Title

Phase-resolved terahertz nanoimaging of WTe2 microcrystals

Journal

Physical Review B, 107(15)

ISSN

2469-9950

Authors

Jing, Ran
Vitalone, Rocco A
Xu, Suheng
et al.

Publication Date

2023-04-01

DOI

10.1103/physrevb.107.155413

Phase-resolved terahertz nanoimaging of WTe_2 microcrystals

Ran Jing,[1,*] Rocco A. Vitalone,[1] Suheng Xu,[1] Chiu Fan Bowen Lo,[1] Zaiyao Fei,[2] Elliott Runburg,[2] Yinming Shao,[1] Xinzhong Chen,[3] Fabian Mooshammer,[1] Alexander S. Mcleod,[1] Mengkun Liu,[3,4] Michael M. Fogler,[5] David H. Cobden,[2] Xiaodong Xu,[2,6] and D. N. Basov[1]

[1]*Department of Physics, Columbia University, New York, New York, USA*
[2]*Department of Physics, University of Washington, Seattle, Washington, USA*
[3]*Department of Physics and Astronomy, Stony Brook University, Stony Brook, New York, USA*
[4]*National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, New York, USA*
[5]*Department of Physics, University of California, San Diego, San Diego, California, USA*
[6]*Department of Materials Science and Engineering, University of Washington, Seattle, Washington, USA*

(Received 16 October 2022; accepted 1 March 2023; published 10 April 2023)

The terahertz (THz) electrodynamics of few-layer WTe_2 is dominated by the plasmon response. However, THz surface plasmons (SPs) with long wavelengths in two-dimensional exfoliated crystals are typically confined by the lateral geometry. Direct visualization of the plasmonic standing wave patterns is challenging due to the spatial confinement and low quality factor of the SP, especially for samples that are only a few monolayers thick. Here, we resolve subtle real-space features of the plasmonic response of WTe_2 by augmenting more common scattering amplitude experiments with the phase contrast accomplished within the time-domain version of THz nanoimaging. Amplitude and phase images allow us to quantitatively evaluate the evolution of the plasmonic response at cryogenic temperatures in samples with variable thickness from 3 to 12 monolayers. The proposed imaging modality is universally applicable to the THz near-field nanoscopy of low-dimensional materials.

DOI: 10.1103/PhysRevB.107.155413

I. INTRODUCTION

Scattering-type near-field microscopy (s-SNOM) is a powerful technique for exploring the nanoscale electrodynamics beyond the diffraction limit [1]. A conducting tip is used to enhance the electric field at the apex and scatters the incident light to the far field, resulting in a spatial resolution that depends only on the size of the tip apex (typically 30–200 nm) instead of the wavelength (typically 1–500 μm) [2,3]. The technique provides a wealth of data in experiments probing nanoscale and mesoscale phenomena, including polaritonic effects in low-dimensional systems [4–10], phase transitions in strongly correlated materials [11–15], the electrodynamics of photonic crystals [9,16,17], and free carrier distribution in nanostructured devices [18,19].

At infrared (IR) and visible frequencies, the s-SNOM technique employs two major kinds of detection strategies: pseudoheterodyne detection for continuous-wave (CW) illumination and Fourier-transformed IR spectroscopy for broadband pulses. Both strategies modulate the tip-scattered signal by oscillating the tip vertically with respect to the sample surface. The near-field (NF) signal is subsequently demodulated at the integer harmonics of the tip oscillation frequency. For an imaging experiment with monochromatic lasers, pseudoheterodyne interferometry is utilized to acquire both the amplitude and the phase of the artifact-free demodulated signal [20–22]. For spectroscopic studies using coherent broadband illumination, a Michelson interferometer is included in the light path to generate an interferogram, which is subsequently Fourier transformed to determine the frequency-dependent NF contrast [23,24]. Combining Fourier-transform IR spectroscopy with real-space scanning produces hyperspectral images that provide both spectrally and spatially dependent information. At terahertz (THz) frequencies, the recent integration of quantum cascade lasers (QCL) [25,26], THz transceivers [27], gas lasers [28,29], and backward-wave oscillators [30] provides possible options for CW imaging. However, the available frequencies are still much more limited than QCLs in the mid-IR.

For nanoscale implementation of time-domain spectroscopy (TDS), high repetition rate broadband pulses generated with photoconductive antennas are widely utilized [10,15,19,31–34]. Recently, high-power broadband THz pulses generated using nonlinear optics, such as optical rectification [35] and tilted-pulse-front schemes, have also been applied in THz NF microscopy [36]. The THz signal is detected via two major approaches: white-light (WL) imaging and TDS. The WL imaging resolves the E field at the main peak $\hat{E}$ of the time-domain THz pulse, corresponding to a spectrally integrated NF contrast [10,15,19,31–33]. Since the materials have a complex optical response, $\hat{E}$ is associated with both the amplitude and the phase of the scattered radiation. While WL imaging provides integrated nanoscale THz contrast, it is impossible to recover the full complex NF signal including both amplitude and phase contrast using this approach. The TDS method [15,31,32,37–39] yields the THz NF amplitude and phase spectra via the Fourier transformation of

*jr.jimmyjing@gmail.com

the time-domain signal. However, TDS is not suitable for fast imaging, especially since hyperspectroscopy (a combination of WL imaging and TDS) demands an exceedingly long data acquisition time. To circumvent these technical issues, we introduce here a phase-resolved detection method for pulsed illumination to rapidly access both the averaged amplitude and the phase information in broadband THz nanoimaging (nano-THz).

As a testbed for our approach, we investigate monolayer and few-layer tungsten ditelluride (WTe_2) microcrystals, which have attracted much attention due to the rich physical phenomena observed in this van der Waals layered material. For example, when patterned into narrow ribbons, WTe_2 microcrystals show hydrodynamic current flow <20 K [40]. When thinned down, the transport properties of WTe_2 reveal a crossover from three-dimensional to a two-dimensional (2D) electronic system. Spin-orbit coupling gradually drives the conduction and valence band away from the Fermi energy, leading to reduced sizes of electron and hole pockets in momentum space [41]. Few-layer WTe_2 further shows various quantum transport and phototransport behaviors [42–45] due to the low symmetry of the material. Additionally, an anisotropic plasma frequency [46–48] gives rise to in-plane hyperbolic plasmonic behavior [49]. To investigate the layer dependence of the Drude behavior on a micron-sized layered WTe_2 structure, we took the advantage of the nano-THz technology [36] to break through the diffraction limit and access the low-energy electron response. In this paper, we report the phase-resolved imaging data collected by this technique. The phase channel not only provides critical information to extract the electrodynamic property of the material but also exhibits a prominent real-space feature of propagating surface plasmon (SP) modes.

A. Detection of the carrier-envelope phase of THz pulses

In a previous study [36], we overcame the diffraction limit and used the traditional WL imaging technique to study the nanoscale THz response of WTe_2. We concluded that the temperature-dependent NF signal of 2L–12L WTe_2 is dominated by the SP. In this paper, we apply the phase-resolved NF imaging technique to acquire more detailed maps of the amplitude and phase of the scattered light. Experimental access to the full complex-valued NF signal has uncovered characteristic features of the plasmonic response of WTe_2 that remained obscured in our preliminary report. Data confirm that the WTe_2 plasmon has longer wavelength λ_P and higher quality factor Q at low temperature. We also found an increase in plasma frequency and a decrease in scattering rate at low temperatures. The complex-valued NF signals also facilitate the extraction of the complex permittivity of WTe_2 by modeling of the real-space patterns.

The schematics of the proposed phase-resolved nano-THz experiments are displayed in Fig. 1. In these experiments, THz radiation is generated through optical rectification of optical pulses at 1030 nm (17 W at 750 kHz repetition rate) with a tilted phase front onto a $LiNbO_3$ single crystal. The generated THz beam is filtered to be a Gaussian beam with a center frequency of 1 THz and a full width at half maximum of 200–300 GHz. The detection is achieved through electro-optic

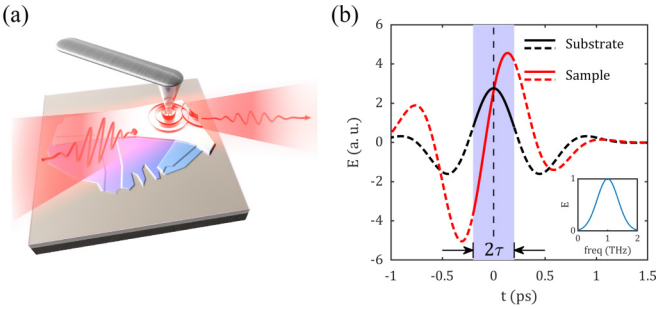

FIG. 1. Sideband detection of THz near-field signal via time-delay modulation. (a) Schematic diagram of the experimental configuration in analogy to Ref. [36]. (b) Schematic time-domain THz signals of an insulating substrate and a conductive sample on the substrate. The signals are different in both amplitude and carrier-envelope phase (CEP). Due to the CEP shift, the main peaks of the sample and the substrate occur at different time points. The shaded area indicates the range of the modulation of the measurement time delay. The inset depicts the amplitude spectrum of the THz pulses.

sampling of the tip-scattered THz pulse in ZnTe crystal gated by 20-fs pulses at 800 nm. The THz beam is focused onto an 80-μm-long PtIr wire tip from Rocky Mountain Nanotechnology. The tip oscillates and modulates the scattered NF signal at $\Omega \sim 70$ kHz. The detected signal is demodulated at the integer multiples of the tip oscillation frequency. Instead of detection at a fixed time delay (as in conventional WL imaging), we periodically modulate the time delay of the gate pulses at a frequency $M \ll \Omega$. Consequently, the detected amplitudes occur at carrier-band frequencies ($n\Omega$) and sideband frequencies ($n\Omega \pm mM$). Here, the carrier bands encode the in-phase part of the signal, and the first sidebands encode the out-of-phase part. By combining signals from different channels, the method provides a reliable estimation of the carrier-envelope phase (CEP). In the Supplemental Material [50], we provide a detailed description of the method.

B. THz NF signal of representative materials

To demonstrate the sensitivity of the THz NF amplitude and phase images to the optical constant of a material (permittivity), we simulated signals at a representative frequency of 1 THz of a gapped insulator and Dirac semimetal at different doping levels (Fig. 2). The parameters of the simulations are summarized in Table I. The sample is assumed to be a 0.3-nm-thick 10 μm×10 μm square stacked on top of a SiO_2/Si substrate. The simulation is based on the real-space NF modeling documented in Ref. [36]. Since the simulation assumes $T = 0$ K, the impact of thermal doping is not considered. In this simulation, we assume the Dirac semimetal has a relatively low scattering rate $\gamma = 10\ \mathrm{cm}^{-1}$ to highlight the doping dependence of the NF signal. In the Supplemental Material [50], we also calculated results with a higher scattering rate, as reported, for example, in graphene [51]. For the gapped insulator with $\in = 80$ [Figs. 2(a) and 2(b)], the difference in amplitude between the sample and the substrate is induced by the difference in $\in_1$ ($\in_2 \tilde{} 0$). The phase of the gapped insulator is identical to that of the substrate due to the lack of absorption. The neutral Dirac [Figs. 2(c) and 2(d)] material shows

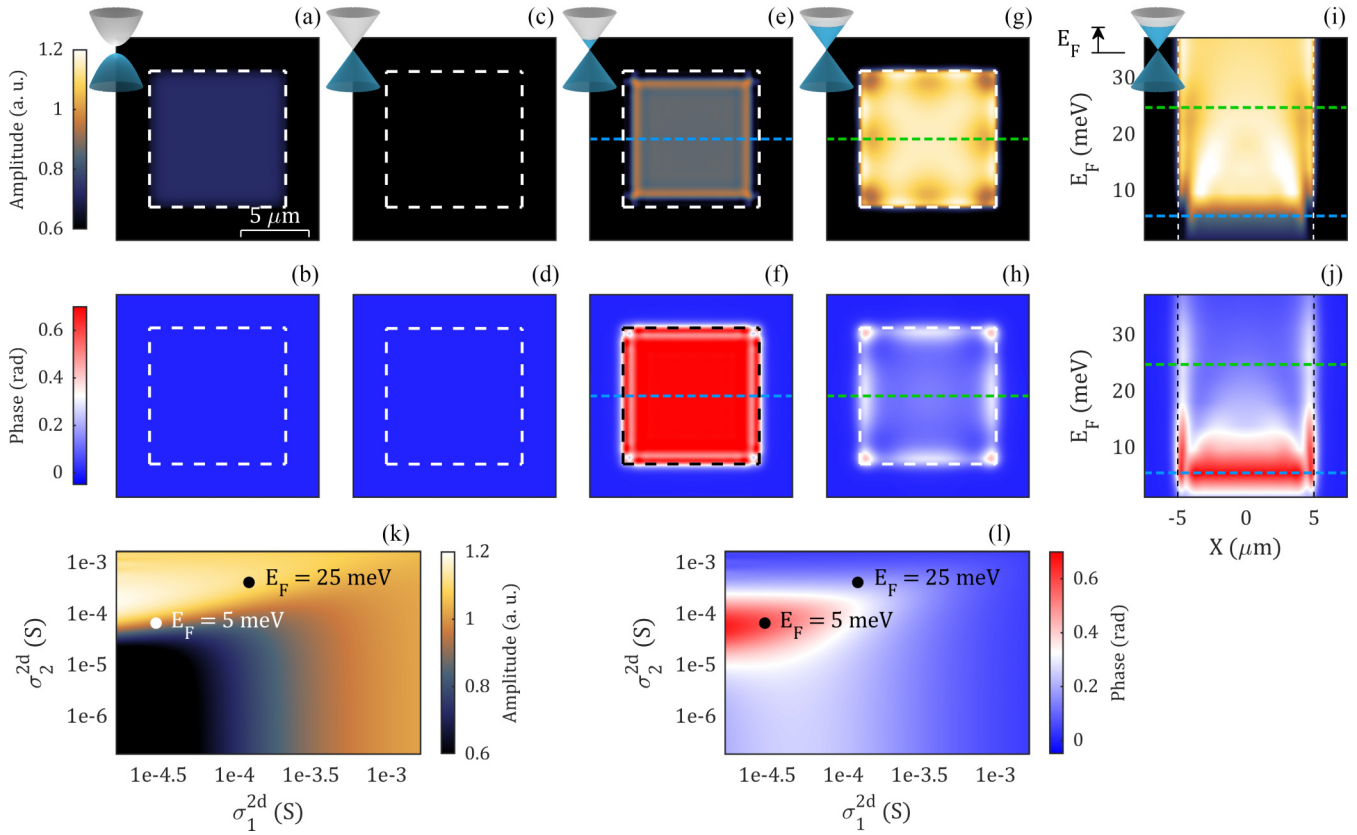

FIG. 2. THz near-field (NF) amplitude and phase for representative materials. The simulated NF amplitude and phase of a 0.3-nm-thin crystal. (a) and (b) correspond to a band insulator, (c) and (d) to Dirac material with $\mu = 0$, (e) and (f) to Dirac material with $\mu = 5$ meV, and (g) and (h) to Dirac material with $\mu = 25$ meV. The optical constants (conductivities) utilized for simulations are listed in Table I. The simulations are calculated at $f = 1$ THz and $T = 0$ K. (i) and (j) Horizontal linecuts ($Y = 0$) of amplitude and phase are extracted for a range of chemical potentials ($\mu = 0$–37.5 meV). (k) and (l) NF signal map of an infinitely large metallic sheet sample with arbitrary sheet conductivity $\sigma^{2d} = \sigma d$ at 1 THz. The dots mark the results of Dirac materials simulated at different chemical potentials.

a negligible contrast with respect to the substrate. Because of the absence of free charges, only the interband transition contributes to the optical response. However, the quantum conductance is not strong enough to generate observable NF contrast in either amplitude or phase channels.

For doped Dirac materials [Figs. 2(c)–2(h)], the surface electrodynamics is dominated by the SPs. At low chemical potentials, the plasma frequency is low and comparable with the probing THz radiation. The weak oscillator strength gives rise to a low amplitude signal [Fig. 2(e)]. However, because the probing frequency is close to the SP resonance, a strong phase response is observed [Fig. 2(f)]. At higher chemical potentials, the plasma frequency increases. The real part of the dielectric function at 1 THz becomes more negative, and the wavelength of the SP and the NF amplitude at 1 THz are increased [Fig. 2(g)]. The upshift of the plasma resonance frequency reduces the overall phase response at simulation frequency 1 THz [Fig. 2(h)]. Because of the long wavelength of the SP screened by the substrate and the low quality factor, the clarity of the SP fringe pattern is suppressed. To study the chemical potential dependence [Figs. 2(i) and 2(j)], we extracted horizontal linecuts along the sample. With increasing chemical potentials, the plasma fringes with growing wavelength on opposite edges gradually merge and weaken in both channels. As a result, the most significant feature of SPs at high chemical potential is the local enhancement of phase images near the sample edge. Considering the entire doping dependence [vertical axes of Fig. 2(j)], the spatial modulation

TABLE I. The permittivity and conductivity of the simulated materials, the unscreened (screened) plasma wavelength λ_P (λ_P^*), and the Q factor of the screened SPs in Fig. 1. The permittivity is calculated based on graphene band structure at $T = 0$ K and $\gamma = 10$ cm^{-1}.

		Dirac semimetal		
	Band insulator	$\mu = 0$	$\mu = 5$ meV	$\mu = 25$ meV
$\in$	80	$1 + 3700i$	$-4200 + 1900i$	$-2500 + 8000i$
σ_1^{2d}(S)	0	1×10^{-3}	5×10^{-4}	2.2×10^{-3}
λ_P (λ_P^*), Q	–	–	4 (1.9) μm, 2	24 (12) μm, 3

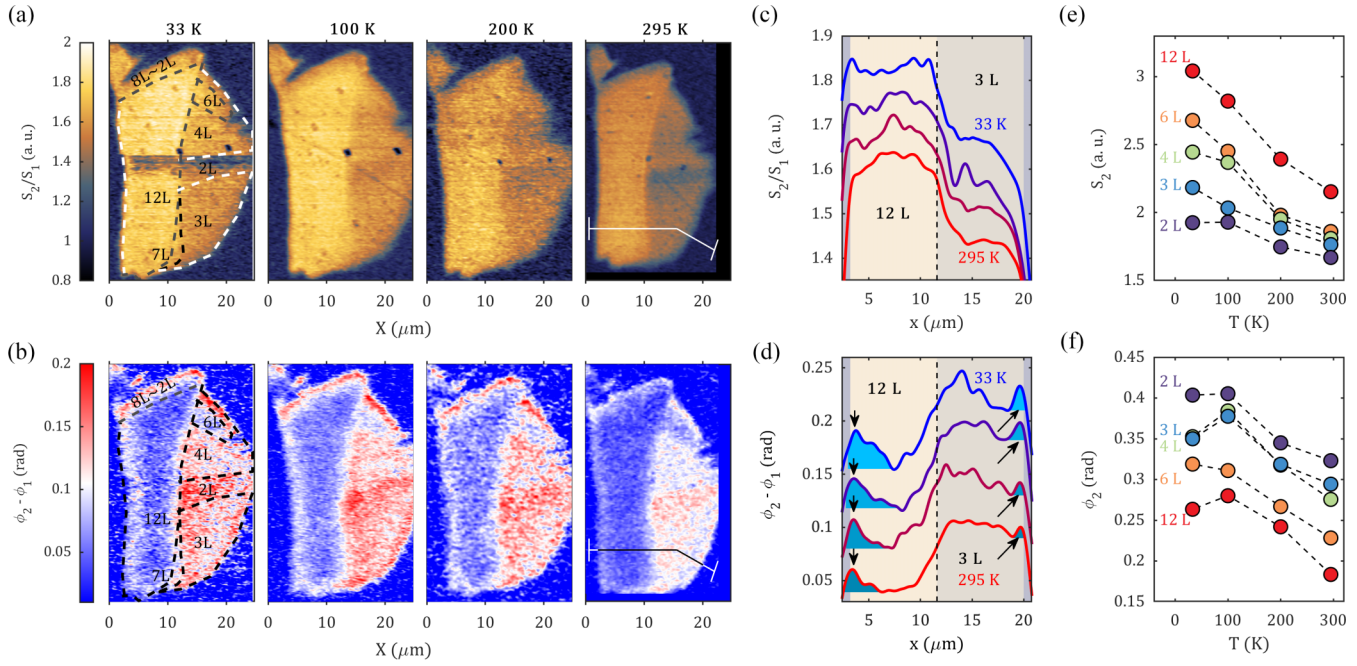

FIG. 3. The phase-resolved THz imaging of a multiterraced WTe_2 microcrystal. (a) and (b) Temperature-dependent normalized amplitude (S_2/S_1) and phase images ($\phi_2 - \phi_1$) of WTe_2 as a function of temperature. The data in each image are subsequently normalized on the SiO_2/Si substrate. The boundaries of all terraces are demarcated with dashed lines in the images recorded at 33 K. Solid lines in the images taken at 295 K reveal the trajectory of the linecuts analyzed in (c) and (d). (c) The amplitude and (d) phase linecuts extracted from (a) and (b). The linecuts in both channels are shifted along the vertical axis for clarity. The modulations near the sample edges (marked with black arrows and highlighted with blue shading) extend into the 12L region; the width of these peaks depends strongly on the temperature. (e) and (f) Temperature-dependent unnormalized amplitude (S_2) and phase (ϕ_2) averaged over all pixels in each terrace.

in scattering response is more apparent in the phase than in the amplitude channel. In Figs. 2(k) and 2(l), we map out the NF amplitude and phase for infinitely large sheet samples with arbitrary sheet conductivities $\sigma_{2d} = \sigma d$. By introducing the phase channel, we can estimate the conductivity or permittivity of the investigated sample based on Figs. 2(k) and 2(l).

II. RESULTS

We now demonstrate the utility of the phase-resolved WL imaging, using multiterraced semimetallic WTe_2 microcrystals as a case study. The configuration of our homebuilt THz-SNOM and THz generation and detection are schematically displayed in the Supplemental Material [50]. In this experiment, we focus on the demodulation at the second harmonic frequency of tip oscillation. The exfoliated microcrystals of WTe_2 have a variety of terraces. The boundaries of all regions are demarcated, and the layer numbers are labeled in Fig. 3(a). The entire microcrystal is encapsulated between the top (6 nm) and bottom (20 nm) hexagonal boron nitride. The substrate of the structure is *p*-doped silicon with a 285-nm SiO_2 cover layer.

The NF amplitude and phase of WTe_2 are acquired from 33 to 295 K [Figs. 3(a) and 3(b)]. In NF imaging, dividing the higher-order signal by the lower-order signal is proven to generate more generic NF information [52]; thus, the images and linecuts are presented as S_2/S_1 and $\phi_2 - \phi_1$. However, we observe that the ratio between signals of different orders depends strongly on the geometry of the NF tip, which may change when scanning different images. As a result, tracking the temperature dependence of the signal level is difficult. Therefore, we use S_2 normalized on the subtrate in the same image to extract the averaged signal in Figs. 3(e) and 3(f). With decreasing temperature [Fig. 3(e)], the amplitude recorded in 3L–12L regions increases, consistent with our earlier nano-THz experiment using conventional data acquisition [36]. The 2L region exhibits a slightly enhanced amplitude at low temperatures, in contrast to the previous study, where the signal on 2L is nearly temperature independent. The phase values for all regions are also higher at lower temperatures. As for the layer dependence, thicker WTe_2 regions show higher NF amplitude. The phase shows the opposite layer dependence.

The inconsistency between the 2L results here and the previous study [36] reveals the importance of acquiring the full complex-valued NF signal. Firstly, by incorporating the phase-resolved technique, we can separate the amplitude and phase degrees of freedom. Since 2L WTe_2 has a large phase compared with the substrate, the signal of 2L is delayed compared with that of the substrate. The WL time delay when scanning the 2L is not on the THz main peak. If we consider the main peak signal as the WL amplitude, the acquired signal on 2L will be significantly lower than the real WL amplitude due to the large phase contrast. Therefore, this phenomenon becomes more apparent at lower temperatures where the phase difference increases in magnitude.

As discussed in Ref. [36], the NF signal of few-layer WTe_2 is dominated by SP. In Figs. 3(c) and 3(d), we extracted linecuts at identical locations in all images in Figs. 3(a) and 3(b) to demonstrate the real-space pattern of SP. At 295 K [Fig. 3(c)],

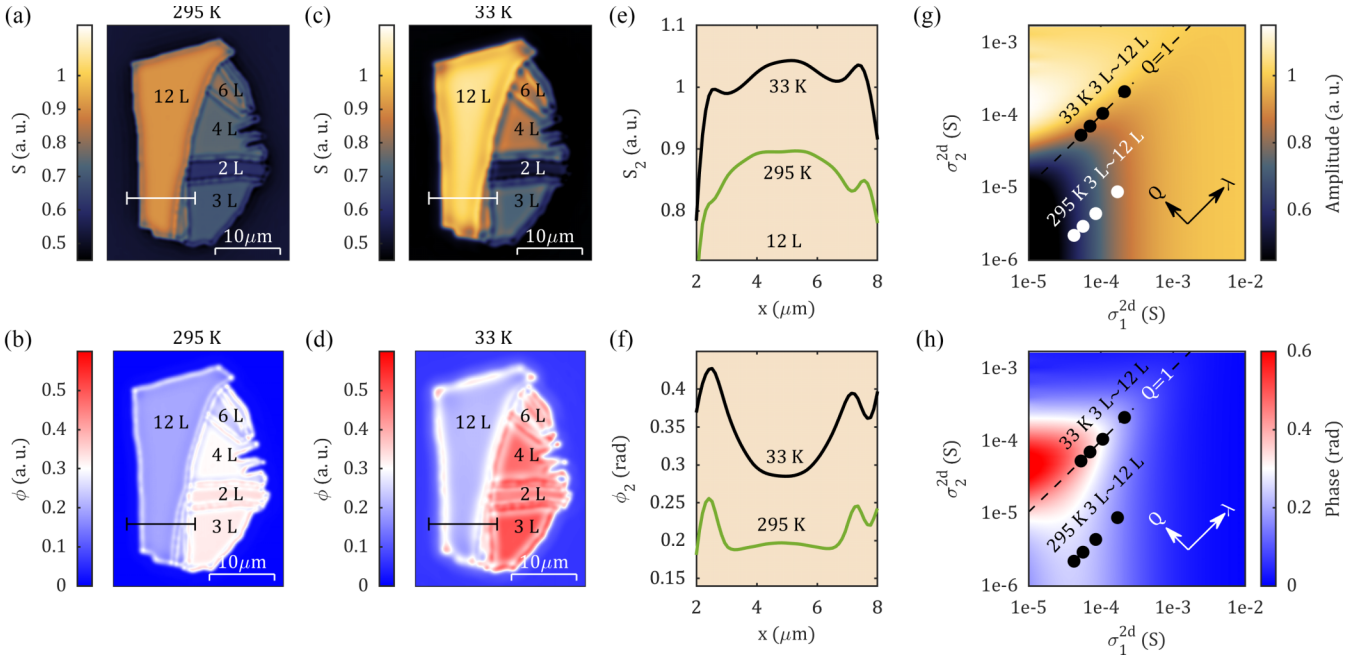

FIG. 4. Real-space modeling of surface plasmon (SP) patterns in WTe_2. (a) The amplitude and (b) phase images. Simulations adequately reproduce key features of the $T = 295$ K experimental data in Fig. 3 using permittivity values listed in Table II. The values are based on Ref. [36]. (c) and (d) The simulated amplitude and phase are calculated at $T = 33$ K. (e) and (f) The simulated amplitude and phase linecuts of 12L WTe_2 at the location indicated in (a)–(d), identical to those extracted from experiments. (g) and (h) We mark the estimated conductivities of 3L–12L WTe_2 on the near-field amplitude and phase map generated by the lightning-rod model [53]. The dashed line highlights conductivities corresponding to $Q = 1$.

the magnitude of the scattering amplitude in the middle of the 12 L WTe_2 region is higher than that at the border. At 33 K, the contrast between the center and the border of 12 L is absent. In the phase channel [Fig. 3(d)], peaks (indicated by arrows) are observed at all temperatures and on both 12L/substrate and 3L/substrate edges. To clearly illustrate the temperature dependence of the extent of the edge peak, phase linecuts in Fig. 3(d) are shifted vertically, and the peaks are highlighted with blue shades. The peak becomes more prominent and extends more deeply into the sample at lower temperatures. As discussed below, the strong and systematic temperature and thickness dependence of the edge peak in the phase profile is an important indicator of the underlying optical properties of the microcrystals. In the Supplemental Material [50], we also provide spectroscopic results showing the consistency between the phase-resolved imaging and TDS.

III. DISCUSSION

A. Simulations of nano-THz contrasts

To understand the imaging results, we simulated the real-space SP patterns at 295 K [Figs. 4(a) and 4(b)] and 33 K [Figs. 4(c) and 4(d)]. The simulation assumes the NF signal of a 2D material is proportional to the z-direction (along the tip) polarization of a dipole at a certain height near the sample surface. The concept of this method has been successfully applied to model previous NF patterns at both THz [36] and mid-IR [54–56] frequencies. The permittivity (conductivity) values of all regions in Figs. 4(a)–4(d) are summarized in Table II. The permittivity values used in the simulation are based on the model and the conclusions drawn in Ref. [36] that the 2L–12L WTe_2 can be described by a semimetal with $\Delta = 10$–20 meV. The real-space simulation focuses on the SP pattern, and the averaged signal is less important than that produced by the lightning-rod model presented in Figs. 4(g) and 4(h). Nevertheless, the simulated signal shows the same layer number and temperature dependence as in the experiment.

We extracted linecuts [Figs. 4(e) and 4(f)] in the simulation to compare with the experimental result on 12L. In both the 33 and 295 K simulation results, we observe the same narrow edge peak in the phase channel as in the experiment. From 295 to 33 K, the wavelength of the simulated SP increased by an order of magnitude even though we do not observe a change in the length scale of the real-space pattern proportional to the wavelength. The reasons are the low quality factor of the plasmon and the small size of the sample. The amplitude is boosted and flattened, which agrees with the experiment as well.

In addition to the real-space pattern, we mark the conductivities (permittivities) for 3L–12L WTe_2 in the NF signal map in Figs. 4(g) and 4(h) together with the increase in wavelength and quality factor. Because we assume 3L–12L WTe_2 shares the same optical constant, the sheet conductivities of different regions are along a diagonal line corresponding to a specific quality factor at each temperature. The maps clearly show that the dominating factor in the increase in amplitude and phase on WTe_2 from 295 to 33 K is the dramatic increase in the quality factor. The fast lengthening of the wavelength at 33 K boosts the value of $\sigma_{1,2d}$, consistent with the transport measurement [57].

TABLE II. The permittivity of each region in the simulation in Figs. 4(a) and 4(b). The permittivity value is estimated based on Ref. [36]. We also estimated the uncertainty of the simulated permittivity based on the quality of the simulation. For 3L–12L, the wavelength is proportional to the layer number N. We note that the difference in the NF signal of 3L–12L is solely caused by the difference in sample thickness.

Layer number	$\in$	$\sigma_{1,2d}$ (μS)	λ_P, λ_P^*, Q of 2L or 12L
2L (295 K)	$-17(3)+i350(100)$	30	80 nm, 28 nm, 0.05
2L (33 K)	$-450(100)+i450(100)$	36	2 μm, 0.69 μm, 1
3L–12L (295 K)	$-30(5)+i600(200)$	$20\times N$	$67\times$N nm, $23\times$N nm, 0.05 0.8 μm, 0.3 μm for 12L
3L–12L (33 K)	$-760(200)+i760(200)$	$30\times N$	$1.7\times$N μm, $0.58\times$N μm, 1 20 μm, 7 μm for 12L

To exhibit the λ_P and Q dependence of the SP pattern, we tuned both parameters in a wide range in the simulation. By increasing the Q factor from 0.2 to 2 [Figs. 5(a) and 5(b)] and maintaining $\lambda_P = 10$ μm, the amplitude fringe [indicated by black arrows in Fig. 5(a)] becomes visible. In the phase channel [Fig. 5(b)], the overall value increases with increasing the Q factor, and the phase fringe (indicated by black arrows) becomes visible as well. In Figs. 5(c) and 5(d), we demonstrate the λ_P dependence of the SP pattern by fixing the Q factor at 0.5. The wavelength dependence of the overall value is analogous to the result shown in Figs. 2(i) and 2(j). The longer λ_P diminishes the visibility of the plasmon fringe in both the amplitude and phase channels due to the limited sample size.

The SP pattern is determined by the phase shift θ acquired by the reflected wave from the sample edge [inset of Fig. 5(e)]. In Figs. 5(e) and 5(f), we demonstrate modeling with $\theta = 0-\pi$. To emphasize the features near the edge, the model calculates an SP with screened wavelength $\lambda_P^* = 8$ μm and $Q = 0.25$ at 1 THz near the left end of a one-dimensional ribbon ($L = 10$ μm). Because of the screening of the p-doped Si, the dispersion of the SP is close to linear, and therefore, the wave packet is not distorted in the propagation. In the amplitude channel [Fig. 5(e)], the signal on the edge gradually decreases with increasing θ. When $\theta = \pi$, the wave reflection is analogous to mechanical waves propagating on a string with a fixed end, leading to a low response near the edge. In the phase channel [Fig. 5(f)], the peak near the edge becomes a valley at a high phase shift. Compared with higher frequencies, THz-range SPs are more easily screened by the metallic gating material beneath the sample due to their long wavelengths. Therefore, the screened plasmon acquires

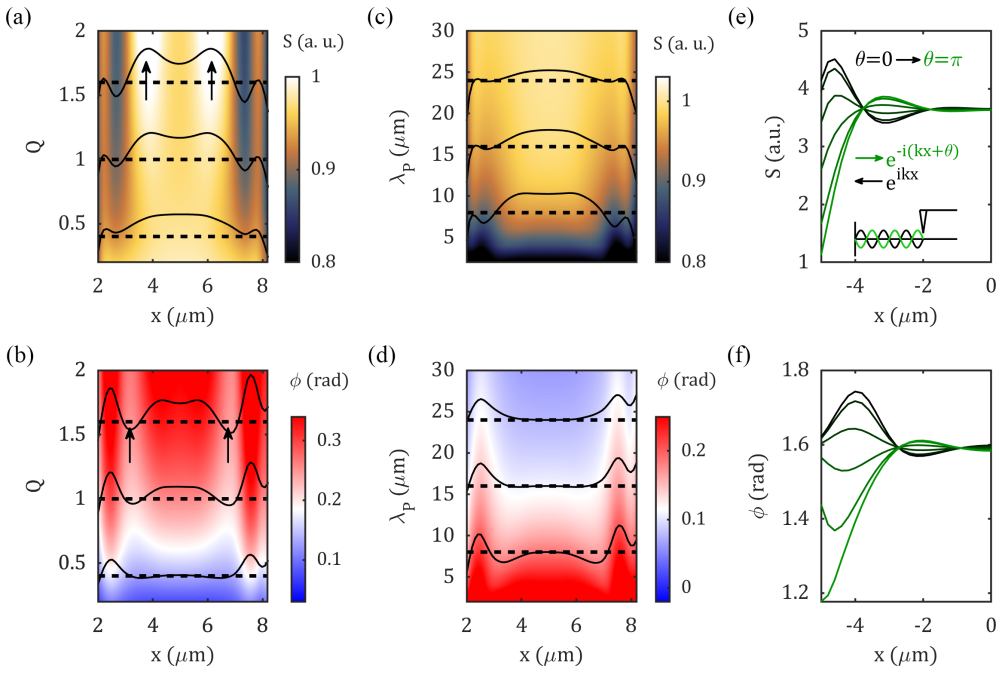

FIG. 5. Simulating THz plasma fringe patterns. In (a) and (b), we maintain the wavelength $\lambda_P = 10$ μm and tune the Q factor. The Q-factor dependences of the amplitude and phase linecuts are shown in false-color images. We highlight three curves with $Q = 0.4$, 1, and 1.6. The plasmon fringe is indicated by black arrows. In (c) and (d), we maintain the Q factor $Q = 0.5$ and tune the wavelength λ_P. We highlight three curves with $\lambda_P = 8$, 16, and 24 μm. (e) and (f) The surface plasmon (SP) pattern depends critically on the phase shift θ in the process of SP reflection from the sample edge. The θ dependence of the SP pattern is plotted. In the inset, we illustrate the phase shift acquired by the reflected SP wave.

a reflection phase shift $\theta = 0$ [58]. In the current sample configuration, the Si substrate is highly p doped and shows moderate metallicity ($\rho = 0.01 - 0.02\,\Omega\,\mathrm{cm}$). Accordingly, the SP reflection phase shift is tuned to 0 in the modeling in Figs. 4 and 5(a)–5(d), which matches reasonably well with the experimental data.

Because the illumination in the experiment is pulsed instead of CW, we must consider the launching and detection of the SP pattern in the time domain. As shown in the inset of Fig. 5(e), the signal is the interference between the tip-launched plasmon and the edge-reflected plasmon. The edge-reflected plasmon is launched by the tip at an earlier time point. For WTe_2, the speed of the propagating SP is estimated to be on the order of $v \sim 10\ \mu\mathrm{m/ps}$ [36]. Therefore, the reflected wave takes a nonnegligible time ($\sim$ picoseconds) to travel back to the tip. In this experiment, we detect near the main peak $\hat{E}$ of the pulse. The reflected wave is launched at an earlier time t when the THz excitation field is lower. This phenomenon induces an artificial damping factor to the observed SP fringe. We therefore predict that SP fringe patterns can be more easily observed at a larger time delay than the main peak. More details are documented in the Supplemental Material [50].

In conclusion, we devised, implemented, and validated a sideband detection method capable of extracting the averaged amplitude and phase of broadband THz signals. We demonstrate the utility of the imaging modalities for few-layer WTe_2. With a small time-delay oscillation, the in-phase and out-of-phase frequency components of the signal encoded in alternating carrier bands and sidebands can be easily detected through lock-in amplification. Applying this method to the nano-THz investigation of a multiterraced semimetallic WTe_2 microcrystal, we demonstrated the NF amplitude and phase pattern of an SP confined in a relatively small sample geometry. We simulated the SP pattern over a broad parameter space of plasmonic wavelength, quality factor, sample thickness, etc. By comparing these simulations with the experiment, we extracted the permittivity of few-layer WTe_2. The advantage of the sideband detection method over traditional WL imaging is that the amplitude and the phase degrees of freedom can be independently extracted. The full complex-valued signal provided by the method is the key to studying the damped plasmonic and polaritonic behaviors at THz frequencies and applies to metallic as well as dielectric systems. Compared with hyperspectral imaging, this represents a practical method to efficiently achieve averaged phase information in the challenging THz frequency range with less beam time and fewer resource requirements.

ACKNOWLEDGMENTS

Research at Columbia on electrodynamics of WTe_2 is solely supported as part of Programmable Quantum Materials, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award No. DE-SC0019443. Development of nano-THz instrumentation is supported by the Vannevar Bush Faculty Fellow program ONR-VB: N00014-19-1-2630.

[1] L. Novotny, The history of near-field optics, Prog. Opt. **50**, 137 (2007).

[2] Y. Inouye and S. Kawata, Near-field scanning optical microscope with a metallic probe tip, Opt. Lett. **19**, 159 (1994).

[3] B. Knoll and F. Keilmann, Enhanced dielectric contrast in scattering-type scanning near-field optical microscopy, Opt. Commun. **182**, 321 (2000).

[4] Z. Fei, A. S. Rodin, G. O. Andreev, W. Bao, A. S. McLeod, M. Wagner, L. M. Zhang, Z. Zhao, M. Thiemens, G. Dominguez *et al.*, Gate-tuning of graphene plasmons revealed by infrared nano-imaging, Nature (London) **487**, 82 (2012).

[5] J. Chen, M. Badioli, P. Alonso-González, S. Thongrattanasiri, F. Huth, J. Osmond, M. Spasenović, A. Centeno, A. Pesquera, P. Godignon *et al.*, Optical nano-imaging of gate-tunable graphene plasmons, Nature (London) **487**, 77 (2012).

[6] S. Dai, Q. Ma, M. K. Liu, T. Andersen, Z. Fei, M. D. Goldflam, M. Wagner, K. Watanabe, T. Taniguchi, M. Thiemens *et al.*, Graphene on hexagonal boron nitride as a tunable hyperbolic metamaterial, Nat. Nanotech. **10**, 682 (2015).

[7] G. X. Ni, A. S. McLeod, Z. Sun, L. Wang, L. Xiong, K. W. Post, S. S. Sunku, B.-Y. Jiang, J. Hone, C. R. Dean *et al.*, Fundamental limits to graphene plasmonics, Nature (London) **557**, 530 (2018).

[8] A. J. Sternbach, S. Latini, S. Chae, H. Hübener, U. De Giovannini, Y. Shao, L. Xiong, Z. Sun, N. Shi, P. Kissin *et al.*, Femtosecond exciton dynamics in WSe_2 optical waveguides, Nat. Commun. **11**, 3567 (2020).

[9] S. S. Sunku, G. Ni, B.-Y. Jiang, H. Yoo, A. Sternbach, A. S. McLeod, T. Stauber, L. Xiong, T. Taniguchi, K. Watanabe *et al.*, Photonic crystals for nano-light in moiré graphene superlattices, Science **362**, 1153 (2018).

[10] J. Zhang, X. Chen, S. Mills, T. Ciavatti, Z. Yao, R. Mescall, H. Hu, V. Semenenko, Z. Fei, H. Li *et al.*, Terahertz nanoimaging of graphene, ACS Photon. **5**, 2645 (2018).

[11] M. M. Qazilbash, M. Brehm, B.-G. Chae, P.-C. Ho, G. O. Andreev, B.-J. Kim, S. J. Yun, A. V. Balatsky, M. B. Maple, F. Keilmann *et al.*, Mott transition in VO_2 revealed by infrared spectroscopy and nano-imaging, Science **318**, 1750 (2007).

[12] M. K. Liu, M. Wagner, E. Abreu, S. Kittiwatanakul, A. McLeod, Z. Fei, M. Goldflam, S. Dai, M. M. Fogler, J. Lu *et al.*, Anisotropic Electronic State via Spontaneous Phase Separation in Strained Vanadium Dioxide Films, Phys. Rev. Lett. **111**, 096602 (2013).

[13] M. Liu, A. J. Sternbach, and D. N. Basov, Nanoscale electrodynamics of strongly correlated quantum materials, Rep. Prog. Phys. **80**, 014501 (2016).

[14] A. S. McLeod, E. Van Heumen, J. G. Ramirez, S. Wang, T. Saerbeck, S. Guenon, M. Goldflam, L. Anderegg, P. Kelly, A. Mueller *et al.*, Nanotextured phase coexistence in the correlated insulator V_2O_3, Nat. Phys. **13**, 80 (2017).

[15] H. T. Stinson, A. Sternbach, O. Najera, R. Jing, A. S. Mcleod, T. V. Slusar, A. Mueller, L. Anderegg, H. T. Kim, M. Rozenberg *et al.*, Imaging the nanoscale phase separation in vanadium

dioxide thin films at terahertz frequencies, Nat. Commun. **9**, 3604 (2018).

[16] L. Xiong, C. Forsythe, M. Jung, A. S. McLeod, S. S. Sunku, Y. M. Shao, G. X. Ni, A. J. Sternbach, S. Liu, J. H. Edgar *et al.*, Photonic crystal for graphene plasmons, Nat. Commun. **10**, 4780 (2019).

[17] F. J. Alfaro-Mozaz, S. G. Rodrigo, P. Alonso-González, S. Vélez, I. Dolado, F. Casanova, L. E. Hueso, L. Martín-Moreno, R. Hillenbrand, and A. Y. Nikitin, Deeply subwavelength phonon-polaritonic crystal made of a van der Waals material, Nat. Commun. **10**, 42 (2019).

[18] A. J. Huber, F. Keilmann, J. Wittborn, J. Aizpurua, and R. Hillenbrand, Terahertz near-field nanoscopy of mobile carriers in single semiconductor nanodevices, Nano Lett. **8**, 3766 (2008).

[19] Z. Yao, V. Semenenko, J. Zhang, S. Mills, X. Zhao, X. Chen, H. Hu, R. Mescall, T. Ciavatti, S. March *et al.*, Photo-induced terahertz near-field dynamics of graphene/InAs heterostructures, Opt. Express **27**, 13611 (2019).

[20] L. Gomez, R. Bachelot, A. Bouhelier, G. P. Wiederrecht, S.-h. Chang, S. K. Gray, F. Hua, S. Jeon, J. A. Rogers, M. E. Castro *et al.*, Apertureless scanning near-field optical microscopy: A comparison between homodyne and heterodyne approaches, J. Opt. Soc. Am. B **23**, 823 (2006).

[21] N. Ocelic, A. Huber, and R. Hillenbrand, Pseudoheterodyne detection for background-free near-field spectroscopy, Appl. Phys. Lett. **89**, 101124 (2006).

[22] A. J. Sternbach, J. Hinton, T. Slusar, A. S. McLeod, M. K. Liu, A. Frenzel, M. Wagner, R. Iraheta, F. Keilmann, A. Leitenstorfer *et al.*, Artifact free time resolved near-field spectroscopy, Opt. Express **25**, 28589 (2017).

[23] S. Amarie, T. Ganz, and F. Keilmann, Mid-infrared near-field spectroscopy, Opt. Express **17**, 21794 (2009).

[24] M. Brehm, A. Schliesser, and F. Keilmann, Spectroscopic near-field microscopy using frequency combs in the mid-infrared, Opt. Express **14**, 11222 (2006).

[25] M. C. Giordano, S. Mastel, C. Liewald, L. L. Columbo, M. Brambilla, L. Viti, A. Politano, K. Zhang, L. Li, A. G. Davies *et al.*, Phase-resolved terahertz self-detection near-field microscopy, Opt. Express **26**, 18423 (2018).

[26] E. A. A. Pogna, L. Viti, A. Politano, M. Brambilla, G. Scamarcio, and M. S. Vitiello, Mapping propagation of collective modes in Bi_2Se_3 and $Bi_2Te_{2.2}Se_{0.8}$ topological insulators by near-field terahertz nanoscopy, Nat. Commun. **12**, 6672 (2021).

[27] C. Liewald, S. Mastel, J. Hesler, A. J. Huber, R. Hillenbrand, and F. Keilmann, All-electronic terahertz nanoscopy, Optica **5**, 159 (2018).

[28] P. Alonso-González, A. Y. Nikitin, Y. Gao, A. Woessner, M. B. Lundeberg, A. Principi, N. Forcellini, W. Yan, S. Vélez, A. J. Huber *et al.*, Acoustic terahertz graphene plasmons revealed by photocurrent nanoscopy, Nat. Nanotechnol. **12**, 31 (2017).

[29] S. Chen, A. Bylinkin, Z. Wang, M. Schnell, G. Chandan, P. Li, A. Y. Nikitin, S. Law, and R. Hillenbrand, Real-space nanoimaging of THz polaritons in the topological insulator Bi_2Se_3, Nat. Commun. **13**, 1374 (2022).

[30] Q. Lu, A. T. Bollinger, X. He, R. Sundling, I. Bozovic, and A. Gozar, Surface Josephson plasma waves in a high-temperature superconductor, npj Quantum Mater. **5**, 69 (2020).

[31] K. Moon, H. Park, J. Kim, Y. Do, S. Lee, G. Lee, H. Kang, and H. Han, Subsurface nanoimaging by broadband terahertz pulse near-field microscopy, Nano Lett. **15**, 549 (2015).

[32] K. Moon, Y. Do, M. Lim, G. Lee, H. Kang, K.-S. Park, and H. Han, Quantitative coherent scattering spectra in apertureless terahertz pulse near-field microscopes, Appl. Phys. Lett. **101**, 011109 (2012).

[33] X. Chen, X. Liu, X. Guo, S. Chen, H. Hu, E. Nikulina, X. Ye, Z. Yao, H. A. Bechtel, M. C. Martin *et al.*, THz near-field imaging of extreme subwavelength metal structures, ACS Photon. **7**, 687 (2020).

[34] A. Pizzuto, D. M. Mittleman, and P. Klarskov, Laser THz emission nanoscopy and THz nanoscopy, Opt. Express **28**, 18778 (2020).

[35] M. Plankl, P. E. Faria Junior, F. Mooshammer, T. Siday, M. Zizlsperger, F. Sandner, F. Schiegl, S. Maier, M. A. Huber, M. Gmitra *et al.*, Subcycle contact-free nanoscopy of ultrafast interlayer transport in atomically thin heterostructures, Nat. Photon. **15**, 594 (2021).

[36] R. Jing, Y. Shao, Z. Fei, C. F. B. Lo, R. A. Vitalone, F. L. Ruta, J. Staunton, W. J.-C. Zheng, A. S. Mcleod, Z. Sun *et al.*, Terahertz response of monolayer and few-layer WTe_2 at the nanoscale, Nat. Commun. **12**, 5594 (2021).

[37] D. Grischkowsky, S. Keiding, M. Van Exter, and C. Fattinger, Far-infrared time-domain spectroscopy with terahertz beams of dielectrics and semiconductors, J. Opt. Soc. Am. B **7**, 2006 (1990).

[38] R. Ulbricht, E. Hendry, J. Shan, T. F. Heinz, and M. Bonn, Carrier dynamics in semiconductors studied with time-resolved terahertz spectroscopy, Rev. Mod. Phys. **83**, 543 (2011).

[39] P. U. Jepsen, D. G. Cooke, and M. Koch, Terahertz spectroscopy and imaging—Modern techniques and applications, Laser Photon. Rev. **5**, 124 (2011).

[40] U. Vool, A. Hamo, G. Varnavides, Y. Wang, T. X. Zhou, N. Kumar, Y. Dovzhenko, Z. Qiu, C. A. C. Garcia, A. T. Pierce *et al.*, Imaging phonon-mediated hydrodynamic flow in WTe_2, Nat. Phys. **17**, 1216 (2021).

[41] F.-X. Xiang, A. Srinivasan, Z. Z. Du, O. Klochan, S.-X. Dou, A. R. Hamilton, and X.-L. Wang, Thickness-dependent electronic structure in WTe_2 thin films, Phys. Rev. B **98**, 035115 (2018).

[42] Q. Ma, S.-Y. Xu, H. Shen, D. MacNeill, V. Fatemi, T.-R. Chang, A. M. M. Valdivia, S. Wu, Z. Du, C.-H. Hsu *et al.*, Observation of the nonlinear Hall effect under time-reversal-symmetric conditions, Nature (London) **565**, 337 (2019).

[43] K. Kang, T. Li, E. Sohn, J. Shan, and K. F. Mak, Nonlinear anomalous Hall effect in few-layer WTe_2, Nat. Mater. **18**, 324 (2019).

[44] Q. Wang, J. Zheng, Y. He, J. Cao, X. Liu, M. Wang, J. Ma, J. Lai, H. Lu, S. Jia *et al.*, Robust edge photocurrent response on layered type II Weyl semimetal WTe_2, Nat. Commun. **10**, 5736 (2019).

[45] S.-Y. Xu, Q. Ma, H. Shen, V. Fatemi, S. Wu, T.-R. Chang, G. Chang, A. M. M. Valdivia, C.-K. Chan, Q. D. Gibson *et al.*, Electrically switchable Berry curvature dipole in the monolayer topological insulator WTe_2, Nat. Phys. **14**, 900 (2018).

[46] A. J. Frenzel, C. C. Homes, Q. D. Gibson, Y. M. Shao, K. W. Post, A. Charnukha, R. J. Cava, and D. N. Basov, Anisotropic electrodynamics of type-II Weyl semimetal candidate WTe_2, Phys. Rev. B **95**, 245140 (2017).

[47] C. C. Homes, M. N. Ali, and R. J. Cava, Optical properties of the perfectly compensated semimetal WTe_2, Phys. Rev. B **92**, 161109(R) (2015).

[48] Y. Shao, R. Jing, S. H. Chae, C. Wang, Z. Sun, E. Emmanouilidou, S. Xu, D. Halbertal, B. Li, A. Rajendran *et al.*, Nonlinear nanoelectrodynamics of a Weyl metal, Proc. Natl. Acad. Sci. USA **118**, e2116366118 (2021).

[49] C. Wang, S. Huang, Q. Xing, Y. Xie, C. Song, F. Wang, and H. Yan, Van der Waals thin films of WTe_2 for natural hyperbolic plasmonic surfaces, Nat. Commun. **11**, 1158 (2020).

[50] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevB.107.155413 for details about the phase-resolved method, additional simulation, and extended data.

[51] P. Gallagher, C.-S. Yang, T. Lyu, F. Tian, R. Kou, H. Zhang, K. Watanabe, T. Taniguchi, and F. Wang, Quantum-critical conductivity of the Dirac fluid in graphene, Science **364**, 158 (2019).

[52] L. Mester, A. A. Govyadinov, and R. Hillenbrand, High-fidelity nano-FTIR spectroscopy by on-pixel normalization of signal harmonics, Nanophotonics **11**, 377 (2022).

[53] A. S. McLeod, P. Kelly, M. D. Goldflam, Z. Gainsforth, A. J. Westphal, G. Dominguez, M. H. Thiemens, M. M. Fogler, and D. N. Basov, Model for quantitative tip-enhanced spectroscopy and the extraction of nanoscale-resolved optical constants, Phys. Rev. B **90**, 085136 (2014).

[54] A. Y. Nikitin, P. Alonso-González, S. Vélez, S. Mastel, A. Centeno, A. Pesquera, A. Zurutuza, F. Casanova, L. E. Hueso, F. H. L. Koppens *et al.*, Real-space mapping of tailored sheet and edge plasmons in graphene nanoresonators, Nat. Photon. **10**, 239 (2016).

[55] D. J. Rizzo, B. S. Jessen, Z. Sun, F. L. Ruta, J. Zhang, J.-Q. Yan, L. Xian, A. S. McLeod, M. E. Berkowitz, K. Watanabe *et al.*, Charge-transfer plasmon polaritons at graphene/α-$RuCl_3$ interfaces, Nano Lett. **20**, 8438 (2020).

[56] S. Xu, A. S. McLeod, X. Chen, D. J. Rizzo, B. S. Jessen, Z. Yao, Z. Wang, Z. Sun, S. Shabani, A. N. Pasupathy *et al.*, Deep learning analysis of polaritonic wave images, ACS Nano **15**, 18182 (2021).

[57] Z. Fei, T. Palomaki, S. Wu, W. Zhao, X. Cai, B. Sun, P. Nguyen, J. Finney, X. Xu, and D. H. Cobden, Edge conduction in monolayer WTe_2, Nat. Phys. **13**, 677 (2017).

[58] B.-Y. Jiang, E. J. Mele, and M. M. Fogler, Theory of plasmon reflection by a 1D junction, Opt. Express **26**, 17209 (2018).

Supplementary Information for Ran, et al., “Phase-resolved Terahertz nano-imaging of WTe_2 microcrystal”

Contents

Supplementary Note 1: Side-band detection of Gaussian beam.

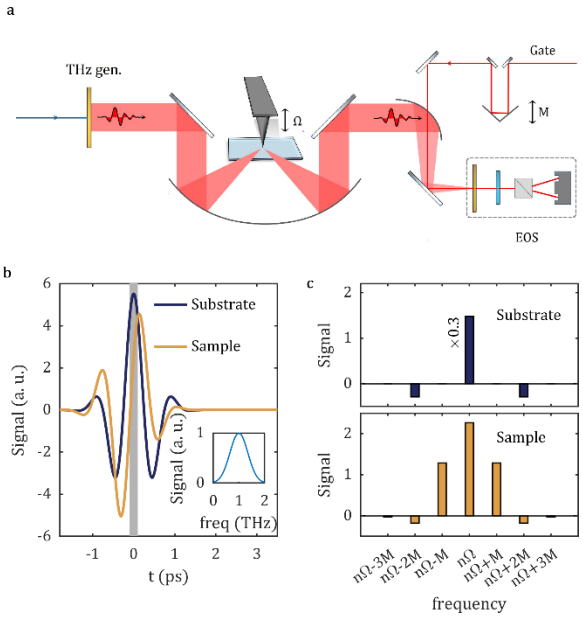

Figure S1|Side-Band detection of THz near-field signal employing the time-delay modulation. **a** The schematic diagram of the experiment configuration. **b** Time-domain spectra of two gaussian pulses with $\phi_0 = 0, \beta = 0, \alpha = 0$ (black, reference) and $\phi_0 = \pi/3, \beta = 0.1, \alpha = 0.01$ (red, sample). Two curves share the same $\omega_0 = 2\pi \cdot 1\ THz, \Delta\omega = 2\pi \cdot 0.35\ THz$. The grey area indicates the range of the time-delay modulation ($\gamma = 0.1\ ps$). **c** Frequency-domain spectrum of signal produced by the time-delay modulation shown in **b**. The carrier-band value of $\phi_0 = 0$ is rescaled by a factor of 0.3.

we assume the signal pulse S(t) can be described by a gaussian wave packet, with a center frequency ω_0 ($2\pi \cdot 1$ THz) and the full width half maximum FWHM $= 2\pi \cdot 0.82$ THz ($\Delta\omega = 2\pi \cdot 0.35$ THz). For THz signal whose phase ϕ does not have strong frequency-dependence, we assume $\phi \approx \phi_0 - \beta\frac{\omega_0(\omega-\omega_0)}{\Delta\omega^2} + \alpha\frac{(\omega-\omega_0)^2}{2\Delta\omega^2}$ ($|\alpha| \ll 1$), where α and β are unitless,

$$S(t) \propto E(t) \approx \Re\{\sqrt{2\pi\Delta\omega^2}e^{-\beta^2\frac{\omega_0^2}{2\Delta\omega^2}}e^{-i(\omega_0 t-\phi_0+\frac{\omega_0^2}{2\Delta\omega^2}\beta^2\alpha)}e^{-\frac{1}{2}\Delta\omega^2(t-t^\star)^2}\}. \quad (1)$$

where $t^\star = \frac{\beta\omega_0}{\Delta\omega^2} = \phi'(\omega_0)$. In the following discussion, the term $\frac{\omega_0^2}{2\Delta\omega^2}\beta^2\alpha$ in the phase is neglected. In practical experiments, the investigated surface consists of a reference material and a sample material. In the current example, we assume the phases of the reference and the sample spectrum are $(\phi_0, \beta, \alpha) = (0,0,0)$ and $(\phi_0, \beta, \alpha) = (\frac{\pi}{3}, 0.1, 0.01)$. The non-zero carrier-envelop phase ϕ_0 and parameter β shifts the main peak of the signal pulse away from the original detection time point at $t = 0$ ps (Fig. SFigure S1b). As a result, a pure sample-induced phase shift of the THz pulse causes observable contrast in the WL signal, which is difficult to differentiate from the contrast due to the change of the overall amplitude. By incorporating an oscillation of the time-delay at frequency M ($M \ll \Omega$),

$$t = t_0 + \tau\cos Mt, \quad (1)$$

where $t_0 = 0$ ps, $\tau = 0.1$ ps, part of the signal is distributed from the carrier-bands (CB) at frequencies $n\Omega$ ($n \in Z^+$) to side-bands (SB) $n\Omega \pm mM$ ($m \in Z^+$) (Figure S1c). The range of the oscillation is highlighted with a grey box in Fig. SFigure S1b. When $t_0 = 0$ ps (Fig. SFigure S1c), the values of $n\Omega + M$ and $n\Omega - M$ side-bands are zero for the reference spectrum with $\phi_0 = 0$ and finite value for the sample spectrum with $\phi_0 = \pi/3$.

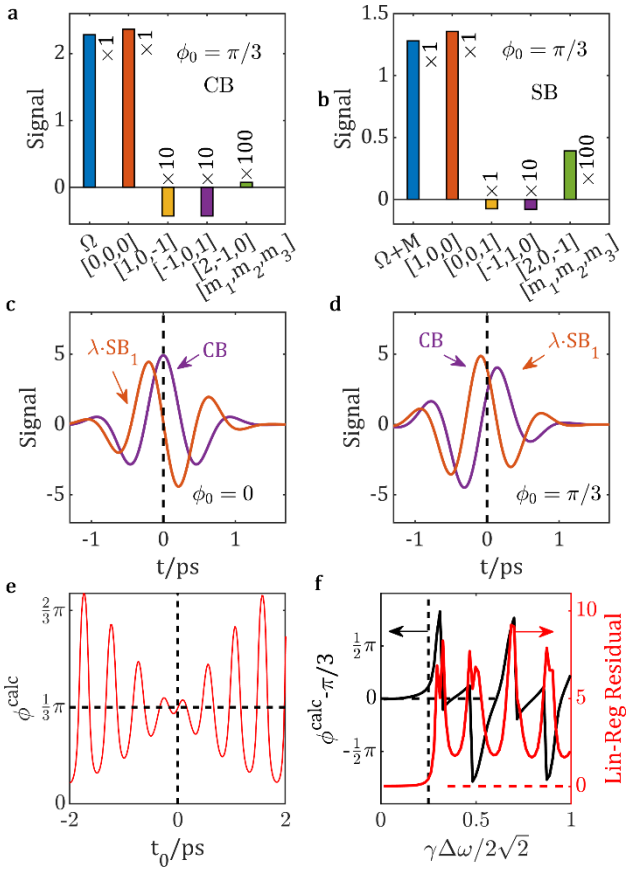

Figure S2|Phase calculation produced by the time-delay modulation. Values of **a**, the carrier-band and **b**, the 1st side-band and the first four contributing terms are compared side-by-side. Terms with small values are rescaled by 10 or 100 times. The rescale factor is displayed on the bar plot. **c** The t_0-dependent spectrum of the carrier-band and the rescaled 1st side-band of the reference spectrum. The dashed line at $t = 0$ ps indicates where the values are collected to calculate the phase. **d** The same with c for sample spectrum. **e** ϕ^{calc} is calculated at different delay t_0. When t_0 is near the interval [0,1.3] ps, the calculation result is close to $\frac{\pi}{3}$. **f** The accuracy and the linearity of the model phase ϕ^{calc} is calculated at different γ value. The linearity is represented by the residue of the linear regression to $\phi^{calc}(\phi_0)$ with $\phi_0 \in [0,2\pi]$. The vertical dash line marks the limit of $\gamma\Delta\omega/2\sqrt{2}$ value below which the proposed method is valid. The horizontal dash lines marks zero level.

With Eq. (1) and Eq. (1), the time-dependent signal can be expanded into

$$S(t) \propto \Re\left\{ e^{-\frac{1}{4}\Delta\omega^2(\tau^2+2(t_0-t^*)^2)} e^{-i(\omega_0 t_0-\phi_0+\frac{\omega_0^2}{2\Delta\omega^2}\beta^2\alpha)} \sum_{m_{1,2,3}} (-1)^{m_1+m_2+m_3} i^{m_1} J_{m_1} I_{m_2} I_{m_3} e^{i(m_1+2m_2+m_3)Mt} \right\}, \quad (2)$$

where J_{m_1}, I_{m_2} and I_{m_3} are Bessel function and modified Bessel functions of the first kind with arguments $\omega_0\tau$, $\Delta\omega^2\tau^2/4$ and $\Delta\omega^2\tau(t_0 - t^*)$ respectively. For a certain side-band $n\Omega + mM$, all terms with $m_1 + 2m_2 + m_3 = m$ in the sum of Eq. (2) contribute to the detected signal. The detailed frequency-domain analysis for $\phi = \pi/3$ spectra is shown in Fig. SFigure

S2a and 2b. The values of the CB $S_{n\Omega}$, the 1st SB $S_{n\Omega+M}$ and their first four contributing terms are compared side-by-side. Here, Ω is the tip tapping frequency. When the conditions

$$\begin{cases} I_{m_2} \sim \Delta\omega^2\tau^2/8 \ll 1 \\ I_{m_3} \sim \dfrac{\Delta\omega^2\tau}{2}|t_0 - t^\star| \ll 1 \end{cases} \tag{3}$$

are satisfied, the contribution from terms with non-zero m_2 and m_3 are negligible. Consequently, the CB and SBs encodes the in-phase and the out-of-phase signal alternatively. Eq. (2) approximately reduces to

$$\begin{bmatrix} S_{n\Omega} \\ S_{n\Omega\pm M} \end{bmatrix} \propto \begin{bmatrix} J_0 I_0 I_0 & 2J_1 I_0 I_{-1} \\ -J_0 I_0 I_1 & J_1 I_0 I_0 \end{bmatrix} \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \approx \begin{bmatrix} J_0 I_0 I_0 & 0 \\ 0 & J_1 I_0 I_0 \end{bmatrix} \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix} \tag{4}$$

where $\theta = -\omega_0 t_0 + \phi_0$. According to Eq. (4), the in-phase and out-of-phase signal can be independently extracted from the CB and the ± 1 SBs. To correctly calculate the phase, we introduce the parameter $\lambda = \frac{J_0(\omega_0\tau)}{J_1(\omega_0\tau)}$ to cancel the pre-factor. In Fig. SFigure S2c and S2d, t_0-dependent spectra of CB and rescaled 1st SB of the reference and sample are displayed. The calculated amplitude and phase can be expressed as

$$\begin{cases} S_n^{calc} = \left(S_{n\Omega}^2 + \left(\lambda \dfrac{S_{n\Omega+M} + S_{n\Omega-M}}{2} \right)^2 \right)^{0.5} \\ \phi_n^{calc}(\phi_0) = -\tan^{-1} \lambda \dfrac{S_{n\Omega+M} + S_{n\Omega-M}}{2S_{n\Omega}} \end{cases}. \tag{5}$$

Now, we discuss the accuracy and the linearity of $\phi^{calc}(\phi_0)$. In Fig. S2d, ϕ^{calc} is calculated by applying the prescribed protocol at varying the time-delay t_0. In the current example, Eq. (3) requires $\tau \ll \frac{2\sqrt{2}}{\Delta\omega} = 1.29$ ps and $|t_0 - t^\star| \ll \frac{2}{\Delta\omega^2\tau} = 4.14$ ps. In our simulation of the protocol we use, $t_R^\star = 0.13$ ps, $t_S^\star = 0$ ps and $\tau = 0.1$. Therefore, we confirm the calculation is accurate for t_0 near the interval [0,0.13] ps (Fig. SFigure S2e). In practical nano-THz experiments, to maximize the amplitude of the measured fields, t_0 is usually chosen to be the time-delay of the WL peak of the reference spectrum. Due to the factor of $e^{-\frac{1}{2}\Delta\omega^2(t-t^\star)^2}$ in Eq. (1), the white light peak is located at $t_0 \approx t_R^\star + \frac{\phi_0 - \omega_0 t_R^\star}{\omega_0}$. Here, $\phi_0 - \omega_0 t_R^\star$ is wrapped in the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$. When $\omega_0 = 2\pi \cdot 1$ THz, $t_0 \in t_R^\star + \left(-\frac{1}{4}, \frac{1}{4}\right)$ ps. Hence, Eq. (3) becomes,

$$\begin{cases} \tau \ll \dfrac{2\sqrt{2}}{\Delta\omega} \\ |t_S^\star - t_R^\star| \ll \dfrac{2}{\Delta\omega^2\tau} \end{cases}. \tag{6}$$

Besides the accuracy, the linearity of the mapping of $\phi_0 \rightarrow \phi^{calc}$ is an important indicator of the validity of the method. The source of the non-linearity arises from the contribution of terms with non-zero m_2 and m_3 in Eq. (2). The linearity is characterized by the residual of the linear-regression on $\phi^{calc}(\phi_0)$. In Fig. S2e, the τ dependent accuracy and the linearity of the mapping $\phi^{calc}(\phi_0)$ share similar behavior. With our residual analysis, we find the side-band detection method is sufficiently valid for $\frac{\tau\Delta\omega}{2\sqrt{2}} \leq 0.25$ (viz. $\tau \leq 0.32$ ps). By including more terms in Eq. (4), it is possible to extend the validity of our protocol to higher values of τ, which increases the value of SBs.

Supplementary Note 2: Real-space THz near-field imaging of graphene.

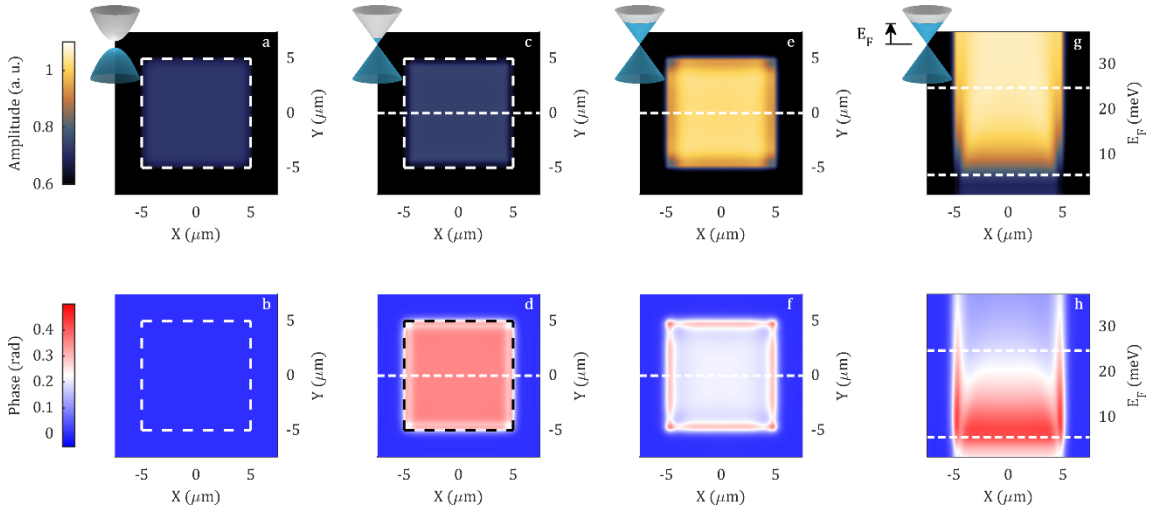

Figure S3| THz near-field amplitude and phase for representative materials. The result is presented in same order and format with Fig. 2 in the main text. Here, we increased the scattering rate from $\gamma = 10\ cm^{-1}$ to $30\ cm^{-1}$. With increased scattering rate, even if the plasmon pattern is smeared, the trend of the signal is identical to the result with lower scattering rate.

At THz frequencies, the scattering rate of graphene is higher than 1 THz at doping level below 50 meV (>77 K) [1]. Higher scattering rate smear the pattern of surface plasmon. In Figure S3, we recalculate the result presented in Fig. 2 in the main text with increased scattering rate $\gamma = 30\ \mathrm{cm}^{-1}$. The results show similar feature to those with lower scattering rate in Fig. 2. The overall amplitude increases monotonically, and the phase is roughly maximized at a frequency slightly higher than the probe frequency (1 THz). Due to the increase of the scattering rate, the quality factor Q of the surface plasmon is reduced to 1/3 of the original value. As a result, the real-space plasmon pattern presented in Fig. S3c to S3h is damped in a shorter distance to the sample edge. For reference, the plasmon wavelength and the Q-factor of different scattering rate is listed and compared in Table. S1.

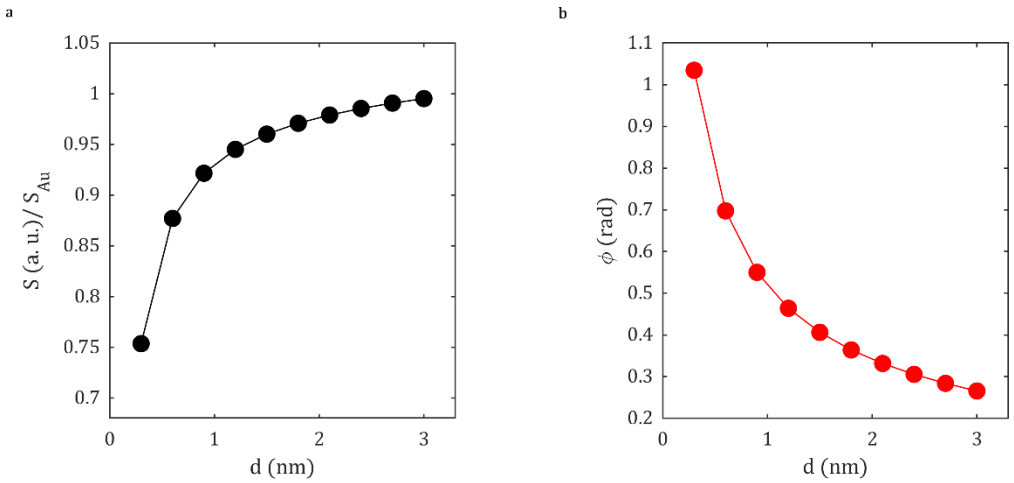

Figure S4| Thickness dependence of THz near-field signal. We assume the material has a Dirac band structure with $E_F = 12.5\ meV$ and $\gamma = 10\ cm^{-1}$. In the model, we only change the thickness of the sample.

		Dirac Semimetal	
		$\mu = 5$ meV	$\mu = 25$ meV
$\gamma = 10\ \mathrm{cm}^{-1}$	$\epsilon(1\ \mathrm{THz})$	$-4200 + 1930i$	$-25800 + 7950i$
	$\lambda_P\ (\lambda_P^*), Q$	3.96 (1.90) μm, 2.18	24.3 (11.7) μm, 3.25
$\gamma = 30\ \mathrm{cm}^{-1}$	$\epsilon(1\ \mathrm{THz})$	$-2270 + 3810i$	$-15450 + 14400i$
	$\lambda_P\ (\lambda_P^*), Q$	2.14 (1.02) μm, 0.60	14.6 (7.00) μm, 1.07

Table S1| The permittivity of the simulated materials, the unscreened (screened) plasma wavelength $\lambda_P\ (\lambda_P^*)$ and Q-factor of the screened SPs.

In Fig. S4, we maintain the dielectric property of the model material and we tune the thickness of the sample. We expect to two major impacts on the near-field signal. Firstly, the wavelength of the surface plasmon is changed due to its straightforward proportionality with the sample thickness. Secondly, the bulk near-field response of the material changes with the thickness. The first part is analyzed in Fig. 5 of the main text. The second part is summarized in Fig. S4. With increasing thickness, the near-field amplitude of the thin metal increases and the phase decreases.

Supplementary Note 3: Topography of WTe_2 microcrystals

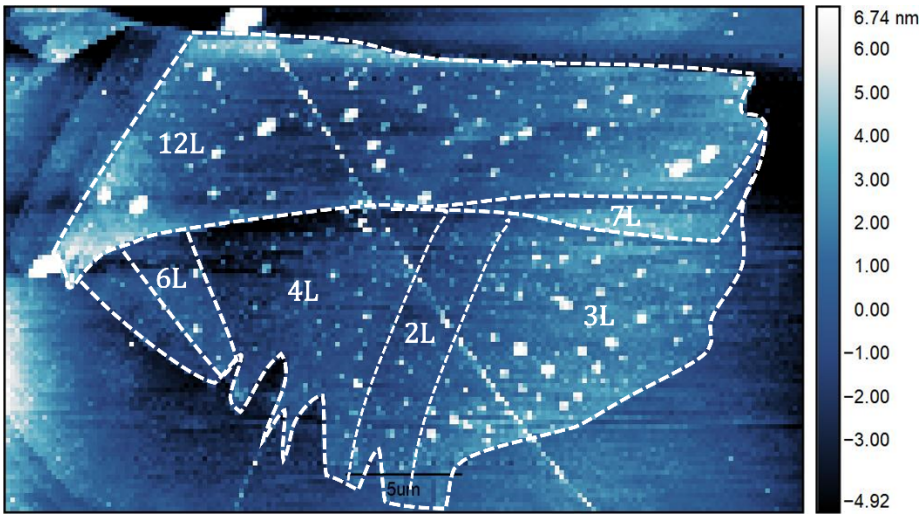

Figure S5|Topography of WTe_2 device.

We show the topography of the WTe_2 device. The topographic noise of our AFM measurement is ~200 pm, enabling us to resolve single layers. In this image, we can clearly observe tiny defects on the device surface. Being able to resolve these defects indicates the good resolution of our AFM. Because all features in the AFM image exhibits as singular feature, rather than double or multiple features, we can ascertain that the AFM tip is a sharp tip with only one apex.

Supplementary Note 4: Spectrum and hyper-spectroscopy.

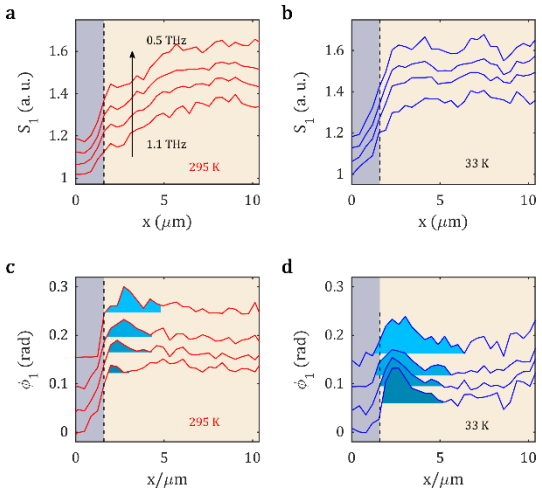

Figure S6| Frequency-dependent linecuts across the 12L WTe_2. The grey area marks the SiO_2/Si substrate. The yellow area marks the 12L WTe_2.

We acquired time-domain spectra across the 12L WTe_2 and performed Fourier-transformation. In Fig. S5, we plot the linecuts at frequencies 0.5 THz to 1.1 THz at 295 K and 33 K. In the amplitude at room temperature (Fig. S5a), the signal is higher in the interior of 12 L. The spatial dependence is nearly absent in the lower temperature result (Fig. S5b). The frequency dependence of the amplitude is small. In the phase channel (Fig. S5c and d), the edge peak is present at both temperatures. The low temperature peaks are wider and more prominent than high temperature ones, in agreement with the WL imaging. More interestingly, we observed a strong negative dependence of the width of edge peak with the frequency at low temperature. According to Fig. 5d in the main text, the surface plasmon (SP) wavelength is longer at lower frequencies compared to higher frequencies.

Supplementary Note 5: Phase-resolved imaging at different time-delay

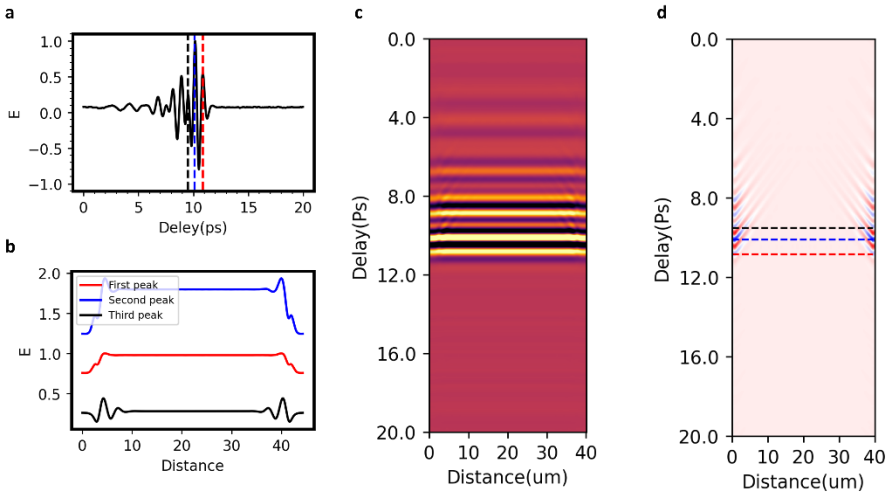

Figure S7| Phase-resolved imaging at different time-delay. **a** The center of the time-delay modulation can be chosen at different location in the time-domain spectra. Here we choose first three peaks to simulate the phase-resolved signal (dashed lines). **b** The simulated phase linecuts with the center of the time-delay modulation is located at the first three peaks. The linecut extends from one side of the sample to the other. **c** The amplitude and **d** the phase simulated at all time-delays.

In the main text, we presented experimental results and simulations with the center of the time-delay modulation at the main peak. In many cases, the main peak (the peak with the maximum E field) is located near the beginning of the whole pulse. Depending on the spectral range and the phase of different colors, the pulse can extend from a broadband half cycle pulse to a quasi-monochromatic multi-cycle pulse. Even though applying the phase-resolved technique at the main peak produces the best signal to noise ratio in both the amplitude and the phase, the plasmonic patterns may be further damped artificially.

The reason of the artificial damping arises from the propagation of the plasmon. When detecting at the main peak, if there is no plasmon launched by the tip, reflected by the sample edge, and propagate back to the tip, the plasmon feature cannot be imaged. In Fig. S6a, we simulated the near-field amplitude and phase with the center of the time-delay modulation at the first three main peaks. The result is shown in Fig. S6b. When the center of the modulation at the first peak (red), we can only see short-ranged oscillations with very small amplitude, which resembles a highly damped plasmon. When the center of the modulation is shifted to the third peak (black), the amplitude of the oscillation is stronger, and the quality factor of the pattern is higher. Before the first peak, the plasmon is almost not launched (E~0), therefore, the curve is close to a flat line. In contrast, when the plasmon is already launched by the first two strong peaks, the signal at the third peak observes the plasmon ripple propagating back to the tip. In Fig. S6c and d, we simulated the amplitude and phase at all time-delays. In the phase channel (Fig. S6d), we can clearly see the propagation (diagonal lines) of the plasmon launched by THz pulse. They are damped after a short period of time. To completely capture the damping process, the data should be collected at later time-delays.

We should note that, even though collecting phase-resolved signal at later time-delays produces more generic plasmon patterns, the absolute values of the amplitude and phase become complex. As is shown in Fig. S6b, the simulations at the first two peaks produce a finite contrast between the sample and the substrate, however, the simulation at the third peak produces nearly no contrast. The phenomenon is caused by the progression of the phase of spectral components with different frequencies. At the main peak, the phases of all frequency components are nearly identical with each other. Therefore, we can approximate the amplitude as an integration of the amplitude in the frequency domain. The phase can be approximated by an average of the phase in the frequency domain if the phases are not drastically different among all frequency components. With the progression of time, the phases of components with different frequencies progress at different pace. Hence, the phases of all frequency components are not nearly in-phase anymore at later time delays than the main peak. As a result, the amplitude and the phase collected at later time-delay is an integration of all frequency components with drastically different phases. Therefore, the absolute values become meaningless.

In conclusion, if we focus on extracting the faithful optical response devoid of plasmon, we should park the time-delay at the earlier time points. If we want to image the plasmonic feature, we prefer extended pulses and collect the signal at later time-delays.

Supplementary Note 6: Estimation of temperature and layer number dependence of permittivity of WTe_2

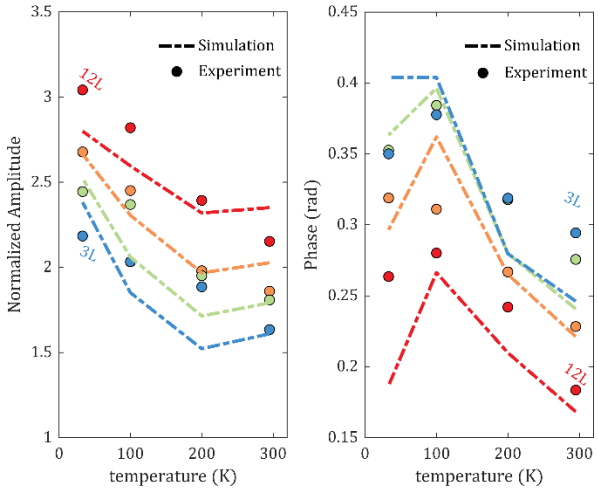

Figure S8|Comparison of the experiment and simulation result. Using the wavelength and Q-factor listed in Table. S2, we simulated the temperature and layer number dependence of the near-field signal (dashed line).

T (K)	33	100	200	295
λ_{12L} (μm)	20	9	2.5	0.8
Q	1	0.75	0.2	0.05
ϵ	$-760(200)$ $+ i760(200)$	$-340(100)$ $+ 450(150)i$	$-90(30)$ $+ i480(150)$	$-30(5)$ $+ 600(200)i$

Table S2|Temperature dependent permittivity used in the simulation.

By comparing the averaged amplitude, phase and their real-space pattern, we estimated the permittivities of WTe_2 (Table. S2). At low temperatures, both the wavelength and Q-factor increase dramatically. In Fig. S8, we compare the averaged amplitude and phase extracted in experiment and simulation. The gradual increase of intensity at low temperature is mainly caused by the increase of surface plasma wavelength (or the decrease of ϵ_1). The increase of the phase is mainly caused by the increase of the Q-factor. Interestingly, At the base temperature, the experiment show that the phase is not higher than that at 100 K. This feature is also captured by the simulation.

Bibliography

[1] P. Gallagher, C.-S. Yang, T. Lyu, F. Tian, R. Kou, H. Zhang, K. Watanabe, T. Taniguchi and F. Wang, "Quantum-critical conductivity of the Dirac fluid in graphene," *Science,* vol. 364, p. 158–162, 2019.