

Large Dataset Generation, Integration and Simulation in Materials Science

Jeff Simmons and J.-C. Zhao

As new and rapid materials characterization tools become widely available, large datasets are being generated at a faster and faster pace, which is both exciting and challenging. The exciting aspects are that large amounts of composition, structure, and spectral information about materials are being gathered routinely. This promises to accelerate the traditional scientific pursuits in materials science of establishing processing–structure and structure–property relationships and hypothesis testing. Even more exciting is the promise of quantitative validation of complex physics-based models with the statistically significant datasets that are now becoming possible. The main challenges presented by the emergence of these techniques are rapid acquisition of larger and larger datasets, using this information to provide predictive capabilities in materials systems, and analysis of the increasing volumes of data. The selected articles highlighted here provide a snapshot of some current activities in this expanding field, but in no way attempt to cover the entire field. Such an attempt would be well beyond the page limit of a *JOM* issue. In fact, this topic, due to the substantial number of material to cover, has been spread over two months: March and July.

Within Part I three articles were chosen to highlight data generation. The article by Uchic et al. describes the current state of the art of automated serial sectioning. That by Barabash et al. describes one of the exciting developments in x-ray characterization: characterization of internal strains in materials. Zhao et al. describe recent developments in femtosecond-laser based measurements of thermal conductiv-

ity, expansion coefficients, and elastic modulus in micron-scale spatial resolution and their applications in rapid collection of composition-phase-property data.

Beyond characterization, materials engineering and materials science are concerned with prediction and explanation of properties from the standpoint of structure. Two articles were chosen to highlight important considerations when making this connection. Simple in concept, the idea of modeling the response of a 3-D structure is actually a complex operation. The article by Lewis et al. highlights the details necessary for accomplishing this. It is often taken for granted that a microstructure can be described, but to provide quantitative descriptions of a fundamentally random input structure suitable for simulations is an unsolved problem and is currently an active area of research. The article by McDowell et al. describes the current state of the art of representing random media for the purposes of modeling properties.

Finally, the increasing data rate of modern microscopy is generating a new focus for research: that of converting measurements to context. The output of a microscope observation is typically an image that the trained eye can interpret. But, microstructure images are not images, per se, but consist of thousands or millions of individual measurements that are displayed as a picture. The actual value stored at each pixel location is the magnitude of a signal under certain conditions. Converting this fundamentally memory-based representation to a context-based representation (e.g., particle sizes, lath thicknesses, morphologies, etc.) suitable for modeling

constitutes an emerging problem in large dataset generation: that of analysis.

Rowenhorst and Lewis describe the issues involved in converting 3-D serial section data to a materials context. Since this is a painstaking process, research is being devoted to developing algorithms that require less human interaction to analyze large datasets. Image segmentation is an exploding field, mostly fueled by the proliferation of medical imagery and the introduction of the digital camera: it is not possible to cover this field with any fidelity. Instead, we highlight several techniques currently in development for performing segmentations in materials images. The articles in the July issue will describe four promising techniques: statistical segmentation (including Bayesian methods), inverse diffusion, graph-cut methods, and level set methods.

The articles in this two-part topic represent only a sampling of the exciting developments in this rapidly expanding field. It is with regrets that we could not cover all of the interesting developments in the field, including 3-D atom probe, MultiModal data collection and Fusion, 3-D + t time series tomography, and the many physics-based “forward models” that can be validated for the first time with the rich datasets being collected, among many others. Keep watching this space for further developments.

Jeff Simmons is with the Air Force Research Laboratory and is a member and former chairman of the Advanced Characterization, Testing and Simulation (ACTS) Committee of the Structural Materials Division of TMS. J.-C. Zhao, professor, is with the Department of Materials Science and Engineering, The Ohio State University and is the advisor to *JOM* from the ACTS Committee of the Structural Materials Division of TMS.