This document is the accepted manuscript version of the following article:
Siday, T., Hayes, J., Schiegl, F. *et al.* All-optical subcycle microscopy on atomic length scales. Nature, 629, 329-334 (2024). https://doi.org/10.1038/s41586-024-07355-7

All-optical subcycle microscopy on atomic lengthscales

T. Siday[1,†], J. Hayes[1,†], F. Schiegl[1,†], F. Sandner[1], P. Menden[1], V. Bergbauer[1], M. Zizlsperger[1], S. Nerreter[1], S. Lingl[1], J. Repp[1], J. Wilhelm[1,], M. A. Huber[1,*], Y. A. Gerasimenko[1,*] and R. Huber[1]*

[1]Department of Physics and Regensburg Center for Ultrafast Nanoscopy (RUN), University of Regensburg, 93040 Regensburg, Germany

[†]*These authors contributed equally to this work*

**Correspondence should be addressed to these authors*

Optical microscopy at the shortest possible length- and timescales has been a long-sought Rosetta stone, connecting nanoscopic elementary dynamics with the macroscopic functionalities of condensed matter. Super-resolution microscopy has circumvented the far-field diffraction limit by harnessing optical nonlinearities[1]. By exploiting linear interaction with tip-confined evanescent light fields[2], near-field microscopy[3,4] has reached even higher resolution, prompting a vibrant research field by exploring the nanocosm in motion[5-19]. Yet the finite radius of the nanometre-sized tip apex has prevented access to atomic resolution[20]. Here, we leverage extreme atomic nonlinearities within tip-confined evanescent fields to push all-optical microscopy to picometric spatial and femtosecond temporal resolution. On these scales, we discover an entirely unprecedented and efficient non-classical near-field response, in phase with the vector potential of light and strictly confined to atomic dimensions. This ultrafast signal is characterised by an optical phase delay of ~$\pi/2$ and facilitates direct monitoring of tunnelling dynamics. We showcase the power of our new optical concept by imaging nanometre-sized defects hidden to atomic force microscopy and by subcycle sampling of current transients on a semiconducting van der Waals material. Our results facilitate a radically new access to quantum light-matter interaction and electronic dynamics at ultimately short spatio-temporal scales in both conductive and insulating quantum materials.

Field-resolving metrology of electromagnetic waves has heralded a revolution in the science of light-matter interaction. By exploiting calibrated nonlinearities to directly map the time evolution of optical carrier waves, electro-optic and streaking detection have accessed the maximal classical information[21,22] as well as quantum field fluctuations[23,24] of light, laying the foundations for strong-field and attosecond science[25]. Precise clocking of light emission has revealed microscopic dynamics of electrons with a resolution faster than a cycle of light, including the build-up of screening[26], lightwave-driven acceleration[27-30], electron tunnelling[31,32] and many-body correlations[33]. Yet, extracting subcycle optical information from single atoms in molecular or crystalline media has remained out of reach.

Near-field microscopy exploits evanescent fields confined by miniscule metallic objects to access light-matter interaction on nanometre lengthscales, orders of magnitude below the diffraction limit[3,4]. It has enabled foundational insights into phase transitions[5], light-matter interaction in quantum materials[10], *in operando* photochemistry[19] and elusive quasiparticles – from Dirac plasmons[6-8,14,15] to hyperbolic[16] and topological phonon[13] polaritons. It has also proven a natural companion to ultrafast spectroscopy, sampling the femtosecond evolution of the dielectric function while retaining nanometre resolution[9,11,17,18,34]. Yet, the spatial resolution of near-field microscopy remains fundamentally limited by the mesoscopic size of the probe apex (~10 nm).

This resolution threshold has recently been surpassed in optical imaging of individual molecules[35-37], exploiting inelastic scattering and luminescence to achieve sub-nanometre lateral resolution. While sampling such responses on femtosecond scales has been out of reach, tantalising glimpses into this domain have become possible with the advent of lightwave-driven scanning tunnelling microscopy (LW-STM)[38-44], which maps time-integrated electric currents rectified by the carrier wave of light. However, measuring the rapid motion of electrons within a single cycle of light remains elusive. Instead, LW-STM relies on pivotal assumptions about the light-matter interaction within the sample itself to reconstruct transient currents (Methods).

Here, we demonstrate a fundamentally new optical microscopy paradigm which capitalises on atomic nonlinearities of mesoscopic near fields. This provides subcycle access to the optical response of matter on femtosecond timescales with picometric spatial resolution. The key to our approach is to combine, for the first time, all-optical detection from near-field microscopy with ultra-high vacuum, cryogenic temperatures, and sub-nanometre tip tapping amplitudes. We observe a remarkable emergent,

unpredicted non-classical response, localised to single atomic orbitals. This unconventional near-field response manifests in scattered electromagnetic fields as a phase delay of ~π/2, driven by the ultrafast motion of tunnelling electrons. We harness this near-field optical tunnelling emission (NOTE) to achieve atomic-scale spatial resolution and to directly sample subcycle tunnelling current transients flowing even within the bandgap of a transition metal dichalcogenide semiconductor.

Atomic-scale optical microscopy

The prevailing approach to near-field microscopy relies on evanescent fields enhanced at the apex of a sharp tip illuminated by light. With the apex positioned in proximity to a material surface, the electric dipole determined by the apex radius (~10 nm) polarises the sample. The emergent dipole radiation encoding the nanoscale dielectric function is scattered to the far field and detected all-optically. To isolate this weak dipole radiation from far fields, the tip is tapped at an amplitude $A \approx 100$ nm and harmonic demodulation is used[20].

While atomic protrusions dominate short-range interactions in atomic force microscopy (AFM)[45], their role in near-field experiments remains poorly understood – particularly the role of tunnelling electrons. To explore this, we control the tip-sample separation at the scale of atomic wavefunctions by mounting an atomically sharp tungsten tip to a stiff qPlus AFM sensor[45]. This enables picometric tapping amplitudes, three orders of magnitude smaller than typical for near-field microscopy. To drive atomic nonlinearities, we focus strong terahertz (THz) pulses to the tip apex, and detect the scattered fields with subcycle resolution using electro-optic sampling (EOS) (Fig. 1a, Methods)[18].

We first explore whether the frontier atoms of our tip can modify the scattering of near fields when positioned within a few angstroms above a gold surface. To this end, we sample the scattered THz fields (Fig. 1b) at a large tapping amplitude ($A = 25$ nm), while approaching the tip to the surface (Fig. 1c). For distances $\gtrsim 1$ nm, the peak signal demodulated at the second harmonic of the tapping frequency ($\hat{E}_2^{\mathrm{scat}}$) decays over tens of nanometers, as expected for conventional near fields[20]. Intriguingly, when the tip approaches the sample extremely closely ($\lesssim 1$ nm), the transient transforms, accruing a phase delay $\Delta\varphi$ (Fig. 1b) and a rapid amplitude increase for all demodulation orders, E_i^{scat} (Fig. 1c, Extended Data Fig. 1). This approach curve can be fitted with two independent exponential functions: the conventional near-field interaction (Fig. 1c, blue dashed line), and a much faster decay

in atomic-scale proximity to the sample (Fig. 1c, red dashed line), indicating a fundamentally different contribution to the scattered fields.

The large tapping amplitude ($A = 25\,\mathrm{nm}$) invariably mixes mesoscopic and atomic-scale contributions by spanning orders of magnitude in tip-sample separation over a single tapping cycle. Yet the optical phase is clearly influenced when the tip apex is in atomic proximity to the sample surface. Hence, we reduce A to $200\,\mathrm{pm}$, approximately a single atomic period (Fig. 2a). Under these conditions, a current emerges for the shortest tip-sample separation, even without any static bias voltage applied (Fig. 2b, left). The picometric decay indicates its origin: electrons driven by the ultrafast bipolar transient bias of the THz pulse traversing back and forth through the tunnelling barrier between tip and sample. Those rectified owing to nonlinearities in the local density of states (DOS) contribute to a time-averaged net current $\langle J_{\mathrm{lw}} \rangle$, which when measured electronically forms the basis for both LW-STM[38-44] and lightwave electronic control of nanodevices[30,46].

To investigate if our phase-shifted optical signal can be correlated with LW-STM currents, we simultaneously sample their decays at $t = 0\,\mathrm{fs}$. Intriguingly, even at picometric tapping amplitudes, we observe a strong signal in the scattered fields which decays by an order of magnitude in $\sim 200\,\mathrm{pm}$ – almost identical to $\langle J_{\mathrm{lw}} \rangle$ (Fig. 2b, right, Extended Data Fig. 2c). To confirm the origin of this signal is the same as in Fig. 1, we also capture the scattered THz transients (Fig. 2c, Extended Data Fig. 3). For tip-sample separation $\Delta z = 0\,\mathrm{pm}$ (Fig. 2b), strong waveforms peaking at $t = 0\,\mathrm{fs}$ occur in all harmonics of the scattered field (red spheres), again delayed by a characteristic phase shift $\Delta\varphi \approx \pi/2$ versus transients measured away from the tunnelling barrier (blue spheres).

The picometric decay of the scattered signal, alongside the characteristic $\Delta\varphi \approx \pi/2$ phase delay acquired for transients measured in this regime exposes the microscopic origin of the atomically confined near fields: an alternating optical polarisation emanating from a.c. tunnelling currents flowing in response to the instantaneous THz electric field. This result is unintuitive, since light emission from this electronic tunnelling process alone would be practically unmeasurable, significantly weaker than the hypothetical emission of an electron on a classical trajectory around the nucleus of a hydrogen atom. Yet, by exploiting the efficient coupling of evanescent fields from the tip apex to the far field, tunnelling emission dominates even the conventional near-field response for atomic-scale tapping amplitudes.

Key NOTE features can be qualitatively understood with a semiclassical dipole model (Methods), where the near-field response is treated as a point dipole $\boldsymbol{p}_{\mathrm{nf}}$ (Fig. 3a, top) above an infinite half-space[3], driven transiently by the external light field. The resultant ultrafast tunnelling current J_{lw} charges a local dipole $\boldsymbol{p}_{\mathrm{lw}}$ which oscillates with a phase delay of $\pi/2$ with respect to $\boldsymbol{p}_{\mathrm{nf}}$ (Fig. 3b, top). By incorporating the spectral response function of tip and detector, this model (Methods, Extended Data Figs. 4-6) already yields good qualitative agreement with the measured transients (Fig. 2c). Yet, the model ultimately assumes that tunnelled electrons form the nanoscale dipole $\boldsymbol{p}_{\mathrm{lw}}$, and that the ultrafast flow of electrons in NOTE follows the same I-V curve as DC tunnelling, which is known to be a fairly coarse approximation[40].

To understand how $\boldsymbol{p}_{\mathrm{lw}}$ forms without any assumptions on tunneling, we performed *ab initio* quantum simulations of the charge density inside a mesoscopic tip-sample structure of 1010 Na atoms, excited with a lightwave pulse (Methods). Electron tunnelling is naturally included through the overlap of atomic orbitals. The quantum simulations reproduce the hallmark $\Delta\varphi \approx \pi/2$ phase delay (Fig. 3c, red and blue curves). At a tip-sample separation of 9 Å, where lightwave-driven currents flow, the simulation provides direct access to the distribution of both tunnelling currents (Fig. 3a, bottom) and the NOTE dipole (Fig. 3b, bottom) in space. The tunnelling current distribution is extremely inhomogeneous owing to the polarisation of individual atomic orbitals, resulting in the decay of both J_{lw} and $\boldsymbol{p}_{\mathrm{lw}}$ by 14 Å (Fig. 3a,b bottom, Extended Data Fig. 7), and illustrating how atomic-scale resolution could be achieved with NOTE. The simulations also reveal the origin of the time-integrated current $\langle J_{\mathrm{lw}} \rangle$ measured in LW-STM[38-44], being simply proportional to the residual strength of $\boldsymbol{p}_{\mathrm{lw}}$ at large electro-optic delay times (Fig. 3c). This demonstrates precisely how much richer the available information from NOTE microscopy is over LW-STM.

Hints into more exotic quantum dynamics can also be seen in the narrowing of experimental curves (Fig. 2c), something entirely missed by the semiclassical model: for $50\ \mathrm{fs} < t < 200\ \mathrm{fs}$, the NOTE dipole deviates from the exact $\pi/2$ phase shift. Likely, screening induced by transferred electrons polarises the surrounding media, affecting the tunnelling barrier for the subsequent electrons. The small size of $\boldsymbol{p}_{\mathrm{lw}}$ predicted by the quantum simulations explains why NOTE was entirely unforeseen. Only through picometre-stable experimental modulation of scattered fields can the atomic-scale NOTE signal

be separated from much larger mesoscopic screening currents. Yet, even our cutting-edge quantum simulations do not paint the entire picture. The orders of magnitude in scale required to quantitatively model NOTE – an inherently non-equilibrium and non-perturbative process – poses major challenges for even cutting-edge quantum theory. While our quantum simulations grasp the key NOTE mechanism, a quantitative description of time-dependent many-body interactions, including screening, scattering and emission of phonons and plasmons, would require further theoretical developments.

Angstrom-scale lateral resolution

To test if atomic-scale lateral confinement of the NOTE signal is accessible experimentally, we focus on topographically flat surface features where contrast should emerge solely from spatial variations in tunnelling probability, if currents are localised to the atomic scale. Figures 4a,b show conventional multi-mode images (Methods) of the smallest structures available to us without topographical artefacts – nanometre-sized packing defects in the herringbone reconstruction of an Au(111) surface. Intriguingly, in quasi-constant height mode (Methods), we measure a local suppression of $\langle J_{\mathrm{lw}} \rangle$ with atomic-scale precision in the vicinity of the defect (Fig. 4c).

To explore whether such defects can be resolved in the NOTE signal, we trace a line cut across a comparable defect (Fig. 4d,e, location in Fig. 4a), while simultaneously acquiring $\langle J_{\mathrm{lw}} \rangle$ and the instantaneous NOTE signal, measured at the peak of the scattered transient ($t = 0$ fs). Both signals trace out the entire feature, indicating how the resolution of both NOTE and LW-STM emerge from the orbital overlap between the frontier atom of the tip and the sample. While the lateral extent of tunnelling into the defect limits the confinement of both NOTE and STM signals, this does not imply a resolution limit; rather, it suggests that the resolution rulebook follows that of state-of-the-art STM, where sub-Angstrom features can be resolved. Unlike STM-related techniques, NOTE requires no electrical current measurements and like AFM[47] and SNOM is, therefore, not inherently limited to electrically conductive samples. NOTE now facilitates atomic and subcycle resolution without any need for DC conductivity.

Subcycle quantum flow of electrons

Figure 5 provides a glimpse into how tunnelling-induced THz emission carries time-domain information about non-equilibrium electronic transport in semiconducting quantum materials. We study

a monolayer of WSe_2 exfoliated on Au(111) (Methods). DC-STM imaging (Fig. 5a) resolves the moiré superlattice formed between gold and WSe_2 alongside the expected semiconductor-like differential conductance (Extended Data Fig. 8, Methods). We sweep the THz field strength and sample the NOTE signal at the peak of the transient ($t = 0$ fs, Fig. 5c). Here, another optical nonlinearity emerges, lining up well with the DC conductance of the monolayer (Fig. 5c, dashed lines). This excellent agreement is characteristic of the subcycle sampling intrinsic to NOTE – the signal at $t = 0$ fs essentially measures the instantaneous tunnelling current at the peak of the incident field $J_{\mathrm{lw}}(t \approx -250\ \mathrm{fs})$ owing to the $\pi/2$ phase shift. Whereas NOTE offers model-free access to subcycle tunnelling spectroscopy, this is not the case for the time-averaged current $\langle J_{\mathrm{lw}} \rangle$ (Fig. 5b). While an onset is also traced out by $\langle J_{\mathrm{lw}} \rangle$, this current emanates only from electrons rectified by our specific combination of waveform and conductance; in the process, subcycle dynamics are lost.

While the simplicity of the monolayer tunnelling junction results in comparable DC and ultrafast I-V curves, in most materials the ultrafast flow of tunnelling electrons is unlikely to line up with DC measurements. To demonstrate this, we switch to a more insulating sample: a native trilayer of WSe_2 exfoliated on Au(111). The extra layers of WSe_2 introduce ultrafast tunnelling pathways hidden from time-integrated measurements. Figure 5d shows the NOTE signal (red) and near fields (blue) on the trilayer. By inverting the tip transfer function (Methods), we trace the near-field and NOTE dipoles as a function of time (Fig. 5e). Taking the time derivative of the NOTE dipole, we directly clock the evolution of ultrafast tunnelling (Fig. 5e, purple). Fascinatingly, electrons appear to enter the trilayer even for voltages inside the DC WSe_2 bandgap. Here, field-induced band bending and subcycle distortion of atomic orbitals in the top WSe_2 layer – owing to charges trapped within the trilayer over the entire THz pulse – may contribute to the signal. These material-specific insights into the atomic-scale a.c. flow of electrons without *a priori* assumptions about ultrafast tunnelling is NOTE's greatest strength, remaining inaccessible to any other technique.

Discussion and conclusions

In conclusion, we have discovered an entirely unforeseen quantum-mechanical contrast mechanism for optical near fields on atomic lengthscales: a.c. tunnelling currents driven by optical near fields emitting tip-enhanced electromagnetic waves. This emission enables direct clocking of subcycle currents in a

picometre-tuneable tunnel junction for the first time, and brings ultrafast optical microscopy to the atomic scale. Unlike STM, NOTE microscopy is inherently compatible with insulating materials by transiently driving electrons to and from the sample[48] within a single THz pulse, leaving essentially zero rectified tunnelling current. In contrast to current-sampling techniques, NOTE's quantitative field sampling of the emitted light relies on the well-defined $\chi^{(2)}$ optical nonlinearity of EOS, thus decoupling the detection scheme from the inherent sample dynamics. This provides experimental access to electronic dynamics in a wide array of quantum materials and future quantum information platforms, where long-lived spin and quasiparticle coherences require reduced coupling to itinerant electrons. With electro-optic sampling being available throughout the infrared and visible domain, we expect NOTE will be widely scalable in frequency, and even possible at room temperature. Hence the power of all-optical subcycle spectroscopy can now be combined with atomic resolution. This opens the door to strong-field dynamics previously only accessible over macroscopic lengthscales, such as lightwave electronics and valleytronics[30], quantum nanoplasmonics[49,50], lightwave engineering of quantum phases, high-harmonic generation, and many other timely aspects of attosecond physics.

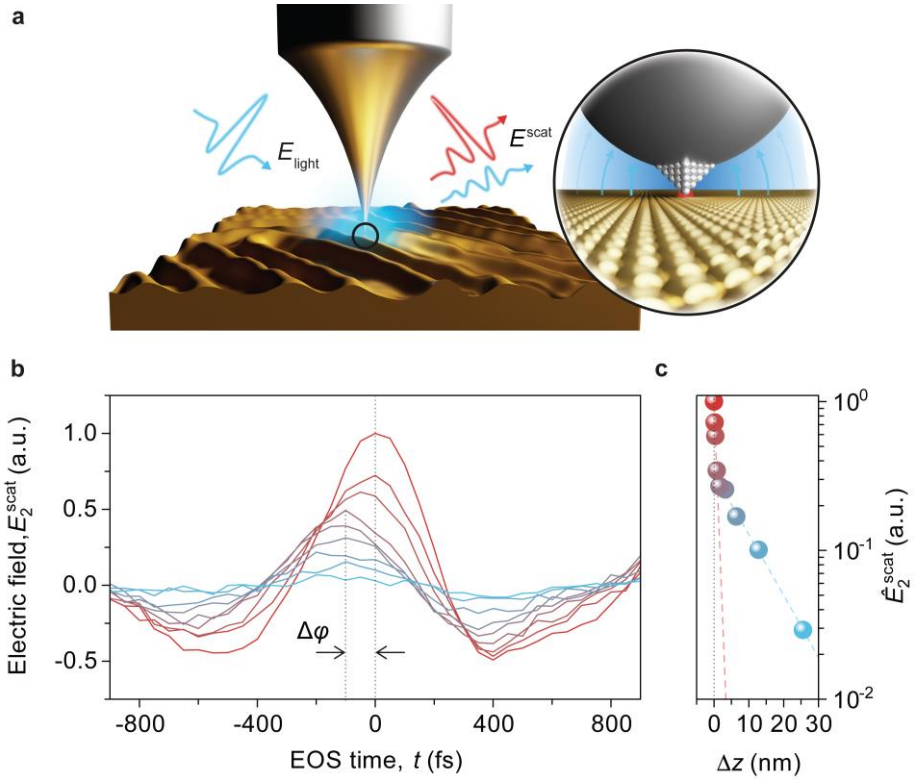
References

1. Hell, S. W. Far-field optical nanoscopy. *Science* **316**, 1153–1158 (2007).
2. Betzig, E. & Trautman, J. K. Near-field optics: microscopy, spectroscopy, and surface modification beyond the diffraction limit. *Science* **257**, 189–195 (1992).
3. Zenhausern, F., Martin, Y. & Wickramasinghe, H. K. Scanning interferometric apertureless microscopy: optical imaging at 10 Angstrom resolution. *Science* **269**, 1083–1085 (1995).
4. Knoll, B. & Keilmann, F. Near-field probing of vibrational absorption for chemical microscopy. *Nature* **399**, 134–137 (1999).
5. Qazilbash, M. M. et al. Mott transition in VO_2 revealed by infrared spectroscopy and nano-imaging. *Science* **318**, 1750–1753 (2007).
6. Fei, Z. et al. Gate-tuning of graphene plasmons revealed by infrared nano-imaging. *Nature* **487**, 82–85 (2012).
7. Chen, J. et al. Optical nano-imaging of gate-tunable graphene plasmons. *Nature* **487**, 77–81 (2012).
8. Lundeberg, M. B. et al. Tuning quantum nonlocal effects in graphene plasmonics. *Science* **357**, 187–191 (2017).
9. Huber, M. A. et al. Femtosecond photo-switching of interface polaritons in black phosphorus heterostructures. *Nat. Nanotechnol.* **12**, 207–211 (2017).
10. Sunku, S. S. et al. Photonic crystals for nano-light in moiré graphene superlattices. *Science* **362**, 1153–1156 (2018).
11. 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).
12. Esmann, M. et al. Vectorial near-field coupling. *Nat. Nanotechnol.* **14**, 698–704 (2019).
13. Hu, G. et al. Topological polaritons and photonic magic angles in twisted α-MoO_3 bilayers. *Nature* **582**, 209–213 (2020).
14. Zhao, W. et al. Efficient Fizeau drag from Dirac electrons in monolayer graphene. *Nature* **594**, 517–521 (2021).
15. Dong, Y. et al. Fizeau drag in graphene plasmonics. *Nature* **594**, 513–516 (2021).
16. Zhang, Q. et al. Interface nano-optics with van der Waals polaritons. *Nature* **597**, 187–195 (2021).

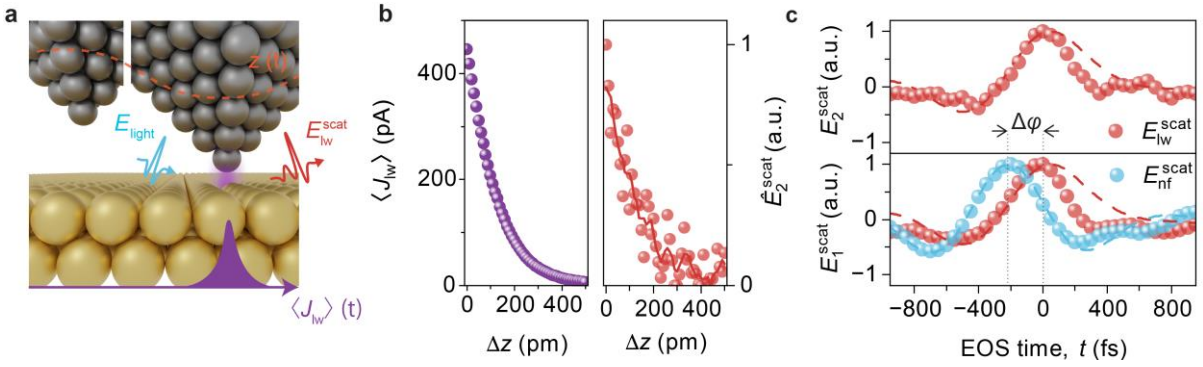
17. Sternbach, A. J. et al. Programmable hyperbolic polaritons in van der Waals semiconductors. *Science* **371**, 617–620 (2021).
18. Plankl, M. et al. Subcycle contact-free nanoscopy of ultrafast interlayer transport in atomically thin heterostructures. *Nat. Photon.* **15**, 594–600 (2021).
19. Meirzadeh, E. et al. A few-layer covalent network of fullerenes. *Nature* **613**, 71–76 (2023).
20. Novotny, L. & Hecht, B. *Principles of Nano-Optics*. (Cambridge, Cambridge University Press, 2012).
21. Leitenstorfer, A. et al. The 2023 terahertz science and technology roadmap. *J. Phys. D: Appl. Phys.* **56**, 223001 (2023).
22. Kienberger, R. et al. Atomic transient recorder. *Nature* **427**, 817–821 (2004).
23. Riek, C. et al. Direct sampling of electric-field vacuum fluctuations. *Science* **350**, 420–423 (2015).
24. Benea-Chelmus, I.-C., Settembrini, F. F., Scalari, G. & Faist, J. Electric field correlation measurements on the electromagnetic vacuum state. *Nature* **568**, 202–206 (2019).
25. Corkum, P. B. & Krausz, F. Attosecond science. *Nat. Phys.* **3**, 381–387 (2007).
26. Huber, R. et al. How many-particle interactions develop after ultrafast excitation of an electron–hole plasma. *Nature* **414**, 286–289 (2001).
27. Wimmer, L. et al. Terahertz control of nanotip photoemission. *Nat. Phys.* **10**, 432–436 (2014).
28. Hohenleutner, M. et al. Real-time observation of interfering crystal electrons in high-harmonic generation. *Nature* **523**, 572–575 (2015).
29. Boolakee, T. et al. Light-field control of real and virtual charge carriers. *Nature* **605**, 251–255 (2022).
30. Borsch, M., Meierhofer, M., Huber, R. & Kira, M. Lightwave electronics in condensed matter. *Nat. Rev. Mater.* **8**, 668–687 (2023).
31. Sainadh, U. S. et al. Attosecond angular streaking and tunnelling time in atomic hydrogen. *Nature* **568**, 75–77 (2019).
32. Shafir, D. et al. Resolving the time when an electron exits a tunnelling barrier. *Nature* **485**, 343–346 (2012).

33. Freudenstein, J. et al. Attosecond clocking of correlations between Bloch electrons. *Nature* **610**, 290–295 (2022).
34. Klarskov, P., Kim, H., Colvin, V. L. & Mittleman, D. M. Nanoscale laser terahertz emission microscopy. *ACS Photonics* **4**, 2676–2680 (2017).
35. Zhang, R. et al. Chemical mapping of a single molecule by plasmon-enhanced Raman scattering. *Nature* **498**, 82–86 (2013).
36. Lee, J., Crampton, K. T., Tallarida, N. & Apkarian, V. A. Visualizing vibrational normal modes of a single molecule with atomically confined light. *Nature* **568**, 78–82 (2019).
37. Yang, B. et al. Sub-nanometre resolution in single-molecule photoluminescence imaging. *Nat. Photon.* **14**, 693–699 (2020).
38. Cocker, T. L. et al. An ultrafast terahertz scanning tunnelling microscope. *Nat. Photon.* **7**, 620–625 (2013).
39. 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).
40. Jelic, V. et al. Ultrafast terahertz control of extreme tunnel currents through single atoms on a silicon surface. *Nat. Phys.* **13**, 591–598 (2017).
41. Luo, Y. et al. Nanoscale terahertz STM imaging of a metal surface. *Phys. Rev. B* **102**, 205417 (2020).
42. Garg, M. et al. Real-space subfemtosecond imaging of quantum electronic coherences in molecules. *Nat. Photon.* **16**, 196–202 (2022).
43. Wang, L., Xia, Y. and Ho, W. Atomic-scale quantum sensing based on the ultrafast coherence of an H_2 molecule in an STM cavity. *Science* **376**, 401–405 (2022).
44. Arashida, Y. et al. Subcycle Mid-Infrared Electric-Field-Driven Scanning Tunneling Microscopy with a Time Resolution Higher Than 30 fs. *ACS Photonics* **9**, 3156–3164 (2022).
45. Giessibl, F. J. Advances in atomic force microscopy. *Rev. Mod. Phys.* **75**, 949–983 (2003).
46. Bionta, M. R. et al. On-chip sampling of optical fields with attosecond resolution. *Nat. Photon.* **15**, 456–460 (2021).

47. Schumacher, Z., Spielhofer, A., Miyahara, Y., Grutter, P. The limit of time resolution in frequency modulation atomic force microscopy by a pump-probe approach. *Appl. Phys. Lett.* **110**, 053111 (2017).
48. Patera, L. L., Queck, F., Scheuerer, P. & Repp, J. Mapping orbital changes upon electron transfer with tunnelling microscopy on insulators. *Nature* **566**, 245–248 (2019).
49. Varas, A., García-González, P., Feist, J., García-Vidal, F. J. & Rubio, A. Quantum plasmonics: from jellium models to ab initio calculations. *Nanophotonics* **5**, 409–426 (2016).
50. Takeuchi, T. & Yabana, K. Extremely large third-order nonlinear optical effects caused by electron transport in quantum plasmonic metasurfaces with subnanometer gaps. *Sci. Rep.* **10**, 21270 (2020).



300 **Figure 1 | Near-field response at atomic-scale tip-sample separation. a,** A THz light pulse (E_{light})
301 is coupled to the apex of a sharp tungsten tip positioned close to the surface of Au(111). This results in
302 a nanoscale dipole that polarizes the gold surface and scatters light into the far field (E^{scat}) where it is
303 detected using electro-optic sampling (EOS). By modulating the height of the STM tip using a qPlus
304 AFM sensor and measuring at harmonics of the tip tapping frequency, E^{scat} can be isolated from the
305 far-field background. We focus on the key open question of how atomic-scale protrusions at the tip
306 apex (zoom-in) contribute to the scattered light fields detected by EOS. **b,** Electro-optically detected
307 scattered THz transients demodulated at the second harmonic of the tip tapping frequency (E_2^{scat}) with
308 a tapping amplitude $A = 25$ nm. From blue to red, the average separation between tip and sample
309 decreases. For the shortest tip-sample separation, the transient is markedly transformed, accruing a
310 phase shift $\Delta\varphi$, and dramatic increase in peak amplitude. On average, ~900 electrons are rectified per
311 pulse. **c,** Peak of the THz transient ($\hat{E}_2^{\mathrm{scat}}$) at $t = 0$ fs, for increasing tip-sample separation Δz. The
312 dashed lines show independent exponential fits for the decay of the near field and the contribution to
313 the measured signal emerging at atomic-scale proximity to the sample. The dramatic increase in $\hat{E}_2^{\mathrm{scat}}$
314 near $\Delta z = 0$ nm (grey dashed vertical line) hints towards contributions from atomic protrusions on the
315 tip apex.



316 **Figure 2 | Picometric decay of the lightwave-driven emission from tunnelling currents. a,**
317 Schematic illustrating the isolation of the atomic-scale near-field signal. The tip tapping amplitude is
318 set far smaller than used in conventional near-field microscopy – to an amplitude on the order of a single
319 atomic period (200 pm). This suppresses the conventional near-field response, as well as far-field
320 contributions. When the distance of closest approach for frontier atoms of the tip extends into the
321 tunnelling region, lightwave-driven tunnelling currents are modulated in time over the tapping cycle as
322 indicated by $\langle J_{\mathrm{lw}} \rangle(\mathrm{t})$ and emit coherent radiation ($E_{\mathrm{lw}}^{\mathrm{scat}}$). **b,** Peak of the electro-optically detected
323 scattered THz transients $\hat{E}_2^{\mathrm{scat}} = E_2^{\mathrm{scat}}(t = 0\ \mathrm{fs})$ measured at the second harmonic of the tip tapping
324 frequency ($A = 200$ pm), alongside the time-integrated lightwave tunnelling current $\langle J_{\mathrm{lw}} \rangle$ measured
325 for increasing tip-sample separation. **c,** Electro-optically detected scattered THz transients measured at
326 the first two harmonics of the tip tapping frequency ($A = 200$ pm). The red curve represents the case
327 of an approached tip ($\Delta z \approx 0$ pm) in **b** where on average ~180 electrons per pulse are rectified. The blue
328 curve shows the scattered near fields, sampled at a larger tip tapping amplitude ($A = 2$ nm). The dashed
329 lines correspond to the semiclassical dipole model described in Methods.

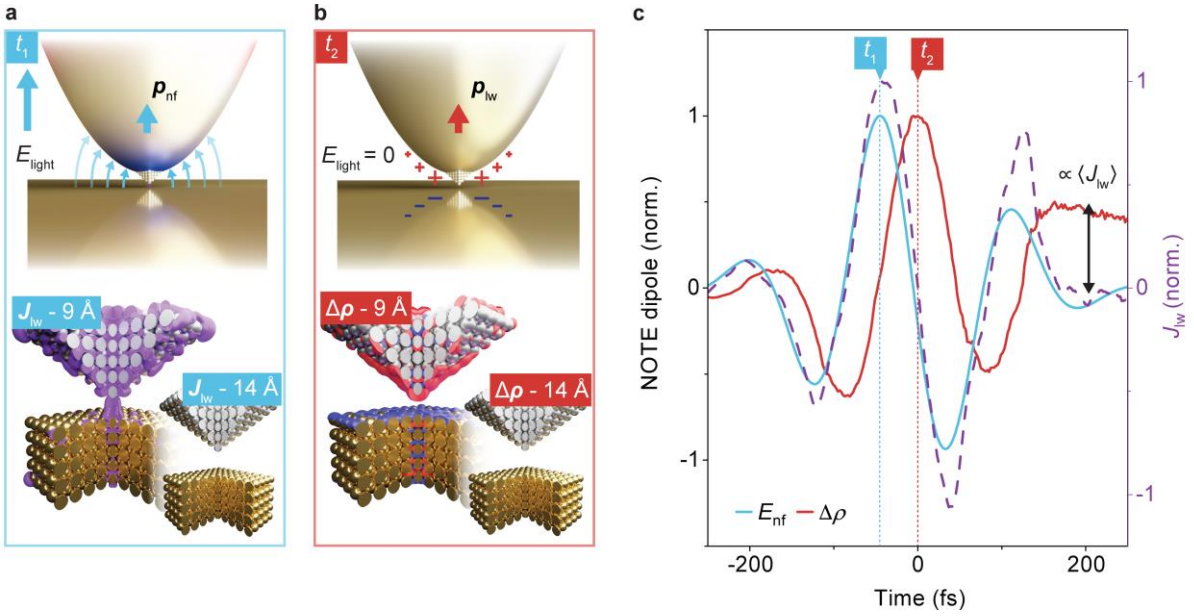


Figure 3 | Microscopic picture of THz emission by lightwave-driven tunnelling currents. a, Formation of a mesoscopic near-field dipole $\boldsymbol{p}_{\mathrm{nf}}$ at the tip apex, driven by the external THz electric field E_{light} at time t_1 in **c** (top). In the bottom part of the figure, the frontier atoms of the tip are shown, where atomically confined tunnelling currents J_{lw} flow in response to the near-field dipole. Time-dependent density functional theory is used here to calculate the ultrafast tunnelling currents at this instant, shown by the purple isosurface for tip-sample separations 9 Å (where lightwave currents flow) and 14 Å (where almost no lightwave currents flow). **b**, At time t_2, there is no near-field dipole $\boldsymbol{p}_{\mathrm{nf}}$, as E_{light} is near zero (top). However, the *ab initio* calculations (bottom) show that even without external fields, a dipole emerges – shown by an isosurface of the relative charge density $\Delta\rho$ – at a tip-sample separation of 9 Å. This is induced by lightwave-driven tunnelling currents, and entirely disappears when the tip is retracted away from the tunnelling contact to 14 Å. **c**, Time evolution of the near fields calculated in the vicinity of the tip apex (blue curve) and $\Delta\rho$ (red curve, proportional to $\boldsymbol{p}_{\mathrm{lw}}$), calculated in the *ab initio* simulations, for a tip-sample separation of 9 Å. The corresponding lightwave-driven tunnelling currents are shown by the purple dashed curve. Nonlinearities in the formation of the tunnelling currents result in the accumulation of charges at large delay times. This charge build-up is proportional to the time-integrated lightwave tunnelling current $\langle J_{\mathrm{lw}} \rangle$.

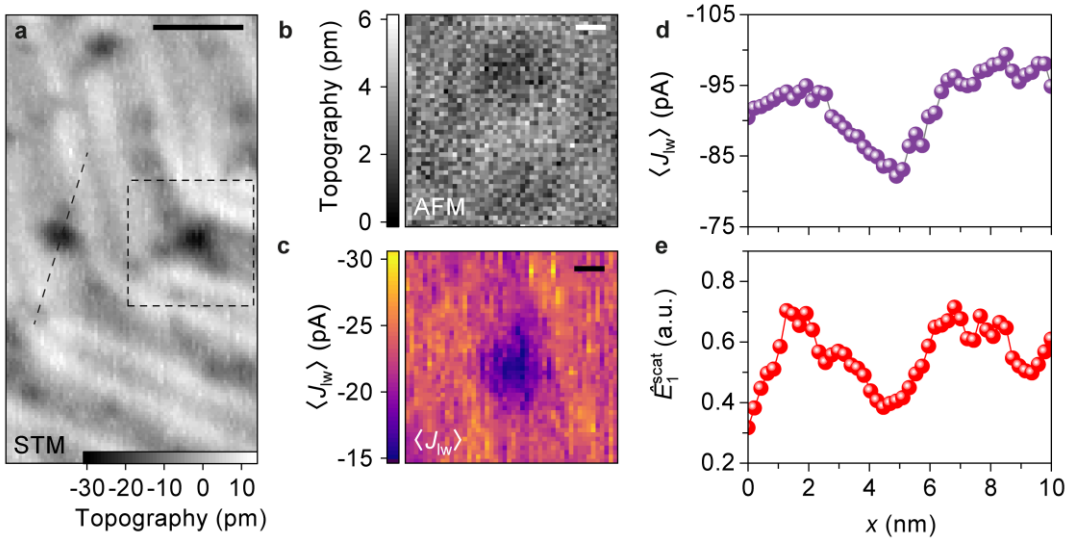


Figure 4 | Spatial resolution of NOTE microscopy. a, Constant-current STM image of the Au(111) surface ($V = -450$ mV, $I_{set} = 100$ pA), showing a herringbone reconstruction and nanometre sized lattice defects. Scale bar: 5 nm. **b,** AFM topography ($A = 250$ pm, $\Delta\nu = -6$ Hz) of the region shown by the dashed square in **a**. Note that in AFM the depression is barely visible, indicating that the origin of the STM signal is related to a modulation in the local DOS. **c,** Quasi-constant height lightwave STM image measured simultaneously with **b** clearly featuring the lattice defect. Scale bars: 1 nm. **d,** Lightwave STM line scan over a similar defect ($A = 250$ pm, $\Delta\nu = -5.8$ Hz) showing the time-averaged lightwave current $\langle J_{lw} \rangle$. **e,** The same defect imaged via its NOTE signal, simultaneously with **d**. Shown here is the first harmonic of the detected electric field, measured at the peak of the THz transient. For clarity, the data is smoothed using a Savitzky-Golay filter with a 5-point window. The linecut in **d, e** is taken along the dashed line shown in **a**.

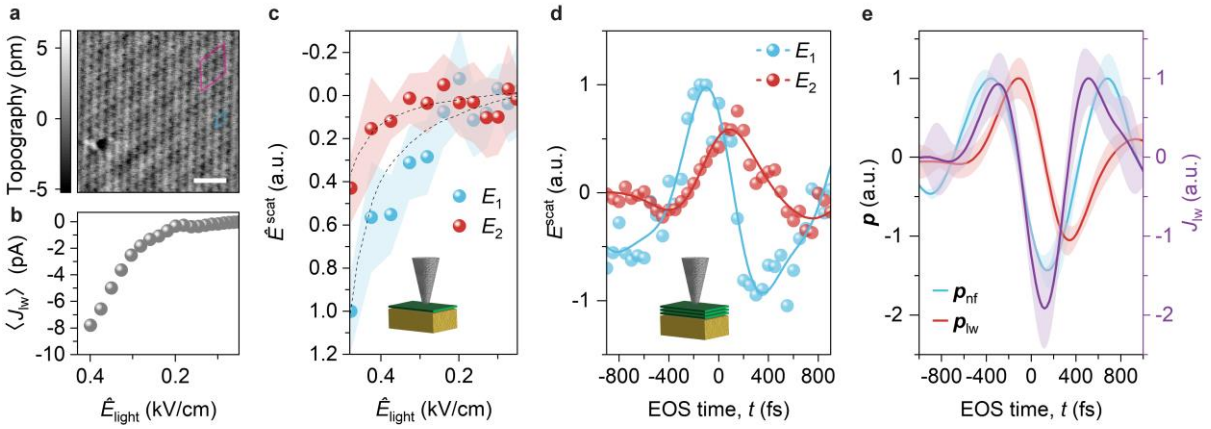


Figure 5 | Subcycle NOTE spectroscopy of WSe_2. a, Constant-current STM image (V = 1 V, I_{set} = 100 pA) of the surface of monolayer WSe_2 exfoliated directly onto an Au(111) surface. Unit cells of the atomic WSe_2 sublattice and moiré superlattice are highlighted in blue and pink, respectively. Scale bar: 1 nm. **b,** Onset of the time-integrated THz-driven tunnelling currents $\langle J_{lw} \rangle$ on monolayer WSe_2 as a function of the peak THz field strength ($\hat{E}_{light}$), showing the onset of tunnelling from the WSe_2 valence band edge. **c,** Subcycle detection of tunnelling currents for increasing $\hat{E}_{light}$, measured at the peak of the EOS transient for the first ($\hat{E}_1^{scat}$, blue) and second ($\hat{E}_2^{scat}$, red) harmonic of the tip tapping frequency ($A = 250$ pm). For the highest field strength shown, on average ~20 electrons were rectified per pulse. The steady-state conductance is overlayed as a dashed black line as guide to the eye, highlighting the band-edge like onset behaviour observed in the NOTE signal. **d,** Scattered near fields sampled on trilayer WSe_2 when the tip is away from the tunnelling region (blue circles) and the NOTE signal when electrons are allowed to tunnel (red circles). Here, ~24 electrons per pulse are being rectified for the tip in the tunnelling region on average. **e,** The near-field response at the tip apex (blue curve), the NOTE dipole (red) and the retrieved ultrafast tunnelling currents, sampled in real time (purple). All error bars in this figure represent one standard deviation.

Methods

Ultra-stable low-temperature near-field microscope. The setup is based on a UNISOKU USM-1400 ultra-high vacuum low-temperature scanning probe microscope, with custom high numerical aperture parabolic mirrors (NA ~0.4) for focusing and collecting light to and from the tip. For qPlus AFM/STM, a tungsten wire is mounted to the prong of the qPlus sensor[45], followed by electrochemical etching. Before loading the sensor assembly into the microscope, tungsten oxides are removed using concentrated hydrofluoric acid.

Optical picoscopy and lightwave STM measurements. We use a high-repetition rate ytterbium fibre amplifier (Satsuma HP³ by Amplitude; centre wavelength, 1030 nm; pulse duration, 260 fs; pulse repetition rate, 13.7 MHz). The output is used to generate phase-locked THz pulses (Extended Data Fig. 9) using optical rectification in a lithium niobate crystal configured for tilted-pulse-front generation. The strength of the THz field is tuned with a pair of wire-grid polarizers. The pulses are then focused onto the tip using the microscope's first parabolic mirror. For NOTE measurements, near fields scattered from the tip are collected using the second parabolic mirror in the microscope, and subsequently detected using electro-optic sampling in a 2-mm thick gallium phosphide (110) crystal (GaP), where the femtosecond gate pulse is provided by the laser fundamental. A lock-in amplifier (Zurich Instruments HF2LI) is used to measure the scattered near fields E_i^{scat} at the i^{th} harmonic of the tip tapping frequency (~20 kHz).

Using qPlus sensors allows us to extract both DC and time-averaged lightwave-driven tunnelling currents, either with an oscillating tip during near-field measurements (Fig. 2b, Fig. 4c,d), or in a more conventional lightwave STM configuration with a static tip (Fig. 5b). For the lightwave STM mode of operation, we modulate the THz pulse train with an optical chopper at a few hundred Hertz. The time-averaged lightwave tunnelling current $\langle J_{\mathrm{lw}} \rangle$ is amplified using a transimpedance preamplifier (Femto DLPCA-200) and subsequently demodulated at the chopper frequency using a lock-in amplifier. For the case where the tip is oscillating with an amplitude $A > 0$ (for simultaneous measurement of E^{scat}) the measurement configuration of $\langle J_{\mathrm{lw}} \rangle$ remains identical. Yet, the tapping motion of the tip modifies the absolute strength of $\langle J_{\mathrm{lw}} \rangle$ since current flows only during a fraction of the mechanical tapping cycle.

As the tip tapping frequency (~20 kHz) is far larger than the bandwidth of the transimpedance preamplifier (–3 dB @ 1.1 kHz), we measure the currents integrated over the tapping cycle. For a tip tapping amplitude of 250 pm, this results in an approximately six-fold reduction in the measured value of $\langle J_{\mathrm{lw}} \rangle$ compared to the current at the shortest distance in the tapping cycle[45]. We do not apply this factor in Fig. 2b and Fig. 4c,d to minimise the number of data processing steps. For recording $\langle J_{\mathrm{lw}} \rangle$ as a function of the position on the sample (Fig. 4c,d), we scan in quasi constant height mode with the AFM feedback set to maintain a constant frequency offset, exploiting the fact that the AFM is not sensitive to certain gold defects (Fig. 4b). We note that in Fig. 4d, $\langle J_{\mathrm{lw}} \rangle$ is measured without a chopper to improve our experimental sensitivity to the NOTE signal. This configuration remains a valid measure of $\langle J_{\mathrm{lw}} \rangle$ in this case as we separately confirmed that no meaningful DC tunnelling current is flowing during the line scans.

The polarity of the lightwave bias induced by the peak of the THz pulse was verified using the WSe_2/Au sample in Fig. 5b. For small THz fields, $\langle J_{\mathrm{lw}} \rangle$ is near-zero and sets in only when the peak bias reaches either the conduction or valence band of WSe_2. As the Fermi level is close to the centre of the gap, leading and trailing peaks of the opposite polarity cannot yield substantial signal, and so the negative sign of the rectified current (Fig. 5b) proves that the peak voltage is also negative with respect to the grounded tip.

Sample preparation. The experiments were performed with the following samples: single crystal Au(111) (Figs. 1 and 2), Au/mica (Fig. 4) and WSe_2/Au/mica (Fig. 5)**.** The atomically flat Au(111) surface with herringbone reconstruction was prepared by several sputter-anneal cycles on a precision-cut Au single crystal followed by quenching at 550°C. Au/mica samples were prepared by thermal evaporation of Au on freshly cleaved mica in a high-vacuum chamber. Atomic flatness of the Au surface was achieved in the same way as for the single crystal sample. WSe_2/Au/mica samples were fabricated by stamping the freshly cleaved bulk WSe_2 onto an atomically flat Au/mica sample in the load-lock chamber of the microscope, which was purged with nitrogen gas. This process yields predominantly monolayer WSe_2 but in some areas, native bilayers and trilayers emerge, which can be distinguished in an optical microscope and by tunnelling spectra. The samples were subsequently annealed at 350°C to

ensure compatibility with UHV conditions. Extended Data Fig. 8 shows the DC conductance of the monolayer and trilayer, which show the expected onsets of conduction and valence bands. In the monolayer, a small but finite conductance within the bandgap originates from hybridized chalcogen-gold states[51].

Dependence of the scattered near fields on the demodulation order. As discussed in the main text, the scattered near fields consist of both a classical near field as well as a non-classical contribution of light emission by ultrafast tunnelling currents. As the nonlinear dependence of the former is much weaker, their relative harmonic strengths evolve differently when the signal is demodulated at higher orders of the tip tapping frequency. For higher tip-sample separations, where the classical near-field contribution dominates the signal, the measured amplitude drops much more rapidly for increased demodulation orders than for smaller distances, where the current induced emission prevails (Fig. 1c and Extended Data Fig. 1).

Modelling the current-induced emission of THz radiation. For a qualitative understanding of the salient features of NOTE microscopy, we exploit a popular model used in near-field microscopy – the point dipole model (PDM)[3]. We approximate the near fields at the tip apex and the emission of lightwave tunnelling currents as the superposition of two electric dipole moments: a near-field dipole $\boldsymbol{p}_{\mathrm{nf}}$ representing the polarisation of the tip apex in the presence of the external THz field E_{light}, and a second dipole $\boldsymbol{p}_{\mathrm{lw}}$, driven by the ultrafast tunnelling current J_{lw} between tip and sample (Fig. 3a,b).

The near-field dipole moment $\boldsymbol{p}_{\mathrm{nf}}$ is calculated using the PDM, assuming a tip radius of curvature of 50 nm, and a realistic negative relative permittivity for both the tip and sample ($\epsilon_r \approx -10^6$). As the PDM does not incorporate field enhancement, we assume a peak external field strength of $\sim$10 MV cm^{-1} to obtain a realistic peak near-field amplitude.

To model $\boldsymbol{p}_{\mathrm{lw}}$, we first calculate the tunnelling currents induced by the electric field at the tip apex. We assume these currents to instantaneously follow the bias voltage induced by E_{light}, and use a nonlinear conductance of the tip-sample junction (Extended Data Fig. 4a). Following charge conservation, $\boldsymbol{p}_{\mathrm{lw}}$ is formed by instantaneous charging of the tip apex – the time integral of J_{lw} – across a distance given by the excursion of electrons in the tip and sample during the THz pulse (here, 40 nm,

see next section below). We then propagate the time-dependent dipoles $\boldsymbol{p}_{\mathrm{nf}}$ and $\boldsymbol{p}_{\mathrm{lw}}$ to the far field using a numerically simulated tip transfer function (Extended Data Fig. 6) and the response function of our electro-optic detection (Extended Data Fig. 4b). Figure 2c shows the temporal evolution of the propagated signals, where $\boldsymbol{p}_{\mathrm{lw}}$ (red dashed line) exhibits the characteristic phase delay of $\pi/2$ with respect to $\boldsymbol{p}_{\mathrm{nf}}$ (blue dashed line).

Estimating the current-induced dipole $\boldsymbol{p}_{\mathrm{lw}}$. To estimate a realistic strength of the current-induced dipole, we use an analytical model to obtain a spatial distribution of charges within the tip apex (Extended Data Fig. 5). We approximate the tip apex with a sphere of radius *r*. Since the tunnelling probability decays exponentially, it is enough to consider the case of the sphere in close proximity ($d \ll r$) to the metallic plane. Under this condition we solve a model of self-consistent mirror images[52]. To obtain an analytical solution, we represent the metallic plane by a sphere of extremely large radius $R \gg r$. As the maximum strength of $\boldsymbol{p}_{\mathrm{lw}}$ occurs at the moment when the external driving field – and near-field dipole $\boldsymbol{p}_{\mathrm{nf}}$ – crosses zero (see delay time t_2 in Fig. 3 of main text) we do not need to include an external driving field into our electrostatic model. The resultant charge distribution is concentrated very close to the apex (Extended Data Fig. 5) and can be written as $\sigma(\theta) = \sigma_{\mathrm{u}} + \sigma_{\mathrm{n}}(\theta)$, where θ is the angle shown in Extended Data Fig. 5a, σ_{u} and σ_{n} are uniform and non-uniform contributions respectively. The current-induced dipole $\boldsymbol{p}_{\mathrm{lw}}$ is formed by: (i) the dipole between σ_{u}, which is equivalent to a point charge $4\pi r^2\sigma_{\mathrm{u}}$ at the sphere centre, and its mirror image, $p_{\mathrm{lw}}^{(1)} = 8\pi r^2(r+d)\sigma_{\mathrm{u}}$, as well as (ii) the dipole of the sphere formed by a non-uniform charge distribution, $p_{\mathrm{lw}}^{(2)} = 2\pi r^3 \int_0^{\pi} \sigma_{\mathrm{n}}(\theta) \sin\theta \cos\theta \, d\theta$, and (iii) the mirror image of $p_{\mathrm{lw}}^{(2)}$ in the plane.

This inherently electrostatic solution is limited by the assumption that the charge forms the calculated distribution instantaneously. In the real tip however, the tunnelling of lightwave-driven charges is constrained by the tunnelling process to the frontier atoms of the tip apex, meaning redistribution of charges after tunnelling plays an important role. A rough estimate is to simply consider the region with highest field enhancement – the tip apex radius (~50 nm). Outside of this region, charge redistribution will be significantly slower. A more accurate estimate can be reached by considering the

drift velocity. We consider that charge transfer has occurred at t_1 (Fig. 3) only, and the charge redistributes through the tip during the next quarter-cycle of the THz pulse. For a centre frequency of 1 THz, using a mobility of $10\ \mathrm{cm^2\ V^{-1}\ s^{-1}}$ and a field strength of $0.1\ \mathrm{MV\ cm^{-1}}$ (1 V voltage drop over $2r = 100$ nm distance) acting on an electron in the tip, we obtain a propagation distance of ~2.5 nm, suggesting the electrostatic model is indeed applicable for small tip apex radii. This simple approximation though cannot be extended to atomic scales and stronger fields, where quantum effects start to play a role and self-consistent microscopic treatment of light-matter interactions is necessary[53].

Ab initio TDDFT simulations. For a consistent microscopic understanding of the subcycle tunnelling dynamics, we performed real-time TDDFT calculations with the CP2K code[54]. We model the metallic substrate and the metallic tip with Na atoms instead of W and Au atoms following previous work[55]. The use of Na significantly reduces the computational cost enabling simulations on long time scales. We model the tip by a finite regular square pyramid (height: 1.93 nm, base: 3.00 nm x 3.00 nm, 349 Na atoms, bcc structure, lattice constant: 0.428 nm) and we model the substrate by a finite square box (3.43 nm x 3.43 nm x 1.71 nm, 661 atoms, bcc structure, lattice constant: 0.428 nm). The distance between the lowest tip atom and the central, top substrate atom was varied between 9 Å and 14 Å. The simulation volume for representing the charge density and the electrostatic potential is a square box of size 4.2 nm x 4.2 nm x 8.0 nm. For the Na atoms, a dual-space, norm-conserving Goedecker-Teter-Hutter pseudopotential[56] treating only the 3s electron explicitly is used. Kohn-Sham orbitals are expanded in a Gaussian DZVP-MOLOPT basis set[57]. We employ the adiabatic Perdew-Burke-Ernzerhof exchange correlation potential[58] and propagate the time-dependent Kohn-Sham equations using an exponential midpoint propagator with a time step of 0.2 fs. The external electric field pulse was represented by a 6 THz wave (period: 166.7 fs) with a Gaussian envelope (FWHM 235.5 fs) with a peak electric field of $1\ \mathrm{V\ nm^{-1}}$. The total simulation time is 1000 fs. Our TDDFT calculations do not include scattering and some aspects of dynamical screening, such as many-body effects and quasiparticle (optical phonons, plasmons etc.) emission, which were suggested to affect tunnelling[59] and, hence, are also encoded in NOTE.

Numerical simulation of the tip-transfer function. To simulate the frequency dependent transfer function from the far field to the near field, we solved Maxwell's equations in the frequency domain with COMSOL Multiphysics, in a similar way to the transfer function calculated in ref. 55. We positioned the tip 1 nm above a flat gold sample in a 2D simulation geometry (0.9 mm x 1.2 mm). By using a scattering boundary condition on one side, we coupled in the THz radiation (Extended Data Fig. 6a). All other boundaries were equipped with perfectly matched layers. To retrieve the complex valued transfer function (Extended Data Fig. 6b), we evaluate the electric field amplitude and phase in the centre of the tunnelling junction and compare it with a second simulation where the tip was replaced with vacuum. We find a field enhancement showing an amplitude response following f^{-1}, and a spectrally uniform phase shift of ~ –0.5π.

Time resolution in NOTE versus LW-STM. The physical quantity measured in LW-STM is given by the rectified charge transferred by the time-integrated action of the entire light pulse. This charge transfer is measured by slow time integrating current detectors. Nonetheless, pump-probe experiments using two THz pulses[38,39,55], or combining optical and THz pulses[44] can generate transient responses, which on first inspection may appear similar to NOTE signals. However, these transient responses are more akin to nonlinear auto- or cross-correlations than subcycle temporal sampling. The dynamics in the junction of a LW-STM are extremely nonequilibrium, hosting a complex interplay of competing nonlinearities which depend on the exact conditions of tunnelling, as well as dynamics occurring within the sample itself. Hence, reconstruction of ultrafast tunnelling currents is only possible by making model assumptions about the subcycle field response of the junction, which in turn are known only under the most well-defined conditions[55].

Conversely, NOTE samples lightwave tunnelling currents directly in time using electro-optic sampling. This is only possible because EOS relies on an instantaneous and well-defined $\chi^{(2)}$ optical nonlinearity for frequency downconversion. Unlike LW-STM, this downconversion is also seperated from the dynamics occurring within the tip-sample junction. It is these conditions which illustrate the key advantage of NOTE and enable true subcycle sampling of ultrafast tunnelling for the first time.

References

51. Sørensen, S. G., Füchtbauer, H. G., Tuxen, A. K., Walton, A. S. & Lauritsen, J. V. Structure and electronic properties of in situ synthesized single-layer MoS_2 on a gold surface. *ACS Nano* **8**, 6788–6796 (2014).
52. Smythe, W. R. *Static and Dynamic Electricity*. (New York, McGraw-Hill, 1950).
53. Jestädt, R., Ruggenthaler, M., Oliveira, M. J. T., Rubio, A. & Appel, H. Light-matter interactions within the Ehrenfest–Maxwell–Pauli–Kohn–Sham framework: fundamentals, implementation, and nano-optical applications. *Adv. Phys.* **68**, 225–333 (2019).
54. Kühne, T. D. et al. CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structure calculations. *J. Chem. Phys.* **152**, 194103 (2020).
55. Peller, D. et al. Quantitative sampling of atomic-scale electromagnetic waveforms. *Nat. Photon.* **15**, 143–147 (2021).
56. Goedecker, S., Teter, M. & Hutter, J. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B* **54**, 1703-1710 (1996).
57. VandeVondele, J. & Hutter, J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.* **127**, 114105 (2007).
58. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865-3868 (1996).
59. Persson, B. N. J. & Baratoff, A. Self-consistent dynamic image potential in tunneling. *Phys. Rev. B* **38**, 9616–9627 (1988).

Acknowledgements. We thank Carlos Bustamante, Franco Bonafé, Angel Rubio, and Franz J. Gießibl for helpful discussions and Martin Furthmeier for technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID, 314695032 – SFB 1277 (Subprojects A05 and B02), major instrumentation grant INST 89/505-1 and through research grants HU1598/8 and HU1598/9. The authors gratefully acknowledge the computing time provided to them on the high performance computers Noctua 2 at the NHR Center PC2. These are funded by the Federal Ministry of Education and Research and the state governments participating on the basis of the resolutions of the GWK for the national high-performance computing at universities (www.nhr-verein.de/unsere-partner).

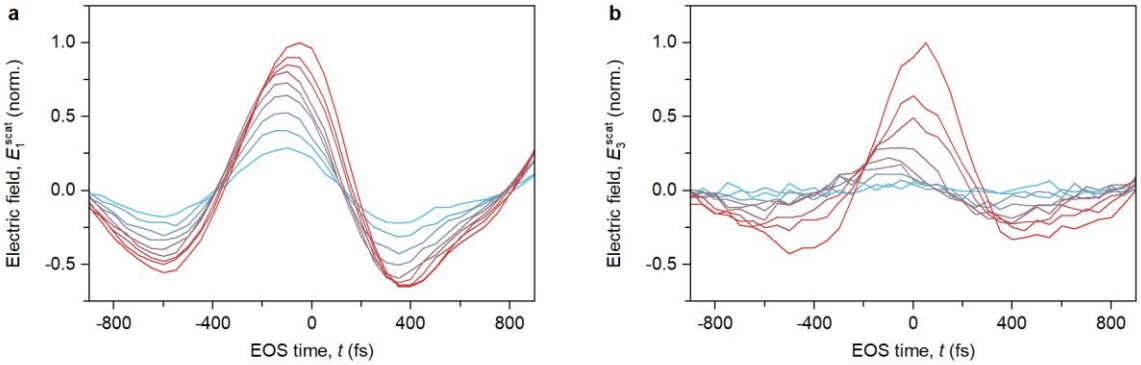
Author contributions. T.S., M.A.H, Y.A.G. and R.H. conceived the study. T.S., J.H., F. Schiegl, F. Sandner, P.M., V.B., S.L., M.A.H. and Y.A.G. performed the experiments and analysed the data. T.S. and Y.A.G. fabricated the samples. T.S., J.H., F. Schiegl, F. Sandner, P.M., V.B., M.Z., S.N., S.L., J.R., J.W., M.A.H., Y.A.G. and R.H. contributed to the discussions of the experimental results. T.S., J.H., M.A.H., Y.A.G., and R.H. developed and computed the dipole emission model. J.W. developed and computed the *ab initio* quantum simulations. The manuscript was written by T.S., Y.A.G. and R.H. with input and contributions from all authors.

Competing interests. The authors declare no competing interests.

Materials & Correspondence. Correspondence and requests for materials should be addressed to J. Wilhelm, M. A. Huber and Y. A. Gerasimenko.

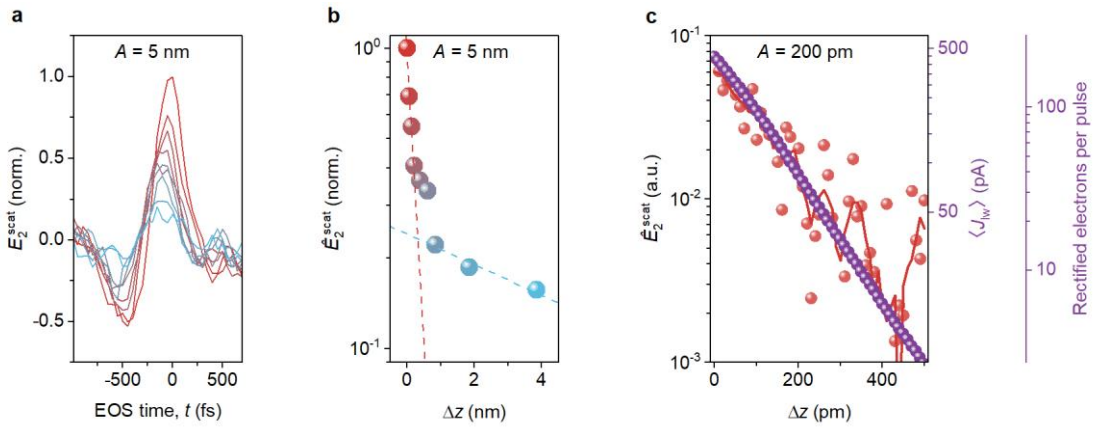
Data availability

All the plotted data are available as source data with this manuscript. All further data are available from the corresponding authors.



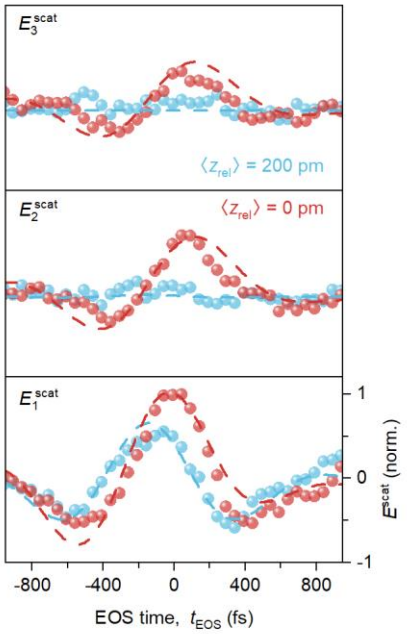
577

578 **Extended Data Fig. 1 | Tip-sample distance dependence of the scattered transients for the first**
579 **and third harmonic.** First (**a**) and third (**b**) harmonic, E_1^{scat} and E_3^{scat}, of the scattered electric fields,
580 measured simultaneously with the transients shown in Fig. 1b (peak far-field strength $\hat{E}_{\mathrm{light}} =$
581 1.3 kV cm^{-1}). For the closest tip-sample distance ~900 electrons are rectified per pulse on average. The
582 colours correspond to the distances shown in Fig. 1c. The strong underlying nonlinearity is highlighted
583 by the increased difference in signal strength for different tip-sample separations when going to higher
584 demodulation orders.



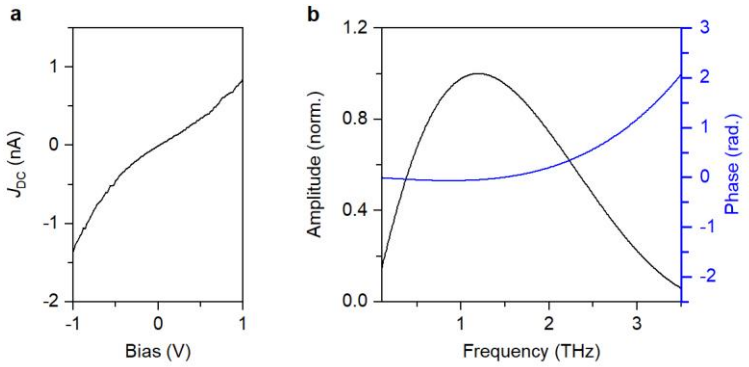
585

586 **Extended Data Fig. 2 | Decay of the NOTE signal at large tip tapping amplitudes. a,** Electro-
587 optically detected scattered THz transients demodulated at the second harmonic of the tip tapping
588 frequency (E_2^{scat}) with A = 5 nm and $\hat{E}_{\mathrm{light}} = 1.1$ kV cm^{-1}. For the transient acquired at the closest tip-
589 sample distance $\langle J_{\mathrm{lw}} \rangle = -180$ pA, on average ~80 rectified electrons per pulse. **b,** Peak of the THz
590 transient ($\hat{E}_2^{\mathrm{scat}}$) at t = 0 fs, measured at a tip tapping amplitude A = 5 nm for increasing tip-sample
591 separation Δz. **c,** Logarithmic-scale plot of the data shown in Fig. 2b: $\hat{E}_2^{\mathrm{scat}}$ measured at the second
592 harmonic of the tip tapping frequency (A = 200 pm, $\hat{E}_{\mathrm{light}} = 0.8$ kV cm^{-1}), alongside the time-
593 integrated lightwave tunnelling current $\langle J_{\mathrm{lw}} \rangle$ measured for increasing tip-sample separation (on average
594 ~200 electrons per pulse are rectified for the closest tip-sample distance).



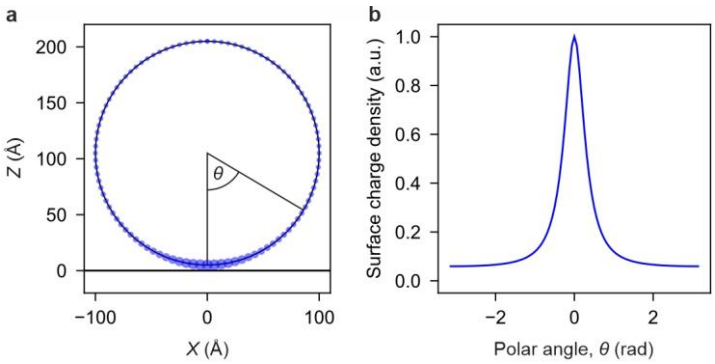
595

596 **Extended Data Fig. 3 | NOTE transients sampled on Au(111).** Electro-optically detected scattered
597 THz transients measured at the first three harmonics of the tip tapping frequency (A = 1 nm). In each
598 panel, the red curve shows the case of an approached tip, where tunnelling currents can flow ($\langle J_{lw} \rangle =$
599 -300 pA, on average ~140 rectified electrons per pulse, $\Delta\nu = -4.8$ Hz, $\hat{E}_{light} = 1.3$ kV cm^{-1}). The
600 blue curve shows the case where the tip is retracted away from the tunnelling barrier.

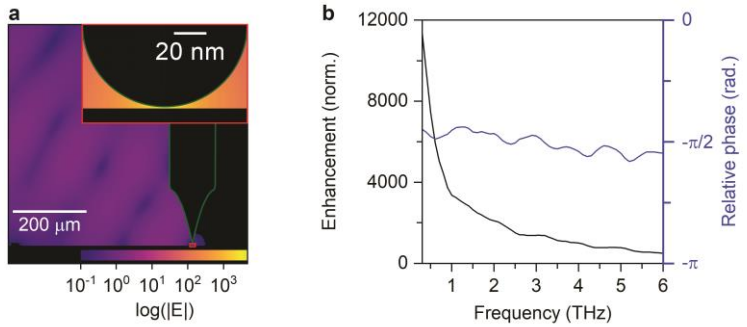


601

602 **Extended Data Fig. 4 | Steady-state tunnelling spectrum of Au(111) and optical detection**
603 **bandwidth. a,** Experimental tunnelling spectrum, measured with steady-state STM on the surface of
604 Au(111). **b,** Amplitude and phase of the response of the EOS detection, including the detector response
605 of GaP and focusing conditions of the parabolic mirrors.

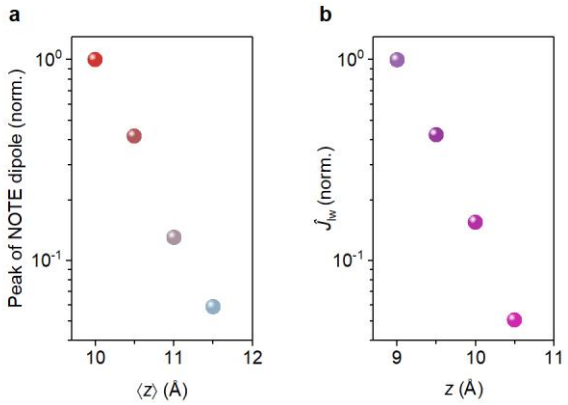
606 

607 **Extended Data Fig. 5 | Distribution of tunnelled charges at the tip apex. a,** Distribution of charges
608 on the surface of a sphere with radius r = 100 Å at distance d = 5 Å from a large sphere with radius R.
609 The local strength of the surface charge is indicated by the diameter of the blue circles. **b,** Surface
610 charge density around the surface of the sphere as a function of the polar angle θ shown in **a**.

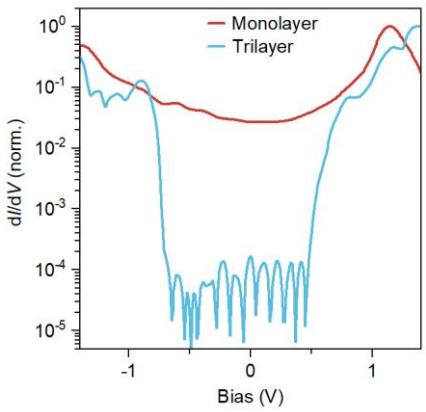


611

612 **Extended Data Fig. 6 | Numerical model of the tip transfer function**. **a,** Spatial distribution of the
613 electric field magnitude on a cross-section of the STM tip. The inset shows the near fields in the vicinity
614 of the nanoscale apex. **b,** Magnitude and phase of the field enhancement at the tip apex as a function of
615 frequency.

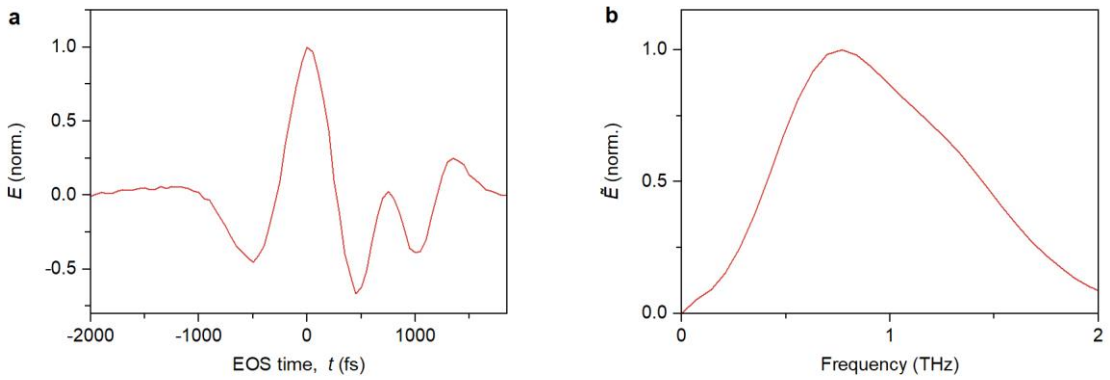


Extended Data Fig. 7 | Atomic-scale decay of the NOTE signal in the quantum simulation. **a,** Amplitude of the simulated NOTE dipoles (A = 1 Å) evaluated at $t = 0$ fs, for increasing average tip-sample separation $\langle z \rangle$. **b,** Simulated peak of the ultrafast tunnelling current ($\hat{J}_{\mathrm{lw}}$) for a static tip as a function of the tip-sample separation z. The angstrom-scale decay of the NOTE dipole in panel **a** closely follows that of the ultrafast tunnelling current.



622

623 **Extended Data Fig. 8 | Bias spectroscopy of mono- and trilayer WSe_2 on Au(111).** Normalized
624 steady-state differential conductance of WSe_2 mono- and trilayer on Au(111). In the monolayer, some
625 electrons tunnel directly from tip to Au(111), resulting in a measurable current within the gap. The
626 additional tunnelling barriers introduced by the trilayer remove this current pathway.



627

628 **Extended Data Fig. 9 | Far-field THz transient and spectrum. a,** Incident THz transient sampled in

629 the far field. **b,** Corresponding spectrum of the incident THz field.

Figure	Peak far-field strength $\hat{E}_{\mathrm{light}}$
1	1.3 kV cm^{-1}
2	0.8 kV cm^{-1}
4c	0.4 kV cm^{-1}
4d,e	0.5 kV cm^{-1}
5d	0.3 kV cm^{-1}

630

631 **Extended Data Table 1 | Experimental peak field strengths in the far field.**