

CPP 44: Condensed-matter simulations augmented by advanced statistical methodologies (joint session DY/CPP)

Time: Wednesday 15:00–18:45

Location: H20

Invited Talk CPP 44.1 Wed 15:00 H20
Quantum Machine Learning — ●ANATOLE VON LILIENFELD —
 Institute of Physical Chemistry, University of Basel

Many of the most relevant chemical properties of matter depend explicitly on atomistic and electronic details, rendering a first principles approach to chemistry mandatory. Alas, even when using high-performance computers, brute force high-throughput screening of compounds is beyond any capacity for all but the simplest systems and properties due to the combinatorial nature of chemical space, i.e. all compositional, constitutional, and conformational isomers. Consequently, efficient exploration algorithms need to exploit all implicit redundancies present in chemical space. I will discuss recently developed statistical learning approaches for interpolating quantum mechanical observables in compositional and constitutional space. Results for our models indicate remarkable performance in terms of accuracy, speed, universality, and size scalability.

CPP 44.2 Wed 15:30 H20
Quantum Monte Carlo method for Vibrational Frequencies —
 ●YU YANG LIU and GARETH CONDUIT — University of Cambridge, UK

Quantum Monte Carlo methods have become a leading contender for high accuracy calculations for the electronic structure. Calculating energy derivatives such as atomic forces and the matrix of force constants is important in relaxing structures, calculating vibrational properties, and performing molecular dynamics simulations. We develop a quantum mechanical expectation value to evaluate the matrix of force constants directly in Quantum Monte Carlo. The approach allows the full modeling of correlation effects such as Van der Waals force, opening new applications to molecules and solids in condensed matter.

CPP 44.3 Wed 15:45 H20
Localized Basis Functions for Variationally Enhanced Sampling — ●BENJAMIN PAMPEL, KURT KREMER, and OMAR VALSSON — Max Planck Institute for Polymer Research, Mainz, Germany

Variationally Enhanced Sampling (VES) is a recently developed method for molecular dynamics simulations. It enhances sampling by introducing a bias potential along certain collective variables that is constructed via minimisation of a convex functional.

This bias potential is usually represented by a linear expansion in some basis set, with delocalised functions such as plane waves or Chebyshev/Legendre polynomials as common choices. However, it is an open question if localised functions perform better. In particular, the wavelet family of Daubechies might be a good choice. These functions offer the favourable property of forming an orthonormal basis with a tunable number of vanishing moments. Furthermore, their intrinsic principle of multiresolution allows increasing the precision of the bias representation at specific points of interest.

We have implemented Daubechies wavelets into the VES code and have tested their performance in various systems. As a direct comparison of the different basis sets is difficult, we have developed a new measure of the error of free energy calculations. The Daubechies wavelets are observed to perform better than both Chebyshev/Legendre polynomials and Gaussian basis functions, resulting in faster convergence and yielding more accurate free energy surfaces without increases in computational cost.

CPP 44.4 Wed 16:00 H20
Representing molecules and materials for accurate interpolation of quantum-mechanical calculations — ●MARCEL LANGER, ALEX GOESSMANN, and MATTHIAS RUPP — Fritz Haber Institute of the Max Planck Society, Berlin, Germany

The search for novel materials, the exploration of phase diagrams and other high-throughput applications require numerical simulations of molecules and materials from first principles, but are limited by their high computational cost. By interpolating between reference calculations, machine learning can act as a fast accurate surrogate for these calculations, greatly increasing the number of accessible systems. [1] This requires a *representation* of molecules and materials suitable for interpolation. We show how the state-of-the-art representations can

be understood in a unified framework [2] based on local descriptions of atomic environments via k -body functions, group averaging and tensor products, and discuss implications. We benchmark predictive accuracy of selected representations by carefully controlling for all other factors, including data distribution, regression method and optimization of free parameters. For the latter, we employ a consistent and fully automatic procedure to optimize both numerical and categorical free parameters, such as the choice of k -body functions, using sequential model-based optimization with tree-structured Parzen estimators. [3]

References: [1] a) Rupp et al, Phys Rev Lett **108**, 058301, 2012. b) Rupp, Int J Quant Chem **115**, 1058, 2015. [2] Willatt et al, Phys Chem Chem Phys, accepted, 2018. [3] a) Bergstra et al, NIPS **24**, 2546, 2011; b) Bergstra et al, ICML **30**, I-115, 2013.

CPP 44.5 Wed 16:15 H20
Machine Learning of Free Energies — ●CLEMENS RAUER and TRISTAN BERAU — Max Planck Institute for Polymer Science, Ackermannweg 10, 55128 Mainz, Germany

Free energies are important molecular properties which can provide an insight into the thermodynamic state of the respective system. Accurate calculations of free energies are an important tool for many biophysical applications, ranging from protein-ligand binding[1] to the insertion of small molecules into a lipid[2]. However, computationally expensive high level simulations are necessary in order to obtain accurate free energy estimates, and therefore, only a small subset of chemical space can be accurately covered. We overcome this problem by building a Δ -machine learning[3] model. Using this approach we can use a "cheap" low level method to predict free energies and learn the correction to a higher level method or experimental value. Then, we can predict high level free energies for significantly larger compound sets than was used in the training of the model. We show that by using only limited high level data, highly accurate free energies can be calculated using this method. As a first system we apply this approach to the prediction of hydration free energies.

[1] Mobley, D.L. & Gilson, M.K. *Annu. Rev. Biophys.* **2017**, 46:531-58

[2] Menichetti, R. et al. *J. Chem. Phys.* **2017**, 147, 125101

[3] Ramakrishnan et al. *J. Chem. Theory Comput.* **2015**, 11, 2087-2096

CPP 44.6 Wed 16:30 H20
Deep Learning for Multiscale Simulations of Soft Matter Systems — ●MARC STIEFFENHOFER¹, TRISTAN BERAU¹, and MICHAEL WAND² — ¹Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz — ²Institute of Computer Science, Johannes Gutenberg University Mainz

The great success of Deep Neural Networks (DNNs) is based on their ability to learn and extract descriptive features directly from training data and to build a hierarchical, abstract representation of the input that takes multiple length scales into account. Such multiscale representations can also be found in soft matter systems where many physical phenomena and properties are governed by a large range of different length- and timescales.

In this work, we explore links between multiscale representations of DNNs and multiscale simulations of soft matter systems. The main focus is to investigate if DNNs can be used to link distribution functions generated at different resolutions. We have applied DNNs to the backmapping of coarse-grained molecular configurations to higher-resolution representations. This requires to reproduce the fine-grained statistics that match the coarse-grained representation.

Our model relies on 3D convolutional neural networks that are trained to generate molecular equilibrium structures. The training of the model is based on the generative adversarial approach and results are discussed for a system of octane molecules.

CPP 44.7 Wed 16:45 H20
Unsupervised machine learning of chemical compound space for hierarchical screening and coarse-graining applications — ●KIRAN KANEKAL, KURT KREMER, and TRISTAN BERAU — Max Planck Institute for Polymer Research, Mainz, Germany

Recently, the number and type of chemical fingerprints used for super-

vised learning of molecular properties has increased significantly, as the importance of properly encoding the 3D structure of a molecule to greatly increase accuracy has been recognized. Some of the most successful of these fingerprints encode a great deal of physical information, including coefficients for potential energy functions commonly used in classical atomistic molecular dynamics simulations [1]. In this work, we incorporate these fingerprints into an unsupervised machine learning (clustering) scheme to define subspaces in the chemical compound space, for which we use the Generated DataBase [2] (GDB) as a proxy. We show that the use of different molecular fingerprints leads to significant differences in the clustering observed, as each fingerprint will highlight specific molecular properties. Therefore, unsupervised learning, when coupled with these fingerprints, naturally enables hierarchical screening approaches for materials design. Furthermore, the presence of strong correlations between clusters identified using different fingerprints implies that a lowering of resolution (i.e. coarse-graining) is viable for that specific region of chemical compound space.

1. B. Huang and O. A. von Lilienfeld, *J. Chem. Phys.* 2016, 145, 161102. 2. T. Fink and Jean-Louis Reymond, *J. Chem. Inf. Model.* 2007, 47, 342.

15 min. break

CPP 44.8 Wed 17:15 H20

Understanding three-body contributions to coarse-grained force fields — ●CHRISTOPH SCHERER, RENE SCHEID, TRISTAN BERAU, and DENIS ANDRIENKO — Max-Planck-Institut für Polymerforschung, Mainz, Germany

Coarse-graining (CG) is a systematic reduction of the number of degrees of freedom (DOF) used to describe a system of interest. CG can be thought of as a projection on the CG DOF and is therefore dependent on the number and type of CG basis functions. We present an extension of the two-body basis set with three-body basis functions [1]. The CG scheme is implemented in the VOTCA-CSG toolkit [2]. We show that naive extensions of the CG force field can result in substantial changes of the two-body interactions making them much more attractive. This is related to the three-body basis functions of the Stillinger-Weber type having a significant two-body component. This interference can be alleviated by CG the two- and three-body contributions separately. Furthermore, we employ Kernel-based Machine learning (ML) [3] overcoming the restrictions of a fixed basis set. The approach is illustrated on liquid water where three-body interactions are essential to reproduce the structural properties, and liquid methanol where two-body interactions are sufficient to reproduce the main features of the atomistic system. [1] Scherer, Andrienko, *PCCP*, 20, 22387 (2018); [2] Rühle, Junghans, Lukyanov, Kremer, Andrienko, *JCTC*, 5, 3211 (2009); [3] Glielmo, Sollich, De Vita, *PRB*, 95, 214302 (2017)

CPP 44.9 Wed 17:30 H20

Reweighting Dynamics of Nonequilibrium Steady States — ●MARIUS BAUSE, TIMON WITTENSTEIN, KURT KREMER, and TRISTAN BERAU — Max Planck Institute for Polymer Research, Ackermannweg 10, 55128 Mainz

Markov State Models (MSM) are a discrete representation of the kinetics of a given system constructed by coarse-graining microtrajectories. While frequently applied to equilibrium systems, a protocol for non-equilibrium steady state systems (NESS) has not been developed due to loss of dynamic properties like detailed balance. We propose to apply the principle of Maximum Caliber by Jayne's, postulating that the distribution of paths is given by the maximal path entropy. The Markovian assumption alleviates the combinatorial explosion of microtrajectories. The entropy production between states is chosen for a physical constraint and defines the NESS ensemble. By choosing a reference system for a prior, reweighting between ensembles becomes possible. The entropy productions become the defining elements of the ensemble. The method is tested on a minimal model under non-conservative forces.

CPP 44.10 Wed 17:45 H20

Free Energy Landscape of Phase Transitions Investigated by Enhanced Sampling over Degenerate Collective Variables — ●BIN SONG, KURT KREMER, and OMAR VALSSON — Max Planck Institute for Polymer Research, Mainz, Germany

Phase transition events such as nucleation of atomic or molecular crystals, self-assembly of micelles etc. have been intriguing scientists not

only for their underlying physical principles, but also for their potential applications. Molecular dynamic simulations are increasingly called upon to either provide the detailed mechanistic insights and to prognosticate such events that may be too expensive or slow to examine with experiments. As potential energy surface of such events are often complex and difficult to sample, free energy surface (FES) are preferred. However, FES presents its own challenge to sample when the height of barriers are higher than or comparable to the thermal energy. Enhanced sampling methods biasing on properly chosen collective variables (CVs) can help overcome these barriers. Sometimes globally defined CVs are not sufficient, and more than a few of locally defined ones are needed. The increased dimensionality of the bias potential could be another source of frustration for achieving properly sampled FES, which we would like to alleviate with our new development of Variational Enhanced Sampling (VES) in cases where the local CVs are permutationally invariant. We demonstrate the capability of the new development by sampling the phase transitions in systems of increased complexity, including Lennard-Jones clusters and liquid-solid transition in materials.

CPP 44.11 Wed 18:00 H20

Efficient Equilibration of Hard Particles with Collective Moves – A Comparison of Computational Methods — ●MARCO KLEMENT and MICHAEL ENGEL — Institute for Multiscale Simulation, Friedrich-Alexander-Universität Erlangen, Germany

Particle simulations are an important method to study the phase behavior of fluids and solids. A common task is structure prediction via thermal equilibration. Examples are crystallization or melting and the aging of glasses. Near the liquid-solid phase transition of hard spheres event driven molecular dynamics is known as most efficient. Monte Carlo simulations are boosted by event-chain in the same region. It is neither clear how equilibration does benefit from Newtonian motion nor how the efficiency of an algorithm should be quantified; and several ways are in use. Here we present further improvement of event-chain Monte Carlo by bringing the chain closer to natural movement with particle reflection events and quantify the improvement using different approaches from literature.

CPP 44.12 Wed 18:15 H20

The temperature dependence of the mechanical unfolding of a supramolecular complex studied by molecular simulations — TAKASHI KATO, KEN SCHÄFER, STEFAN JASCHONEK, and ●GREGOR DIEZEMANN — Institut für Physikalische Chemie, Universität Mainz

The conformational dynamics of supramolecular complexes can be studied on a single molecule level using the techniques of dynamic force spectroscopy and important kinetic information can be extracted. Using models for the free energy landscape of the system as a function of the molecular extension it is possible to obtain parameters like the bare unfolding rate, the activation free energy in the force-free case and also the distance between the folded configuration and the transition state. In the present work we performed molecular simulations of the mechanical unfolding of a model molecular complex, a pair of interlocked calixarene catenanes. From rupture force distributions the kinetic rates for the opening transition were extracted for varying parameters of the pulling device. In order to be able to perform a model-free analysis we performed simulations over a broad range of temperatures and found Arrhenius behavior for the kinetic rates. This allows to determine the activation free energy and the bare rate independently as a function of the force without using specific models for the shape of the energy landscape. We discuss our findings in light of the models that are usually used in the analysis of force dependent kinetic rates obtained from experimental or simulation data.

CPP 44.13 Wed 18:30 H20

An efficient anharmonic free energy method applied to vacancies in ZrC — ●THOMAS MELLAN¹, ANDREW DUFF², BLAZEJ GRABOWSKI³, and MICHAEL FINNIS¹ — ¹Imperial College, London, UK — ²Hartree Centre, Daresbury, UK — ³Max-Planck-Institut für Eisenforschung GmbH, Düsseldorf, Germany

We have developed a thermodynamic integration approach to calculate the anharmonic contribution to the volume and temperature dependent free energy of a crystal. The method provides an effective balance between accuracy and computational efficiency, giving a factor of 15 speed-up on comparable free energy approaches with average errors less than 1 meV/atom. The method is demonstrated with new predictions on the thermodynamics of substoichiometric ZrC, including vacancy concentration and heat capacity.