

## Admission Round 2019

Project Title	<b>Machine Learning meets Theoretical Chemistry: Data-driven Analysis of Grapheneoxide</b>
Project leads / supervisors	Dr. Annika Bande (HZB, Helmholtz) Prof. Dr. Ulf Leser (HU, ECDF)

### Project description

In recent years data science more and more finds its way into the field of materials sciences, usually in the form of regression or classification approaches which are trained on known properties of certain materials and applied to predict these properties for less well characterized yet physically and chemically similar other materials. However, a pertinent problem in this field is the search for materials which are optimal wrt some property, a problem which only partly matches to the above described scenario, since it is impossible to compute properties of the typically exponentially many candidate structures. For such a setting, the methods for calculating the properties of given structures must be complemented by intelligent algorithms which decide which structures are assessed in which order, and use the results of previous assessments to guide this search process through the space of all candidates.

In this project, we want to research such methods and apply them to a particular important problem in the development of Solar Fuels: The search for an optimal material for photocatalysis, namely water oxidation or even water splitting. The searched material class will be sustainable grapheneoxide quantum dots. The desired properties are: (1) a band edge of more than 1.23 eV that is to be centered around the energies for the water-to-oxygen oxidation and the water-to-hydrogen reduction, (2) a good representation of its diverse spectra (X-ray absorption and emission (XAS and XES), UV/vis, and infrared), (3) sizes of the quantum dots, and (4) their oxygen and eventually dopand content. Many of these properties can be computed numerically with standard quantum chemistry software as e.g. the ORCA program, but only at the cost of extremely resource-intensive calculations which currently limits the number of materials that can be characterized.

The PhD candidate shall study methods that allow a goal-driven search through the space of all candidate materials to find those that maximize a cost function derived from the materials properties. To this end, we plan to apply methods for reinforcement learning and heuristic search procedures, possibly combined with Deep Learning network to model correlations between the target properties and other, material-derived features. Particular challenges in this project are the inhomogeneity and incompleteness of data with respect to the target properties. The expected impact is high, because the project will create a complete workflow from high-throughput data production, database generation, and learning of structure-property relationships that can as well be applied to many other interesting materials and classes of properties in the material sciences.

### Prerequisites:

The PhD candidate shall study methods that allow a goal-driven search through the space of all candidate materials to find those that maximize a cost function derived from the materials properties. To this end, we plan to apply methods for reinforcement learning and heuristic search procedures, possibly combined with Deep Learning network to model correlations between the target properties and other, material-derived features. Particular challenges in this project are the inhomogeneity and incompleteness of data with respect to the target properties. The expected impact is high, because

the project will create a complete workflow from high-throughput data production, database generation, and learning of structure-property relationships that can as well be applied to many other interesting materials and classes of properties in the material sciences.

## References

J. Ren, F. Weber, F. Weigert, Y. Wang, S. Choudhury, J. Xiao, I. Lauer mann, U. Resch-Genger, A. Bande, T. Petit "Influence of surface chemistry on optical, chemical and electronic properties of blue luminescent carbon dots", *Nanoscale* (2019), DOI: 10.1039/C8NR08595A.

F. Weber, J. Ren, T. Petit, A. Bande "Theoretical X-ray absorption spectroscopy database analysis for oxidised 2D carbon nanomaterials", under review with *Phys. Chem. Chem. Phys.*

Bande, A., Bravic, I., Minnich, C.M., Puskar, L., Ritter, E., Schade, U., Aziz, E.F. "Isomeric Xylene Molecules in the Terahertz-Far Infrared Regime: Computational Chemistry and Spectral Modeling View", *Vib. Spectrosc.* 92, 220 (2017).

Habibi, M., Weber, L., Neves, M., Wiegandt, D. L. and Leser, U. "Deep Learning with Word Embeddings improves Biomedical Named Entity Recognition." special issue for ISMB/ECCB 2017; *Bioinformatics* 33(14): i37 (2017).