

Big, Deep, and Smart Data in Scanning Probe Microscopy

Sergei V. Kalinin,^{*,†,‡} Evgheni Strelcov,^{†,‡} Alex Belianinov,^{*,†,‡} Suhas Somnath,^{†,‡} Rama K. Vasudevan,^{†,‡} Eric J. Langerfelt,^{†,§} Richard K. Archibald,^{†,§} Chaomei Chen,^{||} Roger Proksch,[⊥] Nouamane Laanait,^{†,‡} and Stephen Jesse^{*,†,‡}

[†]The Institute for Functional Imaging of Materials, [‡]The Center for Nanophase Materials Sciences, and [§]Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831, United States

^{||}College of Computing and Informatics, Drexel University, Philadelphia, Pennsylvania 19104, United States

[⊥]Asylum Research, an Oxford Instruments Company, Santa Barbara, California 93117, United States

ABSTRACT: Scanning probe microscopy (SPM) techniques have opened the door to nanoscience and nanotechnology by enabling imaging and manipulation of the structure and functionality of matter at nanometer and atomic scales. Here, we analyze the scientific discovery process in SPM by following the information flow from the tip–surface junction, to knowledge adoption by the wider scientific community. We further discuss the challenges and opportunities offered by merging SPM with advanced data mining, visual analytics, and knowledge discovery technologies.

Since the invention of scanning tunneling microscopy (STM) and atomic force microscopy (AFM)¹ 35 and 30 years ago, respectively, scanning probe microscopy (SPM) techniques have become irreplaceable for studying and controlling the nanoworld,² making Richard Feynman's statement "there is plenty of room at the bottom"³ ring truer than ever before. The collective efforts of commercial, academic, and government institutions have created a fleet of SPM platforms⁴ and enabled a broad range of studies, such as quantum transport imaging in low-dimensional systems,^{5–7} functional magnetic and ferroelectric studies,^{8–11} atomically resolved imaging of conductors and insulators,^{1,12,13} imaging active device structures,^{14–16} single-molecule reactions,^{17,18} biological recognition imaging,^{19,20} and many others. In fact, it is true to say that the relatively low cost and ease of use of SPMs made them the workhorse of nanoscience.

The crux of all SPM techniques is a localized probe, precisely positioned to interact with a surface, connected to the macroscopic world *via* various sensors and associated electronics to record a plethora of observables (*e.g.*, force, displacement, voltage, current, electromagnetic radiation, *etc.*). Scanning probe microscopy measurements acquire information locally, in order to build an ensemble of the material's functionality block by block. Traditionally, SPM capabilities and performance are delineated in the context of individual components, like the characteristics and functionality of the probe, stability and noise of the positioning system, control of the sample environment, and the type of measurements accessible (*i.e.*, probe excitation schemes and data acquisition modes).



Below, we briefly summarize these functional elements in the framework of information generation at the tip–surface junction and information transfer to the macroscopically observed world.

One of the main SPM technology thrusts is probe development. Early SPM modes focused on exploiting mechanical, van der Waals, and elastic interactions between the probe and the surface.^{21,22} Development of magnetic and chemically functionalized probes enabled new microscopy modes, which essentially filter and amplify probe–surface interactions through probe selectivity.^{23–25} Later, the development of batch-fabricated silicon probes²⁶ enabled much higher reproducibility between individual SPMs. As the field grew, modulated excitation of the SPM probes, coupled with heterodyne detection schemes, enabled high-resolution measurements of mechanical and electrical properties.^{27,28}

SPM platform development has pursued, and continues to pursue, two goals: measurement stability and environmental control. Broadly defined, platform stability describes noise contribution across a wide frequency range, where low frequencies affect results globally, like image drift, and high frequencies corrupt data locally, within a single pixel. Environmental controls, such as temperature, pressure, chemical surroundings, *etc.*, broaden the range of addressable material phenomena. Notably, the environmental controls and the SPM platform stability are linked, as the measurement environment fundamentally affects

Received: June 24, 2016

Published: September 27, 2016

noise and, thus, performance. For example, low temperatures reduce the intrinsic Brownian motion and the coefficients of thermal expansion that affect drift. Environmental factors beyond temperature, such as pressure, humidity, and atmosphere, diminish the unwanted and uncontrolled chemical side reactions and may reduce mechanical degrees of freedom that lead to ill-defined electrochemical reactions²⁹ or poorly understood boundary conditions like water layers on surfaces of materials in ambient SPM systems.³⁰ Additionally, a poorly controlled environment may have deleterious effects similar to improper grounding or gating in transport measurements.

Ultimately, the development of SPM has given rise to a broad variety of imaging modes based on the specificity of the probe and signal modulation. Examples of these imaging schemes include contact, intermittent contact, and noncontact modes for mechanical probe excitation.²⁸ These operational modes may utilize amplitude and frequency modulation, multiple line pass operation, and, in the past decade, multiple frequency excitation modes.²⁷ These broadband excitation modes like Band excitation^{31–33} and G-mode^{34–36} techniques are examples of probe modulation schemes developed to push the limits of frequency- and time-resolved imaging, allowing full information capture at the tip–surface junction. These, and many other similar platform developments, are keeping pace with the underlying developments in information technology, such as faster computer-based acquisition, larger storage capabilities, and a growing arsenal of analyses methods, as described by Mody in *Instrumental Community*.⁴

In this Nano Focus, we analyze the data captured by SPM platforms from the vantage point of information generation, its deconvolution, and transfer from the tip–surface junction to the collective knowledge of the scientific community. We discuss the challenges and opportunities offered by merging SPM with advanced data mining, visual analytics, and knowledge-discovery technologies enabled by information technology.

SCANNING PROBE MICROSCOPY AS A CONTRIBUTOR TO THE SCIENTIFIC PROCESS

Scanning probe microscopy allows material characterization at a continuum of scales from atomic to nano- to mesoscopic lengths and provides two types of information generally categorized as morphological or structural imaging and as functional imaging (Figure 1). These two types of informational modes cover a range of time scales, from nanoseconds—for a single period of electromagnetic interaction in a high-frequency capacitance or near-field optical measurements—to microseconds for cantilever deflection data, to milliseconds of the mechanical response of the microscope and associated feedback systems (Figure 2). Information processing in the first step, from microscope to controller, is typically automatic, and the signals are preconditioned using filters, lock-in amplifiers, and phase-locked loops (PLLs) along with various feedback loops. These signal-processing steps limit and bias the nano- and microtime scale-generated information by averaging the response, in return for an improved signal-to-noise ratio. Early in the development of the SPM systems, due to low platform stabilities, these data-capturing techniques were fully justified. Today, however, the microscopes are much more stable than even a decade ago, yet the basics of the data capture and processing at this first, fundamental level have remained largely unchanged. More importantly, in the classical heterodyne detection, the data are analyzed and captured at a single frequency, yielding a single set of amplitude and phase measurements

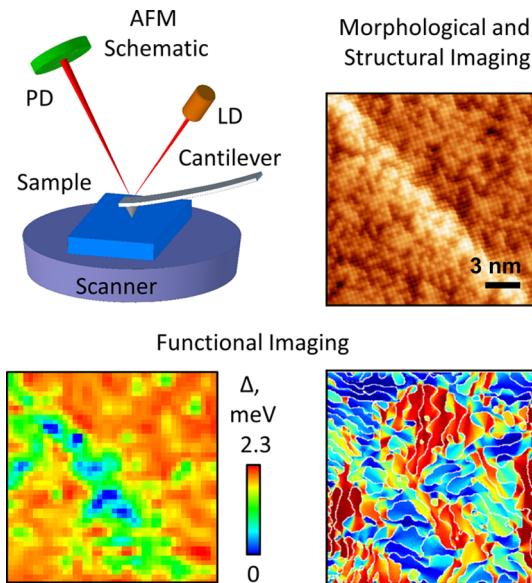


Figure 1. SPM: A schematic of atomic force microscope (PD stands for “photodetector” and LD “laser diode”) and some SPM characterization examples of morphological and functional imaging. The bottom row shows a map of a superconducting gap in $\text{FeTe}_{0.55}\text{Se}_{0.45}$ (reprinted from ref 37. Copyright 2013 American Chemical Society) and a map of elaborate ferroelectric domains in a relaxor ceramic (reprinted with permission from ref 38. Copyright 2016 The Royal Society of Chemistry).

(details of various detection methods are shown Table 1). Consider the case of a basic approximation of the cantilever as the SHO, the knowledge of the driving force is required, rendering the model system underdetermined.^{32,37} Measuring at two or more frequencies, using a technique known as dual-frequency amplitude-resonant tracking mode (DART),³⁸ obviates this limitation, but implicitly relies on the linearity of the system response. These basic approximations in early stage data treatment and other restraints will be discussed in forthcoming sections, with the detection principles described in more detail in Table 2.

At the second level of the measurement process is the interaction between the controller and the human operator (Figure 2). The associated time scales here are more fluid, ranging from seconds to hours, and ill-defined procedures like tuning the microscope, preliminary imaging and interpretation, and selection of regions of interest for detailed studies can all be broadly described as the interactive decision-making process.

The first two data flow steps, microscope to controller and controller to operator, are typically followed by detailed quantitative analysis, interpretation, and finally, dissemination and publication. At the publication stage, the data generated by the microscope are curated and transformed to material-specific information presented in the context of general knowledge. These latter stages can take any amount of time from days to months or more. During the analysis and publication preparation steps, the data and its interpretation are typically only available to the microscope operator and immediate collaborators. As a result, the vast majority of data, acquired at the second step, is rarely reused and re-examined, with final interpretation gravitating toward the original hypotheses and assumptions. Generally, serendipitous phenomena are explored only if they are associated with clearly visible signatures in the images or spectroscopic responses, limited by the knowledge of

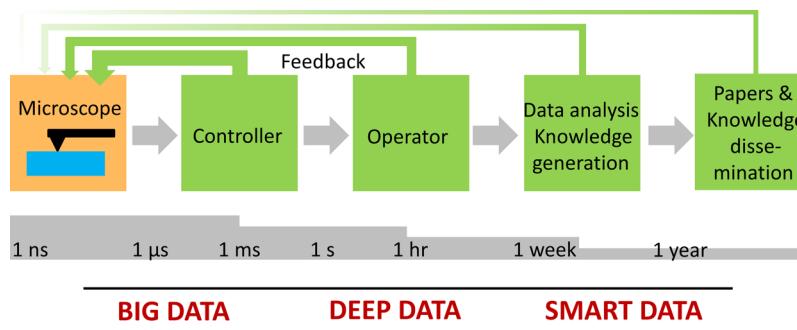


Figure 2. Information flowchart: For effective operation, the data-generating AFM system requires feedback from multiple sources. The controller and operator (on the time scale of microseconds to hours) provide the principal feedback. The feedback from the researcher analyzing the collected data is less impactful (on the already acquired measurement), and feedback from the whole community is weaker still, taking months or years to take full effect. The concept of smart data here implies timely and effective feedback of the scientific community on operation of individual machines *via* machine-learning techniques and real-time dissemination of acquired data.

Table 1. Comparison of AFM Data Acquisition Methods

excitation	detection	feedback	refs
$\omega = \text{const}$	lock-in: $A_{\omega}, \theta_{\omega}$	yes	1
$\omega = \text{const}$	lock-in: $A_{\omega}, \theta_{\omega}$	yes	53
$\omega = \text{const}$	lock-in as a carry along measurement (force modulation)	no	54, 55
	harmonics lock-in: $A_{n\omega}, \theta_{n\omega}$	yes	
	higher resonant mode lock-in: $A_{\omega_2}, \theta_{\omega_2}$	yes	56
ω	phase locked loop: ω_r, A_r	yes	57
ω, Q	ring-down	yes	58, 59
$\omega_1, \omega_2 = \omega_1 + \delta$	intermodulation harmonics, $A_{\omega \pm n\delta}, \theta_{\omega \pm n\delta}$	yes, one amplitude	51
$\omega_1, \omega_2 = \omega_1 + \delta$	DART, ω_r, A_r, Q	yes, but usually on deflection	60
ω_1, ω_2	bimodal excitation	yes, usually on A1	theory: 61; experimental: 62
broad-band excitation	thermal (white) noise	no	63
spectroscopic in omega: $\omega \in [\omega_1, \omega_2]$	Fourier analysis	no	64
spectroscopic in omega: $\omega \in [\omega_1, \omega_2]$	fast sweeps	no	65
spectroscopic in omega: $\omega \in [\omega_1, \omega_2]$	band excitation, $\{A(\omega), \theta(\omega)\}$	no	38
spectroscopic in z:	"z(t) vs z base(t)", otherwise known as a "force curve"		66
spectroscopic in z:	fast force curves (pulsed force, peak force, fast force mapping)		67
spectroscopic in z etc.	specific height (z) position measurements (lift, nap, etc.)		68

^aAmplitude and phase on excitation frequency $A_{\omega}, \theta_{\omega}$; amplitude and resonant frequency, ω_r, A_r ; harmonics of excitation frequency $A_{n\omega}, \theta_{n\omega}$; $[\omega_1, \omega_2]$ denotes interval from ω_1 to ω_2 sampled at N bins.

those immediately involved in data collection, at the worst, and by associated collaborators, at best. Furthermore, if an object or a signature of interest is detected at the analysis stage, replicating the measurement in the same region of interest is often challenging, if not impossible. Therefore, the process of information and knowledge generation (the third step in Figure 2) is associated with significant information loss and artificial researcher-based compression, latency, and interpretation bias.

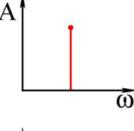
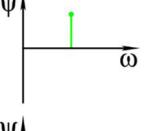
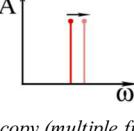
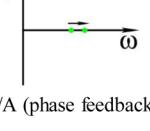
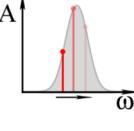
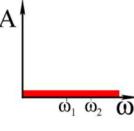
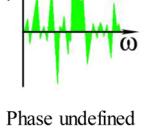
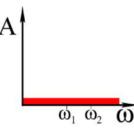
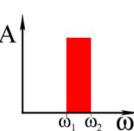
The final level in Figure 2 is knowledge dissemination and its adoption by the broader scientific community. This process is complex and slow, traditionally based on publications, interpersonal communication, conference presentations, and other forms of research-related social interactions. The propagation of new knowledge is often gradual and is strongly affected by scientific and social factors.^{39–42} The practical consequence is a lack of global and timely community feedback since whatever significant feedback does occur happens at the postpublication level, with only the published data subset available for examination. Similarly, venues for feedback and the publication of negative results are limited. Multiple recent initiatives devoted to the open data concepts (beyond “comment” sections in scientific journals) are being suggested now and will be mentioned below.

In this Nano Focus, we discuss the data generation and data flow in SPM from the tip–surface junction to the scientific community, in view of the recent developments in data mining, visual analytics, and information analysis technologies. Primarily, we focus on full information capture in SPM measurements and opportunities opened by the incorporation of faster feedback at all stages of imaging process. We adopt a classification scheme based on big, deep, and smart data concepts.⁴³ Here, *big data* refers to acquisition, data mining, correlation analysis, unsupervised classification, and learning on large data volumes. *Deep data* refers to the transition from correlative to causative data and system analytics, including inferring the mechanism or material specific functionality from the data. Finally, *smart data*, refers to the process of establishing near real-time dissemination and community feedback of the entire data volume that is bias free, with all meta-data open to scrutiny.

BIG DATA IN SCANNING PROBE MICROSCOPY

The nature of the measurement process in SPM offers an ideal model system for analysis of data generation and its flow from the probing volume of the tip–surface interaction, directly to community-wide knowledge. The classical SPM modes are

Table 2. Schematic Comparison of Data Acquisition and Control Methods

Method	Amplitude control	Phase control
Imaging (single frequency)		
Lock-in detection		
Single frequency, phase-locked loop		
Spectroscopy (multiple frequency)		
Lock-in sweep/fast tune		No phase control of excitation signal
Thermal excitation		
Single pulse		Phase undefined
Band excitation		No phase control
		Precise tailoring of the signal

We discuss the data generation and data flow in SPM from the tip–surface junction to the scientific community, in view of the recent developments in data mining, visual analytics, and information analysis technologies.

based on scalar, time-dependent excitation signals, where voltage is applied to the piezo actuators driving the probe, and/or directly to the tip. Similarly, in the simplest cases, the detection signal is also scalar, *e.g.*, vertical and lateral displacement in AFM, along with the tip–surface current in STM. Additionally, these and other, more advanced, information channels can be excited and analyzed simultaneously, giving rise to multimodal data sets. These factors all contribute to the ease with which high-dimensional, large spectroscopic data sets can be acquired in SPM.

Structure and Capacity of the Data Channels in Scanning Probe Microscopy. The data generation rate in SPM is presently detector limited to ~10 MHz for optical detectors and 1–10 kHz for current-based measurements. These throughput values, typically scaled to 32 bits of precision, either by the data acquisition card or the underlying computer operating system, correspond to data flow rates of up to 40MB/s.

On average, a single high-quality SPM image is collected in approximately 5 min at 512×512 pixels, whereas in scanning transmission electron microscopy (STEM) a single $4k \times 4k$ pixel image is captured in well under a second. Furthermore, an image produced by SPM in 5 min, at the full detector capture rate, can take up ~10 GB.^{34,44} For the sake of perspective, in a STEM equipped with a fast, large area pixelated detector, a single image size explodes to more than 500 GB.^{44,45} While these are seemingly astronomical values for a single image, the data flow is amplified by orders of magnitude for any spectral, or otherwise higher dimensionality, experiments in both STEM and SPM; rapidly approaching the raw data output rates of an experiment (*e.g.*, ATLAS) at the Large Hadron Collider.⁴⁶ Remarkably, the data generation rate is so high that capturing the full data stream over typical image times goes well beyond the data-handling capabilities of a dedicated computing workstation. In principle, in AFM, there are many strategies to reduce data set sizes, such as limiting sampling to the inverse resonant frequency, but this process inevitably results in information loss.

The information content in the data stream can be defined in many ways. At the most basic level, it can be linked to the level of materials-specific signal and system noise; however, such definition requires (often *a priori*) materials-specific knowledge, *i.e.*, definition of the “ideal” image for specific imaging modes, conditions, and materials systems. Practically, the achievable information limit in AFM can be estimated based on a typical

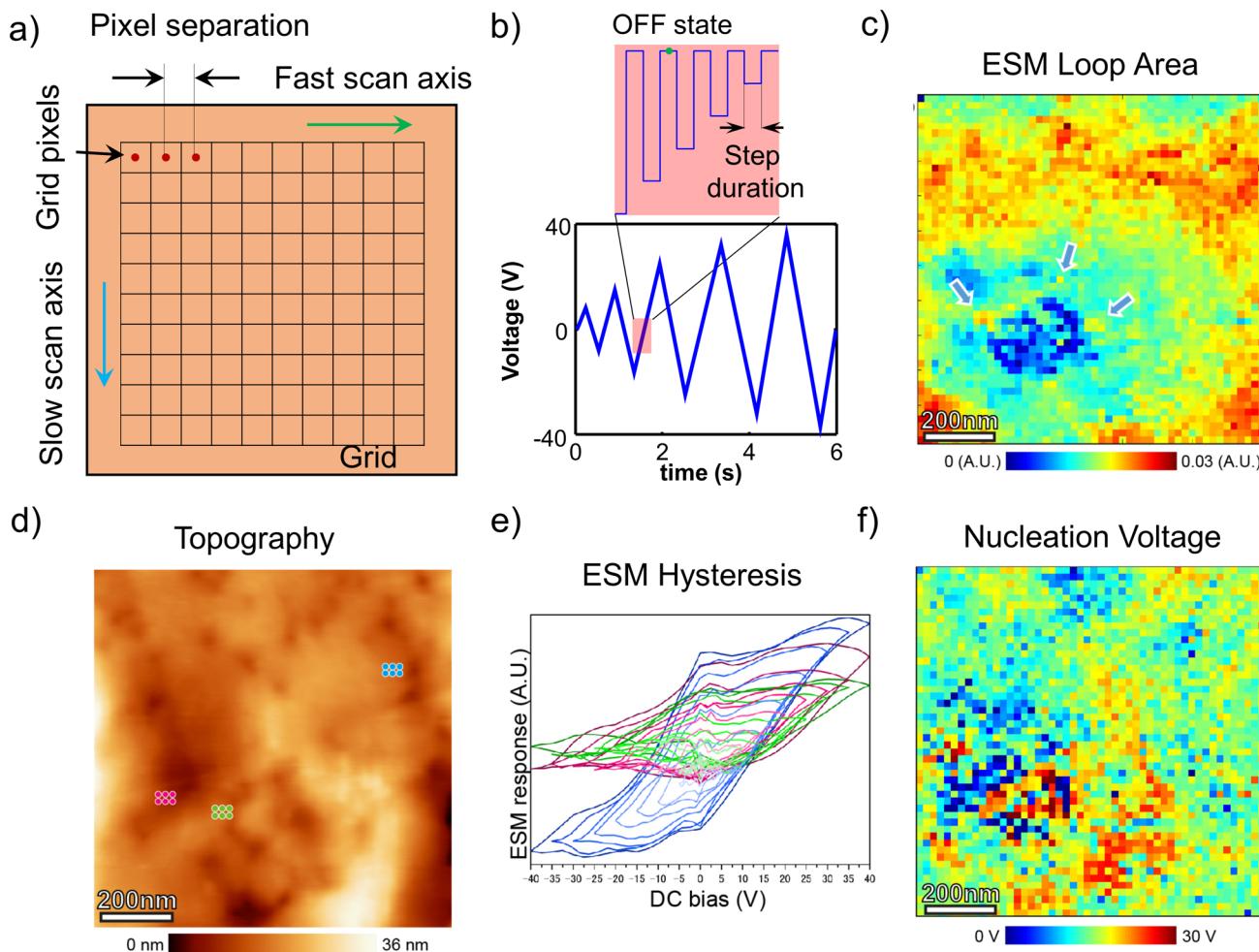


Figure 3. An example of AFM spectral imaging: (a) For spectral imaging, the AFM tip is sequentially positioned at each pixel of a rectangular grid of points that overlays the sample's region of interest; (b) a voltage waveform such as shown here is applied to the tip, and the response of the sample (cantilever deflection) is recorded for each pixel of the grid and each voltage value; (c) response hysteresis loop area or (f) nucleation voltage can be thus mapped out; (e) response hysteresis loops can also be plotted and correlated to (d) the topography of the sample.

oscillation amplitude of the cantilever (~ 0.1 nm in contact modes, 50–100 nm in noncontact modes), in addition to the magnitude of thermal noise in the system. Traditionally, the first step in processing the force-based AFM data has been the heterodyne filtering (lock-in or PLL) that compresses the ~ 10 MHz optical detector data stream to ~ 1 kHz of amplitude/phase or frequency/amplitude data. Heterodyne detection methods have been the bedrock of SPM since its original development and remain the mainstay of detection methods in the vast majority of the commercial systems.

Despite their popularity, single-frequency methods that utilize on-resonance cantilever excitation suffer from the fact that the system is fundamentally under-determined, *i.e.*, even in the simple harmonic oscillator (SHO) approximation, the system has three unknowns, whereas only two are measured.³² In some cases, knowing the driving force provides the remaining unknown, rendering the system fully determined. However, the frequency dispersion in the transfer function of the piezoactuator and the nonlinearity of the system limit the precision with which the driving force can be recovered, which often leads to poorly understood and spurious image contrast,³⁷ with notable examples in magnetic dissipation imaging^{47,48} and contrast in phase imaging.²⁸ In these cases, high-quality images can be readily

generated, but the imaged features depend strongly on the instrument, the probe, and the environmental conditions, which are generally irreproducible. This ambiguity led to the development of a broad spectrum of frequency scanning⁴⁹ or parallel detection³² methods as well as the amplitude-based feedback mode DART.⁵⁰ Correspondingly, the fact that a system can be nonlinear inspired the development of multiple frequency detection modes.^{27,51,52}

Henceforth, despite the large variety of the dynamic-force-based AFMs, they can be classified by their excitation and detection paradigms. Perhaps the most important detail is that the vast majority of existing quantitative data analysis methods in SPM rely on the assumption of the SHO-type response, which in and of itself biases data interpretation. Aside from that issue, the basic data flow requirements for multifrequency measurements are similar to single-frequency experiments, with more complex excitation waveforms and multimodal (post-compression) detection channels.

What followed in the development of the dynamic SPM modes was prompted by the need to capture full cantilever response in a frequency interval. This was achieved by thermal excitation methods,⁶⁹ but white noise led to the loss of the phase information. Ring-down methods enabled the capture of

spectral response, but at the cost of low signal intensity.^{58,59} Frequency sweep methods^{49,70,71} enable full acquisition of frequency response, but are generally slow (note that these techniques are likely to see a resurgence with the adoption of full information capture modes). Finally, the band-excitation (BE) mode³¹ introduced the concept of parallel excitation and multiple frequency detection, akin to multiple parallel lock-in detection.⁵² In BE and rapid frequency sweeps, the data flow is increased to \sim 100 kHz, and the excited band response is stored, begetting spectral image data sets where the response, position, and frequency are captured simultaneously. Notably, the extraction of a cantilever response is achieved by fitting it to the SHO model following the experiment, and only coarsely determined parameters are available immediately.

Finally, the full capture of the data flow in dynamic SPM channel was introduced in the recently developed G-mode.^{34–36,44,72} In G-mode, the full data stream of the photodetector is captured continuously at the highest detector-allowable rate. This mode enables imaging and probing of material properties on the time scales of a single cantilever oscillation, at the cost of large data volumes. The subsequent analysis can yield classical SPM observables (e.g., amplitude of the response at the excitation frequency), more complex parameters including harmonics, and other, generally extant Fourier components. Alternatively, the data can be analyzed and then compressed using classical information theory methods, such as principal component analysis (PCA).⁷³ Remarkably, G-mode allows one to estimate the information content in the data flow, based on either standard information theory criteria or by exploring the presence of spatially correlated features in the loading maps.⁷⁴ Presently, this analysis can only be performed after the data are acquired. However, significant efforts are directed to making G-mode data analysis possible in real-time.^{75,76} Similarly, extension of BE and G-mode to probing responses with laser spots positioned along multiple points of the cantilever will enable quantitative probing of SPM dynamics.

Overall, with significant improvements in the SPM platform stability, environmental control, and the underlying information technology governing data acquisition, storage, and control, we envision a shift away from classical heterodyne detection methods toward broadband signal collection and multimodal analysis.

Scanning and Spectroscopic Modes. The nature of the imaging process in SPM dictates that the measurements are performed over a set of spatial points in a defined parameter space (see Figure 3). These are typically rectangular grids of locations addressed sequentially by the probe, giving rise to the concepts of slow and fast scan axes, labeled in Figure 3a. In the higher dimensional spectroscopic imaging modes, the signal is measured as a function of spatial position and applied *stimulus*, producing hyperspectral data sets. Since the spectroscopic process is sequential, these measurements can be classified based on the sequence in which the parameter space is sampled. For example, (x, y, V) denotes the measurements where the voltage response is measured at each spatial location x, y , with x representing the slow scan, and y representing the fast scan axis, as shown in Figure 3b. On the other hand, (V, x, y) corresponds to a stack of images acquired sequentially at different voltages. The microscope operator identifies and selects the experimental parameters and the way the data are written. This, in turn, requires metadata of experimental details for proper experiment reconstruction and broader data dissemination (to be discussed later), as shown in the reconstruction of the regions of interest (Figure 3c–f).

An important consideration, in imaging and in spectroscopic modes, is data accumulation time for a single point in parameter space. This time value determines the trade-off between data acquisition rate and data quality. Faster acquisition produces a robust system for probing dynamics. Notably, the nature of the system noise, with $1/f$ noise contribution obviating the benefits of image acquisition above 1–10 ms per point, sets the range on this time limit. To elaborate, consider an acquisition rate of 1 ms per pixel (which is a convenient setting for a lock-in amplifier) capturing data at 32 or more bits of precision. For a four-dimensional (4D) parameter space with 256 data points in all four dimensions, the acquisition time will be 1193 h (barring any time necessary to move and to settle the probe), yielding an information package of 17 GB. In the case of G-mode type measurement at 1 MHz data collection rate, 1 ms measurement in the same 4D space with 256 data points in each dimension produces a data volume of 4.2 TB. These data will not contain any additional information over the classical methods. With a reasonable experimental time of a few hours to a day, and a stable microscope platform, the advantage of shifting to experimental techniques with faster acquisition rates is immediately clear. These techniques enable much higher time resolution and are capable of probing otherwise inaccessible material properties. This notion is further supported by the fact that, often, information distributed over parameter space can be collected and compressed using multivariate statistics methods, resulting in much more manageable data file sizes at the final storage step.

Another important topic, briefly mentioned earlier, is the data acquisition sequence and the experimental order of operation, which are set by both the materials response (such as the onset of irreversible phenomena) and by the noise structure of the system. For simple imaging, the slow and fast scan axes come with pronouncedly different noise structures, which are readily identifiable and usable by the operator for fine-tuning imaging conditions. While the (x, y, V) spectroscopic imaging is instrumentally more complex and can suffer from irreversible processes, it brings about high veracity spectroscopic data at well-defined locations, such as local current voltage curves^{77–80} or ferroelectric hysteresis.⁸¹ The measurement veracity here is limited by sample drift during single pixel acquisition time, or the dominant error is in x, y for quickly varying voltage signals. Conversely, for the (V, x, y) imaging case, the error source in the spatial drift ultimately corrupts the spectroscopic voltage sweep, since every pixel in the image has to be acquired before the probe is biased at the next voltage step. This latter case requires careful alignment of the spatial maps,⁸² typically using topographic data for registration, and in cases where the spatial densities of the probed phenomena are higher than the equivalent pixel size⁸³ or in the presence of strong nonuniform drift, this registration is practically impossible.

The wide assortment of spectroscopic imaging modes in SPM explores a variety of material properties through bias-induced (ferroelectric polarization switching,⁸⁴ electrochemical reactions), thermal (glass transition, melting),^{85,86} and mechanical transformations. In addition, dynamic cantilever response (in single frequency, DART, BE, or G-mode) or tip–surface current (conductive AFM) provides detailed, time-resolved information (polarization, domain size, ionic motion, second phase formation, melting) induced by the stimulus. The advantage of faster sampling is in capturing phase transformations and electrochemical reactions at the single-defect level.^{87–89}

Table 3. Development of Multidimensional SPM

technique	dimensionality	target data set ^a	target data size ^b	refs
band excitation piezoresponse force microscopy (BE-PFM)	3D, space and ω	(256 × 256) × 64	32 MB	90, 38, 31, 85, 74, 86, 91
switching spectroscopy PFM (SS-PFM)	3D, space and voltage	(64 × 64) × 128	4 MB	92–97, 87, 98–101
time relaxation PFM (TR-PFM)	3D, space and time	(64 × 64) × 128	4 MB	96, 97, 102
AC sweeps	4D, space, ω , voltage	(64 × 64) × 64 × 256	512 MB	103, 104
BE polarization switching (BEPS)	4D, space, ω , voltage	(64 × 64) × 64 × 128	256 MB	96, 105, 106, 99, 107–109
BE thermal	4D, space, ω , temperature	(64 × 64) × 64 × 256	512 MB	85, 86, 110, 111
time relaxation BE (TR-BE)	4D, space, ω , time	(64 × 64) × 64 × 64	64 MB	106, 112, ^{106,113}
FORC BEPS	5D, space, ω , voltage, voltage	(64 × 64) × 64 × 64 × 16	2 GB	113, 114, 106, 115–117
time relaxation on sweep, BE	5D, space, ω , voltage, time	(64 × 64) × 64 × 64 × 64	16 GB	118, 119
FORC time BE	6D, space, ω , voltage, voltage, time	(64 × 64) × 64 × 64 × 16 × 64	128 GB	(lower resolution realized) 120
FORC IV BEPS	5D, space, ω , voltage, cycle	(64 × 64) × 64 × 64 × 16	4 GB	117, 121
FORC IV and FORC IV-Z	4D, space, voltage, cycle	(64 × 64) × 64 × 20	200 MB	44
time-resolved Kelvin probe force microscopy (KPFM)	3D, space, time	(60 × 20) × 1 × 10 ⁶	8 MB	122, 78, 123
open loop (OL) BE KPFM	4D, space, ω , voltage	(256 × 256) × 32 × 16	256 MB	124, 125
general-mode PFM (G-PFM)	3D, space and voltage	(256 × 256) × 1.6 × 10 ⁴	4 GB	44
G-mode voltage spectroscopy (G-VS)	ND, space, voltage ^d	(256 × 256) × 1.6 × 10 ⁶	400 GB	in development

^aDimensionality is given as (space × space) × frequency × (parameters). Note that the signal can be multimodal (e.g., collect phase and amplitude of response or three-vector component of the signal). Highest number of measured variables to date is eight (phase and amplitude in the on/off state for vertical and lateral signals). The collection of multimodal data multiplies file size by N/2. ^bNot realized yet due to data acquisition and processing limitations, but is the ultimate goal for data acquisition and analysis developments. ^cCurrent data acquisition times are limited by the eigenfrequency of the cantilever in the contact mode. However, expected introductions of fast data acquisition electronics and small cantilevers are expected to push these by a factor of ~10 in next 2–4 years. ^dAdditional output/spectroscopy channel, e.g., photothermal and electrical excitation.

Probing local transformation requires sweeping *stimuli* (tip bias, temperature, set-point, etc.) while measuring the response. Probing intrinsic hysteresis and history dependence at the first-order transitions compels first-order reversal curve (FORC) studies, effectively increasing the dimensionality and size of the data (e.g., probing Preisach densities). These simple physical arguments illustrate that a complete examination of local transformations with SPM techniques requires a 6D (space (x, y) × frequency × (stimulus × stimulus) × time) detection scheme, as compared to 1D molecular unfolding spectroscopy. The associated data sizes for these types of measurements are presented in Table 3 and in Figure 4. Note that despite the large variety in these modes, the data flow rate is quantized with a single 3D data unit of the BE method, and only the structure (or the order) of the scanned parameter space is different. The volume of the data scales with the total acquisition time. In comparison, the transition to full information capture modes (spectroscopic G-modes) will result in a further 100- to 1000-fold increase in data flow, and hence volume, for comparable scanning times.

Data Analysis Requirements in Multidimensional Scanning Probe Microscopies. The acquisition of large hyperspectral data sets brings on new challenges in data storage, dimensionality reduction, visualization, and interpretation. For instance, G-mode data (captured continuously, as the tip is rastered, without subdivision into pixels) can be converted to a pixel-by-pixel analysis or a full ensemble analysis. In the pixel-by-pixel case, the full data set is separated into sections or chunks, corresponding to a spatial location where the level of granularity in the pixel can be tuned *a posteriori* (i.e., a single spatial pixel can be split in two along the fast scan direction). This treatment is highly advantageous, since for areas with low instrumental noise and high material response, the image resolution can be arbitrarily redefined. These data sections can be transformed into the frequency domain *via* a Fourier

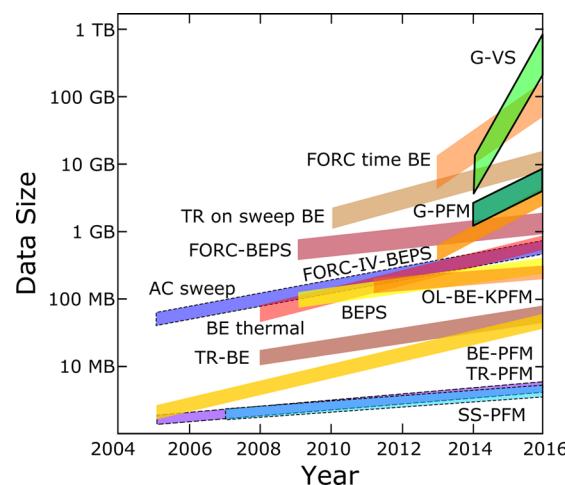


Figure 4. Evolution of information volume in multidimensional scanning probe microscopies. Abbreviations are provided in Table 3. Band excitation piezoresponse force microscopy (BE-PFM), switching spectroscopy PFM (SS-PFM), time relaxation PFM (TR-PFM), alternating current (AC) sweeps, band excitation polarization switching (BEPS), band excitation thermal, time relaxation BE (TR-BE), FORC BEPS, time relaxation (TR) on sweep, BE, FORC time BE/FORC IV BEPS, FORC IV and FORC IV-Z, time-resolved Kelvin probe force microscopy (KPFM), open loop (OL) BE KPFM, general-mode PFM (G-PFM), G-mode voltage spectroscopy (G-VS).

transform and subsequently processed with a multitude of other signal-processing routines to evaluate the material signal quality and noise signature. Alternatively, multivariate statistical analysis methods, such as PCA,^{73,126} independent component analysis (ICA),⁸⁹ and Bayesian inference methods (to name a few), may be applied to the full data set to uncover the statistically significant components and trends.¹²⁷ These two data treatment approaches, (a) functional fits on a data seg-

versus (b) multivariate analysis on the full data set, form the two primary paradigms for the multidimensional SPM techniques and thus set the stage for data analysis requirements.

The acquisition of large hyperspectral data sets brings on new challenges in data storage, dimensionality reduction, visualization, and interpretation.

These new, data-intense SPM modes come at the cost of rather stringent requirements for data-processing capabilities. Figure 5 depicts the time required to process data acquired in BE and G-mode techniques. In BE mode, the initial analysis step involves multiple SHOs, or other functional fits for each acquired pixel, which are fully parallelizable on a modern high-performance computing (HPC) infrastructure. The G-mode data processing utilizes multivariate statistical analysis, which works on the full data set and requires additional dedicated algorithms for parallelization, even in the basic imaging mode. Naturally, complex high-dimensional data sets obfuscate the analysis and data visualization process due to their sheer size, and the computational power necessary to extract, to transform, to correlate, and to present the data to the researcher exceeds generic compute platforms. Consumer-level computing devices are feasible for processing BE data smaller than 1 GB in a timely fashion. Dedicated multicore workstation systems can process large BE and G-mode data sets, however, processing times can extend beyond 24 h for data sets exceeding 10 GB. Visualization of the processed data is even more complex at this entry performance level, due to severely limited visual processing resources available at these data volumes, even for high-end workstations. High-performance computing clusters can easily provide a 100- to 500-fold improvements in processing time and can potentially enable real-time processing with optimization. The real bottleneck in processing SPM data is in the availability of random access memory (RAM) and, to a slightly lesser extent, the number of computational nodes. Therefore, medium-to-small size computational clusters provide significant advantages over workstation-based analysis.

We proceed further with a brief discussion of the analyses used to digest SPM spectral data. In the vast majority of cases to date, this analysis has been carried out using a physics-based or a phenomenological model fit of the data at each response point, as shown in Figure 6. Examples include the fitting of force–distance curves to the contact mechanics model,^{21,128} long-distance interactions to electro- and magneto-static models,^{129–131} or local current–voltage curves to traditional semiconductor physics-based transport mechanisms.¹³² The phenomenological fits are often used when the underlying physical phenomena are uncertain, and the relative noise levels in the parameter range preclude high-veracity model selection. Alternatively, this approach is also useful when the process is poorly understood, but the response has a characteristic shape, as is the case for hysteresis loop measurements in piezoresponse force microscopy.^{77,81,93} Often, however, classical physics is, in essence, a phenomenological fit, since the link between macro- and microscopic transport mechanisms is established after the data have been fitted.

A special case of the fitted-data-processing analysis is a conversion between integral and differential data representations, or in the way the data are represented before the analysis, e.g., real-space or fast Fourier transform (FFT) supported axis. Additionally, this sort of data differentiation can be found between relaxation data and relaxation time distributions or between FORCs and equivalent Preisach densities.¹¹⁴ Often, these coordinate conversions are prone to noise amplification, typical for ill-defined problems, calling for other regularization methods. For multimodal spectroscopies, this type of analysis can be performed iteratively at each analysis level, as exemplified by dynamic PFM and electrochemical strain microscopy measurements.^{120,133,134} Here, responses are measured as a function of voltage and time–yield–time-delay hysteresis loops using proper relaxation function fits (see Figure 6). Subsequent 3D data sets are fitted using phenomenological models to reduce the dimensionality further and to produce 2D images of polarization dynamics and their time dispersion. Note that, despite the complexity of the analysis procedure, 5D data sets can be reduced to 2D images with readily identifiable physical meaning, such as separation of reaction and diffusion

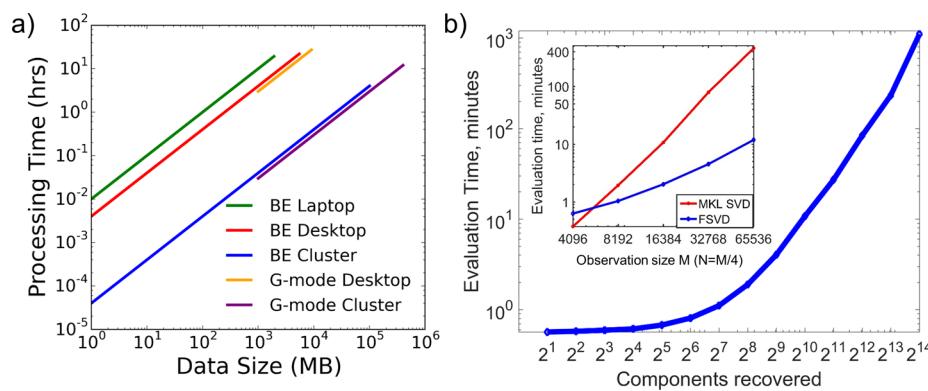


Figure 5. Processing time limitations: (a) variation in computational time for band excitation, G-mode data sets for a range of data sizes; and (b) analysis of a G-mode data file performed *via* reduced singular value decomposition (RSVD): evaluation time *versus* number of recovered components. For a small number of components, RSVD analysis takes less time than the regular singular value decomposition (SVD), but for full recovery, SVD runs faster. The inset shows evaluation time *versus* data size for full SVD and math kernel library (MKL-RSVD). Problem sizes are defined by M = number of observations, while dimensionality of each set is set to $N = M/4$, 1/4 of the number of observations. For each size, full SVD recovers all components, while RSVD only recovers 1024 first components. Thus, the underlying assumption here is that a relatively small number of components can successfully define data. Measurements were made on a cluster, using timings from a single core of Intel Xeon E5-2670 2.6 GHz processor.

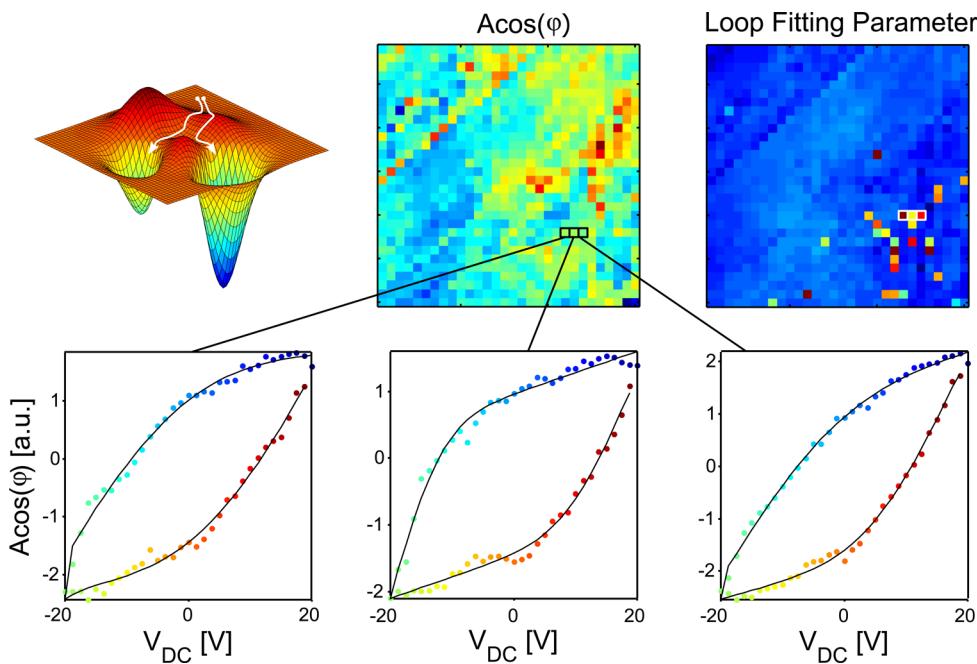


Figure 6. (a) Error surfaces for least-squares fit often contain multiple minima. Gradient descent from even close initial configurations (similar responses) can give rise to significantly different fit parameters. As an example, shown is a PFM image of polycrystalline lead zirconate titanate (PZT) and several hysteresis loops from adjacent locations. Despite close similarity of the locally measured loops, the nonlinear fit converges to different minima, resulting in discontinuous features on parameter maps. Modifying the cost function for fit to include the cost of spatial smoothness of the parameters can alleviate this problem.

in electrochemical systems, or components of relaxation in a ferro-relaxor arising from field-induced phase transitions and ordinary polarization switching.

The general drawback of physics-based interpretation methods is that they are applicable only if the physics of the system is known, or postulated *a priori*; alternatively, derived parameters are just parametrization of local responses. Second, nonlinear functional fits on a per-pixel basis can lead to a different basis of attraction—the behavior that leads to sharp transitions in images, which are often nonphysical. Finally, for multidimensional spectroscopies, each simplification step yields a number of parameter maps in each dimension, giving rise to multimodal data sets with undetermined physical sense.

An alternative approach for analysis of the multidimensional spectroscopic data is exploratory data analysis with the aid of visual analytics. Compared to the complex and often abstract data structures that emerge in the context of the network or semantic analysis in classical computer science problems, imaging data are usually defined on rectangular and uniformly sampled parameter spaces. Such a seemingly small differentiation—the ability to visualize the data on a native support, as opposed to an abstract representation—significantly reduces complications associated with data interpretation. This real-space-based data coordinate system enables the use of a broad spectrum of dimensionality reduction and visualization tools, rooted in multivariate statistics including supervised and unsupervised methods of varying complexity. These approaches can be applied to an entire multidimensional data set or to specified subdimensions, to explore the variability of material response and its relationship to surface morphology and excitation signals. Similarly, clustering and segmentation techniques allow the separation of responses in characteristic groups that can be further cross-correlated to specific signal features captured by other information channels. In some cases, these

types of data treatment can be used for “recognition” of the surfaces, or other predictive functions, where a region of known functionality can be used to train a neural network to look for “fingerprints,” in order to recognize regions with unknown functionality.^{44,86}

While multivariate statistics and visualization tools are excellent approaches for visual exploration of complex data sets, which is often optimal for a specific information-theory criterion, they fall short of providing real physical interpretations. In fact, this lack of “physical sense” has been the strongest argument *against* multivariate tools, hindering their acceptance by individual scientific communities. The fact remains, however, that basic multivariate methods are designed to be nondiscriminatory, as pure statistical procedures, as opposed to fitting to a physics-based model, which makes sense only if the model applies to the specific case. At the lowest level, these statistical methods can be used for filtering and uniformly denoising data along parameter space, avoiding cumbersome single-pixel analyses. More complex methods, like Bayesian unmixing,^{77,127,135,136} allow the introduction of constraints that can be matched to the physics of the problem, making the method more specialized—akin to parametric fits. However, these constraints are typically much less specialized than physics-based models, including boundary conditions, or forcing the components to be positive and sum to one. Notably, these more complex, higher end, multivariate methods require HPCs for timely processing of the 5D and 6D data sets due to memory and computation requirements.

Challenges in Big-Data Scanning Probe Microscopy.

To date, various bottlenecks limit the data generation in SPM experimental methods. These can be broadly classified as instrumental or analytic in nature, as illustrated in Figure 7.

Currently, SPM instrumentation software is capable of transferring a limited number of data samples ($\sim 10^6$) to and from the data-acquisition hardware. This limitation is being addressed

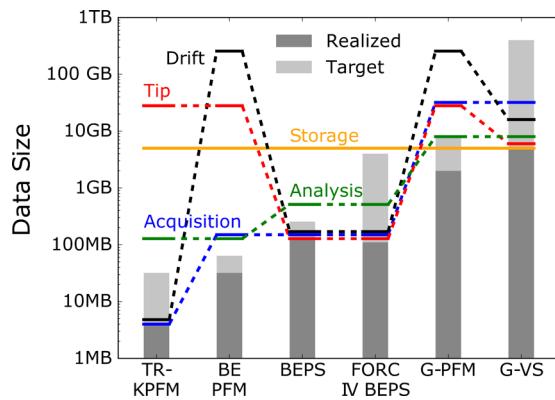


Figure 7. Limitations in realizing the target band excitation and G-mode data sizes can be divided into five categories: long-term microscope drift, degradation of the AFM tip or its coating, rate of data acquisition, data storage capacity, and time needed to perform data analysis.

by equipment manufacturers slowly, as their software and hardware designs are often driven by extant rather than latent markets. Specifically, in high-dimensional spectroscopic imaging in SPM, this slow transfer rate forces trade-offs between the number of actual data points and repeat cycles to improve the signal-to-noise ratio. This limitation precludes larger data sizes in time-resolved measurements, such as Kelvin probe force microscopy (Tr-KPFM) and BEPS. Although the scanner position drift can be neglected in measurements that span just a few minutes (such as single-frequency and contact imaging, BE PFM, or G-PFM), the drift is substantial in slower measurements (such as Tr-KPFM,¹²² BEPS, and FORC-IV BEPS). The slow transfer rate produces a secondary undesired effect where probe tip drift precludes acquisition of larger data sets than currently possible. In addition, probe wear, or general probe instability, continues to be a concern, even after years of platform development. Many techniques, such as BEPS, FORC IV BEPS, and G-VS, apply large biases between the tip and the substrate, which can result in electrochemical reactions at the tip, eroding the conductive metal layer and damaging the surface. Naturally, mechanical abrasion due to friction when scanning also results in probe damage, further deteriorating spatial resolution.

Limitations extrinsic to instrument control and other physical factors are related to the existing capabilities in the transfer, storage, and analysis of multidimensional data sets. Until recently, data generated by most spectral techniques occupied a relatively small amount of space and could be easily stored and shared via common media transfer tools, like external drives, or cloud-based storage. With advances in multidimensional BE and G-mode techniques as well as other fast acquisition detection methods in electron and ion microscopy, it is possible to fill entire storage drives with just a few data sets collected in <1 h. Access to rapid transfer and large-volume cloud-based storage is necessary to enable use and distribution of such large data sets. The analysis time scales approximately linearly with the data size, and current techniques generate data that require much more time to process and to interpret than to collect. Hence, the adoption of distributed computational and data-storage platforms offers immediate potential for addressing the data challenges in the SPM community.

DEEP DATA

Scientific progress is measured by the discovery of knowledge. This process is ultimately a transition, or translation, of correlations

between measured and external variables into a causative understanding of the underlying mechanisms driving the physical phenomena. Remarkably, some authors argue that big data defined as purely correlative studies can substitute for the classical hypothesis-driven science.¹³⁷ While supported by the success of many recent purely correlative studies, this approach encounters severe difficulties in physical sciences, which strive to generalize knowledge through abstract pictures of underlying mechanisms. We believe that data-based science does not obviate the need for hypothesis-driven science. Instead, we suggest that the purely correlative relations at which “big data” excels can invigorate the classical approach of the physical sciences, especially in the face of the increasing complexity of the systems that these disciplines aspire to understand.¹³⁸

Deep-data techniques applied to multimodal imaging promise to extend our empirical knowledge of materials by merging statistical analysis tools with physics-derived constraints.

We expect that every scientific branch will generate its own set of deep-data approaches. Here, we attempt to summarize some of the main classes of deep-data problems that appear in the context of SPM imaging, with the ultimate goal of extracting material-specific information from the imaging and spectroscopic data. This goal is two-fold: to amass unbiased data from an instrument used by a single investigator or group and to compare the obtained results directly with the overall body of prior knowledge established by the wider scientific community. In this context, we enumerate deep-data problems with progressively increasing information dimensionality, starting with single pixel spectral responses, followed by spectral imaging, and, finally, tackling the inevitable future fusion of different spectral imaging responses of materials, through multimodal imaging.

Single Pixel and Local Spectral Responses. Commonly, single-pixel and local spectral response analyses are performed using direct fits to physical models. For example, the paradigmatic SHO model is applicable to noncontact SPM modes operating near single cantilever resonances. Transitioning to contact-mode imaging and multiresonance imaging requires more complex Euler–Bernoulli-type models of the cantilever, that lack simple closed-form analytical solutions. Furthermore, the actual number of control parameters is rather large; for example, the relative position of the tip on the cantilever strongly affects cantilever dynamics,¹³⁹ whereas laser beam position on the lever affects the measured signal.¹⁴⁰ Correspondingly, quantitative measurements require quantification and calibration schemes that allow acquisition of (partial) information on the cantilever dynamics and their subsequent use in the deconvolution of force–distance-type interactions during imaging. We refer to this approach of explicit instrument calibration, followed by closed form analysis to extract the relevant probe parameters, as the classical approach. However, one can formulate and solve this problem in a more general context by employing a deep-data-based methodology.

For instance, the need for analytical closed form solutions can often be obviated by the use of neural networks trained on the numerical solutions,^{108,141} offering advantages in speed and universality. While the training of the neural network can be

time-consuming, and analysis of the data requires careful processing (normalization, etc.), once these steps are complete, the result is a way to accelerate analysis of large data sets. By contrast, the classical approach, with its use of numerical fitting algorithms (e.g., *via* gradient descent), demands the evaluation of the response function (and its derivatives) at each step, which is computationally prohibitive. Furthermore, the classical approach does not guarantee global convergence. On the other hand, training the neural network also requires computational efforts to provide training examples of the response over a sufficiently sampled parameter space. Once the network is trained, analysis of each pixel response becomes a set of matrix operations and simple functions. Finally, a trained neural network can be recycled, further increasing its value.

A particular strength of the deep-data approach is that it enables the creation of a community-wide repository of multimodel response libraries. The utility of this concept can be easily illustrated in the example of current–voltage measurements common to many SPM modes. In most cases, classical semiconductor physics models are used to analyze these data. However, the small scale of the tip–surface junction can affect the transport mechanisms, necessitating numerical solutions for different sets of materials parameters and postulated junction models. The measured responses can be affected by the presence of stray resistances (e.g., I – R drop in material bulk). The development of numerical models, however, requires efforts beyond individual experimentalists; hence, the enormous value of community-wide libraries of calculated responses is immediately apparent. Generally, the most crucial aspect of having access to multimodel response libraries is to enable unbiased modeling. In the classical approach, it is quite common to have large uncertainties in the initial models that may direct the results of the response functions into pre-existing hypotheses. However, in the case of a deep-data approach, this uncertainty, as to which model better describes the problem, is naturally addressed with Bayesian inference methods applied to a library of multimodel responses, or to put it simply, deep-data methods allow unbiased selections of the best models that most precisely describe the data.

A particular strength of the big-data approach is that it enables the creation and explorations of a community-wide repository of multimodel response libraries.

Imaging and Spectral Imaging. While the problems considered in this Nano Focus can be posed for virtually any physical measurements, spatially resolved measurements can provide multiple realizations of the same measurement under slightly different local conditions, giving rise to large statistic sets. In addition, if the grid size of the SPM image is comparable to, or smaller than, the size of characteristic inhomogeneities in the material, the spatial (or spectral) images will contain these spatial features. For 2D images, the presence of the latter can be ascertained by visual inspection, correlation analysis, or more powerful methods described in the following section. For spectroscopic data sets, equivalent information can be obtained by examination of dimensionality-reduced data sets (e.g., *via* PCA). Correspondingly, we refer to the cases when images contain only statistical spatially uncorrelated informa-

tion as sparse spatial grids, and images that contain spatial features as dense grids.

The first broad class of problems arising in the context of any imaging technique is deconvolution, namely, the separation of the material properties from the effects of the probe. The theory and the numerical methods for deconvolution are well developed for linear imaging mechanisms. In these methods, the acquired image can be represented as a convolution of an ideal image, containing materials parameters, and (position independent) probe function as shown in eq 1:

$$I(\mathbf{x}) = \int I_0(\mathbf{x} - \mathbf{y})F(\mathbf{y})d\mathbf{y} + N(\mathbf{x}) \quad (1)$$

where $I(\mathbf{x})$ the measured image and \mathbf{x} is a set of spatial coordinates, $I_0(\mathbf{x} - \mathbf{y})$ is the *ideal image* representing material properties, $F(\mathbf{y})$ is the probe effect function with its own set of noise coordinates \mathbf{y} , and $N(\mathbf{x})$ is the noise function.

Note that even when eq 1 holds, and the probe function is (to some extent) known, the inversion is a nontrivial problem since the noise present in the image is invariably magnified in the process, requiring regularization methods. When the resolution function is unknown but some initial information is available, more complex reconstruction methods, such as Landweber regularization,¹⁴² can be used. However, the linearity of the imaging mechanism has to be established. For KPFM, in the small response approximation, the response is linear. For PFM, it is almost linear within the limitation of the decoupling approximation.¹⁴³ However, for other important techniques such as topography imaging, the imaging mechanism is nonlinear, imposing the need for blind deconvolution.¹⁴⁴

An interesting and almost unexplored avenue is the deconvolution within spectral imaging. If the imaging is linear, then the ideal image, the measured image, and the response function can all be functions of frequency, with an additional function smoothness constraint of the parameter-dependent resolution function. For physics-based deconvolution, the spatial smoothness of extracted parameters can provide additional constraints.

Finally, there are intriguing aspects in cases when the signal has multiple contributions with spatially varying loadings. When the responses of individual components are known, they can be separated using approaches such as multiple linear regression. However, the advantage of large statistical sets is that components can be separated using statistical unmixing methods,^{127,145} even when individual forms are unknown. The simplest examples of such approaches are unsupervised methods such as PCA, which allow decorrelation and dimensionality reduction in the optimal (from an information theory viewpoint) sense. However, PCA components lack well-defined physical meaning. The transition to deep-data analysis, imparting the pre-existing knowledge of the physics of the system, can be achieved by using constrained decompositions based on Bayesian inference that can naturally incorporate a physical model into the data analysis *via* prior distributions. For example, Bayesian linear unmixing (BLU) postulates the linear additivity and non-negativity of endmembers and, as such, is ideally suited for demixing of spectroscopic mixtures.¹²⁷ Applied to 3D electron energy loss spectroscopy or Raman hyperspectral data sets (spectral intensity *versus* x , y , and energy, or Raman shift), BLU yields a user-defined number of pure spectra (endmembers) and corresponding abundance maps that reflect the spatial distribution of endmembers intensity, scaled from 0 to 100%. The pure spectra can be compared to the known signatures of specific elements, phases, or mechanisms, and thus spatial

localization of these effects can be achieved. As mentioned above, BLU is a blind source separation method that requires no prior knowledge of the pure spectra. The results of unmixing voltage spectroscopic data sets (current *versus* x , y , and voltage) are “pure” current–voltage curves that are then evaluated by fitting them to the equations of specific transport mechanisms, as opposed to comparing them to a reference (as is the case for optical and electron spectroscopies). Thus, the “pure” conductivity behavior can be extracted by the means of statistics even if it is unknown *a priori*.^{77,135} Furthermore, additional constraints can be imposed on the analysis methods to accommodate real physical limitations of a system and to decrease the solution uncertainty related to noise corruption.

Multimodal Imaging. The prior two subsections have demonstrated how the increase in the dimensionality of SPM data can be accommodated within a framework rooted in deep data and fast acquisition. This increase in data dimensionality, facilitated by development in SPM instrumentation, will ultimately culminate in a new paradigm based on multimodal imaging. Signs of things to come are already evident in a number of combinatorial instruments available today, such as SPM-microRaman, SPM-mass spectrometry, and SPM-secondary ion mass spectrometry to name a few. Extracting materials properties from this multitude of information channels is at the core of deep data. In this section, we outline the specific deep-data problems that will emerge in the context of multimodal imaging.

Data sets that are acquired with two experimental modalities can differ in either signal provenance (*e.g.*, high-resolution topographic and low-resolution spectral images of the same region) or probe–sample interaction (*e.g.*, SPM and mass spectroscopy images will differ in spatial support (x , y), sampling density Δ , and spatiotemporal resolution Y). We denote these multimodal channels by $A(x, y, \Delta, Y)$ and $B(\bar{x}, \bar{y}, \bar{\Delta}, \bar{Y})$, where A and B are signal measured by different experimental modalities.

The joint analysis of A and B starts with response matching, feature extraction, and identification. Multiple strategies can be utilized to achieve this task, but they must all share the common characteristic of being scale-invariant, *i.e.*, insensitive to Δ and Y . For instance, to transform signals A and B so that they share the same spatial support (x , y), one can: (1) extract features using detectors and descriptors that are scale-invariant such as SIFT¹⁴⁶ and ORB¹⁴⁷ and (2) match the feature descriptors by an appropriate metric (*e.g.*, Hamming distance for ORB) to find the optimal and general geometric transformation $T: B(\bar{x}, \bar{y}) \rightarrow B(x, y)$. Note, however, that the memory storage requirements, due to the high dimensionality of the feature descriptors, (*e.g.*, the SIFT descriptor is a 128-dimensional vector) and the prohibitive computational time of brute force mutual matching of descriptors $\sim O(N^2)$, where N is the number of descriptors per image, require dedicated computational resources.

Once the optimal transformation T is found, A and B share a common spatial support and can be fused or pansharpened.^{148–150} Although the transformation T was found through a scale-invariant approach, mapping $B(\bar{x}, \bar{y}, \bar{\Delta})$ onto the new spatial support (x , y) is a nontrivial problem, especially in the case of general geometric operations such as projective transformations.¹⁵¹ In cases where the signals have comparable sampling densities (*i.e.*, pixel densities in A and B are similar, or $\Delta \approx \bar{\Delta}$), interpolation can be used to “warp” image B to the new spatial grid. In the other limiting case where $\Delta \neq \bar{\Delta}$, B can be considered to be sparse in the spatial domain (x , y). Henceforth, one can reconstruct $B(x, y)$ by hyperspectral

pansharpening¹⁴⁸ or compressed sensing.^{152,153} The latter, in particular, gives a robust reconstruction of images that have missing or corrupted data. Since its introduction,¹⁵² compressed sensing research has developed robust and fast algorithms¹⁵⁴ for image analysis. Compressed sensing uses mixed optimization to determine the sparse set of components that represents the image being analyzed. Compressed sensing approaches search for the best image that minimizes a regularization norm, such as the L_1 norm of total variation, and minimizes the least-squares error as compared to measurements. L_1 optimization has had a long reach across various disciplines of mathematics and statistics, with applications ranging from image processing to multivariate statistics to uncertainty quantification, and it has been shown to have a strong potential in a broad range of microscopy data analysis applications.^{155,156}

An alternative method to unify the process of feature identification and to learn the metric that is used for matching is through recent developments in deep convolutional neural networks such as MatchNet.¹⁵⁷ The advantage of these deep-learning techniques is that they learn the “natural” similarity measure in the feature space that is used to compare learned features from A and B , instead of an imposed metric on the feature descriptors. Note, however, that the development of deep-learning techniques for feature learning is very much in its infancy and requires substantial testing to study their behavior under scale-invariance and robustness to the noise that is inherent in all scientific images. Furthermore, a key aspect of deep-learning techniques is training on large databases of images, preferably scientific, and therefore further reinforces the before-mentioned need for establishing scientific community-based repositories with open accessibility.

Regardless of the feature identification and matching techniques used, features that are extracted and matched from A and B represent different properties of the same material and can be mined directly for correlations (*e.g.*, mutual information) using unsupervised learning techniques from big data. However, far more information can be extracted by employing supervised machine learning techniques in deep data.

Features extracted from A and B represent multiple views of the same material, *i.e.*, they provide information on different properties of the material. Let us denote these views with $\{\chi_i^{A(B)}\}_{i=1,\dots}$ where the superscript $A(B)$ indicates the experimental signal from which the material property χ was obtained. Multiview machine-learning methods,¹⁵⁸ can be used to learn relations between $\chi^A(x', y')$ and $\chi^B(x, y)$ to predict $\chi^A(x', y')$ given a measurement $\chi^B(x', y')$. However, overfitting based on available information in the spatial support (x , y) is often a problem and requires regularization. Techniques in multiview learning overcome this problem through what is commonly known as manifold coregularization, where training on the available information is performed in a joint fashion with the goal of maximizing the mutual agreement of the two views χ^A and χ^B . In conjunction with deep learning or spectral analysis of data features,¹⁵⁹ which can function as coregularization, multiview learning techniques can transform multimodal imaging data into material knowledge that neither signal A nor B could have provided on their own.

Deep-data techniques applied to multimodal imaging promise to extend our empirical knowledge of materials, especially for systems with complex energy landscapes, such as disordered systems, (*e.g.*, spin glass) or strongly coupled systems with competing degrees of freedom, such as high-temperature superconductors. Additionally, the deep-data approaches outlined here point the

way to a new paradigm in experimental investigations that we refer to as “smart data”.

SMART DATA

Deep data translates experimental signals into interpretable information. The methodology that enables us to quantify the incompleteness of our understanding and guides the subsequent scientific quest belongs to the realm of “smart data.” Smart data should enable capabilities to design new experiments based on information accessed through deep data. Smart data is also where information acquired through deep data makes contact with the classical paradigms of the scientific method, such as theory and computation, and enables a seamless interface between experiment and theory through computation/simulation and, finally, physical understanding. Paradigmatic examples of smart-data systems include the GoogleCar, the use of neural nets in medical decision-making, such as with cancerous growth in mammalian tissue, or more ambitious projects such as the general diagnosis of IBM’s Watson.^{160–164} Below, we discuss some of the smart-data applications and developments specific to microscopy as well as pathways for conjoining microscopy data with computation and theory and, ultimately, smooth incorporation into domain knowledge.

Microscope Level. At the operational level of the microscope, the classical feedback systems operating on the real-time signals are the enabling component of SPM imaging. Notably, techniques, such as interleave mode and adaptive feedback, developed for compensation of the piezoelectric transducer drift often incorporate machine-learning algorithms and are not discussed here. Rather, we discuss possible applications of smart-data methods in various SPM modes.

The broad set of smart-data tools can be envisioned in spectroscopic imaging as the real-time feedback of the parametric sweeps, which is in essence the ability to monitor real-time material responses. These approaches are required for high-resolution thermal and electrochemical imaging when the size of the region is undergoing irreversible changes, which limits the spatial resolution and ultimately reduces the quality of the overall data set. Equivalently, it would be hugely beneficial to tune the sampled spatial point density dynamically, such as with a higher density of points around resonance in BE measurements and around regions with strong signal change in voltage spectroscopies to maximize the quality of the derived experimental information.

Just as exciting is the development of techniques that enable adaptable scanning trajectories,¹⁶⁵ which culminate in spatial grids where the information density is dependent on the object of interest. These examples include scanning relatively small areas, such as grain boundaries or nanotubes, with higher grid density around the area of maximal response or notable behavior. These developments require creative engineering to estimate relevant information density during the scan, which also relies on previously imaged areas, and for the simultaneous adaptation of cantilever trajectory and scanning rates. Advances in these techniques will include adaptable trajectories in interleave modes to obviate topographic cross talk in the electrostatic measurements, such as iso-capacitance lines calculated from known or estimated topography.

Data Analytics and Image Analysis. In the last several years, there has been a progressive effort to make data open for analysis. Indeed, in many communities, including astronomy, high-energy physics, crystallography, genetics, and thermodynamics, universal data repositories have long become the norm.

The driving forces between the adoption of these solutions are well explored in the literature.¹⁶⁶ However, practically, the feasibility of such large-scale repositories depends on the availability of computational power as well as a handle on the data flow and storage solutions. While SPM data present handling challenges due to volume, heterogeneity, and extensive metadata, recent progress of databases such as Open Microscopy Environment and their software package Open Microscopy Environment Remote Objects¹⁶⁷ (OMERO) suggests the viability of this pathway.

We note that these community-wide data repositories allow much more than transparency in data analysis and knowledge derivation. They also provide a seamless way to incorporate negative results into the knowledge process, make data open for reuse, and enable community-wide interpretation of the entire results library, as opposed to the selected subsets published in the literature. The success of projects such as ImageNet, PolyMath, and GalaxyZoo are a testament to the validity of this approach.¹⁶⁶

Information to Knowledge. The body of scientific literature is growing at an unprecedented rate and is becoming increasingly accessible. While many benefit from lowered barriers to information access, the challenges of finding the state-of-the-art scientific knowledge effectively remain, and in some aspects have intensified. For example, the retraction rate of scientific publications is also unprecedently high,¹⁶⁸ including high-profile withdrawals.¹⁶⁹ Furthermore, the trustworthiness of scientific literature becomes questionable in light of studies that reveal how common citation distortions are and how arduous they are to capture without a significant effort.¹⁷⁰

Scientific knowledge is a complex adaptive system of facts, hypotheses, beliefs, opinions, and a wide variety of other types of information describing what we know and how we interpret the world. It is adaptive in the sense that the existing scientific knowledge is subject to re-examination in light of new discoveries, and clearer understanding is forged over time by generations of scientists and scholars.^{39,40} The domain expertise of a scientist is characterized not only by their knowledge of various findings in scientific literature but also by their understanding of the uncertainties, constraints, and exceptions that are associated with these findings.

To capture the latest advances, to recognize novel and high-impact scientific discoveries, and to make significant contributions of their own, scientists need to cast wider nets and to augment their expertise and efforts with computational and explanatory tools. Existing tools such as CiteSpace¹⁷¹ can help users understand the structure and dynamics of a scientific field based on patterns and trends computationally derived from bibliographic records, network properties of citations, and entities such as natural language phrases, along with geographic and bibliographic relations. Most of the tools that help analyze the scientific literature focus on bibliographies, as opposed to the information at a finer level of granularity, namely, the full text.¹⁷² However, the increasingly available peer-reviewed and open-access scientific publications provide the necessary basis for synergistic combination of these approaches with the ultimate goal of automating the discovery process. Such trends and analyses do not capture real discoveries at the outset and are thus training indicators.

Several challenges must be conquered in order to assist scientists, to capture state-of-the-art information, and to make solid scientific discoveries. For example, it is often a challenging task, even for a domain expert, to tease out the most critical

The domain expertise of a scientist is characterized not only by their knowledge of various findings in scientific literature but also by their understanding of the uncertainties, constraints, and exceptions that are associated with these findings.

information from an ambiguously or conditionally stated scientific claim in unstructured text. The level of complexity involved in capturing the state-of-the-art scientific knowledge suggests that it is essential to develop tools that also tackle the challenges of distilling the semantics from a large body of scientific literature. Extracting and representing this information is among the most pressing needs in scientific discovery and dissemination.¹⁷³ Several recent studies have demonstrated a great potential for a repository of semantic predications, in literature-based discovery,¹⁷⁴ the public health domain,¹⁷⁵ and in identifying drug-disease relations for drug repurposing.¹⁷⁶ Their experiences have revealed a few areas where semantically focused approaches could be enriched to provide powerful functionalities, especially in areas where predefined domain ontologies may not be readily available. An integrative effort in machine learning and scholarly knowledge exploration will significantly lower the threshold for conducting literature-based discovery and improving the understanding of science and innovation in general.

OUTLOOK

The emergence of powerful computational resources in the 20th century led to the development of the scientific computational disciplines. This was the birth of a new branch of the scientific method, where, alongside theory and experiment, the scientific workflow has acquired a new component in the form of numerical simulations. The ever accelerating power and availability of computation resources have prompted some to append the scientific method with data-intensive science.¹⁷⁷ An alternative historical interpretation is that the proliferation of data-enabled theory gave birth to numerical simulation. The latter has facilitated the interpretation of experimental investigations of ever more complex systems and phenomena. The interactions between theory and experiment through simulation are, at present, limited by the enormous number of information channels provided by modern instruments and the equally vast number of parameters that define a numerical simulation. A far more complete and scientifically rigorous interaction between modern day theory and microscopy is only now becoming a reality. An illustration of this progress is shown in Figure 8. The recent emergence of supercomputing and the accompanying big data makes it feasible, for the first time, for experimental investigations to build a framework that can be integrated seamlessly with theory, *i.e.*, numerical simulations. We believe that this framework will likely transform how we perform and interpret SPM. The change that is sweeping through all empirical investigations comes with a host of challenges, some of which we have discussed, with others yet to be unearthed. However, the opportunity to invigorate scientific disciplines in the information age and finally to unify and to consolidate its various approaches in order to propel our understanding of materials into yet undiscovered realms is exciting to say the least.

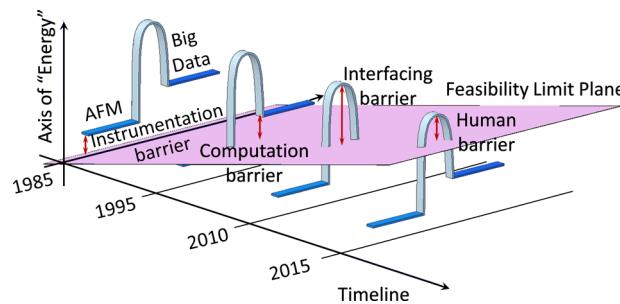


Figure 8. Timeline of SPM and big-data development: An instrumentation barrier hampered realization of SPM in the 1980s. The computational power in the 1990s was not sufficient to implement big-data techniques. By 2010, both SPM and big-data techniques were realized, but interfacing between the two was not developed. Today, the only impediment for full integration of SPM and big-data analytics is the human education factor: The SPM community needs to adopt statistical and machine-learning techniques and to incorporate them into its research.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: sergei2@ornl.gov.

*E-mail: belianinova@ornl.gov.

*E-mail: sjesse@ornl.gov.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The research on developing advanced SPM methods and analysis was conducted at the Center for Nanophase Materials Sciences, which is a DOE Office of Science User Facility (S.V.K., E.S., A.B., S.S., R.K.V., N.L., and S.J.). Research by E.J.L. on developing the HPC data analysis and stewardship infrastructure is sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy. N.L. was supported by the Eugene P. Wigner Fellowship at Oak Ridge National Laboratory. R.K.A. acknowledges the Compute and Data Environment for Science (CADES) for continuous support. The mathematical aspects were sponsored by the applied mathematics program at the DOE and the computational aspects made use of the Oak Ridge Leadership Computing Facility, a DOE Office of Science User Facility at ORNL supported under contract no. DE-AC05-00OR22725. This manuscript has been authored by UT-Battelle, LLC, with the U.S. Department of Energy.

REFERENCES

- (1) Binnig, G.; Quate, C. F.; Gerber, C. Atomic Force Microscope. *Phys. Rev. Lett.* **1986**, 56 (9), 930–933.
- (2) Gerber, C.; Lang, H. P. How the doors to the nanoworld were opened. *Nat. Nanotechnol.* **2006**, 1 (1), 3–5.
- (3) Feynman, R. P. There's Plenty of Room at the Bottom. *Caltech Engineering and Science* **1960**, 23 (5), 22–36.
- (4) Mody, C. C. M. *Instrumental Community*; The MIT Press: Cambridge, MA, 2011.
- (5) Nazin, G.; Zhang, Y.; Zhang, L.; Sutter, E.; Sutter, P. Visualization of charge transport through Landau levels in graphene. *Nat. Phys.* **2010**, 6 (11), 870–874.
- (6) Topinka, M. A.; LeRoy, B. J.; Westervelt, R. M.; Shaw, S. E. J.; Fleischmann, R.; Heller, E. J.; Maranowski, K. D.; Gossard, A. C. Coherent branched flow in a two-dimensional electron gas. *Nature* **2001**, 410 (6825), 183–186.

- (7) Topinka, M. A.; LeRoy, B. J.; Shaw, S. E. J.; Heller, E. J.; Westervelt, R. M.; Maranowski, K. D.; Cossard, A. C. Imaging coherent electron flow from a quantum point contact. *Science* **2000**, *289* (5488), 2323–2326.
- (8) Bonnell, D. A.; Kalinin, S. V.; Khoklin, A. L.; Gruverman, A. Piezoresponse Force Microscopy: A Window into Electromechanical Behavior at the Nanoscale. *MRS Bull.* **2009**, *34* (9), 648–657.
- (9) Alvarez, T.; Kalinin, S. V.; Bonnell, D. A. Magnetic-field measurements of current-carrying devices by force-sensitive magnetic-force microscopy with potential correction. *Appl. Phys. Lett.* **2001**, *78* (7), 1005–1007.
- (10) Albrecht, T. R.; Grutter, P.; Horne, D.; Rugar, D. Frequency-Modulation Detection Using High-Q Cantilevers for Enhanced Force Microscope Sensitivity. *J. Appl. Phys.* **1991**, *69* (2), 668–673.
- (11) Martin, Y.; Wickramasinghe, H. K. Magnetic Imaging by Force Microscopy with 1000-A Resolution. *Appl. Phys. Lett.* **1987**, *50* (20), 1455–1457.
- (12) Sugimoto, Y.; Pou, P.; Abe, M.; Jelinek, P.; Perez, R.; Morita, S.; Custance, O. Chemical identification of individual surface atoms by atomic force microscopy. *Nature* **2007**, *446* (7131), 64–67.
- (13) Binnig, G.; Rohrer, H.; Gerber, C.; Weibel, E. 7 × 7 Reconstruction on Si(111) Resolved in Real Space. *Phys. Rev. Lett.* **1983**, *50* (2), 120–123.
- (14) Tanimoto, M.; Vatet, O. Kelvin probe force microscopy for characterization of semiconductor devices and processes. *J. Vac. Sci. Technol., B: Microelectron. Process. Phenom.* **1996**, *14* (2), 1547–1551.
- (15) Kalinin, S. V.; Bonnell, D. A. Scanning impedance microscopy of electroactive interfaces. *Appl. Phys. Lett.* **2001**, *78* (9), 1306–1308.
- (16) Shao, R.; Kalinin, S. V.; Bonnell, D. A. Nanoimpedance microscopy and spectroscopy. *MRS Online Proc. Libr.* **2003**, *738*, 163–168.
- (17) Rief, M.; Oesterhelt, F.; Heymann, B.; Gaub, H. E. Single molecule force spectroscopy on polysaccharides by atomic force microscopy. *Science* **1997**, *275* (5304), 1295–1297.
- (18) Clausen-Schaumann, H.; Seitz, M.; Krautbauer, R.; Gaub, H. E. Force spectroscopy with single bio-molecules. *Curr. Opin. Chem. Biol.* **2000**, *4* (5), 524–530.
- (19) Hinterdorfer, P.; Dufrene, Y. F. Detection and localization of single molecular recognition events using atomic force microscopy. *Nat. Methods* **2006**, *3* (5), 347–355.
- (20) Stroh, C.; Wang, H.; Bash, R.; Ashcroft, B.; Nelson, J.; Gruber, H.; Lohr, D.; Lindsay, S. M.; Hinterdorfer, P. Single-molecule recognition imaging-microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (34), 12503–12507.
- (21) Meyer, E.; Hug, H. J.; Bennewitz, R. *Scanning Probe Microscopy: The Lab on a Tip*; Springer-Verlag: Berlin, 2003.
- (22) Bonnell, D. *Scanning Probe Microscopy and Spectroscopy: Theory, Techniques, and Applications*. Wiley-VCH: New York, 2000.
- (23) Kalinin, S. V.; Gruverman, A. *Scanning Probe Microscopy: Electrical and Electromechanical Phenomena at the Nanoscale*; Springer Science+Business Media, LLC: New York, 2007; Vols. 1 and 2.
- (24) Kalinin, S. V.; Gruverman, A. *Scanning Probe Microscopy of Functional Materials: Nanoscale Imaging and Spectroscopy*; Springer Science+Business Media, LLC: New York, 2011.
- (25) Bonnell, D. A.; Kalinin, S. V. *Scanning Probe Microscopy for Energy Research: 7 (World Scientific Series in Nanoscience and Nanotechnology)*. World Scientific Printing Co., Pte., Ltd.: Singapore, 2013.
- (26) Albrecht, T. R.; Akamine, S.; Carver, T. E.; Quate, C. F. Microfabrication of Cantilever Styli for the Atomic Force Microscope. *J. Vac. Sci. Technol., A* **1990**, *8* (4), 3386–3396.
- (27) Garcia, R.; Herruzo, E. T. The emergence of multifrequency force microscopy. *Nat. Nanotechnol.* **2012**, *7* (4), 217–226.
- (28) Garcia, R.; Perez, R. Dynamic atomic force microscopy methods. *Surf. Sci. Rep.* **2002**, *47* (6–8), 197–301.
- (29) Arruda, T. M.; Kumar, A.; Jesse, S.; Veith, G. M.; Tselev, A.; Baddorf, A. P.; Balke, N.; Kalinin, S. V. Toward Quantitative Electrochemical Measurements on the Nanoscale by Scanning Probe Microscopy: Environmental and Current Spreading Effects. *ACS Nano* **2013**, *7* (9), 8175–8182.
- (30) Garcia-Martin, A.; Garcia, R. Formation of nanoscale liquid menisci in electric fields. *Appl. Phys. Lett.* **2006**, *88* (12), 123115.
- (31) 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* **2007**, *18*, 435503.
- (32) Jesse, S.; Kalinin, S. V. Band excitation in scanning probe microscopy: sines of change. *J. Phys. D: Appl. Phys.* **2011**, *44* (46), 464006.
- (33) Jesse, S.; Vasudevan, R.; Collins, L.; Strelcov, E.; Okatan, M. B.; Belianinov, A.; Baddorf, A. P.; Proksch, R.; Kalinin, S. V. Band excitation in scanning probe microscopy: recognition and functional imaging. *Annu. Rev. Phys. Chem.* **2014**, *65*, 519–536.
- (34) Somnath, S.; Belianinov, A.; Kalinin, S. V.; Jesse, S. Full information acquisition in piezoresponse force microscopy. *Appl. Phys. Lett.* **2015**, *107* (26), 263102.
- (35) Collins, L.; Belianinov, A.; Somnath, S.; Rodrigues, B.; Balke, N.; Kalinin, S. V.; Jesse, S. Multifrequency spectrum analysis using fully digital G Mode-Kelvin probe force microscopy. *Nanotechnology* **2016**, *27* (10), 105706.
- (36) Belianinov, A.; Kalinin, S. V.; Jesse, S. Complete information acquisition in dynamic force microscopy. *Nat. Commun.* **2015**, *6*, 6550.
- (37) Proksch, R.; Kalinin, S. V. Energy dissipation measurements in frequency-modulated scanning probe microscopy. *Nanotechnology* **2010**, *21* (45), 455705.
- (38) 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* **2007**, *18* (43), 435503.
- (39) Chen, C. *The Fitness of Information: Quantitative Assessments of Critical Evidence* (Wiley Series in Probability and Statistics); Wiley: Hoboken, NJ, 2014.
- (40) Chen, C. *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*. Springer: London, 2013.
- (41) Chen, C. M.; Hicks, D. Tracing knowledge diffusion. *Scientometrics* **2004**, *59* (2), 199–211.
- (42) Kuhn, T.; Perc, M.; Helbing, D. Inheritance Patterns in Citation Networks Reveal Scientific Memes. *Phys. Rev. X* **2014**, *4* (4), 041036.
- (43) Kalinin, S. V.; Sumpter, B. G.; Archibald, R. K. Big-deep-smart data in imaging for guiding materials design. *Nat. Mater.* **2015**, *14*, 973.
- (44) Belianinov, A.; Vasudevan, R.; Strelcov, E.; Steed, C.; Yang, S. M.; Tselev, A.; Jesse, S.; Biegalski, M.; Shipman, G.; Symons, C. Big data and deep data in scanning and electron microscopies: deriving functionality from multidimensional data sets. *Adv. Struct. Chem. Imaging* **2015**, *1* (1), 1–25.
- (45) Belianinov, A.; He, Q.; Kravchenko, M.; Jesse, S.; Borisevich, A.; Kalinin, S. V. Identification of phases, symmetries and defects through local crystallography. *Nat. Commun.* **2015**, *6*, 7801.
- (46) Klimentov, A.; Buncic, P.; De, K.; Jha, S.; Maeno, T.; Mount, R.; Nilsson, P.; Oleynik, D.; Panitkin, S.; Petrosyan, A.; Porter, R. J.; Read, K. F.; Vaniachine, A.; Wells, J. C.; Wenaus, T.; Iop, Next Generation Workload Management System For Big Data on Heterogeneous Distributed Computing. Proceedings from the 16th International Workshop on Advanced Computing and Analysis Techniques in Physics Research, Prague, Czech Republic, September 1–5, 2014; Iop Publishing Ltd: Bristol, U.K., 2015; Vol. 608.
- (47) Liu, Y.; LeBlanc, P.; Grutter, P.; Durig, U. Magnetic dissipation imaging. *J. Appl. Phys.* **1997**, *81* (8), 5024–5024.
- (48) Grutter, P.; Liu, Y.; LeBlanc, P.; Durig, U. Magnetic dissipation force microscopy. *Appl. Phys. Lett.* **1997**, *71* (2), 279–281.
- (49) Kos, A. B.; Hurley, D. C. Nanomechanical mapping with resonance tracking scanned probe microscope. *Meas. Sci. Technol.* **2008**, *19* (1), 015504.
- (50) Rodriguez, B. J.; Callahan, C.; Kalinin, S. V.; Proksch, R. Dual-frequency resonance-tracking atomic force microscopy. *Nanotechnology* **2007**, *18* (47), 475504.

- (51) Hutter, C.; Platz, D.; Tholén, E. A.; Hansson, T. H.; Haviland, D. B. Reconstructing Nonlinearities with Intermodulation Spectroscopy. *Phys. Rev. Lett.* **2010**, *104* (5), 050801.
- (52) Platz, D.; Tholén, E. A.; Pesen, D.; Haviland, D. B. Intermodulation atomic force microscopy. *Appl. Phys. Lett.* **2008**, *92* (15), 153106.
- (53) Martin, Y.; Wickramasinghe, H. K. Magnetic imaging by “force microscopy” with 1000 Å resolution. *Appl. Phys. Lett.* **1987**, *50* (20), 1455–1457.
- (54) Maivald, P.; Butt, H. J.; Gould, S. A. C.; Prater, C. B.; Drake, B.; Gurley, J. A.; Elings, V. B.; Hansma, P. K. Using force modulation to image surface elasticities with the atomic force microscope. *Nanotechnology* **1991**, *2* (2), 103.
- (55) Florin, E. L.; Radmacher, M.; Fleck, B.; Gaub, H. E. Atomic force microscope with magnetic force modulation. *Rev. Sci. Instrum.* **1994**, *65* (3), 639–643.
- (56) Stark, R. W.; Drobek, T.; Heckl, W. M. Tapping-mode atomic force microscopy and phase-imaging in higher eigenmodes. *Appl. Phys. Lett.* **1999**, *74* (22), 3296–3298.
- (57) Albrecht, T. R.; Grütter, P.; Horne, D.; Rugar, D. Frequency modulation detection using high-Q cantilevers for enhanced force microscope sensitivity. *J. Appl. Phys.* **1991**, *69* (2), 668–673.
- (58) Proksch, R.; Dahlberg, E. D. A detection technique for scanning force microscopy. *Rev. Sci. Instrum.* **1993**, *64* (4), 912–916.
- (59) Stark, M.; Guckenberger, R.; Stemmer, A.; Stark, R. W. Estimating the transfer function of the cantilever in atomic force microscopy: A system identification approach. *J. Appl. Phys.* **2005**, *98* (11), 114904.
- (60) Rodriguez, B. J.; Callahan, C.; Kalinin, S. V.; Proksch, R. Dual-frequency resonance-tracking atomic force microscopy. *Nanotechnology* **2007**, *18* (47), 475504.
- (61) Rodriguez, T. R.; Garcia, R. Compositional mapping of surfaces in atomic force microscopy by excitation of the second normal mode of the microcantilever. *Appl. Phys. Lett.* **2004**, *84* (3), 449–451.
- (62) Proksch, R. Multifrequency, repulsive-mode amplitude-modulated atomic force microscopy. *Appl. Phys. Lett.* **2006**, *89* (11), 113121.
- (63) Gannepalli, A.; Sebastian, A.; Cleveland, J.; Salapaka, M. Thermally driven non-contact atomic force microscopy. *Appl. Phys. Lett.* **2005**, *87* (11), 111901.
- (64) Stark, M.; Stark, R. W.; Heckl, W. M.; Guckenberger, R. Spectroscopy of the anharmonic cantilever oscillations in tapping-mode atomic-force microscopy. *Appl. Phys. Lett.* **2000**, *77* (20), 3293–3295.
- (65) Kos, A. B.; Hurley, D. C. Nanomechanical mapping with resonance tracking scanned probe microscope. *Meas. Sci. Technol.* **2008**, *19* (1), 015504.
- (66) Meyer, E.; Heinzelmann, H.; Grütter, P.; Jung, T.; Hidber, H. R.; Rudin, H.; Güntherodt, H. J. Atomic force microscopy for the study of tribology and adhesion. *Thin Solid Films* **1989**, *181* (1–2), 527–544.
- (67) Krotl, H.-U.; Stifter, T.; Marti, O. Concurrent measurement of adhesive and elastic surface properties with a new modulation technique for scanning force microscopy. *Rev. Sci. Instrum.* **2000**, *71* (7), 2765–2771.
- (68) Hosaka, S.; Kikukawa, A.; Honda, Y.; Koyanagi, H.; Tanaka, S. Simultaneous Observation of 3-Dimensional Magnetic Stray Field and Surface Structure Using New Force Microscope. *Jpn. J. Appl. Phys.* **1992**, *31* (7A), L904.
- (69) 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: Condens. Matter Mater. Phys.* **2001**, *64* (4), 045401.
- (70) Rabe, U.; Arnold, W. Acoustic Microscopy by Atomic-Force Microscopy. *Appl. Phys. Lett.* **1994**, *64* (12), 1493–1495.
- (71) Rabe, U.; Kopycinska, M.; Hirsekorn, S.; Saldana, J. M.; Schneider, G. A.; Arnold, W. High-resolution characterization of piezoelectric ceramics by ultrasonic scanning force microscopy techniques. *J. Phys. D: Appl. Phys.* **2002**, *35* (20), 2621–2635.
- (72) Collins, L.; Belianinov, A.; Proksch, R.; Zuo, T.; Zhang, Y.; Liaw, P. K.; Kalinin, S. V.; Jesse, S. G-mode magnetic force microscopy: Separating magnetic and electrostatic interactions using big data analytics. *Appl. Phys. Lett.* **2016**, *108* (19), 193103.
- (73) Bonnet, N. Multivariate statistical methods for the analysis of microscope image series: applications in materials science. *J. Microsc.* **1998**, *190*, 2–18.
- (74) Jesse, S.; Kalinin, S. V. Principal component and spatial correlation analysis of spectroscopic-imaging data in scanning probe microscopy. *Nanotechnology* **2009**, *20* (8), 085714.
- (75) Lingerfelt, E. J.; Messer, O. B.; Desai, S. S.; Holt, C. A.; Lentz, E. J. Near real-time data analysis of core-collapse supernova simulations with bellerophon. *Procedia Computer Science* **2014**, *29*, 1504–1514.
- (76) Lingerfelt, E.; Belianinov, A.; Endeve, E.; Ovchinnikov, O.; Somnath, S.; Borreguero, J.; Grodowitz, N.; Park, B.; Archibald, R.; Symons, C. BEAM: A Computational Workflow System for Managing and Modeling Material Characterization Data in HPC Environments. *Procedia Computer Science* **2016**, *80*, 2276–2280.
- (77) Strelcov, E.; Belianinov, A.; Hsieh, Y.-H.; Chu, Y.-H.; Kalinin, S. V. Constraining data mining with physical models: voltage-and oxygen pressure-dependent transport in multiferroic nanostructures. *Nano Lett.* **2015**, *15* (10), 6650–6657.
- (78) Strelcov, E.; Ievlev, A. V.; Jesse, S.; Kravchenko, I. I.; Shur, V. Y.; Kalinin, S. V. Direct Probing of Charge Injection and Polarization-Controlled Ionic Mobility on Ferroelectric LiNbO_3 Surfaces. *Adv. Mater.* **2014**, *26* (6), 958–963.
- (79) Belianinov, A.; Ganesh, P.; Lin, W.; Sales, B. C.; Sefat, A. S.; Jesse, S.; Pan, M.; Kalinin, S. V. Research Update: Spatially resolved mapping of electronic structure on atomic level by multivariate statistical analysis. *APL Mater.* **2014**, *2* (12), 120701.
- (80) Yang, N.; Belianinov, A.; Strelcov, E.; Tebano, A.; Foglietti, V.; Di Castro, D.; Schlueter, C.; Lee, T.-L.; Baddorf, A. P.; Balke, N. Effect of doping on surface reactivity and conduction mechanism in samarium-doped ceria thin films. *ACS Nano* **2014**, *8* (12), 12494–12501.
- (81) Balke, N.; Maksymovych, P.; Jesse, S.; Herklotz, A.; Tselev, A.; Eom, C.-B.; Kravchenko, I. I.; Yu, P.; Kalinin, S. V. Differentiating Ferroelectric and Nonferroelectric Electromechanical Effects with Scanning Probe Microscopy. *ACS Nano* **2015**, *9* (6), 6484–6492.
- (82) Polomoff, N. A.; Premnath, R. N.; Bosse, J. L.; Huey, B. D. Ferroelectric domain switching dynamics with combined 20 nm and 10 ns resolution. *J. Mater. Sci.* **2009**, *44* (19), 5189–5196.
- (83) Ovchinnikova, O. S.; Tai, T.; Bocharova, V.; Okatan, M. B.; Belianinov, A.; Kertesz, V.; Jesse, S.; Van Berk, G. J. Co-registered Topographical, Band Excitation Nanomechanical, and Mass Spectral Imaging Using a Combined Atomic Force Microscopy/Mass Spectrometry Platform. *ACS Nano* **2015**, *9* (4), 4260–4269.
- (84) Rodriguez, B. J.; Jesse, S.; Baddorf, A. P.; Zhao, T.; Chu, Y. H.; Ramesh, R.; Eliseev, E. A.; Morozovska, A. N.; Kalinin, S. V. Spatially resolved mapping of ferroelectric switching behavior in self-assembled multiferroic nanostructures: strain, size, and interface effects. *Nanotechnology* **2007**, *18* (40), 405701.
- (85) 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.* **2008**, *93* (7), 073104.
- (86) Nikiforov, M. P.; Jesse, S.; Morozovska, A. N.; Eliseev, E. A.; Germinario, L. T.; Kalinin, S. V. Probing the temperature dependence of the mechanical properties of polymers at the nanoscale with band excitation thermal scanning probe microscopy. *Nanotechnology* **2009**, *20* (39), 395709.
- (87) Rodriguez, B. J.; Choudhury, S.; Chu, Y. H.; Bhattacharyya, A.; Jesse, S.; Seal, K.; Baddorf, A. P.; Ramesh, R.; Chen, L. Q.; Kalinin, S. V. Unraveling Deterministic Mesoscopic Polarization Switching Mechanisms: Spatially Resolved Studies of a Tilt Grain Boundary in Bismuth Ferrite. *Adv. Funct. Mater.* **2009**, *19* (13), 2053–2063.
- (88) Rodriguez, B. J.; Chu, Y. H.; Ramesh, R.; Kalinin, S. V. Ferroelectric domain wall pinning at a bicrystal grain boundary in bismuth ferrite. *Appl. Phys. Lett.* **2008**, *93* (14), 142901.

- (89) Marincel, D. M.; Zhang, H. R.; Britson, J.; Belianinov, A.; Jesse, S.; Kalinin, S. V.; Chen, L. Q.; Rainforth, W. M.; Reaney, I. M.; Randall, C. A.; Trolier-McKinstry, S. Domain pinning near a single-grain boundary in tetragonal and rhombohedral lead zirconate titanate films. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2015**, *91* (13), 134113.
- (90) Jesse, S.; Mirman, B.; Kalinin, S. V. Resonance enhancement in piezoresponse force microscopy: Mapping electromechanical activity, contact stiffness, and Q factor. *Appl. Phys. Lett.* **2006**, *89* (2), 022906.
- (91) Nikiforov, M. P.; Thompson, G. L.; Reukov, V. V.; Jesse, S.; Guo, S.; Rodriguez, B. J.; Seal, K.; Vertegel, A. A.; Kalinin, S. V. Double-Layer Mediated Electromechanical Response of Amyloid Fibrils in Liquid Environment. *ACS Nano* **2010**, *4* (2), 689–698.
- (92) Jesse, S.; Baddorf, A. P.; Kalinin, S. V. Switching spectroscopy piezoresponse force microscopy of ferroelectric materials. *Appl. Phys. Lett.* **2006**, *88* (6), 062908.
- (93) Rodriguez, B. J.; Jesse, S.; Baddorf, A. P.; Kalinin, S. V. High resolution electromechanical imaging of ferroelectric materials in a liquid environment by piezoresponse force microscopy. *Phys. Rev. Lett.* **2006**, *96* (23), 237602.
- (94) Rodriguez, B. J.; Jesse, S.; Alexe, M.; Kalinin, S. V. Spatially Resolved Mapping of Polarization Switching Behavior in Nanoscale Ferroelectrics. *Adv. Mater.* **2008**, *20*, 109.
- (95) Kalinin, S. V.; Jesse, S.; Rodriguez, B. J.; Chu, Y. H.; Ramesh, R.; Eliseev, E. A.; Morozovska, A. N. Probing the role of single defects on the thermodynamics of electric-field induced phase transitions. *Phys. Rev. Lett.* **2008**, *100* (15), 155703.
- (96) Tan, Z.; Roytburd, A. L.; Levin, I.; Seal, K.; Rodriguez, B. J.; Jesse, S.; Kalinin, S. V.; Baddorf, A. P. Piezoelectric response of nanoscale PbTiO₃ in composite PbTiO₃–CoFe₂O₄ epitaxial films. *Appl. Phys. Lett.* **2008**, *93*, 074101.
- (97) Bintachitt, P.; Trolier-McKinstry, S.; Seal, K.; Jesse, S.; Kalinin, S. V. Switching spectroscopy piezoresponse force microscopy of polycrystalline capacitor structures. *Appl. Phys. Lett.* **2009**, *94* (4), 042906.
- (98) Seal, K.; Jesse, S.; Nikiforov, M. P.; Kalinin, S. V.; Fujii, I.; Bintachitt, P.; Trolier-McKinstry, S. Spatially Resolved Spectroscopic Mapping of Polarization Reversal in Polycrystalline Ferroelectric Films: Crossing the Resolution Barrier. *Phys. Rev. Lett.* **2009**, *103* (5), 057601.
- (99) Wicks, S.; Seal, K.; Jesse, S.; Anbusathaiyah, V.; Leach, S.; Garcia, R. E.; Kalinin, S. V.; Nagarajan, V. Collective dynamics in nanostructured polycrystalline ferroelectric thin films using local time-resolved measurements and switching spectroscopy. *Acta Mater.* **2010**, *58* (1), 67–75.
- (100) Rodriguez, B. J.; Jesse, S.; Bokov, A. A.; Ye, Z. G.; Kalinin, S. V. Mapping bias-induced phase stability and random fields in relaxor ferroelectrics. *Appl. Phys. Lett.* **2009**, *95* (9), 092904.
- (101) Rodriguez, B. J.; Jesse, S.; Morozovska, A. N.; Svechnikov, S. V.; Kiselev, D. A.; Kholkin, A. L.; Bokov, A. A.; Ye, Z. G.; Kalinin, S. V. Real space mapping of polarization dynamics and hysteresis loop formation in relaxor-ferroelectric PbMg_{1/3}Nb_{2/3}O₃–PbTiO₃ solid solutions. *J. Appl. Phys.* **2010**, *108* (4), 042006.
- (102) Kalinin, S. V.; Rodriguez, B. J.; Budai, J. D.; Jesse, S.; Morozovska, A. N.; Bokov, A. A.; Ye, Z. G. Direct evidence of mesoscopic dynamic heterogeneities at the surfaces of ergodic ferroelectric relaxors. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *81* (6), 064107.
- (103) Bintachitt, P.; Jesse, S.; Damjanovic, D.; Han, Y.; Reaney, I. M.; Trolier-McKinstry, S.; Kalinin, S. V. Collective dynamics underpins Rayleigh behavior in disordered polycrystalline ferroelectrics. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (16), 7219–7224.
- (104) Griggio, F.; Jesse, S.; Kumar, A.; Marincel, D. M.; Tinberg, D. S.; Kalinin, S. V.; Trolier-McKinstry, S. Mapping piezoelectric nonlinearity in the Rayleigh regime using band excitation piezoresponse force microscopy. *Appl. Phys. Lett.* **2011**, *98* (21), 212901.
- (105) MakSYMovich, P.; Balke, N.; Jesse, S.; Huijben, M.; Ramesh, R.; Baddorf, A. P.; Kalinin, S. V. Defect-induced asymmetry of local hysteresis loops on BiFeO₃ surfaces. *J. Mater. Sci.* **2009**, *44* (19), 5095–5101.
- (106) Balke, N.; Jesse, S.; Morozovska, A. N.; Eliseev, E.; Chung, D. W.; Kim, Y.; Adamczyk, R. E.; Garcia, R. E.; Dudney, N.; Kalinin, S. V. Nanometer-scale electrochemical intercalation and diffusion mapping of Li-ion battery materials. *Nat. Nanotechnol.* **2010**, *5*, 749–754.
- (107) McLachlan, M. A.; McComb, D. W.; Ryan, M. P.; Morozovska, A. N.; Eliseev, E. A.; Payzant, E. A.; Jesse, S.; Seal, K.; Baddorf, A. P.; Kalinin, S. V. Probing Local and Global Ferroelectric Phase Stability and Polarization Switching in Ordered Macroporous PZT. *Adv. Funct. Mater.* **2011**, *21* (5), 941–947.
- (108) Kumar, A.; Ovchinnikov, O.; Guo, S.; Griggio, F.; Jesse, S.; Trolier-McKinstry, S.; Kalinin, S. V. Spatially resolved mapping of disorder type and distribution in random systems using artificial neural network recognition. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2011**, *84* (2), 024203.
- (109) Kim, Y.; Kumar, A.; Tselev, A.; Kravchenko, I. I.; Han, H.; Vrejoiu, I.; Lee, W.; Hesse, D.; Alexe, M.; Kalinin, S. V. Non-Linear Phenomena in Multiferroic Nanocapacitors: Joule Heating and Electromechanical Effects. *ACS Nano* **2011**, *5* (11), 9104.
- (110) Nikiforov, M. P.; Gam, S.; Jesse, S.; Composto, R. J.; Kalinin, S. V. Morphology Mapping of Phase-Separated Polymer Films Using Nanothermal Analysis. *Macromolecules* **2010**, *43* (16), 6724–6730.
- (111) Nikiforov, M. P.; Hohlbauch, S.; King, W. P.; Voitsovsky, K.; Contera, S. A.; Jesse, S.; Kalinin, S. V.; Proksch, R. Temperature-dependent phase transitions in zeptoliter volumes of a complex biological membrane. *Nanotechnology* **2011**, *22* (5), 055709.
- (112) Guo, S.; Jesse, S.; Kalnaus, S.; Balke, N.; Daniel, C.; Kalinin, S. V. Direct Mapping of Ion Diffusion Times on LiCoO₂ Surfaces with Nanometer Resolution. *J. Electrochem. Soc.* **2011**, *158* (8), A982–A990.
- (113) Ovchinnikov, O.; Jesse, S.; Guo, S.; Seal, K.; Bintachitt, P.; Fujii, I.; Trolier-McKinstry, S.; Kalinin, S. V. Local measurements of Preisach density in polycrystalline ferroelectric capacitors using piezoresponse force spectroscopy. *Appl. Phys. Lett.* **2010**, *96* (11), 112906.
- (114) Guo, S.; Ovchinnikov, O. S.; Curtis, M. E.; Johnson, M. B.; Jesse, S.; Kalinin, S. V. Spatially resolved probing of Preisach density in polycrystalline ferroelectric thin films. *J. Appl. Phys.* **2010**, *108* (8), 084103.
- (115) Vasudevan, R.; Liu, Y.; Li, J.; Liang, W. I.; Kumar, A.; Jesse, S.; Chen, Y. C.; Chu, Y.-H.; Nagarajan, V.; Kalinin, S. V. Nanoscale-control of phase-variants in strain-engineered BiFeO₃. *Nano Lett.* **2011**, *11*, 3346.
- (116) Arruda, T. M.; Kumar, A.; Kalinin, S. V.; Jesse, S. Mapping Irreversible Electrochemical Processes on the Nanoscale: Ionic Phenomena in Li Ion Conductive Glass Ceramics. *Nano Lett.* **2011**, *11* (10), 4161–4167.
- (117) Kumar, A.; Jesse, S.; Morozovska, A. N.; Eliseev, E.; Tebano, A.; Yang, N.; Kalinin, S. V. Variable temperature electrochemical strain microscopy of Sm-doped ceria. *Nanotechnology* **2013**, *24* (14), 145401.
- (118) 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.* **2011**, *98* (20), 202903.
- (119) Kumar, A.; Ciucci, F.; Morozovska, A. N.; Kalinin, S. V.; Jesse, S. Measuring oxygen reduction/evolution reactions on the nanoscale. *Nat. Chem.* **2011**, *3* (9), 707–713.
- (120) Vasudevan, R. K.; Zhang, S.; Okatan, M. B.; Jesse, S.; Kalinin, S. V.; Bassiri-Gharb, N. Multidimensional dynamic piezoresponse measurements: Unraveling local relaxation behavior in relaxor-ferroelectrics via big data. *J. Appl. Phys.* **2015**, *118* (7), 072003.
- (121) Kim, Y.; Strelcov, E.; Hwang, I. R.; Choi, T.; Park, B. H.; Jesse, S.; Kalinin, S. V. Correlative Multimodal Probing of Ionically-Mediated Electromechanical Phenomena in Simple Oxides. *Sci. Rep.* **2013**, *3*, 2924.
- (122) Strelcov, E.; Jesse, S.; Huang, Y. L.; Teng, Y. C.; Kravchenko, I. I.; Chu, Y. H.; Kalinin, S. V. Space- and Time-Resolved Mapping of

- Ionic Dynamic and Electroresistive Phenomena in Lateral Devices. *ACS Nano* **2013**, *7* (8), 6806–6815.
- (123) Ding, J.; Strelcov, E.; Kalinin, S. V.; Bassiri-Gharb, N. Spatially Resolved Probing of Electrochemical Reactions via Energy Discovery Platforms. *Nano Lett.* **2015**, *15* (6), 3669–3676.
- (124) Guo, S.; Kalinin, S. V.; Jesse, S. Open-loop band excitation Kelvin probe force microscopy. *Nanotechnology* **2012**, *23* (12), 125704.
- (125) Collins, L.; Kilpatrick, J. I.; Weber, S. A. L.; Tselev, A.; Vlassiuk, I. V.; Ivanov, I. N.; Jesse, S.; Kalinin, S. V.; Rodriguez, B. J. Open loop Kelvin probe force microscopy with single and multi-frequency excitation. *Nanotechnology* **2013**, *24* (47), 475702.
- (126) Bonnet, N. Artificial intelligence and pattern recognition techniques in microscope image processing and analysis. In *Advances in Imaging and Electron Physics*; Hawkes, P. W., Ed. Elsevier Academic Press, Inc: San Diego, 2000; Vol. 114, pp 1–77.
- (127) Dobigeon, N.; Brun, N. Spectral mixture analysis of EELS spectrum-images. *Ultramicroscopy* **2012**, *120*, 25–34.
- (128) Cheng, S.; Bocharova, V.; Belianinov, A.; Xiong, S.; Kisliuk, A. M.; Somnath, S.; Holt, A. P.; Ovchinnikova, O. S.; Jesse, S.; Martin, H. Unraveling the Mechanism of Nanoscale Mechanical Reinforcement in Glassy Polymer Nanocomposites. *Nano Lett.* **2016**, *16* (6), 3630.
- (129) 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.* **1997**, *81* (3), 1023–1030.
- (130) Fumagalli, L.; Gramse, G.; Esteban-Ferrer, D.; Edwards, M. A.; Gomila, G. Quantifying the dielectric constant of thick insulators using electrostatic force microscopy. *Appl. Phys. Lett.* **2010**, *96* (18), 183107.
- (131) Gramse, G.; Casuso, I.; Toset, J.; Fumagalli, L.; Gomila, G. Quantitative dielectric constant measurement of thin films by DC electrostatic force microscopy. *Nanotechnology* **2009**, *20* (39), 395702.
- (132) Sze, S. M. *Physics of Semiconductor Devices*, 2nd ed.; Wiley-Interscience: New York, 1981.
- (133) Kumar, A.; Ehara, Y.; Wada, A.; Funakubo, H.; Griggio, F.; Trolier-McKinstry, S.; Jesse, S.; Kalinin, S. V. Dynamic piezoresponse force microscopy: Spatially resolved probing of polarization dynamics in time and voltage domains. *J. Appl. Phys.* **2012**, *112* (5), 052021.
- (134) Yang, N.; Strelcov, E.; Belianinov, A.; Tebano, A.; Foglietti, V.; Schlueter, C.; Lee, T.-L.; Jesse, S.; Kalinin, S. V.; Balestro, G. Effect of Water Adsorption on Conductivity in Epitaxial Sm_{0.1}Ce_{0.9}O_{2-δ} Thin Film for Micro Solid Oxide Fuel Cells Applications. *ECS Trans.* **2015**, *69* (16), 39–48.
- (135) Strelcov, E.; Belianinov, A.; Hsieh, Y. H.; Jesse, S.; Baddorf, A. P.; Chu, Y. H.; Kalinin, S. V. Deep Data Analysis of Conductive Phenomena on Complex Oxide Interfaces: Physics from Data Mining. *ACS Nano* **2014**, *8* (6), 6449–6457.
- (136) Tselev, A.; Ivanov, I. N.; Lavrik, N. V.; Belianinov, A.; Jesse, S.; Mathews, J. P.; Mitchell, G. D.; Kalinin, S. V. Mapping internal structure of coal by confocal micro-Raman spectroscopy and scanning microwave microscopy. *Fuel* **2014**, *126*, 32–37.
- (137) Viktor Mayer-Schönberger, K. C. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*; Eamon Dolan/Houghton Mifflin Harcourt: New York, 2013.
- (138) Laanait, N.; Callagon, E. B. R.; Zhang, Z.; Sturchio, N. C.; Lee, S. S.; Fenter, P. X-ray–driven reaction front dynamics at calcite–water interfaces. *Science* **2015**, *349* (6254), 1330–1334.
- (139) Labuda, A.; Proksch, R. Quantitative measurements of electromechanical response with a combined optical beam and interferometric atomic force microscope. *Appl. Phys. Lett.* **2015**, *106* (25), 253103.
- (140) Huey, B. D.; Ramanujan, C.; Bobji, M.; Blendell, J.; White, G.; Szoszkiewicz, R.; Kulik, A. The importance of distributed loading and cantilever angle in piezo-force microscopy. *J. Electroceram.* **2004**, *13* (1–3), 287–291.
- (141) Ovchinnikov, O. S.; Jesse, S.; Bintacchit, P.; Trolier-McKinstry, S.; Kalinin, S. V. Disorder Identification in Hysteresis Data: Recognition Analysis of the Random-Bond-Random-Field Ising Model. *Phys. Rev. Lett.* **2009**, *103* (15), 157203.
- (142) Landweber, L. An iteration formula for Fredholm integral equations of the first kind. *American journal of mathematics* **1951**, *73* (3), 615–624.
- (143) Morozovska, A. N.; Eliseev, E. A.; Bravina, S. L.; Kalinin, S. V. Resolution-function theory in piezoresponse force microscopy: Wall imaging, spectroscopy, and lateral resolution. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2007**, *75* (17), 174109.
- (144) Villarrubia, J. S. Algorithms for scanned probe microscope image simulation, surface reconstruction, and tip estimation. *J. Res. Natl. Inst. Stand. Technol.* **1997**, *102* (4), 425–454.
- (145) Altmann, Y.; Dobigeon, N.; McLaughlin, S.; Tourneret, J. Y. Nonlinear Spectral Unmixing of Hyperspectral Images Using Gaussian Processes. *Ieee Transactions on Signal Processing* **2013**, *61* (10), 2442–2453.
- (146) Lowe, D. G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* **2004**, *60* (2), 91–110.
- (147) Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G.; IEEE, ORB: an efficient alternative to SIFT or SURF. Proceedings from the 2011 IEEE International Conference on Computer Vision (ICCV), Barcelona, Spain, November 6–13, 2011; IEEE: New York, 2011; pp 2564–2571.
- (148) Loncan, L.; et al. Hyperspectral pansharpening: a review. 2015, arXiv:1504.04531v1arxiv.
- (149) Wei, Q.; Bioucas-Dias, J.; Dobigeon, N.; Tourneret, J. Y. Hyperspectral and Multispectral Image Fusion Based on a Sparse Representation. *Ieee Transactions on Geoscience and Remote Sensing* **2015**, *53* (7), 3658–3668.
- (150) Wei, Q.; Dobigeon, N.; Tourneret, J. Y. Bayesian Fusion of Multi-Band Images. *Ieee Journal of Selected Topics in Signal Processing* **2015**, *9* (6), 1117–1127.
- (151) Szeliski, R. *Computer Vision - Algorithms and Applications*. Springer London: London, 2011.
- (152) Donoho, D. L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52* (4), 1289–1306.
- (153) Candes, E. J.; Plan, Y. A Probabilistic and RIPless Theory of Compressed Sensing. *IEEE Trans. Inf. Theory* **2011**, *57* (11), 7235–7254.
- (154) Goldstein, T.; Osher, S. The Split Bregman Method for L1-Regularized Problems. *Siam Journal on Imaging Sciences* **2009**, *2* (2), 323–343.
- (155) Stevens, A.; Yang, H.; Carin, L.; Arslan, I.; Browning, N. D. The potential for Bayesian compressive sensing to significantly reduce electron dose in high-resolution STEM images. *Microscopy* **2014**, *63* (1), 41–51.
- (156) Stevens, A.; Kovarik, L.; Abellán, P.; Yuan, X.; Carin, L.; Browning, N. D. Applying compressive sensing to TEM video: a substantial frame rate increase on any camera. *Adv. Struct. Chem. Imaging* **2015**, *1* (1), 1.
- (157) Xufeng, H.; Leung, T.; Yangqing, J.; Sukthankar, R.; Berg, A. C. In MatchNet: Unifying feature and metric learning for patch-based matching. Proceedings from the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, June 7–12, 2015; IEEE: New York, 2015; pp 3279–3286.
- (158) Sindhwani, V.; Rosenberg, D. S. In An RKHS for multi-view learning and manifold co-regularization. Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, July 5–9, 2008; ACM: New York, 2008; pp 976–983.
- (159) Reuter, M.; Wolter, F. E.; Peinecke, N. Laplace-Beltrami spectra as 'Shape-DNA' of surfaces and solids. *Computer-Aided Design* **2006**, *38* (4), 342–366.
- (160) Guizzo, E. How google's self-driving car works. *IEEE Spectrum Online*, October 18, 2011, <http://spectrum.ieee.org/automaton/robotics/artificial-intelligence/how-google-self-driving-car-works>.
- (161) Chakraborty, J.; Rangayyan, R. M.; Banik, S.; Mukhopadhyay, S.; Desautels, J. L. Statistical measures of orientation of texture for the detection of architectural distortion in prior mammograms of interval-cancer. *Journal of Electronic Imaging* **2012**, *21* (3), 033010.
- (162) Mazurowski, M. A.; Habas, P. A.; Zurada, J. M.; Lo, J. Y.; Baker, J. A.; Tourassi, G. D. Training neural network classifiers for

- medical decision making: The effects of imbalanced datasets on classification performance. *Neural networks* **2008**, *21* (2), 427–436.
- (163) Murdoch, T. B.; Detsky, A. S. The inevitable application of big data to health care. *Jama* **2013**, *309* (13), 1351–1352.
- (164) Ferrucci, D. A.; Levas, A.; Bagchi, S.; Gondek, D.; Mueller, E. T. Watson: beyond jeopardy! *Artif. Intell.* **2013**, *199-200*, 93–105.
- (165) Ovchinnikov, O. S.; Jesse, S.; Kalinin, S. V. Adaptive probe trajectory scanning probe microscopy for multiresolution measurements of interface geometry. *Nanotechnology* **2009**, *20* (25), 255701.
- (166) Nielsen, M. *Reinventing Discovery: The New Era of Networked Science*. Princeton University Press: Princeton, NJ, 2011.
- (167) Allan, C.; Burel, J.-M.; Moore, J.; Blackburn, C.; Linkert, M.; Loynton, S.; MacDonald, D.; Moore, W. J.; Neves, C.; Patterson, A. OMERO: flexible, model-driven data management for experimental biology. *Nat. Methods* **2012**, *9* (3), 245–253.
- (168) Chen, C.; Hu, Z.; Milbank, J.; Schultz, T. A visual analytic study of retracted articles in scientific literature. *J. Am. Soc. Inf. Sci. Technol.* **2013**, *64* (2), 234–253.
- (169) Mora, C.; Frazier, A. G.; Longman, R. J.; Dacks, R. S.; Walton, M. M.; Tong, E. J.; Sanchez, J. J.; Kaiser, L. R.; Stender, Y. O.; Anderson, J. M.; Ambrosino, C. M. Mora et al. reply. *Nature* **2014**, *511* (7507), E5–E6.
- (170) Greenberg, S. A. How citation distortions create unfounded authority: Analysis of a citation network. *BMJ* **2009**, *339*, b2680.
- (171) Chen, C. M. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *J. Am. Soc. Inf. Sci. Technol.* **2006**, *57* (3), 359–377.
- (172) Reed, J. W.; Jiao, Y.; Potok, T. E.; Klump, B. A.; Elmore, M. T.; Hurson, A. R. TF-ICF: A new term weighting scheme for clustering dynamic data streams. Proceedings from the *4th International Conference on Machine Learning and Applications*, Orlando, FL, December 14–16, 2006; IEEE: New York, 2006; pp 258–263.
- (173) Kilicoglu, H.; Shin, D.; Fiszman, M.; Rosemblat, G.; Rindflesch, T. C. SemMedDB: A PubMed-scale repository of biomedical semantic predications. *Bioinformatics* **2012**, *28* (23), 3158–3160.
- (174) Cameron, D.; Bodenreider, O.; Yalamanchili, H.; Danh, T.; Vallabhaneni, S.; Thirunarayan, K.; Sheth, A. P.; Rindflesch, T. C. A graph-based recovery and decomposition of Swanson's hypothesis using semantic predications. *J. Biomed. Inf.* **2013**, *46*, 238–251.
- (175) Rosemblat, G.; Resnick, M. P.; Auston, I.; Shin, D.; Schneiderman, C.; Fiszman, M.; Rindflesch, T. C. Extending SemRep to the public health domain. *J. Am. Soc. Inf. Sci. Technol.* **2013**, *64* (10), 1963–1974.
- (176) Xu, R.; Wang, Q. Large-scale extraction of accurate drug-disease treatment pairs from biomedical literature for drug repurposing. *BMC Bioinf.* **2013**, *14*, 181–192.
- (177) *The Fourth Paradigm*; Hey, T.; Tansley, S.; Tolle, K., Eds.; Microsoft Research: Redmond, WA, 2009.