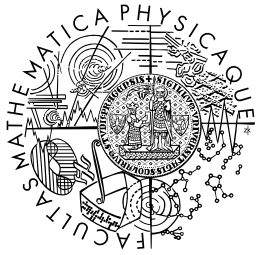

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

HABILITATION THESIS

Mgr. Jiří Klimeš, Ph.D.

Towards a reliable description of cohesive properties of molecular solids

Department of Chemical Physics and Optics

Prague 2023

Preface

The presented thesis consists of 9 published scientific articles and accompanying text which forms the body of the thesis. The selected articles focus on understanding the accuracy and precision of advanced electronic structure methods, primarily of the random phase approximation. In addition, I included two articles which are results of collaboration with experimental groups. Most of the research was performed at the Faculty of Mathematics and Physics of the Charles University, with the exception of one work which resulted from my Marie Sklodowska-Curie fellowship at J. Heyrovský Institute of Physical Chemistry in Prague.

The motivation for the research and the random phase approximation method are presented in the first and second chapters, respectively. The three subsequent chapters present the main findings of our work together with additional background information, related works not included in the thesis, and outlook. The third chapter discusses the results obtained for the random phase approximation method for molecular solids and molecular adsorption in zeolites. Simulations that were used to support experimental findings in two papers are presented in chapter four. Chapter five presents studies in which we try to understand precision of our results.

Acknowledgements

I'm grateful to my wife, kids, and family for their support and understanding. I thank my students, postdocs, and colleagues that contributed to the work presented here and to other studies.

Contents

1. Introduction

The world around us is formed by atoms yet it's very difficult to imagine materials at the atomic level due to the small size of atoms. Fortunately, researchers were able to figure out the physical laws of the atomic world, that is the quantum mechanics, without seeing them. The laws can be used to describe the interactions between atoms and can thus explain the existence of molecules or solids. Later, these laws were used in computer simulations to build collections of atoms and analyse their time evolution or other properties. Nowadays, with powerful supercomputers and accurate methods we can perform "virtual experiments" in which we can see how atoms behave in complex materials. Needless to say, experimental techniques have also made substantial progress and in the past years we could see some beautiful examples of world at the atomic level viewed both by simulations and experiments [1]. In this thesis I'll discuss some of our contributions to the development of methods for materials simulations and applications to systems studied experimentally. But let me first briefly discuss my motivation for this research direction and how it occurred.

I finished my masters by simulations of current flow through a chain of gold atoms between two electrodes, a model for "nanoelectronic" system. This and related "molecular electronics" were very popular topics at the time and I wanted to continue to work on them [2]. After searching I applied for PhD at the London Centre for Nanotechnology and, after some modifications of the plan, I started under the supervision of Angelos Michaelides. Amusingly, I started to work in the field of surface science, one that I haven't realised before that it existed. The project that I was working on was to understand basic natural processes and my part was to study salt dissolution in water. However, after some time it became clear that the methods that we were using lack accuracy, the adsorption energy of a single water molecule was only around 2/3 of that obtained by a reference quality approach [3]. As my work involved a comparison of energies of various surfaces with water clusters the large error of adsorption energy made the results a bit dubious and I felt it's important to take the accuracy seriously and look for a better method.

The scheme that we were using was Kohn-Sham density functional theory (DFT) [4, 5] with approximate exchange-correlation (XC) functional called PBE after its proponents Perdew, Burke, and Ernzerhof [6]. The DFT is a particularly efficient approximation of quantum mechanics that, at the time, allowed us to perform the salt dissolution study involving structure optimisation of hundreds of structures with hundreds of atoms. However, the efficiency comes at the cost of reduced accuracy and while PBE was perceived as the standard for solid state calculations, it was clearly having problems to describe the interaction between salt and water. Fortunately, at that time more and more people realised that the simple XC functionals lack the ability to describe long-range correlations between electrons, also known as dispersion forces. The dispersion interactions decay as r^{-6} and hold together systems such as graphene or noble gas clusters. This change of our perception of the DFT methods was coming from people interested in organic chemistry and biochemistry [7–14] and also from researchers in solid state physics that were interested in systems such as layered materials [15–

17]. Fortunately for us, various methods that these and other groups had been developing were becoming more widely available to be used or to be tested.

To try to improve the adsorption energy of water on salt we used several of the newly developed methods, such as the van der Waals density functional (vdW-DF) of Langreth and Lundqvist groups [17] or pairwise dispersion corrections mostly developed by group of prof. Stefan Grimme [18]. However, the first didn't improve the adsorption energy compared to PBE and the latter gave an adsorption energy too large, likely due to overestimation of the interaction originating from sodium atoms. This is because the early pairwise methods didn't make a difference between sodium atoms that are neutral and have a large polarisability and atoms that are positively charged, as in the salt crystal, and have a small polarisability. The long-range correlation effects in the vdW-DF scheme are calculated directly from density so this approach looked more promising. Apart from a correlation functional that includes dispersion, the vdW-DF method also contains a standard XC functional that should describe short-range exchange and correlation effects. When looking at the vdW-DF method more closely we realised that the original XC functional, called revPBE [19], can be replaced by another one without breaking any physics. We used this fact to try out different parametrisations and, after some time, we got variants of vdW-DF that worked better for intermolecular binding energies and also for the adsorption energy [20]. This scheme is used even nowadays, especially in condensed matter for systems for which it's difficult to describe the different oxidation states of the atoms with the pairwise methods [21].

For myself, the experience with testing and developing vdW-DF-based methods and various XC functionals gave me the feeling that the DFT development can be quite messy. As a next step I hoped to work on something that required less parametrisation and offered hopefully better accuracy. Being lucky again I had a chance to visit group of prof. Georg Kresse in Vienna who at the time published a study about implementation of second-order Møller-Plesset perturbation theory (MP2) for solids [22]. Moreover, his group was working on other correlated methods, such as the random-phase approximation (RPA) [23]. These methods are based on perturbation theory in which one only needs to make a choice which set of perturbation terms to use. Having even more luck again, I started to work in the group in 2012.

In Vienna I was working on different projects related to the development of RPA. The first one was to obtain energies of electronic bands for self-consistent RPA within the optimised effective potential (OEP) method [24]. This looked like a short project for prof. Kresse as he did many of the calculations already and, according to him, we just had to do the calculations for a few more materials to finish. However, while doing some tests I realised that the default settings that we were using lead to imprecise results. The energies of the states were converging only slowly with the basis-set size and, in fact, extrapolation to the complete basis set limit was needed to avoid the basis-set related errors. In fact, I stumbled upon another change of mindset that was going on, now in the solid state community and that highlighted the importance of precision.

The band-structure calculations that I did using OEP-RPA are similar to *GW* calculations that are very often performed to obtain the electronic band-structure of solid materials. The *GW* calculations are usually done on top of DFT states [25,

26]. However, they are significantly more computationally demanding compared to the DFT calculations. Therefore, several simplifications have been employed to reduce their cost and, in fact, the *GW* is itself an approximation of a more general expression for electron self-energy [27]. This wealth of choices meant that while there had been many *GW* calculations it was difficult to find results which actually agreed with each other. A particularly worrying example of disagreement was zinc oxide (ZnO) which even prompted someone to write a blog about this issue (now probably lost). One of the reasons of disagreement was the fact the *GW* energies converge much slower with parameters, such as the basis-set size or the number of k-points, than the DFT energies. Therefore, evaluating the *GW* energies using a set of parameters sufficient to converge the DFT data can give values which are still very imprecise. This was shown shortly before I started with the OEP-RPA calculations and was still in the process of being accepted by the wider community [28–30]. While the necessity to converge our results slowed down the process of getting them, it showed the importance of obtaining precise values and later helped us to contribute to explanation and understanding of their origin [31, 32].

Many of the articles presented in this thesis are devoted to molecular solids. These systems are important but their properties such as binding energy are quite difficult to describe accurately as they depend on a good description of both intramolecular and intermolecular interactions. Moreover, many of them have the right size to be studied by simple correlated methods such as MP2 or RPA within periodic boundary conditions. Hence they are useful test systems for our methods and often give us information that can't be obtained from other tests. For example, we modified the vdW-DF functional to give very accurate binding energies for molecular dimers. However, the results for binding energies of molecular solids contained in tests C21 [33] and X23 [34] were much less satisfactory. Therefore, the good results for dimers are due to some error cancellation which is less effective for the molecular solids.

Molecular solids can be also used to illustrate a qualitative change that appeared in method development and testing. For our vdW-DF paper in 2010 [20] we were able to test the method on few tens of structures, mainly due to the large computational cost of the scheme [35]. Since that time increased computational power and improved algorithms allow one to test methods on data sets with thousands or even more binding energies [36]. For molecular solids we can obtain similar sets by creating dimers, trimers, and larger molecular clusters by choosing an appropriate number of molecules from the solid. We can then obtain reference data and test the specific method for this large set which gives us a very detailed information about the strengths and weaknesses of the method, such as the extent of the error cancellations. In Chapter 3 I show some examples of our recent work along this direction.

Let me mention one more point that motivates part of our research and which relates to replicability and reproducibility of results. On several occasions I spent days, weeks, or even months trying to figure out why results of my calculations deviate from data published in the literature. Sometimes I was able to identify the issue myself, sometimes not, in other cases the disagreement pointed to an interesting observation. In many studies we therefore try to estimate the uncertainty of our results or we even calculate the same property using two different

codes and analyse the differences. This helps us to develop a set of reliable parameters and set-ups for the calculations and hopefully makes our data more reproducible. I think this exercise is particularly useful for calculations involving molecular solids since there is a lot of parameters that one needs to choose and that can affect the result. Some examples of our research that concerns precision are discussed in Chapter 5.

To conclude, I've been working on projects that involved approaches used traditionally within quantum chemistry as well as projects using solid state methods. In some cases it was even convenient to learn to use codes for simulation of biomolecules to get access to longer simulation times. The systems that we study are on the border between the mentioned fields as well and we can therefore use quite a variety of approaches to analyse the problem at hand.

2. Methods

We use a range of simulation methods to describe the various systems and properties that we study. All of them describe nuclei as classical particles, the electrons are then treated at the level of quantum mechanics or neglected within classical force-fields. Within quantum mechanics we use Kohn-Sham density functional theory (DFT) approximations or methods that consider correlations between electrons explicitly such as Møller-Plesset perturbation theory, random-phase approximation (RPA), or coupled clusters (CC) at various levels. This range of theoretical approaches gives us a versatile set of tools that we can use for various tasks, from calculation of highly accurate energies of small systems to molecular dynamics of systems with tens of thousands of atoms. Many of the methods are standard and available in different computational packages and I'll thus not discuss them here. In the following I'll comment on the scheme that we try to develop and which is less widely used and that's the RPA.

2.1 Random phase approximation

The RPA method comes from the works of Bohm and Pines in 1950s [37] in which an approximation that became known as the RPA was made to simplify calculations of electron correlation energy in the uniform electron gas (UEG). Since that time RPA saw years of interest and neglect and nowadays the RPA is used in several contexts for calculating total energies or excitation spectra. The origins and various approaches to evaluate the RPA energy are discussed in more detail in review articles [38–40].

The RPA energy is composed of two terms, the "exact-exchange" energy, E_{EXX}, and the RPA correlation energy, $E_{\mathrm{c}}^{\mathrm{RPA}}$. The E_{EXX} part is the expectation value of the many-body Hamiltonian obtained for a single Slater determinant. The Slater determinant can be built using states obtained by a DFT calculation or it can be the self-consistent Hartree-Fock solution. In the latter case the E_{EXX} energy is simply the Hartree-Fock energy. Note that the name "exact exchange" originates from the fact that the electron exchange energy is not approximated, as opposed to the approximations made in standard DFT exchange-correlation functionals.

There are several ways to derive the different expression for the RPA correlation energy and let me here briefly outline the one used in solid state physics which is based on adiabatic connection and fluctuation-dissipation theorem [23, 41, 42]. Within the adiabatic connection one scales the Coulomb interaction V_{ee} with a parameter λ to go from the non-interacting Kohn-Sham (KS) system ($\lambda = 0$) to the interacting one ($\lambda = 1$) [43–46]. In this way one obtains a formula for the exchange-correlation energy which is exact in principle but contains unknown quantities, such as wavefunctions for different values of λ. The relation can be further modified by what's denoted as fluctuation dissipation theorem leading to an equation for the correlation energy that involves response function $\chi(i\omega, \lambda)$ at imaginary frequency $i\omega$ and coupling strength λ

$$E_{\mathrm{c}} = -\frac{1}{2\pi}\int_0^1 \mathrm{d}\lambda \int_0^\infty \mathrm{d}\omega \mathrm{Tr}\left[V_{\mathrm{ee}}\left[\chi(i\omega,\lambda) - \chi(i\omega,0)\right]\right] . \tag{2.1}$$

The trace in the integral denotes integration over real space and spin coordinates which we don't explicitly write in the equation. The response function for the non-interacting system $\chi(i\omega, 0)$ can be calculated from the occupied and virtual KS states [47] or using KS Green's functions [48]. The response $\chi(i\omega, \lambda)$ at interaction strength λ can be evaluated within time-dependent DFT. To obtain $\chi(i\omega, \lambda)$ exactly one would need to know the exact exchange-correlation kernel $f_{\rm xc}$ [49]. To make the scheme computationally tractable $f_{\rm xc}$ can be approximated [50, 51] or neglected. The latter case then leads to the RPA energy expression commonly used in solid state physics

$$E_{\rm c}^{\rm RPA} = \frac{1}{2\pi}\int_0^\infty {\rm d}\omega {\rm Tr}\left[\ln\left[1-\chi(i\omega,0)V_{\rm ee}\right]+\chi(i\omega,0)V_{\rm ee}\right] . \tag{2.2}$$

In this expression the frequency integral is evaluated on a grid and diagonalisation is used to evaluate the logarithm.

In simple terms, RPA can be understood via interactions of independent electron excitations or density fluctuations [42]. This is illustrated in Fig. 2.1 for two and three atoms. Within perturbation theory one starts from mean field solution (one colour spheres in Fig. 2.1) and considers electron excitations to virtual states. These lead to density fluctuations which are indicated with the red-blue spheres in the figure. In Eq. 2.2 the excitations and density fluctuations are described by the response function $\chi(i\omega, 0)$. The Coulomb interaction then couples the different excitations. Due to the logarithm, the number of possible excitations is infinite. When two excitations appear on atoms in distance r we obtain the long-range electron correlation with the leading order proportional to $-r^{-6}$. This process can be also nicely illustrated with a pair of quantum oscillators that we teach in our classes of introduction to quantum mechanics. A process where electrons on three atoms get excited is shown on the right hand side of Fig. 2.1. This contribution to the correlation energy can be approximated by so-called Axilrod-Teller-Muto formula [52, 53].

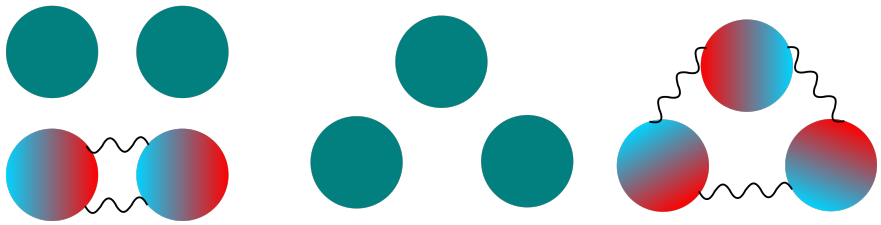

Figure 2.1: A sketch of interactions described by random-phase approximation (RPA) between atomic dimer (left) and trimer (right). Full spheres indicate mean field solutions and red-blue spheres show electron excitations or density fluctuations. The wiggly line then shows the Coulomb interactions between the excitations.

When the logarithm is expanded into series the integral over frequency can be carried out and one obtains the same expressions as within Møller-Plesset perturbation theory (MPPT). The corresponding terms can be drawn with Goldstone diagrams, as illustrated in Fig. 2.2. Note that the first diagram is a contribution

obtained within MP2 and is often called direct MP2 [22]. The third- and fourth-order contributions would be only obtained at MP3 and MP4 levels. The MPPT expressions contain additional terms that also allow for exchange of electrons between the density fluctuations (bubbles) as well as other diagrams.

The RPA energy expression includes diagrams up to infinite order while MP2, MP3, ... expressions are finite. This is important for metals and other systems with delocalised electrons as for them screening by higher-order terms can be significant. This can be imagined with the help of Fig. 2.1 for the case of two atoms. In the second order, the fluctuation on one of the atoms interacts directly with the fluctuation on the other atom. However, in third order one of the fluctuations can interact with another fluctuation on the same atom (we still consider only two atoms). This additional fluctuation will typically reduce the extent of the initial fluctuation as electrons tend to avoid each other. As a consequence, the overall polarisation of the atom and thus the interatomic interaction will be reduced compared to the second-order process. One can also say that the excitation will be screened by the other electrons. Within MPPT the lack of screening leads to overestimated MP2 energies for some systems, e.g., for those with delocalised electrons [54–56].

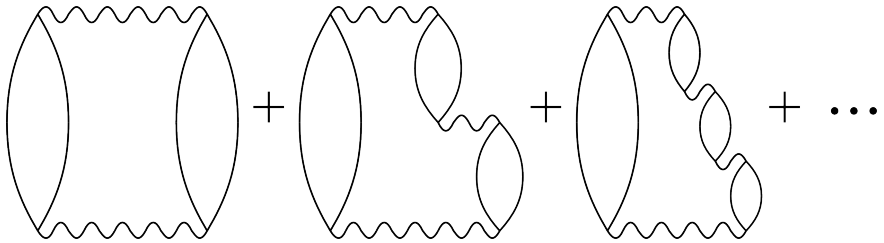

Figure 2.2: Examples of Goldstone diagrams giving low-order contributions to energy at random-phase approximation level. Notably, there is no exchange of electrons between the different bubbles (density fluctuations) and there are no particle-hole interactions (ladder diagrams).

RPA has attracted interest within the DFT community as an advanced exchange-correlation (XC) functional [57]. It's often denoted as a fifth-rung functional, according to the classification of Perdew and Schmidt [58]. More specifically, the fifth rung contains functionals that consider virtual states for the calculation of the exchange-correlation energy. This can mean RPA in its form of Eq. 2.2 or a general, double-hybrid, XC functional that contains some RPA contribution and other terms. Both of these forms are being used and developed. The double hybrid schemes are typically developed to improve the description of long-range correlations within hybrid functionals or, from the other side, to improve the accuracy of RPA for calculation of atomisation energies and similar properties [59–61]. In such functionals one often performs range-separation of the Coulomb interaction [62, 63] and takes only the long-range part of the RPA energy [64, 65]. The double hybrid methods with long-range RPA component have the benefit of a faster convergence with the basis-set size as one avoids the reconstruction of the two-electron cusp [66].

RPA is often evaluated non-self-consistently, meaning that the RPA energy is obtained in a "one shot" manner for a set of states (orbitals and their energies)

calculated by other XC functional or the Hartree-Fock (HF) method. A self-consistent evaluation is possible within the optimized-effective potential scheme. This procedure removes the dependence on the input states but increases the computational cost [67–71]. Within the "one shot" approach, there is therefore an additional flexibility due to the input states and while majority of calculations use the PBE functional [6], there is no reason not to use another one [57, 72]. In our work we also tested the use of different XC functionals to provide input states for RPA to identify if another choice improves the predicted properties.

The RPA energy expression can be also obtained from the CC method. To do this only double (D) excitations are included in the cluster operator $\hat{T}$, integrals are not anti-symmetrised and contributions from ladder diagrams are not considered. The approximation is then called direct ring CCD or drCCD for short. The relation between the methods and the different working equations was discussed by Scuseria and co-workers [73], Jansen *et al.* [74], and others [75, 76].

The neglect of exchange diagrams in RPA allows for a fast evaluation of the energy but it introduces some serious errors in the result. The neglect means that each electron correlates with itself and as a consequence, the correlation energy is usually loo large in magnitude. Most notably, the correlation energy is non-zero even for systems with a single electron, such as hydrogen atom. The accuracy of the total energies can be included by adding the exchange terms, such as in the second-order screened exchange (SOSEX) scheme [75, 76].

There is one more point that deserves to be mentioned here. As discussed, RPA is often evaluated on energies of semi-local DFT, most often the PBE functional. It's also known that PBE and similar "generalised-gradient approximation" (GGA) functionals give a smaller energy difference between the occupied and virtual orbitals than the one obtained by the HF method. The gap appears in denominator when calculating the response function $\chi(i\omega, 0)$ therefore a smaller gap increases the response and should thus increase the interaction energy between atoms or molecules. However, within CC, the larger HF gap is reduced by attractive electron-hole ladder diagrams. These are missing from RPA and there is a notion that the smaller GGA gap partially compensates for the lack of the ladder diagrams. Therefore, the long-range correlations, expressed, e.g., by C_6 coefficients turn out to be described reasonably well at the RPA level based on GGA functionals, at least for closed shell atoms and molecules.

To summarise, RPA is a method that can be either viewed as an advanced DFT functional or a simple CC approximation. As such it avoids some of the problems of simpler DFT approximations at a computational cost typically lower than that of accurate CC approximations. These properties make it suitable for our work which is primarily focused on cohesive properties of molecular clusters or solids. How RPA performs for these systems is discussed in the next chapter of this thesis.

3. Tests of random phase approximations for molecular solids and clusters

3.1 Introduction

In this chapter I will discuss results of tests that we performed to assess the accuracy of the random phase approximation (RPA) scheme for the prediction of binding energies. Before going into the discussion, I comment on the development of RPA-based methods for predicting binding energies. Initial results appeared around 2010 and showed that the RPA binding energies are too weak when compared to reference data. One such result was obtained by Li and co-workers [77] for benzene and methane crystals. Moreover, we also obtained an unsatisfactory result for adsorption energy of a water molecule on NaCl(100) surface [78]. The adsorption energy was less accurate even compared to dispersion corrected density functional theory (DFT) functionals which typically require much smaller computational time. However, there were also some encouraging results. The RPA binding energies of various systems were underestimated by a consistent amount, typically by 10 to 20%. Moreover, RPA gave correct predictions for systems considered to be difficult, such as the CO adsorption on transition metals [79]. Therefore, at that time, the RPA looked promising but more tests were required to understand the results.

The consistent results obtained for RPA suggested that its accuracy could be improved by adding additional terms from perturbation theory. One such scheme is the second-order screened exchange (SOSEX), which adds second-order exchange diagram with one screened Coulomb interaction [75]. This makes the method similar to second-order Møller-Plesset perturbation theory (MP2) with the difference that in MP2 the Coulomb interactions aren't screened. The screening means that RPA+SOSEX keeps its accuracy for systems with small energy gap between occupied and virtual states or bands [76]. The downside of SOSEX is its increased scaling and computational cost compared to RPA when using plane-wave basis sets. As a result and due to other developments it hasn't seen a lot of applications for calculations of binding energies [38, 80], at least so far.

In a work published in 2011 Ren and co-workers noticed that the consistent underbinding can be reduced by replacing the exact exchange (EXX) energy component of RPA by Hartree-Fock (HF) [81]. The EXX evaluates the same Hamiltonian as HF but using DFT states. This difference suggests that additional perturbation theory terms that take into account non-diagonal part of the many-body Hamiltonian would improve the binding energies. This was indeed observed and the energy correction was termed "renormalised singles energies" (RSE) by Ren *et al.* [81].

The RSE corrections to RPA can be evaluated by diagonalization and they thus don't affect the computational cost of RPA significantly. However, adding RSE reduces the accuracy of RPA for lattice constants of metals due to presence of bare Coulomb interaction [42]. During my postdoc in Vienna with prof.

Kresse we developed a singles correction that takes screening of the Coulomb interaction into account and we called it *GW*SE [42]. RPA with *GW*SE, based on PBE orbitals, then shows similar benefits as RPA: consistent accuracy over different classes of materials (metals, semiconductors, insulators). Moreover, it substantially improved the binding energies over RPA which we illustrated for adsorption energy of water monomer on NaCl(100) and for benzene crystal [42]. For the first system, RPA with RSE gave the same adsorption energy as the estimated reference value of −430 meV, using the RPA+*GW*SE scheme lead to a weaker binding of −410 meV. For benzene crystal, we observed a reduction of the errors by around 10 % for either of the methods [42].

The work discussed so far motivated my research presented within this chapter. In short, RPA with the RSE or *GW*SE corrections showed a very high accuracy and modest increase of computational cost (same $O(N^3)$ scaling and a prefactor larger by 6 to 10 times compared to RPA). Therefore, the method looked promising and we decided to analyse its accuracy in a more detail. In our work from 2015 we only tested one molecular solid and one adsorption energy [42]. Therefore, we focused on expanding the tests both for molecular solids and adsorption. As a result we published the articles P1 and P2 included in this thesis which are discussed later in this chapter, in Section 3.2. Furthermore, in that section I also briefly mention published articles to which I contributed but which are not included in this thesis.

The work performed in P1 showed that RPA with RSE or *GW*SE corrections gives a very good binding energies for molecular solids. However, binding energies are only a single number for each solid and it's possible that a good performance is a consequence of various error cancellations. To understand the binding energies in detail one can perform many-body expansion (MBE) of the binding energy. Within MBE the binding energy is obtained as a sum of interaction energies of molecular dimers extracted from the solid and corrections from molecular trimers, tetramers, and possibly even larger clusters. If the crystalline material is formed by a single compound and all the molecules are symmetry equivalent, it suffices to set one molecule as the reference one. The binding energy can be then obtained from interactions of the reference molecule (ref) with the other molecules (indexed by i, j, k, ...) as

$$E_{\mathrm{bind}} = \frac{1}{2}\sum_{i} E_{\mathrm{ref},i} + \frac{1}{3}\sum_{i,j} \Delta E_{\mathrm{ref},i,j} + \frac{1}{4}\sum_{i,j,k} \Delta E_{\mathrm{ref},i,j,k} + \dots \,, \tag{3.1}$$

where $E_{\mathrm{ref},i}$ are the dimer contributions and $\Delta E_{\mathrm{ref},i,j}$ and $\Delta E_{\mathrm{ref},i,j,k}$ are non-additive three- and four-body terms.

Analysis of the RPA errors using MBE was one of the main topics of my ERC grant. We used crystalline hydrocarbons and methane clathrate cluster as test systems. Moreover, we also used a reference set of molecular trimers to understand the three-body errors of RPA for a broader set of interactions [82]. The results of these tests were published in three papers included in this thesis as P3, P4, and P5 and are discussed in Section 3.3.

One of the points we tried to understand is how the RPA energies depend on input states. As discussed in the previous chapter, it is possible to run RPA calculations self-consistently but this comes with a higher computational cost and can introduce other issues [24, 68, 71]. Therefore, for solids, the current practice

is to run self-consistent DFT calculations first and use the resulting states as an input for RPA energy evaluation [23, 50, 83]. To assess the effect of using different input states we evaluated binding energies and many-body terms for RPA based on several DFT functionals.

There is one more reason to understand the many-body errors of a given theoretical method and that is embedding. Imagine that we have a very accurate but computationally expensive method and the system is too large to be treated with this method within a single calculation. For molecular solids it can mean that the energy of the periodic solid cannot be obtained directly, we can get only the energies of small molecular clusters. In this way, we could use the MBE to obtain the binding energy of the solid. This could be rather computationally demanding and prone to slow convergence with the number of fragments considered. An alternative is to start from a binding energy obtained with a simpler scheme and correct only terms where the simpler scheme produces errors. Therefore, we can obtain a good estimate of the binding energy by using:

$$E^{\text{PBC, high}} = E^{\text{PBC, low}} + \sum_i \left(E_i^{\text{MBE, high}} - E_i^{\text{MBE, low}} \right) . \tag{3.2}$$

In this equation, "high" and "low" correspond to the accurate and less accurate methods, respectively, "PBC" is a result corresponding to periodic boundary conditions while "MBE" is many-body expansion. The terms that we sum over can be different dimers, trimers, and so on, as in Eq. 3.1. Therefore, when more and more terms are included in the summation in Eq. 3.2 the initially less accurate result is gradually improved to the value of the high quality method. This scheme will work the better the closer the differences in the sum are to zero. Therefore, one of our motivations is to understand how close are different methods to reference schemes such as CCSD(T) for the two-, three-, four-, and higher-order interactions.

3.2 Tests of RPA binding and adsorption energies

In this part I discuss articles P1 and P2 in which we analysed the accuracy of the RPA method using a general test set of molecular solids and a set of adsorption energies of molecules in zeolites, respectively.

3.2.1 Binding energies of molecular solids

To test RPA for binding energies of molecular solids we selected the C21 test set developed by de la Roza and Johnson [33] and later modified as the X23 test set by Reilly and Tkatchenko [34]. The test set contains molecular crystals for which reliable experimental sublimation enthalpies are available. Zero-point energies and thermal effects are subtracted from the experiment to obtain binding energies that can be directly compared to calculated results. In the set there are hydrogen bonded crystals, such as ammonia or oxalic acid, as well as crystals bound by weaker electrostatic (quadrupole) and dispersion interactions, such as benzene or adamantane. While the C21 contains 21 crystals, in P1 we obtained

the binding energy only for ten systems due to limited computational resources as discussed below.

Our aim in P1 was to obtain the binding energies with a precision of around 1 %. As the binding energies are approximately between 25 and 100 kJ/mol, we aimed at errors of 1 kJ/mol at most. Such high precision is needed to veritably assess the accuracy of RPA. However, reaching high precision of RPA binding energies is not straightforward as the energies depend strongly on the k-point grid and basis-set size. At the same time, these parameters can't be set at values sufficiently high to achieve convergence due to the memory demands of the calculations. For example, we observed that increasing the basis set cut-off by 100 eV approximately doubles the memory requirements. The largest calculations made for P1 needed around 6 TB of RAM distributed over 64 compute nodes on the Salomon supercomputer.

The computational details are described in P1 and here we mention points that might seem minor at first but can be actually important. First, if it is not possible to obtain energy for a large cut-off and dense k-point grid $E^{\mathrm{k}}_{\mathrm{cut}}$, one can approximate it by starting from energy $E^{\mathrm{k\,low}}_{\mathrm{cut}}$ obtained for a less dense k-point grid and correct it with a difference of energies obtained for a smaller basis-set cut-off as follows

$$E^{\mathrm{k}}_{\mathrm{cut}} \approx E^{\mathrm{k\,low}}_{\mathrm{cut}} + E^{\mathrm{k}}_{\mathrm{cut\,low}} - E^{\mathrm{k\,low}}_{\mathrm{cut\,low}}\,. \tag{3.3}$$

The difference of the last two energies can be though of as a k-point correction and can substantially reduce the cost of performing the calculations. However, we found that this k-point correction depends on the basis-set cut-off. Importantly, the difference $E^{\mathrm{k}}_{\mathrm{cut\,low}} - E^{\mathrm{k\,low}}_{\mathrm{cut\,low}}$ is larger in magnitude for smaller cut-offs. For the systems considered, the correction is not reliable for basis-set cut-offs below 600 eV when standard projector-augmented wave (PAW) data sets are used in VASP. Therefore, it is not advisable to use calculations with a small basis-set cut-off, say 400 eV in VASP, to check the k-point convergence. The calculations with this small cut-off are relatively fast but the results will give misleading information about the actual k-point convergence for large cut-offs.

Second, the RPA energy of the solid as well as the energy of the reference isolated molecule depend on the basis-set size. The leading order of the basis-set incompleteness error was derived to be $E^{-3/2}_{\mathrm{cut}}$ [23, 31, 32]. This is simply the inverse number of basis functions and identical to the L^{-3} dependence on largest angular momentum L observed for typical Gaussian basis sets [84–86]. However, when one calculates the binding energy, the leading errors might accidentally subtract out, leading to a $E^{-5/2}_{\mathrm{cut}}$ dependence of the binding energy. The prefactor in the basis-set incompleteness error is determined by the electron density, if the density doesn't change upon formation of the dimer or crystal, the leading orders subtract out. We observed such a behaviour for adamantane which is a crystal bound by dispersion. For other systems the $E^{-3/2}_{\mathrm{cut}}$ convergence of binding energy was observed.

Third, the GWSE calculations can be tricky to evaluate, due to their small magnitude and necessity to extrapolate with the basis-set size and with the k-point set. A tedious but reliable approach was to obtain converged RSE corrections separately and only use the computationally demanding GWSE routines to converge the $E_{GW\mathrm{SE}} - E_{\mathrm{RSE}}$ difference. The benefit is that the difference is small and thus the uncertainties of the extrapolated result are much lower than when

extrapolating the $E_{GW\mathrm{SE}}$ energy directly.

Finally, we performed the calculations and extrapolations with k-point grid and basis-set size using standard PAW potentials distributed with VASP. We then used a finite cells and k-point grids to obtain data for more computationally demanding hard PAW potentials to assess the quality of the standard PAW data. The effects turned out to be non-negligible, using the standard PAW potentials instead of the more reliable hard PAW potentials leads to errors within 1 kJ/mol for dispersion dominated systems. Errors can be larger for hydrogen bonded systems and for the GWSE correction, therefore one should use the hard PAWs for such systems or at least calculate similar correction to what done in P1.

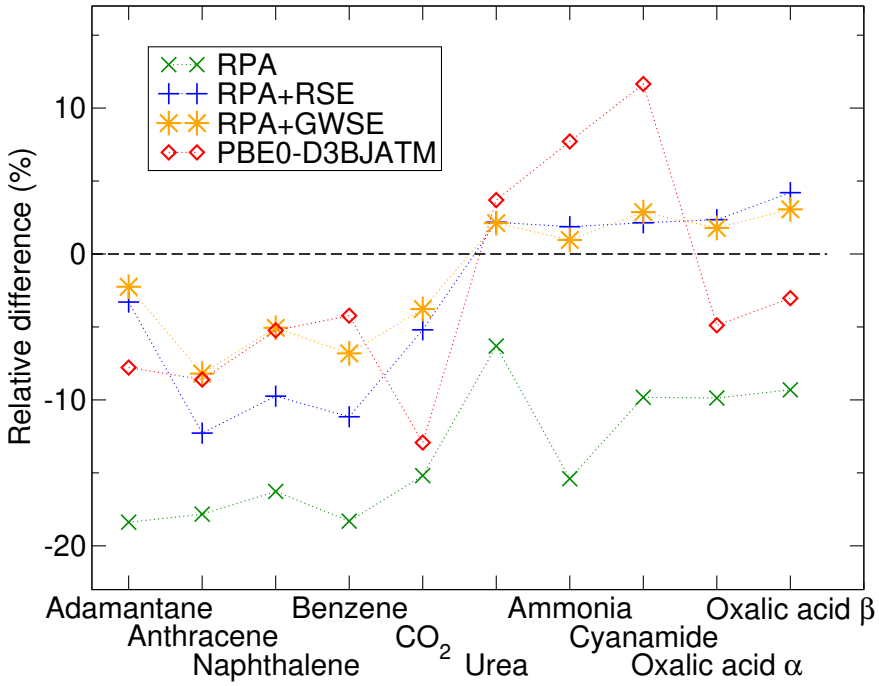

Figure 3.1: Relative differences from reference binding energies of molecular crystals for RPA-based methods and a dispersion corrected hybrid functional PBE0-D3 with three-body correlation and Becke-Johnson damping [87–89]. The values were obtained for ten solids from the C21/X23 test set [33, 34]. All the RPA calculations were based on the PBE functional [6] and results are shown for standard RPA as well as RPA with renormalised singles energies (RSE) [81] and GW singles energies (GWSE) [42].

Overall, we found that, when PBE states are used, singles corrected RPA provides binding energies with an average absolute error of 4.3 and 2.9 kJ/mol for RSE and GWSE corrections, respectively. The errors were lower than those observed for state-of-the-art dispersion corrected hybrid DFT functionals, namely PBE0 [87] with range-separated many-body dispersion (rsMBD) [90] and two versions of the D3 correction [88, 89]. Moreover, the errors were more consistent for all the RPA-based schemes than for the DFT methods, as shown in Fig. 3.1. This is especially visible for hydrogen bonded systems where RPA with RSE or GWSE give consistent errors around 3 %. In contrast, the errors for the DFT functional are both positive and negative.

We have used the RPA-based schemes to study molecular solids also in articles not included in this thesis. In Ref. [91], which was a result of collaboration with my former PhD supervisor and other colleagues, we compared various high quality methods for the prediction of binding energies of molecular solids. RPA+GWSE based on PBE orbitals showed again errors consistent across different binding types and of only few percent when compared to the reference. Moreover, we have RPA+RSE based on SCAN orbitals to rank polymorphs in the 7th crystal structure prediction blind test. The publication is being prepared at the time of the writing of this thesis. We took part in second part of the blind test where we were given a hundred of crystal structures and were asked to rank them in energy. The structures differed in number of molecules per unit cell and the cell shape. This made it difficult to identify k-point grid setting that would give converged energy differences while still being computationally tractable. In a similar spirit, also the basis-set cut-off had to be chosen sufficiently large but not too large due to the quickly rising computational cost.

3.2.2 Molecular adsorption in zeolites

Adsorption energies of molecules on solid surfaces are properties for which RPA-based methods were very promising, as exemplified by the work of Schimka *et al.* [79] on the "CO puzzle" [92]. However, already in 2009 during my stay in Vienna with prof. Kresse we found that RPA underestimates the adsorption energy of water on NaCl(001), a system for which a reference was available [3], by around 10 %. The introduction of RSE and GWSE corrections to RPA then significantly reduced this issue [42]. This motivated us to assess the accuracy of RPA-based methods for adsorption in zeolites, that is, the work performed in P2.

From different possible zeolites we picked chabazite which has a small unit cell and the calculations are thus computationally cheap. We started from a model of chabazite used previously by Florian Göltl and co-workers [93]. The structure is visualised in Fig. 3.2, where we show approximately four unit cells of the material with methane adsorbed. The structure is formed by a network of oxygen connected silicon atoms with one silicon atom replaced by an aluminium atom. Moreover, there is one hydrogen atom bound to one oxygen adjacent to the aluminium and the adsorption site is on the thus formed -OH group.

To test the accuracy of RPA we initially wanted to use the MP2 approach as a reference. Zeolites have a strongly ionic character and an electronic band gap of around 9 eV [94]. This means that electron excitations are mostly local, not extended over several atoms, and the importance of screening is low. For this reason, MP2 can be expected to be a good reference method for the adsorption energy if the adsorbing molecule shares similar characteristics. Moreover, MP2 also seemed sufficient as previous calculations of adsorption energies of methane, ethane, and propane in chabazite showed differences of tens of percent between MP2 and RPA data [93].

For our test we selected seven small molecules: methane, ethane, ethylene, acetylene, propane, carbon dioxide, and water. From all these, we expected that MP2 will give less reliable adsorption energies only for ethylene and acetylene as they have π bonds with delocalised electrons. The largest surprise came very early: the adsorption energies obtained by MP2 and singles-corrected RPA were

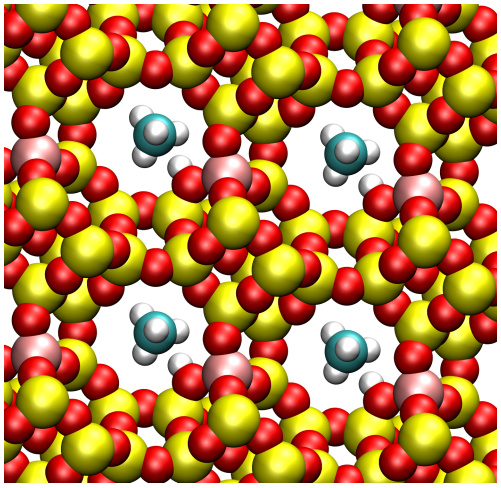

Figure 3.2: Structure of solid chabazite model with adsorbed methane. The figure shows approximately four unit cells.

essentially on top of each other, see Fig. 3.3. The only differences could be observed for ethylene and acetylene for which the most plausible explanation was the larger error of MP2 due to the lack of screening. MP2 was only a good reference for RPA without singles, and, as expected, the RPA errors could be described as a rather consistent underbinding by around 10 %.

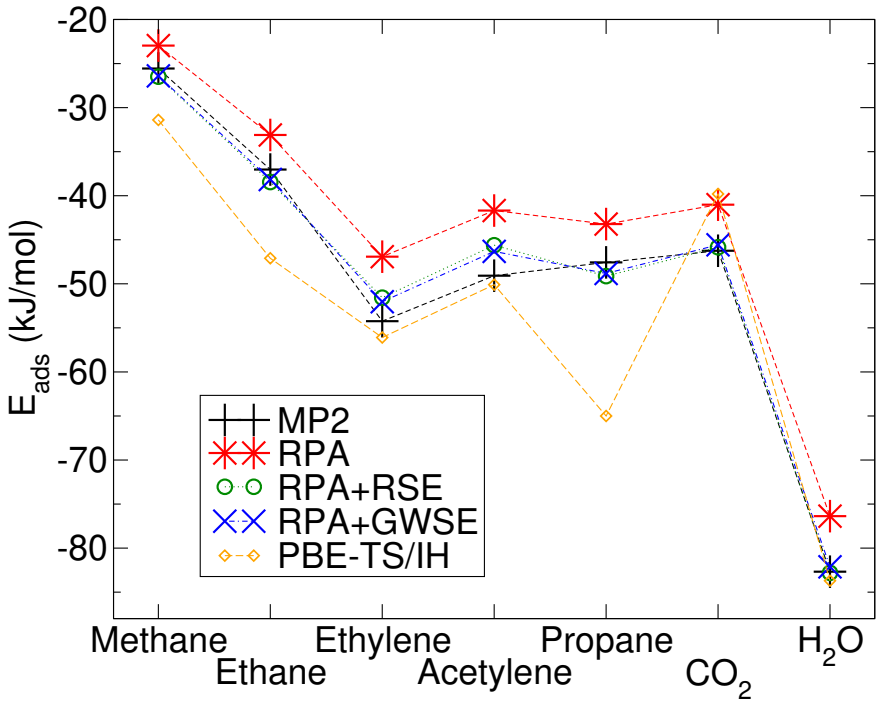

Figure 3.3: Adsorption energies of different molecules in bulk chabazite model with a structure $AlSi_{11}O_{24}H$.

To assess the errors of singles-corrected RPA a better reference than MP2 was needed. At that time some quantum Monte Carlo (QMC) scheme would be

the only and very computationally demanding possibility within periodic boundary conditions [95, 96]. Therefore, we opted for an alternative and commonly used approach: we selected several atoms around the adsorption site from the periodic material, capped broken bonds with hydrogens and formed a finite cluster. A finite cluster with two tetrahedral sites (2T, $AlSiO_7H_7$) was small enough to be treated by the CCSD(T) scheme in an explicitly correlated variant CCSD(T)(F12*), these calculations were done by Dr. Tew using the Turbomole code [97]. The differences of RPA+RSE from the CCSD(T) reference values were extremely low, below 3 % in absolute magnitude, as shown in Fig. 3.4. The RPA+RSE results were also more consistent than those of MP2. For example, MP2 underestimated the adsorption energies of ethane and propane by almost one kJ/mol (around 5 %) but overestimated those of ethylene and acetylene by almost two kJ/mol (around 10 %).

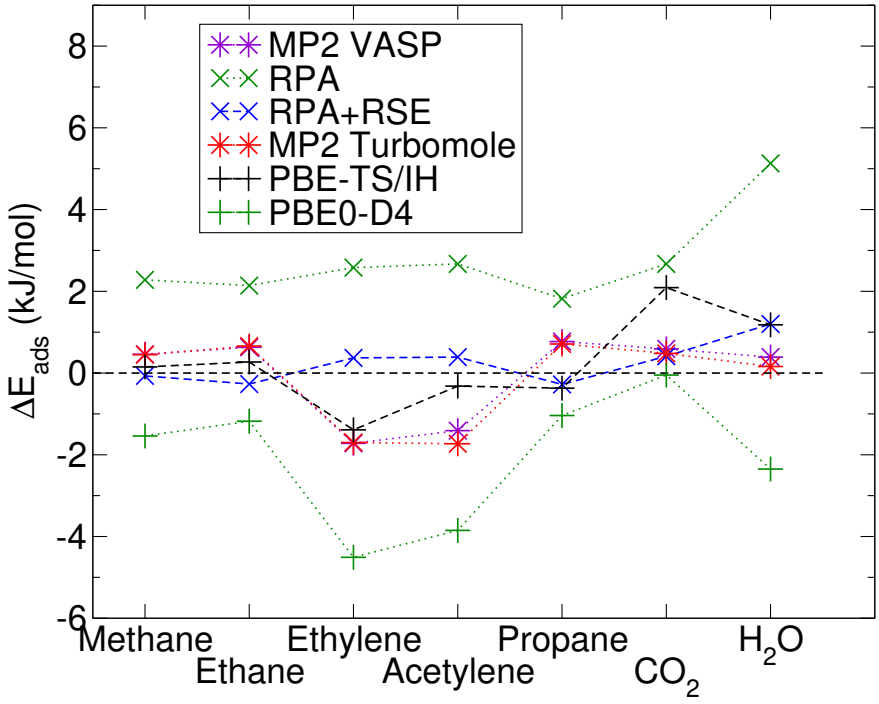

Figure 3.4: Deviation of adsorption energies on a 2T cluster for different methods and molecules. The reference scheme was explicitly correlated coupled clusters approach CCSD(T)(F12*) [97]. The values are in kJ/mol and positive number means underbinding and negative values overbinding.

The low errors of the adsorption energies obtained for the simple correlated methods (RPA+RSE, MP2) made it necessary to consider carefully the precision of the set-up. For example, we used hard PAW potentials in VASP as the binding energies were around 1 kJ/mol smaller in magnitude when standard PAW potentials were used. To be able to compare the VASP data to the reference values obtained for finite clusters using Gaussian basis sets the adsorption energies obtained with VASP had to be extrapolated with the basis-set cut-off and unit cell size. This again required a rather careful and meticulous approach but in the end we found a good agreement between the MP2 adsorption energies obtained by VASP and Turbomole. This is demonstrated in Fig. 3.4 where one can see

that the two approaches (data called "MP2 VASP" and "MP2 Turbomole") differ marginally, by few tenths of a kJ/mol at most. After thorough testing, we didn't compute the *GW*SE corrections for the finite cluster due to precision reasons. The computational cost and memory requirements of *GW*SE corrections made it impossible to obtain enough data to extrapolate reliably the *GW*SE correction with the basis-set size and cell volume. For illustration, for the 2T cluster, RPA+RSE differs from CCSD(T) by −0.1 and 0.4 kJ/mol for methane and acetylene, respectively, the reference energies are −11.0 and −16.7 kJ/mol. The *GW*SE correction beyond RSE would be a fraction of kJ/mol, and we didn't have resources to obtain enough data to extrapolate the values with sufficient precision.

Having adsorption energies of very high quality both for bulk and finite clusters we used them to assess the quality of prediction of different DFT-based methods. In general, approximate DFT functionals, including hybrids, can be nowadays quite easily used to obtain adsorption energies even for zeolites with a larger unit cell. However, DFT dispersion corrections can predict erroneous adsorption energies for zeolites if they do not consider the ionic character of the material. One of the schemes that was developed to take changes of density response due to changing oxidation state is the iterative Hirshfeld approach for the Tkatchenko-Scheffler method [21, 98]. Our tests showed a superiority of this method to other DFT schemes, see the data obtained for 2T cluster in Fig. 3.4. Moreover, a D4 correction of Grimme and co-workers [99, 100] turned out to introduce only a minor change compared to its D3 predecessor [88].

Overall, RPA with RSE or *GW*SE corrections turned out to provide adsorption energies with a very high accuracy and with an acceptable computational cost. Over the years, similar observations were made in other cases as well by other researchers, for example for water on graphene or hexagonal boron nitride [80, 101]. We have also used singles corrected RPA for zeolites in three papers not included in this thesis. First, to study the adsorption behaviour of carbon monoxide in a zeolite with a ferrierite structure [102], then to analyse adsorption properties of various molecules in Ref. [103] and finally to compare stabilities of various structures of zeolites relevant for catalysis in Ref. [104].

3.3 Tests of many-body energies

The small errors of RPA with singles corrections, either RSE or *GW*SE, that we observed in P1 and P2 confirmed the high promise of RPA for highly accurate calculations of binding or adsorption energies. Moreover, they suggested that RPA with singles can be a good starting point for the correction scheme given by Eq. 3.2 mentioned earlier in this Chapter. That is, that to increase the accuracy of the binding energy predicted by RPA one would need to correct only a small number of two- or three-body energies with a higher-level approach. To find out if this is indeed so, one needs to understand the RPA errors in detail. To achieve this, we can perform many-body expansion to see if the errors of four- and higher-order errors are negligible and if for the two- and three-body terms the RPA errors are dominated by a small set of contributions. To understand the many-body errors of RPA we used the 3B-69 test set of Řezáč and co-workers which is a set for testing three-body energies [82]. The results were published in

P3 and are discussed in the next Subsection. Furthermore, we performed MBE for methane clathrate and four molecular solids, obtaining reference CCSD(T) energies for two-, three-, and four-body energies and tested the accuracy of various RPA-based schemes. This work, published in P4 for methane clathrate and P5 for the solids, then forms the subsequent parts of this Chapter.

3.3.1 Accuracy of RPA for three-body energies

When it comes to understanding the errors of theoretical methods for predicting binding energies, we often focus on binding energies of dimers, larger clusters, or solids. There are several tests for the dimers, such as S22, S66, or X40 from prof. Hobza's group [105–107] or those in the "GMTKN" databases of Goerigk, Grimme, and co-workers [108–111]. For larger clusters, the binding energies of water clusters are available [112] and reference binding energies are available also for molecular solids [33, 34]. While testing on the binding energies is important and useful, more information can be obtained for clusters and solids by testing various many-body energies [113, 114]. This can uncover if a good performance for the total binding energy is a result of some cancellation of errors. Despite this importance, reference data for three-body interactions are comparatively rare, the (most likely) only general test set is the 3B-69 of Řezáč and co-workers [82]. We used this test set as a first step on our way to understand the accuracy of RPA for molecular solids. As with the S22 or S66 sets also the 3B-69 test set contains trimers of different molecules and with different dominating contribution to the binding. In Fig. 3.5 we show one of the trimers from the set, succinic anhydrate in a structure where both dispersion and electrostatic interactions are important.

It is known that different exchange-correlation (XC) DFT functionals produce very different predictions for three-body interactions [82, 113]. Moreover, the three-body errors can be rather large with relative errors easily reaching tens of percent [82]. In our set-up we evaluate RPA energies on DFT states and in P3 we were also interested how the RPA result varies when different XC functionals are used. For this we used three XC functionals: PBE [6], PBE0 [87], and SCAN0 [115, 116]. When going from PBE to PBE0 we test the effect of Fock exchange and the SCAN0 allows to test a recent meta-generalized gradient approximation. Note that in the paper as well as in the following we put the name of the XC functional that was used to obtain the RPA energies in brackets like this: RPA(XC). The singles corrections also depend on the DFT functional input, however, we do not repeat the name of the functional after RSE and only write RPA(XC)+RSE instead of RPA(XC)+RSE(XC). One of the downside of the 3B-69 test set is the lack of distance dependent data. Therefore, we calculated also binding energies of neon and argon trimers to gain information about the short- and long-range errors.

A crucial contribution to P3 was done by Dr. Modrzejewski who implemented RPA in his code focusing on scalability and high precision of the three-body interactions. To illustrate the latter point, consider that the RPA energy is often evaluated by integration over a frequency grid [48]. The frequency grid is generated based on the excitation energies in the system and therefore a different grid is usually used for the trimer, dimer, and monomer calculations. To reach a high precision of the three-body interactions, one should therefore use a very dense

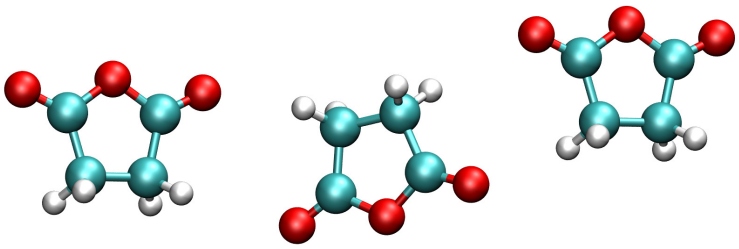

Figure 3.5: Trimer of succinic anhydride, structure 18c of the 3B-69 test set [82].

integration grid to avoid numerical errors from the integration. To reduce these errors the code of Dr. Modrejewski uses a common grid for all the calculations needed to obtain the trimer non-additive energy. This helps to keep the number of integration points low, below twenty. There are other technical points concerning the reduction of numerical errors and they are discussed in detail in P3.

One interesting observation of the study was that the singles corrections (at the RSE level) do not always improve the three-body energies. While we observed a substantial improvement for RPA based on PBE states upon addition of the singles corrections and a smaller reduction of errors for PBE0-based RPA, the errors were actually larger for RPA(SCAN0)+RSE compared to RPA(SCAN0), see Table 3.1. The reason for the worsening is difficult to extract from the data but could be simply due to the fact that RPA(SCAN0) was already giving low errors and the RSE correction is a too crude approximation and overcorrects the energies. Nevertheless, the overall accuracy of RPA was very satisfactory for the 3B-69 test set with RPA(SCAN0) and RPA(PBE0)+RSE giving an accuracy between that of MP3 and CCSD.

Table 3.1: Mean absolute errors (MAE) for the different methods on the whole 3B-69 data set, data in kJ/mol.

Method	MAE (kJ/mol)
RPA(PBE)	0.18
RPA(PBE)+RSE	0.11
RPA(PBE0)	0.11
RPA(PBE0)+RSE	0.10
RPA(SCAN0)	0.08
RPA(SCAN0)+RSE	0.12
MP2	0.19
MP3	0.11
CCSD	0.06

The distance-dependent three-body energies for neon and argon trimers showed also interesting results. A clear observation from these calculations was that the three-body errors of DFT functionals largely transfer to errors of the subsequent RPA calculations. This is illustrated in Fig. 3.6 where we compare the three-body energies obtained with different XC functionals, the singles-corrected

RPA energies calculated on top of these functionals, and the CCSD(T) data which we take as a reference. The structure is an isosceles triangle with R being the length of the base and the base angles are 45°. Both for DFT and the RPA+RSE data the errors are the largest for PBE and consequently decrease for PBE0, SCAN, and SCAN0. In an afterthought, the inheritance of the errors by RPA from the DFT calculation makes sense, large DFT errors are likely related to errors in density and the states, i.e., quantities that form the input for RPA calculations.

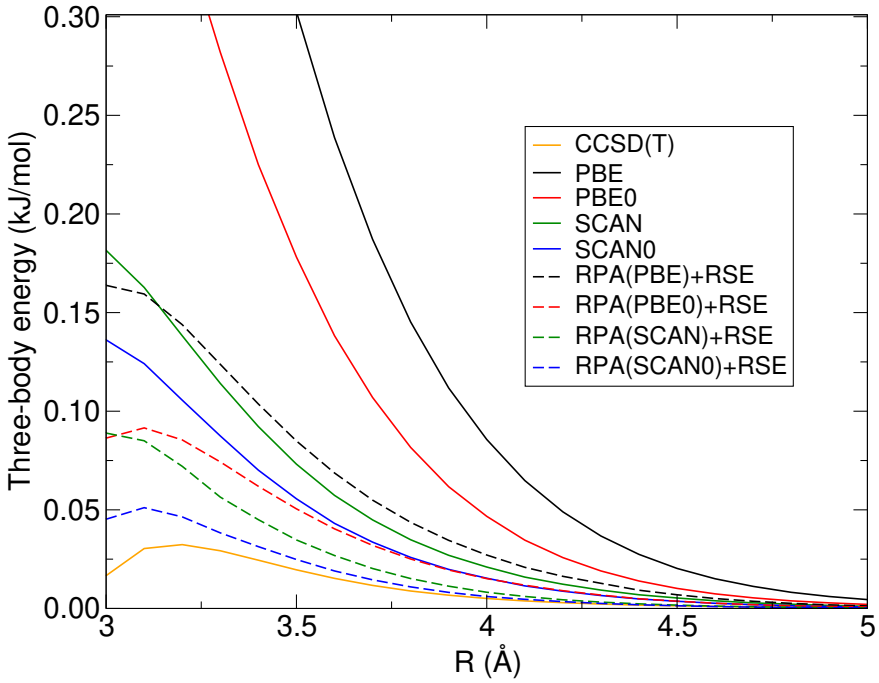

Figure 3.6: Three-body energies of neon trimer for CCSD(T), different exchange-correlation functionals, and singles-corrected RPA on top of these functionals. The structure is an isosceles triangle with base angles of 45° and R being the base length.

Another point that became clear is that the errors for different input states can be deduced already from the EXX data. That is, one doesn't need to compare the RPA energies to reference and can compare EXX to HF to get a basic understanding of the magnitude of the three-body errors. Strictly speaking, the HF should not be considered as a reference for EXX and only the total energies should be compared. However, we observed that if EXX is far from the HF three-body energies, the RPA three-body energies are far from CCSD(T) as well. This point is demonstrated in Fig. 3.7 in which we show the EXX and EXX+RSE data obtained for different XC functionals. One can see that the ordering of the functionals is the same as in Fig. 3.6 with SCAN0 again being the closest to the HF three-body curve. Therefore, the three-body energies seem to offer an independent and very simple and fast test of quality of XC functionals.

Overall, we saw that running RPA using PBE input leads to rather large errors and the singles corrections are crucial to improve the quality of this method. The PBE XC functional is commonly used to provide input states for RPA, at least

in solid state calculations. Our results then highlight that this might be far from optimal choice. However, it is possible that the three-body errors are either balancing some errors in other many-body terms or they are comparatively low to errors in two-body terms. The results presented next for clathrate and molecular solids shed more light on this issue.

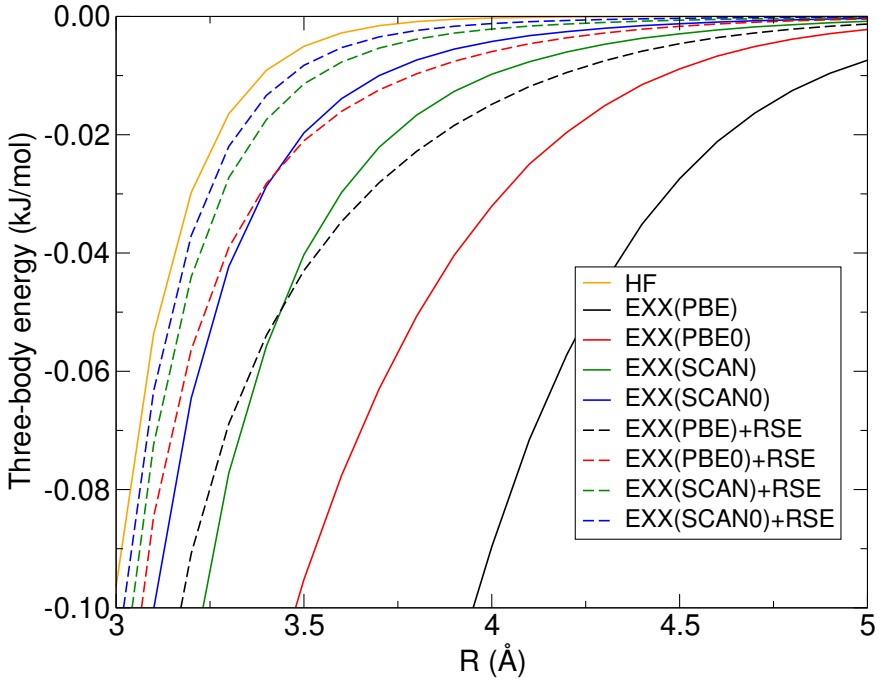

Figure 3.7: Three-body energy of neon trimer for Hartree-Fock (HF), exact exchange (EXX), and EXX with renormalized singles corrections (RSE). The EXX and RSE data were evaluated using different exchange-correlation functionals. The structure is an isosceles triangle with base angles of 45° and R being the base length.

3.3.2 Methane clathrate

In this part I discuss results of tests that we performed on binding energy of methane in methane clathrate. Methane clathrates are materials present on the Earth in seas and oceans where they store large amounts of methane. Their structure is formed by water cages in which methane, or other molecule for different clathrates, is present, see the structure on the left of Fig. 3.8. One of the properties of interest is the binding energy of methane with the water cage [117], E_{bind}, that is obtained as

$$E_{\text{bind}} = E_{\text{cage+CH}_4} - E_{\text{cage}} - E_{\text{CH}_4}\,, \tag{3.4}$$

where the energies on the right hand side are the energy of methane in the cage, the energy of the empty cage, and the energy of isolated molecule, respectively. From the point of view of methods development the binding energy of methane in clathrate represents a nice example of a property which is difficult to be described

accurately with DFT-based methods [117]. For example, the optPBE-vdW functional which performs well on molecular dimers [20] or ice structures [118] predicts E_{bind} twice as large as the reference [117].

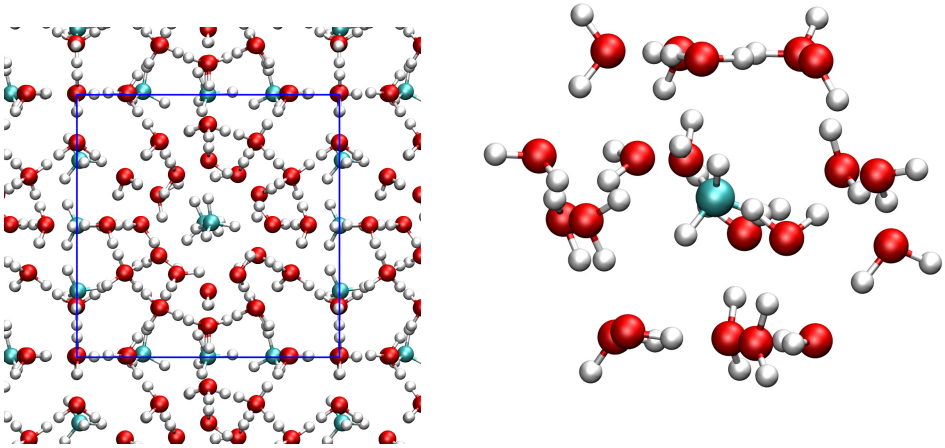

Figure 3.8: Structure of methane clathrate crystal on the left and methane clathrate cluster on the right.

To test RPA and other methods we used a finite cluster model of the clathrate which consists of a single methane molecule bound in a cage of twenty water molecules, as shown on the right hand side of Fig. 3.8. The cluster was used previously by Deible *et al.* [119] who assessed the performance of different DFT functionals for predicting of E_{bind}. There are several benefits of using the cluster compared to the periodic solid. Most of all, the number of MBE contributions that needs to be evaluated is finite. Moreover, the system is small enough so that $E_{\text{cage+CH}_4}$ and E_{cage} can be calculated in a single calculation for DFT, HF, RPA, and MP2. Therefore, the binding energy E_{bind} can be evaluated directly using so-called supermolecular approach. The name "supermolecular" is used because the water cage is considered as a single entity. The results of the supermolecular approach can be then used to compare to the results obtained with MBE. This serves to check the convergence of MBE and also to verify the quality of the set-up. Both the supermolecular approach and MBE for calculating E_{bind} are sketched in Fig. 3.9.

The first step that we needed for the testing was to obtain the reference MBE contributions. Specifically, we needed the interaction energies of 20 methane-water dimers and the non-additive contributions of 190 trimers and 1140 tetramers. To have values which are both highly accurate and precise we used the CCSD(T) scheme and performed a careful analysis of the convergence of the MBE contributions with respect to the basis set size. The details are discussed in P4, but the precision can be illustrated by the fact that E_{bind} obtained by the supermolecular and MBE schemes differ by only 0.05 kJ/mol for the HF method and by 0.1 kJ/mol for MP2. Our final CCSD(T) binding energy equals -19.7 kJ/mol and confirms a good accuracy of a previous estimate of -20.4 kJ/mol obtained by Lao and Herbert using the supermolecular approach and the domain-based local pair natural orbital coupled clusters scheme [DLPNO-CCSD(T)] [120, 121].

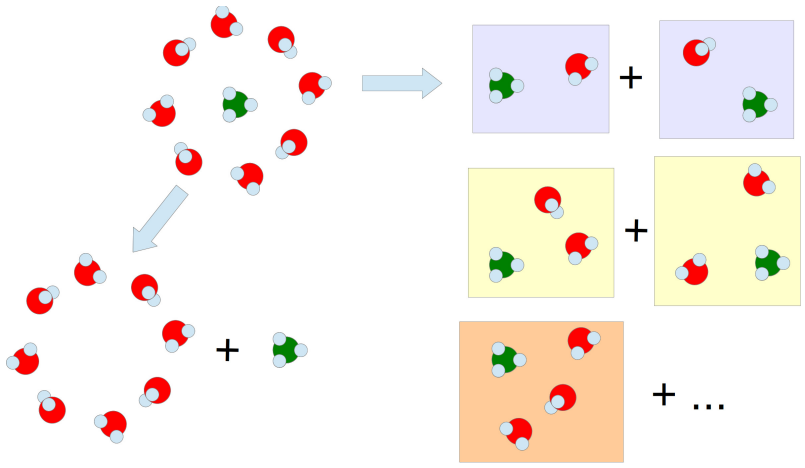

Figure 3.9: Two possible ways to calculate the binding energy: a supermolecular approach on the left and many-body expansion on the right.

In the following we briefly mention some of our observations made for the calculations and results of the clathrate. These are both technical or related to the accuracy of the different methods.

One technical point concerns the calculation of reference quality binding energies using MBE. We usually use the CCSD(T) scheme to obtain reference binding energies and use or at least test its explicitly correlated (F12) version to speed up convergence with the basis-set size [97, 122, 123]. However, the F12 methods affect only the CCSD energy and not directly the triples (T) term. To reduce the basis-set size dependence of the (T) term one can scale it using ratio of MP2-F12 and MP2 energies [124, 125]. This is called scaling and it is quite efficient for the dimer or two-body interaction energies. However, for trimers and tetramers, we observe large errors when the scaled energies from the output are used directly, see Table 3.2. The errors are smaller when a common scaling factor is used [125]. However, the most reliable way to obtain the three- and four-body (T) terms is to use the bare (T) contributions directly as they have almost no dependence on the basis-set size (Table 3.2). This approach is also the simplest as, compared to using a common scaling factor, no additional data processing needs to be performed.

Table 3.2: The (T) many-body contributions to the binding energy of methane in clathrate obtained for different basis-sets. The $(T)_{\text{unscaled}}$ data show the bare (T) terms as obtained by Molpro, $(T)_{\text{indiv. scaled}}$ and $(T)_{\text{common scaled}}$ give (T) contributions which were scaled with MP2-F12/MP2 energy factor by the programme and externally using a common scaling factor [125].

Scaling	Two-body		Three-body		Four-body	
	AVTZ	AVQZ	AVTZ	AVQZ	AVDZ	AVTZ
$(T)_{\text{unscaled}}$	−6.33	−6.65	0.84	0.85	0.08	0.09
$(T)_{\text{indiv. scaled}}$	−6.80	−6.85	0.19	0.48	−6.19	3.75
$(T)_{\text{common scaled}}$	−6.94	−6.93	0.93	0.88	0.11	0.10

As with the 3B-69 reference data we used the clathrate system to test how the RPA energy depends on the input states provided by DFT XC functionals. Specifically, we used PBE, SCAN, PBE0, and SCAN0 which again allowed an assessment of an effect of using Fock exchange in the XC formula. Moreover, we also obtained two- and three-body contributions for RPA based on HF states and two-body contributions based on two optimized effective potentials (OEP). The OEP calculations were done by Dr. Śmiga from Toruń and were done at the level of exact exchange (OEPX) [126] or including second-order perturbation theory terms (OEP2-sc) [127–129].

For the 2-body interactions, the results are somewhat expected (Table 6 in P4), RPA without singles underbinds and this underbinding is corrected by the singles corrections. For hybrids (PBE0 and SCAN0), the singles have a smaller effect as the starting Hamiltonian is already close to the HF one. Interestingly, the OEP2-sc gave EXX, RSE, and RPA 2-body energies similar to those obtained by using SCAN states and both approaches also gave similar energy gap between occupied and virtual states for water and methane. However, the energy of the highest occupied state was −14.0 eV for OEP2-sc while it was only −9.9 eV for SCAN. These data support the idea that the good performance of RPA based on (meta-)GGA states comes from the good estimation of the electronic gap (governing the excitation energies) rather than correct estimation of the orbital energies.

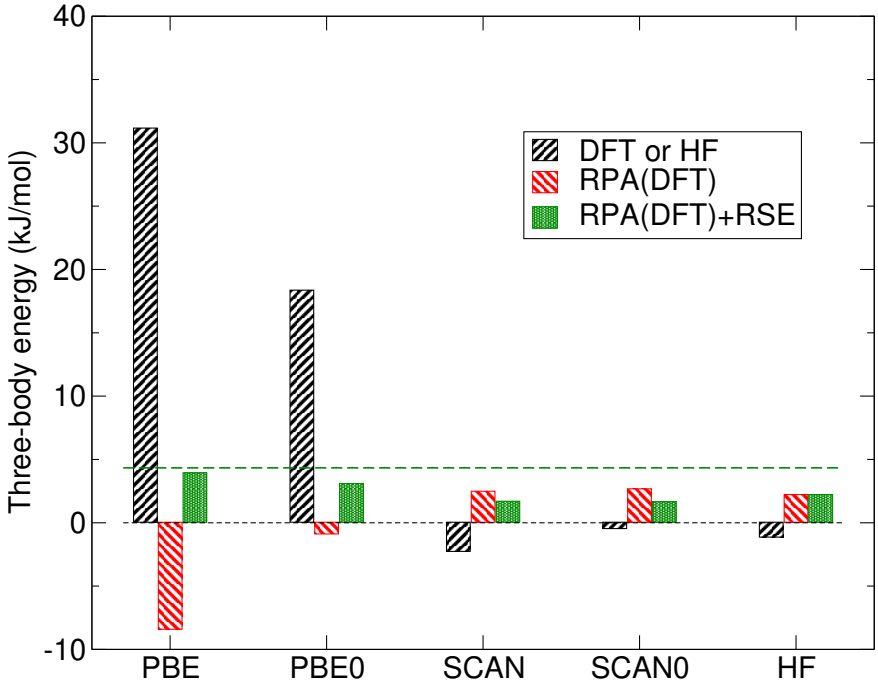

Figure 3.10: Three-body energies for interaction energy of methane with water clathrate cage obtained with different methods. The green dashed line gives the reference value of 4.4 kJ/mol obtained using CCSD(T).

From the 3B-69 test set we know that the RPA errors for three-body energies are strongly affected by the errors of the DFT functionals used to provide the input states. The observations that we made for clathrate are consistent with

the 3B-69 results. To illustrate the results for clathrate, we show the three-body energies obtained by the different DFT functionals and RPA without and with singles in Fig. 3.10. The three-body energies obtained by PBE and PBE0, equal to 31.2 and 18.4 kJ/mol, respectively, grossly overestimate the reference CCSD(T) value of 4.4 kJ/mol which we show with the dashed line. In contrast, the SCAN and SCAN0 data are −2.3 kJ/mol and −0.5 kJ/mol, respectively, straddling the HF value of −1.1 kJ/mol. As a consequence, the EXX, RSE, and RPA contributions are generally larger when PBE states are used compared to using SCAN inputs. For example, RPA correlation energy is 9.8 kJ/mol for PBE input but only 2.7 kJ/mol for SCAN. Despite the larger individual terms, the RPA+RSE three-body energy evaluated on PBE states, equal to 3.9 kJ/mol, is the one closest to the reference. However, without singles the RPA(PBE) three-body energy equals to −8.0 kJ/mol and has the largest error of all the methods tested. The results are much less sensitive for the SCAN inputs and we believe that this is a piece of evidence that makes SCAN (or similar methods with low three-body errors) more suitable to be used as input for RPA.

We found one rather surprising result when testing the predictions of more recent dispersion corrections, such as the D4 from Grimme's group [99, 100], on the clathrate data. Specifically, the three- and four-body contributions were much overestimated in magnitude when correcting HF data and here we briefly discuss a plausible explanation. The two-, three-, and four-body energies for SCAN and HF are given in Table 3.3 in the "Energy" columns together with the corresponding D4 corrections. As mentioned, the SCAN and HF three-body and, to a lesser extent, the four-body energies are close to each other. In contrast, for two-body terms SCAN gives strong binding while HF is repulsive. In both cases, this is the expected behaviour [115]. Therefore, to obtain good two-body energies the dispersion correction needs to be small for SCAN and large for HF. To achieve this, the damping function must reduce the dispersion term substantially more for SCAN than for HF. As a consequence, for SCAN the dispersion terms are small also for the three- and four-body terms while for HF the lack of substantial damping most likely leads to the large many-body D4 energies. While one might not be interested in using dispersion-corrected HF, it is a reminder that one should also test at least three-body energies and not only the total or two-body terms when developing DFT methods and dispersion corrections.

Table 3.3: The SCAN and Hartree-Fock many-body contributions and the respective dispersion corrections as obtained by the D4 method. Data in kJ/mol.

Method	Two-body		Three-body		Four-body	
	Energy	D4	Energy	D4	Energy	D4
SCAN	−21.6	−6.9	−2.3	0.4	7.00	0.3
HF	16.0	−30.3	−1.1	−54.7	2.3	48.6

The evaluation of the three- and four-body terms tends to be more time consuming compared to the evaluation of the two-body interactions. This is due to their larger number as well as larger fragment size. The fact that smaller basis sets suffice to achieve their convergence compared to the two-body terms helps

to reduce the computational demands but typically not enough to make the cost of the two-body terms dominant. One way to reduce the overall computational cost of the binding energy evaluation is to obtain part of the contributions using less demanding method. This can be done either for specific terms or using the correction scheme described by Eq. 3.2 where one starts from a supermolecular calculation. For a method to be useful within the correction scheme, it needs to deviate as little as possible from the demanding scheme which it replaces. For clathrate, the RPA(PBE0)+RSE and RPA(SCAN0) were the methods that gave the smallest errors for the three- and four-body terms. In contrast, the large three- and four-body errors of RPA(PBE) make it unsuitable for the correction scheme.

To compare the results obtained using the correction scheme for different methods we created subgroups for the three- and four-body terms. Specifically, we divided the fragments according to the number of hydrogen-bonded water dimers they contain. The reason is that the clusters containing proximate molecules tend to have larger errors. Therefore, we have two subgroups of trimers: "$L_{3b\,1hb}$" and "$L_{3b\,0hb}$" where in the first group the water molecules are hydrogen bonded and there is no hydrogen bond in the second group. Similarly we have three subgroups of tetramers. According to the correction scheme, we start with the supermolecular binding energy obtained at the RPA level and consecutively replace the RPA terms with their CCSD(T) counterparts. The convergence of the binding energy with the subset included at the CCSD(T) level is shown in Fig. 3.11. Apart from the two RPA-based results we also obtained data for MP2 combined with CCSD(T) (CC/MP2 data) and supermolecular binding energy at HF level combined with MBE expansion at the CCSD(T) level (HF/CC data). All the approaches converge faster with the fragment set included than the bare MBE at the CCSD(T) level. Also the data from the correction scheme are reasonably close (within 1 kJ/mol) to the converged value when the contributions of hydrogen-bonded trimers (3b 1hb set) are included at the CCSD(T) level.

We note that a large part of the errors of MP2 comes from missing three-body correlations and adding it via an approximate term would further improve the MP2 data [130]. Finally, we note that standard DFT functionals have substantial errors for the many-body contributions and they would hardly be useful here. A possible exception are long-range corrected functionals but that needs to be explored [131].

Overall, the data obtained for methane clathrate clearly show that methods that give low error in the supermolecular approach can exhibit large errors in the many-body contributions. Moreover, the SCAN states seem to be preferable to PBE ones due to their lower many-body errors. Finally, the clathrate cluster was a useful system on which we thoroughly tested the computational set-up not only for RPA but also for calculation of reference CCSD(T) energies.

3.3.3 Many-body expansion of binding energies of solids

In the final results part of this Chapter I will discuss MBE calculations for crystals of short hydrocarbons: ethane, ethylene, and acetylene that we published in P5. This work allowed us to test and use our methodology for molecular crystals, systems with an infinite number of many-body contributions. The molecules in

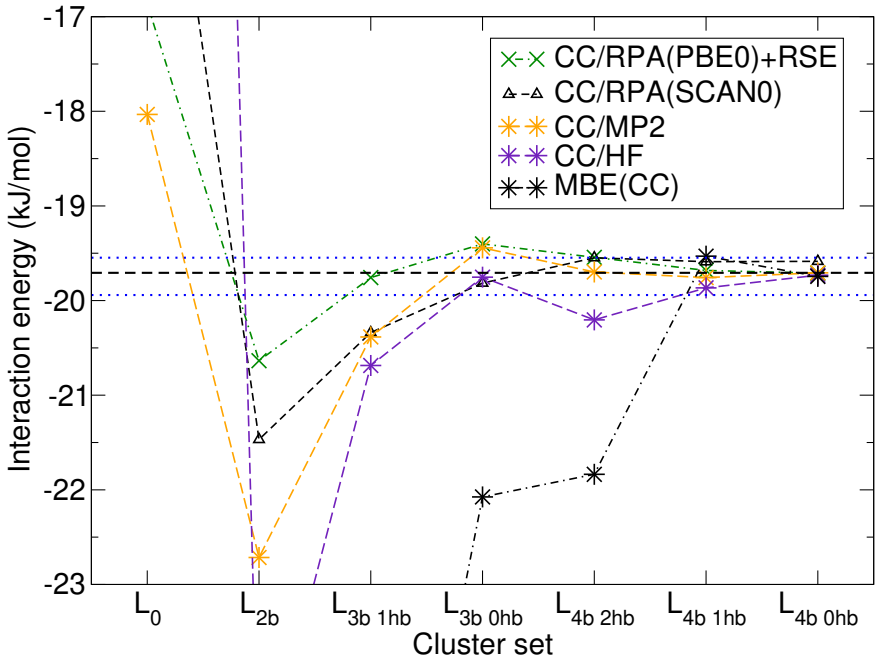

Figure 3.11: Binding energy of methane in clathrate for different terms included in many-body expansion and the correction scheme. The terms are divided according to the n-body order ($n = 2$, 3, and 4) and the number of hydrogen bonds (hb) in the fragments. The standard CCSD(T) many-body expansion is labeled MBE(CC) and it's final value of -19.7 kJ/mol is indicated by the black dashed line with the blue dotted lines indicating values one percent larger and smaller. The correction scheme starts from a supermolecular interaction energy obtained for a method `X` and many-body expansion of a difference between CCSD(T) and `X`. For the simpler method `X` we use two RPA approaches, MP2, and Hartree-Fock.

these crystals have a small size and thus the computational cost of the many-body terms is affordable. Moreover, the molecules have weak but different electrostatic moments which we expected to affect the convergence of the MBE contributions with cut-off distance.

The large number of fragment contributions that we needed to evaluate required a substantial change to the way we prepare and process the data. As a consequence we wrote a Python library called `mbelib` [132] in which we collected subroutines and processing scripts that we had, made them more consistent and added several other functions. With this library, we can take a structure of a crystal (in VASP `POSCAR` format), extract monomers, create a list of dimers, trimers, and tetramers within some cut-off, prepare input files for Molpro, and process output files. Moreover, we can divide the data into separate groups based on intermolecular distances occurring in the fragments, that is, we can group the fragments according to the number of nearest neighbours occurring in them. This is very useful for isolating and understanding the asymptotic behaviour of the errors.

As with the work on methane clathrate we needed to calculate the reference MBE energies, for which we used again CCSD(T). An important point for the evaluation of the reference energies was the size of basis-set that needs to be used

to obtain converged results. A large basis set should yield more precise values but the computational cost can be too large to be practical. In P5 we showed that for two-body contributions large basis sets are needed only for small intermolecular distances. This is understandable as with increasing distance between the molecules the perturbing fields become more and more homogeneous. Here we show an additional example that a similar situation occurs also for trimers.

The three-body energies have a smaller dependence on the basis-set size compared to the two-body terms, at least when all the energies needed to evaluate the three-body contribution of a single fragment are evaluated in the same basis set. For ethylene, we observed differences up to few percent between AVDZ and AVTZ energies. To understand how different trimers contribute to the error we divided them into groups according to the number of nearest neighbours in the trimer. For ethylene, we classify the molecules as nearest neighbours if their average distance is below 5 Å.[1] We then have a small group of trimers with three nearest neighbours and a larger one with two neighbours. The groups with one and zero nearest neighbours are infinite in principle, they are only finite due to the finite cut-off distance imposed on the trimers.

A useful way to look at the errors is to sum the basis-set errors from the largest cut-off distance to the lowest one. In our scripts this is called reverse summation. The differences between AVDZ and AVTZ values for all the trimers and the subgroups are plotted in Fig. 3.12 for CCSD on the left and CCSD-F12b on the right. One should read the graphs from the right hand side, the black line (all data) starts at zero for the largest cut-off and as the cut-off decreases, it accumulates the differences obtained for different trimers. The value for the smallest cut-off is then the final difference between the AVDZ and AVTZ three-body energies. The data show that the basis-set dependence is indeed very small for the trimers with no nearest neighbours, that is, the red line is close to zero for all the cut-offs for which data exist. This is true both for CCSD and its explicitly correlated variant. The explicitly corrected variant clearly reduces the basis-set errors for the sets with a single or two nearest neighbours. This is demonstrated by smaller changes of the values for the CCSD-F12b data compared to CCSD. The contribution to the basis set error due to the set of trimers with three nearest neighbours (orange line) is similar for CCSD and CCSD-F12b. We note that in P5 we surprisingly observed a larger basis-set difference for CCSD-F12b, around 0.03 kJ/mol compared to ≈0.01 kJ/mol for CCSD. The analysis in Fig. 3.12 clearly shows that this is not because of smaller individual differences for CCSD but due to a larger cancellation of positive and negative values.

Our main goal while performing the study was to identify the origins of the good performance of the RPA+RSE scheme based on PBE input states observed previously [91, 133]. The many-body analysis in P5 showed that, at least for the systems considered, the good performance of the RPA(PBE)+RSE is mainly due to error cancellations. Specifically, the two- and four- body terms are too repulsive and the three-body terms are too attractive, as shown for ethylene in Fig. 3.13. While this is rather disappointing, it's rather remarkable how well this error cancellation works across the range of systems studied in Refs. [133] and [91].

[1] Actually, there are no dimers with distance above 5 and below 6 Å, the second nearest neighbours have then distances above 6 Å.

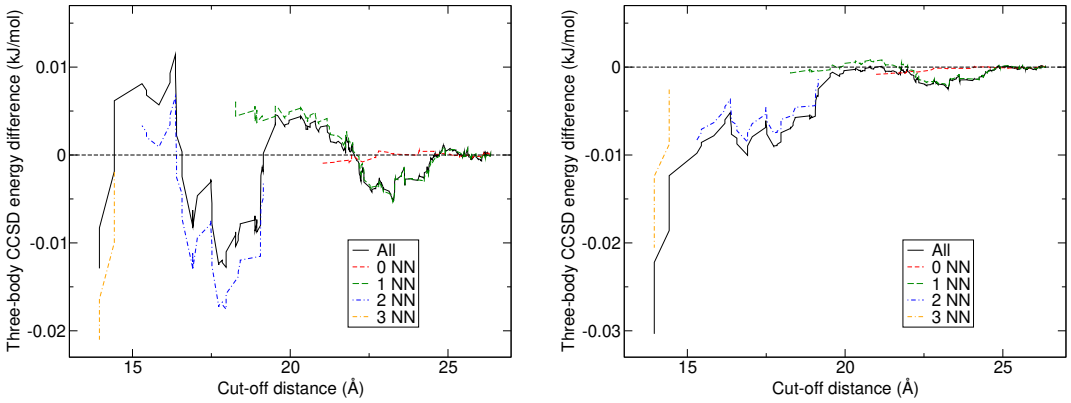

Figure 3.12: Difference between three-body energies obtained by AVDZ and AVTZ basis sets for CCSD on the left and CCSD-F12b on the right. Apart from data for all the trimers we show the data for groups of trimers created according to different number of nearest neighbours (NN).

As with clathrate, the reason for the large many-body errors of RPA(PBE) and RPA(PBE)+RSE can be traced down to the many-body errors of input PBE states. In fact, for PBE the three-body energy of ethylene is equal to 6.86 kJ/mol, far away from the CCSD(T) reference of 1.56 kJ/mol or the HF value of −0.38 kJ/mol. This difference is too large to be corrected by standard dispersion corrections that include three-body terms [88, 100, 134]. Therefore, our data support the idea that the three-body errors of molecular clusters and solids are a crucial tool for assessing or developing DFT functionals and dispersion corrections [82, 113].

In P3 and P4 we observed smaller many-body errors for RPA based on the SCAN functional and the data calculated for molecular solids in P5 show a similar behaviour, see Fig. 3.13. Moreover, when the RSE corrections are added to RPA(SCAN), the errors for the three-body terms increase, again in line with results obtained in P3 and P4. When compared to RPA(PBE), the error for the two-body terms of RPA(SCAN) doesn't decrease as much as the errors for the three- and four-body terms. As a consequence, there is much less error cancellation between the different MBE orders and the overall RPA(SCAN) or RPA(SCAN)+RSE error is larger compared to their PBE-based counterparts.

The many-body contributions sum together hundreds of energies of individual fragments. Therefore, apart from looking at the errors of the many-body energies it is extremely useful to consider errors of the individual contributions. A convenient way to do this is to plot the many-body energy as a function of the cut-off distance. For the two-body terms we found that the errors of RPA methods are large when the intermolecular distance is small. We called this group "proximate" dimers and it included all dimers with intermolecular distance less than 10 Å. Moreover, as mentioned previously, the interaction energy in this group depends strongly on the basis-set size. In contrast, the RPA approaches exhibited low errors for the rest of the dimers, called the "distant" group. We can illustrate the good results for the distant dimers by the fact that the difference between CCSD and RPA(SCAN) two-body energies was 0.01 kJ/mol at most, the actual values being between 0.35 to 0.5 kJ/mol for the four different systems.

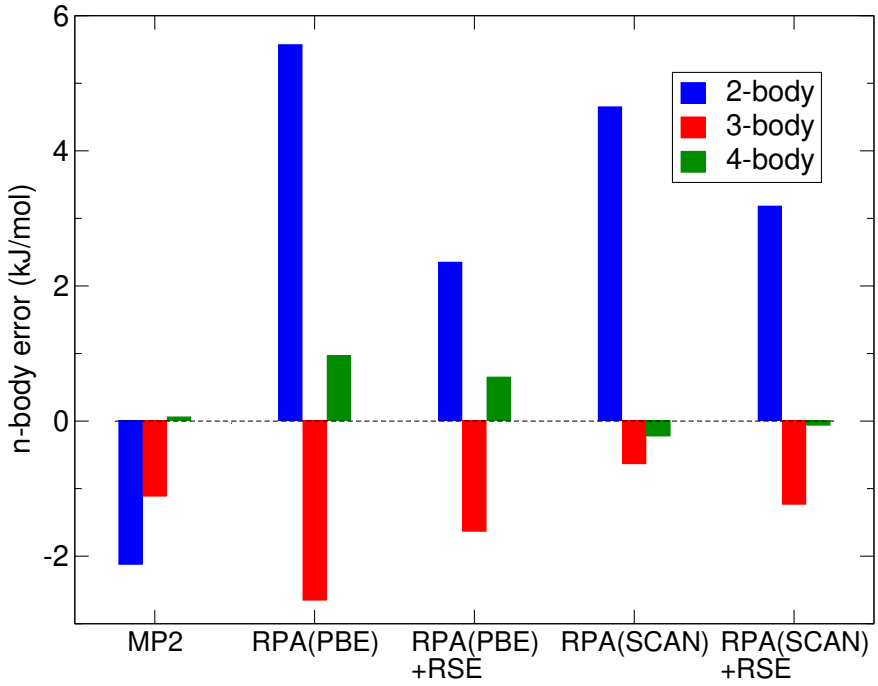

Figure 3.13: Errors of MP2 and different RPA schemes for the n-body contributions to binding energy of ethylene solid. Positive values mean that the terms are too repulsive compared to the CCSD(T) reference.

The (T) contribution is around 0.1 kJ/mol for the distant dimers and for all the systems so the CCSD(T) and RPA(SCAN) differ by approximately this amount. However, the close agreement of CCSD and RPA(SCAN) likely suggests that, at large separations, RPA with SCAN input states is a good approximation to CCSD.

The low errors of RPA-based methods for the distant dimers suggest that the three-body errors could be also low for separated trimers. To analyse this, it is again useful to divide the trimers into four groups according to the number of nearest neighbours in the trimer. We then expect that the errors will decrease with decreasing number of contacts in the fragment. The results that we found for ethane are shown in Fig. 3.14. Interestingly, the fragments with three NNs look to be described better by RPA than those from the group with two NNs. The trimers with a single NN contribute by around 0.1 kJ/mol to the overall error, which is still significant. As expected, the deviations of RPA from the reference are rather small for the trimers without any molecules in contact. However, the overall contribution of these trimers is almost negligible within the cut-off we used.

For comparison, we also included MP2 data in Fig. 3.14. MP2 correlation lacks three-body terms and the Hartree-Fock contributions are also small due to the small quadrupole moment of ethane [135]. Overall, the values are close to zero for all the fragments in all the groups. In fact, the MP2 line for the 0 NN group coincides with the line denoting zero, the contribution of this group is less than 0.001 kJ/mol. Of course, the MP2 results would be likely improved by adding an approximate three-body correlation term but we haven't tested this thoroughly

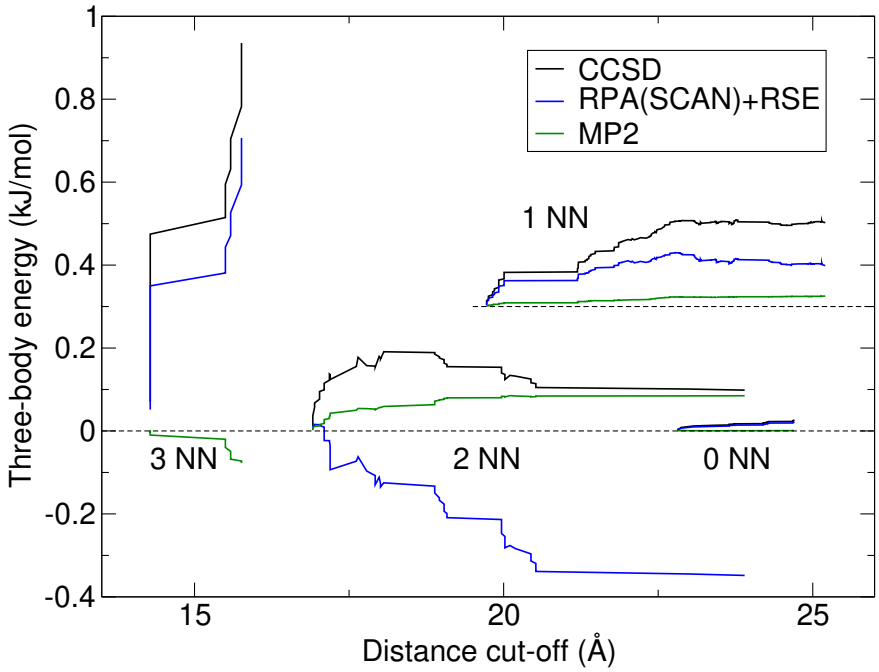

Figure 3.14: Three-body energy of ethane as a function of cut-off distance for different methods. The contributions were divided into subgroups according to number of nearest neighbours (NN) in the fragments with dividing distance of 6.1 Å. For clarity, the data for one NN was shifted upwards by 0.3 kJ/mol and the dashed lines denote zero.

yet [130, 136].

Our study was largely motivated by the possibility of using RPA in the correction scheme of Eq. 3.2 and the promising results obtained for clathrate. The results for molecular solids showed that neither PBE- nor SCAN-based RPA are suitable for the correction due to problems occurring for the three-body terms. There are two main issues that we identified. First, the difference between the reference data and RPA energies converges only slowly with the distance cut-off. This can be seen in Fig. 3.14: the distance between the RPA and CCSD lines for the 1 NN group visibly increases with cut-off distance. It means that the RPA terms are not a very good approximation of their CCSD counterparts and that many trimers would need to be included in the correction scheme to reduce the overall error. Second, the RPA energies have a much larger dependence on the basis-set size compared to the reference data. To illustrate this, we show the cut-off distance convergence of the RPA correlation energies in AVTZ and AVQZ basis sets together with extrapolation to the complete basis-set limit (CBS) for orthorhombic acetylene in Fig. 3.15. For comparison we also included the CCSD energies obtained using the AVDZ and AVTZ basis sets. Clearly the difference between the RPA data is much larger than the one for CCSD, even though the latter uses a smaller basis set. Moreover, as we show in P5, the slow convergence is not restricted to short trimer separations but also affects the data at larger distances. This means that the RPA energies require basis-set extrapolation for the three-body terms to reduce the basis-set errors. This makes the calculations rather costly from the computational point of view.

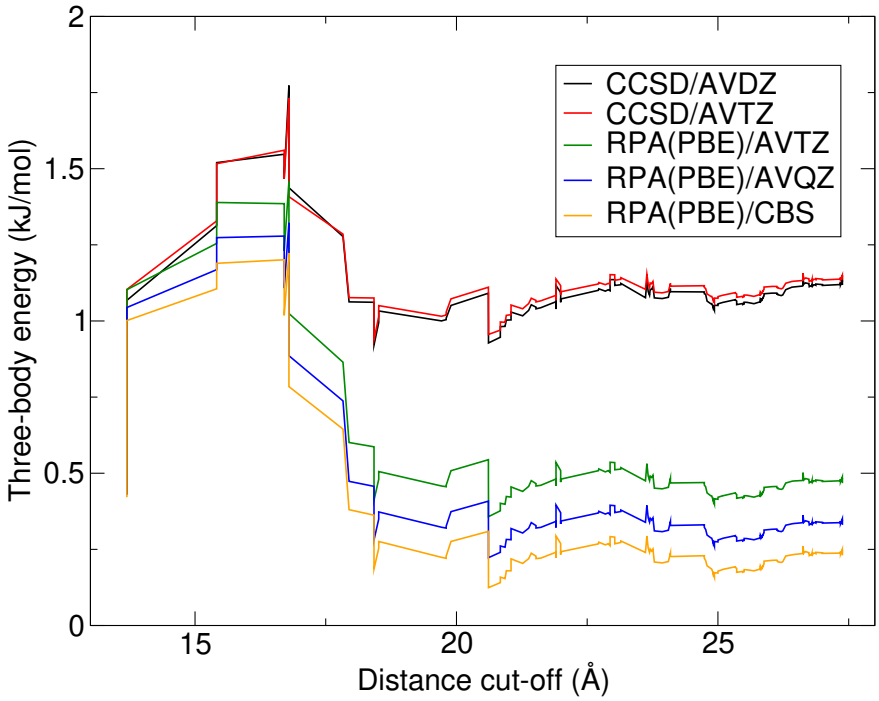

Figure 3.15: Three-body correlation energies obtained by CCSD and RPA as a function of the distance cut-off and for different basis sets for orthorhombic acetylene.

The reason for the slow basis-set convergence and deviations from CCSD is still not completely clear. However, our tests performed after completion of P5 suggest that part of the issue can be attributed to the use of DFT states instead of HF states employed by CCSD. An additional reason is the lack of second-order exchange terms which seem to reduce the basis-set dependence as well. We note that the latter point can be likely supported by analysing the leading terms of basis-set convergence for the appropriate energies, as done in Ref. [32].

Overall, in P5 we assessed the accuracy of different correlated methods using a newly generated test set of many-body energies of four solids. We identified that smaller and thus computationally cheaper basis sets are sufficient to generate the reference data when molecules in the fragment are not in contact. The other results were not so positive or encouraging. We found that the good performance of RPA(PBE)+RSE is mostly due to cancellation of errors between different orders of MBE. Three- and four-body errors of RPA are smaller when SCAN states are used, however, two other issues appear. Specifically, the three-body energies converge slowly with the basis-set size and their difference to CCSD(T) also converges slowly with the cut-off distance. Either of these issues make the SCAN- or PBE-based RPA unsuitable for the correction scheme.

3.4 Summary

Let me summarise this chapter by discussing the results in a broader view. The work performed for P1, P2 as well as related works by us and other researchers showed that RPA with singles corrections (RSE or GWSE) evaluated non-self-

consistently on PBE states predicts accurate binding energies and adsorption energies. The results are close to CCSD(T) or QMC reference and the method treats well both insulators as well as metals, as showed, for example, for adsorption of water molecule on graphene [101] and hexagonal boron nitride [80, 137]. This makes it preferred to use over simpler DFT functionals for such problems if the computational cost allows it. The analysis done in P4 and P5 showed that, unfortunately, the good performance for binding energies of molecular solids is due to error cancellations between different MBE orders. This makes RPA(PBE), with or without singles, unsuitable for the correction scheme of Eq. 3.2. That is, one can't improve the binding energies with few tens of CCSD(T) calculations for the most problematic terms.

The large many-body errors of RPA(PBE)+RSE are inherited from many-body errors of PBE. One can therefore expect that RPA based on DFT functionals with three- and four-body errors smaller than PBE will give improved results for the non-additive energies. In P3 and P4 we found that the three-body errors are lowered for hybrid functionals. In a work performed as a follow-up to P5 and for the same systems we also observe that hybrid functionals tend to reduce the many-body errors over the respective pure functionals. This is illustrated on many-body errors for ethylene crystal shown in Fig. 3.16. Interestingly, the four-body terms of RPA(SCAN0) are small, below 0.05 kJ/mol in absolute magnitude. In contrast both RPA(PBE0) and its singles corrected version give an error over 0.3 kJ/mol.

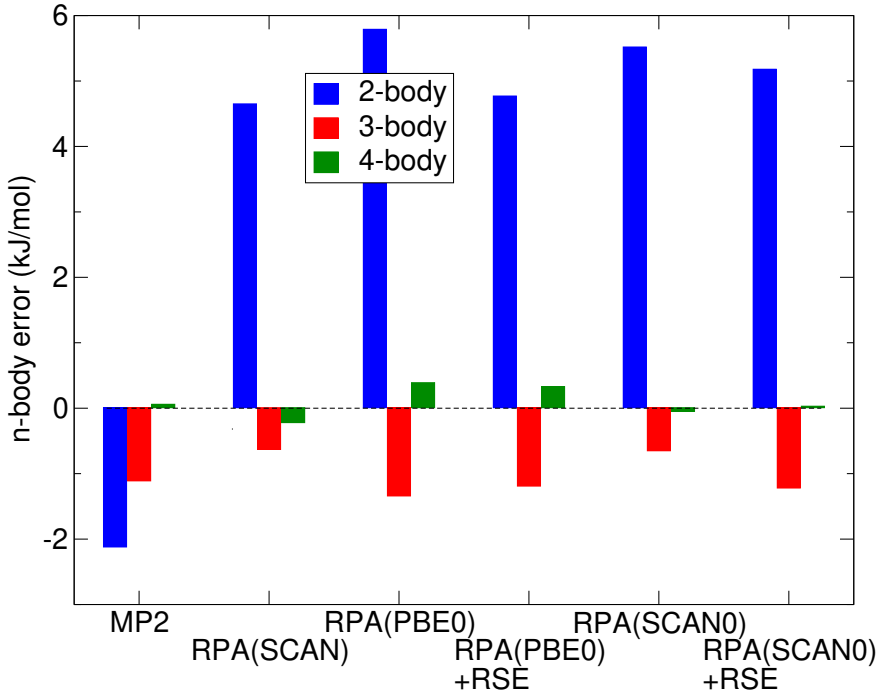

Figure 3.16: Errors of MP2 and different RPA schemes for the many-body energy contributions of crystalline ethylene.

Due to the problems with the description of three- and four-body terms observed for the various DFT functionals, RPA based on HF input states is worth exploring and we are currently following this direction. Would there be any prob-

lem with this approach? Yes, the HF states would likely substantially reduce the accuracy of RPA for two-body energies. The reason is a much larger electronic gap between the occupied and virtual states in HF and thus smaller polarizability. However, the two-body energies need to be recalculated for the proximate dimers anyway. Moreover, one can expect that the dimer interaction energies for distant dimers could be easily corrected by a pairwise formula, similar to dispersion correction. Another possibility would be to use a range-separated hybrid with a long-range HF potential. This could possibly, keep a good description of the two-body terms and reduce the large errors for the three- and four-body energies. A problematic part would be of a technical character, one would have to make sure that the periodic calculations of a range-separated hybrid precisely reproduce finite cluster energies needed in the second part of the correction scheme. This is a potential issue if the PAW approach would be used for the former and Gaussian basis calculations for the latter. Nevertheless, it is another route that is worth exploring based on the data we gathered.

There is another technical issue connected with using hybrid functionals or HF as an input for RPA within periodic boundary conditions and plane-wave basis sets. It is a larger computational cost of HF diagonalisation compared to diagonalisation based on PBE or SCAN. Of course, this issue is less pressing with increasing power of supercomputers but it is a possible problem nevertheless.

The three- and four-body errors of MP2 are quite small, primarily due to the use of HF states. Therefore, one could think to use MP2 for the correction scheme instead of RPA and we saw that for clathrate this approach worked rather well. Moreover, the leading error of MP2 can be attributed to the lack of three-body correlation which can be approximately added via the ATM correction [130]. The problem with using MP2 is again more technical and caused by a larger computational cost compared to RPA, as will be discussed in Chapter 5. Specifically, MP2 has an $O(N_k^3)$ scaling with the number of k-points which is much less favourable compared to the $O(N_k)$ scaling of RPA. While this might not seem like a lot, one needs to realise that the ratio of going from $2\times2\times2$ k-point grid to, say, $3\times3\times3$ grid is not $(3/2)^3 = 3.375$ for MP2 but actually $(3^3/2^3)^3 \approx 38$.

The other unpleasant feature that we observed for RPA based on PBE or SCAN is the slow convergence of the three-body energies with the basis-set size. In another ongoing work that was spurred by the results obtained in P5 we are testing how the basis set errors of change when hybrid or HF states are used to evaluate RPA energy. Moreover, we are also testing how including additional perturbation terms, such as the second-order screened exchange affects the convergence with the basis-set size.

Overall, RPA(PBE)+RSE or $+GW$SE is a suitable and comparatively computationally affordable scheme to reach high accuracy. For reference energies within the correction scheme, smaller many-body errors are needed and the PBE input fails for this task while functionals employing, at least partly, HF states are more promising.

4. Understanding experimental results

4.1 Introduction

In this chapter I'll discuss my contributions to studies that were done together with experimental groups and that were published in P6 and P7. The interaction with experimentalists is very useful for a modeller as real systems can be very far from idealised structures that we usually use to test and develop theoretical methods. Effects that can be neglected in calculations of test systems might be indispensable to account for to understand the experimental data. Finally, the experimentally measured properties can be different from those on which we focus when developing methods. Therefore, understanding the experimental observations can be a detective-like work where we try to find ways how to extract relevant data from simulations and, at the same time, often need to use novel approaches to perform the simulations.

The aforementioned points were very relevant for both P6 and P7. First, the structures were essentially unknown, second, the main effects were believed to be caused by water but there was little to no proof of the presence of water, third, the measured effects were strongly affected by quantities that we often neglect. The first work, P6, was done in collaboration with experimental groups at MFF UK and J. Heyrovský Insitute of Physical Chemistry and studied temperature-dependent behaviour of water between graphene and silica. The second work, P7, was done in collaboration with an experimental group in Taiwan and considered water between graphene layers.

In both studies we analysed the properties of water layer confined between two materials. As the thickness of the water layer is on the nanometre scale, such configurations are also called nano-confined water [138]. Nano-confined water attracts a lot of interest from researchers from various fields as water can be considered nano-confined in cells of organisms, in cracks between rocks, and many other systems. There have been many heated debates about the possible structure and properties of water in general [139] and nano-confined in particular, especially about its thermodynamic properties [140]. One relevant property of water that was important to consider in P6 and P7 was surface premelting [141, 142]. The molecules on a surface of ice become liquid below the melting temperature of the bulk. The shift of the melting temperature is several tens of degrees C for the first surface layer although different experimental techniques can give different values due to the small layer thickness [143]. When the water is confined, the premelting is further affected by bonding to the confining material and by its movements. An interesting question related to nano-confined water is from which cavity size does the confined water start to freeze in structure similar to bulk ice and not in amorphous structure. This question was answered for water droplet in vacuum [144, 145], but the answer most likely depends on the properties of the confining material and the shape of the cavity.

4.2 Graphene wrinkles

Graphene and 2D materials in general have been widely studied in the past two decades using many different techniques [146]. The experimental set-ups often involve supported 2D material, that is, the system is placed on a surface of another material. In P6 the support was silica (SiO_2) and the main technique used to study graphene was Raman scattering at different temperatures. For graphene, Raman spectra give information about charge doping and strain in the material [147–150]. From atomic force microscopy it was known that the graphene on surface is not flat but corrugated, with so-called wrinkles [151]. The Raman spectra was then a combination of contributions from the flat parts and from the wrinkles, but were disentangled to obtain temperature dependent strain and doping [152]. The strain was then used to calculate lattice constant of the graphene. The problem appeared when the temperature dependence of the graphene lattice constant was compared to data available for silica and isolated graphene: the observed lattice expansion was about one or two magnitudes larger than that of graphene or silica. Therefore, the increase in strain could not be explained by expansion of the substrate and there had to be another effect in play. A possible explanation was that there was water between the substrate and the graphene layer. Its increased mobility with increasing temperature and possibly other effects would then drive the expansion of the graphene and affect the doping. My role was then to analyse the role of water in the system and contribute thus to the explanation of the observed effects.

It is known that orientation of water molecules can affect the doping of a graphene layer, I contributed to one such study during my stay at UCL [153]. These effects can be modelled using a single static layer of water molecules and a rather small simulation cell with sides not necessarily exceeding 1 nm. Therefore, density functional theory (DFT) can be used and one can assess the changes in doping from the density of states. In contrast, the height of the wrinkles is several nm, their width can be even larger and molecular dynamics is needed to capture temperature effects. Therefore, much larger system sizes and longer time scales need to be used than for studies like in Ref. [153]. We checked the use of simpler quantum mechanical methods such as tight binding but they had still too large computational cost. Therefore, we opted for the use of classical forcefields to model the system. The clear downside of this option is that one can't access the doping directly, the benefit is that the dynamics of water can be studied on sufficiently long time scales. To get some information about the changes in doping, we analysed how temperature affects the orientation of the water molecules. Moreover, we calculated the average distance of the graphene layer from the surface atoms as it likely affects charge transfer and thus doping.

As stated, the information about the structure of the system was rather scarce. We therefore constructed several models with different number of water molecules between the silica substrate and the graphene layer. The graphene itself was built from a flat part and a wrinkled part with defined width and height, the structure of the substrate was taken from the literature [154]. We then run molecular dynamics for these different structures to gain basic information about how the system behaves. In Fig 4.1 we show result of molecular dynamics at 300 K for the same initial structure of graphene layer and two different numbers of water

molecules. In the (a) structure, water molecules create a double layer in the wrinkle which itself has a shape of a hairpin. Water molecules which were in the flat regions condense so that there are areas where graphene is in direct contact with the substrate. In the (b) structure the number of water molecules is larger leading to a Gaussian or bell-like structure of the wrinkle. Water molecules in the flat region also condense and their number is sufficient to form a double layer. Moreover, for either of the structures we note that the width of water in the wrinkle is larger at the bottom of the wrinkle, this is caused by the presence of OH groups on the surface which bind water more strongly than graphene.

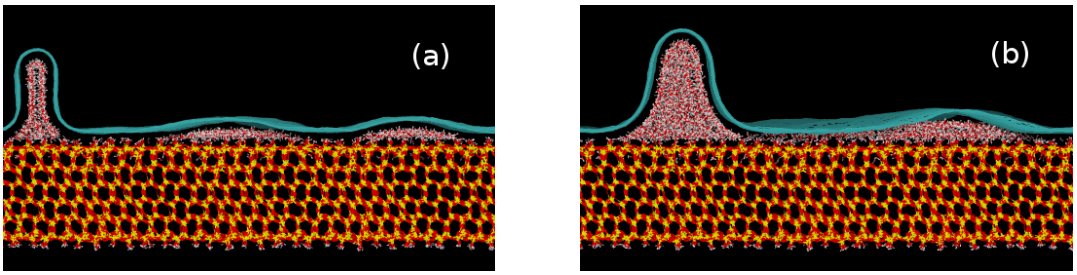

Figure 4.1: Some of the models tested for the graphene wrinkle-water-silica system. System (a) contains 1950 water molecules and shows a "hairpin" structure of wrinkle with a water double layer forming between the hydrophobic graphene surfaces. System (b) contains 3800 water molecules and apart from water-filled wrinkle shows that a water double-layer forms on the silica surface.

Apart from silica substrate we also briefly tested how water behaves when graphene layer is used as a substrate. One of the reasons was that the model of silica surface needed to be rather large to accommodate the wrinkle and contained around 50 thousand atoms (without graphene and water). The graphene substrate then allows one to use a smaller simulation cell and also affects the interfacial water due to lack of atomic charges in the forcefield. The structure turned out to be quite interesting so we show it here in Fig. 4.2. When the number of water molecules is such that a double layer can form in the flat area, a double layer 2D water ice is formed. The structure is not formed by perfect hexagons, we observe rings with smaller as well as larger number of water molecules. This is again consistent with various similar structures that were observed for confined 2D water [155].

Our final set-up to obtain the data at different temperatures was the following: A wrinkle with width and height of 4 nm was build on the silica substrate. The space between the substrate and graphene was filled with water using a structure of water ice Ih. Subsequently we run geometry optimization and several short molecular dynamics at temperatures of few tens of K to obtain structure without short interatomic distances. Long molecular dynamics at the desired temperatures were then run to equilibrate the structures. Several properties, such as average volume per water molecule, were analysed to assess if the structure is equilibrated or not. However, reaching equilibrium is difficult for temperatures below 200 K due to a very slow dynamics of the system.

The final structures at temperatures of 228, 256, and 266 K are shown in Fig. 4.3. Note that we used the TIP4P/Ice water model which has a melting

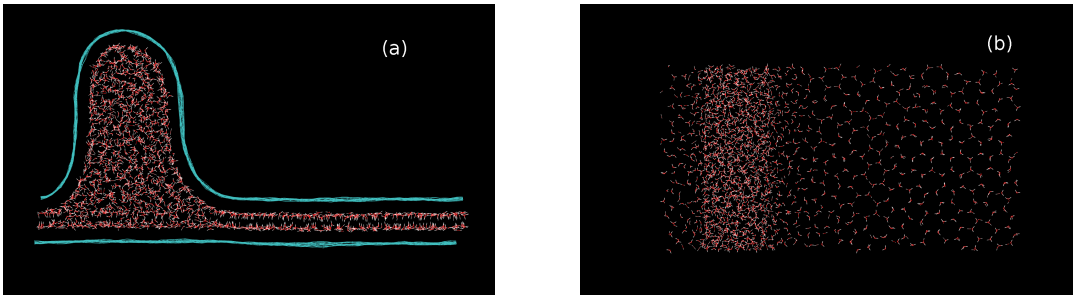

Figure 4.2: Side (a) and top (b) view of water between a flat and wrinkled graphene surfaces. In contact with graphene, formation of layered structures of water was observed.

temperature close to the experimental one [156] thus all the models are at temperatures below that of melting of bulk water. One can see that the ice core persists at 228 K and only melts at higher temperatures, consistent with the observations for finite ice crystal [142]. Moreover, we can observe a formation of water layers for water in contact with graphene, again in agreement with previous works [157, 158].

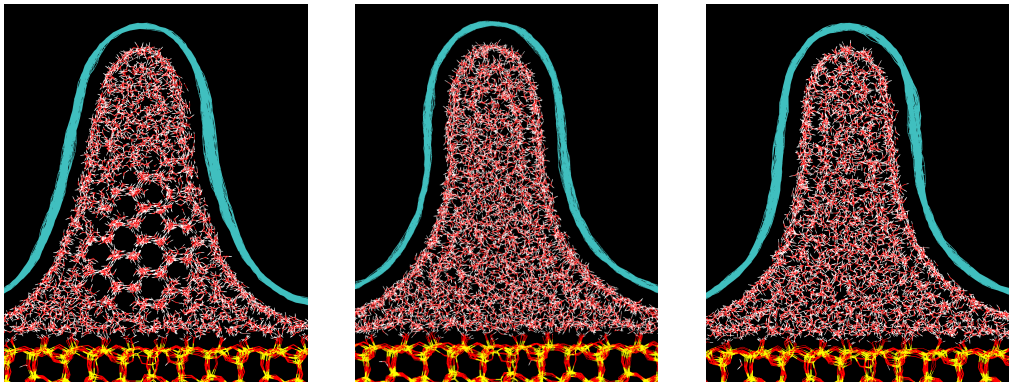

Figure 4.3: Snapshots of water in graphene wrinkle from molecular dynamics simulations. The systems were run at 228 K (left), 256 K (middle), and 266 K (right).

For each temperature we obtained production runs of lengths of at least 5.25 ns. These were used to obtain properties such as the C-C bond length, water orientation and self-diffusion, and others. In agreement with experiment, the C-C bond length increased with increasing temperature. The magnitude of increase was about twice as large compared to experiment which could be considered satisfactory given the long list of approximations and simplifications of the set-up.

Interestingly, there was one approximation which could have a large effect on the results – the neglect of the quantum nature of the nuclei. It was shown that when carbon nuclei are treated classically, graphene expands at all temperatures, when the nuclei are treated using quantum mechanics, graphene actually shrinks at low temperatures [159]. We therefore decided to use the so-called path-integral molecular dynamics (PIMD) to model the dynamics of the carbon

nuclei at the quantum level. This scheme required us to change the code from gromacs [160] to LAMMPS [161] to use the i-Py driver [162] which reduced the accessible timescales. However, we also used the PIGLET thermostats [163] which reduced the number of necessary PIMD beads down to only 8 for temperatures of 60 and 125 K and less for higher temperatures.

The PIMD simulations were started from the last structure of classical MD for each simulated temperature. As shown in Fig. 4.4 the change of the lattice constant was reduced from the classical MD simulations and the increase of strain resembled the experimental one. Moreover, the simulated data show a clear kink in the strain dependence on temperature which appears at temperatures close to the melting of the ice core. Therefore, it is consistent with the assumption that the increased dynamics of melted water drives the increased strain in the graphene wrinkle.

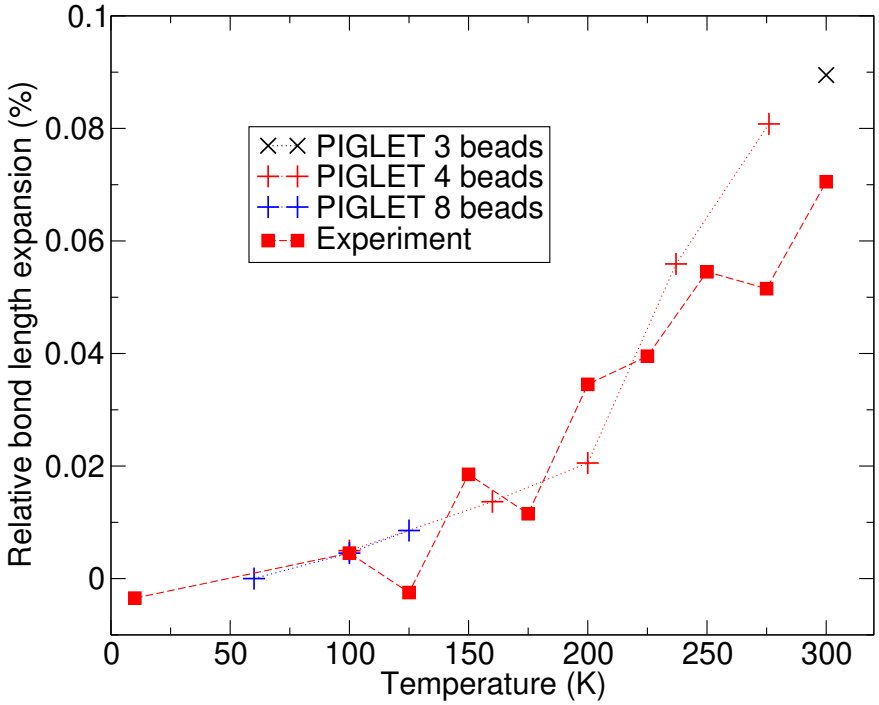

Figure 4.4: The change of lattice constant from the equilibrium value as a function of temperature. The simulated data were obtained for a different number of PIMD beads and the PIGLET coupling.

One of the properties that we calculated was self-diffusion coefficient of water. We compared the results for water in wrinkle and in the flat region. Moreover, we also used data from molecular dynamics runs which contained more water molecules as they started from a structure with two ice layers between silica and graphene and not only one as those used for most of the calculations. The self-diffusion coefficient for the different areas and number of water molecules are shown in Fig. 4.5 as a function of temperature. For all the data sets one can observe a kink in the temperature dependence appearing at around 200 K and caused by the premelting of confined water. The differences between the two data sets for water in wrinkle are small. The water in flat region shows a smaller diffusion for the thicker two layer structure. This is likely due to formation of

larger water "puddles" of condensed water in which water molecules are more strongly bound and thus they are less mobile. The mobility of water is larger in the wrinkle than in the flat parts temperatures above $\approx$ 170 K. For lower temperatures the diffusion is the largest in the flat part of the single layer model. This is due to the fact that the water molecules in this structure aren't typically bound to other water molecules but only to the surface atoms and thus their binding energies are weaker compared to the molecules in wrinkle or in the thicker two layer ice. This likely transfers to a smaller barriers for diffusion and thus slightly faster dynamics of the water molecules in the single layer system.

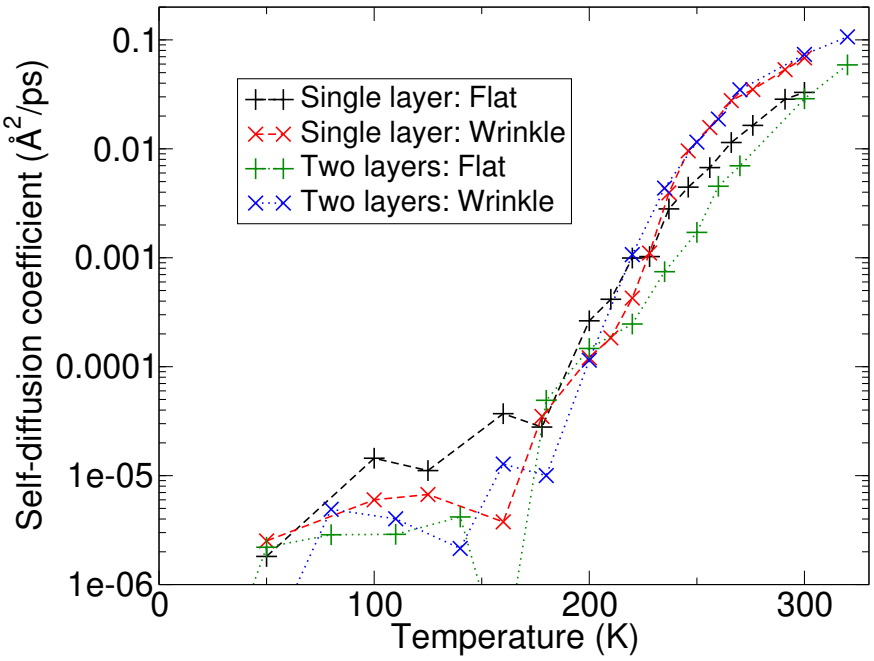

Figure 4.5: Self-diffusion coefficient of water in the wrinkle and flat areas of the system as a function of temperature. We show data for two systems differing in the water coverage in the flat area.

Overall, our simulations were consistent with the experimental data and show a clear influence of water on the properties of graphene layer on silica. Moreover, it was important to consider quantum nature of nuclei which lead to a quantitative agreement with experiment. We note that with around 93 thousands of atoms, our simulations are likely one of the largest PIMD simulations performed so far.

One of the issues that still deserves attention is the structure of water in wrinkle. We assumed that the structure is the same as in ice I_h as the wrinkle size should be large enough to make it stable. It'd certainly be interesting to go from the other side – to freeze liquid water and to analyse how the frozen structure depends on the size of the wrinkle. This would complement studies that are performed for carbon nanotubes [164–166].

4.3 Ferroelectric 2D ice

The system studied by the group from Taiwan consisted of a 40 nm thick layer of aluminium oxide with holes of around 1 μm and two sheets of graphene on

both sides of the oxide. The graphene layers could be brought in close contact by applying voltage with water layer thought to be forming in between them, as sketched in Fig. 4.6. However, the layers stayed in contact even after removing the voltage and different measurements indicated a presence of a ferroelectric ordering of the water layer. Importantly, the effect was reduced with increasing humidity or increasing temperature. This supported the hypothesis that a ferroelectric ice forms between the graphene sheets because for larger temperatures ice would be less stable. Moreover, a thicker water layer wouldn't keep the ferroelectric structure after switching off the voltage. However, as in the previous study there was not enough evidence to support the hypothesis only from experimental data. The simulations that I performed confirmed that the formation of ferroelectric 2D ice is plausible in the system.

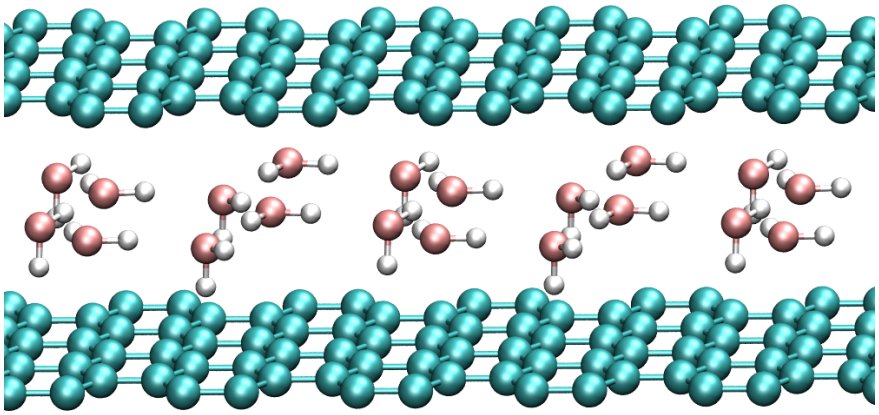

Figure 4.6: One of the models used for 2D ferroelectric water ice confined within graphene layers. For clarity, only some of the atoms are shown.

To obtain a set of plausible structures of water between graphene sheets we tested several strategies. The first is to use different graphene cells and add water molecules randomly between the sheets. For this random structure search [167] we can force different periodicities on the water molecules. For example, each molecule that is put in the unit cell can have a defined number of replicas in x and y direction within the unit cell. This simplifies creation of incommensurate structures. Moreover, we can change the distance between the graphene sheets to increase or decrease the space available for water molecules. When adding a new water molecule, we calculate its distance to molecules already present, and if the distance is too small the molecule is not added and we generate a new random position. We tested several sizes of unit cells and used different periodicities of water molecules and their total number, using up to 16 water molecules in the unit cell. The distances between graphene layers were set to 5 or 6 Å to obtain monolayer structures and to 8 Å to obtain double layer structures.

For each of the settings (cell size, number of water molecules and their periodicity) we generated between 100 and 1000 random structures. The structures were then geometry optimized by LAMMPS [161] using different surface charges. This helps to identify structures stable for zero voltage and structures that could exist for non-zero voltage on the graphene sheets. As with P6 we used the AIREBO

potential for carbon to describe graphene [168]. The low energy structures were then visually inspected and those that were unique were reoptimized with VASP.

The random structure search procedure found some interesting structures. For example, the water molecules often form a hydrogen bonded network and due to the confinement this typically leads to water layers with in-plane dipole, see Fig. 4.7(a). Upon reoptimization with VASP this structure turned into one identified before by Zhao and co-workers and called high density flat rhombic monolayer ice (HD-fRMI) [169] shown in Fig. 4.7(b). For charged surfaces stable structures often contained water molecules oriented upright, that is with water molecule dipole perpendicular to the graphene layers. An example of such structure is shown in Fig. 4.7(c), this structure started from the same initial random geometry as that in Fig. 4.7(a). Interestingly, upon reoptimisation with VASP (in zero external field) we obtained a single layer ice XI, see Fig. 4.7(d). This was the lowest energy structure found, its adsorption energy is −647 meV, for comparison, the adsorption energy of the HD-fRMI structure is −618 meV.

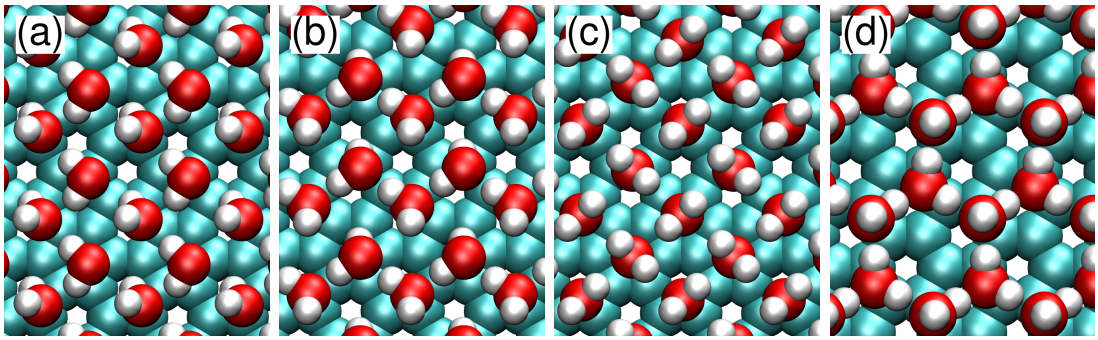

Figure 4.7: Structures of water layer between graphene sheets obtained from random structure search. Structures (a) and (c) started from the same initial random geometry and only used different surface charges on graphene atoms in forcefield optimization (zero and ±0.1 |e|, respectively). Structures (b) and (d) were obtained after reoptimizations of (a) and (c) structures using VASP. Only the bottom graphene layer is shown for clarity.

Setting the distance between graphene layers to 8 Å typically resulted in bilayer structures. In general, for bilayer all water molecules can accept and donate two hydrogen bonds. This is more energetically favourable compared to directing OH bonds to graphene layer, which occurs for the monolayers. The possibility to form more hydrogen bonds in the double layer has two consequences: the low energy bilayer structures tend to be more stable than the low energy monolayer configurations and they also tend to have zero net dipole along the surface normal. The lowest energy structure identified was a bilayer ice with 4-, 6-, and 8-membered rings ($E_{\text{ads}} = -688$ meV), see Fig. 4.8(a) and (b). Another low energy structure contained staggered chains of water molecules ($E_{\text{ads}} = -637$ meV), see Fig. 4.8(c) and (d).

To expand our set of tested structures we also used confined ice geometries proposed by Chen *et al.* [155]. Some of them turned out to be lower in energy compared to those found by the random structure search. A possible reason is that we typically allowed only two water molecules in the cell to be positioned

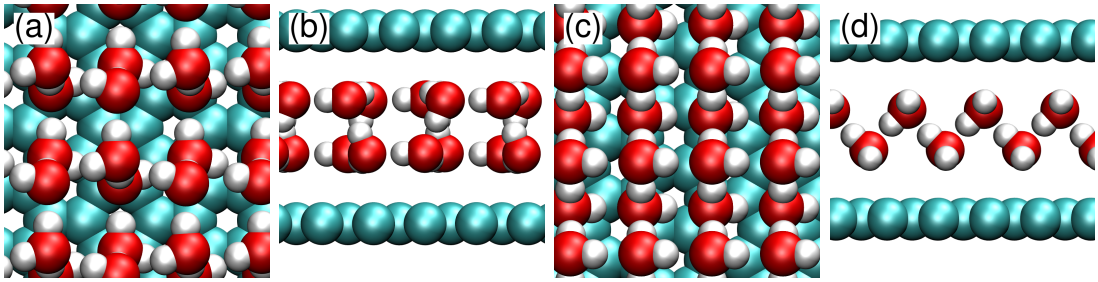

Figure 4.8: Random structures obtained for water between graphene layers. Initial conditions set a space of 8 Å between the layers and 12 water molecules in total. Panels (a) and (b) show a top and side views of a double-layer "ring" structure while a less stable "staggered" structure is shown in panels (c) and (d).

explicitly randomly, the positions of the other molecules were set automatically by assuming periodicity of the structure. While we tried to generate structures with a larger number of randomly placed molecules, we were only able to use a limited number of such structures and they were not as stable as the those obtained from the less flexible settings. Apart from the structures of Chen and co-workers we also used the ice XI monolayer structure.

To compare the ferroelectric structures to structures with zero net dipole we took the structures hex and cairo from Ref. [155] and one monolayer of ice XI. The first two have zero net dipole while ice XI is ferroelectric. We then made ferroelectric versions of the hex and cairo structures and a ice XI with no dipole by manually turning the water molecules pointing to one graphene layer to point to the other layer. The energies of these structures, after geometry optimisation and with the PBE-D3^{BJ} functional, are shown in Table 4.1. The ferroelectric structures are less stable in all the cases, the differences are, however, very small, only 3 meV for the ice XI and 5 meV for the hex phase. This means that there is only a very small energy penalty for stability of the ferroelectric phase and it is possible that it can stay stable once formed, as hinted from the experimental data.

Table 4.1: Binding energies of water molecules in the graphene confined structures. The hex and Cairo structures are from Ref. [155] and have no net dipole along surface normal, thus they are noted as anti-ferroelectric (anti-FE). A ferroelectric phase (FE) was constructed from them by rotating water molecules so that all free hydrogen bonds point to the same graphene. Ice XI is FE and thus an opposite procedure was performed to obtain the anti-FE phase.

Structure	Type	E_{ads} (meV)
Hex	anti-FE	−663
Hex	FE	−658
Cairo	anti-FE	−663
Cairo	FE	−652
Ice XI monolayer	FE	−648
Ice XI monolayer	anti-FE	−651

It is likely that the ferroelectric phases become more stable when the graphene layers are attached to a voltage. In VASP we do not possess the methods to change the potential on the graphene planes and we tried to simulate this process by setting an external constant electrostatic field in the simulation cell. This should induce a charge transfer between the graphene sheets and qualitatively lead to the same effect as setting the charges directly, as done in the forcefield calculations. The total energies of the hex and hex_FE phases in different fields are compared in Fig. 4.9 where one can see that the ferroelectric phase becomes indeed more stable when the field is large enough. One can also note that the curve is symmetric for the hex phase due to the symmetry of the geometry.

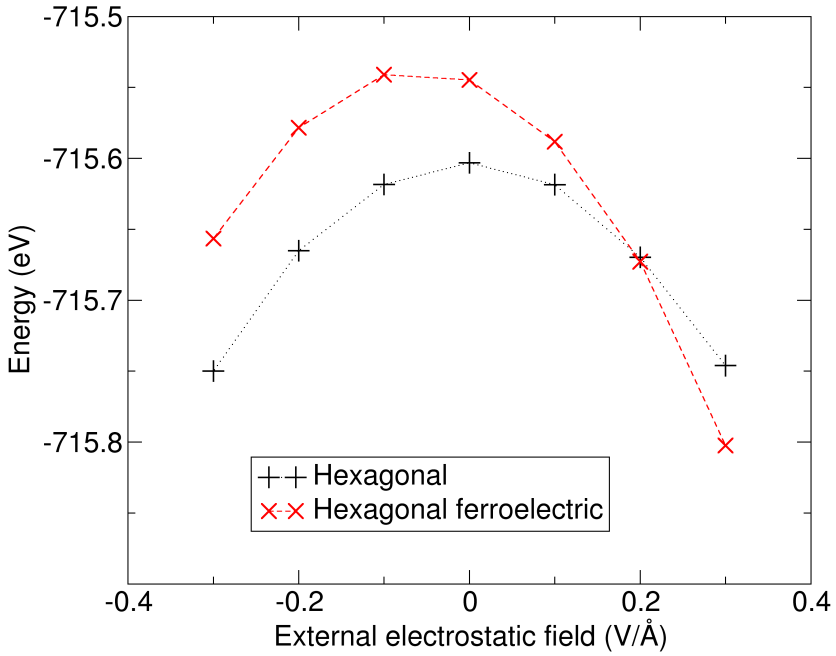

Figure 4.9: Energy of single hexagonal ice layer with zero net dipole and its ferroelectric modification confined between two graphene layers for different values of external electrostatic field.

We now mention some of the results obtained using forcefield molecular dynamics that we used to assess how water behaves for different coverages and temperatures. We set up a system with two graphene layers and different numbers of water molecules between them. The lateral cell size was about 100×100 Å^2 and the number of water molecules between around 400 and 2000, which corresponds to coverages between ≈ 0.4 and 2.1 monolayers. In total eleven different numbers of water molecules were used. For these structures we run simulations at temperatures 260, 300, 320, and 340 K and with surface charges 0.0, 0.001, 0.002, 0.005, 0.01, 0.02, 0.03, and 0.06 $\pm|e|$. Charges were positive on one layer and negative on the other. The initial structure was a layer taken from molecular dynamics of water bulk with some of the molecules removed to obtain the chosen number of waters.

We analyse the simulations using distance between graphene layers which gives an information about thickness of water layer. Examples of the results are shown in Fig. 4.10 in which the distances are averaged over cells with a side of

approx. 5 Å. Note that there is no water between the sheets for values below ≈5 Å, a monolayer is present for distances around 6 Å. Each additional water layer then adds around 2.5 Å to the interlayer distance.

First of all, and not unexpectedly, water molecules tend to condense together, as shown in the upper panels that result from simulation with zero surface charges. At low temperatures (top left panel) the monolayer is preserved while at higher temperatures a water multilayer or a puddle is formed. However, one should be cautious, the condensation would be quite sensitive to relative strength of the adsorption energy of water molecule and the water-water interactions [170]. The situation is markedly different when there is an electric field between the graphene layers (bottom panels). For charges 0.01 $|e|$ and larger a monolayer is formed with some of the water molecules oriented in a way that one OH bond points in the direction of the graphene surface normal. Therefore, the interaction with the electrostatic field is sufficient to prevent the water molecules from clustering.

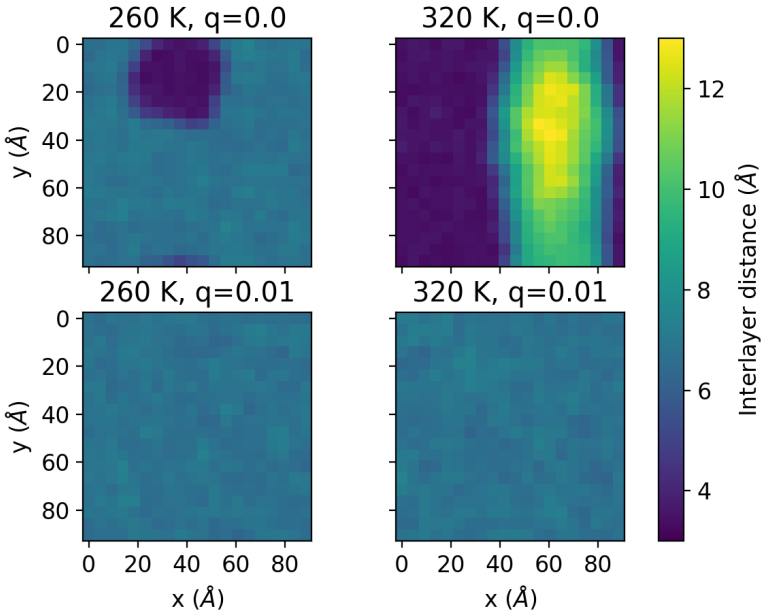

Figure 4.10: Average interlayer distance between two graphene sheets that contain 929 confined water molecules. Data are shown for two temperatures (260 and 320 K) and the graphene atoms have either zero charge or +0.01 |e| on one and −0.01 |e| on the other graphene sheet.

Overall, we see that ferroelectric structures have similar stability to structures with zero net dipole along the graphene normal. Moreover, they become more stable when electric field is added, which we used either using an external electric field in VASP or by setting partial charges on the carbon atoms. Both the DFT and forcefield simulations then support the experimental observations that point to a formation of a ferroelectric water ice layer between the graphene sheets. Finally, we point out that our simulations were performed with flexible graphene sheets which then allow for a formation of water multilayers even when the coverage is equal to one (i.e., monolayer). This is an effect sometimes neglected in other studies.

4.4 Conclusions

The two discussed examples show how simulation can help to understand experimental observations or support some of the hypotheses that explain the observations. The works also showed that there are systems where our theoretical methods need to be further developed to be able to assess some of the properties directly or more easily.

One of the clear issues was the inability to obtain long trajectories of molecular dynamics while taking into account polarisation or dynamic charge transfer, e.g., within graphene or between substrate and graphene. A possible solution would be to use computationally cheap electronic structure methods, such as DFT functionals with small basis sets [171–173] or simpler methods [174–176]. Another possibility would be to use more flexible machine learned potentials that would incorporate the prediction of partial charges on graphene. The importance of these methods is likely to rise as there are many complex materials for which we will need to understand the dynamics as well as the electronic structure.

5. Precision

5.1 Introduction

One of the interesting things about computer simulations of molecules and materials is that our results are almost always wrong compared to reality. Wrong in the sense that the values, primarily energies, that we obtain are not exact because of the range of approximations that appear when setting-up the model, in the Hamiltonian, wavefunction, numerical set-up of the system, implementation, data processing and so on, as illustrated in Fig. 5.1. Of course the fact that our results are not exact is not necessarily bad by itself and it is something that we are usually very well aware of. Moreover, understanding to what extent different approximations affect the results is often a driving force for our research. In Chapter 3 I focused on errors observed for different theoretical methods and in this chapter I focus on understanding some of the errors introduced by numerical set-up.

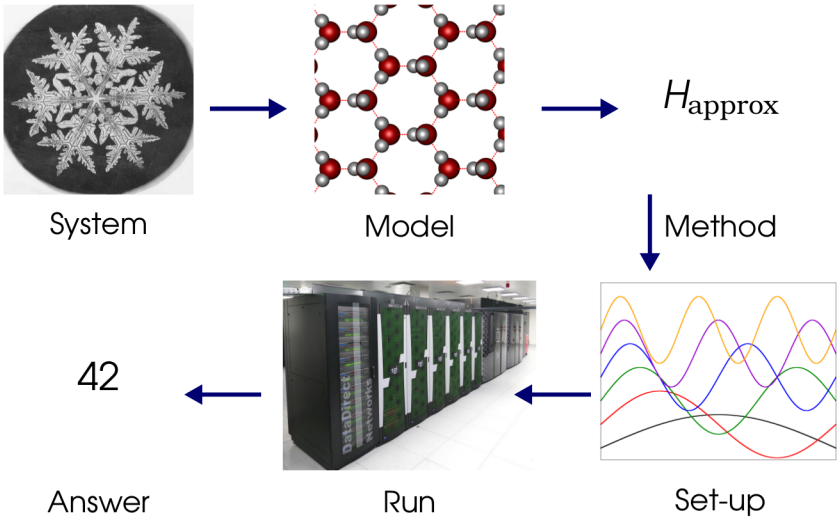

Figure 5.1: There are many steps we have to make when calculating properties of some system before we get the answer. Each of the steps involves some approximations or requires us to choose some parameters. These choices will affect the result.

A large deal of our work concerns understanding the accuracy of different theoretical methods, that is how large deviations of the predicted properties (such as binding energies) are caused by the approximations made to the Hamiltonian or the wavefunction. An example is the accuracy of the random phase approximation which I discussed in Chapter 3. To assess the accuracy, we take reference data for some property and calculate the same property using the method that we are testing. We then compare the results of the tested method to the reference to see how good or bad the tested method is. That is, we seek to find out how *accurate* the method is. However, the differences will usually not come only from the approximations made for the Hamiltonian or wavefunction. There are other

settings and approximations that can lead to deviations as well, such as the basis set size or treatment of core electrons. Moreover, for calculations within periodic boundary conditions (PBC) we have additional parameters such as the density of k-points for solids or the size of simulation cell for isolated-like molecules. The deviations caused by these settings from exact energy for given method are usually described by the term *precision*. It is important to note that to assess the accuracy of some method we need data that are highly precise, both for the reference and the tested data sets, otherwise the comparison will not reflect the true performance of the method.

There are more reasons why we should try to produce precise data. First of all, if reference data are not precise, researchers using them to test their methods will get imprecise and maybe misleading results. Fortunately, for methods such as coupled clusters we are quite aware of the issues with basis-set convergence or neglect of core correlation and higher-order correlation contributions [177, 178]. However, one can come across results where the tested method was run with a less converged or imprecise set-up. This can lead to a misleading assessment of the method's accuracy. This is not ideal as users often quote results of such benchmark studies and they can misjudge the quality of their results. Finally, if imprecise set-up is used in a publication another researcher can have hard time reproducing the results, say with a different code or even with the same code if crucial details of the set-up were not published.

The situation with quality of set-up and thus precision of the results is especially worrying for hybrid density functional theory (DFT) or post-Hartree-Fock (post-HF) methods within PBC. The reason is that their computational cost is high so it is tempting to quickly produce results with reasonable parameters than to perform computationally expensive convergence tests. However, in contrast to semi-local DFT functionals the energy has typically a stronger dependence on some of the parameters. Thus using reasonable parameters, say similar to those used for semi-local functionals, can lead to substantial reduction of the precision. To identify various issues related to the use of hybrid functionals and post-HF methods we often use a strategy in which we calculate the same property by two different codes and analyse the differences. An example of this approach was published in P8 is discussed below. It involves calculation of second-order Møller-Plesset (MP2) binding energies of molecular solids with two methods: many-body expansion (MBE) and directly within PBC.

Since the start of my PhD I've been using the VASP code which uses so-called projector-augmented wave (PAW) method to avoid explicit optimization of core electrons [179, 180]. The PAW is one of the approximations that affect the precision of our calculations. For studies such as diffusion of water on NaCl surface [181] or salt dissolution [182] we used so-called "standard" PAWs which lead to imprecisions of around 2 % for adsorption energies. Such loss of precision was deemed acceptable as there was a large number of structures to optimize and using more precise "hard" PAWs would substantially increase the computational cost. However, in another study [20], where we studied the accuracy of the van der Waals density functional (vdW-DF) [17] the use of the more precise hard PAWs was absolutely necessary. This was because the imprecision of the standard PAWs was close, in terms of relative error, to the inaccuracy of the vdW-DF scheme and thus using less precise standard PAWs would not lead to a fair comparison,

as discussed above. In recent years, we have analysed the errors of different PAW data sets distributed with VASP in order to understand their origin. The first results are presented in P9 and discussed in Subsection 5.3 of this Chapter.

5.2 Comparison of many-body expansion and calculations within periodic boundary conditions

There are many publications discussing the calculations of binding energies of molecular solids. As mentioned before, one wouldn't expect some precision issues of the results if the computational method is some density functional theory (DFT) approximation, such as dispersion corrected semi-local DFT functional [33, 34]. This is because for such methods the energy converges quickly with parameters such as the basis-set size or the number k-points. However, correlated or post-HF methods have a slower convergence with the basis-set size and with the density of the k-point grid [22, 23, 183, 184] and the numerical set-up deserves more attention.

For post-HF methods, the binding energy is often obtained using many-body expansion (MBE) and not within PBC [185–187]. The reasons are two-fold, first, there are not many implementations of post-HF methods within PBC, especially for coupled clusters methods, and, second, the computational cost for a single energy evaluation is larger within PBC than when MBE is used. Nevertheless, both the MBE and PBC approaches should yield identical results if converged with all the parameters. In P8 we used both MBE and PBC calculations to obtain MP2 binding energies of four molecular solids and tried to find out how close agreement one can obtain for the results in practice. Moreover, we asked which of the calculations are easier to perform in practice and what are the main barriers for getting a close agreement between the results.

The systems that we considered in P8 were methane, carbon dioxide, ammonia, and methanol, their structures are shown in Fig. 5.2. The reasons for choosing these systems were that the molecules as well as the unit cells are small, which facilitates the calculations using post-HF methods and large basis sets, and they have different importance of electrostatic interactions – small for methane and large for methanol.

In P8 we were able to obtain PBC- and MBE-based binding energies within 0.1 kJ/mol or better for all the systems. This was probably the first time such an agreement for binding energy of molecular solid was obtained for a post-HF method, for HF such high precision was demonstrated before, e.g., for benzene in Ref. [188]. During the five or so years that took between the start of the work and the publication we identified several problems and issues, most of them are discussed in P8. Here, I comment on some of the interesting or important observations.

One of the problems that one encounters when calculating HF and post-HF energies within PBC is their slow convergence with the k-point grid size. This is caused by so-called Coulomb singularity for Γ-point which occurs due to infinite value of some integrals [22, 189]. For HF methods the problem can be reduced by using a real-space cut-off for the Coulomb interaction [190, 191].

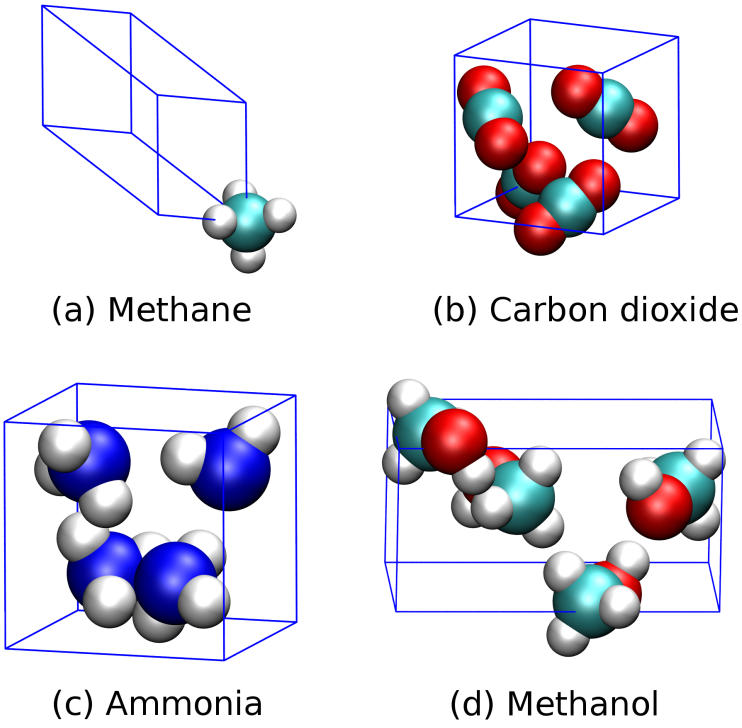

Figure 5.2: Unit cells (including molecules) of systems considered for calculations in P8: methane, carbon dioxide, ammonia, and methanol.

When performing post-HF calculations in VASP, the magnitude of the error is reduced by approximate calculation of the relevant terms [22, 47]. However, the result of the post-HF calculation will also depend on whether the Coulomb cut-off was used in the HF calculation or not.

In Fig. 5.3 we show two sets of MP2 correlation energies of solid methanol as obtained by VASP for different k-point sets. The sets only differ in the way the HF calculation was done, no Coulomb cut-off was used for the first set (black +) while the Coulomb interaction was cut to obtain the second set of data (green $\times$). Note that we did not obtain the data for the $3\times3\times2$ k-point grid for the first set of MP2 energies due to their very high computational cost. Clearly, the convergence is much worse for the first set of data and extrapolation would be required to obtain converged value. For the second set of data, the difference between the values obtained with $2\times2\times1$ and $3\times3\times2$ k-point grids is less than 0.05 kJ/mol (per unit cell) so already the first value could be considered converged. Clearly if the Coulomb cut-off scheme is not used for HF, the states are farther from convergence leading to errors for MP2 correlation energies. Note that one can expect similar issue for random-phase approximation (RPA) energies based on states obtained by global hybrid DFT functionals or HF states. In contrast, the problem doesn't appear for RPA based on pure DFT states as they are not affected by the Coulomb singularity for exchange.

We now discuss some of the observations that we made for the MBE calculations. The standard way to obtain the many-body energies within MBE is to

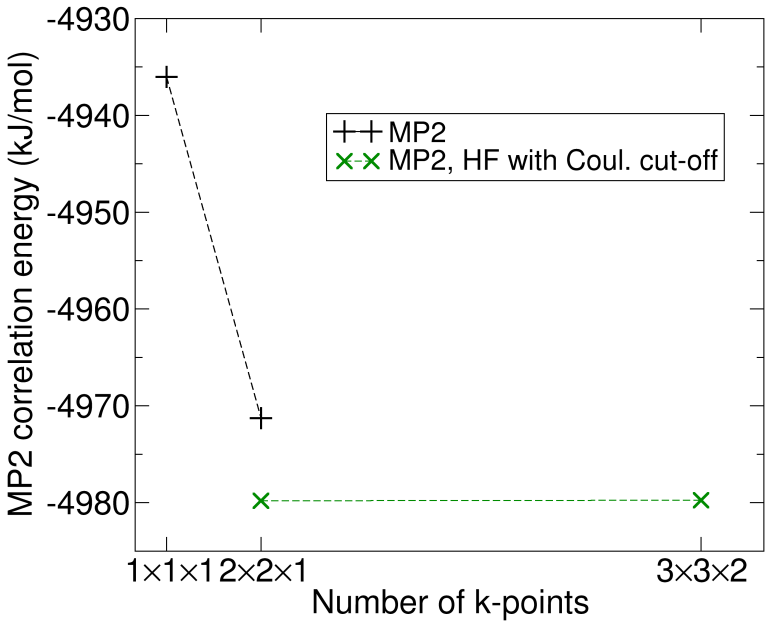

Figure 5.3: MP2 correlation energy of methanol solid as a function of k-point set. The energy was obtained for two sets of HF states, default settings were used for the first and the Coulomb cut-off technique was used for the second.

introduce a cut-off distance and sum all the contributions within the cut-off. To understand if the many-body energies are converged with the cut-off, one usually plots them as a function of the cut-off distance. The red line in Fig. 5.4 is an example of such dependence for the two-body HF energy of methanol crystal. Note that not all contributions are included above a cut-off distance of 30 Å. This is because we sum contributions of all the dimers from given number of shells around the unit cell with the reference molecule.[1] The two-body energy shows large fluctuations with the cut-off distance making it difficult to find a converged value.

The reason for the oscillations in the two-body energy of methanol is that in the crystal the molecules form two chains of hydrogen bonds, some point in one direction and other chains point in the opposite direction. This can be seen in Fig. 5.2(d) and it is schematically illustrated in the left panel of Fig. 5.5 with the reference molecule highlighted with a green colour. In Fig. 5.5 we show the individual molecules as arrows as they have a dipole moment. When the cut-off distance is increased, there will be some dimers with positive contributions and some with negative energies and generally they will not sum to zero. In fact, for some distances one observes more positive contributions and for other more negative contributions due to the structure of the crystal. This then causes the oscillations in the curve showing the two-body energy as a function of the cut-off

[1]To obtain the two-body contributions we consider all dimers between the reference molecule and all molecules contained in a set of unit cells. All cells within some shell around the unit cell are considered. The shells are defined in a way that the unit cell with reference molecule is shell 0, all the 26 unit cells around it form shell 1, the 98 unit cells around shell 1 cells form shell 2 and so on.

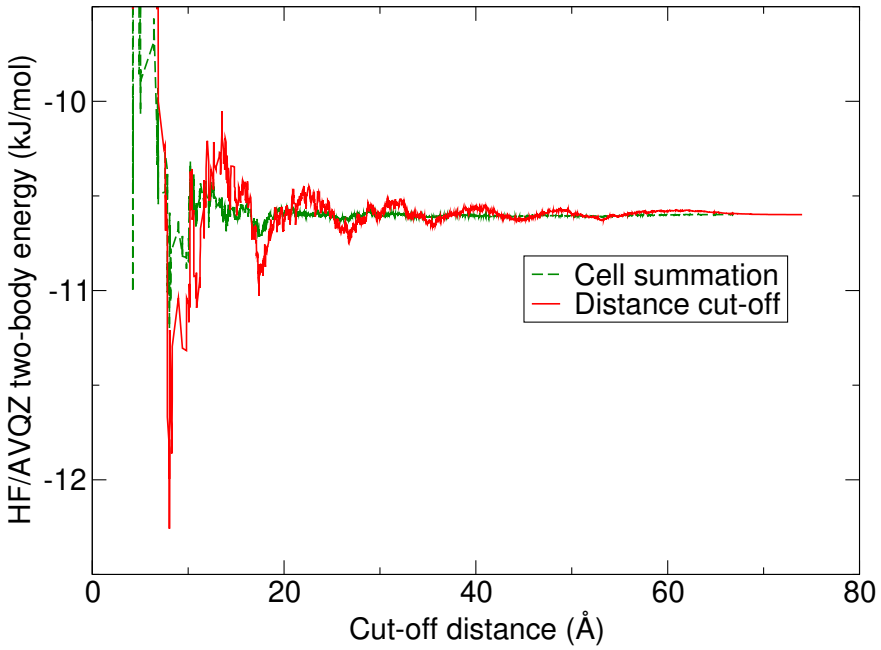

Figure 5.4: HF two body energy of methanol as a function of radial cut-off distance using standard summation and a modified "box" summation in which contributions from all monomers in one unit cell are considered at once.

distance. However, we see that the unit cells don't have a net dipole moment. It is therefore beneficial to bundle all the contributions from a given unit cell together and add them at one cut-off, typically the lowest one. For methanol this means that instead of accumulating dipole-dipole contributions we add together dipole-quadrupole interactions. The convergence of this "cell" summation is shown by the green dashed line in Fig. 5.4. One can see that the oscillations are significantly reduced compared to the standard distance based cut-off summation.

The faster convergence of the box summation makes one think if the summation couldn't be simplified further by adding even more contributions at once. For example, we can add all the contributions from one shell of cells at the same time. The values can be then easily listed in a table which makes it convenient to assess their convergence as well as uncover numerical issues. This is illustrated for the HF and MP2 two-body energies of methanol in Table 5.1. One can see that the dominant contributions originate from the cells in the zeroth and first shell and they are below 0.1 kJ/mol from the third shell onwards. Interestingly, there are no oscillations and the signs are always negative. With such tables we also showed in P8 that smaller basis sets are sufficient when obtaining contributions from the more distant shells. Overall, for molecules with a dipole adding the two-body energies from shells 0 to 2 should yield more precise values than using a distance cut-off and the same number of molecules.

We observed the largest benefits of the cell summation for the three-body interactions. These are more demanding to calculate than the two-body energies and can also show oscillations with the cut-off distance. A comparison of standard summation and the cell summation is shown in Fig. 5.6 for crystal of CO_2. For the standard summations the oscillations are rather small but clearly visible even

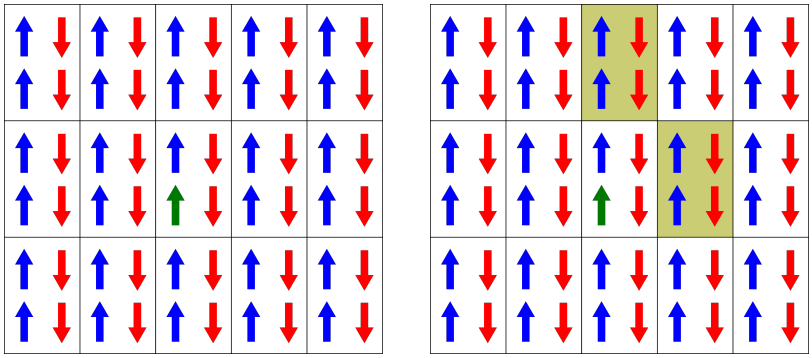

Figure 5.5: A sketch of the structure of methanol crystal. Each unit cell contains two molecules with OH bonds pointing in one direction and two in the opposite direction. The reference molecule used in many-body expansion (green arrow) then necessarily belongs to one of the two groups. Thus for two-body terms the reference molecule interacts with molecules that have dipoles oriented in the same direction and with molecules with dipoles in (approximately) the opposite direction. The different distance at which these contributions occur and their opposite sign leads to oscillations in the distance convergence of the two-body energy. As each unit cell has zero dipole it is advantageous to sum all the contributions from one unit cell at once. This modified "cell" summation is illustrated in the right panel for three-body interactions. All the contributions to the three-body energy involving one molecule from each of the highlighted cells would be added at once, typically when the first contribution from these cells would be encountered.

though isolated CO_2 molecule doesn't have a dipole but only a quadrupole. The oscillations are substantially reduced with the cell summation, both for the HF and MP2 energies.

The data shown in Fig. 5.6 highlight another issue – problems with numerical errors [192]. One can see that the curve for MP2-F12 is clearly going to negative values above a cut-off distance of $\approx$ 30 Å. There are several settings that we identified that can lead to such issues: too loose convergence settings for HF orbitals or the total energy, small fitting basis set used in the F12 correction, or large orbital basis sets. There is only one recommendation that we can make and that is always look at your data. Indeed, the final value obtained from data shown in Fig. 5.6 would look perfectly fine, but the graph uncovers the issues.

The benefits of using cell summation are smaller for the four-body terms. While the standard distance cut-off summation leads to oscillations, the cell summation increases substantially the number of necessary calculations. As for four-body terms we consider the reference molecule and three other unit cells, up-to 64 possible tetramers would need to be added in one contribution. This increases the numerical demands as well as uncertainties due to finite numerical precision and we didn't use the cell summation for tetramers.

In our study we obtained the MBE contributions using rather large distance cut-offs, especially for the dimers and trimers. This uncovered some numerical problems of the calculations and results. However, a large number of calculations should not be necessary in principle as one should be able to fit the results to some simple physical model when the distances between molecules are large and

Table 5.1: Contributions to the two body HF and MP2 correlation energy of methanol from different shells of unit cells. Data in kJ/mol and for the AVTZ basis set.

Shell	No. mol.	HF	MP2
0	3	−6.11	−6.37
1	104	−3.56	−27.72
2	392	−0.61	−0.57
3	872	−0.06	−0.09
4	1544	−0.02	−0.02
5	2408	−0.01	−0.01

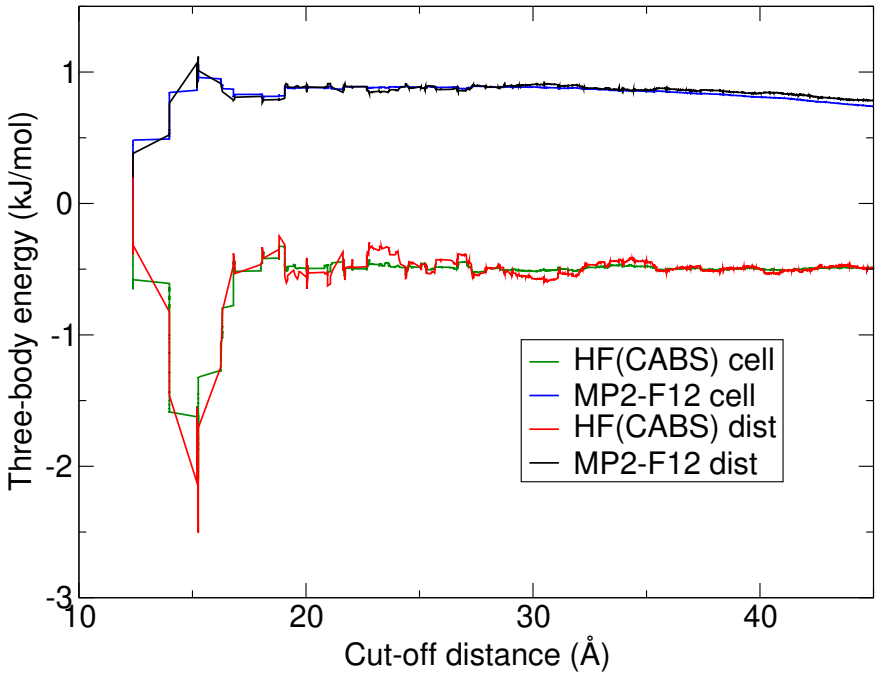

Figure 5.6: Comparison of distance summation and "cell" summation for calculating three-body HF and MP2 energies of carbon dioxide. The "cell" summation includes additional terms and leads to a smoother convergence with the cut-off distance. The CABS and F12 corrections were used for HF and MP2, respectively, to speed up the basis-set convergence.

the molecules are well separated [193, 194]. By large we mean that the molecules are not nearest neighbours or, possibly, second nearest neighbours. For the systems considered here this corresponds to distances above approximately 7 Å. For these separations, the interactions are dominated by the leading terms: dipole-dipole for electrostatics and $-C_6/r^6$ for dispersion, possibly with an anisotropic C_6 coefficient. One can therefore fit a model describing these two interactions to obtain the contributions at large distances. However, dipole-dipole and dispersion interactions with C_6 term are only pairwise and thus will give a zero for three- and higher-body contributions. Using a model for the two-body terms is of only a limited value because the two-body energies typically require less computational

resources than the three- or four-body terms. The minimum requirement for a model to describe three- and four-body contributions of well separated molecules at the HF level is to include their polarisability. For correlation one could use interacting quantum oscillator model, as used by Tkatchenko and co-workers in the context of DFT dispersion corrections [134] or the Axilrod-Teller-Muto term [52, 53, 130].

In P8 we used a simple model based on point charges and point polarisability to fit the data for three-body interactions of methanol as it was difficult to obtain a converged value even with the cell summation. The fit is shown Fig. 5.7 as the red line with the original data shown with a black line. The fit is not very accurate for small cut-offs where the contributions come mostly from trimers with at least two molecules in contact. In contrast, for large cut-offs where the number of trimers with molecules in close contact is small, there is almost a constant off-set between the fit and the original data. To simplify the comparison we also show a line where the fit was shifted to be close to the original data for large distances, this is shown with green line. One can see that the agreement between the original data and the fit is very satisfactory. With the fitted forcefield we could easily estimate the three-body energy for much larger distance cut-off. The estimated HF energies for the larger cut-off differ by only 0.04 kJ/mol from the value obtained for the cut-off used for the explicit HF calculations. This change is much smaller than our estimated uncertainty of the HF energy, equal to 0.3 kJ/mol. Note that we calculate the uncertainty as one half of the difference between the largest and the smallest value of the three-body energy on an interval of last 10 Å below the distance cut-off.

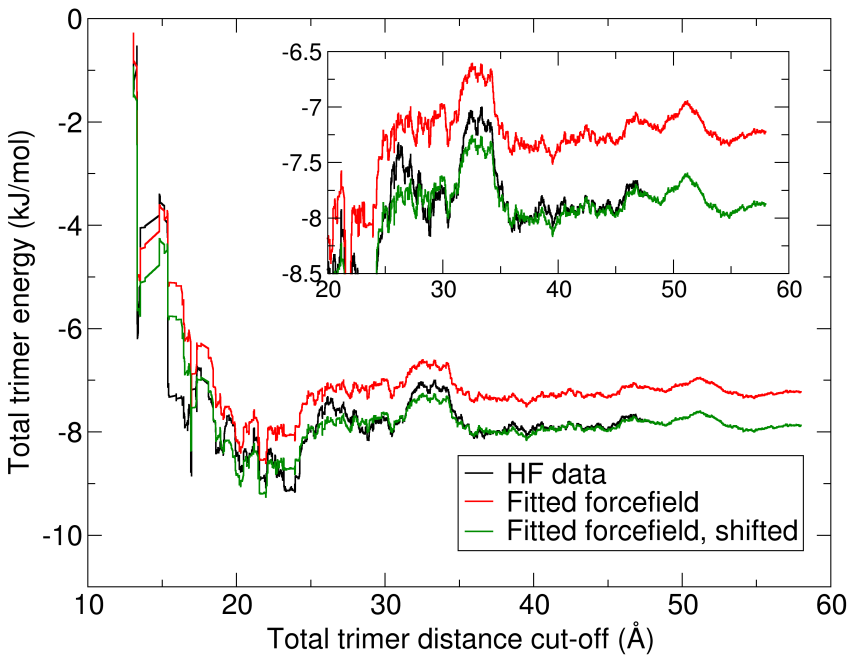

Figure 5.7: Convergence of Hartree-Fock three-body energy of methanol with distance cut-off (black) together with three-body energy obtained from a polarizable forcefield fitted to reproduce the three-body interaction (red). Green line shows the fitted data shifted to lie close to the Hartree-Fock results.

Overall, we found that one can obtain a close agreement between HF and MP2 binding energies obtained within PBC and using MBE with differences around 0.1 kJ/mol. From the users' perspective the biggest gain results from removing or reducing the dependence of energy on different parameters. For example, by using Coulomb real-space cut-off to reduce k-point dependence of HF or MP2 calculations within PBC or by explicit correlation (F12) to reduce basis-set size dependence. However, some of these approaches might involve approximations that affect the precision of the result and one needs to be careful and critically test the obtained values. Moreover, relying on default parameters is also not recommended and one should also test settings for convergence criteria or integral screening thresholds. The default parameters might be perfectly fine for a single calculation but might not be enough for MBE where a large number of small values are added together.

Finally, our recommendation is to use the Coulomb cut-off technique for periodic calculations, extrapolate binding energies first with k-points (if needed) and then with basis-set cut-off. Within MBE, use shell contributions instead of a radial cut-off to obtain two-body energies and for three-body energies use box summation both for the mean-field and correlation contributions. The four-body terms are prone to numerical errors but, as discussed in Chapter 3, can be obtained using small basis sets. The basis-set size can be also reduced for trimer or dimers with separated molecules.

5.3 Precision of Projector-augmented wave potentials

We mostly rely on the code VASP for calculations within periodic boundary conditions. VASP implements the projector-augmented wave (PAW) method to avoid explicit optimization of core electrons and thus to speed-up the calculations. One of benefits of VASP is the availability of well tested PAW data sets (pseudopotentials, partial waves, and other data) for essentially all of the elements of the periodic table. The PAW method and the resulting PAW data sets are also considered to be quite transferable, that is they should achieve similar precision in different chemical environments. It is therefore not surprising that VASP is often used to study molecular crystals or molecular adsorption on solid surfaces [195]. However, when testing the accuracy of, say, different DFT approximations for the prediction of intermolecular binding energies or binding energies of solids we want to reach a precision of around one per cent or better. At the same time intermolecular interactions are several orders of magnitude smaller than total energies or even than energies of covalent bonds and thus are very sensitive to the various computational settings. Therefore, one needs to be careful even when using PAW potentials for intermolecular interactions so as not to introduce significant errors into the results. Together with my post-doc Dr. Yourdkhani we assessed the precision of some of the PAW data supplied with VASP for intermolecular interactions and published them in P9. Here I highlight some of the main findings.

There are several PAW data sets distributed with the VASP code which in essence differ in the precision they offer. For light main group elements (first and

second period) one can use so-called Soft, Standard, and Hard PAWs, the precision of which increases from Soft to Hard.[2] The Standard PAWs are the most widely used but we have often found that in many situations their precision for describing intermolecular binding energies is not sufficient enough. For example, for water dimer Standard PAW leads to an error in binding energy of around 5 meV. This might not seem like a lot, but the binding energy is around −215 meV for the PBE functional, so the relative error is slightly above 2 %. As many dispersion corrected DFT functionals claim to have an average absolute error of 10 to 15 meV the PAW error is, in fact, comparable to the error of the method [20].

To understand the error of the PAW potentials we first used the S22 [105] and S66 [106] data sets and calculated the difference (error) between binding energies predicted for a given PAW potential and an all-electron reference obtained using a large aug-cc-pV5Z basis set. The S22 and S66 data sets contain dimers bound mostly by hydrogen bonds or electrostatics, dispersion, or a mixture of different interactions. The results can thus show the bonding situations for which the PAW potentials lead to the largest and the smallest errors. From the results for the S22 set, shown in Fig. 5.8, we can observe that the errors are marginal for systems with dominant dispersion interactions (dimers 8 to 11), this holds even for the least precise Soft PAW. In contrast, the errors are noticeable even for the Hard PAW for hydrogen bonded systems. This and results obtained for S66 is consistent with previous observations which also showed the largest errors for hydrogen bonded systems [196].

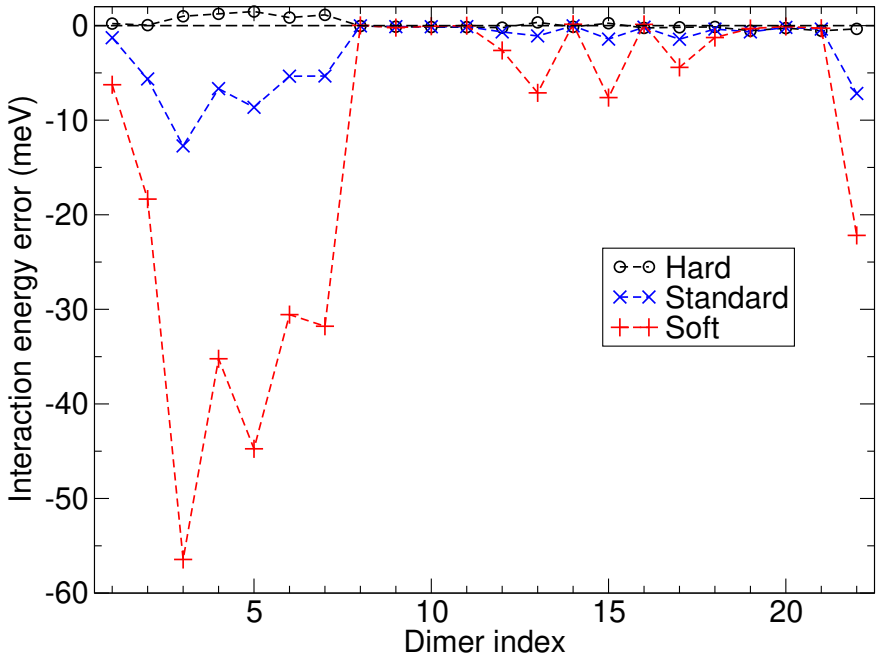

Figure 5.8: Errors of interaction energies obtained for different PAW potentials supplied with VASP for the S22 data set [105].

Deeper analysis of the results shows that the error clearly correlates with

[2]The names originate from pseudopotentials and correspond to the plane-wave basis-set cut-off that is required to obtain converged states. The "harder" the pseudopotential, the larger the cut-off needs to be.

importance of electrostatic interactions for the binding energy. There are not many options how the use of the PAW method could change the dimer interaction energy, the difference needs to be caused by errors in the electron density. We therefore obtained approximate all-electron densities for the different PAWs using the tag `LAECHG=.TRUE.` in VASP input. We then used the density obtained with Hard PAW as the reference and calculated differences for Standard and Soft PAWs. An example of this density error is shown in Fig. 5.9 for water-pyridine dimer, both for the Standard PAW (top) and the Soft PAW (bottom). As expected, the density errors are larger for the Soft PAW. Moreover, one can see that the errors are mostly localised around each atom and appear to be larger for oxygen and nitrogen than for carbon and hydrogen. Importantly, using less precise PAWs doesn't cause a considerable charge transfer between the atoms, even though we are working on a more detailed numerical assessment of this effect. For each atom the density errors can be approximately described as a sum of p-like functions, each lying in a direction of one covalent bond. As a consequence, the total density error is large and dipole-like for oxygen in water. In contrast, for carbons in uracil it resembles more an f orbital, see Fig. 5.9.

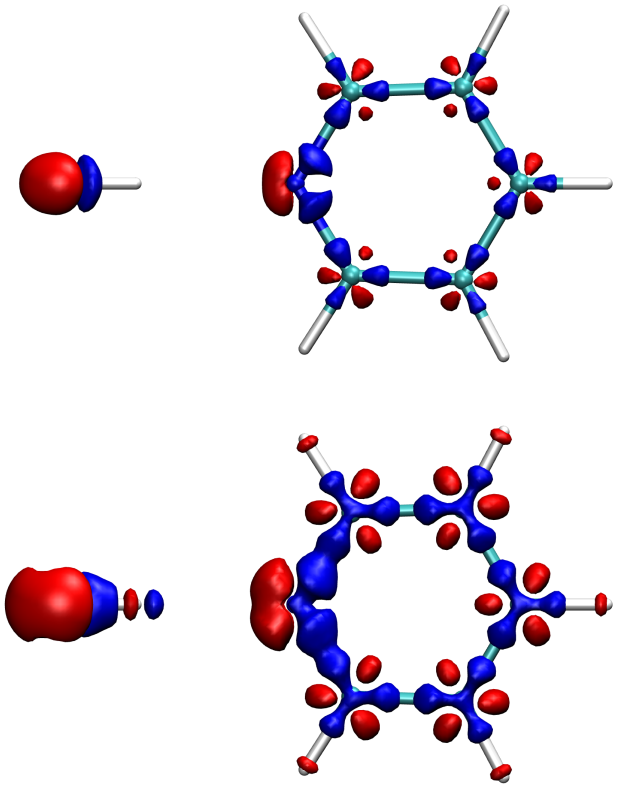

Figure 5.9: Approximate all electron density difference for Standard and Hard PAWs (top) and Soft and Hard PAWs (bottom) obtained for water-pyridine dimer. The isosurfaces are drawn at values of 0.02 Å^{-3} (blue) and -0.02 Å^{-3} (red).

The understanding that the PAW potentials lead to density errors allow to develop several ways that correct either the density error or the subsequent error of the binding energy. One of the simple such models that we proposed uses the fact that the errors are dipole-like and large for oxygen and nitrogen. Therefore,

we use a point dipole to represent the density error and put it on all oxygen atoms and those nitrogen atoms that don't have three neighbouring atoms lying in the same plane as the nitrogen. The dipoles then interact electrostatically with the rest of the system which we represent by partial charges obtained from iterative Hirshfeld partitioning [21, 197, 198]. This model leads to the following estimate of the error in the interaction energy of a dimer

$$\Delta E_{\text{dipole}} = \sum_{i\in\text{A}} \sum_{j\in\text{B}} \left[\frac{\boldsymbol{\mu}_i \cdot \mathbf{r}_{ij} q_j}{r_{ij}^3} + \frac{\boldsymbol{\mu}_j \cdot \mathbf{r}_{ij} q_i}{r_{ij}^3} + \frac{\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j}{r_{ij}^3} - 3\frac{(\boldsymbol{\mu}_i \cdot \boldsymbol{r}_{ij})(\boldsymbol{\mu}_j \cdot \boldsymbol{r}_{ij})}{r_{ij}^5} \right], \tag{5.1}$$

where A and B are the two monomers and the sums go over all the atoms on each of the monomers, q_i and μ_i are partial charges and dipoles assigned to atom i, respectively. In Eq. 5.1 the first two terms represent the interaction of the density error (dipole) with the density (point charges), the last two terms give the dipole-dipole interaction between the density errors. The magnitudes of the dipoles (that is the dipoles of the density errors) can be fitted but we calculated them from errors of dipoles of isolated water, formaldehyde, and ammonia molecules. We used water to obtain error for oxygen with two bonds and formaldehyde for oxygen with a single oxygen bond.

In Fig. 5.10 we show an example of the performance of the correction for binding curve of water-ammonia dimer. The structure is such that the water oxygen is hydrogen bond acceptor. One can see that the simple correction substantially decreases the errors. The predictions obtained with corrected Soft PAW reach the quality of Standard PAW for almost all of the distances and actually surpass it above 3 Å. Standard PAW is then within 1 meV from the Hard PAW for O$\cdots$H distances above 2.4 Å while without correction the error is still 2.5 meV for this distance. Interestingly, the cancellation is not perfect, even though we use the dipole errors for ammonia and water. This is likely due to change of the density errors due to polarization or higher-order contributions.

Apart from the long-range error in electrostatic interaction we also observe an error for short separations which has an exponential decay. This error is likely to affect atomisation energies but so far we have not tested models to correct for it. Moreover, our dipole-point charge model of Eq. 5.1 is very simple and there are many ways how it can be improved, such as adding higher-order moments. If an accurate correction is developed, it would allow one to use the Soft PAWs or soft pseudopotentials to perform routine calculations at reduced computational cost over Standard or Hard PAWs. Such combination would be very helpful in situations in which the energy of a large number of systems needs to be calculated, such as for crystal structure prediction, for systems studied in Chapter 4 or to run molecular dynamics of complex materials [171].

From a broader perspective, we showed that interaction energies of molecular dimers are very helpful to understand errors of PAW data sets or pseudopotentials. So far, other tests of PAWs or pseudopotentials focused more on other than non-covalent interactions and used atomic solids [199, 200] or properties of small molecules [201]. For solids, one can assess the magnitude of deviation from a reference binding curve and make a corresponding statement about the quality of the tested PAW or pseudopotential. However, it is difficult to say more about the cause of the deviation or to perform additional analysis on the atomic solid. Reasons for this are the high symmetry of the solid and the very limited number

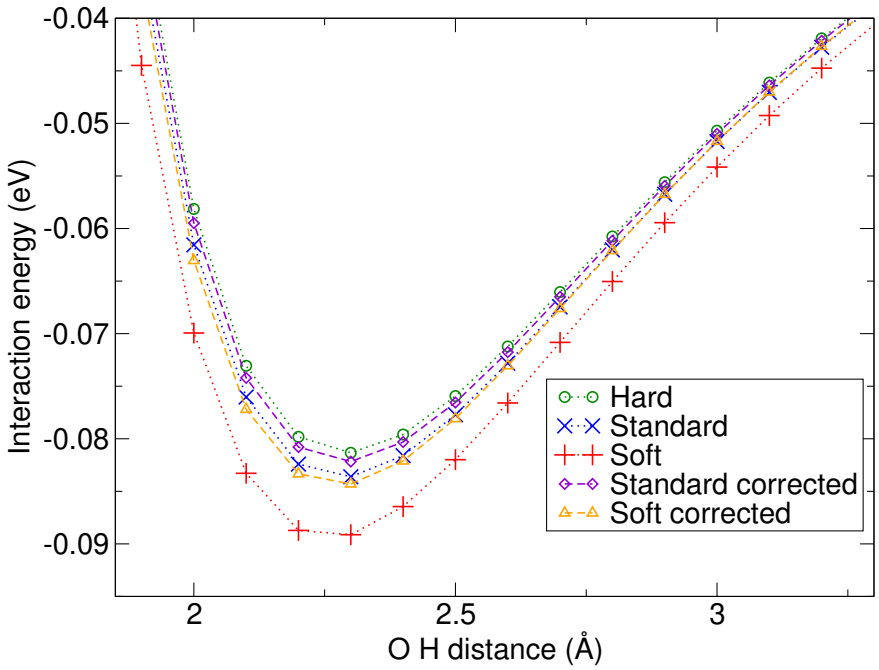

Figure 5.10: Binding curve of water-ammonia dimer obtained for different PAW potentials. The results of Standard and Soft PAWs were corrected using Eq. 5.1 with iterative Hirshfeld charges employed.

of structures that can be analysed. The situation differs for molecular dimers where we can freely change the orientation of the molecules and modify their composition so that the number of data that can be analysed is much larger. Moreover, the relation of the error to the electron density gives the developers a clear measure how to test or improve their PAW data sets or pseudopotentials. In this regard, our initial work focused on VASP with which we have a substantial experience, but we are currently testing data sets within other codes as well.

5.4 Conclusions

I presented two examples of our work that focused on analysis of precision of binding energies of molecular solids and dimers. While precision of results doesn't look like an exciting topic it's crucial to consider it and analyse it in the phase of development of novel approaches. I'd say that for calculations of highly accurate binding energies of molecular solids we are still in this phase. Indeed, the precise calculation of binding energy of benzene was published only a decade ago [188] and diffusion Monte Carlo with low uncertainties for several systems appeared around five years ago [91]. Within this time, there has been a substantial progress in developing coupled clusters methods for periodic materials which also includes considerable analysis of how various settings affect the result [202–204]. The gained understanding will hopefully establish reliable parameters and set-ups which will help researchers to obtain results with minimised uncertainties.

6. Outlook

In this thesis I discussed results obtained within two topics of our research: tests of accuracy and precision of quantum mechanical methods and understanding of experimental observations via computational techniques. The results illustrate the set of methods that we can use and that we are, in some cases, also developing.

The tests of accuracy of the random phase approximation (RPA) presented in Chapter 3 showed that accurate binding energies are obtained when states from the Perdew-Burke-Ernzerhof (PBE) [6] functional and singles corrections [42, 81] are used. We also showed that there are substantial error cancellations between different many-body orders that lead to the good results. However, one should keep in mind that the errors and the extent of error cancellation is much smaller for PBE-based RPA than for PBE itself. In this light, the results are probably expectable, PBE-based RPA is very good but not perfect. In the end, RPA is an approximate method and the simplifications have to appear somewhere. With the gained knowledge, what are the systems and properties that we should look next?

One of the areas where RPA could be useful, even with the PBE input, is the description of energy differences between polymorphs or phases of molecular crystals. We have already shown that RPA with or without the singles corrections describes the energy differences between ice phases well [91, 184]. The reason is that the RPA errors for the many-body terms are likely similar for the different structures and they thus mostly cancel when calculating energy differences. Concerning polymorphs, we have so far tested RPA on one system in the most recent test on crystal structure prediction. The RPA results were good but qualitatively comparable to results obtained with dispersion corrected density functional theory (DFT-D) methods. Nevertheless, we plan to test RPA for other sets of polymorphs to understand more the quality of its predictions and possible ways to improve it. One such way would be to replace monomer terms with a higher-level scheme. This scheme is useful when the lower level method predicts incorrectly energies of different monomer conformations that appear in the different polymorphs. The method was shown to be quite effective for DFT-based approaches [205, 206].

Our work on the many-body expansion (MBE) of RPA was motivated by the possibility to use it within the correction scheme. That is, we would like to correct the periodic RPA binding energy with high level coupled clusters terms in order to obtain reference binding energies. We found that RPA based on states from PBE and other semi-local functionals is less suitable for this task due to a slow convergence of three-body corrections and their substantial dependence on the basis-set size. The situation improves when Fock (exact) exchange is used in the mean-field calculation, that is when RPA is evaluated on states calculated by hybrid functionals or on Hartree-Fock (HF) states. Therefore, in the subsequent work we want to follow this direction for the development of the correction scheme.

The number of calculations that one needs to do within MBE is very large and can be time consuming. We are thus testing how well can be the different contributions calculated by simple models that are based, e.g., on point charges or multipole expansion. These models also allow us to understand the physics

that governs the behaviour of the different methods. In this regard, the three-body errors of RPA are rather intriguing, it'd be very useful to develop a model for the errors that we observed for the short hydrocarbons in P5. Finally, we found it very difficult to converge the four-body terms. An accurate model for these interactions would help to understand how far we were from the converged values in P5 and, especially, for methanol and ammonia in P8.

Our work on accuracy of theoretical methods lead us to recognise the importance of precision of the calculated properties. As part of precision analysis, in P9 we identified issues that appear when using less hard PAW data sets to calculate interaction energies. In the future, we want to develop a more general correction for these errors that would allow researchers to improve the precision of their calculations or reduce their computational cost. Our plan is to expand our data set with reference binding curves of molecular dimers to cover more elements, assess the predictions of more computational packages, and make it widely available, in a similar way to what's available for atomic solids [207, 208]. Hopefully, this could become a very useful resource for users interested in calculation of cohesive properties of molecular clusters, solids, and other systems.

Overall, I hope that within this thesis I demonstrated our contributions to the development of quantum mechanical methods and to understanding of their reliability as well as illustrated some of the possible applications to analyse experimental observations.

Bibliography

(1) Mallada, B.; Gallardo, A.; Lamanec, M.; de la Torre, B.; Špirko, V.; Hobza, P.; Jelinek, P. Real-space imaging of anisotropic charge of σ-hole by means of Kelvin probe force microscopy. *Science* **2021**, *374*, 863–867.

(2) Heath, J. R.; Ratner, M. A. Molecular Electronics. *Physics Today* **2003**, *56*, 43–49.

(3) Li, B.; Michaelides, A.; Scheffler, M. How strong is the bond between water and salt? *Surface Science* **2008**, *602*, L135.

(4) Hohenberg, P.; Kohn, W. Inhomogeneous electron gas. *Physical Review B* **1964**, *136*, B864.

(5) Kohn, W.; Sham, L. J. Self-consistent equations including exchange and correlation effects. *Physical Review* **1965**, *140*, A1133.

(6) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Physical Review Letters* **1996**, *77*, *ibid*, **78**, 1396 (1997), 3865.

(7) Kristyán, S.; Pulay, P. Can (semi)local density functional theory account for the London dispersion forces? *Chemical Physics Letters* **1994**, *229*, 175.

(8) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment. *The Journal of Chemical Physics* **2001**, *114*, 5149.

(9) Wu, X.; Vargas, M. C.; Nayak, S.; Lotrich, V.; Scoles, G. Towards extending the applicability of density functional theory to weakly bound systems. *The Journal of Chemical Physics* **2001**, *115*, 8748.

(10) Wu, Q.; Yang, W. Empirical correction to density functional theory for van der Waals interactions. *The Journal of Chemical Physics* **2002**, *116*, 515.

(11) Van Mourik, T.; Gdanitz, R. J. A critical note on density functional theory studies on rare-gas dimers. *The Journal of Chemical Physics* **2002**, *116*, 9620.

(12) Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of Computational Chemistry* **2004**, *25*, 1463.

(13) Zimmerli, U.; Parrinello, M.; Koumoutsakos, P. Dispersion corrections to density functional for water aromatic interactions. *The Journal of Chemical Physics* **2004**, *120*, 2693.

(14) Neumann, M.; Perrin, M. Energy ranking of molecular crystals using density functional theory calculations and an empirical van der Waals correction. *Journal of Physical Chemistry B* **2005**, *109*, 15531–15541.

(15) Andersson, Y.; Langreth, D. C.; Lundqvist, B. I. van der Waals Interactions in Density-Functional Theory. *Physical Review Letters* **1996**, *76*, 102.

(16) Dobson, J. F.; Dinte, B. P. Constraint Satisfaction in Local and Gradient Susceptibility Approximations: Application to a van der Waals Density Functional. *Physical Review Letters* **1996**, *76*, 1780.

(17) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. Van der Waals density functional for general geometries. *Physical Review Letters* **2004**, *92*, 246401.

(18) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry* **2006**, *27*, 1787.

(19) Zhang, Y.; Yang, W. Comment on "Generalized gradient approximation made simple". *Physical Review Letters* **1998**, *80*, 890.

(20) Klimeš, J.; Bowler, D. R.; Michaelides, A. Chemical accuracy for the van der Waals density functional. *Journal of Physics: Condensed Matter* **2010**, *22*, 022201.

(21) Bučko, T.; Lebègue, S.; Hafner, J.; Ángyán, J. G. Improved density dependent correction for the description of london dispersion forces. *Journal of Chemical Theory and Computation* **2013**, *9*, 4293.

(22) Marsman, M.; Grüneis, A.; Paier, J.; Kresse, G. Second-order Moller-Plesset perturbation theory applied to extended systems. I. Within the projector-augmented-wave formalism using a plane wave basis set. *The Journal of Chemical Physics* **2009**, *130*, 184103.

(23) Harl, J.; Kresse, G. Cohesive energy curves for noble gas solids calculated by adiabatic connection fluctuation-dissipation theory. *Physical Review B* **2008**, *77*, 045136.

(24) Klimeš, J.; Kresse, G. Kohn-Sham band gaps and potentials of solids from the optimised effective potential method within the random phase approximation. *The Journal of Chemical Physics* **2014**, *140*, 054516.

(25) Hybertsen, M. S.; Louie, S. G. Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Physical Review B* **1986**, *34*, 5390–5413.

(26) Shishkin, M.; Kresse, G. Implementation and performance of the frequency-dependent GW method within the PAW framework. *Physical Review B* **2006**, *74*, 035101.

(27) Hedin, L. New Method for Calculating the One-Particle Green's Function with Application to the Electron-Gas Problem. *Phys. Rev.* **1965**, *139*, A796–A823.

(28) Shih, B.; Xue, Y.; Zhang, P.; Cohen, M. L.; Louie, S. G. Quasiparticle band gap of ZnO: High accuracy from the conventional G_0W_0 approach. *Physical Review Letters* **2010**, *105*, 146401.

(29) Baroni, S.; Gebauer, R.; Malcioğlu, O. B.; Saad, Y.; Umari, P.; Xian, J. Harnessing molecular excited states with Lanczos chains. *Journal of Physics: Condensed Matter* **2010**, *22*, 074204.

(30) Friedrich, C.; Müller, M. C.; Blügel, S. Band convergence and linearization error correction of all-electron GW calculations: The extreme case of zinc oxide. *Phys. Rev. B* **2011**, *83*, *ibid*, **84**, 039906(E) (2011), 081101(R).

(31) Gulans, A. Towards numerically accurate many-body perturbation theory: Short-range correlation effects. *The Journal of Chemical Physics* **2014**, *141*, 164127.

(32) Klimeš, J.; Kaltak, M.; Kresse, G. Predictive *GW* calculations using plane waves and pseudopotentials. *Physical Review B* **2014**, *90*, 075125.

(33) Otero-de-la Roza, A.; Johnson, E. R. A benchmark for non-covalent interactions in solids. *The Journal of Chemical Physics* **2012**, *137*, 054103.

(34) Reilly, A. M.; Tkatchenko, A. Understanding the role of vibrations, exact exchange, and many-body van der Waals interactions in the cohesive properties of molecular crystals. *The Journal of Chemical Physics* **2013**, *139*, 024705.

(35) Lazić, P.; Atodiresei, N.; Alaei, M.; Calciuć, V.; Blügel, S. JuNoLo - Jülich nonlocal code for parallel post-processing evaluation of vdW-DF correlation energy. *Comput. Phys. Commun.* **2010**, *181*, 371.

(36) Řezáč, J. Non-Covalent Interactions Atlas Benchmark Data Sets: Hydrogen Bonding. *Journal of Chemical Theory and Computation* **2020**, *16*, 2355–2368.

(37) Bohm, D.; Pines, D. A Collective Description of Electron Interactions. I. Magnetic Interactions. *Physical Review* **1951**, *82*, 625–634.

(38) Ren, X.; Rinke, P.; Joas, C.; Scheffler, M. Random-phase approximation and its applications in computational chemistry and materials science. *Journal of Materials Science* **2012**, *47*, 7447.

(39) Eshuis, H.; Furche, F. Basis set convergence of molecular correlation energy differences within the random phase approximation. *The Journal of Chemical Physics* **2012**, *136*, 084105.

(40) Chen, G. P.; Voora, V. K.; Agee, M. M.; Balasubramani, S. G.; Furche, F. Random-Phase Approximation Methods. *Annual Review of Physical Chemistry* **2017**, *68*, 421–445.

(41) Niquet, Y. M.; Fuchs, M.; Gonze, X. Exchange-correlation potentials in the adiabatic connection fluctuation-dissipation framework. *Physical Review A* **2003**, *68*, 032507.

(42) Klimeš, J.; Kaltak, M.; Maggio, E.; Kresse, G. Singles correlation energy contributions in solids. *The Journal of Chemical Physics* **2015**, *143*, 102816.

(43) Harris, J.; Jones, R. O. Surface-energy of a bounded electron-gas. *Journal of Physics F: Metal Physics* **1974**, *4*, 1170.

(44) Langreth, D. C.; Perdew, J. P. Exchange-correlation energy of a metallic surface. *Solid State Commun.* **1975**, *17*, 1425.

(45) Gunnarsson, O.; Lundqvist, B. I. Exchange and correlation in atoms, molecules, and solids by spin-density functional formalism. *Phys. Rev. B* **1976**, *13*, 4274.

(46) Langreth, D. C.; Perdew, J. P. Exchange-correlation energy of a metallic surface - wave-vector analysis. *Phys. Rev. B* **1977**, *15*, 2884.

(47) Gajdoš, M.; Hummer, K.; Kresse, G.; Furthmüller, J.; Bechstedt, F. Linear optical properties in the projector-augmented wave methodology. *Physical Review B* **2006**, *73*, 045112.

(48) Kaltak, M.; Klimeš, J.; Kresse, G. Low scaling algorithms for the random phase approximation: Imaginary time and Laplace transformations. *Journal of Chemical Theory and Computation* **2014**, *10*, 2498.

(49) Gross, E. K. U.; Kohn, W. Local density-functional theory of frequency-dependent linear response. *Physical Review Letters* **1985**, *55*, 2850–2852.

(50) Olsen, T.; Thygesen, K. S. Random phase approximation applied to solids, molecules, and graphene-metal interfaces: From van der Waals to covalent bonding. *Physical Review B* **2013**, *87*, 075111.

(51) Olsen, T.; Thygesen, K. S. Beyond the random phase approximation: Improved description of short-range correlation by a renormalized adiabatic local density approximation. *Physical Review B* **2013**, *88*, 115131.

(52) Axilrod, B. M.; Teller, E. Interaction of the van der Waals type between three atoms. *The Journal of Chemical Physics* **1943**, *11*, 299.

(53) Muto, Y. Force between nonpolar molecules. *Proceedings of the Physico-Mathematical Society of Japan* **1943**, *17*, 629.

(54) Hobza, P.; Selzle, H. L.; Schlag, E. W. Potential Energy Surface for the Benzene Dimer. Results of ab Initio CCSD(T) Calculations Show Two Nearly Isoenergetic Structures: T-Shaped and Parallel-Displaced. *The Journal of Physical Chemistry* **1996**, *100*, 18790–18794.

(55) Hill, J. G.; Platts, J. A. Spin-Component Scaling Methods for Weak and Stacking Interactions. *Journal of Chemical Theory and Computation* **2007**, *3*, 80–85.

(56) Řezáč, J.; Greenwell, C.; Beran, G. J. O. Accurate Noncovalent Interactions via Dispersion-Corrected Second-Order Møller-Plesset Perturbation Theory. *Journal of Chemical Theory and Computation* **2018**, *14*, 4711–4721.

(57) Eshuis, H.; Furche, F. A parameter-free density functional that works for noncovalent interactions. *Journal of Physical Chemistry Letters* **2011**, *2*, 983.

(58) Perdew, J. P.; Schmidt, K. In *Density functional theory and its applications to materials*, Van Doren, V. E., Van Alsenoy, K., Geerlings, P., Eds.; American Institute of Physics: Melville, NY: 2001.

(59) Grimme, S.; Steinmetz, M. A computationally efficient double hybrid density functional based on the random phase approximation. *Physical Chemistry Chemical Physics* **2016**, *18*, 20926–20937.

(60) Mezei, P. D.; Kállay, M. Construction of a Range-Separated Dual-Hybrid Direct Random Phase Approximation. *Journal of Chemical Theory and Computation* **2019**, *15*, 6678–6687.

(61) Kalai, C.; Mussard, B.; Toulouse, J. Range-separated double-hybrid density-functional theory with coupled-cluster and random-phase approximations. *The Journal of Chemical Physics* **2019**, *151*, 074102.

(62) Adamson, R. D.; Dombroski, J. P.; Gill, P. M. Chemistry without Coulomb tails. *Chemical Physics Letters* **1996**, *254*, 329–336.

(63) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **2003**, *118*, 8207.

(64) Toulouse, J.; Zhu, W.; Ángyán, J. G.; Savin, A. Range-separated density-functional theory with the random-phase approximation: Detailed formalism and illustrative applications. *Phys. Rev. A* **2010**, *82*, 032502.

(65) Bruneval, F. Range-separated approach to the RPA correlation applied to the van der Waals bond and to diffusion of defects. *Physical Review Letters* **2012**, *108*, 256403.

(66) Franck, O.; Mussard, B.; Luppi, E.; Toulouse, J. Basis convergence of range-separated density-functional theory. *The Journal of Chemical Physics* **2015**, *142*, 074107.

(67) Verma, P.; Bartlett, R. J. Increasing the applicability of density functional theory. II. Correlation potentials from the random phase approximation and beyond. *The Journal of Chemical Physics* **2012**, *136*, 044105.

(68) Bleiziffer, P.; Heßelmann, A.; Görling, A. Efficient self-consistent treatment of electron correlation within the random phase approximation. *The Journal of Chemical Physics* **2013**, *139*, 084113.

(69) Nguyen, N. L.; Colonna, N.; de Gironcoli, S. Ab initio self-consistent total-energy calculations within the EXX/RPA formalism. *Physical Review B* **2014**, *90*, 045138.

(70) Hellgren, M.; Caruso, F.; Rohr, D. R.; Ren, X.; Rubio, A.; Scheffler, M.; Rinke, P. Static correlation and electron localization in molecular dimers from the self-consistent RPA and *GW* approximation. *Physical Review B* **2015**, *91*, 165110.

(71) Yu, J. M.; Nguyen, B. D.; Tsai, J.; Hernandez, D. J.; Furche, F. Selfconsistent random phase approximation methods. *The Journal of Chemical Physics* **2021**, *155*, 040902.

(72) Lu, D.; Li, Y.; Rocca, D.; Galli, G. Ab initio Calculation of van der Waals Bonded Molecular Crystals. *Phys. Rev. Lett.* **2009**, *102*, 206411.

(73) Scuseria, G. E.; Henderson, T. M.; Sorensen, D. C. The ground state correlation energy of the random phase approximation from a ring coupled cluster doubles approach. *The Journal of Chemical Physics* **2008**, *129*, 231101.

(74) Jansen, G.; Liu, R.; Ángyán, J. G. On the equivalence of ring-coupled cluster and adiabatic connection fluctuation-dissipation theorem random phase approximation correlation energy expressions. *The Journal of Chemical Physics* **2010**, *133*, 154106.

(75) Freeman, D. L. Coupled-cluster expansion applied to the electron gas: Inclusion of ring and exchange effects. *Physical Review B* **1977**, *15*, 5512–5521.

(76) Grüneis, A.; Marsman, M.; Harl, J.; Schimka, L.; Kresse, G. Making the random phase approximation to electronic correlation accurate. *The Journal of Chemical Physics* **2009**, *131*, 154115.

(77) Li, Y.; Lu, D.; Nguyen, H.-V.; Galli, G. van der Waals Interactions in Molecular Assemblies from First-Principles Calculations. *Journal of Physical Chemistry A* **2010**, *114*, 1944.

(78) Klimeš, J.; Harl, J.; Kresse, G., unpublished results.

(79) Schimka, L.; Harl, J.; Stroppa, A.; Grüneis, A.; Marsman, M.; Mittendorfer, F.; Kresse, G. Accurate surface and adsorption energies from many-body perturbation theory. *Nature Materials* **2010**, *9*, 741.

(80) Al-Hamdani, Y. S.; Rossi, M.; Alfè, D.; Tsatsoulis, T.; Ramberger, B.; Brandenburg, J. G.; Zen, A.; Kresse, G.; Grüneis, A.; Tkatchenko, A.; Michaelides, A. Properties of the water to boron nitride interaction: From zero to two dimensions with benchmark accuracy. *The Journal of Chemical Physics* **2017**, *147*, 044710.

(81) Ren, X.; Tkatchenko, A.; Rinke, P.; Scheffler, M. Beyond the Random-Phase Approximation for the electron correlation energy: The importance of single excitations. *Physical Review Letters* **2011**, *106*, 153003.

(82) Řezáč, J.; Huang, Y.; Hobza, P.; Beran, G. J. O. Benchmark calculations of three-body intermolecular interactions and the perform- ance of low-cost electronic structure methods. *Journal of Chemical Theory and Computation* **2015**, *11*, 3065.

(83) Harl, J.; Kresse, G. Accurate Bulk Properties from Approximate Many-Body Techniques. *Physical Review Letters* **2010**, *103*, 056401.

(84) Kutzelnigg, W.; Morgan, J. D. Rates of convergence of the partial-wave expansions of atomic correlation energies. *The Journal of Chemical Physics* **1992**, *96*, 4484.

(85) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *The Journal of Chemical Physics* **1992**, *96*, 6796.

(86) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence in correlated calculations on water. *Journal of Physical Chemistry* **1997**, *9639*, 106.

(87) Adamo, C.; Barone, V. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of Chemical Physics* **1999**, *110*, 6158.

(88) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.

(89) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry* **2011**, *32*, 1456.

(90) Tkatchenko, A.; DiStasio, R. A.; Car, R.; Scheffler, M. Accurate and Efficient Method for Many-Body van der Waals Interactions. *Physical Review Letters* **2012**, *108*, 236402.

(91) Zen, A.; Brandenburg, J. G.; Klimeš, J.; Tkatchenko, A.; Alfè, D.; Michaelides, A. Fast and accurate quantum Monte Carlo for molecular crystals. *Proceedings of the National Academy of Sciences of U. S. A.* **2018**, *115*, 1724–1729.

(92) Feibelman, P. J.; Hammer, B.; Norskov, J. K.; Wagner, F.; Scheffler, M.; Stumpf, R.; Watwe, R.; Dumesic, J. The CO/Pt(111) puzzle. *Journal of Physical Chemistry B* **2001**, *105*, 4018.

(93) Göltl, F.; Grüneis, A.; Bučko, T.; Hafner, J. Van der Waals interactions between hydrocarbon molecules and zeolites: Periodic calculations at different levels of theory, from density functional theory to the random phase approximation and Møller-Plesset perturbation theory. *The Journal of Chemical Physics* **2012**, *137*, 114111.

(94) Göltl, F.; Hafner, J. Structure and properties of metal-exchanged zeolites studied using gradient-corrected and hybrid functionals. II. Electronic structure and photoluminescence spectra. *The Journal of Chemical Physics* **2012**, *136*, 064502.

(95) Foulkes, W. M. C.; Mitas, L.; Needs, R. J.; Rajagopal, G. Quantum Monte Carlo simulations of solids. *Reviews of Modern Physics* **2001**, *73*, 33.

(96) Dubecký, M.; Jurečka, P.; Derian, R.; Hobza, P.; Otyepka, M.; Mitas, L. Quantum Monte Carlo methods describe noncovalent interactions with subchemical accuracy. *Journal of Chemical Theory and Computation* **2013**, *9*, 4287.

(97) Hättig, C.; Tew, D. P.; Köhn, A. Communications: Accurate and efficient approximations to explicitly correlated coupled-cluster singles and doubles, CCSD-F12. *The Journal of Chemical Physics* **2010**, *132*, 231102.

(98) Tkatchenko, A.; Scheffler, M. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Physical Review Letters* **2009**, *102*, 073005.

(99) Caldeweyher, E.; Bannwarth, C.; Grimme, S. Extension of the D3 dispersion coefficient model. *The Journal of Chemical Physics* **2017**, *147*, 034112.

(100) Caldeweyher, E.; Ehlert, S.; Hansen, A.; Neugebauer, H.; Spicher, S.; Bannwarth, C.; Grimme, S. A generally applicable atomic-charge dependent London dispersion correction. *The Journal of Chemical Physics* **2019**, *150*, 154122.

(101) Brandenburg, J. G.; Zen, A.; Fitzner, M.; Ramberger, B.; Kresse, G.; Tsatsoulis, T.; Grüneis, A.; Michaelides, A.; Alfè, D. Physisorption of water on graphene: Sub-chemical accuracy from many-body electronic structure methods. *Journal of Physical Chemistry Letters* **2019**, *10*, 358–368.

(102) Rubeš, M.; Trachta, M.; Koudelková, E.; Bulánek, R.; Klimeš, J.; Nachtigall, P.; Bludský, O. Temperature Dependence of Carbon Monoxide Adsorption on a High-Silica H-FER Zeolite. *Journal of Physical Chemistry C* **2018**, *122*, 26088–26095.

(103) Marakatti, V. S.; Klimeš, J.; Kasinathan, P.; Sorathia, K.; Tew, D. P.; Gaigneaux, E. M. Insights on hydrogen bond assisted solvent selection in certain acid–base heterogeneous catalysis through acceptor and donor numbers. *Catalysis Science & Technology* **2021**, *11*, 1345–1357.

(104) Klein, P.; Dedecek, J.; Thomas, H. M.; Whittleton, S. R.; Klimes, J.; Brus, J.; Kobera, L.; Bryce, D. L.; Sklenak, S. NMR Crystallography of Monovalent Cations in Inorganic Matrices: Na+ Siting and the Local Structure of Na+ Sites in Ferrierites. *Journal of Physical Chemistry C* **2022**, *126*, 10686–10702.

(105) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Physical Chemistry Chemical Physics* **2006**, *8*, 1985.

(106) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *Journal of Chemical Theory and Computation* **2011**, *7*, 2427.

(107) Řezáč, J.; Riley, K. E.; Hobza, P. Benchmark Calculations of Noncovalent Interactions of Halogenated Molecules. *Journal of Chemical Theory and Computation* **2012**, *8*, 4285–4292.

(108) Goerigk, L.; Grimme, S. A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions – Assessment of Common and Reparameterized (meta-)GGA Density Functionals. *Journal of Chemical Theory and Computation* **2010**, *6*, 107–126.

(109) Goerigk, L.; Grimme, S. A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions. *Physical Chemistry Chemical Physics* **2011**, *13*, 6670.

(110) Goerigk, L.; Grimme, S. Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals—Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions. *Journal of Chemical Theory and Computation* **2011**, *7*, 291–309.

(111) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Physical Chemistry Chemical Physics* **2017**, *19*, 32184–32215.

(112) Temelso, B.; Archer, K. A.; Shields, G. C. Benchmark Structures and Binding Energies of Small Water Clusters with Anharmonicity Corrections. *Journal of Physical Chemistry A* **2011**, *115*, 12034–12046.

(113) Tkatchenko, A.; von Lilienfeld, O. A. Popular Kohn-Sham density functionals strongly overestimate many-body interactions in van der Waals systems. *Physical Review B* **2008**, *78*, 045116.

(114) Góra, U.; Podeszwa, R.; Cencek, W.; Szalewicz, K. Interaction energies of large clusters from many-body expansion. *The Journal of Chemical Physics* **2011**, *135*, 224102.

(115) Sun, J.; Ruzsinszky, A.; Perdew, J. P. Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Physical Review Letters* **2015**, *115*, 036402.

(116) Hui, K.; Chai, J.-D. SCAN-based hybrid and double-hybrid density functionals from models without fitted parameters. *The Journal of Chemical Physics* **2016**, *144*, 044114.

(117) Cox, S. J.; Towler, M. D.; Alfè, D.; Michaelides, A. Benchmarking the performance of density functional theory and point charge force fields in their description of sI methane hydrate against diffusion Monte Carlo. *The Journal of Chemical Physics* **2014**, *140*, 174703.

(118) Santra, B.; Klimeš, J.; Tkatchenko, A.; Alfè, D.; Slater, B.; Michaelides, A.; Car, R.; Scheffler, M. On the Accuracy of van der Waals Inclusive Density-Functional Theory Exchange-Correlation Functionals for Ice at Ambient and High Pressures. *The Journal of Chemical Physics* **2013**, *139*, 154702.

(119) Deible, M. J.; Tuguldur, O.; Jordan, K. D. Theoretical Study of the Binding Energy of a Methane Molecule in a $(H_2O)_{20}$ Dodecahedral Cage. *Journal of Physical Chemistry B* **2014**, *118*, 8257–8263.

(120) Lao, K. U.; Herbert, J. M. A simple correction for nonadditive dispersion within extended symmetry-adapted perturbation theory (XSAPT). *Journal of Chemical Theory and Computation* **2018**, *14*, 5128–5142.

(121) Riplinger, C.; Neese, F. An efficient and near linear scaling pair natural orbital based local coupled cluster method. *The Journal of Chemical Physics* **2013**, *138*, 034106.

(122) Werner, H.-J.; Adler, T. B.; Manby, F. R. General orbital invariant MP2-F12 theory. *The Journal of Chemical Physics* **2007**, *126*, 164102.

(123) Adler, T. B.; Knizia, G.; Werner, H.-J. A simple and efficient CCSD(T)-F12 approximation. *The Journal of Chemical Physics* **2002**, *116*, 3175.

(124) Knizia, G.; Adler T. B.and Werner, H.-J. Simplified CCSD (T)-F12 methods: Theory and benchmarks. *The Journal of Chemical Physics* **2009**, *130*, 054104.

(125) Marchetti, O.; Werner, H.-J. Accurate Calculations of Intermolecular Interaction Energies Using Explicitly Correlated Coupled Cluster Wave Functions and a Dispersion-Weighted MP2 Method. *Journal of Physical Chemistry A* **2009**, *113*, 11580–11585.

(126) Talman, J. D.; Shadwick, W. F. Optimized effective atomic central potential. *Physical Review A* **1976**, *14*, 36.

(127) Bartlett, R. J.; Grabowski, I.; Hirata, S.; Ivanov, S. The exchange-correlation potential in ab initio density functional theory. *The Journal of Chemical Physics* **2005**, *122*, 034104.

(128) Bartlett, R. J. Ab initio DFT and its role in electronic structure theory. *Molecular Physics* **2010**, *108*, 3299–3311.

(129) Śmiga, S.; Marusiak, V.; Grabowski, I.; Fabiano, E. The ab initio density functional theory applied for spin-polarized calculations. *The Journal of Chemical Physics* **2020**, *152*, 054109.

(130) Huang, Y.; Beran, G. J. O. Reliable prediction of three-body intermolecular interactions using dispersion-corrected second-order Møller-Plesset perturbation theory. *The Journal of Chemical Physics* **2015**, *143*, 044113.

(131) Chai, J.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Physical Chemistry Chemical Physics* **2008**, *10*, 6615.

(132) `www.github.com/klimes/mbelib/`.

(133) Klimeš, J. Lattice energies of molecular solids from the random phase approximation with singles corrections. *The Journal of Chemical Physics* **2016**, *145*, 094506.

(134) Ambrosetti, A.; Reilly, A. M.; DiStasio, R. A.; Tkatchenko, A. Long-range correlation energy calculated from coupled atomic response functions. *The Journal of Chemical Physics* **2014**, *140*, 18A508.

(135) `https://cccbdb.nist.gov/quadlistx.asp`, accessed 6.9.2023.

(136) Kennedy, M. R.; McDonald, A. R.; DePrince, A. E.; Marshall, M. S.; Podeszwa, R.; Sherrill, C. D. Communication: Resolving the three-body contribution to the lattice energy of crystalline benzene: Benchmark results from coupled-cluster theory. *The Journal of Chemical Physics* **2014**, *140*, 121104.

(137) Al-Hamdani, Y. S.; Ma, M.; Alfè, D.; von Lilienfeld, O. A.; Michaelides, A. Communication: Water on hexagonal boron nitride from diffusion Monte Carlo. *The Journal of Chemical Physics* **2015**, *142*, 181101.

(138) Corti, Horacio R.; Appignanesi, Gustavo A.; Barbosa, Marcia C.; Bordin, J. Rafael; Calero, Carles; Camisasca, Gaia; Elola, M. Dolores; Franzese, Giancarlo; Gallo, Paola; Hassanali, Ali; Huang, Kai; Laria, Daniel; Menéndez, Cintia A.; de Oca, Joan M. Montes; Longinotti, M. Paula; Rodriguez, Javier; Rovere, Mauro; Scherlis, Damián; Szleifer, Igal Structure and dynamics of nanoconfined water and aqueous solutions. *The European Physical Journal E* **2021**, *44*, 136.

(139) Perakis, F.; Amann-Winkel, K.; Lehmkühler, F.; Sprung, M.; Mariedahl, D.; Sellberg, J. A.; Pathak, H.; Späh, A.; Cavalca, F.; Schlesinger, D.; Ricci, A.; Jain, A.; Massani, B.; Aubree, F.; Benmore, C. J.; Loerting, T.; Grübel, G.; Pettersson, L. G. M.; Nilsson, A. Diffusive dynamics during the high-to-low density transition in amorphous ice. *Proceedings of the National Academy of Sciences* **2017**, *114*, 8193–8198.

(140) Taschin, A.; Bartolini, P.; Marcelli, A.; Righini, R.; Torre, R. A comparative study on bulk and nanoconfined water by time-resolved optical Kerr effect spectroscopy. *Faraday Discuss.* **2013**, *167*, 293–308.

(141) Paesani, F.; Voth, G. A. Quantum Effects Strongly Influence the Surface Premelting of Ice. *Journal of Physical Chemistry C* **2008**, *112*, 324–327.

(142) Pan, D.; Liu, L.-M.; Slater, B.; Michaelides, A.; Wang, E. Melting the Ice: On the Relation between Melting Temperature and Size for Nanoscale Ice Crystals. *ACS Nano* **2011**, *5*, 4562–4569.

(143) Li, Y.; Somorjai, G. A. Surface Premelting of Ice. *Journal of Physical Chemistry C* **2007**, *111*, 9631–9637.

(144) Pradzynski, C. C.; Forck, R. M.; Zeuch, T.; Slavíček, P.; Buck, U. A Fully Size-Resolved Perspective on the Crystallization of Water Clusters. *Science* **2012**, *337*, 1529–1532.

(145) Moberg, D. R.; Becker, D.; Dierking, C. W.; Zurheide, F.; Bandow, B.; Buck, U.; Hudait, A.; Molinero, V.; Paesani, F.; Zeuch, T. The end of ice I. *Proceedings of the National Academy of Sciences of U. S. A.* **2019**, *116*, 24413–24419.

(146) Novoselov, K. S.; Geim, A. K.; Morozov, S. V.; Jiang, D.; Zhang, Y.; Dubonos, S. V.; Grigorieva, I. V.; Firsov, A. A. Electric Field Effect in Atomically Thin Carbon Films. *Science* **2004**, *306*, 666–669.

(147) Malard, L. M.; Pimenta, M. A.; Dresselhaus, G.; Dresselhaus, M. S. Raman Spectroscopy in Graphene. *Physics Reports* **2009**, *473*, 51.

(148) Lazzeri, M.; Mauri, F. Nonadiabatic Kohn Anomaly in a Doped Graphene Monolayer. *Physical Review Letters* **2006**, *97*, 266407.

(149) Yan, J.; Zhang, Y.; Kim, P.; Pinczuk, A. Electric Field Effect Tuning of Electron-Phonon Coupling in Graphene. *Physical Review Letters* **2007**, *98*, 166802.

(150) Das, A.; Pisana, S.; Chakraborty, B.; Piscanec, S.; Saha, S. K.; Waghmare, U. V.; Novoselov, K. S.; Krishnamurthy, H. R.; Geim, A. K.; Ferrari, A. C.; Sood, A. K. Monitoring Dopants by Raman Scattering in an Electrochemically Top-Gated Graphene Transistor. *Nature Nanotechnology* **2008**, *3*, 210.

(151) Verhagen, T. G. A.; Vales, V.; Frank, O.; Kalbac, M.; Vejpravova, J. Temperature-Induced Strain Release via Rugae on the Nanometer and Micrometer Scale in Graphene Monolayer. *Carbon* **2017**, *119*, 535.

(152) Lee, J. E.; Ahn, G.; Shim, J.; Lee, Y. S.; Ryu, S. Optical Separation of Mechanical Strain from Charge Doping in Graphene. *Nature Communications* **2012**, *3*, 1024.

(153) Li, X.; Feng, J.; Wang, E.; Meng, S.; Klimeš, J.; Michaelides, A. Influence of water on the electronic structure of metal supported graphene: Insight from van der Waals density functional theory. *Physical Review B* **2012**, *85*, 085425.

(154) Emami, F. S.; Puddu, V.; Berry, R. J.; Varshney, V.; Patwardhan, S. V.; Perry, C. C.; Heinz, H. Force Field and a Surface Model Database for Silica to Simulate Interfacial Properties in Atomic Resolution. *Chemistry of Materials* **2014**, *26*, 2647–2658.

(155) Chen, J.; Schusteritsch, G.; Pickard, C. J.; Salzmann, C. G.; Michaelides, A. Two Dimensional Ice from First Principles: Structures and Phase Transitions. *Physical Review Letters* **2016**, *116*, 025501.

(156) Abascal, J. L. F.; Sanz, E.; García Fernández, R.; Vega, C. A potential model for the study of ices and amorphous water: TIP4P/Ice. *The Journal of Chemical Physics* **2005**, *122*, 234511.

(157) Taherian, F.; Marcon, V.; van der Vegt, N. F. A.; Leroy, F. What Is the Contact Angle of Water on Graphene? *Langmuir* **2013**, *29*, 1457–1465.

(158) Akaishi, A.; Yonemaru, T.; Nakamura, J. Formation of Water Layers on Graphene Surfaces. *ACS Omega* **2017**, *2*, 2184–2190.

(159) Herrero, C. P.; Ramírez, R. Quantum effects in graphene monolayers: Path-integral simulations. *The Journal of Chemical Physics* **2016**, *145*, 224701.

(160) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19.

(161) Plimpton, S. J. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal of Computational Physics* **1995**, *117*, 1.

(162) Ceriotti, M.; More, J.; Manolopoulos, D. E. i-PI: A Python interface for ab initio path integral molecular dynamics simulations. *Computer Physics Communications* **2014**, *185*, 1019–1026.

(163) Ceriotti, M.; Manolopoulos, D. E. Efficient First-Principles Calculation of the Quantum Kinetic Energy and Momentum Distribution of Nuclei. *Physical Review Letters* **2012**, *109*, 100604.

(164) Takaiwa, D.; Hatano, I.; Koga, K.; Tanaka, H. Phase Diagram of Water in Carbon Nanotubes. *Proceedings of the National Academy of Sciences of U. S. A.* **2008**, *105*, 39.

(165) Agrawal, K. V.; Shimizu, S.; Drahushuk, L. W.; Kilcoyne, D.; Strano, M. S. Observation of Extreme Phase Transition Temperatures of Water Confined inside Isolated Carbon Nanotubes. *Nature Nanotechnology* **2017**, *12*, 267.

(166) Pugliese, P.; Conde, M. M.; Rovere, M.; Gallo, P. Freezing Temperatures, Ice Nanotubes Structures, and Proton Ordering of TIP4P/ICE Water inside Single Wall Carbon Nanotubes. *Journal of Physical Chemistry B* **2017**, *121*, 10371.

(167) Pickard, C. J.; Needs, R. J. Ab initio random structure searching. *Journal of Physics: Condensed Matter* **2011**, *23*, 053201.

(168) Stuart, S. J.; Tutein, A. B.; Harrison, J. A. A Reactive Potential for Hydrocarbons with Intermolecular Interactions. *The Journal of Chemical Physics* **2000**, *112*, 6472.

(169) Zhao, W.-H.; Bai, J.; Yuan, L.-F.; Yang, J.; Zeng, X. C. Ferroelectric hexagonal and rhombic monolayer ice phases. *Chemical Sciences* **2014**, *5*, 1757–1764.

(170) Jaffe, R. L.; Gonnet, P.; Werder, T.; Walther, J. H.; Koumoutsakos, P. Water–Carbon Interactions 2: Calibration of Potentials using Contact Angle Data for Different Interaction Models. *Molecular Simulation* **2004**, *30*, 205–216.

(171) Hintzsche, L. E.; Fang, C. M.; Marsman, M.; Jordan, G.; Lamers, M. W.P. E.; Weeber, A. W.; Kresse, G. Defects and defect healing in amorphous $Si_3N_{4-x}H_y$: An ab initio density functional theory study. *Physical Review B* **2013**, *88*, 155204.

(172) Brandenburg, J. G.; Caldeweyher, E.; Grimme, S. Screened exchange hybrid density functional for accurate and efficient structures and interaction energies. *Physical Chemistry Chemical Physics* **2016**, *18*, 15519–15523.

(173) Brandenburg, J. G.; Bannwarth, C.; Hansen, A.; Grimme, S. B97-3c: A revised low-cost variant of the B97-D density functional method. *The Journal of Chemical Physics* **2018**, *148*, 064104.

(174) Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB). *Journal of Chemical Theory and Computation* **2011**, *7*, 931–948.

(175) Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *Journal of Chemical Theory and Computation* **2012**, *8*, 141–151.

(176) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB–An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions. *Journal of Chemical Theory and Computation* **2019**, *15*, 1652–1671.

(177) Řezáč, J.; Hobza, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the "Gold Standard," CCSD(T) at the Complete Basis Set Limit? *Journal of Chemical Theory and Computation* **2013**, *9*, 2151–2155.

(178) Řezáč, J.; Dubecký, M.; Jurečka, P.; Hobza, P. Extensions and applications of the A24 data set of accurate interaction energies. *Physical Chemistry Chemical Physics* **2015**, *17*, 19268–19277.

(179) Blöchl, P. E. Projector augmented-wave method. *Physical Review B* **1994**, *50*, 17953.

(180) Kresse, G.; Joubert, J. From ultrasoft pseudopotentials to the projector augmented wave method. *Physical Review B* **1999**, *59*, 1758.

(181) Klimeš, J.; Bowler, D. R.; Michaelides, A. A critical assessment of theoretical methods for finding reaction pathways and transition states of surface processes. *Journal of Physics: Condensed Matter* **2010**, *22*, 074203.

(182) Klimeš, J.; Bowler, D. R.; Michaelides, A. Understanding the role of ions and water molecules in the NaCl dissolution process. *The Journal of Chemical Physics* **2013**, *139*, 234702.

(183) Maschio, L.; Usvyat, D.; Schütz, M.; Civalleri, B. Periodic local Møller-Plesset second order perturbation theory method applied to molecular crystals: Study of solid NH3 and CO2 using extended basis sets. *The Journal of Chemical Physics* **2010**, *132*, 134706.

(184) Macher, M.; Klimeš, J.; Franchini, C.; Kresse, G. The random phase approximation applied to ice. *The Journal of Chemical Physics* **2014**, *140*, 084502.

(185) Hermann, A.; Schwerdtfeger, P. Ground-State Properties of Crystalline Ice from Periodic Hartree-Fock Calculations and a Coupled-Cluster-Based Many-Body Decomposition of the Correlation Energy. *Physical Review Letters* **2008**, *101*, 183005.

(186) Beran, G. J. O.; Wen, S. Predicting organic crystal lattice energies with chemical accuracy. *Journal of Physical Chemistry Letters* **2010**, *1*, 3480–3487.

(187) Červinka, C.; Fulem, M. State-of-the-art Calculations of Sublimation Enthalpies for Selected Molecular Crystals and Their Computational Uncertainty. *Journal of Chemical Theory and Computation* **2017**, *13*, 2840–2850.

(188) Yang, J.; Hu, W.; Usvyat, D.; Matthews, D.; Schütz, M.; Chan, G. K. Ab initio determination of the crystalline benzene lattice energy to sub-kilojoule/mole accuracy. *Science* **2014**, *345*, 640.

(189) Gygi, F.; Baldereschi, A. Self-consistent Hartree-Fock and screened-exchange calculations in solids - application to silicon. *Physical Review B* **1986**, *34*, 4405.

(190) Rozzi, C. A.; Varsano, D.; Marini, A.; Gross, E. K. U.; Rubio, A. Exact Coulomb cutoff technique for supercell calculations. *Physical Review B* **2006**, *73*, 205119.

(191) Spencer, J.; Alavi, A. Efficient calculation of the exact exchange energy in periodic systems using a truncated Coulomb potential. *Physical Review B* **2008**, *77*, 193110.

(192) Richard, R. M.; Lao, K. U.; Herbert, J. M. Understanding the many-body expansion for large systems. I. Precision considerations. *The Journal of Chemical Physics* **2014**, *141*.

(193) Stone, A. Distributed multipole analysis, or how to describe a molecular charge distribution. *Chemical Physics Letters* **1981**, *83*, 233–239.

(194) Stone, A., *The Theory of Intermolecular Forces*; Oxford University Press: 2013.

(195) Hafner, J. Ab-initio simulations of materials using VASP: Density-functional theory and beyond. *Journal of Computational Chemistry* **2008**, *29*, 2044–2078.

(196) Witte, J.; Neaton, J. B.; Head-Gordon, M. Push it to the limit: comparing periodic and local approaches to density functional theory for intermolecular interactions. *Molecular Physics* **2019**, *117*, 1298–1305.

(197) Bultinck, P.; Van Alsenoy, C.; Ayers, P. W.; Carbó-Dorca, R. Critical analysis and extension of the Hirshfeld atoms in molecules. *The Journal of Chemical Physics* **2007**, *126*, 144111.

(198) Bultinck, P.; Ayers, P. W.; Fias, S.; Tiels, K.; Van Alsenoy, C. Uniqueness and basis set dependence of iterative Hirshfeld charges. *Chemical Physics Letters* **2007**, *444*, 205–208.

(199) Dal Corso, A. Pseudopotentials periodic table: From H to Pu. *Computational Materials Science* **2014**, *95*, 337–350.

(200) Lejaeghere, K.; Bihlmayer, G.; Björkman, T.; Blaha, P.; Blügel, S.; Blum, V.; Caliste, D.; Castelli, I. E.; Clark, S. J.; Corso, A. D.; de Gironcoli, S.; Deutsch, T.; Dewhurst, J. K.; Marco, I. D.; Draxl, C.; Dułak, M.; Eriksson, O.; Flores-Livas, J. A.; Garrity, K. F.; Genovese, L.; Giannozzi, P.; Giantomassi, M.; Goedecker, S.; Gonze, X.; Grånäs, O.; Gross, E. K. U.; Gulans, A.; Gygi, F.; Hamann, D. R.; Hasnip, P. J.; Holzwarth, N. A. W.; Iuşan, D.; Jochym, D. B.; Jollet, F.; Jones, D.; Kresse, G.; Koepernik, K.; Küçükbenli, E.; Kvashnin, Y. O.; Locht, I. L. M.; Lubeck, S.; Marsman, M.; Marzari, N.; Nitzsche, U.; Nordström, L.; Ozaki, T.; Paulatto, L.; Pickard, C. J.; Poelmans, W.; Probert, M. I. J.; Refson, K.; Richter, M.; Rignanese, G.-M.; Saha, S.; Scheffler, M.; Schlipf, M.; Schwarz, K.; Sharma, S.; Tavazza, F.; Thunström, P.; Tkatchenko, A.; Torrent, M.; Vanderbilt, D.; van Setten, M. J.; Speybroeck, V. V.; Wills, J. M.; Yates, J. R.; Zhang, G.-X.; Cottenier, S. Reproducibility in density functional theory calculations of solids. *Science* **2016**, *351*, aad3000.

(201) Adllan, A. A.; Corso, A. D. Ultrasoft pseudopotentials and projector augmented-wave data sets: application to diatomic molecules. *Journal of Physics: Condensed Matter* **2011**, *23*, 425501.

(202) Liao, K.; Grüneis, A. Communication: Finite size correction in periodic coupled cluster theory calculations of solids. *The Journal of Chemical Physics* **2016**, *145*, 141102.

(203) McClain, J.; Sun, Q.; Chan, G. K.-L.; Berkelbach, T. C. Gaussian-Based Coupled-Cluster Theory for the Ground-State and Band Structure of Solids. *Journal of Chemical Theory and Computation* **2017**, *13*, 1209–1218.

(204) Irmler, A.; Grüneis, A. Particle-particle ladder based basis-set corrections applied to atoms and molecules using coupled-cluster theory. *The Journal of Chemical Physics* **2019**, *151*, 104107.

(205) Taylor, C. R.; Bygrave, P. J.; Hart, J. N.; Allan, N. L.; Manby, F. R. Improving density functional theory for crystal polymorph energetics. *Physical Chemistry Chemical Physics* **2012**, *14*, 7739.

(206) Greenwell, C.; Beran, G. J. O. Inaccurate Conformational Energies Still Hinder Crystal Structure Prediction in Flexible Organic Molecules. *Crystal Growth & Design* **2020**, *20*, 4875–4881.

(207) https://www.materialscloud.org/discover/sssp/table/efficiency/.

(208) https://molmod.ugent.be/deltacodesdft.

List of attached publications

P1 *Lattice energies of molecular solids from the random phase approximation with singles corrections*
J. Klimeš, The Journal of Chemical Physics **145**, 094506 (2016)
doi: 10.1063/1.4962188

P2 *Efficient and accurate description of adsorption in zeolites*
J. Klimeš and D. P. Tew, The Journal of Chemical Physics **151**, 234108 (2019)
doi: 10.1063/1.5123425

P3 *Random Phase Approximation Applied to Many-Body Noncovalent Systems*
M. Modrzejewski, S. Yourdkhani, and J. Klimeš, Journal of Chemical Theory and Computation **16**, 427–442 (2020)
doi: 10.1021/acs.jctc.9b00979

P4 *Random-Phase Approximation in Many-Body Noncovalent Systems: Methane in a Dodecahedral Water Cage*
M. Modrzejewski, S. Yourdkhani, S. Śmiga, and J. Klimeš, Journal of Chemical Theory and Computation **17**, 804–817 (2021)
doi: 10.1021/acs.jctc.0c00966

P5 *Assessment of random phase approximation and second-order Møller-Plesset perturbation theory for many-body interactions in solid ethane, ethylene, and acetylene*
Pham N. K., M. Modrzejewski, and J. Klimeš, The Journal of Chemical Physics **158**, 144119 (2023)
doi: 10.1063/5.0142348

P6 *Anomalous Freezing of Low-Dimensional Water Confined in Graphene Nanowrinkles*
T. Verhagen, J. Klimeš, B. Pacáková, M. Kalbáč, and J. Vejpravová, ACS Nano **14**, 15587 (2020)
doi: 10.1021/acsnano.0c03161

P7 *Ferroelectric 2D ice under graphene confinement*
H.-T. Chin, J. Klimeš, I-F. Hu, D.-R. Chen, H.-T. Nguyen, T.-W. Chen, S.-W. Ma, M. Hofmann, C.-T. Liang, and Y.-P. Hsieh, Nature Communications **12**, 6291 (2021)
doi: 10.1038/s41467-021-26589-x

P8 *Binding energies of molecular solids from fragment and periodic approaches*
J. Hofierka and J. Klimeš, Electronic Structure **3**, 034010 (2021)
doi: 10.1088/2516-1075/ac25d6

P9 *Using non-covalent interactions to test precision of projector-augmented wave data sets*
S. Yourdkhani and J. Klimeš, Journal of Chemical Theory and Computation **19**, 8871 (2023)
doi: 10.1021/acs.jctc.3c00930