

## **Towards Data-Driven Next-Generation Transmission Electron Microscopy**

Steven R. Spurgeon,<sup>1</sup> Colin Ophus,<sup>2</sup> Lewys Jones,<sup>3,4</sup> Amanda Petford-Long,<sup>5</sup> Sergei V. Kalinin,<sup>6</sup> Matthew J. Olszta,<sup>1</sup> Rafal E. Dunin-Borkowski,<sup>7</sup> Norman Salmon,<sup>8</sup> Khalid Hattar,<sup>9</sup> Wei-Chang D. Yang,<sup>10</sup> Renu Sharma,<sup>10</sup> Yingge Du,<sup>11</sup> Ann Chiaramonti,<sup>12</sup> Haimei Zheng,<sup>13</sup> Edgar C. Buck,<sup>1</sup> Libor Kovarik,<sup>14</sup> R. Lee Penn,<sup>15</sup> Dongsheng Li,<sup>11</sup> Xin Zhang,<sup>11</sup> Mitsuhiro Murayama,<sup>16</sup> and Mitra L. Taheri<sup>17</sup>

<sup>1</sup>*Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, Washington, United States<sup>a)</sup>*

<sup>2</sup>*National Center for Electron Microscopy, Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California, United States*

<sup>3</sup>*Advanced Microscopy Laboratory, Centre for Research on Adaptive Nanostructures and Nanodevices (CRANN), Dublin 2, Ireland*

<sup>4</sup>*School of Physics, Trinity College Dublin, The University of Dublin, Dublin 2, Ireland*

<sup>5</sup>*Materials Science Division, Argonne National Laboratory, Argonne, Illinois, United States*

<sup>6</sup>*Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States*

<sup>7</sup>*Ernst Ruska-Centre for Microscopy and Spectroscopy with Electrons and Peter Grünberg Institute, Forschungszentrum Jülich, 52425 Jülich, Germany*

<sup>8</sup>*Hummingbird Scientific, Lacey, Washington, United States*

<sup>9</sup>*Center for Integrated Nanotechnologies, Sandia National Laboratories, Albuquerque, New Mexico, United States*

<sup>10</sup>*Physical Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland, United States*

<sup>11</sup>*Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington, United States*

<sup>12</sup>*Applied Chemicals and Materials Division, Material Measurement Laboratory, National Institute of Standards and Technology, Boulder, Colorado,*

*United States*

<sup>13)</sup> *Materials Sciences Division, Lawrence Berkeley National Lab, Berkeley, California, United States*

<sup>14)</sup> *Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington, United States*

<sup>15)</sup> *Department of Chemistry, University of Minnesota, Minneapolis, Minnesota, United States*

<sup>16)</sup> *Department of Materials Science and Engineering, Virginia Tech, Blacksburg, Virginia, United States*

<sup>17)</sup> *Department of Materials Science and Engineering, Johns Hopkins University, Baltimore, Maryland, United States<sup>b)</sup>*

(Dated: 29 October 2020)

Electron microscopy touches upon 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.

---

<sup>a)</sup>Electronic mail: [steven.spurgeon@pnnl.gov](mailto:steven.spurgeon@pnnl.gov)

<sup>b)</sup>Electronic mail: [mtaheri4@jhu.edu](mailto:mtaheri4@jhu.edu)

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.<sup>1</sup> TEM provides rich, directly-resolved information about the structure and dynamics of phenomena spanning atoms to microns that are of great fundamental and practical significance to society. It has played a key role in protein and drug discovery,<sup>2</sup> redefined our understanding of crystalline solids,<sup>3</sup> and catalyzed the electronics revolution that gave rise to today’s massively interconnected world.<sup>4</sup>

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,<sup>5</sup> 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.<sup>6</sup> 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<sup>7</sup> to transform our un-

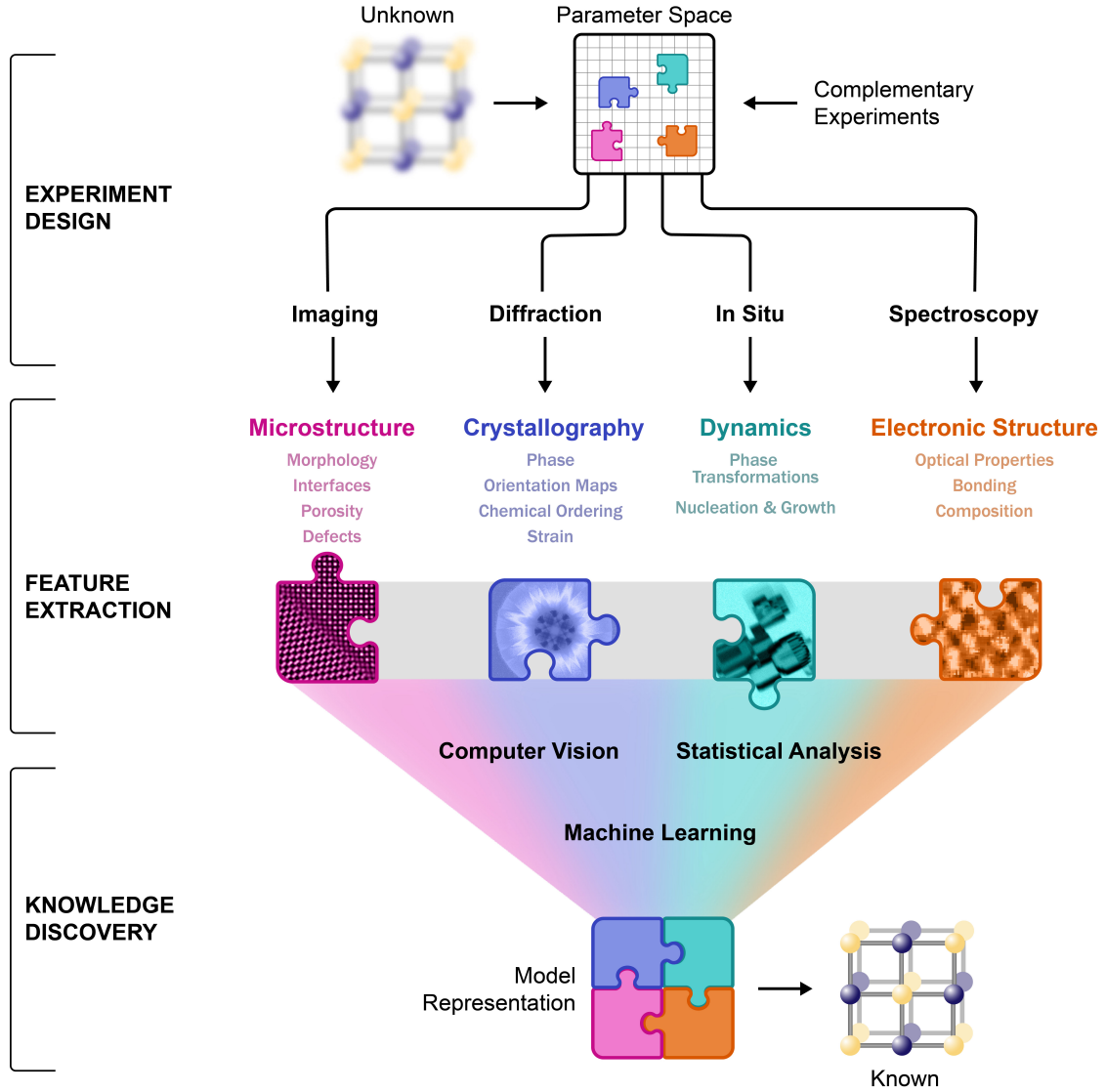


FIG. 1. The electron microscopy framework for translating unknown structures or processes into quantifiable, physically meaningful descriptors and model representations.

derstanding of metals, semiconductors, ceramics, and more.

Sweeping changes precipitated by recent technological innovations and the growth of modern data science tools call for a reexamination of the electron microscopy framework, shown in Figure 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 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, shown in the top of Figure 1. This stage is by its very nature based on preexisting 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,<sup>8</sup> data-rich high-speed detectors,<sup>9</sup> and atomic-scale electron tomography<sup>10</sup> have provided a wealth of new imaging modalities waiting to be exploited. For example, hybrid pixel detectors<sup>9</sup> 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.<sup>11</sup> An alternative detector technology, back-thinned monolithic active pixel sensors that can count individual incident electrons,<sup>12</sup> has greatly improved electron energy loss spectroscopy (EELS) capabilities.<sup>13</sup>

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.<sup>14</sup> The current experiment design process is heavily biased toward 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 Figure 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<sup>15</sup> can estimate detection limits of various imaging methods (e.g. high-angle annular dark field, annular bright field, etc.) and their ability to detect low-contrast single-atom defects of the kind found in diamond qubit materials, for instance.<sup>16</sup> 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 labor-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 behaviors over higher-dimensional parameter spaces, are well suited to this kind of predictive costing analysis.<sup>17</sup> 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 Figure 1. Each technique provides a window into one or more features of the sample, convoluted with artifacts 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 channeling 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 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,<sup>18</sup> in which entire diffraction patterns are collected across the 2D 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 data set. 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 sur-

veying 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 (e.g. 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<sup>19</sup> (e.g. liquid/gas stage, biasing/heating stage, mechanical strain stage, etc.) and those where the microscope itself is directly modified to create a desired environment (e.g. gas, irradiation, deposition, etc.) using a differential pumping mechanism<sup>20</sup> 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.<sup>21</sup> 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.<sup>9,13</sup> Much can be learned from the astronomy community<sup>22</sup> 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.<sup>23</sup> Similar concepts have been successfully developed in the past, but only for a limited range of options (e.g. environmental cell experiments<sup>24</sup> or magnetic imaging using a



low magnetic field pole piece<sup>25</sup>). 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.<sup>26</sup> 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.<sup>27</sup> 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 behavior 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

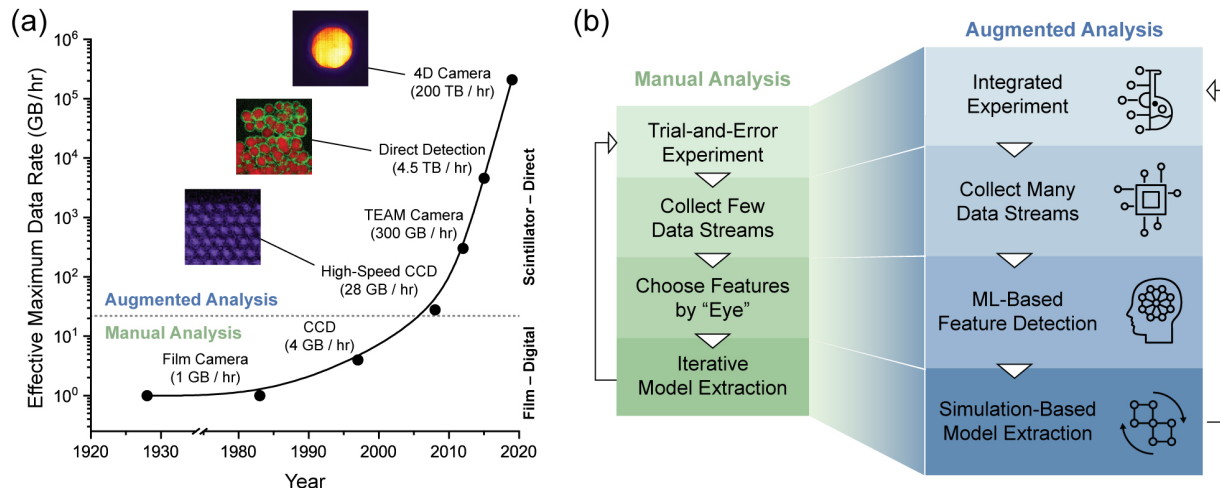


FIG. 2. (a) Effective maximum microscope data production rates by year, showing the rapid increase associated with better detector technologies. Estimates are constrained by the overhead associated with processing and transfer of data. (b) Present and emerging microscopy analysis workflows harnessing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

of possible structures and spectra to aid in knowledge discovery.<sup>28</sup> 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 bottom half of Figure 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 Figure 2.(a), 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 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.<sup>28</sup> 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 Figure 2.(b). 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 (e.g. scattering, mechanical testing, transport, etc.) 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.<sup>17</sup> New software toolsets<sup>29</sup> allow researchers to store and share their analysis workflows in various forms. Following trends in the data science community, Jupyter notebooks and their Google Collab implementations have become more widespread and mainstream, enabling the dissemination of ML code and trained neural networks.<sup>2830</sup> Still, more development of open-source platforms using the FAIR—Findable-Accessible-Interoperable-Reusable—principles<sup>31</sup> 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 nonexistent 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 im-

plement the proposed architecture. We believe the time is right for new national initiatives to catalyze 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 analyzed 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 distill vast multidimensional data sets 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 then 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 Figure 2.(b). 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 groundbreaking discoveries required to solve the

pressing global challenges of the next decade.

## 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 Drs. Andrew Lang, Bethany Matthews, and James 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 multiprogram national laboratory operated for the U.S. 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 U.S. Department of Energy under Contracts No. DE-AC02-05CH11231 (C.O.), No. 10122 (S.R.S. and Y.D.), and No. DE-AC02-05-CH11231 within the KC22ZH program (H.Z.). This work was supported in part by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division (A.P-L.). 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 U.S. DOE. Work at the Molecular Foundry was supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy 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 Department of Energy’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 multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. 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 U.S. DOE or the United States Government.

## COMPETING INTERESTS

The authors declare no competing interests.

## REFERENCES

- <sup>1</sup>E. Ruska, “The development of the electron microscope and of electron microscopy,” *Reviews of Modern Physics* **59**, 627–638 (1987).
- <sup>2</sup>P. S. Shen, “The 2017 Nobel Prize in Chemistry: cryo-EM comes of age,” *Analytical and Bioanalytical Chemistry* **410**, 2053–2057 (2018).
- <sup>3</sup>D. Shechtman, I. Blech, D. Gratias, and J. W. Cahn, “Metallic Phase with Long-Range Orientational Order and No Translational Symmetry,” *Physical Review Letters* **53**, 1951–1953 (1984).
- <sup>4</sup>M. Varela, A. Lupini, K. V. Benthem, A. Borisevich, M. Chisholm, N. Shibata, E. Abe, and S. Pennycook, “Materials Characterization in the Aberration-Corrected Scanning Transmission Electron Microscope,” *Annual Review of Materials Research* **35**, 539–569 (2005).
- <sup>5</sup>K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh, “Machine learning for molecular and materials science,” *Nature* **559**, 547–555 (2018).
- <sup>6</sup>I. Bruno, S. Gražulis, J. R. Helliwell, S. N. Kabekkodu, B. McMahon, and J. Westbrook, “Crystallography and databases,” *Data Science Journal* **16**, 1–17 (2017).
- <sup>7</sup>P. R. Baldwin, Y. Z. Tan, E. T. Eng, W. J. Rice, A. J. Noble, C. J. Negro, M. A. Cianfrocco, C. S. Potter, and B. Carragher, “Big data in cryoEM: automated collection, processing and accessibility of EM data,” *Current Opinion in Microbiology* **43**, 1–8 (2018).
- <sup>8</sup>A. M. Minor, P. Denes, and D. A. Muller, “Cryogenic electron microscopy for quantum science,” *MRS Bulletin* **44**, 961–966 (2019).
- <sup>9</sup>M. W. Tate, P. Purohit, D. Chamberlain, K. X. Nguyen, R. Hovden, C. S. Chang, P. Deb, E. Turgut, J. T. Heron, D. G. Schlom, D. C. Ralph, G. D. Fuchs, K. S. Shanks, H. T. Philipp, D. A. Muller, and S. M. Gruner, “High Dynamic Range Pixel Array Detector for Scanning Transmission Electron Microscopy,” *Microscopy and Microanalysis* **22**, 237–249 (2016).
- <sup>10</sup>J. Zhou, Y. Yang, Y. Yang, D. S. Kim, A. Yuan, X. Tian, C. Ophus, F. Sun, A. K. Schmid, M. Nathanson, H. Heinz, Q. An, H. Zeng, P. Ercius, and J. Miao, “Observing crystal nucleation in four dimensions using atomic electron tomography,” *Nature* **570**, 500–503 (2019).
- <sup>11</sup>Y. Jiang, Z. Chen, Y. Han, P. Deb, H. Gao, S. Xie, P. Purohit, M. W. Tate, J. Park, S. M. Gruner, V. Elser, and D. A. Muller, “Electron ptychography of 2D materials to

- deep sub-ångström resolution,” *Nature* **559**, 343–349 (2018).
- <sup>12</sup>C. Booth, “K2: A Super-Resolution Electron Counting Direct Detection Camera for Cryo-EM,” *Microscopy and Microanalysis* **18**, 78–79 (2012).
- <sup>13</sup>J. L. Hart, A. C. Lang, A. C. Leff, P. Longo, C. Trevor, R. D. Twisten, and M. L. Taheri, “Direct Detection Electron Energy-Loss Spectroscopy: A Method to Push the Limits of Resolution and Sensitivity,” *Scientific Reports* **7**, 8243 (2017).
- <sup>14</sup>DOE BES Workshop Reports Website, <https://science.osti.gov/bes/Community-Resources/Reports>, [Accessed: June 22, 2020].
- <sup>15</sup>C. Ophus, “A fast image simulation algorithm for scanning transmission electron microscopy,” *Advanced Structural and Chemical Imaging* **3**, 13 (2017), arXiv:1702.01904.
- <sup>16</sup>F. Dolde, H. Fedder, M. W. Doherty, T. Nöbauer, F. Rempp, G. Balasubramanian, T. Wolf, F. Reinhard, L. C. L. Hollenberg, F. Jelezko, and J. Wrachtrup, “Electric-field sensing using single diamond spins,” *Nature Physics* **7**, 459–463 (2011).
- <sup>17</sup>P. M. Voyles, “Informatics and data science in materials microscopy,” *Current Opinion in Solid State and Materials Science* **21**, 141–158 (2017).
- <sup>18</sup>C. Ophus, “Four-Dimensional Scanning Transmission Electron Microscopy (4D-STEM): From Scanning Nanodiffraction to Ptychography and Beyond,” *Microscopy and Microanalysis* **25**, 563–582 (2019).
- <sup>19</sup>T. L. Daulton, B. J. Little, K. Lowe, and J. Jones-Meehan, “In Situ Environmental Cell-Transmission Electron Microscopy Study of Microbial Reduction of Chromium(VI) Using Electron Energy Loss Spectroscopy,” *Microscopy and Microanalysis* **7**, 470–485 (2001).
- <sup>20</sup>R. Sharma and P. A. Crozier, “Environmental Transmission Electron Microscopy in Nanotechnology,” in *Handbook of Microscopy for Nanotechnology* (Kluwer Academic Publishers, Boston) pp. 531–565.
- <sup>21</sup>I. M. Robertson, C. A. Schuh, J. S. Vetrano, N. D. Browning, D. P. Field, D. J. Jensen, M. K. Miller, I. Baker, D. C. Dunand, R. Dunin-Borkowski, B. Kabius, T. Kelly, S. Lozano-Perez, A. Misra, G. S. Rohrer, A. D. Rollett, M. L. Taheri, G. B. Thompson, M. Uchic, X.-L. Wang, and G. Was, “Towards an integrated materials characterization toolbox,” *Journal of Materials Research* **26**, 1341–1383 (2011).
- <sup>22</sup>D. G. York, J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, R. Barkhouser, S. Bastian, E. Berman, W. N. Boroski, S. Bracker, C. Briegel, J. W. Briggs, J. Brinkmann, R. Brunner, S. Burles, L. Carey, M. A. Carr, F. J. Castander,



- B. Chen, P. L. Colestock, A. J. Connolly, J. H. Crocker, I. Csabai, P. C. Czarapata, J. E. Davis, M. Doi, T. Dombeck, D. Eisenstein, N. Ellman, B. R. Elms, M. L. Evans, X. Fan, G. R. Federwitz, L. Fiscelli, S. Friedman, J. A. Frieman, M. Fukugita, B. Gillespie, J. E. Gunn, V. K. Gurbani, E. de Haas, M. Haldeman, F. H. Harris, J. Hayes, T. M. Heckman, G. S. Hennessy, R. B. Hindsley, S. Holm, D. J. Holmgren, C.-h. Huang, C. Hull, D. Husby, S.-I. Ichikawa, T. Ichikawa, Ž. Ivezić, S. Kent, R. S. J. Kim, E. Kinney, M. Klaene, A. N. Kleinman, S. Kleinman, G. R. Knapp, J. Korienek, R. G. Kron, P. Z. Kunszt, D. Q. Lamb, B. Lee, R. F. Leger, S. Limmongkol, C. Lindenmeyer, D. C. Long, C. Loomis, J. Loveday, R. Lucinio, R. H. Lupton, B. MacKinnon, E. J. Mannery, P. M. Mantsch, B. Margon, P. McGehee, T. A. McKay, A. Meiksin, A. Merelli, D. G. Monet, J. A. Munn, V. K. Narayanan, T. Nash, E. Neilsen, R. Neswold, H. J. Newberg, R. C. Nichol, T. Nicinski, M. Nonino, N. Okada, S. Okamura, J. P. Ostriker, R. Owen, A. G. Pauls, J. Peoples, R. L. Peterson, D. Petravick, J. R. Pier, A. Pope, R. Pordes, A. Prosapio, R. Rechenmacher, T. R. Quinn, G. T. Richards, M. W. Richmond, C. H. Rivetta, C. M. Rockosi, K. Ruthmansdorfer, D. Sandford, D. J. Schlegel, D. P. Schneider, M. Sekiguchi, G. Sergey, K. Shimasaku, W. A. Siegmund, S. Smee, J. A. Smith, S. Snedden, R. Stone, C. Stoughton, M. A. Strauss, C. Stubbs, M. SubbaRao, A. S. Szalay, I. Szapudi, G. P. Szokoly, A. R. Thakar, C. Tremonti, D. L. Tucker, A. Uomoto, D. Vanden Berk, M. S. Vogeley, P. Waddell, S.-i. Wang, M. Watanabe, D. H. Weinberg, B. Yanny, and N. Yasuda, “The Sloan Digital Sky Survey: Technical Summary,” *The Astronomical Journal* **120**, 1579–1587 (2000).
- <sup>23</sup>F. Borrnert, H. Muller, M. Linck, A. Horst, A. I. Kirkland, B. Buchner, and H. Lichte, “Approaching the Lab in the Gap: First Results from a Versatile In-situ (S)TEM,” *Microscopy and Microanalysis* **21**, 99–100 (2015).
- <sup>24</sup>E. Boyes and P. Gai, “Environmental high resolution electron microscopy and applications to chemical science,” *Ultramicroscopy* **67**, 219–232 (1997).
- <sup>25</sup>N. Shibata, Y. Kohno, A. Nakamura, S. Morishita, T. Seki, A. Kumamoto, H. Sawada, T. Matsumoto, S. D. Findlay, and Y. Ikuhara, “Atomic resolution electron microscopy in a magnetic field free environment,” *Nature Communications* **10**, 2308 (2019).
- <sup>26</sup>F. F. Tao and P. A. Crozier, “Atomic-Scale Observations of Catalyst Structures under Reaction Conditions and during Catalysis,” *Chemical Reviews* **116**, 3487–3539 (2016).

- <sup>27</sup>Examples include: The Potein Databank (<https://www.rcsb.org/>), The Materials Data-bank (<http://www.materialsdatbank.org/>), MaRDaC (<https://www.mardac.org/>), OQMD (<https://www.oqmd.org/>), and Citrination (<https://www.citrination.com/>). [Accessed: June 22, 2020].
- <sup>28</sup>M. Ziatdinov, O. Dyck, X. Li, B. G. Sumpter, S. Jesse, R. K. Vasudevan, and S. V. Kalinin, “Building and exploring libraries of atomic defects in graphene: Scanning transmission electron and scanning tunneling microscopy study,” *Science Advances* **5**, eaaw8989 (2019), arXiv:1809.04256.
- <sup>29</sup>S. Somnath, C. R. Smith, N. Laanait, R. K. Vasudevan, and S. Jesse, “USID and Py-croscopy Open Source Frameworks for Storing and Analyzing Imaging and Spectroscopy Data,” *Microscopy and Microanalysis* **25**, 220–221 (2019), arXiv:1903.09515.
- <sup>30</sup>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.
- <sup>31</sup>M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. ’t Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, “The FAIR Guiding Principles for scientific data management and stewardship,” *Scientific Data* **3**, 160018 (2016).