SCISPACE
formerly Typeset

Open access • Journal Article • DOI:10.1002/WCMS.1159

cp2k: atomistic simulations of condensed matter systems — Source link

Juerg Hutter, Marcella Iannuzzi, Florian Schiffmann, Joost VandeVondele

Institutions: University of Zurich, ETH Zurich

Published on: 01 Jan 2014 - Wiley Interdisciplinary Reviews: Computational Molecular Science (Wiley-Blackwell)

Topics: Massively parallel

Related papers:

- Separable dual-space Gaussian pseudopotentials
- Generalized Gradient Approximation Made Simple
- QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach
- Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases.
- A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu

Share this paper:

View more about this paper here: https://typeset.io/papers/cp2k-atomistic-simulations-of-condensed-matter-systems-12mcj86s5y

Zurich Open Repository and Archive
University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2014

CP2K: Atomistic simulations of condensed matter systems

Hutter, Juerg ; Iannuzzi, Marcella ; Schiffmann, Florian ; VandeVondele, Joost

Abstract: cp2k has become a versatile open-source tool for the simulation of complex systems on the nanometer scale. It allows for sampling and exploring potential energy surfaces that can be computed using a variety of empirical and first principles models. Excellent performance for electronic structure calculations is achieved using novel algorithms implemented for modern and massively parallel hardware. This review briefly summarizes the main capabilities and illustrates with recent applications the science cp2k has enabled in the field of atomistic simulation.

DOI: https://doi.org/10.1002/wcms.1159

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: https://doi.org/10.5167/uzh-89998
Journal Article
Accepted Version

Originally published at:
Hutter, Juerg; Iannuzzi, Marcella; Schiffmann, Florian; VandeVondele, Joost (2014). CP2K: Atomistic simulations of condensed matter systems. Wiley Interdisciplinary Reviews. Computational Molecular Science, 4(1):15-25.
DOI: https://doi.org/10.1002/wcms.1159

CP2K: Atomistic Simulations of Condensed Matter Systems

Jürg Hutter and Marcella Iannuzzi
Physical Chemistry Institute
University of Zurich
Winterthurerstrasse 190
CH-8057 Zurich, Switzerland

Florian Schiffmann and Joost VandeVondele
Nanoscale Simulations
ETH Zurich
Wolfgang-Pauli-Strasse 27
CH-8093 Zurich, Switzerland

December 17, 2012

Abstract

CP2K has become a versatile open source tool for the simulation of complex systems on the nanometer scale. It allows for sampling and exploring potential energy surfaces that can be computed using a variety of empirical and first principles models. Excellent performance for electronic structure calculations is achieved using novel algorithms implemented for modern and massively parallel hardware. This review briefly summarizes the main capabilities and illustrates with recent applications the science CP2K has enabled in the field of atomistic simulation.

1 Introduction

Computer simulation of matter with atomistic detail has become a very prominent tool in chemistry, physics, life sciences and materials sciences. In these fields, simulation results can yield the insights needed to interpret experimental measurements, can be used to predict material properties, or to design new compounds. A precise picture of the structures and dynamical processes at the atomic scale is a valuable starting point to rationally design new experiments and new systems. With sustained exponential growth in computer resources, the impact of simulation will continue to increase.

The CP2K computer program is a powerful tool for atomistic simulation. [1] It aims at providing a broad range of models and simulation methodologies, suitable for large and condensed phase systems, and is able to exploit the most advanced computer hardware. CP2K has a large impact in the field of density functional theory (DFT) based molecular dynamics (MD) simulation, particularly with its capability to describe the dynamics of systems containing hundreds of atoms with relative ease, but has a broader range of capabilities. An important aspect in this respect is the possibility of choosing the adequate method for a given problem and the flexibility of combining methods. The versatility of CP2K appears evident from the numerous applications, also in combination with experiment, which produced high level scientific work. For example Treier et al. [2] investigated the atomically precise fabrication of tailored nanographenes from polyphenylene precursors through cyclodehydrogenation of a prototypical polyphenylene on Cu(111). In this study scanning tunneling microscopy (STM) experiments were complemented by CP2K-based large-scale ab initio simulations, which allowed the atomistic details of the surface-assisted cyclodehydrogenation reaction to be established. The simulations gave insight into the copper-promoted reaction pathway, shedding light on the thermally activated intramolecular aryl-aryl coupling of the cyclic polyphenylene, the role of dispersion forces, and the influence of the substrate.

Nowadays, massively parallel software is key to benefit from the advances in computer hardware, and the development in CP2K aims at combining efficient algorithms with excellent parallel scalability. Staying at the forefront of atomistic simulation is a tremendous task that requires a significant investment. CP2K leverages the contributions and support from an active community within an open source development model.

In the following sections we provide an overview of the main functionality of CP2K. First, various techniques to sample and explore potential energy surfaces are discussed. Subsequently, we provide more detail about the options that are available to compute the potential energy of a system, with an emphasis on DFT. In a next section, computational and software design aspects are discussed. Finally, we highlight selected applications to solar cells, water interfaces, and functionalized metallic surfaces.

2 Exploring the configuration space

Key in many simulations performed with CP2K is the exploration of the potential and free energy surfaces of complex systems. As system size increases, the standard procedures to locate stationary points become less useful, and need to be complemented with methods aimed at sampling and exploring, including kinetic and entropic effects. This is particularly clear for systems that must be described at a finite temperature, such as liquids, where most quantities of interest can only be obtained from statistical averages and time correlation functions. Free energy differences or non-harmonic vibrational spectra are two prominent examples. CP2K provides a variety of simulation methods to explore the potential energy surface, ranging from methods to characterize stationary points to procedures to enhance the sampling. Some methods only require the calculation of the potential energy, but most require the computation of forces as well. Usually, these methods are oblivious to the nature of the underlying potential energy surface, working equally well with first principles and (semi)empirical approaches, or combinations of those.

2.1 Stationary points

CP2K provides methods to optimize atomic positions and cell vectors based on various standard algorithms, including cubically scaling methods that require an explicit inversion of an approximate Hessian as well as linear scaling approaches that are suitable for large systems. Full vibrational analysis based on finite differences is possible for smaller systems. Iterative diagonalization of the Hessian for selected modes[3] is an effective method to study the vibrational spectroscopy of molecules at surfaces [4] or other large systems where only part of the spectrum is of interest. This has been exploited in the simulation of IR spectra of dye molecules at TiO_2 surfaces, see section 5.1 for more information. The nudged elastic band (NEB) and string methods [5] allow for simultaneous optimizations of reaction paths and the location of transition states. The latter methods require multiple calculations on weakly coupled systems and can be efficiently carried out in parallel, yielding good time to solution if large computer resources are available. Constrained geometry optimizations and NEB calculations have been combined [2], for example, to derive a simple mechanism for the complete dehydrogenation of a cyclic polyphenylene adsorbed on Cu(111). By large ab-initio calculations it was shown that the process is catalytically activated and enhanced by adsorbate-substrate van der Waals interactions.

2.2 Molecular Dynamics

Molecular dynamics is one of the most commonly employed simulation methods within CP2K, and with DFT based forces is feasible for systems containing hundreds of atoms. The default integrator of the equations of motion is based on the velocity Verlet algorithm [6], which allows for very useful extentions of the MD scheme, as for example introducing constraints or extended Lagrangian formalisms. Standard methods are available for the most common ensembles (NVE, NVT, NPT) [6, 7], and require the calculation of forces and stresses. Geometrical constraints and restraints are available in all ensembles, allowing for free energy calculations by integration of the Lagrangian multipliers. Thermostats and barostats are of great importance for equilibration and the efficient and correct sampling of different ensembles in MD. Available variations of thermostats include an adaptive Langevin form [8], generalize Langevin equation with colored noise (GLE) [9], stochastic velocity rescaling (CSVR) [10], and the more traditional thermostats based on the work by Nosé and others [11, 7] Different thermostats can be distributed over different regions of the system, which guarantees an optimal control of the temperature and faster equilibration, or can be used with different temperatures to improve sampling [12]. MD simulations with the GLE thermostat have been used for example by Nagata et al. [13] to explore nuclear quantum effects on the water structure near the liquid-vapor interface. The path-integral MD performed at different temperatures for a box of 80 water molecules (H_2O) and deuterated molecules (HDO, D_2O) shows that the broken symmetry of the hydrogen-bond interactions in HDO gives rise to distinct OH and OD bond orientations at the interface. The integration time step is by default fixed, and should be taken such as to guarantee the stability of the integration algorithm. For problematic systems, or hot particles, a variable time step can be adopted, which is tuned on the largest allowed atomic displacement. More advanced MD techniques include the simulation of shock waves using the multi scale shock method by Reed et al. [14], multiple time-scale algorithms (RESPA [15]) in which different potential descriptions are employed to increase MD efficiency [16], and path integral MD to study quantum effects [17, 18].

For the combination of MD with electronic structure methods, Born–Oppenheimer (BO) MD is employed [19, 20]. One advantage of MD is that atomic configurations are generated in a continuous fashion and one can hence predict an initial trial wave function in SCF calculations by multi-linear extrapolation. An effective extrapolation technique is based on the propagation of the density matrix $\mathbf{P}$, more precisely its contra-covariant representation $\mathbf{PS}$, where $\mathbf{S}$ is the overlap matrix. Even for low extrapolation orders, the number of SCF iterations at each MD step can be reduced considerably. Long term stability of the BO MD is obtained by using an approximately time reversible extrapolation scheme for the electronic degrees of freedom [20] based on the ASPC method (always stable predictor corrector) by Kolafa [21]. The obvious advantage of being able to run long and stable MD simulations is that structural and dynamic properties extracted from the sampling, such as pair correlation functions and mean square displacement, are better converged and meaningful. As an alternative to BOMD, a method based on Langevin dynamics [22] can be used. On the other hand, the original Car-Parrinello method is not implemented.

In order to study processes dominated by rare events, like reactions and structural transformations, MD is typically not sufficient due to too slow sampling of the configurations space. Constraints and restraints can be employed to reconstruct the free energy profile along specific reaction pathways by applying methods like thermodynamic integration or umbrella sampling. Metadynamics is a powerful method to accelerate the sampling and reconstruct the free energy surfaces in terms of a few collective variables [23, 24]. In CP2K different versions of metadynamics are implemented, such as extended Lagrangian metadynamics, well-tempered metadynamics, and the multi-walker scheme. The code offers a large choice of possible collective variables and it is rather easy to define new variables as combinations of the available ones. Applications of the method in combination with classical force-fields, DFT, or QM/MM have been published in different fields of molecular simulations. For instance, Michel et al. [25] used DFT-MD augmented by metadynamics to investigate C–C and C–H reductive eliminations from octahedral diphosphine Pt(IV) complexes. In this study different competitive pathways have been disclosed and the computed free energy barriers could explain the kinetics of the system. The self-repair mechanism of photoexcited DNA, consisting of the splitting dynamics of the thymine dimer, has been studied using a QM/MM scheme [26, 27]. The analysis of several statistically independent runs reveals an asynchronously concerted mechanism and identifies the dimer splitting in DNA as an ultrafast reaction. The relevance of entropic effects on fluxional conformational changes of small gold clusters has been highlighted by metadynamics simulations, showing that at thermodynamic equilibrium a large variety of planar/quasi-planar and tridimensional structures with non-intuitive topology coexist. [28]. Applications in other fields as diffusion and reaction at solid interfaces[29, 30], materials science [31], catalysis and more, are also good examples of the versatility of the technique as implemented in the framework of CP2K.

2.3 Monte Carlo

Monte Carlo (MC) simulations [6] are an alternative to MD simulations if sampling is the main goal. One advantage of MC is that this approach is feasible even if forces are not available or expensive to compute. However, MC is usually difficult to perform efficiently in a general and system independent manner. System specific configurational changes (moves) can speedup sampling significantly and allow for simulations that would be inefficient with standard MD procedures, for example establishing the equilibrium between two different phases of a system. In recent years, methodologies to perform NPT and grand canonical Monte Carlo simulations of molecular liquids have been implemented into the CP2K code [32, 33, 34]. These techniques have been used to simulate the vapor-liquid coexistence curve of water [35] and methanol [36]. They were also used extensively in the investigation of the properties of water interfaces (see section 5.2).

2.4 Ehrenfest Dynamics

In order to study the time dependent evolution of electrons in a system, a useful approximation is the assumption that nuclei can be described classically.[37] This leads to the Ehrenfest system of equations that describe the non-adiabatic evolution of a coupled nuclear and electronic system. Within time-dependent Kohn–Sham density functional theory, a generalized action principle leads to consistent and energy conserving equations of motion also in the case of atom-centered basis functions [38]

$$\dot{c}_{\alpha j} = -\sum S^{-1}_{\alpha\beta} \left(iH_{\beta\gamma} + B_{\beta\gamma} \right) c_{\gamma j} \tag{1}$$

$$M_A \ddot{R}_A = -\frac{\partial U(R,t)}{\partial R_A} + \sum_j \sum_{\alpha\beta} c^\star_{\alpha j} \left(\frac{\partial H_{\alpha\beta}}{\partial R_A} + D^A_{\alpha\beta} \right) c_{\beta j} \tag{2}$$

where $c_{\alpha j}$ are the orbital expansion coefficients, $S_{\alpha\beta} = \langle \Phi_\alpha \mid \Phi_\beta \rangle$ and $B_{\alpha\beta} = \langle \Phi_\alpha \mid \frac{d}{dt} \Phi_\beta \rangle$ the basis function overlap and its time derivative, and H the Kohn–Sham matrix. $U(R,t)$ is a possible time-dependent external potential and the term $D^A_{\alpha\beta}$ collects all terms from the explicit position dependence of the basis functions. Methods to propagate the first order differential equations for the orbital coefficients have been investigated extensively [39]. Methods based on matrix exponentials calculated by diagonalization or Padé approximants, as well as Arnoldi subspace iteration methods are implemented in CP2K. Together with the application of time-dependent electrical fields, these methods can be used to study the time-dependent response of the electronic structure to perturbations within periodic boundary conditions. One recent application investigated the photo-induced electron transfer rate between a gold atom and a gold cation solvated in CCl_4 [40].

3 Energy and Force Methods

CP2K aims to provide a wide range of potential energy descriptions, ranging from empirical approaches such as classical force fields to methods based on quantum mechanics, in particular DFT. Furthermore, various descriptions can be combined arbitrarily at the level of the input, so that various combinations of methods are naturally available. Examples include embedding approaches that rely on a potential energy surface that combines two different methods

$$E = E_a[\{A + B\}] - E_a[\{B\}] + E_b[\{B\}] \ , \tag{3}$$

linear combinations of potentials as needed for free energy calculations

$$E_\lambda = \lambda E_a + (1 - \lambda) E_b \ , \tag{4}$$

or propagation on the lowest potential energy surface, for example for a system with various spin states

$$E = \mathbf{MIN}\,\{E_a[\{A\}], E_b[\{A\}], \ldots\} \ . \tag{5}$$

In the following we give a brief description of some of the available methods.

3.1 Classical Force Fields

CP2K is compatible with standard biomolecular force fields (CHARMM, AMBER, GROMOS) and can read the corresponding topology files. On the input level, a formula parser allows for specifying general functional forms for the non-bonded pair potential forces, so that most proposals in literature can be used without coding. More specialized force fields, used commonly for materials, include the embedded atom models (EAM) [41]. Polarizable force fields using shell models [42] or multipole Ewald [43] methods are available. The force field implementations in CP2K cannot compete in efficiency with specialized codes for the simulations of biomolecules or metallic materials. However, its flexibility is of use for non-standard systems and for QM/MM type simulation where it provides full control to the user within a single code. The QM/MM coupling implemented in CP2K supports full electrostatic coupling between the QM charge distribution and the classical point charges using a highly efficient multi grid technique [44]. It has been used for example in the study of thymine dimer radical anion splitting in the self-repair process of duplex DNA [27, 26] or in order to describe metal-supported cyclohexaphenylene dehydrogenation [29]. In the later application a combination of DFT and an EAM potential for the description of the metal support has been used.

3.2 Electronic Structure Methods

Electronic structure methods can differ significantly in performance and computational cost. In CP2K, the computationally most expedient methods are semi-empirical (SE) approaches based on the NDDO Hamiltonian (AM1, MNDO, MNDO/d, PM3, PM6) [45, 46] as well as density functional based tight-binding methods (DFTB) [47]. These methods have been parametrised for small and short range basis sets, particularly suitable for linear scaling applications, and feature a very efficient Hamiltonian matrix construction.

The majority of the applications with CP2K are based on Kohn–Sham (KS) based density functional theory [48] in the generalized gradient approximation (GGA). An extended set of functionals is available through an external library (libxc) [49] or internal routines. Hybrid functionals for periodic systems are accessible for all type of simulations [16, 50]. Normally, hybrid functionals are computationally significantly more demanding than GGA calculations, but the auxiliary density matrix method (ADMM) [51] is an approximation that enables hybrid calculations at comparable cost. Recently, post-Hartree–Fock type calculations such as double hybrids and MP2 have been implemented [52], and with resolution-of-identity methods, MP2 and RPA energy expressions [53, 54] can be efficiently calculated for periodic systems. The KS orbitals can be optimized using the orbital transformation (OT) method [19, 55] for systems with a band gap, or using standard diagonalization or iterative diagonalization schemes for metals. The OT method is highly robust and efficient, but nevertheless scales cubically with systems size, dominating GGA calculations for systems containing a few hundred or more atoms. For metallic systems the smearing of the occupation number of the KS orbitals and the mixing of the electronic density in the reciprocal space are required to optimize the wave functions. When these measures are adopted, the quality of the description of the electronic structure is guaranteed also for metallic systems [56, 57, 58, 59]. Moreover, efficient parallel mathematical libraries for the diagonalization of large matrices, like ELPA [60], allows to carry out such calculations for systems containing thousands of atoms [61].

In CP2K, linear scaling algorithms based on sparse matrix algebra allow for SE and GGA calculations on systems with a band gap containing millions of atoms [62]. The key ingredients for the linear scaling calculation of the KS matrix are the use of a local basis, which is constructed from contracted spherical Gaussian functions, and an efficient technique to obtain the classical Coulomb and exchange-correlation energy, namely and the Gaussian and plane waves (GPW) scheme [63]. The advantage of the Gaussian basis is the analytic nature of all one-electron integrals, while the GPW scheme avoids the need for two-electron four center integrals in GGA calculations and gives direct access to integrals over molecular orbitals for post-Hartree–Fock methods [52]. The GPW scheme can be thought of as an auxiliary basis method, which represents the density in a basis of plane waves, or equivalently, on an equidistant grid in real space.

$$\rho(r) = \sum_{\alpha\beta} P_{\alpha\beta} \chi_\alpha(r) \chi_\beta(r) = \sum_G c_G e^{iG \cdot r} \quad , \tag{6}$$

where $P_{\alpha\beta}$ is the density matrix in the Gaussian basis with basis functions $\chi_\alpha(r)$, and G denotes the lattice vectors in reciprocal space. The calculation of the plane wave expansion coefficients c_G proceeds via the determination of the density on the real space grid, for which screening algorithms and multigrid techniques are used to make this step efficient and linear scaling in system size [19]. Plane waves are thus not part of the primary basis, but are an auxiliary representation of the density that allows for the use of Fourier transform based techniques, standard in plane wave codes [37], to compute the density dependent energies and potentials. In CP2K, matrix elements of these potentials are calculated efficiently and in linear scaling time via numerical integration over their real-space grid representation. Plane waves allow naturally for periodic calculations and calculations with reduced periodicity, such as surfaces and gas phase molecules, are possible using advanced Poisson solvers [64, 65]. In CP2K, periodic calculations are currently only possible using a single k-point (Gamma point), which is an approximation that becomes accurate for sufficiently large unit cells.

The GPW method requires a representation of the electron density on an equidistant grid, and is thus not practical to describe core electrons. A smooth density is obtained by using dual-space pseudopotentials [66, 67, 68] that are analytically represented by Gaussian functions, fully separable in the nonlocal part, and energy optimized. A corresponding series of basis sets covering the non-Lanthanoids up to Rn has been globally optimized on molecules (MOLOPT) using a generally contracted form with a single set of exponents for all angular momenta [69]. These MOLOPT basis set can include diffuse functions yet maintain a good condition number of the overlap matrix, an important property for accurate and efficient calculation of large condensed phase systems. An all-electron extension of the GPW method (GAPW) has been developed [70, 71] building on techniques proposed by Blöchl [72] for his projector augmented-wave (PAW) method. In this way, condensed phase all-electron calculations are possible for properties where core electrons are of importance, namely, X-ray absorption spectra [73], NMR chemical shieldings [74], and g-tensor calculations [75].

4 Computational Aspects

Academic software design at the scale of a simulation program like CP2K is an interesting challenge. With an average growth of 200 lines of code per day, the CP2K code is currently approximately 800'000 lines of code. An open code repository allows the community instant access to all versions of the code, which is available under the Gnu Public License (GPL), and to follow development. Several tens of authors have contributed to the code, with a smaller team of around twenty people from five institutions having write access to the repository. Most of CP2K is written in Fortran95, with elements from Fortran03 and extensions such as OpenMP and CUDA C. It employs various external libraries. In addition to bringing new features, the use of external libraries decreases the complexity of CP2K and enhances the efficiency and robustness of the code. The libraries range from basic functionality such as message passing (MPI) over dense linear algebra (BLAS, LAPACK, ScaLAPACK), Fast Fourier Transforms (FFTW) to more specialized chemical libraries such as electron repulsion integrals (libint) and exchange correlation functionals (libxc). CP2K itself can be built as a library, allowing for easy access to some part of the functionality by external programs. Increasingly, CP2K itself will be structured as independent libraries to facilitate sharing with and contributions from various communities. Currently, the most prominent example is a library for the handling of sparse matrices (DBCSR) that provides a highly efficient and scalable implementation of a Cannon based sparse matrix matrix multiplication. Having lean, library-like interfaces within CP2K has facilitated the implementation of features such as farming (running various inputs within a single job), general input parameter optimization, path-integral MD, or Gibbs ensemble MC. Rapid development and refactoring is facilitated by a growing set of test inputs, currently well over two thousand, which is checked for consistency during development and automatically at every commit to the repository.

Good performance and parallel scalability are key features of CP2K. This is achieved using a multi-layer structure of specifically designed parallel algorithms. On the highest level, parallel algorithms are based on message passing with the MPI and suitable for distributed memory architectures. This level is important, and requires careful design of data structures and algorithms. Increasingly, MPI level parallelism has been augmented with shared memory parallelism based on threading and programmed using OpenMP directives. This combination becomes more and more important as the core count per node increases, and top-level computers feature 100'000 and more cores. Ongoing work aims at porting the main algorithms of CP2K to accelerators and GPUs, as these energy efficient devices become more standard in supercomputers. At the lowest level, auto-generated and auto-tuned code allows for generating CPU-specific libraries that deliver good performance without a need for dedicated code development.

5 Illustrative Applications

5.1 Dye Sensitized Solar Cells (DSSC)

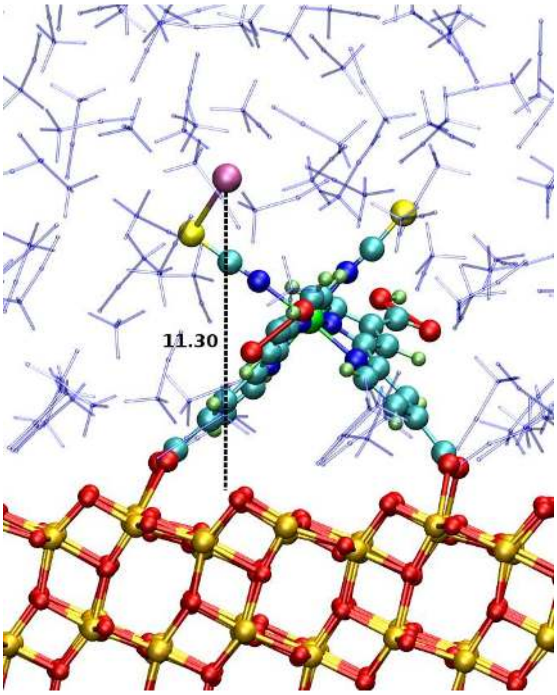

Figure 1: Snapshot of a DFT based simulation of the an intermediate dye-iodide complex attached to the TiO_2 surface in explicit solution. The iodide-surface distance coincides with the maximum concentration of ions found in classical molecular dynamics simulations of the electrolyte near surface.

One application demonstrating the potential of CP2K for simulating complex systems is a study of the active interface in dye sensitized solar cells. In these devices the redox active region consists of a dye attached to a semiconductor surface solvated by a liquid electrolyte. For a detailed understanding this system needs to be modeled with various levels of theory. The wide range of different Hamiltonians provided by CP2K allowed to use the same program for every part of this study

The modeling of the semiconductor-solvent interface exploited CP2K's capability to perform large scale ab initio MD simulations (approx. 500 atoms)[76]. In this way it has been shown that the solvent posses a structural ordering not only at the immediate contact but up to 1.5 nm away from the surface. The DFT MD data have been used to fit a classical force field for the solid-liquid interaction reproducing the structural features. With this force field, the classical module of CP2K has been used to study slowly converging properties such as diffusion properties and dielectric permittivity profiles, showing a strong anisotropic behavior even 2 nm away from the surface.

Atomistic insight in the dye/semiconductor interface has been gained by DFT geometry optimizations and mode selective vibrational analysis[4]. Combining the computed data with experimental IR spectra of this interface the preferred binding mode could be identified. Furthermore a protonation dependent change in the binding has been predicted and evidence was found in the experimental spectra and DFT MD simulations including semiconductor, dye and solvent (800 atoms).

In a third study the regeneration process of the solvated dye molecule after oxidation (electron injection) has been targeted[77]. Free energy methods and first principles MD have been used to compute the formation energies for possible intermediate complexes (see Figure 5.1) In this way, a new regeneration pathway has been identified. Furthermore, classical simulations of the distribution of the redox couple at the interface using explicit models and thermodynamic integration were performed. These results showed that the distribution is very different from the standard double layer model and peaks just at the position of the redox active group of the dye. The combined insight from these studies presents a different view of the processes in DSSC. Instead of a single component, it is shown that favorable interactions between structural aspects (dye/semiconductor, electrolyte/semiconductor) and the distinct regeneration pathway of the dye are essential for high performance devices.

5.2 Water interfaces

Water is undoubtly one of the most intriguing substances on earth, its presence being key to life. Accurate simulation of the water is very challenging as weak interactions dominate the system, and subtle effects such as polarization and directionality of hydrogen bonds are crucial. Furthermore, liquid water at ambient conditions is very close to both its freezing and boiling point, and shows various anomalous properties. Numerous simulations with CP2K have been performed to study bulk liquid water, including structure, dynamics, spectroscopic properties and bulk solvation. Among the important findings of these studies is the importance of dispersion corrections to yield the correct density of liquid water in isobaric simulations.[79] Here, work is highlighted that focuses on air-water interface instead that is crucial for various reactive atmospheric processes.[80] Clearly, the air-water interface is sufficiently different from the bulk to challenge the transferability of empirical models, and merits a first-principles approach augmented with empirical dispersion corrections.[81, 82] The ability of CP2K to

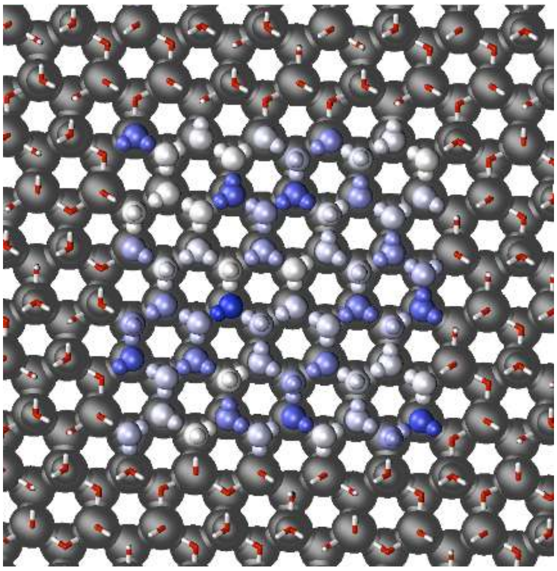

Figure 2: Shown is a model for the (0001) surface of ice Ih in which the molecules in the top layer of the central simulation cell have been colored according to their binding strength ranging from strongly bound (blue) to weakly bound (white). The large ($\approx 80\ kJmol^{-1}$) variation in binding strength is due to the proton disorder of the crystal and facilitates pitting and pre-melting.[78] Models ranging from 288 to 864 molecules have been computed with GGA and hybrid functionals.

deal with large systems and the relatively low cost of treating empty space in a slab model makes CP2K a suitable tool. This is particularly clear for studies of ice(Ih) surface, where large unit cells are necessary to capture reliably the influence of the disordered nature of the hydrogen bonding network on the surface vacancy formation energy.[78] Of particular interest is the question if ions are present at higher concentration at the interface, and how surface propensity correlates with the nature of the ion.[83, 84, 85] Answering this question for the basic charge defects H+ and OH- decides if the surface will be basic or acidic and is particularly demanding for the liquid phase as it requires extensive sampling to obtain free energy profiles.[83, 84] Molecular dynamics simulations furthermore yield the necessary time-correlation functions to compute surface specific spectroscopic information, complementing recent experiments on solvated ions.[86]

5.3 *h*-BN Nanomesh

Monolayers of hexagonal boron nitride (*h*-BN) [88], as well as the isoelectronic carbon structure, graphene, grown on transition metal (TM) surfaces have received much interest as possible templates for use in nano-devices. Chemical vapor deposition of precursor molecules, e.g. borazine in the case of *h*-BN, on a hot metallic surface leads to the spontaneous formation of uniform epitaxial monolayers. Originally, the preparation of a single layer of *h*-BN was achieved on

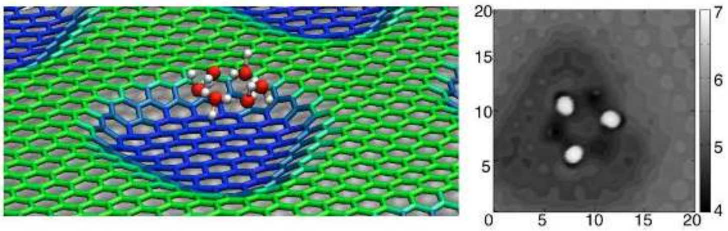

Figure 3: Water cluster adsorbed in the pore of the h-BN/Rh(111) nanomesh. The dipoles of the water molecules arrange in a homodrome hexamer and the simulated STM image, as obtained within the Tersoff-Hamann approximation, (right panel) is consistent with the experimental observation.[87]

the Rh(111) surface [88, 89], but recently similar structures have been grown on many other TM surfaces. The resulting long-range structure of the monolayer depends strongly on a series of factors: The mismatch in the unit cell length l of the free-standing monolayer and the TM surface, the distance-dependence of the *h*-BN to metal interaction, and the deformation energy, both in-plane and out-of-plane, of the monolayer play an important role. In *h*-BN/Rh(111), the balance of these factors, i.e. lattice mismatch of -7.0% and quite strong rhodium-nitrogen interaction, leads to a strongly corrugated structure with a periodicity of 3.22 nm, the so-called *nanomesh* [89]. The highly regular hexagonal arrangement corresponds to a coincidence lattice of 13×13 *h*-BN on 12×12 Rh unit cells. The adsorbate is characterized by "pores" of about 2 nm diameter that strongly interact with the metal and elevated regions, where the interaction with the metal is weak, that form the connected "wire" network. The strong variation in bonding leads to a corresponding variation of the electrostatic potential above the monolayer that is responsible for trapping molecules. The *h*-BN/Rh(111) nanomesh is a very stable structure that withstands temperatures of 1000 K and can be exposed to liquids without losing its properties.

DFT based electronic structure calculations have been instrumental for the understanding of the structure and properties of the *h*-BN and graphene nanomeshes [89, 58]. Using slab model systems with 1000 to 4000 atoms within periodic boundary conditions the nanomesh system has been studied extensively. In a series of experiments the interaction of water with the nanomesh system has been investigated [90, 87]. Simulations have been used for the interpretation of the STM experiments [59, 56]. In this joint effort, both the structure of ice-like as well as small, few molecule clusters in the pore regions have been identified and characterized.

In another experiment it was demonstrated [57] that the nanomesh structure can be reversibly altered using atomic hydrogen. Using DFT calculations it was shown that the atomic hydrogen intercalates the *h*-BN/Rh structure and binds to the top layer of metal atoms. This hydrogen layer weakens the metal-nitrogen bonding, causing a change in the corrugation of the *h*-BN layer.

Acknowledgments

We would like to acknowledge everybody who has supported the CP2K project, too many to name explicitly. Support has come from early users of the code, developers, research facilities and universities, funding bodies, hardware vendors and computer centers.

References

[1] The CP2K developers group. http://www.cp2k.org/(2012).

[2] Treier M, Pignedoli CA, Laino T, Rieger R, Muellen K, Passerone D, Fasel R. Surface-assisted cyclodehydrogenation provides a synthetic route towards easily processable and chemically tailored nanographenes. *Nature Chem* 2011, 3:61–67.

[3] Reiher M, Neugebauer J. A mode-selective quantum chemical method for tracking molecular vibrations applied to functionalized carbon nanotubes. *J Chem Phys* 2003, 118:1634–1641.

[4] Schiffmann F, VandeVondele J, Hutter J, Wirz R, Urakawa A, Baiker A. Protonation-Dependent Binding of Ruthenium Bipyridyl Complexes to the Anatase(101) Surface. *J Phys Chem C* 2010, 114:8398–8404.

[5] Jonsson H, Mills G, Jacobsen KW. Nudged elastic band method for finding minimum energy paths of transitions. In B. J. Berne, G. Ciccotti, and D. F. Coker, editors, *Classical and Quantum Dynamics in Condensed Phase Simulations*, pages 385–404. World Scientific, Singapore, 1998.

[6] Frenkel D, Smit B. *Understanding Molecular Simulation – From Algorithms to Applications*. Academic Press, San Diego, 1996 and 2002.

[7] Mundy CJ, Balasubramanian S, Bagchi K, Tuckerman ME, Martyna GJ, Klein ML. Nonequilibrium molecular dynamics. In *Rev Comp Chem, vol 14*, pages 291–397. 2000.

[8] Jones A, Leimkuhler B. Adaptive stochastic methods for sampling driven molecular systems. *J Chem Phys* 2011, 135:084125.

[9] Ceriotti M, Bussi G, Parrinello M. Langevin Equation with Colored Noise for Constant-Temperature Molecular Dynamics Simulations. *Phys Rev Lett* 2009, 102:020601.

[10] Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys* 2007, 126:014101.

[11] Nosé S. Constant temperature molecular-dynamics methods. *Progress of Theoretical Physics Supplement* 1991, 103:1–46.

[12] VandeVondele J, Röthlisberger U. Canonical adiabatic free energy sampling (CAFES): A novel method for the exploration of free energy surfaces. *Journal of Physical Chemistry B* 2002, 106:203–208.

[13] Nagata Y, Pool RE, Backus EHG, Bonn M. Nuclear quantum effects affect bond orientation of water at the water-vapor interface. *Phys Rev Lett* 2012, 109:226101.

[14] Reed EJ, Fried LE, Joannopoulos JD. A method for tractable dynamical studies of single and double shock compression. *Phys Rev Lett* 2003, 90:235503.

[15] Tuckerman ME, Berne BJ, Martyna GJ. Reversible multiple time scale molecular-dynamics. *J Chem Phys* 1992, 97:1990–2001.

[16] Guidon M, Schiffmann F, Hutter J, VandeVondele J. Ab initio molecular dynamics using hybrid density functionals. *J Chem Phys* 2008, 128:214104.

[17] Tuckerman ME, Marx D, Klein ML, Parrinello M. Efficient and general algorithms for path integral Car–Parrinello molecular dynamics. *J Chem Phys* 1996, 104:5579–5588.

[18] Walewski L, Forbert H, Marx D. Quantum Induced Bond Centering in Microsolvated HCl: Solvent Separated versus Contact Ion Pairs. *J Phys Chem Lett* 2011, 2:3069–3074.

[19] VandeVondele J, Krack M, Mohamed F, Parrinello M, Chassaing T, Hutter J. QUICKSTEP: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comp Phys Comm* 2005, 167:103–128.

[20] Hutter J. Car-Parrinello molecular dynamics. *WIREs Comput Mol Sci* 2012, 2:604–612.

[21] Kolafa J. Time-reversible always stable predictor-corrector method for molecular dynamics of polarizable molecules. *J Comp Chem* 2004, 25:335–342.

[22] Kühne TD, Krack M, Mohamed FR, Parrinello M. Efficient and Accurate Car–Parrinello–like Approach to Born–Oppenheimer Molecular Dynamics. *Phys Rev Lett* 2007, 98:066401.

[23] Laio A, Parrinello M. Escaping free-energy minima. *Proc Natl Acad Sci* 2002, 99:12562–12566.

[24] Barducci A, Bonomi M, Parrinello M. Metadynamics. *WIREs Comput Mol Sci* 2011, 1:826–843.

[25] Michel C, Laio A, Mohamed F, Krack M, Parrinello M, Milet A. Free energy ab initio metadynamics: A new tool for the theoretical study of organometallic reactivity? Example of the C-C and C-H reductive eliminations from platinum(IV) complexes. *Organometallics* 2007, 26:1241–1249.

[26] Masson F, Laino T, Rothlisberger U, Hutter J. A QM/MM Investigation of Thymine Dimer Radical Anion Splitting Catalyzed by DNA Photolyase. *ChemPhysChem* 2009, 10:400–410.

[27] Masson F, Laino T, Tavernelli I, Rothlisberger U, Hutter J. Computational study of thymine dimer radical anion splitting in the self-repair process of duplex DNA. *J Am Chem Soc* 2008, 130:3443–3450.

[28] Santarossa G, Vargas A, Iannuzzi M, Baiker A. Free energy surface of two- and three-dimensional transitions of au 12 nanoclusters obtained by *ab initio* metadynamics. *Phys Rev B* 2010, 81:174205.

[29] Pignedoli CA, Laino T, Treier M, Fasel R, Passerone D. A simple approach for describing metal-supported cyclohexaphenylene dehydrogenation. *EPJ B* 2010, 75:65–70.

[30] Miceli G, Bernasconi M. First-principles study of the hydrogenation process of li2nh. *J Phys Chem C* 2011, 115:13496–13501.

[31] Selli D, Baburin IA, Martonák R, Leoni S. Superhard sp^3 carbon allotropes with odd and even ring topologies. *Phys Rev B* 2011, 84:161411.

[32] Kuo IFW, Mundy CJ, McGrath MJ, Siepmann JI, VandeVondele J, Sprik M, Hutter J, Chen B, Klein ML, Mohamed F, Krack M, Parrinello M. Liquid water from first principles: Investigation of different sampling approaches. *J Phys Chem B* 2004, 108:12990–12998.

[33] McGrath MJ, Siepmann JI, Kuo IFW, Mundy CJ, VandeVondele J, Hutter J, Mohamed F, Krack M. Isobaric-isothermal Monte Carlo simulations from first principles: Application to liquid water at ambient conditions. *ChemPhysChem* 2005, 6:1894–1901.

[34] McGrath MJ, Siepmann JI, Kuo IFW, Mundy CJ, VandeVondele J, Sprik M, Hutter J, Mohamed, Krack M, Parrinello M. Toward a Monte Carlo program for simulating vapor-liquid phase equilibria from first principles. *Comp Phys Comm* 2005, 169:289–294.

[35] McGrath MJ, Siepmann JI, Kuo IFW, Mundy CJ, VandeVondele J, Hutter J, Mohamed F, Krack M. Simulating fluid-phase equilibria of water from first principles. *J Phys Chem A* 2006, 110:640–646.

[36] Kuo IFW, Mundy CJ, McGrath MJ, Siepmann JI. Structure of the methanol liquid-vapor interface: A comprehensive particle-based simulation study. *J Phys Chem C* 2008, 112:15412–15418.

[37] Marx D, Hutter J. *Ab Initio Molecular Dynamics.* Cambridge University Press, Cambridge, 2009.

[38] Kunert T, Schmidt R. Non-adiabatic quantum molecular dynamics: General formalism and case study H-2(+) in strong laser fields. *EPJ D* 2003, 25:15–24.

[39] Castro A, Marques MAL, Rubio A. Propagators for the time-dependent Kohn-Sham equations. *J Chem Phys* 2004, 121:3425–3433.

[40] Chen H, Ratner MA, Schatz GC. Theoretical calculation of the photo-induced electron transfer rate between a gold atom and a gold cation solvated in CCl_4. *J Photochem Photobiol A: Chemistry* 2011, 221:143–147.

[41] Foiles SM, Baskes MI, Daw MS. Embedded-atom-method functions for the fcc metals Cu, Ag, Au, Ni, Pd, Pt, and their alloys. *Phys Rev B* 1986, 33:7983–7991.

[42] Devynck F, Iannuzzi M, Krack M. Frenkel pair recombinations in UO2: Importance of explicit description of polarizability in core-shell molecular dynamics simulations. *Phys Rev B* 2012, 85:184103.

[43] Laino T, Hutter J. Notes on "Ewald summation of electrostatic multipole interactions up to quadrupolar level" [J. Chem. Phys. 119, 7471 (2003)]. *J Chem Phys* 2008, 129:074102.

[44] Laino T, Mohamed F, Laio A, Parrinello M. An efficient linear-scaling electrostatic coupling for treating periodic boundary conditions in QM/MM simulations. *J Chem Theory Comput* 2006, 2:1370–1378.

[45] Jensen F. *Introduction to Computational Chemistry.* Wiley, Chichester, 2007.

[46] Murdachaew G, Mundy CJ, Schenter GK, Laino T, Hutter J. Semiempirical Self-Consistent Polarization Description of Bulk Water, the Liquid-Vapor Interface, and Cubic Ice. *J Phys Chem A* 2011, 115:6046–6053.

[47] Porezag D, Frauenheim T, Köhler T, Seifert G, Kaschner R. Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon. *Phys Rev B* 1995, 51:12947–12957.

[48] Parr RG, Yang W. *Density-Functional Theory of Atoms and Molecules.* Oxford University Press, Oxford, 1989.

[49] Marques MAL, Oliveira MJT, Burnus T. LIBXC: A library of exchange and correlation functionals for density functional theory. *Comp Phys Comm* 2012, 183:2272–2281.

[50] Guidon M, Hutter J, VandeVondele J. Robust Periodic Hartree-Fock Exchange for Large-Scale Simulations Using Gaussian Basis Sets. *J Chem Theory Comput* 2009, 5:3010–3021.

[51] Guidon M, Hutter J, VandeVondele J. Auxiliary density matrix methods for Hartree–Fock exchange calculations. *J Chem Theory Comput* 2010, 6:2348–2364.

[52] Del Ben M, Hutter J, VandeVondele J. Second-order Moller-Plesset perturbation theory in the condensed phase: An efficient and massively parallel Gaussian and plane waves approach. *J Chem Theory Comput* 2012, 8:4177–4188.

[53] Eshuis H, Bates JE, Furche F. Electron correlation methods based on the random phase approximation. *Theor Chem Acc* 2012, 131:1084.

[54] Ren X, Rinke P, Blum V, Wieferink J, Tkatchenko A, Sanfilippo A, Reuter K, Scheffler M. Resolution-of-identity approach to Hartree-Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions. *New J Phys* 2012, 14:053020.

[55] VandeVondele J, Hutter J. An efficient orbital transformation method for electronic structure calculations. *J Chem Phys* 2003, 118:4365–4369.

[56] Ding Y, Iannuzzi M, Hutter J. Nano-ice models for the water aggregates observed on the h-BN/Rh(111) nanomesh. *J Phys: Condens Matter* 2012, 24:445002.

[57] Brugger T, Ma HF, Iannuzzi M, Berner S, Winkler A, Hutter J, Osterwalder J, Greber T. Nanotexture Switching of Single-Layer Hexagonal Boron Nitride on Rhodium by Intercalation of Hydrogen atoms. *Angew Chem Int Ed* 2010, 49:6120–6124.

[58] Iannuzzi M, Hutter J. Comparative study of the nature of chemical bonding of corrugated graphene on Ru(0001) and Rh(111) by electronic structure calculations. *Surf Sci* 2011, 605:1360–1368.

[59] Ding Y, Iannuzzi M, Hutter J. Investigation of Boron Nitride Nanomesh Interacting with Water. *J Phys Chem C* 2011, 115:13685–13692.

[60] Auckenthaler T, Blum V, Bungartz HJ, Huckle T, Johanni R, Krämer L, Lang B, Lederer H, Willems PR. Parallel solution of partial symmetric eigenvalue problems from electronic structure calculations. *Parallel Computing* 2011, 37(12):783–794.

[61] Joshi S, Ecija D, Koitz R, Iannuzzi M, Seitsonen AP, Hutter J, Sachdev H, Vijayaraghavan S, Bischoff F, Seufert K, Barth JV, Auwärter W. Boron nitride on Cu(111): An electronically corrugated monolayer. *Nano Lett* 2012, 12:5821–5828.

[62] Vandevondele J, Borstnik U, Hutter J. Linear scaling self-consistent field calculations with millions of atoms in the condensed phase. *J Chem Theory Comput* 2012, 8:3565–3573.

[63] Lippert G, Hutter J, Parrinello M. A hybrid Gaussian and plane wave density functional scheme. *Mol Phys* 1997, 92:477–487.

[64] Martyna GJ, Tuckerman ME. A reciprocal space based method for treating long range interactions in ab initio and force-field-based calculations in clusters. *J Chem Phys* 1999, 110:2810–2821.

[65] Genovese L, Deutsch T, Neelov A, Goedecker S, Beylkin G. Efficient solution of Poisson's equation with free boundary conditions. *J Chem Phys* 2006, 125:074105.

[66] Goedecker S, Teter M, Hutter J. Separable dual-space Gaussian pseudopotentials. *Phys Rev B* 1996, 54:1703–1710.

[67] Hartwigsen C, Goedecker S, Hutter J. Relativistic separable dual-space Gaussian pseudopotentials from H to Rn. *Phys Rev B* 1998, 58:3641–3662.

[68] Krack M. Pseudopotentials for H to Kr optimized for gradient-corrected exchange-correlation functionals. *Theor Chem Acc* 2005, 114:145–152.

[69] VandeVondele J, Hutter J. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J Chem Phys* 2007, 127:114105.

[70] Lippert G, Hutter J, Parrinello M. The Gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations. *Theor Chem Acc* 1999, 103:124–140.

[71] Iannuzzi M, Chassaing T, Wallman T, Hutter J. Ground and excited state density functional calculations with the Gaussian and augmented-plane-wave method. *Chimia* 2005, 59:499–503.

[72] Blöchl PE. Projector augmented–wave method. *Phys Rev B* 1994, 50:17953–17979.

[73] Iannuzzi M, Hutter J. Inner-shell spectroscopy by the Gaussian and augmented plane wave method. *Phys Chem Chem Phys* 2007, 9:1599–1610.

[74] Weber V, Iannuzzi M, Giani S, Hutter J, Declerck R, Waroquier M. Magnetic linear response properties calculations with the Gaussian and augmented-plane-wave method. *J Chem Phys* 2009, 131:014106.

[75] Declerck R, Pauwels E, Van Speybroeck V, Waroquier M. First-principles calculations of hyperfine parameters with the Gaussian and augmented-plane-wave method: Application to radicals embedded in a crystalline environment. *Phys Rev B* 2006, 74:245103.

[76] Schiffmann F, Hutter J, VandeVondele J. Atomistic simulations of a solid/liquid interface: a combined force field and first principles approach to the structure and dynamics of acetonitrile near an anatase surface. *J Phys: Condens Matter* 2008, 20:064206.

[77] Schiffmann F, VandeVondele J, Hutter J, Urakawa A, Wirz R, Baiker A. An atomistic picture of the regeneration process in dye sensitized solar cells. *Proc Natl Acad Sci* 2010, 107:4830–4833.

[78] Watkins M, Pan D, Wang EG, Michaelides A, VandeVondele J, Slater B. Large variation of vacancy formation energies in the surface of crystalline ice. *Nature Materials* 2011, 10:794–798.

[79] Schmidt J, VandeVondele J, Kuo IFW, Sebastiani D, Siepmann JI, Hutter J, Mundy CJ. Isobaric-Isothermal Molecular Dynamics Simulations Utilizing Density Functional Theory: An Assessment of the Structure and Density of Water at Near-Ambient Conditions. *J Phys Chem B* 2009, 113:11959–11964.

[80] Mundy CJ, Kuo IFW. First-principles approaches to the structure and reactivity of atmospherically relevant aqueous interfaces. *Chem Rev* 2006, 106:1282–1304.

[81] Kuo IFW, Mundy CJ, Eggimann BL, McGrath MJ, Siepmann JI, Chen B, Vieceli J, Tobias DJ. Structure and dynamics of the aqueous liquid-vapor interface: A comprehensive particle-based simulation study. *J Phys Chem B* 2006, 110:3738–3746.

[82] Baer MD, Mundy CJ, McGrath MJ, Kuo IFW, Siepmann JI, Tobias DJ. Re-examining the properties of the aqueous vapor-liquid interface using dispersion corrected density functional theory. *J Chem Phys* 2011, 135:124712.

[83] Mundy CJ, Kuo IFW, Tuckerman ME, Lee HS, Tobias DJ. Hydroxide anion at the air-water interface. *Chem Phys Lett* 2009, 481:2–8.

[84] Watkins M, VandeVondele J, Slater B. Point defects at the ice (0001) surface. *Proc Natl Acad Sci* 2010, 107:12429–12434.

[85] Lewis T, Winter B, Stern AC, Baer MD, Mundy CJ, Tobias DJ, Hemminger JC. Does Nitric Acid Dissociate at the Aqueous Solution Surface? *J Phys Chem C* 2011, 115:21183–21190.

[86] Baer MD, Mundy CJ, Chang TM, Tao FM, Dang LX. Interpreting Vibrational Sum-Frequency Spectra of Sulfur Dioxide at the Air/Water Interface: A Comprehensive Molecular Dynamics Study. *J Phys Chem B* 2010, 114:7245–7249.

[87] Ma HF, Ding Y, Iannuzzi M, Brugger T, Berner S, Hutter J, Osterwalder J, Greber T. Chiral distortion of confined ice oligomers (n = 5,6). *Langmuir* 2012, 28:15246–15250.

[88] Corso M, Auwärter W, Muntwiler M, Tamai A, Greber T, Osterwalder J. Boron nitride nanomesh. *Science* 2004, 303:217–220.

[89] Berner S, Corso M, Widmer R, Groening O, Laskowski R, Blaha P, Schwarz KH, Goriachko A, Over H, Gsell S, Schreck M, Sachdev H, Greber T, Osterwalder J. Boron nitride nanomesh: Functionality from a corrugated monolayer. *Angew Chem Int Ed* 2007, 46:5115–5119.

[90] Ma HF, Brugger T, Berner S, Ding Y, Iannuzzi M, Hutter J, Osterwalder J, Greber T. Nano-ice on Boron Nitride Nanomesh: Accessing Proton Disorder. *ChemPhysChem* 2010, 11:399–403.