Room Temperature Amplification of Terahertz Radiation by Grating-Gate Graphene Structures

Stephane Boubanga-Tombet,1 Wojciech Knap,1,2,3 Deepika Yadav,1 Akira Satou,1 Dmytro B. But,1,2,4 Vyacheslav V. Popov,2,5 Ilya V. Gorbenko,6 Valentin Kachorovskii,2,6 and Taiichi Otsuji1,∗

1Research Institute of Electrical Communication, Tohoku University, Sendai 980-8577, Japan
2CENTERA Laboratories, Institute of High Pressure Physics PAS, Warsaw 01-142, Poland
3Laboratory Charles Coulomb, University of Montpellier and CNRS, Montpellier F-34095, France
4CEZAMAT Warsaw Technical University, Warsaw 02-346, Poland
5Kotelnikov Institute of Radio Engineering and Electronics (Saratov Branch), RAS, Saratov 410019, Russia
6Ioffe Institute, 194021 St. Petersburg, Russia

(Dated: March 24, 2020)

We report on experimental studies of terahertz (THz) radiation transmission through grating-gate graphene-channel transistor nanostructures and demonstrate room temperature THz radiation amplification stimulated by current-driven plasmon excitations. Specifically, with increase of the dc current under periodic charge density modulation, we observe a strong red shift of the resonant THz plasmon absorption, its complete bleaching, followed by the amplification and blue shift of the resonant plasmon frequency. Our results are, to the best of our knowledge, the first experimental observation of energy transfer from dc current to plasmons leading to THz amplification. We present a simple model allowing for phenomenological description of the observed amplification phenomena. This model shows that in the presence of dc current the radiation-induced correction to dissipation is sensitive to the phase shift between THz oscillations of carrier density and drift velocity, and with increase of the current becomes negative, leading to amplification. The experimental results of this work as all obtained at room temperature, pave the way towards the new 2D plasmons based, voltage tuneable THz radiation amplifiers.

I. INTRODUCTION

More than forty years ago, active theoretical and experimental studies of plasma oscillations in two-dimensional electron systems (2DESs) began and plasmonic resonances were observed [1–7]. The interest to this area dramatically increased after seminal work of Dyakonov and Shur [8], who theoretically predicted that the dc current in the channel of a field effect transistor (FET) can become unstable with respect to the excitation of plasma waves and, as a result, to the generation of radiation with the frequency controlled by the gate voltage. For typical values of FETs the oscillation frequency can be easily tuned to be in the terahertz (THz) range. This work, as well as the next publication [9] focused on the plasmon-mediated THz detection aroused great interest in the scientific community due to its fundamental and applied importance in the field of THz physics. Nevertheless, despite more than 25 years of tremendous effort, this instability has not yet been demonstrated experimentally. Initially, numerous attempts were made to find instability in single FETs made of different materials [10–18], but all failed. Quickly enough it was understood that multiple-gate periodic structures are more promising. Such structures interact much better with THz radiation and, as was demonstrated for the first time in a pioneer work of Muravjov et all [19], where high quality plasmon resonances in absorption were observed in GaN/AlGaN 2D grating gate structures showing very good characteristics as THz radiation detectors. Nevertheless, numerous experimental attempts to measure the room temperature current-stimulated emission[20] or amplification of THz radiation from multi-gate structures have also not been successful so far.

A fundamentally new aspect of our work is the use of graphene [21, 22] as an active element of such structures. Graphene shows record mobility at room temperatures, which gives much higher quality factor of plasma resonances than that in conventional materials. This explains growing interest in graphene plasmonics [23–30]. The mediation of plasmons can greatly enhance the interaction between light and graphene due to their slow-wave nature, relatively low level of losses, and a high degree of spatial confinement of their field in graphene. Plasmon enhanced THz graphene devices have recently been investigated [31, 32] and improvement of the device performances on gain modulation [32, 33], sensitivity [34] and emission [31] have been reported, yielding the foundational pillars for future robust graphene based THz technology. Also recently some theoretical and experimental studies of grating gate graphene structures considering plasma resonances were reported [35, 36]. Nevertheless, although intensive studies of the THz resonant detection in graphene-based periodic structures have already been carried out, there has not yet been a detailed study of current-induced effects. Also, there does not exist any report of room temperature experiments showing any efficient energy transfer from dc current to plasmons leading to THz radiation generation or amplification. [37–39].

In this work, we explore THz light-plasmon coupling,
light absorption and amplification by graphene structures focusing on current-driven effects. A very important aspect of our work is a usage of grating gate structures, which are known to increase coupling between radiation and plasmons [19]. These grating gate structures allowed us to create a periodic structure of highly-conducting active regions separated by low-conducting passive regions. The large difference between concentrations allowed us for the first time to localize plasmons in active regions and control their properties by the dc driving current. We demonstrate that plasmons oscillating at THz frequencies can be excited and tuned electrically by the gate voltage and, with increasing dc current, they may amplify THz radiation. More specifically, with the increase of dc current the plasmon spectra shows a strong red shift, complete bleaching of the resonant THz plasmon absorption, followed by the amplification. Most importantly, the plasmon-mediated process is remarkably strong with up to ~ 19% extinction, and up to about ~ 9% amplification coefficients. It is also tunable over a wide THz frequency range. The whole phenomena may take place at a room temperature what is of a paramount importance for potential applications.

We also present a phenomenological theoretical description of observed results. We consider a simple model of periodically alternating stripes of the electron density and plasma wave velocity: active regions with high plasma wave velocity and low resistance and passive regions with low plasma wave velocity and very high resistance. We assume that dc current is applied to this plasmonic crystal and calculate radiation-induced correction to the dissipation in the channel. We show that in the presence of dc current the radiation-induced correction to dissipation becomes sensitive to the phase shift between THz oscillations of carrier density and drift velocity. We demonstrate that with increasing the dc current the radiation-induced correction to dissipation changes sign, which results in amplification of the optical signal. Both our experimental observations and simplified phenomenological description establish clearly a new challenge for developing of microscopical theory. The key concepts in developing such a description are also discussed.

Briefly speaking this work presents two main achievements that have the breakthrough character in the domain of two-dimensional plasmons interacting with THz radiation:

(i) Experiments demonstrating current-driven plasmons resonances at THz range that with the increase of the current undergo redshift, complete bleaching followed by the amplification and blue-shift of the resonant plasmon frequency.

(ii) Theory providing the phenomenological description of the experimentally observed phenomena, in particular, switching from dissipation to amplification at drift velocities smaller than the plasma wave velocity.

It should be stressed also that, the results presented in this work can be extremely important for applications since they show that one can use the gate voltage to control of carrier density (hence THz resonance frequency) and drain current/voltage to control switching between THz radiation dissipation or amplification regimes. All experimental results were obtained at room temperature, and therefore, they pave the way towards new robust graphene/semiconductors based resonant, compact and voltage/current controlled THz absorbers, amplifiers and sources.

II. EXPERIMENTAL

A. Samples

The samples were fabricated with a field-effect transistor structure featuring an interdigitated dual-grating-gate (DGG) where the plasmonic cavities are formed below the gates electrode grating fingers [40–43]. In order to ensure high carrier mobility in our devices, h-BN-encapsulated graphene heterostructures (h-BN/graphene/h-BN) were fabricated. Optically estimated h-BN thickness was in the range from 20 nm to 32 nm for both layers. The first h-BN layer was transferred onto a SiO$_2$/Si wafer serving as a part of the backgate dielectric layer. The process was followed by transferring the monolayer graphene. The second h-BN layer was transferred onto the monolayer graphene and served as the top gates dielectric layer. The layer sequence and the sample architecture are shown in Fig. 1. Raman spectroscopy (sharp mono peak at G and G' band) and the contrast fractions of the optical microscope images of the transferred graphene sheets allowed confirming that they were monolayers. Two device structures were designed with gate fingers width of $L_{g1} = 0.5 \, \mu m$ and $L_{g2} = 1 \, \mu m$ ($L_{g1} = 0.75 \, \mu m$ and $L_{g2} = 1.5 \, \mu m$) separated by $d_1 = 0.5 \, \mu m$ and $d_2 = 1 \, \mu m$ gaps ($d_1 = 0.5 \, \mu m$ and $d_2 = 2 \, \mu m$) referred to as Asymmetric DGG structures A-DGG 3.1 and (A-DGG 3.2), respectively (see Fig. 1). The devices had very similar electrical properties with very close Charge Neutrality Points (CNP), ranging from -0.1 V to +0.15 V.

In the experiments, we manipulated the electron concentration in the channel by gate voltages to create active and passive plasmonic regions with high and low electron concentration, respectively. Specifically, we always biased the gates in such a way that in given experiment configuration the only one type of gates was positively biased (electrical n-type doping) and all other types of gates were set at voltages granting the CNP condition of the graphene layer. In this way by successive biasing the different gates, we defined four types of grating gate structures (C1, . . . , C4) with cavities of length 0.5, 0.75, 1.0, 1.1 $\mu m$, each composed of six fingers with highly doped n-type cavities ($\sim 2 \times 10^{12} \, \text{cm}^{-2}$) separated by six regions at the CNP condition ($n_2 \sim 0.1 \times 10^{12} \, \text{cm}^{-2}$ for both electrons and holes at room temperature). The parameters of these structures are given in Table I. The ex-
ambipolar resistance versus back gate (a), and top gate for C1 cavity (b) of A-DGG3.1 structure. We show also resistance versus inverse swing voltage (1/(U_\text{g} - U_\text{CNP})) that allows extraction of the resistance of the contacts (a) and the resistance of total ungated part of samples for C1 cavity (b). C1 cavity carrier density, $n$ and mobility as a function of the gate voltage (c): lower horizontal-axis corresponds to the top gate bias and upper horizontal-axis is for back gate bias.

From the sample architecture, the biasing of the front gates change the resistance in much smaller degree than the back gating. The resistance change is roughly proportional to the ratio of the length of the biased gate ($\times$6) and the total length of the channel, as given in Tab. I. Knowing the total parasitic resistance $R_0$ we can extract sample averaged conductivity $\sigma = (L_{\text{ch}}/W)/(R - R_0)$, where $L_{\text{ch}}$ is the total channel length, $W$ is the average channel width, $R$ is the total resistance. The carrier mobility $\mu_i$ can be calculated from equation $e\mu_i(n_i + p_i) = \sigma_i$.
(see [26, 44] for details). The electron and hole density $n_i$ and $p_i$ (index $i = 1, 2$ numerates regions with high and low conductivity, respectively) as a function of the back gate voltage can be calculated using the parallel-plate capacitor model with the correction by the quantum capacitance of graphene. Figures 2(c) shows carrier density and mobility as a function of the front gate voltage (lower horizontal scale) and back gate voltage (upper horizontal scale). As can be seen, the mobility is over 30,000 cm$^2$/Vs in a whole investigated range indicating the excellent quality of h-BN-encapsulated graphene.

B. Time-domain spectroscopy experiments

Experimental set-up configuration is shown in Fig. 3(a). THz time-domain spectroscopy (THz-TDS) was employed to measure the changes in the temporal profiles of the THz pulses transmitted through the graphene plasmonic cavities. For the generation of broadband THz waves, a mode locked erbium-doped femtosecond fiber laser with a pulse repetition rate of 80 MHz was used. The pulse width was approximately 80 fs (full width at half maximum) and the central wavelength was 1550 nm. The pulsed laser beam was focused onto a biased, low-temperature-grown InGaAs/InAlAs THz photoconductive antenna (TERA15-TX, Menlosystems). The antenna had a gap spacing of 100 µm and a bias of 15 V was applied with an average optical power of 20 mW. The linearly polarized THz wave emitted by the antenna was collected and refocused onto the sample from 45° oblique angle (and subsequently onto the THz detector) using off-axis parabolic mirrors. A small 250 µm-diameter hole at the end of a tapered aperture of a conical shape was placed close to the sample. We have checked that this aperture was not destroying the polarization of the incoming beam. THz radiation was detected using a second low-temperature-grown InGaAs/InAlAs THz photoconductive antenna made of the 25 µm dipole and 10 µm gap excited with an average optical power of 17 mW (at 1550 nm). The THz-TDS system had an effective bandwidth of 0.1 - 3 THz with a relatively high signal-to-noise ratio (above 10$^4$) in the whole range. The sample stage was rotated to align the THz radiation polarization perpendicular or parallel to the grating fingers direction.

In the first part of experimental studies, we measured the spectra for zero drain-to-source voltage ($U_d = 0$). As mentioned above, the spectra were always recorded with the only one gate voltage biased away from CNP and with all other gates biased at the CNP. In the second part of experimental studies, $U_d$-dependent measurements were performed.

We used the reflection configuration, as shown in Fig. 3(b). The THz pulse was transmitted twice through the h-BN, graphene and thin h-BN/SiO$_2$ dielectric layers. Due to highly doped low resistivity conditions of the Si substrate with the 45° oblique angle incidence, the interface gave total reflection. When the THz pulse traveled through such the path, it reflected and transmitted through at the interfaces between graphene and the h-BN buffer layer and between the h-BN buffer layer and the SiO$_2$ layer. Such a multi-layered vertical structure makes a superposition of the reflected pulses at each interface in its temporal pulse profile. Since the thicknesses of the h-BN buffer layer ($\sim 40$ nm) and SiO$_2$ layer ($\sim 90$ nm) are far smaller than the wavelength of the THz radiation, the artefacts and distortions caused by the multiple reflections on the temporal pulse profile are negligibly small (the round-trip delay time from/to the top graphene surface will be $\sim 3$ fs which is almost two-orders shorter than the THz pulse width). Therefore, what is measured by the TDS is the total pulse after its double passage throughout the aforementioned multi-layered sample.

Some typical representative data are shown in Fig. 4. In each measurement we registered two time traces and calculated spectra: i) the first $T_{2\text{CNP}}$, with all gates biased the way ensuring CNP condition of the whole graphene channel and ii) the second $T_{2\text{P}}$ with only one type of the multi-finger gate ‘strongly biased’(up to $U_g - U_{\text{CNP}} \sim 3$ V) while keeping all other parts of the sample at the CNP condition. Then we calculated $2\tilde{A} = (1 - T_{2\text{P}}^{2}/T_{2\text{CNP}}^{2})$. As it will be shown later $2\tilde{A}$ contains mainly information about the extinction. The left panels in Fig. 4 are examples of some data (time traces, Fourier
TABLE I. Samples and different plasmonic cavities parameters.

| Cavity | C1  | C2  | C3  | C4  |
|--------|-----|-----|-----|-----|
| Structure/sample | A-DGG3.1 | A-DGG3.2 | A-DGG3.1 | A-DGG3.2 |
| Thickness of top h-BN layer (nm) | 32 | 20 | 32 | 20 |
| CNP (V) | +0.15 | -0.12 | +0.10 | -0.06 |
| Biased cavity length (µm) | 0.5 | 0.75 | 1.0 | 1.5 |
| Total channel length (µm) | 26.5 | 24.0 | 26.5 | 24.0 |
| d_1 and d_2 (µm) | 0.5 / 2.0 | 0.5 / 1.0 | 0.5 / 2.0 | 0.5 / 1.0 |
| Channel width (average) (µm) | 4.9 | 1.325 | 4.9 | 1.325 |
| d_3 and d_4 (µm) | 2.0 / 0.5 | 1.0 / 0.5 | 2.0 / 0.5 | 1.0 / 0.5 |

Transform and resulting extinction) obtained without any drain-source biasing ($U_d = 0$), whereas the right panels are the results obtained with a rather strong drain-source biasing at $U_d = 700$ mV. Even if the differences in traces are difficult to see at the first glance they are very well resolved in the experiment because of, aforementioned, very high signal to noise ratio (above $10^5$ in whole presented spectral range). The shadowed regions in Fig. 4 were magnified in the neighbor panels to show the regions where the differences can be observed. First, one can clearly see that the time traces (and spectra) without and with strong biasing (electrical doping) show some well resolved experimentally differences. Second, one can observe that the application of additional drain-source biasing (700 mV) inverts the order of the blue “without carriers” and red with carriers traces suggesting “negative transmission”. This very important point will be systematically investigated and discussed in the next parts of this paper.

Generally, the experimental spectra $T_p$ and $T_{CNP}$ contain information about THz transmission, reflection and absorption. The key parameter that governs the underlying physics in an array of conducting strips is the ratio of the conductivity in the active region to the light velocity: $2\pi\sigma_1/c\sqrt{\epsilon}$ (see Refs. [45, 46] and discussion in Sec. A of the Supplementary material). Here $\epsilon$ is the dielectric constant of the surrounding media.

For the case, when this parameter is small,

$$\frac{2\pi\sigma_1}{c\sqrt{\epsilon}} \ll 1,$$

the transmission coefficient is approximately given by

$$T \approx 1 - A,$$

where $A \ll 1$ is absorption coefficient (see Supplementary Material). The quantity, which is measured in the experimental is proportional to $T^2$, because incoming beam goes through the array of the conducting strips, reflects from the metallic substrate (mirror), and it goes again through the array of the conducting strips. We get

$$T^2 \approx (1 - A)^2 \approx 1 - 2A.$$

As mentioned above the quantity that we interpret in the experiment is $(1 - T_p^2/T_{CNP}^2)$, where $T_{CNP}$ is the spectra registered with whole sample at CNP condition and $T_p^2$ the spectra obtained with only one of the multifinger gates “biased”. Using Eq. 3 one can show that measured quantity is proportional to $2A$. In fact, correction to the transmission coefficient in our experiment is smaller because the size of the beam $S_b$ is bigger than the samples grating area $S_g$. The measured absorption is given by $2A = 2A S_b/S_g$.

The coefficient $A$ can be simply expressed in terms of radiation-induced correction to the dissipation in the system, $\delta P$:

$$A = \frac{\delta P}{S} = \frac{8\pi\delta P}{c\sqrt{\epsilon}E_0^2}.$$  \hspace{1cm} (4)

Here $E_0$ is the amplitude of the incoming radiation and $S = c\sqrt{\epsilon}E_0^2/8\pi$ is the time-averaged radiation Pointing vector for linearly-polarized wave. Using Eq. (3) and Eq. (4), we obtain $(1 - T_p^2/T_{CNP}^2) \approx 2A \propto \delta P$. This is the quantity $2A$ which we plot in all experimental figures. Importantly, this quantity is proportional to the dissipation in the channel—the property, which we will use for theoretical interpretation of our observations.

III. EXPERIMENTAL RESULTS

A. Resonant plasmons at zero drain current conditions

In Fig. 5 we show the results of some systematic measurements with zero drain voltage/current applied to the structures ($U_d = 0$ V).

Comparison of the measurements results for two different polarizations of the THz radiation: parallel and perpendicular to the gate fingers are shown in Figs. 5(a) and 5(b) respectively. This comparison is done for the same electrical doping of cavities ($U_g - U_{CNP} \sim 3$ V). One can see that, in the parallel polarization case, the extinction spectra are characterized by the Drude-like response with a monotonic decrease of the absorption with frequency [46]. In contrast, in the perpendicular polarization, a completely different line shape was observed with a pronounced resonant absorption peak shifting to higher energies for narrower gate fingers.

In the case of the polarisation parallel to the stripes the dissipation is due to the Drude absorption and mea-
FIG. 4. Typical time traces obtained for A-DGG1 sample with (orange) and without gate biasing (green). Left panels (a)-(c) experiments without drain-source biasing ($U_d = 0$), right panels (e)-(g) the results obtained with a strong drain biasing at $U_d = 700$ mV. Panels (b) and (f) are magnification of the shadowed regions in the panels (a) and (e) respectively. One can observe that the drain-source biasing inverses the order of magnitude of orange and green traces in the time traces and spectra row data. Panels (d) and (h) show spectra obtained by Fourier transform for time traces with and without gate biasing, respectively.

FIG. 5. Gate length dependent extinction spectra of the graphene structures for $U_d = 0$ V with incident light polarized parallel (a) and perpendicular (b) to the gates fingers. The measured data points and lines are the fits by the Drude model fit for parallel polarization and the damped oscillator model for perpendicular polarization. Black solid squares are a maximum of experimental resonant plasma frequencies as a function of the wave vector $q = \pi/L_1$. Black straight line through these dots is the best linear fit (right hand scale).

measured extinction spectra can be fitted by using the Drude formula:

$$
\tilde{A} \propto \frac{\gamma_1}{\gamma_1^2 + (\omega - \omega_1^2/\omega)^2} \approx \frac{\gamma_1/4}{\delta \omega^2 + \gamma_1^2/4}. \tag{5}
$$

where $\gamma_1 = 1/\tau_1$ is the plasmon relaxation rate in active area and coefficient in this equation is frequency-independent. The fitted values of the amplitude and plasma frequency varied for each cavity, while the variation of $\tau_1$ was fairly small: $0.12 \sim 0.13$ ps.

Likewise, using the results of the supplementary materials, we fitted the frequency dependencies of the extinction spectra for the perpendicular polarization case by formula of the so-called damped oscillator model [46–49]

$$\tilde{A} \propto \frac{\gamma_1}{\gamma_1^2 + (\omega - \omega_1^2/\omega)^2} \approx \frac{\gamma_1/4}{\delta \omega^2 + \gamma_1^2/4}. \tag{6}
$$

where $\omega_1$ is the resonance frequency in the active region with high conductivity and $\delta \omega = \omega - \omega_1$. The lines in Fig. 5 correspond to theoretical calculations. The calculated curves agree quantitatively very well with the extinction spectra registered in the experiments. It is worth noting that the width of the Drude peak is approximately twice larger as compared to plasmonic resonance in agreement with the theoretical expectations (see Eqs. (5) and (6)).

In Fig. 5(b) we show also the frequency as a function of the wave vector ($q = \pi/L_1$) where $L_1$ is the single finger gate width of the active region with high conductivity (right hand scale). One can see that the resonant frequency scales almost linearly with $q$ determined by the
length of the grating finger, suggesting that in the investigated structures the grating fingers act as independent cavities with 2D gated plasmons with dispersion relation close to $\omega = s_1 q$, with $s_1 \sim 3 \times 10^6$ m/s.

Summarizing, the results from Fig. 5 show that: i) the metallic gates are close enough to 2DEG to grant dispersion very close to linear - strongly gated plasmon dispersion, and ii) it is not the grating period, but the single finger dimension that determines the plasmons wave vectors.

In Fig. 6 we show the results of the study of the plasmon resonances, as a function of the gate voltage (2D carrier density). As in all experiments, the gate voltage-dependent extinction spectra are shown while tuning the only one cavity gate at the time, and keeping all other parts of the sample at the charge neutrality condition. The gate voltage-dependent data show a clear blue-shift and increase of the strength of the absorption peak frequency with increasing carrier density (increasing gate voltage). As shown in Figs. 6(a)-6(e), the fitting curves obtained using damped oscillator model, agree very well with the measured extinction spectra.

Figures 7 shows the gate-voltage dependences of the extracted plasmon frequencies. One can see that a sample with a shorter cavity length has a higher frequency at a given gate voltage. Figure 7 also shows theoretical plasmon frequencies, as functions of the gate voltage calculated as

$$\omega_q = \sqrt{\frac{4e^2 T \ln[2 + 2 \cosh(\varepsilon_F/T)]}{h^2 \varepsilon q[1 + \coth(qd)]}} q,$$

(7)

where $q$ is the plasma wave vector, $d$ is the thickness of the top h-BN layer given in Tab. I, $\varepsilon = 4.5$ is the dielectric constant of h-BN layers [50], $\varepsilon_F$ is the Fermi energy, $T = 300$ K is the temperature. At sufficiently high electron concentration ($\varepsilon_F \gg T$), Eq. (7) simplifies to the well known form [24, 26]:

$$\omega_q = \frac{s q}{\sqrt{qd[1 + \coth(qd)]}}.$$

(8)

Here

$$s = \sqrt{\frac{4e^2 \varepsilon_F d}{\varepsilon h^2}},$$

(9)

is the plasma wave velocity ($s = s_1$ in the active region, and $s = s_2$ in the passive region). The factor $qd[1 + \coth(qd)]$ describes deviation from the linear dispersion and is important when the thickness of the top h-BN layer is comparable with the gate lengths ($L_1$ or $L_2$).

In the calculations we have taken into account the quantum capacitance correction [44], fringing and plasma leakage as predicted by phenomenological model described later in the work and in the supplementary materials. One can see that the theoretical calculations made without any fitting parameters reproduce relatively well
the absolute values and the functional gate voltage dependencies of the plasma frequency for all four cavities, close to \(1/4\) power dependence. The small discrepancies may result from the uncertainty of h-BN layer thickness determination (precision \(\sim 10\%\)) and unknown exact value of the dielectric function. In fact in the literature one can find \(\varepsilon\) in the range from 3.5 to 5 [50–53]. Also in our calculations graphene Fermi velocity was assumed carrier density independent, where in principle it may change in the range, between \(1.3 \times 10^6\) to \(1.0 \times 10^6\) m/s in our experimental conditions [53, 54].

It should be stressed however, that independently of small numerical discrepancies, the observed plasma frequency follows \(1/4\) power dependence on gate voltage (carrier density) typical for plasmons in graphene - as shown by dotted line in Fig. 7 (see also Fig. 6(b) in supplementary materials).

Terahertz plasmon resonances in graphene have been already observed in micro-ribbons [46, 55, 56], rings, disks [47, 48, 57, 58] and grating-gate structures [35]. However, it is worth to mention that: i) here we report the first experimental observations of such plasma standing waves with 100% gate tunable frequency and ii) the plasmon related extinction coefficient going up to 19%.

To summarize this part, the scaling behavior of plasmon frequency versus the plasmon wave vector and the gate voltage shown in Figs. 5, 6 and 7 were successfully calculated using standard physical models of gated 2D plasmons in graphene [24–26].

They clearly confirm the existence of 2D plasmons oscillating at THz frequencies in our graphene/h-BN nanosstructures and allow an unambiguous attribution of the observed resonances to 2D plasmons in the individual fingers of the grating-gate defined cavities C1,..., C4.

![FIG. 7. Gate voltage dependence of the extracted resonance frequency. Continuous lines are results of calculations according Eq. (7) with corrections to fringing and plasma leakage phenomena- see supplementary materials for more details. The dotted line is \(1/4\) power dependence as a function of the gate bias, typical for plasmons in graphene.](image)

B. Plasma resonances in the presence of the dc current

The most important part of the present work concerns the experimental investigations of the influence of the \(dc\) drain-source bias \(U_d\) on the identified earlier THz 2D plasmons resonances.

In Fig. 4 one can see examples of experimental time traces and their Fourier transform for cavity C2 (1 \(\mu\)m) with and without applied drain bias. The results: first \((T_{\text{CNP}}^2)\) with all the gates ‘unbiased’ (set at the CNP) and the second with only C2 cavity gate ‘biased’ at \(\sim 3\) V \(\sim (T'_{\text{CNP}}^2)\) are plotted. Comparing the results with and without drain-source biases one can clearly see that in the range of plasmonic resonances the order of the traces with and without gate biasing is inverted after switching on the drain-source voltage (700 mV). First for zero drain bias, when the cavity C2 is filled with carriers – the outgoing light intensity is lower than the incoming one (plasma resonant absorption). For high drain voltage (700 mV), in contrast, the registered signal is higher when the cavity is filled with carriers (amplification or emission). This is the reason why in the last right panels data plotted as 2A show negative values.

To investigate further this intriguing behavior we performed very systematic measurements versus drain bias for all four cavities. In Fig. 8 we show 3D plot illustrating the typical drain bias dependence of the extinction (C2 plasma cavity). With increase of the \(dc\) current, a strong red shift of the resonant THz plasmon absorption, its complete bleaching, followed by the amplification and blue shift of the resonant plasmon frequency are observed. In Fig. 9 we collect all results in the form of the 2D plots - keeping the colour code as in the Fig. 8. As \(U_d\) increases the absorption peak clearly shifts to lower frequencies along with a noticeable reduction of plasmon resonance amplitude. Then the absorption completely vanishes, as seen for example in the measured extinction spectra for cavity C1, being zero at \(U_d\) \(\sim 45\) mV up to \(U_d\) \(\sim 90\) mV [see Fig. 9(a)]. In this drain voltage range, the plasmonic device C1 becomes perfectly transparent to the incoming THz radiation within the entire experimental bandwidth. It is important to stress that, here we report the first experimental observation of such a strong Doppler shift of plasma resonances in graphene and the transparency behavior over a relatively wide frequency range (0.1 to 3 THz). With increasing \(U_d\) beyond this transparency regime, a resonant “negative absorption” feature appears in the extinction spectra and shifts with increasing drain bias towards higher frequencies (noticeable blue shift).

Throughout all of our experiments, the data from all cavities C1, , ..., C4 were quite similar with the amplification threshold voltage increasing with the cavity length \(L_g\) and important widening of the transparency range (from about 50 mV up to 300 mV).

The results presented in Fig. 9 show some universal behaviour. This is visualised when we present them in re-
scaled mode. In Fig. 10, the vertical axis is the frequency divided by its value at \( U_d = 0 \) and the drain voltage (the horizontal axis) is normalized by the drain voltage \( U_0 \) at which the plasmonic frequency tends to zero (see Fig. 9). Very good quantitative agreement both at low and at high currents can be obtained by fitting experimental curves as

\[
\frac{\omega_2^2(x)}{\omega_1^2} = 1 + \zeta + x^2 - \sqrt{\zeta^2 + 4x^2(1 + \zeta)}
\]  

(10)

where \( \omega_1(x) \) stands for position of the plasmonic resonance, \( x = U_d/U_0 \), \( \omega_1 = \omega_1(0) \) is the experimentally measured value of the plasmonic frequency at zero current. \( U_d \) is the voltage across the structure, \( U_0 \) is the voltage corresponding to onset of the transparency window, and \( \zeta \) is a fitting parameter. The values of \( \zeta \) corresponding to solid lines in Fig. 10 are presented in Table II. The unified universal behavior of frequency at low currents (for \( x < 1 \)) can be obtained by taking in Eq. (10) the limit \( \zeta \to \infty \), which yields

\[
\omega_1(x) = \omega_1 \sqrt{1 - \left( \frac{U_d}{U_0} \right)^2},
\]  

(11)

The similar functional dependence of the plasmonic frequency versus electron drift velocity of 2D electron gas in GaAs/AlGaAs grating gate structures (Eq. 10) was obtained in Ref. [45]. However, the physical model used in this work can not be applied to interpret our results. Validity of the physical model of Ref. [45] for our structures and analogy used to derive Eq. 10 will be discussed in more details in the following section and in the Supplementary materials.

IV. PHENOMENOLOGICAL DESCRIPTION OF THE OBTAINED RESULTS

In phenomenological interpretation we consider a plasmonic crystal composed of two different regions with lengths \( L_1 \) and \( L_2 \), plasma velocities \( s_1, s_2 \), carrier density \( n_1, n_2 \) and drift velocities \( V_1, V_2 \) (see Fig. 11). We assume that \( n_1 \gg n_2 \) and refer region “1” as active and region “2” as passive. The plasma waves velocities \( s_1 \) in the electrically doped active cavities (for gate voltage 3 V) were calculated from experimental data shown in Fig. 7, using Eq. 8. The results are shown in Table II. One can see that for all cavities \( s_1 \sim 3.2 \times 10^8 \) cm/s. Because the plasma velocity is proportional to the
(1/4) power of the carrier density (see Eq. 9), the plasma velocities in CNP passive cavities can be estimated as $s_2 \sim s_1/2 \sim 1.5 \times 10^8$ cm/s. The drift velocities in the electrically doped regions (active regions) $V_1$ are by more then one order of magnitude lower than these in the CNP regions. The ratio $(V_2/V_1 \sim n_1/n_2 \sim 10)$ of the drift velocities stems from the dc current continuity condition $(n_1 V_1 = n_2 V_2)$. We also take into account in this subsection, that damping rates in the passive and active regions can be different. We assume, however, that they have the same order of magnitude: $\gamma_1 \sim \gamma_2$.

Importantly, plasma wave velocities are in all regions higher than Fermi velocity in graphene. As a consequence in our experimental situation the carrier drift velocities in all regions are smaller than the plasma wave velocities.

Therefore the most important theoretical challenge is to find the physical process/mechanism in which the amplification can take place for drift velocities below the plasma wave velocity. In this work, we will develop a phenomenological theory which shows that indeed, amplification is possible for $V_1 < s_1$. This theory shows a possible way of physical interpretation and give universal dependence of the plasmonic frequency on the current in a sense that the only controlling parameter is $V_1/s_1$.

The structure shown in Fig. 11 is illuminated by radiation with the large wavelength, $\lambda \gg L_1$, $\lambda \gg L_2$, so that the electric field of radiation is approximately homogeneous. This field excites plasmonic oscillations in the active regions with the fundamental frequency $\omega_1 = \pi s_1/L_1$ and in the passive regions with the frequency $\omega_2 = \pi s_2/L_2$ (see Fig. 12). In this model, we assume that

$$s_1/L_1 \gg \gamma_1 \sim \gamma_2 \gg s_2/L_2.$$  \hspace{1cm} (12)

In this case, resonances in the passive region strongly overlap. Physically, this means that plasma waves rapidly decay along the passive region, which, in turn, implies that different active regions are disconnected at plasmonic frequencies in an excellent agreement with experimentally observed independence of different active regions. On the other hand, different active regions are connected at zero dc frequency in a sense that dc current flows through the system.

Let us clarify this point in more detail. The oscillations decay into passive region as $\propto \exp(-\gamma \delta x/s_2)$ (here $\delta x$ is coordinate counted from the border between regions). Under the condition (12), this exponent at $\delta x = L_2$ becomes $\propto \exp(-\gamma L_2/s_2) \ll 1$. Hence, the coupling of oscillations in the neighboring active regions is exponentially small and one can reduce the problem to calculation of excitations in a single active strip with proper boundary conditions (BC). Detailed discussion of such approach is

| Cavity Name | C1 | C2 | C3 | C4 |
|-------------|----|----|----|----|
| Structure/sample | A-DGG2 | A-DGG1 | A-DGG2 | A-DGG1 |
| $L_1$ ($\mu$m) | 0.5 | 0.75 | 1.0 | 1.5 |
| $L_2$ ($\mu$m) | 3.5 | 3.0 | 3.0 | 2.25 |
| $U_0$ (mV) | 45 | 69 | 92 | 143 |
| $U_T$ (mV) | 90 | 159 | 267 | 458 |
| $s_1$ ($10^8$ cm/s) | 3.2 | 3.2 | 3.1 | 3.1 |
| $s_2$ ($10^8$ cm/s) | 1.6 | 1.6 | 1.6 | 1.6 |
| $f_1 = \omega_1/2\pi$ (THz) | 2.76 | 1.90 | 1.45 | 0.97 |
| $f_2 = \omega_2/2\pi$ (THz) | 0.11 | 0.23 | 0.24 | 0.32 |
| $\zeta$ | 1.5 | 2.3 | 3.9 | 4.5 |
| $n_1$ ($10^{12}$ cm$^{-2}$) | 2.0 | 2.0 | 2.0 | 2.0 |
| $n_2$ ($10^{12}$ cm$^{-2}$) | 0.1 | 0.1 | 0.1 | 0.1 |

**TABLE II. Properties of different plasma cavities.**

**FIG. 11.** 1D plasmonic crystal driven by dc current in the oscillating field of THz radiation.
presented in the Supplementary material. For zero current, the proper conditions look
\[ \delta j_1(0) = \delta j_1(L_1) = 0, \]  
which means that the current is fixed on both sides of the active region. Such conditions allow us for detailed quantitative description of plasmonic resonances in transmission coefficient.

To check how inequalities Eq. (12) are fulfilled in our experimental situation we have to estimate resonant frequencies in the passive and active plasma cavities. The length of CNP cavities do not vary much and for rough estimations one can take \( L_2 \approx 3 \mu m \) what leads to plasma angular frequency as \( \omega_2 = 2\pi f_2 \approx 1.6 \) THz. One can see that the conditions \( \omega_2 \ll \omega_1 \) and \( (s_1/L_1 \gg s_2/L_2) \) are relatively well fulfilled for all four experimental configurations (active cavities C1,…,C4).

![Fig. 12. Active regions of the plasmonic crystal with large fundamental frequency are separated by passive regions with small frequency](image)

Let us first consider the case of zero current. In this case existence of the two plasma regions has two consequences - the plasma frequency is shifted toward lower frequencies and broadening appears (similar broadening was predicted for a single gate problem in Ref. [59]). These two phenomena are described by following equations (see supplementary materials):
\[ \gamma_{\text{eff}} = \gamma_1 + \frac{2s_1}{L_1} \ln \left( \frac{s_1 + s_2}{s_1 - s_2} \right), \]  
where \( \gamma_1 = 1/\tau_1, \) \( \tau_1 \) is the momentum relaxation time in the active region (in this subsection, we assume that \( \gamma_1 \) can differ from \( \gamma_2 \), but have the same order of magnitude, \( \gamma_1 \sim \gamma_2 \)), so that inequality (12) holds for both \( \gamma_1 \) and \( \gamma_2 \)
\[ \delta \omega_1 = \frac{\pi(\gamma_2 - \gamma_1)s_1s_2}{(s_1^2 - s_2^2) \left[ \pi^2 + \ln^2 \left( \frac{s_1 + s_2}{s_1 - s_2} \right) \right]}, \]  

The effective width of the plasma cavities \( L_1 \) should include also correction related to the fringing effect, that leads to the following correction of the plasmon frequency:
\[ \delta \omega_1^{\text{fr}} = \frac{2d}{L_1} \omega_1. \]  

In Fig. 13, we show the momentum relaxation time \( \tau_1 \) determined from transport measurements together with plasmon relaxation time \( \tau_{\text{eff}} = 1/\gamma_{\text{eff}} \), calculated using Eq. (14). We show also experimental plasmon relaxation time determined from the fit of the resonances (see Fig. 6). One can see that leakage-induced contribution to the damping reduces the relaxation time by approximately one order of magnitude, and calculated relaxation times are in relatively good agreement with experimental data. This explains why widths of plasma resonances are much larger than these predicted from the transport experiments. It should be stressed that in the plasmonic crystal approximation developed here the correction to the resonant frequency (Eq. (15)) is relatively small. Therefore the experiments in the absence of the drain bias can be interpreted taking the standard graphene plasmon description of independent plasmon cavities Eq. (7), (8), (9). For more detailed discussion of plasmons resonances width and corrections to plasma frequency see supplementary materials.

![Fig. 13. Relaxation time as a function of the gate bias. \( \tau_1 \) is the momentum relaxation time in the active region shown as the black diamonds, \( \tau_{\text{eff}} \) is the plasmon relaxation rate calculated using Eq. (14) (solid lines), color dots are experimental \( \tau_{\text{eff}} \).](image)

Next, we calculate dissipation in the channel under the drain biased condition and demonstrate that with increasing the current it decreases and changes sign, which implies amplification. We will discuss the problem in the hydrodynamic regime assuming that electron-electron collisions dominate over other types of the scattering. (For discussion of crossover from hydrodynamic to ballistic regime and of the Cherenkov instability in graphene see Refs. [35, 36].) The suggested mechanism should work for any electronic spectrum, but calculations are more compact and physically transparent for parabolic spectrum characterized by the effective mass \( m \). We thus restrict ourselves to discussion of this case only. Qualitatively, the results are valid also for graphene, where role of \( m \) is played by \( E_F/v_F^2 \). The theoretical model that we develop in this work simplifies if, in addition to inequality (12), we also assume that \( s_1 \gg s_2 \) [this inequality is different from Eq. (12) for \( L_2 \ll L_1 \)]. This condition is not strictly fulfilled and this could be one of the possible reasons of lack of quantitative agreement with experiment.
(see discussion and the supplementary materials).

Since oscillations rapidly decays into passive regions, we consider dissipation in the active region only and put boundary conditions (13) for all active strips. Dissipation per unit area is given by

$$P = \left\langle N \frac{m v^2}{\tau} \right\rangle_{t,x},$$

(17)

where \( \tau \) is the momentum relaxation time, \( N \) is local concentration in the channel and \( v \) is the total velocity given by sum of the constant velocity related to \( dc \) current and the oscillating component [see Ref. \([60]\)]. Equation (17) implies averaging over active and passive regions, where \( N \) and \( \tau \) have different values. However, due to the higher conductivity and resonant excitation, the main contribution comes from the active region, so that we skip contribution of region 2 (in this case, in order to find dissipation averaged over structure as a whole, the dissipation calculated in region 1 should be multiplied by the factor \( L_1/(L_1 + L_2) \)).

We start with the zero-current case. Calculation of dissipation for this case are presented in Supplementary materials. We derived there exact formula for dissipation, Eq. (S21), and demonstrated that it can be with a good precision replaced by simplified equation obtained in the case of damped oscillator model,

$$P_{DO} = \frac{N_1 F_0^2}{2m} \frac{\gamma_1}{\left( \omega - \frac{v_1^2}{\tau_1} \right)^2 + \gamma_1^2},$$

(18)

where \( F_0 = e E_0 \), and \( E_0 \) is the amplitude of the incoming wave, which assumed to be homogeneous is space [see Eq. (S25) of Supplementary materials]. This, in turn, results in Eq. (6) for correction to the transmission coefficient. The latter equation perfectly fits the observed resonances provided that one replaces \( \gamma_1 \) with \( \gamma_{eff} \). This gives additional confirmation of the plasmonic mechanism.

Let us now assume that we apply driving \( dc \) current in the system leading to non-zero electron drift velocity \( V_1 \) in the active region. In this case, the expression for dissipation in the active region should be modified as follows:

$$P_1 = \left\langle N_1 (1 + n_1) \frac{m (V_1 + v_1)^2}{\tau_1} \right\rangle_{t,x},$$

$$\approx \frac{m N_1}{\tau_1} \left( V_1^2 + (v_1^2 + 2V_1 v_1 n_1)_{t,x} \right),$$

(19)

Here, \( v_1 \) and \( n_1 = \delta N_1/N_1 \) are the radiation-induced corrections to drift velocity and normalized concentration, respectively. Hence, radiation-induced correction to dissipation in the active region is given by

$$\delta P_1 = \frac{m N_1}{\tau_1} \left( v_1^2 + 2V_1 v_1 n_1 \right)_{t,x}.$$  

(20)

The second term in this equation arises due to the presence of the current. Remarkably, this term depends on the phase shift between \( n_1 \) and \( v_1 \), and can be negative. This means that Eq. (20) is not positively defined and for special cases \( \delta P \) could become negative. This implies switching from dissipation to amplification. The full theoretical description of the plasmonic crystal with two regions having arbitrary properties is quite cumbersome but can be essentially simplified for the case, when damping of plasmonic oscillations in the passive region is sufficiently large, so that condition Eq. (12) holds.

We use the same approach as we used to study the response at zero current (see Supplementary material). Namely, we will describe the problem phenomenologically, by using the same BC, Eq. (13), as for the zero-current case.

We linearize the hydrodynamic equations and search for solution in the following form \( v_1 = \delta v_1(x)e^{-i\omega t} + \text{h.c.} \) and \( n_1 = \delta n_1(x)e^{-i\omega t} + \text{h.c.} \), where \( \delta v_1(x) \) and \( \delta n_1(x) \) obey

$$\left( V_1 \frac{\partial}{\partial x} - i\omega + \gamma_1 \right) \delta v_1 + s_1 \frac{\partial \delta n_1}{\partial x} = \frac{F_0}{2m},$$

$$\left( V_1 \frac{\partial}{\partial x} - i\omega \right) \delta n_1 + \frac{\partial \delta v_1}{\partial x} = 0.$$  

(21)

(22)

The oscillating correction to the current is given by \( j_1 = \delta j_1(x)e^{-i\omega t} + \text{h.c.} \), where

$$\delta j_1(x) = s_1 \left[ V_1 \delta n_1(x) + \delta v_1(x) \right].$$

(23)

We solve Eqs. (21) and (22) with boundary conditions Eq. (13),

substitute \( v_1 \) and \( n_1 \) into Eq. (20) and average over time and space. Thus obtained solution depends on the current, which is encoded in the drift velocity without radiation, \( V_1 \). Calculations are very similar to conventional calculation of response of a single FET [see Refs. \([9]\) and \([61]\)]. The dependence of total averaged dissipation \( \delta P = [L_1/(L_1 + L_2)]\delta P_1 \) on \( \omega \) for different \( V_1/s_1 \) is shown in Fig. 14. Qualitatively, this dependence is similar to the experimentally observed one. Indeed, as seen, with increasing the current, the frequency and the amplitude of the plasmonic resonance decrease. For sufficiently large current, \( \delta P \) changes sign, which implies
amplification. It is worth noting, however, that there appear some new peaks in the amplification, which are not seen in the experiment. General analytical expression for dissipation simplifies to the form allowing an analytical solution if the quality factor of resonance is high, so that its frequency is much larger than damping. For this case, assuming \( L_1 = L_2 = L \), we get

\[
\delta P = \frac{F_0^2 N_1}{m \pi^2} \frac{\gamma_{\text{eff}}}{\omega - \omega_1(x)^2 + \gamma_{\text{eff}}^2/4^2} A(x),
\]

where

\[
\omega_1(x) = \frac{\pi s_1}{L_1} (1 - x^2),
\]

\[
A(x) = \frac{(1 - 3x^2) \cos^2(\pi x/2)}{(1 - x^2)^2},
\]

is the frequency and amplitude of the resonance, respectively, which depend on \( x = \sqrt{\frac{\gamma}{1/s}} \). Even if this analytical solution was obtained for conditions somehow far from the experimental ones it captures main characteristics. It shows both the Doppler shift and a possibility of amplification. Indeed, the frequency \( \omega_1(x) \) depends on current [see Eq. (25)], while amplitude \( A(x) \) turns to zero at certain value of \( x = x_0 = 1/\sqrt{3} \), which is the dissipation-amplification transition point as shown in Fig. 15.

![Figure 15](image_url)

**FIG. 15.** The amplitude of the dissipation as a function of \( x = \sqrt{\frac{\gamma}{1/s}} \). Dissipation-amplification transition corresponds to \( x = x_0 = 1/\sqrt{3} \).

V. DISCUSSION

Let us briefly discuss other known mechanisms of amplification and instability. First of all, it is instructive to compare our theoretical model with the theory developed in Ref. [45] for grating-gated 2DEG in GaAs. This theory predicted that with the increase of the current (drift velocity) one may observe strong Doppler shift of plasma resonance followed by 100% transparency range and amplification. These predictions are very similar to our observations. Moreover, in principle, Eq. (10) can be obtained from Eq. (52) of [45] (see Supplementary material). However, in spite of the very good description by Eq. (10), the physical interpretation using the models of Ref. [45] is not possible. Actually, this model describes different physical situation, namely, in contrast to our system, the electron concentration is homogeneous in the absence of radiation. The second reason, which is more important, is the fact, that to explain amplification within this model one uses Cherenkov-like effects that requires drift velocity in the active region to be higher than the plasma wave velocity. Indeed, it turns out that \( x = U_d/U_0 = \sqrt{\gamma_{\text{eff}}}/s_1 \), so that amplification is only possible for \( \sqrt{\gamma_{\text{eff}}}/s_1 \). At the same time, our experimentally determined plasma velocity \( s_1 \approx 3.2 \times 10^8 \text{ cm/s} \) is higher than the Fermi velocity in graphene. The latter represents the upper limit for drift velocity. Therefor in our experiments \( \sqrt{\gamma_{\text{eff}}}/s_1 \).

In Ref. [62], the instability in gated plasmonic crystal was predicted for the case when drift velocity exceeds plasma wave velocity in the passive region. This mechanism, however, was developed for ideal system with very low momentum relaxation rate \( \gamma_2 \ll s_2/L \) and can not be applied for our system, where experimental conditions correspond to inequality (12). Also, since experimental data show strong evidence that the threshold value of velocity, corresponding to onset of amplification is smaller than the plasma wave velocity in the active region, we can exclude other mechanisms related to Cherenkov-like plasmon instability [45, 54, 63–65].

There are only a few theoretical predictions of plasma instability for drift velocities smaller than the plasma velocity. Most known is the instability based on amplified plasmon reflection at the cavity boundaries [8]. This instability is called Dyakonov-Shur instability. The evolution of \( dc \) response with the current increasing and approaching to instability threshold was described in Ref. [61]. It was predicted that current leads to decrease of the damping rate in the channel, which turns to zero on the instability threshold [see Eq. (59) and (60) of Ref. [61]]. Evidently, such a scenario is not realized in our experiment, where width of resonances do not essentially depend on the \( dc \) current.

Although one can not fully exclude mechanism based on so-called transit time instability (it is caused by exchange of excitation between active regions due to carriers moving with saturated drift velocity in passive regions) considered in Refs [66–68], there are some experimental facts against this mechanism. Most importantly, this mechanism implies strong connection between active regions via exchange of excitation travelling through passive regions. At the same time, the experimental data show that active regions behave like independent plasmonic resonators.

Importantly, in our work we do not suggest instability mechanism. Instead, we predict that current-driven amplification of the incoming radiation is possible without developing of any kind of plasma instabilities. Then,
limitations $\mathcal{V}_1 > s_1$ (needed for Cherenkov instability) and/or built-in asymmetry of the device (needed for well known Dyakonov-Shur instability [8]) are lifted. We demonstrate theoretically that amplification is possible when velocity in the active region is smaller than the plasma wave velocity ($\mathcal{V}_1 < s_1$).

The reason for the absence of transparency frequency interval in our theory is also clear. This interval arises due to formation of current-tunable bands in the density-modulation-induced plasma crystal. Since we were able to develop simplified theory only in the regime of strong plasmonic mismatch between active and passive regions, $s_1 \gg s_2$, the allowed bands of the crystal are quite narrow $\Delta \omega \sim s_2/L_2$ [see Eq. (2) in Ref. [62]], so that they are smeared due to the strong dissipation [see Eq. (12)]. On the other hand, signature of plasmonic bands is seen in our experiment as is illustrated in Fig. 9, where transparency interval in the transmission spectra is clearly seen.

Our most important experimental observations are two effects induced by dc current: (i) very strong redshift in absorption and (ii) THz radiation amplification and gain. One can see that Eq. (24) provides a good phenomenological description of these experimental results. We notice, that in contrast with theory of Ref. [45], our theory predicts monotonous decrease of the plasmonic peak amplitude in a very good agreement with experimental data. We also predict that the width of the plasmonic resonance is current-independent, which is the case in the experiment. Also, it is worth to stress that the model gives an explanation of the origin of the width in the condition of zero current. It clearly shows that the main contribution to the line widths (plasmon relaxation rate) is coming from the plasmon leakage phenomena [69].

However, in spite of very good qualitative description, the quantitative description is still lacking. For example, Eq. (24) predicts that exactly at the threshold, the resonance frequency is only by a factor $2/3$ smaller than the frequency at the zero current. At the same time, experiment shows much smaller values of the threshold frequency (see Fig. 9). Also we are not able to explain the transparency frequency intervals or to derive Eq. (10), that fits so well frequency versus normalise drain voltage dependence.

Additionally the theoretical model developed in this work assumes that $s_1 \gg s_2$. This condition is not strictly fulfilled and this is one of the possible reasons of lack of quantitative agreement with experiment. The full theory of plasmonic crystal with arbitrary relation between plasma wave velocities in the active and passive regions is an open unsolved problem. At the present stage, its solution following the methods shown in this work leads to very cumbersome equations that can be analyzed only numerically. It is the subject of ongoing work and its results will be published elsewhere [70].

VI. CONCLUSION

We have investigated resonant plasmons and influence of current/voltage bias on plasmonic instabilities in gated 2D graphene systems. Without applied current, we observed 2D plasmons in the grating-gate cavities oscillating at THz frequencies. We have demonstrated that these plasmons follow the scaling laws of plasmon dispersions and the gate voltage dependencies predicted by standard physical models developed for gated 2D systems. Using different linear polarizations we demonstrated that the width of plasmonic peak (observed for polarization perpendicular to the grating) is twice smaller than the peak of the Drude resonance (observed for polarization parallel to the grating) in the excellent agreement with the theoretical predictions. Also, experiment showed strong asymmetry of plasmonic resonances. The shape of such asymmetric resonances is perfectly fitted by damped oscillator model which, in turn, is in very good agreement with exact calculation of plasmonic dissipation peaks in the active regions. The broadening of the resonances' width could also be explained as mainly due to plasmon leakage between the active and passive regions.

In the experiments with applied current, we demonstrated that increase of current leads to a strong Doppler-like shift of the plasma resonances, followed by full transparency and THz light amplification phenomena. We developed a phenomenological model based on a simple physical assumption: we assume that active regions are connected at zero frequency but disconnected at THz frequencies. In fact, this is the only assumption which allowed us to explain qualitatively all the scope of experimental data at zero current. For non-zero current, it provides a simplified model that describes basic tendencies together with qualitative description of physical mechanism of amplification taking place for drift velocities smaller then plasma velocities.

Although the presented model captures basic physics of the observed phenomena it predicts critical velocity and threshold plasma resonance frequency of dissipation-amplification transition significantly higher than these estimated from experimental data. Also it does not describe the transparency frequency intervals or origin of Eq. (10). Therefore, our results show a clear need for more advanced theoretical approach establishing a new challenge for theory of plasmonic crystals.

Summarizing, we list our main results:

- We have demonstrated current-driven excitation of Dirac plasmons in grating-gate graphene-channel transistor nanostructures successfully leading to THz radiation amplification even at room temperature.

- We have observed and described in detail unique graphene plasmon THz dynamics at room temperature, driven by dc current. Specifically, we demonstrated that the plasmon resonances are redshifted,
undergo complete bleaching followed by the amplification with more than 9% of gain and blue-shift of the resonant plasmon frequency in the THz range.

- We have theoretically modeled the 2D Dirac plasmon system to explain the origin of the amplification phenomena due to influence of dc current on the phase shift between the change of carrier density and drift velocity induced by the THz radiation. This radiation-induced correction to dissipation is sensitive to this phase shift and with increasing dc current the dissipation becomes negative, which clearly reproduce the experimentally observed phenomena switching from dissipation to amplification.

- We have demonstrated that in contrast to previous theoretical predictions - transition to amplification regime occurs at drift velocities smaller than the plasma wave velocity. We theoretically found a critical value of drift velocity corresponding to the onset of amplification.

- We have discussed a leakage of plasma waves in the plasmonic crystals, which can limit plasmonic quality factor even in very clean structures. This explains why in all previous experiments on THz detection in gating gate structures, the quality factor of plasmonic resonances was much smaller than the one expected from the transport measurements of the momentum relaxation rate. We derived analytical expression for leakage-dominated damping which shows that leakage can be controlled by gates covering passive regions of the plasmonic crystal.

Concluding, we would like to stress that all results presented in this work were obtained at room temperature ~ 300 K. Therefore, in spite of lacking complete quantitative theoretical description, demonstrated here – for the first time – THz amplification, paves the way towards new tunable compact THz amplifiers and future solid-state based THz plasmonic technology.

ACKNOWLEDGEMENTS

We thank V. Ryzhii, S. Mikhailov, D. Svintsov, K. Maussang, M.S. Shur, M. Dyakonov for many useful discussions.

The work was supported by JSPS KAKENHI (#16H06361, #16K14243, and #18H05331), Japan; the International Research Agendas program of the Foundation for Polish Science co-financed by the European Union under the European Regional Development Fund for CENTERA (No. MAB/2018/9) and by the Foundation for Polish Science through the TEAM project POIR.04.04.00-00-3D76/16 (TEAM/2016-3/25). The work of V.Yu.K. was supported by Russian Foundation of Basic Research (Grant No. 20-02-00490) and by Foundation for the Advancement of Theoretical Physics and Mathematics BASIS. The work of V.V.P. was carried out within the framework of the state task.

[1] A. V. Chaplik, Sov. Phys. JETP 35, 395 (1972).
[2] S. J. Allen, D. C. Tsui, and R. A. Logan, Phys. Rev. Letters 38, 980 (1977).
[3] T. Theis, J. Kotthaus, and P. Stiles, Solid State Communications 24, 273 (1977).
[4] D. Tsui, S. Allen, R. Logan, A. Kamgar, and S. Copper-Smith, Surface Science 73, 419 (1978).
[5] T. Theis, J. Kotthaus, and P. Stiles, Solid State Communications 26, 603 (1978).
[6] T. N. Theis, Surface Science 98, 515 (1980).
[7] D. Tsui, E. Gornik, and R. Logan, Solid State Communications 35, 875 (1980).
[8] M. Dyakonov and M. Shur, Phys. Rev. Letters 71, 2465 (1993).
[9] M. Dyakonov and M. Shur, IEEE Transactions on Electron Devices 43, 380 (1996).
[10] K. Hirakawa, K. Yamanaka, M. Grayson, and D. C. Tsui, Applied Physics Letters 67, 2326 (1995).
[11] W. Knap, J. Lusakowski, T. Parenty, S. Bollaert, A. Cappy, V. V. Popov, and M. S. Shur, Applied Physics Letters 84, 2331 (2004).
[12] J. Lusakowski, W. Knap, N. Dyakonova, L. Varani, J. Mateos, T. Gonzalez, Y. Roelens, S. Bollaert, A. Cappy, and K. Karpierz, Journal of Applied Physics 97, 064307 (2005).
[13] N. Dyakonova, F. Teppe, J. Lusakowski, W. Knap, M. Levinshtein, A. P. Dmitriev, M. S. Shur, S. Bollaert, and A. Cappy, Journal of Applied Physics 97, 114313 (2005).
[14] T. Otsuji, Y. M. Meziani, T. Nishimura, T. Suemitsu, W. Knap, E. Sano, T. Asano, and V. V. Popov, Journal of Physics: Condensed Matter 20, 384206 (2008).
[15] S. Boubanga-Tombet, F. Teppe, J. Torres, A. El Moutaouakil, D. Coquillat, N. Dyakonova, C. Consejo, P. Arcade, P. Nouvel, H. Marinchio, T. Laurent, C. Palermo, A. Penarier, T. Otsuji, L. Varani, and W. Knap, Applied Physics Letters 97, 262108 (2010).
[16] A. El Fatimy, N. Dyakonova, Y. Meziani, T. Otsuji, W. Knap, S. Vandenbrouk, K. Madjour, D. Théron, C. Gauqiére, M. A. Poisson, S. Delage, P. Prystawko, and C. Skierbiszewski, Journal of Applied Physics 107, 24504 (2010).
[17] W. Knap, S. Nadar, H. Videlier, S. Boubanga-Tombet, D. Coquillat, N. Dyakonova, F. Teppe, K. Karpierz, J. Lusakowski, M. Sakowicz, I. Kasalynas, D. Selita, G. Valusis, T. Otsuji, Y. Meziani, A. El Fatimy, S. Vandenbrouk, K. Madjour, D. Théron, and C. Gauqiére, Journal of Infrared, Millimeter, and Terahertz Waves 32, 618 (2011).
[18] T. Otsuji, T. Watanabe, S. A. Boubanga Tombet,
[63] O. Matov, O. Meshkov, O. Polischuk, and V. Popov, Physica A: Statistical Mechanics and its Applications 241, 409 (1997).

[64] M. V. Krasheninnikov and A. V. Chaplik, Zhurnal Eksperimental’noi i Teoreticheskoi Fiziki 79, 555 (1980).

[65] G. R. Aizin, J. Mikalopas, and M. Shur, Physical Review B 93, 195315 (2016).

[66] V. Ryzhii, A. Satou, and M. S. Shur, Physica Status Solidi (A) 202, R113 (2005).

[67] V. Ryzhii, A. Satou, and M. S. Shur, IEICE 89, 1012 (2006).

[68] Y. Koseki, V. Ryzhii, T. Otsuji, V. V. Popov, and A. Satou, Physical Review B 93, 245408 (2016).

[69] V. V. Popov, O. V. Polischuk, W. Knap, and A. El Fatimy, Applied Physics Letters 93, 263503 (2008).

[70] I. V. Gorbenko, V. Y. Kachorovskii, and W. Knap, "Current-driven optical response of plasmonic crystal: from dissipation to amplification", private communication (2019).
A. Transmission through array of conducting strips

To analyze transmission through array of conducting strips one can use theory developed in Ref. [1] (see also Ref. [2]). In the case when parameter $2\pi\sigma_1/c^2\sqrt{\epsilon}$ is small \((2\pi\sigma_1/c^2\sqrt{\epsilon} \ll 1)\) the formulas for \(T, R\) and \(A\) (transmission, reflection, and absorption coefficients, respectively) can be obtained by expansion over this parameter. Indeed, as seen from Eq. (33) of Ref. [1] this parameter coincides with the ratio of the radiative decay \(\Gamma\) to the momentum relaxation rate:

\[
\frac{\Gamma}{\gamma_1} = \frac{2\pi\sigma}{c^2\sqrt{\epsilon}}. \tag{S1}
\]

Since, we assumed that inequality Eq. (1) is fulfilled, we find \(\Gamma \ll \gamma_1\). Then, we find that expressions for \(T, R\) and \(A\) [see Eqs. (40a), (40b) and (40c) of Ref. [1]] simplify. In vicinity of plasmonic resonance, we find:

\[
T = 1 - \frac{\pi\sigma_1}{c^2\sqrt{\epsilon}} \gamma_1^2 \delta\omega^2 + \left(\frac{\gamma_1}{2}\right)^2, \tag{S2}
\]

\[
R = \left(\frac{\pi\sigma_1}{c^2\sqrt{\epsilon}}\right)^2 \delta\omega^2 + \left(\frac{\gamma_1}{2}\right)^2, \tag{S3}
\]

\[
A = \frac{\pi\sigma_1}{c^2\sqrt{\epsilon}} \delta\omega^2 + \left(\frac{\gamma_1}{2}\right)^2. \tag{S4}
\]

Here \(\delta\omega = \omega - \omega_1\) and \(\omega_1\) is the frequency of the plasmonic resonance. As seen from these equations,

\[
R \sim \frac{\pi\sigma_1}{c^2\sqrt{\epsilon}} A \ll A \ll 1, \quad 1 - T \approx A \ll 1, \tag{S5}
\]

provided that condition (1) is fulfilled. This is in good agreement with Fig. 4 of Ref. [1]. Indeed, as seen from this figure, conducting strips, created by gate voltage, change transmission coefficient of the system as a whole by a small correction (much smaller than unity). This implies that transmission through graphene strip is close to unity, which is the case in our experiment. This yields experimental proof that \(2\pi\sigma_1/c \ll 1\). Then, reflection is small, \(\propto (\pi\sigma_1/c^2\sqrt{\epsilon})^2\), and can be neglected, while \(T \approx 1 - A\). Using the Drude formula for conductivity \(\sigma_1 = e^2N_1/m\gamma_1\), one can rewrite Eq. (S4) as follows:

\[
A = \frac{\pi e^2 N_1}{mc^2\sqrt{\epsilon}} \gamma_1 \frac{\gamma_1}{2} \left(\frac{\gamma_1}{2}\right)^2. \tag{S6}
\]

Comparing this equation with Eq. (S26), we find

\[
A \approx \frac{8\pi\delta P}{c\sqrt{\epsilon}E_0^2}. \tag{S6}
\]

where \(\delta P\) is radiation-induced correction to the dissipation. Equations (3) and (S6) allow us to reduce the problem of calculation of transmission coefficient to the calculation of \(\delta P\).

B. Dissipation in the strip of width \(L\)

As we discussed in the main text, active regions are disconnected at THz frequency due to the inequality Eq. (12) The effect of passive regions can be taken into account by the following replacing in the final equations: \(\gamma_1 \rightarrow \gamma_{\text{eff}}\). Hence, we can consider a single active strip. The hydrodynamic equations for velocity \(\mathbf{v}\) and concentration \(N\) in conventional 2D system with parabolic spectrum with mass \(m\) read (for graphene away from the Dirac point, calculations presented below should also work if one replaces \(m\) with \(E_F/\sqrt{\epsilon}\)):

\[
\frac{\partial \mathbf{v}_1}{\partial t} + (\mathbf{v}_1 \nabla) \mathbf{v}_1 + \gamma_1 \mathbf{v}_1 = -\frac{\partial \Phi}{\partial r} + \frac{F_0}{m} \cos(\omega t), \tag{S7}
\]

\[
\frac{\partial N_1}{\partial t} + \nabla (N_1 \mathbf{v}_1) = 0. \tag{S8}
\]
Here $\gamma_1 = 1/\tau_1$, $F_0 = eE_0$, and $E_0$ is the electric field of the radiation,

$$\Phi = \Phi(r, t) = \int d^2r'K(r - r')\delta N_1(r', t)$$  \hspace{1cm} (S9)

is the potential created by redistribution of the electron density, and $\delta N_1(r, t) = N_1(r, t) - N_1$ is the deviation of the concentration from its equilibrium value. Function $K(r)$ depends on geometry of the problem. In particular, for gated 2D system, $K(r) = (s_1^2/N_1)\delta(r)$, where $N_1$ is the homogeneous electron concentration in the absence of radiation and $s_1$ is the plasma wave velocity in the active region.

C. Drude dissipation

Dissipation in the system depends on the orientation of the field. For $F_0 \parallel e_y$, both $n_1$ and $v_1$ do not depend on coordinates. Hence $n = 0$, $v_x = 0$

$$v_{1y} = F_0 \frac{1}{2m} \frac{1}{-i\omega + \gamma_1} e^{-i\omega t} + \text{h.c.}$$  \hspace{1cm} (S13)

From Eqs. (17) and (S13), we get the Drude dissipation: per unit area:

$$P_{Dr} = \frac{N_1 F_0^2}{2m} \frac{\gamma_1}{\omega^2 + \gamma_1^2}$$  \hspace{1cm} (S14)

D. Plasmon-assisted dissipation

For $F_0 \parallel e_x$, external field excites plasmons in the strip. In this case, $v_y = 0$ and $v_x$ and $n$ can be written as

$$v_{1x} = \delta v_1(x)e^{-i\omega t} + \delta v_1^*(x)e^{i\omega t},$$  \hspace{1cm} (S15)

$$n_1 = \delta n_1(x)e^{-i\omega t} + \delta n_1^*(x)e^{i\omega t}.$$  \hspace{1cm} (S16)

The amplitudes $\delta v_1$ and $\delta n_1$ obey

$$(-i\omega + \gamma_1)\delta v_1 + s_1^2 \frac{\partial \delta n_1}{\partial x} = \frac{F_0}{2m},$$  \hspace{1cm} (S17)

$$-i\omega \delta n_1 + \frac{\partial \delta v_1}{\partial x} = 0.$$  \hspace{1cm} (S18)

These equations should be solved with the boundary conditions $\delta v_1(x = 0) = \delta v_1(x = L_1) = 0$ (absence of current at the edges of the system). Direct calculation yields for the velocity amplitude

$$\delta v_1 = \frac{F_0}{m(\gamma_1 - i\omega)} \frac{\sin(kx/2)\sin[k(x - L_1)/2]}{\cos(kL_1/2)},$$  \hspace{1cm} (S19)

where

$$k = \frac{\sqrt{\omega(\omega + i\gamma_1)}}{s_1}.$$  \hspace{1cm} (S20)

Now, dissipation is inhomogeneous, so that we need to average both over $t$ and over $x$. The averaged plasmon-assisted dissipation per unit area looks (we skip contribution of passive regions)
\[ P_{pl} = N_1 L_1 \left( \frac{m \gamma_1^2}{\tau_1} \right)_{x,t} = 2 N_1 m \gamma_1 \int_0^{L_1} |\delta v_1(x)|^2 \]
\[ \approx 2 N_1 F_0^2 \frac{L_1}{4 m \gamma_1^2 (\delta \omega)^2 + \gamma_1/4}, \quad (S23) \]

The plasmonic resonances are encoded in the factor \( \cos(k L_1/2) \cos(k^* L_1/2) \) which turns to zero (for \( \gamma_1 = 0 \)) at the resonant plasmonic frequencies

\[ \omega_1(N) = \omega_1(1 + 2N), \quad N = 0, 1, 2, \ldots \quad (S22) \]

where

\[ \omega_1 = \frac{\pi s_1}{L_1} \]

is the fundamental plasmonic frequency in the active region. It is worth noting that homogeneous external field excites only odd modes with \( 1 + 2N \), while even modes \( 2N \) remain silent although they are present in the full set of resonant excitations. Although Eq. (S21) looks quite complicated, it dramatically simplifies in the resonant regime:

\[ \delta \omega = \omega - \omega_1(N) \ll \omega, \quad \omega_1 \gg \gamma_1. \]

In this case, for \( N = 0 \) (fundamental plasma mode), after some algebra, we find

\[ P_{pl} = \frac{N_1 F_0^2}{\pi^2 m} \frac{\gamma_1}{(\delta \omega)^2 + (\gamma_1/2)^2} \quad (S23) \]

Comparing the amplitudes of the Drude and plasmonic-assisted dissipation, we get

\[ \frac{P_{DO}(\omega = 0)}{P_{pl}(\delta \omega = 0)} = \frac{\pi^2}{8} \approx 1.23. \quad (S24) \]

Let us compare now Eqs. (S14), (S23), and (S24) with Fig. 5 of Ref. [1]. First of all, from Fig. 5 of Ref. [1] we see that amplitude of the Drude peak coincides with the amplitude of dissipation in the plasmonic resonance. This observation is in a reasonably good agreement with Eq. (S24). Second observation related to the widths of the Drude and plasmonic peaks. As seen from Fig. 5 of Ref. [1] the width of the Drude peak is twice larger than the width of the plasmonic resonance. This observation is in excellent agreement with Eqs. (S14), (S23). Hence, we come to the conclusion that peaks in the transmission coefficient observed in Ref. [1] for different polarizations of the radiation can be well interpreted as the Drude (parallel polarization) and plasmonic (perpendicular) peaks. A more detailed analysis of this statement is presented in the next subsection. As one can see from Fig. 5 of Ref. [1], for small resonance frequencies, when quality factor of plasmonic resonance is smaller that unity, the peaks are asymmetrical functions of frequency in contrast to symmetrical Lorentz peak given by Eq. (S23). To describe dissipation for small quality factors one can use exact equation Eq. (S21). The result of calculations are shown in Fig. 2. As seen, asymmetry increases with decreasing the quality factor \( \omega_1/\gamma_1 \) in a very good agreement with experimental data. All curves are normalized to the value

\[ P_m = \frac{4 N_1 F_0^2}{\pi^2 m \gamma_1}, \]

which gives the maximal value of the dissipation for the case of high quality factor [see Eq. (S23)].

**FIG. 2.** Dissipation calculated by the use of Eq. (S21) for different values of \( \omega_1/\gamma_1 = 0.25, 0.5, 0.75, 1.0, 1.25, 1.5 \) increasing from left to right.

One can compare results obtained with the use of the exact formula Eq. (S21) with the results obtained within damped oscillator model. Dissipation in this model is given by the Drude dissipation formula with the plasmonic-induced frequency shift

\[ P_{DO} = \frac{N_1 F_0^2}{2m} \frac{\gamma_1}{\left(\omega - \frac{\omega_0^2}{\omega}\right)^2 + \gamma_1^2}. \quad (S25) \]

For the case of high quality factors, \( \omega_0 \gg \gamma \), this equation simplifies even more

\[ P_{DO} \approx \frac{N_0 F_0^2}{8m} \frac{\gamma_1}{(\delta \omega)^2 + \gamma_1^2/4}, \quad (S26) \]
where $\delta \omega = \omega - \omega_1$. Although resonance equation (S26) is symmetric function of $\delta \omega$, the general equation (S25) shows a strongly asymmetrical peak. Equation (S25) describes very well the shape of the resonance at the fundamental frequency as illustrated in Fig. 3 for the same values of quality factors as in Fig. 2. There are however some differences between exact formula (S21) and Eq. (S25). First of all, the amplitude of the dissipation is slightly different: $P_{DO}^{N}/P_m = \pi^2/8$. Second, the shape of the curves also slightly differ and the difference increases with decreasing the quality factor. We illustrate last statement in the Fig. 4, where we compare relative dissipation obtained from two equations for low quality factor $\omega_1/\gamma_1 = 0.5$. We also note that damped plasmon model does not allow to calculate amplitude of plasmonic mode with $N \neq 0$. By using Eq. (S21), we find in the resonance approximation close to $N$-th resonance

$$P_{pl}^N = \frac{N_1 F_0^2}{\pi^4 m (1 + 2N)^2} \frac{\gamma_1}{(\delta \omega)^2 + (\gamma_1)^2/4},$$

(S27)

where $\delta \omega = \omega - \omega_N$. Hence,

$$\frac{P_{pl}^N(\delta \omega = 0)}{P_{pl}^N(\delta \omega = 0)} = \frac{1}{(1 + 2N)^2}$$

(S28)

Before closing this section, we note that Eq. (S21) describes all resonances. In Fig. 5 we illustrate it for $\omega_1/\gamma_1 = 5$.

FIG. 3. Dissipation calculated by the use of Eq. (S25) for different values of $\omega_1/\gamma_1 = 0.25, 0.5, 0.75, 1.0, 1.25, 1.5$ increasing from left to right.

E. Leakage of plasmons from active to passive region

Above, we assumed that plasmons are localized in the active regions. Let us now briefly discuss the leakage of plasmonic excitation which arise due to finite coupling with the passive regions (such a leakage was previously predicted for a single gate problem in Ref. [3]). We assume that damping rates in the active and passive regions are different, $\gamma_1 < \gamma_2$, which is the case in our experiment. We also assume that $\gamma_2 L_2 \gg s_2$, which means that different active regions are effectively disconnected, so that without loss of generality, we can put $L_2 = \infty$. Then, we have three regions: active region, $|x| < L_1/2$ and two passive regions: $-\infty < x < -L_1/2$ and $L_1/2 < x < \infty$. We use the following BC between regions, $s^2_1 v_1 = s^2_2 v_2$, $s^2_1 n_1 = s^2_2 n_2$ (for $x = \pm L_1/2$) which represent current and energy conservation at the boundary (see more detailed discussion in Ref. [4]). Plasmonic oscillations in the passive region can be described by the same equations as Eq. (S17) and (S18) with the obvious change of parameters. In the active region there are two solutions with wave vectors $\pm \sqrt{\omega (\omega + i \gamma_1)/s_1}$, while in the passive regions, there is a solution with the wave vector $\sqrt{\omega (\omega + i \gamma_2)/s_2}$, for $x > L_1/2$ and solution with wave vector $-\sqrt{\omega (\omega + i \gamma_2)/s_2}$, for $x < L_1/2$. Hence, we have four independent solutions and, therefore, four independent coefficients, which can be written as a vector $a = (a_1, a_2, a_3, a_4)$. This vector should be found from four BC. We skip this quite standard calculation and limit ourselves to discussion of general idea and presenting the final result. Using BC one can write equation for $a$ in the matrix form: $A a = c$, where vector $c \propto E_0$ describes coupling to external field. Writing solution of this equation as $a = A^{-1} c$, we find that resonances in our system corresponds to poles of function $1/det.A$. Standard calculation

FIG. 4. Comparison of resonance shape found from Eq. (S21) and Eq. (S25) for $\omega_1/\gamma_1 = 0.5$.

FIG. 5. Several resonances for $\omega_1/\gamma_1 = 5$. Dashed line corresponds to Eq. (S28)
yields (we skip all irrelevant coefficients) \( \det \hat{A} \propto \exp \left[ i L_1 \sqrt{\omega(\omega + i \gamma_1)/s_1} - (u_1 + u_2)^2/(u_1 - u_2)^2 \right] \), where \( u_{1,2} = s_{1,2}/\sqrt{1 + i \gamma_{1,2}/\omega} \). Assuming that \( \gamma_{1,2} \ll \omega \) and expanding this equation over \( \delta \omega = \omega - \omega_1 \) we find expressions for effective damping and resonance frequency correction:

\[
\gamma_{\text{eff}} = \gamma_1 + \frac{2s_1}{L_1} \ln \left( \frac{s_1 + s_2}{s_1 - s_2} \right), \quad (S29)
\]

\[
\omega_{\text{eff}}^2 = \omega_1 - \frac{\pi(\gamma_2 - \gamma_1)s_1s_2}{(s_1^2 - s_2^2) \left[ \pi^2 + \ln^2 \left( \frac{1 + s_2}{s_1 + s_2} \right) \right]}, \quad (S30)
\]

The effective width of the plasma cavities \( L_1 \) should include also correction related to the fringing effect, that leads to the following correction of the plasmon frequency:

\[
\delta \omega_1 = \frac{2d}{L_1}, \quad (S31)
\]

Second term in the r.h.s. of Eq. (S29) accounts for damping of oscillation caused by leakage effect, while the second terms in the r.h.s. of Eq. (S30) yields small correction to resonance frequency caused by coupling between active and passive regions. (To avoid confusion, we notice that Eqs. (S29) and (S30) are not valid for \( s_2 \to s_1 \). First of all, we assumed that \( (\omega_{\text{eff}}^2 - \omega_1)/\omega_1 \ll 1 \). Also, at \( s_2 \to s_1 \) the width of the resonance becomes much larger than \( \omega_1 \), so that the resonance disappear).

Expressions for \( \gamma_{\text{eff}} \) and \( \omega_{\text{eff}}^2 \) essentially simplify for \( s_2 \ll s_1 \):

\[
\gamma_{\text{eff}} = \gamma_1 + \frac{4s_2}{L_1}, \quad (S32)
\]

\[
\omega_{\text{eff}}^2 = \omega_1 - \frac{(\gamma_2 - \gamma_1)s_2}{\pi s_1}, \quad (S33)
\]

We notice that leakage-induced contribution to the damping is large, which explains why width of resonance is much larger than \( \gamma_1 \) extracted from the transport measurements. On the other hand, negative correction to the real part of the frequency is small but can be responsible for small deviation of the experimentally observed resonance frequency from the theoretical value obtained within the model of an isolated active strip.

Numerical calculations of the plasma frequency versus a gate voltage for the cavity C1 are shown in Fig. 6(a). One can see that calculated values are slightly higher than experimental ones. The black line represents calculations according the Eq. (S29) and the red line is Eq. (S30). One can see that more advanced description (Eq. (S30)) gives correction to plasma frequency only in the region of small carrier densities where Fermi level is comparable to the thermal energy \( kT \). The blue line is calculated according to Eq. (7) with additional correction due to fringing effect approximated by Eq. (S31). Finally the green line is calculated taking into account additionally plasmon leakage phenomena as described in by Eq. (S33). One can see that the plasmon leakage correction to the resonant plasma frequency is relatively small. In Fig. 6(b) the same results are shown in log-log scale demonstrating typical for graphene \( 1/4 \) power dependence. One can see that the theoretical calculations made without any fitting parameters reproduce relatively well the absolute values and the functional frequency versus gate voltage dependency. As already mentioned in the main text the small discrepancies may result from the uncertainty of h-BN layer thickness determination (precision 10%) and unknown exact value of the dielectric function. In fact, in the literature, one can find this value in the range from 3.5 to 5 [5, 6]. It should be stressed however, that independently of small numerical discrepancies, the observed plasma frequency follows closely \( 1/4 \) power dependence on gate voltage (carrier density) typical for plasmons in graphene [7] - as shown by dotted line in Fig. 6(b).

![Graph](attachment:image.png)
F. Frequency of plasmonic excitations in the grating-gate structure

In this subsection, we demonstrate how to derive Eq. (10) from Eq. (52) of Ref. [1]. To this end, one should first determine $\omega_1$ from this equation (by putting drift velocity in the electron liquid to zero) and then rewrite this equation in terms of $\omega_1$:

\[
\frac{\omega_1(x)^2}{\omega_1^2} = \frac{1 - \Delta/2}{1 - \Delta} + x^2 - \sqrt{\frac{\Delta^2}{(1 - \Delta)^2} + 4x^2\frac{1 - \Delta/2}{1 - \Delta}}, \tag{S34}
\]

where $\omega_1(x)$ is the resonance frequency of the plasmon oscillations as a function of and $x = qV_1/\omega_1$, where $q$ is the plasmon wave-vector. Since we demonstrated experimentally that $\omega_1 = s_1q$, we find $x = V_1/s_1$. Equation Eq. (S34) transforms into Eq. (10) if we take

\[
\Delta = \frac{2\zeta}{1 + 2\zeta}. \tag{S35}
\]

Assuming that $V_1 \propto U_d$ we arrive at Eq. (10). It is worth stressing again that the onset of the transparency band corresponds to $V_1 = s_1$, so that amplification starts for $V_1 > s_1$.

[1] S. A. Mikhailov, Physical Review B 58, 1517 (1998).
[2] L. Ju, B. Geng, J. Horng, C. Girit, M. Martin, Z. Hao, H. A. Bechtel, X. Liang, A. Zettl, Y. R. Shen, and F. Wang, Nature Nanotechnology 6, 630 (2011).
[3] V. V. Popov, O. V. Polischuk, W. Knap, and A. El Fatimy, Applied Physics Letters 93, 263503 (2008).
[4] V. Y. Kachorovskii and M. S. Shur, Applied Physics Letters 100, 232108 (2012).
[5] A. Laturia, M. L. Van de Put, and W. G. Vandenberghe, npj 2D Materials and Applications 2, 6 (2018).
[6] G. L. Yu, R. Jalil, B. Belle, A. S. Mayorov, P. Blake, F. Schedin, S. V. Morozov, L. A. Ponomarenko, F. Chiappini, S. Wiedmann, U. Zeitler, M. I. Katsnelson, A. K. Geim, K. S. Novoselov, and D. C. Elias, Proceedings of the National Academy of Sciences 110, 3282 LP (2013).
[7] V. Ryzhii, A. Satou, and T. Otsuji, Journal of Applied Physics 101, 024509 (2007).