Models incorporating molecular fragmentation, error cancellation and machine learning for the prediction of redox potentials

Sarah Maier, Eric M. Collins, and Krishnan Raghavachari

Indiana University, Bloomington, Indiana 47405, USA
e-mail: sarmaier@iu.edu

While accurate wavefunction theories like CCSD(T) are capable of modelling redox processes, steep computational scaling renders them intractable for treating large systems or extensive databases. In contrast, density functional theory (DFT) is more computationally feasible, yet often fails to quantitatively describe redox processes. Over the last decade our group has developed the Connectivity-Based Hierarchy (CBH), an error correction approach based on systematic molecular fragmentation protocols that offers significant improvements to DFT performance. Here we report an efficient delta machine learning (ΔML) model that builds on CBH and achieves coupled cluster accuracy by correcting for deficiencies in DFT. Emphasizing systematic improvements to low levels of theory via corrections to local molecular units, CBH represents a suitable foundation for improved molecular feature attribution in ΔML.

Our project design integrates concepts from molecular fragmentation, systematic error cancellation, and deep learning. First, we show that by using an electrostatic difference map, ionization sites within a molecule may be identified, and CBH correction schemes for ionization processes may be automated. As a central focus of our work, we employ our "FragGraph" QM/ML model, which embeds atom-centered, fragment-wise fingerprints taken from CBH into a computational graph to further increase accuracy in the prediction of ionization potentials. CBH fingerprints systematically capture portions of the molecular environment, making them a useful tool in molecular property prediction. Highlighting a new feature of our ΔML framework, we show that the incorporation of electronic descriptors from DFT improves model performance. In summary, we present a progression of methods aimed at systematically improving low-level redox calculations, and we assess their performance on a large test set of ionization potentials.