Big Data of Materials Science from First Principles -- Critical Next Steps

Matthias SCHEFFLER
Fritz Haber Institute of the Max Planck Society,
Berlin

Using first-principles electronic-structure codes, a huge number of materials has been studied in recent years. The amount of already created data is immense. Thus, the field is facing the challenges of "Big Data", which are often characterized in terms of the *"four V"*: *Volume* (amount of information), *Variety* (heterogeneity of the form and meaning of the data), *Veracity* (uncertainty of the data quality), and *Velocity* at which data may change or new data arrive.

Obviously, the computed data may be used as is: query and read out what was stored. However, for achieving deeper and novel scientific insight, the *four V* should be complemented by an "*A*", the Big-Data *Analysis*. Calculating properties and functions for many materials, (e.g. efficiency of potential photovoltaic, thermoelectric, battery, or catalytic materials) is the necessary first step. Finding the actuating mechanisms (the "causes") of a certain function is the desired science. In fact, such scientific understanding is needed for deciding what new materials should be studied next as most promising novel candidates and for identifying interesting anomalies.

For many, maybe most, material functions, the "cause → property/function" relation is complex and indirect. Let us label the "cause" by a multi-dimensional descriptor d, which is initially unknown. The property/function is a number P (e.g. the thermoelectric figure of merit of a material), or a string of numbers. From statistical-learning theory, it is known that inverting the $d \to P$ mapping, i.e. identifying the cause from known data P, is an ill-posed problem, even when a one-to-one correspondence exists: A little error in the data P may suggest a very different cause d.

From the above-mentioned issues, the *4V & A,* and for first-principles computational materials science and engineering, the two key challenges concern big-data *veracity* and *analysis*. These are at the focus of this talk

Montag 19. Mai 2014, 17:30 Uhr
(ab 17:00 Uhr Kaffee)

Technische Universität Wien
Freihaus, Hörsaal 5 (2. Stock, grüner Bereich)
Wiedner Hauptstraße 8, 1040 Wien

wpk.univie.ac.at

TU WIEN

universität wien