REPORT SERIES IN AEROSOL SCIENCE
N:O 89 (2007)

QUANTUM CHEMICAL STUDIES ON TROPOSPHERIC NUCLEATION MECHANISMS INVOLVING SULFURIC ACID

THEO KURTÉN

Division of Atmospheric Sciences
Department of Physical Sciences
Faculty of Science
University of Helsinki
Helsinki, Finland

Academic dissertation

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public criticism in auditorium A110, A. I. Virtasen aukio 1, on November 2nd, 2007, at 12 o'clock noon.

Helsinki 2007

ISBN 978-952-5027-82-2 (printed version)
ISSN 0784-3496
Helsinki 2007
Yliopistopaino
ISBN 978-952-5027-83-9 (ethesis version)
http://ethesis.helsinki.fi
Helsinki 2007
Helsingin yliopiston verkkojulkaisut

Acknowledgements

The work of this thesis has been carried out at the Department of Physical Sciences in the University of Helsinki. I thank Prof. Juhani Keinonen, the head of the department, for the facilities that allowed me to carry out my work.

I thank Prof. Markku Kulmala, the head of the Division of Atmospheric Sciences and one of the supervisors of this thesis, for his support, interest and valuable advice.

Dr. Hanna Vehkamäki, the leader of the simulation subgroup at the Division of Atmospheric Sciences and the other supervisor of this thesis, has been a constant source of advice, support and mentoring. I am extremely grateful for her guidance and friendship.

I thank Prof. Kari E. J. Lehtinen from the University of Kuopio and Prof. Martti Puska from the Helsinki University of Technology for their efforts in reviewing this thesis, and for their helpful and constructive comments.

Dr. Johanna Blomqvist and Dr. Nino Runeberg at CSC deserve special thanks for their help and patience in answering countless technical questions.

My sincere thanks go to my co-authors at the University of Helsinki: Dr. Boris Bonn, Ms. Martta Salonen, Dr. Markku Sundberg, and Dr. Leena Torpo.

I thank Dr. Madis Noppel from the University of Tartu, Prof. Kari Laasonen at the University of Oulu and Dr. Veli-Matti Kerminen at the Finnish Meteorological Institute for their assistance and co-operation.

The other members of the simulations group are also gratefully acknowledged for their help, friendship and co-operation: Dr. Ismael Kenneth Ortega, Dr. Ismo Napari, Dr. Antti Lauri, Dr. Joonas Merikanto, Mr. Jan Julin and Mr. Ville Loukonen. I thank Mrs. Anca Hienola for help with proofreading, and all my co-workers at the Division of Atmospheric Sciences for their friendship and for our relaxed yet motivating work environment.

I have had many excellent teachers - some of whom are already listed above -, but I would especially like to thank four people who have, at different stages of my education, instilled in me a love of science: Prof. Lauri Halonen, Mr. Juhani Kaila, Dr. Elina Näsäkkälä and Sr. Clare-Ann Litteken.

I thank my parents and my brother for their caring and support, and my wife Kristiina for her love and companionship.

Abstract

Nucleation is the first step of the process by which gas molecules in the atmosphere condense to form liquid or solid particles. Despite the importance of atmospheric new-particle formation for both climate and health-related issues, little information exists on its precise molecular-level mechanisms. In this thesis, potential nucleation mechanisms involving sulfuric acid together with either water and ammonia or reactive biogenic molecules are studied using quantum chemical methods. Quantum chemistry calculations are based on the numerical solution of Schrödinger's equation for a system of atoms and electrons subject to various sets of approximations, the precise details of which give rise to a large number of model chemistries.

A comparison of several different model chemistries indicates that the computational method must be chosen with care if accurate results for sulfuric acid - water - ammonia clusters are desired. Specifically, binding energies are incorrectly predicted by some popular density functionals, and vibrational anharmonicity must be accounted for if quantitatively reliable formation free energies are desired. The calculations reported in this thesis show that a combination of different high-level energy corrections and advanced thermochemical analysis can quantitatively replicate experimental results concerning the hydration of sulfuric acid.

The role of ammonia in sulfuric acid - water nucleation was revealed by a series of calculations on molecular clusters of increasing size with respect to all three co-ordinates; sulfuric acid, water and ammonia. As indicated by experimental measurements, ammonia significantly assists the growth of clusters in the sulfuric acid - co-ordinate. The calculations presented in this thesis predict that in atmospheric conditions, this effect becomes important as the number of acid molecules increases from two to three. On the other hand, small molecular clusters are unlikely to contain more than one ammonia molecule per sulfuric acid. This implies that the average NH_3:H_2SO_4 mole ratio of small molecular clusters in atmospheric conditions is likely to be between 1:3 and 1:1.

Calculations on charged clusters confirm the experimental result that the HSO_4^- ion is much more strongly hydrated than neutral sulfuric acid. Preliminary calculations on $HSO_4^-\bullet NH_3$ clusters indicate that ammonia is likely to play at most a minor role in ion-induced nucleation in the sulfuric acid - water system.

Calculations of thermodynamic and kinetic parameters for the reaction of stabilized Criegee Intermediates with sulfuric acid demonstrate that quantum chemistry is a powerful tool for investigating chemically complicated nucleation mechanisms. The calculations indicate that if the biogenic Criegee Intermediates have sufficiently long lifetimes in atmospheric conditions, the studied reaction may be an important source of nucleation precursors.

Contents

List of publications

This thesis consists of an introductory review, followed by six research articles. The papers are reproduced with the kind permission of the journals concerned.

I T. Kurtén, M. R. Sundberg, H. Vehkamäki, M. Noppel, J. Blomqvist and M. Kulmala: An *ab initio* and density functional theory reinvestigation of gas-phase sulfuric acid monohydrate and ammonium hydrogensulfate, *Journal of Physical Chemistry A*, Vol. **110**, 7178-7188, 2006.

II T. Kurtén, L. Torpo, C.-G. Ding, H. Vehkamäki, M. R. Sundberg, K. Laasonen and M. Kulmala: A density functional study on water-sulfuric-acid-ammonia clusters and implications for atmospheric cluster formation, *Journal of Geophysical Research*, Vol. **112**, D04210 (7 pages), 2007.

III T. Kurtén, M. Noppel, H. Vehkamäki, M. Salonen and M. Kulmala: Quantum chemical studies of hydrate formation of H_2SO_4 and HSO_4^-, *Boreal Environment Research*, Vol **12**, 431-453, 2007.

IV T. Kurtén, L. Torpo, M. R. Sundberg, V.-M. Kerminen, H. Vehkamäki and M. Kulmala: Estimation of the NH_3:H_2SO_4 ratio of nucleating clusters in atmospheric conditions using quantum chemical methods, *Atmospheric Chemistry and Physics*, Vol. **7**, 2765-2773, 2007.

V T. Kurtén, B. Bonn, H. Vehkamäki and M. Kulmala: A computational study of the reaction between biogenic stabilized Criegee intermediates and sulfuric acid, *Journal of Physical Chemistry A*, Vol **111**, 3394-3401, 2007.

VI L. Torpo, T. Kurtén, H. Vehkamäki, K. Laasonen, M. R. Sundberg and M. Kulmala: The significant role of ammonia in atmospheric nanoclusters, *Journal of Physical Chemistry A*, *in press*, 2007.

1. Introduction

From the reactions of nanometer-sized molecular clusters on picosecond timescales to the evolution of the whole Earth's climate over hundreds of millions of years, atmospheric science spans roughly sixteen orders of magnitude in space and twenty-eight orders of magnitude in time. Scientific research on the tiniest end of the spatial and temporal spectrum can have profound effects on the largest scale. For example, the hypothesis [Svensmark 1998] that galactic cosmic rays may influence cloud formation via ion-induced nucleation could, if true, have significant implications for climate variation on the geological time scale. The implications for current climate change would probably remain minor, as there is little evidence for any trend in cosmic rays since 1950 [Benestad 2005]. Incidentally, it would also extend the spatial scale of atmospherically relevant phenomena by a further fourteen orders of magnitude to encompass the motion of our solar system through the entire galactic disc.

Nucleation - the formation of aerosol particles from gas-phase molecules - is one of the most important and challenging research areas within the field of atmospheric science [Kulmala 2003]. On a local scale, aerosol particles can affect human health by causing respiratory and circulatory diseases [Brunekreef and Holgate 2002, von Klot *et al*. 2005, Pope and Dockery 2006], and also have an impact on visibility and weather. It should be noted that aerosols and health care are not related only through the adverse effects of air pollution; for example the design of asthma inhalers also requires information on aerosol formation and behavior. A major part of the particles responsible for air pollution phenomena are thought to be produced by primary particle sources (*e.g.* incomplete combustion) rather than new-particle formation, but in remote continental regions nucleation may well be the main source of atmospheric aerosol [Spracklen *et al*. 2006]. Also, some recent studies [see *e.g.* Brunekreef and Holgate 2002] indicate that a large part of the adverse health effects of particulate pollution may be caused by the smallest (nanometer-scale) particles, even though their contribution to the aerosol mass indicators currently used in legislation (*e.g.* PM_{10}) is negligibly small. As the focus of aerosol research and legislative measures moves toward smaller particle classes, the importance and relevance of new-particle formation studies will continue to grow.

On a global scale, aerosols affect the climate via three different processes. First, aerosols may scatter and absorb both sunlight and (in the case of *e.g.* soot particles) infrared radiation emitted by the Earth. The radiative forcing associated with this so-called "direct effect" can be either negative (cooling) or positive (warming) depending on the aerosol type. The Intergovernmental Panel on Climate Change [IPCC 2007] estimates that the total net direct effect of all anthropogenic emissions of aerosols and their precursor vapors is cooling. Second, aerosol particles can promote the formation of cloud condensation nuclei (CCN), leading to an increase in the reflectivity of clouds, which increases the Earth's albedo and thus exerts a negative radiative forcing. This phenomenon is called the first indirect, or cloud-albedo, effect. Third, the presence of aerosol particles increases the number of CCN but decreases their size, which tends to increase average cloud lifetimes. This second indirect, or cloud-lifetime, effect is presumed to be cooling in nature, but is extremely difficult to quantify. The effect of

nucleation on CCN concentrations and size distributions is likely to be important in most regions of the atmosphere, and modeling these effects is one of the main challenges in the development of climate models [Kulmala 2003, Pirjola *et al.* 2004, Stier *et al.* 2005].

Humans affect the concentrations and properties of atmospheric aerosol in at least three ways. The most obvious effect is the direct emission of aerosol particles into the atmosphere via various anthropogenic primary sources, such as combustion or road dust thrown up into the air by vehicles. Combustion along with industrial and other processes also emit vapors that can nucleate, or react to form nucleating substances, in the atmosphere. The most prominent example is sulfur dioxide (SO_2), which reacts with oxidisers in the atmosphere (mainly OH radicals) to form sulfuric acid (H_2SO_4), one of the most important species believed to participate in atmospheric new-particle formation. The most subtle and poorly understood anthropogenic influence on aerosol distributions is the effect of land-use and land-use changes. For example, aerosol particles formed with the help of vapors emitted by boreal forests may play a very significant role in the local radiative balance [Kurtén *et al.* 2003, Spracklen *et al.* 2006, Tunved *et al.* 2006]. If the forest area (or even just the characteristics of the forest) changes, the aerosol concentrations also change, exerting a radiative forcing. Unfortunately, the influence of different ecosystems and terrain types on aerosol and aerosol precursor emissions are very poorly understood at the moment.

Theoretical studies of nucleation processes have been conducted since the 19th century. Macroscopic theories of nucleation can be said to originate with the work of Laplace, lord Kelvin and Gibbs, who laid the formations for the thermodynamic machinery needed to compute *e.g.* the free energies of cluster formation (see Vehkamäki [2006] for a short historical review). Classical nucleation theory (CNT), derived by Becker, Döring [Becker and Döring 1935] and Zeldovich [1943], and extensively modified by later researchers, is based on the computation of cluster properties using macroscopically determined parameters such as the surface tension of bulk liquids. For clusters composed only of a few molecules, the use of bulk properties is not well justified. The advent of computer simulations in the 1950s made it possible to compute cluster properties using various molecular-scale methods such as Molecular Dynamics (MD) or Monte-Carlo (MC) modeling. In terms of the spatial dimensions, the focus of theoretical nucleation studies thus moved from a scale of milli- or micrometers (corresponding to the smallest measurable bulk samples) to nanometer-scale molecular clusters. In recent decades, the massive increase of computer power has made it possible to perform atmospherically relevant (though limited) nucleation studies at the even more fundamental level of quantum chemistry, where cluster properties are computed from the interactions of individual electrons. However, the predictions of quantum chemical studies concerning the formation and properties of sulfuric acid – water – ammonia clusters in atmospheric conditions have so far been in contradiction with the experimental results.

The main objectives of this thesis are:

1. To explain the discrepancies between experiments and previous quantum chemical studies concerning the hydration of sulfuric acid.

2. To explain the role of ammonia in tropospheric sulfuric acid - water nucleation, for example in terms of the most probable NH_3:H_2SO_4 mole ratios of nucleating clusters.

3. To investigate and compare neutral and ion-induced nucleation in the sulfuric acid - water system with respect to *e.g.* water affinity and the role of ammonia.

4. To investigate the possible relevance of reactions between sulfuric acid and oxidized reactive biogenic molecules for atmospheric nucleation.

2. Atmospheric nucleation: observations and proposed mechanisms

Atmospheric nucleation has been studied since the 19th century, when John Aitken [1881] first reported evidence for new-particle formation in certain suitable atmospheric conditions, stating for example that "sulphur when burned has been shown to be an intensely active fog-producer". However, until recently it was thought that gas-to-particle nucleation is a rare phenomenon in the atmosphere, and that the vast majority of all atmospheric particles are emitted from sources rather than formed *in situ*. The advent of nanometer-scale size distribution measurements changed this picture completely: new-particle formation has now been observed almost everywhere in the atmosphere, from the most polluted city centers to the ultra-pure air of Antarctica [Kulmala *et al.* 2004a]. In the remote marine boundary layer, nucleation is a rare phenomenon [Heintzenberg *et al.* 2004], but in the continental boundary layer it seems to occur quite frequently both in rural and urban areas [Kulmala *et al.* 2004a]. According to Spracklen *et al.* [2006], nucleation may account for around 30% of all aerosol particles over the continents. The clearest example of new-particle formation is illustrated by so-called “banana plots”, where the axes represent size (on a logarithmic scale), time and concentration (usually expressed on a color scale). Figure 1 shows an exceptionally clear new-particle formation event measured at the Hyytiälä field station in Finland. The event, starting around noon, is visible as the formation of a new mode which then grows at an almost constant rate until late at night. Similar new-particle formation events have also been observed for example at Mace Head in Ireland [O’Dowd *et al.* 2002a] and several other locations around the world [Kulmala *et al*. 2004a].

Despite the large and growing number of particle formation observations, the molecular-level mechanisms of atmospheric nucleation are still unknown. One reason for this is that the initial steps of particle formation are beyond the reach of current aerosol instrumentation. State-of-the art instruments can detect neutral clusters of 2.5 nm diameter or above [McMurry 2000, Kulmala *et al.* 2005], corresponding to some hundreds of molecules. Chemical composition measurements are possible only for particles larger than about 10 nm [Nash *et al.* 2006], which already contain on the order of 10^5 molecules, the vast majority of which originate from various growth processes rather than nucleation itself. Experimental investigations into the molecular mechanisms of nucleation are thus restricted to rather indirect approaches, for example measuring correlations of particle growth or formation rates with the concentrations of various trace gases (see Sihto *et al.* [2006] or Riipinen *et al.* [2007] for an application to sulfuric acid)

or other environmental variables. Since it is likely that nucleation at different sites involves different mechanisms, the observed correlations are also likely to vary from place to place. A further problem is that few condensable trace gases have been measured continuously for long periods. For example, continuous long-term data on ammonia concentrations are very rare.

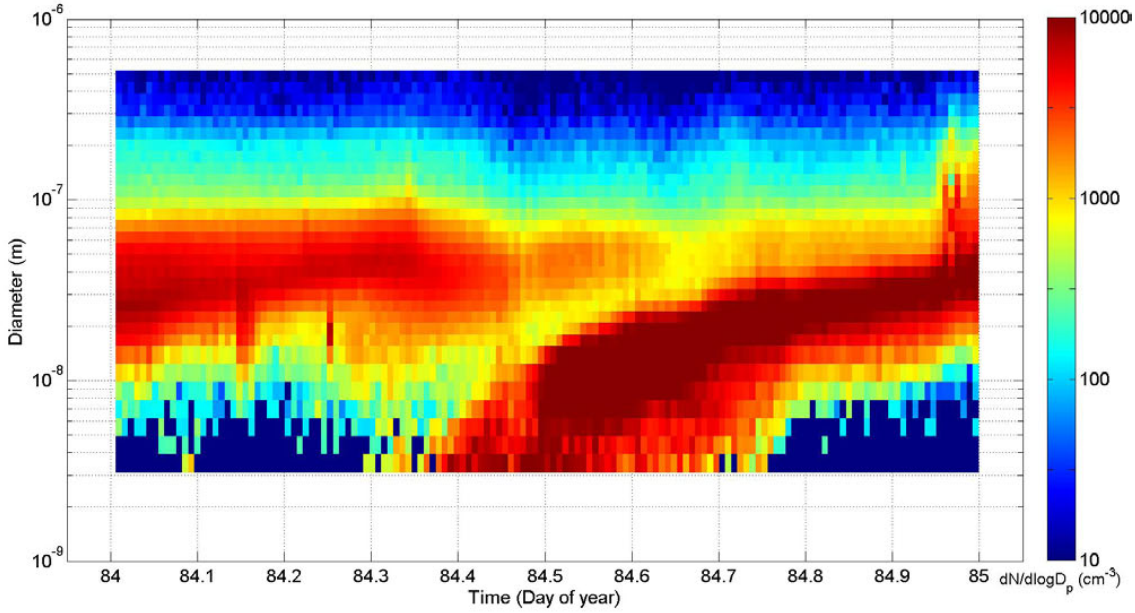

Figure 1. Size distribution plot (courtesy of Ms. Ilona Riipinen) displaying a nucleation event measured at the SMEAR II station in Hyytiälä, Finland on the 25th of March 2003. The *x*-axis is the time in decimal days, running from midnight to midnight. The *y*-axis is the particle diameter, and the color scale indicates the concentration of particles in each diameter range (more precisely: the logarithmic number size distribution function). For details on the SMEAR II measurement station see Hari and Kulmala [2005].

The results of experimental atmospheric new-particle formation studies have been summarized by Kulmala *et al.* [2004a]. As expected, large variation exists between different measurement sites. Nevertheless, particle formation rates have been observed to correlate with the concentration of sulfuric acid in a surprisingly large variety of conditions. Sunlight also seems to be a very important (though not absolutely necessary, see *e.g.* Wiedensohler *et al.* [1997]) requirement for the onset of particle formation, though whether this is due to photochemical or meteorological reasons is still unclear. Based on the observations, the main candidates for nucleation in the troposphere are binary sulfuric acid – water and ternary sulfuric acid – water – ammonia systems [Kulmala *et al.* 2000, Napari *et al.* 2002, Kulmala 2003]. Binary nucleation is likely to take place mainly in industrial plumes, where sulfuric acid concentrations are high enough, or in the free troposphere, where the temperature and background aerosol concentration are low enough to allow significant sulfuric acid – water cluster formation. Ternary nucleation, on the other hand, is expected to be thermodynamically possible in

almost all atmospheric conditions [Anttila *et al.* 2005], and could be the dominant nucleation mechanism in the continental boundary layer. Jung *et al.* [2007] recently reported that for a dataset collected in the Pittsburgh region, a ternary nucleation model predicted the presence or absence of nucleation events almost perfectly, and by far outperformed models based on other nucleation mechanisms. The global importance of ternary nucleation is, however, still unknown, and for example Yu [2006] has claimed that its contribution to boundary-layer new-particle formation is negligible. Ion-induced nucleation of sulfuric acid – water clusters or other compounds [Yu and Turco 2000, Lee *et al.* 2003, Lovejoy *et al.* 2004] is also thermodynamically possible almost everywhere in the atmosphere, and may be an important source of particles in the upper troposphere or lower stratosphere. However, recent modeling and experimental studies seem to indicate that the contribution of these mechanisms to the observed new-particle formation in the lower troposphere is relatively small [Lovejoy *et al.* 2004, Laakso *et al.* 2004, Kulmala *et al.* 2004a, Iida *et al.* 2006]. In coastal areas, iodine – containing molecules are believed to play a role in nucleation [O'Dowd *et al.* 2002a], either with or without sulfuric acid. Various organic compounds are known to participate in particle growth [O'Dowd *et al.* 2002b, Kulmala *et al.* 2004b], but it is still unclear whether they participate in the actual nucleation mechanisms or not. Some recent studies [Zhang *et al.* 2004, Surratt *et al.* 2007, Nadykto and Yu 2007, **Paper V**] have suggested that the clustering or reactions of sulfuric acid with various organic species might be an important nucleation mechanism in the atmosphere. It should further be noted that recent experiments [Berndt *et al.* 2005, 2006] indicate that the mixture of sulfuric acid and some intermediate products of the photochemical SO_2 oxidation chain nucleate much more effectively than sulfuric acid on its own.

From a macroscopic point of view, gas-to-particle nucleation can occur when the vapor pressure of some compound in the gas phase exceeds its saturation vapor pressure in the ambient conditions (which include not only temperature and total pressure but also the concentrations or activities of all other compounds present). In microscopic terms, this implies that the binding energies of the nascent molecular clusters must be great enough to compensate for the negative entropy of the cluster formation. (Here the term "molecule" is used loosely as a synonym for the basic unit of any sort of chemical compound, be it a true covalently bound molecule, air ion or radical.)

The molecular perspective also demonstrates why atmospheric nucleation is almost always multicomponent: if some molecule was extremely strongly bound to other molecules of the same type, it would have scant chances to make it into the gas phase in the first place. The exceptions to the rule are compounds formed in the gas phase through chemical reactions of precursor compounds such as some organic substances. These could conceivably contribute to particle formation events via one-component mechanisms. Thus, the most probable nucleation mechanisms must involve two or more types of molecules which are attracted to each other much more strongly than to themselves. General examples of such compound pairs can be deduced from elementary chemistry: strong acids and strong bases, ions with opposite charge, or appropriate hydrogen bond donor and acceptor pairs are the most prominent types. Indeed, all of

these three cases are represented in the proposed atmospheric nucleation mechanisms, even within the sulfuric acid – water – ammonia system alone.

Sulfuric acid - water vapor mixtures, the main systems of interest in this thesis, are prime examples of multicomponent nucleation. Water alone can not nucleate due to its high saturation vapor pressure, while the concentration of sulfuric acid in the atmosphere is too low for significant numbers of particles to form. The strong attraction of these molecules for each other means that the binary mixture can nucleate even if the partial pressures of the individual compounds are many orders of magnitude too low for them to nucleate on their own. However, it turns out that in most atmospheric conditions this is not enough - some third component is needed to explain observed nucleation events in the atmosphere [Weber *et al.* 1999, Kulmala *et al.* 2003, 2004a]. Laboratory measurements [Ball *et al.* 1999], supported by theoretical calculations [see *e.g.* Vehkamäki *et al.* 2004, Nadykto and Yu 2007 and **Papers I-II**, **IV** and **VI**] and some field observations [Jung *et al.* 2007] show that ammonia is one plausible candidate, as it strongly enhances the binding of sulfuric acid molecules to each other. Electrical charge is another alternative, as demonstrated both by laboratory studies and calculations [Froyd and Lovejoy 2003, **Paper III**]. As sulfuric acid is a good proton donor and a very poor proton acceptor, it comes as no surprise that ion-induced sulfuric acid - water nucleation is observed primarily for negative ions [see *e.g.* Lovejoy *et al.* 2004 and Laakso *et al.* 2007], with positive ions playing little or no role. Some speculations have been made [Lovejoy *et al.* 2004, Yu 2006] about a possible co-operative effect between these two "nucleation enhancers", ammonia and negative charge. However, the preliminary calculations in **Paper III** indicate that ammonia is not likely to significantly promote nucleation of HSO_4^- clusters, as the two compounds are very weakly attracted to each other, if at all. If this is true also for somewhat larger clusters, ammonia and negative charge could be seen as similar but competing "enhancers" in the sulfuric acid - water system. Both act to strengthen the binding of sulfuric acid and/or water molecules to each other, but any given small molecular cluster is likely to contain only one or the other, not both.

3. Theoretical methods used to study nucleation

The textbook definition of nucleation is that it is the first step of a first-order phase transition. The term "first order" implies that some parameter of the system (*e.g.* the density) undergoes a discontinuous change in the transition. This in turn implies that in some conditions, there exists an energy barrier for the phase change. For atmospheric nucleation processes, the barrier is commonly defined in terms of the free energy, or the minimum amount of work required to form a certain cluster. Depending on the boundary conditions assumed for the system (*i.e.* which parameters are kept constant), the relevant free energy may be either the Gibbs or Helmholtz free energies, or the grand potential. In most atmospheric nucleation studies, the total number of particles, pressure and temperature (but not the total volume) are assumed to remain constant during the cluster formation, in which case the appropriate free energy is the Gibbs free energy, G. However, molecular dynamics simulations are often performed in conditions of constant particle number, volume and temperature, in which case the Helmholtz free energy, F, is

the most appropriate potential. It should be noted that in the conditions relevant for atmospheric nucleation, the different free energies will for practical purposes usually be equal [Vehkamäki 2006].

A note on energy barriers.

In physical chemistry or molecular physics, it is common to separate two different types of energy barriers. Thermodynamic barriers are related to the minimum points on the potential energy hypersurface of the atomic nuclei, while kinetic barriers (known as activation energies in chemical kinetics) are related to first-order saddle points. *An almost dogmatic mantra in elementary chemistry is that "kinetics can not be determined from thermodynamics". For example, spontaneous self-combustion of human beings is thermodynamically feasible, but (fortunately) has a quite high kinetic barrier.*

The free energy surfaces encountered in nucleation studies can be viewed as a small subset of the potential energy surfaces studied in molecular physics. Instead of individual nuclear co-ordinates, their co-ordinate axes correspond to the number of molecules in a cluster. The arrangement of atoms in the clusters is not accounted for explicitly, but is instead implicitly included via the method by which the free energies are calculated.

Perhaps due to this difference, the rigorous separation of kinetic and thermodynamics is not often explicitly acknowledged in nucleation studies. For most commonly studied nucleating systems (*e.g.* argon or water) this is not a problem, as the nucleation processes do not usually involve the breaking of chemical bonds, and hence have very low kinetic barriers, if any. "Kinetics" can then be computed from collision rates of ideal gases. However, as shown in **Paper V**, nucleation mechanisms involving *e.g.* sulfuric acid and organic molecules may well involve chemical reactions, which may have nonzero barriers (though in the particular case studied the barrier proved to be close to zero). In such cases, ignorance of the difference between thermodynamics and kinetics may well lead to significant conceptual problems and modeling errors.

The free energy barrier (and also the free energy surface around the barrier) is the central concept for nucleation studies. The nucleation rate (number of new particles formed per unit of time and volume), for example, depends exponentially on the barrier height

$$J \propto e^{-\frac{\Delta\varphi^*}{k_B T}}, \tag{3.1}$$

where J is the nucleation rate and $\Delta\varphi^*$ is the free energy of formation of the so-called critical cluster from its molecular constituents. The critical cluster corresponds to the maximum point of the free energy versus cluster size curve, and it is thus the smallest cluster for which growth is more likely than decay. The size of the critical cluster (*i.e.* the location of the maximum of the free energy curve) depends on the vapor activities as demonstrated by the first nucleation theorem

$$\left(\frac{\partial \ln J}{\partial A_{i,g}}\right)_{A_{j\neq i,g},T} \approx \Delta n_i^*, \qquad (3.2)$$

where Δn_i^* is the number of molecules of type i in the critical cluster and $A_{i,g}$ is the gas-phase activity of the corresponding monomer vapor. It should be noted that though the nucleation theorems are often derived using classical nucleation theory, they are independent of its assumptions, and can be derived directly from statistical mechanics [Vehkamäki 2006].

The central question in applied theoretical nucleation studies is thus: how can we obtain the free energy change for the formation of clusters from monomers? The approaches to this problem can be classified into three groups depending on the scale at which the interactions between the nucleating compounds are treated.

1. Classical nucleation theory (CNT), which treats intermolecular interactions in terms of average bulk properties. In CNT, the free energy changes are obtained from macroscopically measured parameters for the participating compounds. The most important parameters are surface tensions, activities, saturation vapor pressures and molecular volumes.

2. Methods based on classical interaction potentials, in which potential models (also called force fields) are used to describe the interactions between molecules in terms of their functional groups, without explicitly accounting for electrons or quantum mechanical effects. Using classical interaction potentials, free energy changes can be computed via three different approaches: classical density functional theory, Molecular Dynamics (MD) and Monte Carlo (MC) simulations.

3. Quantum chemistry simulations, in which the interactions between molecules are computed by solving the Schrödinger equation (or related sets of equations) for the motion of electrons, and the free energy changes are computed by applying quite simple statistical mechanics, usually to minimum energy configurations of the atomic nuclei.

3.1. Classical nucleation theory

Classical nucleation theory [Becker and Döring 1935, Zeldovich 1943] assumes that molecular clusters can be modeled as spherical droplets with a constant density equal to the bulk liquid density, and a surface tension equal to the bulk liquid surface tension. For reviews of CNT see Abraham [1974], Kashchiev [2000] and Vehkamäki [2006]. The central approximations of CNT are related to the surface tension. In order to simultaneously fulfill the laws of thermodynamics and be able to use bulk surface tension values, we have to assume that the surface of tension (where the derivative of the surface tension with respect to the radius is zero) and the equimolar surface (see equation 3.7 below) of the droplet are identical, and equal to the radius of the droplet. For a one-component droplet, the Gibbs free energy of formation from constituent vapor molecules then becomes

$$\Delta G = A\sigma_{l,v} + n\Delta\mu = 4\pi r^2\sigma_{l,v} - \frac{4}{3v_l}\pi r^3 k_B T \ln(S), \quad (3.3)$$

where A is the surface area and r the radius of the droplet, n is the number of molecules in the droplet, $\sigma_{l,v}$ is the (liquid-vapor) surface tension, $\Delta\mu$ is the difference of the chemical potential in the vapor and liquid computed at the vapor pressure, v_l is the molecular volume (computed from the bulk liquid density) and S is the saturation ratio of the vapor. (Strictly speaking, the equation applies only for clusters in equilibrium with the surrounding vapor, but as shown *e.g.* by Vehkamäki [2006], it can be used to a high accuracy also for non-equilibrium clusters.) The term proportional to r^2 corresponds to the work required to create the liquid-vapor interface, and is responsible for the free energy barrier mentioned above. The term proportional to r^3 corresponds to the interactions between molecules in the droplet, which for $S > 1$ and sufficiently large r lead to the cluster formation being thermodynamically favorable. The maximum point of the ΔG vs r curve corresponds to the critical cluster, for which the formation energy is

$$\Delta G^* = \frac{4}{3}\pi (r^*)^2 \sigma_{l,v}, \quad (3.4)$$

where the superscript "*" denotes the critical cluster. For multicomponent systems, the analog of equation (3.3) still applies, though it should be noted that the surface tension now depends on the composition

$$\Delta G = A\sigma_{l,v} + \sum_i n_i \Delta\mu_i, \quad (3.5)$$

where the index i corresponds to the different components (*e.g.* molecules of a different type) in the cluster. Connecting the chemical potential difference to measurable variables is also complicated by so-called surface excess number corrections, which arise from the above-mentioned assumption that the surface of tension is equal to the equimolar surface of the cluster. This means that the molecular numbers n_i in the expression above correspond to droplet numbers, which are the sum of liquid (also called core) and surface numbers ($n_{i,l}$ and $n_{i,s}$, respectively)

$$n_i = n_{i,l} + n_{i,s}, \quad (3.6)$$

where the equimolarity condition requires that

$$\sum_i n_{i,s} v_{i,l} = 0, \quad (3.7)$$

which in turn means that the surface numbers may be negative for some components. For surface active mixtures, the liquid and surface numbers may be of the same order of magnitude, and CNT predictions may fail dramatically. The simplest expression for the multicomponent formation free energy in terms of measurable variables is

$$\Delta G = A\sigma_{l,v}(x_{i,l}) - k_B T \sum_i \ln(S_i) n_i = A\sigma_{l,v}(x_{i,l}) - k_B T \sum_i \ln\left(\frac{A_{i,g}}{A_{i,l}(x_{i,l})}\right) n_i, \quad (3.8)$$

where the saturation ratios have been expressed in terms of the gas- and liquid phase activities $A_{i,g}$ and $A_{i,l}(x_{i,l})$, $x_{i,l}$ indicates the mole fraction of component i in the liquid phase, and the dependence of the surface tension on the composition has been explicitly indicated. The critical cluster size has to be solved iteratively, as the right-hand-side term contains explicitly the total molecular numbers in the droplet, n_i, but the mole fractions (and thus *e.g.* the surface tension and the activities) depend on the liquid numbers $n_{i,l}$. The location of the critical cluster is then obtained by setting $\partial\Delta G / \partial n_{i,l} = 0$ for each

component. However, equation (3.4) for the formation free energy of the critical cluster still applies.

3.2. Methods based on classical interaction potentials

The central concept of molecular simulations (as opposed to macroscopic theories on one hand, and quantum chemistry on the other hand) is the formulation of molecular interactions in terms of a force field. Numerous different approaches to force fields have been presented over the decades, but most common force fields in use today share some general characteristics with respect to the functional forms used to model the different interactions [Leach 2001, Jensen 2007].

Nonpolar interactions are typically accounted for by Lennard-Jones terms

$$V(\mathrm{A},\mathrm{B})_{LJ} = 4\varepsilon_{A,B}\left[\left(\frac{\sigma_{A,B}}{r_{A,B}}\right)^{12} - \left(\frac{\sigma_{A,B}}{r_{A,B}}\right)^{6}\right], \tag{3.9}$$

where A and B are two functional groups (which can be *e.g.* individual noble gas atoms, atoms bound to a molecule, or entire groups of atoms within a larger molecule), $\varepsilon_{A,B}$ and $\sigma_{A,B}$ are constants, and $r_{A,B}$ is the distance between the groups. For example argon clusters, which have traditionally been the mainstay of theoretical nucleation simulations [see *e.g.* Lauri *et al.* 2006a, 2006b], can be treated using only Lennard-Jones terms, which makes the simulations very efficient and fast. Interactions between partial charges on different molecules are treated using Coulomb terms

$$V(\mathrm{A},\mathrm{B})_{COULOMB} = \frac{Q_A Q_B}{4\pi\varepsilon_{dielec} r_{A,B}}, \tag{3.10}$$

where Q_A and Q_B are the partial charges of functional groups A and B and ε_{dielec} is the dielectric constant of the medium (usually but not always set to be vacuum). Partial charges are often placed at the atomic nuclei, but additional, off-center charges are also frequently used to improve the performance of the models. For example, many popular potentials for water molecules contain four or five partial charges. Traditionally, the partial charges are fixed parameters that do not vary during the simulation. Polarizable models, which allow the charges to vary with the molecular interactions, have been developed (see *e.g.* Guillot and Gissani [2001] for an application to water), but at least in nucleation studies they have not represented a major improvement over the best nonpolarizable models [Merikanto *et al.* 2004].

Intramolecular bonds can either be kept rigid, or treated using harmonic potentials for stretching and bending motions

$$V(\mathrm{A},\mathrm{B})_{BOND} = \frac{k_{A,B,r}}{2}(r_{A,B} - r_0)^2 + \frac{k_{A,B,\vartheta}}{2}(\theta_{A,B} - \theta_0)^2, \tag{3.11}$$

where $k_{A,B,r}$ and $k_{A,B,\theta}$ are constants, $\theta_{A,B}$ is the angle between the two functional groups (if not bonded directly to each other) and r_0 and θ_0 are constants corresponding to the equilibrium bond length and angle, respectively. Torsion potentials can be used to model the interaction of groups further from each other; they are commonly of the form [Leach, 2001]

$$V(\mathrm{A},\mathrm{B})_{TORSION} = \sum_i k_{A,B,i} \cos(\omega_{A,B})^i , \quad (3.12)$$

where $\omega_{A,B}$ is the torsion angle between the groups A and B and the $k_{A,B,i}$ are constants. The total potential for a system of particles is obtained by summing over all the different contributions. The various constants are typically obtained by fitting either to some suitable experimental properties (such as bulk surface tensions or heats of vaporization), or to quantum chemical results, as described in section 4. The major advantage of classical potentials is that they are computationally efficient compared to quantum chemical calculations. This allows the study of larger systems or longer timescales. The major disadvantage, on the other hand, is that classical potentials are rarely transferable - a potential developed for a certain type of molecule in one environment will not be likely to model its interactions correctly in a different environment. Furthermore, most force fields are by construction unable to treat the formation and breaking of chemical bonds.

Classical density functional theory is the computationally most efficient way to calculate cluster formation energies using force fields. In contrast to MD and MC, individual molecules are not simulated as such, but the interaction potential is used to compute average density distributions for atoms or molecules within a cluster. Classical DFT has been successfully applied to surface active systems, and Laaksonen and Napari [2001] have used it to show why CNT fails for these cases. However, classical DFT is practically applicable only for molecular systems with relatively simple structures, and can not directly be applied to modeling *e.g.* sulfuric acid - water nucleation.

The idea of Molecular Dynamics and Monte Carlo simulations (see Frenckel and Smit [2002] for a review) is to generate a large number of distributions for a limited number (typically on the order of 10-10 000) of molecules, and compute the properties of interest by applying advanced statistical methods to this sample set. In MD simulations, the net force acting on each functional group is computed from the force field, and the groups are moved according to the laws of classical mechanics. MD thus provides information on the time evolution of molecular systems, and can be used to study dynamics of processes such as cluster growth and evaporation, at least as long as they do not involve the formation or breaking of chemical bonds. The disadvantage of MD compared to MC is that all the forces must be calculated at each individual step, which is time-consuming. In MC simulations, particles are moved by random steps, and only the total energy (rather than the forces) is computed. The central idea of the MC method (more correctly called Metropolis MC after its original developer [Metropolis *et al.* 1955]) is to compare the potential energy of the system before and after each random step. If the energy is lower after the random step, the new configuration is always accepted. If it is higher, the new configuration is accepted with a probability $e^{-\Delta V/k_B T}$, where ΔV is the potential energy change. This ensures that the distribution of configurations generated by the simulation follow the Boltzmann distribution.

Typically, MD or MC simulations are performed either in conditions of constant particle number, volume and energy (called NVE, or microcanonical ensemble simulations) or constant particle number, volume and temperature (called NVT, or canonical ensemble simulations). Other ensembles such as NTP (corresponding to constant particle number,

temperature and pressure) are also common. NVE is the most natural choice for a MD simulation and NVT for a MC simulation, but several techniques exist to allow sampling from other ensembles (see Leach [2001] for a discussion of ensembles). For example, various types of thermostats may be used in MD runs to keep the temperature constant during the simulation.

In principle, any property of the system can be computed from MD or MC simulations by calculating suitably weighted averages (time averages for MD and ensemble averages for MC) over all simulation steps. For example, partition functions can be computed for a system of *n* indistinguishable particles in the NVT ensemble as [Leach 2001]

$$q_{NVT} = \frac{1}{n!h^{3n}} \iint d\mathbf{p}^n d\mathbf{r}^n e^{\frac{-H_{class}(\mathbf{p}^n, \mathbf{r}^n)}{k_B T}}, \tag{3.13}$$

where H_{class} is the classical Hamiltonian function for the system (in practice equal to the total energy computed as the sum of kinetic and potential energy terms), **p** and **r** are the momenta and locations of the particles and the integration is performed over all 6*n* positions and momenta. This formula is only directly applicable to MD methods, as MC methods do not in general compute particle momenta. Once the partition function is evaluated, thermodynamic properties of interest could then be relatively straightforwardly computed using standard statistical mechanical results, *e.g.* in this case the Helmholtz free energy *F* would be

$$F = -k_B T \ln(q_{NVT}). \tag{3.14}$$

In practice, the partition function is impossible to evaluate as it would require computing every single point in the 6*n*-dimensional phase space of the system. While reasonably accurate estimates can still be computed for mechanical properties such as internal energies and pressures, which depend on the derivatives of the partition function, entropic properties, which depend on the partition function itself, are much more difficult to calculate. The reason for this is that the configurations corresponding to high energies make large contributions to these properties, as illustrated by the explicit expression for the Helmholtz free energy in the NVT ensemble (obtained from equations 3.13 and 3.14 above) [Leach 2001]

$$F = k_B T \ln(\iint d\mathbf{p}^n d\mathbf{r}^n e^{\frac{H_{class}(\mathbf{p}^n, \mathbf{r}^n)}{k_B T}} \rho(\mathbf{p}^n, \mathbf{r}^n)), \tag{3.15}$$

where $\rho(\mathbf{p}^N, \mathbf{r}^N)$ is the probability distribution function, given by

$$\rho(\mathbf{p}^n, \mathbf{r}^n) = \frac{e^{\frac{-H_{class}(\mathbf{p}^n, \mathbf{r}^n)}{k_B T}}}{q_{NVT}}. \tag{3.16}$$

By construction, MD and MC simulations sample the low-energy regions of the phase space, which means that the computed absolute free energies (and other entropic properties) will be inaccurate [Leach 2001].

However, the properties of chemical and physical interest are usually not related to absolute free energies, but to changes in free energies between different systems.

Nucleation simulations, for example, are usually focused on obtaining the free energy changes of cluster formation processes in different conditions. These can be evaluated relatively accurately through various methods, the most common of which are thermodynamic perturbation and thermodynamic integration. Both are based on the concept of a coupling parameter λ, which connects two systems such that λ=0 for one system and λ=1 for the other. By performing several simulations at slightly different values of λ and combining the results using appropriate statistical theories (see Lauri *et al.* [2006b] for an example), the free energy difference between the systems can be calculated. For example, the Helmholtz free energy change for a NVT system can be computed as [Leach 2001, Frenckel and Smit 2002]

$$\Delta F(\lambda: 0 -> 1) = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial H_{class}(\mathbf{p}^n, \mathbf{r}^n; \lambda)}{\partial \lambda} \right\rangle_{\lambda} d\lambda \,, \tag{3.17}$$

where the brackets indicate expectation values and the integral is in practice replaced by a finite summation.

In nucleation studies, the parameter λ is constructed in such a fashion that λ=0 and λ=1 correspond to systems with *n* and *n+1* molecules, respectively. This approach has been applied *e.g.* by Lauri *et al.* [2006a, 2006b] to study the nucleation of Lennard-Jones argon. In these simulations, equation 3.17 was used sequentially to construct the free energy surface for nucleation by adding atoms to the cluster one at a time. The number of points needed for the evaluation of the integral in equation 3.17 is in these cases quite small, typically on the order of 2-10. Free energy changes for cluster formation can also be computed from growth and decay probabilities, without use of a coupling parameters, see *e.g.* Merikanto *et al.* [2004] for an example on water nucleation.

3.3. Quantum chemistry simulations

Quantum chemistry can be defined as computing the properties of some system of atoms and electrons by explicitly evaluating the interactions between the electrons using quantum mechanics. This is done by numerically solving the Schrödinger wave equation or some closely related set of equations. In principle, quantum chemistry could be used to calculate numerically exact values for any parameters of interest. However, as computational resources are inevitably limited with respect to both time and memory, a number of more or less severe approximations have to be made in going from the Schrödinger equation to, say, Gibbs free energies of reactions. These approximations can be made in several of different ways, giving rise to a multitude of computational approaches, also known as "model chemistries".

The starting point of all quantum chemistry methods is the Schrödinger equation for a system of *K* atomic nuclei and *N* electrons

$$\hat{H}\Psi(\mathbf{r}_1 ... \mathbf{r}_{N+K}) = E\Psi(\mathbf{r}_1 ... \mathbf{r}_{N+K}) \,, \tag{3.18}$$

where $\hat{H}$ is the Hamiltonian operator of the system, E is its energy, $\Psi(\mathbf{r}_1..\mathbf{r}_{N+K})$ is its many-body wavefunction and $\mathbf{r}_i$ is the vector coordinate of the i:th particle. If relativistic corrections are ignored, the Hamiltonian operator can be written as [Jensen 2007]

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn}\,, \tag{3.19}$$

where the terms (from left to right) correspond to the nuclear and electronic kinetic energy operators, and the three types of potential energy operators, corresponding to nuclei-electron, electron-electron and nuclei-nuclei interactions, respectively.

In almost all studies on polyatomic systems, the Born-Oppenheimer approximation is used to split the wavefunction into electronic and nuclear components

$$\Psi(\mathbf{r}_{el}, \mathbf{r}_{nuc}) = \chi(\mathbf{r}_{nuc})\Phi(\mathbf{r}_{el})\,, \tag{3.20}$$

where $\chi(\mathbf{r}_{nuc})$ and $\Phi(\mathbf{r}_{el})$ are the nuclear and electronic wavefunctions, respectively. The justification for the Born-Oppenheimer approximation is that the atomic nuclei are much heavier (by a factor of *ca.* 1830 in the case of protons, and more for all other nuclei) than the electrons, and therefore move much slower. From the point-of-view of the electrons, the nuclei can thus be treated as classical point masses. The electronic wavefunction depends on the nuclear positions, but only parametrically - the electronic wavefunction can be solved separately for each given nuclear configuration. The (classical) forces acting on the nuclei can be calculated from the Hellman-Feynman theory once the electronic wavefunction has been computed

$$F_{A,x} = \frac{\partial E}{\partial x_A} = \int \Phi^*(\mathbf{r}_{el})\nabla x_A \hat{H}\Phi(\mathbf{r}_{el})d\tau\,, \tag{3.21}$$

where $F_{A,x}$ is the force acting on nuclei A in the direction x, the superscript "*" indicates a complex conjugate, and the integration is performed over all variables. In principle, quantum chemistry simulations can then proceed in the same fashion as the force-field based molecular dynamics described in section 3.2. In practice, computational considerations limit the dynamics to plain geometry optimizations, which employ a variety of algorithms to reach a minimum energy configuration for the atomic nuclei starting from some initial input geometry, using as few steps as possible. Usually this involves construction of a set of redundant internal co-ordinates, which are more efficient for optimization than Cartesian co-ordinates (see *e.g.* Jensen [2007] or Leach [2001] for a review of optimization routines).

There are two main families of methods by which electronic energies accurate enough for thermochemical predictions can be obtained. The more traditional wavefunction-based methods first iteratively solve the electronic wavefunction $\Phi(\mathbf{r}_{el})$ for a system of electrons which move in each other's static average potential. This so-called Hartee-Fock wavefunction is then improved upon by various excitation operators. These act to include the effects of electron-electron correlation, which plays an important role in chemical interactions. The other approach is based on density functional theory (DFT) developed by Kohn and Sham [1965], in which the Schrödinger equation is reformulated in terms of the electron density, and exchange and correlation interactions are included via an exchange-correlation functional (see page 26 for details).

3.3.1. Wavefunction-based methods

In wavefunction-based methods, two major approximations to the form of the electronic wavefunction have to be made before the Schrödinger equation can be solved. (Analogous approximations are also made for DFT methods as will be seen below.) The first step is to expand the ($3N$-dimensional) electronic wavefunction in terms of a set of one-electron wavefunctions. The most common form for this expansion is the Slater determinant, which is the simplest combination of one-electron orbitals that fulfills the requirement of the Pauli exclusion principle that the electronic wavefunction is antisymmetric with respect to the exchange of two electrons. In the general case, the expansion can be written as

$$\Phi(\mathbf{r}_{el}) = \sum_s \frac{a_s}{\sqrt{N!}} \begin{vmatrix} \phi_{1,s}(\mathbf{r}_1) & \dots & \phi_{N,s}(\mathbf{r}_1) \\ \dots & & \dots \\ \phi_{1,s}(\mathbf{r}_N) & \dots & \phi_{N,s}(\mathbf{r}_N) \end{vmatrix}, \tag{3.22}$$

where the a_s are constants, N is the number of electrons in the system and $\phi_{1,s}(\mathbf{r}_i)...\phi_{N,s}(\mathbf{r}_i)$ are the one-electron spin orbitals corresponding to Slater determinant s. In spin-restricted calculations, the spatial components of the spin orbitals are assumed to be identical for each pair of electrons. In spin-unrestricted calculations, two distinct sets of molecular orbitals are used for α and β spin electrons. Most commonly, the Schrödinger equation is first solved for a single Slater determinant (with $a_0 = 1$) to yield the Hartree-Fock (also known as the self-consistent field, SCF) wavefunction, from which further determinants corresponding to various electronic excitations can then be generated. Accurate treatment of some chemically complicated systems such as biradicals may, however, require the use of multireference methods, in which multiple Slater determinants built from different sets of one-electron spin orbitals are present from the start. Since multireference methods have not been used in any of the papers included in this thesis, they will not be treated further here.

The second main approximation concerning the form of the wavefunction is the basis-set expansion, in which each one-electron spin orbital is expressed as a linear combination of basis functions $f_1,...,f_{m_{basis}}$

$$\phi_i(\mathbf{r}_i) = \sum_{j=1}^{m_{basis}} c_{i,j} f_j(\mathbf{r}_i), \tag{3.23}$$

where $c_{i,j}$ are called orbital expansion coefficients. Several different types of basis functions have been developed for various purposes, but for studies of molecular clusters the most common choice is to use gaussian functions

$$f_{\xi,n,l,m}(r,\theta,\varphi) = N_c Y_{l,m}(\theta,\varphi) r^{(2n-2-l)} e^{-\xi r^2}, \tag{3.24}$$

where N_c is a normalization constant, n, l and m are quantum numbers (see *e.g.* Atkins and Friedman [1997] for definitions), $Y_{l,m}(\theta,\varphi)$ is a spherical harmonic function and ξ is a constant. The co-ordinate system in equation (3.24) is normally atom-centered (though some applications may require the placing of basis functions on other centers than atomic nuclei), and each atom type is normally represented by a certain number of basis functions. Slater-type functions, which resemble gaussian functions except that the

exponential term is proportional to $-r$ instead of $-r^2$, are also often used. Numerical grids fitted to the appropriate exact atomic wavefunctions, or plane waves, are also common, though the latter are more rarely used for molecular cluster studies. Slater-type functions are exact solutions to the Schrödinger equation for isolated one-electron atoms, and are therefore much more accurate *per basis function* than gaussian functions. However, gaussian functions are computationally much more efficient, mainly due to the fact that the product of two or more gaussian functions is still a gaussian function, which makes many-center integrals easier to calculate. For this reason, the majority of quantum chemical studies on gas-phase molecular systems use gaussian basis sets. A common procedure to reduce the errors arising from the incorrect behavior of the $-r^2$ - exponential term is to use fixed linear combinations of several gaussian functions as the basis functions of equation (3.23). This basis-set contraction is used to varying degrees in almost all common gaussian basis sets.

If the number of basis functions $m_{basis} = N/2$ (or $m_{basis} = N$ in case of spin-unrestricted calculations) in equation (3.23), there is precisely one basis function per electron pair (or single electron), in which case the basis set is called minimal. In most applications requiring chemical accuracy, the basis set needs to be significantly larger than minimal, and two kinds of one-electron spin orbitals will be generated: occupied and virtual. The virtual orbitals are used in the treatment of electron correlation as described below.

Non-minimal basis sets are usually classified by the number of basis functions they contain per electron pair, with double-ξ containing two, triple-ξ three and so forth. As chemical reactions usually involve rearrangements of the valence electrons only, with the core electrons acting as "spectators", additional basis functions are often added for the valence electrons only, leading to the classifications valence double-ξ, valence triple-ξ *etc*. To improve the description of chemical bonding, additional polarization functions (corresponding to higher values of the angular momentum quantum number l) are often added to the basis set. For commonly used double- or triple-ξ basis sets, the number of polarization functions typically varies between one and four per atom. For accurate descriptions of weakly bound clusters, diffuse functions (corresponding to low values of ξ) are also often added to the basis set. Diffuse functions can be visualized as corrections for the too fast decay of the e^{-r^2} term in the gaussian basis functions far away from the nucleus as compared to the real e^{-r} decay of the exact one-electron orbitals. The corresponding erroneous behavior close to the nuclei is corrected for by the contraction procedure described above. While the concepts of basis set size and polarization functions are applicable also to Slater-type and numerical basis functions, it can be argued that the concept of diffuse functions is related to gaussian basis sets only.

The solution of the electronic wavefunction is based on the variational principle, which states that any approximate solution to the Schrödinger equation has an energy above or equal to that of the exact energy, with the equality holding only if the approximate solution is the exact solution (for a proof see *e.g.* Jensen [2007]). Thus, the set of coefficients $c_{i,j}$ that give the best approximation to the true ground-state energy and wavefunction are those which minimize the energy, in which case

$$\frac{\partial E}{\partial c_{i,j}} = 0 \tag{3.25}$$

for every $c_{i,j}$. Application of the variational principle to a single-determinant wavefunction leads to a set of N Hartree-Fock equations, expressed here in atomic units*:

$$(\frac{-1}{2}\nabla^2 - \sum_{j=1}^{K}\frac{Q_j}{|\mathbf{r}_i - \mathbf{R}_j|} + \sum_{j\neq i}^{N}\int\frac{|\phi_j(\mathbf{r}')|^2}{|\mathbf{r}_i - \mathbf{r}'|}d\mathbf{r}' - \sum_{j\neq i}^{N}\int\frac{\phi_j^*(\mathbf{r}')\phi_i(\mathbf{r}')}{|\mathbf{r}_i - \mathbf{r}'|}d\mathbf{r}')\phi_i(\mathbf{r}_i) = \varepsilon_i\phi_i(\mathbf{r}_i)\,, \tag{3.26}$$

where $\mathbf{R}_1,...,\mathbf{R}_K$ are the positions of the K nuclei, $Q_1,..,Q_K$ are their charges, ε_i is the energy corresponding to the one-electron orbital ϕ_i and $i = 1...N$ (see *e.g.* Atkins and Friedman [1997] for a derivation). From left to right, the terms on the left-hand side of the equation correspond to the kinetic energy, nuclei-electron Coulombic attraction, electron-electron Coulombic repulsion and electron-electron exchange interactions. The exchange interaction does not have a classical counterpart, but arises from the Pauli exclusion principle through the construction of the Slater determinant (equation 3.22). For practical calculations, the Hartree-Fock equations must be converted into the Roothan-Hall equations, which are expressed in matrix form

$$\mathbf{FC} = \mathbf{SCE}, \tag{3.27}$$

where $\mathbf{C}$ is the matrix of orbital expansion coefficients (see equation 3.23), $\mathbf{E}$ is a diagonal matrix containing the orbital energies ε_i, and $\mathbf{F}$ and $\mathbf{S}$ are called the Fock and overlap matrixes, respectively

$$\begin{aligned} F_{ab} &= \int f_a^*(\mathbf{r})\hat{h}f_b(\mathbf{r})d\mathbf{r} \\ &+ \sum_{c=1}^{m_{basis}}\sum_{d=1}^{m_{basis}} D_{cd}\left(\int f_a^*(\mathbf{r})\,f_c^*(\mathbf{r}')\hat{g}\,f_b(\mathbf{r})\,f_d(\mathbf{r}')d\mathbf{r}d\mathbf{r}' - \int f_a^*(\mathbf{r})\,f_c^*(\mathbf{r}')\hat{g}\,f_d(\mathbf{r})\,f_b(\mathbf{r}')d\mathbf{r}d\mathbf{r}'\right) \end{aligned} \tag{3.28}$$

$$S_{ab} = \int f_a^*(\mathbf{r})\,f_b(\mathbf{r})d\mathbf{r}\,, \tag{3.29}$$

where the operator $\hat{h}$ contains all one-electron interactions

$$\hat{h} = -\frac{1}{2}\nabla^2 - \sum_{j=1}^{K}\frac{Q_j}{|\mathbf{r} - \mathbf{R}_j|}, \tag{3.30}$$

the operator $\hat{g}$ contains the two-electron interactions

$$\hat{g} = \frac{1}{|\mathbf{r} - \mathbf{r}'|} \tag{3.31}$$

and the D_{cd} are elements of the reduced density matrix $\mathbf{D}$, defined as

$$D_{cd} = \sum_{j=1}^{occ} c_{c,j}c_{d,j}\,, \tag{3.32}$$

where the summation is over the occupied orbitals. The integrals in equation (3.28) are performed over products of four basis functions, which means that the number of

* Atomic units are used to simplify the rather lengthy expressions encountered in quantum mechanics by eliminating various constants such as h and c. The atomic unit of length is the Bohr radius, 0.529177249 Å, the atomic unit of energy is the Hartree, $4.3597482 \times 10^{-18}$ J, and the atomic unit of mass is that of an electron, $9.1093897 \times 10^{-31}$ kg. All equations in this section and the next will hereafter be given in atomic units.

integrations needed grows as the fourth power of the basis set size. Thus, the so-called formal scaling (the factor by which the computational effort grows as a function of the system size) of the Hartree-Fock methods is N^4, though in practice the scaling is usually within the interval $N^2 ... N^3$ [Jensen 2007].

The Roothan-Hall equations can be solved to yield the orbital expansion coefficients by diagonalizing the Fock matrix. The total energy can be given as [Jensen 2007]

$$E = \sum_{a=1}^{m_{basis}} \sum_{b=1}^{m_{basis}} D_{ab} \int f_a^*(\mathbf{r}) \hat{h} f_b(\mathbf{r}) d\mathbf{r} + \frac{1}{2} \sum_{a=1}^{m_{basis}} \sum_{b=1}^{m_{basis}} \sum_{c=1}^{m_{basis}} \sum_{d=1}^{m_{basis}} (D_{ab} D_{cd} - D_{ad} D_{cb}) \int f_a^*(\mathbf{r}) f_c^*(\mathbf{r}') \hat{g} f_b(\mathbf{r}) f_d(\mathbf{r}') d\mathbf{r} d\mathbf{r}' + V_{Nuc} \quad , \tag{3.33}$$

where V_{Nuc} contains all the interactions between the nuclei (treated as classical point charges). However, as the wavefunction and energy of each electron depends on the wavefunction and energy or every other electron (*i.e.* the Fock matrix is not known unless all the orbital expansion coefficients are already known), the solving of the Roothan-Hall equations must be done iteratively, starting out from some initial guess matrix, called the trial wavefunction. The trial wavefunction may be obtained in various ways, for example by diagonalizing a "core" Fock matrix containing only the one-electron interactions, or by using semiempirical methods (see Jensen [2007] for details). The trial wavefunction can then be used to compute a new, improved set of orbital expansion coefficients and thus a new **D** matrix. Repeating the process until convergence yields the Hartree-Fock energy and wavefunction. In modern quantum chemistry programs, the Roothan-Hall equations are not solved by simple diagonalizations. Instead, various sophisticated algorithms are used to minimize the number of iterative steps (or the computational time and/or memory) required for convergence.

The solution to the Hartree-Fock equations represents the best possible single-determinant wavefunction formed from the given basis set. However, as the HF approach assumes that each electron moves in the average potential of the other electrons, the correlation between the electrons is neglected. While the correlation accounts for only a small fraction of the total energy, it is vitally important for accurate simulations of chemical reactions and bonding. There are three main approaches to including correlation within wavefunction theory: configuration interaction (CI), Møller-Plesset (MP) perturbation theory and coupled-cluster theory (CC). While their theoretical background is quite different, all methods work in a rather similar way in terms of the wavefunctions presented above. First, a Hartee-Fock calculation is performed to yield a set of occupied and virtual orbitals. Second, a set of additional Slater determinants are formed by substituting one or more occupied orbitals for virtual orbitals in the original Hartree-Fock determinant. (Physically, this corresponds to exciting electrons to higher energy levels.) The number of simultaneous substitutions per Slater determinant is determined by the level of the method: singly excited Slater determinants contain only one substitution, doubly excited ones contain two and so on. Next, the coefficients a_s in equation (3.22) are determined, and the correlation energy is evaluated using these coefficients. In multireference methods the $c_{i,j}$ and a_s coefficients would be optimized simultaneously, but that is beyond the scope of this thesis.

In CI, the a_s coefficients are determined using a variational approach analogous to equation (3.25), with the energies determined by diagonalizing the Hamiltonian CI matrix. Full CI (FCI) corresponds to allowing all possible excitations, and in the infinite basis-set limit yields a numerically exact solution to the Schrödinger equation (given the limitations of the Born-Oppenheimer approximation and the neglect of relativistic effects). However, FCI scales as $N!$, and in practice the CI expansion is restricted to single, double and perhaps triple excitations, giving rise to the CIS, CISD and CISDT methods, respectively. (It should be noted that the CIS energy does not represent an improvement over the HF method, though it is sometimes used for probing excited states.) However, the truncated CI methods are not size extensive, meaning that the energy calculated for n noninteracting systems is not equal to n times the energy of one system. For this reason, CI methods are nowadays not frequently used for energy calculations.

Møller-Plesset (MP) perturbation theory [Møller and Plesset 1934] is based on treating electron correlation as a small perturbation of the HF Hamiltonian

$$\hat{H}_{M-P} = \hat{H}_{HF} + \lambda(\hat{V}_{ee} - 2\langle \hat{V}_{ee} \rangle), \tag{3.34}$$

where $\hat{H}_{HF}$ is the Hartree-Fock Hamiltonian operator, λ is an arbitrary constant, $\hat{V}_{ee}$ is the potential energy operator corresponding to electron-electron interactions (see equation 3.19) and the "<>" brackets indicate an expectation value. The difference of the exact $\hat{V}_{ee}$ operator and twice its expectation value is called the fluctuation potential. The assumption that the perturbation corresponding to the fluctuation potential is small compared to the HF Hamiltonian is not necessarily justified, but unlike all other perturbation-theory based alternatives presented so far, MP theory is size-extensive. The formula for the n:th order perturbative energy and wavefunction can be readily obtained from standard perturbation theory once the form of the perturbation is known, though they rapidly become too complicated to evaluate. The methods are indexed as MPn by the level of perturbation. Like CIS, MP1 does not represent an improvement over the HF energy. By far the most common variant of MP theory is MP2, for which the energy can be expressed as [Leach 2001]

$$E_{MP2} = \sum_{a}^{occ.} \sum_{b>a}^{occ.} \sum_{c}^{virt.} \sum_{d>c}^{virt.} \frac{\int \phi_a^*(\mathbf{r})\phi_b^*(\mathbf{r}') \frac{1}{|\mathbf{r}-\mathbf{r}'|} [\phi_c(\mathbf{r})\phi_d(\mathbf{r}') - \phi_d(\mathbf{r})\phi_c(\mathbf{r}')] d\mathbf{r} d\mathbf{r}'}{\varepsilon_c + \varepsilon_d - \varepsilon_a - \varepsilon_b}, \tag{3.35}$$

where it should be noted that the integration is performed over all occupied and virtual molecular orbitals, not the atomic orbitals. MP2 formally scales as N^5, recovers around 80-90% of the correlation energy [Jensen 2007] and is the most affordable (and therefore probably also the most widely used) correlated wavefunction-based method. For cases where the HF wavefunction is not a good description of the system (*e.g.* biradicals), MP2 often fails badly, though this characteristic is to some extent shared by all non-multireference methods. MP3 does not usually significantly improve upon the MP2 result, and MP4, which formally scales as N^7, is the only other perturbation-based method that is commonly used.

Coupled-cluster methods are based on the parametrization of the wavefunction by an exponential ansatz [Jensen 2007]

$$\Psi_{CC} = e^{\hat{T}}\Phi_{HF}\,, \tag{3.36}$$

where Φ_{HF} is the HF reference configuration (the Slater determinant of equation 3.22 with s=0) and $\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 ...$ is the cluster operator, expressed as the sum of all excitation operators, with $\hat{T}_1$ corresponding to single excitations, $\hat{T}_2$ to double excitations and so on. In practice, the cluster operator is truncated at some point, after which the exponential can be expanded in a Taylor series. The level of truncation is indicated by letter indexes. For example, CCD contains only double, CCSD both single and double, and CCSDT single, double and triple excitation operators. Some other approximations may be made in the construction or evaluation of the excitation operator, resulting in *e.g.* the Brueckner doubles (BD), CC2 and CC3 methods. The CC methods are size-extensive like MP theory, but in contrast to the analytical MP energy expressions the CC energy must be evaluated iteratively, using rather complicated techniques (see *e.g.* Bartlett and Musial [2007] for a comprehensive review of coupled-cluster methods). Also, in contrast to MPn methods, where *n*:th order perturbations are treated to *n*:th order, CC methods treat *n*:th order perturbations to infinite order. This can be demonstrated by suitably grouping the terms of the Taylor expansion of, for example, the exponential *ansatz* for the CCSD method [Jensen 2007]:

$$e^{\hat{T}_1+\hat{T}_2} = 1 + \hat{T}_1 + (\hat{T}_2 + \frac{1}{2}\hat{T}_1^2) + (\hat{T}_2\hat{T}_1 + \frac{1}{6}\hat{T}_1^3) + (\frac{1}{2}\hat{T}_2^2 + \frac{1}{2}\hat{T}_2\hat{T}_1^2 + \frac{1}{24}\hat{T}_1^4) + ... \tag{3.37}$$

This means that the fraction of the correlation energy recovered (and the accuracy of the results) is significantly higher for CC than for MP methods of equivalent order. Unfortunately, the formal scaling of *n*:th order CC methods is N^{2n+2} while that for MPn methods is N^{n+3}, in addition to which the actual computational effort of CC methods tends to be significantly (often around 10-20 times) higher than that of MP methods with the same formal scaling. (For some examples of calculation times see section 3.3.4.) For cases where the HF wavefunction is a good starting point, CCSD(T) (where the triple excitations are evaluated non-iteratively using perturbation theory) is considered to be the state-of-the art method, and usually yields relative energies very close to the FCI limit. If results beyond the CCSD(T) level are desired, other error sources (such as relativistic errors and the Born-Oppenheimer approximation) must usually also be accounted for. For the (small) systems where this can be done, the accuracy of computational results is often higher than that of the most accurate experimental apparatus available. As an example of the accuracy that high-level computational methods can yield for small systems, the experimental vibrational wavenumbers for gas-phase ammonia were recently matched to within 1 cm^{-1} [Rajamäki *et al.* 2004] using a combination of basis-set limit CCSDTQ(P) and FCI energies, relativistic corrections and advanced vibrational models. Unfortunately, the most accurate quantum chemical tools are only applicable to systems containing, at most, a few tens of electrons. The clusters studied in atmospheric applications are therefore far too large for these methods to be used, and the wavefunction-based methods are usually limited to MP2, with possible single-point energy corrections at slightly higher levels.

As mentioned above, chemical reactions are related mainly to the valence electrons, while the core electrons are less interesting from a chemical point of view. For this reason, quantum chemical studies using correlated wavefunction-based methods that focus on topics related to chemical binding (*e.g.* energetics, thermochemistry, kinetics and many types of spectroscopy) usually omit the core electrons from the correlation treatment. While this so-called frozen core (FC) assumption is not a good approximation for the total electronic energies, its influence on relative energies, such as chemical formation and reaction energies, is minimal. Due to the steep scaling of correlated wavefunction-based methods, the computational savings of the FC approximation are significant, especially if second-row or heavier elements are present. It should be noted that simply including the core electrons into the correlation treatment in a standard calculation is not sufficient to model the correlation of core electrons; specialized basis sets are also needed, further increasing the computational cost.

3.3.2. Density functional theory

In 1964, Hohenberg and Kohn proved [Hohenberg and Kohn 1964], based on the variational principle, that for an arbitrary system of electrons, moving in an external potential $V_{ext}(\mathbf{r})$ (and subject to mutual Coulombic repulsions), the ground state electron density $\rho(\mathbf{r})$ uniquely determines the external potential. Since the electron density also determines the total amount of electrons via the Born interpretation of quantum mechanics

$$N = \int \rho(\mathbf{r}) d\mathbf{r} \,, \tag{3.38}$$

it also determines the Hamiltonian of the system and thus its full (ground-state) wavefunction, and all the information contained in it. The ground-state energy is therefore also a functional of $\rho(\mathbf{r})$. However, while the full wavefunction is $3N$ or $4N$ – dimensional (as every particle is described by three spatial and one spin variable), the electron density is a 3-dimensional function regardless of the number of particles in the system. If the functional for the ground-state energy were known, DFT would thus allow the computation of the energy of any system at essentially a fixed cost, regardless of the number of particles in the system [Jensen 2007]. However, so-called orbital-free DFT methods are not (yet) accurate enough to be of general use, and the electron density still has to be expressed in terms of one-electron orbitals analogously to equation (3.22) in the previous section

$$\rho(\mathbf{r}) = \sum_{i=1}^{N} \left|\phi_i(\mathbf{r})\right|^2 \tag{3.39}$$

with the one-electron orbitals further expanded in terms of a basis set just like in equation (3.23). Often, even the basis sets used for wavefunction-based and DFT calculations are the same. The idea of expressing the electron density in terms of an orbital basis was first proposed by Kohn and Sham [1965], and it is the foundation of all modern DFT calculations. Strictly speaking, the DFT used in quantum chemistry should be called Kohn-Sham DFT to distinguish it both from classical DFT and from other quantum mechanical DFT variants, such as the orbital-free methods briefly mentioned above. (In this thesis, the common convention is adopted that "DFT" always refers to Kohn-Sham

DFT unless otherwise mentioned.) It should be noted that though Kohn and co-workers themselves strongly denied that the Kohn-Sham orbitals of equation (3.39) have any physical significance [Kohn *et al.* 1996], some later researchers [Baerends and Gritsenko 1997] have claimed that they are, in fact, even more physically based than the corresponding Hartree-Fock orbitals.

Similarly to wavefunction-based methods, DFT methods can be either restricted or unrestricted depending on the treatment of electron spin. In the latter case, separate Kohn-Sham orbitals are included for α and β spin electrons, and the total density can be expressed as a sum of two components, $\rho(\mathbf{r})_\alpha$ and $\rho(\mathbf{r})_\beta$.

The next step in DFT calculations is to express the energy functional as a sum of four terms, three of which are known

$$E[\rho(\mathbf{r})] = \int V_{ext}(\mathbf{r})\rho(\mathbf{r})d\mathbf{r} + T[\rho(\mathbf{r})] + E_H[\rho(\mathbf{r})] + E_{XC}[\rho(\mathbf{r})], \tag{3.40}$$

where the first term accounts for the external potential (in practice, the interaction between the nuclei and the electrons), the second term is the kinetic energy of a noninteracting system of electrons with density $\rho(\mathbf{r})$, and the third term is the Coulombic repulsion between the electrons. The kinetic energy and Coulomb terms can be given exactly as

$$T[\rho(\mathbf{r})] = \sum_{j=1}^{N} \int \phi_j^*(\mathbf{r}) dr \frac{-\nabla^2}{2} \phi_j(\mathbf{r}) d\mathbf{r} \tag{3.41}$$

$$E_H[\rho(\mathbf{r})] = \frac{1}{2} \int \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}'. \tag{3.42}$$

$E_{XC}[\rho(\mathbf{r})]$ accounts for the exchange and correlation interactions, and it is the sole unknown term in the functional. If $E_{XC}[\rho(\mathbf{r})]$ were known exactly, DFT would yield the exact solution of the Schrödinger equation. However, unlike the excitation operators in correlated wavefunction-based methods, there is no systematic way to develop or even improve the exchange-correlation functionals in DFT, and various approximate (and highly variable) forms have to be used, as described below.

In practice, DFT calculations are very similar to HF calculation. Analogously to the variational principle, Hohenberg and Kohn [1964] proved that the true ground-state density corresponds to a minimum value for the energy, and the best approximation to the ground-state density and energy is obtained by minimizing the energy with respect to the orbital expansion coefficients of the Kohn-Sham orbitals. This leads to the set of *N* Kohn-Sham equations, which are analogous to the Hartree-Fock equations (3.26) [Leach 2001]

$$\left(-\frac{1}{2}\nabla^2 + V_{ext}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + v_{XC}(\mathbf{r})\right)\phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}), \tag{3.43}$$

where

$$v_{XC}(\mathbf{r}) = \frac{\partial E_{XC}[\rho(\mathbf{r})]}{\partial \rho(\mathbf{r})}. \tag{3.44}$$

In computational applications, the Kohn-Sham equations are solved like the Hartree-Fock equations, using linear algebraic techniques and iterating to self-consistency.

The development of DFT since the 1960s has essentially consisted of a search for better and better exchange-correlation functionals. Based on the analysis of Perdew (Perdew *et al.* [2005], also summarized by Jensen [2007]), exchange-correlation functionals can be divided into five rungs (also called "generations", "steps" or "levels") along a Jacob's ladder ascending toward the heaven of chemical accuracy. In Perdew's classification, functionals are grouped depending on the types of terms they contain as follows:

1. The first rung corresponds to LDA, the local density approximation. In the LDA, $E_{XC}[\rho(\mathbf{r})]$ is approximated by the exchange-correlation energy of a uniform electron gas. Exact expressions for this are available both at the low- and high-density limits, and various intermediate values can be interpolated using high-level quantum Monte-Carlo results [*e.g.* Ceperley and Alder 1980]. The spin-unrestricted version of LDA is often called LSDA, local spin density approximation. While the performance of LDA is satisfactory for some metallic systems, it is useless for modeling gas-phase molecules, as bond strengths may be overestimated by several tens of kcal/mol.

2. The next step up along Perdew's Jacob's ladder are the generalized gradient approximation, or GGA, functionals, which in addition to terms proportional to $\rho(\mathbf{r})$ also contain gradient terms proportional to $\nabla\rho(\mathbf{r})$. GGA functionals were first developed in the late 1980s, and have remained popular ever since, as they are relatively simple to evaluate and provide near-chemical accuracy (often defined as an accuracy of 1 kcal/mol with respect to the reaction or binding energies of molecular systems) for many applications. For example the PW91 functional used in **Papers I** and **II** belongs to the GGA family.

3. In Perdew's classification, the third rung of the ladder corresponds to the meta-GGA functionals, which in addition to terms proportional to ρ(r) and $\nabla\rho(\mathbf{r})$ contain the kinetic energy density (or some related term), which is obtained from the occupied Kohn-Sham orbitals. The dependence of the functional on the orbitals increases the computational requirements, but sometimes improves the correspondence with higher-level wavefunction-based or experimental data.

4. The fourth rung corresponds to hybrid DFT functionals (also called hyper-GGA), which contain the Hartree Fock exchange interaction in addition to the above-mentioned components. Since the HF exchange is exact, it could be expected that replacing all the exchange in $E_{XC}[\rho(\mathbf{r})]$ by the HF exchange would lead to the best results. However, as noted by Gritsenko *et al.* [1997], the definitions of exchange in HF-theory and DFT are slightly different, mainly due to the fact that DFT exchange and correlation are (at least in the level 1-4 functionals) inherently local, while HF exchange also contains non-local components. The long-range exchange and correlation interactions are to some extent cancelled out in the construction of DFT functionals, and replacing all exchange terms by HF exchange destroys this error cancellation and leads to erroneous results. Most hybrid functionals (such as the extremely popular B3LYP functional used in **Papers I and V**,

and the heavily parametrized MPW1B95 functional used in **Paper VI**) contain both HF and DFT exchange components, with the exact fraction of each determined for example by fitting to experimental results. It should be noted that Perdew *et al.* [2005] regard functionals containing empirical fitting parameters as highly suspect, and note that functionals such as B3LYP do not even reproduce the correlation energy of a uniform electron gas, which they regard as a "sacrosanct" minimum criteria that all functionals should fulfill.

5. The fifth rung corresponds to DFT functionals that include terms containing the unoccupied (virtual) Kohn-Sham orbitals. This allows the treatment of dispersion (non-local correlation) effects, which are the stumbling block of level 1-4 functionals. However, this also increases the scaling of DFT methods to (at least) N^5 or N^6, corresponding to MP2 calculations. Also, convergence with respect to basis set size is slow. Few fifth-level DFT functionals are in use at present, and experiences of their use are very limited. However, intensive development work on computationally affordable fifth-rung functionals is underway [see *e.g.* Schwabe and Grimme 2006].

The five-step ladder presented above is by no means the only possible. The existence and ordering of rungs 1,2 and 5 are generally acknowledged and agreed upon, but some authors consider levels 3 and 4 to be the same, or classify them by some other category. Perdew *et al.* [2005] emphasize a non-empirical approach to density functional design, and strongly criticize the construction of functionals by extensive parameter fitting to molecular datasets. Unsurprisingly, the authors responsible for such constructions view the issue differently [see *e.g.* Zhao and Truhlar 2004].

3.3.3. Thermochemistry

The starting point for the calculation of thermodynamic parameters in quantum chemistry simulations is almost always an optimized minimum-energy geometry. Usually, it is assumed that the translational, rotational, vibrational and electronic degrees of freedom can be separated from each other. The partition function for the system can then be expressed as a product of four components

$$q = q_{el} q_{vib} q_{rot} q_{trans} . \tag{3.45}$$

The entropy and enthalpy (and all other thermodynamic parameters) corresponding to each component q_X can then be computed separately, using the standard formulae [see *e.g.* Lindler 2004]

$$H_X = k_B T^2 \left(\frac{\partial \ln q_X}{\partial T}\right)_V + k_B T V \left(\frac{\partial \ln q_X}{\partial V}\right)_T \tag{3.46}$$

$$S_X = k_B T \left(\frac{\partial \ln q_X}{\partial T}\right)_V + k_B \ln(q_X) , \tag{3.47}$$

where T denotes temperature and V volume. The four contributions can be added together to give

$$H_{tot} = H_{el} + H_{vib} + H_{rot} + H_{trans} \tag{3.48}$$

$$S_{tot} = S_{el} + S_{vib} + S_{rot} + S_{trans} . \tag{3.49}$$

$$G_{tot} = H_{tot} - TS_{tot} \tag{3.50}$$

In studies of gas-phase molecules or clusters, the ideal gas assumption is usually used to compute the translational component. The translational partition function is then

$$q_{trans} = (\frac{2\pi M k_B T}{h^2})^{\frac{3}{2}} V = (\frac{2\pi M k_B T}{h^2})^{\frac{3}{2}} \frac{k_B T}{P}, \tag{3.51}$$

where M is the mass and P the partial pressure of the molecule or cluster. The corresponding entropy and enthalpy terms are

$$S_{trans} = k_B \left[\frac{5}{2} + \ln(\frac{2\pi M k_B T}{h^2})^{\frac{3}{2}} + \ln(\frac{k_B T}{P}) \right] \tag{3.52}$$

$$H_{trans} = \frac{5}{2} k_B T . \tag{3.53}$$

As the contributions from translational motion depend only on the molecular mass, temperature and pressure, they can be computed without any results from the quantum chemical calculations. In equation (3.51), the partial pressure is often set to the "standard" value of 1 atm. If, for example, free energies at some other pressures are desired, the standard free energies must then be corrected as follows:

$$G(P) = G(P_0) - k_B T \ln(\frac{P}{P_0}), \tag{3.54}$$

where P_0 = 1 atm and P is the ambient pressure. It should be noted that the gas pressure affects only the translational component of the partition function.

The simplest method to calculate the rotational partition function is to assume that the molecule (or cluster) can be treated as a rigid rotor. In this case, the only parameters needed for the calculation are the rotational symmetry number σ_{symm} of the system (typically equal to 1 for larger cluster structures) and the moments of inertia about the principal axes of inertia. These are usually obtained numerically by diagonalizing the 3×3 matrix consisting of the moments of inertia about any set of Cartesian axes passing through the center of mass, as described by Jensen [2007]. The only information required from the quantum chemical computations is the positions of the atomic nuclei. The rotational partition function for a rigid rotor is

$$q_{rot} = \frac{\sqrt{\pi}}{\sigma_{symm}} (\frac{8\pi^2 k_B T}{h^2})^{\frac{3}{2}} \sqrt{I_1 I_2 I_3}, \tag{3.55}$$

where I_1, I_2 and I_3 are the moments of inertia about the principal axes. The corresponding entropy and enthalpy contributions are

$$S_{rot} = k_B \left[(\frac{3}{2} + \ln(\frac{\sqrt{\pi}}{\sigma_{symm}} (\frac{8\pi^2 k_B T}{h^2})^{\frac{3}{2}} \sqrt{I_1 I_2 I_3}) \right] \tag{3.56}$$

$$H_{rot} = \frac{3}{2} k_B T . \tag{3.57}$$

The vibrational contribution is by far the most difficult component of the partition function to calculate. Even the simplest theoretical framework - in which all vibrations are treated as harmonic - requires calculation of the second derivatives of the electronic energy with respect to $3K$ nuclear co-ordinates (where K is the number of atomic nuclei). This may require as much or more computational effort than all the other steps combined.

The first step in calculating the vibrational partition function is to define the normal vibrational modes. This is normally done by computing $3K$ force constants in some suitable co-ordinate system (which may consist of *e.g.* Cartesian co-ordinates for each atom, or combinations of internal co-ordinates such as bond distances, angles and dihedral angles). The calculations can be done numerically using finite difference methods, although analytical expressions by which the force constants can be computed directly from the electronic wavefunction have been developed for many common quantum chemical methods. Analytical frequency calculations are typically very memory-intensive even though they are usually significantly faster than numerical ones. The normal modes are obtained by diagonalizing the mass weighted force $3K \times 3K$ constant matrix, as described by Wilson *et al.* [1955]. The eigenvectors correspond to the normal modes, and the eigenvalues to the harmonic vibrational frequencies. For non-linear molecules, six of the eigenvalues will be zero (or very close to zero), as they correspond to rotational and translational degrees of freedom. The vibrational partition function for a (non-linear) molecular system treated as a collection of harmonic oscillators is

$$q_{vib} = \prod_{i=1}^{3K-6} \frac{e^{-\frac{h\nu_i}{2k_BT}}}{1-e^{-\frac{h\nu_i}{k_BT}}}, \tag{3.58}$$

where ν_i is the frequency corresponding to normal mode i. The corresponding entropy and enthalpy contributions are

$$S_{vib} = k_B \sum_{i=1}^{3K-6} \left(\frac{h\nu_i}{k_BT(e^{\frac{h\nu_i}{k_BT}}-1)} - \ln(1-e^{\frac{-h\nu_i}{k_BT}}) \right) \tag{3.59}$$

$$H_{vib} = k_B \sum_{i=1}^{3K-6} \left(\frac{h\nu_i}{2k_B} + \frac{h\nu_i}{k_B(e^{\frac{h\nu_i}{k_BT}}-1)} \right). \tag{3.60}$$

The first, temperature-independent, component of the enthalpy expression is called the zero-point energy. Its origin is the Heisenberg uncertainty principle, which together with the quantization of vibrational energy requires that a vibrating system can never be completely at rest.

Unfortunately, real molecules - not to mention weakly bound molecular clusters - are not rigidly rotating harmonic oscillators, and application of equations (3.55) – (3.60) may lead to serious errors. Kathmann *et al.* [2007] recently investigated the effect of anharmonicity on the thermochemistry of hydrated ions. In terms of the formation free energies at 298 K, the effect of anharmonicity was found to be notable for $j>4$ for

$Na^+\bullet(H_2O)_j$ clusters and $j>1$ for $Cl^-\bullet(H_2O)_j$ clusters. It should be noted that Kathmann *et al.* use the term "anharmonicity" to cover two different issues: the anharmonicity of vibrational frequencies themselves, and the contribution of other conformers than the minimum energy structure to the thermochemical parameters of the cluster. Within the framework of quantum chemical methods, the solution recommended by Kathmann *et al.* - a complete sampling of the configuration space - is impossible due to the enormous computational effort required. However, various schemes have been developed to at least partially correct for the worst errors of the rigid rotor - harmonic oscillator (RRHO) approximation. In general, the purpose of these schemes is to numerically compute a set of more accurate vibrational or rotational-vibrational energy levels E_i, after which the partition function can be calculated:

$$q_{vib}q_{rot} = q_{vibrot} = \sum_i g_i e^{\frac{-E_i}{k_B T}}, \tag{3.61}$$

where g_i is the degeneracy of the i:th energy level. Entropy and enthalpy contributions can then be computed using equations (3.46) and (3.47).

The simplest method to account for vibrational anharmonicity is to keep the rotational and vibrational degrees of freedom separate, and simply multiply the harmonic frequencies by some scaling factor. The vibrational thermochemical contributions can then be calculated using these scaled frequencies. Scaling factors for DFT-based methods are typically on the order of 0.95-0.99, while those for correlated wavefunction-based methods are usually somewhat lower [Scott and Radom 1996, Foresman and Frisch 1996]. The main problem of the scaling factor approach is that different types of vibrations have different "anharmonicities". For example, vibrations corresponding to stretching motions of X-H bonds (where X is any atom heavier than hydrogen) are typically more anharmonic than stretching motions of other types of bonds. In molecular clusters, intermolecular bonds are normally more anharmonic than intramolecular bonds, as demonstrated by the decrease in scaling factors as a function of cluster size observed in **Paper III**. The vibrational modes of real molecular systems are also coupled to each other, as are the vibrational and rotational modes. Thus, the normal mode model - scaled or not - is in any case not a very good model for their energy levels. This is evident *e.g.* from the scaling factor analysis in **Paper III**, which demonstrates that simply using anharmonic fundamental vibrational frequencies in the expressions derived for harmonic oscillators underestimates the role of anharmonicity by a factor of 2 compared to calculations using more complicated anharmonic partition functions.

Various high-level theoretical models have been developed to account for the anharmonicity of vibrations along a limited number of modes [see *e.g.* Pesonen and Halonen 2003]. However, these methods are normally not applicable to larger molecular clusters with several tens or even hundreds of vibrational modes. The only currently generally available computational methods accounting for anharmonicity that are applicable even to medium-sized molecular clusters are the *ab initio* vibrational self consistent field (VSCF) method developed by Chaban *et al*. [1999], and the perturbative method developed by Barone [2004, 2005].

The VSCF method is based on treating the vibrations of a molecular system as a group of coupled bosons, the energy levels of which can be solved using a basis-set expansion, the variational principle, and linear algebraic techniques, analogously to the SCF method described in section 3.3.1. The basis sets can be *e.g.* eigenfunctions of harmonic oscillators or Morse oscillators [Morse 1929]. Correlation between the oscillators is taken into account using methods analogous to those presented above, such as configuration interaction or coupled-cluster methods [see *e.g.* Christiansen 2004]. The determination of the VSCF Hamiltonian requires on the order of 6K single-point energy evaluations, and is therefore computationally quite expensive. Also, the convergence of the VSCF equations themselves can be problematic for some systems such as weakly bound clusters.

In the algorithm developed by Barone *et al.* [2004, 2005], anharmonicity is treated as a perturbation to the RRHO system, and the anharmonic oscillator energy levels are computed using perturbation theory. The computation requires the evaluation of all third and selected fourth derivatives of the vibrational normal modes and the couplings between them. The number of energy evaluations needed for this method is also on the order of 6K. The perturbative approach may often be slightly faster than the VSCF technique, but as its current implementation in the Gaussian 03 program [Frisch *et al*. 2004] yields only frequencies, not intensities, it is not favored by spectroscopists. There is very little knowledge on the relative qualities of the thermochemical parameters computed by the two methods.

The separation of electronic energy levels is normally large enough that the excited electronic states are not populated at chemically interesting temperatures. The electronic partition function thus only depends on the electronic ground state. If the ground-state energy is taken as the zero of energy, the value of the electronic partition function is simply equal to $g_{0,el}$, the degeneracy of the electronic ground-state wave function. The wave function may be degenerate either due to spin degeneration, as in the case of radical doublets or the triplet oxygen molecule, or spatial symmetry, as in the NO molecule, which is a radical doublet with a spatial degeneracy of 2, yielding total degeneracy of 4. It should be noted that due to the so-called Jahn-Teller effect (see *e.g.* Atkins and Friedman [1997] for details), nonlinear molecules or molecular complexes with spatially degenerate electronic ground states tend to undergo geometric distortions that remove the degeneracy. Stable molecules (and the molecular clusters of atmospheric interest studied in this thesis) usually have spatially non-degenerate singlet ground-state wavefunctions, and it can thus normally be assumed that $g_{0,el} = 1$.

For computing formation free energies of molecular clusters, it is most convenient to set the zero of energy to correspond to the energies of the isolated free atomic nuclei and electrons. In this case the electronic partition function is

$$q_{el} = g_{0,el} e^{\frac{-E_0}{k_B T}}, \tag{3.62}$$

where E_0 (a negative number) is the ground-state electronic energy computed as described in the previous sections. The electronic entropy and enthalpy contributions thus become

$$S_{el} = k_B \ln(g_{0,el}) \tag{3.63}$$

$$H_{el} = E_0 . \tag{3.64}$$

3.3.4. Practical aspects of quantum chemical calculations

As discussed for example by Jensen [2007] and Leach [2001], the level of theory needed for a computation depends both on the property of interest and the accuracy desired. Qualitatively reliable molecular geometries are often produced by simple Hartee-Fock calculations with near-minimal basis sets. For covalently bound systems, third- or fourth-generation density functionals with polarized valence double-ξ basis sets often yield qualitatively very reliable molecular geometries and energies, though valence triple-ξ basis sets with multiple polarization functions may be required for quantitative accuracy. For weakly bound systems (such as the molecular clusters studied in this thesis), DFT methods are less reliable, and correlated wavefunction-based methods are sometimes needed instead. Unfortunately, these methods converge much slower with respect to the basis set size, and accurate energy calculations may require basis sets as large as quadruple-ξ or higher. On the other hand, wavefunction-based methods are two-dimensional, in the sense that increasing the basis set will always, all other things being equal, increase the reliability of the results. This does not apply for DFT methods - different properties do tend to converge toward a limiting value with increasing basis set size, but there is no guarantee that the largest basis set corresponds to the best result. Indeed, as many DFT functionals have been parametrized using computations with quite modest basis sets, it is sometimes argued that medium-size basis sets are preferable for many DFT calculations.

The goal of most computational studies is to obtain the maximum possible accuracy of the properties of interest with the lowest possible computational effort. Toward this end, various cost-reducing techniques are often used. Some of the most important of these, applicable to the type of systems studied in this thesis, are described below.

The minimum-energy geometries and vibrational frequencies are often less sensitive to the level of theory used in the calculation than the energies themselves. However, as described in section 3.3.3, they are often the most computationally expensive properties to calculate. For this reason, it is common practice to compute the geometry and vibrational frequencies at a lower level of theory, and the single-point electronic energy at a higher level of theory. Results from such computations are often reported in the shorthand $method_1/basis_1//method_2/basis_2$ notation, where the subscripts 1 and 2 denote the levels of theory at which the final electronic energy and the minimum-energy geometry have been calculated, respectively. (Calculations in which the two levels are the same are denoted simply by *method/basis*.) Multi-step methods, where the results of several different single-point energy evaluations are combined to yield a final energy value, represent a further extension of the same idea. For example, a calculation using both high-level correlation and a large basis set may be computationally too expensive to perform on a given system. However, its result may be approximated by combining the results of two calculations: one employing a modest level of correlation and a large basis set, and the other employing a high level of correlation and a small basis set. For wavefunction-based methods, analytical formulae have been developed which describe

the convergence of electronic energies with respect to basis set size [see *e.g.* Helgaker 1997]. These can be used to extrapolate the basis-set limit energies, as done in **Papers I, III** and **IV**. Compared to the commonly employed counterpoise correction [Boys and Bernardi 1970] for correcting basis-set related errors in weakly bound systems, the advantage of extrapolation methods is that they are entirely general, and do not require the (sometimes arbitrary) division of the system into components. However, while the counterpoise correction can also be applied to geometry optimizations or vibrational frequency calculations (though the advantage of doing so is often small while the cost may be significant, see *e.g.* **Paper I**), extrapolation procedures can only be applied to the results of completed calculations.

Most quantum chemical studies focus on comparing the energies of different systems (*e.g.* different binding patterns of the same atoms or molecules) rather than on computing the total electronic energies accurately. This allows the use of various approximations that are not well-justified with respect to the total energies, but affect the relative energies only minimally. The frozen-core approximation mentioned in section 3.3.2. is one such example. An analogous method for density functional theory is the use of pseudopotentials, where the explicit treatment of core electrons is replaced by a suitable effective potential function. (This also permits the inclusion of relativistic corrections for heavier atoms.) Recently, various implementations of the Resolution of Indentity (RI) approximation (also known as density fitting, DF, within the DFT framework) have extended the size limit of systems that can be treated at a given limit of chemical accuracy. See *e.g.* Eichkorn *et al.* [1995], Weigend and Häser [1997] and Dunlap [2000] for modern implementations. These methods are based on the use of a second basis set expansion to reduce the scaling of some part of the quantum chemical calculation involving the evaluation of integrals. The RI-MP2 method has been used in **Papers III**, **IV** and **VI**, while Nadykto *et al.* [2007] recently successfully applied density fitting methods to small sulfuric acid - water clusters.

Applied quantum chemical studies are not usually concerned with the mathematical details of the precise techniques by which the equations presented in the previous sections are solved. This work is done using extensively developed computer programs, with specialized (and frequently both complicated and non-intuitive) algorithms taking care of the actual calculations. In practice, computational chemistry is more and more coming to resemble an experimental science, with the computer being used as an experimental tool to probe some molecular system. The main role of the investigator is then to choose his or her tools (*i.e.* some combination of method, basis set and possibly a thermochemical model) based on experience and knowledge of their strengths and weaknesses.

To give the reader some sense of the computational effort required in typical quantum chemical studies, the total CPU hours required for some calculations included in the papers of this thesis are reported in Table 1. The values presented demonstrate that, as mentioned above, the effort required to calculate harmonic vibrational frequencies often exceeds that required by all other job types, while the calculation of anharmonic vibrational frequencies is even more prohibitively expensive. It should be noted that apart

from the algorithm and number or type of processors used, calculation times depend, among other things, on the quality of the input structure or trial wavefunction (or electron density), the amount and type of central and disk memory available and, for multiple-processor jobs, the speed of the inter-node communications net. As a detailed discussion of computer hardware is beyond the scope of this thesis, the values given in Table 1 are intended as order-of-magnitude examples only.

Ref.	molecular system	job type	method/basis set	CPU time (hours)
Paper I	$H_2SO_4 \bullet H_2O$	single-point energy calculation	CCSD(T)/ cc-pV(T+d)Z	35 [a]
Paper III	$H_2SO_4 \bullet (H_2O)_4$	single-point energy calculation	MP4/ aug-cc-pV(D+d)Z	260 [b]
Paper III	$H_2SO_4 \bullet (H_2O)_2$	numerical an-harmonic frequency calculation	MP2/ aug-cc-pV(D+d)Z	8400 [c]
Paper IV	$(H_2SO_4)_2 \bullet (NH_3)_4$	geometry optimization	RI-MP2/ aug-cc-pV(D+d)Z	250 [d]
Paper IV	$(H_2SO_4)_2 \bullet (NH_3)_4$	single-point energy calculation	RI-MP2/ aug-cc-pV(T+d)Z	20 [d]
Paper IV	$(H_2SO_4)_2 \bullet (NH_3)_4$	numerical harmonic frequency calculation	RI-MP2/ aug-cc-pV(D+d)Z	800 [e]
Paper V	$C_{15}H_{24}O_3$ (sCI)	geometry optimization	B3LYP/ 6-311+G(2d,p)	150 [b]
Paper V	$C_{15}H_{24}O_3$ (sCI)	analytical harmonic frequency calculation	B3LYP/ 6-311+G(2d,p)	300 [b]
Paper V	$(CH_3)_2COO \bullet H_2SO_4$	single-point energy calculation	RI-CC2/QZVPP	13 [d]

a) Using the Gaussian 03 program [Frisch *et al.* 2004] on 1 UltraSPARC IV processor
b) Using the Gaussian 03 program [Frisch *et al.* 2004] on 4 UltraSPARC IV processors
c) Using the Gaussian 03 program [Frisch *et al.* 2004] on 24 UltraSPARC IV processors
d) Using the Turbomol 5.8 program [Ahlrichs *et al.* 1989] on 1 dual-core 2.2 GHz AMD Opteron processor
e) Using the Turbomol 5.8 program [Ahlrichs *et al.* 1989] on 16 dual-core 2.2 GHz AMD Opteron processors

Table 1. The total amount of processor-hours (CPU time) needed for various computations carried out in the studies contained in this thesis. For jobs performed on multiple processors, the real time is the CPU time divided by the number of processors. For more details on the calculations see the indicated papers.

4. Review of computational studies on sulfuric acid-containing clusters

As discussed in a recent review by Kurtén and Vehkamäki [2007], the use of quantum chemical methods in nucleation studies can be divided into three categories. The most modest application is to use quantum chemical methods solely to determine part or all of the parameters of a force field, which is then used to compute *e.g.* formation free energies as described in section 3.2. For the sulfuric acid - water system, this has recently been done by Kusaka *et al.* [1998], Kathmann and Hale [2001] and Ding *et al.* [2003a]. In contrast, the most ambitious application would be to use quantum chemical methods to directly calculate the free energy of formation of a critical cluster, and thereby compute *e.g.* nucleation rates without having to resort to any measurement data whatsoever. However, the heavy computational cost of such calculations, together with the error sources related especially to the calculations of thermal entropy and enthalpy contributions (see section 3.3.3) has so far prevented this from being done for any atmospherically relevant nucleation mechanism. An intermediate alternative is to use quantum chemical interaction energies and free energies for small to medium-sized clusters to investigate the fundamental chemical behavior of molecular systems relevant to nucleation, for example in order to qualitatively compare the plausibility of different nucleation mechanisms. Within the context of sulfuric acid - water - ammonia clusters, two research questions which have recently been successfully answered by such an approach concern the formation of sulfuric acid hydrates in ambient conditions, and the role of ammonia in nucleation. Possible atmospheric nucleation mechanisms involving sulfuric acid together with various organic molecules have also been studied, though the results are still far from conclusive.

Experimental studies [Hanson and Eisele 2000] indicate that in atmospheric conditions, sulfuric acid is bound to one or more water molecules. Hydrates correspond to a local minimum on the free energy surface, and the free energy difference between the critical cluster and the most stable hydrates is larger than the difference between the critical cluster and the isolated free molecules. Therefore, sulfuric acid - water nucleation models that do not account for hydrate formation tend to overestimate the nucleation rate [Heist and Reiss 1974]. One of the earliest applications of quantum chemistry to atmospheric nucleation has been to estimate the extent of sulfuric acid hydration in atmospheric conditions. Sulfuric acid - water clusters have previously been studied at the HF (with a moderately small basis set) level by Kurdi and Kochanski [1989], at the BLYP (with a plane-wave basis set) level by Arstila *et al.* [1998], at the B3LYP/6-311++G(2d,2p) level by Bandy and Ianni [1998] and Ianni and Bandy [2000], at the B3LYP/D95++(d,p) level by Re *et al.* [1999], at the PW91/DNP level by Ding *et al.* [2003b, 2004] and at the PW91/ATZ2P level by Al Natsheh *et al.* [2004]. However, the earliest HF and B3LYP - level studies, in contrast to the experiments, predicted that sulfuric acid remains unhydrated in atmospheric conditions. Later studies at the PW91 level qualitatively predicted the formation of hydrates, but did not correctly replicate the experimentally measured extent of hydration. In **Paper I**, we have shown by a series of test calculations that the discrepancy between earlier studies and experimental results are mainly related to the B3LYP functional, which tends to underpredict the binding energies of these hydrogen-bonded clusters compared to higher-level methods. The same has been noted

by Al Natsheh *et al.* [2004] and Nadykto and Yu [2007]. Various basis-set effects and vibrational anharmonicity play smaller but non-negligible roles. In **Paper III**, we have shown that a combination of high-level correlation (at the MP4 level) and large-basis set (aug-cc-pV(T+d)Z) energy corrections with anharmonic frequency calculations at the MP2/aug-cc-pV(D+d)Z level can quantitatively replicate the experimental thermochemical parameters for the very smallest sulfuric acid - water clusters. This is illustrated in Figure 2, which compares the predictions of various computational methods regarding the fraction of sulfuric acid molecules bound to 0, 1 or 2 water molecules at 298.15 K and a relative humidity (RH) of 50%.

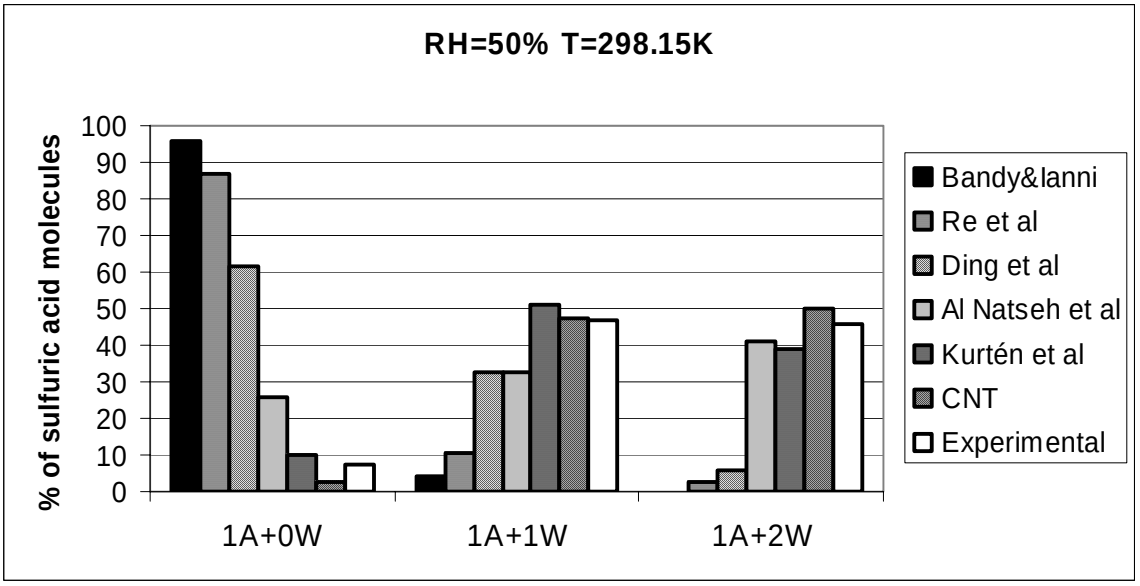

Figure 2. The fraction of sulfuric acid molecules bound to 0 (1A+0W), 1 (1A+1W) or 2 (1A+2W) water molecules, as predicted by different quantum chemical methods (Kurtén *et al.* refers to **Paper III**, see text for other references), classical nucleation theory (CNT; Noppel *et al.* [2002] with Clegg *et al.* [1998] activity model) and experiments [Hanson and Eisele 2000].

Laboratory experiments [Ball *et al.* 1999] have shown that the presence of ammonia in atmospherically representative concentrations has a clear and moderately strong nucleation-enhancing effect in the sulfuric acid - water system. However, different theoretical approaches have yielded very different predictions for the strength of this effect. The first applications of CNT significantly overpredicted the enhancing effect [Napari *et al.* 2002], while updated CNT-based models and state-of-the-art thermodynamics [Anttila *et al.* 2005] predict modest enhancement, in line with the experimental results. In terms of the average NH_3:H_2SO_4 mole ratio of nucleating clusters in some atmospherically realistic conditions (*i.e.* temperatures and partial pressures), CNT and thermodynamics have predicted ratios around 1:1 or even 2:1. On the other hand, earlier quantum chemical studies of clusters containing one sulfuric acid, one ammonia molecule and multiple water molecules (at the B3LYP/6-311++G(2d,2p) level by Ianni and Bandy [1999] and at the MP2/6-311++G(d,p) level by Larson *et al.* [1999]) predicted ratios close to zero, corresponding to no nucleation enhancement by ammonia.

(Larson *et al.* did claim to predict the formation of ammonium bisulfate in the atmosphere, but using formation free energies computed from their reported electronic energies, rotational constants and vibrational frequencies together with atmospherically realistic temperatures and partial pressures for ammonia still leads to very low NH_3:H_2SO_4 mole ratios for the one-acid clusters.) This discrepancy is partially related to the underprediction of binding energies by the B3LYP functional noted *e.g.* in **Paper I**. However, the most important parameter for assessing the role of ammonia in sulfuric acid - water nucleation is not the binding energy of ammonia to the cluster in itself, but the difference in binding energies of the $H_2SO_4 \bullet NH_3$ and $H_2SO_4 \bullet H_2O$ clusters (and their more extensively hydrated counterparts). As discussed in **Paper I,** the differences between different methods tend to cancel out when relative energetics are calculated, so for example the shortcomings of B3LYP are not likely to explain this particular controversy.

More recent studies [**Paper II**, **Paper IV**, **Paper VI**, Nadykto and Yu 2007] have demonstrated that the main drawback of earlier computational studies on ammonia-containing clusters was related to the limitations of the cluster dataset rather than the quantum chemical method. Computations on clusters containing two [**Paper II**, **Paper IV**, Nadykto and Yu 2007] or three [**Paper VI**] sulfuric acid molecules show that the effect of ammonia (indicated *e.g.* by the difference in formation energies of ammonia-containing and ammonia-free clusters of the same molecularity) increases with the number of acids in the cluster. As discussed in **Papers II** and **IV**, ammonia and water can be thought to compete for the same acid molecules. The concentration of water in the atmosphere is around 6-10 orders of magnitude higher than the concentration of ammonia [Seinfeld and Pandis 1998]. Therefore, if ammonia-containing clusters are to dominate the cluster distribution (as indicated by the experimental results), it is not enough for the binding of ammonia to the clusters just to be slightly stronger than that of water. Instead, as demonstrated by equation (3.54), it has to be stronger by at least a factor of $k_B T \times \ln(P_{H_2O}/P_{NH_3})$, where P_{NH_3} and P_{H_2O} are the partial pressures of ammonia and water, respectively. This condition is not fulfilled at all for the one-acid clusters, is fulfilled only in some rare conditions for the two-acid clusters, but is probably fulfilled in almost all conditions for the three-acid clusters. (The remaining uncertainty is related to the effects of full hydration, which has not been studied for the three-acid clusters in **Paper VI**.) In other words, ammonia is expected to significantly assist the growth of two-acid clusters to three-acid clusters in ambient conditions. In terms of the NH_3:H_2SO_4 mole ratio, this implies a lower limit of around 1:3.

In **Paper IV**, the role of ammonia in sulfuric acid - water nucleation was approached from another direction by estimating an upper limit to the NH_3:H_2SO_4 mole ratio of nucleating clusters in atmospheric conditions. This was made possible by the observation [Ianni and Bandy 1999, **Paper II**, Nadykto and Yu 2007] that hydration of sulfuric acid - ammonia clusters affects the acid - ammonia binding relatively weakly, and also more or less systematically, with the addition of water molecules tending to weaken the binding of ammonia to the clusters. NH_3:H_2SO_4 mole ratios computed as weighted averages from cluster distributions of "core clusters" containing only sulfuric acid and ammonia molecules thus represent an upper limit to the real atmospheric mole ratios. Leaving out

the water molecules from the simulations decreased the computational requirements significantly, and the formation free energies for the core clusters containing two sulfuric acid and up to four ammonia molecules could be computed at the relatively high level RI-MP2/aug-cc-pV(T+d)Z//RI-MP2/aug-cc-pV(D+d)Z. Possible systematic errors were accounted for by a sensitivity analysis, in which the results of **Papers I** and **Paper III** were used estimate the maximal effects of neglect of higher-order correlation or anharmonicity. The computed cluster distributions indicate that the NH_3:H_2SO_4 mole ratios are unlikely to exceed 1:1 anywhere in the free atmosphere. However, much larger mole ratios (around or even above 2:1) are routinely measured for larger particles. This indicates that the chemical composition of nucleating particles may differ significantly from the over - 10 nanometer - particles on which measurements can be performed.

Experimental evidence [*e.g.* O'Dowd *et al.* 2002b, Zhang *et al.* 2004, Surratt *et al.* 2007] indicates that organic molecules, possibly together with sulfuric acid, may be involved in atmospheric nucleation in some areas. In urban environments, organic products of combustion processes (*e.g.* from traffic) may account for a large fraction of the measured particle mass. At the other extreme, nucleation over background rural sites may involve biogenic vapors emitted by vegetation. Countless quantum chemical studies have been performed on gas-phase organic reactions potentially relevant to particle formation, but few of these have focused specifically on nucleation mechanisms, and even fewer have considered the simultaneous involvement of sulfuric acid and organics. The first such study was performed by Zhang *et al.* [2004], who investigated complexes of sulfuric acid with several aromatic acids formed in combustion processes in order to help explain the results of their experimental nucleation studies. Using a combination of B3LYP geometries and CCSD(T) energy corrections, Zhang *et al.* found the binding energies of these complexes to be remarkably high, and concluded that such clusters may play a role in pollution-related nucleation events. Nadykto and Yu [2007] recently reported PW91/6-311++G(3df,3pd) level results on complexes of sulfuric acid with formic acid and acetic acid. The formation free energies of complexes was found to be comparable to that of sulfuric acid - ammonia clusters, indicating that the acids are quite strongly bound to each other. While hydrogen-bonded sulfuric acid - organic acid clusters may play a role in the atmosphere, some experimental studies [Surratt *et al.* 2007] indicate that sulfuric acid - organic nucleation mechanisms may involve real chemical reactions as opposed to simple clustering. Such a mechanism is studied in **Paper V**, in which the reactions of stabilized Criegee Intermediates (sCIs; formed in the ozonolysis of biogenic terpenes) with sulfuric acid and water molecules are compared using a combination of B3LYP and RI-CC2 methods. While the results of **Paper V** are only preliminary order-of-magnitude estimates, and their atmospheric relevance depends on the still-unknown atmospheric lifetimes of the biogenic sCIs, the study demonstrates that quantum chemical calculations are powerful tools for exploring nucleation mechanisms involving complicated chemistry.

5. Review of papers

This thesis consists of five articles published in peer-reviewed journals and one article accepted for publication in a peer-reviewed journal. **Papers I, II, IV** and **VI** study sulfuric acid - water - ammonia clusters, while **Paper III** focuses on a comparison of hydrated HSO_4^- and H_2SO_4, and **Paper V** presents a hypothetical mechanism for organic - sulfuric acid nucleation.

Paper I is an error analysis of previous studies on sulfuric acid - water - ammonia clusters. The energetics for the formation of sulfuric acid monohydrate and ammonium hydrogensulfate clusters predicted by various quantum chemical methods and basis sets are compared, and the effect of anharmonic vibrational corrections at the B3LYP level is briefly studied. The discrepancies between previous theoretical and experimental studies concerning the hydration of sulfuric acid are shown to be caused mainly by too low binding energies predicted by the B3LYP density functional. Topological analysis of the electron density (within the Atoms In Molecules framework) is used to identify different bonding patterns and also construct a proton transfer parameter which predicts how different methods will perform for larger clusters

Paper II is a qualitative (PW91/DNP level) study of the energetics of sulfuric acid - water - ammonia clusters containing up to two acid, one ammonia and seven water molecules. The computed formation free energies are used to calculate cluster distributions in various conditions (temperatures and monomer vapor concentrations). The effect of ammonia on the cluster distribution is seen to increase with the number of acid molecules, though the absolute number of ammonia-containing clusters is still very low in almost all atmospherically relevant conditions. This is explained in terms of mass-balance effects due to the large ratio of the partial pressures of water and ammonia in atmospheric conditions.

Paper III is an in-depth, very high-level study of the energetics of $(H_2SO_4)(H_2O)_{1...4}$ and $(HSO_4^-)(H_2O)_{1...4}$ clusters. The study demonstrates that a combination of MP2/aug-cc-pV(D+d)Z geometries, high-level correlation (MP4) and large basis-set (aug-cc-pV(T+d)Z) corrections, perturbative anharmonic vibrational frequency calculations and hindered rotor analysis can quantitatively replicate experimental formation enthalpies and entropies for very small clusters. However, due to the large computational effort involved, the method could not be directly applied to clusters larger than $H_2SO_4 \bullet (H_2O)_2$ or $HSO_4^- \bullet (H_2O)_2$, and scaling-factor based extensions failed severely for the $HSO_4^- \bullet (H_2O)_4$ cluster, presumably due to unidentified hindered rotations. The role of ammonia in ion-induced nucleation around the HSO_4^- core ion is predicted to be small.

Paper IV is a high-level (RI-MP2/aug-cc-pV(T+d)Z//RI-MP2/aug-cc-pV(D+d)Z) study of a relatively small set of clusters containing two sulfuric acid molecules and 0-4 ammonia molecules. Based on the results of earlier studies (such as **Papers I-III**), formation free energies computed for these clusters can be used to estimate an upper limit for the NH_3:H_2SO_4 mole ratio of small clusters in atmospheric conditions. The possible influence of various error sources is taken into account using a sensitivity analysis. The

results indicate that the NH_3:H_2SO_4 mole ratio of small clusters is very unlikely to exceed 1:1 in any conditions encountered in the atmosphere. The possible reasons for the predicted differences between the chemical compositions of small and large clusters (for which mole ratios of 2:1 are often measured) are discussed.

Paper V is a preliminary study of the reaction of biogenic stabilized Criegee Intermediates (sCI) with sulfuric acid. Criegee Intermediates are biradicals formed in the ozonolysis of terpenes, and experimental evidence indicates that their reactions play a significant role in particle formation from biogenic vapors. Thermodynamic parameters are computed at the B3LYP/6-311+G(2d,p) level for the reaction of sulfuric acid with five different sCIs (three model species and two large biogenic molecules), and these are compared to the corresponding parameters for the reaction with water, believed to be the main sink reaction. However, as all reactions are strongly exothermic, thermodynamics alone does not yield information on the atmospheric relevance of the reaction. Activation barrier and intrinsic reaction co-ordinate calculations (using the B3LYP and RI-CC2 methods) on the model species $(CH_3)_2COO$ indicate that while the sink reaction with water has a moderately high barrier, the reaction with sulfuric acid is essentially barrierless. If reaction with water is indeed the main atmospheric sink of biogenic sCIs, the reaction with sulfuric acid could potentially explain some part of the observed new-particle formation events for example in boreal forests. However, the atmospheric importance of the reaction depends crucially on the lifetimes of the biogenic sCIs, which are still unknown.

Paper VI extends the analysis of **Paper II** by including clusters containing three sulfuric acid molecules. Geometry optimizations and frequency calculations are carried out at the MPW1B95/aug-cc-pV(D+d)Z level, and the final electronic energies are computed at the RI-MP2/aug-cc-pV(T+d)Z level. The results confirm the hypothesis of **Paper II** that the binding of ammonia to the clusters is strengthened by the addition of further sulfuric acid molecules. As the effect of full hydration is not modeled due to computational reasons, cluster size distributions can not be directly computed from the data. However, analysis of the formation free energies in atmospheric conditions strongly indicates that the three-acid cluster size distribution, unlike the one- and two-acid distributions, is likely to be dominated by ammonia-containing clusters. This confirms the experimental and classical thermodynamical result that ammonia lowers the free energy barrier for sulfuric acid - water nucleation in atmospheric conditions, and also tentatively implies a lower limit of 1:3 for the NH_3:H_2SO_4 mole ratio.

Author's contribution

I am alone responsible for the summary of the thesis.

In **Paper I**, I was responsible for planning and carrying out the computations as well as most of the analysis of the results and the writing itself. In **Paper II**, the calculations had already been done by others, but I was responsible for the majority of the data analysis as well as almost all of the writing. In **Paper III**, I carried out most of the planning as well as all of the quantum chemical calculations, performed a major part of the data analysis

(with the exception of the scaling factor and anharmonic thermochemistry calculations) and was responsible for most of the writing. In **Paper IV,** I was responsible for the original idea (of computing an upper ratio to the NH_3:H_2SO_4 ratio using unhydrated clusters) as well as the final, published quantum chemical computations and most of the writing. In **Paper V**, I was responsible for the planning and carrying out of the quantum chemical computations, almost all of the data analysis and a major part of the writing. In **Paper VI**, I was responsible for part of the initial planning (*e.g.* constructing initial input structures), a moderate part of the actual calculations (mainly, the final RI-MP2 energy corrections), a major part of the data analysis and a significant part of the writing.

Errata

In **Paper I**, the AIM parameters computed at the MP2 level actually correspond to the Hartree-Fock level. This is not very clearly indicated in the Gaussian 03 manual, but came up in discussions with Gaussian, Inc representatives. The conclusions drawn about the behavior of MP2 with respect to proton transfer and bonding patterns are therefore not valid.

6. Conclusions

The work presented in this thesis demonstrates that quantum chemistry is a powerful tool to study the fundamental, molecular-level processes related to nucleation. However, care must be taken in the planning of studies, as calculations performed with inaccurate methods or with an insufficient dataset (with regard to, for example, the types of clusters included in the study) can easily lead to erroneous conclusions. The role of quantum chemistry in nucleation studies is likely to increase in the future, as the focus of investigations shifts toward chemically complicated mechanisms involving *e.g.* organics or radical intermediates of the SO_2 oxidation chain. However, the restrictions on configurational sampling caused by the high computational effort of the calculations mean that quantum chemical methods will not in the foreseeable future be able to replace computationally cheaper force-field-based methods for the modeling of dynamic processes related to nucleation. Instead, more work is needed to combine the best features of the two approaches, with the ultimate objective being a simulation that is accurate both with respect to molecular interaction energies and configurational sampling.

The main results of this thesis are the following:

- The discrepancies between earlier quantum chemical studies and experiments concerning the hydration of sulfuric acid were shown to be caused mainly by errors in the electronic energies, with basis-set effects and vibrational anharmonicity also playing moderately important roles. State-of-the-art quantum chemical calculations including both high-order correlation, large basis sets and advanced anharmonic thermochemical modeling were shown to quantitatively replicate the experimental results on sulfuric acid hydration.

•The probable NH_3:H_2SO_4 mole ratio of nucleating clusters in most atmospheric conditions was determined to within a factor of 3. On a qualitative level, the role of ammonia in sulfuric acid - water nucleation is thus explained. The calculations reported here indicate that the presence of ammonia at atmospherically realistic partial pressures significantly assists the growth of two-acid clusters to three-acid clusters, implying a lower limit of 1:3 for the mole ratio. On the other hand, calculations on H_2SO_4•NH_3 core clusters indicates that the NH_3:H_2SO_4 mole ratio of the smallest molecular clusters is unlikely to exceed 1:1 in almost any atmospheric conditions. The latter result was shown to be insensitive even to quite large systematic errors in the computed free energies. The sensitivity analysis approach presented here is likely to be useful in future studies on similar issues.

•The HSO_4^- ion was shown to be much more extensively hydrated than neutral H_2SO_4, in accordance with experimental results. Preliminary calculations indicate that ammonia plays little or no role in ion-induced nucleation of the HSO_4^- core ion, contrary to previous speculations.

•The reaction between sulfuric acid and biogenic stabilized Criegee Intermediates was shown to be of potential atmospheric importance as a formation pathway of nucleation precursors. However, the true relevance of the reaction depends on the lifetime of the Criegee Intermediates, which still remains to be reliably determined.

7. References

F. F. Abraham: *Homogeneous Nucleation Theory - The Pretransition Theory of Vapor Condensation*, Academic Press, New York, U.S.A. (1974)

R. Ahlrichs, M. Bär, M. Häser; H. Horn and C. Kölmel: Electronic structure calculations on workstation computers: The program system Turbomole, Chem. Phys. Lett. **162**, 165-169 (1989).

J. M. Aitken: On Dust, Fogs and Clouds, Nature **23**, 195-197 (1881).

A. Al Natsheh, A. B. Nadykto, K. V. Mikkelsen, F. Yu and J. Ruuskanen: Sulfuric Acid and Sulfuric Acid Hydrates in the Gas Phase: A DFT Investigation, J. Phys. Chem. A **108**, 8914-8929 (2004); correction published in J. Phys. Chem. A. **110**, 7982 (2006).

T. Anttila, H. Vehkamäki, I. Napari and M. Kulmala: Effect of ammonium bisulphate formation on atmospheric water-sulphuric acid-ammonia nucleation, Boreal Env. Res. **10**, 511-523 (2005).

H. Arstila, K. Laasonen and A. Laaksonen: *Ab initio* study of gas-phase sulphuric acid hydrates containing 1 to 3 water molecules, J. Chem. Phys. **108**, 1031-1039 (1998).

P. W. Atkins and R. S. Friedman: *Molecular Quantum Mechanics*, 3rd edition, Oxford University Press, U.S.A. (1997).

E. J. Baerends and O. V. Gritsenko: A Quantum Chemical View of Density Functional Theory, J. Phys. Chem. A **101**, 5383-5403 (1997).

S. M. Ball, D. R. Hanson, F. L. Eisele and P. H. McMurry: Laboratory studies of particle nucleation: Initial results for H_2SO_4, H_2O, and NH_3 vapors, J. Geophys. Res. **D 104**, 23709-23718 (1999).

A. R. Bandy and J. C. Ianni: Study of Hydrates of H_2SO_4 Using Density Functional Theory, J. Phys. Chem. A **102**, 6533-6539 (1998).

V. Barone: Vibrational zero-point energies and functions beyond the harmonic approximation, J. Chem. Phys. **120**, 3059-3065 (2004).

V. Barone: Anharmonic vibrational properties by a fully automated second-order perturbative approach, J. Chem. Phys. **122**, 014108 (2005).

R. J. Bartlett and M. Musial: Coupled-cluster theory in quantum chemistry, Reviews of Modern Physics **79**, 291-352 (2007).

R. Becker and W. Döring: Kinetische Behandlung der Keimbildung in übersättigen Dämpfen, Ann. Phys. (Leipzig) **24**, 719-752 (1935).

R. E. Benestad: A review of the solar cycle length estimates, Geophys. Res. Lett. **32**, L15714, 2005.

T. Berndt, O. Böge, F. Stratmann, J. Heintzenberg and M. Kulmala: Rapid Formation of Sulfuric Acid Particles at Near-Atmospheric Conditions, Science **307**, 698-700 (2005).

T. Berndt, O. Böge and F. Stratmann: Formation of atmospheric H_2SO_4/H_2O particles in the absence of organics: A laboratory study, Geophys. Res. Lett. **33**, L15817 (2006).

S. F. Boys and F. Bernardi: The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies, Mol. Phys. **19**, 553-566 (1970).

B. Brunekreef and S. T. Holgate: Air pollution and health, The Lancet **230**, 1233-1242 (2002).

D. M. Ceperley and B. J. Alder: Ground State of the Electron Gas by a Stochastic Method, Phys. Rev. Lett. **45**, 566-569 (1980).

G. M. Chaban, J. O. Jung and R. B. Gerber: *Ab initio* calculation of anharmonic vibrational states of polyatomic systems: Electronic structure combined with self-consistent field, J. Chem. Phys. **111**, 1823-1829 (1999).

O. Christiansen: Vibrational coupled cluster theory, J. Chem. Phys. **210**, 2149-2159 (2004).

S. L. Clegg, P. Brimblecombe and A. S. Wexler: Thermodynamic Model of the System H^+ - NH_4^+ - SO_4^{2-} - NO_3^- - H_2O at Tropospheric Temperatures, J. Phys. Chem. A **102**, 2137-2154 (1998).

C.-G. Ding, T. Taskila, K. Laasonen and A. Laaksonen: Reliable potential for small sulfuric acid-water clusters, Chem. Phys. **287**, 7-19 (2003a).

C.-G. Ding, K. Laasonen and A. Laaksonen: Two Sulfuric Acids in Small Water Clusters, J. Phys. Chem. A **107**, 8648-8658 (2003b).

C.-G. Ding and K. Laasonen: Partially and fully deprotonated sulfuric acid in $H_2SO_4(H_2O)_n$ (n=6-9) clusters, Chem. Phys. Lett. **390**, 307-313 (2004).

B. I. Dunlap: Robust and variational fitting: Removing the four-center integrals from center stage in quantum chemistry, J. Mol. Struct (THEOCHEM) **529**, 37-40 (2000).

K. Eichkorn, O Treutler, H. Öm, M. Häser and R. Ahlrichs: Auxiliary basis sets to approximate Coulomb potentials, Chem. Phys. Lett. **242**, 652-660 (1995).

J. B. Foresman and Æ Frisch: *Exploring Chemistry with Electronic Structure Methods*, 2nd edition, Gaussian, Inc., Wallingford CT, U.S.A. (1996).

D. Frenkel and B. Smit: *Understanding Molecular Simulation - From Algorithms to Applications*, 2nd edition, Academic Press, London, U. K. (2002).

M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople: Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, U.S.A. (2004).

K. D. Froyd and E. R. Lovejoy: Experimental Thermodynamics of Cluster Ions Composed of H_2SO_4 and H_2O. 2. Measurements and *Ab Initio* Structures of Negative Ions, J. Phys. Chem. A **107**, 9812-9824 (2003).

O. V. Gritsenko, P. R. T. Schipper and E. J. Baerends: Exchange and correlation energy in density functional theory: Comparison of accurate density functional theory quantities

with traditional Hartree-Fock based ones and generalized gradient approximations for the molecules Li_2, N_2, F_2, J. Chem. Phys. **107**, 5007-5015 (1997).

B. Guillot and Y. Guissani: How to build a better pair potential for water, J. Chem. Phys. **114**, 6720-6733 (2001).

D. R. Hanson and F. Eisele: Diffusion of H_2SO_4 in humidified nitrogen: hydrated H_2SO_4, J. Phys. Chem. A **104**, 1715-1719 (2000).

P. Hari and M. Kulmala: Station for Measuring Ecosystem–Atmosphere Relations (SMEAR II), Boreal Env. Res. **10**, 315-322 (2005).

J. Heintzenberg, J. Birmili, A. Wiedensohler, A. Nowak and T. Tuch: Structure, variability and persistence of the submicrometer marine aerosol, Tellus **56B**, 357-367 (2004).

R. H. Heist and H. Reiss: Hydrates in supersaturated binary sulfuric acid - water vapor, J. Chem. Phys. **61**, 573-581 (1974).

T. Helgaker, W. Klopper, H. Koch and J. Noga: Basis-set convergence of correlated calculations on water, J. Chem. Phys. **106**, 9639-9646 (1997).

P. Hohenberg and W. Kohn: Inhomogeneous Electron Gas, Phys. Rev. **136**, B864-B871 (1964).

J. C. Ianni and A. R. Bandy: A Density Functional Theory Study of the Hydrates of $NH_3 \bullet H_2SO_4$ and Its Implications for the Formation of New Atmospheric Particles, J. Phys. Chem A. **103**, 2801-2811 (1999).

J. C. Ianni and A. R. Bandy: A theoretical study of the hydrates of $(H_2SO4)_2$ and its implications for the formation of new atmospheric particles, J. Mol. Struct. (THEOCHEM) **497**, 19-37 (2000).

K. Iida, M. Stolzenburg, P. McMurry, M. J. Dunn, J. N. Smith, F. Eisele and P. Keady: Contribution of ion-induced nucleation to new particle formation: Methodology and its application to atmospheric observations in Boulder, Colorado, J. Geophys. Res. **111**, D23201 (2006).

IPCC (Intergovernmental Panel on Climate Change): *Fourth Assesment Report on Climate Change*, Online version available at: http://ipcc-wg1.ucar.edu/wg1/wg1-report.html. The full report will be published by Cambridge Univ. Press, New York, U.S.A. (2007).

F. Jensen: *Introduction to Computational Chemistry*, 2nd edition, John Wiley&Sons Ltd., West Sussex, U. K. (2007).

J. Jung, P. J. Adams and S. N. Pandis: Regional Air Quality - Atmospheric Nucleation Interactions, *extended conference abstract published in*: C. D. O'Dowd and P. E. Wagner (eds.): *Nucleation and Atmospheric Aerosols*, 871–877, Springer (2007), *in pres*s.

D. Kashchiev: *Nucleation - Basic Theory With Applications*, Butterworth-Heinemann, Oxford, U. K. (2000).

S. M. Kathmann and B. N. Hale: *Monte Carlo* simulations of small sulfuric acid - water clusters, J. Phys. Chem B. **105**, 11719-11728 (2001).

S. Kathmann, G. Schenter and B. Garrett: The Critical Role of Anharmonicity in Aqueous Ionic Clusters Relevant to Nucleation, J. Phys. Chem. C **111**, 4977-4983 (2007).

S. von Klot, A. Peters, P. Aalto, T. Bellander, N. Berglind, D. D'Ippoliti, R. Elosua, A. Hörmann, M. Kulmala, T. Lanki, H. Löwel, J. Pekkanen, S. Picciotto, J. Sunyer, F. Forastriere and the HEAPSS study group: Ambient Air Pollution Is Associated With Increased Risk of Hospital Cardiac Readmissions of Myocardial Infarction Survivors in Five European Cities, Circulation **112**, 3073-3079 (2005).

W. Kohn and L. J. Sham: Self-Consistent Equations Including Exchange and Correlation Effects, Phys. Rev. **140,** A1133-A1138 (1965).

W. Kohn, A. D. Becke and R. G. Parr: Density Functional Theory of Electronic Structure, J. Phys. Chem. **100**, 12974-12980 (1996).

M. Kulmala, L. Pirjola and J. M. Mäkelä: Stable sulphate clusters as a source of new atmospheric particles, Nature **404**, 66-69 (2000).

M. Kulmala: How Particles Nucleate and Grow, Science **302**, 1000-1001 (2003).

M. Kulmala, H. Vehkamäki, T. Petäjä, M. Dal Maso, A. Lauri, V.-M. Kerminen, W. Birmili and P. H. McMurry: Formation and growth rates of ultrafine atmospheric particles: a review of observations, J. Aerosol Sci. **35**, 143-176 (2004a).

M. Kulmala, V.-M. Kerminen, T. Anttila, A. Laaksonen and C. D. O'Dowd: Organic aerosol formation via sulphate cluster activation, J. Geophys. Res. **109**, D04205 (2004b).

M. Kulmala, K. E. J. Lehtinen, L. Laakso, G. Mordas and K. Hämeri: On the existence of neutral atmospheric clusters, Boreal Env. Res. **10**, 79-87 (2005).

L. Kurdi and E. Kochanski: Theoretical studies of sulfuric acid monohydrate: neutral or ionic complex, Chem. Phys. Lett. **158**, 111-115 (1989).

T. Kurtén, M. Kulmala, M. Dal Maso, T. Suni, A. Reissell, H. Vehkamäki, P. Hari, A. Laaksonen, Y. Viisanen and T. Vesala: Estimation of different forest-related

contributions to the radiative balance using observation in southern Finland, Boreal Env. Res. **8**, 275-285 (2004).

T. Kurtén and H. Vehkamäki: Investigating atmospheric sulfuric acid- water –ammonia particle formation using quantum chemistry, *manuscript accepted for publication* in Adv. Quantum Chem. (2007).

I. Kusaka, Z.-G. Wang and J. Seinfeld: Binary nucleation of sulfuric acid-water: *Monte Carlo* simulation, J. Chem. Phys. **108**, 6829-6847 (1998).

L. Laakso, T. Anttila, K. E. J. Lehtinen, P. P. Aalto, M. Kulmala, U. Hõrrak, J. Paatero, M. Hanke and F. Arnold: Kinetic nucleation and ions in boreal particle formation events, Atmos. Chem. Phys. **4**, 2353-2366 (2004).

L. Laakso, T. Grönholm, L. Kulmala, S. Haapanala, A. Hirsikko, E. R. Lovejoy, J. Kazil, T. Kurtén, M. Boy, E. D. Nilsson, A. Sogachev, I. Riipinen, F. Stratmann and M. Kulmala: Hot-air balloon measurements of vertical variation of boundary layer new particle formation, Boreal Env. Res. **12**, 279-294 (2007).

A. Laaksonen and I. Napari: Breakdown of the Capillary Approximation in Binary Nucleation: A Density Functional Study, J. Phys. Chem. B 105, 11678-11682 (2001).

L. J. Larson, A. Largent and F.-M. Tao: Structure of the Sulfuric Acid - Ammonia System and the Effect of Water Molecules in the Gas Phase, J. Phys. Chem. A **103**, 6786-6792 (1999).

A. Lauri, J. Merikanto, E. Zapadinsky and H. Vehkamäki: Comparison of Monte Carlo simulation methods for the calculation of the nucleation barrier of argon, Atmos. Res. **82**, 489-502 (2006a).

A. Lauri, E. Zapadinsky, H. Vehkamäki and M. Kulmala: Comparison between the classical theory predictions and molecular nucleation results for heterogeneous nucleation of argon, J. Chem. Phys. **125**, 164712 (2006b).

A. R. Leach: *Molecular Modelling - Principles and Applications*, 2nd edition, Pearson Education Ltd., Harlow, U.K. (2001).

S.-H. Lee, J. M. Reeves, J. C. Wilson, D. E. Hunton, A. A. Viggiano, T. M. Miller, J. O. Ballenthin and L. R. Lait: Particle formation by Ion Nucleation in the Upper Troposphere and Lower Stratosphere, Science **301** 1886-1889 (2003).

B. Lindler.: *Thermodynamics and Introductory Statistical Mechanics,* John Wiley & Sons, Hoboken, U.S.A. (2004).

E. R. Lovejoy, J. Curtius and K. D. Froyd: Atmospheric ion-induced nucleation of sulfuric acid and water, J. Geophys. Res. **109**, D08204 (2004).

P. H. McMurry: A review of atmospheric aerosol measurements, Atmos. Environ. **34**, 1959-1999 (2000).

J. Merikanto, H. Vehkamäki and E. Zapadinsky: Monte Carlo simulations of critical cluster sizes and nucleation rates of water, J. Chem. Phys. **121**, 914-924 (2004).

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth and A. H. Teller: Equation of State Calculations by Fast Computing Machines, J. Chem. Phys. **21**, 1087-1092 (1955).

P. M. Morse: Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels, Phys. Rev. **34**, 57-64 (1929).

C. Møller and M. S. Plesset: Note on an Approximation Treatment for Many-Electron Systems, Phys. Rev. **46**, 618-622 (1934).

A. B. Nadykto, H. Du and F. Yu: Quantum DFT and DF–DFT study of vibrational spectra of sulfuric acid, sulfuric acid monohydrate, formic acid and its cyclic dimer, Vibrational Spectroscopy, *in press, corrected proof, available online* (2007).

A. B. Nadykto and F. Yu: Strong hydrogen bonding between atmospheric nucleation precursors and common organics, Chem. Phys. Lett. **435**, 14-18 (2007).

I. Napari, M. Noppel, H. Vehkamäki and M. Kulmala: An improved model for ternary nucleation of sulfuric acid – ammonia – water, J. Chem. Phys. **116**, 4221-4227 (2002).

D. G. Nash, T. Baer and M. V. Johnson: Aerosol mass spectrometry: An introductory review, Int. J. of Mass Spectrometry **258**, 2-12 (2006).

M. Noppel, H. Vehkamäki and M. Kulmala: An improved model for hydrate formation in sulfuric acid-water nucleation, J. Chem. Phys. **116**, 218-228 (2002).

C. D. O'Dowd, J. L. Jimenez, R. Bahreini, R. C. Flagan, J. H. Seinfeld, K. Hämeri, L. Pirjola, M. Kulmala, S. G. Jennings and T. Hoffman: Marine aerosol formation from biogenic iodide emissions, Nature **417**, 632-636 (2002a).

C. D. O'Dowd, P. Aalto, K. Hämeri, M. Kulmala and T. Hoffman: Atmospheric particles from organic vapours, Nature **416**, 497-498 (2002b).

J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria and G. I. Csonka: Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits, J. Chem. Phys. **123**, 062201 (2005).

J. Pesonen and L. Halonen: Recent advances in the theory of vibration-rotation Hamiltonians, Adv. Chem. Phys. **125**, 269-349 (2003).

L. Pirjola, K. E. J. Lehtinen, H.-C. Hansson and M. Kulmala: How important is nucleation in regional/global modelling?, Geophys. Res. Lett. **31**, L12109 (2004).

C. A. Pope III and D. W. Dockery: Health effects of fine particulate air pollution - lines that connect, J. Air Waste Manage Assoc **56**, 709-742 (2006).

T. Rajamäki, M. Kallay, J. Noga, P. Valiron and L. Halonen: High excitations in coupled-cluster series: vibrational energy levels of ammonia, Mol. Phys. **102**, 2297-2310 (2004).

S. Re, Y. Osamura and K. Morokuma: Coexistence of Neutral and Ion-Pair Clusters of Hydrated Sulfuric Acid $H_2SO_4(H_2O)_n$ (n= 1-5) - A Molecular Orbital Study, J. Phys. Chem. A **103**, 3535-3547 (1999).

I. Riipinen, S.-L. Sihto, M. Kulmala, F. Arnold, M. Dal Maso, W. Birmili, K. Saarnio, K. Teinilä, V.-M. Kerminen, A. Laaksonen and K. E. J. Lehtinen: Connections between atmospheric sulphuric acid and new particle formation during QUEST III-IV campaigns in Heidelberg and Hyytiälä, Atmos Chem Phys. **7**, 1899-1914 (2007).

T. Schwabe and S. Grimme S: Towards chemical accuracy for the thermodynamics of large molecules: new hybrid density functionals including non-local correlation effects, Phys. Chem. Chem. Phys. **8**, 4398-4401 (2006).

A. P. Scott and L. Radom: Harmonic Vibrational Frequencies: An Evaluation of Hartree-Fock, Møller-Plesset, Quadratic Configuration Interaction, Density Functional Theory, and Semiempirical Scale Factors, J. Phys. Chem. **100**, 16502-16513 (1996).

J. H. Seinfeld and S. N. Pandis: *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, Wiley&Sons, New York, U.S.A. (1998).

S.-L. Sihto, M. Kulmala, V.-M. Kerminen, M. Dal Maso, T. Petäjä, I. Riipinen, H. Korhonen, F. Arnold, R. Janson, M. Boy, A. Laaksonen and K. E. J. Lehtinen: Atmospheric sulphuric acid and aerosol formation: implications from atmospheric measurements for nucleation and early growth mechanisms, Atmos. Chem. Phys. **6**, 4079-4091 (2006).

D. V. Spracklen, K. S. Carslaw, M. Kulmala, V.-M. Kerminen, G. W. Mann, and S.-L. Sihto: The contribution of boundary layer nucleation events to total particle concentrations on regional and global scales, Atmos. Chem. Phys. **6**, 5631-5648 (2006).

P. Stier, J. Feichter, S. Kinne, S. Kloster, E. Vignati, J. Wilson, L. Ganzeveld, I. Tegen, M. Werner, Y. Balkanski, M. Schultz, O. Boucher, A. Minikin and A. Petzold: The aerosol-climate model ECHAM5-HAM, Atmos. Chem. Phys. **5**, 1125-1156 (2005).

J. D. Surratt, J. H. Kroll, T. E. Kleindienst, E. O. Edney , M. Claeys, A. Sorooshian, N. L. Ng, J. H. Offenberg, M. Lewandowski, M. Jaoui, R. C. Flagan and J. H. Seinfeld:

Evidence for Organosulfates in Secondary Organic Aerosol, Environ. Sci. Technol. **41**, 517-527 (2007).

H. Svensmark: Influence of Cosmic Rays on Earth's Climate, Phys. Rev. Lett. **81**, 5027-5030 (1998).

P. Tunved, H.-C. Hansson, V.-M. Kerminen, J. Ström, M. Dal Maso, H. Lihavainen, Y. Viisanen, P. P. Aalto, M. Komppula and M. Kulmala: High Natural Aerosol Loading over Boreal Forests, Science **312**, 261-263 (2006).

H. Vehkamäki, I. Napari, M. Kulmala and M. Noppel: Stable ammonium bisulphate clusters in the atmosphere, Phys. Rev. Lett. **93**, 148501 (2004).

H. Vehkamäki: *Classical Nucleation Theory in Multicomponent Systems*, Springer-Verlag, Berlin-Heidelberg, Germany (2006).

R. J. Weber, P. H. McMurry, R. L. Mauldin III, D. J. Tanner, F. L. Eisele, A. D. Clarke and V. N. Kapustin: New particle formation in the remote troposphere: A comparison of observations at various sites, Geophys. Res. Lett. **26**, 307-310 (1999).

F. Weigend and M. Häser: RI-MP2: first derivatives and global consistency, Theor. Chem. Acc. **97**, 331-340, (1997).

A. Wiedensohler, H. C. Hansson, D. Orsini, M. Wendish, F. Wagner, K. N. Bower, T. W. Chourlarton, M. Wells, M. Parkin, K. Acker, W. Wieprecht, M. C. Facchini, J. A. Lind, S. Fuzzi, B. G. Arends and M. Kulmala: Night-time formation and occurrence of new particles associated with orographic cloud, Atmos. Environ. **31**, 2545-2559 (1997).

E. B. Wilson Jr., J. C. Decius and P. C. Cross: *Molecular Vibrations*, McGraw-Hill, New York, U.S.A. (1955).

F. Yu and R. P. Turco: Ultrafine aerosol formation via ion-mediated nucleation, Geophys. Res. Lett. **27**, 883-886 (2000).

F. Yu: Effect of ammonia on new particle formation: a kinetic H_2SO_4-H_2O-NH_3 nucleation model constrained by laboratory measurements, J. Geophys. Res. **111**, D01204 (2006).

R. Zhang, I. Suh, J. Zhao, D. Zhang, E. C. Fortner, X. Tie, L. T. Molina, M. J. Molina: Atmospheric New Particle Formation Enhanced by Organic Acids, Science **304**, 1487-1490 (2004).

J. Zeldovich: On the theory of new phase formation; cavitation, Acta Physiochimica URSS XVIII, 1/22/01 (1943).

Y. Zhao and D. G. Truhlar: Hybrid Meta Density Functional Theory Methods for Thermochemistry, Thermochemical Kinetics, and Noncovalent Interactions: The MPW1B95 and MPWB1K Models and Comparative Assessments for Hydrogen Bonding and van der Waals Interactions, J. Phys. Chem. A. **108**, 6908-6918 (2004).

8. List of symbols and variables

When possible, the use of a single symbol to denote multiple different variables, parameters or operations has been avoided. However, some conventions are so well-established that a change of notation would only serve to increase confusion. Thus, *e.g.* "*n*" denotes both the number of molecules in classical nucleation theory, and the primary quantum number in quantum mechanics, while *V* denotes both potential energy and volume. Similarly, the superscript "*" is used to denote both critical cluster properties and complex conjugation. Dummy variables for summation (*e.g. a, b, c, d, i, j*) are used in several instances as there are more summations in this thesis than there are letters in the alphabet.

A: surface area of a droplet
A, B: functional groups
$A_{i,g}$: gas-phase activity of compound i
$A_{i,l}(x_{i,l})$: liquid-phase activity of compound i
a_s: the coefficient of Slater determinant number s
$c_{i,j}$: orbital expansion coefficients
$\mathbf{C}$: matrix of orbital expansion coefficients
$\mathbf{D}$: reduced density matrix
$\mathbf{E}$: diagonal matrix of orbital energies
E: energy
E_0: ground-state electronic energy
E_i: energy level i
E_{MP2}: MP2 energy
$E[\rho(\mathbf{r})]$: the total energy as a functional of the electron density
$E_H[\rho(\mathbf{r})]$: Coulomb interaction energy functional
$E_{XC}[\rho(\mathbf{r})]$: exchange-correlation energy functional
F: Helmholtz free energy
$F_{A,x}$: the (Hellman-Feynman) force acting on nuclei A in the direction x.
$\mathbf{F}$: Fock matrix
f_j: basis function
G, G_{tot}: Gibbs free energy
g_i: degeneracy of the i:th energy level
$g_{0,el}$: degeneracy of the ground-state electronic wavefunction
$\hat{g}$: two-electron operator
h: Planck's constant, 6.626068×10^{-34} m^2kgs^{-1}
H_{class}: classical Hamiltonian
$\hat{H}$: quantum-mechanical Hamiltonian operator

$\hat{H}_{HF}$: Hartree-Fock Hamiltonian operator

$\hat{H}_{M-P}$: Møller-Plesset perturbation theory Hamiltonian operator

H_{tot}: total enthalpy

H_X: enthalpy contribution from degrees of freedom of type X

$\hat{h}$: one-electron operator

I_1, I_2, I_3: moments of inertia about the principal axes

J: nucleation rate

K: number of atomic nuclei

$k_{A,B,r}$, $k_{A,B,\theta}$, $k_{A,B,i}$: force-field constants related to the bond stretching, bending and torsional motions (respectively) of functional groups A and B

k_B: Bolzmann constant, $1.3806503 \times 10^{-23}$ $m^2kgs^{-2}K^{-1}$

l: angular momentum quantum number

m_{basis}: number of basis functions

M: mass of a molecule

m: magnetic quantum number

n_i: number of molecules of type i

$n_{i,l}$: number of liquid (core) molecules of type i

$n_{i,s}$: number of surface molecules of type i

n: number of molecules in sections 3.1 and 3.2, the primary quantum number in section 3.3.

N: number of electrons

N_c: a normalization constant

P: pressure

$\mathbf{p}_i$**:** momentum vector of particle i

Q_A, Q_B: charges or partial charges of functional groups or nuclei A and B

q: partition function

q_{el}: partition function of electronic degrees of freedom

q_{rot}: partition function of rotational degrees of freedom

q_{trans}: partition function of translational degrees of freedom

q_{vib}: partition function of vibrational degrees of freedom

q_{NVT}: partition function in the NVT ensemble

r: radius. In section 3.1, r denotes the radius of a droplet, while in section 3.3 it denotes the distance from an atomic nuclei.

$\mathbf{r}_i$: position vector of particle i

$\mathbf{R}_i$: position vector of nucleus i

$\mathbf{r}_{el}$: position vector containing the positions of all electrons

$\mathbf{r}_{nuc}$: position vector containing the positions of all nuclei

r,r': integration variables corresponding to electronic position vectors

$r_{A,B}$: distance between functional groups A and B

r_0: equilibrium bond distance

S: saturation ratio

S_{tot}: total entropy

S_X: entropy contribution from degrees of freedom of type X

S: overlap matrix

$\hat{T}_{\acute{e}}$: electronic kinetic energy operator

$\hat{T}_n$: nuclear kinetic energy operator

$\hat{T}$: cluster operator in coupled-cluster theory

$\hat{T}_i$: operator corresponding to *i*:th level excitations in coupled-cluster theory

$T[\rho(\mathbf{r})]$: kinetic energy functional

T: temperature (in Kelvin)

V: potential energy *or* volume

$\hat{V}_{ne}$: potential energy operator corresponding to interactions between nuclei and electrons

$\hat{V}_{ee}$: potential energy operator corresponding to interactions between electrons

$\hat{V}_{nn}$ potential energy operator corresponding to interactions between nuclei

$V(\mathrm{A},\mathrm{B})_X$: force field potential for the interactions of type *X* between functional groups A and B

V_{Nuc}: potential energy term containing all interactions between atomic nuclei

$V_{ext}(\mathbf{r})$: external potential

$v_{XC}(\mathbf{r})$: exchange-correlation functional

$v_{i,l}$: molecular volume of compound *i* corresponding to the bulk liquid density

x: a co-ordinate axis in section 3.3.

$x_{i,l}$: mole fraction of compound *i* in the liquid phase.

$Y_{l,m}(\theta,\varphi)$: spherical harmonic function corresponding to quantum numbers *l* and *m*

$\Delta\varphi^*$: free energy change in critical cluster formation

ΔG^* : Gibbs free energy change in critical cluster formation

Δn_i^* : number of molecules of type *i* in the critical cluster

$\Delta\mu_i$: the difference of the chemical potential of compound *i* in the vapor and liquid, taken at the vapor pressure

ε_i: the energy corresponding to the one-electron orbital ϕ_i

$\varepsilon_{A,B}$: force-field constant related to Wan der Waals interactions between groups A and B

ε_{dielec}: dielectric constant of the medium

ξ: a constant term in the exponential of a gaussian basis function.

θ: azimuthal angle in spherical co-ordinates

θ_0: equilibrium bond angle

$\theta_{A,B}$: angle between the two functional groups A and B

λ: coupling parameter in classical simulations, perturbation parameter in MP theory

ν_i: vibrational frequency corresponding to normal mode *i*

$\rho(\mathbf{r})$: electron density

$\rho(\mathbf{p}^N,\mathbf{r}^N)$: probability distribution function

$\sigma_{A,B}$: force-field constant related to Wan der Waals interactions between groups A and B

$\sigma_{l,v}$: liquid-vapor surface tension

σ_{symm}: rotational symmetry number

$\Phi(\mathbf{r}_{el})$: electronic wavefunction

Φ_{HF} : Hartee-Fock wavefunction

φ: free energy in sections 3. and 3.1; polar angle in spherical co-ordinates in section 3.3

ϕ: general one-electron wavefunction

$\phi_{i,s}(\mathbf{r}_i)$: one-electron wavefunction of the i:th electron in Slater determinant number s
$\chi(\mathbf{r}_{nuc})$: nuclear wavefunction
Ψ: general many-electron wavefunction
Ψ_{CC}: coupled-cluster wavefunction
$\omega_{A,B}$: torsion angle (dihedral angle) between groups A and B

Subscripts, superscripts and operations:

*: critical cluster (in CNT) *or* complex conjugate (in quantum mechanics)
α: corresponding to α - spin electrons
β: corresponding to β - spin electrons
$d\tau$: inside an integral indicates integration over all variables
ΔX: change in variable or parameter X
$\langle X \rangle$: expectation value of X
∇: gradient operator

9. List of abbreviations

6-311++G(d,p): a gaussian valence triple-ξ basis set with one diffuse and one polarization function for each atom
6-311+G(2d,p): a gaussian valence triple-ξ basis set with one polarization function for each hydrogen atom and one diffuse function and two polarization functions for each non-hydrogen atom
6-311++G(2d,2p): a gaussian valence triple-ξ basis set with one diffuse and two polarization functions for each atom
6-311++G(3df,3pd): a gaussian valence triple-ξ basis set with one diffuse and four polarization functions for each atom
AIM: atoms in molecules
ATZ2P: a slater-type valence triple-ξ basis set with one diffuse and two polarization functions for each atom
aug-cc-pV(D+d)Z: a gaussian valence double-ξ basis set with multiple diffuse and polarization functions for each atom
aug-cc-pV(T+d)Z: a gaussian valence triple-ξ basis set with multiple diffuse and polarization functions for each atom
BLYP: a GGA exchange-correlation functional
B3LYP: a hybrid-GGA exchange-correlation functional
BD: Brueckner doubles
CI: configuration interaction
CIS: configuration interaction, singles
CISD: configuration interaction, singles and doubles
CISDT: configuration interaction, singles doubles and triples
CC: coupled cluster
CCN: cloud condensation nuclei
cc-pV(T+d)Z: a gaussian valence triple-ξ basis set with multiple polarization functions for each atom
CCSD: coupled cluster singles and doubles

CCSDT: coupled cluster singles, doubles and triples
CCSD(T): coupled cluster singles and doubles and non-iterative triples
CCSDTQ(P): coupled cluster singles and doubles, triples, quadruples and non-iterative pentuples
CC2: coupled cluster with approximate doubles
CC3: coupled cluster with approximate triples
CNT: classical nucleation theory
CPU: central processing unit
D95++(d,p): a gaussian double-ξ basis set with one polarization and one diffuse function for each atom
DF: density fitting
DFT: density functional theory (Kohn-Sham DFT unless otherwise stated)
DNP: a numerical double-ξ basis set with one polarization function for each atom
FC: frozen-core
FCI: full configuration interaction
GGA: generalized gradient approximation
HF: Hartree-Fock
IPCC: Intergovernmental panel on climate change
LDA: local density approximation
LJ: Lennard-Jones
LSDA: local spin density approximation
MC: Monte Carlo (Metropolis Monte Carlo unless otherwise stated)
MD: molecular dynamics
MPn: n:th degree Møller-Plesset perturbation theory
MPW1B95: a hybrid meta-GGA exchange-correlation functional
NVE: ensemble with constant number of molecules, volume and energy
NVT: ensemble with constant number of molecules, volume and temperature
NPT: ensemble with constant number of molecules, pressure and temperature
PM_{10}: total mass concentration of particles under 10 μm in diameter
PW91: a GGA exchange-correlation functional
QZVPP: a gaussian valence quadruple-ξ basis set with two polarization functions for each atom
RRHO: rigid rotor harmonic oscillator
RI: resolution of identity
SCF: self-consistent field
sCI: stabilized Criegee intermediate
VSCF: vibrational self-consistent field