REPORT SERIES IN AEROSOL SCIENCE
N:o 161 (2015)

MODELLING THE ROLE OF CHARGE IN ATMOSPHERIC PARTICLE FORMATION USING QUANTUM CHEMICAL METHODS

KAI RUUSUVUORI

Division of Atmospheric Sciences
Department of Physics
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 E204, Gustaf Hällströmin katu 2, on February 6th, 2015, at 12 o'clock noon.

Helsinki 2015

Author's Address: Department of Physics
P.O. Box 64
FI-00014 University of Helsinki
kai.ruusuvuori@helsinki.fi

Supervisors: Professor Hanna Vehkamäki, Ph.D.
Department of Physics
University of Helsinki

Docent Theo Kurtén, Ph.D.
Department of Chemistry
University of Helsinki

Reviewers: Professor Jyrki Mäkelä, Ph.D.
Department of Physics
Tampere University of Technology

Application Scientist Nino Runeberg, Ph.D.
CSC - IT Center for Science

Opponent: Professor Henrik Kjærgaard, Ph.D.
Department of Chemistry
University of Copenhagen

ISBN 978-952-7091-10-4 (printed version)
ISSN 0784-3496
Helsinki 2015
Unigrafia

ISBN 978-952-7091-11-1 (pdf version)
http://ethesis.helsinki.fi
Helsinki 2015
Helsingin yliopiston verkkojulkaisut

Careful. We don't want to learn anything from this.
-Bill Watterson, "Calvin and Hobbes"

Acknowledgements

The research presented in this thesis was carried out at the Department of Physics in the University of Helsinki. I want to thank the heads of the department, Prof. Juhani Keinonen and Prof. Hannu Koskinen, for providing the facilities for my work. I also want to thank Prof. Markku Kulmala, head of the Division of Atmospheric Sciences, for the opportunity to work at his division.

I want to thank my supervisors, Prof. Hanna Vehkamäki, Doc. Theo Kurtén and Dr. Kenty Ortega for their guidance, support and patience. I feel fortunate to have had such a trio of talented people guiding me into the world of computational aerosol physics.

I wish to express my gratitude to all my co-authors. Their contributions have been essential for the work presented in this thesis.

I want to thank Prof. Jyrki Mäkelä and Dr. Nino Runeberg for reviewing this thesis. I greatly appreciate the discussions we had and all the helpful and constructive comments I received.

The financial support of the Research Foundation of the University of Helsinki, Academy of Finland and European Research Council is gratefully acknowledged. Furthermore, I wish to thank CSC - IT Center for Science for computational resources and technical assistance.

The division has been a very pleasant place to work at. For this, I thank my colleagues. In particular, I wish to thank the members of the computational aerosol physics group. You have played a crucial part in making work both fun and inspiring.

I thank all my friends for the moments of laughter, discussions, board gaming sessions, movie nights, memorable trips and everything else that make life worth living. A special thank you goes to Dr. Ari Partanen, without whom I might never have finished my master's degree. I am grateful for the friendship of my siblings and the love and support of my parents. Lastly, I wish to thank Anniina Turunen for sharing her life with me. Your love and companionship mean the world to me.

Kai Ilmari Ruusuvuori
University of Helsinki, 2015

Abstract

New particle formation is an important process in the atmosphere. As ions are constantly produced in the atmosphere, the behaviour and role of charged particles in atmospheric processes needs to be understood. In order to gain insight on the role of charge in atmospheric new particle formation, the electron structure of the molecules taking part in this process needs to be taken into account using quantum chemical methods.

Quantum chemical density functional theory was employed in an effort to reproduce an experimentally observed sign preference. While computational results on molecular structures agreed well with results obtained by other groups, the computationally obtained sign preference was opposite to the experimentally observed. Possible reasons for this discrepancy were found in both computational results and experiments.

Simulations of clusters containing water, pyridine, ammonia and a proton were performed using density functional theory. The clusters were found to form a core consisting of ammonium ion and water with the pyridine molecule bonding to the ammonium ion. However, the solvation of the ammonium ion was observed to affect the possibility of proton transfer.

Calculations of proton affinities and gas phase basicities of several compounds, which can be considered as candidates to form atmospheric ions in the boreal forest, were performed. The generally small differences between the calculated gas phase basicites and proton affinities implied only small entropy changes in the protonation reaction. Comparison with experiments resulted in the conclusion that the largest experimentally observed peaks of atmospheric ions most likely corresponded to pyridine and substituted pyridines. Furthermore, a combination of low proton affinity and high observed cation concentration was concluded to imply a high concentration of neutral parent molecules in the atmosphere.

A combination of quantum chemistry and a code for modelling cluster dynamics was employed to study the use of protonated acetone monomers and dimers as the ionization reagent in a chemical ionization atmospheric pressure interface time-of-flight mass spectrometer (CI-APi-TOF). The results showed that the ionization reagents successfully charged dimethylamine monomers. However, there were discrepancies between the simulated and measured cluster distributions. Possible reasons for this discrepancy were found in both measurements and the modelling parameters.

Keywords: Quantum chemistry, computational chemistry, cations, aerosols

Contents

List of publications

This thesis consists of an introductory review, followed by 4 research articles. In the introductory part, these papers are cited according to their roman numerals. **Paper I** is reproduced with the kind permission of Elsevier. **Paper II** is reproduced with the kind permission of American Chemical Society. **Paper III** and **Paper IV** are reproduced under the Creative Commons Attribution 3.0 License.

I Ruusuvuori, K., Kurtén, T., Ortega, I.K., Loukonen, V., Toivola, M., Kulmala, M., and Vehkamäki, H. (2011). Density-functional study of the sign preference of the binding of 1-propanol to tungsten oxide seed particles, *Computational and Theoretical Chemistry*, 996: 322-327.

II Ryding, M.J., Ruusuvuori, K., Andersson, P.U., Zatula, A.S., McGrath, M.J., Kurtén, T., Ortega, I.K., Vehkamäki, H., and Uggerud, E. (2012). Structural rearrangements and magic numbers in reactions between pyridine-containing water clusters and ammonia, *J. Phys. Chem. A*, 116: 4902-4908.

III Ruusuvuori, K., Kurtén, T., Ortega, I.K., Faust, J., and Vehkamäki, H. (2013). Proton affinities of candidates for positively charged ambient ions in boreal forests, *Atmos. Chem. Phys.*, 13: 10397-10404.

IV Ruusuvuori, K., Hietala, P., Kupiainen, O., Jokinen, T., Junninen, H., Sipilä, M., Kurtén, T., and Vehkamäki, H. (2014). The charging of neutral dimethylamine and dimethylamine-sulphuric acid clusters using protonated acetone, *Atmos. Meas. Tech. Discuss.*, 7: 11011-11044.

1 Introduction

Climate change and, perhaps more importantly, human influence on the climate has been under scientific and public discussion for decades. Since the year 1990 the Intergovernmental Panel on Climate Change (IPCC) has published frequent assessment reports on global climate change. In the fifth report, whose Work Group I report was published in 2013 (IPCC, 2013) and Work Group II report was published in 2014 (IPCC, 2014), it was stated that the global mean surface temperatures for the period 2081-2100 may rise 0.3-4.8 °C relative to 1986-2005, depending on the emissions. This does not mean a similar rise in temperature everywhere. Locally the average temperature may be unchanged, change by more than +4.8 °C or even decrease by some amount - and on top of this figure comes the possible seasonal or monthly variation. While it is generally accepted that the consequences will present a challenge for humanity, not all areas will be affected equally. Some areas might benefit from the change, while others could be rendered unfit for human life. In order to better prepare for the coming changes, we need to be able to predict as accurately as possible how the climate will change in different areas.

As climate change depends on various feedback mechanisms and interactions on several different size and time scales, predicting climate change is a multidisciplinary challenge of a notably large scale. However, the basic principles governing the Earth's temperature are quite simple. Since the heat flux from the Earth's core (averaged over the surface of the Earth) is ~ 0.1 W/m^2 (Gando et al., 2011), and since the heat flux from the Sun (averaged over the surface of the Earth) is ~ 340 W/m^2 at the top of the atmosphere (Kopp and Lean, 2011), the former can be ignored as an energy source. In addition to the the Sun as an energy source, there is the surface of the Earth, which can reflect electromagnetic radiation coming from the sun or absorb it, and there is the atmosphere, which can also reflect or absorb electromagnetic radiation coming from the sun (as well as infrared radiation emitted by the Earth itself). If the energy absorbed is larger than the net energy radiated and reflected into space, the global average temperature eventually increases. It is worth mentioning that nearly all of the energy produced by mankind ends up heating the atmosphere. Locally this effect on the radiative forcing may be even 0.68 W/m^2 (Flanner, 2009), which is nearly an order of magnitude larger than the average heat flux from the Earth's core. However, since the global average is only 0.028 W/m^2, this effect can - at least for the moment - be largely neglected as well. Since the Sun is beyond human influence, the only ways to

change the radiative balance is to change the atmosphere and the surface of the Earth. In fact, mankind has been changing them since the dawn of civilization, through e.g. land usage, mining and biomass burning (Ruddiman and Thomson, 2011). The actual magnitude of this influence has been the topic of debate for a long time, dating back to at least the year 1896 (Arrhenius, 1896). While it is true that some of the climate change is due to natural variation, the fifth IPCC report states that human influence on the climate change is not only clear, but it is also the dominant reason for the global warming observed between the years 1951-2010.

The IPCC report on climate change also lists the components of radiative forcing, level of scientific understanding and uncertainty of these different components. One of the few components estimated to have a net cooling effect on the radiative forcing is called "Aerosols and precursors". The net effect of atmospheric aerosols on the radiative forcing is a combination of cooling and warming processes (see for example Haywood and Boucher, 2000, and references therein). The cooling effect is based on the aerosol particles' ability to both scatter light, which is called the direct effect, and to function as cloud condensation nuclei (CCN), which is referred to as the indirect effect. The indirect effect can be further broken down into several factors: because the amount of water available for condensation stays the same, the amount of CCN affects the size of the cloud droplets, which affects the lifetime of the clouds as well as their optical properties - which in turn determine the effect on the Earth's albedo. The more cloud droplets, the whiter the clouds are and the longer they prevail, and thus the larger the cooling effect. The warming effect is based on the absorption properties of atmospheric aerosol particles, which in turn can lead to the heating of the air around them. For example, soot particles absorb long-wavelength radiation much the same way as greenhouse gases (see for example Jacobson, 2004, and references therein). This absorption can apply to long-wavelength radiation coming to the Earth from the Sun, or the radiation that is emitted by Earth itself. The uncertainty related to the effect of atmospheric aerosol particles and their precursors, as well as the uncertainty related to the effect of cloud adjustments due to aerosol particles, is larger than the uncertainties related to any other drivers in climate change. This makes aerosol related effects the biggest contributor to the uncertainty in the total radiative forcing.

1.1 Aerosol particles and condensing vapours

By definition, an aerosol is a system of liquid or solid particles suspended in a gas, such as air. The size of the aerosol particles ranges roughly from 1 nm to 100 micrometers. Figure 1 shows a simplified picture of some of the sources of aerosol particles. These sources can be categorized in two different ways: anthropogenic or natural, and primary or secondary (Seinfeld and Pandis, 2006). Anthropogenic aerosol particles are produced by human activity such as mining, traffic or industry, whereas natural sources include volcanoes, wind picking up dust and sea spray from waves breaking against the shore. Primary aerosol particles, such as Saharan dust, enter the atmosphere as is, whereas secondary aerosol particles are formed in the atmosphere from ambient vapors, such as organic compounds released by vegetation (Slowik et al. 2010). In the past it was thought that the majority of aerosol particles consisted of primary particles, but recent studies suggest that the majority may actually be secondary in nature, depending on the geographical location as well as the atmospheric layer in question (Merikanto et al., 2009, Crippa et al., 2013). This makes particle formation one of the most important processes in the atmosphere (Kulmala et al., 2013).

There are several gas-phase substances in the atmosphere that take part in atmospheric particle formation, although only some of these are relevant to the work presented in this thesis. In addition, since **Paper III** contains tens of molecule species with possible atmospheric relevance, only a few substances will be explicitly mentioned here, namely water, sulphuric acid, pyridine, ammonia, dimethylamine and tungsten oxide. Water vapour is the dominant greenhouse gas and it is abundant in the atmosphere. Even though there is considerable variability in concentrations depending on location, atmospheric layer and ambient temperature (Held and Soden 2000), it is thus likely present in many particle formation reactions. Although the role of water vapour depends on the reaction, it has been linked for example with the formation of sulphuric acid particles. Sulphuric acid, in turn, has been observed to play an important role in atmospheric new particle formation (see Sipilä et al. 2010, and references therein). Pyridine and ammonia are also atmospherically relevant: ammonia is the most abundant gas-phase base in the atmosphere and plays a significant role in atmospheric chemistry (Bouwman et al., 1997), while pyridine containing positive clusters can dominate the positive ion composition (Beig and Brasseur, 2000, Ehn et al., 2010). These two substances were studied in **Paper II**. Amines play a similar role in the atmosphere as ammonia, and

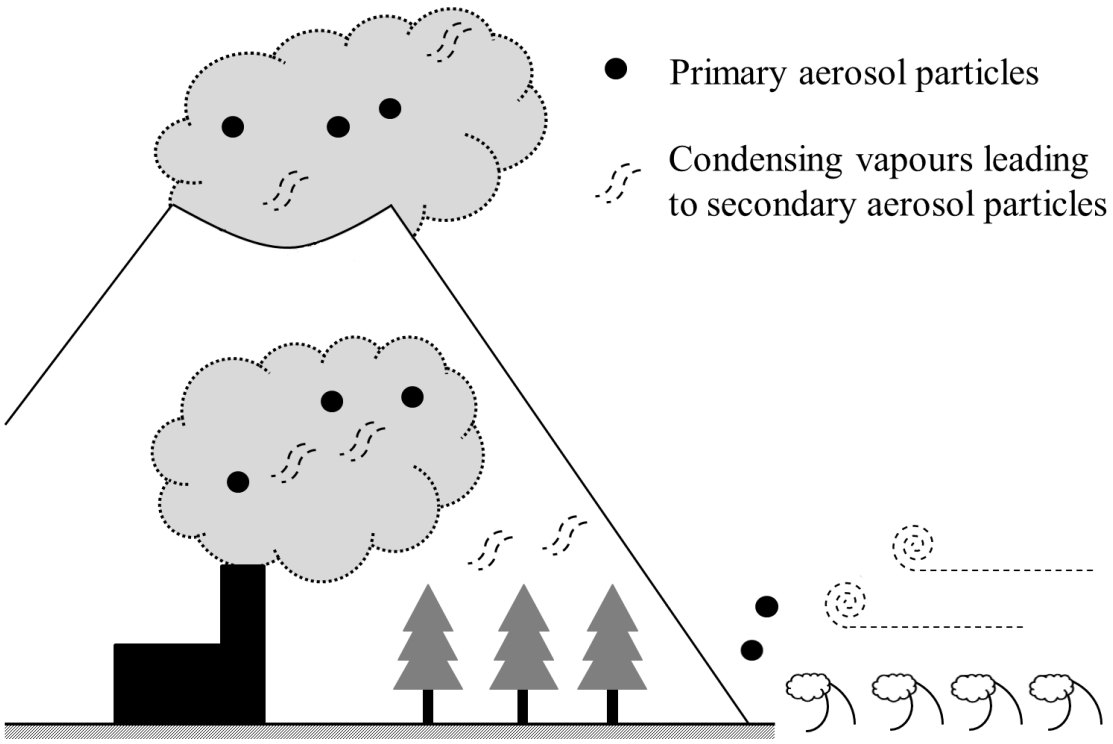

Figure 1: Some of the sources of primary and secondary aerosol particles.

dimethylamine - encountered in **Paper IV** - is one of the most abundant atmopsheric amines (Ge et al. 2011). In addition to these atmospherically relevant substances, also tungsten oxide molecules were studied (**Paper I**). While tungsten oxide molecules are likely not present in the atmosphere in meaningful quantities, tungsten oxide generators provide experimentalists a way to generate monodispersed nanoparticles. Thus, they bear importance for atmospheric science itself.

In addition to playing an important role in the atmosphere, atmospheric aerosol particles also play a role in our day to day lives. Those familiar with smog will know how aerosol particles can affect visibility, and although decreased visibility may often be experienced as a mere nuisance, it can also increase the risk of traffic accidents and hinder air and sea traffic. Atmospheric aerosol particles also have health effects. They have been linked with, for example, respiratory diseases and low birthweight (Raaschou-Nielsen et al., 2013, Pedersen et al., 2013). Furthermore, the World Health Organization (WHO) has linked air pollution to roughly 7 million premature deaths in 2012, making air pollution the single biggest environmental health risk in the world (WHO, 2014). Due to the health effects, there is a need to quantify and limit particle

emissions. There are two common air quality standards, the PM-10 and the PM-2.5, which measure the total particulate mass of particles with a diameter of less than 10 micrometers and less than 2.5 micrometers, respectively. These standards can be used as simple indicators by which the average citizen can assess the risk of health effects. However, discussion is still ongoing which particles harm our health the most (see for example Kreyling et al., 2006 and references therein). It has been suggested that smaller particles penetrate deeper into our respiratory system and even beyond to blood vessels, thus being more harmful than bigger ones. On the other hand, the chemical composition of the particles also plays a role, and since bigger particles can carry more pollutants into our system, they may end up being more harmful than smaller ones. Understanding the way aerosol particles travel in our respiratory system also has importance in the design of inhalators and, by extension, the treatment of diseases such as asthma.

1.2 Electric charge and new particle formation

The study of atmospheric particle formation is the study of the processes and substances that play a role in cluster formation and growth in atmospheric conditions. In this work, the focus will be on cluster formation. Its first steps, often referred to as nucleation, happen at the molecular level. Advances in experimental methods, such as the development of the atmospheric pressure interface time-of-flight mass spectrometer (APi-TOF, Junninen et al., 2010), have pushed detection limits of particles down to about 1 nm. However, direct experimental observation of the formation process itself remains practically impossible, especially if the forming clusters are electrically neutral. This is because electric fields can be used to manipulate the trajectories of ions and electric charges can be measured in small quantities, making charged clusters much easier to detect and characterize than electrically neutral clusters. Thus, electrically neutral samples are often charged to push the limits of measurements further down the size scale. Charging the sample may also be necessary to make detection possible in the first place. For example the aforementioned APi-TOF consists of an APi, which guides sampled ions from atmospheric pressure to the near-vacuum pressure TOF, and the TOF, which measures the ratio of mass and charge of the ion. Thus, the APi-TOF can only measures ions and if the sample is initially electrically neutral, it must be charged before it enters the APi. The problem this presents - and not just for the APi-TOF - is that the addition of charge, and the way it is added, may change the

composition and other properties of the sample.

Computer simulations do not have the same limitations as experimental methods, and can be used to model even the smallest of molecular clusters and their formation. Instead, there are limitations in computational resources as well as practical challenges in modelling molecular clusters containing either a large number of electrons, a large number of molecules or both. Because of this, two types of gaps between experiments and modelling can be identified: either the modelled system or reaction is such that it can not (as yet) be directly measured, or the experimentally observed system is too large to be modelled with sufficient accuracy. In the past there has in practice been little overlap between the regimes of computational modelling and experimental observation. However, the lowering of the experimental detection limits of particles and the increase in computational power, as well as the discovery of more efficient ways of getting accurate results, mean that the gaps are closing. As an example of this development, recent studies have proved the worth of quantum chemical methods in studying new particle formation (Almeida et al., 2013). However, when dealing with clusters with a large amount of molecules, the practical challenges related to finding the best way of arranging those molecules - in other words, the cluster geometry that minimizes the energy of the cluster - largely remain. As yet, a reliable and systematic method for finding this so called global minimum geometry amongst all possible geometries does not exist.

As was mentioned earlier, the electric charge of a cluster means it is easier to measure experimentally. Electric charge also plays a role in atmospheric new particle formation since ions are constantly produced in the atmosphere, mainly by galactic cosmic rays and background radiation (caused by for example radon). As the charging state of a molecule affects its chemical properties, the role these ions play in the chemical reactions related to new particle formation may differ greatly from the role of the original neutral molecule. Charge can also transfer from one molecule to another, or from a molecule to a cluster, which means it can play a role in a chain of chemical reactions, such as the catalytic oxidation of SO_2 (see Bork et al. 2013, and references therein). While experimental observation of, for example, a proton transfer reaction - where a proton is transferred from a molecule to either another molecule or a cluster - is difficult, computational methods allow us to study these kinds of reactions in detail. They also make it possible to accurately determine, for example, electron and proton affinities of molecules - the molecules' "willingness" to accept an electron or a proton,

respectively. Thus, by performing simulations on charged systems we are able to obtain information on how the presence of an electric charge affects atmospheric chemistry.

In addition, simulations can also provide insight on the chemistry that happens inside measurement devices. If the measurement sample is charged, the charging may be done using corona discharge, ionizing radiation or chemical ionization (CI). In chemical ionization, ions are first generated separately and then mixed with the sample flow in the charger. The sample flow goes through the charger and the mixing happens in the beginning of the charger. Thus, the time the sample spends in the charger - which depends on the speed of the flow and the size of the charger - is the duration of the charging process. The idea is to choose the ion species so that it will donate its charge only those molecules in the sample that are the subject of interest. For example in the CI-APi-TOF case studied in **Paper IV** the ions were protonated acetone monomers and dimers (in other words, acetone monomers and dimers with an added proton). The reason acetone was chosen to be the ionization reagent, was its high proton affinity. Since proton transfer is likely to happen only from a molecule with a lower proton affinity to a molecule with a higher proton affinity, these so called charger ions could then be used to selectively charge high proton affinity base molecules such as dimethylamine, as well as clusters containing such molecules. In addition, when employing CI, it is important to produce a high enough concentration of the charger ions so that they are not significantly depleted during the charging process, as this will have a negative effect on the charging efficiency. A poor charging efficiency would in turn lead to a situation where a significant fraction of the substance that is being measured would not be detected. Simulations enable us to study the charging process itself and give insight on how and if the charge is - in theory - transferred to the sample and whether the ion-molecule collisions will also result in clustering. Comparison with experimental results will then show whether theory and practice agree. If discrepancies are found, they can reveal previously unknown issues in either measurements, simulations or both.

The formation and effects of aerosol particles is shown schematically in Figure 2. The figure does not, however, give a comprehensive picture of the field of aerosol study as the study of aerosols encompasses everything from the formation of nanometer-sized molecular clusters to global climate (Kurtén et al., 2008, Stevens et al., 2013), and its applications range from administering drugs via the respiratory system to predicting future weather patterns; from designing particle filters for cars to protecting troops from biological weapons. It is a field of vast importance and complexity. It is also

clear that the uncertainties related to the effects of atmospheric aerosols are not easily solved. Even though the fifth IPCC report shows that scientific understanding of atmospheric aerosols has grown since the last report, there is still a lot of work to be done due to local and seasonal variations in the sinks and sources of primary particles and precursor vapors for secondary particles as well as the atmospheric conditions. In addition, the interactions between, for example, the atmosphere, the cryosphere and the biosphere also depend on the season and the specific local ecosystem; there are direct couplings and feedback effects. For global models, this becomes a problem, since computational resources limit the amount of processes we can explicitly include in the model. For molecular modelling, the problem is the vast amount of molecular species that should be modelled. The key is to try to identify the most important factors in the various processes with the help of experiments and thus narrow down the problem.

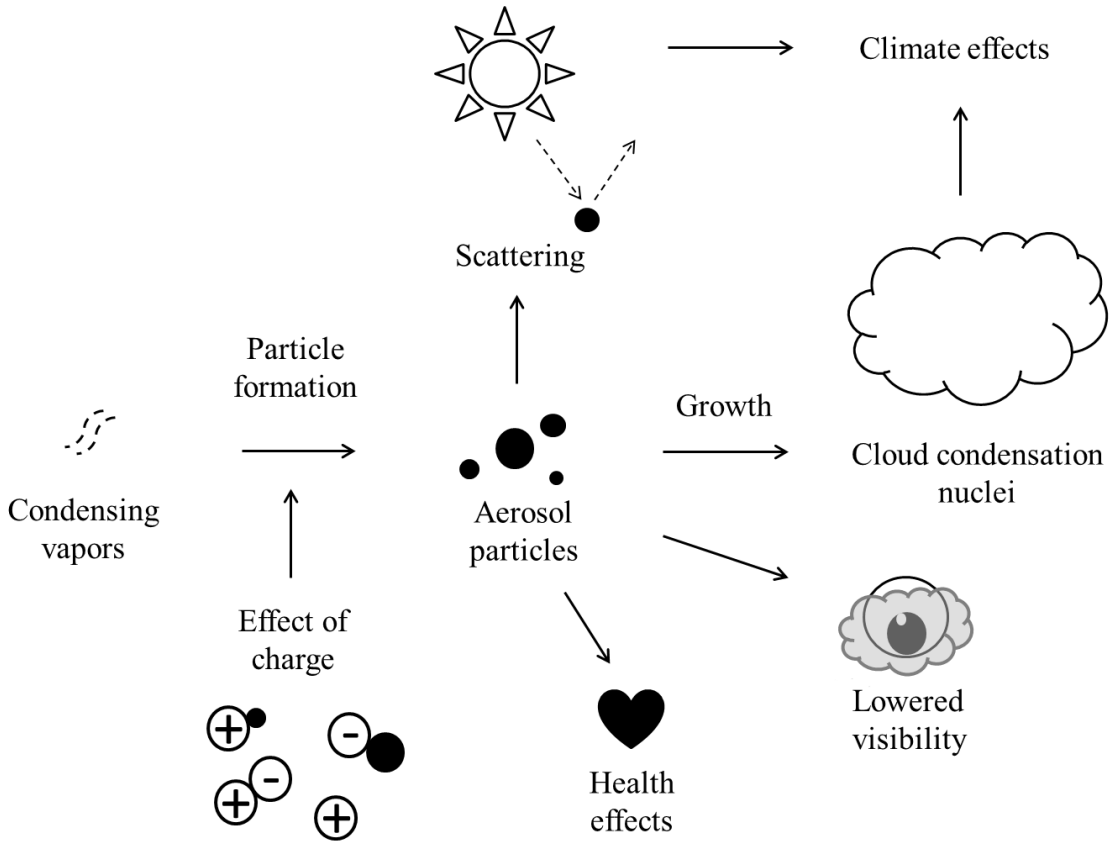

Figure 2: A rough schematic of the formation and effects of aerosol particles.

The main objectives of this thesis are:

1. To describe computational challenges in closing the gap between experiments and theoretical simulations on atmospherically relevant molecular clusters.

2. To study the role and behavior of charge in atmospheric new particle formation.

3. To use computational methods to provide insight on the processes taking place within a chemical ionization atmospheric pressure interface time-of-flight mass spectrometer and to interpret experimental results.

2 Nucleation

Nucleation is a concept that is fundamental for new particle formation. It is the onset of a phase transition, thus governing the first steps of particle formation, and it can be divided into two categories: homogeneous and heterogeneous (see for example Vehkamäki, 2006). In homogeneous nucleation, a supersaturated vapor starts to condense by itself, for example water droplets start to form from water vapor (although this will not happen in atmospheric conditions). In heterogeneous nucleation, a supersaturated vapor starts to condense onto a substrate, which can be anything from a macroscopic surface to a pre-existing aerosol particle, such as a nanometer-sized salt crystal or a molecular cluster.

As microscopic embryos of a new phase are born within the pre-existing phase, some energy is needed to form the interface between these two phases. For example, when water droplets are formed in air, some energy, characterized by surface tension, is needed for the formation of the liquid surface. This means there is an energy barrier which needs to be overcome. Drawn as a function of the amount of molecules in the nucleating cluster, the energy barrier for one-component nucleation without any electric charge is qualitatively depicted in Figure 3.

The cluster size corresponding to the highest point of the barrier is known as the critical cluster. For clusters smaller than the critical cluster size, evaporation is more favorable than growth. This means that, on average, molecules are evaporated from the cluster at a higher rate than they are added. Thus, growth, or going "uphill" along the energy barrier, is an improbable event, where new molecules - or clusters below the critical size - attach themselves to a cluster at a higher rate than molecules are evaporated from the cluster. This is the process of nucleation. For clusters larger than the critical size, growth is favorable, making the clusters stable in the sense that they will unlikely evaporate back into single molecules (although there is still some non-zero probability for evaporation). When evaporation becomes negligible, the process of growth is called condensation. In heterogeneous nucleation, the substrate will replace part of the droplet surface, leading to a smaller surface energy, and thus a lower energy barrier.

Atmospheric particle formation has been traditionally considered to happen by nucleation and subsequent growth. However, to be exact, the particle formation can also be barrierless (Figure 4). This means that every step in the process is energetically favor-

able. As the nucleation process takes the cluster initially in an energetically unfavorable direction, a barrierless reaction does not strictly speaking fall under the category nucleation. It is currently not clear which process, nucleation or barrierless formation, governs atmospheric particle formation or if both processes play an important role.

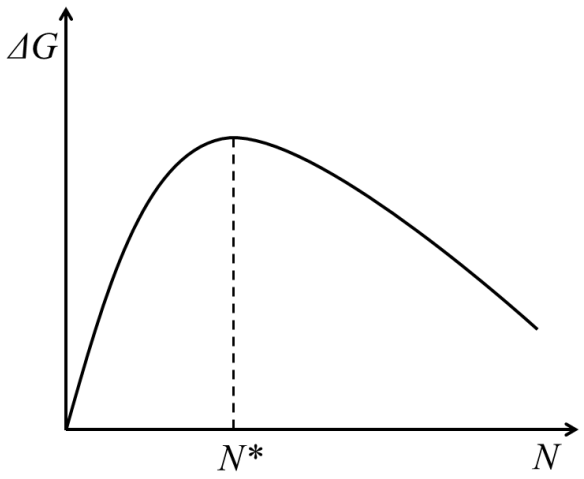

Figure 3: The formation free energy (ΔG) of a one-component system without electric charge as a function of number of molecules (N) in the cluster. The number of molecules in the so-called critical cluster is marked as N^*.

It should be noted that while the curves in Figure 3 and Figure 4 are smooth, in reality the addition of a molecule may not always be energetically favourable even when the growth happens via condensation. Instead, a cluster with a certain number of molecules may be slightly more stable than a cluster with one molecule more or one molecule less. The number of molecules in such a cluster is called a magic number, referring to the fact that the extra stability is anomalous and usually not very well understood.

The traditional theoretical framework for nucleation and growth is known as classical nucleation theory (CNT, see Volmer and Weber, 1925, Farkas, 1927, Becker and Döring, 1935, and Zeldovich, 1942). It has had some success in describing the growth of particles in certain conditions, but fails in predicting the onset pressure for simple substances (see for example Merikanto et al., 2007). One of the main criticisms directed at CNT is that it uses bulk liquid properties to describe liquid droplets also in their earliest stages, when the cluster may consist of only a few molecules. This can easily lead

to errors, since during these earliest steps of nucleation concepts such as density or surface tension are not well defined. Furthermore, classical physics in general may not be suitable to describe systems of this size, since for a system consisting of only a handful of molecules, quantum physical phenomena (such as tunneling) may become important.

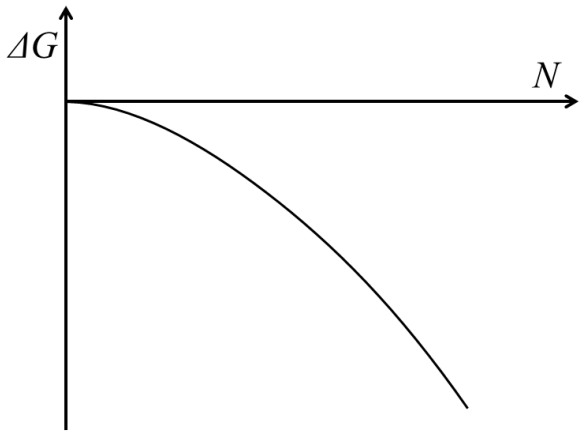

Figure 4: The formation free energy (ΔG) of a one-component system without electric charge as a function of number of molecules (N) in the cluster for the barrierless case. The absence of a free energy barrier means the "critical cluster" is effectively the monomer.

2.1 Ion-induced nucleation

Ion-induced nucleation is a special case of heterogeneous nucleation, where the core ion takes the role of a substrate for the condensing molecules (see Girschick et al, 1996, Chan and Mohnen, 1980, Diamond et al., 1985, Raes and Janssens, 1985, Laakso et al., 2002, and Yu and Turco, 2000). In classical ion-induced nucleation theory (CIINT), the effect of charge has been taken into account according to the classical Kelvin-Thomson equation, which treats positive and negative ions identically. However, its predictions have not always matched experimental results (Mäkelä et al. 1996). This is likely due to the rather simplistic description of the electrostatic potential and the neglect

of the chemical identities of the core ion and the molecules forming the cluster (see for example Kathmann et al. 2005).

It is generally thought that Coulombic attractions may enhance nucleation, either by lowering the free energy barrier shown in Figure 3 or steepening the fall of the curve in Figure 4, as shown in Figure 5 and Figure 6, respectively. However, estimates of the importance of both ion-induced and ion-mediated (where ion-ion recombination is taken into account) nucleation vary. Recent studies suggest that neutral pathways dominate new particle formation at least in the continental boundary layer (Manninen et al. 2010, Hirsikko et al. 2011), but ion-induced nucleation still contributes a non-negligible fraction. Its importance may also grow in other regions and atmospheric layers, where the conditions are such that there are only very small amounts of substances that nucleate without the help of charge.

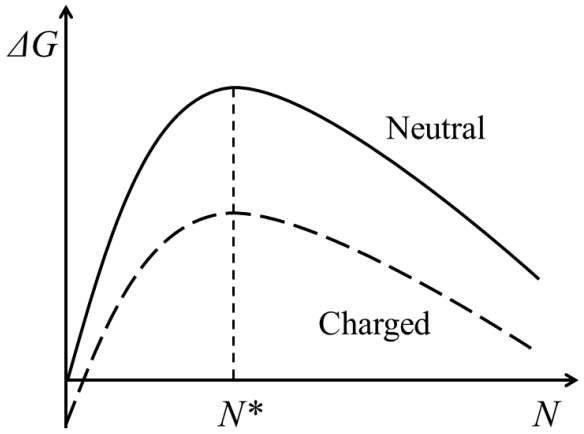

Figure 5: The formation free energy (ΔG) of a one-component system without electric charge (Neutral) and with a charged core ion (Charged) as a function of number of molecules (N) in the cluster. The lower starting point of the curve for the system with a core ion is due to the electrostatic term of the formation free energy. The number of molecules in the so-called critical cluster is marked as N^*.

While it is clear that the role of ions in atmospheric nucleation needs to be studied further, molecular ions themselves have been the subject of scientific study for decades. Although in the classical picture of point charges only the amount of charge matters,

not the sign, in reality negatively charged molecules (anions) differ chemically from positively charged molecules (cations) as well as from electrically neutral molecules (Simons, 2008). This difference is largely due to the valence electrons of anions experiencing a qualitatively different attractive potential, which affects for example the reactivity and the stability of anions. To be more specific, an extra electron in the anion increases its polarizability, which in turn increases the weakest intermolecular attractive forces known as dispersion forces. Furthermore, in laboratory conditions ions of opposite charge are sometimes produced differently. While anions are often produced by adding an electron, cations are often produced by adding a proton - which transforms the neutral molecule into its conjugate acid - instead of removing an electron. This further underlines the differences between anions and cations in experimental measurements.

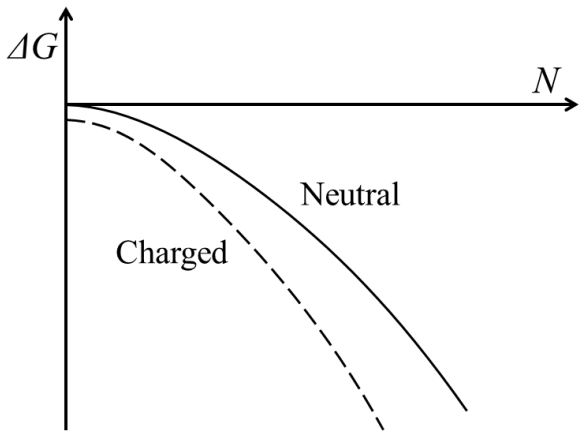

Figure 6: The formation free energy (ΔG) of a one-component system without electric charge (Neutral) and with a charged core ion (Charged) as a function of number of molecules (N) in the cluster for the barrierless case. The lower starting point of the curve for the system with a core ion is due to the electrostatic term of the formation free energy. The absence of a free energy barrier means the "critical cluster" is effectively the monomer.

In reactions between ions and electrically neutral molecules, the effect of the sign of the charge also depends on the chemical properties of the electrically neutral molecule

species. While the net charge of neutral molecules is zero, they still do not have a uniformly distributed electron density, which also leads to a nonuniform electrostatic potential. This gives rise to electrostatic moments such as the dipole and quadrupole moment. The existence of electrostatic moments may be intuitive in cases where the molecule consists of atoms with different electronegativities and the molecular geometry is non-symmetric. An example of such a case is shown in Figure 7. However, even diatomic molecules such as O_2 have electrostatic moments. This is because even their electron density is not completely uniform, even though it does have symmetry. In addition, the electron density of a molecule will react to external electric fields, even ones caused by the presence of molecular ions. As a consequence, some molecules will bind more strongly with negatively charged ions while others will prefer positively charged ions. This so-called sign preference has been observed experimentally as early as 1897, when it was noted that water starts to nucleate earlier when using anions as core ions than when using cations (Wilson, 1897). A negative sign preference was also observed in the study by Winkler et al. (Winkler et al. 2008), which was the starting point for the study presented in **Paper I**. However, the importance of the chemical identity of the core ion should not be forgotten. A schematic picture showing both the classical picture of Coulombic attraction and the effect of the sign of the charge is shown in Figure 8.

The neglect of the specific properties of the core ion, such as molecular structure, also has consequences for the enhancement of particle formation, which are not taken into account in Figure 5 or Figure 6. As a simple example, let us consider a case of barrierless growth, where the addition of a molecule to the core ion is assumed to be energetically favourable initially. Since the core ion has a structure and size, there is only a limited amount of optimal sites for the molecules to attach themselves to. Furthermore, as molecules attach themselves to these sites, the charge of the core ion may be distributed more evenly over the formed cluster. The attached molecules may also surround the core ion so that they shield the surrounding monomers from the effect of the charge of the core ion. Thus, after the energetically most favourable sites have been filled, the effect of the charge may diminish considerably. This may even render the attachment of further molecules energetically unfavourable. Thus, the formation free energy curve of a charged system may resemble the schematic curve shown in Figure 9. It should be noted that since electrically neutral molecules also have structure and sites, which are energetically more favourable to attach to than

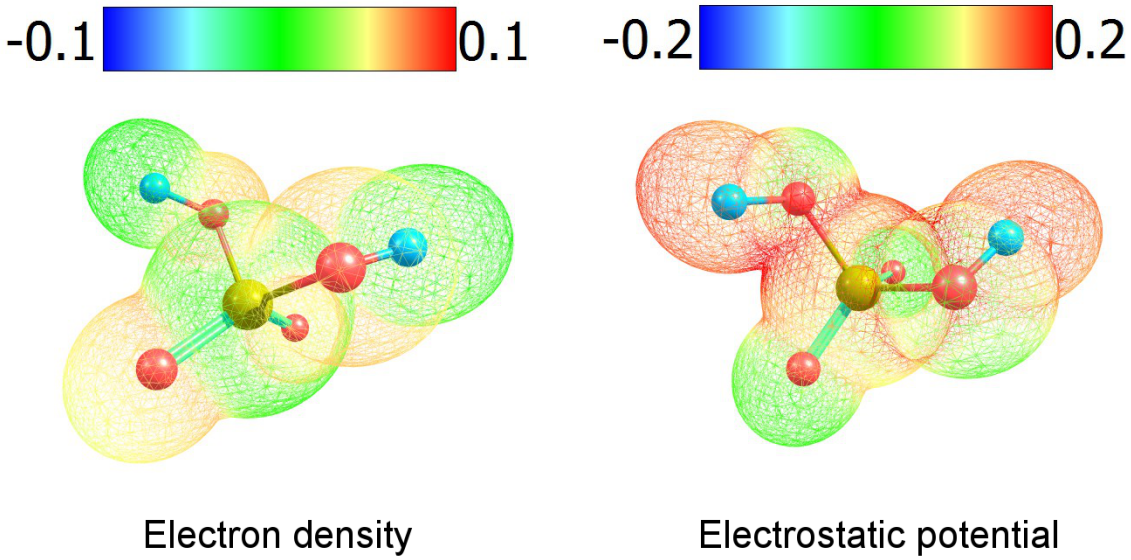

Figure 7: A surface plot of the electron density and electrostatic potential of an electrically neutral sulphuric acid monomer. The numbers are in atomic units and the sizes of the spheres are proportional to van der Waals radii.

others, the same effect can also in some cases be seen for particle formation without ions.

Another aspect of reactions involving ions that the CIINT does not properly take into account is charge transfer. As an example, let us assume the core ion is a cation that has been produced by adding a proton. What happens when the first molecule of the electrically neutral, nucleating substance collides with this ion? Will the molecule stick to the ion and form a thermodynamically stable cluster (in other words, will the free energy change related to the formation of the cluster be negative)? Will the proton transfer from the ion to the molecule? Or will the cluster be thermodynamically unstable, in which case the molecule of the nucleating substance will evaporate? If this happens, will the molecule end up with the proton? This sort of detailed information on the chemistry is not provided by the CIINT. A partial answer may be provided if we know we know the proton affinities of both the neutral parent molecule of the core ion and the nucleating substance. However, while the proton affinity is considered an important thermodynamic quantity, it does not directly describe what really happens when a molecule collides with another molecule that is carrying an extra proton, or

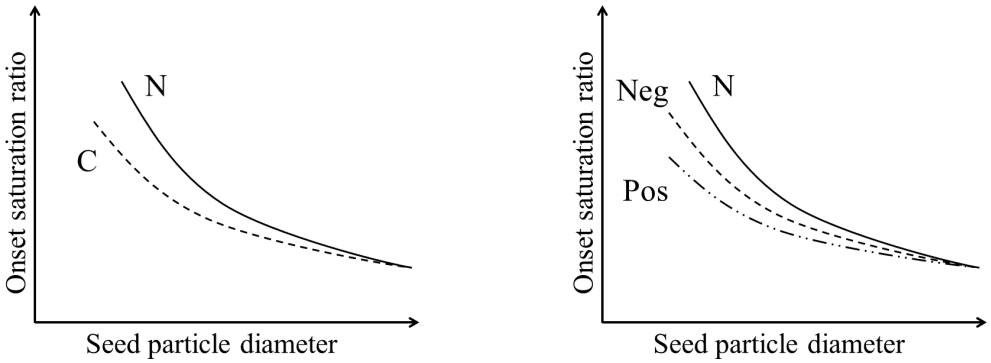

Figure 8: A schematic picture of ion-induced nucleation. The classical picture (left) compared with a more realistic case (right). N denotes electrically neutral seed particles, C denotes charged seed particles of either sign, Neg denotes negatively charged seed particles and Pos denotes positively charged seed particles. The shown positive sign preference does not represent all cases, since the nature of the sign preference depends on the system.

if more than one molecule is involved. For example in the study performed in **Paper II**, water molecules were added to a cluster with a pyridine molecule, an ammonia molecule and a proton. The water molecules were always added to the ammonia part of the core cluster, while the proton was assigned either to the pyridine or the ammonia. Even though pyridine has a higher proton affitinity than ammonia, with two or more water molecules it was energetically favourable for the proton to be attached to the ammonia. Thus, although this qualitative indicator can be of use, it does not give sufficient insight on the behaviour of charge.

It has been shown that in order to understand the role of charge in ion-induced nucleation, the electron structure of the core ion needs to be taken into account (Nadykto et al. 2006). This can be done using a subset of computational methods generally known as quantum chemical methods. Using these methods, we can for example accurately determine the enthalpy and Gibbs free energy changes related to a protonation reaction - and thus determine a molecule's proton affinity and gas-phase basicity, respectively - as was done in **Paper III**. With quantum chemical methods, we can also determine the polarizability of a molecule, which is a measure of how easily the molecule's electron density - and thus its charge density - is distorted by an external electric field. In ad-

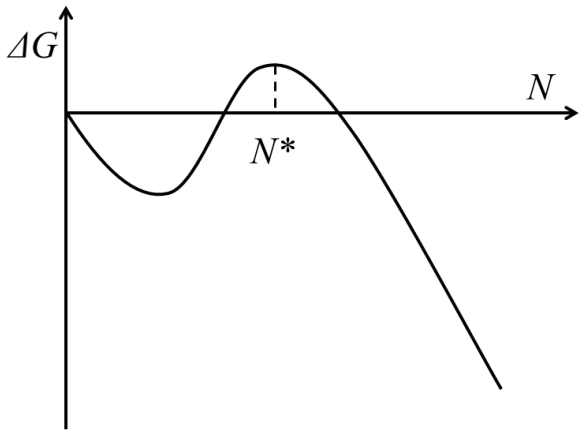

Figure 9: An example of the formation free energy (ΔG) of a system with a charged core ion and a nucleating substance as a function of number of molecules (N). The depicted free energy barrier following a barrierless growth is a possible result of the core ion running out of suitable sites for the monomers to attach themselves to. Since in this case the peak of the barrier is higher than the starting point of the formation free energy curve, the system has a critical cluster size N^*. If the peak of the barrier is lower than the starting point of the curve, the critical cluster size is effectively just the core ion.

dition, proton transfer reactions can be studied by performing a so called relaxed scan for studying the energy barrier related to simply moving the proton from one molecule to another, as was done in **Paper II**. We can also model the dynamics in a given temperature to see how thermal vibrations affect the proton transfer. For example in a study performed by Loukonen et al., it was observed that in one case the proton spent 10% of the time closer to one of the molecules and 90% of the time closer to the other molecule (Loukonen et al. 2014). It is this kind of detailed information that makes quantum chemical methods such an invaluable tool. However, as was previously mentioned, computational methods have challenges of their own. As with measurement equipment, it is important to know the strengths and weaknesses of the computational tools one is employing.

3 Computational science

The increase of computing power of computers, as well as the development of methods and available programs, has led to the increased importance of computational methods for scientists in many different fields. From data mining genomic sequence repeats (Tsirigos et al., 2012) to modelling the global climate (Makkonen et al., 2012) and vehicle frontal impact simulations (Matsumoto et al., 2012), computers have made it possible to perform studies that are otherwise unfeasible due to cost, risk or limitations of experimental methods. From the viewpoint of theory, computational studies can be used to test theories and their consequences as well as provide numerical solutions to equations with no analytical solution. Due to its role and versatility, computational science can be considered the third major field of science, the others being theoretical and experimental science.

The methods used in computational science are as varied as the applications and can be roughly divided into statistical, quantum physical and classical. The employed method needs to be chosen so that it suits the problem at hand. When choosing a method, one needs to consider questions like what is the time scale, how large is the system, is there empirical data available, is dynamics important, what level of accuracy is wanted, are there chemical reactions and what approximations can be made. For example, if the aim is to study a proton transfer reaction the electronic structure of atoms needs to be modelled using quantum physics. On the other hand if the aim is to model real time dynamics of molecular clusters - which can be computationally very demanding - classical methods employing empirical potentials are more suited for the job. Sometimes classical and quantum physical methods can be combined. An example of such a combination is first principles molecular dynamics (FPMD), where the forces between atoms are derived from quantum mechanics and the movement of atoms from the forces using classical Newtonian mechanics.

In the study of atmospheric aerosol particles, simulations are used to study e.g. cluster geometries, vibrational frequencies, chemical reactions (such as proton transfer) and collisions between monomers and clusters. The information obtained from such studies can be used to interpret experimental results or increase our understanding on the reactions taking place inside measurement equipment.

The work presented in this thesis has been done using quantum chemical methods, so statistical and classical methods will not be described in further detail. Even the quan-

tum chemical methods will not be described in full detail here. A more comprehensive description may be found in textbooks such as Jensen (2007) or Koch and Holthausen (2000).

3.1 Quantum chemistry

The electron structure of an element determines its chemical properties (Harrison, 1989). In the classical picture, an atom consists of a nucleus that is orbited by one or more negatively charged point-like particles called electrons. However, in reality the electrons are not point-like particles, but probability distributions as described by quantum physics. In order to model atomic and molecular level interactions properly, the electrons need to be modelled at a quantum physical level. The application of quantum physics to chemical problems is known as quantum chemistry.

In many cases quantum chemistry is about solving the time-independent Schrödinger equation of the system, given in Dirac's bracket notation (see for example Sakurai, 1994) as

$$\hat{H} \left|\Psi_i\right\rangle = E_i \left|\Psi_i\right\rangle \tag{1}$$

where $\hat{H}$ is the Hamiltonian operator, Ψ_i is the wavefunction of state i and E_i is the energy of state i. The wavefunction of state i contains all information there is to know about the system, and after solving the Schrödinger equation we can in principle determine all of the system's attributes. The first step in solving the Schrödinger equation is obtaining the Hamiltonian operator. In order to make this simpler, it can be assumed that the electrons move in a field of fixed nuclei. This approximation is usually good and is justified by the very low mass of the electrons compared to even the lightest nuclei. It allows us to separate the wavefunction into an electronic wavefunction, where the nuclear coordinates are mere parameters, and a nuclear wavefunction, which results in a constant nucleus-nucleus repulsion term, since the fixed nuclei have no kinetic energy. The end results is a so-called electronic Hamiltonian that can be expressed as

$$\hat{H}_{elec} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M}\frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N}\frac{1}{r_{ij}} = \hat{T} + \hat{V}_{Ne} + \hat{V}_{ee} \tag{2}$$

If the nuclei are not fixed, we can still make the same basic approximation, but the movement of the nuclei will result in a mass-polarization term in the total Hamiltonian. In neither of these cases the movement of the electrons depends on the movement of the nuclei, only their position; this is known as the Born-Oppenheimer approximation. In special cases where the electronic Schrödinger equation has several energetically close solutions, the Born-Oppenheimer approximation may break down. Another common approximation is to assume that the Hamiltonian is non-relativistic, as has been done in the equations presented above. This approximation is good for at least first and second row atoms in the periodical table. However, the heavier the modelled atoms are, the more relativistic effects start to matter. These can be taken into account by using so called relativistic effective core potentials (RECPs, see Frenking et al., 1996, Cundari et al., 1996). As the chemical properties of an atom depends mostly on the valence electrons, the core electrons of heavy atoms can be modelled by a suitable function, known as an effective core potential (ECP) or a pseudopotential. For the ECP to be relativistic, it needs to be fitted to atomic calculations that explicitly take into account the relativistic effects. Once this is done, however, the RECP provides a way to both take relativistic effects into account and avoid explicitly modelling the chemically nearly inert core electrons.

The Hamiltonian operator is specific to the system. As can be seen from the form of the non-relativistic Born-Oppenheimer Hamiltonian operator, this property of the operator lies in two terms: the number of electrons and the external potential, which depends on the positions and charges of all nuclei in the system. Once the Hamiltonian operator has been obtained, the next step is to obtain the eigenfunctions Ψ_i and the eigenvalues E_i of the Hamiltonian operator. This opens the door to obtaining any desired property by applying a suitable operator to the wavefunction.

The bad news is that even with the Born-Oppenheimer approximation and non-relativistic Hamiltonian, there is no known way to analytically solve the Schrödinger equation for a system with more than one electron. Thus, the solution has to be done numerically.

3.1.1 The Hartree-Fock method

The description of different quantum chemistry methods starts here with a description of wavefunction based methods. To be more precise, the description begins with that of

the Hartree-Fock (HF) method. It is not only a corner stone of wavefunction methods, but also has conceptual importance, and helps in understanding the Kohn-Sham density functional theory. Thus, the HF method will be described here to some detail, although lenghty derivations are omitted as the work presented in this thesis concentrates on the application of various quantum chemistry methods. Further theory and mathematical formalism can be found in, for example, the books by Jensen (2007) or Koch and Holthausen (2000).

Let us begin with the ground state wavefunction. When the Hamiltonian operator is applied to the ground state wavefunction according to the Schrödinger equation, the lowest energy of the system, E_0, is obtained. The importance of this ground state wavefunction lies in the possibility of approaching it by using what is known as the variational principle.

According to the variational principle, the expectation value of the Hamiltonian operator from any guessed wavefunction is an upper limit to the expectation value obtained using the true ground state wavefunction. Assuming for simplicity fixed nuclei, this can be expressed in Dirac's notation as

$$\langle \Psi_{\text{Trial}} | \hat{H} | \Psi_{\text{Trial}} \rangle = E_{\text{Trial}} \geq \langle \Psi_0 | \hat{H} | \Psi_0 \rangle , \tag{3}$$

where the wavefunctions are the electronic wavefunctions for a system with N electrons. Thus, we need to find the wavefunction that minimizes the energy. In principle this means we need to search through all physically sensible N-electron wavefunctions, which is not feasible. However, the variational principle also applies to any subset of all possible wavefunctions. These subsets can be searched using various algebraic minimizations schemes. The downside is that it is very unlikely that the true ground state wavefunction is included in the subset.

The problem of being unable to solve the Schrödinger equation is thus transformed into finding a suitable subset of wavefunctions. Such a subset should naturally be physicaly reasonable approximation without being computationally unfeasible in practice. In the HF scheme the problem of finding a suitable subset is approached by replacing the N-electron wavefunction by an antisymmetrized product of N one-electron wavefunctions. This product is known as the Slater determinant, and the one-electron wavefunctions are known as spin orbitals.

Now that the wavefunction is transformed into a more manageable form, the next step is to apply the variational principle in an effort to find the best possible Slater determinant. Since the Slater determinant is comprised only of spin orbitals and a normalization factor, only the spin orbitals can be varied. As a result we obtain the HF equations

$$\hat{f}\chi_i = \epsilon_i \chi_i, i = 1, 2, ...N, \tag{4}$$

where $\hat{f}$ is the effective one-electron operator also known as the Fock operator, χ_i is the i^{th} spin orbital and ϵ_i are the orbital energies. The Fock operator, defined as

$$\hat{f}_i = -\frac{1}{2}\nabla_i^2 - \sum_A^M \frac{Z_A}{\vec{r}_{iA}} + V_{HF}(i), \tag{5}$$

includes the HF potential for electron i, $V_{HF}(i)$. The HF potential for electron i depends on its position, $\vec{r}_i$, and the potential consists of two terms

$$V_{HF}(\vec{r}_i) = \sum_j^N \left(\hat{J}_j(\vec{r}_i) - \hat{K}_j(\vec{r}_i) \right), \tag{6}$$

where the first term, known as the Coulomb operator, describes the potential experienced by an electron due to the average charge distribution of other electrons. The latter term is the exchange operator. It has no classical interpretation and is a consequence of the antisymmetric nature of the Slater determinant, which in turn is a result of the need to satisfy the Pauli exclusion principle. However, the exchange operator has an important role in exactly cancelling out the Coulomb self-interaction. This self-interaction arises from the Coulomb operator because the sum includes the value $j = i$, leading to the inclusion of each electron's interaction with itself.

The physical interpretation of HF orbital energies (ϵ_i) is given by Koopmans' theorem (Koopmans, 1934), which states that it is an approximation of the negative of the ionization energy that is associated with the removal of an electron from the orbital in question.

Since the Fock operator depends on the very spin orbitals we would like to obtain from solving the HF equations, the equations cannot be solved in closed form. The usual

technique for going around this problem is known as self-consistent field (SCF). It is an iterative technique, where the first set of orbitals is "guesstimated", then used to solve the HF equations. The resulting new orbitals are then used to solve the HF equations again, and this procedure will go on until some predefined conditions are met.

As electron-electron repulsion has only been treated approximately, a single Slater determinant can never be the true wavefunction of a many-electron system. However, the Slater determinant is the exact wavefunction of N non-interacting particles moving in the field of the effective potential V_{HF}. Thus we can write

$$\hat{H}_{HF}\Phi_{SD} = E^0_{HF}\Phi_{SD} = \sum_i^N \hat{f}_i\Phi_{SD} = \sum_i^N \epsilon_i\Phi_{SD} \tag{7}$$

where the Hamiltonian operator is a sum of the Fock operators given above. As this is an approximation of the true wavefunction, the variational principle tells us that the energy E_{HF} is always higher than the the exact ground state energy. The difference between E_{HF} and E^0_{HF} is called the correlation energy.

The correlation energy is a measure of the error introduced by the HF scheme. The main contributors to this error are the dynamical and non-dynamical electron correlation. The former is due to the fact that electrostatic interaction is treated in an average manner. Thus, there is no true electron correlation, which results in electrons often getting too close to each other. This, in turn, causes an overestimation of electron repulsion. The latter contribution is related to the possibility of other Slater determinants with comparable energies. This may lead to a situation where the ground state Slater determinant may not be a good approximation to the true ground state of the system.

While the single Slater determinant is not ideal, it is a decent approximation to the wavefunction in the sense that it captures most of the physics of a many electron system. It can often produce qualitatively reliable molecular geometries. Furthermore, while the computational effort of performing a HF calculation formally scales with the system size (number of electrons N) as N^4, the actual scaling is often closer to N^3 or even N^2 for some systems. This makes it computationally relatively light, which means it can be used for systems that are too large for other, more accurate methods.

3.1.2 Basis sets

Solving the HF equations encountered in the previous chapter is a highly non-linear optimization problem. This is why in nearly all *ab initio* methods the unknown spin orbitals are expanded as a linear combination of L predefined basis functions (also known as atomic orbitals):

$$\chi_i = \sum_{\alpha}^{L} c_{\alpha i}\phi_\alpha \tag{8}$$

This simplifies the optimization problem into a linear one and the coefficients $c_{\alpha i}$ remain the only variables. If there were an infinite amount of these basis functions, this expansion would not be an approximation, but an exact description of the orbital (known as the complete basis set). However, having an infinite number of functions would not be computationally feasible and only finite basis sets can be used. This naturally affects the accuracy of the description. For example, if we consider the orbital as a function in a coordinate system spanned by the complete basis set, it is apparent that the smaller the basis set is, the worse the description it gives to the orbital, since the components of the orbital along the missing basis functions have no representation. There is a lower limit to the amount of basis functions that can be used to describe a system so that it contains all the electrons of all the (neutral) constituent atoms. Basis sets that have this lowest amount of basis functions are known as minimal basis sets.

The size of the basis set is not the only thing affecting the accuracy of the basis set expansion: also the choice of basis functions themselves affects the accuracy. Since *ab initio* methods generally scale at least at the same rate as the HF method, using as few basis functions as possible would be preferable - and the more accurately the basis function describes the orbital, the fewer are needed. However, basis functions also differ in efficiency. Electronic structure calculations usually employ one of two type basis functions, known as Slater Type Orbitals (STO) (Slater, 1930) and Gaussian Type Orbitals (GTO) (Boys, 1950). The general mathematical form of these basis functions are given by the following equations:

$$\eta^{STO} = Y_{lm}(\Theta, \phi) N r^{n-1} e^{-\zeta r} \tag{9}$$

$$\eta^{GTO} = Nx^l y^m x^n e^{-\alpha r^2} \tag{10}$$

STOs accurately model the true behaviour of the electron cloud and are mostly used for atomic or diatomic systems. They also contain integrals that are computationally challenging. GTOs on the other hand are a lot less challenging computationally. The downside with GTOs is that they have problems in representing the spin orbitals both near the nucleus and far from it. Because of this, roughly three times as many GTOs as STOs are required to get a certain level of accuracy. However, the computational advantage GTOs offer more than makes up for this, which is why STOs are rarely used nowadays. Furthermore, linear combinations of GTOs can be used to construct what are known as contracted Gaussian functions (CGFs). The linear combination can be chosen so that each CFG resembles a single STO as much as possible. This method also reduces the computational cost of the basis functions. Thus, it is no surprise that contracted basis sets are often preferred. There are also a few other types of basis functions: the plane wave basis functions (Slater, 1937) and numerical basis sets (Delley, 1990). The plane wave basis sets take on a slightly different approach compared to using a linear combination of atomic orbitals to describe the system. However, they are best suited for periodic systems or the valence band in a metal. While they can also be used in the study of molecular species, in practice it is usually more efficient to use GTOs. In the numerical basis sets the basis functions are represented numerically on atom centered grids. They are generated by numerically solving the Kohn-Sham equations, which will be presented in chapter 3.1.4. The consequence of solving these equations numerically is that the solution of integrals over basis functions must also be done numerically.

After the type of basis functions has been chosen, improving on the minimal basis set is straightforward; just increase the amount of basis functions. Doubling the number of basis functions for each occupied orbital will result in what is known as a double zeta type basis set, tripling will result in a triple zeta basis set and so on. We can also choose to multiply only the basis functions of the valence electrons, which results in so called split valence basis sets, such as the valence double zeta basis set VDZ. When an atomic or a molecular charge distribution is subjected to an external electric field, the charge density will be distorted. This will also happen when bonds are formed between molecules or atoms. In order to allow the orbitals the flexibility to adapt to an external field, polarization functions - or multiple sets of polarization functions -

need to be added also. These are functions of higher angular momentum - higher l in equations 9 and 10 - than the occupied orbitals in the atom. The addition of such functions may be denoted as P, such as in DZP. Or, if double polarization functions were added to a triple zeta basis set, the corresponding notation might read TZ2P. For example, the DZP basis set for a hydrogen would include two s-functions and one p-function (2s1p) and a TZ2P would include three s-functions and 2 p-functions (3s2p). In order to account for electron correlation - assuming the used method includes it - correlation functions (see for example Woon and Dunning, 1993) need to be used. These are functions of higher angular momentum which are needed in order to describe electron correlation more accurately than in the average mean field approach used in the HF scheme. However, as these also function as polarization functions for HF wavefunctions, they are sometimes called polarization functions as well. Furthermore, when dealing with weak or long range interactions and behavior, additional diffuse basis functions are needed. These are basis functions, where the coefficient in the exponent of the basis function (α in the case of GTOs) is small. Cases which warrant the use of such basis functions include, for example, hydrogen bonds, polarizability calculations and loosely bound electrons in anions and excited states.

Care needs to be taken when adding different types of basis functions to a basis set. For example, too many polarization functions compared to the size of the basis set may result in artefacts (Jensen, 1992). Artefacts and other issues may also rise when using mixed basis sets (a small basis set for the uninteresting atoms and a larger basis set for the more interesting ones) or when trying to describe a neutral molecule with an ionic structure - a strong dipole, where most of the valence electrons are located on one atom - using the same basis set for all atoms of the molecule. On the other hand, it is generally not cost effective to use large basis sets that reduce the absolute error of the energy below chemical accuracy, which is nowadays considered to be ~ 1 kcal/mol. The important thing is to try to keep the error as constant as possible. In short, the basis set needs to be balanced. Even then, the electron density around a nucleus may be described (when the basis functions are centered at the nucleus, as is often the case) by basis functions centered at another nucleus. This is the source for what is known as the basis set superposition error (BSSE). It is introduced when comparing geometries of different molecular systems and can have a qualitative impact on the results when calculating, for example, quantitatively small binding energies. While counterpoise correction (Boys and Bernardi, 1970, van Duijneveldt et al., 1994) and other schemes can be used to reduce the BSSE - at least in intermolecular interactions - it is difficult

to fully get rid of it. Furthermore, the BSSE is not the only source of errors and it can in some cases compensate other errors. Thus, while trying to eliminate BSSE does improve the accuracy of the method, it does not always lead to better results (McMahon and Lane, 2011).

3.1.3 Post Hartree-Fock methods

Improving upon the HF scheme begins with introducing the electron correlation. Methods such as configuration interaction (CI, see Sherrill and Schaefer, 1999) use linear combinations of Slater determinants instead of a single Slater determinant. However, as such methods have not been used in the work presented in this thesis, they will not be described here further.

In perturbation theory, the electron correlation can be introduced also by treating it as a small perturbation to the Hamiltonian operator. Thus, the resulting Hamiltonian operator can be expressed as a sum of two terms:

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}', \tag{11}$$

where λ is a parameter determining the strength of the perturbation, $\hat{H}'$ is the perturbation and $\hat{H}_0$ is the unperturbed Hamiltonian operator. From this the n^{th}-order perturbation equation can be derived:

$$\hat{H}_0 \Psi_n + \hat{H}' \Psi_{n-1} = \sum_{i=0}^{n} W_i \Psi_{n-1}, \tag{12}$$

where W_i is the i^{th}-order correction to the energy when $i \geq 1$. When $i = 0$, the equation returns the Schrödinger equation for the unperturbed case. In the Møller-Plesset (MP) perturbation theory (Møller and Plesset, 1934), the unperturbed Hamiltonian operator is chosen to be the HF Hamiltonian operator, which is a sum over Fock operators:

$$\hat{H}_0 = \sum_{i}^{N} \hat{f}_i, \tag{13}$$

where $\hat{f}_i$ is given by Equation (5). As can be seen with the help of Equations (5) and (6), the sum of Fock operators counts the average electron-electron repulsion twice. The perturbation is then chosen to be the electron-electron repulsion operator minus twice the average electron-electron repulsion operator:

$$\hat{H}' = \hat{H} - \hat{H}_0 = \hat{V}_{ee} - 2\langle\hat{V}_{ee}\rangle \tag{14}$$

Using these equations, the total energy for an order of n can be determined. The second order perturbation - known as MP2 - is the first one to actually include electron correlation and thus improve upon the HF wavefunction and the total energy. As MP3 does not usually offer a notable improvement in accuracy and as the scaling of the computational effort is of the order of N^{n+3} for the nth order perturbation, MP2 is the most often used of all MP methods. With a formal scaling of N^5, MP2 is also the most affordable of all the correlated wavefunction methods. However, as the electron correlation is only treated as a small perturbation, the quality of the original HF wavefunction has a large impact on the success of the MP2 method.

The last wavefunction method to be mentioned here is known as the coupled cluster method (CC, see Bartlett, 1989), where an exponential ansatz is used to parameterize the wavefunction:

$$\hat{\Psi} = e^{\hat{T}}\hat{\Phi}_0, \tag{15}$$

where $\hat{\Phi}_0$ is the HF wavefunction and the exponential operator can be written as:

$$e^{\hat{T}} = \sum_{k=0}^{\infty} \frac{1}{k!}\hat{T}^k. \tag{16}$$

The operator $\hat{T}$ in the exponential is the sum of all the excitation operators:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + ... + \hat{T}_N, \tag{17}$$

where N is the amount of electrons in the system. Although the sum in equation (16) is in principle infinite, in practice the amount of electrons in the system limits the

amount of terms in the sum. This is easy to see if we consider the case where only double excitations are included. Then:

$$\hat{T} = \hat{T}_2, \tag{18}$$

$$e^{\hat{T}} = 1 + \hat{T}_2 + \frac{1}{2}(\hat{T}_2)^2 + + \frac{1}{6}(\hat{T}_2)^3 + ..., \tag{19}$$

where the fourth in term $e^{\hat{T}}$ corresponds to three non-interacting pairs of interacting electrons. If the system contains only five electrons, this term - as well as higher order terms - are unphysical and must thus be omitted. The truncation of equation (17) used in this example is the lowest level approximation of $\hat{T}$ that gives any improvement over HF theory. Due to computational considerations it is common to truncate the sum at some point, and further approximations may also be introduced. The notation reveals which truncations and approximations are in use. For example, CCSD refers to coupled cluster with single and double excitations ($\hat{T} = \hat{T}_1 + \hat{T}_2$), while CC2 (Christiansen et al., 1995) refers to an approximation of CCSD, where a subset of the CCSD equations are used and where the doubles give an MP2 like expression. The upside of using coupled cluster methods is that for an n^{th} order treatment, a significantly higher fraction of the correlation energy is recovered than with the MP methods of the corresponding order. The downside is the computational effort required; not only is the formal scaling of coupled cluster methods N^{2n+2}, but even compared to an MP calculation with the same formal scaling, the computation can still be an order of magnitude more demanding. Due to this reason, high level coupled cluster methods are not usually applicable to atmospherically relevant systems. Methods such as CC2 can, however, be often used for calculating the energy of a specific molecular geometry, also known as the single point energy, due to its N^5 scaling. For the cases where it can be applied, the CCSD(T) method (Raghavachari et al., 1989, Scuseria and Lee, 1990) - where the parentheses mean that the triples contribution is evaluated by perturbation theory and added to the CCSD result - is often considered to be state-of-the-art.

An important property of wavefunction methods is that their accuracy can be increased in a systematic way. The same goes for the basis sets presented in the previous chapter. Thus, the problem of improving the accuracy of electron correlation methods is two-dimensional. However, with the possible exception of the HF method, the scaling of

wavefunction methods limits their use for systems with a large number of electrons, effectively making the limit of complete basis set and full correlation unreachable. At least a partial answer to this problem is provided by density functional theory.

3.1.4 Kohn-Sham Density functional theory

A wavefunction of N electrons has $4N$ variables ($3N$ spatial + N spin), which means it quickly gets complicated, and some simpler way of describing the system would be preferable. As it turns out, the wavefunction can be replaced by an electron density, which only has three spatial variables. Furthermore, this electron density integrates to the number of electrons, it has maxima only at the positions of the nuclei and it contains information about the nuclear charge. Recalling what was previously said about defining the unique Hamiltonian operator of the system, we see that the density fulfills the requirements of uniqueness. This is more clearly stated by the first Hohenberg-Kohn theorem (Hohenberg and Kohn, 1964): The external potential is (to within a constant) a unique functional of $\rho(r)$; since the external potential fixes H, the full many-particle ground state is a unique functional of $\rho(r)$. Since all states of the system are characterized by this single unique Hamiltonian operator, the ground state density is formally enough for a complete determination of all molecular properties (although the ground state itself cannot be used to obtain the properties of e.g. electronically excited states).

The second Hohenberg-Kohn theorem states that the energy obtained from any trial density that fulfills certain boundary conditions is an upper bound to the true ground state energy. In other words, this is the familiar variational principle. Together the first and second Hohenberg-Kohn theorems form a basis for all modern density functional theories. However, the theorems are merely proofs of existence and are of no practical use in calculating the system's properties. In order to make such calculations easier, we may employ the Kohn-Sham scheme (Kohn and Sham, 1965).

Although not the only version of density functional theory (DFT), the Kohn-Sham density functional theory (KS-DFT) is used so widely that it is generally referred to as simply DFT. The idea of the Kohn-Sham scheme is to define a non-interacting reference system of N particles, whose exact ground state wave function can be represented by a Slater determinant. Analogously to the HF case, this approach leads to the so called Kohn-Sham equations that define the spin orbitals of the Slater determinant:

$$\hat{f}^{KS}\varphi_i = \epsilon_i\varphi_i, i = 1, 2, ...N, \tag{20}$$

where the Kohn-Sham operator $\hat{f}^{KS}$ is defined as

$$\hat{f}^{KS} = -\frac{1}{2}\nabla_i^2 + V_S(\vec{r}). \tag{21}$$

The $V_S(\vec{r})$ is an effective potential that can be chosen so that the reference system's density ρ_r equals exactly that of the interacting system ρ_0. It turns out that the potential that produces the correct orbitals of the non-interacting system is of the general form

$$V_S = V_C + V_{Ne} + V_{XC}, \tag{22}$$

where V_C is the classical Coulomb potential, V_{Ne} is the potential due to the nuclei and V_{XC} is the exchange-correlation potential. The V_{XC} is due to the exchange-correlation energy of the system, which contains the contributions of self-interaction correction, exchange and correlation as well as the difference between the kinetic energy of the non-interacting system and the kinetic energy of the interacting system. While the exchange-correlation energy usually represents only a small contribution to the total energy, it is not known exactly for any real world application and needs to be approximated. This approximation is known as the exchange correlation functional.

Analogously to the HF case, the Kohn-Sham orbitals in the Kohn-Sham equation can be expanded in terms of linear combinations of predefined basis functions. It is important to note, however, that even though there are similarities between HF and KS-DFT, care needs to be taken when comparing these two. For example, the exchange-correlation which is present in both theories, does not have the same meaning (Koch and Holthausen, 2000). Another important difference is that the Kohn-Sham approach would be exact if the exact expression for the exchange-correlation energy was known, whereas the HF approach is always an approximation.

The big difference between DFT and wavefunction methods is that in DFT the wavefunction is replaced by an electron density. Thus, there is no wavefunction. The practical advantage of using the electron density instead of a wavefunction is that, regardless

of the system size, the amount of variables stays in principle the same. However, as basis functions are in practice used with DFT, the formal scaling becomes $O(N^4)$. Even so, the scaling is favorable compared with wavefunction methods providing comparable accuracy, which allows larger systems to be modelled. A further advantage is provided by the fact that the DFT energy converges exponentially with respect to basis set size. This means that basis sets larger than a polarized triple zeta are often unnecessary. This helps keep the computational cost low. In addition, as some DFT methods are semi-empirical - meaning that the exchange correlation functional has been approximated by some parameterization based on empirical data - basis sets that differ greatly from those used in the parameterization can lead to larger errors even when the used basis set is larger.

As was already mentioned, with the correct exchange correlation functional, we would obtain the correct eigenvalue of the Hamiltonian operator of the the Schrödinger equation. Unfortunately the form of the correct exchange correlation functional is not known, and there is no known way of systematically improving the approximated functionals, which is why current development of DFT concentrates on finding better exchange correlation functionals. However, there is a rough hierarchy of exchange-correlation functionals, known as the Jacob's ladder (see for example Perdew and Schmidt, 2001), which stretches from the "Hartree world" towards chemical accuracy. The number of rungs on this ladder may vary depending on context, but five rungs are typically found. The first and lowest rung is the local spin density approximation (LSDA, see Kohn and Sham, 1965), the second rung is the generalized gradient approximation (GGA, see Perdew, 1985), the third rung is the meta-generalized gradient approximation (meta-GGA, see for example Perdew et al., 1999, and references therein), the fourth rung is the exact exchange and compatible correlation (for example hyper-GGA, see for example Perdew et al., 2008, and references therein) and the fifth rung is the exact exchange and exact partial correlation. Nowadays the fourth rung often includes any functional with any exact exchange ingredient (hybrid functionals, see for example Haunschild et al., 2012). Widely used functionals include the GGA functional PBE (Perdew et al., 1996, and Perdew et al., 1997), the meta-GGA functional TPSS (Tao et al., 2003) and the hybrid functional B3LYP (Becke, 1993).

3.1.5 Multi-step methods

As mentioned previously, the limit of infinite correlation and infinite basis set is in practice unattainable. There are, however, methods known as theoretical model chemistries, which try to estimate this limit as well as possible. Some of the more known methods belong to either the Complete Basis Set (CBS) or Gaussian-n (Gn) models. These methods employ the fact that different properties of the molecule converge with different rates as the level of sophistication increases. In other words, some properties, such as the minimum energy geometry, are less sensitive to approximations than others, such as the single point energy. For example, the CBS-QB3 method (see Montgomery et al., 1999, and Montgomery et al., 2000, and references therein), used in **Paper III**, includes the following steps:

1. A geometry optimization with the B3LYP density functional using a CBSB7 basis set, which is a triple-zeta basis set.

2. Vibrational frequencies are calculated at the B3LYP/CBSB7 level.

3. A single point energy calculation is performed at the MP2/6-311+G(2df,2p) level, which also yields the corresponding HF energy.

4. A single point energy calculation is performed at the MP4(SDQ)/6-31G(d,p) level in order to estimate the effect of higher order electron correlation.

5. A single point energy calculation is performed at the QCISD(T)/6-31+G(d†) level in order to estimate the effect of higher order electron correlation.

6. Final energies are obtained by extrapolation.

Other model chemistries use different levels of theory in each step. Also the amount of computational steps may vary. Their goal, however, is often the same: to provide chemical accuracy compared to experimental results, such as the G2/97 data set (Curtiss et al., 1998). The weakness of these methods is that they are often quite memory intensive due to the high level energy calculations. Thus, they are not always applicable to larger systems. It is also uncertain whether the model chemistries retain

their accuracy when applied to systems and properties outside the data set used for calibration.

In addition to established model chemistries, one can always construct a multi-step method using any combination of methods and steps that are deemed suitable for the problem at hand. For example, the quantum chemical data in **Paper IV** was produced using a multi-step method consisting of the following steps:

1. A geometry optimization with the B3LYP density functional using a CBSB7 basis set, which is a triple-zeta basis set.

2. Vibrational frequencies are calculated at the B3LYP/CBSB7 level.

3. A single point energy calculation is performed at the RI-CC2/aug-cc-pVTZ level.

This method thus uses the same level of theory for the geometry as the CBS-QB3 method, but calculates the single point energy of the system at a lower level and does not extrapolate it to the basis set limit. As a result, some accuracy is sacrificed in the absolute energies, but this is balanced by the ability to apply this method to systems that are too large to be modelled by CBS-QB3. Furthermore, the property of interest is usually the change in energy in a chemical reaction, which is not as sensitive to approximations as the absolute energy. Naturally, the success of such multi-step approach depends greatly on the chosen methods and the strengths and weaknesses of different methods need to be kept in mind; if the optimized geometry is unrealistic, even the most accurate energy calculation will not yield a useful result. A common - and preferable - practice is performing benchmarking calculations to probe the suitability of the chosen method for the given problem before performing large sets of calculations. The benchmark results are always compared either directly with experimental results or with methods that have already been validated by comparing them with experimental results.

The term multi-step method usually refers to the use of several levels of theory during the course of a simulation. However, the term is somewhat misleading in the sense that a quantum chemical simulation can consist of several steps with varying methods, especially if the aim is to model molecules and molecular cluster (see for example Mitas

et al., 2000, Belbruno et al, 2001, and Vargas et al., 2002). For example, if we want to know the energetics of the formation of a molecular cluster, we begin with trying to find a global minimum energy geometry among all possible local minima by optimizing some initial geometry with the chosen method. Both the global and local minima are stationary points of the energy, expressed as a function of nuclear coordinates. For a point to be stationary, its first derivative must be zero. For the point to be a minimum, the second derivatives along all derivatives need to be positive. However, doing a single geometry optimization is generally not enough, as it is impossible to know beforehand if the stationary point the simulation ends up in is a local minimum or the global minimum. Thus, finding the geometry with the lowest energy starts with conformational sampling or, in other words, studying the ways the molecules can be arranged in the cluster. The larger the cluster, the harder this becomes. However, there are methods for performing conformational sampling other than manually constructing initial cluster geometries and optimizing them using some computational method. Only some of these methods will be shortly described here. We will start with simulated annealing (see for example Bacelo et al., 1999, and references therein), where the basic idea is to give the system a high initial temperature, which will cause large movement of atoms and molecules in the system. The temperature is then then gradually reduced, slowing down the movement and eventually causing the system to end up in a minimum energy geometry. Practical limitations on simulation times means that this method is not guaranteed to lead to a global minimum geometry, but several candidates for a good local minimum geometry can be generated simply by varying the initial temperature and other possible simulation parameters. Another method is replica-exchange molecular dynamics (Sugita and Okamoto, 1999), where the basic idea is to generate several replicas of the system for a range of temperatures (for example one replica for each temperature) and model them simultaneously with molecular dynamics. During the simulation, the temperatures of neighbouring replicas will be swapped several times, which means that the geometry formed at a certain temperature will be subjected to another temperature several times. This leads to the sampling of a wide conformational space, although practical limitations again dictate that the full conformational space will not be covered. Lastly, it is also possible to write codes that automatically generate a large amount of cluster geometries according to some initial conditions, perform optimization simulations on these geometries and then select the best ones according to some set of criteria (see for example Ortega et al., 2012). While a large number of different conformations can in principle be covered this way, such codes are very

dependant on the different parameters that are used (which building block are used for generating clusters, which geometries are classified as suitable, which level of theory is used for optimization etc.). Thus, finding good minimum energy geometries can still be very non-trivial and labour intensive.

After the conformational sampling, further steps are needed in order to get the energetics. The optimization calculation produces the single point energy (at the level of the optimization) at a temperature of 0 K, but in order to get the Gibbs free energy, entropy and enthalpy of the system, a vibrational frequency calculation is needed. The frequency calculation needs to be done using the same quantum chemical method as the optimization. The resulting vibrational frequencies are then used to solve the partition function of the system at a given temperature, from which the needed thermodynamic corrections to the single point energy are calculated. These corrections can then be used with the already obtained single point energy or a value obtained from a separate single point energy calculation with a higher level of theory. All of this - configurational sampling, geometry optimization, vibrational frequency calculation and single point energy calculation - needs to be done to both the resulting cluster and the constituents it is formed from, so that the change in free energy or enthalpy in the reaction can be obtained.

3.2 Computational resources

As has been already seen from the formal scaling of the different methods, the computational cost of the numerical solution is determined by the system size and the accuracy of the method. In practice, computational resources are nearly always a limiting factor and the most accurate methods can only be used for systems consisting of only a few light atoms. This also means that modelling dynamics with a purely quantum chemical method is usually out of the question.

The issue of limited computational resources is not easily solved. A computer simulation needs three things in order to run: computing cores, random access memory (RAM) and hard drive space. As a very rough description, the cores dictate how long the simulation takes, the amount of RAM dictates how much information the cores can handle at one time, and hard drive space dictates how much information can be stored during and after the simulation (for a more extensive description see del Rosario and Choudhary, 1994, and Drepper, 2007). However, not all calculations consume the

resources in an equal way; some may require large amounts of hard disk space for data, while requiring only moderate computing power from the processor, and others may need a lot of computing power in order to finish in a reasonable time, while not requiring a lot of RAM per core.

The evolution of computers has followed what is known as Moore's law (Moore, 1965) for several decades. This (observational) law states that the amount of transistors on integrated circuits - or cores in the context of computers - doubles every two years. As the amount of transistors as well as their speed dictates the computational power of a core, this means the computational power of a single core has (roughly) more than doubled every two years. The practical challenges of including more and faster transistors in a single core have in recent years shifted the focus of increasing performance from creating faster and faster cores to including several cores in a single processor. This has made computational science available also to those who do not have access to high performance computing. For example, a modern desktop computer combined with a commercially available quantum chemistry program will enable the user to perform calculations that would not have been feasible even with a supercomputer not too long ago. High performance computing has not remained the same either, and modern supercomputers, such as the Taito and Sisu clusters at the Finnish IT-center for science, CSC, can have thousands of computing cores (at the time of writing there were 8960 cores on Taito, 11776 cores on Sisu). Users thus have the possibility of running their calculations parallel on up to several thousand cores, assuming that the calculation can be run in this way and the user policies of the cluster allow it. High performance computing can also refer to so called Beowulf clusters, which can be constructed from ordinary desktop computers, even though the performance of Beowulf clusters is far from state-of-the-art supercomputers. However, such clusters can provide a viable and relatively cheap alternative if supercomputers are not available, or if the user wants to have full control over the computing environment.

While parallel computing on several cores has provided a straightforward way to increase computing power, taking full advantage of it is not at all straightforward. In order for parallel computing to provide any advantage, the simulation needs to have parts that can be divided to different processors. Ideally these parts should be as independent of each other as possible, so that a process does not have to wait for another process on some other core to finish before it can be completed. If the different processes need to spend a lot of time waiting for input from other processes, the advantage

of parallellisation may be lost altogether. In computing, the degree to which this parallellisation can be done is called scalability. Ideally, the parallellisation should be linear, so that doubling the amount of processors would halve the amount of time it takes to finish the simulation. Furthermore, the actual code that performs the simulation should be efficient. An inefficient code with linear scaling can in fact be slower than an efficient code with less than favorable scaling. Unfortunately the parallellisation of quantum chemical methods generally leaves something to be desired, especially with the more accurate methods. In practice this means that doubling the amount of cores used from 16 to 32, or even from 8 to 16, may not halve required computation time (see for example Harding et al, 2008). However, it should be noted that the actual behavior of the scaling will depend on the specifics of the calculation (simulated system, method, the used program, how much information needs to be transmitted between the cores and so on).

Other aspects of processor architecture have also changed along the way. Before processors with a 64-bit architecture started to be widely used, most processor architectures were 32-bit, generally limiting the amount of RAM memory they were able to address to 4 gigabytes. In theory, the 64-bit processor architecture enables the use of 16.8 million terabytes of RAM. In practice, the amount of physical space required by the actual memory chips to achieve this makes it impossible to equip any computer with anything close to this amount of RAM. Furthermore, as processor architectures are not fully defined by the classification into 32- or 64-bit, other technical limitations also need to be considered. The limit can be further pushed downwards by operating systems. For example, the maximum limit for RAM varies in different versions of Windows 7 from 8 GB to 192 GB. For quantum chemical methods the increases in usable RAM memory has been a very important development, since in principle only one core is needed to run a simulation - assuming a very lax deadline for the results - but the minimum amount of memory (per core) can easily exceed 4 gigabytes. Even so, the amount of available RAM memory can still be a limiting factor when trying to model anything other than small systems with the more accurate quantum chemical methods with large basis sets. This problem is even more pronounced if (and when) the simulation needs to be run parallel in order to get results in reasonable time, since the amount of requested RAM will then in practice (assuming all the parallel tasks are similar) be divided by the amount of cores that is used for the job. This means that reserving 32 GB of RAM for a calculation using 8 cores will result in 4GB of RAM available for each core. If one would like to have 32 cores in use and the calculation would require

16 GB of RAM per core in order to run, the computing environment would have to be able to provide 512 GB of RAM for the user. This is not an impossible figure, but it does not represent the level that is routinely available to most computational scientists (at the time of writing).

Quantum chemical methods can also lead to large checkpoint or temporary files, which need to be stored on a hard disk. The increase in hard disk space has followed a less famous law of its own, known as the Kryder's law (Walter, 2005), which states that the magnetic disk areal storage capacity increases roughly by an order of magnitude every five years. This has led to a situation where storage space is not often a problem, not least because increasing storage space can be accomplished by adding more hard disks without quite the same challenges as with parallel processing. The challenges are mostly somewhat more practical: heat, space and consumption of electricity. These same challenges naturally apply to other components, and thus to the actual supercomputers, as well (Valentini et al., 2013). However, also the rate at which data can be transferred between different hard disks needs to be taken into account when building large arrays of hard disks. In order to obtain sufficient read/write speeds, it may be necessary to use solid state drives (SSDs) instead of traditional hard drives. SSDs also have other preferable qualities. Due to differences in construction - most notably the lack of moving parts - they are not only faster, but also produce less heat than traditional hard drives. In addition, they are often physically smaller than corresponding traditional hard drives.

All in all, the constant advances made in computer power are of course good news for a computational scientist - if the present technology is not capable of solving a research problem within reasonable time, all he has to do is wait. However, mere waiting is rarely a feasible solution. A better solution would be to use the available resources in an optimal manner. Such optimization is behind e.g. the complete basis set extrapolation, density fitting and resolution of the identity techniques. In the density fitting technique, an (in practice always incomplete) auxiliary basis set is used to describe the electron density (Baerends et al., 1973). Since the molecular orbitals have features the electron density does not have, the auxiliary basis set can be smaller than the basis set used to describe the molecular orbitals. The resolution of the identity (RI) technique, on the other hand, is a mathematical trick used to transform expensive four- and three-center integrals into three- and two-center integrals (Kendall and Früchtl, 1997). The RI does not require the use of an auxiliary basis set. However, the use of such an auxiliary

basis set can improve the computational efficiency (Klopper, 2004). Since density fitting schemes may employ RI, these terms are sometimes used interchangeably.

3.3 The promise of quantum computation

By this point it has become clear that solving the Schrödinger equation is a challenge for conventional computers. The question is, is there some other type of computer which would be more suitable for electron structure calculations. Since we are dealing with calculations concerning quantum physical phenomena, this would imply that a quantum computer might be the solution. As it turns out, a quantum computer does indeed seem promising. Since quantum computers are outside the scope of this work, only a few general points will be mentioned. More information on the subject can be found in the textbook by Nielsen and Chuang (2000) as well as, for example, the review article by Ladd et al. (2010).

One of the potential advantages that makes quantum computers such an interesting proposition over classical computers is quantum parallellisation. This means that we could use one computing core to calculate several values of a function simultaneously as a superposition of these values. On a classical computer a parallel calculation resulting in the same set of values would generally require a core per each calculated value (assuming that techniques such as hyperthreading cannot be utilized in the calculations). The larger the amount of values calculated, the greater the theoretical advantage given by a quantum core. However, the capabilities of any computer are harnessed through various algorithms. It is thus the speed of the algorithm that in the end determines how fast the computation is. At present, there are only three classes of quantum algorithms - although with several possible applications - that provide a performance increase on quantum computers over classical algorithms on classical computers. These are the quantum Fourier transform, the quantum search algorithm and the simulation of systems whose dynamical behaviour is governed by quantum physics (such as the Schrödinger equation) - the last one being good news for those working with quantum chemistry. The caveat is that actually writing efficient quantum algorithms is not a trivial task any more than writing an efficient classical algorithm. Furthermore, while the calculation itself may be swift, it is a difficult task to extract information from the calculated quantum state.

In addition to the challenges in performing quantum algorithms, quantum computers

may need to face the challenge of performing classical algorithms. This challenge may come in the form of the operating system; as the operating system itself does not necessarily have to be faster than on a classical system, it may be more sensible to simulate a classical operating system on a quantum computer than try to code a quantum operating system.

Due to the theoretical and practical challenges related to quantum computers, it is not yet at all clear if quantum computers can live up to the promise of being faster than their classical counterparts.

4 Review of papers

The conventional, yet not insignificant, computational resources provided by CSC were taken advantage of in four different projects probing the effect of charge on the first steps of particle formation.

In **Paper I** we tried to reproduce the experimentally observed sign preference between charged tungsten oxide nanoparticles and n-propanol vapor (Winkler et al. 2008) using DFT. Several different density functionals were employed and the optimized structures agreed with results obtained by other groups. However, the computational sign preference was positive, whereas the experimental sign preference was negative. Possible reasons for this discrepancy were found in both computational results and experiments, highlighting the challenges in computational science.

In **Paper II** we investigated magic numbers observed in clusters containing water, pyridine, ammonia and a proton. As these clusters were quite large, containing roughly 20 water molecules, the use of high-level theory was not possible. Thus, simulations concentrated on small systems of water, pyridine, ammonia and a proton. The aim was to use DFT to obtain insight on how the proton, ammonia and pyridine would behave in the presence of water. The clusters were found to form a core consisting of ammonium ion and water with the pyridine molecule bonding to the ammonium ion.

In **Paper III** we calculated the proton affinities and gas phase basicities of several possible candidates to form atmospheric ions in the boreal forest. The calculations were performed using the CBS-QB3 method, which was chosen based on benchmark calculations, and the proton affinity results were compared to values listed in the NIST WebBook when possible. The agreement with the CBS-QB3 values and the values listed in the NIST WebBook was good. The generally small differences between the calculated gas phase basicites and proton affinities implied that the entropy changes in the protonation reaction were small. The modelled molecules were grouped by their chemical formulae and the largest calculated proton affinity in each group was compared to the experimentally observed cation concentrations in the boreal forest. While information on the sources of the observed cations would have benefitted the comparison, the largest peaks were concluded to most likely correspond to pyridine and substituted pyridines. It was also concluded that a combination of low proton affinity and high observed cation concentration implied a high concentration of neutral parent molecules in the atmosphere.

In **Paper IV** we studied the use of protonated acetone monomers and dimers as the ionization reagent in a chemical ionization atmospheric pressure interface time-of-flight mass spectrometer (CI-APi-TOF). We employed quantum chemical methods to obtain structures, free energies of binding, vibrational frequencies and polarizabilities. We then used the values obtained from quantum chemistry in our cluster kinetics code ACDC to simulate the ionization of various concentrations of sulphuric acid and dimethylamine over a 0.2 s period of time. The simulation results showed that the protonated acetone monomers and dimers successfully charged dimethylamine monomers, forming mainly clusters with one acetone, one dimethylamine and a proton at dimethylamine concentrations of up to 200 ppt. The simulations also showed that with higher dimethylamine concentrations, protonated dimethylamine dimers would be the dominant cluster type. Furthermore, the charger ion concentration started to be depleted at dimethylamine concentrations of over 1000 ppt. This resulted in a "tail", which rose rapidly as the dimethylamine concentration was increased, when the simulation results were normalized with charger ion concentration. The experimental data did not show this behavior. Possible reasons for this discrepancy were found in both measurements and the parameters used in the ACDC code. In addition, the experimental data did not show signs of the protonated dimethylamine dimers even at high dimethylamine concentrations. This could indicate fragmentation of clusters within the measurement device or an underestimation of the measured dimethylamine concentrations. The discrepancies between measurements and simulations need to be studied further before the CI-APi-TOF is a viable option for measuring dimethylamine using acetone as an ionization reagent.

Author's contribution

I am alone responsible for this summary. I have also written the majority of Paper I, Paper III and Paper IV. In Paper II I was responsible for writing the section depicting the computational part. I have performed all of the quantum chemical modelling in Paper I, Paper II and Paper III and a part of the quantum chemical modelling presented in Paper IV. In Paper IV I also used the ACDC code that was developed in-house. However, I did not contribute to its development.

5 Results and discussion

5.1 Sign preference of vapour nucleation on WO_x seed particles

In Chapter 2.1, various differences in the ability to act as nucleation seeds between anions, cations and electrically neutral molecules were briefly discussed, as was the fact that a result of these differences - the sign preference - has also been observed experimentally by Winkler et al. (2008). In the experiment, negatively charged seed particles were observed to have a lower onset saturation ratio (the saturation ratio, where nucleation probability is 50%) than positively charged seed particles. This indicated that there was a negative sign preference between the nucleating substance (n-propanol) and the tungsten oxide seed particles, as well as the smallest negative ions (obtained by ionizing ambient air). In principle, the observed sign preference could be modelled using quantum chemistry methods. However, in our study (**Paper I**), the binding of tungsten oxide molecules to n-propanol molecules seemed to favor positively charged cases. In addition, the size of the tungsten oxide molecule affected the relative ordering of the electrically neutral and negatively charged case: for tungsten dioxide the electrically neutral case was the most weakly bound, for tungsten trioxide and W_3O_9 the negatively charged case was the most weakly bound.

Starting from possible computational reasons for the aforementioned discrepancy, tungsten is a transition metal with 74 electrons per neutral nucleus. As the largest tungsten oxide molecule that was studied had three tungsten atoms and nine oxygen atoms, using a very high level of theory was not computationally feasible. On the other hand, transition metals are somewhat notorious, requiring high accuracy methods for an accurate description of the electron structure due to the possibility of several states with similar energies. The fact that we were modelling not only neutral molecules but anions and cations as well made the study even more challenging. It may well be that the DFT methods used were not accurate enough to reproduce the qualitative sign preference correctly.

Possible reasons for the discrepancy could also be found from the experiment. A tungsten oxide generator was used to produce small nanoparticles, which were then observed and recognized based on their electrical mobility diameter. The electrical mobility diameter was then related to the mass of the particle using Kilpatrick's relation as well as

the particle diameter. However, neither the particle diameter nor Kilpatrick's relation are without issues (see for example Mäkelä et al., 1996). It also later turned out that the experimental setup of Winkler et al. suffered from some contaminations originating from the tubing. Furthermore, unpublished test measurements with another tungsten oxide generator, performed at the Division of Atmospheric Sciences, Department of Physics, University of Helsinki, showed no tungsten oxide ions in the positive channel for particles below 3 nm. While these measurements should not be considered identical to the measurements performed by Winkler et al., they did indicate that some unexpected chemistry involving the positively charged tungsten oxide ions may have been taking place.

It is impossible to say whether any of the mentioned experimental issues are the real reason for the discrepancy. In the experiment, it was not essential to know the exact composition of the seed particles, as it focused on generating and measuring nanometer-sized particles. However, from the viewpoint of the simulations, not knowing the composition of the seed particles diminishes the meaningfulness of the comparison.

In order to fully understand the effect different charging states have, we need to study where the charge is located. Keeping track of the location of extra electrons is less than straightforward, since in quantum physics the location of an electron is a probability distribution. Protons, on the other hand, have a large enough mass to exhibit classical behavior, making them easier to track.

5.2 Formation of protonated water-pyridine-ammonia -clusters

In our study on clusters containing water, pyridine, ammonia and a proton (**Paper II**), we found that the amount of water in the cluster affected the location of the proton: when only one water molecule was present, the proton preferred the pyridine, and when 2-4 water molecules were present, the proton preferred the ammonia. In addition, the minimum energy geometries favored an efficient solvation of the ammonium molecule. In other words, for the case of less than four water molecules, a pyridine molecule, an ammonia molecule and a proton, each of the available water molecules bonded with a hydrogen of the ammonia molecule. When a fourth water molecule was introduced to the system, it preferred to bond with another water molecule, thus starting a second solvation shell.

We performed a relaxed scan of two different proton transfers: From ammonia to pyridine and from pyridine to ammonia, with the two cases having a different configuration of three water molecules. This revealed that the type of solvation of the ammonia molecule affected the possibility of a proton transfer. This piece of information is especially important in atmospheric sciences due to the abundance of water molecules in the atmosphere. Water molecules also present a computational challenge, since adding even a few water molecules to a cluster increases the amount of possible conformations by a significant amount.

The experimental part of the paper studied so called magic numbers as well as the location of charge in protonated clusters containing one ammonia molecule, one pyridine molecule and some water molecules. Although the performed simulations did not include cluster sizes for which magic numbers were experimentally observed, the computational results still provided support for experimental results. Previous experimental results (see Ryding et al., 2012, for several references) implied that in a cluster with a protonated ammonia molecule and 20 water molecules, the protonated ammonia is situated on the surface of the cluster. In the study presented in **Paper II**, it was also concluded that a protonated water cluster with one ammonia and one pyridine molecule will likely have a protonated ammonia. Both of these findings are in line with the computational results. Based on the computational results, it is plausible that a protonated pyridine will donate its proton to an ammonia if the cluster contains more than one water molecule. Thus, starting from clusters with protonated pyridine and water and ending up with clusters with protonated ammonia, water and a pyridine attached to the surface is well within the realm of possibility. This is especially true, if one thinks of the constant, tiny structural rearrangements due to thermal vibrations. However, a definitive answer to the question of magic numbers would benefit from modelling also the larger clusters. Unfortunately the limits of computational resources rule out the use of more accurate methods for this task. The less accurate methods, in turn, might not be able to account for the possible proton transfer. In either case, for a cluster with around 20 molecules the amount of possible cluster geometries means that finding a global minimum energy geometry would require some systematic and reliable way of performing conformational sampling reasonably fast.

5.3 Proton affinities of compounds with atmospheric relevance

The subject of protons was approached from a different perspective in **Paper III**, where we calculated the proton affinities and gas phase basicities of several possible candidates to form atmospheric ions in the boreal forest.

The molecules that were studied were chosen based on mass spectrometer measurements on ambient cations. Even though modern mass spectrometers are able to determine the mass of an observed molecule quite accurately, they are only able to reveal the elemental composition of the ion. This is a problem, since several different molecules, each with different chemical properties, may have exactly the same chemical formula. This limits the extent to which comparisons can be made. If the exact sources for the cations or their neutral parent molecules were known, or if there was some information on their structure, the options could be narrowed down. However, if we know the proton affinities of all molecules with the same chemical formula, we can still say which one of them is most likely retaining the charge in the atmosphere. Quantum chemical calculations of the proton affinities were performed using the CBS-QB3 method, chosen based on benchmarking calculations performed on a handful of molecules, and results were compared to values listed in the NIST WebBook when possible. The modelled molecules were then grouped by their chemical formula and the largest calculated proton affinity in each group was then compared to experimentally observed cation concentrations in the boreal forest. The CBS-QB3 method proved to provide results suitably quickly, and those results compared well with previous results when such were available, with the exception on tropylium cation, whose CBS-QB3 result differed from the NIST WebBook value by 20 kcal mol^{-1}. The tropylium ion is formed by the McLafferty rearrangement (McLafferty, 1959) instead of a simple protonation reaction, which is a likely cause for the discrepancy. However, even though the calculated values seemed reliable, there was no clear correlation between these proton affinities and the observed concentrations. This can be at least partially explained by the fact that the atmospheric cation concentrations do not depend solely on the proton affinity, but also on the concentration of the neutral parent molecules. In other words, a large cation concentration can mean that either a noticeably larger fraction of the neutral parent molecules are charged, or that there are so many more neutral parent molecules around that even a small charging ratio leads to a large ion concentration. This is why drawing definite conclusions from the measurement data and the simulations was challenging.

However, the computational results implied that the change in entropy in the protonation reaction is small. This means the gas phase basicities and proton affinities are close to each other and the following equilibrium holds at least for cases where the difference between the proton affinities is several kcal/mol:

$$\frac{[A]}{[B]} = e^{\frac{1}{RT}[GB(A)-GB(B)]}\frac{[HA^+]}{[HB^+]} \approx e^{\frac{1}{RT}[PA(A)-PA(B)]}\frac{[HA^+]}{[HB^+]} \tag{23}$$

In other words, if the neutral molecule A has a larger proton affinity than the neutral molecule B, then the concentration of the protonated molecule HA^+ will likely be higher than the concentration of the protonated molecule HB^+. If the proton affinities are very close to each other, the relative concentrations of the neutral molecules will the reflect the relative concentrations of the protonated molecules. However, if the neutral molecule A has a smaller proton affinity than the neutral molecule B, yet more of HA^+ is seen in the measurements, then the concentration of neutral molecule A should be considerably larger than the concentration of B. This seemed to be the case for pyridine and the few smaller substituted pyridines, when compared with for example alkyl amines.

5.4 Ion-molecule-cluster reactions in a CI-APi-TOF-instrument

In atmospheric sciences, proton transfer reactions also play a role in measurements, as mentioned in Chapter 2.1. Thus, we studied the use of protonated acetone dimers and monomers in a chemical ionization atmospheric pressure interface time-of-flight mass spectrometer (CI-APi-TOF, see **Paper IV**). Acetone and ethanol had previously been used successfully as ionization reagents in a chemical ionization mass spectrometer (Yu and Lee, 2012). In **Paper IV** the study concentrated on acetone, which has a higher selectivity than ethanol due to a higher proton affinity. To test if acetone could be used as an ionization reagent in a CI-APi-TOF, data provided by quantum chemical calculations was combined with a cluster kinetics code (ACDC) to simulate an idealized situation, where a mixture of dimethylamine and sulphuric acid (with varying concentrations of both) was ionized using protonated acetone dimers and monomers over a 0.2 s period of time. The simulation results showed that the protonated acetone monomers and dimers successfully charged dimethylamine monomers, forming mainly clusters with one acetone, one dimethylamine and a proton. The simulations also showed that with over

200 ppt of dimethylamine in the system, the protonated dimethylamine dimer would become most abundant cluster type. These dimers would be formed through collisions between the dimethylamine monomers and the clusters with an acetone, a dimethylamine and a proton. The sulphuric acid concentration affected the overall charging efficiency at low dimethylamine concentrations, but otherwise only had a non-negligible effect on cluster types containing sulphuric acid. Comparison with measurement data obtained using a CI-APi-TOF indicated that the protonated dimethylamine dimers may fragment within the APi, as they were not seen in the measured signal. The fragmentation would most likely result in a uncharged dimethylamine monomer, which would not be observed by the mass spectrometer, and a protonated dimethylamine. Given the relative magnitudes of the observed amounts of these two cluster types, measurement results could underestimate the total dimethylamine concentration by around 20%. If true, this will have an impact on the interpretation of experimental results.

Another discrepancy in the measurement data was the absence of the upward "tail" that was seen in the simulations at high ($\sim$ 1000 ppt) dimethylamine concentrations. The "tail" is a rapid rise in the ion concentration, normalized with the remaining charger ion concentration after the ionization, and it is caused by the depletion of the charger ions. This means the tail is not a mere artefact of the simulation, but should also be seen in the measurements at some dimethylamine concentration. Thus, possible explanations for this discrepancy were that the simulated "tail" is seen at too low dimethylamine concentrations, or that the dimethylamine concentrations in the experiment were overestimated. Neither possibility could be ruled out. Although the quantum chemical data on the proton transfer reactions should be reliable, tuning parameters like sticking factor could cause the "tail" to shift to higher dimethylamine concentrations. We were able to make the agreement between simulations and measurements better by tuning the sticking factor between dimethylamine monomers and everything else. This could be achieved by using a value of 0.4, which is not an unrealistically small value. While there is no clear physical reason for choosing this value and choosing it only for the dimethylamine monomers, the fact that tuning the sticking factor has a notable effect may imply that some of the collisions do not lead to the formation of a cluster in the experiment. On the other hand, the simulations predicted that the amounts of neutral acetone molecules that were present in the ambient air used to dilution in the experiments should make the agreement between simulations and experiments worse. This discrepancy could not be solved by simply tuning the sticking factor.

As was stated earlier, an overestimation of the experimental dimethylamine concentration could also explain the missing "tail". The dimethylamine concentrations in the experiment were estimated based on the rate with which the permeation tube produced dimethylamine and dilution. Combining this issue with possible wall losses (which were not included in the simulations), the estimated dimethylamine concentrations could be many times the actual dimethylamine concentration. However, a difference of several orders of magnitude is not easily explained by these factors alone. The overall conclusion was that a considerable amount of work is needed to resolve the mismatch between theory and experiments. Before the discrepancies can be better understood, the CI-APi-TOF cannot be recommended for dimethylamine measurements with protonated acetone as the ionization reagent.

The quantum chemical methods used in these studies are listed in Table 1, while Table 2 summarizes shortly the research problems and the obtained results.

Table 1: The quantum chemical methods used in each paper.

Paper	Methods
I	**Density functional theory (functional/basis set):** BLYP/DZP, RPBE/DZP, TPSSTPSS/SDD, TPSSTPSS/def2-QZVPP
II	**Semi-empirical methods:** DFTB **Density functional theory (functional/basis set):** B3LYP/6-31G++(2df,2pd) **Wavefunction methods (method/basis set):** RI-MP2-F12/cc-pVDZ-F12
III	**Multi-step methods:** W1BD, G2, G3, G4, CBS-QB3, CBS-4M, CBS-APNO
IV	**Density functional theory (functional/basis set):** B3LYP/CBSB7 **Wavefunction methods (method/basis set):** RI-CC2/aug-cc-pVTZ **Multi-step methods:** W1BD

Table 2: A summary of research problems and results.

Paper	Summary
I	**Research problem:** Reproducing an experimentally observed sign preference using computational methods. **Results:** Simulations predicted opposite sign preference. Several possible sources for the discrepancy identified.
II	**Research problem:** Studying the behaviour of ammonia, pyridine and a proton in the presence of water. **Results:** The amount of water molecules affected the preferred place of the proton. The type of solvation also affected proton transfer.
III	**Research problem:** Calculating the proton affinities of candidates for ambient cations in the boreal forest. Comparing the results with experimentally observed concentrations. **Results:** The CBS-QB3 proved effective in determining proton affinities. The obtained proton affinities gave insight on the experimentally measured concentrations.
IV	**Research problem:** Studying the viability of using protonated acetone as an ionization reagent in a CI-APi-TOF. Comparing modelling results with measurements. **Results:** Simulations suggest that protonated acetone is a viable ionization reagent in a CI-APi-TOF. However, there was a mismatch between the simulated and experimental results.

6 Conclusions

In the work presented in this thesis, several quantum chemical methods were applied to several different types of research problems related to charge and the effect it has on particle formation. While the effect of charge on certain processes and properties can be measured, simulations of the electron structure of molecules are able to provide more detailed information on the cause and nature of this effect. This makes quantum chemical methods an essential tool in understanding the role of charge in atmospheric processes.

Computational results alone are not enough, however, and need to be validated by direct comparison with experimental data when possible. Despite advances in experimental methods, computational methods and computational resources, there is still a gap between systems that can be directly measured and systems that can be modelled computationally with quantum chemical methods. Closing this gap is likely one of the most important current challenges in aerosol physics. With the help of codes such as the ACDC used in the presented work, some overlap between computational and experimental results is already possible. However, in order to employ quantum chemical methods for producing free energies and polarizabilities for such codes, a certain minimum level of knowledge is required of the modelled system in order to make meaningful comparisons between computational and experimental results. This minimum level is knowing the elemental composition of the molecules under study. In other words, high accuracy mass spectrometers such as the APi-TOF have a key role in bridging the gap. Furthermore, computational methods also have an important role in giving insight on the processes taking place within measurement devices and, by extension, the interpretation of measurement results. This leads to an iterative process where advances in measurement techniques lead to more accurate modelling, which in turn can result in a better understanding of the measurement results.

The study on the sign preference in the binding of tungsten oxide to *n*-propanol (**Paper I**) was unsuccessful in reproducing the experimentally observed sign preference. Possible resons for this discrepancy were found in the experiments as well as the used computational methods, highlighting the need for better computational resources as well as more accurate understanding on the processes taking place inside measurement equipment. However, the magnitude of the computationally obtained positive sign preference implied that the sign of the charge can have a considerable effect on

new particle formation processes. In addition, the relative ordering of the electrically neutral and negatively charged case depended on the type of tungsten oxide molecule.

The sensitivity of the behaviour of charge to the specific properties of the modelled system was also seen in the study of a system with a proton, one pyridine molecule, one ammonia molecule and up to four water molecules (**Paper II**). It was found that with one water molecule attached to the ammonia it was energetically favourable for the pyridine to be protonated. With two to four water molecules it was energetically favourable for the ammonia to be protonated despite it having a lower proton affinity than pyridine. The solvation state of the ammonia molecule was also found to affect the possibility of proton transfer between the pyridine and ammonia molecules.

In **Paper III** and **Paper IV** proton affinity and its practical applications were studied. Comparison between calculated proton affinities and measured ambient cation concentrations in the boreal forest (**Paper III**) implied that pyridine and a few of the smaller substituted pyridines were more abundant in the ambient air than for example alkyl amines. In addition, the CBS-QB3 method was found to provide accurate proton affinities in reasonable time. While proton affinities can also be used for choosing an ionization reagent to be used in chemical ionization (CI), the computational modelling of the charging process of a CI-APi-TOF (**Paper IV**) revealed discrepancies between modelling results and experimental results, showing that proton affinity alone is not enough to understand the chemical processes taking place when sample air is charged. For example, although the charger ions successfully charged the sample in the simulations, the experimentally observed charged cluster types did not show the predicted presence of protonated dimethylamine dimers. If this discrepancy is due to fragmentation within the measurement device, it could lead to roughly a 20% underestimation of ambient dimethylamine concentrations. Further modelling and experiments are needed to better understand the cause of the discrepancies.

In conclusion, charge plays a complex role in molecular interactions and its behaviour depends on the system in question. Proton affinities can be useful in identifying ambient ions and estimating the relative abundancies of their electrically parent molecules in the ambient air based on the observed cation concentrations. Proton affinitien can also be useful in choosing an ionization reagent to be used in chemical ionization. While they can be accurately determined with the help of computational methods, proton affinities should, however, only be considered qualitative indicators. In other words, they should not be expected to capture the details of relevant chemical reactions. This is especially

true in cases where the system consists of more than two molecules and a proton. In practice each specific system needs to be modelled separately to find out how the proton behaves. This applies also to the study of the effect of sign preference. Although present computational resources and modelling tools enable the study of processes inside measurement equipment, the encountered discrepancies between experimental and computational results reveal that there is still work to be done before it can be said that the processes taking place are fully understood. The found discrepancies can, however, help ask the right questions and thus be used to direct the focus of future research.

References

Arrhenius, S. (1896) On the influence of carbonic acid in the air upon the temperature of the ground. *Philosophical Magazine and Journal of Science*, 5, 41:237-276.

Almeida, J., Schobesberger, S., Kürten, A., Ortega, I.K., Kupiainen-Määttä, O., Praplan, A.P., Adamov, A., Amorim, A., Bianchi, F., Breitenlechner, M., David, A., Dommen, J., Donahue, N.M., Downard, A., Dunne, E., Duplissy, J., Ehrhart, S., Flagan, R.C., Franchin, A., Guida, R., Hakala, J., Hansel, A., Heinritzi, M., Henschel, H., Jokinen, T., Junninen, H., Kajos, M., Kangasluoma, J., Keskinen, H., Kupc, A., Kurtén, T., Kvashin, A.N., Laaksonen, A., Lehtipalo, K., Leiminger, M., Leppä, J., Loukonen, V., Makhmutov, V., Mathot, S., McGrath, M.J., Nieminen, T., Olenius, T., Onnela, A., Petäjä, T., Riccobono, F., Riipinen, I., Rissanen, M., Rondo, L., Ruuskanen, T., Santos, F.D., Sarnela, N., Schallhart, S., Schnitzhofer, R., Seinfeld, J.H., Simon, M., Sipilä, M., Stozhkov, Y., Stratmann, F., Tomé, A., Tröstl, J., Tsagkogeorgas, G., Vaattovaara, P., Viisanen, Y., Virtanen, A., Vrtala, A., Wagner, P.E., Weingartner, E., Wex, H., Williamson, C., Wimmer, D., Ye, P., Yli-Juuti, T., Carslaw, K.S., Kulmala, M., Curtius, J., Baltensperger, U., Worsnop, D.R., Vehkamäki, H., and Kirkby, J. (2013) Molecular understanding of sulphuric acid-amine particle nucleation in the atmosphere. *Nature*, 502:359-363, doi: 10.1038/nature12663.

Bacelo, D.E., Binning, R.C. Jr., and Ishikawa, Y. (1999) Ab initio Monte Carlo simulated annealing study of $HCl(H_2O)_n$ $(n = 3, 4)$ clusters *J. Phys. Chem. A*, 103:4631-4640, doi: 10.1021/jp9843003.

Baerends, E. J., Ellis, D. E. and Ros, P. (1973) Self-consistent molecular Hartree-Fock-Slater Calculations: I. The computational procedure. *Chem. Phys.*, 2:41.

Bartlett, R.J. (1989) Coupled-cluster approach to molecular structure and spectra: a step toward predictive quantum chemistry. *J. Phys. Chem.*, 93:1697-1708, doi: 10.1021/j100342a008.

Becke, A.D. (1993) Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98, 7:5648-52, doi: 10.1063/1.464913.

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

Beig, G. and Brasseur, G. P. (2000) Model of tropospheric ion composition: A first attempt. *J. Geophys. Res.*, 105:22671-22684, doi: 10.1029/2000JD900119.

Belbruno, J.J., Tang, Z.-C., Smith, R. and Hobday, S. (2001) The structure and energetics of carbon-nitrogen clusters. *Molecular Physics*, 99, 11:957-967, doi: 10.1080/00268970110040219.

Bork, N., Kurtén, T. and Vehkamäki, H. (1997) Exploring the atmospheric chemistry of O_2SO^{-3} and assessing the maximum turnover number of ion-catalysed H_2SO_4 formation. *Atmos. Chem. Phys.*, 13:3695-3703, doi: 10.5194/acp-13-3695-2013.

Bouwman, A. F., Lee, D. S., Asman, W. A. H., Dentener, F. J., Van Der Hoek, K. W. and Olivier, J. G. J. (1997) A global high-resolution emission inventory for ammonia. *Global Biochemical Cycles*, 11, 4:561-587, doi: 10.1029/97GB02266.

Boys, S.F. (1950) Electronic wave functions. I. A general method of calculation for the stationary states of any molecular system. *Proc. R. Soc. (London) A*, 200:542-554, doi: 10.1098/rspa.1950.0036.

Boys, S.F., and Bernardi, F. (1970) Calculation of small molecular interactions by differences of separate total energies - some procedures with reduced errors. *Mol. Phys.*, 19:553-566, doi: 10.1080/00268977000101561.

Chan, L.Y. and Mohnen, V.A. (1980) The formation of ultrafine ion $H_2O - H_2SO_4$ aerosol particles through ion-induced nucleation process in the stratosphere. *J. Aerosol Sci.*, 11:35-45.

Christiansen, O., Koch, H., and Jorgensen, P. (1995) The second-order approximate coupled cluster singles and doubles model CC2. *Chem. Phys. Lett.*, 243:409-418.

Crippa, M., El Haddad, I., Slowik, J.G., DeCarlo, P.F., Mohr, C., Heringa, M.F., Chirico, R., Marchand, N., Sciare, J., Baltensperger, U., and Prévôt, A.S.H. (2013) Identification of marine and continental aerosol sources in Paris using high resolution aerosol mass spectrometry. *J. Geophys. Res. Atmos.*, 118:1950-1963, doi: 10.1002/jgrd.50151.

Cundari, T.R., Benson, M.T., Lutz, M.L. and Sommerer, S.O. (1996) Effective core potential approaches to the chemistry of the heavier elements. *Rev. Comput. Chem.*, 8, 145.

Curtiss, L.A., Redfern, P.C., Raghavachari, K. and Pople, J.A. (1998) Assessment of Gaussian-2 and density functional methods for the computation of ionization energies and electron affinities. *J. Chem. Phys.*, 109:42.

del Rosario, J.M., and Choudhary, A.N. (1994) High-performance I/O for massively parallel computers: problems and prospects. *Computer*, 27, 3:59-68, doi: 10.1109/2.268887.

Delley, B. (1990) An all-electron numerical method for solving the local density functional for polyatomic molecules. *J. Chem. Phys.*, 92:508-517.

Diamond, G.L., Iribarne, J.V. and Corr, D.J. (1985) Ion-induced nucleation from sulfur dioxide. *J. Aerosol Sci.*, 16:43-55.

Drepper, U. (2007) What every programmer should know about memory. Red Hat, Inc., http://www.akkadia.org/drepper/cpumemory.pdf (10.6.2014)

Ehn, M., Junninen, H., Petäjä, T., Kurtén, T., Kerminen, V.-M., Schobesberger, S., Manninen, H. E., Ortega, I. K., Vehkamäki, H., Kulmala, M. and Worsnop, D. R. (2010) Composition and temporal behavior of ambient ions in the boreal forest. *Atmos. Chem. Phys.*, 10: 8513-8530, doi: 10.5194/acp-10-8513-2010.

Farkas, L. (1927) Keimbildungsgeschwindigkeit in übersättigten Dämpfen. *Z. Physik. Chem.*, 125:236-242.

Flanner, M.G. (2009) Integrating anthropogenic heat flux with global climate models. *Geophys. Res. Lett.*, 36, L02801, doi: 10.1029/2008GL036465

Frenking, G., Antes, I., Böhme, M., Dapprich, S., Ehlers, A.W., Jonas, V., Neuhaus, A., Otto, M., Stegmann, R., Veldkamp, A. and Vyboishchikov, S.F. (1996) Pseudopotential calculations of the transition metal compounds - scope and limitations. *Rev. Comput. Chem.*, 8:63.

Gando, A., Gando, Y., Ichimura, K., Ikeda, H., Inoue, K., Kibe, Y., Kishimoto, Y., Koga, M., Minekawa, Y., Mitsui, T., Morikawa, T., Nagai, N., Nakajima, K., Nakamura, K., Narita, K., Shimizu, I., Shimizu, Y., Shirai, J., Suekane, F., Suzuki, A., Takahashi, H., Takahashi, N., Takemoto, Y., Tamae, K., Watanabe, H., Xu, B.D., Yabumoto, H., Yoshida, H., Yoshida, S., Enomoto, S., Kozlov, A., Murayama, H., Grant, C., Keefer, G., Piepke, A., Banks, T.I., Bloxham, T., Detwiler, J.A., Freedman, S.J., Fujikawa, B.K., Han, K., Kadel, R., O'Donnell, T., Steiner, H.M., Dwyer,

D.A., McKeown, R.D., Zhang, C., Berger, B.E., Lane, C.E., Maricic, J., Miletic, T., Batygov, M., Learned, J.G., Matsuno, S., Sakai, M., Horton-Smith, G.A., Downum, K.E., Gratta, G., Tolich, K., Efremenko, Y., Perevozchikov, O., Karwowski, H.J., Markoff, D.M., Tornow, W., Heeger, K.M. and Decowski, M.P. (2011) Partial radiogenic heat model for Earth revealed by geoneutrino measurements. *Nature Geoscience*, 4:647-651.

Ge, X., Wexler, A. S. and Clegg, S. L. (2011) Atmospheric amines - Part I: A review. *Atmospheric Environment*, 45, 3:524-546, doi: 10.1016/j.atmosenv.2010.10.012.

Girshick, S.L., Rao, N.P. and Kelkar, M. (1996) Model for ion-induced nucleation based on properties of small ionic clusters. *J. Vac. Sci. Technol.*, 14:1/6/01.

Harding, M.E., Metzroth, T., and Gauss, J. (2008) Parallel calculation of CCSD and CCSD(T) analytic first and second derivatives. *J. Chem. Theory Comput.*, 4, 1:64-74, doi: 10.1021/ct700152c.

Harrison, W.A. (1989) Electronic structure and the properties of solids: the physics of the chemical bond. Dover Books on Physics, Dover Publications.

Haunschild, R., Odashima, M.M., Scuseria, G.E., Perdew, J.P., and Capelle, K. (2012) Hyper-generalized-gradient functionals constructed from the Lieb-Oxford bound: Implementation via local hybrids and thermochemical assessment. *J. Chem. Phys.*, 136:184102, doi: 10.1063/1.4712017.

Haywood, J., and Boucher, O. (2000) Estimates of the direct and indirect radiative forcing due to tropospheric aerosols: a review. *Reviews of Geophysics*, 38, 4:513-543, doi: 10.1029/1999RG000078.

Held, I.M. and Soden, B. J. (2000) Water vapor feedback and global warming. *Annu. Rev. Energy Environ.*, 25:441-475, doi: 10.1146/annurev.energy.25.1.441.

Hirsikko, A., Nieminen, T., Gagné, S., Lehtipalo, K., Manninen, H.E., Ehn, M., Hõrrak, U., Kerminen, V.-M., Laakso, L., McMurry, P.H., Mirme, A., Mirme, S., Petäjä, T., Tammet, H., Vakkari, V., Vana, M. and Kulmala, M. (2011) Atmospheric ions and nucleation: a review of observations. *Atmos. Chem. Phys.*, 11:767-798, doi: 10.5194/acp-11-767-2011.

Hohenberg, P. and Kohn, W. (1964) Inhomogeneous electron gas. *Phys. Rev.*, 136, B864.

IPCC [Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V.and Midgley, P.M. (eds.)] (2013) IPCC, 2013: Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

IPCC [Field, C.B., Barros, V.R., Dokken, D.J., Mach, K.J., Mastrandrea, M.D., Bilir, T.E., Chatterjee, M., Ebi, K.L., Estrada, Y.O., Genova, R.C., Girma, B., Kissel, E.S., Levy, A.N., MacCracken, S., Mastrandrea, P.R. and White, L.L. (eds.)])] (2014) IPCC, 2014: Summary for Policymakers. In: Climate Change 2014: Impacts, Adaptation and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1-32.

Jacobson, M.Z. (2004) Climate response of fossil fuel and biofuel soot, accounting for soot's feedback to snow and sea ice albedo and emissivity. *Journal of Geophysical Research: Atmospheres*, 109:2156-2202, doi: 10.1029/2004JD004945.

Jensen, F. (1992) Reaction of organic sulfides with singlet oxygen. A theoretical study including electron correlation. *J. Org. Chem.*, 57, 24:6478-6487, doi: 10.1021/jo00050a022.

Jensen, F. (2007) Introduction to Computational Chemistry. John Wiley & Sons Ltd

Junninen, H., Ehn, M., Petäjä, T., Luosujärvi, L., Kotiaho, T., Kostiainen, R., Rohner, U., Gonin, M., Fuhrer, K., Kulmala, M. and D. R. Worsnop (2010) A high-resolution mass spectrometer to measure atmospheric ion composition. *Atmos. Meas. Tech.*, 3:1039-1053, doi: 10.5194/amt-3-1039-2010.

Kathmann, S. M., Schenter, G. K. and Garrett, B. C. (2005) Ion-induced nucleation: The importance of chemistry. *Phys. Rev. Lett.*, 94:116104-1 - 116104-4, doi: 10.1103/PhysRevLett.94.116104.

Kendall, R. A. and Früchtl, H. A. (1997) The impact of the resolution of the identity approximate integral method on modern ab initio algorithm development. *Theor. Chem. Acc.*, 97:158-163, doi: 10.1007/s002140050249.

Klopper, W. (2004) A hybrid scheme for the resolution-of-the-identity approximation in second-order Møller-Plesset linear-r_{12} perturbation theory. *J. Chem. Phys.*, 120:10890, doi: 10.1063/1.1742904.

Koch, W. and Holthausen, M. C. (2000) A Chemist's Guide to Density Functional Theory. WILEY-VCH

Kohn, W. and Sham, L.J. (1965) Self consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133-A1138, doi: 10.1103/PhysRev.140.A1133.

Koopmans, T.A. (1934) Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den einzelnen Elektronen Eines Atoms. *Physica*, 1:104-113, doi: 10.1016/S0031-8914(34)90011-2.

Kopp, G. and Lean, J.L. (2011) A new, lower value of total solar irradiance: Evidence and climate significance. *Geophys. Res. Lett.*, 38, L01706, doi: 10.1029/2010GL045777.

Kreyling, W.G., Semmler-Behnke, M. and Möller, W. (2006) Ultrafine particle-lung interactions: does size matter? *Journal of Aerosol Medicine*, 19, 1:74-83.

Kulmala, M., Kontkanen, J., Junninen, H., Lehtipalo, K., Manninen, H.E., Nieminen, T., Petäjä, T., Sipilä, M., Schobesberger, S., Rantala, P., Franchin, A., Jokinen, T., Järvinen, E., Äijälä, M., Kangasluoma, J., Hakala, J., Aalto, P.P., Paasonen, P., Mikkilä, J., Vanhanen, J., Aalto, J., Hakola, H., Makkonen, U., Ruuskanen, T., Mauldin III, R.L., Duplissy, J., Vehkamäki, H., Bäck, J., Kortelainen, A., Riipinen, I., Kurtén, T., Johnston, M.V., Smith, J.N., Ehn, M., Mentel, T.F., Lehtinen, K.E.J., Laaksonen, A., Kerminen, V.-M., and Worsnop, D.R, (2013) Direct observations of atmospheric aerosol nucleation. *Science*, 339:943-946, doi: 10.1126/science.1227385.

Kurtén, T., Loukonen, V., Vehkamäki, H., and Kulmala, M. (2008) Amines are likely to enhance neutral and ion-induced sulfuric acid-water nucleation in the atmosphere more effectively than ammonia. *Atmos. Chem. Phys.*, 8:4095-4103, doi: 10.5194/acp-8-4095-2008.

Laakso, L., Mäkelä, J. M., Pirjola, L. and Kulmala, M.(2002) Model studies on ion-induced nucleation in the atmosphere. *J. Geophys. Res. D*, 107, D20:AAC 5-1 - AAC 5-19, doi: 10.1029/2002JD.

Ladd, T. D., Jelezko, F., Laflamme, R., Nakamura, Y., Monroe, C. and O'Brien, J. L. (2010) Quantum computers. *Nature*, 464:45-53, doi: 10.1038/nature08812.

Loukonen, V., Kuo, I.-F. W., McGrath, M. J. and Vehkamäki, H. (2014) On the stability and dynamics of (sulfuric acid)(ammonia) and (sulfuric acid)(dimethylamine) clusters: a first-principles molecular dynamics investigation. *Chem. Phys.*, 428:164-174, doi: ??.

Makkonen, R., Asmi, A., Kerminen, V.-M., Boy, M., Arneth, A., Guenther, A., and Kulmala, M. (2012) BVOC-aerosol-climate interactions in the global aerosol-climate model ECHAM5.5-HAM2. *Atmos. Chem. Phys.*, 12:10077-10096, doi: 10.5194/acp-12-10077-2012.

Manninen, H.E., Nieminen, T., Asmi, E., Gagné, S., Häkkinen, S., Lehtipalo, K., Aalto, P., Vana, M., Mirme, A., Mirme, S., Hõrrak, U., Plass-Dülmer, C., Stange, G., Kiss, G., Hoffer, A., Törö, N., Moerman, M., Henzing, B., de Leeuw, G., Brinkenberg, M., Kouvarakis, G.N., Bougiatioti, A., Mihalopoulos, N., O'Dowd, C., Ceburnis, D., Arneth, A., Svenningsson, B., Swietlicki, E., Tarozzi, L., Decesari, S., Facchini, M.C., Birmili, W., Sonntag, A., Wiedensohler, A., Boulon, J., Sellegri, K., Laj, P., Gysel, M., Bukowiecki, N., Weingartner, E., Wehrle, G., Laaksonen, A., Hamed, A., Joutsensaari, J., Petäjä, T., Kerminen, V.-M. and Kulmala, M. (2010) EUCAARI ion spectrometer measurements at 12 European sites - analysis of new particle formation events. *Atmos. Chem. Phys*, 10:7907-7927, doi: 10.5194/acp-10-7907-2010.

Matsumoto, A.T., Driemeier, L., and Alves, M. (2012) Performance of polymeric reinforcements in vehicle structures submitted to frontal impact. *International Journal of Crashworthiness*, 17, 5:479-496, doi: 10.1080/13588265.2012.678693.

McMahon, J.D., and Lane, J.R. (2011) Explicit correlation and basis set superposition error: The structure and energy of carbon dioxide dimer. *J. Chem. Phys.*, 135:154309-1 - 154309-9, doi: 10.1063/1.3653230.

Merikanto, J., Zapadinsky, E., Lauri, A. and Vehkamäki, H. (2007) Origin of the failure of classical nucleation theory: incorrect description of the smallest clusters. *Phys. Rev. Lett.*, 98:145702-1 - 145702-4, doi: 10.1103/PhysRevLett.98.145702.

Merikanto, J. Spracklen, D.V., Mann, G.W., Pickering, S.J., and Carslaw, K.S. (2009) Impact of nucleation on global CCN. *Atmos. Chem. Phys.*, 9:8601-8616, doi: 10.5194/acp-9-8601-2009.

Mitas, L., Grossman, J.C., Stich, I., and Tobik, J. (2000) Silicon clusters of intermediate size: energetics, dynamics, and thermal effects. *Phys. Rev. Lett.*, 84, 7:1479-1482, doi: 10.1103/PhysRevLett.84.1479.

Montgomery, J.A. Jr., Frisch, M.J., Ochterski, J.W., and Petersson, G.A. (1999) A complete basis set model chemistry, VI. Use of density functional geometries and frequencies. *J. Chem. Phys.*, 110:2822-2827.

Montgomery, J.A. Jr., Frisch, M.J., Ochterski, J.W., and Petersson, G.A. (2000) A complete basis set model chemistry, VII. Use of the minimum population localization method. *J. Chem. Phys.*, 112:6532-6542.

Møller, C., and Plesset, M.S. (1934) Note on an approximation treatment for many-electron systems. *Phys. Rev.*, 46:618-622.

Moore, G.E. (1965) Cramming more components onto integrated circuits. *Electronics*, 38:114.

Mäkelä, J.M., Riihelä, M., Ukkonen, A., Jokinen, V. and Keskinen, J. (1996) Comparison of mobility equivalent diameter with Kelvin-Thomson diameter using ion mobility data. *J. Chem. Phys.*, 105:1562-1571.

McLafferty, F.W. (1959) Mass spectrometric analysis, molecular rearrangements. *Anal. Chem.*, 31:82-87.

Nadykto, A.B., Natsheh, A.A., Yu, F., Mikkelsen, K.V. and Ruuskanen, J. (2006) Quantum Nature of the Sign Preference in Ion-Induced Nucleation. *Phys. Rev. Lett.*, 96:125701-1 - 125701-4, doi: 10.1103/PhysRevLett.96.125701.

Nielsen, M. A. and Chuang, I. L. (2000) Quantum Computation and Quantum Information. Cambridge University Press, Cambridge.

Ortega, I.K., Kupiainen, O., Kurtén, T., Olenius, T., Wilkman, O., McGrath, M.J., Loukonen, V. and Vehkamäki, H. (2012) From quantum chemical formation free energies to evaporation rates. *Atmos. Chem. Phys*, 12:225-235, doi: 10.5194/acp-12-225-2012.

Pedersen, M., Giorgis-Allemand, L., Bernard, C., Aguilera, I., Andersen, A.-M.N., Ballester, F., Beelen, R.M.J., Chatzi, L., Cirach, M., Danileviciute, A., Dedele, A., van Eijsden, M., Estarlich, M., Fernández-Somoano, A., Fernández, M.F., Forastiere,

F., Gehring, U., Grazuleviciene, R., Gruzieva, O., Heude, B., Hoek, G., de Hoogh, K., van den Hooven, E.H., Håberg, S.E., Jaddoe, V.W.V., Klümper, C., Korek, M., Krämer, U., Lerchundi, A., Lepeule, J., Nafstad, P., Nystad, W., Patelarou, E., Porta, D., Postma, D., Raaschou-Nielsen, O., Rudnai, P., Sunyer, J., Stephanou, E., SÃ¸rensen, M., Thiering, E., Tuffnell, D., Varró, M.J., Vrijkotte, T.G.M., Wijga, A., Wilhelm, M., Wright, J., Nieuwenhuijsen, M.J., Pershagen, G. and Brunekreef, B. (2013) Ambient air pollution and low birthweight: a European cohort study (ESCAPE). *The Lancet Respiratory Medicine*, 1, 9:695-704.

Perdew, J.P. (1985) Accurate density functional for the energy: Real-space cutoff of the gradient expansion for the exchange hole. *Phys. Rev. Lett.*, 55:1665-1668, doi: 10.1103/PhysRevLett.55.1665

Perdew, J.P., Burke, K., and Ernzerhof, M. (1996) Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865-3868, doi: 10.1103/PhysRevLett.77.3865

Perdew, J.P., Burke, K., and Ernzerhof, M. (1997) Errata: Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 78:1396, doi: 10.1103/PhysRevLett.78.1396

Perdew, J.P., Kurth, S., Zupan, A., and Blaha, P. (1999) Accurate density functional with correct formal properties: A step beyond the generalized gradient approximation. *Phys. Rev. Lett.*, 82:2544-2547, doi: 10.1103/PhysRevLett.82.2544

Perdew, J.P., and Schmidt, K. (2001) Jacob's ladder of density functional approximations for the exchange-correlation energy. *AIP Conf. Proc.*, 577, 1, doi: 10.1063/1.1390175

Perdew, J.P., Staroverov, V.N., Tao, J., and Scuseria, G.E. (2008) Density functional with full exact exchange, balanced nonlocality of correlation, and constraint satisfaction. *Phys. Rev. A*, 78:052513, doi: 10.1103/PhysRevA.78.052513

Raaschou-Nielsen, O., Andersen, Z.J., Beelen, R., Samoli, E., Stafoggia, M., Weinmayr, G., Hoffmann, B., Fischer, P., Nieuwenhuijsen, M.J., Brunekreef, B., Xun, W.W., Katsouyanni, K., Dimakopoulou, K., Sommar, J., Forsberg, B., Modig, L., Oudin, A., Oftedal, B., Schwarze, P.E., Nafstad, P., De Faire, U., Pedersen, N.L., Östenson, C.-G., Fratiglioni, L., Penell, J., Korek, M., Pershagen, G., Eriksen, K.T., Sørensen, M., Tjønneland, A., Ellermann, T., Eeftens, M., Peeters, P.H., Meliefste, K., Wang, M., Bueno-de-Mesquita, B., Key, T.J., de Hoogh, K., Concin, H., Nagel,

G., Vilier, A., Grioni, S., Krogh, V., Tsai, M.-Y., Ricceri, F., Sacerdote, C., Galassi, C., Migliore, E., Ranzi, A., Cesaroni, G., Badaloni, C., Forastiere, F., Tamayo, I., Amiano, P., Dorronsoro, M., Trichopoulou, A., Bamia, C., Vineis, P. and Hoek, G. (2013) Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE). *The Lancet Oncology*, 14, 9:813-822, doi: 10.1016/S1470-2045(13)70279-1.

Raes, F. and Janssens, A. (1985) Ion-induced aerosol formation in a $H_2O - H_2SO_4$ system - I. Extension of the classical theory and search for experimental evidence. *J. Aerosol Sci.*, 16:217-227.

Raghavachari, K., Trucks, G.W., Pople, J.A., and Head-Gordon, M. (1989) A fifth-order perturbation comparison of electron correlation theories. *Chemical Physics Letters*, 6:479-483, doi: 10.1016/S0009-2614(89)87395-6.

Ruddiman, W.F., and Thomson, J.S. (2001) The case for human causes of increased atmospheric CH_4 over the last 5000 years. *Quaternary Science Reviews*, 20, 18:1769-1777, doi: 10.1016/S0277-3791(01)00067-1.

Sakurai, J.J. (1994) Modern quantum mechanics. Revised edition. Addison-Wesley Publishing Company.

Scuseria, G.E., and Lee, T.J. (1990) Comparison of coupled-cluster methods which include the effects of connected triple excitations. *J. Chem. Phys.*, 93:5851-5855, doi: 10.1063/1.459684.

Seinfeld, J.H., and Pandis, S.N. (2006) Atmospheric chemistry and physics: from air pollution to climate change, 2nd Edition. Wiley-Interscience.

Sherrill, C.D., and Schaefer, H.F. (1999) The configuration interaction method: advances in highly correlated approaches. *Adv. Quant. Chem.*, 34:143-269, doi: 10.1016/S0065-3276(08)60532-8.

Sipilä, M., Berndt, T., Petäjä, T., Brus, D., Vanhanen, J., Stratmann, F., Patokoski, J., Mauldin III, R. L., Hyvärinen, A.-P., Lihavainen, H. and Kulmala, M. (2010) The role of sulfuric acid in atmospheric nucleation. *Science*, 327, 5970:1243-1246, doi: 10.1126/science.1180315.

Slater, J.C. (1930) Atomic shielding constants. *Phys. Rev.*, 36:57-64, doi: 10.1103/PhysRev.36.57.

Slater, J.C. (1937) Wave functions in a periodic potential. *Phys. Rev.*, 51:846-851, doi: 10.1103/PhysRev.51.846.

Slowik, J. G., Stroud, C., Bottenheim, J. W., Brickell, P. C., Chang, R. Y.-W., Liggio, J., Makar, P. A., Martin, R. V., Moran, M. D., Shantz, N. C., Sjostedt, S. J., van Donkelaar, A., Vlasenko, A., Wiebe, H. A., Xia, A. G., Zhang, J., Leaitch, W. R. and Abbatt, J. P. D. (2010) Characterization of a large biogenic secondary organic aerosol event from eastern Canadian forests. *Atmos. Chem. Phys.*, 10:2825-2845, doi: 10.5194/acp-10-2825-2010.

Stevens, B., Giorgetta, M., Esch, M., Mauritsen, T., Crueger, T., Rast, S., Salzmann, M., Schmidt, H., Bader, J., Block, K., Brokopf, R., Fast, I., Kinne, S., Kornblueh, L., Lohmann, U., Pincus, R., Reichler, T., and Roeckner, E. (2013) Atmospheric component of the MPI-M Earth System Model: ECHAM6. *Journal of Advances in Modeling Earth Systems*, 5, 2:146-172, doi: 10.1002/jame.20015.

Sugita, Y., and Okamoto, Y. (1999) Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314:141-151, doi: 10.1016/S0009-2614(99)01123-9.

Tao, J.M., Perdew, J.P., Staroverov, V.N., and Scuseria, G.E. (2003) Climbing the density functional ladder: Nonempirical meta-generalized gradient approximation designed for molecules and solids. *Phys. Rev. Lett*, 91:146401, doi: 10.1103/PhysRevLett.91.146401.

Tsirigos, A., Haiminen, N., Bilal, E., and Utro, F. (2012) GenomicTools: a computational platform for developing high-throughput analytics in genomics. *Bioinformatics*, 28, 2:282-283, doi: 10.1093/bioinformatics/btr646.

Valentini, G.L., Lassonde, W., Khan, S.U., Min-Allah, N., Madani, S.A., Li, J., Zhang, L., Wang, L., Ghani, N., Kolodziej, J., Li, H., Zomaya, A.Y., Xu, C.-Z., Balaji, P., Vishnu, A., Pinel, F., Pecero, J.E., Kliazovich, D., and Bouvry, P. (2013) An overview of energy efficiency techniques in cluster computing systems. *Cluster Computing*, 16, 1:3-15, doi: 10.1007/s10586-011-0171-x.

van Duijneveldt, F.B., van Duijneveldt-van de Rijdt, J.G.C.M., and van Lenthe, J.H. (1994) State of the art in counterpoise theory. *Chem. Rev.*, 94, 7:1873-1885, doi: 10.1021/cr00031a007.

Vargas, R., Garza, J., Hay, B.P., and Dixon, D.A. (2002) Conformational study of the alanine dipeptide at the MP2 and DFT Levels. *J. Phys. Chem. A*, 106, 13:3213-3218, doi: 10.1021/jp013952f.

Volmer, M., and Weber, A. (1925) Keimbildung in übersättigten Gebilden. *Z. Phys. Chem.*, 119:277-301.

Vehkamäki, H. (2006) Classical nucleation theory in multicomponent systems. Springer.

Walter, C. (2005) Kryder's law. *Scientific American*, 293:32-33, doi: 10.1038/scientificamerican0805-32.

Public Health, Environmental and Social Determinants of Health (PHE). Available online at http://www.who.int/phe/health_topics/outdoorair/databases World Health Organization, Geneva.

Wilson, C. T. R. (1897) Condensation of water vapour in the presence of dust-free air and other gases. *Phil. Trans. R. Soc. A*, 189:265.

Winkler, P.M., Steiner, G., Vrtala, A., Vehkamäki, H., Noppel, M., Lehtinen, K.E.J. Reischl, G.P., Wagner, P.E. and Kulmala, M. (2008) Heterogeneous Nucleation Experiments Bridging the Scale from Molecular Ion Clusters to Nanoparticles. *Science*, 319:1374-1377, doi: 10.1126/science.1149034.

Woon, D.E., and Dunning, T.H. Jr. (1993) Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon. *J. Chem. Phys.*, 98, 2:1358-1371.

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

Yu, H. and Lee, S.-H. (2012) Chemical ionisation mass spectrometry for the measurement of atmospheric amines. *Environ. Chem.*, 9(3):190-201, doi:10.1071/EN12020.

Zeldovich, J. (1942) Theory of the formation of a new phase, cavitation. *Zh. Eksp. Theor. Fiz.*, 12:525-538.