

ARTICLE

DOI: 10.1038/s41467-018-04887-1

OPEN

High-veracity functional imaging in scanning probe microscopy via Graph-Bootstrapping

Xin Li^{1,2,3,4}, Liam Collins^{1,2}, Keisuke Miyazawa⁵, Takeshi Fukuma⁶, Stephen Jesse^{1,2} & Sergei V. Kalinin ^{1,2}

The key objective of scanning probe microscopy (SPM) techniques is the optimal representation of the nanoscale surface structure and functionality inferred from the dynamics of the cantilever. This is particularly pertinent today, as the SPM community has seen a rapidly growing trend towards simultaneous capture of multiple imaging channels and complex modes of operation involving high-dimensional information-rich datasets, bringing forward the challenges of visualization and analysis, particularly for cases where the underlying dynamic model is poorly understood. To meet this challenge, we present a data-driven approach, Graph-Bootstrapping, based on low-dimensional manifold learning of the full SPM spectra and demonstrate its successes for high-veracity mechanical mapping on a mixed polymer thin film and resolving irregular hydration structure of calcite at atomic resolution. Using the proposed methodology, we can efficiently reveal and hierarchically represent salient material features with rich local details, further enabling denoising, classification, and high-resolution functional imaging.

¹Institute for Functional Imaging of Materials, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. ²Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. ³Department of Industrial and Manufacturing Engineering, Florida State University, Tallahassee, FL 32306, USA. ⁴Department of Statistics, Florida State University, Tallahassee, FL 32306, USA. ⁵Division of Electric Engineering and Computer Science, Kanazawa University, Kakuma-machi, 920-1192 Kanazawa, Japan. ⁶Nano Life Science Institute (WPI-NanoLSI), Kanazawa University, Kakuma-machi, 920-1192 Kanazawa, Japan. Correspondence and requests for materials should be addressed to S.V.K. (email: sergei2@ornl.gov)

After the first demonstration of atomic force microscopy (AFM) by Binnig and Rohrer¹, AFM and related force-based scanning probe microscopies (SPMs) have rapidly become the tool that opened the nanoworld for exploration and modification². Since the late 1980s, a variety of SPM methods sensitive to magnetic³, electrostatic^{4–8}, mechanical^{9–11}, and piezoelectric^{12–15} properties of surfaces have been realized, followed by the development of a number of spectroscopic modes that provided insight into kinetics and thermodynamics of single-molecule reactions^{16–19} and mechanisms of bias-induced phase transitions^{20–24} on a single defect level.

Despite this progress, for over 20 years, the progress in SPM methods was associated preponderantly with the development of low noise and controlled environment platforms, as well as functionalized probes. At the same time, the basic principles of signal processing and visualization involved in SPM imaging remained the same, namely the use of the single frequency heterodyne detection methods in lock-in and phase-locked loop detection, and plotting-associated amplitude/phase or frequency maps. Considerable progress in SPM instrumentation was achieved in the early 2000s, with the introduction of dual frequency methods by Garcia²⁵ and Proksch²⁶. The next step was the development of the band excitation (BE) method by Jesse and Kalinin^{21, 27–31}, which enabled quantitative measurements of conservative and dissipative interactions on the nanoscale. Since then, an increasing number of multifrequency SPM techniques including, amongst many others, bi-/tri-modal SPM^{32, 33} and intermodulation techniques^{34, 35} as well as multidimensional approaches such as three-dimensional force mapping AFM^{36–38} (3D-AFM), holography³⁹, and time-resolved^{40–42} techniques have been realized. Finally, the general acquisition (G)-Mode^{43–45} approach has recently been developed to enable full capture of the information stream from the photodetector, potentially reaching the information limit of SPM.

However, multifrequency/multidimensional techniques such as intermodulation SPM^{34, 35}, BE^{27–31}, and G-Mode^{43–45} generate large, often complex, datasets, necessitating approaches for

visualizing and converting the data to materials specific information. In prior BE work, we have primarily used functional fitting of the data in the Fourier domain^{27–31}, relying on a prior based on simple harmonic oscillator physics (SHO model) of the tip–surface interactions. This approach by definition ignores the materials behaviors associated with deviations from SHO models, for example, nonlinearities that lead to dynamic stiffening or softening of the tip–surface junction and hence require more complex dynamic models. The introduction of such models, in turn, can lead to significant issues such as spurious growth in the number of free parameters, expansive analysis times, potential overfitting, etc. Meanwhile, the linear unmixing methods based on principal component analysis⁴⁶ show only limited usefulness, since the BE signal is non-linear with respect to the local mechanical properties.

Here, we propose and implement an approach based on low-dimensional embedding of high-dimensional data via a combination of graph analytics and hierarchical clustering, illustrated for BE-SPM and 3D-AFM but generally applicable to other data-rich SPM modalities. We note that fundamental physics of the tip–surface interactions is intrinsically low-dimensional and is determined by a relatively small number of materials parameters. The transfer function of the cantilever is a non-linear^{11, 47} (and generally very complex) function of these parameters, precluding the use of the linear unmixing methods for the analysis. However, we argue that the intrinsic low dimensionality of the physics suggests the presence of the low-dimensional manifold can be derived from the high-dimensional response space of SPM measurements. Figure 1 illustrates the concept of manifold-physic inference applied to functional imaging using SPM.

Results

Traditional BE-SPM Imaging. Unlike single frequency techniques, which excite/detect the tip–sample interaction within a single frequency bin, BE detection utilizes a non-sinusoidal excitation signal having finite amplitudes over a selected frequency space. Practically, this is achieved using a digitally

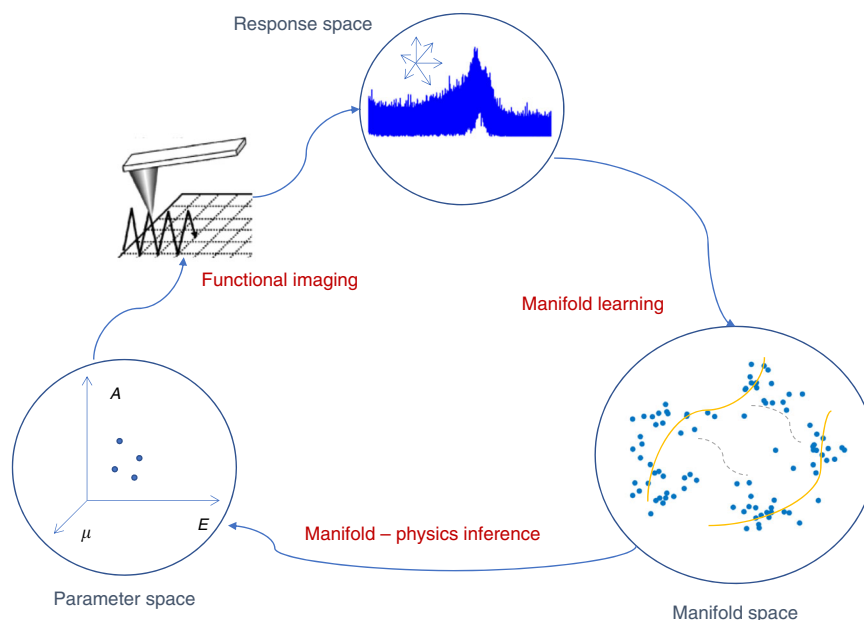


Fig. 1 Concept of manifold-physic inference. The properties of the surface that define the scanning probe microscopy (SPM) signal form a low-dimensional parameter space, for example, defined by the Young and Poisson moduli and the work of adhesion. These properties are translated onto a (very high dimensional) response space by means of the imaging mechanism of the cantilever/microscope. Aside from possible discontinuities due to changes in imaging regimes, points that are in close proximity in the parameter space will generally be in close proximity in the response space, forming a complex non-linear manifold

synthesized signal having a defined band of frequencies (in the Fourier domain), which are subsequently inverse fast Fourier transformed (iFFT) to generate a signal in the time domain, used to modulate the tip-sample interaction. Further, the band of frequencies is typically chosen to be positioned (in the Fourier domain) across one or many contact resonance (CR) peaks, allowing further insight into the cantilever dynamics than accessible in single frequency techniques.

In the linear regime, the behavior of the cantilever CR can be approximated by a SHO and described by three independent parameters: CR frequency, ω_0 , amplitude at the CR, A_0 , and quality factor, Q , which can be deconvoluted and stored as images as given by Eqs. (1) and (2):

$$A(\omega) = \frac{A^{\max} \omega_0^2}{\sqrt{(\omega^2 - \omega_0^2)^2 + (\omega \omega_0 / Q)^2}} \quad (1)$$

$$\tan(\phi(\omega)) = \frac{\omega \omega_0 / Q}{\omega^2 - \omega_0^2}. \quad (2)$$

In this way, knowledge of the full CR behavior in BE allows both conservative ($\Delta\omega_0$) and dissipative (ΔQ) contributions of the tip-sample interaction to be decoupled. In terms of mechanical property measurements, this enables separating the influence of elastic (conservative) and viscous (dissipative) material behavior. This makes BE-SPM (sometimes referred to as CR-SPM) a promising route for mapping local mechanical properties of materials.

Figure 2 represents the topography and the SHO fitting results of the BE measurements on a thin-film polymer blend of polycaprolactone (PCL) in a polystyrene (PS) matrix. In Fig. 2b,

the amplitude response is seen to be reasonably constant across the entire image, with abrupt changes in the amplitude response detected at the interfaces between materials, likely due to sudden changes in sample topography leading to imperfect tip-sample contact. At the same time, clear evidence of differences in viscoelastic behavior of the polymer blend can be observed in the resonance frequency and Q maps, Fig. 2c, d, respectively. The PS matrix clearly shows a higher resonance frequency and Q factor than observed in the PCL inclusions. The shift in the CR to a higher frequency indicates an increase in the stiffness (Young's modulus) of the material. Whereas peak shape (Q) behavior is reflective of energy dissipation in the tip-surface junction, such that a lower relative Q value is indicative of a more viscous/compliant material. In summary, the results in Fig. 2 strongly suggest that the PS matrix is stiffer and less viscous than the PCL material, in agreement with previous works involving nanomechanical mapping by SPM techniques⁴⁸ of these samples.

This analysis based on the simple model of harmonic oscillator is extremely useful; however, as previously discussed, the reliance on a SHO type model precludes the investigation of subtle but important materials non-linearities (dynamic mechanical responses in compliant materials). Hence, further progress requires comprehensive data-driven approaches to visualize the data and elucidate underlying physics, ultimately yielding higher veracity functional imaging.

Graph-Bootstrapping. Over the past 2 decades, networks have become an invaluable tool for extracting insights from large-scale complex systems in many branches of science^{49, 50}. Despite successful applications in numerous areas, network analytics usually require existing graph databases, which are constructed by manually labeling instances over a long time. However, of interest here is the emerging class of SPM techniques involving

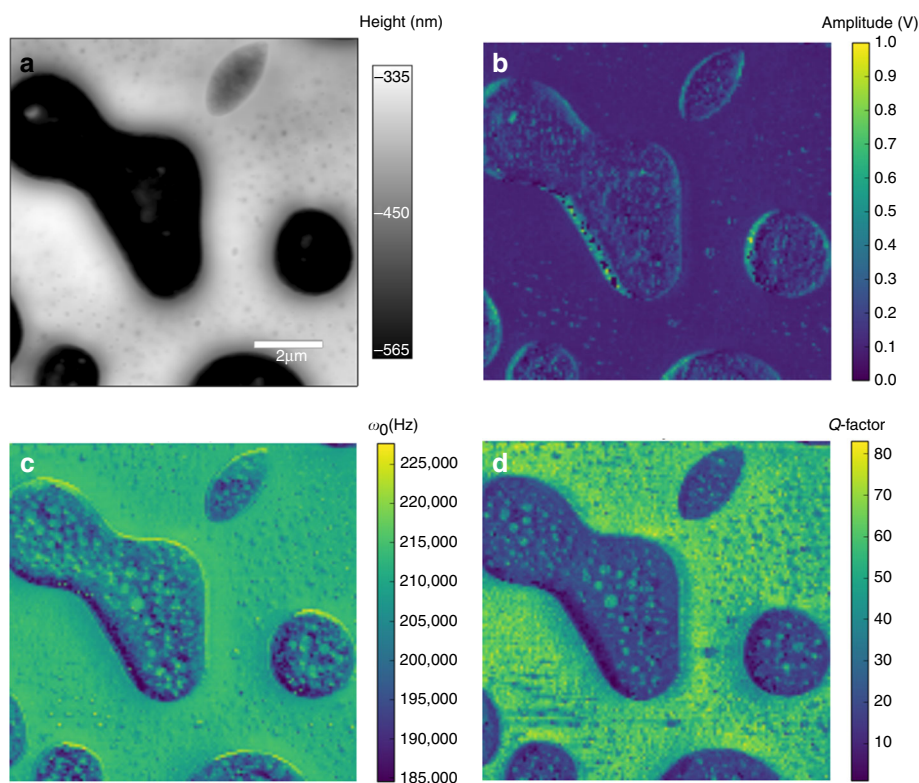


Fig. 2 Traditional BE-SPM imaging on a thin-film polymer blend of PCL in a PS matrix. **a** Band excitation scanning probe microscopy (BE-SPM) topography of polystyrene/polycaprolactone (PS/PCL) sample. Spatial mappings of **(b)** amplitude, A , **(c)** contact resonance (CR) frequency, ω_0 , and **(d)** Q -factor, fitted via the simple model of harmonic oscillator (SHO)

significantly increasing levels of data capture, which often lack prior physical understanding of the tip-sample interactions and their relationship to particular material functionalities. Therefore, to gain insights into the behavior of both global and local relationships within high-dimensional measurements, we first construct the nearest neighborhood graph⁵¹. The complexity of constructing an exact nearest neighbor graph is $O(n^2p)$, which is too expensive for high-dimensional datasets $\{X_1, X_2, \dots, X_n\}$ where $X_i \in R^p$, especially for SPM datasets in which depending on the detection mode (Broad-band, G-Mode) p can reach to 10^4 . Recently, Tang et al.⁵² proposed LargeVis, a very efficient algorithm to build approximate nearest neighbor graph based on random projection tree⁵³ and neighbor-exploring⁵⁴ techniques. The weights of edges are calculated in a similar way to the distributed stochastic neighbor embedding (t-SNE)⁵⁵ method, by converting the Euclidean distances between neighbors into conditional probabilities that represent similarities. Furthermore, LargeVis layouts the nearest neighbor graph in low-dimensional manifold space following a principled probability model solved via asynchronous stochastic gradient descent⁵⁶.

For SPM datasets, especially for high-dimensional detection mode ($p \sim 10^4$), we found LargeVis projection on the low-dimension manifold space tends to be few bulk modules, which can usually distill most salient parts of the material. However, it is difficult to further explore local details in those modules. At the same time, the low-dimensional (2D/3D) coordinates calculated by LargeVis preserve all relationships between measurements. To explore intrinsic structures and present them in a clearer way, we propose reconstructing the graph based on the LargeVis low-

dimensional manifold coordinates and subsequently recalculating the manifold layout positions based on the reconstructed graph, following the same principled probability model. We refer to this method as Graph-Bootstrapping.

Traditionally, manifold embeddings are used for visualization purposes only, that is, to overlay known labels on the manifold points. Here, we expand this approach by learning manifold from high-dimensional measurements and subsequently clustering on the manifold to provide a natural way of visualizing and denoising the data, as well elucidating the relevant physics. This task however brings two main challenges. First, the manifold layout should consist of distinguishable groups which when accessible will enable straightforward clustering to be performed. Second, the clustering algorithm should be robust enough to cover irregular manifold layouts. Figure 3 presents a concrete example of manifold embedding and clustering with simulated Gaussian curves. Specifically, we consider four groups of parameters (Fig. 3a): two ranges of mean in $[0, 1]$, $[10, 11]$ with the step size of 0.01, and two standard deviations: 0.2, 1. Each Gaussian curve has a unique combination of mean and standard deviation (Fig. 3b). We compare two popular visualization methods, Isomap⁵¹ and t-SNE, with our proposed Graph-Bootstrapping (GB) method. For clustering, we use K -means and hierarchical density estimates method (HDBSCAN⁵⁷), which is explained in detail in subsequent sections.

From Fig. 3, we can see that the Isomap method could not differentiate Gaussian curves of different standard deviations. For clustering, K -means works correctly on t-SNE and Graph-Bootstrapping manifolds, given the right tuning parameter, K ,

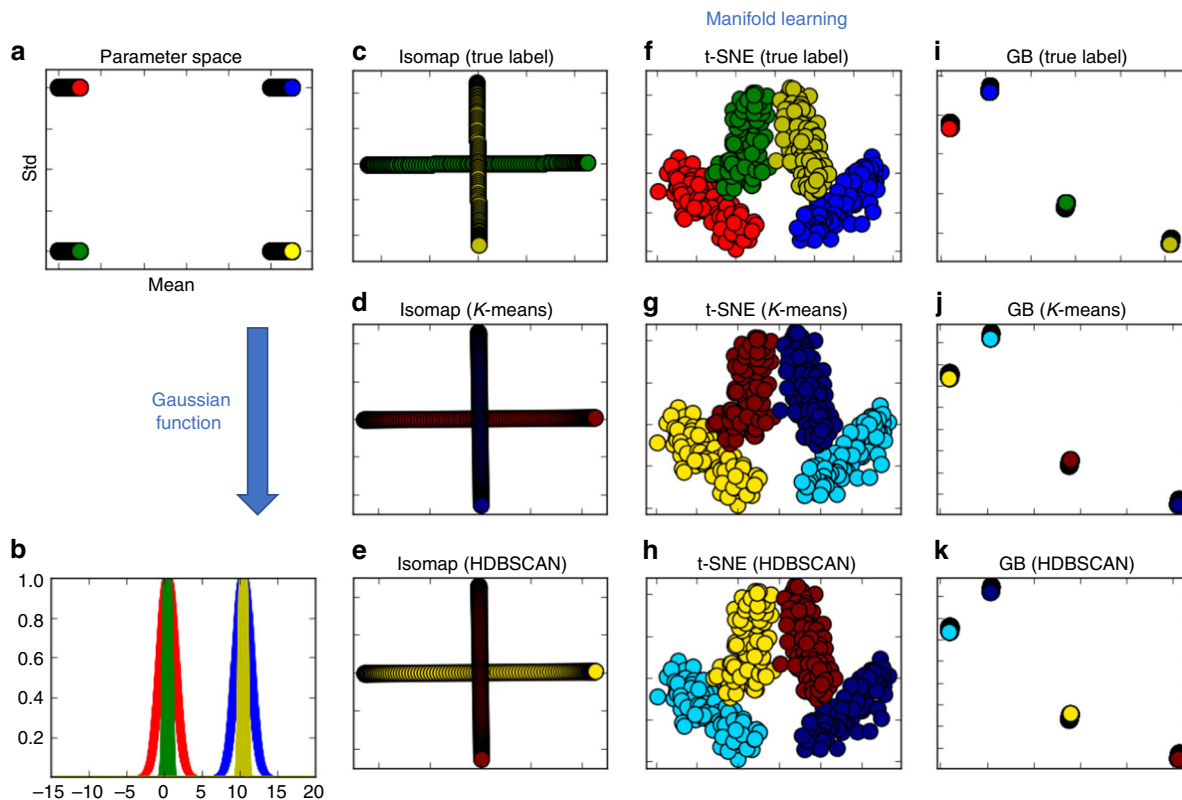


Fig. 3 Illustration of manifold embedding and clustering on simulated Gaussian curves. **a** Four groups of parameters including two ranges of mean in $[0, 1]$, $[10, 11]$ with the step size of 0.01, and two standard deviations of 0.2, 1. **b** Four groups of Gaussian curves, generated from the four groups of parameters in **a**. Each Gaussian curve has a unique combination of mean and standard deviation. **c–e** Manifold points estimated by Isomap⁵¹, colored by true labels and clustered labels via K -means and hierarchical density estimates method (HDBSCAN⁵⁷), respectively. **f–h** Manifold points estimated by t-SNE⁵⁵, colored by true labels and clustered labels via K -means and HDBSCAN, respectively. **i–k** Manifold points estimated by Graph-Bootstrapping (GB), colored by true labels and clustered labels via K -means and HDBSCAN, respectively

the number of clusters. However, the number of clusters is usually an unknown parameter, which in the case of SPM imaging is related to material properties that we would like to estimate. Thus, the manifold learning and clustering should share conjugate hyperparameters that are based on the local structure (such as nearest neighbors), which is the underlying logic behind the Graph-Bootstrapping presented herein.

High-veracity BE-SPM Imaging. Graph-Bootstrapping is firstly applied for analysis of the BE and broadband BE-SPM measurements of the polymer mixture. Graph-Bootstrapping reveals domain differentiation despite the fact that the algorithm is purely statistical in nature and does not require any prior information regarding the material and any differentiating structures. It is also an efficient approach. SHO fittings for a high-dimensional broadband excitation ($p = 15,159$) is prohibitive, yet Graph-Bootstrapping took 20 min to process 4GB of broadband BE-SPM measurements on a single workstation (Intel Xenon E5-1650V3, 32GB DDR3 RAM). Figure 4 compares the results by LargeVis and Graph-Bootstrapping. The relationship between LargeVis and Graph-Bootstrapping clusters is illustrated in following sections. Here we distinguish BE datasets, which contain a single CR peak (narrow band) with broadband resonance BE datasets that capture a broad frequency range spanning several resonance peaks.

Previous studies have shown that complex networks often exhibit hierarchical organizations^{58–60}. Correspondingly, the geometry derived from bootstrapping clearly displays the hierarchical groups within SPM measurements, which allows us to utilize HDBSCAN^{57, 61} to gain insights into many network properties, yielding the spatial mapping of heterogeneous local structures of material. Mathematically, HDBSCAN relies on the mutual reachability distance, which works in a conjugate way with the neighbor exploring stage during graph construction, that is, a

neighbor of a neighbor is also likely to be a neighbor: $D_{\text{mreach},k}(a,b) = \max\{\text{core}_k(a), \text{core}_k(b), d(a,b)\}$, where $d(a,b)$ is the original metric distance between points a and b , $\text{core}_k(x)$ is the core distance of a point x to cover its k nearest neighbors. Supplementary Fig. 1 shows the relationship between the total cluster number and the minimum nearest neighbor number, k . HDBSCAN mostly yields two clusters with a few outliers of three clusters based on LargeVis manifolds (Supplementary Fig. 1). Since there is no clear trend in the resulting cluster number, we choose the majority of total cluster numbers, a value of 2. Supplementary Fig. 2 illustrates the spatial mappings of the two LargeVis clusters.

Compared to LargeVis, the total number of clusters based on the Graph-Bootstrapping manifold decreases continuously as k increases (Supplementary Fig. 1). At the first glance, the resulting number of clusters $P(k)$, varies as a power of k :

$$P(k) = Ck^{-\alpha} \quad (3)$$

where α is called the exponent of the power law that exists in vast branches of natural sciences, computer sciences, economics, and social sciences^{62–67}. To get the fitting of power law distribution, we used the software developed by Alstott et al.⁶⁸. We first normalized the cluster numbers and calculated the log–log plot of probability density function of the empirical data as well as fitted power law (Supplementary Fig. 1). We also fitted the trend with exponential distributions (Supplementary Fig. 1), another popular candidate for heavy-tailed distributions:

$$P(k) = Ce^{-(k-\text{offset})/\tau} \quad (4)$$

where τ is the relaxation constant. For exponential fitting, from $k \approx 3\tau$ (more accurately, it should be $k \approx 3\tau + \text{offset}$, for simplicity, we omit the notation of offset in the rest of paper), the cluster number tends to be stable as k increases.

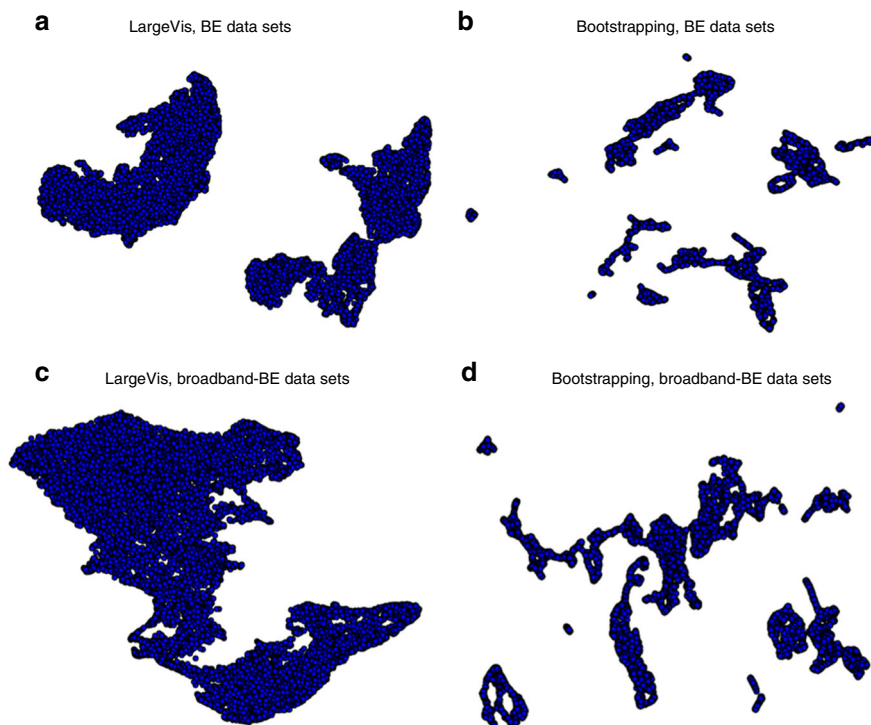


Fig. 4 Manifold layouts by LargeVis and Graph-Bootstrapping calculated from the polymer mixture datasets. **a** Manifold points estimated by LargeVis, based on the band excitation scanning probe microscopy (BE-SPM) datasets. **b** Manifold points estimated via Graph-Bootstrapping by firstly reconstructing the graph from LargeVis manifold in **a** and then recalculating the manifold points based on the reconstructed graph. **c, d** The same analysis procedure with those in **a, b**, but applied on the broadband BE-SPM datasets collected on the same polymer mixture sample

To establish the quality of clustering, corresponding to the number of distinct surface functionalities, we calculated the (cumsum) standard deviations (STD) of SPM measurements within each cluster at different k values. Supplementary Fig. 3 displays the boxplot of cluster STD distributions. Like Supplementary Fig. 1, we can see big jumps around $k \approx 3\tau$, 2τ , τ . Clusters with lower STDs are more stable numerically. On one hand, the newly raised clusters could reveal subtler physical factors that affect materials at higher spatial resolution; on the other hand, the newly raised clusters could be essentially of the same material property, but are divided by SPM measurements noises. Supplementary Fig. 4 displays manifold clusters and the spatial distributions at low- and high-hierarchical levels. We note small inclusions (Supplementary Fig. 4) in the semicrystalline PCL and PS structure, which are likely smaller crystallites formed during the film synthesis, and are useful in highlighting the spatial resolution of the approach. What's more, we see newly bootstrapped clusters as k decreases seem to be randomly located when k is below 2τ

(Supplementary Fig. 4). The above observations are consistent with our conjecture on Supplementary Figs 1, 3. Combining Supplementary Figs. 1, 3, 4, we empirically propose $k \approx 3\tau$ to be the optimal choice between accuracy and overfitting. Or, if we could estimate the measurement noise level, then it may be a good reference. That is to say, we are less faithful in clusters whose standard deviations are less than that of system noise (if obtainable).

Based on above analysis, we can directly denoise the data, visualize it, and extract underlying physics parameters based on the manifold clusters in Supplementary Fig. 5, where the BE signals of bootstrapped clusters and their spatial mappings clearly distill the subtle material surface details, especially for the interface region. One can further calculate the similarity loadings by calculating pairwise distances between the mean SPM curve of the cluster and every raw SPM curve. The resulting similarity loadings (inversion of Euclidean distances) clearly highlight different material regimes (Supplementary Figs. 6, 7), yielding subtler and more enriched local heterogeneity than the SHO

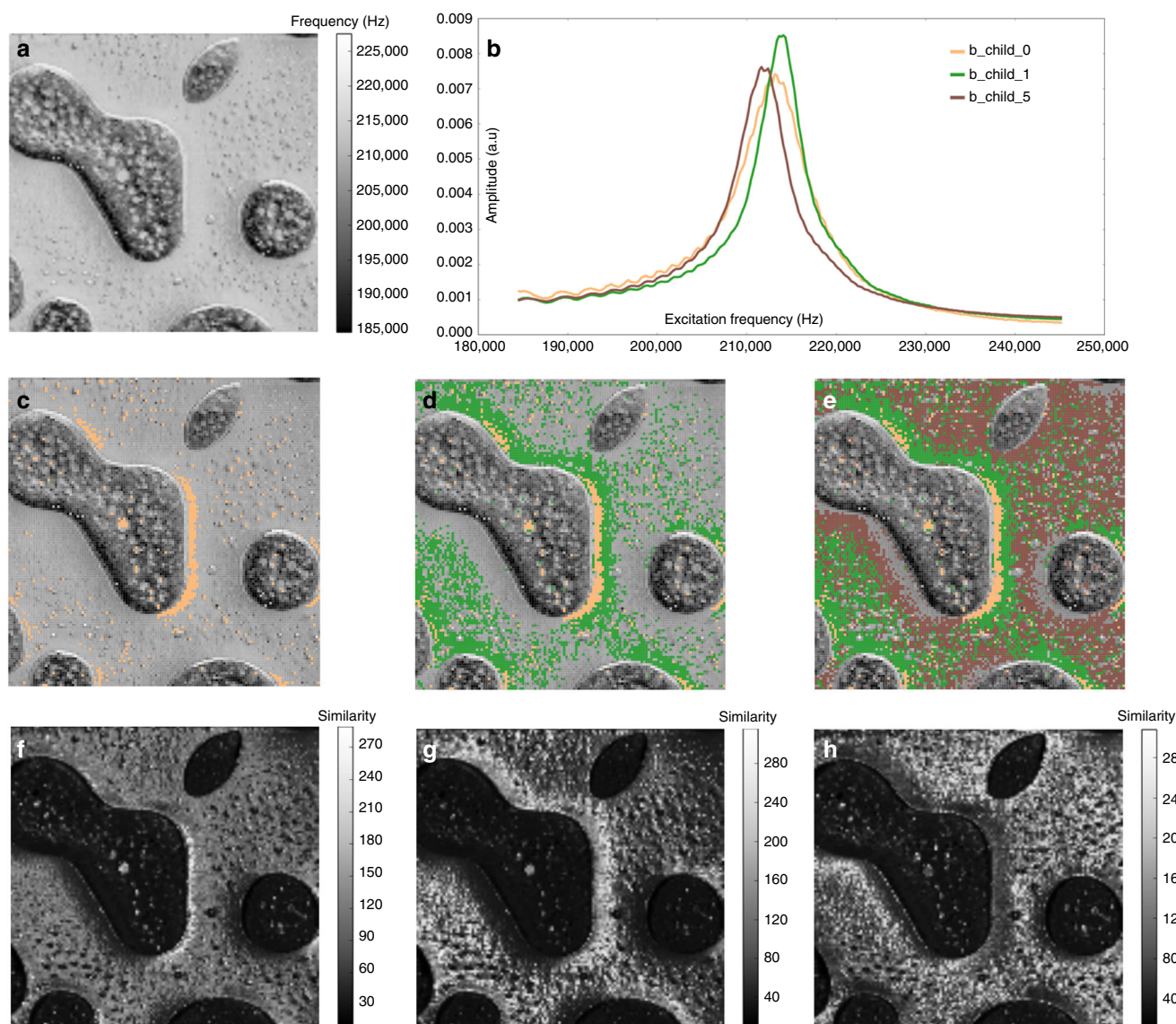


Fig. 5 Application on the polymer mixture sample with band excitation scanning probe microscopy. **a** Spatial mapping of frequency via the simple model of harmonic oscillator (SHO) fitting. **b** Mean SPM response curves of three bootstrapped clusters, b_child_0 , b_child_1 , and b_child_5 . **c–e** Cluster labels (b_child_0 , b_child_1 , b_child_5) cumulatively overlaid on SHO fitted frequency map. **f–h** Similarity loadings of clusters b_child_0 , b_child_1 , and b_child_5 , respectively

fitting. Those similarity loadings can verify the accuracy of the Graph-Bootstrapping method.

In Fig. 5, we present deeper insights obtained via Graph-Bootstrapping. Figure 5a is the spatial mapping of frequency from SHO fittings. Such maps are normally transformed using analytical models into material property maps such as elastic modulus. The brighter area (higher frequency) in Fig. 5a corresponds to the PS matrix of higher stiffness. Noticeably, little structural inhomogeneities are detected within the PS matrix based on Fig. 5a (or in other SHO fitting parameters as shown in Fig. 2). Meanwhile Graph-Bootstrapping illustrates at least three distinguishable clusters (Fig. 5c–e) within the PS matrix with the corresponding SPM response curves (Fig. 5b). To quantitatively check the accuracy of clustering, we provide the similarity loadings in Fig. 5f–h. We do see the similarity loading has higher similarity values at the spatial positions of the cluster. In Fig. 5b, we see the nonlinear ring patterns at lower frequency range which are ignored by definition in the SHO fittings. In this way, our generalized method can be used to unveil hidden details, without the need for an a priori model. We further note the mean SPM curve of cluster b_child_1 has a higher height than the other two clusters, which SHO fittings failed to reveal. We also attached Graph-Bootstrapping results of broad-band BE datasets collected on the same polymer mixture sample in Supplementary Figs. 8, 9, 10, 11.

Application on 3D-AFM Imaging. To further demonstrate the broad capability of Graph-Bootstrapping, we applied it to the 3D-AFM dataset that has been recently investigated by Söngen et al.⁶⁹ to resolve point defects in the hydration structure of calcite (10.4). For each (x, y) surface position, excitation frequencies extracted at different z piezodisplacements (corresponding to various hydration layers) are inputted to the Graph-Bootstrapping.

Figure 6a is the manifold clustering results of the first-order bootstrapping, and correspondingly Fig. 6b is the surface distribution of cluster labels revealing the lattice structure. We performed hierarchical clustering on the second-order bootstrapped manifolds (Fig. 6d) and overlaid the same set of labels onto the first-order manifolds (Fig. 6c) where we can see clusters still aggregate. Figure 6e shows the atomic structure of the (10.4) surface unit cell with a scale bar that applies to all surface spatial mappings. Figure 6f displays the surface positions of the second-order bootstrapped clusters revealing subtle details between the Ca site and the carbonate site, which can be further investigated via the excitation frequency shift profiles of the clusters presented in Fig. 6g. First, we note that the shift curve of cluster_5 consistently exhibits the largest local maxima and the smallest local minima in the third, fourth, and fifth hydration layer, corresponding to the profile of the Ca defect site which has been verified manually by Söngen et al.⁶⁹. Second, Graph-Bootstrapping also elucidated the irregular flattening of the cluster_0 curve between the fourth and fifth hydration layer as well as its large shifts of local maxima and local minima between the second and fourth hydration layer. We can assure this irregularity by comparing surface label sites and similarity loading (Supplementary Fig. 12). We can see that the similarity loading indeed has higher similarity values at the spatial positions of the cluster_0.

Discussion

Over the past 10 years, several groups have made significant contributions towards the quantification of viscoelastic properties by CR SPM^{70–72}. However, even CR-SPM still incessantly undergoes developments to this day, in both modeling and hardware. This is because there is still a lack in knowledge of how the cantilever precisely behaves on certain materials in the

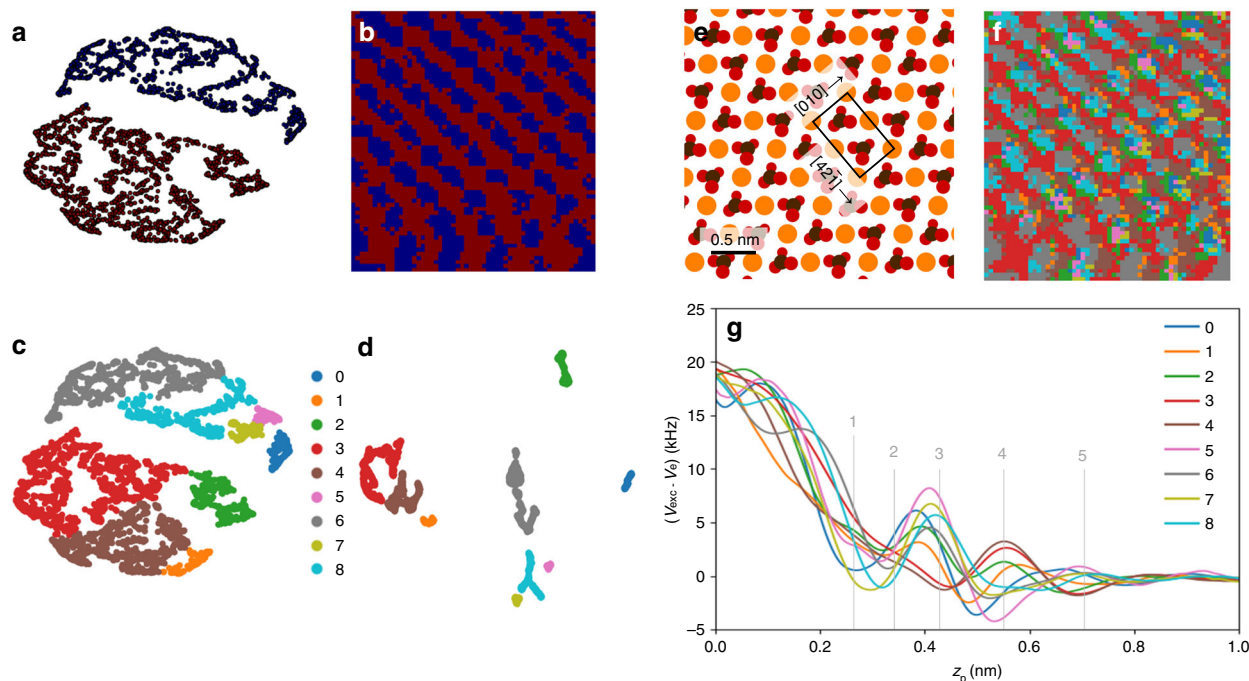


Fig. 6 Application on hydration structure of calcite with 3D-AFM. **a** First-order bootstrapping manifold colored by clustering labels, that is, the manifold derived from the reconstructed graph based on LargeVis manifold and **b** the spatial mapping of the corresponding labels. **c, d** Comparison of manifold patterns (first- and second-order bootstrapping) colored by the same set of clustering labels where hierarchical clustering was performed on the second-order bootstrapping manifold. Second-order bootstrapping manifold is derived from the reconstructed graph based on the first-order manifold. **e** Atomic structure of the calcite (10.4) surface unit cell, reproduced from Ref.⁶⁹. (Copyright [2018], APS). **f** Surface positions and **g** excitation frequency shift profiles of second-order bootstrapping clusters

presence of background forces. We proposed a generalized method to spatially map the local inhomogeneities on the high-dimensional datasets captured in these modes, and to do so without requiring an a priori model or placing any constraints on the data. As we highlight in Fig. 5, even at this stage, we can pinpoint deviations from expected cantilever behavior that are not described by a simple model adopted as the gold standard. This information, in turn, can be used to evaluate the effectiveness of existing approaches. After extracting patterns revealed by Graph-Bootstrapping, further efforts can be made on deeper theoretical and experimental study on correlations between the whole shape of SPM curves and material properties.

To further illustrate the broad capability of Graph-Bootstrapping, we applied it to 3D-AFM and demonstrated its successes on resolving irregular hydration structure of calcite at atomic resolution. In summary, we develop a universal algorithm for representation of high-dimensional SPM data. Based on the combination of network analysis and search for low-dimensional manifold, this approach reveals and hierarchically represents salient material features with rich local details, further enabling denoising, classification, and high-resolution functional imaging. While demonstrated for BE-SPM and 3D-AFM, this approach can be universally applicable to other data-rich imaging methods. We further pose that learning of low-dimensional manifold representing the data will open the pathway for extraction of the relevant physics coupled between material and imaging systems.

Methods

Band-excitation scanning probe microscopy. Mechanical measurements were performed on thin-film polymer blend samples consisting of PCL inclusions embedded in a PS matrix. Samples were prepared by spin coating in 2:1 ratio as described in previous work by Kocun et al.⁴⁸. Film thickness was >300 nm. Mechanical properties of the sample were previously measured using bimodal tapping mode^{48,73} where Young's modulus of the PS was assumed to be 3.0 GPa and used as a reference to determine the stiffness of PCL (~0.85 GPa).

All measurements on polymer sample were performed using a commercial AFM system (Cypher, Asylum Research and Oxford Instruments Company), which was equipped with a laser for photothermal excitation (Bluedrive module) and used for mechanical perturbation of the AFM tip-sample interaction. CR-AFM measurements were performed in contact mode where the tip was held in constant force (~65 nN) with the surface using static deflection feedback. BE-AFM imaging was operated at normal AFM imaging speeds (1 Hz), and the AFM system was coupled with data acquisition and arbitrary waveform generators (National Instruments, NI5122 and NI5412), which were controlled using custom Matlab code (MathWorks). Measurements used a gold-coated NSC36 (Micromasch) cantilever with a calibrated spring constant of 0.89 N/m and free air resonance frequency of 44.7 kHz. The BE excitation center frequency, which was used to excite the photothermal laser position on the AFM cantilever, was chosen to approximately match the CR frequency ($\omega_0 = \sim 210$ kHz) and a bandwidth ~100 kHz using a sampling rate of 4 MHz leading to a data capture of ~20 bins per band. In addition, broadband excitation/detection was implemented, using a center frequency of ~500 kHz and bandwidth of ~490 kHz allowing a larger number of resonance peaks to be captured simultaneously.

Three-dimensional atomic force microscopy (3D-AFM). 3D-AFM measurements on calcite were performed by Söngen et al.⁶⁹ and we refer the reader to this text for additional details.

Code availability. The Graph-Bootstrapping codes used in this study can be made available from the corresponding authors upon request.

Data availability. The data that support the findings of this study are available from the corresponding authors upon request.

Received: 6 January 2018 Accepted: 31 May 2018

Published online: 21 June 2018

References

- Binnig, G., Quate, C. F. & Gerber, C. Atomic force microscope. *Phys. Rev. Lett.* **56**, 930–933 (1986).
- Gerber, C. & Lang, H. P. How the doors to the nanoworld were opened. *Nat. Nanotechnol.* **1**, 3–5 (2006).
- Martin, Y. & Wickramasinghe, H. K. Magnetic imaging by “force microscopy” with 1000 Å resolution. *Appl. Phys. Lett.* **50**, 1455–1457 (1987).
- Belaidi, S., Girard, P. & Leveque, G. Electrostatic forces acting on the tip in atomic force microscopy: modelization and comparison with analytic expressions. *J. Appl. Phys.* **81**, 1023–1030 (1997).
- Bluhm, H., Wadas, A., Wiesendanger, R., Meyer, K.-P. & Szczes-acuteniak, L. Electrostatic force microscopy on ferroelectric crystals in inert gas atmosphere. *Phys. Rev. B* **55**, 4–7 (1997).
- Tanimoto, M. Kelvin probe force microscopy for characterization of semiconductor devices and processes. *J. Vac. Sci. Technol. B, Microelectron. Nanometer. Struct. Process. Meas. Phenom.* **14**, 1547 (1996).
- Vatel, O. & Tanimoto, M. Kelvin probe force microscopy for potential distribution measurement of semiconductor devices. *J. Appl. Phys.* **77**, 2358–2362 (1995).
- Nonnenmacher, M., O’Boyle, M. P. & Wickramasinghe, H. K. Kelvin probe force microscopy. *Appl. Phys. Lett.* **58**, 2921–2923 (1991).
- Liu, X. X., Heiderhoff, R., Abicht, H. P. & Balk, L. J. Scanning near-field acoustic study of ferroelectric BaTiO₃ ceramics. *J. Phys. D: Appl. Phys.* **35**, 74–87 (2002).
- Drobek, T., Stark, R. W. & Heckl, W. M. Determination of shear stiffness based on thermal noise analysis in atomic force microscopy: passive overtone microscopy. *Phys. Rev. B* **64**, 045401 (2001).
- Rabe, U. & Arnold, W. Acoustic microscopy by atomic force microscopy. *Appl. Phys. Lett.* **64**, 1493–1495 (1994).
- Gruverman, A., Auciello, O., Ramesh, R. & Tokumoto, H. Scanning force microscopy of domain structure in ferroelectric thin films: imaging and control. *Nanotechnology* **8**, A38–A43 (1997).
- Gruverman, A. L., Hatano, J. & Tokumoto, H. Scanning force microscopy studies of domain structure in BaTiO₃ single crystals. *Jpn. J. Appl. Phys.* **36**, 2207–2211 (1997).
- Kalinin, S. V., Setter, N. & Kholkin, A. L. Electromechanics on the nanometer scale: emerging phenomena, devices, and applications. *MRS Bull.* **34**, 634–642 (2009).
- Balke, N., Bdikin, I., Kalinin, S. V. & Kholkin, A. L. Electromechanical imaging and spectroscopy of ferroelectric and piezoelectric materials: state of the art and prospects for the future. *J. Am. Ceram. Soc.* **92**, 1629–1647 (2009).
- Clausen-Schaumann, H., Seitz, M., Krautbauer, R. & Gaub, H. E. Force spectroscopy with single bio-molecules. *Curr. Opin. Chem. Biol.* **4**, 524–530 (2000).
- Rief, M. Single molecule force spectroscopy on polysaccharides by atomic force microscopy. *Science* **275**, 1295–1297 (1997).
- Hinterdorfer, P. & Dufrene, Y. F. Detection and localization of single molecular recognition events using atomic force microscopy. *Nat. Methods* **3**, 347–355 (2006).
- Stroh, C. et al. Single-molecule recognition imaging microscopy. *Proc. Natl Acad. Sci. U.S.A.* **101**, 12503–12507 (2004).
- Kalinin, S. V. et al. Local bias-induced phase transitions. *Mater. Today* **11**, 16–27 (2008).
- Jesse, S. et al. Band excitation in scanning probe microscopy: recognition and functional imaging. *Annu. Rev. Phys. Chem.* **65**, 519–536 (2014).
- Arruda, T. M., Kumar, A., Kalinin, S. V. & Jesse, S. The partially reversible formation of Li-metal particles on a solid Li electrolyte: applications toward nanobatteries. *Nanotechnology* **23**, 325402 (2012).
- Kumar, A., Ciucci, F., Morozovska, A. N., Kalinin, S. V. & Jesse, S. Measuring oxygen reduction/evolution reactions on the nanoscale. *Nat. Chem.* **3**, 707–713 (2011).
- Kumar, A., Ovchinnikov, O. S., Funakubo, H., Jesse, S. & Kalinin, S. V. Real-space mapping of dynamic phenomena during hysteresis loop measurements: dynamic switching spectroscopy piezoresponse force microscopy. *Appl. Phys. Lett.* **98**, 202903 (2011).
- García, R. & Pérez, R. Dynamic atomic force microscopy methods. *Surf. Sci. Rep.* **47**, 197–301 (2002).
- Rodriguez, B. J., Callahan, C., Kalinin, S. V. & Proksch, R. Dual-frequency resonance-tracking atomic force microscopy. *Nanotechnology* **18**, 475504 (2007).
- Jesse, S., Nikiforov, M. P., Germinario, L. T. & Kalinin, S. V. Local thermomechanical characterization of phase transitions using band excitation atomic force acoustic microscopy with heated probe. *Appl. Phys. Lett.* **93**, 073104 (2008).
- Jesse, S., Kalinin, S. V., Proksch, R., Baddorf, A. P. & Rodriguez, B. J. The band excitation method in scanning probe microscopy for rapid mapping of energy dissipation on the nanoscale. *Nanotechnology* **18**, 435503 (2007).
- Guo, S., Kalinin, S. V. & Jesse, S. Open-loop band excitation Kelvin probe force microscopy. *Nanotechnology* **23**, 125704 (2012).
- Guo, S., Kalinin, S. V. & Jesse, S. Half-harmonic Kelvin probe force microscopy with transfer function correction. *Appl. Phys. Lett.* **100**, 063118 (2012).

31. Jesse, S. & Kalinin, S. V. Band excitation in scanning probe microscopy: sines of change. *J. Phys. D: Appl. Phys.* **44**, 464006 (2011).
32. Chawla, G. & Solares, S. D. Mapping of conservative and dissipative interactions in bimodal atomic force microscopy using open-loop and phase-locked-loop control of the higher eigenmode. *Appl. Phys. Lett.* **99**, 074103 (2011).
33. Solares, S. D. & Chawla, G. Triple-frequency intermittent contact atomic force microscopy characterization: simultaneous topographical, phase, and frequency shift contrast in ambient air. *J. Appl. Phys.* **108**, 054901 (2010).
34. Platz, D., Tholén, E. A., Pesen, D. & Haviland, D. B. Intermodulation atomic force microscopy. *Appl. Phys. Lett.* **92**, 153106 (2008).
35. Hutter, C., Platz, D., Tholén, E. A., Hansson, T. H. & Haviland, D. B. Reconstructing nonlinearities with intermodulation spectroscopy. *Phys. Rev. Lett.* **104**, 050801 (2010).
36. Fukuma, T., Ueda, Y., Yoshioka, S. & Asakawa, H. Atomic-scale distribution of water molecules at the mica–water interface visualized by three-dimensional scanning force microscopy. *Phys. Rev. Lett.* **104**, 016101 (2010).
37. Fukuma, T. et al. Mechanism of atomic force microscopy imaging of three-dimensional hydration structures at a solid–liquid interface. *Phys. Rev. B* **92**, 155412 (2015).
38. Collins, L. et al. Quantitative 3D-KPFM imaging with simultaneous electrostatic force and force gradient detection. *Nanotechnology* **26**, 175707 (2015).
39. Tetard, L. et al. Imaging nanoparticles in cells by nanomechanical holography. *Nat. Nanotechnol.* **3**, 501–505 (2008).
40. Stark, M., Stark, R. W., Heckl, W. M. & Guckenberger, R. Inverting dynamic force microscopy: from signals to time-resolved interaction forces. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8473–8478 (2002).
41. Giridharagopal, R. et al. Submicrosecond time resolution atomic force microscopy for probing nanoscale dynamics. *Nano Lett.* **12**, 893–898 (2012).
42. Collins, L. et al. Breaking the time barrier in Kelvin probe force microscopy: fast free force reconstruction using the G-Mode platform. *ACS Nano* **11**, 8717–8729 (2017).
43. Belianinov, A., Kalinin, S. V. & Jesse, S. Complete information acquisition in dynamic force microscopy. *Nat. Commun.* **6**, 6550 (2015).
44. Collins, L. et al. Full data acquisition in Kelvin probe force microscopy: mapping dynamic electric phenomena in real space. *Sci. Rep.* **6**, 30557 (2016).
45. Jesse, S., Somnath, S., Collins, L. & Kalinin, S. V. Full information acquisition in scanning probe microscopy. *Microsc. Today* **25**, 34–45 (2017).
46. Jesse, S. & Kalinin, S. V. Principal component and spatial correlation analysis of spectroscopic-imaging data in scanning probe microscopy. *Nanotechnology* **20**, 085714 (2009).
47. Jesse, S., Baddorf, A. P. & Kalinin, S. V. Dynamic behaviour in piezoresponse force microscopy. *Nanotechnology* **17**, 1615–1628 (2006).
48. Kocun, M., Labuda, A., Meinhold, W., Revenko, I. & Proksch, R. Fast, high resolution, and wide modulus range nanomechanical mapping with bimodal tapping mode. *ACS Nano* **11**, 10097–10105 (2017).
49. Newman, M. E. J. Communities, modules and large-scale structure in networks. *Nat. Phys.* **8**, 25–31 (2011).
50. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7821–7826 (2002).
51. Tenenbaum, J. B., de Silva, V. & Langford, J. C. A global geometric framework for nonlinear dimensionality reduction. *Science* **290**, 2319–2323 (2000).
52. Tang, J., Liu, J., Zhang, M. & Mei, Q. Visualizing large-scale and high-dimensional data. In *Proc. of the 25th International Conference on World Wide Web* 287–297 (International World Wide Web Conference Committee, 2016).
53. Dasgupta, S. & Freund, Y. Random projection trees and low dimensional manifolds. In *Proc. of the Fourtieth Annual ACM Symposium on Theory of Computing* 537–546 (ACM, 2008).
54. Dong, W., Moses, C. & Li, K. Efficient k-nearest neighbor graph construction for generic similarity measures. In *Proc. of the 20th International Conference on World Wide Web* 577–586 (International World Wide Web Conference Committee, 2011).
55. Van Der Maaten, L. J. P. & Hinton, G. E. Visualizing high-dimensional data using t-sne. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
56. Recht, B., Re, C., Wright, S. & Niu, F. Hogwild: a lock-free approach to parallelizing stochastic gradient descent. *Adv. Neural Inf. Process. Syst.* **24**, 693–701 (2011).
57. McInnes, L., Healy, J., & Astels, S. hdbSCAN: Hierarchical density based clustering. *J. Open Source Softw.* **2**, 205 (2017).
58. Ravasz, E. et al. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555 (2002).
59. Sales-Pardo, M., Guimera, R., Moreira, A. A. & Nunes Amaral, L. A. Extracting the hierarchical organization. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15224–15229 (2007).
60. Clauset, A., Moore, C. & Newman, M. E. J. Hierarchical structure and the prediction of missing links in networks. *Nature* **453**, 98–101 (2008).
61. Campello, R. J. G. B., Moulavi, D., Zimek, A. & Sander, J. Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Trans. Knowl. Discov. Data* **10**, 1–51 (2015).
62. Newman, M. E. J. Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **46**, 323–351 (2005).
63. West, J. B., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126 (1997).
64. Buzsáki, G. & Mizuseki, K. The log-dynamic brain: how skewed distributions affect network operations. *Nat. Rev. Neurosci.* **15**, 264–278 (2014).
65. Humphries, N. E. et al. Environmental context explains Lévy and Brownian movement patterns of marine predators. *Nature* **465**, 1066–1069 (2010).
66. Bolmatov, D., Brazhkin, V. V. & Trachenko, K. Thermodynamic behavior of supercritical matter. *Nat. Commun.* **4**, 1–7 (2013).
67. Albert, R., Jeong, H. & Barabási, A. L. Diameter of the World-Wide Web. *Nature* **401**, 130–131 (1999).
68. Alstott, J., Bullmore, E., & Plenz, D. powerlaw: A python package for analysis of heavy-tailed distributions. *PLoS ONE* **9**, e85777 (2014).
69. Söngen, H. et al. Resolving point defects in the hydration structure of calcite (10.4) with three-dimensional atomic force microscopy. *Phys. Rev. Lett.* **120**, 116101 (2018).
70. Yuya, P. A., Hurley, D. C. & Turner, J. A. Contact-resonance atomic force microscopy for viscoelasticity. *J. Appl. Phys.* **104**, 074916 (2008).
71. Gannepalli, A., Yablon, D. G., Tsou, A. H. & Proksch, R. Mapping nanoscale elasticity and dissipation using dual frequency contact resonance AFM. *Nanotechnology* **22**, 355705 (2011).
72. Killgore, J. P. et al. Viscoelastic property mapping with contact resonance force microscopy. *Langmuir* **27**, 13983–13987 (2011).
73. Labuda, A., Kocun, M., Meinhold, W., Walters, D. & Proksch, R. Generalized Hertz model for bimodal nanomechanical mapping. *Beilstein J. Nanotechnol.* **7**, 970–982 (2016).

Acknowledgments

The research was performed at the Center for Nanophase Materials Sciences, a U.S. Department of Energy, Office of Science User Facility at Oak Ridge National Laboratory. T.F. and K.M. acknowledge support from the World Premier International Research Center Initiative (WPI), MEXT, Japan. The authors gratefully acknowledge Hagen Söngen and Angelika Kühnle for help and discussion on 3D-AFM datasets.

Author contributions

X.L. and S.V.K. conceived and designed the study. X.L. developed Graph-Bootstrapping methodology and performed the associated data analysis. L.C. and S.J. performed SHO fitting analysis on BE-SPM datasets. K.M. and T.F. provided 3D-AFM measurements of calcite. S.V.K. and S.J. supervised the study. X.L. and S.V.K. co-wrote the article with comments from all authors.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-04887-1>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article 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. The images or other third party material in this article 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 <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018