



# Deep learning for electron and scanning probe microscopy: From materials design to atomic fabrication

Sergei V. Kalinin,\* Maxim Ziatdinov, Steven R. Spurgeon, Colin Ophus, Eric A. Stach, Toma Susi, Josh Agar, and John Randall

Machine learning and artificial intelligence (ML/AI) are rapidly becoming an indispensable part of physics research, with applications ranging from theory and materials prediction to high-throughput data analysis. In parallel, the recent successes in applying ML/AI methods for autonomous systems from robotics through self-driving cars to organic and inorganic synthesis are generating enthusiasm for the potential of these techniques to enable automated and autonomous experiment in imaging. In this article, we discuss recent progress in application of machine learning methods in scanning transmission electron microscopy and scanning probe microscopy, from applications such as data compression and exploratory data analysis to physics learning to atomic fabrication.

## Introduction

Electron and scanning probe microscopy techniques have become a mainstay of research in materials science, condensed-matter physics, biology, and nanotechnology.<sup>1</sup> Electron microscopy (EM) has made atomic resolution imaging,<sup>2</sup> the mapping of plasmon and phonon excitations,<sup>3</sup> the probing of the chemical states of individual atoms,<sup>4</sup> and atomic assembly<sup>5</sup> a reality. Similarly, scanning probe microscopy has enabled mapping of individual chemical bonds, probing of quasiparticles and superconducting order parameters,<sup>6</sup> as well as the exploration of tip-induced reactions and atomic manipulation.<sup>7</sup> These instrumental capabilities have provided insight into atomic structures and functionalities of materials ranging from superconductors and ferroelectrics to macromolecules. However, the reams of data generated by modern imaging and hyperspectral imaging tools surpass the current infrastructure for storage and analysis, limiting our ability to derive actionable physics, chemistry, and materials insights.

Until recently, most imaging techniques have relied on semiquantitative analysis, where human experts interpret

two-dimensional (2D) images or individual one-dimensional (1D) spectra. However, data collected by electron and probe microscopes are often intrinsically quantitative but encoded across various modalities and dimensionalities. This characteristic requires subsequent extraction of features and correlations at various length and time scales. This includes information regarding the relative positions of atoms and bonding networks directly related to the thermodynamics and kinetics of materials synthesis. Minute symmetry-breaking distortions contain information concerning structural and polar-order parameters. Spatial maps of plasmonic and phonon interactions contain information about the dielectric function, convoluted with shape effects. Correspondingly, the second challenge for microscopy is learning the fundamental physics and chemistry of the studied materials from imaging and spectral data.

Finally, the third challenge is inherently linked to the capability of electron and scanning probe microscopes to modify materials in a controllable fashion. The last note left by Richard Feynman on his blackboard was “What I cannot make,

Sergei V. Kalinin, Department of Materials Science and Engineering, The University of Tennessee, Knoxville, USA; sergei2@utk.edu  
Maxim Ziatdinov, Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, USA; ziatdinovma@ornl.gov  
Steven R. Spurgeon, Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, USA; Department of Physics, University of Washington, Seattle, USA; steven.spurgeon@pnnl.gov  
Colin Ophus, National Center for Electron Microscopy, Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, USA; clophus@lbl.gov  
Eric A. Stach, Department of Materials Science and Engineering and Laboratory for Research on the Structure of Matter, University of Pennsylvania, Philadelphia, USA; stach@seas.upenn.edu  
Toma Susi, Faculty of Physics, University of Vienna, Vienna, Austria; tomas.susi@univie.ac.at  
Josh Agar, Department of Mechanical Engineering and Mechanics, Drexel University, Philadelphia, USA; jca92@drexel.edu  
John Randall, Zyvex Labs, Richardson, USA; jrandall@zyvexlabs.com

\*Corresponding author

doi:10.1557/s43577-022-00413-3

I cannot understand.” From this perspective, local probes’ capability to visualize and manipulate matter on the atomic level is the next frontier for nanoscience and nanotechnology. Active learning methods under the guise of machine learning offer a pathway to harness imaging data streams, derive physical insights into the chemistry and physics of materials, and assemble them at the nano- and atomic scales.

### Learning more from data

Since the early days of van Leeuwenhoek, the power of microscopy lays in its ability to visualize objects on progressively smaller length scales. For electron and scanning probe microscopies, an additional step involves the conversion of interactions between the electron or scanning probe and the object to form images. However, data interpretation has been largely driven by human insights. This includes identifying features in images or spectra and qualitative interpretation, in some cases followed by quantitative analysis, ultimately connecting results to physical models and prior knowledge. This approach is inherently limited by human perception and bias. For example, the human eye is remarkably good at identifying well-localized objects, but struggles to detect the emergence of correlated signatures in different parts of the image field or to detect small or gradual changes in periodicity. Furthermore, the eye is extremely sensitive to color scales and can regularly be deceived by the perception of contrast. Even more importantly, interpretation of the data in terms of relevant physics is highly dependent on prior knowledge.

Machine learning (ML) methods offer an opportunity to change this paradigm. While neural networks have been known since the early 1950s, the lack of training algorithms and sufficiently powerful computational tools have led to slow progress, with ebbs and flows. This situation has changed in the last decade.<sup>8</sup> The explosive development of ML tools in computer vision, medicine, and robotics has created the knowledge base and enabling infrastructure that allows its extension to the physical sciences. Next, we identify some of the applications enabled by classical supervised ML methods and new opportunities enabled by unsupervised and physics-driven ML.

### What is in the image: Supervised learning

Artificial intelligence-guided knowledge extraction from raw experimental data streams is a challenge across all materials science, chemistry, and physics disciplines.<sup>9,10</sup> In electron and scanning probe microscopy specifically, we are often concerned with determining key microstructural descriptors that encode atomic-scale properties and processes.<sup>11</sup> These descriptors are challenging to define rigorously but are essential as a key step in transforming raw data into actionable metrics for high-throughput and automated decision-making. The most common characterization modalities for scanning transmission electron microscopy (STEM), scanning electron microscopy (SEM), and scanning probe microscopy (SPM) are various forms of 2D imaging, as shown in **Figure 1**. For TEM, most

measurements compress the signals measured from a three-dimensional (3D) sample volume into a projected 2D image. SEM and SPM are primarily sensitive to material surfaces and, thus, are also well suited for 2D imaging. Fortunately, 2D image analysis is also the sub-domain of deep learning (DL) that has received the most attention, due to immense online image databases in medicine and biology, astronomy, and social media.

Most DL applications in image processing target classification, with the most common subproblem being segmentation. Semantic segmentation is the process of labeling every pixel (or voxel) into discrete classes and is the desired output of many deep learning routines.<sup>12</sup> Segmentation is beneficial for many microscopy applications in both TEM<sup>13</sup> and SPM studies.<sup>14</sup> Image segmentation examples in electron microscopy include separating atomic resolution phases,<sup>15</sup> and rapid determination of microstructural features.<sup>16</sup> Segmentation has also been applied to SPM data sets, where it has been used to detect and avoid atomic surface defects before patterning<sup>17</sup> and detect arbitrarily complex features over many length scales.<sup>18</sup> Deep learning can also produce more profound insights into microscopy experiments by directly extracting relevant properties. SPM examples include extracting mechanical properties of copolymers<sup>19</sup> and automated molecular structure discovery in AFM.<sup>20</sup> The field of TEM also contains many examples, including measurement of carbon nanotube chiral indices<sup>21</sup> and automated classification, and symmetry determination of atomic structures.<sup>22</sup>

Given limited prior knowledge of a system or experiment, we require models that can generalize to novel situations in real-world scenarios. For example, structure determination in the microscope heavily depends on imaging conditions, aberrations, sample orientation, and detector. Existing libraries often poorly capture the resulting variety of potential image types, spectra, and diffraction patterns. Although high-throughput simulations can generate synthetic data for network training, these trained networks are typically material-specific. They cannot yet approximate the vast number of possible real-world imaging conditions. Recent advances in few-shot ML may be used for triaging and classification in scenarios in which we have very limited prior knowledge.<sup>23</sup> This approach is almost entirely unexplored in the context of electron microscopy and analytical characterization more broadly. It can allow us to do more with less and inform otherwise intractable novel situations. Motivated by humans’ ability, especially children, to rapidly learn novel visual concepts by utilizing what they learned in the past, one-shot or few-shot approaches allow human-level performance with fewer and less intensively labeled data (i.e., shots). Studies leveraging these methods in materials science and microscopy are limited,<sup>16,24</sup> but are a critical step toward handling large data streams with few to no annotations. This concept also has significant implications for studying transient behavior and unfolding experiments, where decisions must be made quickly given limited prior knowledge. Old ways of training

and retraining large models on vastly annotated data are not well suited to a rapid “discovery” cycle.

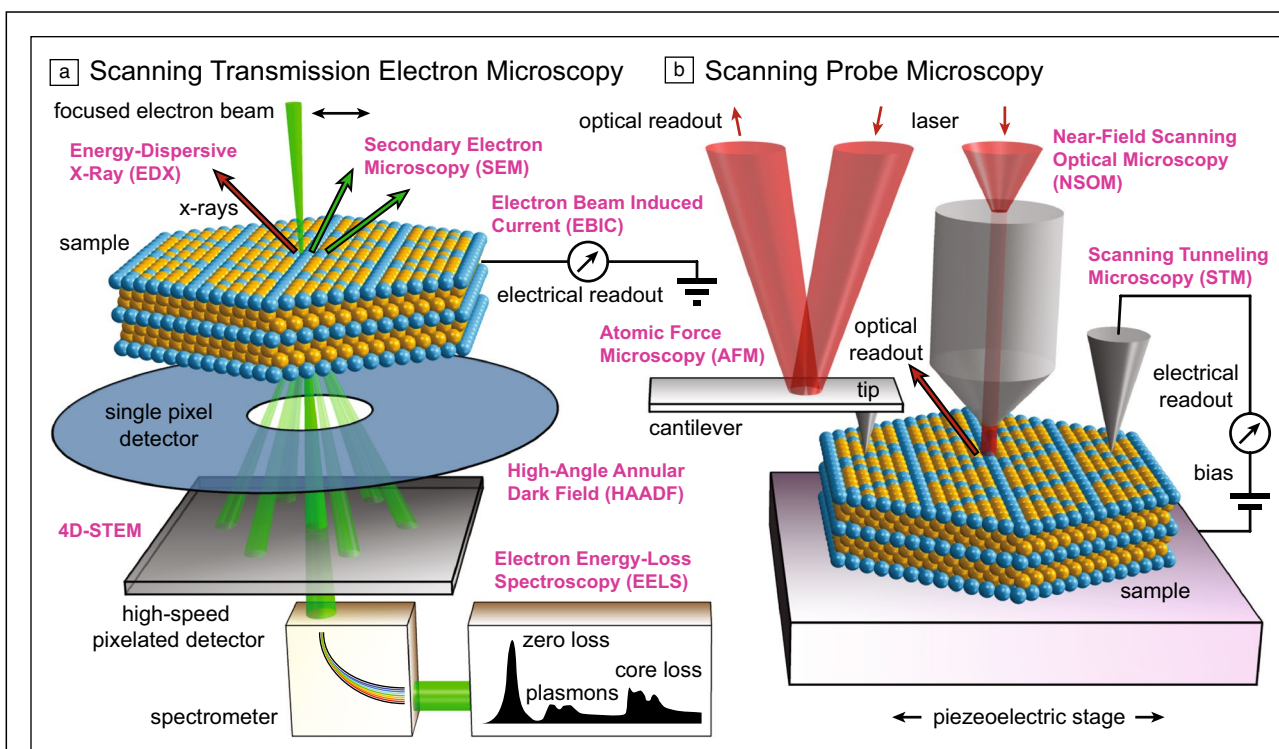
### Discovery via unsupervised learning

The alternative approach for analyzing imaging and spectroscopic data is based on unsupervised learning. In this case, an ML algorithm seeks to discover common traits and variabilities within the high-dimensional data. One approach to address this is to use statistical methods of machine and deep learning to disentangle phenomena in high-dimensional space to bring important physics into focus. Techniques such as clustering, principal component analysis, nonnegative matrix factorization, and dictionary learning are highly effective in quickly probing structure–property relationships in multidimensional imaging but they have an inherent flaw, specifically, probing structure–property relationships in microscopy requires consideration of spatiotemporal relationships in data, whereas the aforementioned ML approaches consider each dimension as independent and uncorrelated. This situation becomes particularly problematic when comparing data translated in space or time.

One partial solution to this problem can be achieved by using autoencoders. Autoencoders generally refer to a class of ML methods aiming to discover low-dimensional representations of the data and are composed of the encoder and decoder parts represented by neural networks. When trained,

the encoder learns a compact statistical representation of the training data distribution that can be decoded back to the original data using the decoder. Because autoencoders are variants of deep neural networks, various neural network building blocks can be used. This allows for incorporation of local spatial dependencies using 2D convolutions and sequential dependencies using 1D convolution or recurrent neural networks.

One rapidly developing example is the variational autoencoder (VAE). In the encoder part of the network, a data set is compressed to a small number of latent variables (latent vector), which the decoder then aims to decode into original data. The training of VAEs balances the reconstruction quality and the proximity of the latent variable distributions to a chosen (typically Gaussian) prior. The unique aspect of autoencoders is that the latent representations of the system often allow one to disentangle the representation of the data (i.e., to discover the systematic traits and align them with the specific latent variables). For example, when disentangling a representation of the hand-written digit (MNIST) data set, one of the latent representations may correspond to the width of the digit, and the other to the tilt. Practically, VAE architectures can be modified to represent variables such as rotation, dilation, and shift as separate latent vectors. The remaining factors of variation can be used to establish order parameters, explore structures, and, in special cases, analyze relevant physical mechanisms.<sup>25</sup>



**Figure 1.** Information channels are available in (a) scanning transmission electron microscopy (STEM), and (b) scanning probe microscopy (SPM) experiments. Measurement techniques are labeled in pink. Figure courtesy of Colin Ophus.

### Probing structure–property relations

Although mere representation of data is important, a central concept in materials science is that of mapping quantitative or semiquantitative structure–property relationships. Advances in spectroscopic modes such as electron energy-loss spectroscopy (EELS) and scanning tunneling spectroscopy (STS) allow us to generalize this concept to atomic- and nanometer-level studies. A variant of scanning tunneling microscopy (STM), STS of individual defects, has been known since the early 1990s. In electron microscopy, advances in EELS allowed the determination of the valence states of the three and fourfold-coordinated Si atom in graphene,<sup>26</sup> followed by P<sup>27</sup> and other elements.<sup>28</sup> In cases when structural descriptors are well defined, this analysis is straightforward. However, in many cases, it is not clear what relevant building blocks define the structure of a solid.

In these cases, ML can be used as a correlative tool to simultaneously simplify the structural and spectral descriptors and establish the relationship between the two. The linear method of this analysis is canonical correlation analysis (CCA), representing the generalization of principal component analysis to two data sources. CCA has been applied to establish the relationship between structural distortions and electron-scattering patterns in graphene with defects.<sup>7</sup> However, in many cases, the image formation mechanisms are nonlinear. In this case, the encoder–decoder architectures allow building correlative relationships as described for the predictability of the plasmon spectra from high-angle annular dark-field (HAADF) STEM data<sup>29</sup> (Figure 2) and decoding relationships between domain structure and functionality in ferroelectric materials.<sup>30</sup>

### Physical discovery

Ultimately, atomic configurations and property maps reflect fundamental information and interactions for atomic assembly. In certain cases, information regarding the processes that lead to the formation of a material can be derived. Hence, a key task for microscopy is to find out whether these laws can be learned from the observational data. A parallel can be drawn to sciences such as astronomy, in which physical laws of celestial mechanics and subsequently Newton's Laws were derived from the observation of planetary motion. Recently, it has been shown that for ferroic materials, including ferroelectrics and ferroelastics, mesoscopic-order parameter fields can be visualized.<sup>31,32</sup> These parameters can be fitted to the Ginzburg–Landau theory prediction to learn the corresponding free energies, gradient terms, and even flexoelectric constants.<sup>33</sup> This approach can be further extended to the atomistic level to learn the Hamiltonians describing atomic interactions.<sup>34</sup>

### Toward automated experimentation

Transformative discoveries in domains such as energy storage, quantum information science, and advanced manufacturing require novel experimental paradigms beyond highly manual,

disconnected, and inefficient experiment architectures. Electron microscopy, a cornerstone of the study of atomic structure, chemistry, and dynamics, exemplifies this challenge. Hardware advances have left us awash with multimodal, high-volume data, leading us to rethink how to make effective decisions in complex, fast-paced experiments.<sup>35</sup> We are now data-rich but faced with increasingly difficult analysis tasks that leave us knowledge-poor. There is presently a transformative opportunity to leverage AI and emerging analytics approaches to accelerate scientific discovery in electron microscopy, laying the groundwork for autonomous experimentation.

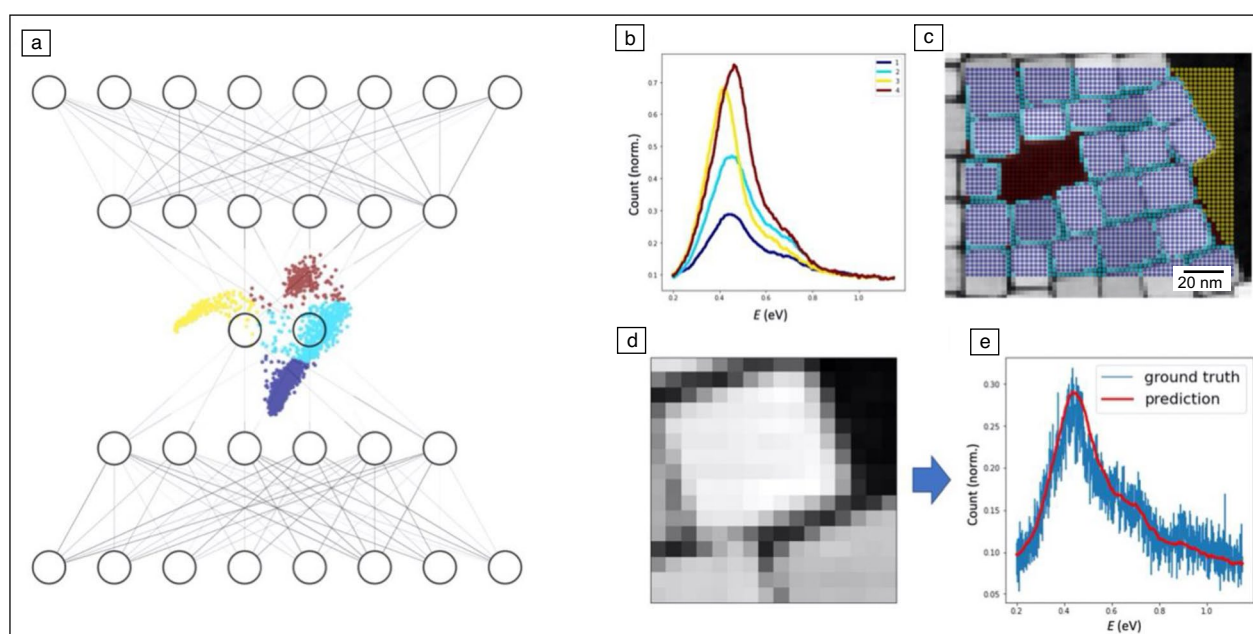
### Self-driving electron microscopy

Fully automated experiments range from simply repeating an (often-debated) “standard recipe” for a given experimental method (open-loop experimentation), to data-driven autonomous discovery platforms, which can, without intervention, optimize measurement or synthesis parameters. Examples of the latter may select samples or regions to probe or modify, or in some cases even, which experiments to perform or which samples to fabricate next (closed-loop experimentation), balancing exploration, and exploitation. One of the key ingredients for the implementation of full automation is tight integration between the experimental hardware, software and algorithms, and control hardware, as well as the ability to modify both to suit experimental needs. Synchrotron beamline experiments are typically custom built, and therefore, many have pioneered full automation. These light source experiments are supported by a range of data science tools, including the Globus data transfer ecosystem<sup>36</sup> and the Data and Learning Hub for ML-enabled modeling.<sup>37</sup> These developments lead to the open question of how the scanning probe and electron microscopy communities will manage the growing volumes of data.

In electron microscopy, the most widespread automation tools are those optimized for biological studies, where samples are relatively uniform and imaging modalities are standardized. One example is the SerialEM platform, which automates EM experiments such as tomographic tilt series collection, as well as automating steps such as stage positioning and microscope defocus tuning.<sup>38</sup> More complex examples include the hardware automation platforms developed for single-particle reconstructions in cryo-EM.<sup>39</sup> These examples only scratch the surface of the possibilities for fully autonomous experiments, which could combine multiple signal channels and measurement modalities to perform arbitrarily complex experiments.

To make meaningful progress toward automation of complex experiments in materials EM and SPM, we must be able to address several gaps in current experimental AI practice. First, we require models capable of analytically enriching large, heterogeneous, and complex data sets in real time with minimal human intervention. Such multimodal data are highly varied in their dimensionality, feature types, and artifacts, motivating the need for more flexible and transferable approaches to classification and segmentation, particularly





**Figure 2.** (a) Illustration of the encoder–decoder model and its latent space. The inputs and outputs can be structural image patches and 1D spectra measured in those patches. In this example, the inputs are high-angle annular dark-field (HAADF) scanning transmission electron microscope (STEM) image patches, and the outputs are corresponding electron energy-loss spectroscopy (EELS) spectra. (b) Distinct plasmonic responses uncovered from the analysis of the latent space and (c) their spatial locations on the HAADF STEM image. (d, e) Prediction of EELS spectra from a structural image using a trained encoder–decoder model. Data from K. Roccapriore (Oak Ridge National Laboratory). Figure courtesy of M. Ziatdinov.

as we move toward self-driving instrumentation. Second, we require models that can generalize and adapt to novel scenarios, given limited prior knowledge of a system or experiment. In contrast to other experimental techniques, the reconstruction of a structure in the electron microscope heavily depends on imaging conditions, aberrations, orientation, and the detector. The wide variety of potential image types, spectra, and diffraction patterns is poorly captured by existing libraries. High-throughput simulations can help generate some synthetic data for network training, but it is still difficult to approximate the vast number of possible real-world imaging conditions. Third, we require physically grounded, interpretable models from which we can derive actionable metrics for control systems and automation. As described in the preceding section, domain knowledge must be integrated into these models from the outset to ensure that meaningful features and behaviors are detected, particularly if a control system is to anticipate and decide on the next steps of an experiment in a robust manner. Recent efforts in this direction include sparse-data-guided electron microscope platforms,<sup>40</sup> which leverage a handful of user-provided examples to guide automated decision-making. These developments show the potential of a flexible and domain-aware approach to analytics and control unconstrained by a lack of prior knowledge. Finally, we need to consider how computational hardware can be co-designed to meet the latency and throughput requirements of an experiment, taking

advantage of both cloud and edge-based computing systems. There is also an emerging opportunity to design automation for custom hardware solutions that can achieve inference times on the order of nanoseconds.<sup>41</sup>

Realization of autonomous experiments necessitates a definition of prior knowledge and goals. These considerations, in turn, determine the selection of enabling algorithms. For example, automated experimentation based on uniform imaging over a large grid of points assumes zero prior knowledge and explores material everywhere. Similarly, spectroscopic measurements at locations with *a priori* known structures of interest require ML methods at the segmentation stage. Conversely, the discovery of specific structural and spectral elements of interest and sparse image reconstructions requires more advanced Gaussian process-based techniques. Recently, it has been shown that deep kernel learning captures a combination of the expressive power of the deep convolutional networks and flexibility of Gaussian processes to enable discovery of structure–property relationships in scanning probe and electron microscopy.<sup>42,43</sup> There is a tremendous opportunity for automated experiments that control perturbation energy and frequency to capture spatiotemporal phenomena associated with specific materials structures.

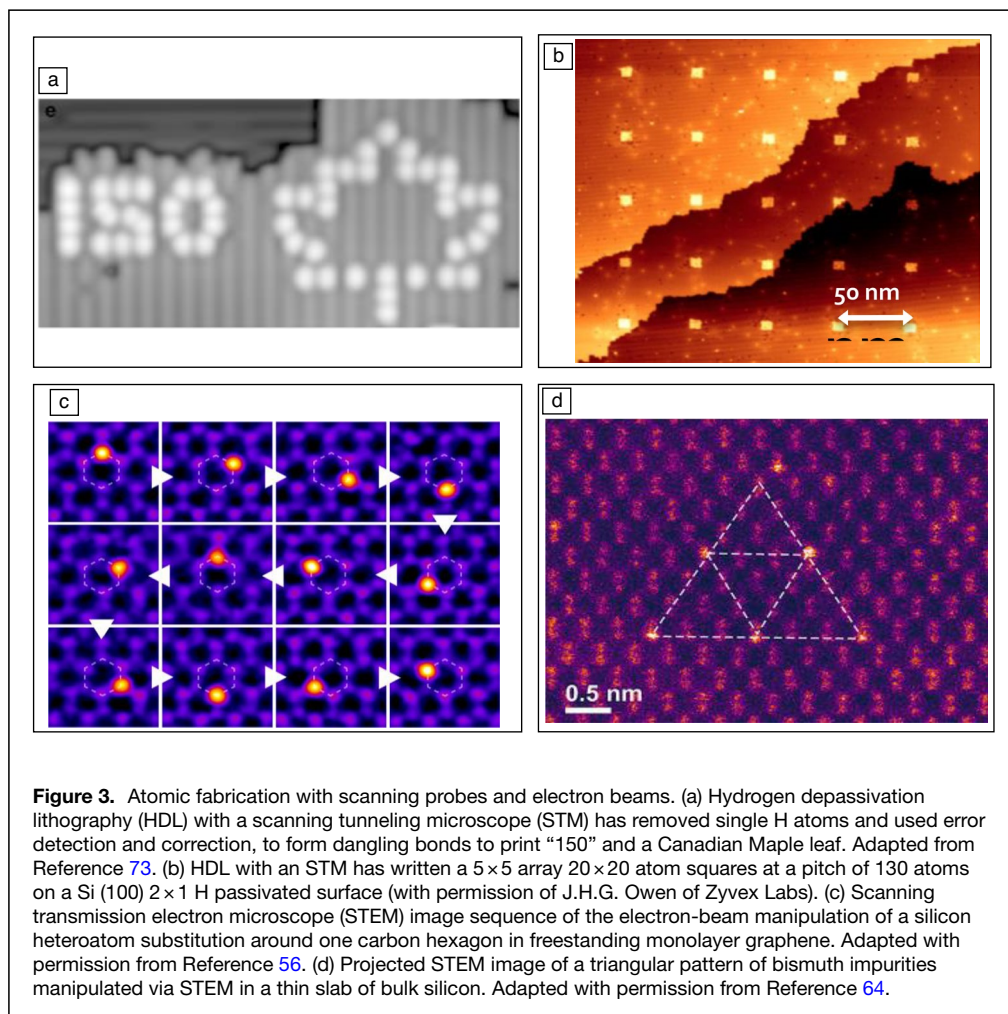
The unique opportunity opened by the quantitative microscopies is the fundamental studies of structure– and composition–property relationships. Even nominally homogeneous materials contain atomic-level fluctuations, thus, encompassing

the range of local chemical composition and environments. Learning local structure–property relationships allows for probing broad regions of chemical spaces. This approach can be further expanded by combining the local characterization and macro- or mesoscale combinatorial libraries. Here, the key requirement is the instantiation of ML methods that contain information on possible physical mechanisms underlying materials properties and refinement of them throughout the experiment.

### Mesosopic and atomic fabrication

A unique niche for automated experimentation is the possibility to not only measure previously fabricated structures, but to also use the imaging modality itself as a manipulation tool. Although light-based lithography is widely used in semiconductor manufacturing and is reaching ever-smaller dimensions, the wavelength of light places a hard limit on the dimensionality. Truly atomic-level manipulation can only be achieved with atomic-sized probes. Richard Feynman pointed out more than 60 years ago that nothing in the laws of physics prohibits this ultimate limit of materials engineering.<sup>44</sup> It took several decades for technological developments to catch up, first with the invention of STM in the early 1980s,<sup>45</sup> and then with the development of effective aberration correctors for STEM by the early 2000s.<sup>46</sup>

The power to manipulate individual atoms was established for the first time by pioneering STM experiments in the early 1990s.<sup>7</sup> This soon led to the creation of atomic assemblies with controllable collective quantum properties.<sup>47,48</sup> Although STM is primarily limited to surfaces held at cryogenic temperatures in ultrahigh vacuum, the technique of hydrogen depassivation lithography (HDL)<sup>49</sup> has enabled heteroatom placement on semiconductor surfaces followed by crystal overgrowth to pattern subsurface donor atoms<sup>50</sup> and to fabricate solid-state qubits.<sup>51</sup> Despite less widely used, atomic force microscopy



**Figure 3.** Atomic fabrication with scanning probes and electron beams. (a) Hydrogen depassivation lithography (HDL) with a scanning tunneling microscope (STM) has removed single H atoms and used error detection and correction, to form dangling bonds to print “150” and a Canadian Maple leaf. Adapted from Reference 73. (b) HDL with an STM has written a  $5 \times 5$  array  $20 \times 20$  atom squares at a pitch of 130 atoms on a Si (100)  $2 \times 1$  H passivated surface (with permission of J.H.G. Owen of Zyvex Labs). (c) Scanning transmission electron microscope (STEM) image sequence of the electron-beam manipulation of a silicon heteroatom substitution around one carbon hexagon in freestanding monolayer graphene. Adapted with permission from Reference 56. (d) Projected STEM image of a triangular pattern of bismuth impurities manipulated via STEM in a thin slab of bulk silicon. Adapted with permission from Reference 64.

(AFM) can also be used for atom manipulation/fabrication, notably in some cases even at room temperature.<sup>52</sup> Despite many successes, the inherent limitation of SPM techniques to a slowly scanning physical tip in contact with a surface has prompted a search for alternatives.

The possibility to manipulate individual atoms using the Å-sized electron probe of an aberration-corrected STEM was discovered in 2014, alongside a mechanism for the nondestructive movement of Si impurities in graphene.<sup>53</sup> Controlled experiments for Si in graphene<sup>5,54–56</sup> and single-walled carbon nanotubes<sup>57</sup> were soon reported, although this proved more challenging for P<sup>58</sup> and the most common dopants in graphene, N and B.<sup>59</sup> From 2016, increasing attention was drawn to the unique advantages of STEM in being able to address atoms within bulk crystals,<sup>60,61</sup> alongside the first experimental demonstrations of directed crystallization and amorphization,<sup>62</sup> dopant movement,<sup>63</sup> and finally, the controlled manipulation of Bi dopants in bulk silicon,<sup>64</sup> the latter exhibiting a novel nondestructive mechanism.<sup>65</sup>

Although SPM automation is already commonplace, the first efforts to apply ML and automation to STEM atomic fabrication are underway, as shown in Figure 3. Critical components, such as neural network tools to automatically locate the atoms

of the lattice<sup>66,67</sup> including any impurities, are already operating with real-time integration into microscope control software,<sup>68</sup> while automated electron-beam placement and various detector feedback schemes are being explored.<sup>56,69</sup> Overall, the modern computerized STEM is already an excellent platform for atomic fabrication,<sup>70</sup> and even more, rapid progress is primarily held back by the difficulty of obtaining high-quality samples with large impurity concentrations and atomically clean surfaces, as well as the presence of competing processes including electron-beam damage<sup>71</sup> and uncontrolled chemical modifications.<sup>57,58,72</sup>

In summary, broad adoption of ML methods in scanning probe and electron microscopy offers a clear pathway to facilitate and automate the analysis of data. These methods also show promise to guide learning of underpinning physical mechanisms. Harnessing ML as a part of the experiment holds promise for enabling automated microscopy experiments, targeting exploration of predefined objects of interest, discovering structure–property relationships, and ultimately testing multiple physical hypotheses. Finally, ML methods in microscopy open the pathway toward direct atomic fabrication via scanning probes and electron beams, which is a critical step toward quantum computing and other serendipitous developments.

## Acknowledgments

This effort (S.V.K.) is based upon work supported by the US Department of Energy (DOE), Office of Science, Basic Energy Sciences (BES), Materials Sciences and Engineering Division, and was performed and partially supported (M.Z.) at the Oak Ridge National Laboratory's Center for Nanophase Materials Sciences (CNMS), a US Department of Energy, Office of Science User Facility. S.R.S. was supported by the Energy Storage Materials Initiative (ESMI), under the Laboratory Directed Research and Development (LDRD) Program at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-program national laboratory operated for the US Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830. Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under Contract No. DE-AC02-05CH11231. This material (J.C.A.) is based upon work supported by the US Department of Energy, Office of Science, Office of Advanced Scientific Computing Research under Award No. DE-SC-0002501, and the US Army Research Laboratory under Grant No. W911NF-19-2-0119. T.S. was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No. 756277-ATMEN). E.A.S. was supported by the National Science Foundation through the University of Pennsylvania Materials Research Science and Engineering Center (MRSEC) (DMR-1720530). The work at Zyvex Labs is based upon work supported by the US Department of Energy Office of Energy Efficiency and Renewable Energy (EERE) under Advanced Manufacturing Office Award No. DEEE0008311.

## Data availability

The data sets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Conflict of interest

The authors have no conflicts of interest.

## References

1. S.J. Pennycook, *Ultramicroscopy* **180**, 22 (2017)
2. Y. Jiang, Z. Chen, Y.M. Hang, P. Deb, H. Gao, S.E. Xie, P. Purohit, M.W. Tate, J. Park, S.M. Gruner, V. Elser, D.A. Muller, *Nature* **559**(7714), 343 (2018)
3. O.L. Krivanek, T.C. Lovejoy, N. Dellby, T. Aoki, R.W. Carpenter, P. Rez, E. Soignard, J.T. Zhu, P.E. Batson, M.J. Lagos, R.F. Egerton, P.A. Crozier, *Nature* **514**(7521), 209 (2014)
4. M. Varela, S.D. Findlay, A.R. Lupini, H.M. Christen, A.Y. Borisevich, N. Dellby, O.L. Krivanek, P.D. Nellist, M.P. Oxley, L.J. Allen, S.J. Pennycook, *Phys. Rev. Lett.* **92**(9), 095502 (2004)
5. O. Dyck, S. Kim, E. Jimenez-Izal, A.N. Alexandrova, S.V. Kalinin, S. Jesse, *Small* **14**(38), 1801771 (2018). <https://doi.org/10.1002/sml.201801771>
6. K.A. Moler, *Nat. Mater.* **16**(11), 1049 (2017)
7. D.M. Eigler, E.K. Schweizer, *Nature* **344**(6266), 524 (1990)
8. Y. LeCun, Y. Bengio, G. Hinton, *Nature* **521**(7553), 436 (2015)
9. R. Batra, L. Song, R. Ramprasad, *Nat. Rev. Mater.* **6**(8), 655 (2021)
10. E. Stach, B. DeCost, A.G. Kusne, J. Hattrick-Simpers, K.A. Brown, K.G. Reyes, J. Schrier, S. Billinge, T. Buonassisi, I. Foster, C.P. Gomes, J.M. Gregoire, A. Mehta, J. Montoya, E. Olivetti, C. Park, E. Rotenberg, S.K. Saikin, S. Smullin, V. Stanev, B. Maruyama, *Matter* **4**(9), 2702 (2021)
11. S.R. Kalidindi, M. De Graef, *Annu. Rev. Mater. Res.* **45**, 171 (2015)
12. G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J. van der Laak, B. van Ginneken, C.I. Sanchez, *Med. Image Anal.* **42**, 60 (2017)
13. J.P. Horwath, D.N. Zakharov, R. Mégret, E.A. Stach, *NPJ Comput. Mater.* **6**, 108 (2020). <https://doi.org/10.1038/s41524-020-00363-x>
14. S.V. Kalinin, E. Strelcov, A. Belianinov, S. Somnath, R.K. Vasudevan, E.J. Lingerfelt, R.K. Archibald, C.M. Chen, R. Proksch, N. Laanait, S. Jesse, *ACS Nano* **10**(10), 9068 (2016)
15. M. Ziatdinov, O. Dyck, A. Maksov, X.F. Li, X.H. San, K. Xiao, R.R. Unocic, R. Vasudevan, S. Jesse, S.V. Kalinin, *ACS Nano* **11**(12), 12742 (2017)
16. S. Akers, E. Kautz, A. Trevino-Gavito, M. Olszta, B.E. Matthews, L. Wang, Y.G. Du, S.R. Spurgeon, *NPJ Comput. Mater.* **7**, 187 (2021). <https://doi.org/10.1038/s41524-021-00621-z>
17. M. Rashidi, R.A. Wolkow, *ACS Nano* **12**(6), 5185 (2018)
18. O.M. Gordon, J.E.A. Hodgkinson, S.M. Farley, E.L. Hunsicker, P.J. Moriarty, *Nano Lett.* **20**(10), 7688 (2020)
19. D. Yablou, I. Chakraborty, H. Passino, K. Iyer, A. Doufas, M. Shivokhin, *MRS Commun.* **11**(6), 962 (2021)
20. B. Alldritt, P. Hapala, N. Oinonen, F. Urtev, O. Krejci, F.F. Canova, J. Kannala, F. Schulz, P. Liljeroth, A.S. Foster, *Sci. Adv.* **6**(9) eaay6913 (2020). <https://doi.org/10.1126/sciadv.aay6913>
21. G.D. Forster, A. Castan, A. Loiseau, J. Nelayah, D. Alloyeau, F. Fossard, C. Bichara, H. Amara, *Carbon* **169**, 465 (2020)
22. S.V. Kalinin, M.P. Oxley, M. Valletti, J.J. Zhang, R.P. Hermann, H. Zheng, W.R. Zhang, G. Eres, R.K. Vasudevan, M. Ziatdinov, *NPJ Comput. Mater.* **7**, 181 (2021). <https://doi.org/10.1038/s41524-021-00621-6>
23. C. Finn, P. Abbeel, S. Levine, *34th International Conference on Machine Learning (ICML 2017)* (Sydney, August 6–11, 2017), pp. 1126–1135
24. K. Kaufmann, H. Lane, X. Liu, K.S. Vecchio, *Sci. Rep.* **11**, 8172 (2021). <https://doi.org/10.1038/s41598-021-87557-5>
25. J.C. Agar, B. Naul, S. Pandya, S. van der Walt, J. Maher, Y. Ren, L.-Q. Chen, S.V. Kalinin, R.K. Vasudevan, Y. Cao, J.S. Bloom, L.W. Martin, *Nat. Commun.* **10**(1), 4809 (2019)
26. W. Zhou, M. Kapetanakis, M. Prange, S. Pantelides, S. Pennycook, J.-C. Idrobo, *Phys. Rev. Lett.* **109**(20), 206803 (2012)
27. T. Susi, T.P. Hardcastle, H. Hofsäss, A. Mittelberger, T.J. Pennycook, C. Mangler, R. Drummond-Brydson, A.J. Scott, J.C. Meyer, J. Kotakoski, *Materials* **4**(2), 021013 (2017)
28. Y.-C. Lin, P.-Y. Teng, P.-W. Chiu, K. Suenaga, *Phys. Rev. Lett.* **115**, 206803 (2015)
29. K.M. Roccapiore, M. Ziatdinov, S.H. Cho, J.A. Hachtel, S.V. Kalinin, *Small* **17**(21), 2100181 (2021). <https://doi.org/10.1002/sml.202100181>
30. S.V. Kalinin, K. Kelley, R.K. Vasudevan, M. Ziatdinov, *ACS Appl. Mater. Interfaces* **13**(1), 1693 (2021)
31. C.L. Jia, S.B. Mi, K. Urban, I. Vrejoiu, M. Alexe, D. Hesse, *Nat. Mater.* **7**(1), 57 (2008)



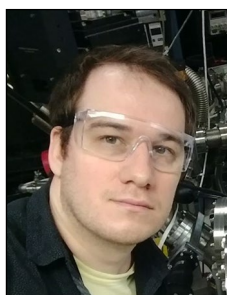
32. A.Y. Borisevich, E.A. Eliseev, A.N. Morozovska, C.-J. Cheng, J.-Y. Lin, Y.H. Chu, D. Kan, I. Takeuchi, V. Nagarajan, S.V. Kalinin, *Nat. Commun.* **3**, 775 (2012). <https://doi.org/10.1038/ncomms1778>
33. Q. Li, C.T. Nelson, S.-L. Hsu, A.R. Damodaran, L.-L. Li, A.K. Yadav, M. McCarter, L.W. Martin, R. Ramesh, S.V. Kalinin, *Nat. Commun.* **8**, 1468 (2017). <https://doi.org/10.1038/s41467-017-01733-8>
34. L. Vleck, M. Ziatdinov, A. Maksov, A. Tselev, A.P. Baddorf, S.V. Kalinin, R.K. Vasudevan, *ACS Nano* **13**(1), 718 (2019)
35. S.R. Spurgeon, C. Ophus, L. Jones, A. Petford-Long, S.V. Kalinin, M.J. Olszta, R.E. Dunin-Borkowski, N. Salmon, K. Hattar, W.D. Yang, R. Sharma, Y. Du, A. Chiaramonti, H. Zheng, E.C. Buck, L. Kovarik, R.L. Penn, D. Li, X. Zhang, M. Murayama, M.L. Taheri, *Nat. Mater.* **20**(3), 274 (2021). <https://doi.org/10.1038/s41563-020-00833-z>
36. I. Foster, C. Kesselman, *Int. J. Supercomput. Appl. High Perform. Comput.* **11**(2), 115 (1997)
37. B. Blaiszik, K. Chard, R. Chard, I. Foster, L. Ward, presented at the 13th International Conference on Synchrotron Radiation Instrumentation (SRI 2018), National Synchrotron Radiation Research Center, Taipei, June 11–16, 2018
38. D.N. Mastrorade, *J. Struct. Biol.* **152**(1), 36 (2005)
39. M. Fislage, A.V. Shkumatov, A. Stroobants, R.G. Efremov, *IUCrJ* **7**, 707 (2020)
40. M. Olszta, D. Hopkins, K.R. Fiedler, M. Oostrom, S. Akers, S.R. Spurgeon, *Microsc. Microanal.* **28**(5), 1611 (2022). <https://doi.org/10.1017/S1431927622012065>
41. A.M. Deiana, N. Tran, J. Agar, M. Blott, G. Di Guglielmo, J. Duarte, P. Harris, S. Hauck, M. Liu, M.S. Neubauer, arXiv preprint, [arXiv:2110.13041](https://arxiv.org/abs/2110.13041) (2021)
42. Y. Liu, K. P. Kelley, R.K. Vasudevan, H. Funakubo, M.A. Ziatdinov, S.V. Kalinin, arXiv preprint, [arXiv:2108.06037](https://arxiv.org/abs/2108.06037) (2021)
43. K.M. Roccapriore, S.V. Kalinin, M. Ziatdinov, arXiv preprint, [arXiv:2108.03290](https://arxiv.org/abs/2108.03290) (2021)
44. R.P. Feynman, *Eng. Sci.* **23**(5), 22 (1960)
45. G. Binnig, H. Rohrer, *Rev. Mod. Phys.* **59**(3), 615 (1987)
46. P.W. Hawkes, *Philos. Trans. R. Soc. Lond. A* **367**(1903), 3637 (2009)
47. M.F. Crommie, C.P. Lutz, D.M. Eigler, *Science* **262**(5131), 218 (1993)
48. A.A. Khajetoorians, D. Wegner, A.F. Otte, I. Swart, *Nat. Rev. Phys.* **1**(12), 703 (2019)
49. J.N. Randall, J.W. Lyding, S. Schmucker, J.R. Von Ehr, J. Ballard, R. Saini, H. Xu, Y. Ding, *J. Vac. Sci. Technol. B* **27**(6), 2764 (2009)
50. T.J.Z. Stock, O. Warschkow, P.C. Constantinou, J. Li, S. Fearn, E. Crane, E.V.S. Hofmann, A. Köler, D.R. McKenzie, S.R. Schofield, N.J. Curson, *ACS Nano* **14**(3), 3316 (2020)
51. Y. He, S.K. Gorman, D. Keith, L. Kranz, J.G. Keizer, M.Y. Simmons, *Nature* **571**(7765), 371 (2019)
52. O. Custance, R. Perez, S. Morita, *Nat. Nanotechnol.* **4**(12), 803 (2009)
53. T. Susi, J. Kotakoski, D. Kepaptsoglou, C. Mangler, T.C. Lovejoy, O.L. Krivanek, R. Zan, U. Bangert, P. Ayala, J.C. Meyer, Q. Ramasse, *Phys. Rev. Lett.* **113**, 115501 (2014). <https://doi.org/10.1103/PhysRevLett.113.115501>
54. T. Susi, J.C. Meyer, J. Kotakoski, *Ultramicroscopy* **180**, 163 (2017)
55. O. Dyck, S. Kim, S.V. Kalinin, S. Jesse, *Appl. Phys. Lett.* **111**(11), 113104 (2017)
56. M. Tripathi, A. Mittelberger, N.A. Pike, C. Mangler, J.C. Meyer, M.J. Verstraete, J. Kotakoski, T. Susi, *Nano Lett.* **18**(8), 5319 (2018)
57. K. Mustonen, A. Markevich, M. Tripathi, H. Inani, E.-X. Ding, A. Hussain, C. Mangler, E.I. Kauppinen, J. Kotakoski, T. Susi, *Adv. Funct. Mater.* **29**(52), 1901327 (2019)
58. C. Su, M. Tripathi, Q.-B. Yan, Z. Wang, Z. Zhang, C. Hofer, H. Wang, L. Basile, G. Su, M. Dong, J.C. Meyer, J. Kotakoski, J. Kong, J.C. Idrobo, T. Susi, J. Li, *Sci. Adv.* **5**, 2252 (2019)
59. T. Susi, D. Kepaptsoglou, Y.-C. Lin, Q.M. Ramasse, J.C. Meyer, K. Suenaga, J. Kotakoski, *2D Mater.* **4**(4), 042004 (2017)
60. S.V. Kalinin, A. Borisevich, S. Jesse, *Nature* **539**, 485 (2016)
61. S. Jesse, A.Y. Borisevich, J.D. Fowlkes, A.R. Lupini, P.D. Rack, R.R. Unocic, B.G. Sumpter, S.V. Kalinin, A. Belianinov, O.S. Ovchinnikova, *ACS Nano* **10**(6), 5600 (2016)
62. S. Jesse, Q. He, A.R. Lupini, D.N. Leonard, M.P. Oxley, O. Ovchinnikov, R.R. Unocic, A. Tselev, M. Fuentes-Cabrera, B.G. Sumpter, S.J. Pennycook, S.V. Kalinin, A.Y. Borisevich, *Small* **11**(44), 5895 (2015)
63. S. Jesse, B.M. Hudak, E. Zarkadoulas, J. Song, A. Maksov, M. Fuentes-Cabrera, P. Ganesh, I. Kravchenko, P.C. Snijders, A.R. Lupini, A.Y. Borisevich, S.V. Kalinin, *Nanotechnology* **29**(25), 255303 (2018)
64. B.M. Hudak, J. Song, H. Sims, M.C. Troparevsky, T.S. Humble, S.T. Pantelides, P.C. Snijders, A.R. Lupini, *ACS Nano* **12**(6), 5873 (2018)
65. A. Markevich, B.M. Hudak, J. Madsen, J. Song, P.C. Snijders, A.R. Lupini, T. Susi, *J. Phys. Chem. C Nanomater. Interfaces* **125**(29), 16041 (2021)
66. M. Ziatdinov, O. Dyck, A. Maksov, X. Li, X. Sang, K. Xiao, R.R. Unocic, R. Vasudevan, S. Jesse, S.V. Kalinin, *ACS Nano* **11**(12), 12742 (2017)
67. J. Madsen, P. Liu, J. Kling, J.B. Wagner, T.W. Hansen, O. Winther, J. Schjøtt, *Adv. Theory Simul.* **1**(8), 1800037 (2018)
68. J. Madsen, A. Postl, T. Susi, *Microsc. Microanal.* **25**(S2), 166 (2019)
69. X. Li, O. Dyck, S.V. Kalinin, S. Jesse, *Microsc. Microanal.* **25**(S2), 1688 (2019)
70. O. Dyck, M. Ziatdinov, D.B. Lingerfelt, R.R. Unocic, B.M. Hudak, A.R. Lupini, S. Jesse, S.V. Kalinin, *Nat. Rev. Mater.* **4**(7), 497 (2019)
71. T. Susi, J.C. Meyer, J. Kotakoski, *Nat. Rev. Phys.* **1**, 397 (2019)

72. M. Tripathi, A. Markevich, R. Böttger, S. Fackso, E. Besley, J. Kotakoski, T. Susi, *ACS Nano* **12**(5), 4641 (2018)
73. R. Achal, M. Rashidi, J. Croshaw, D. Churchill, M. Taucer, T. Huff, M. Cloutier, J. Pitters, R.A. Wolkow, *Nat. Commun.* **9**, 8 (2018). <https://doi.org/10.1038/s41467-018-05171-y> □

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.



**Sergei V. Kalinin** is a professor at The University of Tennessee, Knoxville, following 20 years at Oak Ridge National Laboratory. His research focuses on the applications of machine learning in scanning transmission electron microscopy and scanning probes for physics discovery and atomic fabrication, as well as mesoscopic studies of electrochemical and ferroelectric phenomena. He is a Fellow of the Materials Research Society, American Physical Society, Microscopy Society of America, Institute of Physics, IEEE Foresight Institute, and the American Vacuum Society; a recipient of the Blavatnik Award for Physical Sciences (2018), Burton Medal from the Microscopy Society of America (2010), four R&D100 Awards, and a number of other distinctions. Kalinin can be reached by email at [sergei2@utk.edu](mailto:sergei2@utk.edu).



**Maxim Ziatdinov** received his doctorate. He is a research scientist and team lead at Oak Ridge National Laboratory. His research is directed primarily toward the synergy of the experiment, theory, and machine learning for accelerated discoveries in the fields of materials science and nanoscience. He is particularly interested in developing physics-based machine learning approaches for analyzing complex molecular, atomic, and electronic patterns from scanning probe and electron microscopy experiments on classical and quantum materials. He is also actively working on the development of the machine-learning-enabled “smart” automation of imaging and spectroscopic instrumentation. Ziatdinov can be reached by email at [ziatdinovma@ornl.gov](mailto:ziatdinovma@ornl.gov).



**Steven R. Spurgeon** is a senior research scientist at Pacific Northwest National Laboratory and affiliate associate professor of physics at the University of Washington. His work focuses on developing artificial intelligence and data-driven approaches to accelerate the discovery and design of nanomaterials for next-generation electronics, quantum computing, and energy storage. He has published more than 60 journal articles and book chapters and has received awards from the US Department of Energy, the National Science Foundation, the Materials Research Society, the Microscopy Society of America, and the US Department of Defense. Spurgeon can be reached by email at [steven.spurgeon@pnnl.gov](mailto:steven.spurgeon@pnnl.gov).



**Colin Ophus** is a staff scientist at the National Center for Electron Microscopy at Lawrence Berkeley Laboratory, running both a user program and research group focused on methods, algorithms, and codes for simulation, analysis, and instrument design for transmission electron microscopy. In 2018, he received a US Department of Energy Early Career Research Award, and in 2022 was awarded the Burton Medal from the Microscopy Society of America. He is a project leader for the Prismatic scanning transmission electron microscopy (STEM) simulation and py4DSTEM open-source codes. Ophus can be reached by email at [dlophus@lbl.gov](mailto:dlophus@lbl.gov).





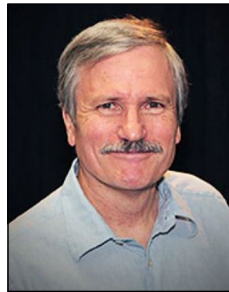
**Eric A. Stach** is the Robert D. Bent Professor of Engineering in the Department of Materials Science and Engineering at the University of Pennsylvania, and Director of the Laboratory for Research on the Structure of Matter. He received his PhD degree in materials science and engineering from the University of Virginia. Stach is a co-founder and Chief Technology Officer of Hummingbird Scientific. He is a Fellow of the American Physical Society, Materials Research Society, and the Microscopy Society of America. Stach can be reached by email at [stach@seas.upenn.edu](mailto:stach@seas.upenn.edu).



**Josh Agar** leads the M3-Learning: Multifunctional Materials and Machine Learning Group and is an assistant professor in the Department of Mechanical Engineering and Mechanics at Drexel University. His research interests center around codesign of experimental synthesis and spectroscopy, machine learning, and heterogeneous computing. Agar earned his PhD degree from the University of Illinois at Urbana-Champaign. He was a postdoctoral fellow at the University of California, Berkeley. His work has been published and highlighted in a range of high-impact journals across scientific disciplines. Agar can be reached by email at [jca92@drexel.edu](mailto:jca92@drexel.edu).



**Toma Susi** is an associate professor of physics at the University of Vienna, Austria. He received his doctorate from Aalto University, Finland, in 2011, and moved to the University of Vienna in 2013. Susi received the ERC Starting Grant in 2017, and researches new ways to manipulate materials at the atomic level. He has worked on materials synthesis, spectroscopy, electron microscopy, and modeling, authoring more than 70 peer-reviewed articles and reviews mostly on heteroatom-doped graphene and carbon nanotubes. Susi can be reached by email at [toma.susi@univie.ac.at](mailto:toma.susi@univie.ac.at).



**John Randall** is CEO of Zyvex Labs, and a Fellow of the American Vacuum Society, IEEE, and the Micro Nano Engineering Society, and has 40 years of experience in Micro-Nano-Fabrication. He joined Zyvex in 2001 after 15 years at Texas Instruments (TI) where he worked in high-resolution processing for integrated circuits, microelectromechanical systems, and quantum devices. Prior to TI, Randall worked at the Massachusetts Institute of Technology's Lincoln Laboratory. He has 119 articles published in refereed journals, 34 issued US patents, and other publications with 5554 citations. Randall can be reached by email at [jrandall@zyvexlabs.com](mailto:jrandall@zyvexlabs.com).