envisioned. The natural next step in developing such devices is to investigate the possibility of topological protection of GSPs, similarly to the way it has been done in microwave and micron-scale photonic structures [10–15].

Several recent proposals [16, 17] utilize large magnetic fields to produce topological GSPs by breaking the time-reversal symmetry in nanopatterned graphene. Among the practical limitations of this approach are the lack of reconfigurability (large magnetic fields cannot be rapidly turned off) and the degradation of graphene’s quality arising from its patterning [17, 18]. In this Letter, we propose an alternative approach that requires neither graphene patterning nor magnetic field. Instead, we demonstrate that a regular landscape of chemical potential can be “imprinted” into graphene by placing it in close proximity of an electrically-biased patterned metagate shown in Fig. 1(a). Because graphene’s optical conductance is related to its chemical potential, such landscapes present periodic patterns of the effective refractive index to the GSPs, causing their Bragg scattering and producing a complex propagation band structure. When the metagate pattern is a two-dimensional (2D) triangular crystal with broken mirror symmetry, topological properties linked to the valley degree of freedom emerge in GSPs.

Historically, such valley crystals (VC) were originally discovered in electronic systems and studied in the context of valley-polarized non-zero Hall conductivities [19–23], giving rise to the emerging field of valleytronics [24, 25]. More recently, the study of valley-based topological phases was extended to bosonic systems, in which the particular interest is the emergence of confined chiral kink states (CKSs) at the domain wall between two VCs that are mirror images of each other. The propagation of the CKSs is valley-locked and topologically protected in the absence of inter-valley scattering [26, 27]. Such topologically robust transport of valley-locked CKSs has been extensively studied in several bosonic—photonic/plasmonic [28–33] and phononic [34, 35]—systems as well as in electronic systems [36–41].

In this Letter, we show that a monolayer graphene electrostatically doped by a metagate provides a nanoscale plasmonic platform for mid-infrared valleytronics. The specific metagate design shown in Fig. 1(a) consists of a half-infinite perfectly conducting metal penetrated by a triangular lattice of infinitely deep equilateral triangular holes (corrections due to finite resolution in patterning and finite depths of the holes are discussed in Supplemental Material [42]). When a static voltage $V = V_{MG} - V_{gr}$ is applied between the metagate (at an electric potential $V_{MG}$) and graphene (at an electrochemical potential $V_{gr}$): $-eV_{gr} = E_F(r) - e\phi_{gr}(r)$ [43], the latter is electrostatically doped by free electrons with a spatially-dependent surface density $n(r)$, where $E_F(r) = \hbar v_F \sqrt{\pi n(r)} (v_F \sim 10^6 \text{m/s} \text{ is the Fermi velocity})$ is the chemical potential. Here $r = (x, y)$ and the graphene sheet is located at $z = 0$. The $v_F$-renormalization from electron-electron interactions [44, 45] is ignored for the discussions below because its influence on the GSP dispersion was calculated to be negligible, see Supplemental Material [42]. An example of the chemical potential landscape $E_F(r)$ in graphene is shown in Fig. 1(b) for the geometric parameters of the metagate listed in the caption. Clearly, the metagate imprints a VC onto graphene with the spatial symmetry ($C_3$) required for realizing a valley-Hall photonic topological insulator [28].

The static electron density $n(r)$ is determined from $n(r) = (\epsilon_0/e) [\partial_z \phi(r, z)]_{z=-0} - [\partial_z \phi(r, z)]_{z=+0}$, where
the static potential \( \phi(r, z) \) is a solution to the Laplace equation \( \nabla^2 \phi = 0 \) with self-consistent boundary conditions: \( \phi(r, z = 0) = \phi_{gr}(r) = V_{gr} + \frac{\hbar v_F}{c} \sqrt{\pi n(r)} \) and \( \phi(MG) = V_{MG} \). Due to graphene quantum capacitance (GQC) [46], the boundary conditions themselves are dependent on the final solution \( n(r) \), thus requiring an iterative method to numerically solve the Laplace problem (see Supplemental Materials [42]). The GQC effect can be neglected in the limit of \( \rho_1 \equiv l_s/h \ll 1 \) (equivalently, \( E_F(r) \ll e|V| \)) [43, 44], where \( l_s = \frac{\pi \hbar^2 e^2}{c^2 E_{F_0}} \) is the graphene static screening length [47] (\( \epsilon_s \) is the static permittivity of the spacer material, and \( E_{F_0} = (E_F(r)) \) is the averaged value over a unit cell). We note that the normalized screening length scale \( \rho_1 \) determines the velocity of the acoustic GSPs \( v_{ac} \equiv \omega_{ac}(q)/|q| \) according to \( v_{ac}/v_F = (1 + \rho_1/2)/(\rho_1 + r_F^2/4)^{1/2} \geq 1 \) [8, 45]. With GQC ignored, graphene in the static limit can be treated as a classical conducting sheet \( \phi_{gr}(r) = V_{gr} \), and the solution \( n(r) \) of the Laplace equation (subject to \( n(r) \)-dependent boundary conditions) is scale-invariant, i.e., the normalized chemical potential profile \( \theta(r/L) \equiv E_F(r/L)/\sqrt{V/h} \) depends only on the geometric proportions (e.g., \( L : r : h \)) of the structure.

The band dispersion of GSPs is obtained from the coupled integral equations for the optical perturbations of the potential \( \delta \phi \) and charge density \( \delta n \) [48]:

\[
\delta \phi(q, \omega) = \sum_{q'} \tilde{\nu}(q, q', \omega) \delta n(q', \omega) \tag{1a}
\]

\[
\delta n(q, \omega) = \sum_{q'} \tilde{\chi}_{nn}(q, q', \omega) \delta \phi(q', \omega) \tag{1b}
\]

where the quantities under a tilde are calculated in the momentum/frequency Fourier domains. Equation (1a) represents the Poisson’s equation for 2D charge distributions, where \( \tilde{\nu} \) describes the 2D screened Coulomb interaction in the presence of surrounding dielectrics and/or metal metagates. While \( \tilde{\nu} \propto \delta q, q' \) for an unpatterned gate, the non-trivial \((q, q')\)-dependence of \( \tilde{\nu} \) arises as the periodic nature of the metagate induces a periodic array of image charges. Thus, a density perturbation with the wavevector \( q \) induces harmonic overtones of the potential with a set of wavevectors \( q' = q + \{G\} \) (\( \{G\} \) is the set of the reciprocal lattice vectors). An exact analytic form of \( \tilde{\nu}(q, q', \omega) \) in the presence of a periodic metagate is derived in the Supplemental Material [42]. Equation (1b) expresses the charge continuity, and contains the density response function \( \tilde{\chi}_{nn} \) that is related to the non-local conductivity \( \tilde{\sigma} \) according to \( \tilde{\chi}_{nn}(q, q', \omega) = \frac{\delta q}{\omega c} \tilde{\sigma}(q, q', \omega) \) (see Supplemental Material [42] for a more detailed derivation). For the simplest example of graphene suspended in vacuum (no metagate) and doped homogeneously \( (E_F(r) = E_{F_0}) \), we have \( \tilde{\nu}(\omega) = \delta q, q' \frac{c^2}{2\epsilon_0|q|} \) and \( \tilde{\chi}_{nn} = \delta q, q' \frac{E_{F_0} q, q'}{\pi \hbar^2 c} \) in local Drude approximation [49]. Substituting these expressions into Eqs. (1)
Therefore, even though Eqs. (local) Drude conductivity, they cannot account for (ii).

When GQC and NLER are negligible in \( \rho_1, \rho_2 \ll 1 \) limits, a scale-invariant \( E_F(r) \) and a local Drude conductivity \( \sigma (r, r', \omega) = \frac{2 e^2 E_F(r) (r - r')}{\pi \hbar^2 \omega} \) (equivalently, \( \chi_{\text{un}}(q, q', \omega) = \frac{E_F(q) (q' - q)}{\pi \hbar^2 \omega} \)) can be assumed. Then, if the permittivity dispersion of the medium encapsulating graphene is negligible as well, Equations (1) can be recast as a linear eigenvalue matrix problem, see Supplemental Material [42], and the resulting plasmonic dispersions are also scale-invariant with the natural frequency scale given by \( \omega_0 = \omega_{\text{amb}}(q) = \frac{2 \pi}{L} \). With \( L = 200 \text{ nm} \) and \( E_F = 0.2 \text{ eV} \), for example, we get \( \omega_0 / 2 \pi = 33 \text{ THz} \sim 0.02 \pi / L \), which confirms that our model lies well in the electrostatic limit. A similar model was used [54] in the absence of a patterned metagate.

On the other hand, if GQC, NLER or the permittivity dispersion of the spacer medium cannot be ignored, the plasmonic dispersions are no longer scalable as the operator of the eigenvalue problem itself acquires a frequency dependence. For simplicity in demonstrating generic concepts of GSP-based valley-topological transport, we start by working in a vacuum and neglecting those non-scalable effects, which is actually valid for experimentally reasonable system sizes and bias field gradients \( (L > 100 \text{ nm} \text{ and } V / h > 0.5 \text{ V/nm}, \text{see Supplemental Material [42]}) \). The example corresponding to significant GQC and NLER corrections in the presence of dispersive graphene-encapsulating materials \( (\text{e.g., hBN}) \) is discussed later. We neglect the hybridization of the metagte-supported spoof surface plasmons \( (\text{SSP}) [55] \) with GSPs because the coupling between SSPs and GSPs is negligible [56] in the strongly-electrostatic limit.

Before examining the effect of \( E_F(r) \) landscaping, we first consider a situation where GSPs propagate over the metagte and graphene homogeneously doped by other means \( (\text{i.e. } V = 0 \text{ in Fig. 1}(a), \text{but } E_F(r) = E_F(0) \neq 0) \). A propagation bandgap opens at \( \Theta \) point of the Brillouin zone \( (\text{BZ}) \) due to the broken mirror symmetry in the metagte structure as observed in Figure. 1(d). The overall effect of the metagte screening and the emergence of the acoustic GSP [8, 57] is also apparent from the linear \( \omega = v_{\text{ac}}|q| \) dispersion near \( \Theta \)-point. The periodic screening from the metagte itself is, however, insufficient for opening a complete bandgap despite its proximity \( (L / h = 25) \) to graphene. The complete bandgap opens over the whole \( \text{BZ} \), as shown in Fig. 1(e), only when a periodic \( E_F(r) \) landscape (Fig. 1(b)) is introduced by biasing \( (V \neq 0) \) metagte respect to graphene.

The magnitude of bandgap depends on the orientation angle \( \theta \) of the triangular holes defined in Fig. 1(a). For \( \theta \equiv 0 \text{ (mod} 60^\circ) \), the system is mirror-symmetric with respect to \( \Theta \) and \( \Theta' \) directions. Therefore, the spatial profiles of the two lowest energy eigenstates at each valley are mirror images of each other (see Fig. 2(a)), thus forming degenerate Dirac points [29]. In contrast, these two states are no longer degenerate (see Fig. 2(b)) when the mirror symmetry is broken for \( \theta \neq 0 \text{ (mod} 60^\circ) \), thereby producing valley-topological bandgaps.

Topological properties of our GSP-based VC are investigated first through the valley Chern numbers, \( C_{\nu}(L,U) \equiv (2 \pi i)^{-1} \int_{L,U} F(L,U)(k) d^2k \), evaluated numerically from the computed eigenstates \( \delta \delta_{\nu}(L,U) [58] \). Here, \( \nu \) is the valley index \( (\Theta \text{ or } \Theta') \), \( L \) or \( U \) corresponds to the lower \( (L) \) and upper \( (U) \) bands with respect to the
lateralized the dispersion relations of the CKSs along the x-direction of a 1D BZ (defined as $k_x \in (-\pi/L, \pi/L)$) aligned with the domain wall separating the two domains, each consisting of 20 unit cells. This is effectively a projection of the original 2D BZ (Fig. 1(c)) into $k_x$ axis, see Fig. 3(b). The two valleys of the 2D BZ correspond to $k_x^{K,K'} = \pm 2\pi/3L$ under this projection. As expected, the CKSs cross the bandgap near $k_x^{K}$ and $k_x^{K'}$ as shown in Fig. 3(c). The phase and group velocities of the CKSs are supported by the P(N)-type domain wall present in the same(opposite) direction(s) as if the waves are propagating in a medium with a positive(negative) refractive index. Figure. 3(d) depicts the confinement of CKSs at two types of domain walls.

Next, we demonstrate that valley-selective CKSs are immune to back-scattering along sharply-curved pathways. A phased array of point dipoles emitting at the mid-gap frequency $\omega_{\text{mid}}$ [42] was used to excite a CKS in the $K$-valley, and the non-radiative losses were modeled by assuming a finite electron scattering rate $\gamma = \omega_{\text{mid}}/300$. The robust topologically-protected propagation is observed in Figs. 4(a)-(c). The key condition for topological protection is the valley conservation [29, 32], which can be analytically proven for the 120° turns (Fig. 4(b)) using the $C^3$-symmetry of the system. On the other hand, valley conservation for the 60° turn (Fig. 4(c)) involves a more complicated mechanism because the domain walls changes between P- and N-type after the turn. At the 60° turn, $\phi_n^P$ state (at $K$) is transferred into $\phi_n^{K'}$ state, which belongs to the $K'$-valley with respect to the coordinate frame rotated by 60°, and, thus, effectively remains in the $K$-valley in the original coordinates. Figure. 4(d) confirms that the Drude loss is the only source of attenuation of the CKSs with or without a structural defect. Abrupt jumps in the attenuation curves at the turning points are numerical artifacts due to abrupt rotation of the integration box. Attenuation in the presence of structural defects appears to be even less than that in linear propagation, which results from overestimated propagation lengths since the CKSs don’t exactly follow the prescribed zigzag domain wall [60].

Finally, we show that the non-scalable effects from GQC and NLER become significant when a realistic heterostructure—graphene encapsulated between thin hBN layers—is considered. Accounting for NLER beyond the Drude approximation requires retaining the terms at least up to the fourth-order in $p_2$ in the Taylor expansion of $\tilde{\chi}_{nn}(\mathbf{q}, \mathbf{q}', \omega)$. For inhomogeneously doped graphene, the results obtained for the homogeneous case [62] were extended using the following substitutions[9, 50]: $|\mathbf{q}|^2 \rightarrow \mathbf{q} \cdot \mathbf{q}'$ and $E_F \delta_{\mathbf{q}, \mathbf{q}'} \rightarrow E_F (\mathbf{q} - \mathbf{q}')$ (see Supplemental Material [42] for more details). We have used the specific parameters (see caption to Fig. 5) of the metagate in order to locate the mid-gap frequency close to one of the principal wavelength ($\lambda = 10.5 \mu m$) of CO$_2$ lasers. First, we observe from Fig. 5(a,b) that the depth and the sharpness
of the chemical potential variation is reduced by the GQC effect. Second, according to Fig. 5(c), the inclusion of the GQC effect red-shifts the frequencies of the GSPs. This effect is attributed to the overall reduction of $E_F$ due to GQC-related charge screening, and the dependence of $\omega_{\text{GQC}}$ on $E_F$. Finally, the inclusion of the NLER results in a large (comparable with the total bandgap size) blue-shift of the topological GSPs, which can be interpreted as enhanced response due to velocity matching ($\rho_2 \rightarrow 1$) between GSPs and electrons [45].

In conclusion, we have described a GSP-based VC based on a deeply sub-wavelength periodic landscape of chemical potential produced by a metagater biased with respect to graphene. Highly suppressed intervalley scattering due to extremely sub-wavelength nature of GSPs enables robust topological valley transports—guiding plasmonic CKs along sharp curves under valley conservation. Our scheme requires neither magnetic fields nor physical patterning of graphene, and enables rapid reconfiguration by controlling the electric bias. Inherently quantum effects in graphene, such as quantum capacitance and non-local electron response, are shown to play an important role in realistic designs involving hBN encapsulation of graphene.

This work was supported by the Army Research Office (ARO) under a Grant No. W911NF-16-1-0319, and by the National Science Foundation (NSF) under the Grants No. DMR-1741788 and DMR-1719875. M.J. was also supported in part by Cornell Fellowship and in part by the Kwanjeong Fellowship from Kwanjeong Educational Foundation.

* mj397@cornell.edu
  gshvets@cornell.edu

[1] A. Woessner, M. B. Lundeberg, Y. Gao, A. Principi, P. Alonso-Gonzalez, M. Carrega, K. Watanabe, T. Taniguchi, G. Vignale, M. Polini, J. Hone, R. Hillenbrand, and F. H. L. Koppens, Nature Mater. 14, 421 (2015).
[2] D. K. Efetov and P. Kim, Phys. Rev. Lett. 105, 256805 (2010).
[3] H. Liu, Y. Liu, and D. Zhu, J. Mater. Chem. 21, 3335 (2011).
[4] Z. Fang, Y. Wang, A. E. Schlather, Z. Liu, P. M. Ajayan, F. J. G. de Abajo, P. Nordlander, X. Zhu, and N. J. Halas, Nano. Lett. 14, 299 (2014).
[5] X. Du, I. Skachko, A. Barker, and E. Y. Andrei, Nat. Nanotech. 3, 491 (2008).
[6] K. I. Bolotin, K. Sikes, Z. Jiang, M. Klima, G. Fudenberg, J. Hone, P. Kim, and H. L. Stormer, Solid State Commun. 146, 351 (2008).
[7] A. S. Mayorov, R. V. Gorbachev, S. V. Morozov, L. Britnell, R. Jalil, L. A. Ponomarenko, P. Blake, K. S. Novoselov, K. Watanabe, T. Taniguchi, and A. K. Geim, Nano Lett. 11, 2396 (2011).
[8] P. Alonso-Gonzalez, A. Y. Nikitin, Y. Gao, A. Woessner, M. B. Lundeberg, A. Principi, N. Forcellini, W. Yan, S. Velez, A. J. Huber, K. Watanabe, T. Taniguchi, F. Casanova, L. E. Hueso, M. Polini, J. Hone, F. H. L. Koppens, and R. Hillenbrand, Nat. Nanotech. 12, 31 (2017).
[9] A. Woessner, Y. Gao, I. Torre, M. B. Lundeberg, C. Tan, K. Watanabe, T. Taniguchi, R. Hillenbrand, J. Hone, M. Polini, and F. H. L. Koppens, Nat. Photonics 11, 421 (2017).
[10] Z. Wang, Y. Chong, J. D. Joannopoulos, and M. Soljacic, Nature 461, 772 (2009).
[11] M. Hafezi, E. A. Demler, M. D. Lukin, and J. M. Taylor, Nature Physics 7, 907 (2011).
[12] A. B. Khanikaev, S. H. Mousavi, W.-K. Tse, M. Kargarian, A. H. MacDonald, and G. Shvets, Nature Mater. 12, 233 (2013).
[13] M. C. Rechtsman, J. M. Zeuner, Y. Plotnik, Y. Lumer, D. Podolsky, F. Dreisow, S. Nolte, M. Segev, and A. Szameit, Nature 496, 196 (2013).
[14] W.-J. Chen, S.-J. Jiang, X.-D. Chen, B. Zhu, L. Zhou, J.-W. Dong, and C. T. Chan, Nature Comm. 5, 5782 (2014).
[15] T. Ma, A. B. Khanikaev, S. H. Mousavi, and G. Shvets, Phys. Rev. Lett. 114, 127401 (2015).
[16] D. Jin, T. Christensen, M. Soljacic, N. X. Fang, L. Lu, and X. Zhang, Phys. Rev. Lett. 118, 245301 (2017).
[17] D. Pan, R. Yu, H. Xu, and F. J. G. de Abajo, Nat. Commun. 8, 1243 (2017).
[18] W. Wang, T. Christensen, A.-P. Jauho, K. S. Thygesen, M. Wubs, and N. A. Mortensen, Sci. Rep. 5, 9535 (2015).
[19] D. Xia, X.-W. Yao, and Q. Niu, Phys. Rev. Lett. 99, 236809 (2007).
[20] E. McCann, K. Kechedzhi, V. I. Falko, H. Suzuura, T. Ando, and B. L. Altshuler, Phys. Rev. Lett. 97, 146805 (2006).
[21] A. Rycerz, J. Tworzydlo, and C. W. J. Beenakker,
1. NUMERICAL METHOD

We elaborate on the numerical method used in this Letter and implemented in an in-house code. As briefly described in the main text (Eqs. (1)), the method is essentially based on the linear response formulation [48] of plasmonic excitations. For a 2D Bloch state with Bloch wavevector $k$, the momentum integral becomes a summation over a countable set, $k + \{G\}$ ($\{G\}$ is the set of the reciprocal lattice vector). Then, Eqs. (1) in the main text are then casted into an eigenvalue problem on discrete matrices:

$$
\hat{V}_k \hat{\Phi}_k = \lambda_{nk} \Phi_{nk}, \text{ with}
$$

(S1a)

$$
[\hat{V}_k]_{\alpha \beta} = v_k (G_\alpha, G_\beta, \omega),
$$

(S1b)

$$
[\hat{X}_k]_{\alpha \beta} = \omega^2 \hat{\chi}_{nn} (k + G_\alpha, k + G_\beta, \omega),
$$

(S1c)

where $\alpha, \beta$ are countable indices, $G_\alpha, G_\beta \in \{G\}$, $\lambda_{nk} = \omega^2 n_k$, and $[\phi_{nk}]_\alpha = \phi_k (k + G_\alpha, \omega n_k)$. $v_k (q, q', \omega)$ is the lattice-summed effective Coulomb interaction. We outline the properties of the Coulomb operator $\hat{V}_k$ and the proper density-density response operator $\hat{X}_k$, which are introduced in Eqs. (S1), and discuss several computational details.

1.1. The Coulomb operator $\hat{V}_k$

As in Eq. (S1b), we define the matrix elements of the Coulomb operator to be $[\hat{V}_k]_{\alpha \beta} = v_k (G_\alpha, G_\beta, \omega)$, where $v_k (q, q', \omega)$ is the momentum-space representation of the lattice-summed effective Coulomb interaction, $G_\alpha, G_\beta \in \{G\}$ ($G$ is the reciprocal lattice vectors), and $\alpha, \beta$ are countable indices for $\{G\}$. The momentum-space representation of a linear response function, $f (q, q', \omega)$, is related to its real-space representation, $f (r, r', \omega)$, by double Fourier transformation:

$$
f (q, q', \omega) = \frac{1}{|\Omega|} \int d^d r \int d^d r' e^{-i q \cdot r} e^{i q' \cdot r'} f (r, r', \omega)
$$

(S2)

where $\Omega$ denotes the unit cell, and $|\Omega|$ is the unit cell area (if there is no periodicity in the system, $\Omega$ extends to the entire space $\mathbb{R}^d$, and the normalizing constant is replaced into $(2\pi)^{-d}$ with the system dimension $d$). The real space representation of the 2D lattice-summed effective Coulomb interaction is defined as:

$$
v_k (r, r', \omega) = \sum_{R} e^{-i k \cdot (r - r' - R)} v (r, r' + R, \omega)
$$

(S3)

where $R = n_1 R_1 + n_2 R_2$ denotes the direct lattice vectors ($n_1, n_2 \in \mathbb{Z}$), and $v (r, r', \omega)$ is the effective Coulomb interaction. The physical meaning of $v (r, r', \omega)$ is the screened coulomb potential energy at position $r$ generated by a source charge placed at position $r'$. In this context, the screening is from surrounding dielectrics and metals, not from other electrons in the plasma. In vacuum, we simply have $v (r, r', \omega) = e^2 / 4\pi \epsilon_0 |r - r'|$. Thus, the Coulomb operator in vacuum is calculated to be:

$$
[\hat{V}_k^{\text{vac}}]_{\alpha \beta} = v_k^{\text{vac}} (G_\alpha, G_\beta, \omega) = \frac{1}{|\Omega|^2} \int d^d r \int d^d r' e^{-i G_\alpha \cdot r} e^{i G_\beta \cdot r'} \sum_{R} e^{-i k \cdot (r - r' - R)} \frac{e^2}{4\pi \epsilon_0 |r - r' - R|} = \frac{e^2}{2\epsilon_0 |k - G|} \delta_{\alpha \beta}.
$$

(S4)

We used $\sum_{R} e^{-i k \cdot (r - R)} / |r - R| = \frac{2\pi}{|k - G|} \sum_{G} e^{-i G \cdot r} / |k - G| [63]$. If graphene is in a homogeneous dielectric, the Coulomb operator becomes dependent of frequency due to the dispersion $\epsilon_0 \rightarrow \epsilon (\omega)$. When the surrounding mediums are layered
and inhomogeneous as in our system design, Eq. (S2) and Eq. (S3) are not very useful for calculating $V_k$. Instead, we start from a Bloch ansatz of the plasma field, $\Phi_k(r, z) = x\hat{x} + y\hat{y}$, in an electrostatic limit ($\nabla^2 \Phi_k = 0$):

$$
\Phi_k(r, z) = \begin{cases} 
\sum_{\alpha} e^{ik_{\alpha} \cdot r} \left( c_{\alpha} \cosh(|k_\alpha| z) - s_{\alpha} \sinh(|k_\alpha| z) \right) & (z \geq 0) \\
\sum_{\alpha} e^{ik_{\alpha} \cdot r} \left( c_{\alpha} \cosh(|k_\alpha| z) + s_{\alpha} \sinh(|k_\alpha| z) \right) & (-h \leq z \leq 0)
\end{cases} \quad (k_{\alpha} \equiv k + G_{\alpha});
$$

(S5)

here, graphene is placed at $z = 0$, and MG is placed at $z = -h$. Then, following the vector representation of state vectors introduced in Eqs. (2), the vector elements of total (screened) potential on graphene $|\phi_k\rangle$ and induced charge density on graphene $|n_k\rangle$ are given as $|\phi_k\rangle|_\alpha = \xi c_{\alpha}$ and $|n_k\rangle|_\alpha = \int d|k_\alpha||s_{\alpha} + s'_{\alpha}|$. If we find relations between $c_{\alpha}$ and $s_{\alpha}, s'_{\alpha}$, by matching boundary conditions, we can derive the expression for $V_k$ from $|\phi_k\rangle = V_k|n_k\rangle$.

For this section, we assume that MG is infinitely deep (a few times $h$ suffices this assumption) and the permittivity of the metal is infinite. For an unstructured metal gate, we have simple boundary conditions: $\Phi_k(r, z = \infty) = \Phi_k(r, z = -h) = 0$. Then, we get $c_{\alpha} = s_{\alpha} = \tanh(|k_\alpha|h)s_{\alpha}$ and, thus, $[\hat{V}_k^{\text{flat}}]_{\alpha, \beta} = \frac{e^2}{\epsilon_0} \xi c_{\alpha} [\xi^2 + \cosh(|k_\alpha| h)].$ As pointed out in [8], the effective coulomb interaction does not diverge even at long-wavelength limit due to the screening from the metal gate: $\lim_{h \rightarrow 0} [\hat{V}_k|_\alpha, \alpha] = \frac{\epsilon^2}{\epsilon_0}$. Now, we turn into our design where the metal gate is periodically carved out with triangular cylindrical holes. We use the set of laplacian eigenmodes in a triangular cylinder, $T_\mu(r)$, [66] to describe the field in the MG domain ($z < -h$):

$$
\Phi_k(r, z < -h) = \sum_{R} e^{i k_{\alpha} \cdot r} \phi_{\text{hole}} (r - R, z), \quad \text{where} \quad \phi_{\text{hole}} (r, z) = \sum_{\mu} t_{\mu} T_\mu (r) e^{\xi_{\mu} (z+h)},
$$

(S6)

and $\mu$ is a countable index for the laplacian eigenmodes. $\xi_{\mu}$ is given to satisfy $\nabla^2_{2D} T_\mu (r) + \xi^2 T_\mu = 0$. $T_\mu (r)$ vanishes outside of the hole, and forms an eigenbasis inside the hole: $\int_S T_\mu (r) T_\nu (r) d\mathbf{r} = |S| \delta_{\mu\nu}$ ($S$ denotes the hole region, and $|S|$ is its section area). For compact description of boundary conditions, we use the vector representation for the linear coefficients: $|c\rangle \equiv \{c_{\alpha}\}, |s\rangle \equiv \{s_{\alpha}\}$, and $|t\rangle \equiv \{t_\mu\}$. Also, we take several groups of repeatedly used numbers in operator forms: $[\hat{\mathcal{C}}_k]_{\alpha, \beta} = \cosh(|k_\alpha|h) \delta_{\alpha\beta}$, $[\hat{\mathcal{S}}_k]_{\alpha, \beta} = \sinh(|k_\alpha|h) \delta_{\alpha\beta}$, $[\hat{K}_k]_{\alpha, \beta} = [k_\alpha]_{\delta\beta\beta}$, $[\hat{\mathcal{P}}_{\mu}]_{\alpha, \beta} = \frac{1}{|S|} \int_S e^{-i(k_{\alpha} - k_\beta) \cdot r} d\mathbf{r}$, $[\hat{B}_k]_{\mu, \alpha} = \frac{1}{|S|} \int_S T_\mu (r) e^{i k_{\alpha} \cdot r} d\mathbf{r}$, and $[\hat{\Xi}]_{\mu, \nu} = \xi_\mu \delta_{\mu\nu}$. $\hat{P}$ does a spatial projection into the hole region $S$, and $\hat{B}_k$ is the basis transformation from $\{e^{i k_{\alpha} \cdot r}\}$ into $\{T_\mu (r)\}$. Then, the boundary conditions at $z = -h$ are expressed as:

$$
\hat{P} \left( \hat{\mathcal{C}}_k |c\rangle - \hat{\mathcal{S}}_k |s\rangle \right) = \hat{\mathcal{C}}_k |c\rangle - \hat{\mathcal{S}}_k |s\rangle \quad (\because \Phi_k(r, z = -h) \text{ vanishes outside the hole}),
$$

$$
|t\rangle = \hat{B}_k \left( \hat{\mathcal{C}}_k |c\rangle - \hat{\mathcal{S}}_k |s\rangle \right) \quad (\because \text{Continuity of } \Phi_k(r, z) \text{ at } z = -h),
$$

(S7)

$$
\hat{\Xi} |t\rangle = \hat{B}_k \left( \hat{\mathcal{C}}_k \hat{\mathcal{K}}_k |s\rangle - \hat{\mathcal{S}}_k \hat{\mathcal{K}}_k |c\rangle \right) \quad (\because \text{Continuity of } \partial_z \Phi_k(r, z) \text{ at } z = -h).
$$

Using $\frac{1}{|S|} \sum_{\mu} T_\mu (r') T_\mu (r) = \delta(r - r')$, we get $\hat{B}_k^{\dagger} \hat{B}_k = \frac{|S|}{|\Omega|} \hat{P}$. Then, Eqs. (S7) is simplified to:

$$
\hat{\mathcal{C}}_k |c\rangle - \hat{\mathcal{S}}_k |s\rangle = \hat{Q}_k \left( \hat{\mathcal{C}}_k \hat{\mathcal{K}}_k |s\rangle - \hat{\mathcal{S}}_k \hat{\mathcal{K}}_k |c\rangle \right) \longrightarrow |s\rangle = \left( \hat{Q}_k \hat{\mathcal{C}}_k \hat{\mathcal{K}}_k + \hat{\mathcal{S}}_k \right)^{-1} \left( \hat{Q}_k \hat{\mathcal{S}}_k \hat{\mathcal{K}}_k + \hat{\mathcal{C}}_k \right) |c\rangle \quad (S8)
$$

where $\hat{Q}_k \equiv \frac{|S|}{|\Omega|} \hat{B}_k^{\dagger} \hat{\Xi}^{-1} \hat{B}_k$. The matrix $\hat{Q}_k$ can be directly calculated as cross-Fourier transform of the green’s function inside the holes $g(r, r') = \frac{|S|}{|\Omega|} \sum \xi_{\mu}^{-1} T_\mu (r) T_\mu (r')$ [66]:

$$
\left[ \hat{Q}_k \right]_{\alpha, \beta} = \frac{|S|}{|\Omega|} \hat{B}_k^{\dagger} \hat{\Xi}^{-1} \hat{B}_k \left[ \frac{1}{|S|} \int_S T_\mu (r) e^{-i k_\alpha \cdot r} d\mathbf{r} \right] \frac{1}{|S|} \int_S T_\mu (r') e^{i k_\beta \cdot r'} d\mathbf{r'}
$$

$$
= \frac{1}{|\Omega|} \int d\mathbf{r} \int_S d\mathbf{r'} e^{-i k_\alpha \cdot r} e^{i k_\beta \cdot r'} g(r, r').
$$

(S9)

Recalling $|c\rangle = \frac{1}{\xi_0} \hat{V}_k \hat{\mathcal{K}}_k (|s\rangle + |s'\rangle)$, we finally obtain the exact expression for $\hat{V}_k$ that captures the MG screening effect:

$$
\hat{V}_k^{\text{MG}} = \frac{\epsilon^2}{\epsilon_0} \left( \hat{K}_k + \hat{K}_k \hat{S}_k \hat{C}_k^{-1} + \left( \hat{C}_k \hat{Q}_k \hat{C}_k + \hat{K}_k^{-1} \hat{S}_k \hat{C}_k \right)^{-1} \right)^{-1}. 
$$

(S10)
We start from evaluating $\hat{X}_k$ in local Drude approximation [49]:

$$\sigma_{\text{Drude}}(\mathbf{r}, \mathbf{r}', \omega) = \frac{e^2}{2\pi e\hbar^2 \omega + i\gamma} E_F(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}');$$

then, Eq. (S2) gives

$$\sigma_{\text{Drude}}(\mathbf{q}, \mathbf{q}', \omega) = \frac{1}{(2\pi)^3} \int \mathcal{D}\mathbf{r} \int_{\Omega} \mathcal{D}\mathbf{r}' e^{-i\mathbf{q}\cdot\mathbf{r}} e^{i\mathbf{q}'\cdot\mathbf{r}'} \frac{e^2}{\pi\hbar^2 \omega + i\gamma} E_F(\mathbf{r}) \delta(\mathbf{r} - \mathbf{r}') = \frac{e^2}{\pi\hbar^2 \omega + i\gamma} \tilde{E}_F(\mathbf{q} - \mathbf{q}'),$$

where $\tilde{E}_F(\mathbf{q}) = \frac{1}{(2\pi)^3} \int_{\Omega} e^{-i\mathbf{q}\cdot\mathbf{r}} E_F(\mathbf{r}) d\mathbf{r}$. Recalling the definition of the proper density-density response function $n(\mathbf{q}, \omega) = \sum_{\mathbf{q}'} \chi_{nm}(\mathbf{q}, \mathbf{q}', \omega) \phi(\mathbf{q}', \omega)$ and using a set of relations, (1) $\mathbf{J}(\mathbf{q}, \omega) = \sum_{\mathbf{q}'} \sigma(\mathbf{q}, \mathbf{q}', \omega) \mathbf{E}(\mathbf{q}', \omega)$, (2)
\[ i \mathbf{q} \cdot \mathbf{J}(\mathbf{q},\omega) - \omega \cdot (-e)n(\mathbf{q},\omega) = 0, \quad \text{and} \quad (3) \quad -eE(\mathbf{q},\omega) = -i\mathbf{q}\phi(\mathbf{q},\omega), \]

we obtain the fundamental relation between the conductivity and the proper density-density response function: 
\[ \chi_{nn}(\mathbf{q},\mathbf{q}',\omega) = \frac{9e^2 \sigma(\mathbf{q},\mathbf{q}')}{\pi \omega^2}. \]

Thus, \( \tilde{X}_k \) in local Drude approximation is given as
\[ [\tilde{X}_k^{\text{Drude}}]_{\alpha,\beta} = \omega(\omega + i\gamma) \chi_{nn}^{\text{Drude}}(k_\alpha, k_\beta, \omega) = \frac{k_\alpha \cdot k_\beta \tilde{E}_F(\mathbf{k}_\alpha - \mathbf{k}_\beta)}{\pi \hbar^2}. \]  

(S13)

As mentioned in the main text, \( \omega^2 \) is substituted by \( \omega(\omega + i\gamma) \) as the Drude loss \( \gamma \) is included in the analysis.

A more precise description of \( \chi_{nn}(\mathbf{q},\mathbf{q}',\omega) \) beyond Drude approximation is obtained in Random Phase Approximation (RPA), in which the proper density-density response function is simply replaced by the Hartree response function: 
\[ \chi_{nn}^{\text{RPA}}(\mathbf{q},\mathbf{q}',\omega) = \chi_{\text{H}}(\mathbf{q},\mathbf{q}',\omega) \]  

[48]. In a homogeneous electron gas, the Hartree response function reduces to the density-density response function in non-interacting limit, and the density-density response function of non-interacting 2D massless electron gas has an exact analytic expression in a non-relativistic limit [62]:

\[ \chi_{\text{H}}^{\text{hom}}(\mathbf{q},\mathbf{q}',\omega) = \frac{E_F}{\pi \hbar^2 v_F^2} \left( \frac{x^2 y}{1 - x^2} (F_+(x, y) - F_-(x, y)) - 2 \right), \]

where
\[ x \equiv v_F |\mathbf{q}| / \omega, \quad y \equiv \hbar \omega / 2E_F, \quad \text{and} \quad F_{\pm}(x, y) = \int_1^{1+y} du \sqrt{1 - u^2}. \]  

(S14)

Here, we assume \( x < 1 \) (suppressed intraband transitions), \( y(1 + x) < 1 \) (suppressed interband transitions), and the zero temperature limit (\( 2E_F \gg k_B T \)). For our purpose, we introduce the Taylor expansion of \( \chi_{\text{H}}^{\text{hom}}(\mathbf{q},\mathbf{q}',\omega) \) of a homogeneous 2D massless electrons respect to \( x \) and \( y \) up to the 6th order terms:

\[ \chi_{\text{H}}^{\text{hom}}(\mathbf{q},\mathbf{q}',\omega) = \delta_{\mathbf{q},\mathbf{q}'} \frac{E_F}{\pi \hbar^2 v_F^2} \left( x^2 - x^2 y^2 + \frac{3}{4} x^4 - \frac{1}{3} x^2 y^4 - \frac{1}{2} x^4 y^2 + \frac{5}{8} x^6 + O(x^n y^m; n + m > 6) \right) \]

\[ = \delta_{\mathbf{q},\mathbf{q}'} \left( \frac{E_F}{\pi \hbar^2 \omega^2} \left[ \frac{E_F^{-1}}{4\pi} |\mathbf{q}|^2 + \frac{3E_F^2 v_F^2 |\mathbf{q}|^4}{4\pi \hbar^2 \omega^4} \right] - \frac{h^2 \omega^2 E_F^{-1} |\mathbf{q}|^2}{8\pi \omega^2} - \frac{v_F^2 E_F^2 |\mathbf{q}|^6}{8\pi \hbar^2 \omega^6} + O(x^n y^m; n + m > 6) \right). \]  

(S15)

The Hartree response function in inhomogeneously doped graphene, however, has not been microscopically calculated due to the difficulty in quantifying the interband transitions with a spatially varying Fermi level [50]. Instead, Ref. [50] suggested a semi-phenomenological expression for \( \chi_{\text{H}}^{\text{hom}}(\mathbf{q},\mathbf{q}',\omega) \) in an inhomogeneous case at small \( x \) and \( y \):

\[ \chi_{\text{H}}^{4\text{th}}(\mathbf{q},\mathbf{q}',\omega) = \chi_{nn}^{\text{Drude}}(\mathbf{q},\mathbf{q}',\omega) - \frac{E_F^{-1}}{4\pi} (\mathbf{q} - \mathbf{q}') \cdot \mathbf{q} - \frac{3E_F^2 v_F^2 E_F^{-1} (\mathbf{q} - \mathbf{q}') (\mathbf{q} \cdot \mathbf{q}')^2}{4\pi \hbar^2 \omega^4}, \]  

(S16)

where \( E_F^\mu(\mathbf{r}) = \frac{1}{|\Omega|} \int_{\Omega} e^{-i \mathbf{q} \cdot \mathbf{r}} E_F^\mu(\mathbf{r}) d\mathbf{r} \). This expression is acquired by substituting \( \delta_{\mathbf{q},\mathbf{q}'} E_F \) to \( \tilde{E}_F(\mathbf{q} - \mathbf{q}') \), \( \delta_{\mathbf{q},\mathbf{q}'} E_F^{-1} \) to \( E_F^{-1}(\mathbf{q} - \mathbf{q}') \), and \( |\mathbf{q}|^2 \) to \( \mathbf{q} \cdot \mathbf{q}' \) from Eq. (S15), so that it reduces back to Eq. (S15) in homogeneous limit while minimally capturing the non-local effects through the fourth-order dependence on momentum in the third term. Eq. (S16) was successfully used in [9] to predict the plasmonic scattering in an inhomogeneously doped graphene. We also use Eq. (S16) to examine the significance of non-local effects plasmonic dispersions in Fig. 5 in the main text and in the section 2.1. of this Supplemental Material. Similarly, we can extend this semi-phenomenological expression up to the 6th-order terms:

\[ \chi_{\text{H}}^{6\text{th}}(\mathbf{q},\mathbf{q}',\omega) = \chi_{\text{H}}^{4\text{th}}(\mathbf{q},\mathbf{q}',\omega) - \frac{h^2 \omega^2 E_F^{-3} (\mathbf{q} - \mathbf{q}') \cdot \mathbf{q} - \frac{3E_F^2 v_F^2 E_F^{-1} (\mathbf{q} - \mathbf{q}') (\mathbf{q} \cdot \mathbf{q}')^2}{8\pi \omega^2} + \frac{5v_F^4 E_F(\mathbf{q} - \mathbf{q}') (\mathbf{q} \cdot \mathbf{q}')^3}{8\pi \hbar^2 \omega^6}. \]  

(S17)

1.3. 2D bulk dispersion

The eigenvalue problem given in Eq. (S1a) is generally nonlinear, because the operators themselves are not independent of the eigenvalues. Only in local Drude approximation and in vacuum, as mentioned above, both \( \tilde{X}_k \) (Eq. (S10) and \( X_k \) (Eq. (S13)) are independent of the frequency, and Eq. (S1a) becomes an ordinary linear eigenvalue problem. \( \tilde{X}_k \) and \( X_k \) are discrete, but infinite matrices; thus, we need to truncate the matrices beyond certain orders in numerical calculations. Figure S2 depicts the speed of convergence of the numerical results. If the truncation order of \( \alpha \) is \( N \), we use \( (2N + 1)^2 \) planar modes; \( |n_1|, |n_2| \leq N \) with \( G_{\alpha=(n_1,n_2)} = n_1G_1 + n_2G_2 \). Similarly, if the truncation order
of μ is M, we use $M^2$ laplacian eigenmodes of a triangular hole for computing $\hat{Q}_k$; $0 < p, q \leq M$ with $T_{\mu=(p,q)}(r)$. We use $N = M = 10$ in calculating the bulk dispersions of Fig. 1(d)-(e) and Fig. 2(b) in the main text.

When we use Eq. (S16) for the density-density response operator, instead of Eq. (S13), for evaluating the non-local effects in the section 2.1, the eigenvalue problem obtains a quadratic form:

$$\left( \lambda_{nk}^2 \left[ \hat{X}_k^{(2)} V_{MG}^{(2)} - I \right] + \lambda_{nk} \hat{X}_k^{(1)} V_{MG}^{(1)} + \hat{X}_k^{(3)} V_{MG}^{(3)} \right) \phi_{nk} = 0,$$

(S18)

where $\hat{X}_k^{(1)} = \hat{X}_k^{Drude}$, $\hat{X}_k^{(2)}_{\alpha,\beta} = -\frac{1}{4\pi\epsilon_0} E_F^{-1} (k_\alpha - k_\beta) k_\alpha \cdot k_\beta$, and $\hat{X}_k^{(3)}_{\alpha,\beta} = \frac{3\alpha^2}{4\pi\epsilon_0} E_F (k_\alpha - k_\beta) (k_\alpha \cdot k_\beta)^2$. We solve this quadratic eigenvalue equation for calculating plasmonic dispersions including the non-local effects in Fig. S5(a).

For the calculation of Fig. 5 in the main text, $V_{MG}^{(1)}$ is slightly modified to account for the anisotropic ($\epsilon_\parallel \neq \epsilon_\perp$) dispersion of hBN:

$$V_{MG}^{(1)} = \frac{\epsilon_0^2}{\epsilon_{eff}} \left( k_{\parallel}^2 + k_{\parallel}^2 \hat{S}_{k}^{\parallel} C_{\parallel}^{-1} + \left( C_{\parallel}^{\prime}\epsilon_{eff} \hat{Q}_k \hat{C}_{\parallel} + \hat{k}_{\parallel}^{-1} \hat{S}_{k}^{\parallel} C_{\parallel}^{-1} \right) \right)^{-1},$$

(S19)

where $[\hat{S}_{k}]_{\alpha,\beta} = \sinh(\eta) |k_\alpha| h) \delta_{\alpha,\beta}$, $[\hat{C}']_{\alpha,\beta} = \cosh(\eta) |k_\alpha| h) \delta_{\alpha,\beta}$, $\epsilon_{eff} = \sqrt{\epsilon_\parallel/\epsilon_\perp}$, and $\eta = \sqrt{\epsilon_\parallel/\epsilon_\perp}$. Then, $V_{MG}^{(1)}$ acquires non-polynomial frequency dependence due to the dispersive nature of the permittivity. This non-polynomial type of eigenvalue problems is then solved by iterations just like in the Newton’s method.

### 1.4. 1D Edge dispersion

In the edge dispersion calculation of Fig. 3(c), we essentially perform a 2D bulk dispersion calculation with a nanoribbon-like unit cell that consists of the upper domain of 40 up-triangles ($\theta = 30^\circ$) and the lower domain of 40 down-triangles ($\theta = -30^\circ$), see Fig. 3(b). The length of the unit cell is $L_x = L$ along the $x$-axis, and $L_y = 20\sqrt{3}L$ along the $y$-axis. The $y$-component of Bloch wavevector is taken to be zero, since we are interested only in the periodicity along the $x$-axis. Within the unit cell, the linear coefficients $t_{\mu}$ of each hole are completely independent. Thus, Eq. (S6) needs to be modified as:

$$\Phi_{k=\theta,\pi}(r, z < -h) = \sum_{R} e^{ik_{\parallel} R} \sum_{j} \phi_{\text{hole}}^j (r - r_c - R, z),$$

with

$$\phi_{\text{hole}}^j (r, z) = \sum_{\mu} t_{\mu}^j T_{\mu}^j (r) e^{i\alpha_\mu(z+h)} \quad \text{and} \quad T_{\mu}^j (x, y) = \begin{cases} T_{\mu} (x, y) & (\theta_j = 30^\circ) \\ T_{\mu} (x, -y) & (\theta_j = -30^\circ) \end{cases}$$

(S20)

where $j$ is the index for 80 different triangular holes within a unit cell, $r_c^j$ is the relative displacement of the center of the hole $j$ from the origin of the unit cell, and $\theta_j$ is the rotation angle of the hole $j$. Here, $T_{\mu}(r)$ specifically represents the triangular eigenmodes of a triangular hole with the orientation $\theta = 30^\circ$. Then, the matrix elements of the basis transformation operator $\hat{B}_{k=\theta,\pi}$ changes accordingly:

$$[\hat{B}_{k=\theta,\pi}]_{\alpha, j} = \frac{1}{|S|} \int_S T_{\mu}^j (r - r_c^j) e^{ik_{\parallel} \cdot r} \, dr.$$

(S21)
Here, the number pair of $\mu$ and $j$ comprises of the row index of $\hat{B}_{k=k,x}$ operator. In the calculation of Fig. 3(c), the truncation order of $\alpha$ is 5 along the $x$-axis and 200 along the $y$-axis: $|n_1| \leq 5, |n_2| \leq 200$ with $G_{\alpha=(n_1,n_2)} = \frac{2\pi}{L_x} n_1 \hat{x} + \frac{2\pi}{L_y} n_2 \hat{y}$. We still use the truncation order of $\mu$ of $M = 10$ per triangle.

### 1.5. Valley-selective excitation and topological valley transport

In the presence of an external field $|\phi_{\text{ext}}\rangle$ at a specific frequency $\omega_{\text{ext}}$, the plasmonic equation becomes a linear problem:

$$|\phi\rangle = |\phi_{\text{ext}}\rangle + (\omega_{\text{ext}}(\omega_{\text{ext}} + i\gamma))^{-1} \hat{V}_{k\to0} \hat{X}_{k\to0} |\phi\rangle \implies |\phi\rangle = \left( I - (\omega_{\text{ext}}(\omega_{\text{ext}} + i\gamma))^{-1} \hat{V}_{k\to0} \hat{X}_{k\to0} \right)^{-1} |\phi_{\text{ext}}\rangle. \quad (S22)$$

We calculate $|\phi\rangle_{\text{ext}}$ from the array of point dipoles (see Fig. S3(a)), considering the MG screening in a similar method described in the section 1.1. The matrix elements of the basis transformation operator $\hat{B}_{k\to0}$ are evaluated using Eq. (S21).

As mentioned in the main text, our numerical method intrinsically imposes periodic boundary conditions; thus, as depicted in Fig. S3(b), we are actually simulating over a 2D periodic array of the dipole arrays. To minimize the effect of the dipole array from a unit cell into the adjacent cells, we use a large domain size ($50 \times 32 = 1600$ triangles), and

**FIG. S3.** Computational details in the calculations of Fig. 4. (a) Dipole array. (b) Illustration of the periodic boundary conditions in the example of Fig. 4(d): the positive-type domain walls are in red dotted lines, and the negative-type domain walls are in blue dotted lines (the black dotted lines are borders of unit cells). (c) Full computational geometry of Fig. 4(d).
introduce a finite Drude loss ($\omega_{\text{ext}}/\gamma = 300$) so that the excited edge mode decays to a negligible amplitude during the propagation within a domain (Fig. S3(c)). In the calculations of Fig. 4, the truncation order of $\alpha$ is 150 along the $x$-axis and 96 along the $y$-axis; thus, $\hat{V}_k \to 0$ and $\hat{X}_k \to 0$ are $58093 \times 58093$ complex matrices. In order to reduce the appreciable memory requirement, we lower the truncation order of $\mu$ to 5 per triangle; still, $\hat{B}_k \to 0$ is a complex $64000 \times 58093$ matrix.

2. NON-SCALEABLE CORRECTIONS AND LANDAU DAMPING

The plasmonic dispersions are not scalable under consideration of non-scalable corrections. In this section, we discuss the significance of these corrections and estimate the lower bound of the system size, above which graphene plasmons don’t suffer from additional losses due to intra-/interband electronic transitions.

2.1. Quantum capacitance of graphene and non-local effects

In electrochemical equilibrium, the static potential energy $\phi_{\text{ES}}(r)$ and the static charge density on graphene $n_{\text{ES}}(r)$ have to satisfy:

$$\phi_{\text{ES}}(r) + E_F(r) = eV \rightarrow \phi_{\text{ES}}(r) + \hbar v_F \sqrt{\pi n_{\text{ES}}(r)} = eV,$$

(S23)

where $V$ is the voltage difference applied between graphene and MG. The Fermi energy term can be interpreted as the contribution from an additional capacitance that is intrinsic to graphene, which is called the quantum conductance (QC) of graphene [46]. Eq. (S11) and Eq. (S23) altogether give a self-consistent Thomas-Fermi solution. The difference between this self-consistent solution and the simple electrostatic simulation is negligible when $eV$ is much greater than $E_F(r)$. If we assume $L/r = 4$, $L/h = 25$, and $V/h = 1 \text{ V/nm}$, the correction from quantum capacitance of graphene is not very significant unless $L$ goes below 100 nm ($h = 4 \text{ nm}$), see Fig. S4.

We calculate the non-scalable corrections in $\omega_{1K}$ and $\omega_{2K}$ considering both QC of graphene and non-local effects (Eq. (S18)), see Fig. S5(b). These corrections reduces the bandgap size and the mid-gap frequency, and their effects decreases as the system size $L$ increases or as the induced Fermi energy $E_F \propto \sqrt{V/h}$ increases. As shown in Fig. S5(a), the non-scalable corrections are more significant at higher bands.
FIG. S5. (a) Blue: Band dispersion calculated without non-scalable corrections ($L/r = 4$, $L/h = 25$, and $\theta = 30^\circ$). Red: Band dispersion calculated with non-scalable corrections at $L = 100\text{nm}$ and $V/h = 0.5V/\text{nm}$. (b) Non-scalable corrections in $\omega_{\mathbf{K}}$ and $\omega_{2\mathbf{K}}$ ($L/r = 4$, $L/h = 25$, and $\theta = 30^\circ$). (c) Threshold parameters for landau damping ($x$, dark blue) and interband EHPs ($y(1+x)$, brown) evaluated with the mid-gap frequency at $\mathbf{K}$ and the minimum value of $E_F(r)$ ($L/r = 4$, $L/h = 25$, and $\theta = 30^\circ$).

2.2. Intraband and interband transitions

According to Ref. [62], even when the Drude loss $\gamma$ is zero, the imaginary part of the Hartree function of 2D massless electrons becomes non-zero under certain energy-momentum ranges:

$$\hbar \omega < \hbar v_F |\mathbf{q}|$$

$$\hbar \omega + \hbar v_F |\mathbf{q}| > 2E_F,$$

which implies a finite loss of plasmonic excitations regardless of the quality of graphene growth. If we use the dimensionless parameters used in Eq. (S16), $x \equiv v_F |\mathbf{q}|/\omega$ and $y \equiv \hbar \omega/2E_F$, those conditions are simplified to $x > 1$ and $y(1+x) > 1$. Actually, these are also the conditions where Eq. (S18) is no longer valid. Physically, Eq. (S24a), or $x > 1$, corresponds to the condition for severe landau damping, and Eq. (S24b), or $y(1+x) > 1$, corresponds to the condition for the excitation of interband electron-hole pairs (EHP). When the plasmon frequency and momentum lie in those ranges, the plasmonic excitations cannot be sustained even in perfectly clean ($\gamma = 0$) graphene. Figure. S5(b) depicts these threshold parameters, $x$ and $y(1+x)$, at several different system sizes and different doping levels in our system. In the calculations, we take $\omega = \frac{1}{2}(\omega_{\mathbf{K}} + \omega_{2\mathbf{K}})$ and $|\mathbf{q}| = |\mathbf{K}| = \frac{4\pi}{L}$ for rough estimation of those parameters near the topological bandgap. Because the induced Fermi energy is not homogeneous across graphene, we take the minimum value of $E_F(r)$ for the estimation of $y$. As clearly seen in Fig. S5(b), our system is far from regions of Landau damping as long as the system size stays reasonable—accessible in the capabilities of up-to-date nanofabrication technologies. Note that $x$ is equal to $\rho_2$ defined in the main text.

2.3. Higher order corrections in non-local response and $v_F$-renormalization

We confirm that it is sufficient to take the terms upto the fourth order in $x$ and $y$ (equivalent to the second order in $|\mathbf{q}|^2$, as mentioned in the main text) in $\chi_H$ for precise determination of the bandgap size and location; Figure. S6(b) shows that there is no significant difference between the dark green and the red curve.

We also confirm that the effect of $v_F$-renormalization is negligible. To model $v_F$-renormalization, we use the density-dependent renormalized Fermi velocity expression from Ref. [44]:

$$v_F(n) = v_F(n_0) \left( 1 + \frac{1}{4\pi} \frac{\epsilon_{gr}}{\epsilon_{gr} + \epsilon_{BN}} \ln(n_0/n) \right),$$

where $n_0 = 10^{15}\text{cm}^{-2}$ is the cut-off density, $v_F(n_0) = 0.85 \times 10^6\text{m/s}$, $\epsilon_{gr} = e^2/8\hbar v_F \epsilon_0 \sim 4$ is the graphene screening, and $\epsilon_{BN} \sim 3.5$ is the static permittivity of hBN along z-direction. When we ignore $v_F$-renormalization in Fig.5. in
FIG. S6. An example of band dispersion in a dispersive medium (hBN) (a) hBN permittivity data from [7, 61] (b) The dark green curve is the dispersion calculated using $\chi^{4\text{th}}$ (Eq. S16) without considering $v_F$-renormalization; the red curve is calculated using $\chi^{6\text{th}}$ (Eq. S18) without considering $v_F$-renormalization; the blue curve is calculated using $\chi^{4\text{th}}$ with $v_F$-renormalization (Eq. S25) (the graphene quantum capacitance is accounted for all curves)

the main text, or in the green and red curves in Fig. S6, we take the value of $v_F$ to be the average value of $v_F(n(r))$ over a unit cell, which was $\sim 1.02 \times 10^6$ m/s. As clearly seen in the negligible difference between the dark green and blue curves in Fig. S6, the effect of $v_F$-renormalization can be effectively ignored.

3. HERMITICITY OF THE PLASMONIC EIGENVALUE EQUATION AND VALLEY-CHERN NUMBER

Here, we confirm that the plasmonic eigenvalue equation given in Eqs. (2) in the main text is Hermitian under an appropriate inner product. Our discussion given in this section shares similarities to the discussion given in Ref. [16] (see the second section of their supplemental material).

3.1. Generalized Hermitian eigenvalue problem

The plasmonic eigenvalue equation in Eqs. (2) is $\hat{V}_k \hat{X}_k |\phi_{nk}\rangle = \lambda_{nk} |\phi_{nk}\rangle$, where $|\phi_{nk}\rangle = \phi_{sc}(k + G_{\alpha}, \omega_{nk})$, $|\phi_{nk}\rangle = \phi_{sc}(k + G_{\alpha}, \omega_{nk})$, and $\lambda_{nk}$ is the eigenvalue. In previous sections, we observed that both $\hat{V}_k$ and $\hat{X}_k$ are Hermitians; however, their product is not Hermitian under the standard inner product. Nevertheless, we can simply recast this problem as a generalized eigenproblem by multiplying the both sides by the inverse of the Coulomb operator $\hat{V}_k^{-1}$:

$$\hat{X}_k |\phi_{nk}\rangle = \lambda_{nk} \hat{V}_k^{-1} |\phi_{nk}\rangle.$$  \hspace{1cm} \text{(S26)}

Then, Eq. (S26) is a generalized eigenvalue equation with Hermitian definite pencil $\{\hat{X}_k, \hat{V}_k^{-1}\}$ [64].

3.2. Berry curvature and valley-Chern number

In the calculations of Berry curvature, we use the method described in Ref. [58] is. The detailed description of the method can be also found in the supplemental material of Ref. [16]. Here, we briefly note two important comments regarding the calculations of Berry curvature and valley-Chern numbers. First, when we evaluate an inner product between two state vectors, let’s say $|\phi\rangle$ and $|\phi'\rangle$, during the calculations of Berry curvature, their inner product is defined respect to the Hermitian pencil $\{\hat{X}_k, \hat{V}_k^{-1}\}$:

$$\langle \phi | \phi' \rangle \equiv \sum_{\alpha,\beta} \left[ \hat{V}_k^{-1} \right]_{\alpha,\beta} |\phi\rangle^*_\alpha |\phi'\rangle^*_\beta.$$  \hspace{1cm} \text{(S27)}

Second, as mentioned in the main text, we define the valley-Chern number $C_\nu$ to be the integration of the berry curvature over a triangular area centered at the valley, where the vertices of the triangle are given as the nearest
FIG. S7. Effect of the finite depth of the triangular holes in MG onto the topological bandgap (a) A realistic MG structure achievable via etching; side view. (b) Shifts in the plasmonic frequencies $\omega_{1,2K}$ as a function of the depth of the holes ($L/r = 4, L/h = 25, \theta = 30^\circ$); the dotted lines indicates the calculated frequencies under the assumption $H_{MG} \to \infty$.

Γ points from the valley. We come up with this artificial definition because two valleys actually share the same momentum space in contrary to the assumption considered in the calculation of $C_\nu$ using effective Hamiltonians [19] that two valleys are separated into two totally decoupled subspaces. The triangular area defined above is the largest momentum subspace, in which the valley of our interest in $C_\nu$ calculation is the closest, and occupies exactly the half of the whole Brillouine zone. The calculated values of $|C_\nu|$ in Fig. 3(a) in the main text show small deviations from 0.5 for three reasons: (1) our choice of the integration range cannot exclude the contributions from Γ-points that are not relevant to valley-based dynamics, (2) our system is not a two-level system, and (3) the contributions from the adjacent valleys are still present because the valleys share the same momentum space.

4. POSSIBLE PERTURBATIONS ON METAGATE STRUCTURE

So far, we have assumed that the holes in MG are infinitely deep and perfectly triangular. In this section, however, we break these two assumptions, thereby perturbing the geometry of the holes in MG. Then, we examine how much the size of the topological bandgap is robust against possible perturbations on the structure of MG that may occur due to imperfection in nanofabrication.

4.1. Finite aspect ratio of the triangular holes

First, we evaluate the effect of finite depth of the holes in MG. We consider a structure depicted in Fig. S7(a) that the other end of the holes are terminated by the same metallic material. In order to take the finite depth of the holes

FIG. S8. Effect of the finite radius of curvature at the corners of the triangular holes (a) A realistic MG structure achievable via etching following after electron-beam lithography; top view. (b) Shifts in the plasmonic frequencies $\omega_{1,2K}$ as a function of the radius of curvature at the corners of the triangle ($L/r = 4, L/h = 25, \theta = 30^\circ$).
into account in our numerical model, we simply modify the ansatz for the potential in the holes $\phi_{\text{hole}}(r, z)$ in Eq. (S6) so that $\phi_{\text{hole}}(r, z = -h - H_{\text{MG}}) = 0$:

$$
\phi_{\text{hole}}(r, z) = \sum_{\mu} t_{\mu} T_{\mu}(r) \left( \cosh (\xi_{\mu}(z + h)) + \coth (\xi_{\mu} H_{\text{MG}}) \sinh (\xi_{\mu}(z + h)) \right),
$$

where $H_{\text{MG}}$ is the depth of the holes. Then, the matrix elements of $\hat{\Xi}$ needs to be redefined accordingly for the third boundary condition matching in Eq. (S7): $[\hat{\Xi}]_{\mu,\nu} = \xi_{\mu} \coth (\xi_{\mu} H_{\text{MG}}) \delta_{\mu\nu}$. The rest of the discussion in the section 1.1. following after Eq. (S7) applies in the same way.

Figure. S7(b) illustrates that the topological bandgap at $K$ point is highly insensitive to $H_{\text{MG}}$ as long as it is greater than $\sim 5$ times $h$. This is due to the ultrahigh field confinement of acoustic plasmons in graphene with metal [1]. Given that $L/r = 4$, $L/h = 25$, and $H_{\text{MG}} = 5h$, the required aspect ratio of the triangular holes $2\sqrt{3}r/H_{\text{MG}}$ (the ratio of the side length of the triangle to the height of the holes) is estimated to be around 4. For example, if $L$ is 200 nm, the side length of the triangle is 173 nm, and the required height of the holes is 40 nm (assuming $H_{\text{MG}} = 5h$). Thus, the assumption that the holes are infinitely deep works well as long as the triangular shape of the holes are maintained for a relatively shallow depth.

### 4.2. Round corners of the triangle

Next, we introduce a finite radius of curvature at the corners of the triangle as illustrated in Fig. S8(a). In calculation of Fig. S8(b), we choose to use COMSOL Multiphysics Electromagnetic Waves, Frequency Domain module instead of our in-house codes since it is tricky to compute the matrix elements of $\hat{B}_K$ when the analytic expression for the Laplacian eigenfunctions is not known. Assuming $L = 200$ nm and $r = 50$ nm, the reduction in the size of the topological bandgap, $\omega_{2K} - \omega_{1K}$, is calculated to be only around 3% with $d = 10$ nm ($d/r = 0.2$). Thus, our theoretical predictions made with the assumption that the holes are perfectly triangular are still sufficiently precise despite roundy corners of the triangle in real samples, given that a resolution of $d = 10$ nm (the minimum diameter of a dot is 20 nm) is well within experimental capabilities in electron-beam lithography [65] or focused ion-beams.