Giant enhancement of third-harmonic generation in graphene–metal heterostructures

Irati Alonso Calafell [1 ✉], Lee A. Rozema[1], David Alcaraz Iranzo[2], Alessandro Trenti [1], Philipp K. Jenke [1], Joel D. Cox[3,4], Avinash Kumar [2], Hlib Bieliaiev[1], Sébastien Nanot[2,5], Cheng Peng[6], Dmitri K. Efetov [2], Jin-Yong Hong[6], Jing Kong[6], Dirk R. Englund [6], F. Javier García de Abajo [2,7], Frank H. L. Koppens [2,7] and Philip Walther [1]

Nonlinear nanophotonics leverages engineered nanostructures to funnel light into small volumes and intensify nonlinear optical processes with spectral and spatial control. Owing to its intrinsically large and electrically tunable nonlinear optical response, graphene is an especially promising nanomaterial for nonlinear optoelectronic applications. Here we report on exceptionally strong optical nonlinearities in graphene–insulator–metal heterostructures, which demonstrate an enhancement by three orders of magnitude in the third-harmonic signal compared with that of bare graphene. Furthermore, by increasing the graphene Fermi energy through an external gate voltage, we find that graphene plasmons mediate the optical nonlinearity and modify the third-harmonic signal. Our findings show that graphene–insulator–metal is a promising heterostructure for optically controlled and electrically tunable nano-optoelectronic components.

The strong light–matter coupling regime can be reached by concentrating light into nanometric volumes, which opens a wide range of applications that extend from optical sensing[1] to quantum technologies[2]. To this end, metallic nanostructures that support plasmonic excitations—the coherent oscillations of conduction electrons—are widely used to intensify electromagnetic near fields, and turn out to be particularly important for enhancing nonlinear optical processes on the nanoscale[3]. Although some aspects of plasmons can be modified by the geometry and optical properties of the host conductive media[4], to create an actively tunable plasmonic platform remains an open challenge when relying on traditional plasmonic materials such as noble metals[5–7]. In addition, plasmons in noble metals suffer from intrinsic ohmic losses that limit both their lifetimes and their optical nonlinearities[8], motivating schemes to mitigate loss by, for example, employing lattice resonances in engineered nonlinear metasurfaces[9]. Recently, a variety of nonlinear optical effects in graphene—including third-harmonic generation (THG)[10–13], four-wave mixing[11,14], the optical Kerr effect[15] and high-harmonic generation[16,17]—have been observed with high efficiency, confirming that graphene exhibits an intrinsically strong and actively tunable Nonlinear optical response. Although the optical nonlinearity of graphene is relatively efficient when normalized to the number of carbon atoms in a sample, its atomically-thin character reduces the light–matter interaction volume in which to accumulate a large signal. In this context, the use of photonic waveguides has been shown to substantially improve the nonlinear response of graphene[18].

Besides possessing an intrinsically large optical nonlinearity, highly-doped graphene can support surface plasmon polaritons that exhibit exceptionally long lifetimes, are highly confined, and can be electrically tuned across a wide spectral range[19–22].

Graphene–insulator–metal heterostructures have been demonstrated to enable strong optical field confinements, down to single-atom length scales[23], as well as near-perfect absorption of impinging light beams[24–26]. It has also been argued that such systems can reach the strong-coupling quantum regime[20,27,28]. Nonetheless, despite extensive theoretical predictions of coherent nonlinear plasmonic effects in graphene[27,29–34], very few experiments have explored graphene plasmons in the nonlinear response regime[35–37]. Moreover, prior observations have thus far relied on nonlinear mixing[35], terahertz radiation[37] or directly patterned graphene[36] to excite plasmons.

Here we demonstrate efficient THG in graphene assisted by metallic elements that enhance light coupling into plasmons of the monatomic carbon layer. We studied heterostructures made up of nanometre-thick gold nanoribbons deposited on graphene and separated by an insulating spacer layer. In our experiment, metal nanoribbons play a double role: to intensify the electric field of a far-field mid-infrared incident light beam into graphene and to launch graphene surface plasmons[23]. Our experiments demonstrate that gold nanoribbons serve as efficient nanoantennas, capable of increasing the observed THG intensity by three orders of magnitude above that of bare graphene. We confirmed that the THG originates in the graphene layer, as evidenced by our ability to actively tune the enhanced nonlinear signal by controlling the graphene Fermi energy (E_F) using an externally applied voltage; such a degree of tunability is simply not possible in conventional noble metal plasmonics because the involved carrier densities are too large to be substantially modified through gating. Additionally, our experiments revealed the role of graphene plasmons on the third-harmonic signal, which in particular emerged when varying the carrier concentration. We further observed that when the incident photon wavelength was tuned, the observed plasmonic
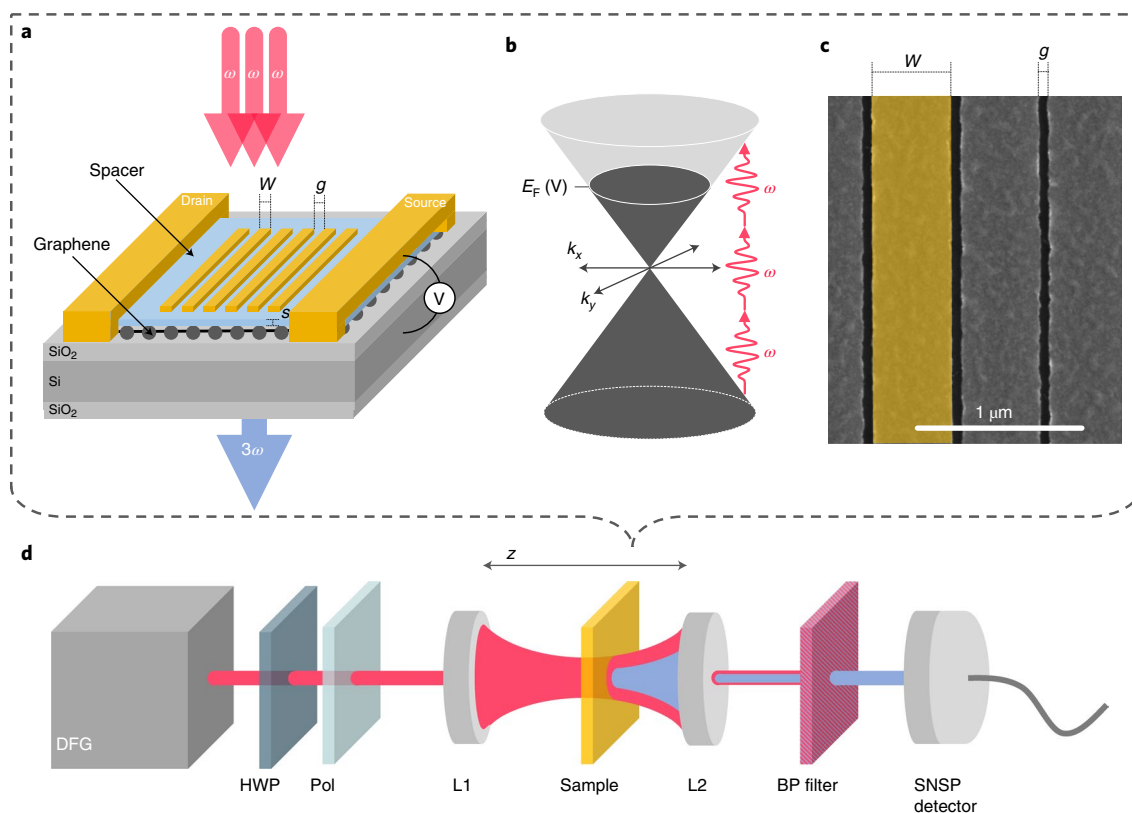
Fig. 1 | Sketch of gate-tunable graphene heterostructures. a, Graphene is encapsulated by a few-nm-thick Al_2O_3 or a monolayer h-BN film, which sets the space s between the graphene and gold nanoribbons. The gold nanoribbon arrays are characterized by the ribbon width W and the inter-ribbon gap g. Normally incident light of frequency ω undergoes THG, which is collected in transmission. A gate voltage (V) tuned from −150 to +150 V sets the graphene Fermi Energy E_F. **b**, The conical electron dispersion relation of graphene can be tuned in resonance with one, two or three incident photons. **c**, Scanning electron microscopy image of one of our high-quality gold nanoribbon arrays. **d**, Sketch of our experimental set-up. Difference-frequency generation (DFG) between the signal and idler beams of an optical parametric oscillator (not shown) provides mid-infrared of ~60 fs pulses. A half-wave plate (HWP), together with a polarizer (pol), selectively rotate the linear polarization of the incident light, which is then focused onto the sample by the L1 lens. A second lens (L2) collimates the incident and outgoing third-harmonic light. A bandpass (BP) filter isolates the third-harmonic signal, which is coupled into a multimode fibre and sent to a superconducting-nanowire single-photon (SNSP) detector. The sample is moved in the z direction (z scan), in and out of the focal point of the beam.

feature appeared for the carrier concentrations predicted by our simulations, which describe acoustic graphene plasmons excited by the nanoribbons and confined below the metal. Moreover, eliminating the plasmons from our model resulted in a poor agreement with experimental data. These signatures of plasmon-enhanced and -suppressed THG provide a new route towards the amplification and control of light at deep subwavelength scales.

THG in graphene heterostructures

Our samples were van der Waals heterostructures, which consist of a graphene sheet with a metallic nanoribbon array placed a few nanometres above it and separated by either an insulating Al_2O_3 (3–20 nm) or a monolayer hexagonal boron nitride (h-BN) spacer, as depicted in Fig. 1a. All of the gold nanoribbons were 8 nm thick, with an additional 2 nm titanium adhesion layer between the spacer and the gold nanoribbons. To isolate the nonlinear signal from the heterostructure, we used a modified z-scan set-up with a tight depth of focus (Fig. 1d). In our configuration, the sample was moved through the focal point of a femtosecond-pulsed mid-infrared incident light beam (with a wavelength of 5.5 or 3.9 μm) and a third-harmonic signal (at 1.833 or 1.3 μm, respectively) was measured in transmission (as detailed in Methods). All the measurements in this work were performed under ambient pressure and temperature conditions. A set of representative z-scan

measurements is presented in Fig. 2a, which shows that we only observed signals from bare graphene and from gold + graphene heterostructures (see Supplementary Fig. 2 for more detail). The spectrum of the nonlinear signal (with a 3.9 μm incident light) is presented in Fig. 2c, which shows a clear peak at the third-harmonic wavelength. The wavelength of the THG signal with a 5.5 μm incident light is confirmed in Supplementary Fig. 1.

Our data clearly show that the THG signals from the heterostructures are greatly enhanced compared with those of bare graphene, with additional control experiments (Supplementary Fig. 2) demonstrating that the metal structures alone do not produce a measurable THG signal. Moreover, the THG signal is maximized when the polarization is perpendicular to the direction of the nanoribbons (blue squares in Fig. 2d,e). Additionally, as shown in Fig. 2d, the THG signal of bare graphene is co-polarized with the incoming light (red triangles), and the THG signal of the heterostructures is perpendicular to the nanoribbons (blue squares). In both cases, the strongly polarized signal indicates a coherent nonlinear process.

We quantified the enhancement and verified the third-order nature of our signal by measuring the power dependence of bare graphene (red triangles in Fig. 2b) and the gold + graphene heterostructures (squares in Fig. 2b). The slope of the linear fits on a log–log scale is fixed to 3 and the y intercepts are free parameters, which we use to calculate the enhancement of the heterostructures over bare
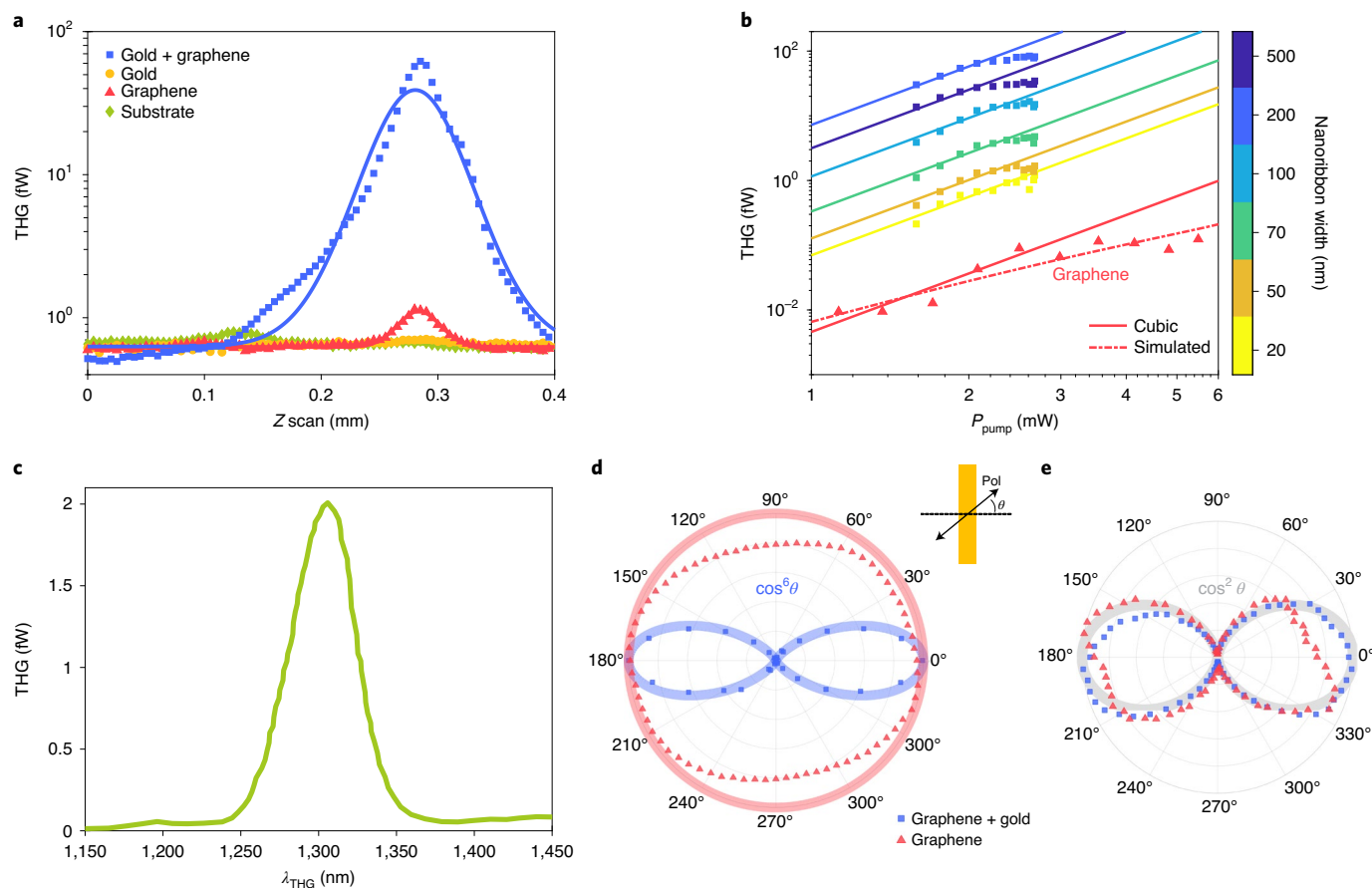
Fig. 2 | Characterization of the third-harmonic signal. a, *z*-scan measurements on four different regions of the sample (symbols). Gaussian fits to the data (curves) provide visual guides. **b**, Power scaling of the measured third-harmonic signal (symbols) in bare graphene and in gold-nanoribbon–graphene heterostructures for different nanoribbon widths with a monolayer h-BN spacer between the gold and graphene. The incident wavelength is 3.9 μm and $E_F \approx 150$ meV for this measurement. Linear fits to the data (solid lines) are used to determine the enhancement of the heterostructure relative to bare graphene. The dashed curve, which models the increase of the electron temperature with the incident light power, explains the observed saturation, which slightly differs from the expected cubic dependence. **c**, Spectrum of THG from bare graphene, measured with 3.9 μm incident light. **d**, Third-harmonic signal of bare graphene and the heterostructures measured as the input polarization is rotated. Bare graphene (red) is isotropic with respect to the incident polarization, whereas the gold nanoribbons (blue) result in a $\cos^6 \theta$ dependence on the polarization angle θ relative to the direction perpendicular to the ribbons (that is, as expected from a third-order power scaling). **e**, Third-harmonic emission measured for a fixed input polarization when a polarizer placed after the sample is rotated. For bare graphene (red), the third-harmonic signal is co-polarized with the input light, whereas for the heterostructures (blue) the third-harmonic polarization is always orthogonal to the nanoribbons. In both cases, the polarization of the third-harmonic signal is coherent relative to the incident light, which is indicated by the grey $\cos^2 \theta$ line. For the data presented in **d** and **e**, the incident light wavelength is 5.5 μm and the Fermi energy is $E_F \approx 150$ meV. The studied heterostructure has an Al_2O_3 spacer of $s = 5$ nm, and ribbons with $W = 200$ nm and $g = 50$ nm.

graphene. Although at higher powers a small saturation effect can be observed in the gold + graphene data, a clear third-order power scaling is supported by the data. To explain the saturation, we modelled the effect of the increasing incident-light power raising the electronic temperature (see the section Electron temperature (Methods)). The result for bare graphene, plotted as the dashed curve, fits our data well. We experimentally found the maximum enhancement for a 3.9 μm incident wavelength with a monolayer h-BN spacer and a ribbon width of $W = 200$ nm. Under these conditions, the heterostructures produce a THG signal that is $1,600 \pm 800$ times larger than that of bare graphene, which corresponds to a maximum THG conversion efficiency of $2 \times 10^{-7}\%$ at 1.9 mW of incident power.

To understand the enhancement mechanism and the role of the metal, we performed rigorous coupled-wave analysis (RCWA) simulations, which are presented in Fig. 3a and the detailed in Methods. Notice that due to the form of the nonlinear field integral (equation (8)), the analogous expression for higher harmonics predicts an even larger enhancement. The simulations show a strong concentration

of the electric field in the gap between the nanoribbons only when the polarization of the incident field is perpendicular to the ribbons. In contrast, the bare graphene signal is independent of the incident polarization (red triangles in Fig. 2d). Note that the slight asymmetry is caused by a polarization-dependent detection efficiency of our superconducting detector. Therefore, we conclude that the enhancement is mediated by the gold nanoribbons, which amplify the electric near-field in the graphene layer. The simulations in Fig. 3a also show that the field strength in the gap depends on the width of the nanoribbons. To verify this experimentally, we performed a series of THG measurements for different nanoribbon widths, with a spacer thickness of $s = 5$ nm and an incident light wavelength of 5.5 μm (results for other spacers are shown in Supplementary Fig. 4). From the THG signals, we estimate an effective third-order susceptibility $\chi^{(3)}$, as described in Methods. The result is shown in Fig. 3a. The experimental data agree well with our RCWA simulations, which assume an uncertainty of $\pm 20\%$ on the gap size (nominally set to 50 nm) caused by experimental imperfections.

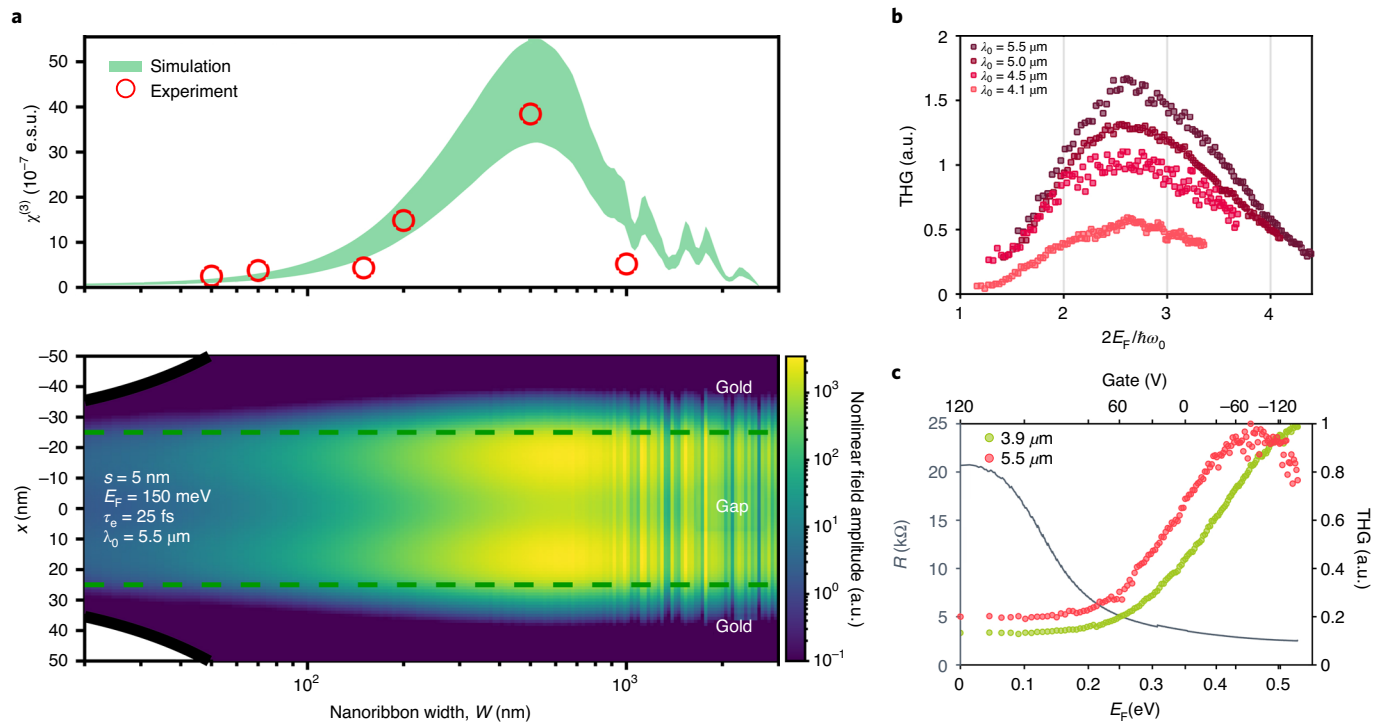**Fig. 3 | Third-order nonlinear susceptibility and electrical gating dependence. a**, Top: effective nonlinearity of gold-nanoribbon–graphene heterostructures versus ribbon width, with an $s = 5$ nm Al_2O_3 spacer and an incident light wavelength of 5.5 μm. Symbols show the experimentally estimated values, and the shaded area shows the result of our simulation in which an uncertainty of 20% on the inter-ribbon gap was introduced to account for manufacturing imperfections. The simulated $\chi_{sim}^{(3)}$ is the result of integrating the nonlinear fields (bottom) along x. Bottom: field confinement in the gap between the gold nanoribbons plotted along the gap (vertical axis) for different nanoribbon widths (horizontal axis). The colour plot indicates the calculated product of electric fields of the third-harmonic and incident light $|\eta_x^3(\omega)\eta_x(3\omega)|$, relevant for THG. **b**, Gate dependence of the third-harmonic signal in bare graphene for various incident-light wavelengths. **c**, Gate dependence of the third-harmonic signal in heterostructures with $W = 200$ nm, $g = 50$ nm and $s = 5$ nm, and incident light wavelengths of 3.9 and 5.5 μm (green and red symbols, respectively). The grey curve is the measured resistance (R) of the sample, the peak of which indicates the charge neutrality point.

Electrical tunability

Although gold nanoribbons of the appropriate width can greatly enhance the nonlinear response of the system, their geometry cannot be actively changed. In contrast, the optical nonlinearity in graphene can be electrically tuned, providing a practical route to active control. The optical response in graphene depends on the strength of intraband and interband transitions[11,13,34,38], which in turn depend on the ratio of the impinging light energy $\hbar\omega_0$ to the graphene E_F. The latter can be tuned in situ by applying an external voltage to the graphene layer relative to the silicon substrate.

Conceptually, we can understand the effect of E_F tuning on THG as illustrated in Fig. 1b. By sweeping the gate voltage, we can match E_F to be resonant to an interband transition for one, two or three incident photons. Each transition results in a different resonance in the third-order nonlinear susceptibility given by:

$$\chi^{(3)} \propto \left[-17G\left(\frac{\hbar\omega_0}{2|E_F|}\right) + 64G\left(\frac{2\hbar\omega_0}{2|E_F|}\right) - 45G\left(\frac{3\hbar\omega_0}{2|E_F|}\right) \right] \quad (1)$$

where $G(x) = \ln\left|\frac{1+x}{1-x}\right| + i\pi H(|x| - 1)$ and H is the Heaviside step function[39]. However, the resonances enter the expression with different signs. Thus, for small Fermi energies $2E_F < \hbar\omega_0$, although the three transitions can occur, the total nonlinear susceptibility nearly cancels out. For large Fermi energies, $2E_F > 3\hbar\omega_0$, all three of these transitions are Pauli blocked and there is only a small non-resonant intraband contribution[11,13]. At intermediate Fermi energies, $\hbar\omega_0 < 2E_F < 3\hbar\omega_0$, it is possible to increase the THG by, for example, Pauli blocking the one-photon and two-photon transitions, so that

only the three-photon transition is allowed and the other two transitions no longer cancel it out. For low electron temperatures, the gate response is expected to result in several sharp features as the system is tuned in and out of resonance. However, thermal broadening turns these features into broad shoulders[11,13].

The observed THG signal in bare graphene is plotted in Fig. 3b as a function of E_F for four different incident wavelengths, $\lambda_0 = 5.5$, 5.0, 4.5 and 4.1 μm (0.225, 0.248, 0.276 and 0.302 eV, respectively). A prominent peak in THG is found to emerge at a larger E_F for shorter wavelengths and thus larger $\hbar\omega_0$; we examine this feature more closely in Supplementary Fig. 5d, where we plot the Fermi energy at which the THG is maximized versus the incident photon energy. As we show in the Supplementary Information, the exact location of the maximum E_F is affected by the electron temperature, which depends on the incident fluence. Importantly, these gating data show that we can actively modulate the THG signal in bare graphene. In particular, over the four datasets for bare graphene presented in Fig. 3b, we achieved an average intensity modulation of the THG by a factor of ~10 ± 3. As illustrated in Supplementary Fig. 5, the high electron temperature (which we estimate to be ~1,100 K (Methods)) is the main limitation on this modulation.

Similar gating measurements on the heterostructure for the geometry in which we obtained a maximum field enhancement ($W = 200$ nm, $g = 50$ nm) are shown in Fig. 3c. Once again, we observe the expected shift as a function of the energy of the incident photons. We could modulate the THG by a factor of 7.4 ± 0.2 with a 3.9 μm incident light wavelength, and by a factor of 4.7 ± 0.2 with a 5.5 μm one. We stress that this active tunability comes from
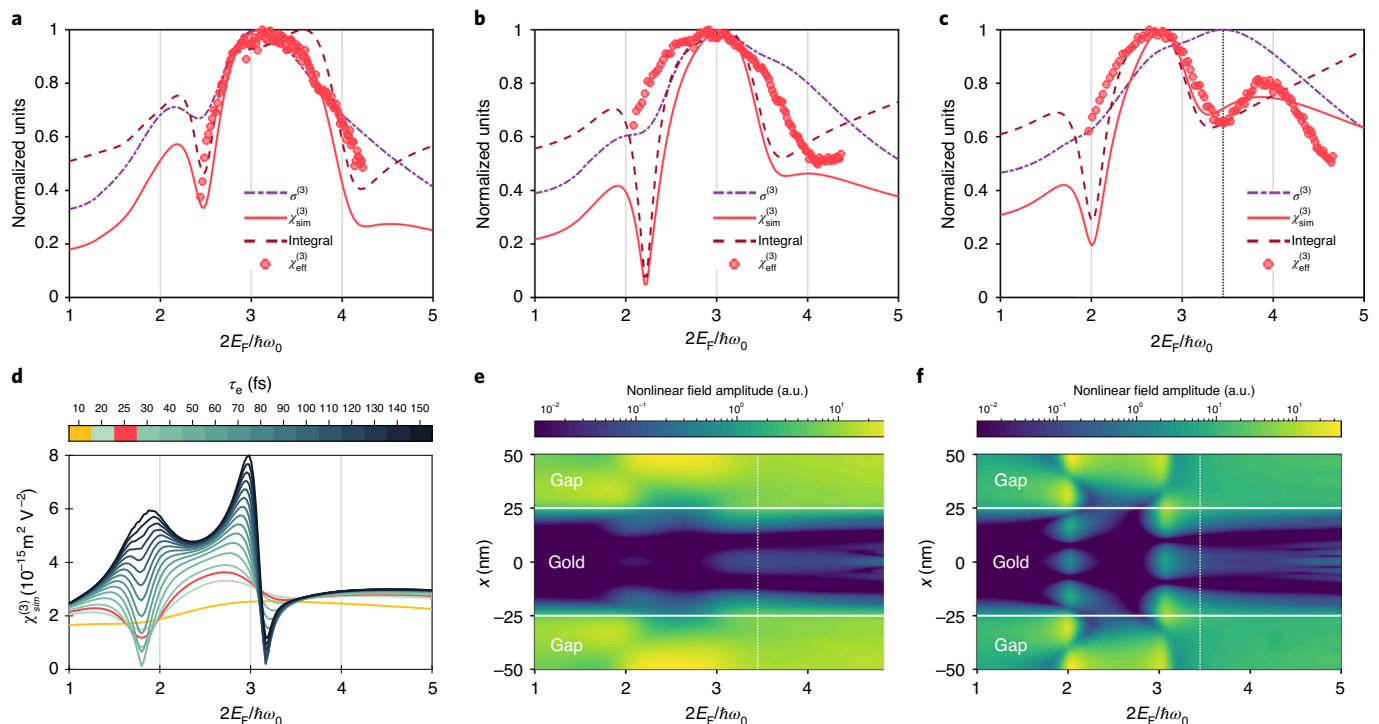
Fig. 4 | The role of graphene plasmons in THG. a–c, Gate dependence of the nonlinear response functions for incident light wavelengths $\lambda_0 = 5.5\,\mu m$ (**a**), 5.0 μm (**b**) and 4.5 μm (**c**) with a nanoribbon width of 55 nm and a gap of 45 nm. All the curves are normalized to their maxima to qualitatively compare their features. Data in these three panels correspond to the measured $\chi_{eff}^{(3)}$ (red circles) and simulated $\chi_{sim}^{(3)}$ (solid red curves) third-order nonlinear susceptibility, the third-order nonlinear conductivity $\sigma_{sim}^{(3)}$ (dashed–dotted violet curves) and the nonlinear in-plane field integral (dashed mauve curve). As the incident photon energy increases (wavelength decreases), the dip associated with the excitation of an acoustic graphene plasmon moves to higher values of E_F. **d**, Simulated gate dependence of the third-order nonlinear susceptibility $\chi_{sim}^{(3)}$ for various electron relaxation times τ_e (at a wavelength $\lambda_0 = 5.5\,\mu m$). For very low τ_e, no plasmonic effects are predicted by our model, whereas for larger τ_e plasmonic effects become evident. Our measurements are explained well by an intermediate value of τ_e of 25 fs (red curve). **e,f**, Product of field amplitudes $|\eta_x^3(\omega)\eta_x(3\omega)|$ in the graphene layer below the gold nanoribbons and in the gap as a function of E_F for $\lambda_0 = 5.5\,\mu m$ and plasmon lifetimes $\tau_e = 25\,fs$ (**e**) and $\tau_e = 150\,fs$ (**f**), respectively. An acoustic plasmon can be observed below the gold nanoribbon at $3.45 \times 2E_F/\hbar\omega_0$. As the field of the acoustic plasmon has a sign opposite to that of the field in the gap (see the phase plots in Supplementary Fig. 6), the net integrated field decreases, which leads to a dip in the effective nonlinearity ($\chi_{eff}^{(3)}$ and $\chi_{sim}^{(3)}$). All simulations were carried out with the corresponding electron temperatures specified in Supplementary Fig. 7 and with a nanoribbon width and gap both equal to 50 nm.

the unique combination of atomic thickness and linear electronic dispersion in graphene, which cannot be achieved in standard noble metal plasmonics[3].

Plasmon mediation

Interestingly, the THG signal for a larger range of E_F and for structures with smaller nanoribbon widths reveals an intriguing gate response. We plot the THG gate response of a heterostructure with $W = 55\,nm$ and $g = 45\,nm$ for different incident wavelengths in Fig. 4a–c. When the incident wavelength is 5.5 μm, two peaks and a dip are clearly present in the data (Fig. 4a). As expected, when the wavelength (energy) is decreased (increased) these features shift to higher Fermi energies. The dip is still visible with an incident wavelength of 5.0 μm at $2E_F/\hbar\omega_0 \approx 4$ (Fig. 4b). However, it shifts beyond our accessible gating range at a wavelength of 4.5 μm (Fig. 4c). None of these features are evident in the bare graphene data presented in Fig. 3b.

To provide an interpretation, we note that the graphene third-order nonlinearity associated with THG is determined by the interplay between the third-order Nonlinear graphene conductivity $\sigma_{3\omega}^{(3)}$ and the Nonlinear field integral $\int_L \eta_x^3(\omega)\eta_x(3\omega)\mathrm{d}x$ (that is, the third power of the linear field at the fundamental frequency drives the THG current amplitude, whereas the field at 3ω represents the emission enhancement produced by the heterostructure at the THG frequency). Tuning the Fermi level to higher energies can lead to

the excitation of acoustic graphene plasmons under the metal[23], which also affects the nonlinear response. The third-order conductivity $\sigma_{3\omega}^{(3)}$ depends on the excitation frequency ω, E_F, τ_e and electron temperature T_e associated with incident light absorption; the T_e can reach high values relative to the ambient room temperature, which causes anomalous behaviour, such as shifts in the features of the optical conductivity as a function of doping, compared with that of a constant T_e (Supplementary Fig. 4). Similarly, these parameters affect the nonlinear field integral through the linear graphene conductivity $\sigma^{(1)}$, although in a different manner, and thus lead to different field dependencies on E_F in the gap and below the metal (see the simulations in Fig. 4e,f). In particular, the fields from different spatial regions can have opposing signs, and thus affect the overall nonlinear performance.

It is thus important to stress that, as we show in the Fig. 4a–c, the third-order conductivity alone, despite its rich dependence on these parameters, cannot fully explain the observed THG signal. This is more evident in Supplementary Fig. 5, in which we plot the third-order conductivity at various electron temperatures and observe no features qualitatively similar to the dips in our experimental data. However, the peaks and dips that appear in the response at specific values of $2E_F/\hbar\omega_0$ correlate well with the integral of the electric fields, which partly stems from plasmonic interferences, and drives the nonlinear response. This integral, shown as a dashed line in Fig. 4a–c, exhibits the same trend as that of the data.

From our simulations, we conclude that the nonlinear response in our experiment is mainly driven by the field in the gap region in which it takes comparatively larger values (Fig. 4e). In particular, with an incident wavelength of $\lambda = 5.5\,\mu m$, $W = 55\,nm$, $g = 45\,nm$ and $s = 5\,nm$, we find a dip between the $3\hbar\omega_0$ and $4\hbar\omega_0$ transitions, and a peak just below the $4\hbar\omega_0$ transition, at $3.45 \times 2E_F/\hbar\omega_0$ and $\sim 3.9 \times 2E_F/\hbar\omega_0$, respectively (Fig 4e). The dip in the data can be explained by the partial cancellation of positive and negative complex components of the incident field in the gap by the field under the gold nanoribbons, which produces a reduction of the observed $\chi^{(3)}$.

In a more intuitive picture, at the dip we excite acoustic graphene plasmons under the metal ribbons[23,24]. As shown in Fig. 4d, for short lifetimes $\tau_e < 100\,fs$, exciting an acoustic plasmon decreases the net integrated nonlinear field (dashed maroon curves in Fig. 4a–c, labelled 'integral'). This manifests as a dip in the nonlinear signal. In Fig. 4a–c, the model with $\tau_e = 25\,fs$ fits best to our data, which confirms that the dips at 1.8 and $3.45 \times 2E_F/\hbar\omega_0$ in Fig. 4d correspond to the excitation of acoustic plasmons. These plasmonic features do not appear in the $W = 200\,nm$ heterostructure data presented in Fig. 3c because the nanoribbon width is too large, and exciting acoustic plasmons in these structures would require more doping or a longer incident wavelength. As an additional experimental confirmation, we repeated the same gate-dependent THG measurements for higher incident photon energies 0.25 eV (5.0 μm) and 0.28 eV (4.5 μm) (Fig. 4b,c). This shifted the acoustic plasmon resonance to higher gate voltages, which caused the dips to move to higher Fermi energies in our simulations. The observed shift is a key signature of graphene plasmons, which we confirmed experimentally, observing that the dip disappears for high incident photon energies.

Finally, when graphene plasmons were explicitly suppressed in our simulations (by decreasing the plasmon lifetime τ_e), we were unable to reproduce the dips and peaks in our data (red curve in Fig. 4a), which further confirms the excitation of plasmons in our graphene heterostructures.

Conclusion

Recent studies reported a wide range in the estimate of $\chi^{(3)}$ using THG[10–13,40,41]. We measure an effective nonlinearity of $\chi^{(3)}_{eff} \approx 3.4 \times 10^{-6}$ e.s.u. for the $W = 200\,nm$ heterostructure with a monolayer h-BN spacer, $E_F = 0.45$ eV, irradiated with 5.5 μm light, and an electron temperature of approximately 1,100 K, which is an order of magnitude larger than that of bare graphene, for which we observed a maximum value of $\chi^{(3)}_{eff} \approx 3.9 \times 10^{-7}$ e.s.u. (with $E_F = 0.39$ eV and irradiated at 5.5 μm). Moreover, all of our experimental measurements agree well with simulations based on the third-order nonlinear conductivity taken from Rostami et al.[34] and Mikhailov[42].

Unlike in metal plasmonics, we can actively modulate the nonlinearity of our graphene heterostructures by controlling E_F with an external gate voltage. Graphene-based linear optical devices have already been shown to operate at gigahertz frequencies[43], and hence our system provides a new route towards high-speed Nonlinear optoelectronic switches and frequency converters. Additionally, our measurements reveal intriguing plasmonic effects supported by simulations in which graphene surface plasmon polaritons appear to directly modify the nonlinear optical response of our structures. These plasmonic excitations potentially provide a novel approach to the manipulation and amplification of light at subwavelength scales. In the present work, acoustic graphene plasmons seem to modulate the nonlinear response of graphene, whereas our simulations suggest that improving the plasmon lifetime by a factor of five would increase the nonlinear response by one order of magnitude. We also observed that the maximum field enhancement (and hence the maximum nonlinearity) was obtained in a device that did not support acoustic graphene plasmons. However, one could,

in principle, design different metal nanostructure geometries to simultaneously enhance the field and launch graphene plasmons. Engineering smaller nanostructures also has the potential to excite plasmons at shorter wavelengths. Our findings suggest that graphene plasmonic devices could provide unprecedentedly strong nonlinearities, and potentially result in nonlinear optical effects at the single-photon level[2,27,28].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41565-020-00808-w.

References

1. Choi, I. & Choi, Y. Plasmonic nanosensors: review and prospect. *IEEE J. Sel. Top. Quantum Electron.* **18**, 1110–1121 (2011).
2. Tame, M. S. et al. Quantum plasmonics. *Nat. Phys.* **9**, 329–340 (2013).
3. Kauranen, M. & Zayats, A. V. Nonlinear plasmonics. *Nat. Photon.* **6**, 737–748 (2012).
4. Stockman, M. I. Nanoplasmonics: the physics behind the applications. *Phys. Today* **64**, 39–44 (2011).
5. Chikkaraddy, R. et al. Single-molecule strong coupling at room temperature in plasmonic nanocavities. *Nature* **535**, 127–130 (2016).
6. Abd El-Fattah, Z. M. et al. Plasmonics in atomically thin crystalline silver films. *ACS Nano* **13**, 7771–7779 (2019).
7. Echarri, A. R., Cox, J. D. & de Abajo, F. J. G. Quantum effects in the acoustic plasmons of atomically thin heterostructures. *Optica* **6**, 630–641 (2019).
8. Maier, S. A. *Plasmonics: Fundamentals and Applications* (Springer Science & Business Media, 2007).
9. Kravets, V. G., Kabashin, A. V., Barnes, W. L. & Grigorenko, A. N. Plasmonic surface lattice resonances: a review of properties and applications. *Chem. Rev.* **118**, 5912–5951 (2018).
10. Kumar, N. et al. Third harmonic generation in graphene and few-layer graphite films. *Phys. Rev. B* **87**, 121406 (2013).
11. Jiang, T. et al. Gate-tunable third-order nonlinear optical response of massless Dirac fermions in graphene. *Nat. Photon.* **112**, 430–436 (2018).
12. Hong, S.-Y. et al. Optical third-harmonic generation in graphene. *Phys. Rev. X* **3**, 021014 (2013).
13. Soavi, G. et al. Broadband, electrically tunable third-harmonic generation in graphene. *Nat. Nanotechnol.* **13**, 583–588 (2018).
14. Hendry, E., Hale, P. J., Moger, J., Savchenko, A. K. & Mikhailov, S. A. Coherent nonlinear optical response of graphene. *Phys. Rev. Lett.* **105**, 097401 (2010).
15. Dremetsika, E. et al. Measuring the nonlinear refractive index of graphene using the optical Kerr effect method. *Opt. Lett.* **41**, 3281–3284 (2016).
16. Yoshikawa, N., Tamaya, T. & Tanaka, K. High-harmonic generation in graphene enhanced by elliptically polarized light excitation. *Science* **356**, 736–738 (2017).
17. Baudisch, M. et al. Ultrafast nonlinear optical response of Dirac fermions in graphene. *Nat. Commun.* **9**, 1018 (2018).
18. Ooi, K. J., Ang, L. K. & Tan, D. T. Waveguide engineering of graphene's nonlinearity. *Appl. Phys. Lett.* **105**, 111110 (2014).
19. Ju, L. et al. Graphene plasmonics for tunable terahertz metamaterials. *Nat. Nanotechnol.* **6**, 630 (2011).
20. Koppens, F. H., Chang, D. E. & García de Abajo, F. J. Graphene plasmonics: a platform for strong light–matter interactions. *Nano Lett.* **11**, 3370–3377 (2011).
21. Grigorenko, A., Polini, M. & Novoselov, K. Graphene plasmonics. *Nat. Photon.* **6**, 749–758 (2012).
22. Garcia de Abajo, F. J. Graphene plasmonics: challenges and opportunities. *ACS Photon.* **1**, 135–152 (2014).
23. Iranzo, D. A. et al. Probing the ultimate plasmon confinement limits with a van der Waals heterostructure. *Science* **360**, 291–295 (2018).
24. Lee, I.-H., Yoo, D., Avouris, P., Low, T. & Oh, S.-H. Graphene acoustic plasmon resonator for ultrasensitive infrared spectroscopy. *Nat. Nanotechnol.* **14**, 313 (2019).
25. Thongrattanasiri, S., Koppens, F. H. & García de Abajo, F. J. Complete optical absorption in periodically patterned graphene. *Phys. Rev. Lett.* **108**, 047401 (2012).
26. Kim, S. et al. Electronically tunable perfect absorption in graphene. *Nano Lett.* **18**, 971–979 (2018).

27. Gullans, M., Chang, D. E., Koppens, F. H. L., García de Abajo, F. J. & Lukin, M. D. Single-photon nonlinear optics with graphene plasmons. *Phys. Rev. Lett.* **111**, 247401 (2013).

28. Alonso Calafell, I. et al. Quantum computing with graphene plasmons. *npj Quantum Inf.* **5**, 37 (2019).

29. Mikhailov, S. A. Theory of the giant plasmon-enhanced second-harmonic generation in graphene and semiconductor two-dimensional electron systems. *Phys. Rev. B* **84**, 045432 (2011).

30. Gorbach, A. V. Nonlinear graphene plasmonics: amplitude equation for surface plasmons. *Phys. Rev. A* **87**, 013830 (2013).

31. Cox, J. D. & García de Abajo, F. J. Electrically tunable nonlinear plasmonics in graphene nanoislands. *Nat. Commun.* **6725**, 5725 (2014).

32. Manzoni, M. T., Silveiro, I., García de Abajo, F. J. & Chang, D. E. Second-order quantum nonlinear optical processes in single graphene nanostructures and arrays. *New J. Phys.* **17**, 083031 (2015).

33. Cox, J. D., Marini, A. & García de Abajo, F. J. Plasmon-assisted high-harmonic generation in graphene. *Nat. Commun.* **8**, 14380 (2017).

34. Rostami, H., Katsnelson, M. I. & Polini, M. Theory of plasmonic effects in nonlinear optics: the case of graphene. *Phys. Rev. B* **95**, 035416 (2017).

35. Constant, T. J., Hornett, S. M., Chang, D. E. & Hendry, E. All-optical generation of surface plasmons in graphene. *Nat. Phys.* **12**, 124–127 (2016).

36. Kundys, D. et al. Nonlinear light mixing by graphene plasmons. *Nano Lett.* **18**, 282–287 (2018).

37. Jadidi, M. M. et al. Nonlinear terahertz absorption of graphene plasmons. *Nano Lett.* **16**, 2734–2738 (2016).

38. Mikhailov, S. A. Quantum theory of the third-order nonlinear electrodynamic effects of graphene. *Phys. Rev. B* **93**, 085403 (2016).

39. Cheng, J. L., Vermeulen, N. & Sipe, J. Third-order nonlinearity of graphene: effects of phenomenological relaxation and finite temperature. *Phys. Rev. B* **91**, 235320 (2015).

40. Saynatjoki, A. et al. Rapid large-area multiphoton microscopy for characterization of graphene. *ACS Nano* **7**, 8441–8446 (2013).

41. Woodward, R. et al. Characterization of the second-and third-order nonlinear optical susceptibilities of monolayer MoS_2 using multiphoton microscopy. *2D Mater.* **4**, 011006 (2016).

42. Mikhailov, S. A. Quantum theory of the third-order nonlinear electrodynamic effects of graphene. *Phys. Rev. B* **93**, 085403 (2016).

43. Phare, C. T., Lee, Y.-H. D., Cardenas, J. & Lipson, M. Graphene electro-optic modulator with 30 GHz bandwidth. *Nat. Photon.* **9**, 511 (2015).

Methods

Experimental details. We carried out our THG measurements using a modified z-scan set-up in which the third-harmonic signal was measured as the sample was moved along the z axis through the focus of the laser beam (Fig. 1d). Our incident light beam consisted of linearly polarized pulses of ~260 fs in duration with a tunable carrier wavelength of 5.5 μm (0.225 eV) at an 80 MHz repetition rate, which we created by performing DFG between the signal and the idler beams of an optical parametric oscillator. The typical acquisition time for a gate-dependent THG measurement was around 30 min; over this time we observed <2% power fluctuations. We used a HWP to tune the polarization of the incoming beam to that set by the polarizer. By rotating the HWP and the polarizer together, we could rotate the incident light polarization relative to the orientation of the nanoribbons. Then, a lens with a 5.26 mm focal length focused the incident light down to a waist of ~20 μm for 5.5 μm light and ~13 μm for 3.9 μm light. When the sample was moved parallel to the incident light beam (along the z axis), the nonlinear emission occurred most efficiently where the fluence was maximum (at the focal point). Afterwards, a lens with an 11 mm focal length collimated the beam, which was then sent through a BP filter to separate the THG signal from the excitation beam. Finally, the signal was coupled into a multimode fibre and sent to a large-area SNSP detector, with SNSP about a 20% detection efficiency at the third-harmonic wavelength, 5.5/3 μm = 1.833 μm.

We verified the wavelength of this signal by removing all the spectral filters, with the sample kept in focus, and using a NIREOS GEMINI interferometer to perform Fourier transform spectroscopy on the signal.

Sample fabrication and electrical doping. Here we used a heterostructure that consisted of wet-transferred chemical vapour deposition graphene stripes shaped by dry ion etching with an ultraviolet-resist positive mask. The standard chemical vapour deposition graphene used in these samples had a typical mobility below 1,000 m^2 V^{-1} s^{-1} and ~25 fs plasmon lifetimes, which is much shorter than those found in exfoliated graphene, where the lifetime can reach up to ~500 fs at room temperature[44]. An insulator, which can consist of either wet-transferred chemical vapour deposition h-BN or Al$_2$O$_3$ deposited by atomic layer deposition, defined the space between the substrate and the nanometre-thick gold nanoribbons, which were deposited by thermal evaporation onto a positive polymethyl methacrylate mask shaped using electron-beam lithography. Note that for adhesive purposes, there was a 2-nm-thick Ti layer below the gold nanoribbons. These nanoribbons concentrated the electric field of a far-field incident light into graphene[23].

We performed the gating measurements by applying a backgate voltage with a sourcemeter that allowed us to monitor the current leakage between the Si and graphene layers as we increased the applied voltage. A 1 mV voltage between the source and drain allowed us to measure the graphene resistance, which we then used to estimate the induced E_F. To calculate the induced E_F given an applied voltage V, we considered the SiO$_2$ and graphene layers to behave as a parallel-plate capacitor, in which the SiO$_2$ dielectric has a relative permittivity of $\epsilon_d \approx 4$ and a thickness d_d in between. The conversion is given by:

$$E_F(V) = \text{sign}\{n\}\hbar\sqrt{\pi|n|} \tag{2}$$

with $n = C\Delta V/e$, where $C = \epsilon_0\epsilon_d/d_d$ is the surface capacitance and $\Delta V = V - V_{CNP}$ is the difference between the applied voltage and the voltage at which the charge neutrality points (CNP) is found.

Extracting the third-order susceptibility. Experimentally, we estimated $\chi^{(3)}_{exp}$ starting with the expression of the input (i) and output (o) average power as a function of the field:

$$P(\omega_{i,o}) = \frac{1}{8}\left(\frac{\pi}{\ln 2}\right)^{3/2} f\tau W^2 n_{\omega_{i,o}} \epsilon_0 c \frac{|E(\omega_{i,o})|^2}{2} \tag{3}$$

where we assume laser pulses with repetition rate f, duration τ, waist W on the sample and Gaussian profile, $n_{\omega_{i,o}} = 2.4$ is the refractive index, and ϵ_0 (c) is the permittivity (speed of light) in a vacuum[11]. Additionally, we can write the THG process as a function of the input and output fields as follows and solve for $\chi^{(3)}_{exp}$:

$$E(\omega_o) = \frac{1}{4}\frac{i\omega_i}{2\pi c}\chi^{(3)}_{exp} d_{gr} E^3(\omega_i) \tag{4}$$

where $\omega_i = \omega_o/3$ and $d_{gr} = 0.33$ nm is the effective thickness of graphene.

Electron temperature. To estimate the electron temperature in our samples, we performed gating measurements on bare graphene for different incident wavelengths (Fig. 3b) and determined the E_F at which the maximum third-order susceptibility is found. As shown in Supplementary Fig. 5d, this E_F at which the maximum $\chi^{(3)}$ is found shifts linearly with the photon energy of the excitation light. We fitted the electron temperature of the simulations to best match the experimental data and determined a value of 1,100 K, which is in good agreement with the simulations shown in Supplementary Fig. 7. Note that, as mentioned in literature[11,13,45,46], the electron temperature changes with both the excitation

wavelength and the E_F. However, owing to the small E_F shift of about 80 meV, this effect was neglected in our simulations.

For the power-dependent measurements, it is necessary to describe the electron temperature as a function of fluence. To a good approximation, we assumed a chemical potential $\mu > k_B T_e$, which allows us to describe the electron temperature as[47]:

$$T_e = \sqrt{1 + \frac{2\eta_a F_0}{\alpha T_0^2}} \tag{5}$$

where

$$\alpha = \frac{2\pi}{3}\frac{k_B^2 \mu}{\hbar^2 v_{F_0}}^2 \tag{6}$$

T_0 is the ambient temperature, F_0 is the energy per pulse of the incident light and η_a is the amount of energy absorbed that leads to hot electrons, which we considered to be 3.5×10^{-3}, consistent with previously reported values[47].

Given the different electron temperatures for different excitation powers, as discussed in Soavi et al.[45], we were able to explain the power dependence of the third-harmonic signal (Fig. 2b).

Simulating the third-order susceptibility. We write the third-harmonic susceptibility as:

$$\chi^{(3)}_{sim} = \left|\frac{\alpha^{(3)}_{3\omega}}{\epsilon_0 d_{gr} L}\right| \tag{7}$$

where $\alpha^{(3)}_{3\omega}$ is the third-order polarizability given by:

$$\alpha^{(3)}_{3\omega}(\omega) = \frac{i\sigma^{(3)}_{3\omega}}{3\omega}\int_L \eta^3_\omega(x)\eta_{3\omega}(x)dx \tag{8}$$

where $\sigma^{(3)}_{3\omega}$ is the analytical third-order conductivity in graphene derived by Mikhailov[42], L is the length of the simulated region used for integration along the direction defining the nanoribbon width and $\eta^3_\omega(x)$ is a dimensionless quantity that represents the enhancement in the electric field amplitude that acts on the graphene layer relative to the incident field amplitude, so when multiplied by $E^3(\omega_i)$ in equation (4) it yields the actual field amplitude as a function of position x; likewise, $\eta_{3\omega}(x)$ represents the factor by which the THG field that reaches the detector is modified by the presence of the structure that surrounds the graphene. Equation (8) represents the contribution of the THG current to the far field. In particular, the $\eta^3_\omega(x)$ factor times $\sigma^{(3)}_{3\omega}$ is the THG current, which we represent as a polarization density at the emission frequency 3ω through the continuity equation. Reciprocity means that $\eta_{3\omega}(x)$ is exactly given by the enhancement amplitude in the near field relative to free space when the structure is illuminated with 3ω radiation. More precisely, we calculated $\eta_{3\omega}(x)$ as the complex factor of enhancement relative to the incident field in the field that acts on the graphene layer when it is illuminated by a 3ω plane wave that impinges from the detector direction. The electric field enhancements $\eta_\omega(x)$ and $\eta_{3\omega}(x)$ in the above expression are obtained using a RCWA Matlab script[48], adapted to include graphene as an interface material that adopts the non-local two-dimensional linear optical conductivity of graphene $\sigma(Q, \omega)$ (ref. [20]) that depends on the chemical potential μ and electron temperature T_e. We note that these enhancement factors depend on both the geometry of the heterostructures and the T_e- and μ-dependent linear conductivity. Here, we considered the influence of electronic heating by the incident light pulse in the optical response of the graphene–metal hybrid system. In particular, following the procedure in Yu et al.[49], we made use of the implicit relation between E_F, T_e and μ obtained from conservation of doping charge,

$$\left(\frac{E_F}{k_B T_e}\right)^2 = 2\int_0^\infty dx x\left[\left(e^{x-\mu/k_B T_e} + 1\right)^{-1} - \left(e^{x+\mu/k_B T_e} + 1\right)^{-1}\right] \tag{9}$$

along with the graphene heat capacity:

$$F = \beta\frac{(k_B T)^3}{(\hbar v_F)^3} \tag{10}$$

where F is the energy absorbed from the excitation pulse (that is, $F = \eta_a F_0$ where η_a is the absorbed fraction of pulse energy F_0). The energy absorbed into graphene was obtained from the linear RCWA simulations at an ambient temperature of 300 K using

$$\beta = \frac{2}{\pi}\left\{\int_0^\infty dx x^2\left[\left(e^{x+\mu/k_B T_e} + 1\right)^{-1} + \left(e^{x-\mu/k_B T_e} + 1\right)^{-1}\right] - \frac{1}{3}\left(\frac{E_F}{k_B T_e}\right)^3\right\} \tag{11}$$

The linear conductivity can thus be determined directly and the temperature-dependent nonlinear conductivity can be computed by using Maldague's identity:

$$\sigma_{3\omega}^{(3)}(\omega, \tau, \mu, T) = \frac{1}{4k_B T} \int_{-\infty}^{\infty} dE \frac{\sigma_{3\omega}^{(3)}(\omega, \tau, 0)}{\cosh^2\left(\frac{E-\mu}{2k_B T}\right)} \tag{12}$$

Data availability

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

References

44. Woessner, A. et al. Highly confined low-loss plasmons in graphene–boron nitride heterostructures. *Nat. Mater.* **14**, 421–425 (2015).
45. Soavi, G. et al. Hot electrons modulation of third-harmonic generation in graphene. *ACS Photon.* **6**, 2841–2849 (2019).
46. Mikhailov, S. A. Theory of the strongly nonlinear electrodynamic response of graphene: a hot electron model. *Phys. Rev. B* **100**, 115416 (2019).
47. Shi, S.-F. et al. Controlling graphene ultrafast hot carrier response from metal-like to semiconductor-like by electrostatic gating. *Nano Lett.* **14**, 1578–1582 (2014).
48. Manceau, J.-M., Zanotto, S., Sagnes, I., Beaudoin, G. & Colombelli, R. Optical critical coupling into highly confining metal–insulator–metal resonators. *Appl. Phys. Lett.* **103**, 091110 (2013).
49. Yu, R., Manjavacas, A. & García de Abajo, F. J. Ultrafast radiative heat transfer. *Nat. Commun.* **8**, 2 (2017).

Author contributions

I.A.C., L.A.R., F.H.L.K. and A.T. designed the experiment, performed the measurements and analysed the results. I.A.C., L.A.R., D.A.I., F.H.L.K. and J.D.C. wrote the manuscript. D.A.I., S.N., A.K. and F.H.L.K. fabricated and characterized the samples. D.A.I., J.D.C., P.K.J., H.B. and F.J.G.A. provided theoretical support and simulations. P.W., F.H.L.K. and F.J.G.A. supervised the project. All the authors read and commented on the manuscript.

competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41565-020-00808-w.

Correspondence and requests for materials should be addressed to I.A.C.

Peer review information *Nature Nanotechnology* thanks the anonymous reviewers for their contribution to the peer review of this work.

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