Informatics-Enabled Design of Structural Materials

JENNIFER L.W. CARTER [1,3] and AMIT K. VERMA [2,4]

1.—Department of Materials Science and Engineering, Case Western Reserve University, Cleveland, OH 44106, USA. 2.—NextManufacturing Center, Carnegie Mellon University, Pittsburgh, PA 15213, USA. 3.—e-mail: jwc137@case.edu. 4.—e-mail: amitk2@andrew.cmu.edu

Informatics-enabled design is a paradigm shift for materials engineering and has led to many breakthroughs within the last decade.[1] The term "materials informatics" joined the publication keyword vernacular around 2005[2,3] (Fig. 1). Fifteen years later, there have been 3000+ articles published on the application and development of informatics, data science, and machine learning techniques to materials exploration and design, and informatics is being taught in a few undergraduate materials engineering curriculums to balance courses on statistics, experimental study design, and physics-based computational modeling. There are pervasive challenges of applying these informatics techniques to any particular class of material. These include: the construction of open Big Data repositories, the implementation of machine learning to sparse datasets, and platform design for materials discovery.[4] Additionally, there are an array of unique challenges that arise in advancing the informatics-enabled design of structural materials because of the need for high-throughput quantitative evaluation of performance metrics across multiple time and length scales that require destructive measures.[5]

This special topic aims to present some of the needs and limitations of the informatics toolsets for the design of structural materials. We have invited authors to speak on three challenges specifics to high-throughput metric evaluations necessary for informatics-enabled structural materials design: incorporation of cluster expansion theory with first-principles calculations; the value of novel and

industrial-scale additive manufacturing for high-throughput synthesis, and characterization of mechanical properties; and the need to quantify microstructural distributions and extremes from three-dimensional datasets. Finally, we present a review of a fringe, yet pervasive, topic (only because of the still large and unknown challenges) of expediting literature reviews by creating algorithms for natural language search of heritage publication data.

Density functional theory (DFT) is the workhorse of modern computational materials research but is computationally expensive and, in practice, even with state-of-the-art supercomputers, is limited to a relatively small number of atoms/unit cells. This computational expense limits the ability to apply DFT to explore the design microstructures needed for structural materials. In contrast, cluster expansion (CE) theory provides a direct approximation of the lattice free energy, or other thermodynamic variables, in terms of a discrete cluster function, making it one of the most widely used approaches for phase diagram calculations. The CE theory enables the application of DFT-based simulations for exploring the design of multicomponent systems and systems with chemical disorder. In the paper "The Cluster Expansion of Alloy Theory: Historical Development and Modern Innovations," Kadkhodaei (Univ. of Illinois, Chicago) and Munoz (Univ. of Texas, El Paso) map the theoretical developmental origins of CE that culminated in its current formalism to provide insights into how CE theory can be applied for the DFT design of structural materials. Examples include the interaction of antiphase boundaries with the lattice, stacking fault energy calculations, surface phenomena such as adsorption and segregation, as well as precipitation with coherent and incoherent interfaces.

While CE–DFT coupled theories allow us to explore a larger number of questions and provide a framework for high-throughput computational screening, high-throughput synthesis and
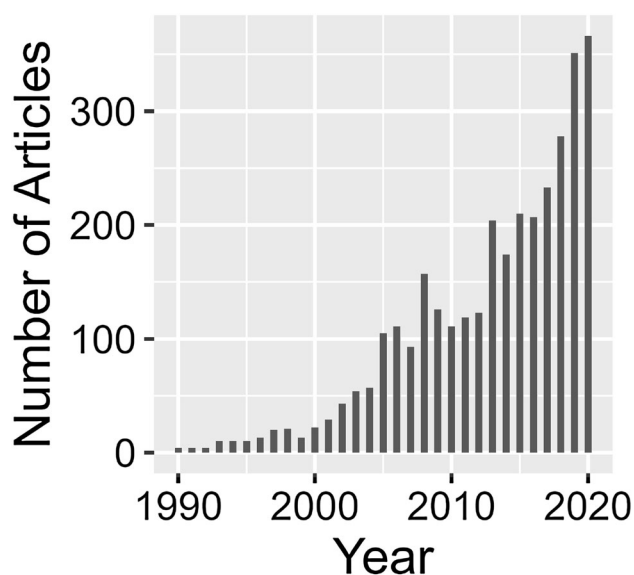
Fig. 1. Materials informatics emerged as a recurring keyword in 2005, and 15 years later there have been 3282 articles in Scopus published with this term.

characterization methods are needed to experimentally validate novel systems and assess questions of manufacturability. In the paper "Experimental Methods to Enable High-Throughput Characterization of New Structural Materials," Ellendt and his colleagues (Univ. of Bremen, Germany) have put forward *Farbige Zustände*, as a novel approach to achieve this. The approach features a high-temperature on-demand droplet generator to produce spherical microsized samples, which are then heat-treated and subjected to various short-time characterizations, which yield a large number of physical, mechanical, technological, and electrochemical descriptors. Through the case study, Ellendt et al. show that this method can be used to synthesize, heat treat, and characterize 6,000+ different steel samples within 1 week. Moreover, more than 90,000 descriptors can be derived to specify the material profiles of the different alloys during this period. These descriptors can be extrapolated, or correlated with material properties at the macroscale, and simultaneously push forward the informatics-enabled design of structural materials.

Extending beyond the understanding of a material behavior/properties lies the synthesis route through which the material comes to life. Metal additive manufacturing (AM) holds the promise of revolutionizing manufacturing by democratizing the technology needed for synthesizing high-performance high-temperature materials and enabling the manufacturing of complex-geometry components for extreme environments from materials (such as titanium aluminides) that are conventional unmachinable due to low ductility at room temperature. Currently, metal AM has been under development through various means, and machine learning (ML)/artificial intelligence (AL) is

prominent in the cycle. In the paper titled "High-Throughput Statistical Interrogation of Mechanical Properties with Build Plate Location and Powder Reuse in AlSi10Mg," Carroll et al. (Sandia National Laboratory) argue the need for statistically significant datasets to develop insights about the AM processes, and how attributes that are inherent to AM can be leveraged in achieving these. The paper presents a high-throughput tensile testing technique that provides efficient data to answer questions such as whether powder reuse, or location on the build plate, or the size affects the final material properties.

Further, informatics is providing efficient means of probing 3D microstructural evolution as measured in materials during in situ thermal–mechanical conditions. One such example of 3D microstructural evolution is high-energy x-ray diffraction microscopy (HEDM), which provides unprecedented information on local field evolution under applied stress and has revealed new physical mechanisms that were previously not observed while being valuable for developing microstructure-aware models for accurate material property predictions. However, these measurements are extremely slow, and as a result, material kinetics cannot be studied due to limited temporal resolution. In the paper "Physics-Informed Data-Driven Surrogate Modeling for Full-Field 3D Microstructure and Micromechanical Field Evolution of Polycrystalline Materials," Pokharel et al. (Los Alamos National Laboratory) developed an ML-based crystal plasticity model to provide fast inference of spatially resolved 3D microstructure, its micromechanical fields, and their evolution. As a result, this framework coupled with HEDM can be used to carry out targeted beamline experiments, thereby increasing the overall output of the limited beamtime. Many similar ML/AI-based surrogate models are successfully replacing computationally expensive, physics-based models, further accelerating the informatics-driven design of materials.

Finally, the question of legacy data/knowledge, embedded in the plain text of published articles, for the needs of ML/AI algorithms is being approached by a few members of the community. With recent advances in the field of natural language processing (NLP), the extraction of knowledge is possible from some colloquial-speech plain text, although the NLP tools are not conducive to scientific texts. Individual communities have taken up this challenge and are bridging the gaps to make this a reality. The paper "Challenges and Advances in Information Extraction from Scientific Literature: A Review" by Hong et al., (Univ. of Chicago in collaboration with Argonne National Laboratory) presents an exhaustive review detailing the challenges that the materials science community will need to overcome to leverage the modern NLP tools for future research. The upshot of these efforts is hard to put into perspective, though given we are approaching an

era of autonomous experimentation, intelligent systems with a wide knowledge base are a prerequisite to avoid the pitfalls that may arise with narrow-focused AI.

All titles and authors of the articles are published under the topic "Informatics-Enabled Design of Structural Materials" in the November 2021 issue (vol. 73, no. 11) of *JOM*. The articles can be accessed fully via the journal's page at: http://link.springer.com/journal/11837/73/11/page/1

REFERENCES

1. R. Ramprasad, R. Batra, G. Pilania, et al., Machine learning in materials informatics: recent applications and prospects. *npj Comput. Mater.* 3, 1–13. https://doi.org/10.1038/s41524-017-0056-5 (2017).
2. U.S. Schubert, Materials informatics: from data to knowledge towards integrated eScience approaches. *QSAR Comb. Sci.* 24, 5. https://doi.org/10.1002/qsar.200590002 (2005).
3. K. Rajan, Materials informatics. *Mater. Today* 8, 38–45. https://doi.org/10.1016/S1369-7021(05)71123-8 (2005).
4. L. Himanen, A. Geurts, A.S. Foster, and P. Rinke, Data-driven materials science: status, challenges, and perspectives. *Adv. Sci.* 6, 1900808. https://doi.org/10.1002/advs.201900808 (2019).
5. J.L.W. Carter, A.K. Verma, and N.M. Senanayake, Harnessing legacy data to educate data-enabled structural materials engineers. *MRS Adv.* 5, 319–327. https://doi.org/10.1557/adv.2020.132 (2020).