

Direction-Resolved Nanoscale Optical Imaging with Near-Nanometer Resolution by Emerging Infrared Torsional Force Microscopy

Corresponding Author: Professor Min Ouyang

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The manuscript presents a transformative approach to infrared nanospectroscopy through the development of Infrared Torsional Force Microscopy (TFM-IR). By combining the torsional dynamics of an atomic force microscopy (AFM) cantilever with a nonlinear frequency-mixing scheme, the authors successfully mapped both in- and out-of-plane photothermal signals, overcoming the "blind spot" of traditional AFM-IR: the inability to capture in-plane (lateral) dynamics. The agreement between the experimental anisotropy-sensitive mapping and the theoretical calculation is striking.

The achievement of ~1 nm spatial resolution is truly remarkable, surpassing the standard 10–20 nm limits of s-SNOM. The application to twisted bilayer graphene (TBG) is a masterstroke. This leap in resolution facilitates site-resolved spectroscopy, enabled the direct observation of physics within a single moiré unit cell and offered new insights into how local strain and stacking orders influence light-matter interactions in quantum systems.

Overall, this is a seminal contribution to the fields of scanning probe microscopy and nanoscale optics. The paper is well-written, though a few technical points (detailed below) require further clarification. I fully support its publication in Nature Communications once these minor concerns are addressed.

1. The schematic of the AFM in Fig. 1b is confusing. Is it a perspective view from the end of the cantilever? I understand torsion angle θ is defined as the angle between the planes of the untwisted and twisted cantilever, but the authors should consider revising the figure and/or adding clarification in the text.

2. In Fig 2, for the TFM images of nanobubbles, both the in-plane and out-of-plane signals appear asymmetric. Is this asymmetry intrinsic to the nanobubble geometry, or does it arise from other factors (e.g., measurement or tip effects)?

3. In the wavenumber-dependent TFM-IR measurements of TBG, is there any observable dependence on the tip used for imaging?

4. The exceptionally high spatial resolution of the TFM-IR imaging is impressive and holds great promise for studying a broad range of emerging quantum materials. Analogous to the anisotropy-sensitive mapping of nanobubbles in mica, are there any observable differences between the f_{in} and f_{out} imaging channels in TBG moiré lattice? Although this may be experimentally challenging, the ability to resolve optical anisotropy in nanoscale quantum materials, would immediately open exciting new avenues of research with rich physics building on this pioneering work.

Reviewer #2

(Remarks to the Author)

This manuscript introduces infrared torsional force microscopy (TFM-IR), a new type of AFM-based IR imaging technique that detects cantilever torsional resonant oscillation as a result of nonlinear mechanical frequency mixing from carefully set laser excitation repetition rate. The authors derived a formalism to resolve both in-plane and out-of-plane photothermal

responses and provided experimental data to support its feasibility. Several representative samples are used for demonstrations: hexagonal boron, mica, and twisted bilayer graphene.

The novelty of this work is that it proposed a new method for AFM-IR microscopy. The TFM-IR provides different types of information and is sensitive to in-plane anisotropy. The authors presented a clear theory foundation for using certain combination of repetition rate to access in-plane or out-of-plane components.

The drawback of the work is that the images of the work are quite noisy and is not on par with typical AFM-IR images the community often see. Also, the torsional force mode is typically in AFM contact mode with a large vertical force setpoint (~ 20 nN). This measurement condition inherits the limitation of the contact mode. If the sample are susceptible to destruction/displacement due to tip dragging on it, then this method would not be suitable.

Critically, the bold claim of near-nanometer (~1 nm) spatial resolution, highlighted in twisted bilayer graphene imaging (e.g., Fig. 4), appears overstated. The provided data, while showing high spatial contrast, is likely due a certain mechanism of cross talk between photothermal signal and mechanical transduction pathways. If the authors insist ~1 nm spatial resolution, they should provide a level of understanding why such a resolution is generally possible in the TFM-IR and why it is not seen in various kind of AFM-IR techniques. This component is important, because it would mean if such high spatial resolution is generally present for TFM-IR or it is only for the type of bilayer graphene sample measured. After all, bilayer graphene's Moire patterns that can already be revealed by regular lateral force microscopy without IR laser.

Figure 5d's signal appears to be very strong (indicated from scale bar), but the noise level remains very high. This is strange. The images of Figure 5 do not show much difference across different IR energies, given the high noise background. If the author can articulate why there should be an IR wavelength dependent difference in the Moire patterns in a physical picture with beyond empirical understanding, that would be more convincing.

Some minor comments:

More experimental details should be included in either method section or supplementary information. For example, I could not even find which AFM the authors used in their work. If it is a custom built AFM, the author should say so.

In the introduction, the authors says:

"This allows the optical signal intensity to be tuned and optimized via the torsional drive amplitude without disturbing the tip-sample feedback, offering a key advantage over conventional AFM-IR and sSNOM."

In fact, sSNOM's signal channel and AFM feedback are naturally decoupled. Applying a CW laser in sSNOM does not perturb the tapping mode feedback in sSNOM. For AFM-IR techniques, except for the original contact mode AFM-IR, new embodiment such as PiFM, tapping mode AFM-IR, and PFIR all separate the AFM feedback from the photothermal signal read out channels. This statement from the author is not true.

The author should cite a few previous works on AFM-IR revealing h-BN's polariton fringes. A few more references on the combinatory heterodyne schemes in AFM-IR/PiFM/surface sensitive AFM-IR should also be referenced.

Reviewer #3

(Remarks to the Author)

The authors have extended the technique of Torsional Force Microscopy (TFM) by adding optical probing in the mid-infrared, thus establishing TFM-IR, and claim the ability to map strain in nanobubbles and wavelength-dependent features within a moire unit cell in twisted samples with a resolution approaching 1 nm. The technique is interesting, but there are some important issues with this work that need to be addressed before this can be published in any journal. Here are some of the main ones:

1. From the introduction and abstract it seems as if the authors developed a new technique from scratch, whereas in reality the TFM part already existed and they added optical excitation to this. They should be more clear about this.
2. The authors list disadvantages of sSNOM and AFM-IR, such as unavoidable tip induced perturbations and sophisticated apparatuses. I don't see how that's any different with TFM-IR, which in addition requires a laser with tunable repetition rate.
3. The authors claim a much better resolution than sSNOM, citing Refs. 53-55 to have a spatial resolution of several hundred nanometers. This seems incorrect, because at least Ref. 55 seems to have a sub-10 nm resolution.
4. The authors created the impression that they are studying twisted bilayer graphene, where one graphene layer is twisted w.r.t. the other layer, but it seems that they just transferred bilayer graphene onto hBN. So is it just the twist angle between bilayer graphene and hBN? The authors should be clear about this and provide sufficient detail. It would also be nice to verify the obtained twist angle by an independent measurement.
5. The authors claim a super-resolution approaching 1 nm, but their data in Figures 1-3 don't seem to support this. The data in Figures 4-5 is more convincing in this regard, but it's very disappointing that the reason for this resolution is not discussed even with some attempt of an explanation. Does it not have anything to do with the amplitude of the motion of the tip?
6. The descriptions of the figures are either lacking details (1a for example) or containing superfluous details (1c for

example).

7. The authors seem to show results for only one nanobubble, calling it a "typical" nanobubble, which created the expectation of seeing more similar data on other nanobubbles in the Supplement, but these are absent. This makes this an $N = 1$ experiment. Seeing more and similar data would make this work much more convincing.

8. The data in Figure 3 shows a dependence on excitation wavelength, but the explanation for the observations is not entirely convincing. And are there no artefacts when changing wavelength? Different collimation, focus position, etc? Here, data on more nanobubbles could strengthen the case again.

9. There is also not much explanation of the observed wavelength-dependent features within a moire unit cell.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

The authors have reasonably and satisfactorily addressed my comments and concerns in their response and the revised manuscript. I believe the manuscript is now suitable for publication in Nature Communications.

Reviewer #2

(Remarks to the Author)

The revision improved the clarity of this manuscript, and expanded the background information of this manuscript.

While I acknowledge the novelty of this work on method implementation and interpretation of signal original in their original submission, I am still not convinced that the spatial resolution is from the IR/spectroscopy rather than due to the mechanical readout of the torsional force. The additional example shown in the supplementary still lacks spectroscopic contrast, thus not convincing.

I would only agree to the author's claim of 1 nm only if they can measure a sample with compositional/spectroscopic heterogeneity not mechanical heterogeneous response. One test example would be a cellular surface that has intrinsic chemical heterogeneity, or a protein aggregate, like amyloid fibrils, under different IR excitation wavelength.

Reviewer #3

(Remarks to the Author)

The authors have addressed my main concerns. I do still see quite some typos, and would therefore advise a careful proofreading, in particular of the Supplement.

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)

The wavelength dependent TFM-IR images have addressed my concerns.

Although the original of these spatial signal contrast in Figure 5 is still a question, as a demonstration of method, it indeed means that the TFM-IR maintains good spatial resolution. I think this manuscript can be published.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.

The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Response to Reviewer #1:

“The manuscript presents a transformative approach to infrared nanospectroscopy through the development of Infrared Torsional Force Microscopy (TFM-IR). By combining the torsional dynamics of an atomic force microscopy (AFM) cantilever with a nonlinear frequency-mixing scheme, the authors successfully mapped both in- and out-of-plane photothermal signals, overcoming the “blind spot” of traditional AFM-IR: the inability to capture in-plane (lateral) dynamics. The agreement between the experimental anisotropy-sensitive mapping and the theoretical calculation is striking.

The achievement of ~1 nm spatial resolution is truly remarkable, surpassing the standard 10–20 nm limits of s-SNOM. The application to twisted bilayer graphene (TBG) is a masterstroke. This leap in resolution facilitates site-resolved spectroscopy, enabled the direct observation of physics within a single moiré unit cell and offered new insights into how local strain and stacking orders influence light-matter interactions in quantum systems.

Overall, this is a seminal contribution to the fields of scanning probe microscopy and nanoscale optics. The paper is well-written, though a few technical points (detailed below) require further clarification. I fully support its publication in Nature Communications once these minor concerns are addressed.”

Response: We thank the reviewer for taking time carefully reading the manuscript, for the positive assessment of our results, and for recommending publication of our work. Below, we address the points raised by the reviewer.

“1. The schematic of the AFM in Fig. 1b is confusing. Is it a perspective view from the end of the cantilever? I understand torsion angle θ is defined as the angle between the planes of the untwisted and twisted cantilever, but the authors should consider revising the figure and/or adding clarification in the text.”

Response: We thank the reviewer for the suggestion and apologize for the oversight. The bottom schematic of AFM tip in Fig.1b is indeed a perspective view. We have revised the corresponding figure caption and increased the font size of θ in the figure to clarify this point.

“2. In Fig 2, for the TFM images of nanobubbles, both the in-plane and out-of-plane signals appear asymmetric. Is this asymmetry intrinsic to the nanobubble geometry, or does it arise from other factors (e.g., measurement or tip effects)?”

Response: We thank the reviewer for this insightful question. We have performed multiple control experiments to safely exclude tip-related effects or measurement artefacts and attribute the observed asymmetry primarily to intrinsic variations in the nanobubble structure.

First, for the same nanobubble (e.g., Fig. 2), the asymmetric features differ between the in-plane and out-of-plane maps. If these asymmetries originated from tip-effects or instrumental artefacts, one would expect similar spatial patterns to appear in both channels. The distinct asymmetry observed in the two channels instead indicates sensitivity to different physical components of the local response.

Second, the systematic wavenumber-dependent measurements presented in Fig. 3 show a clear evolution of these asymmetric features with excitation energy. Such energy-dependent behavior is inconsistent with static artefacts (e.g., tip geometry or feedback conditions), and instead supports an intrinsic origin associated with local optical or mechanical properties of nanobubble.

Third, we find that these asymmetric features are robust against variations in experimental conditions for the same nanobubble, including changes in probe tips and imaging parameters. This reproducibility further suggests that they are not artefact related.

In addition, we have performed measurements on multiple nanobubbles and have included results from a different nanobubbles in the new Supplementary Fig. 9. While fine asymmetric and inhomogeneous features may vary among different nanobubbles, the key characteristics, such as the annular distribution and the characteristic spatial profiles of in-plane and out-of-plane responses, are consistently observed. This indicates that the asymmetry likely reflects local variations in individual nanobubble geometry, strain distribution, or underlying substrate interactions, superimposed on a common structural motif.

These observations underscore the importance of direction-resolved, high-spatial-resolution imaging as enabled by TFM-IR, which allows both universal features and structure-specific variations of individual nanobubbles to be resolved. We have included additional discussion on this topic in the revised main text.

“3. In the wavenumber-dependent TFM-IR measurements of TBG, is there any observable dependence on the tip used for imaging?”

Response: No, these are not tip-dependent features. Similar features are reproducible across different tips, provided they are of the same type (e.g., with the same spring constant).

“4. The exceptionally high spatial resolution of the TFM-IR imaging is impressive and holds great promise for studying a broad range of emerging quantum materials. Analogous to the anisotropy-sensitive mapping of nanobubbles in mica, are there any observable differences between the f_{in} and f_{out} imaging channels in TBG moiré lattice? Although this may be experimentally challenging, the ability to resolve optical anisotropy in nanoscale quantum materials, would immediately open exciting new avenues of research with rich physics building on this pioneering work.”

Response: We thank the reviewer for raising this insightful question and fully agree with the perspective that anisotropy-resolved imaging in moiré quantum materials could open exciting new directions.

In fact, we have performed preliminary measurements probing both in-plane (f_{in}) and out-of-plane (f_{out}) responses in TBG moiré lattices, which reveal clear indications of anisotropy-dependent optical contrast at the nanoscale. For the reviewer's reference, we provide representative unpublished results (see "Confidential Fig. 1"), which suggest that the two channels can exhibit distinct symmetry and spatial features for TBG moiré lattice, reflecting the underlying anisotropic optical response.

However, a comprehensive and systematic investigation of anisotropy-resolved optical properties in moiré lattices, particularly including energy dependence and quantitative interpretation, requires substantial additional experimental and theoretical analysis. We therefore consider this direction to be beyond the scope of the present work, which focuses on establishing the TFM-IR methodology and demonstrating its key capabilities, including direction-selective detection and near-nanometer spatial resolution. To maintain clarity and focus, we have chosen to present the out-of-plane (f_{out}) channel in the current manuscript to highlight the spatial resolution and energy-dependent imaging capability. We agree with the reviewer that extending this approach to fully resolve anisotropic optical responses in moiré systems represents a highly promising direction, and we are actively pursuing this in ongoing work.

We believe that the revised manuscript has been significantly strengthened by addressing the thoughtful points raised in the reviewer's report, and we sincerely thank the reviewer for their constructive comments and suggestions.

Response to Reviewer #2:

“This manuscript introduces infrared torsional force microscopy (TFM-IR), a new type of AFM-based IR image technique that detects cantilever torsional resonant oscillation as a result of nonlinear mechanical frequency mixing from carefully set laser excitation repetition rate. The authors derived a formalism to resolve both in-plane and out-of-plane photothermal responses and provided experimental data to support its feasibility. Several representative samples are used for demonstrations: hexagonal boron, mica, and twisted bilayer graphene.

The novelty of this work is that it proposed a new method for AFM-IR microscopy. The TFM-IR provides different types of information and is sensitive to in-plane anisotropy. The authors presented a clear theory foundation for using certain combination of repetition rate to access in-plane or out-of-plane components.”

Response: We thank the reviewer for taking the time to evaluate our work and for their positive comments on its novelty and results. We also appreciate the reviewer’s thoughtful questions on specific aspects of the study and welcome the opportunity to address and clarify these points.

“The drawback of the work is that the images of the work are quite noisy and is not on par with typical AFM-IR images the community often see. Also, the torsional force mode is typically in AFM contact mode with a large vertical force setpoint (~ 20 nN). This measurement condition inherits the limitation of the contact mode. If the sample are susceptible to destruction/displacement due to tip dragging on it, then this method would not be suitable.”

Response: We thank the reviewer for this thoughtful comment and the opportunity to clarify these important points.

First, because TFM-IR is a newly developed imaging modality, the full understanding and optimization of its imaging characteristics are still evolving. In the present work, we adopted feedback conditions similar to those used in prior TFM studies (particularly Ref. 11), to enable direct comparison of TFM image and to ensure stable operation in initial demonstrations on TFM-IR. Under these conditions, TFM-IR achieves sufficiently high spatial resolution to resolve nanoscale features that are not accessible with conventional approaches. To quantitatively assess image quality, we have performed a signal-to-noise ratio (SNR) analysis for the data in Figs. 4 and 5 across multiple wavenumbers and summarized the results in **Table** below. The averaged SNR is ~ 4.42 , which, while not yet fully optimized, is sufficient to clearly resolve the spatial features discussed in the manuscript.

We also emphasize that the apparent image quality depends strongly on the intrinsic infrared response of the sample. In particular, the IR signal from moiré graphene is relatively weak, as it arises primarily from indirect or symmetry-breaking mechanisms, in contrast to materials such as hexagonal boron nitride, which exhibit strong phonon resonances. For example, the TFM-IR image of hBN shown in Fig. 1c exhibits signal quality comparable to prior AFM-IR studies (e.g., Refs.

19, 20), supporting that the current implementation is not fundamentally limited by noise. In fact, our results from the hBN have revealed beating pattern that is caused by the interference of edge-launched and tip-launched phonon polaritons, while conventional tapping AFM-IR can only reveal edge-launched phonon polaritons, highlighting the sensitivity of current TFM-IR technique.

Table: Summary of the SNR analysis for the TFM-IR images presented in Figures 4 and 5.

Figure	Average Background (V)	RMS Background (V)	Average Signal Intensity (V)	SNR = (Avg Signal-Avg Background)/(RMS Background)
4a	0.832	0.205	1.800	4.72
5a	0.271	0.070	0.522	4.01
5b	0.261	0.061	0.535	4.49
5c	0.183	0.047	0.402	4.46
5d	0.879	0.211	1.809	4.41

Second, the visual appearance of increased “noise” in TFM-IR images reflects the fundamentally different detection mechanism compared to conventional AFM-IR. Standard contact or tapping-mode AFM-IR measures vertical photothermal expansion, effectively averaging the optical response over the tip-sample interaction area and acting as a spatial low pass filter; In contrast, TFM-IR detects the torsional response of the cantilever, which is driven by lateral force asymmetry and is therefore sensitive to spatial gradients in the photothermal response. As a result, uniform or slowly varying signals are suppressed, while localized variations are enhanced. This gradient-sensitive detection increases sensitivity to high spatial frequency components, which can give rise to a visually more textured appearance, but in fact reflects enhanced detection of nanoscale heterogeneity rather than random noise.

Third, we agree with the reviewer that the current implementation operates under conditions close to contact mode. However, we observed no evidence of surface damage or sample displacement during repeated scans on all materials systems studied in the current work. Importantly, contact-mode operation is not an inherent requirement of TFM-IR. As described in the manuscript, the torsional excitation and detection are independent of the feedback mechanism. Therefore, future implementations can readily be extended to reduced-force or intermittent-contact modes, which will further broaden the applicability of the technique to more delicate samples. We have revised the Discussion section in the main text to incorporate these clarifications.

“Critically, the bold claim of near-nanometer (~1 nm) spatial resolution, highlighted in twisted bilayer graphene imaging (e.g., Fig. 4), appears overstated. The provided data, while showing high spatial contrast, is likely due a certain mechanism of cross talk between photothermal signal and mechanical transduction pathways. If the authors insist ~1 nm spatial resolution, they should provide a level of understanding why such a resolution is generally possible in the TFM-

IR and why it is not seen in various kind of AFM-IR techniques. This component is important, because it would mean if such high spatial resolution is generally present for TFM-IR or it is only for the type of bilayer graphene sample measured. After all, bilayer graphene's Moire patterns that can already be revealed by regular lateral force microscopy without IR laser.”

Response: We thank the reviewer for raising this important point. We have carefully examined the origin of the observed high spatial resolution and provide additional clarification and supporting evidence below.

First, we have conducted several control experiments to rule out the possibility that the observed contrast arises from mechanical cross-talk. The anisotropic in-plane and out-of-plane signals observed in both mica nanobubbles (e.g., Figs. 2 and 3) and moiré lattice measurements (see “Confidential Data 1”) are accessed through distinct nonlinear frequency-mixing conditions. Switching between these operating conditions gives systematic and reproducible changes in image contrast, contradicting a fixed mechanical cross-talk pathway hypothesis.

Second, the measured signals exhibit clear wavenumber dependence (e.g., Figs. 3 and 5). We have also performed control experiments by comparing TFM-IR under condition of off-resonant optical excitation. A new Supplementary Fig. 10 is added in the revised Supplementary Information, showing TFM-IR imaging of same moiré lattice presented in Fig.4a but under off-resonant optical excitation (1607 cm^{-1}), where the optical response is strongly suppressed and the resulting image becomes featureless. Same control experiment has also been conducted in the study of mica nanobubble (see new Supplementary Fig. 9). This spectral selectivity is inconsistent with a mechanical cross-talk mechanism, which would not be expected to depend on excitation wavenumber. Together, these results strongly support that the detected signal is of photothermal origin rather than arising from mechanical transduction artifacts.

Regarding the origin of the near-nanometer spatial resolution, we provide the following qualitative understanding. The TFM-IR detection mechanism is inherently sensitive to lateral force asymmetries and spatial gradients of the photothermal response. In conventional AFM-IR or s-SNOM, where the measured signal is effectively proportional to a spatially averaged optical response over the tip-sample interaction volume, resulting in spatial resolution limited by the tip apex size and thermal diffusion. This corresponds to a convolution of the true optical response with a finite point-spread function. In TFM-IR, the detected signal originates from torsional cantilever motion, which is driven by lateral force asymmetries (i.e., torque) within the tip-sample contact region (see Supplementary Note 2). This detection mechanism is inherently sensitive to spatial variations and gradients in the photothermal response, rather than its absolute magnitude. As a result, it can be viewed as a derivative-like measurement of the spatially averaged signal. This gradient-sensitive detection is effectively analogous to a derivative operation on the spatially varying signal. As a result, slowly varying background contributions are suppressed, while localized spatial variations are enhanced, leading to an apparent sharpening of features beyond what is typically observed in amplitude-based measurements. Mathematically, this corresponds to

a derivative-like operation on the spatially averaged signal. In addition, the torsional response arises from asymmetric tip-sample coupling and lateral force components, which can reduce the effective sensing volume compared to conventional flexural detection. Furthermore, the TFM-IR signal is generated through a nonlinear mixing process between the cantilever torsional dynamics and the laser repetition rate. This detection pathway can further enhance localized responses while suppressing distributed background contributions, thereby improving the effective spatial localization of the measured signal. Together, these factors enable TFM-IR to access highly localized variations in optical response, manifesting as ultrahigh spatial resolution as we have experimentally demonstrated.

We emphasize that the ~ 1 nm spatial resolution reported in this work is not an arbitrary claim. It is extracted from representative line profiles (e.g., Fig. 4a) and is consistent across multiple measurements and samples. This value reflects the effective spatial sensitivity of the TFM-IR contrast under the specific conditions employed, rather than a strict diffraction- or tip size- limited optical resolution.

Finally, we would like to share with the reviewer that our group has obtained unpublished results indicating that even higher spatial resolution can be achieved in TFM-IR imaging under ambient conditions from other material systems. In particular, these results reveal optical features unambiguously below 1 nm (see “Confidential Fig. 2”). Taken together, these observations further support our conclusion that the TFM-IR can achieve ultrahigh spatial sensitivity in appropriate regimes across a range of materials.

We have added further discussion in the revised main text to provide a qualitative understanding of the ultrahigh spatial resolution achieved in TFM-IR imaging.

“Figure 5d’s signal appears to be very strong (indicated from scale bar), but the noise level remains very high. This is strange. The images of Figure 5 do not show much difference across different IR energies, given the high noise background. If the author can articulate why there should be an IR wavelength dependent difference in the Moire patterns in a physical picture with beyond empirical understanding, that would be more convincing.”

Response: We thank the reviewer for this insightful comment. We first clarify that the SNRs for the wavenumber-dependent images in Fig. 5 are comparable across different excitation conditions, as quantified in our analysis (see **Table** provided above). The apparent high noise background arises primarily from the weak intrinsic signal levels in the mid-infrared regime for twisted bilayer graphene moiré lattice, rather than from inconsistencies between measurements. Despite the modest signal level, the data in Fig. 5 exhibit clear energy-dependent characteristics of the moiré lattice. For example, two distinct types of local optical motifs are revealed: dot-like hotspots with two-fold symmetry (Figs. 5a and 5c) and triangular patch-like features with three-fold symmetry (Fig. 5b and 5d). The intensity of these local motifs also varies with excitation wavelength, with

some conditions producing sharper, more pronounced hotspots, while others appear more diffuse with reduced amplitude.

While a complete microscopic understanding of the wavelength-dependent moiré contrast is beyond the scope of the present work and will require systematic experimental and theoretical investigation (including, e.g., anisotropy-resolved measurements demonstrated in “Confidential Fig. 1”), we outline below a physical picture to account for the observed behavior:

(1) Local stacking configuration within the moiré unit cell: Within a moiré superlattice, the local atomic registry varies continuously across the unit cell (e.g., AA, AB, and intermediate stacking configurations). These different stacking configurations are known to produce spatial variations in interlayer hybridization, symmetry, and local electronic structure. Such variations directly affect local optical absorption and electron-phonon coupling, leading to spatially modulated infrared responses, with supporting evidence from the site-resolved spectroscopy in Fig. 4b. Importantly, such variations can alter both the spectral weight and selection rules of optical transitions, resulting in spatially varying infrared responses that evolve with excitation energy (see Refs. 62, 63).

(2) Spatially varying strain distribution: Moiré superlattices are typically accompanied by non-uniform strain fields arising from lattice relaxation and interlayer interactions. Such strain is spatially modulated across the unit cell and is often concentrated near domain boundaries or specific stacking regions. These strain variations can locally shift phonon frequencies and modify electron-phonon coupling strength, thereby changing the resonance conditions for infrared absorption. As a result, different regions within a moiré unit cell can exhibit distinct spectral response, giving rise to energy-dependent contrast in TFM-IR imaging. The role of strain in shaping local electronic and optical properties in moiré systems has been theoretically predicted and indirectly supported experimentally, although direct spatially resolved experimental characterization remains challenging (see Refs. 64, 65). In this context, TFM-IR may provide sensitivity and spatial resolution to such strain-induced variations in future studies.

(3) Interlayer electronic coupling and symmetry breaking: The interlayer coupling varies spatially across the moiré unit cell, leading to modulation of the local band structure, including the emergence of flat bands and spatially localized states. These variations can influence optical transitions, even in the mid-infrared regime where dipole-allowed transitions are generally weak. Symmetry breaking and interlayer hybridization can activate otherwise weak or forbidden transitions through indirect processes such as phonon-assisted absorption or coupling to localized electronic states, contributing to spatially heterogeneous optical responses, which are inherently energy dependent. (see Refs. 63, 66).

Importantly, the three mechanisms discussed above are strongly intertwined in moiré systems and collectively govern the local optical and electron-phonon response. Their combined effect provides a natural physical basis for the observed wavelength-dependent contrast in the TFM-IR images. We anticipate that more quantitative microscopic modelling and simulations will be developed to

further elucidate the experimental observations presented in Fig. 4 and 5 in future work following the publication of the present study.

We have revised the manuscript with addition of a few new references to include additional discussion reflecting these considerations.

Finally, we acknowledge the reviewer's observation that the contrast differences across excitation wavelengths are modest. This is expected for twisted bilayer graphene in the mid-infrared regime, where the optical response is generally weak and dominated by indirect or symmetry-breaking mechanisms rather than strong dipole-allowed transitions. Consequently, spectral variations are subtle in magnitude but physically meaningful, reflecting changes in local coupling conditions rather than large amplitude differences. The observed wavelength-dependent contrast, together with the site-resolved spectroscopy within a moiré unit cell (Fig. 4b), is reproducible and spatially structured at the moiré length scale, supporting its intrinsic origin rather than being dominated by noise.

“Some minor comments:

More experimental details should be included in either method section or supplementary information. For example, I could not even find which AFM the authors used in their work. If it is a custom built AFM, the author should say so.”

Response: We thank the reviewer for pointing out this oversight. We have expanded the Methods section to include a complete description of the experimental setup, sample preparation, relevant experimental parameters, acquisition conditions, and measurement procedures to improve clarity and reproducibility.

“In the introduction, the authors says:

“This allows the optical signal intensity to be tuned and optimized via the torsional drive amplitude without disturbing the tip-sample feedback, offering a key advantage over conventional AFM-IR and sSNOM.”

In fact, sSNOM's signal channel and AFM feedback are naturally decoupled. Applying a CW laser in sSNOM does not perturb the tapping mode feedback in sSNOM. For AFM-IR techniques, except for the original contact mode AFM-IR, new embodiment such as PiFM, tapping mode AFM-IR, and PFIR all separate the AFM feedback from the photothermal signal read out channels. This statement from the author is not true.”

Response: We thank the reviewer for this insightful comment and apologize for the lack of clarity in our original wording.

We agree that in techniques such as sSNOM and several modern AFM-IR implementations (e.g., PiFM, tapping-mode AFM-IR, and PFIR), the feedback loop and optical detection channels are generally designed to be separated. Our intention was not to suggest that these methods lack

channel separation, but rather to highlight a specific aspect of the torsional force mode implementation used in this work.

In TFM-IR, the AFM feedback is maintained through the vertical deflection signal (with appropriate filtering to isolate the low-frequency feedback component), while the optical signal is extracted from the orthogonal torsional resonance response. In this configuration, the torsional detection channel can be independently optimized (e.g., via torsional drive amplitude) without directly perturbing the vertical feedback loop used for maintaining tip-sample interaction. This provides a practical degree of flexibility in tuning the optical signal channel independently within the same measurement framework. In contrast, in both sSNOM and AFM-IR, the photothermal response can influence the tip-sample interaction in a manner that indirectly affects both the imaging conditions and signal readout (e.g., through changes in oscillation amplitude, contact conditions, or tapping dynamics). As noted by the reviewer, modern implementations have introduced various strategies to mitigate such coupling; however, residual interdependence between interaction conditions and signal formation can still be present depending on the specific mode and implementation (see e.g., Ref. 14).

To avoid any ambiguity, we have revised the related statements in both the “Principles of TFM-IR and Direction-Resolved nanoscale Optical Imaging” section and the “Resolving Anisotropic Strain at the Nanoscale: Experiment and Simulation” section.

“The author should cite a few previous works on AFM-IR revealing h-BN’s polariton fringes. A few more references on the combinatory heterodyne schemes in AFM-IR/PiFM/surface sensitive AFM-IR should also be referenced.”

Response: We thank the reviewer for the suggestion and apologize for the oversight. We have revised the manuscript to include additional relevant references on AFM-IR studies of phonon polaritons in hBN, as well as prior works describing heterodyne and combinatory detection schemes in AFM-based infrared techniques. These references have been incorporated into the revised main text (Refs. 7, 19, 20) to better acknowledge prior work and to place our study in the appropriate context.

We believe that the revised manuscript has been significantly strengthened by addressing the thoughtful comments raised by the reviewer, and we sincerely thank the reviewer for all valuable feedback and suggestions.

Reviewer #3:

"The authors have extended the technique of Torsional Force Microscopy (TFM) by adding optical probing in the mid-infrared, thus establishing TFM-IR, and claim the ability to map strain in nanobubbles and wavelength-dependent features within a moire unit cell in twisted samples with a resolution approaching 1 nm. The technique is interesting, but there are some important issues with this work that need to be addressed before this can be published in any journal. Here are some of the main ones:"

Response: We thank the reviewer for taking the time to evaluate our work and for the positive comments. We welcome the opportunity to address the reviewer's questions point by point below.

"1. From the introduction and abstract it seems as if the authors developed a new technique from scratch, whereas in reality the TFM part already existed and they added optical excitation to this. They should be more clear about this."

Response: We fully acknowledge the prior seminal work on TFM, which forms the foundational basis for the TFM-IR approach presented in this work. We have revised the Introduction to more clearly reflect this prior development.

At the same time, we would like to emphasize that the novelty of the present work lies not in a simple addition of light to the TFM platform itself, but in its extension to infrared nanospectroscopy through a new torsional detection framework. Specifically, we introduce a nonlinear frequency-mixing scheme that enables the extraction of photothermal signals via torsional cantilever dynamics, allowing simultaneous access to both in-plane and out-of-plane responses. This approach leverages the unique sensitivity of torsional motion to lateral force asymmetries, enabling a distinct detection mechanism compared to conventional AFM-IR implementations. In addition, the method demonstrates enhanced spatial sensitivity and ultrahigh spatial resolution in optical imaging.

"2. The authors list disadvantages of sSNOM and AFM-IR, such as unavoidable tip induced perturbations and sophisticated apparatuses. I don't see how that's any different with TFM-IR, which in addition requires a laser with tunable repetition rate."

Response: We would like to clarify the context in which the comparison was made and further elaborate on the distinctions.

First, in the main text, the above comparison between TFM-IR and sSNOM/AFM-IR was intended specifically in the context of anisotropic and polarization-dependent optical imaging. In such measurements using sSNOM or AFM-IR, tip-induced perturbations to the local electromagnetic field or polarization state can become a concern and often require careful consideration or

additional complexity in the experimental configuration (e.g., electrically modulated photothermal force schemes as introduced in the Ref. 9). In the present work, TFM-IR addresses these challenges by employing a simple nonlinear frequency-mixing scheme involving the laser repetition rate, cantilever torsional excitation, and detection frequencies, which enables simultaneous access to in-plane and out-of-plane photothermal responses without directly relying on polarization-sensitive optical detection. This provides a practical advantage for studying anisotropic responses within the measurement framework used here.

Second, when considering the broader comparison of measurement architectures, TFM-IR provides a distinct separation between the feedback and signal channels. Specifically, the AFM feedback is maintained via the vertical deflection signal (with appropriate filtering to isolate the feedback loop from the torsional resonance), while the TFM-IR signal is extracted independently from the orthogonal torsional resonance amplitude. This allows the optical signal channel to be tuned (e.g., via torsional drive conditions and laser repetition rate) without directly affecting the feedback loop that controls the tip-sample interaction; In contrast, in conventional sSNOM and several AFM-IR implementations (including tapping AFM-IR), the tapping dynamics and tip-sample interaction parameters can influence both the topographic feedback and the optical signal channel. As a result, the feedback conditions and optical response are not always fully independent. For example, changes in tapping amplitude or tip-sample coupling can simultaneously affect signal generation and imaging conditions (see e.g., Ref. 14).

We have revised the main text to clarify that these statements are made in the context of specific measurement considerations and to better reflect the distinctions between these techniques.

“3. The authors claim a much better resolution than sSNOM, citing Refs. 53-55 to have a spatial resolution of several hundred nanometers. This seems incorrect, because at least Ref. 55 seems to have a sub-10 nm resolution.”

Response: We thank the reviewer for this comment and for pointing out the need for clarification.

We respectfully note that our interpretation of original Ref. 55 (now Ref. 59 in the revised main text) is based on the authors’ explicit statement regarding the spatial resolution of their s-SNOM measurements. In that work, the authors state in the Methods section: *“The spatial resolution of the s-SNOM is predominantly set by the tip radius of curvature a (~ 20 nm in our experiment) and therefore allows us to resolve optical and optoelectronic properties at a length scale below that of the ferroelectric domains. First, the laser was focused on a metalized AFM tip using a parabolic mirror. Then the back-scattered light was registered. With this approach, we obtained the genuine near-field signal at a resolution of ~ 20 nm”* (Ref. 59). Based on their presented data and the authors’ own description, the reported spatial resolution is on the order of the tip radius (~ 20 nm), rather than sub-10 nm as suggested by the reviewer.

We acknowledge that different s-SNOM implementations, sample conditions, and analysis methods can lead to variations in the effective spatial resolution. However, for consistency with the cited reference and its reported experimental parameters, we have revised the corresponding statement in our manuscript to more accurately reflect the spatial resolution described in Ref. 59 and to avoid potential overgeneralization. The revised text now reads: “For example, sSNOM studies of twisted bilayer graphene typically achieve spatial resolutions of a few tens to hundreds of nanometers, with the tip apex size imposing a fundamental limitation⁵⁴⁻⁵⁹”.

“4. The authors created the impression that they are studying twisted bilayer graphene, where one graphene layer is twisted w.r.t. the other layer, but it seems that they just transferred bilayer graphene onto hBN. So is it just the twist angle between bilayer graphene and hBN? The authors should be clear about this and provide sufficient detail. It would also be nice to verify the obtained twist angle by an independent measurement.”

Response: We thank the reviewer for this important question and for the opportunity to clarify our sample configuration. We apologize for the lack of clarity in our original description of the sample preparation. We would like to clarify that the samples studied in this work are indeed twisted bilayer graphene structures, in which the twist is introduced between the two graphene layers, rather than between graphene and the hBN substrate. The hBN served as a supporting substrate and was intentionally misaligned during the fabrication process to avoid introducing additional moiré patterns.

To avoid any ambiguity, we have revised the Methods section to include a detailed description of the fabrication procedure. We have also added a new Supplementary Figure (Supplementary Fig. 11) characterizing the step-by-step assembly of the moiré lattice and confirming the fabrication process. Briefly, the samples were prepared using a modified tear-and-stack method (Refs. 79-81). A PDMS/PPC stamp mounted on a glass slide was first used to pick up a thick hBN flake, followed by the first graphene layer. The second graphene layer was then picked up after a controlled rotation by approximately 2.4° , thereby defining the twist angle between the two graphene layers. The hBN flake was intentionally misaligned relative to the graphene layers to minimize substrate-induced moiré effects. The nominal twist angle defined during fabrication is consistent with the periodicity of the moiré lattice observed in our AFM measurements (Fig. 4a).

“5. The authors claim a super-resolution approaching 1 nm, but their data in Figures 1-3 don't seem to support this. The data in Figures 4-5 is more convincing in this regard, but it's very disappointing that the reason for this resolution is not discussed even with some attempt of an explanation. Does it not have anything to do with the amplitude of the motion of the tip?”

Response: We thank the reviewer for raising this important question and for the opportunity to clarify both the scope of our claims and our understanding of the underlying mechanism.

First, we would like to clarify the roles of different datasets in this work. The three main contributions of the present study are: (i) the introduction of TFM-IR and its comparison with prior techniques using hBN phonon polaritons as a representative example (Fig. 1); (ii) the demonstration of simultaneous access to in-plane and out-of-plane photothermal responses enabled by a nonlinear frequency-mixing scheme, illustrated using mica nanobubbles (Figs. 2 and 3); and (iii) the observation of near-nanometer spatial resolution in TFM-IR imaging, demonstrated using moiré patterns in twisted bilayer graphene (Figs. 4 and 5), where the intrinsic lattice periodicity provides a suitable length scale for evaluating spatial resolution.

Because the relevant spatial features of hBN polaritons and mica nanobubbles extend hundreds of nanometers to micrometer scales, near-nanometer spatial resolution is neither required nor critical for interpreting the underlying physics. In contrast, the moiré lattice in twisted bilayer graphene provides a well-defined nanoscale structure, enabling a more direct assessment of the spatial resolving capability of the technique.

Second, while a comprehensive microscopic theory of the ultimate spatial resolution of TFM-IR will require further systematic investigation, we provide the following qualitative understanding. In conventional sSNOM and AFM-IR, the measured signal is effectively averaged over the finite tip-sample interaction volume and is typically limited by the tip apex size and thermal diffusion, corresponding to a convolution of the true optical response with a finite point-spread function; In TFM-IR, the detected signal originates from torsional cantilever motion, which is driven by lateral force asymmetries (i.e., torque) within the tip-sample contact region (see Supplementary Note 2). This detection mechanism is inherently sensitive to spatial variations and gradients in the photothermal response, rather than its absolute magnitude. As a result, it can be viewed as a derivative-like measurement of the spatially averaged signal. Such gradient-sensitive detection suppresses slowly varying background contributions while enhancing localized variations, thereby increasing sensitivity to high spatial-frequency components. Consequently, features that appear broadened in conventional AFM-IR can appear sharper in TFM-IR, leading to a spatial resolution that can be much smaller than the nominal tip-sample interaction size. In addition, the torsional interaction involves lateral force components and asymmetric force distributions within the contact region, which can further reduce the effective sensing volume compared to flexural detection modes. The TFM-IR signal is also generated through a nonlinear mixing process involving the cantilever torsional dynamics and the laser repetition rate, which can selectively enhance localized responses while suppressing distributed background contributions. Together, these factors contribute to the high spatial sensitivity observed in TFM-IR under favorable conditions.

Regarding the reviewer's question on the role of the oscillation amplitude of the tip, we agree that the driven torsional oscillation amplitude can influence the effective interaction and signal strength. In practice, optimization of the torsional drive amplitude is part of achieving optimal signal-to-noise and spatial sensitivity. This represents one key advantage of TFM-IR that the torsional detection channel is decoupled from the vertical deflection-based feedback that controls the tip-sample interaction, which allows the optical signal to be optimized via torsional excitation

parameters without directly perturbing the feedback loop, providing additional flexibility in balancing sensitivity and stability.

We have added a dedicated paragraph in the revised main text to clarify these points and to better explain the mechanisms underlying the observed spatial resolution in TFM-IR.

“6. The descriptions of the figures are either lacking details (1a for example) or containing superfluous details (1c for example).”

Response: We appreciate the criticism about figure captions, and have carefully revised all figure captions in the manuscript to improve clarity, consistency, and level of detail. In particular, we have expanded caption of Fig. 1a and streamlined those that contained overly detailed descriptions (e.g., Fig. 1c), to achieve a better balance between completeness and readability.

“7. The authors seem to show results for only one nanobubble, calling it a "typical" nanobubble, which created the expectation of seeing more similar data on other nanobubbles in the Supplement, but these are absent. This makes this an $N = 1$ experiment. Seeing more and similar data would make this work much more convincing.”

Response: We thank the reviewer for this important comment. In fact, nanobubble structures can be easily found in our samples. In the Supplementary Fig. 6a, we showed that mica nanobubbles are commonly present across exfoliated natural mica samples, indicating that the specific structure presented in the Figs. 2 and 3 are not unique but representative of a broader class of features observed in our samples. In addition, we have added a new Supplementary Fig. 6b with corresponding TM-IR image of the same large region in the Supplementary Fig. 6a to highlight consistent optical features observed for all nanobubbles.

To further address the reviewer’s concern, we have provided an additional independent set of individual mica nanobubble in the new Supplementary Fig. 9. The specific geometric parameters (e.g., height and width) of that nanobubble differ from the one presented in the main text. Nevertheless, their in-plane and out-of-plane photothermal responses exhibit consistent spatial patterns and strain-related features.

These additional data support the reproducibility of our observations across multiple nanobubbles and strengthen the generality of the conclusions drawn in the manuscript. We have added a dedicated paragraph in the revised main text to clarify these points.

“8. The data in Figure 3 shows a dependence on excitation wavelength, but the explanation for the observations is not entirely convincing. And are there no artefacts when changing wavelength? Different collimation, focus position, etc? Here, data on more nanobubbles could strengthen the case again.”

Response: We thank the reviewer for this thoughtful comment. The primary purpose of Fig. 3 (and Fig. 5 as well) is to experimentally demonstrate the tunable optical excitation capability of our TFM-IR approach and to illustrate its potential for probing anisotropic, wavelength-dependent phenomena, such as spatial redistribution of strain. A complete quantitative interpretation of the wavelength-dependent redistribution of strain would require detailed modeling that incorporates wavelength-dependent optical absorption, as well as structural variations and possible defects in individual nanobubbles. Such an analysis is beyond the scope of the present manuscript, which focuses on establishing the experimental capability and demonstrating representative behavior.

Regarding potential experimental artifacts associated with wavelength tuning and optical alignment, we have taken several measures to minimize and assess such effects. First, we employ broadband optical components throughout the optical path, avoiding the need to replace optics when tuning wavelength (see revised Methods). This helps maintain consistent beam alignment and reduces wavelength-dependent variations in collimation and focusing conditions. Second, any residual misalignment or focus drift typically leads to smooth, global intensity variations rather than localized contrast changes within individual nanostructures, whereas the observed features in Fig. 3 correspond to spatially resolved, structure-specific variations.

In addition, as described in the Methods section, we utilize two different laser sources and observe consistent results at overlapping wavelengths, indicating that the observed wavelength dependence is not dependent on a particular optical source. We have also performed control experiments to rule out mechanical cross-talk effects- for example, the disappearance of the observed features under off-resonance optical excitation (see new Supplementary Fig. 9 for mica nanobubbles and Supplementary Fig. 10 for moiré lattices) supports the optical origin of the signal.

Importantly, in our wavelength-dependent measurements (both Fig. 3 for nanobubbles and Fig. 5 for moiré lattices), we verify that the corresponding height and TFM amplitude images remain unchanged throughout the measurements. This consistency confirms that the sample morphology remains stable, providing that the observed variations arise from intrinsic wavelength-dependent optical responses rather than changing sample conditions or damage.

Taken together, these experimental controls and consistency checks support our conclusion that the observed wavelength-dependent TFM-IR signals are intrinsic to the samples rather than artifacts associated with the tuning of excitation wavelength or mechanical cross-talk.

We have provided more discussion in the revised main text and Methods to reflect these considerations.

“9. There is also not much explanation of the observed wavelength-dependent features within a moire unit cell.”

Response: While a complete microscopic understanding of the wavelength-dependent moiré contrast is beyond the scope of the present work and will require systematic experimental and theoretical investigations (including, e.g., anisotropy-resolved measurements demonstrated in “Confidential Fig. 1”), we outline below a physical picture to account for the observed behavior:

(1) Local stacking configuration within the moiré unit cell: Within a moiré superlattice, the local atomic registry varies continuously across the unit cell (e.g., AA, AB, and intermediate stacking configurations). These different stacking configurations are known to produce spatial variations in interlayer hybridization, symmetry, and local electronic structure. Such variations directly affect local optical absorption and electron-phonon coupling, leading to spatially modulated infrared responses, with supporting evidence from the site-resolved spectroscopy in Fig. 4b. Importantly, such variations can alter both the spectral weight and selection rules of optical transitions, resulting in spatially varying infrared responses that evolve with excitation energy (see Refs. 62, 63).

(2) Spatially varying strain distribution: Moiré superlattices are typically accompanied by non-uniform strain fields arising from lattice relaxation and interlayer interactions. Such strain is spatially modulated across the unit cell and is often concentrated near domain boundaries or specific stacking regions. These strain variations can locally shift phonon frequencies and modify electron-phonon coupling strength, thereby changing the resonance conditions for infrared absorption. As a result, different regions within a moiré unit cell can exhibit distinct spectral response, giving rise to energy-dependent contrast in TFM-IR imaging. The role of strain in shaping local electronic and optical properties in moiré systems has been theoretically predicted and indirectly supported experimentally, although direct spatially resolved experimental characterization remains challenging (see Refs. 64, 65). In this context, TFM-IR may provide sensitivity and spatial resolution to such strain-induced variations in future studies.

(3) Interlayer electronic coupling and symmetry breaking: The interlayer coupling varies spatially across the moiré unit cell, leading to modulation of the local band structure, including the emergence of flat bands and spatially localized states. These variations can influence optical transitions, even in the mid-infrared regime where dipole-allowed transitions are generally weak. Symmetry breaking and interlayer hybridization can activate otherwise weak or forbidden transitions through indirect processes such as phonon-assisted absorption or coupling to localized electronic states, contributing to spatially heterogeneous optical responses, which are inherently energy dependent. (see Refs. 63, 66).

Importantly, the three mechanisms discussed above are strongly intertwined in moiré systems and collectively govern the local optical and electron-phonon response. Their combined effect provides a natural physical basis for the observed wavelength-dependent contrast in the TFM-IR images.

We anticipate that more quantitative microscopic modelling and simulations will be developed to further elucidate the experimental observations presented in Fig. 4 and 5 in future work following the publication of the present study.

We have added a new paragraph in the revised main text, including several new references, to incorporate additional discussion reflecting these considerations.

We believe the revisions made in response to the reviewer's comments and questions have significantly strengthened the manuscript, and we sincerely thank the reviewer for the thoughtful feedback and careful review of our work.

Response to Reviewer #1:

“The authors have reasonably and satisfactorily addressed my comments and concerns in their response and the revised manuscript. I believe the manuscript is now suitable for publication in Nature Communications.”

Response: We sincerely thank the reviewer for the careful evaluation of our manuscript and for the positive assessment of our revision. We greatly appreciate the reviewer’s constructive feedback throughout the review process, which has helped us improve the clarity and quality of the manuscript. We are pleased that the revised version is now considered suitable for publication.

Response to Reviewer #2:

“The revision improved the clarity of this manuscript, and expanded the background information of this manuscript.

While I acknowledge the novelty of this work on method implementation and interpretation of signal original in their original submission, I am still not convinced that the spatial resolution is from the IR/spectroscopy rather than due to the mechanical readout of the torsional force. The additional example shown in the supplementary still lacks spectroscopic contrast, thus not convincing.

I would only agree to the author's claim of 1 nm only if they can measure a sample with compositional/spectroscopic heterogeneity not mechanical heterogeneous response. One test example would be a cellular surface that has intrinsic chemical heterogeneity, or a protein aggregate, like amyloid fibrils, under different IR excitation wavelength.”

Response: We thank the reviewer for the continued thoughtful evaluation of our work and for recognizing the novelty of the method. We understand that the remaining concern is whether the observed ~1 nm spatial resolution in TFM-IR reflects a genuine optical (spectroscopic) signal or arises from mechanical crosstalk or from a mechanically mediated readout of tip-sample forces.

We first respectfully clarify a potential misunderstanding regarding the additional data provided in our previous revision. The new datasets (Supplementary Fig. 9a and Supplementary Fig. 10) were intentionally acquired under off-resonant excitation conditions, for which the absence of contrast in TFM-IR images is expected. These measurements, performed on two distinct systems (mica nanobubbles and twisted bilayer graphene moiré), consistently show featureless signals under off-resonant excitation. Rather than lacking spectroscopic contrast, as noted in the reviewer’s comment, these results provide direct evidence of spectral selectivity, supporting the conclusion that the observed TFM-IR signal originates from optical absorption rather than mechanical crosstalk. We apologize for not making this point sufficiently clear in the previous revision.

To further address the reviewer’s concern, we clarify and strengthen our evidence along several key directions:

(1) Spectral selectivity. Nanoscale features in TFM-IR images appear only at resonant excitation wavelengths and disappear under off-resonant conditions. A purely mechanical crosstalk mechanism would not produce such wavelength-dependent behavior, as the mechanical response of the cantilever-sample system is not intrinsically selective to infrared excitation wavelength.

(2) Mode-dependent (anisotropic) response. We observe distinct in-plane and out-of-plane signals that evolve differently with excitation wavelength. As described in the manuscript, anisotropic detection in TFM-IR arises from a nonlinear frequency-mixing scheme and does not rely on modulation of tip-sample mechanical interaction. A fixed mechanical readout pathway would therefore be expected to yield a consistent spatial response independent of detection channel. In contrast, the observed mode-dependent contrast tracks the optical excitation, consistent with anisotropic photothermal forces and difficult to reconcile with a purely mechanical origin.

(3) Decoupling from mechanical/topographic channels. To facilitate comparison, we have compiled the data from Figs. 4, 5 and Supplementary Fig. 10 and included the simultaneously acquired TFM images corresponding to each TFM-IR image at different excitation wavelengths. The TFM (mechanical) signals remain unchanged across all measurements, while the TFM-IR signal exhibits pronounced changes in both amplitude and spatial distribution. Importantly, this evolution is not limited to global intensity scaling; rather, we observe clear spatial redistribution of contrast, including changes in relative intensity and the emergence or suppression of distinct local motifs (e.g., two-fold *vs* three-fold features at different excitation energies). If the signal originated from a mechanically mediated readout of a spatially fixed response, varying the excitation wavelength would primarily result in uniform amplitude scaling while preserving the spatial pattern. Instead, the observed wavelength-dependent reorganization of spatial features indicates that the underlying optical response itself evolves spatially with excitation energy. Such behavior cannot be explained by sharpening of a static response.

Taken together, the combination of spectral selectivity, mode-dependent contrast, and independence from mechanical/topographic signals provides strong evidence that the observed nanoscale features originate from genuine optical (photothermal) contrast rather than mechanical crosstalk/artifacts or from a mechanically mediated readout of tip-sample forces.

We also appreciate the reviewer's suggestion to examine chemically heterogeneous systems (e.g., biological samples). We note that such biological systems involve substantially different sample preparation, properties, and experimental conditions, and are beyond the scope of the present study. Nevertheless, we are pleased to share unpublished confidential data acquired from a different two-dimensional material containing defects (see "Confidential Data for Editor and Reviewers Only.pdf"). These measurements demonstrate that TFM-IR can indeed resolve spatially localized, wavelength-selective optical features with chemically heterogeneous defects within a mechanically homogeneous background (i.e., featureless in the simultaneously acquired TFM channel). This dataset directly addresses the reviewer's suggestion in a non-biological system and further supports the optical origin of the observed TFM-IR contrast while demonstrating its capability to resolve features with ultrahigh spatial resolution.

More broadly, we emphasize that in photothermal AFM-based techniques (including both conventional AFM-IR and TFM-IR), the detected signal inherently involves mechanical transduction of optically induced forces or torques. Consequently, even in chemically heterogeneous samples, variations in local optical absorption will manifest through mechanically detected signals. The key criterion for establishing the optical origin is therefore whether the observed contrast follows the spectral dependence of optical excitation rather than static mechanical properties. In this regard, our results consistently demonstrate: (i) disappearance of optical contrast under off-resonant excitation, (ii) strong wavelength dependence of both amplitude and spatial distribution, (iii) invariance of the mechanical (TFM) channel, and (iv) the ability to resolve optical features in mechanically homogeneous systems (confidential data). These combined observations show that the measured contrast with ultrahigh spatial resolution cannot be attributed to sharpening of a static mechanical response but instead reflects intrinsically spatially and spectrally varying optical behavior.

Finally, we emphasize that, as in other scanning probe techniques where the measured signal originates from a specific physical interaction while the spatial resolution is determined by the nonlinear detection mechanism, the ultimate spatial resolution in TFM-IR is governed by the

detection mechanism that converts local interactions into measurable signals. In conventional AFM-IR, localized photothermal forces are transduced into flexural motion of the cantilever, whereas in TFM-IR they are transduced into torsional motion. The observed ultrahigh spatial resolution likely reflects the intrinsic sensitivity of torsional detection to local force gradients and asymmetric tip-sample interactions, as discussed in the revised manuscript and supported by Ref. 11. Importantly, this does not diminish the optical nature of the measurement; rather, it highlights a key strength of TFM-IR: its ability to transduce local optical responses into highly localized measurable signals.

In summary, we respectfully maintain that our data provide strong and internally consistent evidence that TFM-IR measures genuine optical responses with nanoscale spatial resolution, and that the observed contrast cannot be explained by mechanical crosstalk or static mechanical readout alone. We have revised the main text accordingly to further clarify this point. We again thank the reviewer for their insightful comments and the opportunity to further strengthen the manuscript.

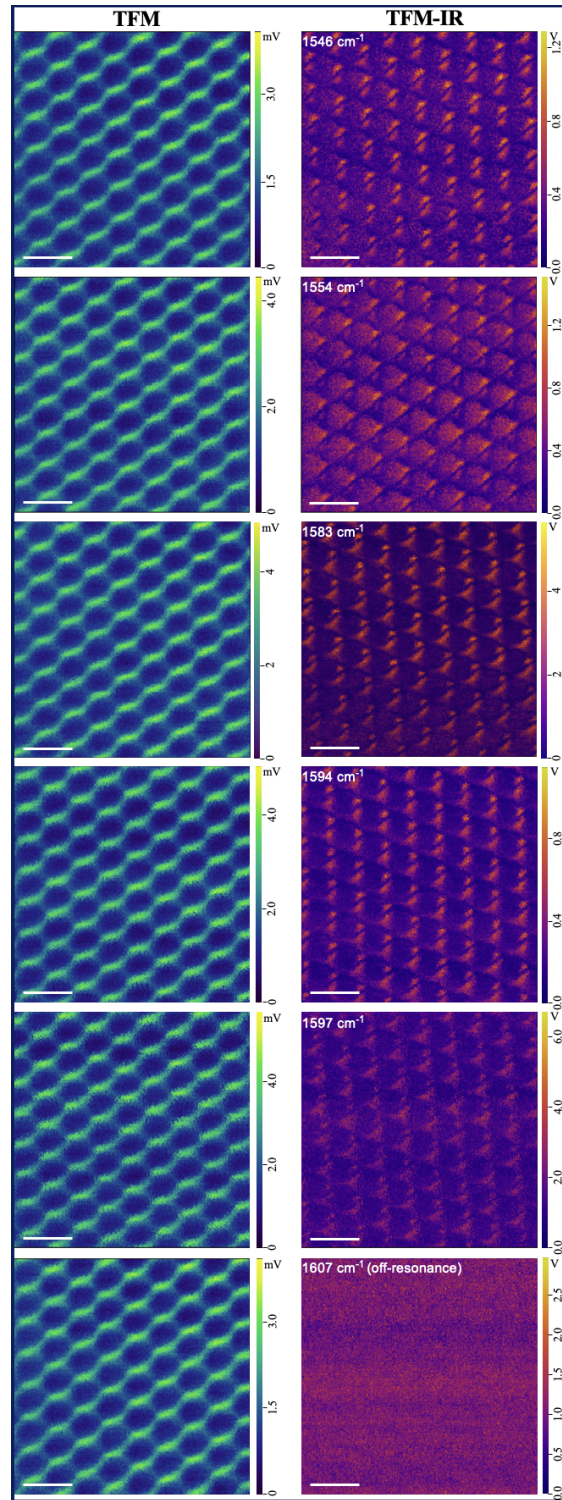


Figure | Comparison of simultaneously acquired TFM and TFM-IR images under different optical excitation wavelengths. The TFM (mechanical) images remain unchanged across all conditions, while the TFM-IR signal exhibits pronounced variations in both amplitude and spatial distribution. The optical contrast in TFM-IR disappears under off-resonant excitation. Data are compiled from Fig. 4a, Fig. 5, and Supplementary Fig. 10. Scale bar, 5 nm.

Response to Reviewer #3:

“The authors have addressed my main concerns. I do still see quite some typos, and would therefore advise a careful proofreading, in particular of the Supplement.”

Response: We thank the reviewer for the positive assessment of our revised manuscript and for acknowledging that our responses have addressed the main concerns. We also appreciate the helpful suggestion regarding remaining typographical errors. We have carefully proofread the manuscript and the Supplementary Information and corrected these issues in the revised version.