GENERATING A FAIR CRYSTAL-STRUCTURE DATABASE WITH THE AiiDA INFORMATICS PLATFORM

Marnik Bercx[1,*], Flaviano José dos Santos[1,*], Nicola Marzari[1]

[1]Theory and Simulation of Materials (THEOS), and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
(*) marnik.bercx@epfl.ch, flaviano.dossantos@epfl.ch

Computer simulations that use powerful electronic-structure techniques are widely used to characterize or predict materials' properties. Such efforts rely on databases of measured or calculated data, with structural data being especially useful. Here, we develop and validate a set of protocols to generate a comprehensive structural database of 3D materials abiding to the FAIR data principles. We start from three major experimental structure databases: the Pauling file (MPDS), the inorganic crystal structure database (ICSD), and the crystallography open database (COD). Structures are refined with density-functional theory calculations using the open-source SIRIUS accelerated library together with Quantum ESPRESSO. Since calculations are driven by the AiiDA (http://aiida.net) materials' informatics infrastructure, all curated workflows, the entire provenance of the simulations and the resulting structural data can be shared openly on the Materials Cloud (http://materialscloud.org). We present our protocols and their validation, together with the use of AiiDA's advanced automation and error handling features to create robust workflows for electronic-structure simulations. As the conductivity character of a material is unknown a-priori, we employ smearing techniques for all materials. Smearing is widely used for metallic and magnetic systems, where they improve the accuracy of Brillouin zone sampling and lessen the impact of level-crossing instabilities. Advanced smearing techniques, such as Methfessel-Paxton and Cold smearing are constructed to make the system's total free energy temperature independent at least up to the third order. In doing so, these end up with non-monotonic occupation functions (and, for Methfessel-Paxton, not positive definite), which can result in the chemical potential not being uniquely defined. Thus, we propose a protocol combining different root-finding methods to implement a data-driven solution to determine the material's correct Fermi energy. We validate the method by calculating the Fermi energy of thousands of materials and comparing them with the results of previous approaches.

Keywords
High-throughput, 3-D materials, Fermi energy, smearing.

Funding
H2020 MaX CoE Grant no. 824143.
EU H2020-NMBP-TO-IND-2018 project "INTERSECT" Grant No. 814487.