Towards data-driven next-generation transmission electron microscopy

Electron microscopy touches on nearly every aspect of modern life, underpinning materials development for quantum computing, energy and medicine. We discuss the open, highly integrated and data-driven microscopy architecture needed to realize transformative discoveries in the coming decade.

Steven R. Spurgeon, Colin Ophus, Lewys Jones, Amanda Petford-Long, Sergei V. Kalinin, Matthew J. Olszta, Rafal E. Dunin-Borkowski, Norman Salmon, Khalid Hattar, Wei-Chang D. Yang, Renu Sharma, Yingge Du, Ann Chiaramonti, Haimei Zheng, Edgar C. Buck, Libor Kovarik, R. Lee Penn, Dongsheng Li, Xin Zhang, Mitsuhiro Murayama and Mitra L. Taheri

From its inception nearly a century ago, transmission electron microscopy (TEM) has emerged as a cornerstone of characterization in materials science, chemistry, physics and medicine[1]. TEM provides rich, directly resolved information about the structure and dynamics of phenomena spanning atoms to micrometres that are of great fundamental and practical significance to society. It has played a key role in protein and drug discovery[2], redefined our understanding of crystalline solids[3] and catalysed the electronics revolution that gave rise to today's massively interconnected world[4].

In spite of these numerous successes, many grand materials challenges remain outside of our present capabilities. Mastery of quantum phenomena, for example, requires insight into subtle and dilute electronic perturbations that can only be probed through sensitive multi-modal analyses closely linked to theory. Control of chemical reaction pathways in catalysts depends on access to interchangeable, finely tuned environments and the cumulative knowledge of a large library of prior experiments. True combinatorial engineering of high-entropy alloys demands on-the-fly experimental decision-making based on automated characterization. In these domains and more, a reimagined microscopy paradigm is needed to unlock entirely new classes of materials and functionality.

As in many other areas of science[5], advances in TEM instrumentation now permit the rapid generation of vast data sets across a range of modalities, in which important connections might be more easily overlooked. Counterintuitively, microscopists focus on the methods already familiar to them, rather than harnessing more suitable tools from the full suite at their disposal. This situation is compounded by the growing complexity and closed-source nature of modern microscopes, which limit our ability and motivation to fully understand and customize their operation. Due to these barriers, the much-lauded promise of artificial intelligence (AI) and machine learning (ML) to revolutionize TEM experiment design, execution and analysis has not yet been realized. In contrast, other fields such as X-ray crystallography that have adopted open, standardized methods and data exchanges have witnessed enormous success[6]. Automated X-ray experimentation is now routinely conducted at scale, aided by easily accessible libraries of past work to plan and interpret future studies. Additive manufacturing is another area in which shared repositories of blueprints and techniques have empowered end users to conduct experimentation never imagined by their original creators. In electron microscopy, the growth of single-particle cryo-imaging demonstrates the untapped potential of automated 'big data' tools[7] to transform our understanding of metals, semiconductors, ceramics and more.

Sweeping changes precipitated by recent technological innovations and the growth of modern data science tools call for a re-examination of the electron microscopy framework, shown in Fig. 1. This framework aims to discover knowledge about an unknown materials structure or process, employing a priori assumptions and an array of microscopy tools to probe different features of the unknown system. These features can then be distilled into salient physical mechanisms and quantifiable metrics through the eyes of various scientific disciplines. While there are many ways to define this framework, we broadly divide it into three overlapping categories: experiment design, feature extraction and knowledge discovery. These generally applicable categories provide a basis to understand the present state-of-the-art and its shortcomings, with a focus here on the application of data techniques to the physical sciences. In particular, we argue that an open, highly integrated and data-driven framework will transform characterization in the next-generation transmission electron microscope, benefitting both the physical and the biological sciences.

Experiment design

The first step in the analysis process involves the definition of unknowns and the selection of appropriate techniques to explore them, as shown in the top of Fig. 1. This stage is by its very nature based on pre-existing knowledge of a system or process, derived through prior work, intuition and an understanding of the characterization tools available to the researcher. In such an analysis, the investigator leverages the complementary strengths of both parallel-beam and scanning TEM (STEM) imaging and spectroscopy, which have enabled study of the structure and local properties of materials at high resolution. Recent developments such as ultra-stable cryo stages[8], data-rich high-speed detectors[9], and atomic-scale electron tomography[10] have provided a wealth of new imaging modalities waiting to be exploited. For example, hybrid pixel detectors[9] now offer sufficient dynamic range and sensitivity to record full diffraction patterns at each point on a sample, enabling ptychographic reconstruction and unprecedented spatial resolution at low voltages[11]. An alternative detector technology, back-thinned monolithic active pixel sensors that can count individual incident electrons[12], has

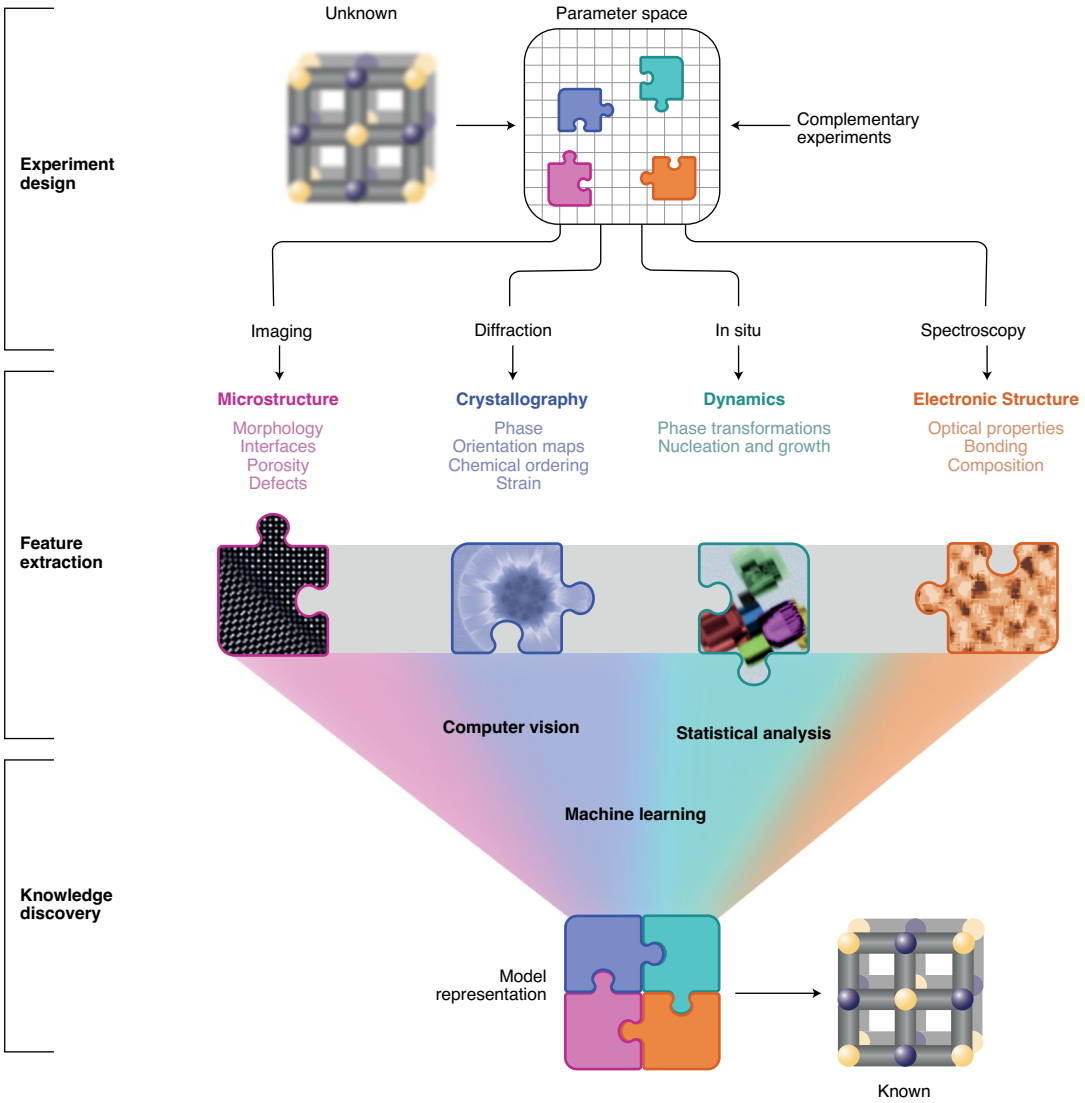


Fig. 1 | The electron microscopy framework. The framework for translating unknown structures or processes into quantifiable, physically meaningful descriptors and model representations.

greatly improved electron energy-loss spectroscopy (EELS) capabilities[13].

However, to harness these developments to solve evolving materials challenges, experimentation must become far more data-driven, integrated and automated, as noted in recent agency reports[14]. The current experiment design process is heavily biased towards techniques and features already familiar to the human operator. Microscopists are prone to rote analyses based on their prior experience; while this allows them to rapidly triage complex, novel scenarios, it can also blind them to more optimal approaches that may be outside their expertise. To adequately bound the parameters of an experiment (represented by the grid at the top of Fig. 1), we must carefully consider the full array of tools at our disposal.

Here, simulations and data science can help intelligently plan experiments by quantifying the strengths and weaknesses of each technique before the first sample is measured. Newly developed high-throughput graphics processing unit (GPU)-accelerated simulations[15] can estimate detection limits of various imaging methods (for example, high-angle annular dark field, annular bright field, and so on) and their ability to detect low-contrast single-atom defects of the kind found in diamond qubit materials, for instance[16]. The speed, cost and efficacy of these methods may then be compared against spectroscopic approaches such as EELS, considering the effect of beam parameters, sample characteristics and ionization edges of interest. Performing a quick simulation before labour-intensive experimentation can yield both tremendous cost and time savings.

The overarching goal of successful experiment design is to build a pipeline to translate microscope- and experiment-specific data (starting from raw data streams from detectors and cameras) into materials-specific descriptors and functionalities. As the number of imaging and sample parameters grows, it becomes increasingly difficult for a human operator alone to select the best combination of techniques. AI and ML methods, which

can efficiently evaluate behaviours over higher-dimensional parameter spaces, are well suited to this kind of predictive costing analysis[17]. Prior to undertaking an experiment, ML could be used to mine open databases of past work, harvesting appropriate imaging techniques and experimental parameters from related systems. These parameters could then be compared to the specifics of the system under study, validated against simulations, and presented to the user in real time to estimate what descriptors could be confidently measured. At present, no widely used database of prior work exists and such a highly integrated level of planning is simply not possible, leading to failure-prone or information-poor experimentation. The proposed approach leverages the strength of AI to very quickly operate with large volumes of data, augmenting the intrinsic depth and domain expertise of the human operator. Ongoing active research in human–computer interaction will continue to define best practices in this area. Beneficially, this approach will unlock the full range of analytical modes available on modern instrumentation for many more users of all experience levels.

Feature extraction

After the experimental parameter space has been defined and techniques have been selected to probe those parameters, we consider the process of feature extraction, shown in the middle of Fig. 1. Each technique provides a window into one or more features of the sample, convoluted with artefacts introduced during the recording process. For example, atomic-scale STEM imaging and spectroscopy create a two-dimensional projection of a three-dimensional crystal structure, but beam broadening and channelling can degrade resulting data fidelity, thereby complicating inverse structure determination. In addition, the beam itself may change (damage) the sample, or the instrument alignment may drift, effectively introducing noise that obscures the original object. For this reason, a combination of several complementary analysis techniques is usually required to arrive at more unique solutions by probing different characteristics of the underlying sample.

Unfortunately, data collection is presently highly disconnected and prescriptive. We choose imaging modes and detectors based on the features we expect to find and then acquire data in a linear and serial fashion, overlooking higher-dimensional or low-contrast correlations by neglecting to use all available data streams. We contrast this with the emerging data-rich 4D-STEM technique[18], in which entire diffraction patterns are collected across the two-dimensional space of a sample and then post-processed to generate particular contrast modes and signals. In effect, nearly all the transmitted beams from a sample are recorded, which can be used to reconstruct multiple signals on the fly or after the fact. This capability improves our ability to detect features that may be weakly represented in any one isolated dataset. Access to complete data streams is essential, especially during initial acquisition when the experimental parameters can still be adjusted.

Feature extraction is also increasingly constrained by the manual nature of traditional experimentation. To their credit, vendors have improved automation of microscope alignment, now offering software that can optimize the instrument faster and more accurately than most human operators. However, flexible and truly automated data collection integrating a full suite of modalities is far from being realized. At present, most investigations of unknown samples are similar: we manually scan many regions, searching for predetermined features of interest or deviations from known structures. One can envision batch experimentation, where the stage movement, alignment, focusing and image capture allow for the rapid surveying of a sample overnight. Key regions could then be highlighted without human bias using a ML network trained on the sum of prior knowledge and presented to the operator in the morning for further examination. This kind of pipeline has been established to a degree in semiconductor failure analysis and single-particle imaging, where repeated sample configurations lend themselves well to automation. However, to extend this approach to other domains, such as metallurgy and catalysis, we must have direct access to open, low-level microscope controls (for example, at a minimum stage position, tilt and defocus) and analysis routines to define flexible characterization workflows. In turn, this ability will help enrich the efficiency and quality of the entire characterization process.

Probing the temporal evolution of materials requires additional considerations, such as the need to precisely correlate observations with experimental parameters that may themselves be difficult to accurately measure. Understanding many important phenomena, including electrochemical cycling of batteries and the nature of low-temperature electronic phase transitions, requires a high degree of control over multiple experimental sub-systems. At present, instruments explore a limited in situ parameter space and there is little interoperability between platforms supplied by different vendors. In situ experiments can be roughly divided into two main groups: those where an experiment is built into a special holder[19] (for example, liquid/gas stage, biasing/heating stage, mechanical strain stage and so on) and those where the microscope itself is directly modified to create a desired environment (for example, gas, irradiation, deposition and so on) using a differential pumping mechanism[20] or other means. The microscope pole piece gap is a limiting factor that determines the kind of experiments that can be performed in the materials science toolbox[21]. A wide variety of holders and experiment types exist that all require different sample configurations, so tightly integrated, cross-correlative work is challenging and more hardware co-development is needed. Beyond holders, the gains made in detectors have allowed for in situ data to be accessed and integrated in a more efficient fashion, even at speeds as high as 4,000 frames per second[9,13]. Much can be learned from the astronomy community[22] as we continue to advance this aspect of microscopy hardware.

As a promising alternative, a modular system in which in situ capabilities are built into the objective lens pole piece could be built to accommodate interchangeable 'lab-in-the-gap' modules[23]. Similar concepts have been successfully developed in the past, but only for a limited range of options (for example, environmental cell experiments[24] or magnetic imaging using a low-magnetic-field pole piece[25]). The advantage of this approach is that the whole pole piece gap can be utilized, allowing for more complex in situ capabilities, but also requiring more intricate engineering for the necessary gas, liquid, electrical and other connections. Such an instrument would permit high-throughput TEM for a variety of experiments, for example in the area of catalysis[26]. One could consider a microscope fitted with several pole piece modules in a carousel (analogous to an optical microscope with several objective lenses) or a cartridge system, which would enable several experiments to be queued up ahead of time. Just as important, an open library of prior experiments and conditions should be developed alongside hardware to guide the planning and execution of in situ experiments, analogous to biological protein and emerging materials science data banks (examples include: the Protein Databank (https://www.rcsb.org/), the Materials Data Bank (http://www.materialsdatabank.org/), the Materials Research Data Council (https://www.mardac.org/), the Open Quantum Materials Database (http://www.oqmd.org/)

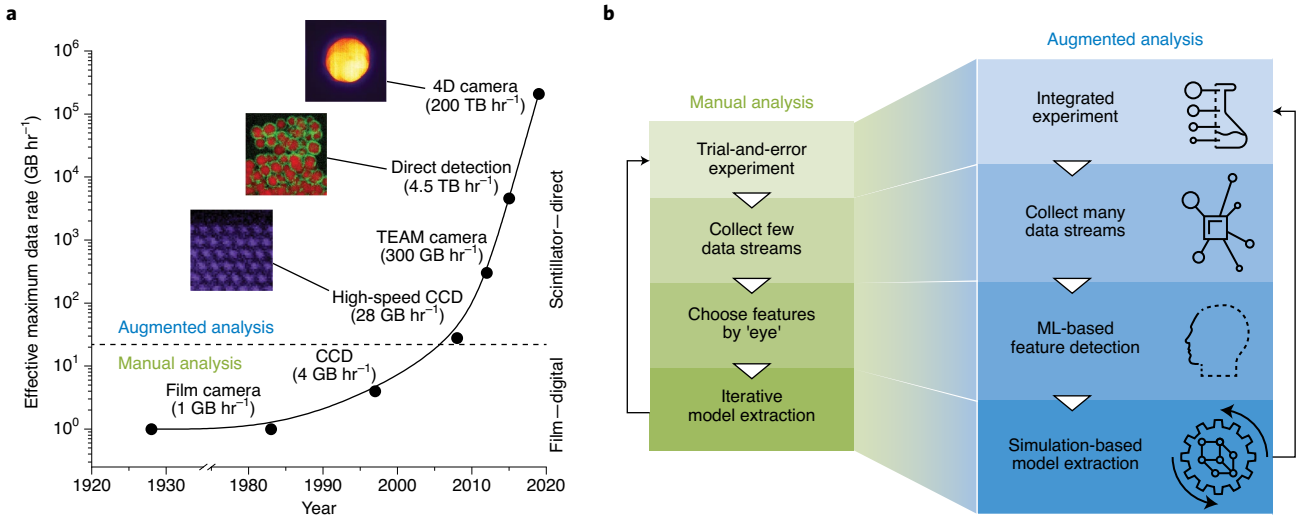


Fig. 2 | Microscopy data-production rates and analysis workflows. a, Effective maximum microscope data-production rates by year, showing the rapid increase associated with better detector technologies. Estimates are constrained by the overhead associated with processing and transfer of data. **b**, Present and emerging microscopy analysis workflows harnessing new methods of data collection and interpretation.

and Citrination (https://www.citrination.com/)). Once such a framework is broadly established, AI could optimize experimental conditions and acquisition parameters on the fly. With standardization of methods and analysis, multiple laboratories around the world could contribute to ambitious 'crowdsourced' experimentation to more quickly and effectively tackle problems, such as combinatorial materials screening. These developments would help realize unprecedented experiments to target a wide range of impactful questions, including:

1. What are the far-from-equilibrium states and rate parameters of materials during fast and non-repeatable phase transformations?
2. What is the nature of the soft–hard (liquid–solid) interfaces present in chemical and biological reactions?
3. What are the chemical states and bonding structure of materials during reactions, especially considering complex reaction dynamics and radiation chemistry?
4. How can we reliably characterize beam-sensitive materials?
5. What behaviour do light atoms, vacancies and point defects exhibit in extreme, reactive environments where they are hard to visualize?

Knowledge discovery

Throughout the characterization process our goal is to identify statistically significant features in large, noisy and potentially incomplete data streams, aiming to build libraries of possible structures and spectra to aid in knowledge discovery[27]. All domains of electron microscopy are producing ever-expanding amounts of data spanning a range of formats that must be appropriately distilled through the interpretive frameworks shown in the lower half of Fig. 1. While, in principle, more data is a positive development, our ability to process and extract meaning from ballooning data sets has not kept pace. As shown in Fig. 2a, epochs in data production have been punctuated by advances in detector technology. Following the initial development of the microscope, data volumes remained relatively flat until the advent of digital imaging (because of the rate limiting time and cost of film processing), after which they experienced rapid growth during the transition from slow scintillators to fast direct detector technologies. Computing power improved at a similar rate and data production is now several orders of magnitude higher than it was a few decades ago. Technological upgrades are becoming more frequent and disruptive, motivating an urgent need for new analysis methods.

As already mentioned, growing data volumes are well suited to interpretation by AI and ML methods trained on established physical models. When grounded in physically meaningful frameworks, these approaches can apply constraints to the classification of multidimensional features, using domain knowledge from materials science, chemistry, physics and biology. Recently introduced deep learning methods have demonstrated extreme efficiency in such feature finding problems[27]. However, a basic problem is that data are typically encoded in limited proprietary formats and there is no good way to assess the amount of useful information obtained in an experiment. We currently lack metrics for data quality and guidelines to determine whether present measurements are even comparable to past work, leading to the manual analysis workflow shown in Fig. 2b. While this workflow has yielded important scientific discoveries, it is limited in its scalability and its ability to efficiently incorporate a wide variety of multidimensional data streams. At the heart of this issue is poor cross-platform support and dialogue between the different microscope, camera, holder and other hardware vendors in terms of signal channels, naming conventions, file formats and metadata. The absence of interoperability severely hampers experiment repeatability, portability and user training. Without a common language for experimental electron microscopy, we cannot properly curate data acquisition and analysis processes in order to ensure scientific integrity of our experiments. This situation also makes it extremely difficult to integrate electron microscopy data with other techniques (for example, scattering, mechanical testing, transport and so on) that would help ML algorithms arrive at more unique solutions for a structure or process. We emphasize that data science can only augment human intuition and domain expertise, but not replace it. Rather,

we must strive to achieve synergy between conventional and data-driven methods, seeking to harness the unique strengths of each analysis approach for the problem at hand.

While the challenges for data interpretation are great, there has been some progress[17]. New software toolsets[28] allow researchers to store and share their analysis workflows in various forms. Following trends in the data science community, Jupyter Notebooks and their Google Colab implementations have become more widespread and mainstream, enabling the dissemination of ML code and trained neural networks[27]. (Commercial equipment, instruments, software or materials are identified only in order to adequately specify certain procedures. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products identified are necessarily the best available for the purpose.) Still, more development of open-source platforms using the FAIR—findable, accessible, interoperable and reusable—principles[29] is needed to standardize best practices, as well as streamline the training of early-career researchers. A repository for data of all formats would also help address the crisis of experimental reproducibility by unlocking a whole class of meta-analyses, which are almost non-existent in microscopy, but routine in fields such as astronomy, high-energy physics, scattering, thermodynamics, genomics and medicine. The community has also recognized the need for greater convergence of microscopists, data scientists and manufacturers to implement the proposed architecture. We believe the time is right for new national initiatives to catalyse adoption of a common experimental language, establish standards and shared methodologies, and provide the infrastructure for data-driven partnerships between hardware and software developers. Inspiration can be drawn from light and neutron sources, which showcase the power of highly integrated, interoperable and open experimentation. The benefits to microscopy and the broader scientific community will be enormous.

Looking to the future

Today's microscopes are capable of producing so much data that it can no longer be effectively analysed by human intuition and experience alone. Next-generation microscopy will require entirely new ways of thinking about experiment design, execution, analysis and sharing. Data science tools must become more tightly integrated into the operation of the instrument, helping to distil vast multidimensional datasets into meaningful descriptors linked to underlying physical models. An open platform for data collection and analysis will intelligently highlight latent features and help extract deep insight from complex, multifaceted observations. Importantly, this platform must continue to evolve to meet domain needs and keep pace with instrumentation developments.

Truly adaptive microscopy, where data dynamically inform the next steps of an experiment on the fly has not yet been realized. In such a microscope, for example, tracking a reaction in a liquid cell would be done by comparing multiple, automatically selected signals quantitatively interpreted through fast simulations based on theory models. A ML network would control the stage and imaging parameters to best highlight features of interest, providing guidance at each stage of the experiment. Data capture, storage and distribution would all be routed through an open framework accessible to the broad community. Data and metadata could be compared in real-time to large databases of similar experiments to predict possible next steps, augmenting human intuition and experience. This stage in the analysis process would then iteratively inform experiment design to build a virtuous augmented workflow, as shown in Fig. 2b. The outcome of such an experiment would be richly quantifiable, repeatable and meaningful at a level far beyond our present capabilities. Bold national initiatives and visionary leadership are strongly needed to realize this future. Collectively, these efforts will enable the ground-breaking discoveries required to solve the pressing global challenges of the next decade. ❐

Steven R. Spurgeon[1]✉, Colin Ophus[2], Lewys Jones[3,4], Amanda Petford-Long[5], Sergei V. Kalinin[6], Matthew J. Olszta[1], Rafal E. Dunin-Borkowski[7], Norman Salmon[8], Khalid Hattar[9], Wei-Chang D. Yang[10], Renu Sharma[10], Yingge Du[11], Ann Chiaramonti[12], Haimei Zheng[13], Edgar C. Buck[1], Libor Kovarik[11], R. Lee Penn[14], Dongsheng Li[11], Xin Zhang[11], Mitsuhiro Murayama[15] and Mitra L. Taheri[16]✉

[1]Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA, USA. [2]National Center for Electron Microscopy, Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [3]Advanced Microscopy Laboratory, Centre for Research on Adaptive Nanostructures and Nanodevices (CRANN), Dublin, Ireland. [4]School of Physics, Trinity College Dublin, The University of Dublin, Dublin, Ireland. [5]Materials Science Division, Argonne National Laboratory, Argonne, IL, USA. [6]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN, USA. [7]Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons and Peter Grünberg Institute, Forschungszentrum Jülich, Jülich, Germany. [8]Hummingbird Scientific, Lacey, WA, USA. [9]Center for Integrated Nanotechnologies, Sandia National Laboratories, Albuquerque, NM, USA. [10]Physical Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA. [11]Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland, WA, USA. [12]Applied Chemicals and Materials Division, Material Measurement Laboratory, National Institute of Standards and Technology, Boulder, CO, USA. [13]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [14]Department of Chemistry, University of Minnesota, Minneapolis, MN, USA. [15]Department of Materials Science and Engineering, Virginia Tech, Blacksburg, VA, USA. [16]Department of Materials Science and Engineering, Johns Hopkins University, Baltimore, MD, USA.

✉*e-mail: steven.spurgeon@pnnl.gov; mtaheri4@jhu.edu*

Published online: 26 October 2020
https://doi.org/10.1038/s41563-020-00833-z

References

1. Ruska, E. *Rev. Mod. Phys.* **59**, 627–638 (1987).
2. Shen, P. S. *Anal. Bioanal. Chem.* **410**, 2053–2057 (2018).
3. Shechtman, D., Blech, I., Gratias, D. & Cahn, J. W. *Phys. Rev. Lett.* **53**, 1951–1953 (1984).
4. Varela, M. et al. *Annu. Rev. Mater. Res.* **35**, 539–569 (2005).
5. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. *Nature* **559**, 547–555 (2018).
6. Bruno, I. et al. *Data Sci. J.* **16**, 38 (2017).
7. Baldwin, P. R. et al. *Curr. Opinion Microbiol.* **43**, 1–8 (2018).
8. Minor, A. M., Denes, P. & Muller, D. A. *MRS Bull.* **44**, 961–966 (2019).
9. Tate, M. W. et al. *Microscopy Microanal.* **22**, 237–249 (2016).
10. Zhou, J. et al. *Nature* **570**, 500–503 (2019).
11. Jiang, Y. et al. *Nature* **559**, 343–349 (2018).
12. Booth, C. *Microscopy Microanal.* **18**, 78–79 (2012).
13. Hart, J. L. et al. *Sci. Rep.* **7**, 8243 (2017).
14. BES workshop reports (DOE, accessed 22 June 2020); https://science.osti.gov/bes/Community-Resources/Reports
15. Ophus, C. *Adv. Struct. Chem. Imaging* **3**, 13 (2017).
16. Dolde, F. et al. *Nat. Phys.* **7**, 459–463 (2011).
17. Voyles, P. M. *Curr. Opinion Solid State Mater. Sci.* **21**, 141–158 (2017).
18. Ophus, C. *Microscopy Microanal.* **25**, 563–582 (2019).
19. Daulton, T. L., Little, B. J., Lowe, K. & Jones-Meehan, J. *Microscopy Microanal.* **7**, 470–485 (2001).
20. Sharma, R. & Crozier, P. A. In *Handbook of Microscopy for Nanotechnology* (eds Yao, N. & Wang, Z. L.) 531–565 (Kluwer Academic Publishers, 2005).
21. Robertson, I. M. et al. *J. Mater. Res.* **26**, 1341–1383 (2011).
22. York, D. G. et al. *Astron. J.* **120**, 1579–1587 (2000).
23. Borrnert, F. et al. *Microscopy Microanal.* **21**, 99–100 (2015).
24. Boyes, E. & Gai, P. *Ultramicroscopy* **67**, 219–232 (1997).
25. Shibata, N. et al. *Nat. Commun.* **10**, 2308 (2019).
26. Tao, F. F. & Crozier, P. A. *Chem. Rev.* **116**, 3487–3539 (2016).
27. Ziatdinov, M. et al. *Sci. Adv.* **5**, eaaw8989 (2019).
28. Somnath, S., Smith, C. R., Laanait, N., Vasudevan, R. K. & Jesse, S. *Microscopy Microanal.* **25**, 220–221 (2019).
29. Wilkinson, M. D. et al. *Sci. Data* **3**, 160018 (2016).

Acknowledgements

This commentary is the result of discussions from the first in a series of Next-Generation Transmission Electron Microscopy (NexTEM) workshops, held at Pacific Northwest National Laboratory in October 2018. S.R.S.

thanks A. Lang, B. Matthews and J. Hart for reviewing the manuscript. This work was supported by the Laboratory Directed Research and Development (LDRD) Nuclear Processing Science Initiative (NPSI) at Pacific Northwest National Laboratory (PNNL). PNNL is a multi-programme national laboratory operated for the US Department of Energy (DOE) by Battelle Memorial Institute under contract no. DE-AC05-76RL0-1830. This work was supported in part by the Office of Science, Office of Basic Energy Sciences, of the US DOE under contracts no. DE-AC02-05CH11231 (C.O.), no. 10122 (S.R.S. and Y.D.), no. KC0201010 ERKCS89 (S.V.K.), no. KC0203020:67037 (D.L.) and no. DE-AC02-05-CH11231 within the KC22ZH programme (H.Z.). This work was supported in part by the US DOE, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division (A.P.-L.). M.M. acknowledges the Virginia Tech National Center for Earth and Environmental Nanotechnology Infrastructure (NanoEarth), a member of the National Nanotechnology Coordinated Infrastructure (NNCI), supported by NSF (ECCS 1542100). This project has received funding from the European Research Council under the Horizon 2020 Research and Innovation Programme (grant no. 856538, project 3D MAGiC; and grant no. 823717, project ESTEEM3 (R.E.D.-B.)). X.Z. acknowledges support from the DOE BES Geosciences Program at PNNL (FWP 56674). The work was partly performed at the Center for Nanophase Materials Sciences (S.V.K.) and the Center for Integrated Nanotechnologies (K.H.), which are Office of Science User Facilities operated for the US DOE. Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the US DOE under contract no. DE-AC02-05CH11231. A portion of the microscopy shown was performed at the Environmental Molecular Sciences Laboratory (EMSL), a national scientific user facility sponsored by the DOE's Office of Biological and Environmental Research and located at PNNL. L.J. acknowledges SFI grants AMBER2-12/RC/2278-P2 and URF/RI/191637. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the US DOE's National Nuclear Security Administration under contract DE-NA-0003525. The views expressed in the article do not necessarily represent the views of the US DOE or the US government.

Competing interests

The authors declare no competing interests.