⟨S|P|S⟩ swiss physical society

Nr. 59
Oktober 2019

SPG MITTEILUNGEN
COMMUNICATIONS DE LA SSP

AUSZUG - EXTRAIT

Artificial intelligence in materials science - hype or revolution? PT 5/2019

Claudia Draxl, Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin and Fritz Haber Institute of the Max Planck Society, Berlin

Artificial intelligence in materials science - hype or revolution? PT 5/2019

Claudia Draxl, Physics Department and IRIS Adlershof, Humboldt-Universität zu Berlin and Fritz Haber Institute of the Max Planck Society, Berlin

The growth of data from simulations and experiments is expanding beyond a level that is addressable by established scientific methods. The so-called "4 *V* challenge" of Big Data – *Volume* (the amount of data), *Variety* (the heterogeneity of form and meaning of data), *Velocity* (the rate at which data may change or new data arrive), and *Veracity* (uncertainty of quality) – is clearly becoming eminent also in the sciences. Controlling our data, in turn, sets the stage for explorations and discoveries.

Novel approaches and tools of Artificial Intelligence (AI) can find patterns and correlations in data that cannot be obtained from individual calculations or experiments and not even from high-throughput studies. In fact, data-driven research is adding a new research paradigm to the scientific landscape, as indicated in Figure 1, which shows the historical development of the research paradigms in our field.

For a real breakthrough, *Open Data* and sharing, as well as an efficient data infrastructure is key. In other words, for shaping this forth paradigm and contributing to the development or discovery of improved and novel materials, data must be what is now called FAIR - Findable, Accessible, Interoperable and Reusable [2].

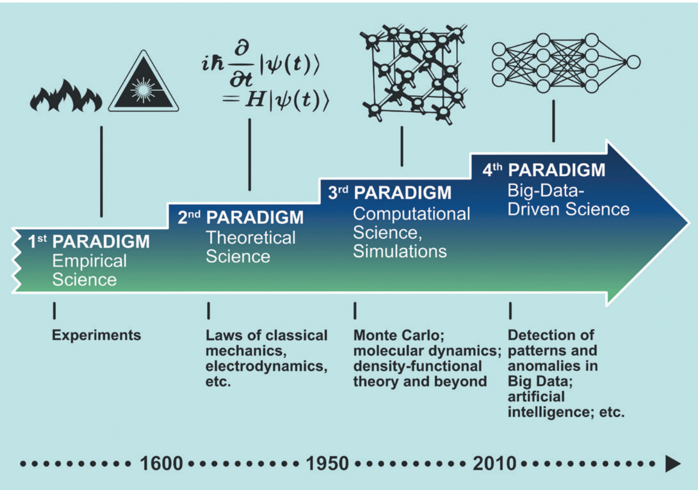

Figure 1: Development of research paradigms of materials science and engineering (from [1]).

The NOMAD Laboratory [3,4] is a living example for such infrastructure in computational materials science, comprising the NOMAD *Repository* (raw data) and its *Archive* (normalized, i.e. code-independent data). The NOMAD *Encyclopedia* is a web-based public platform that visualizes the properties of this large variety of materials. The NOMAD Analytics Toolkit provides a collection of examples and tools to demonstrate how materials data can be turned into knowledge by AI approaches (e.g. [5,6]).

What are the challenges [7] ahead of us to make the data revolution happen? On the computational side, reliable error quantification (for a very first step, see [8]) and trust levels that help making data comparable; efficient algorithms and implementations of high-level methods for creating the *right* data; and novel AI tools are among them. Figure 2 sketches this situation.

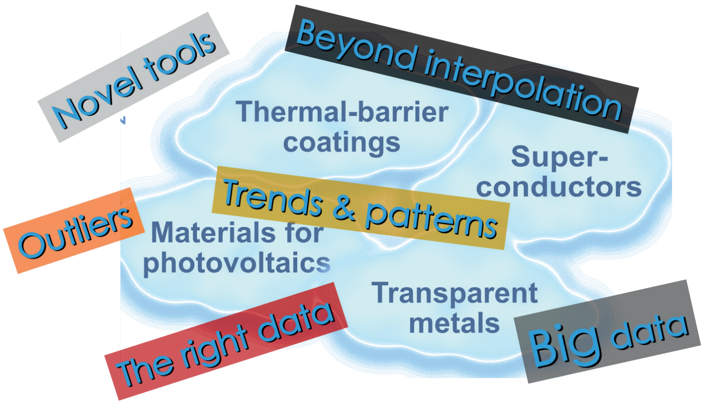

Figure 2: Challenges on the way to novel discoveries.

It also indicates NOMAD's long-term vision, namely to provide (high-dimensional) materials maps that tell where in the composition and compound space one can find materials for a certain application. The very big next step, that has indeed already been started [9], is exploring FAIRness in data of experimental materials science, synthesis, and a number of functional materials.

*Work in collaboration with Matthias Scheffler and the entire NOMAD Team.

References:

[1] C. Draxl and M. Scheffler, *Big-Data-Driven Materials Science and its FAIR Data Infrastructure*, Invited Perspective in Handbook Andreoni W., Yip S. (eds) Handbook of Materials Modeling. Springer, Cham (2019).
[2] M. D. Wilkinson et al., The FAIR Guiding Principles for scientific data management and stewardship, Sci Data 3, 160018 (2016).
[3] The NOMAD Laboratory, https://nomad-coe.eu/, including metadata and various software for parsing, analysis, and visualization.
[4] C. Draxl and M. Scheffler, NOMAD: The FAIR Concept for Big-Data-Driven Materials Science, MRS Bulletin 43, 676 (2018).
[5] L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl, and M. Scheffler, *Big Data of Materials Science - Critical Role of the Descriptor*, Phys. Rev. Lett. 114, 105503 (2015).
[6] L. M. Ghiringhelli, J. Vybiral, E. Ahmetcik, R. Ouyang, S. V. Levchenko, C. Draxl, and M. Scheffler, *Learning physical descriptors for materials science by compressed sensing*, New J. Phys. 19, 023017 (2017).
[7] C. Draxl and M. Scheffler, *The NOMAD Laboratory: From Data Sharing to Artificial Intelligence*, J. Phys. Mater. 2, 036001 (2019).
[8] K. Lejaeghere et al., *Reproducibility in density-functional theory calculations of solids*, Science 351, aad3000 (2016).
[9] FAIR-DI - FAIR Data Infrastructure for Physics, Chemistry, Materials Science, and Astronomy e.V., is a non-profit association, https://fair-di.eu).