Developing a machine-learned band gap density functional to facilitate high-throughput searches for optical materials

Brian Kolb and Alexie M. Kolpak

Density functional theory (DFT) is built on the idea that the ground-state electronic charge density is a fundamental variable for an atomic system, meaning that properties of the system can be determined by functionals of this density. Approximations to the total energy functional, for example, have been extremely successful, and are the cornerstone of much materials research. Some key properties, such as the electronic band gap, are not well treated in a Kohn-Sham scheme, and require expensive higher-order methods such as GW or TD-DFT to compute accurately. Such difficulties can preclude high-throughput screening for these properties, or their computation in large complex systems such as heterostructures or systems with dilute defects. While, in principle, properties other than energy can be written as density functionals, this remains a largely unexplored possibility since the form of these functionals is completely unknown *a priori*. Progress can be made, however, by computing densities via DFT and band gap via the GW method for a set of materials and leveraging machine learning techniques to find a connection between these. Such a connection would allow the band gap to be predicted with accuracy nearing GW, but at the computational cost of DFT. Solving such a problem *(i.e.* finding predictive relationships within large and complex data sets) is the forte of machine learning approaches, which have already proven their utility in similar problems. In this work, we construct a training set of ≈1000 small non-metallic systems for which we compute the ground-state charge density within DFT and the electronic band gap within the GW approximation. We use an in-house developed code to train a neural-network-based density functional to predict the GW band gap from the ground-state density. The aim of this approach is to provide a cheap alternative to expensive GW calculations for use in high-throughput searches for optical materials.