nature reviews methods primers

https://doi.org/10.1038/s43586-026-00474-7

Primer

Check for updates

Ultrafast nano-imaging and nano-spectroscopy

Branden L. Esses[1], **Daniel Sandner**[2], **Patrick A. McKee**[1], **Roland Wilcken**[1], **Jun Nishida**[3,4], **Richard L. Puro**[1] ✉ & **Markus B. Raschke**[1] ✉

Abstract

Ultrafast pump–probe nano-imaging combines scanning probe-based optical near-field microscopy with ultrafast spectroscopy to enable imaging with deep sub-wavelength spatial resolution, femtosecond temporal resolution and simultaneous spectral resolution. Ultrafast nano-imaging has gained increased attention for its ability to provide far-from-equilibrium excitation and excited-state contrast. With coherent and nonlinear probing, coupled electron, spin and lattice dynamics on elementary timescale and length scale can be resolved. Through nano-movies, ultrafast nano-imaging visualizes correlated quantum dynamics underlying the properties of solid-state materials, semiconductors, molecular electronic, photonic, photovoltaic and other functional materials. With nanometre spatial resolution, this method probes elementary dynamic processes across multiple length scales that are otherwise obscured in conventional ultrafast spectroscopy in which heterogeneities are spatially averaged. This Primer describes the theoretical background and experimental implementation of ultrafast nano-imaging; signal interpretation and modelling; representative examples and a perspective for the future development of the field.

Sections

---

[1]Department of Physics, University of Colorado, Boulder, CO, USA. [2]Physics Department, Technical University of Munich, Garching, Germany. [3]Institute for Molecular Science, National Institutes of Natural Sciences, Okazaki, Aichi, Japan. [4]The Graduate University for Advanced Studies, SOKENDAI, Hayama, Kanagawa, Japan. ✉e-mail: richard.puro@colorado.edu; markus.raschke@colorado.edu

Introduction

Many classes of materials exhibit heterogeneities on the nanoscale that can affect their macroscopic properties and associated device performance. For example, molecular, semiconductor, photonic and 2D van der Waals materials, among others, exhibit nanoscale heterogeneities arising from domain formation, grain boundaries, defects and impurities. Understanding and ultimately controlling the relationship between macroscopic functionality and underlying nanoscale material structure and dynamics has been a long-standing challenge. The fields of materials science, condensed matter and molecular physics have faced a lack of suitable experimental techniques to provide the desired combination of spectroscopic sensitivity, nanometre spatial resolution and femtosecond temporal information to probe the elementary excitations and processes that influence material properties.

Optical spectroscopy from the terahertz (THz) to extreme ultraviolet (UV) selectively probes quantum states associated with electronic, spin and lattice degrees of freedom of a material. However, far-field plane wave techniques such as electronic, Raman or infrared (IR) spectroscopies, and their associated pump–probe implementations, spatially average across macroscopic length scales (Fig. 1a). Recently, the combination of tip-based scanning probe optical microscopy with ultrafast spectroscopy has emerged as a frontier in optical imaging[1–11], providing nanometre spatial resolution through near-field localization together with femtosecond temporal resolution (Fig. 1d). The resulting ultrafast nano-movies can then resolve the spatiotemporal evolution of the electronic, spin and vibrational degrees of freedom and their coupled dynamics, which underlie material properties on elementary length scale and timescale. Ultrafast spectroscopy aims to characterize the electronic response, spin dynamics and vibrational, phonon or atomic motion that are induced and probed by optical excitations[12–19]. Importantly, it provides access to dynamics and coupling of these excitations by selectively exciting and temporally resolving their relative contributions to the material response, which is not readily possible with continuous wave spectroscopy.

In pump–probe techniques, a pump pulse (of shorter duration than the dynamics to be resolved) excites the material out of its equilibrium state. The subsequent probe pulse of varying time delays then interrogates the temporal evolution of the pump-induced transient excitations on timescales from femtoseconds to nanoseconds or longer (Fig. 1b,c). These time-resolved signatures of the material response provide insight into the physical processes underlying complex phenomena such as phase transitions (including melting, insulator–metal transitions (IMTs) and superconductivity); carrier dynamics and many-body correlations in semiconductors and their heterostructures; polaron formation in oxides and perovskites; photochemical reaction pathways; or biophysical processes (such as photosynthesis, vision or protein function). In addition, the study of coherent phenomena through nonlinear optical interactions[18,19] and extension to strong-field phenomena with intense ultrafast pulses[20,21] have been established.

Electronic processes typically occur on femtosecond timescales, with lattice and vibrational dynamics becoming prevalent over picoseconds, then thermalization and transport at the nanosecond scale and slower collective and structural processes occurring over microseconds or more (Fig. 1e). Many of these processes are defined by the low-energy scales of free-carrier dynamics (such as the Drude response), many-body electron dynamics, intersubband transitions, charge transfer and energy transfer, intermolecular coupling, polaron formation and vibrational and lattice motion (typically in the 1–100 meV range). Therefore, probing in the mid-IR (MIR) to THz spectral range (3–300 μm) is of particular interest.

Despite the proliferation of ultrafast spectroscopy, the extension to microscopy for simultaneous spatial resolution has been challenging[22–24]. The spatial resolution of plane wave spectroscopy is limited by the Abbe diffraction limit to hundreds of nanometres for visible wavelengths or micrometres for MIR wavelengths, making it difficult to resolve relevant structural features or sample heterogeneities typically in the nanometre range. Modifications to the optical, electronic, magnetic, structural or chemical properties of materials arise when the size of the system or its structural elements become comparable to elementary length scales, such as the mean-free path of electrons, coherence length of phonons, dipole–dipole interactions or other characteristic correlation lengths. Furthermore, as the dimensions of the medium shrink to the 1–100 nm range, the timescales of elementary excitations can also change, known as finite size effects. Many nanostructures gain their unique functionality from these correlations and interactions on nanometre length scales. Despite new promising developments of ultrafast microscopy based on labelling or scattering[25,26], most methods have remained limited in spatial resolution and spectroscopic specificity, let alone being able to obtain the full spatiotemporal-spectral picture of a sample response.

Optical nano-probe imaging, in the form of tip-enhanced Raman or scattering-type scanning near-field optical microscopy (s-SNOM), has emerged as a frontier in optical imaging, enabling few-nanometre spatial resolution through near-field enhancement, localization and coupling between the tip and sample[1,10,27–30]. Probing in the optical near field overcomes the conventional diffraction limit of far-field microscopy, providing wavelength-independent, deep sub-diffraction-limited spatial resolution[28]. In combination with a wide range of optical spectroscopies — including linear (visible to far-IR or THz), inelastic (Raman), coherent and nonlinear techniques of second harmonic generation (SHG), four-wave mixing, coherent anti-Stokes Raman spectroscopy and ultrafast spectroscopy — optical nano-probe imaging provides as fast as few-femtosecond temporal resolution to probe and control structure, coupling and dynamics of elementary excitations in materials on the nanoscale (Fig. 1). Several other nano-imaging techniques with ultrafast capabilities are described in Box 1.

In s-SNOM, a pointed metallic tip provides the desired nanometre spatial resolution through near-field localization and enhancement. The tip can thus be viewed as providing a locally confined light source of nanometre dimensions and thus smaller than the micrometre-scale incident laser focus[31–34]. The incident light field induces an oscillating optical dipole in the apex of the metallic tip, which can be viewed as an optical antenna. Scanning the tip in close proximity to the sample, by means of atomic force or tunnelling current feedback, the mutual near-field tip–sample coupled polarization enables optical imaging contrast. When the tip apex is within the range of its reactive near field to the sample, a surface charge or polarization density is induced in the sample laterally confined on the scale of the apex size. The field of this induced sample polarization recursively acts back on the tip, modifying its polarization.

This coupled tip–sample polarization can formally be described by solving the optical wave equation in the quasi-static approximation with boundary conditions of the tip and sample surface[35,36]. It is equivalent to an image dipole located within the sample coupled to the tip, thus containing information about the polarizability of the sample material and its frequency-dependent dielectric response. The light scattered into the far field by the tip then contains spectroscopic

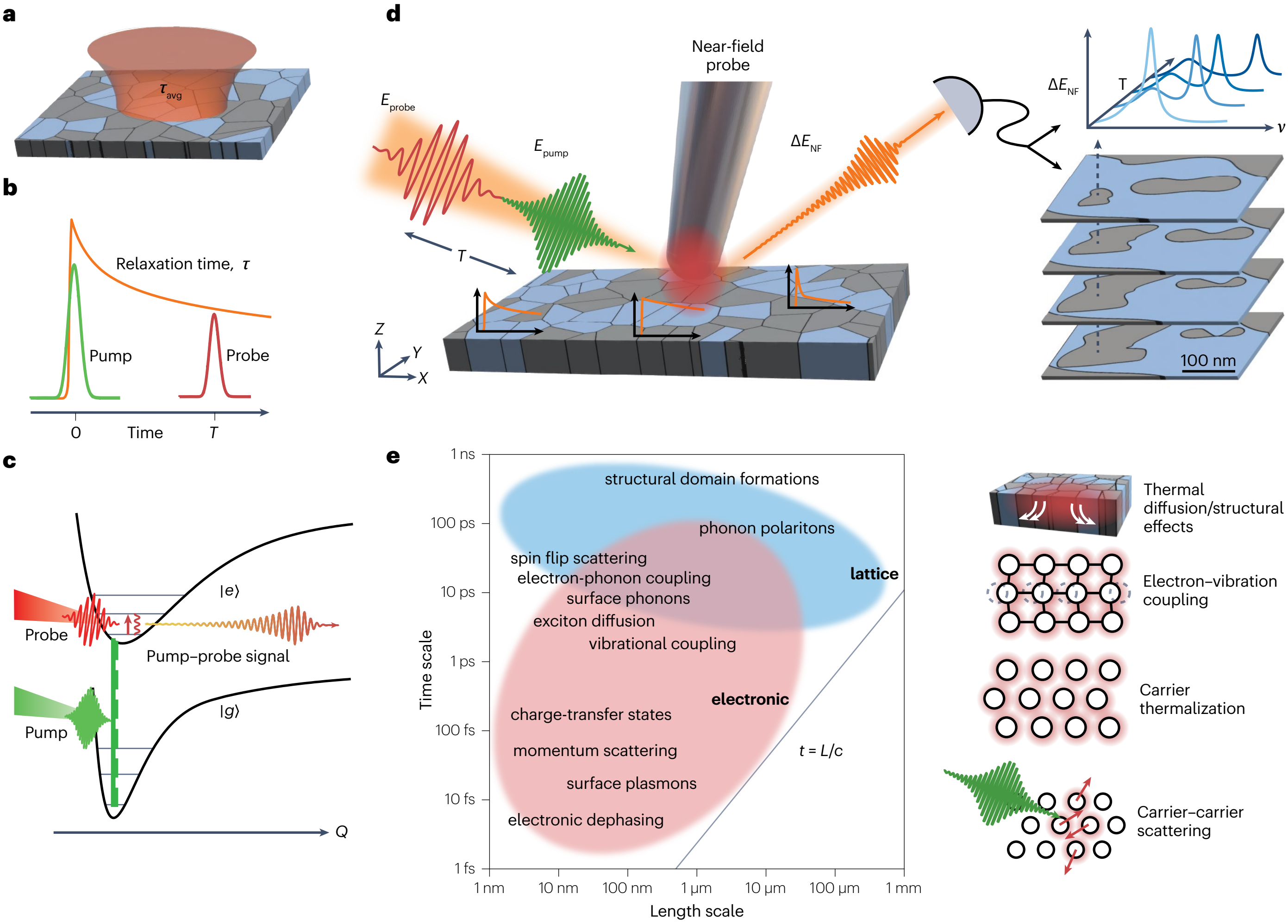

Fig. 1 | Ultrafast nano-imaging combining infrared scattering-type scanning near-field optical microscopy with pump–probe optical spectroscopy. **a**, Schematic of a far-field optical probe sampling a heterogeneous material in which the contributions of many domains are sampled simultaneously, resulting in a spatially averaged decay rate, τ_{avg}. **b**, Illustration of pump–probe spectroscopy in which a material excitation (orange) is induced by a pump pulse (green), and a subsequent probe pulse (red) at variable time delay, T, samples the relaxation to the ground state, τ. **c**, Energy level diagram showing a high-energy (green, visible or near-infrared (IR)) pump exciting a material to its electronic excited state and a subsequent low-energy (red, mid-IR or terahertz) probe exciting a vibrational transition within the excited electronic manifold with relaxation dynamics observable in the scattered pump–probe signal (orange). **d**, Ultrafast scattering-type scanning near-field optical microscopy (s-SNOM) in which a high-energy (green, visible or near-IR) pump and low-energy (red, mid-IR or terahertz) probe are focused onto a metallic atomic force microscopic tip used to spatially localize the optical fields and map the spatio-spectral dynamics (spectral ν versus temporal T versus spatial response (X, Y)) encoded in the transient near-field signal ΔE_{NF}. **e**, Representative material excitations accessible through ultrafast s-SNOM plotted based on their characteristic timescale and length scale with select examples illustrated. Panel **c** adapted from ref. 7, CC BY 4.0.

information about the sample from a region laterally and vertically confined by the dimensions of the tip apex radius, which in turn defines the spatial resolution[37–39]. The spatially localized near-field signal is then discriminated from an unspecific scattering background by modulating the tip–sample distance, which subsequently modulates the tip–sample-coupled polarization, or related gating techniques.

As a near-field technique, s-SNOM is conceptually distinct from far-field microscopy. The intrinsic spatial resolution limit in the far field, owing to diffraction of propagating radiation, is overcome in the near field through the extrinsic probe tip, defining the spatial confinement of the evanescent field. The spatial resolution is then not related to the free-space wavelength, but instead tied to the curvature of the tip apex defining the boundary conditions for the electromagnetic field. With improved tip designs and tip–sample modulation techniques, spatial resolution has progressively improved into the few-nanometre regime[40]. Furthermore, the high-momentum states accessible through s-SNOM[23,41–45] extend the selection rules compared with plane wave far-field spectroscopy beyond the aspect of spatial resolution opening probing of dipole and momentum-forbidden transitions such as surface plasmon polaritons[46,47] or surface phonon polaritons[48].

In combination with femtosecond pulsed laser sources for probing and excitation, ultrafast s-SNOM has advanced to provide simultaneous ultrahigh spatial, spectral and temporal resolution used to study a wide range of dynamic processes on nanometre length scales in

Box 1 | Nanoscale imaging techniques complementary to ultrafast scattering-type scanning near-field optical microscopy

Near-field scanning optical microscopy

Early attempts to combine fibre-based near-field scanning optical microscopy with ultrafast spectroscopy have been foundational for the development of ultrafast scattering-type scanning near-field optical microscopy[113,179–181]. On the basis of the large transition moments of excitons in low-dimensional semiconductors[113], carrier diffusion[181,182], exciton spin dynamics[183] and coherent evolution could be observed[184]. However, dispersion in the optical fibres and their narrow spectral bandwidth make aperture probes difficult to apply beyond the visible and near-infrared. Additionally, the low collection efficiency of aperture probes limits spatial resolution to ~100 nm (refs. 185,186).

Scanning tunnelling microscopy

With the ability of scanning tunnelling microscopy to probe and manipulate matter at the atomic scale, it has been desirable to combine its superior spatial resolution with ultrafast spectroscopy[187–189]. Despite many early challenges[190,191], recent success by transient terahertz modulation of the tunnelling current[192] provides ultrafast terahertz spectroscopy and imaging. Although achieving atomic resolution, this method requires conductive substrates. Additionally, extending into the infrared and visible spectral range, in which transient thermal expansion of the tip can conceal the signal of interest, is still being addressed[193].

Electron and X-ray diffraction

Ultrafast scattering-type scanning near-field optical microscopy as an all-optical technique complements many recent parallel developments in ultrafast pump–probe nano-imaging based on various electron, extreme ultraviolet and X-ray microscopies that benefit from the intrinsically short wavelength of pulses of high-energy photons or electron wave packets. These include femtosecond electron diffractive imaging[140,194,195], spatiotemporal photoemission electron microscopy[196,197] and ultrafast transmission electron microscopy[198,199]. However, although high-energy photons or electrons are desirable in some applications (such as for probing structural and lattice dynamics, resolving core levels or certain spin states directly), these sources make it difficult to resolve low-energy phenomena of molecular vibrations, intermolecular coupling, carrier dynamics and associated many-body phenomena. Furthermore, ultrafast electron and X-ray microscopy are experimentally demanding, requiring in-vacuum operation, typically with facility-scale instruments.

Light-induced force microscopy

In this group of techniques, a (non)linear optical excitation or polarization is induced in the sample and detected as a nanomechanical force response with the tip. These techniques do not require optical signal detection. Although promising, signal interpretation is still challenging because of the convolution of different light-induced tip–sample responses, measured by the force sensor (such as DC polarization, surface photovoltage and thermal expansion)[200,201].

semiconductor, molecular and quantum materials[1–3,49]. Spatiotemporal imaging resolves inhomogeneities in the ultrafast local material response owing to variations in composition, strain, doping, defects or other perturbations from nanometre to micrometre length scales[4,9]. Furthermore, it is possible to image the transient behaviour of surface phonon, plasmon, exciton or other polaritons, which are processes that cannot readily be accessed with conventional ultrafast far-field techniques[49,50]. Many surface polariton phenomena are either intrinsically confined to sub-wavelength scales or require photon momenta for their excitation larger than those provided by propagating far-field light, yet accessible in the near field.

Ultrafast s-SNOM requires a delicate balance of various experimental parameters necessitating careful consideration. This includes the choice of laser source and probe parameters, tip designs, overall optical layout, modulation and signal detection schemes and a thorough understanding of the optical contrast mechanism for artefact-free near-field signal extraction and interpretation[35,51]. Following a series of works by different groups to validate the approach in proof-of-concept experiments, ultrafast s-SNOM is now at the cusp of becoming a routine technique with an ever-growing application space across different research areas and is at the transition from method development and instrumentation science to material science application focused. Table 1 summarizes the most important imaging parameters for ultrafast s-SNOM in current implementation, with values expected to extrapolate the ongoing developments.

In this Primer, we describe the background of ultrafast pump–probe nano-imaging and suggest guidance for its implementation; discuss signal interpretation and analysis; highlight relevant example applications and provide a perspective for the future development of the field.

Experimentation

In this section, we describe the experimental principles of ultrafast pump–probe nano-imaging. The choice of laser system, the pump and probe preparation and pulse alignment, including their coupling with the scanning probe instrument and focusing onto the tip, are discussed (Fig. 2a). We also outline near-field signal collection, amplification and detection, as well as pump and probe pulse modulation and signal demodulation for background contrast (Fig. 2b,c).

As in conventional far-field ultrafast spectroscopy, different excitation regimes must be distinguished, from weak ground-state perturbation at low pump fluence to strong and far-from-equilibrium excited-state pumping at high pump fluence. This distinction is particularly important in ultrafast s-SNOM and dictates much of the experimental protocol. In far-field ultrafast spectroscopy, wavevector conservation in reflection or transmission typically allows for the spatial filtering of residual pump and probe light from the transient signal of interest. By contrast, wavevector conservation is lifted in near-field scattering. Therefore, scattered pump, probe and transient signal light all reach the detection system and must be separated by spectral or temporal modulation filtering or gating.

A useful parallel can be drawn to far-field phase-cycling techniques, which isolate nonlinear signals by shifting the optical phase of the pump and probe beams to select particular excitation pathways based on their unique phase dependence[52,53]. Phase cycling disentangles the material response by assigning each nonlinear pathway a unique phase signature, allowing them to be isolated during detection. In the near field, however, all fields scattered from the tip enter the same channel. For ultrafast nano-imaging, separation commonly relies on double modulation in which detection at harmonics of the tip-tapping frequency, ω_t (typically 50–500 kHz), isolates the near-field response from far-field background scattering, whereas pump modulation at Ω_M (from 1 kHz to 10 kHz) provides contrast between the ground-state and excited-state material response. Combining both modulations generates doubly modulated sidebands ($n\omega_t \pm m\Omega_M$ with n and m positive integers) that contain only the pump-induced nano-localized perturbation. Lock-in detection at these mixed harmonics has a role analogous to far-field phase cycling. Both move the desired nonlinear response into a distinct detection channel.

In the development of ultrafast nano-imaging, early work focused on pump–probe s-SNOM performed with high-repetition rate oscillator-based femtosecond laser systems with low pulse energies in a weakly perturbative regime[4,10,54]. By contrast, the application of high pulse energy, amplified laser systems with lower repetition rate allows for far-from-equilibrium optical excitation to induce and probe longer-lived transient phenomena such as band structure collapse and Mott transition[4,9]; to separate coupling between different electronic degrees of freedom[55,56]; and to probe polaron and excited-state molecular and lattice vibrations[7]. In any experiment, one of the main challenges is to isolate the spatially localized pump-induced excited-state response from both the transient far-field response and the nano-localized ground-state response. A combination of phenomena can mask and superimpose the transient near-field signal of interest, or a far-field transient signal may crosstalk into the near-field response, leading to its misinterpretation as a nano-localized transient response.

Experimental set-up

In ground-state IR s-SNOM, the metallic tip of an atomic force microscope (AFM) is illuminated with IR light, and nanoscale optical contrast is obtained by detecting the tip-scattered near field[27,29,57,58]. This is illustrated in Fig. 2a by the red MIR probe beam. The AFM is configured in sample-scanning mode so the tip remains stationary in the laser focus during scanning. Ground-state s-SNOM can be performed with continuous wave lasers, even incoherent light sources, as well as with various pulsed lasers used for non-time-resolved measurements, which benefit from the spectral bandwidth for nano-spectroscopy. Most conveniently, both continuous wave and ultrafast s-SNOM can be set up in an epi-illumination detection configuration, in which the near field from the AFM tip is collected in the collinear backscattering direction by the same element, which focuses the light, such as an off-axis parabolic (OAP) mirror. The backscattered near-field light is then detected and can be spectrally resolved with different photodetection schemes shown in Fig. 2a. These include interferometry in which the near-field light is interfered with an MIR reference pulse of variable delay t_{ref} and detected by a photodiode such as HgCdTe (MCT), or electro-optical sampling (EOS) techniques, in which the near-field light is interfered with a near-IR (NIR) or visible reference pulse (yellow gate beam in Fig. 2a) in a nonlinear optical crystal. Interferometric and EOS detection are discussed in more detail later in this section. Ultrafast s-SNOM involves the same experimental set-up as ground-state s-SNOM, but with an additional NIR or visible-pump beam incident on the sample collinear with the probe beam at variable time delay T (green visible-pump beam in Fig. 2a).

Specific hardware requirements for ultrafast s-SNOM include pulsed laser sources, with additional nonlinear and parametric conversion applied to generate selected or tunable pump and probe wavelengths. The sample is first excited by the ultrashort pump pulse of nanosecond, picosecond or femtosecond duration depending on the desired time resolution or material resonances to be excited. The subsequent probe pulse at variable time delay, T, then samples the temporal evolution of the pump-induced perturbation and its relaxation to either the ground state or into a new photoinduced chemical, structural or electronic phase.

Mode-locked titanium–sapphire (Ti:S) or erbium-doped or ytterbium-doped fibre oscillators and subsequent laser amplifiers are commonly used as femtosecond or picosecond laser sources. To generate tunable pump and probe pulses of desired wavelengths, nonlinear processes are used for frequency conversion, combining as needed second or third harmonic generation for visible pulse generation, optical parametric amplifiers (OPAs) or optical parametric oscillators followed by difference frequency generation (DFG) for MIR pulse generation or intrapulse optical rectification for THz pulse generation. At the facility scale, picosecond and femtosecond free-electron lasers can be used as either pump or probe beams, synchronized with a table-top femtosecond laser source.

As established in ultrafast spectroscopy, and equally critical for the phase-sensitive technique of s-SNOM, accurate pre-characterization of both pump and probe pulses is important for determining the spectral phase and amplitude, using autocorrelation or cross-correlation techniques such as frequency-resolved optical gating (FROG) or its cross-correlation variant (X-FROG) (Fig. 2d,e). The X-FROG trace shown in Fig. 2d is measured using sum frequency generation (SFG) between the NIR laser output (before the visible pump is generated through SHG) and MIR-probe beams. Because X-FROG requires a known reference field, autocorrelation FROG is first performed on the NIR beam to retrieve its temporal intensity and spectral phase. The NIR pulse is then treated as a characterized gate pulse in the subsequent SFG X-FROG measurement of the MIR field. In the X-FROG acquisition shown in Fig. 2d, the relative delay T between the MIR pulse and the gated NIR pulse is scanned, and a spectrometer records the SFG spectrum for each delay, producing a 2D spectrogram $I(\omega, T)$. An SFG-based FROG retrieval algorithm is then used to recover the MIR pulse duration and spectral phase[59] (Fig. 2e). In combination with passive or active dispersion control and pulse shaping, either transform-limited pump and probe pulses or shaped pulses of a known transient waveform are desired.

For the s-SNOM instrument itself, an AFM is operated in tapping mode with metal-coated cantilever tips to facilitate the near-field localization and coupled tip–sample polarization. The microfabricated

Table 1 | General spatio-spectral-temporal characteristics of pump–probe scattering-type scanning near-field optical microscopy set-ups

Parameter	Demonstrated value (values in development)
Spatial resolution	10–20 nm (1 nm)
Time resolution	50 fs, ps and ns range (10 fs)
Spectral range	Visible, mid-IR, THz (UV–THz, especially FIR)
Drift[a]	0.1 nm min^{-1}
Field of view	Up to 20 μm

FIR, far infrared; IR, infrared; THz, terahertz; UV, ultraviolet. [a]Effective drift can be reduced to zero by rescanning and tracking.

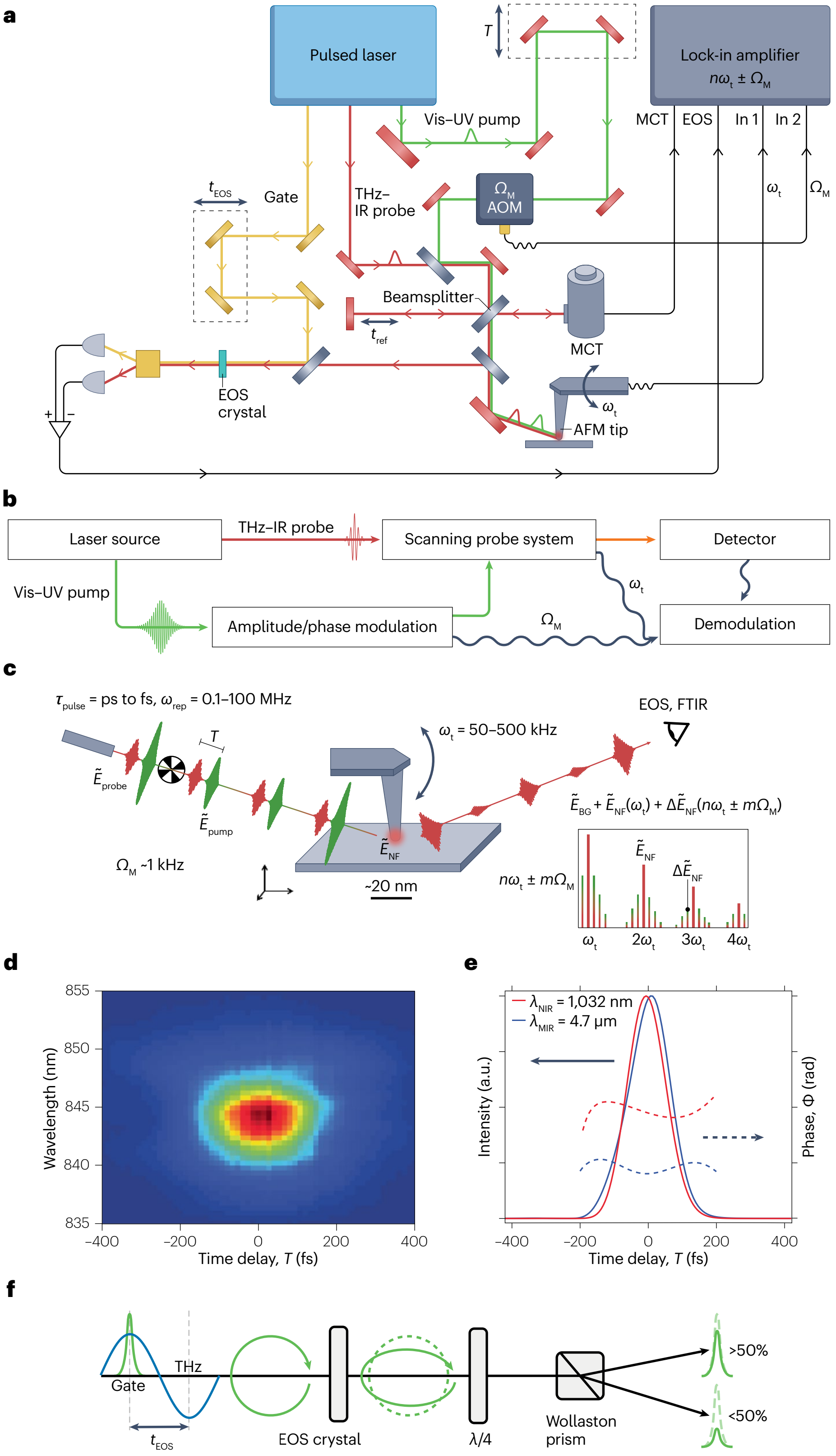

Fig. 2 | Experimental set-up for ultrafast scattering-type scanning near-field optical microscopy. a, Visible (Vis) to ultraviolet (UV) pump light (green), modulated by an acousto-optic modulator (AOM), excites a sample surface at the apex of an atomic force microscope (AFM) cantilever tip in tapping mode. Broadband and mid-infrared (IR) probe light delayed from the pump provides time-resolved access to the optical material response. The tip-scattered nano-localized near field is detected either through Fourier transform infrared (FTIR) interferometry by a photodiode (commonly with a HgCdTe (MCT) element) or through electro-optical sampling (EOS) by a gate pulse. Ω_M is the modulation frequency of the pump and ω_t is the tip-tapping frequency. **b**, The block diagram summarizes the key components of an ultrafast scattering-type scanning near-field optical microscopy set-up and its connections to one another. **c**, Diagram of modulations leading to Fourier decomposition of tip-scattered field, which is composed of far-field background scattering $\tilde{E}_{BG}$, the ground-state nano-localized near-field response $\tilde{E}_{NF}$ and the transient excited-state nano-localized near-field response $\Delta\tilde{E}_{NF}$. The pump beam (green) and thus only $\Delta\tilde{E}_{NF}$ are modulated at frequency Ω_M, whereas all scattered fields are modulated at frequency ω_t by the tip oscillation. Selecting sidebands $n\omega_t \pm m\Omega_M$ for n and m non-negative integers, lower right, isolates the transient tip-scattered field $\Delta\tilde{E}_{NF}$ from other background contributions. The pulse duration is denoted as τ. **d**,**e**, Cross-correlation frequency-resolved optical gating trace of probe pulse (part **d**) with corresponding temporal amplitude (solid lines) and phase (dashed lines) of both pump and probe pulses (part **e**). **f**, Schematic of the Pockels effect in a nonlinear crystal in which the detected terahertz (THz) field modifies the polarization of a visible gate field for EOS detection. a.u., arbitrary unit; NIR, near-infrared. Panel **a** adapted from ref. 7, CC BY 4.0. Panels **c** and **d** adapted with permission from ref. 9, American Chemical Society.

cantilever with a sharp probe tip is electrically driven by a piezo actuator to oscillate near its mechanical resonance frequency (first mode, denoted ω_t) with the tip intermittently touching the sample surface. Changes in oscillation amplitude and phase arising from tip–sample force interactions are monitored by electronic or optical readout, often by a laser beam reflected from the backside of the cantilever onto a segmented photodiode.

The AFM is designed such that the probe tip is stationary within the pump and probe beam focus, hence configured in sample scanning mode. A feedback loop controls the vertical z-piezo position of the sample scanner to keep the cantilever deflection or oscillation amplitude at a set point, and the resulting piezo-voltage is used to reconstruct the surface topography when raster scanning the sample in the xy plane. Soft materials typically require lower tapping forces, set by driving the cantilever slightly below its mechanical resonance, using high set points (80–90% of the tapping amplitude of the tip in free space), soft cantilevers and low tapping amplitudes. Harder surfaces can be imaged with higher tapping forces by driving closer to or slightly above cantilever resonance, using lower set points (60–70%), stiffer cantilevers and high tapping amplitudes. Proper AFM operation and understanding of the mechanical tip–sample interaction are critical for the success of s-SNOM imaging and even more so for ultrafast s-SNOM. The tip–sample force interaction and associated dynamic cantilever deflection, including its anharmonicity, dependence on tapping amplitude, variations with sample material and the mechanical properties themselves (such as adhesion and Young's modulus), are all critical as the near-field interaction follows the motion of the tip within the tip–sample mechanical interaction potential.

With AFM scanning, the surface topography is characterized to locate sample features of interest, as well as other mechanical properties such as modulus, adhesion, electrical conductance, surface potential or magnetic properties, which can be correlated with the local optical response. For the majority of s-SNOM applications, commercially available metallic tips coated with a Pt/Ir alloy or Pt silicide are commonly used, as well as gold-coated and silver-coated tips, which have larger field enhancement and scattering cross-section, yet lower damage threshold.

Tip illumination. The tip serves to both localize the incident optical field, $\tilde{E}_{\text{in}}$, and scatter the induced evanescent near-field radiation into propagating far-field radiation, $\tilde{E}_{\text{NF}}$. The stationary AFM tip is positioned at the focal point of preferably an OAP mirror of high numerical aperture (NA > 0.3) to avoid chromatic and spherical aberrations. Practical size constraints imposed by typical AFM heads limit the working distance to >10 mm, necessitating an incident beam diameter of >10 mm. To avoid coma aberration, the optical axis of the OAP must be aligned parallel to the incident beam. This is typically achieved with a plane mirror mounted normal to the optical axis next to the OAP, allowing the incident beam to be positioned on-axis via its back reflection. Notably, even angle deviations as small as 0.03° reduce the intensity at the tip by a factor of two[60].

The tip-scattered probe light is then collected and collimated in the backscatter direction with the same OAP or in the forward scatter direction by another one, although this set-up is less common. The light is detected either by EOS or interferometrically to obtain spectral resolution. The choice of detection scheme mainly depends on the probe wavelength with interferometric detection preferable for the MIR and NIR above 500 cm^{-1}, whereas EOS is preferable for the far-IR or THz below 500 cm^{-1}.

Pump beam. As in far-field pump–probe spectroscopy, the synchronized pump pulse must overlap spatially and temporally with the probe pulse. To achieve spatial overlap in non-degenerate pump–probe s-SNOM, it is most convenient to align the pump beam collinearly with the probe beam. This is typically achieved by using a dichroic beamsplitter to transmit the probe but reflect the pump, such as undoped germanium or silicon windows. For most practical purposes, it is preferred to polarize the pump beam perpendicular to and the probe beam parallel to the tip axis.

Interferometric near-field detection. For interferometric detection, a Michelson-type layout is used, in which the beamsplitter is inserted into the incident beam before it is focused onto the tip (Fig. 2a, red beam path). The pulse in the sample arm is focused onto the tip, followed by the collimated backscattered light, $\tilde{E}_{\text{NF}}$, directed to and focused on the photodetector. Typically for MIR, a cryogenically cooled MCT element is used, but depending on the spectral range, indium antimonide, indium arsenide or bolometric detectors can be used as well[61]. Importantly, the speed of the detector must exceed the largest desired harmonic of the tip-tapping frequency, ω_t.

The reference arm of the interferometer provides a reference field, $\tilde{E}_{\text{ref}}$, with tunable relative phase. A retroreflector mounted on a translation stage is used to adjust the length of the reference arm relative to the sample arm, generating a time-integrated intensity interferogram. The light from the reference arm is focused on the detector collinear with the sample beam.

In contrast to conventional Fourier transform infrared (FTIR) spectroscopy, the interferometer described earlier is asymmetric owing to the location of the tip and sample in one arm of the interferometer. Scanning the reference arm, the resulting interferograms are asymmetric containing a tail in the case of resonant sample interactions only at positive delays. Asymmetric interferometry provides both the amplitude and the phase of $\tilde{E}_{\text{NF}}$. Only the frequency components contained in both the reference and sample signals can be reconstructed through Fourier analysis, limiting the bandwidth of detection to that of the incident field.

Electro-optic near-field detection. In EOS of the tip-scattered field, $\tilde{E}_{\text{NF}}$, a short gate pulse of variable delay interacts with the signal field in a nonlinear crystal. In the THz spectral range, this is commonly implemented via second-order ($\chi^{(2)}$) nonlinearities using the Pockels effect, in which the signal field induces a birefringence in the nonlinear optical crystal (such as lithium niobate, zinc telluride, gallium arsenide or gallium phosphide)[62]. Importantly, the gate pulse duration must be shorter than a half cycle of the highest frequency of the IR or THz pulse to be characterized, limiting the approach for most practical purposes to wavelengths >10 µm (30 fs cycle period necessitating ~15 fs gate pulse duration). By use of a Wollaston prism or a polarizing beamsplitter, the polarization state of the gate pulse is projected onto two photodiodes whose difference signal represents the gate pulse polarization and thus the strength and direction of $\tilde{E}_{\text{NF}}$ at the time where both pulses overlap in the crystal (Fig. 2f). The amplitude and phase of $\tilde{E}_{\text{NF}}$ are then reconstructed by scanning the time delay of the gate pulse across its duration.

Comparison between interferometric intensity and direct field measurement. EOS projects the THz detection into the visible to NIR where detectors are faster, more sensitive and operate at room temperature, compared with IR detectors. Additionally, EOS can resolve

Table 2 | Comparison of different detection schemes for pump–probe scattering-type scanning near-field optical microscopy

Detection scheme	Time resolution[a] (fs)	Bandwidth[a] (cm^{-1})	Spectral range[b,c] (cm^{-1})	Spectral acquisition time	Delay scan acquisition time[d]	Image frame acquisition time[d]
Interferometric detection (nano-FTIR)	50–200	150–250 (laser-based) or ~1,000 (synchrotron-based)	>800	30 min (depends on spectral resolution)	5–10 h	Not applicable
Interferometric detection (two-phase)	50–200	100–200 (<10% laser frequency)	>800	Not applicable	1–10 min	50 ms per pixel, 10–30 min per frame
Electro-optic sampling	10–100	20–50 (at few THz) or 100–400 (far-IR to mid-IR)	1 to $\omega_{\mathrm{max}} = \frac{1}{\tau_{\mathrm{Gate}}}$	150 min	9 min	40–100 min per frame
Pseudo-heterodyne	100	150	>800	Not applicable	Similar to two-phase acquisition time	Similar to two-phase acquisition time

THz, terahertz. [a]Pseudo-heterodyne works best for small spectral width ($\Delta\omega < \omega/100$, in which ω is the frequency), thus requiring longer pulses. [b]Fourier transform infrared (FTIR) works from far infrared (IR) up to near-IR or visible; low frequencies are limited by detectors. [c]Assuming 3 points per optical cycle, maximum resolvable frequency is limited by $\bar{\omega}_{\mathrm{max}} = 1/\tau_{\mathrm{Gate}}$, for τ_{Gate} the electro-optic sampling gate pulse duration. [d]In simplest implementation, two similar measurements (time traces per movie frame) at different positions of the reference arm.

THz radiation with photon energies below the bandgap of common detectors and is not limited to detecting frequencies contained in the probe pulse incident on the tip (for example, for nonlinear near-field signals). However, EOS needs specific nonlinear optical crystals in which the phonon absorption bands either do not overlap with the detected THz field, or the crystals are thin enough to permit transmission which necessitates larger THz or gate fields. In the case of MIR frequencies, EOS needs ultrafast pulses (such as pulse duration, $\tau \lesssim 15$ fs for wavelength, $\lambda < 10$ µm) with precise timing, making it less versatile for broadband or incoherent sources compared with interferometry with MCTs. For these reasons, EOS is typically used to detect THz fields, whereas an interferometric intensity measurement is typically used for MIR wavelengths. Table 2 provides a quantitative comparison between interferometric detection and EOS-based near-field detection in terms of spectral range (bandwidth) and acquisition times at the present stage of development. Pseudo-heterodyne detection, a subcategory in interferometric detection discussed in more detail subsequently, is also included.

Pump modulation. In contrast to static s-SNOM, in which the signal emerges from an ensemble of the material within the near-field probe volume in its unperturbed ground state, in ultrafast s-SNOM, the probe signal of interest is that of a sub-ensemble perturbed from the ground state. Depending on the intensity of the pump and its absorption cross-section with a sample resonance, this excited-state sub-ensemble may only be a small fraction of the otherwise ground-state-dominated probe signal. Therefore, a method is required to discriminate the transient excited-state signal from other contributions.

Similar to conventional far-field pump–probe spectroscopy, excited-state contrast is achieved by modulation of pump or probe beam or both (see Box 2 for a comparison between near-field and far-field pump–probe spectroscopy). In ultrafast s-SNOM, periodic modulation of the pump intensity at frequency, Ω_{M}, via a chopper, acousto-optic modulator (AOM) or electro-optic modulator (EOM), allows for demodulation of the detected IR signal and the extraction of the transient pump-induced change to the near-field probe signal $\Delta\tilde{E}_{\mathrm{NF}}$. The near-field localized photoexcited material response is encoded in the time-dependent (t) doubly modulated near-field probe signal by:

$$\Delta\tilde{E}_{\mathrm{NF}} \propto \sin(\omega_{\mathrm{t}}t)\cdot\sin(\Omega_{\mathrm{M}}t) = \frac{1}{2}\{\cos[(\Omega_{\mathrm{M}}-\omega_{\mathrm{t}})t]-\cos[(\Omega_{\mathrm{M}}+\omega_{\mathrm{t}})t]\}, \quad (1)$$

giving rise to frequency sidebands at $n\omega_{\mathrm{t}} \pm m\Omega_{\mathrm{M}}$. AOMs or EOMs are preferred owing to their ability of generating arbitrary waveforms, including pure sinusoidal intensity modulations. Conventional mechanical choppers create a rectangular modulation of the pump intensity, thus spreading the pump-induced signal into higher harmonic sidebands of Ω_{M}.

Signal demodulation. Lock-in detection is used to detect signal components at harmonics and sidebands of the tip-tapping frequency, ω_{t}, and pump modulation frequency, Ω_{M} (Fig. 2c). Figure 2c illustrates the detection scheme in which the periodically modulated pump pulses (with frequency Ω_{M}) and harmonic tip-tapping motion (with frequency ω_{t}) result in periodic modulation of $\tilde{E}_{\mathrm{NF}}$. The Fourier decomposition of $\tilde{E}_{\mathrm{NF}}$ is shown as an inset in Fig. 2c, in which harmonics of ω_{t} alone contain the nano-localized ground-state response $\tilde{E}_{\mathrm{NF},0}$, whereas sidebands modulated by both frequencies ω_{t} and Ω_{M} contain the nano-localized excited-state response $\Delta\tilde{E}_{\mathrm{NF}}$. The optical chopper, EOM or AOM, and the AFM-segmented photodiode which measures the live tip-tapping frequency, provides the corresponding reference signals for the lock-in amplifier. The nano-localized ground-state signal, $\tilde{E}_{\mathrm{NF}}$, is isolated from far-field background scattering by detecting harmonics of the tip-tapping frequency, and the nano-localized transient, $\Delta\tilde{E}_{\mathrm{NF}}$, is isolated through the sidebands $n\omega_{\mathrm{t}} \pm m\Omega_{\mathrm{M}}$ (generally choosing $m = 1$). The lock-in bandwidth is determined by the acquisition time constant, τ, with $\tau = 1$ ms resulting in a detection bandwidth of generally <1 kHz (dependent on the filter setting used by the lock-in). Care should be taken to sample slow enough that multiple sidebands do not crosstalk but fast enough for efficient acquisition time. With typical EOM, AOM or chopper frequencies of few kilohertz, a time constant on the order of 1–10 ms (larger time constant for lower value of Ω_{M}) is thus sufficient to avoid crosstalk between the pump–probe sidebands. Table 3 summarizes key frequency scales for ultrafast s-SNOM and considerations for choosing each properly.

Signal theory

In this section, we describe how to separate the tip-scattered near-field signal, $\tilde{E}_{\mathrm{NF}}$, from unwanted far-field background scattering, $\tilde{E}_{\mathrm{BG}}$, and to isolate the transient excited-state near-field response, $\Delta\tilde{E}_{\mathrm{NF}}$, from the ground state.

Near-field signal isolation. This section describes the basic principles of interferometric ground-state s-SNOM near-field signal detection and applies analogous to pump–probe s-SNOM for excited-state spectroscopy. In the case of an intensity measurement (in which a time-integrating photodetector is used), the detector signal, I, results from the superposition of three optical fields: the tip-scattered near-field, $\tilde{E}_{\mathrm{NF}}$; the reference or local oscillator field, $\tilde{E}_{\mathrm{LO}}$; and all other contributions from sample or tip scattering composing an effective background field, $\tilde{E}_{\mathrm{BG}}$, represented by

$$I = |\tilde{E}_{\mathrm{LO}} + \tilde{E}_{\mathrm{NF}} + \tilde{E}_{\mathrm{BG}}|^2, \tag{2}$$

and expanded to

$$I = |\tilde{E}_{\mathrm{LO}}|^2 + |\tilde{E}_{\mathrm{NF}}|^2 + |\tilde{E}_{\mathrm{BG}}|^2 + 2\mathfrak{R}[\tilde{E}_{\mathrm{NF}}\,\tilde{E}^*_{\mathrm{LO}} + \tilde{E}_{\mathrm{LO}}\,\tilde{E}^*_{\mathrm{BG}} + \tilde{E}_{\mathrm{NF}}\,\tilde{E}^*_{\mathrm{BG}}]. \tag{3}$$

Using lock-in demodulation at harmonics of the tip-tapping frequency, $n\omega_{\mathrm{t}}$ with $n > 1$, isolates the terms with $\tilde{E}_{\mathrm{NF}}$, to first approximation, as the magnitude of $\tilde{E}_{\mathrm{NF}}$ varies cubically with tip–sample distance owing to the $1/r^3$ component of the near-field intensity. The nonlinear tip–sample distance dependence leads to large higher harmonic components of the detected $\tilde{E}_{\mathrm{NF}}$ as the tip oscillates sinusoidally. By contrast, $\tilde{E}_{\mathrm{BG}}$ arises from phase-sensitive interference between sample scattering (independent of tip height) and scattering from the tip shaft. $\tilde{E}_{\mathrm{BG}}$ varies approximately linearly with tip–sample distance as the phase between these two components varies linearly with tip height. The detected intensity of $\tilde{E}_{\mathrm{BG}}$ modulates approximately sinusoidally as the tip oscillates sinusoidally, leading to faster fall-off with increasing harmonic order compared with $\tilde{E}_{\mathrm{NF}}$.

The relative contributions of $\tilde{E}_{\mathrm{NF}}$ and $\tilde{E}_{\mathrm{BG}}$ to the total scattered signal are also wavelength-dependent, with $\tilde{E}_{\mathrm{BG}}$ generally larger at shorter wavelengths for a given demodulation order[10]. In the MIR spectral range, measurements of the third or higher ($n \geq 3$) harmonics are desirable to isolate $\tilde{E}_{\mathrm{NF}}$ if the tapping amplitude of the AFM tip is >40 nm (refs. 27,63). Higher harmonic detection is generally necessary in the visible spectral range owing to the increased $\tilde{E}_{\mathrm{BG}}$, whereas the second harmonic is sufficient in the THz range[10]. In general, because the intensity decreases by a factor of 5–10 for each harmonic, it is important to choose the lowest viable harmonic while still effectively isolating $\tilde{E}_{\mathrm{NF}}$.

Selecting only the terms with frequency components at the $n\omega_{\mathrm{t}}$ demodulation order and neglecting the near-field intensity term $|\tilde{E}_{\mathrm{NF}}|^2$ as it is much smaller than any of the other contributions, equation (3) reduces to

$$I^{n\omega_{\mathrm{t}}}(t_{\mathrm{ref}}) = 2\mathfrak{R}[\tilde{E}^{n\omega_{\mathrm{t}}}_{\mathrm{NF}}\tilde{E}^*_{\mathrm{LO}}(t_{\mathrm{ref}}) + \tilde{E}^{n\omega_{\mathrm{t}}}_{\mathrm{NF}}\tilde{E}^*_{\mathrm{BG}}], \tag{4}$$

in which t_{ref} is the interferometer delay, given by $2d_{\mathrm{ref}}/c$, with reference arm travel distance, d_{ref}, and the speed of light, c. As only $\tilde{E}^{n\omega_{\mathrm{t}}}_{\mathrm{NF}}\tilde{E}^*_{\mathrm{LO}}(t_{\mathrm{ref}})$ depends on the interferometer position, $\tilde{E}^{n\omega_{\mathrm{t}}}_{\mathrm{NF}}\tilde{E}^*_{\mathrm{BG}}$ appears as an intensity offset in the interferogram and can be removed by Fourier transform.

The following sections describe how to isolate the transient nano-localized excited-state signal $\Delta\tilde{E}_{\mathrm{NF}}$ and then to detect either the broadband $\Delta\tilde{E}_{\mathrm{NF}}(\omega)$ for ultrafast spectroscopy or the single-frequency $\Delta\tilde{E}_{\mathrm{NF}}(\omega = \mathrm{const})$ for spatiotemporal imaging at a fixed frequency.

Near-field Fourier-transform infrared spectroscopy (nano-FTIR). In contrast to conventional FTIR spectroscopy with the sample interaction outside the Michelson interferometer (known as symmetric interferometry), in interferometric s-SNOM (commonly referred to as nano-FTIR), the light–matter interaction only occurs in one arm of the interferometer[64–66]. In symmetric interferometry, the resulting intensity interferogram is generally an even, real-valued function and thus its Fourier transform is purely real. In asymmetric interferometry, the interferogram is not even, thus its Fourier transform $\mathcal{F}\{I^{n\omega_{\mathrm{t}}}\}(\omega)$ is a complex quantity, representing both amplitude and phase of $\tilde{E}_{\mathrm{NF}}$ as

$$\mathcal{F}\{I^{n\omega_{\mathrm{t}}}\}(\omega) \propto \tilde{E}^{n\omega_{\mathrm{t}}}_{\mathrm{NF}}(\omega) = E_{0,n}(\omega)e^{i\phi_n(\omega)}, \tag{5}$$

in which the near-field spectral amplitude $E_{0,n}(\omega)$ and phase $\phi_n(\omega)$ relate to the real and imaginary parts of the local dielectric function of the material $\tilde{\varepsilon}_s(\omega)$. In asymmetric interferometry, the spectral information for resonant interactions in $\tilde{\varepsilon}_s(\omega)$ manifests itself as a tail on the positive delay side of the interferogram. For this reason, the interferogram only needs to be acquired for positive time delays except what is necessary to record the full centre burst.

The spectral resolution, $\Delta\omega$, of $\mathcal{F}\{I\}(\omega)$ is inversely proportional to the length, L_{ref}, of the measured interferogram given by $L_{\mathrm{ref}} = c/2\Delta\omega$. The resolution of the scan ΔL_{ref} relates to the highest frequency resolvable as $\omega_{\mathrm{max}} = c/2\Delta L_{\mathrm{ref}}$ by the Nyquist theorem. In practice, L_{ref} and ΔL_{ref} should be chosen for a given sample based on the desired values ω_{max} and $\Delta\omega$, the expected sample resonance frequencies and their linewidths, respectively.

Post-processing the FTIR data includes zero-padding and apodization. Zero-padding involves adding zeroes to both ends of the interferogram to artificially enhance the resolution, effectively interpolating between measured data points. Apodization is the process of weighing the intensity values based on their delay from $t_{\mathrm{ref}} = 0$, in which data near the interference maximum are weighed more than data far into the tail. This suppresses noise at interferometer delays where the signal is weak. Different apodization functions can be used depending on the spectral

Box 2 | Differences between near-field and far-field pump–probe spectroscopy

In conventional far-field pump–probe spectroscopy, the pump beam is chosen to be larger than the probe focus to ensure a mostly homogeneous pump fluence within the probe volume and to reduce the influence of mechanical drift in the delay line. This consideration is not applicable in ultrafast scattering-type scanning near-field opical microscopy (s-SNOM) because both the pump and probe focus spots are much larger than the tip near-field region. However, just as the infrared probe light is localized by the tip apex, the pump light may also be locally enhanced. Therefore, the excited-state dynamics need to be understood as a convolution of sample regions or sub-ensembles with potentially locally varying pump excitation. The tip-induced pump field enhancement can be limited by setting the polarization of the pump perpendicular to the tip axis. Owing to the sinusoidal modulation of the pump intensity in s-SNOM, the photoinduced effect and its dynamics are again an average across ultrafast processes occurring at different fluences. Differences between far-field time traces and those acquired in ultrafast s-SNOM may also occur if the relaxation dynamics are different in the near-surface region preferentially probed in s-SNOM compared with dynamics in the material bulk. The exact pump fluence experienced in the near-field probe region is difficult to quantify owing to possible multiple scattering and interference associated with tip and sample scattering.

Table 3 | Key frequency scales and constraints for ultrafast scattering-type scanning near-field optical microscopy

Property	Typical range	Considerations
Laser repetition rate	1–100 MHz	Time between consecutive pump pulses must be long enough for the excitation and system response to have fully relaxed back to the initial equilibrium state
Atomic force microscopy cantilever frequency, ω_t	50–500 kHz	Avoid undersampling by choosing ω_t so that the laser repetition rate exceeds twice the highest detected harmonic frequency (known as the Nyquist limit)
Pump modulation, Ω_M	1–10 kHz	Large Ω_M allows for fast acquisition, but ensures that $\Omega_M \ll \omega_t$ to avoid overlap between sidebands $n\omega_t \pm m\Omega_M$ with different values of n and m[a]
Lock-in time constant, τ	1–10 ms	Must be sufficiently high so that the frequency resolution is lower than Ω_M to avoid crosstalk between sidebands

[a]Both n and m are non-negative integers defining the demodulation and sideband harmonics.

bandwidth and nature of the material response. The same approach applies to processing heterodyne pump–probe (HPP) IR s-SNOM data, to obtain the excited state described in the next section[67].

Heterodyne pump–probe IR s-SNOM (HPP IR s-SNOM). This section describes the extension of nano-FTIR for isolating the excited-state transient near-field spectrum, $\Delta\tilde{E}_{NF}(\omega, T)$, as a function of pump–probe delay time, T, a process known as HPP IR s-SNOM. When the ground-state sample response, $\tilde{\varepsilon}_{s,0}(\omega)$, is modified by a pump excitation, a transient material response, $\Delta\tilde{\varepsilon}_s(\omega, T)$, is generated. By modulating the intensity of the pump beam at frequency Ω_M (Fig. 2c), $\Delta\tilde{\varepsilon}_s(\omega, T)$ will be time-dependent with

$$\tilde{\varepsilon}_s(\omega, \Omega_M, T) = \tilde{\varepsilon}_{s,0}(\omega) + \Delta\tilde{\varepsilon}_s(\omega, m\Omega_M, T) \tag{6}$$

in which the transient near-field response is modulated at harmonics of Ω_M, with a DC ground-state response. As $\tilde{E}_{NF}(\omega)$ is related to $\tilde{\varepsilon}_s(\omega)$, the transient near-field response is given by

$$\tilde{E}_{NF}(\omega, \omega_t, \Omega_M, T) = \tilde{E}_{NF,0}(\omega, n\omega_t) + \Delta\tilde{E}_{NF}(\omega, n\omega_t, m\Omega_M, T), \tag{7}$$

with $\Delta\tilde{E}_{NF}$ modulated both at tip-tapping frequency and at frequency by the modulated material response, giving frequency sidebands in $\tilde{E}_{NF}(\omega, \omega_t, \Omega_M)$ (Fig. 2c). The unmodulated term ($m = 0$) in equations (6) and (7) represents the ground-state material response if the instantaneous pump intensity is zero during its modulation, as is the case with an optical chopper but not necessarily with an AOM. In this case, the nano-localized ground-state and excited-state responses can be extracted simultaneously by recording the relevant $m = 0$ and $m = 1$ sideband signals. If the instantaneous pump intensity during modulation is always non-zero, the $m = 0$ sideband contains mixed excited-state and ground-state character, but the transient $\Delta\tilde{E}_{NF}$ is still isolated in the $m \neq 0$ sidebands.

Demodulating the detector signal I as described in equation (4) by sideband lock-in detection at $n\omega_t \pm m\Omega_M$ (for brevity, indicated by the superscript $\tilde{E}^{n,m}$) provides the transient signal, ΔI, given by

$$\Delta I^{n,m}(T, t_{ref}) = 2\Re[\Delta\tilde{E}^{n,m}_{NF}(T)\tilde{E}^*_{LO}(t_{ref}) + \Delta(\tilde{E}^{n,m}_{NF}(T)\tilde{E}^{*m}_{BG})], \tag{8}$$

which expands to

$$\Delta I^{n,m}(T, t_{ref}) = 2\Re[\Delta\tilde{E}^{n,m}_{NF}(T)\tilde{E}^*_{LO}(t_{ref}) \tag{9}$$

$$+\tilde{E}^{n}_{NF,0}(T)\Delta\tilde{E}^{*m}_{BG} \tag{10}$$

$$+\Delta\tilde{E}^{n,m}_{NF}(T)\tilde{E}^*_{BG,0}]. \tag{11}$$

For a given value of T, the excited-state spectrum is obtained by varying t_{ref} with the difference of detecting at sidebands $n\omega_t \pm m\Omega_M$. As only the first term in equation (8) is dependent on t_{ref}, the other terms appear as constant offsets in the interferogram and are removed by Fourier transformation. Recording the $m = \pm 1$ sideband is sufficient to determine excited-state dynamics as the $|m| > 1$ sidebands only differ from $m = 1$ if there is outstanding nonlinearity in the sample response with a sinusoidally modulated pump intensity. The choice of tip harmonic, n, follows the same considerations as for ground-state near-field spectroscopy. The interferograms, $\Delta I^{n,m}(T, t_{ref})$, are then processed.

The transient signal, $\Delta\tilde{E}_{NF}$, can also contain undesirable contributions from pump-induced excitations in the sample substrate or the tip material, which are difficult to distinguish a priori. This effect can be minimized by choosing sample substrates that are only weakly absorbing at the pump fundamental or second harmonic frequency (two-photon absorption), such as CaF_2 or other wide bandgap materials. The photoinduced carriers in the tip result in a broad Drude response, adding a broad background contribution[68]. The latter can be discriminated spectrally if, for example, only narrow vibrational or phonon spectral features are of interest in the sample response. More generally, careful normalization of the sample response against a reference material response is necessary.

Two-phase heterodyne detection. When the spectral phase and amplitude of a material response are only weakly wavelength-dependent across the spectral bandwidth of the probe pulse, it is not necessary to record the entire FTIR interferogram. This is the case for the broad free-carrier Drude or polaron response in metals or semiconductors, or to an extent, short-lived electronic, molecular or interband transitions. Here, the acquisition time can be reduced by orders of magnitude by recording $\Delta I(T, t_{ref})$ for only two or three selected values of t_{ref}. In case of a two-phase measurement, two measured intensity values suffice to eliminate the background terms from the heterodyne term, $\Delta E^{n,m}_{NF}(T)E^*_{LO}(t_{ref})$, as this is the only term in equation (8) which depends on t_{ref}. The intensities for two-phase values chosen such that ΔE_{NF} and E_{LO} interfere either constructively ($\phi = 0$) or destructively ($\phi = \pi$) are then given by

$$\Delta I^{n,m}(\omega_{pr}t_{ref} \equiv \phi = 0) \approx 2\Re[\Delta\tilde{E}^{n,m}_{NF}\tilde{E}^*_{LO} + \Delta\tilde{E}^{n,m}_{NF}\tilde{E}^*_{BG} + \tilde{E}^{n}_{NF}\Delta\tilde{E}^{*m}_{BG}] \text{ and} \tag{12}$$

$$\Delta I^{n,m}(\omega_{pr}t_{ref} \equiv \phi = \pi) \approx 2\Re[-\Delta\tilde{E}^{n,m}_{NF}\tilde{E}^*_{LO} + \Delta\tilde{E}^{n,m}_{NF}\tilde{E}^*_{BG} + \tilde{E}^{n}_{NF}\Delta\tilde{E}^{*m}_{BG}], \tag{13}$$

with ω_{pr} the centre frequency of the probe pulse. Taking the difference then isolates the pump-induced change in the amplitude of the near-field signal R_{HPP} as

$$R_{HPP} = \Delta I^{n,m}(\phi = 0) - \Delta I^{n,m}(\phi = \pi) \approx 4|\tilde{E}_{LO}||\Delta\tilde{E}^{n,m}_{NF}(T)|. \tag{14}$$

This approximation is valid only if the two-phase measurements are performed for consecutive values of t_{ref} corresponding to constructive and destructive interference as established by measuring one interferogram before applying two-phase detection. Once the appropriate values of t_{ref} are established, $\Delta I^{n,m}(\phi)$ can be imaged with spatial (x, y) and temporal (T) resolution for these two-phase values. Selecting the two values of t_{ref} as close as possible to an interference maximum or minimum is important to maximize signal-to-noise ratio (SNR). Interferometer drift between measured pixels (an unwanted change in t_{ref}) and especially between images at different pump delays may lead to artefacts as drift off the interference maxima makes R_{HPP} appear to decay faster than it really does. This can lead to artificially fast time constants or incorrect comparison of signal intensity between spatial positions if drift is larger than ~$\lambda/10$ during the measurement. For this reason, remeasuring the full interference pattern before each image is recommended. Additionally, the interferometer position can be tracked in real time using an external laser diode or internal position controls of the interferometer delay stage.

In practice, two consecutive scan images are taken recording $\Delta I^{n,m}$ at each image pixel with the reference arm fixed at each of the two delay positions. A time series of images can then be recorded at increments of T, corresponding to frames of an ultrafast movie. Thus, two-phase HPP s-SNOM provides spatiotemporal imaging of excited-state dynamics[7]. However, no spectral information is provided, and the signal can be viewed as a spectral average of the excited-state response from a superposition of resonant and non-resonant contributions. This approach is applicable not only for a broad spectral response with respect to the probe pulse bandwidth but also if the response is spectrally more narrow as long as no spectral information is necessary.

Full excited-state spectroscopy as a function of T is needed if a time-dependent spectral shift is expected and shall be resolved during the relaxation process. However, such full temporal–spectral resolution is only practical at selected points of interest and a limited number of pump–probe delay times. Recording one R_{HPP} value (with two interferometer positions) is orders of magnitude faster than one high spectral resolution FTIR spectrum (containing several thousand interferometer positions) and thus enables many more spatial or temporal points to be measured at the expense of spectral resolution. In practice, the choice of desired spatial, spectral and temporal resolution within the 4D space of x, y, T and ω, requires some form of survey spectroscopy to first identify the parameters of maximal physical value for the science question to be addressed. Then, compromises must be made to limit full temporal–spectral s-SNOM to selected but representative sample features and to limit the spatiotemporal imaging to two-phase heterodyne s-SNOM imaging, or even just line transects along or across sample regions of interest.

To accelerate data acquisition, yet still obtain a spectrally resolved excited-state response, different computational spectroscopy approaches can be applied. For pre-characterized excited state spectral features, compressed sensing combined with prior knowledge algorithms allow for deep sub-Nyquist sampling of the interferogram, decreasing acquisition time by up to a factor of 10 while still obtaining the same spectral information compared with conventional interferometry[69]. Alternatively, sampling in the rotating frame can be applied and demonstrated to provide spatio-spectral imaging at massively shortened acquisition times[70].

Three-phase heterodyne detection. The two-phase approach is only applicable if the phase of $\Delta\tilde{E}_{\mathrm{NF}}$ is constant both spatially and temporally. If the transient near-field spectral phase $\Delta\phi_{\mathrm{NF}}$ varies either spatially or temporally, $\Delta I^{n,m}(t_{\mathrm{ref}})$ must be recorded for a minimum of three phase delays. By measuring $\Delta I^{n,m}$ for $\omega_{\mathrm{pr}} t_{\mathrm{ref}} = \phi, \phi + \pi/2$ and $\phi + \pi$, R_{HPP} can be determined along with $\Delta\phi_{\mathrm{NF}}$. In contrast to the two-phase approach, the value of ϕ can now be arbitrary and does not need to correspond to the exact delay positions of constructive or destructive interference, although choosing delay positions near $t_{\mathrm{ref}} = 0$ is still preferable to maximize SNR.

The detector intensities for each delay are then given by

$$\Delta I^{n,m}(\phi) \approx 2\Re[\Delta\tilde{E}_{\mathrm{NF}}^{n,m}\tilde{E}_{\mathrm{LO}}^{*}\cos(\phi + \Delta\phi_{\mathrm{NF}}) + \Delta\tilde{E}_{\mathrm{NF}}^{n,m}\tilde{E}_{\mathrm{BG}}^{*} + \tilde{E}_{\mathrm{NF}}^{n}\Delta\tilde{E}_{\mathrm{BG}}^{*m}], \quad (15)$$

$$\begin{aligned}\Delta I^{n,m}&(\phi + \pi/2)\\ &\approx 2\Re[\Delta\tilde{E}_{\mathrm{NF}}^{n,m}\tilde{E}_{\mathrm{LO}}^{*}\sin(\phi + \Delta\phi_{\mathrm{NF}}) + \Delta\tilde{E}_{\mathrm{NF}}^{n,m}\tilde{E}_{\mathrm{BG}}^{*} + \tilde{E}_{\mathrm{NF}}^{n}\Delta\tilde{E}_{\mathrm{BG}}^{*m}],\end{aligned} \quad (16)$$

$$\begin{aligned}\Delta I^{n,m}&(\phi + \pi)\\ &\approx 2\Re[-\Delta\tilde{E}_{\mathrm{NF}}^{n,m}\tilde{E}_{\mathrm{LO}}^{*}\cos(\phi + \Delta\phi_{\mathrm{NF}}) + \Delta\tilde{E}_{\mathrm{NF}}^{n,m}\tilde{E}_{\mathrm{BG}}^{*} + \tilde{E}_{\mathrm{NF}}^{n}\Delta\tilde{E}_{\mathrm{BG}}^{*m}].\end{aligned} \quad (17)$$

The sum $\Delta I^{n,m}(\phi) + \Delta I^{n,m}(\phi + \pi)$ gives twice the background contribution to the signal, which can then be eliminated from the first two terms. Let $D \equiv [\Delta I^{n,m}(\phi) + \Delta I^{n,m}(\phi + \pi)]/2$ so that

$$\Delta I^{n,m}(\phi) - D \approx 2\Re[\Delta\tilde{E}_{\mathrm{NF}}^{n,m}\tilde{E}_{\mathrm{LO}}^{*}\cos(\phi + \Delta\phi_{\mathrm{NF}})], \text{ and} \quad (18)$$

$$\Delta I^{n,m}(\phi + \pi/2) - D \approx 2\Re[\Delta\tilde{E}_{\mathrm{NF}}^{n,m}\tilde{E}_{\mathrm{LO}}^{*}\sin(\phi + \Delta\phi_{\mathrm{NF}})]. \quad (19)$$

The HPP amplitude R_{HPP} is then found as the quadrature sum of equations (18) and (19), whereas the pump–probe transient phase $\Delta\phi_{\mathrm{NF}}$ is found as the arctangent of their ratio

$$R_{\mathrm{HPP}} = \sqrt{(\Delta I^{n,m}(\phi) - D)^2 + (\Delta I^{n,m}(\phi + \pi/2) - D)^2}, \text{ and} \quad (20)$$

$$\phi + \Delta\phi_{\mathrm{NF}} = \arctan\left(\frac{\Delta I^{n,m}(\phi + \pi/2) - D}{\Delta I^{n,m}(\phi) - D}\right). \quad (21)$$

The absolute phase of $\Delta\phi_{\mathrm{NF}}$ cannot be determined owing to its convolution with ϕ. However, as only the relative phase between two locations or time delays is physically meaningful, the contribution from the constant value of ϕ can be ignored. The result of the three-phase approach is rapid imaging of the transient near-field pump–probe response to generate spatiotemporal nano-movies with both phase and amplitude contrast. Two-phase imaging should be applied on samples in which the phase of $\Delta\phi_{\mathrm{NF}}$ is constant (for example, the interference maxima in an interferogram are always at the same reference arm positions) owing to its faster acquisition time. However, if the interferogram phase changes with position or pump–probe delay, three-phase imaging may be necessary. In practice, several interferograms can be recorded at various positions and pump–probe delay times to determine whether the interference maxima appear at a constant reference arm position to determine which approach to apply.

Pseudo-heterodyne detection. In addition to the two-phase and three-phase detection schemes discussed earlier, pseudo-heterodyne detection[29,71] can be applied to measure the amplitude and phase of $\tilde{E}_{\mathrm{NF}}$ in the single frequency bandwidth limit and is commonly applied to ground-state s-SNOM. Pseudo-heterodyne detection involves the same interferometric set-up as the other heterodyne schemes discussed, but with a piezo-driven mirror in the reference arm oscillating at frequency $\Omega_{\mathrm{r}} \approx 0.1$–1kHz. The tip-scattered signal is then detected at sidebands of the tip-tapping harmonics, with the first and second sidebands ($n\omega_t \pm m\Omega_r$, $m = 1,2$) used together to determine the amplitude and phase of $\tilde{E}_{NF}$ (for details, see refs. 29,71).

Because the frequency Ω_r is of similar magnitude to the pump amplitude modulation, Ω_M, applying pseudo-heterodyne detection to ultrafast s-SNOM with both modulations simultaneously is challenging owing to crosstalk between the numerous sidebands. However, ultrafast s-SNOM has been demonstrated with pseudo-heterodyne detection with a different approach to achieve excited-state contrast[10]. Rather than modulating the pump arm at kilohertz frequencies, pulse picking can be applied with every other pump pulse blocked using electro-optic modulation. This results in an alternating sequence of ground-state (probe pulse without a pump pulse) and excited-state (probe pulse following a pump pulse at controllable delay) amplitude and phase measurements. The two signals are then separated by two digital boxcars to simultaneously measure the pump-on (s_P) and pump-off (s_R) near-field responses. The normalized difference between the two, defined as $R_{HPP} = (s_P - s_R)/s_R$, is then the excited-state near-field signal. In principle, this boxcar approach can be applied to any detection scheme discussed earlier in place of the pump modulation and sideband detection, but with pseudo-heterodyne detection a boxcar or similar gating approach is necessary.

Ultrafast nano-movies and relaxation time images. The techniques described earlier can be used to produce spatial maps of the pump–probe amplitude signal R_{HPP} in real space with nanometre resolution, referred to as ultrafast nano-movies. Both spatial heterogeneities in the magnitude of R_{HPP} and corresponding variations in the temporal evolution of R_{HPP} as a function of the pump–probe delay T can be resolved in this way. The sideband-demodulated signal at frequency $n\omega_t \pm m\Omega_M$ representing $\Delta I^{n,m}$ and as obtained from the lock-in amplifier is then plotted for each image pixel. By taking scans of this signal at two reference phase values ($\phi = 0$ and $\phi = \pi$), and varying the time delay T, a spatially resolved animation of R_{HPP} can be constructed to visualize spatial heterogeneities in the dynamic response of a material system. By processing and modelling the decay dynamics, time constants can be derived for each pixel, providing a spatially resolved decay rate map.

Electro-optic sampling. EOS reconstructs $\Delta\tilde{E}_{NF}$ by sampling the tip-scattered field in time domain with a femtosecond gate pulse in a nonlinear crystal[1]. Several studies have applied this approach in nano-localized visible-pump IR-probe experiments[72]. In practice, the pump–probe delay, T, is fixed, whereas the gate–MIR delay, τ_{EOS}, is scanned to retrieve the full amplitude and phase of $\tilde{E}_{NF}$ for each value of T. This enables measurements of temporal dynamics at a single sample location, or imaging, to obtain local values of $\tilde{E}_{NF}$ and $\Delta\tilde{E}_{NF}$. Practical considerations include the weak near-field signal (often offset by a higher gate-pulse intensity) and the gate-duration constraint noted earlier in which EOS becomes challenging for wavelengths below 10 µm, where sub-10 fs gate pulses are required.

Results

Fundamentally, coherent light–matter interactions are described by the complex-valued optical susceptibility tensor, $\tilde{\chi}(\omega)$, related to the relative dielectric permittivity $\tilde{\varepsilon}_r(\omega) \equiv 1 + \tilde{\chi}(\omega) \equiv \varepsilon_1(\omega) + i\varepsilon_2(\omega)$. The relative dielectric permittivity determines the complex-valued macroscopic optical polarization density, $\tilde{P}(\omega)$, induced in a material by an optical driving field, $\tilde{E}(\omega) = E_0(\omega)e^{i(\omega t - kr)}$, expressed as $\tilde{P}(\omega) = \varepsilon_0\tilde{\chi}(\omega)\tilde{E}(\omega)$. Equivalently, the microscopic polarizability $\tilde{\alpha}(\omega)$ can be defined such that $\tilde{P}(\omega) = N\tilde{\alpha}(\omega)\tilde{E}(\omega)$ with N being the number density of polarized atoms, unit cells or molecules. Inserting the plane wave ansatz $\tilde{E}(\omega) = E_0(\omega)e^{i(\omega t - kr)}$, with r position and t time, into the wave equation

$$\nabla^2 E(r,t) - \tilde{\mu}_r(\omega)\tilde{\varepsilon}_r(\omega)\frac{\partial^2 E(r,t)}{\partial t^2} = 0, \qquad (22)$$

yields the dispersion relation $k^2(\omega) = (\frac{\omega}{c})^2\mu_r(\omega)\varepsilon_r(\omega)$. The term $\tilde{\mu}_r(\omega)$ is the relative magnetic permeability, which is commonly assumed to be negligibly different from unity ($\tilde{\mu}_r(\omega) = 1$), thus the complex refractive index is given by $\tilde{n} = \sqrt{\tilde{\varepsilon}_r} = n + i\kappa$ with dispersion, n, and absorption, κ, expressed in terms of $\tilde{\varepsilon}_r$ as $\varepsilon_1 = n^2 - \kappa^2$ and $\varepsilon_2 = 2n\kappa$.

The polarization induced in the sample material by the near field of the probe tip modifies the tip-scattered field, $\tilde{E}_{NF}$, which is the measured quantity in s-SNOM. The phase and amplitude of $\tilde{E}_{NF}$ are modified relative to a non-resonant reference material by the absorption and dispersion in the sample encoding the frequency and linewidth of electronic and vibrational resonances in $\tilde{E}_{NF}$. Access to both the phase and amplitude of the nano-localized near-field response in s-SNOM provides a more complete and less artefact-prone picture of the light–matter interaction, reflecting the actual optical response function of the material. Moreover, the high near-field momentum provided by the probe tip enables access to otherwise momentum-forbidden phonon, plasmon and other collective excitations, thus extending conventional far-field optical selection rules.

Several techniques such as spectroscopic ellipsometry, variations of FTIR spectroscopy, THz and other spectroscopies have been established in far field for the reliable determination of the complex value of $\tilde{\varepsilon}_r(\omega)$. Deriving complex values of the local optical permittivity from the near-field spectra remains an outstanding challenge in the field. However, just like in most conventional far-field ultrafast spectroscopy, for most spectroscopic s-SNOM applications, the absolute value of $\tilde{\varepsilon}_r(\omega)$ is not needed. $\tilde{E}_{NF}(\omega)$ qualitatively relates to $\tilde{\varepsilon}_r(\omega)$ in an established fashion, giving reliable information about relative changes of amplitude, frequency, linewidth and other parameters of the ultrafast temporal evolution of material excitations.

To correct for artefacts from tip-induced perturbations to the sample response for accurate relation of $\tilde{E}_{NF}(\omega)$ to $\tilde{\varepsilon}_r(\omega)$, several models have been developed[3,35,36,73]. These involve different levels of approximation and their applicability depends on the type of materials and resonances, and they have only been reliably tested for select examples. The challenge is in part related to the wide distribution of optical momentum states in the near field, the tip acting as a strongly interacting extrinsic perturbation, and complex spatial and spectral variation in the scattering response. Although the underlying mutual tip–sample polarization governing the s-SNOM signal can be calculated using Maxwell's equations with appropriate boundary conditions at both tip and sample surfaces, such a full electrodynamic calculation needs to include all nanoscale geometric tip and sample properties, which are in part not precisely known[74–76]. This problem is fundamentally similar to far-field spectroscopy of heterogeneous or scattering media, with the optical response a (phase-sensitive) superposition of the intrinsic dielectric response of the material and its extrinsic size distribution, with both a priori not know. Therefore, different models depending on various assumptions have been developed to determine scaling, qualitative comparison or at least semi-quantitative determination of the spatio-spectral dielectric function from s-SNOM (Fig. 3a). The appropriate choice of the model to relate $\tilde{E}_{NF}$ and $\Delta\tilde{E}_{NF}$ (Fig. 3b) to $\tilde{\varepsilon}_r$ and $\Delta\tilde{\varepsilon}_r$ (Fig. 3c) in ultrafast s-SNOM is therefore dependent on the material and the nature of the resonant interaction to be studied.

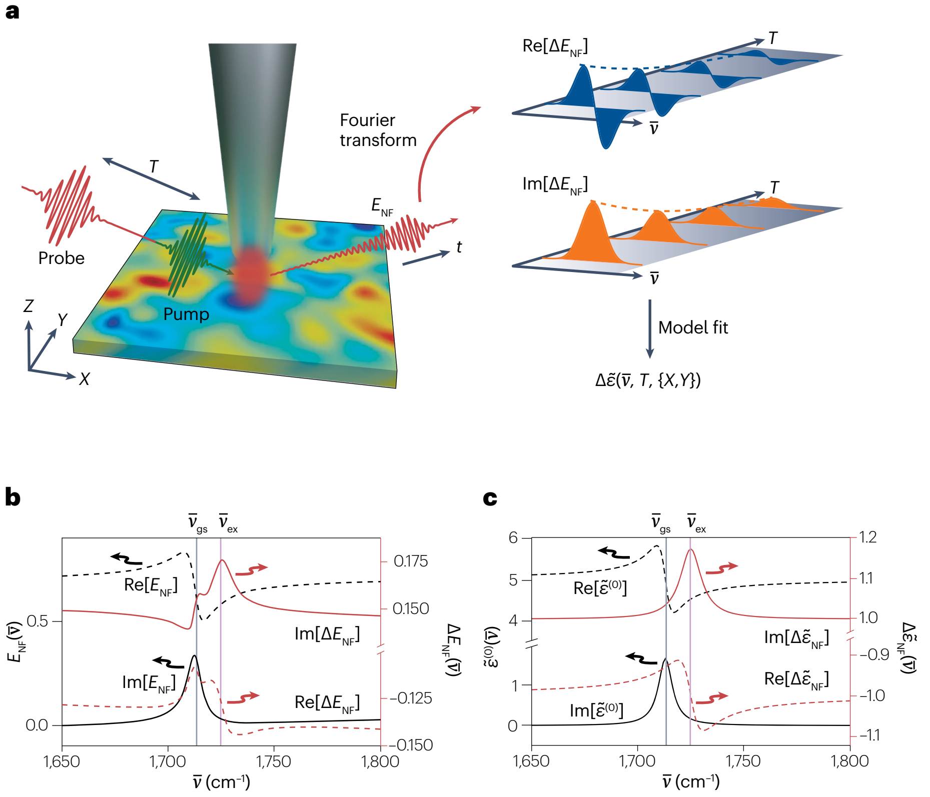

Fig. 3 | Example of near-field spectra calculated through the finite-dipole model with a four-layer reflection scheme. a, Real and imaginary components of the transient near-field response, $\Delta E_{\mathrm{NF}}(\bar{\nu}, T)$, and its evolution with pump–probe delay time, T, related to the transient dielectric response, $\Delta\tilde{\varepsilon}(\bar{\nu}, T)$. **b**, Simulated real and imaginary components of the ground-state tip-scattered near-field signal, $\tilde{E}_{\mathrm{NF}}(\omega)$ (black), and its pump-induced change, $\Delta\tilde{E}_{\mathrm{NF}}(\omega)$ (red), for a model vibrational response. **c**, The ground-state resonance is indicated by $\bar{\nu}_{\mathrm{gs}}$ and the frequency-shifted excited-state resonance by $\bar{\nu}_{\mathrm{ex}}$, corresponding to resonances in the ground-state (black, real and imaginary components) and excited-state (red, real and imaginary components) dielectric functions $\tilde{\varepsilon}_s$ and $\Delta\tilde{\varepsilon}_s$. Adapted from ref. 7, CC BY 4.0.

Although it is often challenging to quantitatively reconstruct $\tilde{\varepsilon}_r$ and $\Delta\varepsilon_r$ from near-field spectra, s-SNOM has matured as a reliable technique to determine qualitative spectral information, typically sufficient to analyse the material response in both equilibrium ground state and ultrafast pump–probe s-SNOM. The following sections illustrate simple ways to interpret $\Delta\tilde{E}_{\mathrm{NF}}$ in terms of the transient change in dielectric permittivity $\Delta\tilde{\varepsilon}_r$, which have been successfully applied to study nano-localized excited-state dynamics in a growing number of materials. We start with the simplest case of materials with small positive permittivity, in which $\Delta\tilde{E}_{\mathrm{NF}} \propto \Delta\tilde{\varepsilon}_r$ to good approximation. Expanding further, models representing the probe tip as an optical dipole are described to interpret spectra for samples with strong resonance features and polariton-resonant tip–sample interactions.

Ultrafast near-field response for low-permittivity materials

For many molecular samples with either dilute or weakly resonant vibrational modes, $\tilde{\varepsilon}_r(\omega)$ is small and positive across the IR spectral region. The referenced near-field response is defined as $\tilde{E}_{\mathrm{NF}} = \tilde{E}_{\mathrm{NF,sample}}/\tilde{E}_{\mathrm{NF,ref}} = E_{0,\mathrm{NF}}e^{i\phi_{\mathrm{NF}}}$, in which the sample response is normalized against a reference with a non-resonant optical response such as gold or doped silicon.

In weak oscillator materials in which the sample response is perturbed negligibly by tip interactions, $\tilde{E}_{\mathrm{NF}}$ is related to the absorption coefficient and dispersion of the sample as $\mathrm{Im}\{\tilde{E}_{\mathrm{NF}}\} \propto \kappa$ and $\mathrm{Re}\{\tilde{E}_{\mathrm{NF}}\} \propto n$ to good approximation[27,66,77]. For small phase responses in weakly resonant samples, $\mathrm{Im}\{\tilde{E}_{\mathrm{NF}}\} \propto \phi_{\mathrm{NF}} \propto \kappa$. Correspondingly, the transient near-field response $\Delta\tilde{E}_{\mathrm{NF}} = \Delta E_{0,\mathrm{NF}}e^{i\Delta\phi_{\mathrm{NF}}}$ is related to $\Delta\kappa$ and Δn as

$$\mathrm{Im}\{\Delta\tilde{E}_{\mathrm{NF}}\} \propto \Delta\phi_{\mathrm{NF}} \propto \Delta\kappa \qquad (23)$$

and

$$\mathrm{Re}\{\Delta\tilde{E}_{\mathrm{NF}}\} \propto \Delta E_{0,\mathrm{NF}} \propto \Delta n. \qquad (24)$$

The transient near-field response can then be interpreted in a similar way to the transient absorption in ultrafast far-field spectroscopy. s-SNOM spectra are often plotted as $\mathrm{Im}\tilde{E}_{\mathrm{NF}}(\omega)$ or $\mathrm{Re}\tilde{E}_{\mathrm{NF}}(\omega)$, in which the real and imaginary components are related to each other through the Kramers–Kronig relation[78,79].

Figure 3b shows examples of the real and imaginary spectra of $\tilde{E}_{\mathrm{NF}}(\omega)$ in the ground state and transient $\Delta\tilde{E}_{\mathrm{NF}}(\omega)$, modelled by a material response described by ground-state $\tilde{\varepsilon}^{(0)}(\omega)$ and transient $\Delta\tilde{\varepsilon}_{\mathrm{NF}}(\omega)$, as shown in Fig. 3c. The ground-state s-SNOM spectra $\tilde{E}_{\mathrm{NF}}(\omega)$ are qualitatively similar to $\tilde{\varepsilon}^{(0)}(\omega)$, whereas the excited-state spectra $\Delta\tilde{E}_{\mathrm{NF}}(\omega)$ are closely related to $\Delta\tilde{\varepsilon}_{\mathrm{NF}}(\omega)$. Extrinsic tip–sample interactions can cause small changes to the centre frequency, ω_0, and linewidth, Γ, of resonance features in $\tilde{E}_{\mathrm{NF}}(\omega)$ and $\Delta\tilde{E}_{\mathrm{NF}}(\omega)$, compared with those in $\tilde{\varepsilon}^{(0)}(\omega)$ and $\Delta\tilde{\varepsilon}_{\mathrm{NF}}(\omega)$. Far-field reflections can lead to spectral features from a material outside the near-field region appearing in s-SNOM spectra; hence, the small feature at ground-state resonance frequency ω_{gs} appears in $\Delta\tilde{E}_{\mathrm{NF}}(\omega)$. Models of tip–sample interactions and far-field reflection effects more accurately relate s-SNOM spectra to the corresponding resonance features in $\tilde{\varepsilon}^{(0)}(\omega)$ and $\Delta\tilde{\varepsilon}_{\mathrm{NF}}(\omega)$, compared with equations (23) and (24), enabling precise determination of ω_0 and Γ, which describe material resonances. Applying such models is particularly important when measuring strong, coupled or collective modes such as phonon, plasmon and polariton resonances. The following sections describe common approaches to modelling the tip–sample interactions and far-field reflections for accurate interpretation of s-SNOM spectra.

Point dipole and multilayer reflection model

For samples with strong resonances in which $\varepsilon_1(\omega)$ may approach zero or even become negative, increasingly refined models must be used to accurately determine the centre frequency and linewidth of material resonances from the corresponding resonance features in s-SNOM spectra. Additionally, artefacts can arise from far-field reflections interacting with the tip requiring consideration of the Fresnel coefficient of the sample[80] (Fig. 3b,c). To address these issues, various dipole coupling models that describe the tip–sample interactions have been established for both ground state[81–83] and ultrafast s-SNOM[1,7,49].

The simplest models describe the tip apex as a sphere of radius, R_{tip}, with an effective polarizability represented by a point dipole (point-dipole model (PDM))[28,35]. Other models describe the tip as an elongated ellipsoid (finite dipole) with length and apex radius used as fit parameters[36]. More recent numerical methods are less reliant on simplifying approximations[75].

Point-dipole model. In this section, we illustrate the application of the PDM to interpret $\Delta\tilde{E}_{\text{NF}}$ with a four-layer model to include sample reflections. In free space, the polarizability of the tip, α_0, is described by the Clausius–Mossotti equation as

$$\alpha_0 = 4\pi R_{\text{tip}}^3(\tilde{\varepsilon}_{\text{tip}}-1)/(\tilde{\varepsilon}_{\text{tip}}+2), \tag{25}$$

in which $\tilde{\varepsilon}_{\text{tip}}$ is the relative dielectric function of the tip material. When the tip is brought into contact with a polarizable sample material and illuminated, a polarization, p, is induced in the tip, which then induces a polarization in the sample, modifying the tip polarization. The interfacial polarization density distribution induced in the sample is effectively described by an image dipole of dipole moment $p_{\text{samp}} = ((\tilde{\varepsilon}_{\text{r}}-1)/(\tilde{\varepsilon}_{\text{r}}+1))p \equiv \beta p$, for $\tilde{\varepsilon}_{\text{r}}$ the relative dielectric function of the sample medium. The tip polarization modified by the near field of this effective image dipole can then be described by an effective polarizability $\tilde{\alpha}_{\text{eff}}$ as

$$\tilde{\alpha}_{\text{eff}} = \frac{\alpha_0}{1-\dfrac{\alpha_0\beta}{16\pi(R_{\text{tip}}+Z_{\text{tip}})^3}}, \tag{26}$$

in which Z_{tip} is the tip–sample distance[35]. The near-field signal $\tilde{E}_{\text{NF}}$ is then proportional to $\tilde{\alpha}_{\text{eff}}$, given by

$$\tilde{E}_{\text{NF}} \propto \tilde{\alpha}_{\text{eff}}\tilde{E}_{\text{in}}, \tag{27}$$

for $\tilde{E}_{\text{in}}$ the incident optical field. With the vertical tip oscillation approximately sinusoidal and the dielectric function of the tip known, one can use the Fourier decomposition of $\tilde{\alpha}_{\text{eff}}$ to relate $\tilde{E}_{\text{NF}}$ to $\tilde{\varepsilon}_{\text{r}}$. The transient $\Delta\tilde{E}_{\text{NF}}$ is instead related to the transient $\Delta\tilde{\alpha}_{\text{eff}}$, with the sample response to the pump-induced perturbation $\Delta\tilde{\varepsilon}_{\text{r}}$ and thus $\Delta\beta$ assumed to be the only parameter changed by pumping.

In samples in which far-field reflections cannot be neglected, the model can be expanded to incorporate modifications of the local field at the tip apex by Fresnel effects. The sample reflections can be modelled by describing the sample via multiple layers in which reflections occur off each interface, for example, the sample–air and sample–substrate interface. The interfaces that need to be considered depend on the penetration depth of the probe and thus on the material-specific dielectric functions at the probe frequency. Far-field reflections modify the driving field in equation (27) to

$$\tilde{E}_{\text{NF}} \propto \tilde{\alpha}_{\text{eff}}\,[(1+r_{\text{p}})^2\tilde{E}_{\text{in}}], \tag{28}$$

in which r_{p} is the total reflection coefficient of the multilayer sample, which can be calculated through transfer matrix methods. The factor $(1+r_{\text{p}})^2$ arises because a component of the incoming field is reflected from the sample and onto the tip, increasing the field incident on the tip by $(1+r_{\text{p}})$. The tip-scattered field is then reflected from the sample, which further increases the collection by the same factor. The transient $\Delta\tilde{E}_{\text{NF}}$ can then be modelled with far-field reflections accounted for using

$$\Delta\tilde{E}_{\text{NF}} \propto \Delta\tilde{\alpha}_{\text{eff}}\,[(1+r_{\text{p}})^2\tilde{E}_{\text{in}}] + (2\Delta r_{\text{p}})\tilde{\alpha}_{\text{eff}}\,[(1+r_{\text{p}})\tilde{E}_{\text{in}}], \tag{29}$$

in which $\Delta\tilde{\alpha}_{\text{eff}}$ and Δr_{p} cannot be calculated solely from the transient $\Delta\tilde{\varepsilon}_{\text{s}}$ owing to the nonlinear dependence on $\tilde{\varepsilon}_{\text{s}}$ and are instead calculated as

$$\Delta\tilde{\alpha}_{\text{eff}} = \tilde{\alpha}_{\text{eff}}\,(\tilde{\varepsilon}_{\text{s}}+\Delta\tilde{\varepsilon}_{\text{s}}) - \tilde{\alpha}_{\text{eff}}\,(\tilde{\varepsilon}_{\text{s}}),\ \text{and} \tag{30}$$

$$\Delta r_{\text{p}} = r_{\text{p}}(\tilde{\varepsilon}_{\text{s}}+\Delta\tilde{\varepsilon}_{\text{s}}) - r_{\text{p}}(\tilde{\varepsilon}_{\text{s}}). \tag{31}$$

In practice, s-SNOM spectra and ultrafast s-SNOM spectra can be fit with equations (28) and (29), respectively, choosing an appropriate model for $\tilde{\varepsilon}_{\text{r}}$ based on the sample of interest. Parameters describing the nano-localized material response, such as centre frequency and linewidth for samples with narrow resonances, or the plasma frequency, ω_{p}, and carrier density, n, for samples described by a Drude response, can then be extracted from measured spectra. Modelling the tip–sample interaction with a point-dipole approach and accounting for sample reflections has been used to successfully interpret the ultrafast nano-localized near-field response in perovskite films[84], semiconductor nanowires (NWs)[1] and black phosphorus[4]. Further efforts have been made to treat excitation heterogeneity within the near-field localization volume by an additional layer[85].

Extended models of tip–sample interactions

The PDM effectively describes tip–sample interactions in molecular[28], semiconducting[86] and phonon-resonant samples[27]. However, the frequency and linewidth of measured spectral features can be modelled with higher accuracy through phenomenological models, which extend the tip dipole model using different effective geometries leading to modified $\tilde{\alpha}_{\text{eff}}$ and in some cases better agreement with experimental results. A widely used approach is to represent the probe tip as an elongated multipole in the shape of an ellipsoid with multiple point charges along its major axis, known as the finite-dipole model[36]. The dimensions of the ellipsoid and positions of the charges are then additional model parameters to better represent the actual tip polarization.

Both PDM and finite-dipole model generally assume a flat topography and homogeneous vertical composition of the samples. Reflections and scattering from sample inhomogeneities, however, can greatly influence near-field spectra and images[75]. To account for these effects, comparing near-field spectra of different tip harmonics by analysing the ratios of adjacent harmonics $\frac{\tilde{E}_{\text{NF},n+1}}{\tilde{E}_{\text{NF},n}}$ has been proposed[87]. Alternatively, full numerical simulations can be implemented with arbitrarily complex sample geometries[74–76]. Numerical approaches, however, can be computationally intensive and therefore struggle to extract an unknown $\tilde{\varepsilon}_{\text{r}}$ from a measured $\tilde{E}_{\text{NF}}$ or equivalently the transient $\Delta\tilde{\varepsilon}_{\text{r}}$ from the measured $\Delta\tilde{E}_{\text{NF}}$. Machine learning for backward extraction has recently been demonstrated in ground-state s-SNOM[88–90]. Future applications to the transient near-field response may provide further insights into the nano-localized ultrafast response in a broad range of materials.

Spectral modelling

In pump–probe spectroscopy, the transient spectral response encodes the perturbed electronic or vibrational states of the material following photoexcitation relating to the formation and dynamics of new quasi-particle states. This section describes how to fit ultrafast near-field data including the basic models of a free-electron response using the Drude model and interband electronic or IR dipole-active vibrational resonances using the Lorentz-oscillator model.

Drude model. For metals and many photoexcited or doped semiconductors, the Drude model delivers an accurate description of the optical properties over a broad frequency range, although it reduces carrier–carrier and carrier–lattice interactions to a single frequency-independent unspecific (momentum) scattering rate given by the inverse of the Drude scattering time, τ_D. The Drude model is based on the assumption that electrons of an effective mass, m^*, and charge, e, are driven by an optical field of frequency, ω, and are subject to a damping rate, $\gamma \propto 1/\tau_D$. The equation of motion reads:

$$m^* \frac{\mathrm{d}\upsilon(t)}{\mathrm{d}t} + m^* \gamma\ \upsilon(t) = -e\ E_0 e^{-i\omega t}. \tag{32}$$

Its solution can be used to calculate the current density and ultimately the frequency-dependent conductivity, $\tilde{\sigma}(\omega)$

$$\tilde{\sigma}(\omega) = \frac{ne^2}{m^*(\gamma - i\omega)}, \tag{33}$$

or dielectric function, $\tilde{\varepsilon}(\omega)$

$$\tilde{\varepsilon}(\omega) = \varepsilon_\infty + \frac{i\ \tilde{\sigma}(\omega)}{\varepsilon_0\ \omega} = \varepsilon_\infty + \frac{i\ (ne^2/m^*)}{\varepsilon_0\ \omega(\gamma - i\omega)}. \tag{34}$$

Alternatively, $\tilde{\varepsilon}(\omega)$ is also commonly expressed by the plasma frequency $\omega_p^2 = \frac{ne^2}{\varepsilon_0 m^*}$, determined by the carrier density, n, and collision time, $\tau_D = 1/\gamma$, as

$$\tilde{\varepsilon}(\omega) = \varepsilon_\infty - \frac{\omega_\mathrm{p}^2}{\omega^2 + i\omega/\tau_D}. \tag{35}$$

Depending on whether one follows the conventions of IR or THz spectroscopy, the Drude model is typically written in terms of the dielectric function or the optical conductivity, respectively[91]. The effective mass, m^*, differs from the free-electron mass primarily because the electronic band structure modifies the dispersion relation, $E(k)$. Many-body interactions can further renormalize m^*, depending on the importance of electron–electron interactions. Similarly, ε_∞ different from one arises from the effective contributions from phonons or bound electronic states at higher energies. The Drude permittivity diverges for $\omega \to 0$, resulting in the so-called Drude peak, as is characteristic for a relaxation of free as opposed to bound charges with optical conductivity increasing with decreasing frequency. The validity of the Drude model is commonly assumed when the transient absorption scales quadratically with the IR probe wavelength.

This model is generally applicable to intraband transitions in metals, doped semiconductors, including semiconductor nanostructures, or molecular systems, in which it captures the free-carrier response in the IR to THz spectral range for low carrier densities[92,93]. However, it assumes that carriers move in a continuous, homogeneous medium with isotropic scattering and that a single, frequency-independent scattering time, τ_D, describes the entire carrier dynamics. In low-dimensional systems, disordered materials or systems with strong quantum confinement, the carrier response at small frequencies may deviate from this simple description. The Drude–Smith model extends the Drude approach by including an additional phenomenological parameter that accounts for carrier backscattering and localization. This modification reduces the low-frequency conductivity relative to the standard Drude response and has been successfully applied to NWs, nanoparticle films and other confined systems in which localization effects are prevelant[94,95].

Harmonic oscillator model. Unlike the Drude-based models for free or quasi-free carriers, the harmonic oscillator, or Lorentz-oscillator model, describes bound excitations. In the Lorentz-oscillator model, a restoring force yields the classical driven and damped harmonic oscillator. The dielectric function, considering multiple uncoupled transitions, j, reads:

$$\tilde{\varepsilon}(\omega) = \varepsilon_\infty + \sum_j \frac{S_j \omega_{0,j}^2}{\omega_{0,j}^2 - \omega^2 - i\omega\Gamma_j} \tag{36}$$

with individual oscillator strength, S_j, resonance frequencies, $\omega_{0,j}$, and linewidths, Γ_j.

Although the model predicts a Lorentzian lineshape for the absorption coefficient, observed lineshapes are typically convoluted to a normal distribution owing to static disorder and stochastic fluctuations. The resulting absorbance spectra are then classified in terms of homogeneous and inhomogeneous broadening, with Gaussian, Lorentzian or intermediate Voigt profiles. In the inhomogeneous limit, in which correlation times are long and the environment is effectively static during the optical interaction, the transition frequency distribution is preserved and the absorption profile assumes a Gaussian form. By contrast, in the homogeneous limit, corresponding to short correlation times (and motional narrowing), rapid fluctuations average over all accessible frequencies, leading to exponential decay of the dipole correlation function and a Lorentzian lineshape. The characteristic timescale governing this crossover is the correlation time, τ_c, with the fluctuation rate determined by its inverse, $1/\tau_c$ (refs. 96,97).

The Lorentz model is strictly valid only for a homogeneous material response, but nano-localized spectra, despite contributions from inhomogeneous broadening, are frequently well described by Lorentzian lineshapes, suggesting that homogeneous sample size limits are reached in some experiments. Fits are typically performed for each pump–probe delay time to track the temporal evolution of resonance frequencies, spectral linewidths or relaxation rates. When appropriate, global or constrained fitting is used to improve parameter stability and physical consistency across time delays. These dynamics of the extracted spectral parameters then provide insight into the fundamental processes that govern the material properties and allow comparison with theoretical models.

Drude–Lorentz model. In the presence of multiple discrete vibrational or electronic transitions superimposed on top of a free-carrier response, the transient spectral response may be described by a superposition of multiple Lorentz and Drude terms from above, with the overall complex valued dielectric response given by:

$$\tilde{\varepsilon}(\omega) = \varepsilon_\infty - \frac{\omega_\mathrm{p}^2}{\omega^2 + i\omega\gamma} + \sum_j \frac{S_j \omega_{0,j}^2}{\omega_{0,j}^2 - \omega^2 - i\omega\Gamma_j}. \tag{37}$$

Fitting this Drude–Lorentz model allows simultaneous tracking of carrier and polaron dynamics and excitonic or vibrational resonances. This model has been applied to various semiconductor and nanostructured materials such as transition metal dichalcogenides[10], vanadium dioxide[98], the photophysics of hybrid perovskites[6,7] and other materials[99,100].

Electron–phonon coupling and the displaced oscillator model. The Lorentz model assumes that each resonance is uncoupled and harmonic, neglecting interactions between different oscillators and nonlinear effects. In many molecular and solid-state systems, electronic excitations are coupled to vibrational degrees of freedom. This coupling can be described by extending the simple harmonic oscillator model to a coupled oscillator framework, which accounts for vibration–vibration or electron–phonon (el–ph) interactions, for example.

For a molecule in an electronic excited state, the equilibrium bond distance of the nuclear coordinate is typically displaced relative to the ground-state configuration. The Hamiltonian that describes this electronic two-level system can be written as displaced harmonic oscillators using the Born–Oppenheimer approximation, with the Hamiltonian given by

$$\begin{aligned} H &= H_{\mathrm{el}} + H_{\mathrm{ph}} + H_{\mathrm{el-ph}} \\ &= \hbar\omega_0\left(b^{\dagger}b + \frac{1}{2}\right) + \hbar\omega_e|e\rangle\langle e| + \hbar g(b^{\dagger} + b)|e\rangle\langle e| + \hbar\omega_0\Delta^2|e\rangle\langle e|, \end{aligned} \tag{38}$$

in which $b^{\dagger}$ and b are the phonon creation and annihilation operators, $|e\rangle$ is the excited electronic state and $\Delta = g/\omega_0$ represents the displacement of the vibrational potential in the excited state. The term $\hbar g(b^{\dagger} + b)|e\rangle\langle e|$ then captures the el–ph coupling with coupling constant g, whereas $\hbar\omega_0\Delta^2|e\rangle\langle e|$ represents the associated reorganization energy ΔE_{RO}, related to each other by the Huang–Rhys factor S:

$$S = \frac{\Delta E_{\mathrm{RO}}}{\hbar\omega_0} = \left(\frac{g}{\omega_0}\right)^2 \tag{39}$$

with the vibrational frequency, ω_0.

A larger Huang–Rhys factor indicates a larger shift of the nuclear configuration upon electronic excitation, resulting in more pronounced vibronic features in optical absorption and emission spectra of the electronic transition. This framework provides an intuitive picture of how electronic excitation reshapes the nuclear potential energy surface, giving rise to the characteristic vibronic structure observed in molecular and solid-state optical spectra[101].

As discussed earlier, the relative dielectric permittivity $\tilde{\varepsilon}_{\mathrm{r}}(\omega)$ describes how a material responds to an optical field and is governed by the energies and strengths of its allowed optical transitions. When electronic and vibrational degrees of freedom couple, referred to as vibronic coupling in molecules or el–ph coupling in solids, the electronic transition is no longer isolated but interacts with a ladder of vibrational states. This interaction reshapes the spectral profile of the transition and generates a characteristic vibronic structure. Each vibrational level contributes to the overall response, with its strength determined by the Franck–Condon factors, which quantify how much the nuclear configuration shifts upon electronic excitation. As a result, the susceptibility and therefore the dielectric function develop multiple phonon-assisted resonances. These manifest as a series of spectrally resolved peaks in addition to the zero-phonon line (0–0 transition, when present), or, when the peaks are not individually resolved, as an overall asymmetric broadening of the spectrum. In either case, the absorption features gain additional sidebands and a broadened lineshape.

Polarons. In polar materials, el–ph coupling is dominated by the long-range Coulomb interaction between longitudinal optical (LO) phonons and charge carriers described by the Fröhlich interaction[102]. In nonpolar materials, the deformation potential gives rise to el–ph coupling via modulation of the interatomic distance. Given sufficiently strong el–ph coupling, the electron can no longer be treated as predicted by band theory. Instead, it becomes dressed by a large polarization cloud through lattice deformation, forming a quasi-particle known as a polaron.

Polarons exhibit increased effective mass and altered transport and optical properties. Depending on the strength of the el–ph coupling and the recursive nature of the interaction, it is often distinguished between small and large polarons. Large polarons, extending over many unit cells of the lattice, feature an effective mass on the order of a few m_e and can be considered as mobile. By contrast, for small polarons, the charge-induced deformation causes an additional localization of the electronic wavefunction, resulting in even stronger deformations, effectively trapping the charge carrier to a single lattice site.

Importantly, polarons are a bound state with optically allowed transitions, giving rise to polaron absorption in the THz and IR range[103]. The absorption coefficient α per density of phonons n_p for large polarons is given by

$$\frac{\alpha}{n_p} = \frac{128\pi e^2}{3m\omega c}\frac{(kR)^3}{[1 + (kR)^2]^4} \tag{40}$$

with polaron mass, m, and radius, R[103]. The polaron wavevector $k \equiv \sqrt{2m(\hbar\omega - 3E_p)}/\hbar$ incorporates the polaron binding energy, E_p. Small polarons feature higher binding energies and are thus optically active in the visible and NIR. Small polarons are similar to solvated electrons in liquids[104,105].

Modelling ultrafast s-SNOM transients

Scanning the pump–probe time delay in ultrafast s-SNOM captures the relaxation dynamics of the induced ultrafast perturbation. If two-phase, three-phase or pseudo-heterodyne detection is used, the amplitude of the transient near-field signal, R_{HPP}, represents the pump-induced sample perturbation. In broadband nano-FTIR or EOS detection, spectral characteristics such as the pump-induced change in amplitude or shift in centre frequency of a vibrational resonance would be the figure of merit, dependent on the system being studied. The time dynamics of the pump-induced sample perturbation, referred to here as $f(T)$ for generality, describe the dynamic relaxation of charge carriers, conductivity, interband excitations, molecular vibrations, phonon population, spin or magnetization, or indirectly, temperature, quantum phases or evolving chemical species or other photoinduced responses.

Instrument response function. The relaxation dynamics measured in ultrafast spectroscopy are convoluted with the instrument response owing to the photo-injection dynamics of carriers and the finite cross-correlation between pump and probe pulses. Modelling ultrafast dynamics, in general, require separating the intrinsic dynamics of the material from the instrument response.

The instrument response function (IRF) captures the temporal profile of the excitation pulse convoluted with the probe pulse.

The temporal profile of the pump and probe pulses can be determined using autocorrelation or cross-correlation techniques such as FROG or XFROG. The cross-correlation between the two pulses then determines the IRF[106]. The IRF can be approximated by a Gaussian function:

$$\mathrm{IRF}(T) = A\exp\left(-\frac{4\ln 2\, T^2}{\Delta T_{\mathrm{IRF}}^2}\right), \tag{41}$$

in which ΔT_{IRF} is the width of the IRF, given by the quadrature sum of the widths of the pump and probe pulses as

$$\Delta T_{\mathrm{IRF}}^2 = \Delta T_{\mathrm{pump}}^2 + \Delta T_{\mathrm{probe}}^2. \tag{42}$$

Alternatively, ΔT_{IRF} is often characterized more directly by measuring the temporal dynamics of a reference sample, which relaxes much faster than the pulse duration, such as for gold where $\tau \sim 10$ fs. The measured time trace is then dominated by the IRF with only a small modification from carrier relaxation.

For markedly chirped pulses, the effective instrument response can deviate from a Gaussian and exceed the transform-limited cross-correlation width[107]. In the special case of ultrafast, passively mode-locked oscillators, and related femtosecond lasers, the temporal profile of the pump pulse is better described by a hyperbolic secant squared (sech^2). The measured signal is then modelled as the convolution of the intrinsic dynamics $f_{\mathrm{intr}}(T)$ with the IRF:

$$f(T) = f_{\mathrm{intr}}(T) \circledast \mathrm{IRF}(T). \tag{43}$$

The intrinsic dynamics, $f_{\mathrm{intr}}(T)$, are then fit with various decay models with the IRF removed.

Exponential decay dynamics. A simple model to describe non-interacting excited-state populations, i, uses a multiexponential decay:

$$f(T) = \sum_i A_i e^{-k_i T} + C, \tag{44}$$

with amplitude, A_i, relaxation rate, $k_i = 1/\tau_i$, and with each term corresponding to a distinct physical process with its own characteristic dynamics. This approach is useful when multiple mechanisms act sequentially, such as initial fast carrier cooling, followed by slower carrier–lattice equilibration and, on longer timescales, multistate trapping dynamics[108]. This model comes with the risk of over-parameterizing if too many terms are used, causing the amplitudes and decay rates to become interdependent in the fitting procedure, especially when the timescales are similar and thus physically meaningless. Moreover, for many processes such as carrier cooling, the assumption of time-independent relaxation rates, which implies a carrier density independence, is not justified.

For heterogeneous materials, such as amorphous semiconductors or nano-materials, the relaxation rates may even vary continuously rather than being discrete quantities. In these cases, the dynamics can be described by a stretched exponential

$$f(T) = Ae^{-(k_{\mathrm{str}}T)^{\beta}},\ 0 < \beta < 1, \tag{45}$$

in which β is the stretching exponent. The stretched exponential has the advantage of describing time trace dynamics that would otherwise require multiple exponential terms with a reduced number of parameters, k_{str} and β to obtain a satisfactory fit. However, although qualitatively representing a time-varying dynamics, the fit parameters themselves do not necessarily map directly onto microscopic physical quantities[109].

Generalized rate equation. In the case of interacting excited states or quasi-particles, one can apply a general rate equation considering each subpopulation, n_j, and the corresponding decay rates

$$\frac{\mathrm{d}n_j}{\mathrm{d}T} = S_j(T) - \sum_{i,d} f_{d,ij}(n_i), \tag{46}$$

in which $f_{d,ij}$ represents the functional dependence of different processes, d, for the decay or conversion rate between populations, n_i and n_j, and $S_j(t)$, the generation term. For linear excitation of the sample, $S_j(t)$ is equal to the temporal envelope of the pump pulse.

A special case of the general rule is the so-called ABC model (in which A, B and C are coefficients), commonly applied in semiconductors, which treats the decay of a population of electrons, n_1, and holes, n_2, by different orders of their density[110]. In the ABC model, $f_1 = k_1 n_1$ corresponds to the Shockley–Read–Hall recombination (which is density-independent); $f_2 = k_2 n_1 n_2$ corresponds to radiative recombination (or other bimolecular recombination process, with decay rate depending on the carrier density to second order); and $f_3 = k_{3,e}(n_1)^2 n_2 + k_{3,h} n_1 (n_2)^2$ corresponds to Auger recombination, with the decay rate depending on the density to third order. Equation (46) for the ABC model is then written as:

$$\frac{\mathrm{d}n}{\mathrm{d}T} = S(T) - k_1 n - k_2 n^2 - k_3 n^3. \tag{47}$$

This approach connects the observed phenomenological decay constants to different carrier density-dependent recombination mechanisms. This deconvolutes the contributions of mechanisms to the relaxation dynamics, such as in InGaAs NWs in which Auger, bimolecular and Shockley–Read–Hall recombination all occur[111]. Different models can be distinguished by probing a wide dynamic range over several timescales (for example, 50 fs to few nanoseconds) and performing complementary analyses, such as fluence dependence.

In the application to ultrafast s-SNOM for the study of semiconductors, the ABC model (along with fluence-dependent data acquisition) is commonly used to distinguish different microscopic processes. For III–V semiconductors, it was found that radiative recombination (k_2) dominates[112]. Similarly, ultrafast s-SNOM of perovskites under far-from-equilibrium excitation was also analysed with the ABC model to spatially resolve Auger recombination, which is crucial for high power optoelectronic devices such as LEDs and lasers[7].

Applications

Ultrafast carrier dynamics in semiconductor nanostructures

The continuous miniaturization of semiconductor devices together with the search for faster yet more energy-efficient microelectronic devices makes the nanoscale characterization of dynamic processes in semiconductor nanostructures and nano-devices highly desirable. Semiconductor nanostructures are characterized by large surface-to-volume ratios, thickness comparable to the mean-free path of carriers and phonons, and surfaces prone to defects due to dangling bonds and oxidation. These effects all lead to enhanced carrier scattering.

Ultrafast s-SNOM provides the desired nanometre spatial and femtosecond temporal resolution to track the dynamics of carrier cooling, transport, and recombination on their characteristic time

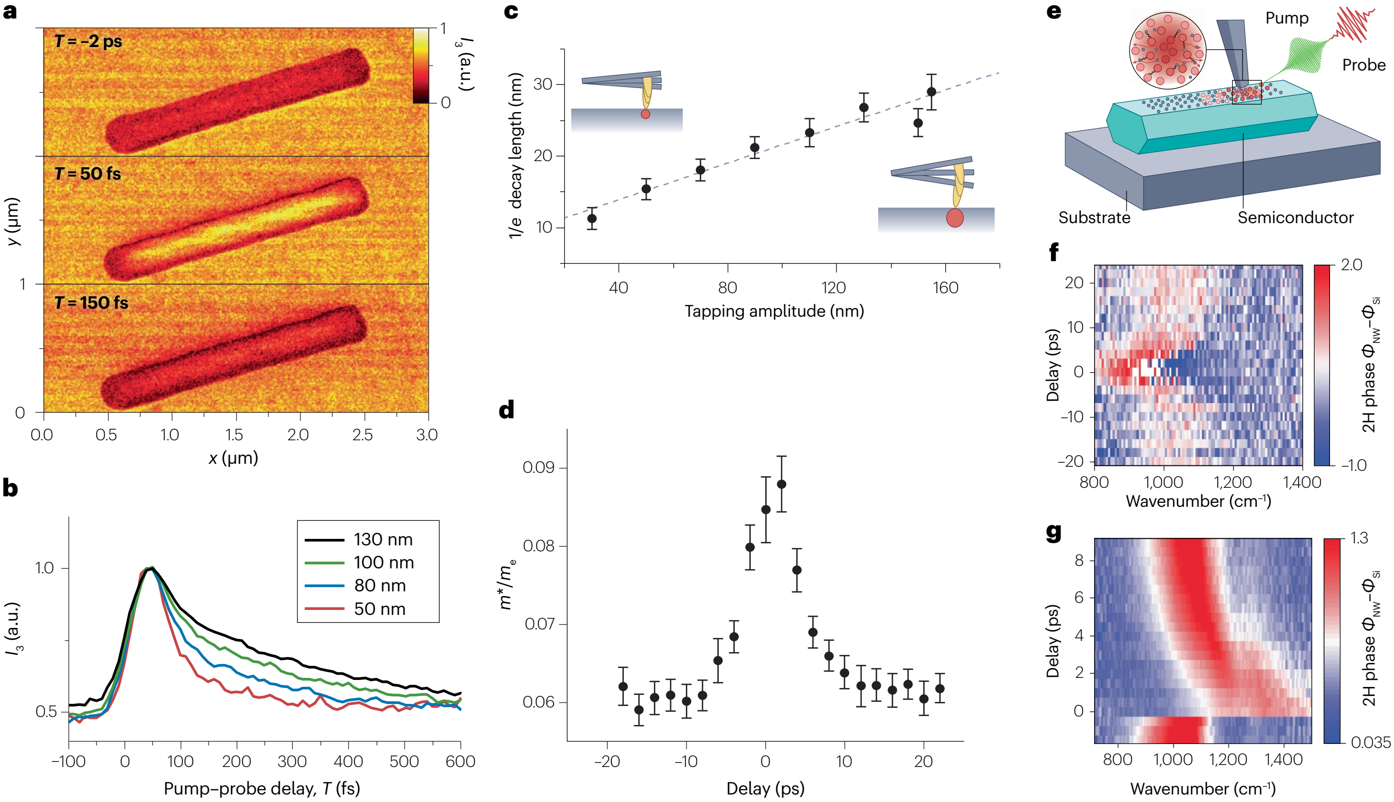

Fig. 4 | Ultrafast pump–probe nano-imaging of carrier dynamics in semiconductor nanostructures. **a**, Ultrafast nano-movie frames of an InAs nanowire (NW) photoexcitation. Here, *x* and *y* are the spatial coordinates of the image pixels, I_3 is the third harmonic near-field intensity, and *T* is the pump–probe delay time. **b**, Time traces of the transient near-field intensity measured on InAs NWs for various tapping amplitudes. **c**, Decay length of near-field intensity as a function of tip-tapping amplitude from vertical scans over a gold reference sample. **d**, Time dependence of total effective mass in $In_{0.44}Ga_{0.56}As$ NWs. **e**, Schematic of hot-electron dynamics in semiconductor nanostructures. **f**,**g**, Colour maps showing the evolution of the NW near-field phase referenced to silicon $\phi_{NW}(\omega) - \phi_{Si}(\omega)$ spectra of a doped InGaAs NW upon THz (part **f**) and near-IR photoexcitation (part **g**). a.u., arbitrary unit. Panels **a**–**c** adapted from ref. 1, Springer Nature Limited. Panels **d** and **f** adapted from ref. 112, CC BY 4.0. Panel **g** adapted from ref. 114, CC BY 3.0.

and length scales. Owing to the large transition dipole moment of electronic interband transitions and the discrete quantum states arising from finite size effects, semiconductor nanostructures served as an early testbed in the development of ultrafast near-field spectroscopy already before the advent of s-SNOM[113].

Ultrafast nano-imaging in the weakly perturbing regime. Ultrafast s-SNOM has been applied to InAs as a representative small direct bandgap semiconductor to study the time evolution of plasma frequency and carrier density (Fig. 4e). The momentum provided by the probe tip enables the study of surface plasmons and their dispersion in pristine, high-quality, flat surfaces without the need for nanofabricating gratings[2]. Far-field spectroscopy can only couple to high-momentum excitations, such as surface plasmons, by studying nanostructured samples, which alter the material response. By nano-FTIR probing in the spectral range of 5–16 μm, the transient tunability of surface plasmons in InAs at low fluences following pump excitation at 1.56 μm could be demonstrated[2].

Extending this initial demonstration of the sensitivity of ultrafast s-SNOM to semiconductor surface plasmons, semiconductor nanostructures were studied for their technological relevance[1,112,114,115]. Initially, a visible or NIR pump excitation was used to create electron–hole pairs far above the bandgap, which subsequently relax to the band edge and recombine[1,114,115]. Instead of pump modulation and sideband detection, in early experiments, pump-induced effects were isolated by taking the difference between near-field spectra acquired for positive and negative pump–probe delays[1]. Using the PDM, the dielectric function of a semiconductor NW could be extracted and described by the Drude model, yielding the spatially and temporally resolved plasma frequency and thus carrier density. An unusually fast initial decay of the carrier density, as shown in Fig. 4c, is analysed by introducing femtosecond tomography, in which the tip-tapping amplitude is varied to control the depth of s-SNOM probing to resolve carrier dynamics in regions of varying thickness below the surface[1] (Fig. 4a,b). As can be seen in Fig. 4c, the fastest dynamics are observed for a more strongly localized near field, pointing to the ultrafast formation of a depletion layer on the NW surface.

Although the tapping amplitude can be adjusted to control the probe depth, the near-field localization can be equally tuned by the lock-in demodulation harmonic, in which higher-order demodulation yields stronger spatial confinement[35,57]. Because the tip–sample distance is harmonically modulated, higher harmonics selectively extract scattering components with a nonlinear dependence on tip–sample distance, isolating the most highly localized near-field contribution closer to the surface. Carrier dynamics in GaAs nanostructures have been shown to differ across demodulation orders, reflecting distinct carrier behaviour at the semiconductor surface[116]. By contrast, in conventional ground-state s-SNOM, simultaneous analysis across multiple demodulation orders has enabled depth-resolved characterization of dielectric layers, including subsurface structures[82,117]. The same multiharmonic approach is potentially useful for ultrafast s-SNOM as well, leading to 4D tomography of photoinduced carriers and excitons.

Ultrafast s-SNOM was extended to study the effect of doping, diameter and pump fluence dependence in NWs of other semiconductors also using excitation far above the bandgap[114,115]. For silicon NWs, a linear dependence of the charge carrier lifetime with the NW diameter was observed[115]. The surface-to-volume ratio of NWs with diameter, d, can be written as $4/d$ and thus the surface-related decay rate as $4\,k_{\mathrm{SRV}}/d$ with the surface recombination velocity k_{SRV}. As the lifetime is the inverse of the sum of all decay rates, a linear dependence, $\tau(d)$, indicates that surface recombination dominates in the silicon NWs studied[115]. Near-field spectroscopy was used to select a single NW, rather than an ensemble average of NWs with different diameters as in ultrafast far-field spectroscopy.

Most recently, doped semiconductor NWs have been studied with pump fluences spanning two orders of magnitude to provide a broader picture of carrier-mediated recombination in III–V semiconductor NWs, at the example of $\mathrm{In_{0.44}Ga_{0.56}As}$ (ref. 114). In this work, sideband demodulation was shown to provide sufficient excited-state contrast to isolate the small pump-induced change of the dielectric function and thus the tip-scattered field even for the weakest pump fluence studied of 15 μJ cm^{-2}. The results were analysed using a generalized rate equation model observing an inverse relationship between carrier lifetime and carrier density, which is the sum of intrinsic doping and photoexcited charge carriers, attributed to bimolecular or radiative recombination[114]. Furthermore, ultrafast s-SNOM with THz excitation has been applied to $\mathrm{In_{0.44}Ga_{0.56}As}$ NWs to study the dynamics of hot carrier cooling[112]. Typically, excitation in the visible or NIR frequency range is far above the bandgap of the studied semiconductors, convoluting the dynamics of hot carrier cooling and recombination. The required, intense THz beam was generated by a free-electron laser, showcasing the variety of laser sources suitable for ultrafast near-field optical microscopy. To study the pure hot carrier dynamics, without the convoluting effect of Auger recombination heating the carrier population, for example, a THz-pump MIR-probe experiment was performed. The few-picosecond THz-pump pulse with a wavelength of 25 μm was absorbed by the intrinsic carriers owing to intraband transitions, also referred to as the free-carrier absorption, to raise the temperature of the electronic subsystem without creating additional electron–hole pairs. Owing to the non-parabolicity of the Γ valley, which is the conduction-band minimum at the Γ symmetry point of the Brillouin zone, and the occupation of the higher-energy X and L valleys, the effective mass is a function of carrier temperature. By extracting the effective mass via the Drude model from MIR spectra in the region of the plasma frequency, the electron gas temperature evolution can be studied. Figure 4d shows the effective electron mass m^{*}/m_e in $\mathrm{In_{0.44}Ga_{0.56}As}$ NWs plotted against delay, T, derived from the temporo-spectral map shown in Fig. 4f (the pump–probe phase response $\Delta\phi_{\mathrm{NF}}(T)$ plotted against T as a 2D colour map). By incorporating a broadband laser source with a range of probe energies, the intrinsic carrier doping in the NWs can be studied (Fig. 4g). Figure 4g shows another temporo-spectral map of $\phi_{\mathrm{NF}}(T)$ indicating ~300 cm^{-1} blue shift of the NW plasma frequency upon THz excitation owing to the increase in carrier density. The shift of the resonance back to before-excitation is typically assigned to charge carrier recombination.

These studies on semiconductors and semiconductor nanostructures have established the capability of ultrafast s-SNOM to provide momentum allowing for excitation of surface plasmons[2], to provide nanoscale spatial resolution useful for selecting single nanostructures[115] and as a tomographic approach to measure depth-dependent ultrafast dynamics[1]. Ultrafast s-SNOM studies have not identified any considerable spatial heterogeneity in the dynamics in the semiconductor nanostructures studied, which is an additional benefit of the method applicable to samples exhibiting nanoscale heterogeneity.

Strong-coupling and strong-field effects. Experiments using ultrafast s-SNOM on semiconductor nanostructures have so far been carried out in the weak perturbation regime, that is, without modification of the quantum states of the material themselves through the optical field. In this section, we describe the extension of ultrafast s-SNOM into the strong-field regime based on the strong transition dipole moment of the intersubband electronic transition in semiconductor multiquantum well (MQW) heterostructures. MQWs are semiconductor heterostructures with engineered electronic states, tunable based on the thickness of each semiconductor layer, offering strong optical nonlinearity, ultrafast response and chip-scale integration[118,119]. When resonantly coupled to metal NW antennas, strong coupling between the MQW intersubband transition and metal NW antennas results in hybridization of the MQW electronic resonance with the NW plasmonic mode leading spectroscopically to a peak splitting between the upper and lower polariton modes. On resonance, the splitting between the polariton peaks is described by the Rabi frequency:

$$\Omega \propto \mu \sqrt{\frac{2N\hbar\omega_{\mathrm{MQW}}}{\epsilon V}}, \tag{48}$$

in which μ is the transition moment, ω_{MQW} is the MQW transition centre frequency (degenerate with the NW resonance) and V is the cavity mode volume.

The high fluence of picosecond and femtosecond pulses allows for saturation of the intersubband transitions leading to a ground-state bleach and thus reduction in the total transition moment of the 0–1 transition[120,121]. Reduction of μ subsequently reduces Ω, tuning the coupled system from the strong-coupling to weak-coupling regime based on pump fluence[120]. In the application of s-SNOM, the position of the probe tip along the MQW or antenna structure modified the interference between driving field and antenna resonance controlling saturation by tip position alone. Ultrafast switching from strong to weak coupling has been studied in far-field IR-pump IR-probe spectroscopy[120], and a clear next step for the field of ultrafast nano-spectroscopy is to extend to the degenerate IR-pump IR-probe regime where also the tip-induced interference dependence of the ultrafast dynamics can be studied.

Beyond semiconductor heterostructures, ultrafast switching may also be applicable to molecular vibrations strongly coupled to plasmonic antennas rather than the MQW intersubband transitions. Although the MQW intersubband transition has a much larger transition dipole moment μ compared with molecules ($\mu/e = 1.76$ nm for the MQW studied in ref. 120 compared with 0.02 for the C–O vibration of poly(methyl methacrylate), the oscillator density of vibrations is much larger than the doping density of MQWs ($\Delta n_0 = 10^{18}$ cm^{-3} for the MQWs compared with $n = 10^{22}$ cm^{-3} for the C–O vibration in poly(methyl methacrylate). On the basis of equation (48), the two factors cancel leading to a similar coupling rate if the MQWs were replaced by a molecular film. Degenerate ultrafast s-SNOM can then be extended beyond semiconductor heterostructures to also probe and control the ultrafast dynamics of strongly coupled molecular solids in the MIR spectral range.

Ultrafast and multimodal nano-imaging of quantum materials

Correlated electron and other quantum materials display a wide range of phenomena emerging from many-body electron, lattice and spin interactions giving rise to superconductivity, metal–insulator transitions, magnetoelectric effects and other quantum effects. These materials also exhibit associated quantum phases and phase transitions, which can be induced by strain, temperature, electric and magnetic fields. Doping and impurities can sensitively influence these properties, which are often related to domain formation or phase coexistence on length scales ranging from the atomic scale to the mesoscale. However, the role of these and other spatial inhomogeneities is yet poorly understood, even the question to which extent certain phenomena are intrinsic or extrinsic to the material. Optical nano-imaging has therefore been widely adopted to complement the range of scanning probe, electron and X-ray imaging techniques. Taking advantage of the range of optical contrast mechanisms sensitive to low-energy electron, phonon, polariton, magneto-optical and other degrees of freedom and symmetries, they can probe and resolve the role of heterogeneities on the equilibrium quantum properties and their association with local strain or doping[122].

The implementation of ultrafast s-SNOM with far-from-equilibrium excitation extends the insight that can be gained, in particular with regard to coupling and dynamics of the electronic, lattice and spin interactions on the nanoscale. The associated quasi-particle excitations can further hybridize to form mixed modes, such as plasmon polaritons in graphene[3,50], Kane plasmon polaritons in HgCdTe[123], exciton polaritons in WSe_2 (ref. 124), phonon polaritons[125–127] and electron–photon modes in Dirac fluids[128]. Many of these interactions are often themselves transient and localized, making nanoscale spatial, spectral and temporal resolution desirable in their investigation. In this section, we discuss ultrafast s-SNOM to image spatial heterogeneity and associated dynamics of the IMT in VO_2 and exciton dynamics in WSe_2, as two prominent examples to illustrate the application potential of the technique to quantum materials.

Insulator-to-metal transition in vanadium dioxide. VO_2 exhibits a reversible IMT phase transition that can be induced both thermally (T_{IMT} ~ 340 K)[129,130] and optically[17]. The thermal IMT is a sharp first-order phase transition at 340 K, from a monoclinic insulating phase to a rutile metallic phase, associated with strong electron–lattice coupling[4,7,9,131,132]. The details of this phase transition behaviour are complex and sensitive to a wide range of parameters such as local strain, pressure and doping[23,133,134] (Fig. 5c). Despite extensive research, the role of coupling between the electronic and structural degrees of freedom and their influence on the IMT have remained inconclusive[17,129].

Initial studies focusing on polycrystalline films[135,136] have been extended to single VO_2 microcrystals[137] of well-defined crystallographic orientation, strain and temperature-dependent IMT[138]. With strong near-field contrast based on the Drude response of the metallic phase, continuous wave IR s-SNOM has been used to study the nano-domain behaviour during the thermally induced insulator-to-metal phase transition, with an example of s-SNOM image of a VO_2 microcrystal[139] shown in Fig. 5a, which shows mesoscopic metallic stripe domains perpendicular to the *c* axis before breaking into smaller domains at lower temperature. The complex domain texture indicates that beyond strain, other structural heterogeneities or electronic defects sensitively influence the IMT, which is unresolvable in static nano-imaging. This hypothesis was further supported by ultrafast microscale pump–probe measurements of individual VO_2 microcrystals[23,140]. To further address not only the nucleation dynamics of the IMT on the nanoscale but also develop ultrafast pump–probe nano-imaging, VO_2 has served as a benchmark for studying non-equilibrium phase transitions and to establish excited-state contrast in ultrafast s-SNOM.

Visible-pump MIR-probe s-SNOM was applied to VO_2 at variable temperatures both below and above the IMT threshold, observing nanoscale correlations between the carrier dynamics and steady-state switching associated with strain[4]. However, the photoinduced IMT was not driven directly and was limited by low pump fluence and the use of a high-repetition rate (20–40 MHz) laser source, creating a quasi-continuous wave situation. The application of 10–100 MHz repetition rate laser sources with low pulse energies would only induce a weak carrier perturbation and thus primarily drive the IMT thermally. Furthermore, the repetition rate is too high to induce and probe the tens of nanosecond lifetime of the true photoinduced IMT, which requires low-repetition rate yet high pulse energies. This motivated the extension of established far-field micro pump–probe spectroscopy with far-from-equilibrium excitation towards the development and demonstration of ultrafast nano-imaging using s-SNOM with sub-megahertz to megahertz amplified laser sources to image long-lived photoinduced quantum phases in matter.

Ultrafast s-SNOM with high-fluence, low-repetition-rate pump and MIR probe spectroscopy was demonstrated using VO_2 nanobeams. Following their pre-characterization via static IR s-SNOM to spatially distinguish between long-range strain and local defects, ultrafast s-SNOM imaging under far-from-equilibrium excitation was established and resolved the spatiotemporal evolution of the long-lived transient metallic phase with nanometre spatial and femtosecond temporal resolution[9]. Experiments have been carried out in femtosecond non-degenerate visible-pump MIR-probe s-SNOM based on a 1 MHz Yb-potassium gadolinium tungstate-amplified laser system with OPA and DFG to generate tunable IR probe pulses from 2 µm to 16 µm to measure the Drude response of the above bandgap (~0.6–0.7 eV) interband transitions of the pump-induced non-thermal transient metallic phase of the VO_2 microcrystals. The pump pulse energy has been varied from weak carrier excitation to high excitation densities, leading to a complete, ultrafast, photoinduced collapse of the band structure into the transient metallic phase.

Within a single microcrystal, spatial variations of the threshold fluence required for the non-thermal transition into the metallic state were observed, as shown in Fig. 5b. Within regions of just few tens of nanometres, the VO_2 behaviour varies locally from subthreshold excitation to excitation into the long-lived (greater than tens of picoseconds) metallic state even for the same excitation fluence. Importantly, these transient domains of varying threshold fluence appear parallel with respect to the crystallographic *c* axis. This is in contrast to the domain formation measured in the thermal IMT of the same VO_2 microcrystals, which is predominantly perpendicular with respect to the crystallographic *c* axis[138,139,141], as illustrated in Fig. 5a. This implies that the ultrafast photoinduced and thermal IMT is sensitive to different aspects of the electronic and lattice structure of VO_2. The results suggest that not strain but primarily the electronic degrees of freedom determine

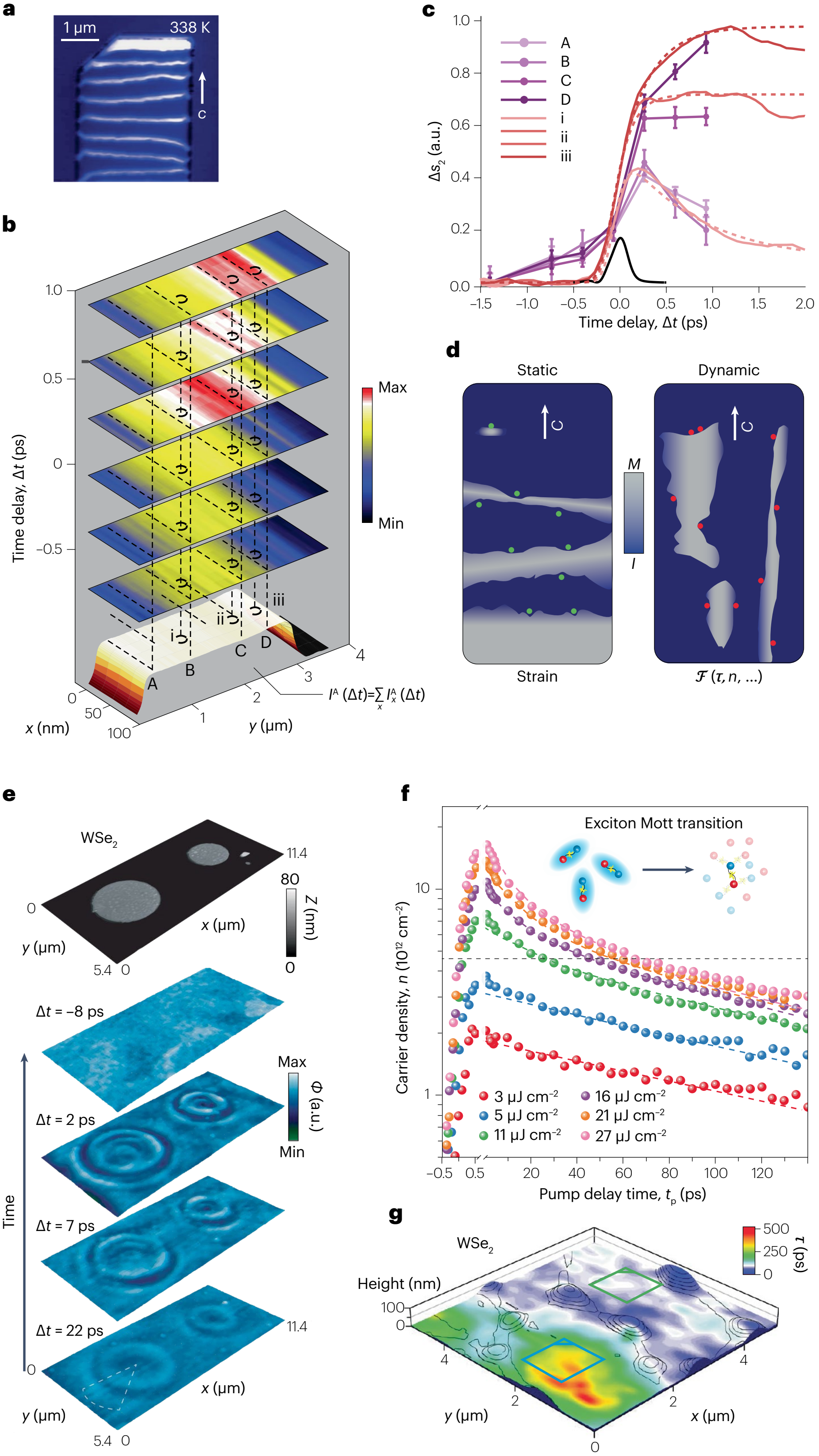

Fig. 5 | Ultrafast pump–probe nano-imaging of quantum phenomena in quantum materials. **a**, Static scattering-type scanning near-field optical microscopy image of a VO$_2$ nanobeam, demonstrating pinning and roughening of domain walls in the thermal transition, with the crystalline *c* axis noted. **b**, Subset of images showing spatial and temporal variations in insulator-to-metal dynamics, in which the last frame represents the summed intensity $I^A(\Delta t) = \sum_x I_x^A(\Delta t)$, obtained by summing the pump–probe signal $I_x^A(\Delta t)$ over all lateral positions *x* within region A at each pump–probe delay Δt. **c**, Near-field pump–probe time traces indicating from weak perturbation to full-phase transition at different sample locations within 100 nm at nominally same pump fluence, with A, B, C, D referring to the summed intensities $I^A(\Delta t)$ in part **b**, and I, II, III referring to the three sample locations in part **b**. **d**, Different mechanisms between thermal insulator–metal transition (strain inhomogeneity controlled) and photoinduced insulator–metal transition (defect and doping controlled). The photoinduced case depends on parameters such as pump–probe delay τ and excitation (carrier) density $n(\mathcal{F}(\tau, n, \dots))$. Colour scale denotes local state from metallic (M) to insulating (I). **e**, Time-resolved nano-imaging of hyperbolic polariton propagation across a WSe$_2$ flake. **f**, Pump-induced carrier response in WSe$_2$ showing carrier dynamics as a function of pump delay. **g**, Derived characteristic lifetime τ of photo-generated carriers in a WSe$_2$/WS$_2$ heterostructure. a.u., arbitrary unit. Panel **a** adapted with permission from ref. 139, American Chemical Society. Panels **b** and **c** adapted with permission from ref. 9, American Chemical Society. Panel **d** adapted from ref. 23, Springer Nature Limited. Panel **e** adapted with permission from ref. 131, AAAS. Panel **f** reprinted with permission from ref. 145, American Chemical Society. Panel **g** reprinted from ref. 56, Springer Nature Limited.

the ultrafast photoinduced IMT possibly associated with defect states near the bandgap of the insulating phase. Owing to the clear spatial separation of the domains (in contrast to random variations), one may expect that doping and stoichiometry possibly due to crystal growth, rather than random structural inhomogeneities, are responsible for the variation in the ultrafast dynamics[142,143]. Ultrafast s-SNOM thus allows distinguishing if and how the ultrafast photoinduced phase transition is in fact related to the thermally induced phase transition and if it is sensitive to the same defects and strain or other perturbations of the electronic or lattice structure.

The ultrafast photoinduced IMT s-SNOM work mentioned earlier[9] has been carried out in self-homodyne non-phase resolved detection and avoided background artefacts with only one VO_2 nanobeam in the pump focus and tip vicinity and benefited from the large excited-state contrast of the metallic versus insulating phases, as visualized by the ultrafast movie in Fig. 5d. The extension to transient pseudo-HPP nano-imaging[10], by sinusoidally modulating the reference arm of the interferometer to retrieve both amplitude and phase of the near-field signal, eliminated pump-induced background effects for extended sample surfaces. In a different approach, in s-SNOM with a cascaded lock-in detection scheme, demodulating higher tip-tapping harmonics, and combined with a visible pump (780 nm and 150 fs) and an MIR probe (31.5 THz and 60 fs) pulse train, the spatial correlation between the local switching and the ultrafast semiconducting response below the IMT threshold was investigated in a strained VO_2 nanobeam[4]. The VO_2 nanobeams have subsequently served again as a model system to develop and demonstrate HPP IR s-SNOM with pump modulation and sideband detection as a generalized approach for selective excited-state and background free contrast[7]. Adopting the concept of heterodyne probe detection as described earlier and as established in conventional far-field ultrafast spectroscopy, HPP provides phase-resolved and amplitude-resolved detection of ΔE_{NF}.

The phase-resolved detection in HPP s-SNOM then allowed to determine the actual photoinduced change in the dielectric permittivity at the MIR probe frequency to gain further insights into the differences in nature between the photoinduced and the thermal IMT. The observed differences in dynamic heterogeneities with the transient domain formation oriented preferentially parallel with respect to the crystallographic c axis suggest that the photoinduced IMT is highly sensitive with respect to local perturbations in electronic structure through defects or doping, compared with the static equilibrium domain texture primarily controlled by strain.

The high sensitivity of the photoinduced IMT to nanoscale inhomogeneities, compared with thermal IMT, was shown in another study. NIR-pump 8-µm probe s-SNOM showed distinct fluence thresholds to achieve the photoinduced IMT in VO_2 thin films, whereas temperature-dependent s-SNOM found indistinguishable transition temperatures for the same locations[132]. This highlights the role of nucleation sites for the ultrafast light-induced transition and allows for energy-efficient control of the material by keeping it close to the transition temperature.

These results highlight the ability, in particular, of combined static ground-state and excited-state dynamic s-SNOM imaging as an enabling tool to address heterogeneity, and the role of defects, doping and strain in quantum phase transitions towards the goal of understanding the origin and mechanisms underlying quantum phases in complex materials in general. With the extension to variable and cryogenic temperature, s-SNOM and MIR to THz spectroscopic access to the coupled quantum degrees of freedom of electrons, phonons and quasi-particle states, ultrafast s-SNOM could prove as a key enabling technique to resolve many elementary processes in quantum materials.

Exciton dynamics in 2D semiconductors. In 2D van der Waals materials, from carrier to thermal transport, the underlying elementary dynamics are distinctly different from corresponding bulk materials owing to finite size and quantum confinement effects. Furthermore, from transistor to opto-electronic devices based on 2D materials, these systems are inherently heterogeneous, which requires nanometre spatial resolution together with sub-picosecond temporal resolution to locally resolve carrier generation, decay and transport across the device. In this section, using WSe_2 as an example, we highlight how time-resolved THz s-SNOM can reveal how heterogeneity affects the excitonic and polaritonic dynamics and their coupling.

Ultrafast visible (650–700 nm) pump NIR (700–1,050 nm) probe s-SNOM was applied to bulk 60 nm WSe_2 on SiO_2 (ref. 124). In this study, the near-field tip launches exciton–polariton waves and observes their spatiotemporal evolution within the WSe_2 slab waveguide with ultraslow group velocity of $v_g \approx 0.017c$. Similarly, on monolayer WSe_2 and $MoSe_2$ with ultrafast s-SNOM with pump photon energies tuned near the A-exciton resonance (~1.66 eV), the exciton population dynamics could be resolved across the flake[144]. These examples demonstrate strong excitonic light–matter coupling and how ultrafast near-field s-SNOM can probe non-equilibrium polaritonic dynamics and slow-light phenomena in van der Waals semiconductors on the nanoscale.

Ultrafast THz s-SNOM applied to WSe_2/WS_2 heterostructures[56] revealed the ultrafast, spatially inhomogeneous dynamics of optically dark interlayer excitons, with the pump-induced signal being modulated by local variations in charge-transfer efficiency, that is, the efficiency with which photoexcited carriers tunnel between the two monolayers to form interlayer excitons. These variations were traced back to heterogeneities in the interface quality and local disorder, demonstrating how nanoscale structural variability directly impacts exciton formation and recombination dynamics[56]. In the study, the incident THz probe pulse is spectrally modified by the tip–sample near-field coupling, thus encoding the transient polarizability of intralayer versus interlayer excitons, probe by the tip-scattered THz emission of the ultrafast interlayer charge-transfer tunnelling generated by coherent THz transients.

At elevated pump fluence, high enough exciton densities can be achieved, giving rise to many-body effects. The associated Mott transition of hybrid excitons has been studied in twisted WSe_2 bilayers using visible-pump/THz-probe s-SNOM. Utilizing the pump–probe s-SNOM allows probing all photoexcited carriers, regardless of interband selection rules, which enables mapping the recombination dynamics across nanoscale inhomogeneities in WSe_2. From the pump-fluence-dependent variation in decay dynamics (Fig. 5f), a continuous crossover from monomolecular recombination of dark excitons to a radiative recombination of an unbounded electron–hole plasma could be distinguished, a behaviour in agreement with the Mott criterion from the exciton Bohr radius calculations (Fig. 5f,g). Furthermore, the nano-imaging revealed local variations in the density and transition width (Δn, the full-width at half-maximum of the density-driven Mott transition) due to disorder, twist-angle fluctuations and the interaction between the non-local dielectric environments[145], as visualized in Fig. 5f and the temporal evolution in Fig. 5g.

At exciton densities below the Mott threshold, Wannier–Mott exciton between valence and conduction bands was studied in monolayer

WSe_2 (ref. 131). Anisotropic Rydberg transitions in the excitonic series transiently drive the real part of the in-plane dielectric constant negative while keeping the out-of-plane component positive, satisfying the hyperbolicity condition ($\epsilon_{ab}\epsilon_c < 0$) for probe frequencies below 1,100 cm^{-1}. This condition then allows launching of hyperbolic exciton polaritons that couple to the excitonic transitions. The spatiotemporal evolution of the near-field response, measured with a visible pump (1.51 eV) and an MIR probe at ~1,200 cm^{-1}, revealed concentric features emanating from gold nano-antenna edges beneath the WSe_2, as a direct real-space signature of hyperbolic wavefronts (Fig. 5e). These rings persist for tens of picoseconds after photoexcitation, with their propagation direction and confinement dynamically tunable via pump fluence and through gating, thus demonstrating control of nanoscale light routing in a quantum material.

These examples demonstrate the utility of ultrafast THz to IR s-SNOM to probe the complex, spatially heterogeneous quantum dynamics in 2D semiconductors ranging from interlayer exciton formations (Fig. 5g) to density-driven Mott transitions, and photoinduced hyperbolic polaritons, with nanometre spatial and femtosecond temporal resolution. Many extensions of ultrafast s-SNOM – both in the quantum states that can be probed and in the range of 2D quantum materials and their heterostructures with associated nanoscale phenomena – can be envisioned, including studies of Moiré physics. This includes spectroscopy of the 1*s* to 2*p* intraexcitonic transition, which has so far primarily been studied in far-field spectroscopy[55,146] in 2D materials, and is accessible in s-SNOM as demonstrated in carbon nanotubes[8] and in monolayer CrSBr[147]. Ultrafast s-SNOM can thus provide a crucial experimental link between microscopic heterogeneity and macroscopic device function in 2D quantum materials.

Ultrafast nano-imaging of coupled structure and dynamics in photovoltaic perovskites

Perovskites are a class of materials with the general structure ABX_3 (Fig. 6b) with cations A and B and anion X[148]. In contrast to their all-inorganic counterparts, hybrid organic–inorganic metal halide perovskites have attracted much recent attention owing to several exceptional optoelectronic properties, enabling high-performance photodetectors, phototransistors, lasing and above all a new generation of solar cells. Despite chemical heterogeneities, grain boundaries and defects across multiple length scales, hybrid organic–inorganic metal halide perovskites exhibit exceptional carrier lifetimes and mobility usually only found in crystalline semiconductors. This, together with broad absorption across the visible to the NIR spectral range, leads to >25% photovoltaic conversion efficiency for single-layer devices and even higher for tandem solar cells in combination with silicon[149].

Stable over a wide range of cation substitutions, metal ions and halides allow for tuning crystal structure, bandgap or quantum confinement[150]. Specifically, lead halide perovskites (LHPs), with the A cation commonly a small organic molecule such as methylammonium or formamidinium (FA), the B cation lead and the halide iodine or bromine, have become established for most photovoltaic applications[151]. Improving photovoltaic efficiency as well as ambient and photostability requires understanding the physical nature of charge carriers, their transport and their interplay with the crystal lattice. The lack of scientific consensus on the elementary photophysical processes in LHPs appears to be due to a complex interplay of the roles of both static and dynamic disorder. Notably, the rapid crystallization in solution-based synthesis and film deposition results in a plethora of defects and stoichiometric and structural inhomogeneities on length scales from the atomic unit cell level, to nanoscale grains and up to micrometre-sized and millimetre-sized domains[152].

Perovskite films and devices have been investigated with almost every conceivable spectroscopic method in recent years, including time-resolved spectroscopy with excitation and probe frequencies throughout the electromagnetic spectrum[153–156]. A key result of these efforts is the discovery of the formation of large polarons in LHPs in response to the photoexcitation and on timescales up to 100 ps. In contrast to excitons, which are prone to non-radiative recombination notably at defects and grain boundaries and thus require highly crystalline materials, polaron formation in LHP-based solar cells has been perceived as responsible for protecting electrons and holes from recombination, making LHP solar cells less susceptible to defects. The ability to address the role and effect of spatial heterogeneity on polaron formation makes ultrafast s-SNOM a suitable method for such materials. It allows nano-imaging with IR spectral resolution and specificity the fast carrier dynamics in the photoexcited state on intragrain and intergrain scales, such as probing deep into the heterogeneity of the material in which conventional techniques only provide an ensemble-averaged effective dynamic response. In particular, spectrally resolving the transient evolution of phonon bands or vibrational modes as reporters of their local chemical and structural environment, for example, complements probing the electronic carrier response and can further provide insight into the associated coupled degrees of freedom of lattice disorder and crystallinity, polaron formation and carrier mobility.

Polaron dynamics in hybrid perovskites. In visible-pump IR-probe ultrafast s-SNOM imaging, spatial heterogeneities in charge carrier dynamics have been observed, attributed to grain-to-grain variations in polaron lifetimes, their stabilization energy and radius[7]. This work extends spatially averaged studies, which established the photoinduced deformation in the soft perovskite lattice observed with diffused X-ray scattering and attributed to polaron formation[157]. The evolving carrier dynamics can be described with rate equation modelling separating bimolecular recombination (scales with molecular number density as n^2) and Auger recombination (scales as n^3). Thus, at early delay times and high pump fluences ($n \approx n_0$), Auger recombination dominates, which results in a rapid initial population decay on a 10 ps timescale, followed at later times (>100 ps) by a linear recombination regime, associated with further increase in polaron size. Spatiotemporal maps of the carrier recombination dynamics reveal differences in the recombination rate between nanograins and to a lesser degree even within grains[6,84,85], as can be seen from the ultrafast nano-movie and lifetime map in Fig. 6a,d. Correlative far-field transient absorption spectroscopy with a non-colinear optical parametric amplifier shows a transient narrowing of the broad polaron response interpreted as an increase in correlation lengths and scattering times of photoinduced charge carriers[84]. This observation can be modelled, and as the polaron density decreases with time, the perovskite lattice supports larger polarons with a corresponding increase in polaron stabilization energy. However, the specific relationship between the characteristics of the perovskite lattice (such as structure, crystalline order and composition) and the associated polaron lifetime could not be established in this work.

Correlative nano-spectroscopy and transient vibrational solvatochromism. To gain further insight into the relationship of lattice structure and polaron formation, molecular vibrations can be used as a

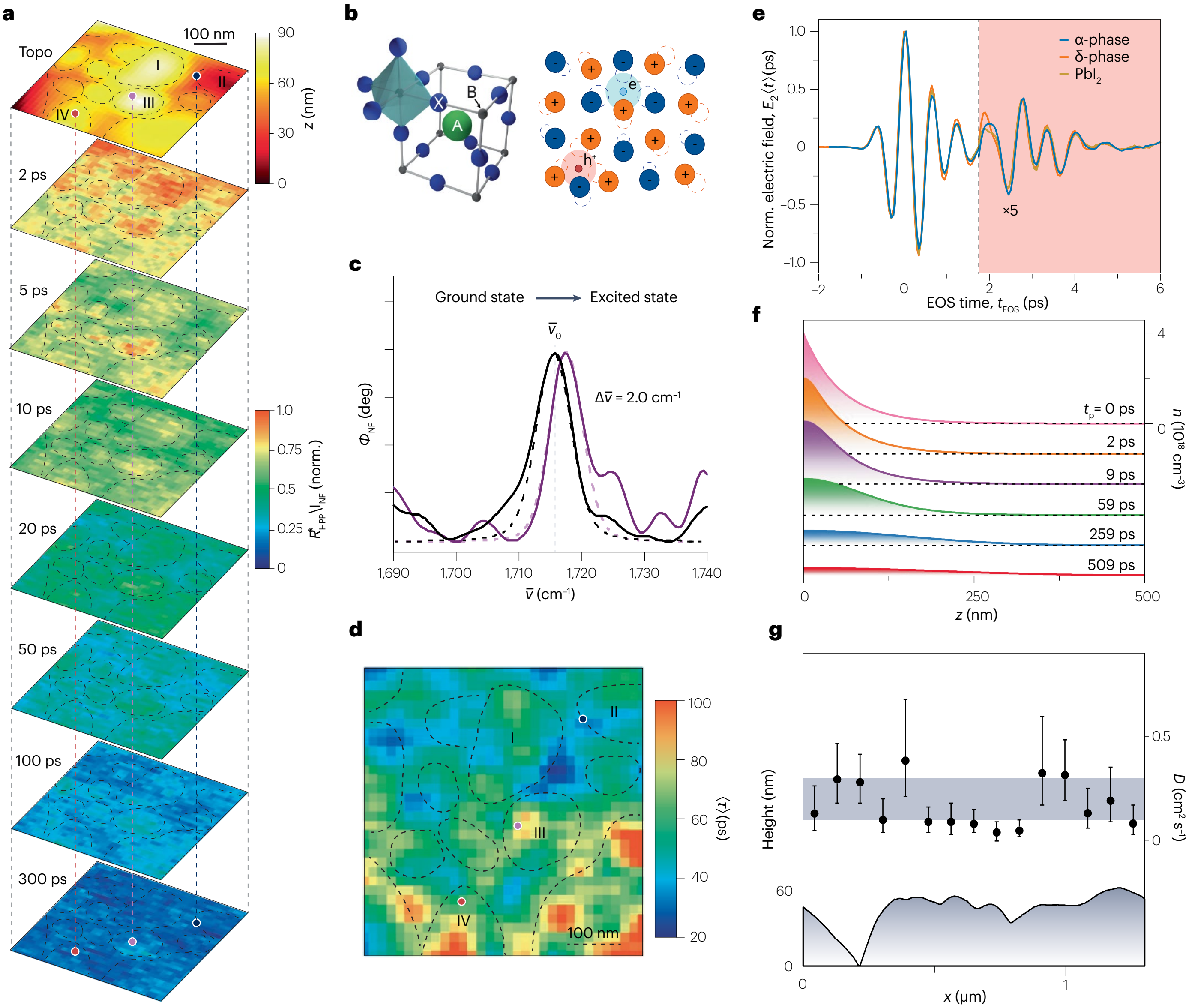

Fig. 6 | Perovskite ultrafast nanoscale dynamics. a, Topography and normalized two-phase heterodyne amplitude (R^*_{HPP}/I_{NF}) nano-images at increasing delay times, in which I_{NF} is the second harmonic ground-state image. The amplitude represents the spectral averaging over the very broad polaron spectrum in the infrared[178]. **b**, Unit cell of a general ABX_3 perovskite and schematic of polaron formation in the presence of photoexcited carriers. **c**, Unpumped ground-state formamidinium near-field phase Φ_{NF} (black) and the corresponding excited-state spectrum (purple) at the same sample location, demonstrating the transient vibrational solvatochromism (blue shift $\Delta\bar{\nu}$) of the formamidinium vibration, which serves as a local reporter of polaron dynamics. **d**, Decay map associated with the ultrafast nano-movie (part **a**). **e**, Scattered terahertz waveform amplitude E_2 (second harmonic) showing specific signatures for different perovskite phases, α and δ, as a function of electro-optical sampling (EOS) time. **f**, Vertical profile of the 3D time-dependent charge carrier density in the material determined by a diffusion model for a diffusion coefficient of $D = 0.3\ \mathrm{cm^2\,s^{-1}}$. **g**, Extracted diffusion coefficient, D, along a topographic cross-section. HPP, heterodyne pump–probe. Panels **a**–**d** reprinted from ref. 6, CC BY 4.0. Panels **e**–**g** reprinted from ref. 85, CC BY 4.0.

sensitive reporter of the local dielectric environment. For LHPs, the use of the FA antisymmetric C=N stretch vibration at ~1,720 cm^{-1} has been established in that regard responding sensitively to photoexcited lattice distortion in far-field pump MIR-probe spectroscopy[158,159], and as measure of compositional and structural variations in static nano-FTIR[160]. Here, ground-state continuous wave vibrational nano-spectroscopy with derived amplitude, A, linewidth, Γ, and peak position, ν_0, can serve as qualitative metrics of local composition, vibrational lifetime and structural disorder, respectively[160]. Correspondingly transient vibrational shifts in the excited state after photoexcitation can be interpreted as vibrational solvatochromism reflecting changes to the local dielectric environment and lattice polarizability associated with polaron formation building on corresponding demonstration of s-SNOM locally resolving vibrational Stark shifts at molecular hetero-interfaces.

In the application of LHP, a transient vibrational blue shift in pump–probe s-SNOM has been observed with pronounced spatial heterogeneity and interpreted to results from an increase in the local reaction field within the of polarons evolving after photoexcitation[7] (Fig. 6c). However, the relationship between perovskite lattice characteristics and the subpopulation of long-lived, stabilized large polarons has not been directly addressed[7], which will be discussed in the next section.

Ultrafast nano-imaging resolving cation–polaron coupling. Using vibrational resonances or phonons as local reporters relaxes the need for large spectral bandwidth that would be required for a full characterization of spectrally broad electronic Drude or interband excitations in general or the broad polaron response in particular. Instead, nano-spectroscopy can be performed with high spectral resolution focusing on vibrational modes of the organic cation in the MIR or lattice phonons in the THz spectral range that are sensitively coupled to transient pump-induced changes in lattice disorder, charge or dielectric environment associated with polaron formation[6].

In the application of ultrafast s-SNOM and the extension of the experiments described earlier, a pronounced spatial dependency of the transient blue shift[6,7] was observed and interpreted as a Stark-like vibrational solvatochromism of the FA resonance and its response to an increase in local reaction field induced by the evolving polaron (Fig. 6c). This change in reaction field is induced by an alteration of the polarizability of the perovskite lattice, which can be described by the Onsager continuum solvation model[161,162].

Importantly, correlations between both the transient Stark shift and the fraction of residual photoinduced carriers at >1 ns with the ground-state spectral heterogeneity have been observed, linking structure and dynamics. This suggested that an initially softer, more polarizable lattice supports stable polarons and longer-lived residual carriers[6]. This result perhaps even suggests that more structural disorder facilitates more effective polaron formation as explanation for high photovoltaic performance in perovskites of even low degree of crystallinity. Similarly, low-frequency lattice phonon modes in the THz spectral range can be examined to locally differentiate between crystallographic α-phase and δ-phase domains as well as non-stoichiometric PbI_2 islands by identifying differences in the dielectric function observed via pump–probe THz s-SNOM[85], as shown in Fig. 6e–g.

In corresponding pump–probe THz s-SNOM, information about carrier transport can be obtained. As discussed earlier, the tip near field penetrates into the sample to a depth of few tens of nanometres[163]. This information can be used for ultrafast nano-tomography modelling the diffusion of photoexcited carriers and relating them to a phase shift in the THz-EOS signal[85]. Diffusion into the sample plane occurs due to the inhomogeneous distribution of photoexcited carriers by the finite penetration depth of the pump beam, as seen in Fig. 6f. Therefore, the out-of-plane carrier mobility could be retrieved locally, which was surprisingly homogeneous in the light of the considerable inhomogeneity of the material (Fig. 6g), but in line with the high device performance in what seems unimpeded by grain boundaries.

The combination of results obtained by ultrafast s-SNOM to date already demonstrates the potential to reveal the coupled charge, molecular cation, lattice dynamics and carrier transport on the nanoscale. These studies show how different combinations of ultrafast nano-imaging of polaron recombination dynamics, transient vibrational nano-spectroscopy and THz nano-tomography can thus provide guidance for synthesis and device engineering to improve photophysical properties of perovskites or related materials.

Reproducibility and data deposition

Performing reproducible experiments, especially those that are interlaboratory comparable, has been a major challenge in near-field imaging as highly customized and home-built experimental set-ups can be difficult to operate. The challenge has been amplified by the strict requirements for stable scanning probe microscopy, special sample preparation and optical spectroscopy at low power. However, the field has recently progressed towards commercial scanning probe platforms with defined specifications (such as AFM tips from NanoAndMore and Bruker), standardized protocols for performing the experiments, a better understanding of near-field light–matter interaction and metrics for identification and discrimination from imaging artefacts. This has improved the reproducibility of s-SNOM in general.

For ultrafast s-SNOM, reproducibility relies on the stability of the AFM and the laser system, the tip and laser modulation, spatial overlap of pump and probe pulses, alignment of the interferometer or EOS detection and, most importantly, the relative position of illumination focus, tip and sample.

Scanning probe tips

Commercial availability of metal-coated (for example, gold, platinum–iridium and Pt–Si) cantilever tips has improved reproducibility in both imaging and spectroscopy between tips. These cantilever tips are batch-produced by standardized microfabrication protocols (such as photolithography and etching) and are broadly suitable for ultrafast s-SNOM from the visible to THz. However, commercially distributed tips are often shipped in polymer gel packs, in which tip contamination with molecules released from the gel, in particular polydimethylsiloxane, has been observed and is evident in corresponding IR vibrational signatures[164]. The tips can be cleaned via UV ozone or RCA processing to improve signal level and reproducibility[165]. A contaminated tip-based variation in spectral phase and amplitude can also be corrected through signal normalization against a spectrally flat reference sample, such as gold for continuous wave s-SNOM or germanium for ultrafast s-SNOM. In germanium referencing, the pump excitation of the interband transition in the visible and subsequent broadband Drude response in IR s-SNOM probing allows determination of T_0 and optimization of the pump–probe overlap. Depending on pump and probe wavelength, other semiconductors such as GaAs or InAs can also be used for this purpose. Referencing against germanium or gold also help to assess day-to-day variations in the spectral phase of the MIR or THz probe pulse and variations in the phase response between individual probe tips.

Signal variations between tips also arise from the differences in apex radius and other geometric details, damage and abrasion over extended measurement times and different tapping amplitudes for the same piezo driving parameters owing to variations in cantilever stiffness or mechanical coupling. Sharper probe tips generally lead to larger field enhancement at the tip apex affecting both pump excitation and probe signal, and larger tip-tapping amplitudes lead to a larger near-field volume probing deeper into the sample medium[1]. Maintaining a constant tapping amplitude over time and between tips ensures a constant probe depth, but accounting for shape variations is challenging and thus datasets measured with different tips on samples with fluence-dependent relaxation dynamics are difficult to compare.

Box 3 | Checklist for achieving artefact-free pump–probe nano-imaging

1. Choose sample substrates ideally transparent for both pump and probe beam wavelength, avoiding interband or defect-induced carrier (including by two-photon excitation) absorption and interaction with substrate phonons (such as CaF_2, BaF_2 or undoped silicon).
2. Be aware of hot carrier excitation in the tip, especially for strong and short wavelength pump, in which photocarrier excitation in the tip occurs. Although it can be normalized by referencing against gold or germanium substrates, attention needs to paid to differences in fluence dependence and carrier lifetime.
3. Clean tips to remove contamination from organics outgassed during gelpack storage.
4. Document and maintain a fixed tapping amplitude and atomic force microscopic set point, to ensure comparability of measurements between different tips.
5. Reacquire reference spectra every few hours and especially after each realignment or tip exchange to correct for drift in laser spectrum, spectral phase and tip-dependent carrier contributions.
6. Allow time for thermal equilibration after loading a sample (generally >1 h) and purge the sample chamber with dry air or N_2 to minimize sample drift. Correct for any remaining drift with an automated feature alignment algorithm to ensure pixels from different nano-movie frames are properly aligned.

Laser stability

Laser stability both in terms of noise and drift critically affects the quality of pump–probe nano-imaging. Visible-pump IR-probe studies are often performed with a NIR femtosecond pump laser that provides the pump and probe beams, with each generated in cascaded nonlinear processes of SHG, optical parametric oscillators/OPAs, DFG or optical rectification. The intensity, pointing stability, pulse duration and exact transient (such as chirp or higher order dispersion), phase front and beam profile are sensitive to details of the alignment and prone to thermal drifts especially during long measurement times. Overall, with active temperature control of the laboratory and experimental set-up, laser and set-up enclosure with nitrogen or dry air purge, achieving long-term stability of ~1% for the power of both pump and probe beams at the tip can readily be achieved.

Lock-in noise suppression

For s-SNOM lock-in signal detection, the only relevant noise contributions are those that overlap in frequency with the pump–probe sidebands (typically tens to hundreds of kilohertz, depending mainly on ω_t) used for demodulation. Although the laser exhibits broadband intensity noise and slow drift, only the spectral components within the lock-in bandwidth ($\lesssim$1 kHz, depending on the lock-in time constant) at these harmonics affect the measured SNR. High-frequency noise far from $n\omega_t$ and low-frequency drift are strongly suppressed by the demodulation scheme. For this reason, comparing the relative intensity noise of the laser at the specific demodulation frequencies within the lock-in detection bandwidth provides a more accurate estimate of its impact on s-SNOM SNR.

Sample drift and degradation

AFM tip and sample drift, tip-induced or laser-induced sample damage, and other sample degradations are other important considerations in particular for ultrafast s-SNOM. With measurement times of even single time traces exceeding minutes, especially for low pump excitation, the sample drift of <1 nm min^{-1} is desirable, or in general less than the tip radius during the timescale of a local probe measurement. Furthermore, for full spatiotemporal imaging, at least for each frame, sample drifts must be small enough to still allow for image stacking for subsequent time frames.

The drift can be reduced by temperature control of the laboratory environment and especially the AFM and sample area, by allowing the system to thermally equilibrate after loading new tips or samples and by also purging the sample chamber with nitrogen or dry air of constant temperature. In general, different drift correction procedures can be applied, as large spatiotemporal or spatio-spectral datasets require measurement times of hours, with the goal of being repeatable for different pump fluences. The repeated AFM topography images acquired in parallel with the s-SNOM optical data inform on the amount of drift and can be used for drift correction, either manually or through automated feature alignment, to align and even remove distortion in different frames of ultrafast nano-movies. Measuring transients, either just local probe, transects or arrays, periodic re-scanning of the sample topography and re-positioning of the tip will be important. Proper protocol is to also frequently assess for emergent tip or sample damage, mechanical or photoinduced. Mitigating sample degradation is highly sample-dependent and may require limiting pump power, operate under dry nitrogen purge, extend measurement time or reassess the overall measurement protocol. Box 3 summarizes key considerations for artefact-free pump–probe nano-imaging.

Limitations and optimizations

Beam fluence and localization

In ultrafast far-field microscopy, it is desirable to probe a reduced spatial region within a large pump area using a probe focus size smaller than that of the pump. The Gaussian, or related, profile of the pump beam ensures that an approximately homogeneous excitation volume is sampled by the probe. Similarly, in ultrafast s-SNOM, if the pump field were as strongly localized by the metallic tip as the probe, the transient near-field signal $\Delta\tilde{E}_{NF}$ would be an average across an inhomogeneous excitation profile. With the field localization determined by tip geometry affecting pump and probe equally to first order, unlike far-field microscopy, this problem can only indirectly be mitigated.

To minimize this effect, although the probe beam is preferably polarized parallel with respect to the tip axis, the pump beam can be polarized perpendicular. Polarization of the pump perpendicular to the tip axis and choosing a tip material such as Pt–Ir, which does not support surface plasmon polaritons, limits the field enhancement of the pump to a factor of 2–3 (refs. 29,58,166).

Despite the still inhomogeneous pump fluence within the probe region, the excited-state dynamics remain insensitive to this variation in the low-density regime. At these fluence levels, where carrier–carrier interactions are negligible and recombination is dominated by monomolecular processes, even a twofold to threefold change in pump fluence produces essentially the same temporal dynamics. As a result, the measured signal effectively represents an intensity-weighted average over this narrow fluence distribution, with most of the probe response arising from the locally enhanced, highest-fluence region. However, for a detailed fluence-dependent interpretation and for strong pump excitation, careful modelling of the near-field distribution and enhancement or additional far-field control experiments are necessary.

Balancing SNR with acquisition time

The time for data acquisition in ultrafast s-SNOM is a balance between desirable SNR, tip and sample degradation, and tip, sample, optical and laser drift. The time required to obtain a single interferogram, a full time trace or a complete image is set by the lock-in integration time, image size and pixel resolution and desirable spectral and temporal resolution. Shorter lock-in integration times allow for faster measurements, but at the expense of SNR. Longer integration times improve SNR, although they increase the risk that sample drift or degradation will occur before the measurement is complete.

Another key factor in acquisition speed is the detection scheme itself. FTIR and EOS provide spectral resolution of $\Delta\tilde{E}_{\mathrm{NF}}(\omega)$, but require scanning hundreds to thousands of delay positions at each image pixel depending on desired spectral resolution and signal quality. By contrast, two-phase or three-phase measurements require measuring only two or three delay positions at each image pixel, greatly reducing measurement time, but sacrificing spectral resolution.

The choice between these methods depends on whether the transient response contains spectral information or whether it can be treated as independent of frequency within the probe bandwidth. Data acquisition speed is not a fixed instrument property but a tunable parameter space defined by integration time, detection scheme and sample stability. Strategies to improve acquisition speed include: minimizing the number of temporal or spatial points by adaptive or undersampling techniques[167–169], using faster lock-in electronics or multichannel detection schemes and applying single-frequency detection schemes for spatiotemporal imaging and spectral detection schemes for spectral-temporal imaging to gain insights into the spatio-spectral-temporal parameter space without fully exploring all dimensions.

Choice of laser repetition rate

The choice of repetition rate for pulsed laser sources must take into consideration the tapping frequency of the AFM tip and the relaxation time of the pump excitation. By the Nyquist theorem, signal modulations faster than half the laser repetition rate cannot be efficiently detected. Standard cantilever AFM tips often resonate at 200–300 kHz, with higher harmonics necessary for near-field isolation than approaching megahertz frequencies. Sub-megahertz laser repetition rates can be used, but should be paired with low-frequency AFM tips, with resonances in the 50 kHz range available. By contrast, thermal conductivity of the sample (depending on size, dimension and substrate) leads to continuous wave heating, which can be reduced by limiting the pump and probe power accordingly. Additionally, the repetition rate should not be higher than the inverse of the lifetime of the excitation, for example, the pump-induced excitation should be fully relaxed and re-equilibrated before arrival of the next pump pulse. High-repetition rate lasers in the 80–100 MHz are thus only applicable for studies of short-lived excitations, with strong absorption or weak pump perturbation. Therefore, the choice of laser repetition rate and cantilever resonance must be balanced to avoid photodamage and thermal load but with adequate pump fluence to observe the desired excitation dynamics, while also not undersampling the tip-tapping modulation (Table 3). Several additional considerations, which may be relevant depending on the sample of interest, are described in Box 4.

Outlook

Ultrafast s-SNOM and related nonlinear, coherent and dynamic nano-imaging has been one of the most vibrant frontiers in imaging science. On the basis of work over the past decade, with growing contributions in particular over the past few years, ultrafast s-SNOM has transitioned from method development to material focused. Although currently implemented and pursued by still few groups worldwide, with the commercial availability of ultrafast laser systems and s-SNOM platforms from different suppliers that are well suited for ultrafast nano-imaging, we expect ultrafast s-SNOM to be adopted by progressively more research groups for a wide range of applications. Essentially, all conventional ultrafast spectroscopy of solids and interfaces, where heterogeneity has limited interpretation and modelling, could be extended to the nanoscale with ultrafast s-SNOM. Furthermore, with the evolving standardization of instrumentation and measurement protocols, it could also attract industrial applications for nano-metrology, quality control and manufacturing.

A natural approach to further extend the application space, as with any spectroscopic technique, is the extension of the spectral range. Experiments so far have largely focused on selected visible and NIR pump frequencies with probing in the THz, MIR and NIR. Large spectral gaps are yet underused in the far-IR (300–800 cm^{-1}) with characteristic molecular fingerprint vibrations and phonon resonances, as well as parts of the NIR (2–4 µm) with hydrogen stretch vibrations and plasmonic resonances. Ultrafast s-SNOM is compatible with any pump and probe frequencies across the UV to THz. For example, the far-IR region between 300 cm^{-1} and 700 cm^{-1} has been largely underexplored; however, it represents important phonon and molecular vibrations. Another challenge for future development is the implementation of degenerate pump–probe nano-imaging, such as IR-pump and IR-probe spectroscopy. In contrast to far-field spectroscopy with the ability to separate pump and probe light spatially, degenerate pump–probe s-SNOM requires further probe modulation to isolate the signal. This will allow selective excitation and probing the coupling

Box 4 | Failure modes and safety thresholds

Sample heating in pump–probe scattering-type scanning near-field optical microscopy arises primarily from the pump pulse. In some materials, high fluences can drive the system into saturation or even induce photodamage, as is well documented in perovskites. These effects restrict the accessible excitation densities and thereby limit the usable dynamic range of the experiment. Systems that exhibit only small pump-induced changes in polarizability or absorption may also fall below the detectable near-field modulation depth, reducing the effective signal and potentially degrading the achievable temporal resolution. Furthermore, when the relaxation times of long-lived excited states exceed the laser repetition period, the material can be re-excited before fully returning to equilibrium. This leads to convolution of the measured relaxation dynamics and may produce cumulative carrier population buildup that complicates quantitative interpretation. A final limitation arises from tip-induced or tip-mediated artefacts, such as far-field scattering backgrounds or local perturbations of the evanescent electromagnetic field, which can influence or convolve the intrinsic spectral response of the material.

Glossary

Abbe diffraction limit
The fundamental resolution limit of a diffraction-limited optical system, which is given by approximately $\lambda/2$.

Drude response
The characteristic frequency-dependent behaviour of a material's conductivity or dielectric function, as described by the Drude model of free electrons with a finite scattering time.

Evanescent field
Non-propagating electromagnetic field that decays exponentially away from the interface where it is generated.

Finite size effects
Changes in electronic, magnetic or optical properties that occur in small or low-dimensional systems (on the order of the mean-free path of electrons, for example) owing to quantum confinement and the increased role of surfaces or boundaries compared with the bulk.

Free-carrier dynamics
Behaviour of mobile charge carriers (electrons and holes) after excitation, including generation, transport, scattering and recombination.

Intersubband transitions
Optical transitions of charge carriers between quantized energy levels, called subbands, within the same conduction or valance band of a nanostructure such as a nanowire or quantum well.

Microscopic polarizability
Dipole moment induced in an individual atom, molecule or nanostructure by an applied field, reflecting the intrinsic electronic or vibrational response.

Mott transition
A metal–insulator transition driven by strong electron–electron interactions, in which a material that would be metallic according to band theory becomes insulating owing to Coulomb repulsion localizing the charge carriers.

Tip–sample-coupled polarization
Modelling of near-field optics in which the probe tip is represented as a point dipole (point-dipole model) or extended dipole (finite-dipole model) to describe sample coupling.

Tip-tapping frequency
Mechanical oscillation frequency of an atomic force microscopy cantilever operated in tapping mode.

Ultrafast nano-movies
Time-resolved imaging techniques that capture dynamics with high temporal and spatial resolution by combining ultrashort pump–probe schemes with the nanoscale spatial resolution of an atomic force microscope to visualize electronic, structural and vibrational evolution frame by frame.

and dynamics of phonons, magnons, low-energy quantum phases as well as collective and intermolecular vibrations.

In addition, ultrafast s-SNOM is able to not only probe and image ultrafast dynamics locally but also observe transport and diffusion of charge carriers, phonons and polaritons across interfaces and in low-dimensional materials[170]. To resolve the elementary steps how nanoscale heterogeneities affect transport phenomena would be of value to improve device performance of solar cells, microelectronics, optoelectronics and batteries.

Pump–probe s-SNOM can also be extended to coherent and nonlinear spectroscopy with different wave-mixing schemes based on four-wave mixing, coherent anti-Stokes Raman spectroscopy or stimulated Raman scattering, with the prospect towards coherent multidimensional spectroscopy. Their implementation is associated with additional experimental challenges requiring additional pulse and modulation schemes for isolation of increasingly small signals.

To date, ultrafast s-SNOM has been mostly applied under ambient conditions at room temperature. The implementation under vacuum and cryogenic conditions[30,60,171,172] will be essential for probing quantum phases in correlated electron systems, topological and low-dimensional materials and their heterostructures, superconductors or other low-temperature phenomena. Temperature-dependent ultrafast spectroscopy on the nanoscale is expected to distinguish intrinsic and extrinsic effects of heterogeneity and thus systematically disentangle the complex coupled dynamics underlying many complex quantum systems and their emergent properties. The increased complexity in instrumentation and challenges imposed by constrained free space optical access and vibration control are expected to be compensated by the scientific advances that can be expected. This includes the extension of s-SNOM to study magnetic phenomena in quantum materials with large external magnetic fields posing large experimental challenges; however, recent proof-of-concept demonstrations have made progress towards performing s-SNOM in up to 5 T magnetic fields[173]. Additionally, the ultrahigh vacuum environment enables maintaining clean surfaces, minimizes photodamage and generally improves tip–sample stability, which are also beneficial as s-SNOM approaches atomic resolution and few molecules spectroscopic sensitivity.

To perform experiments under in situ and in operando conditions, and also under controlled external stimuli such as electric fields, strain or chemical environments is still challenging and would require additional developments. Extension of s-SNOM to in-liquid environments will also expand the range of material classes accessible to ultrafast s-SNOM, particularly for in situ spectroscopy of biological systems[174] and eventually towards biomedical applications[175]. Ultrafast s-SNOM may also contribute to electrochemistry, catalysis and biophysics when applied to samples in solvents and electrolytes[176].

In parallel, theoretical frameworks and data analysis must evolve in tandem with experimental progress to pave the way for ultrafast s-SNOM becoming an established technique used by research groups beyond those versed in laser science in chemistry, biology or materials science. Incorporating machine learning may offer novel strategies for extracting dynamic spectral signatures from high-dimensional datasets with poor SNR, particularly where traditional fitting fails[88,177]. With its broad interdisciplinary appeal and continuing advances in such experimental and data analysis techniques, ultrafast s-SNOM is uniquely positioned to make nanoscale ultrafast dynamics broadly accessible across diverse material classes, informing device design and optimization.

References

1. Eisele, M. et al. Ultrafast multi-terahertz nano-spectroscopy with sub-cycle temporal resolution. *Nat. Photon.* **8**, 841–845 (2014).
   An early demonstration of NIR pump/THz probe with EOS on InAs nanowires to reconstruct a time-dependent dielectric function of the material.
2. Wagner, M. et al. Ultrafast dynamics of surface plasmons in InAs by time-resolved infrared nanospectroscopy. *Nano Lett.* **14**, 4529–4534 (2014).
3. Wagner, M. et al. Ultrafast and nanoscale plasmonic phenomena in exfoliated graphene revealed by infrared pump–probe nanoscopy. *Nano Lett.* **14**, 894–900 (2014).
4. Huber, M. A. et al. Ultrafast mid-infrared nanoscopy of strained vanadium dioxide nanobeams. *Nano Lett.* **16**, 1421–1427 (2016).
5. Jiang, T., Kravtsov, V., Tokman, M., Belyanin, A. & Raschke, M. B. Ultrafast coherent nonlinear nanooptics and nanoimaging of graphene. *Nat. Nanotechnol.* **14**, 838–843 (2019).

6. Wilcken, R. et al. Correlated nanoimaging of structure and dynamics of cation–polaron coupling in hybrid perovskites. *Sci. Adv.* https://doi.org/10.1126/sciadv.ads3706 (2025).
An example showcasing the advantages of combining both ground-state IR s-SNOM and ultrafast nano-imaging to correlate the influence of structural and chemical factors on excited-state dynamics and thus photovoltaic performance in lead-halide perovskites.
7. Nishida, J. et al. Ultrafast infrared nano-imaging of far-from-equilibrium carrier and vibrational dynamics. *Nat. Commun.* **13**, 1083 (2022).
An early work demonstrating developments in ultrafast nano-imaging utilizing a novel sideband demodulation and two-phase heterodyne technique, with key applications to materials systems.
8. Nishida, J., Otsuka, K., Minato, T., Kato, Y. K. & Kumagai, T. Ultrafast infrared nano-imaging of local electron–hole dynamics in CVD-grown single-walled carbon nanotubes. *Sci. Adv.* **11**, eaadv9584 (2025).
An application of ultrafast IR s-SNOM on carbon nanotubes, demonstrating nanoscale spatial heterogeneity in carrier dynamics resulting from variations in local strain and environmental conditions, despite homogeneous topographic features.
9. Dönges, S. A. et al. Ultrafast nanoimaging of the photoinduced phase transition dynamics in VO_2. *Nano Lett.* **16**, 302–307 (2016).
NIR-pump/MIR-probe experiment on VO_2 microcrystals, demonstrating excited-state contrast obtained without pump modulation, which is possible as the signal is dominated by the response to the pump.
10. Sternbach, A. J. et al. Artifact free time resolved near-field spectroscopy. *Opt. Express* **25**, 28589 (2017).
A theoretical description of a transient pseudo-heterodyne detection method (Tr-pHD) utilizing digital boxcar detection to obtain the amplitude and phase of the near-field signal, including an application to resolving the IMT in vanadium dioxide.
11. Sternbach, A. J. et al. Femtosecond exciton dynamics in WSe_2 optical waveguides. *Nat. Commun.* **11**, 3567 (2020).
12. Mukamel, S. *Oxford Series in Optical and Imaging Sciences* Vol. 6 (Oxford Univ. Press, 1995).
13. Fayer, M. D. (ed.) *Ultrafast Infrared and Raman Spectroscopy*. *Practical Spectroscopy* (CRC Press, 2001).
14. Asplund, M. C., Zanni, M. T. & Hochstrasser, R. M. Two-dimensional infrared spectroscopy of peptides by phase-controlled femtosecond vibrational photon echoes. *Proc. Natl Acad. Sci. USA* **97**, 8219–8224 (2000).
15. Nibbering, E. T. J., Fidder, H. & Pines, E. Ultrafast chemistry: using time-resolved vibrational spectroscopy for interrogation of structural dynamics. *Annu. Rev. Phys. Chem.* **56**, 337–367 (2005).
16. Zewail, A. H. *Femtochemistry*: *Ultrafast Dynamics of the Chemical Bond*. *World Scientific Series in 20th Century Chemistry* (World Scientific, 1994).
17. Cavalleri, A. et al. Femtosecond structural dynamics in VO_2 during an ultrafast solid-solid phase transition. *Phys. Rev. Lett.* **87**, 237401 (2001).
18. Lindberg, M. & Koch, S. Theory of coherent transients in semiconductor pump–probe spectroscopy. *J. Opt. Soc. Am. B* **5**, 139–146 (1988).
19. Cundiff, S. T. & Mukamel, S. Optical multidimensional coherent spectroscopy. *Phys. Today* **66**, 44–49 (2013).
20. Brabec, T. & Krausz, F. Intense few-cycle laser fields: frontiers of nonlinear optics. *Rev. Mod. Phys.* **72**, 545 (2000).
21. Dombi, P. et al. Strong-field nano-optics. *Rev. Mod. Phys.* **92**, 025003 (2020).
22. Gross, N. et al. Progress and prospects in optical ultrafast microscopy in the visible spectral region: transient absorption and two-dimensional microscopy. *J. Phys. Chem. C* **127**, 14557–14586 (2023).
23. O'Callahan, B. T. et al. Inhomogeneity of the ultrafast insulator-to-metal transition dynamics of VO_2. *Nat. Commun.* **6**, 6841 (2015).
24. Zhu, T., Snaider, J. M., Yuan, L. & Huang, L. Ultrafast dynamic microscopy of carrier and exciton transport. *Annu. Rev. Phys. Chem.* **70**, 219–244 (2019).
25. Delor, M., Weaver, H. L., Yu, Q. & Ginsberg, N. S. Imaging material functionality through three-dimensional nanoscale tracking of energy flow. *Nat. Mater.* **19**, 56–62 (2020).
26. Ginsberg, N. S., Hsieh, C.-L., Kukura, P., Piliarik, M. & Sandoghdar, V. Interferometric scattering microscopy. *Nat. Rev. Methods Primers* **5**, 23 (2025).
27. Hillenbrand, R., Abate, Y., Liu, M., Chen, X. & Basov, D. N. Visible-to-THz near-field nanoscopy. *Nat. Rev. Mater.* **10**, 285–310 (2025).
A key review paper that provides a comprehensive and up-to-date summary of IR s-SNOM, including an overview of fundamental mechanisms driving its operation, modelling and detection techniques and applications to material systems.
28. Knoll, B. & Keilmann, F. Near-field probing of vibrational absorption for chemical microscopy. *Nature* **399**, 134–137 (1999).
29. Atkin, J. M., Berweger, S., Jones, A. C. & Raschke, M. B. Nano-optical imaging and spectroscopy of order, phases, and domains in complex solids. *Adv. Phys.* **61**, 745–842 (2012).
A comprehensive review that provides a rigorous theoretical framework of IR s-SNOM and complimentary techniques, including optical selection rules and interferometric detection techniques.
30. Chen, X. et al. Modern scattering-type scanning near-field optical microscopy for advanced material research. *Adv. Mater.* **31**, 1804774 (2019).
31. Wessel, J. Surface-enhanced optical microscopy. *J. Opt. Soc. Am. B* **2**, 1538–1540 (1985).
32. Fischer, U. C. & Pohl, D. W. Observation of single-particle plasmons by near-field optical microscopy. *Phys. Rev. Lett.* **62**, 458–461 (1989).
33. Specht, M. et al. Scanning plasmon near-field microscope. *Phys. Rev. Lett.* **68**, 476–479 (1992).
34. Novotny, L. & Stranick, S. J. Near-field optical microscopy and spectroscopy with pointed probes. *Annu. Rev. Phys. Chem.* **57**, 303–331 (1992).
35. Raschke, M. B. & Lienau, C. Apertureless near-field optical microscopy: tip–sample coupling in elastic light scattering. *Appl. Phys. Lett.* **83**, 508–511 (2003).
36. Cvitkovic, A., Ocelic, N. & Hillenbrand, R. Analytical model for quantitative prediction of material contrasts in scattering-type near-field optical microscopy. *Opt. Express* **15**, 8550–8565 (2007).
37. Hillenbrand, R., Taubner, T. & Keilmann, F. Phonon-enhanced light–matter interaction at the nanometre scale. *Nature* **418**, 159–162 (2002).
38. Denk, W. & Pohl, D. W. Near-field optics: microscopy with nanometer-size fields. *J. Vac. Sci. Technol. B* **9**, 510–513 (1991).
39. Sánchez, E. J., Novotny, L. & Xie, X. S. Near-field fluorescence microscopy based on two-photon excitation with metal tips. *Phys. Rev. Lett.* **82**, 414–417 (1999).
40. Shiotari, A. et al. Scattering near-field optical microscopy at 1-nm resolution using ultralow tip oscillation amplitudes. *Sci. Adv.* **11**, eadu1415 (2025).
41. Damodaran, R. et al. Phase coexistence and electric-field control of toroidal order in oxide superlattices. *Nat. Mater.* **16**, 13–19 (2017).
42. Pollard, B., Maia, F. C. B., Raschke, M. B. & Freitas, R. O. Infrared vibrational nanospectroscopy by self-referenced interferometry. *Nano Lett.* **16**, 551–557 (2016).
43. Park, K.-D. & Raschke, M. B. Polarization control with plasmonic antenna tips: a universal approach to optical nanocrystallography and vector-field imaging. *Nano Lett.* **18**, 2912–2917 (2018).
44. Khatib, O., Bechtel, H. A., Martin, M. C., Raschke, M. B. & Carr, G. L. Far infrared synchrotron near-field nano-imaging and -spectroscopy. *ACS Photonics* https://doi.org/10.1021/acsphotonics.8b00565 (2018).
45. Olmon, R. L., Krenz, P. M., Jones, A. C., Boreman, G. D. & Raschke, M. B. Near-field imaging of optical antenna modes in the mid-infrared. *Opt. Express* **16**, 20295–20305 (2008).
46. Park, K.-D., Raschke, M. B., Atkin, J. M., Lee, Y. H. & Jeong, M. S. Probing bilayer grain boundaries in large-area graphene with tip-enhanced Raman spectroscopy. *Adv. Mater.* **29**, 16361–16367 (2017).
47. Park, K.-D. et al. Hybrid tip-enhanced nanospectroscopy and nanoimaging of monolayer WSe_2 with local strain control. *Nano Lett.* **16**, 2621–2627 (2016).
48. Bylinkin, A. et al. Real-space observation of vibrational strong coupling between propagating phonon polaritons and organic molecules. *Nat. Photon.* **15**, 197–202 (2021).
49. Huber, M. A. et al. Femtosecond photo-switching of interface polaritons in black phosphorus heterostructures. *Nat. Nanotechnol.* **12**, 207–211 (2017).
50. Ni, G. et al. Ultrafast optical switching of infrared plasmon polaritons in high-mobility graphene. *Nat. Photon.* **10**, 244–247 (2016).
51. Roth, R. M. et al. Resonant-plasmon field enhancement from asymmetrically illuminated conical metallic-probe tips. *Opt. Express* **14**, 221–231 (2006).
52. Shim, S.-H. & Zanni, M. T. How to turn your pump–probe instrument into a multidimensional spectrometer: 2D IR and Vis spectroscopies via pulse shaping. *Phys. Chem. Chem. Phys.* **11**, 748–761 (2009).
53. Munoz, M. F. et al. Fast phase cycling in non-collinear optical two-dimensional coherent spectroscopy. *Opt. Lett.* **45**, 5852–5855 (2020).
54. Xu, X. G. & Raschke, M. B. Near-field infrared vibrational dynamics and tip-enhanced decoherence. *Nano Lett.* **14**, 4613–4617 (2014).
55. Steinleitner, P. et al. Direct observation of ultrafast exciton formation in a monolayer of WSe_2. *Nano Lett.* **17**, 1455–1460 (2017).
56. Plankl, M. et al. Subcycle contact-free nanoscopy of ultrafast interlayer transport in atomically thin heterostructures. *Nat. Photon.* **15**, 594–600 (2021).
An application of s-SNOM with visible pump and THz probe with EOS to sample nanoscale spatial heterogeneities in both interlayer charge-transfer efficiency and carrier lifetime in transition metal dichalcogenide heterostructures.
57. Hillenbrand, R. & Keilmann, F. Complex optical constants on a subwavelength scale. *Phys. Rev. Lett.* **85**, 3029 (2000).
58. Novotny, L. The history of near-field optics. *Prog. Opt.* **50**, 137–184 (2007).
59. Trebino, R. et al. Measuring ultrashort laser pulses in the time-frequency domain using frequency-resolved optical gating. *Rev. Sci. Instrum.* **68**, 3277–3295 (1997).
60. Yang, H. U., Hebestreit, E., Josberger, E. E. & Raschke, M. B. A cryogenic scattering-type scanning near-field optical microscope. *Rev. Sci. Instrum.* https://doi.org/10.1063/1.4789428 (2013).
61. Jing, R. et al. Bolometric superconducting optical nanoscopy (BOSON). *Phys. Rev. X* **15**, 031027 (2025).
This work details the development of bolometric detection methods as an extension of SNOM to create bolometric superconducting optical nanoscopy, which improves sensitivity in certain applications in which sample and tip heating may limit SNOM performance.
62. Boyd, R. W. *Nonlinear Optics* (Elsevier Science & Technology, 2020).
63. Nishida, J. et al. Sub-tip-radius near-field interactions in nano-FTIR vibrational spectroscopy on single proteins. *Nano Lett.* **24**, 836–843 (2024).
64. Amarie, S., Ganz, T. & Keilmann, F. Mid-infrared near-field spectroscopy. *Opt. Express* **17**, 21794–21801 (2009).
65. Huber, A. J., Kazantsev, D., Keilmann, F., Wittborn, J. & Hillenbrand, R. Simultaneous infrared material recognition and conductivity mapping by nanoscale near-field microscopy. *Adv. Mater.* **19**, 2209–2212 (2007).
66. Taubner, T., Hillenbrand, R. & Keilmann, F. Nanoscale polymer recognition by spectral signature in scattering infrared near-field microscopy. *Appl. Phys. Lett.* **85**, 5064–5066 (2004).

67. Ebel, A., Dreher, W. & Leibfritz, D. Effects of zero-filling and apodization on spectral integrals in discrete Fourier-transform spectroscopy of noisy data. *J. Magn. Reson.* **182**, 330–338 (2006).
68. Zhao, Z. et al. Applications of ultrafast nano-spectroscopy and nano-imaging with tip-based microscopy. *eLight* https://doi.org/10.1186/s43593-024-00079-1 (2025).
69. Labouesse, S., Johnson, S. C., Bechtel, H. A., Raschke, M. B. & Piestun, R. Smart scattering scanning near-field optical microscopy. *ACS Photon.* **7**, 3346–3352 (2020).
70. Johnson, S. C. et al. Infrared nanospectroscopic imaging in the rotating frame. *Optica* **6**, 424–429 (2019).
71. Ocelic, N., Huber, A. & Hillenbrand, R. Pseudoheterodyne detection for background-free near-field spectroscopy. *Appl. Phys. Lett.* **89**, 101124 (2006).
72. Fan, Y. et al. Direct measurement of the real strength of near-field electric field. *Appl. Phys. Lett.* **125**, 211105 (2024).
73. Aizpurua, J., Taubner, T., García de Abajo, F. J., Brehm, M. & Hillenbrand, R. Substrate-enhanced infrared near-field spectroscopy. *Opt. Express* **16**, 1529–1545 (2008).
74. Chen, X. et al. Rapid simulations of hyperspectral near-field images of three-dimensional heterogeneous surfaces — part II. *Opt. Express* **30**, 11228–11242 (2022).
75. Babicheva, V. E., Gamage, S., Stockman, M. I. & Abate, Y. Near-field edge fringes at sharp material boundaries. *Opt. Express* **25**, 23935–23944 (2017).
76. McArdle, P., Lahneman, D., Biswas, A., Keilmann, F. & Qazilbash, M. Near-field infrared nanospectroscopy of surface phonon-polariton resonances. *Phys. Rev. Res.* **2**, 023272 (2020).
77. Huth, F. et al. Nano-FTIR absorption spectroscopy of molecular fingerprints at 20 nm spatial resolution. *Nano Lett.* **12**, 3973–3978 (2012).
78. Németh, G., Bechtel, H. & Borondics, F. Origins and consequences of asymmetric nano-FTIR interferograms. *Opt. Express* **32**, 15280–15294 (2024).
79. Herrera, E. Z. et al. Interpreting synchrotron infrared nano-spectra of hydroxyapatite through Kramers–Kronig relations and complex Lorentzian oscillators. *Sci. Rep.* **15**, 39259 (2025).
80. Zhang, L. M. et al. Near-field spectroscopy of silicon dioxide thin films. *Phys. Rev. B-Condens. Matter Phys.* **85**, 075419 (2012).
81. Govyadinov, A. A., Amenabar, I., Huth, F., Carney, P. S. & Hillenbrand, R. Quantitative measurement of local infrared absorption and dielectric function with tip-enhanced near-field microscopy. *J. Phys. Chem. Lett.* **4**, 1526–1531 (2013).
82. Mooshammer, F. et al. Nanoscale near-field tomography of surface states on $(\mathrm{Bi}_{0.5}\mathrm{Sb}_{0.5})_2\mathrm{Te}_3$. *Nano Lett.* **18**, 7515–7523 (2018).
83. Wirth, K. G. et al. Tunable s-SNOM for nanoscale infrared optical measurement of electronic properties of bilayer graphene. *ACS Photon.* **8**, 418–423 (2021).
84. Nishida, J. et al. Nanoscale heterogeneity of ultrafast many-body carrier dynamics in triple cation perovskites. *Nat. Commun.* **13**, 6582 (2022).
85. Zizlsperger, M. et al. In situ nanoscopy of single-grain nanomorphology and ultrafast carrier dynamics in metal halide perovskites. *Nat. Photon.* **18**, 975–981 (2024).
An example using ground-state THz s-SNOM and EOS to differentiate crystalline phases in perovskite and then resolving the degree of spatial heterogeneity in carrier diffusion in a visible pump/THz probe experiment.
86. Huber, A. J., Kazantsev, D., Keilmann, F., Wittborn, J. & Hillenbrand, R. Simultaneous IR material recognition and conductivity mapping by nanoscale near-field microscopy. *Adv. Mater.* **19**, 2209–2212 (2007).
87. Mester, L., Govyadinov, A. A. & Hillenbrand, R. High-fidelity nano-FTIR spectroscopy by on-pixel normalization of signal harmonics. *Nanophotonics* **11**, 377–390 (2022).
88. Zhao, Y., Chen, X., Yao, Z., Liu, M. & Fogler, M. M. Deep-learning-aided extraction of optical constants in scanning near-field optical microscopy. *J. Appl. Phys.* **133**, 133105 (2023).
89. Chen, X. et al. Machine learning for optical scanning probe nanoscopy. *Adv. Mater.* **35**, e2109171 (2023).
90. Chen, X., Ren, R. & Liu, M. Validity of machine learning in the quantitative analysis of complex scanning near-field optical microscopy signals using simulated data. *Phys. Rev. Appl.* **15**, 014001 (2021).
91. Dressel, M. & Grüner, G. *Electrodynamics of Solids*: *Optical Properties of Electrons in Matter.* Illustrated edn (Cambridge Univ. Press, 2002).
92. Parkinson, P. et al. Transient terahertz conductivity of GaAs nanowires. *Nano Lett.* **7**, 2162–2165 (2007).
93. Norman, S. et al. Resonance-amplified terahertz near-field spectroscopy of a single nanowire. *Nano Lett.* **24**, 15716–15723 (2024).
94. Baxter, J. B. & Schmuttenmaer, C. A. Conductivity of ZnO nanowires, nanoparticles, and thin films using time-resolved terahertz spectroscopy. *J. Phys. Chem. B* **110**, 25229–25239 (2006).
95. Cocker, T. L. et al. Microscopic origin of the Drude–Smith model. *Phys. Rev. B* **96**, 205439 (2017).
96. Baiz, C. R., Schach, D. & Tokmakoff, A. Ultrafast 2D IR microscopy. *Opt. Express* **22**, 18724–18735 (2014).
97. Anderson, P. W. A mathematical model for the narrowing of spectral lines by exchange or motion. *J. Phys. Soc. Jpn.* **9**, 316–339 (1954).
98. Gavdush, A. A. et al. Insulator–metal transition in VO_2 film on sapphire studied by broadband dielectric spectroscopy. *Sci. Rep.* **15**, 3500 (2025).
99. Richter, C. & Schmuttenmaer, C. A. Exciton-like trap states limit electron mobility in TiO_2 nanotubes. *Nat. Nanotechnol.* **5**, 769–772 (2010).
100. Di Pietro, P. et al. Impact of terahertz short pulses on the oxygen defect state in TiO_{2-x}. *Phys. Rev. Res.* **7**, 023011 (2025).
101. Huang, K. & Rhys, A. Theory of light absorption and non-radiative transitions in F-centres. *Proc. R. Soc. Lond. Ser. A Math. Phys. Sci.* **204**, 406–423 (1950).
102. Fröhlich, H. Electrons in lattice fields. *Adv. Phys.* **3**, 325–361 (1954).
103. Emin, D. Optical properties of large and small polarons and bipolarons. *Phys. Rev. B* **48**, 13691–13702 (1993).
104. Ghalgaoui, A., Fingerhut, B. P., Reimann, K., Elsaesser, T. & Woerner, M. Terahertz polaron oscillations of electrons solvated in liquid water. *Phys. Rev. Lett.* **126**, 097401 (2021).
105. Elsaesser, T., Woerner, M. & Fingerhut, B. P. Coherent polaron dynamics of electrons solvated in polar liquids. *PNAS Nexus* **1**, pgac078 (2022).
106. Megerle, U., Pugliesi, I., Schriever, C., Sailer, C. F. & Riedle, E. Sub-50 fs broadband absorption spectroscopy with tunable excitation: putting the analysis of ultrafast molecular dynamics on solid ground. *Appl. Phys. B* **96**, 215–231 (2009).
107. Polli, D., Brida, D., Mukamel, S., Lanzani, G. & Cerullo, G. Effective temporal resolution in pump–probe spectroscopy with strongly chirped pulses. *Phys. Rev. A At. Mol. Opt. Phys.* **82**, 053809 (2010).
108. Cleveland, J. M. et al. A framework for extracting the rates of photophysical processes from biexponentially decaying photon emission data. *J. Phys. Chem. C* **129**, 5893–5904 (2025).
109. Milovanov, A., Rasmussen, J. J. & Rypdal, K. Stretched-exponential decay functions from a self-consistent model of dielectric relaxation. *Phys. Lett. A* **372**, 2148–2154 (2008).
110. Johnston, M. B. & Herz, L. M. Hybrid perovskites for photovoltaics: charge-carrier recombination, diffusion, and radiative efficiencies. *Acc. Chem. Res.* **49**, 146–154 (2016).
111. Esmaielpour, H., Isaev, N., Finley, J. J. & Koblmüller, G. Influence of Auger heating and Shockley–Read–Hall recombination on hot-carrier dynamics in InGaAs nanowires. *Phys. Rev. B* **109**, 235303 (2024).
112. Luferau, A. et al. Hot-electron dynamics in a semiconductor nanowire under intense THz excitation. *ACS Photon.* **11**, 3123–3130 (2024).
A THz-pump/MIR-probe study of carrier dynamics in semiconductor nanowires, using the FELBE free-electron laser for THz pulse generation to resolve the dynamics and to infer an increase in the electron effective mass.
113. Lienau, C., Guenther, T., Unold, T., Mueller, K. & Elsaesser, T. Ultrafast near-field spectroscopy of single semiconductor quantum dots. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **362**, 861–879 (2004).
114. Luferau, A. et al. Time-resolved nanospectroscopy of III–V semiconductor nanowires. *Nanoscale Adv.* **7**, 3692–3700 (2025).
115. Li, J. et al. Ultrafast optical nanoscopy of carrier dynamics in silicon nanowires. *Nano Lett.* **23**, 1445–1450 (2023).
116. Pushkarev, V. et al. Charge transport in single-crystalline GaAs nanobars: impact of band bending revealed by terahertz spectroscopy. *Adv. Funct. Mater.* **32**, 2107403 (2022).
117. Mester, L., Govyadinov, A. A., Chen, S., Goikoetxea, M. & Hillenbrand, R. Subsurface chemical nanoidentification by nano-FTIR spectroscopy. *Nat. Commun.* **11**, 3359 (2020).
118. Pirotta, S. et al. Fast amplitude modulation up to 1.5 GHz of mid-IR free-space beams at room-temperature. *Nat. Commun.* **12**, 799 (2021).
119. Chung, H. et al. Electrical phase modulation based on mid-infrared intersubband polaritonic metasurfaces. *Adv. Sci.* **10**, 2207520 (2023).
120. Mann, S. A. et al. Ultrafast optical switching and power limiting in intersubband polaritonic metasurfaces. *Optica* **8**, 606–613 (2021).
121. Wang, Y. et al. Tip-enhanced imaging and control of infrared strong light–matter interaction. *Laser Photon. Rev.* **18**, 2301148 (2024).
122. Tapar, J. K., Kishen, S., Prashant, K., Nayak, K. & Emani, N. K. Enhancement of the optical gain in GaAs nanocylinders for nanophotonic applications. *J. Appl. Phys.* **127**, 153102 (2020).
123. Charnukha, A. et al. Ultrafast nonlocal collective dynamics of Kane plasmon-polaritons in a narrow-gap semiconductor. *Sci. Adv.* **5**, eaau9956 (2019).
124. Mrejen, M., Yadgarov, L., Levanon, A. & Suchowski, H. Transient exciton-polariton dynamics in WSe_2 by ultrafast near-field imaging. *Sci. Adv.* https://doi.org/10.1126/sciadv.aat9618 (2019).
125. Hu, H. et al. Doping-driven topological polaritons in graphene/α-MoO_3 heterostructures. *Nat. Nanotechnol.* **17**, 940–946 (2022).
126. Hu, G., Liu, J., Qiu, C.-W., Alù, A. & Fan, S. Phonon polaritons and hyperbolic response in van der Waals materials. *Adv. Opt. Mater.* **8**, 1901393 (2020).
127. Kowalski, R. A. et al. Ultraconfined terahertz phonon polaritons in hafnium dichalcogenides. *Nat. Mater.* https://doi.org/10.1038/s41563-025-02345-0 (2025).
128. Xu, S. et al. Electronic interactions in Dirac fluids visualized by nano-terahertz spacetime interference of electron–photon quasiparticles. *Sci. Adv.* **10**, 5553 (2024).
129. Morrison, V. R. et al. A photoinduced metal-like phase of monoclinic VO_2 revealed by ultrafast electron diffraction. *Science* **346**, 445–448 (2014).
130. Zylbersztejn, A. & Mott, N. F. Metal–insulator transition in vanadium dioxide. *Phys. Rev. B* **11**, 4383–4395 (1975).
131. Sternbach, A. J. et al. Programmable hyperbolic polaritons in van der Waals semiconductors. *Science* **371**, 617–620 (2021).
132. Sternbach, A. J. et al. Inhomogeneous photosusceptibility of VO_2 films at the nanoscale. *Phys. Rev. Lett.* **132**, 186903 (2024).
133. Cavalleri, A., Dekorsy, T., Chong, H. H. W., Kieffer, J. C. & Schoenlein, R. W. Evidence for a structurally-driven insulator-to-metal transition in VO_2: a view from the ultrafast timescale. *Phys. Rev. B* **70**, 161102 (2004).
134. Brady, N. F. et al. Heterogeneous nucleation and growth dynamics in the light-induced phase transition in vanadium dioxide. *J. Phys. Condens. Matter* **28**, 125603 (2016).

135. Qazilbash, M. M. et al. Mott transition in VO$_2$ revealed by infrared spectroscopy and nano-imaging. *Science* **318**, 1750–1753 (2007).
136. Stinson, H. et al. Imaging the nanoscale phase separation in vanadium dioxide thin films at terahertz frequencies. *Nat. Commun.* **9**, 3604 (2018).
137. Guiton, B. S., Gu, Q., Prieto, A. L., Gudiksen, M. S. & Park, H. Single-crystalline vanadium dioxide nanowires with rectangular cross sections. *J. Am. Chem. Soc.* **127**, 498–499 (2005).
138. Atkin, J. M. et al. Strain and temperature dependence of the insulating phases of VO$_2$ near the metal–insulator transition. *Phys. Rev. B* **85**, 020101 (2012).
139. Jones, A. C., Berweger, S., Wei, J., Cobden, D. & Raschke, M. B. Nano-optical investigations of the metal–insulator phase behavior of individual VO$_2$ microcrystals. *Nano Lett.* **10**, 1574–1581 (2010).
140. Chapman, H. N. et al. Femtosecond diffractive imaging with a soft-X-ray free-electron laser. *Nat. Phys.* **2**, 839–843 (2006).
141. Li, P. et al. Optical nanoimaging of hyperbolic surface polaritons at the edges of van der Waals materials. *Nano Lett.* **17**, 228–233 (2017).
142. Fan, L. et al. Comprehensive studies of interfacial strain and oxygen vacancy on metal–insulator transition of VO$_2$ film. *J. Phys. Condens. Matter* **28**, 255002 (2016).
143. Kim, H., Charipar, N., Osofsky, M., Qadri, S. B. & Piqué, A. Optimization of the semiconductor–metal transition in VO$_2$ epitaxial thin films as a function of oxygen growth pressure. *Appl. Phys. Lett.* **104**, 081913 (2014).
144. Zhang, S. et al. Nano-spectroscopy of excitons in atomically thin transition metal dichalcogenides. *Nat. Commun.* **13**, 542 (2022).
145. Siday, T. et al. Ultrafast nanoscopy of high-density exciton phases in WSe$_2$. *Nano Lett.* **22**, 2561–2568 (2022).
146. Poellmann, C. et al. Resonant internal quantum transitions and femtosecond radiative decay of excitons in monolayer WSe$_2$. *Nat. Mater.* **14**, 889–893 (2015).
147. Meineke, C. et al. Ultrafast exciton dynamics in the atomically thin van der Waals magnet CrSBr. *Nano Lett.* **24**, 4101–4107 (2024).
148. Peña, M. A. & Fierro, J. Chemical structures and performance of perovskite oxides. *Chem. Rev.* **101**, 1981–2018 (2001).
149. Todorov, T., Gunawan, O. & Guha, S. A road towards 25% efficiency and beyond: perovskite tandem solar cells. *Mol. Syst. Des. Eng.* **1**, 370–376 (2016).
150. Ishihara, T., Takahashi, J. & Goto, T. Exciton state in two-dimensional perovskite semiconductor (C$_{10}$H$_{21}$NH$_3$)$_2$PbI$_4$. *Solid State Commun.* **69**, 933–936 (1989).
151. Kulkarni, S. A. et al. Band-gap tuning of lead halide perovskites using a sequential deposition process. *J. Mater. Chem. A* **2**, 9221–9225 (2014).
152. Motti, S. G., Rappe, A. M., Saliba, M., Grätzel, M. & Nazeeruddin, M. K. Heterogeneity at multiple length scales in halide perovskite thin films. *Nat. Rev. Mater.* **4**, 456–470 (2019).
153. Davies, M. L., Fermin, D. J., Stranks, S. D. & Snaith, H. J. Interpreting time-resolved photoluminescence of perovskite materials. *Phys. Chem. Chem. Phys.* **22**, 3623–3631 (2020).
154. Zhang, B. et al. Ultrafast terahertz modulation characteristics of organolead halide perovskite films revealed by time-resolved terahertz spectroscopy. *Sci. Rep.* https://doi.org/10.1038/s41598-018-30396-3 (2018).
155. Khan, J. I. et al. Evaluation of interfacial photophysical processes by time-resolved optical spectroscopy in perovskite solar cells. *Matter* **8**, 316–325 (2024).
156. Panuganti, S. et al. Transient X-ray diffraction reveals nonequilibrium phase transition in thin films of CH$_3$NH$_3$PbI$_3$ Perovskite. *ACS Energy Lett.* **8**, 691–698 (2023).
157. Wieghold, S. et al. Stressing halide perovskites with light and electric fields. *ACS Energy Lett.* **7**, 2211–2218 (2022).
158. Taylor, V. C. et al. Investigating the role of the organic cation in formamidinium lead iodide perovskite using ultrafast spectroscopy. *J. Phys. Chem. Lett.* **9**, 895–901 (2018).
159. Sandner, D. et al. Hole localization in bulk and 2D lead-halide perovskites studied by time-resolved infrared spectroscopy. *J. Am. Chem. Soc.* **146**, 19852–19862 (2024).
160. Nishida, J., Alfaifi, A. H., Gray, T. P., Shaheen, S. E. & Raschke, M. B. Heterogeneous cation–lattice interaction and dynamics in triple-cation perovskites revealed by infrared vibrational nanoscopy. *ACS Energy Lett.* **5**, 1636–1643 (2020).
161. Pollard, B., Muller, E. A., Hinrichs, K. & Raschke, M. B. Vibrational nano-spectroscopic imaging correlating structure with intermolecular coupling and dynamics. *Nat. Commun.* **5**, 3587 (2014).
162. Onsager, L. Electric moments of molecules in liquids. *J. Am. Chem. Soc.* **58**, 1486–1493 (1936).
163. Behr, N. & Raschke, M. B. Optical antenna properties of scanning probe tips: plasmonic light scattering, tip–sample coupling, and near-field enhancement. *J. Phys. Chem. C* **112**, 12069–12075 (2008).
164. Bechtel, H. A., Johnson, S. C., Khatib, O., Muller, E. A. & Raschke, M. B. Synchrotron infrared nano-spectroscopy and -imaging. *Surf. Sci. Rep.* https://doi.org/10.1016/j.surfrep.2020.100493 (2020).
165. Shcherbakov, M. R. et al. Photo-induced force microscopy. *Nat. Rev. Methods Primers* **5**, 34 (2025).
166. Neacsu, C. C., Steudle, G. & Raschke, M. B. Plasmonic light scattering from nanoscopic metal tips. *Appl. Phys. B* **80**, 295–300 (2005).
167. Metzner, S. et al. Assessment of subsampling schemes for compressive nano-FTIR Imaging. *IEEE Trans. Instrum. Meas.* **71**, 1–8 (2022).
168. Cheng, P. et al. Adaptive under-sampling strategy for fast imaging in compressive sensing-based atomic force microscopy. *Ultramicroscopy* **261**, 113964 (2024).
169. Fu, M. et al. Accelerated nano-optical imaging through sparse sampling. *Nano Lett.* **24**, 2149–2156 (2024).
170. Yang, R., Li, R., Blankenship, B. W., Li, J. & Grigoropoulos, C. P. Decoupling carrier dynamics and energy transport in ultrafast near-field nanoscopy. *Nano Lett.* **25**, 2242–2247 (2025).
171. Ni, G. et al. Fundamental limits to graphene plasmonics. *Nature* **557**, 530–533 (2018).
172. Dapolito, M. et al. Infrared nano-imaging of Dirac magnetoexcitons in graphene. *Nat. Nanotechnol.* **18**, 1409–1415 (2023).
173. Kim, R. H., Park, J.-M., Haeuser, S. J., Luo, L. & Wang, J. A sub-2 kelvin cryogenic magneto-terahertz scattering-type scanning near-field optical microscope (cm-THz-sSNOM). *Rev. Sci. Instrum*. https://doi.org/10.1063/5.0130680 (2023).
174. O'Callahan, B. T. et al. In liquid infrared scattering scanning near-field optical microscopy for chemical and biological nanoimaging. *Nano Lett.* **20**, 4497–4504 (2020).
175. Chen, Z. et al. A CRISPR/Cas12a-empowered surface plasmon resonance platform for rapid and specific diagnosis of the Omicron variant of SARS-CoV-2. *Natl Sci. Rev.* **9**, nwac104 (2022).
176. Koopman, M. et al. Near-field scanning optical microscopy in liquid for high resolution single molecule detection on dendritic cells. *FEBS Lett.* **573**, 6–10 (2004).
177. Chen, X. et al. Hybrid machine learning for scanning near-field optical spectroscopy. *ACS Photon.* **8**, 2987–2996 (2021).
This work demonstrates utilizing artificial and hybrid neural networks to analyse the tip-scattered signal in IR s-SNOM and improve model fitting, allowing for easier extraction of optical material properties such as the dielectric function.
178. Munson, K. T., Swartzfager, J. R. & Asbury, J. B. Lattice anharmonicity: a double-edged sword for 3D perovskite-based optoelectronics. *ACS Energy Lett.* **4**, 1888–1897 (2019).
179. Pohl, D. W., Denk, W. & Lanz, M. Optical stethoscopy: image recording with resolution λ/2. *Appl. Phys. Lett.* **44**, 651–653 (1984).
180. Courjon, D. *Near-Field Microscopy and Near-Field Optics* (Imperial College Press, 2003).
181. Achermann, M. et al. Direct experimental observation of different diffusive transport regimes in semiconductor nanostructures. *Phys. Rev. B* **60**, 2101–2105 (1999).
182. Achermann, M. et al. Quantization energy mapping of single V-groove GaAs quantum wires by femtosecond near-field optics. *Appl. Phys. Lett.* **76**, 2695–2697 (2000).
183. Levy, J. et al. Spatiotemporal near-field spin microscopy in patterned magnetic heterostructures. *Phys. Rev. Lett.* **76**, 1948–1951 (1996).
184. Guenther, T., Unold, T., Lienau, C., Mueller, K. & Elsaesser, T. Coherent nonlinear optical response of single quantum dots studied by ultrafast near-field spectroscopy. *Phys. Rev. Lett.* **88**, 057401 (2002).
185. Emiliani, V., Guenther, T., Lienau, C., Nötzel, R. & Ploog, K. H. Femtosecond near-field spectroscopy: carrier relaxation and transport in single quantum wires. *J. Microsc.* **202**, 229–233 (2001).
186. Nishiyama, Y., Imura, K. & Okamoto, H. Observation of plasmon wave packet motions via femtosecond time-resolved near-field imaging techniques. *Nano Lett.* **15**, 7657–7665 (2015).
187. Feldstein, M. J. & Scherer, N. F. in *Ultrafast Phenomena X* Vol. 62 of *Springer Series in Chemical Physics* 139–140 (Springer, 1996).
188. Hamers, R. J. & Cahill, D. G. Ultrafast time resolution in scanned probe microscopies: surface photovoltage on Si(111)-(7x7). *J. Vac. Sci. Technol. B* **9**, 514 (1991).
189. Weiss, S., Ogletree, D. F., Botkin, D., Salmeron, M. & Chemla, D. S. Ultrafast scanning probe microscopy. *Appl. Phys. Lett.* **63**, 2567–2569 (1993).
190. Alonso-González, P. et al. Acoustic terahertz graphene plasmons revealed by photocurrent nanoscopy. *Nat. Nanotechnol.* **12**, 31–35 (2017).
191. Ambrosio, A. et al. Mechanical detection and imaging of hyperbolic phonon polaritons in hexagonal boron nitride. *ACS Nano* **11**, 8741–8749 (2017).
192. Cocker, T. L., Peller, D., Yu, P., Repp, J. & Huber, R. Tracking the ultrafast motion of a single molecule by femtosecond orbital imaging. *Nature* **539**, 263–267 (2016).
193. Gutzler, R., Garg, M., Ast, C. R., Kuhnke, K. & Kern, K. Light–matter interaction at atomic scales. *Nat. Rev. Phys.* **3**, 441–453 (2021).
194. Sciaini, G. & Miller, R. D. Femtosecond electron diffraction: heralding the era of atomically resolved dynamics. *Rep. Prog. Phys.* **74**, 096101 (2011).
195. Miao, J., Ishikawa, T., Robinson, I. K. & Murnane, M. M. Beyond crystallography: diffractive imaging using coherent X-ray light sources. *Science* **348**, 530–535 (2015).
196. Kubo, A. et al. Femtosecond imaging of surface plasmon dynamics in a nanostructured silver film. *Nano Lett.* **5**, 1123–1127 (2005).
197. Kubo, A., Pontius, N. & Petek, H. Femtosecond microscopy of surface plasmon polariton wave packet evolution at the silver/vacuum interface. *Nano Lett.* **7**, 470–475 (2007).
198. Feist, A. et al. Ultrafast transmission electron microscopy using a laser-driven field emitter: femtosecond resolution with a high coherence electron beam. *Ultramicroscopy* **176**, 63–73 (2017).
199. Zewail, A. H. Four-dimensional electron microscopy. *Science* **328**, 187–193 (2010).
200. Schumacher, Z. et al. Nanoscale force sensing of an ultrafast nonlinear optical response. *Proc. Natl Acad. Sci. USA* **117**, 19773–19779 (2020).
201. Gruber, C. et al. Imaging nanowire plasmon modes with two-photon polymerization. *Appl. Phys. Lett.* https://doi.org/10.1063/1.4913470 (2015).

Acknowledgements

B.L.E., P.A.M. and M.B.R. acknowledge funding from the US Department of Energy, Office of Basic Sciences, Division of Material Sciences and Engineering, under award no. DESC0008807. R.L.P., R.W., J.N. and M.B.R. further acknowledge support from the NSF Science and Technology Center on Real-Time Functional Imaging (STROBE) grant no. DMR-1548924. D.S. acknowledges funding support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy-EXC 2089 (e-Conversion).

Author contributions

Introduction (M.B.R., D.S. and B.L.E.); Experimentation (B.L.E., D.S., R.L.P. and J.N.); Results (R.L.P., D.S. and J.N.); Applications (B.L.E., D.S., P.A.M., M.B.R. and R.W.); Reproducibility and data deposition (R.L.P.); Limitations and optimizations (R.L.P. and P.A.M.); Outlook (D.S., R.L.P., P.A.M. and B.L.E.); overview of the Primer (M.B.R., R.W. and B.L.E.). All authors discussed and edited the full manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Peer review information *Nature Reviews Methods Primers* thanks Mengkun Liu, who co-reviewed with Yuanchi Qing; Han Zhang; and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Related links

Bruker: https://www.brukerafmprobes.com/c-15-probes.aspx
NanoAndMore: https://www.nanoandmore.com/afm-probes