

The Marriage of Electron Microscopy and Modern Data Science: Entering a New Era in the Study of Materials



Koji Kimoto

Doctor of Science
Director
Center for Basic Research on Materials
National Institute for Materials Science

1. Electron Microscopy and the Study of Materials

There are innumerable examples in which the macroscopic properties and performance of a material are governed by its microscopic structure; for this reason, electron microscopes, which offer various kinds of characterization techniques with high spatial resolution¹⁾, have made crucial contributions to the study of materials. To list just a few examples: the mechanical properties of structural materials have been elucidated by in-situ observations of crystal defects via ultra-high-voltage electron microscopy and similar techniques; advanced semiconductor devices have been characterized by localized cross-sectional observations using focused ion beam (FIB) and microsampling techniques; scanning electron microscopy (SEM) has become a critical tool for manufacturing semiconductor devices; electron holography has been used to visualize quantum-mechanical phenomena in superconductors and other exotic materials; cryo-electron microscopy (cryo-EM, for which the 2017 Nobel Prize in Chemistry was awarded) has been used to reveal the structure and function of proteins and biomolecules; and environmentally-controlled transport systems for anaerobic samples have become common tools for characterizing lithium secondary battery materials. A glance at the historical evolution of electron microscopy shows that the needs of advanced material developers have consistently driven advances in characterization techniques—and that cutting-edge microscopic techniques, just as consistently, have spurred progress in materials research and development. The history also clearly shows that Hitachi and Hitachi High-Tech have made seminal contributions to many of the most important technological breakthroughs, including those mentioned above: ultra-high-voltage electron microscopy, FIB techniques, microsampling, critical dimension (CD) SEM, field-emission electron guns, electron holography, and more.

Electron microscopes may be instruments for observing the *physically* small, but the results of these observations often reveal new insights that are *conceptually* enormous; the opportunity to see unexpected new microscopic structures or atomic arrangements *with one's own eyes* often provides scientists researching novel materials and engineers designing advanced devices with clues for a fundamental understanding of the materials. Discoveries of unexpected new materials, structures, or phenomena are often announced in journal publications trumpeting *a successful R&D program targeting just this finding* or *an innovative new conceptual proposal*; this phrasing may seem little more than a self-serving interpretation, but in fact such unexpected results are a typical form of serendipity, and advanced characterization techniques accelerate materials innovation.

2. The Explosion of Electron-microscopy Data—and the Need for Data Science

Up until the early 1990s, electron-microscopy images were exposed on photographic film; researchers would select 20 or so images at a time to develop in a darkroom. In the years that followed, improvements in the performance of system components (detectors, electron guns, spherical aberration correctors, electron spectrometers, monochromators, X-ray detectors, in-situ observation holders, and more), combined to ensure that the data acquired by electron microscopes became multi-modal—and vastly more voluminous. There were even attempts (such as RDE²⁾) to convert measurement

data into structured formats and save it for later use. Recent researchers need to control complicated scientific instruments, then sift through the resulting big data to extract meaningful information; as a result, basic skills in software engineering and data science have become essential prerequisites for *every* stage of materials characterization, from experimental planning to decision-making. Those of us who were educated in earlier eras will protest that we were taught to respect the hardware and revere the raw data above all. Accurate though this may be—and this is my personal opinion with a sense of self-admonition—to my mind it is still no excuse for lacking, in today's world, a solid grounding in data science and numerical analysis.

Indeed, to the modern researcher, electron microscopes and other advanced measurement instruments are more than just metrology systems—they are *platforms for research and development*. As long as instrument vendors continue to provide users with open software development environments accompanying their instruments, we can be sure that forward-thinking researchers and ambitious startup companies around the world will continue to develop new tools to enhance materials characterization.

3. Data Science as a Mandatory Component of Basic Literacy for Modern Researchers

In my student days, I joined a research group in which high-level technical staff was developing laboratory instruments under professors' supervision. That department of university had access to a workshop, which—to my amazement—successfully fabricated an ultra-high-vacuum component from a rather incompetent sketch I had made. By fusing *science* with *engineering*, we were able to do cutting-edge research; I learned that the pace of R&D can be accelerated by a commitment to testing each and every idea—no matter how big or how small—as soon as possible. To my mind, the research organizations most likely to survive and thrive going forward are the ones that implement and test new ideas at early stages—and here I have in mind not just ideas for fabricating and observing samples, but also ideas for improving laboratory instruments, for control systems, for analysis software, and so forth.

I would like to share a topic from my research that illustrates the marriage of electron microscopy and data science. Recent years have seen growing interest in 4D-STEM, a form of scanning transmission electron microscopy (STEM) that captures many diffraction patterns while scanning the incident probe³⁾. Our research group had applied nonnegative matrix factorizations (a dimensionality-reduction technique used in machine learning) to analyze large numbers of diffraction patterns⁴⁾; this yields a small collection of diffraction patterns from which the full set of experimentally-measured diffraction patterns can be approximately reconstructed, and we used this approach successfully to detect nanocrystalline precipitates in a metallic glass, identifying the crystal structure and the mean grain size⁵⁾. When we began this research project, we asked a colleague in another research group to develop the computational code; this seemed faster and less expensive than outsourcing the coding to an external contractor. However, we quickly found ourselves inundated with new ideas we wanted to try, and would have felt entirely too *guilty* asking him to develop the code because most of our ideas wound up not working well. Thus, before long I bowed to the inevitable and began developing the code myself. This significantly accelerated the "Plan-Do-Check" loop of our research, in which we would implement a new idea, test its effectiveness, then connect it to a follow-up idea; developing the code myself also enhanced my understanding of existing machine-learning libraries (such as Scikit-learn), and at present we are studying the possibility of adding constraints about electron microscopy. We plan to distribute our code freely via GitHub and other open-source repositories to allow other researchers to benefit from our work. Within the machine-learning community, specialized knowledge possessed by members of outside research disciplines—like us—is called *domain knowledge*; going forward, it seems certain that R&D initiatives to embed domain knowledge within existing machine learning will proceed apace throughout a broad range of research fields.

Incidentally, it is amusing to note that research on non-negative matrix factorizations seems to have begun in earnest after the publication of a Nature article in 1999⁶⁾—a date that I, now aged 60, cannot help but think of as rather recent. Of course, the technique is ultimately based on the linear algebra that I supposedly learned as an undergraduate some 40 years ago—and which I am currently in the process of reviewing. Around 10 years ago, I was complaining that, for any given topic or paper of interest, "there are hopelessly too many related papers—it's *impossible* to keep up." In response,

a professor I admired told me: "Studying and learning are lifelong pursuits" which made me reflect on my attitude: if even this *genius* would need to keep studying throughout his life, then *my* brain would require studies lasting well beyond death. Meanwhile, another colleague I respect insists that "researchers these days don't study *nearly* enough." If we do not update our knowledge continuously, our research activities are not sustainable. I am currently affiliated with a research center populated both by researchers studying advanced materials characterization and researchers studying materials informatics^{7,8)}, and the need for characterization researchers to acquire interdisciplinary data-science skills—as a component of basic literacy—has never seemed to me more urgent than it is today.

4. Conclusions

The 2024 Nobel prizes in both physics and chemistry recognized work related to artificial intelligence (AI). The significance of this historical confluence seems clear: In research fields—very much including the study of materials and the study of characterization techniques—the question of how to incorporate AI is *the* central challenge of our time. In the years to come, advanced measurement systems will continue to produce multi-dimensional datasets of ever greater enormity, thus requiring a continuing infusion of machine-learning techniques to handle the data explosion; consequently, this infusion will not be a one-time transient event, but will instead evolve into a long-term process of entrenchment. Electron microscopes play a key role as advanced scientific measurement instruments, but require precise alignments to ensure correct results; this is one area in which the infusion of AI techniques promises significant advances in coming years. Ever since the invention of the electron microscope by Ernst Ruska (recipient of the 1986 Nobel Prize in Physics) in 1931—and continuing right up to the present—the field has witnessed continual technological progress on many simultaneous fronts. I am confident that the fusion of advanced electron microscopy techniques with modern methods of data science, and the further incorporation of other disciplines through open R&D environments, promises a wealth of unexpected discoveries driving material innovation.

References

- 1) K. Kimoto, K. Mitsuishi, T. Hara, M. Mitome, T. Nagai, Transmission Electron Microscope for Materials Research, Kodansha, 2020 (ISBN978-4065203866) (in Japanese).
- 2) Research Data Express (RDE), <https://dice.nims.go.jp/services/RDE/>
- 3) K. Kimoto and K. Ishizuka, Spatially resolved diffractometry with atomic-column resolution, *Ultramicroscopy*, **111**, 1111-1116 (2011).
- 4) F. Uesugi, S. Koshiya, J. Kikkawa, T. Nagai, K. Mitsuishi, K. Kimoto, Non-negative matrix factorization for mining big data obtained using four-dimensional scanning transmission electron microscopy, *Ultramicroscopy*, **221**, 113168 (2021).
- 5) K. Kimoto, J. Kikkawa, K. Harano, O. Cretu, Y. Shibasaki, F. Uesugi, Unsupervised machine learning combined with 4D scanning transmission electron microscopy for bimodal nanostructural analysis, *Scientific Reports*, **14**, 2901 (2024).
- 6) D. D. Lee and H. S. Seung, Learning the parts of objects by non-negative matrix factorization, *Nature*, **401**, 788-791 (1999).
- 7) Author information
https://samurai.nims.go.jp/profiles/kimoto_koji?locale=en
- 8) Research group information
<https://www.nims.go.jp/AEMG/index.html>