Materials design and optimization for next-generation photovoltaics

Simone SALA – ENI S.p.A.

Computational tools ranging from HPC-scale simulations to machine learning models trained on experimental or synthetic databases have become widespread in materials science research. This talk will focus on applications to next-generation photovoltaic devices. Recently, an open-access and community-driven database containing over 40000 perovskite solar cells was published. This resource has enabled the application of predictive data-driven models to correlate device structure with photovoltaic performance, whereas the literature so far usually focused on specific device layers. In this context, we will discuss the potential of machine learning models for two main tasks: (i) to rationalize the experimental results encoded in the database and (ii) to predict the performance of hole transport materials not included in the training database. The main pitfalls and conceptual limitations of the approach will be discussed and correlated with the database structure and dimension, by comparing the performance of different choices of descriptors, regression algorithms and dataset size. When validated against experimental data available in the published literature, task (ii) was found to be particularly challenging, as the ability of the model to generalize to a new chemical space is limited by several factors. Our results are suggestive of the potential of in-silico simulation at an HPC scale to produce more comprehensive databases with more detailed descriptors that could replace or complement experimental data and enable effective materials discovery and optimization for such complex devices.