CHEMICAL REVIEWS

pubs.acs.org/CR

Editorial

Introduction: Machine Learning at the Atomic Scale

Cite This: *Chem. Rev.* 2021, 121, 9719–9721

Read Online

ACCESS | Metrics & More | Article Recommendations

Looking for patterns in structures and trends in data is one of the defining features of science in general, and chemistry is no exception. Traditionally based upon background knowledge from experiments, chemical intuition, and physics-based derivations, over the past decades chemical research relied increasingly on the use of computers due to the ever accelerating digital transformation of society. This has resulted in the formation of a large, vibrant, and diverse atomistic simulation community which by now boasts multiple decades worth of expertise and tradition. It speaks to the encouraging progress that nowadays many archetypical problems in chemistry can be tackled successfully by simulation, such as accurately predicting vibrational and electronic spectra of small systems in vacuum or calculating protein binding affinities of ligands in explicit solvent which compare well to experimental measurements. The increasing availability of experimental and calculated data, as well as the growing confidence in statistical learning and mathematical data processing schemes, makes it nowadays possible to also automate some of the more repetitive chemical research tasks, allowing us all to go beyond what traditionally could have been done when relying on conventional techniques. For instance, by using machine learning on large amounts of previously acquired data (calculated or measured), available from the rapidly growing body of literature, or through open access/source protocols from well-established online repositories, it is nowadays not out of the ordinary to search, detect, and exploit hitherto unknown patterns and trends which were "hidden" in the high dimensional nonlinear spaces directly relevant to chemistry.

This is not to say that physics-based simulation techniques and intuition-driven research has become moot. On the contrary, the understanding of chemistry through the laws of quantum mechanics and statistical mechanics, as well as through extensive scholarship documenting decades, if not centuries, of experimental observation, has proven invaluable to further guide, develop, and inform the conception and implementation of data-driven techniques. More specifically, the synergistic application of machine learning and traditional atomistic modeling continues to serve as an accelerator of discovery, e.g. extending the reach, quality, and number of high-end electronic structure calculations. But also scalable approximate machine-learning potentials, trained on small systems and used to infer solutions for problems of much larger time and length scales, have had considerable success. Last but not least, statistical sampling techniques have been revisited, enabling a much accelerated and automatized analysis of complex transitions or improved predictions and understanding of order paramters. In conclusion there is ample opportunity for optimism that historically cumbersome tasks can be streamlined, harsh approximations of the quantum behavior of electrons and nuclei can be overcome, and harmful neglect of higher dimensions can be mitigated. We strongly believe that this will open up countless possibilities to pursue computational chemistry research projects at new levels, i.e. with unprecedented accuracy, reliability, and scope. In the remainder, we briefly outline the content of this thematic issue which can loosely be grouped into two categories: First those reviews which deal with machine learning at the atomistic scale in a more universal sense, i.e. dealing with the statistically rigorous definition of structural motifs at the atomic scale, the detection of recurring patterns, and the rigorous determination of quantitative structure−property trends throughout diverse chemistries. The second category contains reviews which focus on the specific problem of predicting accurate potential energies (and derivatives), which are of extraordinary importance for all molecular modeling applications.

I. PATTERNS THROUGHOUT CHEMISTRY

Unsupervised Learning Methods for Molecular Simulation Data

Laio and co-workers contribute an account of unsupervised learning studies in the context of the increasing availability of large data sets stemming from molecular simulations, covering the mathematical foundations of the methods as well as the main ideas that have been used to adapt them to atomistic modeling.

Physics-Inspired Structural Representations for Molecules and Materials

Ceriotti and co-workers summarize the current understanding of the process of building a mathematical representation of an atomistic structure, oftentimes the first step into the application of machine-learning algorithms. They highlight the deep similarities that unite the most widely adopted families of descriptors and the interplay between regression and classification models and the underlying description of their inputs.

Special Issue: Machine Learning at the Atomic Scale

Published: August 25, 2021

Combining Machine Learning and Computational Chemistry for Predictive Insights into Chemical Systems

Keith and co-workers discuss the wider context within which computational chemistry and machine learning methods can be united. They also include concise tutorials of methods at the intersection between computational and physical sciences.

Machine Learning for Electronically Excited States of Molecules

Westermayr and Marquetand summarize the state of the art on using machine learning to predict quantum properties. In particular, the focus lies on electronically excited states, that play a fundamental role in photochemical and photophysical processes. Multiple aspects are covered ranging from the calculation of absorption spectra to dynamics in the excited-state and nonadiabatic effects.

Computational Discovery of Transition-Metal Complexes: From High-Throughput Screening to Machine Learning

The contribution by Kulik and co-workers is a focused overview of the application of computational chemistry and machine-learning techniques to the computational discovery of transition-metal complexes. From high-throughput data generation, to the identification of structure−property relations, it highlights the role played by automated data analytics and the specific considerations that make this area of application particularly suitable for machine learning.

Ab Initio Machine Learning in Chemical Compound Space

Huang and von Lilienfeld give a comprehensive historical account of machine learning related work aimed at the exploration of chemical compound space in ways that are consistent with quantum mechanics. The review emphasizes the importance of efficient and accurate surrogate models in removing the bottlenecks imposed by the high computational requirements of state-of-the-art quantum mechanical calculations, extending the portion of this enormous space that can be investigated.

II. POTENTIAL ENERGIES

Four Generations of High-Dimensional Neural Network Potentials

The comprehensive review of Behler details the historic evolution of neural network potentials for materials, discussing four generations of models that have incorporated increasingly rich types of physical interactions and that are applicable to systems with many degrees of freedom.

Gaussian Process Regression for Materials and Molecules

Csányi and co-workers provide an introduction to Gaussian process regression, focusing in particular on the construction of Gaussian approximation potentials—that learn from and predict energy and forces between atoms—but covering also symmetry-adapted extensions of the method that can be applied to fit vectorial and tensorial properties.

Machine Learning Force Fields

In their review, Müller and co-workers present an overview of the development and application of force fields trained on quantum chemistry data, that combine the accuracy of first-principles calculations with the efficiency of empirical force fields. They emphasize the universality of machine-learning force fields that do not rely on prior definition of chemical bonds and how their accuracy is only limited by the quality and amount of reference data.

Neural Network Potential Energy Surfaces for Small Molecules and Reactions

Manzhos and Carrington outline the progress made possible by the application of machine-learning techniques to generate highly accurate approximations of the potential energy surfaces for small molecules and reactions. They also discuss the relevance for quantum dynamics calculations and the recent advances that make it possible to use these methods for larger, more complex molecules.

Machine Learning for Chemical Reactions

Meuwly focuses his review on studies related to reactive processes. He summarizes the long history of machine-learning applications in the field, that range from the use of Bayesian inference for the incorporation of experimental information to the explicit simulation of reactive networks using machine-learning potentials.

Michele Ceriotti orcid.org/0000-0003-2571-2832
Cecilia Clementi
O. Anatole von Lilienfeld

AUTHOR INFORMATION

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.chemrev.1c00598

Notes

Views expressed in this editorial are those of the authors and not necessarily the views of the ACS.

Biographies

Michele Ceriotti is Associate Professor at the Institute of Materials at the École Polytechnique Fédérale de Lausanne. He received his Ph.D. in Physics from ETH Zürich in 2010, under the supervision of Professor Michele Parrinello. He spent three years in Oxford as a Junior Research Fellow at Merton College and joined EPFL in 2013, where he leads the laboratory for Computational Science and Modeling. His research interests focus on the development of methods for molecular dynamics and the simulation of complex systems at the atomistic level, as well as their application to problems in chemistry and materials science—using machine learning both as an engine to drive more accurate and predictive simulations and as a conceptual tool to investigate the interplay between data-driven and physics-inspired modeling.

Cecilia Clementi is Einstein Professor of Physics at Freie Universität (FU) Berlin, Germany. She joined the faculty of FU in June 2020 after 19 years as a Professor of Chemistry at Rice University in Houston, Texas. Cecilia obtained her Ph.D. in Physics at SISSA and was a postdoctoral fellow at the University of California, San Diego, where she was part of the La Jolla Interfaces in Science program. Her research focuses on the development and application of methods for the modeling of complex biophysical processes, by means of molecular dynamics, statistical mechanics, coarse-grained models, experimental data, and machine learning. Cecilia's research has been recognized by a National Science Foundation CAREER Award (2004) and the Robert A. Welch Foundation Norman Hackerman Award in Chemical Research (2009). Since 2016 she has also been a co-Director of the National Science Foundation Molecular Sciences Software Institute.

O. Anatole von Lilienfeld is a full university professor of computational materials discovery at the Faculty of Physics at the University of Vienna. Research in his laboratory deals with the development of improved methods for a first-principles-based understanding of chemical compound space using perturbation theory, machine

https://doi.org/10.1021/acs.chemrev.1c00598
Chem. Rev. 2021, 121, 9719−9721

learning, and high-performance computing. Previously, he was an associate and assistant professor at the University of Basel, Switzerland, and at the Free University of Brussels, Belgium. From 2007 to 2013, he worked for Argonne and Sandia National Laboratories after postdoctoral studies with Mark Tuckerman at New York University and at the Institute for Pure and Applied Mathematics at the University of California Los Angeles. In 2005, he was awarded a Ph.D. in computational chemistry from EPF Lausanne under the guidance of Ursula Rothlisberger. His diploma thesis work was done at ETH Zurich with Martin Quack and the University of Cambridge with Nicholas Handy. He studied chemistry at ETH Zurich, the Ecole de Chimie Polymers et Materiaux in Strasbourg, and the University of Leipzig. He serves as editor in chief of the IOP journal *Machine Learning: Science and Technology* and on the editorial board of *Science Advances*. He has been on the editorial board of *Nature's Scientific Data* from 2014 to 2019. He was the chair of the long IPAM "UCLA program 'Navigating Chemical Compound Space for Materials and Bio Design'" which took place in 2011. He is the recipient of multiple awards including the Swiss National Science foundation postdoctoral grant (2005), Harry S. Truman postdoctoral fellowship (2007), Thomas Kuhn Paradigm Shift award (2013), Swiss National Science professor fellowship (2013), Odysseus grant from Flemish Science foundation (2016), ERC consolidator grant (2017), and Feynman Prize in Nanotechnology (2018).

ACKNOWLEDGMENTS

We would like to thank all authors of the reviews, the reviewers who ensured the clarity and scholarship of each contribution, and the editors of *Chemical Reviews* for their support and patience in preparing this collection. In memoriam Alessandro de Vita (1965−2018).

https://doi.org/10.1021/acs.chemrev.1c00598
Chem. Rev. 2021, 121, 9719−9721