LETTERS
https://doi.org/10.1038/s41566-017-0054-7

nature photonics

Broadband gate-tunable terahertz plasmons in graphene heterostructures

Baicheng Yao[1,2,7*], Yuan Liu[3,4], Shu-Wei Huang[1,8], Chanyeol Choi[1], Zhenda Xie[1,9], Jaime Flor Flores[1], Yu Wu[2], Mingbin Yu[5,10], Dim-Lee Kwong[5,11], Yu Huang[3,4], Yunjiang Rao[2], Xiangfeng Duan[4,6*] and Chee Wei Wong[1*]

Graphene, a unique two-dimensional material comprising carbon in a honeycomb lattice[1], has brought breakthroughs across electronics, mechanics and thermal transport, driven by the quasiparticle Dirac fermions obeying a linear dispersion[2,3]. Here, we demonstrate a counter-pumped all-optical difference frequency process to coherently generate and control terahertz plasmons in atomic-layer graphene with octave-level tunability and high efficiency. We leverage the inherent surface asymmetry of graphene for strong second-order nonlinear polarizability[4,5], which, together with tight plasmon field confinement, enables a robust difference frequency signal at terahertz frequencies. The counter-pumped resonant process on graphene uniquely achieves both energy and momentum conservation. Consequently, we demonstrate a dual-layer graphene heterostructure with terahertz charge- and gate-tunability over an octave, from 4.7 THz to 9.4 THz, bounded only by the pump amplifier optical bandwidth. Theoretical modelling supports our single-volt-level gate tuning and optical-bandwidth-bounded 4.7 THz phase-matching measurements through the random phase approximation, with phonon coupling, saturable absorption and below the Landau damping, to predict and understand graphene plasmon physics.

The discovery of graphene has spurred dramatic advances ranging from condensed matter physics and materials science to physical electronics, mechanics and thermal processes. In optics[6,7], the additional chiral symmetry of the Dirac fermion quasiparticles of graphene[8] enables optical conductivity defined only by the fine structure constant $\pi\alpha$ (ref. [9]), one that is remarkably charge-density-tunable[10,11] and with broadband nonlinearities[12–15]. The collective oscillations of the two-dimensional correlated quasiparticles in graphene[16] naturally make for a fascinating cross-disciplinary field in graphene plasmonics[17], with applications ranging from tight-field-enhanced modulators, detectors, lasers and polarizers to biochemical sensors[18–22]. Unlike conventional noble metal plasmons, graphene plasmons are dominant in the terahertz and far-infrared frequencies[23]. To excite and detect these plasmons, specialized techniques such as resonant scattering nanoscale antennae near-wfield microscopy or micro- and nanoscale scattering arrays have been pursued, albeit still using terahertz/far-infrared sources[24–28]. Recently, nonlinear optical processes, only with free-space experiments, have proven especially effective in generating graphene plasmons with efficiencies up to 10^{-5} (refs [4,5]). However, it has been challenging to generate, detect and control on-chip graphene plasmons all-optically, a key step towards planar integration and next-generation high-density optoelectronics.

Terahertz generation has recently been revisited in a number of studies for imaging, spectroscopy and communications[29]. Wide terahertz tunability can provide new grounds for broadband stand-off spectroscopy and wavelength-agile ultrahigh-bandwidth communications, but tunability in terahertz materials has been limited so far (Supplementary Table 1). Here, we demonstrate experimentally the charge- and gate-tunability of terahertz plasmons over an octave, from 4.7 THz to 9.4 THz, bounded only by the pump amplifier optical bandwidth. Through the surface asymmetry of dual-layer graphene heterostructures and tight plasmon field confinement, we leverage the intrinsically strong second-order nonlinear polarizability $\chi^{(2)}$ for difference-frequency coherent terahertz generation. We implement a chip-integrated counter-pumped resonant process for frequency- and phase-matching over an octave of the full-scale terahertz carrier frequency. Our designed heterostructure achieves widely tunable terahertz generation via gating at the single volt level, matching our conductivity models and numerical predictions.

Figure 1a shows the graphene on silicon nitride waveguide (GSiNW) architecture studied in this work. The GSiNW has a bottom atomic layer of graphene connecting the drain and source contacts, a layer of alumina functioning as a thin dielectric barrier, and a second atomic layer of top graphene connecting the gate. The silicon nitride waveguide is 1 μm wide and 725 nm in height, and the graphene–Al_2O_3–graphene hybrid heterostructure is assembled with direct contact to the nitride core, enabling effective light–graphene interaction along the ~80 μm waveguide overlap region. The waveguide input–output regions are tapered for effective on/off-chip coupling. Details of the nanofabrication are provided in Supplementary Section 3. The graphene layers serve simultaneously

[1]Fang Lu Mesoscopic Optics and Quantum Electronics Laboratory, University of California, Los Angeles, CA, USA. [2]Key Laboratory of Optical Fiber Sensing and Communications (Education Ministry of China), University of Electronic Science and Technology of China, Chengdu, China. [3]Department of Materials Science and Engineering, University of California, Los Angeles, CA, USA. [4]California Nanosystems Institute, University of California, Los Angeles, CA, USA. [5]Institute of Microelectronics, Singapore, Singapore. [6]Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, USA. Present addresses: [7]Cambridge Graphene Center, University of Cambridge, Cambridge, UK. [8]Department of Electrical, Computer, and Energy Engineering, University of Colorado Boulder, Boulder, CO, USA. [9]National Laboratory of Solid State Microstructures and School of Electronic Science and Engineering, Nanjing University, Nanjing, China. [10]Shanghai Institute of Microsystem and Information Technology, and Shanghai Industrial Technology Research Institute, Shanghai, China. [11]Institute for Infocomm Research, Singapore, Singapore. *e-mail: yaobaicheng@uestc.edu.cn; xduan@chem.ucla.edu; cheewei.wong@ucla.edu

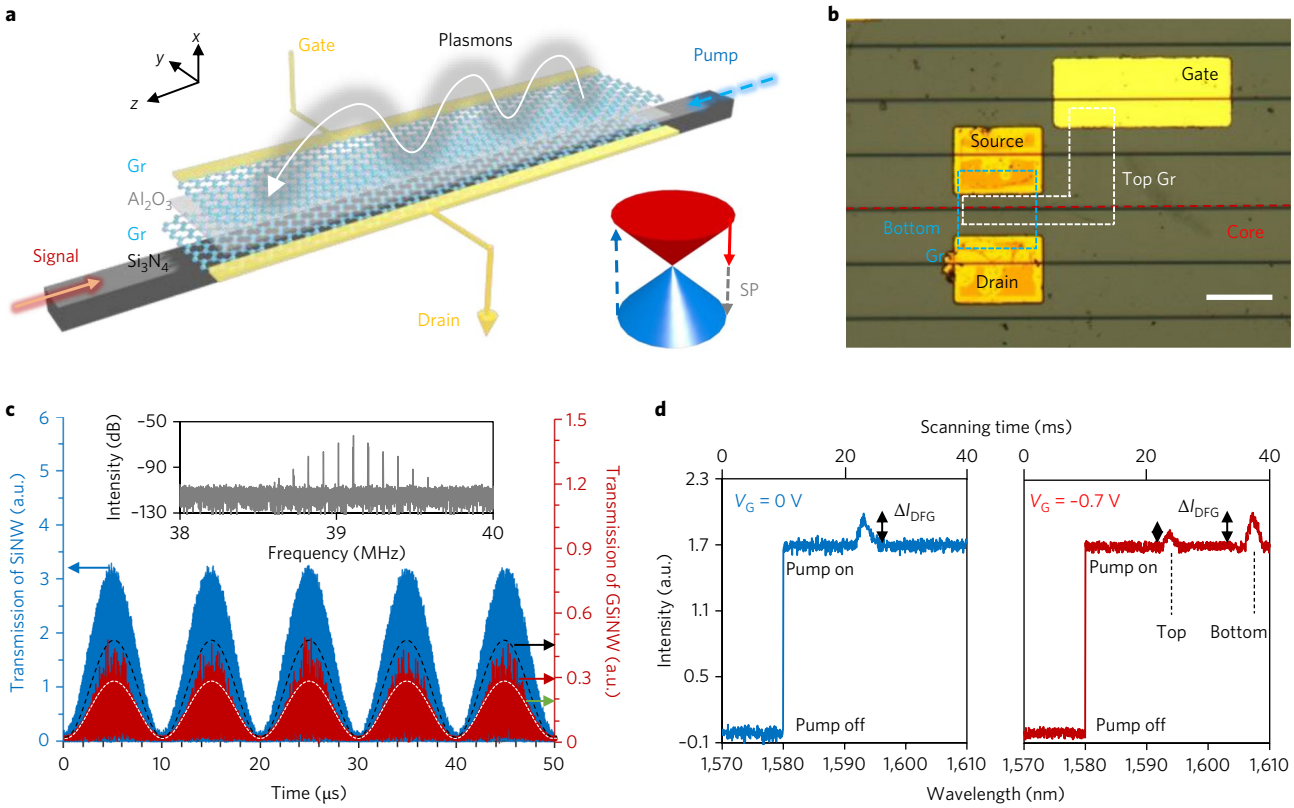

Fig. 1 | Generating and controlling terahertz plasmons in graphene heterostructures via counter-pumped all-optical nonlinear processes. a, Schematic of the dual-layer graphene hybrid for difference frequency generation (DFG) via counter-pumped surface $\chi^{(2)}$ nonlinearity. Dual-layer graphene is deposited onto a silicon nitride waveguide core (GSiNW) with an Al_2O_3 spacer. Inset: Dirac cone structure of the DFG process. **b**, Top-view microscope image of the GSiNW. Bottom and top graphene layers are indicated by blue and white dashed boxes, respectively. Dark left-to-right horizontal lines are the Si_3N_4 waveguides (the selected waveguide is indicated by a red dashed line), orange rectangles are the source and drain contacts, and the gate contact on the surface is bright yellow. Scale bar, 50 μm. **c**, Measured optical transmissions. Blue shaded curve: SiNW without graphene layers. Red shaded curve: GSiNW. White and black dashed curves: modelled linear transmissions. Inset: Modulated pump RF spectrum, with the 39.1 MHz peak from the mode-locked pulse train, and 100 kHz sideband harmonics from the sinusoidal modulation. **d**, Measured DFG-based signal enhancement in the optical spectra. With f_s scanning, DFG induces an additional 0.3 a.u. measured enhancement, as marked by 'ΔI_{DFG}'. When $V_G = 0$ V (left), a distinct peak is observed at 1,593.2 nm. When $V_G = -0.7$ V (right), two enhanced DFG peaks are observed separately (1,593.7 nm and 1,607.2 nm), corresponding to the top- and bottom-layer graphene plasmons, respectively.

as active electrodes, the second-order nonlinearity medium and the nanoscale plasmon waveguides. The surface asymmetry of graphene has an effective second-order nonlinear polarizability $\chi_{eff}^{(2)}$ described approximately as

$$\chi_{eff}^{(2)} = \frac{e^3}{4\pi\hbar^2}\frac{1}{k_{SP}\sqrt{f_s f_p}}\left[\frac{\pi}{2} + \arctan\left(\frac{2\pi\sqrt{f_s f_p} - 2\nu_F k_F}{\gamma}\right)\right] \quad (1)$$

where f_s, f_p are the frequencies of the pump and signal laser, $e = -1.6\times10^{-19}$ C is the unit charge, $\hbar$ is the reduced Plank's constant, γ is the scattering rate of graphene, ν_F is the Fermi velocity, $\hbar k_F = \hbar(2m_e E_F)^{1/2}$ is the Fermi momentum, and $\hbar k_{SP}$ is the counter-pumped phase-matched momentum. The detailed derivation is shown in Supplementary Section 2. The pump and signal lasers are launched into the GSiNW in opposite directions, as shown in Fig. 1.

The on-chip difference frequency generation (DFG) process for terahertz plasmon generation is shown in the inset to Fig. 1a and in Fig. 2a. Driven by the $\chi_{eff}^{(2)}$ polarizability, the energy of one pump photon is distributed into the lower-energy signal photon and a plasmon. In the GSiNW, the photonic modes propagate along the Si_3N_4 core in opposite directions while the plasmonic mode co-propagates along the graphene interface in the same direction as the pump. Our counter-pumped nonlinear phase matching scheme[30] satisfies both energy conservation ($f_{SP} + f_s = f_p$) and momentum conservation ($k_{SP} = k_p + k_s$), where f_{SP}, f_s and f_p are the plasmon, signal and pump frequencies, respectively. $\hbar k_{SP}$, $\hbar k_s$ and $\hbar k_p$ are the momenta of the plasmon, signal and pump, respectively. Based on the optical dispersion $k = 2\pi fn/c$, the phase-matching condition is achieved when $f_s n_s - f_{SP} n_{SP} = -f_p n_p$, where n_p, n_s and n_{SP} are the effective indices of the plasmon, signal and pump, respectively. In addition, we note that both the optical pump and signal wavelength modes propagating along the graphene–Al_2O_3–graphene–Si_3N_4 heterostructure are transverse magnetic (TM) polarized. Optical TM polarization enables strong evanescent field interactions with the graphene layers, and both being TM enables DFG plasmon generation.

The surface plasmon polariton frequency f_{SP} is determined by the Fermi-level-based graphene dispersion, with a plasmon frequency each for the top and bottom graphene layers. First consider the case of zero gate voltage. In our dual-layer graphene–Al_2O_3–graphene capacitor at zero gate voltage, both the top- and bottom-layer graphene are intrinsically positively charged (p-doped; top Fermi level $E_{T0} \approx$ bottom Fermi level $E_{B0} \approx -50$ meV) due to carrier trapping. With the small (30 nm) interlayer distance between the top- and bottom-layer graphene, the plasmon modes weakly couple to form two hybrid modes—symmetric

and antisymmetric. In the low-frequency regime, the dispersions of the symmetric and antisymmetric modes are consequently described as

$$f_{\mathrm{sym}} = \frac{1}{2\pi}\sqrt{\frac{2e^2}{\varepsilon}(E_{\mathrm{T}}+E_{\mathrm{B}})k_{\mathrm{SP}}} \quad (2)$$

$$f_{\mathrm{asym}} = \frac{1}{2\pi}\sqrt{\frac{4e^2E_{\mathrm{T}}E_{\mathrm{B}}d}{\varepsilon(E_{\mathrm{T}}+E_{\mathrm{B}})}}k_{\mathrm{SP}} \quad (3)$$

Here, ε is the background permittivity and d is the dielectric layer thickness. The symmetric mode is observed in our measurements while the antisymmetric mode is strongly damped and hence hard to observe experimentally. Detailed numerical calculations are shown in Supplementary Section 2.

Next we consider the case of applied gate voltages V_{G}. The Fermi levels of the top and bottom graphene are significantly and oppositely tuned, with $|E_{\mathrm{T}}|=|E_{\mathrm{T0}}+E_{\mathrm{G}}|$ and $|E_{\mathrm{B}}|=|E_{\mathrm{B0}}-E_{\mathrm{G}}|$, where E_{G} $(=\hbar|v_{\mathrm{F}}|(\pi N)^{-1/2})$ is the quasi-Fermi level determined by the gate-injected electron density N and Fermi velocity v_{F}. This modulates the plasmon mode dispersion—and hence DFG phase matching—with an order-of-magnitude larger modulation than the mode-splitting frequency change at zero voltage. The difference in the top and bottom graphene Fermi levels (when gated) leads to different top and bottom plasmon frequencies, and thus negligible interlayer plasmon coupling. Figure 1b presents a top-view optical micrograph of the nanofabricated source–drain and gate electrodes on the graphene–Al_2O_3–graphene hybrid heterostructure. The top and bottom graphene layers are denoted by white and blue dashed boxes, respectively, the selected core waveguide is denoted by a red dashed line, and the yellow areas show the gold contact electrodes (with the top gate brighter than the source and drain electrodes because it is patterned above the Al_2O_3 dielectric).

The DFG experimental set-up consists of a mode-locked picosecond pump pulse at a repetition rate of 39.1 MHz launched into the GSiNW from the left, with an amplified continuous-wave (c.w.) signal counter-launched from the right, both in TM polarizations. The detailed experimental set-up is provided in Supplementary Section 4. Figure 1c shows an example time-domain modulated transmission, with ~200 W pump peak power and at 1,532 nm. With graphene saturable absorption, the broadband transmission of the high-peak-power pulsed pump is ~2.1 dB higher than under a c.w. pump of the same average power. With the modulated pump pulses together with a counter-pumped c.w. signal, the resulting observed transmission envelope waveform is sinusoidal with saturable absorption-induced modulation. The modelled saturable absorption transmission and modulation are also shown in the black dashed curves of Fig. 1c.

As noted in Fig. 1a, we search for plasmon generation experimentally by monitoring the signal intensity as a function of the swept c.w. signal frequency, with the pulsed pump laser switched on. With the presence of the plasmon in the DFG phase-matching process, the signal intensity will rise as shown in Fig. 1d, left (ΔI_{DFG}). In this case, the pump laser is fixed at 1,532 nm (195.82 THz) and the c.w. signal laser is swept from 1,570 nm to 1,610 nm (191.08 THz to 186.34 THz) at a scanning rate of 1 nm ms^{-1}. To directly detect this DFG plasmon signal over noise, we implemented a 100 kHz modulation on the mode-locked picosecond laser, with lock-in filtering, amplification and balanced detection. In Fig. 1d, left, the measurements were carried out at $V_{\mathrm{G}}=0$ V. The plasmon is detected when the signal photon is at 1,593.2 nm (188.4 THz) in this case—this corresponds to a plasmon frequency f_{SP} of 7.5 THz. The pump on–off intensity contrast ratio is measured to be ~1.7 a.u. (with the pump off lock-in signal referenced to zero), which arises from the residual saturable absorption modulation as mentioned above[21]. With the presence of the DFG plasmon, an additional 0.3 a.u. peak intensity contrast (ΔI_{DFG}) is observed. In Fig. 1d, right, when $V_{\mathrm{G}}=-0.7$ V we observe two enhanced peaks at 1,593.7 nm (188.2 THz) and 1,607.2 nm (186.7 THz). This corresponds to a top and bottom graphene plasmon frequency at ~7.6 THz and 9.2 THz, respectively.

With the gate voltage applied, the Fermi level is tuned from E_{F} to E_{F}' and f_{SP} changes to f_{SP}'. Correspondingly, the enhanced signal f_{s} shifts to $f_{\mathrm{s}}'=f_{\mathrm{p}}-f_{\mathrm{SP}}'$. We measure the correlation of E_{F} with V_{G} through I_{SD}–V_{G} measurements, as shown in Fig. 2b. Here, I_{SD} is the source–drain current and R_{SD} is the source–drain resistance. In our chip, when tuning V_{G} up to ±4 V, I_{SD} changes from 1.29 μA to 1.47 μA; correspondingly, R_{SD} is in the range of 6.9–7.8 kΩ. (At $V_{\mathrm{G}}=0$, $E_{\mathrm{T0}}=E_{\mathrm{B0}}=-50$ meV due to intrinsic doping.) When close to the Dirac point ($V_{\mathrm{Dirac}}=0.25$ V), graphene has the highest sheet resistance. Accordingly, $|E_{\mathrm{F}}|$ is estimated to be tuned in the range 0 to ~270 meV. When V_{G} approaches 0.25 V (−0.25 V), E_{F} of the bottom (top) graphene reaches close to the Dirac point (Supplementary Section 4.4). Furthermore we note that, when E_{F} changes, the effective second-order nonlinear susceptibility $\chi_{\mathrm{eff}}^{(2)}$ of the graphene also changes, as shown in equation (1). Consequently, the plasmon intensity (I_{SP}) and signal intensity enhancement (ΔI_{DFG}) would also be gate-tuned.

With the gate tunability of E_{F}, we observe tuning of the graphene terahertz plasmon signal, as shown in Fig. 2c. In the top panel of Fig. 2c, when V_{G} increases from −0.7 V to −0.3 V, the DFG enhanced signal peak of the bottom layer ($\lambda_{\mathrm{s,Bot}}$) blueshifts from 1,607.2 to 1,601.3 nm, with intensity increasing from 0.28 to 0.37 a.u (at $V_{\mathrm{G}}=-0.4$ V), then decreasing back to ~0.32 (at $V_{\mathrm{G}}=-0.3$ V). Simultaneously $\lambda_{\mathrm{s,Top}}$ and the intensity of the top layer decreases significantly. In the bottom panel of Fig. 2c, as we further increase V_{G} from −0.2 V to 0.2 V, the signal peak generated by the bottom-layer graphene plasmons blueshifts from 1,598.6 to 1,578.1 nm, with intensity decreasing from 0.35 to 0.09 a.u. However, the signal peak generated by the top-layer graphene plasmons begins to reappear from −0.05 V and then redshifts to 1,601.6 nm with intensity increasing to 0.11 a.u. In Fig. 2c, the measured ΔI_{DFG} of bottom (top) layer graphene is 0.37 a.u. at $V_{\mathrm{G}}\approx-0.4$ V (0.11 a.u. at $V_{\mathrm{G}}\approx0.4$ V). The ΔI_{DFG} of the top layer graphene is approximately three times weaker than the ΔI_{DFG} of the bottom layer, because the top-layer graphene is farther from the waveguide core, where the evanescent field overlapping of the pump and signal are weaker.

Figure 2d analyses the observed gate tuning of the terahertz graphene plasmons. Blue circles and red diamonds are measured results and solid curves are theoretical fittings. With the bottom- and top-layer graphene having the same carrier densities (for example, n_{electron} on the bottom layer equals n_{hole} on the top layer), they have symmetrical curves. Limited by the L-band amplifier spectral window and the ±0.04 a.u. noise, plasmons with f_{SP} higher than 9.4 THz or lower than 4.7 THz are difficult to determine rigorously. We note that our demonstrated plasmon tuning range from 4.7 to 9.4 THz already spans over an octave, the full scale of the carrier plasmon frequency. Based on the measured results, Fig. 2d also provides the tunable effective index n_{SP} and the tunable plasmon wavelength λ_{SP}, by making use of the dispersion relationship $n_{\mathrm{SP}}=(f_{\mathrm{p}}n_{\mathrm{p}}+f_{\mathrm{s}}n_{\mathrm{s}})/(f_{\mathrm{p}}-f_{\mathrm{s}})$ and $\lambda_{\mathrm{SP}}=c/(f_{\mathrm{SP}}n_{\mathrm{SP}})=c/(f_{\mathrm{p}}n_{\mathrm{p}}+f_{\mathrm{s}}n_{\mathrm{s}})$. Here f_{p} is fixed at 195.8 THz. For the bottom-layer graphene when V_{G} approaches V_{Dirac}, n_{SP} increases from ~69 to 116; correspondingly, λ_{SP} changes in the range 460–466 nm during gate voltage modulation. The top-layer graphene has a symmetrical n_{SP}–V_{G} measured dependence. This result supports that a lower Fermi level (closer to the Dirac point) could bring better plasmonic confinement.

To further understand gate-tunable plasmon generation, we investigated its phase-matching conditions (Fig. 3). To generate the terahertz graphene plasmons, the phase-matching conditions

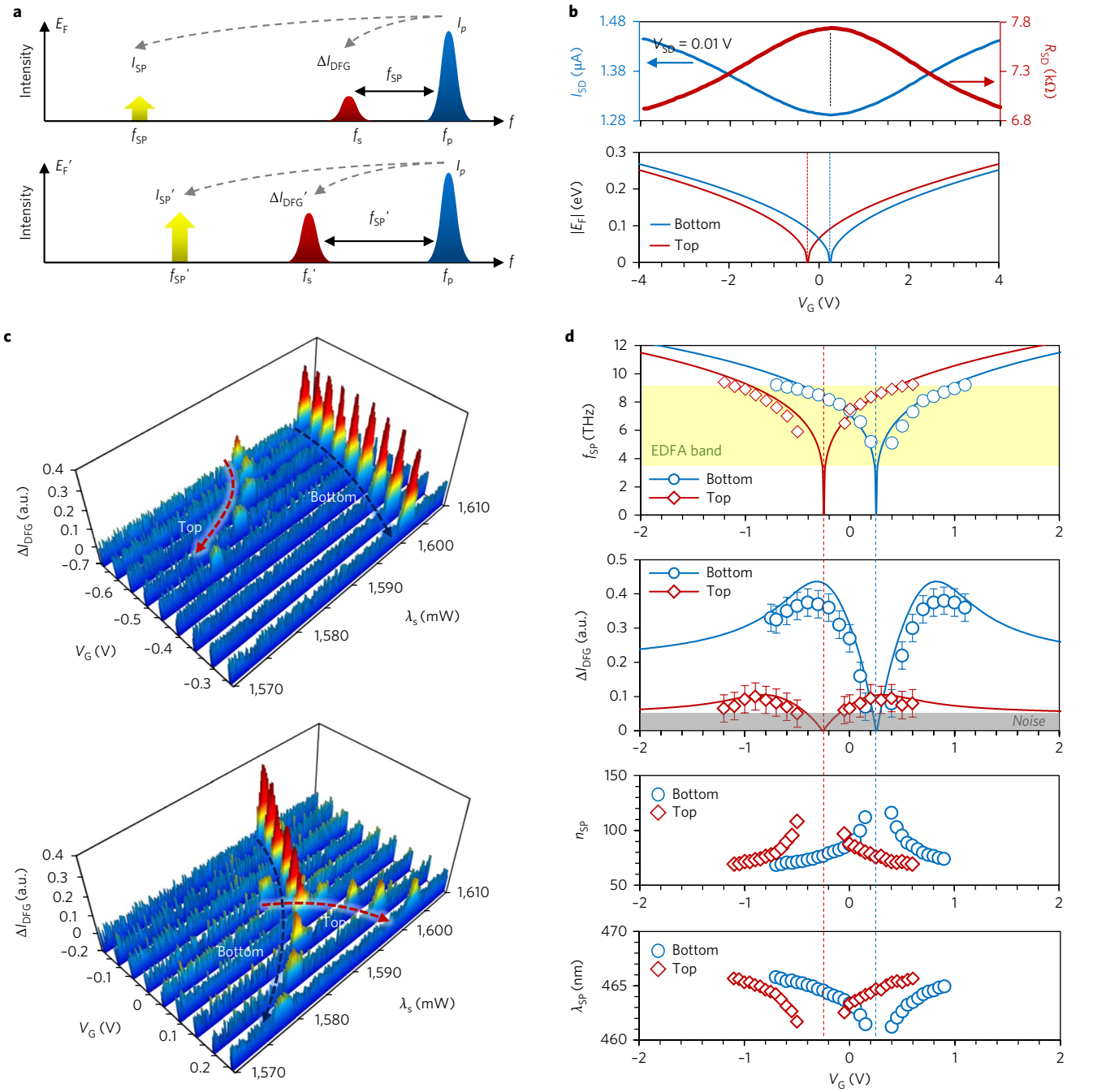

Fig. 2 | Observation and gate tunability of the DFG graphene plasmons. a, Nonlinear process of the DFG. Under different tuning of graphene E_F, both the intensity and frequency of the plasmons could be tuned, which corresponds to a nonlinear enhancement on the signal intensity in the 1,570–1,610 nm wavelengths (*L* band). **b**, Top: Measured I_{SD}–V_G correlation and sheet resistance of the GSiNW (blue dots) under a fixed V_{SD} of 10 mV. Bottom: Derived Fermi level of the bottom-layer graphene (blue) and top-layer graphene (red), based on the measured I_{SD}–V_G correlation. **c**, Measured spectra of ΔI_{DFG}. The V_G mapping step is 50 mV, with the bottom- and top-layer graphene DFG peaks marked by blue and red arrows, respectively. **d**, Gate-tunable parameters of the DFG plasmons. From top to bottom: observed frequency f_{SP}, intensity ΔI_{DFG}, effective index n_{SP} and wavelength λ_{SP}, as a function of gate voltage V_G. In **d**, solid curves are theoretical fittings, the band limitation of our *L*-band erbium-doped fibre amplifier (EDFA) (1,570–1,610 nm) is marked by the yellow region and the background noise up to approximately ±0.04 a.u. is marked by a grey area and error bars.

of the counter-pumped DFG and dispersion of the plasmonic modes must be satisfied simultaneously. By using random phase approximation (RPA), we map the graphene plasmon dispersion at $V_G = 0$ V in Fig. 3a,b; here, both the top- and bottom-layer graphene have the same $|E_F|$. The photon–electron interaction loss L_{SP} is normalized. In the regime where the graphene f_{SP} is much lower than the Landau damping, the dispersion of graphene plasmons $k_{SP}(f_{SP})$ behaves approximately as a quadratic function ($k_{SP} \propto f_{SP}^2$). In the f_{SP}–k_{SP} band-structure map, the phase-matching condition of the DFG could be written as $(c/2\pi)k_{SP} = -f_{SP}n_s + f_p(n_p + n_s)$, shown as grey solid lines (near-vertical lines) in Fig. 3a,b. Here, n_s and n_p are the effective indices of the pump and signal. Hence, in Fig. 3a,b,

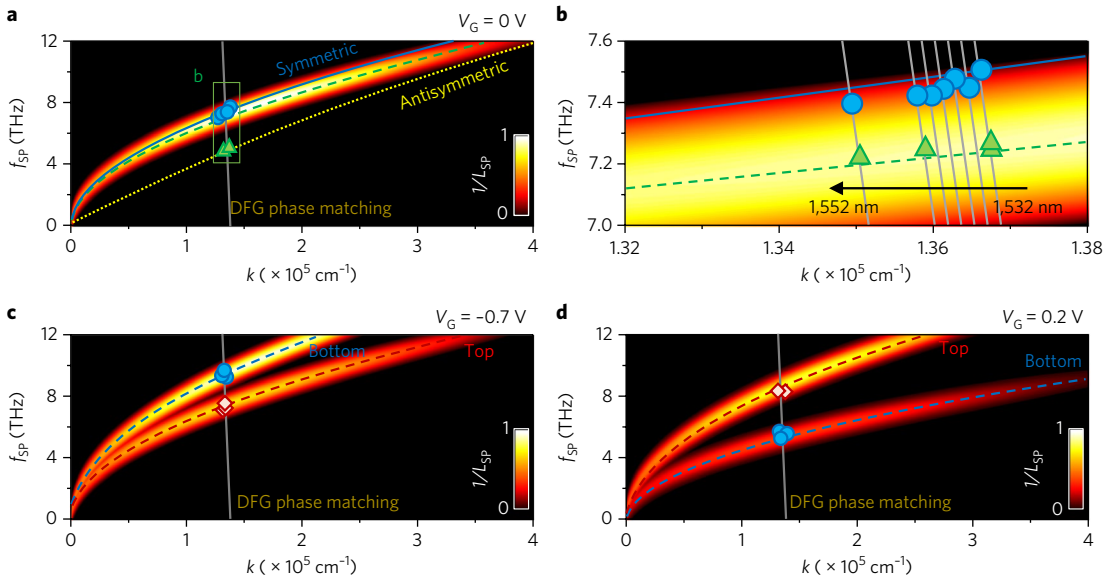

Fig. 3 | Counter-pumped phase-matching conditions. a, Counter-pumped phase-matching condition meets the graphene plasmonic dispersion, for $V_G = 0$ V. Green dashed curve: calculated dispersion of each graphene atomic layer. Blue solid curve and yellow dotted curve: symmetric and antisymmetric mode dispersions, respectively, when interlayer plasmonic coupling is taken into consideration. Top-down grey solid line: DFG phase-matching condition. Blue circles and green triangles: measured data points from Fig. 2 and Supplementary Fig. 5.3, respectively. **b**, Zoom-in of **a**. Grey solid lines: phase-matching cases on tuning λ_p from 1,532 nm to 1,552 nm, with f_{SP} from 7.5 THz to 7.4 THz. **c**,**d**, Under $V_G = -0.7$ V and 0.2 V, respectively, dispersions of the bottom-layer graphene (blue dashed curves) and the top-layer graphene (red dashed curves) are tuned. Measured DFG frequencies on the bottom and top graphene layers are marked as blue circles and red diamonds, respectively. Note that the change in polarity of V_G swaps the bottom- and top-layer dispersion curves. In all panels, intensity is normalized.

plasmons are generated only at the intersections of the graphene dispersion curves and the DFG phase-matching lines.

At $V_G = 0$ V, the Fermi levels of the top and bottom graphene layers are nearly the same. As a result, f_{SP} of both the top and bottom graphene layers are essentially the same, at ~7.5 THz in the experiment, as the measured blue dots show in Fig. 3b. This is ~0.3 THz higher than the numerical calculation (green dashed curve), which did not consider the weak interlayer plasmon coupling between the bottom and top graphene layers. When this interlayer plasmon coupling is taken into account in the numerical calculations, the green dashed curve moves up to the blue solid line in Fig. 3a,b. This is the symmetric mode of the dual-layer graphene (the antisymmetric mode has strong damping and is thus hard to observe). There is a good match between the measured blue dots and the numerically calculated blue solid line. Detailed theoretical discussions are provided in Supplementary Section 2.4. For comparison and verification we also measured f_{SP} at $V_G = 0$ V in another GSiNW sample with a 60 nm Al_2O_3 layer, which then has negligible interlayer coupling. This is shown as green triangles in Fig. 3b, which match the green dashed curve well. Details of the 60 nm Al_2O_3 dielectric device are also shown in Supplementary Fig. 5.3.

Moreover, to verify that f_{SP} matches the plasmonic dispersion curve, we changed f_p to vary the DFG phase-matching points, as illustrated in the zoomed-in Fig. 3b. By tuning λ_p from 1,532 nm to 1,542 nm (195.8 THz to 194.6 THz), we observe that the enhanced signal peak location λ_s changes from 1,593.2 nm to 1,603.0 nm (188.3 THz to 187.2 THz). Hence, f_{SP} decreases from 7.5 THz to 7.4 THz, and the trace follows the graphene plasmonic dispersion well. The measured spectra and λ_p–λ_s correlation is described in Supplementary Section 5.4. When we tune the gate voltage, the dispersion curves of the bottom- and top-layer graphene move independently. As a result, the generated plasmons on the bottom and top atomic layers have different f_{SP}s, supporting the results of Fig. 2. For instance, Fig. 3c,d, shows the scenarios of $V_G = -0.7$ V and $V_G = 0.2$ V. Here, f_p is fixed at 195.8 THz, and blue circles and red diamonds show the measured results for the bottom- and top-layer graphene, respectively (notice the change in polarity of V_G swaps the bottom and top f_{SP}s).

We next examine the intensity of the DFG plasmons (I_{SP}). In the DFG process, I_{SP} is proportional to the intensity of the pulsed pump I_p and the intensity of the c.w. signal I_s, as

$$I_{SP} = \frac{(\chi^{(2)}_{eff})^2 I_p I_s}{L^2_{SP}} \quad (4)$$

Detailed theoretical derivations are shown in Supplementary Section 2.3. In the experiment, I_{SP} could be directly estimated from the measured ΔI_{DFG}. Applying the Manley–Rowe relation, that is, the conservation of photon numbers, we can rewrite this relationship as

$$I_{SP} = \Delta I_{DFG} \frac{f_{SP}}{f_p - f_{SP}} \quad (5)$$

By fixing the signal intensity at 1.4 W and f_p at 195.8 THz, we illustrate the spectra of ΔI_{DFG} in Fig. 4a,b. In Fig. 4a, $V_G = 0$ V, and in Fig. 4b, $V_G = -0.7$ V. For either the bottom- or top-layer graphene, ΔI_{DFG} increases linearly when I_p is increased from 0 to 32 mW. The insets of Fig. 4a,b summarize the $\Delta I_{DFG}/\Delta I_p$ correlations, with a slope on the order of 10^{-3} a.u. mW^{-1}. Considering the optical loss and amplifications, we estimate the plasmons generated on-chip are on the order of single nanowatts.

We also examine the conversion efficiency of the second-order nonlinearity-based plasmon generation in Fig. 4c,d. Here, we define the conversion efficiency $\eta = I_{SP}/I_s I_p$. From equation (1), by

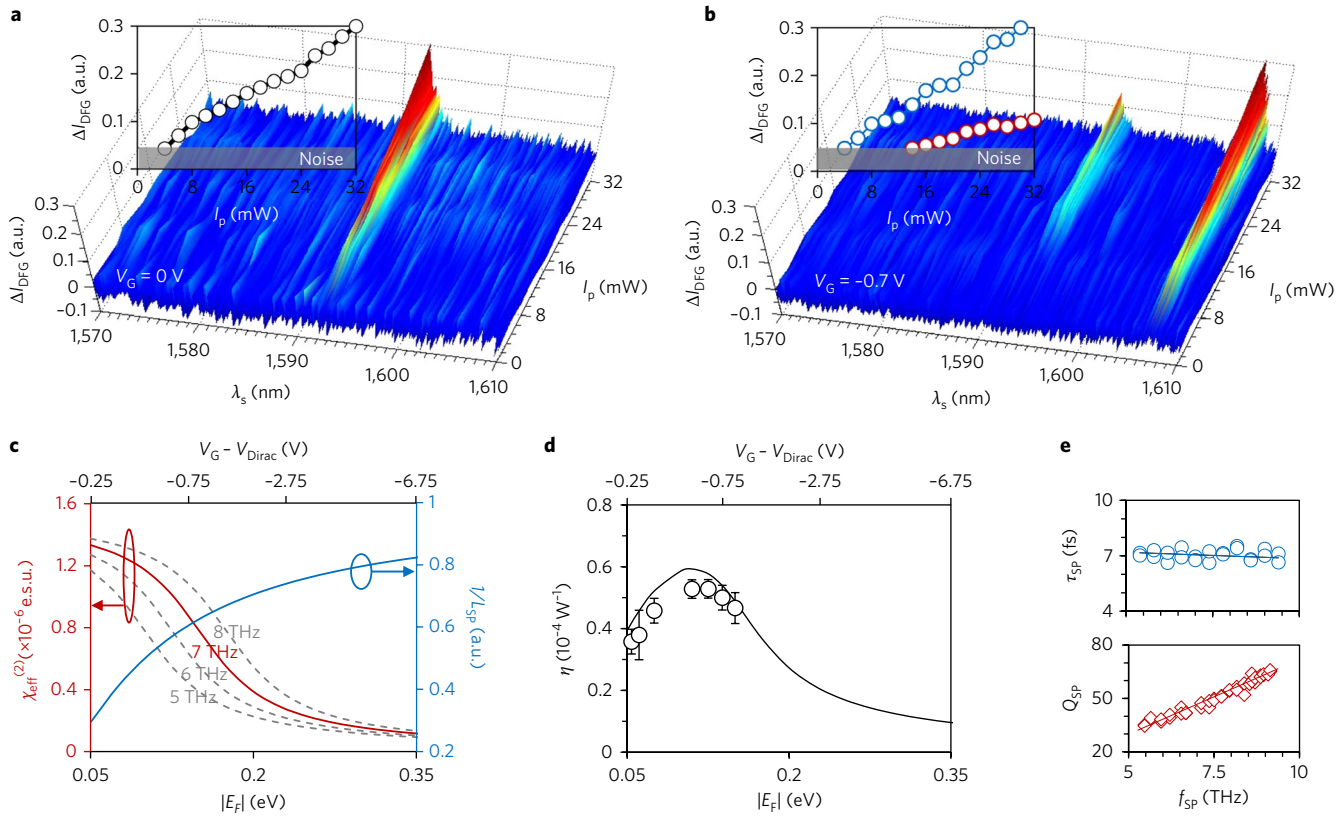

Fig. 4 | Conversion efficiency. a, Under $V_G = 0$ V, the measured spectra of the 1,593.2 nm enhanced peak are plotted. The peak intensity increases linearly from 0 to 0.3 a.u., with I_p rising from 0 to 32 mW. **b**, Under $V_G = -0.7$ V, the measured spectra of the enhanced peaks located at 1,593.7 nm (bottom layer) and 1,607.4 nm (top layer) are plotted. Peak intensities also increase linearly. Insets in **a** and **b** show the measured I_p–ΔI_{DFG} correlations. **c**, Calculated $\chi_{eff}^{(2)}$ (red curve) and $1/L_{SP}$ (blue curve) during gate tuning. With increasing E_F, graphene offers a lower $\chi_{eff}^{(2)}$ but higher $1/L_{SP}$. **d**, Overall conversion efficiency η of the DFG graphene plasmon, proportional to $\chi_{eff}^{(2)}/L_{SP}$. η increases first, and then gradually decreases when $|E_F|$ is higher than 0.13 eV. Black circles: measured results, with a peak value of ~0.6×10^{-4}. Error bars are standard error arising from the intensity noise in the measurement. **e**, Lifetime (τ_{SP}) and Q factor of the generated plasmons, according to the measured ΔI_{DFG} linewidth.

normalizing the pump power, $\eta = (\chi_{eff}^{(2)}/L_{SP})^2$. For $f_p \approx f_s \gg f_{SP}$, η could be approximately written as

$$\eta = \frac{e^3}{2\pi^3\hbar^2 k_{SP}\sqrt{f_s f_p}} \left\{\frac{\pi}{2} + \arctan\left(\frac{2\pi\sqrt{f_s f_p} - 2v_F\sqrt{2m_e E_F}}{\gamma}\right)\right\} \arctan\left(\frac{E_F}{\varepsilon\gamma}\right) \quad (6)$$

where m_e is the electron rest mass, γ is the scattering rate, $\hbar$ is the reduced Plank's constant and v_F is the Fermi velocity at ~10^6 m s^{-1}. In the GSiNW, $\chi_{eff}^{(2)}$ and $1/L_{SP}$ are of opposite trend: when $E_F = 0$ eV, graphene has the largest $\chi_{eff}^{(2)}$; however, its carrier density is minimal and $L_{SP}\to\infty$, and the graphene plasmon is completely damped. When E_F is high, due to the high carrier density, L_{SP} could be low, but $\chi_{eff}^{(2)}$ approaches 0. As a balanced trade-off, with increasing Fermi level, η rises first and then drops gradually when $E_F > 130$ meV. This means, by selecting a proper $|V_G - V_{Dirac}| \approx 0.5$ V to ensure $E_F \approx 130$ meV, we can find a highest $\eta \approx 6\times10^{-5}$ W^{-1}. Under a tuned E_F, η of the bottom and top graphene layers can indeed be degenerate.

In this Letter, by using counter-pumped $\chi^{(2)}$ DFG, we demonstrate terahertz plasmon generation and control in chip-scale integrated graphene. With a dual-layer graphene heterostructure, our counter-pumped configuration enables phase and frequency matching, with robust DFG signal detection through lock-in and balanced detection. The coherent DFG excitation is gate-tunable for both graphene layers, with symmetric–antisymmetric frequency crossing between the two layers and with tunability from 4.7 to 9.4 THz, a tunability over an octave. Under positive and negative gate voltages, we observe the plasmon dispersion frequencies exchange between the top and bottom graphene layers, with a linewidth quality factor that increases slightly with f_{SP}, up to ~60. The optimal planar terahertz plasmon generation efficiency approaches 10^{-4} when the Fermi level is 0.13 eV, drawing from a trade-off with larger $\chi^{(2)}$ but larger plasmon loss from phonon coupling (L_{SP}) with smaller E_F. These observations on chip-scale graphene terahertz plasmon generation and control open a new architectural platform for widely tunable terahertz sources, gate-tunable metasurfaces and two-dimensional atomic crystal optoelectronics.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41566-017-0054-7.

Received: 18 May 2017; Accepted: 24 October 2017;
Published online: 11 December 2017

References

1. Geim, A. & Novoselov, K. The rise of graphene. *Nat. Mater.* **6**, 183–191 (2007).
2. Zhang, Y., Tan, Y., Stormer, H. & Kim, P. Experimental observation of the quantum Hall effect and Berry's phase in graphene. *Nature* **438**, 201–204 (2005).

3. Novoselov, K. et al. A roadmap for graphene. *Nature* **490**, 192–200 (2012).
4. Constant, T., Hornett, S., Chang, D. & Henry, E. All-optical generation of surface plasmons in graphene. *Nat. Phys.* **12**, 124–127 (2016).
5. Yao, X., Tokman, M. & Belyanin, A. Efficient nonlinear generation of THz plasmons in graphene and topological insulators. *Phys. Rev. Lett.* **112**, 055501 (2014).
6. Abajo, F. Graphene nanophotonics. *Science* **339**, 917–918 (2013).
7. Bonaccorso, F., Sun, Z., Hasan, T. & Ferrari, A. Graphene photonics and optoelectronics. *Nat. Photon.* **4**, 611–622 (2010).
8. Katsnelson, M., Novoselov, K. & Geim, A. Chiral tunnelling and the Klein paradox in graphene. *Nat. Phys.* **2**, 620–625 (2006).
9. Nair, R. et al. Fine structure constant defines visual transparency of graphene. *Science* **320**, 1308–1310 (2008).
10. Vakil, A. & Engheta, N. Transformation optics using graphene. *Science* **332**, 1291–1294 (2011).
11. Cox, J. & Abajo, F. Electrically tunable nonlinear plasmonics in graphene nanoislands. *Nat. Commun.* **5**, 5725 (2014).
12. Hendry, E., Hale, P., Moger, J., Savchnko, A. & Mikhailov, S. Coherent nonlinear optical response of graphene. *Phys. Rev. Lett.* **105**, 097401 (2010).
13. Bostwick, A. et al. Observation of plasmarons in quasi-freestanding doped graphene. *Science* **328**, 999–1002 (2010).
14. Ju, L. et al. Graphene plasmonics for tunable terahertz metamaterials. *Nat. Nanotech.* **6**, 630–634 (2011).
15. Gu, T. et al. Regenerative oscillation and four-wave mixing in graphene optoelectronics. *Nat. Photon.* **6**, 554–559 (2012).
16. Bostwick, A., Ohta, T., Seyller, T., Horn, K. & Rotenberg, E. Quasiparticle dynamics in graphene. *Nat. Phys.* **3**, 36–40 (2007).
17. Grigorenko, A., Polini, M. & Novoselov, K. Graphene plasmonics. *Nat. Photon.* **6**, 749–758 (2012).
18. Ansell, D. et al. Hybrid graphene plasmonic waveguide modulators. *Nat. Commun.* **6**, 8846 (2015).
19. Freitag, M. et al. Photocurrent in graphene harnessed by tunable intrinsic plasmons. *Nat. Commun.* **4**, 1951 (2013).
20. Chakraborty, S. et al. Gain modulation by graphene plasmons in aperiodic lattice lasers. *Science* **351**, 246–248 (2016).
21. Yao, B. et al. Graphene based widely-tunable and singly-polarized pulse generation with random fiber lasers. *Sci. Rep.* **5**, 18526 (2015).
22. Rodrigo, D. et al. Mid-infrared plasmonic biosensing with graphene. *Science* **349**, 165–168 (2015).
23. Boltasseva, A. & Atwater., H. Low-loss plasmonic metamaterials. *Science* **331**, 290–291 (2011).
24. Alonso-González, P. et al. Controlling graphene plasmons with resonant metal antennas and spatial conductivity patterns. *Science* **334**, 1369–1373 (2014).
25. Alonso-González, P. et al. Acoustic terahertz graphene plasmon revealed by photocurrent nanoscopy. *Nat. Nanotech.* **12**, 31–35 (2017).
26. Nikitin, A. et al. Real-space mapping of tailored sheet and edge plasmons in graphene nanoresonators. *Nat. Photon.* **10**, 239–243 (2016).
27. Ni, G. et al. Ultrafast optical switching of infrared plasmon polaritons in high-mobility graphene. *Nat. Photon.* **10**, 244–247 (2016).
28. Yan, H. et al. Tunable infrared plasmonic devices using graphene/insulator stacks. *Nat. Nanotech.* **7**, 330–334 (2012).
29. Editorial. Terahertz optics taking off. *Nat. Photon.* **7**, 665 (2013).
30. Lan, S. et al. Backward phase-matching for nonlinear optical generation in negative-index materials. *Nat. Mater.* **14**, 807–811 (2015).

Acknowledgements

The authors thank E. Kinigstein, M. Hoff, J. Lim, J. Yang, Y.-P. Lai and Y. Li for discussions and H. Liu for help with the buffered oxide etch. Chip fabrication was also supported by the Nanoelectronics Research Facilities (NRF) of UCLA. The authors acknowledge support from the National Science Foundation (DMR-1611598, CBET-1520949 and EFRI-1433541), the Office of Naval Research (N00014-15-1-2368) and the University of California National Laboratory programme. This work was also supported by the National Science Foundation of China (61705032) and the 111 project of China (B14039).

Author contributions

B.Y. designed and led the work. B.Y., S.-W.H. and Z.X. performed optical measurements and data analysis. B.Y., C.C., J.F.F., M.Y. and D.-L.K. performed silicon nitride chip processing. Y.L., Y.H. and X.D. conducted graphene material growth and characterizations, designed and fabricated the graphene devices, and conducted the relevant electrical measurements. B.Y., Y.L. and C.C. contributed the device characterizations. C.W.W. and Y.R. supported and supervised the research. B.Y., S.H., Z.X. and Y.W. discussed and performed the theoretical analysis and simulations. B.Y., S.H., Y.R. and C.W.W. prepared the manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41566-017-0054-7.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to B.Y., X.D. or C.W.W.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Plasmon generation via counter-pumped surface $\chi^{(2)}$ nonlinearity. Detailed theoretical analysis is provided in Supplementary Section 2, which describes dispersion of the silicon nitride waveguides, phase-matching conditions on the GSiNWs, graphene index modulation, the DFG process, nonlinear conversion efficiency and dual-layer graphene plasmonic coupling.

Nanofabrication of the graphene-based semiconductor chip. The top oxide cladding of the silicon nitride waveguide core was chemically etched with a wet buffered oxide etch to increase the evanescent field coupling to the graphene. After etching, the distance between the core and the top surface was less than 20 nm, ensuring a good light–graphene interaction. Monolayer graphene was then grown by chemical vapour deposition (CVD) and transferred onto the chip using a conventional wet transfer technique, followed by patterning by photolithography and oxygen plasma etching. This graphene layer is regarded as the bottom-layer graphene with dimensions of 100 μm × 40 μm. After graphene transfer, a Ti/Au (20/50 nm) pad was deposited using electron-beam evaporation to serve as source–drain electrode. Using the source and drain, the resistance of the bottom-layer graphene could be measured. Subsequently, a thin 40 nm layer of Al_2O_3 was deposited using atomic layer deposition, providing sufficient capacitance for the graphene based semiconductor chip. Another graphene layer deposited on top of the Al_2O_3 insulator was linked to the gate electrode. The fabrication process for the graphene-based semiconductor chip is shown in Supplementary Fig. 3.1. Characterization of the nanofabrication is shown in Supplementary Section 3.2.

Experimental arrangement. To enable detection of the DFG plasmon signal, four experimental techniques were implemented: (1) both the pump and the signal were TM polarized, maximizing the graphene–light interaction; (2) a mode-locked picosecond pulsed laser served as the pump, allowing the maximum effective energy density in the GSiNW to reach 50 mJ cm^{-2}, ensuring that the DFG process was fully excited; (3) the c.w. signal was amplified to have a maximum power of 1.6 W, which can effectively pre-saturate the graphene layers, further increasing the surface $\chi^{(2)}$ efficiency, decreasing the loss and reducing the effect of a high-peak-power pulse-induced optical modulation; (4) balanced photodetection (BPD; New Focus 2017) and lock-in amplification (Stanford Research Systems 830) were implemented so that the original d.c. background of the signal was suppressed, with the balanced signal further filtered and integrated (up to 100 μs) in a lock-in amplifier, amplifying the selected signal dramatically (up to 60 dB) and suppressing white noise effectively. The bandwidth of our BPD and lock-in amplifier was on the scale of hundreds of kilohertz, so we used a 100 kHz sinusoidal waveform to modulate the pump pulses in a slow envelope. Details of the experimental architecture are shown in Supplementary Section 4.

Data availability. The data that support the plots within this paper and other findings of this study are available from the corresponding authors upon reasonable request.