Molecular Systems Design & Engineering

EDITORIAL

View Article Online
View Journal | View Issue

Cite this: *Mol. Syst. Des. Eng.*, 2018, 3, 429

DOI: 10.1039/c8me90007h

rsc.li/molecular-engineering

Machine learning and data science in materials design: a themed collection

Andrew Ferguson [abc] and Johannes Hachmann [def]

Guest Editors Andrew Ferguson and Johannes Hachmann introduce this themed collection of papers showcasing the latest research leveraging data science and machine learning approaches to guide the understanding and design of hard, soft, and biological materials with tailored properties, function and behaviour.

The application of data-driven modeling and machine learning in the materials domain is opening new paths to the understanding, design, and engineering of next-generation materials systems. Traditionally, physical laws that define the fundamental connections between a material's composition and its structure and function are used as the foundation for analytical or numerical models, and these physics-based approaches provide a route to assess candidate compounds with respect to properties of interest. The inverse problem – engineering a novel material with particular properties – has become a focus of cutting-edge research efforts aimed at accelerating the discovery and design process. Inverse engineering is more challenging as it is generally not possible to simply "invert" a physics-based model and run it in reverse. Instead, as large-scale data generated by modern experimental and computational approaches are becoming more readily available, paradigms and tools from data science offer a new way to engage both the forward and inverse modes of inquiry.

In forward problems, informatics techniques can facilitate high-throughput virtual screening studies, and data mining approaches can help uncover latent correlations, or even the underlying mechanisms, governing a system's behavior. Such relationships are typically not intuitively apparent or readily accessible from massive and/or high-dimensional data sets. Machine learning allows for the construction of inexpensive data-derived prediction models to circumvent, or reduce the reliance upon, expensive physics-based modeling or experimentation. In inverse problems, the structure–property relationships resulting from forward analyses can be utilized for the rational, *de novo* design of new materials with tailored features. Statistical inference techniques are invaluable in performing a principled interpolation between sparse observations within chemical or materials space, and in directing the exploration of this space towards promising candidates.

This collection of invited papers showcases a diverse set of investigations in which the integration of data science tools with domain expertise has led to advances in materials research. These include new insights into the properties, functionality, and behaviors of hard, soft, and biological materials, as well as the acceleration of discovery and design efforts. These contributions demonstrate the immense potential of data science techniques in materials and chemical science and engineering, and are emblematic of a rapidly growing body of work implementing these paradigms and tools in all corners of the discipline.

[a] *Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, 1304 West Green Street, Urbana, IL 61801, USA. E-mail: alf@illinois.edu; Fax: +1 217 333 2736; Tel: +1 217 300 2354*
[b] *Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, 600 South Mathews Avenue, Urbana, IL 61801, USA*
[c] *Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, IL 61801, USA*
[d] *Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260, USA. E-mail: hachmann@buffalo.edu*
[e] *New York State Center of Excellence in Materials Informatics, Buffalo, NY 14203, USA*
[f] *Computational and Data-Enabled Science and Engineering Graduate Program, University at Buffalo, The State University of New York, Buffalo, NY 14260, USA*

Andrew Ferguson

Andrew Ferguson is Associate Professor of Materials Science and Engineering, and Chemical and Biomolecular Engineering, and an Affiliated Associate Professor of Physics, and Computational Science and Engineering at the University of Illinois at Urbana-Champaign. His research group combines molecular simulation, statistical thermodynamics and machine learning to understand and engineer soft and biological materials and antiviral vaccines. He is the recipient of a 2017 UIUC College of Engineering Dean's Award for Excellence in Research, 2016 AIChE CoMSEF Young Investigator Award for Modeling & Simulation, 2015 ACS OpenEye Outstanding Junior Faculty Award, 2014 NSF CAREER Award, 2014 ACS PRF Doctoral New Investigator and was named the Institution of Chemical Engineers North America 2013 Young Chemical Engineer of the Year.

Johannes Hachmann

Johannes Hachmann is an Assistant Professor of Chemical Engineering at the University at Buffalo (UB), and a Faculty Member of the UB Computational and Data-Enabled Science and Engineering graduate program and the New York State Center of Excellence in Materials Informatics. His research spans the areas of computational chemistry, computational materials science, and applied data science in chemistry and materials. He received a Dipl.-Chem. degree (2004) after undergraduate studies at the Universities of Jena and Cambridge, earned MSc (2007) and PhD (2010) degrees in chemistry from Cornell University, and conducted postdoctoral research at Harvard University before joining the UB faculty in 2014.