Phase-resolved terahertz self-detection near-field microscopy

MARIA C. GIORDANO,1 STEFAN MASTEL,2 CLEMENS LIEWALD,3,4 LORENZO L. COLUMBO,5,6 MASSIMO BRAMBILLA,5,7 LEONARDO VITI,1 ANTONIO POLITANO,6 KAI ZHANG,6 LIANHE LI,10 A. GILES DAVIES,10 EDMUND H. LINFIELD,10 RAINER HILLENBRAND,11,12 FRITZ KEILMANN,3 GAETANO SCAMARCIO,5,7,* AND MIRIAM S. VITIELLO1

1NEST, CNR - Istituto Nanoscienze and Scuola Normale Superiore, Piazza San Silvestro 12, I-56127 Pisa, Italy
2CIC nanoGUNE Consolider, 20018 Donostia—San Sebastian, Spain
3Fakultät für Physik & Center for NanoScience (CeNS), Ludwig-Maximilians-Universität München, Geschwister-Scholl-Platz 1, 80539 Munich, Germany
4Nanosystems Initiative Munich, Schellingstrasse 4, 80799 Munich, Germany
5CNR- Istituto di Fotonica e Nanotecnologie, Sede di Bari, via Amendola 173, I-70126 Bari, Italy
6Dipartimento di Elettronica e Telecommunicazioni, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
7Dipartimento Interateneo di Fisica, Università degli Studi e Politecnico di Bari, via Amendola 173, I-70126 Bari, Italy
8Fondazione Istituto Italiano di Tecnologia, Graphene Labs, via Morego 30, 16163, Genova, Italy
9i-Lab, Sichou Institute of Nano-Tech and Nano-Bionics, Chinese Academy of Sciences, Sichou, Jiangsu, China
10School of Electronic and Electrical Engineering, University of Leeds, Leeds LS2 9JT, UK
11IKERBASQUE, Basque Foundation for Science, 48013 Bilbao, Spain
12CIC nanoGUNE and UPV/EHU, 20018 Donostia-San Sebastian, Spain

Abstract: At terahertz (THz) frequencies, scattering-type scanning near-field optical microscopy (s-SNOM) based on continuous wave sources mostly relies on cryogenic and bulky detectors, which represents a major constraint for its practical application. Here, we devise a THz s-SNOM system that provides both amplitude and phase contrast and achieves nanoscale (60-70nm) in-plane spatial resolution. It features a quantum cascade laser that simultaneously emits THz frequency light and senses the backscattered optical field through a voltage modulation induced inherently through the self-mixing technique. We demonstrate its performance by probing a phonon-polariton-resonant CsBr crystal and doped black phosphorus flakes.

© 2018 Optical Society of America under the terms of the OSA Open Access Publishing Agreement

OCIS codes: (180.4243) Near-field microscopy; (180.5810) Scanning microscopy; (140.5965) Semiconductor lasers, quantum cascade; (110.6795) Terahertz imaging; (110.3175) Interferometric imaging; (310.6628) Subwavelength structures, nanostructures.

References and links
1. Introduction

Terahertz (THz) frequency electromagnetic waves (30–300 µm wavelength) can resonantly interact with fundamental excitations of molecules and solids and thus offer an ideal tool for the optical characterization of emerging low-dimensional materials and biological-systems. Scattering-type scanning near-field optical microscopy (s-SNOM) displayed an exceptional potential for nanoscale imaging of material properties [1] as has been demonstrated also at 2.5 THz [2]. In s-SNOM, an incident light beam is focused on a sharp atomic force microscope (AFM) metallic tip strongly confining the radiation to the near-field region of its nanometric apex. Nanoscale resolved (10 - 100 nm) optical images can be retrieved by analyzing the scattered radiation as a function of tip position, placing the tip in close proximity to the sample surface.

Coherent imaging, i.e. detection of amplitude and phase contrast, is crucial to get information on the real and imaginary parts of the dielectric response and hence several interferometric approaches have been developed allowing amplitude and phase resolved s-SNOM imaging in the visible and infrared spectral ranges [3, 4]. Solutions based on fs pulsed laser sources and electro-optic sampling detection in the THz range [5, 6] or solutions based on microwave circuitry in the sub-THz range [7] have been reported. However, in spite of a strong demand to extend in the far-infrared the spectral coverage of coherent nano-imaging based on continuous-wave (CW) compact sources, conventional interferometric techniques [2, 8, 9] have suffered from the poor dynamic range of cryogenically cooled bolometric detectors needed to measure the typically small s-SNOM signals [2]. Progress in coherent THz nano-imaging would therefore greatly benefit from compact, room-temperature operating and fast detection systems for THz frequency operation.

Here we tackle the problem by conceiving a simple, potentially fast and compact s-SNOM system based on a THz quantum cascade laser (QCL) operated CW in the self-detection (SD) mode. A similar approach has been recently exploited to produce THz images with sub-wavelength spatial resolution [10, 11]. In this work, we demonstrate amplitude- and phase-resolved background-free SD-s-SNOM imaging with 60-70 nm spatial resolution comparable to the scattering tip size, providing a key step forward to make THz nanoscopy a widely used tool. Our approach, based on a simple 2-parameter fitting of self-mixing interferograms, outperforms by far all previously reported attempts based on the use of self-mixing and either attaining incoherent near-field s-SNOM imaging, or using a rather complex 6-parameter fitting procedure for amplitude-like and phase-like images at diffraction-limited resolutions (>100 µm) [12].

2. Results and discussion

2.1 The general concept of SD-s-SNOM

In our experimental configuration, we use the radiation source, a THz quantum cascade laser (QCL) [13], (see Experimental Methods) not only for illuminating the tip but also for detecting the radiation field scattered from the tip, therefore avoiding the use of bulky THz cryogenic detectors. The near-field scattering is measured through the voltage change (ΔV) across the electrical contacts of the THz QCL, induced by the self-mixing effect [14, 15]. This effect is based on the reinjection of a small fraction (10⁻⁴ to 10⁻²) of the emitted field that coherently interferes within the laser cavity.
The inherent stability of QCLs against optical feedback [16] has been recently exploited in a number of self-mixing interferometry (SMI) configurations, providing an interesting method to control the emission of THz QCLs by reconfigurable photo-generated anisotropic metamaterials [17], to trace the free carrier distribution in a semiconductor target [18], and to map the real and imaginary refractive index of polymeric materials [12], for example.

In our system (Fig. 1(a)), the 2.7 THz QCL radiation is focused onto the apex of a Pt tip, which is sinusoidally dithered, normally and in close proximity to the sample surface at frequency $\Omega$. The scattered field $E_s$ is collected by a parabolic mirror and focused back onto the QCL front facet along the same incident optical path, inducing the changes in the QCL voltage ($\Delta V$) (see details on the experimental arrangement in Appendix 1). Hence, $\Delta V$ measures the optical response of the sample. To quantify this properly, we relied on the well-established Lang-Kobayashi (LK) model [19], and extended it to encompass a complex sample permittivity. In the very weak feedback limit, the voltage change at the QCL terminals can be written (see Appendix 2):

$$\Delta V \propto s \cos(\omega_0 \frac{2L}{c} - \varphi)$$  \hspace{1cm} (1)

where $s$ and $\varphi$ are the amplitude and phase, respectively, of the ratio between scattered and incident electrical fields of THz radiation, $\omega_0 = 2\pi c / \lambda$ is the unperturbed laser frequency, and $L$ is the laser-to-tip distance which varies by the tip dithering and by the piezoelectric mirror (PZM) displacement. By using an optical attenuator ($A$), the feedback remains sufficiently low to keep the system in the validity range of Eq. (1). In this regime, the proposed experimental layout is similar to a single-arm homodyne interferometer [14], in which the laser output facet and the tip define the external cavity of our self-mixing interferometer.

Equation (1) contains all necessary physical quantities to predict observable amplitude and phase contrasts in SD-s-SNOM images. Although in conventional s-SNOM, the phase contrast is extracted by analyzing the detector signal variations induced by changing the optical path length of the reference beam of an interferometer, in SD-s-SNOM the amplitude and phase information is obtained from the analysis of the $\Delta V$ as a function of $\Delta L$ as explained below.
Fig. 1. Self-detection scattering type near field optical microscope with nanometer resolution at terahertz frequencies. (a) Schematic diagram showing the experimental arrangements. The same optical components (two paraboloid mirrors, an attenuator (A), one fixed mirror and one piezo-actuated mirror (PZM)) focus the THz beam emitted by the QCL onto the apex of a Pt tip sinusoidally dithered at frequency $\Omega$ in close proximity to the sample surface and feeds the scattered radiation back into the QCL cavity to produce self-mixing. The distance $L \approx 60$ cm between the QCL front facet and the tip can be varied by translating the PZM. (b) Self-mixing induced modulation from a Au surface of the voltage drop across the QCL obtained using the experimental arrangement (a) and measured by a lock-in amplifier to the $n = 3$ demodulation order, as a function of $\Delta L$. The dots are the experimental data and the solid line is the best-fitted curve obtained using Eq. (2). The $\Delta V_3$ signal show the expected $\lambda/2$ periodicity characteristic of self-mixing. (c) AFM topographic and (d) 3rd harmonic near-field THz imaging of the Au-on-silicon sample at a fixed mirror position. (e) Edge response profile extracted from (d) by averaging seven adjacent rows of pixels along the green horizontal line. The inset shows the corresponding spatial derivative, which has been fitted by a Gaussian function (red curve) with a full-width-at-half-maximum of 65 nm.

2.2 The implementation of SD-s-SNOM

In our arrangement, as is typical in s-SNOM for effectively distinguishing the near-field scattering from background scattering \cite{1}, $\Delta V$ is demodulated at low-order harmonics of the dither frequency using a lock-in amplifier, and the measured harmonic amplitudes $\Delta V_n$ are recorded up to the order $n = 5$ as a function of the displacement $\Delta L$. Decomposing $\Delta V$ into harmonic components $\Delta V = \sum_{n=0}^{\infty} \Delta V_n \cos(n\Omega t)$, we show that the demodulated signals equivalently carry information on the optical amplitude and phase components $s_n$ and $\varphi_n$ (see Appendix 2):

$$\Delta V_n = s_n \cos(\omega_0 \frac{2L(t)}{c} - \varphi_n)$$  \hspace{1cm} (2)

The experimental $\Delta V_n$ vs $\Delta L$ curves (e.g. $\Delta V_3$ in Fig. 1(b)) demonstrates that our SD-s-SNOM system behaves similarly to an external interferometer and produces approximately
sinusoidal interference fringes, as predicted by Eq. (2). Figures 1(c)-(d) provide direct proof that the SD-s-SNOM signal is sufficiently stable for homodyne THz imaging.

2.3 Deep sub-wavelength in-plane spatial resolution

Comparison between the AFM (Fig. 1(c)) and the self-detected (Fig. 1(d)) images of a gold film deposited on a SiO₂-coated silicon substrate demonstrates the capability of our THz imaging system to achieve an in-plane resolution comparable with the employed tip apex sizes (50-70nm). Figure 1(e) shows the $\Delta V_3$ signal collected along the green horizontal line in panel 1d traversing the edge of the gold-coated region. The inset of Fig. 1(e) shows the corresponding first-order spatial derivative. From the full-width-at-half-maximum of the Gaussian curve interpolating the derivative function (red curve in the inset of Fig. 1(e)), we can retrieve a remarkable spatial resolution of $\sigma_x = 68$ nm along the horizontal (fast) axis, corresponding to $\sim \lambda/1500$.

2.4 Amplitude and phase contrast imaging capability

The phase-contrast at THz frequencies between gold and SiO₂ is negligibly small, as is usual for all combinations of non-resonant materials [9]. Hence, to provide a proof-of-principle of the amplitude and phase contrast imaging capability of the SD-s-SNOM, we selected a polar crystal (CsBr), which exhibits a strong phonon-polariton (Reststrahlen) resonance in the range 2.2 – 3.3 THz [20]. We simulated the SD-s-SNOM near-field amplitude ($s_3$) (Fig. 2(a)) and phase ($\phi_3$) signals (Fig. 2(b)) for both CsBr and Au as a function of wave number, using a single phonon oscillator model in the framework of the finite dipole model (see Appendix 2). Our findings show that, although Au exhibits the expected flat amplitude response and negligible phase throughout the THz frequency range investigated, CsBr shows a four-times stronger $s_3$ peak at 90 cm⁻¹, and a large phase signal $\phi_3 = 150^\circ$ at this wave number. Hence, we prepared a suitable sample comprising a CsBr crystal, coated with a thin (100 nm) Au film, and selected a THz QCL that operated single mode in continuous wave at 90.3 cm⁻¹ (2.7 THz). These choices should ensure a strong near-field amplitude and phase contrast at CsBr/Au steps obtained by mechanically scratching the Au thin film to expose the underlying CsBr surface.

Figures 2(c)-2(d) compare the AFM topography (Fig. 2(c)) and the SD-s-SNOM image (Fig. 2(d)) of a CsBr/Au step, simultaneously obtained while raster-scanning in the x-y plane underneath the tip and recording $\Delta V_3$ at each pixel. To show the $L$ dependence of both the $\Delta V_3$ signal and the optical contrast between CsBr and Au, we changed the optical phase $0^\circ - \pi$ over a $\pi/3$ range in $\pi/2$ steps, every $\Delta y = 0.5 \mu m$ during the acquisition. We thus observe 13 striped regions in Fig. 2(d). In order to retrieve amplitude and phase from the collected optical images, each line in Fig. 2(c) and 2(d) were preliminarily horizontally shifted to straighten the CsBr/Au edge. Then the signal $\Delta V_3$ was vertically averaged in each stripe and associated with the corresponding $\Delta L$ value. Finally, sinusoidal functions (Eq. (2)) were fitted to the $<\Delta V_3>$ vs $\Delta L$ curves to extract the amplitude $s_3$ and phase $\phi_3$ as a function of the horizontal distance from the CsBr/Au edge (Fig. 2(f) and Fig. 2(g)). Then the height (Fig. 2(e)) and the signal $\Delta V_3$ were vertically averaged on each stripe and associated with the corresponding $L$ value. From the steep rise of the optical signal at the edge the Au film on the CsBr crystal, we can estimate the optical spatial resolution. To address the issue we have followed the same procedure detailed for Fig. 1(e). The full-width-at-half-maximum of the Gaussian curves interpolating the derivative function of the amplitude and phase signals (red curves in the insets of Figs. 2(f) and 2(g)), give estimates of the spatial resolutions: 54 nm and 60 nm for the amplitude and phase signals, respectively.
In Fig. 3, we demonstrate the strong sensitivity of SD-s-SNOM to changes in the amplitude and phase of the effective polarizability of the coupled tip-sample system. For this, we exploited the strong dependence of $s_3$ and $\phi_3$ on the tip-sample distance $z$ (see Appendix 2) \cite{21} and recorded $\Delta V_3$ during approach curves obtained by progressively stepping the PZM position by $\Delta L = 9.25 \, \mu\text{m}$. For each value of $z$, $s_3$ (Fig. 3(a)) and $\phi_3$ (Fig. 3(b)) were extracted by fitting sinusoidal functions (Eq. (2)) to the $\Delta V_3$ vs $L$ curves. The prevalent near-field nature of 3rd harmonic signals is shown by the monotonic decrease of both the amplitude and phase signals for increasing $z$, with $s_3$ becoming negligible and $\phi_3$ saturating for $z > 60 \, \text{nm}$. The approach curves of Fig. 3(a)-3(b) are nicely reproduced by the calculated near-field amplitudes ($s_3$) (Fig. 3(c)) and phases ($\phi_3$) (Fig. 3(d)) of CsBr and Au, solving the full set of transcendental coupled LK equations, with no approximations, at each experimental $L$ value (see Appendix 2). Similarly, our technique is expected to measure changes of the real and imaginary part of the effective polarizability as a function of the incident laser wavelength.
2.5 Application for 2D material imaging

Finally, we demonstrate the capability of SD-s-SNOM to image doped van der Waals layered materials. We selected black phosphorus (BP), an emerging layered semiconductor, which has been recently demonstrated to be a suitable material system in the THz for photodetectors [22, 23], and near-field optical probes [25]. Here we study BP flakes with different thickness and carrier densities (see the estimate of free carrier density in black phosphorus flakes in Experimental Methods section) that were transferred onto a Si/SiO$_2$ substrate by mechanical exfoliation [23, 26]. A thin (5 nm) SiO$_2$ protection layer was deposited via sputtering to encapsulate the material system and avoid degradation under ambient exposure, ensuring that flakes remained clean and stable for several months [22, 23].

Figure 4 compares the topography images (Figs. 4(a), 4(c), and 4(e) and the corresponding $\Delta V_3$ signals (Figs. 4(b), 4(d), and 4(f) collected on Se-doped BP flakes. A near-field contrast is clearly observed between the Se-doped BP flakes and the SiO$_2$ substrate in the 2D scans of $\Delta V_3$ shown in Fig. 4(b), 4(d). This effect is also evident in the profile of $\Delta V_3$ measured across the edge of the Se-doped BP flakes (Figs. 4(l) and 4(m)). In Fig. 4(f) the optical phase was changed at constant steps of 0.39 $\pi$ ($\Delta L = 5.5 \mu m$), every $\Delta y = 0.21 \mu m$ during the scan.

To extract the amplitude and phase from the optical images, each line in Fig. 4(e) and 4(f) were preliminarily horizontally shifted to straighten the BP/SiO$_2$ edge, in analogy to the analysis performed on the CsBr/Au sample (Figs. 2(c)-2(g)). Then, the height (Fig. 4(g)) was vertically averaged on each stripe and associated with the corresponding $L$ value. From the striped image of Fig. 4(f) we extracted the near field amplitude $s_3$ (Fig. 4(h)) and phase $\varphi_3$ (Fig. 4(i)) profiles at the BP/SiO$_2$ step edge. Apparently, no phase contrast between BP and SiO$_2$ is observed, due to the off-resonance excitation of both materials, as expected since in both cases optical phonon resonances fall in the mid-IR [27]. Incidentally, the detected near-field scan results highly sensitive to the presence of oxidation-induced clusters on the surface of the investigated BP flake, which causes an abrupt change of the detected optical signal.
Fig. 4. SD s-SNOM nanoscopy at THz frequencies of doped black-phosphorus. (a, c, e) AFM tomographic images and (b, d, f) corresponding $\Delta V_3$ near field optical signal collected on a set of representative Se-doped BP flakes having thickness 130 nm (a), 300 nm (c) 100 nm (e); (f) Self- detected near field 3rd order signal. Every $\Delta y = 0.21 \mu m$ the PZM mirror position is changed in order to shorten the optical path by $\Delta L = 5.5 \mu m$. (g-i) average topography (g), amplitude $s_3$ (h), and phase $\phi_3$ (i) as a function of the position with respect to the Se-doped BP/ SiO$_2$ edge. The sharp features in panels (h) and (i) at $\Delta x = 0$ are known edge artifacts arising at the sharp edges of the mechanical exfoliated black phosphorus flakes. The dashed vertical line in panels (h-i) corresponds to the BP/SiO$_2$ edge. The horizontal dashed lines in panel (g) mark the amplitude contrast between the two materials. (l-m) $\Delta V_3$ profiles averaged over seven adjacent rows of pixels along the green lines in panels (b) and (d), respectively.

3. Conclusion

In conclusion, we have demonstrated a detector-less s-SNOM system operating at THz frequencies that provides both amplitude and phase contrast with deep sub-wavelength (60-70 nm) in-plane spatial resolutions. These achievements are expected to have a profound impact on the flourishing field of QCL-based THz imaging, in which near-field coherent detection has been missing, so far. The ability of SD-s-SNOM to resolve both the amplitude and phase of the THz field opens up the possibility of mapping the complex permittivity of a target using QCLs with a high spatial resolution, down to a few tens of nanometers. The proved sensitivity to thin-layered samples of our novel THz SD s-SNOM opens the way to further investigations of resonant 2D-materials and combined Van der Waals heterostructures, with potential impacts in plasmonics and optoelectronics.

Future experiments could make use of the tunable bandgap of BP [28] to control its mode-activation energy as well as exploiting resonant excitation to retrieve plasmonic features. The in-plane asymmetry of BP [29] may also allow for further tunability. Looking forward, complex heterostructures that combine BP with graphene, transition-metal dichalcogenides and hexagonal boron nitride have the potential to provide a robust technological platform for THz nanophotonics and ultrafast plasmonics.

Further development of SD-s-SNOM will also benefit from the availability of broadly tunable THz QCLs [30] and THz QCL combs [31, 32]. Also, the inherent ultrafast response of QCLs to optical feedback perturbations, associated with their unipolar nature and the ps-long lifetimes of inter-subband transitions, raises a compelling perspective for new time-resolved hyper spectral THz imaging systems with deep sub-wavelength spatial resolutions (< 100 nm or $\lambda/1000$).
The unique features of SD-s-SNOM, i.e. the inherent field sensitivity and compactness, can in principle lead to further elimination of intermediate optical elements and true access to near-field radiation, opening the way to many applications in fields ranging from biosensing to quantum optics [33]. Also, a full integration of sample and nanoscopy set-up inside the same cryostat may become possible, thereby allowing to exploit in situ the potential of coherent THz nanoscopy at cryogenic temperatures and under high magnetic fields, for understanding fundamental excitations, such as phonons, magnons, polaritons, and phase transitions in exotic semiconductor and organic nanostructures, superconductors, multiferroics, and metamaterials, just to mention a few examples.

Appendix 1

Experimental arrangement

A bound-to-continuum single longitudinal mode THz QCL emitting, at a wavelength $\lambda = 111 \mu m$ (2.7 THz, 90.3 cm$^{-1}$) with a surface plasmon waveguide, was grown by molecular beam epitaxy employing a GaAs/Al$_{0.15}$Ga$_{0.85}$As heterostructure on a nominally undoped GaAs substrate. The 10-$\mu$m-thick active region was embedded between a doped ([Si] = 3.0 $\times$ 10$^{18}$ cm$^{-3}$) 700-nm-thick GaAs bottom layer, and a doped ([Si] = 5.0 $\times$ 10$^{18}$ cm$^{-3}$) 200-nm-thick GaAs top contact layer. The GaAs semi-insulating substrate was lapped down to a thickness of 150 $\mu$m. Laser bars 1.5 mm long and 150 $\mu$m wide were cleaved and mounted on a copper bar.

The QCL was mounted in a liquid helium continuous-flow cryostat fitted with a polymethylpentene window, and maintained at a fixed heat sink temperature of 15 K. To maximize the sensitivity to coherent optical feedback [14, 15], the QCL was driven in continuous wave at a current $I = 450$ mA ($J = 132$ A/cm$^2$) which is just 5% higher than the threshold current density ($J_{th} = 125$A/cm$^2$), using a highly stable current generator (Lightwave Electronics, mod. QCL 2000). The emitted THz beam was collimated using a 90° off-axis parabolic mirror with an effective focal length of 50 mm, and reflected by a mirror mounted on a piezo-controlled translator stage into the entrance optical port of a commercial near-field microscope (Mod. NeaSNOM, Neaspec, Martinsried, Germany). A 5 dB substrate-free, metal-mesh diffractive attenuator (Mod. 224, Lasnix, Berg, Germany) was placed in the beam. In the microscope, a second paraboloid mirror with an equivalent focal length of 25 mm focused the beam onto a Pt tip (Bruker, mod. RMN-25PT300) having a nominal apex radius of 50 nm, a shank length of 80 $\mu$m, and which was sinusoidally dithered at its resonant frequency $\Omega = 15380$ Hz. The laser polarization lay in the plane containing the tip in order to induce an oscillating dipole in the tip efficiently. The average optical path length from the QCL front facet to the tip was $L \approx 60$ cm. The scattered radiation was collected by the same focusing paraboloid mirror and coupled back into the laser cavity along the same incident optical path. The voltage modulation across the QCL terminals produced by the self-mixing effect was pre-amplified using a low-noise amplifier (DL Instruments, mod. 1201) and demodulated up to the highest harmonic order ($n = 5$) allowed by the electronic card of the NeaSNOM.

The density of free carriers $n_e$ in the black phosphorus samples was estimated by performing transport analysis on field effect transistors FET having flakes with identical thickness as active channels [20–22, 24, 25]. $n_e$ is connected with the FET threshold voltage $V_{th}$ via the relation $n_e = \frac{C_G \Delta V_{th}}{q A_G}$, where $C_G$ is the gate-to-channel capacitance, $A_G$ the gated BP area, and $q$ is the electron charge. We measured $n_e = 4.0 \times 10^{17}$ cm$^{-3}$ for the flake shown in Figs 4(a)-4(b), $n_e = 7.0 \times 10^{17}$ cm$^{-3}$ for the flake shown in Figs. 4(c)-4(d) and $n_e = 1.5 \times 10^{17}$ cm$^{-3}$ for the flake shown in Figs. 4(e)-4(f).
Appendix 2

Theory of self-detection scattering-type scanning near-field optical microscopy (SD s-SNOM)

The near-field scattering is modeled according to the finite-dipole model, in which the tip, described as a spheroid having an effective length $L_t$ and apex radius $a$, polarizes the sample. The latter, in turn, acts back on the tip, yielding a near-field interaction specific to the sample material. The radiation scattered by a tip, placed with its apex at a distance $z$ from a target surface, is proportional to the effective polarizability $\alpha_{\text{eff}}$ and can be written as

$$\sigma = \frac{E_s}{E_i} = (1 + r_p)^2 \alpha_{\text{eff}},$$

where

$$\alpha_{\text{eff}} = \frac{2}{L_t} \left[ (\frac{L_t}{2L})^{\frac{3}{2}} - \left(\frac{L_t}{2L}\right)^2 \ln \left(\frac{L_t}{2L}\right) \right]$$

(3)

Here, $E_i$ and $E_s$ are the incident and scattered fields, respectively, and $r_p$ is the far-field Fresnel reflection coefficient for p-polarized incidence. The sample dielectric permittivity $\varepsilon_s$ enters through the so-called surface response function $\beta = (\varepsilon_s - 1) / (\varepsilon_s + 1)$, and $\varepsilon_t$ is the tip dielectric function. The apex-sample surface distance $z$ oscillates with frequency $\Omega$ and amplitude $2z_a$ as $z(t) = z_a + z_s \cos(\Omega t)$; $g$, is a complex factor related to the fraction of the total charge induced in the spheroid.

The complex effective scattered efficiency is represented in terms of amplitude and phase as $\sigma(t) = s(t)e^{i \phi(t)}$, where both $s$ and $\phi$ show a nonlinear dependence as a function of $z$. We have found a good agreement with the experimental results by using the values $L_t = 530$ nm, and $g = 0.98 e^{0.08i}$.

Coherent optical feedback in a semiconductor laser cavity is described by the Lang-Kobayashi (LK) model [14, 15, 35], in which two coupled differential equations describe the evolution of the excited state population $N(t)$ and the optical field $E(t)$. The crucial feedback contribution in this model is provided by an additive term, representing the retarded field re-entering the laser cavity.

The LK equations read [35]:

$$\frac{dE(t)}{dt} = \frac{1}{2} \left( 1 + i \alpha_H \right) \left[ G_a \left( N(t) - N_0 \right) - \frac{1}{\tau_p} \right] E(t) + \frac{\gamma}{\tau_c} E(t-\tau) e^{-i\omega \tau}$$

$$\frac{dN(t)}{dt} = \frac{J \eta}{q d} - \frac{N(t)}{\tau_c} - G_a \left( N(t) - N_0 \right) |E(t)|^2$$

(4)

where $N_0$ is the carrier density at transparency, $G_a$ is the gain coefficient, $\tau = 2L/c$, $\tau_p = 30$ ns and $\tau_c = 35$ ns are the photon life time and the cavity round trip time inside the QCL cavity, respectively, and $\tau_c$ is the carriers decay time ($1-10 \, \text{ps}$), $J$ is the current density, $\eta$ is the internal quantum efficiency, $d$ is the active layer thickness and $q$ is the electron charge, $\gamma$ is the feedback strength parameter, i.e. the fraction of the back-scattered field that efficiently
couples with the lasing mode. A value $\alpha_H = 1.5$ for the linewidth enhancement factor was used, which is appropriate for a THz QCL close to threshold [36, 37].

In our theoretical approach, we consider a feedback field provided only by the tip-scattered field $E_s$ and neglect any contribution from the background field. The latter is a component well-known in s-SNOM that can be experimentally suppressed by recording signals at higher harmonics of $\Omega$. Considering that the feedback path $2L$ depends on the oscillating tip position, we assume:

$$\gamma(t) \approx \sigma(t) e^{i\varphi(t)} = \sum_{n=-\infty}^{\infty} \sigma_n e^{i\Delta n} = \sum_{n=-\infty}^{\infty} s_n e^{i\varphi_n} e^{i\Delta n}.$$ (5)

The optical feedback coefficient $C$ is defined as $C = |\gamma|\sqrt{1 + \alpha_H^2} \frac{2L}{c \tau_s}$. In the very weak feedback limit, when $C = 10^{-2}$, as appropriate under our experimental condition, the perturbation induced on the laser frequency can be neglected and one can assume that the frequency of the retrojected laser coincides with $\omega_0$, radically simplifying the self-mixing formalism [14–16]. In fact, the sinusoidal line shapes unveiled in the experiments of Fig. 1 confirm that the system is in the very weak feedback self-mixing regime [19], with estimated values of the feedback coefficient $C = 0.03 – 0.1$. Hence, the LK equations are solved for the steady-state value of the carrier variation $\Delta N$ induced by the self-mixing (with respect to the free running laser value at threshold), assuming the tip and the PZM dynamics are slow with respect to field and carrier timescales:

$$\Delta N = -2 \left| \gamma \right| \cos(a_0 \frac{2L(t)}{c} - \varphi(t))$$ (6)

noting that $\Delta V$ is proportional to $\Delta N$ [16], we find the following expression:

$$\Delta V(t) \propto s(t) \cos(a_0 \frac{2L(t)}{c} - \varphi(t))$$ (7)

Due to the tip oscillation, the laser-tip distance varies as $L(t) = L_0 + 2s_\lambda t - z(t)$, where $L_0 + 2s_\lambda t$ is the laser-target distance which may vary, if the PZM is moved, on time scales much longer than the tip period $2\pi/\Omega$. Moreover, we approximate $L = L_0$ in the following, since in our experimental configuration $z = \lambda$.

By introducing the spectral representation of the scattered amplitude in Eq. (4), we can rewrite the steady state equations accordingly, where the RHS term of Eq. (7) becomes:

$$\Delta V[t] = \sum_{n} s_n \cos(\frac{2\pi}{\lambda} - \phi_n)$$

(8)

where $\Omega_n = n\Omega$, and $s_n = s_{-n}$, since $\Delta V$ is an even function of time ($\Delta V$ depends on time only through $z$). Equation (8) allows information retrieval on the medium optical response since it provides the link between the demodulated SMI signal and the phase and amplitude spectral components of the scattering amplitude $\sigma$.

$$\Delta V_n \propto s_n \cos(a_0 \frac{2L}{c} - \varphi_n).$$ (9)

The simulated plots presented in Figs. 2(a)-2(b) and 3(c)-3(d) have been obtained by solving in steady state Eq. (7) in the case of moving mirror PZM, assuming a total
displacement $\Delta L$ of three interferometric fringes, where each fringe is $\frac{\lambda}{2} = 55.5 \mu m$ corresponding to a phase shift $\frac{\omega_0 2\Delta L}{c} = 2\pi$ (Fig. 1). The oscillating laser voltage $\Delta V$ is then processed with a fast Fourier transform algorithm (equivalent to the lock-in filtering in the experiment) to recover the components $\Delta V_n$ corresponding to each mirror displacement. The amplitude and phase values $s_n$ and $\varphi_n$, $n = 1...5$, are finally extracted by a fitting procedure based on Eq. (9).

**Funding**

EU Graphene Flagship. M.S.V. acknowledges support from the European Union ERC Consolidator Grant SPRINT and from the second half of the Balzan Prize 2016 in applied photonics delivered to Federico Capasso. G.S. acknowledges support from PON SISTEMA (MIUR). A.G.D., E.H.L., and L.H.L. acknowledge support from the EPSRC (UK) (EP/J017671/1, EP/P021859/1), and from the Royal Society and Wolfson Foundation. R.H. acknowledges support from the Spanish Ministry of Economy and Competitiveness (National plan MAT2015-65525-R).